Survival Models

for Censored Point Processes

by

Benjamin John Cowling

A thesis submitted for the degree of

Doctor of Philosophy

University of Warwick, Department of Statistics

November 2003

Contents

List of Figures

List of Tables

0.1 Acknowledgements

I would like to acknowledge the sponsorship provided by the Engineering and Physical Sciences Research Council. I would like to thank my supervisors, Jane Hutton and Ewart Shaw, for all their advice and motivation. The support of my wife Pauline, my family and friends is also gratefully acknowledged.

I would also like to acknowledge Paula Williamson, Tony Marson and David Chadwick as co-grant holders of the data set used for illustration.

0.2 Declaration

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.

0.3 Summary

In studies of recurrent events, there can be a lot of information about a cohort over a period of time, but it may not be possible to extract as much information from the data as would be liked. This thesis considers data on individuals experiencing recurrent events, before and after they are randomised to treatment. The pre-randomisation outcome is a period count, while the post-randomisation outcome is a survival time. Standard survival analysis may treat the pre-randomisation period count as a covariate, but it is proposed that point process models will give a more precise estimate of the treatment effect.

A joint model is presented, based on a Poisson process with individual frailty. The pre-randomisation seizure counts are distributed as Poisson variables with rate depending on explanatory variables as well as a random frailty. The model for the post-randomisation survival times is the exponential distribution with the same individual seizure rate, modified by a multiplicative treatment effect. A conjugate mixing distribution (frailty) is used, and alternative mixing distributions are also discussed.

The model is motivated by and illustrated on individual patient data from five randomised trials of two treatments for epilepsy. The data are presented, and the standard analyses are contrasted with the results of the joint model.

This thesis also considers the relative efficiency of the joint model compared to other survival models. Finally, some extensions to the model are considered, including a more general non-conjugate mixing distribution, and alternative ways of including explanatory variables in the joint model.

0.4 General Notation

n	Number of individuals in study population
$i = 1, \ldots, n$	Subscript for individuals
X	a period event count
x	observed value of X
Y	a survival time
y	observed value of Y
$\mathbf{Z}$	a matrix of explanatory variables (covariates)
$\mathbf{z}$	observed value of $\mathbf{Z}$
$\mathbf{Z}_1$	a matrix containing general covariate information
$\mathbf{z}_1$	observed value of $\mathbf{Z}_1$
$\mathbf{Z}_2$	a matrix containing treatment covariate information
$\mathbf{z}_2$	observed value of $\mathbf{Z}_2$

Chapter 1

Introduction

This thesis considers a new model for longitudinal data which include period counts and survival times. The work is motivated by individual patient data from five randomised trials of two common treatments for epilepsy, included in Marson *et al.* (2002). A baseline pre-randomisation seizure count is recorded, as well as the individuals' times to first post-randomisation seizure, as an internationally agreed outcome (ILAE Commission on Antiepileptic Drugs, 1998). Primary interest lies in the contrast between the treatment effects, and a possible interaction between the treatment and the covariates *age* and *epilepsy type*. Standard survival analysis may treat the pre-randomisation period count as a covariate, specifically as a fixed covariate (Verity *et al.*, 1995; Kwong & Hutton, 2003). However, such a model ignores the variation in counts within individuals. It is suggested that it is preferable to consider the pre-randomisation count as a second outcome, rather than an explanatory variable.

It is interesting to consider the analysis of data which are a mixture of counts and

times, because recurrent event data sometimes come in this form. The endpoint of medical trials is often defined as a time-to-event outcome, even for treatment of a recurrent event. If such data were collected, then information on previous event history could be used sensibly to give a more precise estimate of the treatment effects. In addition, the interpretation of this type of model is very appealing, as will be demonstrated later. Data of this form may be found in healthcare (e.g. treatments for asthma, HIV, chronic granulotamous disease, epilepsy), engineering, psychology and economics.

1.1 Overview of Thesis

In chapter 2, an overview is given of the current literature on count models, and survival models, and models for a mixture of longitudinal data and survival data, which is a developing area. One important consideration is the distinction between *true* and *apparent contagion*. These two alternatives are the underlying causes of overdispersion in count data, which cannot be attributed to known explanatory variables: if there is *true contagion*, then the overdispersion is caused because the events are clustered; while if there is *apparent contagion*, then the overdispersion is caused by unexplained differences between individuals (as in 'frailty' models).

Chapter 3 gives an overview of the epilepsy data, including information on the distribution of covariates. Standard parametric and non-parametric analyses of the data are also presented.

In chapter 4, a joint model for data consisting of pre-randomisation event counts and post-randomisation survival times is derived and discussed. The full log-

likelihood is given with first- and second-derivatives, and a maximum likelihood approach is suggested. A corresponding Bayesian model is also presented, and the use of MCMC methods is discussed.

The results of the joint model applied to the epilepsy data are given in chapter 5. To investigate the model fit, some diagnostics are presented. In addition, the data are reanalysed excluding data from one of the original trials, and with a reclassification scheme for the covariate *epilepsy type*, which is suspected to have been misclassified for some individuals.

Chapter 6 investigates the relative efficiency of the joint model compared to a related survival model. The results of a simulation study are presented, and a theoretical approach is discussed.

In chapter 7, an extension to the joint model is investigated, that is, using a more general non-conjugate family of distributions for the frailty. The power variance family of Hougaard (1986b) is described, and the full log-likelihood for a count model with this frailty distribution is derived. Such a count model, with covariates, has not been well covered in the literature. Generalising the joint model of chapter 4 by incorporating the power variance family as the mixing distribution is discussed, but not presented.

In chapter 8, another extension to the joint model is investigated. Here, the model is modified to allow covariates to affect the shape of the mixing distribution. The new model is illustrated on two subsets of the epilepsy data, using MCMC inference. Also considered in chapter 8 is the problem of a missing informative covariate. The missing covariate is simulated, and the results are discussed.

Finally, chapter 9 concludes the thesis, and gives a lengthy discussion of the strengths and weaknesses of the joint model, and the suitability of the underlying assumptions. The analyses of the epilepsy data are compared and contrasted. Extensions to the model are discussed, and areas for further work are suggested.

Appendix A contains information on the Pareto survival distribution, including the log-likelihood and derivatives. Appendix B contains more detailed information about the simulation study presented in chapter 6. Appendix C contains the `s-plus` functions used to fit the maximum likelihood joint model, and appendix D contains the WinBUGS (Spiegelhalter *et al.*, 2000) code used to implement MCMC for the Bayesian joint model.

Chapter 2

Literature Review

There is a great amount of literature on survival analysis, and the analysis of count data. A developing area in the literature is models for the joint analysis of longitudinal data and survival data. This chapter describes the current literature in each of these areas.

2.1 Analysis of Count Data

A well-known choice of model to apply to count data is the Poisson Generalised Linear Model (McCullagh & Nelder, 1989). To account for the overdispersion in the counts, a random mixture distribution may be applied to the mean. The most convenient choice of mixture distribution is the gamma, which leads to the negative binomial distribution for the counts (Greenwood & Yule, 1920). Many other mixture distributions have been suggested, including the inverse Gaussian distribution (Dean *et al.*, 1989), and log-normal (Shaban, 1988).

An interesting family of distributions is the power variance family, described by Hougaard (1986b), which includes the gamma, positive stable, and inverse Gaussian distributions as special cases. Hougaard, Lee and Whitmore (1997) apply this distribution to data on counts of epileptic seizures, but without allowing for covariate effects.

Luceño (1995) creates overdispersion in a Poisson model by assuming that events are clustered. Gourieroux and Visser (1997) introduce heterogeneity through the individual exponential waiting times making up the count distribution. Winkelmann (1995) derives a count model based on an underlying point process with gamma waiting times, and finds that under this model, overdispersion in the counts occurs if the waiting time distribution has decreasing hazard. Toscas and Faddy (2003) generalise a Poisson process by changing the transition probabilities, to give an overdispersed Poisson distribution for period counts.

Hougaard (2000) gives a comprehensive discussion of various choices of mixture distribution for models of overdispersed period count data. An extensive discussion of the analysis of count data is given by Cameron and Trivedi (1998). Diggle *et al.* (2002) and Clayton (1994) give good overviews of the analysis of recurrent event data.

2.1.1 Analysis of Multivariate Longitudinal Count Data

Cameron and Trivedi (1998) suggest methods for investigating the heterogeneity in a bivariate Poisson distribution, and they prefer to have the heterogeneity components between two times as correlated, but not identical.

There is a large amount of literature concerning the analysis of a combination of pre-randomisation and post-randomisation event counts, for example the epilepsy data described by Thall and Vail (1990), and the subsequent re-analyses (Zeger & Liang, 1992; Lindsey, 1993; Diggle *et al.*, 2002).

Marshall and Olkin (1990) generate a bivariate negative binomial distribution by using a univariate heterogeneity term, using the following relationship:

$$f(y_1,\, y_2 \mid z_1,\, z_2) = \int_0^\infty f_1(y_1 \mid z_1,\, \nu)\, f_2(y_2 \mid z_2,\, \nu)\, g(\nu)\, \mathrm{d}\nu,$$

where y_1 and y_2 are both counts, f_1 and f_2 are univariate densities, and ν may be interpreted as common unobserved heterogeneity affecting both counts. They let $f_1(y_1)$ and $f_2(y_2)$ be Poisson with parameters $\mu_1\nu$ and $\mu_2\nu$ respectively, where ν has gamma distribution with parameter α.

Diggle *et al.* (2002), Cook and Lawless (2002), and Clayton (1994) give good overviews of the analysis of repeated measures and recurrent event data. Also related is the theory of point processes (Daley & Vere-Jones, 1988; Cox & Isham, 1980), and renewal processes as described by Smith (1958) and Cox (1962). At present, renewal theory is widely used in the modelling of stochastic failure processes.

2.1.2 True and Apparent Contagion

There is much discussion in the statistical literature about the recognition of the distinction between true and apparent contagion. The definitions of these two conditions are given below.

True contagion is the occurrence of an event which affects the probability of a subsequent event (unlike a Poisson process), and so there is a dependence between the occurrence of successive events. If the occurrence of an event shortens the expected waiting time for the next occurrence of an event (and so the events are clustered), that is known as *true positive contagion.* The reverse case is known as *true negative contagion.*

Apparent contagion is when the sampled individuals come from a heterogeneous population in which individuals have constant but differing propensity to experience events, and this difference cannot be explained solely by the covariates, as in *frailty models*. For a given individual, occurrence of an event does not make it more or less likely that another event will occur.

Feller (1943) observed that the same negative binomial model had been derived by Greenwood and Yule (1920) under the assumption of population heterogeneity (apparent contagion), and by Eggenberger and Polya (1923) under the assumption of true contagion. He noted that it is therefore possible to interpret the negative binomial distribution in two ways, which are quite different in their nature as well as their implications. To differentiate between true and apparent contagion, longitudinal data are required.

The joint model proposed in chapter 4 will assume that there is apparent contagion, but not true contagion, in the recurrent event process.

2.2 Analysis of Survival Data

A typical analysis of the epilepsy data might apply standard survival techniques to the post-randomisation times alone, treating the pre-randomisation event counts as a covariate. Good overviews of standard methods for the analysis of survival data are given in Cox and Oakes (1984), Klein and Moeschberger (1997), and Collett (2003).

In survival data, it is often the case that there is additional variance between individuals, which cannot be attributed to the explanatory variables alone. This is known as heterogeneity in the sample. For a broad discussion of heterogeneity in survival analysis, see Aalen (1988), and Pickles and Crouchley (1995).

Survival analysis can also be put in the framework of counting processes, a thorough description is given by Andersen *et al.* (1993).

2.2.1 Robust Model Selection

The literature contains various suggestions for model selection procedures. Automatic routines such as *forward selection*, *backward selection*, and the combination of these, *stepwise selection*, are implemented in the survival models in statistical packages such as `s-plus` and `SAS`. Collett (2003, pp. 81-3) notes some of the disadvantages of these automated procedures, and describes a general strategy for model selection. The strategy is based around comparing the values of the deviance ($-2\log\widehat{L}$) of nested models with a prespecified amount, to decide whether or not to include a particular explanatory variable, or covariate. The procedure is

summarised in four steps below:

- Fit the model just using one covariate at a time, and record which variables significantly decrease the deviance.[1] Call the recorded set of variables P.

- Fit the model including all the variables P, and then exclude one variable at a time. Keep only the variables which give a significant increase in the deviance when they are excluded from the model. Some variables may cease to be important in the presence of other variables. If more than one variable is non-significant, the variable giving the least raise in deviance when excluded should be omitted first, and the whole step repeated, until a set Q is obtained where leaving out any of the variables in Q will give a significant increase in the deviance.

- Starting with the variables Q, add all other variables one at a time, to see if any now give a significant reduction in the deviance. Interaction terms may also be included at this stage, making sure that all necessary lower-order terms are also included in the model. Combine Q with all the variables selected by this procedure, to form a set R.

- Finally, consider the variables in R to check if the omission of any will lead to a significant increase in the deviance. Repeat this step if any variables are selected for exclusion. The resulting set S of variables is the final selection of this procedure.

[1] A typical measure of a significant decrease (at the 95% level) is a decrease of more than 2 or 3 times the difference in degrees of freedom between the two models; some flexibility in the selection rule should be allowed.

2.2.2 Analysis of Multivariate Survival Data

In the literature, there do not seem to be examples of the analysis of pre- and post-randomisation times together, or in the language of renewal theory, backward and forward recurrence times. However, bivariate survival analysis (Oakes, 1982, 1989) is in some ways similar.

Lindeboom and van den Berg (1994) consider bivariate survival models in which the dependence between two survival times is by way of stochastically related unobserved components. Their results suggest that it may be hazardous to estimate bivariate survival models in which the mixing distribution is univariate. This is because a univariate random variable may not be able to account both for the mutual dependence of the survival times and for the change in the composition of the sample over time due to unobserved heterogeneity.

Hougaard (1987) gives a good overview of the analysis of multivariate survival data, and also discusses some aspects of recurrent event data in the form of counts, and Poisson mixture models. Hougaard (2000) gives a comprehensive discussion of multivariate survival analysis.

Prentice *et al.* (1981) describe a stratified proportional hazards model, to model recurrent event data where a small number of failure times are recorded for a large number of individuals. They relate the underlying hazard or intensity function to covariates, and preceding failure time history.

2.3 Joint Modelling of Longitudinal Data and Survival Data

Diggle *et al.* (2002, ch. 14) give some references of literature describing the joint modelling of recurrent event or repeated measures data, with survival data. This is an area of active research. These models use clinical information on a repeated measure (such as measures on a biomarker over time), or a recurrent event, to infer a distribution on the time to some other clinical event.

A typical example is the work of Xu and Zeger (2001), who are interested in the time to discontinuation of a treatment for schizophrenia, and use the patients' PANSS score to try to estimate the discontinuation time. The PANSS score is related to the severity of the patients' symptoms. They use a latent variable model, where the recurrent event process is modelled by a GLM with linear predictor following a Gaussian stochastic process (Diggle, 1988). They use an MCMC algorithm to make inference on the parameters. Another example is the work of Faucett and Thomas (1996), who describe MCMC methods to jointly model a repeatedly measured covariate (CD4 count) and censored survival data (time to onset of AIDS, for HIV patients).

No literature has so far been found on the joint modelling of event counts and event times, which is the form of the epilepsy data motivating this thesis. One possibility for the epilepsy data would be to treat the counts as a covariate measured with error (Carroll *et al.*, 1995), in a survival model.

2.4 Bayesian Methods

With the advent of more accessible powerful computing, Bayesian methods have become more popular. Gilks *et al.* (1996) give a good overview of Markov Chain Monte Carlo methods. For a general reference on the application of Bayesian methods, see Congdon (2001). Gamerman (1997) considers the application of Bayesian methods to generalised linear mixed models. Ibrahim *et al.* (2001) gives a good overview of applications of Bayesian methods to survival data. The software package WinBUGS (Spiegelhalter *et al.*, 2000) may be used to implement MCMC methods.

In Bayesian methods, the choice of prior is a very important consideration, and there is substantial discussion on this in the literature. Natarajan and Kass (2000) suggest that a shrinkage prior would be the best choice of prior for second stage variance components, but the advantage over a vague prior is fairly small with only a univariate random effect.

2.5 Non-Parametric Frailty Models

One alternative to the parametric frailty models discussed above is to use a non-parametric frailty. Recent years have seen a lot of development in Bayesian non-parametric methods, for example Polya trees (Lavine, 1992, 1994). Walker and Mallick (1997) describe the use of Polya trees to model the random mixing distribution in hierarchical generalised linear models, as an alternative to parametric frailties. When they apply their method to the kidney infection data of McGilchrist

and Aisbett (1991) they discover that the frailty distribution is bimodal, due to a difference between the sexes.

These methods seems very useful, particularly for exploratory analysis of data, where the estimated non-parametric frailty distribution could be used to specify a sensible parametric frailty model, and for model-checking, where the estimated non-parametric frailty can be used to check for bimodality or other problems with the parametric frailty. For predictive purposes, it would seem that the use of a parametric frailty is preferable, where possible.

Frequentist non-parametric frailty models, on the other hand, are not so well developed. Walker and Malick (1997) describe some of these methods, which are based on the work of Laird (1978).

2.6 Discussion

Standard models for the analysis of longitudinal data and survival data have been discussed in this chapter. An area of active research is in models for the joint analysis of longitudinal data comprising a repeated measure and a survival time. However, no literature has yet been found on the joint analysis of a period count followed by a survival time, for a single recurrent event process. A typical survival analysis would treat the count as a fixed covariate (Kwong & Hutton, 2003). It is proposed that a more sensible analysis of such data should try to model the process as a whole, treating the period count and the survival time as dual outcomes.

Chapter 3

Introduction to the Epilepsy Data

This thesis is motivated by the individual patient data from five trials comparing *sodium valproate* (VPS) with *carbamazepine* (CBZ), as initial treatments for epilepsy, as described in Marson *et al.* (2002). In this chapter, an overview of the data is given, and the results of some standard analyses are presented. The results of standard survival models fitted to the first post-randomisation seizure times are also presented in Kwong and Hutton (2003).

3.1 Overview

The International Dictionary of Medicine and Biology (Landau, 1986) defines epilepsy as *"a neurological disorder characterized by the tendency to suffer recurrent seizures or fits, whether minor or major."*

A recent meta-analysis by Marson *et al.* (2002) used individual patient data from

randomised controlled trials of two drugs, carbamazepine (CBZ) and valproate (VPA), given to newly diagnosed patients with either partial-onset epilepsies (also known as focal epilepsies) or generalised-onset epilepsies. The authors found some evidence to support the prior clinical belief of an interaction between treatment and epilepsy type (Wallace *et al.*, 1997), when they took the outcome as 'time to first post-randomisation seizure'. The authors also investigated two other measured outcomes, 'time to 12 month remission', and 'time to withdrawal from treatment'.

This thesis considers the individual patient data from five of the larger trials included in the meta-analysis of Marson *et al.* (2002), comprising 1225 individuals in total:

- Trial 1: Heller *et al.* (1995);
- Trial 2: De Silva *et al.* (1996);
- Trial 3: Richens *et al.* (1994);
- Trial 4: Verity *et al.* (1995);
- Trial 5: Mattson *et al.* (1992).

Informative covariates include a binary indicator of the *type of epilepsy* of the individual (generalised-onset or partial-onset); *sex*; *age at randomisation*; and a factor indicating which of the five trials the individual took part in. In terms of seizure information, a 6-month pre-randomisation seizure count was recorded, followed by the time to first post-randomisation seizure (possibly censored).

In the data of Marson *et al.*, the individuals have been classified into two broad epilepsy syndromes. It is noted that there is a possibility of misclassification of the epilepsy types of individuals, indeed Williamson *et al.* (2002) investigated this problem in the same data.

It has been decided to excluded some individuals from the analyses in this thesis, due to missing values, or because they are clearly outliers. Thirty-nine (3%) with missing pre-randomisation seizure counts had to be excluded, with the majority of these individuals from the fifth (Mattson) trial. A further 3 individuals with missing ages were also excluded. Sixteen individuals with first post-randomisation seizure times of less than 1 day were excluded as outliers. In addition, 23 of the remaining individuals with 6-month pre-randomisation seizure counts of 100 or more were chosen to be excluded, as outliers. This choice was made because the individuals with very large counts were found to have a disproportional effect on the results of the following analyses. Thus a subset of size 1144 is studied, with no missing information in this subset.

Thirty-five of the thirty-nine individuals with missing pre-randomisation seizure counts are from the fifth (Mattson) trial, and are therefore all individuals with partial epilepsy, and are generally older men. In addition, the decision to exclude 'outliers' means that there are some issues with the representativeness of the chosen subset of 1144 individuals. However, the subset does contain 93% of the original sample, so the results are of clinical interest.

3.2 Distribution of Variables

In this section, the distribution and association between the important variables are investigated. Table 3.1 gives some information about the distribution of ages, sexes, and types of epilepsy. The fifth trial (Mattson *et al.*, 1992) is clearly different to the other four trials, because it contains mainly older men, all with partial-onset epilepsies.

Table 3.2 gives a summary of the distribution of pre-randomisation seizure counts, by *epilepsy type* and randomised treatment. The data show that individuals with partial-onset epilepsies typically have seizures more frequently than individuals with generalised-onset epilepsies. Two histograms illustrating the distribution of 6-month pre-randomisation seizure counts are presented later, in figure 7.1 on page 124.

Table 3.3 gives a summary of the distribution of pre-randomisation seizure counts, by trial. The fifth trial is different to the other four trials, with only a little overdispersion in the seizure counts.

The association between age at randomisation and epilepsy type is illustrated in figure 3.1. According to clinicians, generalised-onset epilepsies typically arise in childhood, and should rarely, if ever, be diagnosed in adults over the age of 30. There is a suggestion that many individuals have misclassified epilepsy types in this data, since 22% of individuals with generalised-onset epilepsies have an age of onset greater than 30. Some investigation int this misclassification is given in Williamson *et al.* (2002), and is also considered in Section 5.5 on page 58.

Table 3.1: Distribution of aetiological covariates.

Trial	n	mean age at onset (s.d.)	% male	% partial
1	115	30.9 (15.0)	50.4	37.4
2	87	10.3 (3.6)	48.3	48.3
3	282	33.4 (15.1)	50.7	51.4
4	235	10.1 (2.9)	46.8	44.2
5	425	46.9 (16.4)	92.7	100.0
Total	1144	31.6 (19.9)	65.3	66.3

Table 3.2: Distribution of 6-month pre-randomisation seizure counts by epilepsy type and drug.

Type	Drug	n	6-month Pre-randomisation count				
			mean	s.d.	median	min.	max.
general'd	CBZ	196	4.61	8.61	3	0	99
general'd	VPS	189	5.86	11.41	2	0	98
partial	CBZ	372	8.70	16.84	4	0	99
partial	VPS	387	8.75	16.87	4	0	99
	Total	1144	7.55	15.00	3	0	99

Table 3.3: Distribution of 6-month pre-randomisation seizure counts by trial.

Trial	n	6-month Pre-randomisation count				
		mean	s.d.	median	min.	max.
1	115	8.33	15.63	2	0	89
2	87	11.17	16.59	4	0	98
3	282	9.17	14.15	4	2	98
4	235	11.15	24.06	3	1	99
5	425	3.52	2.27	3	1	10
Total	1144	7.55	15.00	3	0	99

Table 3.4: Distribution of post-randomisation times to first seizure.

Type	Drug	%	Times to first post-randomisation seizure				
		obs.	mean	s.d.	median	min.	max.
general'd	CBZ	71.9	517.6	654.1	249	1	4070
general'd	VPS	65.1	635.2	824.9	282	1	4520
partial	CBZ	67.2	354.5	520.3	76	1	2348
partial	VPS	73.6	269.8	463.4	49	1	2704
	Total	69.8	400.2	602.7	103	1	4520

'% obs.' = percentage of individuals with observed survival times.

Figure 3.1: Box-plot of age by epilepsy type.

3.3 Kaplan-Meier Results

In this section, non-parametric estimates of the survival function are considered. Figure 3.2 shows a Kaplan-Meier plot of the times to first seizure, where individuals are classified into four groups by epilepsy type and by the treatment given. This plot shows some evidence for an interaction between treatment and epilepsy type. A steep initial drop in the estimated survival curves is very noticeable. In figure 3.3, the Kaplan-Meier plot for the first year after randomisation is presented. The lines for the two treatments for generalised-onset epilepsy lie close together, but the lines for the two treatments for partial-onset epilepsy are further apart.

A log-rank test for a difference between the 4 survival curves gives a χ^2-value of 35 on 3 degrees of freedom, which is highly significant ($p < 0.001$). A test only for individuals with generalised-onset epilepsies gives no evidence of a difference

between CBZ and VPS ($p = 0.23$). On the other hand, considering only individuals with partial-onset epilepsies, there is strong evidence that CBZ is superior to VPS ($p < 0.01$). These results might be expected by looking at figure 3.3.

The Kaplan-Meier estimates of the 1-, 3-, 6- and 12-month 'survival' rates are shown in table 3.5.

Table 3.5: Proportion of individuals who have not experienced any post-randomisation seizures, after 1, 3, 6 and 12 months.

Type	Drug	Proportion 'surviving'			
		1 month	3 months	6 months	12 months
generalised	CBZ	0.82	0.67	0.55	0.47
generalised	VPS	0.79	0.68	0.57	0.48
partial	CBZ	0.70	0.54	0.45	0.37
partial	VPS	0.61	0.44	0.35	0.29
	Total	0.70	0.55	0.45	0.38

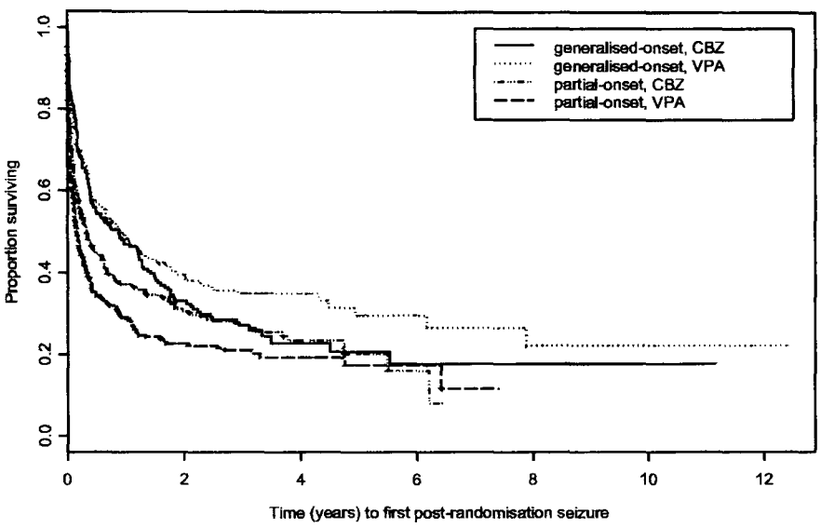

Figure 3.2: Plot of the Kaplan-Meier Estimate of the survivor function, stratified by epilepsy type and treatment.

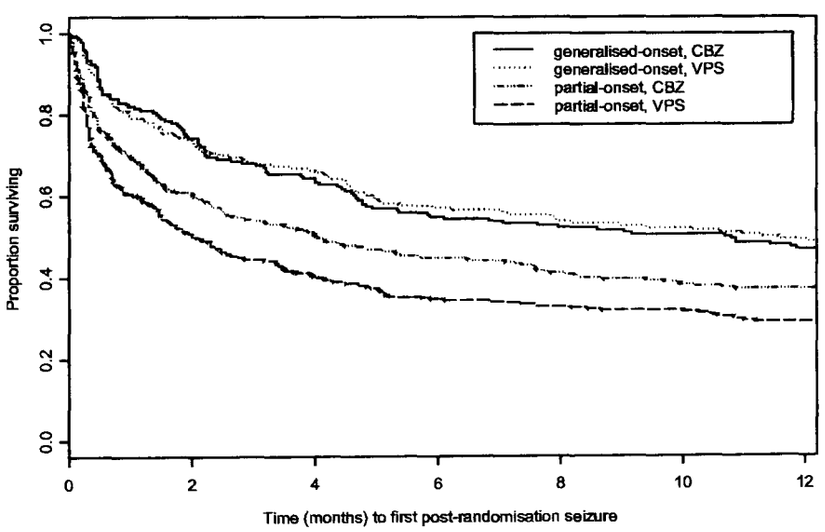

Figure 3.3: Plot of the Kaplan-Meier Estimate of the survivor function, stratified by epilepsy type and treatment, for the first year of randomisation.

3.4 Analysis of Pre-Randomisation Counts

The pre-randomisation seizure counts X_i may be modelled by a Poisson distribution. To account for variability in the counts, explanatory variables may be incorporated in the model, and a standard way to do this is with a generalised linear model (McCullagh & Nelder, 1989). The Poisson GLM uses a log-link to relate the covariates to the mean event count. However, the Poisson distribution specifies that the mean is the same as the variance, but often count data are overdispersed, and a common modification is to incorporate a random effect in the mean. Using a gamma random effect gives the negative binomial distribution.

The negative binomial model may be specified by the equations:

$$f(x_i \mid \lambda_i, u_i, \nu_i) = \frac{(\lambda_i u_i \nu_i)^{x_i} \exp(-\lambda_i u_i \nu_i)}{x_i!},$$

$$g(\nu_i \mid \alpha) = \frac{\alpha^\alpha \nu_i^{\alpha-1} \exp(-\alpha \nu_i)}{\Gamma(\alpha)},$$

where

$$\lambda_i = \exp(\boldsymbol{\beta}_1' \mathbf{z}_{1i}).$$

Here $\mathbf{z}_{1i}$ is a vector of covariates for individual i, and $\boldsymbol{\beta}_1$ is a vector of regression coefficients, including an intercept term. For all individuals $u_i = 182$, and the Poisson case (with no overdispersion) arises when $\alpha \to \infty$, that is, $\nu_i = 1$ for all individuals i.

Both models may be applied to the epilepsy data, and the maximum likelihood

Table 3.6: Estimates (standard errors) for Poisson and negative binomial GLM

Regression Coefficient	Poisson GLM estimate (s.e.)	NB GLM estimate (s.e.)
α	∞	1.221 (0.055)
β_0	−3.093 (0.033)	−3.059 (0.092)
β_{type}	0.541 (0.013)	0.557 (0.037)
β_{age}	0.035 (0.009)	0.025 (0.022)
β_{trial2}	0.257 (0.050)	0.385 (0.147)
β_{trial3}	−0.059 (0.038)	−0.130 (0.110)
β_{trial4}	0.296 (0.043)	0.189 (0.122)
β_{trial5}	−1.447 (0.044)	−1.479 (0.119)
−Log-likelihood (df)	7489 (1137)	3311 (1136)

Type: $-1/+1$ for generalised/partial-onset epilepsy

Age: original age $-$ 30, in decades

estimates are given in table 3.6. The large drop in log-likelihood for just one extra parameter shows that the negative binomial provides a much better fit than the Poisson. The small value of $\widehat{\alpha}$ shows considerable heterogeneity.

It is noted that the negative binomial GLM can be improved upon, and this topic is explored further in chapter 7. The advantage of the negative binomial is that the likelihood and its derivatives are tractable, which is not the case with many other typical models for count data with covariates.

3.5 Analysis of Post-Randomisation Times

A variety of parametric accelerated failure time (AFT) models may be fitted to the post-randomisation times to first seizure. This section considers three typical survival models, specified by the equations below. The models are the exponential (3.1), Weibull (3.2) and the Pareto (3.3).

$$f(y_i \mid \mu_i) = \mu_i \exp(-\mu_i y_i); \tag{3.1}$$

$$f(y_i \mid \mu_i, \gamma) = \mu_i \gamma y_i^{\gamma-1} \exp(-\mu_i y_i^{\gamma}); \tag{3.2}$$

$$f(y_i \mid \mu_i, \gamma) = \mu_i \left(\frac{\gamma}{\gamma + \mu_i y_i}\right)^{\gamma+1}; \tag{3.3}$$

where in each model $\mu_i = \exp(\boldsymbol{\theta}' \mathbf{w}_i)$ for a vector $\boldsymbol{\theta}$ of regression coefficients, and a vector $\mathbf{w}_i$ of covariates for each individual i including an intercept term. The parameter γ in models (3.2) and (3.3) is a scale parameter.

The parameter estimates for these three survival models are presented in table 3.7. It is noted that gamma, log-logistic and log-normal survival models give similar results to the Pareto model presented here (Kwong & Hutton, 2003, p. 156). In their paper, Kwong and Hutton prefer to include an interaction between *treatment* and *age* than an interaction between *treatment* and *epilepsy type*. As noted elsewhere (Williamson *et al.*, 2002), there is a problem with the misclassification of *epilepsy type* in these data, and it is also known that *age at randomisation* is strongly associated with *epilepsy type*.

The results in table 3.7 show some evidence for a treatment-type interaction,

specifically a beneficial effect of VPS over CBZ, for generalised epilepsies, and a beneficial effect of CBZ over VPS, for partial epilepsies. However, in the model which fits best, the Pareto model, this interaction is non-significant. Goodness-of-fit diagnostics reveal that these distributions do not fit the data particularly well, mainly because they cannot model the steep initial drop in the survivor function, as shown in figure 3.2 on page 23. Kwong and Hutton (2003) also fit proportional hazards models to the times to first seizure, but conclude that parametric accelerated life models are more suitable for these data.

Table 3.7: Estimates (standard errors) for typical survival models fitted to the times to first seizure

Regression Coefficient	Exponential Estimate (s.e.)	Weibull Estimate (s.e.)	Pareto Estimate (s.e.)
θ_0	−6.984 (0.074)	−3.345 (0.151)	−5.109 (0.245)
$\theta_{\log(count)}$	0.406 (0.022)	0.302 (0.036)	0.540 (0.067)
θ_{type}	0.184 (0.032)	0.138 (0.047)	0.412 (0.097)
θ_{age}	−0.144 (0.018)	−0.098 (0.027)	−0.171 (0.059)
θ_{trial2}	0.188 (0.095)	0.101 (0.167)	−0.127 (0.364)
θ_{trial3}	−0.150 (0.082)	−0.210 (0.135)	−0.055 (0.282)
θ_{trial4}	−0.158 (0.089)	−0.228 (0.146)	−0.245 (0.313)
θ_{trial5}	0.509 (0.084)	0.175 (0.146)	0.663 (0.299)
θ_{trt}	−0.004 (0.026)	0.008 (0.038)	0.101 (0.080)
$\theta_{trt\times type}$	0.182 (0.026)	0.115 (0.038)	0.153 (0.080)
Scale	1 (0)	0.482 (0.014)	0.364 (0.021)
−Log-lik. (df)	5736 (1134)	5269 (1133)	5179 (1133)

Trt: $-1/+1$ for CBZ/VPS

3.6 Discussion

The epilepsy data contain information on over 1200 individuals randomised to two common treatments for epilepsy, *carbamazepine* (CBZ) and *sodium valproate* (VPS). The data is largely complete, although 39 individuals had the missing outcome of a 6-month pre-randomisation seizure count. A small number of other individuals were chosen to be excluded as outliers. It is noted that although the results are not completely robust to alternative arbitrary cut-offs for outliers (such as excluding all counts larger than 90), the conclusions are not altered by such a change.

Current clinical belief is that VPS is the preferred treatment for generalised-onset epilepsies, and CBZ is the preferred treatment for partial-onset epilepsies (Wallace *et al.*, 1997). However, this hypothesis has not yet been proven beyond reasonable doubt, and studies comparing the two treatments are ongoing. The original meta-analysis of Marson *et al.* (2002) found only some evidence to support the clinical belief. Kwong and Hutton (2003) fitted a variety of typical survival models to the times to first post-randomisation seizure, and found some evidence that VPS is better for younger patients, while CBZ is better for older patients. However, it is known that *age at randomisation* is confounded by *epilepsy type*, and their paper does not discuss this problem.

The non-parametric Kaplan-Meier estimates of the survival curves, in figures 3.2 and 3.3 on page 23 show some evidence in favour of VPS being the preferred treatment for generalised-onset epilepsies, and strong evidence that CBZ is the preferred treatment for partial-onset epilepsies.

In table 3.7, the parameter estimates of standard survival models show that there is only some evidence for a treatment-type interaction. In the best-fitting survival models, the Pareto, gamma, log-logistic and log-normal models, the interaction term is barely significant.

The Kaplan-Meier estimates indicated that the interaction may be unbalanced, that is, the improvement of CBZ over VPS for partial-onset epilepsies is *greater* than the improvement of VPS over CBZ for generalised-onset epilepsies. Therefore it might be useful to consider the two epilepsy syndromes separately.

One final consideration is that the standard survival models have included the (log-transformed) pre-randomisation seizure count as a covariate. However, this information is really an outcome rather than an explanatory variable, and it would be preferable to use it as such. In the next chapter, a joint model based on a Poisson process is derived for these data, and the pre-randomisation seizure count and the post-randomisation time to first seizure are modelled as dual outcomes.

Chapter 4

A Joint Model for Event Data

The motivation for this thesis is individual patient data from a randomised trial of two treatments for an illness which causes recurrent events. Associated with each individual i ($i = 1, \ldots, n$) in the study, there is an event count X_i, over a pre-randomisation time period u_i. Also recorded is the time, Y_i, from randomisation to the first post-randomisation event with a censoring indicator δ_i ($\delta_i = 1$ indicates that Y_i is observed, while $\delta_i = 0$ indicates censoring). For each individual there is also background information and a treatment indicator. In this chapter a joint model is derived, for data of this form.

Figure 4.1 below shows an example of a recurrent event process. Each of the eight lines represents an individual, with an 'x' marking an event time. The individuals have been randomised over two treatments, in a controlled trial. Observation begins at 6 months prior to the start of the trial, and each individual's event history is recorded until randomisation, and then for a set period after randomisation, during which time only the time until the first event occurs, for each individual, is

observed. If no event has occurred by the end of the trial, or if that individual is lost to follow up, then the time at which they were last known to be event free is marked with an 'o'.

In figure 4.1, a dashed line in the trial period represents treatment with drug 'A', and a dotted line represents treatment with drug 'B'. In a controlled trial, it would be expected that both treatments would have some effect in reducing the underlying event rate, so interest lies in devising a model to assess which treatment is more effective in this respect.

It is of interest to consider data which are a mixture of counts and times in this way, because recurrent event data sometimes come in this form. The endpoint of medical trials is often defined as a time-to-event outcome, even for treatment of a recurrent event. Indeed, in studies of epilepsy, time to first post-randomisation seizure is an internationally agreed outcome (ILAE Commission on Antiepileptic Drugs, 1998). Data of this form may be found in healthcare (e.g. treatments for asthma, HIV, chronic granulotamous disease, epilepsy), engineering, psychology and economics.

4.0.1 Building a Joint Model

The simplest model for such data is a homogeneous Poisson process. That is, all individuals experience events according to a Poisson process with rate λ. The event count X_i for individual i will then be Poisson with mean λu_i, and with no treatment effect, the inter-event times will be exponential with rate λ. Because of the memoryless property of the exponential distribution, the post-randomisation

seizure time Y_i would also be exponential with the same rate λ.

However, count data are often overdispersed, that is, the variance is greater than the mean. Some of this overdispersion may be attributed to the covariates such as *age at randomisation* and *sex*, and thus the rate may be allowed to vary with the covariates, so that the rate for individual i is λ_i, where λ_i depends in some way on that individual's covariates. However, there may remain some additional unexplained variance, perhaps due to heterogeneity in the population. This is known as *apparent contagion* (p. 7). A common model for overdispersed count data is the negative binomial distribution (Greenwood & Yule, 1920), where each individual experiences events according to a Poisson process with event rate $\lambda_i \nu_i$, where λ_i depends on the covariates, and ν_i is a random term, which follows a gamma dis-

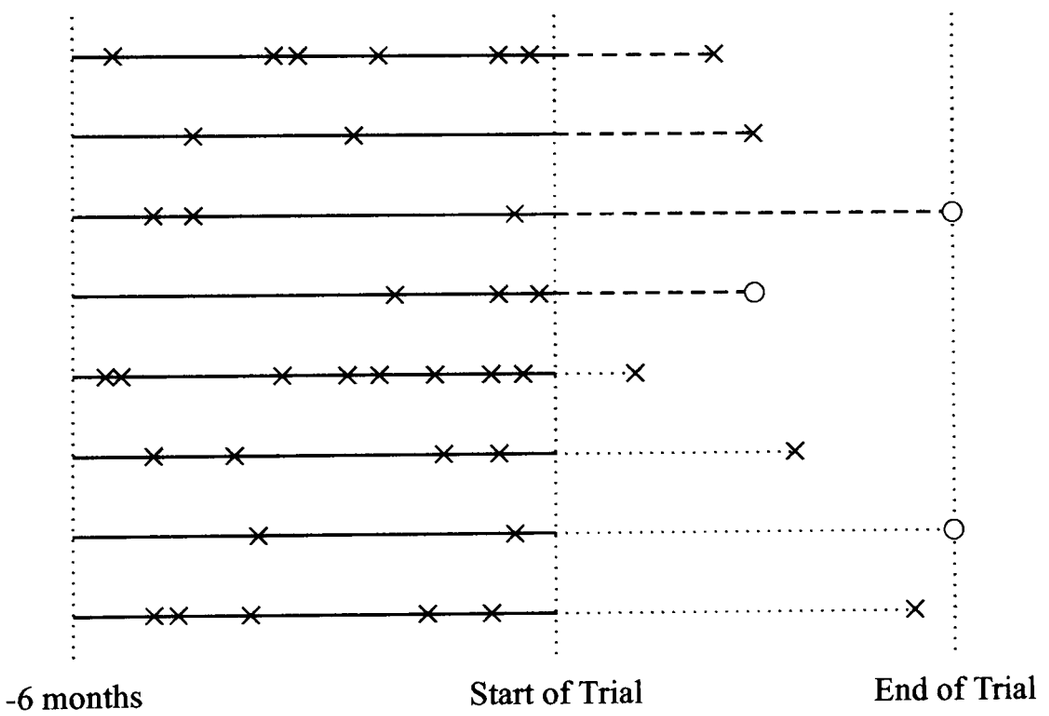

Figure 4.1: Example of Data

tribution. Let the important explanatory variables be entered in a covariate $\mathbf{Z}_{1i}$. Then relating λ_i to $\mathbf{Z}_{1i}$ using a log-link gives the negative binomial Generalised Linear Model (McCullagh & Nelder, 1989).

If the underlying point process were modelled as a Poisson process with individual rate $\lambda_i\nu_i$, where λ_i depends on the covariates of individual i, while ν_i is random, then an inter-event time would be exponential with the same rate. However, this joint model must allow for a treatment effect. It is assumed that the treatment acts *multiplicatively* on the event rate. That is, the event rate for individual i will become $\lambda_i\psi_i\nu_i$, where ψ_i depends in some way on the treatment information. A log-link is used to relate a treatment covariate $\mathbf{Z}_{2i}$ to the multiplicative factor ψ_i, and further work could consider alternative assumptions for the impact of treatment on the event rate. The treatment covariate $\mathbf{Z}_{2i}$ will contain an intercept term as well as a treatment indicator, and may also contain other explanatory variables and interaction terms.

It is well known that one derivation of the Pareto distribution is as a gamma mixture of exponentials (see also appendix A on page 172). Here, the unconditional distribution of Y_i is Pareto, with survivor function $S(y_i \mid \lambda_i, \psi_i, \alpha) = (1 + \lambda_i\psi_i y_i/\alpha)^{-\alpha}$. The hazard function is $h(y_i \mid \lambda_i, \psi_i, \alpha) = \alpha\lambda_i\psi_i/(\alpha + \lambda_i\psi_i y_i)$, which is always decreasing.

Figure 4.2 shows a graphical model of the process described. The circular nodes represent variables, either parameters or data, and the square nodes are logical.

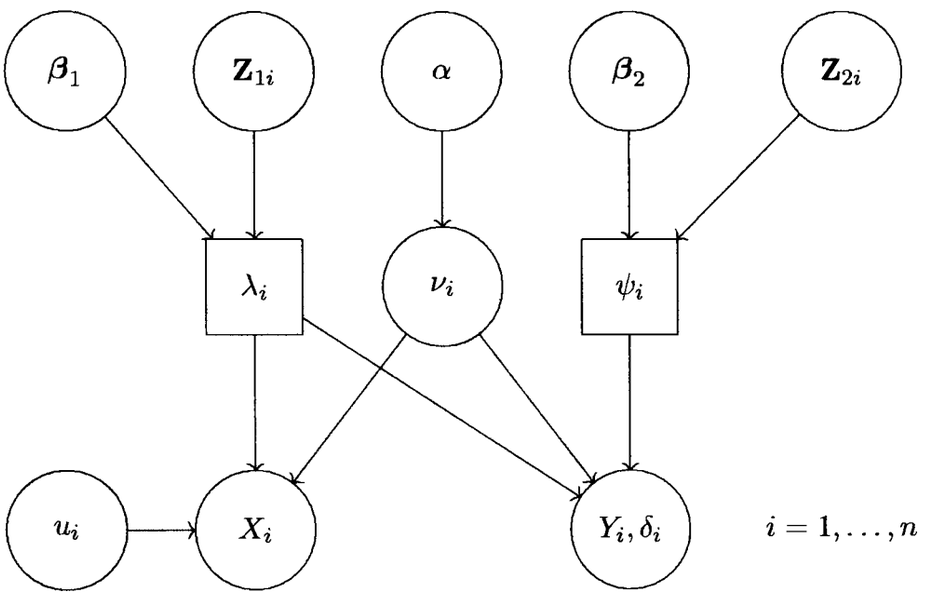

Figure 4.2: Graphical Model of the underlying point process

The model is specified by the equations:

$$f_X(x_i \mid \lambda_i,\, u_i,\, \nu_i) = \frac{(\lambda_i u_i \nu_i)^{x_i} \exp(-\lambda_i u_i \nu_i)}{x_i!} \tag{4.1}$$

$$f_Y(y_i \mid \lambda_i,\, \psi_i,\, \nu_i) = \lambda_i \psi_i \nu_i \exp(-\lambda_i \psi_i \nu_i y_i) \tag{4.2}$$

$$g_\nu(\nu_i \mid \alpha) = \frac{\alpha^\alpha \nu_i^{\alpha-1} \exp(-\alpha \nu_i)}{\Gamma(\alpha)},$$

where

$$\lambda_i = \exp(\boldsymbol{\beta}_1' \mathbf{z}_{1i}),$$

$$\psi_i = \exp(\boldsymbol{\beta}_2' \mathbf{z}_{2i}).$$

Here, $\alpha > 0$ measures the degree of heterogeneity (a large value of α would indicate only a small amount of heterogeneity). The mean of the heterogeneity

term, ν_i, is fixed to 1 for identifiability. The parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are vectors of regression coefficients, $\mathbf{z}_{1i}$ will include an intercept term and $\mathbf{z}_{2i}$ will generally be parameterised to include an average treatment effect as well as a treatment contrast, and may contain other explanatory variables and interaction terms. The use of log-links ensures that λ_i and ψ_i are always positive.

It is noted that the inclusion of the treatment effect term ψ_i should avoid the problems encountered by Lindeboom and van den Berg (1994) when using a univariate heterogeneity. The average treatment effect in $\mathbf{z}_{2i}$ should allow for the change in the population over time, while the random effect allows for differences between individuals.

An alternative model could use a correlated bivariate random effect, although the average treatment effect would no longer be identifiable. This alternative model will be discussed further in section 9.5 on page 168.

A major difference between survival models, and the point process model described above, is that a typical survival model would treat the pre-randomisation counts $\mathbf{X}$ as a covariate, rather than an outcome. Figure 4.3 shows a graphical representation of a typical survival model applied to this type of data.

The typical survival model in figure 4.3 may be represented by the equations:

$$Y_i \,|\, \mu_i, \boldsymbol{\gamma} \sim f(\mu_i, \boldsymbol{\gamma})$$

where

$$\mu_i = g(\boldsymbol{\theta}, \mathbf{w}_i, x_i)$$

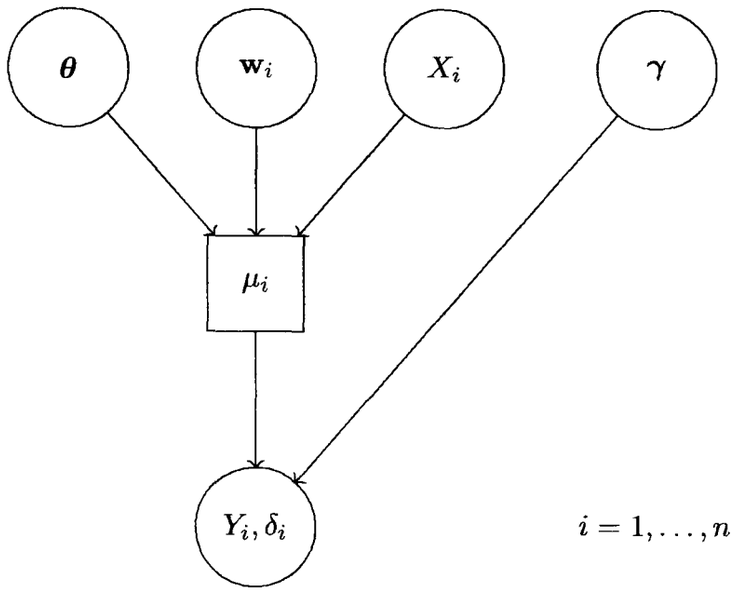

Figure 4.3: Graphical Depiction of a Typical Survival Model

for some survival distribution $f(\cdot)$ such as exponential, gamma, or Pareto, where $g(\cdot)$ is some link function, μ_i represents the covariate effects, and γ represents the scale or shape parameters of the distribution $f(\cdot)$. The vector $\boldsymbol{\theta}$ represents regression coefficients, and $\mathbf{w}_i$ is a vector of covariates (including the treatment covariate z_i). The Pareto survival model is discussed in appendix A on page 172, and applied to the epilepsy data in section 3.5 on page 26.

4.1 Derivation of the Joint Distribution

In the following sections, the derivation of the log-likelihood for this joint model is presented. The log-likelihood is derived in two stages, first considering the contribution of an individual with an observed survival time, and then considering the contribution of an individual with a censored survival time. Finally, the full log-likelihood is presented.

First, note that the complete gamma function $\Gamma(r)$ is defined as

$$\Gamma(r) = \int_0^\infty \theta^{r-1} \exp(-\theta)\, \mathrm{d}\theta \qquad r > 0.$$

In the following work, The observed data is written as $\mathcal{D} = (\mathbf{x},\ \mathbf{y},\ \boldsymbol{\delta},\ \mathbf{u},\ \mathbf{Z}_1,\ \mathbf{Z}_2)$, and the data for individual i as $\mathbf{d}_i = (x_i,\ y_i,\ \delta_i,\ u_i,\ \mathbf{z}_{1i},\ \mathbf{z}_{2i})$.

4.2 Joint Distribution with Y_i Observed

If the survival time for individual i, y_i, is observed (i.e. $\delta_i = 1$), then

$$\begin{aligned} f(x_i, y_i | u_i, \lambda_i, \psi_i, \alpha) &= \int_0^\infty f_X(x_i \mid u_i,\ \lambda_i,\ \nu_i) f_Y(y_i \mid \lambda_i,\ \psi_i,\ \nu_i) g_\nu(\nu_i \mid \alpha)\, \mathrm{d}\nu_i \\ &= \frac{\alpha^\alpha \lambda_i^{x_i+1} u_i^{x_i} \psi_i}{x_i!\, \Gamma(\alpha)} \times \\ &\qquad \int_0^\infty \nu_i^{x_i+\alpha} \exp(-\nu_i(\lambda_i u_i + \lambda_i \psi_i y_i + \alpha))\, \mathrm{d}\nu_i \\ &= \frac{\Gamma(x_i + \alpha + 1)\alpha^\alpha (\lambda_i u_i)^{x_i} \lambda_i \psi_i}{x_i!\, \Gamma(\alpha)(\lambda_i u_i + \lambda_i \psi_i y_i + \alpha)^{x_i+\alpha+1}}. \end{aligned} \tag{4.3}$$

Note that if there is no heterogeneity in the sample, by letting $\alpha \to \infty$ in (4.3),

the joint distribution is given by

$$\lim_{\alpha\to\infty} f(x_i,\, y_i \mid u_i,\, \lambda_i,\, \psi_i,\, \alpha)$$

$$= \frac{(\lambda_i u_i)^{x_i}\lambda_i\psi_i}{x_i!} \lim_{\alpha\to\infty} \frac{\Gamma(x_i+\alpha+1)}{\Gamma(\alpha)} \frac{\alpha^\alpha}{(\lambda_i u_i + \lambda_i\psi_i y_i + \alpha)^{x_i+\alpha+1}}$$

$$= \frac{(\lambda_i u_i)^{x_i}\lambda_i\psi_i}{x_i!} \lim_{\alpha\to\infty} \frac{\prod_{j=0}^{x_i}(\alpha+j)}{(\lambda_i u_i + \lambda_i\psi_i y_i + \alpha)^{x_i+1}} \left(\frac{1}{\frac{\lambda_i u_i+\lambda_i\psi_i y_i}{\alpha}+1}\right)^\alpha$$

$$= \frac{(\lambda_i u_i)^{x_i}\lambda_i\psi_i}{x_i!} \times 1 \times \exp(-\lambda_i u_i - \lambda_i\psi_i y_i)$$

$$= \frac{(\lambda_i u_i)^{x_i}\exp(-\lambda_i u_i)}{x_i!} \lambda_i\psi_i \exp(-\lambda_i\psi_i y_i),$$

which is the product of the Poisson and exponential densities (4.1) and (4.2).

4.3 Joint Distribution with Y_i Censored

If instead it is only recorded that $Y_i > y_i$, i.e. $\delta_i = 0$, then the survivor function $S_Y(y_i \mid \lambda_i,\, \psi_i,\, \nu_i) = \exp(-\lambda_i\psi_i\nu_i y_i)$ is required.

Therefore, for an individual with a censored time,

$$\begin{aligned}
&\Pr(X_i = x_i,\, Y_i > y_i \mid u_i,\, \lambda_i,\, \psi_i,\, \alpha) \\
&\quad = \int_0^\infty f_X(x_i \mid u_i,\, \lambda_i,\, \nu_i)\, S_Y(y_i \mid \lambda_i,\, \psi_i,\, \nu_i)\, g_\nu(\nu_i \mid \alpha)\, \mathrm{d}\nu_i \\
&\quad = \frac{\alpha^\alpha \lambda_i^{x_i} u_i^{x_i}}{x_i!\,\Gamma(\alpha)} \int_0^\infty \nu_i^{x_i+\alpha-1} \exp(-\nu_i(\lambda_i u_i + \lambda_i \psi_i y_i + \alpha))\, \mathrm{d}\nu_i \\
&\quad = \frac{\Gamma(x_i+\alpha)\alpha^\alpha(\lambda_i u_i)^{x_i}}{x_i!\,\Gamma(\alpha)(\lambda_i u_i + \lambda_i \psi_i y_i + \alpha)^{x_i+\alpha}}.
\end{aligned}$$

4.4 Marginal Distributions

In this model, the marginal distribution of the pre-randomisation seizure counts X_i is the negative binomial distribution with parameters α and $\alpha/(\lambda_i u_i + \alpha)$:

$$f(x_i \mid u_i,\, \lambda_i,\, \alpha) = \frac{\Gamma(x_i+\alpha)}{x!\,\Gamma(\alpha)} \left(\frac{\alpha}{\lambda_i u_i + \alpha}\right)^\alpha \left(\frac{\lambda_i u_i}{\lambda_i u_i + \alpha}\right)^{x_i};$$

note that the parameter ψ_i is not involved.

Straightforward manipulation also shows that the marginal distribution of the survival times Y_i is the Pareto:

$$f(y_i \mid \lambda_i,\, \psi_i,\, \alpha) = \lambda_i \psi_i \left(\frac{\alpha}{\alpha + \lambda_i \psi_i y_i}\right)^{\alpha+1}.$$

4.5 The Full Log-Likelihood and Derivatives

The full log-likelihood ℓ_j for the data $\mathcal{D}$ on all the n individuals is given by

$$\begin{aligned}\ell_j(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \alpha\,|\mathcal{D}) = \sum_{i=1}^{n}\Big\{ \sum_{j=0}^{x_i-1} \ln(\alpha+j) + \delta_i \ln(\alpha+x_i) + x_i \ln(u_i) \\ +\alpha\ln(\alpha) + (x_i+\delta_i)\ln(\lambda_i) + \delta_i \ln(\psi_i) \\ - \ln(x_i!) - (x_i+\alpha+\delta_i)\ln(\lambda_i u_i + \lambda_i\psi_i y_i + \alpha)\Big\}.\end{aligned}$$

The saturated log-likelihood for particular data will be given by ℓ_{js}, where

$$\ell_{js}(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{n}\Big\{x_i\ln(x_i) - x_i - \ln(x_i!) - \delta_i\ln(y_i) - \delta_i\Big\}.$$

This is because in the saturated model, the parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ will be defined such that $\lambda_i u_i = x_i$ and $\lambda_i\psi_i y_i = \delta_i$, and with no heterogeneity unexplained by the $2n$ parameters in $\mathbf{z}_1$ and $\mathbf{z}_2$, $\alpha \to \infty$.

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

$$\frac{\partial\ell_j}{\partial\boldsymbol{\beta}_1} = \sum_{i=1}^{n}\left(\frac{\alpha(x_i - \lambda_i u_i + \delta_i - \lambda_i\psi_i y_i)}{(\lambda_i u_i + \lambda_i\psi_i y_i + \alpha)}\right)\mathbf{z}_{1i}, \tag{4.4}$$

$$\frac{\partial\ell_j}{\partial\boldsymbol{\beta}_2} = \sum_{i=1}^{n}\left(\frac{\delta_i\lambda_i u_i - x_i\lambda_i\psi_i y_i + \alpha(\delta_i - \lambda_i\psi_i y_i)}{(\lambda_i u_i + \lambda_i\psi_i y_i + \alpha)}\right)\mathbf{z}_{2i}, \tag{4.5}$$

$$\begin{aligned}\frac{\partial\ell_j}{\partial\alpha} = \sum_{i=1}^{n}\Big\{\sum_{j=0}^{x_i-1}\frac{1}{(\alpha+j)} + \frac{\delta_i}{\alpha+x_i} + \ln(\alpha) + 1 \\ - \ln(\lambda_i u_i + \lambda_i\psi_i y_i + \alpha) - \frac{(x_i+\alpha+\delta_i)}{(\lambda_i u_i + \lambda_i\psi_i y_i + \alpha)}\Big\}.\end{aligned}$$

The second-order derivatives are:

$$\frac{\partial^2 \ell_j}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1'} = -\sum_{i=1}^{n} \left(\frac{\alpha(x_i + \alpha + \delta_i)(\lambda_i u_i + \lambda_i \psi_i y_i)}{(\lambda_i u_i + \lambda_i \psi_i y_i + \alpha)^2} \right) \mathbf{z}_{1i} \mathbf{z}_{1i}',$$

$$\frac{\partial^2 \ell_j}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2'} = -\sum_{i=1}^{n} \left(\frac{\alpha(x_i + \alpha + \delta_i)\lambda_i \psi_i y_i}{(\lambda_i u_i + \lambda_i \psi_i y_i + \alpha)^2} \right) \mathbf{z}_{1i} \mathbf{z}_{2i}',$$

$$\frac{\partial^2 \ell_j}{\partial \boldsymbol{\beta}_1 \partial \alpha} = \sum_{i=1}^{n} \left(\frac{(\lambda_i u_i + \lambda_i \psi_i y_i)(x_i - \lambda_i u_i + \delta_i - \lambda_i \psi_i y_i)}{(\lambda_i u_i + \lambda_i \psi_i y_i + \alpha)^2} \right) \mathbf{z}_{1i},$$

$$\frac{\partial^2 \ell_j}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2'} = -\sum_{i=1}^{n} \left(\frac{(x_i + \alpha + \delta_i)(\lambda_i u_i + \alpha)\lambda_i \psi_i y_i}{(\lambda_i u_i + \lambda_i \psi_i y_i + \alpha)^2} \right) \mathbf{z}_{2i} \mathbf{z}_{2i}',$$

$$\frac{\partial^2 \ell_j}{\partial \boldsymbol{\beta}_2 \partial \alpha} = \sum_{i=1}^{n} \left(\frac{(x_i - \lambda_i u_i + \delta_i - \lambda_i \psi_i y_i))\lambda_i \psi_i y_i}{(\lambda_i u_i + \lambda_i \psi_i y_i + \alpha)^2} \right) \mathbf{z}_{2i},$$

$$\frac{\partial^2 \ell_j}{\partial \alpha \partial \alpha} = -\sum_{i=1}^{n} \left\{ \sum_{j=0}^{x_i - 1} \frac{1}{(\alpha + j)^2} + \frac{\delta_i}{(\alpha + x_i)^2} - \frac{1}{\alpha} \right.$$

$$\left. - \frac{(x_i + \delta_i - \alpha - 2(\lambda_i u_i + \lambda_i \psi_i y_i))}{(\lambda_i u_i + \lambda_i \psi_i y_i + \alpha)^2} \right\}.$$

It is clear from the first-order partial derivatives (4.4) and (4.5) that the observed pre-randomisation counts x_i and post-randomisation event times y_i both contribute to the estimation of the parameters in $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$. Thus, pre-randomisation information about the event rate is contributing to the estimation of the treatment effect.

4.6 Maximum Likelihood Estimation

For a given set of data, it is straightforward to perform maximum likelihood estimation of the parameters α, β_1 and β_2, using a numerical method such as the Newton-Raphson algorithm. With this method, given suitable starting values for the parameters, the first and second derivatives of the log-likelihood may be used in an iterative scheme, to find the maximum of the log-likelihood function.

Starting values for β_1 may be chosen by applying a Poisson GLM (McCullagh & Nelder, 1989) to the count data alone. The regression coefficient estimates under the Poisson GLM may then be used as initial estimates of $\widehat{\lambda}_i$, and also utilised to find an initial estimate of α as

$$\widehat{\alpha} = \frac{1}{n-k} \sum_{i=1}^{n} \frac{[(x_i - \widehat{\lambda}_i u_i)^2 - \widehat{\lambda}_i u_i]}{(\widehat{\lambda}_i u_i)^2}.$$

This estimator was proposed by Gourieroux, Monfort and Trognon (1984), note that division by $(n-k)$ is a degrees-of-freedom correction.

Choosing the starting value of β_2 is more difficult, and the choice of 0 will not always be suitable (particularly if the treatments are very effective, since at least one parameter will then be quite a long way from 0). The Newton-Raphson algorithm is quite sensitive to the choice of starting values, but converges quickly when the initial values are close to the maximum likelihood solution.

In practice, it has been quicker to use the bounded minimising function `nlminb` in `s-plus`, applied to the negative of the log-likelihood function, to find the maximum likelihood solution. This solution can then be used with the second

derivatives of the log-likelihood function to give the observed information matrix, from which the variance-covariance matrix may be derived.

The `s-plus` functions to fit the joint model are given in appendix C on page 198.

4.6.1 Model Selection

The nature of the joint model, with covariates included in λ_i and in ψ_i, means that some thought must be given to the method of selection of which covariates to include in a final model, from the explanatory variables in the data. Let the complete set of general explanatory variables be denoted $\mathbf{Z}_g$, and the complete set of treatment-specific variables be denoted $\mathbf{Z}_t$.

A procedure along the same lines as the one described in section 2.2.1 is suggested. It is proposed that any variable being included in ψ_i, other than treatment-related variables, should also be included in λ_i. This is similar to the idea that when interaction terms are included in a regression model, all the associated lower-order terms should also be included. The suggested procedure has five steps:

- Fit the model just including in λ_i one variable from $\mathbf{Z}_g$ at a time. Record which variables significantly decrease the deviance.[1] Call the recorded set of variables P_λ.
- Starting with the variables P_λ included in λ_i, include in ψ_i just one variable at a time from P_λ and $\mathbf{Z}_t$, to see if any give a significant decrease in the

[1] A typical measure of a significant decrease (at the 95% level) is a decrease of more than 2 or 3 times the difference in degrees of freedom between the two models; some flexibility in the selection rule should be allowed.

deviance. Also see if there are any variables in $\mathbf{Z}_t$ but not in P_λ which, when included in λ_i *and* ψ_i, give a significant reduction in the deviance. Call the variables selected for inclusion in λ_i and ψ_i P_λ^* and P_ψ^* respectively.

- Fit the model including all the variables P_λ^* and P_ψ^*, and then exclude one variable at a time from ψ_i and λ_i, remembering the restriction that terms from $\mathbf{Z}_g$ in ψ_i should also appear in λ_i. If more than one variable is non-significant, the variable giving the least raise in deviance when excluded should be omitted first, and the whole step repeated, until sets Q_λ and Q_ψ are obtained, where leaving out any of the variables in these sets will give a significant increase in the deviance.

- Starting with the variables Q_λ and Q_ψ, add all other variables one at a time, to see if any now give a significant reduction in the deviance. Interaction terms may also be included at this stage, making sure that all necessary lower-order terms are also included in the model. Denote the sets produced at this step R_λ and R_ψ.

- Finally test all the variables in R_λ and R_ψ to see if the omission of any will lead to a significant increase in the deviance. Repeat this step if any variables are selected for exclusion. The resulting sets S_λ and S_ψ of variables are the final selection of this procedure.

4.6.2 Model Checking

The three main areas of model checking can be thought of as the following questions:

- Does the model fit the data adequately?
- Do the assumptions of the model seem sensible?
- Does the model make clinical sense?

To answer the first two questions, some thought needs to be given to diagnostic plots and tests. Some suggestions for model checking are given in section 5.3 of the following chapter, where the joint model is applied to the epilepsy data. Consideration can also be given to generalisations of the joint model. One possible generalisation is explored in chapter 7 of this thesis, and another is considered in chapter 8.

The third of these questions concerns the suitability of the model in modelling particular data. For the epilepsy data, a Poisson process with individual frailty is a natural choice, and it certainly seems preferable to use the pre-randomisation seizure counts as an outcome rather than an explanatory variable, as in standard survival analyses. Some discussion is given in section 5.7 of the following chapter to the suitability of the model for the epilepsy data. Chapter 9 includes a lengthy discussion about some of the assumptions of the joint model, and problems with the epilepsy data.

4.7 Bayesian Estimation

The graphical model in figure 4.2 may be thought of as specifying a Bayesian model, and in this case Markov Chain Monte Carlo simulation may be used to make inference on the parameters. Indeed, using the computer package WinBUGS

(Spiegelhalter *et al.*, 2000), it is easy to specify this model, and fit it to a set of data. The code for this is given in appendix D on page 209.

One problem may be the choice of prior. First consider the negative binomial model specified by

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

where

$$\lambda_i = \exp(\boldsymbol{\beta}_1' \mathbf{z}_{1i}).$$

In this model, a standard choice of priors would be a gamma prior for α, and a normal prior for $\boldsymbol{\beta}_1$ (Congdon, 2001).

For the extension to the joint model, where

$$Y_i \mid \lambda_i, \psi_i, \nu_i \sim \text{Exponential}(\lambda_i \psi_i \nu_i),$$

and

$$\psi_i = \exp(\boldsymbol{\beta}_2' \mathbf{z}_{2i}),$$

it is suggested to use a normal prior for $\boldsymbol{\beta}_2$.

Chapter 5

Application of Joint Model to the Epilepsy Data

An overview of the epilepsy data was given in chapter 3. Included in that chapter were the application of standard count models to the pre-randomisation count data, and standard survival models to the post-randomisation times to first seizure (treating the pre-randomisation counts as a covariate). In this chapter, the joint model derived in chapter 4 is applied to the epilepsy data. The parameter estimates are used to derive estimates of the multiplicative treatment effect $\widehat{\psi}_i$ for various covariate combinations, revealing clinically interesting results.

Some discussion is given to model-checking, and diagnostic plots are presented. In addition, some reanalyses of the data are performed. Finally, conclusions are drawn from the analyses of the epilepsy data.

5.1 Implementing the Joint Model

In this section, the joint model is applied to the epilepsy data. To fit the model, the `s-plus` bounded minimising function `nlminb` was used to find the maximum likelihood solutions, and the second derivatives of the log-likelihood were used to estimate the variance-covariance matrix. A function to run a full Newton-Raphson procedure was also written, but found to be computationally slower.

The maximum likelihood estimates for two models are presented in table 5.1. 'Model 1' includes only a treatment effect, and 'Model 2' includes interactions between *treatment* and *epilepsy type*, and between *treatment* and *age at randomisation*.

The regression coefficient β_{t0} measures the average treatment effect over all individuals, and the coefficient β_{trt} measures the contrast between treatments, in reducing the individual event rate. In Model 2, the coefficients β_{type2} and β_{age2} measure the post-randomisation effect of the respective covariates. The improvement in log-likelihood of the second model is large enough to prefer 'Model 2' to 'Model 1'. The saturated log-likelihood is -5707.4. The correlations of the regression coefficients in 'Model 1' are given in table 5.2. Some of the correlations are quite large, particularly between the regression coefficients of the *trial* indicators, and other terms in $\widehat{\boldsymbol{\beta}}_1$.

Table 5.1: Maximum likelihood parameter estimates for full joint models

Term	Regression Coefficient	Model 1 Estimate (s.e.)	Model 2 Estimate (s.e.)
	α	1.279 (0.057)	1.277 (0.056)
	β_0	−3.081 (0.090)	−3.077 (0.090)
	β_{type}	0.551 (0.036)	0.549 (0.036)
	β_{age}	0.010 (0.022)	0.009 (0.022)
λ_i	β_{trial2}	0.388 (0.143)	0.381 (0.143)
	β_{trial3}	−0.138 (0.107)	−0.136 (0.107)
	β_{trial4}	0.175 (0.118)	0.167 (0.119)
	β_{trial5}	−1.355 (0.115)	−1.360 (0.115)
	β_{t0}	−2.492 (0.042)	−2.496 (0.045)
	β_{trt}	0.050 (0.041)	−0.023 (0.044)
ψ_i	β_{type2}	–	0.026 (0.046)
	$\beta_{trt\times type}$	–	0.229 (0.045)
	β_{age2}	–	0.011 (0.021)
	$\beta_{trt\times age}$	–	0.064 (0.021)
−Log-likelihood (df)		9127 (1134)	9104 (1130)

5.2 Interpretation of Results

By comparing table 5.1 with table 3.6 on page 25, it may be seen that the regression coefficients in the full joint model are very similar to those of the negative binomial GLM. It is noted that the standard errors of the regression coefficients in λ_i are slightly smaller in the joint model, because the added information of the post-randomisation survival times helps in the estimation of the factors affecting the underlying seizure rate.

Table 5.2: The correlations of the regression coefficients in Model 1

Coefficient	α	β_0	β_{type}	β_{age}	β_{trial2}	β_{trial3}	β_{trial4}	β_{trial5}	β_{t0}
β_0	-0.005								
β_{type}	-0.001	0.091							
β_{age}	0.003	-0.042	-0.041						
β_{trial2}	-0.003	-0.635	0.009	0.326					
β_{trial3}	0.007	-0.840	-0.128	-0.058	0.502				
β_{trial4}	0.006	-0.774	-0.088	0.392	0.600	0.618			
β_{trial5}	0.004	-0.798	-0.373	-0.255	0.396	0.715	0.516		
β_{t0}	-0.108	-0.032	0.029	-0.008	0.007	-0.011	-0.003	-0.007	
β_{trt}	0.010	-0.011	0.011	0.001	0.009	0.009	0.008	0.005	-0.014

In table 5.1, the parameters $\beta_{trt \times type}$ and $\beta_{trt \times age}$ in Model 2, representing interactions between *treatment* and *epilepsy type*, and between *treatment* and *age*, are both highly statistically significant. This is in contrast to the best-fitting standard survival models, where the interactions are non-significant.

One way in which the values of the regression coefficients in the joint model may be interpreted is by considering the predictive properties of the model. In the joint model, the treatment reduces multiplicatively the rate at which each individual's seizures occur. Therefore the estimates of the regression coefficients may be used to calculate estimates of this overall multiplicative treatment effect, $\widehat{\psi}$, for a new patient, given the age and epilepsy type of that patient. The lower the value of $\widehat{\psi}$ is, the better the treatment is expected to be in lowering the event rate, and hence in increasing the expected time to first post-randomisation seizure.

Estimates of $\widehat{\psi}$ are given in table 5.3 for 'typical' patients of various ages and epilepsy types, as suggested by a clinician. By considering the pattern of the numbers in the table, a general idea of the interaction of epilepsy type and age with treatment should become clear. For example, the model predicts that a child of age 15 with generalised-onset epilepsy who is treated with CBZ may expect seizures to occur about 11% as often as before they were treated.

From the top of table 5.3, note that VPS is the better treatment for individuals with generalised-onset epilepsies, while CBZ is the better treatment for individuals with partial-onset epilepsies where onset is above the age of 20. It is also noted that CBZ seems to be more efficacious in older patients, while the reverse is true of VPS.

These results support the clinical belief (Wallace *et al.*, 1997) that VPS is the

Table 5.3: Predicted multiplicative effect of treatments on the underlying seizure rate of 'typical' individuals, depending on epilepsy type and age

Age	Epilepsy type	CBZ $\widehat{\psi}$ (95% C.I.)	VPS $\widehat{\psi}$ (95% C.I.)
5	generalised	0.118 (0.095, 0.147)	0.052 (0.041, 0.065)
15	generalised	0.112 (0.092, 0.137)	0.056 (0.045, 0.069)
25	generalised	0.106 (0.087, 0.130)	0.060 (0.049, 0.074)
5	partial	0.079 (0.063, 0.098)	0.086 (0.070, 0.105)
15	partial	0.075 (0.062, 0.089)	0.093 (0.079, 0.110)
25	partial	0.071 (0.061, 0.082)	0.100 (0.087, 0.115)
50	partial	0.062 (0.052, 0.074)	0.121 (0.102, 0.143)

better treatment for generalised epilepsies, and that CBZ is the better treatment for partial epilepsies.

5.3 Diagnostic Plots

Many different diagnostic plots can be used to assess the goodness-of-fit of the joint model to the epilepsy data. Diagnostic plots can also be used to look for outlying points, or clusters of points, which unduly affect the fit of the model. Other uses for diagnostic plots are to investigate the inclusion or exclusion of particular covariates, or the way in which the covariates are included, whether transformed in some way, or using a different link function.

One way to investigate the goodness-of-fit of the model is to consider how well the distribution of the survival times is modelled. The survival times may be rescaled

by using the model estimates of $\widehat{\lambda}_i$ and $\widehat{\psi}_i$, given the covariates for each individual i. For censored times, 1 is added to the rescaled time, so that survival times y_i^* are formed, where $y_i^* = \widehat{\lambda}_i \widehat{\psi}_i y_i + 1 - \delta_i$. These variables y_i^* should theoretically be exponentially distributed with mean 1.

Figure 5.1 shows a plot of the rescaled survival times y_i^* against the quantiles of an exponential distribution with mean 1, stratified by *treatment* and *epilepsy type*. The rescaled times are constructed using the estimates from Model 2 in table 5.1 on page 49. The lines in all four plots are fairly straight, but the lower two plots give cause for concern. This suggests that there is some scope for improvement of the model, possibly by using a mixture of two distributions for the survival times. In addition, the results are quite sensitive to the inclusion of further outliers. For instance, with the inclusion of only the individuals with counts less than 90, the estimate of the *treatment-age* interaction becomes non-significant.

A contour plot may also be constructed, of the profile log-likelihoods of the joint model, for various bivariate combinations of parameters. An example is shown in figure 5.2, which is an illustration using Model 2 in table 5.1 on page 49. Here, β_{t0} and β_{trt} are fixed at various levels, and in each case the profile maximum likelihood solution is found. In figure 5.2, 11 values of each parameter are taken, so the contour plot is composed using 121 observed profile log-likelihoods. The maximum likelihood solution is at $(-2.496, -0.023)$, with log-likelihood -9104.2. The oval shape of the contour lines in figure 5.2 suggest that the model behaves well. Further work could investigate many more contour plots of the profile likelihoods, for other pairs of parameters.

Another class of graphical diagnostic is plots which show some measure of in-

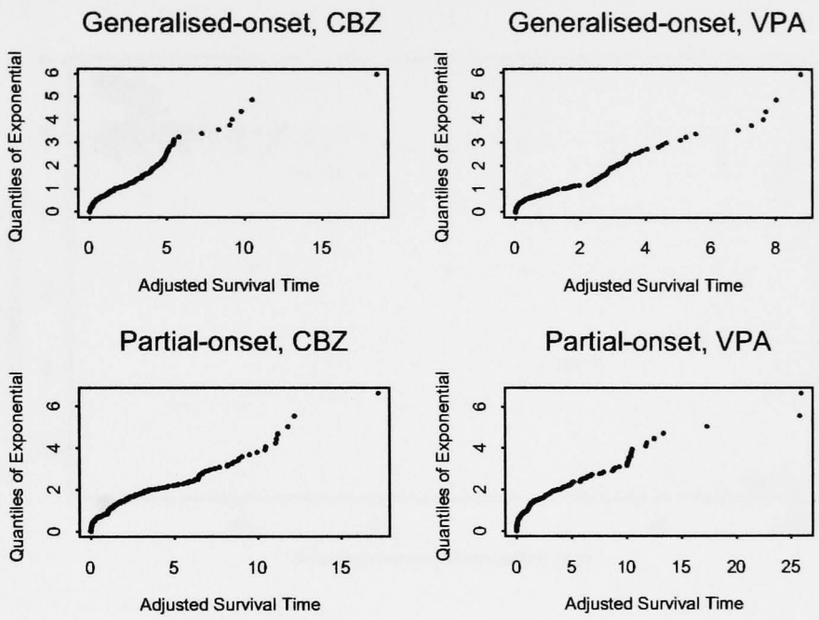

Figure 5.1: Plots of the ordered rescaled survival times against quantiles of a standard exponential.

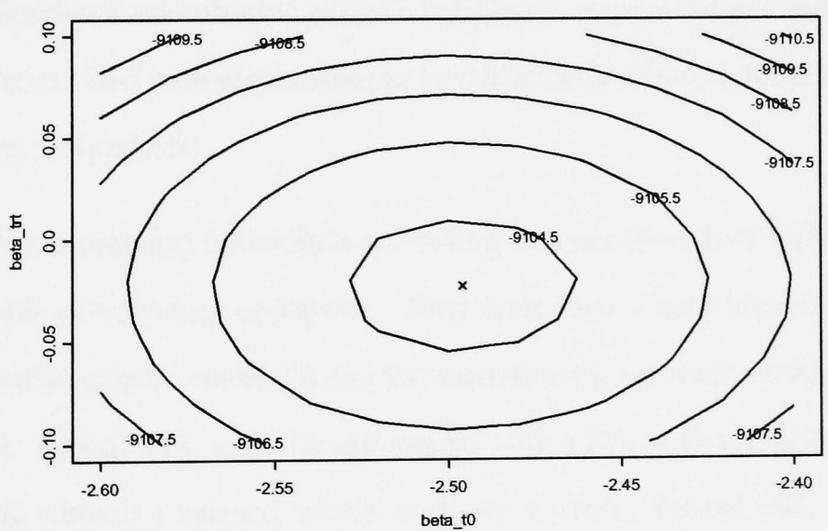

Figure 5.2: Contour plot of the profile log-likelihood, the maximum likelihood solution is marked with an 'x'.

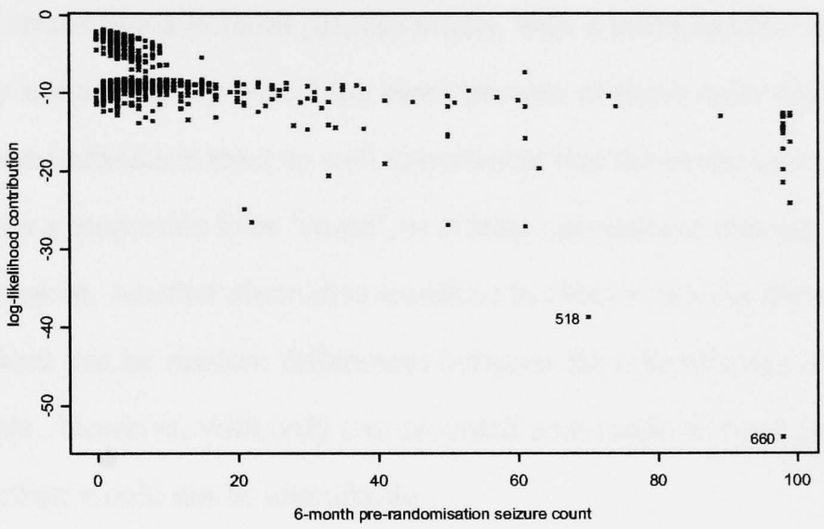

Figure 5.3: Plot of log-likelihood contributions against 6-month pre-randomisation seizure count. The patient ID of two outlying points are labelled.

fluence against index, or covariate information. In figure 5.3, a plot is shown of the log-likelihood contribution of each individual, against the pre-randomisation seizure count. Two individuals seem to be particularly unusual, these are labelled as patients 518 and 660.

Both of these outlying individuals are young females from trial 4 (Verity *et al.*, 1995), with partial-onset epilepsies. They both have a very high 6-month pre-randomisation seizure count (70 and 98 respectively), but react amazingly well to treatment. Patient 518, aged 12 and treated with CBZ, is lost to follow-up after 1431 days without a seizure, which is nearly 4 years. Patient 660, aged 8 and treated with VPS, has her first post-randomisation seizure after 1144 days, which is just over 3 years.

There are other individuals in the data like the two mentioned here, with large

pre-randomisation seizure counts, but very long times to first post-randomisation event. It seems that a Poisson process model, with a multiplicative treatment effect, may not adequately model the event process of these individuals. In some sense, these individuals react so well to treatment that the model could be adapted to allow for a proportion to be 'cured', or at least a proportion who react much better to treatment. Another alternative would be to allow a random treatment effect, so that there can be random differences between the effectiveness of treatments on patients. However, with only one recorded post-randomisation event, such a random effect would not be identifiable.

5.4 Exclusion of Veterans Trial

The previous analysis has included all 5 of the larger trials from the meta-analysis of Marson *et al.* (2002). However, one of these trials, the Veterans' Affairs trial (Mattson *et al.*, 1992), is different to the other four, in many respects. One particularly noticeable difference is that the count data from this trial do not appear to be overdispersed, as can be seen from table 3.3 on page 20.

The results of two joint models fitted to the full data, excluding the Veterans' Affairs trial, are shown in table 5.4. The first model includes only a treatment intercept and contrast, while the second model also includes interactions between the treatments and *epilepsy type*, and between the treatments and *age at randomisation*. Some of the estimates are quite different to those shown in table 5.1 on page 49, although the conclusions remain the same. That is, both age and epilepsy type seem to interact with the treatments, and CBZ is better for partial-onset

epilepsies, while VPS is better for generalised-onset epilepsies. The full predictions based on these data are shown in table 5.5, and may be compared to those in table 5.3 on page 52.

Table 5.4: Maximum likelihood parameter estimates for full joint models, for data excluding 5th trial

Regression Coefficient	Model 1 Estimate (s.e.)	Model 2 Estimate (s.e.)
α	0.973 (0.051)	0.963 (0.050)
β_0	−3.067 (0.101)	−3.070 (0.102)
β_{type}	0.541 (0.041)	0.555 (0.041)
β_{age}	0.064 (0.036)	0.073 (0.036)
β_{trial2}	0.496 (0.170)	0.508 (0.171)
β_{trial3}	−0.151 (0.121)	−0.149 (0.121)
β_{trial4}	0.283 (0.144)	0.286 (0.145)
β_{t0}	−2.735 (0.050)	−2.777 (0.057)
β_{trt}	−0.011 (0.050)	−0.021 (0.056)
β_{type2}	–	−0.187 (0.051)
$\beta_{trt\times type}$	–	0.220 (0.051)
β_{age2}	–	−0.113 (0.032)
$\beta_{trt\times age}$	–	0.097 (0.032)
−Log-likelihood (df)	6196 (710)	6162 (706)

Table 5.5: Predicted multiplicative effect of treatments on the underlying seizure rate of 'typical' individuals, depending on epilepsy type and age, for data excluding 5th trial

Age	Epilepsy type	CBZ $\widehat{\psi}$ (95% C.I.)	VPS $\widehat{\psi}$ (95% C.I.)
5	generalised	0.155 (0.122, 0.196)	0.064 (0.049, 0.083)
15	generalised	0.125 (0.103, 0.153)	0.063 (0.050, 0.078)
25	generalised	0.102 (0.082, 0.125)	0.062 (0.050, 0.076)
5	partial	0.069 (0.053, 0.089)	0.068 (0.054, 0.086)
15	partial	0.056 (0.045, 0.069)	0.067 (0.055, 0.082)
25	partial	0.045 (0.037, 0.055)	0.066 (0.055, 0.080)
50	partial	0.027 (0.020, 0.036)	0.064 (0.047, 0.087)

5.5 Reclassification of Epilepsy Type

The results presented so far in this chapter suggest treatment interactions both with *epilepsy type* and with *age at randomisation*. It is noted that these two explanatory variables are strongly associated. The reliability of classification of *epilepsy type* in these data has been questioned (Williamson *et al.*, 2002). It may be that *age* is acting as a surrogate for the misclassified *epilepsy type*, and this is why the *treatment* by *age* interaction is significant.

Williamson *et al.* (2002) use two alternative reclassification schemes:

1. All individuals with generalised-onset epilepsies and *age at randomisation* greater than 30 are reclassified to have a partial-onset epilepsy. This is because generalised-onset epilepsies are thought to arise very rarely beyond

the age of 30.

2. All individuals are reclassified by an expert into 7 categories of *epilepsy type*.

This section presents the analysis of the data with the first reclassification scheme. The exact definitions of the 7 expert reclassifications are currently unknown, so

Table 5.6: Maximum likelihood parameter estimates for full joint models, for data excluding 5th trial, with reclassification scheme 1

Regression Coefficient	Model 1 Estimate (s.e.)	Model 2 Estimate (s.e.)
α	0.916 (0.047)	0.910 (0.047)
β_0	−3.304 (0.104)	−3.315 (0.105)
β_{type}	0.510 (0.047)	0.529 (0.041)
β_{age}	−0.036 (0.039)	−0.029 (0.039)
β_{trial2}	0.524 (0.172)	0.539 (0.172)
β_{trial3}	0.006 (0.124)	0.010 (0.124)
β_{trial4}	0.324 (0.144)	0.330 (0.145)
β_{t0}	−2.735 (0.050)	−2.725 (0.061)
β_{trt}	−0.015 (0.050)	0.011 (0.061)
β_{type2}	–	−0.222 (0.056)
$\beta_{trt\times type}$	–	0.105 (0.055)
β_{age2}	–	−0.093 (0.035)
$\beta_{trt\times age}$	–	0.089 (0.034)
−Log-likelihood (df)	6221 (710)	6193 (706)

Type: $-1/+1$ for generalised/partial-onset epilepsy

Age: original age $-$ 30, in decades

Table 5.7: Predicted multiplicative effect of treatments on the underlying seizure rate of 'typical' individuals, depending on epilepsy type and age, for data excluding 5th trial, with reclassification scheme 1

Age	Epilepsy type	CBZ $\widehat{\psi}$ (95% C.I.)	VPS $\widehat{\psi}$ (95% C.I.)
5	generalised	0.142 (0.112, 0.179)	0.075 (0.058, 0.097)
15	generalised	0.118 (0.095, 0.147)	0.075 (0.059, 0.095)
25	generalised	0.098 (0.077, 0.126)	0.075 (0.058, 0.095)
5	partial	0.074 (0.056, 0.098)	0.060 (0.046, 0.077)
15	partial	0.062 (0.050, 0.076)	0.059 (0.049, 0.073)
25	partial	0.051 (0.043, 0.061)	0.059 (0.050, 0.070)
50	partial	0.032 (0.024, 0.044)	0.058 (0.044, 0.077)

the analysis of these data with the second reclassification scheme are not presented in this thesis. For the first reclassification scheme, 86 (22%) of the 385 individuals with generalised-onset epilepsies are reclassified to have partial-onset epilepsies. Table 5.6 shows the maximum likelihood estimates for the joint model fitted to the reclassified data. The fifth trial has been excluded from these data, so the results may be compared with those in table 5.4. The new predictions are shown in table 5.7, and may be compared to the predictions in table 5.5.

The conclusions are not changed a great deal by the reclassification of some individuals. CBZ is still the treatment of choice for older patients with partial-onset epilepsies, and VPS is the treatment of choice for younger patients with generalised-onset epilepsies.

5.6 Reanalysis Stratified by Type

The difference in distribution of seizure counts and other variables between the two epilepsy types suggests a stratified analysis. The maximum likelihood parameter estimates of this stratified analysis are shown in table 5.8, and may be compared to the results in table 5.4 on page 57 ('Model 2'). It should be noted that it would be preferable to fit fewer parameters to these data, with only a few hundred individuals in each strata. The larger number of parameters may cause instability in some estimates, but these results are presented anyway because interest lies in comparing them with earlier results.

The differences between the estimates of $\widehat{\alpha}$ across epilepsy types are not too surprising, since this difference in overdispersion was one of the reasons for performing a stratified analysis. Perhaps more interesting is the fluctuation in the estimates $\widehat{\beta}_{trial2}$, $\widehat{\beta}_{trial3}$ and $\widehat{\beta}_{trial4}$. It is also noted that the *treatment-age* interaction seems much stronger for those with partial-onset epilepsies, although this may be connected with the misclassification investigated in section 5.5.

It is reassuring that the scale of the treatment contrast is very similar across the two groups. The original Kaplan-Meier plot in figure 3.3 on page 23 suggested that the treatment contrasts were possibly unbalanced. Figure 3.3 suggested that the degree of improvement of CBZ over VPS for partial-onset epilepsies was far greater than the degree of improvement of VPS over CBZ for generalised-onset epilepsies. However, the estimates of β_{trt} in table 5.8 are very similar in absolute value, suggesting that the degree of improvement of CBZ over VPS for partial-onset epilepsies is very similar to the degree of improvement of VPS over CBZ

Table 5.8: Maximum likelihood parameter estimates for full joint models, for data excluding 5th trial, modelling the two epilepsy types separately.

Regression Coefficient	generalised-onset estimate (s.e.)	partial onset estimate (s.e.)
α	1.174 (0.092)	0.838 (0.060)
β_0	−3.587 (0.120)	−2.569 (0.172)
β_{age}	0.084 (0.046)	0.046 (0.059)
β_{trial2}	0.795 (0.213)	0.067 (0.269)
β_{trial3}	−0.254 (0.149)	−0.026 (0.200)
β_{trial4}	0.183 (0.179)	0.377 (0.231)
β_{t0}	−2.630 (0.089)	−2.937 (0.075)
β_{trt}	−0.354 (0.087)	0.328 (0.075)
β_{age2}	−0.178 (0.052)	−0.090 (0.041)
$\beta_{trt\times age}$	−0.096 (0.052)	0.220 (0.041)
−Log-likelihood (df)	3014 (375)	3124 (324)

Age: original age − 30, in decades

for generalised-onset epilepsies, for individuals of the same age. The difference observed in the Kaplan-Meier plot in figure 3.3 may be explained by the different distribution of ages among the two epilepsy syndromes.

The predictions of the stratified model are shown in table 5.9. In contrast to earlier results, these estimates do not suggest a superior treatment for generalised-onset epilepsies in patients with age of onset lower than 20. The trend is still for VPS to be the preferred treatment for generalised-onset epilepsies, and CBZ the preferred treatment for most partial-onset epilepsies.

Table 5.9: Predicted multiplicative effect of treatments on the underlying seizure rate of 'typical' individuals, depending on epilepsy type and age, for data excluding 5th trial, modelling epilepsy types separately.

Age	Epilepsy type	CBZ $\widehat{\psi}$ (95% C.I.)	VPS $\widehat{\psi}$ (95% C.I.)
5	generalised	0.126 (0.095, 0.166)	0.100 (0.072, 0.139)
15	generalised	0.116 (0.094, 0.143)	0.076 (0.061, 0.096)
25	generalised	0.107 (0.086, 0.132)	0.058 (0.047, 0.072)
5	partial	0.083 (0.062, 0.112)	0.053 (0.041, 0.069)
15	partial	0.061 (0.049, 0.075)	0.061 (0.049, 0.075)
25	partial	0.045 (0.037, 0.054)	0.069 (0.057, 0.083)
50	partial	0.021 (0.014, 0.031)	0.095 (0.069, 0.131)

5.7 Discussion

In this chapter, the joint model was applied to the epilepsy data of Marson *et al.* (2002). The results give strong evidence that VPS is preferable to CBZ, in terms of seizure control, for generalised-onset epilepsies, which agrees with clinical opinion (Wallace *et al.*, 1997). The results also give strong evidence that CBZ is preferable to VPS, in terms of seizure control, for partial-onset epilepsies, where the *age at randomisation* is above 20, which also agrees with clinical opinion (Wallace *et al.*, 1997).

However, there are some doubts about the appropriateness of the joint model for the epilepsy data. There is quite a large difference between the fitted log-likelihood and the saturated log-likelihood (p. 49), which suggests that the joint model may not fit the data particularly well. This view is supported by figure 5.3

on page 55, where many individuals contribute a value lower than -10 to the log-likelihood. Further concerns about the goodness-of-fit of the joint model are raised by the Q-Q plots shown in figure 5.1 on page 54, particularly for partial-onset epilepsies.

On the other hand, the typical survival models for these data are unattractive, because they treat the 6-month pre-randomisation seizure count as an explanatory variable rather than an outcome variable. The attraction of a joint model is that the outcomes are treated as such, and the underlying recurrent event process is modelled in a sensible way.

There are many possible areas to include more flexibility in the joint model. A more general mixture distribution is considered in chapter 7. Allowing for covariates to affect the shape of the mixture distribution is demonstrated in chapter 8. Further extensions, and alternative point process models, are discussed in chapter 9, which also contains a more lengthy criticism of the joint model presented in this chapter.

Chapter 6

Discussion of Relative Efficiency

In this thesis it has previously been suggested that the joint model is more sensible than a typical survival model, for data such as the epilepsy data. In addition to this, it is proposed that the joint model provides more precise estimates of the treatment effect than typical survival models. This chapter investigates the relative efficiency of the joint model derived in chapter 4, compared to other survival models.

Recall that this thesis considers data that are a combination of event counts and survival times. More specifically, there are data such that for each individual i from a population of n individuals, the following are known:

- A pre-randomisation event count $X_i = x_i$, over a period u_i, often taken as the same period u for all individuals.
- A post-randomisation survival time $Y_i = y_i$, which may be censored, with an indicator of censoring δ_i so that $\delta_i = 1$ indicates an observed survival time, while $\delta_i = 0$ indicates a censored survival time.

- An indicator z_i of the treatment given (individuals are randomly assigned to one of two treatments, and $z_i = \pm 1$).

- Some other covariates which may be useful (for example, age and sex). The treatment indicator, and the other covariates, are entered in vectors which here are denoted by by $\mathbf{Z}_{1i}$, $\mathbf{Z}_{2i}$ and $\mathbf{W}_i$, depending on context.

Interest lies in comparing the variance of the maximum likelihood estimator of the treatment effect, under different models. First it is shown that under a simplified joint model (with no allowance for the heterogeneity between individuals), the variance of the maximum likelihood estimator of the treatment effect is the same as under a corresponding exponential survival model.

Then the full joint model (allowing for heterogeneity) is compared with a Pareto survival model, treating the pre-randomisation counts as a covariate. By considering the expected information matrices under each model, and making some additional assumptions about the maximum likelihood estimators, evidence is presented that the variance of the maximum likelihood estimator under the joint model will generally be smaller than under the Pareto model. The results of a simulation study are also presented, to support this hypothesis.

6.1 Joint Model with No Heterogeneity

First consider a simplification of the model given in chapter 4, where there is no extra-Poisson heterogeneity between individuals. This simplified joint model is

specified by the equations

$$\begin{aligned} X_i \,|\, \lambda_i,\, u_i &\sim \text{Poisson}(\lambda_i u_i), \\ Y_i \,|\, \lambda_i,\, \psi_i, &\sim \text{Exponential}(\lambda_i \psi_i), \end{aligned}$$

where

$$\begin{aligned} \lambda_i &= \exp(\boldsymbol{\beta}_1' \mathbf{z}_{1i}) \\ \psi_i &= \exp(\boldsymbol{\beta}_2' \mathbf{z}_{2i}). \end{aligned}$$

The log-likelihood may be derived, for the data $\mathcal{D} = (\mathbf{x},\, \mathbf{y},\, \boldsymbol{\delta},\, \mathbf{u},\, \mathbf{Z}_1,\, \mathbf{Z}_2)$ on all n individuals:

$$\begin{aligned} \ell_1(\boldsymbol{\beta}_1,\, \boldsymbol{\beta}_2,\, \alpha \,|\mathcal{D}) = \sum_{i=1}^{n} \Big\{ & x_i \ln(u_i) + (x_i + \delta_i) \ln(\lambda_i) \\ & + \delta_i \ln(\psi_i) - \ln(x_i!) - \lambda_i u_i - \lambda_i \psi_i y_i) \Big\}. \end{aligned}$$

The first derivatives are given by:

$$\frac{\partial \ell_1}{\partial \boldsymbol{\beta}_1} = \sum_{i=1}^{n} \Big(x_i - \lambda_i u_i + \delta_i - \lambda_i \psi_i y_i \Big) \mathbf{z}_{1i}, \tag{6.1}$$

$$\frac{\partial \ell_1}{\partial \boldsymbol{\beta}_2} = \sum_{i=1}^{n} \Big(\delta_i - \lambda_i \psi_i y_i \Big) \mathbf{z}_{2i}. \tag{6.2}$$

And the second derivatives are given by:

$$\frac{\partial^2 \ell_1}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1'} = -\sum_{i=1}^{n} (\lambda_i u_i + \lambda_i \psi_i y_i) \mathbf{z}_{1i} \mathbf{z}_{1i}', \tag{6.3}$$

$$\frac{\partial^2 \ell_1}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2'} = -\sum_{i=1}^{n} \lambda_i \psi_i y_i \mathbf{z}_{1i} \mathbf{z}_{2i}', \tag{6.4}$$

$$\frac{\partial^2 \ell_1}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2'} = -\sum_{i=1}^{n} \lambda_i \psi_i y_i \mathbf{z}_{2i} \mathbf{z}_{2i}'. \tag{6.5}$$

These derivatives may be used to find maximum likelihood estimates of the parameters contained in the vectors $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2$, and to give the observed information matrix. Particular interest lies in the factors affecting the estimated variance of the parameters contained in $\widehat{\boldsymbol{\beta}}_2$.

6.1.1 Model with Treatment Contrast

Here, only the simplest case is considered, where $\mathbf{z}_{1i}$ is one-dimensional, containing just an intercept term, and $\mathbf{z}_{2i}$ is two-dimensional, containing an overall treatment intercept term and a treatment contrast term ($z_i = \pm 1$). Thus $\lambda_i = \lambda = \exp(\beta_{10})$ for all i, and $\psi_i = \exp(\beta_{20} + \beta_z z_i)$.

First, let

$$
\begin{aligned}
r_0 &= \sum_{i=1}^{n} \delta_i \, \mathrm{I}(z_i = -1) \\
r_1 &= \sum_{i=1}^{n} \delta_i \, \mathrm{I}(z_i = 1) \\
t_0 &= \sum_{i=1}^{n} y_i \, \mathrm{I}(z_i = -1) \\
t_1 &= \sum_{i=1}^{n} y_i \, \mathrm{I}(z_i = 1) \\
k &= \sum_{i=1}^{n} x_i
\end{aligned}
$$

where $\mathrm{I}(A)$ is the indicator function of set A. So r_0 and r_1 are the number of observed survival times in each treatment group, t_0 and t_1 are the totals of all the recorded times (observed or censored) in each treatment group, and k is the total number of recorded pre-randomisation seizure counts.

The maximum likelihood solutions are found by setting the first derivatives (6.1) and (6.2) to zero, giving:

$$\sum_{i=1}^{n} (\widehat{\lambda} u_i + \widehat{\lambda}\widehat{\psi}_i y_i) = k + r_0 + r_1 \tag{6.6}$$

$$\sum_{i=1}^{n} \widehat{\lambda}\widehat{\psi}_i y_i = r_0 + r_1 \tag{6.7}$$

$$\sum_{i=1}^{n} \widehat{\lambda}\widehat{\psi}_i y_i z_i = -r_0 + r_1. \tag{6.8}$$

Maximum Likelihood Estimate of the Treatment Effect

Brief consideration is given to the maximum likelihood estimate of the treatment effect. From equations (6.7) and (6.8) it is clear that

$$\sum_{i=1}^{n} \widehat{\lambda}\widehat{\psi}_i y_i \mathrm{I}(z_i = -1) = \widehat{\lambda}\exp(\widehat{\beta}_{20} - \widehat{\beta}_z)t_0 \;\; = \;\; r_0$$

$$\sum_{i=1}^{n} \widehat{\lambda}\widehat{\psi}_i y_i \mathrm{I}(z_i = 1) = \widehat{\lambda}\exp(\widehat{\beta}_{20} + \widehat{\beta}_z)t_1 \;\; = \;\; r_1.$$

It follows that the maximum likelihood estimate of the treatment effect, $\widehat{\beta}_z$, is given by

$$\widehat{\beta}_z = \frac{1}{2}\log\left(\frac{r_1 t_0}{r_0 t_1}\right).$$

Therefore, the pre-randomisation event information has not contributed to the estimate of the treatment effect, in this model. It is noted that Collett (1994, pp. 127-8) arrives at effectively the same answer for the maximum likelihood estimator of a treatment effect in a simple exponential survival model. Collett, however, parameterises the treatment covariate as $0/1$ (rather than ± 1 used here), so his expression is different by a factor of 2.

The Variance of this Maximum Likelihood Estimator

Here, consideration is given to the variance of the maximum likelihood estimate derived above. Interest lies in whether the estimate is affected by the pre-randomisation information.

The negatives of equations (6.3), (6.4) and (6.5), at the maximum likelihood solution, form the 3×3 observed information matrix I_1. However, noting that $z_i^2 = 1$, it is clear that just three different terms make up the elements of I_1. To simplify working, define the variables a_1, b_1 and c_1 where

$$\begin{aligned} a_1 &= \sum_{i=1}^{n} \widehat{\lambda}_i u + \widehat{\lambda}_i \widehat{\psi}_i y_i \\ b_1 &= \sum_{i=1}^{n} \widehat{\lambda}_i \widehat{\psi}_i y_i \\ c_1 &= \sum_{i=1}^{n} \widehat{\lambda}_i \widehat{\psi}_i y_i z_i. \end{aligned}$$

Now the 3×3 information matrix is of the form

$$I_1 = \begin{pmatrix} a_1 & b_1 & c_1 \\ b_1 & b_1 & c_1 \\ c_1 & c_1 & b_1 \end{pmatrix}.$$

The determinant of I_1 is

$$\begin{aligned} \det(I_1) &= a_1 b_1^2 + 2 b_1 c_1^2 - a_1 c_1^2 - b_1^3 - b_1 c_1^2 \\ &= b_1^2(a_1 - b_1) + c_1^2(b_1 - a_1) \\ &= (a_1 - b_1)(b_1^2 - c_1^2). \end{aligned}$$

Interest lies in the lower-right corner of the adjoint matrix, that is $[\mathrm{adj}(I_1)]_{33}$,

where

$$\left[\text{adj}(I_1)\right]_{33} = a_1 b_1 - b_1^2.$$

Under the maximum likelihood solutions (6.6), (6.7) and (6.8), it is known that

$$\begin{aligned} a_1 &= k + r_0 + r_1 \\ b_1 &= r_0 + r_1 \\ c_1 &= -r_0 + r_1. \end{aligned}$$

So

$$\begin{aligned} \left[\text{adj}(I_1)\right]_{33} &= (r_0 + r_1)k \\ \det(I_1) &= (4r_0 r_1)k. \end{aligned}$$

And so, in this model, the variance of the maximum likelihood estimator of the treatment effect $\widehat{\beta}_z$ is given by

$$\text{Var}(\widehat{\beta}_z) = \frac{r_0 + r_1}{4\, r_0\, r_1}.$$

Note that if the treatment groups were balanced, and there was no censoring, the variance of the maximum likelihood estimate of the treatment effect would be given by $\text{Var}(\widehat{\beta}_z) = 1/n$.

Collett (1994, pp. 127-8) arrives at effectively the same answer for the variance

of the maximum likelihood estimator of a treatment effect in a simple exponential survival model. Collett, however, parameterises the treatment covariate as $0/1$, so his expression is different by a factor of 4.

Thus, in this simple joint model with no heterogeneity, the pre-randomisation information does not contribute to the estimation of the treatment effect, nor to the variance of that estimate. This is perhaps not surprising, because in this simple joint model, a homogeneous Poisson process is assumed, and the 'memoryless' property of the exponential distribution applies.

This model has not been investigated further, although including useful covariate information may in general cause this simple joint model to give different estimates to an exponential survival model with those same covariates included. It is known that including covariate information will not necessarily reduce the variance of the treatment effect, and this is recorded by Ford *et al.* (1995). They show that including an informative covariate into an exponential regression model can, in fact, only increase the variance of the estimates of the regression coefficients of the other covariates.

It is clear that including a random term for heterogeneity in the model will introduce more uncertainty, and therefore increase the variance of the estimated treatment effect. And, of course, the estimated treatment effect will have a slightly different meaning, depending on the model used. Thus the full joint model derived in chapter 4 will not be more efficient than a simple exponential survival model, in terms of the variance of the maximum likelihood estimate of the treatment effect. However, allowing for heterogeneity (overdispersion) in the joint model should provide a much better fit to most observed data than a simple exponential survival

model.

6.2 Relative Efficiency of the full Joint Model over a Pareto Survival Model

This section considers the comparison of the full joint model against other survival models, focussing on the relative efficiency of the treatment estimates under the different models. Relative efficiency is defined as the ratio of the precision of the maximum likelihood estimate of the treatment effect under the joint model, to the precision of the maximum likelihood estimate of the treatment effect under an alternative survival model. In terms of variances, this relative efficiency may be expressed as

$$RE = \frac{\mathrm{Var}(\widehat{\theta}_z)}{\mathrm{Var}(\widehat{\beta}_z)},$$

where $\widehat{\beta}_z$ is the maximum likelihood estimate of the treatment effect under the joint model, and $\widehat{\theta}_z$ is the maximum likelihood estimate of the treatment effect under an alternative model. A value of $RE > 1$ would indicate that the joint model is more precise, and thus more efficient, than the alternative model.

It has previously been shown that the joint model will not, in general, provide more efficient estimates of a treatment effect than a simple exponential survival model. However, simulation studies have shown that the joint model is generally more efficient than other typical survival models, such as the log-logistic, Weibull, gamma or Pareto. Here, because of the similarities between the joint model and the Pareto survival model, the Pareto model is chosen for comparison. In this

way, the maximum likelihood estimates of the treatment effect should also be comparable. A summary of the Pareto survival model is presented in appendix A, including the log-likelihood and derivatives.

The following subsections consider the relative efficiency of the joint model over a simple Pareto survival model (treating the pre-randomisation seizure count as a covariate). First, some motivating examples are given, and then the underlying theory is examined by considering the entries in the expected information matrices.

6.2.1 Example – Simulation Studies

Simulation studies have shown that the joint model is much more efficient than the Pareto survival model. The results of two such simulation studies are presented here. The simulation studies are discussed in more detail in appendix B. Each study involves a total of 2400 datasets, each with 200 individuals. Each study was performed within `s-plus` 2000, and took about 120 hours on a Pentium II 233MHz PC running Windows NT4.

Study 1 with True Joint Model

In the first study, data are simulated based on a true underlying joint model, with various parameter combinations. For each combination of parameters, 100 sets of data are simulated, each containing 200 individuals. The joint model of chapter 4, and a Pareto survival model, are fitted to the data, and the results recorded. In this section, primary interest lies in the comparison between the estimates of $\widehat{\beta}_{trt}$ and

$\widehat{\theta}_{trt}$, and particularly the variability of these estimates. Tables 6.1 and 6.2 give the particular parameter combinations, and the relevant estimates. More information on the simulation study, and more detailed results, are presented in appendix B.

The first thing to notice from tables 6.1 and 6.2 are the numbers in the final column. In the studies in table 6.1, which are the studies with a mild treatment effect, the joint model gave a more precise estimate of the treatment effect about 97% of the time. In the studies with a stronger treatment effect, in table 6.2, the joint model gave more precise estimates of the treatment effect 98% of the time.

From table 6.1 and 6.2, the treatment estimates under the joint model, $\widehat{\beta}_{trt}$, seem to be unbiased. However, the treatment estimates under the Pareto model, $\widehat{\theta}_{trt}$, always seem to be a slight underestimate, but only by up to 5% of the 'true' parameter value (either 0.4 or 0.8).

Study 2 with True Pareto Model

In the second study, data are simulated based on a true underlying Pareto model, with various parameter combinations. In addition, the pre-randomisation counts follow a negative binomial distribution with fixed parameters. For each combination of parameters, 100 sets of data are simulated, each containing 200 individuals. The joint model and a Pareto survival model are fitted to the data, and the results recorded. Again, primary interest lies in the comparison between the estimates of β_{trt} and θ_{trt}, and particularly the variability of these estimates. Tables 6.3 and 6.4 give the particular parameter combinations, and the relevant estimates. More information, and further results, may be found in appendix B.

Table 6.1: Results of relative efficiency simulation study 1, with 'true' joint model. Part 1 ($\beta_{trt} = 0.4$)

α	β_{sex}	β_{t0}	Joint Model			Pareto Model			C
			A	s.d.($\widehat{\beta}_{trt}$)	m(e.s.e.($\widehat{\beta}_{trt}$))	B	s.d.($\widehat{\theta}_{trt}$)	m(e.s.e.($\widehat{\theta}_{trt}$))	
0.8	0.0	−1	−0.003	0.091	0.086	−0.014	0.112	0.106	1.00
0.8	0.0	−2	0.008	0.092	0.090	−0.012	0.114	0.104	0.97
0.8	0.4	−1	0.000	0.088	0.086	−0.022	0.110	0.107	1.00
0.8	0.4	−2	−0.003	0.085	0.091	−0.017	0.103	0.107	0.99
0.8	0.8	−1	−0.008	0.094	0.086	−0.016	0.113	0.112	1.00
0.8	0.8	−2	−0.014	0.087	0.090	−0.042	0.103	0.111	0.98
1.2	0.0	−1	−0.011	0.081	0.083	−0.021	0.090	0.094	0.95
1.2	0.0	−2	−0.023	0.087	0.086	−0.029	0.094	0.092	0.82
1.2	0.4	−1	0.000	0.086	0.083	−0.014	0.101	0.095	0.96
1.2	0.4	−2	0.000	0.087	0.087	−0.017	0.094	0.096	0.92
1.2	0.8	−1	−0.009	0.089	0.084	−0.012	0.103	0.100	1.00
1.2	0.8	−2	−0.003	0.086	0.087	−0.014	0.095	0.098	0.97

A = mean($\widehat{\beta}_{trt} - \beta_{trt}$); B = mean($\widehat{\theta}_{trt} - \beta_{trt}$); m(e.s.e(·)) is median estimated standard error

C = the proportion of studies where { e.s.e.($\widehat{\beta}_{trt}$) < e.s.e.($\widehat{\theta}_{trt}$) }

Table 6.2: Results of relative efficiency simulation study 1, with 'true' joint model. Part 2 ($\beta_{trt} = 0.8$)

α	β_{sex}	β_{t0}	Joint Model			Pareto Model			C
			A	s.d.($\widehat{\beta}_{trt}$)	m(e.s.e.($\widehat{\beta}_{trt}$))	B	s.d.($\widehat{\theta}_{trt}$)	m(e.s.e.($\widehat{\theta}_{trt}$))	
0.8	0.0	−1	−0.003	0.099	0.087	−0.032	0.110	0.106	1.00
0.8	0.0	−2	0.008	0.094	0.091	−0.017	0.115	0.109	0.99
0.8	0.4	−1	0.005	0.082	0.086	−0.015	0.098	0.108	1.00
0.8	0.4	−2	0.010	0.087	0.092	−0.023	0.100	0.111	0.99
0.8	0.8	−1	−0.006	0.083	0.088	−0.014	0.114	0.113	1.00
0.8	0.8	−2	−0.001	0.087	0.092	−0.013	0.112	0.114	1.00
1.2	0.0	−1	0.003	0.079	0.083	−0.012	0.089	0.095	0.97
1.2	0.0	−2	−0.004	0.070	0.088	−0.028	0.082	0.097	0.98
1.2	0.4	−1	0.002	0.074	0.084	−0.022	0.096	0.098	0.99
1.2	0.4	−2	−0.005	0.089	0.087	−0.029	0.104	0.100	0.99
1.2	0.8	−1	−0.007	0.084	0.084	−0.020	0.105	0.101	0.99
1.2	0.8	−2	−0.003	0.088	0.088	−0.030	0.093	0.102	1.00

A = mean($\widehat{\beta}_{trt} - \beta_{trt}$); B = mean($\widehat{\theta}_{trt} - \beta_{trt}$); m(e.s.e(·)) is median estimated standard error

C = the proportion of studies where { e.s.e.($\widehat{\beta}_{trt}$) < e.s.e.($\widehat{\theta}_{trt}$) }

The first thing to notice from tables 6.3 and 6.4 are the numbers in the final column. In both tables, representing both mild and strong treatment effects, the joint model gave more precise estimates of the treatment effect around 98% of the time.

From table 6.3, for combinations including a mild treatment effect, the treatment estimates under the joint model, $\widehat{\beta}_{trt}$, and under the Pareto model, $\widehat{\theta}_{trt}$, seem to be unbiased. However in table 6.4, for studies including a stronger treatment effect, it seems that $\widehat{\beta}_{trt}$ slightly underestimates the treatment effect θ_{trt}.

6.2.2 Example – Epilepsy Data

The joint model and Pareto survival model have also been applied to the epilepsy data, and these results are presented in tables 6.5 and 6.6 below. It is clear that, whether or not the informative covariate *type* is included, the variance of the maximum likelihood estimate of the treatment effect in the joint model is much lower than in the Pareto model. Similar results may be seen when comparing the joint model to other standard choices of parametric survival models such as the Weibull, gamma, log-logistic models. It is noted that this result does not apply to the exponential survival model, where the variance of the maximum likelihood estimator of the treatment effect is much lower, but the model does not fit at all well.

The relative efficiency of the joint model over the Pareto for the epilepsy data may also be calculated. The relative efficiency is the precision of the maximum likelihood estimate of the treatment effect in the joint model divided by that in the Pareto model. From Tables 6.5 and 6.6, the estimate of the relative efficiency here

Table 6.3: Results of relative efficiency simulation study 2, with 'true' Pareto model. Part 1 ($\beta_{trt} = 0.4$)

α	β_{sex}	β_{t0}	Joint Model			Pareto Model			C
			A	s.d.($\widehat{\beta}_{trt}$)	m(e.s.e.($\widehat{\beta}_{trt}$))	B	s.d.($\widehat{\theta}_{trt}$)	m(e.s.e.($\widehat{\theta}_{trt}$))	
0.8	0.0	−1	−0.003	0.091	0.086	−0.014	0.112	0.106	1.00
0.8	0.0	−2	0.008	0.092	0.090	−0.012	0.114	0.104	0.97
0.8	0.4	−1	0.000	0.088	0.086	−0.022	0.110	0.107	1.00
0.8	0.4	−2	−0.003	0.085	0.091	−0.017	0.103	0.107	0.99
0.8	0.8	−1	−0.008	0.094	0.086	−0.016	0.113	0.112	1.00
0.8	0.8	−2	−0.014	0.087	0.090	−0.042	0.103	0.111	0.98
1.2	0.0	−1	−0.011	0.081	0.083	−0.021	0.090	0.094	0.95
1.2	0.0	−2	−0.023	0.087	0.086	−0.029	0.094	0.092	0.82
1.2	0.4	−1	0.000	0.086	0.083	−0.014	0.101	0.095	0.96
1.2	0.4	−2	0.000	0.087	0.087	−0.017	0.094	0.096	0.92
1.2	0.8	−1	−0.009	0.089	0.084	−0.012	0.103	0.100	1.00
1.2	0.8	−2	−0.003	0.086	0.087	−0.014	0.095	0.098	0.97

A = mean($\widehat{\beta}_{trt} - \beta_{trt}$); B = mean($\widehat{\theta}_{trt} - \beta_{trt}$); m(e.s.e(·)) is median estimated standard error

C = the proportion of studies where { e.s.e.($\widehat{\beta}_{trt}$) < e.s.e.($\widehat{\theta}_{trt}$) }

Table 6.4: Results of relative efficiency simulation study 2, with 'true' Pareto model. Part 2 ($\beta_{trt} = 0.8$)

α	β_{sex}	β_{t0}	Joint Model			Pareto Model			C
			A	s.d.$(\widehat{\beta}_{trt})$	m(e.s.e.$(\widehat{\beta}_{trt})$)	B	s.d.$(\widehat{\theta}_{trt})$	m(e.s.e.$(\widehat{\theta}_{trt})$)	
0.8	0.0	−1	−0.003	0.099	0.087	−0.032	0.110	0.106	1.00
0.8	0.0	−2	0.008	0.094	0.091	−0.017	0.115	0.109	0.99
0.8	0.4	−1	0.005	0.082	0.086	−0.015	0.098	0.108	1.00
0.8	0.4	−2	0.010	0.087	0.092	−0.023	0.100	0.111	0.99
0.8	0.8	−1	−0.006	0.083	0.088	−0.014	0.114	0.113	1.00
0.8	0.8	−2	−0.001	0.087	0.092	−0.013	0.112	0.114	1.00
1.2	0.0	−1	0.003	0.079	0.083	−0.012	0.089	0.095	0.97
1.2	0.0	−2	−0.004	0.070	0.088	−0.028	0.082	0.097	0.98
1.2	0.4	−1	0.002	0.074	0.084	−0.022	0.096	0.098	0.99
1.2	0.4	−2	−0.005	0.089	0.087	−0.029	0.104	0.100	0.99
1.2	0.8	−1	−0.007	0.084	0.084	−0.020	0.105	0.101	0.99
1.2	0.8	−2	−0.003	0.088	0.088	−0.030	0.093	0.102	1.00

A = mean$(\widehat{\beta}_{trt} - \beta_{trt})$; B = mean$(\widehat{\theta}_{trt} - \beta_{trt})$; m(e.s.e$(\cdot)$) is median estimated standard error

C = the proportion of studies where $\{$ e.s.e.$(\widehat{\beta}_{trt}) <$ e.s.e.$(\widehat{\theta}_{trt})\}$

is around 3.2. Thus the joint model gives a much more efficient estimate of the treatment effect than the Pareto model, for the epilepsy data.

Table 6.5: Maximum likelihood parameter estimates for Pareto survival model and joint model on subset of epilepsy data

Regression Coefficient	Pareto Model Estimate (s.e.)	Regression Coefficient	Joint Model Estimate (s.e.)
γ	0.410 (0.029)	α	0.794 (0.043)
θ_0	−5.582 (0.160)	β_{10}	−2.910 (0.043)
θ_x	0.812 (0.074)		
		β_{20}	−2.733 (0.050)
θ_z	0.052 (0.090)	β_{2z}	−0.019 (0.050)
$-\log(L)$ (df)	3552 (716)	$-\log(L)$ (df)	6293 (716)

For the Pareto model: γ is the shape parameter; θ_0 is the intercept;
θ_x is the coefficient for the 6-month pre-randomisation count X_i;
and θ_z measures the treatment contrast.
For the joint model: α is the shape parameter; β_{10} is the overall intercept;
β_{20} is the treatment intercept; and β_{2z} measures the treatment contrast.

Table 6.6: Maximum likelihood parameter estimates for Pareto survival model and joint model on subset of epilepsy data

Regression Coefficient	Pareto Model Estimate (s.e.)	Regression Coefficient	Joint Model Estimate (s.e.)
γ	0.418 (0.030)	α	0.943 (0.049)
θ_0	−5.462 (0.162)	β_{10}	−3.012 (0.041)
θ_x	0.727 (0.078)	β_{1type}	0.540 (0.041)
θ_{type}	0.314 (0.094)		
		β_{20}	−2.698 (0.050)
θ_z	0.046 (0.089)	β_{2z}	−0.052 (0.050)
		β_{2type}	−0.022 (0.050)
$\theta_{z\times type}$	0.123 (0.089)	$\beta_{2z\times type}$	0.262 (0.050)
$-\log(L)$ (df)	3527 (714)	$-\log(L)$ (df)	6193 (713)

For the Pareto model, parameters as before, and additionally:
θ_{type} is the coefficient for the epilepsy type;
and $\theta_{z\times type}$ measures the treatment-type interaction.
For the joint model, parameters as before, and additionally:
β_{1type} is for the overall effect of epilepsy type;
β_{2type} measures the post-randomisation effect of type.
and $\beta_{2z\times type}$ measures the treatment-type interaction.

6.2.3 The Pareto Survival Model

Consider the Pareto survival model

$$\begin{aligned} Y_i \,|\, \mu_i,\, \nu_i &\sim \text{Exponential}(\mu_i \nu_i), \\ \nu_i \,|\, \gamma &\sim \text{Gamma}(\gamma, \gamma) \end{aligned}$$

where

$$\mu_i = \exp(\boldsymbol{\theta}' \mathbf{w}_i).$$

The case where $\mathbf{w}_i = (1, x_i, z_i)'$ is considered here. The vector of regression coefficients $\boldsymbol{\theta} = (\theta_0, \theta_x, \theta_z)'$. The pre-randomisation event count x_i could be centred around its mean, or transformed in some way (a standard choice would be to take the natural logarithm). The treatment covariate $z_i = \pm 1$.

This model is described further in appendix A, where the log-likelihood, and the first and second partial derivatives of the log-likelihood, are presented. Next, the expected information matrix for this Pareto survival model is considered.

The Expected Information Matrix I_p

The maximum likelihood solutions $\widehat{\gamma}$ and $\widehat{\mu}_i$ obey the first-order conditions specified by equations (A.2) and (A.3) on page 173. That is,

$$\sum_{i=1}^{n} \left\{ \frac{\widehat{\gamma}(\delta_i - \widehat{\mu}_i y_i)}{\widehat{\gamma} + \widehat{\mu}_i y_i} \right\} = 0$$

$$\sum_{i=1}^{n} \left\{ \frac{\widehat{\gamma}(\delta_i - \widehat{\mu}_i y_i)}{\widehat{\gamma} + \widehat{\mu}_i y_i} \right\} x_i = 0$$

$$\sum_{i=1}^{n} \left\{ \frac{\widehat{\gamma}(\delta_i - \widehat{\mu}_i y_i)}{\widehat{\gamma} + \widehat{\mu}_i y_i} \right\} z_i = 0$$

$$\sum_{i=1}^{n} \left\{ \ln(\widehat{\gamma}) + 1 + \frac{\delta_i}{\widehat{\gamma}} - \ln(\widehat{\gamma} + \widehat{\mu}_i y_i) - \frac{\widehat{\gamma} + \delta_i}{\widehat{\gamma} + \widehat{\mu}_i y_i} \right\} = 0.$$

Consider the 4×4 expected information matrix I_p formed by the expected values of the negative second derivatives of the log-likelihood, which are given by equations (A.4) to (A.6) on page 174. The matrix is composed in order $(\widehat{\gamma}, \widehat{\theta}_0, \widehat{\theta}_x, \widehat{\theta}_z)$ and is of the form

$$I_p = \begin{pmatrix} \cdot & \cdot & \cdot & a_2 \\ \cdot & \cdot & \cdot & b_2 \\ \cdot & \cdot & \cdot & c_2 \\ a_2 & b_2 & c_2 & d_2 \end{pmatrix}.$$

In this matrix, interest lies in the lower-right element of I_p^{-1}, to give the variance of the estimate of θ_z. The entries in the matrix denoted $(\cdot)$ have been ignored. Since

it is later shown that the matrix is block-diagonal, these entries are not relevant. The values a_2 to d_2 are given by

$$a_2 = \mathbb{E}\left[\sum_{i=1}^{n}\left(\frac{(\widehat{\mu}_i y_i - \delta_i)\widehat{\mu}_i y_i}{(\widehat{\gamma}+\widehat{\mu}_i y_i)^2}\right) z_i\right]$$

$$b_2 = \mathbb{E}\left[\sum_{i=1}^{n}\left(\frac{\widehat{\gamma}(\widehat{\gamma}+\delta_i)\widehat{\mu}_i y_i}{(\widehat{\gamma}+\widehat{\mu}_i y_i)^2}\right) z_i\right]$$

$$c_2 = \mathbb{E}\left[\sum_{i=1}^{n}\left(\frac{\widehat{\gamma}(\widehat{\gamma}+\delta_i)\widehat{\mu}_i y_i}{(\widehat{\gamma}+\widehat{\mu}_i y_i)^2}\right) x_i z_i\right]$$

$$d_2 = \mathbb{E}\left[\sum_{i=1}^{n}\left(\frac{\widehat{\gamma}(\widehat{\gamma}+\delta_i)\widehat{\mu}_i y_i}{(\widehat{\gamma}+\widehat{\mu}_i y_i)^2}\right) z_i^2\right] = \mathbb{E}\left[\sum_{i=1}^{n}\left(\frac{\widehat{\gamma}(\widehat{\gamma}+\delta_i)\widehat{\mu}_i y_i}{(\widehat{\gamma}+\widehat{\mu}_i y_i)^2}\right)\right].$$

It is asserted that, assuming that the individuals have been randomised to either treatment, and the treatment groups are balanced, the approximation $a_2 = b_2 = c_2 = 0$ is valid. This is because in each case the sum may be divided into two halves (by treatment group) and each of these two partial sums will have the same expected value. Thus the two partial sums will cancel each other out, in each case. Therefore the matrix is block-diagonal, with the lower-right element of I_p^{-1} given by $1/d_2$.

The probability density function $f(y_i)$ for the Pareto distribution is given in equation (A.1) in appendix A. Making the further assumption that $\widehat{\mu}_i$ and $\widehat{\gamma}$ are unbiased estimates of μ_i and γ, the expectation over y_i of the term inside the summa-

tion of d_2 may be written as

$$\begin{aligned}
\int_0^\infty \left(\frac{\widehat{\gamma}(\widehat{\gamma}+\delta_i)\widehat{\mu}_i y_i}{(\widehat{\gamma}+\widehat{\mu}_i y_i)^2}\right) f(y_i)\,\mathrm{d}y_i &= \int_0^\infty \frac{\widehat{\gamma}(\widehat{\gamma}+\delta_i)\widehat{\mu}_i y_i}{(\widehat{\gamma}+\widehat{\mu}_i y_i)^2}\,\frac{\widehat{\gamma}^{\widehat{\gamma}+1}\widehat{\mu}_i}{(\widehat{\gamma}+\widehat{\mu}_i y_i)^{\widehat{\gamma}+1}}\,\mathrm{d}y_i \\
&= \int_0^\infty \frac{\widehat{\gamma}^{\widehat{\gamma}+2}(\widehat{\gamma}+\delta_i)\widehat{\mu}_i^2 y_i}{(\widehat{\gamma}+\widehat{\mu}_i y_i)^{\widehat{\gamma}+3}}\,\mathrm{d}y_i \\
&= \left[-\frac{\widehat{\gamma}^{\widehat{\gamma}+2}(\widehat{\gamma}+\delta_i)(\widehat{\gamma}+2\widehat{\mu}_i y_i+\widehat{\gamma}\widehat{\mu}_i y_i)}{(\widehat{\gamma}+1)(\widehat{\gamma}+2)(\widehat{\gamma}+\widehat{\mu}_i y_i)^{\widehat{\gamma}+2}}\right]_0^\infty \\
&= \frac{\widehat{\gamma}(\widehat{\gamma}+\delta_i)}{(\widehat{\gamma}+1)(\widehat{\gamma}+2)}.
\end{aligned}$$

Since the sum is convergent, the expectation and summation may be swapped, and thus

$$\begin{aligned}
d_2 &= \mathbb{E}\left[\sum_{i=1}^n \left(\frac{\widehat{\gamma}(\widehat{\gamma}+\delta_i)\widehat{\mu}_i y_i}{(\widehat{\gamma}+\widehat{\mu}_i y_i)^2}\right)\right] \\
&= \sum_{i=1}^n \mathbb{E}\left[\left(\frac{\widehat{\gamma}(\widehat{\gamma}+\delta_i)\widehat{\mu}_i y_i}{(\widehat{\gamma}+\widehat{\mu}_i y_i)^2}\right)\right] \\
&= \sum_{i=1}^n \frac{\widehat{\gamma}(\widehat{\gamma}+\delta_i)}{(\widehat{\gamma}+1)(\widehat{\gamma}+2)}.
\end{aligned}$$

If there are no censored times (for simplicity),

$$\begin{aligned}
d_2 &= \sum_{i=1}^n \frac{\widehat{\gamma}}{\widehat{\gamma}+2} \\
&= \frac{n\widehat{\gamma}}{\widehat{\gamma}+2}.
\end{aligned}$$

And in this case, the variance of the maximum likelihood estimator of the treatment effect $\widehat{\theta}_z$ is given by

$$\mathrm{Var}(\widehat{\theta}_z) = \frac{\widehat{\gamma}+2}{n\widehat{\gamma}}. \tag{6.9}$$

If censored times are included,

$$\begin{aligned} d_2 &= \sum_{i=1}^{n}\left[\frac{\widehat{\gamma}}{\widehat{\gamma}+2}\mathrm{I}(\delta_i=1)+\frac{\widehat{\gamma}^2}{(\widehat{\gamma}+1)(\widehat{\gamma}+2)}\mathrm{I}(\delta_i=0)\right] \\ &= \frac{r\widehat{\gamma}}{(\widehat{\gamma}+2)}+\frac{(n-r)\widehat{\gamma}^2}{(\widehat{\gamma}+1)(\widehat{\gamma}+2)} \\ &= \frac{r\widehat{\gamma}+n\widehat{\gamma}^2}{(\widehat{\gamma}+1)(\widehat{\gamma}+2)}, \end{aligned}$$

where $\mathrm{I}(A)$ is the indicator function of A, r is the number of individuals with observed survival times, and $n-r$ is the number of individuals with censored survival times.

In this case, the variance of the maximum likelihood estimator of the treatment effect $\widehat{\theta}_z$ is given by

$$\mathrm{Var}(\widehat{\theta}_z) = \frac{(\widehat{\gamma}+1)(\widehat{\gamma}+2)}{r\widehat{\gamma}+n\widehat{\gamma}^2}. \tag{6.10}$$

Note that these expressions are independent of x_i (which enters the model through μ_i). A small numerical example is given in table 6.7 on page 95.

6.2.4 The Joint Model

Now consider the joint model of chapter 4, allowing for heterogeneity in the population. The model is specified by

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

where

$$\begin{aligned} \lambda_i &= \exp(\boldsymbol{\beta}_1' \mathbf{z}_{1i}) \\ \psi_i &= \exp(\boldsymbol{\beta}_2' \mathbf{z}_{2i}). \end{aligned}$$

Let $\mathbf{z}_{1i} = 1$ for all individuals, and $\mathbf{z}_{2i} = (1, z_i)'$ where the treatment covariate $z_i = \pm 1$. Then $\boldsymbol{\beta}_1 = \beta_{10}$ and $\boldsymbol{\beta}_2 = (\beta_{20}, \beta_z)'$.

The log-likelihood and derivatives for this model are given in chapter 4 on pages 40 to 41.

The Expected Information Matrix I_j

The expected information matrix is made up of the expected values of the negative second derivatives of the log-likelihood. The second derivatives are given by equations (4.6) to (4.6) on page 41. The information matrix is composed in order

$(\widehat{\alpha}, \widehat{\beta}_{10}, \widehat{\beta}_{20}, \widehat{\beta}_z)$ and is given by

$$I_j = \begin{pmatrix} \cdot & \cdot & \cdot & a_3 \\ \cdot & \cdot & \cdot & b_3 \\ \cdot & \cdot & \cdot & c_3 \\ a_3 & b_3 & c_3 & d_3 \end{pmatrix}.$$

Interested lies in the lower-right element of I_j^{-1}, to give the variance of the maximum likelihood estimator of β_z. The entries marked $\cdot$ have been ignored because it will be assumed below that the matrix is block-diagonal, and thus the values of the other elements are not relevant. The values a_3 to d_3 are given by

$$\begin{aligned}
a_3 &= \mathbb{E}\left[\sum_{i=1}^{n}\left(\frac{(x_i - \lambda_i u_i + \delta_i - \lambda_i\psi_i y_i))\lambda_i\psi_i y_i}{(\lambda_i u_i + \lambda_i\psi_i y_i + \alpha)^2}\right) z_i\right] \\
b_3 &= \mathbb{E}\left[\sum_{i=1}^{n}\left(\frac{\widehat{\alpha}(x_i + \widehat{\alpha} + \delta_i)\widehat{\lambda}_i\widehat{\psi}_i y_i}{(\widehat{\lambda}_i u_i + \widehat{\lambda}_i\widehat{\psi}_i y_i + \widehat{\alpha})^2}\right) z_i\right] \\
c_3 &= \mathbb{E}\left[\sum_{i=1}^{n}\left(\frac{(x_i + \widehat{\alpha} + \delta_i)(\widehat{\lambda}_i u_i + \widehat{\alpha})\widehat{\lambda}_i\widehat{\psi}_i y_i}{(\widehat{\lambda}_i u_i + \widehat{\lambda}_i\widehat{\psi}_i y_i + \widehat{\alpha})^2}\right) z_i\right] \\
d_3 &= \mathbb{E}\left[\sum_{i=1}^{n}\left(\frac{(x_i + \widehat{\alpha} + \delta_i)(\widehat{\lambda}_i u_i + \widehat{\alpha})\widehat{\lambda}_i\widehat{\psi}_i y_i}{(\widehat{\lambda}_i u_i + \widehat{\lambda}_i\widehat{\psi}_i y_i + \widehat{\alpha})^2}\right) z_i^2\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n}\left(\frac{(x_i + \widehat{\alpha} + \delta_i)(\widehat{\lambda}_i u_i + \widehat{\alpha})\widehat{\lambda}_i\widehat{\psi}_i y_i}{(\widehat{\lambda}_i u_i + \widehat{\lambda}_i\widehat{\psi}_i y_i + \widehat{\alpha})^2}\right)\right],
\end{aligned}$$

since $z_i^2 = 1$.

The same argument as before is used to justify the assumption that the matrix is block diagonal, namely that these sums may be split by treatment, and each partial

sum will have the same expected value, so in each case the total sum should be zero (because $z_i = \pm 1$). Therefore the matrix I_j is block-diagonal, and only d_3 is relevant for the variance of the maximum likelihood estimator of the treatment effect $\widehat{\beta}_z$.

Making the further assumption that $\widehat{\lambda}_i$, $\widehat{\psi}_i$ and $\widehat{\alpha}$ are unbiased estimates of λ_i, ψ_i and α, the expectation over x_i and y_i of the term inside the sum in d_3, denoted by

d_{3i}, may be written as:

$$
\begin{aligned}
d_{3i} &= \sum_{x_i=0}^{\infty} \int_{y_i=0}^{\infty} \frac{(x_i+\widehat{\alpha}+\delta_i)(\widehat{\lambda}_i u_i+\widehat{\alpha})\widehat{\lambda}_i\widehat{\psi}_i y_i}{(\widehat{\lambda}_i u_i+\widehat{\lambda}_i\widehat{\psi}_i y_i+\widehat{\alpha})^2} f(x_i,y_i)\,\mathbf{d}y_i \\
&= \sum_{x_i=0}^{\infty} \int_{y_i=0}^{\infty} \frac{(x_i+\widehat{\alpha}+\delta_i)(\widehat{\lambda}_i u_i+\widehat{\alpha})\widehat{\lambda}_i\widehat{\psi}_i y_i}{(\widehat{\lambda}_i u_i+\widehat{\lambda}_i\widehat{\psi}_i y_i+\widehat{\alpha})^2} \\
&\qquad \times \frac{\Gamma(x_i+\widehat{\alpha}+1)\widehat{\alpha}^{\widehat{\alpha}}(\widehat{\lambda}_i u_i)^{x_i}\widehat{\lambda}_i\widehat{\psi}_i}{x_i!\,\Gamma(\widehat{\alpha})(\widehat{\lambda}_i u_i+\widehat{\lambda}_i\widehat{\psi}_i y_i+\widehat{\alpha})^{x_i+\widehat{\alpha}+1}}\,\mathbf{d}y_i \\
&= \sum_{x_i=0}^{\infty} \frac{\Gamma(x_i+\widehat{\alpha}+1)(x_i+\widehat{\alpha}+\delta_i)\widehat{\alpha}^{\widehat{\alpha}}(\widehat{\lambda}_i u_i)^{x_i}}{x_i!\,\Gamma(\widehat{\alpha})} \\
&\qquad \times \int_{y_i=0}^{\infty} \frac{(\widehat{\lambda}_i u_i+\widehat{\alpha})(\widehat{\lambda}_i\widehat{\psi}_i)^2 y_i}{(\widehat{\lambda}_i u_i+\widehat{\lambda}_i\widehat{\psi}_i y_i+\widehat{\alpha})^{x_i+\widehat{\alpha}+3}}\,\mathbf{d}y_i \\
&= \sum_{x_i=0}^{\infty} \frac{(x_i+\widehat{\alpha}+\delta_i)}{(x_i+\widehat{\alpha}+1)(x_i+\widehat{\alpha}+2)} \\
&\qquad \times \frac{\Gamma(x_i+\widehat{\alpha}+1)}{x_i!\,\Gamma(\widehat{\alpha})}\left(\frac{\widehat{\lambda}_i u_i}{\widehat{\lambda}_i u_i+\widehat{\alpha}}\right)^{x_i}\left(\frac{\widehat{\alpha}}{\widehat{\lambda}_i u_i+\widehat{\alpha}}\right)^{\widehat{\alpha}} \\
&= \sum_{x_i=0}^{\infty} \frac{(x_i+\widehat{\alpha}+\delta_i)(x_i+\widehat{\alpha})}{(x_i+\widehat{\alpha}+1)(x_i+\widehat{\alpha}+2)} \\
&\qquad \times \frac{\Gamma(x_i+\widehat{\alpha})}{x_i!\,\Gamma(\widehat{\alpha})}\left(\frac{\widehat{\lambda}_i u_i}{\widehat{\lambda}_i u_i+\widehat{\alpha}}\right)^{x_i}\left(\frac{\widehat{\alpha}}{\widehat{\lambda}_i u_i+\widehat{\alpha}}\right)^{\widehat{\alpha}}. \qquad (6.11)
\end{aligned}
$$

Note that this final term is simply an expectation over X_i where the distribution of X_i is negative binomial with parameters $\widehat{\alpha}$ and $\widehat{\alpha}/(\widehat{\lambda}_i u_i+\widehat{\alpha})$. To be more precise,

the sum (6.11) can be expressed as

$$\mathbb{E}\left[\frac{(x_i+\widehat{\alpha}+\delta_i)(x_i+\widehat{\alpha})}{(x_i+\widehat{\alpha}+1)(x_i+\widehat{\alpha}+2)}\right]. \tag{6.12}$$

For binary δ_i, this expectation (6.12) will clearly be between 0 and 1. Consider first the case where $\delta_i = 1$, that is, where an individual has an observed survival time, and then the case where $\delta_i = 0$.

Now, for $\delta_i = 1$,

$$d_{3i}^{1} = \widehat{\alpha}\left(\frac{\widehat{\alpha}}{\widehat{\lambda}_i u_i+\widehat{\alpha}}\right)^{\widehat{\alpha}}\Gamma(2+\widehat{\alpha})\,\xi\left(1+\widehat{\alpha},2+\widehat{\alpha},3+\widehat{\alpha},\frac{\widehat{\lambda}_i u_i}{\widehat{\lambda}_i u_i+\widehat{\alpha}}\right),$$

where $\xi(a,b,c,z)$ is a regularised hypergeometric function, and may be expressed in terms of its series expansion,

$$\xi(a,b,c,d) = \sum_{k=0}^{\infty}\frac{(a)_k\,(b)_k\,z^k}{(c)_k\,k!\,\Gamma(c)}.$$

For $\delta_i = 0$,

$$\begin{aligned} d_{3i}^{0} \;=\; & \widehat{\alpha}\left(\frac{\widehat{\alpha}}{\widehat{\lambda}_i u_i+\widehat{\alpha}}\right)^{\widehat{\alpha}}\left\{\widehat{\alpha}^2\Gamma(\widehat{\alpha})\,\xi\left(1+\widehat{\alpha},2+\widehat{\alpha},3+\widehat{\alpha},\frac{\widehat{\lambda}_i u_i}{\widehat{\lambda}_i u_i+\widehat{\alpha}}\right)\right. \\ & \left.+\frac{(1+\widehat{\alpha})\widehat{\lambda}_i u_i\Gamma(2+\widehat{\alpha})}{\widehat{\lambda}_i u_i+\widehat{\alpha}}\,\xi\left(2+\widehat{\alpha},2+\widehat{\alpha},4+\widehat{\alpha},\frac{\widehat{\lambda}_i u_i}{\widehat{\lambda}_i u_i+\widehat{\alpha}}\right)\right\}. \end{aligned}$$

Note that, in the simplest case, $\widehat{\lambda}_i u_i = \widehat{\lambda} u$ for all individuals, and thus, **if there**

are no censored times (for simplicity),

$$\begin{aligned} d_3 &= \sum_{i=1}^{n} d_{3i}^1 \\ &= n\,\widehat{\alpha}\left(\frac{\widehat{\alpha}}{\widehat{\lambda}u+\widehat{\alpha}}\right)^{\widehat{\alpha}} \Gamma(2+\widehat{\alpha})\,\xi\left(1+\widehat{\alpha}, 2+\widehat{\alpha}, 3+\widehat{\alpha}, \frac{\widehat{\lambda}u}{\widehat{\lambda}u+\widehat{\alpha}}\right) \end{aligned} \quad (6.13)$$

In this case, the variance of the maximum likelihood estimator of the treatment effect $\widehat{\beta}_z$ is given by the reciprocal of (6.13).

If censored times are included, the expression for the variance of the treatment effect is rather more complicated, and may be written

$$d_3 = \sum_{i=1}^{n} \left(\delta_i d_{3i}^1 + (1-\delta_i) d_{3i}^0\right), \quad (6.14)$$

where d_{3i}^1 and d_{3i}^0 are as given above.

In this case, the variance of the maximum likelihood estimator of the treatment effect $\widehat{\beta}_z$ is given by the reciprocal of (6.14).

A small numerical example of these relationships is given in table 6.8 on page 97.

6.2.5 Comparing the Variances

The precisions of the estimated treatment effect under the Pareto model and the joint model may be compared. An expression for the relative efficiency of the full joint model over a Pareto survival model, in terms of variance rather than

Table 6.7: Estimated variance of $\widehat{\theta}_z$ for different combinations of $\widehat{\gamma}$, with data that have no censored times, 25% censoring, and 50% censoring.

Parameters	Estimated Variance		
	amount of censoring		
$\widehat{\gamma}$	none	25%	50%
0.50	0.0250	0.0300	0.0375
0.75	0.0183	0.0214	0.0257
1.00	0.0150	0.0171	0.0200
1.25	0.0130	0.0146	0.0167
1.50	0.0117	0.0130	0.0146
1.75	0.0107	0.0118	0.0131
2.00	0.0100	0.0109	0.0120

precision, is given by

$$RE = \frac{\mathrm{Var}(\widehat{\theta}_z)}{\mathrm{Var}(\widehat{\beta}_z)} \tag{6.15}$$

where under the assumption of no censoring, $\mathrm{Var}(\widehat{\theta}_z)$ is given by equation (6.9) and $\mathrm{Var}(\widehat{\beta}_z)$ is given by the reciprocal of equation (6.13). If the data includes censoring, then $\mathrm{Var}(\widehat{\theta}_z)$ is given by equation (6.10) and $\mathrm{Var}(\widehat{\beta}_z)$ is given by the reciprocal of equation (6.14).

The condition $RE > 1$ holds if the variance of the treatment estimate under the full joint model is smaller. To give some impression of the expected variances that these equations specify, numerical evaluations may be made for various combinations of parameters.

First, for the Pareto model, estimates are shown of the variance of the treatment effect, for various choices of $\widehat{\gamma}$, and different proportions of censoring. These

estimates are shown in table 6.7. In table 6.8 the estimates of the variance of the treatment effect under the joint model are presented for various choices of $\widehat{\alpha}$ and $\widehat{\lambda u}$ and different proportions of censoring. In both cases, the examples use a sample size of 200.

Typically, the value of $\widehat{\lambda u}$ would be similar to the mean observed count, so values of 2, 4 and 6 have been chosen here. The parameters $\widehat{\alpha}$ and $\widehat{\gamma}$ both measure, in some sense, the amount of overdispersion in the data, and should be fairly similar to each other. In the results of the epilepsy data, $\widehat{\alpha}$ was larger than $\widehat{\gamma}$. However, in the simulation studies, $\widehat{\alpha}$ was generally smaller than $\widehat{\gamma}$.

From tables 6.7 and 6.8, it may be seen that the estimated variance of $\widehat{\theta}_z$ is almost always higher than the estimated variance of $\widehat{\beta}_z$. In addition, the variance of $\widehat{\beta}_z$ seems to be much less affected by an increase in the proportion of censored times.

6.3 Discussion

This chapter has discussed the relative efficiency of the joint model compared to typical survival models, for data of the form of the epilepsy data described in chapter 3. In particular, the joint model has been compared to a Pareto survival model, treating the pre-randomisation event count as a covariate. Results of two simulation studies have been presented, of which more details are given in appendix B. A theoretical approach has also been described.

The simulation studies have suggested that whether the 'true' underlying model is a joint model or a Pareto model, the joint model will provide more precise esti-

Table 6.8: Estimated variance of $\widehat{\beta}_z$ for different combinations of $\widehat{\alpha}$ and $\widehat{\lambda}u$, with data that have no censored times, 25% censoring, and 50% censoring.

Parameters		Estimated Variance		
		amount of censoring		
$\widehat{\alpha}$	$\widehat{\lambda}u$	none	25%	50%
0.50	2	0.0119	0.0129	0.0142
0.75	2	0.0103	0.0112	0.0122
1.00	2	0.0095	0.0102	0.0110
1.25	2	0.0089	0.0095	0.0102
1.50	2	0.0085	0.0090	0.0096
1.75	2	0.0081	0.0086	0.0092
2.00	2	0.0079	0.0083	0.0088
0.50	4	0.0098	0.0105	0.0112
0.75	4	0.0087	0.0092	0.0098
1.00	4	0.0081	0.0085	0.0090
1.25	4	0.0077	0.0081	0.0085
1.50	4	0.0074	0.0078	0.0082
1.75	4	0.0072	0.0076	0.0079
2.00	4	0.0071	0.0074	0.0077
0.50	6	0.0089	0.0094	0.0099
0.75	6	0.0079	0.0083	0.0087
1.00	6	0.0074	0.0078	0.0081
1.25	6	0.0071	0.0074	0.0077
1.50	6	0.0069	0.0072	0.0075
1.75	6	0.0068	0.0070	0.0073
2.00	6	0.0066	0.0069	0.0071

mates of the treatment effect. Thus using the joint model, where appropriate, will give an increased probability of detecting a significant treatment effect. The theoretical approach, concluded in tables 6.7 and 6.8, supports this argument, although the assumptions made by this approach may be questionable. In particular, the assumption that $\widehat{\gamma}$ and $\widehat{\mu}_i$ are unbiased estimators of γ and μ_i is called into question by the results of the simulation studies in appendix B. The results suggest that when data is generated according to a Pareto model with specified parameters, the Pareto model does not give unbiased estimates of those parameters.

Unfortunately, without the assumption of unbiased estimators, a theoretical approach such as section 6.2 would not seem to be possible. This is an area for future research.

Chapter 7

Poisson Mixture Models and the PVF Family

In chapter 4, a joint model for event counts and event times was derived, and this model was fitted to the epilepsy data in chapter 5. Further investigation suggested that this joint model does not fit the epilepsy data particularly well. There are a few possible explanations, such as:

- The use of a gamma mixing distribution does not give enough flexibility to model the structure of the underlying population heterogeneity.
- The events for a given individual are not independent of each other, perhaps because there is *true contagion* as well as *apparent contagion* (see page 7).
- The data include a 6-month event count for each individual, rather than event counts over shorter sub-periods, and the assumption of a constant individual event rate over this entire period may be unrealistic. However, this

assumption cannot be tested by period count data – exact times would be required.

- The model has assumed a multiplicative and instantaneous treatment effect, but perhaps there is a delayed response to treatment.

The first of these reasons concerns the mixing distribution, while the second and third concern the assumption of an underlying Poisson process with individual frailty. The last concerns the way in which a treatment effect has been included in the model.

This chapter considers alternative mixing distributions, to investigate whether the joint model could be improved in this respect. Rather than working with the full joint model, this chapter considers only the distribution of the pre-randomisation seizure counts X_i. Thus, this chapter considers Poisson mixture models of the form

$$\begin{aligned} X_i \,|\, \lambda_i,\, u_i,\, \nu_i &\sim \text{Poisson}(\lambda_i u_i \nu_i), \\ \nu_i \,|\, \boldsymbol{\vartheta} &\sim g(\nu_i \,|\, \boldsymbol{\vartheta}), \end{aligned}$$

where

$$\lambda_i = \exp(\boldsymbol{\beta}_1' \mathbf{z}_{1i}).$$

The parameter λ_i depends on the covariates, u_i is a constant, ν_i is random and $g(\nu_i|\boldsymbol{\vartheta})$ is some distribution (i.e. gamma, inverse Gaussian, PVF, etc.) depending on the parameters $\boldsymbol{\vartheta}$. Let the data be denoted $\mathcal{D} = (\mathbf{x},\, \mathbf{u},\, \mathbf{Z}_1)$.

In the following sections, the Poisson models using one- and two-parameter mixing distributions are derived. Then the power variance family (Hougaard, 1986a) is explained, and the Poisson-PVF mixture model is derived with covariates. Finally, the models are fitted to the epilepsy data and compared.

7.1 One-Parameter Gamma Mixture Model

Consider the gamma mixing distribution, with parameters (γ, γ), i.e. with mean 1 and variance $1/\gamma$. The model is defined by

$$f(x_i \mid \lambda_i, u_i, \nu_i) = \frac{(\lambda_i u_i \nu_i)^{x_i} \exp(-\lambda_i u_i \nu_i)}{x_i!} \qquad u_i > 0$$

$$g(\nu_i \mid \gamma) = \frac{\gamma^\gamma \nu_i^{\gamma-1} \exp(-\gamma \nu_i)}{\Gamma(\gamma)} \qquad \gamma > 0$$

where

$$\lambda_i = \exp(\boldsymbol{\beta}_1' \mathbf{z}_{1i}),$$

Recall that the complete gamma function $\Gamma(r)$ is defined as

$$\Gamma(r) = \int_0^\infty k^{r-1} \exp(-k)\, \mathrm{d}k \qquad r > 0.$$

A simple transformation of the variable gives

$$\frac{\Gamma(r)}{a^r} = \int_0^\infty k^{r-1} \exp(-ak)\, \mathrm{d}k \qquad r > 0.$$

Integrating over ν_i to find the unconditional distribution of X_i,

$$\begin{aligned} f_X(x_i \mid \lambda_i,\, u_i,\, \gamma) &= \int_0^\infty \frac{(\lambda_i u_i \nu_i)^{x_i} \exp(-\lambda_i u_i \nu_i) \gamma^\gamma \nu_i^{\gamma-1} \exp(-\gamma \nu_i)}{x_i!\, \Gamma(\gamma)} \, \mathrm{d}\nu_i \\ &= \frac{(\lambda_i u_i)^{x_i} \gamma^\gamma}{x_i! \Gamma(\gamma)} \int_0^\infty \nu_i^{x_i+\gamma-1} \exp(-\nu_i(\lambda_i u_i + \gamma)) \, \mathrm{d}\nu_i \\ &= \frac{\Gamma(x_i+\gamma)}{x_i! \Gamma(\gamma)} \left(\frac{\gamma}{\lambda_i u_i + \gamma}\right)^\gamma \left(\frac{\lambda_i u_i}{\lambda_i u_i + \gamma}\right)^{x_i}, \end{aligned}$$

which is the negative binomial with parameters γ and $\gamma/(\lambda_i u_i + \gamma)$.

The log-likelihood is given by

$$\begin{aligned} \ell_2(\boldsymbol{\beta}_1,\, \gamma \,|\mathcal{D}) = \sum_{i=1}^n \Big\{ \sum_{j=0}^{x_i-1} \ln(\gamma + j) + \gamma \ln(\gamma) + x_i \ln(\lambda_i) + x_i \ln(u_i) \\ - \ln(x_i!) - (x_i + \gamma) \ln(\lambda_i u_i + \gamma) \Big\}. \end{aligned}$$

The first-order derivatives of ℓ_2 are:

$$\frac{\partial \ell_3}{\partial \boldsymbol{\beta}_1} = \sum_{i=1}^n \left(\frac{\gamma(x_i - \lambda_i u_i)}{(\lambda_i u_i + \gamma)}\right) \mathbf{z}_{1i}$$

$$\frac{\partial \ell_3}{\partial \gamma} = \sum_{i=1}^n \left\{ \sum_{j=0}^{x_i-1} \frac{1}{(\gamma + j)} + 1 + \ln(\gamma) - \ln(\lambda_i u_i + \gamma) - \frac{(x_i + \gamma)}{(\lambda_i u_i + \gamma)} \right\}.$$

The second-order derivatives of ℓ_2 are:

$$\frac{\partial^2 \ell_3}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1'} = -\sum_{i=1}^{n} \left(\frac{\gamma(x_i + \gamma)\lambda_i u_i}{(\lambda_i u_i + \gamma)^2} \right) \mathbf{z}_{1i} \mathbf{z}_{1i}'$$

$$\frac{\partial^2 \ell_3}{\partial \boldsymbol{\beta}_1 \partial \gamma} = -\sum_{i=1}^{n} \left(\frac{\lambda_i u_i (x_i - \lambda_i u_i)}{(\lambda_i u_i + \gamma)^2} \right) \mathbf{z}_{1i}$$

$$\frac{\partial^2 \ell_3}{\partial \gamma \partial \gamma} = \sum_{i=1}^{n} \left\{ \sum_{j=0}^{x_i - 1} \frac{-1}{(\gamma + j)^2} + \frac{1}{\gamma} - \frac{1}{(\lambda_i u_i + \gamma)} - \frac{(\lambda_i u_i - x_i)}{(\lambda_i u_i + \gamma)^2} \right\}.$$

7.2 Two-Parameter Gamma Mixture Model

Now consider the case where the mean of the gamma distribution is not fixed to 1, but allowed to vary.

$$X_i \mid \lambda_i, u_i, \nu_i \sim \text{Poisson}(\lambda_i u_i \nu_i),$$

$$\nu_i \mid \theta, \gamma \sim \text{Gamma}(\theta, \gamma),$$

where

$$\lambda_i = \exp(\boldsymbol{\beta}_1' \mathbf{z}_{1i}).$$

In this model, λ_i will not include an intercept term, to allow identifiability of the second shape parameter in the gamma mixing distribution.

The densities are specified by

$$f(x_i \mid \lambda_i,\, u_i,\, \nu_i) \;=\; \frac{(\lambda_i u_i \nu_i)^{x_i} \exp(-\lambda_i u_i \nu_i)}{x_i!}$$

$$g(\nu_i \mid \theta,\, \gamma) \;=\; \frac{\theta^{\gamma} \nu_i^{\gamma-1} \exp(-\theta \nu_i)}{\Gamma(\gamma)}$$

The unconditional distribution of X_i is

$$f_X(x_i \mid \lambda_i,\, u_i,\, \theta,\, \gamma) = \frac{\Gamma(x_i+\gamma)}{x_i!\Gamma(\gamma)} \left(\frac{\theta}{\lambda_i u_i + \theta}\right)^{\gamma} \left(\frac{\lambda_i u_i}{\lambda_i u_i + \theta}\right)^{x_i}$$

which is the negative binomial with parameters γ and $\theta/(\lambda_i u_i + \theta)$.

The log-likelihood for this model is given by

$$\ell_3(\boldsymbol{\beta}_1,\, \theta,\, \gamma \,|\mathcal{D}) \;=\; \sum_{i=1}^{n} \Big\{ \sum_{j=0}^{x_i-1} \ln(\gamma+j) + \gamma \ln(\theta) + x_i \ln(\lambda_i) + x_i \ln(u_i) - \ln(x_i!) - (x_i+\gamma)\ln(\lambda_i u_i + \theta) \Big\}.$$

The first-order derivatives of ℓ_3 are:

$$\frac{\partial \ell_3}{\partial \boldsymbol{\beta}_1} \;=\; \sum_{i=1}^{n} \left(\frac{\theta x_i - \gamma \lambda_i u_i}{(\lambda_i u_i + \theta)}\right) \mathbf{z}_{1i}$$

$$\frac{\partial \ell_3}{\partial \theta} \;=\; \sum_{i=1}^{n} \left\{ \frac{\gamma}{\theta} - \frac{(x_i+\gamma)}{(\lambda_i u_i + \theta)} \right\}$$

$$\frac{\partial \ell_3}{\partial \gamma} \;=\; \sum_{i=1}^{n} \left\{ \sum_{j=0}^{x_i-1} \frac{1}{(\gamma+j)} + \ln(\theta) - \ln(\lambda_i u_i + \theta) \right\}.$$

The second-order derivatives of ℓ_3 are:

$$\frac{\partial^2 \ell_3}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1'} = -\sum_{i=1}^{n} \left(\frac{\theta(x_i+\gamma)\lambda_i u_i}{(\lambda_i u_i+\theta)^2} \right) \mathbf{z}_{1i}\mathbf{z}_{1i}'$$

$$\frac{\partial^2 \ell_3}{\partial \boldsymbol{\beta}_1 \partial \theta} = \sum_{i=1}^{n} \left(\frac{(x_i+\gamma)\lambda_i u_i}{(\lambda_i u_i+\theta)^2} \right) \mathbf{z}_{1i}$$

$$\frac{\partial^2 \ell_3}{\partial \boldsymbol{\beta}_1 \partial \gamma} = -\sum_{i=1}^{n} \left(\frac{\lambda_i u_i}{(\lambda_i u_i+\theta)^2} \right) \mathbf{z}_{1i}$$

$$\frac{\partial^2 \ell_3}{\partial \theta \partial \theta} = \sum_{i=1}^{n} \left\{ -\frac{\gamma}{\theta^2} + \frac{(x_i+\gamma)}{(\lambda_i u_i+\theta)^2} \right\}$$

$$\frac{\partial^2 \ell_3}{\partial \theta \partial \gamma} = \sum_{i=1}^{n} \left\{ \frac{1}{\theta} - \frac{1}{(\lambda_i u_i+\theta)} \right\}$$

$$\frac{\partial^2 \ell_3}{\partial \gamma \partial \gamma} = \sum_{i=1}^{n} \left\{ \sum_{j=0}^{x_i-1} \frac{-1}{(\gamma+j)^2} \right\}.$$

7.3 Power Variance Family

This section presents the power variance family (PVF) mixing distribution described by Hougaard (1986a) and Hougaard, Lee and Whitmore (1997). A revised version of their description is presented, and the modifications required to include covariates in the model are also discussed. The inclusion of covariates is not well documented in the literature, and derivatives of the log-likelihood, as will be described in this chapter, have not been found elsewhere.

The power variance family of distributions, denoted $G(\alpha, \gamma, \theta)$ is usually described in terms of its Laplace transform,[1] which for $\alpha \neq 0$ is

$$L(s) = \exp(-\gamma\{(\theta + s)^\alpha - \theta^\alpha\}/\alpha).$$

The limit for $\alpha \to 0$ is the Laplace transform of the gamma, which is $L(s) = (\theta/(\theta + s))^\gamma$.

The parameter space is $\alpha \leq 1$, $\gamma > 0$, with $\theta > 0$ for $\alpha \leq 0$, and $\theta \geq 0$ for $0 < \alpha \leq 1$. Special cases are the positive stable distributions ($\theta = 0$), the gamma distributions ($\alpha = 0$), the inverse Gaussian distributions ($\alpha = 1/2$), and the noncentral gamma distributions of zero shape ($\alpha = -1$). For $\alpha = 1$, a degenerate distribution at γ is obtained independently of θ.

Hougaard *et al.* (1997) define the distribution $P\text{-}G(\alpha, \gamma, \theta)$ as the distribution of X when the conditional distribution of X given ν is Poisson with mean ν, where ν has distribution $G(\alpha, \gamma, \theta)$. The derivation of the probability mass function is

[1] The Laplace transform of a random variable ν is given by $L(s) = \mathbb{E}[\exp(-s\nu)]$, which is similar to the moment generating function of ν, $M(t) = \mathbb{E}[\exp(t\nu)]$.

given in the literature, so is not repeated here. The probability mass function for $\alpha \neq 0$ is

$$\begin{aligned}
\Pr[X = 0] &= p(0) = \exp(-\gamma[(1+\theta)^{\alpha} - \theta^{\alpha}]/\alpha) \\
\Pr[X = x] &= p(0) \left\{ \sum_{j=1}^{x} c_{x,j}(\alpha) \gamma^{j} (1+\theta)^{j\alpha - x} \right\} / x!, \qquad x = 1, 2, \ldots
\end{aligned}$$

where the factors $c_{x,j}(\alpha)$ are polynomials in α, given recursively by

$$\begin{aligned}
c_{x,1}(\alpha) &= \frac{\Gamma(x-\alpha)}{\Gamma(1-\alpha)}, \\
c_{x,j}(\alpha) &= c_{x-1,j-1}(\alpha) + c_{x-1,j}(\alpha)[x - 1 - j\alpha], \\
c_{x,x}(\alpha) &= 1.
\end{aligned}$$

The first few terms are

$$c_{1,1}(\alpha) = 1$$

$$c_{2,1}(\alpha) = 1 - \alpha, \qquad c_{2,2}(\alpha) = 1$$

$$c_{3,1}(\alpha) = (1-\alpha)(2-\alpha), \qquad c_{3,2}(\alpha) = 3(1-\alpha), \qquad c_{3,3}(\alpha) = 1.$$

For $\alpha = 0$ the expression is the negative binomial,

$$\Pr[X = x] = \frac{\Gamma(\gamma + x)}{x!\Gamma(\gamma)} \left(\frac{\theta}{1+\theta} \right)^{\gamma} \left(\frac{1}{1+\theta} \right)^{x}, \qquad x = 0, 1, 2, \ldots.$$

For $\alpha = 1$ it is the Poisson

$$\Pr[X = x] = \frac{\gamma^x \exp(-\gamma)}{x!}, \qquad x = 0, 1, 2, \ldots.$$

For $\theta > 0$ the moment $\mathbb{E}[X^k]$ is finite for all $k \geq 0$, and

$$\begin{aligned} \mathbb{E}[X] &= \gamma\theta^{\alpha-1}, \\ \mathrm{Var}[X] &= \gamma\theta^{\alpha-1} + \gamma(1-\alpha)\theta^{\alpha-2}. \end{aligned}$$

A very useful property is that if ν is distributed $G(\alpha, \gamma, \theta)$, then $\lambda u\nu$ is distributed $G(\alpha, (\lambda u)^\alpha\gamma, \theta/(\lambda u))$. Therefore if the conditional distribution of $X \mid \nu$ is Poisson with mean $\lambda u\nu$, and ν is random with distribution $G(\alpha, \gamma, \theta)$, then the unconditional distribution of X is P-$G(\alpha, (\lambda u)^\alpha\gamma, \theta/(\lambda u))$.

For the gamma special case ($\alpha = 0$), X is distributed P-$G(0, \gamma, \theta/(\lambda u))$, and as expected,

$$\Pr[X = x] = \frac{\Gamma(\gamma + x)}{x!\,\Gamma(\gamma)} \left(\frac{\theta}{\lambda u + \theta}\right)^\gamma \left(\frac{\lambda u}{\lambda u + \theta}\right)^x, \qquad x = 0, 1, 2, \ldots$$

which is the negative binomial as described in section 7.2.

The derivatives of $c_{x,j}(\alpha)$ with respect to α will be required later. Using the

digamma function $\Psi(\cdot)$,

$$\begin{aligned} c'_{x,1}(\alpha) &= c_{x,1}(\alpha)\{\Psi(1-\alpha) - \Psi(x-\alpha)\}, \\ c'_{x,j}(\alpha) &= c'_{x-1,j-1}(\alpha) + c'_{x-1,j}(\alpha)[x-1-j\alpha] - jc_{x-1,j}(\alpha), \\ c'_{x,x}(\alpha) &= 0. \end{aligned}$$

The second derivatives are given by the recurrence relations

$$\begin{aligned} c''_{x,1}(\alpha) &= c_{x,1}(\alpha)\{(\Psi(x-\alpha))^2 - 2\Psi(x-\alpha)\Psi(1-\alpha) + (\Psi(1-\alpha))^2 \\ &\qquad +\Psi'(x-\alpha) - \Psi'(1-\alpha)\}, \\ c''_{x,j}(\alpha) &= c''_{x-1,j-1}(\alpha) + c''_{x-1,j}(\alpha)[x-1-j\alpha] - 2jc'_{x-1,j}(\alpha), \\ c''_{x,x}(\alpha) &= 0. \end{aligned}$$

7.4 The P-G Mixture Model with Covariates

The P-G model with covariates is specified by

$$\begin{aligned} X_i \,|\, \lambda_i,\, u_i,\, \nu_i &\sim \text{Poisson}(\lambda_i u_i \nu_i) \\ \nu_i \,|\, \gamma &\sim G(\alpha,\, (\lambda_i u_i)^{\alpha}\gamma,\, \theta/(\lambda_i u_i)), \end{aligned}$$

where

$$\lambda_i = \exp(\boldsymbol{\beta}_1' \mathbf{z}_{1i}).$$

Note that λ_i will not contain an intercept term, to have identifiability. Here, X_i is unconditionally distributed as $P\text{-}G(\alpha,\, (\lambda_i u_i)^{\alpha}\gamma,\, \theta/(\lambda_i u_i))$. This distribution has mean and variance given by

$$\begin{aligned} \mathbb{E}[X_i] &= (\lambda_i u_i)^{\alpha}\gamma \left(\frac{\theta}{\lambda_i u_i}\right)^{\alpha-1} = (\lambda_i u_i)\gamma\theta^{\alpha-1}, \\ \text{Var}[X_i] &= (\lambda_i u)^{\alpha}\gamma \left(\frac{\theta}{\lambda_i u_i}\right)^{\alpha-1} + (\lambda_i u_i)^{\alpha}\gamma(1-\alpha)\left(\frac{\theta}{\lambda_i u_i}\right)^{\alpha-2} \\ &= \mathbb{E}[X_i]\left(1 + \frac{(1-\alpha)\lambda_i u_i}{\theta}\right). \end{aligned}$$

For $\alpha \neq 0$, the point mass at $x_i = 0$ is

$$\begin{aligned} \Pr[X_i = 0] = p(0_i) &= \exp\left(-\frac{(\lambda_i u_i)^\alpha \gamma[(1+\theta/\lambda_i u_i)^\alpha - (\theta/\lambda_i u_i)^\alpha]}{\alpha}\right) \\ &= \exp\left(-\frac{\gamma[(\lambda_i u_i + \theta)^\alpha - \theta^\alpha]}{\alpha}\right). \end{aligned}$$

And for $x_i > 0$ the mass function is

$$\begin{aligned} \Pr[X_i = x_i] &= \frac{p(0_i)}{x_i!(1+\theta/\lambda_i u_i)^{x_i}} \left\{ \sum_{j=1}^{x_i} c_{x_i,j}(\alpha)(\lambda_i u_i)^{j\alpha}\gamma^j(1+\theta/\lambda_i u_i)^{j\alpha} \right\} \\ &= \frac{p(0_i)(\lambda_i u_i)^{x_i}}{x_i!(\lambda_i u_i + \theta)^{x_i}} \left\{ \sum_{j=1}^{x_i} c_{x_i,j}(\alpha)\gamma^j(\lambda_i u_i + \theta)^{j\alpha} \right\}, \end{aligned}$$

where the factors $c_{x_i,j}(\alpha)$ are polynomials in α as described earlier.

The log-likelihood is given by

$$\ell_4 = \sum_{i=1}^{n} \left\{ \ln p(0_i) + x_i \ln(\lambda_i u_i) - \ln(x_i!) - x_i \ln(\lambda_i u_i + \theta) + \ln Q_i \right\},$$

where

$$Q_i = \sum_{j=1}^{x_i} c_{x_i,j}(\alpha)\gamma^j(\lambda_i u_i + \theta)^{j\alpha}.$$

Note that only the first term of ℓ_4 applies for the special case $x_i = 0$. Let this special case be defined as

$$\ell_4^* = \sum_{i=1}^{n} \ln p(0_i).$$

The partial derivatives of $\ln Q_i$ are shown first, to simplify later working. The

first-order terms are:

$$\frac{\partial \ln Q_i}{\partial \boldsymbol{\beta}_1} = \frac{\sum_{j=1}^{x_i} c_{x_i,j}(\alpha)\gamma^j(\lambda_i u_i + \theta)^{j\alpha-1} j\alpha\lambda_i u_i \mathbf{z}_{1i}}{Q_i}$$

$$\frac{\partial \ln Q_i}{\partial \alpha} = \frac{\sum_{j=1}^{x_i} c'_{x_i,j}(\alpha)\gamma^j(\lambda_i u_i + \theta)^{j\alpha}}{Q_i} + \frac{\sum_{j=1}^{x_i} c_{x_i,j}(\alpha)\gamma^j(\lambda_i u_i + \theta)^{j\alpha} j \ln(\lambda_i u_i + \theta)}{Q_i}$$

$$\frac{\partial \ln Q_i}{\partial \gamma} = \frac{\sum_{j=1}^{x_i} c_{x_i,j}(\alpha) j\gamma^{j-1}(\lambda_i u_i + \theta)^{j\alpha}}{Q_i}$$

$$\frac{\partial \ln Q_i}{\partial \theta} = \frac{\sum_{j=1}^{x_i} c_{x_i,j}(\alpha)\gamma^j j\alpha(\lambda_i u_i + \theta)^{j\alpha-1}}{Q_i}.$$

The second-order terms are:

$$\frac{\partial^2 \ln Q_i}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1} = \frac{\sum_{j=1}^{x_i} c_{x_i,j}(\alpha)\gamma^j(\lambda_i u_i + \theta)^{j\alpha-2} j\alpha(j\alpha-1)(\lambda_i u_i)^2 \mathbf{z}_{1i}\mathbf{z}'_{1i}}{Q_i} + \frac{\sum_{j=1}^{x_i} c_{x_i,j}(\alpha)\gamma^j(\lambda_i u_i + \theta)^{j\alpha-1} j\alpha\lambda_i u_i \mathbf{z}_{1i}\mathbf{z}'_{1i}}{Q_i} - \frac{\partial \ln Q_i}{\partial \boldsymbol{\beta}_1} \cdot \frac{\partial \ln Q_i}{\partial \boldsymbol{\beta}'_1}$$

$$\frac{\partial^2 \ln Q_i}{\partial \boldsymbol{\beta}_1 \partial \alpha} = \frac{\sum_{j=1}^{x_i} c'_{x_i,j}(\alpha)\gamma^j(\lambda_i u_i + \theta)^{j\alpha-1} j\alpha\lambda_i u_i \mathbf{z}_{1i}}{Q_i}$$

$$+ \frac{\sum_{j=1}^{x_i} c_{x_i,j}(\alpha)\gamma^j(\lambda_i u_i + \theta)^{j\alpha-1} j\lambda_i u_i \mathbf{z}_{1i}}{Q_i}$$

$$+ \frac{\sum_{j=1}^{x_i} c_{x_i,j}(\alpha)\gamma^j(\lambda_i u_i + \theta)^{j\alpha-1} j^2\alpha\lambda_i u_i \ln(\lambda_i u_i + \theta)\mathbf{z}_{1i}}{Q_i}$$

$$- \frac{\partial \ln Q_i}{\partial \boldsymbol{\beta}_1} \cdot \frac{\partial \ln Q_i}{\partial \alpha}$$

$$\frac{\partial^2 \ln Q_i}{\partial \boldsymbol{\beta}_1 \partial \gamma} = \frac{\sum_{j=1}^{x_i} c_{x_i,j}(\alpha)\gamma^{j-1}(\lambda_i u_i + \theta)^{j\alpha-1} j^2\alpha\lambda_i u_i \mathbf{z}_{1i}}{Q_i}$$

$$- \frac{\partial \ln Q_i}{\partial \boldsymbol{\beta}_1} \cdot \frac{\partial \ln Q_i}{\partial \gamma}$$

$$\frac{\partial^2 \ln Q_i}{\partial \boldsymbol{\beta}_1 \partial \theta} = \frac{\sum_{j=1}^{x_i} c_{x_i,j}(\alpha)\gamma^j(\lambda_i u_i + \theta)^{j\alpha-2} j\alpha(j\alpha-1)\lambda_i u_i \mathbf{z}_{1i}}{Q_i}$$

$$- \frac{\partial \ln Q_i}{\partial \boldsymbol{\beta}_1} \cdot \frac{\partial \ln Q_i}{\partial \theta}$$

$$\frac{\partial^2 \ln Q_i}{\partial\alpha\partial\alpha} = \frac{\sum_{j=1}^{x_i} c''_{x_i,j}(\alpha)\gamma^j(\lambda_i u_i+\theta)^{j\alpha}}{Q_i} + \frac{2\sum_{j=1}^{x_i} c'_{x_i,j}(\alpha)\gamma^j(\lambda_i u_i+\theta)^{j\alpha} j\ln(\lambda_i u_i+\theta)}{Q_i} + \frac{\sum_{j=1}^{x_i} c_{x_i,j}(\alpha)\gamma^j(\lambda_i u_i+\theta)^{j\alpha}[j\ln(\lambda_i u_i+\theta)]^2}{Q_i} - \frac{\partial \ln Q_i}{\partial\alpha}\cdot\frac{\partial \ln Q_i}{\partial\alpha}$$

$$\frac{\partial^2 \ln Q_i}{\partial\alpha\partial\gamma} = \frac{\sum_{j=1}^{x_i} c'_{x_i,j}(\alpha) j\gamma^{j-1}(\lambda_i u_i+\theta)^{j\alpha}}{Q_i} + \frac{\sum_{j=1}^{x_i} c_{x_i,j}(\alpha)\gamma^{j-1}(\lambda_i u_i+\theta)^{j\alpha} j^2\ln(\lambda_i u_i+\theta)}{Q_i} - \frac{\partial \ln Q_i}{\partial\alpha}\cdot\frac{\partial \ln Q_i}{\partial\gamma}$$

$$\frac{\partial^2 \ln Q_i}{\partial\alpha\partial\theta} = \frac{\sum_{j=1}^{x_i} c'_{x_i,j}(\alpha)\gamma^j j\alpha(\lambda_i u_i+\theta)^{j\alpha-1}}{Q_i} + \frac{\sum_{j=1}^{x_i} c_{x_i,j}(\alpha)\gamma^j j^2\alpha(\lambda_i u_i+\theta)^{j\alpha-1}\ln(\lambda_i u_i+\theta)}{Q_i} + \frac{\sum_{j=1}^{x_i} c_{x_i,j}(\alpha)\gamma^j j(\lambda_i u_i+\theta)^{j\alpha-1}}{Q_i} - \frac{\partial \ln Q_i}{\partial\alpha}\cdot\frac{\partial \ln Q_i}{\partial\theta}$$

$$\frac{\partial^2 \ln Q_i}{\partial\gamma\partial\gamma} = \frac{\sum_{j=1}^{x_i} c_{x_i,j}(\alpha) j(j-1)\gamma^{j-2}(\lambda_i u_i + \theta)^{j\alpha}}{Q_i}$$

$$- \frac{\partial \ln Q_i}{\partial\gamma} \cdot \frac{\partial \ln Q_i}{\partial\gamma}$$

$$\frac{\partial^2 \ln Q_i}{\partial\gamma\partial\theta} = \frac{\sum_{j=1}^{x_i} c_{x_i,j}(\alpha)\gamma^{j-1} j^2 \alpha(\lambda_i u_i + \theta)^{j\alpha-1}}{Q_i}$$

$$- \frac{\partial \ln Q_i}{\partial\gamma} \cdot \frac{\partial \ln Q_i}{\partial\theta}$$

$$\frac{\partial^2 \ln Q_i}{\partial\theta\partial\theta} = \frac{\sum_{j=1}^{x_i} c_{x_i,j}(\alpha)\gamma^{j} j\alpha(j\alpha-1)(\lambda_i u_i + \theta)^{j\alpha-2}}{Q_i}$$

$$- \frac{\partial \ln Q_i}{\partial\theta} \cdot \frac{\partial \ln Q_i}{\partial\theta} \cdot$$

The first-order derivatives of ℓ_4 for the special case $x_i = 0$ are:

$$\frac{\partial \ell_4^*}{\partial \boldsymbol{\beta}_1} = \sum_{i=1}^{n} \left(-\gamma\lambda_i u_i (\lambda_i u_i + \theta)^{\alpha-1}\right) \mathbf{z}_{1i}$$

$$\frac{\partial \ell_4^*}{\partial \alpha} = \sum_{i=1}^{n} \left\{ \frac{\gamma\{(\lambda_i u_i + \theta)^{\alpha}[1 - \alpha \ln(\lambda_i u_i + \theta)] + \theta^{\alpha}[\alpha \ln\theta - 1]\}}{\alpha^2} \right\}$$

$$\frac{\partial \ell_4^*}{\partial \gamma} = \sum_{i=1}^{n} \left\{ \frac{\theta^{\alpha} - (\lambda_i u_i + \theta)^{\alpha}}{\alpha} \right\}$$

$$\frac{\partial \ell_4^*}{\partial \theta} = \sum_{i=1}^{n} \left\{\gamma[\theta^{\alpha-1} - (\lambda_i u_i + \theta)^{\alpha-1}]\right\}.$$

The first derivatives of ℓ_4 for $x_i > 0$ are:

$$\frac{\partial \ell_4}{\partial \boldsymbol{\beta}_1} = \frac{\partial \ell_4^*}{\partial \boldsymbol{\beta}_1} + \sum_{i=1}^{n} \left\{ \left(x_i - \frac{x_i \lambda_i u_i}{\lambda_i u_i + \theta} \right) \mathbf{z}_{1i} + \frac{\partial \ln Q_i}{\partial \boldsymbol{\beta}_1} \right\}$$

$$\frac{\partial \ell_4}{\partial \alpha} = \frac{\partial \ell_4^*}{\partial \alpha} + \sum_{i=1}^{n} \left\{ \frac{\partial \ln Q_i}{\partial \alpha} \right\}$$

$$\frac{\partial \ell_4}{\partial \gamma} = \frac{\partial \ell_4^*}{\partial \gamma} + \sum_{i=1}^{n} \left\{ \frac{\partial \ln Q_i}{\partial \gamma} \right\}$$

$$\frac{\partial \ell_4}{\partial \theta} = \frac{\partial \ell_4^*}{\partial \theta} - \sum_{i=1}^{n} \left\{ \frac{x_i}{\lambda_i u_i + \theta} - \frac{\partial \ln Q_i}{\partial \theta} \right\}.$$

The second derivatives for the special case $x_i = 0$ are given by:

$$\frac{\partial^2 \ell_4^*}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1'} = -\sum_{i=1}^{n} \gamma \lambda_i u_i (\lambda_i u_i + \theta)^{\alpha-2} (\alpha \lambda_i u_i + \theta) \, \mathbf{z}_{1i} \mathbf{z}_{1i}'$$

$$\frac{\partial^2 \ell_4^*}{\partial \boldsymbol{\beta}_1 \partial \alpha} = -\sum_{i=1}^{n} \gamma \lambda_i u_i (\lambda_i u_i + \theta)^{\alpha-1} \ln(\lambda_i u_i + \theta) \, \mathbf{z}_{1i}$$

$$\frac{\partial^2 \ell_4^*}{\partial \boldsymbol{\beta}_1 \partial \gamma} = -\sum_{i=1}^{n} \lambda_i u_i (\lambda_i u_i + \theta)^{\alpha-1} \, \mathbf{z}_{1i}$$

$$\frac{\partial^2 \ell_4^*}{\partial \boldsymbol{\beta}_1 \partial \theta} = -\sum_{i=1}^{n} (\alpha - 1) \gamma \lambda_i u_i (\lambda_i u_i + \theta)^{\alpha-2} \, \mathbf{z}_{1i}$$

$$\frac{\partial^2 \ell_4^*}{\partial\alpha\partial\alpha} = \sum_{i=1}^{n} \left\{ \frac{\gamma}{\alpha^3} \Big[2\theta^\alpha - 2(\lambda_i u_i + \theta)^\alpha - 2\alpha\theta^\alpha \ln\theta + \alpha^2\theta^\alpha (\ln\theta)^2 \right.$$

$$\left. +2\alpha(\lambda_i u_i + \theta)^\alpha \ln(\lambda_i u_i + \theta) - \alpha^2(\lambda_i u_i + \theta)^\alpha (\ln(\lambda_i u_i + \theta))^2 \Big] \right\}$$

$$\frac{\partial^2 \ell_4^*}{\partial\alpha\partial\gamma} = \sum_{i=1}^{n} \left\{ \frac{(\lambda_i u_i + \theta)^\alpha - \theta^\alpha + \alpha\theta^\alpha \ln\theta - \alpha(\lambda_i u_i + \theta)^\alpha \ln(\lambda_i u_i + \theta)}{\alpha^2} \right\}$$

$$\frac{\partial^2 \ell_4^*}{\partial\alpha\partial\theta} = \sum_{i=1}^{n} \left\{ \gamma[\theta^{\alpha-1} \ln\theta - (\lambda_i u_i + \theta)^{\alpha-1} \ln(\lambda_i u_i + \theta)] \right\}$$

$$\frac{\partial^2 \ell_4^*}{\partial\gamma\partial\gamma} = 0$$

$$\frac{\partial^2 \ell_4^*}{\partial\gamma\partial\theta} = \sum_{i=1}^{n} \left\{ \theta^{\alpha-1} - (\lambda_i u_i + \theta)^{\alpha-1} \right\}$$

$$\frac{\partial^2 \ell_4^*}{\partial\theta\partial\theta} = \sum_{i=1}^{n} \left\{ \gamma(\alpha - 1)[\theta^{\alpha-2} - (\lambda_i u_i + \theta)^{\alpha-2}] \right\}.$$

And the full second-order derivatives of ℓ_4 are given by:

$$\frac{\partial^2 \ell_4}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1'} = \frac{\partial^2 \ell_4^*}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1'} + \sum_{i=1}^{n} \left\{ \frac{x_i \lambda_i u_i \theta}{(\lambda_i u_i + \theta)^2} \mathbf{z}_{1i} \mathbf{z}_{1i}' - \frac{\partial^2 \ln Q_i}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1} \right\}$$

$$\frac{\partial^2 \ell_4}{\partial \boldsymbol{\beta}_1 \partial \alpha} = \frac{\partial^2 \ell_4^*}{\partial \boldsymbol{\beta}_1 \partial \alpha} + \sum_{i=1}^{n} \frac{\partial^2 \ln Q_i}{\partial \boldsymbol{\beta}_1 \partial \alpha}$$

$$\frac{\partial^2 \ell_4}{\partial \boldsymbol{\beta}_1 \partial \gamma} = \frac{\partial^2 \ell_4^*}{\partial \boldsymbol{\beta}_1 \partial \gamma} - \sum_{i=1}^{n} \frac{\partial^2 \ln Q_i}{\partial \boldsymbol{\beta}_1 \partial \gamma}$$

$$\frac{\partial^2 \ell_4}{\partial \boldsymbol{\beta}_1 \partial \theta} = \frac{\partial^2 \ell_4^*}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1'} + \sum_{i=1}^{n} \left\{ \frac{x_i \lambda_i u_i}{(\lambda_i u_i + \theta)^2} \mathbf{z}_{1i} + \frac{\partial^2 \ln Q_i}{\partial \boldsymbol{\beta}_1 \partial \theta} \right\}$$

$$\frac{\partial^2 \ell_4}{\partial \alpha \partial \alpha} = \frac{\partial^2 \ell_4^*}{\partial \alpha \partial \alpha} + \sum_{i=1}^{n} \frac{\partial^2 \ln Q_i}{\partial \alpha \partial \alpha}$$

$$\frac{\partial^2 \ell_4}{\partial \alpha \partial \gamma} = \frac{\partial^2 \ell_4^*}{\partial \alpha \partial \gamma} + \sum_{i=1}^{n} \frac{\partial^2 \ln Q_i}{\partial \alpha \partial \gamma}$$

$$\frac{\partial^2 \ell_4}{\partial \alpha \partial \theta} = \frac{\partial^2 \ell_4^*}{\partial \alpha \partial \theta} + \sum_{i=1}^{n} \frac{\partial^2 \ln Q_i}{\partial \alpha \partial \theta}$$

$$\frac{\partial^2 \ell_4}{\partial \gamma \partial \gamma} = \frac{\partial^2 \ell_4^*}{\partial \gamma \partial \gamma} + \sum_{i=1}^{n} \frac{\partial^2 \ln Q_i}{\partial \gamma \partial \gamma}$$

$$\frac{\partial^2 \ell_4}{\partial \gamma \partial \theta} = \frac{\partial^2 \ell_4^*}{\partial \gamma \partial \theta} + \sum_{i=1}^{n} \frac{\partial^2 \ln Q_i}{\partial \gamma \partial \theta}$$

$$\frac{\partial^2 \ell_4}{\partial \theta \partial \theta} = \frac{\partial^2 \ell_4^*}{\partial \theta \partial \theta} + \sum_{i=1}^{n} \left\{ \frac{x_i}{(\lambda_i u_i + \theta)^2} + \frac{\partial^2 \ln Q_i}{\partial \theta \partial \theta} \right\}.$$

7.5 Application of Poisson Mixture Models to the Epilepsy Data

In this section, the Poisson-gamma and Poisson-PVF models are fitted to the pre-randomisation counts in the epilepsy data. The epilepsy data has previously been described and analysed in chapters 3 and 5. The various mixture models are fitted to the counts X_i, allowing for effects of covariates such as *epilepsy type*, *age at randomisation*, and an indicator of which of the five trials the individual took part in.

The results of the simple Poisson GLM, and the negative binomial GLM (one-

Table 7.1: Estimates (standard errors) for Poisson and negative binomial GLM

Regression Coefficient	Poisson GLM estimate (s.e.)	NB GLM estimate (s.e.)
γ	–	1.221 (0.055)
β_0	−3.099 (0.033)	−3.059 (0.092)
β_{type}	0.541 (0.013)	0.557 (0.037)
β_{age}	0.035 (0.009)	0.025 (0.022)
β_{trial2}	0.256 (0.050)	0.385 (0.147)
β_{trial3}	−0.059 (0.038)	−0.131 (0.110)
β_{trial4}	0.296 (0.043)	0.189 (0.122)
β_{trial5}	−1.447 (0.044)	−1.479 (0.119)
−Log-likelihood (df)	7489 (1138)	3311 (1137)

Type: $-1/+1$ for generalised/partial-onset epilepsy

Age: original age $-$ 30, in decades

parameter gamma mixture), are shown in table 7.1. An intercept term β_0 is included in the regression model, and the negative binomial distribution just has one scale parameter γ. Note from table 7.1 that the log-likelihood is vastly increased by the introduction of the scale parameter in the negative binomial model, while the regression coefficients remain very similar.

The Poisson-PVF mixture family are also fitted to the epilepsy data. The parameter estimates for these new models are shown in tables 7.2 and 7.3. In these models, an intercept term is not included in the regression coefficients, to have

Table 7.2: Estimates (standard errors) for P-G models (1). The models include the covariates *epilepsy type*, *age at randomisation*, and an indicator of which trial the individual was part of.

Regression Coefficient	Model 1 estimate (s.e.)	Model 2 estimate (s.e.)	Model 3 estimate (s.e.)
α	0	0.5	0.829 (0.012)
γ	1.221 (0.030)	0.119 (0.010)	0.027 (0.002)
θ	26.006 (2.353)	11.620 (0.833)	0.778 (0.448)
β_{type}	0.557 (0.037)	0.426 (0.047)	0.296 (0.031)
β_{age}	0.025 (0.022)	0.015 (0.030)	0.004 (0.019)
β_{trial2}	0.385 (0.145)	0.400 (0.123)	0.432 (0.139)
β_{trial3}	−0.131 (0.108)	0.251 (0.091)	0.546 (0.104)
β_{trial4}	0.189 (0.120)	0.274 (0.063)	0.351 (0.114)
β_{trial5}	−1.479 (0.117)	−0.807 (0.056)	−0.138 (0.112)
−Log-lik (df)	3311 (1137)	3182 (1137)	3074 (1136)

Model 1: negative binomial model: P-$G(0,\ \gamma,\ \theta/(\lambda_i u_i))$

Model 2: inverse-Gaussian model: P-$G(0.5,\ (\lambda_i u_i)^{0.5}\gamma,\ \theta/(\lambda_i u_i))$

Model 3: Full mixture model: P-$G(\alpha,\ (\lambda_i u_i)^{\alpha}\gamma,\ \theta/(\lambda_i u_i))$

identifiability of all the shape parameters of the mixture distribution. Table 7.2 shows the results for models including *type*, *age* and *trial*, while table 7.3 shows results for models including type. Observe that the negative binomial models in table 7.1 (second column) and table 7.2 (first column) differ only in parameterisation, where the intercept term $\widehat{\beta}_0$ in the first table is replaced by a second scale parameter $\widehat{\theta}$ in the second table. The exact relationship is $\widehat{\gamma}/\widehat{\theta} = \exp(\widehat{\beta}_0)$, and, from the table, $\log(1.221/26.006) = -3.059$.

In table 7.3 the model estimates of $\mathbb{E}[X]$, $\mathrm{Var}[X]$ and the median count, $\mathrm{med}[X]$ are also presented. The negative binomial model fits closest to the observed mean, but does not fit the observed variance and median very well. The full P-G model gives the best fit to the observed median count. These model fits will be investigated in more detail.

In tables 7.2 and 7.3, the fitted log-likelihood for the full P-G models is much higher than for the inverse Gaussian or negative binomial special cases. To investigate further the fit of these various models, histograms may be created of the observed and fitted distributions of counts. This is done for the models shown in table 7.3, where the only covariate included in the model is epilepsy type (taking values ± 1). Figure 7.1 shows the observed distribution of counts in the two groups. Note that in the generalised group, nearly 40% of the individuals experienced exactly 2 seizures, which is a massive proportion for a Poisson mixed model to cope with.

Figure 7.2 shows the fitted distribution of counts for the negative binomial model. Notice that in these models, the mode is at 0, in contrast to the observed distributions where the mode is at 2. It is clear that this fitted distribution is heavily

influenced by a few large counts.

Figure 7.3 shows the fitted distribution of counts for the Poisson-inverse-Gaussian mixture models. These models looks more like the observed distribution, although the generalised epilepsy model still fails to pick up the enormous spike at 2.

Figure 7.4 shows the fitted distribution of counts for the full P-G model. These

Table 7.3: Estimates (standard errors) for P-G Models (2). The models include just one covariate, *epilepsy type*.

Regression Coefficient	Model 1 estimate (s.e.)	Model 2 estimate (s.e.)	Model 3 estimate (s.e.)
α	0	0.5	0.823 (0.012)
γ	0.898 (0.023)	0.104 (0.003)	0.033 (0.001)
θ	24.209 (0.857)	7.510 (0.746)	0.383 (0.251)
β_{type}	0.257 (0.035)	0.191 (0.046)	0.144 (0.027)
$-$Log-lik (df)	3498 (1141)	3263 (1141)	3127 (1140)
$\mathbb{E}[X]$ generalised	5.22	5.71	6.10
$\mathbb{E}[X]$ partial	8.73	8.36	8.13
Var$[X]$ generalised 0	35.58	62.89	449.46
Var$[X]$ partial	93.55	130.99	797.12
med$[X]$ generalised	3	3	3
med$[X]$ partial	6	5	4

Model 1: negative binomial model: P-$G(0,\ \gamma,\ \theta/(\lambda_i u_i))$

Model 2: inverse-Gaussian model: P-$G(0.5,\ (\lambda_i u_i)^{0.5}\gamma,\ \theta/(\lambda_i u_i))$

Model 3: Full mixture model: P-$G(\alpha,\ (\lambda_i u_i)^{\alpha}\gamma,\ \theta/(\lambda_i u_i))$

For type 0, observed $\mu_x = 5.22$, $\sigma_x^2 = 101.83$, med$[X] = 3$

For type 1, observed $\mu_x = 8.73$, $\sigma_x^2 = 283.71$, med$[X] = 4$

look better than the inverse-Gaussian plots; note that now the mode has increased to 2 and 3, respectively, and the right tail fits more closely to the observed data. However, even this model does not give a good fit for the spike at $X_i = 2$.

Of all three models, the full P-G model looks least influenced by the few large outlying counts (demonstrated by the lowest fitted mean in table 7.3). In fact, this P-G model doesn't look very different from a standard Poisson, although the fitted log-likelihood reveals that it fits much better than that simpler model. However, none of these models give a particularly good fit to the observed count data, particularly to the large spike where 40% of individuals in the generalised-onset epilepsy group experienced exactly two seizures in 6 months.

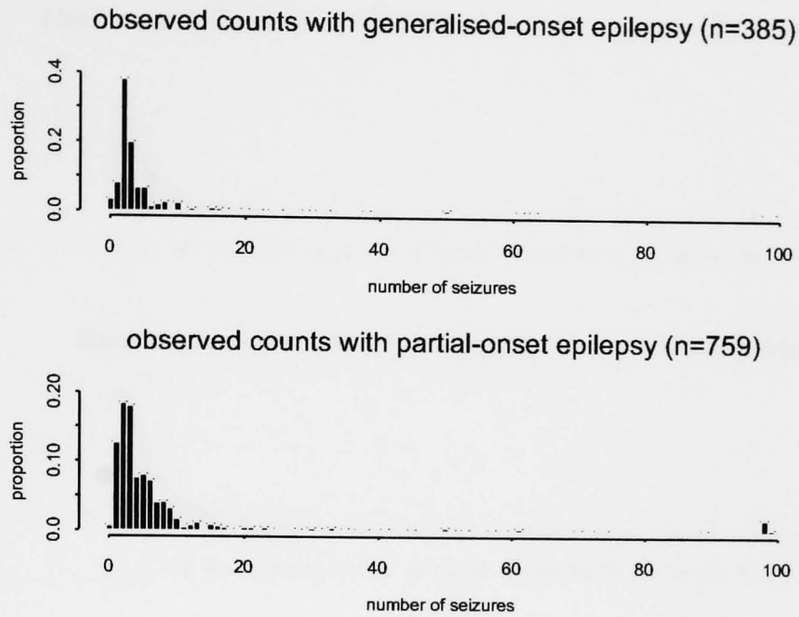

Figure 7.1: Observed distribution of seizure counts

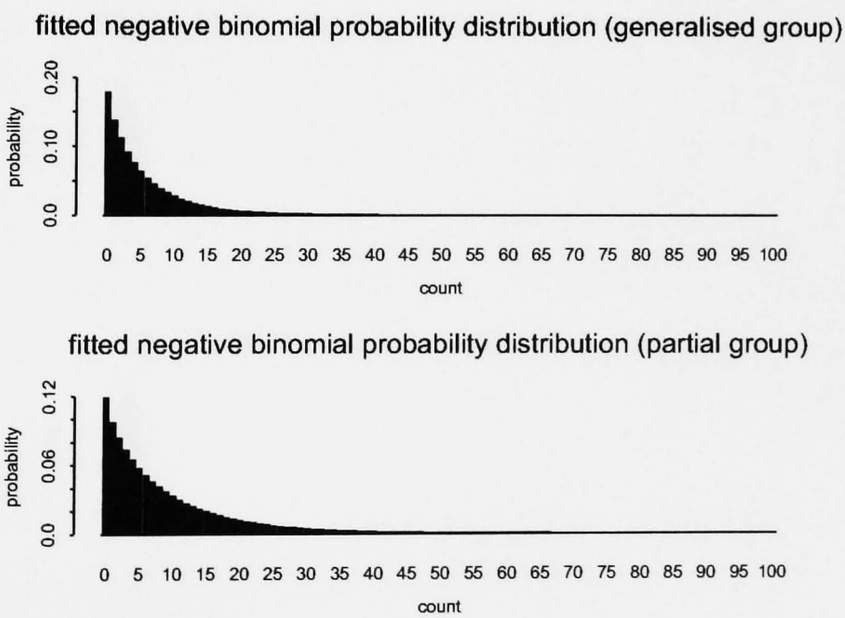

Figure 7.2: Fitted negative binomial distributions

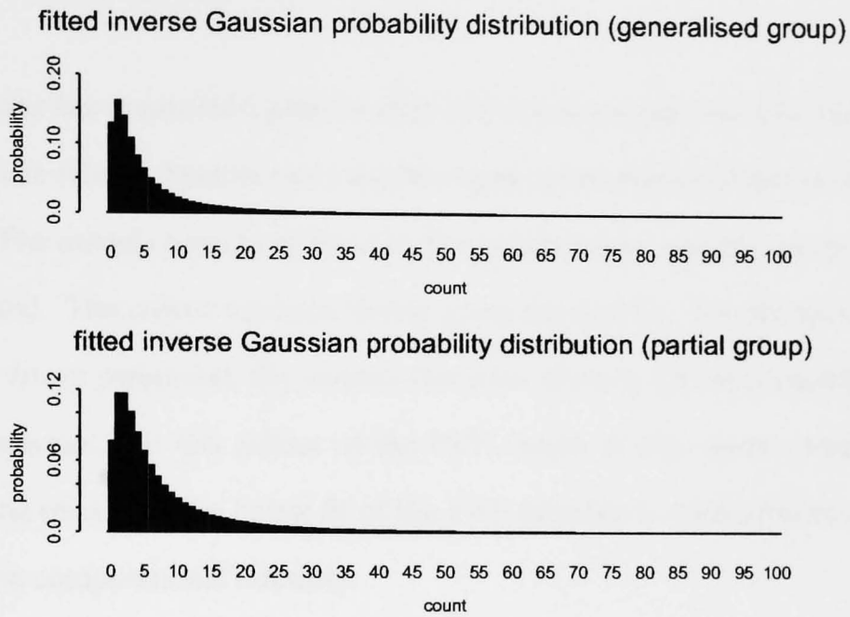

Figure 7.3: Fitted Poisson-inverse-Gaussian mixture distributions

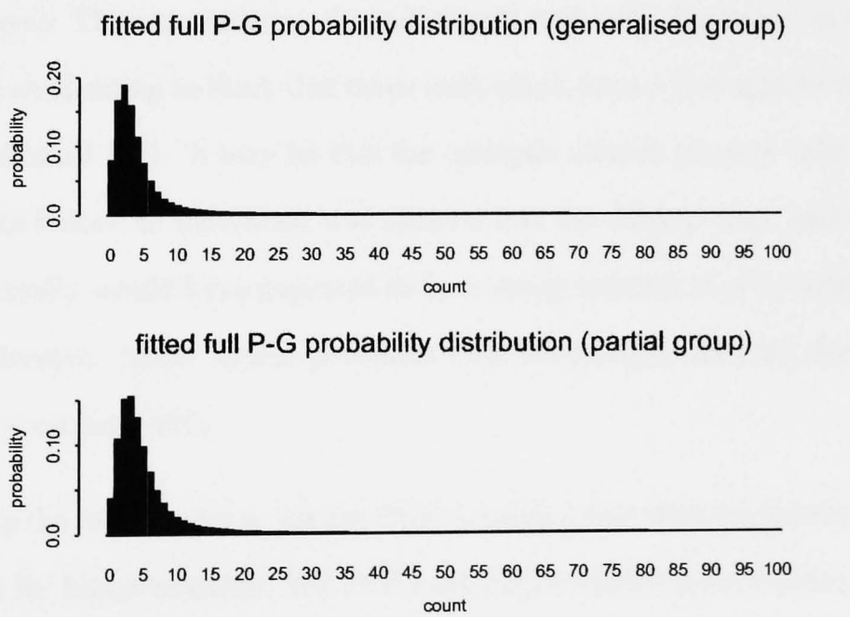

Figure 7.4: Fitted Full P-G mixture distributions

7.6 Discussion

This chapter has presented a general class of Poisson mixture models which allow for covariate effects. Special cases are the negative binomial and inverse-Gaussian models. The models have been fitted to the epilepsy data, and the goodness-of-fit investigated. The power variance family gives the best fit. For the special cases with one fewer parameter, the inverse-Gaussian mixture seems preferable to the gamma mixture, but this subset of the PVF family is also more complex than the gamma mixture. The better fit of the PVF mixture is counterbalanced by an increase in computational intensity.

One problem with the epilepsy data is that, although the seizure count is supposedly a 6-month count, most of the individuals are newly diagnosed epileptics. Epilepsy is generally diagnosed after two seizures, and treatment is given soon after diagnosis. Thus, considering the individuals with only 2 seizures in 182 days, it may be misleading to think that these individuals have a low seizure rate going into the clinical trial. It may be that the epileptic seizure process only started a few weeks before an individual was entered into the clinical trial, and that individual actually would have expected to have many seizures in a 6-month period, if left untreated. Some further problems with the epilepsy data are discussed in section 9.4 on page 166.

Extending the joint model to use the PVF mixture rather than the gamma mixture is an area for future research. The PVF mixture is certainly more flexible, but also more computationally demanding. A Bayesian approach to this problem might be computationally easier than a maximum likelihood approach.

Chapter 8

Further Extensions to the Joint Model

This chapter considers two extensions to the joint model of chapter 4. The first extension is the introduction of covariate effects to the overdispersion parameter α. This model is described in section 8.1, and illustrated on two subsets of the epilepsy data in sections 8.2 and 8.3.

The second extension considers the implication of clinical information on the results of fitting the joint model to the epilepsy data in chapter 5. It has been suggested that there is a missing binary covariate, recording the particular type of seizure experienced, which would be present in nearly all the high-seizure individuals, but missing in nearly all of the low-seizure individuals. Such a covariate, if included in the model, would probably facilitate a much better fit of the joint model. Section 8.4 explores various schemes for the regeneration of this missing covariate, and the possible change in conclusions which follow.

8.1 Regression on α

The epilepsy data, first described in chapter 3, contain individual patient data from five large randomised controlled trials of two treatments for epilepsy. In the application of the joint model to the epilepsy data, in chapter 5, it became clear that one of these five trials (Mattson *et al.*, 1992) was quite different to the other trials, in terms of the distribution of seizure counts. As can be seen from table 3.3 on page 20, there is very little evidence of overdispersion in the 6-month pre-randomisation counts for this trial, compared to the massive overdispersion in the other four trials. The joint model described in chapter 4 allows for covariates to affect the underlying individual event rates. However, the joint model does not allow covariates to affect the *amount of difference* between individuals. It is interesting to investigate whether the estimated heterogeneity is different depending on the value of an explanatory variable.

This section proposes the introduction of covariate effects through the heterogeneity parameter α of the joint model. Recalling the graphical model shown in figure 4.2 on page 34 with associated equations, the model proposed in this section relates the heterogeneity parameter α to covariates through a log-link. Thus the model specifies $\alpha_i = \exp(\boldsymbol{\xi}' \mathbf{z}_{3i})$, where $\boldsymbol{\xi}$ is a vector of regression coefficients, and the vector of covariates $\mathbf{z}_{3i}$ is chosen in some way from the set of pre-randomisation covariates.

Figure 8.1 shows the proposed model. The circular nodes represent variables, either parameters or data, and the square nodes are logical. The model is specified

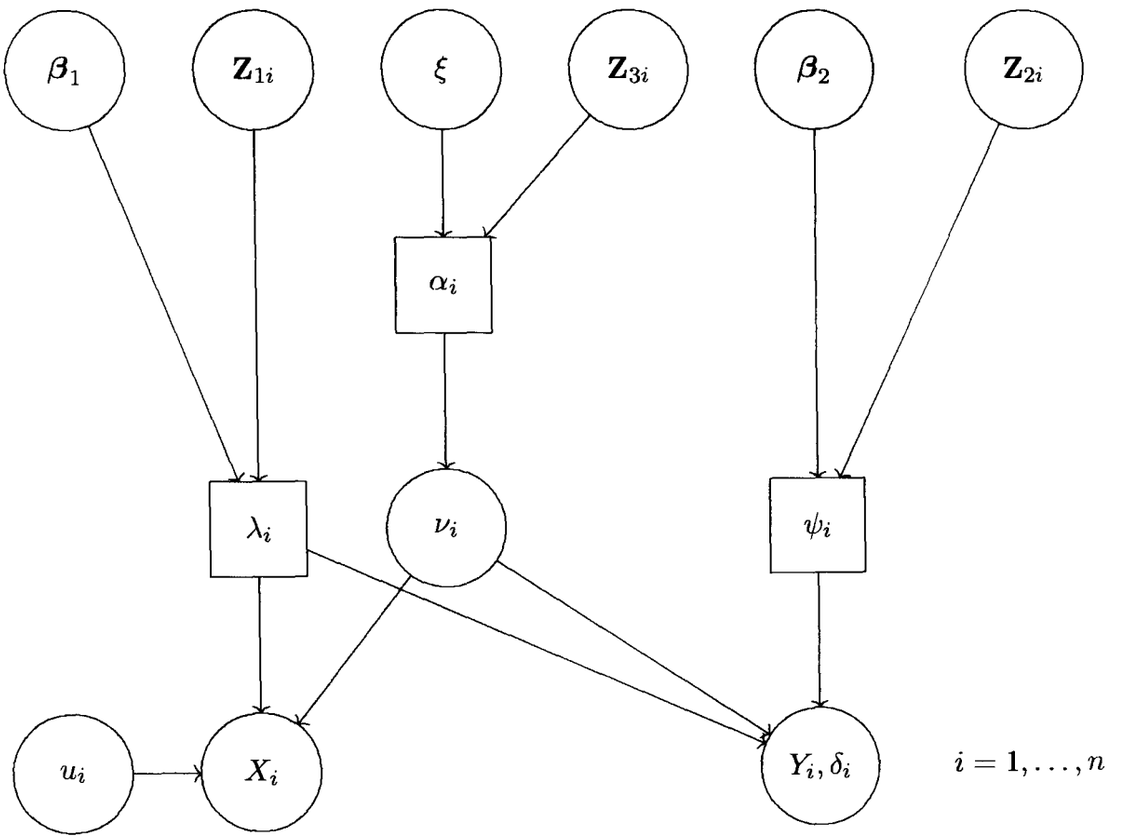

Figure 8.1: Graphical Model of New Joint Model

by the equations:

$$f_X(x_i \mid \lambda_i, u_i, \nu_i) = \frac{(\lambda_i u_i \nu_i)^{x_i} \exp(-\lambda_i u_i \nu_i)}{x_i!}, \tag{8.1}$$

$$f_Y(y_i \mid \lambda_i, \psi_i, \nu_i) = \lambda_i \psi_i \nu_i \exp(-\lambda_i \psi_i \nu_i y_i), \tag{8.2}$$

$$g_\nu(\nu_i \mid \alpha_i) = \frac{\alpha_i^{\alpha_i} \nu_i^{\alpha_i - 1} \exp(-\alpha_i \nu_i)}{\Gamma(\alpha_i)}, \tag{8.3}$$

where

$$\lambda_i = \exp(\boldsymbol{\beta}_1' \mathbf{z}_{1i}), \tag{8.4}$$

$$\psi_i = \exp(\boldsymbol{\beta}_2' \mathbf{z}_{2i}) \tag{8.5}$$

$$\alpha_i = \exp(\boldsymbol{\xi}' \mathbf{z}_{3i}). \tag{8.6}$$

The log-likelihood may be derived following a similar procedure to the one given in chapter 4, see pages 36 to 40. The full log-likelihood ℓ_n for the data $\mathcal{D}$ on all the n individuals is given by

$$\begin{aligned}\ell_n(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\xi} \,|\mathcal{D}) = \sum_{i=1}^{n} \Big\{ \sum_{j=0}^{x_i-1} & \ln(\alpha_i + j) + \delta_i \ln(\alpha_i + x_i) + x_i \ln(u_i) \\ & + \alpha_i \ln(\alpha_i) + (x_i + \delta_i) \ln(\lambda_i) + \delta_i \ln(\psi_i) \\ & - \ln(x_i!) - (x_i + \alpha_i + \delta_i) \ln(\lambda_i u_i + \lambda_i \psi_i y_i + \alpha_i) \Big\},\end{aligned}$$

where λ_i, ψ_i and α_i are specified by equations (8.4), (8.5) and (8.6). The first and second derivatives are straightforward but messy.

To complete the Bayesian model, in addition to the relationships specified in equations (8.1) through (8.6), the prior distributions of the parameters $\boldsymbol{\xi}$, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ must be specified. In the absence of clinical or other expert information, it is suggested that vague normal priors are used, so that

$$\begin{aligned}\pi(\boldsymbol{\xi}) &\sim \mathbf{N}(\mathbf{0}, m\mathbf{I}) \\ \pi(\boldsymbol{\beta}_1) &\sim \mathbf{N}(\mathbf{0}, m\mathbf{I}) \\ \pi(\boldsymbol{\beta}_2) &\sim \mathbf{N}(\mathbf{0}, m\mathbf{I})\end{aligned}$$

for some large m, where $\mathbf{I}$ is the identity matrix. Using these vague priors, it is strongly recommended that covariates are suitably scaled.

The model may be fitted either using maximum likelihood methods, or Bayesian methods. In the following sections, the model is illustrated on two subsets of the epilepsy data, with MCMC simulation using the software package WinBUGS (Spiegelhalter *et al.*, 2000).

8.2 Example 1: Epilepsy Data

The joint model as specified by the graphical model in figure 4.2 on page 34 may also be implemented using Bayesian methodology. The model may be fitted to the epilepsy data using the software package WinBUGS (Spiegelhalter *et al.*, 2000) to perform MCMC simulation. In this section, the results for a subset of the epilepsy data are given, and compared to the maximum likelihood estimates for the same subset.

A small subset of the data has been chosen to illustrate the method. The first subset contains 450 of the 1144 individuals, where inclusion was based on these rules:

- must have an observed survival time rather than a censored survival time;
- not in the fifth (Veterans' Affairs) trial;
- not with a very high seizure count and a very long survival time;
- not with a very low seizure count and a very short survival time.

Table 8.1: Division of individuals among categories of 'low' or 'high' 6-month pre-randomisation seizure count, and 'short' and 'long' time to post-randomisation failure.

	generalised-onset		partial-onset	
	'short' time	'long' time	'short' time	'long' time
'low' count	25	198	19	137
'high' count	31	10	84	27

The 425 patients in the fifth trial are excluded from this illustration because they are known to be different to those in the other four trials. The restriction of excluding the 188 remaining individuals with censored times allows improvement in computational time, and simplifies the model slightly. The other two conditions exclude 82 of the remaining 531 individuals, the exact breakdown is given in table 8.1. Thirty-five individuals with generalised-onset epilepsy, and 46 individuals with partial-onset epilepsy, are excluded. The divisions between categories was arbitrarily set so that a 'very high' count is more than 8 pre-randomisation seizures, and a 'very long' time is over 800 days to first post-randomisation event. A 'very low' count is less than 2 seizures, and a 'very short' time is less than 10 days to first seizure.

Some summary statistics of this subset are given in table 8.2, and a Kaplan-Meier estimate of the survival function is shown in figure 8.2. This analysis will ignore the possible interaction between treatment and epilepsy type, and focus on the difference in overdispersion between the two epilepsy types. This difference can be seen in the distribution of pre-randomisation seizure counts, in table 8.2.

Table 8.2: Distribution of 6-month pre-randomisation seizure counts in subset of epilepsy data, by epilepsy type.

Type	n	6-month Pre-randomisation count				
		mean	s.d.	median	min.	max.
generalised	229	6.16	12.07	3	0	99
partial	221	17.10	25.57	7	0	99

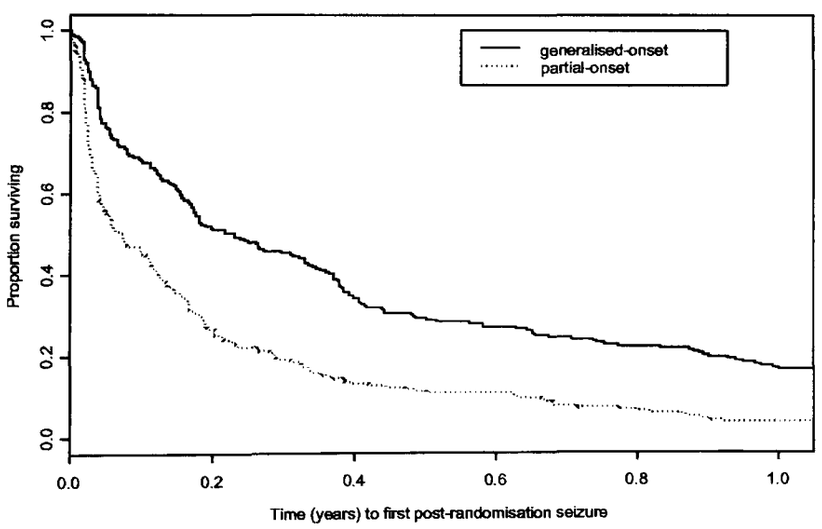

Figure 8.2: Kaplan-Meier estimate of the survival function for a subset of the epilepsy data, stratified by epilepsy type.

Table 8.3: Maximum likelihood and MCMC parameter estimates for subset of epilepsy data

Regression Coefficient	Max. Lik. Estimate (s.e.)	MCMC Estimate (s.e.)
α	0.931 (0.059)	0.927 (0.059)
β_0	−2.870 (0.051)	−2.870 (0.051)
β_{type}	0.510 (0.051)	0.514 (0.051)
β_{age}	−0.029 (0.032)	−0.028 (0.032)
β_{t0}	−0.946 (0.054)	−0.946 (0.054)
β_{trt}	0.099 (0.054)	0.099 (0.054)

Type: $-1/+1$ for generalised/partial-onset epilepsy
Age: original age $-$ 20, in decades
Trt: $-1/+1$ for CBZ/VPA

8.2.1 Results of Joint Model

The code for fitting the joint model in WinBUGS is given in appendix D, and the Bayesian results are given with the corresponding maximum likelihood estimates in table 8.3. The Bayesian model was run in WinBUGS, with the first 4000 iterations discarded as a 'burn-in' period, and the following 40000 iterations sampled. For the normal priors for β_1 and β_2, m was taken as 1000, and $\pi(\alpha) \sim \text{Gamma}(0.001, 0.001)$ was used as the prior for α.

The batch means method outlined by Roberts (1996, p. 50) was used to estimate the standard error of the MCMC estimates.

It is reassuring to notice that the maximum likelihood and MCMC estimates in

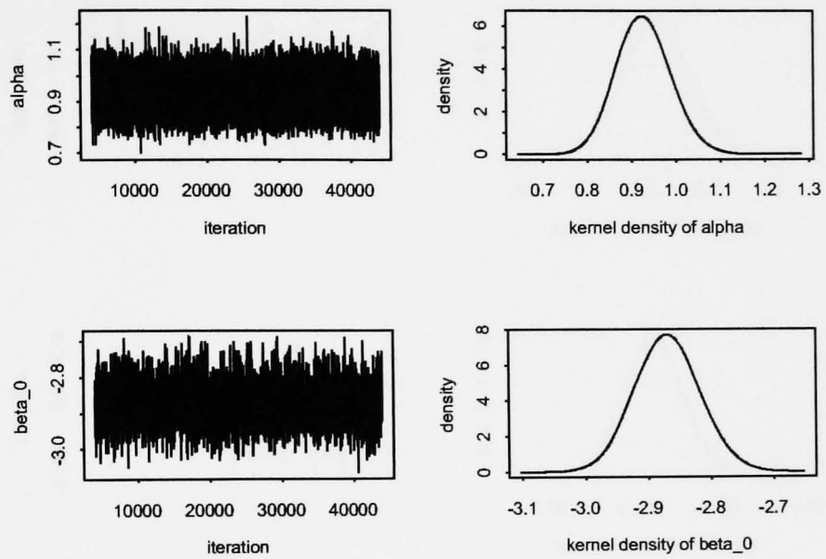

Figure 8.3: MCMC Diagnostics: trace and kernel density plots of α and β_0.

table 8.3 are almost identical, to three decimal places. Some diagnostic plots for the MCMC estimates are given in figures 8.3, 8.4 and 8.5. The diagnostics show nicely bell-shaped kernel density estimates, and traces indicating convergence.

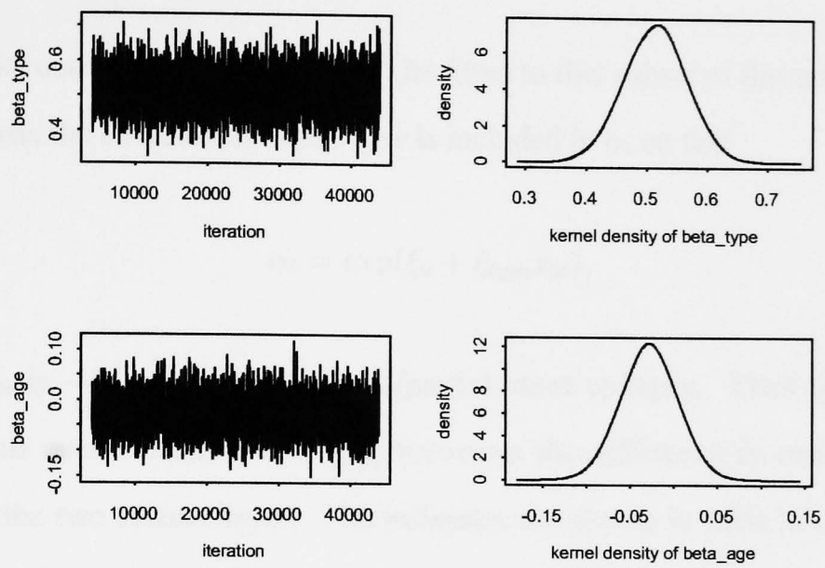

Figure 8.4: MCMC Diagnostics: trace and kernel density plots of β_{type} and β_{age}.

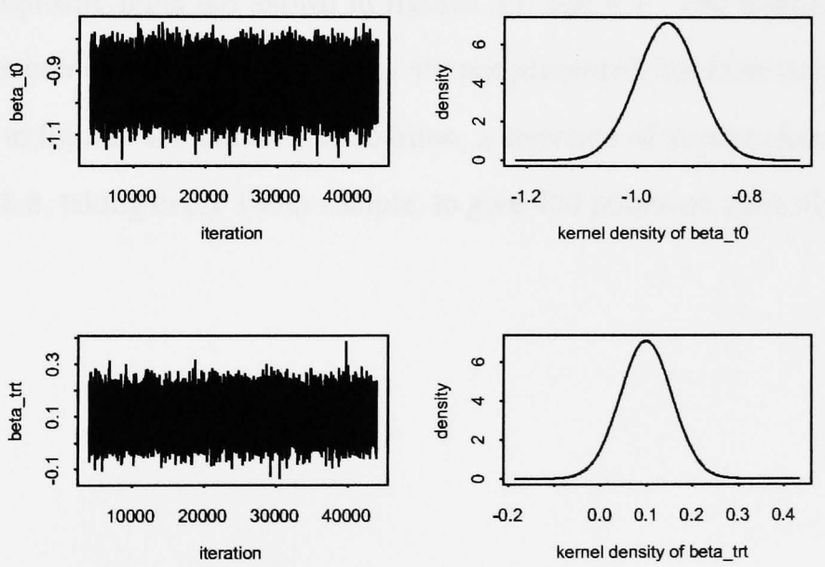

Figure 8.5: MCMC Diagnostics: trace and kernel density plots of β_{t0} and β_{trt}.

8.2.2 Results varying α by epilepsy type

The model derived in section 8.1 may be fitted to this subset of the epilepsy data. In this case, the covariate *epilepsy type* is included in α, so that

$$\alpha_i = \exp(\xi_0 + \xi_{type} z_{3i}), \tag{8.7}$$

where z_{3i} is $-1/+1$ for generalised-/partial-onset epilepsy. Thus ξ_0 measures the overall overdispersion, and ξ_{type} measures the difference in overdispersion between the two seizure types. The estimates are shown in table 8.4. Note that $\exp(\widehat{\xi_0}) = 0.936$, which is very close to the estimated $\widehat{\alpha}$ in table 8.3. Note also that $\Pr(\widehat{\xi_{type}} < 0) \approx 0.025$, so epilepsy type seems to be important in considering the overdispersion between individuals.

Some diagnostic plots are shown in figures 8.6 and 8.7. The traces and kernel density estimates for β_{age}, β_{t0} and β_{trt} are not presented, but look very similar to the plots in figures 8.4 and 8.5. In addition, a selection of scatter plots are shown in figure 8.8, taking every 100th sample, to give 400 points on each plot.

Table 8.4: MCMC parameter estimates allowing overdispersion to vary by type

Regression Coefficient	MCMC Estimate (s.e.)
ξ_0	-0.066 (0.064)
ξ_{type}	-0.124 (0.065)
β_0	-2.868 (0.051)
β_{type}	0.511 (0.053)
β_{age}	-0.029 (0.032)
β_{t0}	-0.952 (0.055)
β_{trt}	0.100 (0.054)

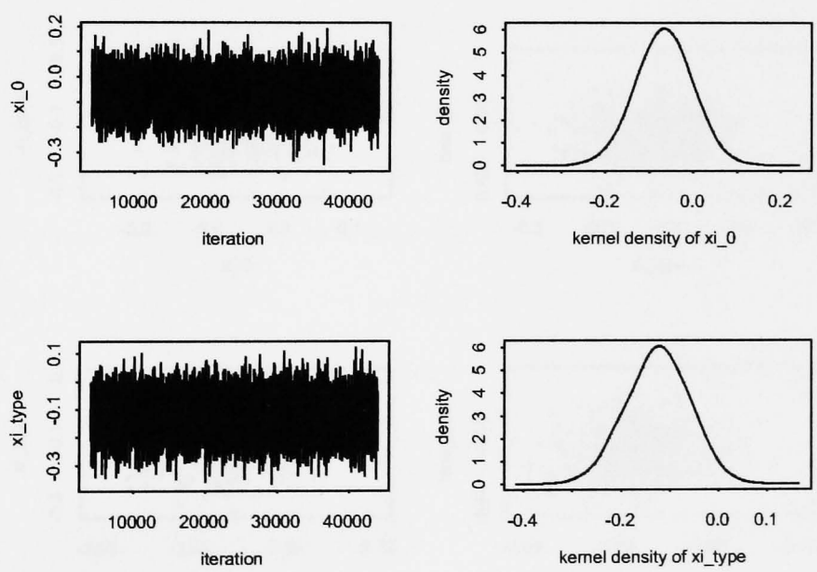

Figure 8.6: MCMC Diagnostics: trace and kernel density plots of ξ_0 and ξ_{type}.

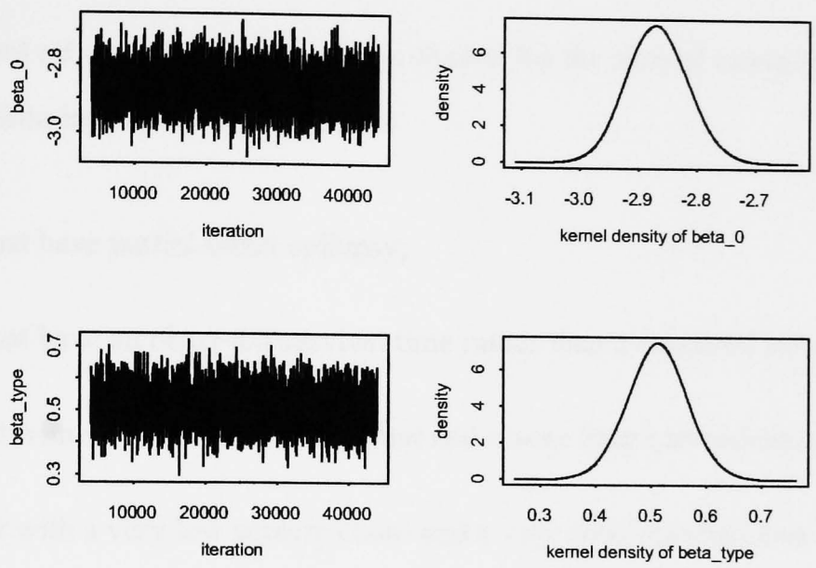

Figure 8.7: MCMC Diagnostics: trace and kernel density plots of β_0 and β_{type}.

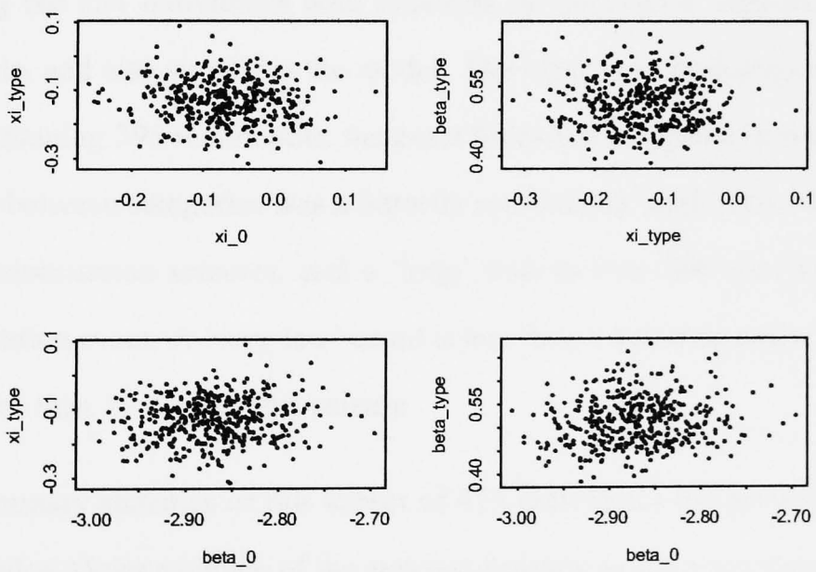

Figure 8.8: Bivariate scatter plots of selected variables.

8.3 Example 2: Epilepsy Data

A different subset of the data has been chosen for the second example. The rules for inclusion in this second subset are:

- must have partial-onset epilepsy;
- must have an observed survival time rather than a censored survival time;
- not with a very high seizure count and a very long survival time;
- not with a very low seizure count and a very short survival time.

Only patients with partial-onset epilepsy are included in this illustration to give a more suitable comparison with trial 5, which contains only partial-onset epileptics. This excludes 325 of the original 1144 patients. The second condition of excluding the 224 individuals with censored survival times improves computational time, and also simplifies the model. The other two conditions exclude 180 of the remaining 595 individuals, the exact breakdown is given in table 8.5. The divisions between categories was arbitrarily set so that a 'high' count is more than 8 pre-randomisation seizures, and a 'long' time is over 800 days to first post-randomisation event. A 'very low' count is less than 2 seizures, and a 'very short' time is less than 10 days to first seizure.

Some summary statistics of this subset of 415 individuals are given in table 8.6, and a Kaplan-Meier estimate of the survival function is shown in figure 8.9. This example focuses on the difference in overdispersion between the five trials, which can be seen in the distribution of pre-randomisation seizure counts, particularly for

Table 8.5: Division of individuals among categories of 'low' or 'high' 6-month pre-randomisation seizure count, and 'short' and 'long' time to post-randomisation failure.

	'short' time	'long' time
'low' count	120	331
'high' count	84	60

Table 8.6: Distribution of 6-month pre-randomisation seizure counts in subset of epilepsy data, by trial.

Trial	n	6-month pre-randomisation count				
		mean	s.d.	median	min.	max.
1	33	14.76	21.74	6	0	89
2	37	12.43	15.88	7	0	73
3	86	16.20	20.63	9	2	98
4	70	23.03	36.46	5	2	99
5	189	3.64	2.34	3	1	10

trial 5 (the Veterans' Affairs trial) in table 8.6. Primary interest is in the difference between trial 5 and the other 4 trials, so for the following analyses, the first four trials are grouped together, and a single covariate is used to indicate whether or not the individual was in trial 5.

In subsection 8.3.1 the maximum likelihood and MCMC results for the joint model are presented. In subsection 8.3.2, the MCMC results are presented for the model including trial in the regression coefficient α_i.

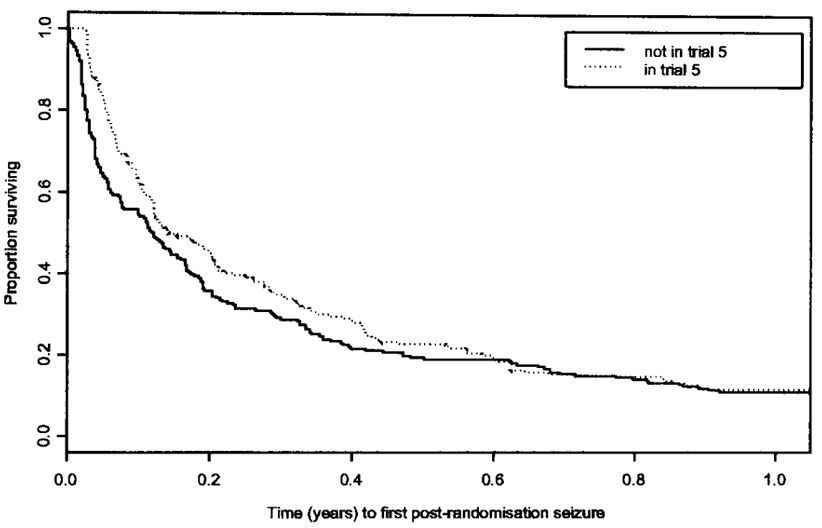

Figure 8.9: Kaplan-Meier estimate of the survival function for a subset of the epilepsy data, stratified by trial.

8.3.1 Results of Joint Model

The MCMC results for the joint model fitted to these data are given with the corresponding maximum likelihood estimates in table 8.7. The estimates are very similar. The MCMC chain was run for 44000 iterations, discarding the first 4000 as a 'burn-in' period.

Some diagnostic plots for the MCMC estimates are given in figures 8.10, 8.11 and 8.12. The diagnostics show bell-shaped kernel density estimates, and traces indicating good mixing.

Table 8.7: Maximum likelihood and MCMC parameter estimates for subset of epilepsy data

Regression Coefficient	Max. Lik. Estimate (s.e.)	MCMC Estimate (s.e.)
α	1.165 (0.080)	1.159 (0.079)
β_0	−3.082 (0.051)	−3.079 (0.051)
β_{age}	−0.029 (0.030)	−0.031 (0.030)
β_{trial5}	−0.701 (0.060)	−0.700 (0.062)
β_{t0}	−1.210 (0.057)	−1.211 (0.058)
β_{trt}	0.043 (0.057)	0.043 (0.057)

Age: original age − 30, in decades

Trial5: $-1/+1$ for excluded/included in fifth trial

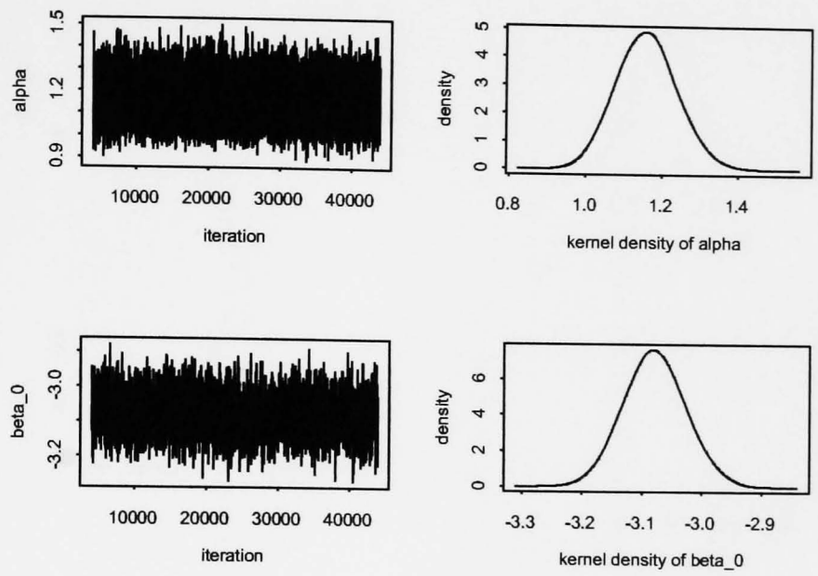

Figure 8.10: MCMC Diagnostics: trace and kernel density plots of α and β_0.

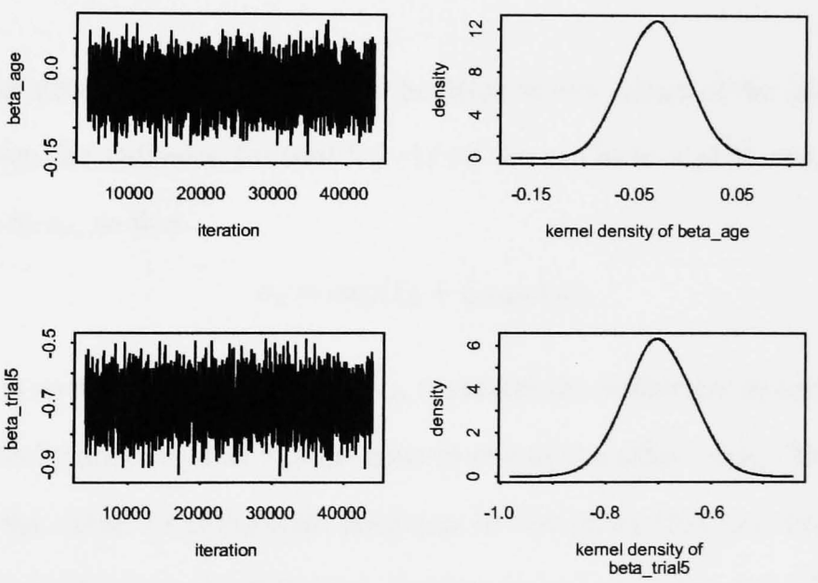

Figure 8.11: MCMC Diagnostics: trace and kernel density plots of β_{age} and β_{trial5}.

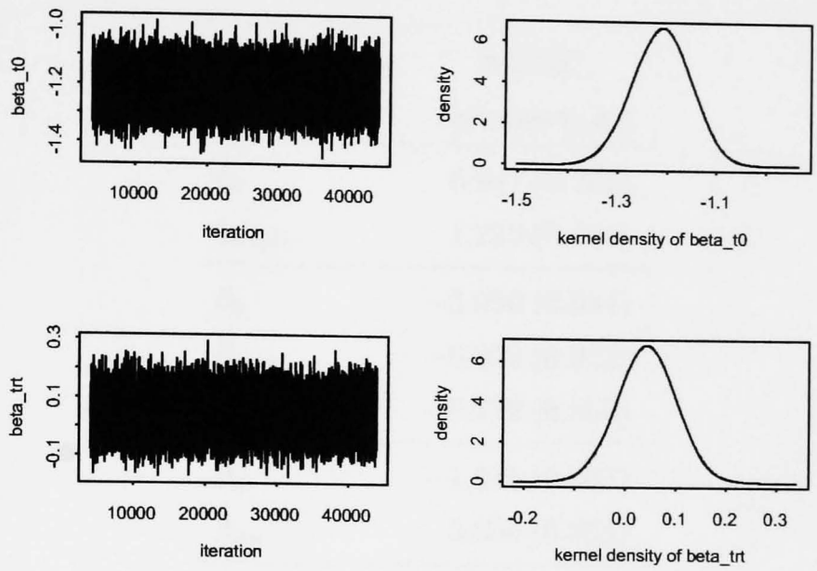

Figure 8.12: MCMC Diagnostics: trace and kernel density plots of β_{t0} and β_{trt}.

8.3.2 Results varying α by trial

The model derived in section 8.1 may be fitted to this subset of the epilepsy data. In this case, the indicator for trial 5 ($-1/+1$ for not in/in trial 5) was chosen for inclusion in α_i, so that

$$\alpha_i = \exp(\xi_0 + \xi_{trial5} z_{3i}), \tag{8.8}$$

where ξ_0 is an intercept term, and ξ_{trial5} measures the difference in overdispersion between individuals in trial 5, and those in one of the other trials. The covariate z_{3i} takes the value -1 if the individual was in one of the first four trials, and $+1$ if the individual was in the fifth trial. Further investigation into the differences in overdispersion between the five trials may be investigated in the future.

The MCMC estimates are shown in table 8.8. As expected, the estimate of $\widehat{\xi}_{trial5}$

Table 8.8: MCMC parameter estimates allowing overdispersion to vary by trial

Regression Coefficient	MCMC Estimate (s.e.)
ξ_0	0.997 (0.162)
ξ_{trial5}	1.253 (0.161)
β_0	−3.090 (0.044)
β_{age}	−0.022 (0.023)
β_{trial5}	−0.722 (0.050)
β_{t0}	−1.245 (0.057)
β_{trt}	0.056 (0.055)

is highly significant, since trial 5 is much less overdispersed than the others, as can be seen from table 8.6.

For individuals not included in trial 5, the estimate $\widehat{\alpha}_i = 0.77$, indicating a high level of overdispersion. For individuals in trial 5, the estimate $\widehat{\alpha}_i = 9.49$, indicating very little overdispersion.

Some diagnostic plots are shown in figures 8.13 and 8.14. The traces and kernel density estimates for β_{age}, β_{t0} and β_{trt} are not presented, but look very similar to the plots in figures 8.11 and 8.12. In addition, a selection of scatter plots are shown in figures 8.15 and 8.16, taking every 100th sample, to give 400 points on each plot.

The shape of the first scatter plot in figure 8.15 shows a strong association between the sampled values of ξ_0 and ξ_{trial5}. One explanation for this is that α_i is well determined for individuals in trials 1-4, but not for those in trial 5.

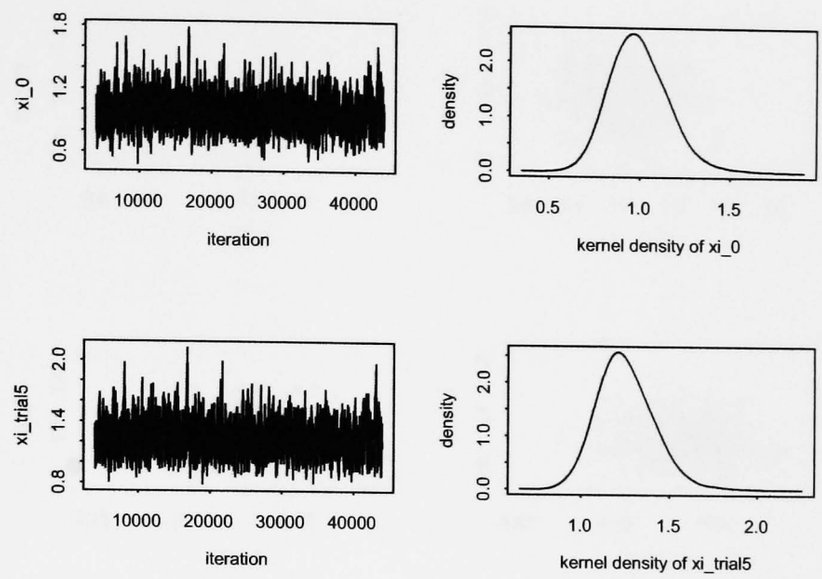

Figure 8.13: MCMC Diagnostics: trace and kernel density plots of ξ_0 and ξ_{trial5}.

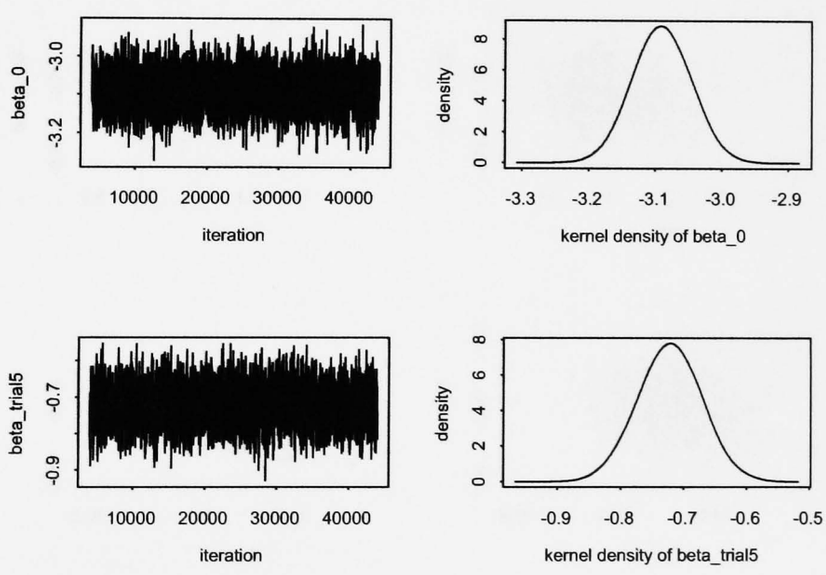

Figure 8.14: MCMC Diagnostics: trace and kernel density plots of β_0 and β_{trial5}.

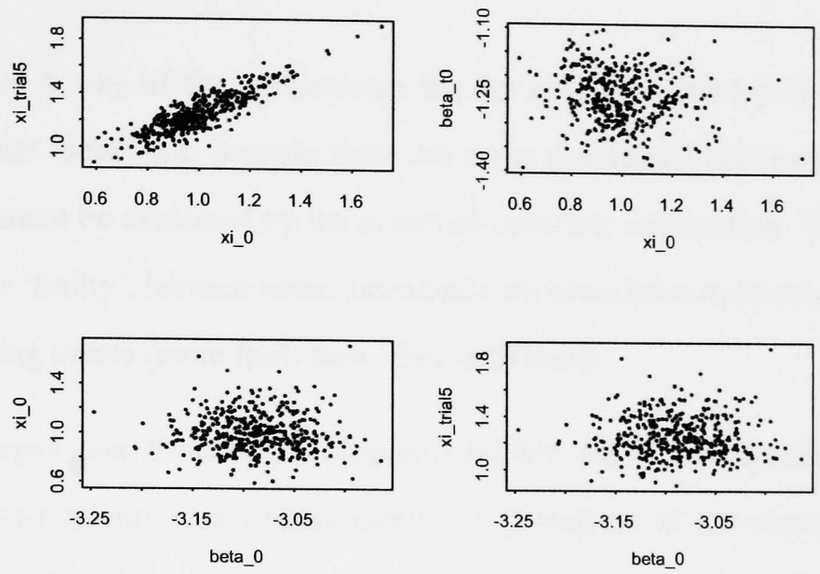

Figure 8.15: Bivariate scatter plots of selected variables.

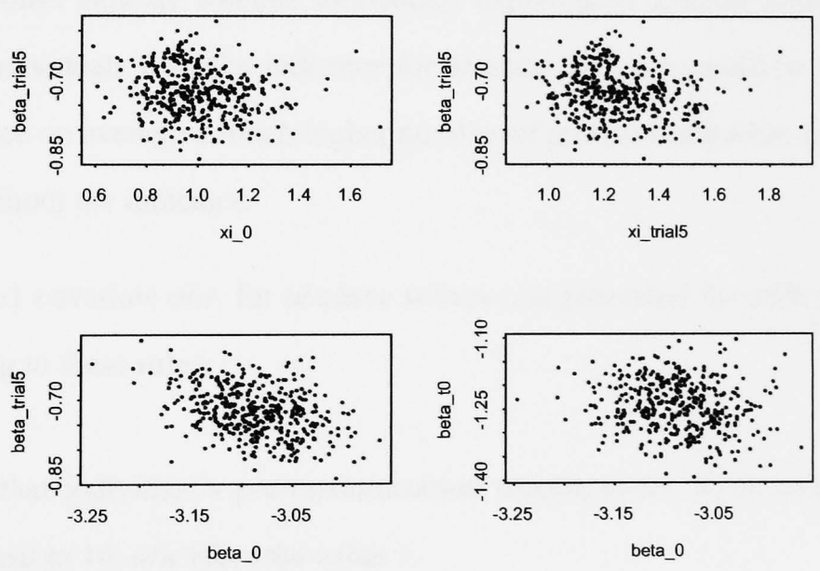

Figure 8.16: Bivariate scatter plots of selected variables.

8.4 Missing Covariate

In chapter 4, one of the explanations for the need to allow for overdispersion in the joint model was because there are some differences between individuals which cannot be explained by the observed covariate information. This is often known as 'frailty', because some individuals are considered to be at a higher risk of suffering events (more frail) than other individuals.

The epilepsy data, described in chapter 3 exhibits clear overdispersion in the 6-month pre-randomisation seizure counts. A histogram of the observed seizure counts, stratified by seizure type, was shown in figure 7.1 on page 124.

In a discussion with Dr Tony Marson, a neurologist at the Walton Centre, Liverpool, it was suggested that there is a binary covariate missing from the data, which would indicate whether individuals experienced absence seizures or not. Those individuals with the indicator for absence seizures would be expected to experience on average a much higher number of pre-randomisation seizures than those without the indicator.

A new ± 1 covariate *abs*, for *absence seizures*, is generated for each individual i according to these rules:

- If that individual's pre-randomisation seizure count X_i is greater than or equal to 10, *abs* takes the value 1.
- If that individual's pre-randomisation seizure count X_i is between 0 and 9, *abs* takes the value 1 with probability $(X_i/10)$, and -1 otherwise. So an individual with 9 pre-randomisation seizures has probability $9/10$ of being

classified into the *abs*=1 group.

For the full epilepsy data, this scheme assigns approximately 500 of the 1144 individuals to have *abs*=1. There are only 160 individuals with 10 or more pre-randomisation seizures.

Histograms of the observed distribution of seizure counts, stratified by absence seizures and the first imputation of the new covariate, are shown in figures 8.17 and 8.18.

It is of interest to generate the new covariate *abs* more than once, and investigate the joint model estimates, including the new covariate. For the purposes of this illustration, the covariate will be imputed just five times. An overview of the five imputations of this new covariate are given in table 8.9.

In each imputation, 160 individuals are automatically assigned to *abs*=1. Taking the first two rows of the table, for the first imputation, 496 of the 1144 individuals were assigned to have the indicator of absence seizures, and in the second imputation, 469 individuals were assigned this indicator. Since 160 individuals were pre-specified to have the indicator, and the other 984 individuals had varying probability of being assigned to the absence seizure group. Thus in the first two imputations, 336 and 309 individuals were randomly assigned to the absence seizure group. If these 645 individuals were completely different, the entry in the second column of the first row would read 645. In fact, table 8.9 shows that 367 individuals were given different values of the indicator, in the first two imputations. The table also shows that 299 individuals were assigned to have *abs*=1 in both groups, of which 139 individuals were randomly assigned to that group in

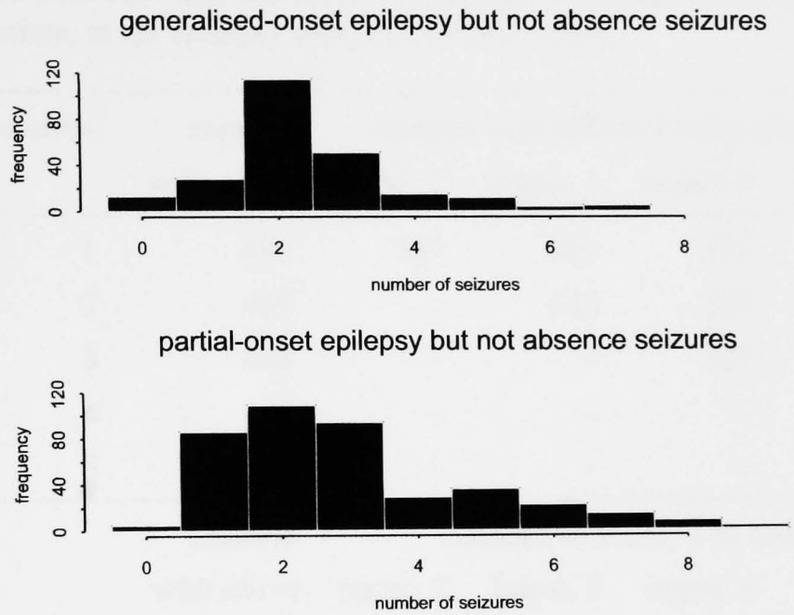

Figure 8.17: Observed seizure counts, for first imputation of new covariate.

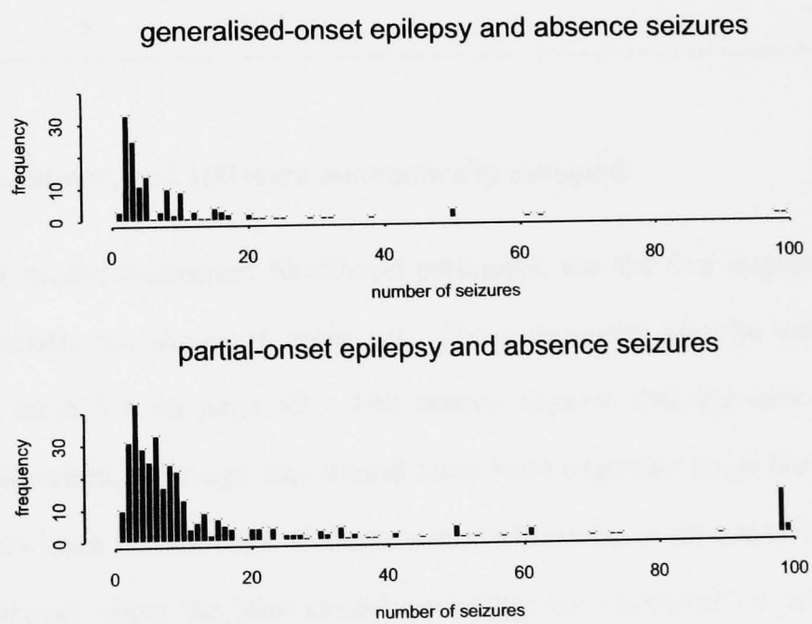

Figure 8.18: Observed seizure counts, for first imputation of new covariate.

Table 8.9: Overview and summary of correlation between five imputations of a new covariate, in the epilepsy data of 1144 individuals.

Imputation	number	number with different value of *abs* in			
	with *abs*=1	imput. 2	imput. 3	imput. 4	imput. 5
1	496	367	386	353	373
2	469	–	315	328	336
3	498	–	–	357	355
4	493	–	–	–	358
5	477	–	–	–	–
	number	number with *abs*=1 in both			
	with *abs*=1	imput. 2	imput. 3	imput. 4	imput. 5
1	496	299	304	318	300
2	469	–	326	317	305
3	498	–	–	317	310
4	493	–	–	–	306
5	477	–	–	–	–

both imputations, and 160 were automatically assigned.

The joint model maximum likelihood estimates, for the five imputations of the new covariate, are shown in table 8.9. These estimates may be compared with those in table 5.1 on page 49. The results suggest that the new covariate is highly important, although this would have been expected from the way the covariate has been constructed. Perhaps more interesting is the massive increase in log-likelihood, since the joint model excluding the simulated covariate has log-likelihood -9127. Thus the additional covariate explains a lot of variation in the epilepsy data, and knowledge about the true classification would be very useful.

Table 8.10: New covariate: maximum likelihood parameter estimates for joint model

Term	Regression Coefficient	Imputation 1 m.l.e. (s.e.)	Imputation 2 m.l.e. (s.e.)	Imputation 3 m.l.e. (s.e.)	Imputation 4 m.l.e. (s.e.)	Imputation 5 m.l.e. (s.e.)
	α	1.913 (0.094)	2.009 (0.100)	1.882 (0.092)	1.884 (0.091)	1.891 (0.092)
	β_0	−3.357 (0.078)	−3.332 (0.076)	−3.327 (0.079)	−3.336 (0.079)	−3.326 (0.078)
	β_{type}	0.376 (0.032)	0.348 (0.032)	0.353 (0.032)	0.349 (0.032)	0.341 (0.032)
	β_{age}	−0.014 (0.019)	−0.006 (0.019)	0.006 (0.019)	−0.004 (0.019)	−0.008 (0.019)
λ_i	β_{trial2}	0.240 (0.122)	0.252 (0.121)	0.199 (0.123)	0.214 (0.123)	0.201 (0.123)
	β_{trial3}	−0.005 (0.093)	−0.034 (0.091)	−0.070 (0.093)	−0.047 (0.093)	−0.040 (0.093)
	β_{trial4}	0.226 (0.102)	0.213 (0.101)	0.234 (0.103)	0.206 (0.103)	0.203 (0.102)
	β_{trial5}	−0.748 (0.103)	−0.717 (0.101)	−0.762 (0.104)	−0.721 (0.104)	−0.692 (0.104)
	β_{abs}	0.614 (0.027)	0.647 (0.027)	0.615 (0.028)	0.619 (0.028)	0.621 (0.028)
ψ_i	β_{t0}	−2.502 (0.041)	−2.515 (0.041)	−2.508 (0.041)	−2.507 (0.041)	−2.510 (0.041)
	β_{trt}	0.047 (0.041)	0.046 (0.041)	0.047 (0.041)	0.042 (0.041)	0.052 (0.041)
−Log-likelihood (df)		8907 (1133)	8882 (1133)	8914 (1133)	8912 (1133)	8912 (1133)

Further interest lies in investigating for an interaction between treatment and the new simulated covariate, representing absence seizures. The maximum likelihood estimates for a joint model including interaction terms are given in table 8.11. In this table, results are given for models excluding *age* and *trial* information, as the models are only for illustrative purposes. The very significant post-randomisation absence seizures effect is noticed. The values for β_{abs2} imply that individuals with *abs*= 1 have, in general, a much lower event rate after treatment. That is, individuals with supposed absence seizures, and therefore a generally high rate of seizures, react much better to treatment than those who do not experience absence seizures.

The non-significant treatment-absence seizures interaction in each case is also noted, suggesting that neither treatment is preferable for high-seizure individuals compared to low-seizure individuals.

8.5 Discussion

In the first part of this chapter, an extension to the joint model for the epilepsy data was investigated. In section 8.1, a model was proposed including covariates in the overdispersion parameter α. This model was illustrated on two subsets of the epilepsy data, in sections 8.2 and 8.3. The results for both subsets suggested that the inclusion of covariates in α may be an important improvement of the joint model. This also reveals that the original joint model may have problems modelling the structure of overdispersion in some data, perhaps because of the inflexibility of the gamma distribution.

Table 8.11: New covariate: maximum likelihood parameter estimates for joint model, including treatment interaction

Term	Regression Coefficient	Imputation 1 m.l.e. (s.e.)	Imputation 2 m.l.e. (s.e.)	Imputation 3 m.l.e. (s.e.)	Imputation 4 m.l.e. (s.e.)	Imputation 5 m.l.e. (s.e.)
	α	1.556 (0.072)	1.642 (0.077)	1.560 (0.072)	1.580 (0.073)	1.604 (0.074)
	β_0	−3.452 (0.030)	−3.434 (0.029)	−3.458 (0.030)	−3.456 (0.030)	−3.431 (0.029)
λ_i	β_{type}	0.166 (0.030)	0.143 (0.029)	0.147 (0.030)	0.148 (0.030)	0.139 (0.030)
	β_{abs}	0.787 (0.028)	0.833 (0.027)	0.807 (0.028)	0.813 (0.028)	0.813 (0.027)
	β_{t0}	−2.442 (0.045)	−2.449 (0.046)	−2.428 (0.045)	−2.431 (0.045)	−2.449 (0.045)
	β_{trt}	0.017 (0.044)	−0.006 (0.044)	−0.023 (0.044)	−0.010 (0.044)	−0.026 (0.044)
ψ_i	β_{type2}	0.098 (0.046)	0.164 (0.046)	0.143 (0.046)	0.153 (0.046)	0.138 (0.047)
	$\beta_{trt\times type}$	0.240 (0.045)	0.246 (0.046)	0.266 (0.045)	0.277 (0.046)	0.216 (0.046)
	β_{abs2}	−0.460 (0.043)	−0.612 (0.044)	−0.581 (0.043)	−0.594 (0.043)	−0.556 (0.044)
	$\beta_{trt\times abs}$	0.022 (0.042)	−0.035 (0.043)	−0.071 (0.042)	−0.105 (0.043)	−0.050 (0.043)
−Log-likelihood (df)		8923 (1134)	8845 (1134)	8883 (1134)	8872 (1134)	8879 (1134)

In the second part of this chapter, in section 8.4, it was suggested that an informative covariate was missing from the epilepsy data, and a method was proposed to investigate the impact of knowing this additional covariate. The new covariate was imputed five separate times, and the results of the joint model applied to each set of data suggest that it is very important additional information. However, this may only be an indication that the original joint model does not fit the data well, in terms of modelling the overdispersion structure or the underlying point process. These issues are discussed further in chapter 9.

It would be useful to have access to information on absence seizures, if only to confirm the results of section 8.4 that this covariate is informative. Further work could then include this covariate in the overdispersion parameter α, in the model of section 8.1. Information on seizure type should certainly be collected in future epilepsy trials.

Chapter 9

Conclusions

9.1 Overview of thesis

In chapter 3, the epilepsy data (Marson *et al.*, 2002) were described, and standard parametric and non-parametric analyses of the data were presented. In chapter 4, a joint model for data consisting of pre-randomisation event counts and post-randomisation survival times was derived and discussed. Methods for maximum likelihood and MCMC inference were suggested. The results of the joint model applied to the epilepsy data were given in chapter 5. To investigate the model fit, some diagnostics were presented. In addition, the data were reanalysed excluding data from one of the original trials, and with a reclassification scheme for the covariate *epilepsy type*, which is suspected to have been misclassified for some individuals.

In chapter 6, the relative efficiency of the joint model compared to a related sur-

vival model was discussed. The results of a simulation study were presented, accompanied by a theoretical approach. The results suggested that the joint model nearly always provides a more precise estimate of a treatment effect, and is therefore a very useful model. A discussion of the assumptions of the joint model is presented in this chapter, and it is suggested there are some situations where the joint model will be an appropriate and sensible model for a mixture of count data and survival data.

In chapters 7 and 8, some extensions to the joint model were explored. In chapter 7, a more general non-conjugate family of distributions was used for the frailty in a Poisson mixture model for the pre-randomisation seizure counts. Generalising the joint model of chapter 4 by incorporating the power variance family as the mixing distribution was discussed. However, considering the complexity of this count model, it was suggested that the results of the new count model are not so good as to encourage the derivation of a joint model based on this frailty. In chapter 8, another extension to the joint model was investigated. The joint model was modified to allow covariates to affect the shape of the mixing distribution. The new model was illustrated on two subsets of the epilepsy data, using MCMC inference. Also in chapter 8, the problem of a missing informative covariate was considered. The missing covariate was simulated, with interesting results.

9.2 Conclusions about the Epilepsy Data

The epilepsy data (Marson *et al.*, 2002) were first presented in chapter 3. The data concern five randomised controlled trials of two common drugs for epilepsy, *car-*

bamazepine (CBZ) and *sodium valproate* (VPS). Explanatory variables include *age at randomisation*, *sex*, *epilepsy type* and an indicator of which of the five trials the patient took part in. Current clinical opinion (Wallace *et al.*, 1997) is that VPS is preferable for individuals with generalised-onset epilepsies, while CBZ is preferable for individuals with partial-onset epilepsies.

A simple non-parametric analysis (p. 21) suggested that CBZ was preferable for patients with partial-onset epilepsies, but that neither drug was preferable for patients with generalised-onset epilepsies. Plots of the Kaplan-Meier estimate of the survival function were presented in figures 3.2 and 3.3 on page 23.

More complex regression models were applied in section 3.5 on page 26, but little evidence was found to support the clinical opinion, for either *epilepsy type*.

In chapter 5, the joint model was applied to the epilepsy data, and the results strongly suggested a preference for VPS over CBZ, for generalised-onset epilepsies, and for CBZ over VPS, for partial-onset epilepsies. This agrees with clinical opinion. Re-analyses were also presented, with the data excluding one of the trials, and with a reclassification scheme for the covariate *epilepsy type*, and of the epilepsy data stratified by *epilepsy type*. The results of these analyses did not differ a great deal from the original results, and did not change the conclusions. However, some concerns were raised about the appropriateness of the joint model for the epilepsy data, and some discussion to the assumptions of the joint model will be presented in the following sections.

Finally, it should be mentioned that efficacy is not the only important property of a treatment for epilepsy. Both CBZ and VPS can have undesirable side-effects, although the side-effect profiles are different. In particular, VPS will rarely be

prescribed to a woman aged between 15-35, because of possible complications if the patient becomes pregnant.

9.3 Assumptions of the Joint Model

In this section, consideration is given to the assumptions used in the joint model derived in chapter 4. Discussion begins with the suitability of assumptions about the model at the individual level, and continues to the population-level assumptions.

9.3.1 Joint Model at Individual Level

Figure 9.1 depicts the underlying hazard rate, $h(t)$, of an individual, under the assumptions of the joint model presented in chapter 4. The model assumes that, for this individual, events occur according to a Poisson process with a constant rate. The individual has a constant hazard before treatment, and then a different constant hazard after treatment. If the treatment is effective, the post-randomisation hazard will be lower than the pre-randomisation hazard.

This section considers two of the basic assumptions of the joint model. Firstly, the assumption of a constant hazard, and secondly, the assumption of an instantaneous treatment effect.

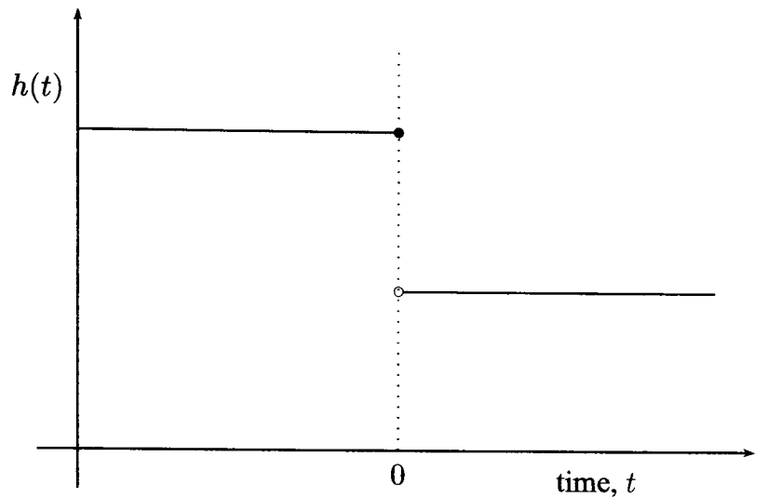

Figure 9.1: Individual hazard rate according to the joint model. Randomisation to treatment occurs at $t = 0$.

Assumption of Constant Hazard

For a chronic disease such as epilepsy, it may not be reasonable to assume that there is a constant underlying event rate. Clinical opinion suggests that seizures may be clustered, due to positive duration dependence. For further details of this idea, see the discussion of *true contagion* on page 7. In figure 9.2, an illustration of true contagion is given. At each event, the hazard instantaneously increases, and takes some time to settle back down to the underlying baseline hazard.

Clinical opinion also suggests that in some cases of epilepsy, particularly in untreated epilepsies, the condition deteriorates as each event occurs. Two suggestions are immediately apparent from this observation: the underlying hazard may be increasing rather than constant; or alternatively, the underlying hazard may be constant between events, and 'jump' at each event. It is possible that both apply, and this is shown in figure 9.3.

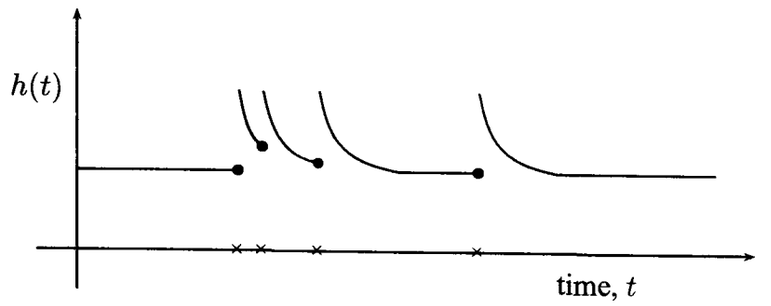

Figure 9.2: Illustration of the effect of true contagion (positive duration dependence) on the hazard. Observed event times are marked by filled circles, and crosses on the x-axis.

In some cases, particularly in individuals who are responding well to treatment, it may be the case that the underlying event rate is decreasing. It would not make clinical sense for the rate to jump down at an event time, but the underlying rate may be decreasing between events. It may even be the case that the underlying rate is decreasing, while the occurrence of an event increases the rate.

One important result of the Poisson process assumption is that knowing the exact event times is unnecessary, it is sufficient to know only the event count over a given period. An example of a varying-rate model would be an underlying point process generated by a gamma renewal distribution, rather than the exponential renewal distribution which generates a Poisson process. However, with such a point process, the exact event times would be needed, because the length of time between the last pre-randomisation event and randomisation would have an effect on the event rate at randomisation. This in turn would affect the estimation of the likelihood of a first post-randomisation event at a given time.

In conclusion, without the exact event times, it is not possible to test the assumption of a constant underlying event rate. With such data, it would also be possible to consider the possibility of an increasing (or decreasing) event rate, or true con-

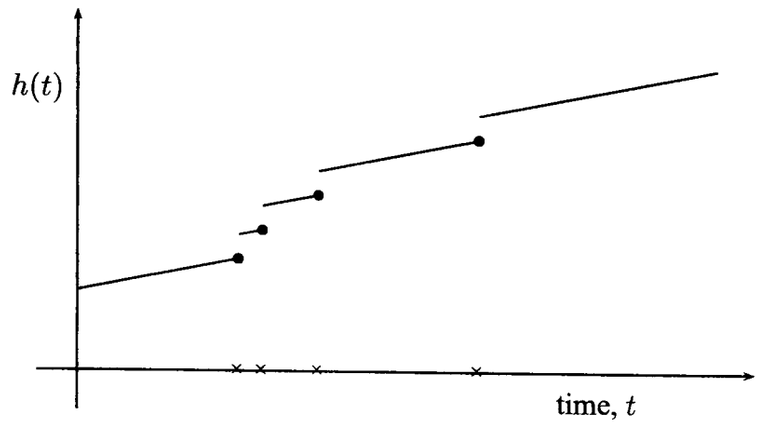

Figure 9.3: Illustration of an increasing underlying hazard for a given individual, with jumps at event times. Observed event times are marked by filled circles, and crosses on the x-axis.

tagion.

Assumption of an Instantaneous Multiplicative Treatment Effect

The joint model derived in chapter 4 also assumes that, for a given individual, the treatment reduces the event rate multiplicatively, with immediate effect. This is illustrated in figure 9.1 on page 161. However, this may not be reasonable for a number of reasons. For instance, in epilepsy, when individuals begin a new regime of an anti-epileptic drug at a certain dose, they will generally not be started at the full dosage, but rather take a lower dose for an initial period. In addition to this, it may take some time for a given dose to reach full effectiveness.

In figure 9.4, an example of a delayed treatment effect is given. Here, the treatment only reaches full effectiveness at time τ, either because the dosage is started low and gradually increased, or because the drug takes a while to reach full effectiveness, or due to a combination of these factors.

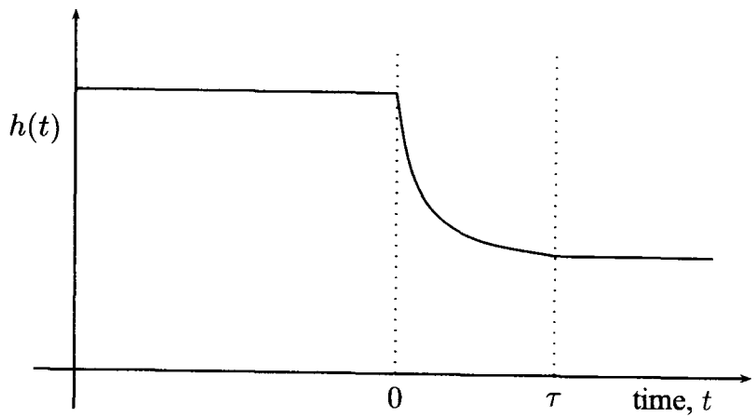

Figure 9.4: Individual hazard rate with delayed treatment effect. Randomisation to treatment occurs at $t = 0$.

In the epilepsy data described in chapters 3 and 5, nearly 25% of the individuals experience their first post-randomisation seizure within two weeks of randomisation, and thus the model of the impact of treatment on the underlying seizure rate is very important. If the time τ in figure 9.4 were 14 days, the way that the impact of treatment is modelled would be very important.

9.3.2 Joint Model at Population Level

The joint model derived in chapter 4 also makes some assumptions about the impact of covariates. In the joint model, allowance is made for individuals in the population to have differing underlying event rates. Some of the difference may be explained by the covariates such as *age*, *sex* and *epilepsy type*, but some unexplained difference remains. The joint model specifies that this difference follows a gamma distribution, which is a common assumption in the analysis of count data (Cameron & Trivedi, 1998), and leads to a marginal negative binomial distribution. In chapter 7 an alternative mixture distribution was described. A regres-

sion model for count data was derived, and fitted to the epilepsy data. The results suggested that a 2-parameter inverse-Gaussian mixture, or indeed a 3-parameter PVF mixture extra parameter, fitted the epilepsy data better than the 2-parameter gamma mixture, but none of the mixtures fitted the data very well.

In addition to the assumption of a particular parametric 'frailty', the joint model of chapter 4 also makes assumptions about the impact of covariates on the event rate. A log-link is used to relate the descriptive covariates to the underlying event rate. The log-link is attractive because it restricts the rate parameters to be positive, which is a necessary condition. There are alternative ways to incorporate covariate information, but they have not been investigated in this work.

Another assumption made by the joint model is on the impact of treatment on the underlying event rates. In the previous section, the assumption of an instantaneous treatment effect was discussed. At the population level, the joint model assumes that the treatment affects each individual event rate λ_i multiplicatively, inflating or reducing it by some proportion ψ, so that the new individual rate is $\lambda_i\psi$. This may not be a realistic model in some data, because perhaps the treatment will work 'better' on individuals with the most serious conditions (highest rates), or vice-versa. However, this problem may be reduced somewhat because the joint model does allow covariates, and treatment-covariate interactions, to have an effect in the estimation of the post-randomisation change in event rate.

9.4 Problems with the Epilepsy Data

In the course of the analyses in this thesis, various problems with the epilepsy data were encountered. The epilepsy data were first described in chapter 3, and the joint model is fitted to these data in chapter 5. The data also appeared in chapters 6, 7 and 8, where they were used to illustrate ideas and extensions to the joint model.

9.4.1 Pre-Randomisation Information

The data contain a 6-month seizure count for almost every individual – those individuals without this information have been excluded from all the analyses in this thesis. There are two major problems with the recording of this particular outcome.

Firstly, this particular outcome was not collected for all individuals. Some patients were asked for alternative information, such as the number of seizures in the preceding 3 months, or the total number of seizures they had ever experienced (along with the date of the first seizure). Some were only asked for the date of their most recent seizure. In every case, this information was used to interpolate or extrapolate a 6-month seizure count.

The models in this thesis allow for event counts over different periods, for different individuals, so it would not be a problem to use the original counts. Unfortunately the data is only available in its current form, as described in chapter 3.

Secondly, this data is subject to measurement error. The seizure counts were self-

reported, and so the patients could have forgotten about some of their seizures, or perhaps not even observed them – for example if they had nocturnal seizures, and slept alone. These information biases may affect some of the results, and further work could look at the sensitivity to measurement error of the models explored in this thesis.

9.4.2 Post-Randomisation Information

The major problem with the times to first post-randomisation seizure is with the assumptions about the drug effect. In figure 3.3 on page 23, it can be seen that about 25% of the first post-randomisation seizures occurred within the first two weeks of treatment. Since the individuals were not put on a full dose immediately, but rather given increasing doses over the first few weeks of treatment, it would be preferable to have a seizure count over a long post-randomisation period. It would be even better if exact seizure times were recorded, over a post-randomisation period.

9.4.3 Explanatory Variables

Some discussion of the possibility of misclassification of the covariate *epilepsy type* was given in section 5.5 on page 58. This problem has also been discussed by Williamson *et al.* (2002), for the epilepsy data.

In chapter 8, another problem was discussed, that is, the lack of information about *seizure type*. The analyses in section 8.4 suggested that *seizure type* would be an

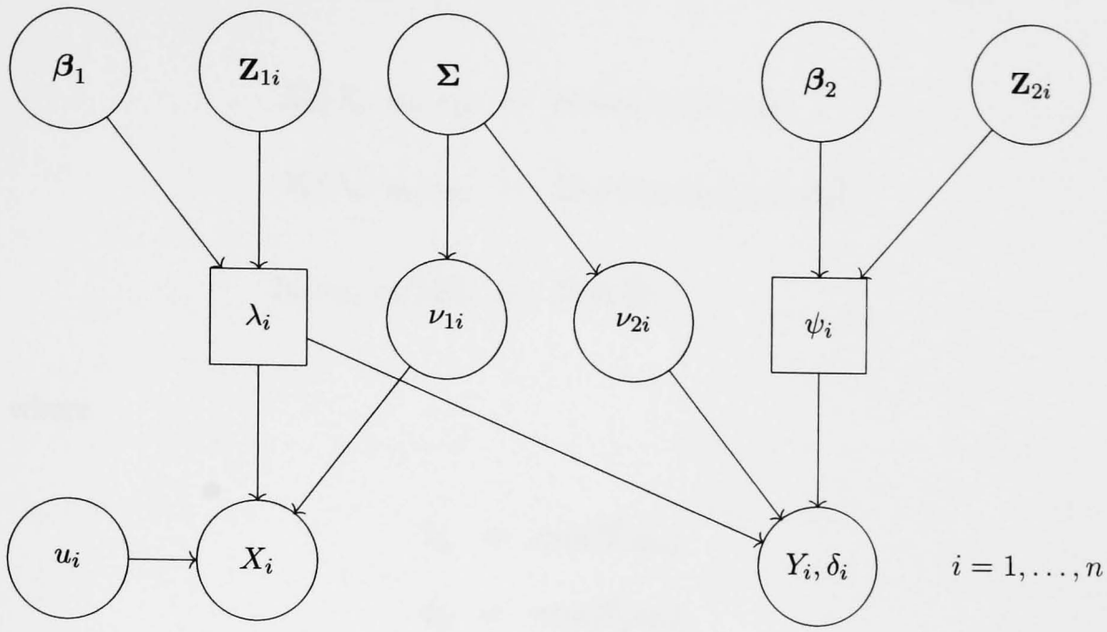

Figure 9.5: Graphical Model of the underlying point process

important covariate. It is recommended that in future studies of treatments for epilepsy, this information is collected.

9.5 A Dual-Outcome Model with a Bivariate Random Effect

An alternative to the joint model of chapter 4 is a model with a bivariate random effect. Correlation between the two random effects would induce dependence between the counts X_i and the times Y_i. An example of such a model is shown in figure 9.5. The circular nodes represent variables, either parameters or data, and

the square nodes are logical. The model is specified by the relationships:

$$\begin{aligned} X_i \,|\, \lambda_i,\, u_i,\, \nu_{1i} &\sim \text{Poisson}(\lambda_i u_i \nu_{1i}) \\ Y_i \,|\, \lambda_i,\, \psi_i,\, \nu_{2i} &\sim \text{Exponential}(\lambda_i \psi_i \nu_{2i}) \\ \ln(\nu_{1i},\, \nu_{2i} \,|\, \boldsymbol{\Sigma}) &\sim \text{N}(0, \boldsymbol{\Sigma}), \end{aligned}$$

where

$$\begin{aligned} \lambda_i &= \exp(\boldsymbol{\beta}_1' \mathbf{z}_{1i}), \\ \psi_i &= \exp(\boldsymbol{\beta}_2' \mathbf{z}_{2i}). \end{aligned}$$

Here, the two diagonal elements of $\boldsymbol{\Sigma}$ measure the degree of heterogeneity, and dependence between X_i and Y_i is induced if the off-diagonal elements are non-zero. The means of ν_{1i} and ν_{2i} are fixed as 1 for identifiability. The parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are vectors of regression coefficients, $\mathbf{z}_{1i}$ will include an intercept term and $\mathbf{z}_{2i}$ will generally be parameterised to include an average treatment effect as well as a treatment contrast, and may contain other explanatory variables and interaction terms. The use of log-links ensures that λ_i and ψ_i are always positive.

In the model presented in figure 9.5, the random effects ν_{1i} and ν_{2i} follow a bivariate log-Normal distribution. However, maximum likelihood estimation of such models would require numerically intensive methods. For most choices of distribution for $(\nu_{1i},\, \nu_{2i})$, a closed form expression will not exist for the joint or marginal distributions of X_i and Y_i.

9.6 Further Work

In this section, some extensions to the joint model of chapter 4 are discussed. In chapter 7, one possible area for extension was outlined, and initial explorations made. In that chapter, the power variance family (Hougaard, 1986b) was used in place of the gamma distribution, to model the unknown heterogeneity in the population. The power variance family was chosen because it incorporates the gamma distribution, as well as other common frailty distributions. However, only the univariate distribution of pre-randomisation counts was investigated in this way in chapter 7, and it remains as further work to develop the joint model using the power variance family as the frailty distribution. It is thought that a full joint model with this frailty distribution would be very computationally demanding. MCMC inference might be preferable to maximum likelihood estimation.

Another extension to the joint model is to allow a progressive treatment effect, as discussed in section 9.3.1, where an example is shown in figure 9.4 on page 164. To modify the joint model in this way in order to fit particular data, good clinical information would be needed on the form of the progressive treatment effect (that is, the shape of the curve shown in figure 9.4 on page 164). Without good clinical information, any assumption about a non-instantaneous treatment effect would need to be carefully tested.

The assumption of an underlying Poisson process may be clinically dubious in some data, but departing from this assumption would require much more detailed information, and any model would be very complex. Some discussion on this topic was given in section 9.3.1, including the suggestion of using a point process

generated by gamma-distributed waiting times (Winkelmann, 1995), rather than exponential-distributed waiting times as in the Poisson process.

In section 5.3 on page 52, some diagnostics for the joint model were presented, and applied to the joint model fitted to the epilepsy data. Further work could consider more diagnostic tests, as well as sensitivity analyses, and goodness-of-fit tests.

9.7 Summary

In conclusion, this thesis has presented a joint model for repeated event data consisting of a period count followed by randomisation to a treatment, and a recorded survival time. The joint model is based on a Poisson process with individual frailty, and assumes a multiplicative treatment effect. The joint model has been illustrated on epilepsy data. In addition, some discussion has been given to the relative efficiency of the joint model compared to typical survival models, and some extensions to the joint model have also been explored.

Appendix A

The Pareto Survival Model

Consider the Pareto survival model generated as an Exponential survival model with a gamma-distributed random effect acting multiplicatively on the rate. That is, each individual i has a survival time Y_i, and associated covariates $\mathbf{w}_i$, and the model is specified by the relationships

$$\begin{aligned} Y_i \,|\, \mu_i,\, \nu_i &\sim \text{Exponential}(\mu_i \nu_i), \\ \nu_i \,|\, \gamma &\sim \text{Gamma}(\gamma, \gamma) \end{aligned}$$

where

$$\mu_i = \exp(\boldsymbol{\theta}' \mathbf{w}_i).$$

Here ν_i is a random effect with mean 1 and variance $1/\gamma$, and $\boldsymbol{\theta}$ is a vector of regression coefficients. Integrating ν_i out, it is found that Y_i follows a Pareto(γ, μ_i)

distribution, with hazard, density and survivor functions specified by

$$\begin{aligned} h(y_i \,|\, \mu_i, \gamma) &= \frac{\gamma \mu_i}{\gamma + \mu_i y_i} \\ f(y_i \,|\, \mu_i, \gamma) &= \mu_i \left(\frac{\gamma}{\gamma + \mu_i y_i} \right)^{\gamma+1} \end{aligned} \tag{A.1}$$

$$S(y_i \,|\, \mu_i, \gamma) = \left(\frac{\gamma}{\gamma + \mu_i y_i} \right)^{\gamma}.$$

A.1 Likelihood and Derivatives

The log-likelihood is given by:

$$\ell_p(\boldsymbol{\theta}, \gamma \,|\mathcal{D}) = \sum_{i=1}^{n} \Big\{ \delta_i \ln(\mu_i) + (\gamma + \delta_i) \ln(\gamma) - (\gamma + \delta_i) \ln(\gamma + \mu_i y_i) \Big\}.$$

The first-order derivatives of ℓ_p are:

$$\frac{\partial \ell_p}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{n} \left\{ \delta_i \mathbf{w}_i - \frac{(\gamma + \delta_i)\mu_i y_i}{\gamma + \mu_i y_i} \mathbf{w}_i \right\} = \sum_{i=1}^{n} \left\{ \frac{\gamma(\delta_i - \mu_i y_i)}{\gamma + \mu_i y_i} \mathbf{w}_i \right\} \tag{A.2}$$

$$\frac{\partial \ell_p}{\partial \gamma} = \sum_{i=1}^{n} \left\{ \ln(\gamma) + 1 + \frac{\delta_i}{\gamma} - \ln(\gamma + \mu_i y_i) - \frac{\gamma + \delta_i}{\gamma + \mu_i y_i} \right\}. \tag{A.3}$$

The second-order derivatives of ℓ_p are:

$$\frac{\partial^2 \ell_p}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = -\sum_{i=1}^{n} \left(\frac{\gamma(\gamma + \delta_i)\mu_i y_i}{(\gamma + \mu_i y_i)^2} \right) \mathbf{w}_i \mathbf{w}_i' \tag{A.4}$$

$$\frac{\partial^2 \ell_p}{\partial \boldsymbol{\theta} \partial \gamma} = -\sum_{i=1}^{n} \left(\frac{(\mu_i y_i - \delta_i)\mu_i y_i}{(\gamma + \mu_i y_i)^2} \right) \mathbf{w}_i \tag{A.5}$$

$$\frac{\partial^2 \ell_p}{\partial \gamma \partial \gamma} = \sum_{i=1}^{n} \left\{ \frac{1}{\gamma} - \frac{\delta_i}{\gamma^2} - \frac{1}{\gamma + \mu_i y_i} - \frac{\mu_i y_i - \delta_i}{(\gamma + \mu_i y_i)^2} \right\}. \tag{A.6}$$

To fit a Pareto survival model to a set of survival times $\mathbf{y_i}$, a numerical method such as Newton-Raphson may be used.

Appendix B

Two Simulation Studies

In chapter 4, a joint model was presented for data which are a mixture of counts and times. This appendix gives further information on two large simulation studies comparing the joint model with a Pareto survival model. Some of the results were presented in section 6.2.1.

B.1 Methods

The simulation studies compares two models for data on n individuals, where for each individual there is:

- A pre-randomisation event count $X_i = x_i$, over a period of 182 days.
- A post-randomisation survival time $Y_i = y_i$, which may be censored, with an indicator of censoring δ_i so that $\delta_i = 1$ indicates an observed survival time, while $\delta_i = 0$ indicates a censored survival time.

- An indicator trt_i of the treatment given (individuals are randomly assigned to one of two treatments, and $trt_i = \pm 1$).
- A possibly informative covariate $sex_i = \pm 1$.

The first model fitted to these data is the joint model of chapter 4, specified by the equations:

$$\begin{aligned} f_X(x_i \mid \lambda_i, u_i, \nu_i) &= \frac{(\lambda_i u_i \nu_i)^{x_i} \exp(-\lambda_i u_i \nu_i)}{x_i!}, \\ f_Y(y_i \mid \lambda_i, \psi_i, \nu_i) &= \lambda_i \psi_i \nu_i \exp(-\lambda_i \psi_i \nu_i y_i), \\ g_\nu(\nu_i \mid \alpha) &= \frac{\alpha^\alpha \nu_i^{\alpha-1} \exp(-\alpha \nu_i)}{\Gamma(\alpha)}, \end{aligned}$$

with

$$\begin{aligned} \lambda_i &= \exp(\boldsymbol{\beta}_1' \mathbf{z}_{1i}), \\ \psi_i &= \exp(\boldsymbol{\beta}_2' \mathbf{z}_{2i}), \end{aligned}$$

where for each individual $\mathbf{z}_{1i} = (1, sex_i)$, and $\mathbf{z}_{2i} = (1, trt_i)'$. The regression parameters are α, $\boldsymbol{\beta}_1 = (\beta_0, \beta_{sex})'$ and $\boldsymbol{\beta}_2 = (\beta_{t0}, \beta_{trt})'$.

The second model fitted to these data is the Pareto survival model, described further in appendix A, specified by the equations

$$\begin{aligned} Y_i \mid \mu_i, \nu_i &\sim \text{Exponential}(\mu_i \nu_i) \\ \nu_i \mid \gamma &\sim \text{Gamma}(\gamma, \gamma), \end{aligned}$$

with

$$\mu_i = \exp(\boldsymbol{\theta}' \mathbf{w}_i).$$

where for each individual $\mathbf{w}_i = (1, sex_i, X_i, trt_i)'$. The regression parameters are γ and $\boldsymbol{\theta} = (\theta_0, \theta_{sex}, \theta_x, \theta_{trt})'$.

In the first study, data were generated according to the joint model specified above, with 'true' parameters from the sets:

$\alpha \in \{0.8, 1.2\}$,

$\beta_0 \in \{-3\}$,

$\beta_{sex} \in \{0.0, 0.4, 0.8\}$,

$\beta_{t0} \in \{-1, -2\}$,

$\beta_{trt} \in \{0.4, 0.8\}$.

Therefore the first study comprises 24 different parameter combinations. For each combination, 100 sets of data were generated, with 200 individuals in each set of data, 100 in each treatment arm.

The second study assumes the Pareto survival model is true, with underlying parameters from the set:

$\gamma \in \{2, 3\}$,

$\theta_0 \in \{-5\}$,

$\theta_{sex} \in \{0.0, 0.4, 0.8\}$,

$\theta_x \in \{0.05, 0.10\}$,

$\theta_{trt} \in \{0.4, 0.8\}$.

Additionally, the counts X_i were generated using the model

$$\begin{aligned} X_i \,|\, v_i &\sim \text{Poisson}(182 \exp(-3) v_i) \\ v_i &\sim \text{Gamma}(1, 1). \end{aligned}$$

The model for X_i may also be expressed as a negative binomial distribution with parameters 1 and $\exp(3)/183)$. Therefore the second study also comprises 24 different parameter combinations. For each combination, 100 sets of data were generated, with 200 individuals in each set of data, 100 in each treatment arm.

B.2 Implementation

The simulation was performed within `s-plus` 2000, on a 233MHz Pentium II PC, with 48MB RAM, running Windows NT4. For each parameter combination, to simulate 100 datasets, and fit both models, took around 5 hours. In total, each simulation study took about 120 hours of computational time.

B.3 Results

The results are presented in tables B.1 to B.16 at the end of this appendix. Each row in the table represents a different parameter combination, and describes the results for 100 simulated datasets with that particular combination.

For the first study, where data was simulated according to a joint model, the joint model estimates are given in tables B.1 to B.4, and the Pareto model estimates are given in tables B.5 to B.8. For the second study, where data was simulated according to a Pareto model, the joint model estimates are given in tables B.9 to B.12, and the Pareto model estimates are given in tables B.13 to B.16.

The average bias of the parameter estimate is given for the joint model estimates in study 1, and for the Pareto model estimates in study 2, and also for all estimates of the treatment effect, because this are expected to be similar under either model, whichever is the 'true' underlying model. For the non-treatment parameters under the model which is not the 'true' model, the mean of the 100 estimates is given. Also given for each parameter is the standard deviation of the 100 estimates, and the median of the 100 estimated standard errors. The average fitted log-likelihood is also presented, as well as the proportion of studies where the estimated standard error of the treatment effect under the joint model is lower than the corresponding e.s.e. under the Pareto model.

B.4 Discussion

In study 1, where the joint model is the 'true' underlying model, the joint model seems to perform fairly well. The results in tables B.1 to B.4 show that the estimates are mostly unbiased, although $\widehat{\alpha}$ does seem to be generally overestimated, and $\widehat{\beta}_0$ generally underestimated.

The Pareto model was also fitted to these data where the joint model is the 'true' model, and the estimates are given in tables B.5 to B.8. The unstable estimates of $\widehat{\gamma}$ are immediately noticeable, in the second and eighth rows of table B.5, where there is a small treatment effect, no sex effect, and a large treatment intercept β_{t0}. It is also of interest that the treatment effect β_{trt} is consistently underestimated by $\widehat{\theta}_{trt}$, although not to a great degree.

Finally, from the final columns of table B.2 and B.4, it can be seen that in 98% of the studies, the joint model gave a more precise estimate of the treatment effect than the Pareto model.

In study 2, where the Pareto model is the 'true' underlying model, the joint model also seems to perform fairly well. The estimates are presented in tables B.9 to B.12. The estimates of the treatment effect seem fairly unbiased, in that $\widehat{\beta}_{trt}$ is very close to the true θ_{trt} when $\theta_{trt} = 0.4$, and generally a slight underestimate when $\theta_{trt} = 0.8$.

The Pareto model estimates are presented in tables B.13 to B.16. Surprisingly, the model does not seem to perform very well. The model generally overestimates $\widehat{\gamma}$, and the estimates are particularly unstable in two of the combinations, in the eighth row of table B.13 and the ninth row of table B.15. The Pareto model also

seems to overestimate $\widehat{\theta}_0$, $\widehat{\theta}_{sex}$ and $\widehat{\theta}_x$. On the other hand, the estimates of $\widehat{\theta}_{trt}$ seem fairly unbiased.

Finally, from the final columns of table B.10 and B.12, it can be seen that again in 98% of the studies, the joint model gave a more precise estimate of the treatment effect than the Pareto model. This is very interesting, given that the 'true' underlying model was a Pareto.

Table B.1: Results of relative efficiency simulation study 1, assuming joint model to be true. Joint model estimates, part 1 (mild treatment effect)

α	β_0	β_{sex}	β_{t0}	β_{trt}	$\widehat{\alpha}$			$\widehat{\beta}_0$			$\widehat{\beta}_{sex}$		
					bias	s.d.	m(e.s.e.)	bias	s.d.	m(e.s.e.)	bias	s.d.	m(e.s.e.)
0.8	−3	0.0	−1	0.4	0.006	0.077	0.082	−0.005	0.087	0.083	0.012	0.086	0.082
0.8	−3	0.0	−2	0.4	0.011	0.092	0.086	−0.025	0.086	0.083	0.002	0.082	0.082
0.8	−3	0.4	−1	0.4	0.023	0.078	0.084	−0.006	0.082	0.082	0.003	0.089	0.082
0.8	−3	0.4	−2	0.4	0.005	0.087	0.085	−0.019	0.085	0.083	−0.007	0.078	0.083
0.8	−3	0.8	−1	0.4	0.026	0.085	0.086	−0.007	0.084	0.083	0.002	0.090	0.082
0.8	−3	0.8	−2	0.4	0.014	0.086	0.086	−0.005	0.070	0.083	−0.002	0.093	0.083
1.2	−3	0.0	−1	0.4	0.037	0.135	0.131	−0.011	0.075	0.069	−0.008	0.070	0.068
1.2	−3	0.0	−2	0.4	0.008	0.122	0.131	−0.003	0.076	0.069	−0.004	0.065	0.069
1.2	−3	0.4	−1	0.4	0.026	0.130	0.133	−0.001	0.071	0.068	−0.006	0.069	0.068
1.2	−3	0.4	−2	0.4	0.024	0.113	0.132	−0.017	0.071	0.069	−0.005	0.063	0.069
1.2	−3	0.8	−1	0.4	0.002	0.123	0.132	0.000	0.068	0.070	0.010	0.064	0.070
1.2	−3	0.8	−2	0.4	0.029	0.129	0.136	0.007	0.071	0.069	0.000	0.071	0.069

'bias' is mean($\widehat{p} - p$) for some parameter p; m(e.s.e.) is the median estimated standard error

Table B.2: Results of relative efficiency simulation study 1, assuming joint model to be true. Joint model estimates, part 1 (continued)

α	β_0	β_{sex}	β_{t0}	β_{trt}	$\widehat{\beta}_{t0}$			$\widehat{\beta}_{trt}$			mean(lik)	C
					bias	s.d.	m(e.s.e.)	bias	s.d.	m(e.s.e.)		
0.8	−3	0.0	−1	0.4	0.009	0.085	0.087	−0.003	0.091	0.086	−1593.7	1.00
0.8	−3	0.0	−2	0.4	0.014	0.088	0.091	0.008	0.092	0.090	−1579.3	0.97
0.8	−3	0.4	−1	0.4	−0.005	0.085	0.087	0.000	0.088	0.086	−1588.7	1.00
0.8	−3	0.4	−2	0.4	−0.004	0.097	0.091	−0.003	0.085	0.091	−1572.2	0.99
0.8	−3	0.8	−1	0.4	−0.001	0.082	0.088	−0.008	0.094	0.086	−1567.2	1.00
0.8	−3	0.8	−2	0.4	−0.012	0.095	0.091	−0.014	0.087	0.090	−1553.5	0.98
1.2	−3	0.0	−1	0.4	0.009	0.099	0.084	−0.011	0.081	0.083	−1629.9	0.95
1.2	−3	0.0	−2	0.4	−0.001	0.086	0.087	−0.023	0.087	0.086	−1662.5	0.82
1.2	−3	0.4	−1	0.4	0.001	0.089	0.084	0.000	0.086	0.083	−1623.8	0.96
1.2	−3	0.4	−2	0.4	−0.003	0.093	0.088	0.000	0.087	0.087	−1634.2	0.92
1.2	−3	0.8	−1	0.4	0.001	0.086	0.085	−0.009	0.089	0.084	−1603.1	1.00
1.2	−3	0.8	−2	0.4	−0.012	0.080	0.088	−0.003	0.086	0.087	−1602.1	0.97

'bias' is mean$(\widehat{p} - p)$ for some parameter p; m(e.s.e.) is the median estimated standard error

'lik' is fitted log-likelihood; C = the proportion of studies where $\{$ e.s.e.$(\widehat{\beta}_{trt}) <$ e.s.e.$(\widehat{\theta}_{trt})\}$

Table B.3: Results of relative efficiency simulation study 1, assuming joint model to be true. Joint model estimates, part 2 (strong treatment effect)

α	β_0	β_{sex}	β_{t0}	β_{trt}	$\widehat{\alpha}$			$\widehat{\beta}_0$			$\widehat{\beta}_{sex}$		
					bias	s.d.	m(e.s.e.)	bias	s.d.	m(e.s.e.)	bias	s.d.	m(e.s.e.)
0.8	−3	0.0	−1	0.8	0.006	0.081	0.082	0.000	0.077	0.083	0.006	0.080	0.083
0.8	−3	0.0	−2	0.8	0.006	0.088	0.084	−0.006	0.086	0.083	0.009	0.078	0.082
0.8	−3	0.4	−1	0.8	0.032	0.074	0.085	0.010	0.097	0.082	0.003	0.071	0.082
0.8	−3	0.4	−2	0.8	0.022	0.092	0.086	−0.008	0.072	0.082	0.008	0.083	0.082
0.8	−3	0.8	−1	0.8	0.017	0.090	0.085	−0.005	0.087	0.083	−0.005	0.086	0.083
0.8	−3	0.8	−2	0.8	0.009	0.094	0.085	−0.002	0.087	0.084	0.016	0.081	0.084
1.2	−3	0.0	−1	0.8	0.023	0.151	0.131	−0.014	0.076	0.068	0.001	0.066	0.068
1.2	−3	0.0	−2	0.8	0.034	0.146	0.133	−0.008	0.069	0.069	0.006	0.068	0.068
1.2	−3	0.4	−1	0.8	0.018	0.146	0.132	−0.012	0.063	0.069	−0.001	0.062	0.069
1.2	−3	0.4	−2	0.8	0.034	0.150	0.135	0.008	0.068	0.068	0.002	0.067	0.068
1.2	−3	0.8	−1	0.8	0.021	0.136	0.131	−0.015	0.072	0.070	0.010	0.070	0.070
1.2	−3	0.8	−2	0.8	0.002	0.135	0.132	0.068	0.175	0.070	−0.073	0.164	0.070

'bias' is mean($\widehat{p} - p$) for some parameter p; m(e.s.e.) is the median estimated standard error

Table B.4: Results of relative efficiency simulation study 1, assuming joint model to be true. Joint model estimates, part 2 (continued)

α	β_0	β_{sex}	β_{t0}	β_{trt}	$\widehat{\beta}_{t0}$			$\widehat{\beta}_{trt}$			mean(lik)	C
					bias	s.d.	m(e.s.e.)	bias	s.d.	m(e.s.e.)		
0.8	−3	0.0	−1	0.8	0.014	0.103	0.088	−0.003	0.099	0.087	−1567.0	1.00
0.8	−3	0.0	−2	0.8	0.007	0.085	0.092	0.008	0.094	0.091	−1553.2	0.99
0.8	−3	0.4	−1	0.8	0.008	0.090	0.087	0.005	0.082	0.086	−1568.7	1.00
0.8	−3	0.4	−2	0.8	0.004	0.103	0.093	0.010	0.087	0.092	−1540.1	0.99
0.8	−3	0.8	−1	0.8	−0.014	0.095	0.089	−0.006	0.083	0.088	−1543.1	1.00
0.8	−3	0.8	−2	0.8	0.003	0.099	0.093	−0.001	0.087	0.092	−1514.3	1.00
1.2	−3	0.0	−1	0.8	0.011	0.084	0.085	0.003	0.079	0.083	−1605.0	0.97
1.2	−3	0.0	−2	0.8	0.001	0.085	0.088	−0.004	0.070	0.088	−1615.4	0.98
1.2	−3	0.4	−1	0.8	0.016	0.087	0.085	0.002	0.074	0.084	−1599.1	0.99
1.2	−3	0.4	−2	0.8	0.003	0.093	0.088	−0.005	0.089	0.087	−1612.2	0.99
1.2	−3	0.8	−1	0.8	−0.009	0.089	0.085	−0.007	0.084	0.084	−1573.8	0.99
1.2	−3	0.8	−2	0.8	−0.008	0.085	0.089	−0.003	0.088	0.088	−1593.0	1.00

'bias' is mean($\widehat{p} - p$) for some parameter p; m(e.s.e.) is the median estimated standard error

'lik' is fitted log-likelihood; C = the proportion of studies where { e.s.e.($\widehat{\beta}_{trt}$) < e.s.e.($\widehat{\theta}_{trt}$) }

Table B.5: Results of relative efficiency simulation study 1, assuming joint model to be true. Pareto model estimates, part 1 (mild treatment effect)

α	β_0	β_{sex}	β_{t0}	β_{trt}	$\widehat{\gamma}$			$\widehat{\theta}_0$			$\widehat{\theta}_{sex}$		
					mean	s.d.	m(e.s.e.)	mean	s.d.	m(e.s.e.)	mean	s.d.	m(e.s.e.)
0.8	−3	0.0	−1	0.4	1.922	0.677	0.427	−5.240	0.188	0.159	−0.010	0.101	0.106
0.8	−3	0.0	−2	0.4	6.576	34.462	0.991	−6.245	0.178	0.150	0.004	0.091	0.104
0.8	−3	0.4	−1	0.4	1.701	0.483	0.382	−5.123	0.205	0.158	0.080	0.115	0.113
0.8	−3	0.4	−2	0.4	2.579	1.938	0.740	−6.165	0.192	0.154	0.060	0.129	0.111
0.8	−3	0.8	−1	0.4	1.527	0.474	0.318	−4.928	0.183	0.160	0.336	0.135	0.126
0.8	−3	0.8	−2	0.4	2.159	1.585	0.602	−5.935	0.194	0.152	0.339	0.133	0.122
1.2	−3	0.0	−1	0.4	3.615	1.977	1.036	−5.127	0.175	0.150	0.001	0.096	0.094
1.2	−3	0.0	−2	0.4	9.610	21.680	2.713	−6.129	0.179	0.146	−0.003	0.082	0.092
1.2	−3	0.4	−1	0.4	3.347	1.919	0.934	−5.029	0.189	0.150	0.112	0.105	0.102
1.2	−3	0.4	−2	0.4	5.348	6.727	1.848	−6.043	0.179	0.147	0.082	0.106	0.103
1.2	−3	0.8	−1	0.4	2.531	1.189	0.632	−4.835	0.170	0.151	0.383	0.141	0.116
1.2	−3	0.8	−2	0.4	3.918	2.754	1.441	−5.860	0.170	0.143	0.363	0.126	0.113

m(e.s.e.) is the median estimated standard error

Table B.6: Results of relative efficiency simulation study 1, assuming joint model to be true. Pareto model estimates, part 1 (continued)

α	β_0	β_{sex}	β_{t0}	β_{trt}	$\widehat{\theta}_x$			$\widehat{\theta}_{trt}$			mean(lik)
					mean	s.d.	m(e.s.e.)	bias	s.d.	m(e.s.e.)	
0.8	−3	0.0	−1	0.4	0.098	0.014	0.010	−0.014	0.112	0.106	−986.0
0.8	−3	0.0	−2	0.4	0.094	0.015	0.010	−0.012	0.114	0.104	−968.7
0.8	−3	0.4	−1	0.4	0.082	0.014	0.009	−0.022	0.110	0.107	−985.1
0.8	−3	0.4	−2	0.4	0.082	0.012	0.009	−0.017	0.103	0.107	−955.2
0.8	−3	0.8	−1	0.4	0.054	0.010	0.007	−0.016	0.113	0.112	−965.8
0.8	−3	0.8	−2	0.4	0.051	0.010	0.006	−0.042	0.103	0.111	−943.1
1.2	−3	0.0	−1	0.4	0.096	0.013	0.010	−0.021	0.090	0.094	−1009.3
1.2	−3	0.0	−2	0.4	0.090	0.012	0.009	−0.029	0.094	0.092	−1026.7
1.2	−3	0.4	−1	0.4	0.080	0.015	0.009	−0.014	0.101	0.095	−1002.4
1.2	−3	0.4	−2	0.4	0.079	0.013	0.009	−0.017	0.094	0.096	−1013.2
1.2	−3	0.8	−1	0.4	0.052	0.009	0.007	−0.012	0.103	0.100	−986.8
1.2	−3	0.8	−2	0.4	0.051	0.009	0.007	−0.014	0.095	0.098	−980.5

'bias' is mean($\widehat{p} - p$) for some parameter p; m(e.s.e.) is the median estimated standard error

'lik' is fitted log-likelihood

Table B.7: Results of relative efficiency simulation study 1, assuming joint model to be true. Pareto model estimates, part 2 (strong treatment effect)

α	β_0	β_{sex}	β_{t0}	β_{trt}	$\widehat{\gamma}$			$\widehat{\theta}_0$			$\widehat{\theta}_{sex}$		
					mean	s.d.	m(e.s.e.)	mean	s.d.	m(e.s.e.)	mean	s.d.	m(e.s.e.)
0.8	−3	0.0	−1	0.8	1.937	0.768	0.455	−5.233	0.196	0.158	0.021	0.111	0.106
0.8	−3	0.0	−2	0.8	2.638	2.247	0.775	−6.193	0.171	0.156	0.023	0.106	0.106
0.8	−3	0.4	−1	0.8	1.728	0.508	0.390	−5.069	0.195	0.157	0.081	0.103	0.113
0.8	−3	0.4	−2	0.8	2.183	1.731	0.579	−6.099	0.180	0.157	0.077	0.110	0.116
0.8	−3	0.8	−1	0.8	1.422	0.386	0.302	−4.914	0.191	0.163	0.336	0.135	0.128
0.8	−3	0.8	−2	0.8	1.844	0.667	0.501	−5.923	0.190	0.155	0.353	0.135	0.127
1.2	−3	0.0	−1	0.8	4.725	11.424	0.907	−5.112	0.181	0.149	0.016	0.096	0.095
1.2	−3	0.0	−2	0.8	4.311	2.541	1.894	−6.103	0.167	0.150	0.002	0.088	0.095
1.2	−3	0.4	−1	0.8	2.919	2.302	0.738	−4.988	0.185	0.150	0.103	0.099	0.104
1.2	−3	0.4	−2	0.8	4.511	6.347	1.306	−5.979	0.147	0.152	0.103	0.102	0.106
1.2	−3	0.8	−1	0.8	2.428	1.021	0.572	−4.834	0.152	0.150	0.403	0.123	0.117
1.2	−3	0.8	−2	0.8	3.672	2.515	1.228	−5.832	0.185	0.148	0.315	0.165	0.114

m(e.s.e.) is the median estimated standard error

Table B.8: Results of relative efficiency simulation study 1, assuming joint model to be true. Pareto model estimates, part 2 (continued)

α	β_0	β_{sex}	β_{t0}	β_{trt}	$\widehat{\theta}_x$			$\widehat{\theta}_{trt}$			mean(lik)
					mean	s.d.	m(e.s.e.)	bias	s.d.	m(e.s.e.)	
0.8	−3	0.0	−1	0.8	0.098	0.014	0.010	−0.032	0.110	0.106	−958.1
0.8	−3	0.0	−2	0.8	0.091	0.012	0.010	−0.017	0.115	0.109	−936.1
0.8	−3	0.4	−1	0.8	0.079	0.014	0.009	−0.015	0.098	0.108	−959.2
0.8	−3	0.4	−2	0.8	0.080	0.012	0.009	−0.023	0.100	0.111	−927.3
0.8	−3	0.8	−1	0.8	0.054	0.011	0.007	−0.014	0.114	0.113	−942.5
0.8	−3	0.8	−2	0.8	0.052	0.010	0.007	−0.013	0.112	0.114	−905.3
1.2	−3	0.0	−1	0.8	0.095	0.015	0.010	−0.012	0.089	0.095	−985.3
1.2	−3	0.0	−2	0.8	0.093	0.013	0.010	−0.028	0.082	0.097	−984.5
1.2	−3	0.4	−1	0.8	0.080	0.013	0.010	−0.022	0.096	0.098	−982.4
1.2	−3	0.4	−2	0.8	0.076	0.011	0.009	−0.029	0.104	0.100	−980.1
1.2	−3	0.8	−1	0.8	0.052	0.009	0.007	−0.020	0.105	0.101	−958.2
1.2	−3	0.8	−2	0.8	0.053	0.010	0.006	−0.030	0.093	0.102	−954.4

'bias' is mean($\widehat{p} - p$) for some parameter p; m(e.s.e.) is the median estimated standard error

'lik' is fitted log-likelihood

Table B.9: Results of relative efficiency simulation study 2, assuming Pareto model to be true. Joint model estimates, part 1 (mild treatment effect)

γ	θ_0	θ_{sex}	θ_x	θ_{trt}	$\widehat{\alpha}$			$\widehat{\beta}_0$			$\widehat{\beta}_{sex}$		
					mean	s.d.	m(e.s.e.)	mean	s.d.	m(e.s.e.)	mean	s.d.	m(e.s.e.)
2	−5	0.0	0.05	0.4	1.310	0.144	0.139	−3.006	0.070	0.066	0.008	0.069	0.066
2	−5	0.0	0.10	0.4	1.212	0.118	0.126	−3.006	0.081	0.069	−0.004	0.070	0.069
2	−5	0.4	0.05	0.4	1.322	0.143	0.139	−3.002	0.075	0.066	0.013	0.074	0.066
2	−5	0.4	0.10	0.4	1.210	0.134	0.125	−2.996	0.074	0.069	0.013	0.071	0.069
2	−5	0.8	0.05	0.4	1.334	0.124	0.141	−3.002	0.079	0.066	0.033	0.067	0.066
2	−5	0.8	0.10	0.4	1.205	0.114	0.126	−3.000	0.078	0.069	0.035	0.075	0.068
3	−5	0.0	0.05	0.4	1.336	0.143	0.143	−3.001	0.075	0.066	0.006	0.079	0.066
3	−5	0.0	0.10	0.4	1.198	0.129	0.124	−2.997	0.068	0.069	0.005	0.084	0.069
3	−5	0.4	0.05	0.4	1.328	0.117	0.142	−3.001	0.079	0.066	0.035	0.063	0.066
3	−5	0.4	0.10	0.4	1.187	0.125	0.124	−3.020	0.083	0.070	0.020	0.074	0.069
3	−5	0.8	0.05	0.4	1.336	0.125	0.141	−3.019	0.071	0.066	0.030	0.077	0.065
3	−5	0.8	0.10	0.4	1.200	0.110	0.126	−3.011	0.073	0.069	0.037	0.080	0.069

'bias' is mean($\widehat{p} - p$) for some parameter p; m(e.s.e.) is the median estimated standard error

Table B.10: Results of relative efficiency simulation study 2, assuming Pareto model to be true. Joint model estimates, part 1 (continued)

γ	θ_0	θ_{sex}	θ_x	θ_{trt}	$\widehat{\beta}_{t0}$			$\widehat{\beta}_{trt}$			mean(lik)	C
					mean	s.d.	m(e.s.e.)	bias	s.d.	m(e.s.e.)		
2	−5	0.0	0.05	0.4	−0.663	0.135	0.082	−0.010	0.119	0.081	−1664.1	0.99
2	−5	0.0	0.10	0.4	−0.098	0.119	0.084	−0.016	0.128	0.083	−1568.8	1.00
2	−5	0.4	0.05	0.4	−0.444	0.132	0.082	−0.006	0.129	0.081	−1627.7	1.00
2	−5	0.4	0.10	0.4	0.094	0.133	0.084	−0.011	0.101	0.083	−1536.8	1.00
2	−5	0.8	0.05	0.4	−0.309	0.142	0.082	0.002	0.135	0.082	−1605.2	1.00
2	−5	0.8	0.10	0.4	0.229	0.129	0.084	−0.002	0.117	0.083	−1514.7	0.99
3	−5	0.0	0.05	0.4	−0.500	0.108	0.080	−0.006	0.100	0.080	−1640.4	0.98
3	−5	0.0	0.10	0.4	−0.950	0.108	0.084	−0.007	0.106	0.083	−1682.9	0.93
3	−5	0.4	0.05	0.4	−0.310	0.112	0.081	−0.002	0.111	0.080	−1606.1	1.00
3	−5	0.4	0.10	0.4	−0.738	0.108	0.084	0.005	0.110	0.083	−1662.8	0.91
3	−5	0.8	0.05	0.4	−0.154	0.126	0.081	−0.026	0.123	0.080	−1578.6	0.99
3	−5	0.8	0.10	0.4	−0.581	0.107	0.084	−0.004	0.114	0.083	−1641.0	0.92

'bias' is mean($\widehat{p} - p$) for some parameter p; m(e.s.e.) is the median estimated standard error

'lik' is fitted log-likelihood; C = the proportion of studies where $\{$ e.s.e.$(\widehat{\beta}_{trt}) <$ e.s.e.$(\widehat{\theta}_{trt})\}$

Table B.11: Results of relative efficiency simulation study 2, assuming Pareto model to be true. Joint model estimates, part 2 (strong treatment effect)

γ	θ_0	θ_{sex}	θ_x	θ_{trt}	$\widehat{\alpha}$			$\widehat{\beta}_0$			$\widehat{\beta}_{sex}$		
					mean	s.d.	m(e.s.e.)	mean	s.d.	m(e.s.e.)	mean	s.d.	m(e.s.e.)
2	−5	0.0	0.05	0.8	1.322	0.143	0.141	−3.012	0.064	0.066	−0.001	0.075	0.066
2	−5	0.0	0.10	0.8	1.198	0.116	0.124	−2.999	0.084	0.069	−0.008	0.075	0.069
2	−5	0.4	0.05	0.8	1.314	0.116	0.142	−3.010	0.075	0.066	0.013	0.072	0.066
2	−5	0.4	0.10	0.8	1.198	0.109	0.125	−3.009	0.075	0.069	0.014	0.079	0.069
2	−5	0.8	0.05	0.8	1.317	0.139	0.140	−2.995	0.077	0.066	0.031	0.073	0.066
2	−5	0.8	0.10	0.8	1.201	0.114	0.124	−3.009	0.078	0.069	0.034	0.080	0.069
3	−5	0.0	0.05	0.8	1.327	0.120	0.141	−3.013	0.077	0.066	0.000	0.078	0.066
3	−5	0.0	0.10	0.8	1.192	0.121	0.126	−3.014	0.073	0.069	−0.008	0.063	0.069
3	−5	0.4	0.05	0.8	1.338	0.148	0.144	−2.999	0.068	0.066	0.019	0.069	0.065
3	−5	0.4	0.10	0.8	1.199	0.122	0.125	−3.008	0.073	0.069	0.016	0.074	0.069
3	−5	0.8	0.05	0.8	1.344	0.142	0.145	−3.019	0.078	0.066	0.031	0.085	0.065
3	−5	0.8	0.10	0.8	1.214	0.122	0.126	−3.006	0.072	0.069	0.028	0.067	0.069

'bias' is mean($\widehat{p} - p$) for some parameter p; m(e.s.e.) is the median estimated standard error

Table B.12: Results of relative efficiency simulation study 2, assuming Pareto model to be true. Joint model estimates, part 2 (continued)

γ	θ_0	θ_{sex}	θ_x	θ_{trt}	$\widehat{\beta}_{t0}$			$\widehat{\beta}_{trt}$			mean(lik)	C
					mean	s.d.	m(e.s.e.)	bias	s.d.	m(e.s.e.)		
2	−5	0.0	0.05	0.8	−0.659	0.132	0.082	−0.012	0.125	0.081	−1647.7	1.00
2	−5	0.0	0.10	0.8	−0.110	0.135	0.084	−0.031	0.117	0.083	−1556.9	1.00
2	−5	0.4	0.05	0.8	−0.462	0.138	0.082	−0.026	0.119	0.081	−1621.5	1.00
2	−5	0.4	0.10	0.8	0.052	0.126	0.084	−0.034	0.117	0.083	−1535.3	1.00
2	−5	0.8	0.05	0.8	−0.333	0.156	0.082	−0.008	0.121	0.081	−1601.4	1.00
2	−5	0.8	0.10	0.8	0.227	0.126	0.084	−0.032	0.106	0.083	−1507.1	0.99
3	−5	0.0	0.05	0.8	−0.497	0.109	0.081	−0.002	0.117	0.080	−1626.1	0.97
3	−5	0.0	0.10	0.8	−0.937	0.121	0.084	0.001	0.093	0.083	−1652.1	0.98
3	−5	0.4	0.05	0.8	−0.325	0.108	0.081	−0.005	0.104	0.080	−1598.4	0.96
3	−5	0.4	0.10	0.8	−0.756	0.118	0.084	0.012	0.105	0.083	−1638.1	0.96
3	−5	0.8	0.05	0.8	−0.152	0.117	0.081	−0.007	0.111	0.080	−1570.7	0.97
3	−5	0.8	0.10	0.8	−0.596	0.121	0.084	−0.012	0.104	0.083	−1624.4	0.95

'bias' is mean($\widehat{p} - p$) for some parameter p; m(e.s.e.) is the median estimated standard error

'lik' is fitted log-likelihood; C = the proportion of studies where $\{$ e.s.e.$(\widehat{\beta}_{trt}) <$ e.s.e.$(\widehat{\theta}_{trt})\}$

Table B.13: Results of relative efficiency simulation study 2, assuming Pareto model to be true. Pareto model estimates, part 1 (mild treatment effect)

γ	θ_0	θ_{sex}	θ_x	θ_{trt}	$\widehat{\gamma}$			$\widehat{\theta}_0$			$\widehat{\theta}_{sex}$		
					bias	s.d.	m(e.s.e.)	bias	s.d.	m(e.s.e.)	bias	s.d.	m(e.s.e.)
2	−5	0.0	0.05	0.4	0.297	1.242	0.483	0.995	0.151	0.157	0.017	0.096	0.100
2	−5	0.0	0.10	0.4	0.397	0.678	0.557	1.046	0.144	0.154	−0.017	0.093	0.098
2	−5	0.4	0.05	0.4	0.247	0.550	0.507	1.224	0.164	0.157	−0.040	0.102	0.100
2	−5	0.4	0.10	0.4	0.443	0.816	0.526	1.279	0.161	0.154	−0.017	0.100	0.099
2	−5	0.8	0.05	0.4	0.257	0.650	0.470	1.388	0.172	0.155	−0.058	0.086	0.100
2	−5	0.8	0.10	0.4	0.636	1.381	0.553	1.454	0.167	0.154	−0.035	0.110	0.098
3	−5	0.0	0.05	0.4	0.760	1.835	1.107	1.008	0.132	0.147	−0.019	0.097	0.091
3	−5	0.0	0.10	0.4	4.575	30.849	1.087	0.028	0.165	0.149	0.017	0.093	0.092
3	−5	0.4	0.05	0.4	0.594	1.342	1.036	1.217	0.146	0.147	0.014	0.099	0.091
3	−5	0.4	0.10	0.4	1.296	2.413	1.301	0.228	0.148	0.144	0.000	0.079	0.090
3	−5	0.8	0.05	0.4	0.849	1.554	1.041	1.377	0.134	0.144	0.034	0.093	0.091
3	−5	0.8	0.10	0.4	0.931	2.083	1.101	0.431	0.128	0.147	−0.010	0.093	0.092

m(e.s.e.) is the median estimated standard error

Table B.14: Results of relative efficiency simulation study 2, assuming Pareto model to be true. Pareto model estimates, part 1 (continued)

γ	θ_0	θ_{sex}	θ_x	θ_{trt}	$\widehat{\theta}_x$			$\widehat{\theta}_{trt}$			mean(lik)
					bias	s.d.	m(e.s.e.)	bias	s.d.	m(e.s.e.)	
2	−5	0.0	0.05	0.4	−0.001	0.011	0.010	0.002	0.105	0.100	−970.6
2	−5	0.0	0.10	0.4	−0.011	0.010	0.010	−0.018	0.101	0.098	−893.8
2	−5	0.4	0.05	0.4	−0.002	0.010	0.010	−0.014	0.103	0.100	−933.2
2	−5	0.4	0.10	0.4	−0.015	0.012	0.010	−0.031	0.081	0.098	−857.3
2	−5	0.8	0.05	0.4	0.000	0.011	0.010	0.014	0.090	0.099	−901.3
2	−5	0.8	0.10	0.4	−0.018	0.011	0.010	−0.006	0.098	0.097	−828.2
3	−5	0.0	0.05	0.4	−0.002	0.010	0.010	0.001	0.083	0.091	−953.5
3	−5	0.0	0.10	0.4	−0.006	0.010	0.009	−0.007	0.095	0.092	−1011.0
3	−5	0.4	0.05	0.4	−0.002	0.009	0.010	0.011	0.090	0.090	−915.4
3	−5	0.4	0.10	0.4	−0.007	0.010	0.009	0.003	0.089	0.090	−990.9
3	−5	0.8	0.05	0.4	−0.002	0.009	0.010	−0.020	0.097	0.091	−881.7
3	−5	0.8	0.10	0.4	−0.007	0.009	0.010	−0.007	0.090	0.091	−959.7

'bias' is mean($\widehat{p} - p$) for some parameter p; m(e.s.e.) is the median estimated standard error

'lik' is fitted log-likelihood

Table B.15: Results of relative efficiency simulation study 2, assuming Pareto model to be true. Pareto model estimates, part 2 (strong treatment effect)

γ	θ_0	θ_{sex}	θ_x	θ_{trt}	$\widehat{\gamma}$			$\widehat{\theta}_0$			$\widehat{\theta}_{sex}$		
					bias	s.d.	m(e.s.e.)	bias	s.d.	m(e.s.e.)	bias	s.d.	m(e.s.e.)
2	−5	0.0	0.05	0.8	0.360	0.799	0.561	0.983	0.175	0.155	−0.036	0.112	0.099
2	−5	0.0	0.10	0.8	0.387	0.720	0.524	1.069	0.162	0.155	0.006	0.097	0.098
2	−5	0.4	0.05	0.8	0.259	0.557	0.532	1.197	0.157	0.155	−0.005	0.098	0.099
2	−5	0.4	0.10	0.8	0.249	0.541	0.505	1.264	0.136	0.154	−0.021	0.102	0.099
2	−5	0.8	0.05	0.8	0.344	0.717	0.525	1.400	0.157	0.156	0.003	0.093	0.099
2	−5	0.8	0.10	0.8	0.599	1.176	0.551	1.443	0.157	0.152	−0.076	0.098	0.097
3	−5	0.0	0.05	0.8	1.282	2.714	1.207	0.973	0.145	0.144	0.029	0.095	0.090
3	−5	0.0	0.10	0.8	0.714	2.363	1.083	0.055	0.150	0.147	−0.004	0.092	0.093
3	−5	0.4	0.05	0.8	3.480	25.046	1.069	1.184	0.165	0.147	0.025	0.087	0.091
3	−5	0.4	0.10	0.8	1.119	1.891	1.281	0.205	0.153	0.147	0.030	0.083	0.091
3	−5	0.8	0.05	0.8	1.048	1.916	1.153	1.370	0.145	0.145	−0.035	0.099	0.091
3	−5	0.8	0.10	0.8	1.228	2.742	1.317	0.437	0.170	0.146	−0.030	0.096	0.090

m(e.s.e.) is the median estimated standard error

Table B.16: Results of relative efficiency simulation study 2, assuming Pareto model to be true. Pareto model estimates, part 2 (continued)

γ	θ_0	θ_{sex}	θ_x	θ_{trt}	$\widehat{\theta}_x$			$\widehat{\theta}_{trt}$			mean(lik)
					bias	s.d.	m(e.s.e.)	bias	s.d.	m(e.s.e.)	
2	−5	0.0	0.05	0.8	−0.002	0.011	0.010	0.009	0.102	0.098	−956.6
2	−5	0.0	0.10	0.8	−0.015	0.009	0.010	−0.036	0.094	0.099	−880.1
2	−5	0.4	0.05	0.8	−0.002	0.010	0.010	−0.019	0.100	0.099	−927.4
2	−5	0.4	0.10	0.8	−0.016	0.009	0.010	−0.046	0.088	0.099	−857.4
2	−5	0.8	0.05	0.8	−0.005	0.010	0.010	−0.005	0.096	0.099	−894.5
2	−5	0.8	0.10	0.8	−0.019	0.010	0.010	−0.049	0.089	0.097	−825.3
3	−5	0.0	0.05	0.8	−0.002	0.010	0.009	−0.004	0.096	0.089	−942.0
3	−5	0.0	0.10	0.8	−0.008	0.009	0.010	0.001	0.088	0.093	−983.2
3	−5	0.4	0.05	0.8	−0.001	0.010	0.010	0.000	0.092	0.091	−908.2
3	−5	0.4	0.10	0.8	−0.006	0.010	0.009	0.006	0.085	0.091	−964.4
3	−5	0.8	0.05	0.8	−0.003	0.010	0.010	−0.008	0.084	0.090	−876.2
3	−5	0.8	0.10	0.8	−0.010	0.011	0.009	−0.019	0.093	0.090	−942.9

'bias' is mean($\widehat{p} - p$) for some parameter p; m(e.s.e.) is the median estimated standard error

'lik' is fitted log-likelihood

Appendix C

S-Plus Code for Joint Model

The function `joint1` may be used to find maximum likelihood estimates for the joint model fitted to a set of data. The function `joint1` makes use of a second function `joint2`, in each Newton-Raphson iteration.

```
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:
#  "count" = pre-randomisation seizure count;
#  "time" = post-randomisation time to first seizure;
#  "cens" = censoring indicator, 1 indicating observed time;
#  "age" as a continuous covariate, transformed if necessary;
#  "type" 0/1 indicating epilepsy type;
#  "sex"  0/1 indicating sex;
#  "trt"  0/1 indicating treatment.
```

```
#
#  When calling the function, "incl1" and "incl2" decide which
#  covariates to include in lambda and psi respectively, and the
#  initial vectors for beta1 and beta2 should be of the correct
#  length.  The code for "incl1" and "incl2" is:
#
#  incl1 = 1  =>  only include intercept term in lambda
#  incl1 = 2  =>  only include intercept and "type" in lambda
#  incl1 = 3  =>  only include intercept and "age" in lambda
#  incl1 = 4  =>  include intercept, "type" and "age" in lambda
#  incl1 = 5  =>  include intercept, "type", "age" and "sex"
#  incl1 = 6  =>  include intercept, "type", "age" and "trial"
#  incl1 = 7  =>  intercept, "type", "age", "sex" and 4 "trial"s
#  incl1 = 8  =>  intercept, "type", "age" and 5 "trial"s
#  incl1 = 9  =>  intercept, "type", "age", "sex" and 5 "trial"s
#
#  incl2 = 1  =>  only include intercept and trt contrast in psi
#  incl2 = 2  =>  intercept, trt contrast and type2 and trt*type
#  incl2 = 3  =>  intercept, trt contrast and age2 and trt*age
#  incl2 = 4  =>  intercept, trt contrast and both type and age
#              interactions in psi (type2,trt*type,age2,trt*age)
#
#
#
# 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$covmat)),
        digits = 3), covmat = newcov$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$type)
    beta1 <- matrix(beta1vec, nrow = k1)
    beta2 <- matrix(beta2vec, nrow = k2)
    modage <- data$age
    newtype <- data$type            # to make type -1/+1 not 0/1
    newsex <- 2 * data$sex - 1      # to make sex -1/+1
```

```
    trt <- 2 * data$trt - 1          # to make trt -1/+1
    trttype <- trt * newtype         # the trt/type interaction
    count <- data$count
    time <- data$time
    cens <- data$cens    #
#
#
# 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), newtype), byrow = T, nrow = k1)
    if(incl1 == 3)
        z1 <- matrix(c(rep(1, n), modage), byrow = T, nrow = k1)
    if(incl1 == 4)
        z1 <- matrix(c(rep(1, n), newtype, modage), byrow = T,
            nrow = k1)
    if(incl1 == 5)
        z1 <- matrix(c(rep(1, n), newtype, modage, newsex),
            byrow = T, nrow = k1)
    if(incl1 == 6)
        z1 <- matrix(c(rep(1, n), newtype, modage, data$trial2,
            data$trial3, data$trial4), byrow = T, nrow = k1)
    if(incl1 == 7)
        z1 <- matrix(c(rep(1, n), newtype, modage, newsex, data$
            trial2, data$trial3, data$trial4), byrow = T,
```

```
            nrow = k1)
    if(incl1 == 8)
        z1 <- matrix(c(rep(1, n), newtype, modage, data$trial2,
            data$trial3, data$trial4, data$trial5), byrow
             = T, nrow = k1)
    if(incl1 == 9)
        z1 <- matrix(c(rep(1, n), newtype, modage, newsex, data$
            trial2, data$trial3, data$trial4, data$trial5),
            byrow = T, nrow = k1)
#
    if(incl2 == 1)
        z2 <- matrix(c(rep(1, n), trt), byrow = T, nrow = k2)
    if(incl2 == 2)
        z2 <- matrix(c(rep(1, n), trt, newtype, trttype),
            byrow = T, nrow = k2)
    if(incl2 == 3)
        z2 <- matrix(c(rep(1, n), trt, modage, modage *
            data$trt), byrow = T, nrow = k2)
    if(incl2 == 4)
        z2 <- matrix(c(rep(1, n), trt, newtype, trttype,
            modage, modage * trt), 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) #lambda is the individual rate
    psi <- exp(t(beta2) %*% z2)    #psi is the treatment effect
#
    bit1 <- 182 + psi * time
    bit2 <- lambda * bit1 + alpha  #these two "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 * (count[i] + alpha + cens[i]) *
            lambda[i] * bit1[i])/(bit2[i] * bit2[i])) * outer(
            z1[, i], z1[, i])
        mat2 <- mat2 - (((count[i] + alpha + cens[i]) * (182 *
            lambda[i] + alpha) * lambda[i] * psi[i] * time[i])/(
            bit2[i] * bit2[i])) * outer(z2[, i], z2[, i])
        mat12 <- mat12 - ((alpha * (count[i] + alpha + cens[i])
            * lambda[i] * psi[i] * time[i])/(bit2[i] * bit2[i]))
            * outer(z1[, i], z2[, i])
#
        term1 <- term1 + ((alpha * (count[i] + cens[i] -
            lambda[i] * bit1[i]))/bit2[i]) * z1[, i]
        term2 <- term2 + ((cens[i] * (182 * lambda[i] + alpha) -
            (count[i] + alpha) * lambda[i] * psi[i] * time[i])/
            bit2[i]) * z2[, i]
#
# for gradient and Hessian contributions for alpha
#
        if(count[i] > 0) {
            for(j in 0:(count[i] - 1)) {
                aterm3[i] <- aterm3[i] + 1/((alpha + j)^
                  2)
                aterm4[i] <- aterm4[i] + 1/(alpha + j)
```

```
            }
        }
        aterm1[i] <- aterm3[i] + cens[i]/((alpha + count[i]) * (
            alpha + count[i])) - 1/alpha - (count[i] + cens[
            i] - alpha - 2 * lambda[i] * bit1[i])/(bit2[i] *
            bit2[i])
        aterm2[i] <- aterm4[i] + cens[i]/(alpha + count[i]) +
            log(alpha) + 1 - log(bit2[i]) - (count[i] +
            alpha + cens[i])/bit2[i]
#
# for Hessian contribution of correlation between alpha and beta
#
        mat1a <- mat1a + ((lambda[i] * bit1[i] * (count[i] +
            cens[i] - lambda[i] * bit1[i]))/(bit2[i] * bit2[i]))
            * z1[, i]
        mat2a <- mat2a + (((count[i] + cens[i] - lambda[i] *
            bit1[i]) * lambda[i] * psi[i] * time[i])/(bit2[i] *
            bit2[i])) * z2[, i] #
#
# for log-likelihood contribution of individual i
#
        if(count[i] == 0)
            llterm1[i] <- 0
        else {
            for(j in 0:(count[i] - 1)) {
                llterm1[i] <- llterm1[i] + log(alpha +
                  j)
            }
        }
        ll[i] <- llterm1[i] + cens[i] * log(alpha + count[i]) +
```

```
            alpha * log(alpha) + (count[i] + cens[i]) * log(
            lambda[i]) + count[i] * log(182) + cens[i] * log(
            psi[i]) - lgamma(count[i] + 1) - (count[i] +
            alpha + cens[i]) * log(bit2[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))
}
```

To illustrate the usage of these functions, the commands used to generate the maximum likelihood estimates for 'model 1' and 'model 2' in table 5.1 on page 49 were, respectively:

```
joint1(1,c(-3,0,0,0,0,0,0), c(0,0), incl1=8, incl2=1,
data=epilepsy)
```

```
joint1(1,c(-3,0,0,0,0,0,0), c(0,0,0,0,0,0), incl1=8, incl2=4,
data=epilepsy)
```

Appendix D

WinBUGS Code for Joint Model

In this appendix, some code for the software WinBUGS (Spiegelhalter *et al.*, 2000) is presented. This program is very user-friendly, and it is easy to encode a graphical model, in just a few lines, to allow MCMC inference on a set of data. WinBUGS code for standard survival models is given in Congdon (Congdon, 2001).

D.1 Joint Model

In this section, the code is presented for the joint model of chapter 4. The joint model is specified by figure 4.2 on page 34.

In this particular model, the pre-randomisation covariates are *type* and *age*, and *trt* contains the treatment information. Other variables in the data are *survtime*, the observed time to first post-randomisation seizure, and *count*, the 6-month pre-

randomisation seizure count.

The parameters of the Bayesian model are: ALPHA, measuring the overdispersion; BETA0 as the intercept term in λ_i; BETAage and BETAtype as the regression coefficients corresponding to the covariates *age* and *type*; BETAt0 as the intercept term in ψ_i; and BETAtrt as the regression coefficient measuring the contrast between treatments.

This model was used to produce the MCMC results presented in table 8.3 on page 134. Notice that a vague gamma prior is used for ALPHA, while vague normal priors are used for the BETA parameters.

The code for the model is:

```
# MODEL
{
   ALPHA ~ dgamma( 1.0E-4 , 1.0E-4);
   BETA0 ~ dnorm( -3 , 0.001 );
   BETAage ~ dnorm( 0 , 0.001 );
   BETAtype ~ dnorm( 0 , 0.001 );
   BETAt0 ~ dnorm( -1 , 0.001 );
   BETAtrt ~ dnorm( 0 , 0.001 );

   N <- 450
   u <- 182

   for( i in 1:N ){ # make type -1/+1 instead of 0/1
                    # and modify the age variable
```

```
        newtype[i] <- 2*type[i] - 1 ;
        modage[i] <- ( age[i] - 20 ) / 10 ;
    }

    for( i in 1:N ) {
        betaz1[i] <- BETA0 + BETAage * modage[i]
          + BETAtype * newtype[i] ;
        betaz2[i] <- BETAt0 + BETAtrt * trt[i] ;

        nu[i] ~ dgamma( EACHALPHA[i] , EACHALPHA[i] ) ;
        EACHALPHA[i] <- ALPHA ;

        lambdau[i] <- u * exp( betaz1[i] ) * nu[i] ;
        lambdapsi[i] <- exp( betaz1[i] ) * exp( betaz2[i] )
          * nu[i] ;

        count[i] ~ dpois( lambdau[i] ) ;
        survtime[i] ~ dexp( lambdapsi[i] ) ;
    }
}
```

D.2 Joint Model Regressing on α

In this section the code for a modified joint model, including regression coefficients in the overdispersion parameter α, is presented. This model was presented in section 8.1 on page 128. The model is specified by figure 8.1 on page 129.

Only a few small modifications are made to the code presented in the previous section. The parameter `XI0` is the intercept term in α_i, and `XItype` measures the difference in overdispersion between the two epilepsy types.

This model was used to produce the MCMC results presented in table 8.4 on page 138. Notice that vague normal priors are used for the `XI` and `BETA` parameters.

The code for the model is:

```
# MODEL
{
   XI0 ~ dnorm( 0 , 0.001 );
   XItype ~ dnorm( 0 , 0.001 );
   BETA0 ~ dnorm( -3 , 0.001 );
   BETAage ~ dnorm( 0 , 0.001 );
   BETAtype ~ dnorm( 0 , 0.001 );
   BETAt0 ~ dnorm( -1 , 0.001 );
   BETAtrt ~ dnorm( 0 , 0.001 );

   N <-450
   u <-182

   for( i in 1:N ){ # make type -1/+1 instead of 0/1
                    # and modify the age variable
      newtype[i] <- 2 * type[i] - 1 ;
      modage[i] <- ( age[i] - 20 ) / 10 ;
   }
```

```
  for( i in 1:N ) {
     betaz1[i] <- BETA0 + BETAage * modage[i]
       + BETAtype * newtype[i] ;
     betaz2[i] <- BETAt0 + BETAtrt * trt[i] ;

     nu[i] ~ dgamma( EACHALPHA[i] , EACHALPHA[i] ) ;
     EACHALPHA[i] <- exp( XI0 + XItype * newtype[i] ) ;

     lambdau[i] <- u * exp( betaz1[i] ) * nu[i] ;
     lambdapsi[i] <- exp( betaz1[i] ) * exp( betaz2[i] )
       * nu[i] ;

     count[i] ~ dpois( lambdau[i] ) ;
     survtime[i] ~ dexp( lambdapsi[i] ) ;
  }
}
```

References

Aalen, O. O. (1988). Heterogeneity in Survival Analysis. *Statistics in Medicine*, **7**, 1121–1137.

Andersen, P. K., Borgan, Ø., Gill, R. D., & Keiding, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer-Verlag.

Cameron, A. C., & Trivedi, P. K. (1998). *Regression Analysis of Count Data*. Cambridge, UK: Cambridge University Press.

Carroll, R. J., Ruppert, D., & Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models*. London: Chapman and Hall.

Clayton, D. (1994). Some Approaches to the Analysis of Recurrent Event Data. *Statistical Methods in Medical Research*, **3**, 244–262.

Collett, D. (1994). *Modelling Survival Data in Medical Research*. London: Chapman and Hall.

Collett, D. (2003). *Modelling Survival Data in Medical Research*. Second edn. London: Chapman and Hall.

Congdon, P. (2001). *Bayesian Statistical Modelling*. New York: John Wiley and Sons.

Cook, R. J., & Lawless, J. F. (2002). Analysis of repeated events. *Statistical Methods in Medical Research*, **11**, 141–166.

Cox, D. R. (1962). *Renewal Theory*. London: Methuen.

Cox, D. R., & Isham, V. (1980). *Point Processes*. Monographs on Applied Statistics and Probability, vol. 12. London: Chapman and Hall.

Cox, D. R., & Oakes, D. (1984). *Analysis of Survival Data*. Monographs on Applied Statistics and Probability, vol. 21. London: Chapman and Hall.

Daley, D. J., & Vere-Jones, D. (1988). *An Introduction to the Theory of Point Processes*. Berlin: Springer-Verlag.

De Silva, M., MacArdle, B., McGowan, M., Hughes, E., Stewart, J., Neville, B. G., & Johnson, A. L. (1996). Randomised comparative monotherapy trial of phenobarbitone, phenytoin, carbamazepine, or sodium valproate for newly diagnosed childhood epilepsy. *Lancet*, **347**, 709–713.

Dean, C., Lawless, J. F., & Willmot, G. E. (1989). A Mixed Poisson-Inverse Gaussian Regression Model. *Canadian Journal of Statistics*, **17**, 171–182.

Diggle, P. J. (1988). An approach to the analysis of repeated measures. *Biometrics*, **44**, 959–971.

Diggle, P. J., Heagerty, P., Liang, K. Y., & Zeger, S. L. (2002). *Analysis of Longitudinal Data*. Second edn. Oxford Statistical Science, vol. 25. Oxford, UK: Oxford University Press.

Eggenberger, F., & Polya, G. (1923). Über die Statistik Verketteter Vorgänge. *Zeitschrift für Angerwandte Mathematik and Mechanik (Journal of Applied Mathematics and Mechanics)*, **1**, 279–289.

Faucett, C. L., & Thomas, D. C. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in Medicine*, **15**(15), 1663–85.

Feller, W. (1943). On a General Class of 'Contagious' Distributions. *Annals of Mathematical Statistics*, **14**, 389–400.

Ford, I., Norrie, J., & Ahmadi, S. (1995). Model Inconsistency, Illustrated by the Cox Proportional Hazards Model. *Statistics in Medicine*, **14**, 735–746.

Gamerman, D. (1997). Sampling from the Posterior Distribution in Generalized Linear Mixed Models. *Statistical Computing*, **7**, 57–68.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (eds). (1996). *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.

Gourieroux, C., & Visser, M. (1997). A count data model with unobserved heterogeneity. *Journal of Econometrics*, **79**, 247–268.

Gourieroux, C., Monfort, A., & Trognon, A. (1984). Pseudo Maximum Likelihood Methods: Theory. *Econometrica*, **52**, 681–700.

Greenwood, M., & Yule, G. U. (1920). An Inquiry into the Nature of Frequency Distributions of Multiple Happenings, with Particular Reference to the Occurrence of Multiple Attacks of Disease or Repeated Accidents. *Journal of the Royal Statistical Society*, **83**, 255–279.

Heller, A. J., Chesterman, P., Crawford, P., Chadwick, D., Johnson, A. L., & Reynolds, E. H. (1995). Phenobarbitone, phenytoin, or sodium valproate for newly diagnosed epilepsy: a randomized comparative monotherapy trial. *Journal of Neurology, Neurosurgery, and Psychiatry*, **58**, 44–50.

Hougaard, P. (1986a). Survival Models for Heterogeneous Populations Derived from Stable Distributions. *Biometrika*, **73**, 387–396. Correction *Biometrika*, **75**, 395.

Hougaard, P. (1986b). A Class of Multivariate Failure Time Distributions. *Biometrika*, **73**, 671–678. Correction *Biometrika*, **75**, 395.

Hougaard, P. (1987). Modelling Multivariate Survival. *Scandinavian Journal of Statistics*, **14**, 291–304.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data.* New York: Springer-Verlag.

Hougaard, P., Lee, M. L. T., & Whitmore, G. A. (1997). Analysis of Overdispersed Count Data by Mixtures of Poisson Variables and Poisson Processes. *Biometrics*, **53**, 1225–1238.

Ibrahim, J. G., Chen, M. H., & Sinha, D. (2001). *Bayesian Survival Analysis.* New York: Springer-Verlag.

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**, 799–803.

Klein, J. P., & Moeschberger, M. L. (1997). *Survival Analysis*. New York: Springer-Verlag.

Kwong, G. P. S., & Hutton, J. L. (2003). Choice of parametric models in survival analysis: applications to monotherapy for epilepsy and cerebral palsy. *Applied Statistics*, **52**, 153–168.

Laird, N. (1978). Nonpaametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association*, **73**, 805–811.

Landau, S. (ed). (1986). *International Dictionary of Medicine and Biology*. Vol. 1. New York: John Wiley and Sons.

Lavine, M. (1992). Some Aspects of Polya Tree Distributions for Statistical Modeling. *Annals of Statistics*, **20**, 1222–1235.

Lavine, M. (1994). More Aspects of Polya Tree Distributions for Statistical Modeling. *Annals of Statistics*, **22**, 1161–1176.

Lindeboom, M., & van den Berg, G. J. (1994). Heterogeneity in Models for Bivariate Survival: the Importance of the Mixing Distribution. *Journal of the Royal Statistical Society, Series B*, **56**(1), 49–60.

Lindsey, J. K. (1993). *Models for Repeated Measurements*. Oxford Statistical Science, vol. 10. Oxford, UK: Oxford University Press.

Luceño, A. (1995). A family of partially correlated Poisson models for overdispersion. *Computational Statistics and Data Analysis*, **20**, 511–520.

Marshall, A. W., & Olkin, I. (1990). Multivariate Distributions Generated from Mixtures of Convolution and Product Families. *Pages 371–393 of:* Block,

H. W., Sampson, A. R., & Savits, T. H. (eds), *Topics in Statistical Dependence*. IMS Lecture Notes Monograph Series, vol. 16. Hayward, CA: Institute of Mathematical Sciences.

Marson, A. G., Williamson, P. R., Clough, H., Hutton, J. L., & Chadwick, D. W. (2002). Carbamazepine versus Valproate Monotherapy for Epilepsy: A Meta-analysis. *Epilepsia*, **43**(5), 505–513.

Mattson, R. H., Cramer, J. A., & Collins, J. F. (1992). A comparison of Valproate with Carbamazepine for the treatment of complex partial seizures and secondary generalized tonic-clonic seizures in adults. *New England Journal of Medicine*, **327**(11), 765–771.

McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models*. Second edn. Monographs on Applied Statistics and Probability, vol. 37. London: Chapman and Hall.

McGilchrist, C. A., & Aisbett, C. W. (1991). Regression With Frailty in Survival Analysis. *Biometrics*, **47**, 461–466.

Natarajan, R., & Kass, R. E. (2000). Reference Bayesian Methods for Generalized Linear Mixed Models. *Journal of the American Statistical Association*, **95**(449), 227–237.

Oakes, D. (1982). A Model for Association in Bivariate Survival Data. *Journal of the Royal Statistical Society, Series B*, **44**, 414–422.

Oakes, D. (1989). Bivariate Survival Models Induced by Frailties. *Journal of the American Statistical Association*, **84**, 487–493.

Pickles, A., & Crouchley, R. (1995). A Comparison of Frailty Models for Multivariate Survival Data. *Statistics in Medicine*, **14**, 1447–1461.

Prentice, R. L., Williams, B. J., & Peterson, A. V. (1981). On the Regression Analysis of Multivariate Failure Time Data. *Biometrika*, **68**, 373–379.

Richens, A., Davidson, D. L., Cartlidge, N. E., & Easter, D. J. (1994). A multicentre comparative trial of sodium valproate and carbamazepine in adult onset epilepsy. Adult EPITEG Collaborative Group. *Journal of Neurology, Neurosurgery, and Psychiatry*, **57**(6), 682–687.

Roberts, G. O. (1996). Markov chain concepts related to sampling algorithms. *In: Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.

Shaban, S. A. (1988). Poisson-Lognormal Distributions. *Chap. 7, pages 195–210 of:* Crow, E. L., & Shimizu, K. (eds), *Lognormal Distributions: Theory and Applications*. Statistics: Textbooks and Monographs, vol. 88. New York: Marcel Dekker.

Smith, W. L. (1958). Renewal Theory and its Ramifications. *Journal of the Royal Statistical Society, Series B*, **20**, 243–302.

Spiegelhalter, D., Thomas, A., & Best, N. (2000). *WinBUGS: Bayesian inference Using Gibbs Sampling. (User Manual version 1.3)*. MRC Biostatistics Unit. URL: `http://www.mrc-bsu.cam.ac.uk/bugs/`.

Thall, P. F., & Vail, S. C. (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics*, **46**, 657–671.

Toscas, P. J., & Faddy, M. J. (2003). Analysis of Longitudinal Count Data Using a Generalised Poisson Model. *Statistical Modelling*, **3**, 99–108.

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

Walker, S. G., & Mallick, B. K. (1997). Hierarchical Generalised Linear Models and Frailty Models with Bayesian Nonparametric Mixing. *Journal of the Royal Statistical Society, Series B*, **59**, 845–860.

Wallace, H., Shorvon, S. D., Hopkins, A., & O'Donoghue, M. (1997). Adults with poorly controlled epilepsy: part 1 - Clinical guidelines for treatment. *Royal College of Physicians of London 1997*, 2–30.

Williamson, P. R., Clough, H. E., Hutton, J. L., Marson, A. G., & Chadwick, D. W. (2002). Statistical issues in the assessment of the evidence for an interaction between factors in epilepsy trials. *Statistics in Medicine*, **21**, 2613–2622.

Winkelmann, R. (1995). Duration Dependence and Dispersion in Count-Data Models. *Journal of Business and Economic Statistics*, **13**(4), 467–474.

Xu, J., & Zeger, S. L. (2001). Joint analysis of longitudinal data comprising repeated measures and times to events. *Applied Statistics*, **50**(3), 375–388.

Zeger, Scott L., & Liang, K.-Y. (1992). An Overview of Methods for the Analysis of Longitudinal Data. *Statistics in Medicine*, **11**, 1825–1839.