9th month Thesis Report

Boryana C. Lopez K. Supervisor: Dr. Jane Hutton

10 of June 2013

Chapter 1 Introduction

In this research work, the main interest has been placed on modelling the seizure recurrence of epilepsy, and providing useful prognostic information for patients under treatment for such disease. The statistical approach to this problem has been done through Survival analysis theory, for which some basic definitions are stated in the first part of the Appendix. Although the Appendix comprises complementary calculations and discussions on the work presented here, it is not essential for the understanding of the first three chapters.

The first chapter presents the object of study, which focuses on a multicentre trial that follows-up 1443 patients with different types of Epilepsy. The data collected before, at and after randomization of the patients to two treatment schemes is presented in a summarized fashion. The objectives of the study are made clear, and from this we proceed to the second chapter. Chapter two comprises the research study of a succession of books and papers that have addressed recurrent event data analysis, and the properties of the models are discussed and compared among each other, in an attempt to show the relevance of the last model presented by Rogers & Hutton[9]. This survival model is strongly considered for the development of future work, and hence the third chapter presents this, and two other possible lines of work.

1.1 Mess Study introduction

The Multicentre study of early Epilepsy and Single Seizures (MESS) consists of a study and its resulting data set. Patients in the study had a history of epileptic seizures and their clinicians were unsure of the need for an anti-epileptic drug (AED). Patients were recruited for over five years, and randomized to one of two policies: deferred or immediate treatment. The aims of this study are:

- to measure the differences in policies
- to define prognostic factors for seizure recurrence
- to define psychosocial outcomes of the policies, and
- to make results available in a form which allows patients to make informed decisions.

When treating with AEDs, we expect certain advantages such as the reduction of the number of seizures in the short term, and the remission from epilepsy in the future. However, from the use of such AEDs the patients risk physical and psychological damage, sometimes serious. This study developed as an attempt to assess which policies are optimal for the diverse groups of epilepsy patients. It is of particular interest to understand what the risk of recurrence is, once a first seizure has happened, and how the treatment alters that risk.

The patients were eligible for the trial if they presented a history of one or more spontaneous, unprovoked epilepsy seizures, were at least one month old, had not taken AEDs before, and they were, along with their clinicians, uncertain of needing treatment. Patients with a progressive disease were not considered for the study.

MESS randomized patients with single seizures, subjects with infrequent seizures and patients with more frequent seizures with less severe symptomatology. Thus, the MESS study randomized patients presenting tonic-clonic seizures pre-randomization (including primary and secondary generalized seizures), and with seizures that were simple partial, complex partial, or generalized absence and myoclonic seizures.

Between 1993 and 2000, 1847 patients were invited to join the trial, from UK and non-UK medical centres. Of this 1847 patients, 404 did not consent to randomization, while subsequently 722 were randomly assigned to immediate treatment, and 721 to deferred treatment. Over five years a total of 1443 patients were recruited. The randomization process was undertaken by an independent randomization centre, that balanced across two factors: centre or region, and number of seizures at randomization. The recruitement resulted in 56% of the patients having a single seizure in their history, and the remaining 44% had a history of at least two epileptic seizures. There were 717 (49.6%) subjects from the UK, and 726 (50.3%) from other countries. Further details are shown in the Table 1.1.

For patients allocated to the immediate policy, their clinicians determinated the optimal AED for the patient and prescribed it as soon as possible. Most of this group were treated with carbamazepine (CBZ, 45%), or with valproate (VPS, 45%) and 10% with other AEDs. The remaining patients in the deferred treatment policy received no drugs until the clinician and patient determined that treatment was necessary. Of such deferred-located patients, 332 started treatment during the course of the trial: 134 (40%) with carbamazepine, 142 (43%) with valproate, and the remaining 37 with another AED.

Of the patient's history pre-randomization, data such as date of birth, the times to seizures and total number of seizures prior to randomization, and the type of epilepsy for each individual was collected. The information gathered at randomization included imaging, Electro-encephalogram (EEG) outcome if it was available, age at randomization, date of randomization. The types of EEG abnormalities, the allocation to treatment policy, type and dose of drug were also recorded. The patients' follow-up lasted up to 8 years. The outcomes were the times to first, second and fifth seizures post-randomization and the types of seizures. The following table was extracted from Marson et al [1].

	Immediate treatment ($n = 722$)		Deferred treatment ($n = 721$)	
	Number	Percentage	Number	Percentage
Sex				
Male	403	56%	423	58%
Female	319	44%	298	42%
Centre				
UK	363	50%	354	49%
Non-UK	359	50%	367	51%
EEG abnormalities				
Non-specific abnormality only	83	11%	88	12%
Generalized	131	18%	105	15%
Focal	184	25%	200	28%
Imaging abnormal	71	10%	69	10%
Seizure types before randomization				
Simple partial	15	2%	20	3%
Complex partial	36	5%	32	4%
Secondary generalized Tonic-Clonic	239	33%	215	30%
Myoclonus only	6	$< 1\%$	5	$< 1\%$
Absence only	3	$< 1\%$	3	$< 1\%$
Tonic-Clonic seizures	375	52%	406	56%
Combinations of generalized seizures	21	3%	19	3%
Other seizures	17	2%	13	2%
Number of seizures before randomization				
1	404	56%	408	57%
2	183	25%	165	23%
3	50	7%	58	8%
4	28	4%	18	2%
$5 - 9$	30	4%	36	5%
≥ 10	17	2%	28	4%
Clinical and family history				
Developmental delay/learning disability	34	5%	23	3%
Neurological deficit	52	7%	40	6%
Previous neurological insult	99	14%	90	12%
Previous febrile seizures	53	7%	52	7%
Previous acute symptomatic seizures	14	2%	19	3%
First-degree family history of seizures	76	11%	86	12%

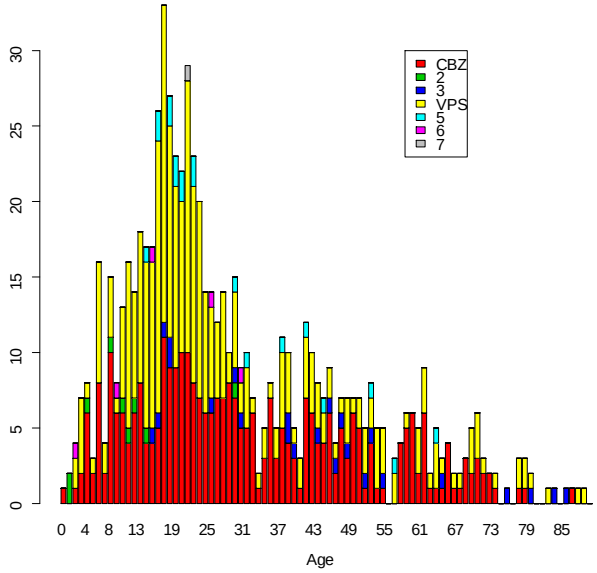
(1.1)

Table 1.1: MESS study data outcome.

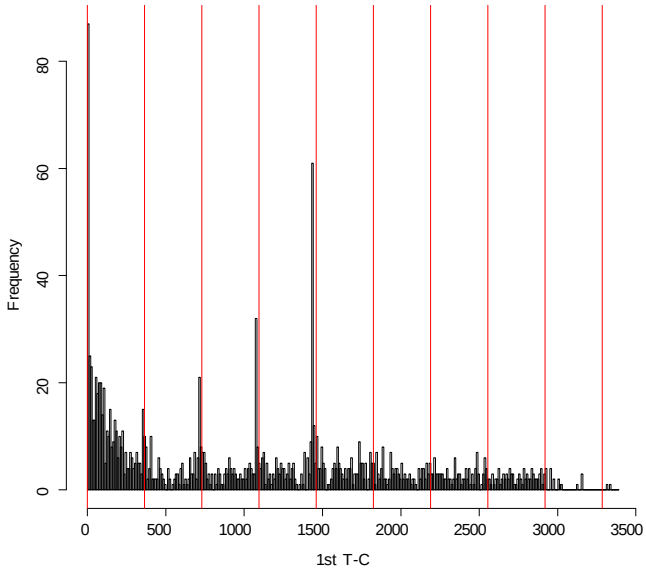
Since with the passing of time, patients in the trial died or left the study, it can be observed that the proportion of patients receiving an AED in the two groups becomes gradually smaller. At 5 years from randomization, 60% of the patients in the immediate policy group are still receiving treatment, contrasted with 41% in the deferred policy group.

In the histograms presented in Figure 1.2, the proportion of patients that received a type of drug at a certain age is shown. Observe that young patients ranging from 5 to 28 years receive a considerable more amount of VPS drug. In the second histogram, the red lines represent the start of a new year. The initial high peaks of the Tonic-Clonic seizures post-randomization appear consistently very near the ending of each year, but this is most likely explained by the censoring

mechanism, as some institutions might have reported such seizures at the end of the year.



Age at randomization by AED



Tonic-Clonic seizures by day and year

Figure 1.2 (1.2)

1.2 MESS study first proposed analyses

According to Marson et al.[1], almost all patients were prescribed CBZ or VPZ, 92% in the immediate group, and 83% of those receiving AEDs in the deferred group. This paper introduces and describes the MESS study and endeavours to supply a prognosis for patients within the study. It is mentioned that the study was conducted as a randomized, unmasked and pragmatic trial, but later on it was found that the unmasking does not appear to generate a bias for the study, judging by the comparison between groups on first seizure recurrence, such recurrence did not appear to be different between them.

Although the initial intention was to recruit 3000 patients for the study in order to perform a split-half cross-validation analysis, only 1443 patients were included and accepted to randomization. As a consequence, this paper focuses on performing the analysis with the population provided, and suggests that performing a cross-validation procedure is not carried out, since the power for generating a validation sub-population diminishes greatly. However, it is noticeable that in a subsequent paper by Kim et al (2006)[2], this same MESS study is re-addressed with a cross-validation analysis.

The analysis of the first time to event is performed with a log-rank test, or Cox's proportional hazards model when adjusting single versus multiple seizures at entry. All analyses were by intention to treat, and the results are reported as absolute differences in proportions or hazard ratios with 95% confidence intervals.

For the times to first and second seizures post-randomization, the difference between treatment groups (immediate versus deferred) is highly significant, while for the time to the fifth seizure, there is no perceptible difference between groups. As for the remission period, the actuarial estimate of

achieving a 2 year remission by the 8th year is very high (over 94% for both treatment groups and single versus multiple seizures). Although the immediate treatment scheme increased the times to first and second seizures post-randomization, and decreased the time to 2 year remission, there is no significant difference between the immediate and the deferred treatments for the time to 5 year remission (Table 1.2).

	Immediate (%)		Deferred (%)		Difference
	Single	Multiple	Single	Multiple	(95% CI)
Time to first seizure					
6 months	18	26	26	44	12 (7.4, 16.5)
2 years	32	43	39	61	11 (6.2, 16.7)
5 years	42	57	51	69	10 (4.5, 16)
8 years	46	60	52	72	9 (2.6, 15.3)
Time to second seizure					
6 months	14		19		5 (1.1, 8.9)
2 years	24		32		8 (3.6, 13.3)
5 years	34		40		6 (0.9, 12)
8 years	38		44		5 (−1.4, 12)
Time to fifth seizure					
6 months	6		7		1 (−1.3, 3.8)
2 years	12		15		3 (−1.1, 6.2)
5 years	19		22		3 (−1.6, 7.6)
8 years	26		25		−1 (−9.3, 7.8)
2-year remission					
2 years	69	57	61	39	12 (6.3, 17.4)
5 years	92	91	92	87	2 (−1.2, 6.1)
8 years	95	94	96	95	1 (−2.5, 3.9)

Table 1.2: MESS study seizure counts.

As a general result, in Marson et al[1], it is mentioned that the benefits from immediate treatment do not impact on the long-term conditions of the patients, but findings from a subpopulation of 441 UK adult patients show that they do come at some cost for the patient. Indeed, from the 527 eligible UK patients, 441 of them returned a quality of life (QOL) question sheet at 100 days from randomization. This data was later on used to analyse the changes between baseline and 2-year follow-up for anxiety, depression and mastery (an individual's perception of control over his life). The immediate treatment patients reported more adverse events that were related to the treatment, such as depression, anxiety, gastrointestinal symptoms, among others.

One year later, Kim et al[2] considered the effects of the two main anti-epileptic drugs (AEDs), carbamazepine (CBZ) and valproate (VPS), with the objective of studying the roles of treatment policies and patient characteristics in the subsequent risk of seizure recurrence. They took a cross-validation sample using only a subset of the total population to assess the model, while the remaining subset was then used to validate such model. From the 1443 patients, 885 of them were assigned as a test sample, 535 patients were used for validation, and the remaining subjects were discarded from the study, as they had incomplete data. The population was then split into three categories (low, medium and high risk), and stratified by randomized treatment policy.

The test sample was used to formulate the prognostic model, which considered the time from randomization to the first seizure hence, based on a Cox regression and stratified by treatment allocation. Both backward and forward stepwise regressions were used to find the predictive value of the baseline covariates of interest. Each stepwise regression gave place to a different model, given that a covariate was included in the model if they were significant in univariate

analysis at $p \leq 0.05$ and it was excluded if $p \geq 0.1$. Since the number of patients with several types of neurological disorder was low, all conditions (neurological deficit or impairment, delayed development and learning disability) were represented under one variable only. In the same way, the variable called EEG contained three Electro-encephalogram (EEG) measurements (specific local, generalized epileptiform and slow wave abnormality).

The resulting models were Model 1 for backward stepwise elimination, and Model 2 for forward stepwise elimination, each stratified by treatment. See Table 1.3.

Variables	Model 1		Model 2	
	Hazards ratios	p-value	Hazards ratios (95% CI)	p-value
Neurological disorder	$1.35(1.07, 1.72)$	$p = 0.013$	$1.36(1.07, 1.73)$	$p = 0.012$
Abnormal EEG	$1.54(1.27, 1.86)$	$p < 0.0001$	$1.53(1.26, 1.86)$	$p < 0.0001$
Number of seizures pre-randomization (log transformation)	$1.56(1.42, 1.72)$	$p < 0.0001$	$1.61(1.46, 1.77)$	$p < 0.0001$
Years between recent seizure and randomization			$0.92(0.84, 1.02)$	$p = 0.1$
First degree relative with epilepsy			$1.27(0.96, 1.7)$	$p = 0.1$

Table1.3 takenfromKimetal[2]. (1.3)

The hazard ratios can be interpreted as the relative change in risk for a unit increase in the prognostic factor. Observe that the second model contains the variables of the first model and it has two aditional variables. However, the identification of two large outliers in the data resulted in the exclusion of the variable of years between recent seizure and randomization. Furthermore, since the two variables added in the second model had a borderline significance of $p = 0.1$, for simplicity they were both omited, which resulted in considering only the first model as the final model.

A prognostic index was defined as the linear predictor resulting from the final model. It was the result of the sum of covariate values for a particular patient, weighted by the corresponding estimated regression coefficients. According to this prognostic index, the population was split into the three following categories:

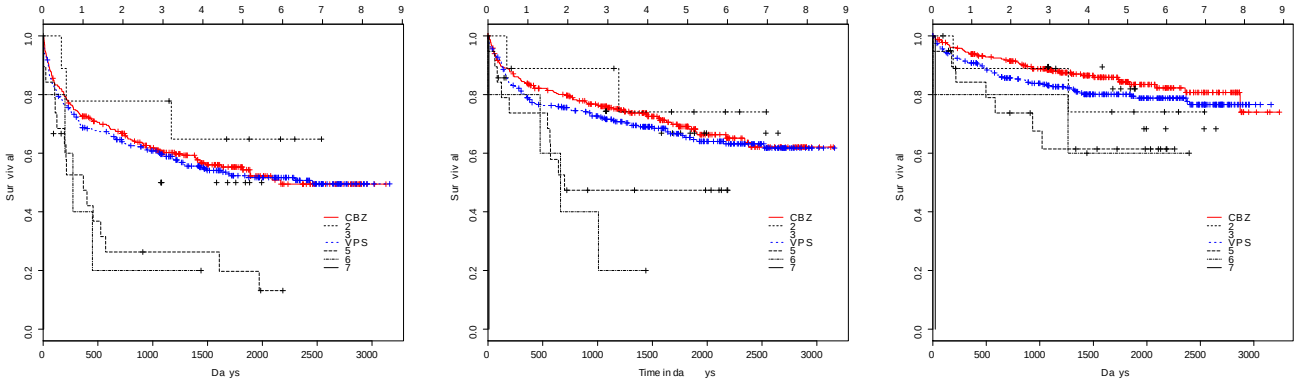
Prognostic index= 0 (Low risk of recurrence)	Prognostic index= 1 (Medium risk recurrence)	Prognostic index= 2 to 4 (High risk recurrence)
- A single seizure	- 2 to 3 seizures, normal EEG, no neurological abnormalities	- 1 seizure, abnormal EEG, neurological disorder
- A normal EEG	- 1 seizure, abnormal EEG, no neurological disorder	- 2-3 seizures, abnormal EEG, no neurological disorder
- Absence of neurological disorder	- 1 seizure, normal EEG, neurological disorder	- 2-3 seizures, normal EEG, neurological disorder - 4 or more seizures.

Table1.4 (1.4)

Subsequently, the validation sample was also split into these three categories, and used to measure the predictive accuracy of the model by plotting the observed proportions of individuals with seizure recurrence, separated by treatment policy and stratified by risk recurrence levels. These plots were compared to the corresponding plots from the test sample.

The model and the Kaplan-Meier plots (Tables 1.5 and 1.6) give an indication that, although the benefit of immediate treatment is not obvious for the low risk population, it does present a

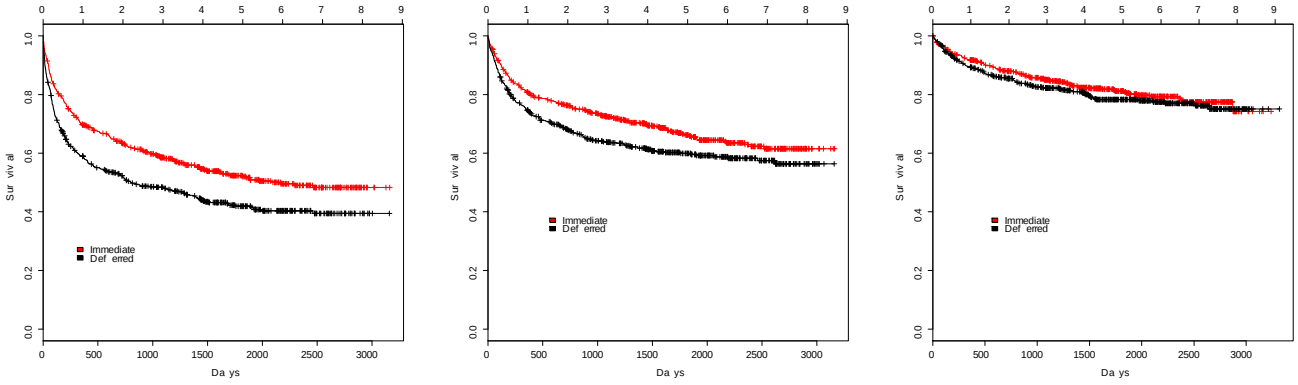
delay in the immediate seizure recurrence for the medium and high risk of recurrence populations. The results also support Marson et al.'s findings, regarding that there is no clinical benefit from immediate treatment in late risks, such as three to five years remission. The risk of seizure recurrence increases with the number of seizures at presentation, abnormal EEG, and the presence of neurological disorders.



Survival curves for 1^{st} seizure Survival curves for 2^{nd} seizure Survival curves for 5^{th} seizure

Figure 1.5 Survival curves by AED administered post-randomization. (1.5)

In Figure 1.5, we can observe the Kaplan-Meier curves for the time from randomization to the times to first, second and fifth seizures, for each one of the different drugs. Observe that for the two drugs of main interest, CBZ and VPS, the differences between survival functionss are not apparent until the fifth seizures. However, we must keep in mind that the number of patients who experience the fifth seizure is considerably less than the initial number of patients. In Figure 1.6 the difference between the survival curves is much more clear. Notice that in both figures, the lower x-axis measures time in days, while the upper x-axis represents the time in years.



1^{st} seizures post-randomization 2^{nd} seizures post-randomization 5^{th} seizures post-randomization

Figure 1.6 Kaplan-Meier survival curves by treatment allocation. (1.6)

It is important to notice that, as warned by this paper[2], the MESS study is unusual in its including patients with frequent and minor seizures, and this results may not apply to other studies

of similar nature. It would be desirable to perform an external validation on other data sets, as a mean for a more rigorous model validity.

Chapter 2 Epilepsy Models

An important approach to medical recurrent event problems lies in the use of survival model analysis. In a large number of cases, due to familiarity and convenience, researchers tend to prefer choosing Cox's proportional hazards model as the underlying model behind such problems. These types of models are defined and further explained in the first section of the Appendix. However, more recent work has suggested that accelerated life models might provide a better fit in some cases. This poses the question of what is the impact of survival regression model misspecification on the parameter estimates and the conclusions they can lead to.

Indeed, the paper by Kwong and Hutton[5] addresses this problem, and mentions that it is not only important to consider the effects of model misspecification, but that it should also be noted that the choice of the baseline family distribution is often ignored. When working with epilepsy data, several important questions arise, such as which factors influence the recurrence of seizures, which base-line distributions best fit this disease, which assumptions must be made about the model, and how robust the results and interpretations are under base-line and survival regression model misspecification.

2.1 Background

In 1997, Hutton and Solomon[6] studied the robustness of regression estimates under misspecification of the survival models. In other words, assuming a regression analysis on survival data, they were interested in the changes that the regression parameters would experience if the estimation were carried out under a proportional hazards model assumption, when the true underlying model was an accelerated life model, and viceversa. For this purpose, the main interest lies in using the orthogonality of parameters to study the effects of model misspescification. By orthogonalizing the parameters of interest from the nuisance parameters, they were able to obtain an expression of the regression parameters relating their value under the true model, and the value under the assumed model.

Using the same form of Cox & Reid's[3] mixture model, the model proposed in Hutton & Solomon[6] is a mixture of an accelerated life (AL) model, and a proportional hazards (PH) model:

$$m(t) = \left\{ g_0(t)\, e^{\beta' z} G_0(t)^{\exp(\beta' z)-1} \right\}^{\psi} \left\{ f_0\left(te^{\gamma' z}\right) e^{\gamma' z} \right\}^{1-\psi},$$

where the survivor functions $F_0\left(te^{\gamma' z}\right)$ and $G_0(t)^{\exp(\beta' z)}$ correspond to the accelerated life model and the proportional hazards model respectively; F_0 and G_0 are the base-line survivor functions with corresponding densities f_0 and g_0 . Here z is the vector of fixed covariates, and β and γ are the associated vectors of unknown regression parameters for the proportional hazards and accelerated life families of models. The interest lies in finding expressions for the parameters of interest, β and γ, while considering ψ as a nuisance parameter.

Now let us consider the following

Definition 1 *If $\theta = (\theta_1, \theta_2)$ where θ_1 and θ_2 partition θ, and $i_{\theta_i \theta_j}$ corresponds to the ij^{th} entry of the information matrix, then we define θ_1 to be orthogonal to θ_2 if*

$$i_{\theta_s \theta_t} = \frac{1}{n} E\left[\frac{dl(\theta; y)}{d\theta_s} \frac{dl(\theta; y)}{d\theta_t}; \theta \right] = \frac{1}{n} E\left[-\frac{d^2 l(\theta; y)}{d\theta_s d\theta_t}; \theta \right] = 0$$

for all $\theta_s \in \theta_1$ and for all $\theta_t \in \theta_2$.

We have from the Cox & Reid (1987) paper, shown in the Appendix under "Parameter orthogonalization ", two parameters (ψ, ϕ) are orthogonal if the following equation holds

$$\sum_{r=1}^{q} i^*_{\phi_r \phi_s} \frac{d\phi_r}{d\psi} = -i^*_{\psi\phi_s},$$

where $i^*_{\psi\phi_s} = E\left[\frac{d^2 l^*(\psi,\phi)}{d\psi d\phi_s}\right]$, $i^*_{\phi_r\phi_s} = e\left[\frac{d^2 l^*(\psi,\phi)}{d\phi_r d\phi_s}\right]$ and l^* is the log-likelihood function corresponding to ψ and ϕ. These are also called orthogonalizing equations of the parameters ϕ_r and ϕ_s. Among several other convenient properties of two orthogonal parameters ψ and λ, if $\hat{\psi}$ and $\hat{\lambda}$ are the maximum likelihood estimators of ψ and λ respectively, then they are asymptotically independent.

In the case of this problem, the exact expressions of this orthogonalizing equations are not obtainable, hence the mixture function $m(t)$ is normalized and equated to the likelihood function

$$L(\psi, \beta, \gamma; \alpha, t, z) = \frac{m(t)}{\int_0^\infty m(t)dt};$$

where α, which is assumed as known, are the parameters of the base-line survival distributions.

While the formal computations are shown in the Appendix, it is worth mentioning that in order to approximate $\int_0^\infty m(t)dt$, a Taylor series is used about $(\beta, \gamma, \psi) = 0$. Observe that this requires the assumption that β and γ are small.

After several approximation steps, it was found that, if the true model is the accelerated life model, which implies that $\psi = 0$, but the assume model is a proportional hazards model, then the orthogonalizing equations result in the expression

$$\beta_s = -\gamma_s \left(1 + \frac{c_4}{c_2}\right),$$

where

$$c_2 = E_{f_0}[\log G_0(T)] \quad \text{and} \quad c_4 = E_{f_0}\left[T\dot{y}(T) \log G_0(T)\right],$$

and T is the time to ocurrence random variable.

The expression for the regression parameters are obtained in a similar manner when the true model is a proportional model, and the accelerated life model is assumed, which corresponds to the case when $\psi = 1$. The equation in question is of the form

$$\gamma_s = -\frac{\beta_s d_6}{d_4},$$

where

$$d_4 = E_{g_0}\left[T\dot{y}(T) \log\left\{g_0(T)/f_0(T)\right\}\right] \quad \text{and} \quad d_6 = E_{g_0}\left[T\dot{y}(T)\left\{1 + \log G_0(T)\right\}\right].$$

Observe that the relationship between the regression parameters under model misspecification, in both cases, is proportional to the first degree. It can be shown that the expressions can be simplified to

$$\frac{\beta_s}{\beta_1} = \frac{\gamma_s}{\gamma_1} \text{ for } s = 2, 3, ..., 1,$$

which means that the ratios of regression coefficients are consistent.

Important results from Hutton & Solomon seem to indicate that, under the assumption of small regression coefficients, these coefficients are consistent when facing model misspecification. This result is then reconsidered in the paper by Kwong & Hutton[5] six years later, with the

purpose of investigating the robustness of such parameter estimates under model misspecification, also assuming in this work that the covariate coefficients are small.

In order to investigate the properties of misspecification in a practical manner this paper addresses two applications: monotherapy for epilepsy and cerebral palsy studies. For each one of them, four distributions are considered for the base-lines: the gamma, log-normal, log-logistic and Weibull distributions. Both proportional hazards and accelerated life models are adjusted and studied.

Although the cerebral palsy data produced results that were consistent with Hutton and Solomon's findings, the epilepsy data analysis presented a more complex case. The epilepsy study consisted of 1183 patients from five different trials, in which each person was randomly assigned either the drug carbamazepine (CBZ) or valproate (VPS). The outcomes were the times to the first seizure after the randomization, and the variables of interest were either categorical (trial, drug and seizure type) or continuous (age, drug and age interaction, and the number of seizures six months before randomization). Such variables of interest were selected by comparing the values of $-2\log(\hat{L})$ for all linear combinations of explanatory variables in Cox proportional hazards models, where L is the maximum likelihood function. Although the Drug variable was not significant in this case, it was considered for the analysis as it was of special medical interest. From the 1183 patients in the study, 836 (71%) showed a first seizure post-randomization, and the remaining 347 (29%) patients had censored observations.

Since neurologists believed that CBZ is a more effective drug for patients with partial seizures, and VPS is better for cases with general seizures, the analysis was directed to compare the survival between the groups combining CBZ and VPS drugs with generalized and partial seizures. In all four models considered, it was found that patients with generalized epilepsy had twice as long until their first seizure post-randomization than those patients with partial epilepsy.

The proportional hazards model was shown to be a bad model for this study, as the proportional hazards constraint on the Weibull base-line distribution failed to reflect the difference in hazard rates for generalized *versus* partial epilepsy. If proportionality constraint were true, the hazards ratio between two groups should be constant with time, but in the life-table obtained, the estimates of the hazard function of the generalized epilepsy group and the estimates corresponding to the partial seizures group did not appear to be this way. The smaller AICs in the gamma, log-normal and log-logistic models indicated that accelerated life models fitted the data better than proportional hazards models. When considering the 95th, 90th and 85th percentiles of the survival, the Weibull model is shown to greatly underestimate the survival.

Finally, for this data, it was observed with a Kaplan-Meier approximation, that the failure distribution was highly skewed, since there was a large number of early seizures. As a result, after considering small covariate coefficients, there is evidence to suggest that the regression estimates under misspecification of the models were not proportional to first-degree. Thus for this study, the results from Hutton & Solomon (1997) do not hold. Indeed, observe that, if h_{AL} is the accelerated life hazard function, then

$$h_{AL}(t) = e^{\beta' z} h_0(e^{\beta' z} t) \simeq e^{\beta' z} h_0 \left[(1 + \beta' z)t\right]$$

when approximating $e^{\beta' z}$ with its first Taylor series term. Furthermore, when approximating h_0 by its Taylor expansion about t, we obtain

$$\begin{aligned} h_{AL}(t) &\simeq e^{\beta' z} \left\{ h_0(t) + \beta' z t h_0'(t) \right\} \\ &= e^{\beta' z} h_0(t) + \beta' z t e^{\beta' z} h_0'(t). \end{aligned} \tag{2.1}$$

Observe that the first term of (2.1) is an approximation to a Cox proportional hazard function, and thus Hutton & Solomon's result will hold only if the second term of the equation is zero. If the hazard function is decreasing (increasing) rapidly, $h_0'(t)$ will be positive (negative) and large, and cannot be ignored, then the function $h_{AL}(t)$ is not approximately a Cox's proportional hazard, and consequentially, Hutton & Solomon's consistency of the parameters does not hold. This result also proposes that estimates of regression parameters corresponding to survival models with late survival times are robust under misspecification, but if the model was such that the survival distribution is strongly right-skewed, with a short median survival, the parameters are not robust.

Precisely in this example of epilepsy data, the survival Kaplan-Meier plots showed that the hazard decreases rapidly due to the abundancy of early seizures. Indeed, for this case the regression parameters do not appear to be consistent for proportional hazards and accelerated life models.

After calculating the maximum likelihood parameter estimates for the four survival models, it was found that the effects of partial epilepsy, age, the natural logarithm of the number of seizures within 6 months before randomization, and the drug-age interaction, tend to reduce the time to a first seizure post-randomization, while VPS and higher age at randomization increase the time to first seizure.

It is worth mentioning that, although there was a small hint that for the drug effect, VPS is better than CBZ, this effect should not be considered on its own since the interaction age-drug was very significant. The estimated effect of age is much more marked with CBZ than with VPS in all models. For young people, the age below which is best to use VPS varies from 23 years (for gamma and Weibull models) to 17 years old (log-logistic model).

Taking this result by Kwong and Hutton, the same year Hutton & Monaghan[7] investigated the effect of misspecifying fully parametric proportional hazards and accelerated life models, taking special interest in observing the impact on the covariate, shape and scale parameters under such models. An important contribution from this paper is the fact that they do not assume that the regression coefficients are "small".

Three categories of misspecification are considered for this paper: when the base-line distribution is false, but the covariate effects are well specified, when the covariate effects are false but the right base-line distribution is considered, and finally when both covariate effects and base-line distribution are misspecified. Additionally, the three distributions considered here for the base-line distribution are: log-normal, log-logistic and Weibull distributions. In this paper, they study the asymptotical distribution of the maximum likelihood (ML) estimator, under the misspecification assumption. Once again, further details can be observed in the Appendix, but the main idea develops from a previous result by (Cox, 1972)[4]; it states that under misspecification, the ML estimator $\hat{\beta}$ of the covariate, shape and location parameters β, are asymptotically distributed as $N(\beta_\alpha, \frac{1}{n}C(\beta_\alpha))$, where$C(\beta_\alpha) = A^{-1}BA^{-1}$, A is the Fisher information matrix and $B(\alpha)_{jk} = E_f\left[\frac{\delta \log L}{\delta \beta_j}\frac{\delta \log L}{\delta \beta_k}\right]$, with L the likelihood function. Additionally, β_α is the solution of a set of expected score equations.

After all different types of misspecifications are considered and the parameters are estimated, a large series of detailed results are stated for every base-line distribution considered. Among the most important results, it was found here that the shape and regression parameters are biased when proportional hazards models are fitted to accelerated life models. Furthermore, the accelerated life model is more robust to misspecification because of its log-linear form. Consistently with Hutton & Solomon[6], it was found that the effect of misspecification decreases as the centre of density of survival times move away from zero, which means that there are not too many early failures.

2.2 Epilepsy seizure recurrence models

From the previous background, we know that the choice of a survival model for medical recurrent events is important, as well as the choice of the parametric base-line distribution. When modelling for epilepsy seizure recurrence, the large amount of information can present a challenge, as it has been demonstrated by previous models that fail to utilize all of the available information.

Let us remember that, among the greatest concerns of proposing a prognosis for epilepsy, is that of understanding which types of epilepsy benefit the most from treatment, and which treatment is best. Although there are seven different antiepileptic drugs (AEDs) under consideration for the MESS study, previous medical analyses show that more attention is directed towards two of them: carbamazepine (CBZ) and valproate (VPS). Additionally, medical epilepsy studies tend to have information of the patient's epilepsy history, previous to being subjected to a treatment. This provides an additional source of information, that in the past has been considered merely as a covariate for the model. Later models have considered to approach the epileptic seizure counts previous to treatment, jointly with the seizure observations after treatment, as time dependent variables.

Previous approaches have been made to medical problems, where the main interest of study is to model phenomena with recurrent events. Some useful approaches were considered:

- Hougaart et. al (1997) used an overdispersed Poisson distribution to model epileptic seizure counts, since in this case the mean is usually smaller than the variance. They used the power variance family as the mixing distribution, but the models did not adjust for explinatory variables.

- Marshall & Olkin (1990) proposed that, if we considered Y_1 as a count variable before randomization, and Y_2 as the count variable post-randomization, then they could be jointly model (Y_1, Y_2) as

$$f(y_1, y_2|z_1, z_2) = \int_0^\infty f_{Y_1}(y_1|z_1, \nu) f_{Y_2}(y_2|z_2, \nu) g(\nu) d\nu,$$

  where f is distributed as a bivariate negative binomial, f_1 is a Poisson distribution with parameter $\mu_1\nu$, f_2 is a Poisson distribution with parameter $\mu_2\nu$ and g is a gamma distribution with parameters α and $1/\alpha$.

- Cameron & Trivedi (1998) considered a bivariate Poisson distribution, and take the heterogeneity between two event times as correlated, not identically.

- Cook & Lawless (2002), reviewed the analysis for repeated data with intensity functions, but did not discuss an analysis of the joint variables (seizure counts, gap times between seizures).

- Cook & Wei (2003) described a bivariate conditional semiparametric approach for count data, mainly, for the variables event counts and event times.

One paper of particular interest to us is Cowling, Hutton & Shaw[8]. The main aims of this study were to present a contrast between the treatment effects, and study the interaction between treatments and the covariates. In this work, the covariates are the age, sex, trial indicator, and type of epilepsy. Although the standard survival analysis method uses the seizure counts pre-randomization as a fixed covariate, the model proposed in this paper treats these counts and the

post-randomization counts as jointly distributed. This model is chosen in order to account for the variation within individuals.

In this paper, the meta-analysis of longitudinal survival data from five trials of two epilepsy treatments is considered. In such treatments, patients were asked to recall seizures from at least six months prior a controlled randomization of the drugs to be administered. In this randomization, the patient would be allocated either the drug carbamazepine (CBZ) or the drug valproate (VPS), which are two common treatments for epilepsy, and in this case, the main focus of the investigation. The data set contains the number of seizures suffered per patient during the six months prior to randomization, drug indicator, age at randomization, sex, type of epilepsy and the individual's times to first seizure post-randomization.

It is desired to model, for each individual, an event count over a fixed initial period, jointly with a survival time following a possible change in event rate. They denote by Z_{1i} be the covariate that contains all the important explanatory variables, for the individual i; Y_i was the time from randomization to the first post-randomization event, while X_i was the event count of such type of event. Then, let us consider that the underlying model is an overdispersed homogeneous Poisson process. The event count X_i over a period of time u_i is considered to be Poisson distributed with mean $\lambda_i u_i$, where λ_i depends on the covariates, and consider the overdispersion to be partly a result of the covariates.. Furthermore, they relate λ_i to Z_{1i} via a log-link, that in turn gives us a negative binomial generalized linear model (McCullagh & Nelder (1989)). A common model for overdispersed data is the Negative Binomial linear model(Greenwood and Yule (1920)), where each individual experiences events according to a Poisson process with event rate $\lambda_i \nu_i$; here, ν_i is a random term, which follows a gamma distribution as proposed in Marshall & Olkin (1990). Furthermore, they consider the overdispersed count data as distributed by a negative binomial, as proposed in Greenwood & Yule (1920).

Let us recall that the inter-event time was distributed as an Exponential with rate $\lambda_i \nu_i$. Because of the memoryless property of the exponential distribution, the variable Y_i (the time from randomization to the first post-randomization event) will also an exponential random variable. However, after the randomization they consider that the treatment acts multiplicatively on the event rate, hence, the event rate for individual i will be $\lambda_i \psi_i \nu_i$, where ψ_i captures dependence on the treatment that was applied.

They denote by Z_{2i} the treatment covariates, which contains an intercept term as well as a treatment indicator. It may also contain other explanatory variables and interaction terms. It is known that a derivation of a Pareto distribution is the Gamma mixture of exponentials, hence, if ν_i is integrated out, Y_i is Pareto with survival function

$$S(y_i|\lambda_i, \psi_i, \alpha) = (1 - \lambda_i \psi_i y_i/\alpha)^{-\alpha},$$

and then the hazard function is given by

$$h(y_i|\lambda_i, \psi_i, \alpha) = -\frac{d}{dy_i} \log\left[S(y_i|\lambda_i, \psi_i, \alpha)\right] = \frac{\alpha \lambda_i \psi_i}{\alpha + \lambda_i \psi_i y_i}.$$

Note that this form of hazard function was previously proposed in Kwong & Hutton[5].

As a summary of the previous model proposition, we have then that the joint model is specified by the equations

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

where

$$\begin{aligned}\lambda_i &= e^{\beta_i{}' z_{1i}},\\ \psi_i &= e^{\beta_2{}' z_{2i}}.\end{aligned}$$

Here, $\alpha > 0$ measures the degree of heterogeneity, the parameters β_1 and β_2 are vectors of regression coefficients, z_{1i} includes an intercept term, and z_{2i} is be parameterized to include an average tratment effect as well as a treatment contrast. The use of log-links ensures that λ_i and ψ_i are always positive.

The full log-likelihood for the data D on the n individuals is given by

$$\begin{aligned}l(\beta_1, \beta_2, \alpha|D) &= \sum_{i=1}^{n}\{\sum_{j=0}^{x_i-1}\ln(\alpha+j) + \delta_i \ln(\alpha+x_i) + x_i \ln(u_i) + \alpha\ln(\alpha) + (x_i+d_i)\ln(\lambda_i)\\ &\quad +\delta_i\ln(\psi_i) - \ln(x_i!) - (x_i+\alpha+\delta_i)\ln(\lambda_i u_i + \lambda_i\psi_i y_i + \alpha)\},\end{aligned}$$

where δ_i is a censoring indicator, taking value 1 if an event is observed during follow-up and 0 otherwise.

For a given set of data, the likelihood can be maximized in order to make inference on the parameters β_1 and β_2, using a numerical method. The observed information matrix can be then used to calculate standar errors of the parameter estimates.

While in Kwong & Hutton[5], they preferred to include an interaction between treatment and age, rather than an interaction between treatment and type of epilepsy, in Cowling et al[8] they found some evidence for an interaction between treatment and type of epilepsy. While this paper fitted three models (exponential, Weibull and Pareto), a goodness-of-fit diagnostics reveal that these distributions do not fit the data very well, mainly because they cannot model the initial steep drop in the survival function. Kwong and Hutton also fitted proportional hazards models to these data, but they concluded that an accelerated life model would fit these better.

The work from Cowling et al. gives an indication that VPS is the better treatment for patients with general onset epilepsies, while CBZ is a better treatment for patients with partial onset epilepsies, where onset is considered above the age of 20 years. Also, CBZ seems to improve the efficacy with age, whereas the reverse is true for VPS.

The advantage gained by considering this joint model for pre and post-randomization seizure counts, is the power to study better the interaction effects and using non-proportional hazards. Observe also that the diagram which describes the relationship of the proposed variables may be thought of as specifying a Bayesian model, and in this case Markov Chain Monte Carlo simulation may be used to make inferences on the parameters.

Cowling et. al. proposes a fully parametric model that considers the joint distribution of an individual's pre-randomization event count, and a single post-randomization failure time under a Poisson process framework. In contrast, the Rogers & Hutton[9] paper considers the joint modelling of a pre-randomization event count and multiple post-randomization survival times with cure rates.

Rogers and Hutton propose a methodology that models the recurrent event data from the MESS study. In this paper, the methodology is developed in order to model the treatment policy effects on first and second seizure times, and the pre-randomization history is exploited as a covariate. Also the analysis of the data was focused on the times from randomization to first post-randomization seizure, and times from first to second post-randomization seizures, with a subsequent follow-up period subjected to right censoring. In order to handle within subject effects and censoring, models were fitted with individual-specific independent and identically distributed

random effects to induce associations among gap times. This assumes that, given the random effect, the gap times for an individual are independent.

Some of the baseline covariates that were collected were age at randomization, sex, the patient's pre-randomization history, electoencephalogram (EEG). As part of the patient's history, their pre-randomization seizures are categorized as:

1. Tonic-Clonic seizures: which comprise tonic-clonic seizures only,
2. Secondary Tonic-Clonic seizures: this consist of partial seizures accompanied by second tonic-clonic seizures,
3. Generalized: which includes any types of generalized seizures, including any combinations of tonic-clonic and other generalized seizures,
4. Partial group: contains patients with simple or complex partial seizures only, and
5. Other: this group contains all the remaining seizures that could not be categorized in any of the other groups.

In Rogers & Hutton's work it is assumed that an individual i experiences seizures according to a Poisson process with rate $\lambda_i\nu_i$, where λ_i is a function of the baseline covariates, and ν_i is the frailty term. They denote by X_i the pre-randomization event count during a period u_i, T_{1i} and T_{2i} are the times for the first and second post-randomization seizure times, and set $Y_{1i} = T_{1i}$ and $Y_{2i} = T_{2i} - T_{1i}$, the gap times between post-randomization seizures. As a result of discussions with the clinicians, it was determined that the number of days at risk before randomization, u_i, would be adjusted so that, if t_i is the time of the first seizure (pre-randomization) and T_i is the time of randomization for individual i, it would take the following values

$$u_i = \max(182, |T_i - t_i|).$$

In this work, the model states that $X_i|\nu_i$ has a Poisson distribution with parameters $\lambda_i u_i \nu_i$, while the gap times follow an Exponential distribution with the same rate. Note that the unconditional density of X_i, $f_X(x_i; \lambda_i, u_i, \alpha)$, is distributed as a Negative Binomial. The randomization effect was considered, as in Cowling et al. (2006), to be multiplicative and denoted by ν_i, which is distributed as $\nu_i \sim Gamma(\alpha, 1/\alpha)$. Additionally, the post-randomization gap times Y_{1i} and Y_{2i} are both Exponentially distributed with rate $\lambda_i\psi_i\nu_i$, and the two survival times are independent given ν_i. Finally we introduce the cure rates p_{1i} and p_{2i}, which depend on κ_1 and κ_2. Here κ_i represents the rate of increase or decrease of seizures relative to the susceptible proportion in the reference group, and p_{1i} is the probability that individual i has a first seizure, and p_{2i} is the probability that individual i has a second seizure. All this can be stated in the equations that follow:

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

where

$$\begin{aligned}
\lambda_i &= \exp(\beta_1' z_{1i}), \quad \psi_i = \exp\left(\beta_2' z_{2i}\right), \\
p_{1i} &= \frac{\exp\left(\kappa_1' \omega_{1i}\right)}{1+\exp\left(\kappa_1' \omega_{1i}\right)}, \\
p_{2i} &= \frac{\exp\left(\kappa_2' \omega_{2i}\right)}{1+\exp\left(\kappa_2' \omega_{2i}\right)},
\end{aligned}$$

$(\beta_1, \beta_2, \kappa_1, \kappa_2)$ are regression paramenters, and z_{1i}, z_{2i}, ω_{1i}, ω_{2i} are vectors of covariates, not necessarily distinct. The unconditional joint distribution of the Y_{ij}, $j = 1, 2$, is a bivariate Lomax distribution, with each of the marginal Y_{ij} having a univariate Lomax distribution with shape and scale parameters α and $\alpha/\lambda_i\psi_i$ respectively. This results in an accelerated failure time model with a Lomax baseline distribution.

In order to formulate the maximum likelihood function, four different types of censoring cases must be considered:

1. Y_{1i} and Y_{2i} are both observed,
2. Y_{1i} is observed but Y_{2i} is censored,
3. Y_{1i} is censored, but exists, hence Y_{2i} is taken to be censored at zero, and
4. Y_{1i} is censored, and the individual is cured, then Y_{2i} doesn't exist.

For the purpose of indicating these cases in the likelihood function, let us define the allocation variable q_i such that

$$q_i = \begin{cases} 1 & \text{if the individual is subjected to post-randomization recurrence,} \\ 0 & \text{if the individual is immune,} \end{cases}$$

and let δ_{ji} the indicator function of the jth survival time, i.e.

$$\delta_{ji} = \begin{cases} 1 & \text{if the } jth \text{ seizure is observed,} \\ 0 & \text{if the } jth \text{ survival time is censored.} \end{cases}$$

From this, we have that the log-likelihood for the observed data D, for all of the n individuals, is given by

$$\begin{aligned}
l(\alpha, \beta_1, \beta_2, \kappa_1, \kappa_2 | D) &= \sum_{i=1}^{n} \{ \sum_{k=0}^{x_i - 1} \ln(\alpha + k) + x_i \ln(u_i) - \ln(x_i!) + \alpha \ln(\alpha) \\
&\quad + (x_i + \delta_{1i}(1 + \delta_{2i})) \ln(\lambda_i) + \delta_{1i}(1 + \delta_{2i}) \ln(\psi_i) + \delta_{1i} \ln(p_{1i}) \\
&\quad + \delta_{1i}\delta_{2i} \ln(p_{2i}) + \delta_{1i} \ln(x_i + \alpha) + \delta_{1i}\delta_{2i} \ln(x_i + \alpha + 1) \\
&\quad - \delta_{1i}\delta_{2i}(x_i + \alpha + 2) \ln(\lambda_i u_i + \lambda_i \psi_i \left(y_{1i} + y_{2i}\right) + \alpha) \\
&\quad + \left(1 - \delta_{1i}\right) q_i \ln\left(\frac{p_{1i}}{\left(\lambda_i u_i + \lambda_i \psi_i y_{1i} + \alpha\right)^{x_i + \alpha}}\right) \\
&\quad + (1 - \delta_{1i})(1 - q_i) \ln\left(\frac{1 - p_{1i}}{\left(\lambda_i u_i + \alpha\right)^{x_i + \alpha + 1}}\right) \\
&\quad + \delta_{1i}(1 - \delta_{2i}) \ln(\frac{1 - p_{2i}}{\left(\lambda_i u_i + \lambda_i \psi_i y_{1i} + \alpha\right)^{x_i + \alpha + 1}} \\
&\quad + \frac{p_{2i}}{\left(\lambda_i u_i + \lambda_i \psi_i \left(y_{1i} + y_{2i}\right) + \alpha\right)^{x_i + \alpha + 1}}) \}.
\end{aligned}$$

This log-likelihood can then be differentiated and the parameters α, β_1 and β_2 can be estimated. However, since the allocation variable q_i, for all i, is not observed, an EM algorithm was used to maximize the log-likelihood function.

An exploratory analysis revealed that clustering the types of seizures might be a natural approach to this problem. There is also evidence to suggest that the covariate effects may vary for times to first and second seizures post-randomization. To allow for this variation in the model, the paper later on considers a joint model that includes seizure rate modifiers, both at randomization and following a first post-randomization seizure. In this scenario, the joint density of Y_{1i} and Y_{2i}, conditional on the frailty term ν, would be

$$f_{Y_1,Y_2|\nu}(y_{1i}, y_{2i}|\nu_i; \lambda_i, \psi_i, p_{1i}, p_{2i}) = p_{1i}p_{2i}\left(\lambda_i\psi_i\nu_i\right)^2 \psi_{2i}e^{-\lambda_i\psi_i\nu_i(y_{1i}+\psi_{2i}y_{2i})},$$

where in this case, $\psi_{1i} = \exp(\beta_2' z_{2i})$, $\psi_{2i} = \exp(\beta_3' z_{3i})$ and z_{2i}, z_{3i} are vectors of covariates, not necessarily distinct. From this change, we have now that the new log-likelihood function for the data D is given by

$$\begin{aligned}
l(\alpha, \beta_1, \beta_2, \beta_3, \kappa_1, \kappa_2|D) \;=\; & \sum_{i=1}^{n}\{\sum_{k=0}^{x_i-1}\ln(\alpha+k) + x_i\ln(u_i) - \ln(x_i!) + \alpha\ln(\alpha) \\
& +(x_i + \delta_{1i}(1+\delta_{2i}))\ln(\lambda_i) + \delta_{1i}(1+\delta_{2i})\ln(\psi_{1i}) + \delta_{1i}\delta_{2i}\ln(\psi_{2i}) \\
& +\delta_{1i}\ln(p_{1i}) + \delta_{1i}\delta_{2i}\ln(p_{2i}) + \delta_{1i}\ln(x_i+\alpha) + \delta_{1i}\delta_{2i}\ln(x_i+\alpha+1) \\
& -\delta_{1i}\delta_{2i}(x_i+\alpha+2)\ln(\lambda_i u_i + \lambda_i\psi_{1i}y_{1i} + \lambda_i\psi_{1i}\psi_{2i}y_{2i} + \alpha) \\
& +(1-\delta_{1i})\,q_i\ln\left(\frac{p_{1i}}{(\lambda_i u_i + \lambda_i\psi_{1i}y_{1i} + \alpha)^{x_i+\alpha}}\right) \\
& +(1-\delta_{1i})(1-q_i)\ln\left(\frac{1-p_{1i}}{(\lambda_i u_i + \alpha)^{x_i+\alpha}}\right) \\
& +\delta_{1i}(1-\delta_{2i})\ln(\frac{1-p_{2i}}{(\lambda_i u_i + \lambda_i\psi_{1i}y_{1i} + \alpha)^{x_i+\alpha+1}} \\
& +\frac{p_{2i}}{(\lambda_i u_i + \lambda_i\psi_{1i}\left(y_{1i}+\psi_{2i}y_{2i}\right) + \alpha)^{x_i+\alpha+1}})\}
\end{aligned}$$

In order to compare and interpret the parameter estimates from this model, the paper uses a reference group, which is defined as individuals presenting partial seizures pre-randomization, a normal EEG and randomized to deferred treatment. The functions ψ_{1i} and ψ_{2i} contain parameter estimates relative to event rates on the gap times Y_{1i} and Y_{2i} respectively. A positive regression coefficient would indicate an increase in the seizure rate relative to the reference group, with an analogous interpretation if the regression coefficient is negative. Similarly, a positive estimate of κ would indicate an increase in the susceptible proportion of patients, relative to the reference group. Interpretation of the remaining parameters is rather straight forward.

Stepwise backwards elimination concluded that the optimal joint model included treatment policy, seizure type, EEG outcome and their interactions for the change in seizure rate at randomization and following a first seizure post-randomization.

This last model was fitted also for the same MESS data by Rogers et al[10], but considering the two treatment policies applied to the partitioned population. It was separated into the groups: tonic-clonic only, tonic-clonic with partial seizures, tonic-clonic with generalized seizures, partial seizures, myoclonic and absence, and all the remaining types of seizures were grouped under the name "other seizures ". From this paper, it was concluded that immediate treatment was considerably relevant for the expected time of the first seizure for tonic-clonic and secondary tonic-clonic groups, although immediate treatment has less influence in the seizure rates for the partial group.

It is brought to attention that MESS was not a population based study but rather it was a trial comparing the two treatment policies. Since the patients in the study and their clinicians were uncertain about the need of an AED, the data from MESS may not immediately generalize to patients for whom there is no uncertanty about the need of treatment.

Chapter 3 Future work

During the course of research through previous works in the epilepsy framework, specifically for the MESS study, questions on the existing models naturally arise, as well as questions on possible new models.

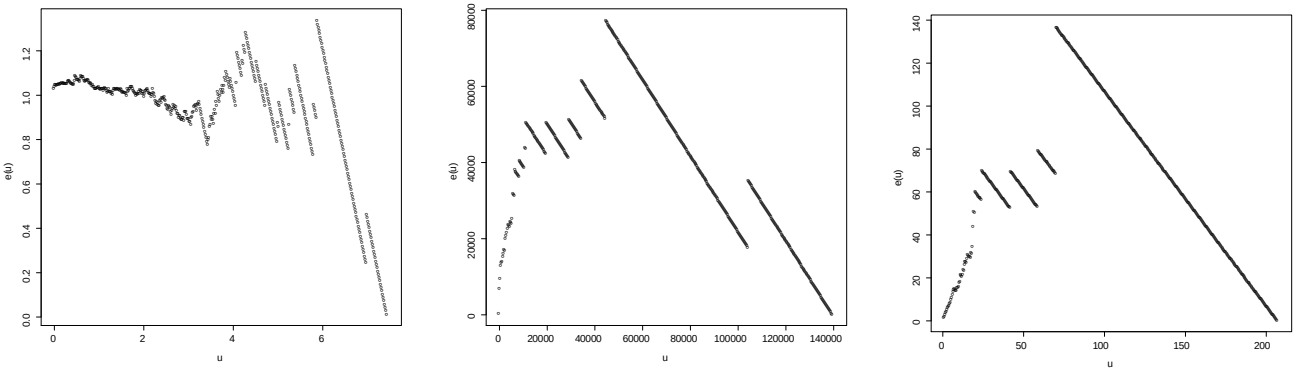
The first problem is based on the papers by Kim et al[2] and Rogers & Hutton[9]. While in the first work, the proposed model is based on a Cox regression model with prognostic indexes, it would be of special interest to implement the Rogers & Hutton's model, which considers individual joint distributions for pre-randomization and post-randomization seizures, along with the Kim et al.'s prognostic index as a covariate, which splits the population into three categories: low, medium and high risk groups. After fitting and comparing this model, we would desire to propose a model in which we propose new prognostic indexes, and finally compare these new indexes with the ones proposed by Kim et al.

A second interesting project arises from the book by Mikosch[11], which introduces the notion of the tail thickness of a distribution, through the mean excess function.

Definition 2 *let $Y > 0$ be a random variable with $E(Y) < \infty$, distribution F and $x_l = \inf\{x : F(x) > 0\}$ and $x_r = \sup\{x : F(x) < 1\}$. Then its* ***mean excess*** *or* ***mean residual life function*** *is given by*

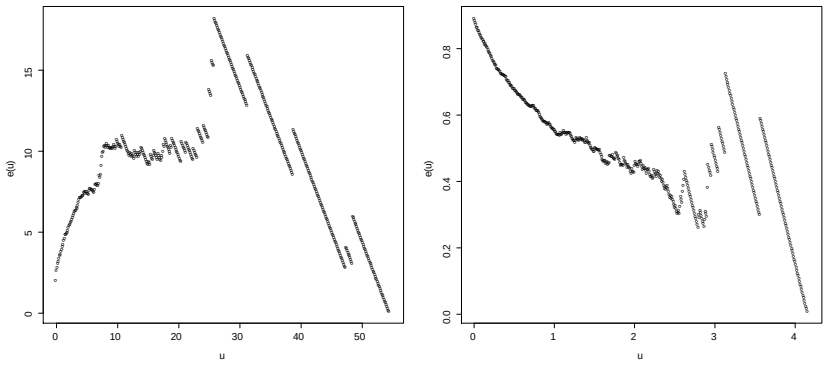
$$e_F(u) = E(Y - u | Y > u),\ u \in (x_l, x_r).$$

In reliability or medical context, this function is also referred to as the *mean residual life function*. This is a useful tool for observing the thickness of distribution tails, as seen in the following examples of 1000 simulated data (Figure 3.1) where the Log-normal distribution is such that $\log X \sim N(0, 4)$, and a Weibull distribution (Figure 3.2), considering the cases in which the shape parameter is $\tau < 1$ and $\tau > 1$.



Standard Exponential Log-normal Pareto with tail index 1.7

Figure 3.1 (3.1)



Weibull(0.5,1) Weibull(1.5,1) (3.2)

Figure 3.2

In an epileptic seizure framework, this mean residual life function can be interpreted as the remaining expected time of a patient being seizure-free, given that he has been seizure-free for u =1 year. For the time to first seizures randomization from the MESS data, the mean residual function is shown in Figure 3.3. Observe that the decay in the mean excess life function resembles that of a Weibull distribution with shape parameter $\tau > 1$ (Figure 3.2). This suggests that a possible work could be to develop, from the model from Rogers & Hutton, a more general model. Let us remember that Rogers & Hutton's model proposes a negative binomial process as the underlying process of the seizure recurrences. The proposed next step would be to model the seizure recurrences with a renewal process, which could possibly consider a Weibull distribution for the inter-event times.

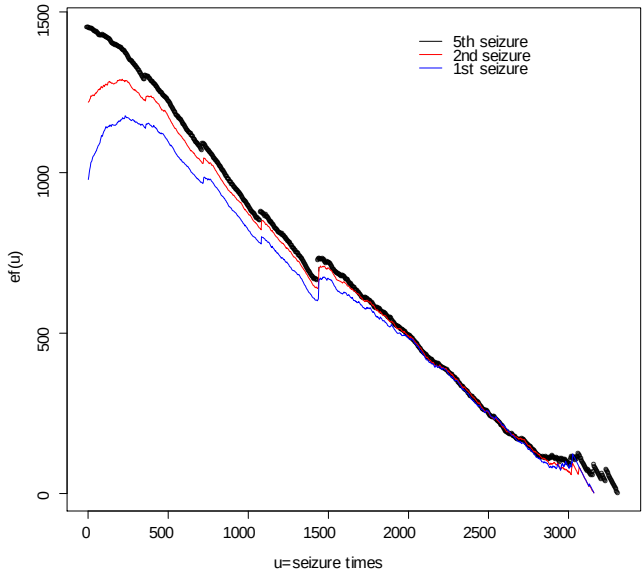
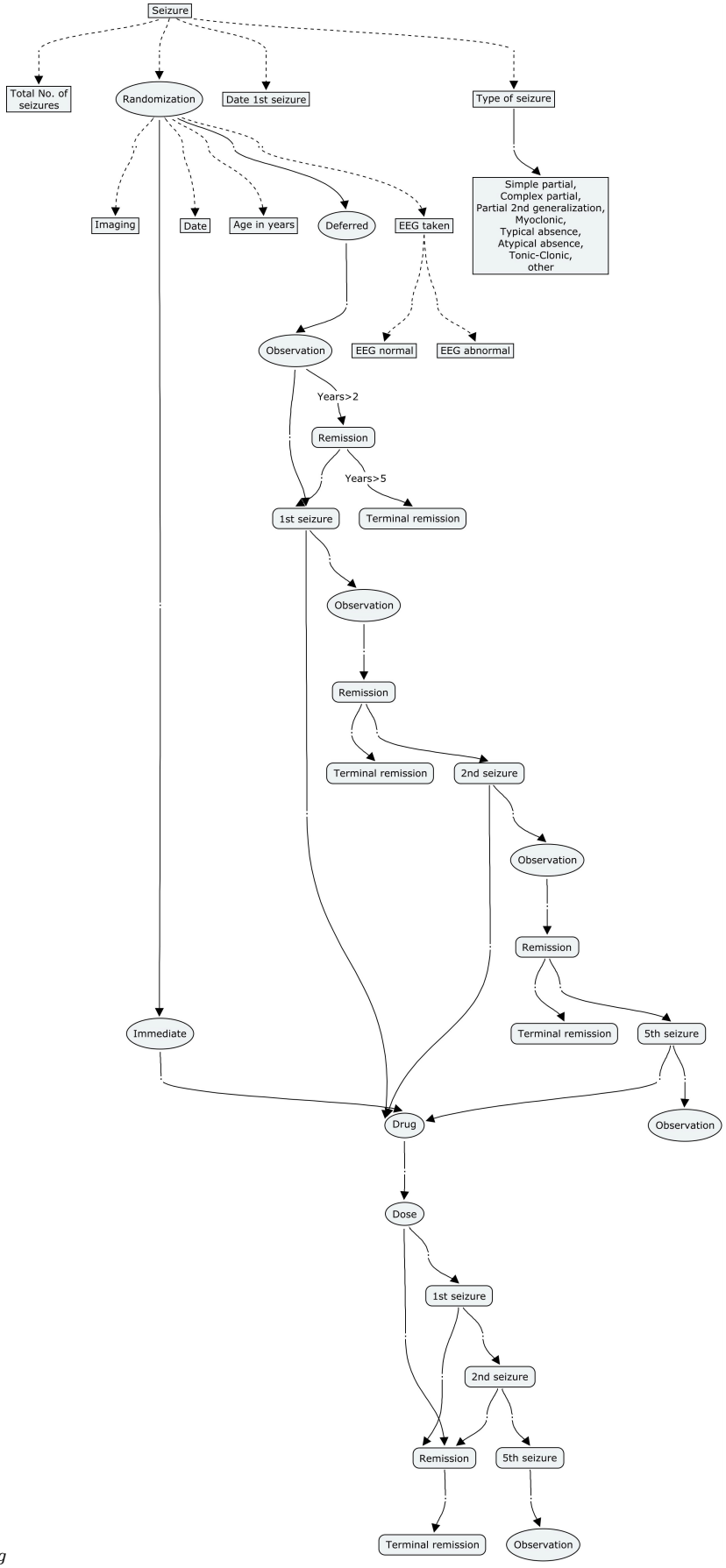


Figure 3.3 Mean excess life function plots for first, second and fifth post-randomization seizures. (3.3)

A final interest for future work evolves from Cowling et al's idea of approaching the model

from a Bayesian point of view. The process of decision making for the patients involved in the MESS study could be represented through a preliminary form of a decision tree. Indeed, consider for each individual the option of either being randomly located to either the deferred or the immediate policy, having hence a decision to make depending on whether or not more recurrences or treatment second effects manifest in the patient. After observing the patient's behaviour under the policy he has been allocated to, and since MESS is a pragmatic study, the physician might also decide to change the dose or even the type of AED of the patient. Such an approach could provide a new point of view on the subject, to propose a model and compare with the existing models. In Figure 3.4 we show an example of the decission tree that could represent the decission making process for each patient in the study. Observe that in the tree, the words in ovals, rounded squares and sharp squares represent the decision to be made, the outcomes and the data collected respectivelly. The dotted-lined arrows indicate the moment at which the data was collected.

decision tree



15.*jpg*

Figure 3.4 Preliminary decision tree for MESS study. (3.4)

Chapter 4 Appendix

4.1 Survival analysis models

For convenience, some basic definitions will be presented in this section. Let T denote the random variable of failure times under study. Assuming that T is a continuous random variable, the *survival function* is defined as

$$S(t) = P(T > t) = \int_t^\infty p(s)ds,$$

where $p(\cdot)$ denotes the probability density function of T. In industrial and financial context, $S(t)$ may also be denoted by $R(t)$. Furthermore, the function

$$h(t) = \lim_{dt \to 0} \frac{P(t \leq T < t + dt | T \geq t)}{dt}, \ t > 0$$

is defined as the *hazard function*, and describes the instantaneous risk of an event. Under these two definitions, it can be seen that a natural relationship between the two of them is given by

$$S(t) = \exp(-H(t)) = \exp\left\{ - \int_0^t h(s)ds \right\},$$

where $H(\cdot)$ is called the *cumulative hazard function*.

If we consider a set of covariates $x = (x_1, x_2, ..., x_p)^T$, then the *general proportional hazards model* is defined as

$$h(t|x_1, ..., x_p) = h_0(t)g(x_1, ..., x_p),$$

where $h_0(\cdot)$ is the *underlying hazard function*, also called the *base-line hazard function*, and $g(x_1, ..., x_p)$ represents the effect from the covariates. As a more particular model, the Cox's proportional hazards model assumes that $g(x) = \exp\left(\sum_{j=1}^p b_j x_j\right) = \exp\left(b'x\right)$, where $b = (b_1, ..., b_p)$ are the coefficients of the covariates. This means that, under Cox's proportional hazards, the model is of the form

$$h(t|x) = h_0(t) \exp\left(b'x\right).$$

Another popular survival model is the accelerated failure time model, also known as the accelerated life model. It considers once more that, if b and x are considered as in the previous definition, then the model is defined as

$$h(t) = \exp(b'x)h_0(t \exp(b't)),$$

where $h_0(\cdot)$ is the base-line hazard function.

4.2 Parameter orthogonalization

The importance of parameter orthogonalization, and the means to obtain orthogonality, were developed and discussed in the paper by Cox & Reid[3], in 1987. The main goal of this paper is to address parametric probability models, which in turn are in term of a set of parameters θ, consisting of a subset of parameters of interest θ_1, and a subset of nuisance parameters θ_2. Although it is mentioned that both sets of parameters could be orthogonalized from each other,

the paper focuses in the direct problem when θ_1 is a scalar, and θ_2 is the set of remaining nuisance parameters.

Previously, the maximum likelihood estimator (MLE) $\hat{\theta}_2$ of θ_2 is used in the profile likelihood, causing then that such profile likelihood becomes a function only of θ_1, as desired. The difficulty with this method lies in the fact that, when the number of nuisance parameters is high, the estimators obtained can be inefficient or even inconsistent. For this and severeal other reasons mentioned later on, Cox & Reid (1986) proposed using the conditional likelihood function, given the MLE of the orthogonalized parameters, θ_2. Let us then consider the following notation: let y be a $n \times 1$ vector of observations from a random variable Y, which has a pdf $f_Y(y;\theta)$, and where θ is a $1 \times p$ vector of unknown parameters. Then, let $L(\theta; y)$ and $l(\theta; y)$ denote the Maximum likelihood function and the log-likelihood function of θ respectively. From this notation, we can now proceed to the following definition.

Definition 3 *If $\theta = (\theta_1, \theta_2)$ where θ_1 and θ_2 partition θ, and $i_{\theta_i\theta_j}$ corresponds to the ij^{th} entry of the information matrix, then we define θ_1 to be orthogonal to θ_2 if*

$$i_{\theta_s\theta_t} = \frac{1}{n}E\left[\frac{dl(\theta;y)}{d\theta_s}\frac{dl(\theta;y)}{d\theta_t};\theta\right] = \frac{1}{n}E\left[-\frac{d^2l(\theta;y)}{d\theta_s d\theta_t};\theta\right] = 0$$

for all $\theta_s \in \theta_1$ and for all $\theta_t \in \theta_2$.

This definition can be extended to the case where more than two sets of parameters are orthogonal.

Some interesting properties inherent to orthogonality are that, if we suppose that $\theta = (\psi, \lambda)$ and ψ and λ are orthogonal, then:

1. if $\hat{\psi}$ and $\hat{\lambda}$ are the Maximum likelihood estimators of ψ and λ respectively, then they are asymptotically independent.
2. the asymptotic standard error of ψ is equal independently of either λ's value being known or unknown.
3. there might be some simplifications in the numerical calculation of $(\hat{\psi}, \hat{\lambda})$.
4. if we denote by $\hat{\psi}_\lambda = \hat{\psi}(\lambda)$ the ML estimate of ψ when λ is given, we also have that if λ and ψ are orthogonal, then $\hat{\psi}_\lambda$ varies only slowly with λ.
5. if $\hat{\psi}_\lambda = \hat{\psi}$ for all λ, then λ and ψ are orthogonal.

4.2.1 Construction of orthogonal parameters

Consider a parameter of interest ψ and a vector of nuisance parameters $\lambda = (\lambda_1, \lambda_2, ..., \lambda_q)$, which we want to orthogonalize from one another. Let ϕ be a function, such that $\phi_i = \phi_1(\psi, \lambda)$ for $i = 1, 2, ..., q$, and suppose the likelihood function is in terms of the parameters $(\psi, \phi_1, ..., \phi_q)$, in other words,

$$l(\psi, \lambda) = l^*(\psi, \phi_1, \psi_2, ..., \psi_q),$$

then the first and second derivatives of the likelihood function have the form

$$\frac{dl(\psi,\lambda)}{d\psi} = \frac{dl^*(\psi,\phi)}{d\psi} + \sum_{r=1}^{q} \frac{dl^*(\psi,\phi)}{d\phi_r}\frac{d\phi_r}{d\psi},$$

hence we obtain

$$\frac{d^2l(\psi,\lambda)}{d\psi d\lambda_t} = \sum_{s=1}^{q} \frac{d^2l^*(\psi,\phi)}{d\psi d\phi_s}\frac{d\phi_s}{d\lambda_t} + \sum_{r=1}^{q}\sum_{s=1}^{q} \frac{d^2l^*(\psi,\phi)}{d\phi_r d\phi_s}\frac{d\phi_s}{d\lambda_t}\frac{d\phi_r}{d\psi} + \sum_{r=1}^{q} \frac{dl^*(\psi,\phi)}{d\phi_r}\frac{d^2\phi_r}{d\psi d\lambda_t}. \tag{4.1}$$

By taking the expectation of equation 4.1, we have

$$\begin{aligned} E\left[\frac{d^2l(\psi,\lambda)}{d\psi d\lambda_t}\right] &= \sum_{s=1}^{q} E\left[\frac{d^2l^*(\psi,\phi)}{d\psi d\phi}\right]\frac{d\phi_s}{d\lambda_t} + \sum_{r=1}^{q}\sum_{s=1}^{q} E\left[\frac{d^2l^*(\psi,\phi)}{d\phi_r d\phi_s}\right]\frac{d\phi_s}{d\lambda_t}\frac{d\phi_r}{d\psi} + \sum_{r=1}^{q} E\left[\frac{dl^*(\psi,\phi)}{d\phi_r}\frac{d^2\phi_r}{d\psi d\lambda_t}\right] \\ &= \sum_{s=1}^{q} \frac{d\phi_s}{d\lambda_t} i^*{}_{\psi\phi_s} + \sum_{s=1}^{q}\sum_{r=1}^{q} \frac{d\phi_r}{d\psi}\frac{d\phi_s}{d\lambda_t} i^*_{\phi_r\phi_s}, \end{aligned}$$

where $i^*_{\psi\phi_s} = E\left[\frac{d^2l^*(\psi,\phi)}{d\psi d\phi_s}\right]$, $i^*_{\phi_r\phi_s} = e\left[\frac{d^2l^*(\psi,\phi)}{d\phi_r d\phi_s}\right]$ and $\sum_{r=1}^{q} E\left[\frac{dl^*(\psi,\phi)}{d\phi_r}\frac{d^2\phi_r}{d\psi d\lambda_t}\right]$ is equal to zero.
By definition, ψ and λ are orthogonal if

$$E\left[\frac{d^2l(\psi,\lambda)}{d\psi d\lambda}\right] = 0;$$

which leads to the equation

$$0 = E\left[\frac{d^2l(\psi,\lambda)}{d\psi d\lambda}\right] = \sum_{s=1}^{q} \frac{d\phi_s}{d\lambda_t}\left(i^*_{\psi\phi_s} + \sum_{r=1}^{q} i^*_{\phi_r\phi_s}\frac{d\phi_r}{d\psi}\right),$$

but since the Jacobian of the transformation from (ψ,ϕ) to (ψ,λ) cannot be zero, then $\frac{d\phi_s}{d\lambda_t}$ can't be zero. This finally leads us to the equation

$$i^*_{\psi\phi_s} + \sum_{r=1}^{q} i^*_{\phi_r\phi_s}\frac{d\phi_r}{d\psi} = 0\ ,$$

and thus the orthogonalizing equation, which determines the dependence between ϕ and ψ, is given by

$$\sum_{r=1}^{q} i^*_{\phi_r\phi_s}\frac{d\phi_r}{d\psi} = -i^*_{\psi\phi_s}.$$

Observe that, if ψ had a higher dimension, there is no guarantee, for instance in the case that $\psi = (\psi_1,\psi_2)$, that the condition $\frac{d^2\phi_s}{d\psi_1 d\psi_2} = \frac{d^2\phi_s}{d\psi_2 d\psi_1}$ holds. For this reason, global orthogonality of parameters cannot always be obtained.

4.3 Deviations from correctly specified models

Let us remember that the model proposed in Hutton & Solomon[6] is a mixture of an accelerated life (AL) model, and a Cox's proportional hazards (PH) model:

$$m(t) = \left\{ g_0(t)\, e^{\beta' z} G_0(t)^{\exp(\beta' z)-1} \right\}^{\psi} \left\{ f_0\left(te^{\gamma' z}\right) e^{\gamma' z} \right\}^{1-\psi},$$

where the survivor functions $F_0\left(te^{\gamma' z}\right)$ and $G_0(t)^{\exp(\beta' z)}$ correspond to the accelerated life model and the proportional hazards model respectively. F_0 and G_0 are the base-line survivor functions with corresponding densities f_0 and g_0 . Here z is the vector of fixed covariates, and let β and γ be the associated vectors of unknown regression parameters for the proportional hazards and accelerated life families of models. The interest lies in finding expressions for the parameters of interest, β and γ, while considering ψ as a nuisance parameter.

Let us remember from the Cox & Reid (1987) paper, that the orthogonalizing equation between the parameters (ψ, ϕ) is of the form

$$\sum_{r=1}^{q} i^*_{\phi_r \phi_s} \frac{d\phi_r}{d\psi} = -i^*_{\psi \phi_s},$$

where $i^*_{\psi\phi_s} = E\left[\frac{d^2 l^*(\psi,\phi)}{d\psi d\phi_s}\right]$, $i^*_{\phi_r\phi_s} = e\left[\frac{d^2 l^*(\psi,\phi)}{d\phi_r d\phi_s}\right]$ and l^* is the log-likelihood function corresponding to ψ and ϕ. In the case of this problem the exact expressions of this orthogonalizing equations are not obtainable, hence the mixture function $m(t)$ is normalized and equated to the likelihood function

$$L(\psi, \beta, \gamma; \alpha, t, z) = \frac{m(t)}{\int_0^\infty m(t)dt};$$

where α, which is assumed as known, are the parameters of the base-line survival distributions.

4.3.1 Accelerated Life model true

Supose that the true model is the Accelerated Life model, which implies that $\psi = 0$, and that we have the independent failure times $t = (t_1, t_2, ..., t_n)$ and z_i a vector of $1 \times q$ fixed covariates for the i^{th} individual. Then, in order to calculate $\int_0^\infty m(t)dt$, we aproximate

$$m(t)e^{-\psi\beta' z} e^{-(1-\psi)\gamma' z}$$

with a first-order Taylor series about $(\beta, \gamma, \psi) = 0$, and find that

$$\int_0^\infty m(t)dt \approx (1 - \gamma' z + c_1\psi) e^{\psi\beta' z} e^{-(1-\psi)\gamma' z},$$

where $c_1 = E_{f_0}\left[\log\left\{g_0(t)/f_0(t)\right\}\right]$. Observe that this requires the assumption that β and γ are small.

From this, we obtain that the likelihood function is

$$L(\psi, \beta, \gamma; \alpha, t, z) \approx \prod_{i=1}^{n} \frac{m(t_i)}{\int_0^\infty m(s)ds} = \prod_{i=1}^{n} \frac{\left[g_0(t_i) G_0(t_i)^{\exp(\beta' z_i)-1}\right]^{\psi} \left[f_0(t_i e^{\gamma z_i})\right]^{(1-\psi)}}{1 - \gamma z_i + c_1\psi};$$

which, with some calculations leads to the log-likelihood expression

$$l = l(\psi, \beta, \gamma; \alpha, t, z) \approx \psi \sum_{i=1}^{n} \left[\log \{g_0(t_i)\} + (e^{\beta' z_i} - 1) \log \{G_0(t_i)\} \right] + (1-\psi) \sum_{i=1}^{n} \log[f_o(t_i e^{\gamma' z_i})] - \sum_{i=1}^{n} \log(1 - \gamma' z_i + c_1 \psi).$$

From this equation, the first and second derivatives of $l(\psi, \beta, \gamma; \alpha, t, z)$ are found with respect to β, γ and ψ, but the expectations of the second derivatives are analitically intractable, and therefore, they are once again approached by first-order Taylor aproximations. By doing this, two sets of orthogonalizing equations are found, but one set is non-informative on the transformation between γ and β, since the corresponding normalizing constant does not include β. In order to simplify the expressions, the explanatory variables are orthonormalized, so that

$$\sum_{i=1}^{n} z_{is} = 0, \quad \sum_{i=1}^{n} z_{is} z_{ir} = \delta_{rs}, \quad r, s = 1, ..., q,$$

where q is the dimension of the explanatory variables, and δ_{rs} is the Kronecker delta. Finally, from this simplifications the orthogonalization equation is expressed as

$$-c_2 \psi \frac{\delta \beta_s}{\delta \psi} = (c_2 + c_4)\gamma_s + c_2 \beta_s;$$

where

$$c_2 = E_{f_0}[\log G_0(T)] \quad \text{and} \quad c_4 = E_{f_0}\left[T\dot{y}(T) \log G_0(T)\right].$$

A solution of such partial differential equation is

$$\beta_s = -\gamma_s \left(1 + \frac{c_4}{c_2}\right), \tag{4.2}$$

and from this it can be observed that when the proportional hazards assumption is used, when the true model is a life accelerated model, the coefficients are proportional to the first degree. Ultimatly, the expression can be further simplified to

$$\frac{\beta_s}{\beta_1} = \frac{\gamma_s}{\gamma_1} \text{ for } s = 2, 3, ..., 1,$$

which means that the ratios of regression coefficients are consistent. As a special case, when the true model is an accelerated life model and the analysis is performed under a proportional hazards model, but the base-line distribution is correct, (or so to speak, $f_0 = g_0$), then the regression coefficients have an even more simple expression,

$$\beta_s = -\gamma_s(1 - c_4).$$

4.3.2 Cox Proportional Hazards model true

For the case in which the true model is proportional hazards, but for the analysis it is assumed accelerated life model, the procedure is exactly the same as seen for the converse case. Once again, β and γ are assumed to be small, and after following the same steps around the value $\psi = 1$, we reach the orthogonalization equation

$$d_4 (\psi - 1) \frac{\delta \gamma_s}{\delta \psi} = d_4 \gamma_s + d_6 \beta_s,$$

where

$$d_4 = E_{g_0}\left[T\dot{y}(T)\log\left\{g_0(T)/f_0(T)\right\}\right] \quad \text{and} \quad d_6 = E_{g_0}\left[T\dot{y}(T)\left\{1+\log G_0(T)\right\}\right].$$

A solution of such partial differential equation is given by

$$\gamma_s = -\frac{\beta_s d_6}{d_4}; \tag{4.3}$$

observe that this implies that the regression coefficients are linearly related under model misspecification. Furthermore, this relationship can be re-expressed as

$$\frac{\gamma_s}{\gamma_1} = \frac{\beta_s}{\beta_1} \text{ for } s = 2, 3, ..., q,$$

meaning that the ratios of the estimated regression parameters are consistent.

4.3.2.1 Inclusion of random censoring

Under the assumption of now having censoring in the data, the likelihood function is modified in order to include the censoring. The new orthogonalization equations are approximated by first-order Taylor series, with ψ, β and γ near 0. After calculating the first and second derivatives of the log-likelihood function, it is found that the orthogonalizing equations coincide with the orthogonalizing equations corresponding to the uncensored case. An analogous result is reached when it is considered $\psi = 1$.

It is concluded that the proportional properties found before still hold when right-censoring is included in the model.

4.3.3 Importance of Model selection

Observe that if the orthogonalizing equations (4.2) and (4.3) hold simultaneously, then it follows that

$$1 + \frac{c_4}{c_2} = \frac{d_4}{d_6}.$$

Furthermore, if the base-line distributions are the same, $g_0 = f_0$, then it $c_2 = -1$, $d_4 = c_5 - 1$ and $d_6 = c_4 - 1$, where $c_5 = E_{f_0}\left\{T^2 y''(T)\right\}$. Under such a case, the resulting expression

$$P = \frac{(1-c_4)^2}{1-c_5}$$

is proposed as a measure of how far are the hazards from proportionality. A value of P far from 1 would indicate a clearer departure from proportionality for the chosen model. This formula is proposed and named the proportionality factor.

4.4 Hutton & Monaghan[7] results

In this paper, the main objective is to study the properties of misspecification of two sorts, given that Cox's proportional hazards and accelerated life models are to be compared. As it was investigated in previous works (Hutton & Solomon), given that covariate coefficients are assumed small and that baseline parameters are considered known, and if a Cox's proportional hazards is misspecified as an accelerated life model, then the regression coefficients are consistent between the true and the assumed models. The result is shown to hold for the converse situation, in which the assumed model is a proportional hazards, and the true model is an accelerated life one. When orthonormalizing covariates, it was found that the importance of the covariates is preserved, although in a following work, Kwong & Hutton (2002), they suggest that orthonormalization of the covariates is not essential. An important difference from previous works, is that this paper does not assume that the regression coefficients are "small".

Three categories of misspecification are considered for this paper: when the base-line distribution is false, but the covariate effects are well specified, when the covariate effects are false but the right base-line distribution is considered, and finally when both covariate effects and base-line distribution are misspecified. The three distributions considered here for the base-line distribution are: log-normal, log-logistic and Weibull distributions.

Suppose there are two treatments to be compared, which are the standard ($x = 0$) and the new ($x = 1$) treatment. If the true model is given by $f(t;\alpha)$ while the fitted model is $g(t;\beta)$, then for a sample of size n, $t_{1,...,t_n}$, the likelihood function will be

$$L_g = \prod_{i=1}^{n} g(t_i;\beta).$$

As quoted from Cox (1961) then the maximum likelihood estimator $\hat{\beta}$ under the misspecified model, will be asymtotically distributed as $N(\beta_\alpha, \frac{1}{n}C(\beta_\alpha))$, where β_α is the solution of the set of equations

$$E_f\left[\frac{\delta}{\delta\beta_j}\log g(T;\beta)\right] = 0 \text{ for } j = 1, ..., p$$

and

$$C(\beta_\alpha) = A^{-1}BA^{-1}$$

where A is the Fisher information matrix with $(j,k)^{th}$ element $-E\left[\frac{\delta^2}{\delta\beta_j\beta_k}\log L\right]$; and B has $(j,k)^{th}$ component $B(\alpha)_{jk} = E_f\left[\frac{\delta\log L}{\delta\beta_j}\frac{\delta\log L}{\delta\beta_k}\right]$. In order to calculate such derivatives, the log-likelihood was split into the failure times belonging to the standard an the new treatment. The expectation was taken on both sides of the equation and the asymptotic average was estimated for each treatment.

$$E_f\left[\log g(T)\right] = a_0 E_{f_0}\left[\log g_0(T)\right] + a_1 E_{f_1}\left[\log g_1(T)\right],$$

where g_0 and f_0 are the assumed and true distributions under the standard treatment, g_1 and f_1 are the assumed and true distributions under the new treatment and a_0 and a_1 are the proportion of observations for the standard and new treatmens respectively. In order to perform the desired asymptotic average estimation, the resulting equations were solved analitically if possible, but otherwise the Newton-Raphson procedure was carried out.

After all different types of misspecifications are considered and the parameters are estimated, a large series of detailed results are stated for every base-line distribution considered. Among the most important results found by this study are:

- The shape and regression parameters are biased when proportional hazards models are fitted to accelerated life models.
- For proportional hazards models, the direction of the bias also depends on the covariate distribution and effect size.
- For orthonormalized covariates, the regression coefficients are proportional (this was originally obtained in Hutton and Solomon (1997)).
- Consistently with Hutton & Solomon (1997), it was found that the effect of misspecification decreases as the centre of density of survival times move away from zero, which means that there are not too many early failures.
- When the fitted model is misspecified, the bias in the lower and upper percentiles is usually more substancial than the bias in the median.
- Fully parametric models give narrower confidence limits for quartiles than non-parametric methods, but they are subjected to the bias of assuming an incorrect model.
- The accelerated life model is more robust to misspecification because of its log-linear form.

Bibliography

[1] Marson A., Jacoby A., Johnson A., Kim L., Gamble C., Chadwick D., "Immediate versus deferred antiepileptic drug treatment for early epilepsy and single seizures: a randomised controlled trial", *Lancet* 2005; **365**, *pp.*2007-13.

[2] Lois G. Kim, Tony L. Johnson, Anthony G. Marson, David W. Chadwick, "Prediction of risk of seizure recurrence after a single seizure and early epilepsy: further results from the MESS trial", *Lancet* 2006; **5**, *pp.*317-22.

[3] D. R. Cox, N. Reid, "Parameter Orthogonality and Approximate Conditional Inference", *Journal of Royal Statistical Society*, **B** (1987), **49**,No. 1, *pp.* 1-39.

[4] D. R. Cox, "Regression models and life-tables (with discussion)", *Journal of the Royal Statistical Society*, 1972, **B**, *34*, *pp.*187-220.

[5] G. P. S. Kwong, J. L. Hutton, "Choice of parametric models in survival analysis: applications to monotherapy for epilepsy and cerebral palsy", *Applied Statistics*, 2003, **52**, *Part 2*, *pp.* 153-168.

[6] J. L. Hutton, P. J. Solomon, "Orthogonality in Mixed Regression Models for Survival Data", *Journal of Royal Statistical Society*, **B** (1997), **59**, No.1, *pp.*125-136.

[7] J. L. Hutton, P. F. Monaghan, "Choice of parametric Accelerated Life and Proportional Hazards Models for Survival Data: Asymptotic results", *Life Data Analysis*, 2002 Kluwer Academic Publishers, 2002, **8**, *pp.*375-393.

[8] B. J. Cowling, J. L. Hutton, J. E. H. Shaw, "Joint modelling of event counts and survival times", *Applied Statistics*, 2006, **55**, *Part* 1, *pp.*31-39.

[9] J. K. Rogers, J. L. Hutton, "Joint modelling of pre-randomization event counts and multiple randomization survival times with cure rates: application to data for early epilepsy and single seizures", *Journal of Applied Statistics*, 2012, **3**, *pp.*546-562.

[10] J. K. Rogers, J. Hutton, A. G. Marson and D. W. Chadwick, "Assessing the risk of subsequent tonic-clonic seizures in patients with a history of simple or complex partial seizures", J*ournal of Neurology, Neurosurgery and Psychiatry*, 2012, **83**, *pp.*803-809.

[11] Thomas Mikosch, "Non-life Insurance Mathematics, An introduction with the Poisson Process", Ed. Springer, 2nd. Edition (2009).

[12] W. N. Venables, B. D. Ripley, "Modern Applied Statistics with S-PLUS", Ed. Springer, 3rd. Edition (1999).

[13] D. Collett, "Modelling Survival Data in Medical Research", Ed. Chapman & Hall, 1st. Edition (1994).

[14] Lee Wang, "Statistical Methods for Survival Data Analysis", Ed. Wiley Interscience, 3rd. Edition (2003).