Statistical Models for Censored Point Processes with Cure Fractions

by

Jennifer Kaye Rogers

Thesis

Submitted for the degree of

Doctor of Philosophy

Department of Statistics

January 2011

Contents

List of Tables

List of Figures

Acknowledgments

I would like to acknowledge the sponsorship I received from the Engineering and Physical Sciences Research Council. I would like to thank my supervisor, Professor Jane Hutton; the advice and encouragement that she has given me throughout has been invaluable. The clinical guidance that has been provided by Professor Tony Marson is also gratefully acknowledged.

I would also like to thank my family and friends for the support they have given me, not just during my PhD, but throughout my entire academic career to date.

Declarations

I hereby declare that this is my own work, except where explicitly stated, and that it has not been submitted for a degree at any other university.

Abstract

We consider the analysis of data from the MRC Multicentre Trial for Early Epilepsy and Single Seizures (MESS), which was undertaken to assess the differences between two policies: immediate, or deferred treatment for patients in early epilepsy. In studies of recurrent events, like epileptic seizures, there is typically lots of information about individuals' seizure patterns over a period of time, which is often not fully utilised in analysis. We develop methodology that allows pre-randomisation seizure counts and post-randomisation times to first seizure, and from first to second seizure, to be jointly modelled, assuming that these outcomes are predicted by (unobserved) seizure rates.

The joint model was found to be superior to standard survival methods. The model had more power to detect statistically significant covariate effects not found by standard survival analysis, however, interesting characteristics within the data, not present in the model were also highlighted. The simple joint model was extended to acknowledge these characteristics.

The results suggested that the identically distributed assumption for the survival times may not be accurate. Instead we adjusted the model to allow for changes in seizure rate both at randomisation and following a seizure post-randomisation.

There is evidence to suggest that there may be a substantial subset of the MESS sample containing individuals who we would not expect to experience seizures post-randomisation. If survival data has a proportion that are immune to the event of interest, a model that ignores this may give misleading results. We considered a cure rate model that allows the separation of individuals who will never experience seizure recurrence and those who are at risk of future seizures. We can then formulate probabilistic models for the 'at risk' individuals. These modifications to the simple joint model have been considered both in isolation and together in a full model.

Abbreviations

n	Number of individuals in study population
$i = 1, \ldots, n$	Subscript for individuals
$j = 1, 2$	Subscript for post-randomisation seizure
X	Pre-randomisation event count
x	Observed value of X
u	Number of days over which the pre-randomisation event count was observed
T_1	Post-randomisation time from randomisation to first seizure
t_1	Observed value of T_1
T_2	Post-randomisation time from randomisation to second seizure
t_2	Observed value of T_2
Y_1	Post-randomisation time to first seizure
y_1	Observed value of Y_1
Y_2	Post-randomisation time from first to second seizure
y_2	Observed value of Y_2
$\boldsymbol{Z}_1$	Matrix containing pre-randomisation covariate information
$\boldsymbol{z}_1$	Observed value of $\boldsymbol{Z}_1$
$\boldsymbol{Z}_2$	Matrix containing post-randomisation covariate information
$\boldsymbol{z}_2$	Observed value of $\boldsymbol{Z}_2$
$\boldsymbol{Z}_3$	Matrix containing covariate information following a first seizure post-randomisation
$\boldsymbol{z}_3$	Observed value of $\boldsymbol{Z}_3$
$\boldsymbol{W}_1$	Matrix containing cure rate covariate information
$\boldsymbol{w}_1$	Observed value of $\boldsymbol{w}_1$
$\boldsymbol{W}_2$	Matrix containing cure rate covariate information
$\boldsymbol{w}_2$	Observed value of $\boldsymbol{w}_2$

$f(\cdot)$	Full Probability Density Function (pdf)
$F(\cdot)$	Full Cumulative Density Function (cdf)
$S(\cdot)$	Full Survivor Function
$h(\cdot)$	Full Hazard Function
$H(\cdot)$	Full Cumulative Hazard Function
$g(\cdot)$	pdf for susceptibles (for cure rate model)
$G(\cdot)$	cdf for susceptibles (for cure rate model)
$R(\cdot)$	Survivor Function for susceptibles (for cure rate model)
$L(\cdot)$	Likelihood Function
$\ell(\cdot)$	Log-likelihood
α	Degree of heterogeneity
λ_i	Pre-randomisation seizure rate for individual i
ψ_i	Post-randomisation seizure rate modifier for individual i
ν_i	Random effect term for individual i
δ_{1i}	Censoring indicator for first seizure
δ_{2i}	Censoring indicator for second seizure

Chapter 1

Introduction

This thesis builds a number of models that can be used in the analysis of data that arrive in the form of event counts and survival times. This work is motivated by individual patient data from a randomised controlled trial that was undertaken to address the question of whether immediate or deferred treatment should be favoured for patients that are in the early stages of epilepsy. A baseline seizure count is recorded for each individual, with the associated number of days over which the seizures were observed and the seizure types experienced. We consider post randomisation times to first and second seizure. Time to first seizure is an internationally agreed outcome in epilepsy trials (ILAE Commission on Antiepileptic Drugs 1998). The primary interest lies in the contrast between immediate and deferred treatment, with possible interactions with age, sex, pre-randomisation seizure type experienced and the outcome of an electroencephalogram (EEG).

In studies of recurrent events, like epileptic seizures, there is frequently lots of information about individuals' seizure patterns over a period of time that

is generally not fully utilised in analysis. Additionally, epilepsy is characterised by recurrent seizures, not a single, isolated event, yet in many treatment studies it is often only time to first seizure that is analysed. We develop methodology that allows the pre-randomisation seizure counts and multiple post-randomisation survival times to be jointly modelled. This method assumes that all these outcomes are predicted by (unobserved) seizure rates. We assume that each patient has an underlying constant seizure rate, which we allow to vary depending on baseline attributes, and suppose that their subsequent post-randomisation seizure rate will be reduced relative to their associated baseline seizure rate. A greater reduction in the seizure rate results in a longer time to seizure post-randomisation, indicating a better therapy.

The class of statistical models that will be developed consider some of the issues arising when analysing data of this type. More specifically, our data exhibits cure rates and there is evidence to suggest that individuals' seizure rates change not only at randomisation to treatment policy, but also following a post-randomisation seizure.

As previously stated, we consider the development of new statistical methodology, which is applied to an epilepsy dataset. The methods developed in this thesis, however, are not restricted to randomised controlled trials relating to epilepsy. We consider novel statistical models that can be applied to any recurrent event data that combine event counts and survival times.

1.1 Overview of Thesis

An overview of the current literature on count models and survival models is provided in Chapter 2. The discussion of survival analysis includes non-parametric, semi-parametric and fully parametric approaches to modelling. Methods used to analyse data that is in the form of recurrent event gap times are also presented. A complication in the analysis of survival data occurs when there is a proportion of the population not susceptible to the event of interest. When presented with data of this type cure rate models are appropriate. These models are also discussed in Chapter 2, as is the existing literature on a model that considers the analysis of a pre-randomisation event count and a single post-randomisation survival time jointly. Finally, this chapter also provides an overview of statistical analysis in the presence of missing data.

Chapter 3 provides an overview of epilepsy and introduces the dataset that motivates the work in the thesis. Standard non-parametric analyses of the data are also provided. We examine the clinical features of the individuals included in the trial and generate Kaplan-Meier curves.

Standard parametric models are considered in Chapter 4. The event counts and survival times are considered separately, with the Negative Binomial Generalised Linear Model being applied to the event counts. The Log-logistic and Lomax survival distributions are considered for the analysis of the post-randomisation times to first seizure and times from first to second seizure.

Chapter 5 develops a model for the joint analysis of pre-randomisation event

counts and post-randomisation times to first seizure and times from first to second seizure. The log-likelihood is derived, with the first and second derivatives, so that numerical methods can be applied to find the maximum likelihood parameter estimates. This joint model is then applied to the dataset and the results are interpreted in terms of pre-randomisation and post-randomisation seizure rates.

Limitations to the simple joint model developed in Chapter 5 are considered in Chapter 6. The simple joint model assumes that the post-randomisation seizure rate remains constant, but a simple modification relaxes this assumption. We also allow for the inclusion of cure rates for each of the survival times, that is, we consider the scenario where individuals who enter the trial may be immune to seizures post-randomisation. These two extensions are considered both in isolation, and together in a single full model.

The performance of the final full model is compared to standard survival methods in Chapter 7. We assess the goodness-of-fit of the models, considering how well the distribution of the survival times is modelled. We use a method developed by Maller and Zhou (1996) for testing the goodness-of-fit of parametric distributions to survival data.

Further extensions to the joint model are considered in Chapter 8. We consider the appropriateness of the zero-truncated, one-inflated Poisson distribution for the count data and the inclusion of further post-randomisation survival times. We discuss further possible analyses of the dataset that consider not only the treatment policy to which an individual was randomised, but also which

antiepileptic drug they subsequently received. Finally, Chapter 8 also considers the analysis of long-term prognosis.

Chapter 9 concludes the thesis, discussing the strengths and weaknesses of the various models that have been developed and the suitability of the underlying assumptions.

Appendix A contains detailed information on the Lomax survival distribution. The log-likelihood, first derivatives and second derivatives are presented, allowing inference on the parameters using numerical methods. Appendix B contains the R functions used to fit both the maximum likelihood simple joint model that was initially developed and the final full joint model that incorporates all the extensions considered. Finally, Appendix C contains a summary of a clinical paper that is currently being written with neurologists at the University of Liverpool. This paper considers the post-randomisation times to first seizure of any type and first tonic-clonic seizure, using the joint model that incorporates cure rates, that we have developed.

Chapter 2

Literature Review

There is an abundance of literature that addresses the analysis of count data and survival times separately. Some preliminary work has considered the analysis of data where a recurrent event process is recorded in the form of a pre-randomisation event count, followed by a post-randomisation time to first event, for each individual.

We shall present commonly used approaches for the analysis of event counts and survival times and then outline methods that have been developed to jointly analyse event counts and survival times in a single model.

We shall also present an introduction into missing data techniques. Likelihood inference generally proceeds to derive the maximum likelihood estimate of a parameter, θ, by maximising the observed likelihood. There might be two problems with this: the observed likelihood could be impossible, or difficult to derive due to the integration involved or, alternatively, it may be hard to maximise. Methods used to handle missing data that focus on likelihood based

techniques can also be used when presented with problems of this type.

2.1 Analysis of Count Data

A model typically used for the analysis of count data is the Poisson Generalised Linear Model (McCullagh and Nelder 1989). A restrictive property of the Poisson distribution is that the mean and variance are equal. When presented with count data that are overdispersed, random effect mixture distributions are often used. The most convenient choice of random effects distribution is the Gamma, which consequently yields the Negative Binomial distribution (Greenwood and Yule 1920).

Alternatives to the Gamma distribution as the mixing distribution are considered by Hougaard et al. (1997), with special consideration being given to the analysis of the frequency of epileptic seizures. A larger family of mixture distributions are considered, including the Inverse Gaussian mixture distribution. This paper, however, does not consider the inclusion of covariates.

2.2 Analysis of Survival Data

Survival analysis is the analysis of data which is in the form of times from a well defined start point, up to a particular event of interest. A comprehensive and thorough discussion of survival techniques is given by Collett (2003). The actual survival time, t, for an individual, is a realisation of the random variable T, which can take any non-negative value. This random variable has associated with it a probability distribution, with an underlying probability

density function $f(t)$.

There are generally two functions that are of central interest in survival analysis, namely the survivor function and the hazard function. The survivor function is defined to be the probability that an individuals' survival time is greater than or equal to some value t, expressed as

$$S(t) = \mathbb{P}(T \geq t) = 1 - F(t), \tag{2.1}$$

where $F(t)$ is the cumulative distribution function of the random variable T. The hazard function can be thought of as the instantaneous death rate. To derive its form, first consider the probability that an individuals' survival time lies in the interval t and $t + \xi t$, conditional on survival to time t, for some $\xi > 0$. Dividing this probability by ξt then gives a probability per unit time. The hazard function is simply the limiting value of this quantity, as ξt tends to zero:

$$h(t) = \lim_{\xi t \to 0} \left\{ \frac{\mathbb{P}(t \leq T \leq t + \xi t \mid T \geq t)}{\xi t} \right\}. \tag{2.2}$$

A further function that may be of interest when considering survival data is the cumulative hazard function, simply defined as

$$H(t) = \int_0^t h(u)du. \tag{2.3}$$

2.2.1 Censoring

An important issue in survival analysis is that of censoring. Censoring occurs when an individuals' actual survival time cannot be measured, but we have instead some measurable censored time associated with them. There are generally three types of censoring: (i) right censoring, occurring when the censored survival time is less than the actual, unknown survival time, (ii) left censoring, occurring when the observed, censored survival time is greater than the actual, unknown survival time, and (iii) interval censoring, which is evident if the actual survival time is only known up to some interval. We consider the analysis of data that is subject to right censoring.

The standard survival techniques outlined in this section are only applicable if the censoring is non-informative. This essentially means that the censoring is not related to any factors associated with the actual survival time.

2.2.2 Non-Parametric Procedures

Often the first step in survival analysis would be to produce either graphical or numerical summaries of the data based on non-parametric, or distribution-free estimates of the survivor and hazard functions. These estimates are significant in their own right, but are generally used as a pre-cursor to more detailed analysis. The most widely used non-parametric estimate of the survivor function, that allows for censoring, is the Kaplan-Meier estimate.

Given n individuals, let the number of individuals alive just before time $t_{(j)}$ be denoted by n_j, with d_j denoting the number of deaths at this time. It

follows that the estimated probability of survival through the interval $t_{(j)} - \epsilon$ to $t_{(j)}$, for small ϵ, is $(n_j - d_j)/n_j$. The probability of survival between $t_{(j)}$ and $t_{(j+1)} - \epsilon$, the time immediately before the next death, is simply unity as there are no deaths in this interval. The joint probability of surviving in the intervals $t_{(j)} - \epsilon$ to $t_{(j)}$ and $t_{(j)}$ to $t_{(j+1)} - \epsilon$ is therefore estimated by $(n_j - d_j)/n_j$. It follows that in the limit, as ϵ tends to zero, $(n_j - d_j)/n_j$ becomes an estimate of surviving from $t_{(j)}$ to $t_{(j+1)}$.

Now suppose that there are r death times, among the n individuals, that are ordered $t_{(1)} < t_{(2)} < \ldots < t_{(r)}$. The estimated survivor function at time t, for $t_{(k)} \leq t < t_{(k+1)}$ is simply the product of the probability of surviving in the interval $t_{(k)}$ to $t_{(k+1)}$, and all preceding intervals. This leads to the Kaplan-Meier estimate of the survivor function, given by

$$\hat{S}(t) = \prod_{j=1}^{k} \Big(\frac{n_j - d_j}{n_j} \Big), \qquad k = 1, 2, \ldots r. \tag{2.4}$$

2.2.3 Modelling Survival Data

In many situations, individuals' survival times will be accompanied by a number of explanatory variables, or covariates. Interest is most commonly concerned with how one or more of these covariates may affect an individual's survival time. When these situations arise, simple non-parametric approaches are not sufficient, and more sophisticated modelling is necessary. Many of the principles and procedures of linear modelling lend themselves easily to the modelling of survival data.

Cox Proportional-Hazards Model

Probably the most widely adopted and well known of the survival models is the Cox proportional-hazards model (Cox 1972). This is a semi-parametric model, as it assumes that the hazard function for an individual is proportional to some baseline hazard, but does not assume a probability distribution for the survival times. We consider data that violates the proportional hazards assumption, rendering the Cox model unsuitable. A brief discussion is included here however, for completeness.

Consider a vector of m explanatory variables, $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)$, assumed to have been collected at time zero, and let $h_0(t)$ be the baseline hazard function, with $\boldsymbol{x} = (0, 0, \ldots, 0)$. Under the Cox proportional hazards model the hazard function for the ith individual can then be written as

$$h_i(t) = e^{\eta_i} h_0(t),$$

where $\eta_i = \boldsymbol{\beta}' \boldsymbol{x}_i$ is the linear component of the model, also known as the risk score or prognostic index, for the ith individual. Fitting the proportional hazards model to a given data set involves estimating the coefficients of the explanatory variables, $\boldsymbol{\beta}$, and the baseline hazard, $h_0(t)$.

The Cox proportional hazards model is in many cases advantageous due to its widespread applicability, owing to its lack of restriction to a specific functional form. If, however, an assumed probability distribution is valid, inferences from a fully parametric analysis will yield more precise results than the Cox semi-parametric approach.

Parametric Models

Superiority over the Cox proportional hazards model can be obtained through parametric analysis, when the probability distribution assumed is accurate. Preliminary study of the validity of a range of probability distributions can be carried out using Kaplan-Meier estimates of the survival curves. Transforming the survivor function to produce a plot that should give a straight line if the assumed model is appropriate is one way of assessing the suitability of parametric models. Common survival distributions include the Weibull, Exponential, Lognormal and Log-logistic distributions.

Assume that a suitable parametric model has been adopted, and that the density function of the random variable associated with the survival times is $f(t)$. If there are no censored observations, then the likelihood for n observations is simply $\prod_{i=1}^{n} f(t_i)$. Now suppose that right censoring is present and consider a censoring indicator, δ_i, which takes the value zero if the ith survival time, t_i, $i = 1, 2, \ldots, n$ is censored, and unity if the survival time is observed. The likelihood is now given by

$$\prod_{i=1}^{n} \{f(t_i)\}^{\delta_i} \{S(t_i)\}^{1-\delta_i}.$$

Parametric models can be categorised as proportional-hazards, proportional-odds, accelerated failure time, or a combination of the three. Exploratory analysis suggests that our data is best modelled through an accelerated failure time model (Wei 1992), so we focus our attention on the family of distributions satisfying this property. The accelerated failure time model assumes that for individual i, the covariates act multiplicatively on the time scale, and so affect

the rate at which individuals progress along time. If a distribution displays the accelerated failure time property, the survivor function for individual i can be expressed as

$$S_i(t) = S_0(t/e^{\eta_i}),$$

where $S_0(t)$ is the baseline survivor function. Common survival distributions that are accelerated failure time are the Weibull, Exponential, Log-logistic, Lognormal, Gamma, and Inverse Gaussian distributions.

Suitability of the accelerated failure time family of distributions is best assessed using a Q-Q plot of the Kaplan-Meier estimate of the survivor functions. If we consider the simple case where there are two groups, then the accelerated failure time model says that $S(t)$ in one group is equal to $S(\phi t)$ in another. A plot of the percentiles of the Kaplan-Meier estimated survivor curves should produce a straight line, with slope ϕ, that passes through the origin if the accelerated failure time model is appropriate.

2.2.4 Recurrent Event Gap Times

Often, interest may lie in studying processes that generate events repeatedly through time. Such processes are known as recurrent event processes and the data they provide are referred to as recurrent event data (Cook and Lawless 2007). Typically, in medical settings there is recurrent event data available on a large number of individuals, exhibiting a relatively small number of events. Examples of such settings may include asthma attacks in respiratory studies or epileptic seizures in neurology studies. Methods for handling recurrent event data tend to be based on counts and rate functions, or the analysis of gap

times, which we consider here.

We begin by considering the methodology currently available for renewal processes. Renewal processes are ones in which the gap times between successive events, $Y_j = T_j - T_{j-1}$, $(j = 1, 2, \ldots)$, are independent and identically distributed. It is usually assumed that the time origin, $t = 0$, corresponds to an event, but this may be relaxed to allow Y_1 to have a different distribution from $(Y_2, Y_3, \ldots)$. Cook and Lawless (2007) consider $X(s,t)$, the number of events over $(s,t]$. The distribution for counts, $X(s,t)$, in renewal processes is intractable except for the renewal process in which the Y_j are exponential random variables. That is, when the process is a homogeneous Poisson process.

Methods Based on Gap Times

Analyses based on gap times are often useful when events are relatively infrequent, when there is some type of renewal after an event, or when prediction of the time to the next event is of interest.

We assume that individual i is observed over the time interval $[0, \tau_i]$ and that m_i events are observed at times $0 < t_{i,1} < \ldots < t_{i,m_i} < \tau_i$, with $y_{i,j} = t_{i,j} - t_{i,j-1}$ and $y_{i,m_i+1} = \tau_i - t_{i,m_i}$, where $t_{i,0} = 0$. The $y_{i,j}$ are the observed gap times for individual i, with the final time being possibly censored. The likelihood for n individuals can be written in terms of the density and survivor functions:

$$\prod_{i=1}^{n} \prod_{j=1}^{m_i} f(y_{i,j}) S(y_{i,m_i+1}). \tag{2.5}$$

If $y_{i,m_i+1} = 0$, that is, if observation for individual i terminates after the m_ith event, the term $S(y_{i,m_i+1})$ disappears.

2.2.5 Survival Analysis with Cure Rates

A proper survival distribution should have total mass 1, with the resulting Kaplan-Meier curve having its asymptote at zero. That is, in standard survival analysis we assume that every individual in the sample is susceptible to the event of interest. In some situations however, there may be a number of individuals who would never experience the event of interest, regardless of the time for which they were followed. Maller and Zhou (1996) encourage us to think of these individuals as cured, or immune to the event of interest. If survival data does indeed have a proportion that are immune to the event of interest, considering a proper survival model that ignores this may give misleading results. An improper survival distribution allows, formally, infinite survival times. Cure rate models allow the quantity $p = F(\infty) = \lim_{t\to\infty} F(t)$ (where $F(t)$ is the cumulative distribution function of the survival times) to be strictly less than 1, corresponding to the presence of immunes in the population.

Suppose t_i^* is the true survival time for individual i, which is only observed if it does not exceed their associated censoring time, c_i, otherwise we observe c_i. Consequently, the actual, observed survival time for individual i can be expressed as $t_i = \min(t_i^*, c_i)$. To formulate the probabilistic mechanism that allows the true survival times t_i^* to be infinite first assume that individual i has an associated Bernoulli random variable, B_i, taking the value 1 if individ-

ual i is susceptible to the event of interest, and with $B_i = 0$ corresponding to an immune individual. Additionally, $p < 1$ represents the proportion of susceptibles in the population, so that

$$B_i = \begin{cases} 1 & \text{with probability} \quad p, \\ 0 & \text{with probability} \quad 1-p. \end{cases}$$

In reality we do not know whether an individual is immune or not, so B_i is not observed. Susceptible individuals are assumed to have a proper cumulative distribution function, $G(t)$, with $G(\infty) = 1$. Formally, individuals with $B_i = 0$ have $t_i^* = \infty$, hence, for all $t \geq 0$

$$\begin{aligned} \mathbb{P}\{t_i^* \leq t \mid B_i = 1\} &= G(t), \\ \mathbb{P}\{t_i^* \leq t \mid B_i = 0\} &= 0, \end{aligned}$$

These probabilities imply that, for all $t \geq 0$, the cumulative distribution function of the true survival times t_i^* is

$$\begin{aligned} F(t) &= \mathbb{P}\{t_i^* \leq t\} \\ &= \mathbb{P}\{t_i^* \leq t \mid B_i = 1\}\mathbb{P}\{B_i = 1\} + \mathbb{P}\{t_i^* \leq t \mid B_i = 0\}\mathbb{P}\{B_i = 0\} \\ &= pG(t) + 0 \\ &= pG(t). \end{aligned}$$

Consequently, for all $t \geq 0$

$$G(t) = \frac{F(t)}{p} = \frac{F(t)}{F(\infty)}. \tag{2.6}$$

To ensure that p remains within the interval $[0, 1]$ the following reparameterisation is often considered:

$$\kappa = \ln\left(\frac{p}{1-p}\right).$$

We can also allow the cure rate to depend on individuals' covariates:

$$p_i = \frac{\exp(\boldsymbol{\kappa}'\boldsymbol{w}_i)}{1+\exp(\boldsymbol{\kappa}'\boldsymbol{w}_i)}.$$

The explanatory variables are entered into the covariate $\boldsymbol{w}_i$ and the parameter $\boldsymbol{\kappa}$ is the corresponding vector of regression coefficients.

2.3 Joint Modelling of Event Counts and a Single Survival Time

Consider the case where for each individual we have a pre-randomisation event count and a post-randomisation survival time. Most standard survival analysis may treat the pre-randomisation event count information as a covariate (Verity et al. 1995); this strategy however ignores any existing variation between individuals. The pre-randomisation event count is an outcome in its own right and Cowling et al. (2006) proposed a technique that jointly analysed an individual's pre-randomisation seizure count, and a single post-randomisation failure time under a Poisson process framework, in a single model.

We assume that each individual experiences events according to a Poisson process with rate $\lambda_i\nu_i$, where the parameter λ_i relates to the baseline covari-

ates, with additional heterogeneity in the population being modelled through ν_i, assumed to follow a Gamma distribution with expectation 1 and variance $1/\alpha$. Smaller values of α are indicative of higher levels of heterogeneity. Consequently, the pre-randomisation event count over a period u_i, for individual i, X_i, follows a Poisson distribution with mean $\lambda_i u_i \nu_i$. If the underlying point process is modelled through a Poisson process, then gap times between events will be Exponential with the same rate. Consequently, post-randomisation, the time to first seizure will be Exponential, but the rate is updated to allow for a treatment effect. It is assumed that the treatment effect acts multiplicatively on the rate, so that the time to first seizure for individual i post-randomisation, Y_i, is Exponential with event rate $\lambda_i \psi_i \nu_i$, where ψ_i depends on the treatment in some way. In summary:

$$\begin{aligned} X_i \mid \nu_i &\sim \text{Poisson}(\lambda_i u_i \nu_i), \\ Y_i \mid \nu_i &\sim \text{Exponential}(\lambda_i \psi_i \nu_i), \\ \nu_i &\sim \text{Gamma}(\alpha, \alpha). \end{aligned}$$

Therefore the joint model is specified by the following equations:

$$\begin{aligned} f_{X|\nu}(x_i \mid \nu_i; \lambda_i, u_i) &= \frac{(\lambda_i u_i \nu_i)^{x_i} \exp(-\lambda_i u_i \nu_i)}{x_i!}, \\ f_{Y|\nu}(y_i \mid \nu_i; \lambda_i, \psi_i) &= \lambda_i \psi_i \nu_i \exp(-\lambda_i \psi_i \nu_i y_i), \\ f_{\nu}(\nu_i; \alpha) &= \frac{\alpha^{\alpha} \nu_i^{\alpha-1} \exp(-\alpha \nu_i)}{\Gamma(\alpha)}, \end{aligned}$$

where $\lambda_i = \exp(\boldsymbol{\beta}_1' \boldsymbol{z}_{1i})$ and $\psi_i = \exp(\boldsymbol{\beta}_2' \boldsymbol{z}_{2i})$. The data enter the model through $\boldsymbol{z}_{1i}, \boldsymbol{z}_{2i}, u_i, x_i, y_i$ and δ_i, the censoring indicator taking the value zero if the time to first seizure post-randomisation is censored and unity if the survival time

is observed. The baseline explanatory variables are entered into the covariate $\boldsymbol{z}_{1i}$, and the treatment covariate $\boldsymbol{z}_{2i}$ contains a treatment indicator, and may also contain other explanatory variables and interaction terms. The parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are the corresponding vectors of regression coefficients.

If the random effect term is integrated out of the joint density of X_i and ν_i, then the resulting unconditional density, $f_X(x_i; \lambda_i, u_i, \alpha)$, is simply the Negative Binomial. The unconditional distribution of Y_i, obtained when the random effect term is integrated out of the joint density of Y_i and ν_i, is the Lomax distribution (Johnson and Kotz 1970), with density

$$\begin{aligned} f_Y(y_i; \lambda_i, \psi_i, \alpha) &= \int_0^\infty f_{Y|\nu}(y_i \mid \nu_i; \lambda_i, \psi_i) g_\nu(\nu_i; \alpha) d\nu_i \\ &= \frac{\lambda_i \psi_i}{(1 + \lambda_i \psi_i y_i / \alpha)^{\alpha+1}}. \end{aligned} \tag{2.7}$$

2.4 Missing Data

Standard statistical methods are well developed for the analysis of complete rectangular data sets, where the the rows of a data matrix represent subjects. Missing data arises when some of the observations in the data matrix are not observed. One approach to handling missing data is to simply omit, from the data matrix, those rows that contain missing values. This technique, commonly referred to as 'complete-case analysis', although simple to carry out, is generally inappropriate. The disadvantages of adopting this approach stem from the potential loss of precision and induction of bias in discarding the incomplete cases.

A comprehensive discussion on statistical analysis in the presence of missing data is given by Little and Rubin (2002). They identify the problems associated with the presence of missing data in standard statistical analysis, consider how missingness can arise, present different missing data patterns and discuss the various methods for handling datasets that exhibit missing data.

First we define $Y = (y_{ij})$ to be the complete data matrix, and introduce the missing data indicator matrix, $M = (m_{ij})$, which is unity if y_{ij} is missing, and zero if y_{ij} is observed. Now the complete data Y can be partitioned into $Y = (Y_{obs}, Y_{mis})$ where

$$\begin{aligned} Y_{obs} &= Y[m_{ij} = 0], \\ Y_{mis} &= Y[m_{ij} = 1]. \end{aligned}$$

The full probability density function for the observed data, Y_{obs}, missing data, Y_{mis}, and the missing data indicator, M, is $f(Y_{obs}, Y_{mis}, M; \theta, \varphi)$. This full probability density function can be factorised as:

$$f(Y_{obs}, Y_{mis}, M; \theta, \varphi) = f(Y_{obs}, Y_{mis}; \theta) f(M \mid Y_{obs}, Y_{mis}; \varphi), \quad (\theta, \varphi) \in \Omega_{\theta,\varphi}, \tag{2.8}$$

where $\Omega_{\theta,\varphi}$ is the parameter space of (θ, φ). The actual observed data consists of the variables (Y_{obs}, M), hence the probability density function for the observed data can be obtained by integrating Y_{mis} out of equation 2.8:

$$f(Y_{obs}, M; \theta, \varphi) = \int f(Y_{obs}, Y_{mis}; \theta) f(M \mid Y_{obs}, Y_{mis}; \varphi) dY_{mis}. \tag{2.9}$$

Treating the missing data indicator as a random variable, the missing data mechanism is characterised by $f(M \mid Y_{obs}, Y_{mis}; \varphi)$, the conditional distribution of M given (Y_{obs}, Y_{mis}), where φ denotes unknown parameters. The missing data mechanisms, more specifically the question of whether the values that are missing are related to the underlying variable values, are an extremely important issue in the area of missing data. Missing data mechanisms generally determine how the missing data analysis should proceed. There are three distinct types of missing data mechanism: missing completely at random, missing at random and not missing at random.

The missing data are said to be missing completely at random (MCAR) if missingness does not depend on the values of Y, missing or observed. That is,

$$f(M \mid Y_{obs}, Y_{mis}; \varphi) = f(M; \varphi), \qquad \text{for all } Y, \varphi. \tag{2.10}$$

An assumption that is less restrictive than MCAR is missing at random (MAR). If missingness depends on those values of Y that are observed, and not on the components that are missing, the data are said to be MAR. That is,

$$f(M \mid Y_{obs}, Y_{mis}; \varphi) = f(M \mid Y_{obs}; \varphi), \qquad \text{for all } Y_{obs}, \varphi. \tag{2.11}$$

Finally if data is neither MAR or MCAR, that is the distribution of M depends on the missing values in the data matrix Y, we call the data not missing at random (NMAR).

2.4.1 Missing Data Methods

Methods proposed for handling missing data can be grouped into the following categories:

Complete-Case Analysis: Analysis is carried out only on those cases for which there is no missing data.

Weighting Procedures: A modification to complete-case analysis that differentially assigns weights to the complete cases to adjust for bias. This strategy is found most commonly in sample surveys to handle nonrespondents.

Imputation-Based Procedures: Essentially missing values are filled in, with the subsequent completed data analysed using standard statistical methods. For the resulting inferences to be valid, modifications to standard analyses are necessary to allow for the differing status of the real and imputed values.

Model-Based Procedures: A broad class of procedures can be generated by defining a model for the observed data and basing subsequent inferences on the likelihood or posterior distribution under that model.

We have already discussed the relative simplicity of discarding the incompletely recorded units and the associated disadvantages. Complete-case analysis may be justified, however, when the amount of missing data is small, so that the loss of precision and bias is minimal. Weighted complete-case estimators are often simple to compute, but the corresponding standard errors are less straightforward to derive.

Imputation techniques are an attractive and flexible method for handling missing data. However, one should be wary of the potential pitfalls. Users of imputation methods should be considerate of the fact that the data are not complete, and imputed values should not be regarded in the same way as real values. Imputations are means or draws from a predictive distribution of the missing values and so a method of creating a predictive distribution is required. Modelling methods include, but are not limited to, mean modelling, regression modelling or last observation carried forward. Single imputation methods replace missing values with a single imputed value, but multiple imputation replaces missing values with a vector of $D \geq 2$ imputed values. The D values are ordered in such a way that D complete datasets are generated from the vectors of imputations. Standard complete-data methods are used to analyse each dataset, with the subsequent D complete-data inferences combined to form one inference which reflects the uncertainty due to nonresponse.

Likelihood based methods are more attractive than complete-case analysis or imputation techniques (Schafer and Graham 2002). These methods, however rely on a few crucial assumptions. The sample may have to be larger than usual as missing data reduces the sample size and the sample needs to remain large enough for the maximum likelihood estimates to be approximately unbiased and normally distributed. Additionally, the likelihood function comes from an assumed parametric distribution for the complete data and likelihood methods may not necessarily be robust to deviations from the model assumptions.

2.5 Discussion

This chapter has discussed the standard statistical models that currently exist for the analysis of event counts and survival times. We have also outlined models that can be used for the analysis of survival data when there are cure rates present in the population and models that can be used in the analysis of recurrent events.

Cook and Lawless (2007) discuss the use of models that are appropriate in the specification and testing of treatment effects in recurrent events (§ 8.4.1). They suggest that methods based on rates and mean functions, rather than gap times, offer the most straightforward specification of treatment effects for recurrent events and outline the mixed Poisson model as a natural framework for analyses. Although methods based on rates and mean functions may be the most advantageous for the specification and testing of treatment effects in recurrent events, Cook and Lawless (2007) recognise that there may be situations where analyses based on gap times are more natural. Additionally, if datasets exhibit cure rates, models that focus on rates and mean functions are not the most conclusive. Such models would assess post-randomisation event rates that are zero, but to consider this, focus would need to be on a suitable length of observation time, or gap time, with no events recorded.

An additional consideration in the analysis of recurrent event data, that is considered by Cook and Lawless (2007) in § 8.4.3, is the use of baseline count data. Interest is focussed on the use of mixed Poisson processes for the analysis of recurrent event data that incorporates a period of observation

in which subjects are monitored prior to randomisation to treatment. Interest in this instance would be focussed on the change in event rate pre and post-randomisation. There is no discussion, however, of the analysis of recurrent event data that arrive in the form of a pre-randomisation baseline count and post-randomisation gap times. Recall that time to first seizure is an internationally agreed outcome in epilepsy trials (ILAE Commission on Antiepileptic Drugs 1998), so methods pertaining to the analysis of gap times, that fully make use of the pre-randomisation event counts, seem the most sensible.

A model that considers the analysis of pre-randomisation event counts and a single post-randomisation survival time, proposed by Cowling et al. (2006), has been discussed in this chapter. It is suggested that these methods may form the basis of a model that allows pre-randomisation event counts and multiple post-randomisation survival times to be jointly analysed in a single model.

Finally, this chapter has briefly presented missing data methods that are commonly adopted when presented with incomplete data. The different missing data mechanisms and methods for handling missing data have been discussed.

Chapter 3

Introduction to the Epilepsy Data

We shall now present an overview of epilepsy and introduce the dataset that will be analysed using both standard statistical methods and the new joint models that we shall develop in later chapters. We shall conduct exploratory analysis on the data and produce a number of Kaplan-Meier plots that will form an initial examination of the effect of the explanatory variables on times to first and second seizure.

3.1 Overview

Epilepsy is formally defined as the occurrence of recurrent and unprovoked seizures (Warrell et al. 2003). An epileptic seizure itself is caused by excessive neuronal activity which will manifest itself in an alteration of consciousness, or motor, sensory, autonomic, or psychic events. The International League Against Epilepsy (ILAE) classification scheme divides seizures into partial,

generalised or unclassified seizures (Berg et al. 2010). Partial seizures are localised and involve only part of the brain; they can be either simple (consciousness not impaired) or complex (consciousness impaired).

Partial seizures include motor, sensory, occipital, frontal lobe and temporal lobe seizures. Motor seizures focally affect parts of the body that are correlated with their area of representation in the motor cortex. Following motor seizures there can be (on rare occasions) paralysis of the affected part. Sensory seizures comprise paraesthesias or numbness in focal areas of the body. Symptoms of occipital seizures are visual, altering the size, shape or depth of objects. Jerking or forced closure of the eyelids may also occur. Frontal lobe seizures are composed of pelvic thrusting, rocking of the body and head movements that are nocturnal. Vocalisation may also occur. Sufferers of temporal lobe seizures may sense a loss of personal or environmental reality and can experience intense familiarity or unfamiliarity, psychic symptoms and sensations pertaining to smell, taste and vertigo.

Partial epilepsy can sometimes occur with secondary generalisation. Generalised seizures involve all of the brain and are categorised as tonic-clonic (grand mal epilepsy), absence (petit mal epilepsy), myoclonic or atonic. Tonic-clonic seizures consist of a tonic phase where an individual becomes stiff as muscles contract, and a clonic phase where limbs jerk caused by the muscles contracting and relaxing in quick succession. During absence seizures, activity ceases and individuals simply stare blankly and are unresponsive, without loss of posture. Patients are totally unaware of their absence seizures. Symptoms of myoclonic seizures are brief, shock-like contractions of muscles, which may be

generalised or focal. Atonic seizures comprise a sudden loss of muscle tone; episodes are brief and recovery rapid.

Epilepsy can be genetically determined, caused by migration disorders or trauma. Epilepsy syndromes, such as juvenile myoclonic epilepsy, usually occurring between the ages of 12 and 18, are inherited. A certain proportion of children who suffer epilepsy syndromes will go on to develop epilepsy at a later age. Epilepsy can also be linked with the presence of tumours, incidence of strokes, infection, cerebral degeneration, multiple sclerosis or metabolic disorders. Epileptic seizures can be encouraged by lack of adequate sleep, alcohol abuse and the ingestion of certain drugs; seizures may also be confined to the menstrual period. Epilepsy attacks can additionally be triggered by particular stimuli, such as noise and movement.

The question of whether to start patients on a course of anticonvulsants after a single epileptic seizure remains an area of uncertainty. Several studies have shown that intervention after a single seizure reduces the risk of short-term recurrence, but does not affect the long-term remission rates in individuals with single or infrequent seizures (Marson et al. 2005; Chandra 1992). Warrell et al. (2003) state that seizure recurrence after a single untreated seizure is around 80%, but Berg and Shinnar (1991) put seizure recurrence at 50%. It is suspected that seizure recurrence may be different for different seizure types, which may account for the difference in these two values. It is also thought that the risk of future seizures increases with the number of previous seizures, with around 30% of epilepsy sufferers never achieving long-term remission (Cockerell et al. 1995).

Antiepileptic drugs (AEDs) often come with extremely unpleasant side effects, which include weight loss or weight gain, altered mood, drowsiness, hair loss, polycystic ovarian disease, visual field defects and teratogenicity. For most epilepsy sufferers, the benefits of AEDs will far outweigh the associated risks. For those individuals, however, who have had only a single seizure, or have infrequent and mild epileptic seizures the question of whether to withhold treatment until absolutely necessary becomes clinically important.

Typically, the questions asked by individuals in early epilepsy might be: 'If I have a seizure, am I likely to have another one, and if so, when?' or 'If I have a second seizure am I likely to have more, and with increased frequency?'. We aim to develop methodology that attempts to provide answers to these questions and ascertains how these questions might be dependent on important covariates such as seizure type or EEG outcome. We have already presented some of the existing literature for the analysis of recurrent event data. This work would encourage us to think of epileptic seizures in terms of counts, possibly comparing the difference in counts pre and post-randomisation. Clearly, when presented with questions of this type it is more natural, and clinically relevant, to analyse recurrent event data in terms of the survival times. Additionally, recall that time to first seizure is an internationally agreed outcome (ILAE Commission on Antiepileptic Drugs 1998) in epilepsy trials. Standard methodology exists for the analysis of recurrent survival data, but, as we have previously discussed, there is no literature that provides a parametric modelling strategy for pre-randomisation event counts and multiple post-randomisation survival times.

3.1.1 The MESS Trial

The MRC Multicentre Trial for Early Epilepsy and Single Seizures (MESS) was undertaken to address the question of immediate versus deferred treatment with AEDs in those patients that have had one, or very few seizures. Interest lay in both the effects on short-term recurrence, and long term prognosis.

The MESS trial randomised 1443 patients worldwide, across 83 centres. The eligibility criteria were: being aged at least 1 month, having a suitably documented history of at least one clinically definite, spontaneous and unprovoked epileptic seizure (excluding febrile convulsions), and there being genuine uncertainty in both clinician and patient as to whether treatment with AEDs should commence. Patients were excluded from the trial if they had previously received treatment with AEDs. Patients were randomised to either immediate or deferred treatment using the minimisation method, balancing across two factors: centre or region, and number of seizures prior to randomisation (defined as either single or multiple).

The trial was a pragmatic trial, meaning that those individuals assigned to immediate treatment were administered the most appropriate antiepileptic drug, determined by the clinician. Those patients randomised to deferred treatment received no drugs until both clinician and patient agreed that treatment was absolutely necessary, in which case the clinician decided the optimum anticonvulsant, dose and duration of treatment.

Baseline covariates collected for each individual included age, sex, demographic information, any history of existing neurological disorders and information on past seizures including the number of seizures experienced and seizure type. An electroencephalogram (EEG) was requested for each individual, CT and MRI scans were performed if clinically indicated. Follow-up occurred at 3 months, 6 months, 1 year, and then at yearly intervals (more regularly if clinically necessary). At each follow-up information was collected about any seizures the patient had experienced since the previous follow-up, along with information about AEDs currently being taken, including dose and any side-effects experienced. In the event of death, the date and cause of death were recorded. The outcomes measured included times to first, second and fifth seizure, as well as times to one and two year remission.

Recruitment for the trial ran from 1st January 1993 to 31st December 2000. Final follow-up was attempted between 31st December 2001 and 30th June 2002. Statistical analyses were by intention to treat, interest lay in the treatment policy to which an individual was assigned rather than whether an individual was receiving treatment or not when they experienced future seizures. Detailed methods and primary analyses can be found in Marson et al. (2005) and Kim et al. (2006). We shall present a brief summary of each of the trials here.

To analyse the times to each outcome event, Marson et al. (2005) used the log-rank test (Peto and Peto 1972), or the Cox proportional-hazards model when adjusting for the number of seizures pre-randomisation. The number of seizures pre-randomisation was taken to either be single or multiple. The key

demographic and clinical features of participants are given in Marson et al. (2005), including the number of individuals experiencing particular types of seizures pre-randomisation, stratified by the treatment policy to which they were subsequently assigned. Also included is the number of individuals presenting an abnormal EEG and whether the abnormalities were non-specific, generalised or focal. This analysis of the MESS data does not consider pre-randomisation seizure types, or EEG outcome, in the analysis of the post-randomisation times to events of interest. For times to first seizure and times to second seizure the differences between the two treatment groups were found to be statistically significantly different ($\chi^2 = 21.4$, $p < 0.0001$ and $\chi^2 = 9.2$, $p = 0.0025$ respectively).

Kim et al. (2006) developed a prognostic model to categorise individuals as low risk, medium risk or high risk of seizure recurrence, using Cox regression, stratified by treatment policy. For the prognostic model, the number of groups likely to maintain reasonable separation was found to be three. In this analysis of the MESS data, EEG outcome and pre-randomisation seizure type were considered. The seizure types considered were 'tonic-clonic seizures only' and 'simple or complex partial seizures only'. An abnormal EEG was defined as a focal or general, excluding non-specific abnormalities. Backwards stepwise regression was used and found three potentially important prognostic factors: existence of a neurological disorder, total number of seizures pre-randomisation over all seizure types and an abnormal EEG. Seizure type pre-randomisation was not found to be significant for the prognostic model.

The statistically significant prognostic factors were used to identify individuals

as at a low risk, medium risk or high risk of seizure recurrence as follows:

- Low risk - one seizure only pre-randomisation, no neurological disorder and a normal EEG.
- Medium risk - two or three seizures pre-randomisation, no neurological disorder and a normal EEG; or one seizure pre-randomisation and either a neurological disorder, or an abnormal EEG.
- High risk - one seizure pre-randomisation, a neurological disorder and an abnormal EEG; or two or three seizures pre-randomisation and either a neurological disorder, or an abnormal EEG, or both; or any individual with 4 or more seizures pre-randomisation.

For low-risk individuals, there was no significant difference found between the treatment policies. For individuals in the medium and high risk groups, immediate treatment was favoured when considering seizure recurrence.

We carry out exploratory analysis on 1425 individuals; 18 were removed due to missing information, assumed missing completely at random. It is important to note that 812 of the 1425 individuals included in the exploratory analysis presented only a single seizure pre-randomisation. The period of time from this single seizure to randomisation, for these individuals, ranged from the same day to 464 days, with the median number of days being 27. For the majority of those individuals with only one seizure pre-randomisation, their associated period of time from first seizure to randomisation may be inaccurately small, possibly representing how long it took for them to arrange an appointment with their GP. This results in imprecise estimates of their associated underlying seizure rates and an ensuing overestimation of the seizure rate reductions.

Following discussions with clinicians, we subsequently made adjustments to the values of u_i, the number of days from the first pre-randomisation seizure to randomisation, in the dataset so that $u_i \geq 182$. That is, any value of $u_i < 182$, we replaced with 182.

As a sensitivity analysis to the choice of 182 days as the minimum period pre-randomisation, the data were re-analysed with $u_i \geq 91$ and $u_i \geq 365$. The resulting regression coefficients from these adjustments can be found in Rogers et al. (2009), along with their associated $\widehat{\lambda}_i$ and $\widehat{\psi}_i$. The magnitudes of differences observed in seizure rates between the groups were maintained through each adjustment.

The log-likelihoods associated with each model would suggest that having a minimum pre-randomisation period of 365 days is optimal. Further inspection of the log-likelihoods however, suggested that the likelihood function is very flat, hence the decision was made to take the clinicians' suggestion of a minimum pre-randomisation period of 182 days. All future analysis of the MESS data will be carried out with this adjustment.

3.2 Distribution of Variables

Of the 1425 individuals included in the exploratory analysis, 691 (48.91%) experienced at least one seizure following randomisation, with a subsequent 480 (69.46%) of these experiencing a second. Later findings conclude that the variables age and sex are not statistically significant in determining pre-randomisation seizure rates or post-randomisation seizure rate reductions, so

we exclude these variables completely here. EEG outcome is simply defined as being abnormal or normal; further investigation concluded that there was no statistically significant differences between the effects of the various types of EEG abnormality on post-randomisation survival times ($\chi^2 = 2$, p= 0.569), only the presence of a normal EEG or abnormal EEG of any type was found to be statistically significant. The pre-randomisation seizure types are categorised as follows:

Tonic-Clonic: Those individuals presenting with tonic-clonic seizures only pre-randomisation.

2° Tonic-Clonic: Those individuals presenting with partial seizures with secondary tonic-clonic seizures pre-randomisation.

Generalised: Those individuals presenting with any combination of generalised seizures pre-randomisation (this group could include those having a combination of tonic-clonic and other generalised seizures).

Partial: Those presenting with partial seizures only pre-randomisation (either simple or complex).

Other: Those presenting with seizures pre-randomisation that do not fit into any of the above categories.

Immediate treatment reduces the risk of a seizure post-randomisation (one year relative risk 0.734 [95% C.I. (0.63,0.85)], 8 year relative risk 0.820 [95% C.I. (0.74,0.91)]). Additionally, having an abnormal EEG will increase the risk of seizure recurrence (one year relative risk 1.325 [95% C.I. (1.15,1.53)], 8 year

relative risk 1.302 [95% C.I. (1.17,1.45)]). For those who have a seizure post-randomisation, neither treatment policy nor EEG outcome are statistically significant in determining the risk of experiencing a second post-randomisation seizure.

Table 3.1 shows the clinical features for the 1425 individuals included in the exploratory analysis. Over half of the sample experienced tonic-clonic seizures only pre-randomisation (54.8%) and almost a third experienced tonic-clonic seizures with partial seizures (31.9%). Just 7.2% of those randomised had only partial seizures pre-randomisation, with the other seizure groups making up the remaining 6.1%.

Those with tonic-clonic seizures only and tonic-clonic seizures with partial seizures have statistically significantly lower abnormal EEG outcomes than those with partial seizures only pre-randomisation.

By examining the percentages with one and two seizures at different points in time, we can see that numbers appear to be levelling off. The fact that these figures seem to be levelling off suggests that those individuals susceptible to first and second seizures post-randomisation have presented with these seizures by 8 years. These figures also fall well below 100%, suggesting that there may be a cure fraction in the population, that is, not everyone is susceptible to seizures post-randomisation.

	Tonic-Clonic only		Tonic-Clonic with Partial		Generalised Seizures		Partial Seizures		Other Seizures	
	Imm	Def	Imm	Def	Imm	Def	Imm	Def	Imm	Def
Number at randomisation	375	406	239	215	26	23	51	52	21	17
Number with at least one seizure post-rand	155	204	106	117	14	17	29	32	7	10
Number with at least two seizures post-rand	109	123	75	93	11	17	19	24	3	6
Percentage with abnormal EEG (s.e.)	41.6 (2.5)	40.9 (2.4)	40.6 (3.2)	40.9 (3.3)	73.1 (8.7)	52.2 (14.4)	68.6 (6.3)	61.5 (6.5)	38.1 (9.8)	41.2 (10.8)
Percentage with a seizure at:										
6 months	18.4	27.8	21.8	34.9	26.9	56.5	29.4	46.2	19.0	29.4
1 year	26.1	34.4	29.3	41.9	42.3	65.2	43.1	51.9	23.8	52.9
2 years	31.7	41.4	35.6	47.9	46.2	69.6	52.9	55.8	23.8	58.8
5 years	40.3	49.5	42.7	54.0	53.8	69.6	56.9	61.5	33.3	58.8
8 years	41.3	50.2	44.3	54.4	53.8	73.9	56.9	61.5	33.3	58.8
Percentage with two seizures at:										
6 months	12.0	12.6	12.6	20.9	19.2	47.8	21.6	32.7	9.5	23.5
1 year	16.3	18.0	18.8	27.9	30.8	60.9	27.5	40.4	9.5	29.4
2 years	19.2	23.2	24.3	35.3	38.5	69.6	33.3	44.2	9.5	35.3
5 years	26.9	29.3	30.5	42.8	42.3	69.6	37.3	46.2	14.3	35.3
8 years	29.1	30.3	31.4	43.3	42.3	73.9	37.3	46.2	14.3	35.3

Table 3.1: Clinical features for the 1425 individuals for which exploratory analysis was carried out.

3.3 Nonparametric Estimation of Gap Time Distributions in the Analysis of Recurrent Event Data

When data consists of repetitions of the same event through time, there are essentially two possible time scales that may be of interest: the total time, measured from the start of the follow-up, to the occurrence of all the events, or the gap times, that is, the durations between two successive events. Analysis of the MESS data focusses on the analysis of the times from randomisation to first seizure, and the times from first to second seizure, with the overall follow-up time subject to right censoring. When dealing with gap time distributions of recurrent events in this censoring scenario, all the gap times, except the first one, may be subject to dependent censoring (Lin et al. 1999). We consider data where the duration of the time to first seizure will have an effect on the potential censoring value of the second duration. A long time to first seizure post-randomisation implies a short observation period for the time from first to second seizure post-randomisation, and vice versa.

Recall that when examining the percentages of individuals with first and second seizures at different points in time post-randomisation, in Table 3.1, we observed a levelling off of figures. It was proposed that this levelling off suggested that everyone susceptible to two seizures post-randomisation had presented with both by 8 years. This result subsequently means that dependent censoring may not be an issue in the analysis of this dataset. It would appear that the individuals are followed up for the required amount of time such that

a long time to first seizure will not bear on the observation period for the time from first to second seizure.

Let T_{1i} and T_{2i} be the time to first and second seizure respectively, for individual i, $i = 1 \ldots n$. Now set $Y_{1i} = T_{1i}$ and $Y_{2i} = T_{2i} - T_{1i}$, so that Y_{1i} is the actual time to first seizure and Y_{2i} is the actual time from first seizure to the second. Nonparametric estimation of the marginal gap time distributions can be difficult, but Visser (1996) proposes a nonparametric estimator of the conditional survivor function of $Y_2 \mid Y_1$. First recall that we are considering data where the censoring mechanism bears on the sum of the times to first and second seizure, rather than on each time separately. Suppose that Y_1^* and Y_2^* are the random variables associated with the true gap times. The observed random variables are therefore

$$Y_1 = \min(Y_1^*, C), \quad Y_2 = \min(Y_2^*, C - Y_1^*)\mathbb{I}(Y_1^* \leq C),$$

where $\mathbb{I}(A)$ is the indicator function of the event A and the random variable C represents the censoring time. The observed random variables, for n individuals, are $(Y_{1i}, Y_{2i}, \delta_i)$, $i = 1, \ldots, n$, where δ is the censoring indicator, taking the values

$$\delta = \begin{cases} 1, & \text{if } C < Y_1^*, \\ 2, & \text{if } Y_1^* \leq C < Y_1^* + Y_2^*, \\ 3, & \text{if } Y_1^* + Y_2^* \leq C. \end{cases}$$

Visser (1996) assumes that (Y_1^*, Y_2^*, C) are discrete random variables and that they take values in $(0, 1, 2, \ldots, K)$. It follows that Y_1 and Y_2 are also discrete

random variables, taking values in $(0, 1, 2, \ldots, K)$. They denote the bivariate survivor function of the pair (Y_1, Y_2) by $S_{Y_1,Y_2}(k, l) = \mathbb{P}(Y_1 \geq k, Y_2 \geq l)$. Additionally, let $S_{Y_1}(k) = \mathbb{P}(Y_1 \geq k)$ and $S_{Y_2|Y_1}(l \mid k) = \mathbb{P}(Y_2 \geq l \mid Y_1 \geq k)$ denote the survivor function of Y_1 and the conditional survivor function of Y_2 given that $Y_1 \geq k$ respectively.

In addition to the survivor functions, Visser (1996) denotes the hazard function of Y_1 and the conditional hazard function of Y_2 given that $Y_1 \geq k$ by $h_{Y_1}(k) = \mathbb{P}(Y_1 = k \mid Y_1 \geq k)$ and $h_{Y_2|Y_1}(l \mid k) = \mathbb{P}(Y_2 = l \mid Y_1 \geq k, Y_2 \geq l)$ respectively. It is straightforward to evaluate the subsequent survivor function of Y_1, using the associated hazard function:

$$S_{Y_1}(k) = \{1 - h_{Y_1}(0)\} \ldots \{1 - h_{Y_1}(k-1)\},\ k = 1, 2, \ldots, K,\ S_{Y_1}(0) = 1. \quad (3.1)$$

The conditional survivor function and conditional hazard function of Y_2 given that $Y_1 \geq k$ are related in a similar way.

The estimator of the conditional hazard function of Y_2 given that $Y_1 \geq k$, proposed by Visser (1996), is given by

$$\widetilde{h}_{Y_2|Y_1}(l \mid k) = \sum_{i=1}^{n} \mathbb{I}(Y_{1i} \geq k, Y_{2i} = l, \delta_i = 3) \Big\{ \sum_{i=1}^{n} \mathbb{I}(Y_{1i} \geq k, Y_{2i} \geq l) \Big\}^{-1}. \quad (3.2)$$

This estimator is in general a biased estimator for $h_{Y_2|Y_1}(l \mid k)$, except when $k = K$, or when Y_1 and Y_2 are independent. Substituting this estimator into the expression relating the conditional hazard function to the conditional survivor function would result in inconsistent estimates of $S_{Y_2|Y_1}(l \mid k)$. Instead

Visser (1996) considers the estimators of $h_{Y_1}(k)$ and $h_{Y_2|Y_1=k}(l) = \mathbb{P}(Y_2 = l \mid Y_1 = k, Y_2 \geq l)$, given by

$$\widehat{h}_{Y_1}(k) = \sum_{i=1}^{n} \mathbb{I}(Y_{1i} = k, \delta_i \geq 2) \Big\{ \sum_{i=1}^{n} \mathbb{I}(Y_{1i} \geq k) \Big\}^{-1}, \tag{3.3}$$

$$\widehat{h}_{Y_2|Y_1=k}(l) = \sum_{i=1}^{n} \mathbb{I}(Y_{1i} = k, Y_{2i} = l, \delta_i = 3) \Big\{ \sum_{i=1}^{n} \mathbb{I}(Y_{1i} = k, Y_{2i} \geq l) \Big\}^{-1} \tag{3.4}$$

It is then straightforward to obtain estimators, $\widehat{S}_{Y_1}(k)$ and $\widehat{S}_{Y_2|Y_1=k}(l)$, of the survivor functions of Y_1, and Y_2 given that $Y_1 = k$ respectively. The subsequent estimator of $S_{Y_2|Y_1}(l \mid k)$, proposed by Visser (1996), is based on the following transformation:

$$S_{Y_2|Y_1}(l \mid k) = \{S_{Y_1}(k)\}^{-1} \sum_{j=k}^{K} S_{Y_2|Y_1=j}(l)\{S_{Y_1}(j) - S_{Y_1}(j-1)\}. \tag{3.5}$$

We can determine the severity of the dependent censoring in our data set by comparing estimates of $S_{Y_2|Y_1}(l \mid k)$ with Kaplan-Meier estimates of the marginal survivor function for Y_2.

Figure 3.1 shows the estimates of $S_{Y_2|Y_1}(l \mid k)$ plotted against the corresponding Kaplan-Meier estimates of the marginal survivor function for Y_2. We can see that in general Kaplan-Meier estimates of the survivor function are lower than the conditional estimates of the survivor function. We have already discussed that underestimates of survival may be due to the fact that a longer time to first seizure, results in a shorter subsequent observation period for second seizure. This consequently means that more observations may be censored, which would lead to underestimates in survival probabilities.

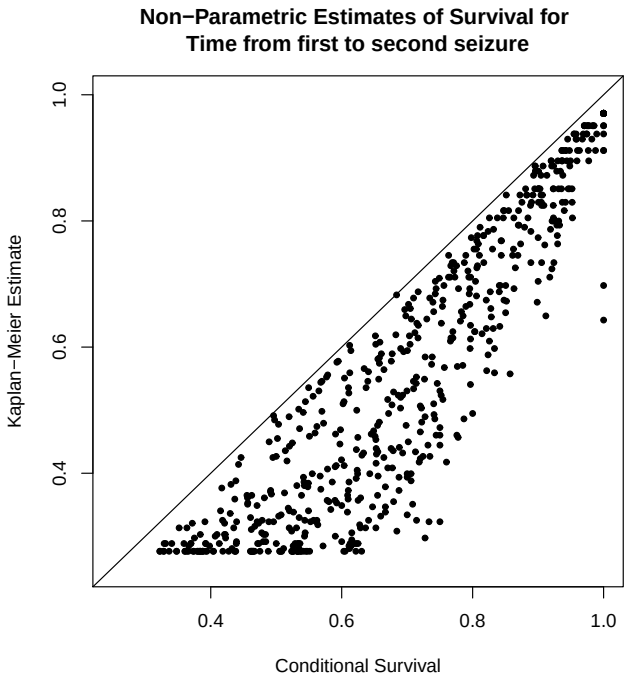


Figure 3.1: Estimates of $S_{Y_2|Y_1}(l \mid k)$ plotted against the corresponding unconditional marginal Kaplan-Meier estimates.

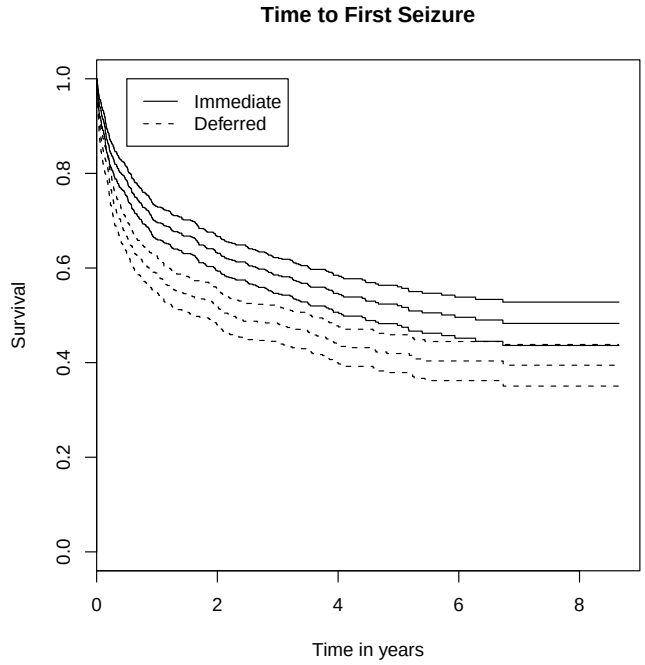
Cook and Lawless (2007) address the issue of dependent censoring and suggests ways to facilitate the examination of marginal gap time distributions. One approach is to fit random effects models, which use individual-specific independent and identically distributed random effects to induce associations among gap times. Such models assume that given a random effect, the gap times for an individual are independent. This is the approach that we shall be considering in our modelling strategy. Other methods considered are the specification of a multivariate model for a specified set of gap times or conditional models.

3.3.1 Kaplan-Meier Plots

We consider Kaplan-Meier plots of the outcomes time to first seizure and time from first seizure to second, examining possible treatment policy, EEG outcome and seizure type effects.

The Kaplan-Meier curves in Figure 3.2 highlight immediately that treatment policy appears to be influential in determining an individual's time to first seizure post-randomisation, but not their time from first to second seizure. A plausible explanation for this is that those individuals randomised to deferred treatment who experience a seizure post-randomisation would most likely receive subsequent treatment with AEDs, bringing them in line with those allocated to immediate treatment thereafter.

Note that the each of the Kaplan-Meier curves in Figure 3.2 have their asymptotes well above zero. This suggests that the associated survivor functions may not be proper and that there may be a proportion of the individuals' included in the MESS trial 'immune' from seizure recurrence post-randomisation. These cure rates are apparent for those allocated to immediate and deferred treatment.



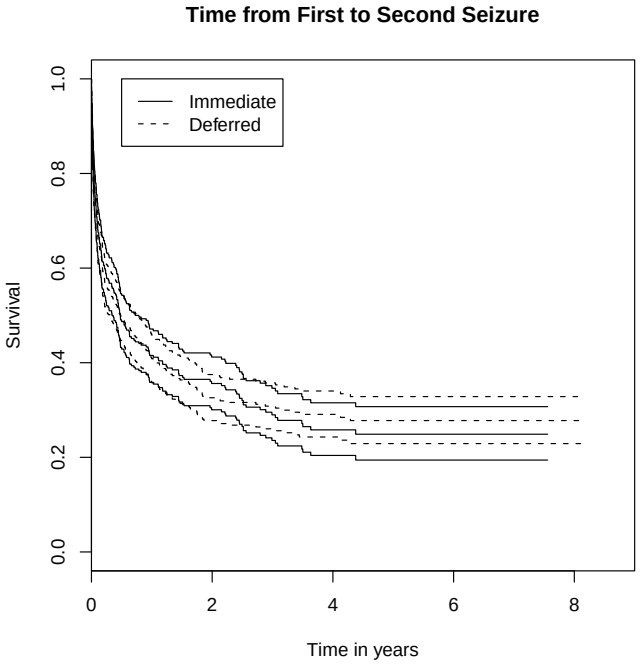


Figure 3.2: Kaplan-Meier curves for time to first seizure and time from first to second seizure (with 95% CI), stratified by treatment policy.

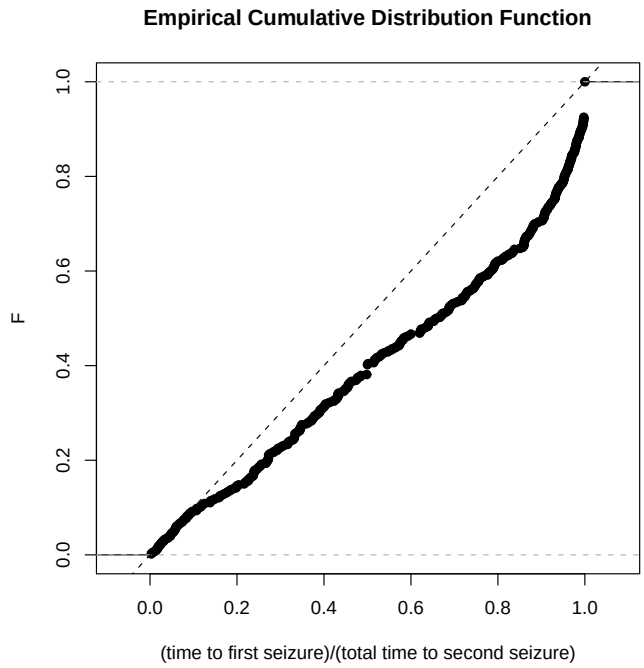


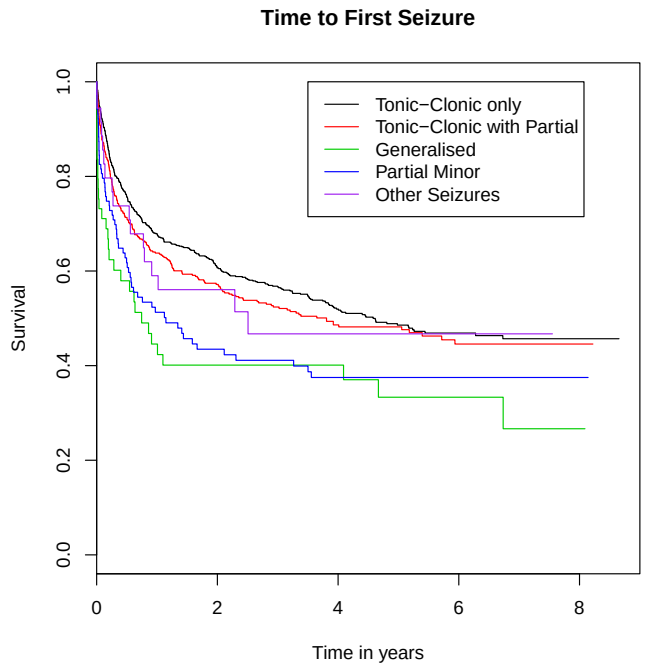
Figure 3.3: Empirical cumulative distribution function for Y_1/T_2.

Figure 3.3 shows the empirical cumulative distribution function for Y_1/T_2. Note that this is only considered for those individuals presenting at least two seizures post-randomisation, as there is evidence to suggest that dependent censoring is present. We observe that this plot has its median at 0.663, suggesting that for those experiencing at least two seizures post-randomisation, their time from first seizure to second is typically shorter than their time from randomisation to first seizure. Around 60% of those having at least two seizures post-randomisation have $Y_1 > Y_2$, with approximately 30% having Y_1 around nine times bigger than Y_2. These results suggest that there may be clustering within seizures.

Figure 3.4 shows that those individuals presenting with generalised or partial seizures pre-randomisation typically have a shorter time to first seizure post-randomisation than the other seizure types. Those individuals with generalised seizures pre-randomisation also present their second seizure post-randomisation much sooner than other seizure types. Additionally, the differences between the Kaplan-Meier curves appear to be more pronounced for the second seizure post-randomisation, than for time to first seizure.

Figure 3.5 suggests that for those participants presenting with partial seizures only pre-randomisation, treatment policy appears to have no effect on their time to first seizure post-randomisation. For all other seizure types immediate treatment is favoured.

When considering EEG outcome, Figure 3.6 indicates that for those presenting a normal EEG, treatment policy has no effect on their associated time to first seizure. For those with an abnormal EEG, allocation to immediate treatment brings their expected time to first seizure in line with those having a normal EEG. Those randomised to deferred treatment, following an abnormal EEG outcome can expect a much shorter time to first seizure post-randomisation. For time from first to second seizure post-randomisation there appears to be no difference in the four Kaplan-Meier curves.



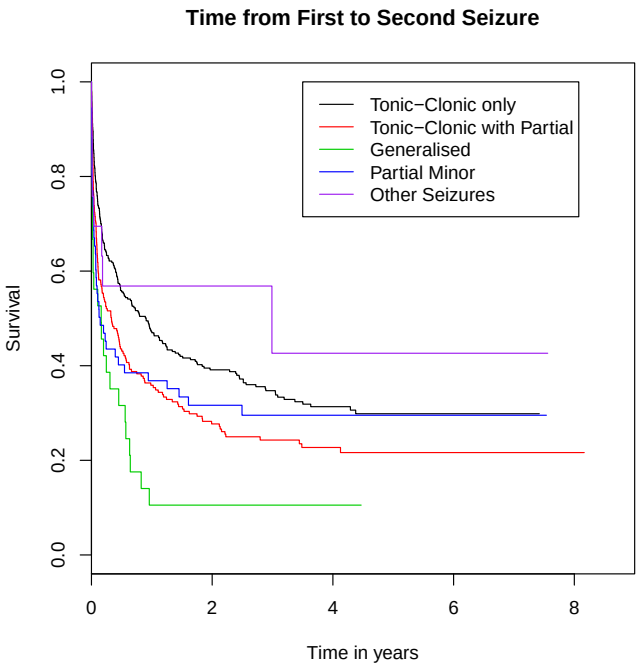
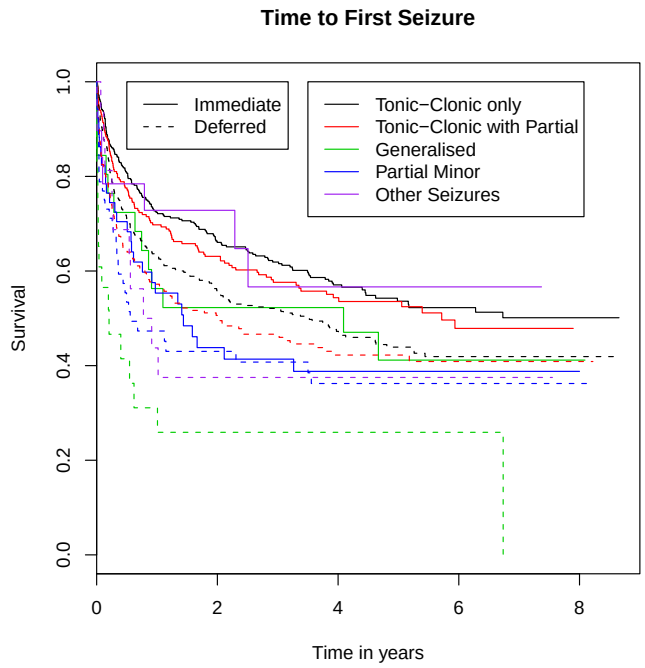


Figure 3.4: Kaplan-Meier curves for time to first seizure and time from first to second seizure, stratified by seizure type pre-randomisation.



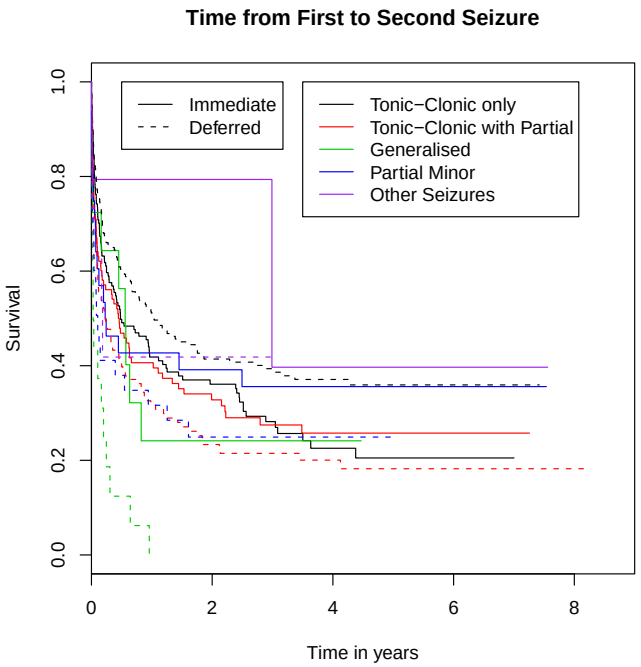


Figure 3.5: Kaplan-Meier curves for time to first seizure and time from first to second seizure, stratified by seizure type pre-randomisation and treatment policy.

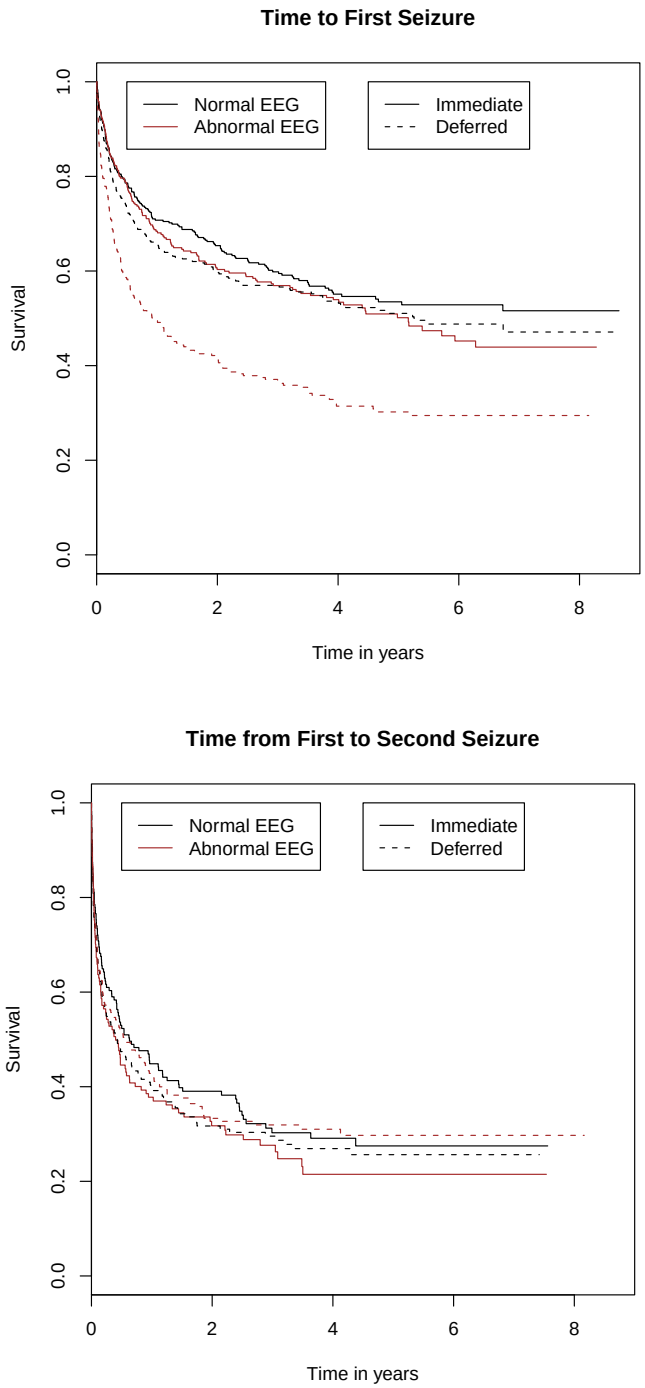
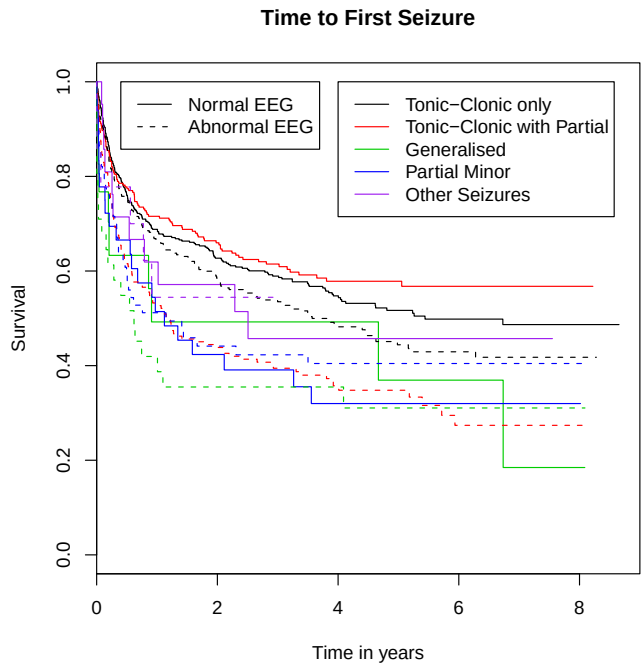


Figure 3.6: Kaplan-Meier curves for time to first seizure and time from first to second seizure, stratified by EEG outcome and treatment policy.

Finally, Figure 3.7 gives an indication as to any interactions between EEG outcome and pre-randomisation seizure types that may be present. For time to first seizure, EEG outcome appears to be influential for those with secondary tonic-clonic seizures pre-randomisation, with those having a normal EEG faring better. EEG outcome also seems to have a slight impact on time to first seizure for those with tonic-clonic seizures only pre-randomisation, and possibly for those with generalised seizures pre-randomisation. These interactions are not seen for time from first to second seizure.



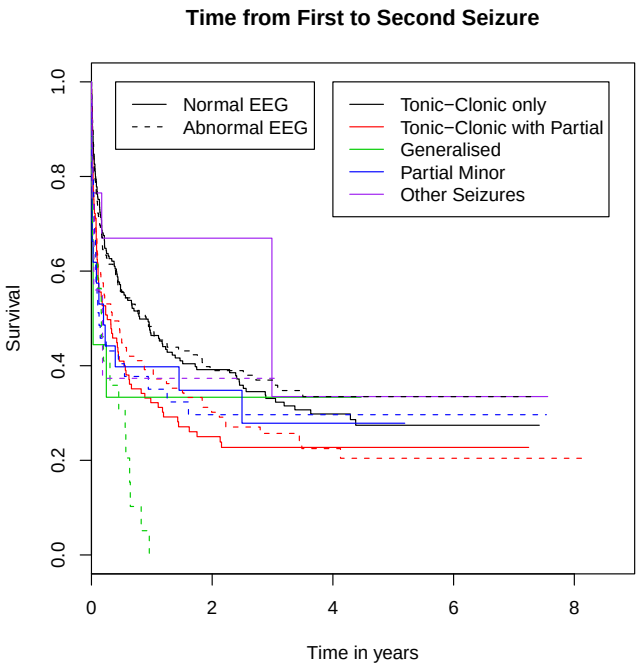


Figure 3.7: Kaplan-Meier curves for time to first seizure and time from first to second seizure, stratified by seizure type pre-randomisation and EEG outcome.

Transformations of the Kaplan-Meier curves can indicate which parametric distributions may be most suitable for formal statistical modelling of survival data. Transforming the Kaplan-Meier estimate of the survivor function to produce a plot that should give a straight line if the assumed model is appropriate is one way of assessing the suitability of parametric models. The survivor function of the Log-logistic distribution with shape parameter γ and scale $1/\mu$ is given by:

$$S(y) = \{1 + (\mu y)^{\gamma}\}^{-1}. \tag{3.6}$$

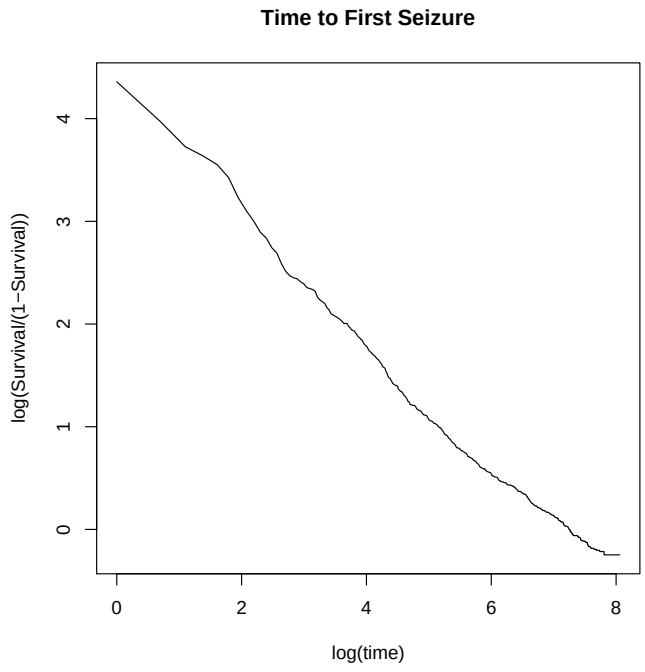
From Equation 3.6, the log-odds of survival beyond y can be expressed as:

$$\ln\left\{\frac{S(y)}{1 - S(y)}\right\} = -\gamma \ln \mu - \gamma \ln y. \tag{3.7}$$

It follows from Equation 3.7 that if the the survivor function is estimated using the Kaplan-Meier estimate, and the subsequent estimated log-odds of survival beyond y are plotted against $\log(y)$, a straight line with intercept $-\gamma \ln \mu$ and slope $-\gamma$ will be observed if the Log-logistic distribution is appropriate.

Figure 3.8 shows the estimated log-odds of survival beyond y plotted against $\ln(y)$ for times to first seizure and from first to second seizure. We observe that both of these plots are straight lines, which supports the suitability of the Log-logistic distribution in the parametric modelling of the MESS data.

Recall that the Log-logistic distribution belongs to the accelerated failure time family of distributions.



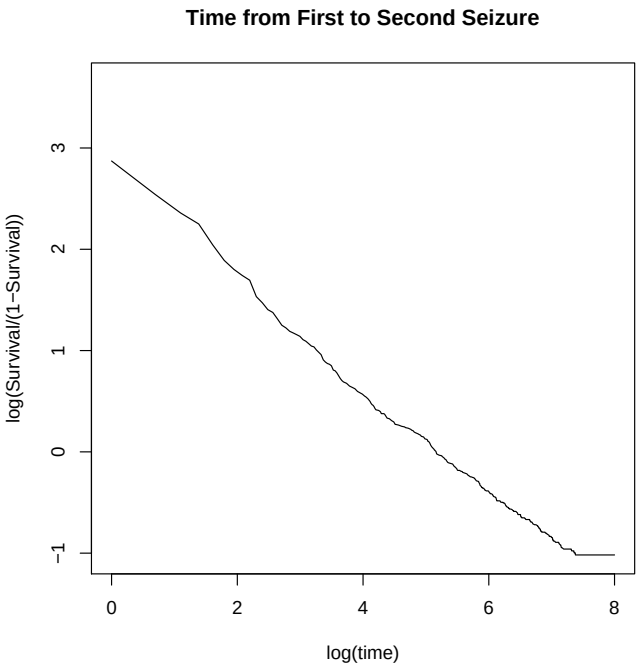


Figure 3.8: Log-odds of Kaplan-Meier estimates of survival for time to first seizure and time from first to second seizure, stratified by seizure type pre-randomisation and EEG outcome.

3.4 Discussion

The MESS data contains information on 1443 patients, randomised to either immediate or deferred treatment. The MESS trial was a pragmatic trial, so that individuals received the optimal type and dose of AED, as determined by the clinician.

Exploratory analysis of the MESS data was carried out on 1425 individuals. There were 18 individuals with missing information, assumed missing completely at random, excluded from the analyses. Previous analyses of the data have concluded that treatment policy is statistically significant in determining times to first and second seizure, but this analysis did not include information available on seizure types and used a log-rank test, or the Cox proportional-hazards model (Marson et al. 2005), despite the data violating the proportional-hazards assumption. An alternative analysis of this data developed a prognostic model based on the existence of a neurological disorder, total number of seizures experienced pre-randomisation and the presence of an abnormal EEG (Kim et al. 2006). The prognostic model was used to categorise individuals as at a low risk, medium risk or high risk of seizure recurrence. For those at a low risk, there was no statistically significant difference in treatment policy on times to first seizure post-randomisation. For those determined as medium or high risk, immediate treatment was favoured.

Non-parametric estimates of the Kaplan-Meier curves are presented in this chapter to provide an initial indication as to possible covariate effects that may exist. Treatment policy seems to be significant overall in determining

times to first seizure post-randomisation, but not times from first to second seizure. We also note that there may be cure rates present in the data. For those who have at least two seizures post-randomisation, their time from first to second seizure is typically shorter than their time to first seizure, suggesting that clustering within seizures may be present.

Our analysis of the MESS data takes into consideration the types of seizures an individual has experienced pre-randomisation. The corresponding Kaplan-Meier curves suggest that seizure type may be significant in determining times to first and from first to second seizure. Additionally, there is evidence to suggest that there may be seizure type interactions with treatment policy and EEG outcome, on post-randomisation survival times. A strong interaction between EEG outcome and treatment policy has also been observed for time to first seizure post-randomisation.

When basing the analysis of recurrent event data on gap times, an important issue that must be addressed is that of dependent censoring. The marginal Kaplan-Meier estimates of the survival function, for the times from first to second seizure, have been compared with conditional estimates of the survivor function of $Y_2 \mid Y_1$, proposed by Visser (1996), to determine the severity of the effect of dependent censoring in this dataset. Methods used to conduct gap time analysis in recurrent event data, when dependent censoring is present, have been discussed.

Following the non-parametric, exploratory analysis of the MESS data that has been carried out in this chapter, we shall first formally analyse each of the

outcomes separately, using standard statistical techniques. In later chapters we shall analyse the pre-randomisation event counts and post-randomisation survival time jointly, in a single model.

Chapter 4

Univariate Analysis of the Event Counts and Survival Times

We shall begin by considering the univariate analysis of the pre-randomisation counts, using standard statistical techniques. We shall then go on to analyse each of the post-randomisation survival times separately, using two standard survival distributions. We shall comment on the statistical significance of the covariates and compare the two survival distributions.

Note that there were five of the 1425 individuals considered in the exploratory analysis with incomplete information on their associated seizure history. These individuals were excluded from the formal statistical modelling. Additionally, Table 3.1 showed that only 3.4% and 2.7% of those randomised experienced either generalised seizures or other seizures respectively pre-randomisation. These groups are sufficiently small that any statistically significant covariate effects associated with these seizure type groups are unlikely to be confirmed in subsequent statistical modelling. The decision was made to also exclude

these individuals from final analyses. This left us with a final sample size of 1334.

4.1 Analysis of Pre-Randomisation Counts

We consider the Negative Binomial Generalised Linear Model as a marginal model for the pre-randomisation event counts, specified by the following probability density function:

$$f_X(x_i; \lambda_i, u_i, \alpha) = \frac{\Gamma(x_i + \alpha)}{x_i! \Gamma(\alpha)} \left(\frac{\lambda_i u_i}{\alpha + \lambda_i u_i} \right)^{x_i} \left(\frac{\alpha}{\alpha + \lambda_i u_i} \right)^{\alpha}, \qquad (4.1)$$

where $\lambda_i = \exp(\boldsymbol{\beta}_1' \boldsymbol{z}_{1i})$. Here $\boldsymbol{z}_{1i}$ is a vector of covariates for individual i, and $\boldsymbol{\beta}_1$ is a vector of regression coefficients, including an intercept term.

All analyses in this thesis have been carried out in R. The estimated regression coefficients for the Negative-Binomial marginal count model are given in Table 4.1. The small value of α suggests that there is substantial heterogeneity within the population. A regression coefficient > 0 (< 0) would indicate an increased (decreased) seizure count relative to the seizure count in the reference group, which contains those individuals presenting with partial seizures only pre-randomisation.

Examination of the coefficients in Table 4.1 tells us that individuals with tonic-clonic seizures only and secondary tonic-clonic seizures pre-randomisation have statistically significantly lower pre-randomisation seizure rates than those individuals presenting with partial seizures only pre-randomisation.

Regression Coefficient	Estimates (standard errors) for Negative Binomial GLM
α	2.088 (0.113)
$\beta_{1,0}$	-4.119 (0.084)
$\beta_{1,t-c^*}$	-1.086 (0.093)
$\beta_{1,2^\circ t-c}$	-0.686 (0.095)
$\beta_{1,partial}$	reference
-Log-likelihood (d.f.)	2486 (1330)

Table 4.1: Estimated regression coefficients for the Negative Binomial GLM. *tonic-clonic.

4.2 Analysis of Post-Randomisation Survival Times

Recall from Figure 3.8 that transformations of the Kaplan-Meier estimates of the survivor function suggested that the survival data may be well modelled through the Log-logistic distribution, which belongs to the accelerated failure time family of distributions. Additionally, recall that in the joint model, developed by Cowling et al. (2006), the unconditional distribution of the post-randomisation times to first seizure, Y_i, was the Lomax distribution. We therefore consider these survival distributions for the two post-randomisation survival times separately, namely time to first seizure and time from first to second seizure. The Log-logistic and Lomax distributions are defined by the following probability density functions:

- Log-logistic (shape=γ, scale=$1/\mu_i$)

$$f_Y(y_i;\mu_i,\gamma) = \frac{\mu_i\gamma(\mu_i y_i)^{\gamma-1}}{(1+(\mu_i y_i)^\gamma)^2}, \tag{4.2}$$

- Lomax (shape=γ, scale=γ/μ_i)

$$f_Y(y_i;\mu_i,\gamma) = \mu_i\Big(\frac{\gamma}{\gamma+\mu_i y_i}\Big)^{\gamma+1}, \tag{4.3}$$

where in each model $\mu_i = \exp(\boldsymbol{\theta}'\boldsymbol{d}_i)$ for a vector $\boldsymbol{\theta}$ of regression coefficients, and a vector $\boldsymbol{d}_i$ of covariates for each individual i, including an intercept term. Increasing values of the m regression coefficients, θ_k, $k = 0,\ldots,m$, correspond to an increase in the acceleration factor, and hence a decrease in the expected time to seizure. Conversely, negative values of θ_k, $k = 0,\ldots,m$, correspond to deceleration and an increase in the expected time to seizure. The parameter $\gamma > 0$ is a shape parameter and represents the degree of additional heterogeneity within the population, with smaller values indicating higher levels of heterogeneity.

Recall that exploratory analysis supported the use of the Log-logistic distribution and note that equations (4.2) and (4.3) are equivalent when $\gamma = 1$. Hence a value of γ close to 1 (indicating that there is considerable heterogeneity in the population) would suggest that the data could be sufficiently modelled through the Lomax distribution, validating the use of the joint model proposed by Cowling et al. (2006) and presented in Chapter 2. We now present the estimated regression coefficients obtained when each of these distributions is fit to the post-randomisation survival times, time to first seizure and time from first to second seizure.

Time to First Seizure

The parameter estimates for time to first seizure only, for the Log-Logistic and Lomax distributions are given in Table 4.2. The reference group contains individuals with partial seizures pre-randomisation, with a normal EEG and randomised to deferred treatment.

Regression Coefficient	Estimates (standard errors) for the following models:			
	Log-logistic		Lomax	
θ_0	-0.381	(1.156)	0.572	(1.016)
θ_{trt}	-1.217	(1.421)	-0.927	(1.319)
θ_{t-c}	-0.567	(0.604)	-0.494	(0.602)
$\theta_{2^\circ t-c}$	-0.853	(0.620)	-0.660	(0.624)
$\theta_{partial}$	reference		reference	
θ_{eeg}	-0.340	(1.399)	0.072	(1.283)
$\theta_{\ln(rate)}$	1.227	(0.217)	1.138	(0.189)
$\theta_{t-c\times trt}$	-0.968	(0.648)	-1.040	(0.663)
$\theta_{2^\circ t-c\times trt}$	-0.976	(0.665)	-0.898	(0.684)
$\theta_{partial\times trt}$	reference		reference	
$\theta_{eeg\times trt}$	-1.118	(0.347)	-1.138	(0.357)
$\theta_{\ln(rate)\times trt}$	-0.366	(0.266)	-0.327	(0.251)
$\theta_{t-c\times eeg}$	0.598	(0.666)	-0.032	(0.246)
$\theta_{2^\circ t-c\times eeg}$	1.668	(0.685)	0.497	(0.685)
$\theta_{partial\times eeg}$	reference		reference	
$\theta_{\ln(rate)\times eeg}$	-0.117	(0.264)	1.190	(0.708)
γ	0.617	(0.021)	0.233	(0.008)
-Log-likelihood (d.f.)	5112	(1321)	5096	(1321)

Table 4.2: Estimated regression coefficients, for the two survival models, fitted to the times to first seizure.

We begin by conducting the Wald test (Wald 1943) on each of the estimated regression coefficients given in Table 4.2 and find that for both survival distributions the only statistically significant covariates are $\theta_{\ln(rate)}$ and $\theta_{eeg\times trt}$. Furthermore, those experiencing tonic-clonic only and secondary tonic-clonic

seizures pre-randomisation can typically expect to have a longer time to first seizure post-randomisation, than those with partial seizures only. Additionally $\theta_{2^\circ t-c\times eeg}$ is significant in the Log-logistic model only. The lack of statistically significant covariates in these models is contrary to the observations made through the investigation of the Kaplan-Meier curves. Probably most surprising is that the exploratory analysis suggested that treatment policy should be significant, but this is not supported by the coefficient estimates.

The Log-logistic and Lomax distributions can not be compared using the standard likelihood ratio test, as they are non-nested models. The Akaike Information Criterion (AIC) (Sahamotoa et al. 1986) is a method for comparing two non-nested models and is given by $2(m-\ell)$, where m is the number of parameters in the model and ℓ is the maximised log-likelihood associated with the model. Computing the AIC for each of the two survival distributions we have that for time to first seizure the AIC for the Log-logistic model is 10250, and for the Lomax distribution the corresponding AIC is 10217. This indicates that the Lomax distribution is preferred over the Log-logistic distribution.

Time from First to Second Seizure

The parameter estimates for the times from first to second seizure, for the Log-Logistic and Lomax distributions are given in Table 4.3. If we compute the AIC for each of these distributions we have that for the Log-logistic distribution the AIC is 5652, and the AIC for the Lomax distribution is 5614, meaning that again the Lomax distribution is the preferred of the two distributions.

If we now compute the Wald statistics for each of the estimated regression coefficients presented in Table 4.3, we see that only θ_{t-c} is statistically significant in the Lomax survival distribution. Additionally, only $\theta_{\ln(rate)}$ is significant in the Log-logistic distribution. This time, however, the lack of statistically significant covariates is not as surprising. The exploratory analysis suggested that some covariates that were statistically significant in determining the times to first seizure failed to be statistically significant when considering the times from first to second seizure.

Regression Coefficient	Estimates (standard errors) for the following models:			
	Log-logistic		Lomax	
θ_0	-0.234	(1.497)	-0.085	(1.294)
θ_{trt}	-2.183	(1.893)	-1.650	(1.708)
θ_{t-c}	-1.770	(0.894)	-2.188	(0.783)
$\theta_{2^\circ t-c}$	-0.364	(0.901)	-1.063	(0.802)
$\theta_{partial}$	reference		reference	
θ_{eeg}	-2.013	(1.800)	-0.918	(1.616)
$\theta_{\ln(rate)}$	0.780	(0.294)	0.440	(0.250)
$\theta_{t-c\times trt}$	1.747	(0.932)	1.121	(0.874)
$\theta_{2^\circ t-c\times trt}$	0.647	(0.944)	0.540	(0.887)
$\theta_{partial\times trt}$	reference		reference	
$\theta_{eeg\times trt}$	0.492	(0.499)	0.382	(0.481)
$\theta_{\ln(rate)\times trt}$	-0.175	(0.369)	-0.148	(0.342)
$\theta_{t-c\times eeg}$	0.192	(0.953)	0.920	(0.902)
$\theta_{2^\circ t-c\times eeg}$	0.043	(0.966)	0.716	(0.917)
$\theta_{partial\times eeg}$	reference		reference	
$\theta_{\ln(rate)\times eeg}$	-0.301	(0.354)	0.040	(0.325)
γ	0.595	(0.024)	0.270	(0.012)
-Log-likelihood (d.f.)	2813	(1321)	2794	(1321)

Table 4.3: Estimated regression coefficients, for the two survival models, fitted to the times from first to second seizure.

4.3 Discussion

In Chapter 3, exploratory analysis was carried out on 1425 individuals. In the formal analysis of the MESS data however, five individuals, for whom there was no information on their pre-randomisation seizure history, were removed, as were all those individuals presenting with either generalised or other seizures pre-randomisation. This left us with a final sample size of 1334 individuals all presenting with tonic-clonic and/or partial seizures.

The Negative Binomial Generalised Linear Model was considered as a marginal model for the event counts, with the Log-logistic and Lomax accelerated failure time models being adopted for the marginal survival times.

Seizure type was found to be statistically significant in determining an individuals' pre-randomisation seizure count. Those individuals experiencing partial seizures only, typically have more seizures pre-randomisation, than those presenting with tonic-clonic seizures.

In the analysis of the post-randomisation survival times, the Lomax distribution was found to be more suitable than the Log-logistic distribution for both times to first, and from first to second seizures. A lack of statistically significant exploratory variables was observed. This surprising result is in stark contrast to the conclusions drawn following the examination of the Kaplan-Meier curves in Chapter 3.

Chapter 5

The Joint Model

Analysis of the post-randomisation survival times in a univariate setting, using standard survival distributions, was unable to pick up on covariate effects that were suggested by the examination of the Kaplan-Meier curves.

Cowling et al. (2006) compared the joint model for pre-randomisation event counts and post-randomisation survival times with the best fitting standard survival distribution that treated the pre-randomisation event count information as a covariate. A simulation study of power (Cowling 2003) indicated that the joint model provided more precise estimates of the treatment effect than the standard parametric survival models. The joint model also had more power to identify interaction effects not affirmed by the standard survival models.

We shall use the model developed by Cowling et al. (2006) as the basis for a joint model that analyses pre-randomisation event counts and two post-randomisation survival times together.

5.1 Building the Joint Model for the Event Counts Pre-Randomisation and Two Post-Randomisation Survival Times

In many treatment studies it is often time to first event that is measured, however epilepsy is characterised by recurrent seizures, not a single, isolated event. Our data arrives in three parts: a pre-randomisation event count, post-randomisation time to first seizure and post-randomisation time from first to second seizure. Our aim is to develop methodology that analyses these outcomes jointly, in a single model.

Let T_{1i} and T_{2i} be the times from randomisation to first and second seizure respectively, for individual i, $i = 1 \dots n$. Now, setting $Y_{1i} = T_{1i}$ and $Y_{2i} = T_{2i} - T_{1i}$, gives that Y_{1i} is subsequently the time to first seizure, and Y_{2i} is the time from first seizure to the second. We shall assume individuals experience seizures according to a Poisson process with rate $\lambda_i \nu_i$, where the parameter λ_i relates to the baseline covariates, with additional heterogeneity in the population being modelled through ν_i, assumed to follow a Gamma(α, α) distribution. Smaller values of α are indicative of higher levels of heterogeneity. Consequently, the pre-randomisation event count, for individual i, over period u_i, X_i, follows a Poisson distribution with mean and variance $\lambda_i u_i \nu_i$.

A consequence of the Poisson process is that interevent times are Exponential, so that post-randomisation survival times to first seizure, and from first to second seizure, Y_{1i} and Y_{2i}, will be independent, conditional on the random

effect term, and Exponentially distributed with rate $\lambda_i\psi_i\nu_i$. The parameter ψ_i is a post-randomisation seizure rate modifier, related to the individuals' treatment in some way. In summary:

$$\begin{aligned} X_i \mid \nu_i &\sim \text{Poisson}(\lambda_i u_i \nu_i), \\ Y_{ji} \mid \nu_i &\sim \text{Exponential}(\lambda_i \psi_i \nu_i), \qquad j = 1, 2, \\ \nu_i &\sim \text{Gamma}(\alpha, \alpha). \end{aligned}$$

The joint density of the survival times is the product of the densities of Y_{1i} and Y_{2i}, so that the joint model is specified by the following equations:

$$\begin{aligned} f_{X|\nu}(x_i \mid \nu_i; \lambda_i, u_i) &= \frac{(\lambda_i u_i \nu_i)^{x_i} \exp(-\lambda_i u_i \nu_i)}{x_i!}, \\ f_{Y_1,Y_2|\nu}(y_{1i}, y_{2i} \mid \nu_i; \lambda_i, \psi_i) &= (\lambda_i \psi_i \nu_i)^2 \exp(-\lambda_i \psi_i \nu_i (y_{1i} + y_{2i})), \\ f_\nu(\nu_i; \alpha) &= \frac{\alpha^\alpha \nu_i^{\alpha-1} \exp(-\alpha \nu_i)}{\Gamma(\alpha)}, \end{aligned}$$

where $\lambda_i = \exp(\boldsymbol{\beta}_1' \boldsymbol{z}_{1i})$, $\psi_i = \exp(\boldsymbol{\beta}_2' \boldsymbol{z}_{2i})$ and $\boldsymbol{z}_{1i}$, $\boldsymbol{z}_{2i}$ are vectors of covariates, not necessarily distinct.

5.1.1 Marginal Distributions

If the random effect term is integrated out of the joint density of X_i and ν_i, then the resulting unconditional density, $f_X(x_i; \lambda_i, u_i, \alpha)$, is simply the Negative Binomial (Equation 4.1). The unconditional joint distribution of the Y_{ji}, $j = 1, 2$, obtained when the random effects are integrated out of the joint density of the survival times, Y_{1i} and Y_{2i}, and ν_i, is the bivariate Lomax distribution (Nayak 1987). This distribution has the following density and survivor

functions:

$$\begin{aligned} f_{Y_1,Y_2}(y_{1i}, y_{2i}; \lambda_i, \psi_i, \alpha) &= \int_0^\infty f_{Y_1,Y_2|\nu}(y_{1i}, y_{2i} \mid \nu_i; \lambda_i, \psi_i) g_\nu(\nu_i; \alpha) d\nu_i \\ &= \frac{\alpha+1}{\alpha}(\lambda_i\psi_i)^2\left\{1 + \frac{\lambda_i\psi_i(y_{1i}+y_{2i})}{\alpha}\right\}^{-(\alpha+2)} \end{aligned} \tag{5.1}$$

$$\begin{aligned} S_{Y_1,Y_2}(y_{1i}, y_{2i}; \lambda_i, \psi_i, \alpha) &= \int_{y_{2i}}^\infty \int_{y_{1i}}^\infty f_{Y_1,Y_2}(u, v; \lambda_i, \psi_i, \alpha) du\, dv \\ &= \left\{1 + \frac{\lambda_i\psi_i(y_{1i}+y_{2i})}{\alpha}\right\}^{-\alpha}. \end{aligned} \tag{5.2}$$

Each of the Y_{ji} have univariate Lomax marginal distributions, with shape and scale parameters α and $\alpha/\lambda_i\psi_i$ respectively, with density:

$$f_{Y_j}(y_{ji}; \lambda_i, \psi_i, \alpha) = \frac{\lambda_i\psi_i}{(1+\lambda_i\psi_i y_{ji}/\alpha)^{\alpha+1}}, \qquad j = 1, 2.$$

5.1.2 The Full Log-Likelihood and Derivatives

When formulating the likelihood, we need to consider the different ways that censoring can occur. There are three different ways censoring can arise in this setting, namely: (i) Y_{1i} and Y_{2i} are both observed, (ii) Y_{1i} is observed, but Y_{2i} is censored, and (iii) Y_{1i} is censored, so Y_{2i} is taken to be censored at zero. We now consider these three situations separately.

Joint Distribution with Y_{1i} and Y_{2i} Observed

In this situation the joint density of Y_{1i} and Y_{2i} contributes towards the likelihood, giving

$$\int_0^\infty f_{X|\nu}(x_i \mid \nu_i; \lambda_i, u_i) f_{Y_1,Y_2|\nu}(y_{1i}, y_{2i} \mid \nu_i; \lambda_i, \psi_i) f_\nu(\nu_i; \alpha) d\nu_i$$

$$= \frac{(\lambda_i u_i)^{x_i}}{x_i!} \frac{(\lambda_i \psi_i)^2 \alpha^\alpha}{\Gamma(\alpha)} \frac{\Gamma(x_i + \alpha + 2)}{(\lambda_i u_i + \lambda_i \psi_i (y_{1i} + y_{2i}) + \alpha)^{x_i + \alpha + 2}}. \tag{5.3}$$

Joint Distribution with Y_{1i} Observed and Y_{2i} Censored

In this situation the density of Y_{1i} and the survivor function for Y_{2i} contribute to the likelihood, giving

$$\int_0^\infty f_{X|\nu}(x_i \mid \nu_i; \lambda_i, u_i) f_{Y_1|\nu}(y_{1i} \mid \nu_i; \lambda_i, \psi_i) S_{Y_2|\nu}(y_{2i} \mid \nu_i; \lambda_i, \psi_i) f_\nu(\nu_i; \alpha) d\nu_i$$

$$= \frac{(\lambda_i u_i)^{x_i}}{x_i!} \frac{\lambda_i \psi_i \alpha^\alpha}{\Gamma(\alpha)} \frac{\Gamma(x_i + \alpha + 1)}{(\lambda_i u_i + \lambda_i \psi_i (y_{1i} + y_{2i}) + \alpha)^{x_i + \alpha + 1}}. \tag{5.4}$$

Joint Distribution with Y_{1i} Censored, so Y_{2i} Taken to be Censored at Zero

In this situation it is the survivor functions of Y_{1i} and Y_{2i} that will contribute to the likelihood, however we assume that the second survival time is censored at zero, giving $S_{Y_2|\nu}(0 \mid \nu_i; \lambda_i, \psi_i) = 1$.

$$\int_0^\infty f_{X|\nu}(x_i \mid \nu_i; \lambda_i, u_i) S_{Y_1|\nu}(y_{1i} \mid \nu_i; \lambda_i, \psi_i) f_\nu(\nu_i; \alpha) d\nu_i$$

$$= \frac{(\lambda_i u_i)^{x_i}}{x_i!} \frac{\alpha^\alpha}{\Gamma(\alpha)} \frac{\Gamma(x_i + \alpha)}{(\lambda_i u_i + \lambda_i \psi_i y_{1i} + \alpha)^{x_i + \alpha}}. \tag{5.5}$$

Conversely, by keeping $S_{Y_2|\nu}(y_{2i} \mid \nu_i; \lambda_i, \psi_i)$ in the calculations, we subsequently obtain a simpler likelihood function, so we proceed in this way to obtain

$$\int_0^\infty f_{X|\nu}(x_i \mid \nu_i; \lambda_i, u_i) S_{Y_1|\nu}(y_{1i} \mid \nu_i; \lambda_i, \psi_i) S_{Y_2|\nu}(y_{2i} \mid \nu_i; \lambda_i, \psi_i) f_\nu(\nu_i; \alpha) d\nu_i$$

$$= \frac{(\lambda_i u_i)^{x_i}}{x_i!} \frac{\alpha^\alpha}{\Gamma(\alpha)} \frac{\Gamma(x_i + \alpha)}{(\lambda_i u_i + \lambda_i \psi_i (y_{1i} + y_{2i}) + \alpha)^{x_i + \alpha}}. \tag{5.6}$$

Note that equations (5.5) and (5.6) are equivalent when $y_{2i} = 0$.

Log-likelihood

Let δ_{ji} be the indicator function for the jth survival time, taking the value 1 if the seizure is observed, and zero if the survival time is censored. Combining these indicator functions with equations (5.3)-(5.6) allows us to formulate the log-likelihood for the observed data $\mathcal{D}$, for all the n individuals, given by

$$\begin{aligned}\ell(\alpha, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \mid \mathcal{D}) = \sum_{i=1}^{n} \Bigg\{ & \left[\sum_{k=0}^{x_i-1} \ln(\alpha + k) \right] + (x_i + \delta_{1i}(1 + \delta_{2i})) \ln(\lambda_i) + x_i \ln(u_i) \\ & - \ln(x_i!) + \alpha \ln(\alpha) + \delta_{1i}(1 + \delta_{2i}) \ln(\psi_i) + \delta_{1i} \ln(x_i + \alpha) \\ & + \delta_{1i}\delta_{2i} \ln(x_i + \alpha + 1) - (1 - \delta_{1i})(x_i + \alpha) \ln\left(\lambda_i u_i + \lambda_i \psi_i y_{1i} + \alpha\right) \\ & - \delta_{1i}(\delta_{2i} + x_i + \alpha + 1) \ln\left(\lambda_i u_i + \lambda_i \psi_i (y_{1i} + y_{2i}) + \alpha\right) \Bigg\}. \end{aligned} \tag{5.7}$$

First and second derivatives of this log-likelihood can easily be obtained, allowing inference on the parameters α, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ using a numerical method such as Newton Raphson.

First Derivatives

The first-order derivatives of the full log-likelihood are

$$\begin{aligned}
\frac{\partial \ell}{\partial \boldsymbol{\beta}_1} &= \sum_{i=1}^{n} \left\{ \frac{\alpha(x_i + \delta_{1i}(1 + \delta_{2i}) - \lambda_i u_i - \lambda_i \psi_i (y_{1i} + y_{2i}))}{\lambda_i u_i + \lambda_i \psi_i (y_{1i} + y_{2i}) + \alpha} \right\} \boldsymbol{z}_{1i}, \\
\frac{\partial \ell}{\partial \boldsymbol{\beta}_2} &= \sum_{i=1}^{n} \left\{ \frac{\delta_{1i}(1 + \delta_{2i})(\lambda_i u_i + \alpha) - \lambda_i \psi_i (y_{1i} + y_{2i})(x_i + \alpha)}{\lambda_i u_i + \lambda_i \psi_i (y_{1i} + y_{2i}) + \alpha} \right\} \boldsymbol{z}_{1i}, \\
\frac{\partial \ell}{\partial \alpha} &= \sum_{i=1}^{n} \Bigg\{ \left[\sum_{k=0}^{x_i-1} \frac{1}{\alpha + k} \right] + \frac{\delta_{1i}}{x_i + \alpha} + \frac{\delta_{1i}\delta_{2i}}{x_i + \alpha + 1} + \ln(\alpha) + 1 \\
&\quad - \ln(\lambda_i u_i + \lambda_i \psi_i (y_{1i} + y_{2i}) + \alpha) - \frac{x_i + \alpha + \delta_{1i}(1 + \delta_{2i})}{\lambda_i u_i + \lambda_i \psi_i (y_{1i} + y_{2i}) + \alpha} \Bigg\}.
\end{aligned}$$

Second Derivatives

The second-order derivatives of the full log-likelihood are

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1 \boldsymbol{\prime}} &= -\sum_{i=1}^{n} \left\{ \frac{\alpha(x_i + \delta_{1i}(1+\delta_{2i}+\alpha)(\lambda_i u_i + \lambda_i \psi_i(y_{1i}+y_{2i}))}{(\lambda_i u_i + \lambda_i \psi_i(y_{1i}+y_{2i}) + \alpha)^2} \right\} \boldsymbol{z}_{1i} \boldsymbol{z}_{1i} \boldsymbol{\prime}, \\
\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2 \boldsymbol{\prime}} &= -\sum_{i=1}^{n} \left\{ \frac{\alpha(\alpha + x_i + \delta_{1i}(1+\delta_{2i}))\lambda_i \psi_i(y_{1i}+y_{2i})}{(\lambda_i u_i + \lambda_i \psi_i(y_{1i}+y_{2i}) + \alpha)^2} \right\} \boldsymbol{z}_{1i} \boldsymbol{z}_{2i} \boldsymbol{\prime}, \\
\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}_1 \partial \alpha} &= \sum_{i=1}^{n} \left\{ \frac{(\lambda_i u_i + \lambda_i \psi_i(y_{1i}+y_{2i}))(x_i + \delta_{1i}(1+\delta_{2i}) - \lambda_i u_i}{(\lambda_i u_i + \lambda_i \psi_i(y_{1i}+y_{2i}) + \alpha)^2} \right. \\
&\quad \left. - \frac{\lambda_i \psi_i(y_{1i}+y_{2i}))}{(\lambda_i u_i + \lambda_i \psi_i(y_{1i}+y_{2i}) + \alpha)^2} \right\} \boldsymbol{z}_{1i}, \\
\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2 \boldsymbol{\prime}} &= -\sum_{i=1}^{n} \left\{ \frac{(x_i + \alpha + \delta_{1i}(1+\delta_{2i}))(\lambda_i u_i + \alpha)\lambda_i \psi_i(y_{1i}+y_{2i})}{(\lambda_i u_i + \lambda_i \psi_i(y_{1i}+y_{2i}) + \alpha)^2} \right\} \boldsymbol{z}_{2i} \boldsymbol{z}_{2i} \boldsymbol{\prime}, \\
\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}_2 \partial \alpha} &= \sum_{i=1}^{n} \left\{ \frac{(x_i - \lambda_i u_i + \delta_{1i}(1+\delta_{2i}))\lambda_i \psi_i(y_{1i}+y_{2i})}{(\lambda_i u_i + \lambda_i \psi_i(y_{1i}+y_{2i}) + \alpha)^2} \right. \\
&\quad \left. - \frac{(\lambda_i \psi_i(y_{1i}+y_{2i}))\lambda_i \psi_i(y_{1i}+y_{2i})}{(\lambda_i u_i + \lambda_i \psi_i(y_{1i}+y_{2i}) + \alpha)^2} \right\} \boldsymbol{z}_{2i}, \\
\frac{\partial^2 \ell}{\partial \alpha \partial \alpha} &= -\sum_{i=1}^{n} \left\{ \left[\sum_{k=0}^{x_i - 1} \frac{1}{(\alpha+k)^2} \right] + \frac{\delta_{1i}}{(x_i+\alpha)^2} + \frac{\delta_{1i}\delta_{2i}}{(x_i+\alpha+1)^2} - \frac{1}{\alpha} \right. \\
&\quad \left. - \frac{(x_i + \delta_{1i}(1+\delta_{2i}) - \alpha - 2(\lambda_i u_i + \lambda_i \psi_i(y_{1i}+y_{2i})))}{(\lambda_i u_i + \lambda_i \psi_i(y_{1i}+y_{2i}) + \alpha)^2} \right\}.
\end{aligned}$$

5.2 Implementing the Joint Model

We consider two versions of the joint model: the joint model proposed by Cowling et al. (2006), which considers the pre-randomisation event counts and times to first post-randomisation seizure only (Joint Model A), and the joint model we have developed here that models the pre-randomisation event counts and post-randomisation seizure times to first and second seizure (Joint Model B). The estimated regression coefficients for the two fitted models are given

in Table 5.1. A regression coefficient > 0 (< 0) would indicate an increased (decreased) seizure rate relative to the seizure rate in the reference group. The reference group contains individuals with partial seizures pre-randomisation, with a normal EEG and randomised to deferred treatment.

It is encouraging to note that the estimated regression coefficients contained in λ_i are very similar to those obtained through the Negative-Binomial marginal count model, presented in Table 4.1.

If we conduct the Wald test on each of the regression coefficients in ψ_i we can see that in the joint models there are numerous significant covariates, indeed all but the pre-randomisation seizure types and EEG outcome are significant. It may be surprising to conclude that the pre-randomisation seizure types are not significant in the model, but note that the interaction terms are highly significant.

If we compare the significant variables appearing in ψ_i, for each of the two models considered, the estimated regression coefficients observed in Joint Model B are closer to zero than those estimates observed in Joint Model A. Recall that the exploratory analysis suggested that covariates that were significant in determining the times to first seizure post-randomisation, may not be significant when analysing the times from first to second post-randomisation seizure. This may explain the averaging down effect observed here and suggests that we should not assume that the ψ_i remains constant between post-randomisation seizures.

	Regression Coefficient	Estimates (standard errors) for the following models:			
		Joint Model A		Joint Model B	
	α	1.942	(0.105)	1.738	(0.092)
λ_i	$\beta_{1,0}$	-4.133	(0.087)	-4.145	(0.091)
	$\beta_{1,t-c}$	-1.074	(0.097)	-1.054	(0.100)
	$\beta_{1,2^\circ t-c}$	-0.694	(0.100)	-0.697	(0.103)
	$\beta_{1,partial}$	reference		reference	
ψ_i	$\beta_{2,0}$	-2.759	(0.327)	-2.067	(0.252)
	$\beta_{2,trt}$	0.979	(0.331)	0.433	(0.263)
	$\beta_{2,t-c}$	0.549	(0.337)	0.057	(0.262)
	$\beta_{2,2^\circ t-c}$	0.242	(0.347)	0.067	(0.270)
	$\beta_{2,partial}$	reference		reference	
	$\beta_{2,eeg}$	-0.251	(0.338)	-0.537	(0.268)
	$\beta_{2,t-c\times trt}$	-1.214	(0.333)	-0.491	(0.265)
	$\beta_{2,2^\circ t-c\times trt}$	-1.340	(0.346)	-0.854	(0.274)
	$\beta_{2,partial\times trt}$	reference		reference	
	$\beta_{2,eeg\times trt}$	-0.629	(0.184)	-0.377	(0.148)
	$\beta_{2,t-c\times eeg}$	0.777	(0.347)	0.875	(0.277)
	$\beta_{2,2^\circ t-c\times eeg}$	1.478	(0.359)	1.272	(0.286)
	$\beta_{2,partial\times eeg}$	reference		reference	
-Log-likelihood (d.f.)		7872	(1320)	11233	(1320)

Table 5.1: Estimated regression coefficients for the joint models. The term λ_i contains parameter estimates corresponding to the effect of covariates on the underlying event rate and ψ_i contains parameter estimates corresponding to the effect of covariates on the post-randomisation reduction in event rates.

5.3 Interpretation of Results

To gain a better understanding of the estimated regression coefficients, given in Table 5.1, we can obtain subsequent estimates of the pre-randomisation seizure rates and the post-randomisation seizure rate modifiers, for the different seizure types, EEG outcomes and treatment policies. We shall use the estimates given by Joint Model B.

Table 5.2 gives the expected pre-randomisation seizure rates per unit time,

for the different seizure types. We can see that those individuals presenting with partial seizures pre-randomisation can typically expect to have the highest seizure rate, with those experiencing tonic-clonic seizures only and secondary tonic-clonic seizures having statistically significantly lower rates.

Seizure Type	$\widehat{\lambda}_i$ (95% C.I.)		Expected yearly rate
Tonic-Clonic	0.0055	(0.005,0.006)	2
2° Tonic-Clonic	0.008	(0.007,0.009)	3
Partial	0.016	(0.013,0.019)	6

Table 5.2: The expected pre-randomisation seizure rate per unit time and the corresponding expected yearly seizure rate, using the estimated regression coefficients from Joint Model B.

Table 5.3 gives estimates of the expected post-randomisation change in seizure rate, stratified by seizure type, EEG outcome and treatment policy. As an example consider a person presenting with tonic-clonic seizures only pre-randomisation, with an abnormal EEG and randomised to deferred treatment. Table 5.2 tells us that their expected pre-randomisation seizure rate per unit time, $\widehat{\lambda}_i$, is 0.0055, which equates to a seizure approximately every 182 days. Their subsequent $\widehat{\psi}_i$, from Table 5.3, is 0.188, meaning that post-randomisation they should expect to have seizures about 19% as often as they had experienced pre-randomisation. Recall that the post-randomisation seizure rate per unit time is given by $\widehat{\lambda}_i\widehat{\psi}_i = 0.0055 \times 0.188 = 0.0010$, which equates to one seizure approximately every 970 days.

Looking at the values of $\widehat{\psi}_i$ presented in Table 5.3, we can see that treatment policy is not statistically significant for those individuals with a normal

EEG. Additionally, those individuals having an abnormal EEG, but allocated to immediate treatment, can expect to have a post-randomisation seizure rate in line with those presenting a normal EEG. We can see that for those with an abnormal EEG, immediate treatment is favoured for all groups except partial, where no significant difference between treatment policies is observed. This is in line with what was suggested by the exploratory analysis.

Seizure Type	$\widehat{\psi_i}$ (95% C.I.)			
	Abnormal EEG			
	Immediate		Deferred	
Tonic-Clonic	0.122	(0.10,0.15)	0.188	(0.15,0.23)
2° Tonic-Clonic	0.127	(0.10,0.16)	0.282	(0.22,0.36)
Partial	0.078	(0.05,0.12)	0.074	(0.05,0.11)
	Normal EEG			
	Immediate		Deferred	
Tonic-Clonic	0.127	(0.10,0.15)	0.134	(0.11,0.16)
2° Tonic-Clonic	0.089	(0.07,0.11)	0.135	(0.11,0.17)
Partial	0.195	(0.12,0.32)	0.127	(0.08,0.21)

Table 5.3: The expected change in seizure rate post-randomisation, using the estimated regression coefficients from Joint Model B.

If we further consider the values of the expected post-randomisation seizure rate reductions, presented in Table 5.3, we observe large reductions in the seizure rates across all groups. Note that the estimates presented suggest that even those individuals with an abnormal EEG and randomised to deferred treatment should expect to see considerable reductions in their seizure rate post-randomisation. It seems unrealistic that an individual receiving no treatment should expect to see such dramatic reductions in their seizure rate post-randomisation.

5.4 Discussion

In this chapter we have built a model that allows pre-randomisation event counts and post-randomisation times to first, and from first to second seizure, to be analysed jointly. This model is proposed as an alternative to standard survival analysis, which treats the pre-randomisation event count information as a covariate, and was based on a joint model developed by Cowling et al. (2006). We have built our joint model under a Poisson process framework, with an assumed underlying individual event rate. This underlying seizure rate is modified at randomisation to allow for the treatment effects. We have assumed that post-randomisation survival times are independent, conditional on the individual-specific random effects, and identically distributed.

The joint model that has been developed in this chapter has subsequently been used to analyse the MESS data. Two versions of the joint model were considered: one that jointly analyses the pre-randomisation event count and time to first seizure post-randomisation only (Joint Model A), and a second that additionally incorporates the times from first to second seizure (Joint Model B). On fitting the joint models, we concluded that the seizure type interactions with treatment and EEG outcome were highly statistically significant.

To gain a better understanding of the regression coefficients presented, subsequent estimates of the pre-randomisation seizure rates and post-randomisation seizure rate modifiers were derived. These estimates revealed possible limitations of the joint model presented in this chapter. First, the magnitudes of

the seizure rate reductions presented in Table 5.3 are a cause for concern. Non-parametric analysis of the data, carried out in Chapter 3, highlighted the possibility of cure rates being prevalent in the dataset, which would explain the unrealistic reductions in seizures rates observed here.

Secondly, we have noted that the estimated regression coefficients observed in Joint Model B are closer to zero than those estimates observed in Joint Model A. This result has suggested that the assumption of a constant ψ_i post-randomisation may be violated.

The joint models for the pre-randomisation and post-randomisation seizure rates, developed in this chapter, seem to provide an improvement over standard survival models. The inclusion of additional information in the joint models has resulted in an increase in power, which consequently means that statistically significant covariate effects, not recognised by the standard survival distributions, have been affirmed.

Previously analyses of the MESS data, by Marson et al. (2005) and Kim et al. (2006), concluded that the risk of seizure recurrence increased with the number of seizures pre-randomisation and an abnormal EEG, and that immediate treatment increased times to first and second seizures. These findings are consistent with our analysis of the MESS data. Nonetheless, neither of these analyses considered differences between types of epileptic seizures, or interactions between the covariates, which we have found to be statistically significant in determining underlying seizure rates and post-randomisation seizure rate reductions.

To summarise, despite an observed improvement, the joint model does not incorporate other characteristics evident within the data, and discussed here. Extensions to the joint model that accommodate these interesting characteristics in the dataset shall be considered in the next chapter.

Chapter 6

Extensions to the Simple Joint Model

As previously discussed, examination of the results obtained following the implementation of Joint Models A and B, has highlighted possible limitations. The assumption of a constant seizure rate post-randomisation may not be accurate, and there is evidence to suggest that cure rates may be present.

In this next section we discuss each of these limitations separately and explore possible solutions to the problems that they present. We shall then proceed to build a model that encompasses all the interesting characteristics present in the data, in one complete model.

6.1 Varying Post-Randomisation Seizure Rate

There is evidence to suggest that the seizure rates may change not only at randomisation, but also following a first seizure post-randomisation. We shall

account for this by first considering the following adjustment:

$$\begin{aligned} X_i \mid \nu_i &\sim \text{Poisson}(\lambda_i u_i \nu_i), \\ Y_{1i} \mid \nu_i &\sim \text{Exponential}(\lambda_i \psi_{1i} \nu_i), \\ Y_{2i} \mid \nu_i &\sim \text{Exponential}(\lambda_i \psi_{1i} \rho \nu_i), \\ \nu_i &\sim \text{Gamma}(\alpha, \alpha), \end{aligned}$$

where $\lambda_i = \exp(\boldsymbol{\beta}_1' \boldsymbol{z}_{1i})$, $\psi_{1i} = \exp(\boldsymbol{\beta}_2' \boldsymbol{z}_{2i})$ and $\boldsymbol{z}_{1i}$, $\boldsymbol{z}_{2i}$ are vectors of covariates, not necessarily distinct. The parameter ρ is a constant, which determines the change in seizure rate following a post-randomisation seizure.

Integrating the random effect term out of the joint density of the survival times, Y_{1i} and Y_{2i}, and ν_i, the unconditional joint distribution of the Y_{ji}, $j = 1, 2$, remains the bivariate Lomax distribution, with univariate Lomax marginal distributions. Proceeding in the same manner as before, considering the different censoring patterns separately, allows us to formulate the log-likelihood for the observed data $\mathcal{D}$, on all the n individuals, obtaining

$$\begin{aligned} \ell(\alpha, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \rho \mid \mathcal{D}) = \sum_{i=1}^{n} \Bigg\{ & \left[\sum_{k=0}^{x_i - 1} \ln(\alpha + k) \right] + \{x_i + \delta_{1i}(1 + \delta_{2i})\} \ln(\lambda_i) \\ & + x_i \ln(u_i) - \ln(x_i!) + \alpha \ln(\alpha) + \delta_{1i}(1 + \delta_{2i}) \ln(\psi_{1i}) + \delta_{1i}\delta_{2i} \ln(\rho) \\ & + \delta_{1i} \ln(x_i + \alpha) + \delta_{1i}\delta_{2i} \ln(x_i + \alpha + 1) \\ & - \{x_i + \alpha + \delta_{1i}(1 + \delta_{2i})\} \ln\left(\lambda_i u_i + \lambda_i \psi_{1i}(-y_{1i} + \rho y_{2i}) + \alpha\right) \Bigg\}. \end{aligned} \tag{6.1}$$

The derivation of this adjusted model allows us to conduct a hypothesis test, comparing Joint Model B with the model that we consider here. The estimate

for ρ was 2.2 (standard error 0.16)[1], with the corresponding log-likelihood ratio test statistic of 2468 providing overwhelming support for its inclusion.

We do not believe that this model is sufficient for modelling the data, we merely use this example to illustrate that the identically distributed assumption for the two survival times is violated. Instead, we now consider a joint model that includes seizure rate modifiers, both at randomisation and following a first post-randomisation seizure, that depend on covariates. This modification is implemented by considering the following adjustment to the joint density for the post-randomisation survival times:

$$f_{Y_1,Y_2|\nu}(y_{1i}, y_{2i} \mid \nu_i; \lambda_i, \psi_{1i}, \psi_{2i}) \quad = \quad (\lambda_i \psi_{1i} \nu_i)^2 \psi_{2i} \exp(-\lambda_i \psi_{1i} \nu_i (y_{1i} + \psi_{2i} y_{2i})),$$

where $\lambda_i = \exp(\boldsymbol{\beta}_1' \boldsymbol{z}_{1i})$, $\psi_{1i} = \exp(\boldsymbol{\beta}_2' \boldsymbol{z}_{2i})$, $\psi_{2i} = \exp(\boldsymbol{\beta}_3' \boldsymbol{z}_{3i})$ and $\boldsymbol{z}_{1i}$, $\boldsymbol{z}_{2i}$, $\boldsymbol{z}_{3i}$ are vectors of covariates, not necessarily distinct.

Table 6.1 shows the subsequent estimated pre-randomisation seizure rates and the expected yearly seizure rates, stratified by seizure type. It is not surprising to observe that the figures presented in Table 6.1 are the same as those presented in Table 5.2. The extensions that we consider in this chapter correspond to interesting characteristics present in the post-randomisation survival times only. There is no reason why, by considering extensions to the simple joint model, we should expect parameter estimates concerning the pre-randomisation event counts to change.

[1]Maximum likelihood estimates of the regression coefficients for the different models considered in this chapter can be found at http://www.warwick.ac.uk/go/jenniferrogers/research/thesis.
This page is password protected, the password is 'thesisrogers2010'.

Seizure Type	$\widehat{\lambda_i}$ (95% C.I.)		Expected yearly rate
Tonic-Clonic	0.0054	(0.005,0.006)	2
2° Tonic-Clonic	0.008	(0.007,0.009)	3
Partial	0.016	(0.013,0.019)	6

Table 6.1: The expected pre-randomisation seizure rate per unit time and the corresponding expected yearly seizure rate, using the estimated regression coefficients from Joint Model B.

Table 6.2 shows the maximum likelihood estimates for ψ_{1i} and ψ_{2i}. The values of ψ_{1i} represent the change in seizure rate at randomisation, with ψ_{2i} representing the change in rate following a first post-randomisation seizure. As in Chapter 5, we see that treatment policy does not appear to be statistically significant in determining the estimate of ψ_{1i} for those individuals with a normal EEG. Additionally, those individuals having an abnormal EEG, but allocated to immediate treatment can expect to have a seizure rate following randomisation in line with those presenting a normal EEG. For those with an abnormal EEG immediate treatment is favoured for all groups except partial, where no statistically significant difference between treatment policies is observed.

Following a first seizure post-randomisation we see that, in general, seizure rates increase. Those with an abnormal EEG, and allocated to deferred treatment, typically have the smallest increase in seizure rate following a first seizure post-randomisation. A possible explanation for this is that these individuals typically see the smallest reduction in seizure rate following randomisation. Recall that those randomised to deferred treatment were simply withheld AEDs until it was deemed absolutely necessary by the clinician. It seems reasonable to suggest that those individuals who were randomised to deferred

treatment, who experience a seizure post-randomisation, may start a subsequent course of treatment with AEDs, thus bringing them in line with those randomised to immediate treatment.

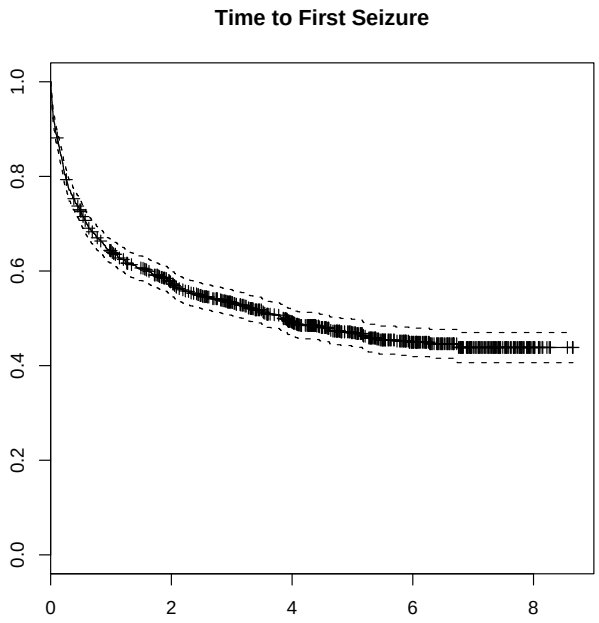
Seizure Type	$\widehat{\psi}_{1i}$ (95% C.I.) first seizure			
	Abnormal EEG			
	Immediate		Deferred	
Tonic-Clonic	0.086	(0.07,0.11)	0.157	(0.12,0.20)
2° Tonic-Clonic	0.109	(0.08,0.14)	0.227	(0.17,0.31)
Partial	0.052	(0.03,0.09)	0.065	(0.04,0.10)
	Normal EEG			
	Immediate		Deferred	
Tonic-Clonic	0.094	(0.08,0.12)	0.107	(0.09,0.13)
2° Tonic-Clonic	0.067	(0.05,0.09)	0.087	(0.07,0.11)
Partial	0.119	(0.06,0.22)	0.093	(0.05,0.17)
	$\widehat{\psi}_{2i}$ (95% C.I.) second seizure			
	Abnormal EEG			
	Immediate		Deferred	
Tonic-Clonic	3.755	(2.48,5.70)	1.413	(0.99,2.01)
2° Tonic-Clonic	1.708	(1.12,2.62)	1.361	(0.89,2.08)
Partial	2.202	(0.99,4.91)	2.134	(1.01,4.53)
	Normal EEG			
	Immediate		Deferred	
Tonic-Clonic	3.116	(2.19,4.43)	1.983	(1.41,2.78)
2° Tonic-Clonic	3.640	(2.31,5.73)	4.902	(3.18,7.56)
Partial	1.814	(0.74,4.46)	2.972	(1.19,7.44)

Table 6.2: The expected change in seizure rates following randomisation and following the first post-randomisation seizure, using the estimated regression coefficients from the joint model that incorporates a varying post-randomisation seizure rate.

6.2 Cure Rate Models

The magnitude of the reductions in seizure rates post-randomisation, observed in Chapter 5 suggest that there may be a substantial proportion of the population that we should regard as cured. This was also discussed in the exploratory analysis, presented in Chapter 3. Berg and Shinnar (1991) noted that, on average, around 50% of people do not experience seizure recurrence after a single, untreated seizure. Recall that over half of the 1425 individuals for whom exploratory analysis was carried out presented only a single seizure pre-randomisation. It is therefore not unreasonable to suspect that a substantial proportion of the individuals included in the MESS trial would never have a seizure post-randomisation, regardless of the length of time for which they were followed. It has already been acknowledged that if survival data does indeed have a proportion that are immune to the event of interest, a model that ignores this may give misleading results. More specifically, ignoring any potential cure fraction could result in underestimates of the post-randomisation seizure rates, thus contributing to the magnitude of seizure rate reductions that have been observed.

Recall that a proper survival distribution should have total mass 1, with the resulting Kaplan-Meier curve having its asymptote at zero. The Kaplan-Meier curves for times to first seizure, and from first to second seizure, are presented in Figure 6.1. Both of the Kaplan-Meier curves level off well above zero, suggesting that there may be an immune component present for both time to first seizure post-randomisation and time from first to second seizure.



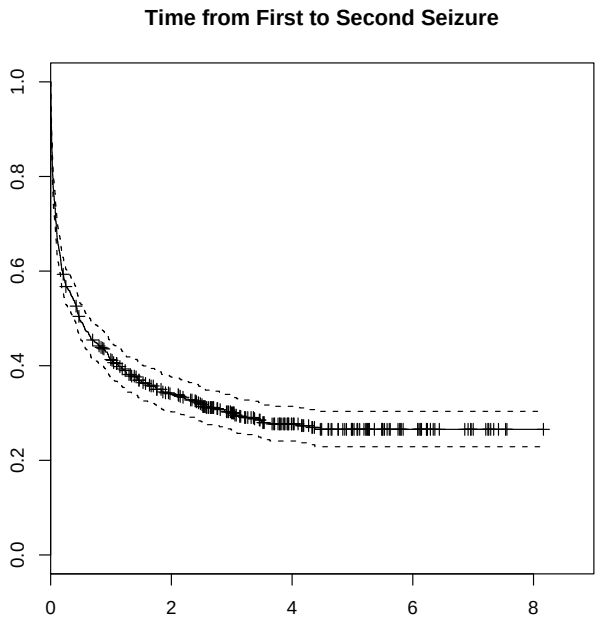


Figure 6.1: Kaplan-Meier curves for time to first and second seizures (with 95% CI). Curves are marked at each censoring time.

We have presented a cure rate model, proposed by Maller and Zhou (1996), which takes into consideration the fact that there may be a proportion of individuals in the population immune to the event of interest. We shall now proceed to consider the analysis of the MESS data using cure rate models.

6.2.1 Standard Survival Analysis with Cure Rates

Standard software in R (Peng) allows the fitting of various parametric mixture models, including the Log-logistic, for the estimation of cure rates. We therefore proceed to fit a Log-logistic mixture model that incorporates a cure fraction. This model was considered for both times to first seizure post-randomisation, and times from first to second seizure.

We initially consider mixture models that allow the associated cure rates to depend on an intercept term only. The derivation of these adjusted models allows us to conduct a hypothesis test, comparing the Log-logistic cure rate models with the standard Log-logistic models presented in Chapter 4. The estimated cure fractions are 23.9% for time to first seizure, and 19.3% for time from first to second seizure. The corresponding log-likelihood ratio test statistics are 6236 and 3268 for the models for time to first seizure and time from first to second seizure respectively. These two highly statistically significant test statistics provide overwhelming evidence for the inclusion of cure rates, when analysing the survival data. These estimated values, however, are slightly lower than suggested by the Kaplan-Meier curves in Figure 6.1, but still provide sufficient evidence to suggest that we should incorporate a cure fraction into the joint model.

Further investigation concluded that the cure rates associated with times to first seizure post-randomisation are dependent on seizure type, EEG outcome and the logarithm of the pre-randomisation seizure rate for each individual. For the times from first to second post-randomisation, no covariate effects were found to be statistically significant in determining the proportion cured.

6.2.2 Cure Rate for Single Post-Randomisation Survival Time

We can adjust our original, simple joint model, to allow for the inclusion of cure rates. We first consider a model that jointly models the pre-randomisation seizure counts and post-randomisation time to first seizure only, which has the following density and survivor functions:

$$\begin{aligned} f_{Y_1|\nu}(y_{1i} \mid \nu_i; \lambda_i, \psi_i, p) &= p g_{Y_1|\nu}(y_{1i} \mid \nu_i; \lambda_i, \psi_i) \\ &= p\lambda_i\psi_i\nu_i \exp(-\lambda_i\psi_i\nu_i y_{1i}), \\ S_{Y_1|\nu}(y_{1i} \mid \nu_i; \lambda_i, \psi_i, p) &= 1 - p + pR_{Y_1|\nu}(y_{1i} \mid \nu_i; \lambda_i, \psi_i) \\ &= 1 - p + p\exp(-\lambda_i\psi_i\nu_i y_{1i}), \end{aligned}$$

where $\lambda_i = \exp(\boldsymbol{\beta}_1' \boldsymbol{z}_{1i})$, $\psi_i = \exp(\boldsymbol{\beta}_2' \boldsymbol{z}_{2i})$ and $\boldsymbol{z}_{1i}$, $\boldsymbol{z}_{2i}$ are vectors of covariates, not necessarily distinct. The term p represents the susceptible proportion in the population, so that $1 - p$ is the cure fraction. The density and survivor functions for the susceptibles are given by $g(\cdot)$ and $R(\cdot)$ respectively.

On fitting the joint model allowing for the cure rate, we obtain a maximum likelihood estimate for p of 0.574 (standard error 0.02). Comparing this esti-

mate with the Kaplan-Meier curve for time to first seizure in Figure 6.1, this estimate seems sensible. We can formally test for the inclusion of p by comparing the log-likelihood of Joint Model A with the log-likelihood obtained here. A highly statistically significant likelihood-ratio test statistic of 442 supports the inclusion of the cure rate, and the corresponding estimated value suggests that there is a substantial proportion of the population 'immune' to seizures post-randomisation.

6.2.3 Allowing the Cure Rates to Depend on Covariates

We can allow the cure rate to depend on individuals' covariates by considering the following parameterisation:

$$p_i = \frac{\exp(\boldsymbol{\kappa}'\boldsymbol{w}_i)}{1+\exp(\boldsymbol{\kappa}'\boldsymbol{w}_i)}.$$

The explanatory variables associated with individual i are entered into the covariate $\boldsymbol{w}_i$, with $\boldsymbol{\kappa}$ denoting the corresponding vector of regression coefficients. Estimates of $\widehat{\lambda}_i$ will remain the same as those presented in Tables 5.2 and 6.1. Subsequent parameter estimates can be used to obtain estimates of the post-randomisation seizure rate modifiers and cure rates, denoted by $\widehat{\psi}_i$ and $1-\widehat{p}_i$ respectively. Table 6.3 shows the estimated cure rates, whilst Table 6.4 presents the estimated seizure rate modifiers for the subsequent susceptible proportion. The results in Tables 6.3 and 6.4 allow us to estimate an individuals' probability of being immune to seizure recurrence, dependent on their pre-randomisation seizure types EEG outcome and their associated interaction terms. If an individual is susceptible to seizure recurrence, we can predict the magnitude of their seizure rate reductions post-randomisation, dependent

on their pre-randomisation seizure types, EEG outcome and, additionally, the treatment policy to which they are assigned.

Treatment policy was not found to be significant in determining whether an individual is susceptible or immune to seizure recurrence post-randomisation. Model selection was carried out using stepwise backwards elimination (Hocking 1976). Examination of Table 6.3 highlights that those individuals with an abnormal EEG can expect to have a statistically significantly lower cure rate than those with a normal EEG, except for those individuals experiencing Partial seizures only pre-randomisation, where no statistically significant difference is observed.

Seizure Type	$1-\widehat{p}_i$ (95% C.I.)			
	Abnormal EEG		Normal EEG	
Tonic-Clonic	0.352	(0.28,0.43)	0.495	(0.44,0.55)
2° Tonic-Clonic	0.235	(0.16,0.33)	0.549	(0.48,0.62)
Partial	0.391	(0.27,0.53)	0.280	(0.14,0.48)

Table 6.3: The expected cure rates when considering the times to first seizure, using estimates from the joint model that incorporates the post-randomisation times to first seizure only, with associated cure rates.

Looking at the estimated values of $\widehat{\psi}_i$ presented in Table 6.4, we can again see that treatment policy does not appear to be significant for those individuals with a normal EEG. Those individuals having an abnormal EEG, but allocated to immediate treatment, still expect to have a post-randomisation seizure rate in line with those presenting a normal EEG. We can see that for those with an abnormal EEG immediate treatment is favoured for all groups except partial, where no significant difference between treatment policies is observed.

Finally, we observe that the magnitudes of the seizure rate reductions post-randomisation are not as large as those presented in Table 5.3. Additionally, note that those with Partial seizures, with or without secondary Tonic-Clonic seizures, with an abnormal EEG, and allocated to deferred treatment have a value of $\widehat{\psi}_i$ not statistically significantly different from unity, which corresponds to the seizure rate not changing post-randomisation.

Seizure Type	$\widehat{\psi}_i$ (95% C.I.)			
	Abnormal EEG			
	Immediate		Deferred	
Tonic-Clonic	0.189	(0.12,0.30)	0.638	(0.48,0.86)
2° Tonic-Clonic	0.186	(0.13,0.27)	0.761	(0.55,1.06)
Partial	0.431	(0.23,0.79)	0.683	(0.40,1.18)
	Normal EEG			
	Immediate		Deferred	
Tonic-Clonic	0.487	(0.36,0.66)	0.626	(0.47,0.84)
2° Tonic-Clonic	0.360	(0.23,0.55)	0.559	(0.40,0.79)
Partial	0.461	(0.23,0.92)	0.278	(0.13,0.59)

Table 6.4: The expected change in seizure rates following randomisation for the susceptible proportion, using estimates from the joint model that incorporates the post-randomisation times to first seizure only, with associated cure rates.

6.2.4 Cure Rates for Both Post-Randomisation Survival Times

We now develop a model that considers the pre-randomisation event counts and both of the post-randomisation survival times, allowing each of the survival times to have an associated cure rate.

In this scenario, censored times to first seizure post-randomisation may be censored because individuals are immune to seizure recurrence or, alternatively, their period of follow up may not be long enough to observe any seizures post-randomisation, that is, the observation time may be less than their actual, unobserved time to first seizure.

These two distinct events have an effect on the subsequent post-randomisation time from first to second seizure. If an individual is immune to seizures post-randomisation, then their time from first to second seizure simply does not exist. Alternatively, if the time to first seizure exists but we do not observe it, we know that an individual could either have a second post-randomisation seizure (which we again do not observe), or they could, despite having a first seizure, be immune to further seizures.

We introduce, for each individual, an allocation variable, q_i, which is an indicator function taking the value 1 if the individual is susceptible to post-randomisation seizure recurrence, and zero if the individual is immune.

Considering the different censoring patterns separately allows us to formulate the log-likelihood for the observed data $\mathcal{D}$, on all the n individuals, obtaining

$$
\begin{aligned}
\ell(\alpha, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\kappa}_1, \boldsymbol{\kappa}_2 \mid \mathcal{D}) = \sum_{i=1}^{n} \Bigg\{ & \left[\sum_{k=0}^{x_i-1} \ln(\alpha + k) \right] + x_i \ln(u_i) - \ln(x_i!) + \alpha \ln(\alpha) \\
& + (x_i + \delta_{1i}(1 + \delta_{2i})) \ln(\lambda_i) + \delta_{1i}(1 + \delta_{2i}) \ln(\psi_i) + \delta_{1i} \ln(p_{1i}) \\
& + \delta_{1i}\delta_{2i} \ln(p_{2i}) + \delta_{1i} \ln(x_i + \alpha) + \delta_{1i}\delta_{2i} \ln(x_i + \alpha + 1) \\
& - \delta_{1i}\delta_{2i}(x_i + \alpha + 2) \ln(\lambda_i u_i + \lambda_i \psi_i (y_{1i} + y_{2i}) + \alpha) \\
& + (1 - \delta_{1i}) q_i \ln \left(\frac{p_{1i}}{(\lambda_i u_i + \lambda_i \psi_i y_{1i} + \alpha)^{x_i + \alpha}} \right) \\
& + (1 - \delta_{1i})(1 - q_i) \ln \left(\frac{1 - p_{1i}}{(\lambda_i u_i + \alpha)^{x_i + \alpha}} \right) \Bigg\} \\
& + \delta_{1i}(1 - \delta_{2i}) \ln \left(\frac{1 - p_{2i}}{(\lambda_i u_i + \lambda_i \psi_i y_{1i} + \alpha)^{x_i + \alpha + 1}} \right. \\
& \left. + \frac{p_{2i}}{(\lambda_i u_i + \lambda_i \psi_i (y_{1i} + y_{2i}) + \alpha)^{x_i + \alpha + 1}} \right),
\end{aligned} \tag{6.2}
$$

where $\lambda_i = \exp(\boldsymbol{\beta}_1' \boldsymbol{z}_{1i})$, $\psi_i = \exp(\boldsymbol{\beta}_2' \boldsymbol{z}_{2i})$ and $\boldsymbol{z}_{1i}$, $\boldsymbol{z}_{2i}$ are vectors of covariates, not necessarily distinct. In addition,

$$
\begin{aligned}
p_{1i} &= \frac{\exp(\boldsymbol{\kappa}_1' \boldsymbol{w}_{1i})}{1 + \exp(\boldsymbol{\kappa}_1' \boldsymbol{w}_{1i})}, \\
p_{2i} &= \frac{\exp(\boldsymbol{\kappa}_2' \boldsymbol{w}_{2i})}{1 + \exp(\boldsymbol{\kappa}_2' \boldsymbol{w}_{2i})}.
\end{aligned}
$$

In practice, we do not observe the allocation variable, q_i, for each individual, instead we can regard it as a missing value. We must now proceed to perform likelihood inference in the presence of missing data. Little and Rubin (2002) note that, in a formal sense, there is no difference between maximum likelihood for incomplete data and maximum likelihood for complete data: the likelihood for the parameters, based on the incomplete data, is derived and the maximum likelihood estimates are then found by solving the likelihood equation.

Recall that the full model treats the missing data indicator as a random variable, and considers the joint distribution of M and $Y = (Y_{obs}, Y_{mis})$. We have already discussed how the full probability density function can be re-written as the product of the probability density function of Y, and the conditional distribution of M given Y. This conditional probability density function, $f(M \mid Y_{obs}, Y_{mis}; \varphi)$, is indexed by the unknown parameter φ, and is the distribution of the missing-data mechanism. Recall that the actual observed data are (Y_{obs}, M), and the distribution of the observed data is obtained through integrating Y_{mis} out of the the joint density of Y and M (Equation 2.9). The full likelihood of θ and φ is then any function of θ and φ proportional to the joint probability distribution of Y_{obs} and M:

$$L_{full}(\theta, \varphi \mid Y_{obs}, M) \propto f(Y_{obs}, M; \theta, \varphi), \quad (\theta, \varphi) \in \Omega_{\theta,\varphi}. \tag{6.3}$$

We can similarly let the likelihood of θ, based on Y_{obs}, ignoring the missing-data mechanism, be any function of θ proportional to $f(Y_{obs}; \theta)$:

$$L_{ign}(\theta \mid Y_{obs}) \propto f(Y_{obs}; \theta), \quad \theta \in \Omega_{\theta}. \tag{6.4}$$

The subsequent question of importance is when inference for θ should be based on the full likelihood in Equation (6.3), and when it can be based on the simpler, ignorable likelihood, given by Equation (6.4). The missing-data mechanism is ignorable for likelihood inference if the data are MCAR, or MAR, and the parameters θ and φ are distinct, in the sense that the joint parameter space of (θ, φ) is the product of the parameter space of θ and the parameter space of φ.

A common likelihood based technique, that is adopted when presented with mixture models of the type presented in Equation 6.2, is the EM-algorithm.

The EM Algorithm

The EM algorithm is a general iterative algorithm for maximum likelihood estimation, in the presence of missing data. This technique is based on a somewhat ad-hoc idea of: (1) Replacing missing values by estimated values, (2) Estimating parameters, (3) Re-estimating the missing values according to the new parameter estimates, (4) Re-estimating the parameters, iterating until convergence.

Suppose that the complete data are Y, with associated density $f(Y;\theta)$. We write $Y = (Y_{obs}, Y_{mis})$, where Y_{obs} represents the observed part of Y, and Y_{mis} denotes the missing part. The objective is to maximise the ignorable likelihood,

$$L_{ign}(\theta \mid Y_{obs}) = \int f(Y_{obs}, Y_{mis}; \theta) dY_{mis},$$

with respect to θ.

The EM algorithm comprises an E-step and an M-step. The E-step finds the conditional expectation of the 'missing data', given the observed data and the current estimated parameters. These expectations are then substitutes for the 'missing data'. We write 'missing data' in quotation marks because, in reality, the EM algorithm does not substitute the missing values themselves, but regards the functions of Y_{mis} appearing in the complete-data log-likelihood, $\ell(\theta \mid Y)$, as missing. The subsequent M-step simply comprises finding the

maximum likelihood estimates of θ, as if there were no missing data. The M-step uses exactly the same computational methods as maximum likelihood estimation from $\ell(\theta \mid Y)$.

The algorithm is formally defined as:

1. Choose initial value $\theta^{(0)}$, set t=0.

2. E-step: calculate

$$\begin{aligned} Q(\theta, \theta^{(t)}) &= \mathbb{E}[\ln(L(\theta \mid Y_{obs}, Y_{mis}))] \\ &= \int \ell(\theta \mid Y) f(Y_{mis} \mid Y_{obs}, \theta = \theta^{(t)}) dY_{mis} \end{aligned} \tag{6.5}$$

   (at this stage $\theta^{(t)}$ is fixed and $Q(\theta, \theta^{(t)})$ is a function of θ).

3. M-step: find $\theta^{(t+1)}$ which maximises $Q(\theta, \theta^{(t)})$ as a function of θ.

4. Set $t = t + 1$ and go to step 2.

Essentially, each iteration of the EM algorithm updates $\theta^{(t)}$ to $\theta^{(t+1)}$ by solving the following equation:

$$\frac{\partial}{\partial \theta} \mathbb{E}[\ln(L(\theta \mid Y_{obs}, Y_{mis}))] = 0,$$

where $\theta^{(t)}$ is used to find $\mathbb{E}[\ln(L(\theta \mid Y_{obs}, Y_{mis}))]$, and $\theta^{(t+1)}$ is the solution to the equation. The main feature of the algorithm is that

$$L(\theta^{(t)} \mid Y_{obs}) \leq L(\theta^{(t+1)} \mid Y_{obs}).$$

That is, the likelihood function is increased at each iteration of the algorithm and consequently, the sequence, $\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, \ldots$, is guaranteed to converge to the location of a local maximum of the likelihood.

The complete-data log-likelihood is given by Equation (6.2). The parameters to be maximised are $\theta = \{\alpha, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\kappa}_1, \boldsymbol{\kappa}_2\}$. The observed data comprises $Y_{obs} = \{X, Y_1, Y_2, Z_1, Z_2, W_1, W_2, \delta_1, \delta_2\}$, and the missing data is $Y_{mis} = \{Q\}$. To implement the EM-algorithm, using the iterative scheme outlined, we first carry out the E-step:

$$\begin{aligned} Q(\theta, \theta^{(t)}) = \sum_{i=1}^{n} \Bigg\{ \Bigg[\sum_{k=0}^{x_i-1} \ln(\alpha + k) \Bigg] + x_i \ln(u_i) - \ln(x_i!) + \alpha \ln(\alpha) \\ +(x_i + \delta_{1i}(1 + \delta_{2i})) \ln(\lambda_i) + \delta_{1i}(1 + \delta_{2i}) \ln(\psi_i) + \delta_{1i} \ln(p_{1i}) \\ +\delta_{1i}\delta_{2i} \ln(p_{2i}) + \delta_{1i} \ln(x_i + \alpha) + \delta_{1i}\delta_{2i} \ln(x_i + \alpha + 1) \\ -\delta_{1i}\delta_{2i}(x_i + \alpha + 2) \ln(\lambda_i u_i + \lambda_i \psi_i (y_{1i} + y_{2i}) + \alpha) \\ +(1 - \delta_{1i}) \mathbb{E}(q_i \mid Y_{obs}, \theta^{(t)}) \ln \left(\frac{p_{1i}}{(\lambda_i u_i + \lambda_i \psi_i y_{1i} + \alpha)^{x_i + \alpha}} \right) \\ +(1 - \delta_{1i})(1 - \mathbb{E}(q_i \mid Y_{obs}, \theta^{(t)})) \ln \left(\frac{1 - p_{1i}}{(\lambda_i u_i + \alpha)^{x_i + \alpha}} \right) \\ +\delta_{1i}(1 - \delta_{2i}) \ln \Bigg(\frac{1 - p_{2i}}{(\lambda_i u_i + \lambda_i \psi_i y_{1i} + \alpha)^{x_i + \alpha + 1}} \\ + \frac{p_{2i}}{(\lambda_i u_i + \lambda_i \psi_i (y_{1i} + y_{2i}) + \alpha)^{x_i + \alpha + 1}} \Bigg) \Bigg\}, \end{aligned} \tag{6.6}$$

where $\mathbb{E}(q_i \mid Y_{obs}, \theta^{(t)})$ is the expected value of q_i, given the observed data and the current values of the parameters of interest. Completion of the E-step requires the derivation of an expression for this expectation. Clearly, as q_i is an indicator variable, we have $\mathbb{E}(q_i \mid Y_{obs}, \theta^{(t)}) = \mathbb{P}(q_i = 1 \mid Y_{obs}, \theta^{(t)})$. Recall that $q_i = 1$ corresponds to an individual having a post-randomisation time to

first seizure that is censored, but exists. We can now formulate an expression for $\mathbb{E}(q_i \mid Y_{obs}, \theta^{(t)})$ as so:

$$\begin{aligned}\mathbb{E}(q_i \mid Y_{obs}, \theta^{(t)}) &= \mathbb{P}(q_i = 1 \mid Y_{obs}, \theta^{(t)}) \\ &= p_{1i} R_{Y_1}(y_{1i}; \lambda_i, \psi_{1i}, \alpha) \\ &= p_{1i}\left(1 + \frac{\lambda_i \psi_{1i} y_{1i}}{\alpha}\right)^{-\alpha}.\end{aligned}$$

The subsequent first derivatives of Equation 6.6 are simple to obtain, allowing the M-step, derivation of the maximum likelihood estimates of the parameters contained in θ, to be carried out, using numerical methods.

We consider two versions of this model, the first simply allows the cure rates to depend on intercept terms, so that we consider only the overall cure rates, across all individuals, for the times to first seizure, and from first to second seizure post-randomisation. Following this, we shall allow the cure rates to depend on statistically significant covariates, using stepwise backwards elimination to evaluate the optimum model.

On fitting the joint model that considers only the overall cure rates, across all individuals, for each of the post-randomisation survival times, we obtain maximum likelihood estimates for p_1 and p_2 of 0.478 (standard error 0.06) and 0.235 (standard error 0.11) respectively. Comparing both these estimates with the Kaplan-Meier curves for time to first seizure, these estimates seem sensible.

The formulation of this joint model allows us to formally test for the inclusion of p_1 and p_2, comparing the subsequent log-likelihood with that of Joint

Model B. A highly statistically significant likelihood-ratio test statistic of 649 supports the inclusion of both cure rates. The maximum likelihood estimates of p_1 and p_2 suggest that there is a substantial proportion of the population 'immune' to post-randomisation seizure recurrence, and a significant proportion who experience a single seizure post-randomisation 'immune' to further seizure recurrence. Also note that $\widehat{p}_1 > \widehat{p}_2$, which supports the conjecture that seizures beget seizures, that is, the more seizures an individual has, the more likely they are to continue experiencing seizures in the future.

The optimal joint model that considers the pre-randomisation event counts and both post-randomisation survival times, with cure rates, was derived using stepwise backwards elimination. This model concluded that the parameters statistically significant in determining the cure rates associated with the times to first seizure were again seizure type, EEG outcome and their associated interaction terms. Additionally, the optimal model found that treatment, and its interaction with EEG outcome, were statistically significant in determining the probability of being 'immune' to a first seizure post-randomisation. There were no covariates that were found to be statistically significant in determining the cure rates associated with the times from first to second seizure.

Tables 6.5 and 6.6 show the estimated cure rates for times to first seizure, and the subsequent estimated seizure rate modifiers for the susceptible proportion respectively. Examination of Table 6.5 highlights that those individuals allocated to deferred treatment, with an abnormal EEG, can expect to have statistically significantly lower cure rates than those with a normal EEG, except for those individuals experiencing partial seizures only pre-randomisation,

where no statistically significant difference is observed. For those with a normal EEG, no statistically significant treatment effect on the cure rates is observed.

Seizure Type	$1-\widehat{p}_{1i}$ (95% C.I.)			
	Abnormal EEG			
	Immediate		Deferred	
Tonic-Clonic	0.537	(0.46,0.61)	0.365	(0.30,0.43)
2° Tonic-Clonic	0.398	(0.32,0.48)	0.247	(0.18,0.32)
Partial	0.500	(0.37,0.63)	0.331	(0.22,0.49)
	Normal EEG			
	Immediate		Deferred	
Tonic-Clonic	0.533	(0.47,0.59)	0.514	(0.46,0.57)
2° Tonic-Clonic	0.593	(0.52,0.66)	0.574	(0.50,0.64)
Partial	0.361	(0.22,0.54)	0.344	(0.20,0.52)

Table 6.5: The expected cure rates for the times to first seizure, using the estimated regression coefficients from the joint model that incorporates the post-randomisation times to first seizure and from first to second seizure, with cure rates.

Recall that for times from first to second seizure, the associated cure rate depended on an intercept term only. The overall estimated cure rate was 24%.

Looking at the values of $\widehat{\psi}_i$ presented in Table 6.6, we again see that treatment policy is not statistically significant for those individuals with a normal EEG, and once again, those individuals having an abnormal EEG, but allocated to immediate treatment have a post-randomisation seizure rate in line with those presenting a normal EEG. We can see that for those with an abnormal EEG, immediate treatment is favoured. We again observe that the magnitudes of the seizure rate reductions post-randomisation are not as large as those presented in Table 5.3. Additionally, note that a number of groups have a value of $\widehat{\psi}_i$ not statistically significantly different from unity, which corresponds to

the seizure rate not changing post-randomisation.

Seizure Type	$\widehat{\psi_i}$ (95% C.I.)			
	Abnormal EEG			
	Immediate		Deferred	
Tonic-Clonic	0.536	(0.41,0.69)	0.832	(0.66,1.04)
2° Tonic-Clonic	0.386	(0.30,0.50)	0.735	(0.58,0.94)
Partial	0.726	(0.47,1.13)	0.887	(0.58,1.35)
	Normal EEG			
	Immediate		Deferred	
Tonic-Clonic	0.701	(0.56,0.88)	0.746	(0.60,0.92)
2° Tonic-Clonic	0.663	(0.50,0.88)	0.866	(0.68,1.11)
Partial	0.743	(0.43,1.27)	0.622	(0.36,1.07)

Table 6.6: The expected change in seizure rates following randomisation for the susceptible proportion, using the estimated regression coefficients from the joint model that incorporates the post-randomisation times to first seizure and from first to second seizure, with cure rates.

6.3 Building a Full Model that Incorporates Varying Post-Randomisation Seizure Rates and Cure Rates

We now proceed to develop a model that jointly models the pre-randomisation event counts and post-randomisation times to first seizure, and from first to second seizure, incorporating both of the extensions that have already been considered in isolation. That is, we shall develop a joint model that allows the seizure rate to vary post-randomisation and incorporates cure rates for each of the survival times.

We assume that the pre-randomisation event count, for individual i, over period u_i, X_i, follows a Poisson distribution with mean and variance $\lambda_i u_i \nu_i$, as in the initial simple joint model. The probability density function for the pre-randomisation event counts is given by

$$f_{X|\nu}(x_i \mid \nu_i; \lambda_i, u_i) = \frac{(\lambda_i u_i \nu_i)^{x_i} \exp(-\lambda_i u_i \nu_i)}{x_i!}.$$

The parameter λ_i relates to the baseline covariates, with additional heterogeneity in the population being modelled through ν_i, again assumed to follow a Gamma(α, α) distribution.

The probability density functions and survivor functions for the times to first seizure, and from first to second seizure post-randomisation, are specified by the following equations:

$$\begin{aligned}
f_{Y_1|\nu}(y_{1i} \mid \nu_i; \lambda_i, \psi_{1i}, p_{1i}) &= p_{1i} g_{Y_1|\nu}(y_{1i} \mid \nu_i; \lambda_i, \psi_{1i}) \\
&= p_{1i} \lambda_i \psi_{1i} \nu_i \exp(-\lambda_i \psi_{1i} \nu_i y_{1i}), \\
S_{Y_1|\nu}(y_{1i} \mid \nu_i; \lambda_i, \psi_{1i}, p_{1i}) &= 1 - p_{1i} + p_{1i} R_{Y_1|\nu}(y_{1i} \mid \nu_i; \lambda_i, \psi_{1i}) \\
&= 1 - p_{1i} + p_{1i} \exp(-\lambda_i \psi_{1i} \nu_i y_{1i}), \\
f_{Y_2|\nu}(y_{2i} \mid \nu_i; \lambda_i, \psi_{1i}, \psi_{2i}, p_{2i}) &= p_{2i} g_{Y_2|\nu}(y_{2i} \mid \nu_i; \lambda_i, \psi_{1i}, \psi_{2i}) \\
&= p_{2i} \lambda_i \psi_{1i} \psi_{2i} \nu_i \exp(-\lambda_i \psi_{1i} \psi_{2i} \nu_i y_{1i}), \\
S_{Y_2|\nu}(y_{2i} \mid \nu_i; \lambda_i, \psi_{1i}, \psi_{2i}, p_{2i}) &= 1 - p_{2i} + p_{2i} R_{Y_2|\nu}(y_{2i} \mid \nu_i; \lambda_i, \psi_{1i}, \psi_{2i}) \\
&= 1 - p_{2i} + p_{2i} \exp(-\lambda_i \psi_{1i} \psi_{2i} \nu_i y_{2i}).
\end{aligned}$$

The density and survivor functions for the susceptibles are given by $g(\cdot)$ and $R(\cdot)$ respectively. The term p_{ji} represents the probability that individual i is susceptible to post-randomisation seizure j, so that $1-p_{ji}$ is the cure fraction. The following parameterisations for p_{1i} and p_{2i} are considered:

$$\begin{aligned} p_{1i} &= \frac{\exp(\boldsymbol{\kappa}_1'\boldsymbol{w}_{1i})}{1+\exp(\boldsymbol{\kappa}_1'\boldsymbol{w}_{1i})}, \\ p_{2i} &= \frac{\exp(\boldsymbol{\kappa}_2'\boldsymbol{w}_{2i})}{1+\exp(\boldsymbol{\kappa}_2'\boldsymbol{w}_{2i})}. \end{aligned}$$

Additionally, $\lambda_i = \exp(\boldsymbol{\beta}_1'\boldsymbol{z}_{1i})$, $\psi_{1i} = \exp(\boldsymbol{\beta}_2'\boldsymbol{z}_{2i})$, $\psi_{2i} = \exp(\boldsymbol{\beta}_3'\boldsymbol{z}_{3i})$ and $\boldsymbol{z}_{1i}$, $\boldsymbol{z}_{2i}$, $\boldsymbol{z}_{3i}$, $\boldsymbol{w}_{1i}$, $\boldsymbol{w}_{2i}$ are vectors of covariates, not necessarily distinct.

6.3.1 Marginal Distributions

If the random effect term is integrated out of the joint density of X_i and ν_i, then the resulting unconditional density, $f_X(x_i;\lambda_i,u_i,\alpha)$, is, as before, the Negative Binomial (Equation 4.1). The unconditional joint distribution of the Y_{ji}, $j=1,2$, for the susceptible proportion, obtained when the random effects are integrated out of $g_{Y_1,Y_2|\nu}(y_{1i},y_{2i} \mid \nu_i;\lambda_i,\psi_{1i},\psi_{2i})$, the joint density of the survival times, Y_{1i} and Y_{2i}, and ν_i, is the bivariate Lomax distribution.

The marginal distribution of Y_{1i}, for those that are susceptible, is the univariate Lomax distribution with shape and scale parameters α and $\alpha/\lambda_i\psi_{1i}$ respectively. The marginal distribution of Y_{2i}, for those that are susceptible, is also the univariate Lomax distribution, but now with shape and scale parameters α and $\alpha/\lambda_i\psi_{1i}\psi_{2i}$ respectively.

6.3.2 The Full Log-Likelihood and Derivatives

When formulating the likelihood we need to consider the different ways that censoring can occur. There are four different ways censoring can arise in this setting, namely: (i) Y_{1i} and Y_{2i} are both observed, (ii) Y_{1i} is observed, but Y_{2i} is censored, (iii) Y_{1i} is censored, but exists, so Y_{2i} is taken to be censored at zero, and (iv) Y_{1i} is censored, and cured, so Y_{2i} does not exist. We now consider these four situations separately:

Joint Distribution with Y_{1i} and Y_{2i} Observed

In this situation the joint density of Y_{1i} and Y_{2i} contributes towards the likelihood, giving

$$\int_0^\infty f_{X|\nu}(x_i \mid \nu_i; \lambda_i, u_i) p_{1i} p_{2i} g_{Y_1,Y_2|\nu}(y_{1i}, y_{2i} \mid \nu_i; \lambda_i, \psi_{1i}, \psi_{2i}) f_\nu(\nu_i; \alpha) d\nu_i$$

$$= \frac{(\lambda_i u_i)^{x_i}}{x_i!} \frac{p_{1i} p_{2i} \psi_{2i} (\lambda_i \psi_{1i})^2 \alpha^\alpha \Gamma(x_i + \alpha + 2)}{\Gamma(\alpha)(\lambda_i u_i + \lambda_i \psi_{1i}(y_{1i} + \psi_{2i} y_{2i}) + \alpha)^{x_i + \alpha + 2}}. \tag{6.7}$$

Joint Distribution with Y_{1i} Observed and Y_{2i} Censored

In this situation the density function of Y_{1i} and survivor functions of Y_{2i} contribute to the likelihood, giving

$$\int_0^\infty f_{X|\nu}(x_i \mid \nu_i; \lambda_i, u_i) p_{1i} g_{Y_1|\nu}(y_{1i} \mid \nu_i; \lambda_i, \psi_{1i})(1 - p_{2i} + p_{2i} R_{Y_2|\nu}(y_{2i} \mid \nu_i; \lambda_i, \psi_{1i}, \psi_{2i})) f_\nu(\nu_i; \alpha) d\nu_i$$

$$= \frac{(\lambda_i u_i)^{x_i}}{x_i!} \frac{p_{1i} p_{2i} \lambda_i \psi_{1i} \alpha^\alpha \Gamma(x_i + \alpha + 1)}{\Gamma(\alpha)} \left(\frac{1 - p_{2i}}{(\lambda_i u_i + \lambda_i \psi_{1i} y_{1i} + \alpha)^{x_i + \alpha + 1}} + \frac{p_{2i}}{(\lambda_i u_i + \lambda_i \psi_{1i}(y_{1i} + \psi_{2i} y_{2i}) + \alpha)^{x_i + \alpha + 1}} \right). \tag{6.8}$$

Joint Distribution with Y_{1i} Censored but Exists

In this scenario only the survivor function of Y_{1i} contributes towards the likelihood, giving

$$\int_0^\infty f_{X|\nu}(x_i \mid \nu_i; \lambda_i, u_i) p_{1i} R_{Y_1|\nu}(y_{1i} \mid \nu_i; \lambda_i, \psi_{1i}) f_\nu(\nu_i; \alpha) d\nu_i$$

$$= \frac{(\lambda_i u_i)^{x_i}}{x_i!} \frac{p_{1i} \alpha^\alpha}{\Gamma(\alpha)} \frac{\Gamma(x_i + \alpha)}{(\lambda_i u_i + \lambda_i \psi_{1i} y_{1i} + \alpha)^{x_i + \alpha}}. \tag{6.9}$$

Joint Distribution with Y_{1i} Censored and Cured

In this scenario we only have the pre-randomisation event counts and random effects for each individual, with the cure rate associated with the time to first seizure post-randomisation:

$$\int_0^\infty f_{X|\nu}(x_i \mid \nu_i; \lambda_i, u_i)(1 - p_{1i}) f_\nu(\nu_i; \alpha) d\nu_i$$

$$= \frac{(\lambda_i u_i)^{x_i}}{x_i!} \frac{(1 - p_{1i})\alpha^\alpha}{\Gamma(\alpha)} \frac{\Gamma(x_i + \alpha)}{(\lambda_i u_i + \alpha)^{x_i + \alpha}}. \tag{6.10}$$

Log-likelihood

Let δ_{ji} be the indicator function for the jth survival time, taking the value 1 if the seizure is observed, and zero if the survival time is censored. Additionally, we consider an allocation variable, q_i, which is an indicator function taking the value 1 if the individual is susceptible to post-randomisation seizure recurrence, and zero if the individual is immune. Combining these indicator functions with equations (6.7)-(6.10) allows us to formulate the log-likelihood for the observed data $\mathcal{D}$, for all the n individuals, given by

$$
\begin{aligned}
\ell(\alpha, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\kappa}_1, \boldsymbol{\kappa}_2 \mid \mathcal{D}) = \sum_{i=1}^{n} \Bigg\{ & \left[\sum_{k=0}^{x_i - 1} \ln(\alpha + k) \right] + x_i \ln(u_i) - \ln(x_i!) \\
& + \alpha \ln(\alpha) + (x_i + \delta_{1i}(1 + \delta_{2i})) \ln(\lambda_i) + \delta_{1i}(1 + \delta_{2i}) \ln(\psi_{1i}) + \delta_{1i}\delta_{2i} \ln(\psi_{2i}) \\
& + \delta_{1i} \ln(p_{1i}) + \delta_{1i}\delta_{2i} \ln(p_{2i}) + \delta_{1i} \ln(x_i + \alpha) + \delta_{1i}\delta_{2i} \ln(x_i + \alpha + 1) \\
& - \delta_{1i}\delta_{2i}(x_i + \alpha + 2) \ln(\lambda_i u_i + \lambda_i \psi_{1i}(y_{1i} + \psi_{2i} y_{2i}) + \alpha) \\
& + (1 - \delta_{1i}) q_i \ln \left(\frac{p_{1i}}{(\lambda_i u_i + \lambda_i \psi_{1i} y_{1i} + \alpha)^{x_i + \alpha}} \right) \\
& + (1 - \delta_{1i})(1 - q_i) \ln \left(\frac{1 - p_{1i}}{(\lambda_i u_i + \alpha)^{x_i + \alpha}} \right) \\
& + \delta_{1i}(1 - \delta_{2i}) \ln \Bigg(\frac{1 - p_{2i}}{(\lambda_i u_i + \lambda_i \psi_{1i} y_{1i} + \alpha)^{x_i + \alpha + 1}} \\
& + \frac{p_{2i}}{(\lambda_i u_i + \lambda_i \psi_{1i}(y_{1i} + \psi_{2i} y_{2i}) + \alpha)^{x_i + \alpha + 1}} \Bigg) \Bigg\}. \qquad (6.11)
\end{aligned}
$$

As before, we do not observe the allocation variable, q_i, for each individual, instead we again regard it as a missing value and perform likelihood inference in the presence of missing data. We adopt the EM-algorithm, implementing the algorithm outlined in § 6.2.4.

Equation 6.11 is the complete-data log-likelihood, from which we can derive the subsequent $Q(\theta, \theta^{(t)}) = \mathbb{E}[\ln(L(\theta \mid Y_{obs}, Y_{mis}))]$. Again we have

$$
\mathbb{E}(q_i \mid Y_{obs}, \theta^{(t)}) = p_{1i} \left(1 + \frac{\lambda_i \psi_{1i} y_{1i}}{\alpha} \right)^{-\alpha}.
$$

The first derivatives of the complete-data log-likelihood are straightforward to derive, allowing subsequent maximisation using standard numerical methods.

First Derivatives

The first-order derivatives of the complete-data log-likelihood are

$$
\begin{aligned}
\frac{\partial \ell}{\partial \boldsymbol{\beta}_1} &= \sum_{i=1}^{n} \Bigg\{ x_i + \delta_{1i}(1+\delta_{2i}) - \frac{\delta_{1i}\delta_{2i}(x_i+\alpha+2)(\lambda_i u_i + \lambda_i \psi_{1i}(y_{1i}+\psi_{2i}y_{2i}))}{\lambda_i u_i + \lambda_i \psi_{1i}(y_{1i}+\psi_{2i}y_{2i}) + \alpha} \\
&\quad -(1-\delta_{1i})(x_i+\alpha)\left(\frac{q_i(\lambda_i u_i + \lambda_i \psi_{1i} y_{1i})}{\lambda_i u_i + \lambda_i \psi_{1i} y_{1i} + \alpha} + \frac{(1-q_i)\lambda_i u_i}{\lambda_i u_i + \alpha} \right) \\
&\quad -\delta_{1i}(1-\delta_{2i})(x_i+\alpha+1)\Bigg(\frac{(1-p_{2i})(\lambda_i u_i + \lambda_i \psi_{1i} y_{1i})}{(\lambda_i u_i + \lambda_i \psi_{1i} y_{1i} + \alpha)^{x_i+\alpha+2}} \\
&\quad + \frac{p_{2i}(\lambda_i u_i + \lambda_i \psi_{1i}(y_{1i}+\psi_{2i}y_{2i}))}{(\lambda_i u_i + \lambda_i \psi_{1i}(y_{1i}+\psi_{2i}y_{2i}) + \alpha)^{x_i+\alpha+2}} \Bigg(\frac{1-p_{2i}}{(\lambda_i u_i + \lambda_i \psi_{1i} y_{1i} + \alpha)^{x_i+\alpha+1}} \\
&\quad + \frac{p_{2i}}{(\lambda_i u_i + \lambda_i \psi_{1i}(y_{1i}+\psi_{2i}y_{2i}) + \alpha)^{x_i+\alpha+1}} \Bigg)^{-1} \Bigg\} \boldsymbol{z}_{1i}, \\
\frac{\partial \ell}{\partial \boldsymbol{\beta}_2} &= \sum_{i=1}^{n} \Bigg\{ \delta_{1i}(1+\delta_{2i}) - \frac{\delta_{1i}\delta_{2i}(x_i+\alpha+2)\lambda_i \psi_{1i}(y_{1i}+\psi_{2i}y_{2i})}{\lambda_i u_i + \lambda_i \psi_{1i}(y_{1i}+\psi_{2i}y_{2i}) + \alpha} \\
&\quad - \frac{(1-\delta_{1i})(x_i+\alpha) q_i \lambda_i \psi_{1i} y_{1i}}{\lambda_i u_i + \lambda_i \psi_{1i} y_{1i} + \alpha} \\
&\quad -\delta_{1i}(1-\delta_{2i})(x_i+\alpha+1)\Bigg(\frac{(1-p_{2i})\lambda_i \psi_{1i} y_{1i}}{(\lambda_i u_i + \lambda_i \psi_{1i} y_{1i} + \alpha)^{x_i+\alpha+2}} \\
&\quad + \frac{p_{2i}\lambda_i \psi_{1i}(y_{1i}+\psi_{2i}y_{2i})}{(\lambda_i u_i + \lambda_i \psi_{1i}(y_{1i}+\psi_{2i}y_{2i}) + \alpha)^{x_i+\alpha+2}} \Bigg(\frac{1-p_{2i}}{(\lambda_i u_i + \lambda_i \psi_{1i} y_{1i} + \alpha)^{x_i+\alpha+1}} \\
&\quad + \frac{p_{2i}}{(\lambda_i u_i + \lambda_i \psi_{1i}(y_{1i}+\psi_{2i}y_{2i}) + \alpha)^{x_i+\alpha+1}} \Bigg)^{-1} \Bigg\} \boldsymbol{z}_{2i}, \\
\frac{\partial \ell}{\partial \boldsymbol{\beta}_3} &= \sum_{i=1}^{n} \Bigg\{ \delta_{1i}\delta_{2i} - \frac{\delta_{1i}\delta_{2i}(x_i+\alpha+2)\lambda_i \psi_{1i}\psi_{2i}y_{2i}}{\lambda_i u_i + \lambda_i \psi_{1i}(y_{1i}+\psi_{2i}y_{2i}) + \alpha} \\
&\quad - \frac{\delta_{1i}(1-\delta_{2i})(x_i+\alpha+1)p_{2i}\lambda_i \psi_{1i}\psi_{2i}y_{2i}}{(\lambda_i u_i + \lambda_i \psi_{1i}(y_{1i}+\psi_{2i}y_{2i}) + \alpha)^{x_i+\alpha+2}} \Bigg(\frac{1-p_{2i}}{(\lambda_i u_i + \lambda_i \psi_{1i} y_{1i} + \alpha)^{x_i+\alpha+1}} \\
&\quad + \frac{p_{2i}}{(\lambda_i u_i + \lambda_i \psi_{1i}(y_{1i}+\psi_{2i}y_{2i}) + \alpha)^{x_i+\alpha+1}} \Bigg)^{-1} \Bigg\} \boldsymbol{z}_{3i}, \\
\frac{\partial \ell}{\partial \boldsymbol{\kappa}_1} &= \sum_{i=1}^{n} \Big\{ (1-p_{1i}) - (1-\delta_{1i})(1-q_i) \Big\} \boldsymbol{w}_{1i},
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \ell}{\partial \boldsymbol{\kappa}_2} &= \sum_{i=1}^{n} \Bigg\{ \delta_{1i}\delta_{2i}(1-p_{2i}) - \delta_{1i}(1-\delta_{2i})p_{2i}(1-p_{2i}) \\
&\quad \times \Bigg(\frac{1}{(\lambda_i u_i + \lambda_i \psi_{1i} y_{1i} + \alpha)^{x_i+\alpha+1}} \\
&\quad - \frac{1}{(\lambda_i u_i + \lambda_i \psi_{1i}(y_{1i} + \psi_{2i} y_{2i}) + \alpha)^{x_i+\alpha+1}} \Bigg) \\
&\quad \times \Bigg(\frac{1-p_{2i}}{(\lambda_i u_i + \lambda_i \psi_{1i} y_{1i} + \alpha)^{x_i+\alpha+1}} \\
&\quad + \frac{p_{2i}}{(\lambda_i u_i + \lambda_i \psi_{1i}(y_{1i} + \psi_{2i} y_{2i}) + \alpha)^{x_i+\alpha+1}} \Bigg)^{-1} \Bigg\} \boldsymbol{w}_{2i}, \\
\frac{\partial \ell}{\partial \alpha} &= \sum_{i=1}^{n} \Bigg\{ \Bigg[\sum_{k=0}^{x_i-1} \frac{1}{\alpha+k} \Bigg] + \frac{\delta_{1i}}{x_i+\alpha} + \frac{\delta_{1i}\delta_{2i}}{x_i+\alpha+1} + \ln(\alpha) + 1 \\
&\quad - \delta_{1i}\delta_{2i} \ln(\lambda_i u_i + \lambda_i \psi_{1i}(y_{1i} + \psi_{2i} y_{2i}) + \alpha) \\
&\quad - \frac{\delta_{1i}\delta_{2i}(x_i+\alpha+2)}{\lambda_i u_i + \lambda_i \psi_{1i}(y_{1i} + \psi_{2i} y_{2i}) + \alpha} \\
&\quad - (1-\delta_{1i}) q_i \Bigg(\frac{x_i+\alpha}{\lambda_i u_i + \lambda_i \psi_{1i} y_{1i} + \alpha} + \ln(\lambda_i u_i + \lambda_i \psi_{1i} y_{1i} + \alpha) \Bigg) \\
&\quad - (1-\delta_{1i})(1-q_i) \Bigg(\frac{x_i+\alpha}{\lambda_i u_i + \alpha} + \ln(\lambda_i u_i + \alpha) \Bigg) \\
&\quad - \delta_{1i}(1-\delta_{2i}) \Bigg[\Bigg(\frac{x_i+\alpha+1}{\lambda_i u_i + \lambda_i \psi_{1i} y_{1i} + \alpha} + \ln(\lambda_i u_i + \lambda_i \psi_{1i} y_{1i} + \alpha) \Bigg) \\
&\quad \times \frac{1-p_{2i}}{(\lambda_i u_i + \lambda_i \psi_{1i} y_{1i} + \alpha)^{x_i+\alpha+1}} \\
&\quad + \Bigg(\frac{x_i+\alpha+1}{\lambda_i u_i + \lambda_i \psi_{1i}(y_{1i} + \psi_{2i} y_{2i}) + \alpha} \\
&\quad + \ln(\lambda_i u_i + \lambda_i \psi_{1i}(y_{1i} + \psi_{2i} y_{2i}) + \alpha) \Bigg) \\
&\quad \times \frac{p_{2i}}{(\lambda_i u_i + \lambda_i \psi_{1i}(y_{1i} + \psi_{2i} y_{2i}) + \alpha)^{x_i+\alpha+1}} \Bigg] \\
&\quad \times \Bigg(\frac{1-p_{2i}}{(\lambda_i u_i + \lambda_i \psi_{1i} y_{1i} + \alpha)^{x_i+\alpha+1}} \\
&\quad + \frac{p_{2i}}{(\lambda_i u_i + \lambda_i \psi_{1i}(y_{1i} + \psi_{2i} y_{2i}) + \alpha)^{x_i+\alpha+1}} \Bigg)^{-1} \Bigg\}.
\end{aligned}
$$

6.4 Implementing the Full Joint Model

Tables 6.7 and 6.8 give the estimated regression coefficients for the full joint model. The reference group contains individuals with partial seizures pre-randomisation, with a normal EEG and randomised to deferred treatment. A β regression coefficient > 0 (< 0) would indicate an increased (decreased) seizure rate relative to the seizure rate in the reference group. A κ regression coefficient > 0 (< 0) would indicate an increase (decrease) in the susceptible proportion relative to the susceptible proportion in the reference group and, therefore, a subsequent decrease (increase) in the corresponding cure rate.

We can see that the regression coefficients in λ_i are similar to those observed in Table 5.1 for Joint Model A and Joint Model B. This is not surprising as the extensions considered in this chapter are concerned with the post-randomisation seizure rates, rather than the pre-randomisation seizure rates.

The value of α we observe in Table 6.7 is larger than the value observed in Table 5.1. This increase in α indicates that there is less natural heterogeneity in the full joint model than in the simple joint models developed in Chapter 5. This is increase in α is unsurprising as the inclusion of more covariates explains more of the observed variation between individuals.

	Regression Coefficient	Estimates (standard errors)
	α	2.023 (0.107)
λ_i	$\beta_{1,0}$	-4.131 (0.086)
	$\beta_{1,t-c}$	-1.076 (0.096)
	$\beta_{1,2^\circ t-c}$	-0.701 (0.098)
	$\beta_{1,partial}$	reference
ψ_{1i}	$\beta_{2,0}$	-0.958 (0.320)
	$\beta_{2,trt}$	0.307 (0.347)
	$\beta_{2,t-c}$	0.577 (0.334)
	$\beta_{2,2^\circ t-c}$	0.483 (0.344)
	$\beta_{2,partial}$	reference
	$\beta_{2,eeg}$	0.595 (0.350)
	$\beta_{2,t-c\times trt}$	-0.468 (0.352)
	$\beta_{2,2^\circ t-c\times trt}$	-0.594 (0.362)
	$\beta_{2,partial\times trt}$	reference
	$\beta_{2,eeg\times trt}$	-0.593 (0.202)
	$\beta_{2,t-c\times eeg}$	-0.518 (0.363)
	$\beta_{2,2^\circ t-c\times eeg}$	-0.361 (0.374)
	$\beta_{2,partial\times eeg}$	reference
ψ_{2i}	$\beta_{3,0}$	1.537 (0.524)
	$\beta_{3,trt}$	-0.393 (0.537)
	$\beta_{3,t-c}$	-1.219 (0.544)
	$\beta_{3,2^\circ t-c}$	-0.590 (0.551)
	$\beta_{3,partial}$	reference
	$\beta_{3,eeg}$	-0.820 (0.545)
	$\beta_{3,t-c\times trt}$	0.599 (0.549)
	$\beta_{3,2^\circ t-c\times trt}$	0.450 (0.555)
	$\beta_{3,partial\times trt}$	reference
	$\beta_{3,eeg\times trt}$	0.690 (0.315)
	$\beta_{3,t-c\times eeg}$	0.944 (0.564)
	$\beta_{3,2^\circ t-c\times eeg}$	-0.266 (0.573)
	$\beta_{3,partial\times eeg}$	reference

Table 6.7: Estimated regression coefficients in λ_i, ψ_{1i} and ψ_{2i} for the full joint model. The term λ_i contains parameter estimates corresponding to the effect of covariates on the underlying event rate. The terms ψ_{1i} and ψ_{2i} contain parameter estimates corresponding to the effect of covariates on the post-randomisation reduction in event rates at randomisation and following a first seizure post-randomisation respectively.

Stepwise backwards elimination was used to derive the optimal full joint model, which concluded that treatment policy, seizure type, EEG outcome and their interactions were important in determining the change in seizure rate at randomisation and following a first seizure post-randomisation. Treatment policy and seizure type were statistically significant in determining the proportion immune from seizure recurrence post-randomisation, as was EEG outcome and its interactions with treatment policy and seizure type. No covariate effects were statistically significant in determining the cure rate for a second post-randomisation seizure, this term was found to be dependent on an intercept term only.

Wald tests were carried out on each of the estimated regression coefficients, and concluded that seizure type is statistically significant in determining the underlying, pre-randomisation seizure rates. Only $\beta_{2,eeg\times trt}$ is found to be statistically significant when considering the change in seizure rate following randomisation. When looking at those explanatory variables corresponding with the change in seizure rate following a seizure post-randomisation, we see that only $\beta_{3,t-c}$ and $\beta_{3,eeg\times trt}$ are statistically significant.

When looking at the regression coefficients in Table 6.8 we see that seizure type is statistically significant in determining whether an individual is susceptible to seizure recurrence post-randomisation. Additionally $\kappa_{1,2^\circ t-c\times eeg}$ and $\kappa_{1,eeg\times trt}$ are statistically significant in determining the cure rates associated with the times to first seizure post-randomisation. For the cure rates associated with the times from first to second seizure post-randomisation, recall that no statistically significant covariate effects were observed.

	Regression Coefficient	Estimates (standard errors)
p_{1i}	$\kappa_{1,0}$	0.706 (0.359)
	$\kappa_{1,trt}$	-0.067 (0.146)
	$\kappa_{1,t-c}$	-0.750 (0.365)
	$\kappa_{1,2^\circ t-c}$	-0.979 (0.374)
	$\kappa_{1,partial}$	reference
	$\kappa_{1,eeg}$	-0.006 (0.448)
	$\kappa_{1,eeg\times trt}$	-0.582 (0.228)
	$\kappa_{1,t-c\times eeg}$	0.624 (0.458)
	$\kappa_{1,2^\circ t-c\times eeg}$	1.378 (0.478)
	$\kappa_{1,partial\times eeg}$	reference
p_{2i}	$\kappa_{2,0}$	1.037 (0.099)
-Log-likelihood (d.f.)		10855 (1301)

Table 6.8: Estimated regression coefficients in p_{1i} and p_{2i} for the full joint model. The term p_{1i} contains parameter estimates corresponding to the effect of covariates on the cure rate for the first seizure post-randomisation. The term p_{2i} contains parameter estimates corresponding to the effect of covariates on the cure rate for the second seizure post-randomisation.

6.4.1 Interpretation of the Results

To gain a better understanding of the estimated regression coefficients, given in Tables 6.7 and 6.8, we can obtain subsequent estimates of the cure rates and the pre and post-randomisation seizure rates for the different seizure types, EEG outcomes and treatment policies.

Table 6.9 shows that those individuals with a normal EEG, presenting with partial seizures can expect to have a cure rate of around 30%, irrespective of treatment policy, whilst those with tonic-clonic seizures can expect to have a cure rate of around 50%. For those with an abnormal EEG, higher cure rates are observed for those individuals randomised to immediate treatment, rather

than deferred. Recall that Warrell et al. (2003) put seizure recurrence after a single untreated seizure at around 80%, but Berg and Shinnar (1991) stated that seizure recurrence was 50%. The results that we have observed in Table 6.9 may provide an explanation for the difference in these values.

Seizure Type	$1-\widehat{p}_{1i}$ (95% C.I.)			
	Abnormal EEG			
	Immediate		Deferred	
Tonic-Clonic	0.518	(0.45,0.59)	0.360	(0.20,0.43)
2° Tonic-Clonic	0.389	(0.31,0.48)	0.250	(0.19,0.33)
Partial	0.487	(0.36,0.62)	0.332	(0.22,0.46)
	Normal EEG			
	Immediate		Deferred	
Tonic-Clonic	0.528	(0.47,0.59)	0.511	(0.45,0.57)
2° Tonic-Clonic	0.584	(0.51,0.65)	0.568	(0.50,0.64)
Partial	0.345	(0.20,0.52)	0.330	(0.19,0.50)

Table 6.9: The expected cure rates associated with the post-randomisation times to first seizure, using the estimated regression coefficients from the full joint model.

Recall that no significant covariate effects were observed for the cure rate associated with the times from first to second seizure. The overall cure rate for this survival time is 26%.

Table 6.10 shows the maximum likelihood estimates of ψ_{1i} and ψ_{2i} under the full joint model. The values of $\widehat{\psi}_{1i}$ represent the change in seizure rate following randomisation, with $\widehat{\psi}_{2i}$ representing the change in rate following first post-randomisation seizure. We again see that treatment policy does not appear to be statistically significant in determining the estimate of $\widehat{\psi}_{1i}$ for those individuals with a normal EEG. Additionally, for those individuals with an

abnormal EEG, immediate treatment is favoured.

Seizure Type	$\widehat{\psi}_{1i}$ (95% C.I.) first seizure			
	Abnormal EEG			
	Immediate		Deferred	
Tonic-Clonic	0.347	(0.26,0.47)	0.738	(0.57,0.95)
2° Tonic-Clonic	0.326	(0.24,0.44)	0.786	(0.58,1.06)
Partial	0.522	(0.31,0.88)	0.695	(0.41,1.18)
	Normal EEG			
	Immediate		Deferred	
Tonic-Clonic	0.582	(0.45,0.75)	0.683	(0.53,0.87)
2° Tonic-Clonic	0.467	(0.34,0.65)	0.622	(0.46,0.84)
Partial	0.522	(0.27,1.00)	0.384	(0.20,0.73)
	$\widehat{\psi}_{2i}$ (95% C.I.) second seizure			
	Abnormal EEG			
	Immediate		Deferred	
Tonic-Clonic	3.806	(2.43,5.97)	1.554	(1.00,2.41)
2° Tonic-Clonic	1.835	(1.17,2.88)	0.870	(0.56,1.36)
Partial	2.754	(1.17,6.49)	2.047	(0.98,4.26)
	Normal EEG			
	Immediate		Deferred	
Tonic-Clonic	1.688	(1.13,2.51)	1.373	(0.93,2.02)
2° Tonic-Clonic	2.727	(1.69,4.41)	2.577	(1.67,3.97)
Partial	3.138	(1.28,7.70)	4.649	(1.63,13.27)

Table 6.10: The expected change in seizure rates following randomisation and following the first post-randomisation seizure, using the estimated regression coefficients from the full joint model.

Following a first seizure post-randomisation we again see that, in general, seizure rates increase. We observe that those individuals allocated to immediate treatment see more of an increase in seizure rate following a first seizure post-randomisation than those allocated to deferred treatment. Recall that a possible explanation for this may be that those allocated to deferred treatment may subsequently be started on a course of AEDs following a seizure post-randomisation, bringing them in line with those allocated to immediate

treatment. Also note that these individuals typically experience the smallest reduction in seizure rate following randomisation.

6.5 Discussion

This chapter has examined extensions to the simple joint model that was developed in Chapter 5, both in isolation, and together in a single full model.

The simple joint model assumed that, post-randomisation, an individual's seizure rate remained constant. This assumption was relaxed and a model was considered that allowed the seizure rate to change both at randomisation, and following a first post-randomisation seizure. We let ψ_{1i} and ψ_{2i} be the seizure rate modifiers at randomisation and following a first post-randomisation seizure respectively. We found that seizure rates were generally reduced after randomisation, but then increased following a first post-randomisation seizure. The change in seizure rates both at randomisation and following a first post-randomisation seizure were found to be dependent on seizure type, treatment policy and EEG outcome.

We next considered the use of cure rate models in analysing the MESS data. Standard statistical software exists for the implementation of Log-logistic mixture survival models, that allow for the existence of cure rates. This model was considered for the times to first seizure and from first to second seizure separately. Initial models, that allowed the cure rates to depend on an intercept term only, concluded that there was highly statistically significant evidence to support the inclusion of cure rates in modelling. Further investigation found

that the cure rates associated with time to first seizure post-randomisation were dependent on seizure type, EEG outcome and the logarithm of an individual's pre-randomisation seizure rate. For the cure rates associated with the post-randomisation times from first to second seizure however, no statistically significant covariate effects were found.

A model that jointly analysed the pre-randomisation event counts and post-randomisation times to first seizure only, with cure rates, was included for completeness. A joint model for the pre-randomisation event counts and both of the post-randomisation survival times, allowing each of the survival times to have an associated cure rate, was then considered. It was noted that if an individual's time to first seizure post-randomisation is censored, it is unknown whether censoring is due to the individual being 'immune' to seizures post-randomisation, or due to their period of follow up not being sufficient to observe the survival time. An unobserved allocation variable was introduced to distinguish between these two scenarios and the EM-algorithm was adopted for the maximisation of the subsequent log-likelihood.

The optimal joint model that considers the pre-randomisation event counts and both post-randomisation survival times, with cure rates, concluded that those individuals allocated to deferred treatment, with an abnormal EEG, can expect to have statistically significantly lower cure rates than those with a normal EEG, except for those individuals experiencing partial seizures only pre-randomisation, where no statistically significant difference is observed. For those with a normal EEG, no statistically significant treatment effect on the cure rates was observed. There were no covariates that were found to be sta-

tistically significant in determining the cure rates associated with the times from first to second seizure.

After considering each of the extensions applicable to the MESS data in isolation, a model was developed that allows the pre-randomisation event counts and post-randomisation times to first seizure, and from first to second seizure to be analysed jointly. This full joint model allowed the seizure rate to vary post-randomisation and incorporated cure rates for each of the post-randomisation survival times.

Those individuals with a normal EEG, presenting with partial seizures can expect to have a cure rate of around 30%, irrespective of treatment policy, whilst those with tonic-clonic seizures can expect to have a cure rate of around 50%. For those with an abnormal EEG, higher cure rates are observed for those individuals randomised to immediate treatment rather than deferred. No statistically significant covariate effects were concluded in determining the cure rate for a second post-randomisation seizure, this term was found to be dependent on an intercept term only. Treatment policy was not statistically significant in determining the estimate of the change in seizure rate following randomisation, for those individuals with a normal EEG. For those individuals with an abnormal EEG, immediate treatment was favoured. Following a first seizure post-randomisation we observed a general increase in seizure rates, however, those allocated to deferred treatment, possibly surprisingly, have the smallest increase in seizure rate following a first seizure post-randomisation. Possible explanations for this observation were discussed.

Chapter 7

Model Checking

This thesis discusses the use of joint models for event counts and survival times as an alternative to standard survival models. Chapter 5 developed a simple joint model that had more statistically significant covariate effects than standard survival analysis. Further investigation of the subsequent estimated pre-randomisation seizure rates and post-randomisation seizure rate modifiers, however, highlighted possible limitations to the model.

Chapter 6 considered each of the interesting characteristics, present in the data, separately and together in a final full joint model. This full joint model allowed the pre-randomisation event counts and post-randomisation times to first seizure and from first to second seizure to be analysed together. Furthermore, this full joint model assumed a change in seizure rate both at randomisation and following a first post-randomisation seizure, with each of the survival times having an associated proportion cured.

This chapter will assess the performance of this full joint model, compared with standard survival analysis, by investigating how well the distribution of the survival times is modelled under the different models considered.

7.1 Kaplan-Meier Curves

We will present a number of Kaplan-Meier curves along with the subsequent fitted estimates of survival, using the final full joint model that was developed in Chapter 6. These survival curves will be compared to the estimated survival curves obtained through implementing the standard survival methods.

It has already been noted that standard software (Peng) exists that allows for the analysis of survival data using the Log-logistic model with cure rates. We therefore use the estimates obtained from the Log-logistic mixture model, that incorporated a cure fraction, to produce fitted survival curves for the times to first seizure, and from first to second seizure post-randomisation. The estimated survivor functions are derived using

$$\widehat{S}_{Y_j}(y_{ji};\mu_{ji},\gamma_j,p_{ji}) \;=\; 1-\widehat{p}_{ji}+\widehat{p}_{ji}(1+(\widehat{\mu}_{ji}y_{ji})^{\widehat{\gamma}_j})^{-1}, \quad j=1,2,$$

where in each model $\widehat{\mu}_{ji} = \exp(\widehat{\boldsymbol{\theta}}_j'\boldsymbol{d}_i)$ for a vector $\widehat{\boldsymbol{\theta}}_j$ of estimated regression coefficients for survival time j, and a vector $\boldsymbol{d}_i$ of covariates for each individual i. The parameter $\widehat{\gamma}_j > 0$ is the estimated shape parameter associated with survival time j.

The term $\widehat{p}_{ji}$ represents the estimated susceptible proportion associated with

survival time j post-randomisation, so that $1-\widehat{p}_{ji}$ is the estimated cure fraction. The following parameterisations for $\widehat{p}_{1i}$ and $\widehat{p}_{2i}$ are considered:

$$\begin{aligned}\widehat{p}_{1i} &= \frac{\exp(\widehat{\boldsymbol{\kappa}}_1'\boldsymbol{w}_{1i})}{1+\exp(\widehat{\boldsymbol{\kappa}}_1'\boldsymbol{w}_{1i})},\\ \widehat{p}_{2i} &= \frac{\exp(\widehat{\boldsymbol{\kappa}}_2'\boldsymbol{w}_{2i})}{1+\exp(\widehat{\boldsymbol{\kappa}}_2'\boldsymbol{w}_{2i})}.\end{aligned}$$

The estimated Log-logistic fitted survival curves will be compared with the fitted estimates of survival obtained using the full joint model that allows the seizure rate to vary post-randomisation, and incorporates cure rates for each of the survival times. Log-likelihood ratio tests concluded that the full joint model, presented in § 6.3, was the optimal model for analysing the MESS data. Stepwise backwards elimination was used to determine which explanatory variables were statistically significant in the model. The estimated survivor functions for the survival times, under the full joint model are given by

$$\begin{aligned}\widehat{S}_{Y_1}(y_{1i};\lambda_i,\psi_{1i},\alpha,p_{1i}) &= 1-\widehat{p}_{1i}+\widehat{p}_{1i}\left(1+\frac{\widehat{\lambda}_i\widehat{\psi}_{1i}y_{1i}}{\widehat{\alpha}}\right)^{-\widehat{\alpha}},\\ \widehat{S}_{Y_1}(y_{1i};\lambda_i,\psi_{1i},\psi_{2i},\alpha,p_{2i}) &= 1-\widehat{p}_{2i}+\widehat{p}_{2i}\left(1+\frac{\widehat{\lambda}_i\widehat{\psi}_{1i}\widehat{\psi}_{2i}y_{2i}}{\widehat{\alpha}}\right)^{-\widehat{\alpha}}.\end{aligned}$$

Additionally, $\widehat{\lambda}_i=\exp(\widehat{\boldsymbol{\beta}}_1'\boldsymbol{z}_{1i})$, $\widehat{\psi}_{1i}=\exp(\widehat{\boldsymbol{\beta}}_2'\boldsymbol{z}_{2i})$, $\widehat{\psi}_{2i}=\exp(\widehat{\boldsymbol{\beta}}_3'\boldsymbol{z}_{3i})$, where $\widehat{\boldsymbol{\beta}}_1$, $\widehat{\boldsymbol{\beta}}_2$ and $\widehat{\boldsymbol{\beta}}_3$ are vectors of estimated regression coefficients and $\boldsymbol{z}_{1i}$, $\boldsymbol{z}_{2i}$, $\boldsymbol{z}_{3i}$ are vectors of covariates, not necessarily distinct.

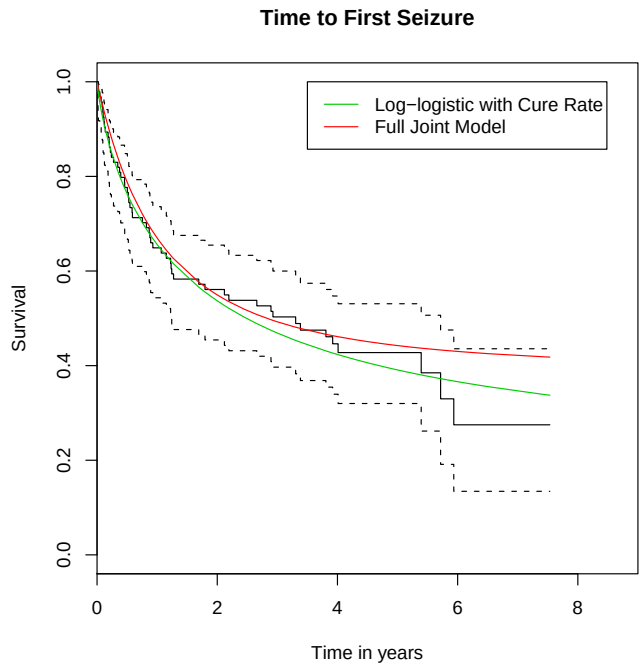
We will examine the Kaplan-Meier and fitted estimates of the survival curves for different subgroups of individuals in our dataset. We confine our investigation of the survival curves to those individuals presenting with secondary

tonic-clonic seizures only pre-randomisation. We then consider four different scenarios for these individuals separately:

- Normal EEG and allocated to immediate treatment.
- Normal EEG and allocated to deferred treatment.
- Abnormal EEG and allocated to immediate treatment.
- Abnormal EEG and allocated to deferred treatment.

Recall that the standard survival analysis of the MESS data included the logarithm of an individual's pre-randomisation seizure rate as an explanatory variable in determining their subsequent times to first seizure and from first to second seizure. In the subsequent derivation of the estimated fitted survival curves we shall just consider the mean of the logarithm of the pre-randomisation seizure rates for each individual.

Figures 7.1-7.4 present the Kaplan-Meier estimates of the survival curves, for each of the subgroups considered, with their corresponding 95% confidence intervals. The estimated fitted survival curves from the two parametric models considered are also presented. We can see that for times to first seizure, and from first to second seizure, for each of the subgroups, the fitted full joint model, and the fitted Log-logistic survival model with cure rates, both seem to model the distribution of survival very well. We observed that for the most part, the estimates of the fitted survival curves for the full joint model, and Log-logistic model with cure rates, tend to remain within the 95% confidence intervals of the corresponding Kaplan-Meier estimates.



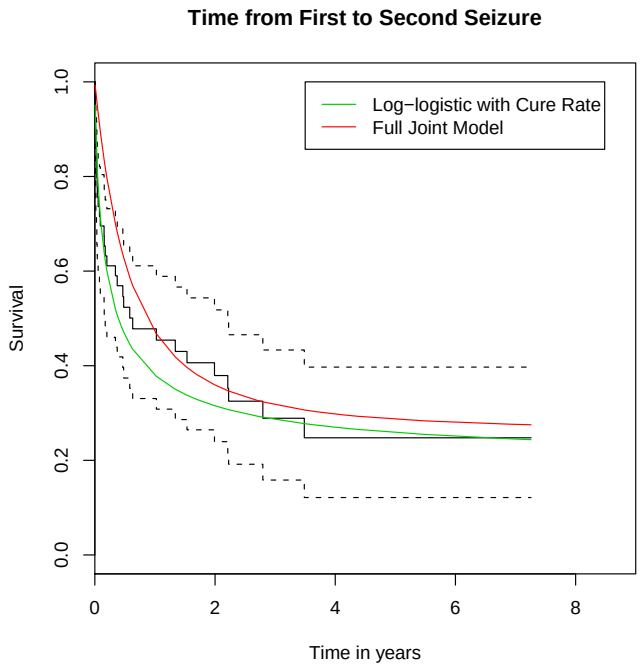
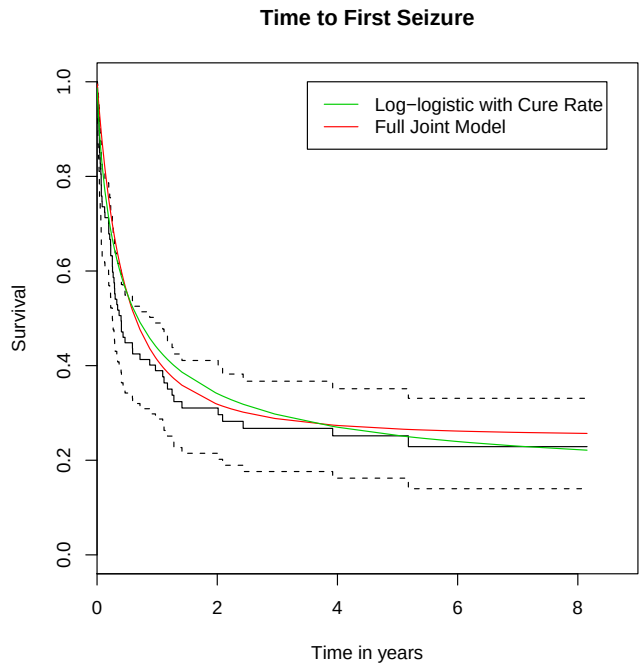


Figure 7.1: Kaplan-Meier curves and fitted curves for time to first seizure, and time from first to second seizure, for those with an abnormal EEG and allocated to immediate treatment.



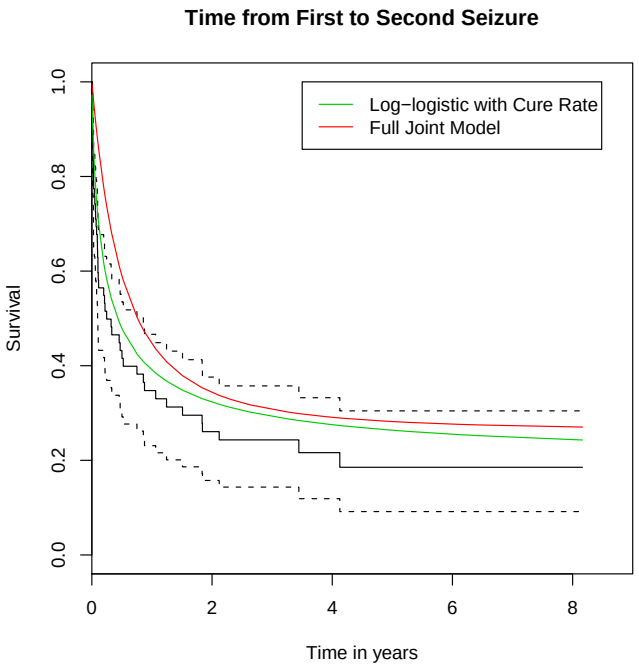
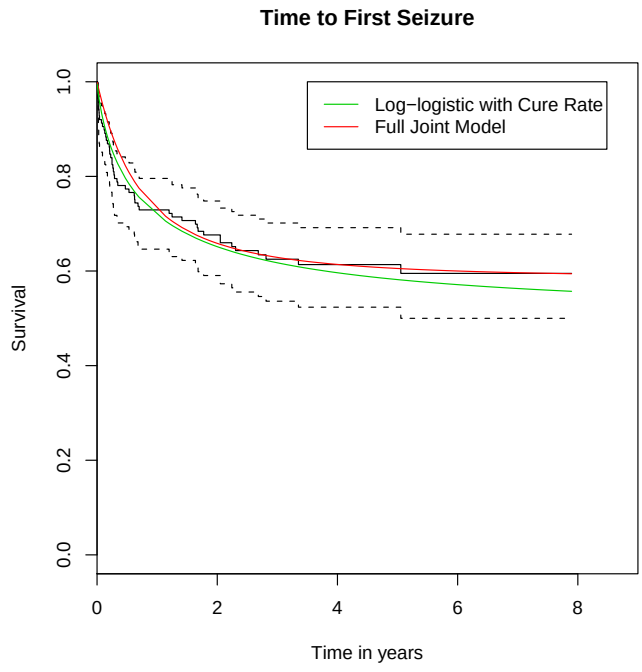


Figure 7.2: Kaplan-Meier curves and fitted curves for time to first seizure, and time from first to second seizure, for those with an abnormal EEG and allocated to deferred treatment.



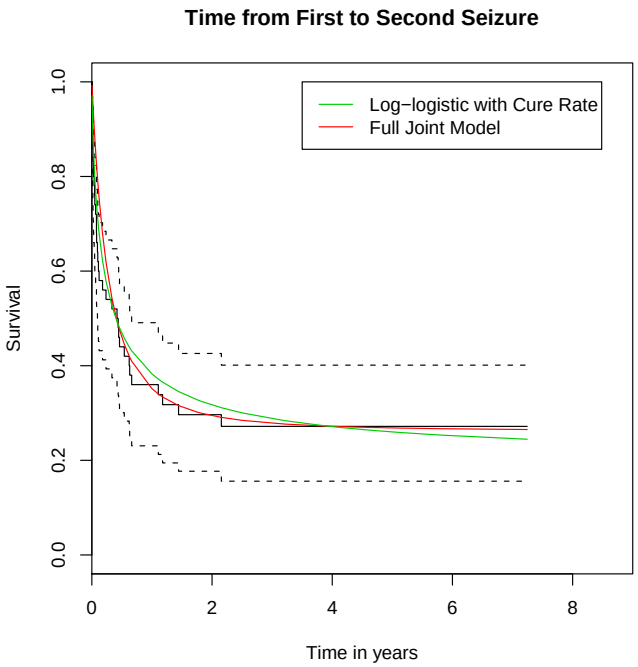
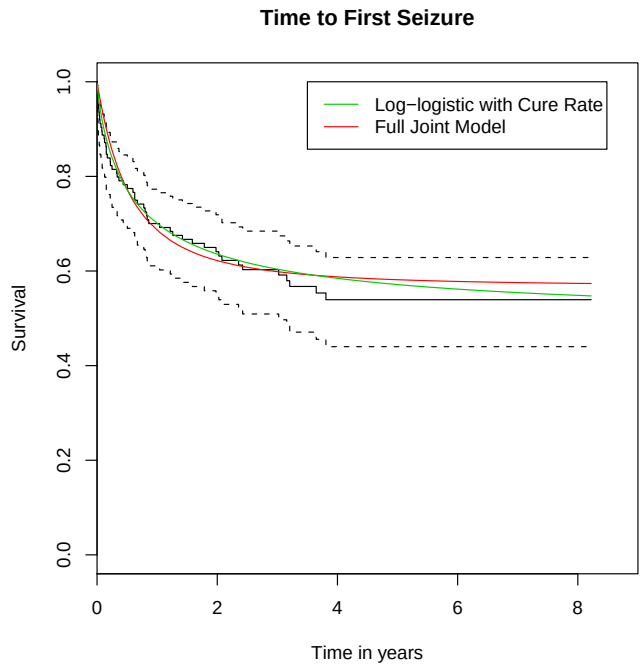


Figure 7.3: Kaplan-Meier curves and fitted curves for time to first seizure, and time from first to second seizure, for those with a normal EEG and allocated to immediate treatment.



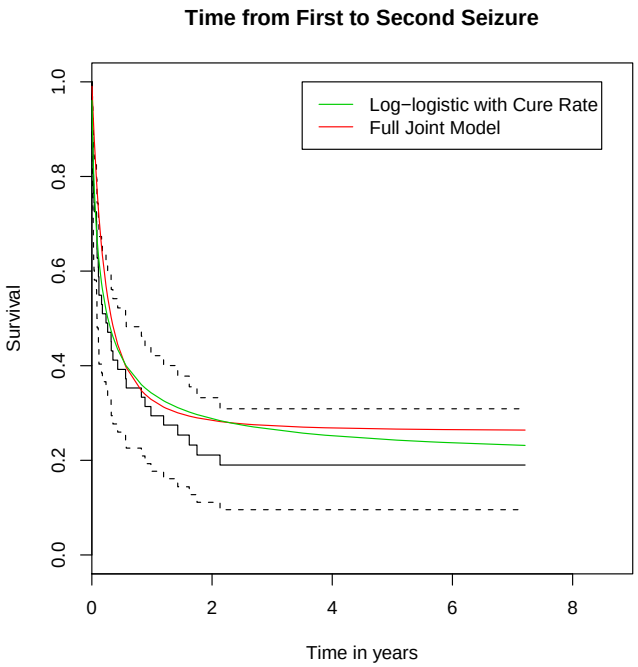


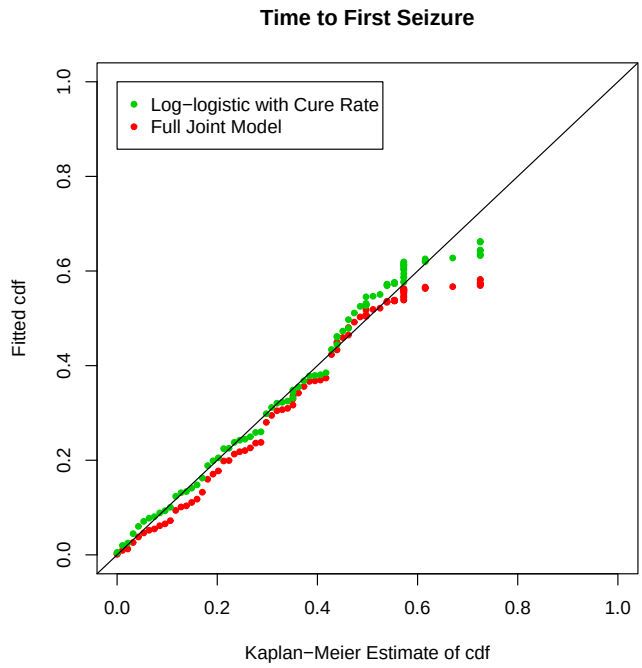
Figure 7.4: Kaplan-Meier curves and fitted curves for time to first seizure, and time from first to second seizure, for those with a normal EEG and allocated to deferred treatment.

7.2 Goodness-of-Fit

A method for testing goodness-of-fit of parametric distributions to survival data was developed by Maller and Zhou (1996), which is a variant to the method devised by Filliben (1975) for testing the normality of uncensored data. Filliben (1975) proposed calculating the correlation coefficient between the order statistics of a sample and the expected values of the order statistics for a sample of the same size from a standard Normal distribution. The subsequent correlation coefficient then forms the basis of the test, with values close to 1 being indicative of a good fit and values close to zero suggesting a poor fit.

The test for censored survival times introduced by Maller and Zhou (1996) considers the hypothesis $H_0 : F = \widehat{F}$, where $\widehat{F}$ is some specified distribution function. Let $y_1 \leq y_2 \leq \ldots \leq y_n$ be the ordered sample of survival times. If there is no censoring present, then a plot of $\widehat{F}(y_{ji})$ against the empirical distribution function, $\tilde{F}(y_{ji}) = i/n$, under H_0, should produce a near-straight line with slope close to 1. When censoring is present the same argument would lead us to expect to obtain, under H_0, a near-straight line with slope close to 1, by plotting $\widehat{F}(y_{ji}) = 1 - \widehat{S}(y_{ji})$ against $\tilde{F}(y_{ji})$, where $\tilde{F}$ is taken as the Kaplan-Meier estimate.

Figures 7.5-7.8 present plots of the estimated Kaplan-Meier estimates of the cumulative distribution function of the survival times, against the corresponding fitted estimates for the parametric models considered.



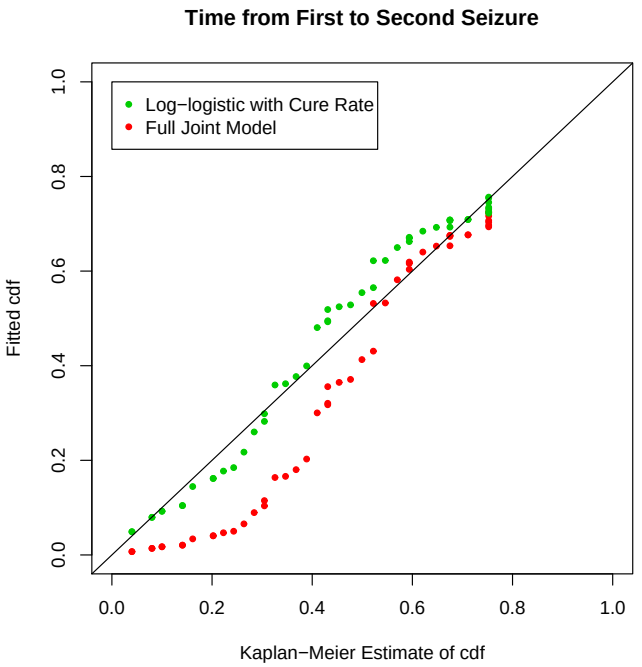


Figure 7.5: Kaplan-Meier estimate and fitted estimates of the cumulative distribution function for time to first seizure, and time from first to second seizure, for those with an abnormal EEG and allocated to immediate treatment.

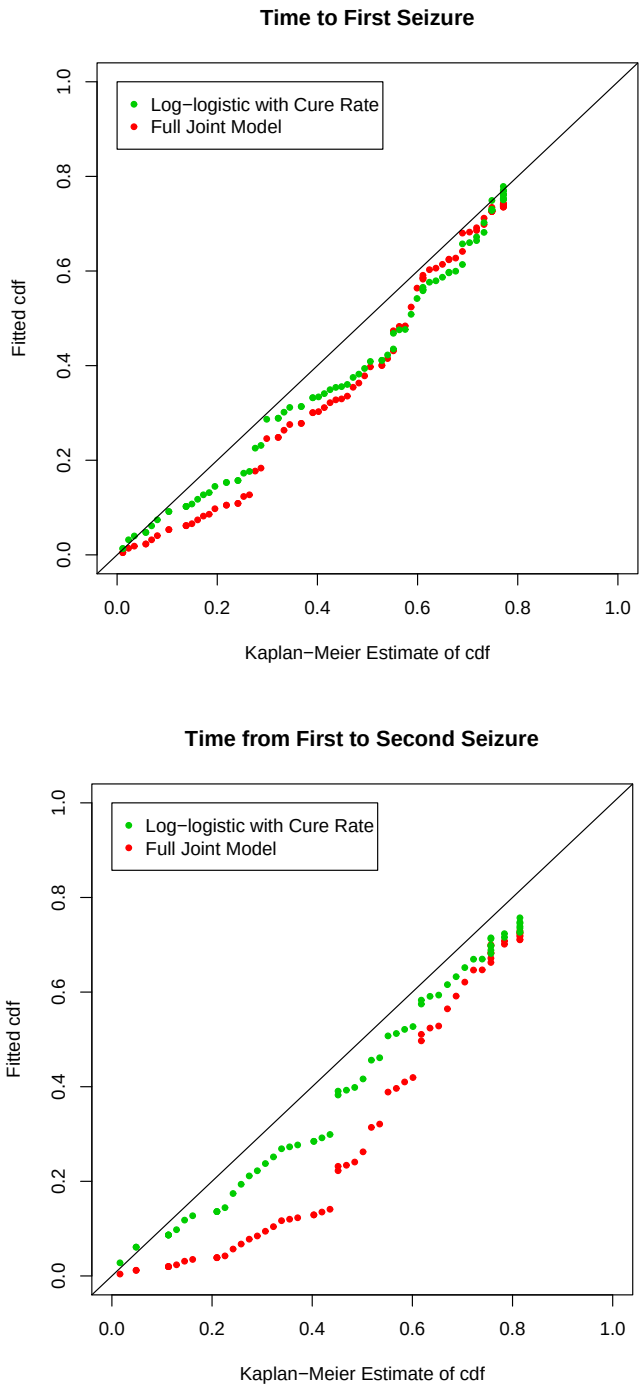
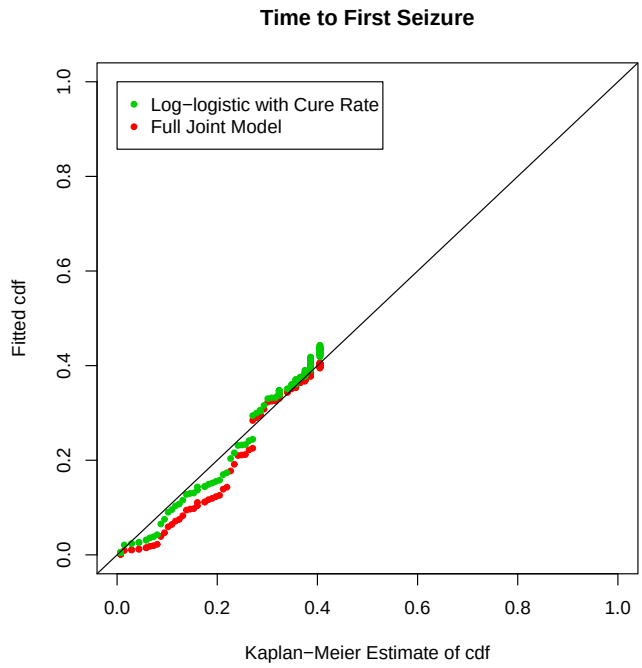


Figure 7.6: Kaplan-Meier estimate and fitted estimates of the cumulative distribution function for time to first seizure, and time from first to second seizure, for those with an abnormal EEG and allocated to deferred treatment.



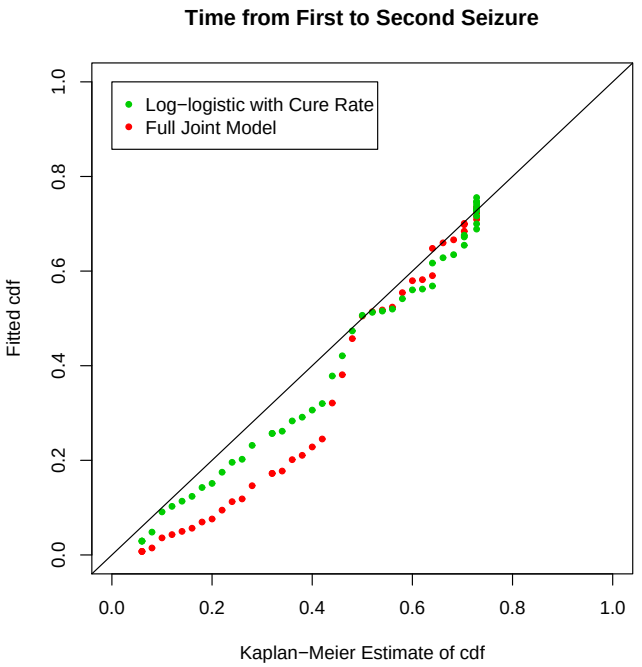


Figure 7.7: Kaplan-Meier estimate and fitted estimates of the cumulative distribution function for time to first seizure, and time from first to second seizure, for those with a normal EEG and allocated to immediate treatment.

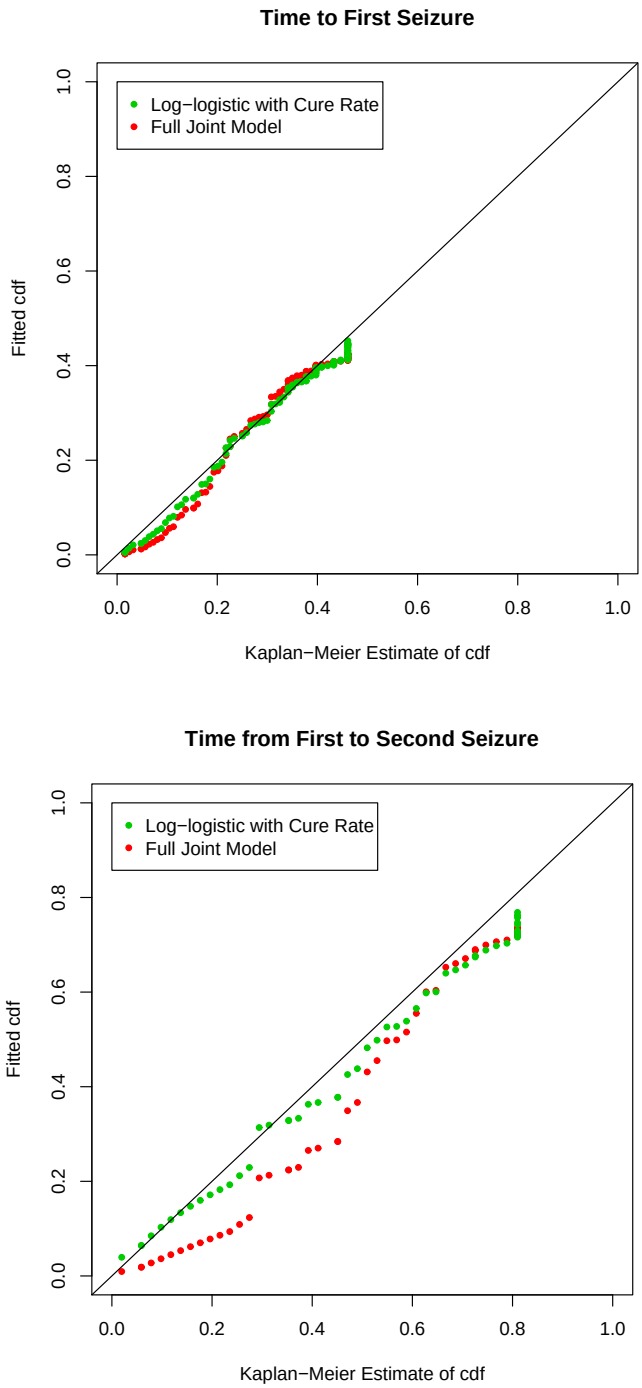


Figure 7.8: Kaplan-Meier estimate and fitted estimates of the cumulative distribution function for time to first seizure, and time from first to second seizure, for those with a normal EEG and allocated to deferred treatment.

The figures suggest that the joint model performs very well when considering the times to first seizure post-randomisation. The Log-logistic distribution, with cure rates, however, generally appears to model the survival times from first to second seizure better.

Maller and Zhou (1996) propose using the correlation coefficient, r, between $\tilde{F}(y_{ji})$ and $\hat{F}(y_{ji})$, $1 \leq i \leq n$ as an appropriate measure of the goodness-of-fit. The correlation coefficient however, is a measure of a linear relationship between two sets of data, this is not necessarily useful when trying to assess the goodness of fit. We do not wish to assess the strength of linearity between $\tilde{F}(y_{ji})$ and $\hat{F}(y_{ji})$, we wish to test how far the points deviate from the line of equality.

7.3 Discussion

We have formally shown that the full joint model performs very well when considering how well the distribution of the survival times is modelled. We have shown, however, that the Log-logistic distribution is also very good at modelling the distribution of the survival times. Comparisons between the two models have simply comprised of a visual examination of Figures 7.1-7.8. No formal comparisons between the full joint model and the standard Log-logistic survival model, with cure rates, have been carried out.

We have not formally assessed the performance of the full joint model in its own right. Maller and Zhou (1996) propose using the correlation coefficient as a means of testing model performance. We have noted however, that correla-

tion coefficients are used to assess the strength of a linear relationship between two sets of data. We wish to assess the strength of equality between two sets of data which renders the correlation coefficient unsuitable.

To assess model performance more accurately, we propose that a statistic which assesses the levels of deviations from equality of two sets of data needs to be derived.

Chapter 8

Further Extensions Applicable to MESS

We shall now present further analyses of the MESS data that could be considered, but that are not carried out in this thesis. We develop a zero-truncated, one-inflated Poisson distribution for the pre-randomisation count data as an alternative to the standard Poisson distribution. We discuss analyses that consider the type of AED an individual is assigned and include further post-randomisation survival times. We also consider the analysis of long-term prognosis.

8.1 Zero-Truncated, One-Inflated Poisson Distribution

It has already been noted that over half of the participants recruited to MESS presented only a single seizure pre-randomisation. This excess of ones that the data displays is not accounted for in any of the models considered in

this thesis. Additionally, recall that the eligibility criterion for the MESS trial specified that participants should have had at least one epileptic seizure pre-randomisation. A one-inflated, zero-truncated Poisson distribution could therefore be considered for the pre-randomisation event counts. The zero-truncated Poisson is a model for count data that is truncated at zero (Finney and Varley 1955). The density function for the zero-truncated Poisson($\lambda_i u_i \nu_i$) distribution is

$$\text{ZTP}(x_i; \lambda_i, u_i, \nu_i) = \frac{(\lambda_i u_i \nu_i)^{x_i} \exp(-\lambda_i u_i \nu_i)}{x_i!(1 - \exp(-\lambda_i u_i \nu_i))} = \frac{(\lambda_i u_i \nu_i)^{x_i}}{x_i!(\exp(\lambda_i u_i \nu_i) - 1)}.$$

The one-inflated, zero-truncated Poisson distribution is a model for data that exhibits excess ones and is truncated at zero. The model assumes that, with probability π, the only possible observation is 1, and with probability $1 - \pi$ a zero-truncated Poisson($\lambda_i u_i \nu_i$) random variable is observed. Hence,

$$\begin{aligned} X_i = 1 \quad &\text{with probability} \quad \pi + (1 - \pi)\frac{\lambda_i u_i \nu_i}{(\exp(\lambda_i u_i \nu_i) - 1)}, \\ X_i = k > 1 \quad &\text{with probability} \quad (1 - \pi)\frac{(\lambda_i u_i \nu_i)^k}{k!(\exp(\lambda_i u_i \nu_i) - 1)}, \end{aligned}$$

giving, for $x_i \geq 1$,

$$f_X(x_i; \lambda_i, u_i, \nu_i, \pi) = \pi \mathbb{I}_{[x_i=1]} + (1 - \pi)\text{ZTP}(x_i; \lambda_i u_i \nu_i), \tag{8.1}$$

where $\mathbb{I}_{[x_i=1]}$ is the indicator function taking the value 1 when $x_i = 1$ and zero otherwise.

8.2 Different Antiepileptic Drugs

MESS was initially designed to investigate the difference between two policies: immediate versus deferred treatment. The randomisation scheme was reevaluated part way through the trial, allowing the relative merits of specific drugs to be investigated. Consequently, during the trial, two randomisation procedures were utilised. Initially the clinicians declared which drug a patient would be administered only if they were allocated to immediate treatment. A consequence of this randomisation scheme is that for those individuals randomised to deferred treatment, it is not known what drug they would have been given had they been randomised to immediate treatment. Subsequently, any analysis confined to a particular drug using this data will be potentially confounded.

Part way through the trial the randomisation scheme was altered so that clinicians had to declare which drug would be most appropriate for participants prior to randomisation. Specifying the drug prior to randomisation creates control groups for individual drugs, but analysing only those individuals that allow for comparisons within specific drugs ignores a substantial amount of costly data. Of the 1425 individuals considered in the exploratory analysis, 614 were randomised using the first randomisation scheme, leaving 811 randomised using the second, updated scheme. We propose adopting missing data techniques to complete the data matrix, allowing analysis across all individuals, subsequently giving more reliable results.

8.2.1 Multiple Imputation

Our data exhibits univariate non-response, as missingness is confined to a single variable, namely the type of drug a patient was assigned at randomisation.

One strategy for handling missing data is to impute the missing values and then use standard statistical methods on the completed data matrix. Suppose there are two variables Z_J and Z_K, and suppose further that Z_J is completely observed, but some of the Z_K are missing. If the two variables are strongly correlated it may be sensible to use Z_J to predict those values of Z_K that are missing. Methods of this type can be used to impute either single or multiple values for each of the missing items, producing complete data matrices which can then be analysed using standard techniques. One method of imputation is regression imputation, which replaces missing values by predicted values from a regression of the missing item on items observed for the unit.

In the case of epilepsy, the drug a patient is assigned is strongly dependent on a number of the baseline covariates that were collected in the MESS trial, namely age, sex and covariates concerning the type of epilepsy a patient has and the nature of the seizures. This would enable us to use regression imputation methods, regressing the missing items on those influential baseline covariates we have observed, which, in turn, should allow us to approach the interesting question of differences between specific drug types.

We have already presented a brief discussion of multiple imputation (MI) in Chapter 2. Recall that MI is the term given to the procedure of replacing

missing values by $D \geq 2$ imputed values. This policy produces D complete data sets, and standard statistical techniques can be used to analyse each of these. If the D sets of imputations are repeated random draws from the same predictive model the D complete-data inferences can be combined to form one inference properly reflecting the uncertainty caused by the nonresponse.

8.3 Inclusion of further Post-Randomisation Survival Times

The outcomes concerning short-term seizure recurrence, that were measured in the MESS trial, were times to first, second and fifth seizures. However, owing to the fact that we have the raw data, it may be possible, and more informative, to establish times to the intermittent third and fourth seizures. For example, if a second seizure is observed, but the fifth seizure censored, we do not know if this censoring occurred before the third, fourth or fifth seizure. Hence, consideration of the raw data would boost the quality of the data we were working with. It would then be useful to generalise the model we have developed to allow for the joint modelling of the pre-randomisation event counts and a general m number of post-randomisation survival times.

8.4 Analysis of Long-Term Prognosis

In studies of epilepsy, interest is often not restricted to the analysis of the risk of short-term seizure recurrence, but also the long-term prognosis. The MESS trial was conducted to assess not only short-term outcomes, times to one and two year remission were also considered.

Figure 8.1 shows the Kaplan-Meier curves for the times to one and two year remission, stratified by treatment policy. Additionally, the Kaplan-Meier curves are marked at each censoring time which is not also a death time.

Standard analysis would ignore the first sections of the plots in Figure 8.1 and simply report the percentage of people achieving one year and two year remission at 365 and 730 days respectively. Marson et al. (2005) discusses time to two year remission and simply reports that at two years 64% of those in the immediate treatment group and 52% in the deferred group achieved immediate remission, further discussing how this difference diminishes in time.

Figure 8.1 shows substantial drop-out within the first one and two years. An immediate consequence of this is that the percentages presented will be exaggerated. For example, Marson et al. (2005) interpret their reported 64% as the percentage of patients randomised achieving immediate remission. What that figure actually represents is the percentage of people randomised who have either achieved immediate remission at two years, or have dropped out of the trial within the two years. An analysis that does not take this into consideration may give misleading results. The reasons for drop-out could include a patient being randomised to deferred treatment and not experiencing a further seizure, hence regarding follow-up as pointless, but could also include patients moving house, or death. Clearly careful investigation of patient drop-out must be considered.

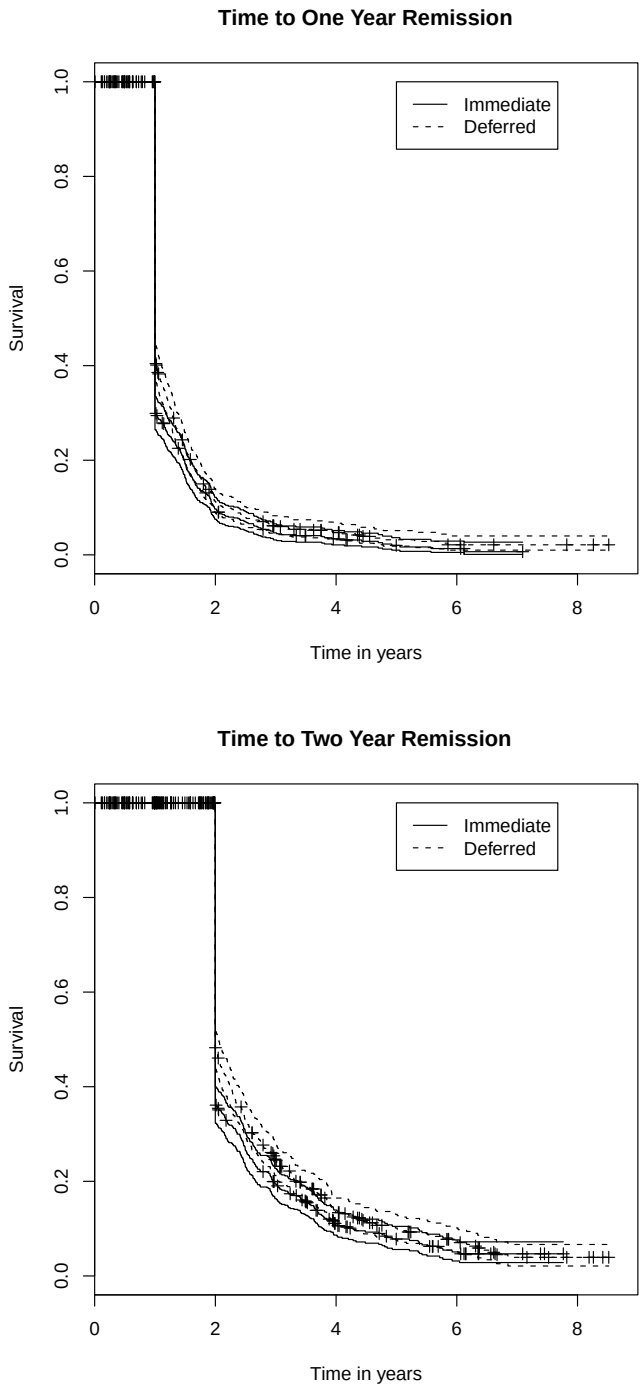


Figure 8.1: Kaplan-Meier plots of the times to one and two remission.

8.5 Discussion

This chapter has outlined some further extensions that could be considered in the analysis of the MESS dataset. Firstly, a zero-truncated, one-inflated Poisson distribution was proposed as an alternative to the standard Poisson distribution. This distribution accounts for the fact that all patients included in the MESS trial had to have had at least one clinically definite, unprovoked epileptic seizure prior to randomisation, and that over half of those included presented with only a single seizure pre-randomisation.

Two randomisation procedures were used in the MESS trial, initially a clinician would declare which drug a patient were to receive only if they were randomised to immediate treatment. The second randomisation procedure, implemented around half way through the recruitment of patients, required clinicians to declare which drug a patient would receive, prior to randomisation to a treatment policy. The second of these randomisation policies allows comparisons between the different AEDs to be carried out. For those individuals randomised using the first treatment policy, it is proposed that missing data methods are adopted to recover information missing about subsequent choices of AEDs.

Outcomes of the MESS trial, not included in the analyses presented in this thesis, are time to fifth seizure and times to one and two year remission. It is proposed that times to third and fourth seizure may be obtainable from the raw data, allowing the joint model to be generalised, to jointly model pre-randomisation event counts and a general number of m post-randomisation

survival times. Investigation into the effects of treatment policy on the long-term prognosis of epilepsy may also be carried out.

Chapter 9

Conclusions

9.1 Overview of Thesis

This thesis has built a number of models that can be used to analyse data that arrive in the form of event counts, and survival times, following a change in event rate. Chapter 3 provided an overview of the epilepsy dataset which motivated the statistical models that have been developed in this thesis. Non-parametric analyses of the data were carried out and non-parametric estimates of the survivor functions, for the two survival times, were transformed in order to provide an indication as to which parametric model may be suitable for further analyses.

In Chapter 4 standard parametric analyses of the pre-randomisation event counts and post-randomisation times to first seizure, and from first to second seizure were presented. Following the analysis of each of the outcomes separately, a joint model was developed in Chapter 5 that allowed the pre-randomisation event counts and post-randomisation survival times to be anal-

ysed in a single model. This joint model comprised a simple extension of the model developed by Cowling et al. (2006). This simple joint model was used in the analysis of the MESS data, and was found to be superior to standard survival techniques. Further examination of the estimated pre-randomisation seizure rates, and post-randomisation seizure rate modifications indicated that there may be interesting characteristics within the MESS data, not accounted for in the simple joint model.

It was proposed in Chapter 6 that the assumption of a constant seizure rate post-randomisation may not be accurate. There was evidence to suggest that cure rates may be present in the dataset, this was also investigated in Chapter 6. After examining each of these extensions to the simple joint model in isolation, a full joint model that incorporated both of the extensions was developed. This joint model allowed the seizure rate to vary post-randomisation and incorporated cure rates for each of the survival times.

Chapter 7 assessed the performance of the full joint model, compared with standard survival analysis, by investigating how well the distribution of the survival times is modelled under the different models considered. The Kaplan-Meier estimates of the survival curves were examined alongside the fitted estimates of survival for the two parametric models considered.

9.2 Conclusions about Epilepsy Data

We can see that those individuals presenting with partial seizures only pre-randomisation typically experience the highest seizure rates. Those experienc-

ing tonic-clonic seizures only and secondary tonic-clonic seizures have statistically significantly lower rates.

For the cure rates associated with the times to first seizure post-randomisation, treatment policy has no effect for those individuals presenting with a normal EEG. Those individuals with a normal EEG, presenting with partial seizures can expect to have a cure rate of around 30%, irrespective of treatment policy, whilst those with tonic-clonic seizures can expect to have a cure rate of around 50%. For those with an abnormal EEG, higher cure rates are observed for those individuals randomised to immediate treatment, rather than deferred.

The optimal full joint model concluded that no explanatory variables were statistically significant in determining the cure rates associated with the times from first to second seizure post-randomisation. The overall estimated cure rate for these survival times was 26%.

When considering the estimated values of the seizure rate modifiers following randomisation, we conclude that treatment policy does not appear to be statistically significant for those individuals with a normal EEG. Those with an abnormal EEG, but allocated to immediate treatment have estimated post-randomisation seizure rate modifiers in line with those presenting with a normal EEG. Additionally, for those individuals with an abnormal EEG and allocated to deferred treatment, the seizure rate modifiers following randomisation are generally not statistically different from unity, which is indicative of no change in seizure rate post-randomisation.

Following a first seizure post-randomisation we see that, in general, seizure rates increase. We observe that those individuals allocated to immediate treatment see a more substantial increase in seizure rate following a first seizure post-randomisation than those allocated to deferred treatment. This contrast is greater for those with an abnormal EEG. It has been proposed that a possible explanation for this is that those allocated to deferred treatment may subsequently start a course of AEDs following a seizure post-randomisation, bringing them in line with those allocated to immediate treatment. Also note that those individuals with an abnormal EEG and allocated to deferred treatment typically experience the smallest reduction in seizure rate following randomisation.

9.3 Conclusions about the Joint Models

We shall provide an overview of the assumptions that were made in the simple joint model. We shall also discuss how the model was extended to accommodate those assumptions that were violated in the MESS data. We shall also provide an overview of additional assumptions that are made in all of the joint models that have been developed in this thesis.

9.3.1 Assumptions in the Simple Joint Model

The simple joint model, developed in Chapter 5, assumed that individuals experience seizures according to a Poisson process, so that the pre-randomisation event counts follow a Poisson distribution and interevent times are Exponential. Each individual that was recruited to MESS had an underlying baseline seizure rate, which was updated at randomisation to allow for treatment ef-

fects. This simple joint model was used to analyse the MESS data, which subsequently cast doubt on assumptions that had been made.

Secondly, note that the simple joint model assumed that post-randomisation, seizure rates remained constant. It has been stated that clinicians believe that epileptic seizures beget epileptic seizures, that is, the more seizures an individual presents with, the more likely they are to carry on having seizures, with increased frequency. We also noted that, following the implementation of the simple joint model, the estimated regression coefficients observed in Joint Model B were closer to zero than those estimates observed in Joint Model A. This result has suggested that the assumption of a constant post-randomisation seizure rate may be violated.

The simple joint model assumed that, post-randomisation, everyone in the sample was susceptible to seizure recurrence. We know that seizure recurrence following a single untreated seizure is around $50\% - 80\%$, and as MESS was a study of early epilepsy it was not unreasonable to assume that a proportion of individuals included in the trial would be 'immune' from seizure recurrence post-randomisation. The magnitude of the seizure rate reductions observed following the implementation of the simple joint model highlighted that if survival data does indeed have a proportion that are immune to the event of interest, considering a proper survival model that ignores this, may give misleading results.

This thesis proceeded to address each of the assumptions discussed above first in isolation, and then together, in a full joint model.

9.3.2 Violation of the Post-Randomisation Survival Time IID Assumption

It was proposed that seizure rates may change not only at randomisation, but also following a first seizure post-randomisation. We considered a joint model that included seizure rate modifiers, both at randomisation and following a first post-randomisation seizure, and allowed these terms to depend on a number of explanatory variables.

We observed that at randomisation seizure rates either did not change, or decreased and following a first seizure post-randomisation, we saw that, in general, seizure rates increased.

9.3.3 Incorporation of Cure Rates

The magnitude of the reductions in seizure rates post-randomisation, observed in Chapter 5, suggested that there may be a substantial proportion of the population that we should regard as cured. Additionally, the Kaplan-Meier curves for each of the post-randomisation survival times have their asymptotes well above zero. This suggested that there may be an immune component present in the MESS data for both times to first seizure post-randomisation and from first to second seizure.

Models were developed that allowed for the inclusion of cure rates for each of the survival times.

9.3.4 Further Assumptions

Cowling (2003) outlines some of the assumptions made when developing a joint model that considers pre-randomisation event counts and a single post-randomisation survival time. Some of the assumptions associated with this joint model are also relevant when we consider the joint models that have been developed in this thesis.

First, all of the joint models assume that the seizure rate between successive seizures remains constant. Clinical opinion, however, is that following a seizure there is an instantaneous increase in the risk of future seizures, which is why we typically observe clustering of seizures. Recall that Figure 3.3 displayed the empirical cumulative distribution function for Y_1/T_2, and showed that for those experiencing at least two seizures post-randomisation, their time from first seizure to second was typically shorter than their time from randomisation to first seizure. It is thought that the immediate increase in risk following a seizure diminishes in time, but that in some cases of epilepsy the condition deteriorates after each event occurs. In the full joint model that we have developed, we have simply assumed that seizure rates change after each event, remaining constant between events. We may need to consider the possibility that seizure rates also change between events.

A further assumption that has been made is that there is an instantaneous multiplicative treatment effect at randomisation. Recall that MESS was a pragmatic trial, meaning that recruited individuals received treatment with AEDs in line with the clinicians' usual practice. When patients start a course

of AEDs, they typically start with smaller doses for an initial period, rather than starting immediately on the full dose. A consequence of this is that we may observe a delayed treatment effect which would need to be accounted for.

9.3.5 Further Work

Chapter 8 outlined further possible extensions to the analysis of the MESS data, that have been considered, but not carried out in this thesis. A zero-truncated, one-inflated Poisson distribution has been proposed as an alternative to the standard Poisson distribution, to reflect the fact that all individuals entering the trial had experienced at least one seizure pre-randomisation, although many presented with a single seizure only.

It was proposed that missing data methods could be adopted to recover information about the specific types of AED that were administered. This information could then be included as an explanatory variable in analyses. Additional outcomes that could be considered are times to third, fourth and fifth seizure post-randomisation, and times to one and two year remission.

One of the big questions in epilepsy is what causes drug refractoriness. We have stated in this thesis that around 30% of epilepsy sufferers will never achieve long-term remission from epileptic seizures. Clinical opinion is that there may be genetic determinants of this, as well as clinical ones. Epilepsy data exists that considers the outcomes time to first seizure and time to 12 month remission. This epilepsy dataset additionally contains information on individuals' DNA. It has been proposed that the models we have developed in

this thesis may be applicable to the analysis of this data.

Chapter 7 compared the performance of the full joint model with the Log-logistic model that incorporated cure rates, by considering how well the fitted estimates modelled the distribution of the survival curves. This comparison, however, did not formally assess the performance of the full joint model in its own right. Maller and Zhou (1996) propose using the correlation coefficient, for the Kaplan-Meier estimates of the cumulative distribution function and the corresponding fitted estimates from the full joint model, as a means of testing model performance. We have noted however, that correlation coefficients would simply assess the strength of a linear relationship between these two sets of data, but we wish to assess the strength of equality. Further work may consider statistics that may be appropriate for this.

9.4 Summary

In conclusion, this thesis has developed a number of statistical models that can be used in the analysis of data that arrives in the form of pre-randomisation event counts and two post-randomisation survival times. These joint models have been motivated by and illustrated on epilepsy data. The final full joint model that was concluded as optimal has been compared with standard survival models that incorporated cure rates. Despite initial indications being that the full joint model was superior to standard modelling strategies, visual comparisons of how the well the fitted estimates model the distribution of the survival times have shown that standard methods also perform very well.

It would be interesting to explore further if there are certain scenarios whereby the full joint model provides a significant improvement over standard survival methods.

Appendix A

Lomax Survival Distribution

Standard software for the analysis of survival data using the Lomax distribution does not exist. By considering the density and survivor functions of the Lomax distribution, we can derive the log-likelihood. First and second derivatives of this log-likelihood can easily be obtained, allowing inference on the parameters γ and $\boldsymbol{\theta}$, using a numerical method such as Newton Raphson.

A.1 The Full Log-Likelihood and Derivatives

Recall that the probability density function for the Lomax distribution is

$$f_Y(y_i;\mu_i,\gamma) = \mu_i\Big(\frac{\gamma}{\gamma+\mu_i y_i}\Big)^{\gamma+1}, \tag{A.1}$$

where $\mu_i = \exp(\boldsymbol{\theta}'\boldsymbol{d}_i)$ for a vector $\boldsymbol{\theta}$ of regression coefficients, and a vector $\boldsymbol{d}_i$ of covariates for each individual i, including an intercept term. It is trivial to derive the corresponding survivor function from the probability density

function:

$$\begin{aligned} S_Y(y_i;\mu_i,\gamma) &= \int_{y_i}^{\infty} f_Y(u;\mu_i,\gamma)du \\ &= \Big(\frac{\gamma}{\gamma+\mu_i y_i}\Big)^{\gamma}. \end{aligned} \tag{A.2}$$

Log-likelihood

Let δ_i be the indicator function for the survival time, taking the value 1 if the seizure is observed, and zero if the survival time is censored. Combining these indicator functions with density and survivor functions, (A.1) and (A.2) respectively, allows us to formulate the log-likelihood for the observed data $\mathcal{D}$ on all the n individuals, given by

$$\begin{aligned} \ell(\gamma,\boldsymbol{\theta} \mid \mathcal{D}) &= \sum_{i=1}^{n}\{\delta_i \ln f_Y(y_i;\mu_i,\gamma) + (1-\delta_i)\ln S_Y(y_i;\mu_i,\gamma)\} \\ &= \sum_{i=1}^{n}\{\delta_i \ln(\mu_i) + (\delta_i+\gamma)\ln(\gamma) - (\delta_i+\gamma)\ln(\gamma+\mu_i y_i)\} \end{aligned} \tag{A.3}$$

First Derivatives

The first-order derivatives of the full log-likelihood, (A.3) are:

$$\begin{aligned} \frac{\partial \ell}{\partial \boldsymbol{\theta}} &= \sum_{i=1}^{n}\left\{\frac{\gamma(\delta_i-\mu_i y_i)}{\gamma+\mu_i y_i}\right\}\boldsymbol{d}_i \\ \frac{\partial \ell}{\partial \gamma} &= \sum_{i=1}^{n}\left\{\ln(\gamma)+1+\frac{\delta_i}{\gamma}-\ln(\gamma+\mu_i y_i)-\frac{\delta_i+\gamma}{\gamma+\mu_i y_i}\right\} \end{aligned}$$

Second Derivatives

The second-order derivatives of the full log-likelihood are:

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}\prime} &= -\sum_{i=1}^{n} \left\{ \frac{\gamma \mu_i y_i (\delta_i + \gamma)}{(\gamma + \mu_i y_i)^2} \right\} \boldsymbol{d}_i \boldsymbol{d}_i\prime \\
\frac{\partial^2 \ell}{\partial \gamma \partial \gamma\prime} &= -\sum_{i=1}^{n} \left\{ \frac{1}{\gamma} - \frac{\delta_i}{\gamma^2} - \frac{1}{\gamma + \mu_i y_i} + \frac{\delta_i - \mu_i y_i}{(\gamma + \mu_i y_i)^2} \right\} \\
\frac{\partial^2 \ell}{\partial \boldsymbol{\theta} \partial \gamma} &= \sum_{i=1}^{n} \left\{ \frac{\mu_i y_i (\delta_i - \mu_i y_i)}{(\gamma + \mu_i y_i)^2} \right\} \boldsymbol{d}_i
\end{aligned}$$

Appendix B

R Code

B.1 Simple Joint Model

```
joint1 <- function(alphainit = 1, beta1initvec = 0, beta2initvec
 = c(0, 0), incl1 = 1, incl2 = 1, data, maxiter = 50)
{
#
#  Function to find maximum likelihood estimates for joint model
#  Data should be in a data frame including:
#  "nseiz" = pre-randomisation seizure count;
#  "period" = period in days from first pre-randomisation
#           seizure to randomisation
#  "time1" = post-randomisation time to first seizure;
#  "time2" = post-randomisation time from first seizure
#           to second seizure;
#  "cens1" = censoring indicator (first seizure);
#  "cens2" = censoring indicator (second seizure);
```

```
#  "type1" 0/1 tonic-clonic seizures only pre-randomisation;
#  "type2" 0/1 partial with 2-degree t-c pre-randomisation;
#  "type3" 0/1 generalised seizures pre-randomisation
#           (including tonic-clonic and generalised);
#  "type4" 0/1 partial minor seizures only pre-randomisation;
#  "type6" 0/1 other seizures pre-randomisation;
#  "ager"  = age at randomisation minus 30 as a continuous
#           covariate in years;
#  "sex"  0/1 indicating sex (male/female);
#  "trt"  0/1 indicating treatment (deferred/immediate).
#  "eeg"  0/1 indicating eeg outcome (normal/abnormal);
#
#  When calling the function, "incl1" and "incl2" decide which
#  covariates to include in lambda and psi respectively.
#
#
# the first section initialises the variables in the model
#
    k1 <- length(beta1initvec)
    k2 <- length(beta2initvec)
    beta1out <- matrix(rep(NA, k1 * maxiter), ncol = k1)
    beta2out <- matrix(rep(NA, k2 * maxiter), ncol = k2)
    alphaout <- matrix(rep(NA, maxiter), ncol = 1)
    beta1out[1,  ] <- beta1initvec
    beta2out[1,  ] <- beta2initvec
    alphaout[1,  ] <- alphainit
```

```
    maxx <- 1
    i <- 1
#
#
# the next section is a Newton-Raphson loop, repeatedly calling
# the function "joint2" until every estimate is within 0.00001
# of its value in the previous iteration
#
#
    while((i <= maxiter - 1) && (maxx > 1e-005)) {
        i <- i + 1
        newests <- joint2(alphaout[i - 1,  ], beta1out[i - 1,
            ], beta2out[i - 1,  ], incl1, incl2, data)
        alphaout[i,  ] <- newests$alpha
        beta1out[i,  ] <- newests$beta1
        beta2out[i,  ] <- newests$beta2
        maxad <- abs(alphaout[i,  ] - alphaout[i - 1,  ])
        maxbd1 <- max(abs(beta1out[i,  ] - beta1out[i - 1,  ]))
        maxbd2 <- max(abs(beta2out[i,  ] - beta2out[i - 1,  ]))
        maxx <- max(maxad, maxbd1, maxbd2)
    }
#
#
# the final section uses the maximum likelihood solution to
# generate the variance-covariance matrix (also using "joint2")
# and then output the estimates and related information
```

```
#
#
    newcov <- joint2(alphaout[i,  ], beta1out[i,  ], beta2out[i,
        ], incl1, incl2, data)
    list(alpha = alphaout[i,  ], beta1 = beta1out[i,  ], beta2
        = beta2out[i,  ], sd = round(sqrt(diag(newcov$covmat)),
        digits = 3), wald = round(c(beta1out[i,  ],beta2out[i,
        ])^2/diag(newcov$covmat)[-1],
        digits = 3), covmat = newcov$covmat, iter = i, loglik
         = newcov$loglik)
}


joint2 <- function(alpha, beta1vec, beta2vec, incl1,incl2, data)
{
#
# Function to help "joint1" by finding the log-likelihood,
# gradient, and Hessian, for a single Newton-Raphson iteration
#
# The first section initialises some parameters for the model,
# and reparameterises some covariates
#
    k1 <- length(beta1vec)
    k2 <- length(beta2vec)
    n <- length(data$type1)
    beta1 <- matrix(beta1vec, nrow = k1)
```

```
beta2 <- matrix(beta2vec, nrow = k2)
ager <- data$ager
type1 <- data$type1
type2 <- data$type2
type3 <- data$type3
type4 <- data$type4
type6 <- data$type6
trt <- data$trt
eeg <- data$eeg
sex <- data$sex
trttype1 <- trt * type1         # trt/type interactions
trttype2 <- trt * type2
trttype3 <- trt * type3
trttype4 <- trt * type4
trttype6 <- trt * type6
trtager <- trt * ager
eegtype1 <- eeg * type1         # eeg/type interactions
eegtype2 <- eeg * type2
eegtype3 <- eeg * type3
eegtype4 <- eeg * type4
eegtype6 <- eeg * type6
trteeg <- trt * eeg
cens1 <- data$cens1
time1 <- data$time1
cens2 <- data$cens2
time2 <- data$time2
```

```
    nseiz <- data$nseiz
    period <- data$period
#
#
# the next section uses the initially specified variables "incl1"
# and "incl2" to construct the covariate matrices which will
# later be used to give lambda and psi
#
#
    if(incl1 == 1)
        z1 <- matrix(rep(1, n), byrow = T, nrow = k1)
    if(incl1 == 2)
        z1 <- matrix(c(rep(1, n), type1, type2, type3, type6),
            byrow = T, nrow = k1)
    if(incl1 == 3)
        z1 <- matrix(c(rep(1, n), ager), byrow = T, nrow = k1)
    if(incl1 == 4)
        z1 <- matrix(c(rep(1, n), type1, type2, type3, type6,
            ager),
            byrow = T, nrow = k1)
    if(incl1 == 5)
        z1 <- matrix(c(rep(1, n), type1, type2, type3, type6,
            ager, sex), byrow = T, nrow = k1)
    if(incl1 == 6)
        z1 <- matrix(c(rep(1, n), type1, type2),
            byrow = T, nrow = k1)
```

```
#
    if(incl2 == 0)
        z2 <- matrix(rep(1, n), byrow = T, nrow = k2)
    if(incl2 == 1)
        z2 <- matrix(c(rep(1, n), trt), byrow = T, nrow = k2)
    if(incl2 == 2)
        z2 <- matrix(c(rep(1, n), trt, type1, type2, type3,
            type6), byrow = T, nrow = k2)
   if(incl2 == 3)
        z2 <- matrix(c(rep(1, n), trt, type1, type2, type3,
            type6, ager), byrow = T, nrow = k2)
    if(incl2 == 4)
        z2 <- matrix(c(rep(1, n), trt, type1, type2, type3,
            type6, ager, eeg), byrow = T, nrow = k2)
    if(incl2 == 5)
        z2 <- matrix(c(rep(1, n), trt, type1, type2, type3,
            type6, eeg), byrow = T, nrow = k2)
    if(incl2 == 6)
        z2 <- matrix(c(rep(1, n), trt, type1, type2, type3,
            type6, trttype1, trttype2, trttype3, trttype6,
            eeg, trteeg, eegtype1, eegtype2, eegtype3,
            eegtype6), byrow = T, nrow = k2)
    if(incl2 == 7)
        z2 <- matrix(c(rep(1, n), trt, type1, type2),
            byrow = T, nrow = k2)
    if(incl2 == 8)
```

```
            z2 <- matrix(c(rep(1, n), trt, type1, type2,
            trttype1, trttype2, eeg, trteeg, eegtype1,
            eegtype2), byrow = T, nrow = k2)
#
#
# the next section initialises the matrices and vectors
# which will store the values of the likelihood contributions,
# and the contributions to the gradient and Hessian, for
# each individual observation
#
#
    mat1 <- matrix(rep(0, k1 * k1), nrow = k1)  # for the Hessian
    mat2 <- matrix(rep(0, k2 * k2), nrow = k2)  # for the Hessian
    mat12 <- matrix(rep(0, k1 * k2), nrow = k1) # for the Hessian
    mat1a <- matrix(rep(0, k1), nrow = k1)      # for the Hessian
    mat2a <- matrix(rep(0, k2), nrow = k2)      # for the Hessian
#
    term1 <- matrix(rep(0, k1), nrow = k1)     # for the gradient
    term2 <- matrix(rep(0, k2), nrow = k2)     # for the gradient
#
    bigmat <- matrix(rep(0, (k1 + k2 + 1) * (k1 + k2 + 1)),
        nrow = k1 + k2 + 1)
    invbigmat <- bigmat     # for the observed information matrix
#
    aterm1 <- rep(NA, n)    # for the Hessian
    aterm2 <- aterm1        # for the gradient
```

```
    aterm3 <- rep(0, n)
    aterm4 <- aterm3
#
    ll <- rep(NA, n)
    llterm1 <- rep(0, n)
#
    lambda <- exp(t(beta1) %*% z1) # individual rate
    psi <- exp(t(beta2) %*% z2)    # treatment effect
#
    cens <- cens1 * (1 + cens2)
    time <- time1 + time2
    bit1 <- lambda * period
    bit2 <- lambda * psi * time
    bit3 <- bit1 + bit2 + alpha  #these "bits" come up a lot
#
#
# the next section is a loop for each observation in the data,
# calculating the individual contribution to the log-likelihood,
# the gradient, and the Hessian
#
#
    for(i in 1:n) {
#
# for Hessian and gradient contributions for beta1 and beta2
#
        mat1 <- mat1 - ((alpha * (nseiz[i] + cens[i] + alpha) *
```

```
            (bit1[i] + bit2[i]))/(bit3[i] * bit3[i])) * outer(
            z1[, i], z1[, i])
        mat2 <- mat2 - (((nseiz[i] + alpha + cens[i]) * (bit1[i]
            + alpha) * bit2[i])/(bit3[i] * bit3[i])) * outer(
            z2[, i], z2[, i])
        mat12 <- mat12 - ((alpha * (alpha + nseiz[i] + cens[i])
            * bit2[i])/(bit3[i] * bit3[i])) * outer(z1[, i],
            z2[, i])
#
        term1 <- term1 + ((alpha * (nseiz[i] + cens[i] -
            bit1[i] - bit2[i]))/bit3[i]) * z1[, i]
        term2 <- term2 + ((cens[i] * bit1[i] - nseiz[i] * bit2[i]
            + alpha * (cens[i] - bit2[i]))/bit3[i]) * z2[, i]
#
# for gradient and Hessian contributions for alpha
#
        for(j in 0:(nseiz[i] - 1)) {
                aterm3[i] <- aterm3[i] + 1/((alpha + j)^
                  2)
                aterm4[i] <- aterm4[i] + 1/(alpha + j)
            }
        aterm1[i] <- aterm3[i] + (cens1[i]/((alpha + nseiz[i]) *
            (alpha + nseiz[i]))) + ((cens1[i] * cens2[i])/
            ((alpha + nseiz[i] + 1) * (alpha + nseiz[i] + 1))) -
            (1/alpha) - ((nseiz[i] + cens[i] - alpha - 2 *
            (bit1[i] + bit2[i]))/(bit3[i] * bit3[i]))
```

```
        aterm2[i] <- aterm4[i] + (cens1[i]/(alpha + nseiz[i])) +
            ((cens1[i] * cens2[i])/(alpha + nseiz[i] + 1)) +
            log(alpha) + 1 - log(bit3[i]) - ((nseiz[i] +
            alpha + cens[i])/bit3[i])
#
# for Hessian contribution of correlation between alpha and beta
#
        mat1a <- mat1a + (((bit1[i] + bit2[i]) * (nseiz[i] +
            cens[i] - bit1[i] - bit2[i]))/(bit3[i] * bit3[i])) *
            z1[, i]
        mat2a <- mat2a + (((nseiz[i] + cens[i] - bit1[i] -
            bit2[i]) * bit2[i])/(bit3[i] * bit3[i])) * z2[, i]
#
# for log-likelihood contribution of individual i
#
        for(j in 0:(nseiz[i] - 1)) {
                llterm1[i] <- llterm1[i] + log(alpha +
                  j)
            }
        ll[i] <- llterm1[i] + cens1[i] * log(alpha + nseiz[i]) +
            (cens1[i] * cens2[i]) * log(alpha + nseiz[i] + 1) +
            nseiz[i] * log(period[i]) + alpha * log(alpha) +
            (nseiz[i] + cens[i]) * log(lambda[i]) + cens[i] *
            log(psi[i]) - lgamma(nseiz[i] + 1) - (nseiz[i] +
            alpha + cens[i]) * log(bit3[i])
    }
```

```
#
#
# the next section combines the individual second-derivative
# matrices into the Hessian, and then the observed information
# matrix.
#
#
    bigmat[2:(k1 + 1), 2:(k1 + 1)] <- mat1
    bigmat[2:(k1 + 1), (k1 + 2):(k1 + k2 + 1)] <- mat12
    bigmat[(k1 + 2):(k1 + k2 + 1), 2:(k1 + 1)] <- t(mat12)
    bigmat[(k1 + 2):(k1 + k2 + 1), (k1 + 2):(k1 + k2 + 1)] <-
       mat2
    bigmat[1, 1] <- 0 - sum(aterm1)
    bigmat[1, 2:(k1 + 1)] <- mat1a
    bigmat[1, (k1 + 2):(k1 + k2 + 1)] <- mat2a
    bigmat[2:(k1 + 1), 1] <- t(mat1a)
    bigmat[(k1 + 2):(k1 + k2 + 1), 1] <- t(mat2a)
#
    invbigmat <- solve(bigmat)
#
#
# the final section finds the updated parameter estimates using
# a Newton-Raphson step, and outputs the new parameter values,
# the observed information matrix, and the fitted log-likelihood
#
#
```

```
    newabvec <- t(c(alpha, beta1, beta2) - invbigmat %*% c(sum(
        aterm2), term1, term2))
#
    list(alpha = newabvec[, 1], beta1 = newabvec[, 2:(k1 + 1)],
        beta2 = newabvec[, (k1 + 2):(k1 + k2 + 1)], covmat =  -
        invbigmat, loglik = sum(ll))
}
```

B.2 Full Model

```
two_cure_rate1 <- function(alphainit, beta1init, beta2init,
 beta3init, kappa1init, kappa2init, incl1, incl2, incl5, incl3,
 incl4, data, maxiter = 50, iterations=500)
{
#
# the first section initialises the variables in the model
#
#
    k1 <- length(beta1init)
    k2 <- length(beta2init)
    k3 <- length(kappa1init)
    k4 <- length(kappa2init)
    k5 <- length(beta3init)
    beta1out <- matrix(rep(NA, k1 * maxiter), ncol = k1)
    beta2out <- matrix(rep(NA, k2 * maxiter), ncol = k2)
    beta3out <- matrix(rep(NA, k5 * maxiter), ncol = k5)
    kappa1out <- matrix(rep(NA, k3 * maxiter), ncol = k3)
    kappa2out <- matrix(rep(NA, k4 * maxiter), ncol = k4)
    alphaout <- matrix(rep(NA, maxiter), ncol = 1)
    beta1out[1,  ] <- beta1init
    beta2out[1,  ] <- beta2init
    kappa1out[1,  ] <- kappa1init
    kappa2out[1,  ] <- kappa2init
    alphaout[1,  ] <- alphainit
```

```
    beta3out[1,  ] <- beta3init
    maxx <- 1
    i <- 1
#
#
# the next section is an optim loop, repeatedly calling
# the function "two_cure2" until every estimate is within
# 0.00001 of its value in the previous iteration
#
#
    while((i <= maxiter - 1) && (maxx > 1e-005)) {
        i <- i + 1
        newests <- two_cure_rate2(alphaout[i - 1,  ], beta1out[i
            - 1,  ], beta2out[i - 1,  ], beta3out[i - 1,  ],
            kappa1out[i - 1,  ], kappa2out[i - 1,  ], incl1,
            incl2, incl5, incl3, incl4, data, iterations)
        alphaout[i,  ] <- newests$alpha
        beta1out[i,  ] <- newests$beta1
        beta2out[i,  ] <- newests$beta2
        beta3out[i,  ] <- newests$beta3
        kappa1out[i,  ] <- newests$kappa1
        kappa2out[i,  ] <- newests$kappa2
        maxad <- abs(alphaout[i,  ] - alphaout[i - 1,  ])
        maxbd1 <- max(abs(beta1out[i,  ] - beta1out[i - 1,  ]))
        maxbd2 <- max(abs(beta2out[i,  ] - beta2out[i - 1,  ]))
        maxbd3 <- max(abs(beta3out[i,  ] - beta3out[i - 1,  ]))
```

```
        maxkd1 <- max(abs(kappa1out[i,  ] - kappa1out[i - 1,
          ]))
        maxkd2 <- max(abs(kappa2out[i,  ] - kappa2out[i - 1,
          ]))
        maxx <- max(maxad, maxbd1, maxbd2, maxbd3, maxkd1,
        maxkd2)
    }
#
#
# the final section uses the maximum likelihood solution to
# generate the variance-covariance matrix (also using
# "two_cure2") and then output the estimates and related
# information
#
#
    newcov <- two_cure_rate2(alphaout[i,  ], beta1out[i,  ],
        beta2out[i,  ], beta3out[i,  ], kappa1out[i,  ],
        kappa2out[i,  ], incl1, incl2, incl5, incl3, incl4,
        data, iterations)
    list(alpha = alphaout[i,  ], beta1 = beta1out[i,  ], beta2 =
        beta2out[i,  ], beta3 = beta3out[i,  ], kappa1 =
        kappa1out[i,  ], kappa2 = kappa2out[i,  ], sd =
        round(sqrt(diag(newcov$covmat)),digits = 3), wald =
        round(c(beta1out[i,  ],beta2out[i,  ],beta3out[i,  ],
        kappa1out[i,  ], kappa2out[i, ])^2/diag(newcov$covmat)[
        -1], digits = 3), covmat = newcov$covmat, iter = i,
```

```
        loglik = newcov$loglik)
}


two_cure_rate2 <- function(alphain, beta1vec, beta2vec,
 beta3vec, kappa1vec, kappa2vec, incl1, incl2, incl5, incl3,
 incl4, data, iterations)
{
#
# Function to help "two_cure1" by finding the log-likelihood,
# and gradient, for a single optim iteration
#
#
#
# the first section initialises some parameters for the model,
# and reparameterises some covariates
#
#
    k1 <- length(beta1vec)
    k2 <- length(beta2vec)
    k3 <- length(kappa1vec)
    k4 <- length(kappa2vec)
    k5 <- length(beta3vec)
    n <- length(data$type1)
    ager <- data$ager
    type1 <- data$type1
```

```
    type2 <- data$type2
    type3 <- data$type3
    type4 <- data$type4
    type6 <- data$type6
    trt <- data$trt
    eeg <- data$eeg
    sex <- data$sex
    trttype1 <- trt * type1         # trt/type interactions
    trttype2 <- trt * type2
    trttype3 <- trt * type3
    trttype4 <- trt * type4
    trttype6 <- trt * type6
    trtager <- trt * ager
    eegtype1 <- eeg * type1         # eeg/type interactions
    eegtype2 <- eeg * type2
    eegtype3 <- eeg * type3
    eegtype4 <- eeg * type4
    eegtype6 <- eeg * type6
    trteeg <- trt * eeg
    cens1 <- data$cens1
    time1 <- data$time1
    cens2 <- data$cens2
    time2 <- data$time2
    nseiz <- data$nseiz
    period <- data$period
#
```

```
#
# the next section uses the initially specified variables "incl1"
# and "incl2" to construct the covariate matrices which will
# later be used to give lambda and psi
#
#
    if(incl1 == 1)
        z1 <- matrix(rep(1, n), byrow = T, nrow = k1)
    if(incl1 == 2)
        z1 <- matrix(c(rep(1, n), type1, type2, type3, type6),
            byrow = T, nrow = k1)
    if(incl1 == 3)
        z1 <- matrix(c(rep(1, n), ager), byrow = T, nrow = k1)
    if(incl1 == 4)
        z1 <- matrix(c(rep(1, n), type1, type2, type3, type6,
            ager), byrow = T, nrow = k1)
    if(incl1 == 5)
        z1 <- matrix(c(rep(1, n), type1, type2, type3, type6,
            ager, sex), byrow = T, nrow = k1)
    if(incl1 == 6)
        z1 <- matrix(c(rep(1, n), type1, type2), byrow = T,
        nrow = k1)
    if(incl1 == 7)
        z1 <- matrix(c(rep(1, n), type1, type3, type6),
            byrow = T, nrow = k1)
#
```

```
 if(incl2 == 0)
     z2 <- matrix(rep(1, n), byrow = T, nrow = k2)
 if(incl2 == 1)
     z2 <- matrix(c(rep(1, n), trt), byrow = T, nrow = k2)
 if(incl2 == 2)
     z2 <- matrix(c(rep(1, n), trt, type1, type2, type3,
         type6), byrow = T, nrow = k2)
if(incl2 == 3)
     z2 <- matrix(c(rep(1, n), trt, type1, type2, type3,
         type6, ager), byrow = T, nrow = k2)
 if(incl2 == 4)
     z2 <- matrix(c(rep(1, n), trt, type1, type2, type3,
         type6, ager, eeg), byrow = T, nrow = k2)
 if(incl2 == 5)
     z2 <- matrix(c(rep(1, n), trt, type1, type2, type3,
         type6, eeg), byrow = T, nrow = k2)
 if(incl2 == 6)
     z2 <- matrix(c(rep(1, n), trt, type1, type2, type3,
         type6, trttype1, trttype2, trttype3, trttype6, eeg,
         trteeg, eegtype1, eegtype2, eegtype3, eegtype6),
         byrow = T, nrow = k2)
 if(incl2 == 7)
     z2 <- matrix(c(rep(1, n), trt, type1, type2), byrow = T,
         nrow = k2)
 if(incl2 == 8)
     z2 <- matrix(c(rep(1, n), trt, type1, type2, trttype1,
```

```
            trttype2, eeg, trteeg, eegtype1, eegtype2),
            byrow = T, nrow = k2)
#
    if(incl3 == 0)
        w1 <- matrix(rep(1, n), byrow = T, nrow = k3)
    if(incl3 == 1)
        w1 <- matrix(c(rep(1, n), trt), byrow = T, nrow = k3)
    if(incl3 == 2)
        w1 <- matrix(c(rep(1, n), trt, type1, type2, type3,
            type6), byrow = T, nrow = k3)
   if(incl3 == 3)
        w1 <- matrix(c(rep(1, n), trt, type1, type2, type3,
            type6, ager), byrow = T, nrow = k3)
    if(incl3 == 4)
        w1 <- matrix(c(rep(1, n), trt, type1, type2, type3,
            type6, ager, eeg), byrow = T, nrow = k3)
    if(incl3 == 5)
        w1 <- matrix(c(rep(1, n), trt, type1, type2, type3,
            type6, eeg), byrow = T, nrow = k3)
    if(incl3 == 6)
        w1 <- matrix(c(rep(1, n), trt, type1, type2, type3,
            type6, trttype1, trttype2, trttype3, trttype6, eeg,
            trteeg, eegtype1, eegtype2, eegtype3, eegtype6),
            byrow = T, nrow = k3)
    if(incl3 == 7)
        w1 <- matrix(c(rep(1, n), trt, type1, type2), byrow = T,
```

```
            nrow = k3)
    if(incl3 == 8)
        w1 <- matrix(c(rep(1, n), trt, type1, type2, trttype1,
            trttype2, eeg, trteeg, eegtype1, eegtype2),
            byrow = T, nrow = k3)
    if(incl3 == 9)
        w1 <- matrix(c(rep(1, n), type1, type2, eeg, eegtype1,
            eegtype2), byrow = T, nrow = k3)
#
    if(incl4 == 0)
        w2 <- matrix(rep(1, n), byrow = T, nrow = k4)
    if(incl4 == 1)
        w2 <- matrix(c(rep(1, n), trt), byrow = T, nrow = k4)
    if(incl4 == 2)
        w2 <- matrix(c(rep(1, n), trt, type1, type2, type3,
            type6), byrow = T, nrow = k4)
   if(incl4 == 3)
        w2 <- matrix(c(rep(1, n), trt, type1, type2, type3,
            type6, ager), byrow = T, nrow = k4)
    if(incl4 == 4)
        w2 <- matrix(c(rep(1, n), trt, type1, type2, type3,
            type6, ager, eeg), byrow = T, nrow = k4)
    if(incl4 == 5)
        w2 <- matrix(c(rep(1, n), trt, type1, type2, type3,
            type6, eeg), byrow = T, nrow = k4)
    if(incl4 == 6)
```

```
        w2 <- matrix(c(rep(1, n), trt, type1, type2, type3,
            type6, trttype1, trttype2, trttype3, trttype6, eeg,
            trteeg, eegtype1, eegtype2, eegtype3, eegtype6),
            byrow = T, nrow = k4)
    if(incl4 == 7)
        w2 <- matrix(c(rep(1, n), trt, type1, type2), byrow = T,
            nrow = k4)
    if(incl4 == 8)
        w2 <- matrix(c(rep(1, n), trt, type1, type2, trttype1,
            trttype2, eeg, trteeg, eegtype1, eegtype2), byrow = T,
            nrow = k4)
#
    if(incl5 == 0)
        z3 <- matrix(rep(1, n), byrow = T, nrow = k5)
    if(incl5 == 1)
        z3 <- matrix(c(rep(1, n), trt), byrow = T, nrow = k5)
    if(incl5 == 2)
        z3 <- matrix(c(rep(1, n), trt, type1, type2, type3,
            type6), byrow = T, nrow = k5)
    if(incl5 == 3)
        z3 <- matrix(c(rep(1, n), trt, type1, type2, type3,
            type6, ager), byrow = T, nrow = k5)
    if(incl5 == 4)
        z3 <- matrix(c(rep(1, n), trt, type1, type2, type3,
            type6, ager, eeg), byrow = T, nrow = k5)
    if(incl5 == 5)
```

```
        z3 <- matrix(c(rep(1, n), trt, type1, type2, type3,
            type6, eeg), byrow = T, nrow = k5)
    if(incl5 == 6)
        z3 <- matrix(c(rep(1, n), trt, type1, type2, type3,
            type6, trttype1, trttype2, trttype3, trttype6, eeg,
            trteeg, eegtype1, eegtype2, eegtype3, eegtype6),
            byrow = T, nrow = k5)
    if(incl5 == 7)
        z3 <- matrix(c(rep(1, n), trt, type1, type2), byrow = T,
            nrow = k5)
    if(incl5 == 8)
        z3 <- matrix(c(rep(1, n), trt, type1, type2, trttype1,
            trttype2, eeg, trteeg, eegtype1, eegtype2),
            byrow = T, nrow = k5)
#
#
#
#
lambdaq<-exp(t(beta1vec)%*%z1)
psiq<-exp(t(beta2vec)%*%z2)
p1q<-exp(t(kappa1vec)%*%w1)/(1+exp(t(kappa1vec)%*%w1))
alphaq<-alphain
q <- p1q * ((1 + (lambdaq * psiq * time1)/alphaq)^(-alphaq))
#
#
like <- function(par){
```

```
  alpha <- par[1]
  beta1 <- matrix(par[2:(k1 + 1)], nrow = k1)
  beta2 <- matrix(par[(k1 + 2):(k1 + k2 + 1)], nrow = k2)
  beta3 <- matrix(par[(k1 + k2 + 2):(k1 + k2 + k5 + 1)],
    nrow = k5)
  kappa1 <- matrix(par[(k1 + k2 + k5 + 2):(k1 + k2 + k5 + k3
    + 1)], nrow = k3)
  kappa2 <- matrix(par[(k1 + k2 + k5 + k3 + 2):(k1 + k2 + k5
    + k3 + k4 + 1)], nrow = k4)
#
  ll <- rep(NA, n)
  llterm1 <- rep(0, n)
#
  lambda <- exp(t(beta1) %*% z1) # individual rate
  psi1 <- exp(t(beta2) %*% z2)    # treatment effect 1
  psi2 <- exp(t(beta3) %*% z3)    # treatment effect 2
  p1 <- exp(t(kappa1) %*% w1)/(1 + exp(t(kappa1) %*% w1))
  p2 <- exp(t(kappa2) %*% w2)/(1 + exp(t(kappa2) %*% w2))
   # p is the susceptible proportion
#
  cens <- cens1 * (1 + cens2)
  bit1 <- lambda * period
  bit2 <- lambda * psi1 * time1
  bit3 <- lambda * psi1 * psi2 * time2
  bit4 <- bit1 + bit2
  bit5 <- bit3 + bit4
```

```
    bit6 <- nseiz + alpha
    bit7 <- ((1 - p1)/((bit1 + alpha)^(bit6)))
    bit8 <- (p1/((bit4 + alpha)^(bit6)))
    bit9 <- ((1 - p2)/((bit4 + alpha)^(bit6 + 1))) + (p2/((bit5
        + alpha)^(bit6 + 1)))
        # these "bits" come up a lot
#
#
 for(i in 1:n) {
  for(j in 0:(nseiz[i] - 1)) {
                llterm1[i] <- llterm1[i] + log(alpha +
                  j)
            }
        ll[i] <- llterm1[i] + (nseiz[i] * log(period[i])) -
lgamma(nseiz[i] + 1) + (alpha * log(alpha)) + ((nseiz[i] +
cens[i]) * log(lambda[i])) + (cens[i] * log(psi1[i])) +
(cens1[i] * log(p1[i])) + (cens1[i] * cens2[i] * log(p2[i])) +
(cens1[i] * log(bit6[i])) + (cens1[i] * cens2[i] * log(bit6[i]
+ 1)) - (cens1[i] * cens2[i] * (bit6[i] + 2) * log(bit5[i] +
alpha)) + (cens1[i] * (1 - cens2[i]) * log(bit9[i])) + ((1 -
cens1[i]) * q[i] * log(bit8[i])) + ((1 - cens1[i]) * (1 - q[i])
* log(bit7[i])) + (cens1[i] * cens2[i] * log(psi2[i]))
    }
return(sum(ll))
}
#
```

```
#
#
#
grad <- function(par){
      alpha <- par[1]
      beta1 <- matrix(par[2:(k1 + 1)], nrow = k1)
      beta2 <- matrix(par[(k1 + 2):(k1 + k2 + 1)], nrow = k2)
      beta3 <- matrix(par[(k1 + k2 + 2):(k1 + k2 + k5 + 1)],
        nrow = k5)
      kappa1 <- matrix(par[(k1 + k2 + k5 + 2):(k1 + k2 + k5 + k3
        + 1)], nrow = k3)
      kappa2 <- matrix(par[(k1 + k2 + k5 + k3 + 2):(k1 + k2 + k5
        + k3 + k4 + 1)], nrow = k4)

#
    term1 <- matrix(rep(0, k1), nrow = k1)   # for the gradient
    term2 <- matrix(rep(0, k2), nrow = k2)   # for the gradient
    term3 <- matrix(rep(0, k5), nrow = k5)   # for the gradient
    kterm1 <- matrix(rep(0, k3), nrow = k3)   # for the gradient
    kterm2 <- matrix(rep(0, k4), nrow = k4)   # for the gradient
#
    aterm1 <- rep(NA, n)
    aterm2 <- rep(0, n)     # for the gradient
#
    lambda <- exp(t(beta1) %*% z1) # individual rate
    psi1 <- exp(t(beta2) %*% z2)    # treatment effect 1
```

```
    psi2 <- exp(t(beta3) %*% z3)    # treatment effect 2
    p1 <- exp(t(kappa1) %*% w1)/(1 + exp(t(kappa1) %*% w1))
    p2 <- exp(t(kappa2) %*% w2)/(1 + exp(t(kappa2) %*% w2))
     # p is the susceptible proportion
#
    cens <- cens1 * (1 + cens2)
    bit1 <- lambda * period
    bit2 <- lambda * psi1 * time1
    bit3 <- lambda * psi1 * psi2 * time2
    bit4 <- bit1 + bit2
    bit5 <- bit3 + bit4
    bit6 <- nseiz + alpha
    bit7 <- ((1 - p1)/((bit1 + alpha)^(bit6)))
    bit8 <- (p1/((bit4 + alpha)^(bit6)))
    bit9 <- ((1 - p2)/((bit4 + alpha)^(bit6 + 1))) + (p2/((bit5
        + alpha)^(bit6 + 1)))
        # these "bits" come up a lot
#
#
 for(i in 1:n) {
# for gradient contributions for beta1, beta2
#
      term1 <- term1 + (nseiz[i] + cens[i] - ((cens1[i] *
cens2[i] * (bit6[i] + 2) * bit5[i])/(bit5[i] + alpha)) - ((1 -
cens1[i]) * bit6[i] * (((q[i] * bit4[i])/(bit4[i] + alpha)) +
(((1 - q[i]) * bit1[i])/(bit1[i] + alpha)))) - ((cens1[i] * (1 -
```

```
cens2[i]) * (bit6[i] + 1) * ((((1 - p2[i]) * bit4[i])/((bit4[i]
+ alpha)^(bit6[i] + 2))) + ((p2[i] * bit5[i])/((bit5[i] + alpha)
^(bit6[i] + 2)))))/bit9[i])) * z1[, i]
#
      term2 <- term2 + (cens[i] - ((cens1[i] * cens2[i] *
(bit6[i] + 2) * (bit2[i] + bit3[i]))/(bit5[i] + alpha)) -
(((1 - cens1[i]) * q[i] * bit6[i] * bit2[i])/(bit4[i] + alpha)) -
((cens1[i] * (1 - cens2[i]) * (bit6[i] + 1) * ((((1 - p2[i]) *
bit2[i])/((bit4[i] + alpha)^(bit6[i] + 2))) + ((p2[i] * (bit2[i]
+ bit3[i]))/((bit5[i] + alpha)^(bit6[i] + 2)))))/bit9[i])) *
z2[, i]
#
kterm1 <- kterm1 + ((1 - p1[i]) - ((1 - cens1[i]) * (1 - q[i])))
* w1[, i]
#
kterm2 <- kterm2 + (cens1[i] * cens2[i] * (1 - p2[i]) -
((cens1[i] * (1 - cens2[i]) * p2[i] * (1 - p2[i]) * (((bit4[i] +
alpha)^(- bit6[i] - 1)) - ((bit5[i] + alpha)^(- bit6[i] - 1))))/
bit9[i])) * w2[, i]
#
term3 <- term3 + ((cens1[i] * cens2[i]) - ((cens1[i] * cens2[i] *
(bit6[i] + 2) * bit3[i])/(bit5[i] + alpha)) - (((cens1[i] * (1 -
cens2[i]) * (bit6[i] + 1) * p2[i] * bit3[i])/((bit5[i] + alpha)^
(bit6[i] + 2)))/bit9[i])) * z3[, i]
#
# for gradient contributions for alpha
```

```
#
        for(j in 0:(nseiz[i] - 1)) {
            aterm2[i] <- aterm2[i] + 1/(alpha + j)
            }
            aterm1[i] <- aterm2[i] + log(alpha) + 1 + (cens1[i]/
bit6[i]) + ((cens1[i] * cens2[i])/(bit6[i] + 1)) - (cens1[i] *
cens2[i] *
log(bit5[i] + alpha)) - ((cens1[i] * cens2[i] * (bit6[i] +
2))/(bit5[i] + alpha)) - ((cens1[i] * (1 - cens2[i]) *
((((((bit6[i] + 1)/(bit4[i] + alpha)) + log(bit4[i] + alpha)) * (1 -
p2[i]))/((bit4[i] + alpha)^(bit6[i] + 1))) + (((((bit6[i] + 1)/
(bit5[i] + alpha)) + log(bit5[i] + alpha)) * p2[i])/((bit5[i] + alpha)^
(bit6[i]+ joint1)))))/bit9[i]) - ((1 - cens1[i]) * q[i] * ((bit6[i]/
(bit4[i] + alpha)) + log(bit4[i] + alpha))) - ((1 - cens1[i]) * (1 -
q[i]) * ((bit6[i]/(bit1[i] + alpha)) + log(bit1[i] + alpha)))
#
    }
return(c(sum(aterm1),term1,term2,term3,kterm1,kterm2))
}
#
#
#
joint<-optim(c(alphain,beta1vec,beta2vec,beta3vec,kappa1vec,
 kappa2vec),like,grad,hessian=T,method='Nelder-Mead',
 control=list(maxit=iterations,fnscale=-1))
#
```

```
#
alpha = joint$par[1]
beta1 = joint$par[2:(k1 + 1)]
beta2 = joint$par[(k1 + 2):(k1 + k2 + 1)]
beta3 = joint$par[(k1 + k2 + 2):(k1 + k2 + k5 + 1)]
kappa1 = joint$par[(k1 + k2 + k5 + 2):(k1 + k2 + k5 + k3 + 1)]
kappa2 = joint$par[(k1 + k2 + k5 + k3 + 2):(k1 + k2 + k5 + k3 +
        k4 + 1)]
#
    list(alpha = alpha, beta1 = beta1, beta2 = beta2, beta3 =
beta3, kappa1 = kappa1, kappa2 = kappa2, sd =
round(sqrt(diag(-solve(joint$hessian))),digits = 3), wald =
round(c(beta1,beta2,beta3,kappa1,kappa2)^2/diag(-solve(joint$hessian))[
-1],digits = 3),covmat = -solve(joint$hessian), loglik = joint$value,
convergence = joint$convergence, message = joint$message)
}
```

Appendix C

Clinical Paper

In addition to the work that has been presented in this thesis, I have collaborated with clinicians at the University of Liverpool. As part of this collaboration, I am currently working on a clinical paper that analyses the MESS data, but attention is focussed on post-randomisation times to first seizure of any type and first tonic-clonic seizure. This paper is currently in the draft stage, but I include a brief summary here[1].

One of the most important decisions for a person with newly diagnosed epileptic seizures will be whether to start treatment with an antiepileptic drug (AED). This will be dependent on an analysis that requires consideration of the risk of seizures on the one hand and the side effects associated AED treatment on the other. Existing literature tends to focuss primarily on the risks associated tonic-clonic seizures in patients who have presented with a tonic clonic seizure previously. For patients who have received a diagnosis of

[1]The most up to date draft of the clinical paper discussed here can be found at http://www.warwick.ac.uk/go/jenniferrogers/research/thesis.
This page is password protected, the password is 'thesisrogers2010'.

epilepsy, but have never experienced a tonic-clonic seizure, information about the risk of future tonic-clonic seizures may be clinically relevant. The MRC Multicentre Trial for Early Epilepsy and Single Seizures (MESS) was a randomized controlled trial which compared the treatment policies of immediate and deferred treatment with AEDs, in those patients considered to be in the early stages of epilepsy. Because of its broad entry criteria, MESS contained a number of patients with a history of partial seizures only. In this paper we explore the outcomes time to first seizure of any type and time to first tonic-clonic seizure in these patients, and compare them with that of other subjects who had at least one tonic-clonic seizure before randomisation.

The data arrives in the two parts: a pre-randomisation seizure count, along with the associated number of days over which these seizures were observed, and post-randomisation survival times to first seizure of any type and first tonic-clonic seizure. We adopt methodology that allows the pre-randomisation seizure counts and post-randomisation survival times to be jointly modelled (Cowling et al. 2006). This method assumes that both these outcomes are predicted by (unobserved) seizure rates, assuming that each patient has an underlying constant seizure rate that we allow to vary depending on the following baseline attributes: age at randomisation, sex and seizure type. Additionally we suppose that the post-randomisation seizure rates will be reduced relative to the baseline seizure rate. A greater reduction in the seizure rate results in a longer time to seizure post-randomisation, indicating a better therapy. We have modified this methodology however, to incorporate the inclusion of possible cure rates.

The data indicate that the risk of secondary generalised tonic-clonic seizures in those patients that have presented with partial seizures only pre-randomisation is low, regardless of EEG outcome. As would be expected, immediate treatment with AEDs also had little benefit in reducing the incidence of secondary generalised tonic-clonic seizures post-randomisation. For this reason, the main issue to be considered in this group of patients will be the effect of immediate treatment on the frequency of their partial seizures. When this is examined, treatment policy is not statistically significant in reducing post-randomisation seizure rates for those individuals presenting with partial seizures only. For those groups of patients with tonic-clonic seizures pre-randomisation, immediate treatment is favoured. This observation is in keeping with the hypothesis that partial seizures are generally more resistant to AEDs than tonic-clonic seizures, be these generalised at onset or secondarily generalised. Indeed, it could be argued that the main effects of currently available AEDs are to limit the spread of seizure discharge within the brain, rather than to prevent the initiation of seizures.

The findings that we have observed may have some regulatory implications. There has been some debate about the licensing of new AEDs. These are brought to market through trials that compare add-on drugs with a placebo in subjects with a history of pharmacologically resistant partial seizures. The key outcomes will be the reduction in seizure frequency compared to placebo, but the patients included are likely to have many more simple or complex partial seizures than secondary generalised seizures. Thus it can be asked whether such trials provide reasonable evidence of effectiveness against secondary generalised seizures.

Bibliography

Berg, A. T., S. F. Berkovic, M. J. Brodie, J. Buchhalter, J. H. Cross, W. van Emde Boas, J. Engel, J. French, T. A. Glauser, G. W. Mathern, S. L. Mosh, D. Nordli, P. Plouin, and I. E. Scheffer (2010). Revised terminology and concepts for organization of seizures and epilepsies: Report of the ILAE commission on classification and terminology, 2005-2009. *Epilepsia 51*(4), 676–685.

Berg, A. T. and S. Shinnar (1991). The risk of seizure recurrence following a first unprovoked seizure: A quantitative review. *Neurology 41*(7), 965–972.

Chandra, B. (1992). First seizure in adults: to treat or not to treat. *Clinical Neurology and Neurosurgery 94(Suppl.)*, S61–S63.

Cockerell, O. C., A. L. Johnson, J. W. A. S. Sander, Y. M. Hart, and S. D. Shorvon (1995). Remission of epilepsy: results from the national general practice study of epilepsy. *The Lancet 346*, 140–144.

Collett, D. (2003). *Modelling Survival Data in Medical Research* (2 ed.). Texts in Statistical Science. Chapman and Hall/CRC CRC Press.

Cook, R. J. and J. F. Lawless (2007). *The Statistical Analysis of Recurrent Events*. Statistics for Biology and Health. Springer.

Cowling, B. J. (2003). *Survival models for censored point processes*. PhD Thesis, Department of Statistics, University of Warwick, Coventry.

Cowling, B. J., J. L. Hutton, and J. E. H. Shaw (2006). Joint modelling of event counts and survival times. *J. R. Statist. Soc. C 55*(1), 31–39.

Cox, D. R. (1972). Regression models and life-tables. *J. R. Statist. Soc. B 34*(2), 187–220.

Filliben, J. J. (1975). The probability plot correlation coefficient test for normality. *Technometrics 17*(1), 111–117.

Finney, D. J. and G. C. Varley (1955). An example of the truncated Poisson distribution. *Biometrics 11*(3), 387–394.

Greenwood, M. and G. U. Yule (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurence of multiple attacks of disease or of repeated accidents. *Journal of the Royal Statistical Society 83*(2), 255–279.

Hocking, R. R. (1976). A Biometrics invited paper. the analysis and selection of variables in linear regression. *Biometrics 32*(1), 1–49.

Hougaard, P., M.-L. Lee, and G. A. Whitmore (1997). Analysis of overdispersed count data by mixtures of Poisson variables and Poisson processes. *Biometrics 53*(4), 1225–1238.

ILAE Commission on Antiepileptic Drugs (1998). Considerations on designing clinical trials to evaluate the place of new antiepileptic drugs in the treatment of newly diagnosed and chronic patients with epilepsy. *Epilepsia 39*(7), 799–803.

Johnson, N. L. and S. Kotz (1970). *Distributions in Statistics: Continuous*

Univariate Distributions, Volume 1 (1 ed.). Wiley series in Probability and Statistics. John Wiley and Sons.

Kim, L., A. Johnson, A. Marson, and D. Chadwick (2006). Prediction of risk of seizure recurrence after a single seizure and early epilepsy: further results from the MESS trial. *The Lancet Neurology 5*, 317–322.

Lin, D. Y., W. Sun, and Z. Ying (1999). Nonparametric estimation of the gap time distributions for serial events with censored data. *Biometrika 86*(1), 59–70.

Little, R. J. A. and D. B. Rubin (2002). *Statistical Analysis with Missing Data* (2 ed.). Wiley series in Probability and Statistics. John Wiley and Sons.

Maller, R. and X. Zhou (1996). *Survival Analysis with Long-Term Survivors*. Wiley series in Probability and Statistics. John Wiley and Sons.

Marson, A., A. Jacoby, A. Johnson, L. Kim, C. Gamble, and D. Chadwick (2005). Immediate versus deferred antiepileptic drug treatment for early epilepsy and single seizures: a randomised control trial. *The Lancet 365*, 2007–2013.

McCullagh, P. and J. A. Nelder (1989). *Generalised Linear Models* (2 ed.). Monographs on Statistics and Applied Probability. Chapman and Hall.

Nayak, T. K. (1987). Multivariate Lomax distribution: properties and usefulness in reliability theory. *Journal of Applied Probability 24*, 170–177.

Peng, Y. gfcure. (Available from http://post.queensu.ca/∼pengp/ software.html). Downloaded 14 July 2009.

Peto, R. and J. Peto (1972). Asymptotically efficient rank invariant test

procedures. *J. R. Statist. Soc. A 135*(2), 185–207.

Rogers, J. K., J. L. Hutton, and K. Hemming (2009). *Joint Modelling of Event Counts and Survival Times*. CRiSM Working Paper 44, University of Warwick, (Available from http://www2.warwick.ac.uk/fac/sci/statistics/crism/research/2009/paper09-44).

Sahamotoa, Y., M. Ishiguro, and G. Kitagawa (1986). *Akaike Information Criterion Statistics*. KFT Scientific.

Schafer, J. L. and J. W. Graham (2002). Missing data: Our view of the state of the art. *Psychological Methods 7*(2), 147–177.

Verity, C. M., G. Hosking, and D. J. Easter (1995). A multicentre comparitive trial of sodium valproate and carbamazepine in paediatric epilepsy. *Developmental Medicine and Child Neurology 37*(2), 97–108.

Visser, M. (1996). Nonparametric estimation of the bivariate survival function with an application to vertically transmitted AIDS. *Biometrika 83*(3), 507–518.

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society 54*(3), 426–482.

Warrell, D. A., T. M. Cox, J. D. Firth, and E. J. Benz (Eds.) (2003). *Oxford Textbook of Medicine* (4 ed.), Volume 3. Oxford University Press.

Wei, L. J. (1992). The accelerated failure time model: A useful alternative to the Cox regression model in survival analysis. *Statistics in Medicine 11*(14-15), 1871–1879.