2010
2011

FACULTY OF SCIENCES

Master of Statistics: Biostatistics

Masterproef

Do we need frailty modeling in the development of prognostic models?

Promotor :
Prof. dr. Roel BRAEKERS

Promotor :
Dr. YVONNE VERGOUWE
Prof. EWOUT STEYERBERG

Sammy Chebon

Master Thesis nominated to obtain the degree of Master of Statistics , specialization Biostatistics

Maastricht University

universiteit hasselt
UNIVERSITEIT VAN DE TOEKOMST

De transnationale Universiteit Limburg is een uniek samenwerkingsverband van twee universiteiten in twee landen: de Universiteit Hasselt en Maastricht University

universiteit hasselt
UNIVERSITEIT VAN DE TOEKOMST

Maastricht University

Universiteit Hasselt | Campus Diepenbeek | Agoralaan Gebouw D | BE-3590 Diepenbeek
Universiteit Hasselt | Campus Hasselt | Martelarenlaan 42 | BE-3500 Hasselt

2010
2011

FACULTY OF SCIENCES

Master of Statistics: Biostatistics

Masterproef

Do we need frailty modeling in the development of prognostic models?

Promotor :
Prof. dr. Roel BRAEKERS

Promotor :
Dr. YVONNE VERGOUWE
Prof. EWOUT STEYERBERG

Sammy Chebon

Master Thesis nominated to obtain the degree of Master of Statistics , specialization Biostatistics

Maastricht University

universiteit hasselt
UNIVERSITEIT VAN DE TOEKOMST

Abstract

Prognostic models are developed to assist in clinical decision making particularly in regards to patient treatment and follow-up time. The Cox proportional hazards model is the most extensively used in analysis. As such almost all prognostic models in survival data are developed based on Cox models. However, overreliance on the the Cox model alone is common even in the presence of clustering ignoring the intrinsic heterogeneity and dependency present in such data. This project highlight the need for prognostic model building to account for clustering with a case study in Traumatic Brain Injury(TBI). Two approaches are used to account for clustering, the fixed effect and the frailty models. By comparing model parameter estimates and standard errors for these methods against the naive Cox, we demonstrate that lack of accounting for heterogeneity may lead to invalid inference as model parameter estimates have underestimated standard errors leading to inflated type I error. Comparison of model predictive performance in terms of discrimination and calibration showed that accounting for clustering substantively improved the discriminating ability as assessed by the C-index. However no improvement was found in terms of calibration. By use of a novelle idea on calculation of between and within cluster c-indices, we demonstrate how much model discrimination ability is improved by inclusion of the random effect. Our results showed again that the random effect substantively improves the models discriminative performance. Though the use of fixed effect model corrects for the clustered nature of the data as well, we argue that the model has limitations such as assumption that the clusters used in fitting the model form the population. As a result the model cannot be generalized to the outside population.

Key Words: Prognostic Models; Cox Proportional Hazard models; Fixed effect Models; Frailty Models; Prognostic Model validation; Calibration; Discrimination.

Acknowledgments

I would like to thank my internal supervisor Prof. Roel Braekers and external supervisors, Dr. Yvonne Vergouw and Dr. Ewout Steyberg of the Erasmus University Medical Centre, Rotterdam for their critique and advice. I am also grateful for the support of Dr. Hester Lingsma, as well as the conducive working environment facilitated by the support staff at Erasmus MC during my working period there. My acknowledgement also goes to my fellow student, Fitsum Baye, who was researching on a similar area and with whom our extensive discussions gave birth to fruitful results.

I'm grateful for the immeasurable support and encouragement I received from my family members. The same goes to my friends and class-mates, the class of 2011. Lastly, I acknowledge all the professors, students and the rest of the staff at Censtat for all the support for the past two years of my studies at Uhasselt.

Most of all I would like to thank God for the smooth completion of this project and my master programme altogether.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Analysis of time to event data

Analysis of time to event data arises more often in various areas of study, for instance in engineering where reliability of equipment is judged by time to failure or in clinical trials where interest is in the time until disease healing, pain cessation, or time till death in terminal diseases such as cancer. The latter case scenario is well reflected in the common terminology: 'Survival Data Analysis' arising from mortality studies. Data arising from such study designs are characterized by censoring which renders the standard statistical methods such as linear models and generalized linear models inapplicable. Further still, the population considered is often changing as only individuals at risk at a given time point are considered, an aspect known in survival data analysis as conditioning (Duchateau and Janssen, 2008). Only right censoring is considered here as it is the most relevant and also the most common in clinical trial settings. Right censoring occurs when an event fails to be realized within the stipulated study timeframe (Wienke, 2011). The other types of censoring are left and interval censoring; and are extensively discussed in (Getachew et al, 2009).

The Kaplan Meier curve (Kaplan and Meir, 1958) is the most common non-parametric estimation approach applied in survival data analysis. In this approach, the data are summarized in a plot of survival functions from which summary statistics such as median time to event or event free rate at a time point can be estimate (Legrand, 2005). The Kaplan Meier curve assumes independence and homogeneity between observations. However, in most cases, observations in Survival data analysis exhibit a form of heterogeneity (Korosteleva, 2009). For instance in clinical trials, patients may

differ in terms of age, gender and many other characteristics. It is often of more interest to study the relationship between survival times and such covariates often measured at baseline particularly if they can be controlled. Though stratified Kaplan Meier curves can be applied, it is not capable to simultaneously take many such variables, into account. Furthermore, it is difficult to quantify the effect of the stratifying variable on the survival times (Fox, 2002). The Proportional hazards model introduced by Cox (1972) is a regression model with event time as the response variable. Unlike the Kaplan Meier approach, the model can simultaneously take several covariates, often referred to as risk factors, into account. The Cox model is perhaps the most widely used in analysis of time to event data today. The model assumes a common study population risk of death or event (hazard) function known as the baseline hazard. Individual subject information is embedded in the model through the covariates acting multiplicatively on the baseline hazard increasing or decreasing individual risk based on the prognostic information (Wienke, 2011).

1.1.1 Clustered Data Analysis

Clustered data arises from many applications, for instance in clinical trials where patients are often treated in groups in given hospital centers or countries. There has been a spiraling approval worldwide for randomized control multi-center clinical trials spanning various countries, virtually making them a gold standard in clinical research. Such trials are favored in terms of timely patient recruitment as well as ease of generalizability of the trial results based on the diversity of the study population (Buyse et al., 1984), although Yamaguchi et al. (2002) as quoted by Legrand (2005) warn of over-optimism as most of the individual centers only recruit a non-random subset of the target population. The key feature of clustered data is that outcomes from the same cluster are likely to be positively correlated. The proper analysis of clustered data requires that this correlation be taken into consideration. Ignorance of such correlation can bias the statistical inference Ying and Liu (2006). Analysis of clustered data assuming independence may lead to correct derivation of effect estimates (e.g. mean differences, odds ratios) without adjusting correlation. However, the variability of these effects would likely be biased, leading to incorrect test statistics and confidence intervals Ying and Liu (2006) and thus invalid inference.

In survival data, adjustment for the clustering is often done by either introducing the cluster as a fixed effect using dummies, or a random effect known as a frailty term. The fixed effect model assumes that the considered clusters are the all clusters in the study while the random effect model assumes that the clusters considered in the model are only a random sample from the population.

Like in Linear Mixed models and Generalized Linear mixed models (See Verbeke and Molenberghs, 2000; Molenberghs and Verbeke, 2005), frailty models accounts for the variability between the clusters which is modeled with a random effect known as frailty: a term first introduced by Vaupel et al. (1979) in univariate survival data analysis to cater for unobserved covariates. It was later first applied to multivariate survival data by Clayton (1978) as shared (within a cluster) frailty. This project focuses on the shared frailty and its application on prognostic modeling.

By allowing individuals in a cluster to share the same frailty variable, we induce positive statistical dependence between the individuals in a cluster. The parameter of the frailty distribution then acts as an association parameter since the event times become more strongly associated as the frailty variance increases. Vaupel et al. (1979) postulate that conditional on the frailty term, the survival times are independent. Because of its simple interpretation and mathematical tractability, the gamma frailty model has been extensively studied (Nielsen et al., 1992, Oakes, 1982). The distribution allows for key joint quantities such as the Kendalls tau which estimates the within cluster dependence to be estimated.

Further discussion on these two approaches are given in section 2.1.

1.2 Prognostic Modeling

Prognostic models are developed to assist in clinical decision making particularly in regards to patient treatment and follow-up time Taktka et al. (2009). The models investigate the relationship between patient characteristics and outcome Legrand et al. (2009). They are often of regression form. Factors found to be univariately or jointly associated with the response are known as prognostic factors Harrell et al. (1996)

A key aspect of the prognostic models is the predictive performance of the model thus validation plays an important role in prognostic modeling. More often a prognostic index is developed from

the model based on weighted sum of the prognostic factors, where weights are the multivariate regression coefficients or a form of transformation (see Legrand et al. (2009)). This new variable is used to classify patients into different risk levels e.g. poor, moderate and good. As noted by Altman and Royston (2000), many prognostic indices fail to be utilized due to lack of clinical credibility and evidence of additional information to inform decision making. Thus as stated by Legrand et al. (2009), the main determinant in the quality of a prognostic model is its validity and generalizability. Justice et al. (1999) remark that the two main components of validation are calibration and discrimination. This two aspects are discussed further in Section 2.2.

1.3 Traumatic Brain Injury (TBI)

Traumatic brain injury (TBI) also known as intracranial injury occurs when the brain sustains traumatic injuries arising from external mechanical force such as acceleration, or deceleration, impact, blast waves or penetration by a projectile Maas et al. (2008). The most common TBI cases arise from falls, car accidents and violence. According to Saatman et al. (2008) the severity of TBI can be classified into mild, moderate and severe based on the Glasgow Coma Scale quantifications. According to Thurman et al. (1999), TBI is a leading global cause of death and disability in children and young adults. It is estimated that about 14 million people are estimated to sustain TBI each year in the USA, of whom 50 000 succumb to their injuries. Furthermore about 2 percent of the US population is believed to live with TBI related disabilities today. In Europe, TBI is related annual fatality incidence or hospital admissions stands at about 235 per 100 000 persons. In spite of this, there is little concerted effort to deal with TBI particularly in pharmaceutical related research investments. This has largely been attributed to the poor knowledge of its pathological pathways and prognostication implying that majority of the people on whom the therapies are tested are often less likely to benefit from the intervention. As such it is of great importance to determine the relevant prognostic factors in determining the possible patient outcome to determine the suitable therapy. More importantly, how these factors combine to affect the patient's outcome. Heterogeneity among care centers is also quoted as one of the barriers in understanding TBI (Saatman et al., 2008).

1.3.1 Previously identified TBI Prognostic factors

Lingsma et al. (2010) give a vivid review of literature on factors previously found to affect patient TBI outcome. Though they discuss various outcomes such as mild disability vs recovery, death or vegetative state, this paper focuses only on the death outcome. Among the factors cited to influence TBI outcome are demographic factors such as age. Almost all the studies reviewed by Lingsma et al. (2010) found older patients more associated to poorer outcomes. Gender and ethnicity have also been weakly linked to TBI. Meta analysis studies by Faraceand and Alves (2000) associated women to poorer outcome compared to men. Similarly black patients are weakly associated to poorer outcomes compared to their asian and white counterparts. Clinical severity related factors have also been closely associated to variation in TBI outcomes. Studies by Gennarelli et al. (1989) and McMahon et al. (1999) have associated coexistence of moderate TBI and extra-cranial injuries with high mortality and morbidity. However, there is no consensus on the prognostic value of extra cranial injury on TBI outcome with some studies Sarrafzadeh et al. (2001) attributing the effect on primary cerebral damage and not extra cranial injury. The severity of extra-cranial injuries is indicated by consciousness level assessed with GCS.

Balestreri et al. (2004) associate abnormality in pupillary reactivity to poorer outcomes. The abnormality is attributed to the brain cistern damage or compression. Structural abnormalities observed from CT-scan have also shown varied effects on the TBI outcomes. These abnormalities include status of basal cisterns, midline shift, the presence and type of intracranial lesions and traumatic subarachnoid haemorrhage. Other prognostic factors that have been widely studied in literature include biomarkers such as putative serum and laboratory values such as high glucose concentration and platelets coagulation.

This paper focuses on easily measurable factors under normal clinical conditions obtained at patient admission, similar to those considered by Steyerberg et al. (2010). Particular interest is paid to the core prognostic factors: motor GCS score; age and pupils reactivity.

1.4 Objective

This report aims to study the application of shared frailty models in to prognostic modeling for TBI. We target to demonstrate the need to account for heterogeneity in prognostic modeling of clustered data where the shared frailty model is the natural modeling tool. As such we fit models ignoring the clustered nature of the data and compare to models considering clusters as fixed effects and models with clusters as frailties. Models are compared in terms of their parameter estimates and associated standard errors. Of interest is how the models compare in terms of their predictive performance based on calibration and discrimination.

1.5 The data

The data set considered arose from the CRASH study and consisted of patients' time of randomization and time of death or censoring and several covariates that could possibly explain the patient's time to death since randomization. The covariates included the three major prognostic factors for TBI: age, pupils reaction (both reactive vs 1 or none); motor score (gcs_motor - categorical,1-5). Other covariates were gender; mechanism of injury (*cause*), neurosurgical operation; CT scan (*ct_scan*) cortical contusion (*cc*), patient's country income classification (*income*, high or low), major extracranial injury. Abnormal scan related variables: One or more petechial haemorrhages within the brain (*nesrip*); Obliteration of the third ventricle or basal cisterns(*otvbs*); Subarachnoid bleeding(*sb*); Midline shift over 5mm(*mso5*); Non evacuated haematoma (*neh*); and Evacuated haematoma (*eh*). Also provided was the treatment center (*hospital _id*) for each patient. A description of all the variables considered is provided in table B.6

1.6 Software

The R-program v2.13.1 was extensively used. The Package 'Survival' was used to fit the models. Calibration and discrimination based on resampling was done using the rms package by Frank Harrell (1996). The Package Coxme was used to calculate the between and within cluster c-indices. SAS procedures LIFETEST, PHREG and LIFEREG were also used for Kaplan Meier, PH and parametric models respectively.

Chapter 2

Methodology

2.1 Modelling Time to Event Data

Analysis of time to event data, commonly known as Survival Data Analysis, deals with failure time as the dependent variable representing duration until the occurrence of a well defined event such as death Klein and Moeschberger (2005). A distinguishing characteristic between survival data analysis and other statistical methods is the presence of censoring. We focus on right censored data only. Let N represent the number of patients under study. Survival data is usually recorded in the form (Y_j, δ_j), with $Y_j = min(T_j, C_j)$ where T_j is the time to event for the j^{th} patient, while C_j is the censoring time; $\delta_j = I(T_j < C_j) = 1$ for uncensored observation and 0 otherwise. A series of interrelated functions can be used to describe survival data. The survival function $S(t) = P(T > t)$, denotes the probability of surviving beyond time t. Let $f(t)$ be the density of function of time to event T, and $F(t)$ the corresponding distribution function. Further still denote the instantaneous failure rate or risk of failure conditional on having survived to time t also the known as the hazard function by

$$h(t) = \lim_{\Delta t \to 0^+} \frac{(P(t \leq T < t + \Delta | T > t))}{\Delta t} \tag{2.1}$$

The cumulative hazard function is derived from the hazard function as

$$H(t) = \int_0^t h(s)ds \tag{2.2}$$

Further

$$h(t) = \frac{f(t)}{S(t)} = \frac{-d}{dt} logS(t) \tag{2.3}$$

and

$$S(t) = exp(-H(t)) \tag{2.4}$$

2.1.1 The Kaplan Meier

The Kaplan Meier estimate for survival at time t is given by

$$\hat{S}(t) = \prod_{k<t} \frac{R_k - d_k}{R_k} \tag{2.5}$$

where R_k is number at risk just before the event at time k while d_k is the number of events at time k. It is a step function that decreases at time of deaths with $\times$ representing censoring. The survival function can also be derived from the Nelson and Alen (1972) estimator of the cumulative hazard function given as

$$H(t) = \sum_{k<t} (d_k)/R_k = -logS(t) \tag{2.6}$$

2.1.2 Cox Regression

This model is also often referred to as proportional Hazards model for the fact that for any two individuals at any time point t, their hazards are related by a proportionality constant that does not depend on t but only the regression coefficients. The model proposes a common hazard function $h_0(t)$ also referred to as the baseline hazard for all subjects under study. The hazard, $h(t)$, for an individual with a k-dimensional covariate vector X is then given by

$$h(t|X) = h_0(t) exp(\boldsymbol{\beta}' X) \tag{2.7}$$

A common assumption is that no parametric form is imposed on the baseline. Thus, the model often is referred to as semi-parametric because it allows the baseline hazard to take any unspecified form while the covariates only enter the model linearly (Wienke, 2011). The Cox model parameters can thus be estimated by the method of partial likelihood presented alongside by Cox (1972) in the

same paper he proposed the model. Assuming no ties in the data, the partial likelihood is given as:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \{ \frac{exp(\boldsymbol{\beta}' X_i)}{\sum_{j<t} exp(\boldsymbol{\beta}' X_j)} \}^{\delta_i} \tag{2.8}$$

Although the estimates are less efficient than the maximum likelihood estimates, having not to impose a parametric form on the baseline hazard serves as a remedial virtue against misspecification (Keele, 2007). In the presence of ties, estimates of the partial likelihood parameters are provided by Breslow, Cox, or Efron approximation is used as full likelihood estimation is time consuming (Pham, 2005) The Assumption of non-parametric baseline can be relaxed to allow the baseline to take a parametric form. The most common being the Weibull distribution. Other possible distributions include the Gamma, lognormal and exponential. The betas in the model are interpreted as the log-hazard ratios.

2.1.3 The Fixed Effect Model.

Consider the hazard function,

$$h_{ij}(t) = h_0(t) exp(X_i' \beta) \tag{2.9}$$

Where $h_{ij}(t)$ is the hazard function at time t for subject $j(j = 1, \cdots, n_i)$ from cluster i with covariate information X_{ij}, and h_0 the baseline hazard. The fixed effect model introduces the cluster effect c_i in the model as a fixed effect with a restriction $c_1 = 0$ to avoid over-parameterizatiion. The model therefore becomes,

$$h_{ij}(t) = h_0(t) exp(X_i' \beta + ci) \tag{2.10}$$

2.1.4 The shared frailty

The intuitive idea behind frailty modelling is that some subjects are more or less susceptible to observe an event than can be accounted by the model specification, i.e. more or less "frail" Keele (2007). In the univariate frailty setting, a random effect (frailty term) is introduced for each subject to account for the unobserved heterogeneity. This concept is covered in detail by Nelson (1986) alongside the concept of selection where the most frail individuals die out first and systematic

selection of less frail ones follows as a consequence. In this paper we focus on the multivariate shared frailty which assumes the random effect to operate at a group level.

Consider individuals drawn from clustered data with a total of K clusters of sizes $n_1, n_2, \cdots n_n$, With $\sum_{i=1}^{n} n_i = N$, the conditional hazard function for individual i in cluster k is then given by

$$h_{ij}(t) = h(t|x_{ij}, u_i) = u_i h_0(t) exp(\beta' x_{ij}) \tag{2.11}$$

a product of cluster specific random effect u_i representing the frailty and baseline hazard function $h_0(t)$ common to all individuals and individual specific covariate effects. Individuals in cluster i are assumed to share the same frailty, u_i which is the cause of dependence between lifetimes within the cluster. The Frailty model further assumes that given the random effect u_k, the proportional hazards assumption holds. $h_{ij}(t)$ is thus called the conditional hazard given u_i. The dependence is positive unless variance $u_i = 0$ meaning no variability between clusters, in which case the frailty is said to be degenerate. The corresponding conditional survival function for subject j in cluster i is then given by

$$S_c(t|x_{ij}, u_i) = [S_0(t)]^{u_i exp(\boldsymbol{\beta}' x_{ij})} \tag{2.12}$$

which represents the probability of being alive at time t given that a patient is in the i^{th} cluster. Though the random effect is unobserved, it is believed to follow a certain distribution $f(U)$, the most common being the gamma frailty, the log-normal frailty and the constant stable frailty. The shared frailty model assumes correlations between lifetimes of randomly selected pairs are always the same implying a symmetric situation equivalent to that in the compound symmetry assumption in mixed models.

Shared Gamma Frailty

The one parameter gamma frailty is the most widely applied particularly due to its mathematical tractability and ease of parameter interpretation. Fitting of the gamma frailty depends on whether a parametric assumption is made on the baseline survival or not. For the parametric gamma, the unconditional likelihood can easily be derived by integrating out the clusters. This yields simple

expressions of the likelihood function, which are important for ML parameter estimation. As noted by Wienke (2011), its closed form can easily be obtained from

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}, \lambda) = \prod_{i=1}^{n} \int_0^{\infty} \prod_{j=1}^{n_i} \Big[u_i h_0(t_{ij}; \boldsymbol{\theta}) exp(\boldsymbol{\beta'} \boldsymbol{X}_{ij}) \Big]^{\delta_i} exp(-u_i H_0(t_{ij}; \boldsymbol{\theta}) e^{\boldsymbol{\beta'} \boldsymbol{X}_{ij}} f(u_i; \lambda) du_i \tag{2.13}$$

with $f(u) \sim gamma(\lambda, \lambda)$

$$= \prod_{i=1}^{n} \frac{\Gamma(\frac{1}{\lambda} + d_i) \prod_{j=1}^{n_i} \Big[h_0(t_{ij}; \theta) exp(\boldsymbol{\beta'} \boldsymbol{X}_{ij}) exp(\boldsymbol{\beta'} \boldsymbol{X}_{ij}) \Big]^{\delta_i}}{\Big(\frac{1}{\lambda} + \sum_{j=1}^{n_i} H_o(t_{ij}; \boldsymbol{\theta}) e^{\boldsymbol{\beta'} \boldsymbol{X}_{ij}} \Big)^{(\frac{1}{\lambda} + d_i)} \lambda^{\frac{1}{\lambda}} \Gamma(\frac{1}{\lambda})} \tag{2.14}$$

The log-likelihood is obtained by taking the logarithm of equation 2.14. The frailty term is explicitly estimated as

$$u_i = \frac{(\frac{1}{\lambda} + d_i)}{\frac{1}{\lambda} + \sum_{j=1}^{n_i} H_o(t_{ij}; \theta) e^{\boldsymbol{\beta'} \boldsymbol{X_{ij}}}} \tag{2.15}$$

where d_i is the number of events in cluster is i , θ is the baseline parameter vector and λ is the single parameter gamma variance. In semi-parametric frailty models, the baseline hazard is treated as a nuisance parameter. As such the Expectation Maximization (EM) algorithm (Dempster, 1977) first adopted for survival data by Nielsen et al. (1992) is the ideal method for parameter estimation treating the random effects as unobserved variables. Execution of the EM algorithm is however slow as it requires many iterations. An estimation approach which uses penalized partial likelihood (PPL) is presented by Therneau and Grambsch (2000) where the random effect is treated as a penalty term. In the Gamma frailty model both the PPL as well as EM algorithm lead to the same result. This does not hold for other frailty distributions. The frailty term in the semi-parametric case is given by

$$u_i = \frac{(\frac{1}{\lambda} + d_i)}{\frac{1}{\lambda} + \sum_{j=1}^{n_i} H_o(t_{ij}) e^{\beta' X_{ij}}} \tag{2.16}$$

where $H_0(t_{ij})$ in this case is a non-parametric estimator of the baseline hazard. The parameters are adjusted for the correlation in the clusters, and the frailty variance is interpreted as a measure of the correlation between the lifetimes in the clusters. Other measures of dependence such as the

Kendall's Tau, $K = \theta/(2+\theta)$, can also be derived from the parameters of the model (Shoukri et al., 2010). K is interpreted as the intra-cluster correlation.

2.1.5 The Weibull parametric model

Although this report is intended to focus on semi-parametric Cox proportional hazards model, results of model fit are compared to those from the parametric fit, in particular the weibull, which is the most commonly used distribution on the baseline. The Weibull hazard function is of the form :

$$h_i(t) = exp(x_i'\beta)^{\rho}\lambda\rho t^{\rho-1}) \tag{2.17}$$

and the corresponding survival function is of the form:

$$S_i(t) = exp(-\lambda t^{p} exp(\rho x_i t'\beta)) \tag{2.18}$$

2.2 Prognostic model performance for Survival Data

2.2.1 Calibration

Calibration is the degree of correspondence between the model estimated probabilities and the actual observed probability Dreiseitl and Ohno-Machado (2002). This is usually presented using calibration curves which plot predicted versus observed outcomes. A calibration curve for time to event data plots the Kaplan Meier estimates at a certain time point against the predicted probabilities at that time point (Vergouwe et al., 2002). Subjects are usually grouped for instance by deciles of predicted probability. The Kaplan Meier is calculated for each group and plotted against model predicted survival. Calibration can also be assessed by plotting the observed survival time object against the predicted and assessing the slope. i.e. $\boldsymbol{Y}$ v/s $\boldsymbol{\beta' X}$, where $\boldsymbol{Y}$ is a survival object. However, this approach may be limited by censoring and the first is the most applied. In either cases, the slope should be equal to one in case of ideal calibration.

2.2.2 Discrimination: the C-index

Discrimination is the model ability to correctly classify patients into different groups (Steyerberg et al., 2010). In survival data analysis, discrimination assess if the relative ranking of individual predicted risk is in the correct order as of the observed. The c-index is the most common tool in evaluation of survival model discrimination. It is estimated by looking at all pairs of samples which are comparable and calculating the probability of the pairs being concordant. Two subjects $i\&j$ are comparable if their event times are not tied i.e $(T_i < T_j|D_i = 1)$, both are not censored, or the censoring occurs for one member after the other experiences an event. The ties are neglected. Two event times are concordant if further the predicted survival probability of subject i above with shorter event time is lower than that of subject j. That is $(S_i > S_j|T_i < T_j)$, where S represents the survival probability.

2.2.3 Validation of PI performance

Prognostic model performance can be studied in sample of patients different from the developing sample. Internal validation is when a model is fit on a section of a data set and validated on the remainder Steyerberg et al. (2010). External validation is carried out on a data set independent from the development sample e.g. data from different centres collected by independent investigators (Altman et al., 2008, Harrell, 2001); thus model generalizability is tested on a more diverse scenario broadening the conclusions.

Model Performance in clustered data

In the presence of clustering, van Oirbeek and Lesaffre (2010) suggested a novelle approach that uses the random effect in its calculations and can be used to judge the significance of accounting for covariates and clusters apart as well as jointly by comparing the model discriminating abilities in terms of the c-indices. Define the c-index as

$$C = P(\widehat{T}_i < \widehat{T}_j|T_i < T_j) = P\big(S(t|X_i) < S(t|X_j)|T_i < T_j\big) \tag{2.19}$$

for any $t > 0$

In a PH model this can also be expressed as

$$S_i(t) < S_j(t) \Leftrightarrow S_0(t)^{e^{\beta' X_i}} < S_0(t)^{e^{\beta' X_j}} \Leftrightarrow e^{\beta' X_i} > e^{\beta' X_j} \tag{2.20}$$

Lessafre and Oirbeek demonstrate that two classes of concordance can be inferred: Between and within cluster concordance depending on the pairs considered. These can further be broken into marginal, conditional and overall concordance. The marginal is derived by considering the marginal survival function from which the frailties have been integrated out. Let $S_c(t|x, u_k = S_0(t))^{u_k e^{\beta' X_i}}$ be the conditional survival for subject in cluster k, the between cluster concordance is defined as

$$Sc(t|x, u_k) < Sc(t|x, u_l) \Leftrightarrow S_0(t)^{u_k e^{\beta' X_i}} < S_0(t)^{u_l e^{\beta' X_j}} \Leftrightarrow u_k e^{\beta' X_i} > u_l e^{\beta' X_j} \tag{2.21}$$

This reduces to equation 2.20 in the case of within cluster. The overall concordance can then be expressed as $C_0 = C_w \pi_w + C_B \pi_B$ where C_w and C_B are within and between concordance probabilities respectively; while π_w and π_B are probabilities that pairs are within and between cluster respectively. It is further noted the C_o converges to C_B if the cluster size is fixed and the number of clusters increases infinitely and converges to C_w if the number of clusters is fixed and the cluster size increases infinitely. As such it is not a suitable measure of model performance as it is a function of the number and size of clusters. C_w measures the intra-cluster predictive ability of β while marginal C_B measures the inter-cluster predictive ability of β . The inter-cluster predictive ability improvement of the model can be checked by comparing the Conditional C_B and marginal C_B.

2.3 Missing data

To deal with missing data, multiple imputations were generated by Gibbs sampling as described in van Buuren et al. (2005). This algorithm imputes an incomplete column (the target column) by generating 'plausible' synthetic values given other columns in the data. Each incomplete column is both a target for imputation and a predictor to other incomplete columns. The most recent imputations are re used to complete the predictors before the target column is imputed. A separate univariate imputation model is specified for each column, depending on its measurement level. Dummy variables are created for categorical data before the corresponding variable is imputed.

Chapter 3

Results

3.1 Exploratory Analysis

The main data analyzed in this project was a part of a larger data set from the CRASH study on TBI. The original data consisted of 10,008 patients recruited into 239 centers. However, the data used in this analysis comprised of the top 20 of the original clusters. These clusters were each randomized, 10 to the model development set and 10 to the validation set. Preference for the subset was based on study objective which was to compare models with different cluster handling techniques. Particularly it was important to assess the performance of the validation indices within each cluster hence clusters of substantial size would be required. It is also believed that the C-index is influenced by the size and number of the clusters (van Oirbeek and Lesaffre, 2010), and therefore it was important to validate the model on an almost equal data set. It is a common practice to analyze an informative subset of clusters. For instance Legrand et al. (2009) analyze a subset of clusters with at least 5 observations each, they refer to as informative clusters. It's also cited in literature that the fixed effects model becomes unstable with increase in the number of clusters particularly when the amount of censoring is very high (Heinze and Schemper, 1993), which was the case in the full dataset. Thus for demonstrative purposes, it was found reasonable to base the results on this subset although the models were later also assessed on the full set. The minimum and maximum number of patients per cluster was 118 and 841 respectively for the training set and 113 and 525 for the test set.

Table 3.1: The baseline variable distribution

		Train		Test	
	Variable	**Frequency**	**Percent**	**Frequency**	**Percent**
cc	1	366	16.97	287	14.21
	2	1034	47.94	1201	59.46
	NA	757	35.1	532	26.34
	Missing	339		291	
ct_scan	1	1732	69.59	1772	76.91
	2	757	30.41	532	23.09
	Missing	7		7	
gender	1	550	22.04	459	19.86
	2	1946	77.96	1852	80.14
income	0	382	15.3	542	23.45
	1	2114	84.7	1769	76.55
major_extracranial_injury	0	1872	75.54	1780	77.8
	1	606	24.46	508	22.2
	Missing	18		23	
neurosurgical_operation	1	413	16.69	422	18.43
	2	2061	83.31	1868	81.57
	Missing	22		21	
neh	1	441	17.73	388	16.85
	2	1290	51.85	1383	60.05
	NA	757	30.43	532	23.1
	Missing	8		8	
eh	1	108	4.34	279	12.11
	2	1623	65.23	1492	64.79
	NA	757	30.43	532	23.1
	Missing	8		8	
nesrip	1	396	15.92	414	17.98
	2	1335	53.66	1357	58.92
	NA	757	30.43	532	23.1
	Missing	8		8	
Ns	1	451	18.13	484	21.02
	2	1280	51.45	1287	55.88
	NA	757	30.43	532	23.1
	Missing	8		8	
otvbs	1	329	13.22	475	20.63
	2	1402	56.35	1296	56.27
	NA	757	30.43	532	23.1
	Missing	8		8	
pupils	0	2245	89.94	2118	91.65
	1	251	10.06	193	8.35
	Missing	2		8	
mso5	1	459	18.45	626	27.18
	2	1272	51.13	1145	49.72
	NA	757	30.43	532	23.1
	Missing	8		8	
Gcs_motor	1	170	7	113	6
	2	95	4	82	5
	3	138	6	124	7
	4	252	10	238	13
	5	1841	74	1240	69

Table 3.1 presents baseline covariate distribution for the training and the test data sets. Age was the only continuous covariate. The median age was 30 while the mean was 35.1 with a standard deviation of 16.5.

The analysis began by graphical exploration of the data to visualize the distribution of the survival times with respect to the clusters and the recorded variables. Figure 3.1 below illustrates the overall survival curve for all subjects in the training set. From the Kaplan Meier Curve, it can be concluded

that most of the deaths that occurred in the six month study were recorded within the first month. Most of the individuals who lived beyond one month survived to the end of the study.

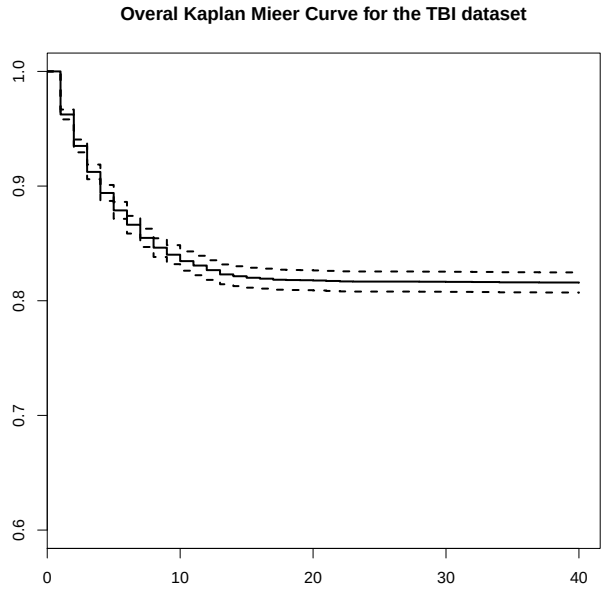

Figure 3.1: Kaplan Meier Curve for Overall Population Survival

Figure 3.2 presents survival curves for the selected 10 clusters in the training set. From the plots we can judge that while some survival curves overlap, some show clearly differing trends indicating possible cluster effect thus the need to account for it. The Kaplan Meier was also used to assess survival variation by variable levels (results omitted)

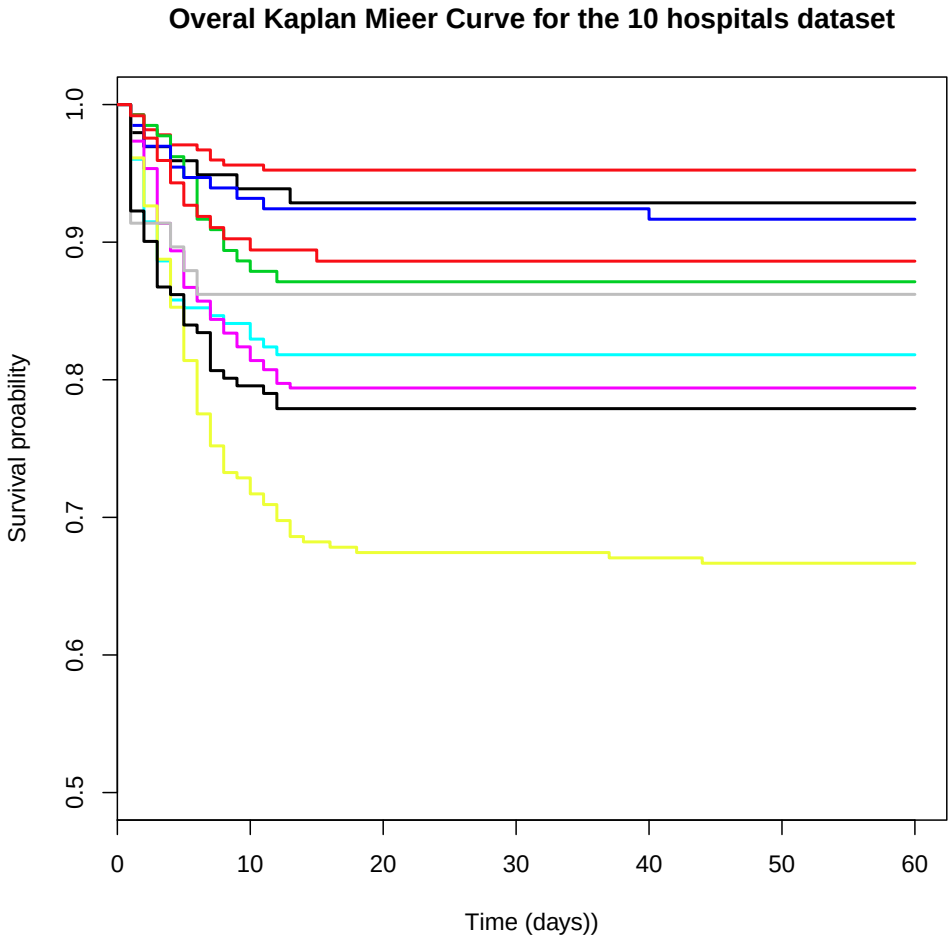

Figure 3.2: Survival Curves for the 10 clusters in the training set

3.2 Univariate Models

Model fitting commenced with univariate models on the training set. This was important in not only identifying factors to be included in the multivariate model, but also in identifying relevant prognostic factors influencing TBI related deaths independent of the other factors. Due to the substantial amount of data as stated by Courgeau and Levievre (2001), the proportional hazards model was preferred in this initial analysis. Table 3.2 below presents the results from the univariate analysis.

Table 3.2: Univariate result for the Cox PH model, frailty and fixed effects model

Parameter	Marginal	frailty	Fixed
age	0.019 (0.003 ; 0.007)	0.021 (0.003)	0.021 (0.003)
factor(ct_scan)2	-2.21 (0.265 ; 0.795)	-0.976 (0.319)	-0.896 (0.321)
factor(pupils)1	1.909 (0.119 ; 0.387)	1.48 (0.127)	1.469 (0.126)
as.factor(motor_gcs)2	0.453 (0.198 ; 0.307)	-0.311 (0.207)	-0.335 (0.206)
as.factor(motor_gcs)3	-0.008 (0.198 ; 0.261)	-0.387 (0.208)	-0.405 (0.207)
as.factor(motor_gcs)4	-0.54 (0.19 ; 0.282)	-0.972 (0.202)	-0.99 (0.201)
as.factor(motor_gcs)5	-1.977 (0.165 ; 0.412)	-2.185 (0.178)	-2.2 (0.177)
factor(cause)2	0.113 (0.168 ; 0.214)	0.078 (0.17)	0.075 (0.17)
factor(cause)3	-0.125 (0.132 ; 0.251)	0.036 (0.136)	0.037 (0.136)
factor(major_extracranial)	0.446 (0.122 ; 0.169)	0.333 (0.125)	0.33 (0.125)
as.factor(ns)2	2.166 (0.308 ; 0.335)	2.1 (0.314)	2.098 (0.318)
as.factor(nesrip)2	-0.559 (0.139 ; 0.186)	-0.547 (0.145)	-0.549 (0.145)
as.factor(otvbs)2	-1.346 (0.127 ; 0.214)	-1.18 (0.134)	-1.156 (0.137)
as.factor(neh)2	-0.862 (0.128 ; 0.203)	-0.832 (0.131)	-0.833 (0.132)
as.factor(mso5)2	-1.284 (0.148 ; 0.215)	-1.26 (0.151)	-1.269 (0.152)
factor(cc)2	-0.635 (0.13 ; 0.187)	-0.574 (0.131)	-0.567 (0.131)
factor(gender)2	0.421 (0.156 ; 0.222)	0.171 (0.157)	0.168 (0.157)
income	0.364 (0.18 ; 0.488)	0.671 (0.462)	0.465 (0.453)

From the univariate analysis, the Cox PH model fitted without adjustment for clustering found all variables including gender ($p-val = 0.007$) and income ($p-value = 0.0427$) significant except the cause of the injury. On the other hand the frailty and the fixed effect models found both gender and income non-significant. This is in line with literature as stated by Borecki and Province (2008) that failing to account for clustering in the Cox models may lead to inflation of type I error. Adjustment of these errors was done in the Cox model using robust sandwich estimator for standard errors (Huber, 1967; White, 1982), with model parameters interpreted at marginal level. After adjustment the two variables were borderline insignificant. The fixed effect model and the frailty model give very close parameter estimates as well as standard errors as they both capture the effect of clustering.

3.3 The Core Model

Special interest was in the performance of model comprising of the core factors : Age, Pupils and GCS motor. These factors have extensively been analyzed in literature based on other methods and are often the basis of prognostic models in TBI. For instance, (Lingsman et al, 2010) analyzed the same data using Generalized Linear Models based on binary mortality outcome at 6-months to determine cluster influence on treatment effect. There were no missing data for the variables in the

core model. The results for the fitted model are presented in Table 3.3 below.

Table 3.3: Parameter estimates for the Core model

Parameter	Marginal (Naïve; Robust)	Frailty	Fixed Effect
age	0.017 (0.003 ; 0.006)	0.019 (0.003)	0.019 (0.003)
factor(pupils)1	1.109 (0.133 ; 0.186)	0.743 (0.138)	0.73 (0.138)
as.factor(motor_gcs)2	0.511 (0.199 ; 0.337)	-0.125 (0.208)	-0.146 (0.207)
as.factor(motor_gcs)3	0.155 (0.199 ; 0.277)	-0.125 (0.211)	-0.141 (0.21)
as.factor(motor_gcs)4	-0.236 (0.193 ; 0.232)	-0.638 (0.205)	-0.657 (0.204)
as.factor(motor_gcs)5	-1.489 (0.178 ; 0.336)	-1.733 (0.19)	-1.75 (0.19)
factor(hospital_id)774			-0.807 (0.472)
factor(hospital_id)776			0.459 (0.451)
factor(hospital_id)802			0.707 (0.486)
factor(hospital_id)815			1.393 (0.418)
factor(hospital_id)847			1.518 (0.405)
factor(hospital_id)854			1.842 (0.399)
factor(hospital_id)1838			1.173 (0.492)
factor(hospital_id)1941			1.91 (0.408)
factor(hospital_id)2113			1.199 (0.458)

From table 3.3, similar to the univariate case, the results from the core model show consistency between the frailty and the fixed effect model both in terms of parameter estimates and standard errors. Again despite the Cox model parameter estimates being as large as the frailty model and fixed effect models in most cases, their variances are underestimated and the risk of type I error is inflated.

3.3.1 Test for Cluster effect

The test for random effect for the frailty model was done by comparing the partial log-likelihood for the models with and without the frailty term. The change in the partial log-likelihood was -2(-2024.42 - -1973.678) = 101.48. This was compared to the chi-square with one degree of freedom and its p-value divided by 2 as this is a $\chi^2_{0;1}$ boundary problem. The resulting p-value was below 0.001 indicating that indeed the clustering needed to be accounted for. The variance of the frailty term was 0.775 giving Kendall's tau of 0.28 interpreted as intra-cluster dependency.

For the fixed effect model, the usual likelihood ratio test was used to compare the model with and without the clusters. The change in partial log-likelihood was -2(-2024.42 - -1973.4) = 102.06.

Comparison with the chi-square of 9 degrees of freedom gave a $p - value < 0.0001$.

3.3.2 Model Diagnostics

Test for Cox Proportionality Assumption

The Cox proportionality assumption was tested by including an interaction term of log-transform of time with each covariate to be tested. Table 3.4 presents the results of model checking for the Cox PH and the fixed effect model.

Table 3.4: Cox Proportional Odds Assumption test for Fixed and Naive Cox PH models

Parameter	**Fixed Effect**			**Cox PH**		
	rho	chisq	p	rho	chisq	p
age	-0.014	0.076	0.782	0.012	0.046	0.831
pupils	-0.120	4.750	0.029	-0.100	3.160	0.076
as.factor(motor_gcs)2	0.011	0.039	0.843	0.049	0.716	0.398
as.factor(motor_gcs)3	-0.004	0.004	0.949	0.000	0.000	0.996
as.factor(motor_gcs)4	0.021	0.140	0.708	0.021	0.131	0.718
as.factor(motor_gcs)5	0.043	0.654	0.419	0.051	0.809	0.368
factor(hospital_id)774	-0.006	0.011	0.917			
factor(hospital_id)776	0.064	1.205	0.272			
factor(hospital_id)802	0.009	0.026	0.873			
factor(hospital_id)815	-0.037	0.415	0.520			
factor(hospital_id)847	0.033	0.318	0.573			
factor(hospital_id)854	0.071	1.604	0.205			
factor(hospital_id)1838	-0.095	2.791	0.095			
factor(hospital_id)1941	-0.029	0.259	0.611			
factor(hospital_id)2113	0.017	0.087	0.768			
GLOBAL	NA	35.925	0.002	NA	6.260	0.394

From Table 3.4 it can be concluded that none of the four covariates showed any violation of the proportionality assumption in the PH model. However for the model with fixed effects, the introduction of hospital ID factors caused the other covariates to show proportionality assumption violation. It is also advised to examine the proportionality assumption by use of schoenfeld residual plots to determine the presence of any trend (Thernau and Grambsch, 2000). Figure 3.3 below presents diagnostic plots for the three covariates in the PH model. As can be seen none of the plots seems to indicate any time trend to warrant corrections. Similar findings were found for the fixed effect model (results not presented).

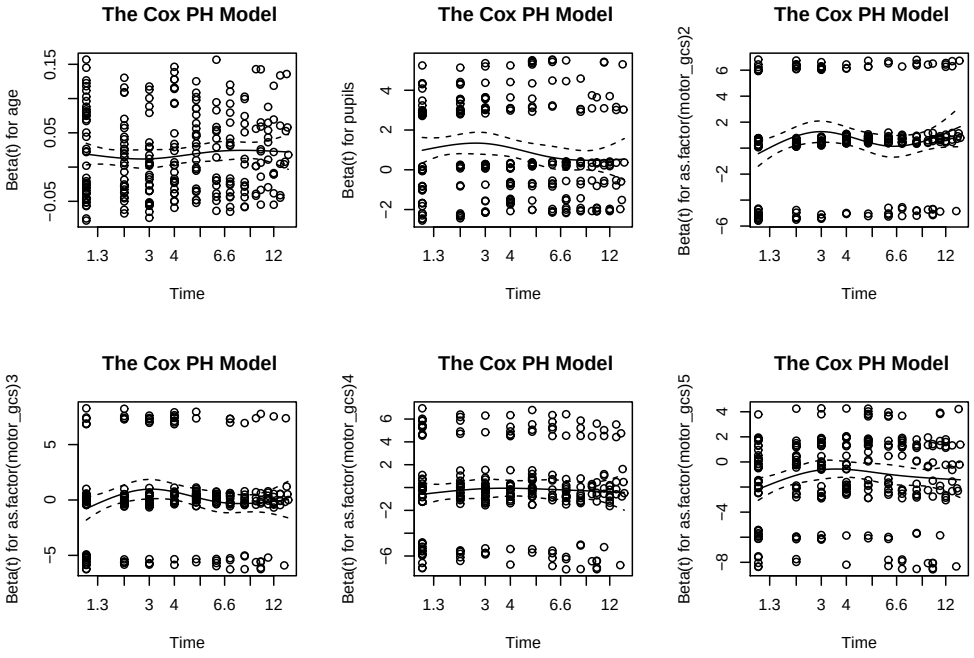

Figure 3.3: Test for Proportional Odds Assumption for the Cox Model

Sensitivity to the Frailty distribution misspecification;

The sensitivity to misspecification for the frailty distribution was assessed by fitting a log-normal distributed frailty model and comparing the parameter estimates to those of the gamma frailty model. Table 3.5 below presents results comparing the two models. It can be inferred from Table 3.5 that the parameter estimates were not highly sensitive to the misspecification of the frailty distribution; particularly if the true frailty distribution was log-normal. However the log-normal seemed to fit the data a little better based on the AIC values.

Table 3.5: Comaparison of the Gamma and Log-normal Frailty model parameters

Parameter	Gamma	Log-normal
age	0.019 (0.003)	0.019 (0.003)
pupils	0.743 (0.138)	0.747 (0.138)
as.factor(motor_gcs)2	-0.125 (0.208)	-0.114 (0.208)
as.factor(motor_gcs)3	-0.125 (0.211)	-0.118 (0.21)
as.factor(motor_gcs)4	-0.638 (0.205)	-0.628 (0.205)
as.factor(motor_gcs)5	-1.733 (0.19)	-1.726 (0.19)
Var. Frailty	0.775	0.602
AIC	4228.198	4228.064

3.3.3 Parameter interpretation

It should be born in mind that the parameter interpretations for these three models differ and examining their magnitude alone is of no relevant consequence. The marginal model parameters express a population average risk, while the frailty model expresses cluster specific risk. The fixed effect model as well assumes independence of observations within clusters. The Cox model parameters are interpreted as log risk rates on population average level. For instance, for the marginal Cox model, the risk of an individual with at least one or both pupils unreactive dying was exp (1.109)= 3.0 times higher compared to an individual who had both pupils reactive to light. For the frailty, for a given hospital, the risk of that individual with at least one eye unreactive dying is exp (0.746) = 2.1 i.e. twice as high compared to an individual in the same hospital but with both pupils reactive.

3.3.4 Parametric model form

The above models were also fitted under the parametric assumption to compare the hazard ratios predicted to that from the semi-parametric model. The extensively applied Weibull parametric model was chosen for convenience purposes although lognormal distributed baseline hazard models were also fitted to assess extend of possible baseline hazard misspecification. The results from the log-normal baseline fit [not presented] showed only marginal disparity from the Weibull baseline model. Table 3.6 presents the results for the Cox PH; Fixed effect and the Frailty Models under Weibull parametric assumption.

Table 3.6: Parametric form for the core model

Parameter	Marginal (Robust; Naive)	Gamma Frailty	Fixed Effect Model
Intercept	12.319 (1.591 ; 0.822)	12.356 (0.989)	14.264 (1.447)
age	-0.063 (0.02 ; 0.011)	-0.06 (0.011)	-0.063 (0.012)
pupils	-3.816 (0.472 ; 0.475)	-2.168 (0.456)	-2.167 (0.468)
as.factor(motor_gcs)2	-2.309 (0.971 ; 0.69)	-0.399 (0.668)	-0.22 (0.686)
as.factor(motor_gcs)3	-0.648 (0.973 ; 0.704)	0.264 (0.687)	0.371 (0.71)
as.factor(motor_gcs)4	0.65 (0.832 ; 0.679)	1.902 (0.674)	2.125 (0.699)
as.factor(motor_gcs)5	3.506 (0.831 ; 0.635)	4.818 (0.649)	5.126 (0.68)
factor(hospital_id)774			0.898 (1.523)
factor(hospital_id)776			-1.345 (1.458)
factor(hospital_id)802			-1.962 (1.574)
factor(hospital_id)815			-4.202 (1.374)
factor(hospital_id)847			-4.753 (1.333)
factor(hospital_id)854			-5.982 (1.317)
factor(hospital_id)1838			-5.209 (1.714)
factor(hospital_id)1941			-5.497 (1.36)
factor(hospital_id)2113			-4.129 (1.523)
Log(scale)	1.227 (0.05 ; 0.055)	1.147 (0.053)	1.173 (0.054)

Similar to the Proportional odds assumptions model, the parametric Cox model with a Weibull baseline hazard showed very little variation in parameters and standard errors between the frailty and the fixed effect regression models. The intercept for the fixed effect model was higher than the rest of the models as it also accommodates for the clustering reference point. The interpretation for the parameters similarly differs for the three different models as previously stated.

The parameter interpretation is based on the transform $\beta_j = \frac{-\gamma_j}{\sigma}$, where γ_j is the Weibull representation of j^{th} covariate coefficient while σ is the scale parameter. Thus the effect of hazard ratio for the individual with at least one pupil unreactive compared to both pupils reactive is given by $exp(-3.256/exp(1.215) = 3.04$ which is similar to the hazard ratio obtained from the proportional hazards model previously from table 3.3. Through similar calculations for the frailty model we get a risk rate of 2.0 times as high for individual with at least one pupil unreactive to light compared to individual with both pupils reactive to light. Again interpretation is conditional on the random effect. These results again are in agreement with the semi-parametric model fitted before. The rest of the parameters can be interpreted in the same manner. The change in partial log-likelihood when the random effect was included was 107.8 and compared to $\chi_{0;1}$ gave $p - value < 0.0001$ indicating significance.

3.4 Extended Multivariate Models

The extended multivariate model was developed by incorporating more covariates found significant in the univariate analysis into the core model. Model building was done by backward elimination while retaining the core variables. This was done in the three modeling approaches, the marginal Cox, the frailty and fixed effect models. The final model consisted of four more covariates found significant all the models: cortical contusion (cc), neh, otvbs and mso5. The fitted model parameter estimates are presented in Table 3.7 below. A major feature observed in the data set was the missing covariate values for individuals who didn't have a CT-scan. As such all the other variables related to CT-scan were not recorded. Attempts to seek for details on the reasons for missing CT-scan were not fruitful. As such, two separate analysis were done, one based on the observed data and one based on multiple imputation. The R library MICE (Multiple imputation chain equations) was used in the imputation. Attempts to treat the missing data as a category were thwarted by convergence problems.

Table 3.7: Extended multivariate Cox model excluding data without CT-scan

Parameter	**Marginal**	**Frailty**	**Fixed Effect**
ge	0.018 (0.003 ; 0.005)	0.0163 (0.0038)	0.017 (0.004)
factor(motor_gcs)2	0.41 (0.215 ; 0.292)	0.1914 (0.2297)	0.027 (0.223)
factor(motor_gcs)3	0.163 (0.229 ; 0.219)	-0.0143 (0.2421)	-0.113 (0.24)
factor(motor_gcs)4	-0.19 (0.217 ; 0.245)	-0.4827 (0.232)	-0.606 (0.231)
factor(motor_gcs)5	-0.929 (0.198 ; 0.206)	-1.2448 (0.2181)	-1.391 (0.214)
factor(pupils)1	0.649 (0.145 ; 0.125)	0.5112 (0.1544)	0.506 (0.155)
factor(cc)2	-0.24 (0.134 ; 0.211)	-0.2864 (0.1408)	-0.279 (0.14)
factor(otvbs)2	-0.888 (0.133 ; 0.132)	-0.795 (0.1383)	-0.796 (0.138)
factor(ns)2	1.193 (0.327 ; 0.277)	1.0455 (0.3388)	1.065 (0.34)
factor(neh)2	-0.259 (0.133 ; 0.162)	-0.3706 (0.1382)	-0.348 (0.136)
factor(hospital_id)774			-0.121 (0.665)
factor(hospital_id)776			0.53 (0.654)
factor(hospital_id)802			0.323 (0.685)
factor(hospital_id)815			1.139 (0.616)
factor(hospital_id)847			1.294 (0.605)
factor(hospital_id)854			1.493 (0.599)
factor(hospital_id)1838			1.88 (0.687)
factor(hospital_id)1941			1.611 (0.608)
factor(hospital_id)2113			1.533 (0.643)

It was equally notable that the fixed effect and frailty model have relatively close parameter

estimates as well as standard errors. Test for proportionality assumption showed no indication for violations by the marginal Cox model (Global $p - value = 0.27$) while for the fixed effect model, only the clustering variable (hospital_id) seemed to violate the proportionality assumption. The results for the assumptions test are presented in Table B.3 in the appendix. An associated graphical examination (results omitted) showed no reasons for concern in either of the models in terms of proportionality assumption violation. Table 3.8 presents results from the multiply imputed covariates model.

Table 3.8: Extended multivariate model with Multiple imputation

Parameter	**Marginal**	**Frailty**	**Fixed**
age	0.016 (0.003 ; 0.005)	0.018 (0.004)	0.018 (0.003)
as.factor(motor_gcs)2	0.331 (0.201 ; 0.29)	-0.146 (0.208)	-0.163 (0.207)
as.factor(motor_gcs)3	0.166 (0.2 ; 0.207)	-0.1 (0.212)	-0.117 (0.212)
as.factor(motor_gcs)4	-0.167 (0.196 ; 0.202)	-0.575 (0.206)	-0.596 (0.205)
as.factor(motor_gcs)5	-1.303 (0.181 ; 0.26)	-1.531 (0.191)	-1.55 (0.191)
pupils	0.794 (0.137 ; 0.145)	0.533 (0.142)	0.521 (0.142)
factor(cc)2	-0.309 (0.127 ; 0.167)	-0.372 (0.131)	-0.371 (0.131)
factor(otvbs)2	-1.01 (0.128 ; 0.162)	-0.826 (0.134)	-0.818 (0.133)
factor(ns)2	0.97 (0.333 ; 0.306)	0.832 (0.336)	0.832 (0.337)
factor(neh)2	-0.361 (0.127 ; 0.135)	-0.464 (0.126)	-0.471 (0.126)
factor(hospital_id)774			-0.983 (0.474)
factor(hospital_id)776			0.317 (0.453)
factor(hospital_id)802			0.323 (0.492)
factor(hospital_id)815			1.127 (0.421)
factor(hospital_id)847			1.268 (0.409)
factor(hospital_id)854			1.381 (0.405)
factor(hospital_id)1838			1.13 (0.509)
factor(hospital_id)1941			1.685 (0.415)
factor(hospital_id)2113			1.204 (0.465)

3.5 Model Predictive Performance

The predictive performances of the models were compared on the training set, the test set as well as based on re-sampling techniques. Due to the incompleteness of most of the variables in the extended model, it was decided that predictive performance of the models be compared based on the core model. This was reasonable as most of the TBI models are based on these core factors. Predictive performance on the extended model was also performed but the results did not show a great difference. Harrell's C- index (Harrell, 1996) was used to assess model discrimination while plots of model predicted survival against the observed survival were used for model calibration.

Calibration and discrimination based on the Training and test set

Table 3.9 below presents the C-indices for each of the three models for the 10 test set clusters.

Table 3.9: Concordance indices for the Core model based on the training and test set

	Training Set			Test Set	
	naïve	**Frailty**	**Fixed effect**	**naïve**	**Frailty**
	0.721	0.743	0.7428	0.898	0.892
	0.947	0.95	0.95	0.882	0.888
	0.621	0.628	0.628	0.745	0.752
	0.818	0.822	0.821	0.712	0.722
	0.825	0.816	0.815	0.723	0.723
	0.667	0.665	0.664	0.809	0.816
	0.774	0.773	0.772	0.754	0.754
	0.934	0.934	0.935	0.77	0.776
	0.747	0.755	0.756	0.732	0.729
	0.802	0.799	0.8	0.76	0.759
Mean	**0.786**	**0.789**	**0.788**	**0.778**	**0.781**
stdev	0.104	0.102	0.102	0.064	0.064
Overal	**0.797**	**0.854**	**0.853**	**0.741**	**0.743**

Based on the training set, a marginally higher c-index per cluster was noted for the frailty model followed by the fixed effect model while the naïve Cox model had the lowest average though still only marginally different. There was no notable difference in variability of the c-index across clusters. The overall c-index was however clearly higher for the frailty model and the fixed effect model compared to the naïve Cox. This can be attributed to the use of the cluster fixed and random effects in ordering of observations across clusters.

The C-indices based on the test set were only slightly lower than for the training set. This is reasonable given that the two data sets are not completely independently collected.

To assess model calibration patients were first grouped based on their 30 day model predicted survival probability into 10 groups by deciles. The Kaplan Meier 30 day survival probability was then computed for each group and plotted against the model predicted survival. Ideally, this should fall on the 45^o line. Figure 3.4 (a) and (b) displays smooth functions of the calibration curves based on Training set and test set respectively. From Figure 3.4 (a) it is clear that based on the training set the three models performed equally well with the calibration curve close to the ideal line. Figure 3.4 (b) compares the calibrations for the frailty and the naïve Cox model based on the test set. Again

no clear difference could be told in terms of modal calibration. The drawback here again is that the fixed effect model cannot be used in a standard way in prediction for clusters outside those used in the development set.

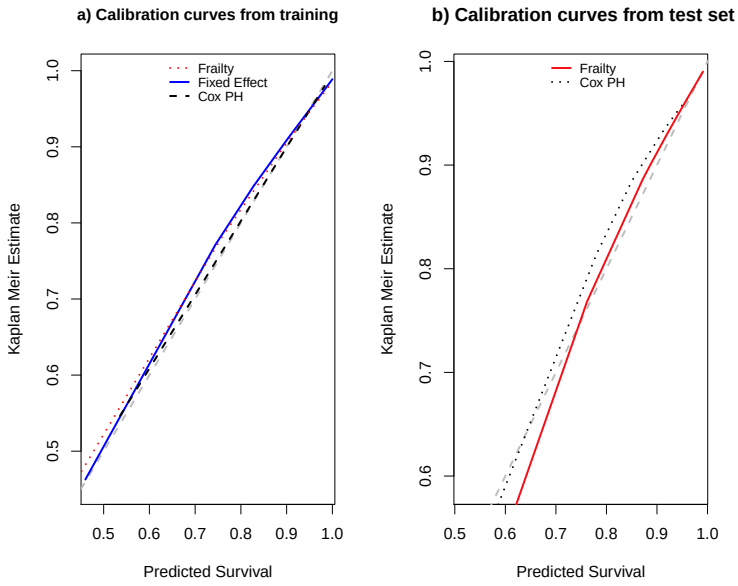

Figure 3.4: Calibration Curves for the Training and the test Set

Calibration and Discrimination based on the re-sampling

Model validation by re-sampling is aimed at correcting for over-optimism that is expressed when model predictive performance is assessed on the same data it was developed (*See* Harrell, 1996). Though One can also use cross-validation, bootstrapping is a more common approach for generating validation sets. Spline functions can be used to estimate the baseline hazard and provide smoothed calibration curves alongside the the correction for over optimism. Discriminative performance of the model was quantified based on the c-index which is calculated from the Somers D-statistic as C=0.5D_{xy}+0.5. Again, based on the results in Table 3.10, it is apparent that taking into account the clustering gives a better discrimination. Results for the prognostic performance based on the frailty model were omitted as the software packages developed for this could not fit the frailty.

Table 3.10: Resampling validation statistics for the three models

	C-index	R2	Slope
Cox	0.751	0.134	0.969
Fixed effect	0.788	0.184	0.959

Figure 3.5 illustrates the calibration results based on the 200 bootstrap samples for the Cox PH model, the fixed effect model. From the calibration curves the 90th quantile of error was 0.008, for the Cox PH and 0.018 for the fixed effect model. This also indicates that neither of the models showed a distinctively poor or better calibration than the other.

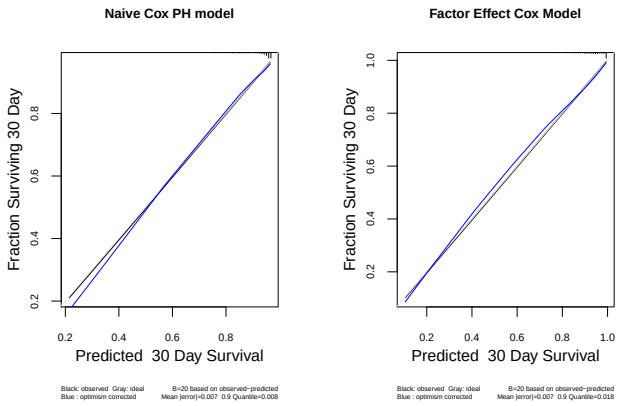

Figure 3.5: Calibration plots for the three models based on re-sampling

3.5.1 The Between and within cluster discrimination based the frailty model

A novelle idea for quantifying the added effect of accounting for clusters in model predictive performance in terms of the c-index and utilizes the estimated frailty terms was applied to the data at model based on the training set. The cluster contribution is estimated from the the difference between marginal between cluster and the conditional between c-indices. Table 3.11 shows the C indices for the core model. Based on the difference between conditional between (0.863) and marginal between (0.793), it can be judged that the introduction of clusters greatly improved the concordance performance of the model.

Table 3.11: frailty based C-indices for the core model

Overal conditional	0.862
Overal marginal	0.793
Conditional Between	0.863
Marginal Between	0.793
Within	0.786

3.6 Extended Analysis

3.6.1 Extended model Predictive performance

Model predictive performance was assessed similar to the core model above. However, due to the missingness in the data, model calibration and discrimination was performed on a singly imputed data set as this could not be achieved with multiple imputation. Table 3.12 presents the concordance indices from the model. There was an increase in concordance compared to the core model for all the models.

Table 3.12: C-indices per cluster for Extended model

	Train			Test	
	Frailty	**Naïve**	**Factor**	**Frailty**	**Naïve**
Overal	0.820	0.812	0.820	0.756	0.754
Mean per cluster	0.875	0.831	0.875	0.800	0.784

3.6.2 Analysis on the full data set

An extended analysis based on the full data was also performed but the results were not found to differ from the the subset data analyzed. Part of the analysis results are presented in the appendix. Table B.4 illustrates the results of the fitted core model. Compared to Table 3.3 of subset data, the frailty model had relatively close parameter estimates. However, the parameter estimates for the fixed effect model were different depending on whether the clustering indicator came first or last in model specification, a problem also noted by Duchateau and Jansen(2008). A common problem problem faced was censoring of almost all the subjects, which gave extreme c-index values or no results at all for some clusters.

Chapter 4

Discussion

This paper aimed at discussing the relevance of frailty models in prognostic modelling for survival data. A common phenomenon in clinical trials is the presence of multivariate failure data arising due to the clustered nature of the data where patients are treated in different centres. These clusters contain possibly dependent failure times within. As such Glidden (1993) re-affirms that such dependence must be taken into account in any statistical analysis. Although in prognostic studies, interest is often in the predictive performance of the model, examined by means of calibration and discrimination, these aspects require that the prognostic model is correctly specificied which in turn requires the correct covariates identification. Thus there is need to account for both dependence and the heterogeneity that is often induced by clustering. As stated by (Bretagnolle and Huber-Caroll (1988) , in survival analysis, unobserved variation if ignored, can lead to serious bias in both parameter estimates and in the estimate of the hazard rate. This theoretical results have been extended to the analysis of correlated failure times as stated by Pertersen, Andersen and Gill, Utrecht [Technical Paper].

In this paper we analyzed the Traumatic Brain Injury data using three approaches, the Cox model without correcting for clustering, the gamma frailty model and the fixed effect model. These three approaches were applied to univariate analysis, to the core model comprising of age, GCS motor and patients pupil reactivity; and to an extended set of factors. Interest was in comparing the model performance in terms of parameter estimates relative to standard errors, predictive performance was assessed by calibration and discrimination.

In all the fitted models we observed consistently that though almost equally as large as the

parameters from the frailty and the fixed effect models, the Cox PH model underestimates the parameter standard errors thus inflating the type I error. For instance in the univariate analysis, the naive Cox PH model found both gender and income significant, however, both the frailty and the fixed effect models found these variables to be insignificant. When adjusted for standard errors, the marginal model found this covariates insignificant too, though on the borderline. Next we fitted the Core model comprising of the three variables identified in literature as the core-prognostic factors in TBI. Various analysis models based on these factors have been carried out based on other outcomes such as survival at 6 months (Lingsma et al., 2010). Accounting for clustering was tested for both the frailty and the fixed effect model. The significance of the frailty term was assessed using a mixture of Chi-square of zero and one degree of freedom, while the likelihood ratio test was used for the fixed effect model. In both models, it was found that the cluster effect was significant and thus it was necessary to account for it. Lingsma et al. (2010) had similar conclusions on analysis of the same data set but on the binary mortality outcome at six months. However, based on the fitted gamma frailty model, the parameter estimate of 0.775 for the variance of the model meant that model estimated intra-cluster dependence was of about 0.28.

Based on the training set, the c-indices were calculated for each cluster. It was found that the frailty model had the highest C-index per cluster though very close to the fixed effect model. The weakness of the fixed effect model is the lack of portability to clusters where it was not developed as it assumes that the clusters where it was developed is the entire enumeration thus making it less generalizable. Thus validation for this model was only done by re-sampling. Based on re-sampling techniques, we demonstrated again that the fixed effect had a higher C-index values compared to the naive Cox model although again the difference was quite marginal. The marginal difference may be attributed to the low dependence showed in the data. Similarly, based on the between and within cluster c-indices as illustrated by the difference between the between marginal and between conditional c-indices. The advantage of this approach is that it is not affected by the variation in number and of clusters like the overall c-index.

All the models, showed a good performance in terms of calibration. However, no clear difference was noted in the performance of either the frailty or the fixed effect model in terms of calibration compared to the Cox PH. Although by small margin, the frailty model consistently showed better

predictive performance in terms of discrimination index compared to the Cox PH model when validated on the test set. Calibration did not seem to differ a lot for the two models, with the frailty having a slope of 1.06 while the the Cox PH had 0.94 Again the fixed effect model could not be extended to the test set as it assumes fixed number of clusters used in model development. When the the extended model was considered, the c-indices all went up but still the the models considering the clustering showed better discrimination. However, due to high-censoring, most clusters could not allow for calibration plots and discrimination indices calculation making the results spurious.

Although the Fixed effect and Frailty model both perform equally well, when the number of centres is very large, the fixed effect model overfits leading to possible bias in treatment effect or covariate. According to Heinze and Schemper (1993), when interest is in estimation of cluster effect, then the fixed effect model with a Firth correction gives a less biased estimator. Firths correction Firth (1993) is needed incase monotone likelihood is observed in the fitting process, i.e. when the likelihood converges to a finite value while at least one parameter estimate diverges to infinity. Monotone likelihood primarily occurs in small samples compared to number of clusters. However, if interest is in the treatment or other covariates effect, which is often the case in prognostic modeling, then the Random effect gives unbiased estimates.

Chapter 5

Conclusions

In this study we sought to determine the need for frailty models in prognostic modeling. To arrive at our conclusions, we compared the model estimates relative to standard errors for the naïve Cox PH, the fixed effect and frailty models. Based on the TBI data set, we showed that the Cox PH model underestimates parameter estimate standard errors inflating type I error risk in clustered data. In univariate analysis we found that gender and cause of injury did not affect TBI related deaths. Both the fixed effect and frailty models found the cluster effect significant. The full multivariate model could be sufficiently described by adding only CT-scan abnormality variables to the core model. Using the C-index to quantify model discriminative ability, the frailty model, though in some cases only by small margins, consistently had better predictive performance compared to the Cox PH model. The Fixed effect model performed almost equally as well as the frailty. However, the fixed effect model has limitations of not being generalizable to clusters from which the model was not developed. It's also cited in literature that the fixed effect model gives biased parameters when the number of clusters grows large relative to the sample size. This together with the monotone likelihood problems often encountered when fitting the model, makes the fixed effect less favorable compared to the frailty model. Calibration did not show any variation in model performance among all the three models. In conclusion, frailty modeling is important not only in correct identification of factors that influence an outcome of interest and correct model specification, but it also improves model predictive performance in terms of discrimination. More research is recommended for ways to quantify clearly model calibration in presence of clustering. We recommend that other analysis approaches such as Bayesian where all clusters can easily be incorporated into the analysis.

Bibliography

Altman, D., Vergoouwe, Y., Roystone, P., and Moons, K. G. M. (2008). Prognosis and prognostic research: validating a prognostic models. *BMJ*, 11:221236.

Altman, D. G. and Royston, P. (2000). What do we mean by validating a prognostic model? *Statistics in Medicine*, 19:59–78.

Balestreri, M., Czosnyka, M., and Chatfield, D. A. (2004). redictive value of glasgow coma scale after brain trauma: change in trend over the past ten years. *P. J Neurol Neurosurg Psychiatry*, 75:161162.

Borecki, I. B. and Province, M. A. (2008). Genetic and genomic discovery using family studies. *The Circulation*, 119:1057–1063.

Buyse, M., , Staquet, M. J., and Sylvester, R. (1984). Quality control in multicentre cancer clinical trials. in: Cancer clinical trials. methods and practice. *Oxford University Press*, 58(282):102–123.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *JBiometrika*, 65:141151.

Courgeau, D. and Levievre, E. (2001). *Event history analysis in demography*. Oxford University Press, New York.

Cox, D. (1972). Regression models and life tables. *Journal of the Royal Statistical Society*, 34(197):187 220.

Dreiseitl, S. and Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: a methodology review. *J. Biomed.Inform.*, 35(5):352–359.

Duchateau, L. and Janssen, P. (2008). *The Frailty Model*. Springer, New York, 1 edition.

Faraceand, E. and Alves, W. M. (2000). Do women fare worse? a metaanalysis of gender differences in outcome after traumatic brain injury. *Neurosurg Focus*, 8:6.

Firth, D. (1993). Bias correction of maximum likelihood estimates. *Biometrika*, 80:27–38.

Fox, J. (2002). Cox proportional-hazards regression for survival data. Technical report, Appendix to An R and S-PLUS Companion to Applied Regression.

Gennarelli, T. A., Champion, H., Copes, W. S., Sacco, W. J., and Alves, W. M. (1989). Mortality of patients with head injury and extracranial injury treatedin trauma centers. *J Trauma*, 29:1193–1202.

Harrell, F., Lee, K. L., and Mark, D. B. (1996). Multivariable prognostic models: Issues in development models, evaluating assumptions and adequacy of, and measuring and reducing errors. *Statistics in Medicine*, 11:221236.

Harrell, F. E. (2001). *Regression modeling strategies*. Springer, New York.

Heinze, G. and Schemper, M. (1993). A solution to the problem of monotone likelihood in cox regression. *Biometrika*, 57(1):114–119.

Justice, A. C., Covinsky, K. E., and Berlin, J. A. (1999). Assessing the generalizability of prognostic information. *Ann Intern Med*, 130:515–524.

Kaplan, E. L. and Meir, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 58(282):457–481.

Keele, L. (2007). Cross validation tests for frailty models. Technical report, Ohio State University.

Klein, J. P. and Moeschberger, M. (2005). *Survival Analysis - Techniques for Censored and Truncated Data*. Springer, London.

Korosteleva, O. (2009). *Clinical Statistics: Introducing Clinical Trials, Survival Analysis, and Longitudinal Data Analysis*. Jones and Bartlett, Otario, 1 edition.

Legrand, C. (2005). *Heterogeneity in multicenter clinical trials using the frailty model*. PhD thesis, Universiteit Hasselt, Campus Diepenbek.

Legrand, C., DUCHATEAU, L., Janssen, P., and SYLVESTER, V. D. R. (2009). Validation of prognostic indices using the frailty model. *Lifetime Data Analysis*, 15(1):59–78.

Lingsma, H., Roozenbeek, B., Steyberg, E., Murray, G. D., and Maas, A. I. R. (2010). Early prognosis in traumatic brain injury: from prophesies to predictions. *Lancet Neurol*, 9:543–554.

Maas, A. I. R., Stocchetti, N., and Bullock, R. (2008). Moderate and severe traumatic brain injury in adults. *Lancet Neurol*, 7:728–749.

McMahon, C. G., Yates, D. W., Campbell, F. M., Hollis, S., and M.Woodford (1999). Unexpected contribution of moderate traumatic brain injury to death after major trauma. *J Trauma*, 47:89195.

Nelson, W. (1986). Survival models for heterogeneous populations derived form stable distributions. *Biometrika*, 73:387–396.

Nielsen, G., Gill, R., and Srensen, P. A. T. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Statistics in Medicine*, 19:2543.

Oakes, D. (1982). A concordance test for independence in the presence of censoring. *Biometrics*, 38:451455.

Pham, H. (2005). *Springer Handbook of Engineering*. Springer Verlag, London.

Saatman, K. E., Duhaime, A.-C., Ross Bullock, A., Maas, I., Valadka, A., and Manley, G. (2008). Classification of traumatic brain injury for targeted therapies. *Journal of neurotrauma*, 25:719–738.

Sarrafzadeh, A. S., Peltonen, E. E., Kaisers, U., Kuchler, I., Lanksch, W. R., and Unterberg, A. W. (2001). Secondary insults in severe head injurydo multiply injured patients do worse. *Crit Care Med*, 29:11623.

Shoukri, M. M., Kumar, P., and Colak, D. (2010). Analyzing dependent proportions in cluster randomized trials: Modeling inter-cluster correlation via copula function. *Computational Statistics and Data Analysis*, 55(2011):1226–1235.

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M. J., , and Kattane, M. W. (2010). Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiology*, 21(1):515–524.

Taktka, A. F. G., Lake, A. E. S. P., and Fishe, A. C. (2009). Evaluation of prognostic models: Discrimination and calibration performance. Technical report, Dept of Clinical Engineering, Royal Liverpool University Hospitla, Liverpool, UK.

Therneau, T. and Grambsch, P. M. (2000). *Modeling Survival data: Extended Cox modeling*. Springer, New York.

Thurman, D. J., Alverson, C., Browne, D., Dunn, K. A., Guerrero, J., Jackson, R., Johnson, V., Langlois, J., Pilkey, D., Sniezek, J., and Toal, S. (1999). Traumatic brain injury in the united states. Technical report, US Congress.

van Buuren, S., Brand, J., Groothuis-Oudshoorn, C., and Rubin, D. (2005). Fully conditional specification in multivariate imputation. Technical report.

van Oirbeek, R. and Lesaffre, E. (2010). An application of harrell's c-index to ph frailty models. *Statistics in Medicine*, 29:31603171.

Vaupel, J., Manton, K., and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Scandinavian Statistics in Medicine*, 16:43954.

Vergouwe, Y., Steyerberg, E. W., Eijkemans, M. J., and Habbema, J. D. (2002). Validity of prognostic models: when is a model clinically useful? *Lancet Neurol*, 20:96–107.

Wienke, A. (2011). *Frailty Models in Survival data*. Chapman and Hall, New York, 1 edition.

Yamaguchi, T., Ohashi, Y., and Matsuyama, Y. (2002). Proportional hazards models with random effects to examine centre effects in multicentre cancer clinical trials. *Statistics Methods in Medical Research*, 11:221236.

Ying, G. S. and Liu, C. (2006). Statistical analysis of clustered data using sas. Technical report, Center for Preventive Ophthalmology and Biostatistics, Department of Ophthalmology, University of Pennsylvania.

Appendix A

Analysis Code

A.1 R-function for calculation of within and between c-indices

```
library(foreign)
library(coxme)
age<-train["age"][,1]
motor_gcs<-train["motor_gcs"][,1]
pupils<-train["pupils"][,1]
time<-train["time"][,1]
event<-train["event"][,1]
hospital_id<-train["hospital_id"][,1]

#Define the survival object we will use later on
SurvObject <- Surv(time,event)

#Fit the data using the function coxme with hospital_id as the frailty term
fit <- coxme(SurvObject~age+pupils+factor(motor_gcs)+(1|hospital_id), x=T, data=train)
x.fit <- fit$x

x<-cbind(age, pupils,motor_gcs)

#Coefficients of the covariates
beta     <- matrix(nrow=1,ncol=6)
beta[1,] <- fit$coef$fixed

#The frailty terms
n.cluster <- length(unique(hospital_id))
w <- exp(fit$frail$hospital_id)

C <- frailty.c(x.fit,beta,w,time,event,hospital_id)
```

Appendix B

Additional Analysis

B.1 Additional Analysis

Table B.1: Calibration slopes for the core model based on the test set

frailty	1.41	1.50	1.18	0.90	0.86	1.31	0.90	0.62	0.87	1.10
naïve	1.28	1.40	1.08	0.77	0.75	1.04	0.81	0.55	0.79	0.96

Table B.2: Parametric form for extended model on main data set

Parameter	**Marginal(Naïve; Robust)**	**Frailty**	**Fixed**
(Intercept)	11.452 (1.915 ; 1.473)	10.835 (1.5)	12.982 (2.25)
age	-0.061 (0.017 ; 0.011)	-0.054 (0.011)	-0.055 (0.011)
pupils	-2.071 (0.288 ; 0.47)	-1.43 (0.465)	-1.387 (0.48)
as.factor(motor_gcs)2	-1.459 (0.913 ; 0.683)	-0.376 (0.66)	-0.128 (0.678)
as.factor(motor_gcs)3	-0.484 (0.738 ; 0.724)	0.189 (0.707)	0.429 (0.732)
as.factor(motor_gcs)4	0.587 (0.814 ; 0.688)	1.656 (0.683)	2.025 (0.71)
as.factor(motor_gcs)5	2.982 (0.756 ; 0.649)	4.071 (0.657)	4.481 (0.691)
factor(neh)2	0.896 (0.536 ; 0.424)	1.122 (0.403)	1.197 (0.414)
factor(otvbs)2	3.01 (0.47 ; 0.441)	2.584 (0.419)	2.627 (0.432)
factor(cc)2	0.902 (0.718 ; 0.426)	0.912 (0.411)	0.997 (0.426)
factor(ns)2	-3.718 (0.741 ; 1.057)	-3.155 (1.019)	-3.158 (1.05)
Log(scale)	1.151 (0.065 ; 0.058)	1.085 (0.056)	1.109(0.057)

Table B.3: Cox Assumption for the extended models

	Fixed Effect			**naïve Cox PH**		
Parameter	rho	chisq	p	rho	chisq	p
age	0.014	0.062	0.803	0.057	0.91	0.34
as.factor(motor_gcs)2	0.041	0.435	0.509	-0.089	2.155	0.142
as.factor(motor_gcs)3	0.015	0.06	0.807	0.079	1.603	0.205
as.factor(motor_gcs)4	0.012	0.037	0.848	0.02	0.104	0.748
as.factor(motor_gcs)5	0.034	0.351	0.553	0.013	0.044	0.834
pupils	-0.104	3.13	0.077	0.051	0.709	0.4
factor(neh)2	0.07	1.25	0.263	0.054	0.705	0.401
factor(otvbs)2	0.009	0.023	0.879	-0.019	0.095	0.758
factor(cc)2	-0.059	0.964	0.326	-0.085	1.943	0.163
factor(ns)2	-0.109	3.08	0.079	-0.112	3.078	0.079
factor(hospital_id)774	-0.037	0.348	0.555			
factor(hospital_id)776	0.009	0.021	0.885			
factor(hospital_id)802	-0.054	0.756	0.385			
factor(hospital_id)815	-0.075	1.47	0.225			
factor(hospital_id)847	-0.027	0.186	0.666			
factor(hospital_id)854	0	0	0.996			
factor(hospital_id)1838	-0.126	4.15	0.042			
factor(hospital_id)1941	-0.071	1.32	0.25			
factor(hospital_id)2113	-0.037	0.351	0.554			
GLOBAL	NA	39.1	0.004	NA	12.226	0.27

Table B.4: Core model for full data

Parameter	**Marginal(Naïve; Robust)**	**Frailty**	**Fixed effect**
age	0.02 (0.002 ; 0.002)	0.022 (0.002)	0.023 (0.002)
pupils	0.847 (0.072 ; 0.086)	0.8 (0.075)	0.8 (0.076)
as.factor(motor_gcs)2	0.5 (0.113 ; 0.135)	0.289 (0.12)	0.18 (0.122)
as.factor(motor_gcs)3	0.146 (0.114 ; 0.126)	-0.017 (0.122)	-0.105 (0.126)
as.factor(motor_gcs)4	-0.396 (0.114 ; 0.127)	-0.539 (0.121)	-0.619 (0.124)
as.factor(motor_gcs)5	-1.126 (0.104 ; 0.119)	-1.385 (0.113)	-1.508 (0.116)

Table B.5: Extended Model for full data

Parameter	Marginal(Naïve; Robust)	Frailty	Fixed
age	0.017 (0.002 ; 0.002)	0.019 (0.002)	0.02 (0.002)
pupils	0.2 (0.122 ; 0.135)	0.095 (0.127)	-0.005 (0.131)
as.factor(motor_gcs)2	-0.088 (0.123 ; 0.132)	-0.11 (0.131)	-0.18 (0.136)
as.factor(motor_gcs)3	-0.539 (0.123 ; 0.132)	-0.515 (0.131)	-0.594 (0.135)
as.factor(motor_gcs)4	-1.137 (0.113 ; 0.132)	-1.24 (0.122)	-1.366 (0.127)
as.factor(motor_gcs)5	0.607 (0.078 ; 0.081)	0.613 (0.081)	0.643 (0.083)
factor(neh)2	-0.287 (0.069 ; 0.083)	-0.312 (0.072)	-0.309 (0.074)
factor(otvbs)2	-0.692 (0.071 ; 0.121)	-0.81 (0.077)	-0.806 (0.081)
factor(cc)2	-0.297 (0.069 ; 0.074)	-0.278 (0.073)	-0.271 (0.075)
factor(ns)2	1.119 (0.183 ; 0.186)	1.102 (0.186)	1.083 (0.188)

B.1.1 Variable Names

Table B.6: Variable names

FIELD NAME	DESCRIPTION	CODES
pcode	Unique patient ID	From 1 to 10792 (not all integers included)
pupils	Pupil reactiveness	0 = Both pupils reactive to light 1 = One or both pupils un-reactive
gcs_total	Total GCS score	Total GCS score (from 3 to 15)
dob	Date of birth	Date of birth (dd/mm/yyyy)
gender	Gender	1=Female 2=Male
cause	Mechanism of injury	1=RTA 2=Fall ¿2metres 3=Other
major_extracranial_injury	Major extracranial injury	1 = Yes 2 = No
ct scan	Head CT scan performed	1 = Yes 2 = No
ns	Normal scan	1 = Yes 2 = No 8 = Not applicable
nesrip	Abnormal scan: One or more petechial haemorrhages within the brain	1 = Yes 2 = No 8 = Not applicable
otvbs	Abnormal scan: Obliteration of the third ventricle or basal cisterns	1 = Yes 2 = No 8 = Not applicable
sb	Abnormal scan: Subarachnoid bleed	1 = Yes 2 = No 8 = Not applicable
mso5	Abnormal scan: Midline shift over 5mm	1 = Yes 2 = No 8 = Not applicable
neh	Abnormal scan: Non evacuated haematoma	1 = Yes 2 = No 8 = Not applicable
eh	Abnormal scan: Evacuated haematoma	1 = Yes 2 = No 8 = Not applicable
cc	Cortical Contusion	1 = Yes 2 = No 8 = Not applicable
hospital_id	Hospital CRASH ID	Unique hospital ID (from 3 to 2362)
income	High/Mid-Low income country	0 = High income 1 = Low or Middle income

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Do we need frailty modeling in the development of prognostic models?

Richting: **Master of Statistics-Biostatistics**
Jaar: **2011**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Chebon, Sammy

Datum: **12/09/2011**