University of Leicester

The application and development of relative survival methods in coronary heart disease

Christopher Paul Nelson

BSc (Hons), MSc

September 2008

Thesis submitted for the degree of Doctor of Philosophy

Centre for Biostatistics and Genetic Epidemiology

Department of Health Sciences

University of Leicester

The application and development of relative survival methods in coronary heart disease

Christopher Paul Nelson

Abstract

Relative survival is an estimate of net-survival without the need for cause-of-death information. This is achieved by matching the study cohort to the general population by various covariates, including age, sex and year of hospitalisation, in order to obtain an expected mortality rate. In this thesis relative survival methodology will be applied in heart disease where the form of the excess hazard rate is known to be very different from cancer, where this methodology originates.

The dataset presented is from the Leicester Royal Infirmary coronary care unit where all admissions to the unit were recorded between 1993 and 2006, which includes all patients in Leicestershire. Only patients who present with an ST-elevated acute myocardial infarction will be studied.

Relative survival is a new methodology in heart disease and this thesis will describe some of the problems that are encountered including the increased prevalence of the disease in the population and the very high early excess mortality rate that is not present in most cancers. Also investigated are period analysis models, which are also new to heart disease and allow the estimation of up-to-date information. An analysis of admission blood glucose levels and diabetic status is performed to examine the potential impact on patient prognosis in the short and long term, which involves the use of relative survival.

A new methodology is developed in this thesis for relative survival that fits spline based flexible parametric models on the log cumulative excess hazard scale. This methodology holds many advantages over current relative survival techniques due to the use of non-split-time data. This thesis demonstrates these advantages.

This thesis details how current relative survival methods have been extended to heart disease. A new model is developed, which is suitable in heart disease and cancer that fits flexible parametric spline based models.

Contents

Chapter 1

Introduction

1.1 Primary aims of the thesis

In this thesis the primary aim is to develop suitable models and move currently established relative survival techniques from cancer into heart disease. Survival from heart disease and survival from cancer share similar properties, including an extended follow-up time and a high proportion of deaths. Relative survival is a population based method that is used to estimate net survival without the need for cause of death information. Cause of death is known to be inaccurately recorded in CHD (1).

1.2 The application of relative survival in heart disease

The second chapter of this thesis will highlight the basics of relative survival and introduce a simple model which will be used to assess the relative survival of patients from the Leicester Royal Infirmary (LRI) coronary care unit (CCU) dataset that will be used throughout this thesis. Comparisons between relative survival and a standard Cox model will be made to highlight how estimates differ when compared to an all-cause approach. The basis of this chapter was published in the *European Heart Journal* (2).

Relative survival is a new methodology in heart disease and the third chapter of this thesis will review the literature in heart disease to determine if and where some or all of the relative survival methodology has been incorporated in heart disease. This could range from plotting an expected survival alongside an observed (or all-cause) survival rate up to fitting full relative survival models with an aim to estimate net survival.

Chapter four details an investigation of the models that are currently available in relative survival and more importantly how these fit to post acute myocardial infarction (AMI)

patients who are known to have a different hazard to most cancers. The potential problem is the time following initial hospitalisation as patients experience a very high mortality rate in the first 30 days following AMI. Of all heart disease deaths following AMI around 50% of these will be within six-months. The models fitted range from standard piecewise models up to the use of partitioned spline models on split-time data.

1.3 The development of relative survival methods

Part of the aims for this thesis are to develop the methodology of relative survival and this includes the development of flexible parametric models for relative survival using restricted cubic splines on non-split-time data. The development of this methodology and the motivation for it are discussed in the fifth chapter of this thesis. This method has been published in *Statistics in Medicine* (3) and combats some of the problems evident in the current methods. The chapter also discusses the limitations of the model and where future development could lie.

The model described in chapter five was implemented in Stata for use by other researchers and released via Stata SSC. How the method was implemented is detailed in chapter six, including how to run the command and obtain the relevant output. The aim was to allow all Stata users the option of using the package and so it was made fully generalisable and included a help file.

1.4 Assessing the assumptions in relative survival

There are several key assumptions in relative survival, with the main one revolving around the assumption that no-one in the general population, from which the expected survival rate is obtained, has the disease of interest. Chapter seven discusses how this is never true but, in cancer, generally assumed to be true as the prevalence of the specific disease is so low that it has very little impact on the results.

Heart disease is the biggest killer in the United Kingdom and as such the prevalence of the disease is larger than a specific cancer. Therefore the assumption may not be true and a correction may need to be made. The seventh chapter of this thesis investigates a simulation of patients using the standard expected survival rates obtained from routine data and an expected survival that excludes death from AMI. Analysing in this way will highlight the amount of bias that is experienced in practice when no adjustments are made.

1.4.1 Obtaining up-to-date estimates in heart disease

The eighth chapter investigates how period analysis can be used to obtain up-to-date estimates of patient survival, and can be used under a relative survival framework. Period analysis is also new to heart disease and the potential benefits that could be gained by using the methodology are investigated.

1.4.2 Application of methodology with genuine clinical endpoints

After the introduction of the methods and simple applications of relative survival the ninth chapter details results from a genuine research objective that investigated the impact of diabetes and glucose on short and long term patient survival. This involved using standard logistic models for the short-term analysis and relative survival models for long-term survival. This allowed the potential benefit of relative survival in practice to be assessed and allowed the new models described in chapter five to be used for a real research question.

1.4.3 Discussion

The final chapter of the thesis aims to tie together this research in the form of a discussion. This includes ideas for further research and the final conclusions that can be drawn from the thesis.

Chapter 2

Introduction to Relative Survival in Coronary Heart Disease

2.1 Introduction

The application and development of relative survival techniques in coronary heart disease encompasses both an applied and a theoretical approach in bringing a standard method of survival analysis from cancer to a new therapeutic area. A paper, published in the European Heart journal, that introduced these methods in heart disease (2) (shown in appendix D) forms the basis of this chapter.

This chapter aims to introduce relative survival methodology including modelling using piecewise models, as described in more detail in chapter 4. The discussion of current uses of relative survival in heart disease will be discussed in chapter 3 as this chapter aims to give a brief overview of an application of relative survival in heart disease.

2.1.1 Coronary Heart Disease

Coronary heart disease (CHD) is the leading cause of mortality in industrialised societies (4; 5), accounting for over 105,000 deaths in the UK in 2004 (6). Understanding both short and long term patient survival may help to inform improved management after presentation with CHD. Much of what is known about survival for patients with CHD comes from randomised controlled trials (RCT). Such trials usually recruit relatively selected populations followed over relatively short times, and the findings of RCTs may not easily be generalisable to the general population (7). Thus the assessment of patient survival in unselected clinical populations can be informative to the patient, the clinician, and in terms of future health-service provision. This leads towards the assessment of population-

based studies where the restrictive inclusion and exclusion criteria present in RCTs is greatly reduced providing a cohort of patients representative of the population.

Important prognostic factors related to survival can be evaluated along with comparisons over time and between centres. Assessing the survival of patients following an acute myocardial infarction (AMI) is important as AMI is the main contributing factor to the total number of deaths from coronary heart disease and therefore as one of the primary causes of death in the UK.

2.1.2 Measuring mortality in heart disease

To understand the natural history of a disease or condition of interest, and the influence of risk-factors and co-morbidity properly, it is essential to use appropriate statistical techniques. Ideally, the impact on mortality of a particular disease or condition would be measured by assessment of mortality specifically due to, or associated with, the disease of interest. However in broad population-based studies, cause of death is often difficult to establish with certainty.

Cause-specific survival considers as events only deaths that can be directly attributed to the disease of interest, with deaths from all other causes being censored. The probability of survival can be evaluated, whilst avoiding consideration of competing risks (8; 9). When considering cause-specific survival, it is usually of interest to fit statistical models to investigate simultaneously the influence of covariates, such as age and sex.

The main limitation of cause-specific survival is its dependence upon reliable coding of information on the cause of death. This reliance is not well founded, particularly when the source of information is the death certificate. For example a patient may be recorded as dying of renal failure where this is the result of cardiogenic shock following AMI. Three trained physician-adjudicators assessed cause of death in 2686 Framingham Heart Study participants (1). Of cases adjudicated as attributable to CHD, only 83.8% were recorded as such on the death certificates. Moreover, of cases recorded as due to CHD on the death certificate, this was confirmed in less than 70% by the physician-adjudicators. In this context, arguments have been presented that are critical of cause-specific survival. Lauer *et al.* (5) argue that data obtained from death certificates or from medical records are haphazard, biased and often inaccurate. These authors also suggest that all-cause mortality should be assessed as the primary end point, as it is both objective and unbiased. More recently, Mant *et al.* (10) argued that cause-specific methods are also flawed when attempting to differentiate among causes of cardiac death. These authors highlighted that this was the result of the inability of clinicians to agree on a cause of death; even where autopsy information was available, disagreement remained for a third of all deaths.

This leads to the most common survival analysis used in CHD research, one which does

not adjust for deaths due to other diseases unrelated to CHD. All-cause (or crude) survival includes all deaths within the cohort under investigation, and does not separate those due to the disease of interest from those due to other causes. All deaths are considered as events with only those surviving the follow-up period, or those lost to follow-up, being censored. Age adjustment is crucial in observational studies.

Clearly, in a cohort of patients with a given condition, some deaths will occur which are unrelated to the disease of interest. However all-cause analysis cannot disentangle deaths related to the disease of interest from deaths due to competing risks. A further, and important, limitation of all-cause survival methodology is its relative inability to disentangle the effect of strong covariates. For example, age is a strong confounder in most conditions. When adjusting for age at diagnosis in an all-cause model an adjustment is made for the combination of the impact of age on mortality associated with the disease of interest, and its impact on mortality from all other causes, the magnitudes of which may differ. However in relative survival the effect of age on other causes is treated separately to that on the disease of interest.

2.1.3 What can cancer share with CHD?

Analysis of survival in population based cancer studies often includes *relative survival*, used alongside, or instead of, crude and cause-specific methods (11; 12). Relative survival estimates the mortality rate for patients with the condition of interest after correcting for estimated mortality from all other causes. This methodology considers survival in patients with a specific malignancy compared to survival in a comparator population. Large scale examples can be found in the `EUROCARE` studies (13) and `CONCORD` studies (14). As some parallels exist between heart disease and cancer; for example, extended survival and follow-up is common to both disease areas, consideration of relative survival methodology may be worthwhile in CHD. An investigation of the practical application in CHD of relative survival methodology would be informative regarding the impact on outcome of CHD compared to what might be expected in the absence of that condition. In particular, relative survival allows for the use of appropriate statistical models to adjust adequately for confounders.

To date there are limited numbers of published relative survival applications and related methods in cardiovascular disease, most involving heart surgery. Of these, Norman *et al.* describes a review of the limitations of current methods that use expected survival and how many authors focus on observed survival (15). Others describe various relative survival statistical models applied in patients undergoing cardiac surgery (16; 17; 18; 19). Few such reports pertain to patients with non-surgical conditions, such as atrial fibrillation (20), MI (21) and stroke (22; 23). For the most part these analyses fail to model the influence of covariates and none fully utilises relative survival methodology. A full literature review

was performed and is discussed in chapter 3.

As an example a recent population-based study (24) investigated changes in colorectal cancer survival, using all-cause and relative survival methods. Using all-cause analysis, the analysis showed that for patients aged ≥75 years, for all stages of disease, 5-year survival was 41.4% in 1976-87 and 43.3% in 1988-99, suggesting improved survival between these periods. By comparison, the estimates of relative survival were much higher and, moreover, nearly identical for the two periods (68.0% and 67.9%), suggesting no improvement. This observation could be explained by improved survival in the background population, as well as in the cohort of interest and/or a change in the age distribution of patients with colorectal cancer. In other words, the apparent improvement over time in survival in the cancer cohort could represent improved survival in the general population, rather than cause-specific survival improvement in the cohort of interest.

2.2 Methodology

Modelling time to event data allows the investigation of a clearly defined endpoint. However when people are followed up over time in a study they can achieve one of two endpoints, survival or death. If a patient dies during the study then their survival time since their follow-up began is known. If a patient survives until the end of the study then it is necessary to censor this patient at the study end date, this is to ensure that the patients known survival time is recorded and that there is nothing known following this date. Similarly patients who are lost to follow-up, which could include patients who emigrate and die in another country, are treated as alive up to the last known contact at which point they become right-censored.

Most models for survival data are on the hazard scale where

$$h(t) = h_0(t)\exp(\beta x) \tag{2.1}$$

Where the total hazard $h(t)$ is made up from the baseline hazard $h_0(t)$ and modelled covariates βx. In a Cox model (25) the baseline, or underlying, hazard is arbitrary and not actually estimated. This is a proportional hazards model as the covariate effects do not depend on time. If the proportional hazards assumption holds (or can be assumed to hold) then it is possible to estimate the parameter effects without considering the baseline hazard function. The general proportional hazards model in equation 2.1 can be extended to relative survival where

$$h(t) = h^*(t) + h_0(t)\exp(\beta x) \tag{2.2}$$

where the addition of an expected mortality rate ($h^*(t)$) converts to modelling on the excess hazard scale described in more detail in section 2.2.2.

2.2.1 Defining the likelihood for survival data

The contribution to the likelihood for survival data for the i^{th} individual is given by

$$L_i = f(t_i)^{d_i} S(t_i)^{1-d_i} \tag{2.3}$$

Where d_i is an indicator of death. This is how censoring is dealt with, as d_i can be either zero (alive) or one (dead). Within this equation $f(t_i)$ is the probability density function and can be expressed in terms of the survival, $S(t)$, and hazard rate as

$$f(t_i) = h(t_i)S(t_i) \tag{2.4}$$

Where, $h(t_i)$ is the hazard function. Equation 2.3 can be rearranged to give

$$L_i = h(t_i)^{d_i} S(t_i) \tag{2.5}$$

The total likelihood is the product over the n individuals

$$L = \prod_{i=1}^{n} h(t_i)^{d_i} S(t_i) \tag{2.6}$$

It is generally easier to work on the log scale and this, an individuals contribution to the log likelihood from a standard all-cause survival is given as

$$\ln L_i = d_i \ln[h(t_i)] + \ln[S(t_i)] \tag{2.7}$$

$$\ln L = \sum_{i=1}^{n} \ln L_i \tag{2.8}$$

In a standard all-cause survival model the survival function, $S(t)$, can be expressed as a function of the cumulative hazard, $H(t)$ and is a useful relationship in survival data that will be used later on.

$$S(t) = \exp(-H(t)) \tag{2.9}$$

2.2.2 What is relative survival?

Total survival can be expressed by

$$S(t) = S^*(t)R(t) \tag{2.10}$$

Relative survival, $R(t)$, attempts to separate mortality from the disease of interest from mortality due to all other causes. To do this the ratio of the observed (all-cause) survival, $S(t)$, in the cohort of interest and the expected survival, $S^*(t)$, in a similar group in the general population is calculated (26) as shown in equation 2.11.

$$R(t) = \frac{S(t)}{S^*(t)} \tag{2.11}$$

The cohort of interest may comprise a sample of individuals with a specific diagnosis, for example acute myocardial infarction (AMI). The comparator group is obtained from routine data, matched to the cohort of interest by age, sex, period, or year of diagnosis, and other potentially important covariates. Deprivation has been used in cancer (12) to determine the effect of socio-economic status on mortality, which may prove to be interesting in heart disease. When modelling relative survival it is usual to convert 2.11 to the hazard scale. In a relative survival model, the observed mortality rate within the cohort of interest is made up of the background mortality rate in the general, comparator population (i.e. deaths due to other causes) plus the excess mortality rate associated with the condition of interest.

$$h(t) = h^*(t) + \lambda(t) \tag{2.12}$$

Where $h(t)$ is the total mortality made up from both the expected mortality rate, $h^*(t)$, and the excess mortality rate, $\lambda(t)$, due to the disease of interest. So when modelling relative survival the idea is to estimate the excess mortality experienced by patients diagnosed with the condition of interest directly by drawing a comparison to that experienced in the general population, thereby obtaining an estimate of net survival in the absence of reliable cause of death information.

An advantage of relative survival is that information on individual cause of death is not required; this removes the main problem associated with cause-specific mortality. However

the method assumes that deaths due to the disease of interest are independent of mortality in the general population (26), which will be returned to later in chapter 7. Relative survival yields excess hazard ratios, as opposed to the standard hazard ratios obtained in Cox and other survival models. Excess hazard ratios can also be used to estimate variability in the excess risk of death due to the disease of interest when, for example, comparing one demographic group to another, or temporal patterns of survival after AMI. As patterns of survival in AMI change, with increasing interest in long-term survival, the use of relative survival methods are expected to become increasingly important and relevant.

2.2.3 Relative survival modelling

In relative survival $R(t)$ the ratio of the observed survival rate, $S(t)$, defined in 2.9 with the expected survival rate, $S^*(t)$. There is a change in scale when assessing relative survival models and by converting this to the hazard scale the overall hazard, which is made up from two components, can be found as shown in equation 2.12. Using the total hazard given in equation 2.12 the log likelihood can be written as

$$\ln L_i = d_i \ln[h^*(t_i) + \lambda(t_i)] + \ln S^*(t_i) + \ln R(t_i) \tag{2.13}$$

Since $S^*(t_i)$ is constant it will not affect the maximum likelihood and as such can be ignored so an individuals contribution to the log likelihood can be shown as

$$\ln L_i = d_i \ln[h^*(t_i) + \lambda(t_i)] + \ln R(t_i) \tag{2.14}$$

This is for data fitted at an individual level, whereas most of the models fitted in practice are at the grouped level. This relationship was used by Esteve *et al.* (27) and the cure models of De Angelis *et al.* (28) and Lambert *et al.* (29).

2.2.4 Assumptions of relative survival

An assumption when using relative survival is that the matched population group used to obtain expected mortality is appropriate for the particular disease under study. The mortality rates should thus represent the expected mortality if the subject did not have the disease under study. A further issue is that when using national or regional life tables to obtain the expected mortality rates, the disease under study is included in these figures. If the prevalence of disease is low then this will have little impact on the estimates and for even the most common cancers it has been shown that this introduces negligible bias

(26). In older age groups AMI has a higher prevalence than most cancers and thus the prevalence issue needs to be considered. In subjects aged 75 years or older, there is negligible bias in the estimate of relative survival. However, the bias increases with age as the prevalence of AMI increases and the estimate of relative survival in the most elderly groups, for example those over 90 years old, are potentially biased. Fortunately, there is very little bias in excess hazard ratios when comparing groups as any bias will be in the same direction. A simulation has been performed in order to assess this assumption and will be discussed in chapter 7.

2.3 The Leicester Royal Infirmary Dataset

The dataset used for the analyses in this thesis was set-up in 1987 at the Leicester Royal Infirmary in the UK. The data that will be used are taken from 1993 onwards for consistency as there was a change in the database management as many more covariates were recorded from this time. All admissions to the coronary care unit (CCU) in Leicestershire, UK, are recorded and followed up to death.

Throughout this thesis models will be fitted on the CCU dataset from the LRI. Expected survival will be matched on year of diagnosis, age and sex to the government actuary department mortality statistics for England and Wales (30). There are two important decisions that have been made. After a heart attack the risk of almost immediate death is high, so this yields many deaths at day zero which were recoded to be given as day 0.1 i.e. it was assumed that on average the patients survived 0.1 days allowing for the probability that death occurred quickly following the heart attack. This was checked via a sensitivity analysis using 0.5 days which found no change in the observed results. This was performed in order to include these patients in the analysis as the Stata package, like all other statistical packages, removes these patients from the analysis when using survival time data. There was a total of 225 patients who died on day zero and were recoded in this way.

Acute myocardial infarction is described using characteristics of the electrocardiogram (ECG) and the results of blood tests. During the period of study, changes occurred in the diagnostic criteria for the definition of MI. These changes, based upon blood tests, led to many patients with lesser severity of cardiac injury being classified as *heart attack* toward the latter part of the observation period. Thus for consistency, only patients with electrocardiograph (ECG) evidence of ST elevation were included, a criterion which did not change over the study period. ST elevation is observed by an elevation between the S and T segments of an ECG as shown in figure 2.1. Moreover, ST elevated MI (STEMI) is associated with higher mortality. Thus, patients included are those experiencing their first known acute myocardial infarction (AMI) and demonstrating ST elevation on the ECG.

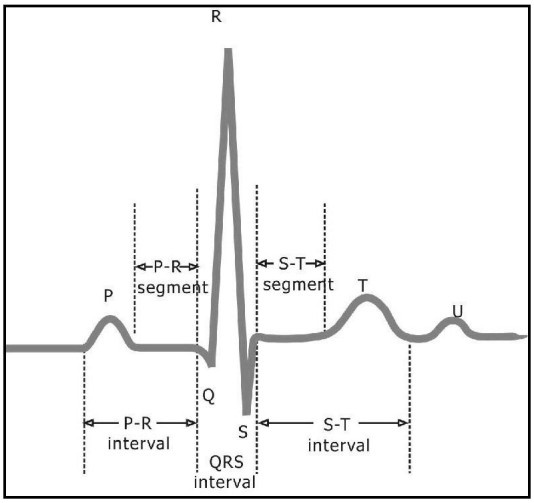


Figure 2.1: Electrocardiograph showing the S-T segment. *Obtained from National Instruments* (31)

A total of 5859 observations include 2383 deaths (40.7%) over the thirteen year period. Of these the cohort used for the analyses consists of 4748 observations as only patients who are experiencing their first known MI and have ST elevation are included. A large proportion, 1760 (37.1%), of this cohort died within the six year follow-up used in the investigation. A subject who is ST elevated is indicative of a blockage in their artery which can be categorised as a more severe heart attack. The cohort includes 3229 men (68.1%), 3965 of whom reported themselves to be White (83.6%) and 761 Asian (16.0%). The mean age of the cohort is 66.7 with men being eight years younger than women on average, (64.1 :Men, 72.2 :Women).

In comparison to a cancer registry a heart disease database is likely to be smaller in size. The databases are generally smaller because there is no national database and the dataset used here is only from one centre. This may change with the introduction of MINAP which will be discussed in section 10.6. However heart disease differs from cancer in a very high increased mortality rate in the first 30 days, with $40 - 50\%$ of all deaths occurring in this time period, there being 843 (47.9%) for this cohort, therefore models need to capture this.

Figure 2.2 shows a simple Kaplan Meier curve for the ST-elevated patients used in the following analyses. Follow-up is restricted to six years as shown and 95% confidence intervals are given. This is the baseline survival where no covariates are considered. There is no censoring due to loss of follow-up in this data. The figure shows that there is a large decrease in patient survival durng the first month or so of follow-up. After this has passed the survival rate decreases gradually over time at a fairly constant rate.

Figure 2.3 shows the underlying hazard rate with 95% confidence intervals. This shows that the hazard rate decreases very rapidly after the initially high hazard experienced

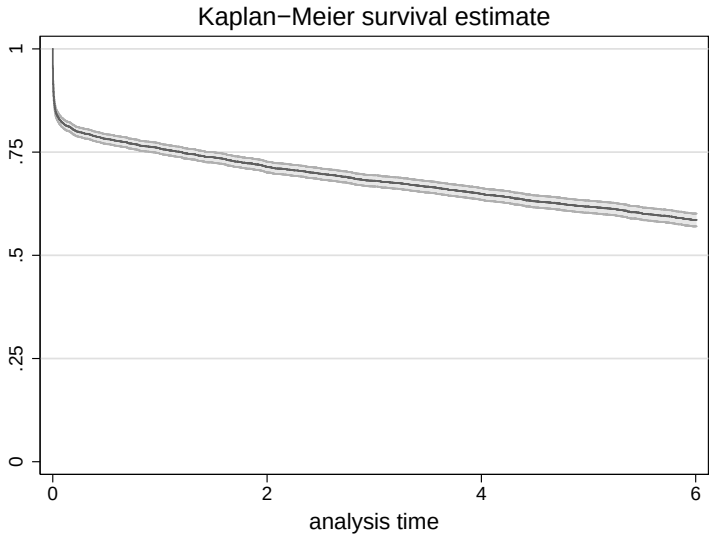


Figure 2.2: Kaplan-Meier curve for the LRI dataset

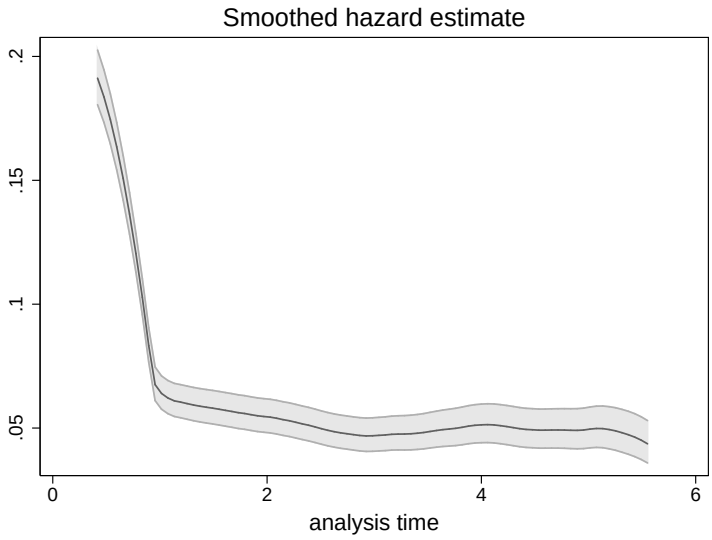


Figure 2.3: Hazard curve for the LRI dataset

immediately following hospitalisation. The hazard rate then remains fairly constant, not decreasing over time. These two figures show what is generally expected from a heart disease dataset. Note that the tails of the hazard function have been removed. This is the default in Stata as these are usually poorly estimated. The models developed in chapter 5 overcome this problem.

2.3.1 Expected mortality

The expected mortality rate is obtained from the Government Actuary Department (GAD) (30) for the England and Wales population. The data is matched on the age of the patient at hospitalisation, sex and year of hospitalisation. The simplest method of estimating the expected survival proportion is the Ederer I method (26) which is where the expected survival proportion is estimated as the average of the expected survival probabilities for every individual in the life table, even those individuals censored in the first intervals. Although the Ederer I method provides unbiased estimates of the expected survival proportion, its application, together with a potentially biased observed survival proportion, results in biased estimates of relative survival (32). This is because the Ederer I method does not allow for the potentially unequal lengths of patient follow-up times.

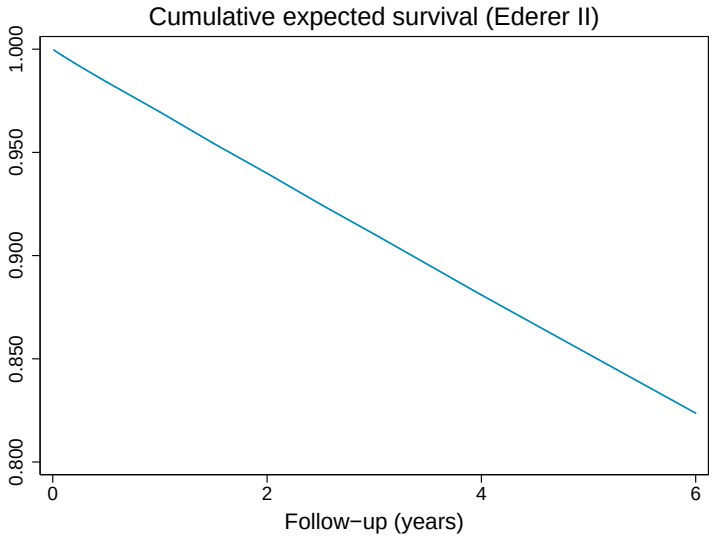


Figure 2.4: Expected mortality curve for the LRI dataset

Ederer and Heise (33) proposed an alternative approach which allows for heterogeneous follow-up times. That is, interval-specific expected survival proportions are estimated for each yearly interval, based on only those patients alive at the start of the interval. The cumulative expected survival proportion is then estimated as the product of the interval-specific survival proportions, however, the expected survival proportion is dependent on the observed mortality, again potentially leading to biased estimates of relative survival (32). The observed mortality in a given interval determines the patients who will form the basis for the calculation in the next interval. The cumulative expected survival proportion of the patients is dependent on the mortality of the disease in question during preceding intervals. Although the Ederer II method is not recommended for estimating long term cumulative expected survival proportions, it is a good estimator for the interval-specific expected survival proportion.

Figure 2.4 shows the Ederer II estimates which are used by piecewise analyses for the LRI dataset. Here it is clear that the expected survival decreases at a consistent rate over time. This does not account for any covariates and is the given expected survival rate for a six year follow-up for the entire cohort.

2.4 Simple example

2.4.1 Background

Data pertaining to patients admitted to the LRI CCU between 1993-2006 were selected as the cohort of interest, with a presentation of STEMI. The analyses performed on the dataset looked at the time from hospitalisation and were not conditional on 30 day survival as is the case with many other heart disease mortality analyses (34). The expected mortality was calculated using rates from the United Kingdom GAD matched on age and sex to the England and Wales population (30).

2.4.2 Methods used

In this chapter a simple analysis in order to illustrate the methods is adopted. This starts with a simple life table and compares the findings from all-cause and relative survival approaches, stratified by age groups, defined as $\leq$60 years, 60-75 years and >75 years old. The effect of age is then assessed using a relative survival model with proportional excess hazards (PEH), which is then compared to a standard (all-cause) Cox proportional hazards model. Non-proportional models for both the relative survival and Cox approach are investigated briefly, with a full investigation of non-PEH models coming later in chapter 4. Finally proportional models investigating the effect of sex, with and without adjustment for age groups, are performed for both relative survival and all-cause approaches. The variables that have been analysed here are also used as matching variables. However, it is important to note that other variables, for example glucose levels at admission, can easily be incorporated into the statistical model, as shown in chapter 9.

2.4.3 Lifetable results

Relative survival estimates can be presented in life tables similar to those for standard survival analyses but with emphasis on relative survival and evaluating survival probabilities. The life tables for each age group are shown in Table 2.1. In this table, in addition to the interval specific standard (observed) survival results, the expected [in *italics*] and interval-specific relative survival estimates [in **bold**] are also shown. If the observed number of

Age group	Start of interval	End of interval	No alive at the start of the interval n	Withdrawals in interval w	Deaths in interval d	Expected deaths in interval d^*	Observed interval survival p	Expected interval survival p^*	Relative survival of interval r	Cumulative observed survival cp	Cumulative expected survival cp_e2	Cumulative relative survival cr_e2
	0	0.0833	1373	5	88	*0.5*	0.936	*1.000*	**0.936**	0.936	*1.000*	**0.936**
	0.0833	0.5	1280	25	10	*2.4*	0.992	*0.998*	**0.994**	0.928	*0.998*	**0.931**
	0.5	1	1245	52	11	*2.8*	0.991	*0.998*	**0.993**	0.920	*0.995*	**0.924**
<60	1	2	1182	115	20	*5.7*	0.982	*0.995*	**0.987**	0.904	*0.990*	**0.912**
	2	3	1047	95	14	*5.5*	0.986	*0.994*	**0.992**	0.891	*0.985*	**0.905**
	3	4	938	88	18	*5.5*	0.980	*0.994*	**0.986**	0.873	*0.979*	**0.892**
	4	6	832	198	18	*9.8*	0.975	*0.986*	**0.989**	0.852	*0.966*	**0.882**
	0	0.0833	1938	4	274	*2.9*	0.858	*0.998*	**0.860**	0.858	*0.998*	**0.860**
	0.0833	0.5	1660	32	66	*13.6*	0.960	*0.992*	**0.968**	0.824	*0.990*	**0.832**
	0.5	1	1562	36	39	*15.5*	0.975	*0.990*	**0.985**	0.803	*0.980*	**0.820**
60-75	1	2	1487	112	73	*31.2*	0.949	*0.978*	**0.970**	0.762	*0.958*	**0.795**
	2	3	1302	120	45	*30.0*	0.964	*0.976*	**0.988**	0.735	*0.935*	**0.786**
	3	4	1137	113	42	*28.9*	0.961	*0.973*	**0.988**	0.706	*0.910*	**0.776**
	4	6	982	233	79	*48.9*	0.909	*0.942*	**0.965**	0.642	*0.857*	**0.749**
	0	0.0833	1436	0	481	*7.6*	0.665	*0.993*	**0.670**	0.665	*0.993*	**0.670**
	0.0833	0.5	955	18	115	*31.0*	0.878	*0.964*	**0.911**	0.584	*0.957*	**0.610**
	0.5	1	822	19	55	*32.6*	0.932	*0.959*	**0.972**	0.545	*0.918*	**0.593**
75+	1	2	748	63	101	*60.6*	0.859	*0.913*	**0.941**	0.468	*0.839*	**0.558**
	2	3	584	49	71	*51.7*	0.873	*0.907*	**0.963**	0.409	*0.761*	**0.537**
	3	4	464	66	52	*42.9*	0.879	*0.899*	**0.978**	0.359	*0.684*	**0.525**
	4	6	346	74	87	*59.4*	0.718	*0.799*	**0.899**	0.258	*0.546*	**0.472**

Table 2.1: Relative survival Lifetables by age group. 0.8333 is equal to one month.

deaths is equal to the expected number of deaths during an interval then the probability of survival for patients with the condition of interest is the same as that in the general population. Investigation of the ≤60 years old age group shows the observed number of deaths dropping from 88 in the first month to just 14 in year 2-3 of follow-up. However the expected numbers of deaths by 30 days and during year 2-3 of follow up are 0.5 and 5.8 respectively. Thus, acute ST elevation MI is associated with almost all excess deaths in the 30 day period following the event, and with an excess of less than 10 deaths in year 2-3.

Clearly, AMI is associated with a marked excess in mortality in the immediate, post AMI period. However, how much of the later mortality excess, and indeed of overall mortality in the cohort, can be ascribed to the index AMI? Consideration of the cumulative survival data illustrates the information to be gained from relative survival analysis. By the end of follow-up, an all cause approach shows cumulative survival of 0.258 for the >75 years old group. However taking expected mortality into account, the cumulative relative survival rate is 0.460. In other words, if the expected, or background, mortality in the population is removed then the the survival proportion is over 20% lower in absolute terms, as mortality due to other causes is included in the all-cause analysis.

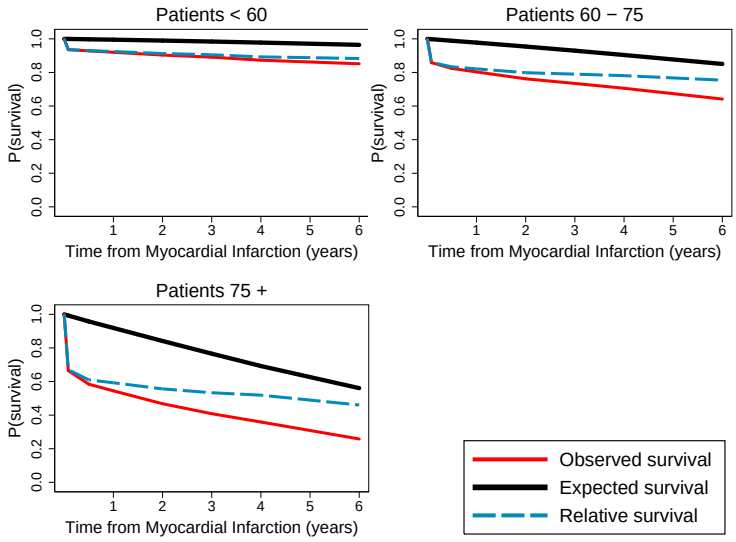


Figure 2.5: The Observed, Expected and Relative survival split by age groups

Using the table the expected survival is shown to decrease with age. By 4-6 years of follow-up, the youngest age group had a cumulative expected survival rate of 0.964 and the oldest age group 0.561. The relative survival rate is strictly a ratio, but is often called a rate. This can be found as the ratio of the observed survival and expected survival shown in the table. Note how cumulative relative survival is greater compared to cumulative observed survival at all times, for all ages. Also worthy of note is that the interval specific relative survival is lowest during the first thirty days after AMI, in spite of this being the

shortest time period considered in the analysis. The same is true for absolute survival (not shown).

Lifetable information is often easier to interpret in graphical form. Figure 2.5 shows the relative survival curve, along with the expected and observed (all cause) survival probabilities split by age groups, (under 60 years old, 60-75 and >75 years). The figure highlights the difference between a relative survival and an all cause survival approach. The observed values shown on the figure represent all of the deaths in the cohort, i.e. the survival proportion that would be obtained if an all-cause approach was assumed. The expected survival is the survival proportion that would be expected if the cohort was in the general population without experiencing the index AMI, and the relative survival is the ratio of these two curves.

The initial drop in survival is smallest for the youngest age group. Moreover, relative and observed survival are very similar in this age group, as expected survival is close to 1.0. As background mortality is very low in this age group, nearly all deaths experienced within this cohort are likely to be attributable to the index AMI. In the other age groups, particularly the eldest, there is also very little difference between the observed and relative survival rates for the first year post MI. However there is a relatively high expected mortality rate in this oldest age group; some of the deaths in this population are not due to the index AMI. Later, the relative survival curve begins to decrease at a slower rate, although never flattening completely. If the relative survival curve flattens out completely then the mortality rate of these surviving individuals is the same as expected in the general population. When this effect occurs it is known as statistical or population cure (28; 29; 35).

2.4.4 Modelling

The influence of cofactors or comorbidity on relative survival can be obtained by modelling, and expressed as excess hazard ratios; their interpretation is straightforward. For example, an excess hazard ratio of 2 for males would suggest that the excess mortality rate in men (i.e. deaths associated to the disease of interest) is twice as high as in women. There are many models available in relative survival as detailed in chapter 4, including piecewise models (section 4.3), polynomials (section 4.5), fractional polynomials (section 4.6), and splines (section 4.8). A new relative survival model is developed in chapter 5. Models are also used in chapter 9, where a full analysis of the dataset is presented.

Table 2.2 illustrates the impact of age on hazard ratios estimated from a Cox proportional hazards model and on excess hazards ratios from a relative survival model. Using <60 years as the comparator group, for both 60-75 and >75 age groups, estimates of the excess hazard ratios are lower with relative survival modelling than with the hazard ratios from

Age group	Hazard ratios (95% CI)	
	Cox Model	Relative survival model
< 60	1.000	1.000
60-75	2.713 (2.30 , 3.20)	2.426 (1.97 , 2.99)
75+	7.842 (6.68 , 9.21)	6.919 (5.67 , 8.44)

Table 2.2: Models comparing the hazard ratios and excess hazard ratios from a Cox proportional hazards model and a relative survival model with proportional excess hazards (with 95% confidence intervals).

the Cox model. Once again this illustrates the increasing influence of competing risks in a population as that population ages.

The effect of age can also be assessed by the excess mortality rate. During the first month the excess mortality rate ranges from 794, in the under sixty age group, to 5490, in the over seventy-five age group, excess deaths per thousand person years. It is therefore possible to conclude that during the first month there is a huge increased risk of death associated with the index event. In the final interval the youngest age group excess mortality rate is 6.57 and the oldest age group is 45.47 per thousand person years.

2.4.5 Modelling time dependent covariates

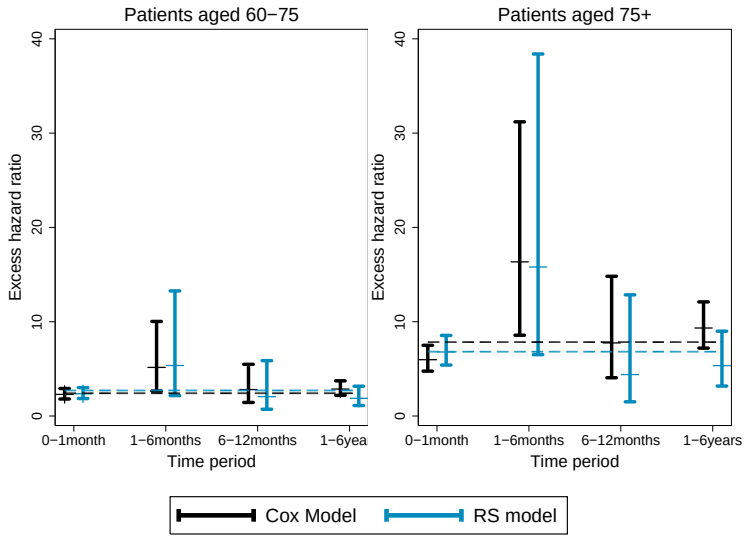


Figure 2.6: Time-dependent effects. Excess hazard ratio at intervals after index STEMI. The graphs illustrate excess risk of death for patients aged $60-75$ or >75 years, each compared with patients aged <60 years. Vertical lines show the 95% confidence limits and point estimates of the excess hazard ratios from a relative survival model and hazard ratios from a Cox model. Horizontal lines show the point estimate of the excess hazard rate ratio from a relative survival model shown in Table 2 and PH mortality rate ratio from a Cox model.

In their simplest form the models assume the impact on mortality to be proportional over

time. This means that the relative effect of a covariate is assumed to be the same at 1 week as at 1 year and so on. As with Cox models, it is possible to investigate whether the excess hazard ratio changes over time (Figure 2.6). For the 6-12 month follow-up interval, and using the all-cause method, it would be estimated that patients aged 60-75 are at approximately 3-fold greater risk of death, and patients aged >75 are near 8-fold greater risk, compared to a patient aged <60. In comparison, the relative survival method estimates of excess risk are rather lower, at 2-fold for patients aged 60-75, and 4-fold for those aged >75 years. Over time the subjects in the cohort are ageing and the expected mortality rate is also increasing with age. Relative survival methodology takes account of this, giving a more realistic estimate of the mortality associated with the previous AMI. This is shown for illustrative purposes even though a likelihood ratio test proved non-significant for relative survival ($p = 0.3571$) suggesting that the proportional model is more appropriate, although the Cox model did significantly improve ($p = 0.0199$).

2.4.6 Adjusting for more than one covariate

An unadjusted Cox model found that male gender was associated with a hazard ratio of 0.60 (0.55 , 0.66), suggesting a 40% lower risk of death for males in the cohort. Using the relative survival estimate, which allows for the fact that the females (on average older) are more likely to die of other causes, the excess hazard ratio is 0.52 (0.45 , 0.58) for males, a 48% lower risk compared to females. As with Cox models it is possible to adjust for multiple covariates, and for illustration both sex and age were considered in the models. The adjusted Cox estimate for males was 0.90 (0.82 , 1.00), indicating a 10% lower risk of mortality after AMI. However, as shown, this methodology fails to separate death due to the index AMI from deaths due to other causes. Females in this cohort are older (mean 72.2) than males (mean of 64.1) and are thus more likely to die from causes other than the AMI. The adjusted relative survival model gives an estimate of 0.78 (0.69 , 0.88) for males, suggesting a 22% lower excess mortality due to CHD for males compared to females post MI, after adjustment for age. The advantage of this model is that it adjusts for the disease associated mortality associated with age separately from the expected mortality experienced in the general population.

2.4.7 Conclusion

Relative survival provides clinically relevant information. Investigation of excess mortality rates has potential implications regarding patients' long term survival. For example while much resource is dedicated to the management of *young* patients with CHD, the initial estimates from the proportional excess hazards model suggest that for patients aged <60 years, the excess risk due to the index MI is around 30 additional deaths per 1000 person

years over the first 6 months, but as few as 6 excess deaths estimated per 1000 person years after 3 years.

2.5 Limitations

Relative survival can be interpreted as a measure of mortality due to the disease of interest only if deaths due to the disease of interest are independent of the mortality in the general population. Even if this is not the case, relative survival still provides a useful comparison with mortality in the comparator population, which is usually the general population (36). It may be worth considering alternative comparator groups when using relative survival methods in CHD to derive the expected mortality rates. For example, patients with CHD have a high prevalence of co-morbidities and of risk-factors which put them at risk of mortality from causes other than CHD, such as pulmonary disease. Expected mortality from the general population has been adopted, but use of expected mortality from a population with similar co-morbidities could be selected. For example, to assess survival after AMI in patients treated with insulin during the index admission, one could consider matching with expected survival by diabetic status. However, obtaining reliable information from such a population is often difficult.

A potentially important issue in the use of relative survival to the assessment of CHD survival, is that in using population life tables to derive the expected mortality rates, deaths due to the condition of interest are included. If the prevalence of that condition in the background population is low enough, then this will have little impact, a reasonable assumption for individual malignancies (26). However, given the predominant contribution of heart disease to mortality in industrialised society, the appropriateness of this assumption in CHD needs to be assessed, in particular for oldest age groups.

Some of the simpler statistical models for relative survival have been presented. However, there are several extensions that would also be applicable to CHD. For example, modelling time-dependent effects on a continuous scale through the use of splines (37; 38; 3) and fractional polynomials (39).

Relative survival methodology merits attention in observational and population-based assessments of CHD mortality. Application of relative survival methods is still in early development in heart disease but its application in cancer is common, and informative. In order to obtain an estimate of net survival in population based studies, relative survival applies sensible assumptions based on the deaths that would be expected to occur in a cohort of patients if they were from the general population. One development relevant in CHD may be calculating expected mortality in groups that are at a potentially higher risk of mortality from CHD, such as ethnic minorities, or social deprivation cohorts.

2.6 Discussion

There is a very high excess mortality experienced by these patients in the first week or so, which is expected following an acute ST-elevated myocardial infarction, but this may cause some issues with the various models. The piecewise model used here may be sensitive to the choice of time split points and needs investigation. The models will be explored fully in chapter 4.

Due to the high mortality a continuous function may be more appropriate and provide a better fit. However this also needs to provide flexibility in order to capture the very high excess hazard rate early on in the follow-up and one such model is proposed in chapter 5. This allows the analysis of continuous covariates which has been assessed in practice, as shown in chapter 9.

The survival curves shown in this chapter and results includes data from thirteen years of follow-up. Within this thirteen year period many improvements have been made, both in medical treatments and in clinical care and for this reason it may be more informative to obtain up-to-date estimates of patient survival using period analysis. This is explored in chapter 8.

Relative survival is new to heart disease and the use of the methods in published research is explored in chapter 3. Part of the reason it may be avoided in research is due to some necessary assumptions. The key assumption is that the expected mortality rate excludes the disease of interest and this is not true. In fact it is never true, even in cancer, but it has a negligible effect as long as the prevalence of the disease is low, which will be explored in chapter 7.

Chapter 3

The Use of Relative Survival in Published Heart Disease Research

3.1 Introduction

Relative survival has been used routinely in cancer studies for many years with the help of large cancer registries and the development of software using common statistical packages. The main aim of this chapter is to determine the use of the methodology outside of cancer studies, heart disease in particular, to determine what elements of relative survival have been adopted.

Relative survival has been used regularly for many years with many models being developed for use in statistical packages since 1987 in the now dated GLIM package (40) up to implementing splines in SAS (41; 42), R (38) and Stata (3) twenty years later. However, rather than review all of the literature that is available to do with relative survival modelling in general, as the models will be investigated in a following chapter, it would be more interesting to identify where relative survival has been adopted in heart disease.

3.2 Search Strategy

There are several medical research databases that could be used to search for papers, and for this literature review Medline was the primary database used with further less complex searches using various others, such as web of science. To determine useful search terms the Cochrane library was investigated for a list of terms that could be used as a starting point. Papers that had been collected up to the starting point of the literature review were also used to determine any further search terms. In medline all search terms were fully *exploded* to include all sub categories and specific search terms were allowed to be

within five words of each other. Searches were limited to only include English language and human studies and various truncations were used, i.e. cardi$, to allow the search for all terms that begin with the truncated word.

The terms used for the relative survival part of the search were:

1. Survival analysis
2. Survival rate
3. Excess hazard/mortality
4. Background hazard/mortality
5. Expected hazard/mortality
6. Statistical cure
7. Relative survival

It is important to note that in order to obtain papers that have a higher chance of being related to relative survival a series of conditions were placed on these search terms. Either of term 1 or term 2 were allowed and were combined with any of terms 3-7, thus ensuring that the paper used survival analyses *or* survival rates *and* included a term relevant to relative survival. These search words were selected as they cover all conceivable phraseology that an author may use when discussing an approach to survival that compared their findings with another group. Survival analysis and survival rate were two headings that produced half of the resulting papers each. Also as noted earlier these words were allowed to be up to five words apart allowing phrases that could potentially incorporate these terms to be included, naturally this also reported phrases that were not related to relative survival also.

The heart disease terms used in the search were much simpler due to being able to *explode* the following terms to include all relevant sub-groups:

- Heart disease$
- Coronary disease$
- Cardio$
- Cardiac$

These terms covered a very wide area and included searching for the terms exactly and using the given spelling in case the paper was not listed in the relevant category. These terms as stated included all categorised fields of Heart disease and Coronary disease known to the Medline database and as such no other terms were necessary.

3.2.1 Search results

Using these terms two hundred and fifty-four papers were found using Medline alone, of which all abstracts were read. Of these any papers which were clearly unrelated to relative survival were excluded, examples of the most common phrases include:

- '...mortality was higher than expected.', but with no indication the authors had used an expected group.
- Some abstracts labelled the introduction as background creating '*Background*: Mortality is known to be...'.
- Excess mortality was often used as a phrase when describing the effect of a covariate to another rather than in terms of relative survival.

There were also several papers which did not appear to be related to heart disease in any way and many were discounted for this reason, especially where the outcome was for cancer survival. Other papers were investigated in other disease areas if the outcome in an MI group was a possibility. After these exclusions and after other smaller searches had been accomplished there were found to be seventy-eight papers that required further reading beyond the abstract.

3.3 Summary of papers

3.3.1 Criteria

There are six key areas that warranted investigation of the papers, they were:

- What comparison group was used
- Was the ratio of observed/expected survival taken
- How was the observed survival modelled
- What was plotted
- Was relative survival modelled
- What method was used for modelling relative survival

Other features were also noted including any important statistical tests, therapeutic area and other potentially important factors associated with the use of relative survival methods.

The key areas are important as they start with the very simplest of relative survival approaches in the comparison to an expected survival/hazard group, moving onto whether the ratio of observed to expected survival was taken, in effect the relative survival being calculated even if not labelled as such. The modelling of the observed survival was deemed important as it was found that for the majority of studies all modelling, if performed, was only done at an all-cause level. If relative survival was modelled this was very important information that needed to be recorded along with the methodology used. The plotted results are also very important, papers may show only the observed survival or even the observed and expected survival which is indicative of a direct comparison being made.

3.3.2 Initial results

Paper methodology		Total
Question	Response	Frequency
Was a Comparison group used	Yes	69
	No	8
Was the Ratio of O/E calculated	Yes	48
	No	29
How was observed survival modelled	Cox	35
	Logistic	4
	Cox and Logistic	11
	Not modelled	27
Was Relative Survival modelled	Yes	4
	No	73
What survival rates were plotted†	None plotted	22
	O	11
	R	3
	O and E	25
	O and R	5
	O and E and R	9
	Other	2
What was the Therapeutic area of the research	Heart surgery	39
	AMI	7
	Stroke	5
	Diabetes	4
	Other heart disease	16
	Other	7

Table 3.1: Table of frequency counts for the relevant criteria that was investigated. †Where O=observed, E=expected and R=Relative; survival

The papers found are listed in the bibliography with references from 15 to 23, and 43 to 111. The results shown in table 3.1 are the frequencies for the papers found for review, these papers were published between 1990 and 2008. There were only eight papers without a comparison group, meaning 69 papers were of interest as they potentially compare an observed survival in their cohort with that of an expected group, which could be considered the first stages of relative survival. Note that one paper was a review and as such only 77 papers could be compared in this way. Surprisingly the ratio of the observed to the

expected was taken in the majority of these papers ($n = 48$), which is 69.6% of those that had a comparative group.

Of the 77 papers only 27 did not model at all with the rest modelling the observed survival either via logistic regression on short term follow-up, generally defined as between 30 days and one year, Cox proportional hazards modelling on long term follow-up or a combination of both. As is also shown only four papers attempted to model relative survival in any way, meaning that the vast majority of approaches that used an expected survival group still only investigated the effect of covariates in an all-cause approach, commonly with a Cox model.

Many of the papers used the expected survival as a comparator in a figure to the observed survival (25 papers) while 17 papers plotted the relative survival either on its own, with the observed survival or both the observed and expected. It should be noted that in no instances were the observed, expected and relative survival estimates plotted on the same figure. Of those listed as other these include one which plotted the standardised mortality rate and another plotted the excess mortality. A common feature of those that plotted the relative survival was that the interval specific relative survival was plotted rather than the cumulative relative survival. This means that the relative survival plots appeared very different to the standard survival plots as the curve was allowed to increase over time.

Finally the most common therapeutic area found was in heart surgery with half of the papers being in this field. Those following acute myocardial infarction (AMI) were the second most common result with seven papers. Stroke and diabetes were also recurring with many papers being from various other fields either in heart disease or other, ranging from specific diseases such as Alzheimers and asthma to specific heart defects and deaths in general.

3.3.3 SMR

In population based research the standardised mortality ratio, or SMR, is defined as the ratio of observed deaths to expected deaths where the observed deaths are obtained from a cohort of patients under study and the expected is from the general population or equivalent. The calculation used to determine the SMR is very simple

$$\frac{O_d}{E_d} \tag{3.1}$$

where O_d is the observed number of deaths and E_d is the expected number of deaths. This is similar to relative survival which is the ratio of the observed and expected survival rates. One of the key differences is that an SMR, when modelled, has a multiplicative

effect on the hazard scale whereas with relative survival, and more importantly the excess mortality rate, the effect is additive on the hazard scale and also has useful properties on the survival scale.

$$SMR: \ h(t) = h^*(t)\exp(\beta x) \tag{3.2}$$

where $h^*(t)$ is the expected hazard and βx represents that which is estimated from the model. Whereas excess mortality $\lambda(t)$ is modelled as follows.

$$\lambda(t): \ h(t) = h^*(t) + \exp(\beta x) \tag{3.3}$$

It is important to make this differentiation between an SMR and an excess mortality rate. There is an interest in CHD in comparison to the general population via an SMR and these were not used in the search terms as these were not of direct interest and would have provided several more papers that would have provided little relevant information.

3.4 In-depth review of key papers

The seventy-eight papers found will be discussed in more depth where necessary in order to discuss any key features in each paper. The first point would be the group to which these papers compare, which in many cases reveals the geographical location of the study.

3.4.1 Comparison groups used

There were many examples of different comparison groups used in the papers found, with the most common being the general population of Sweden (16; 17; 18; 19; 43; 44; 45; 46; 47; 48) with another in Sweden but only the Vasterbötten area of northern Sweden (49). Six of these papers came from the same group in the department of thoracic and cardiovascular surgery, university hospital in Uppsala, Sweden, i.e. Stahle *et al.* A lot of papers came from Norway and the general population was used as a comparison group for nine studies (50; 51; 52; 53; 54; 55; 56; 57; 58), although they all appear to be from the same group.

Studies from North America used data from several different groups, some used only the general white US population (59; 60; 61; 62), whereas others used the general US census data (63; 64; 65). Many studies used Minnesota residents (20; 66; 67; 68; 69; 70; 71), quite often restricted to a white population. One used Washington state (72) and another the Canadian insurance alliance (73).

The Danish general population was also used (74; 75; 76; 77; 78; 79) and a suspected further two by Bronnum-Hansen *et al.* (80; 21) which were from Denmark but only stated

the general population rather than the country specifics. The Finnish (81; 82; 83), the Dutch (84; 85) and the German (86), general population were used in other studies. Other specific areas were Norfolk (87), Rio de Janerio (88) and the Somme region of France (89). Three papers, all after stroke, used the general population of western Australia (22; 23; 90).

Only seven of the papers did not use any comparison group at all (91; 92; 93; 94; 95; 96; 97). Two papers did not state what comparison group was used (98; 99). This leaves only thirteen papers which did something slightly different.

One paper used the expected values that reflected preoperative risk factors but gave no further detail as to how this was calculated or obtained (100). In a study of asthma and mortality the expected survival group used was simply the patients survival of those without asthma (101). Life expectancy at birth was used in one study but it was not clear how this information was gathered or incorporated (102). However this paper was investigating a DNA abnormality so may be a sensible methodology.

Papadimos *et al.* (103) calculated the ratio of O/E but it is not clear how the expected values were calculated as there is simply a reference to another paper stating that the latest STS CABG risk model was used in the calculation. There is a suggestion that the expected values were calculated for patients with the disease in order to make a comparison with the observed results. Similarly three other papers calculated an expected mortality for patients with the disease (104; 105; 106), however this comparison is of no interest in relative survival.

The coefficients from the multivariable risk factor equation for hospital deaths after PTCA (percutaneous transluminal coronary angioplasty) in New York State were used to calculate a risk score and corresponding probability of death for each procedure in one study (107). The probabilities were then summed to obtain an expected number of deaths for the population, based on this model. Finally three of the papers found were not even survival analysis (108; 109; 110).

The expected stroke rate in a study by Witt *et al.* (111) was calculated by applying the sex, age, and period-specific stroke rates in the general Rochester population to the person-time follow-up of the study population. This is different as it does not use expected mortality rates rather the expected stroke rate.

Finally, there was a review paper by Norman *et al.* (15) that investigated several heart surgery papers therefore not specifically calculating a comparative group themselves. However the authors did comment on how the majority of papers only investigated observed mortality even if they had used an expected mortality group.

It is worth noting that the majority of authors obtained expected mortality rates from an appropriate group that was relevant to their own country or area.

3.4.2 The AMI papers

The main reason for the review was to investigate the use of relative survival in coronary heart disease to determine if the methods had been investigated in any way in this field of interest, particularly in those which investigated the effects after a heart attack, or AMI. As stated earlier seven papers were found using the search methods described.

Two of these were not relevant as in the paper by Beattie *et al.* (94) where the term excess mortality was used where the authors were simply referring to an increase in mortality. The paper by Van de Werf *et al.* (106) also predicted results from a model and compared these to those that were observed to determine the model suitability rather than to adjust for the number of deaths due to other causes.

Laundberg *et al.* (76; 75) provided two of the papers looking at survival post AMI as an outcome measure, both papers using the general Danish population for the expected survival. One of these papers did no modelling at all and only assessed the results using the Actuarial method (76). The second paper did model but only via an all-cause approach (75). Both papers from these sets of authors did calculate the ratio of observed to expected survival in the form of a standardised mortality ratio (SMR). Both papers were also from the same study, one being a three-year follow-up and the other a ten-year follow-up. Observed and Expected survival were plotted alongside each other as a comparison but no adjustment for other covariates was performed.

In a paper investigating reasons for death after admission to an intensive care unit Niskanen *et al.* (82) investigated cardiac arrest and cardiovascular disease as separate reasons for the initial hospitalisation. For a comparative group the general population of Finland was used and relative survival was calculated and plotted along with the observed and expected survival. The only modelling that was performed was using a Cox model. SMRs were defined as O/E and relative survival as the Actuarial value/E, which was a different way of reporting the relative survival. A standard lifetable, known as the actuarial method, was primarily used.

Bronnum-Hansen *et al.* (21) whose work recurred in the search as part of the Danish MONICA study group which relates to, *MONItoring of trends and determinants in CArdiovascular diseases*. More specifically these authors assess survival following Stroke in the other papers referenced here. This would explain why the the expected number of deaths in the general population were estimated for each gender by calculating the age-time-specific person-years of observation multiplied by the similar age-time-specific population death rate from the Danish MONICA population. Confidence intervals were determined by Poisson regression but no relative survival modelling was performed. The authors plotted the observed and expected survival and calculated SMRs and excess death rates EDRs, which are the same as excess mortality rates per 1000 person years.

The final paper (105) found following AMI did not plot the expected survival but did calculate the ratio of observed and expected mortality. The authors developed and externally validated a risk model with a demonstrable clinical utility in stratifying patients after myocardial infarction. The model calculates an individual's probability of death at 30 days on the basis of three robust admission characteristics (age, heart rate [hr], and systolic blood pressure). The calculation is given by:

$$P_{30} = \frac{1}{1 + \exp(-L_{30})} \tag{3.4}$$

where

$$L_{30} = \beta_0 + \beta_1(age) + \beta_2(hr)\beta_3(sbp) \tag{3.5}$$

This was used to calculate SMRs by taking the ratio of observed mortality over expected mortality but no modelling was performed. This was an expected survival that was used for prediction and was calculated for patients with the disease rather than the general population and is therefore of no interest as this is not relative survival.

3.4.3 Common statistical tests

A large number of papers that investigated the ratio of observed to expected calculated a standardised mortality ratio (SMR). This is calculated as the quotient of the observed to the expected numbers of deaths per 1000 person-years. Some authors plotted these results which was in effect the relative survival per 1000 person-years. Similarly the excess death rate (EDR) was also calculated in some of these papers which is equal to the observed minus expected number of deaths per 1000 person-years, i.e. this is the same as an excess mortality rate per 1000 person-years. In some of the papers, for example the paper by Aune *et al.* (51), the SMRs were compared between time intervals, for example the difference ($d = a - b$) in SMRs at 1 year compared to the SMR at 5 years would be used to show a change in relative survival (of d).

In other papers an SMR was called a Mortality ratio (MR), which was calculated as $100(O/E)$ and an EDR was called a mean annual mortality rate (MAMR) but was calculated in the same way.

The log rank test was used in around 10 papers which compares one survival rate to another over the follow-up period. In this case the observed survival was compared to the expected survival and a p-value was obtained and used as evidence that mortality was in excess of what was expected. The ratio of O/E was calculated in many papers but some

authors calculated 95% confidence intervals using Poisson distribution tables, for example a paper by Hankey *et al.* (23). The confidence intervals were also calculated using a normal approximation to the Binomial distribution in a paper by Rihal *et al.* (107) when comparing the observed and expected survival. This paper also used the expected survival rate as a covariate in a logistic model for survival.

The actuarial assumption was used in the majority of relative survival analyses as relative survival was only considered using a lifetable approach in all but four papers. There were other tests performed to assess the divergent survival rates from the expected by using a Mantel-Haenszel's test (112). The differences in relative survival between subgroups were evaluated with a proportional hazards test based on a multiplicative model. This method was originally proposed by Norman Breslow (113) but was later adopted by Buckley (114) in a relative survival framework.

3.4.4 Relative survival models used

There were four papers found that modelled the ratio of observed and expected in some way. The first paper, by Kvidal *et al.* (16) investigates survival after aortic valve replacement in Sweden which was published in the year 2000. Logistic and Cox proportional hazards models were used and the observed survival was plotted with the expected survival and with the relative survival. The method used for modeling is described as a Poisson approach using grouped data which is similar to those proposed by Dickman *et al.* (115) albeit not published until 2004. It is not explicitly clear what has been modelled but it may be that the authors modelled SMRs instead of excess mortality and as stated earlier in section 3.3.3, this is not relevant to the current investigation.

Stahle *et al.* (48) in 1997 also used a grouped Poisson approach to modelling with reference to a paper by Breslow *et al.* (116) which is a statistical paper detailing the use of observed and expected survival along with the use of SMRs along with the analysis of grouped data and Poisson regression models. This paper investigates survival after a coronary artery bypass graft (CABG). Similarly to the previous paper the type of modelling is difficult to determine but appears to be on relative survival in this case. The authors reference a relative survival program (117), but this is not capable of modelling.

Stahle *et al.* also performed further relative survival modelling after CABG earlier in a paper published in 1994 (18). However in this paper the methods proposed by Hakulinen and Tenkanen were adopted (40) which analysed the grouped level data using a Binomial distribution. This method is more akin to what would be expected from a relative survival model, although now outdated by the Poisson models (115).

Finally the last paper found that used a form of relative survival modelling was that by Verheul *et al.* (85) published in 1995. However this did not model the ratio of the observed

and expected survival. The authors take the difference between observed and expected survival as the excess mortality. The excess mortality was then modelled using a Poisson GLM using GLIM with the package written by Hakulinen (40). Similarly to the the paper by Stahle *et al.* (18) this method is also what could be classed as proper relative survival.

3.5 Discussion

There is clear interest in expected mortality as many papers attempted to look at this. Of the initial papers found using the search methods described above, less than eighty papers were found to be of potential interest, and all of these were published between 1990 and 2008 despite the method having its roots in the 1960s. The likely reason why relative survival was not overly popular before this time period may be that the method is fairly computational in terms of matching an observed survival group to an expected survival group which has become much easier in many countries around the world in recent times.

The lack of modelling is the key fault with the research investigated here as only four papers modelled the ratio or excess mortality allowing for covariate effects and confounding. There is also a lack of clarity in the modelling methods. There is an inconsistency between comparisons using life tables with expected mortalities to then model univariate and multivariate all-cause survival rather than using the expected survival. Relative survival is used for descriptive analysis in some papers, but is not generally extended to modelling. The main message being that there has been some, but not very much, work in the relative survival area, but there is a need to fully explore the issues and potential advantages that the method could provide to CHD research.

The lack of relative survival methods in common statistical packages may be one of the major drawbacks as user written programs would be required before the Hakulinen approach used in GLIM in 1987. However this program did not appear to make any grounds in heart disease as only two papers utilised it. Today's analyst has relative survival methods easily accessible to them in common statistical packages, with SAS macros and Stata ado files regularly available. This has still not seen a rise in the use of relative survival outside of Cancer and that is the reason this thesis was investigated.

Chapter 4

Relative survival model fitting in coronary heart disease

4.1 Introduction

This chapter aims to assess current modelling approaches available in relative survival. The ability to model coronary heart disease data, which has a very high mortality rate in the first few days as described in chapter 2, will be assessed. Most current models are fitted on the log excess hazard scale within relative survival, and two approaches are available when modelling. Most current approaches are based on split-time data and take the form, $h^*(t) + \exp(\beta x)$, and only differ in the way the baseline excess hazard is modelled, cure models do not split time (29; 28). The models developed in chapter 5 do not split the time-scale. This chapter will give an overview of the models using individual and grouped level data.

The first approach utilises a piecewise model (115; 40) where the data are split by time and the time scale is split with fairly large time intervals. This is the most common approach in practice. An individual level approach includes a separate row of data for each individual's time intervals, i.e if $1,000$ patients survived for ten time intervals there will be $10,000$ rows of data. The grouped approach collapses the data into a smaller dataset with one row of data per time interval and covariate pattern, i.e. using the previous example there would only be ten rows of data.

The second, more flexible approach, investigates time as approximately continuous, while still using split-time data. This is achieved by having regular intervals with many more splits than would be used in a piecewise approach. This gives the impression of a continuous model and could be extended, in theory, to have a time split at every unique event time to create a genuine continuous model.

4.1.1 Section layout

The next section investigates piecewise models as they are easy to fit and the most commonly used in practice. This will start with the Hakulinen (40) method which was the first model to be developed that was available in standard computer software. The Dickman model (115) will then be fitted at the grouped level and then the individual level, which is equivalent to the Esteve (27) model.

The continuous models, starting with standard polynomials ranging from linear up to quartic, will then be assessed. Following this will be multivariate fractional polynomials (MFPs) which are more flexible and have been used in relative survival before (39). The MFP section will also investigate the confidence intervals via bootstrapping and model averaging. Splines (37; 38; 118) will be investigated, then partitioned spline models, before the final discussion section, which will detail the conclusions of the investigation.

4.2 Methodology

In order to assess these models a generic structure will be used for each and the same models will be fitted for all approaches. The piecewise models will use seven split-time intervals to obtain estimates for these data at the grouped and individual level. The continuous models will be assessed using three different datasets, one named *long* intervals which has wide and consistent time splits every six months. The second dataset, *short* intervals, has fairly narrow but consistent time intervals at 0.05 year-bands for the first year of follow-up then 0.1 year-bands up to the end of follow-up (6 years). Both *long* and *short* intervals were proposed by Remontet *et al.* (118). The final dataset adopts *user-defined* intervals, which places more splits where there is more information and less splits where the number of events is limited. These are placed at the following times in years: `0 0.001 0.01 0.025 0.05 0.075 0.1 0.15 0.2 0.25 0.3` then every `0.1` years up to `1`, `1.2` then every `0.2` years up to `6`. This creates ten intervals up to 0.3 years then a split every 0.1 years up to a year then every 0.2 years from 1 to 6 years modelling a total of 41 intervals.

It is fairly common to ignore the first 30-days of follow-up in CHD (34), conditional of survival to 30-days as this is the time frame with the highest experienced mortality. The start date is still the date of hospitalisation but any patient who died within the first month is excluded from the analysis. This is investigated in order to assess if the very high early excess mortality rate is leading to some of the problems encountered. Within relative survival proportional excess hazard models can be fitted where the excess hazard rates are assumed to be proportional to each other, similarly to a Cox proportional hazards model. In order to assess time dependent effects non proportional excess hazards can also

be fitted in relative survival.

4.2.1 Data setup

The data used were described in section 2.3. Age was selected as the covariate to assess for simplicity of illustration, it has also been shown to be one of the most influential covariates when assessing relative survival (119; 36). Age groups were defined as < 60, $60 - 75$ and ≥ 75 with the average age of patients being 66.7. The youngest age group included $1,373(28.92\%)$ patients, the middle age group $1,938(40.83\%)$ and the oldest age group $1,436(30.25\%)$. Due to the need to split-time in piecewise approaches and in continuous approaches assessed at the grouped level it is not possible to assess continuous covairates. Age will be assessed continuously in chapter 5 when split-time data will not be used.

Age group	Survived	Died	Total
<65	1,194 (86.96)	179 (13.04)	1,373
65-75	1,320 (68.11)	618 (31.89)	1,938
≥ 75	474 (33.01)	962 (66.99)	1,436
Total	2,988 (62.95)	1,759 (37.05)	4,747

Table 4.1: Number of events (%) in each age group

Table 4.1 shows the number of events in each age group which highlights that older patients have a much higher proportion of death than the two younger age groups. The table also shows that there are more patients who died than survived in the oldest age group but a surprisingly small number of patients in the youngest age group died (only 13%).

The data for this chapter were split using a Stata program called `STRS` (120). An example of the use of `STRS` is shown below:

```
strs using "Expected Survival.dta", breaks(0 0.08333333 0.5 1 2 3 4 6)
mergeby(sex year age) by(agegp) maxage(100) diagage(age_mi)
diagyear(yeardiag) attage(age) attyear(year) survprob(prob)
savind(i_agegp,replace) savgroup(g_agegp,replace)
```

This example splits time at one and six months, one, two, three, four and six years creating an individual level dataset (`i_agegp`) with several rows of data per subject and a grouped (`g_agegp`) dataset which groups all subjects together. The `mergeby` option shows that the expected mortality rate is matched by patient sex, year of hospitalistion and age at hospitalisation. The `by` option splits the dataset by age groups. The other options include `diagage` and similarly `diagyear`, which contain the patient's age at hospitalisation and year of hospitalisation, whereas `attage` and `attyear` will be created to contain age and year for the patient at each time interval to which the expected survival rates will be matched. This means that the expected survival changes for every year older the patient becomes whilst in the study.

Patient	Age	Died	Start	Survival time
Patient I †	47	Yes	0	3.4606
Patient I	47	No	0	0.08333
Patient I	47	No	0.83333	0.5
Patient I	47	No	0.5	1
Patient I	48	No	1	2
Patient I	49	No	2	3
Patient I	50	Yes	3	3.4606

Table 4.2: Individual level data for a patient before and after splitting time. Before† splitting time.

As an example of individual patient data is shown in table 4.2. The information in the top row for one patient is split by time to create six rows of data where the patient died aged 50 in the interval ending at four years.

4.3 Piecewise modelling

For a piecewise analysis time is split into intervals that are generally equal, but can be unequal, and the baseline excess hazard is assumed constant within each interval. Covariate effects are assumed constant for all intervals in a proportional excess hazards model. The first interval includes the first month which has the effect of appearing to remain constant for a month. In reality it is likely that the effect is much larger initially and much smaller in the later half of the interval but this flexibility has not been incorporated into the model. For seven intervals and three age groups there are nine parameters to estimate. If age is time dependent, i.e. non-proportional, then a non-PEH model for age groups would include 21 parameters (14 age group interactions and 7 baseline effects). It is clear that assessing several covariates over time or using more intervals quickly results in a model with many parameters to estimate.

Dickman *et al.* (115) wrote that an additive excess hazards model or relative survival model can be written as

$$S(t;x) = S^*(t;x) \times R(t;x) \tag{4.1}$$

where S, S^* and R are cumulative observed, expected and relative survival respectively, x is a covariate vector, where the expected hazard may only depend on a subset of these covariates, i.e. age and sex. The basic relative survival PEH model can be written on the hazard scale as:

$$h(t) = h^*(t) + \exp(\beta x) \tag{4.2}$$

Where $h(t)$ is the total mortality rate and $h^*(t)$ is the expected mortality rate. The interest when using relative survival is in the survival relative to the population but also in the excess mortality rate due to modelling on the excess hazard scale. The excess hazard is normally assumed to be a multiplicative function of the covariates written as $\exp(\beta x)$.

4.3.1 Hakulinen binomial models

In 1987 Hakulinen and Tenkanen proposed a relative survival regression model (40) which outlines how a proportional excess hazards regression model may be adapted to estimate relative survival rates using standard generalised linear models (GLM) (121). The proposed model is based on grouped level data, obtained from lifetables, in the framework of GLMs using a binomial error function with a complementary log-log link for the number of observed deaths where the number of survivors, ns can be written as:

$$ns \sim Bin(nr, p_r) \tag{4.3}$$

.

Where p_r is the probability of survival per interval and nr is the effective number at risk, calculated using the same notation as the life table (table 2.1) as

$$nr = n - \frac{w}{2} \tag{4.4}$$

.

Censored survival times are assumed to occur uniformly throughout the interval so that the average number of patients at risk in the interval is given by 4.5, this is known as the actuarial assumption and ns is calculated for each interval as

$$ns = n - d - \frac{w}{2} \tag{4.5}$$

.

This model was the first that could be fitted using common statistical software and was developed for use in GLIM (122) which was considered a great advantage of the method, as unlike other programs specifically for relative survival GLIM was available throughout the world. This also meant that standard predictions and model validations were available and, more importantly, easy to model.

The Hakulinen model can be written as:

$$\ln\left[-\ln\frac{p}{p^*}\right] = \beta_0 + \beta x \tag{4.6}$$

Where p^* is the expected probability of survival, β represents the coefficient values for covariate x and β_0 is a constant (intercept). This model is easily used in standard statistical software by creating a user defined link function.

The advantage of the methodology, particularly in 1987, is that relative survival rates could be obtained using a GLM and excess hazard rate ratios related to the prognostic factors are also obtainable. However the use of grouped level data, rather than individual level data, and the proportional excess hazards (PEH) assumption are fairly restrictive. Statistical packages have greatly developed in the last decade allowing user defined links to be easily adopted in mainstream packages.

Esteve et al(27) discussed the Hakulinen approach and particularly the use of a binomial distribution for the number of observed deaths in each follow-up interval. The population is usually heterogeneous for expected survival at the beginning of each interval and the probability of surviving to the end of it is different from person to person. This is taken into account in Hakulinen's method by allowing different dispersion from that anticipated by the binomial distribution, for example estimating the true variance with a scale parameter.

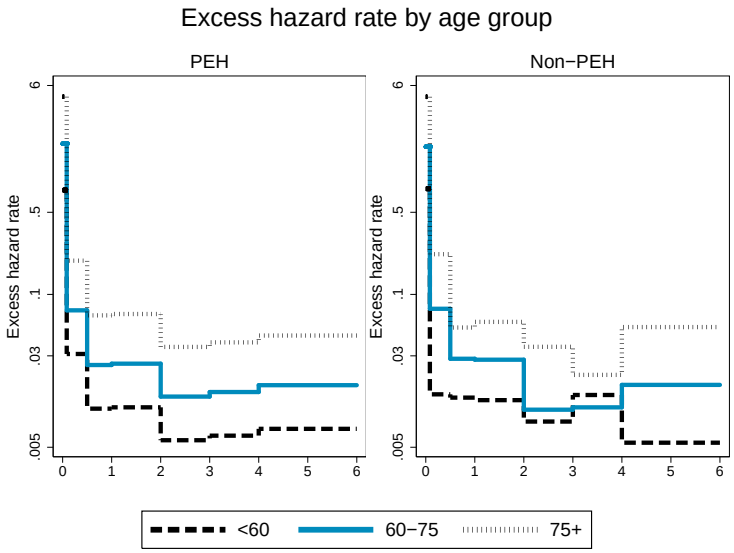


Figure 4.1: Plot of the excess hazard rate from a Hakulinen piecewise model

Fitting this model to the LRI dataset the excess hazard ratios for the two age groups are given as 2.35 for the patients aged 65-75 and 6.23 for the patients aged $\geq$ 75 both in comparison to the youngest age group (<65), as shown later on in table 4.3. This model assumes proportional excess hazards and the left hand side of figure 4.1 shows the estimated hazard rates from this model. There is a lot of variation over time when using

this model as the estimates start high and decrease rapidly. The estimates reach their lowest at the end of 2 years but then rise again during follow-up. This may be due to the effect in specific age groups and choice of time interval splits. The same effect in the final two years is most likely due to the low number of events during this time period, and is probably due to chance. A non-PEH model can be fitted by adding an interaction between age groups and time intervals. The non-PEH results are plotted on the right hand side of figure 4.1 which appear to be quite different to the PEH model after two years. A likelihood ratio test with 12 degrees of freedom gave a value of $p = 0.3125$ showing a non-significant result which suggests no evidence of age being time dependent.

4.3.2 Dickman Poisson models

It is possible to use the Dickman (115) approach on either grouped or individual level data. The approach uses the Poisson distribution where $d_j \sim Poisson(\mu_j)$ and $\mu_j = h_j y_j$ where h_j is the total hazard and where y_j is person time at risk so that $\mu_j = (h_j^* + \lambda_j)y_j$ which can be expressed as $\mu_j = (h_j^* + \exp(\beta x_j))$. Note that the j^{th} subscript refers to one row of the data and is therefore representative of the grouped or individual level approach. The expected number of deaths d_j^* is calculated as $d_j^* = h_j^* y_j$. The model can then be written as

$$\frac{\mu_j}{y_j} = \frac{d_j^*}{y_j} + \exp(x_j\beta) \quad \text{which is equivalent to} \qquad \ln(\mu_j - d_j^*) = \ln(y_j) + x_j\beta \tag{4.7}$$

This uses a Poisson distribution with non-standard link function ($\ln[\mu_j - d_j^*]$) and an offset ($\ln[y_j]$). A disadvantage of the piecewise approach is the inability to model continuous covariates when the data are split and the inability to deal with zero deaths in an interval subgroup. Zero cells can cause convergence problems and as such estimating several covariates can be difficult in small datasets as there are not enough events. There is also a loss of information as individual level data will include exact survival times which should always be used where available, but are clearly not used when all individuals are grouped together. The main advantage is speed, these models generally converge in a few seconds or less. It has been found in large cancer datasets that these results often concur with the individual level results and as such the added benefit of speed has made these models fairly standard in applied work.

This model gave very similar estimates to those from the Hakulinen method with an excess hazard rate ratio of 2.41 for the middle age group and 6.79 for patients aged ≥ 75, as shown later on in table 4.3, where it is also shown that the confidence intervals are fairly narrow. Figure 4.2 shows the estimates from a piecewise Dickman model using grouped level split-

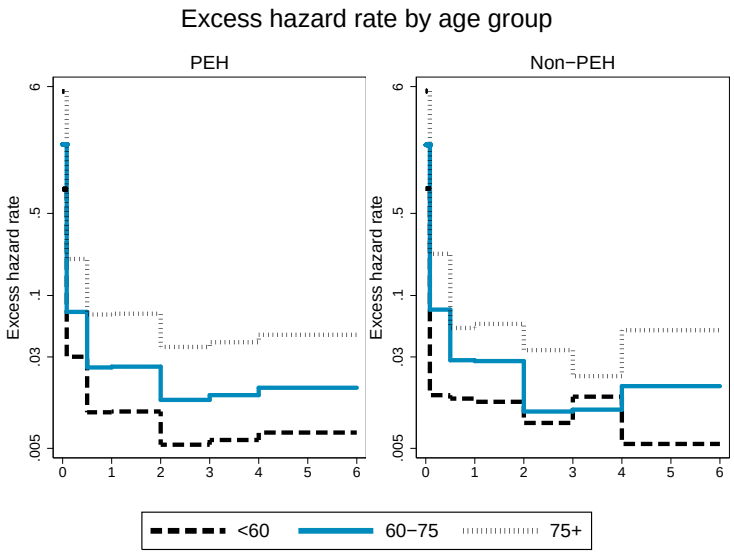


Figure 4.2: Plot of the excess hazard rate from a piecewise *grouped* Dickman approach

time data for PEH on the left hand side and non-PEH on the right hand side. The PEH model appears sensible as it starts fairly high, indicating the high excess mortality rate in the first month of follow-up and then drops to a lower more consistent level of mortality. The non-PEH model shows a fairly similar picture to the PEH model up to two years. The likelihood ratio test gave a value of $p = 0.2998$ with 12 degrees of freedom suggesting that an assumption of PEH is appropriate. Both of these plots are very similar to those given by the Hakulinen model shown in figure 4.1.

If the Dickman model is estimated from subject-band or individual level data then the estimates are identical to those obtained via the full-likelihood approach proposed by Esteve *et al.* (27). A grouped approach enables the use of standard GLM diagnostics.

Figure 4.3 shows the estimates from a piecewise Dickman model using individual level split-time data, these models take longer to converge than the grouped level data and consists of 23320 rows of data compared to only 21 in the grouped level dataset. The PEH plot (left side) and the non-PEH plot (right side) both appear very similar to the grouped level plots shown in figure 4.2. This is supported by the excess hazard rate ratios of 2.42 for the middle age group and 6.82 for the oldest age group. This supports the notion that the grouped level analysis is a very good approximation to the individual level analysis which is in turn equivalent to a full likelihood approach. There is very little difference between the grouped and individual level plot and it is only after around 3-4 years that any difference becomes noticeable. The likelihood ratio test on the individual data gave a value of $p = 0.3645$, which does not provide strong evidence against PEH.

All of the estimates from the three piecewise models are shown in table 4.3. This table highlights the model similarities between the Hakulinen grouped estimates and the

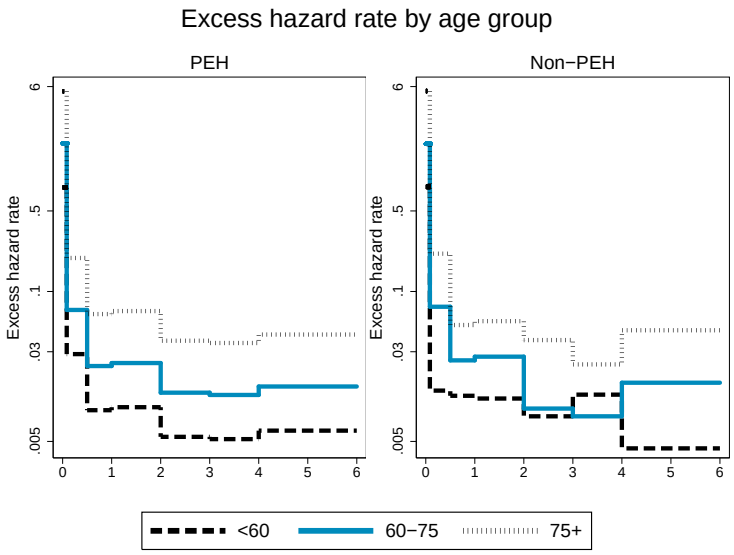


Figure 4.3: Plot of the excess hazard rate from a piecewise *individual* level Dickman approach

Modelling approach	Age group	
	60-75	75+
Hakulinen Grouped	2.352 (1.91 , 2.89)	6.225 (5.11 , 7.59)
Dickman Grouped	2.408 (1.96 , 2.96)	6.794 (5.57 , 8.28)
Dickman Individual	2.416 (1.96 , 2.98)	6.825 (5.59 , 8.34)

Table 4.3: Table of covariate estimates for age groups and AICs for the three piecewise models with confidence intervals using patients aged < 60 as the reference group

Dickman grouped estimates and in turn the Dickman individual level estimates with the Dickman grouped estimates. It is evident that these models are estimating the same thing as all estimates are similar. This is not surprising as they are fitting the same underlying model and only differ in the way in which they estimate the baseline excess hazard and time-dependent effects.

4.3.3 Piecewise conclusions

The three approaches described here, that of Hakulinen (40), Dickman (115) and Esteve (27), via an individual level Dickman approach, are all in effect estimating the same effect. All of the estimates are similar and consistent with the same conclusions and effect size. All of these models are on the log excess hazard scale, and due to the nature of a piecewise approach the results are not very flexible as a constant effect is assumed for each interval.

In all approaches age was not found to be time dependent but it should also be noted that this test is not without its problems due to the parameterisation of time and the use of a global test based on the degrees of freedom. For example by allowing age groups, a three level categorical covariate, to vary over time there were twelve extra parameters to estimate which makes it difficult to assess small changes in model fit.

In conclusion these models are quick to converge and are commonly used in applied research so they are clearly understood. Of the three models fitted the individual level piecewise approach would be the model of choice based on the model's equivalence to a full likelihood approach. However piecewise models are biologically implausible and as such a continuous time model may prove to be more appropriate.

4.4 Continuous modelling

4.4.1 Introduction

Modelling time as a continuous function in relative survival is performed using split times similarly to the piecewise models discussed in section 4.3. There are three sets of split times adopted which were described in section 4.2. It should be noted that the models fitted here all use the same underlying model but have a different way of modelling the baseline excess hazard and time dependent effects. For censored patients, time is calculated as the midpoint of the split-time interval, as shown in equation 4.8,

$$time = \frac{(end + start)}{2} \tag{4.8}$$

Where *end* is the end of the interval and *start* is the beginning of the interval. However where there is an event, *time* is equal to the event time, in this case, death, as this was found to be a good approximation to the likelihood they obtained previously using numerical integration (118). All figures are shown on the log scale as the initial excess hazard is very high and transforming the scale (y-axis) allows the whole model to be shown graphically.

4.5 Polynomials

4.5.1 Model introduction

Polynomials are probably the most common method of modelling non-linear continuous functions, but they have not been used in published relative survival research. Polynomial

Model	Age group		AIC
	60-75	75+	
Quartic	1.028	1.051	11709.05
	(0.82 , 1.29)	(0.85 , 1.30)	
Cubic	1.189	1.259	11595.72
	(0.95 , 1.49)	(1.01 , 1.57)	
Quadratic	1.011	1.016	11754.22
	(0.80 , 1.27)	(0.82 , 1.26)	
Linear	1.015	1.017	11808.47
	(0.81 , 1.28)	(0.82 , 1.27)	

Table 4.4: Table of covariate estimates with 95% confidence intervals for age groups and AICs for four polynomial models

terms are very easy to model but they are restrictive as pre-determined shapes are forced onto a model. It is a sensible starting point as these models are simple to investigate and may provide reasonable motivation for more complex models. Linear, quadratic, cubic and quartic models will be investigated, for example a quartic model can be written as:

$$\ln\left[\lambda(t)\right] = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3 + \alpha_4 t^4 + \beta x \tag{4.9}$$

Further polynomial degrees could be used by forcing more turns into the model but the relative survival would not be expected to change direction more than three times.

4.5.2 Proportional excess hazards

The first model was fitted using the long intervals dataset under a proportional excess hazards assumption. The models are fitted using similar methods to the piecewise model, but here time is included with up to four terms in the model, $time^4$, $time^3$, $time^2$ and $time$ are included, where time is calculated as shown in equation 4.8.

Figure 4.4 shows the results from the fitted model, with the quartic model in the top-left, cubic top-right, quadratic bottom-left and finally the linear model in the bottom-right. The first obvious conclusion is that the linear and cubic models show evidence of non-convergence, although they were reported to have converged. The plots, shown on the log scale, show that the underlying excess hazard for the quadratic and quartic models has not been modelled well as the polynomial turning points appear to be forced and unfeasible for these data. The quadratic and the quartic models are forced to turn upwards at the end of the follow-up forcing the excess hazard to rise unrealistically. A rise in excess mortality may be due to the limited amount of events at the end of the follow-up but these models estimate the excess to be nearly as high as the initial mortality which is clearly wrong.

The covariate estimates for the long interval proportional excess hazards model are shown in table 4.4 where it can be seen that the estimates are very similar in three of the models.

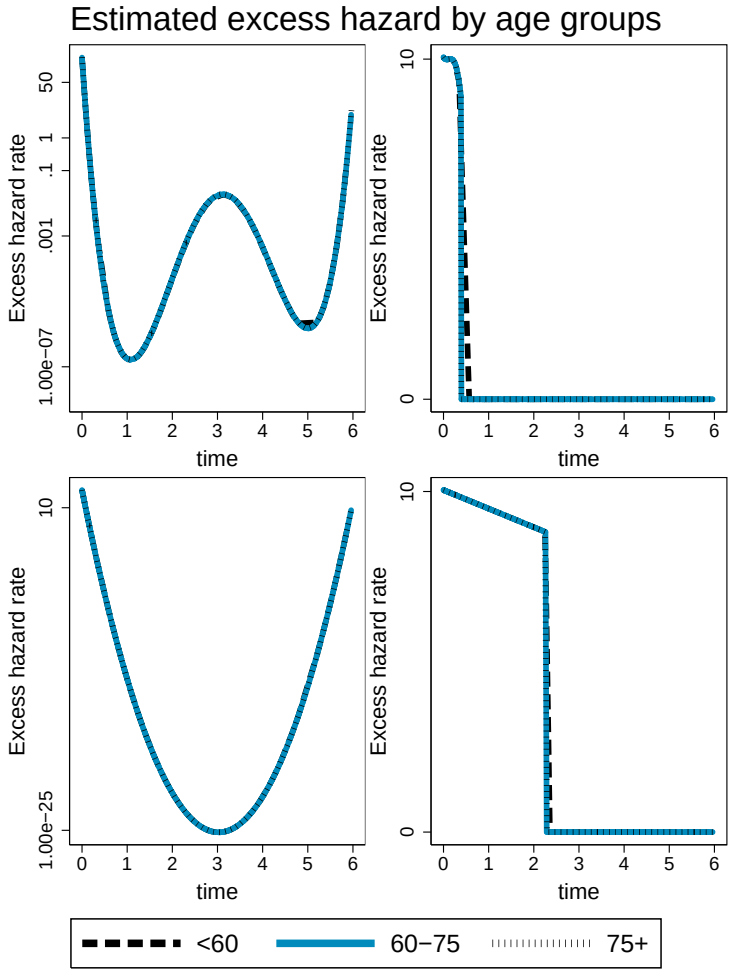


Figure 4.4: Plot of excess hazard curve for the PEH polynomial models using long intervals

The estimates for both age groups from these models are very different to the estimates from the piecewise models. The cubic model estimates are slightly higher than the other three models, with the oldest age group producing a statistically significant result as shown by the 95% confidence intervals. The cubic model AIC (123) is the lowest suggesting that this is the best model out of the four fitted under this criterion. It was shown in figure 4.4 that all models are poor. While it is possible that these models are either overly complex or simply unable to estimate the results due to the forced shape imposed upon the model it is still evidence that polynomials are very poor.

The short and user defined intervals datasets were all fitted under a PEH assumption and all models showed similar problems and very poor fits and are therefore not shown. Non-proportional excess hazards were also fitted and were found to be very poor with similar evidence of non-convergence even though they reported as converged. For this reason the non-PEH models for the long, short and user-defined datasets are not shown.

4.5.3 Removing the first 30-days

Part of the problem with the previous polynomial analyses may be the very large excess hazard rate experienced by patients in the first 30-days of follow-up as this may be overly influencing the model. In many instances a solution to this is to remove the first months data from the analysis, as seen in the paper by Stare *et al.* (34). Analysis investigates survival conditional on surviving for a month post hospitalisation, it may also allow for a more stable model as the large excess mortality of the first month is removed.

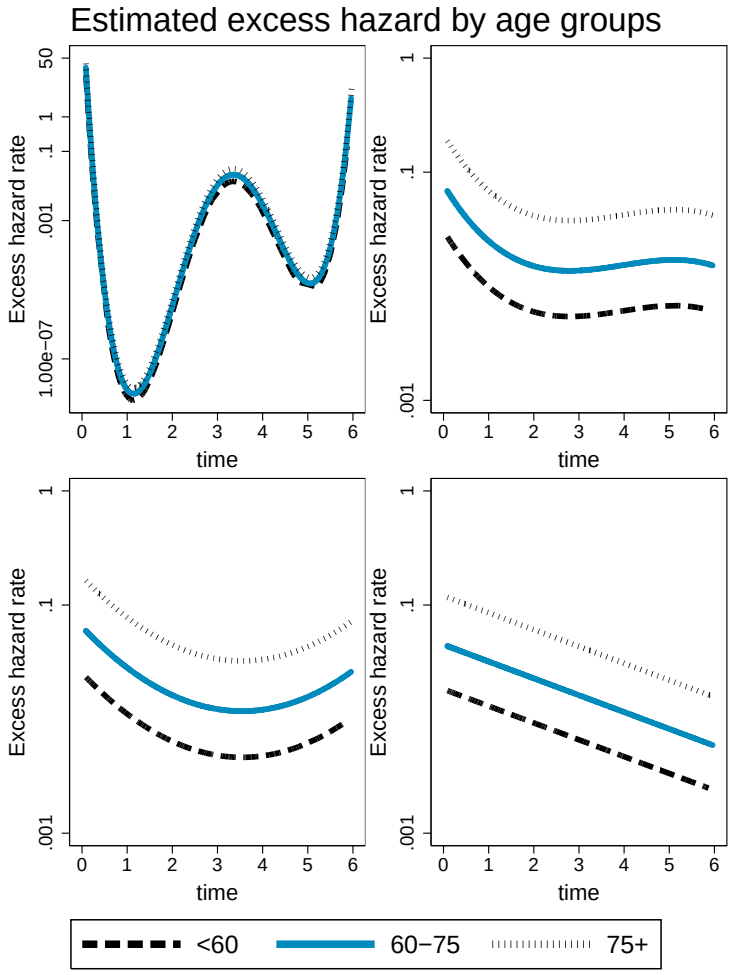


Figure 4.5: Plot of excess hazard curve for the PEH polynomial models using long intervals and removing the first months data

Figure 4.5 shows the results from the long interval polynomial PEH models that exclude the first 30 days of hospitalisation from the analysis. The quartic model (top-left) appears to be very poor, again forcing a shape onto the plot which is not feasible. The other three models manage to estimate the model fairly well, or at least better than when using the full dataset. The cubic model (top-right) shows the most feasible shape as it decreases for the first two years of follow-up and then levels out. The linear (bottom-right) and

quadratic (bottom-left) models do not appear flexible enough.

The same models were fitted to the short and user defined intervals but were found to show no improvement of fit and are not shown. Similarly the non-PEH models were also poor and are not shown.

Polynomial model conclusions

In order to determine why the polynomials did not provide a reasonable fit several attempts were made to try and resolve this. Firstly the polynomials were orthogonalised to remove any colinnearity between the *time* covariates as they are all a function of time. This did not improve the model fit or the model estimates. Secondly as the estimates all appear to be heading towards zero and negative numbers, the ability to be negative was removed by forcing a restriction on the maximum likelihood link function but this prevented models from converging. Finally the time points were altered so that equation 4.8 was adopted for all patients, including those that died, and this provided reasonable estimates for the age groups that were similar to the piecewise estimates. However when it came to model fit, the plots showed a similarly poor shape due to the functions that are forced upon these models. This suggests that the problem is in the data due to the initial high excess mortality.

The shapes forced onto these models do not seem plausible and often have a forced increase in excess hazard rate where one is not expected i.e. in figure 4.4 the quartic model shows an increase after four or five years because the underlying excess mortality is forced to change direction even if not appropriate. None of the approaches adopted when investigating the polynomial models have provided a suitable model from which accurate inferences could be drawn. This was alleviated slightly by removing the first 30-days of follow-up from the analysis but this is not ideal as the first 30-days holds important information and should be modelled in practice.

The inaccuracy and poor fit associated with polynomial models for these types of data was fairly expected as the shapes are not consistent with relative survival data or to the functional forms of the baseline excess hazards given by the previous piecewise models, shown in figure 4.3. Further, more flexible, models may be able to estimate the data more appropriately.

4.6 Multivariate Fractional Polynomials

4.6.1 Model introduction

Fractional polynomials (124) are used in regression models to fit non-linear functions and these functions are more flexible than the functions obtained from *standard* polynomials, which were shown in section 4.5. Fractional polynomials provide smooth estimates of both the baseline excess hazard rate and time dependent effects. The assumption of proportional excess mortality rates is often violated in relative survival and these types of models can fit time dependent effects using fewer parameters than the piecewise models.

The GLM Dickman model (115) can be extended to incorporate fractional polynomials for both the baseline excess mortality rate and for time-dependent effects by using more time-splits than were shown in section 4.3. The use of multivariate fractional polynomials (MFPs) has been advocated as the analysis overcomes the assumption of linearity and does not have the arbitrary selection of knots found when using splines.

The methods are highly computational and use of the methods on individual level data requires a large amount of memory and time. Therefore they are proposed for use on grouped data by Lambert *et al.* (39). This does not allow continuous covariates and reduces the complexity available in other approaches. Models fitted to the individual level take a long time to converge due to the amount of split times required for the models to be suitable. The models are fitted using the `MFP` prefix command in Stata which adopts an algorithm to select the best fitting model.

4.6.2 Fractional polynomial methodology

Fractional polynomials were initially defined by the linear predictor for a fractional polynomial of order M for covariate x as,

$$\beta_0 + \sum_{m=1}^{M} \beta_m x^{\rho_m} \tag{4.10}$$

Where each power, ρ_m, is chosen from a restricted set of powers. The usual set of powers is `-2, -1, -0.5, 0 , 0.5, 1, 2, 3` where x^0 is taken as $\ln(x)$. Other sets of powers can be selected, but this list has been shown to work well in practice (124). When using MFPs it is possible to fit FP1, FP2 and FP3 models, these represent single powers (FP1), two (FP2) and three sets of powers (FP3) for each covariate, i.e. time in an FP3 model would be fitted with three terms, such as $\ln(time)$, $time^{-2}$ and $time^{-0.5}$, all of which would be included in the model. The best fitting model is determined automatically by

using either the AIC or an α level. An FP2 model with powers $(-2, 0)$ can be written as

$$\beta_0 + \beta_1 x^{-2} + \beta_2 \ln(x) \tag{4.11}$$

If the number of powers, $M > 1$, then repeated powers, i.e. $(-0.5, -0.5)$ could be selected from an FP2 model, where if $\rho_1 = \ldots = \rho_M$, then

$$\beta_0 + \beta_1 x^{\rho_1} + \sum_{m=2}^{M} \beta_m x^{\rho_1} \{\ln(x)\}^{m-1} \tag{4.12}$$

Combinations of unique and repeated powers can be incorporated so that, for example, powers of $-2, -2, 0$ would give

$$\beta_0 + \beta_1 x^{-2} + \beta_2 x^{-2} \ln(x) + \beta_3 \ln(x) \tag{4.13}$$

All combinations of powers are fitted and the *best* fitting model is generally selected using an α level of 0.05 or the AIC. Using the default set of powers for an FP2 model there are 8 FP1 Models and 36 FP2 models which includes 8 repeated powers, giving 44 different models to assess. The best fitting model for fractional polynomials of the same degree can be obtain by minimising the deviance. The AIC and deviance are often used when comparing FP1, FP2 and FP3 models with each other.

MFPs have a fitting algorithm which consists of the following four stages:

1. Fit FP1, FP2 and FP3 models for the baseline excess hazard rate, while assuming linear effects for any time-dependent effects. The choice between the best fitting FP1, FP2 and FP3 is based on a pre-specified significance level (e.g. $\alpha = 0.05$) or the AIC.

2. Using the selected FP model for the baseline hazard, each time-dependent effect is fitted in turn using FP1, FP2 and FP3 models. The order in which this is done can be pre-specified (e.g. decreasing age group) or defined in order of univariate significance. The choice between the best fitting FP1, FP2 and FP3 is based on a pre-specified significance level (e.g. $\alpha = 0.05$) or the AIC.

3. FP1, FP2 and FP3 models for the baseline excess hazard (from step 1) are refitted, but with the time-dependent effects selected in step 2.

4. Steps 2 and 3 are then repeated until there is no change in the selected powers. This usually occurs within 2-4 cycles.

Time	Deviance	Powers
Linear	11800.472	1
FP1	11304.639	0.5
FP2	10841.867	-0.5 , 2

Table 4.5: First cycle for time

This algorithm is only applicable for time dependent effects, if a PEH model is fitted then the cycles are not needed as only time is modelled as continuous.

4.6.3 Individual level data

Modelling MFPs is a computationally heavy process when using individual level data. For an FP2 model there are 44 different models fitted on $33,103$ rows of data in the long interval dataset, $193,835$ in the short interval dataset and $126,086$ in the user defined interval dataset. However using the individual level data utilises all the information available after splitting the time scale and will be investigated first.

Long intervals

The first model fitted was a PEH model using long intervals and was fitted using Stata. The model is built by separating the continuous estimates for which a fractional polynomial term will be fitted from the categorical covariates. The AIC was used for model selection. The first step in the model building process determines the best fitting model by running through a cycle and the results of the first cycle are shown in table 4.5.

This details how an FP2 model for time was fitted with powers of $(-0.5, 2)$. The predicted baseline excess hazard rate for this model is obtained from:

$$\beta_0 + \beta_1 time^{-0.5} + \beta_2 time^2 \tag{4.14}$$

The estimates from this long interval PEH model are given as 1.08 for the 65-75 group and 1.06 for the oldest age group. These results are explained by the fit of the model which was very poor and are thus not shown. Note that these are *not* log excess hazard rate ratios and when compared with the piecewise estimates of 2.4 and 6.8 (table 4.3) it shows that the model is very poor. Several `convergence not achieved` messages were given from the program, providing evidence of the models inability to effectively estimate the covariates required in some of the 44 models assessed. This is not surprising given that there were many problems with the polynomial models, which are a subset of fractional polynomials.

Covariate	MFP	Deviance	Powers
	Cycle One		
	Linear	11795.633	1
Time	FP1	11249.033	0.5
	FP2	10822.968	-0.5 , 3
	Linear	10822.968	1
Age 65-75	FP1	10813.134	0.5
	FP2	10773.903	-1 , -0.5
	Linear	10773.903	1
Age 75+	FP1	10744.267	0.5
	FP2	10666.162	-1 , -0.5
Cycle one deviance = 10666.162			
	Cycle Two		
	Linear	10985.570	1
Time	FP1	10873.632	3
	FP2	10666.162	-0.5 , 3
	Linear	10718.931	1
Age 65-75	FP1	10709.667	0.5
	FP2	10666.162	-1 , -0.5
	Linear	10773.903	1
Age 75+	FP1	10744.267	0.5
	FP2	10666.162	-1 , -0.5
Cycle two deviance = 10666.162			

Table 4.6: Non-PEH model MFP cycle results for time and the age*time interactions

The non-PEH model for long intervals converged in two cycles and these are shown in table 4.6. This table shows the process adopted to converge on the *best* fitting model. The deviances shown in cycle two are different to those in cycle one even though the same fractional polynomials have been selected, for example in FP2 time. This is because the model is using the fractional polynomial terms in cycle two that were determined in cycle one, rather than an assumed linear function. Once the 75+ age group is assessed the deviance stays the same and this allows the model to converge. The final model powers were $(-0.5, 3)$ for time $(-1, -0.5)$ for the $65-75$ age group and $(-1, -0.5)$ for the oldest age group. The model itself was found to be very poor and is not shown.

Removing the first month's data as before in section 4.5 is an important process as this not only replicates other analyses of CHD data but removes the huge excess mortality associated with this time period and may result in a model which has fewer convergence issues.

Figure 4.6 shows the PEH model (left plot) excluding the first month's information. The initial excess hazard starts at around 50 excess deaths with the model converging to

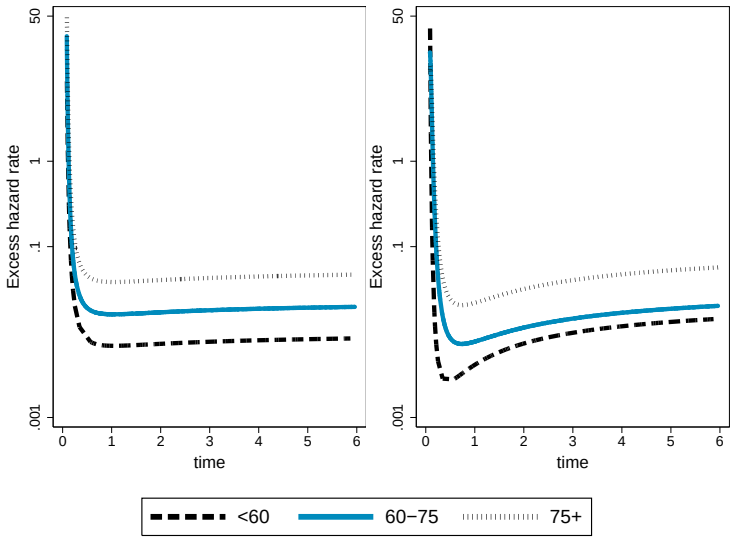


Figure 4.6: Plot of the excess hazard curve using long intervals and removing month one in a fractional polynomial PEH and non-PEH model

produce feasible results. The hazards begin to level out after six months with older patients giving the highest excess hazard and the youngest patients the lowest excess mortality rate. This model shows that the three age groups excess hazards are gradually increasing over time, which may not be feasible and could be a product of the model. The figure also indicates that the high excess hazard rate in the first month appears to be causing the problems.

Also shown in figure 4.6 are the non-PEH model estimates (right plot). The estimates drop further than the PEH model after the initial high mortality rate but then rise sharply at a higher rate than in the PEH model (left plot). The turning points on the plot, particularly the under sixty age group, are quite sharp and are thus likely to be an artifact of the model. It is unlikely that a patient's excess hazard would drop sharply only to rise over time. These models struggled to converge and it is possible that all the results from the long interval analysis are of little use. This lack of convergence may be due to time being modelled as continuous but only being available as 0.5 year time bands, which will not allow much flexibility.

Short intervals

The same four models, PEH and non-PEH both with and without the first 30 days, were fitted to the short interval data. The main issue with fitting these models is the computational time, as the four short interval models took almost three-and-a-half days to fit. The results for the analysis of the complete dataset are shown in figure 4.7. The PEH model (left plot) appears to be able to estimate the results for the entire follow-

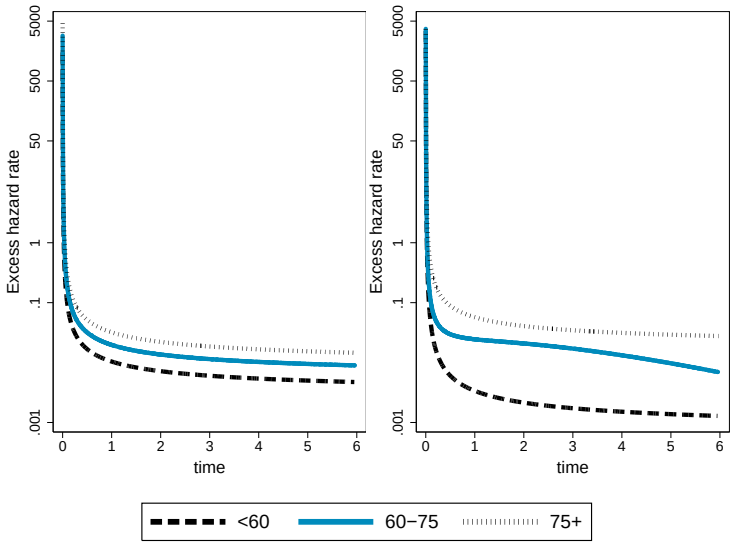


Figure 4.7: Short interval plot of the excess hazard curve from a fractional polynomial PEH and non-PEH model

up period fairly well inclusive of the high risk first month data. The powers chosen for time were $(-0.5, -0.5)$ and the initial excess hazard rate is estimated at near 5000 excess deaths, note that these are not per 1000-person years.

When fitting non-PEH models it is important to allow the two new time dependent covariate interactions to be zero in the model and to treat the value zero as such rather than a regular observation. The final model powers were $(-0.5, -0.5)$ for time $(0.5, 1)$ for the 65-75 age group and $(-0.5, -0.5)$ for the oldest age group. The non-PEH results (right plot of figure 4.7) show the middle age group does not begin to decrease until after around 2-years of follow-up whereas the oldest and youngest age group both decrease at a steady rate. Both the PEH and non-PEH models suffered from convergence issues with regular `convergence not achieved` messages. There is strong evidence that age is not proportional as a likelihood ratio test gave a value of $p < 0.0001$. This is different to the piecewise model findings.

Removing the first month from the analysis gave the results shown in figure 4.8 and the PEH model (left plot) estimates appear to be at a near constant level after the initial drop in excess mortality. This may be indicative of statistical cure, which is unlikely in CHD datasets. The non-PEH model (right plot) is also shown. The 60-75 year old age group excess hazard continues to drops after the initial sharp drop in mortality, which is still present once the first 30-days are removed. The oldest age group shows a reduction in mortality at a slower rate whereas the youngest age group estimates suggests an increase in excess hazard over time, with a strange *dip* after the initial drop in excess mortality. This *dipping* is likely to be an artifact of the model.

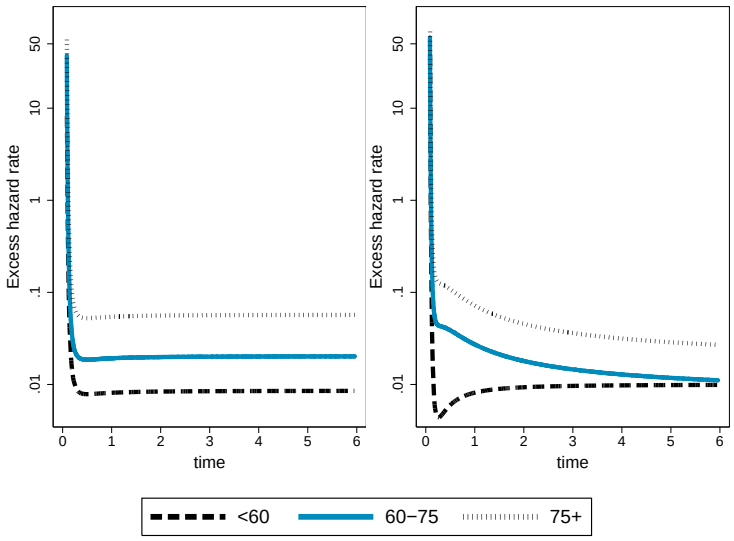


Figure 4.8: Plot of the excess hazard curve using short intervals and removing month one in a fractional polynomial PEH and non-PEH model

User defined intervals

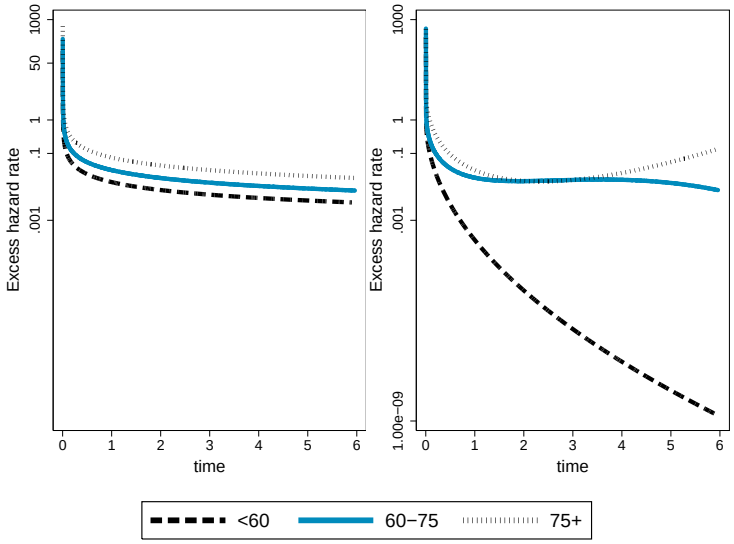


Figure 4.9: User defined intervals plot of the excess hazard curve from a fractional polynomial PEH and non-PEH model

The short intervals showed some success, therefore unlike with the polynomial models in section 4.5 the MFPs are showing potential use when there are many splits in the data. With the addition of more time splits early on in the user defined interval, where the majority of the information is, it may be possible to improve upon the short interval models. Figure 4.9 shows the results for the PEH model (left plot) and the non-PEH model (right plot). First impressions would suggest that the PEH model is a good fit as

the rate is shown to start high and then decrease gradually. The powers selected for time were $(-2, 0.5)$. A likelihood ratio test assessing the proportionality assumption in these user defined interval MFP models gave a value of $p < 0.0001$ suggesting strong evidence of non-PEH, but the non-PEH model appears unable to estimate the youngest age group adequately. The final non-PEH model powers were $(0, 0.5)$ for time $(-2, -2)$ for the 65-75 age group and $(-2, -2)$ for the oldest age group.

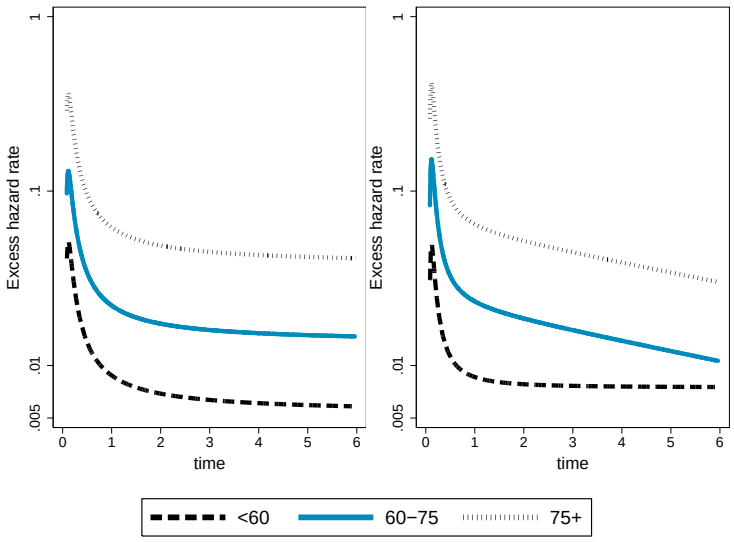


Figure 4.10: User defined intervals removing month one plot of the excess hazard curve from a fractional polynomial PEH and non-PEH model

The results from removing the first 30-days from the model are shown in figure 4.10. A strange shape has appeared at the beginning of the plots for both PEH (left) and non-PEH (right) where the rate appears to increase before it decreases. This is most likely to be an artifact of the model due to removing the information from the start of follow-up, for which this model has placed more time splits. The majority of the perceived benefit of using these user defined intervals is removed when the first 30-days is not assessed.

The PEH model shows that the youngest age group have the lowest excess hazard rate with the oldest age group having the worst excess hazard rate but all of the age groups start to level out after two years follow-up. The non-PEH model is different because it estimates the youngest age group as near constant after two years but models the older two age groups as reducing sharply over time at a similar rate after a year of follow-up.

Individual level data conclusions

There appears to be a problem with estimation in these models, and this is likely to be related to the very high early excess mortality rate, the choice of split times for the intervals and zero events in some of the time intervals.

Interval length	Estimates	Individual level	Grouped level
Long	Aged 60-75	1.076	2.454
		(0.86 , 1.35)	(1.99 , 3.03)
	Aged 75+	1.060	7.441
		(0.85 , 1.32)	(6.09 , 9.09)
Short	Aged 60-75	1.878	2.377
		(1.52 , 2.32)	(1.93 , 2.92)
	Aged 75+	3.001	6.595
		(2.42 , 3.72)	(5.42 , 8.03)
User defined	Aged 60-75	2.255	2.365
		(1.79 , 2.84)	(1.93 , 2.90)
	Aged 75+	5.191	6.281
		(4.16 , 6.48)	(5.17 , 7.63)

Table 4.7: Table of covariate estimates for age groups and AICs for PEH MFP models using individual level and grouped level data

4.6.4 Grouped level data

One of the main problems when modelling relative survival with MFPs is the computational time needed to fit them. It is not feasible to fit an MFP on short interval data during regular working practice as these took up to 24 hours using the relatively small LRI dataset. This leads to the investigation of grouped level data, which is much quicker. Unlike in standard survival analysis where assessing grouped data is equivalent to assessing individual level data, relative survival now includes an expected mortality which is dependent on a subject's age, sex and year of hospitalisation and as such is not equivalent. The first 30-days were retained in these models. In the individual level analysis only FP2 models were assessed as FP3 models could potentially take weeks to converge. At the grouped level FP3 models were assessed as they are now much quicker to run.

The PEH model estimates, which were not shown previously, are shown in table 4.7. This allows comparisons to be drawn from the three individual level PEH models and three grouped level PEH models, using long, short and user defined intervals. It should be noted that the non-PEH model estimates are only comparable when the model fitting algorithm (described in section 4.6.2) selects the same fractional polynomial terms for the covariate interactions, which in this case did not happen. For example, when using individual level data the long interval model selected powers of `-0.5, 3` for `time`, `-1, -0.5` for `age1time` and `-1, -0.5` for `age2time`, whereas the short term model selected powers of `-0.5, -0.5` for `time`, `0.5, 1` for `age1time` and `-0.5, -0.5` for `age2time`.

Table 4.7 shows that the individual level data results are very different from each other, whereas the grouped level approaches only show small variations. More interesting is the comparison to the piecewise estimates shown in table 4.3. The individual level piecewise estimates (Dickman approach) of $2.42(65-75)$ and $6.82(75+)$ are fairly well matched by the grouped approaches, whereas the individual level models are quite different.

The table supports the use of grouped data as the individual level data clearly encountered numerical issues and has produced some fairly unreliable results. The quality of the model fit cannot be assessed using table 4.7, so it is necessary to look at plots of the grouped models to determine if these estimates are from feasible models.

Long intervals

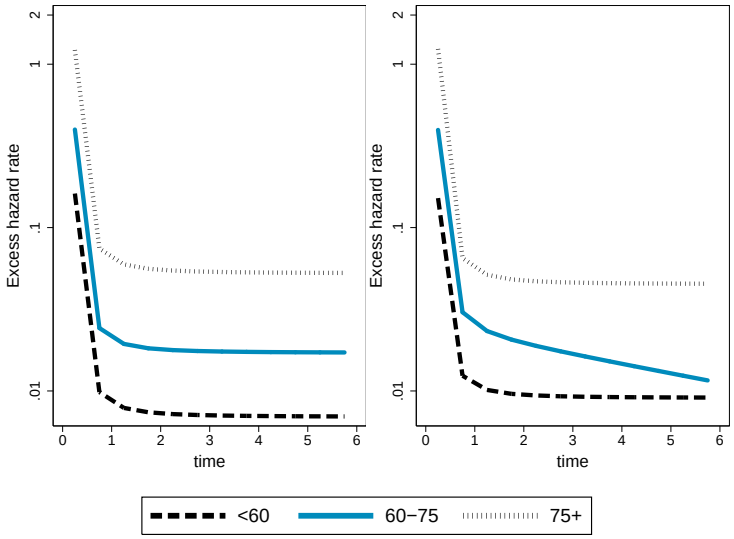


Figure 4.11: Long intervals plot of the excess hazard curve from a grouped fractional polynomial PEH and non-PEH model

Figure 4.11 highlights how using grouped data with long intervals does not provide estimates for the first three months as the value is assumed to have a constant mortality and as such the excess mortality rate starts fairly low at around 1.5 or less. This is because the first interval midpoint of time is at 3 months, using equation 4.8, and this initial estimate is an average of the first six months. The PEH model shown on the left of figure 4.11 shows a high mortality rate following an MI, which reduces quickly and begins to level out after a year or so. The age groups are stacked as expected with the oldest having the worst excess mortality and the youngest the lowest excess mortality rate. The powers chosen for time were $(-0.5, 2)$.

Figure 4.11 also shows results from the non-PEH models on the right hand side, which is very similar to the PEH model shown on the left hand side. The middle age group has a faster reduction in excess hazard compared to the other two groups. The width of the split-time intervals creates a curve that is not overly smooth, suggesting that the model may not be explaining the data adequately enough, certainly not in the first six months where most of the deaths following an MI occur. The final model powers were $(-0.5, 3)$ for time $(-1, -0.5)$ for the 65-75 age group and $(-1, -0.5)$ for the oldest age group.

Short intervals

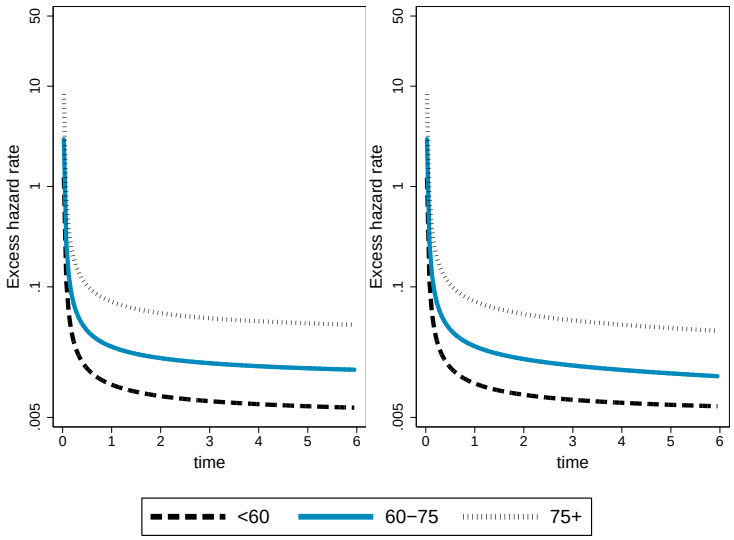


Figure 4.12: Short intervals plot of the excess hazard curve from a grouped fractional polynomial PEH and non-PEH model

Figure 4.12 shows the grouped level data approach using short intervals and shows a smoother looking curve than the long interval grouped data plots. The PEH model curve (left plot) appears appropriate as the excess mortality rate starts quite high and then drops rapidly to show only a small excess hazard after six months. The excess hazard rates are also shown to decrease slowly over time. The powers chosen for time were $(-0.5, -0.5)$. The non-PEH model results are also shown in figure 4.12 (right plot) which is very similar to the PEH model. This model appears to fit well but is in fact an estimate based on 0.05 year bands in the first year which may not be flexible enough to estimate that huge mortality in the first six months after an AMI. The final model powers were $(-0.5, -0.5)$ for time $(0.5, 1)$ for the 65-75 age group and $(-0.5, -0.5)$ for the oldest age group.

User defined intervals

The results for the user defined interval PEH model using grouped are shown in the left hand side of figure 4.13. The first thing to notice is the larger estimate of the excess mortality rate at the start of follow-up when compared to short intervals (figure 4.12) but the remainder of the model is very similar. The powers chosen for time were $(-2, 0.5)$. The non-PEH estimates are shown on the right hand side which is very similar to the short interval model on grouped data and their conclusions also apply here. The final model powers were $(0, 0.5)$ for time $(-2, -2)$ for the 65-75 age group and $(-2, -2)$ for the oldest age group.

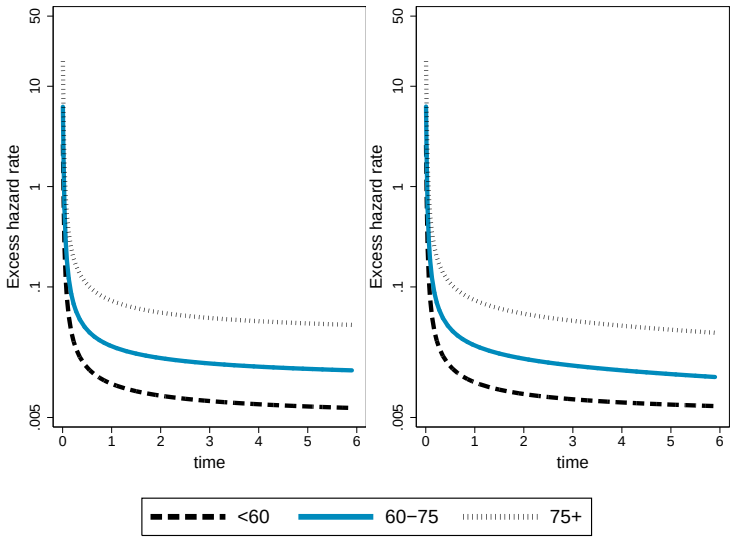


Figure 4.13: User defined intervals plot of the excess hazard curve from a grouped fractional polynomial PEH and non-PEH model

These plots show evidence of proportional excess hazards, with a likelihood ratio test giving a value of $p = 0.2267$ for the user defined intervals. The long and short interval likelihood ratio tests for the age group fractional polynomial terms, i.e. those that model non-proportionality, gave values of $p = 0.0640$ and $p = 0.0726$ respectively, which show evidence of non-PEH even if it is strictly non-significant.

MFP model conclusions

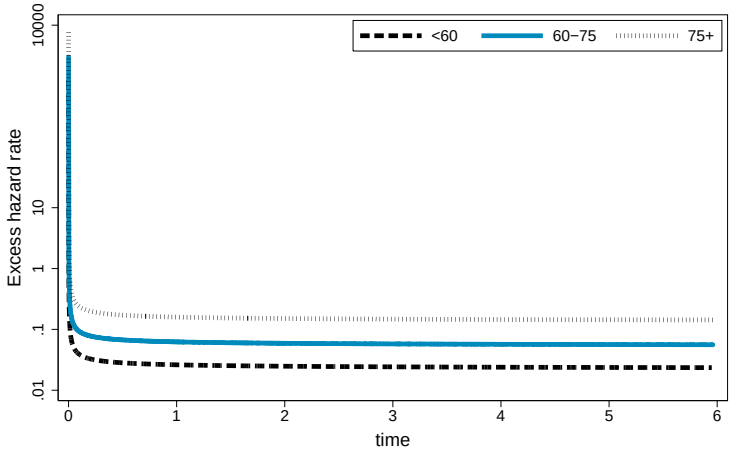


Figure 4.14: Plot of excess hazard curve for the best fitting grouped level model applied to individual level data

It was found that the user defined and short intervals gave estimates that were quite similar to the piecewise estimates when using the grouped level model, with the long

interval models being the most different. The PEH short interval model, which was fitted using a power of -0.5 for time in a grouped data approach was fitted at the individual level using a standard GLM. This gave the results shown in figure 4.14, and provided estimates of 2.37 (1.95 , 2.88 95% CI) and 5.96 (4.92 , 7.22 95% CI) for the $65-75$ and 75+ age groups respectively. These are slightly different to the estimates obtained from the grouped level data which may be due to using mid-time rather than the time at the end of the interval. The very high early excess hazard appears to be the cause of the problems at the individual level analysis.

Using grouped data drastically speeds up the model fitting process, i.e. from 3.5 days for four short interval models to around 3 minutes. However due to grouped data the results become a summary of the information that is available from individual level data. It may be appropriate to investigate the models at a grouped level before applying the best fitting model in individual level data, as shown here. However caution must be taken after the estimation problems that were encountered when using individual level data in this dataset.

While fitting these models there were several issues with model convergence, even the non-PEH short interval *grouped* data model produced a `non-convergence` message. This suggests that while the MFP model selection algorithm is helpful in providing the best fitting model they may not be fitting as well as would be hoped. This leads to another modelling direction in the use of splines described in section 4.8. Before this there is a section discussing the given confidence intervals from these MFP models and the issues related to the selection of only one from several good fitting MFP models.

What has been shown quite clearly is that the long intervals, proposed by Remontet *et al.* (118), do not work well for these data. It is not realistic to model continuous covariates as there are several issues with the current models. There are problems in using split-time and with modelling at an individual level.

4.7 MFP confidence limits

When fitting a fractional polynomial model only one of the possible models is chosen and hence the confidence limits are estimated from this selected model. This is taken to be the *true* model. The chosen model is only one of several well-fitting models and the standard errors will not contain the uncertainty by which it was selected. There are two approaches to obtaining an effect size and confidence intervals that show the uncertainty in model selection. The first is to use bootstrapping so that various models will be selected and produce the bootstrap intervals and the second is model averaging.

4.7.1 Bootstrapping

Bootstrapping is a re-sampling method that samples from the observed data. It is assumed that the data have been obtained by random sampling from some unknown distribution, F. Bootstrapping is useful when estimating the uncertainty in the estimate of the model parameter(s) without having to make any distributional assumptions as the distribution of F is unknown.

In a dataset of size N a sample is taken, with replacement, of size N, meaning that the same observation can appear numerous times, while others will not be sampled at all. The parameters are then estimated and the results stored. This is repeated for hundreds/thousands of bootstrap samples with all of the parameter estimates eventually being used to quantify the uncertainty. The point estimate from the original dataset is used and the confidence interval is calculated using the bootstrap samples.

There are several methods of estimating confidence limits using bootstrapping and for this approach two confidence limits have been selected, the first, a normal approximation, where a $(1-\alpha)\%$ confidence interval is calculated as

$$\theta \pm t_{1-\alpha,B-1}\sqrt{\frac{\sum_{i=1}^{B}\left[\theta_i^* - \bar{\theta^*}\right]^2}{B-1}} \tag{4.15}$$

Where θ is the parameter of interest, θ_i^* is the estimated parameter for the i^{th} sample, B is the number of bootstrap samples taken and $\bar{\theta^*}$ is the average value of the parameter over all bootstrap samples,

$$\bar{\theta^*} = \frac{\sum_{i=1}^{B}\theta_i^*}{B} \tag{4.16}$$

The sample taken (B) is usually large enough so that for a 95% confidence interval a z value of 1.96 can be used. The normal method is useful as relatively few samples are required. If it is reasonable to assume that the confidence interval is symmetric around the estimate then use of the normal method can lead to fewer bootstrap samples being needed. This is obtained using

$$z_0 = \Phi^{-1}\left(\frac{\#\theta_i^* \leq \theta}{B}\right) \tag{4.17}$$

Where $\#\theta_i^* \leq \theta$ is the number of elements of the bootstrap distribution that are less than

or equal to the observed statistic, i.e.

$$\rho_1 = \Phi(2z_0 - z_{1-a/2}) \tag{4.18}$$

$$\rho_2 = \Phi(2z_0 + z_{1-a/2}) \tag{4.19}$$

The percentile method gives

$$[\theta^*_{\alpha/2} \text{ to } \theta^*_{1-\alpha/2}] \tag{4.20}$$

where α_ρ is the ρ^{th} quantile (the $100\rho^{th}$ percentile) of the bootstrap distribution. It is often better to use percentiles than assume normality, however there is a small potential bias using this method. Therefore the second method that was adopted uses bias-corrected percentile confidence intervals, i.e. selects 2.5% and 97.5% as the interval when using an α-level of 95%. So the bias corrected method gives

$$[\theta^*_{\rho_1} \text{ to } \theta^*_{\rho_2}] \tag{4.21}$$

The three methods will generally give similar results in large samples. Bootstrapping is described in detail elsewhere (125). Bootstrapping has previously been used to assess the MFP algorithm (126). This paper assesses the possible instability in MFP models by first applying the algorithm repeatedly in many bootstrap replicates. Log-linear models are then used to investigate dependencies among the inclusion fractions for each predictor and among the simplified classes of fractional polynomial function chosen in the bootstrap samples.

Both proportional and non proportional models were assessed using bootstrapping. 200 and 1000 bootstrap samples were used on both the short and user defined interval data. The long interval data was disregarded as the models performed very poorly at the individual level. The best fitting model is determined using the grouped approach as described in section 4.6.4. It should be noted that different fractional polynomial models, i.e. using different polynomial terms, are selected for each bootstrap sample.

Figure 4.15 shows the first 30 model estimates from the short interval and user defined intervals. The models are very similar and the short intervals show that only a couple of bootstraps stray from the bulk (2 from 30 = 6.67%) whereas the user defined intervals only appears to have one model which is slightly different.

The first figure (figure 4.16) shown here gives the estimates from the short interval models

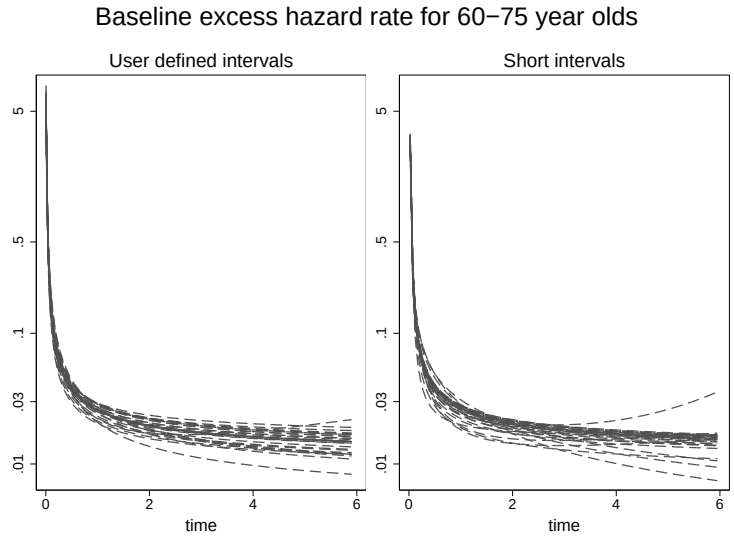


Figure 4.15: Plot of first 30 bootstrap model estimates for age group 2

with only 200 samples from the bootstrap. The figure consists of four plots, the top row shows the baseline excess hazard rate from a PEH model and the bottom row the excess hazard rate ratios from a non-PEH model. The left hand side of the figure shows the results for age group 2, which is for patients aged 60-75 years old, while the right hand side shows the oldest age group (patients aged ≥ 75).

There are three sets of confidence intervals: 1- the shaded area indicates the best fitting model intervals, i.e. these are the intervals that are observed for the best fitting model selected by the algorithm. 2- Dashed lines represent the intervals obtained under a normal approximation. 3- Solid lines represent the bias corrected percentile intervals as detailed earlier in this section. Also on the figure the dashed curve in the centre of the intervals gives the point estimate from the best fitting model.

The baseline excess hazard estimates (top row) are similar for both the normal and bias corrected bootstrap estimates, which are both in turn similar to the best fitting final model. After four years of follow-up the uncertainty becomes slightly larger in the bootstrap estimates suggesting that the final fitting model is not fully representing this. However this is to be expected as there are very few events in later time periods, but this does confirm the doubts about the relatively optimistic levels suggested by taking the best fitting model on its own.

The confidence intervals for the excess hazard rate ratio from a non-PEH model show that the normal estimates increase rapidly and the bias corrected results appear more sensible as they stay around the best fitting model estimates rather than becoming very large. The large intervals in the normal models is possibly due to some models having very high or low estimates towards the end of follow-up as the bootstrap sample may only have a

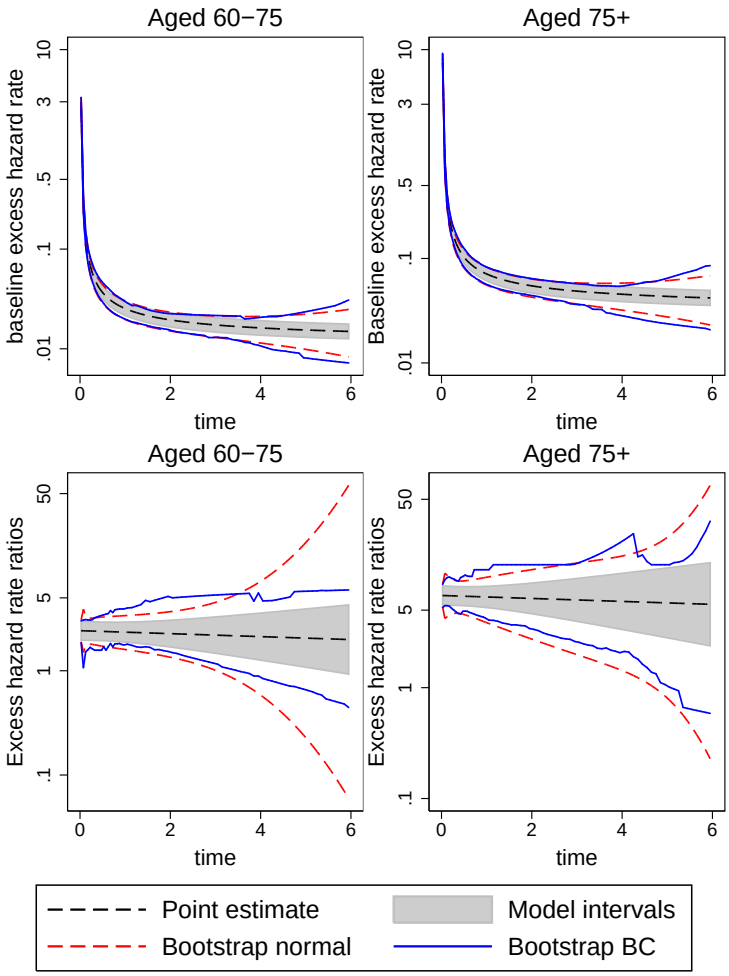


Figure 4.16: Plot of bootstrap confidence limits compared to the fitted model by age group using short intervals with 200 bootstrap samples

few events in the extremes. There does appear to be a problem with the bias corrected confidence intervals as they are not smooth. This is because only 200 samples are taken.

A second bootstrap was performed on the short interval data, this time with 1000 bootstrap samples, and the results are shown in figure 4.17. The figure is set out in the same way as the 200 samples figure shown earlier. Similarly to the 200 bootstrap samples shown in figure 4.16 the baseline excess hazard estimates (top-row) are similar for both the normal and bias corrected bootstrap estimates and these are both similar to the best fitting model that was selected by the fractional polynomial algorithm.

The normal estimates of the excess hazard rate ratio from a non-PEH model (bottom-row) increases rapidly to values that are not feasible. They become much larger than the bias corrected estimates in the middle age group. The estimates from the bias-corrected approach suggest a larger amount of uncertainty when compared to the best fitting model. The bias corrected estimates are more sensible than the normal estimates as they are not

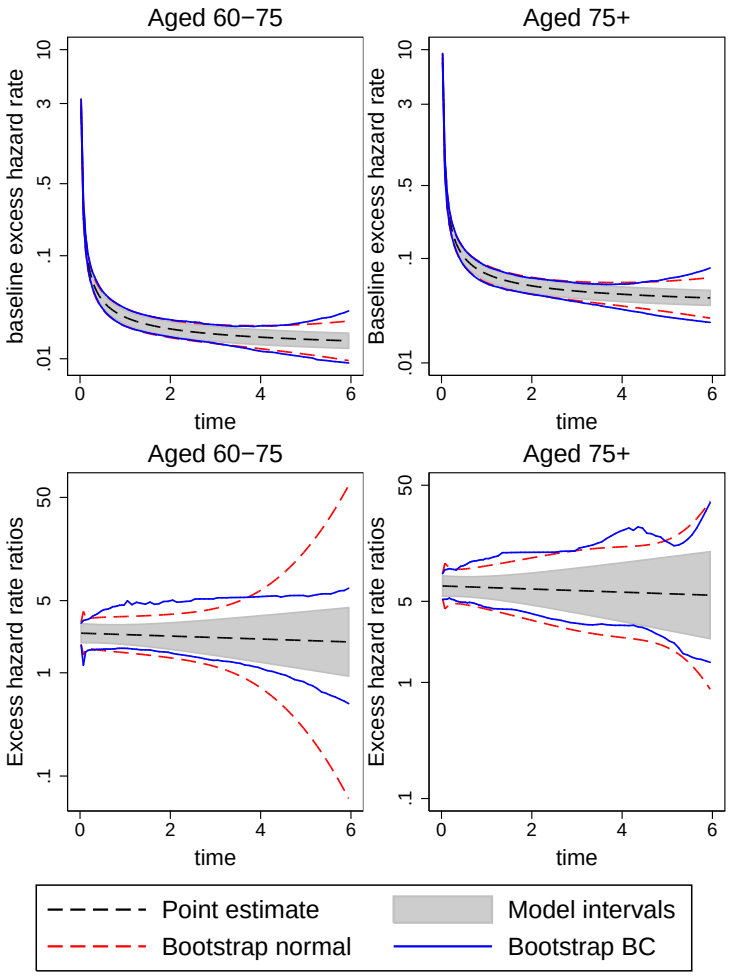


Figure 4.17: Plot of bootstrap confidence limits compared to the fitted model by age group using short intervals with 1000 bootstrap samples

as extreme in the middle age group. However the normal estimates are smooth in the oldest age group but the bias corrected estimates are not, even though there are 1000 bootstrap samples. This suggests that the bootstrapping has not been very successful as the strange looking confidence intervals are not explainable. The bias corrected estimates are slightly smoother than the 200 sample estimates but it may be necessary to take more than 1000 samples, although this becomes unfeasible in practice.

Bootstrapping with 200 samples were performed on the user defined intervals but were found to be very similar to the 1000 samples on the user defined intervals and as such do not add anything extra, and are therefore not shown. The final bootstrapping was performed using 1000 samples on the user defined intervals. The user defined intervals are tailored to be more specific to heart disease data and have proved to be the most reliable when using MFPs as they are better at estimating the high early mortality experienced by patients at the start of follow-up. In figure 4.18 the results of the final bootstrap are shown. The baseline excess hazard rate from a PEH model plots are very similar (top

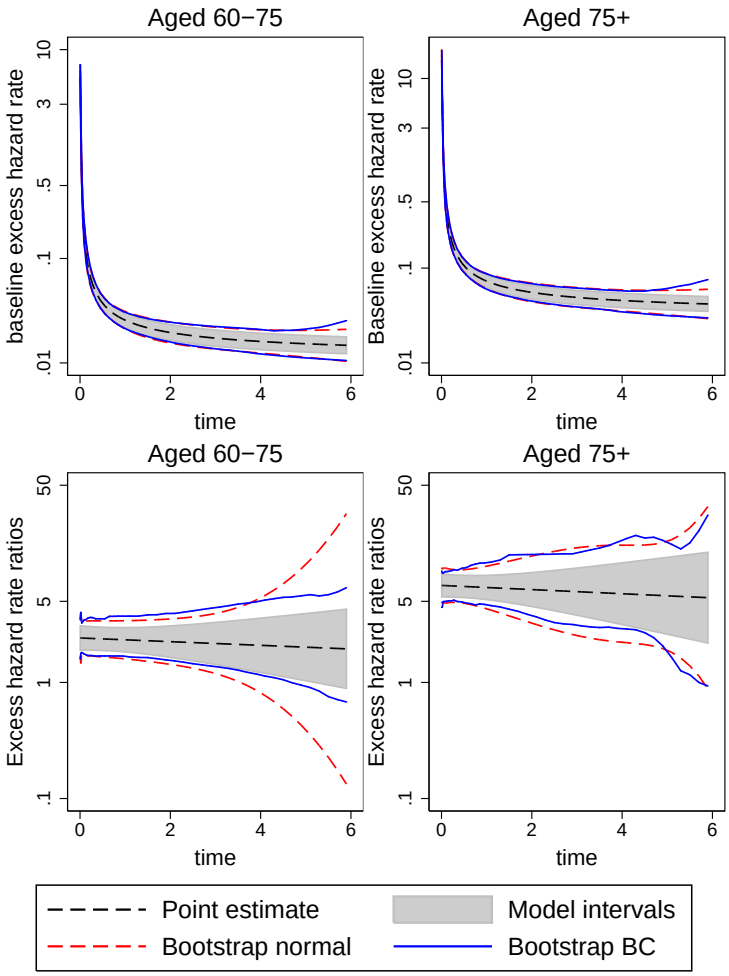


Figure 4.18: Plot of bootstrap confidence limits compared to the fitted model by age group using user defined intervals with 1000 bootstrap samples

row) with the confidence intervals being very similar to the best fitting model.

The excess hazard rate ratios for the middle age group $(60 - 75)$ shows the normal estimates to be much larger than both the bias corrected bootstrap and best fitting model estimates. The estimates are no smoother than the 200 samples user defined approach (*not shown*), suggesting that increasing the number of bootstrap samples has only increased the computational time, rather than precision. Taking only two-hundred bootstrap samples is still very time consuming in comparison to running the MFP algorithm once.

Using bootstrapping, the best fitting PEH models have been found to be fairly reliable as far as the estimation of the baseline excess hazard point estimate and confidence intervals are concerned as there is very little variation around this. The non-PEH models, giving estimates of excess hazard rate ratios, are much less certain and the results are more concerning. The uncertainty shown by the normal estimates is very large, therefore bootstrapping appears not to have worked as well as hoped. This may be due to there

only being a small number of events towards the end of follow-up. The use of another method may be appropriate.

4.7.2 Fractional Polynomial Model Averaging

Short and user defined interval models using grouped analysis all provided similar estimates and reasonably good fits to the data, however this does not truly represent the model selection that has happened. Bootstrapping in the previous section proves to be very time consuming as up to one thousand models are fitted which can prove to be restrictive in the assessment of the uncertainty in selecting one model from numerous good fitting models. A much faster method is to use model averaging (127), which requires that the algorithm only be performed once.

Model Averaging is a popular Bayesian research area (128; 129) but there is now increasing interest from a frequentist perspective (130; 131; 129; 132). The method averages over all of the models using weights where the best fitting models, determined using the AIC, are given more weight. Usually interest lies in model averaging for a parameter or for the functional form obtained from different models.

If there are K contending models, M_k ; $k = 1 \ldots K$ with weights, w_k, which are scaled so that $\Sigma w_k = 1$, then the estimate of a parameter or quantity, θ (assumed to be common to all models) is taken to be,

$$\widehat{\theta}_a = \sum_{k=1}^{K} w_k \widehat{\theta}_k \tag{4.22}$$

Where $\widehat{\theta}_k$ is the estimated value of θ for the k^{th} model, and the variance of $\widehat{\theta}_a$ is

$$\text{var}(\widehat{\theta}_a) = \sum_{k=1}^{K} w_k^2 \left(\text{var}(\widehat{\theta}_k | M_k) + (\widehat{\theta}_k - \widehat{\theta}_a^2) \right) \tag{4.23}$$

The AIC (123) can be used to derive the model weights (131)

$$AIC_k = \ln(L_k) - 2p \tag{4.24}$$

Recently Faes et al (132) used the AIC to derive model weights for fractional polynomial models using

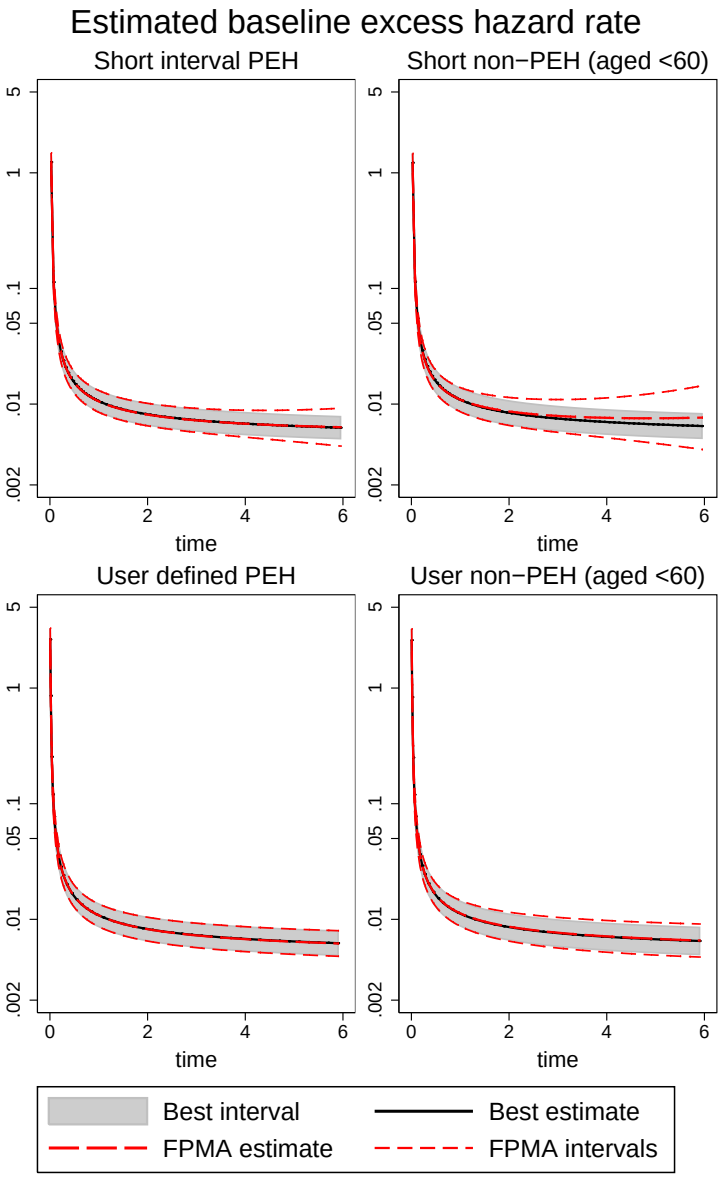


Figure 4.19: Plot of model average confidence limits compared to the fitted model for time

$$w_k = \frac{\exp(\frac{1}{2}\Delta_k)}{\sum_{j=1}^{k}\exp(\frac{1}{2}\Delta_j)} \tag{4.25}$$

where,

$$\Delta_k = AIC_k - AIC_{min} \tag{4.26}$$

Model averaging for fractional polynomials is available in Stata using `FPMA` (127) and was performed on the grouped level MFP models using short and user defined intervals.

The results are given for the baseline excess hazard rate as shown in figure 4.19. In the figure the top row relates to the short interval model and the bottom row the user defined intervals. The left hand side is for proportional models and the right hand side fits time dependent effects. The dashed lines represent the model averages whereas the shaded area represents the best fitting model, which was selected using the AIC.

The short interval results show a larger difference between the model average and the best fitting model whereas the user defined intervals appear to be very similar suggesting that the best fitting model includes all the uncertainty associated with the model. The short interval models are also very similar to each other with the non-PEH model showing that there is more uncertainty towards the end of follow up that is not accounted for in the best fitting model.

Only one covariate can be assessed when using the `FPMA` software for model averaging in non-PEH models, and the remaining covariates are fixed using the best fitting model. This means that the fitted MFP model can assess only one time-dependent covariate at any time with the forms chosen for the other time dependent covariates in the model assumed to be adequate. Figure 4.20 shows the excess hazard rate ratios for the 60-75 age group (left hand side) and 75+ age group (right hand side) short (top row) and user defined (bottom row) interval models.

It is here that a difference is noticed. The short term and user defined estimates using model averaging appear very similar, as do the best fitting models, but this only suggests that using the user defined intervals and short intervals will give similar results. The model averaging confidence intervals narrow half way through the follow-up period whereas the best fitting model increases over time. This is observed in both age groups. Similarly the point estimates rise near the end of follow-up suggesting that the best fitting model may not be giving the correct estimates. This was also observed in the bootstrapping, although not shown, and suggests that this may be a problem with using MFPs without correction. It appears as though the best fitting model is not capturing all of the information at the start of follow-up, when compared to the model averages.

Fractional polynomials are useful for these data, but the uncertainty expressed by the bootstrapping and model averaging should not be ignored. It is for this reason that the next models, using splines, may prove to be more useful, in that there is only one model being fitted.

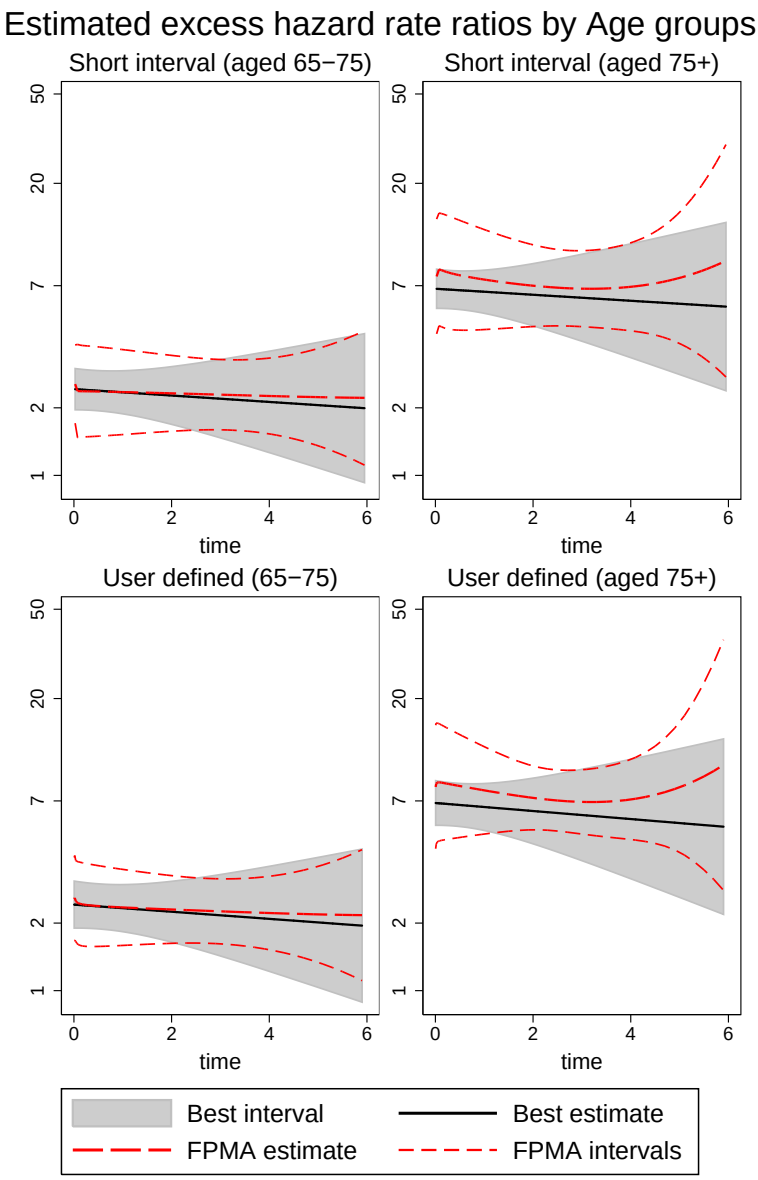


Figure 4.20: Plot of model average confidence limits compared to the fitted model for Age

4.8 Splines

Spline definitions

Splines allow for a more parsimonious model than a piecewise model and enables modelling of the baseline excess hazard and time dependent excess hazard ratios as a smooth function of time. A cubic spline is a smooth piecewise cubic polynomial function, i.e. a combination of a number of polynomial functions, each being defined over a different range or argument. This combination is smooth because the function and its first two derivatives are continuous at the points (knots) where the adjacent polynomial pieces join

together. If $u_+ = u$ if $u > 0$ and $u_+ = 0$ if $u \leq 0$ an unrestricted cubic spline survival function $S(t)$ (order n=4) with ν knots at times $t_i (i = 1 \dots \nu)$ can be written as

$$S(t) = \gamma_0 + \gamma_1 t + \gamma_2 t^2 + \gamma_3 t^3 + \sum_{i=1}^{\nu} \theta_i [(t - t_i)_+]^3 \tag{4.27}$$

There are several types of spline available including restricted cubic splines and B-splines. When considering equation 4.27 there are ν+4 regression coefficients ($\gamma_0 \dots \gamma_3$ and $\theta_1 \dots \theta_\nu$), however these are generally unstable before the first and after the last knot. Restricted cubic splines impose a constraint to force linearity before the first and after the last knot and with ν=3 knots a restricted cubic spline function (133) can be written as

$$S(t) = \gamma_0 + \gamma_1 t + \gamma_2 \left[(t - t_1)_+^3 - \frac{(t_3 - t_1)}{(t_3 - t_2)} (t - t_2)_+^3 + \frac{(t_2 - t_1)}{(t_3 - t_2)} (t - t_3)_+^3 \right] \tag{4.28}$$

Where $t_1 \dots t_3$ are the knotted time points. Data may be sparse in the extremes, which is why linearity may be more sensible. By increasing the number of knots the flexibility of the spline function is increased, however this induces a risk of over fitting. Restricted cubic splines offer the possibility to fit simple and complex patterns of changes in relative risks over time within a single model without a-priori definition of the functional form. In short restricted cubic splines are as follows:

1. Forced to join at the knots
2. Ensure gradients are equal at the knots by forcing the 1^{st} derivatives to be equal
3. Also make the 2^{nd} derivatives equal, i.e. the rate of change
4. Force linearity in the tails

Bolard *et al.* (37) first proposed the use of restricted cubic splines in relative survival which use split-time data as in the fractional polynomial analysis described in section 4.6. *B*-splines are another approach that has applications in relative survival (38). Restricted cubic spline functions are constrained to be linear in the tails, which potentially limits their flexibility. For these reasons quadratic *B*-spline functions are proposed in relative survival (38), and Giorgi *et al.* state that they have shown reasonable flexibility when modelling time-dependent effects.

B-splines are polynomial regression splines, which are piecewise polynomial functions of order q. B-splines can provide a good level of flexibility for fitting data. The splines join together smoothly at the knots due to the continuity of a spline function and of the $(q-2)$ first derivatives. To implement B-splines time is divided into $m + 1$ intervals and the B-spline function is given by.

$$g(t) = \sum_{j=-(q-1)}^{m} \alpha_j \left[B_{j,q}(t) = \frac{t - t_j}{t_{j+(q-1)} - t_j} B_{j,q-1}(t) + \frac{t_{j+q} - t}{t_{j+(q)} - t_{j+1}} B_{j+1,q-1}(t) \right], t \in (t_0, t_{m-1}) \tag{4.29}$$

where $j = -(q-1), \ldots, m$ and $B_{j,1} = 1$ if $t \in (t_j, t_{j+1})$ and $B_{j,1} = 0$ otherwise. Giorgi *et al.* state that each basis function is non-zero in each interval spanned by q adjacent knots, which leads to stable estimates and reduces computations. Estimates of the Esteve (27) PEH model are constrained to remain constant, which provides an average over time of the B-spline estimates. So if the true underlying predictor is time-dependent then this can result in a systematic underestimation of the effect in some time intervals and an overestimation in other intervals. B-splines have been implemented by Giorgi *et al.* (38) for use in the *R* statistical package (134).

4.8.1 Introduction to the models applied

Four different restricted cubic spline models will be fitted and discussed here. The first model will have nine knots at $0.003, 0.1, 0.25, 0.5, 1, 2, 3, 4$ and 5.5 time points (in years), this model incorporates many knots. The second model has seven knots, placed at $0.003, 0.05, 0.1, 0.5, 1, 2$ and 5.5 years. The majority of knots are in the first six months as this is where the highest number of events is present. The third model is more balanced but has only five knots placed at time $0.01, 0.5, 2, 4$ and 5.5 (in years). The final model has only three knots and focuses solely on the tails of the model, the most important point being at the start of follow-up with a knot at 0.003 and 0.1 and nothing until the final knot at 5.5. The knots will be indicated in the figures by vertical dashed lines.

It is important to remember that restricted cubic splines are forced to be linear in the extremes, i.e. for models one, two and four the effect is linear up to time 0.003, approximately 1 day, whereas for model two it is linear up to time 0.01, approximately 4 days. The long interval analyses will not be shown as they have already been found to be inadequate for these data, and were once again found to be poor when using splines.

To exclude the first 30 days different spline points will be required as the first knot is within the excluded time frame and thus needs changing. Model one changes to $0.1, 0.25, 0.5, 0.75, 1, 2, 3, 4$ and 5.5, removing the 0.003 knot but adding another at 0.75. Model two changes to $0.1, 0.25, 0.5, 1, 2, 3$ and 5.5 increasing the flexibility at a later point in the model now that the high 30 day mortality is excluded. Model three changes the first knot from 0.01 to 0.1 and model four changes to $0.1, 0.25$ and 5.5 again focusing on the extremes but now with a point later on.

Model knots	Age group		AIC
frequency	60-75	75+	
9 Knots	1.768	2.312	16501.71
	(1.43 , 2.18)	(1.85 , 2.88)	
7 Knots	1.586	1.708	16775.62
	(1.27 , 1.99)	(1.35 , 2.17)	
5 Knots	1.341	1.368	16918.36
	(1.04 , 1.72)	(1.07 , 1.75)	
3 Knots	1.314	1.325	16945.41
	(1.02 , 1.70)	(1.03 , 1.70)	

Table 4.8: Table of covariate estimates for age groups and AICs for four spline models using short intervals with confidence intervals

Model knots	Age group		AIC
frequency	60-75	75+	
9 Knots	2.341	6.178	20045.21
	(1.90 , 2.89)	(5.06 , 7.55)	
7 Knots	2.346	6.190	20042.03
	(1.90 , 2.89)	(5.07 , 7.56)	
5 Knots	2.245	5.844	20648.28
	(1.77 , 2.85)	(4.66 , 7.32)	
3 Knots	2.336	6.030	20677.81
	(1.83 , 2.99)	(4.77 , 7.62)	

Table 4.9: Table of covariate estimates for age groups and AICs for four spline models using user-defined intervals with confidence intervals

4.8.2 Stage one - PEH models

The short interval estimates using a PEH model are shown in table 4.8. Although it is not shown in the table there are $k-1$ parameters used to model the spline terms (k is the number of knots). The four models all produce an estimate of the age effects and a model AIC (123). Using the AIC the 9 knot model, although it has to estimate more parameters than the other models, has the lowest AIC. The estimates for the age groups in the model are different to the piecewise models and fairly different to each other which may be explained when comparing the fit of the models. These model estimates are very poor when compared to the piecewise estimates of 2.4 and 6.8 (table 4.3) and suggest a poor fit to these data.

The user defined interval estimates using a PEH model are shown in table 4.9 and these estimates appear to be very similar to each other and the piecewise estimates which is a very good sign that these models are producing something fairly reliable. The piecewise estimates were 2.42 and 6.82 for patients aged $65-75$ and ≥ 75 at hospitalisation respectively. Comparison with the short interval models as shown in table 4.8 suggest that the user defined intervals are likely to be a better fit. Using the AIC as a guide the 7 knot model appears to be the best fitting model when using the user-defined intervals in a PEH model. The choice of knots does appear crucial if the primary interest is in the

estimated excess hazard rate ratios. The estimates compare well to the piecewise estimates (table 4.3) and this suggests that the short interval model problems were caused by the split-times adopted.

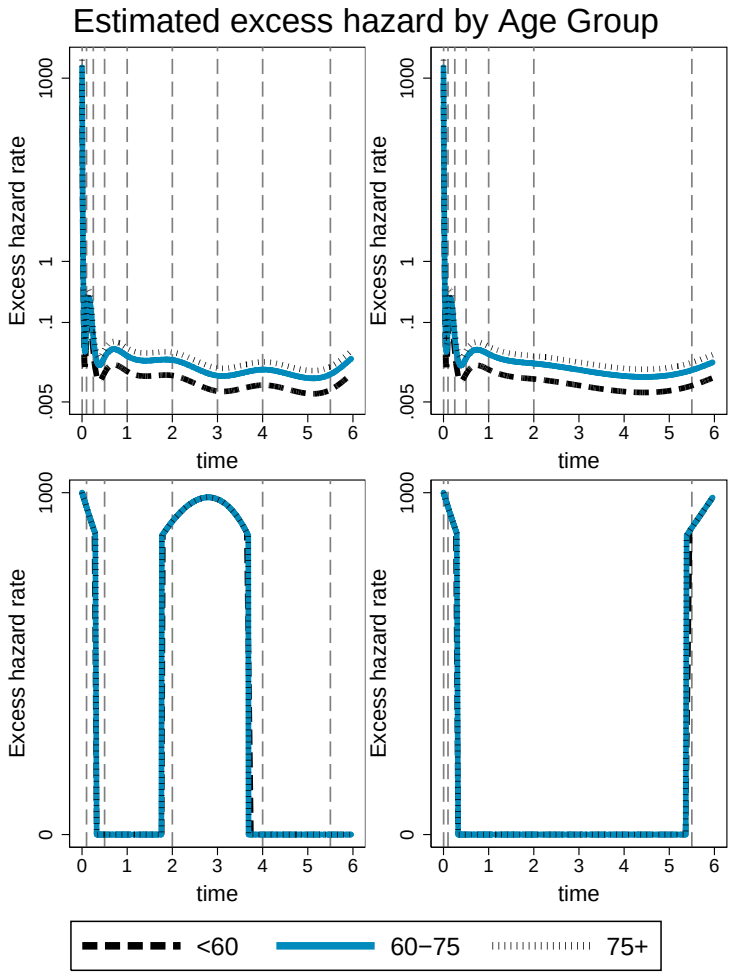


Figure 4.21: Plot of excess hazard curve for the four spline models using short intervals

Figure 4.21 shows the PEH restricted cubic spline models using short intervals. This figure, and all other figures in this spline section, is ordered from the most complex with 9 knots in the top left, 7 knots in the top right, then 5 knots in the bottom left and finally 3 knots in the bottom right of the figure. The bottom row models do not fit and can be disregarded. The top row models show a high excess mortality rate at the start of follow up and a fairly constant excess mortality after six months. This seems feasible, but they also appear over-fitted as the curves are not very smooth, especially for the first six months with several changes in direction.

Figure 4.22 shows the user defined interval results for the PEH model. These plots show a fairly good fit for the 9 and 7 knot models. Whereas the 3 and 5 knot models appear to be very poor even though the estimates obtained and shown in table 4.9 are similar to the

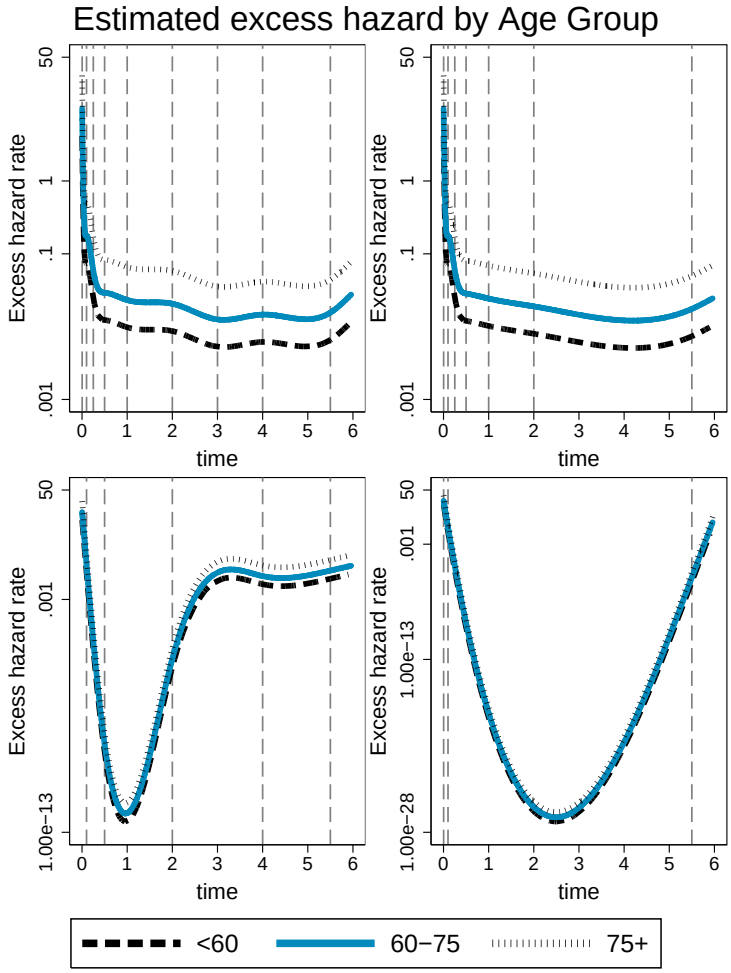


Figure 4.22: Plot of excess hazard curve for the four spline models using user defined intervals

more complex models. It appears as though the addition of more knots in the early time period has been successful here as they are much more behaved than the short interval models. These models assume PEH but there was some evidence that age groups are not proportional in previous models and so this needs to be tested in the next section.

4.8.3 Stage two - non-PEH models

The estimates of the AIC for the non-PEH model for the short intervals and user defined intervals are shown in table 4.10. The non-PEH model produces estimates for the main effects and interactions of age with the spline terms. However the fitted estimates themselves are not comparable, which includes the main effects as they are superseded by the interaction terms and have thus not been shown. Using the AIC values shown in table 4.10 a model with seven knots is preferable for both the user defined and short interval data. A large difference between the models using the two sets of split time intervals was

Number of Knots	AIC	
	Short	User defined
Nine	16419.05	20062.66
Seven	16410.53	20060.42
Five	16822.92	20654.21
Three	16918.70	20674.26

Table 4.10: Table of AIC estimates for the four time dependent spline models using short and user defined intervals

shown for the PEH models.

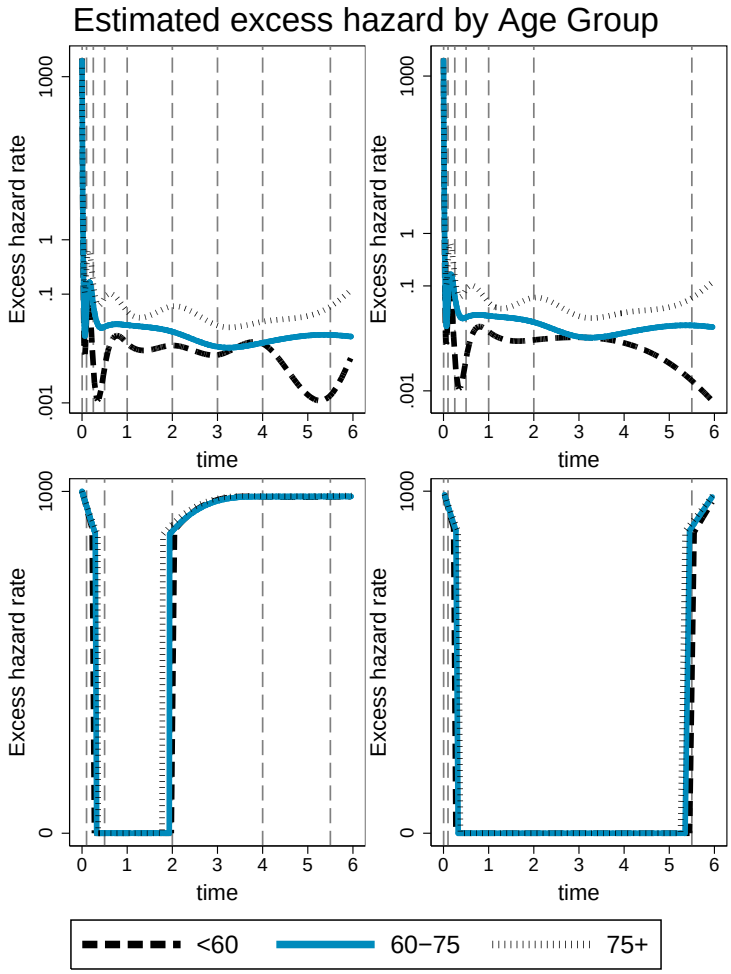


Figure 4.23: Plot of excess hazard curve for the four time dependent spline models using short intervals

The short interval plots are shown in figure 4.23. This plot is quite different to the PEH plots shown in figure 4.21. The 3 and 5 knots models are once again a bad fit to these data, however this time the 7 and 9 knot models also appear to be fairly bad, they are not smooth and appear over fitted. The age groups appear to start low, after the initial drop, and then loop upwards greatly increasing the excess hazard before dropping again in the first six months. This may be due to problems with the models not being able to

turn sharply enough after the large initial drop after the high early excess hazard, which may cause problems in changing direction. The model reported convergence but is clearly not giving sensible results.

The user defined interval models were equally as poor for the 3 and 5 knot models and also overfitted for the 7 and 9 knot models and are therefore not shown.

4.8.4 Stage three - removing the first month

As previously noted the knot positions for the spline models had to change as the first knot was previously contained within the first 30 days of follow-up and this is reflected in the vertical lines on the plots, which indicate the new positions.

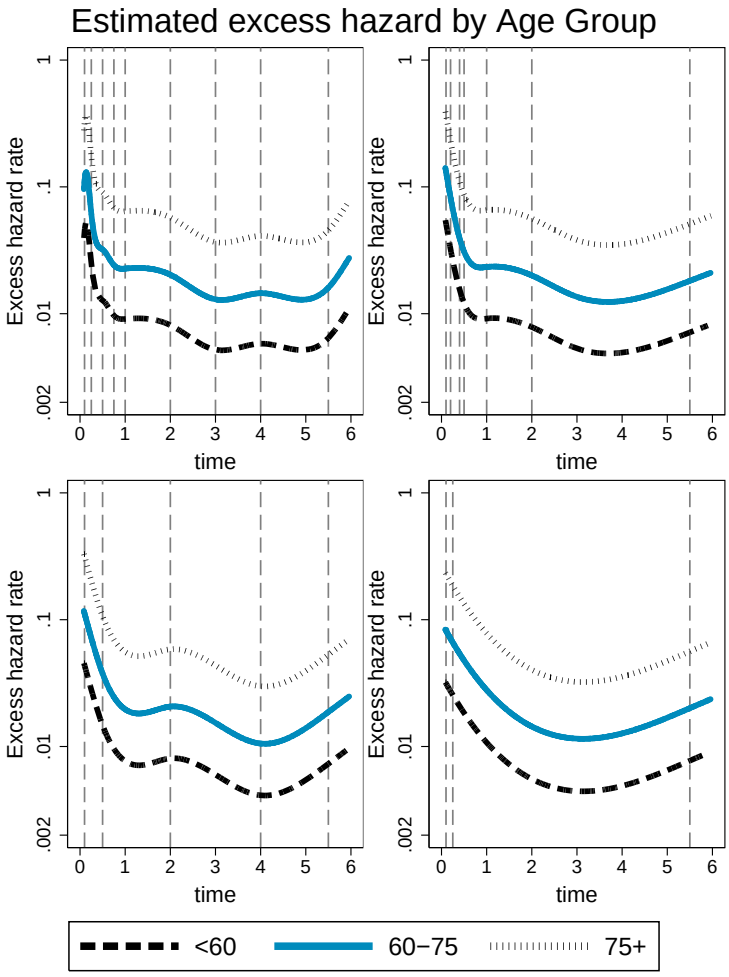


Figure 4.24: Without first month - User intervals proportional excess hazards

The short interval PEH model without the first 30-days is not shown as the results were very similar to the user defined intervals shown in figure 4.24. The figure shows four very similar models with the three knot model producing the worst fit due to the lack

of flexibility where needed. The other three models all appear to be overly influenced by the knot positions which suggests that these are over fitting in some way, even the five knot model changes direction at the knots. The non-PEH models excluding the first 30 days are much worse and are greatly influenced by the knots and are therefore considered unreliable and are thus not shown.

There has been a partial success when removing the first 30 days in the PEH models. However when all of the data were included in the analysis the more flexible models appeared to produce a drop and increase in excess hazard after the initial 30 days as seen clearly in figure 4.21 for the seven and nine knot models. The problem was considered as a possible effect of the restrictions imposed by using restricted cubic splines and forcing linearity in the tails. It was thought that the model was not able to turn quickly enough and so produced a dip before the model settled into a more consistent excess hazard rate. Therefore unrestricted cubic splines, removing the linearity rules, were attempted, but these proved very difficult when attempting to converge models with more than three knots even when using orthogonalised splines and were therefore abandoned.

4.8.5 Stage four - B-Splines

B-splines, proposed by Giorgi *et al.* (38) for use in relative survival, were assessed to see if the models could avoid the potential issues with the large excess hazard during the first 30 days of follow-up. These models were developed for use in R (134) and use individual patient data. However the analysis used here was performed in Stata using individual level split-time data. Quadratic B-splines were adopted, rather than cubic splines, as these were proposed by Giorgi *et al.* (38). For the purpose of comparison the same models used in the restricted cubic splines analysis (with 9, 7, 5 and 3 knots) were fitted.

The short interval data did not fit well in these models and are therefore not shown. Figure 4.25 shows the estimates from the user defined intervals PEH B-spline model. The nine knot model (top left) and the seven knot model (top right) appear to be the better of the four models. They appear to be influenced by the knots used, in that they may be over fitted, and the initial drop in the excess hazard rate appears to fall in two stages, which is not overly feasible. The five knot model (bottom left) appears to have problems estimating the beginning of follow-up and has not removed the strange dipping that the restricted cubic spline models suffered from. The three knot model (bottom right) is overly simple and can be disregarded.

The PEH models produced thus far have been fairly poor, but further investigation may be warranted. Non-PEH models were assessed and were found to be very poor. There were similar problems suggesting that, although convergence was achieved, the results were not sensible, so these are not shown.

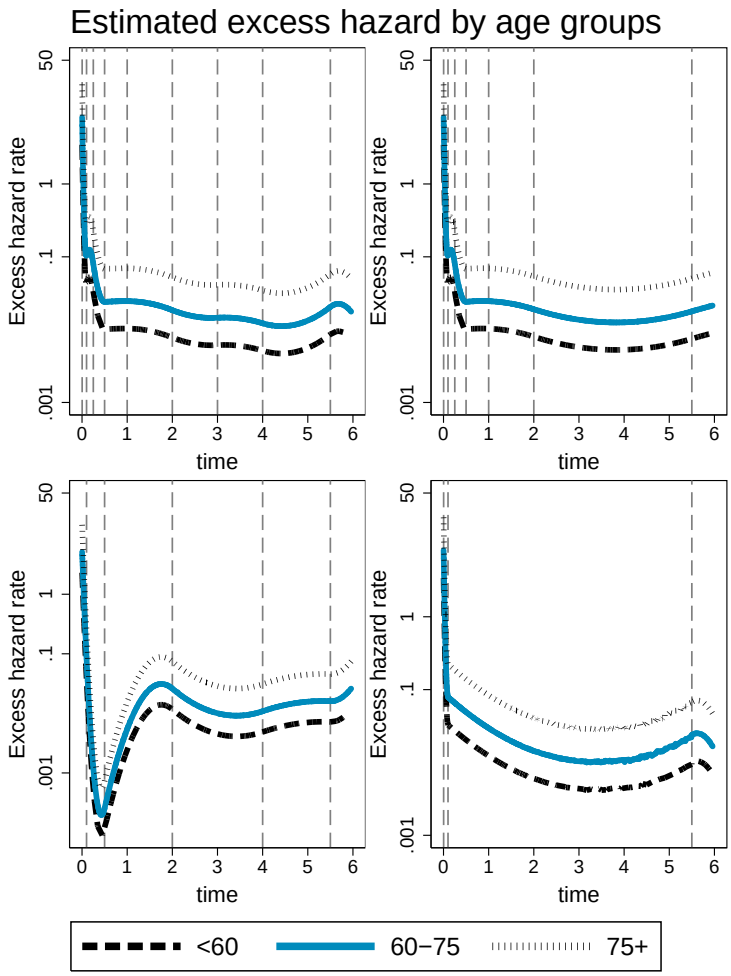


Figure 4.25: Plot of excess hazard curve for the four PEH B-spline models - User intervals

B-splines are arguably worse than restricted cubic splines for these data, although not by much, as dips and bumps appear around the knots and particularly during the large excess hazard experienced in the first few months of follow-up. It was therefore decided that no further models should be fitted using B-splines. However the restricted cubic splines did provide a reasonable fit to these data for the user defined intervals under a PEH assumption. This leads us to the idea of a partitioned model, as splines have thus far had problems estimating a continuous model using all of the data, but after removing the first thirty days this has on occasion proved quite successful. Therefore the following section will fit two restricted cubic spline models to one dataset.

4.8.6 Stage five - Partitioned models

A partitioned model is in effect fitting two models, which are then combined into one. The model will use restricted cubic splines but the first part (0-3 months) will be modelled

using one set of spline terms and the second part (3 months to 6 years) another. The first part of the model has two knots one at 0.001 and 0.1. The second part of the model has knots at $0.3, 2$ and 5. These sets of splines are then modelled within the same GLM model. A key point is that the two parts of the model are not forced to meet at the three month cut-off and by doing this the model is not clinically plausible. This can make the figures quite strange looking. The model split is highlighted by a solid vertical line at 3-months, with the knots indicated by vertical dashed lines as before.

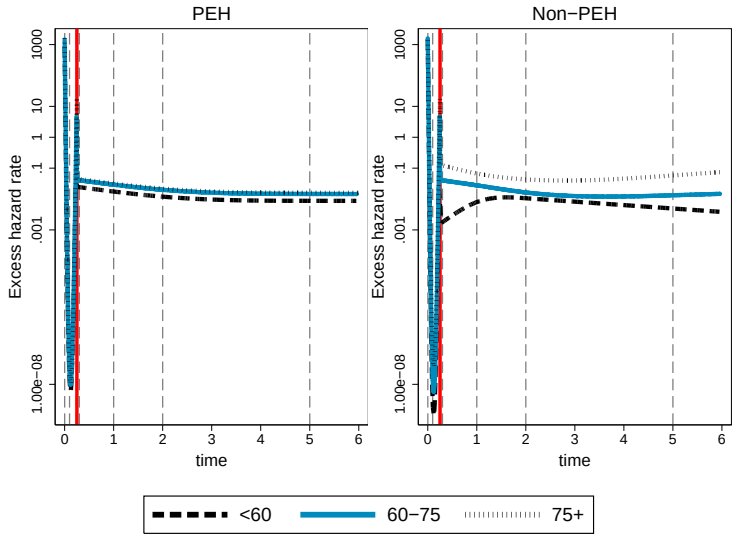


Figure 4.26: Partitioned model excess hazard rates using short intervals

Figure 4.26) shows the short interval dataset results from using a partitioned model, with a PEH assumption (left hand side) and time dependent effects (right hand side) which are fitted using an interaction between the spline terms and age group. The scale of the figure covers a large range and as such the estimates for the second part of the model are not easy to observe. The estimates from the first part of the model are very varied, but the estimate provided by the second part appears reasonable. The non-PEH estimates are quite different for the second part of the model as there is a much greater difference between the age groups.

The user-defined intervals were then modelled. This dataset has more split times in the early follow-up periods, which is now modelled separately, allowing for more information in this part of the model than in the short interval model shown in figure 4.26. The results are shown in figure 4.27 and shows the proportional (left side) and non-proportional (right side) excess hazards model. Both models show a smooth estimate which gradually decreases and then begins to level out in the second part of the model. The estimates seem to be increasing very slightly but this may be due to the lack of information at the end of follow-up rather than a true effect.

The non-PEH plot shows a potential element of over fitting as a different function for each

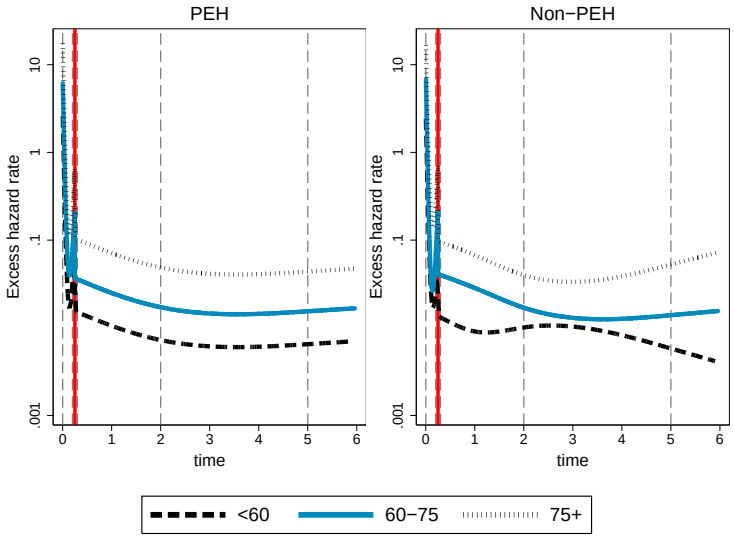


Figure 4.27: Partitioned model excess hazard rates using user defined intervals

Model fitted		Age group 60-75	Age group 75+
Short Intervals	Piecewise	2.355 (1.91 , 2.90)	6.632 (5.43 , 8.10)
	Splines	1.642 (1.31 , 2.06)	1.775 (1.40 , 2.25)
User defined Intervals	Piecewise	2.337 (1.90 , 2.88)	6.232 (5.10 , 7.61)
	Splines	2.338 (1.89 , 2.89)	6.147 (5.02 , 7.53)

Table 4.11: Table of covariate estimates for age groups from piecewise and partitioned spline models

age group is fitted between the two internal knots at 2 and 5 years. It is known that there are few events after five years and as such the knot at this position in a time dependent model may be overly influential. These two plots appear fairly feasible.

A comparison between the partitioned PEH restricted cubic spline models and the estimates from a piecewise model was drawn by comparing the estimates shown in table 4.11. The piecewise estimates are obtained using the short and user defined interval datasets at the individual level unlike the estimates given in the piecewise modelling section (part 4.3) which had only seven intervals. The table shows a fairly large disparity between the results from the partitioned model and those from the piecewise model when using short interval data, where these should be relatively similar if the partitioned model was fitting something sensible. This suggests that there is a potential issue related to the number and location of split-time points, i.e there are not enough split times in the first few months. This conclusion is fairly simple to make as the user defined interval estimates are very similar, which also suggests that this model is useful.

4.8.7 Spline model conclusions

There are clearly problems with model convergence, sensitivity of the model to the knot location and to the frequency of knots. However the restricted cubic splines were found to be much more reliable than quadratic B-splines for modelling these data in a relative survival framework. The short intervals, which has the most splits, were found to be lacking in all of the spline models, although they performed adequately without the first month's data. However, when assessing the partitioned models with short intervals the models did not provide a reasonable fit to the data, nor did it provide reasonable estimates. This suggests that the short intervals proposed by Remontet *et al.* (118) are not overly useful in these data.

The user defined intervals proved to be quite successful when modelling all of the data at the individual level as the model estimates under a PEH assumption were very similar to those from a PEH piecewise approach, as shown in table 4.9. The partitioned models also proved successful as the estimates were very similar to a piecewise approach. These appear to be the useful solution for these data in that the restricted cubic splines on user defined intervals have provided a good fit to these data. However there are still problems associated with using split-time data as only the user defined intervals performed adequately. In chapter 5 a model is developed which does not use split time data, negating this problem.

4.9 Discussion of models

4.9.1 Finding the analysis path

In order to analyse the majority of the available models in relative survival it was essential to provide a comparative analysis for each method. This meant analysing each approach in the same manner. The piecewise models were only used once using the same split-time dataset as these are known to be fairly robust and also fitted very well to these data, with success in practice (2). The continuous approaches needed a more rigorous approach as these methods have many issues that arose during the analysis stage. It was decided that due to the very large mortality experienced by patients in the first 30-days after an ST-elevated myocardial infarction that each method should be assessed by including and excluding these first 30-days. By dropping the first 30-days the aim was to investigate the problem of a very high early mortality rate, but for a final analysis the first 30-days should be included.

4.9.2 Piecewise models

Three models were fitted to the data using a piecewise approach, two at a grouped level and one at the individual level. All three of these models were quick to converge and provided very similar estimates of the excess hazard rate ratios. These models are used frequently in applied research as they are straightforward and can provide fairly reliable estimates of excess mortality and 5- or 10-year relative survival rates, which are very common in practice.

The issues associated with piecewise models are that the hazard rate during each time interval is considered to be constant, which is biologically implausible as the excess mortality is not likely to change only at selected time points. When there are many covariate groups and time dependent effects the number of intervals has to be quite restrictive due to having zero events in subgroups. If PEH is a reasonable assumption then the piecewise models are probably adequate.

4.9.3 Continuous models overall issues

The first continuous models that were adopted fitted polynomials to the data, and these proved to be very poor. Fractional polynomials proved much more successful but still had problems converging. The individual level MFP models were poor as around half of the attempted models when using individual level data did not fit correctly and produced several errors in the model fitting stage. The *difficult* option was adopted frequently as most models would not converge without it, which shows that these models encountered problems during the maximisation process. Even when models converged many provided poor fits suggesting non-convergence, including the MFPs, but especially the polynomials which showed a poor fit in all examples.

Grouped level data proved to be a success in fractional polynomials as models converged quickly and provided believable fits. The problem was with the algorithm selecting a best fitting model from a possible 44 (under an FP2 assumption). Looking at the bootstrapping and model averaging that was performed on the fractional polynomials the uncertainty in both age groups were not correctly estimated and were overly optimistic. However the baseline excess hazard was shown to be a good estimate when using model averaging which is a positive finding. It should be noted that using grouped level data with standard polynomials in the long, short and user defined intervals provided no real improvement over the individual level models.

All models are generally subject to a sensitivity analysis, but no corrections are then made to the differences observed from these when calculating confidence intervals. For example when adopting splines a sensitivity analysis using different knot location and frequency

can provide very different results but a best fitting model would usually be chosen by the analyst without consideration that this is only one of the best fitting models available. Sometimes there is very little difference between the AICs of a model with five knots and four knots but no attempt is made to measure the uncertainty in the final selection of, say, the four knot model over the five knot model. The theory of model averaging could be applied for different knots.

The analysis moved onto splines which once again proved to give very poor estimates, especially when compared to the piecewise estimates in all but the user defined intervals, and for this reason a partitioned model was selected. This provided very good estimates but fairly poor plots, as the figures are in effect split in two and so look very strange. It is for all of these reasons that a new methodology was undertaken as described fully in chapter 5.

4.9.4 Final model selection

The polynomial models fitted were all very poor fits to the data and the individual level MFP models were also fairly poor. MFP models fitted to the grouped level data were fairly successful but it is not possible to assess continuous covariates. This leaves the spline models, which when using user-defined intervals proved to be fairly useful, as shown in table 4.9. The partitioned models using splines gave reasonable estimates when using the user defined intervals. So all models assessed have some issues.

The three datasets used; long, short and user-defined, were all tested using the same fitted models. In almost every case the long interval data was a very poor fit and as such was abandoned part way through the analysis. The short term models fitted well for grouped MFP models but not for others. The user-defined intervals were found to be the most appropriate as they gave reasonable estimates in grouped MFP models and spline models suggesting that the intervals proposed by Remontet *et al.* (118) are not appropriate for all datasets. Naturally a heart disease dataset is different to a cancer dataset in that the majority of information is in the first year of follow-up and this was likely the reason a more CHD tailored dataset proved to be more successful.

Therefore of the available models a spline model with around seven knots fitted to the user-defined intervals appears to be the most appropriate way of assessing the data as continuous.

4.9.5 Further improvements

It was decided not to adopt cure models because of the inappropriate assumptions needed to be able to model these. Cure models (29) assume that at some point the excess mortality

reaches zero, and this is simply not true in heart disease. However it may be possible to model using cure models where the asymptote is not forced to reach zero, rather a level at which the excess hazard becomes constant, this may prove to be interesting and would have great clinical implications if possible. Additive models (114) were also not used but may prove to be useful if investigated.

The literature review (chapter 3) found that there were very few papers which used relative survival in heart disease, this may be due to the problems observed when modelling as shown in this chapter. These few published papers have been supported by an introduction to relative survival by Nelson *et al.* (2) that was described in chapter 2. Simple piecewise models were adopted and this investigation into models has found that these may be the most appropriate models to use above more complex split-time continuous approaches.

The problems observed with the split-time models, including convergence, poor fit, sensitivity of model selection and sensitivity of knot selection has led to the need for a new methodology to be developed. Royston and Parmar (135) proposed a flexible parametric model for survival using splines that can be extended to relative survival as is shown in chapter 5. This approach is modelled on a scale that is less sensitive to knot selection and does not use split time data.

Chapter 5

Flexible Parametric Models for Relative Survival. Using Spline Functions on the Log Cumulative Excess Hazard Scale

5.1 Introduction

The work presented here has been published in the Statistics in Medicine journal (3) (shown in appendix C) and presented at various conferences including the International Society for Clinical Biostatisticians and a young statisticians meeting. The work was also presented in a Stata workshop held at the Karolinska Institutet in Stockholm, Sweden.

5.1.1 Relative survival likelihood

Relative survival $R(t)$ is defined as the ratio of the observed, all-cause, survival $S(t)$ with the expected survival $S^*(t)$ as previously described and defined by Ederer *et al.* (26):

$$R(t) = \frac{S(t)}{S^*(t)} \tag{5.1}$$

So that on the hazard scale the overall hazard rate, $h(t)$ is given by

$$h(t) = h^*(t) + \lambda(t) \tag{5.2}$$

where $h^*(t)$ is the expected mortality/hazard rate and $\lambda(t)$ is the excess mortality rate associated with the disease of interest.

In survival analysis an individuals contribution to the log likelihood can be written as

$$\ln L_i = d_i \ln [f(t_i)] + (1 - d_i) \ln [S(t_i)] \tag{5.3}$$

where $f(t_i)$ is the probability density function associated with $S(t_i)$ and $d_i = 1$ for an event (i.e. death here) and 0 otherwise. Given that the relationship between the distribution, survival and hazard functions are known to be

$$f(t) = h(t)S(t) \tag{5.4}$$

the likelihood can be written as

$$\ln L_i = d_i \ln [h(t_i)] + \ln [S(t_i)] \tag{5.5}$$

For relative survival the survival and hazard functions are defined in equation 5.1 and 5.2. This leads to

$$\ln L_i = d_i \ln [h^*(t_i) + \lambda(t_i)] + \ln [S^*(t_i)] + \ln [R(t_i)] \tag{5.6}$$

As $S^*(t_i)$ is not dependent on any of the unknown model parameters it can be removed from the log likelihood , thus,

$$\ln L_i = d_i \ln [h^*(t_i) + \lambda(t_i)] + \ln [R(t_i)] \tag{5.7}$$

Expected survival for the i^{th} subject is not difficult to calculate, rather it is time-consuming and it requires splitting the data. The importance of equation 5.7 is that the expected mortality is only needed at the time of death. This is easy as the age and year of the patient when they died are all that is required, as well as sex and other matched covariates. This likelihood is general and can be used with various models, including Weibull parametric survival models, fractional polynomials or restricted cubic splines. Parametric models have not been used in relative survival before but it is appropriate for any parametric model where the survival and hazard functions can be written.

5.1.2 Parametric models

There are many parametric models that have been applied in survival data including the Weibull, Poisson, Gompertz and log Normal. However these models have not been used at any length in relative survival with the exception of cure models (29; 28). Weller *et al.* (136) used a Weibull relative survival model but this was based on lifetable data which utilises both splitting and collapsing. Cure models are a special type of parametric model and are the area in which parametric models have been incorporated fully. However the cure models are not appropriate here as they force the relative survival function to plateau which indicates a level of population or statistical cure. This plateau is not feasible for heart disease as patients with a history of MI are at a continued increased risk of mortality when compared to the general population without the disease. Cure models are only really appropriate for certain cancers where it is believed that the excess mortality eventually approaches zero, e.g. colon and rectum cancer (137; 28; 35; 29).

These parametric functions have their benefits, such as the ability to model time as a continuous covariate and not split the data. Continuous data is available for other models if they do not collapse data over covariate patterns, but the splitting of time introduces computational problems. Stepwise estimates utilise a semi-parametric technique as exponential survival times are assumed within the intervals but these are not clinically plausible, whereas fully parametric models give smooth fitted functions, which are useful for prediction and extrapolation. The survival and hazard functions are easily obtainable and the models remain comparable to other parametric models. The piecewise approach used routinely in applied relative survival research can be considered as semi-parametric.

5.1.3 Maximising the likelihood

The likelihood functions for the models were maximised using the Newton Raphson technique with the first and second derivatives estimated numerically as implemented using the `ml` command in Stata (138).

The Newton Raphson technique is essentially a numerical root finder using iterative techniques. An initial value is obtained, which is generally an educated guess, b_0 and this is updated as follows for a one dimension case where b and it's derivative, $g(b)$, are both scalars (see also figure 5.1):

1. Take current guess b_0
2. Calculate the first derivative of b_0 to give $g(b_0)$, which is the gradient
3. Calculate the second derivative of b_0 to give $g'(b_0)$ where $g'(b_0) = (dg/db)$ at b_0. This gives the direction of the slope

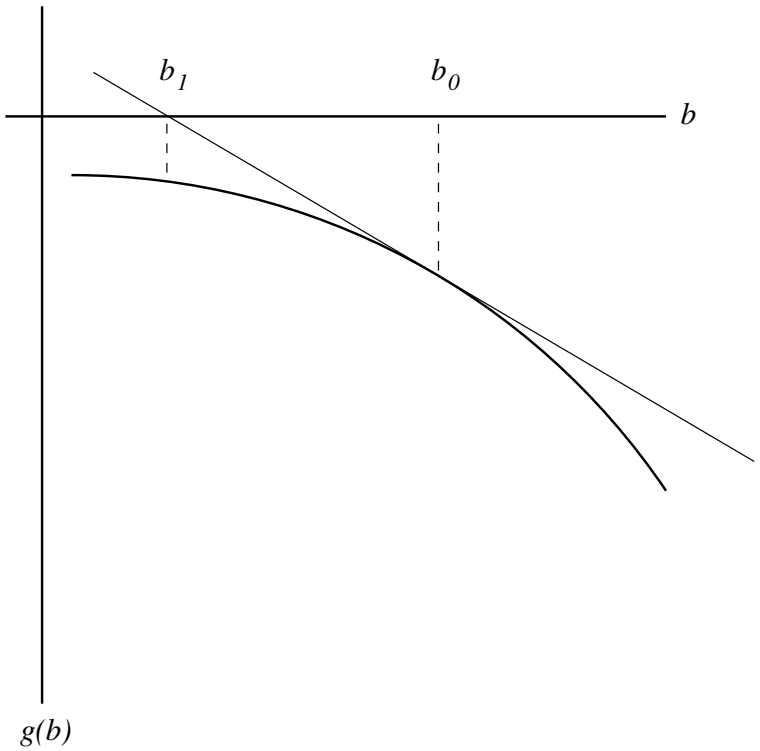


Figure 5.1: One dimension Newton Raphson algorithm.

4. Draw line through point $(b_0, g(b_0)$ with a slope of $g'(b_0)$

5. New estimate, b_1, is the point where the line is zero.

$$b_1 = b_0 - \frac{g(b_0)}{g'(b_0)}$$

6. Repeat steps using new estimate

Now that b_1 is found the process can be repeated

$$b_2 = b_1 - \frac{g(b_1)}{g'(b_1)}$$

and so on until the process converges, which is not guaranteed. There are other maximisation techniques available in Stata (138) but the Newton Raphson technique is used as the default as it is the most consistent method. Stata calculates the likelihood numerically using this technique which has the benefit of being fast and easy to determine.

The default initial values given by Stata were generally found to be fairly poor for the models that were being fitted using the splines. For this reason the first term of the splines, i.e. the linear function of log time was used, which is equivalent to a Weibull model, to obtain a suitable starting point.

5.1.4 Weibull models

The Weibull model is a fairly popular choice for standard survival analysis as it can be written as a proportional hazards model where,

$$h_i(t) = \pi\gamma t^{\gamma-1}\exp(\beta x_i) \tag{5.8}$$

The Weibull model was fitted where the cumulative survival is given by:

$$S(t) = \exp\left(-\pi t^{\gamma}\right) \tag{5.9}$$

with π and γ being the parameters of a Weibull distribution, but given the relationship between the survival function and the cumulative hazard function $[H(t) = -\log\left(S(t)\right)]$ the cumulative hazard function is given as:

$$H(t) = \pi t^{\gamma} \tag{5.10}$$

Modelling on the log cumulative hazard scale would give:

$$\log\left(H(t)\right) = \log(\pi) + \gamma\log(t) \tag{5.11}$$

Here it is clear that the model is fitted as a linear function of log time. Equation 5.8 can also be written as a proportional cumulative hazards model,

$$H_i(t) = \pi t^{\gamma}\exp(\beta x_i) \tag{5.12}$$

This can be extended to relative survival to give the proportional excess hazards as

$$h_i(t) = h^*(t) + \pi\gamma t^{\gamma-1}\exp(\beta x_i) \tag{5.13}$$

and the proportional cumulative excess hazard as,

$$H_i(t) = H^*(t)\pi t^{\gamma}\exp(\beta x_i) \tag{5.14}$$

The cumulative excess hazards model is shown as this scale will be used later in a more

flexible approach. It is important to note that in the Weibull model the excess mortality is monotonic, meaning it can only increase or decrease and cannot change direction.

The total hazard ($h(t)$) of the relative survival model is as follows:

$$h(t) = h^*(t_i) + \log(\pi) + \gamma \log(t) \tag{5.15}$$

This can be substituted into the log likelihood of the relative survival Weibull model, which can be written as:

$$\ln L_i = d_i \ln \left[h^*(t_i) + \log(\pi) + \gamma \log(t)\right] - \ln \left[\pi(t)^\gamma\right] \tag{5.16}$$

The two parameters of a Weibull distribution (π and γ) have to be positive as shown in the definition of a cumulative hazard, that it is only allowed to increase. Here it has been shown how easy it is to fit a parametric survival model, but something more flexible is required.

Covariate	Parameter estimates	95% CI
$\ln(\pi)$	-1.4502	(-1.511 , -1.390)
$\ln(\gamma)$	-1.5465	(-1.602 , -1.491)

Table 5.1: Parameter estimates from a Weibull parametric relative survival model.

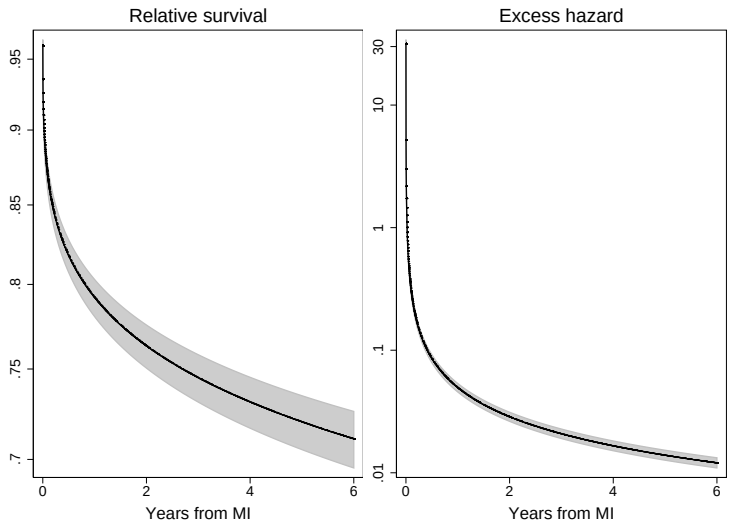


Figure 5.2: Use of the Weibull parametric distribution for relative survival

By fitting a relative survival Weibull model it is possible to obtain the fit shown in figure 5.2 This gave the parameter estimates shown in table 5.1. These values can be used to determine the log excess hazard rate using:

$$\ln(h) = \ln(\gamma) + \ln(\pi) + (\exp(\ln(\gamma)) - 1)\ln(t)$$
$$\ln(h) = -1.5465 + -1.4502 + (\exp(-1.5465) - 1)\ln(t)$$

So if time was taken as one year then the log excess hazard rate = -2.9967 which when exponentiated gives 0.05 which is shown in figure 5.2. The log cumulative excess hazard estimate can be obtained by:

$$\ln(-\ln(R(t))) = \ln(\pi) + \exp(\ln(\gamma))\ln(t)$$
$$\ln(-\ln(R(t))) = -1.4502 + (\exp(-1.5465))\ln(t)$$

Which gives an answer of -1.4502 at one year. To convert this back to the survival scale and obtain relative survival estimates the following is used:

$$RS = \exp(-\exp(\ln(\pi) + \exp(\ln(\gamma))\ln(t)))$$
$$RS = \exp(-\exp(-1.4502)) = 0.791$$

This is also shown in figure 5.2.

Model	Covariate	Excess Hazard ratio	95% CI
Piecewise	Aged 60-75	2.490	(2.01 , 3.08)
	Aged 75+	7.633	(6.24 , 9.34)
Weibull	Aged 60-75	2.442	(1.99 , 2.99)
	Aged 75+	6.822	(5.63 , 8.27)

Table 5.2: Excess hazard rate ratios from a piecewise and Weibull parametric relative survival model

Table 5.2 shows the estimates of the excess hazard rate ratios from a standard piecewise approach and from a PEH Weibull model. Here the estimates are fairly similar for both age groups but there is potentially more uncertainty in the Weibull model, shown by the width of the confidence interval. However the standard error also increases with the hazard ratio and these estimates are broadly similar.

Parameter	Covariate	Parameter estimates	95% CI
π	Aged 65-75	2.420	(1.97 , 2.97)
	Aged 75+	6.752	(5.56 , 8.21)
γ	Aged 65-75	1.026	(0.84 , 1.25)
	Aged 75+	1.143	(0.95 , 1.37)

Table 5.3: Parameter estimates from a Weibull parametric non-proportional excess hazards relative survival model.

There is only one example of the use of a Weibull model in relative survival that is published (136). The difference between the Weibull model adopted here and the Weller *et al.* model

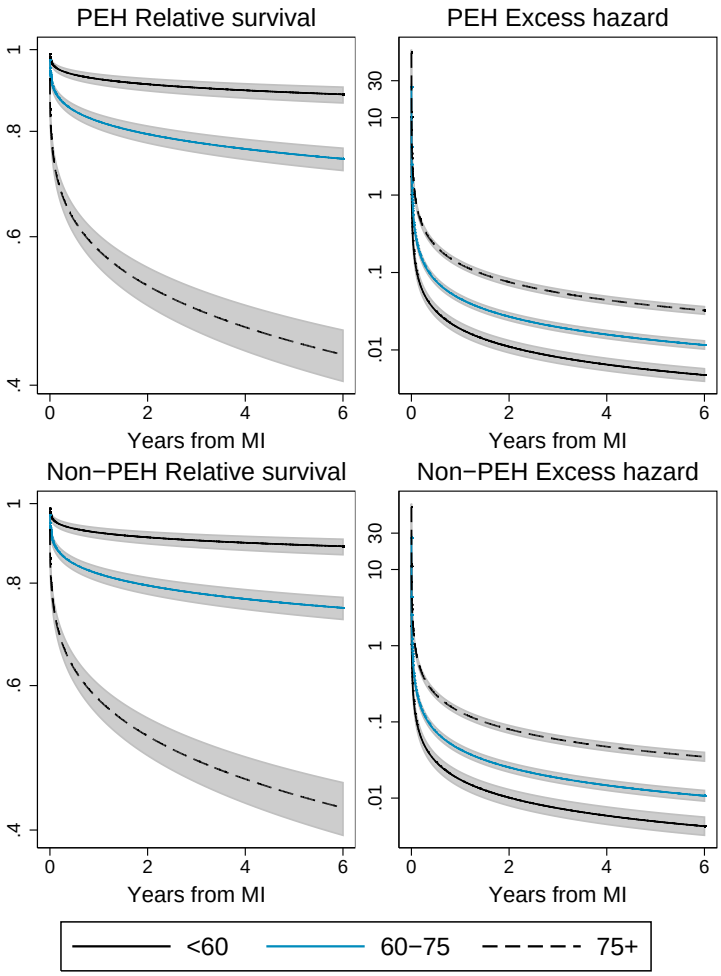


Figure 5.3: Use of covariates and the Weibull parametric distribution for relative survival

is the use of non-split time data to allow a continuous time covariate. These models can be extended to include covariates under both proportional and non-proportional excess hazards assumptions. To fit a proportional model a covariate is fitted to the baseline parameter (π) and for a non-proportional model to the time dependent parameter (γ). These models have been fitted using age groups as covariates and the results are shown in figure 5.3. These estimates appear to be the same but there are very small differences between them.

Table 5.3 shows the parameter estimates for the non-PEH model. This shows that the time dependent effects modelled using the γ parameter are both very small and non-significant suggesting no evidence of non-proportionality. These can be interpreted as a 2.6% and 14.2% unit increase for both the middle and oldest age groups respectively, when compared to the youngest age group. The shapes fitted are smooth and are fitted to non-split time data which are in turn the advantages hoped to achieve by using parametric models over the models already used in practice.

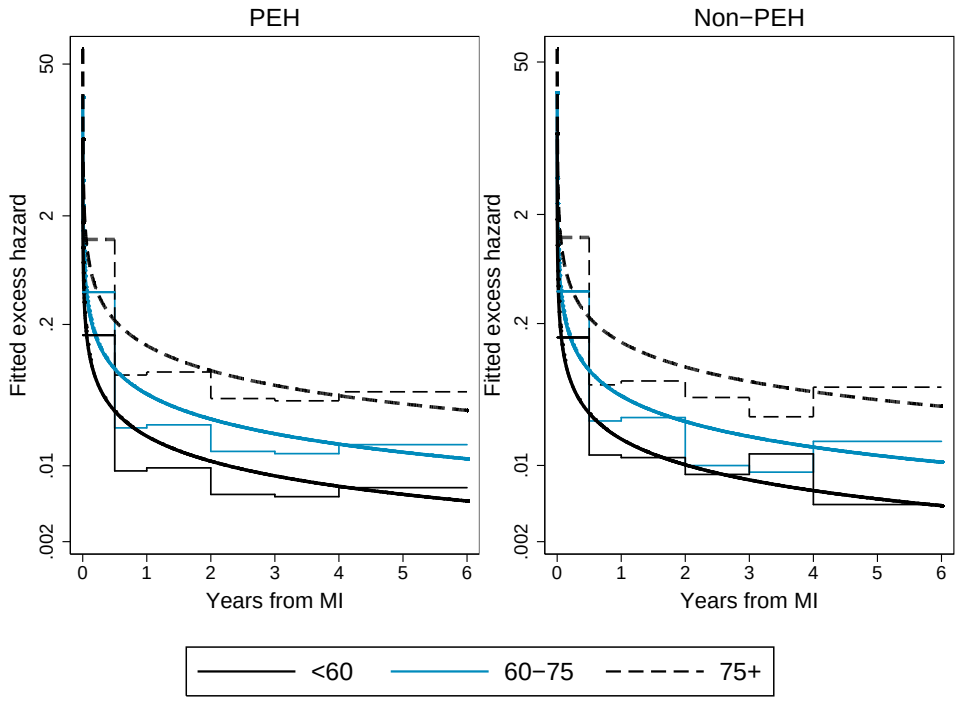


Figure 5.4: Piecewise model comparison with Weibull models for the excess hazard

Figure 5.4 shows a comparison of the Weibull model and a piecewise model using a Poisson model on individual level split-time data. It is fairly clear that the shape of the Weibull is not entirely feasible in that it has a continual reduction in excess mortality where the piecewise models appear to flatten out after around a year of follow-up. This shape is fairly inflexible and is forced upon the model. As stated earlier the excess hazard rate in a Weibull model is monotonic and cannot change direction, which may not occur in this dataset but may in others. It performs reasonably well given the fairly restrictive nature of the curve but is not fitting the data well.

The Weibull model is fairly limited in the shape of the hazard that the model can capture and it is necessary to allow more complex functions to be modelled. One option would be to adopt a more complex parametric model, like the Gamma or generalized F (139), but there are often convergence issues with these. The modelling of proportional excess hazards is also required and a model that has PEH as a special case, even if the underlying shape of the hazard is complex, is generally desirable for application in relative survival. The advantage of the Weibull model is that it can be fitted quickly (*the non-PEH model converged in 3.1 seconds*) and easily, but it is simply not flexible enough.

The non-proportional models shown in figure 5.3 highlight that the assumption of PEH is valid, which is supported by the parameter estimates for γ that are non-significant. However, as determined in figure 5.4, these models do not fit the model well and that the

non-PEH model should be investigated using a more flexible approach.

Equation (5.16) shows that the model is just a linear function of log time on the log cumulative hazard scale and this can be generalised to be a non-linear function of time using splines, as proposed in all-cause survival by Royston and Parmar (135), to be more flexible than the Weibull model and more reliable than the more complex parametric functions like the Gamma model.

5.1.5 Problems with current methods

There have been several approaches proposed to model the excess mortality, $\lambda(t)$. The most commonly used in applied research is a piecewise approach wherein there is a need to split the timescale as shown in section 4.3. Hakulinen and Tenkanen (40) estimate the model based on grouped data using a generalised linear model (GLM) framework with a binomial assumption for the number of observed deaths. Esteve *et al.* (27) adopted a full likelihood approach based on individual level data. Dickman *et al.* (115) proposed a GLM with a Poisson error structure using exact survival times and individual level data which is equivalent to the Esteve *et al.* full likelihood approach. Models can also be fitted at the grouped level (i.e. collapsed over categorical covariate patterns). Relative survival models have been extended to more flexible models such as MFPs (shown in section 4.6) (39) and cubic splines (118) that also use split-time data. B-splines have also been suggested for relative survival to model the hazard ratio as a flexible function of time by Giorgi *et al.* (38) available in R (134) as have restricted cubic splines by Bolard *et al.* (37), both using non-split time data.

The individual level piecewise approach yields several observations for each individual in the dataset as there is a record for each subdivision of time. This can have computational implications in very large datasets. For example Lambert *et al.* (39) used grouped data when fitting fractional polynomial models due to the computation time needed for individual level models. A disadvantage of using grouped data is that continuous covariates need to be categorised, which is not a preferred practice due to the loss of power and subjectivity in choice of cut-points (140). Piecewise models rely on a subjective choice of the number and location of split points and lack clinical and biological plausibility. Using a piecewise approach also implies generation of many parameters for time dependent effects. If a dichotomous covariate is perceived to vary over time then there is a need to create $[h-1]$ more parameters where h is the number of splits in the timescale. This can lead to problems in *small* datasets for example in producing cells with zero events. It is possible to simplify this by making assumptions, for example that excess hazards are proportional after 2 years. B-splines (38), like the restricted cubic spline approach proposed by Bolard *et al.* (37), obtain the survival function using numerical integration greatly increasing computation time. All current approaches model on the log excess hazard scale.

For full details on the problems associated with the current methods available see chapter 4.

5.2 Methodology

5.2.1 The use of splines

In section 4.8 splines were used to model age groups, and splines will be used again here. When the functional relationship between a response, Y, and an explanatory covariate, X, is non-linear it is possible to assess this in many ways, including the use of polynomials and other transformations. One method would be to adopt piecewise polynomials, also called regression splines. These can be defined as piecewise polynomials of degree n whose function and the first $n-1$ derivatives agree at the points where these polynomials join, known as knots. The most common regression spline is the cubic spline, where the functional form is visibly smooth with a cubic polynomial fitted between each knot, but with continuous first and second derivatives. When a fixed number of knots are used it is possible to obtain parameter estimates using standard linear regression in common statistical packages, however the fitted model is dependent on the number and location of the knots selected.

Regression splines (37; 118; 141) are part of a family of smoothing functions encompassed by the term *splines*. B-splines (38; 134) are also used as they can be considered as more computationally efficient. A related non-parametric technique that falls into this family are smoothing splines. The difference between these and regression splines is in the knot placement. Smoothing splines place a knot at each data point while including a penalty term in the likelihood to control the smoothness of the fitted curve, which leads to overly complex estimation methods and will not be used here.

Restricted cubic splines were first proposed by Stone and Koo (133) which were further discussed by Durrleman and Simon (142) and are a simple extension of regression splines, or standard cubic splines (143). The key difference is that the restricted cubic splines are forced to be linear in the tails, before the first knot and after the last. There is little information in the tails of the data and this could lead to over-fitting in standard cubic splines. These two boundary knots can be considered as the external knots, with the knots in between the internal knots. Restricted cubic splines are generally better behaved than standard regression splines and have shown to be fairly robust to knot selection.

The splines used in section 4.8 used split time data and were fitted to the log excess hazard scale.

5.2.2 Royston and Parmar flexible parametric model

In standard all-cause analysis survival data are often modelled using a Cox proportional hazards model, which has the advantage of estimating covariate effects as log hazard ratios without the need to estimate the baseline hazard. However, the behaviour of the hazard function may be of medical interest because it is directly related to the time-course of an illness. The baseline hazard rate can also help in understanding the natural history of the disease through the way the hazard rate changes over time. Selected parametric models such as the Weibull model are an alternative to the Cox model. However, these models are not as popular in practice due to concerns over the restrictions imposed on the shape of the hazard function.

A flexible parametric model using restricted cubic splines was proposed by Royston and Parmar (135) for censored survival data. The advantages of this approach include the ability to model the baseline hazard which may hold clinically important information. The restricted cubic splines offer greater flexibility in the shape of the hazard function when compared to standard parametric models. In addition time by covariate interactions are easy to include and non-proportional models are easy to implement and assess. One of the key advantages of the method is that the survival and hazard functions are obtained analytically, speeding up computational time over methods adopting numerical integration. The application of the method has so far been limited (144; 145; 146) although it is available to run using Stata (147).

Royston and Parmar (135) suggested using a more general transformation of the survival function based on a suggestion by Aranda-Ordaz (148) and extended by Younes and Lachin (149) where

$$g\left[S(t) \mid \theta\right] = \ln\left(\frac{S(t)^{-\theta} - 1}{\theta}\right) \tag{5.17}$$

Where a proportional hazards ($\theta \to 0$) or proportional odds ($\theta = 1$) model is available when θ meets the required conditions. Further extensions to allow non-proportional effects of some or all of the covariates (other values of θ) are available and can be entered by the user or estimated from the data. The disadvantage of such an approach is that covariate effects are difficult to interpret, although it is still possible to transform to a more interpretable scale. The method proposed here is an extension to the Royston and Parmar method used in standard survival analyses to relative survival. The flexible parametric approach adopted models on the log cumulative excess hazard scale, which can have advantages in terms of model stability.

5.2.3 Adapting to relative survival

The current methods fit models on the log excess hazard scale i.e.

$$h(t) = h^*(t) + \exp(\beta x) \tag{5.18}$$

where β is a vector of parameters to be estimated for covariates x . If the hazard function given in 5.2 is considered then integrating both sides gives the cumulative hazard $H(t)$

$$H(t) = H^*(t) + \Lambda(t) \tag{5.19}$$

where $H(t)$ is the cumulative overall hazard $H^*(t)$ is the cumulative expected hazard and $\Lambda(t)$ is the cumulative excess hazard. The flexible parametric approach adopts models on the log cumulative excess hazard scale, which can have advantages in terms of model stability. The hazard and survival function are obtained analytically and non-split time data is used, i.e. a single record per subject. The models are initially restricted to only one time scale. Modelling on the log cumulative excess hazard scale for relative survival analysis leads to:

$$\ln(-\ln R(t;x)) = \ln(\Lambda(t)) = \ln(\Lambda_0(t)) + \beta z \tag{5.20}$$

where $\Lambda_0(t)$ is the baseline cumulative excess hazard function and $R(t;x)$ is the relative survival function. Note that proportional cumulative excess hazards implies proportional excess hazards. The Weibull is a linear function of log time and this is a stable function so, following Royston and Parmar, the notation $z = \ln(t)$ is used and the log cumulative excess hazard is written as:

$$\ln(\Lambda(t)) = s(z;\gamma) + \beta x \tag{5.21}$$

where $s(z;\gamma)$ can be any smoothing function such as the various regression spline functions or fractional polynomials. For a proportional excess hazards model, let $\eta = s(z;\gamma) + \beta x$. Then the associated relative survival and excess hazard functions are given by

$$R(t) = \exp\left[-\exp(\eta)\right] \tag{5.22}$$

and

$$\lambda(t) = \frac{1}{t}\frac{ds(z;\gamma)}{dz}\exp(\eta) \tag{5.23}$$

This implies that the overall survival is

$$S(t) = S^*(t)\exp\left[-\exp(\eta)\right] \tag{5.24}$$

with associated hazard

$$h(t) = h^*(t) + \frac{1}{t}\frac{ds(z;\gamma)}{dz}\exp(\eta) \tag{5.25}$$

Substituting 5.22 and 5.25 into 5.7 gives the log-likelihood contribution for each subject,

$$\ln L_i = d_i \ln\left[h^*(t_i) + \frac{1}{t_i}\frac{ds(z_i;\gamma)}{dz_i}\exp(\eta_i)\right] - \exp(\eta_i) \tag{5.26}$$

As is shown in equation 5.26 the expected survival, $S^*(t)$, is not in the likelihood. Time by covariate interactions are easy to include by modelling γ separately for each covariate level, which removes the assumption of proportional excess hazards which are investigated in section 5.4. The model thus far has been generalisable as any smoothing function could be used for $s(z;\gamma)$, but for this approach the flexible parametric model for censored survival data that was proposed by Royston and Parmar (135) was followed which incorporates the use of natural/restricted cubic splines in models on the cumulative hazard or cumulative odds scale (modelling on the cumulative excess hazard scale is considered here). Following Royston and Parmar the splines are defined by

$$s(z;\gamma) = \gamma_0 + \gamma_1 z + \gamma_2\nu_1(z) + \ldots + \gamma_{m+1}\nu_m(z) \tag{5.27}$$

where the j^{th} basis function is defined for $j = 1,\ldots,m$ as:

$$\nu_j(z) = (z-k_j)_+^3 - \lambda_j(z-k_{min})_+^3 - (1-\lambda_j)(z-k_{max})_+^3 \tag{5.28}$$

where $u_+ = u$ if $u > 0$ and $u_+ = 0$ if $u \leq 0$, k_{min} and k_{max} are the boundary knots where k_{min} is the position of the first knot, k_{max} of the last knot and $\lambda_j = \frac{k_{max}-k_j}{k_{max}-k_{min}}$. These two equations (5.27 and 5.28) are standard restricted cubic splines as were adopted in section 4.8. The derivative of the spline function is needed to calculate the excess hazard rate. However these are simply calculated as follows,

$$\frac{ds(z;\gamma)}{dz} = \gamma_1 + \sum_{j=2}^{m} \gamma_j \left[(3z - k_j)_+^2 - 3\lambda_j(z - k_{min})_+^2 - 3(1-\lambda_j)(z - k_{max})_+^2\right] \quad (5.29)$$

Modelling is performed on the log cumulative excess hazard scale as the survival and hazard functions can then both be estimated analytically and numerical integration is not required, which speeds up computation time. It is more likely to retain stability; for example, when modelling on this scale using log time and a Weibull distribution the relationship will be linear and as such departures from the Weibull will be under more stable conditions. This model does not use time-split data and expands upon a current method for *standard* survival. The models are fitted in a Stata program called `strsrcs`, described in chapter 6. Note that proportional cumulative excess hazards implies proportional excess hazards.

5.2.4 Knot location

Degrees of Freedom	Knot Position*
2	50
3	33 67
4	25 50 75
5	20 40 60 80
6	17 33 50 67 83

Table 5.4: Default knot positions. *Percentiles of all death times, as proposed by Royston and Parmar

The placement of the internal knots is an issue, however, knot selection does not appear critical for a good fit. Over-fitting with too many knots is undesirable for use in prediction in that the fitted curve may follow small, fairly unimportant, features of the data. Furthermore, uncertainty in parameter estimates due to knot optimisation should be reflected in enlarged standard errors and confidence intervals, but this is not straightforward to do. Following Royston and Paramar (135) the boundary knots were placed at the extreme uncensored log survival times, k_{min} and k_{max}. For internal knots the centile-based positions given in table 5.4 were selected. The positions given are essentially those adopted by Royston and Parmar and recommended by Durrleman and Simon (142). The rationale for choosing knots not too far from the median uncensored log survival time is to allow the data to be most closely modelled in the region of greatest density and hence usually of lowest variance.

The key factor is that the values given in table 5.4 are based on percentiles of survival time for those who experience an event, i.e. for those who died during follow-up. For these data the majority of events are within the first 6 months and as such many of the models

will have the majority of the knots within one year where the greatest density of events occurs.

5.2.5 Initial application

Covariate	Parameter estimates	95% CI
Baseline (γ_0)	1.548	(1.170 , 1.926)
Spline 1 (γ_1)	0.688	(0.624 , 0.752)
Spline 2 (γ_2)	0.017	(0.013 , 0.021)
Spline 3 (γ_3)	-0.001	(-0.006 , 0.004)
Spline 4 (γ_4)	-0.017	(-0.025 , -0.008)

Table 5.5: Parameter estimates from a flexible parametric relative survival model.

A model with four degrees of freedom (knots at 25, 50, 75) can be fitted using this methodology without any covariates to give the results shown in table 5.5. The method produces four parameter estimates ($\gamma_1 - \gamma_4$) that are associated with the variables created by the model and these are shown in the table. The linear predictor is for the log cumulative excess hazard. These results can be used to obtain the relative survival estimate by transforming from the log cumulative excess hazard scale to the survival scale:

$$RS(t) = \exp(-\exp(\gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 + \gamma_3 z_3 + \gamma_4 z_4)$$
$$RS(t) = \exp(-\exp(1.548 + 0.688 z_1 + 0.017 z_2 - 0.001 z_3 - 0.017 z_4)$$

The excess hazard is more complex as it uses the derivatives of the spline functions as defined earlier in equation 5.29.

$$\lambda(t) = \exp(\ln\left(\tfrac{1}{t}\right) + \ln(\gamma_1 z_1' + \gamma_2 z_2' + \gamma_3 z_3' + \gamma_4 z_4') + (\gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 + \gamma_3 z_3 + \gamma_4 z_4))$$

$$\begin{aligned}\lambda(t) = \exp(&\ln\left(\tfrac{1}{t}\right) + \ln(0.688 z_1' + 0.017 z_2' - 0.001 z_3' - 0.017 z_4') \\ &+ (1.548 + 0.688 z_1 + 0.017 z_2 - 0.001 z_3 - 0.017 z_4))\end{aligned}$$

Figure 5.5 shows the scale that the data is modelled on, the log cumulative excess hazards scale , which was shown in table 5.5. This plot also shows that the complex hazard function is approximately straight for the most part on the log hazard vs. log time scale (right hand side). If this was straight then the Weibull distribution would be a reasonable model to use. This figure also helps to explain why the splines are calculated on the log time scale as it is much harder when using splines to capture the shape on the left-hand side (time) of figure 5.5 when compared with the right-hand side (log time).

Figure 5.6 shows the fitted estimates of the relative survival and excess hazard functions

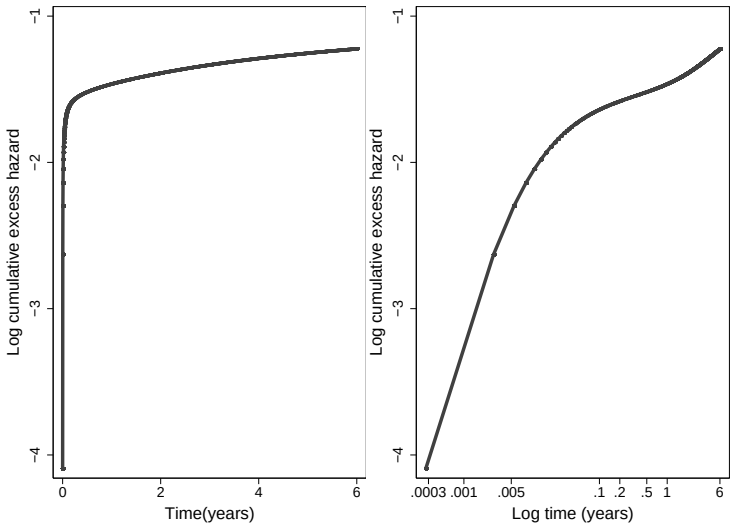


Figure 5.5: Plot on the log cumulative excess hazard scale.

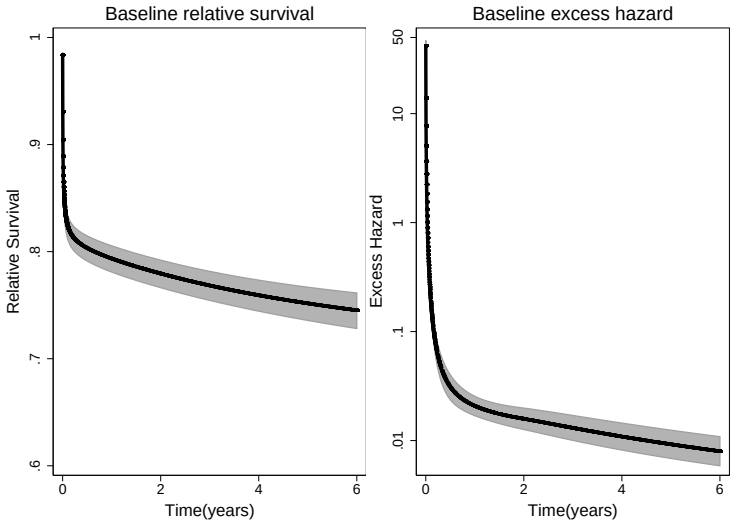


Figure 5.6: Plot of the baseline relative survival and excess hazard rate.

with confidence intervals for the baseline. These plots show a smooth function that can be compared to the Weibull models (shown earlier in figure 5.2). The confidence intervals are obtained using the δ method which is described and tested in section 5.3.6. The relative survival plot decreases at a slower rate after a year than the Weibull model and the excess hazard turns quicker than the Weibull model after the initial very high mortality at the beginning of follow-up.

This application used the default knot positions for a 4df model (default centile positions shown in table 5.4) but naturally a different number of knots could have been used at different knot positions if required. This highlights the potential problem with using

splines as they can be dependent on the knot selection and in order to assess this sections 5.3.1 and 5.3.2 investigate the model sensitivity to knots under a PEH assumption with non-PEH models also assessed later on.

Under the assumption of proportional excess hazards, covariates can be fitted in this model using spline functions for the baseline cumulative excess hazard. For example when assessing age groups it is possible to obtain log excess hazard ratios from this model as a baseline effect under the assumption that the age groups have proportional excess hazards. The linear predictor for the derivatives shares the same coefficients as for the linear predictor on the log cumulative hazard scale.

5.2.6 Comparisons with piecewise models

Piecewise example

A piecewise model was fitted using individual level split-time data with six time intervals on the time scale split annually for the 6 year follow-up, except for the first year which was split into two 6 month intervals and years 5 and 6 which were combined. Using a proportional excess hazards (PEH) assumption gave estimates of the excess hazard ratios (with 95% confidence intervals) as $2.48(2.01, 3.08)$ and $7.63(6.24, 9.34)$ for $60-75$ and 75 years or older respectively using < 60 age group as the comparative baseline. Both age groups have a significantly increased risk of death over the baseline group with the oldest age group being the worst.

This can be compared directly with the results from the spline approach as is shown in figure 5.7. This figure shows the estimates of the excess mortality rate from both a proportional and non-proportional excess hazards approach using both the standard piecewise and spline model approach with 4 degrees of freedom on the top row and 6 degrees of freedom on the bottom row. Using a proportional excess hazards (PEH) assumption gave estimates of the excess hazard ratios (with 95% confidence intervals) for the 4df spline model as $2.37(1.92, 2.91)$ and $6.34(5.20, 7.73)$ for $60-75$ and 75 years or older respectively using < 60 age group as the comparative baseline. These results compare favourably with those from the piecewise model listed above. The difference between the piecewise and the spline model is likely due to the split points used in the piecewise model as this is known to influence the estimates as seen in other chapters (4 and 7). In chapter 7 it is shown that piecewise models with insufficient intervals can lead to biased excess hazard ratios.

In both of the proportional models the estimates from both approaches appear very similar for the majority of the model with the main difference being in the first six months of follow-up. In this early time period the piecewise approach is not flexible enough to be able to estimate the very large excess hazard experienced in the first thirty days following

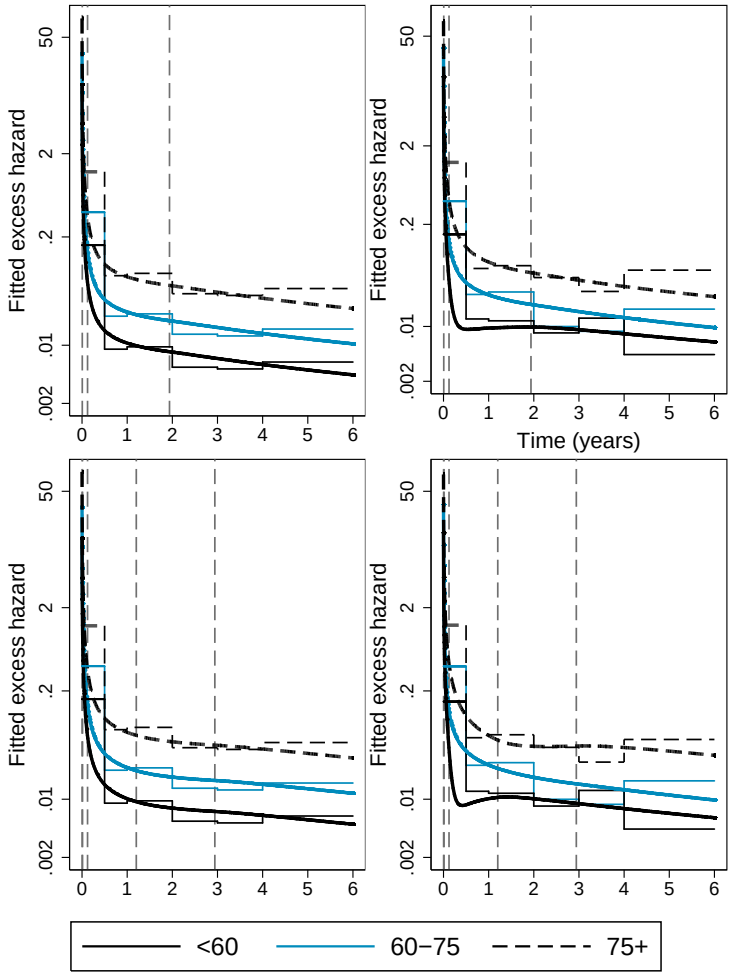


Figure 5.7: Comparison of the fitted excess hazard rates for the selected spline model using non-split data with the piecewise individual level split-time data for PEH (left) and non-PEH models (right) for both a 4df (top) and 6df (bottom) model. Internal knots are shown by vertical dashed lines. It is not possible to distinguish the first 2 knot lines.

an MI or the sharp drop that occurs after these initial 30 days. This is due to the model being forced to be constant for this time period which could be resolved by creating more split-times which would increase the number of parameters. For a PEH model increasing the number of parameters to try and capture this information is not too difficult but with non-PEH the number of parameters increases rapidly. The spline based models appear to capture the same shape from the piecewise model but are also flexible enough to capture the information available from the first six months.

The non-PEH models are more complex and will be discussed in detail later but it is worth comparing the estimates from the spline model to the piecewise model when the assumption of time dependent covariates is introduced. The bottom right plot shows a strange effect in the spline model in that it appears to drop and increase again in the first six months of follow-up which appears quite unnatural. This *dipping* effect may be

an artifact of the splines or it may be true but will be discussed later in this chapter. The splines appear to be forming the same underlying pattern as the piecewise model but naturally the splines are smooth.

The plots in figure 5.7 have shown that for these examples the spline based models compare well with the current standard approach to relative survival.

5.2.7 Orthogonalising the splines

It is possible to create the splines to be orthogonalised basis functions using the Gram-Schmidt orthogonalisation procedure (150). For a restricted cubic spline model with k knots there are k derived covariates $x_0 \dots x_{k-1}$, obtained using equation 4.28. forming a $N \times x$ design matrix X. In order to orthogonalise these sets of covariates a new set of covariates $s_0 \dots s_{k-1}$ are computed. Let s_i be the transformation of the i^{th} derived covariate x_i. The j^{th} transformed covariate, s_j, is obtained by regressing x_j on $s_0 \dots s_{j-1}$ and then obtaining the predicted values $\hat{x_j}$. The s_i are obtained as follows.

$$\begin{aligned} s_0 &= x_0 \\ s_i &= x_i - \hat{x_i} \end{aligned} \tag{5.30}$$

All basis functions higher than the first (linear) function are forced to be uncorrelated and have mean 0 and standard deviation 1. The linear function is also uncorrelated with the higher-basis functions and this leads to models converging where they would have previously had difficulty due to the spline terms being highly correlated.

Orthogonal variables are useful to achieve numerical accuracy for highly collinear variables. Stata's estimation commands can face much collinearity and still produce accurate results. However these commands will drop variables because of collinearity. If it is known with certainty that the variables are not perfectly collinear, as in the case of these spline terms, then retaining all the effects in the model is only possible by producing a set of orthogonal variables.

5.3 Proportional Excess Hazards Investigation

5.3.1 Sensitivity of knot frequency

In order to determine the potential effects of the splines the general arguments put against splines need to be addressed. These are mainly that spline models can be dependent on the frequency and locations of the knots selected which is in itself entirely subjective; unlike in

the MFP models proposed by Lambert *et al.* (39) which determine the best fitting models by way of an algorithm, although these models currently split the time scale.

The first area to investigate would be the number of knots selected as this is most likely to have an effect in terms of under and over fitting a model to the data. The proportional excess hazard models used here include only age groups. This could be included as continuous, but for the purpose of illustration in this sensitivity analysis groups were used splitting age into < 60, $60-75$ and ≥ 75 years old. The models were fitted using `strsrcs` which will be described in chapter 6.

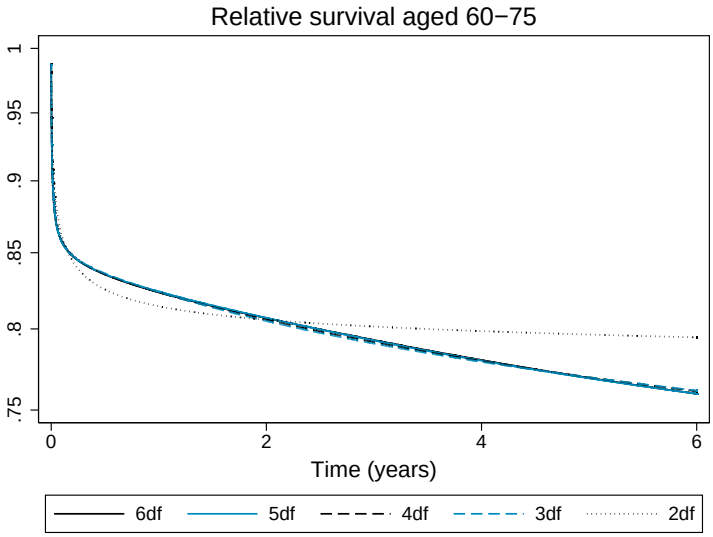


Figure 5.8: Sensitivity of the relative survival estimates to the number of knots using STRSRCS

Degrees of Freedom	Knot Position (time in years)
2	0.122
3	0.011, 1.202
4	0.005, 0.122, 1.964
5	0.003, 0.025, 0.598, 2.576
6	0.003, 0.011, 0.122, 1.202, 2.942

Table 5.6: Knot positions on the timescale

The five models all fit knots at the positions shown in table 5.4 which relate to the points in time shown in table 5.6. Here it is clear that all of the models put all internal knots before three years of follow-up with the majority of knots within the first year. This is expected as the majority of deaths are within a year and these knots are fitted as the percentiles of events.

The plot shown in figure 5.8 gives the estimated relative survival for the middle age group for various degrees of freedom ranging from 2 to 6 which includes 1 internal knot up to 5. As the model is proportional it is not necessary to show all three age groups. The baseline

excess hazard will be shown later.

Figure 5.8 shows that there is only a small difference between the models with 3 to 6 degrees of freedom. The only model that is distinctly different is the 2 df model which places only one internal knot at the 50th percentile, i.e. the median of all death times. But even with this amount of flexibility the model is still not overly different in the general interpretation that would be given. However by adding just a second knot the model appears to fit as well as a model with five knots. It is clear however that the models fitted are all providing smooth fits to these data with no deviations, due to small and unimportant changes in the data, as would be expected in over fitted models.

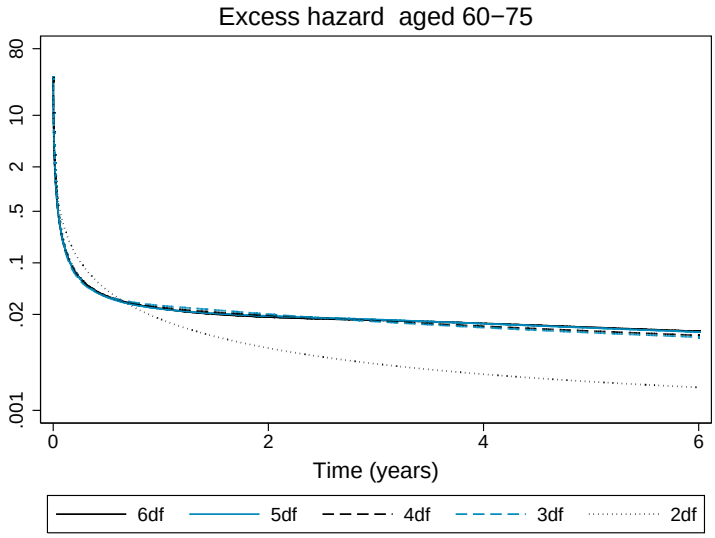


Figure 5.9: Sensitivity of the excess hazard estimates to the number of knots using STRSRCS

It is also possible to investigate how the estimates of the excess hazard are influenced by the change in frequency of knot points. The plot shown in figure 5.9 allows for a similar interpretation to that given for figure 5.8. It should be noted that this plot is on the log scale, as are all excess hazard plots as the differences in fits are apparent when the excess hazard is small and as such this information is lost on the untransformed scale. All models with more than one knot give a very similar fit for the excess hazard as would be expected given the known functional relationship between survival and hazard. This confirms that the PEH models are fairly robust to changes in the frequency of knots selected.

The fitted PEH model provides estimates of the excess hazard rate ratios for the age groups modelled adopting the youngest age group as the baseline and these estimates are shown in table 5.7 along with the knot positions used and the AICs from the PEH.

The spline knots are defined by the percentiles of all death times, as were adopted by Royston and Parmar in the standard survival paper (135) from which this extension to

Spline Model DF	Proportional Excess Hazards			
	Location of knots*	AIC	Age group	
			60-75	75+
2	50	3316.54	2.390(1.92, 2.97)	6.410(5.21, 7.89)
3	33 67	3121.97	2.356(1.91, 2.90)	6.252(5.13, 7.63)
4	25 50 75	3120.22	2.356(1.91, 2.90)	6.252(5.13, 7.63)
5	20 40 60 80	3121.65	2.356(1.91, 2.90)	6.259(5.13, 7.63)
6	17 33 50 67 83	3123.42	2.357(1.91, 2.90)	6.261(5.13, 7.64)

Table 5.7: AIC and estimates of the excess hazard rate ratio with 95% confidence intervals for spline models with various degrees of freedom using non-split data. * Percentiles of all death times, as proposed by Royston and Parmar.

relative survival is taken. Using the AIC, a model with 4 degrees of freedom would be selected as the best fitting model but as is clear by the estimates, all models provide a good fit. There is very little variation between the estimates of the hazard ratio for the $60-75$ age group and even the 2df model gives the same answer to one decimal place including a confidence interval that is almost as narrow as the best fitting model. There are minor differences between the estimates of the oldest age group excess hazard ratios but all are similar and give near identical confidence limits.

This sensitivity analysis is fairly straightforward to do and would be recommended as a method to determine the robustness of the final model selected. Here it is clear that nearly every model selected, excluding the model with a single knot, has given a similar estimate of the relative survival and excess hazard rates and estimates of the excess hazard rate ratios for the covariate analysed.

5.3.2 Sensitivity of knot location

It has now been shown that the frequency of the knots is not important in determining the correct model, and that using the AIC criterion a sensible model is selected. This can be assessed further by using knots with four degrees of freedom (the 4df model was selected by the AIC previously) but not placed at the default percentile points, i.e. not at $25, 50$ and 75.

Five variations on the default knot positions were used at $(15, 25, 80)$, $(20, 70, 90)$, $(15, 45, 85)$, $(50, 70, 90)$ and $(15, 30, 50)$. These were selected based on two reasons, the first was to potentially cause problems for the splines by selecting knots at extreme positions, i.e. all knots before or after the 50^{th} percentile. The second was to assess if changes to the default values $(25, 50, 75)$ that covered the whole range of the data would cause problems also. It would be safe to assume that changing the default values to $(20, 45, 70)$ would have a very minor affect on the estimates. The predicted relative survival using these knot positions and the default knot positions using 4df are shown in figure 5.10 for the middle age group. As before due to the proportional excess hazards assumption there is no need to view the

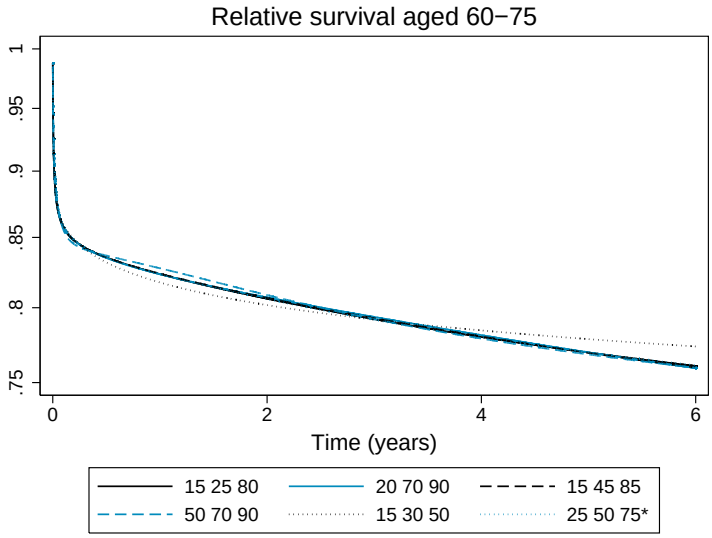


Figure 5.10: Sensitivity of the relative survival estimate to the location of knots using STRSRCS in a proportional model. *The default knot location using 4 degrees of freedom

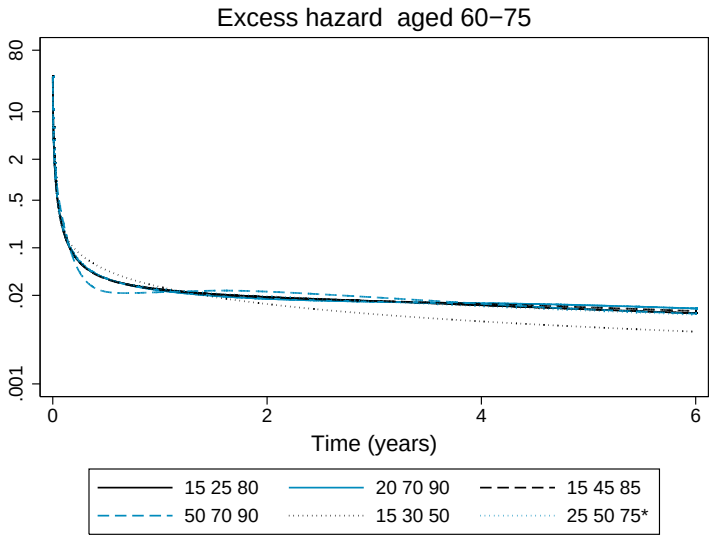


Figure 5.11: Sensitivity of the excess hazard estimate to the location of knots using STRSRCS in a proportional model. *The default knot location using 4 degrees of freedom

other two age group results.

Figure 5.10 shows the fitted relative survival estimates from the six models which vary by their knot location. From this plot it is clear that there are two sets of knots that differ to the other four. The first is the $(50, 70, 90)$ model, which places it's knots at approximately 1.5 months, 1.5 years and 4 years. This means that there are no knots in the first month and a half meaning that for the majority of the information there is no knot. This is not a sensible model to fit as the first internal knot is placed at the median with the emphasis

Proportional Excess Hazards			
Location of		Age group	
Knots position*	AIC	60-75	75+
15 25 80	3119.76	2.356(1.91, 2.90)	6.254(5.13, 7.63)
20 70 90	3119.98	2.358(1.92, 2.90)	6.266(5.14, 7.64)
15 45 85	3119.85	2.356(1.91, 2.90)	6.258(5.13, 7.63)
50 70 90	3187.58	2.343(1.90, 2.88)	6.276(5.15, 7.65)
15 30 50	3138.33	2.366(1.92, 2.92)	6.247(5.11, 7.64)
25 50 75	3120.22	2.356(1.91, 2.90)	6.252(5.13, 7.63)

Table 5.8: Comparing the model sensitivity to knot location using three knots. * Percentiles of all death times, as proposed by Royston and Parmar.

on the final half of the model where it is known that this is not the case for patients following an MI. However even in this supposedly poor model the actual fit is fairly good and very similar to the other, more sensible, models. The model with knots placed at $(15, 30, 50)$ covers the start of follow-up and may be expected to provide a better fit than the $(50, 70, 90)$ model. However this model appears to be the most different model of all, even if it is still very similar to the other five fits. The knots are placed at approximately days one, three and forty-five.

The excess hazard rate estimates are shown in figure 5.11 for these models, which also supports the previous findings that there is little difference between all of the models. However the difference between the suspect model, $(50, 70, 90)$, and the other models is made slightly more apparent in the early stages of follow-up. The estimate for this model drops further than the other five and then turns more sharply in order to model the remainder of the model better. The previous good-fitting four models are fairly indistinguishable in the plot and as such suggest that with a sensible model the results are robust, and even with a seemingly poor model the fit is still fairly good. The model with knots at $(15, 30, 50)$ is also different but in this case it is different at the end of follow-up as the model has no more internal knots later on to model this data appropriately.

Table 5.8 shows the excess hazard rate ratios for patients aged $60 - 75$ and 75+ produced by the five models and AICs. The AIC shows that the worst fitting model is clearly the suspect model, $(50, 70, 90)$, that was shown to be different in figures 5.10 and 5.11, with the other suspect model being the next worse. The best fitting model is not the model that uses the default values but rather one with knots at $(15, 45, 85)$. The estimates for both age groups are all fairly similar across the models with even the worst two models providing very similar estimates which suggests that even silly knot selection is not too bad, although this would not be recommended in practice.

This has shown that for a proportional excess hazards model the location of the knots has very little impact on the final estimates and even when a poor model is selected, in terms of poor knot selection, the estimates are still fairly reliable. This now suggests that under a PEH assumption the flexible parametric models are fairly robust to changes in both

knot frequency and knot location as well as being comparable to the standard piecewise approach. The AICs are all remarkably similar and one would not be overly concerned with any of the models except the two models that are suspect of potentially poorer fits.

5.3.3 Random knot placement

In order to fully justify the use of splines, a model with three internal knots was fitted twenty times to the data with each model using three knots that were randomly allocated using the univariate distribution. The only limitation used was in the lowest knot point being above the 15^{th} percentile as values lower than this caused the model convergence to fail. Also the number of iterations was limited to ten meaning that some of the models may not fully converge, this was to ensure that the loop could complete with twenty observations. To prevent knots being placed at the same time point an adjustment of +1 was made to a matching knot.

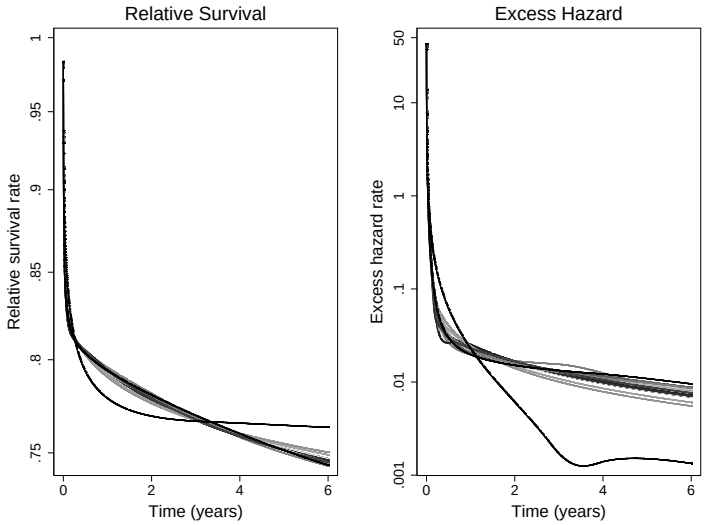


Figure 5.12: Sensitivity of the relative survival estimate to the location of random knots.

Knot positions			
17,50,62	40,42,79	36,83,89	24,38,72
27,59,78	28,38,64	47,62,63	26,53,77
22,40,55	45,57,86	32,60,89	82,88,89
49,54,64	35,46,83	32,35,81	24,53,84
20,26,59	37,58,94	37,61,70	15,36,88

Table 5.9: Random knot locations.

Figure 5.12 shows the fitted relative survival and excess hazard rate for all twenty models selected. It should also be noted that no covariates were modelled and that these are the baseline results. The knots that were randomly selected are shown in table 5.9. As is

clear in the figure, nineteen of these models fitted a very similar response with one being particularly bad. The model that is very poor is the model with the internal knots placed at $(82, 88, 89)$ and is clearly not a sensible model to be fitting. This model has convergence issues and when fitted separately fails to converge so this estimate is of no genuine concern. However other models appear fine and as listed in table 5.9, these are quite varied and some would not be a sensible choice in practice.

5.3.4 Proportional Odds

Royston and Parmar (135) looked at proportional odds models alongside proportional hazards models. The proportional odds assumption for covariate effects was found to be more appropriate than the proportional hazards assumption in two of their example datasets and as such may prove of interest in the LRI dataset. The log-logistic distribution has the following survival and hazard functions:

$$S(t) = \frac{1}{1 + (\pi t)^{\gamma}} \quad h(t) = \frac{\pi\rho(\pi t)^{\rho-1}}{1 + (\pi t)^{\rho}} \tag{5.31}$$

Where π and ρ are parameters. The log-logistic distribution can be extended to include a vector of covariates (x) by reforming the survival and hazard functions as follows:

$$S(t; x) = \frac{1}{1 + (\pi t)^{\rho}\exp(\beta x)} \quad h(t, x) = \frac{\pi\rho(\pi t)^{\rho-1}\exp(\beta x)}{1 + (\pi t)^{\rho}\exp(\beta x)} \tag{5.32}$$

The ratio of the hazards of any two individuals with covariate patterns x_1 and x_2 approaches 1 when the value of t becomes large. Therefore the the impact of a covariate on survival decreases with time. The log odds function is linearly related to the log of time. A general class of proportional odds models have a distribution function such that the odds function is

$$\frac{1 - S(t; x)}{S(t; x)} = g(x)f(t) \tag{5.33}$$

where $g(x)$ is a function of x but not of t and $f(t)$ is a function of t but not of x (151). Proportional odds models have been advocated in situations where the hazard ratios are expected to decrease over time. For example in clinical trials where a treatment is evaluated with long follow-up time.

The log-logistic distribution is a linear function of log time on the odds scales as is shown by:

$$\log\left[\frac{1-S(t;x)}{S(t;x)}\right] = \rho\log\pi + \rho\log(t) + \beta x \tag{5.34}$$

where ρ is the slope and the intercept depends on βx. In the same way that the Weibull model was extended to use splines on the log cumulative excess hazards scale, the log-logistic model can be extended using restricted cubic splines on the log odds scale.

5.3.5 Proportional odds for relative survival

There are similarities between the proportional odds and the proportional hazards models. Equation 5.34 is a linear function and this is replaced by a spline function. Splines are used for the baseline effects whilst assuming a linear predictor in both models. The difference is in the use of a different link function to transform to the scale that is being modelled on. To model proportional excess odds the following survival function was used

$$S(t) = \frac{1}{(1+\exp(s(z;\gamma)+\beta x))} \tag{5.35}$$

where $s(z;\gamma)$ is shown in equation 5.27 with the hazard rate given by

$$h(t) = \frac{1}{t}\left[\frac{ds(z;\gamma)}{dz}\right]\exp\left[(s(z;\gamma)+\beta x))\,\frac{1}{1+\exp(s(z;\gamma)+\beta x)}\right] \tag{5.36}$$

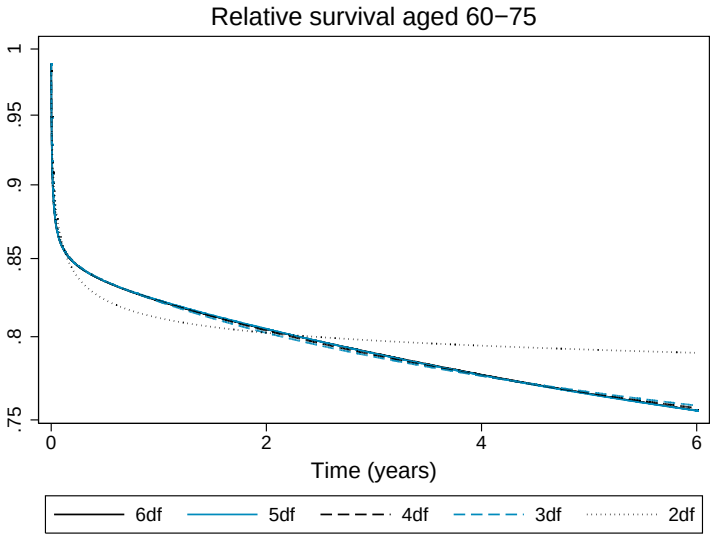


Figure 5.13: Sensitivity of the relative survival estimate to the location of knots using STRSRCS in a proportional odds model.

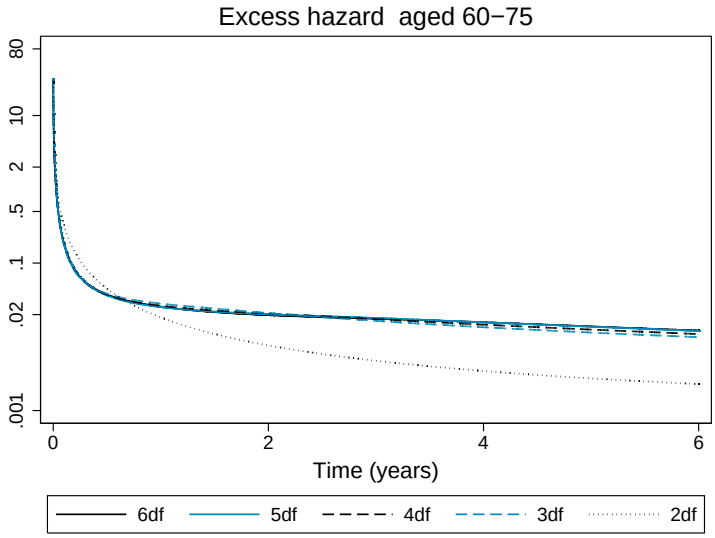


Figure 5.14: Sensitivity of the excess hazard estimate to the location of knots using STRSRCS in a proportional odds model.

The proportional hazards model is well known, but the proportional odds model for survival data, potentially a competitor to the Cox model, also has a fairly long history. It was first described in a semi-parametric framework by Bennett (152) and was later adapted into the flexible parametric framework by Royston and Parmar (135). The feasibility of the proportional odds model in relative survival comes into question when considering the assumption of proportional excess odds (PEO).

Figure 5.13 shows the estimates of relative survival from models with varying degrees of freedom, ranging from 2df to 6df as with the PEH models. Here proportional excess odds are assumed. The most noticeable points come in comparison to figure 5.8 in that the shape of the plot is very similar for all models, even the 2df model fit is almost identical and that would be the worst fitted model here. This goes to suggest that while the idea of proportional excess odds may not be a sensible one the models themselves appear to fit something that may be of some use. They may be a suitable back-up if the hazard scale models fail to converge.

Similarly the estimates shown for the excess hazard rate are shown in figure 5.14. A direct comparison with the PEH model shown in figure 5.9 also shows a clear similarity in the estimates produced by this PEO model. All but the 2df model produced very similar results for the excess hazard rate and even the 2df model may potentially provide sensible estimates even though it is different.

Table 5.10 shows the estimates from these PEO models and it is only here that a difference between these estimates and the PEH estimates is found, but these estimates are not comparable directly as these are odds ratios of excess mortality. In comparison to those

Spline Model DF	Proportional Odds			
	Location of knots*	AIC	Age group	
			60-75	75+
2	50	3298.83	2.554(2.03, 3.21)	7.936(6.35, 9.92)
3	33 67	3126.54	2.538(2.04, 3.16)	7.728(6.23, 9.59)
4	25 50 75	3126.44	2.537(2.04, 3.16)	7.726(6.22, 9.59)
5	20 40 60 80	3127.88	2.539(2.04, 3.16)	7.738(6.23, 9.60)
6	17 33 50 67 83	3129.84	2.539(2.04, 3.16)	7.738(6.23, 9.60)

Table 5.10: AIC and estimates of the excess hazard rate ratio with 95% confidence intervals for spline models with various degrees of freedom using non-split data on the odds scale. * Percentiles of all death times, as proposed by Royston and Parmar.

shown in table 5.7 these estimates are all higher for both age groups and confidence intervals are also wider for all models and all ages.

The parameters shown in the model can be interpreted as the odds of mortality (due to the disease) up to any time, t, are approximately 2.5 times higher in the middle age group than in the youngest age group used as the baseline. The 4df model has the lowest AIC and would be selected for use if PEO models were being used in a true application rather than a sensitivity analysis. The AICs are all higher than the PEH AICs, suggesting that these models are not as useful and are more difficult to explain.

The use of proportional odds in standard survival analysis is fairly limited, though it is potentially a very useful technique. In order to change current habits and for proportional odds models to be used regularly in practice it would be essential for there to be a clear concise list of advantages gained by using the approach. This is especially true in cases where proportional (excess) hazards may not be appropriate. One argument in trials is that a single number is required to describe a treatment effect, but surely it is more important to give an estimate on the scale of most relevance even if it does change over time. However there are no examples in published literature of the use of proportional odds in relative survival and this is likely due to the added complexity in interpretation along with the lack of evidence pertaining to the models usefulness when modelling excess mortality.

5.3.6 Bootstrapping vs. the delta method

The δ method (153) is an approximation for the mean and variance of a non-linear function when there is a known mean and variance for a random variable(s) x. It is a useful approximation when the coefficient of variation of x is small. Denote the function of x by y, where $y = f(x)$ so that

$$\text{var}(y) \approx \left(\frac{dy}{dx}\right)^2_{x=E(x)} \text{var}(x) \tag{5.37}$$

If y is a function of two variables, x_1 and x_2:

$$\text{var}(y) \approx \left(\frac{\partial y}{\partial x_1}\right)^2 \text{var}(x_1) + 2\left(\frac{\partial y}{\partial x_1}\right)\left(\frac{\partial y}{\partial x_2}\right)\text{cov}(x_1, x_2) + \left(\frac{\partial y}{\partial x_2}\right)^2 \text{var}(x_2) \qquad (5.38)$$

This can be extended to any number of variables. In Stata `predictnl` (154) can be used for non-linear predictions after fitting a model. The δ method requires the use of differentiation and this is performed numerically in `predictnl`. This method of approximation is known as the δ method.

By contrast the other way of obtaining estimates of the variance is by bootstrapping. The standard error of the sample mean is found as the standard deviation of the hypothetical population of sample means obtained by repeatedly drawing a sample of a given size. As samples can be repeatedly drawn the mean's of each sample can be found and the standard deviations of these samples computed. This is a bootstrap estimate of the standard error of the mean and can be used in more complex situations such as this. Equation 5.38 can be extended to many variables and the δ method used to find non-linear functions of parameters estimated from a model as the parameter estimate and variance-covariance matrix are known.

The derivatives of complex functions can be obtained accurately using numeric methods using standard Stata functions (154). The log excess hazard function is a non-linear function of the parameter estimates, the spline function and their derivatives, i.e.

$$-\log(t) + \log\left[\frac{ds(z;\gamma)}{dz}\right] + [s(z;\gamma) + \beta x] \qquad (5.39)$$

This is how the software implements the estimation of confidence intervals. If the standard error of the relative survival function is estimated on the cumulative hazard scale $[\log(-\log)]$, i.e. the scale modelled on, then this is a linear function of the parameter estimates and there is no need to use the δ method for the relative survival confidence intervals as these can be obtained by back transformation.

The δ method and bootstrap method were used to determine if the much quicker δ method provides a good estimate of the baseline hazard and variance compared to the computationally intensive bootstrap approach. The bootstrapping for the 6df model took over ten and a half hours to finish for 1000 replications whereas the δ method took forty-five seconds.

The plots shown in figure 5.15 show the estimates for a 2df(top), 4df(middle) and 6df(bottom) model under the PEH assumption. In these plots are the estimates from the delta method,

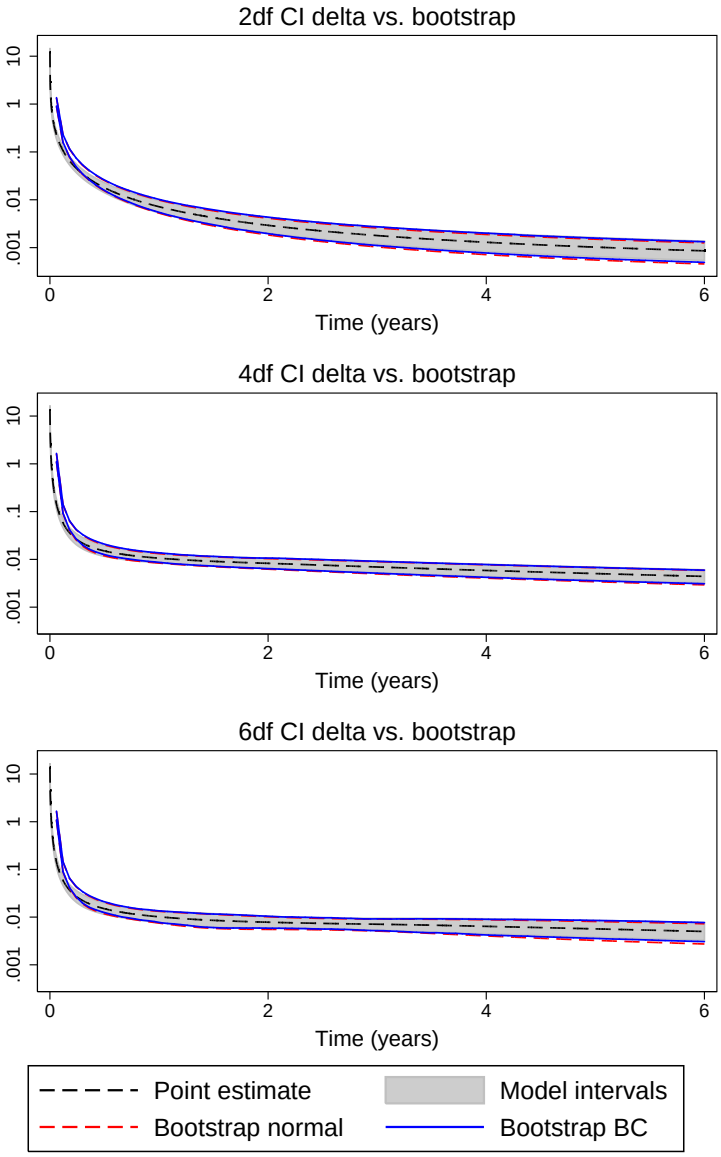


Figure 5.15: Comparison of the confidence interval estimates using the δ method and bootstrapping with both normal and bias corrected intervals.

indicated by the top row in the legend and the bootstrap estimates indicated by the bottom row of the figure legend. There are two bootstrap estimates, normal and bias corrected estimates of the confidence intervals. All three estimates are near identical for the majority of the model with some discrepancy during the first few months.

Table 5.11 includes estimates of the excess hazard rate ratio confidence intervals from the standard method used by the program (δ) and those obtained by bootstrapping (bias-corrected) for the three models shown in figure 5.15. The point estimates, by definition, will be identical, but the confidence intervals are also very similar. The only difference is that the bootstrap estimates of the confidence intervals are slightly wider than the δ

DF used	Method used	60 − 75 years old	75 years or older
6 df	δ method	2.357 (1.91 , 2.91)	6.261 (5.14 , 7.63)
	Bootstrap	2.357 (1.91 , 2.90)	6.261 (5.13 , 7.64)
4 df	δ method	2.356 (1.91 , 2.91)	6.252 (5.13 , 7.62)
	Bootstrap	2.356 (1.91 , 2.90)	6.252 (5.13 , 7.63)
2 df	δ method	2.390 (1.92 , 2.98)	6.410 (5.22 , 7.87)
	Bootstrap	2.390 (1.92 , 2.97)	6.410 (5.21 , 7.89)

Table 5.11: Comparing excess hazard rate ratios using the δ method for obtaining standard errors to bootstrapping

method estimates. The estimates are very similar and as such there is no reason to doubt the δ method as a useful estimator.

It should be noted that in smaller datasets there may be more problems than were experienced here, but these models would not generally be used on much smaller datasets. In contrast the area in which these models have the potential to be used regularly is cancer, and these datasets are generally of a very large size in the form of national registries. The use of bootstrapping becomes impractical in these large datasets.

5.4 Non-PEH models

In a proportional model with one covariate (x_0) and j spline terms (z_j) where for simplicity $j = 2$ the log cumulative excess hazard is written as:

$$\ln(\Lambda(t)) = \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 + \beta x_0 \tag{5.40}$$

However if the assumption of proportional excess hazards is not reasonable then the coefficients associated with the spline variables can be modelled as time dependent by allowing them to vary by covariate values giving a log cumulative excess hazard of:

$$\ln(\Lambda(t)) = \gamma_0 + (\gamma_{10} + \gamma_{11} x_0) z_1 + (\gamma_{20} + \gamma_{21} x_0) z_2 + \beta x_0 \tag{5.41}$$

This is still a linear predictor on the log cumulative excess hazards scale. Following this everything proceeds as before under a PEH assumption. It is possible to transform the linear predictor to the relative survival scale and to the excess hazard scale. These then feed into the likelihood as before. The log cumulative excess hazard shown in equation 5.41 can be generalised for the i^{th} covariate with j knots to give:

$$\ln(H(t)) = \gamma_0 + \Sigma((\gamma_{j0} + \gamma_{j1} x_i) z_j) + \beta x_i \tag{5.42}$$

End of interval	60 − 75 years old	75 years or older
6 months	2.644 (2.09 , 3.34)	8.263 (6.61 , 10.33)
Year 1	2.053 (0.72 , 5.87)	4.398 (1.51 , 12.85)
Year 2	2.330 (1.07 , 5.10)	5.044 (2.24 , 11.36)
Year 3	1.206 (0.31 , 4.65)	5.075 (1.66 , 15.55)
Year 4	0.680 (0.16 , 2.97)	2.189 (0.55 , 8.65)
Years 5 and 6	3.794 (0.96 , 15.03)	11.914 (2.97 , 47.72)

Table 5.12: Estimates of the excess hazard rate ratio with 95% confidence intervals for the individual level piecewise non-proportional excess hazard model using split-time data.

Extending the model to non-proportional excess hazards (non-PEH) is fairly easy to employ in the standard piecewise methods by fitting an interaction between each time interval and each covariate. Doing this gave estimates for the excess hazard rate ratios as shown in table 5.12.

This highlights a large increased uncertainty for the parameter estimates as follow-up time increases. There is a higher excess hazard ratio initially for both age groups. An excess hazard rate ratio up to 6 months of $8.263(6.61, 10.33)$ for patients aged 75 years or older indicates that compared with a patient 60 years or younger they have an excess mortality rate that is over eight times higher during the first 6 months after MI. However, a likelihood ratio test between the PEH and non-PEH model showed a non-significant difference ($\chi^2_{10} = 9.47, p = 0.488$).

Although there is little evidence of time dependent effects in these data there often is in these types of models and it is important that it is possible to fit these models. To model non-PEH using non-split time data and splines the process is fairly straightforward as each time-dependent covariate that is modelled can be allowed to vary by each spline term. For example, it is possible to allow age groups to interact with the spline terms so that age is allowed to vary over time.

The sensitivity analyses thus far have advocated the use of flexible parametric models in relative survival. However these examples shown have all assumed proportional excess hazards, i.e. that the covariates (age groups) do not vary over time. This assumption may or may not hold true but it is important to assess the model's ability to produce valid estimates of time dependent covariates.

Time dependent effects are currently only implemented to have as many knots as the baseline effect, which is following the ideas of Royston and Parmar (135), meaning that each time dependent covariate needs to estimate the same amount of parameters as the baseline.

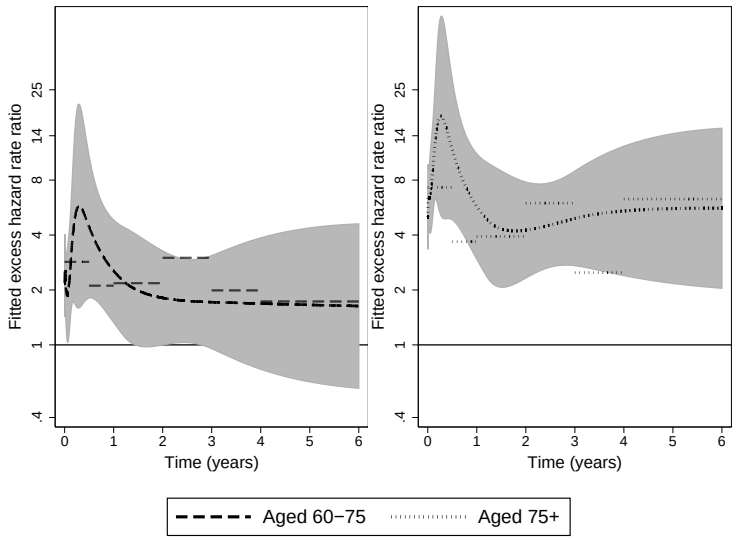


Figure 5.16: The fitted excess hazard rate ratio with 95% confidence interval alongside the piecewise estimates, where — is an excess hazard rate ratio=1 for patients aged 65-75 (left) years and aged 75+ (right) years.

5.4.1 Calculating excess hazard rate ratios

The excess hazard rate when $x_0 = 1$, assuming x_0 is a binary covariate, with only two spline variables can be written as,

$$\lambda(t) = [(\gamma_{10} + \gamma_{11}x_0)z_1 + (\gamma_{20} + \gamma_{21}x_0)z_2] \exp[\gamma_{00} + (\gamma_{10} + \gamma_{11}x_0)z_1 + (\gamma_{20} + \gamma_{21}x_0)z_2 + \beta x_0] \tag{5.43}$$

However when $x_0 = 0$ with only two spline variables then the excess hazard rate can be written as,

$$\lambda(t) = \gamma_{10}z_1 + \gamma_{20}z_2 \exp[\gamma_{00} + \gamma_{10}z_1 + \gamma_{20}z_2] \tag{5.44}$$

This means that the log excess hazard ratio can be estimated using

$$\log(EHR) = \log[(\gamma_{10} + \gamma_{11}x_0)z_1 + (\gamma_{20} + \gamma_{21}x_0)z_2] - \log(\gamma_{10}z_1 + \gamma_{20}z_2) + \gamma_{11}z_1 + \gamma_{12}z_2 + \beta x_0 \tag{5.45}$$

This is a non-linear function of model parameters as it involves the log of the derivatives of the spline function and therefore uses the δ method to calculate log hazard ratio standard

error. The δ method is only used for calculating the standard error of a non-linear function, not the non-linear function itself which is a straightforward calculation. If there is no time dependence then γ_{11} and γ_{21} will both be close to zero and the two terms will cancel out.

In order to compare the fit of the spline based models an example is shown in figure 5.16 where a non-proportional excess hazards model has been fitted with five internal knots and has been compared with that of a Poisson piecewise model based on individual level split time data. The plots show a general level of agreement with the piecewise models being a fair approximation to the flexible approach as this is the most restricted method of the two due to the assumption of a constant effect during each split time interval.

5.4.2 Knot frequency sensitivity

There are three main output measures to investigate from a non proportional model, the relative survival and excess mortality as for a PEH model but also the excess hazard rate ratios. The relative survival estimates are shown in figure 5.17. Unlike in the PEH models it is necessary to show the results for all age groups as they are now not proportional to each other. The plots are split by age groups with the youngest age group at the top and the oldest at the bottom. All three plots are very similar indicating a potential lack of time dependent effects.

In all three plots the 2df model gives a different estimate of relative survival when compared to the 3df-6df models. The remaining models produce a very similar estimate which is similar to that which was observed in the PEH models. There is no evidence here that the estimate of relative survival is overly dependent on the number of knots selected in a non-proportional model.

The estimates of the excess hazard rate for the various models with varying degrees of freedom are shown in figure 5.18 split by age groups as before. The first point of interest is that like before with the relative survival estimates the 2df model appears to be the most different model to the others as it fits a slightly different shape to the excess hazard. There is very little difference between the other models in all of the three age groups and this is indicative of a fairly robust model which does not depend greatly on the number of knots.

Finally the plot shown in figure 5.19 gives the excess hazard rate ratios for the patients aged between 60 and 75 compared to those under 60 in the top two rows of the figure and the oldest age group (≥ 75) compared with the youngest in the bottom two rows of the plot. This plot may appear confusing as there are many sets of results but by separating the results it is possible to show the confidence intervals associated with the estimates.

The middle age group show a potential problem with the method in that there is a very

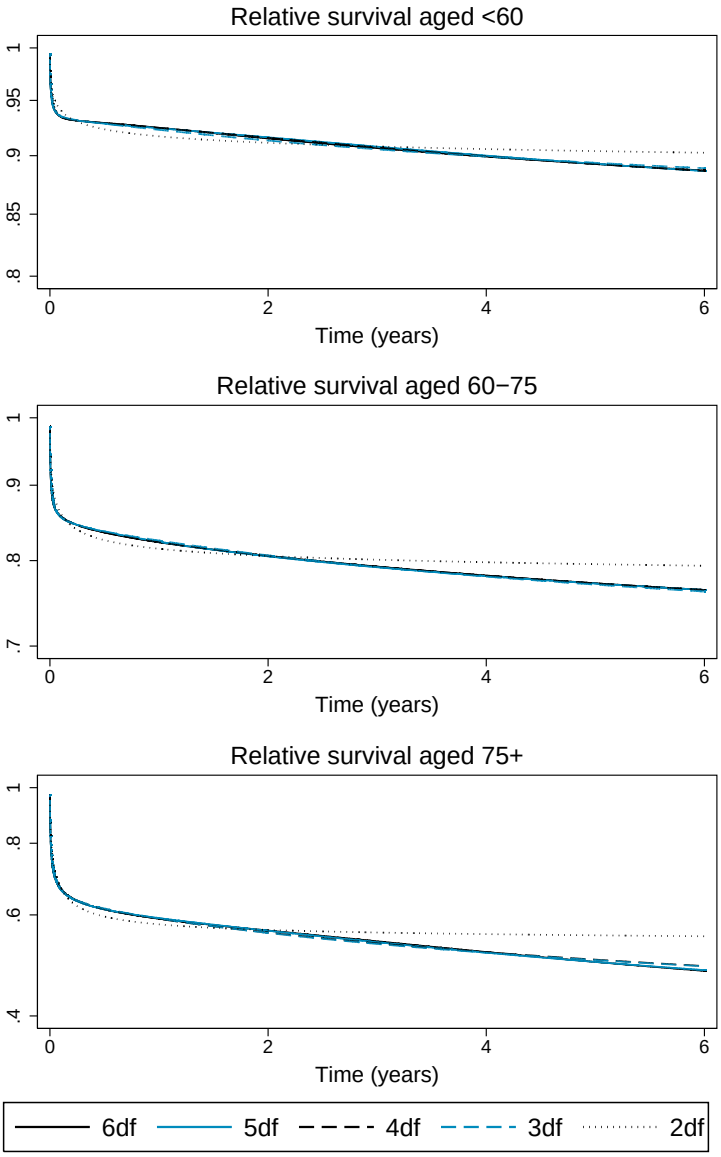


Figure 5.17: Sensitivity of the relative survival estimates to the number of knots using STRSRCS in a non-proportional model

small drop before a large rise in excess mortality when using models with 4 or more degrees of freedom. However as these are ratios it may be representative of the data as the oldest age group also shows a rise in excess mortality during the first year. These results are in comparison to the patients aged under sixty years old and this may be the reason for this. The mortality rate in the first 30 days is so high that the three groups are very similar for this period, but from this point on the two older age groups are at a higher risk of mortality than the younger patients.

The 6df model appears to be slightly overfitted in the oldest age group as it appears to rise, fall, rise and then stabilise unlike the 3df-5df models which rise and then fall only.

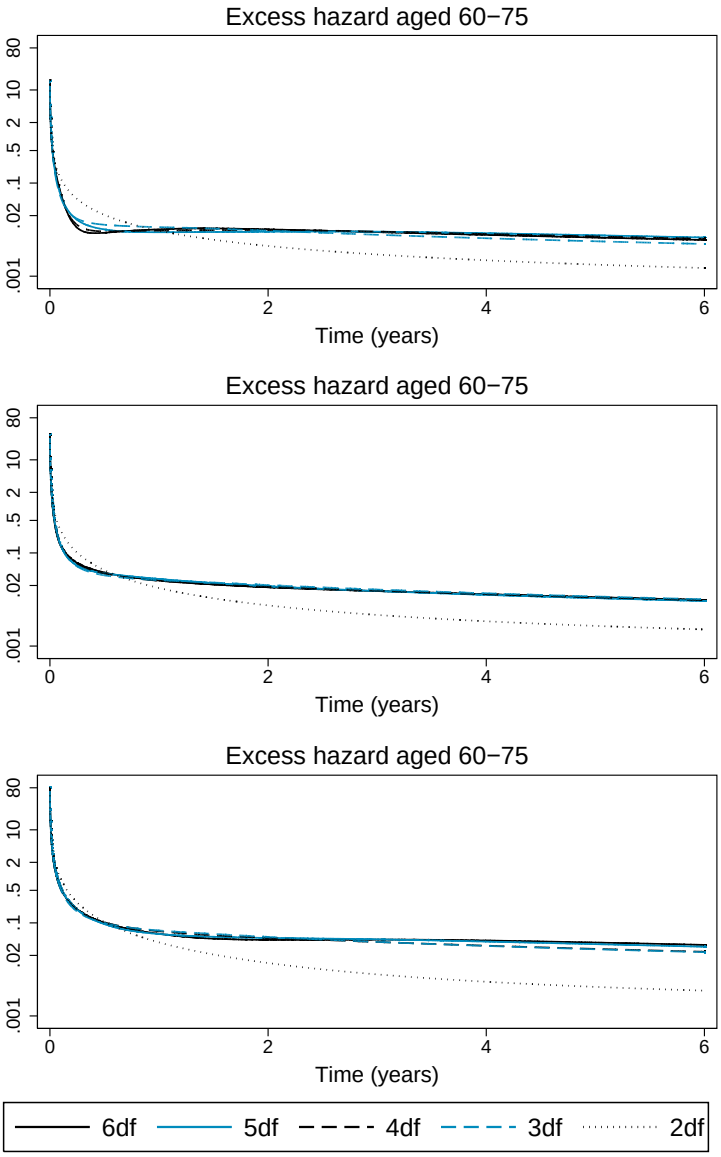


Figure 5.18: Sensitivity of the excess hazard estimates to the number of knots using STRSRCS in a non-proportional model

The 4df model appears to give a fairly consistent fit to the excess hazard rate ratios in both age groups and may be a good model to select.

Table 5.13 shows the AIC estimates from these fitted models. The 3df and 4df have the best fitting models under this criterion and after investigation of the plots would appear to show a very similar fit to each other. The worst fitting model is the 2df model as this has the highest AIC and is the most different fit as shown in the figures.

After this analysis the approach appears robust to changes in the number of knots used when modelling, although including only one internal knot does not appear as similar as

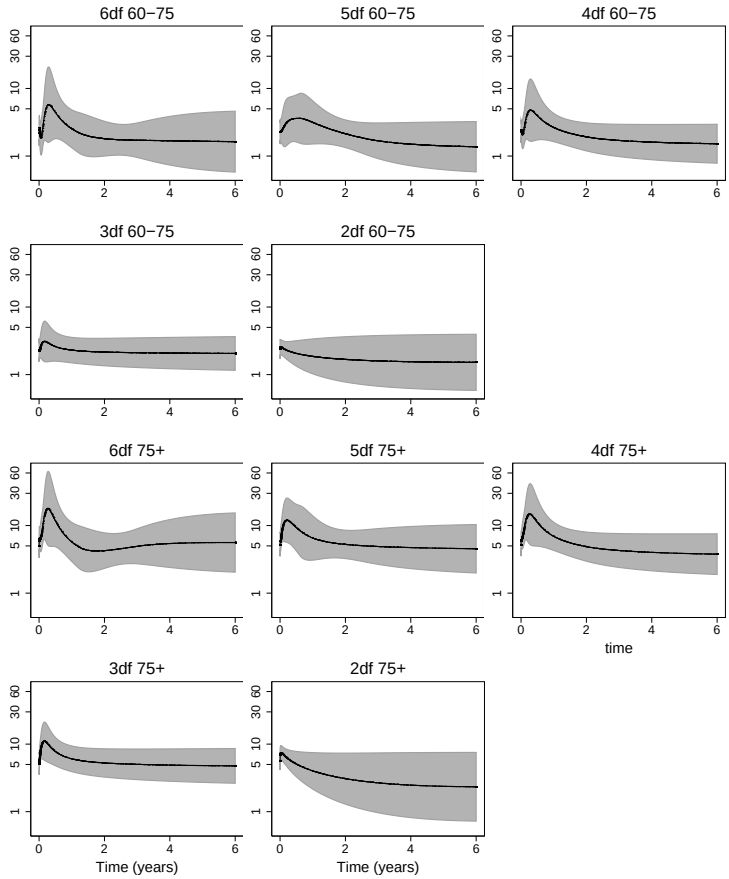


Figure 5.19: Sensitivity of the excess hazard ratio estimates to the number of knots using STRSRCS in a non-proportional model. Aged 60-75 in the top half and aged 75+ in the bottom half

the others.

5.4.3 Knot location sensitivity

While the model is robust to the number of knots, it is also important to test that it is robust to the location of knots.

The results shown in figure 5.20 are from six different models, all with four degrees of freedom and three internal knots. The difference between these six models are the location of the knots. On the left hand side of the figure are the relative survival estimates and the excess hazard rates are on the right side. These are also split by age groups as non-proportional excess hazards are assumed for the age groups.

For the two younger age groups all of the models produce near identical estimates of the relative survival rate as there is no visible separation between the models. There are variations visible in the youngest age group but these are very small. In the oldest age

Model df	Knot centile location*	AIC
2	50	3316.97
3	33 67	3126.32
4	25 50 75	3126.29
5	20 40 60 80	3131.45
6	17 33 50 67 83	3135.56

Table 5.13: AIC estimates from a non-proportional excess hazards model with various knot frequency. * Percentiles of all death times, as proposed by Royston and Parmar.

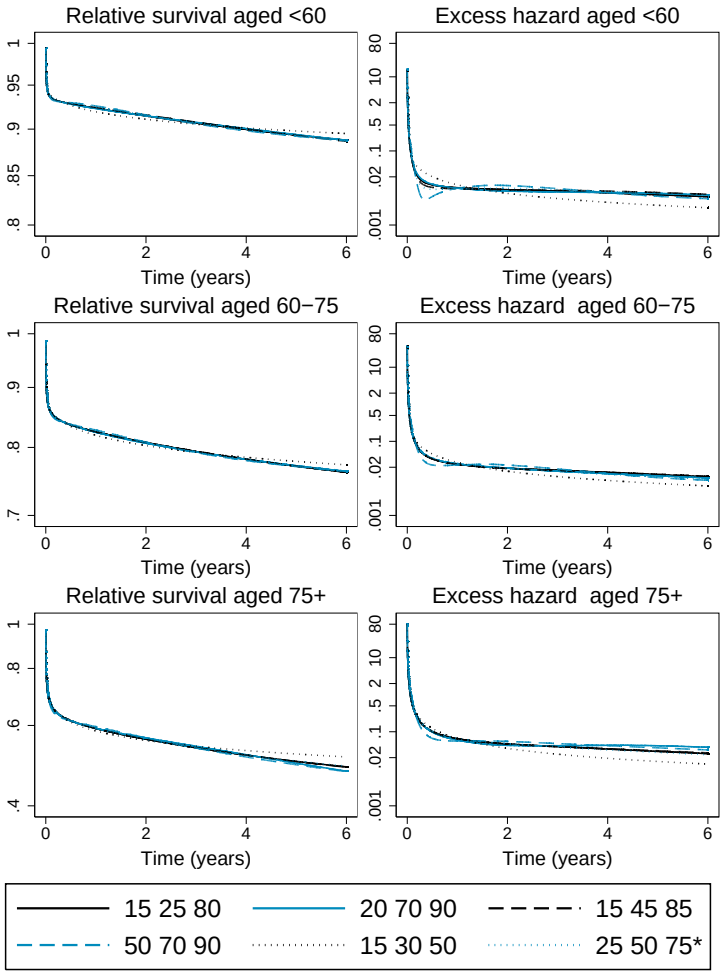


Figure 5.20: Sensitivity of the relative survival and excess hazard rate estimates to the location of knots using STRSRCS in a non-proportional model

group this is also true for the most part with the final year of follow-up showing slight variations between the models with knots including 70 and 90 although this difference is very slight. The most different model is the one with all knots on or before the 50^{th} percentile as this does not appear as capable of modelling the data for the entire follow-up period.

The excess hazard rates show more information and highlight the model with knots at

the 50, 70 and 90 percentiles as different early on after the initial drop in excess hazard at the beginning of follow-up for all age groups. In the youngest age group this is evident as a sharp dip which may be due to the splines inability to turn smoothly after such a sharp drop in mortality, which may be a limitation of the model and will be discussed in section 5.5.2. In the oldest age group, as in the relative survival estimates, it is the models with knots at the 70 and 90 percentiles that are different to the other models which have either a more consistent spread of knots or have more knots early on where the high excess mortality is measured. The (15, 30, 50) knot model is quite the opposite as this models adequately up to two years of follow-up but then tails off towards the end due to all of the knots being placed within the first 2 months of follow-up.

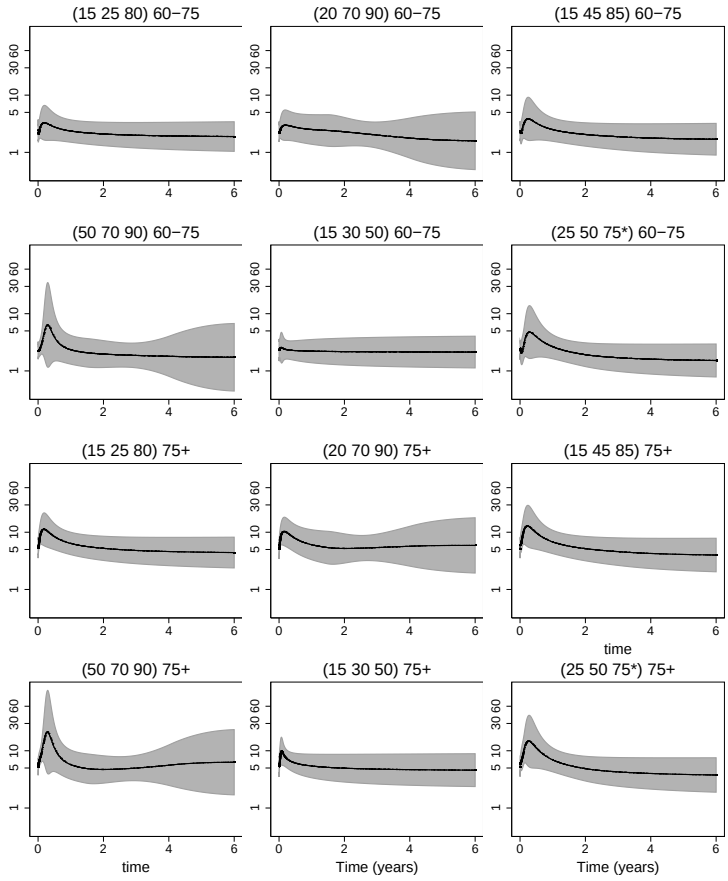


Figure 5.21: Sensitivity of the excess hazard ratio estimates to the location of knots using STRSRCS in a non-proportional model

Figure 5.21 shows the estimates of the excess hazard rate ratios from the five models with varying knot position. The figure includes both the middle age group as the top two rows and the oldest age group as the bottom two rows with the youngest age group being used as the comparator group. The middle age group shows that all models are giving similar estimates for the majority of the model with the bigger differences being in the first six months were the mortality rate begins to level after starting so high.

Knot centile location*	AIC
15 25 80	3127.71
20 70 90	3127.71
15 45 85	3127.21
50 70 90	3194.86
15 30 50	3146.42
25 50 75	3126.29

Table 5.14: AIC estimates from a non-proportional excess hazards model with various knot location. * Percentiles of all death times, as proposed by Royston and Parmar.

The oldest age group are shown to have a difference between the models fitted that was also found in the excess hazard rate and relative survival estimates. Here it is clear that the 50,70 and 90 percentile model is quite different to the others both early and later in the model. The fitted curves are not largely different but it would suggest that without a knot at an earlier point a slightly different model is fitted giving different estimates. However unlike the previous plots the $(15, 30, 50)$ knot model is shown to be fairly consistent with the other models.

Table 5.14 shows the AIC estimates from these fitted models. There are four models with low AICs, one being the model using default values and the other three being $(15, 25, 80)$, $(20, 70, 90)$ and $(15, 45, 85)$. The default value model has the best fitting model under this criterion and after investigation of the plots would appear to show a very similar fit to the other three low AIC models. The worst fitting model is the $(50, 70, 90)$ model as this has the highest AIC and is one of the most different fits as shown in the figures.

These findings suggest that without assuming proportional excess hazards the effect of knot placement is still of little bearing on the final fit of the model. If the user selects knots only to be placed on the final half of all deaths knowing that the large excess mortality is experienced in the first six months then the fitted model is still fairly reliable even if the model itself should not be considered as entirely logical. Similarly failing to place a knot in the later half of the model also appears to produce a poorer fitting model than one with knots across the full range of the data.

5.4.4 Bootstrapping time-dependent standard errors

The plots shown in figure 5.22 show the estimates for a 2df(top), 4df(middle) and 6df(bottom) model without assuming proportionality. In these plots are the estimates from the delta method, indicated by the top row in the legend and the bootstrap estimates indicated by the bottom row of the figure legend. There are two bootstrap estimates, normal and bias corrected estimates of the confidence intervals. All three estimates are near identical for the majority of the model with some discrepancy during the first few months. The normal bootstrap estimate for the lower confidence interval is not shown for the 6df model as it

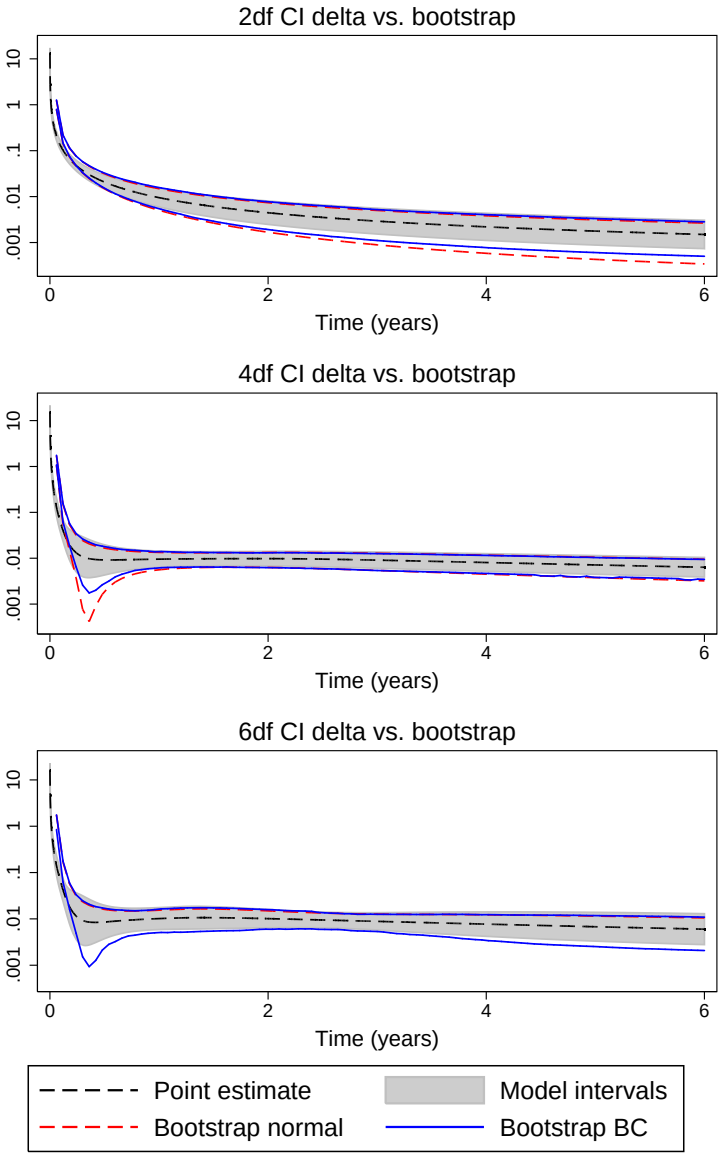


Figure 5.22: Comparison of the confidence interval estimates using the δ method and bootstrapping with both normal and bias corrected intervals for a non-PEH model.

did not appear to correctly estimate the lower interval after 1000 reps. The 6df model took over sixteen and a half hours to converge and highlights the speed of the δ method which for the same model took only fifty-six seconds.

5.4.5 Random knots

In order to fully justify the use of splines when adopting time-dependent effects a model with three internal knots was fitted twenty times to the data with each model using three knots that were randomly allocated using the univariate distribution. The only limitation

used was in the lowest knot point being above the 15^{th} percentile as values lower than this caused the model convergence to fail. Also the number of iterations was limited to fifteen meaning that some of the models may not of converged, this was to ensure that the loop could complete with twenty observations. This is essentially the same approach that was performed on the baseline model in section 5.3.3, but this time the number of iterations was increased from the baseline assessment as these models are more complex.

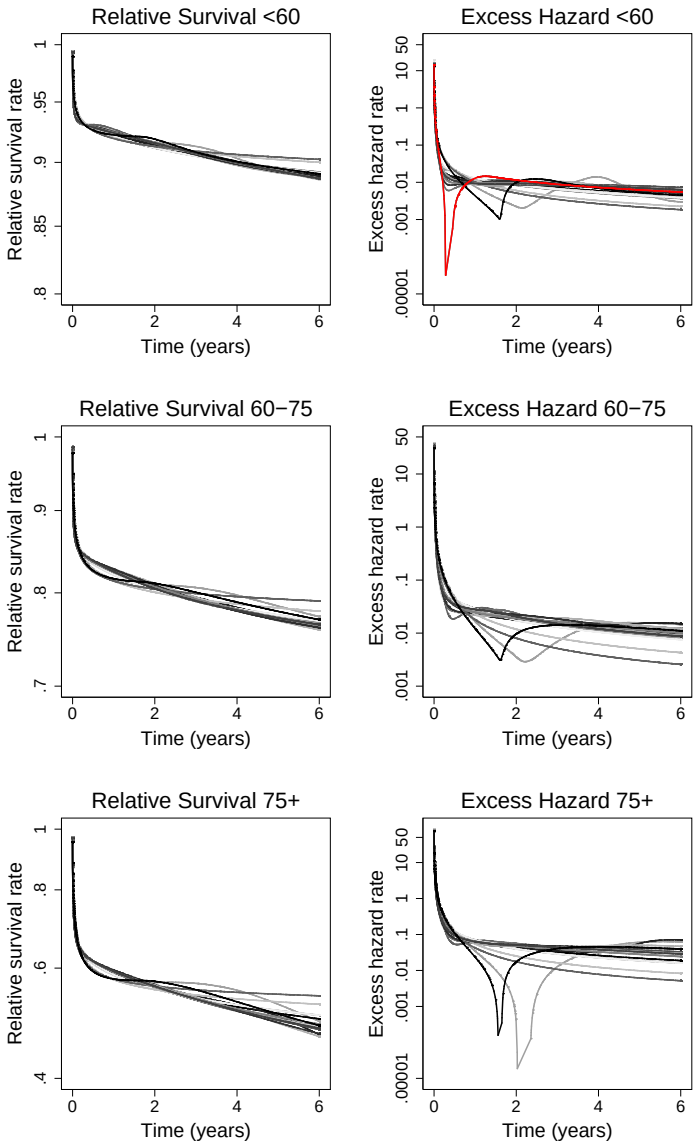


Figure 5.23: Sensitivity of the relative survival estimate to the location of random knots in non-PEH models.

Figure 5.23 shows the fitted relative survival and excess hazard rate for all twenty models selected. The knots that were randomly selected are shown in table 5.15. It was decided to use different knots to those used in the baseline assessment (section 5.3.3) in order to allow these to be truly random also. As is shown in the figure, five of the models appear

Knot positions			
54,65,70	76,88,91	24,25,87	38,59,74
28,35,62	21,22,64	17,89,95	49,70,75
32,34,56	45,83,91	47,56,86	27,47,55
46,73,95	17,23,60	25,60,89	32,33,44
22,69,73	21,48,75	32,67,75	71,72,83

Table 5.15: Random knot locations.

to be bad to very bad fits, suggesting that using these randomly selected knots up to five models are not very useful. Fifteen models fitted a very similar response for both relative survival and excess mortality across all age groups.

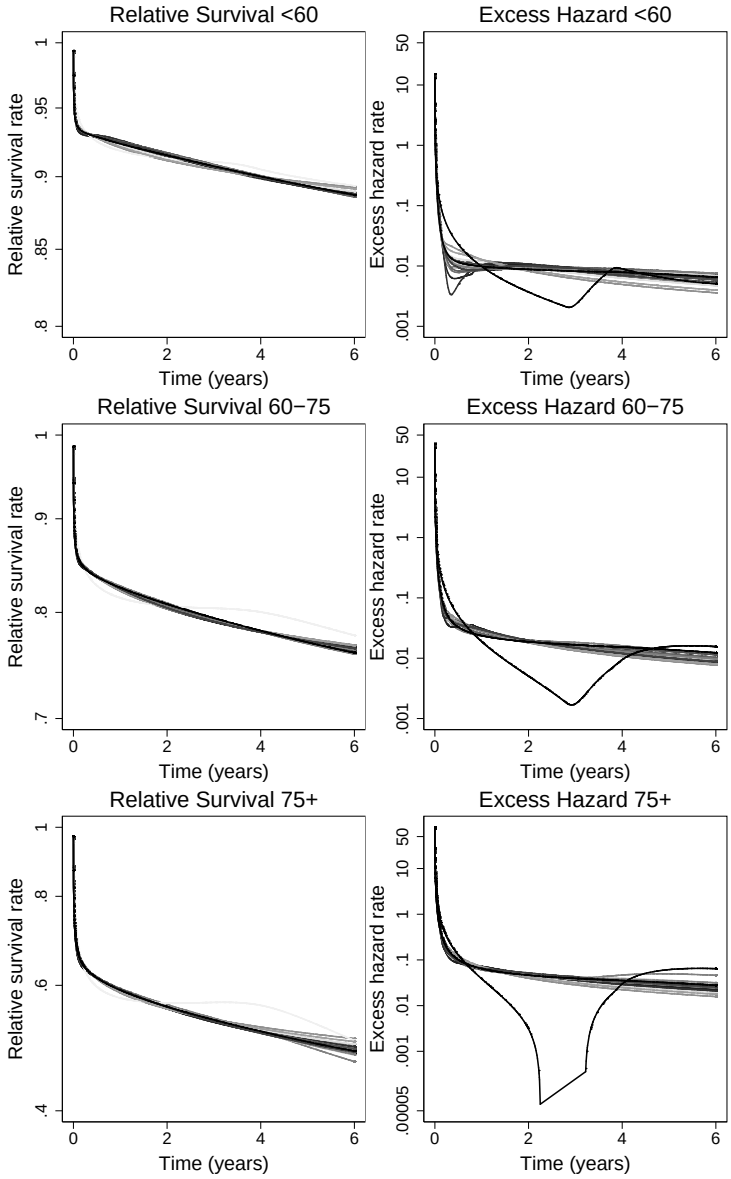


Figure 5.24: Sensitivity of the relative survival estimate to the location of random knots in non-PEH models using previous knots.

There are two models listed in table 5.15 that have their first knot after the 70^{th} percentile at $71, 72, 83$ and $76, 88, 91$ and these two models did not converge and show the worst fit of the twenty models as would be expected without convergence. Another two models did not converge with knots at $17, 89, 95$ and $32, 33, 44$ both of these also show a poor fit to the data. The final poor model did converge that placed the knots at $54, 65, 70$. but still gave a poor fit as the first half of the event data did not have a knot in it. However the other fifteen models appear fine and as listed in table 5.15, these are quite varied and some would not be a sensible choice in practice with knots very close together of starting at the 49^{th} percentile.

The same random knots that were used in the baseline effects in section 5.3.3 were applied to this data and produced the plots shown in figure 5.24. This shows that there are nineteen models fitting a similar curve to the relative survival and the excess hazard rate for all age groups. However there is one model that is clearly a poor fit with knots placed at $82, 88, 89$ this model does not converge and is therefore not a model that could even be chosen by accident.

This has shown that even in complex cases, where there are time dependent effects, the models are producing sensible fits to the data even where the knots are simply placed at random. This has exceptions where some models do not converge and produce poor fits but these models should not be considered in practice due to their absurd knot placement.

5.5 Limitations

5.5.1 Two time dependent effects

The analysis here has been on a single covariate, age groups using the youngest age group as the baseline effect. However in practice it is quite feasible that there will be more than one time-dependent effect which will need to be assessed along with other covariates. This is easy to fit but has some problems when using this approach on the log cumulative excess hazard scale. When modelling on the log excess hazard scale the effects of a time dependent covariate would be assumed to be the same at all levels of a second time dependent covariate, assuming there is no interaction. This is not the case when modelling on the log cumulative excess hazard scale.

A PEH model with two spline terms (z_1 and z_2) and two covariates (x_1 and x_2) has a linear predictor for the hazard given by:

$$\gamma_{01} z_1 + \gamma_{02} z_2 + \beta_0 + \beta_1 x_1 + \beta_2 x_2 \tag{5.46}$$

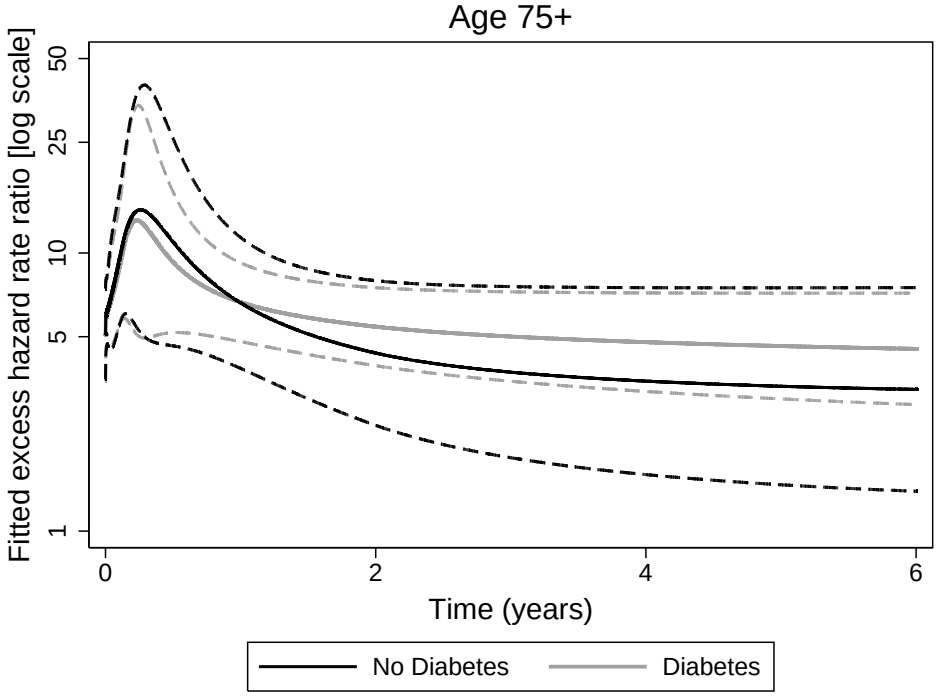


Figure 5.25: Example of the >1 time dependent covariate problem with excess hazard rate ratios

But when extended to time dependent effects this becomes more complex

$$\gamma_{01}z_1 + \gamma_{02}z_2 + \gamma_{11}z_1x_1 + \gamma_{12}z_2x_1 + \gamma_{21}z_1x_2 + \gamma_{22}z_2x_2 + \beta_0 + \beta_1x_1 + \beta_2x_2 \tag{5.47}$$

So the following is used to determine the log excess hazard ratio for x_1 when $x_2 = 0$ where z' is for the derivatives of the splines.

$$\log\left[\gamma_{01}z_1' + \gamma_{02}z_2' + \gamma_{11}z_1'x_1 + \gamma_{02}z_2'x_1\right] - \log\left[\gamma_{01}z_1' + \gamma_{-2}z_2'\right] + \gamma_{11}z_1x_1 + \gamma_{12}z_2x_1 \tag{5.48}$$

Here the effect of covariate x_2 is not present as it is zero but when it equals one the log excess hazard ratio for x_1 is given by:

$$\begin{aligned}&\log\left[\gamma_{01}z_1' + \gamma_{02}z_2' + \gamma_{11}z_1'x_1 + \gamma_{02}z_2'x_1 + \gamma_{21}z_1x_2 + \gamma_{22}z_2x_2\right]\\ &\quad - \log\left[\gamma_{01}z_1' + \gamma_{-2}z_2' + \gamma_{21}z_1x_2 + \gamma_{22}z_2x_2\right]\\ &\quad + \gamma_{11}z_1x_1 + \gamma_{12}z_2x_1\end{aligned} \tag{5.49}$$

As is clear the estimate for x_1 when $x_2 = 1$, as shown in equation 5.49, includes terms for both covariates.

The problem observed is that the levels of one time dependent covariate are not independent of the second time dependent covariate due to the scale being used. In fact they are dependent on each other. For example a model with age groups and diabetes as time dependent effects could be fitted to these data using the flexible parametric models proposed here and the hazard ratio for patients aged seventy-five years or older compared to patients under sixty years old would be different for diabetic patients compared to non-diabetic patients.

This is best explained in figure 5.25 which shows the estimated excess hazard rate ratio for patients aged ≥ 75 compared to patients < 60, but also by diabetic status with both variables being time-dependent. Usually, on the log excess hazard scale, the estimates of the excess hazard ratio would be identical as the terms for diabetes would cancel out. The plot shows that the ratio depends on diabetes as the estimate and confidence interval are all different dependent upon the patients diabetic status. This is not an error of the model, but is a feature of modelling on the log cumulative excess hazard scale. However, users need to be aware of this and to think about the best way to present the information.

5.5.2 Dipping

The second limitation to highlight is one where a fitted model appears to show a strange dip, for example a high excess mortality rate will drop sharply and then rise again before leveling out. This can look very strange as the fitted curve does not appear smooth and is potentially an artifact of the splines rather than an observed trend in the data.

An example of this is shown in figure 5.26. This was shown earlier in figure 5.20 for the-non proportional excess hazards model with knots placed at the $50, 70, 90$ percentiles for patients aged under sixty years old as the two older age groups did not present the same dipping effect. On this plot the confidence intervals are also shown along with vertical lines indicating the knot position. This model was found to be the worst fitting of all models but is only presented as an example of the effect.

The effect is likely due to the models inability to turn quickly enough after such a sharp drop in excess mortality following the first month of follow-up. Without an extra spline term during this interval the model is unable to smoothly estimate the large change in excess mortality that occurs after this point. Therefore it is sensible to suggest that these types of model are not overly suitable and knots should be placed where the majority of the information is. This indicates that a sensitivity analysis should be an important part of fitting these models.

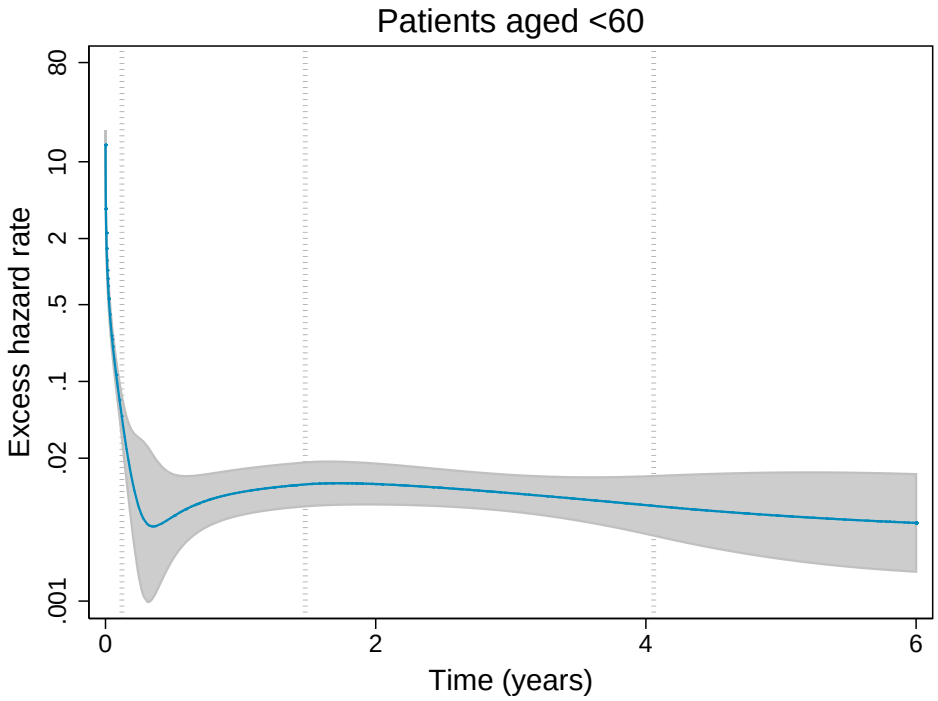


Figure 5.26: Example of the dipping effect in an excess hazard rate

It is also possible that the splines restrictions in the tails of the distribution are the cause of the problem as it is forced to be linear which could be a factor in the model's inability to turn sharply at the start of follow-up. Therefore it may be possible to remove these restrictions by using standard cubic regression splines rather than restricted cubic splines.

5.6 Unrestricted cubic regression splines

In this section different and less complex splines will be investigated by adopting standard cubic splines as defined in regression modeling strategies (143). While linear splines can approximate many common relationships they are not smooth and will not fit the highly curved functions needed for these types of data, which leads to the use of cubic spline functions. Cubic polynomials have been found to have useful properties with an ability to fit sharply curving shapes. Cubic splines can be made to be smooth at the knots by forcing the first and second derivatives of the function to agree at the knots and as such a cubic spline function with 3 knots (z_a, z_b, z_c) is given by

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 (x - z_a)_+^3 + \beta_5 (x - z_b)_+^3 + \beta_6 (x - z_c)_+^3 \qquad (5.50)$$

with the following constructed variables:

$$\begin{aligned} x_1 &= x & x_2 &= x^2 \\ x_3 &= x^3 & x_4 &= (x - z_a)_+^3 \\ x_5 &= (x - z_b)_+^3 & x_6 &= (x - z_c)_+^3 \end{aligned} \tag{5.51}$$

The derivatives of the spline function are also needed and are calculated as:

$$\begin{aligned} x_1 &= 1 & x_2 &= 2x \\ x_3 &= 3x^2 & x_4 &= 3(x - z_a)_+^2 \\ x_5 &= 3(x - z_b)_+^2 & x_6 &= 3(x - z_c)_+^2 \end{aligned} \tag{5.52}$$

The model is then fitted as before. Extension to more knots is simple (143).

If the cubic spline function has k knots then the unrestricted cubic spline will estimate $k+6$ parameters, including the baseline, whereas a restricted cubic spline model will estimate $k + 2$ parameters, including the baseline (β_0). It is also possible to use B-splines instead of the truncated power basis used here. However B-splines are more complex and do not allow for extrapolation beyond the outer knots.

The main problem with cubic splines is their poor behaviour in the tails before the first knot and after the last which is tackled when using restricted cubic splines as used in the `strsrcs` program and the Royston model (135). The reason for employment here is due to the limitations discussed in section 5.5.2. There may be arguments for using standard regression splines as restricted cubic splines are constrained to be linear in the tails for the reason that there is usually little data in the tail. This is not generally the case in the left-hand tail in survival analysis as this is usually where the majority of the information is. Thus it is worthwhile to investigate splines that do not impose this restriction.

Using non-restricted cubic spline models it is necessary to estimate four more parameters than restricted cubic spline models. So for example a three (internal) knot restricted cubic spline approach will model four parameters plus the baseline, whereas a standard cubic spline would model eight parameters plus the baseline. The parameters are not directly linked to the number of internal knots, i.e. a three knot model in both cases will fit three knots at the $25, 50, 75$ centiles. Due to the constraints imposed on restricted cubic splines they will have fewer parameters than standard cubic splines.

5.6.1 PEH models

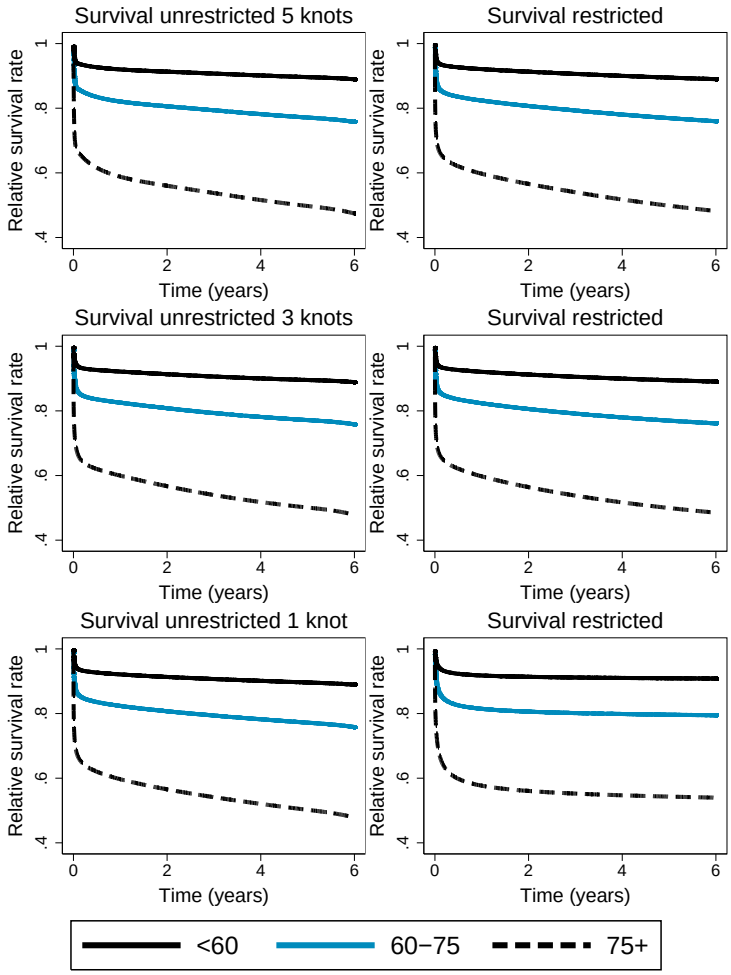


Figure 5.27: Comparison between PEH cubic splines and PEH restricted cubic splines predicted relative survival

A PEH model comparison using one internal knot (bottom row), three internal knots (middle row) and five internal knots (top row) are shown in figure 5.27. As stated above, although there are the same number of knots in both methods the standard unrestricted splines estimate four more parameters than the restricted splines approach.

In figure 5.27 there are estimates from a restricted cubic spline model on the right hand side and an unrestricted cubic regression spline on the left hand side, the plots are labelled as such. The 5 knot and 3 knot models show that by removing the restrictions in the tails there is no visible difference between the models shown side by side in terms of the relative survival estimates. The one knot model is slightly different in that the restricted model curves level out for all age groups whereas the unrestricted model is more similar to the 3 knot and 5 knot models in that the rate continues to decrease at a higher rate than the restricted 1 knot model. This may be evidence that the 1 knot unrestricted model is

better than the restricted one as it matches models with a higher degree of flexibility.

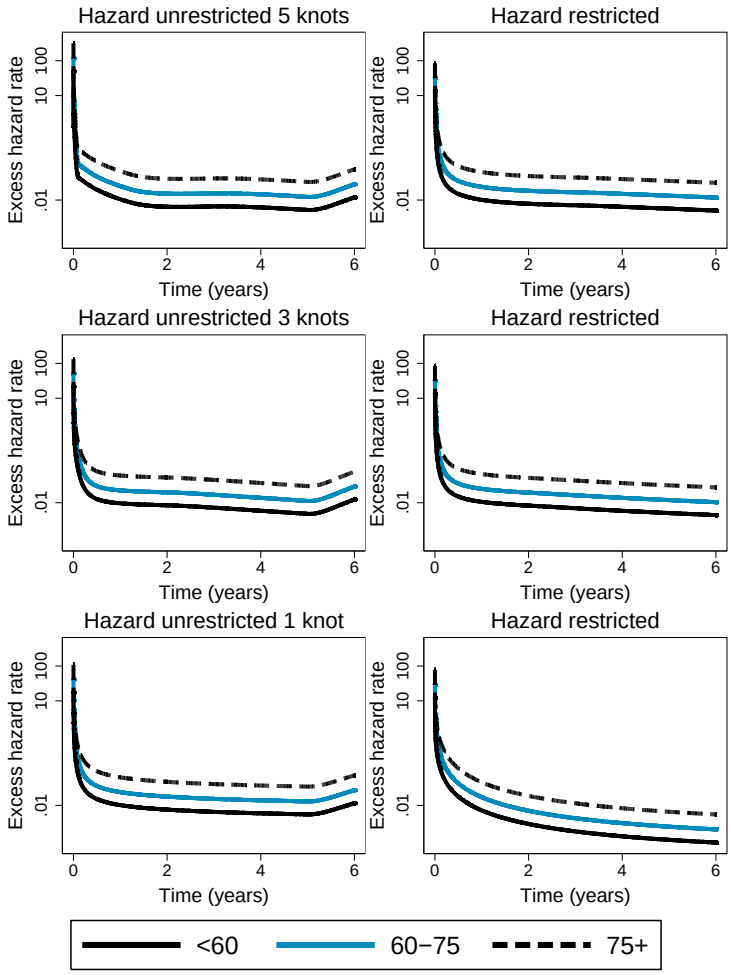


Figure 5.28: Comparison between PEH cubic splines and PEH restricted cubic splines predicted excess hazard

An excess hazard rate PEH model comparison using one, three and five internal knots are shown in figure 5.28. This figure is set up in the same way as figure 5.27, however unlike in the relative survival plots these excess hazards appear different for all models. The 5 knot unrestricted model appears to be fitting a curve that begins to rise at the end of follow-up, i.e. the tail is not behaving as well in the restricted model as an increase in excess mortality after five years of follow-up is not easily explainable and has not been observed before. The 3 knot models are similar to each other but the restricted model is decreasing at a faster rate and again shows an increase after five years of follow-up. The 3 knot unrestricted model is very similar to the 1 knot unrestricted model in its shape. However the 1 knot models do not match each other, as the restricted model appears to be fitting a fairly poor model which decreases at a steady rate without levelling out. The unrestricted models do not appear to be behaving in the right-hand tail of the distribution.

Table 5.16 shows the AICs from the PEH model along with the number of parameters that

Knot position*	Cubic regression spline		Restricted cubic spline	
	N par†	AIC	N par†	AIC
13 25 50 67 83	13	2817.74	9	3165.96
25 50 75	11	3085.86	7	3162.95
50	9	3104.58	5	3364.88

Table 5.16: AIC estimates from restricted and unrestricted cubic spline models. * Percentiles of all death times, as proposed by Royston and Parmar. † Number of parameters estimated by the PEH model for Age groups

are estimated for both a standard cubic spline model and a restricted cubic spline model. There are three baseline parameters estimated, one for each level of the age group covariate that is being modelled. Using these AICs a different number of knots would be selected depending upon the spline model used. The unrestricted splines with five knots has the lowest AIC and the three knot model has the lowest AIC for the restricted cubic spline approach. Even though more parameters are used in the unrestricted models they still produce lower AIC values than the restricted cubic spline models. This may suggest that these models have captured extra variation early on in the timescale, but given the poor fit of the excess hazard shown here may be indicative that the AIC is not an appropriate measure.

There were many problems with convergence as the Newton Raphson technique did not work. This was altered to another technique, `dfp` (*Davidon-Fletcher-Powell*) or `bfgs` (*Broyden-Fletcher-Goldfarb-Shanno*) which did converge eventually. The two external knots were also placed at the 5^{th} and 95^{th} percentiles.

5.6.2 Non-PEH models

A non-PEH model comparison using five (top row), three (middle row) and one (bottom row) internal knots are shown in figure 5.29 predicting relative survival via both methods. The standard cubic regression splines are shown on the left hand side of the figure with the restricted cubic splines shown on the right. Even though these models vary in complexity all six show a very similar estimate of relative survival. The over seventy-five age group has the largest number of events and the highest mortality rate, and the estimates differ to some extent, whereas the other two age groups are very similar.

The unrestricted models show the relative survival estimate for the oldest age group decreasing towards the end of follow-up whereas the restricted approach is starting to decrease, but at a slower rate. The one knot model in the restricted approach is the most different of the six models and is the least likely estimate as this suggests that the estimates level-out into a plateau which is not likely following an MI and is also not suggested in any other models fitted thus far. The relative survival estimate is important but the excess hazard is of more interest as this has been found in previous analyses to be more

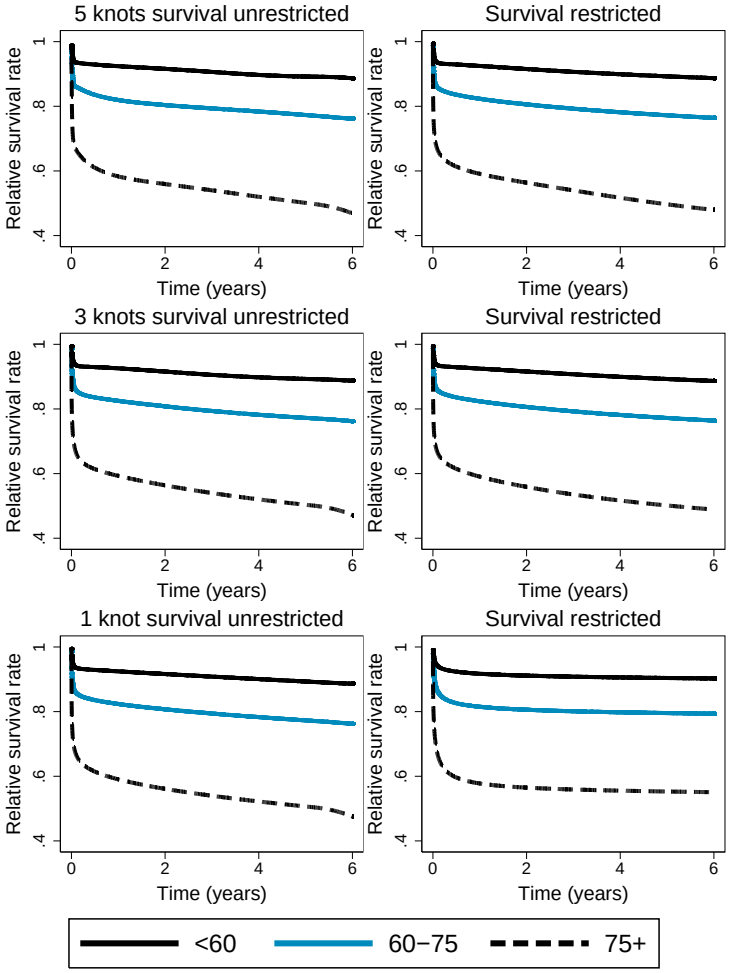


Figure 5.29: Comparison between non-PEH cubic splines and non-PEH restricted cubic splines predicted relative survival

susceptible to poor fitting models.

A non-PEH model comparison using one and three and five internal knots are shown in figure 5.30 predicting the excess hazard rate and is set out in the same way as the previous figure with both unrestricted and restricted splines shown. The restricted and unrestricted spline models are quite different to each other and none of them are very similar (when comparing left to right). The immediate difference between the two approaches is shown in the final tail of the model as the unrestricted model turns to show an increasing hazard at the end of follow-up.

The three restricted cubic spline models all produced similar effects where the excess hazard rate is shown to be continually decreasing. However the one knot model appears to be a poorer fit than the other two as it more curved and the excess hazard rate drops to a much lower point than the other five models. There is also more evidence of potential *dipping* in the youngest age group in the five knot model at the beginning of follow-up.

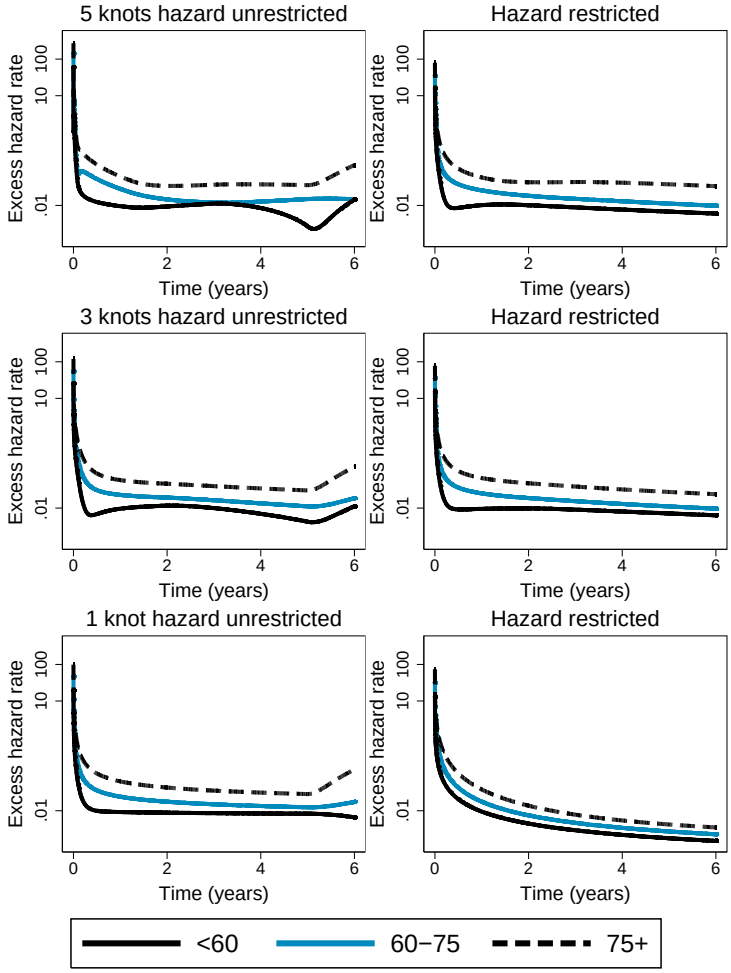


Figure 5.30: Comparison between non-PEH cubic splines and non-PEH restricted cubic splines predicted excess hazard

The five knot unrestricted model appears to change direction around two years as the excess hazard is shown to start increasing again, not something which has been observed in good fitting models and is likely due to being over-fitted with too many knots. The same model after four years produces a severe dip in excess hazard which is much worse than the restricted cubic spline model. The three knot unrestricted model shows a similar effect but it is not as severe, however there is *dipping* at the start of follow-up. This suggests that removing the restrictions has created further problems without producing a solution to the initial problem of dipping. The five and three knot restricted cubic spline models appear the best fits as they are smooth and show a decreasing excess hazard at a similar rate to each other.

There are some serious convergence problems with these splines and the non-PEH models shown here did not estimate a standard error for the baseline effects suggesting further convergence issues. This may be resolvable but as the intention was to determine if the *dipping* effect could be reduced and that has so far been found not to be the case. It was

therefore deemed unnecessary to further investigate the standard regression splines.

5.7 Future development

5.7.1 Allowing for different values of theta

Royston and Parmar (135) developed flexible parametric models based initially on the assumption of either proportional hazards or proportional odds scaling of covariate effects. Generically, the class of such models was based on a transformation of the survival function by a given link function $g(.)$

$$g[S(t;x)] = g[S_0(t)] + \beta x \tag{5.53}$$

where $S_0(t) = S(t;0)$ i.e. the baseline survival. Royston and Parmar used natural cubic splines to model $g[S_0(t)]$ within the Aranda-Ordaz (148) family of link functions. Where the parameterised link function is given as:

$$g(z;\theta) = \log \frac{z^{-\theta} - 1}{\theta} \tag{5.54}$$

where $\theta = 1$ corresponds to the proportional odds model and $\theta \to 0$ to the proportional hazards model. There are further extensions to allow non-proportional effects of some or all of the covariates (other values of θ) but these have not been investigated here. The advantage is that given the potential complexity of the model it would be possible to obtain results for models that could give a better fit to these data.

While this is possible to extend to relative survival the use in standard survival analysis is very limited. This is mainly due to the difficulty in interpreting the results when θ is not 0 or 1, i.e. a hazards or proportional odds model. For other values of θ the model results are difficult to comprehend. If it is not possible to quantify the differences in results then it has very little use in practical application and as such has not been used here. However if required for theoretical application it is quite possible.

5.7.2 Time dependent flexibility

Time dependent effects are currently only implemented to have as many knots as the baseline effect, meaning that each time dependent covariate needs to estimate the same amount of parameters as the baseline. This is due to the way the method is implemented in Stata, as it is in `stpm` (147) also. However this could be extended so that it is possible

to have fewer parameters for time dependent effects. If the sex of the patient is known to be time dependent for the first year of follow-up then it is not necessary to assume non-proportional excess hazards for the remaining follow-up or to estimate these parameters in the model.

A given PEH model that, for example, contains sex and age could be fitted with six degrees of freedom, but sex was then found to vary by time, this would require a non-PEH model to be fitted. However it is possible that sex varies over time in a much simpler fashion than the baseline, perhaps needing only two degrees of freedom. Currently the model would still need six degrees of freedom for the time dependent effects (the same as the baseline effect) rather than offering the ability to model time-dependent effects with less complexity. This is a possible extension to the model that has not been performed here and may prove useful in practical research.

5.7.3 Modelling of statistical cure

Statistical cure is of interest to patients and is a useful trend in the survival of diseases that are curable. However in Heart disease this is not plausible, it is also of note that statistical cure is from a population perspective not an individual.

Mixture cure models (35) assume that a proportion of patients are cured (π) and are not at risk of the outcome, with the remaining proportion being uncured (1-π), with these subjects eventually experiencing the event of interest and the survival function tends to zero for these subjects. Standard mixture cure models are of the form where $S_u(t)$ is usually a standard parametric survival curve function, i.e. Weibull. It is also possible to fit non-parametric functions (155).

$$S(t) = \pi + (1 - \pi)S_u(t) \tag{5.55}$$

The cure fraction can be interpreted as the proportion of cured patients whose survival pattern is similar to the general disease-free population. Long follow-up time is required for these models to obtain an accurate cure fraction estimate usually after median survival time. If the fitted model is semi-parametric with proportional excess hazards then the bias of cure fraction estimates are substantially positive (35). Mixture cure models assume a zero-excess hazard from day one for those who are cured and are irrespective of other factors such as treatment.

The second type of cure model is the bounded cumulative hazard cure model (BCH) (29), which defines an asymptote for the cumulative hazard and hence for the cure fraction, one of the key advantages is that a PEH model is a special case. After treatment it is assumed

that an individual is left with N_i = a number of cells that have the potential to become diseased, particularly cancer subjects. The BCH model can be written as

$$S(t) = S^*(t)\left[\eta + (1-\pi)\left(\frac{\pi^{F_z(t)} - \pi}{1-\pi}\right)\right] \tag{5.56}$$

Where $F_z(t)$ is a cumulative distribution function, generally chosen to be $1 - S_u(t)$. Thus the survivor function has an asymptote at the cure fraction and the cumulative hazard has an asymptote at $\ln(\eta)$. The choice of link function, for π, is important as it leads to different assumptions. Cure models should only be used when it is believed a cure fraction exists but is a useful mathematical function for modelling relative survival. When there is a high excess mortality in the first few weeks there is a chance the models may fail to converge with parameters *wandering to infinity*.

Cure models require an asymptote where the excess hazard rate will eventually reach zero which may be possible using splines but is likely to prove difficult. However cure models use parametric models and often quite simplistic distributions, including the Weibull. Therefore it is possible that the spline based models have the potential to fit a more flexible parametric model for those who are bound to die in order to remove the current parametric restrictions of cure models. It must be noted that this has not been attempted here as cure is not relevant for these data.

5.7.4 Using fractional polynomials

The model used here adopted splines to model $s(z;\gamma)$ but any smoothing function could be used, including fractional polynomials (124; 39). The number of potential fractional polynomials that could be fitted to each covariate are quite numerous, and for a number of covariates the possible combination of powers that can be adopted rise rapidly. It is possible to allow the use of a Stata algorithm to select the best fitting models, which is currently used when fitting these models. This has problems with the uncertainty, as this is only one model out of potentially hundreds and as shown in the fractional polynomial section of chapter 4, the uncertainty is not easy to replicate without model averaging.

It is also problematic due to the number of fitted models that will undoubtedly be poor as they may force a negative hazard onto the data causing convergence issues. It is often necessary to fit FP3 models, which is where three parameters for a non-linear relationship are included in the same model, and not always with a different power as repeated powers are available. This is therefore computationally intensive and not as quick as the current spline method.

The advantage of this would be the speed when compared to modelling these fractional

polynomials on the split time data. They would be much quicker and would use non split-time data which holds further advantages. Another advantage is the ease of obtaining the derivatives as each term is a power where using basic calculus the derivative of x is

$$x' = px^{p-1} \tag{5.57}$$

These are therefore easy to determine and feed into the likelihood. However the advantages are only really held over current fractional polynomial models in relative survival and the spline based models proposed would be much quicker. It would also be possible to create a Stata algorithm to auto-select the best fitting model based on the number and location of knots similar to the fractional polynomials but this has not been incorporated as selection procedures are often used without full consideration of the data. There is a program available in Stata that goes some way to achieving this, which uses split-time data only (156). This program (`mvrs`) requires knot positions in advance and investigates if any knots can be removed. It does not look for alternative locations of the knots.

5.8 Discussion

The ability to use continuous covariates is lost when using grouped data and as such they are often criticised for not being biologically plausible. Using this model it is easy to model continuous covariates and this has been shown in a further chapter (chapter 9) which investigates glucose as a continuous covariate.

This leads onto the comparison between piecewise models and these flexible parametric models in relative survival. The key advantages of piecewise models are their speed in converging and ease of use with the only decision required being the number and location of split-time points to use. However, this is the first problem with split-time piecewise models in that the more splits used the more parameters required and with the preference in these models to use data at the individual level this can lead to huge datasets consisting of many rows of data for the same subject. The more splits used the higher the chance that a zero event will be encountered within a specific covariate level at a certain time point which causes issues with model convergence. It is therefore normal to adopt few time splits but this is problematic in the assumption that the effect is constant for each time interval leading to an approximation of the hazard being given rather than a smooth function over time. This is discussed at length in chapter 4.

The flexible parametric models do not split time and do not therefore suffer from these problems, but they are still very quick to converge, much quicker than the current B-spline (38) approach and restricted cubic spline approach (37) which are based on the log excess hazard scale and use numerical integration to obtain the survival function which increases

the computation time. As a comparison a simple B-spline model with default values, using age groups, was fitted using `RSURV` (134) on our data and this model converged in 148.58 seconds whereas the flexible parametric spline model using the same amount of knots (2 internal) also using a PEH assumption converged in 6.67 seconds. In fact a model with 5 internal knots and with age used as a time-dependent covariate converged in 27.86 seconds (2 knots = 7.59), compared to the B-splines non-PEH time of 211.71 seconds using `RSURV` with only 2 internal knots as this is restricted to only allow two knots. These models, if used with cancer registry data, will be fitted to potentially tens or hundreds of thousands of observations and so speed is an important factor.

These spline based models are not without their problems and the first one is in the estimation of the excess hazard rate ratios when there is more than one time-dependent effect as described in section 5.5.1. This is problematic where the estimate of one time-dependent covariate depends on the level of the other time dependent covariates. This is a problem with modelling on the log *cumulative* excess hazard scale and is potentially a reason to adopt a more time-consuming method on the log excess hazard scale. The other problem is with dipping and the potential artefact produced by using splines, as discussed in section 5.5.2. This is less of a problem and is very uncommon in the models fitted thus far. Where it has been observed it has been in the same age group in different models suggesting that it may be an effect of the data, not the splines, but should always be investigated. A potential third issue is that the time-dependent effects have to be shown graphically rather than being able to put the results into a table which may be problematic for some publications. However, it would be possible to select a range of times a-priori, such as 30-days, 6 months, 1 year, and 5 years, and report the estimated excess hazard rate ratio at these time points in a table.

Relative survival has not been used frequently in CHD as discussed in chapter 3. This method is explored in CHD and was originally developed to handle smaller datasets that are associated with CHD as current methods produced problems when splitting the timescale and running out of events. The method was published (3), which also detailed that the work was performed in heart disease, increasing the awareness of both the relative survival methodology in CHD and the spline methodology in relative survival.

The method is not restricted to heart disease and can be applied in other therapeutic ares. As a test this method has been applied in cancer, which uses relative survival techniques routinely, and was found to be a good fit to the respective datasets. A presentation of the method was made at a Stata workshop where the therapeutic area was cancer and the slides are shown in appendix B. The cancer datasets are much larger and the speed advantage is very important here.

The most challenging part of fitting these splines is in the selection of knots. The frequency and location of knots is often the talking point of any discussion on splines, however the

investigation shown here shows that this is not a big issue. The evidence strongly supports the notion that as long as the knots are close to sensible then the model produced will be suitable and the estimates reliable. There are better fitting models to be had by tweaking knot positions but it is very unlikely that a model produced will show poor and unreliable estimates to the *default-values* user. In fact the default values used in these models often provide the best fit to these data. If the knots are all placed in one half of the data then the fit is not as good, as shown in sections 5.3.2 and 5.4.3.

The method is also extendable to include period analysis which is so far a fairly new concept within CHD and this will be explored in chapter 8. The method has been shown to be unbiased in its estimation of the predicted survival, hazard and hazard rate ratios as is discussed in chapter 7, where the assumptions made in relative survival are assessed using a simulation study. The method has been fully implemented in Stata and this is described in chapter 6, which follows.

Chapter 6

Implementing Flexible Parametric Models in Stata

6.1 Introduction

In the previous chapter (chapter 5) the introduction of a new relative survival model was described that utilises splines in a flexible parametric model. This model was shown to hold various advantages over current methods that model covariates in relative survival. This chapter will describe how the methodology was implemented in Stata and how the program can be used in practice. The program is named `strsrcs` which roughly relates to Survival Time using Relative Survival with Restricted Cubic Splines and consists of five `.ado` files that are required for the program to be used. These are as follows:

- `strsrcs` which is the wrapper holding the main program file
- `rcs` is used to calculate the restricted cubic spline terms
- `strsrcs_mlh` uses `ml` to maximise the likelihood on the log cumulative hazards scale
- `strsrcs_mlo` uses `ml` to maximise the likelihood on the log cumulative odds scale
- `strsrcs_pred` allows the user to predict post-estimation

Rather than develop a piece of software that could be used in the dataset required for this thesis it was deemed necessary to provide the software for other users with very different datasets. In order to do this the program needed to be fully generalisable to other datasets. This was achieved by using Stata's `ml` machinery to implement the maximum likelihood with the program, `strsrcs`, being a *wrapper* around this. However this was not the only thing necessary for a successful program as post-estimation is just as important. A user-friendly post-estimation enables the user to easily obtain predictions of relative survival,

excess hazards and contrasts between groups in the form of excess hazard rate ratios.

This program has been submitted to SSC (157) and is available to describe or install from within Stata using:

```
ssc describe strsrcs
ssc install strsrcs
```

6.2 Calling the program

6.2.1 Setting up a dataset

In order to model the excess mortality the dataset needs to contain expected mortality rates, and as the data is not split by time this is a straightforward process. The expected mortality rates need to be merged onto the dataset for each individual at the event time as was shown in equation 5.26 where $h^*(t_i)$ is the expected mortality at the event time for the i^{th} individual. So for example an expected mortality rate by age, sex and period would need to be merged on at the age and year of the event using Stata code similar to that shown below:

```
gen age_event  = int(min(age_hosp + _t,99))
gen year_event = int(min(year_hosp + _t,2006))
sort  year_event sex age_event
merge year_event sex age_event using expected_mortality_dataset, nokeep
```

where `_t` is generated by using the Stata command `stset` as the time in the study. Using this code a new integer is created for age at event using the minimum value of either the age at event (age at hosp + time in study) or an age of 99. This is to ensure that values of $\geq$100 are set equal to 99 as there are no expected mortality rates available for those with values $\geq$100. A similar integer is created for year in which the minimum of 2006 or year of hospitalisation + time in the study is used, as 2006 is the latest date for which expected mortality data are available at the time of writing.

For each row, which relates to a single individual there is now an expected mortality rate, at the time of death or censoring, which can be used in relative survival.

6.2.2 Program syntax

Stata syntax for the `strsrcs` command is fairly straightforward as it follows standard Stata syntax for estimation commands, as shown below:

```
strsrcs   [varlist] [if exp] [in range]
          [, model_complexity bhazard(varname) scale(hazard|odds)
          strata(strata_varlist) noconstant snoconstant
          inits(name) eform orthog maximize_options ]

where model_complexity  is one of

          { df(#)} knots(knotlist)} }
```

This can be found in the help file associated with this command along with details on each option and how to run the program, which will also be detailed here.

Required options

In order to run the command three options have to be selected, the `bhazard` , `scale` and either the `df` or `knots` are required. The first option `bhazard` is where the variable that contains the expected mortality rate, or background hazard at death, is stated. `scale` is to select models on the cumulative hazards or cumulative odds scales, generally the cumulative hazards scale is suggested as the appropriate scale, but there will be more on proportional odds, and the feasibility of proportional odds modelling, later in this chapter. The final necessary option is one of `df` or `knots` which indicate the number and location of the spline joining points referred to as knots. To use the `df` option an integer from 2 to 6 is used and refers to the degrees of freedom to be adopted by the model, 2df creates one internal knot and 6df five internal knots. These knot positions were adopted by Royston and Parmar(135) and take the form shown in table 6.1. The knot positions are calculated at the centile of the distribution of log survival times for those that die.

Degrees of Freedom	Knot Position†
2	50
3	33 67
4	25 50 75
5	20 40 60 80
6	17 33 50 67 83

Table 6.1: Knot position as selected when using the df option. † Percentiles of all death times.

Use of the `knots` option leaves the user able to define the number and location of knots as per their requirements. The values in knotlist are taken to be centile positions in the distribution of the uncensored event [i.e. where `_d==1`] log times. For example `knots(20 40 60 80)` would produce identical results to the option `df(5)`.

Optional options

The option `strata` allows the modelling of time-dependent effects. For example to allow age to vary by the splines it would be included in the strata option. The `noconstant` and `snoconstant` prevents the inclusion of the constant in the baseline group and across the splines respectively. The final specific option is `inits` which allows the user to give initial values for the maximum likelihood. Allowing the `maximize_options` means that the user can use various options available when using maximum likelihood including different convergence methods and setting maximum numbers of iterations.

The `eform` option is also available to report excess hazard rate ratios, rather than log excess hazard ratios. A further option, `orthog`, is available to orthogonalise the splines if there is too much correlation between them thus solving some problematic models convergence issues. This can also speed up the estimation process and convergence.

6.3 Program methodology

There are seven stages that the wrapper program, `strsrcs`, must process during each call of the program. The program itself is shown in appendix A.1.

The first stage checks that the options are appropriate. One stage of this is to check that one of `df` or `knots` are used and that `scale` and `bhazard` are also specified. The `bhazard` option is also checked to confirm that there are no missing values. The final check is for the time origin to determine if delayed entry models have been fitted, which are useful for period analysis. This will be explained further in chapter 8.

The next stage creates the spline variables based on either the degrees of freedom or some user defined knots using one of the two related options. The external knots are placed at 0^{th} and 100^{th} centiles with all other knots being placed at the default locations or where the user selected. This is performed by calling the `rcs.ado` program using log times created using the centile positions requested. The program is shown in appendix A.2. The third stage calculates the derivatives for the spline terms using the same program.

The fourth stage orthogonalises the splines if the user used the `orthog` option. This is performed within the `rcs` program using a built in Stata command `orthog` to orthogonalise the spline terms. The orthogonalisation is performed using the Gram-Schmidt method (150) described in section 5.2.7. To call the `rcs` program the following syntax is used:

```
rcs varname [if] [in , knots(numlist) [gen](stubname)
                       [dgen](stubname) orthog
```

This is fairly straightforward for the main part with the compulsory option `knots` being where the location of each knot is to be placed, in log time. To generate the spline terms `gen` is required but the derivatives are optional using the `dgen` option. The final option is to orthogonalise the splines by removing the correlation between them and forcing a mean of zero and standard deviation of one.

Once the spline terms and their derivatives have been calculated, and perhaps orthogonalised, the next stage is to obtain the initial values. The default initial values, obtained by Stata, were found to be fairly poor and for this reason the first term of the splines, i.e. the linear function of log time was used, which is equivalent to a Weibull model, to obtain a suitable starting point. The program can be seen in appendix A.2.

The penultimate stage fits the model to the data using either `strsrcs_mlh` for a hazards model or `strsrcs_mlo` for an odds model. When these are called they set up the likelihood function using the created spline terms and covariates that are required which is then maximised by `ml`. The final stage is set out to display the results to be similar to a standard Stata regression model output.

After the final stage model information is stored so that the user can check this using the built in `ereturn` function in Stata. Using this the model scale (hazard or odds) and the variables modelled, including the background mortality variable used, are all returned. The most important information returned is stored in `'e(knots)'` which gives the location of the knots (in log time). The maximisation methods and results (stored in matrices) can also be obtained from these returned values.

6.3.1 The ML programs

The `strsrcs_mlh.ado` program is as shown below for models on the hazard scale, note that the lines have been numbered in order to explain the workings of the code more clearly:

```
01. program strsrcs_mlh
02.   args lnf xb s1 s2 s3 s4 s5 s6 s7 s8 s9 s10 s11 s12
03.   tempvar ht st  st0
04.   local p: word count $ML_y
05.   local q=‘p’-1
06.   local del_entry = 0
07.   qui summ _t0 , meanonly
08.   if r(max)>0 local del_entry = 1
09.   forvalues i = 1/‘q’ ///
10.   {
11.      local ds "‘ds’ ‘s‘i’’ * _d${ML_y‘i’}"
12.      if ‘i’ != ‘q’ local ds "‘ds’ + "
13.      local sp "‘sp’ ‘s‘i’’ * ${ML_y‘i’}"
14.      if ‘i’ != ‘q’ local sp "‘sp’ + "
15.      local sp0 "‘sp0’ ‘s‘i’’ * _s0${ML_y‘i’}"
16.      if ‘i’ != ‘q’ local sp0 "‘sp0’ + "
17.   }
18.   quietly generate double ‘st’=exp(-exp(‘sp’ + ‘xb’))
19.   qui generate double ‘ht’=${ML_y‘p’}+(1/_t)*(‘ds’)*exp(‘sp’+‘xb’)
20.   qui replace ‘lnf’ = _d*ln(‘ht’)+ln(‘st’)
21.   if ‘del_entry’ == 1 ///
22.   {
23.      qui generate double ‘st0’ = exp(-exp(‘sp0’ + ‘xb’)) if _t0>0
24.      qui replace ‘lnf’ = ‘lnf’ - ln(‘st0’) if _t0>0
25.   }
26. end
```

The first line of the code defines the program name and the last line ends the program. Line two limits the number of knots to twelve, as the need for more than five is fairly rare due to over fitting, although this could be altered by the user if more were required for use in other datasets. The fourth line counts the number of 'y' terms in the model, which includes the spline variables. The next line subtracts one term as required. For example, a five knot model with three internal knots only requires four parameters to be calculated. Lines 11 to 14 set up the linear predictor for the derivatives of the splines (`ds`) and the spline terms themselves (`sp`), which are stored in a local macro. This is achieved by looping over the number of spline variables and multiplying them by the relevant term in the model.

Line 18 calculates the relative survival rate as defined in the previous chapter (chapter 5) and similarly the overall hazard rate is calculated in line 19. The log likelihood function is then calculated in line 20 using the values calculated in the two previous lines. Lines

6-8, 15-16 and 21-25 are all used for delayed entry which will be discussed in chapter 8 when using period analysis. The only difference between the standard and delayed entry models in the code are that a new time zero is created that means entry into the model does not have to begin at time zero.

For a model on the log cumulative odds scale the code is identical except for lines 18-19 and line 23 which are changed as shown below:

```
18.   quietly generate double ‘st’=1/(1+exp(‘sp’ + ‘xb’))
19.   quietly generate double ‘ht’=${ML_y‘p’} + (1/_t)*(‘ds’)* ///
              exp(‘sp’ + ‘xb’)*(1/(1+exp(‘sp’ + ‘xb’)))
...
23.   quietly generate double ‘st0’ = 1/(1+exp(‘sp0’ + ‘xb’)) if _t0>0
```

Both programs are shown in the appendix, sections A.3 and A.4 for the log cumulative hazards and log cumulative odds programs respectively.

6.4 Example of model fitting

6.4.1 Fitting a PEH model

To fit a proportional excess hazards (PEH) model with 3 internal knots using age groups the following command could be used:

```
strsrcs age_group1 age_group2, scale(hazard)
          bhazard(rate) knots(25 50 75) eform
```

which gives the following output when using the heart disease dataset described in chapter 2 where `age_group1` for patients aged between 60 and 75 years old and `age_group2` for those aged 75+.

```
------------------------------------------------------------------------------
             |     exp(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
xb           |
  age_group1 |   2.596439   .2450187    10.11   0.000     2.158006    3.123946
  age_group2 |   5.782711   .4949262    20.50   0.000     4.889669    6.838856
-------------+----------------------------------------------------------------
s1           |
       _cons |   .6924478     .03286    21.07   0.000     .6280433    .7568523
-------------+----------------------------------------------------------------
s2           |
       _cons |   .0171798   .0021053     8.16   0.000     .0130535    .0213062
-------------+----------------------------------------------------------------
s3           |
       _cons |   .0004994   .0028727     0.17   0.862    -.0051311    .0061299
-------------+----------------------------------------------------------------
s4           |
       _cons |  -.0225072   .0046971    -4.79   0.000    -.0317132   -.0133011
------------------------------------------------------------------------------
```

The output is useful for assessing covariate effects and investigating significance using confidence limits or p-values. For this output it can be shown that the excess hazard ratios are both significantly different to the baseline age group (patients aged < 60) as both the p-value is highly significant and the confidence limits do not contain 1. Patients aged ≥ 75 have an increased risk of excess mortality around 5.8 times that of a patient aged < 60.

6.4.2 Fitting a non-PEH model

To fit a non-proportional excess hazards (non-PEH) model with 3 internal knots using age groups split as described above we could use the following command:

```
strsrcs, scale(hazard) bhazard(rate) df(4)
         strata(age_group1 age_group2) eform
```

Running this gives the following output when using the LRI CCU heart disease dataset:

```
------------------------------------------------------------------------------
             |     exp(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
xb           |
  age_group1 |   3.750178   2.193575     2.26   0.024     1.191687    11.80162
  age_group2 |   7.482594   3.957859     3.80   0.000      2.65348    21.10029
-------------+----------------------------------------------------------------
s1           |
  age_group1 |   .0524651   .0989708     0.53   0.596    -.1415141    .2464443
  age_group2 |   .0648032   .0895356     0.72   0.469    -.1106834    .2402898
       _cons |   .6429139   .0774252     8.30   0.000     .4911633    .7946645
-------------+----------------------------------------------------------------
s2           |
  age_group1 |    .006089   .0061495     0.99   0.322    -.0059638    .0181418
  age_group2 |   .0028161   .0056025     0.50   0.615    -.0081646    .0137968
       _cons |   .0140816   .0047668     2.95   0.003     .0047388    .0234244
-------------+----------------------------------------------------------------
s3           |
  age_group1 |  -.0100201   .0080703    -1.24   0.214    -.0258377    .0057974
  age_group2 |  -.0068299   .0075098    -0.91   0.363    -.0215487     .007889
       _cons |   .0065658   .0062363     1.05   0.292    -.0056571    .0187888
-------------+----------------------------------------------------------------
s4           |
  age_group1 |    .014816   .0127068     1.17   0.244    -.0100889    .0397209
  age_group2 |   .0149752    .012049     1.24   0.214    -.0086404    .0385908
       _cons |  -.0337768   .0097189    -3.48   0.001    -.0528256   -.0147281
------------------------------------------------------------------------------
```

This output is more complex as both age groups are allowed to vary within each spline term but the actual values presented are difficult to interpret. The p-values suggest that the age groups could be considered as proportional as none of them are significant except the linear predictor for the covariate effects (`xb`). This output is less useful for assessing covariate effects than in a PEH model but it is possible to investigate significance using confidence limits or p-values. However by using flexible models we are more interested in the fitted shape of the model. To do this the fitted survival, hazard and, for the non-PEH model, the excess hazard rate ratios are required.

6.5 Post-estimation output

Post-estimation commands are commonplace with all statistical packages to make the assessment of model fit more accessible. The `strsrcs` command is no different and allows the prediction of the relative survival rate, excess hazard rate and excess hazard rate ratios. There are five options available for the predict statement, the first being the option to predict the confidence intervals, `ci`, and the second to change the significance level for these intervals, `level()`. The other three options allow the prediction of either

the excess hazard, relative survival or, in non PEH models, the excess hazard rate ratios. The program used to produce these results is shown in appendix A.5.

6.5.1 Prediction commands

Within the `strsrcs_pred.ado` file, shown in appendix A.5, the calculation of the predictions are generated through `predictnl`. As a reminder from chapter 5, the relative survival and excess hazard rates are obtained from:

$$R(t) = \exp\left[-\exp(s(z;\gamma) + \beta x)\right] \tag{6.1}$$

$$\lambda(t) = \frac{1}{t}\frac{ds(z;\gamma)}{dz}\exp(s(z;\gamma) + \beta x) \tag{6.2}$$

So on the hazards scale the relative survival is obtained in Stata using

```
qui predictnl double ‘lnH’=xb(xb)+‘rcs’ if ‘touse’, ‘survci’ ‘levelopt’
qui gen double ‘newvarname’=exp(-exp(‘lnH’)) if ‘touse’
```

The relative survival is predicted using the baseline values $[\beta x]$ and the spline terms $[s(z;\gamma)]$ which is on the log cumulative excess hazards scale. This is then converted to the survival scale using the `exp(-exp())` transformation. The excess hazard is calculated as

```
qui predictnl double ‘lh’=-ln(‘timevar’)+ln(‘drcs’)+(xb(xb)+‘rcs’) ///
                      if ‘touse’, ‘hazci’ ‘levelopt’
qui gen double ‘newvarname’=exp(‘lh’)
```

Where `‘timevar’` is t and `‘drcs’` are the derivatives of the splines $\left[\frac{ds(z;\gamma)}{dz}\right]$. As is clear these predictions use the formulae shown above for both the relative survival and excess hazard rate to determine the predicted values. Both `‘rcs’` and `‘drcs’` are the relevant linear predictors stored in local macros, see appendix A for more details. The excess hazard rate ratios are also determined using `predictnl`, although their calculation is more complex as shown in appendix A.5 and the calculation is described in section 5.4.1. Confidence intervals are calculated if the `ci` option is used.

6.5.2 Estimating relative survival

In order to predict the relative survival function a predict statement is required, for example:

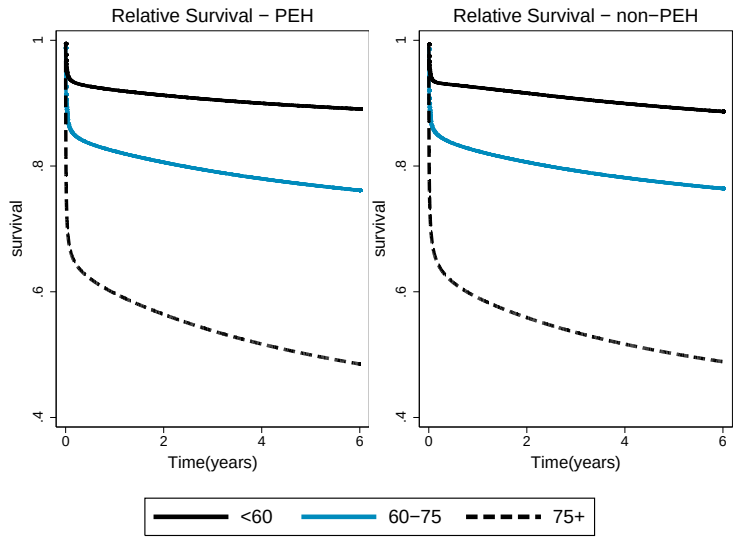


Figure 6.1: Predicted relative survival estimates using STRSRCS

```
predict survival, survival
```

This generates a new variable labelled `survival` which can be plotted as shown in figure 6.1. This gives the relative survival estimate for the fitted covariates, plotted using a `twoway line` graph with the `if` statement. In this figure a PEH and non-PEH plot are shown with the PEH plot on the left hand side. The `predict` command is the same for a PEH and non-PEH model. A simple interpretation of the figure would show that the oldest age group has a much worse relative survival than the two younger age groups, with the youngest age group having the best survival relative to the general population. Confidence intervals are calculated using the `ci` option.

6.5.3 Investigating the excess hazard

In order to predict the excess hazard function a predict statement is required, for example:

```
predict ehaz, hazard
```

This generates a new variable labelled `ehaz` which can be plotted as shown in figure 6.2. This figure shows a PEH and non-PEH plot as fitted from the models given as examples in the previous sections above. Here it is clear that when fitting a PEH or a non-PEH model the same `predict` statement does provide the relevant predictions for time-dependent effects if they have been fitted as the non-PEH plot is clearly different to the PEH plot. A simple interpretation would lead to the conclusion that under both a PEH and non-PEH model the oldest age group has the highest excess hazard rate and the youngest age group the lowest excess hazard rate as was suggested in the relative survival plot shown in figure

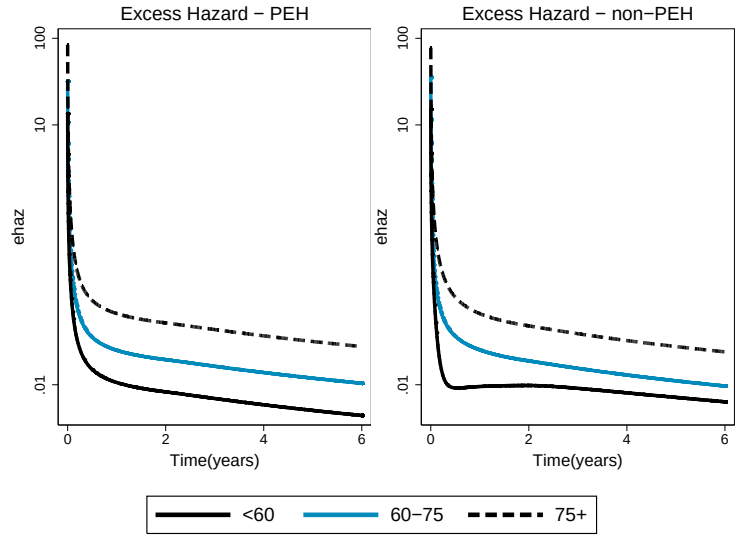


Figure 6.2: Predicted excess hazard estimates using STRSRCS

6.1.

6.5.4 Excess hazard rate ratios

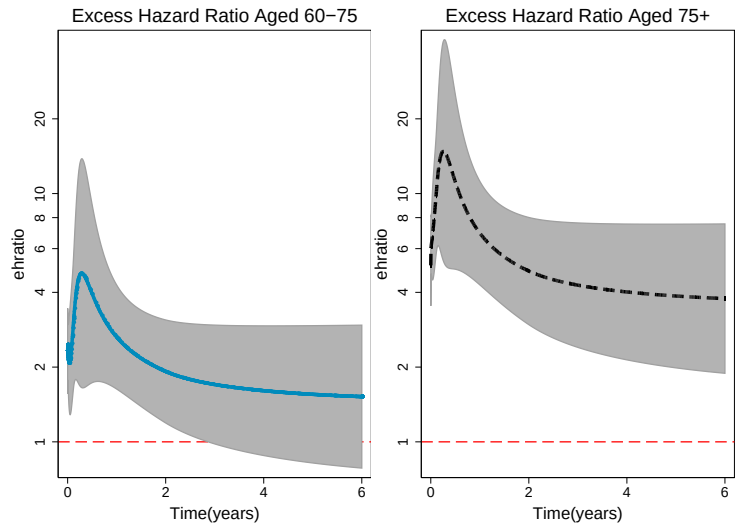


Figure 6.3: Predicted excess hazard ratios with confidence intervals estimates using STRSRCS

In order to predict the excess hazard rate ratio function a predict statement is required, with the calculation for these shown in appendix A.5. This is only valid for models which include time dependent effects, for example:

Method used	$60-70$ years old	75 years or older	
Hazards	Relative survival STRSRCS	2.356 (1.91 , 2.90)	6.252 (5.13 , 7.63)
	All cause STPM	2.715 (2.30 , 3.21)	7.877 (6.71 , 9.24)
	All cause STRSRCS*	2.715 (2.30 , 3.21)	7.877 (6.71 , 9.25)
Odds	Relative survival STRSRCS	2.537 (2.04 , 3.16)	7.726 (6.22 , 9.59)
	All cause STPM	3.120 (2.60 , 3.75)	11.695 (9.72 , 14.07)
	All cause STRSRCS*	3.120 (2.60 , 3.75)	11.695 (9.72 , 14.07)

Table 6.2: Comparing STRSRCS to STPM. * Using an expected mortality rate of 0 in comparison to STPM on the log cumulative hazards and log cumulative odds scale

```
predict ehratio, hratio ci
```

Here the `ci` option is also used. This option is available for all predict statements if required and gives the confidence interval for the required estimates. The default value is to give a 95% interval but this can be changed using the `level()` option. For this example where there are two groups this generates two new variables labelled `ehratio1` and `ehratio2`, which are the point estimates and two other new variables `ehratio1_lci` and `ehratio1_uci`, which are the lower and upper limits of the confidence interval for `ehratio1` respectively with a further two varibles created for the confidence limits of `ehratio2`. These can be plotted as shown in figure 6.3.

It is fairly complex to generalise due to the different number of knots and which factors are time-dependent and the code is shown in appendix A.5. The interpretation is fairly strightforward. In comparison to the youngest age group the $60-75$ year olds, shown on the left hand side of the plot are shown to have a slightly higher excess mortality but after three years the confidence interval crosses one suggesting that there is no significant difference from this point on. The oldest age group has a much higher excess mortality than both the younger age groups and the confidence interval is much bigger than one at its lowest point suggesting strong evidence that the oldest age group are at a significantly higher excess mortality than the youngest age group.

6.6 Model checks

6.6.1 Comparing to STPM

To determine that the effects modelled are accurate in terms of the code used it is possible to compare it to the current all-cause survival using the Stata command `stpm`. To do this the expected hazard rate can be set to zero and the models should be the same. The models are confirmed as equivalent in table 6.2 where the bottom two rows show the results from `stpm` (147), the all-cause program, and `strsrcs` the program written for relative survival but using a zero expected hazard.

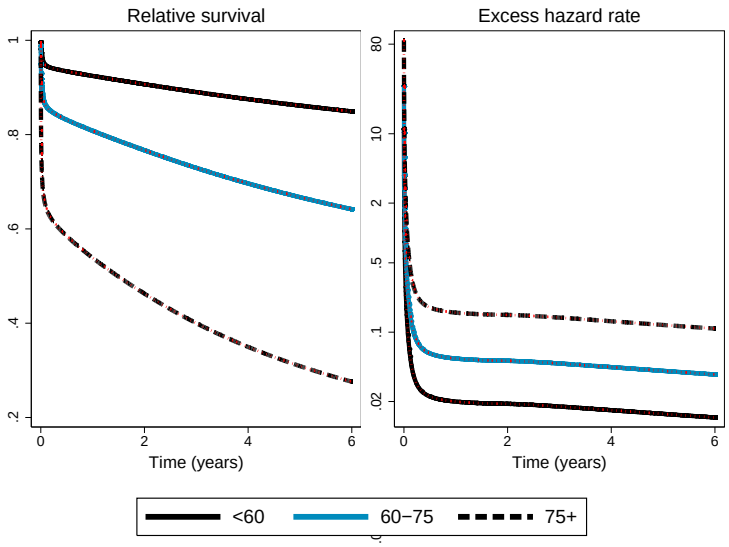


Figure 6.4: Predicted relative survival and excess hazard rate using STPM and STRSRCS with no background mortality rate. The estimates are completely overlaid

Also in table 6.2 are the results from a relative survival model, i.e. `strsrcs` using an expected mortality rate from the general population. Here it is clear that the effect sizes are reduced when the expected mortality rate is included by a significant amount in the oldest age group as the new interval does not include the standard survival estimate. This is shown on the log cumulative hazards scale and log cumulative odds scale.

Figure 6.4 shows the estimates from the models shown in table 6.2 on the bottom two rows of data from the hazards model where a zero background hazard rate was used. This figure shows that the estimates are identical with the thick lines representing `stpm` and the dotted lines `strsrcs`. The estimates are completely overlaid and this confirms that the code is accurate as the estimates are identical when using a standard survival model to that of `stpm`.

6.7 Conclusions

The program `strsrcs` is generalisable to all datasets in terms of its features, whether the models themselves are suitable is up to the user to determine. The output presented is in a standard Stata format and the predict command is also used, which is a regular theme for regression models, and particularly survival models. The method for calling `strsrcs` is not alien to Stata and as such all users familiar with Stata syntax should have no difficulty in running the models. The code itself produces results in agreement with `stpm` when used as an all-cause model allowing the user to have confidence that the method is not bugged. The program has been used in heart disease and cancer with success and in terms of heart

disease with greater success than standard relative survival methods. It is also useful for population-based cancer studies as shown in appendix B. Finally there is a help file for `strsrcs` which is included in the package that is available for download on SSC to all Stata users.

Chapter 7

Simulation Study of the Impact of Disease Prevalence in Relative Survival

7.1 Aims of the simulation

This chapter aims to investigate the impact that a high disease prevalence has on absolute (relative survival) and relative (excess hazard rate ratio) effects. In order to achieve this simulated Weibull survival time data was obtained and estimates were compared to the values obtained using the nationally obtained expected mortality rates (30) that are used throughout this thesis. During the investigation a bias was observed in the excess hazard rate ratios when using piecewise models with insufficient split times, which was an unexpected side effect that further supported the use of the spline based models developed in chapter 5, as these were not biased.

7.2 Motivation

The expected survival rate used in the calculation of relative survival is defined as the rate for a population similar to the patient group but free of the disease under study. However, as population lifetables include all causes of death it may seem desirable to adjust these values to remove the deaths due to the disease of interest. However, in practice it was found that the prevalence of each specific disease in the population was so small, in cancer, that it had little to no impact on the findings (26). In fact Ederer found that by removing all cancer deaths from the expected survival rates that this in fact over-corrected the expected rate. Although the control group should be free of the disease of interest at

entry to the study it does not follow that the control group should not be subject to the risk of the disease and subsequently die from it. This highlights one of the key assumptions that is made in relative survival.

These assumptions were assessed by Ederer *et al.* (26) in cancer over fifty years ago and is routinely referenced as a reason why this assumption is never assessed. This appears fair in cancer studies as the prevalence of a specific cancer under investigation is very small in terms of the general population. Heart disease is very different however, as the largest single killer in the UK it has a much higher prevalence when compared to the most common cancers. Also previous AMI patients are excluded from the cohort as they are assumed to be at a higher risk of death from their successive AMIs but there will be many people in the general population that are in this category. Therefore the prevalence and increased risk of death that this brings to the general population needs to be assessed in order to determine if adjustments are necessary to the expected mortality/survival rates used in heart disease.

The relative survival is not the only outcome of interest, as excess hazard rate ratios will also be assessed. This is important as the bias when comparing two groups, for example treatment and placebo, may be in the same direction and only investigating absolute effects would not detect this. It should also be noted that the prevalence of the disease in the general population will mostly be made up of subjects who have survived their MI for a fairly long time, i.e. in the main, they will be healthy survivors.

7.3 The impact of prevalence

In a lifetable the expected probability of surviving the next year, p^*, is merged in from the general population lifetable. This will include deaths due to the disease. If the prevalence of the disease is θ then p^* can be broken into two components made up of the probability that someone with or without a history of MI will survive so that,

$$p^*(t) = (1-\theta)p^N(t) + \theta p^{MI}(t) \tag{7.1}$$

Where $p^N(t)$ is the survival probability for those that have no history of MI and death is due to a disease other than that under study, which is something that would be used ideally in practice but is difficult to obtain. This leaves $p^{MI}(t)$, which is the survival probability for those that have a history of MI and when θ is low p^* is a good approximation as θp^{MI} becomes small, which is the assumption made by relative survival. Therefore the value that is obtained from routine data, p^*, includes deaths due to an MI as well as other causes. In cancer the value of θ is so low that an assumption of $p^* = p^N$ is usually reasonable.

It should also be noted that if the probability of death is the same, i.e. $p^N = p^{MI}$, then using p^* is fine as $p^* = p^N = p^{MI}$.

Equation 7.1 could be rearranged in terms of the probability of survival for those with a history of MI,

$$p^{MI}(t) = \frac{p^*(t) - (1-\theta)p^N(t)}{\theta} \tag{7.2}$$

This can also be arranged in terms of the survival probability for those without a history of MI

$$p^N(t) = \frac{p^*(t) - \theta p^{MI}(t)}{1-\theta} \tag{7.3}$$

If the values of θ and p^{MI} are known then it is possible to estimate p^N for various ages, sex and periods by incorporating p^*, which can be obtained from routine data. The value of p^N is of major interest as the value of p^* is known.

An assumption will be made that for those with a history of MI an assumption of increased risk will be made. For simplicity it will be assumed that this is constant over time. It is therefore possible to examine the range of p^N for various possible prevalences of MI by including p^{MI} in terms of p^N i.e.

$$logit\left(p^{MI}(t)\right) = \alpha + \beta \tag{7.4}$$

where $\alpha = logit\left(p^N(t)\right)$ and β relates to the log odds ratio of survival, i.e. if $\exp(\beta) = 0.5$ then this would represent a doubling of the odds of mortality for people with a previous MI. This has the effect of reducing the probability of survival. If the log odds ratio is equal to zero then $p^N = p^{MI}$. Therefore p^* can be written as

$$p^*(t) = (1-\theta)p^N(t) + \theta\left(invlogit\left[(\alpha) + log(\beta)\right]\right) \tag{7.5}$$

Included in table 7.1 are the probabilities of one year survival for a 60, 70, 80 or 90 year old male or female in 2003. The values given as p^* are probabilities that were obtained from routine data using GAD tables (30). This highlights the difference between older men and women's expected survival and the natural reduction in expected survival as age increases. Note that the changes in prevalence will not affect these values. The table also shows estimates of p^N which are calculated using the given prevalence and odds ratio of mortality. Here it is clear that the largest differences between p^* and p^N are in the very

Prevalence	Patient Age	Male			Female		
		p^*	p_1^N	p_2^N	p^*	p_1^N	p_2^N
7%	60	0.9900	0.9893	0.9896	0.9938	0.9933	0.9935
	70	0.9738	0.9720	0.9729	0.9839	0.9829	0.9834
	80	0.9243	0.9197	0.9220	0.9488	0.9456	0.9472
	90	0.8150	0.8061	0.8103	0.8470	0.8391	0.8428
15%	60	0.9900	0.9885	0.9892	0.9938	0.9928	0.9933
	70	0.9738	0.9700	0.9719	0.9839	0.9816	0.9828
	80	0.9243	0.9145	0.9193	0.9488	0.9419	0.9453
	90	0.8150	0.7959	0.8048	0.8470	0.8301	0.8381

Table 7.1: Table of expected probabilities for surviving the next year by sex and age. p_1^N is for an OR of 0.5, p_2^N is for an OR of 0.67.

elderly patients with the highest prevalence (15%) and highest odds ratio of mortality, ($\exp(\beta) = 0.5$).

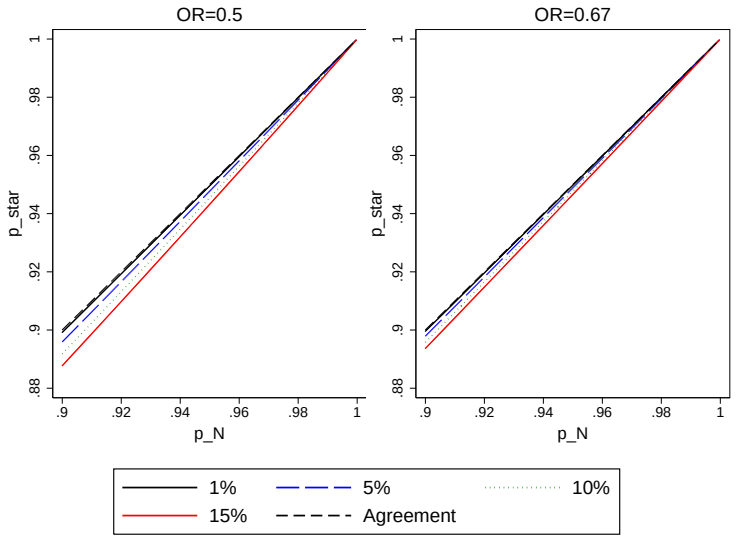


Figure 7.1: Plot of agreement between p^* and p^N assuming a range of prevalences and different odds ratio of mortality

It is possible to use the relationship between p^* and p^N that was shown in equation 7.5 to plot the level of agreement for several different levels of prevalence. This has been shown in figure 7.1 for four different prevalences, 1%, 5%, 10% and 15%, and an odds ratio of mortality increase for those with MI of double ($\exp(\beta) = 0.5$) and a 50% increase ($\exp(\beta) = 0.67$). This is performed by allowing p^N to be within the range of 0.9-1 which is a fair range for patients who are around 80 years old or younger as indicated by table 7.1. The line of perfect agreement is shown and this is where the estimates of p^N and p^* are identical in that there is a zero prevalence of MI or, more importantly, the odds ratio is one. The increase in prevalence increases the disparity between p^N and p^* towards the lower end of the scale. It should be noted that for the majority of patients, i.e. those

under 75 years old, the probabiilty is estimated as 95.3% for males and 97.0% for females which, as shown in figure 7.1, remains fairly unbiased for all of the assumed prevalences. However where the odds ratio of mortality is twice that of the non-MI population the lines of agreement clearly seperate for patients with a lower expected probability of survival.

For patients with an expected survival that is given as 0.9 the estimate will be biased by around 1.2% where the prevalence is high (15%) and the odds ratio are high ($\exp(\beta) = 0,5$), i.e. $p^N = 0.8877$ whereas the estimate used in the model is $p^* = 0.9000$ which would result in an overestimate of the MI patients relative survival. These estimates are only assessed for a single year and these differences will accumulate over time and the best way to estimate how these differences will stack up is to run a simulation.

7.4 Simulation of prevalence over time

The aim of this section is to simulate the survival times under different assumptions of prevalence and odds ratio of mortality. This will be assessed by different age groups as age is the primary factor in determining the expected mortality for a member of the general population. Simulation will be used to provide estimates of p^N which can be compared to p^* in order to determine how biased the estimate of p^* is once all deaths due to MI are removed. In practice if p^N could be used then this would make the assumption being assessed here true, however this value is not easily obtained as adjustments would be needed to the standard lifetable results and values of θ and p^{MI} are unlikely to be known, making adjustments very difficult.

7.4.1 Methodology

Let β be the odds ratio for increased/decreased odds of survival for those diagnosed with MI compared to those without a diagnosis, so that

$$p^{MI} = \frac{\left(\frac{p^N}{1-p^N}\right)\beta}{\left(\frac{p^N}{1-p^N}\right)\beta + 1} \tag{7.6}$$

Which when substituted into p^* gives

$$p^* = (1-\theta)p^N + \frac{\theta p^N \beta}{(1-p^N)\left(\frac{p^N\beta}{1-p^N} + 1\right)} \tag{7.7}$$

This can be simplified to

$$p^* = p^N - p^N\theta + \frac{\theta p^N \beta}{p^N\beta + 1 - p^N} \tag{7.8}$$

Using the same denominator would give

$$p^* = \frac{{p^N}^2\beta - {p^N}^2\beta\theta + p^N - p^N\theta - {p^N}^2 + {p^N}^2\theta + \theta p^N\beta}{p^N\beta + 1 - p^N} \tag{7.9}$$

Rearrange this to remove the denominator and set equal to zero

$$0 = {p^N}^2\beta - {p^N}^2\beta\theta + p^N - p^N\theta - {p^N}^2 + {p^N}^2\theta - p^N\beta p^* - p^* + p^N p^* + \theta p^N\beta \tag{7.10}$$

Which can be simplified to

$$0 = {p^N}^2(\beta - \beta\theta - 1 + \theta) + p^N(1 - \theta - \beta p^* + p^* + p^*\beta) - p^* \tag{7.11}$$

This is a quadratic equation and this can be solved using,

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{7.12}$$

To give the final equation for p^N as

$$p^N = \frac{-(1 - \theta + \beta\theta + p^* - p^*\beta) + \sqrt{(1 - \theta + \beta\theta + p^* - p^*\beta)^2 - 4(1 - p^*)(\beta - \beta\theta - 1 + \theta)}}{2(\beta - \beta\theta - 1 + \theta)} \tag{7.13}$$

7.4.2 The algorithm used for simulation

A hypothetical study will be simulated where all patients are of the same age. This is unrealistic in practice, but helps to show the features that are of interest here. For a simulation to be effective it needs to follow a set algorithm in order to remain consistent. The algorithm used to simulate the data is shown below:

1. Obtain values for p^* for a known male of a certain age using routine data
2. Set the prevalence and odds ratio of mortality

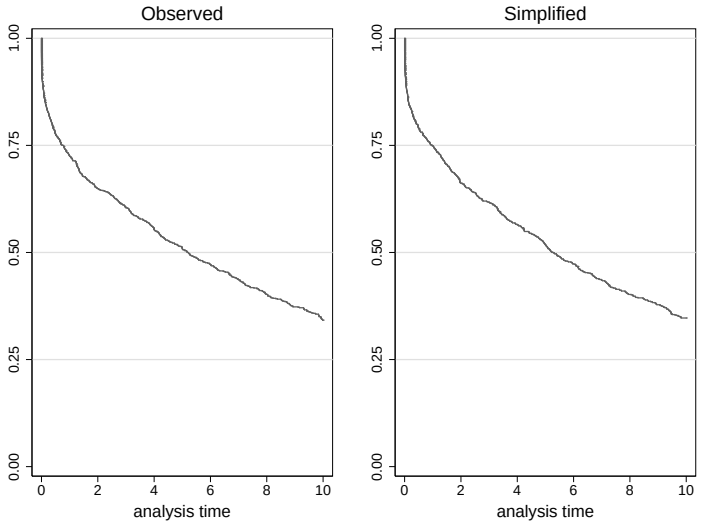


Figure 7.2: Plot of fitted survival curve from a Weibull distribution using the observed values and simplified values for λ and γ

3. Generate p^N using equation 7.13 with a known p^*

4. Obtain p^* and p^N for each year in a ten year follow-up. It is necessary to do this for each year as people age over follow-up time.

5. Generate a dummy covariate using a uniform distribution where there is an equal chance of being in group a or b, for example - treatment groups.

6. Generate survival time for those with MI (t_{MI}) using a Weibull distribution as shown by Bender (158)

$$T = -\left(\frac{\log U}{\lambda \exp(\beta' x)}\right)^{1/\gamma} \tag{7.14}$$

   Where U is from a uniform$(0, 1)$ distribution. It should be noted that the Weibull distribution was used to simulate the relative survival for those that die from an MI.

7. The values for λ and γ were determined by fitting a Weibull model to the LRI CCU dataset, which gave $\lambda = 0.3109$ and $\gamma = 0.2781$. This was simplified to 0.3 for both values which made very little difference to the fitted curve as shown in figure 7.2.

8. Generate survival time for non MI group (t_{non}) using the expected survival probabilities obtained for p^N which includes deaths from all other causes.

9. Store time as the minimum of that obtained from either t_{MI} or t_{non}

10. Obtain relative survival estimates from lifetables using both p^* and p^N as the expected probabilities separately, i.e. obtain the estimates of relative survival twice

11. Model using Dickman Grouped models (115) for speed to obtain excess hazard ratios using both p^* and p^N as the expected probabilities separately, i.e. obtain the excess hazard ratio estimates twice

12. Model using the flexible parametric models described in chapter 5 producing relative survival, excess hazard and excess hazard ratio estimates.

Point 2 states that the prevalence and odds ratio were set. The prevalence was set at either 7% or 15% which means that the prevalence of patients with an MI in the general population is either seven or fifteen percent. The majority of these will be subjects who have survived a long time after their MI and are thus in a low risk group. These two values were decided upon after searching through *HeartStats.org* for figures of prevalence, where two values were found at 6.5% in males (159) and 15.7% in males (160) both aged seventy-five or older. While these numbers are very different it was decided that the 15% value would fairly represent the worst prevalence of MI and 7% as the best case. Note that in women these values were 7% and 8.1% for the same age group.

The values for the odds ratio were much harder to determine as there is no clear reference available or known value of the increased odds ratio of death a patient with MI will experience when compared to a patient with the same demographics but no MI. Therefore a 50% and 100% increase were used as the simulated best and simulated worst case respectively. Note that these are not the best or worst cases in actuality, but are simply labelled as such to make the interpretation easier. After consulation with Iain Squire (a clinical supervisor) it was decided that these values are feasible and higher estimates would be excessive, especially in today's environment where so much early intervention takes place and secondary prevention is extensive and effective.

In step 7 reference is made to figure 7.2, this plot shows the observed survival rate that was plotted under a Weibull distribution to the Leicester Royal Infirmary CCU dataset in order to determine the appropriate values of λ (given as 0.3109) and γ (given as 0.2781) as is shown on the left-hand side of the figure. These values were then simplified to one decimal place ($\lambda = \gamma = 0.3$) and these results are shown on the right hand side of the figure, highlighting that there is no difference in shape between the simulated Weibull model and the observed Weibull model.

More detail is needed to understand parts 8 and 9 of the algorithm. After the binary covariate, x, is generated this is multiplied by 1.0986 to give an expected hazard rate ratio of three. The time for patients with an MI is then calculated as shown in equation 7.14. This generates survival times over the ten year follow-up so the minimum of T and 10 was used for t_{MI}.

To account for the aging process for each of the ten years of follow-up a probability of survival excluding MI cases was generated using

$$T_i^N = \frac{-\log(U)}{-\log(p_i^N)} \tag{7.15}$$

Where $i = 1$ to 10. The minimum of T_i^N and 1 was stored as the survival for each of the ten years, unless the survival was < 1 in which case the patient died and had no further survival. The total time survived was then calculated as $t_{non} = \Sigma T_i^N$. The minimum value of t_{MI} and t_{non} was taken as the patient's survival time. Therefore if a patient had an MI this was determined by the death time in t_{MI} being smaller than t_{non}. Exponential survival times are assumed within each yearly time interval. In the analysis no distinction was made between mortality due to MI or other causes.

By excluding people with a history of MI from the expected hazard the model is in fact over correcting as p^N assumes that these non-MI patients are never going to have an MI. This is similar to a matched cohort study where an MI patient would be matched to a subject without history of MI where this subject may not develop an MI in the future, whereas in reality this is unlikely to be true for all controls.

Point 12 of the algorithm applies spline based flexible parametric models which used two internal knots placed at the 33^{rd} and 67^{th} percentiles of event times (3df model). This is likely to be over parameterised as the data is simulated from a Weibull distribution. In reality, the true distribution of the survival times is not known.

7.4.3 Simulating relative survival

There is a potentially huge range of scenarios that could be simulated for these data, with age, sex, year, prevalence, increased odds ratio and follow-up period all being able to change. To simplify this process two levels of prevalence (7% and 15%) and two levels of the odds ratio (0.5 and 0.67) have been adopted, as described in section 7.4.2. The plotted results are all from the spline based models, rather than the piecewise GLMs. It should be noted that when the GLMs were plotted the five-year relative survival estimates were very similar to the spline based models and it is unnecessary to show both, however, differences are shown in the tables.

Men hospitalised in the **year 1995** were also chosen as this allowed a **ten year** follow-up. The choice of year and sex was not seen as crucial for this simulation, as the aim of this chapter was to assess the impact of differential expected mortality. **Age** was varied using 60, 70, 80 and 90 year olds at the start of follow-up. The simulations were generated and produced lots of output and the following plots are shown for the four ages by varying levels of prevalence and odds ratio. Five year estimates are shown in the following tables and figures, with the figures giving results from the spline-based approach.

Sixty year olds

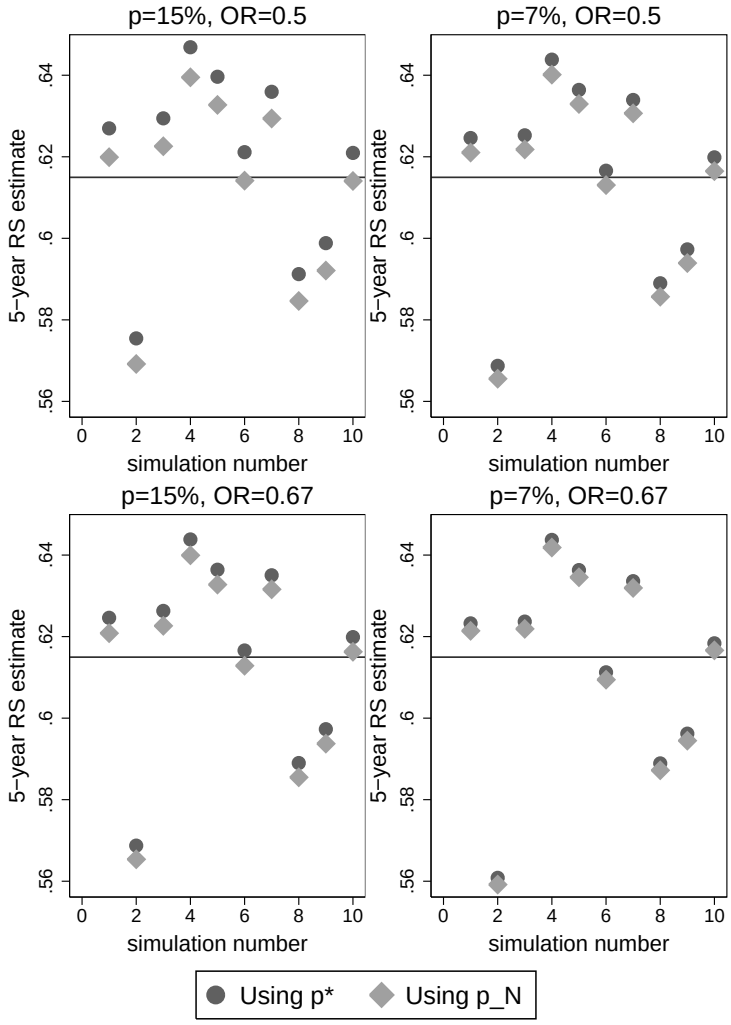


Figure 7.3: Plot of the first ten simulations estimated 5-year relative survival using both p^* and p^N for 60-year old males

Figure 7.3 shows the spline based estimates of relative survival for patients aged 60 years old from the first ten simulations observed. Note that these are limited to ten in order to allow a simple visual interpretation of the difference between using the standard expected mortality rate from lifetables (p^*) and an expected mortality rate that excludes all deaths due to the disease of interest, p^N.

The figure shows an odds ratio where $\exp(\beta) = 0.5$, which is a doubling in the odds of mortality for MI patients, on the top row and an odds ratio of mortality where $\exp(\beta) = 0.67$ (this equates to a 50% increased odds for MI patients) on the bottom row. The left hand column shows a prevalence level of 15% and the right hand column a prevalence rate of 7%. These are the youngest age group assessed. It should be noted that the horizontal line is the true value of the five year relative survival rate which is 0.61496 and is obtained

from a Weibull distribution as described in the algorithm (section 7.4.2).

The top left plot shows the results from the worst case scenario that was simulated in that the prevalence is high and there is a high odds ratio of mortality and shows the largest difference between p^* and p^N of approximately 0.68%. On the other hand the bottom-right scenario represents the best case scenario out of the simulations that were generated where the prevalence is relatively low and the odds ratio of mortality for patients with a history of MI is only increased by 50% when compared with patients with no history of MI. The other two plots relate to the non-extreme middle scenarios where either the prevalence or the odds ratio are high and the other is low.

Prevalence		OR	5-year RS		difference	Coverage		MSE	
			p^*	p^N		p^*	p^N	p^*	p^N
15%	LT	0.5	0.6213	0.6155	0.0057	0.9410	0.9440	0.0006	0.0006
	Splines		0.6221	0.6153	0.0068	0.9430	0.9470	0.0005	0.0005
7%	LT	0.5	0.6183	0.6155	0.0029	0.9480	0.9460	0.0005	0.0006
	Splines		0.6186	0.6152	0.0034	0.9420	0.9480	0.0005	0.0005
15%	LT	0.67	0.6185	0.6155	0.0030	0.9480	0.9450	0.0005	0.0006
	Splines		0.6188	0.6152	0.0036	0.9410	0.9470	0.0005	0.0005
7%	LT	0.67	0.6169	0.6154	0.0015	0.9480	0.9460	0.0005	0.0006
	Splines		0.6169	0.6152	0.0017	0.9440	0.9470	0.0005	0.0005

Table 7.2: Five year relative survival estimates, coverage and MSE after 1000 reps for sixty-year-olds for a lifetable approach and estimates from a spline approach. True value = 0.6149

Table 7.2 shows relevant statistics from 1000 simulations that can be used to fully assess the fit of the models. A piecewise approach and spline-based approach was fitted for all four variations of prevalence and odds ratio of mortality and these eight model statistics are shown here. The relative survival estimates for both p^* and p^N are given, as is the difference $(p^* - p^N)$ between these two values. The coverage of p^* and p^N are given as the proportion of estimates that included the, known, true value of relative survival in the confidence interval for these values. Finally the mean square error (MSE) is shown, which is calculated for p^* as,

$$MSE = Var(p^*) + \left[\frac{\Sigma p^*}{1000} - \exp(-0.3 * 5^{0.3})\right]^2 \tag{7.16}$$

This is an estimate of the variation in the estimates of p^*, and the calculation is performed in the same way for p^N by replacing it for p^*. The coverage and MSE are calculated for p^* because this is the value that would be used in practice and is being assessed to determine the potential bias. These values are also shown for p^N. Ideally the coverage will be close to 95% and the MSE small with little difference between p^* and p^N.

It is clear in table 7.2 that all the models produced similar estimates with the largest difference being that for the worst case model when using spline based models, which,

as stated earlier, is around 0.68%. This shows that for the youngest age group assessed during the simulation there is very little difference between the estimates of p^* and p^N. It is also clear that the estimates from the spline based models are similar to the lifetable estimates.

The coverage is high in all groups at around 95%. This is a sign that by adjusting the values of the prevalence and the odds ratio of mortality the true value of the mean is included a large proportion of the time. The MSE for the eight models are also very small and very similar for both a lifetable and spline based approach. The coverage and MSE for p^N are very similar to those given for p^*.

Seventy year olds

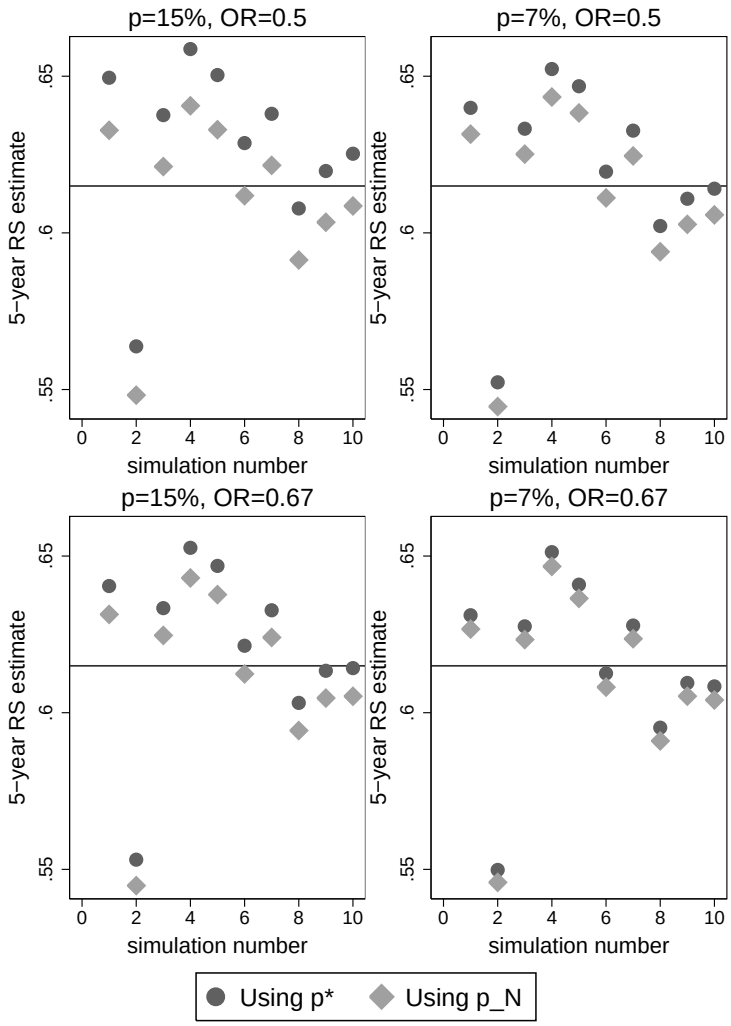


Figure 7.4: Plot of first ten simulations estimated 5-year relative survival using both p^* and p^N for 70-year old males

The previous simulation investigated sixty year olds who would then age to sixty-nine years if they survived the ten-year follow-up period. These patients have a fairly low expected mortality as shown in table 7.1, where the expected mortality from the general population for a sixty year old male is given as 0.9900. When this age is increased ten years the given expected mortality for males is reduced to 0.9738 for surviving the next year. It is at this age that expected mortality becomes increasingly important as year on year the expected mortality rate increases at a faster rate than it would in younger age groups.

Figure 7.4 shows the first ten simulations from the spline based models and is set out in the same way as the sixty-year old figure shown previously with the best case in the bottom-right cell and the worst case in the top left cell. This scale has altered as the differences have become larger than the sixty-year olds, but the scale is the same for the four plots. In this figure it does not appear that the two middle scenarios are different to each other, i.e. a high prevalence and low odds ratio of mortality appears to be similar to a low prevalence and high odds ratio of mortality. The worst case difference was the largest at approximately 1.67% for these ten estimates, which is about 1% larger than the sixty-year olds. The best case once again shows very little difference between p^* and p^N.

Prevalence		OR	5-year RS		difference	Coverage		MSE	
			p^*	p^N		p^*	p^N	p^*	p^N
15%	LT	0.5	0.6325	0.6158	0.0167	0.9080	0.9480	0.0011	0.0007
	Splines		0.6315	0.6153	0.0163	0.9060	0.9500	0.0009	0.0006
7%	LT	0.5	0.6242	0.6159	0.0082	0.9450	0.9490	0.0007	0.0008
	Splines		0.6234	0.6153	0.0081	0.9380	0.9510	0.0007	0.0006
15%	LT	0.67	0.6247	0.6159	0.0088	0.9430	0.9480	0.0007	0.0008
	Splines		0.6240	0.6153	0.0087	0.9350	0.9510	0.0007	0.0006
7%	LT	0.67	0.6201	0.6159	0.0042	0.9510	0.9550	0.0006	0.0008
	Splines		0.6195	0.6153	0.0042	0.9480	0.9460	0.0006	0.0006

Table 7.3: Five year relative survival estimates, coverage and MSE after 1000 reps for seventy-year-olds for a lifetable approach and estimates from a spline approach. True value = 0.6149

It is easier to assess the models using the available statistics from the 1000 simulations, which are shown in table 7.3. The estimates of five-year relative survival are very similar for both methods and the differences are therefore fairly small too. This evidence strengthens the use of the spline based models as they have thus far estimated similar estimates to the lifetable approach. The worst case model shown here gives similar estimates of coverage and MSE for the spline and lifetable approach, but both estimates of coverage are around 90%, which is around 5% below the value that would be hoped for. Coverage will be poor when there is bias, as there is here.

The MSE values and coverage values for the remaining three models are all very similar when comparing approaches but the best case model shows a higher coverage than the other three models. The largest difference is in the worst case model and is estimated at

1.63% after 1000 replications which is slightly lower than the average observed in the ten observations shown in figure 7.4. The coverage for p^N shows that it is unbiased as the coverage is high and the MSE is also low. The coverage is higher than the p^* estimate mainly due to the bias in the p^* estimates.

Eighty year olds

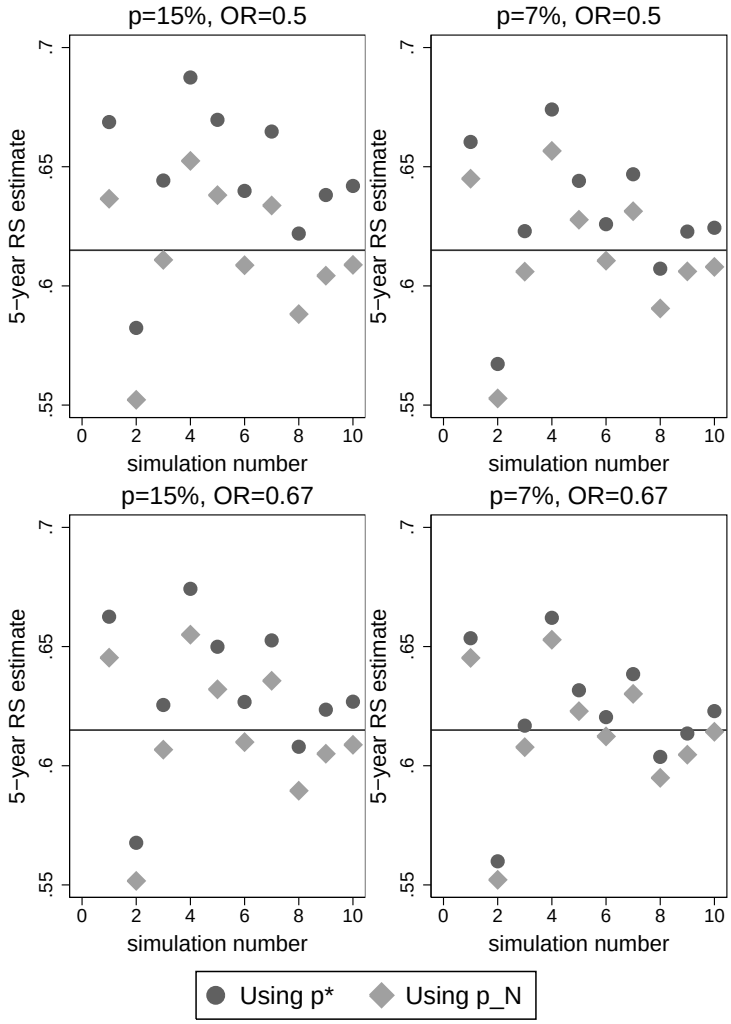


Figure 7.5: Plot of first ten simulations estimated 5-year relative survival using both p^* and p^N for 80-year old males

The age was increased another ten years to now simulate patients who are eighty years old at the start of follow-up where the expected probability of surviving until the patient is eighty-one for a male is given as 92.43% (table 7.1). The plots for the first ten simulations from the four scenarios are shown in figure 7.5. The scale has changed again as the differences have increased again.

The previous findings when using sixty and seventy year olds holds true here also. The worst case gives the largest difference, the best case gives the smallest difference and the middle scenarios are fairly indistinguishable from each other. The difference here is that the best case now shows a visible difference between p^* and p^N and the worst case model shows an average difference of 3.25% which is almost 2% more than patients who started follow-up when they were seventy years old.

Prevalence		OR	5-year RS		difference	Coverage		MSE	
			p^*	p^N		p^*	p^N	p^*	p^N
15%	LT	0.5	0.6576	0.6161	0.0415	0.8090	0.9580	0.0033	0.0013
	Splines		0.6470	0.6153	0.0316	0.8220	0.9460	0.0019	0.0008
7%	LT	0.5	0.6360	0.6159	0.0201	0.9140	0.9590	0.0011	0.0013
	Splines		0.6310	0.6153	0.0157	0.9220	0.9440	0.0011	0.0009
15%	LT	0.67	0.6380	0.6159	0.0221	0.9090	0.9560	0.0012	0.0013
	Splines		0.6326	0.6153	0.0173	0.9160	0.9450	0.0012	0.0009
7%	LT	0.67	0.6265	0.6160	0.0105	0.9410	0.9530	0.0009	0.0014
	Splines		0.6236	0.6153	0.0083	0.9510	0.9480	0.0009	0.0009

Table 7.4: Five year relative survival estimates, coverage and MSE after 1000 reps for eighty-year-olds for a lifetable approach and estimates from a spline approach. True value = 0.6149

Table 7.4 shows the information given by the four models. The spline estimates are similar to the lifetable estimates although they are consistently smaller estimating a lower relative survival rate than the lifetable method. Over a thousand simulations the difference has been reduced to an average of 3.16% which is lower than the initial ten observations shown in figure 7.5. There is a 1% difference between the lifetable estimate (higher) and the spline estimate (lower) when using p^*. This may suggest that the splines are becomming slightly better at estimating the relative survival in older patients than with lifetables.

The coverage is low for the worst case model with only eighty to eighty-two percent being given. The coverage is slightly higher in the spline model but the values are still lower than hoped for, but unsurprising given the bias in the estimated relative survival. The MSE is highest for the worst case scenario and lowest for the best case scenario suggesting that there is less variation when prevalence and odds ratio are small. The coverage is still fairly high in the best case scenario, but the other three show signs that the coverage is not high enough. The coverage and MSE for p^N are both good with the coverage being higher than the p^* estimate suggesting that p^* is potentially biased.

Ninety year olds

The final simulation estimated the relative survival for patients who started follow-up when they were ninety years old and could potentially age up to ninety-nine years old if they survived up to the end of the ten-year follow-up. Table 7.1 gives the expected probability of surviving for a year at 81.50% for males who are ninety with this number

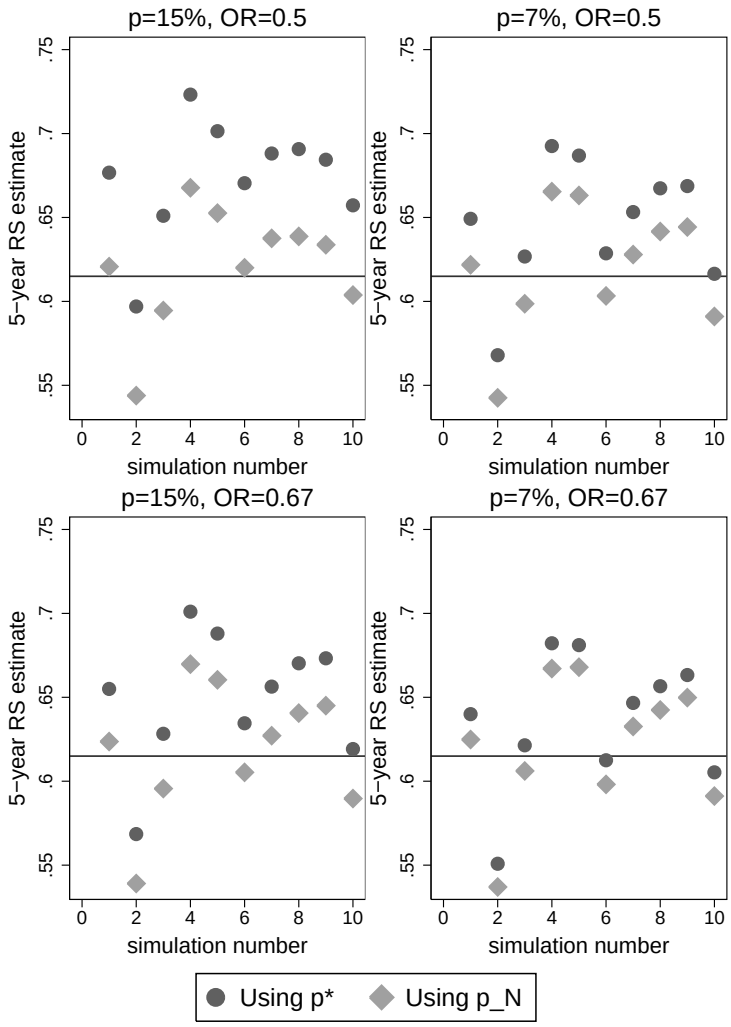


Figure 7.6: Plot of first ten simulations estimated 5-year relative survival using both p^* and p^N for 90-year old males

greatly reducing each year.

The first ten simulation estimates of 5-year relative survival are shown in figure 7.6 where once again the differences between p^* and p^N have increased. This time the worst case scenario has an average difference of 5.27% for these ten observations, which is the highest it has been. The higher prevalence appears to be a more important factor in this age group as the lower left-hand model shows a larger difference then the upper right-hand scenario. This is the first time that these two scenarios have been visibly different. The best case scenario (lower right) also shows a difference but is still the smallest of the four scenarios.

As for the other age groups it is easier to look at the results in a table as is shown in table 7.5. It is here that the performance of the worst case scenario is shown to be better when using spline based models over lifetables. The coverage is only 71% in a lifetable approach

Prevalence		OR	5-year RS		difference	Coverage		MSE	
			p^*	p^N		p^*	p^N	p^*	p^N
15%	LT	0.5	0.7050	0.6165	0.0885	0.7140	0.9520	0.0125	0.0033
	Splines		0.6659	0.6140	0.0519	0.7460	0.9430	0.0039	0.0013
7%	LT	0.5	0.6580	0.6167	0.0412	0.9010	0.9590	0.0019	0.0036
	Splines		0.6392	0.6138	0.0253	0.9280	0.9450	0.0019	0.0014
15%	LT	0.67	0.6647	0.6166	0.0480	0.8840	0.9570	0.0021	0.0036
	Splines		0.6431	0.6138	0.0293	0.9060	0.9470	0.0021	0.0014
7%	LT	0.67	0.6395	0.6170	0.0224	0.9370	0.9540	0.0016	0.0037
	Splines		0.6279	0.6139	0.0140	0.9470	0.9460	0.0016	0.0014

Table 7.5: Five year relative survival estimates, coverage and MSE after 1000 reps for ninety-year-olds for a lifetable approach and estimates from a spline approach. True value = 0.6149

and rises to almost 75% when using spline based models because of the reduced bias in the spline estimates. These values are still low and it is only the best case scenario that shows a coverage that is close to 95%. The difference between p^* and p^N is much higher in all of the lifetable estimates compared to the spline estimates.

The two middle scenarios are not very different when compared in this table but the high prevalence scenario does give a slightly more biased estimate than the high odds ratio of mortality approach as the difference value is higher. The worst case simulation gives an average difference of 5.19% which is lower than the inital ten estimates suggested, as plotted in figure 7.6. The coverage and MSE for p^N are both good.

7.4.4 Simulating excess hazard rate ratios

During the simulation two models were fitted, a piecewise approach using a Poisson distribution (115) on grouped data and a flexible parametric model using spline based models (3). Each model fitted a covariate which was uniform with an equal probability of being in one of two groups. The excess hazard rate ratio for the simulation was set at 3 and proportional excess hazards were assumed. Only the spline based models are shown in the figures for the excess hazard rate ratios as the GLM estimates proved to be slightly biased. This bias is important and is quantified in the tables of results.

The fitted model is given by

$$\begin{aligned} h(t) &= h^*(t) + \exp(\mu_j + \beta x) \quad \text{For the piecewise GLM} \\ H(t) &= H^*(t) + \exp\left[s(z;\gamma) + \beta x\right] \quad \text{For the spline based model} \end{aligned} \tag{7.17}$$

This applies to both models where $\exp(\beta x)$ is modelled in both approaches, see equations 4.2 and 5.25 for piecewise GLM and splines respectively.

Sixty year olds

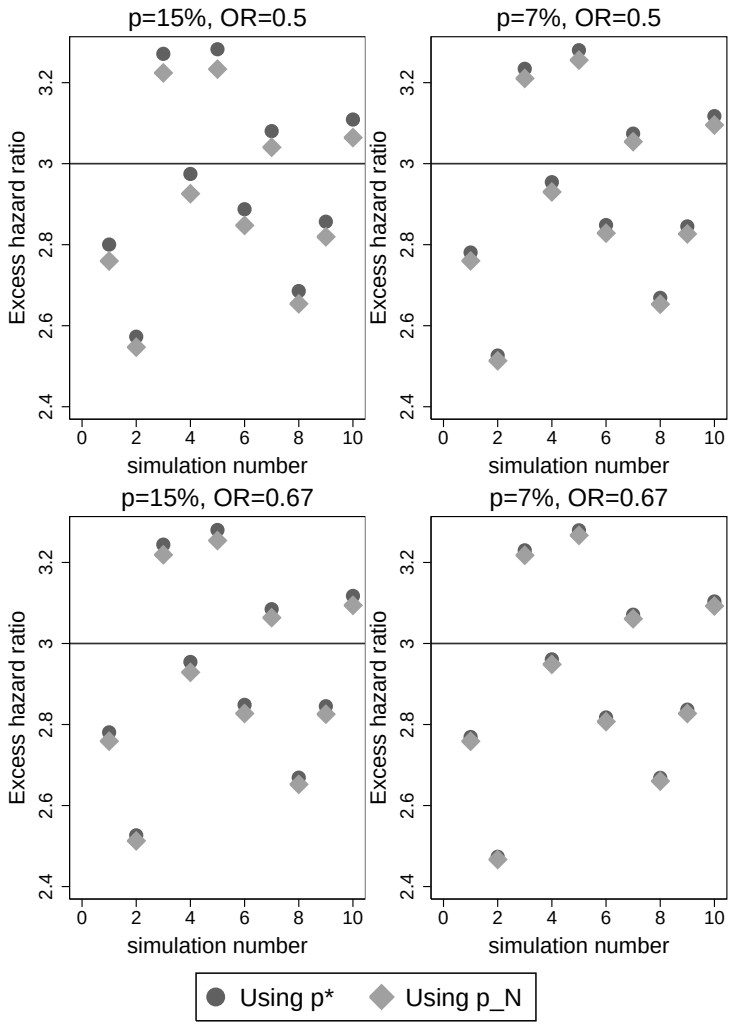


Figure 7.7: Plot of first ten simulations estimated excess hazard ratios using both p^* and p^N for 60-year old males

Figure 7.7 shows the estimates of the excess hazard rate ratio from the first ten simulations for 60-year old males from the spline based models. The figure is set up in the same way as the previous relative survival estimates with the worst case scenario (high prevalence and high odds ratio of mortality) in the top left cell and the best case scenario in the lower right cell.

There is very little difference between any of the estimates and it is clear that, in the worst case scenario, the difference between the two estimates vary from a small difference in observation 2 to the largest difference seen in observation 4. The other three scenarios show a very small difference as the estimates are very similar.

The coverage, difference and MSE were calculated for the excess hazard rate ratios and

Prevalence	OR	EHRR estimate		difference	Coverage		MSE	
		p^*	p^N		p^*	p^N	p^*	p^N
15% GLM	0.5	3.1772	3.1305	0.0466	0.9010	0.9230	0.1221	0.1021
Splines		3.0590	3.0170	0.0420	0.9440	0.9510	0.0820	0.0742
7% GLM	0.5	3.1539	3.1308	0.0231	0.9150	0.9170	0.1129	0.1035
Splines		3.0374	3.0166	0.0208	0.9460	0.9510	0.0786	0.0752
15% GLM	0.67	3.1551	3.1306	0.0245	0.9160	0.9190	0.1134	0.1034
Splines		3.0385	3.0165	0.0221	0.9460	0.9510	0.0788	0.0751
7% GLM	0.67	3.1433	3.1315	0.0118	0.9150	0.9210	0.1091	0.1044
Splines		3.0275	3.0169	0.0106	0.9500	0.9500	0.0774	0.0757

Table 7.6: Average values of the excess hazard rate ratio, coverage and MSE after 1000 reps for sixty-year-olds from a piecewise GLM approach and a spline approach. True value = 3

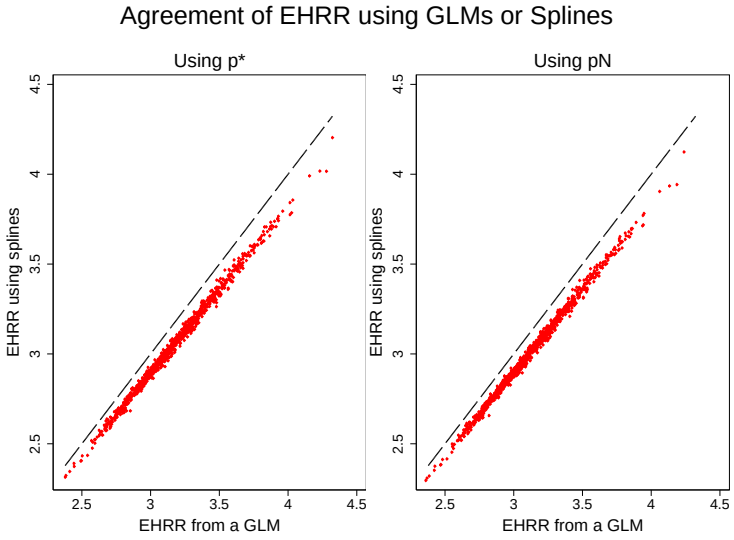


Figure 7.8: Plot of agreement for both p^* and p^N when using the two modelling approaches for patients aged 60 with 15% prevalence and an odds ratio of 0.5

are shown in table 7.6. In this table the estimates when using p^* and p^N from the piecewise model appear too high, and when the confidence intervals were assessed they did not include the known value of 3. This suggests that the piecewise models are biased and overestimate the ratio. The spline based models appear to estimate the value at approximately 3 and are thus unbiased.

Figure 7.8 shows the estimates of the excess hazard rate ratios from the piecewise GLM approach (x-axis) and the spline based approach (y-axis). On this plot is a line of perfect agreement (dashed) showing that these estimates do not agree for either p^* or p^N. In fact it appears that the GLMs always overestimate the excess hazard rate ratios.

This has been investigated further and it was found that if the split-time intervals were altered then the estimate would change and so by adding more in the early part of follow-up the piecewise model became unbiased. Another solution was to keep the split times

but alter the parameters (λ and γ) used in the Weibull simulation to make the hazard rate not decrease so rapidly in the first few weeks after hospitalisation, and by altering these the piecewise model could also be made unbiased (when γ was close to 1).

Age	Method	EHRR estimate		difference	Coverage		MSE	
		p^*	p^N		p^*	p^N	p^*	p^N
Changing γ to be one								
60	GLM	3.0323	3.0205	0.0118	0.9540	0.9590	0.0471	0.0458
	Splines	3.0321	3.0204	0.0118	0.9550	0.9550	0.0477	0.0464
70	GLM	3.1164	3.0299	0.0865	0.9460	0.9570	0.0832	0.0629
	Splines	3.1160	3.0294	0.0866	0.9460	0.9550	0.0835	0.0633
80	GLM	3.0573	3.0228	0.0345	0.9590	0.9610	0.0546	0.0495
	Splines	3.0570	3.0226	0.0344	0.9550	0.9630	0.0551	0.0500
90	GLM	3.2220	3.0338	0.1882	0.9190	0.9570	0.1709	0.0950
	Splines	3.2200	3.0318	0.1882	0.9180	0.9560	0.1705	0.0952
Changing split times								
60	GLM	3.0921	3.0486	0.0434	0.9360	0.9470	0.0907	0.0795
	Splines	3.0590	3.0170	0.0420	0.9440	0.9510	0.0820	0.0742
70	GLM	3.2114	3.0738	0.1377	0.9190	0.9460	0.1845	0.1205
	Splines	3.1636	3.0305	0.1330	0.9370	0.9460	0.1575	0.1088
80	GLM	3.1511	3.0606	0.0904	0.9180	0.9340	0.1279	0.0965
	Splines	3.1115	3.0236	0.0879	0.9350	0.9460	0.1119	0.0885
90	GLM	3.2484	3.0794	0.1691	0.9140	0.9430	0.2493	0.1543
	Splines	3.1883	3.0320	0.1563	0.9360	0.9550	0.2054	0.1365

Table 7.7: Average values of the excess hazard rate ratio, coverage and MSE after 1000 replications with a prevalence of 15% and odds of 0.5 Where γ is changed to 1 or the number of split times are increased in the first six months. True value = 3

The estimates for the excess hazard rate ratios are shown in table 7.7 for when γ and the split times were altered. The results from the simulated worst case scenario (high prevalence and odds) for the four starting ages are shown. The GLM results fall in line with the spline results and become unbiased (estimating p^* and p^N at close to 3). Changing the split times did not work as well as changing the value of γ as p^* is still consistently higher than the spline estimate, but p^N is unbiased. It is probable that by further adjustments to the split times that a much better fit would eventually be found. This is not an ideal situation and gives strength to the use of the splines as they are unbiased by the survival time distribution that is used in the simulation and does not use split time data.

Figure 7.9 shows the agreement between the GLM and spline estimates after the adjustments to either γ or the split times are made. The spline model advantages do not stop there as the coverage is also higher in all scenarios. In comparison to the relative survival estimates of the MSE the excess hazard rate ratios MSE initially appear to be fairly high, and whilst they are larger the values are still fairly small. The largest spline difference is in the worst case scenario, as expected, and is only 0.04 so is very small. The coverage and MSE for p^N both show similar results to p^*, which suggests that even p^N is biased by the use of GLMs.

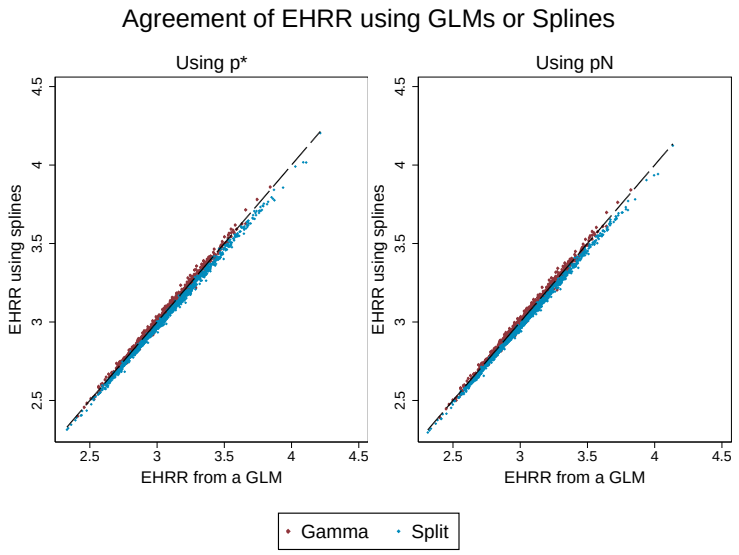


Figure 7.9: Plot of agreement for both p^* and p^N when using the four modelling approaches, two GLMs and two splines, one changing γ and the other changing split times. Only patients aged 60 with 15% prevalence and an odds ratio of 0.5 are shown

Seventy year olds

Increasing the starting age by ten years to seventy has the effect of increasing the difference between the estimates using p^* and p^N, shown in figure 7.10, for this reason the scale is altered from the one shown for sixty year olds. The best case scenario shows a very small difference and the two middle scenarios appear to give very similar differences between p^* and p^N to each other. The worst case scenario, is as expected the scenario that shows the largest differences.

Prevalence	OR	EHRR estimate		difference	Coverage		MSE	
		p^*	p^N		p^*	p^N	p^*	p^N
15% GLM	0.5	3.2515	3.1538	0.0978	0.8830	0.9110	0.1812	0.1274
Splines		3.1115	3.0236	0.0879	0.9350	0.9460	0.1119	0.0885
7% GLM	0.5	3.2042	3.1558	0.0484	0.8960	0.9120	0.1531	0.1287
Splines		3.0672	3.0239	0.0433	0.9390	0.9410	0.0983	0.0888
15% GLM	0.67	3.2076	3.1557	0.0519	0.8950	0.9120	0.1548	0.1285
Splines		3.0704	3.0239	0.0465	0.9390	0.9400	0.0991	0.0887
7% GLM	0.67	3.1806	3.1556	0.0250	0.9040	0.9070	0.1414	0.1295
Splines		3.0456	3.0232	0.0223	0.9420	0.9410	0.0938	0.0895

Table 7.8: Average values of the excess hazard rate ratio, coverage and MSE after 1000 reps for seventy-year-olds from a piecewise GLM approach and a spline approach. True value = 3

Table 7.8 shows the statistics for the seventy year old simulation for the estimates of the excess hazard rate ratio. Once again the piecewise models are biased, but now slightly more so. The coverage is high for the spline based models, higher than the piecewise coverage, and is around 95% which is good. The MSE values are increasing slightly but

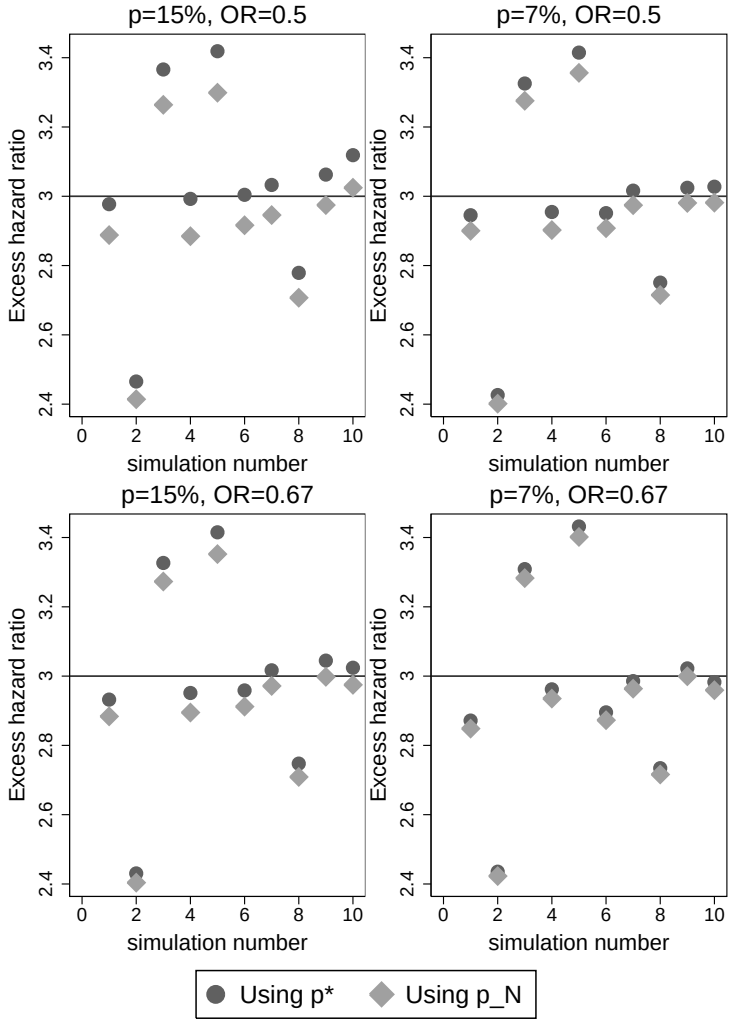


Figure 7.10: Plot of first ten simulations estimated excess hazard ratios using both p^* and p^N for 70 year-old males

are still fairly low. The estimates when using p^N are also biased along with those for p^* for the piecewise models.

The biggest difference from the spline model is 0.09 which is still fairly small and indicates that prevalence and odds ratio of mortality are not having the same affect on the relative effect (excess hazard ratio) as they were on the absolute effect (five year relative survival). This is good news, but needs to be assessed in older age groups as the absolute effects became worse in the two older age groups assessed. The coverage and MSE for p^N are both fine for the spline based models when using p^N.

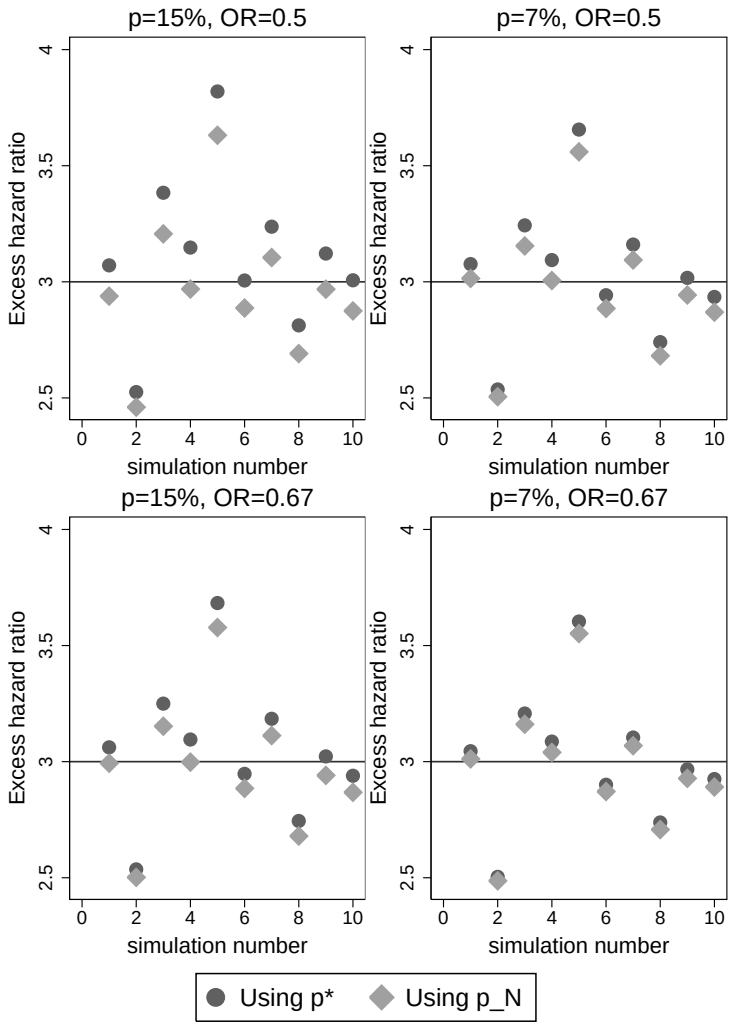


Figure 7.11: Plot of first ten simulations estimated excess hazard ratios using both p^* and p^N for 80 year-old males

Eighty year olds

Figure 7.11 shows the first ten estimates of the excess hazard rate ratios using p^* and p^N for the four variations of prevalence and odds ratio of mortality. The best case scenario (bottom-right plot) shows a very small difference and the estimates appear unaffected in cases with a lower prevalence and odds ratio of mortality. This is different to the relative survival estimates which started to show differences even in the best case scenario. The worst case scenario shows larger differences and observation 2 shows a very similar estimate with little difference.

As for the two age groups assessed earlier the eighty year old piecewise model is biased in its estimation of the excess hazard ratios, as shown in table 7.9. The worst case scenario shows a difference of 0.13 for the spline model, the piecewise estimates are still biased,

Prevalence	OR	EHRR estimate		difference	Coverage		MSE	
		p^*	p^N		p^*	p^N	p^*	p^N
15% GLM	0.5	3.3346	3.1835	0.1510	0.8510	0.9210	0.2718	0.1644
Splines		3.1636	3.0305	0.1330	0.9370	0.9460	0.1575	0.1088
7% GLM	0.5	3.2587	3.1840	0.0747	0.8950	0.9280	0.2151	0.1680
Splines		3.0952	3.0295	0.0657	0.9440	0.9470	0.1301	0.1109
15% GLM	0.67	3.2656	3.1834	0.0822	0.8900	0.9270	0.2194	0.1671
Splines		3.1016	3.0292	0.0723	0.9450	0.9450	0.1321	0.1105
7% GLM	0.67	3.2253	3.1858	0.0395	0.9110	0.9270	0.1937	0.1702
Splines		3.0650	3.0302	0.0348	0.9460	0.9410	0.1208	0.1118

Table 7.9: Average values of the excess hazard rate ratio, coverage and MSE after 1000 reps for eighty-year-olds from a piecewise GLM approach and a spline approach. True value = 3

and this value is still small. The coverage for the spline based models remains consistently better than the piecewise models, and fairly consistent with itself staying at around 95%.

The MSE values are also much lower in the spline based models than in the piecewise models. The two middle scenarios show that high prevalence appears to be slightly worse than high odds ratio of mortality as the difference is slightly higher. The worst case scenario has the largest difference which is twice the difference of the two middle scenarios, which are themselves twice the best case difference. It is interesting to see the effect that the values of prevalence and odds ratio of mortality values have affected the difference in estimates when using p^*, as in practice, and p^N which would be used in preference. The coverage for p^N still show that the estimate is unbiased for the spline based models. The coverage is fairly good in the GLM models but not as high as hoped for.

Ninety year olds

The final figure shown here (figure 7.12) is for the final group being assessed, males aged ninety at the start of follow-up. These patients have a high expected mortality rate (p^*) and this value is also high when deaths excluding MI are removed. The best case scenario is still unbiased as there is no clear difference between the two estimates. The results are on the same scale as the eighty year olds which shows how the estimates have not become too different, unlike in the absolute estimates of relative survival shown in figure 7.6. The differences are largest in the worst case scenario with many of these first ten simulations showing differences that are similar in size to the two middle scenarios.

Table 7.10 shows the relevant statistics from the 1000 simulations for males who were ninety years old at the start of follow-up. This was found to be the worst age group when assessing the absolute effect, and has been found to be the same for the relative estimates here. The main value to assess is the difference in the estimate between p^* that would be used in practice and p^N that would be used if it were available. The largest difference is in the worst case scenario where the prevalence is high and the odds ratio of mortality

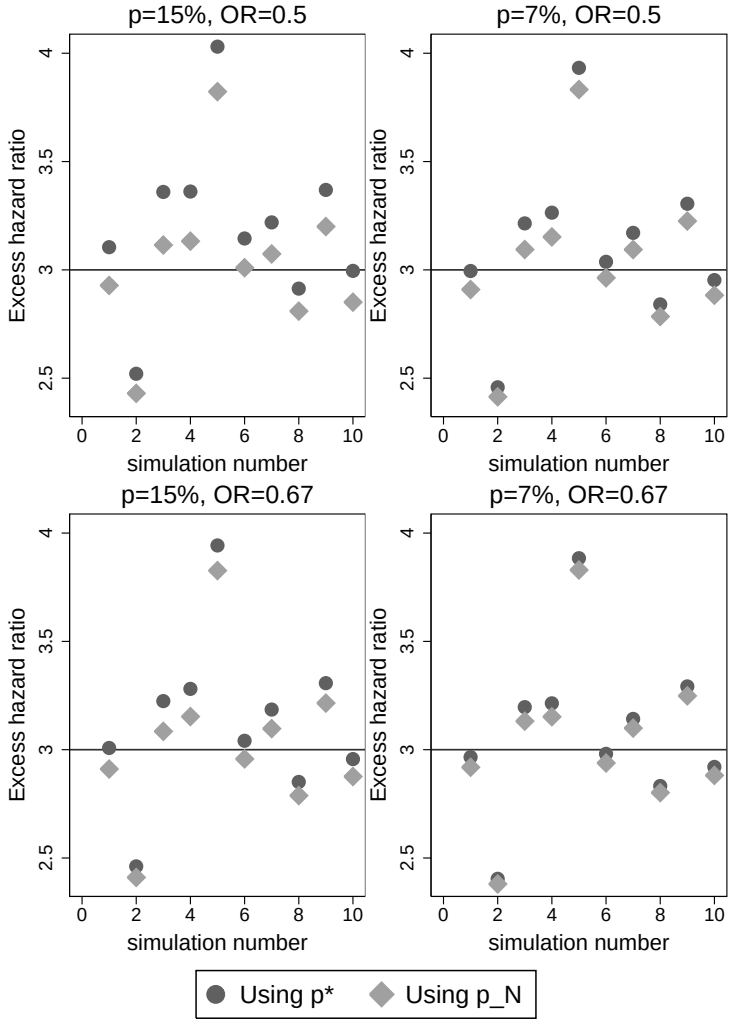


Figure 7.12: Plot of first ten simulations estimated excess hazard ratios using both p^* and p^N for 90 year-old males

Prevalence	OR	EHRR estimate		difference	Coverage		MSE	
		p^*	p^N		p^*	p^N	p^*	p^N
15% GLM	0.5	3.3983	3.2085	0.1898	0.8430	0.9120	0.3788	0.2155
Splines		3.1883	3.0320	0.1563	0.9360	0.9550	0.2054	0.1365
7% GLM	0.5	3.3014	3.2094	0.0920	0.8890	0.9120	0.2895	0.2197
Splines		3.1090	3.0320	0.0770	0.9510	0.9580	0.1656	0.1385
15% GLM	0.67	3.3154	3.2094	0.1060	0.8820	0.9130	0.3010	0.2192
Splines		3.1203	3.0318	0.0885	0.9480	0.9560	0.1703	0.1381
7% GLM	0.67	3.2610	3.2105	0.0505	0.9030	0.9120	0.2565	0.2206
Splines		3.0748	3.0324	0.0424	0.9550	0.9600	0.1524	0.1391

Table 7.10: Average values of the excess hazard rate ratio, coverage and MSE after 1000 reps for ninety-year-olds from a piecewise GLM approach and a spline approach. True value = 3

are high. The piecewise model estimate is biased and has lower coverage than the splines. The spline gives a difference of 0.1563, which is the highest observed so far. This value is

still relatively small and the use of p^* appears to cause the estimate to be overestimated, i.e. the estimate is 0.16 higher than if p^N had been used. This is still a small bias and does not have a large impact on the estimate.

The coverage for the spline based models are all good at around 95% with the MSE estimates being the largest observed. This is to be expected as there is much larger uncertainty in the estimates when the expected mortality is so high, as it would be in males over ninety years old. In the best case scenario the difference is still very low at 0.04. The coverage for p^N is fine for the spline based models but are not quite as good for the GLM models.

Using all of the simulation scenarios it has been found that the relative survival estimates can become slightly biased in cases of high prevalence and high odds ratio of mortality when a patient is aged over eighty years old. However the excess hazard rate ratios have been found to have only a small amount of bias, even in the very elderly worst case scenario simulation.

7.5 Discussion

The primary finding was that even in the worst cases where the prevalence and the odds ratio of mortality are at their highest and the oldest age group used the excess hazard rate ratios only show a small level of bias. This is a very promising result and shows that for relative effects the impact of prevalence is fairly small.

The absolute effect of relative survival shows that under the worst case scenario the estimates became biased by 5% to 9% (ninety year olds) depending on whether the lifetable or spline approach was used in the calculation of the estimate. This is a fairly large value in the context of the underlying value, which is around 60% suggesting an overestimte of five-year relative survival up to 69%. This value is not too high in sixty and seventy year old patients ($< 1.7\%$), whereas eighty year olds have a bias of between 3 and 4.2, which is not too bad.

There is less bias in the spline based models than in the GLM piecewise approaches, as the spline based models were not affected by the parameters of the Weibull distribution used for the simulation and do not use split-time data. This is important and shows that there can be problems in the traditional GLM methods and more thought should be given to the choice of split-times.

It should be noted that the conclusions drawn here relate to the worst case scenario where the prevalence of heart disease in the general population is 15% with these patients at twice the risk of mortality than all other people in the general population without heart

disease. These values may be too high in practice as the prevalence values were obtained for patients over 75 and these ranged between 7% and 15%. Prevalence also seemed to have a bigger impact on the bias than the odds ratio of mortality although this was a small difference that was observed between the two middle scenarios. Prevalence will also be related to age.

As stated before is section 7.4.2, by excluding MI the model is in fact over correcting as p^N assumes that these non-MI patients are never going to have an MI. This is similar to a matched cohort study where an MI patient would be matched to a subject without history of MI where this subject may not develop an MI in the future.

The odds ratios used in this simulation were 1.5 and 2. However after discussion with a clinician it was concluded that these values are feasible. The risk of death after MI decreases exponentially with time from the event. It is at its highest in the first 72 hours and falls rapidly thereafter. The general population will be formed with the majority of patients in the chronic phase of the disease, where the degree of increased risk is likely to be at its lowest. Further to this, the risk varies enormously, and is influenced strongly by a number of factors, in particular the extent of residual left ventricular failure. The latter would double risk of death, even in the long-term, other factors would increase risk but not by as much (161). Therefore, if left ventricular failure doubles the risk of death in the long term, and this is the worst factor, then the use of a doubled risk of death from MI is high.

The level of the odds ratio is in effect assuming that all patients who have had an MI in the population have this condition, which is unlikely, and if anything, is possibly too high for the simulation in practice. Thus, although the assumption of the odds ratio is unrealistically simplistic, it does demonstrate how a high prevalence with a big increase in the odds of mortality can introduce bias in absolute effects.

7.5.1 Further assumptions

The control group, the general population, are assumed to be free of the disease at entry to the study. However, it does not follow that these people will not then develop the disease and die from it, but this has been assumed for the simulation.

During this simulation is has been assumed that the increase in odds ratio of mortality for MI patients is constant over time at either a 50% or a 100% increase when compared to the general population. However in reality it would be fair to expect this value to reduce over time, for example if a patient survives for five years after their initial MI then they may have a survival probability that is more similar to the general population. It also highlights how the increased rate is unlikely to be sufficient for patients immediately after their MI as this rate is much higher than twice that of the general population for the first

month at least. One argument for this however is that this time period of huge excess mortality is thankfully short and the relative survival estimate is almost identical to the all-cause estimate and as such is unlikely to be problematic.

7.5.2 Further work

If the value of (θ) is known, i.e. the prevalence, and the value of the odds ratio are known then it is possible to correct the expected probabilities to ensure that they are disease free. However these are unlikely to be known in practice, but the work described here enables sensitivity analyses to be performed to further assess the impact of changes in the prevalence and odds ratios.

It would potentially be possible to estimate θ and the odds ratios from the literature and then feed in the uncertainty of these estimates in to the model, which would probably require a Bayesian analysis (162; 163) in order to take all relevant uncertainty into account.

Chapter 8

Up-to-date Estimates of Survival After an MI Using Period Analysis

8.1 Introduction

Long term survival rates are frequently used outcome measures in heart disease as they provide a way of monitoring progress of heart disease care over time. Patients are now able to find survival statistics that relate to their condition over the internet (164) and this knowledge helps both the clinician with their management of the disease and the patient in their ability to cope with their condition. However traditional long-term survival rates are in effect estimating the survival expectations of patients diagnosed many years ago. These estimates are often severely outdated at the time they became available as they fail to take ongoing improvements into account. With the improvement in short-term survival attention is likely to shift to monitoring long-term survival.

Period analysis was first proposed by Hermann Brenner and Olaf Gefeller in 1996 under the name *period monitoring* (165) which was later renamed period analysis in 1997 by the same authors (166). Period analysis is a method that was proposed in order to obtain more up-to-date survival estimates and is a method that has been transferred from the previously well-established use of period life tables. These were used in demography to estimate current life expectancy, where period life tables are constructed for synthetic cohorts of individuals on the basis of the most recent available mortality data as was described by Brenner and Gefeller (165).

A review of methods by Brenner *et al.* in 2004 (167) found that it was not a method that had been greatly adopted by the scientific community. In the same year a paper by Smith *et al.* (168) highlighted how easy it is to model period analysis data using delayed entry models in common statistical packages, which now includes Stata. In this chapter

the flexible parametric spline based models defined in chapter 5 are used to fit delayed entry models alongside more traditional life table period estimates of relative survival.

Up-to-date measures are needed in order to communicate with patients that are hospitalised today about their prospective experience. It is also very useful when planning services. These methods have been found to be useful in providing more up-to-date estimates of long-term survival of cancer patients in Europe and the United States (169; 170; 171; 172; 173; 174; 168). The methods are also proposed for use in all-cause and relative survival and results have found that long-term survival rates achieved by the end of the 20^{th} century are much higher than previous traditional analyses had suggested. This evidence was presented by Brenner *et al.* (167).

Period analysis, like any method that attempts to predict survival for patients diagnosed today, must make some assumptions. Empirical investigations using retrospective data has shown period analysis to be useful in cancer and a transferral of the methodology to heart disease warrants investigation.

8.2 Period analysis explained

Standard survival, or relative survival, analyses include all available information on the survival experience of patients diagnosed with a disease of interest. So for these data that is patients who experienced a myocardial infarction with ST-elevation in a certain period. Period analysis is based on a smaller subset of these patients looking at the survival experience in a more recent time interval. For some patients only part of their survival experience will be included. The standard approach will follow all patients from hospitalisation until death, or censoring at a specific date. Therefore patients who experienced events at the beginning of the study have the most influence over the long-term survival as they are followed up for longer. The proposed analysis considers the survival experience of patients during some recent time period, excluding the short term survival of patients recruited at the start of the study.

Period analysis defines a window of observation of patients experience by truncating at the beginning of their follow-up and censoring at the end of the chosen time-frame. This excludes short term survival that exists before the start of the period under study. The short term survival of those diagnosed a long time ago is likely to be worse. This is best explained by assessing figure 8.1. Here six potential patients from the LRI CCU dataset are given to highlight the information utilised by period analysis. The patients are numbered on the left-hand side of the figure and the the two vertical dashed lines indicate the period which is being adopted here (2002-2006).

Patients have been followed up since 1993 with the last update on the 22^{nd} of March in

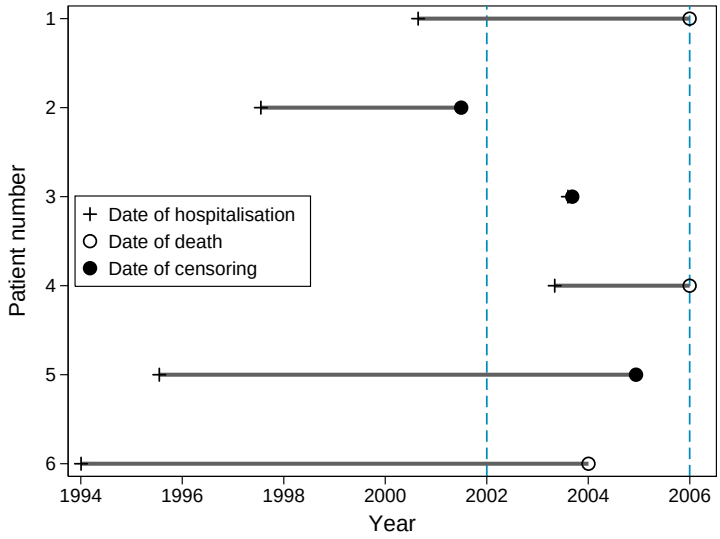


Figure 8.1: Analysis by year showing relative survival and excess hazard rates

2006 making this the censoring date. For this period analysis a ten year follow-up will be used to incorporate more information than the previous six year experience used in previous chapters. This is because period analysis can be a very useful tool for prediction over longer terms. Under a standard survival analysis all six patients given as an example in figure 8.1 are included in the analysis. However under a period analysis only five patients' information will be used with just two patients providing their full experience to the analysis.

Patient one was hospitalised towards the end of the year 2000 and survived up to the censoring date. For this patient the experience from 2002 up to 2006 is analysed meaning that this patient's short term survival, experienced prior to 2002, of just over a year is not included in the analysis but their longer term survival of around one year up to five years is. Patient two is very simple, as they experienced their MI in 1997 before dying in 2001 this patient is excluded from the analysis completely. Patients three and four have their full information included with patient four providing up to three years survival information and patient three their very short survival experience. Patient five was hospitalised in July 1995 and died in October 2004, so this patient contributes their long term survival estimates experienced (approximately 6-9 years) after their initial MI. The final patient was followed up for ten years, but having been hospitalised in 1994 this patient only contributes their 8-10 year survival experience at which point they become censored.

It is important to note that the removal of information outside of the period under investigation leads inevitably to a loss in precision, however the benefits are in the more up-to-date estimates of survival as patients diagnosed prior to the period are not contributing to the short-term survival. Period analysis therefore excludes information that

is not a true reflection of current health management strategies, such as treatment, surgical and clinical advances. A patient diagnosed in 2006 has a very different expected short-term survival compared to a similar patient (demographically) diagnosed in 1993 as several treatments, i.e. statins and thrombolysis, and disease managements, i.e. greater awareness of diabetes in the public space, have been discovered or implemented over this relatively short time interval. This suggests that coronary heart disease may too benefit from more up-to-date estimates of patient survival using period analysis.

In theory, period estimates of long-term patient survival may become overly optimistic. This is where advancements in early detection of therapy do not increase the chance of cure, but merely postpone the death of a patient, i.e. a short term increase in survival does not lead to an increase in long-term survival. This theoretical possibility has been assessed in practice where it appeared to have little impact, as described in detail by Brenner *et al.* (167). One area where this was assessed was in the retrospective analyses that have been performed. This is where period analysis has been used retrospectively in an old cohort to predict future survival which is then compared to the observed survival for the patients originally assessed.

Period estimates that have been derived are found to quite closely predict long-term survival rates observed many years later for patients diagnosed in the period of interest, thereby enabling early detection of recent trends (175; 176; 177). For example Brenner and Hakulinen (175) performed an empirical evaluation of period analysis' ability to detect time trends in the long term by deriving 10-, 15- and 20-year survival rates that might have been obtained using a traditional cohort analysis and by period analysis in various time intervals between 1963 and 1997 and the differences were compared. Here it was found that the detection of time trends in 10-year survival can be advanced by 5-10 years using period analysis compared to a cohort analysis. Traditional cohort analysis is simplistic as it excludes recent data as only patients that have the potential to be followed-up for the entire study period are included in the model.

Brenner and Spix (178) proposed a mixture of period and cohort analysis, where the value of a mixed analysis increases with increasing length of patient follow-up. There is not enough information available in the LRI dataset as complete data is only available from 1993. For this reason a retrospective analysis to assess period analysis' ability to predict time trends in heart disease has not been investigated here.

8.3 Period analysis methods

There are two approaches to standard period analysis, the first and perhaps most common, is a lifetable approach which were originally proposed by Brenner *et al.* (166). These are calculated using the same approach as standard lifetables. The difference is in the

definition of those at risk, in period analysis a patient is at risk during the interval being assessed, meaning that the number at risk and the number of deaths in each time interval need to take the period of interest into account. This can be achieved by direct calculation or, more commonly, by performing two separate, standard, analyses and calculating the difference. However the first run through is as normal with all information included up to the end of the period (i.e. 22/03/2006) and the second run-through censors at the beginning of the period (i.e. 01/01/2002) as described by Brenner and Gefeller (166).

The second approach is delayed entry models, where patients do not contribute to the model until the start of the period of interest. This means that patients are not followed up from time zero but rather from the start of the period of interest. So using the LRI CCU dataset the first five observations using the delayed entry model will have the following results shown in table 8.1.

Patient number	Patient Died	Date of hospitalisation	Date of death	End date	Start time		End time
					St	Pe	
1	No	19Aug1998	n/a	22Mar2006	0	3.3703	7.5893
2	Yes	21Sep2001	04May2005	04May2005	0	0.2793	3.6167
3	No	20Jul1996	n/a	22Mar2006	0	5.4511	9.6701
4	No	02Jul2002	n/a	22Mar2006	0	0	3.7207
5	Yes	05Apr1997	31Dec2004	31Dec2004	0	4.7420	7.7399

Table 8.1: Five patients data in a delayed entry model where St=Standard, Pe=Period.

Patient four was hospitalised after the start of the period and spent 3.7 years in follow-up before being censored. The four other patients all began follow-up before the period of interest and because of this their start time for the delayed entry model is greater than zero. For the patients that died in the interval the end date is earlier as would be expected when compared with those who survived until the end of follow-up. Delayed entry models are easy to achieve in Stata by simply using a start time when setting the data as survival time data with `stset`, this way all Stata's `st` commands allow delayed entry models. This is true for the spline based model developed in chapter 5, which is implemented in Stata as shown in chapter 6. Also in this table are the start times from a standard approach, here it is clearly shown the start times for all patients is zero and that the only difference between a standard analysis and a delayed entry model is the start time.

The likelihood for a delayed entry model can be written as

$$\ln L_i = d_i \log\left[h(t_i)\right] + \log\left[S(t_i)\right] - \log\left[S(t_{0i})\right] \tag{8.1}$$

This differs from the previous likelihood in 5.5, with the addition of the $\log\left[S(t_{0i}\right])$ term, where t_{0i} is the time at which the subject becomes at risk. Therefore the calculations need to be derived using both t_0 and t, which are the time at entry and the event time respec-

tively. For the spline based models that are fitted here this means that the spline derived variables need to incorporate both times. To do this new spline terms are calculated using the same knots using t_0 alongside those already calculated at the event time (t), as the likelihood incorporates both.

8.4 Period analysis results

8.4.1 Initial investigation

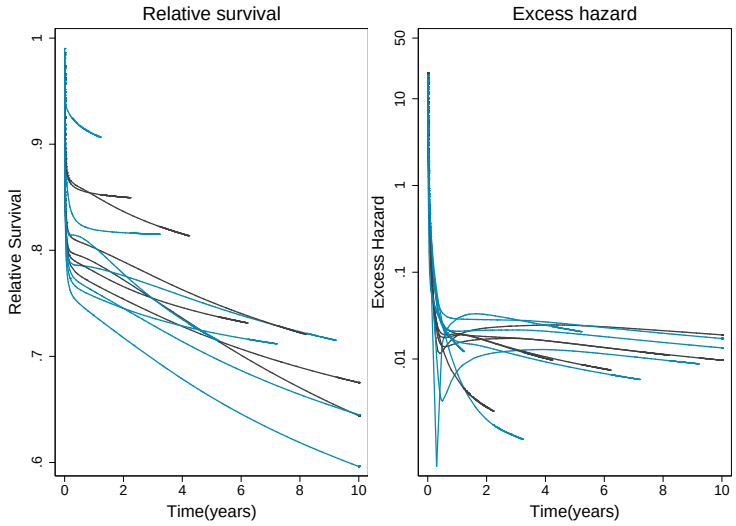


Figure 8.2: Analysis by year showing relative survival and excess hazard rates

In order to determine whether period analysis is appropriate it is a good idea to plot the estimates of the baseline results for each year by individually modelling the years as time dependent effects. However for these data year 2006 would not converge using the various maximisation techniques available as there is only follow-up until 22/MAR/2006.

Figure 8.2 shows the results from the univariate year analysis using the spline based models with only two internal knots. On the right-hand side the excess hazard rate is shown for each model where the individual year was the only covariate. The excess hazards are difficult to differentiate from one another but there appears to be a tendency for the shorter lines, i.e. the more recent years, to have a lower excess hazard than the longer lines. It is far more apparent in the relative survival plot on the left-hand side of the figure as it is easier to see the total shape of the relative survival curve. Using this the four longest lines are all shown to have the lowest relative survival and the four shortest lines are shown to have the best relative survival. There are also large differences in the long term relative survival estimates. However the excess hazard rates appear similar in long-term follow-up, which is slightly distorted by the log-scale.

Due to these results period analysis does appear to have potential use in these data. Short term survival of patients has clearly improved in the last four years that are modelled here when compared to estimates from up to ten years previously. Therefore the use of period analysis has a potentially important impact on the predicted relative survival for patients hospitalised today when compared to estimates that would be obtained from a standard relative survival analysis which gives more power to those with a longer follow-up.

8.4.2 Assessing age groups

Age groups were used again as there are several issues with the investigation of treatments which may initially appear to be an interesting option. Follow-up of ten years is required to analyse treatments and this is not possible as most treatments used today have not been around for ten years or have not been used routinely for ten years. There are many factors that can contribute to changes over time and by using age groups the aim is to obtain a measure of all the changes. It is difficult to differentiate the treatment effects from each other and from clinical implications and so it is better to give an overall measure to quantify the general improvement in survival that a patient today can experience. There may be improvements over time that vary between different age groups which is an indirect measure of the improvements made in caring for ST elevated AMI patients.

Delayed entry models used here with the relative survival spline based models hold the advantage over a lifetable period analysis approach in their ability to obtain smooth estimates by assuming constant effects. For example, if proportional excess hazards (PEH) is a reasonable assumption then it would be expected that more precise estimates are obtainable by modelling than by a lifetable approach. This is because the lifetable approach performs a separate analysis for each sub-group.

Under the assumption of PEH a spline model with five internal knots was fitted to these data using a relative survival approach and a delayed entry period analysis relative survival approach. By contrast the standard approach uses all of the information given by 4748 patients, whereas the period analysis used full and partial information from only 3474 patients. Figure 8.3 shows the estimated relative survival (left-hand side) and excess hazard rates (right-hand side) split by age groups as this enables a comparison of period and the non-period *standard* relative survival approaches.

The time windows are different for the two approaches as all patients are included in the standard approach for up to ten years of follow-up, whereas the period analysis only includes patient survival experience within $2002 - 2006$ up to ten years, but does include patients diagnosed prior to 2002 if they were alive during the time window. The standard relative survival estimates are all lower than the respective period analysis estimates and the excess hazard rates cross after a number of years.

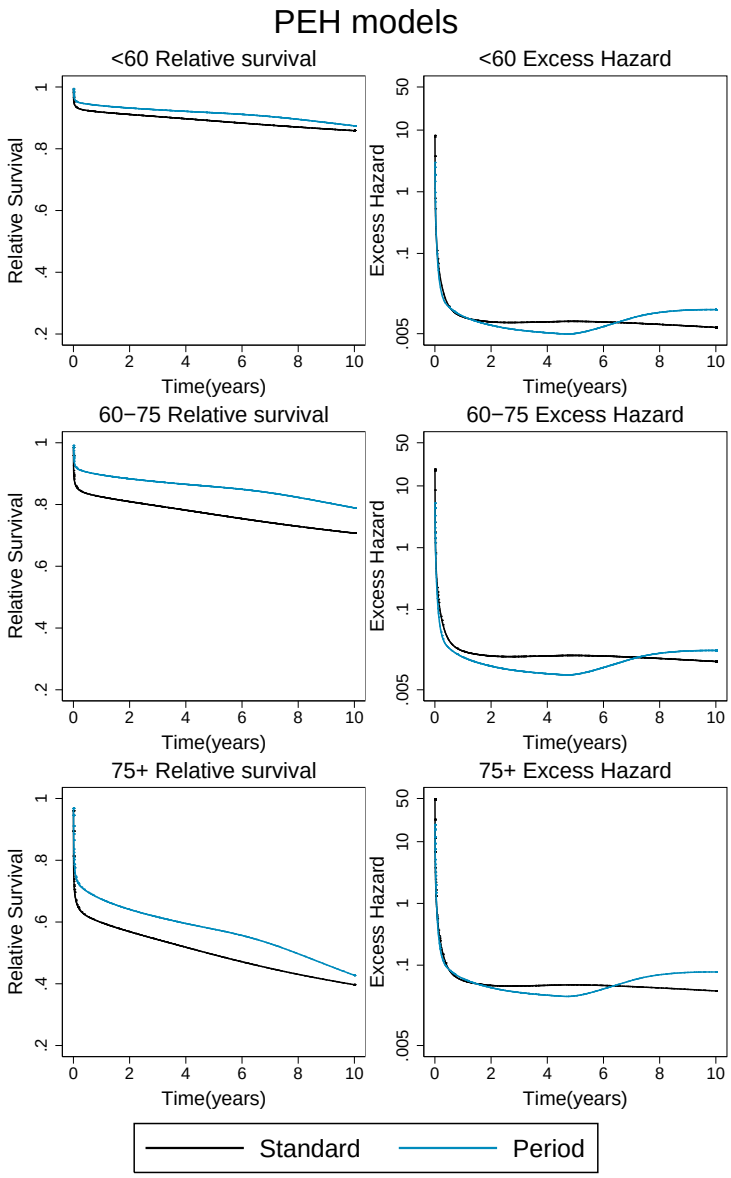


Figure 8.3: Assessment of Age groups using Period analysis under the assumption of proportional excess hazards

There is a difference between period estimates of relative survival and standard estimates for each age group. The middle age group appears to have a much larger difference than the youngest age group and slightly larger than the oldest age group. The youngest age group shows very similar estimates from a period analysis and standard analysis. This suggests that the improvements in survival for patients under sixty years old are much smaller than the improvements in older age groups. The excess mortality rates for the three age groups are all quite similar in shape, as expected under a PEH assumption. It is difficult to read on the log scale but the difference in excess mortality at the start of follow-up (first few days) is quite large suggesting a better mortality rate immediately following an ST elevated AMI.

It is important to note that at the end of the follow-up period the excess mortality rate estimates from the standard and period analysis are becoming very similar to each other. This is expected as the impact of period analysis is in the short term estimates and the right hand side of the plot will be very similar to each other as patients surviving 8-10 years will be in both datasets.

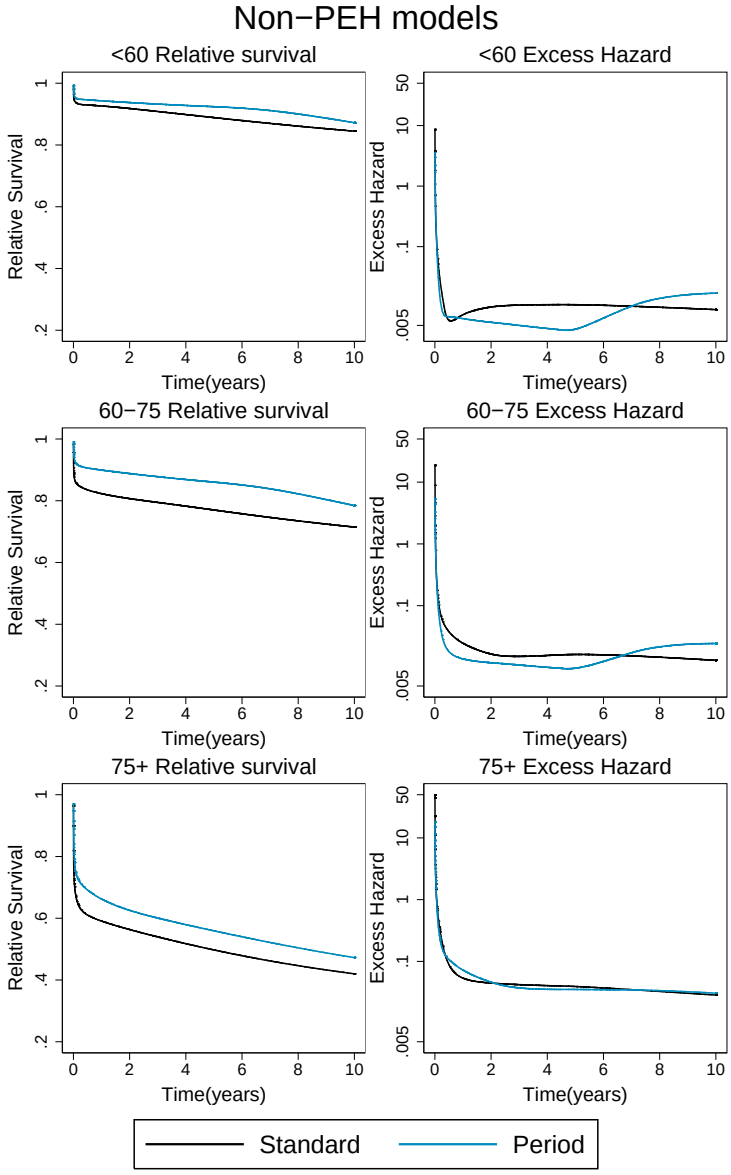


Figure 8.4: Assessment of Age groups using Period analysis fitting time dependent effects for age

It is also possible when using delayed entry models to obtain estimates of time dependent effects under a non-PEH assumption. This is achieved in period analysis the same way as in standard analysis by allowing the age groups to vary by spline terms as shown in chapter 5. The results from this are shown in figure 8.4 with period and standard relative survival models fitted with estimates of the relative survival and excess hazard rate shown.

A likelihood ratio test gave a value of χ^2_{12} $p = 0.2450$ and χ^2_{12} $p = 0.9230$ for the standard relative survival and period survival analysis respectively.

Excess hazard rate ratios are not shown as it is generally considered that the benefits of period analysis is in the estimation of absolute effects. Using all of the data to estimate relative effects is generally accepted if it is reasonable to assume that the effect has stayed constant over time. Naturally when this is the case the estimates will be the same but the precision will be lost in a period analysis so a standard model with all of the data would be better.

The first thing to note from figure 8.4 is that there is very little difference between the PEH (figure 8.3) and non-PEH models. The relative survival estimates are almost identical to the PEH models and all show the same relationship as previously observed. The excess hazard rates for the youngest age group are slightly different as the period estimates give a lower excess mortality rate than the standard models under a non-PEH assumption. The standard model shows some *dipping* as discussed in section 5.5.2, whereas the period estimate does not.

The middle age group estimates a sharper drop in excess mortality than in the PEH model and still crosses the standard estimate after approximately seven years, as do the youngest age group estimates. The oldest age group estimates are very similar suggesting no improvement in long term mortality rates for patients aged seventy-five years or older. The initial excess mortality is once again lower in the period analysis suggesting an improved short term mortality rate in all age groups.

8.5 Sensitivity analysis

Table 8.2 shows the estimates from a sensitivity analysis where the start of the period was changed from 2002 to 2003 and then 2004. Also shown here are the lifetable estimates as a comparison between the standard approach proposed by Brenner and Gefeller (166). The table highlights that the period analysis confidence intervals are wider than those obtained from a standard survival analysis. This is due to the decrease in data and the inevitable reduction in precision that is caused by this. An important part of period analysis is understanding the balance. As the window becomes narrower it has the potential to be more up-to-date but the estimates are likely to increase due to the discarding of information. The choice of window width is a balance between bias and precision, as long as the assumptions are reasonable.

In this table are the estimates from a standard relative survival model using spline based models and all of the models and lifetables including age groups. Five and ten year relative survival is shown as this is a standard estimate that is generally given when using period

Period Start	Method	Rel surv estimate	Age group		
			<60	60-75	75+
	Standard	5 year	0.890 (0.87, 0.91)	0.768 (0.74, 0.79)	0.494 (0.46, 0.53)
All	Splines	10 year	0.859 (0.83, 0.88)	0.707 (0.67, 0.74)	0.397 (0.35, 0.44)
data	Non-PEH	5 year	0.889 (0.87, 0.91)	0.770 (0.75, 0.79)	0.497 (0.46, 0.53)
	Splines	10 year	0.845 (0.81, 0.87)	0.715 (0.67, 0.75)	0.420 (0.35, 0.49)
	Life	5 year	0.919 (0.89, 0.94)	0.863 (0.83, 0.90)	0.532 (0.47, 0.59)
2002	tables	10 year	0.863 (0.82, 0.90)	0.784 (0.73, 0.83)	0.415 (0.33, 0.51)
up	Period	5 year	0.917 (0.89, 0.94)	0.858 (0.82, 0.89)	0.577 (0.52, 0.63)
to	splines	10 year	0.874 (0.83, 0.91)	0.789 (0.74, 0.83)	0.428 (0.35, 0.50)
2006	Non-PEH	5 year	0.924 (0.89, 0.95)	0.860 (0.82, 0.89)	0.559 (0.49, 0.62)
	splines	10 year	0.872 (0.83, 0.91)	0.785 (0.72, 0.83)	0.473 (0.36, 0.57)
	Life	5 year	0.933 (0.90, 0.96)	0.847 (0.80, 0.89)	0.551 (0.48, 0.62)
2003	tables	10 year	0.874 (0.83, 0.91)	0.773 (0.72, 0.83)	0.434 (0.34, 0.53)
up	Period	5 year	0.925 (0.89, 0.95)	0.852 (0.81, 0.89)	0.580 (0.51, 0.64)
to	splines	10 year	0.888 (0.84, 0.92)	0.783 (0.72, 0.83)	0.437 (0.35, 0.52)
2006	Non-PEH	5 year	0.936 (0.90, 0.96)	0.846 (0.80, 0.88)	0.575 (0.50, 0.65)
	splines	10 year	0.881 (0.83, 0.92)	0.782 (0.71, 0.84)	0.489 (0.36, 0.60)
2004†	Life	5 year	0.938 (0.90, 0.97)	0.886 (0.84, 0.93)	0.560 (0.48, 0.64)
up	tables	10 year	0.870 (0.82, 0.91)	0.845 (0.78, 0.90)	0.439 (0.34, 0.55)
to	Period	5 year	0.926 (0.88, 0.95)	0.884 (0.83, 0.92)	0.596 (0.50, 0.68)
2006	splines	10 year	0.887 (0.82, 0.93)	0.825 (0.75, 0.88)	0.447 (0.34, 0.55)

Table 8.2: Proportional and non-proportional excess hazards estimates from a delayed entry spline based model and from lifetables for five and ten year relative survival using three different periods. † Non-PEH models would not converge and cannot be shown.

analysis, and was also used in chapter 7, on simulation. In this table are PEH estimates labelled as period splines and non-PEH estimates labelled as non-PEH splines. As is shown in the table legend the 2004 to 2006 period non-PEH estimates did not converge using any of the available techniques offered by Stata or by altering the knot frequency.

By making the period narrower the estimates all show an improvement in survival, i.e. the five and ten year relative survival rates increase. However this has resulted in a loss of precision as the confidence intervals also become wider and as explained before there is a balance between precision and obtaining the most up-to-date estimates. In comparison to the models using all of the data, i.e. a standard analysis, the estimates show a fairly large improvement in both five and ten year survival rates under a PEH and non-PEH assumption.

The lifetable estimates are slightly different to the PEH modelling estimates but the non-PEH estimates are more similar. This would be expected as lifetables give estimates for specific time intervals and so by preventing age groups to vary by time in the PEH models it is expected that these estimates would be slightly different.

8.6 Adding patient sex

Patient Sex	Method	Rel surv estimate	Age group <60	60-75	75+
Male	Life tables	5 year	0.919 (0.88, 0.94)	0.916 (0.87, 0.95)	0.579 (0.49, 0.66)
		10 year	0.864 (0.82, 0.90)	0.856 (0.79, 0.92)	0.456 (0.33, 0.61)
	Period splines	5 year	0.922 (0.89, 0.94)	0.873 (0.84, 0.90)	0.641 (0.57, 0.70)
		10 year	0.884 (0.84, 0.92)	0.814 (0.76, 0.86)	0.508 (0.41, 0.60)
	Splines† interact	5 year	0.916 (0.88, 0.94)	0.899 (0.86, 0.93)	0.610 (0.53, 0.68)
		10 year	0.875 (0.83, 0.91)	0.851 (0.79, 0.90)	0.473 (0.37, 0.57)
Female	Life tables	5 year	0.922 (0.83, 0.97)	0.746 (0.67, 0.81)	0.480 (0.40, 0.56)
		10 year	0.862 (0.76, 0.93)	0.637 (0.55, 0.72)	0.368 (0.27, 0.49)
	Period splines	5 year	0.885 (0.84, 0.92)	0.816 (0.76, 0.86)	0.513 (0.44, 0.58)
		10 year	0.831 (0.76, 0.88)	0.733 (0.66, 0.79)	0.362 (0.28, 0.44)
	Splines† Interact	5 year	0.914 (0.82, 0.96)	0.765 (0.69, 0.82)	0.537 (0.45, 0.61)
		10 year	0.873 (0.74, 0.94)	0.667 (0.57, 0.75)	0.390 (0.30, 0.48)

Table 8.3: Proportional excess hazards estimates from a spline based delayed model and from Lifetables for five and ten year relative survival using three age groups and sex. † Including a sex and age group interaction

The estimates shown in table 8.2 are quite similar in their precision for a lifetable and spline based delayed entry model. This is because there are only three groups (age) and these groups all have a good number of patients in them. However if sex is added to the model and the estimation of the lifetables then this is in effect stratifying by another covariate so each group could be reduced by around 50%.

Table 8.3 shows the results from a period analysis of age groups and sex between 2002 and 2006. The lifetable results show more uncertainty than the modelled estimates as their intervals are wider for the five year estimates and even more so for ten years. This is an advantage of modelling in that it is possible to obtain adjusted estimates or estimates for certain covariate patterns by adding further covariates to the model.

The problem with the estimates in table 8.3 is that for some groups the difference between the lifetable and period estimates is quite large, for example 10-year relative survival estimates for female patients aged 60-75 differ by 10%. For this reason an interaction was fitted as there was some evidence to suggest (although not statistically significant) that the middle age group and sex were interacting. These results are also shown in table 8.3 and for the middle age groups the lifetable estimates are very similar to the period estimates (with interaction). This unfortunately suggests that an interaction may be neccesary, thus complicating the model and acting against the advantages gained through modelling. The confidence intervals have increased in width due to the extra parameters being modelled and both methods appear equally useful.

8.7 Discussion

The use of period analysis is fairly new to cancer and has yet to be adopted for heart disease. The evidence shown in this chapter suggests that this method could provide more up-to-date estimates of survival following an ST elevated AMI. Delayed entry models appear suitable and the splines have fitted well to the data, with the exception of the non-PEH period analysis of 2004-2006. Thus the new models developed in this thesis are also capable of using the increasingly popular method of period analysis.

Modelling may be more appropriate than lifetables as more precise estimates can be obtained. This is only true if reasonable assumptions can be made, such as proportional excess hazards. When the models include interactions, as shown in the previous section, the results will be similar to those obtained using lifetable stratification. A period analysis investigating genuine clinical research will be performed in chapter 9.

For period analysis to be accepted a retrospective assessment of the methods needs to be performed in heart disease as it has been performed in cancer (175; 176; 177). The dataset used here does not have enough follow-up data, but there are larger databases available that may be able to provide enough data from twenty years or so ago in order to determine the ability of period analysis to predict survival rates.

8.7.1 Further period approaches

When there is delayed recording of incident cases a period analysis is not feasible, for example incidence cases of cancer are often delayed by up to a couple of years in a cancer registry due to the time consuming nature of the process (179). Therefore hybrid analysis was proposed by Brenner and Rachet (179), as an approach that combines elements of both traditional and period analysis which may still be feasible and useful in such situations. As this is not the case in the LRI dataset this approach is not relevant for this thesis. However in large databases, perhaps MINAP (see section 10.6), this may prove to be true and as such the hybrid approach will become relevant.

A regression model can be fitted to survival rates, using year of diagnosis as a regression parameter, and extrapolating fitted trends not yet reported to obtain survival estimates for patients diagnosed in the current calender year. The developers of this method dubbed this as the *projection* method (180), but was later renamed as *cohort* modelling. This proved to be successful at predicting current year survival better than period analysis, especially when there is a moderate-to-large improvement in survival. It should be noted that period analysis was not designed to specifically estimate the survival of patients diagnosed in the current year.

In the same year period trend analysis was proposed by Brenner and Hakulinen (181) which in the following year extended the time window used for modelling to obtain more precise estimates that are accurate and up-to-date (182). This method has also been used in practice (183; 184). This model was motivated by the lack of precision found using a period analysis on just the most recent year, as the only way to achieve a higher precision is to enlarge the period under investigation, which in turn loses up-to-dateness. This involves modelling a trend on period analysis data which would be fairly easy to incorporate into the spline-based models.

The cohort (180) and period (181; 182) trend models were compared with each other in 2008 (185), which is the first time the two approaches were directly compared. The two methods were empirically evaluated and their performance compared at providing up-to-date estimates of 5-year relative survival. It was found that period modelling had much lower standard errors than those estimated by cohort models. For a clear majority of cancers, period modelling also provided better prediction of 5-year relative survival than was later observed for patients diagnosed in the 5-year period around the calender year for which modelling was performed. The period models retained their advantage when fitting models to predict survival by extrapolation of linear trends.

No trend models have been fitted to the LRI data and this may prove to be interesting in the future, allowing extrapolation of survival for patients diagnosed in the current calender year and the ability to achieve a good level of precision may prove to be highly clinically beneficial. Period trend models appear to have advantages over cohort trend models, but it would still be the method of choice for other purposes, such as estimate cohort effects or trends, i.e. changes in survival expectation between subsequent cohorts, for example where the effect of a new treatment, only used in the most recent cohort, is of interest. The disadvantage of both modelling approaches is the added complexity of the analysis in comparison to standard period approaches. However these trend models would be fairly easy to implement using the spline based models.

Chapter 9

Relative Prognostic Impact of Diabetes Diagnosis and Admission Blood Glucose Concentration After ST-Elevation Acute Myocardial Infarction: A Standard Survival Analysis and Extension to Relative Survival

9.1 Introduction

This chapter will investigate short term and long term survival after an ST-elevated MI (STEMI), with the primary investigation being the impact of diabetic status and admission blood glucose on mortality. This begins with an investigation of short term survival over 30-days and up to a year, for which no censoring occurred and thus logistic regression was adopted. Long term follow-up, of six-years, was assessed using relative survival. Piecewise and MFP models were used with split-time data, and flexible parametric spline based models were used on non-split-time data. Modelling of several covariates will be performed to adjust for potential confounders, including a glucose and diabetes interaction.

The main aim was to explore the use of the spline based flexible parametric models developed in chapter 5, to answer a real research question. Continuous covariates will be assessed along with proportional and non-proportional excess hazards models. A compar-

ison with all-cause models will be performed to highlight the impact that relative survival has on the estimation of mortality. Finally, period analysis will be applied to obtain up-to-date estimates of survival.

9.2 Short-term diabetes and glucose

9.2.1 Paper summary

A paper was submitted in 2008 (186), shown in appendix E, investigating survival trends in 4702 patients from routine practice for which the analysis was performed by myself. The aims of this paper were to study the relative influence upon survival, and temporal trends therein, of the diagnosis of diabetes and of admission blood glucose concentration in patients with STEMI. Glucose was classified into quartiles to obtain simple odds ratios. 30-day mortality was the primary endpoint of the study with one-year mortality as a secondary endpoint. The short term data were modelled using a logistic GLM in an MFP model. The impact of calender year was of interest as a way of accounting for different treatments introduced throughout the follow-up period.

Short-term mortality is known to be very highly associated with an MI. The proportion of patients who died was assessed using a logistic model whose fitted values created a trend line. While adjusting for glucose and other covariates, including, sex, previous acute myocardial infarction (AMI), Creatine Kinase (CK), creatinine, thrombolysis and year, the effect of diabetes was found to be non-significant as glucose was found to be a better predictor of short-term mortality.

9.2.2 Introduction

Abnormalities of glucose metabolism are known to have a powerful influence upon prognosis for patients with coronary heart disease (187; 188). In the setting of a patient with an AMI, a prior diagnosis of diabetes is associated with an increase in risk of adverse outcome. Elevated blood glucose is generally common among patients hospitalised with AMI irrespective of diabetic status. Known as 'stress hyperglycaemia', it is similarly associated with patient prognosis (187; 188; 189; 190; 191; 192).

Current algorithms for risk assessment after AMI do not consider blood glucose (193; 194; 195) and contemporary AMI management guidelines do not include recommendations on therapeutic blood glucose targets (196; 197; 198). Changes in criteria for the diagnosis of diabetes, and for AMI, as well as developments in the pharmacological management of AMI, make the comparison of outcomes for patients experiencing AMI at different periods difficult.

9.2.3 Methods

The LRI CCU dataset was used as described in section 2.3. The two pre-defined aims of the analysis are, (i) the relative impact on 30-day and 1-year all-cause mortality of both antecedent diabetes and of admission blood glucose concentration and (ii) trends over this time period in the association with mortality of these parameters.

Admission blood glucose was recorded and is assessed here. As mean admission blood glucose concentration fell over the study period, admission glucose was divided into quartiles of the range observed over the entire study period: Quartile 1, $<$ 7mmol/L; Quartile 2, $7 - 8.2$ mmol/L; Quartile 3, $8.3 - 10.9$ mmol/L; Quartile 4, $\geq$ 11 mmol/L. Continuous glucose will be assessed in the long-term follow up in section 9.8.

9.2.4 Results

30-day mortality proportions were calculated for the complete population with follow-up censored at January 30th 2006. One-year mortality proportions were calculated for patients admitted up to December 31st 2004, censored at December 31st 2005, this allowed full follow-up for all patients.

During the study period (1993-2005), in line with accumulating evidence and clinical experience, increasing proportions of patients received treatment with statin, beta-blocker or renin-angiotensin system inhibitor agents as shown in figure 9.1. Such major changes render difficult the examination of individual treatment effects. However some correction for increasing secondary prevention therapy was included by multivariable analyses of year of hospitalisation as seen in figure 9.1 the proportions receiving various treatments vary greatly over time. Treatment with thombolysis was included in the model in view of the lack of change in the indications for, and the relatively constant proportion of patients receiving this therapy over the period (70% - 77% annually).

Antecedent Diabetes

Antecedent diabetes was recorded for 749 (15.9%) patients, increasing from 13.9% in 1993 to 20.3% in 2005 as shown in figure 9.2. Patients with antecedent diabetes were older by an average of 1.4 years and a higher proportion were female, compared to patients without this diagnosis. Mean admission glucose and creatinine concentrations were higher, but peak CK was lower, in patients with diabetes as shown in table 9.1. Data are presented as differences in means and proportions, with 95% confidence intervals (CI). Patients with antecedent diabetes were less likely to receive thrombolysis or to be prescribed aspirin or beta-blocker during the index admission. However they were more likely to receive diuretic

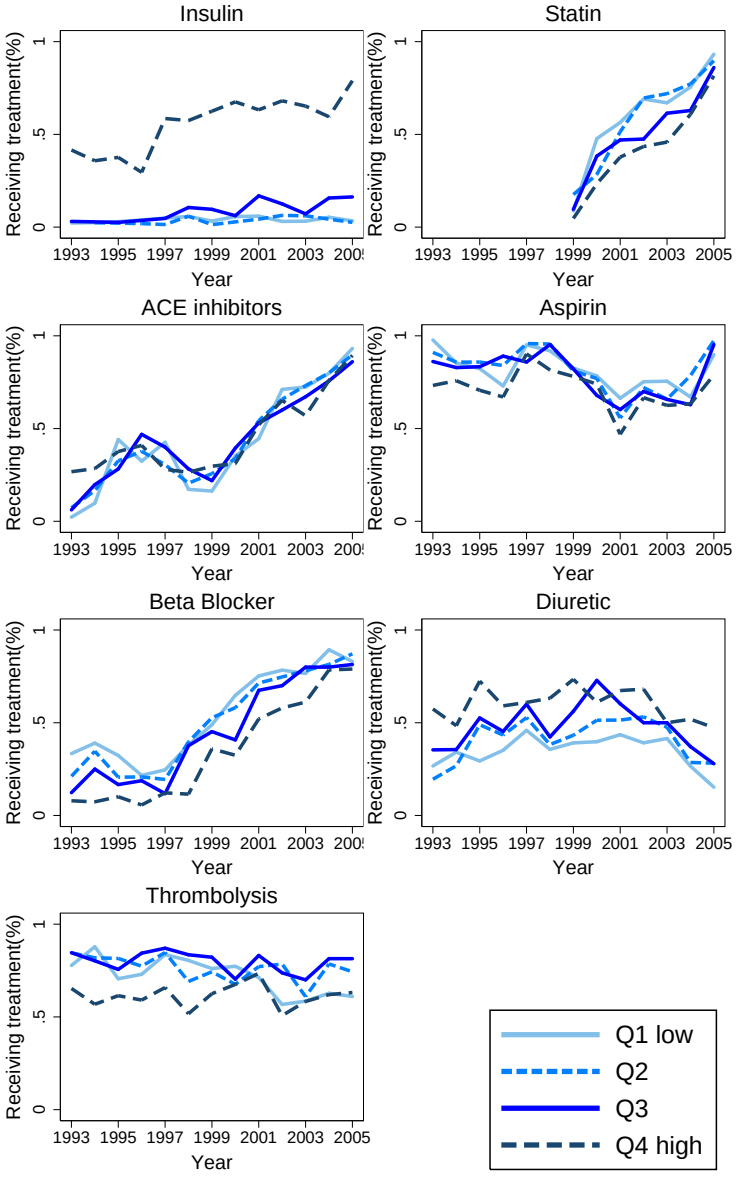


Figure 9.1: Proportion of patients receiving treatments over the study period.

therapy, inhibitors of the renin-angiotensin system and insulin as shown in table 9.1.

Admission blood glucose

Baseline differences between groups were examined using independent two-sample t-tests for continuous variables and chi-squared tests for categorical variables. In the overall population, mean admission blood glucose concentration declined from 10.1 mmol/L in 1993 to 9.2 mmol/L in 2005 ($p = 0.017$) (Figure 9.2). The fall was greater for those with diabetes, falling by an average of 0.39 (95% CI $0.30, 0.48$) mmol/L per year, compared to those without, 0.09 (95% CI $0.05, 0.12$) mmol/L per year (test for interaction $p < 0.001$).

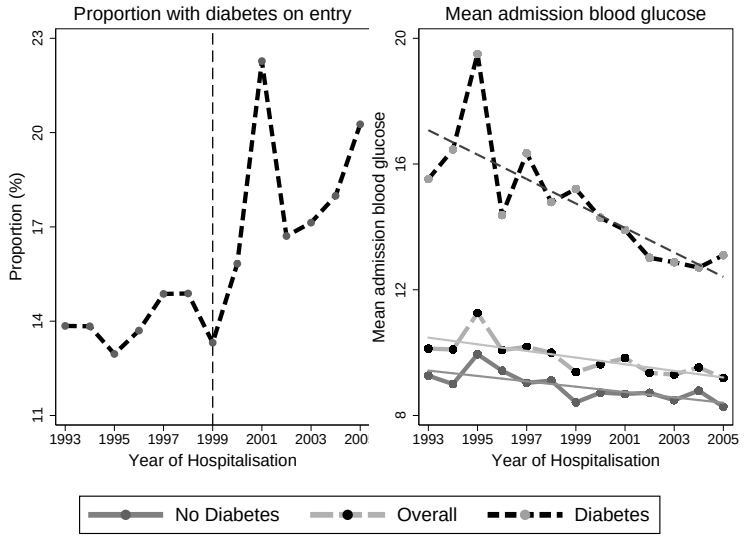


Figure 9.2: Proportion of patients with a-priori diagnosis of diabetes (left) and mean plasma glucose concentration (right) by year of admission. Linear curves show estimates from a regression model including year, diabetic status and year*diabetes interaction.

Variable	Antecedent Diabetes		Difference
	Yes	No	(Diabetes-No Diabetes)
	N=759(15.9%)	N=3951(84.1%)	
Male (%)	460 (61.4%)	2737 (69.3%)	-7.9 (-11.6 , -4.1)
Age (years)	67.9 ± 0.4	66.5 ± 0.2	1.4 (0.5 , 2.3)
Plasma glucose (mmol/L)[1]	14.7 ± 0.3	8.9 ± 0.1	5.7 (5.2 , 6.3)
Creatinine (mol/L)[2]	120.7 ± 2.3	109.2 ± 0.7	11.5 (6.8 , 16.3)
Peak CK (IU/L)[3†]	1818 ± 62	2115 ± 31	-296 (-433 , -159)
History of			
Smoking			
Current (%)	156 (20.8%)	1312 (33.2%)	-12.4 (-15.7 , -9.1)
Former (%)	183 (24.4%)	992 (25.1%)	-0.7 (-4.0 , 2.7)
Never (%)	410 (54.7%)	1645 (41.6%)	13.1 (9.2 , 17.0)
Not known (%)	0	2	
Previous AMI	175 (23.4%)	571 (14.5%)	8.9 (5.7 , 12.1)
Angina (%)	230 (30.7%)	862 (21.8%)	8.9 (5.3 , 12.4)
Hypertension (%)	368 (49.1%)	1296 (32.8%)	16.3 (12.5 , 20.2)
Hyperlipidaemia (%)	151 (20.2%)	451 (11.4%)	8.7 (5.7 , 11.8)
Cerebrovascular disease (%)	83 (11.1%)	191 (4.8%)	6.2 (3.9 , 8.6)
Treatment			
Thrombolysis (%)	468 (62.5%)	2883 (73.0%)	-10.5 (-14.2 , -6.8)
Diuretic (%)	457 (61.0%)	1743 (44.1%)	16.9 (13.1 , 20.7)
Aspirin (%)	528 (70.5%)	3099 (78.4%)	-7.9 (-11.5 , -4.4)
Beta Blocker (%)	282 (37.7%)	1798 (45.5%)	-7.9 (-11.7 , -4.1)
Insulin (%)	510 (68.1%)	280 (7.1%)	61.0 (57.6 , 64.4)
ACE Inhibitor/ARB (%)	385 (51.4%)	1580 (40.0%)	11.4 (7.5 , 15.3)
Statin (%)	196 (26.2%)	1054 (26.7%)	-0.5 (-3.9 , 2.9)

Table 9.1: Demographic and in-hospital treatment characteristics of patients with and without antecedent diagnosis of diabetes. Table shows mean standard error or number (%). CK = Creatine Kinase; ACE = Angiotensin converting Enzyme; ARB = Angiotensin Receptor Blocker. [1] 101 missing values for diabetes and 581 missing values for no diabetes; [2] 9 missing values for diabetes and 39 missing values for no diabetes; [3] 9 missing values for diebetes and 39 missing values for no diabetes; [†]Normal range < 200

		GQ1 <7		GQ2 7-8.3		GQ3 8.3-11		GQ4 ≥11		Missing Glucose		Overall	
		n	%	n	%	n	%	n	%	n	%	n	%
Sex	Male	713	76.7	776	72.1	609	63.9	643	60.7	457	66.9	3,198	68
	Female	217	23.3	300	27.9	344	36.1	417	39.3	226	33.1	1,504	32
Diabetes	No	882	94.8	1,027	95.4	855	89.7	606	57.2	581	85.1	3,951	84
	Yes	48	5.2	49	4.6	98	10.3	453	42.7	101	14.8	749	15.9
	Missing	0	0	0	0	0	0	1	0.1	1	0.1	2	0.04
Statin	No	567	61	803	74.6	714	74.9	859	81	508	74.4	3,451	73.4
	Yes	363	39	273	25.4	239	25.1	201	19	175	25.6	1,251	26.6
ACE or ARB	No	499	53.7	655	60.9	547	57.4	608	57.4	427	62.5	2,736	58.2
	Yes	431	46.3	421	39.1	406	42.6	452	42.6	256	37.5	1,966	41.8
Diuretic	No	596	64.1	645	59.9	483	50.7	418	39.4	360	52.7	2,502	53.2
	Yes	334	35.9	431	40.1	470	49.3	642	60.6	323	47.3	2,200	46.8
Beta Blocker	No	373	40.1	578	53.7	534	56	733	69.2	403	59	2,621	55.7
	Yes	557	59.9	498	46.3	419	44	327	30.8	280	41	2,081	44.3
Thrombolysis	No	272	29.2	248	23	193	20.2	408	38.5	229	33.5	1,350	28.7
	Yes	658	70.8	828	77	760	79.8	652	61.5	454	66.5	3,352	71.3
Insulin	No	895	96.2	1046	97.2	877	92	493	46.5	600	87.9	3,911	83.2
	Yes	35	3.8	30	2.8	76	8	567	53.5	83	12.2	791	16.8
Death in 30 days	No	846	91	962	89.4	782	82.1	731	69	540	79.1	3,861	82.1
	Yes	84	9	114	10.6	171	17.9	329	31	143	20.9	841	17.9
30-day survival total population		930	19.78	1,076	22.88	953	20.27	1,060	22.54	683	14.53	4,702	
Death in 1 year	No	749	86	857	82.6	682	75	616	60.3	458	72.1	3,362	75.2
	Yes	122	14	180	17.4	228	25.1	405	39.7	177	27.9	1,112	24.8
	n	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Overall mean(sd)**	
Age	4,702	63.8	13.27	65.3	12.88	67.6	12.1	69.4	11.54	67.4	13.06	66.7(12.69)	
SBP	4,627	141.8	28.06	139.8	28.49	137.6	31.71	135.4	36.48	137.2	31.02	138.4(31.44)	
Creatinine (mol/L)	4,654	103.3	47.33	102.6	35.78	109.1	38.73	126.3	56.83	113.8	5.64	111.0(48.19)	

Table 9.2: Demographic and in-hospital treatment characteristics of patients according to glucose quartile. ACE = Angiotensin converting Enzyme; ARB = Angiotensin Receptor Blocker; SBP = Systolic Blood Pressure. 1-year survival was assessed on patients admitted up to 31st December 2004, being the last date for which 1-year follow-up was available. Therefore the total number of patients included in the one-year analysis (n=4474) is fewer than in the 30-day analysis (n=4702).

Table 9.2 shows patient demographics by glucose quartile, including those for whom a value of glucose was missing. Patients in glucose quartile 4 were on average 5.6 years older than those in quartile 1. Predictably, as glucose increased, so did the proportion of patients with diabetes. However even in glucose quartiles 3 (10.3%) and 4 (42.7%), a minority of patients had a diabetes diagnosis. Serum creatinine increased, and admission systolic blood pressure fell as glucose quartile increased.

Thrombolysis was administered in 61.5% of patients in quartile 4 compared to $70-80\%$ in glucose quartiles 1 to 3. In-hospital diuretic prescription increased and beta-blocker prescription fell as glucose increased. There was little variation among glucose quartiles in the use of inhibitors of the renin-angiotensin system. Where glucose values were missing, demographic and treatment features were broadly similar to those seen in the population as a whole. In terms of proportions with antecedent diabetes, treatment with insulin, and death during follow-up, these patients most closely resembled those in quartile 3 (Table 9.2).

Survival

During follow up (mean 5.11 years, range $0-4684$ days), overall case fatality was 42.2% (1992/4702). Thirty day mortality was 17.9% (N=841, 42.2% of deaths), and 1-year mortality 24.2% ($N = 1136, 57.0\%$ of deaths). Thirty-day mortality was 20.7% for patients admitted in 1993 compared to 5.6% for those admitted in 2005, a mean reduction in the odds of mortality of 6.3% (95% CI $4.1, 8.4; p < 0.001$) each year. Over the same period 1-year mortality fell from 29.1% to 18.2%, a mean annual reduction in the odds of mortality of 4.3% (95% CI $2.4, 6.2$) ($p < 0.001$). The association between all-cause mortality and year of index AMI was assessed using unconditional logistic regression analysis with covariate effects reported as odds ratios (OR) with 95% CI. While non-linear effects (log odds scale) of year of diagnosis were assessed using fractional polynomials, in all cases a linear effect (log odds scale) provided the best fitting model using the AIC (123).

Antecedent Diabetes and Survival

Figure 9.3 shows temporal trends in all-cause mortality according to antecedent diabetes status. Overall 30-day and 1-year mortality was 22.8% and 31.3% respectively for patients with, compared to 16.3% and 23.0% for those without, antecedent diabetes. From $1993-2005$, 30-day mortality fell from 28.3% to 8.8% in patients with antecedent diabetes (mean annual reduction in the odds of mortality $9.2\%, 95\%$ CI $4.4, 13.7$) and from 19.4% to 4.8% (mean annual reduction $5.9\%, 95\%$ CI$3.4, 8.2$) in patients without this diagnosis.

Over this period, 1-year mortality fell from 39.1% to 27.1% in patients with antecedent

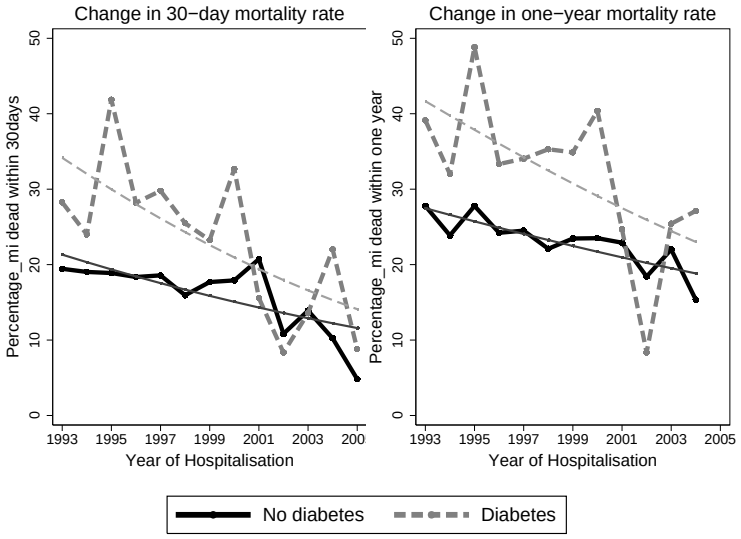


Figure 9.3: Change in overall 30-day (left) and 1-year (right) all-cause mortality proportion over time for patients with and without antecedent diagnosis of diabetes. Trend lines estimated using logistic regression.

diabetes (mean annual reduction in odds of mortality $7.6\%, 95\%$ CI $2.9, 12.1$), and from 27.8% to 15.3% (mean annual reduction $4.4\%, 95\%$ CI $2.06, 6.60$) in those without this diagnosis. These differences in annual risk reduction between patients with and without antecedent diabetes did not achieve statistical significance for 30-day ($p = 0.223$) or 1-year ($p = 0.217$) mortality.

Blood glucose and Survival

Table 9.2 shows overall 30-day and 1-year mortality, and figure 9.4 temporal patterns in these outcomes, by glucose quartile. Overall 30-day mortality for patients with glucose in the highest quartile was over 3-times higher (31%) than in quartile 1 (9.0%). It was also shown that even modest elevation of blood glucose above the median value of 8.3mmol/L was associated with increased mortality. From 1993-2005, mortality risk reduction was proportionately similar for all blood glucose quartiles. Mean annual reduction in the odds of 30-day mortality was $5.41\%(95\%$ CI $-1.00, 11.4)$, $4.2\%(95\%$ CI $-0.97, 9.2)$, $4.5\%(95\%$ CI $-0.01, 8.9)$ and $6.6\%(95\%$ CI $3.2, 10.0)$ for glucose quartile $1, 2, 3$ and 4 respectively. The similar relative risk reductions translated to greater absolute improvements for patients with the greatest elevation of glucose: from 1993-2005, absolute 30-day mortality risk fell by 3.3% in quartile 1, 8.0% in quartile 2, 18.4% in quartile 3 and 26.8% in quartile 4. Mean annual reduction in 1-year mortality risk was $2.80\%(95\%$ CI $-3.2, 8.4)$, $1.19\%(95\%$CI $-3.4, 5.6)$, $3.50\%(95\%$CI $-0.9, 7.7)$ and $6.00\%(95\%$ CI $2.5, 9.4)$ for glucose quartile $1, 2, 3$ and 4 respectively. The absolute reduction in 1-year mortality risk was $6.0\%, 2.0\%, 19.8\%$ and 16.2% for quartiles 1 to 4 respectively.

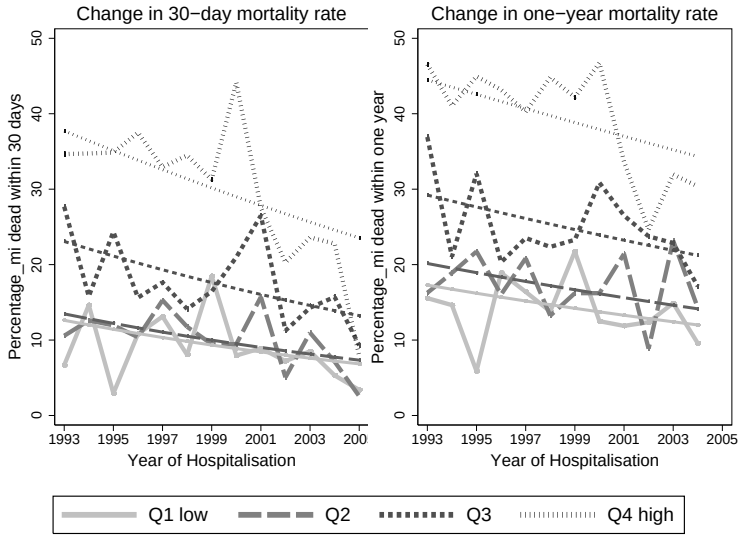


Figure 9.4: Change in overall mortality proportion over time by admission blood glucose quartile. 30-day (left) and 1-year (right) all-cause mortality. Trend line estimated via logistic regression.

An interaction was fitted between year of diagnosis and glucose to determine if the effect of glucose was different over the different follow-up periods. The relative impact on mortality of blood glucose concentration remained approximately constant from 1993-2005, as indicated by the non-significant interaction between year of diagnosis and glucose concentration for both the 30-day ($p = 0.835$) and 1-year mortality analyses ($p = 0.3870$) as determined by a likelihood ratio test. For 30-day mortality, compared to quartile 1, the OR for quartile 2 was $1.1(95\%$ CI $0.80, 1.46)$, for quartile 3 was $2.08(95\%$ CI $1.57, 2.75)$ and quartile 4 was $4.19(95\%$CI $3.23, 5.45)$. For 1-year mortality, compared to quartile 1, the OR for quartile 2 was $1.21(95\%$ CI $0.94, 1.56)$, for quartile 3 was $1.98(95\%$ CI $1.55, 2.52)$ and quartile 4 was $3.83(95\%$CI$3.04, 4.83)$. The likelihood ratio tests were non-significant and the AICs given by the 30-day model were 3488.421 and 3493.559 for the model with and without the interaction respectively. Similarly the one-year analysis gave an AIC of 4057.094 and 4060.065 for the model with and without the interaction respectively. These models were unadjusted for any other covariates.

Glucose and diabetes relative impact on prognosis

Using multivariable, unconditional logistic regression models, the impact upon prognosis of antecedent diabetes, glucose quartile (relative to quartile 1), and year of index AMI, in various combination were assessed (table 9.3). Unadjusted analyses for diabetes and glucose quartile were followed by analysis adjusted for age at AMI, sex, peak CK, creatinine, previous AMI, year of hospitalisation and administration of thrombolysis, including diabetes and glucose individually and together. Fractional polynomials were used to model

Parameter		Unadjusted	Adjusted[1] (Excluding Diabetes)	Adjusted[2] (Excluding Glucose)	Adjusted[3]
		(n=4018)	(n=4002)	(n=4002)	(n=4002)
			30 DAY		
Glucose	**Q1**	1	1	-	1
	Q2	1.193 (0.887 , 1.606)	1.077 (0.774 , 1.500)	-	1.080 (0.776 , 1.503)
	Q3	2.202 (1.667 , 2.910)	1.739 (1.274 , 2.373)	-	1.749 (1.281 , 2.387)
	Q4	4.539 (3.502 , 5.882)	2.504 (1.865 , 3.361)	-	2.604 (1.912 , 3.546)
Diabetes	**No**	1	-	1	1
	Yes	1.518 (1.236 , 1.863)	-	1.282 (1.008 , 1.630)	0.897 (0.691 , 1.164)
Year		0.937 (0.916 , 0.959)	0.934 (0.909 , 0.960)	0.926 (0.901 , 0.951)	0.935 (0.910 , 0.961)
	AIC		2866.548	2917.873	2867.872
Parameter		**Unadjusted**	**Adjusted[1]**	**Adjusted[2]**	**Adjusted[3]**
		(n=3839)	(n=3825)	(n=3825)	(n=3825)
			ONE YEAR		
Glucose	**Q1**	1	1	-	1
	Q2	1.289 (1.004 , 1.656)	1.238 (0.927 , 1.654)	-	1.240 (0.928 , 1.655)
	Q3	2.052 (1.610 , 2.617)	1.637 (1.237 , 2.168)	-	1.643 (1.240 , 2.175)
	Q4	4.036 (3.211 , 5.074)	2.283 (1.742 , 2.991)	-	2.340 (1.759 , 3.112)
Diabetes	**No**	1	-	1	1
	Yes	1.520 (1.258 , 1.837)	-	1.253 (0.998 , 1.573)	0.936 (0.729 , 1.201)
Year		0.953 (0.933 , 0.973)	0.952 (0.927 , 0.977)	0.944 (0.920 , 0.969)	0.952 (0.927 , 0.978)
	AIC		3216.212	3254.922	3217.941

Table 9.3: Results of modelling proportions surviving to 30 days and one year; Adjusted models are adjusted for age at MI, sex, previous AMI, CK, creatinine, thrombolysis and year of hospitalisation and: [1] for blood glucose concentration (quartile); [2] antecedent diabetes and [3]for both blood glucose concentration (quartile) and antecedent diabetes; Year = Odds ratio per year 1993-2005 (30-day) or 1993-2004 (1-year); AIC = Akaike's information criterion (lower values indicate a better fitting model).

continuous variables (124).

After adjustment for covariables, there was a reduction of approximately 7% in the odds of 30-day mortality each year from 1993-2005, and a reduction of approximately 5% per year for 1-year mortality. The unadjusted odds of both 30-day and 1-year mortality was approximately 52% higher in patients with diabetes. Adjustment for covariates, excluding glucose, reduced this to approximately 25% at each time period. However, when glucose quartiles were also included, diabetes was associated with a non-significant lower risk (table 9.3). In contrast, high glucose retained a powerful association with mortality risk, although this was reduced when adjusting for other covariates.

Compared to quartile 1, adjusted odds of mortality was similar in quartile 2, approximately 80% higher in quartile 3 and $150-200$% higher in quartile 4. The 30-day adjusted model with glucose had an AIC of 2866.548 but when diabetes was added to this the AIC was 2867.872 with a likelihood ratio test given as p=0.4112 and the one-year test gave a p-value of 0.6025 showing no significant improvement over the model excluding diabetes. There is not an interaction between year of follow-up and glucose, which means the effect of glucose is fitted to be constant over all years as the interaction was not found to be significant.

9.2.5 Discussion

Between 1993 and 2005, the absolute mortality associated with the antecedent diagnosis of diabetes fell markedly. Similarly, the mortality associated with elevated blood glucose also fell, particularly for patients in whom admission glucose was most elevated. However they both retain a strong influence on patient prognosis. Finally, and importantly, in patients with STEMI, blood glucose concentration at admission to hospital has greater prognostic relevance than antecedent diabetes with regard to short term survival.

While these results are from a single centre, the observations reflect outcomes in routine practice and are likely to be representative of survival trends. It is also possible that the outcome related to glucose concentration may simply be representative of unrecognised diabetes. There were no adjustments for the use of individual secondary prevention therapies. The impact of these therapies was not the main focus of the analysis, and consideration of individual treatment effects may introduce bias for a number of reasons. In particular, in the most severely ill patients, early mortality and adverse clinical features in survivors will limit treatment prescription.

9.3 Extending short-term to long-term

The previous analysis was justifiable for short-term follow-up as a subject's expected mortality due to other causes will be very small and not change by much in this time. This is because the chance of dying in the next thirty days is fairly low in such a short period. However investigation over a longer follow-up period will involve an aging patient and, as time goes on, fewer deaths are likely to be attributable to the initial AMI and thus relative survival may prove to be useful.

The aims of this chapter are to create an extension of the previous work into long-term survival by adopting a relative survival approach. Long term relative survival will be assessed by both diabetic status and glucose quartiles using piecewise models (115), fractional polynomials (39) and spline based flexible parametric models (3).

The primary aim is to determine if similar results to the short-term analysis are observed over a long period of time. The analysis will initially investigate piecewise and MFP approaches using split-time data. It is difficult to adjust for various factors when using split-time data (see chapter 4) and as such a full analysis of the data using the spline based flexible parametric models described in section 5 will be used to assess the relationship between diabetes and glucose on patient mortality. This provides a *test* of the spline based models, developed in chapter 5, ability to investigate a complex genuine research question.

9.3.1 Data setup and description

A six year follow-up was observed for the same population used for the 30-day analysis, i.e. only patients who were hospitalised up to the end of 2005 are included, where $n = 4018$ as shown in the short term table for the unadjusted 30-day model in table 9.3. If a patient died on the date of hospitalisation (n=225) then survival was set to half a day. The spline based models (chapter 5) do not split the timescale, but use traditional relative survival models the following split times were adopted:

```
0.01 0.025 0.05 0.075 0.1 0.15 0.2 0.25   0.3(0.1)1   1.2(0.2)6
```

These relate to the time in years, but can be converted to intervals in days:

```
        0-3 4-9 10-18 19-27 28-36 37-54 55-73 73-91 92-109
every 36 days up to a year then every 73 days up to 6 years
```

This provided lots of splits during the first three months (91.3 days) and then relatively fine splits for the remainder of the period similar to the user defined intervals described in section 4.2. These splits are suitable for the fractional polynomials, but are likely to be over-parameterised in piecewise GLMs using more splits than necessary (as shown in

section 4.3). Less splits were used at:

```
0 0.08333 0.2 0.5 1 1.5 2 6
```

By combining years two to six it is assumed that the baseline hazard is constant over this time period. Given that there is little information in the final few years of follow-up it is reasonable to combine the data in order to reduce the number of parameters that need estimation.

By group	Non-diabetics	Diabetics	Overall
Glucose Q1	63.7	65.6	63.8
Glucose Q2	65.1	69.0	65.3
Glucose Q3	67.8	65.9	67.6
Glucose Q4	70.6	67.7	69.4
Overall	66.4	67.4	66.6

Table 9.4: Average age of patients by glucose and diabetes

Table 9.4 shows the average age at hospitalisation within the four glucose quartiles. This is shown to increase by around two years per quartile (per row in the table) and diabetics are younger in the high quartile groups and older in the low glucose groups. These differences are not large and the expected survival rates were found to be fairly similar across the quartiles. It is interesting to look at the number at risk in each quartile group and diabetic group as shown in table 9.5 at the start of the first time interval (0 years) and the start of the last time interval (5.8 years).

Glucose Quartile	First interval (0-0.01)			Last interval (5.8-6)		
	Diabetic					
	No	Yes	Overall	No	Yes	Overall
Q1	882 (94.8)	48 (5.2)	930	304 (97.4)	8 (2.6)	312
Q2	1027 (95.4)	49 (4.6)	1076	451 (97.8)	10 (2.2)	461
Q3	855 (89.7)	98 (10.3)	953	306 (92.4)	25 (7.6)	331
Q4	606 (57.2)	453 (42.8)	1059	146 (58.6)	103 (41.4)	249
overall	3370 (83.9)	648 (16.1)	4018	648 (81.6)	146 (18.4)	794

Table 9.5: No at risk (with % diabetic) at the start of the follow-up and the end of the follow-up

Table 9.5 shows that over 300 patients survived the follow-up out of 930 patients who were hospitalised in glucose quartile one. There are only 146 diabetic patients alive at the end of follow-up, this group also has the smallest population in the cohort at the beginning of follow-up (648). Of the 4018 individuals at the start of the follow-up 83.9% were non-diabetic and for the final interval 81.6% were non-diabetic, which is a slight reduction. There are only 8 diabetic patients at risk in the final interval in glucose quartile one, which is only 2.6% of the patients at risk in this glucose group.

The estimates of observed, expected and relative survival for the final interval, taken as

Diabetic status		Glucose			
		Quartile 1	Quartile 2	Quartile 3	Quartile 4
Non-Diabetic	O	0.733	0.670	0.601	0.381
	E	0.842	0.829	0.807	0.790
	R	0.871	0.809	0.744	0.483
Diabetic	O	0.555	0.616	0.616	0.411
	E	0.841	0.841	0.860	0.837
	R	0.660	0.733	0.716	0.491

Table 9.6: Observed, Expected and Relative survival rates for the final interval ($5.8 - 6$ years)

5.8 to 6 years from a lifetable, are shown in table 9.6 by glucose quartile and diabetic status. Here it shows that the expected survival is fairly similar in diabetics across the quartiles with only a 5% difference in non-diabetics between quartiles one and four. The relative survival (O/E) is lowest in the fourth quartile.

9.4 Piecewise modelling

The aim of this analysis is to model glucose and diabetes using relative survival with comparison to a stratified Cox model, which is used to calculate all-cause survival. This will enable the adjustment of diabetes or glucose for each other and allow time dependent effects to be assessed. An individual level piecewise model using a Poisson GLM (115) was adopted as it is equivalent to a full likelihood approach (27) as discussed in section 4.3.2.

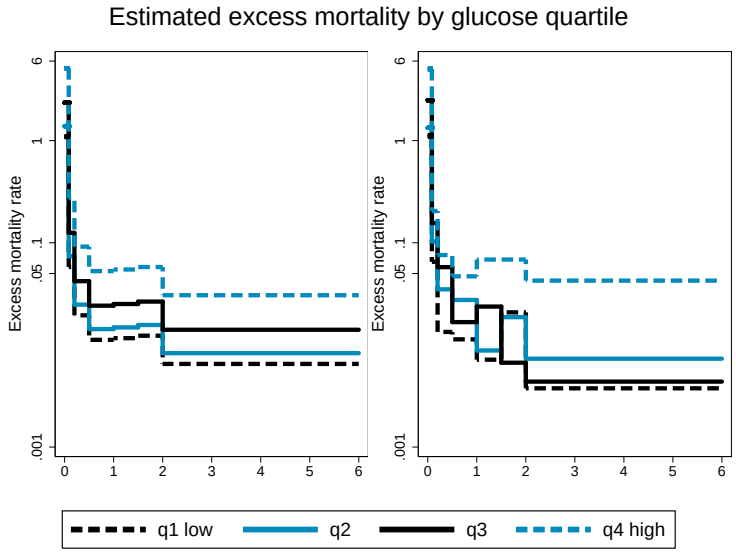


Figure 9.5: Plot of the excess mortality rate by glucose quartile for the PEH (left) and non-PEH model(right)

The first fitted model only included glucose quartiles as a covariate and the excess mortality

rate plots are shown in figure 9.5. There are two plots in this figure the left-hand side is for a PEH model whereas the right-hand plot allows each glucose quartile to vary by time. The time dependent effects were assessed using a global likelihood ratio test and was found to be non-significant with a value of $\chi^2_{18} = 23.44$ where $p = 0.1741$. Similar results were found from a grouped analysis, not shown, to those from this individual level analysis. It is easier to interpret these estimates in 1000 person years, for example in glucose quartile four there are 30.7 excess deaths per thousand person years in the $2-6$ year interval from the PEH model shown in figure 9.5.

The proportional excess hazards (PEH) model estimates the fourth quartile as having the worst excess mortality, which in a PEH model is forced to be proportional to the other three quartile groups, however this difference is emphasised in the non-PEH model where quartile four (high glucose) is even more separated from the lower groups. The excess mortality is low at around five excess deaths per 1000 person years for those in the lower glucose levels for those who survived two years. In both models glucose quartiles 1 and 2 are very similar. The excess hazard rate ratios are shown shortly in table 9.7 for the PEH model which show that quartile 4 has a very large excess hazard rate ratio (4.771) when compared to the lowest glucose quartile group.

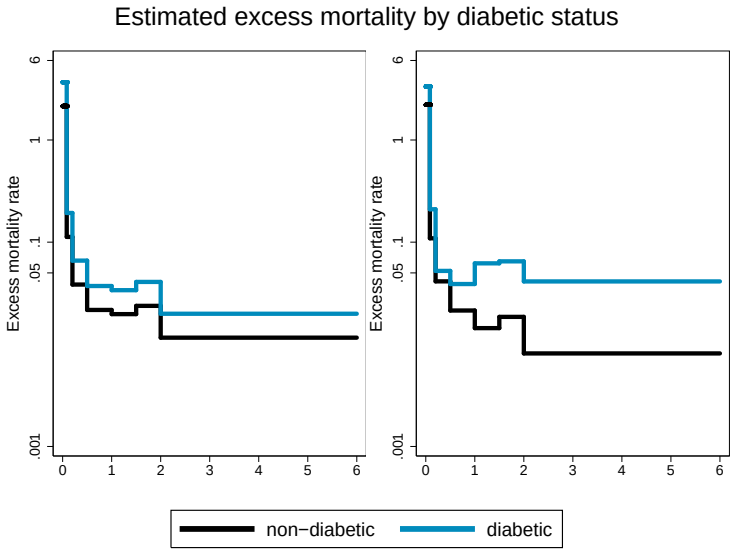


Figure 9.6: Plot of the excess mortality rate by diabetic status for the PEH (left) and non-PEH model(right)

The next model only included diabetic status and the results are shown in figure 9.6. A likelihood ratio test was performed to see if the addition of time dependency for diabetes was statistically beneficial to the model, and with a value of $\chi^2_6 = 19.42$ where $p = 0.0035$ it was found to be of statistical importance.

The PEH plot suggests around 30 excess deaths (per 1000py) for both groups (fewer in the non diabetic group) after 6-months then much less following two years. The exact

Method	Model	Proportional					Non-prop
		Glucose quartile			Diabetes	AIC[1]	AIC[2]
		Q2	Q3	Q4			
Dickman	Glucose	1.274	2.159	4.703	-	13987.5	14000.2
		(0.97 , 1.67)	(1.68 , 2.77)	(3.74 , 5.91)	-		
	Diabetes	-	-	-	1.715	14250.7	14243.5
		-	-	-	(1.46 , 2.02)		
	G and D	1.276	2.186	5.046	0.841	13985.9	14000.8
		(0.98 , 1.67)	(1.70 , 2.81)	(3.97 , 6.41)	(0.70) , (1.01)		
Cox	Glucose	1.240	1.631	2.984	-	23391.7	23393.2
		(1.05 , 1.47)	(1.38 , 1.93)	(2.56 , 3.48)	-		
	Diabetes	-	-	-	1.523	23613.6	23622.6
		-	-	-	(1.34 , 1.73)		
	G and D	1.240	1.639	3.082	0.922	23392.4	23399.3
		(1.04 , 1.47)	(1.38 , 1.94)	(2.61 , 3.63)	(0.80 , 1.06)		

Table 9.7: Estimates for glucose quartiles and diabetes with AICs for three models using two methods

number of estimated excess deaths is difficult to determine from the plot as the results are plotted on a log scale, but the given estimates were 19.27 for diabetics down to 11.23 for non-diabetics (per 1000py). The non-PEH model shows a higher excess mortality after 6 months for diabetics than in the PEH model, in comparison the excess mortality rate does drop below thirty excess deaths (per 1000 person years) unlike the PEH model. By contrast the non diabetics have a lower number of deaths than was estimated in the proportional model. It would seem that forcing a proportionality assumption onto diabetes has resulted in the under-estimation of one group and the over estimation of the other, and as such proportionality is not a fair assumption, as strongly supported by the likelihood ratio test.

The final piecewise model includes both diabetes and glucose, however there is no interaction fitted between diabetes and glucose. In the short term analysis the effect of diabetes was not statistically important once glucose was added to the model. There are three possible likelihood ratio tests that can be performed on this model, the first would be to see if adding diabetes to the glucose only models has added anything ($p = 0.0448$ for PEH and $p = 0.0484$ for non-PEH), the second to see if adding glucose to the diabetes only model has added information ($p < 0.001$ for PEH and $p < 0.001$ for non-PEH) and finally to assess if glucose and diabetes are time dependent. This final test gave a value of $\chi^2_{24} = 33.37$ where $p = 0.0965$ which is not strictly statistically significant, but the large increase in degrees of freedom should be noted. It should also be noted that even by moving these split times by a relatively small fraction can cause problems in these models as there are not enough events to split the timescale further when investigating more than one time dependent effect.

A stratified Cox model (stratified using the split-time intervals) was fitted as a comparison to the relative survival model and the results are shown, with 95% confidence intervals, in

table 9.7. Three models are fitted, including glucose only, diabetes only and glucose and diabetes together, labelled as *G and D* in the table. The top half shows the estimates from a piecewise model proposed by Dickman (115) and the bottom half from a Cox model. There are two columns of AIC (123) values the first is from a PEH model and the second is from a non-PEH model while all covariate estimates are from a PEH model.

The glucose and diabetes model would be selected for the PEH and only glucose in the non-PEH model using the AIC, with the PEH model having the lowest AIC overall (13985.9). If using a stratified Cox model in a standard all-cause analysis then a glucose only model would be selected under a proportional and non-proportional approach.

The effect sizes are larger in the relative survival models, particularly quartile four, and the confidence limits do not overlap in comparison to the Cox model. This result is feasible as glucose is not so important for people with diseases other than heart disease and thus shows a clear advantage of using relative survival. Now the analysis is looking at deaths in excess of other diseases it is potentially higher because glucose increases the risk of death after MI but does not increase the risk of death from other causes, for example glucose is not influential on cancer mortality rates.

9.4.1 Restricted non-PEH

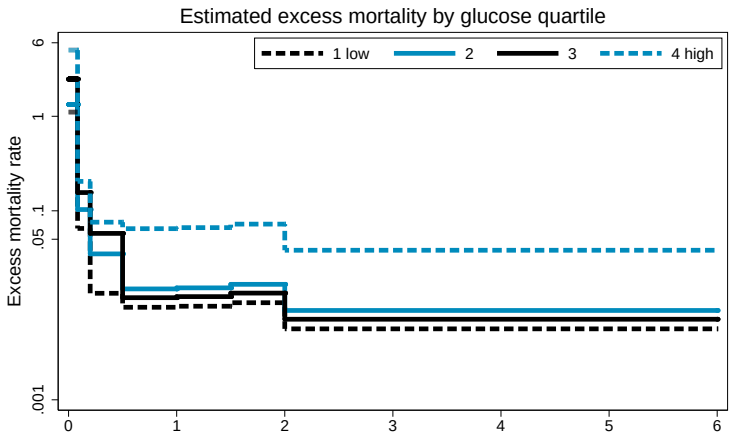


Figure 9.7: Plot of the excess mortality rate by glucose quartile assuming non-PEH for the first 3 time intervals

It is possible to reduce the numbers of parameters being estimated by assuming that only the first three time intervals interact with glucose. This is achieved quite easily in practice. Using a likelihood ratio test a value of $p = 0.4751$ was given when compared against a full non-PEH glucose model, suggesting that this assumption is appropriate. When assessed against a PEH model the likelihood ratio test gave a value of $p = 0.0939$, which is not quite statistically significant.

Time Interval (months)	Glucose quartile		
	Q2	Q3	Q4
0-1	1.203	2.233	4.526
	(0.89 , 1.62)	(1.70 , 2.93)	(3.52 , 5.82)
1-2.4	1.585	2.398	3.147
	0.50 , 5.05)	(0.80 , 7.16)	(1.08 , 9.17)
2.4-6	2.611	4.303	5.658
	(0.32 , 21.28)	(0.57 , 32.61)	(0.78 , 41.19)
6-72	1.565	1.264	6.789
	0.66 , 3.71)	(0.48 , 3.35)	(3.18 , 14.48)

Table 9.8: Estimates of the mortality ratios for the restricted time dependent glucose quartiles effects

Figure 9.7 shows the estimated excess hazard rates using the restricted non-PEH model described above. The effect of glucose is modelled proportionally after 6-months and non proportionally up to this point. The model allows the freedom of the non-PEH model for the first six months, which is where the majority of the information is. Table 9.8 shows the estimates of the excess hazard rate ratios from this model. Here it shows how the estimates are allowed to be different during the first three intervals but are then constant for the remaining follow-up.

9.5 Relative survival fractional polynomial modelling

In the short term paper multivariate fractional polynomial (MFP) models were used to estimate odds ratios using logistic regression, which does not take censoring into account as the follow-up was short. For long term follow-up MFPs can be fitted using relative survival, as described is section 4.6, which accounts for censoring. As fairly fine splits are being used it is not feasible to adjust for many covariates as there will be many zeros in sub-groups, especially when investigating time-dependent effects. Grouped level data will be used as individual level data is computationally intensive taking a long time to converge, also discussed in section 4.6. Functions from fractional polynomials are generally more flexible than more standard polynomial models which, as shown in section 4.5, do not always fit well for these data.

9.5.1 MFP modelling

Glucose and diabetes are categorical covariates and as such are modelled using dummy variables, i.e. they are not transformed. Time however is modelled as continuous and for each of the models was transformed with powers of $(-.5, -.5)$, which were selected using the MFP algorithm described in section 4.6.2. The MFP PEH curves are shown in figure 9.8. The top row of the figure shows the two models which include only a single covariate,

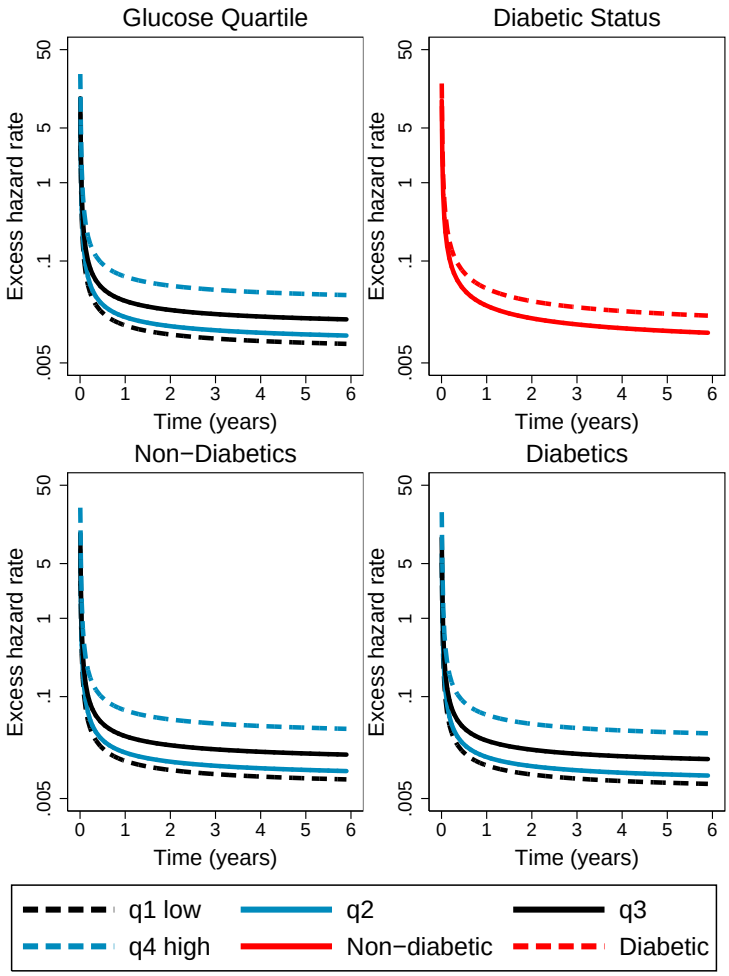


Figure 9.8: Plot of the excess mortality rate by glucose quartile and diabetic status for the PEH model

glucose on the left hand side and diabetes on the right hand side. The bottom row shows the model with both diabetes and glucose in it with non-diabetics on the left and diabetics on the right with the four glucose quartile groups excess mortality rates shown.

The model that includes diabetes only levels out at a high excess mortality rate, even the non-diabetics have a similarly high excess mortality when compared directly to the glucose plots which show that only glucose quartile 4 has has a similarly high mortality rate to the diabetes group. The estimates from the model with glucose on its own is very similar to the estimates with and without diabetes shown on the bottom row. Quartile 4 is slightly separated out from the other three quartiles.

Table 9.9 shown later on confirms the similarity between the glucose estimates with and without diabetes in the model. A Likelihood ratio test of glucose only compared to glucose and diabetes gave a value of $p = 0.1169$ suggesting that the addition of diabetes is non-significant and that the results of the short term analysis apply to the long term in this

PEH model.

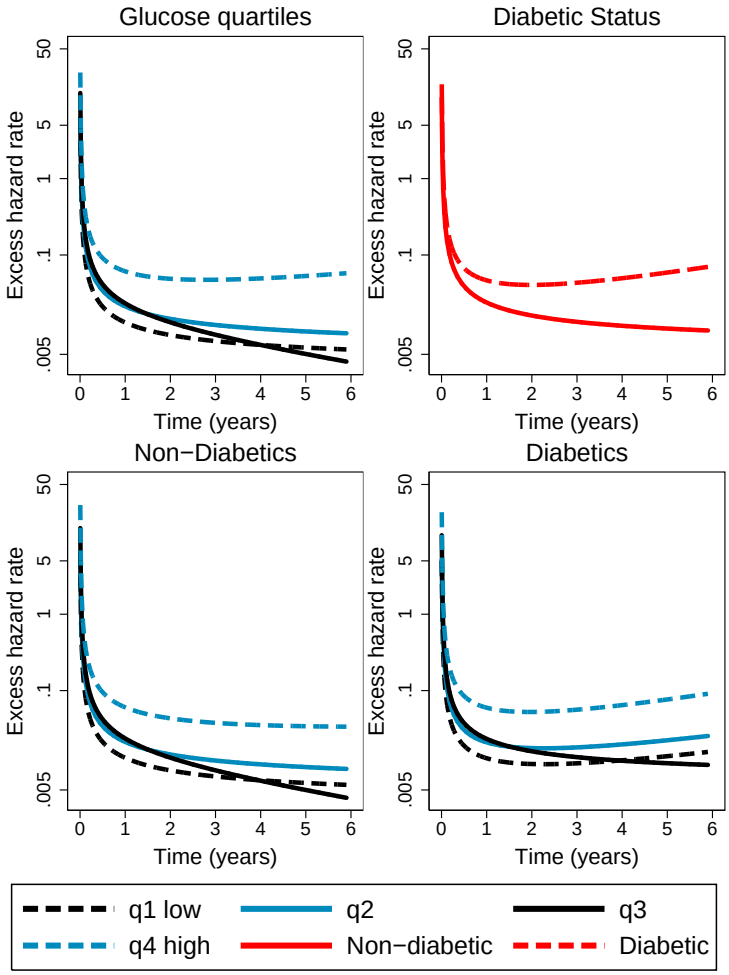


Figure 9.9: Plot of the excess mortality rate by glucose quartile and diabetic status for the non-PEH model

Figure 9.9 shows the results from a non-PEH model and as such the model including both diabetes and glucose forces time dependent effects on both covariates. The figure is set-up as for the PEH model. Glucose quartile three is shown to continually decrease in all three models to a point below even quartile one. The diabetic only model shows that the diabetic group increases after the initial drop in excess mortality. After a year following AMI these patients continue to have a high level of excess mortality, which also continues to rise over the six year follow-up. Similarly the bottom-right plot showing glucose quartiles for diabetics shows that quartiles one, two and four have a steadily increasing excess mortality rather than the non-diabetic group where all four quartiles steadily decrease. The lower three quartiles give fairly similar estimates with quartile four separated from these.

The differences observed between the bottom two plots of figure 9.9 suggest that, for non-PEH models, diabetes may hold statistically important information. This is best assessed using likelihood ratio tests and AICs and this information along with the estimates from

Model	Glucose quartile			Diabetes	AIC	AIC	LR test
	Q2	Q3	Q4		PEH	non-PEH	p-value
Glucose	1.277	2.054	4.197	-	1116.28	1106.45	0.0012
	(0.99 , 1.65)	(1.62 , 2.61)	(3.37 , 5.22)	-			
Diabetes	-	-	-	1.655	1353.67	1341.11	0.0001
	-	-	-	(1.41 , 1.94)			
G and D	1.277	2.073	4.425	0.877	1116.00	1101.20	0.0001
	(0.99 , 1.65)	(1.63 , 2.63)	(3.52 , 5.56)	(0.74 , 1.04)			

Table 9.9: Estimates for glucose quartiles and diabetes from MFP models, Likelihood Ratio test is for time dependent effects

a proportional model are shown in table 9.9. The likelihood ratio test comparing the glucose and diabetes non-PEH model to the glucose only model gave a value of $p = 0.0095$, suggesting that diabetes should be included in the model. This has strong implications that diabetes is important and also that diabetes is involved in an interaction with time.

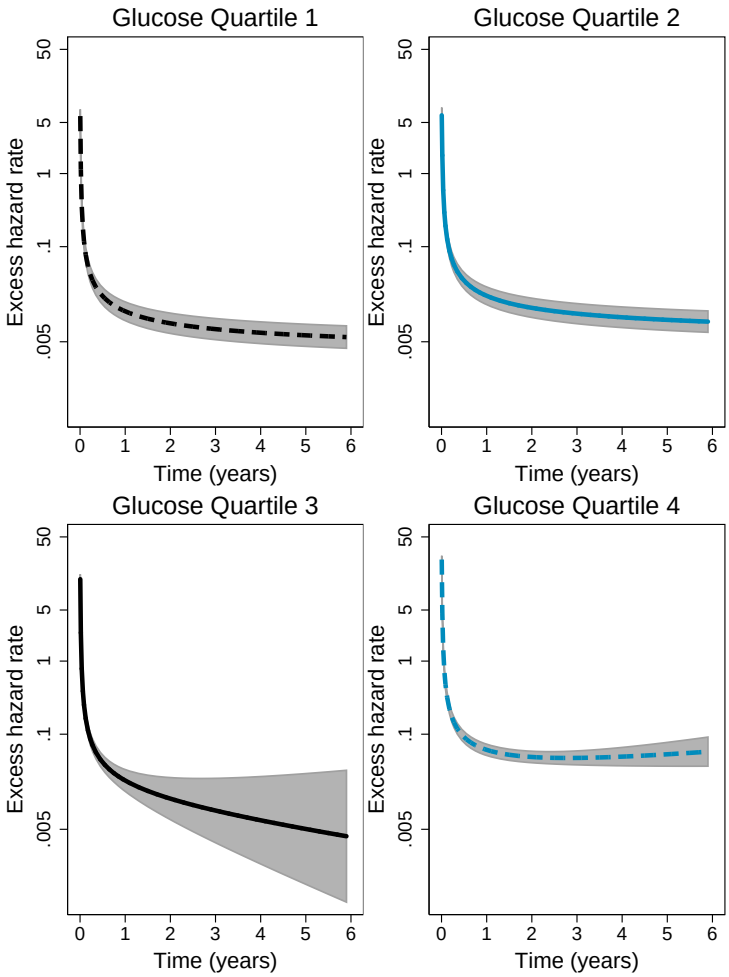


Figure 9.10: Plot of the excess mortality rate by glucose quartile (top-left of figure 9.9) unadjusted for diabetes

Using figure 9.9 it is possible to split each of the four plots into a single plot for each

covariate level to allow the confidence intervals to be plotted clearly. A plot of the excess hazard rate with 95% confidence intervals by glucose quartile using fractional polynomials are shown in figure 9.10. There is a large amount of uncertainty which is likely due to a low number of events in later time periods. These plots highlight that glucose quartile 3 is decreasing over time to have the lowest overall excess mortality, however the intervals are very wide. The width of the intervals may be due to the fractional polynomial function, see section 4.7.2. The likelihood ratio test is highly significant with a value of $p = 0.0012$ (compared to a PEH model), as shown in table 9.9.

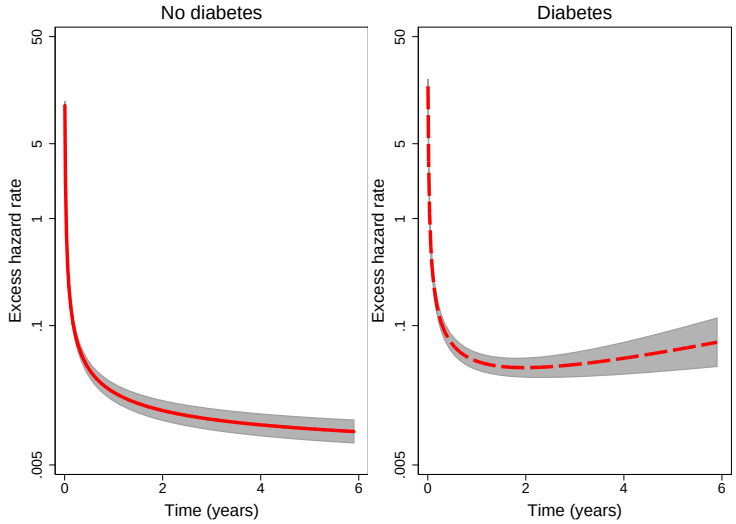


Figure 9.11: Plot of the excess mortality rate by diabetes (top-right of figure 9.9) unadjusted for glucose

A plot of the excess hazard rates and their confidence intervals by diabetic status using fractional polynomials are shown in figure 9.11. This shows that there is less uncertainty associated in these two groups as the confidence intervals are fairly narrow, it also shows that the two groups are statistically significant from each other as their confidence limits appear well separated and would not cross. Both of these plots appear to be affected by time as the non-diabetics excess hazard reduces over time whereas a diabetic patient's excess hazard increases from about one year.

A plot of the excess hazard rate and their confidence intervals by glucose quartile in non-diabetics (left-hand side) and in diabetics (right-hand side) using fractional polynomials are shown in figure 9.12. This plot shows the differences between each quartile group dependent on whether the subject is labelled as diabetic or not. The main differences are in glucose quartiles 2 and 4 whose curves seem to change direction, the diabetics have an increasing excess mortality and the non-diabetics a decreasing excess mortality. There is no interaction between glucose and diabetes but the difference between the two plots shows that diabetes is having a strong effect.

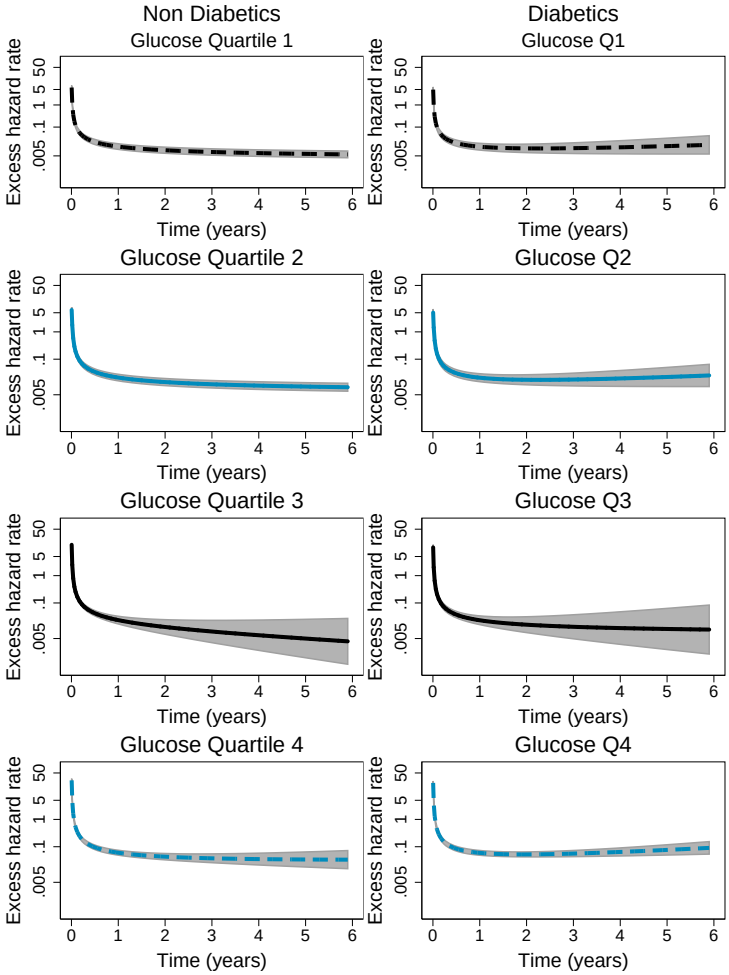


Figure 9.12: Plot of the excess mortality rate by glucose quartile and diabetes (bottom-row of figure 9.9) adjusted for glucose or diabetes

9.5.2 Excess hazard rate ratios

A plot of the excess hazard rate ratio for the glucose only and diabetes only models are shown in figure 9.13, note that these are not on a log-scale. Glucose quartiles 3 and 4 show more uncertainty later on as the amount of data decreases, as does diabetes. Glucose 4 shows a large amount of uncertainty after 4 years. Also the confidence intervals of both glucose quartile 4 and diabetes do not cross 1 showing a statistically significant difference to their comparative groups.

A plot of the excess hazard rate ratio for the models, which are adjusted for both glucose and diabetes, are shown in figure 9.14. The glucose quartile 4 effect appears stable around 4 but has a high level of uncertainty towards the end of follow-up where the lower bound of the 95% confidence interval is moving towards 1. Adjustment for diabetes has reduced the effect size of quartile four and removed the apparant time dependent interaction as

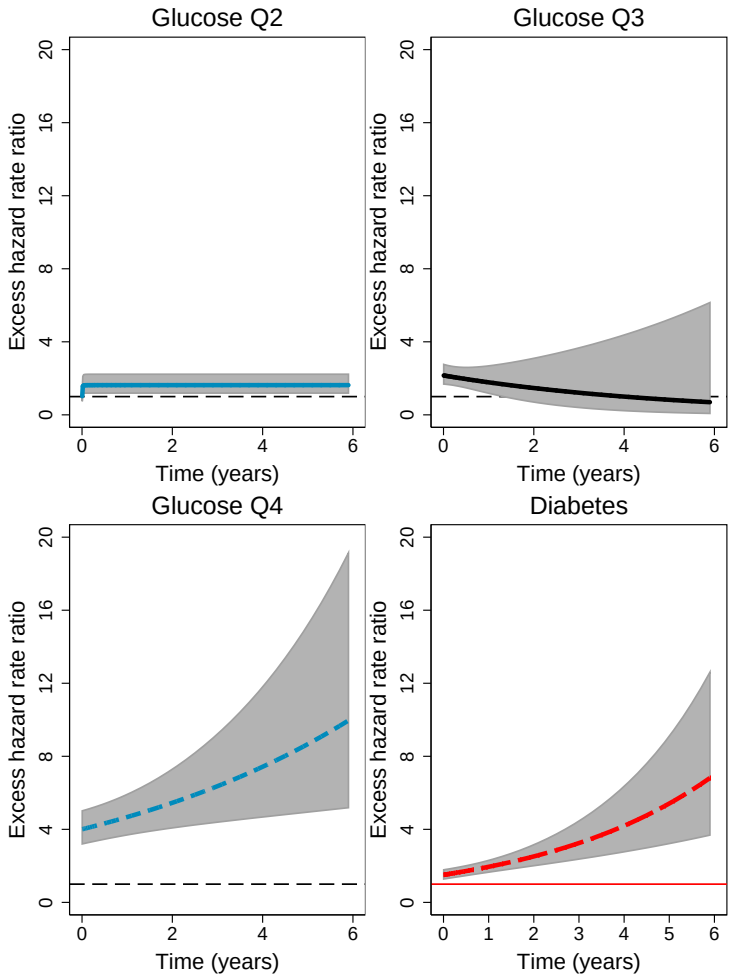


Figure 9.13: Plot of the excess mortality rate ratio by glucose quartile and diabetic status for unadjusted models

the effect shown previously in figure 9.13 was increasing over time in a non-linear way.

The estimated excess hazard rate ratio of diabetes is around one for the first six months to a year, and then the estimate increases, however once confidence intervals are added they overlap 1 for the majority of the follow-up meaning this is not statistically significant after an adjustment for glucose was made. After around three and a half years the effect of diabetes becomes significant as the confidence interval does not contain 1 and is the reason a significant value for the likelihood ratio test was obtained ($p = 0.0095$) when comparing to a glucose only non-PEH model. Diabetes has been found to be important in long term follow-up. Note that for these plots diabetes is shown for the lowest glucose quartile group.

In order to assess the sensitivity of these models a different selection procedure was assessed using a standard 5% α level of assessment. Using this method the same model was selected in all but two cases, in comparison to using the AIC. These were models that included

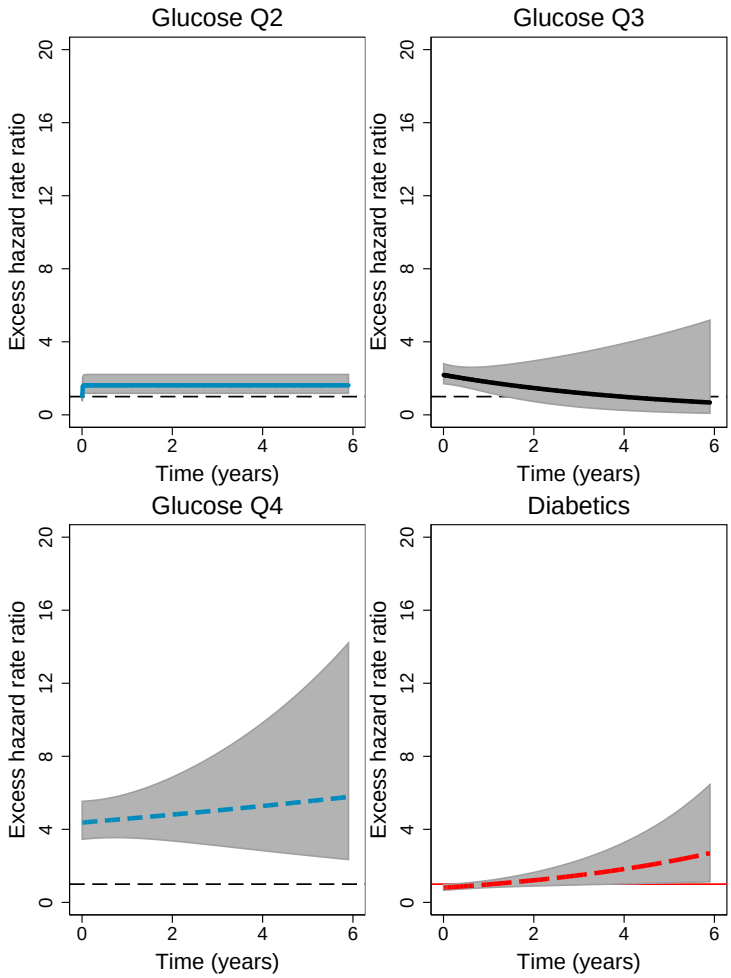


Figure 9.14: Plot of the excess mortality rate ratio by glucose quartile and diabetic status for adjusted models

glucose quartiles as time-dependent, so the AIC appears fairly robust. Two FP3 models were assessed, the first included all baseline and time dependent effects and the second allowed only the baseline effects to be fitted using an FP3 model. An FP2 model was selected from both approaches so FP3 models were not needed.

9.5.3 Interaction between Diabetes and Glucose

In this section an investigation into whether a diabetic or non-diabetic patient has a different mortality rate depending on their glucose group is carried out. These models were fitted by creating an interaction term between glucose quartiles and diabetes.

A plot of the excess hazard rates for the PEH model by glucose quartile and diabetes are shown in figure 9.15. The covariate estimates have little meaning on their own, which is why the results are only presented graphically. The reference category is glucose quartile 1

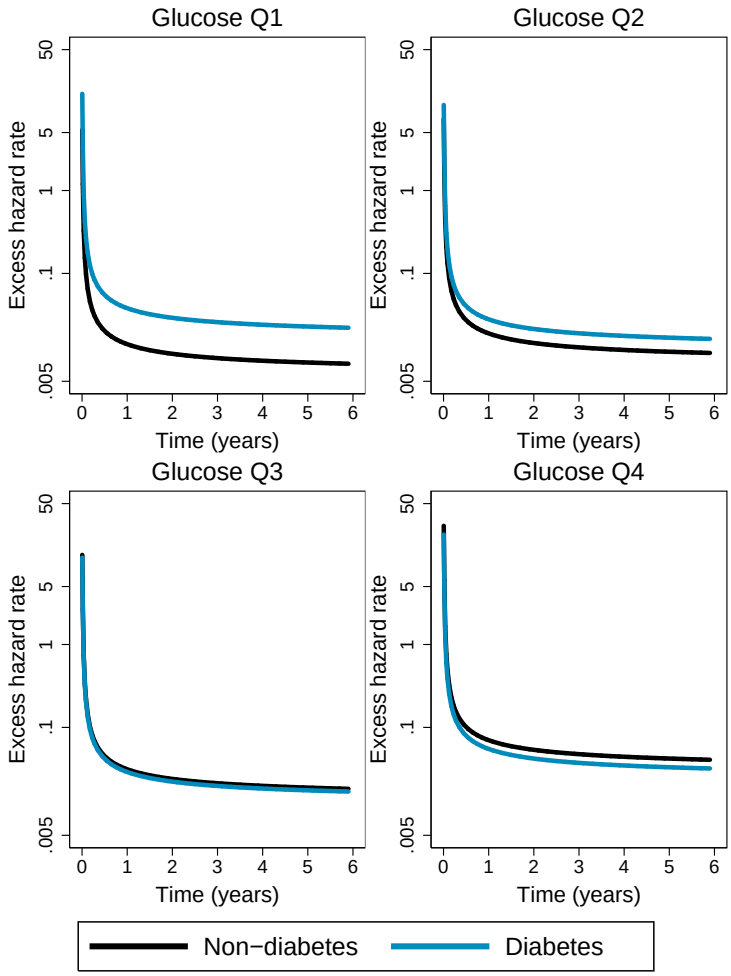


Figure 9.15: Plot of the excess mortality rate by glucose quartile and diabetic status in a PEH model with an interaction

and so a significant result for an interaction between glucose quartile 4 and diabetes would suggest that the effect of glucose quartile 4 in diabetics is significantly different to that of diabetics in glucose quartile 1. The plot shows the four glucose quartiles separately but with diabetics and non-diabetics shown within each quartile. It is possible to assess the significance of the interactions on an individual level and this gives values of $p = 0.174$ for glucose quartile 2, $p = 0.005$ for glucose quartile 3, and $p < 0.0001$ for glucose quartile 4. However it is more suitable to perform a global test and the likelihood ratio test gave a value of $p = 0.0041$ suggesting that the interactions were significant and improve upon the PEH model without the interaction.

The figure is very interesting in that it shows the diabetes patients in glucose quartiles 1 and 2 have a higher excess mortality rate than non-diabetics, whereas in glucose quartile three they are near identical. The interesting part is that glucose quartile 4 shows that non-diabetics have a slightly higher mortality rate. Non-diabetics also have a higher mortality rate at the beginning of glucose quartile 4. This may possibly be indicative of

clinical practice which may treat a patient labelled as diabetic with insulin but a patient not labelled as diabetic may not be given insulin. This also supports the short clinical papers findings which indicate that admission glucose is a very important measure, maybe more so than diabetes.

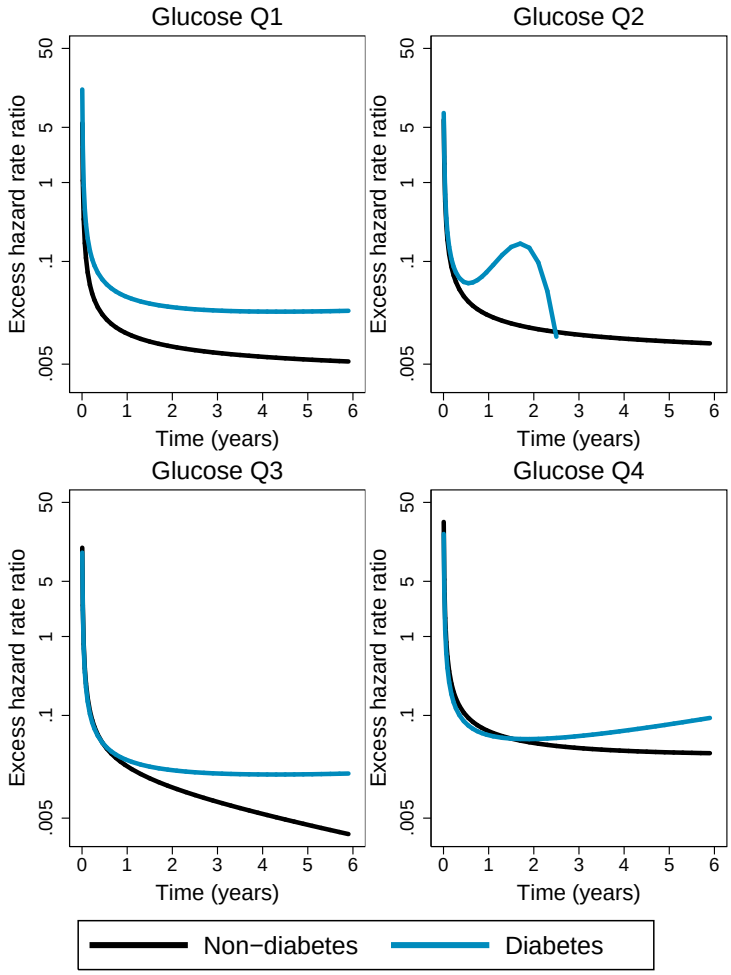


Figure 9.16: Plot of the excess mortality rate by glucose quartile and diabetic status with an interaction

With the assumption of PEH not being valid in previous models it was necessary to investigate time dependent effects. The model fitted is very complex and allows glucose, diabetes and the interaction to vary by time (shown in figure 9.16). This model was compared, using a likelihood ratio test to the PEH model with the interaction and gave a value of $p = 0.0001$.

Glucose quartile 4 shows that non-diabetics have a slightly higher excess mortality rate at the beginning of the model, but now that the model is allowed to vary by time the diabetes group is shown to have a much higher mortality rate than non-diabetics. Glucose quartile 2 has not fitted correctly as there is little data and the plot only estimates the first 2.5 years, so is of no real use. In glucose quartile 3 the non-diabetic and diabetic

group are very different as the diabetics have a near constant excess mortality rate after a year while non-diabetics excess mortality continues to decrease. Glucose quartile 1 shows similar estimates to the PEH model. It should also be noted that these models do not seem reasonable as, although the global test is significant, the individual effects are non-significant, quartile two does not fit correctly and estimates are overly complex.

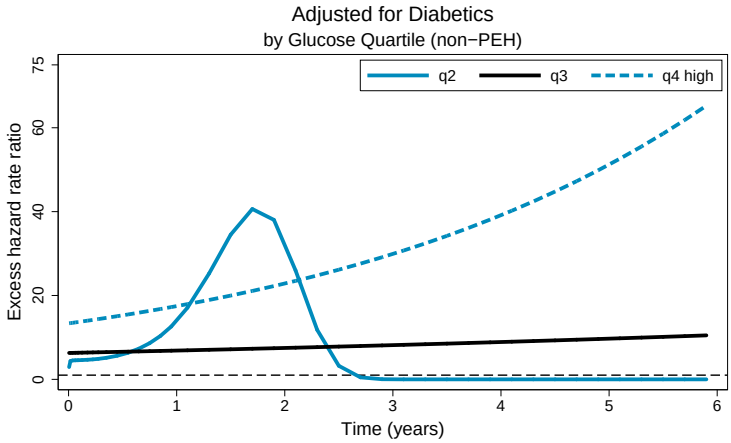


Figure 9.17: Plot of the excess hazard rate ratios by glucose quartile and diabetic status with an interaction

A plot of the excess hazard rate ratios are shown in figure 9.17. The first point of note is that the model is overly complex and this has resulted in a fairly poor fit for quartile 2 in diabetics. Quartile 3 is increasing over time and this affect appears near linear. The confidence intervals, which are not shown here, show that there is huge uncertainty past 4 years for all quartile groups and are not shown because the scale becomes too wide for the main effects to be visibly clear. The glucose quartile 4 excess mortality rate ratio is shown to be increasing rapidly over time starting at around 15 and increasing to around 70 in six years. This suggests that diabetics with a high level of blood glucose at admission after 6-years of follow-up are at approximately seventy times the risk of excess mortality compared to diabetics with low admission blood glucose. This estimate seems very high and does have very wide confidence intervals.

The analyses performed using MFPs has shown that there is a possibility of a significant interaction between diabetes and glucose, but the main finding thus far is that diabetes has significant implications on a STEMI patient's long-term survival. Due to these models having issues and because the method forces the use of split-time data the approach was changed to allow the modelling of multiple covariates and continuous covariates using the spline based models (chapter 5).

Method	Model	Glucose quartile			Diabetes
		Q2	Q3	Q4	
MFP	Glucose	1.277	2.054	4.197	-
		(0.99 , 1.65)	(1.62 , 2.61)	(3.37 , 5.22)	-
	Diabetes	-	-	-	1.655
		-	-	-	(1.41 , 1.94)
	G and D	1.277	2.073	4.425	0.877
		(0.99 , 1.65	1.63 , 2.63)	(3.52 , 5.56)	(4.41 , 6.22)
Spline	Glucose	1.248	2.098	4.373	-
		(0.95 , 1.63)	(1.63 , 2.69)	(3.48 , 5.49)	-
	Diabetes	-	-	-	1.665
		-	-	-	(1.41 , 1.96)
	G and D	1.250	2.121	4.652	0.856
		(0.96 , 1.64)	(1.65 , 2.72)	(3.67 , 5.90)	(4.41 , 6.33)

Table 9.10: Estimates for glucose quartiles and diabetes from MFP and Spline models assuming PEH

9.6 Flexible Parametric Modelling

9.6.1 Introduction

In order to investigate various other covariates the use of fractional polynomials became unfeasible due to using split-time grouped data where all continuous covariates needed to be categorised, which is never a recommended practice (140). It is possible to use individual level data but these models take many hours to converge and have convergence problems as found in chapter 4. In order to assess other covariates, without running out of data when using split times, flexible parametric spline based models (3) were adopted as were developed in chapter 5.

An initial investigation looked at using three internal knots (4df) to assess the results in comparison with those from the MFP models and these results are shown in table 9.10. Three PEH models are fitted, the first has glucose only, the second diabetes only and the last has glucose and diabetes in the model. The estimates and CIs from the spline based models are very similar to the MFP models, e.g. in a model with only diabetes an estimate of the excess hazard rate ratio of 1.655 with a 95% CI of $(1.41, 1.94)$ and $1.665(1.41, 1.96)$ are given for an MFP model and a splines model respectively.

The next task was to investigate the number of degrees of freedom to use when modelling splines, so a sensitivity analysis was performed using a PEH and non-PEH model with both diabetes and glucose, where the non-PEH model allowed both diabetes and glucose quartiles to vary over time. The PEH estimates are shown in table 9.11 with their AIC values alongside the AIC values obtained from the non-PEH model where both glucose and diabetes are time dependent. On the left hand side of the table are the degrees of freedom used in determining the flexibility of the model as shown previously in section

Model	Glucose quartile			Diabetes	AIC[1]	AIC[2]
df	Q2	Q3	Q4			
2	1.270	2.171	4.599	0.832	3483.0	3478.9
	(0.97 , 1.66)	(1.69 , 2.78)	(3.62 , 5.84)	(0.69 , 1.00)		
3	1.241	2.106	4.674	0.862	3120.6	3123.6
	(0.95 , 1.62)	(1.64 , 2.70)	(3.69 , 5.93)	(0.72 , 1.03)		
4	1.250	2.121	4.652	0.856	3047.9	3048.7
	(0.96 , 1.64)	(1.65 , 2.72)	(3.67 , 5.90)	(0.71 , 1.02)		
5	1.255	2.130	4.656	0.859	2944.8	2947.8
	(0.96 , 1.64)	(1.66 , 2.74)	(3.67 , 5.91)	(0.72 , 1.03)		
6	1.249	2.122	4.650	0.859	2904.6	error
	(0.95 , 1.63)	(1.65 , 2.73)	(3.66 , 5.90)	(0.72 , 1.03)		

Table 9.11: Estimates for glucose quartiles and diabetes with AICs for the simple proportional[1] and non-proportional[2] spline models

5.2.4.

Table 9.11 shows that all estimates are similar for glucose and diabetes across all models. There is little change in the width of the confidence intervals and as such it is safe to assume that the models are fairly robust to changes in knot frequency and location. The lowest AIC in the PEH models, shown in AIC[1], is for the 6df model. As is shown in the 6df model AIC[2] for the non-PEH models, there was a problem with model convergence, possibly due to having too many knots with not enough data to support time dependent effects. In both the PEH and non-PEH models the AIC lowers as the degrees of freedom increases. The PEH models have lower AICs than the non-PEH models (excluding the 2df model). The 5df model will be used for the following analyses unless otherwise stated.

Figure 9.18 shows the estimated excess hazard rates with various degrees of freedom for glucose quartile one by diabetic status. This figure shows that for all but the 2df model the fitted baseline hazard is very similar for all models, it is worth noting again that the 6df model is not shown for the non-PEH models as it did not converge. The models fitted both diabetes and glucose as proportional or time dependent to assess the fit, but it may be more appropriate to assess them independently.

The reason for the non-PEH 6df non-convergence may be due to the lack of data in the subgroups at the start of follow-up, where more knots are fitted. However, the lowest frequency of patient death in any of the time periods (i.e. between knots) for either of the covariates was 17, which is enough for convergence. The standard knot placements are at the 17, 33, 50, 67 and 83 percentiles so as an experiment these were changed to 20, 40, 50, 60 and 80. Using these new knot locations the model converged quickly giving an AIC of 2947.739 which when compared to the AICs in table 9.11 is almost identical to the 5df non-PEH model. This shows how it is possible to work around problematic splines and how non-convergence issues can be solved by altering knot placement when using these models.

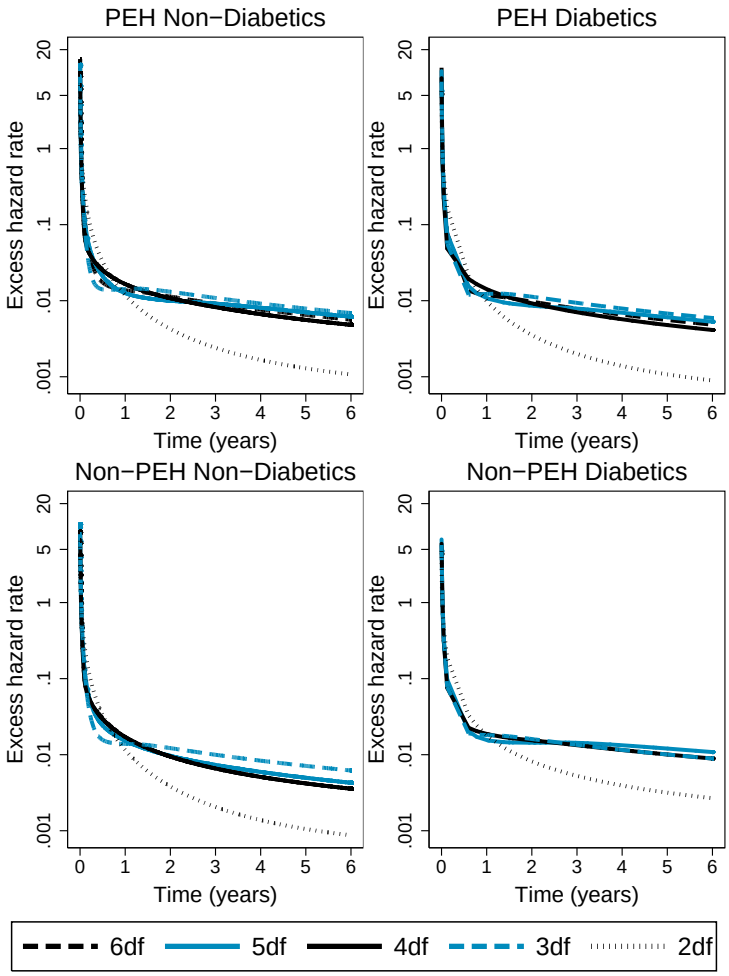


Figure 9.18: Plot of the excess hazard rates in glucose quartile 1 by diabetic status using spline based models with varying degrees of freedom

9.6.2 Glucose as time dependent

As discussed in section 5.5.1 fitting more than one time dependent covariate can make interpretation difficult. The following analyses look at glucose as it holds more interest, and is therefore allowed to vary over the spline terms. Figure 9.19 shows the excess hazard rates for non-diabetics from four spline based models, as indicated in the figure.

The 2df model did not fit sensibly for glucose quartile 3 and tends to zero after two and a half years. Of these four models the 5df model has the lowest AIC (2947.083). A likelihood ratio test comparing these models to a PEH model gave the following results: $5df$ $p = 0.0236$, $4df$ $p = 0.0496$, $3df$ $p = 0.0861$ and $2df$ $p = 0.0419$. These values suggest that glucose is time-dependent. Glucose quartile four is different to the other quartiles in all of the models, with the other quartiles more similar. However in the 4df model, quartile three has the lowest excess mortality rate after two years. This was also shown

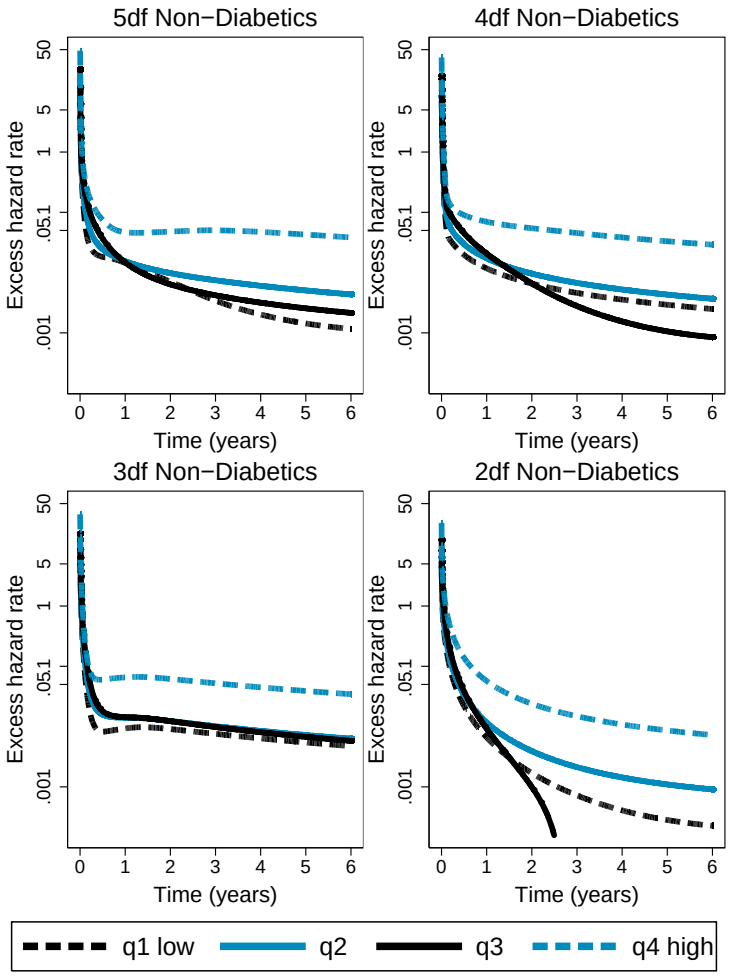


Figure 9.19: Plot of the excess hazard rates by glucose quartile in non-diabetics using spline based models

in the MFP non-PEH plots shown in figure 9.9.

The plot shown in figure 9.20 shows the excess hazard rate ratios. The 5df model, which has the lowest AIC, estimates high ratios after two to three years. In comparison, the 4df and 3df models have a fairly constant ratio of much less than 20. The 2df model has issues estimating the third glucose quartile. It is important to note that the 2df model is only fitted with one internal knot and is therefore likely to be a poor fit for these data. The potential issue with the 5df model could be over-fitting due to over parameterisation for time dependent effects. This model could be extended to have fewer knots for the time dependent effects similarly to the models shown in section 9.4.1, but it is difficult to adapt the program (described in chapter 6) used for these models.

The large estimates of the excess hazard rate ratios in the 5df model are of no great concern due to a large amount of uncertainty associated with it at the end of follow-up. The model was restricted to a 5-year follow-up and the excess hazard rates are shown in figure 9.21.

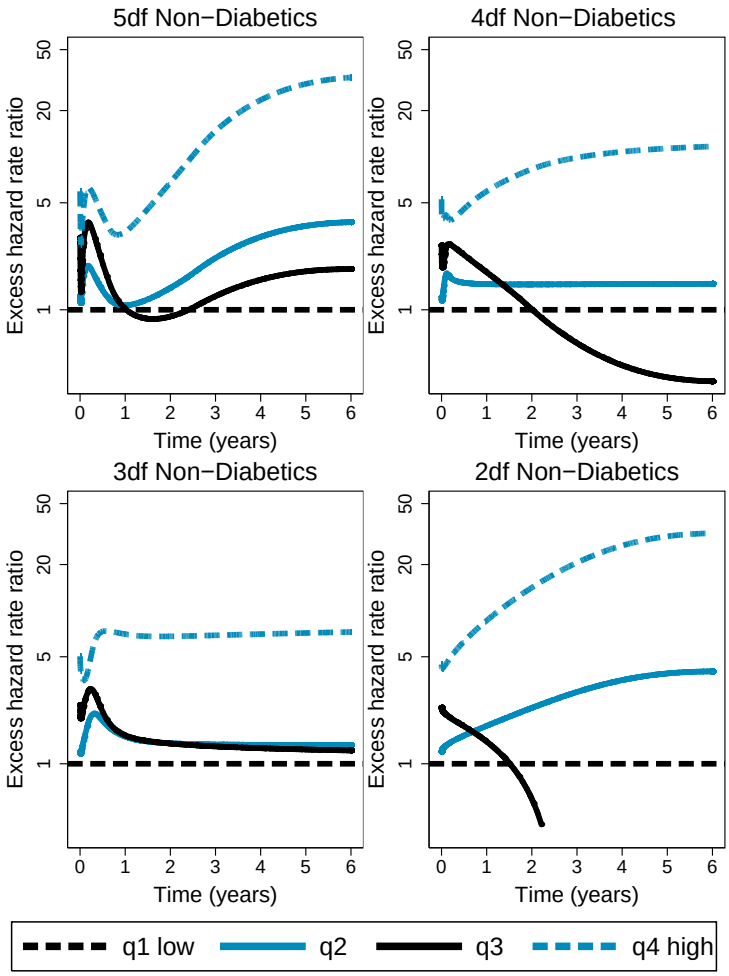


Figure 9.20: Plot of the excess hazard rate ratios by glucose quartile in non-diabetics using spline based models

The 2df model did not fit sensible results for glucose quartile 3 as before. The lowest AIC was for the 5df model (2506.912) and likelihood ratio tests were performed comparing these models with PEH models with glucose and diabetes in them. All p-values were found to be non-significant($p > 0.1$), except the 5df model($p = 0.0254$). Glucose quartiles one to three are grouped together in the 5df and 3df models but the 4df model still estimates quartile 3 as the lowest excess mortality rate, suggesting that the problem is not in the lack of data in the final year. Glucose quartile 4 is the highest in all models and is estimated at a similar level in all the models.

Figure 9.22 shows the excess hazard rate ratios from a five year follow-up. The problem with the six year follow-up was with glucose quartile 4 in that the excess hazard rate ratio seemed very high in the 5df model, although there was a very wide confidence interval around the estimate. Glucose quartile 4 is now lower and more in line with the other model estimates. However this model is now suggesting that glucose quartiles 2 and 3 both have preventative effects in long term survival after 2 years. All of the models start with a

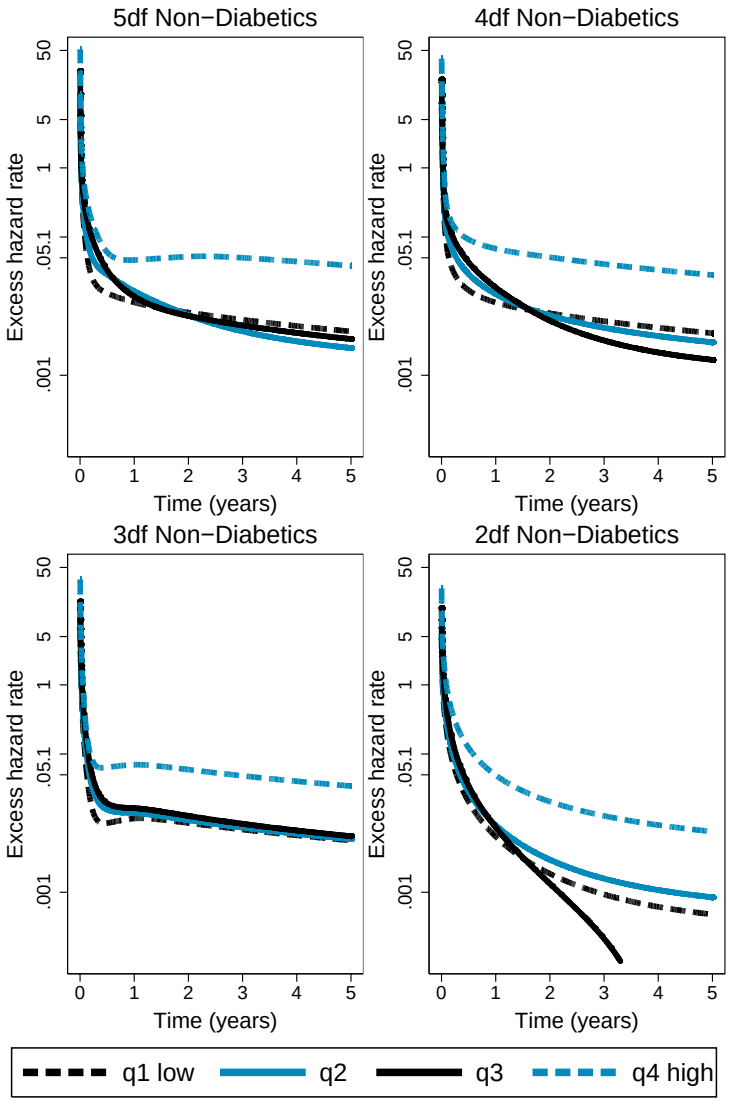


Figure 9.21: Plot of the excess hazard rates by glucose quartile and diabetic status using spline based models

strange hump at the beginning as seen in the six year follow-up. The 3df model shows glucose quartiles 2 and 3 as similar and slightly above one which may be more feasible.

Glucose	Diabetic		
Quartile	No	Yes	Total
Glucose Q1	46	1	47
Glucose Q2	99	0	99
Glucose Q3	54	7	61
Glucose Q4	46	35	81
Total	245	43	288

Table 9.12: No of deaths between the fifth and sixth year of follow-up

The impact on glucose quartile 4 in the 5df model by reducing follow-up to 5 years was

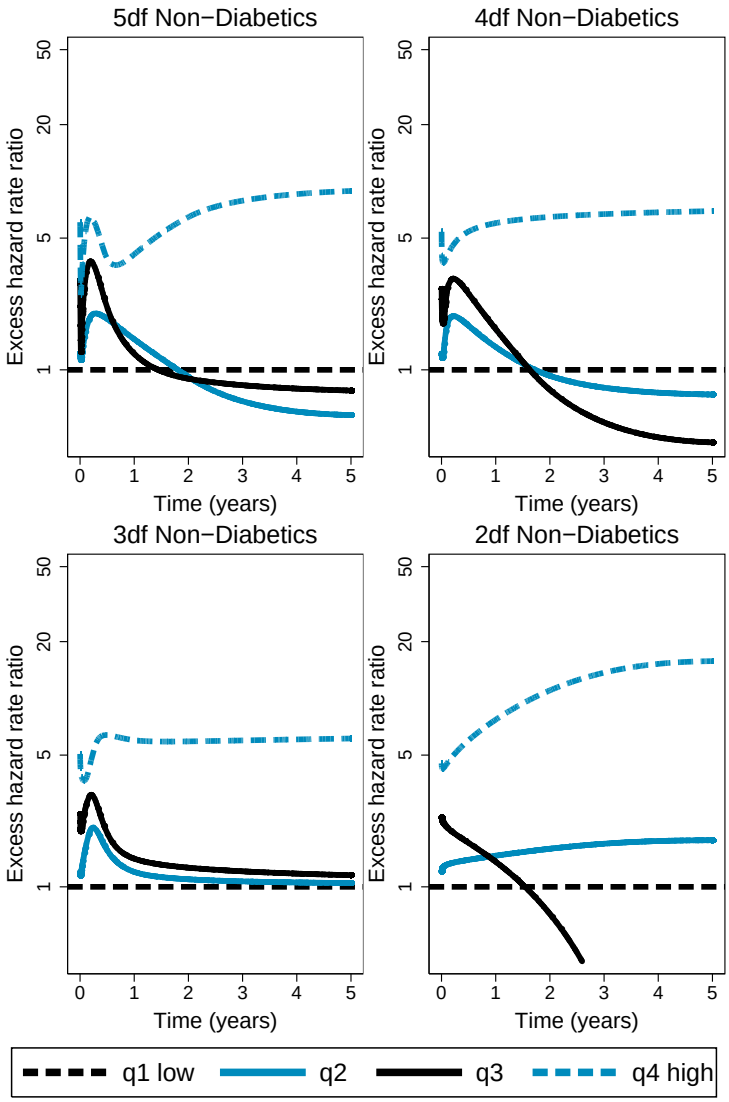


Figure 9.22: Plot of the excess hazard rate ratios by glucose quartile for a 5-year follow-up

quite large. The number of deaths by glucose quartile between the fifth and sixth year of follow-up are shown in table 9.12 in order to determine if this is the reason why. There were 288 deaths in the final year of follow-up, with the most being in glucose quartile 2 and 4. The main finding is that glucose quartile 4 has 81 deaths in the final year of follow-up with a high percentage of these being in diabetics unlike the other three glucose quartiles.

Figure 9.23 shows the results from the 5df plot shown in the top left of figure 9.22 but now with 95% confidence intervals. This was to determine if the large uncertainty found in the previous six year follow-up was due to the very small number of events seen in table 9.12. The figure shows that the intervals are still fairly wide. There is fairly strong evidence that glucose quartile four is different. There is not enough evidence to suggest

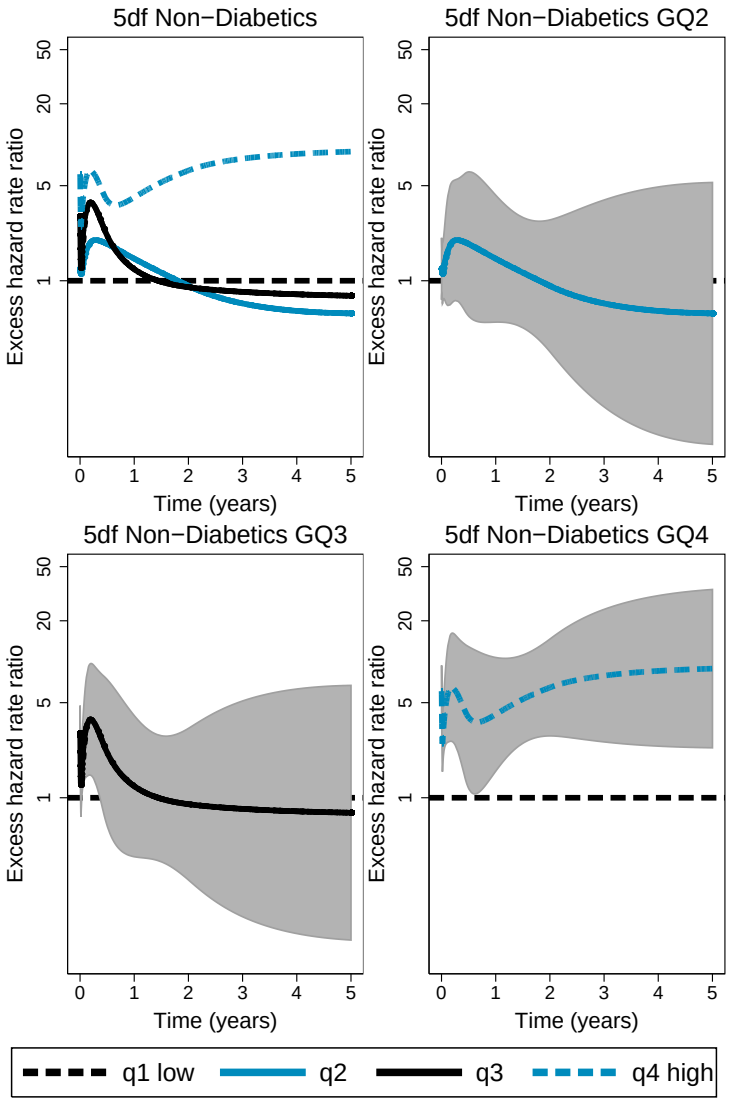


Figure 9.23: Plot of the excess hazard rates by glucose quartile and diabetic status using 5df splines for a 5-year follow-up and with 95% confidence intervals

that the models are superior when restricting the timescale to five years and as quartiles 1-3 remain largely unchanged, and given they have the least number of events, it was not deemed necessary to restrict the timescale for further analyses.

9.6.3 Adding covariates

The main aim of using the flexible parametric splines for relative survival was to obtain estimates of glucose and diabetes after adjustment for other covariates and so the models when stated as adjusted now adjust for the following covariates, which were the same covariates adjusted for in the short term analysis:

Method	Model	Glucose quartile			Diabetes
		Q2	Q3	Q4	
All Cause	Glucose	1.221	1.526	2.279	-
		(1.02 , 1.46)	(1.28 , 1.81)	(1.94 , 2.67)	-
	Diabetes	-	-	-	1.365
		-	-	-	(1.20 , 1.55)
	G and D	1.223	1.531	2.319	0.959
		(1.03 , 1.46)	(1.29 , 1.82)	(1.96 , 2.75)	(0.83 , 1.11)
Relative Survival	Glucose	1.313	2.082	3.232	-
		(1.00 , 1.72)	(1.62 , 2.67)	(2.57 , 4.06)	-
	Diabetes	-	-	-	1.372
		-	-	-	(1.17 , 1.62)
	G and D	1.328	2.122	3.482	0.843
		(1.02 , 1.74)	(1.65 , 2.73)	(2.73 , 4.43)	(0.70 , 1.01

Table 9.13: Estimates for glucose quartiles and diabetes from all-cause (using STPM) and relative survival (using STRSRCS) spline based models assuming PEH after adjustment for other covariates

- Sex
- Year of MI
- Thrombolytic treatment
- Previous acute MI
- Creatinine
- Age at hospitalisation
- CK

These models were fitted using the flexible parametric model as discussed in chapter 5. This is directly comparable to the standard all-cause survival version (135; 147) from which the extended model is based, therefore the same models were fitted using this and the relative survival approach. A PEH model was fitted that included all of the covariates listed above without any time dependent effects. The results from a relative survival and an all-cause approach are shown in table 9.13. The rows listed as *glucose* do not contain diabetic status and the rows labelled as *diabetes* do not contain glucose in the model. The final row *G and D* contains both glucose and diabetes.

This table shows that the relative survival model is producing slightly larger effect sizes for quartiles 3 and 4 of glucose. The effect of diabetes, in the model with glucose in, was estimated as a protective effect in both the all-cause and relative survival model. Using a likelihood ratio test comparing the model without diabetes to one with it in (G and D) the relative survival model test gave a value of $p = 0.0611$ and the all-cause survival model gave a value of $p = 0.5658$ so the effect of diabetes is, strictly speaking, non-significant in

Adjusted for	Glucose quartile		
	Q2	Q3	Q4
Unadjusted	1.237	2.132	4.591
	(0.94 , 1.62)	(1.66 , 2.74)	(3.62 , 5.83
Sex	1.207	1.972	4.166
	(0.92 , 1.58)	(1.54 , 2.53)	(3.28 , 5.29)
Thrombolysis	1.311	2.336	4.347
	(1.00 , 1.71)	(1.82 , 3.00)	(3.43 , 5.50)
Previous AMI	1.269	2.138	4.608
	(0.97 , 1.66)	(1.67 , 2.74)	(3.64 , 5.84)
Creatinine	1.442	2.358	4.596
	(1.10 , 1.90)	(1.83 , 3.04)	(3.61 , 5.86)
Age at MI	1.179	1.897	3.892
	(0.90 , 1.54)	(1.48 , 2.43)	(3.07 , 4.93)
Year of MI	1.120	2.008	4.181
	(0.85 , 1.47)	(1.56 , 2.58)	(3.29 , 5.32)
CK	1.279	2.243	4.839
	(0.98 , 1.68)	(1.74 , 2.88)	(3.81 , 6.15)

Table 9.14: Estimates for glucose quartiles from a model with glucose and diabetes as proportional adjusted for various covariates

both all-cause and relative survival PEH models.

One of the most noticeable changes observed in table 9.13 is seen when compared to table 9.11 as the estimated effect sizes are now reduced. This is due to the modelling of other covariates. In order to determine which had the most impact a simple analysis including glucose and diabetes with each covariate was fitted and the effects are shown in table 9.14. The table shows the estimates for glucose from a glucose and diabetes model with adjustment for one other covariate. This found that age at MI was the most influential covariate reducing the estimate of the glucose quartiles the most.

9.6.4 Diabetes as time dependent

As with the MFPs glucose was not shown to be time dependent in the majority of the spline based models, i.e. only glucose quartile 4 in the 5df model showed statistical significance. In the MFP analysis diabetes showed strong signs that it may be influenced by time. This was assessed by fitting diabetes and glucose in the same model but allowing diabetes to be non-proportional and the results are shown in figure 9.24. A likelihood ratio test gave a value of $p = 0.0072$ when compared to a PEH model.

The estimates for glucose are shown in table 9.15, in the *Relative survival* column labelled '*unadj*'. These show that the effect size for the glucose quartile 4 is quite high compared to those shown when adjusted for many covariates in column '*adj*'. Looking at figure 9.24 there are obvious differences between the two plots. The plot on the left hand side is for non-diabetics and the hazard rate decreases for all quartiles, remembering that the

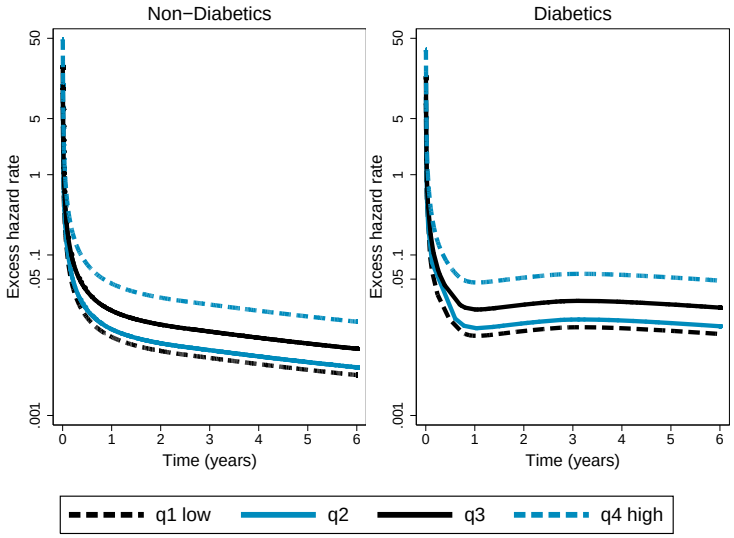


Figure 9.24: Plot of the excess hazard rate by glucose quartile and diabetic status using spline based models

Estimate	All-cause		Relative survival	
	Unadj	Adj	Unadj	Adj
Gluc Q2	1.243	1.225	1.248	1.324
	(1.05 , 1.48)	(1.03 , 1.46)	(0.95 , 1.64)	(1.01 , 1.74)
Gluc Q3	1.638	1.531	2.130	2.128
	(1.38 , 1.94)	(1.29 , 1.82)	(1.66 , 2.74)	(1.65 , 2.74)
Gluc Q4	3.013	2.313	4.622	3.449
	(2.58 , 3.52)	(1.95 , 2.74)	(3.64 , 5.88)	(2.70 , 4.40)
AIC	11042.6	2938.9	10297.9	2567.0

Table 9.15: Estimates for glucose quartiles with AICs for the unadjusted and adjusted (for other covariates) spline based (relative and all-cause) models and when diabetes is time dependent but glucose is not

quartiles are forced to be proportional. The right-hand side plot is for the diabetics excess hazard rate that stays broadly similar over the follow-up time. The four diabetes estimates are all higher than the non-diabetes estimate for glucose quartiles 1 to 3 with only quartile 4 in non-diabetics remaining fairly high in comparison to diabetics.

The excess hazard rate ratio for diabetic status can be seen in figure 9.25. The plot shows that the ratio is relatively flat up to a year with an estimate of around one, which then increases. The confidence intervals do not cross one after a year and a half showing that the effect of time in diabetic patients is clearly different from non-diabetic patients. This result supports the short-term findings where diabetes is not important, as no statistical difference in the ratios was found until 18-months had passed.

This model was adjusted for other covariates and compared to an all-cause approach, and the estimates for glucose are shown in table 9.15. The estimates for diabetes are not shown

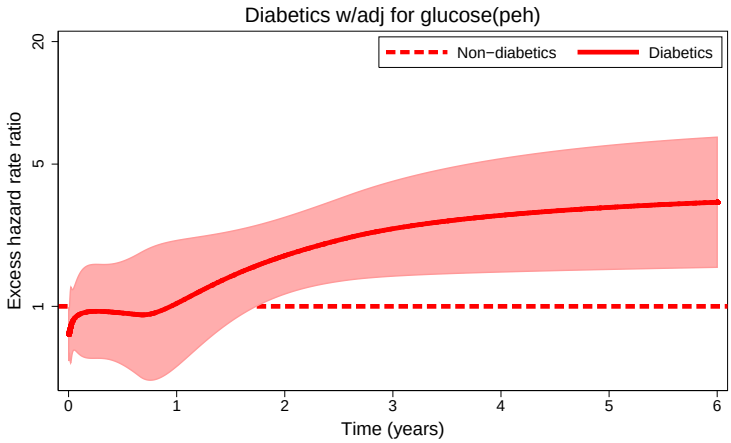


Figure 9.25: Plot of the excess hazard rate ratio by diabetic status using spline based models

as diabetes is involved in an interaction with time. The relative survival excess hazard rate ratios are higher than the all-cause hazard rate ratios for quartiles 3 and 4, whereas quartile 2 is similar under both approaches. Adjusting for covariates has reduced the effect sizes in both all-cause and relative survival models.

9.6.5 Diabetes and glucose as time dependent

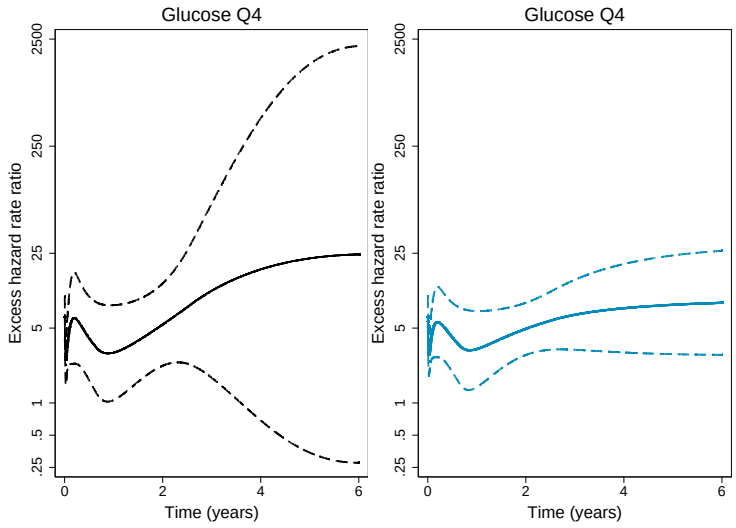


Figure 9.26: Plot of glucose quartile four excess mortality rate ratio with 95% confidence interval by diabetic status using spline based models

It is possible to fit the flexible parametric model to assess more than one time dependent effect but the interpretation is made increasingly difficult because the effects become dependent on each other (see section 5.5.1). However it is possible to fit the model and look

at it, for example, in terms of glucose quartiles at each level of diabetes as shown in figure 9.26.

This plot shows how the time dependent effects of glucose quartile 4 are clearly influenced by diabetic status. The start of the plot is fairly similar where the majority of the information lies but after two years the estimate for glucose quartile four in patients that are non-diabetic show a much larger degree of uncertainty compared to the diabetics.

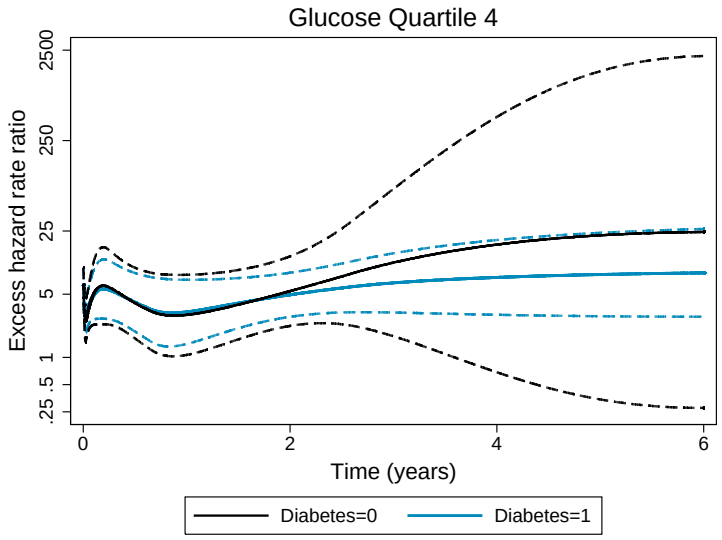


Figure 9.27: Plot of the glucose quartile excess mortality rate ratios overlaid by diabetic status using spline based models

It is possible to overlay these two plots as shown in figure 9.27. This gives a much clearer comparison between the two estimates. It would usually be expected that these estimates would be the same only altered by a constant difference, however because this model is fitted on a log cumulative excess hazard scale the terms for the time by diabetes interaction do not cancel out as if this was on the log excess hazard scale. This is discussed in more detail in section 5.5.1. The differences shown here are quite pronounced and have been found to be much smaller in other databases.

Another reason to assess this model is to assess whether fitting both diabetes and glucose as time dependent effects is suitable. The model fitted with diabetes as the non-proportional covariate had an AIC of 2940.59, whereas the model with glucose as the time dependent effect had an AIC of 2948.795. The new model has an AIC of 2949.203 which is higher than both of the simpler model AICs suggesting that this model is not adding anything. The likelihood ratio test comparing diabetes as time dependent to both glucose and diabetes as time dependent gave a non-significant value of $p = 0.0877$. Thus diabetes should be modelled as time-dependent, but glucose quartiles can assume PEH.

9.6.6 Interaction of diabetes and glucose

When investigating MFP models in section 9.5.3 an interaction between diabetes and glucose was found to be of interest. This model was fitted as a PEH model with only glucose quartiles, diabetes and their interaction included, and when compared to a PEH model without the interaction term the likelihood ratio test was given as $p = 0.0072$, showing a highly significant effect with a lower AIC (2969.818). Note that these values take into account the missing data present in the covariates that will be adjusted for in the following.

It was previously discovered that diabetes should be modelled as time dependent so this was added to the model which gave an AIC of 2963.428 with a likelihood ratio test giving a value of $p = 0.0058$. This addition is also beneficial to the model which now includes glucose as a proportional effect diabetes and glucose interaction as a proportional effect and diabetes as a time dependent effect. Adding the covariates to this model gives an AIC of 2559.303, which is much lower due to the addition of other covariates.

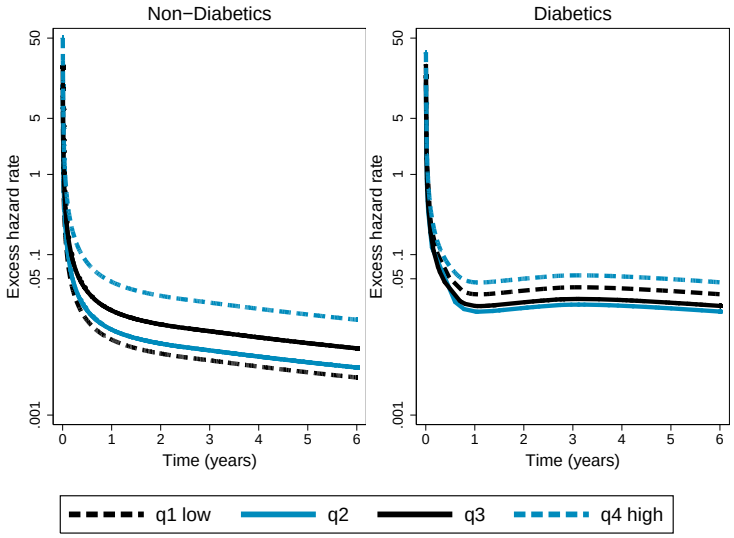


Figure 9.28: Plot of the glucose quartile excess mortality rates by non-diabetics (left) and diabetics (right) using spline based models with an interaction

Figure 9.28 shows the unadjusted model with an interaction term in it, with diabetes as time dependent. The glucose quartile groups are spread further apart in the non-diabetic group. The diabetic group shows that the two mid-level glucose groups have the lowest mortality rates with quartile 4 the worst. All four glucose groups in diabetic patients have a higher excess mortality rate than the worst non-diabetic group.

Figure 9.29 shows the excess hazard rate ratios for diabetes from the model that includes an interaction with glucose. Estimates of the diabetes excess hazard rate ratio without the interaction term can be seen on the left-side of figure 9.25. It is necessary to show

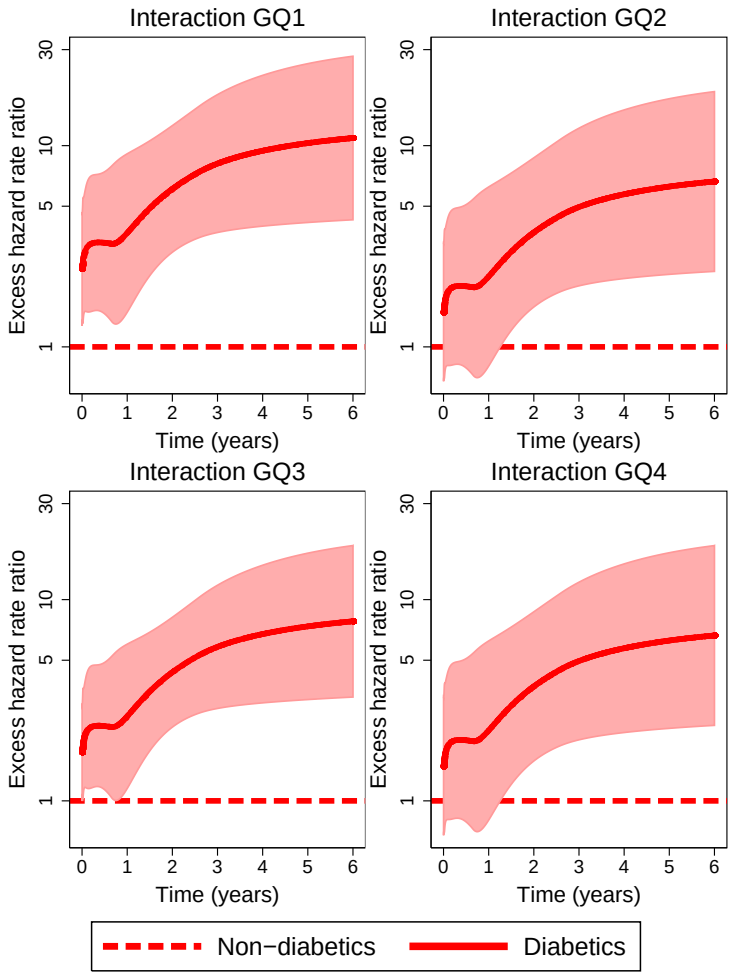


Figure 9.29: Plot of the diabetes excess mortality rate ratios by glucose quartile using spline based models with an interaction

the effect for each of the four glucose groups as the interaction is involved with diabetes. This now shows that a diabetic patient is at a significantly increased risk for the entire follow-up in glucose quartiles 1 and 3, whereas quartiles 2 and 4 show that the effect is non-significant for the first year of follow-up, which supports the short term findings.

The model could be assessed for further improvements by making the interaction time dependent. However this failed to converge and so could not be assessed for these data.

9.7 Adjusting for multiple covariates

9.7.1 Full covariate analysis

This section will further investigate the modelling process by allowing other adjustable covariates to be time dependent. So far the other covariates to be adjusted for have all been considered as proportional and this will now be investigated to see if this was a fair assumption. A simple process was used to determine if the covariates were time dependent. This started by assessing each covariate for a significant time-dependent effect while adjusting for all other covariates as proportional effects, i.e. Diabetes, Glucose quartiles, Sex, Year of MI, Thrombolytic treatment, Previous AMI, Creatinine, Age at event and CK were all included in the model but each term was taken in turn and made time dependent while all others remained proportional. These models were then tested using likelihood ratio tests compared to when all covariates were modelled as proportional.

There were problems with the assessment of creatinine levels as the model failed to converge and altering the degrees of freedom and orthogonalising the splines had no benefit. This meant that this could not be assessed as time dependent, but was still included as proportional. It had already been determined that diabetes was time dependent and glucose was not. Age $p = 0.0137$ and CK $p < 0.0001$ both gave a significant result from the global test. Previous AMI also gave a significant result from the global test $p = 0.0017$ but each term within the model, across the splines, was non-significant at the 5% level. Thrombolysis, sex and year of hospitalisation were non-significant with a value of $p = 0.3279$, $p = 0.4207$ and p=0.4017 respectively.

All of the significant effects were put into a model as time dependent with glucose quartiles, thrombolysis, sex, year and creatinine included as proportional effects. This model gave an AIC of 2500.014. To reduce the complexity of this model the time-dependent covariates were made proportional one at a time to see if the model benefitted. The first to be removed was age at hospitalisation, which gave a value of $p = 0.0039$ from a global test and as such cannot be removed. CK was also kept in the model as once removed the likelihood ratio test gave a value of $p < 0.0001$. Removing previous AMI gave a value of $p = 0.0007$ and diabetes $p = 0.0009$ so none of the time dependent effects could be fitted as proportional effects.

Now the interaction found earlier between glucose and diabetes needed to be added to the model as a proportional effect and this gave an AIC of 2456.911. Testing this model against the model without the interaction in it gave a value of $p = 0.0050$. This final model includes Age at hospitalisation, CK, previous acute MI and diabetic status as time dependent effects and all other covariates as proportional, including a diabetes and glucose interaction in a 5df model which includes 4 internal knots.

A five year relative survival rate by glucose quartile are estimated as 0.864 (0.744 , 0.930) for quartile 2, 0.812 (0.725 , 0.874) for quartile 3, and 0.755 (0.706 , 0.798) for quartile 4. This has to be calculated under several conditions, firstly the year of hospitalisation was set to 2000, thrombolysis was given, creatinine (110.57) and ck (2079.8) are at their mean values, and the male patients are 70 years old at hospitalisation with no previous AMI. This highlights how difficult it can be to generalise results from complex models as all other covariates need to be accounted for.

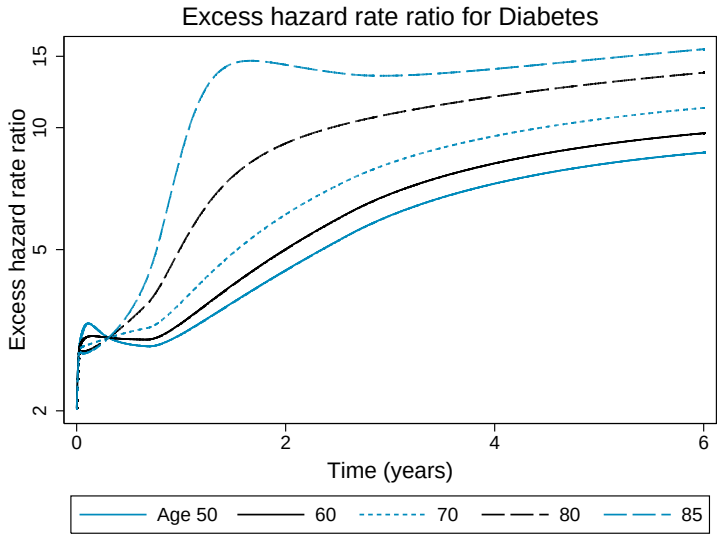


Figure 9.30: Plot of the diabetes excess mortality rate ratios at different ages where CK is set to the mean (2079.842) and there is no previous AMI

Diabetes is time dependent and as discussed earlier (section 5.5.1 and 9.6.5) it is dependent on the other time-dependent covariates. It is possible to estimate the excess hazard rate ratio at set levels of the other time dependent covariates, as figure 9.30 shows. CK is set to the mean value of 2079.8 and there are no previous AMIs. Five ages at hospitalisation are shown; 50, 60, 70, 80 and 85 (90 was not estimated well after two years of follow-up, possibly due to the lack of data). The excess hazard rate ratios are for diabetics and show that increasing age increases the excess hazard rate ratio.

Figure 9.31 shows the excess hazard rate ratio for diabetes with age set at 70 with no history of AMI while CK is set to 500 and 5000. The figure also includes 95% confidence intervals which show a fairly large amount of uncertainty. The estimates are fairly similar after a year of follow-up.

Table 9.16 shows that the estimates from two all-cause methods and a relative survival approach. The Cox estimates and the flexible parametric all-cause spline based models provide similar estimates for all of the glucose quartiles with similar 95% confidence intervals in both diabetics and non-diabetics. The relative survival model shows that these estimates have been raised for the third and fourth glucose quartile in non-diabetics. From

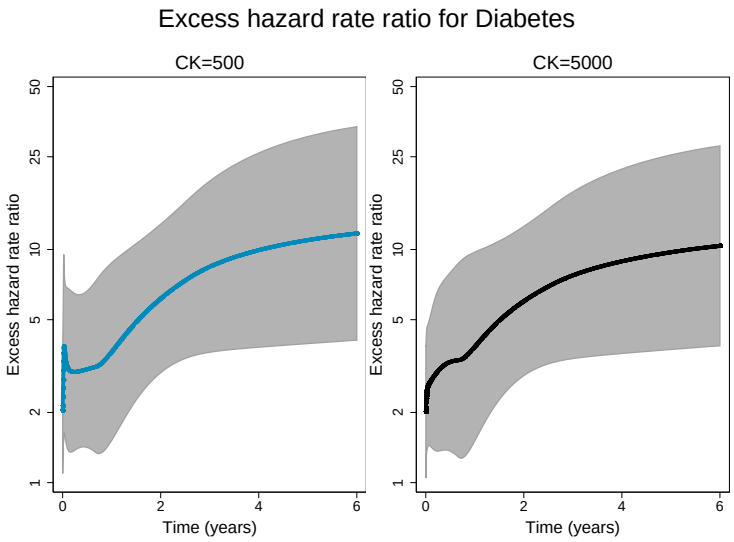


Figure 9.31: Plot of the diabetes excess mortality rate ratios at different values of CK where age is set to 70 and there is no previous AMI

these results the conclusions would be that patients in glucose quartile 4, when compared to quartile 1, have an average excess mortality of approximately 300% higher in non-diabetics. This suggests that survival is worse for patients with high glucose who are not diagnosed as diabetic, which is feasible due to the way patients who are diabetic may be more likely to be treated quickly with insulin.

In diabetics, glucose quartiles 2 and 3 are lower than one. However, the effect of glucose in diabetics is non-significant in the three modelling approaches, suggesting that admission blood glucose levels are not important in a diabetics long-term prognosis. Admission blood glucose is more important in non-diabetics.

9.8 Continuous glucose

One of the key advantages of modelling using spline based models on the log cumulative excess hazard scale is the ability to model continuous covariates as split-time data is not used. In the previous section looking at fitting a model after adjustment for other covariates CK, creatinine, year of MI and age at MI were all modelled as continuous covariates. During this modelling stage when using spline based models the covariates are assumed to be linear and so if a covariate is found to be time-dependent the effect is modelled as linear across all of the splines, for example, an assumption that age is linear is made, but the gradient is a function of time. To get around this assumption, a transformation of the covariate can be performed if linearity does not seem appropriate. As well as investigating glucose as a linear effect, glucose was also transformed into a

Diabetic Status	Model df	Glucose quartile		
		Q2	Q3	Q4
Non Diabetic	Relative Survival	1.416 (1.07 , 1.88)	2.255 (1.73 , 2.94)	3.854 (2.98 , 4.98)
	All Cause	1.265 (1.05 , 1.52)	1.574 (1.31 , 1.89)	2.501 (2.09 , 2.99)
	Cox	1.256 (1.05 , 1.51)	1.562 (1.30 , 1.87)	2.457 (2.06 , 2.94)
Diabetic	Relative Survival	0.553 (0.23 , 1.33)	0.788 (0.40 , 1.56)	1.064 (0.61 , 1.86)
	All Cause	0.783 (0.41 , 1.51)	0.816 (0.46 , 1.46)	1.112 (0.69 , 1.80)
	Cox	0.790 (0.41 , 1.52)	0.827 (0.46 , 1.48)	1.131 (0.70 , 1.83)

Table 9.16: Estimates of the excess hazard rate ratios for glucose quartiles for the final adjusted spline (relative and all-cause) models and Cox (all-cause) model

spline function with the internal knots fitted to the already determined glucose quartile points, which were $7, 8.3$ and 11 with the external knots set at 0 and 40.

To begin with continuous glucose was modelled with only diabetes in a PEH model. The results of this are shown in figure 9.32. The top row of this figure shows the results where glucose was included as a linear function, i.e. not transformed. The bottom row shows where glucose was transformed into a spline function to allow it to be more flexible (using the program shown in appendix A.2). This figure also shows the estimates from a relative survival model on the left-hand side using `strsrcs` and an all-cause survival model on the right-hand side using `stpm`.

It is also worth noting that the reference value is indicated by the vertical dashed line in both cases. The linear effect of glucose is shown to be very similar from a relative and all-cause approach. The relative survival model has a slightly higher excess hazard ratio and the confidence intervals are slightly wider at the extremes. The reference point for the linear effect and the spline based models is 7 as this is the point in which a subject may be considered as diabetic. Putting this reference point higher caused issues in the early parts of the model as most patients have a glucose level under the mean.

The spline results are on the same scale as the linear terms but there is a lot more uncertainty early on in these models that the linear effects have not managed to find, but for glucose values above the reference value the estimates appear very similar from both approaches. When using splines there is a difference between the all-cause and relative survival models as the relative survival model has more uncertainty. Both plots suggest that the lowest glucose and the higher levels of glucose both have a higher excess hazard

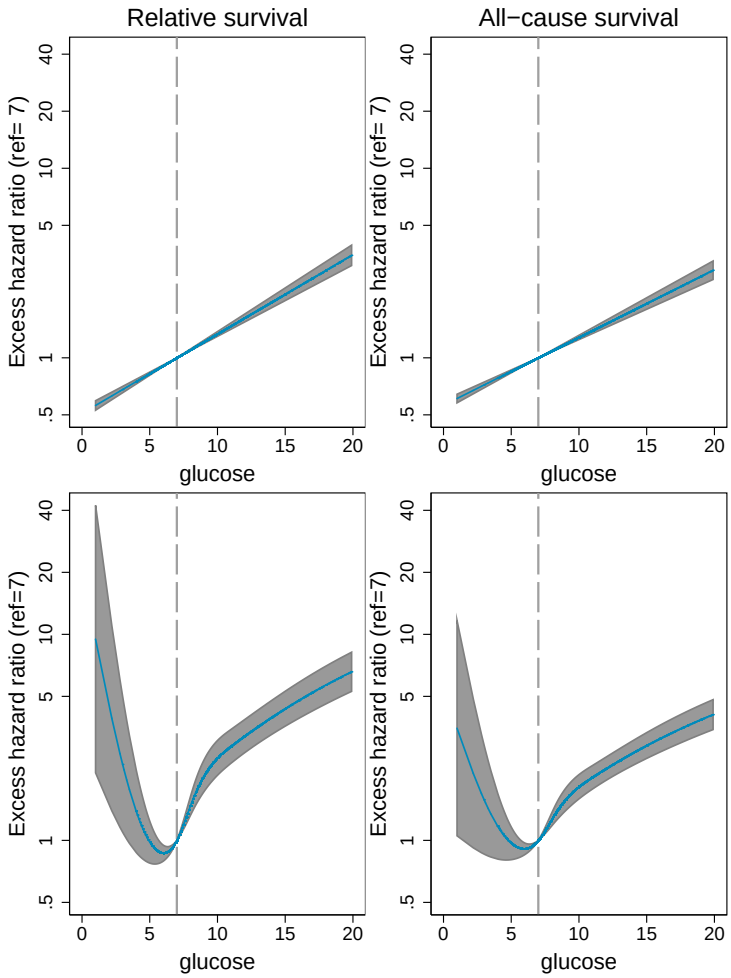


Figure 9.32: Plot of the excess hazard rate for glucose using spline based models in a relative survival (right) and standard survival (left) model where glucose is linear (top) and modeled by splines (bottom)

rate than patients with a glucose level of 7. Linearity may not be appropriate as low glucose is also dangerous.

A five year relative survival rate can be obtained using continuous glucose and for a diabetic patient with an admission blood glucose of 7mmol/L, assuming glucose is linear, gave an estimate of 0.8366 (95%CI 0.8051 , 0.8635). An all-cause model gave a 5-year survival estimate of 0.7141 (95%CI 0.6775 , 0.7474). Using splines the relative survival estimate was given as 0.9005 (95%CI 0.8751 , 0.9210) and the all-case model was 0.7772 (95%CI 0.7410 , 0.8091). These estimates are quite different to the glucose as linear model and shows how the linear model may not be reasonable.

Continuous glucose can be modelled as time dependent with diabetes in the model as shown in figure 9.33. On the left hand side an admission blood glucose level of 9 is compared to 8, whereas on the right hand side a value of 20 is compared with 19. It shows

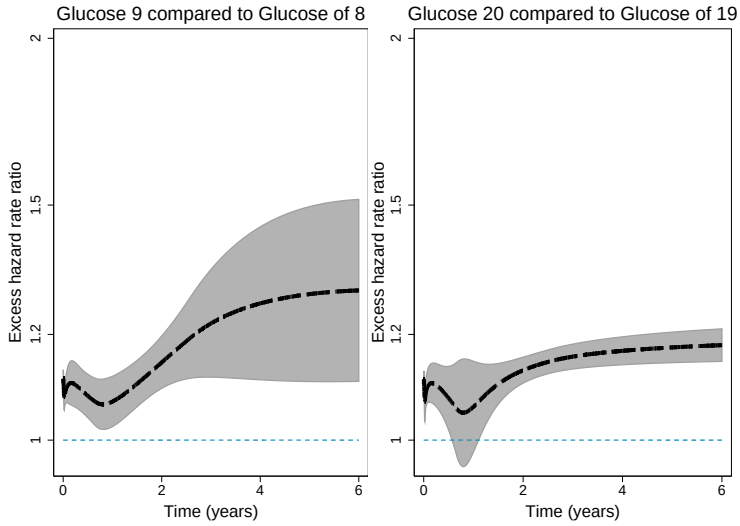


Figure 9.33: Plot of the excess hazard rate ratio for continuous glucose, assuming linearity, using spline based models in a relative survival approach where admission blood glucose of 9 is compared to 8 and 20 is compared to 19

that even though both comparisons are only a single unit increase the effect of glucose is not the same. The figures both suggest that increasing blood glucose increases the excess mortality rate ratio.

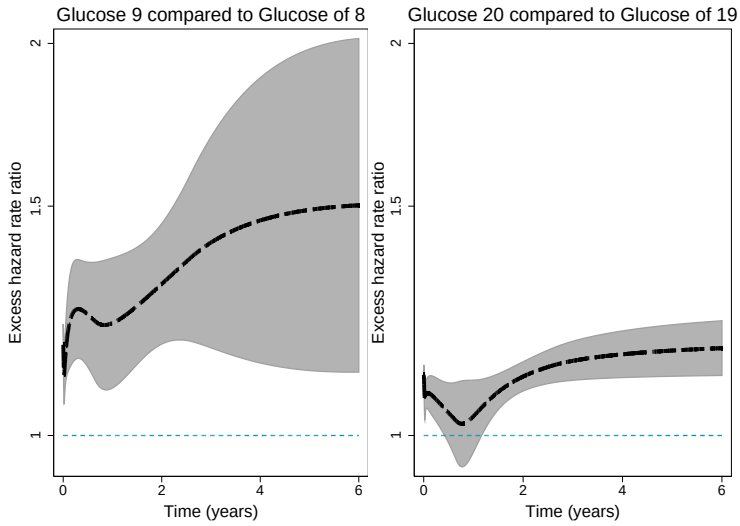


Figure 9.34: Plot of the excess hazard rate ratio for continuous glucose using a quadratic function for glucose in spline based relative survival models where admission blood glucose of 9 is compared to 8 and 90 is compared to 19

In figure 9.32 it was shown that the linear effect was not picking up the shape shown by the approach using glucose spline transformations. The spline approach appeared quadratic

and this was assessed as shown in figure 9.34 where a quadratic function of continuous glucose was assessed. Here the scale is larger than in the linear model, shown in figure 9.33, as the estimates have increased. The shape of the curves are similar to the linear effects model but the model is improved upon, over the linear model, as a global likelihood ratio test gave a value of $p < 0.0001$. This shows that interpreting continuous covariates can be difficult but is possible even where there are complex non-linear time dependent effects of continuous covariates.

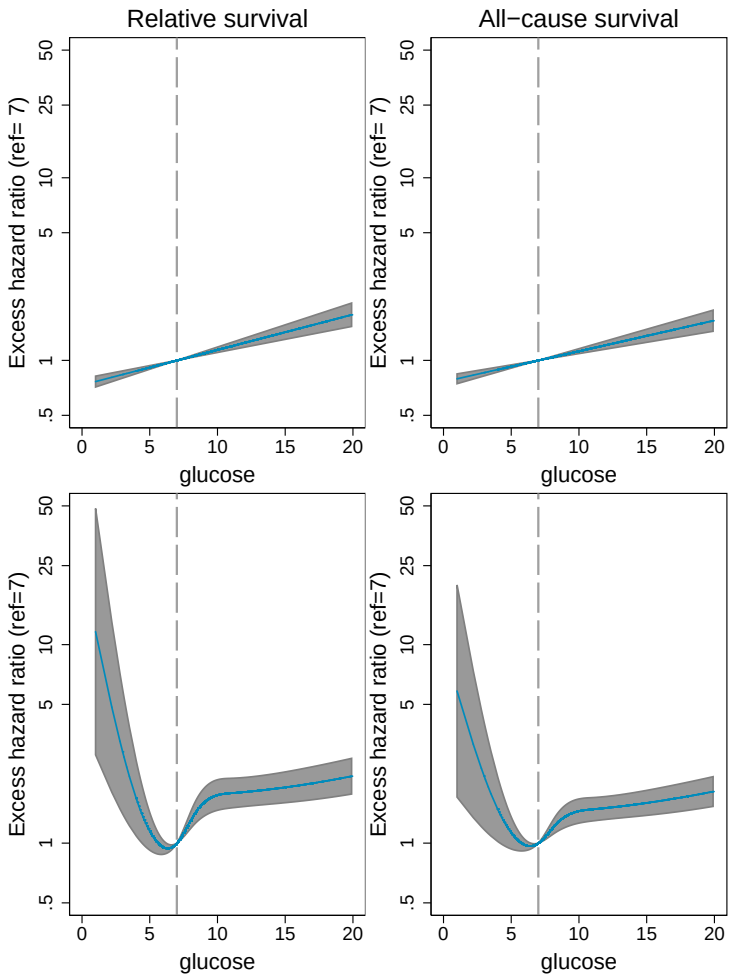


Figure 9.35: Plot of the adjusted excess mortality rate for glucose using spline based models in a standard survival (left) and relative survival (right) model where glucose is linear (top) and modeled by splines (bottom)

When attempting to fit the model found in the previous sections there were problems with convergence, so to adjust the estimates for other covariates a PEH model was built that included; glucose as continuous, sex, year of MI, thrombolytic treatment, previous AMI, creatinine, age at event and CK as proportional effects but diabetes was allowed to be time-dependent. These model results are shown in figure 9.35.

There is very little difference in the linear glucose plots after adjustment for the other

Diabetic Status	Model df	Glucose quartile		
		Q2	Q3	Q4
Non-diabetic	Relative Survival	1.333 (1.00 , 1.77)	2.310 (1.77 , 3.02)	5.251 (4.07 , 6.77)
	Period Analysis	2.194 (1.02 , 4.71)	3.889 (1.92 , 7.89)	6.841 (3.41 , 13.73)
Diabetic	Relative Survival	0.608 (0.25 , 1.51)	0.715 (0.35 , 1.46)	1.408 (0.79 , 2.52)
	Period Analysis	0.427 (0.03 , 5.31)	0.518 (0.10 , 2.65)	1.649 (0.52 , 5.22)

Table 9.17: Estimates of the excess hazard rate ratios for glucose quartiles from a standard relative survival and a period analysis model

covariates, the relative survival approach has slightly wider intervals but the conclusions would be unchanged. Modelling glucose using spline based models has a similar effect as seen in the unadjusted models. The estimate is at its lowest around the reference point with the highest ratios in the low glucose patients. The adjustment has increased the interval width in both approaches when compared to the unadjusted models. After the reference point the top row of the figure provides similar estimates to the bottom row. However before the reference point the linear effects are not capable of estimating the increased risk associated with low values of glucose.

It was not possible to model glucose as time dependent while adjusting for covariates due to model convergence problems, which therefore needs further work. An advantage of the time dependent approach is this ability to include interactions and transformations of individual level covariates.

9.9 Period analysis

Using the model that only includes glucose and diabetes it was found that an interaction was necessary and that diabetes was time dependent. This model can be presented simplistically in figures, and as such is used to now investigate up-to-date survival estimates using period analysis (see chapter 8). The period is defined from 01.01.2002 with only a six year follow-up considered. This means that patients who died, or completed six-years of follow-up before this date are excluded from the analysis.

Figure 9.36 shows the results from a standard relative survival approach on the right side of the figure and from a period analysis on the left side. The plots are shown for non-diabetics only as there are much fewer diabetic patients. The models can be compared left to right and this shows that the period analysis estimates a higher relative survival rate in all glucose quartiles, by up to 20% in quartile four. The excess hazard rates drops

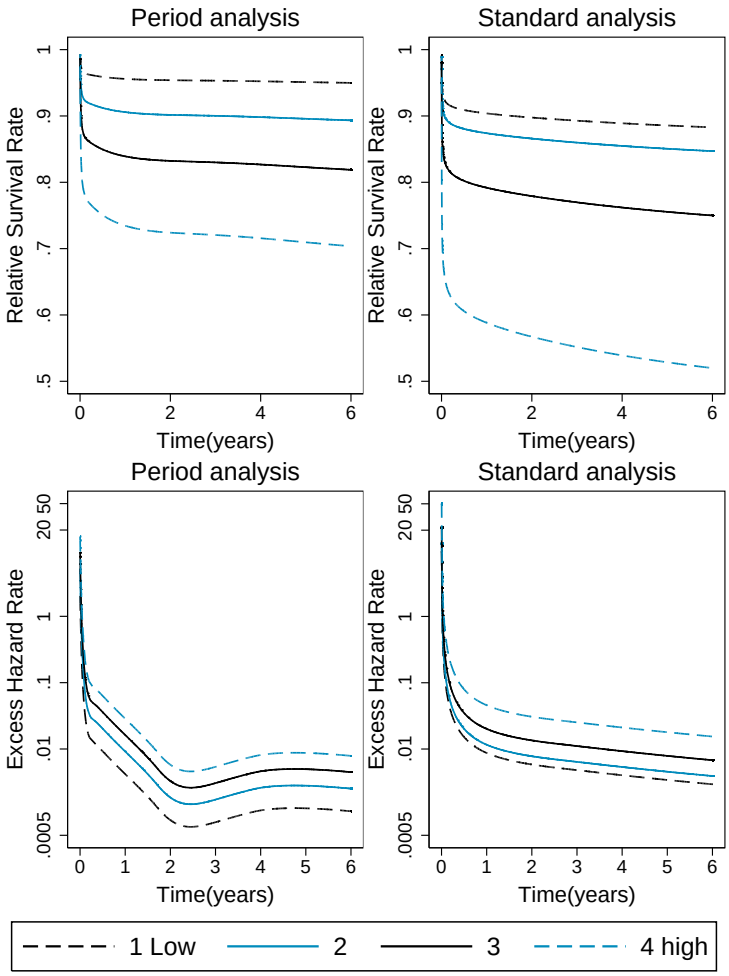


Figure 9.36: Plot of the relative survival (top row) and excess hazard rate (bottom row) from a period analysis (left) and standard relative survival (right) model using spline based models with four internal knots

quite sharply to around two years of follow-up in the period analysis, which drops much lower than in the standard plot. In the relative survival plot, quartiles two and three do not show as big a difference as seen in quartile 1, when using period analysis. However, there is a very large difference in quartile 4.

The excess hazard rate ratios for glucose quartiles are shown in table 9.17 and are shown by diabetic status as glucose is involved in an interaction. Here period analysis has increased the effect size of the excess hazard rate ratios in non-diabetics across all quartiles and in the highest quartile group in diabetics. However the estimate in quartiles 2 and 3 for diabetics are lower in the period analysis approach than in a standard relative survival approach. It should be pointed out that the effect of glucose in diabetics is non-significant and should not be over-interpreted. The reference group, glucose quartile one, only has a small proportion of diabetics. The lack of precision from a period analysis is highlighted in this table as the confidence intervals are much wider than the standard relative survival

approach.

These results could prove to be clinically important as these initial suggestions show that by obtaining more up-to-date estimates the initial excess hazard is much lower, and the survival is better over a longer follow-up period. It has been shown that period analysis can be incorporated into any flexible parametric model, even if it is complex. If relative effects are of primary interest, and if it is possible to assume that these remain consistent over time, then a standard analysis may be more powerful. However if absolute effects, that better reflect current practice, are of interest then a period analysis may be better.

9.10 Discussion

9.10.1 Short term analysis

The initial interest in diabetes and glucose was to investigate the impact of these two factors on short term survival in order to inform clinicians of the potential importance that admission blood glucose could have on a patient's immediate survival and up to 30-days, which is where a very high proportion of all deaths related to the initial MI occur. For this investigation the Leicestershire CCU dataset was used, which has a broad spectrum of ages, ethnicity and the survival rate following MI is around the average observed within the UK. The patient demographics, shown in tables 9.1 and 9.2 highlight the relative similarity of covariate proportions by diabetes and glucose quartiles respectively.

There are two main findings from the short term analysis, firstly during the study period, the absolute mortality associated with diabetes fell markedly. Similarly, the mortality associated with elevated blood glucose concentration also fell, particularly for patients in whom admission glucose was highest. Secondly the association with mortality after STEMI for admission blood glucose concentration was found to be independent of diabetic status and the effect of diabetes became statistically non-significant after adjustment for glucose. This suggests that in the context of STEMI in routine practice, admission blood glucose concentration provides a consistent and more robust marker of prognosis than diabetes.

Clinically it is possible that adverse outcome related to glucose concentration may simply be representative of an undiagnosed diabetic. There were no adjustments for the use of individual secondary prevention therapies. The impact of these therapies was not the main focus of this analysis, and consideration of individual treatment effects may introduce bias for a number of reasons. In particular, in the most severely ill patients, early mortality and adverse clinical features in survivors will limit treatment prescription.

There are potential issues with the methodological approach used in the short term analysis in that there is no accounting for possible censoring. Each patient was allowed to be in

the follow-up for the minimum amount of time by having a strict inclusion criteria set by the final dates at which this would be possible. However mortality data are informed by the Office for National Statistics and as such any deaths from patients who left the Leicestershire area, or migrated during this follow-up period, are assumed to be alive. This was considered as too few for a short term analysis and so adjustment for censoring was not performed.

Death in the short term is adequately modelled using these methods as there is little or no censoring and the odds ratios obtained are able to adequately adjust for all covariates giving good estimates of patient mortality. These methods would not be appropriate for a long term analysis.

9.10.2 Long term analysis

Assessing long term survival allows the use of relative survival due to mortality from other causes having a higher probability of occurrence over a 6-year period than over a 30-day period. Standard all-cause analysis was also performed as a comparator. The piecewise relative survival approach investigated the survival plots by glucose and diabetes and highlighted the difference between all-cause and relative survival as well as the differences between the covariate levels. Glucose quartile 4 and diabetes were shown to have the lowest survival rate and the observed and relative survival rates were also very similar.

The problem with modelling relative survival using the standard methods is the need to use split-time data which was necessary when modelling using the piecewise and MFP approaches used here. For the piecewise models this meant that using time dependent effects created many parameters to analyse and the dataset was not large enough to cope with adjustment for other covariates. Whereas in the MFP analysis, grouped level models were used due to computational limitations as the models take hours to run at an individual level, which meant that no continuous covariates could be assessed. Fine split-times were used so there was not enough data to adjust for other covariates. A small dataset can create problems due to zero cells at various time-points especially when investigating time dependent effects.

Taking these issues into account diabetes and glucose were modelled in PEH and non-PEH models. The piecewise models were compared with stratified Cox models and the results were found to be quite different for glucose quartile 4. The effects are much higher in the relative survival approach which is modelling the deaths in excess of what is expected. These effect sizes are reduced slightly when using the MFP models as they are allowed to be more flexible but on the whole the piecewise and MFP models are broadly similar. During the piecewise modelling the PEH and non-PEH models both attributed a significant effect to diabetes after glucose had been adjusted for, and when investigating MFPs this was

also confirmed. This finding contradicts the short term finding in that diabetes becomes equally as important as glucose. However both results should be considered as correct as the MFP section goes on to show that the effect of diabetes is not important in the short term, but does have influence in the long term reflecting that diabetes is a chronic condition.

The potential interaction between diabetes and glucose was also tested using fractional polynomials and found to be significant in both the PEH and non-PEH models. However this model was overly complex as glucose, diabetes and their interaction all varied by time and this model did not adjusted for other covariates.

9.10.3 Use of flexible parametric models

The first thing to do when changing the methodology used for the modelling was to compare these models to the previous best to make sure that the results are sensible and when compared with the MFPs in table 9.10 the results were near identical. One of the potential weaknesses of modelling with splines is the model's sensitivity to the frequency and location of it's knots, however these types of splines are generally quite robust as was shown when comparing models with 2 degrees of freedom up to 6, all of which gave very similar estimates for glucose and diabetes. The model AIC implied, for these data, that more knots provided a better fit. However there were problems when using non-PEH modelling as the 6df models did not converge even when orthogonalising the splines to remove any correlation between them. This can be remedied by altering the placement of the knots but, in the example given here, did not improve the model over the 5df.

Time dependent effects were analysed first for glucose then for diabetes. When assessed separately they were found to be time dependent when using 5df models. There was an issue with glucose as time dependent in the 5df model as it gave a large estimate for glucose quartile 4, but after reducing follow-up and assessing how many events there were in the final year of follow-up it was determined that the lack of data was not the cause.

Covariate adjustment is much easier than in the previous relative survival models because there is no split-time data and no categorising of covariates is required. The covariates adjusted for in the short term analysis were adjusted for in the long-term analysis. These were all modelled under a PEH assumption and were compared to the estimates obtained from an all-cause model using the same spline methodology. The results (table 9.13) show that these models are very similar but the effect of glucose in the third and fourth quartiles are larger in the relative survival approach, i.e. their excess risk is higher than the estimated overall risk when compared to the baseline group in both models.

The interaction between diabetes and glucose was assessed and found to be highly significant as a proportional effect but the addition of time dependency to the interaction

was not beneficial. Diabetes and glucose were time dependent individually and there is a problem with this method as shown in figures 9.26 and 9.27 where the glucose effect is not independent of diabetic status. A model with glucose as proportional and diabetes as time dependent was found to be a better fit. This model was then adjusted for other covariates and these covariates were assessed for time dependent effects. The final model included:

- Age at hospitalisation, CK, previous AMI and diabetes as time dependent effects
- Glucose quartiles, thrombolysis, creatinine, sex and year of hospitalisation as proportional,
- and included a proportional diabetes and glucose interaction.

Glucose was modelled as a continuous covariate, an advantage over other relative survival models, and was shown to be problematic when investigating time dependency as the models failed to converge when adjusting for the covariates listed above, which is likely to be due to the size of the dataset. It was also shown how to model continuous covariates without the assumption of linearity across the spline terms which may or may not be a feasible assumption.

9.10.4 Conclusions

Long term analysis has shown that the short term results do not hold over a longer follow-up. Glucose is still an important factor but diabetes is as important as glucose in long term survival, however this result could be expected as diabetics are known to have a higher mortality rate than the general population.

Relative survival has added information regarding the expectation of death and this has raised the effect sizes of the glucose estimates to be above those of an all-cause analysis. This has highlighted the impact upon patient survival for those in glucose quartile 4, which was also observed in the short term study. The estimates from an all-cause approach may be considered too conservative as they include all deaths regardless of their expected mortality, which is dependent on age and sex. The patients in glucose quartile 4, having adjusted for expected mortality, are shown to be at greater risk of death due to the disease of interest than under an all-cause approach.

Out of all the methods used the most appropriate is the flexible parametric model as it provides a flexible approach to relative survival allowing the assessment of smaller datasets, continuous covariates and does not take a long time to fit computationally. It is not perfect in that modelling more than one time dependent effect leads to interpretation difficulties and there are issues with some model convergence, particularly in models with many

internal knots. A brief investigation using period analysis showed that this methodology may provide clinically important results as, in the example shown here, the effects were reduced by up to 20%.

The extension to long term relative survival from a short term mortality investigation has supported the evidence that the impact of glucose on patient survival is of key importance and clinically this also has implications not only for the first 30-days following MI but also for up to 6-years and potentially even longer. The long-term study also found that diabetes is involved in an interaction with time and glucose and is a very important factor in a patient's survival.

Chapter 10

Discussion

10.1 Achieving the aims of the thesis

The primary aim of this thesis was to develop suitable models and move currently established relative survival techniques from cancer into heart disease. This started with an initial introduction in chapter 2 to the methodology that was published in the European heart journal (2). An investigation into the use of relative survival in published cardiovascular research (chapter 3) discovered that there were very few applications of only parts of the approach and only a couple papers applied a relative survival model (18; 85).

Moving standard models across therapeutic areas, from cancer to heart disease was expected to create problems due to the different shape of the hazard, especially the very large excess hazard in the first 30-days of follow-up. However, many of the current methods failed to even provide feasible estimates, let alone good fitting models as shown in chapter 4. This led to the development of a new relative survival model in chapter 5, which did not split time and is fitted to the log cumulative excess hazards scale. This model was implemented in Stata (see chapter 6 and appendix A) and proved to be robust to knot frequency and location. It also provides a good fit to the LRI CCU dataset. The model will also be useful for population based cancer studies.

There is a key assumption that the expected survival rate used in the calculation of relative survival is defined as the rate for a population similar to the patient group but free of the disease under study. This is generally considered as satisfied in most cancers (26) except in, for example, lung cancer where smoking is highly linked with lung cancer, but also to other causes of death, including cardiovasuclar disease. This assumption needed to be assessed in heart disease. In chapter 7 it was found that absolute effects (i.e. relative survival estimates), in the very elderly, became slightly biased whereas relative effects had only a small degree of bias. This was taken to be a positive result and relative survival

does appear to have a use in heart disease. Period analysis was also assessed in chapter 8 and found to offer improved estimates of relative survival in patients diagnosed in more recent periods than the traditional methodology.

The final section of the thesis merged these approaches in order to assess a clear clinical question relating to the level of a patient's admission blood glucose (chapter 9). The short term analysis was submitted (186) and a long term follow-up was assessed using relative survival methodology. This incorporated standard models, flexible parametric models, period analysis and the assessment of continuous covariates. The splines performed better than the standard methods and were modelled using several covariates rather than only age groups, as had been adopted throughout the thesis.

10.2 Finding a poor fit from current methods

It was essential to assess the use of relative survival models in heart disease that are currently used in cancer. This included modelling piecewise GLM models (40; 115; 27) which all proved to be good, quick fitting models that provided feasible estimates. The piecewise models provided very similar results to each other and have become the standard models in applied research. For these reasons the piecewise models were used as the benchmark for all other models assessed. There are problems with piecewise models in their assumption of a constant excess hazard within each interval, and the need to estimate many parameters for time dependent effects, which in small datasets can create zero cells with no events. These models were also found to have a bias when used in the simulation (chapter 7) with an inappropriate choice of intervals.

For the (approximately) continuous models based on splitting the time scale three sets of split-times were adopted. Two were proposed by Remontet et al. (118) which were named as long and short intervals. These were initially proposed to be suitable for most datasets in cancer, but as discovered fairly quickly they are not entirely suitable in the heart disease dataset from the LRI CCU. The long interval dataset was a poor fit in almost every instance whereas the short interval dataset proved to be successful when using grouped level data. This disagrees with Remontet *et al.* (118) who state that the long and short intervals are adequate. Models appeared to fit better when using user defined intervals, which allowed for more splits early on and fewer splits towards the end of follow-up where there is little information. However, it should be acknowledged that this is data-driven to some extent and other datasets may require different splits, illustrating one of the main problems with the use of split time data.

Polynomials (section 4.5), MFPs (section 4.6) and splines (section 4.8) were investigated, and at the individual level all of them were poor, i.e. convergence issues, poor fits and long computational times. The MFPs took almost four days to converge and then provided

unfeasible estimates when investigating the whole dataset, including the first 30-days. Models were improved slightly, in all cases, by removing the first 30-days, and this suggested that the models were having estimation problems due to the very high early excess mortality. The desired model needed to use all of the data and so these results were fairly disheartening.

Fitting the three continuous approaches to grouped level data proved to be successful for MFPs, whereas the splines and polynomials were still fairly poor. The MFPs are much quicker to fit at the grouped level and found that both the short and user-defined intervals provided feasible estimates that were similar to the piecewise estimates obtained earlier. However the problem with MFPs is in the model selection, where the final, *best* fitting model is only one of a possible forty-four when using an FP2 model, with many of these likely to have very similar AICs and potentially fitting very different models with different confidence intervals. This was assessed using bootstrapping (126) (section 4.7.1) and model averaging (127; 130; 131; 129; 132) (section 4.7.2) which both showed fairly robust estimates of the baseline excess hazard rate for these data. However, the excess hazard rate ratios showed more uncertainty at the start of follow-up than the *best* fitting model estimated.

Spline model selection when using split-time data was raised in the discussion (section 4.9.3) as several models could be selected using different frequencies and locations of knots which generally results in only one, or the *best* fitting, model being presented. Further work could investigate model selection using model averaging as was performed on the MFPs. Model selection algorithms are available, such as that proposed by Zhou and Shen (199) alongside Bayesian model averaging and stepwise procedures as discussed by Hansen and Kooperberg (200). The models investigation concluded that the problems were caused by using split-time data which caused difficulty when modelling the high excess mortality rate experienced post AMI.

10.3 Fitting on the log cumulative excess hazards scale

Due to the problems encountered in the models investigation a new approach was developed for relative survival in chapter 5 that extended the spline based flexible parametric survival models proposed by Royston and Parmar (135). This approach models on the log cumulative excess hazards scale and does not use split-time data. This overcomes some of the issues with split-time models discussed above and also has a number of other advantages.

The models are much quicker to fit than individual level split-time continuous models. Spline models with five internal knots on individual level data using user defined intervals took 49 seconds to converge, whereas the new approach with the same numbers of knots

fitted in 26 seconds. The models provide a smooth estimate over time and the log cumulative excess hazard scale is stable when using log time. The survival and hazard are obtained analytically, as are predictions i.e. the relative survival rate, excess hazard rate and time-dependent excess hazard rate ratios, and as such are quicker to estimate than other spline models on the log excess hazard scale (37; 38).

The estimates were found to be robust to the frequency and location of knots as shown in section 5.3.3 and 5.4.5 where random knots were fitted to the data, with some at very strange locations that are unlikely to be chosen in practice. With the exception of some very poor knot selections, i.e. at the $(82, 88, 89)$ centiles of all death times, these models all estimated similar model fits. This suggests that model selection criteria is not essential for these models.

A Stata program called `strsrcs` was created and released onto SSC (157) for general use. This allows other research groups to take advantage of the methodology, which also allows the modelling of continuous covariates and time dependent effects. The approach is not without disadvantages. When more than one time dependent covariate is modelled they become dependent on each other due to the scale on which the model is fitted, this means interpretation is more difficult. There are also more common issues with splines to be considered, i.e. knot placement, but these were found to have very little impact on the data analysed here (sections 5.3.1, 5.3.2, 5.4.2 and 5.4.3).

During the simulation the spline models were found to be unbiased, unlike the piecewise GLM models, when calculating the excess hazard rate ratios. This was found to be due to the split times used and the use of a Weibull distribution for the survival times. When the parameters of the Weibull distribution were altered so that the initial drop in excess hazard was not as steep, or when several more splits were added to the first six months of follow-up, the results became very similar. However no adjustments were necessary for the spline based models as they remained unbiased regardless of the Weibull parameters. This shows how the use of split-times can cause problems in the piecewise models which are not evident in the spline based models as the data is not split, and how the shape of the underlying distribution does not create bias in the spline based models either.

The spline models were fitted during the diabetes and glucose investigation and proved to be very successful. There were problems with model convergence when assessing some time-dependent effects but the models appeared sensible even with many covariates, although there are interpretation difficulties as discussed in section 5.5.1 and 9.6.5. An interaction term was fitted along with several time-dependent effects, which have dependent excess hazard rate ratios, and the models converged on feasible estimates. Continuous glucose was also assessed in section 9.8, and while the interpretation is more difficult the ability to assess this is a distinct advantage (140). The use of continuous covariates is also important for prediction (201).

10.4 Cardiovascular disease has a higher prevalence in the population than cancer

In relative survival the expected population are assumed to be disease free, which is not true, but can be assumed as the prevalence of each specific disease (e.g. lung cancer) in the general population is usually low. Heart disease is the UK's single biggest killer and as such this assumption needed to be assessed. This was performed using a simulation study, which found that the excess hazard rate ratios had a low bias regardless of age, high prevalence and high increased odds of mortality due to the disease. The results were not so positive in terms of 5-year relative survival estimates as these showed signs of bias in the very elderly with high prevalence and high odds.

The prevalence values were high at 15% as this value was found for males over seventy-five years old, whereas females and younger patients have a much lower prevalence and as such this value was considered to be high enough for the simulation. The odds of mortality was a more difficult value to determine as there are no clear values published that relate to the increased risk of death that heart disease patients in the general population face compared to people with no history of heart disease. Given that these people are likely to be healthy survivors, as the excess mortality is only of major concern for the first 6-months, a value of twice the odds of death was used. After discussion with the clinical supervisor for the thesis this value was considered as high with only patients with left-ventricular failure achieving this high level of prolonged risk. So both of these values were at the high end of the scale and as such in practice may be too high. This suggests that the 5-year relative survival rates may well be unbiased in practice for patients up to 90 years old.

It is possible to adjust, or at least perform a sensitivity analysis, in order to determine the level of prevalence in the population that would create unacceptable levels of bias in further research.

10.5 Obtaining more up-to-date estimates

Period analysis is a fairly new approach, and has not been used in heart disease, especially as the method generally incorporates relative survival. The spline models developed in this thesis can be extended to incorporate delayed entry (168). This allows the start time of patients to be values other than zero so that a specific calender period can be assessed, which is usually the most recent one. By using survival experience in the most recent period it is possible to obtain more up-to-date estimates of patient relative survival.

Period analysis was assessed in chapter 8 using age groups and showed a good improvement over the standard relative survival approach. It was also found that changing the start

date of the period, from 2002 to 2003 and 2004 had little influence on the estimates and the uncertainty did not rise as would usually be expected. The approach was also applied to the diabetes and glucose model in section 9.9, which showed up to a 20% increase in relative survival for the highest glucose quartile in non-diabetics. This information is very important and has shown the potential that period analysis can bring to heart disease. Simple examples have been shown but these can be easily extended to more complex analyses. It would also be possible to model a trend analysis (180; 181; 182), which would be feasible to incorporate into the spline based model.

There are many treatments for heart disease, i.e. statins, that have only recently been incorporated and as such the effect of time can be very important. It is not easy to differentiate between all of the treatments and year of hospitalisation is often modelled as a covariate as a measure of the progress made over time. The treatment effects are of strong clinical importance and by using period analysis it is possible to obtain up-to-date estimates that are in effect taking account of the most recent medical advances. Period analysis could be very important in determining current heart disease prognosis, future planning of services and feedback to patients.

10.6 Myocardial Infarction National Audit Project

The dataset used for this thesis was collected from one county within England and as such is only a small dataset, certainly in comparison to a cancer registry. The Myocardial Infarction National Audit Project (MINAP) (202; 203) was established in 1999, in response to the national service framework for coronary heart disease, to examine the quality of management of AMI in England and Wales hospitals. However, the MINAP dataset is known to suffer from incomplete data entry and submission, and selectivity in the patients identified for inclusion in MINAP in a large proportion of centres.

Initially MINAP focused on the hospital management of STEMI. However the dataset has been expanded to cover other acute coronary syndromes and the introduction of pre-hospital thrombolysis and primary percutaneous intervention. All hospitals in England and Wales that admit patients with acute coronary syndromes contribute data. At the end of April 2008 the database contained over 596,000 records. This dataset is over one-hundred times larger than the one used for this thesis. In the long term application of the methods used in this thesis should be applied to this large national database.

10.7 Further work

A greater awareness of relative survival in cardiovascular disease is necessary for the methodology to be adopted. In cancer, a relative survival approach is often used as the primary response alongside or instead of all-cause methods. While there are more issues in heart disease the work shown here gives evidence of the potential benefit that could be gained in heart disease. To create the awareness more research in CHD using relative survival needs to be performed and published. The short introductory paper in the European heart journal (2) is available to reference, and this paper addresses some of the basic issues that would be addressed by reviewers. Further publications from this thesis will help this aim, especially the simulation chapter, as this gives evidence against one of the largest arguments against the methodology in heart disease.

With improvements over the last 10-15 years in short-term survival the focus in cardiovascular disease may shift to investigation of longer term survival. This would make relative survival methods more applicable as it holds advantages over both an all-cause and cause-specific approach.

The spline models that were developed on the log cumulative excess hazard scale could be extended to include fractional polynomials which are likely to fit better on non-split-time data. The advantage of this would be the removal of an arbitrary knot selection process, which is often the biggest drawback of splines. However as discussed in chapters 5 and 9 the fitted values are robust to knot location. The spline methodology could be extended further to create an algorithm that would select the best fitting model (199; 200), similarly to MFPs that could then adopt some form of model averaging (127; 130; 131; 129; 132) in order to account for the increased uncertainty of the *best* fitting model. However it was found that in these models, for these data, knot selection was generally not important.

The spline based models are applicable in cancer as they have advantages over the standard methods that use split-time data at the grouped and individual level. This has been shown in a presentation that was given at a workshop on the methods for studying cancer patient survival with application in Stata, to which I was invited and presented this approach. The slides presented are given in appendix B. All of the advantages that apply to CHD will apply to cancer and perhaps the flexible parametric approach may be more readily taken up in cancer, where relative survival methodology is used routinely.

Period analysis also showed potential as relative survival estimates were increased by around 10% in some groups. Publishing the results from the period analysis in this thesis may create a greater awareness of the methodology for cardiovascular researchers, which would be an important step. However, in order for period analysis to be fully accepted in heart disease it is likely that a full retrospective period analysis needs to be performed to give evidence of the ability to predict accurately as has been performed in cancer

(175; 176; 177). The difficulty with this is the lack of data that is available. The dataset used here started in 1993 but does not have enough data in the earlier years to perform this type of analysis. The MINAP dataset may well be able to provide enough data from 1999 onwards, which is in itself less than ten years ago. For this reason a retrospective period analysis using UK data may not be possible for a number of years, which could prevent period analyis from being accepted.

The simulation work performed in this thesis was fairly basic in that the effects of prevalence and the increased odds of mortality were assumed to be constant over time, and the values chosen are likely to be a bit too high in practice. This may have the effect of overestimating the impact on the relative survival estimates. One way around this is to remove all known patients with history of an MI from the expected population, which would be very difficult, and may prove to be overly conservative as suggested by Ederer et al. (26).

There is a second assumption in relative survival which has not been assessed. This assumes that MI does not increase the risk of death from other causes. This could be another problem with the use of relative survival in heart disease as patients suffering an AMI are more likely to be overweight making them more at risk of diabetes and other cardiovascular diseases, such as angina or stroke. These other diseases may not be directly attributable to the index AMI but the link, and therefore the assumption, should be investigated further. It should be noted that this assumption does not invalidate the use of relative survival as it still provides a measure of excess mortality compared to the general population (36).

10.8 Final Conclusions

The application and development of relative survival methods in coronary heart disease is in the early stages and this thesis has shown that relative survival has great potential to have a large impact in this therapeutic area. An introduction to the methodology has been published in heart disease making other researchers in heart disease aware of the methodology so they may too investigate relative survival models. Several models were assessed and problems were encountered but piecewise models fitted very well to the CHD data. New methodology has been developed and shown to fit well to heart disease data and to cancer data. With this new method it is hoped that analysts in other therapeutic areas will begin to fit relative survival models to their, potentially smaller, datasets as before it would have been difficult using split-time data.

Appendix A

STRSRCS.ADO

A.1 STRSRCS.ado

```
*! Version 1.0
*! Chris Nelson 19/MAY/2008

program strsrcs,  sortpreserve
  version 9.2
  if replay()   ///
  {
    syntax  [, DF(int 0) KNOTS(string) BHAZard(varname numeric)
               ORTHOG EForm *]
    if `df' > 0 | "`knots'" != "" | "`bhazard'" != "" ///
    {
      Estimate`0'
    }
    else if (`""`e(cmd)'"' != "strsrcs") error 301
    else Replay `0'
  }
  else Estimate `0'
end


program Estimate, eclass
   st_is 2 analysis
   syntax  [varlist(default=empty)]  [if] [in] ///
   [, DF(int 0) KNOTS(string) SCale(string) BHAZard(varname numeric)
     STRATA(varlist) ORTHOG NOCONStant SNOCONStant INITS(name)
     EForm] ///
   [                                ///
    noLOg                           /// -ml model- options
    noLRTEST                        ///
    Level(integer `c(level)')       /// -Replay- option
    *                               /// -mlopts- options
   ]
   tempvar lnt lnt0
```

```
   tempname initmat

   marksample touse
   mlopts mlopts, ‘options’

   qui replace ‘touse’ = 0  if _st==0

************ DROP EXISTING _RCS AND _D_RCS VARIABLES

   capture drop _rcs*
   capture drop _d_rcs*
   capture drop _s0_rcs*

************ CHECK IF DF OR KNOTS PRESENT

   if "‘df’"== "0" & "‘knots’" == "" ///
   {
     display as error "DF or KNOTS must be specified"
     exit
   }

************ CHECK BASELINE HAZARD IS SPECIFIED

   if "‘bhazard’" == "" ///
   {
      display as error "The baseline hazard must be specified"
      exit
   }

************ CHECK FOR MISSING VALUES

   if ‘touse’ & missing(‘bhazard’)==1 ///
   {
     display as error "The baseline hazard contains missing values"
     exit
   }

************ CHECK TIME ORIGIN FOR DELAYED ENTRY MODELS

   local del_entry = 0
   qui summ _t0 if ‘touse’ , meanonly
   if r(max)>0 ///
   {
     display in green  "note: delayed entry models are being fitted"
     local del_entry = 1
   }


************ DEFINE KNOTS

   gen ‘lnt’ = ln(_t)
   tokenize ‘knots’
   local temp0 : word count ‘knots’

************ CHECK ONLY DF OR KNOTS IS SPECIFIED
```

```
   if ‘df’>0 & ‘temp0’>0 ///
   {
     display as error "Only one of DF OR KNOTS can be specified"
     exit
   }

************ CHECK SCALE OPTION SPECIFIED

   if "‘scale’" =="" ///
   {
     display as error "The scale must be specified"
     exit
   }

************ DEFINE SCALE USING SUBSTR

   if substr("‘scale’", 1, 1)=="h" ///
   {
     local _sc=‘"Hazard"’
   }
   else if substr("‘scale’", 1, 1)=="o" ///
   {
     local _sc=‘"Odds"’
   }
************ CHECK SCALE OPTION SPECIFIED

   else ///
   {
     display as error
        "The scale must be specified as either hazard or odds"
     exit
   }

************ CALCULATE KNOT PLACEMENT

   if ‘temp0’==0 ///
   {
     if ‘df’==2 ///
     {
       qui centile ‘lnt’ if ‘touse’ & _d==1, centile(0 50 100)
       local allknots  ‘r(c_1)’ ‘r(c_2)’ ‘r(c_3)’
     }
     else if ‘df’==3 ///
     {
       qui centile ‘lnt’ if ‘touse’ & _d==1, centile(0 33 67 100)
       local allknots  ‘r(c_1)’ ‘r(c_2)’ ‘r(c_3)’ ‘r(c_4)’
     }
     else if ‘df’==4 ///
     {
       qui centile ‘lnt’ if ‘touse’ & _d==1, centile(0 25 50 75 100)
       local allknots  ‘r(c_1)’ ‘r(c_2)’ ‘r(c_3)’ ‘r(c_4)’ ‘r(c_5)’
     }
     else if ‘df’==5 ///
     {
       qui centile ‘lnt’ if ‘touse’ & _d==1, ///
              centile(0 20 40 60 80 100)
```

```
      local allknots  `r(c_1)' `r(c_2)' `r(c_3)'
                      `r(c_4)' `r(c_5)' `r(c_6)'
    }

    else if `df'==6 ///
    {
      qui centile `lnt' if `touse' & _d==1, ///
              centile(0 17 33 50 67 83 100)
      local allknots  `r(c_1)' `r(c_2)' `r(c_3)'
             `r(c_4)' `r(c_5)' `r(c_6)' `r(c_7)'
    }
    else ///
    {
      display as error "DF must be between 2 and 6"
      exit
    }
  }

  else ///
  {
    qui centile `lnt' if `touse' & _d==1, centile(0 `knots' 100)
    local temp1 = `temp0' + 2
    forvalues i = 1/`temp1' ///
    {
      local allknots "`allknots' `r(c_`i')'"
    }
  }

************ CALL RCS AND RCSDERIV TO CALCULATE SPLINES

  if "`orthog'" != "" ///
  {
    rcs `lnt' if `touse', knots(`allknots') gen(_rcs) ///
         dgen(_d_rcs) orthog
  }

  if "`orthog'" == "" ///
  {
    rcs `lnt' if `touse', knots(`allknots') gen(_rcs) dgen(_d_rcs)
  }

  if `del_entry' == 1 ///
  {
    qui gen `lnt0' = ln(_t0) if _t0>0
    qui rcs `lnt0', knots(`allknots') gen(_s0_rcs)
  }

  local nk : word count `allknots'
  local df = `nk' - 1

************ NOCONSTANT OPTION FOR LINEAR PREDICTOR AND FOR STRATA

  if "`noconstant'" != "" ///
  {
    local xb_nocons = `", noconstant"'
  }
```

```
   if "‘snoconstant’" != "" ///
   {
     local strata_nocons = ‘", noconstant"’
     local xb_nocons = ‘", noconstant"’
   }

************ DEFINE MODEL TERMS

   local loopmax = ‘nk’ - 1
   forvalues i = 1/‘loopmax’ ///
   {
     local rcsterms "‘rcsterms’ _rcs‘i’"
     local drcsterms "‘drcsterms’ _d_rcs‘i’"
     local rcseq "‘rcseq’ (s‘i’: ‘strata’ ‘strata_nocons’)"
   }

************ OBTAIN INITIAL VALUES

   if "‘_sc’" =="Hazard" ///
   {
     if "‘inits’" == "" ///
     {
       display in green "Obtaining Initial Values"
       qui ml model lf strsrcs_mlh ///
         (xb: _rcs1 ‘bhazard’ = ‘strata’ ‘varlist’ ‘xb_nocons’) ///
       (s1: ‘strata’ ‘strata_nocons’) , ‘mlopts’ maximize
       matrix ‘initmat’ = e(b)
       display in green "Initial Values Obtained"
     }
     else matrix ‘initmat’ = ‘inits’
   }
   else if "‘_sc’" =="Odds" ///
   {
     if "‘inits’" == "" ///
     {
       display in green "Obtaining Initial Values"
       qui ml model lf strsrcs_mlo
          (xb: _rcs1 ‘bhazard’ = ‘strata’ ‘varlist’ ‘xb_nocons’) ///
       (s1: ‘strata’ ‘strata_nocons’) , ‘mlopts’ maximize
       matrix ‘initmat’ = e(b)
       display in green "Initial Values Obtained"
     }
     else matrix ‘initmat’ = ‘inits’
   }

************ FIT THE MODEL

   if "‘_sc’" =="Hazard" ///
   {
      ml model lf strsrcs_mlh ///
      (xb:‘rcsterms’ ‘bhazard’=‘strata’ ‘varlist’ ‘xb_nocons’)   ///
      ‘rcseq’ if ‘touse’ ,  ///
      init(‘initmat   ’) ///
      ‘mlopts’ ///
      maximize
```

```
   }
   else if "`_sc'" =="Odds" ///
   {
      ml model lf strsrcs_mlo ///
      (xb:`rcsterms' `bhazard'=`strata' `varlist' `xb_nocons')   ///
      `rcseq' if `touse' ,  ///
      init(`initmat   ') ///
      `mlopts' ///
      maximize
   }

   ereturn local rcs_xb `rcsterms'
   ereturn local drcs_xb `drcsterms'
   ereturn local knots `allknots'
   ereturn local df `df'
   ereturn local predict strsrcs_pred
   ereturn local cmd strsrcs
   ereturn local depvar "_d _t"
   ereturn local strata `strata'
   ereturn local varlist `varlist'
   ereturn local bhazard `bhazard'
   ereturn local scale `_sc'
   Replay, level(`level') `eform'
end

program Replay
        syntax [, Level(int `c(level)') EForm]
        ml display, level(`level') `eform'
end
```

A.2 RCS.ado

```
*! version 1.0
*! orthog option added by Paul Lambert 28/Sep/2007
*! Chris Nelson 16/May/2008


program define rcs
  version 9.2
  syntax  varlist(max=1) [if] [in]                ///
     [, Gen(string) DGen(string) Knots(string) Orthog]      ///
     marksample touse


*************** CHECK FOR GENERATE NAME

  if "`gen'" == "" ///
  {
     di in red "Must specify name for cubic splines basis"
     exit 198
  }
  tempvar x
  gen `x'=`varlist'
```

```
*************** CREATE FIRST REGRESSION VARIABLE RCS1 = X
  tempvar rcs1 orcs1
  quietly gen double `rcs1' = `x' if `touse'

*************** CREATE FIRST REGRESSION VARIABLE DRCS1 = 1

  if "`dgen'" != "" ///
  {
    tempvar drcs1 odrcs1
    quietly gen double `drcs1' = 1 if `touse'
  }

*************** EXTRACT KNOTS

  local nk : word count `knots'
  local i = 1
  tokenize "`knots'"
  while "``i''" != "" ///
  {
     local k`i' ``i''
     local i = `i'+1
  }

*************** CREATE REGRESSION VARIABLES

  local kmin = `k1'
  local kmax = `k`nk''
  local interior  = `nk'-1
  local rcslist `rcs1'
  local orcslist `orcs1'
  local drcslist `odrcs1'
  local odrcslist `odrcs1'

  forvalues j=2/`interior' ///
  {
    ******** CREATE LAMBDA
    local lambda = (`kmax' - `k`j'')/(`kmax' - `kmin')
    ******** CREATE RCS
    tempvar rcs`j' orcs`j'
    quietly gen double `rcs`j'' = ((`x'-`k`j'')^3)*(`x'>`k`j'') - ///
              `lambda'*((`x'-`kmin')^3)*(`x'>`kmin') - ///
              (1-`lambda')*((`x'-`kmax')^3)*(`x'>`kmax')  if `touse'
    local rcslist `rcslist' `rcs`j''
    local orcslist `orcslist' `orcs`j''

*************** IF DERIVATIVES ARE SPECIFIED

    if "`dgen'"!="" ///
    {
      ******** CREATE DRCS
      tempvar drcs`j' odrcs`j'
      qu gen double `drcs`j'' = (3*(`x'-`k`j'')^2)*(`x'>`k`j'') - ///
             `lambda'*(3*(`x'-`kmin')^2)*(`x'>`kmin') - ///
             (1-`lambda')*(3*(`x'-`kmax')^2)*(`x'>`kmax')  if `touse'
      local drcslist `drcslist' `drcs`j''
```

```
      local odrcslist ‘odrcslist’ ‘odrcs‘j’’

    }
  }

  if "‘orthog’" != "" {
    tempname R Rinv
    orthog ‘rcslist’ if ‘touse’, gen(‘orcslist’) matrix(‘R’)
    if "‘dgen’" != "" {
       matrix ‘Rinv’ = inv(‘R’)
       forvalues i = 1/‘interior’ {
         local ortheqlist 0
           forvalues j = 1/‘i’ {
             local ortheqlist "‘ortheqlist’ + ‘Rinv’[‘j’,‘i’]*‘drcs‘j’’"
           }
           gen ‘odrcs‘i’’  if ‘touse’ = ‘ortheqlist’
         }
       }
    }

/* create returned variables */

   forvalues i = 1/‘interior’ {
     if "‘orthog’" != "" {
       local addo o
     }
     gen ‘gen’‘i’ if ‘touse’ = ‘‘addo’rcs‘i’’
     if "‘dgen’" != "" {
       gen ‘dgen’‘i’ if ‘touse’ = ‘‘addo’drcs‘i’’
     }
   }

   if "‘dgen’"!="" ///
   {
     di in green "Variables ‘gen’1 to ‘gen’‘interior’ and
         ‘dgen’1 to ‘dgen’‘interior’ were created"
   }
   else ///
   {
     di in green "Variables ‘gen’1 to ‘gen’‘interior’ were created"
   }
end
```

A.3 STRSRCS_MLH.ado

```
program strsrcs_mlh
   version 9.1
   args lnf xb s1 s2 s3 s4 s5 s6 s7 s8 s9 s10 s11 s12
   tempvar ht st  st0
   local p: word count $ML_y
   local q=‘p’-1
   local del_entry = 0
   qui summ _t0 , meanonly
```

```
   if r(max)>0 local del_entry = 1
   forvalues i = 1/‘q’ ///
   {
      local ds "‘ds’ ‘s‘i’’ * _d${ML_y‘i’}"
      if ‘i’ != ‘q’ local ds "‘ds’ + "
      local sp "‘sp’ ‘s‘i’’ * ${ML_y‘i’}"
      if ‘i’ != ‘q’ local sp "‘sp’ + "
      local sp0 "‘sp0’ ‘s‘i’’ * _s0${ML_y‘i’}"
      if ‘i’ != ‘q’ local sp0 "‘sp0’ + "
   }
   quietly generate double ‘st’=exp(-exp(‘sp’ + ‘xb’))
   quietly generate double ‘ht’=${ML_y‘p’}+(1/_t)*(‘ds’)*exp(‘sp’+‘xb’)
   qui replace ‘lnf’ = _d*ln(‘ht’)+ln(‘st’)
   if ‘del_entry’ == 1 ///
   {
      quietly generate double ‘st0’ = exp(-exp(‘sp0’ + ‘xb’)) if _t0>0
      qui replace ‘lnf’ = ‘lnf’ - ln(‘st0’) if _t0>0
   }
end
```

A.4 STRSRCS_MLO.ado

```
program strsrcs_mlo
   version 9.1
   args lnf xb s1 s2 s3 s4 s5 s6 s7 s8 s9 s10 s11 s12
   tempvar ht st  st0
   local p: word count $ML_y
   local q=‘p’-1
   local del_entry = 0
   qui summ _t0 , meanonly
   if r(max)>0 local del_entry = 1
   forvalues i = 1/‘q’ ///
   {
      local ds "‘ds’ ‘s‘i’’ * _d${ML_y‘i’}"
      if ‘i’ != ‘q’ local ds "‘ds’ + "
      local sp "‘sp’ ‘s‘i’’ * ${ML_y‘i’}"
      if ‘i’ != ‘q’ local sp "‘sp’ + "
      local sp0 "‘sp0’ ‘s‘i’’ * _s0${ML_y‘i’}"
      if ‘i’ != ‘q’ local sp0 "‘sp0’ + "
   }
   quietly generate double ‘st’=1/(1+exp(‘sp’ + ‘xb’))
   quietly generate double ‘ht’=${ML_y‘p’} + (1/_t)*(‘ds’)* ///
    exp(‘sp’ + ‘xb’)*(1/(1+exp(‘sp’ + ‘xb’)))
   qui replace ‘lnf’ = _d*ln(‘ht’)+ln(‘st’)
   if ‘del_entry’ == 1 ///
   {
      quietly generate double ‘st0’ = 1/(1+exp(‘sp0’ + ‘xb’)) if _t0>0
      qui replace ‘lnf’ = ‘lnf’ - ln(‘st0’) if _t0>0
   }
end
```

A.5 STRSRCS_PRED.ado

```
*! version 1.0
*! Chris Nelson 16/MAY/2008

program strsrcs_pred
   version 9.2
   syntax newvarname [if] [in], [Survival] [Hazard] [HRatio] [CI] [LEVel(string)]
   marksample touse, novarlist
   local newvarname `varlist'
   qui count if `touse'
   if r(N)==0 ///
   {
     error 2000          /* no observations */
   }

*************** CHECK THAT EITHER SURVIVAL OR HAZARD IS SPECIFIED

   if `"`survival'"'=="" & `"`hazard'"' =="" & `"`hratio'"' ==""  ///
   {
     display as error "You must specify either the survival,
                                    hazard or hratio option"
     exit 198
   }
   if (`"`survival'"'!="" & `"`hazard'"' !="") ///
   {
     display as error "Only one of the survival,
                  hazard or hratio options can be specified"
     exit 198
   }
   if (`"`survival'"'!="" & `"`hratio'"' !="") ///
   {
     display as error "Only one of the survival,
                      hazard or hratio options can be specified"
     exit 198
   }
   if (`"`hazard'"'!="" & `"`hratio'"' !="") ///
   {
     display as error "Only one of the survival,
                      hazard or hratio options can be specified"
     exit 198
   }

*************** LEVEL FOR CONFIDENCE INTERVAL

   tokenize 'level'
   local level_n : word count `level'
   if `level_n' == 0 ///
   {
      local levelopt `"level(95)"'
   }
   else if `level_n' == 1 ///
   {
      local levelopt `"level(`level')"'
   }
```

```
*************** OBTIAN LINEAR PREDICTORS

   local p: word count `e(rcs_xb)'
   forvalues i=1/`p' ///
   {
      local rcs "`rcs' xb(s`i')*_rcs`i'"
      if `i' != `p' local rcs "`rcs' + "
      local drcs "`drcs' xb(s`i')*_d_rcs`i'"
      if `i' != `p' local drcs "`drcs' + "
   }

*************** PREDICT SURVIVAL

   if "`e(scale)'" =="Hazard"  ///
   {
      if "`survival'" != "" ///
      {
         tempvar lnH lnHlci lnHuci
         if `"`ci'"' != "" ///
         {
            local survci "ci(`lnHlci' `lnHuci')"
         }
         qui predictnl double `lnH' = xb(xb) + `rcs' ///
             if `touse', `survci' `levelopt'
         qui gen double `newvarname' = exp(-exp(`lnH')) if `touse'
         if `"`ci'"' != "" ///
         {
            qui gen double `newvarname'_lci = exp(-exp(`lnHlci')) if `touse'
            qui gen double `newvarname'_uci = exp(-exp(`lnHuci')) if `touse'
         }
*************** REPORT NEW VARIABLE CREATION

         display in green "note: New variable `newvarname' has been created"
         if `"`ci'"' != "" ///
         {
            display in green "      lower bound in `newvarname'_lci"
            display in green "      upper bound in `newvarname'_uci"
         }
      }
   }
   else if "`e(scale)'" =="Odds" ///
   {
      if "`survival'" != "" ///
      {
         tempvar lnO lnOlci lnOuci
         if `"`ci'"' != "" ///
         {
            local osurvci "ci(`lnOlci' `lnOuci')"
         }
         qui predictnl double `lnO' = xb(xb) + `rcs' if `touse', ///
             `osurvci' `levelopt'
         qui gen double `newvarname' = 1/(1+exp(`lnO')) if `touse'
         if `"`ci'"' != "" ///
         {
            qui gen double `newvarname'_lci = 1/(1+exp(`lnOlci')) if `touse'
```

```
            qui gen double ‘newvarname’_uci = 1/(1+exp(‘lnOuci’)) if ‘touse’
         }

*************** REPORT NEW VARIABLE CREATION

         display in green "note: New variable ‘newvarname’ has been created"
         if ‘"‘ci’"’ != "" ///
         {
            display in green "      lower bound in ‘newvarname’_lci"
            display in green "      upper bound in ‘newvarname’_uci"
         }
      }
   }

*************** TEMPORARY TIMEVAR

   tempvar timevar
   quietly generate double ‘timevar’=_t

*************** PREDICT HAZARD

   if "‘e(scale)’" =="Hazard" ///
   {
      if "‘hazard’" != "" ///
      {
         tempvar  lh lhlci lhuci
         if ‘"‘ci’"’ != "" ///
         {
            local hazci "ci(‘lhlci’ ‘lhuci’)"
         }
         qui predictnl double ‘lh’=-ln(‘timevar’) + ln(‘drcs’) +  ///
             (xb(xb) +‘rcs’) if ‘touse’, ‘hazci’ ‘levelopt’
         qui gen double ‘newvarname’=exp(‘lh’)
         if ‘"‘ci’"’ != "" ///
         {
            qui gen double ‘newvarname’_lci = exp(‘lhlci’) if ‘touse’
            qui gen double ‘newvarname’_uci = exp(‘lhuci’) if ‘touse’
         }

*************** REPORT NEW VARIABLE CREATION

         display in green "note: New variable ‘newvarname’ has been created"
         if ‘"‘ci’"’ != "" ///
         {
            display in green "      lower bound in ‘newvarname’_lci"
            display in green "      upper bound in ‘newvarname’_uci"
         }
      }
   }
   else if "‘e(scale)’" =="Odds" ///
   {
      if "‘hazard’" != "" ///
      {
         tempvar  lo lolci louci
         if ‘"‘ci’"’ != "" ///
         {
```

```
            local ohazci "ci(‘lolci’ ‘louci’)"
         }
         qui predictnl double ‘lo’=-ln(‘timevar’) + ln(‘drcs’) +  ///
                     (xb(xb) +‘rcs’) -ln(1+exp(xb(xb) +‘rcs’))  ///
                     if ‘touse’, ‘ohazci’ ‘levelopt’

         qui gen double ‘newvarname’=exp(‘lo’)
         if ‘"‘ci’"’ != "" ///
         {
            qui gen double ‘newvarname’_lci = exp(‘lolci’) if ‘touse’
            qui gen double ‘newvarname’_uci = exp(‘louci’) if ‘touse’
         }

*************** REPORT NEW VARIABLE CREATION

         display in green "note: New variable ‘newvarname’ has been created"
         if ‘"‘ci’"’ != "" ///
         {
            display in green "      lower bound in ‘newvarname’_lci"
            display in green "      upper bound in ‘newvarname’_uci"
         }
      }
   }

*************** PREDICT HAZARD RATIO

   if "‘e(scale)’" =="Hazard" ///
   {
      if "‘hratio’" != ""  ///
      {
         local nk : word count ‘e(strata)’
         local i=1
         tokenize "‘e(strata)’"
         while "‘‘i’’" != "" ///
         {
            local k‘i’ ‘‘i’’
            local i=‘i’+1
         }

*************** CREATE REGRESSION VARIABLES

         tempvar  t
         forvalues i=1/‘e(df)’ ///
         {
            local rcs0 "‘rcs0’ [s‘i’][_cons]*_rcs‘i’"
            forvalues j= 1/‘nk’ ///
            {
               tempvar lhr‘j’ lhr‘j’_lci lhr‘j’_uci
               local rcs‘j’ "‘rcs‘j’’ [s‘i’][_cons]*_rcs‘i’ + ///
                     [s‘i’][‘‘j’’]*_rcs‘i’"
            }
            if ‘i’ != ‘e(df)’ ///
            {
               local rcs0 "‘rcs0’ + "
               forvalues j= 1/‘nk’ ///
               {
```

```
               local rcs‘j’ "‘rcs‘j’’ + "
            }
         }
         local drcs0 "‘drcs0’ [s‘i’][_cons]*_d_rcs‘i’"
         forvalues j= 1/‘nk’ ///
         {
            local drcs‘j’ "‘drcs‘j’’ [s‘i’][_cons]*_d_rcs‘i’ + ///
                        [s‘i’][‘‘j’’]*_d_rcs‘i’"
         }
         if ‘i’ != ‘e(df)’ ///
         {
            local drcs0 "‘drcs0’ + "
            forvalues j= 1/‘nk’ ///
            {
               local drcs‘j’ "‘drcs‘j’’ + "
            }
         }
      }
      gen ‘t’=_t
      forvalues j= 1/‘nk’ ///
      {
         qui predictnl double ‘lhr‘j’’= (-ln(‘t’) + ln(‘drcs‘j’’) + ///
               ([xb][_cons]+ [xb][‘‘j’’] +‘rcs‘j’’)) - ///
               (-ln(‘t’) + ln(‘drcs0’) + ([xb][_cons] +‘rcs0’)) ///
               if ‘touse’ , ci(‘lhr‘j’_lci’ ‘lhr‘j’_uci’) ‘levelopt’
         qui gen double ‘newvarname’‘j’=exp(‘lhr‘j’’)
         if ‘"‘ci’"’ != "" ///
         {
            qui gen double ‘newvarname’‘j’_lci=exp(‘lhr‘j’_lci’)  if ‘touse’
            qui gen double ‘newvarname’‘j’_uci=exp(‘lhr‘j’_uci’)  if ‘touse’
         }
        display in green "note: new variable ‘newvarname’‘j’ has been created"
         if ‘"‘ci’"’ != "" ///
         {
            display in green "      lower bound in ‘newvarname’‘j’_lci"
            display in green "      upper bound in ‘newvarname’‘j’_uci"
         }
      }
   }
   else if "‘e(scale)’" =="Odds" ///
   {
      if "‘hratio’" != "" ///
      {
         display "NOT YET AVAILABLE !!"
      }
   }

end
```

Appendix B

Invited talk: Using Spline Functions on the Log Cumulative Hazard Scale

Using Spline Functions on the Log Cumulative Hazard scale, the STRSRCS Command

Christopher P Nelson
Center for Biostatistics and Genetic Epidemiology
University of Leicester
cn46@le.ac.uk

Methods for studying cancer patient survival, with application in Stata:
September 6 2007

University of Leicester

Contents

Issues with current methods

- Current methods in practice split the timescale
- Piecewise models could be considered as biologically implausible
- Many parameters for time dependent effects
- Collapsed data does not allow modelling of continuous covariates
- Can lead to problems in *small datasets, e.g. zero cells*
- Current spline methods that do not split time use numerical integration
  - this slows the process
  - Giorgi et al. (2003)
  - Bolard et al. (2002)

A difference in scale

- Current relative survival models are fitted to the log excess hazard scale, i.e.

$$h(t) = h^*(t) + \exp(\beta z)$$

- β = a vector of parameters to be estimated for covariates z
- whereas we are interested in modelling on the log cumulative excess hazard scale, i.e.

$$H(t) = H^*(t) + \Lambda(t)$$
$$\ln\left(-\ln R(t;z)\right) = \ln\left(\Lambda(t)\right) = \ln\left(\Lambda_0(t)\right) + \beta z$$

- $H(t)$ = cumulative overall hazard, $H^*(t)$ = cumulative expected hazard , $\Lambda(t)$ = cumulative excess hazard & $\Lambda_0(t)$ = baseline cumulative excess hazard
- Note that proportional cumulative excess hazards implies proportional excess hazards

A Flexible Parametric Model for Survival analyses Royston and Parmar (2002)

- Time by covariate interactions easy (non-PH)
- Ability to model the baseline hazard
- Proportional Hazards ($\theta \to 0$) or Proportional Odds ($\theta = 1$) models are available
- Further extensions to allow non-proportional effects of some or all of the covariates (other values of θ)
- And can be extended to relative survival to give the following likelihood

$$L_i = d_i \ln\left[h^*(t_i) + \frac{1}{t_i}\frac{ds(x_i;\gamma)}{dx_i}\exp(\eta_i)\right] - exp(\eta_i)$$

- Hazard and survival are both obtained analytically.

Stata Code

Merge expected mortality

```
gen _age = int(min(agehosp + _t,99))
gen _year = int(min(yearhosp + _t,2006))
sort _year sex  _age
merge _year sex _age using popmort, nokeep
gen rate = -ln(prob)
```

- Building a model

PEH model

```
xi: strsrcs i.year8594 i.sex i.agegrp, bhaz(rate) scale(hazard) df(4)
```

Age non-PEH model

```
xi: strsrcs i.year8594 i.sex , bhaz(rate) scale(hazard) df(4) strata(i.agegrp)
```

Stata Output

PEH output

```
Variables _rcs1 to _rcs4 and _d_rcs1 to _d_rcs4 were created
                                                Number of obs   =      15564
                                                Wald chi2(5)    =     256.69
Log likelihood = -20984.928                     Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.       z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
xb           |
_1year8594_1 |  -.1744355   .0248468    -7.02   0.000    -.2231342   -.1257367
      _1sex_2 |    -.00843   .0255469    -0.33   0.741    -.0585009    .041641
  _1agegrp_1 |   .0730139   .0626038     1.17   0.243    -.0496873    .1957152
  _1agegrp_2 |   .2006328   .0578447     3.47   0.001     .0872592    .3140064
  _1agegrp_3 |   .5164176   .0586398     8.81   0.000     .4014858    .6313495
       _cons |   .1206518   .0843953     1.43   0.153      -.04476    .2860636
-------------+----------------------------------------------------------------
s1           |
       _cons |   1.354656   .0317743    42.63   0.000     1.292379    1.416932
-------------+----------------------------------------------------------------
s2           |
       _cons |   .1464346   .0107499    13.62   0.000     .1253652    .1675039
-------------+----------------------------------------------------------------
s3           |
       _cons |  -.1662767   .0158295   -10.50   0.000     -.197302   -.1352515
-------------+----------------------------------------------------------------
s4           |
       _cons |   .1074269   .0106424    10.09   0.000     .0865682    .1282855
------------------------------------------------------------------------------
```

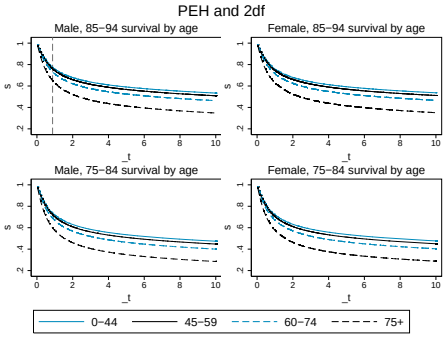

Example relative survival plot

Obtaining relative survival estimates

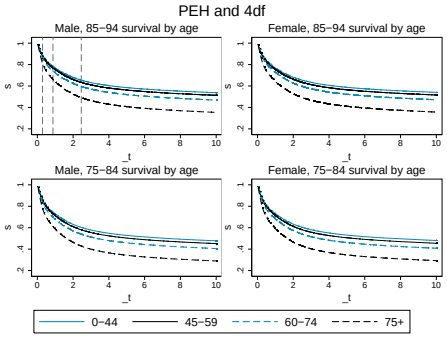
```
predict survival , survival


note: New variable survival has been created
```

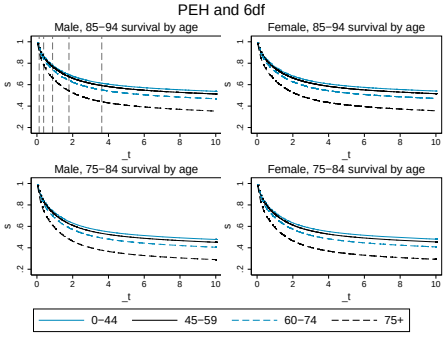

Example relative survival plot



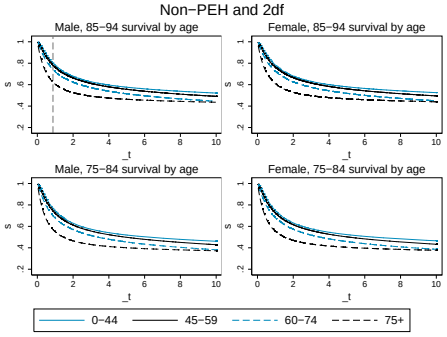
Example relative survival plot



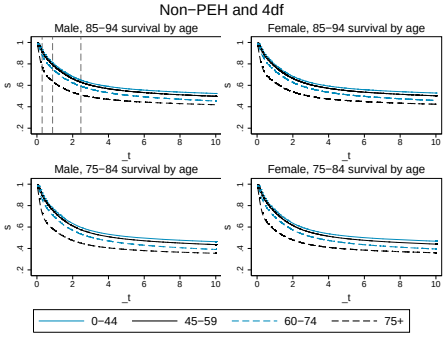
Example relative survival plot



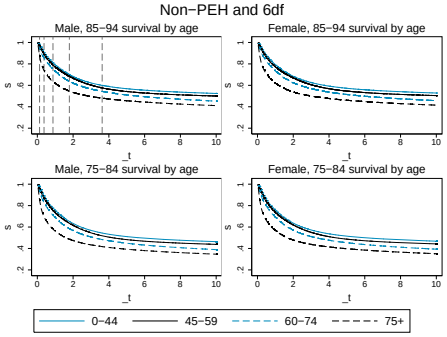
Example relative survival plot



Example relative survival plot



Example relative survival plot



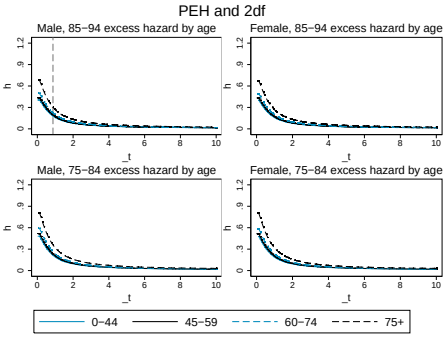
Example excess mortality plot

Obtaining excess mortality estimates

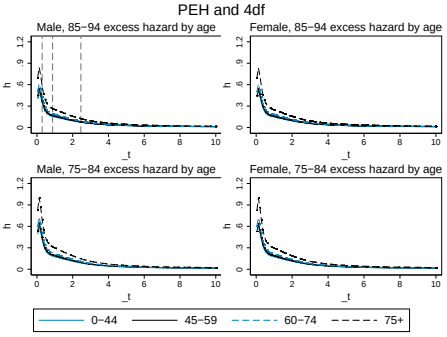
```
predict hazard , hazard


note: New variable hazard has been created
```

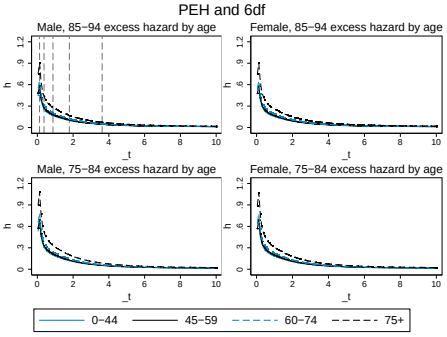

Example excess mortality plot



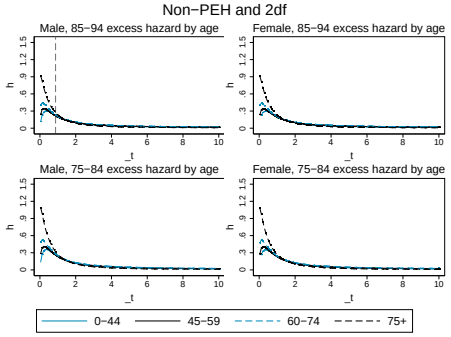
Example excess mortality plot



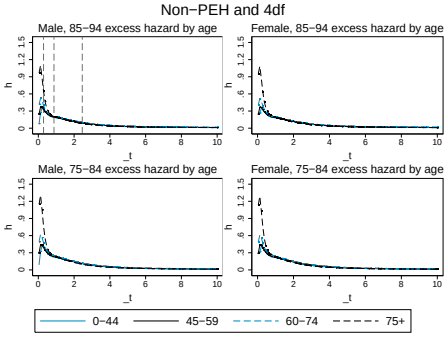
Example excess mortality plot



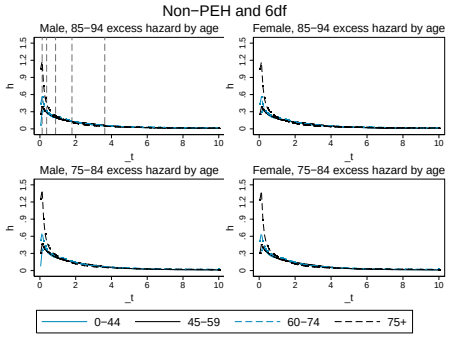
Example excess mortality plot



Example excess mortality plot



Example excess mortality plot



Model selection

Using the AIC

	AIC	
DF	PEH	non-PEH
2	42172.07*	41936.92
4	41989.86	41720.98
6	41937.64	41670.46†

- $*$ took 18.0 seconds to run
- † took 76.4 seconds to run

Excess Hazard Rate Ratios

- Calculating ratios is not easy as it is not currently implemented in the program

Stata code

```
. program hratio
  1.         tempvar lhr1 lhr1_lci lhr1_uci t lhr2 lhr2_lci lhr2_uci lhr3 lhr3_lci lhr3_uci
  2.
.         forvalues i=1/`e(df)' ///
>         {
  3.                 local rcs0 "`rcs0' [s`i']_cons]*_rcs`i'"
  4.                 local rcs1 "`rcs1' [s`i']_cons]*_rcs`i' + [s`i']agegp2]*_rcs`i'"
  5.                 local rcs2 "`rcs2' [s`i']_cons]*_rcs`i' + [s`i']agegp3]*_rcs`i'"
  6.                 local rcs3 "`rcs3' [s`i']_cons]*_rcs`i' + [s`i']agegp4]*_rcs`i'"
  7.
.
.
.               if `i' != `e(df)' ///
>               {
  8.                         local rcs0 "`rcs0' + "
  9.                         local rcs1 "`rcs1' + "
 10.                         local rcs2 "`rcs2' + "
 11.                         local rcs3 "`rcs3' + "
 12.                 }
 ...
REPEAT FOR drcs
 23.                 }
```

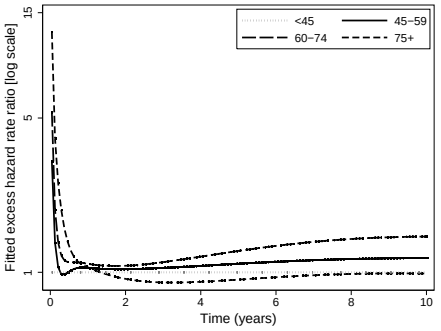

Excess Hazard Rate Ratios

Stata code continued

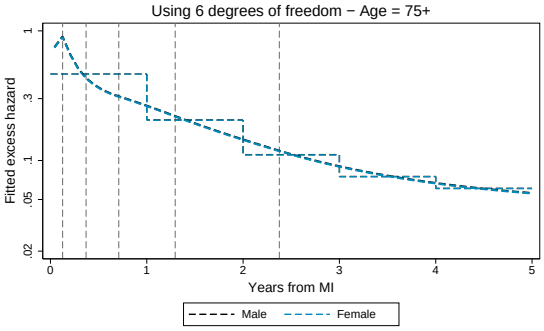
```
.         gen `t'=_t
 25.      qui predictnl double `lhr1'= (-ln(`t') + ln(`drcs1') + ([xb][_cons]+ [xb][agegp2] +`rcs1')) - ///
>                         (-ln(`t') + ln(`drcs0') + ([xb][_cons] +`rcs0')) , ci(`lhr1_lci' `lhr1_uci')
 26.      qui predictnl double `lhr2'= (-ln(`t') + ln(`drcs2') + ([xb][_cons]+ [xb][agegp3] +`rcs2')) - ///
>                         (-ln(`t') + ln(`drcs0') + ([xb][_cons] +`rcs0')) , ci(`lhr2_lci' `lhr2_uci')
 27.      qui predictnl double `lhr3'= (-ln(`t') + ln(`drcs3') + ([xb][_cons]+ [xb][agegp4] +`rcs3')) - ///
>                         (-ln(`t') + ln(`drcs0') + ([xb][_cons] +`rcs0')) , ci(`lhr3_lci' `lhr3_uci')
 28.         qui gen double hratio`e(df)'1=exp(`lhr1')
 29.         qui gen double hratio`e(df)'1_lci=exp(`lhr1_lci')
 30.         qui gen double hratio`e(df)'1_uci=exp(`lhr1_uci')
 31.         qui gen double hratio`e(df)'2=exp(`lhr2')
 32.         qui gen double hratio`e(df)'2_lci=exp(`lhr2_lci')
 33.         qui gen double hratio`e(df)'2_uci=exp(`lhr2_uci')
 34.         qui gen double hratio`e(df)'3=exp(`lhr3')
 35.         qui gen double hratio`e(df)'3_lci=exp(`lhr3_lci')
 36.         qui gen double hratio`e(df)'3_uci=exp(`lhr3_uci')
 37.
.         di "note: new variable hratio`e(df)'1 , hratio`e(df)'2 & hratio`e(df)'3 have been created"
 38. end
```

- The new variables can be plotted
- Potential issue when there are two or more time dependent effects

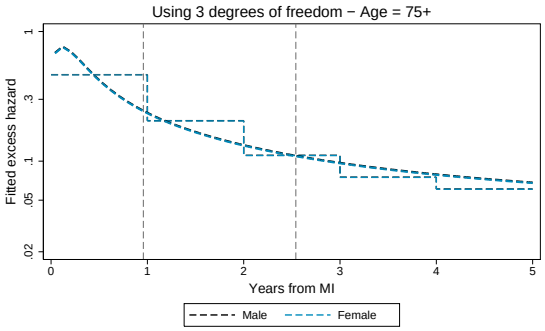
Hazard Rate Ratios



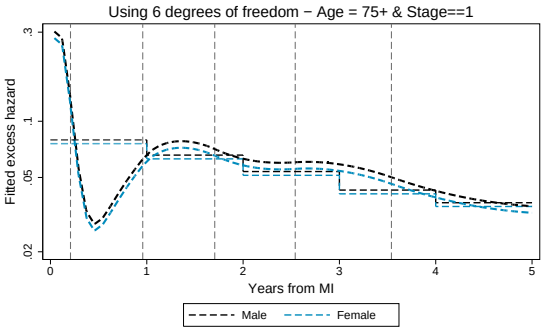
Compare to GLM, Dickman et al. (2004)



Compare to GLM, Dickman et al. (2004)



Compare to GLM, Dickman et al. (2004)



Age continuous

- For a PEH or non-PEH model we could assume age is linear

Stata output

```
strsrcs agehosp , df(4) bhazard(rate) scale(h)

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
xb           |
    agehosp  |   .0131148   .0030013     4.37   0.000     .0072323    .0189972
       _cons |  -3.190079   .2206072   -14.46   0.000    -3.622461   -2.757697
-------------+----------------------------------------------------------------
s1           |
       _cons |   .7594209   .0480404    15.81   0.000     .6652633    .8535784
-------------+----------------------------------------------------------------
s2           |
       _cons |   .1064705   .0297691     3.58   0.000     .0481241    .1648169
-------------+----------------------------------------------------------------
s3           |
       _cons |  -.4495315   .0972077    -4.62   0.000    -.6400551   -.2590078
-------------+----------------------------------------------------------------
s4           |
       _cons |    .564345   .1118058     5.05   0.000     .3452096    .7834805
------------------------------------------------------------------------------
```

- i.e. for every increase in age by one, the excess hazard rate increases by $\approx 14\%$

Age continuous

- Or we could assume that age could also be modelled using splines

```
. rcs agehosp, knots(20 40 60 70 95) gen(agercs)

Variables agercs1 to agercs4 were created

. strsrcs agercs* , df(4) bhazard(rate) scale(h)
```

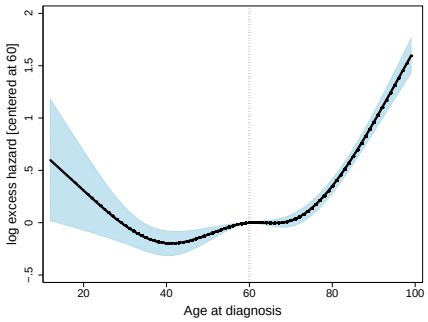
- To get a prediction we can center at age 60 and use predictnl

```
. summ agercs* if agehosp == 60

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
     agercs1 |       125          60           0        60         60
     agercs2 |       125   -38933.33           0  -38933.33  -38933.33
     agercs3 |       125   -29866.67           0  -29866.67  -29866.67
     agercs4 |       125   -21333.33           0  -21333.33  -21333.33

predictnl xb2 = [xb][agercs1]*(agercs1 - 60) + [xb][agercs2]*(agercs2 + 38933.33) ///
              + [xb][agercs3]*(agercs3 +29866.67) + [xb][agercs4]*(agercs4 + 21333.33), ci(lo hi)
note: Confidence intervals calculated using Z critical values
```

Age continuous



Issues with the method

- Dipping early on - artefact??
- Proportional Odds modeling in relative survival
- Assessing more than one time dependent covariate

References I

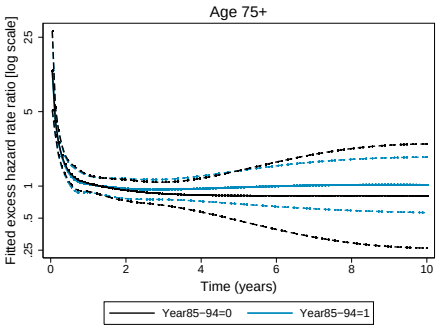
Bolard, P., Quantin, C., Abrahamowicz, M., Esteve, J., Giorgi, R., Chadha-Boreham, H., Binquet, C., and Faivre, J. (2002). Assessing time-by-covariate interactions in relative survival models using restrictive cubic spline functions. *Journal of Cancer Epidemiology & Prevention*, 7(3):113–122.

Dickman, P. W., Sloggett, A., Hills, M., and Hakulinen, T. (2004). Regression models for relative survival. *Statistics in Medicine*, 23(1):51–64.

Giorgi, R., Abrahamowicz, M., Quantin, C., Bolard, P., Esteve, J., Gouvernet, J., and Faivre, J. (2003). A relative survival regression model using b-spline functions to model non-proportional hazards. *Statistics in Medicine*, 22(17):2767–2784.

Royston, P. and Parmar, M. (2002). Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine*, 21(15):2175–2197.

Hazard Rate Ratios



Appendix C

Statistics in Medicine: Flexible parametric Models for Relative Survival, with Application in Coronary Heart Disease

STATISTICS IN MEDICINE
Statist. Med. 2007; **26**:5486–5498
Published online 24 September 2007 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.3064

Flexible parametric models for relative survival, with application in coronary heart disease

Christopher P. Nelson[1,*,†], Paul C. Lambert[1], Iain B. Squire[2] and David R. Jones[1]

[1]*Centre for Biostatistics and Genetic Epidemiology, Department of Health Sciences, University of Leicester, Leicester, U.K.*
[2]*Department of Cardiovascular Sciences, University of Leicester, Leicester, U.K.*

SUMMARY

Relative survival is frequently used in population-based studies as a method for estimating disease-related mortality without the need for information on cause of death. We propose an extension to relative survival of a flexible parametric model proposed by Royston and Parmar for censored survival data. The model provides smooth estimates of the relative survival and excess mortality rates by using restricted cubic splines on the log cumulative excess hazard scale. The approach has several advantages over some of the more standard relative survival models, which adopt a piecewise approach, the main being the ability to model time on a continuous scale, the survival and hazard functions are obtained analytically and it does not use split-time data. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS: relative survival; restricted cubic splines; survival analysis; heart disease

1. INTRODUCTION

Survival data are often modelled using a Cox proportional hazards model, which has the advantage of estimating covariate effects as log hazard ratios without the need to estimate the baseline hazard. However, the behaviour of the hazard function may be of medical interest because it is directly related to the time course of an illness. The baseline hazard rate can also help in understanding the natural history of the disease by the way the hazard rate changes over time. Selected parametric models such as the Weibull model are an alternative to the Cox model. However, these models are not as popular in practice due to concerns over the restrictions imposed on the shape of the hazard function. A flexible parametric model using restricted cubic splines was proposed by Royston and

*Correspondence to: Christopher P. Nelson, Centre for Biostatistics and Genetic Epidemiology, Department of Health Sciences, University of Leicester, Leicester, U.K.
†E-mail: cn46@le.ac.uk

Contract/grant sponsor: British Heart Foundation PhD Studentship; contract/grant number: FS/05/080/19415

Received 1 August 2007
Accepted 8 August 2007

Parmar [1] for censored survival data, enabling the baseline hazard to be directly modelled. The restricted cubic splines offer greater flexibility in the shape of the hazard function when compared with standard parametric models. In addition they can be easily extended to time-dependent effects. One of the key advantages of the method is that the survival and hazard functions are obtained analytically speeding up computational time over methods adopting numerical integration. The application of the method has been limited so far [2–4] although it is available to run using Stata [5]. Here we extend the flexible parametric models to relative survival models.

Relative survival $R(t)$ is defined as the ratio of the observed, all-cause, survival $S(t)$ with the expected survival $S^*(t)$ [6]:

$$R(t) = \frac{S(t)}{S^*(t)} \tag{1}$$

Hence, that on the hazard scale the overall hazard rate, $h(t)$ is given by

$$h(t) = h^*(t) + \lambda(t) \tag{2}$$

where $h^*(t)$ is the expected mortality/hazard rate and $\lambda(t)$ is the excess mortality rate associated with the disease of interest. The expected mortality rate is assumed to be known, often obtained from routinely collected data and matched to the population under study by age, sex, time period and potentially other factors. Relative survival is usually estimated using life tables, but there has been recent interest in models for relative survival.

Comparison with an expected mortality in the population is the key difference between relative survival and the standard methods in that it provides a measure of mortality associated with a disease without the need for information on cause of death. Cause-specific analysis can be an alternative; however, the cause of death is often difficult to state with certainty leading to misclassification. In population-based research, for example, determining the cause of death is usually obtained from death certificates which can be unreliable in both cancer and heart disease [7, 8].

Relative survival is used routinely in population-based cancer studies [9]. Use of relative survival in therapeutic areas other than cancer is limited although there are a few applications of the idea for survival following heart surgery [10, 11], after a myocardial infarction (MI) [12, 13] and after a stroke [14, 15].

There have been several approaches proposed to model $\lambda(t)$. The most commonly used in applied research is a piecewise approach, wherein there is a need to split the timescale. Hakulinen and Tenkanen [16] estimate the model based on grouped data using a generalized linear model (GLM) framework with a binomial assumption for the number of observed deaths. Estève *et al*. [17] adopted a full likelihood approach based on individual level data. Dickman *et al*. [18] proposed a GLM with a Poisson error structure using exact survival times and individual level data which is equivalent to the Estève *et al*. full likelihood approach. Models can also be fitted at the grouped level (i.e. collapsed over categorical covariate patterns). Relative survival models have been extended to more flexible models such as multivariate fractional polynomials (MFP) [19] and cubic splines [20] that also use split-time data. B-splines have also been suggested for relative survival to model the hazard ratio as a flexible function of time by Giorgi *et al*. [21] as have restricted cubic splines by Bolard *et al*. [22], both using non-split-time data.

The individual level piecewise approach yields several observations for each individual in the data set as there is a record for each subdivision of time. This can have computational implications in very large data sets. For example, Lambert *et al*. [19] used grouped data when fitting fractional polynomial models due to the computation time needed for individual level models. A disadvantage

DOI: 10.1002/sim

of using grouped data is that continuous covariates need to be categorized, which is not a preferred practice due to the loss of power [23] and subjectivity in the choice of cut points. Piecewise models rely on a subjective choice of the number and location of split points and lack clinical plausibility. Using a piecewise approach also implies generation of many parameters for time-dependent effects (if a dichotomous covariate is perceived to vary over time then there is a need to create $(h-1)$ more parameters where h is the number of splits in the timescale. It is possible to simplify this by making assumptions, for example, that excess hazards are proportional after 2 years. B-splines [21] and as the restricted cubic spline approach proposed by Bolard *et al*. [22], obtain the survival function using numerical integration increasing the computation time. All current approaches smooth on the log excess hazard scale.

The flexible parametric approach we adopt models on the log cumulative excess hazard scale, which can have advantages in terms of model stability. The hazard and survival functions are obtained analytically and non-split-time data are used, i.e. a single record per subject. The models are thus restricted to only one time scale. This paper proposes an extension of the Royston and Parmar flexible parametric model [1] to relative survival. This method is illustrated in a heart disease data set for the Leicestershire area in the U.K. in which modelling the difference between three age groups will be illustrated. In Section 2, the methodology of relative survival and the extension of Royston and Parmar's method are described. The data set is described in Section 3. Section 4 details analyses of the data set focusing on the fit of the spline models and a comparison with standard relative survival analyses. Section 5 is a discussion of implications along with the advantages and limitations of this method.

2. METHODS

In survival analysis an individual's contribution to the log likelihood can be written as

$$\ln L_i = d_i \ln[h(t_i)] + \ln[S(t_i)] \tag{3}$$

where $d_i = 1$ for an event (i.e. death here) and 0 otherwise. For relative survival the survival and hazard functions are defined in equations (1) and (2). This leads to

$$\ln L_i = d_i \ln[h^*(t_i) + \lambda(t_i)] + \ln[S^*(t_i)] + \ln[R(t_i)] \tag{4}$$

As $S^*(t_i)$ is not dependent on any of the unknown model parameters it can be removed from the log likelihood; thus,

$$\ln L_i = d_i \ln[h^*(t_i) + \lambda(t_i)] + \ln[R(t_i)] \tag{5}$$

when using split-time data the term $\ln[R(t_i)]$ is the one that is split, into one for each follow-up interval. The current methods fit models on the log excess hazard scale, i.e.

$$h(t) = h^*(t) + \exp(\beta z) \tag{6}$$

where β is a vector of parameters to be estimated for covariates z. If we consider the hazard (2) then integrating both sides gives the cumulative hazard $H(t)$

$$H(t) = H^*(t) + \Lambda(t) \tag{7}$$

DOI: 10.1002/sim

where $H(t)$ is the cumulative overall hazard, $H^*(t)$ is the cumulative expected hazard and $\Lambda(t)$ is the cumulative excess hazard. We are interested in modelling on the log cumulative excess hazard scale for relative survival analysis:

$$\ln(-\ln R(t;z)) = \ln(\Lambda(t)) = \ln(\Lambda_0(t)) + \beta z \tag{8}$$

where $\Lambda_0(t)$ is the baseline cumulative excess hazard function and $R(t;z)$ is the relative survival function. Note that proportional cumulative excess hazards imply proportional excess hazards. Following Royston and Parmar we write $x = \ln(t)$ and the log cumulative excess hazard as

$$\ln(\Lambda(t)) = s(x;\gamma) + \beta z \tag{9}$$

where $s(x;\gamma)$ can be any smoothing function such as the various regression spline functions or fractional polynomials. For a proportional excess hazards (PEH) model, let $\eta = s(x;\gamma) + \beta z$. Then the associated relative survival and excess hazard functions are given by

$$R(t) = \exp[-\exp(\eta)] \tag{10}$$

and

$$\lambda(t) = \frac{1}{t}\frac{\mathrm{d}s(x;\gamma)}{\mathrm{d}x}\exp(\eta) \tag{11}$$

This implies that the overall survival is

$$S(t) = S^*(t)\exp[-\exp(\eta)] \tag{12}$$

with associated hazard

$$h(t) = h^*(t) + \frac{1}{t}\frac{\mathrm{d}s(x;\gamma)}{\mathrm{d}x}\exp(\eta) \tag{13}$$

Substituting (10) and (11) into (5) gives the log-likelihood contribution for each subject:

$$\ln L_i = d_i \ln\left[h^*(t_i) + \frac{1}{t_i}\frac{\mathrm{d}s(x_i;\gamma)}{\mathrm{d}x_i}\exp(\eta_i)\right] - \exp(\eta_i) \tag{14}$$

Time by covariate interactions are easy to include by modelling γ separately for each covariate level, which removes the assumption of PEH.

The model thus far has been generalizable as any smoothing function could be used for $s(x;\gamma)$, but we follow the flexible parametric model for censored survival data that was proposed by Royston and Parmar [1] which incorporates the use of natural/restricted cubic splines in models on the cumulative hazard or cumulative odds scale (here we are just concerned with modelling on the cumulative excess hazard scale). Following Royston and Parmar, the splines are defined by

$$s(x;\gamma) = \gamma_0 + \gamma_1 x + \gamma_2 v_1(x) + \cdots + \gamma_{m+1} v_m(x) \tag{15}$$

where the jth basis function is defined for $j = 1, \ldots, m$ as

$$v_j(x) = (x - k_j)_+^3 - \lambda_j (x - k_{\min})_+^3 - (1 - \lambda_j)(x - k_{\max})_+^3 \tag{16}$$

DOI: 10.1002/sim

where $u_+ = u$ if $u > 0$ and $u_+ = 0$ if $u \leqslant 0$, $k_{\min}$ is the position of the first knot, $k_{\max}$ of the last knot and

$$\lambda_j = \frac{k_{\max} - k_j}{k_{\max} - k_{\min}}$$

where the derivatives are obtained from (15) and (16) as

$$\frac{\mathrm{d}s(x;\gamma)}{\mathrm{d}x} = \gamma_1 + \sum_{j=2}^{m} \gamma_j [3(x-k_j)_+^2 - 3\lambda_j (x-k_{\min})_+^2 - 3(1-\lambda_j)(x-k_{\max})_+^2] \tag{17}$$

Modelling is performed on the log cumulative excess hazard scale as the survival and hazard functions can then be estimated analytically, and numerical integration is not required, which speeds up computation time. It is more likely to retain stability; for example, when modelling on this scale using log time and a Weibull distribution the relationship will be linear and as such departures from the Weibull will be under more stable conditions. This model does not use time-split data and expands upon a current method for 'standard' survival. The models are fitted in a Stata program available from the authors on request. The likelihood functions for the models were maximized using the Newton Raphson technique with the first and second derivatives estimated numerically as implemented using the '*ml*' command in Stata [24].

The piecewise approach adopted in this paper uses split-time data at the individual level, i.e. with many rows per person with the split times placed at: 0.5, 1, 2, 3, 4 and 6 years. Piecewise models are fitted using the Dickman *et al*. [18] approach using Poisson GLMs.

3. DATA

The data set used for the analyses was set up in 1987 at the Leicester Royal Infirmary in the U.K. All admissions to the coronary care unit in Leicestershire, U.K., are recorded and followed up to death. The data used in this paper are taken from 1993 to 2006 for consistency as there was a change in the database management as many more covariates were recorded from this time.

A total of 5859 observations include 2383 deaths (40.7 per cent) over the 13-year period. Of these, the cohort used for the analyses consists of 4747 (1759 deaths) observations and follow-up was censored after six years. Acute MI is described using characteristics of the electrocardiogram (ECG) and the results of blood tests. Only patients with ECG evidence of ST elevation, a criterion that did not change over our study period, were included. Moreover, ST elevation MI is associated with higher mortality. Thus, patients included are those experiencing their first known acute MI and demonstrating ST elevation on the ECG. Of the 4747 included in the analysis cohort there are 3231 men (68.1 per cent); 3965 (83.6 per cent) of the cohort self-reported as White and 763 Asian (16.1 per cent). The mean age of the cohort is 66.7 with men being 8 years younger than women on average (64.1: Men, 72.2: Women). One thousand seven hundred and fifty-nine (37.1 per cent) of this cohort died within the 6-year follow-up used in the investigation.

Heart disease has a very high increased mortality rate in the first 30 days, with perhaps 40–50 per cent of all deaths occurring in this time period, there being 843 (47.9 per cent) for this cohort. The models need to capture this important feature of the data adequately. After admission for MI the risk of almost immediate death is high, yielding many deaths at day zero; 252 (4.7 per cent) in this cohort. Deaths on day zero were recorded as day 0.5.

DOI: 10.1002/sim

Follow-up for analysis began at hospitalization and not after 30 days survival as is the case with other heart disease mortality analyses [12]. The expected mortality was calculated using rates from the United Kingdom Government Actuary's Department [25] matched on age, sex and period of hospitalization to the England and Wales population. We illustrate the methodology by estimating differences in relative survival/excess mortality by age group. We define our age groups as <65, 65–75 and 75 years or older at entry.

There are two fitting principles used in this paper, one that uses *split-time* data at the individual level used in the piecewise approach and another that uses *non-split*-time data which has one row of data for each individual with their survival time kept continuous as used in the flexible parametric approach.

4. RESULTS

4.1. Piecewise models

A piecewise model was fitted using individual level split-time data with six time intervals on the time scale split annually for the 6-year follow-up, except for the first year which was split into two 6-month intervals and years 5 and 6 were combined. Using a PEH assumption gave estimates of the excess hazard ratios (with 95 per cent confidence intervals) as 2.71 (2.25, 3.27) and 6.77 (5.71, 8.03) for 65–75 and 75 years or older, respectively, using <65 age group as the comparative baseline. Both age groups have a significantly increased risk of death over the baseline group with the oldest age group being the worst. Extending the model to non-PEH by fitting an interaction between time interval and each covariate gave estimates for the excess hazard rate ratios as shown in Table I. This highlights a large increased uncertainty for the parameter estimates as follow-up time increases. There is a higher excess hazard ratio initially for both age groups. An excess hazard rate ratio up to 6 months of 7.30 (6.06, 8.79) for patients aged 75 years or older indicates that compared with a patient 65 years or younger they have an excess mortality rate that is over seven times higher during the first 6 months after MI. However, a likelihood ratio test between the PEH and non-PEH model showed a non-significant difference ($\chi^2_{10} = 7.12$, $p = 0.714$).

4.2. PEH-flexible parametric models

The flexible parametric models using spline functions for the baseline cumulative excess hazard were fitted to the non-split heart disease data. The estimated relative survival curve with five internal knots split by age group adopting a PEH assumption is shown in Figure 1. Vertical

Table I. Estimates of the excess hazard rate ratio with 95 per cent confidence intervals for the individual level piecewise non-proportional excess hazard model using split-time data.

End of interval	65–75 years old, excess hazard ratio (CI)	75 years or older, excess hazard ratio (CI)
6 months	2.85 (2.32, 3.50)	7.30 (6.06, 8.79)
Year 1	2.11 (0.78, 5.75)	3.68 (1.37, 9.91)
Year 2	2.18 (1.05, 4.49)	3.93 (1.89, 8.14)
Year 3	3.00 (0.90, 9.98)	5.98 (1.81, 19.80)
Year 4	1.99 (0.61, 6.51)	2.49 (0.51, 12.11)
Years 5 and 6	1.73 (0.46, 6.48)	6.30 (2.17, 18.26)

DOI: 10.1002/sim

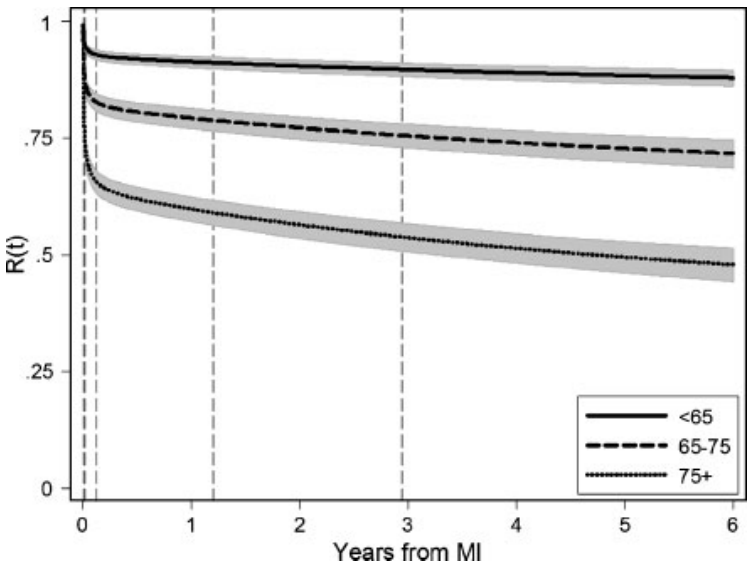


Figure 1. Flexible parametric models using six degrees of freedom and a PEH assumption: the relative survival curve with 95 per cent confidence interval. Internal knots are shown by vertical dashed lines. It is not possible to distinguish the first 2 knot lines.

Table II. AIC and estimates of the excess hazard rate ratio with 95 per cent confidence intervals for spline models with various degrees of freedom using non-split data.

	Proportional excess hazards model				
Spline model df	Location of knots*	AIC	65–75 years old, excess hazard ratio (CI)	75 years or older, excess hazard ratio (CI)	Non-PEH AIC
2	50	3769.36	2.61 (2.15, 3.16)	5.90 (4.95, 7.02)	3768.151
3	33, 67	3375.56	2.61 (2.15, 3.16)	5.90 (4.95, 7.02)	3382.641
4	25, 50, 75	3283.68	2.59 (2.15, 3.12)	5.74 (4.85, 6.80)	3289.849
5	20, 40, 60, 80	3161.10	2.59 (2.15, 3.11)	5.75 (4.86, 6.81)	3170.899
6	17, 33, 50, 67, 83	3108.17	2.59 (2.15, 3.11)	5.73 (4.84, 6.78)	3118.672

*Percentiles of all death times, as proposed by Royston and Parmar.

dashed lines on the figures indicate the position of the internal knots. Table II shows the Akaike information criterion (AIC) values for PEH models with different degrees of freedom (df) along with estimates of the excess hazard rate ratios. Following Royston and Parmar the knot locations are based on percentiles of survival time for those who experience an event. All five models gave similar estimates for both excess hazard ratios. The PEH model with 6 df was determined as the optimal model using the AIC. Therefore, the model was fitted using five internal knots placed at the 17, 33, 50, 67 and 83 percentiles of time of all deaths as adopted by Royston and Parmar [1]. The first two knots are fitted to days one and four with knot three at day 44 and the last two knots at days 439 and 1074. The relative survival for the under 65 year olds is better than the 65–75-year

DOI: 10.1002/sim

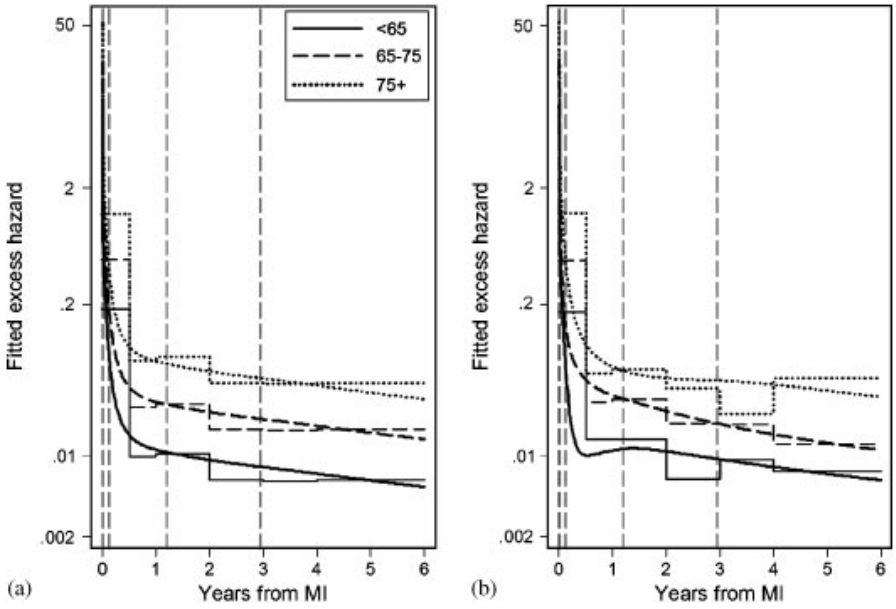


Figure 2. Comparison of the fitted excess hazard rates for the selected spline model using non-split data with the piecewise individual level split-time data for (a) PEH and (b) non-PEH models. Internal knots are shown by vertical dashed lines. It is not possible to distinguish the first 2 knot lines.

group and much better than the over 75-year group. There is a large decrease in relative survival during the first 6 months and after this point the curves are still decreasing but at a slower rate. It is evident that the curves do not reach a plateau, suggesting that these patients do not achieve the same mortality rate as the general population to which they are compared. Figure 2(a) shows the excess hazard rate compared with the standard piecewise approach described in Section 4.1. The estimates of the model are in broad agreement. However the piecewise interval in the first 6 months does not capture the sharp reduction in the hazard rate adequately. The excess hazard shows the high excess mortality rate experienced by the three age groups, with the over 75-year group having the highest excess mortality. Table II shows the estimate of the excess hazard ratio for the 6 df model which is greater than one for both age groups. The estimate of the excess hazard rate ratio for patients aged 65–75 years old is 2.59 which is similar to the piecewise model estimate (2.72). For patients aged 75 years or older the excess hazard rate ratio is estimated as 5.73 which is lower than the piecewise estimate of 6.80.

4.3. Non-PEH-flexible parametric models

Using the same model selection criteria and using age groups as a covariate, a non-PEH model was fitted to the non-split data, which was obtained by fitting an interaction between time and a fixed covariate thus allowing the spline terms to vary by age group. The model with six df was again found to be the best of the non-PEH models using the AIC, as shown in Table II. Using the AIC values to compare the PEH and non-PEH models, for all but the 2-df model a PEH model would be selected. The excess hazard rate appears to give an adequate fit when compared with the standard piecewise non-PEH piecewise approach as shown in Figure 2(b). By including 6 df

DOI: 10.1002/sim

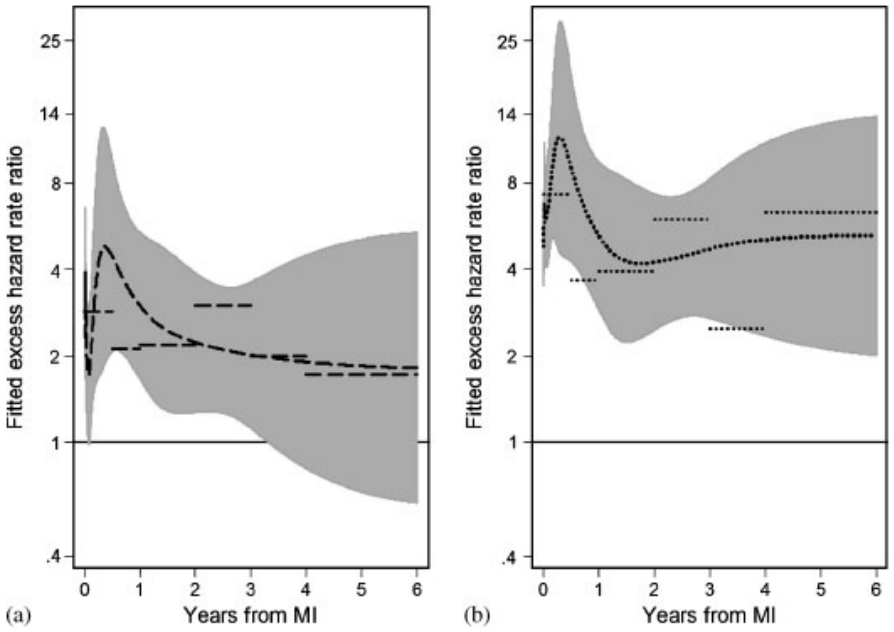


Figure 3. The fitted excess hazard rate ratio with 95 per cent confidence interval alongside the piecewise estimates, where — is an excess hazard rate ratio = 1 for (a) aged 65–75 years and (b) aged 75+ years.

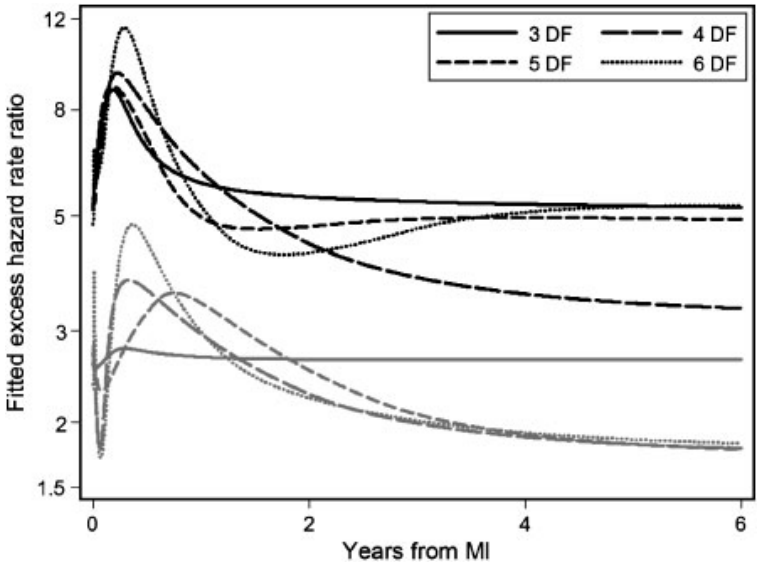


Figure 4. Varying the degrees of freedom in the spline model and the effect on the fitted excess hazard rate ratio for patients aged 65–75 (grey) and patients aged 75 or older (black), both in comparison with patients aged 65 years or younger.

for the spline terms and allowing age to vary over time, the estimated excess hazard rate for the youngest age group appears to stabilize earlier than in the PEH model in Figure 2(a). This is reflected by the fact that in Figure 3(a) and (b) the excess hazard rate ratios appear to increase

DOI: 10.1002/sim

and decrease sharply with a peak around the six-month period. The estimates from the piecewise model assuming non-PEH have been overlaid on the figure and are in broad agreement with the estimates from the piecewise model.

Figure 4 provides an illustration of the various excess hazard rate ratios for models with 3–6 df. For the 65–75-age group (grey lines) the 4-, 5- and 6-df models show similar shapes with a large increase in risk for the first six months and then decrease to an excess hazard rate ratio of approximately 1.8. The 3-df model does not capture any of the changes in excess hazard. Estimates for the over 75 group (black lines) show a similar shape for all four models to begin with but the 4-df model decreases more than the other three models. However, it should be acknowledged that the time-dependent effects are non-significant and there may be a degree of over-fitting here.

5. DISCUSSION

Relative survival has advantages over the standard methods when dealing with population-based censored survival data as the cause of death is not required. We have investigated the application of restricted cubic splines [1] on the cumulative excess hazard scale to allow modelling of time continuously and have compared these with piecewise models proposed by Dickman *et al*. [18]. Age was categorized into three groups for both the piecewise and the proposed model, although there are advantages to also treating age as a continuous variable, we chose to categorize it for simplicity.

Although piecewise models are most popular in practice they are not clinically realistic which has led to interest in models that treat the timescale as continuous. To fit smooth functions of time, fractional polynomials [19] and splines [20–22] have previously been used to model relative survival. These models are fitted on the log excess hazard scale, which requires either numerical integration or approximating the likelihood by splitting the data finely, both of which can be computationally intensive. The model we use is fitted on the log cumulative excess hazard scale. The advantages of the extended Royston and Parmar model proposed here are that there is no splitting of the timescale, avoiding the need to arbitrarily choose split points, and the survival and hazard functions are estimated analytically and hence computationally quicker. As an illustration of computational time, when using RSURV in R [26] which uses a PEH B-spline approach described by Giorgi *et al*. [21] using age groups and default values on our data converged in 148.58 s, whereas our approach implemented in Stata using the same amount of knots (2 internal) also using a PEH assumption converged in 6.67 s and 5 internal knots with a non-PEH assumption converged in 27.86 s on a Pentium 4 central processing unit (CPU) 3.4 GHz PC. A key advantage is the ability to model the baseline hazard [27, 28]. A flexible parametric approach may provide an appropriate estimate for the baseline hazard whose behaviour may be of medical interest.

Using the AIC, a model with 5 internal knots was selected. However, this investigation has shown that various df in the spline models have produced very similar covariate effects using a PEH model. For these data there was no evidence to support rejection of the PEH model, either from the AIC or the likelihood ratio test for the piecewise models. The estimated coefficients were very similar amongst the various PEH models but different from the piecewise model estimates for the oldest age group of 6.60; all of the spline models estimated the excess hazard ratio in the range of 5.73–5.92. If the piecewise model had more splits in the first 6 months then they

DOI: 10.1002/sim

become more similar; for example, splitting the first 6 months into 5 intervals produced an estimate of 5.74. This shows that when the piecewise models are allowed to be more flexible the results become very similar to the spline results. However, this requires more parameters to be estimated. A sensitivity analysis using the splines approach is always recommended especially in complex models. For PEH models the number and location of knots are unlikely to be crucial for estimating the excess hazard ratios, unless you miss the change in the first 6-month period, i.e. miss the extremes.

There is an interest in relative survival models with time-dependent effects. Flexible parametric models can be extended by allowing the covariate(s) to vary within each spline term by fitting an interaction. However, in our data set there is little evidence of non-PEH. If we were still interested in allowing the effect of age to vary over time then model selection becomes less intuitive as shown in Figure 4. All of these models are modelling a similar effect but the 3-df and 4-df models do not appear to model the excess hazard rate ratios as accurately as required. Knot frequency and location are likely to be more important than in PEH models as illustrated in Figure 4. However, these differences should not be over-interpreted for this particular case as the AIC favours PEH models. During the investigation of non-PEH models shown in Figure 3, a sharp increase and decrease around the first 6 months were observed for the excess hazard rate ratios. This effect is likely due to the youngest age group stabilizing earlier and creating a larger difference earlier on around the 6-month period. Further sensitivity analyses were performed on non-PEH models varying the number and location of the knots. The method was found to be robust to small changes in the choice of knot location, for example, fitting at 10, 30, 50 and 70 per cent rather than the 20, 40, 60 and 80 per cent percentiles. This shows that the impact of knots depends on the structural form of the data and as such the type of data should be considered when fitting splines along with sensitivity analyses.

Relative survival in heart disease is in early development and there are limited examples in the literature for the use of relative survival methods in heart disease, although Stare *et al*. [29] have developed relative survival models using a heart disease data set. Analyses conditional on survival for a month post MI [12], or conditional on discharge from hospital [29] do not have the problem of the very high excess hazard rate experienced in this time period. This is a key problem; if piecewise models were to be used then there is a requirement to split the timescale finely within the first few weeks to capture the shape. There are benefits of treating time as continuous, as the model proposed here does, but the high early mortality rate must be adequately modelled. It is important when modelling in observational studies of heart disease that one investigates the effect of covariates in order to obtain adjusted estimates and investigate effect modifiers. It should be noted that this approach is also suitable for other therapeutic areas including cancer. Here risk factors other than age associated with heart disease, such as sex and economic status have not been modelled, but these could be included in extended models.

The models could be extended in a number of ways. The models originally proposed by Royston and Parmar allowed the modelling of proportional hazards and proportional odds. Models for relative survival shown here have been fitted to the cumulative excess hazards scale, but could be extended to the odds scale. An automated selection of knot location may prove a useful extension but may result in clinically implausible fits [5]. The model could also be extended to other smoothing functions such as fractional polynomials. Assessment of the model fit is currently performed by overlaying the piecewise estimates at the individual level over the spline models as shown in Figure 2. This along with the AIC is currently the main method of assessing the fit of the model and the criterion needs further attention.

DOI: 10.1002/sim

ACKNOWLEDGEMENTS

We thank Patrick Royston for his comments and involvement with the extension to the Royston and Parmar model, and Paul Dickman for his contributions to discussions surrounding the methodology. We also thank the four referees for their useful comments and suggestions. Chris Nelson was funded by a British Heart Foundation PhD studentship (FS/05/080/19415).

REFERENCES

1. Royston P, Parmar MKB. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine* 2002; **21**(15):2175–2197.
2. Pillay D, Bhaskaran K, Jurriaans S, Prins M, Masquelier B, Dabis F, Gifford R, Nielsen C, Pedersen C, Balotta C, Rezza G, Ortiz M, Mendoza Cd, Kücherer C, Poggensee G, Gill J, Porter K. CASCADE virology collaboration: the impact of transmitted drug resistance on the natural history of HIV infection and response to first-line therapy. *AIDS* 2006; **20**(1):21–28.
3. Griffin JT, Fraser C, Gras L, de Wolf F, Ghani AC. The effect on treatment comparisons of different measurement frequencies in human immunodeficiency virus observational databases. *American Journal of Epidemiology* 2006; **163**(7):676–683.
4. Royston P, Parmar MKB, Qian W. Novel designs for multi-arm clinical trials with survival outcomes with an application in ovarian cancer. *Statistics in Medicine* 2003; **22**(14):2239–2256.
5. Royston P. st0001. Flexible parametric alternatives to the Cox model, and more. *Stata Journal* 2001; **1**(1):1–28.
6. Ederer F, Axtell LM, Cutler SJ. The relative survival rate: a statistical methodology. *National Cancer Institute Monographs* 1961; **6**:101–121.
7. Lauer MS, Blackstone EH, Young JB, Topol EJ. Cause of death in clinical research: time for a reassessment? *Journal of the American College of Cardiology* 1999; **34**(3):618–620.
8. Mant J, Wilson S, Parry J, Bridge P, Wilson R, Murdoch W, Quirke T, Davies M, Gammage M, Harrison R, Warfield A. Clinicians didn't reliably distinguish between different causes of cardiac death using case histories. *Journal of Clinical Epidemiology* 2006; **59**(8):862–867.
9. Coleman M, Babb P, Damiecki P, Grosclaude P, Honjo S, Jones J, Knerer G, Pitard A, Quinn M, Sloggett A, Stavola BD. Cancer survival trends in England and Wales, 1971–1995: deprivation and NHS region. *Office for National Statistics*, London, 1999.
10. Kvidal P, Bergstrom PR, Horte L-G, Stahle E. Observed and relative survival after aortic valve replacement. *Journal of the American College of Cardiology* 2000; **35**(3):747–756.
11. Norman PE, Semmens JB, Lawrence-Brown MMD. Long-term relative survival following surgery for abdominal aortic aneurysm: a review. *Cardiovascular Surgery* 2001; **9**(3):219–224.
12. Bronnum-Hansen H, Jorgensen T, Davidsen M, Madsen M, Osler M, Gerdes LU, Schroll M. Survival and cause of death after myocardial infarction: the Danish MONICA study. *Journal of Clinical Epidemiology* 2001; **54**(12):1244–1250.
13. Patel PJ, Keating RJ, Gersh BJ, Hodge DO, Hammill SC, Shen WK. Outcome of patients with newly diagnosed atrial fibrillation at the Mayo Clinic and residing in that area. *The American Journal of Cardiology* 2004; **94**(11):1379–1382.
14. Hardie K, Hankey GJ, Jamrozik K, Broadhurst RJ, Anderson C. Ten-year survival after first-ever stroke in the Perth community stroke study. *Stroke* 2003; **34**(8):1842–1846.
15. Hankey GJ, Jamrozik K, Broadhurst RJ, Forbes S, Burvill PW, Anderson CS, Stewart-Wynne EG. Five-year survival after first-ever stroke and related prognostic factors in the Perth Community Stroke Study. *Stroke* 2000; 2080–2086.
16. Hakulinen T, Tenkanen L. Regression analysis of relative survival rates. *Applied Statistics* 1987; **36**(3):309–317.
17. Estève J, Benhamou E, Croasdale M, Raymond L. Relative survival and the estimation of net survival: elements for further discussion. *Statistics in Medicine* 1990; **9**(5):529–538.
18. Dickman PW, Sloggett A, Hills M, Hakulinen T. Regression models for relative survival. *Statistics in Medicine* 2004; **23**(1):51–64.
19. Lambert PC, Smith LK, Jones DR, Botha JL. Additive and multiplicative covariate regression models for relative survival incorporating fractional polynomials for time-dependent effects. *Statistics in Medicine* 2005; **24**(24):3871–3885.

DOI: 10.1002/sim

20. Remontet L, Bossard N, Belot A, Estève J. The French network of cancer registries FRANCIM. An overall strategy based on regression models to estimate relative survival and model the effects of prognostic factors in cancer survival studies. *Statistics in Medicine* 2007; **26**(10):2214–2228.
21. Giorgi R, Abrahamowicz M, Quantin C, Bolard P, Esteve J, Gouvernet J, Faivre J. A relative survival regression model using B-spline functions to model non-proportional hazards. *Statistics in Medicine* 2003; **22**(17):2767–2784.
22. Bolard P, Quantin C, Abrahamowicz M, Esteve J, Giorgi R, Chadha-Boreham H, Binquet C, Faivre J. Assessing time-by-covariate interactions in relative survival models using restrictive cubic spline functions. *Journal of Cancer Epidemiology and Prevention* 2002; **7**(3):113–122.
23. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in Medicine* 2005; **25**(1):127–141.
24. Gould W, Pitblado J, Sribney W. *Maximum Likelihood Estimation with Stata* (2nd edn). Stata Press: Texas, 2003.
25. GAD. *Interim life tables of mortality rates for England and Wales*. http://www.gad.gov.uk/Life_Tables/Interim_life_tables.htm (30 October 2006).
26. Giorgi R, Payan J, Gouvernet J. RSURV: a function to perform relative survival analysis with S-PLUS or R *Computer Methods and Programs in Biomedicine* 2005; **78**(2):175–178.
27. Carstensen B. Who needs the Cox model anyway? http://staff.pubhealth.ku.dk/∼bxc/frmain.html (30 October 2006).
28. Hjort N. On inference in parametric survival data models. *International Statistical Review* 1992; **60**:355–387.
29. Stare J, Henderson R, Pohar M. An individual measure of relative survival. *Applied Statistics* 2005; **54**(1):115–116.

DOI: 10.1002/sim

Appendix D

European Heart Journal: Relative survival: What Can Cardiovascular Disease Learn From Cancer?

European Heart Journal (2008) **29**, 941–947
doi:10.1093/eurheartj/ehn079

SPECIAL ARTICLE

Relative survival: what can cardiovascular disease learn from cancer?

Christopher P. Nelson[1]*, Paul C. Lambert[1], Iain B. Squire[2], and David R. Jones[1]

[1]Centre for Biostatistics and Genetic Epidemiology, Department of Health Sciences, University of Leicester, 2nd Floor, Adrian Building, University Road, Leicester LE1 7RH, UK; and [2]Department of Cardiovascular Sciences, University of Leicester, Leicester, UK

Received 10 October 2007; revised 31 January 2008; accepted 8 February 2008; online publish-ahead-of-print 6 March 2008

Aims To illustrate the application of relative survival to observational studies in coronary heart disease (CHD) and potential advantages compared with all-cause survival methods. Survival after myocardial infarction (MI) is generally assessed using all-cause or cause-specific methods. Neither method is able to assess the impact of the disease or condition of interest in comparison with expected survival in a similar population. Relative survival, the ratio of the observed and the expected survival rates, is applied routinely in cancer studies and may improve on current methods for assessment of survival in CHD.

Methods and results Using a cohort of subjects after a first recorded acute MI, we discuss the application of relative survival in CHD and illustrate a number of the key issues. We compare the findings from relative survival with those obtained using Cox proportional and non-proportional hazards models in standard all-cause survival. Estimated survival rates are higher using relative survival models compared with all-cause methods.

Conclusion Estimates obtained from all-cause mortality fail to disentangle mortality associated with the condition of interest from that due to all other causes. Relative survival gives an estimate of survival due to the disease of interest without the need for cause of death information.

Keywords Relative survival • Survival analysis • Population-based research

Introduction/methods

Why measure mortality in heart disease?

Coronary heart disease (CHD) is the leading cause of mortality in industrialized societies,[1,2] accounting for over 105 000 deaths in the UK in 2004.[3] Understanding both short- and long-term patient survival may help to inform improved management after presentation with CHD. Much of what we know about survival for patients with CHD comes from randomized controlled trials (RCT). Such trials usually recruit relatively selected populations followed over relatively short times, and the findings of RCTs may not easily be generalizable to the general population. Thus the assessment of patient survival in unselected clinical populations can be informative to the patient, the clinician, and in terms of future health-service provision. Important prognostic factors related to survival can be evaluated along with comparisons over time and between centres.

How would we ideally assess mortality?

To understand the natural history of a disease or condition of interest, and the influence of risk-factors and co-morbidity properly, it is essential to use appropriate statistical techniques. Ideally, the impact on mortality of a particular disease or condition would be measured by assessment of mortality specifically due to, or associated with, the disease of interest. However in broad population-based studies, cause of death is often difficult to establish with certainty.

Cause-specific survival analysis

Cause-specific (or net) survival considers as events only deaths that can be directly attributed to the disease of interest, with deaths from all other causes being censored. The probability of survival can be evaluated, but avoiding consideration of competing risks. When considering cause-specific survival, it is usually of interest to fit statistical models to investigate simultaneously the influence of covariates, such as age and sex.

*Corresponding author. Tel: +44 116 229 7254, Fax: +44 116 229 7250, Email: cn46@le.ac.uk

What are the limitations of cause-specific analyses?

The main limitation of cause-specific survival is its dependence upon reliable coding of information on the cause of death. This reliance is not well founded, particularly when the source of information is the death certificate. For example a patient may be recorded as dying of renal failure where this is the result of cardiogenic shock following acute MI. Three trained physician-adjudicators assessed cause of death in 2686 Framingham Heart Study participants.[4] Of cases adjudicated as attributable to CHD, only 83.8% were recorded as such on the death certificates. Moreover, of cases recorded as due to CHD on the death certificate, this was confirmed in less than 70% by the physician-adjudicators. In this context, arguments have been presented that are critical of cause-specific survival. Lauer *et al*.[2] argue that data obtained from death certificates or from medical records are haphazard, biased, and often inaccurate. These authors also suggest that all-cause death should be assessed as the primary end point, as it is both objective and unbiased. More recently, Mant *et al*.[5] argued that cause-specific methods are also flawed when attempting to differentiate among causes of cardiac death. These authors highlighted that this was the result of the inability of clinicians to agree on a cause of death; even where autopsy information was available, disagreement remained for a third of all deaths.

All-cause survival analysis

All-cause (or crude) survival includes all deaths within the cohort under investigation, and does not separate those due to the disease of interest from those due to other causes. All deaths are considered as events with only those surviving the follow-up period, or those lost to follow-up, being censored.

How are all-cause methods also limited?

Clearly, in a cohort of patients with a given condition, some deaths will occur which are unrelated to the disease of interest. However all-cause analysis cannot disentangle deaths related to the disease of interest from deaths due to competing risks. A further, and important, limitation of all-cause survival methodology is its relative inability to disentangle the effect of strong covariates. For example, age is a strong confounder in most conditions. When we adjust for age at diagnosis in an all-cause model we are adjusting for the combination of the impact of age on mortality associated with the disease of interest, and its impact on mortality from all other causes, the magnitudes of which may differ.

Relative survival analysis: what can cancer share with coronary heart disease?

Analysis of survival in population-based cancer studies often includes *relative survival*, used alongside, or instead of, crude and cause-specific methods.[6,7] Relative survival estimates the mortality rate for patients with the condition of interest after correcting for estimated mortality from all other causes. This methodology considers survival in patients with a specific malignancy compared with survival in a comparator population. Large-scale examples can be found in the EUROCARE studies.[8] As some parallels exist between heart disease and cancer; for example, extended survival and follow-up is common to both disease areas, consideration of relative survival methodology may be worthwhile in CHD. An investigation of the practical application in CHD of relative survival methodology would be informative regarding the impact on outcome of CHD compared with what might be expected in the absence of that condition. In particular, relative survival allows for the use of appropriate statistical models to adjust adequately for confounders.

To date there are limited numbers of published relative survival applications and related methods in cardiovascular disease, most involving heart surgery. Of these, Norman *et al*.[9] describe a review of the limitations of current methods that use expected survival and how many authors focus on observed survival. Others describe various relative survival statistical models applied in patients undergoing cardiac surgery.[10–13] Few such reports pertain to patients with non-surgical conditions, such as atrial fibrillation,[14] MI[15] and stroke.[16,17] For the most part these analyses fail to model the influence of covariates and none fully utilizes relative survival methodology.

What is relative survival?

As noted above, relative survival attempts to separate mortality from the disease of interest from mortality due to all other causes. To do this the ratio of the observed (all-cause) survival in the cohort of interest and the expected survival in a similar group in the general population is calculated.[18]

$$\text{Relative survival} = \frac{\text{Observed (all-cause) survival in cohort studied}}{\text{Expected survival based on rates in a comparator population}}$$

The cohort of interest may comprise a sample of individuals with a specific diagnosis, for example, acute myocardial infarction (AMI). The comparator group is obtained from routine data, matched to the cohort of interest by age, sex, deprivation and other potentially important covariates. In a relative survival model, the observed mortality rate within the cohort of interest is made up of the background mortality rate in the general, comparator population (i.e. deaths due to all causes) plus the excess mortality rate associated with the condition of interest, i.e. all deaths resulting from the AMI.

$$\text{Total observed mortality rate} = \text{Expected mortality rate} + \text{Excess mortality rate}$$

So when modelling relative survival we attempt to estimate *directly* the excess mortality experienced by patients diagnosed with the condition of interest, compared with that of the general population, thereby obtaining an estimate of net survival.

An advantage of relative survival is that information on individual cause of death is not required; this removes the main problem associated with cause-specific mortality. However the method assumes that deaths due to the disease of interest are independent of mortality in the general population,[18] to which we will return later. Relative survival yields excess hazard ratios, as opposed to the standard hazard ratios obtained in Cox and other survival models. Excess hazard ratios can also be used to estimate

variability in the excess risk of death due to the disease of interest when, for example, comparing one demographic group with another, or temporal patterns of survival after AMI. As patterns of survival in AMI change, with increasing interest in long-term survival, we can expect the use of relative survival methods to become increasingly important and relevant.

What are the assumptions made within relative survival?

An assumption when using relative survival is that the matched population group used to obtain expected mortality is appropriate for the particular disease under study. The mortality rates should thus represent the expected mortality if the subject did not have the disease under study. A further issue is that when using national or regional life tables to obtain the expected mortality rates, the disease under study is included in these figures. If the prevalence of disease is low then this will have little impact on the estimates and for even the most common cancers it has been shown that this introduces negligible bias.[18] In older age groups AMI has a higher prevalence than most cancers and thus the prevalence issue needs to be considered. In subjects aged 75, there is negligible bias in the estimate of relative survival. However, the bias increases with age as the prevalence of AMI increases and the estimate of relative survival in the most elderly groups, for example those over 90 is potentially biased. Fortunately, there is very little bias in excess hazard ratios when comparing groups, as any bias will be in the same direction.

Relative survival: an example analysis in cancer

A recent population-based study[19] investigated changes in colorectal cancer survival, using all-cause and relative survival methods. Using all-cause analysis, the analysis showed that for patients aged $\geq$75 years, for all stages of disease, 5-year survival was 41.4% in 1976–87 and 43.3% in 1988–99, suggesting improved survival between these periods. In comparison, the estimates of relative survival were much higher and, moreover, nearly identical for the two periods (68.0% and 67.9%), suggesting no improvement. This observation could be explained by improved survival in the background population, as well as in the cohort of interest and/or a change in the age distribution of patients with colorectal cancer. In other words, the apparent improvement over time in survival in the cancer cohort could represent improved survival in the general population, rather than cause-specific survival improvement in the cohort of interest.

Relative survival in coronary heart disease: an illustration

Using information from a large database, we here illustrate the application of relative survival methodology in CHD. The data come from Leicester Royal Infirmary Coronary Care Unit. For all admissions to this CCU, details are recorded regarding presentation, comorbidity, and other clinical data. Patients are followed prospectively for mortality. Here we use data pertaining to patients admitted between 1993 and 2006, with a presentation of acute ST elevation MI. The total of 4747 individual events includes 3231 men. Men (64.1 years) had a mean age 8 years younger than women (72.2 years), the average age of the population being 66.7 years. Of this cohort, 1759 (37.1%) died within the 6 year follow-up period used here. Follow-up began at hospitalization and not after 30 days survival as is sometimes the case.[20] The expected mortality was calculated using rates from the United Kingdom Government Actuary's Department with each individual in the study cohort matched using age, year of hospitalization, and sex to the England and Wales population.[21]

What statistical methods are used?

In this paper we conduct a simple analysis in order to illustrate the methods. We start with simple lifetables and compare the findings from all-cause and relative survival approaches, stratified by age groups, defined as $<$60 years, 60–75 years and $>$75 years old. We then look at the effect of age using a relative survival model with proportional excess hazards and compare this with a standard (all-cause) Cox proportional hazards model. We also investigate non-proportional models for both the relative survival and Cox approach. Finally proportional models investigating the effect of sex, with and without adjustment for age groups, are performed for both relative survival and all-cause approaches. The variables we have analysed in this paper are also used as matching variables. However, it is important to note that other variables, for example glucose levels at admission, can easily be incorporated into the statistical model.

Results

What do relative survival lifetables contain?

Relative survival estimates can be presented in lifetables similar to those for standard survival analyses but with emphasis on relative survival and evaluating survival probabilities. The lifetables for each age group are shown in *Table 1*. In this table, in addition to the interval specific standard (observed) survival results, the expected (in italics) and interval-specific relative survival estimates (in bold) are also shown. If the observed number of deaths is equal to the expected number of deaths during an interval then the probability of survival for patients with the condition of interest is the same as that in the general population. Investigation of the $<$60 years old age group shows the observed number of deaths dropping from 88 in the first month to just 14 in year 2–3 of follow-up. However, the expected numbers of deaths by 30 days and during year 2–3 of follow-up are 0.5 and 5.8, respectively. Thus, acute ST elevation MI is associated with almost all excess deaths in the 30 day period following the event, and with an excess of less than 10 deaths in year 2–3.

Clearly, AMI is associated with marked excess in mortality in the immediate, post AMI period. However, how much of the later mortality exceed, and indeed of overall mortality in our cohort, can be ascribed to the index AMI? Consideration of the cumulative survival data illustrates the information to be gained from relative survival analysis. By the end of follow-up, an all cause approach shows cumulative survival of 0.258 for the $>$75 years old group. However, taking expected mortality into account, the cumulative

Table 1 Abbreviated lifetables for the youngest and oldest age group

Age group	Start of interval (years)	End of Interval (years)	Not alive at the start of the interval	Deaths in interval	Cumulative observed (all-cause) survival	Expected deaths in interval	Cumulative expected survival	Relative survival for interval	Cumulative relative survival
<60	0	0.0833	1373	88	0.936	*0.5*	*1.000*	**0.936**	**0.936**
	0.0833	0.5	1280	10	0.928	*2.6*	*0.998*	**0.994**	**0.931**
	0.5	1	1245	11	0.920	*3*	*0.995*	**0.993**	**0.925**
	1	2	1182	20	0.904	*6*	*0.990*	**0.987**	**0.913**
	2	3	1047	14	0.891	*5.8*	*0.984*	**0.992**	**0.905**
	3	4	938	18	0.873	*5.7*	*0.978*	**0.986**	**0.893**
	4	6	832	18	0.852	*10.1*	*0.964*	**0.989**	**0.883**
75+	0	0.0833	1436	481	0.665	*7.5*	*0.993*	**0.670**	**0.670**
	0.0833	0.5	955	115	0.584	*30.7*	*0.958*	**0.910**	**0.610**
	0.5	1	822	55	0.545	*32.4*	*0.919*	**0.972**	**0.593**
	1	2	748	101	0.468	*9*	*0.841*	**0.938**	**0.556**
	2	3	584	71	0.409	*49.6*	*0.766*	**0.959**	**0.533**
	3	4	464	52	0.359	*40.8*	*0.692*	**0.973**	**0.519**
	4	6	346	87	0.258	*56*	*0.561*	**0.887**	**0.460**

0.0833 years = 1 month.

relative survival rate is 0.460. In other words, if we ignore expected, background mortality in the population, the survival proportion is over 20% lower in absolute terms, as mortality due to other causes is included in the all-cause analysis.

Using the table we can also see that expected survival decreases with age. By 4–6 years of follow-up, the youngest age group had a cumulative expected survival rate of 0.964 and the oldest age group 0.561. The relative survival rate can be found as the ratio of the observed survival and expected survival shown in the table. Note how cumulative relative survival is greater compared with cumulative observed survival at all times, for all ages. Also worthy of note is that the interval specific relative survival is lowest during the first 30 days after AMI, in spite of this being the shortest time period considered in the analysis. The same is true for absolute survival (data not shown).

Can we see this graphically?

Lifetable information is often easier to interpret in graphical form. *Figure 1* shows the relative survival curve, along with the expected and observed (all-cause) survival probabilities split by age groups, (under 60 years old, 60–75 and >75 years). The figure highlights the difference between a relative survival and an all-cause survival approach. The observed values shown on the figure represent all of the deaths in the cohort, i.e. the survival rate if we assumed an all-cause approach. The expected survival is the survival rate we would expect if the cohort was in the general population without experiencing the index AMI, and the relative survival is the ratio of these two lines.

The initial drop in survival is smallest for the youngest age group. Moreover, relative and observed survivals are very similar in this age group, as expected survival is close to 1. As background mortality is very low in this age group, nearly all deaths experienced within this cohort are likely to be attributable to the index AMI. In the other age groups, particularly, the eldest, there is also very little difference between the observed and relative survival rates for the first year post-MI. However, there is a relatively high expected-mortality rate in this oldest age group; some of the deaths in this population are not due to the index AMI. Later, the relative survival curve begins to plateau, although never flattening completely. If the relative survival curve flattens out completely then the mortality rate of these surviving individuals is the same as expected in the general population. When this effect occurs it is known as statistical or population cure.

How do we use statistical models in relative survival?

The influence of cofactors or comorbidity on relative survival can be obtained by modelling, and expressed as excess hazard ratios; their interpretation is straightforward. For example, an excess hazard ratio of 2 for males would suggest that the excess mortality rate in men (i.e. deaths associated with the disease of interest) is twice as high as in women.

Table 2 illustrates the impact of age on hazard ratios estimated from a Cox proportional hazards model and on excess hazards ratios from a relative survival model. Using <60 years as the comparator group, for both 60–75 and >75 age groups, estimates of the excess hazard ratios are lower with relative survival modelling than with the hazard ratios from the Cox model. Once again this illustrates the increasing influence of competing risks in a population as that population ages.

The effect of age can also be assessed by the excess mortality rate. During the first month the excess mortality rate ranges from 800 (<60 age group) to 5450 (>75 age group) excess deaths per 1000 person years. We can conclude that during the first month there is a huge increased risk of death associated with the index event.

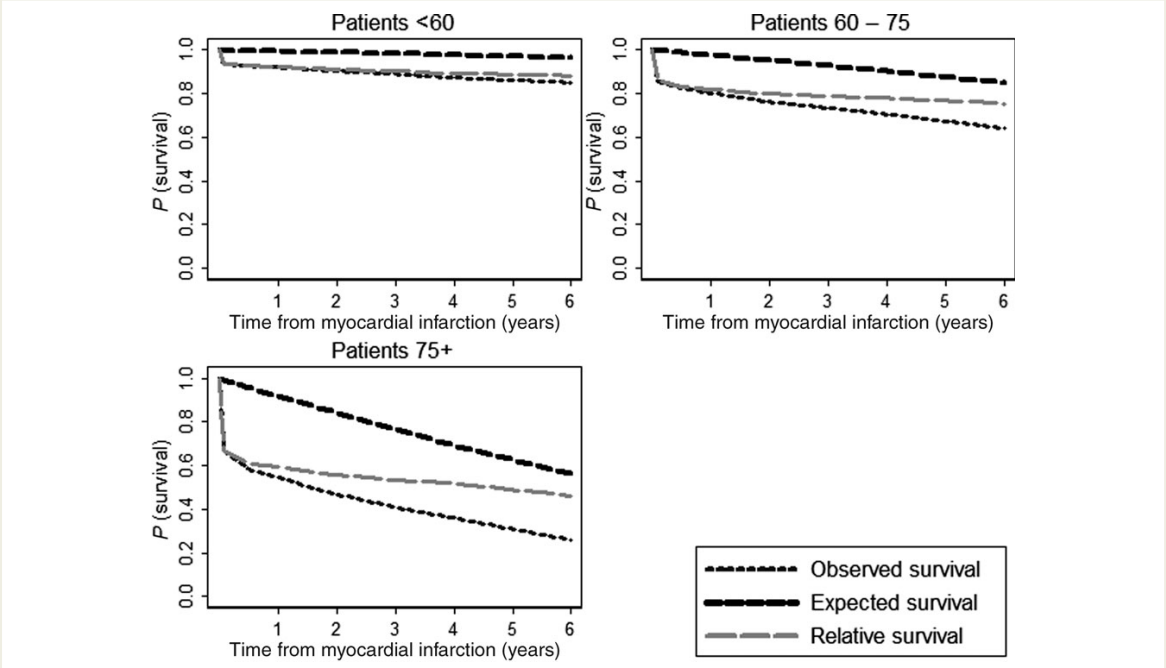


Figure 1 Observed (all-cause), expected, and relative survival split by age groups.

Table 2 Comparing results from a Cox model and a relative survival model

Age group	Hazard ratios (95% CI)	
	Cox model	Relative survival model
<60	1.00	1.00
60–75	2.71 (2.30, 3.20)	2.43 (1.97, 2.99)
75+	7.84 (6.68, 9.21)	6.92 (5.67, 8.44)

Models comparing the hazard ratios and excess hazard ratios from a Cox proportional hazards model and a relative survival model with proportional excess hazards (with 95% confidence intervals).

How do we look at time-dependent effects?

In their simplest form the models assume the impact on mortality to be proportional over time, which may not be the case. As with Cox models, it is possible to investigate whether the excess hazard ratio changes over time (*Figure 2*). For the 6–12 month follow-up interval, and using the all-cause method, we would estimate that patients aged 60–75 are at approximately three-fold greater risk of death, and patients aged >75 are near eight-fold greater risk, compared with a patient aged <60. In comparison, the relative survival method estimates of excess risk are rather lower, at two-fold for patients aged 60–75, and four-fold for those aged >75 years. Over time the subjects in the cohort are ageing and the expected mortality rate is also increasing with age. Relative survival methodology takes account of this, giving a more realistic estimate of the excess mortality associated with the previous AMI. This is shown for illustrative purposes even though a likelihood ratio test proved non-significant suggesting that the proportional model is more appropriate.

What are the results from further analyses?

An unadjusted Cox model found that male gender was associated with a hazard ratio of 0.60 (0.55, 0.66), suggesting a 40% lower risk of death for males in our cohort. Using the relative survival estimate, which allows for the fact that the females (on an average older) are more likely to die of other causes, the excess hazard ratio is 0.52 (0.45, 0.58) for males, a 48% lower risk compared with females. As with Cox models it is possible to adjust for multiple covariates, and for illustration we considered both sex and age in the models. The adjusted Cox estimate for males was 0.90 (0.82, 1.00), indicating a 10% lower risk of mortality after AMI. However, as we have seen, this methodology fails to separate death due to the index AMI from deaths due to other causes. Females in this cohort are older (mean 72.2) than males (mean of 64.1) and are thus more likely to die from causes other than the AMI. The adjusted relative survival model gives an estimate of 0.78 (0.69, 0.88) for males, suggesting a 22% lower excess mortality due to CHD for males compared with females post-MI, after adjustment for age. The advantage of this model is that it adjusts for the disease-associated mortality associated with age separately

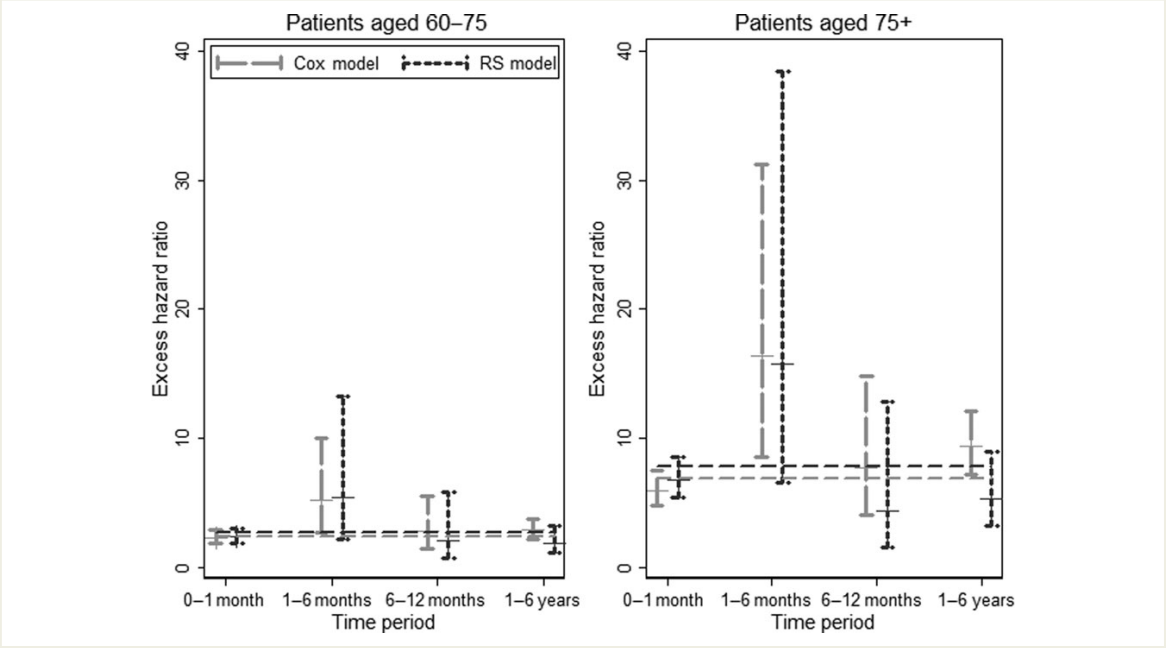


Figure 2 Time-dependent effects. Excess hazard ratio at intervals after index STEMI. The graphs illustrate excess risk of death for patients aged 60–75 or >75 years, each compared with patients aged <60 years. Vertical lines show the 95% confidence limits and point estimates of the excess hazard ratios from a relative survival model and hazard ratios from a Cox model. Horizontal lines show the point estimate of the excess hazard rate ratio from a relative survival model shown in *Table 2* and PH mortality rate ratio from a Cox model.

from the expected mortality experienced in the general population.

Discussion

What can relative survival methodology add to the assessment of coronary heart disease?

Relative survival provides clinically relevant information. Investigation of excess mortality rates has potential implications regarding patients' long-term survival. For example while much resource is dedicated to the management of 'young' patients with CHD, our initial estimates from the proportional excess hazards model suggest that for patients aged <60 years, the excess risk due to the index MI is around 30 additional deaths per 1000 person years over the first 6 months, but as few as six excess deaths estimated per 1000 person years after 3 years.

What are the limitations of relative survival?

Relative survival can be interpreted as a measure of mortality due to the disease of interest only if deaths due to the disease of interest are independent of the mortality in the general population. Even if this is not the case, relative survival still provides a useful comparison with mortality in the comparator population, which is usually the general population.[22] It may be worth considering alternative comparator groups when using relative survival methods in CHD to derive the expected mortality rates. For example, patients with CHD have a high prevalence of co-morbidities and of risk-factors, which put them at risk of mortality from causes other than CHD, such as pulmonary disease. We have used expected mortality from the general population, but use of expected mortality from a population with similar co-morbidities could be selected. For example, to assess survival after AMI in patients treated with insulin during the index admission, one could consider matching with expected survival by diabetic status. However, obtaining reliable information from such a population is often difficult.

A potentially important issue in the use of relative survival to the assessment of CHD survival is that in using population lifetables to derive the expected mortality rates, deaths due to the condition of interest are included. If the prevalence of that condition in the background population is low enough, then this will have little impact, a reasonable assumption for individual malignancies.[18] However, given the predominant contribution of heart disease to mortality in industrialized society, the appropriateness of this assumption in CHD needs to be assessed, in particular for oldest age groups.

We have presented some of the simpler statistical models for relative survival. However, there are several extensions that would also be applicable to CHD, for example, modelling time-dependent effects on a continuous scale through the use of splines[23–25] and fractional polynomials.[26]

Relative survival methodology merits attention in observational and population-based assessments of CHD mortality. Application of relative survival methods is still in early development in heart disease but its application in cancer is common, and informative. In order to obtain an estimate of net survival in population-based studies, relative survival applies sensible assumptions based on the deaths that we would expect to occur in a cohort of patients if they were from the general population. One development relevant in CHD may be calculating expected mortality in groups that are at a potentially higher risk of mortality from CHD, such as ethnic minorities, or social deprivation cohorts.

Conflict of interest: there are no conflicts of interest.

Funding

C.P.N. was funded by a British Heart Foundation PhD studentship (FS/05/080/19415).

References

1. Unal B, Critchley JA, Capewell S. Small changes in United Kingdom cardiovascular risk factors could halve coronary heart disease mortality. *J Clin Epidemiol* 2005;**58**:733–740.
2. Lauer MS, Blackstone EH, Young JB, Topol EJ. Cause of death in clinical research: Time for a reassessment? *J Am Coll Cardiol* 1999;**34**:618–620.
3. BHF. CHD 2006: Mortality. http://wwwheartstatsorg/temp/Chaptersp1%281%29hs1hspdf (accessed 31/10/2006).
4. Lloyd-Jones DM, Martin DO, Larson MG, Levy D. Accuracy of death certificates for coding coronary heart disease as the cause of death. *Ann Intern Med* 1998;**129**:1020–1026.
5. Mant J, Wilson S, Parry J, Bridge P, Wilson R, Murdoch W, Quirke T, Davies M, Gammage M, Harrison R, Warfield A. Clinicians didn't reliably distinguish between different causes of cardiac death using case histories. *J Clin Epidemiol* 2006;**59**:862–867.
6. Dickman PW, Adami HO. Interpreting trends in cancer patient survival. *J Intern Med* 2006;**260**:103–117.
7. Coleman M, Babb P, Damiecki P, Grosclaude P, Honjo S, Jones J, Knerer G, Pitard A, Quinn M, Sloggett A, Stavola BD. *Cancer Survival Trends in England and Wales, 1971–1995: Deprivation and NHS Region*. London: Office for National Statistics; 1999.
8. Coleman MP, Gatta G, Verdecchia A, Esteve J, Sant M, Storm H, Allemani C, Ciccolallo L, Santaquilani M, Berrino F, Group EW. EUROCARE-3 summary: cancer survival in Europe at the end of the 20th century. *Ann Oncol* 2003;**14**:128–149.
9. Norman PE, Semmens JB, Lawrence-Brown MMD. Long-term relative survival following surgery for abdominal aortic aneurysm: a review. *Cardiovasc Surg* 2001;**9**:219–224.
10. Kvidal P, Bergstrom PR, Horte L-G, Stahle E. Observed and relative survival after aortic valve replacement. *J Am Coll Cardiology* 2000;**35**:747–756.
11. Stahle E, Bergstrom R, Edlund B, Frostfeldt G, Lagerquist B, Sjogren I, Hansson HE. Influence of left ventricular function on survival after coronary artery bypass grafting. *Ann Thorac Surg* 1997;**64**:437–444.
12. Stahle E, Bergstrom R, Holmberg L, Edlund B, Nystrom SO, Sjogren I, Hansson HE. Survival after coronary artery bypass grafting. Experience from 4661 patients. *Eur Heart J* 1994;**15**:1204–1211.
13. Stahle E, Bergstrom R, Nystrom SO, Edlund B, Sjorgren I, Holmberg L. Surgical treatment of left ventricular aneurysm–assessment of risk factors for early and late mortality. *Eur J Cardiothorac Surg* 1994;**8**:67–73.
14. Patel PJ, Keating RJ, Gersh BJ, Hodge DO, Hammill SC, Shen WK. Outcome of patients with newly diagnosed atrial fibrillation at the Mayo Clinic and residing in that area. *Am J Cardiol* 2004;**94**:1379–1382.
15. Bronnum-Hansen H, Jorgensen T, Davidsen M, Madsen M, Osler M, Gerdes LU, Schroll M. Survival and cause of death after myocardial infarction: the Danish MONICA study. *J Clin Epidemiol* 2001;**54**:1244–1250.
16. Hardie K, Hankey GJ, Jamrozik K, Broadhurst RJ, Anderson C. Ten-year survival after first-ever stroke in the perth community stroke study. *Stroke* 2003;**34**:842–1846.
17. Hankey GJ, Jamrozik K, Broadhurst RJ, Forbes S, Burvill PW, Anderson CS, Stewart-Wynne EG. Five-year survival after first-ever stroke and related prognostic factors in the Perth Community Stroke Study. *Stroke* 2000;**31**:2080–2086.
18. Ederer F, Axtell LM, Cutler SJ. The relative survival rate: a statistical methodology. *Natl Cancer Inst Monogr* 1961;**6**:101–121.
19. Mitry E, Bouvier A-M, Esteve J, Faivre J. Improvement in colorectal cancer survival: a population-based study. *Eur J Cancer* 2005;**41**:2297–2303.
20. Stare J, Henderson R, Pohar M. An individual measure of relative survival. *Appl Stat* 2005;**54**:115–116.
21. GAD. Interim life tables of mortality rates for England and Wales. http://wwwgadgovuk/Life_Tables/Interim_life_tableshtm (accessed 30/10/2006).
22. Hakulinen T. On long-term relative survival rates. *J Chronic Dis* 1977;**30**:431–443.
23. Bolard P, Quantin C, Abrahamowicz M, Esteve J, Giorgi R, Chadha-Boreham H, Binquet C, Faivre J. Assessing time-by-covariate interactions in relative survival models using restrictive cubic spline functions. *J Cancer Epidemiol Prev* 2002;**7**:113–122.
24. Giorgi R, Abrahamowicz M, Quantin C, Bolard P, Esteve J, Gouvernet J, Faivre J. A relative survival regression model using B-spline functions to model non-proportional hazards. *Stat Med* 2003;**22**:767–2784.
25. Nelson CP, Lambert PC, Squire IB, Jones DR. Flexible parametric models for relative survival, with application in coronary heart disease. *Stat Med* 2007;**26**:5486–5498.
26. Lambert PC, Smith LK, Jones DR, Botha JL. Additive and multiplicative covariate regression models for relative survival incorporating fractional polynomials for time-dependent effects. *Stat Med* 2005;**24**:3871–3885.

Appendix E

In Press: Influence of diabetes diagnosis and admission blood glucose concentration on survival after acute myocardial infarction; results from 4702 index cases in routine practice

Influence of diabetes diagnosis and admission blood glucose concentration on survival after acute myocardial infarction; results from 4702 index cases in routine practice

Iain B. Squire, M.D.

Christopher P. Nelson, M.Sc.

Leong L. Ng, M.D.

David R. Jones, Ph.D.

Kent L. Woods, M.D.

Paul C. Lambert, Ph.D.

First Author : Squire
Short title : Glucose, diabetes and prognosis after STEMI

From: Departments of Cardiovascular Sciences (I.B.S., L.L.N., K.L.W.) and Health Sciences (C.P.N., D.R.J., P.C.L.) University of Leicester, Leicester, UK.

Address for correspondence:

Dr Iain B Squire
Department of Cardiovascular Sciences
Clinical Sciences Building
Leicester Royal Infirmary
Leicester LE2 7LX

Tel: +44 116 252 3125
Fax: +44 116 252 3108
e-mail: is11@le.ac.uk

Running Title: Diabetes, glucose and survival after myocardial infarction

Word Count: 3752 Tables:4

Sources of support: Christopher Nelson was supported by the British Heart Foundation

Conflicts of interest: None

ABSTRACT

Objective To compare the relative association with survival after ST-elevation AMI of the diagnosis of diabetes and of admission blood glucose concentration.

Research Design & Methods A retrospective cohort study in 4702 consecutive patients with ST elevation AMI occurring April 1st 1993 - December 31st 2005 assessed for mortality at 30 days and 1 year. Patients were classified according to antecedent diabetes, and by blood glucose concentration at admission (quartile 1, <7mmol/L; quartile 2, 7 - 8.2 mmol/L; quartile 3, 8.3 - 10.9 mmol/L; quartile 4, ≥11 mmol/L). Multivariable models were constructed for determinants of mortality, including year of AMI and demographic variables, entering admission blood glucose concentration and antecedent diabetes individually and together.

Results All-cause 30-day and 1-year mortality was 22.8% and 31.3% for patients with antecedent diabetes, compared to 16.3% and 23.0% respectively for those without. For glucose quartiles 1, 2, 3, and 4, 30-day mortality was 9.0%, 10.6%, 17.9% and 31.0%. Compared to glucose quartile 1, adjusted odds of mortality was similar in quartile 2, higher by >80% in quartile 3 and by >150% in quartile 4. Antecedent diabetes was associated with unadjusted odds for mortality of 1.52 (95% CI 1.24 , 1.86). On multivariable analysis (excluding glucose quartile), this excess reduced to 1.28 (1.01 , 1.63) and disappeared completely when glucose quartile was added to the analysis (adjusted OR 0.90 (0.69 , 1.16)). In contrast, inclusion of antecedent diabetes in multivariable models did not add to the predictive value for mortality of glucose quartile (p=0.368). Similar relationships were observed for 1 year mortality.

Conclusions In patients with STEMI, blood glucose concentration is of greater prognostic relevance than antecedent diabetes diagnosis. Moderate elevation of blood glucose, below levels previously considered to be clinically relevant, is associated with adverse impact on survival.

Abnormalities of glucose metabolism have powerful association with adverse prognosis for patients with coronary heart disease. In the setting of acute myocardial infarction (AMI), a prior diagnosis of diabetes is associated with increased risk of adverse outcome (1, 2). Moreover, elevated blood glucose is common among patients hospitalised with AMI, irrespective of diabetes status. This phenomenon,'stress hyperglycaemia', is also associated with adverse outcome (2,3,4,5,6,7). In patients with diabetes (8) and in unselected cohorts (9), persisting hyperglycaemia in the first 24-48 hours after AMI is associated with adverse outcome.

On this background, it has been suggested that early reduction of blood glucose may improve prognosis after AMI. However results from randomised trials assessing the impact of intensive, insulin-based therapy in this setting have been inconsistent. In the first such trial, intensive blood glucose management was associated with improved survival (10), a finding which was not replicated in a second study (11). Current algorithms for risk assessment after AMI do not consider blood glucose (12, 13) and contemporary AMI management guidelines do not include recommendations on target glucose concentrations (14, 15, 16). However, a recent position statement from the American Heart Association recommended assessment of blood glucose concentration in the routine management of ACS, and the "active management of significant hyperglycaemia" (17).

Most studies considering the prognostic impact of diabetes in patients with AMI have categorised patients based upon a history of the condition prior to the index event, an approach likely to underestimate prevalence. Prior studies assessing hyperglycaemia have for the most part considered blood glucose concentration as a dichotomised variable, using a variety of cut-off values (5, 18, 19). Moreover, most studies considering blood glucose as a graded variable were confined to selected cohorts, considering only patients with prior diabetes diagnosis (3, 7, 18) or those aged > 65 years (6), or used fasting glucose (19, 20) or glycated haemoglobin (21).

As already noted, previous studies in this area have largely investigated the association with outcome of either hyperglycaemia or previous diagnosis of diabetes; the comparative influence upon outcome of these factors in routine practice has not been assessed. The aim of the current study was to assess the relative association with short (30-day) and medium–term (1-year) survival after ST segment elevation AMI (STEMI) of antecedent diabetes diagnosis and of admission blood glucose concentration. We also wished to assess the lower limit of blood glucose concentration associated with adverse outcome, in an attempt to estimate "significant hyperglycaemia" in this context.

Research Design & Methods

Data are from consecutive admissions to the coronary care unit (CCU) of a large teaching hospital (Leicester Royal Infirmary), one of two serving the population of Leicestershire, UK, approximately 946,000 residents in 2004. For all admitted patients, we record clinical and demographic data including information on diagnosis (STEMI/NSTEMI), electrocardiographic (ECG) site of infarct and details of medical history, coronary heart disease risk-factors, and medication prescribed prior to and during admission. Mortality is recorded prospectively, as described previously (22). The current analysis uses data from patients admitted January 1st 1993 – December 31st 2005. For patients with more than 1 recorded admission, only the first was considered. In view of lack of routine data on mortality for patients resident outside of our Health Authority region we excluded these from analysis. Pre-defined outcome measures were the individual and relative strength of association with 30-day and 1-year, all-cause mortality of antecedent diabetes and admission blood glucose concentration.

Patient identification

On a background of changes during the study period in the definition of AMI, we maintained consistent inclusion criteria by restricting the analysis to patients with STEMI. The diagnosis required (i) ECG evidence of dynamic ST segment elevation together with (ii) appropriate

symptoms and (iii) increase in serum levels of CK to greater than twice the upper limit of the laboratory reference range (ie >400 IU/L). Troponin concentrations were not available for the full period of this study. Patients were categorised according to a history at the time of index admission of the diagnosis of diabetes, and by blood glucose concentration measured at that time. Antecedent diabetes was recorded if self-reported by the patient, or on the basis of prescribed medication.

Statistical analysis

Baseline differences between groups were examined using independent two-sample t-tests for continuous variables and chi-squared tests for categorical variables. Data are presented as differences in means and proportions, with 95% confidence intervals (CI). We calculated 30-day mortality proportions for the complete population with follow-up censored at January 30th 2006. One-year mortality proportions were calculated for patients admitted up to December 31st 2004, the most recent date at which complete 1-year follow-up data were available. During the study period 1993-2005, in line with accumulating evidence and clinical experience, increasing proportions of patients received treatment with statin, beta-blocker and renin-angiotensin system inhibitor agents. Such major changes render difficult the examination of individual treatment effects. However we made some correction for such trends by the inclusion in analyses of year of index AMI. As indications for, and proportions of patients receiving, thrombolysis (70% - 77% annually) were consistent throughout the study period, this variable was included in the model. The association between year of index AMI and all-cause mortality was assessed using unconditional logistic regression analysis with covariate effects reported as odds ratios (OR) with 95% CI. In all cases linear effect (log odds scale) provided the best fitting model (using Akaike information criterion). Other continuous variables (age, glucose, creatinine and peak creatine kinase) showed non-linear relationships with outcome and for exploratory, univariate analyses were categorised to ease

interpretation. Glucose, creatinine and peak creatine kinase were grouped by quartile, and age was divided by <65, 65-74, and ≥75 years.

The blood glucose concentration used was that first recorded for the index admission, assayed as part of routine investigations. Glucose was divided by quartiles of the range of values observed over the study period: Quartile 1, <7mmol/L; Quartile 2, 7 - 8.2 mmol/L; Quartile 3, 8.3 - 10.9 mmol/L; Quartile 4, ≥11 mmol/L. We initially assessed the unadjusted, univariate association with mortality of antecedent diabetes and of glucose quartile. Differential effects over time were assessed by fitting interactions between these covariates and year of diagnosis. We then fitted models adjusted for age, sex, peak CK, creatinine, previous AMI, year of index AMI and administration of thrombolysis. Fractional polynomials were used to model continuous variables. To investigate the effects of antecedent diabetes diagnosis and of glucose quartile, separate models were fitted including these individually, and in combination. Demographic features and outcomes were also assessed for the cohort of patients for whom admission glucose was not recorded. Analyses were performed using Stata 9 (StataCorp. 2005. Stata Statistical Software: Release 9. College Station, TX, USA).

RESULTS

From January 1st 1993 – December 31st 2004, we recorded 4876 consecutive, index admissions with STEMI. Excluding from analysis 174 (3.6%) individuals not normally resident locally, the study population consisted of 4702 patients (3198, 68.0% male; mean age 66.7, SD 12.7, range 21-107 years). Females (72.2 ± 11.4) were older than male patients (64.1 ± 12.4, $p<0.001$). For 683 (14.0%) and 2 patients respectively, information pertaining to admission blood glucose concentration and antecedent diabetes status was not available.

Antecedent Diabetes

Antecedent diabetes was recorded for 749 (15.9%) patients, increasing from 13.9% in 1993 to 20.3% in 2005. Patients with antecedent diabetes were older by an average of 1.4 years and more often female (Table 1). Mean admission glucose and creatinine concentrations were higher, but peak CK lower, in patients with antecedent diabetes. These patients were less likely to receive thrombolysis or to be prescribed aspirin or beta-blocker during the index admission, but more likely to receive diuretic therapy, inhibitors of the renin-angiotensin system and insulin.

Admission blood glucose

Mean admission glucose concentration was lower in 2005 (9.2 mmol/L) compared to 1993 (10.1 mmol/L, p=0.017), falling by an average of 0.39 mmol/L/year (95% CI 0.30 - 0.48) in those with, and 0.09 mmol/L/year (95% CI 0.05 - 0.12; p<0.001) in those without antecedent diabetes. Table 2 shows characteristics of patients according to glucose quartile, and for the subcohort in whom admission blood glucose was not recorded. Patients with glucose in quartile 4 were on average 5.6 years older than those in quartile 1. Even in quartiles 3 (10.3%) and 4 (42.7%) of blood glucose, a minority of patients had antecedent diabetes. Higher glucose quartile was associated with higher serum creatinine and lower admission systolic blood pressure. Thrombolysis was administered in fewer patients in quartile 4 compared to quartiles 1-3. In-hospital diuretic prescription increased, and beta-blocker prescription fell, as glucose quartile increased. There was little variation among quartiles in the use of inhibitors of the renin-angiotensin system.

In patients for whom admission blood glucose was not recorded, demographic and treatment features were broadly similar to those seen in the population as a whole. In terms of proportions with antecedent diabetes, treatment with insulin, and case-fatality, these patients most closely resembled those in quartile 3 (Table 2).

Survival

During follow up, case fatality was 42.2% (1992/4702). Eight hundred and forty one deaths (17.9% of the population, 42.2% of observed deaths), occurred by 30 days. By one year, 1136 patients (24.2% of the population, 57.0% of deaths) had died. The univariate strengths of association with important clinical and demographic variables are shown in Table 3. These were much as expected, with higher mortality risk associated with antecedent diabetes, previous MI, age at MI, blood glucose and creatinine, and lower risk associated with male sex and thrombolysis.

Antecedent Diabetes and Survival

Overall 30-day mortality was 22.8% and 1-year mortality was 31.3% for patients with antecedent diabetes, compared to 16.3% and 23.0% respectively for those without.

Blood glucose and Survival

Table 3 shows 30-day and 1-year mortality by glucose quartile. Blood glucose above the median (8.3mmol/L) was associated with increased mortality. For patients with glucose in quartile 4 (≥11 mmol/L), overall 30-day mortality was over 3-fold higher (31%) than in quartile 1 (9.0%). For 30-day mortality, compared to quartile 1, the OR for quartile 2 was 1.20 (95% CI 0.89 , 1.61), for quartile 3 was 2.20 (95% CI 1.67 , 2.91) and quartile 4 was 4.54 (95% CI 3.50 , 5.88).

Similar associations were evident for 1-year mortality. Compared to quartile 1, the OR for quartile 2 was 1.29 (95% CI 1.00 , 1.66), for quartile 3 was 2.05 (95% CI 1.61 , 2.62) and quartile 4 was 4.04 (95% CI 3.21 , 5.07).

Glucose concentration and antecedent diabetes – relative impact on prognosis

Using multivariable, unconditional logistic regression models, we assessed the impact upon prognosis of antecedent diabetes, glucose quartile (relative to quartile 1), and year of index AMI (Table 4). Unadjusted analyses for diabetes and glucose quartile were followed by analysis adjusted for age, sex, previous MI, peak CK, creatinine, thrombolysis, and year of index AMI, and including either diabetes or glucose individually. Finally, analysis was carried out adjusting for the same covariables and including both diabetes and glucose.

For patients with antecedent diabetes, unadjusted 30-day and 1-year mortality was approximately 50% higher compared to patients without this diagnosis (Table 4). Adjustment for covariables, excluding glucose quartile, reduced this excess to approximately 25%. However, with inclusion of glucose quartile in multivariable analyses, antecedent diabetes was associated with statistically non-significant lower risk of mortality at both 30 days (OR 0.90; 0.69 , 1.16) and 1-year (OR 0.94; 0.73 , 1.20).

In contrast, although attenuated by covariables adjustment, higher blood glucose concentrations retained powerful association with mortality. Compared to quartile 1, adjusted risk was similar in quartile 2, approximately 75% and 150% higher in quartile 3 and 4 respectively. The addition of antecedent diabetes did not materially alter the odds ratio associated with each level of glycaemia, or the overall model fit (Table 3).

Discussion

The aim of this study was to assess the relative prognostic relevance after AMI of two readily available measures of dysglycaemia in a large cohort of patients in routine practice. Our study presents two novel findings. First, blood glucose concentration at admission to hospital with STEMI had more powerful association with prognosis than did antecedent diabetes diagnosis. Second, even minor elevation of blood glucose concentration above the normal range was associated with adverse impact on outcome. Moreover, excess case fatality associated with antecedent diabetes was markedly attenuated when adjusted for covariables, and abolished entirely when correction included admission blood glucose. In contrast, the prior diagnosis of diabetes had no meaningful impact upon the risk associated with glucose concentration.

It is not clear whether hyperglycaemia contributes directly to adverse outcome after STEMI, or is simply a marker of risk. In this context, our observation of associations between higher blood glucose and a number of markers of adverse prognosis, eg lower blood pressure, higher creatinine and greater age, may support the latter. However the benefit of active reduction of blood glucose after AMI has been demonstrated in a number of studies (8, 9, 10, 23). Indeed the greater fall in blood glucose in the period following admission with AMI has been associated with better survival (8, 9, 10). Moreover, prescription of insulin in patients with AMI and not previously known to have diabetes is associated with improved outcome (23), and in both DIGAMI studies, effective glucose lowering soon after AMI was associated with improved survival (10, 11)**.** Indeed, it is noteworthy that the only controlled study of glucose lowering after AMI which lowered blood glucose compared to standard management was the first DIGAMI study; this was also the only such study in which active intervention was associated with improved outcome.

A potential explanation for our findings is that in patients in whom the diagnosis of diabetes is established, physicians may be more likely to employ active management of elevated admission blood glucose concentration. Such behaviour would explain in part previous reports suggesting that a given level of hyperglycaemia appears to have greater influence upon outcome in patients without, compared to those with, antecedent diabetes (5). While this point requires further investigation, the important point is that clinicians should consider blood glucose concentration irrespective of diabetes diagnosis.

While we observed little variation in outcome for blood glucose below 8.2mmol/L, mortality risk was increased above this level. These findings may be considered in the context of the target, and achieved, concentrations in trials of intensive glycaemic control after AMI. These studies recruited patients with admission glucose >11mmol/L (10,11), had target levels of 7-10.9 mmol/L (10) or 7-10 mmol/L (11), and achieved mean glucose levels of 9-10mmol/L (10,11). In one previous study, mortality was lowest in patients for whom blood glucose at 24

hours after AMI was less than 7mmol/L (8). Taken together, these observations suggest that in patients with STEMI both the threshold at which active intervention is applied and target blood glucose levels should be lower than previously considered. Our data lend indirect support to the recent statement from the American Heart Association that optimum blood glucose levels after AMI have yet to be established and should be the focus of future clinical trials (17).

Study limitations

Our database lacks information on some important variables, including objective assessment of left ventricular function and regarding clinical evidence of heart failure. Although blood glucose was unrecorded in a proportion of patients, we assessed demographics and outcome for this cohort, which we are confident did not bias our observations. In a proportion of our population, elevated blood glucose undoubtedly represents hitherto unrecognised diabetes. This limitation applies to the vast majority of studies in this area, classifying patients based upon antecedent diagnosis. It may be argued that some of our cohort, patients in quartile 4 for example, should be considered as having diabetes. However in many patients with elevated blood glucose concentration at admission with AMI, the diagnosis of diabetes is not confirmed on formal testing (24). Moreover or study was designed to be a pragmatic assessment of the impact on prognosis of admission blood glucose values, as it is this rather than later values, which is likely to guide acute management of blood glucose.

We did not adjust for prescription of individual secondary prevention therapies, the impact of which was not the focus of this analysis. Consideration of individual treatment effects may introduce bias for a number of reasons. In the most severely ill patients, early mortality and adverse clinical features in survivors will limit treatment prescription. Our study is limited by assessment of survival up to 1-year, and consideration of the influence of dysglycaemia and diabetes on longer term prognosis would be appropriate. Finally, we have no information on therapy or interventions after discharge. However, as the risk of death was greatest in the first

30 days, it is unlikely that changes after discharge impacted on overall outcome in a major way.

In summary, admission blood glucose concentration is a powerful, readily available marker of adverse outcome after ST elevation AMI, and is prognostically more informative than consideration of antecedent diabetes status. Minor elevation of blood glucose outwith the normal range is associated with adverse impact upon survival. Admission blood glucose concentration should be considered in the early assessment of prognosis after AMI. Further studies of intensive blood glucose management after AMI are merited.

Acknowledgements

Funding source: Christopher Nelson was funded by a British Heart Foundation PhD studentship (FS/05/080/19415).
Authors roles:
Study conception; IB Squire, KL Woods
Database management: KL Woods, IB Squire
Data analysis: CP Nelson, PC Lambert, DR Jones
Interpretation: IB Squire, PC Lambert, DR Jones, CP Nelson, LL Ng
Manuscript preparation: All authors
Conflicts of Interest: None

REFERENCES

1. Malmberg K, Yusuf S, Gerstein HC, Brown J, Zhao F, Hunt D, Piegas L, Calvin J, Keltai M, Budaj A. Impact of diabetes on long-term prognosis in patients with unstable angina and non-Q-wave myocardial infarction: results of the OASIS Registry. Circulation 2000; 102: 1014-1019.
2. McGuire DK, Emanuelsson H, Granger CB, Magnus Ohman E, Moliterno DJ, White HD, Ardissino D, Box JW, Califf RM, Topol EJ. Influence of diabetes mellitus on clinical outcomes across the spectrum of acute coronary syndromes. Findings from the GUSTO IIb study. Eur. Heart J. 2000; 21: 1750-1758
3. Svensson A-M, McGuire DK, Abrahamsson P, Dellborg M. Association between hyper- and hypoglycaemia and 2-year all-cause mortality risk in diabetic patients with acute coronary events. Eur. Heart J. 2005 26; 1255-1261
4. Capes SE, Hunt D, Malmberg K, Gerstein HC. Stress hyperglycaemia and increased risk of death after myocardial infarction in patients with and without diabetes: a systematic overview. Lancet 2000; 355: 773-778
5. Wahab NN, Cowden EA, Pearce NJ, Gardner MJ, Merry H, Cox JL. Is blood glucose an independent predictor of mortality in acute myocardial infarction in the thrombolytic era ? J. Am. Coll. Cardiol. 2002; 40: 1748- 1754
6. Kosiborod M, Rathore SS, Inzucchi SE, Masoudi FA, Wang Y, Havranek EP, Krumholz HM. Admission glucose and mortality in elderly patients hospitalised with acute myocardial infarction. Circulation 2005; 111: 3078-3086.
7. Cao JJ, Hudson M, Jankowski M, Whitehouse F, Weaver WD. Relation of chronic and acute glycaemic control on mortality in acute myocardial infarction with diabetes mellitus. Am. J. Cardiol. 2005; 96: 183-186
8. Ghoyal A, Mahaffey KW, Garg J, Nicolau JC, Hochman JS, Weaver WD, Theroux P, Oliveira GB, Todaro TG, Mojcik CF, Armstrong PW, Granger CB. Prognostic significance of the change in glucose level in the first 24h after acute myocardial infarction: results from the CARDINAL study. Eur. Heart J. 2006; 27: 1289-1297
9. Kosiborod M, Inzucchi SE, Krumholz HM, Xiao L, Jones PG, Fiske S, Masoudi, FA, Marso SP, Spertus JA. Glucometrics in Patients Hospitalized With Acute Myocardial Infarction. Circulation 2008; 117; 1018-1027
10. Malmberg K, Ryden L, Efendic S, Herlitz J, Nicol P, Waldenström A, Wedel H, Welin L. Randomised trial of insulin-glucose infusion followed by subcutaneous insulin treatment in diabetic patients with acute myocardial infarction (DIGAMI study): effects on mortality at 1 year. J.Am.Coll. Cardiol. 1995; 26: 57-65

11. Malmberg K, Ryden L, Wedel H, Birkeland K, Bootsma A, Dickstein K, Efendic S, Fisher M, Hamsten A, Herlitz J, Hildebrandt P, MacLeod K, Laakso M, Torp-Pedersen C, Waldenström A; DIGAMI 2 Investigators. Intense metabolic control by means of insulin in patients with diabetes mellitus and acute myocardial infarction (DIGAMI 2): effects on mortality and morbidity. Eur. Heart J. 2005; 26: 650-661
12. Morrow DA, Antman EM, Giugliano RP, Cairns R, Charlesworth A, Murphy SA, de Lemos JA, McCabe CH, Braunwald E. A simple risk index for rapid triage of patients with ST elevation myocardial infarction: an InTIME II substudy. Lancet 2001; 358: 1571-1575
13. Granger CB, Goldberg RJ, Dabbous O, Pieper KS, Eagle KA, Cannon CP, Van De Werf F, Avezum A, Goodman SG, Flather MD, Fox KA; Global Registry of Acute Coronary Events Investigators. Predictors of hospital mortality in the global registry of acute coronary events. Arch. Int.Med 2003; 163: 2345-2353
14. Van de Werf F, Ardissino D, Betriu A, Cokkinos DV, Falk E, Fox KA, Julian D, Lengyel M, Neumann FJ, Ruzyllo W, Thygesen C, Underwood SR, Vahanian A, Verheugt FW, Wijns W; Task Force on the Management of Acute Myocardial Infarction of the European Society of Cardiology. Management of acute myocardial infarction in patients presenting with ST-segment elevation. The Task Force on the Management of Acute Myocardial Infarction of the European Society of Cardiology. Eur. Heart J. 2003; 24: 28-66
15. Antman EM, Anbe DT, Armstrong PW, Bates ER, Green LA, Hand M, Hochman JS, Krumholz HM, Kushner FG, Lamas GA, Mullany CJ, Ornato JP, Pearle DL, Sloan MA, Smith SC Jr, Alpert JS, Anderson JL, Faxon DP, Fuster V, Gibbons RJ, Gregoratos G, Halperin JL, Hiratzka LF, Hunt SA, Jacobs AK; American College of Cardiology/American Heart Association Task Force on Practice Guidelines. ACC/AHA guidelines for the management of patients with ST-elevation myocardial infarction--executive summary. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to revise the 1999 guidelines for the management of patients with acute myocardial infarction). J Am Coll Cardiol. 2004; 44: 671-719
16. Garber AJ, Moghissi ES, Bransome ED Jr, Clark NG, Clement S, Cobin RH, Furnary AP, Hirsch IB, Levy P, Roberts R, Van den Berghe G, Zamudio V; American College of Endocrinology Task Force on Inpatient Diabetes Metabolic Control. Endocr. Pract. 2004; 10: 77-82

17. Deedwania P, Kosiborod M; Barrett E, Ceriello A, Isley W, Mazzone T, Raskin P. Hyperglycemia and Acute Coronary Syndrome. A Scientific Statement From the American Heart Association DiabetesCommittee of the Council on Nutrition, Physical Activity, and Metabolism. Circulation 2008; 117: 1610-1619
18. Lynch M, Gammage D, Lamb P, Nattrass M, Pentecost BL. Acute myocardial infarction in diabetic patients in the thrombolytic era.Diabetes. Med. 1994; 11: 162-165
19. Schiele F, Descotes-Genon V, Seronde MF, Blonde MC, Legalery P, Meneveau N, Ecarnot F, Mercier M, Penfornis A, Thebault L, Boumal D, Bassand JP; Investigators of the Réseau Franc Comtois de Cardiologie. Predictive value of admission hyperglycaemia on mortality in patients with acute myocardial infarction. Diab. Med. 2006; 23: 1370-1376
20. Suleiman M, Hammerman H, Boulos M, Kapeliovich MR, Suleiman A, Agmon Y, Markiewicz W, Aronson D. Fasting glucose is an important independent risk factor for 30 day mortality in patients with acute myocardial infarction. Circulation 2005; 111: 754-760
21. Soler NG, Frank S. Value of glycosylated haemoglobin measurements after acute myocardial infarction. JAMA 1981: 246: 1690-1693
22. Blackledge HM, Newton J, Squire IB. Prognosis for South Asian and white patients newly admitted to hospital with heart failure in the United Kingdom: historical cohort study. BMJ 2003; 327: 526-31
23. Weston C, Walker L, Birkhead J. Early impact of insulin treatment on mortality for hyperglycaemic patients without known diabetes who present with an acute coronary syndrome. Heart 2007; 93: 1542-1546
24. Tenerz A, Lonnberg I, Berne C, Nilsson G, Leppert J. Myocardial infarction and prevalence of diabetes mellitus. Is increased casual blood glucose at admission a reliable criterion for the diagnosis of diabetes? Eur Heart J 2001;22:1102–1110

	Antecedent Diabetes		
	Yes N=749 (15.9%)	**No N=3951 (84.1%)**	**Difference (Diabetes – No Diabetes)**
Male (%)	460 (61.4%)	2737 (69.3%)	-7.9 (-11.6 , -4.1)*
Age (years)	67.9 ± 0.4	66.5 ± 0.2	1.4 (0.5 , 2.3)†
Plasma glucose (mmol/L)[1]	14.7 ± 0.3	8.9 ± 0.1	5.7 (5.2 , 6.3)†
Creatinine (μmol/L)[2]	120.7 ±2.3	109.2 ±0.7	11.5 (6.8 , 16.3)†
Peak CK (IU/L)[3] ; Normal range < 200	1818 ± 62	2115 ± 31	-296 (-433 , -159)†
History of			
Smoking			
Current (%)	156 (20.8%)	1312 (33.2%)	-12.4 (-15.7 , -9.1)*
Former (%)	183 (24.4%)	992 (25.1%)	-0.7 (-4.0 , 2.7)*
Never (%)	410 (54.7%)	1645 (41.6%)	13.1 (9.2 , 17.0)*
Not known (%)	0	2	
Previous AMI	175 (23.4%)	571 (14.5%)	8.9 (5.7 , 12.1)*
Angina (%)	230 (30.7%)	862 (21.8%)	8.9 (5.3 , 12.4)*
Hypertension (%)	368 (49.1%)	1296 (32.8%)	16.3 (12.5 , 20.2)*
Hyperlipidaemia (%)	151 (20.2%)	451 (11.4%)	8.7 (5.7 , 11.8)*
Cerebrovascular disease (%)	83 (11.1%)	191 (4.8%)	6.2 (3.9 , 8.6)*
Treatment			
Thrombolysis (%)	468 (62.5%)	2883 (73.0%)	-10.5 (-14.2 , -6.8)*
Diuretic (%)	457 (61.0%)	1743 (44.1%)	16.9 (13.1 , 20.7)*
Aspirin (%)	528 (70.5%)	3099 (78.4%)	-7.9 (-11.5 , -4.4)*
Beta Blocker (%)	282 (37.7%)	1798 (45.5%)	-7.9 (-11.7 , -4.1)*
Insulin (%)	510 (68.1%)	280 (7.1%)	61.0 (57.6 , 64.4)*
ACE Inhibitor / ARB (%)	385 (51.4%)	1580 (40.0%)	11.4 (7.5 , 15.3)*
Statin (%)	196 (26.2%)	1054 (26.7%)	-0.5 (-3.9 , 2.9)*

Table 1: Demographic and in-hospital treatment characteristics of patients with and without antecedent diagnosis of diabetes. Table shows mean ± standard error or number (%). CK = Creatine Kinase; ACE = Angiotensin converting Enzyme; ARB = Angiotensin Receptor Blocker.

[1] 101 missing values for diabetes and 581 missing values for no diabetes;

[2] 9 missing values for diabetes and 39 missing values for no diabetes;

[3] 9 missing values for diabetes and 39 missing values for no diabetes

* Calculated difference in proportions (with 95% confidence intervals)

† Calculated difference in means (with 95% confidence intervals)

		GQ1 <7		GQ2 7-8.3		GQ3 8.3-11		GQ4 ≥11		Missing Glucose		Overall	
		n	%	n	%	n	%	n	%	n	%	n	%
Sex	Male	713	76.7	776	72.1	609	63.9	643	60.7	457	66.9	3,198	68.0
	Female	217	23.3	300	27.9	344	36.1	417	39.3	226	33.1	1,504	32.0
Diabetes	No	882	94.8	1,027	95.4	855	89.7	606	57.2	581	85.1	3,951	84.0
	Yes	48	5.2	49	4.6	98	10.3	453	42.7	101	14.8	749	15.9
	Missing	0	0	0	0	0	0	1	0.1	1	0.1	2	0.04
Statin	No	567	61.0	803	74.6	714	74.9	859	81.0	508	74.4	3,451	73.4
	Yes	363	39.0	273	25.4	239	25.1	201	19.0	175	25.6	1,251	26.6
ACE or ARB	No	499	53.7	655	60.9	547	57.4	608	57.4	427	62.5	2,736	58.2
	Yes	431	46.3	421	39.1	406	42.6	452	42.6	256	37.5	1,966	41.8
Diuretic	No	596	64.1	645	59.9	483	50.7	418	39.4	360	52.7	2,502	53.2
	Yes	334	35.9	431	40.1	470	49.3	642	60.6	323	47.3	2,200	46.8
Beta Blocker	No	373	40.1	578	53.7	534	56.0	733	69.2	403	59.0	2,621	55.7
	Yes	557	59.9	498	46.3	419	44.0	327	30.8	280	41.0	2,081	44.3
Thrombolysis	No	272	29.2	248	23.0	193	20.2	408	38.5	229	33.5	1,350	28.7
	Yes	658	70.8	828	77.0	760	79.8	652	61.5	454	66.5	3,352	71.3
Insulin	No	895	96.2	1046	97.2	877	92.0	493	46.5	600	87.9	3911	83.2
	Yes	35	3.8	30	2.8	76	8.0	567	53.5	83	12.2	791	16.8
	n	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Overall mean(sd)**	
Age	4,702	63.8	13.27	65.3	12.88	67.6	12.10	69.4	11.54	67.4	13.06	66.7(12.69)	
SBP	4,627	141.8	28.06	139.8	28.49	137.6	31.71	135.4	36.48	137.2	31.02	138.4(31.44)	
Creatinine (µmol/L)	4,654	103.3	47.33	102.6	35.78	109.1	38.73	126.3	56.83	113.8	5.64	111.0(48.19)	

Table 2: Demographic and in-hospital treatment characteristics of patients according to glucose quartile. ACE = Angiotensin converting Enzyme; ARB = Angiotensin Receptor Blocker; SBP = Systolic Blood Pressure. † 1-year survival was assessed on patients admitted up to 31st December 2004, being the last date for which 1-year follow-up was available. Therefore the total number of patients included in the one-year analysis (n=4474) is fewer than in the 30-day analysis (n=4702).

Variable		30 Day Mortality		1 Year Mortality	
		No. Deaths/ No. at risk (%)	Odds Ratio (95% CI)	No. Deaths/ No. at risk (%)	Odds Ratio (95% CI)
Antecedent Diabetes	No	550/3370 (16.3)	1	743/3225 (23.0)	1
	Yes	148/648 (22.8)	1.52(1.24 , 1.86)	192/614 (31.3)	1.52 (1.24 , 1.84)
Sex	Female	316/1278 (24.7)	1	403/1224 (32.9)	1
	Male	382/2740 (13.9)	0.49 (0.42 , 0.58)	532/2615 (20.3)	0.52 (0.45 , 0.61)
Thrombolysis	No	307/1120 (27.4)	1	405/1065 (38.0)	1
	Yes	391/2898 (13.5)	0.41 (0.35 , 0.49)	530/2774 (19.1)	0.38 (0.33 , 0.45)
Previous MI	No	541/3388 (16.0)	1	704/3231 (21.7)	1
	Yes	157/630 (24.9)	1.75 (1.43 , 2.14)	231/608 (38.0)	2.20 (1.83 , 2.64)
Age (years)					
	<65	117/1669 (7.0)	1	150/1574 (9.5)	1
	65-74	187/1153 (16.2)	2.57 (2.01 , 3.02)	256/1111 (23.0)	2.84 (2.28 , 3.54)
	≥75	394/1196 (32.9)	6.52 (5.21 , 8.15)	529/1154 (45.8)	8.04 (6.55 , 9.86)
Glucose (quartile)					
	<7.3	84/930 (9.0)	1	122/871 (14.0)	1
	7.3-8.2	114/1076 (10.6)	1.20 (0.89 , 1.61)	180/1037 (17.4)	1.29 (1.00 , 1.66)
	8.3-10.9	171/953 (17.9)	2.20 (1.67 , 2.91)	228/910 (25.1)	2.05 (1.61 , 2.62)
	≥11	329/1059 (31.1)	4.54 (3.50 , 5.88)	405/1021 (39.7)	4.04 (3.21 , 5.07)
Creatinine (quartile)					
	<87	80/972 (8.2)	1	114/927 (12.3)	1
	87-100.9	97/1032 (9.4)	1.16 (0.85 , 1.58)	135/971 (13.9)	1.15 (0.88 , 1.50)
	101-120.9	139/1000 (13.9)	1.80 (1.35 , 2.41)	203/961 (21.1)	1.91 (1.49 , 2.45)
	≥121	381/1009 (37.8)	6.76 (5.21 , 8.79)	483/976 (49.5)	6.99 (5.54 , 8.82)
Peak CK (quartile)					
	<752	253/997 (25.4)	1	299/927 (32.3)	1
	752-1567	131/989 (13.3)	0.45 (0.36 , 0.57)	196/940 (20.9)	0.55 (0.45 , 0.68)
	1568-2808	140/1000 (14.0)	0.48 (0.38 , 0.60)	202/971 (20.8)	0.55 (0.45 , 0.68)
	≥2809	168/1019 (16.5)	0.58 (0.47 , 0.72)	232/990 (23.4)	0.64 (0.53 , 0.79)
Year	Annual		0.94 (0.92 , 0.96)		0.95 (0.93 , 0.97)

Table 3. Univariate association with 30-day and all-cause mortality

Parameter		Unadjusted (n=4018)	Adjusted[1] (Excluding Diabetes) (n=4002)	Adjusted[2] (Excluding Glucose) (n=4002)	Adjusted[3] (n=4002)
30 DAY					
Glucose	Q1	1	1	-	1
	Q2	1.20 (0.89 , 1.61)	1.08 (0.77 , 1.50)	-	1.08 (0.78 , 1.50)
	Q3	2.20 (1.67 , 2.91)	1.74 (1.27 , 2.37)	-	1.75 (1.28 , 2.397)
	Q4	4.54 (3.50 , 5.88)	2.50 (1.86 , 3.36)	-	2.60 (1.91 , 3.55)
Diabetes	No	1	-	1	1
	Yes	1.52 (1.24 , 1.86)	-	1.28 (1.01 , 1.63)	0.90 (0.69 , 1.16)
Year		0.94 (0.92 , 0.96)	0.93 (0.919 , 0.96)	0.93 (0.90 , 0.95)	0.94 (0.91 , 0.96)
AIC			0.7162787	0.7291038	0.7166097
Parameter		Unadjusted (n=3839)	Adjusted (n=3825)	Adjusted (n=3825)	Adjusted[1] (n=3825)
ONE YEAR					
Glucose	Q1	1	1	-	1
	Q2	1.29 (1.00 , 1.66)	1.24 (0.93 , 1.65)	-	1.24 (0.93 , 1.66)
	Q3	2.05 (1.61 , 2.62)	1.64 (1.24 , 2.17)	-	1.64 (1.24 , 2.18)
	Q4	4.04 (3.21 , 5.07)	2.28 (1.74 , 2.99)	-	2.340 (1.759 , 3.112)
Diabetes	No	1	-	1	1
	Yes	1.52 (1.268 , 1.84)	-	1.25 (1.00 , 1.57)	0.94 (0.73 , 1.20)
Year		0.95 (0.93 , 0.97)	0.95 (0.93 , 0.98)	0.94 (0.92 , 0.97)	0.95 (0.93 , 0.98)
AIC			0.8408397	0.8509601	0.8412917

Table 4:
Results of modelling proportions surviving to 30 days and one year.
Adjusted models are adjusted for age at MI, sex, previous AMI, CK, creatinine, thrombolysis and year of hospitalisation : [1] for blood glucose concentration (quartile); [2] antecedent diabetes and [3] for both blood glucose concentration (quartile) and antecedent diabetes.
Year = Odds ratio per year 1993-2005 (30-day) or 1993-2004 (1-year)
AIC = Akaike's information criterion (lower values indicate a better fitting model).

Bibliography

[1] D. M. Lloyd-Jones, D. O. Martin, M. G. Larson, and D. Levy. Accuracy of death certificates for coding coronary heart disease as the cause of death. *Annals of Internal Medicine*, 129(12):1020–1026, 1998.

[2] C. P. Nelson, P. C. Lambert, I. B. Squire, and D. R. Jones. Relative survival: what can cardiovascular disease learn from cancer? *European Heart Journal*, 29(7):941–947, 2008.

[3] C. P. Nelson, P. C. Lambert, I. B. Squire, and D. R. Jones. Flexible parametric models for relative survival, with application in coronary heart disease. *Statistics in Medicine*, 26(30):5486–5498, 2007.

[4] B. Unal, J. A. Critchley, and S. Capewell. Small changes in united kingdom cardiovascular risk factors could halve coronary heart disease mortality. *Journal of Clinical Epidemiology*, 58(7):733–40, 2005.

[5] M. S. Lauer, E. H. Blackstone, J. B. Young, and E. J. Topol. Cause of death in clinical research: Time for a reassessment? *Journal of the American College of Cardiology*, 34(3):618–620, 1999.

[6] BHF. Chd 2006: Mortality. *http://www.heartstats.org/temp/Chaptersp1%281%29hs1hs.pdf*, (accessed 15/04/2008).

[7] P. Vineis and F. Faggiano. Commentary: Epidemiological models and prevention of cancer. *Annals of Oncology*, 2(8):559–563, 1991.

[8] C.L. Chiang. Introduction to stochastic process in biostatistics. *Wiley, New York*, 1968.

[9] A. Mathew and M. Pandey. Attributing death to cancer: cause-specific survival estimation. *Journal of Postgraduate Medicine*, 48(4):322–326, 2002.

[10] J. Mant, S. Wilson, J. Parry, P. Bridge, R. Wilson, W. Murdoch, T. Quirke, M. Davies, M. Gammage, R. Harrison, and A. Warfield. Clinicians didn't reliably distinguish between different causes of cardiac death using case histories. *Journal of Clinical Epidemiology*, 59(8):862–867, 2006.

[11] P. W. Dickman and H. O. Adami. Interpreting trends in cancer patient survival. *Journal of Internal Medicine*, 260(2):103–117, 2006.

[12] M.P. Coleman, P. Babb, P. Damiecki, P. Grosclaude, S. Honjo, J. Jones, G. Knerer, A. Pitard, M. Quinn, A. Sloggett, and B. De Stavola. Cancer survival trends in england and wales, 1971-1995: Deprivation and nhs region. *Office for National Statistics, London.*, 1999.

[13] M. P. Coleman, G. Gatta, A. Verdecchia, J. Esteve, M. Sant, H. Storm, C. Allemani, L. Ciccolallo, M. Santaquilani, F. Berrino, and Eurocare Working Group. Eurocare-3 summary: cancer survival in europe at the end of the 20th century. *Annals of Oncology*, 14(5):v128–149, 2003.

[14] M. P. Coleman, M. Quaresma, F. Berrino, J-M. Lutz, Roberta. De Angelis, Riccardo. Capocaccia, P. Baili, B. Rachet, G. Gatta, T. Hakulinen, A. Micheli, M. Sant, H.K. Weir, J. M. Elwood, H. Tsukuma, S. Koifman, G. A. e Silva, S. Francisci, M. Santaquilani, A. Verdecchia, H.H. Storm, and J.L. Young. Cancer survival in five continents: a worldwide population-based study (concord). *The Lancet Oncology*, 9(8):730–756, 2008.

[15] P. E. Norman, J. B. Semmens, and M. Lawrence-Brown. Long-term relative survival following surgery for abdominal aortic aneurysm: a review. *Cardiovascular Surgery*, 9(3):219–224, 2001.

[16] P. Kvidal, R. Bergstrom, L-G. Horte, and E. Stahle. Observed and relative survival after aortic valve replacement. *Journal of the American College of Cardiology*, 35(3):747–756, 2000.

[17] E. Stahle, P. Kvidal, S. O. Nystrom, and R. Bergstrom. Long-term relative survival after primary heart valve replacement. *European Journal of Cardio-Thoracic Surgery*, 11(1):81–91, 1997.

[18] E. Stahle, R. Bergstrom, L. Holmberg, B. Edlund, S. O. Nystrom, I. Sjogren, and H. E. Hansson. Survival after coronary artery bypass grafting. experience from 4661 patients. *European Heart Journal*, 15(9):1204–1211, 1994.

[19] E. Stahle, R. Bergstrom, S. O. Nystrom, B. Edlund, I. Sjorgren, and L. Holmberg. Surgical treatment of left ventricular aneurysm–assessment of risk factors for early and late mortality. *European Journal of Cardio-Thoracic Surgery*, 8:67–73, 1994.

[20] P. J. Patel, R. J. Keating, B. J. Gersh, D. O. Hodge, S. C. Hammill, and W. K. Shen. Outcome of patients with newly diagnosed atrial fibrillation at the mayo clinic and residing in that area. *The American Journal of Cardiology*, 94(11):1379–1382, 2004.

[21] H. Bronnum-Hansen, T. Jorgensen, M. Davidsen, M. Madsen, M. Osler, L. U. Gerdes, and M. Schroll. Survival and cause of death after myocardial infarction: the danish monica study. *Journal of Clinical Epidemiology*, 54(12):1244–50, 2001.

[22] K. Hardie, G. J. Hankey, K. Jamrozik, R. J. Broadhurst, and C. Anderson. Ten-year survival after first-ever stroke in the perth community stroke study. *Stroke*, 34(8):1842–6, 2003.

[23] G. J. Hankey, K. Jamrozik, R. J. Broadhurst, S. Forbes, P. W. Burvill, C. S. Anderson, and E. G. Stewart-Wynne. Five-year survival after first-ever stroke and related prognostic factors in the perth community stroke study. *Stroke*, 31:2080–6, 2000.

[24] E. Mitry, A-M. Bouvier, J. Esteve, and J. Faivre. Improvement in colorectal cancer survival: a population-based study. *European Journal of Cancer*, 41(15):2297–303, 2005.

[25] D.R. Cox. Regression models and life tables. *Journal of the Royal Statistical Society Series B*, 34:187–220, 1972.

[26] F. Ederer, L. M. Axtell, and S. J. Cutler. The relative survival rate: a statistical methodology. *National Cancer Institute Monographs*, 6:101–121, 1961.

[27] J. Esteve, E. Benhamou, M. Croasdale, and L. Raymond. Relative survival and the estimation of net survival: elements for further discussion. *Statistics in Medicine*, 9(5):529–538, 1990.

[28] R. De Angelis, R. Capocaccia, T. Hakulinen, B. Soderman, and A. Verdecchia. Mixture models for cancer survival analysis: application to population-based data with covariates. *Statistics in Medicine*, 18(4):441–454, 1999.

[29] P. C. Lambert, J. R. Thompson, C. L. Weston, and P. W. Dickman. Estimating and modelling the cure fraction in population-based cancer survival analysis. *Biostatistics*, 8(3):576–594, 2007.

[30] GAD. Interim life tables of mortality rates for england and wales. *http://www.gad.gov.uk/Demography_Data/Life_Tables/Interim_life_tables.asp*, (accessed 14/04/2008).

[31] National Instruments. Electrocariograph figure. *http://zone.ni.com/cms/images/devzone/tut/2007-07-09_141618.jpg*, (accessed 04/09/2008).

[32] T. Hakulinen. Cancer survival corrected for heterogeneity in patient withdrawal. *Biometrics*, 38(4):933–942, 1982.

[33] F. Ederer and H. Heise. Instructions to ibm 650 programmers in processing survival computations. *Methodological note No. 10, End Results Evaluation Section, National Cancer Institute*, Bethesda MD, 1959.

[34] J. Stare, R. Henderson, and M. Pohar. An individual measure of relative survival. *Applied Statistics*, 54(1):115–116, 2005.

[35] B. Yu, R. C. Tiwari, K. A. Cronin, and E. J. Feuer. Cure fraction estimation from the mixture cure models for grouped survival data. *Statistics in Medicine*, 23(11):1733–1747, 2004.

[36] T. Hakulinen. On long-term relative survival rates. *Journal of Chronic Diseases*, 30(7):431–43, 1977.

[37] P. Bolard, C. Quantin, M. Abrahamowicz, J. Esteve, R. Giorgi, H. Chadha-Boreham, C. Binquet, and J. Faivre. Assessing time-by-covariate interactions in relative survival models using restrictive cubic spline functions. *Journal of Cancer Epidemiology & Prevention*, 7(3):113–122, 2002.

[38] R. Giorgi, M. Abrahamowicz, C. Quantin, P. Bolard, J. Esteve, J. Gouvernet, and J. Faivre. A relative survival regression model using b-spline functions to model non-proportional hazards. *Statistics in Medicine*, 22(17):2767–2784, 2003.

[39] P. C. Lambert, L. K. Smith, D. R. Jones, and J. L. Botha. Additive and multiplicative covariate regression models for relative survival incorporating fractional polynomials for time-dependent effects. *Statistics in Medicine*, 24(24):3871–3885, 2005.

[40] T. Hakulinen and L. Tenkanen. Regression analysis of relative survival rates. *Applied Statistics*, 36(3):309–317, 1987.

[41] H. Brenner, O. Gefeller, and T. Hakulinen. A computer program for period analysis of cancer patient survival. *European Journal of Cancer*, 38(5):690–695, 2002.

[42] V. Arndt, M. Talback, O. Gefeller, T. Hakulinen, and H. Brenner. Modification of sas macros for a more efficient analysis of relative survival rates. *European Journal of Cancer*, 40(5):778–779, 2004.

[43] A. Bergenfelz, A. Bladstrm, M. Their, E. Nordenstrm, S. Valdemarsson, and J. Westerdahl. Serum levels of uric acid and diabetes mellitus influence survival after surgery for primary hyperparathyroidism: A prospective cohort study. *World Journal of Surgery*, 31(7):1393–1400, 2007.

[44] G. Dellgren, M. J. Eriksson, L. A. Brodin, and K. Radegran. Eleven year's experience with the biocor stentless aortic bioprosthesis: clinical and hemodynamic follow-up with long-term relative survival rate. *European Journal of Cardio-Thoracic Surgery*, 22(6):912–921, 2002.

[45] L. Hellgren, P. Kvidal, L-G. Horte, U-B. Krusemo, and E. Stahle. Survival after mitral valve replacement: Rationale for surgery before occurrence of severe symptoms. *The Annals of Thoracic Surgery*, 78(4):1241–1247, 2004.

[46] D. Lindblom, U. Lindblom, J. Qvist, and H. Lundstrom. Long-term relative survival rates after heart valve replacement. *Journal of the American College of Cardiology*, 15(3):566–573, 1990.

[47] C. Linde-Edelstam, B. O. Gullberg, R. Norlander, S. K. Pehrsson, M. Rosenqvist, and L. Ryden. Longevity in patients with high degree atrioventricular block paced in the atrial synchronous or the fixed rate ventricular inhibited mode. *Pacing and Clinical Electrophysiology*, 15(3):304–313, 1992.

[48] E. Stahle, R. Bergstrom, B. Edlund, G. Frostfeldt, B. Lagerquist, I. Sjogren, and H. E. Hansson. Influence of left ventricular function on survival after coronary artery bypass grafting. *Annals of Thoracic Surgery*, 64(2):437–444, 1997.

[49] A. Uddhammar, AL. Eriksson, L. Nystrom, R. Stenling, and S. Rantapaa-Dahlqvist. Increased mortality due to cardiovascular disease in patients with giant cell arteritis in northern sweden. *Journal of Rheumatology*, 29(4):737–742, 2002.

[50] M. Abdelnoor, S. Nitter-Hauge, and S. Trettli. Relative survival of patients after heart valve replacement. *European Heart Journal*, 11(1):23–8, 1990.

[51] S. Aune, S. R. Amundsen, J. Evjensvold, and A. Trippestad. Operative mortality and long-term relative survival of patients operated on for asymptomatic abdominal aortic aneurysm. *European Journal of Vascular and Endovascular Surgery*, 9(3):293–298, 1995.

[52] S. Aune, S. R. Amundsen, and A. Trippestad. Early morbidity and patterns of survival after carotid endarterectomy. *European Journal of Surgery*, 163(2):101–5, 1997.

[53] S. Aune, S.R. Amundsen, and A. Trippestad. The influence of age on long-term survival pattern of patients operated on for lower limb ischaemia. *European Journal of Vascular and Endovascular Surgery*, 12(2):214–217, 1996.

[54] S. Aune, S. R. Amundsen, J. Evjensvold, and A. Trippestad. The influence of age on operative mortality and long-term relative survival following emergency abdominal aortic aneurysm operations. *European Journal of Vascular and Endovascular Surgery*, 10(3):338–341, 1995.

[55] S. Aune and A. Trippestad. Operative mortality and long-term survival of patients operated on for acute lower limb ischaemia. *European Journal of Vascular and Endovascular Surgery*, 15(2):143–146, 1998.

[56] S. Aune. Risk factors and operative results of patients aged less than 66 years operated on for asymptomatic abdominal aortic aneurysm. *European Journal of Vascular and Endovascular Surgery*, 22(3):240–243, 2001.

[57] S. Aune, E. Laxdal, G. Pedersen, and E. Dregelid. Patient characteristics, operative complications and long-term survival of patients aged 75 years or older subjected to carotid endarterectomy. *International Angiology*, 22(4):421–5, 2003.

[58] E.S. Haug, P. Romundstad, S. Aune, T.B.J. Hayes, and H.O. Myhre. Elective open operation for abdominal aortic aneurysm in octogenarians–survival analysis of 105 patients. *European Journal of Vascular and Endovascular Surgery*, 29(5):489–495, 2005.

[59] K. S. Dujardin, M. Enriquez-Sarano, H. V. Schaff, K. R. Bailey, J. B. Seward, and A. J. Tajik. Mortality and morbidity of aortic regurgitation in clinical practice. a long-term follow-up study. *Circulation*, 99(14):1851–7, 1999.

[60] R. B. Singer. Comparative mortality in medically treated aortic regurgitation. *Journal of Insurance Medicine*, 36:10–15, 2004.

[61] C. M. Tribouilloy, M. Enriquez-Sarano, H. V. Schaff, T. A. Orszulak, K. R. Bailey, A. J. Tajik, and R. L. Frye. Impact of preoperative symptoms on survival after surgical correction of organic mitral regurgitation: rationale for optimizing surgical indications.[see comment]. *Circulation*, 99(3):400–405, 1999.

[62] P. Sorajja, R. A. Nishimura, S. R. Ommen, Ackerman M. J., A J. Tajik, and B. J. Gersh. Use of echocardiography in patients with hypertrophic cardiomyopathy: Clinical implications of massive hypertrophy. *Journal of the American Society of Echocardiography : official publication of the American Society of Echocardiography*, 19(6):788–795, 2006.

[63] E. Klodas, M. Enriquez-Sarano, A. J. Tajik, C. J. Mullany, K. R. Bailey, and J. B. Seward. Aortic regurgitation complicated by extreme left ventricular dilation: long-term outcome after surgical correction.[see comment]. *Journal of the American College of Cardiology*, 27(3):670–677, 1996.

[64] D. Messika-Zeitoun, H. Thomson, M. Bellamy, C. Scott, C. Tribouilloy, J. Dearani, A. J. Tajik, H. Schaff, and M. Enriquez-Sarano. Medical and surgical outcome of tricuspid regurgitation caused by flail leaflets. *Journal of Thoracic and Cardiovascular Surgery*, 128(2):296–302, 2004.

[65] R. B. Singer. Randomized trial of carvedilol in treatment of congestive heart failure. *Journal of Insurance Medicine*, 29:82–90, 1997.

[66] J. F. Avierinos, B. J. Gersh, 3rd Melton, L. J., K. R. Bailey, C. Shub, R. A. Nishimura, A. J. Tajik, and M. Enriquez-Sarano. Natural history of asymptomatic

mitral valve prolapse in the community.[see comment]. *Circulation*, 106(11):1355–61, 2002.

[67] J. W. Hallet Jr, J. M. Naessens, and D. J. Ballard. Early and late outcome of surgical repair for small abdominal aortic aneurysms: a population-based analysis. *Journal of Vascular Surgery*, 18(4):684–691, 1993.

[68] B. J. Maron, S. A. Casey, L. C. Poliac, T. E. Gohman, A. K. Almquist, and D. M. Aeppli. Clinical course of hypertrophic cardiomyopathy in a regional united states cohort. *Journal of the American Medical Association*, 281(7):650–655, 1999.

[69] R. J. Pokorski. Effect of age on mortality experience in patients with hypertrophic cardiomyopathy. *Journal of Insurance Medicine*, 29:43–48, 1997.

[70] C. S. Rihal, B. J. Gersh, J. P. Whisnant, T. W. Rooke, Jr. Sundt, T. M., W. M. O'Fallon, and D. J. Ballard. Influence of coronary heart disease on morbidity and mortality after carotid endarterectomy: a population-based study in olmsted county, minnesota (1970-1988)[erratum appears in j am coll cardiol 1992 nov 1;20(5):1304]. *Journal of the American College of Cardiology*, 19(6):1254–1260, 1992.

[71] W-K. Shen, D. L. Hayes, S. C. Hammill, K. R. Bailey, D. J. Ballard, and B. J. Gersh. Survival and functional independence after implantation of a permanent pacemaker in octogenarians and nonagenarians: A population-based study. *Annals of Internal Medicine*, 125(6):476–480, 1996.

[72] R. B. Singer. Mortality derived from 5-year survival in patients with alzheimer disease. *Journal of Insurance Medicine (Seattle)*, 37(4):264–71, 2005.

[73] A. Naslafkih. Expected versus observed survival in 3 large population studies with hmg-coa reductase inhibitors. *Journal of Insurance Medicine*, 32:155–162, 2000.

[74] S. Jacobsen, J. Petersen, S. Ullman, P. Junker, A. Voss, J. M. Rasmussen, U. Tarp, L. H. Poulsen, G. van Overeem Hansen, B. Skaarup, T. M. Hansen, J. Podenphant, and P. Halberg. Mortality and causes of death of 513 danish patients with systemic lupus erythematosus. *Scandinavian Journal of Rheumatology*, 28(2):75–80, 1999.

[75] J. Launbjerg, P. Fruergaard, J. K. Madsen, L. S. Mortensen, and J. F. Hansen. Ten year mortality in patients with suspected acute myocardial infarction. *British Medical Journal*, 308(6938):1196–1199, 1994.

[76] J. Launbjerg, P. Fruergaard, J. K. Madsen, and J. F. Hansen. Three-year mortality in patients suspected of acute myocardial infarction with and without confirmed diagnosis. *American Heart Journal*, 122(5):1270–1273, 1991.

[77] O. Lund, H. K. Pilegaard, L. B. Ilkjaer, S. L. Nielsen, H. Arildsen, and O. K. Albrechtsen. Performance profile of the starr-edwards aortic cloth covered valve,

track valve, and silastic ball valve. *European Journal of Cardio-Thoracic Surgery*, 16(4):403–413, 1999.

[78] O. Lund, K. Magnussen, M. Knudsen, H. Pilegaard, T.T. Nielsen, and O.K. Albrechtsen. The potential for normal long term survival and morbidity rates after valve replacement for aortic stenosis. *The Journal of heart valve disease*, 5(3):258–267, 1996.

[79] K. H. Olesen, I. H. Rygg, A. Wennevold, and J. Nyboe. Aortic valve replacement with the lillehei-kaster prosthesis in 262 patients: an assessment after 9 to 17 years. *European Heart Journal*, 12(6):680–689, 1991.

[80] H. Bronnum-Hansen, M. Davidsen, P. Thorvaldsen, and Monica Study Group Danish. Long-term survival and causes of death after stroke. *Stroke*, 32(9):2131–6, 2001.

[81] M. Ketonen, P. Pajunen, H. Koukkunen, P. Immonen-Raiha, J. Mustonen, M. Mahonen, M. Niemela, K. Kuulasmaa, P. Palomaki, M. Arstila, T. Vuorenmaa, A. Lehtonen, S. Lehto, H. Miettinen, J. Torppa, J. Tuomilehto, J. Airaksinen, K. Pyorala, and V. Salomaa. Long-term prognosis after coronary artery bypass surgery. *International Journal of Cardiology*, 124(1):72–79, 2008.

[82] M. Niskanen, A. Kari, and P. Halonen. Five-year survival after intensive care–comparison of 12,180 patients with the general population. finnish icu study group. *Critical Care Medicine*, 24(12):1962–1967, 1996.

[83] M. Lehecka, M. Niemel, J. Seppnen, H. Lehto, T. Koivisto, A. Ronkainen, J. Rinne, R. Sankila, J. Jskelinen, and J. Hernesniemi. No long-term excess mortality in 280 patients with ruptured distal anterior cerebral artery aneurysms. *Neurosurgery*, 60(2):235–240, 2007.

[84] G.M. Shahin, G.J. van der Heijden, J.C. Kelder, M. Boulaksil, P.J. Knaepen, and A.J. Six. Long-term follow-up of mitral valve repair: a single-center experience. *Medical Science Monitor*, 12(7):CR308–314, 2006.

[85] H. A. Verheul, R. B. van den Brink, B. J. Bouma, G. Hoedemaker, A. C. Moulijn, E. Dekker, P. Bossuyt, and A. J. Dunning. Analysis of risk factors for excess mortality after aortic valve replacement. *Journal of the American College of Cardiology*, 26(5):1280–6, 1995.

[86] J. Iacovino. Mortality outcome of surgically treated atrial septal defects.[see comment]. *Journal of Insurance Medicine (Seattle)*, 33(1):37–41, 2001.

[87] N. J. Goodson, N. J. Wiles, M. Lunt, E. M. Barrett, A. J. Silman, and D. P. Symmons. Mortality in early inflammatory polyarthritis: cardiovascular mortality is increased in seropositive patients.[see comment]. *Arthritis & Rheumatism*, 46(8):2010–9, 2002.

[88] G. F. Salles, K. V. Bloch, and C. R. Cardoso. Mortality and predictors of mortality in a cohort of brazilian type 2 diabetic patients. *Diabetes Care*, 27(6):1299–305, 2004.

[89] D. Rusinaru, I. Saaidi, S. Godard, H. Mahjoub, C. Battle, and C. Tribouilloy. Impact of chronic obstructive pulmonary disease on long-term outcome of patients hospitalized for heart failure. *The American Journal of Cardiology*, 101(3):353–358, 2008.

[90] P. E. Norman, J. B. Semmens, M. M. Lawrence-Brown, and C. D. Holman. Long term relative survival after surgery for abdominal aortic aneurysm in western australia: population based study. *British Medical Journal*, 317(7162):852–856, 1998.

[91] Jr. Adams, K. F., J. H. Patterson, W. A. Gattis, C. M. O'Connor, C. R. Lee, T. A. Schwartz, and M. Gheorghiade. Relationship of serum digoxin concentration to mortality and morbidity in women in the digitalis investigation group trial: a retrospective analysis.[see comment]. *Journal of the American College of Cardiology*, 46(3):497–504, 2005.

[92] M. N. Ashraf, A. Mortasawi, A. D. Grayson, and A. Y. Oo. Effect of smoking status on mortality and morbidity following coronary artery bypass surgery. *Thoracic & Cardiovascular Surgeon*, 52(5):268–73, 2004.

[93] J. I. Barzilay, R. A. Kronmal, V. Bittner, E. Eaker, C. Evans, and E. D. Foster. Coronary artery disease and coronary artery bypass grafting in diabetic patients aged ¿ or = 65 years (report from the coronary artery surgery study [cass] registry). *American Journal of Cardiology*, 74(4):334–9, 1994.

[94] J. N. Beattie, S. S. Soman, K. R. Sandberg, J. Yee, S. Borzak, M. Garg, and P. A. McCullough. Determinants of mortality after myocardial infarction in patients with advanced renal dysfunction. *American journal of kidney diseases : the official journal of the National Kidney Foundation*, 37(6):1191–1200, 2001.

[95] D. F. Del Rizzo, W. D. Boyd, R. J. Novick, F. N. McKenzie, N. D. Desai, and A. H. Menkis. Safety and cost-effectiveness of midcabg in high-risk cabg patients.[see comment]. *Annals of Thoracic Surgery*, 66(3):1002–7, 1998.

[96] L. Hiltunen. Ten-year mortality and glucose tolerance status in an elderly finnish population. *Diabetes Research and Clinical Practice*, 69:81–87, 2005.

[97] H. H. Tan, R. R. McAlpine, P. James, P. Thompson, M. E. McMurdo, A. D. Morris, J. M. Evans, and Darts Memo collaboration. Diagnosis of type 2 diabetes at an older age: effect on mortality in men and women. *Diabetes Care*, 27(12):2797–2799, 2004.

[98] N. Kalant and I. Shrier. Volume and outcome of coronary artery bypass graft surgery: are more and less the same? *Canadian Journal of Cardiology*, 20(1):81–86, 2004.

[99] R. J. Pokorski. Long-term survival of patients with infective endocarditis. *Journal of Insurance Medicine*, 30:76–87, 1998.

[100] G. D'Ancona, H. Karamanoukian, T. Lajos, M. Ricci, J. Bergsland, and T. Salerno. Posterior thoracotomy for reoperative coronary artery bypass grafting without cardiopulmonary bypass: Perioperative results. *The Heart Surgery Forum*, 3(1):18–23, 2000.

[101] R. de Marco, F. Locatelli, L. Cazzoletti, M. Bugianio, A. Carosso, and A. Marinoni. Incidence of asthma and mortality in a cohort of young adults: a 7-year prospective study. *Respiratory Research*, 6:95, 2005.

[102] K. Majamaa-Voltti, J. Turkka, M. L. Kortelainen, H. Huikuri, and K. Majamaa. Causes of death in pedigrees with the 3243a¿g mutation in mitochondrial dna. *J Neurol Neurosurg Psychiatry*, 79(2):209–211, 2008.

[103] T. J. Papadimos, R. H. Habib, A. Zacharias, T. A. Schwann, C. J. Riordan, S. J. Durham, and A. Shah. Early efficacy of cabg care delivery in a low procedure-volume community hospital: operative and midterm results. *BMC Surgery*, 5:10, 2005.

[104] G. I. Barbash, M. Modan, U. Goldbourt, H. D. White, and F. Van de Werf. Comparative case fatality analysis of the international tissue plasminogen activator/streptokinase mortality trial: variation by country beyond predictive profile. the investigators of the international tissue plasminogen activator/streptokinase mortality trial. *Journal of the American College of Cardiology*, 21(2):281–6, 1993.

[105] R. A. Lawrance, M. F. Dorsch, R. J. Sapsford, A. F. Mackintosh, D. C. Greenwood, B. M. Jackson, C. Morrell, M. B. Robinson, and A. S. Hall. Use of cumulative mortality data in patients with acute myocardial infarction for early detection of variation in clinical practice: observational study.[see comment]. *British Medical Journal*, 323(7308):324–327, 2001.

[106] F. Van de Werf, E. J. Topol, K. L. Lee, L. H. Woodlief, C. B. Granger, P. W. Armstrong, G. I. Barbash, J. R. Hampton, A. Guerci, R. J. Simes, and al et. Variations in patient management and outcomes for acute myocardial infarction in the united states and other countries. results from the gusto trial. global utilization of streptokinase and tissue plasminogen activator for occluded coronary arteries. *Journal of the American Medical Association*, 273(20):1586–1591, 1995.

[107] C. S. Rihal, D. E. Grill, M. R. Bell, P. B. Berger, K. N. Garratt, and Jr. Holmes, D. R. Prediction of death after percutaneous coronary interventional procedures. *American Heart Journal*, 139(6):1032–1038, 2000.

[108] H. Karabulut, F. Toraman, C. Alhan, G. Camur, S. Evrenkaya, S. Dagdelen, and S. Tarcan. Euroscore overestimates the cardiac operative risk. *Cardiovascular Surgery*, 11(4):295–298, 2003.

[109] T. Laatikainen, J. Critchley, E. Vartiainen, V. Salomaa, M. Ketonen, and S. Capewell. Explaining the decline in coronary heart disease mortality in finland between 1982 and 1997. *American Journal of Epidemiology*, 162(8):764–773, 2005.

[110] B. C. Kross, L. F. Burmeister, L. K. Ogilvie, L. J. Fuortes, and C. M. Fu. Proportionate mortality study of golf course superintendents.[see comment]. *American Journal of Industrial Medicine*, 29(5):501–506, 1996.

[111] B. J. Witt, R. D. Brown Jr, S. J. Jacobsen, S. A. Weston, K. V. Ballman, R. A. Meverden, and V. L. Roger. Ischemic stroke after heart failure: A community-based study. *American Heart Journal*, 152(1):102–109, 2006.

[112] N. Mantel and W. Haenzsel. Statistical aspects of the analysis of data from retrospective studies of disease. *National Cancer Institute*, 22:719–748, 1959.

[113] N.E. Breslow. Analysis of survival data under the proportional hazards model. *International Statistical Review*, 43:45–57, 1975.

[114] J. D. Buckley. Additive and multiplicative models for relative survival rates. *Biometrics*, 40(1):51–62, 1984.

[115] P. W. Dickman, A. Sloggett, M. Hills, and T. Hakulinen. Regression models for relative survival. *Statistics in Medicine*, 23(1):51–64, 2004.

[116] N.E. Breslow, J.H. Lubin, P. Marek, and B. Langholz. Multiplicative models and cohort analysis. *Journal of The American Statistical Association*, 78:1–12, 1983.

[117] T. Hakulinen and K.H. Abeywickrama. A computer program package for relative survival analysis. *Computer Programs in Biomedicine*, 19(2-3):197–207, 1985.

[118] L. Remontet, N. Bossard, A. Belot, J. Estve, and the French network of cancer registries FRANCIM. An overall strategy based on regression models to estimate relative survival and model the effects of prognostic factors in cancer survival studies. *Statistics in Medicine*, 26(10):2214–2228, 2007.

[119] H. Brenner and T. Hakulinen. On crude and age-adjusted relative survival rates. *Journal of Clinical Epidemiology*, 56(12):1185–91, 2003.

[120] P. W. Dickman, E. Coviello, and M. Hills. Estimating and modelling relative survival. *Stata Journal, (in press)*.

[121] P. McCullagh and J.A. Nelder. Generalized linear models, second edition. *Monographs on Statistics and Applied Probability*.

[122] M. Aitkin, D. Anderson, B. Francis, and J. Hinde. Statistical modelling in glim. *Oxford University Press*, 1989.

[123] H. Akaike. A new look at the statistical model identification. *IEEE Transactions On Automatic Control*, 19(6):716–723, 1974.

[124] P. Royston and D. Altman. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Journal of the Royal Statistical Society: Series A*, 43(3):429–467, 1994.

[125] B. Efron and R.J. Tibshirani. An introduction to the bootstrap. *Monographs on Statistics and Applied Probability*, 1994.

[126] P. Royston and W. Sauerbrei. Stability of multivariable fractional polynomial models with selection of variables and transformations: a bootstrap investigation. *Statistics in Medicine*, 22(4):639–659, 2003.

[127] P.C. Lambert. Fractional polynomials and model averaging. *http://www.stata.com/meeting/2sweden/lambert_fpma.pdf*, (accessed 18/07/2008), 2007.

[128] J. A. Hoeting, D.E.R.A. Madigan, and C.T. Volinsky. Bayesian model averaging: A tutorial. *Statistical Science*, 14(4):382–417, 1997.

[129] P. Congdon. Model weights for model choice and averaging. *Statistical Methodology*, 4:143–157, 2007.

[130] K. P. Burnham and D. R. Anderson. Multimodal inference: Understanding aic and bic in model selection. *Sociological Methods and Research*, 33(2):261–304, 2004.

[131] S. Buckland, K. Burnham, and N. Augustin. Model selection: An intergral part of inference. *Biometrics*, 53(2):603–618, 2007.

[132] C. Faes, M.H.G. Aerts, and G. Molenberghs. Model averaging using fractional polynomials to estimate a safe level of exposure. *Risk Analysis*, 27(1):111–123, 2007.

[133] C. J. Stone and C. Y. Koo. Additive splines in statistics. in proceedings of the statistical computing section 45–48. *Journal of the American Statistical Association*, 1986.

[134] R. Giorgi, J. Payan, and J. Gouvernet. Rsurv: A function to perform relative survival analysis with s-plus or r. *Computer Methods and Programs in Biomedicine*, 78(2):175–178, 2005.

[135] P. Royston and M.K.B. Parmar. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine*, 21(15):2175–2197, 2002.

[136] E. A. Weller, E. J. Feuer, C. M. Frey, and M. N. Wesley. Parametric relative survival regression using generalized linear models with application to hodgkin's lymphoma. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 48(1):79–89, 1999.

[137] H. Lan, K.A. Cronin, K.A. Johnson, A.B. Mariotto, and E.J. Feuer. Improved survival time: What can survival cure models tell us about population-based survival improvements in late-stage colorectal, ovarian, and testicular cancer? *Cancer*, 112(10):2289–2300, 2008.

[138] W. Gould, J. Pitblado, and W. Sribney. Maximum likelihood estimation with stata: second edition. *Stata Press*, 2003.

[139] C. Cox. The generalized F distribution: An umbrella for parametric survival analysis. *Statistics in Medicine*, 27(21):4301–4312, 2008.

[140] P. Royston, D.G. Altman, and W. Sauerbrei. Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in Medicine*, 25(1):127–141, 2005.

[141] B. Rachet, M. Abrahamowicz, A. J. Sasco, and J. Siemiatycki. Estimating the distribution of lag in the effect of short-term exposures and interventions: adaptation of a non-parametric regression spline model. *Statistics in Medicine*, 22(14):2335–2363, 2003.

[142] S. Durrleman and R. Simon. Flexible regression-models with cubic-splines. *Statistics in Medicine*, 8(5):551–561, 1989.

[143] F. E. Harrell Jr. Regression modelling strategies. with application to linear models, logistic regression and survival analysis. *Springer series in Statistics*, 2001.

[144] D. Pillay, K. Bhaskaran, S. Jurriaans, M. Prins, B. Masquelier, F. Dabis, R. Gifford, C. Nielsen, C. Pedersen, C. Balotta, G. Rezza, M. Ortiz, C. de Mendoza, C. Kücherer, G. Poggensee, J. Gill, and K. Porter. Cascade virology collaboration: The impact of transmitted drug resistance on the natural history of hiv infection and response to first-line therapy. *AIDS*, 20(1):21–28, 2006.

[145] P. Royston, M. K. B. Parmar, and W. Qian. Novel designs for multi-arm clinical trials with survival outcomes with an application in ovarian cancer. *Statistics in Medicine*, 22(14):2239–2256, 2003.

[146] J. T. Griffin, C. Fraser, L. Gras, F. de Wolf, and A. C. Ghani. The effect on treatment comparisons of different measurement frequencies in human immunodeficiency virus observational databases. *American Journal of Epidemiology*, 163(7):676–683, 2006.

[147] P. Royston. st0001. flexible parametric alternatives to the cox model, and more. *Stata Journal*, 1(1):1–28, 2001.

[148] F.J. Aranda-Ordaz. On two families of transformations to additivity for binary response data. *Biometrika*, 68(2):357–363, 1981.

[149] N. Younes and J. Lachin. Link-based models for survival data with interval and continuous time censoring. *Biometrics*, 53(4):1199–1211, 1997.

[150] C. Radhakrishna Rao. Linear statistical inference and its application. *Wiley-Interscience; 2^{nd} edition*, 2002.

[151] E. Marubini and M.G. Valsecchi. Analysing survival data from clinical trials and observational studies. *Wiley Blackwell*, 1995.

[152] S. Bennett. Analysis of survival data by the proportional odds model. *Statistics in Medicine*, 2(2):273–277, 1983.

[153] P. Armitage, G. Berry, and J.N.S. Matthews. Statistical methods in medical research - fourth edition. *Blackwell Science*, 2001.

[154] Stata. Stata help for predictnl. *http://www.stata.com/help.cgi?predictnl*, (accessed 24/04/2008).

[155] Y. Li and J. Feng. A nonparametric comparison of conditional distributions with nonnegligible cure fractions. *Lifetime Data Analysis*, 11(3):367–387, 2005.

[156] P. Royston and W. Sauerbrei. st0120. multivariable modeling with cubic regression splines: A principled approach. *Stata Journal*, 7(1):45–70, 2007.

[157] Stata. Stata help for statistical software components (ssc). *http://www.stata.com/help.cgi?ssc*, (accessed 19/08/2008).

[158] R. Bender, T. Augustin, and M. Blettner. Generating survival times to simulate cox proportional hazards models. *Statistics in Medicine*, 24(11):1713–1723, 2005.

[159] HeartStats.org. Prevalence of myocardial infarction, low values. *http://www.heartstats.org/temp/Tabsp2.10spweb07.xls*, (accessed 20/06/2008).

[160] HeartStats.org. Prevalence of myocardial infarction, high values. *http://www.heartstats.org/temp/Tabsp2.6spweb07.xls*, (accessed 20/06/2008).

[161] R. Stevenson, K. Ranjadayalan, P. Wilkinson, R. Roberts, and A. D. Timmis. Short and long term prognosis of acute myocardial infarction since introduction of thrombolysis. *British Medical Journal*, 307(6900):349–353, 1993.

[162] P. Congdon. Bayesian statistical modelling - second edition. *Wiley Series in Probability and Statistics*, 2006.

[163] J. G. Ibrahim and Sinha D. Chen, M-H. Bayesian survival analysis. *Springer series in Statistics*, 2001.

[164] BHF. Heartstats.org. *http://www.heartstats.org/homepage.asp*, (accessed 10/07/2008).

[165] H. Brenner and O. Gefeller. An alternative approach to monitoring cancer patient survival. *Cancer*, 78(9):2004–2010, 1996.

[166] H. Brenner and O. Gefeller. Deriving more up-todate estimates of long-term patient survival. *Journal of Clinical Epidemiology*, 50(2):211–216, 1997.

[167] H. Brenner, O. Gefeller, and T. Hakulinen. Period analysis for 'up-to-date' cancer survival data: theory, empirical evaluation, computational realisation and applications. *European Journal of Cancer*, 40(3):326–335, 2004.

[168] L. K. Smith, P. C. Lambert, J. L. Botha, and D. R. Jones. Providing more up-to-date estimates of patient survival: a comparison of standard survival analysis with period analysis using life-table methods and proportional hazards models. *Journal of clinical epidemiology*, 57(1):14–20, 2004.

[169] H. Brenner, C. Stegmaier, and H. Ziegler. Trends in survival of patients with breast cancer in saarland, germany. *British Journal of Cancer*, 78(5):694–697, 1998.

[170] H. Brenner, C. Stegmaier, and H. Ziegler. Trends in survival of patients with ovarian cancer in saarland, germany, 19761995. *Journal of Cancer Research and Clinical Oncology*, 125(2):109–113, 1999.

[171] H. Brenner, O. Gefeller, C. Stegmaier, and H. Ziegler. More up-to-date monitoring of long-term survival rates by cancer registries: an empirical example. *Methods of Information in Medicine*, 40(3):248–253, 2001.

[172] H. Brenner and T. Hakulinen. Long-term cancer patient survival achieved by the end of the 20th century: most up-to-date estimates from the nationwide finnish cancer registry. *British Journal of Cancer*, 85(3):367–371, 2001.

[173] H. Brenner. Long-term survival rates of cancer patients achieved by the end of the 20th century: a period analysis. *Lancet*, 360(9340):1131–1135, 2002.

[174] T. Aareleid and H. Brenner. Trends in cancer patient survival in estonia before and after the transition from a soviet republic to an open-market economy. *International Journal of Cancer*, 102(1):45–50, 2002.

[175] H. Brenner and T. Hakulinen. Advanced detection of time trends in long term cancer patient survival: experience from 50 years of cancer registration in finland. *American Journal of Epidemiology*, 156(6):566–577, 2002.

[176] H. Brenner and T. Hakulinen. Up-to-date long-term survival curves of patients with cancer by period analysis. *Journal of Clinical Oncology*, 20(3):826–832, 2002.

[177] H. Brenner, B. Soderman, and T. Hakulinen. Use of period analysis for providing more up-to-date estimates of long-term survival rates: empirical evaluation among 370 000 cancer patients in finland. *International Journal of Epidemiology*, 31(2):456–462, 2002.

[178] H. Brenner and C. Spix. Combining cohort and period methods for retrospective time trend analyses of long-term cancer patient survival rates. *British Journal of Cancer*, 89(7):1260–1265, 2003.

[179] H. Brenner and B. Rachet. Hybrid analysis for up-to-date long-term survival rates in cancer registries with delayed recording of incident cases. *European Journal of Cancer*, 40(16):2494–2501, 2004.

[180] A. B. Mariotto, M. N. Wesley, K. A. Cronin, K. A. Johnson, and E. J. Feuer. Estimates of long-term survival for newly diagnosed cancer patients. *Cancer*, 106(9):2039–2050, 2006.

[181] H. Brenner and T. Hakulinen. Up-to-date and precise estimates of cancer patient survival: Model-based period analysis. *American Journal of Epidemiology*, 164(7):689–696, 2006.

[182] H. Brenner and T. Hakulinen. Maximizing the benefits of model-based period analysis of cancer patient survival. *Cancer Epidemiology Biomarkers & Prevention*, 16(8):1675–1681, 2007.

[183] A. Gondos, F. Bray, D. H. Brewster, J. W. W. Coebergh, T. Hakulinen, M. L. G. Janssen-Heijnen, J. Kurtinaitis, and H. Brenner. Recent trends in cancer survival across europe between 2000 and 2004: A model-based period analysis from 12 cancer registries. *European Journal of Cancer*, 44(10):1463–1475, 2008.

[184] H. Brenner, A. Gondos, and V. Arndt. Recent major progress in long-term cancer patient survival disclosed by modeled period analysis. *Journal of Clinical Oncology*, 25(22):3274–3280, 2007.

[185] H. Brenner and T. Hakulinen. Period versus cohort modeling of up-to-date cancer survival. *International Journal of Cancer*, 122(4):898–904, 2008.

[186] I.B. Squire, C.P. Nelson, L.L. Ng, D.R. Jones, K.L. Woods, and P.C. Lambert. Prognostic impact of dysglycaemia in acute myocardial infarction; a comparison of admission blood glucose and antecedent diabetes in 4702 consecutive admissions. *Submitted*, 2008.

[187] D.K. McGuire, H. Emanuelsson, C.B. Granger, O.E. Magnus, D.J. Moliterno, H.D. White, D. Ardissino, J.W. Box, R.M. Califf, and E.J. Topol. Influence of diabetes mellitus on clinical outcomes across the spectrum of acute coronary syndromes. findings from the gusto iib study. *European Heart Journal*, 21(21):1750–1758, 2000.

[188] A. Svensson, D. K. McGuire, P. Abrahamsson, and M. Dellborg. Association between hyper- and hypoglycaemia and 2 year all-cause mortality risk in diabetic patients with acute coronary events. *European Heart Journal*, 26(13):1255–1261, 2005.

[189] S. E. Capes, D. Hunt, K. Malmberg, and H. C. Gerstein. Stress hyperglycaemia and increased risk of death after myocardial infarction in patients with and without diabetes: a systematic overview. *The Lancet*, 355(9206):773–778, 2000.

[190] N. N. Wahab, E. A. Cowden, N. J. Pearce, M. J. Gardner, H. Merry, J. L. Cox, and ICONS Investigators. Is blood glucose an independent predictor of mortality in acute myocardial infarction in the thrombolytic era? *Journal of the American College of Cardiology*, 40(10):1748–1754, 2002.

[191] M. Kosiborod, S. S. Rathore, S. E. Inzucchi, F. A. Masoudi, Y. Wang, E. P. Havranek, and H. M. Krumholz. Admission Glucose and Mortality in Elderly Patients Hospitalized With Acute Myocardial Infarction: Implications for Patients With and Without Recognized Diabetes. *Circulation*, 111(23):3078–3086, 2005.

[192] J.J. Cao, M. Hudson, M. Jankowski, F. Whitehouse, and W.W. Douglas. Relation of chronic and acute glycemic control on mortality in acute myocardial infarction with diabetes mellitus. *The American Journal of Cardiology*, 96(2):183–186, 2005.

[193] D. A. Morrow, E. M. Antman, R. P. Giugliano, R. Cairns, A. Charlesworth, S. A. Murphy, J. A. de Lemos, C. H. McCabe, and E. Braunwald. A simple risk index for rapid initial triage of patients with st-elevation myocardial infarction: an intime ii substudy. *The Lancet*, 358(9293):1571–1575, 2001.

[194] C. B. Granger, R. J. Goldberg, O. Dabbous, K. S. Pieper, K. A. Eagle, C. P. Cannon, F. Van de Werf, A. Avezum, S. G. Goodman, M. D. Flather, and K. A. A. Fox. Predictors of Hospital Mortality in the Global Registry of Acute Coronary Events. *Archives of Internal Medicine*, 163(19):2345–2353, 2003.

[195] K.A.A. Fox, O. H. Dabbous, R. J. Goldberg, K. S. Pieper, K. A. Eagle, F. Van de Werf, A. Avezum, S. G. Goodman, M. D. Flather, F. A. Jr Anderson, and C. B Granger. Prediction of risk of death and myocardial infarction in the six months after presentation with acute coronary syndrome: prospective multinational observational study (GRACE). *British Medical Journal*, 333(7578):1091–, 2006.

[196] Cardiology The Task Force on the Management of Acute Myocardial Infarction of the European Society of, F. Van de Werf, D. Ardissino, A. Betriu, D/ V. Cokkinos, E. Falk, K. A. A. Fox, D. Julian, M. Lengyel, F.J. Neumann, W. Ruzyllo, C. Thygesen, R.S. Underwood, A. Vahanian, F. W. A. Verheugt, and W. Wijns. Management of acute myocardial infarction in patients presenting with st-segment elevation. *European Heart Journal*, 24(1):28–66, 2003.

[197] E. M. Antman, D. T. Anbe, P. W. Armstrong, E. R. Bates, L. A. Green, M. Hand, J. S. Hochman, H. M. Krumholz, F. G. Kushner, G. A. Lamas, C. J. Mullany, J. P. Ornato, D. L. Pearle, M. A. Sloan, Jr. Smith, S. C., E. M. Antman, Jr. Smith, S. C., J. S. Alpert, J. L. Anderson, D. P. Faxon, V. Fuster, R. J. Gibbons, G. Gregoratos, J. L. Halperin, L. F. Hiratzka, S. A. Hunt, A. K. Jacobs, and J. P. Ornato. Acc/aha guidelines for the management of patients with st-elevation myocardial infarction–executive summary: A report of the american college of cardiology/american heart association task force on practice guidelines (writing committee to revise the 1999 guidelines for the management of patients with acute myocardial infarction). *Circulation*, 110(5):588–636, 2004.

[198] A.J. Garber, E.S. Moghissi, E.D. Jr Bransome, N.G. Clark, S. Clement, R.H. Cobin, A.P. Furnary, I.B. Hirsch, P. Levy, R. Roberts, G. Van den Berghe, and V. Zamudio. American college of endocrinology position statement on inpatient diabetes and metabolic control. *American Association of Clinical Endocrinologists*, 10(1):77–82, 2004.

[199] S. Zhou and X. Shen. Spatially adaptive regression splines and accurate knot selection schemes. *Journal of the American Statistical Association*, 96(13):247–259, 2001.

[200] M.H. Hansen and C. Kooperberg. Spline adaptation in extended linear models (with comments and a rejoinder by the authors). *Statistical Science*, 17(1):2–51, 2002.

[201] P. Royston and W. Sauerbrei. Building multivariable regression models with continuous covariates in clinical epidemiology. with an emphasis on fractional polynomials. *Methods of Information in Medicine*, 44(4):561–571, 2004.

[202] Royal College of Physicians. Myocardial infarction national audit project (minap). *http://www.rcplondon.ac.uk/clinicalstandards/organisation/partnership/Pages/MINAP.aspx*, (accessed 20/08/2008).

[203] Healthcare Commission. Information for healthcare providers (minap). *http://www.healthcarecommission.org.uk/serviceproviderinformation/nationalclinicalaudit/furtherinformation.cfm?cit_id=378*, (accessed 20/08/2008).