Parametric Joint Modelling for Longitudinal and Survival Data

Paraskevi Pericleous

Doctor of Philosophy
School of Computing Sciences
University of East Anglia

July 2016

Abstract

Joint modelling is the simultaneous modelling of longitudinal and survival data, while taking into account a possible association between them. A common approach in joint modelling studies is to assume that the repeated measurements follow a linear mixed effects model and the survival data is modelled using a Cox proportional hazards model. The Cox model, however, requires a strong proportionality assumption, which seems to be violated quite often. We, thus, propose the use of parametric survival models. Additionally, joint modelling literature mainly deals with right-censoring only and does not consider left-truncation, which can cause bias. The joint model proposed here considers left-truncation and right-censoring.

Contents

List of Figures

List of Tables

Acknowledgements

When I was starting the PhD I never thought that I would be so exhausted by the end of it. To all the people that helped me in any way, a huge thank you.

Separate thanks to my husband who made my PhD more difficult and my life happier...

1 Literature Review

1.1 Introduction

This chapter describes the fundamentals of time-to-event and longitudinal analysis. For time-to-event data the basic quantities, missing data mechanisms, the methods for time-to-event data and diagnostics are described. For longitudinal analysis the missing data mechanisms and the most common model used are mentioned. For joint analysis, different approaches to parameter estimation and different modelling strategies are discussed. A gap in the research of joint modelling to do with the use of parametric time-to-event models in joint modelling is identified. This thesis will try to fill in that gap.

1.2 Time-To-Event Analysis

Survival Analysis (also known as time-to-event analysis) is the analysis of time from a start point until the event of interest occurs (20). The survival time is the difference between the two time-points i.e. the starting point and the endpoint (63).

1.2.1 Basic Quantities

Let T be a continuous random variable taking on values on $[0, \infty]$, which denotes time interval until a predetermined event occurs. This event could be cancer appearance or cancer metastasis, start time of smoking, death, etc.

1.2.1.1 Cummulative Distribution Function

The cummulative distribution function is the probability that the event will occur before time t, (85), (68):

$$F(t) = P(T \leq t). \tag{1}$$

This is an increasing function of t, taking values from 0 to 1.

1.2.1.2 Survival Function

The survival function at time t is the probability that the event will not occur before this time t (141), or the probability of surviving beyond time t (9):

$$S(t) = P(T > t), \tag{2}$$

where $t \geq 0$. Survival function is a decreasing function equal to 1 at $t = 0$ and as the time tends to infinity, it approaches zero (1).

The relation between the cummulative distribution function and the survival function is:

$$S(t) = 1 - F(t). \tag{3}$$

Thus, the survival function is the complement of the cummulative distribution function.

1.2.1.3 Probability Density Function

The probability density function is the derivative of the cummulative distribution function(6):

$$f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt}. \tag{4}$$

1.2.1.4 Hazard Function

The hazard function is the instantaneous probability of an event to occur within a small time interval $t + \Delta t$ given that it has not occurred until time

t, (71), (68):

$$h(t) = \lim_{\Delta t \to 0} \frac{P[t \leq T < t + \Delta t | T \geq t]}{\Delta t}. \tag{5}$$

It is also called "conditional hazard rate" or "force of mortality" (73). The hazard function can be obtained by dividing the density $f(t)$ by the survival function $S(t)$ (52):

$$h(t) = \frac{f(t)}{S(t)}. \tag{6}$$

1.2.1.5 Cummulative Hazard Function The cummulative hazard function is the integrated hazard function up to t (93) (68):

$$H(t) = \int_0^t h(u)du = -\log\left[S(t)\right]. \tag{7}$$

1.2.2 Missing Data Mechanisms

"Survival time" is the distance between the starting point and time when the event takes place. There are cases, where there is a lack of observation of survival time. This lack of observation is caused by two different mechanisms. The lack of observation may be due to random factors (censoring) for each subject, or due to the selection process (truncation) in the study (63). There are various cases of censoring and truncation, which are discussed below.

Suppose that a subject takes part in a study. Suddenly, he decides to move to another city. His last visit contains the latest information that is held about him. After he moved, the information about what happened to him is missing. So, there is a lack of observation on the right tail of the time-axis. This lack of observation is called right-censoring (63). Let C_i and T_i denote the censoring and survival times respectively, observed for the i^{th}

individual. Censoring indicator $\delta_i = I(T_i \leq C_i)$, where I is the indicator of an event, thus when right-censoring is present i.e. $C_i < T_i$, the censoring indicator $\delta_i = 0$. When the subject dies and is not right-censored $T_i \leq C_i$, the censoring indicator $\delta_i = 1$.

Consider a study where the subjects are interviewed about the age when they have first used marijuana. One possible answer is that the respondents have used it, but they cannot remember when this happened. So, the event of interest occurred at an age, prior to the age that the interview is done. The exact age, however, is not known (71). This type of lack of time observation is called left-censoring (63).

Another type of censoring, is the interval censoring. This happens when the subjects are contacted at time intervals. Suppose that the individuals are contacted every two months. The event of interest has not occurred by the second month. By the fourth month, however, it has occurred. So, it is known that the event of interest has occurred at some time between the second and fourth months, but the exact date is not known (63).

The left-truncation is caused by delayed entry of the individuals to the study. The most common case of left-truncation is the retirees taking retirement from the age of 60+ and their age is recorded at that point. However, there are individuals that died before they retired and did not make it to the retirement. This type of incomplete observation is called left-truncation (71). The survival of the individuals is conditional on the survival up to point of the study entry.

The right truncation happens when the selection process of the study stops at some point. For example, if a study deals with people who got infected within a specific time period, there is a lack of observation for infections past this specific time period. This lack of observation is considered to be right-truncated (71).

1.2.3 Non-Parametric Methods of Survival Analysis

The main characteristic of the non-parametric methods is that no explicit assumptions are made about the distribution of the survival times (20).

1.2.3.1 Kaplan-Meier Product Limit Estimator The most common tool for non-parametric survival analysis is the Kaplan-Meier product limit estimator, which was suggested by Kaplan and Meier (1958) (69).

If there are p distinct survival time observations, they are put in ascending order $t_1, t_2, ..., t_p$. Let n_{t_i} be the number of individuals at risk at each t_i and let d_{t_i} be the number of individuals who died, at each t_i, where $i = 1, 2, ..., p$. The Kaplan-Meier estimator of the survival function is (71) (69):

$$\hat{S}(t) = \prod_{t_i \leq t}(1 - \frac{d_{t_i}}{n_{t_i}}), \quad t \geq 0. \tag{8}$$

At the start of each study i.e. at t_0, $d_{t_0} = 0$, and the estimator is equal to one.

1.2.4 Semi-Parametric Methods of Survival Analysis

A popular statistical method used for semi-parametric survival analysis is the Cox proportional hazards model, suggested by Cox (1972) (24). This is a regression modelling the hazard as:

$$h(t, x, \beta) = h_0(t) \exp\{x\beta\}, \tag{9}$$

where $h_0(t)$ is the baseline hazard, x denotes a vector of covariates and β the vector of regression coefficients, linking the covariates to the hazard function.

The proportional hazards model leaves the baseline hazard undefined, but it assumes that the ratio of the hazards is constant over time i.e. does not depend on time (63).

To analyse the proportional hazards model, Cox (1972) (24), suggested the partial likelihood function:

$$L_p(\beta) = \prod_{i=1}^{n} [\frac{\exp\{x_i\beta\}}{\sum_{j\in R(t_i)} \exp\{x_j\beta\}}]^{\delta_i}, \tag{10}$$

where δ_i is the censoring indicator and the summation in the denominator is over all subjects in the set of risk ($R(t_i)$) at time t_i. (10) assumes that there are no tied times.

Breslow (1974) (14) and Efron (1977) (35) suggested approximations for the partial likelihood function, taking into account tied times. These are:

$$L_{p1}(\beta) = \prod_{i=1}^{n} [\frac{\exp\{x_{(i)+}\beta\}}{[\sum_{j\in R(t_i)} \exp\{x_j\beta\}]^{\delta_i}}] \tag{11}$$

and

$$L_{p2}(\beta) = \prod_{i=1}^{n} [\frac{\exp\{x_{(i)+}\beta\}}{\prod_{k=1}^{d_i} [\sum_{j\in R(t_{(i)})} \exp\{x_j\beta\} - \frac{k-1}{d_i} \sum_{j\in D(t_{(i)})} \exp\{x_j\beta\}]}]. \tag{12}$$

respectively. d_i denotes the number of subjects with survival time $t_{(i)}$, $x_{(i)+}$ is equal to the sum of the covariate values over the d_i subjects and $D(t_{(i)})$ denotes the individuals that have survival times equal to $t_{(i)}$.

Despite Cox proportional hazards model's popularity, its proportionality assumption often fails to be satisfied. Thus, there is a need for other models that do not use so strong assumptions.

1.2.5 Parametric Methods of Survival Analysis

Parametric methods make use of the assumption that the survival times come from a specific distribution (20) (71) (63). The main distributions used are: Exponential, Weibull, Log-Logistic, Log-Normal, Extreme Value and Logistic. Additionally, Gompertz, Perks, Beard, Makeham, Makeham-Perks and Makeham-Beard distributions, are popular on actuarial applications and mostly used for the cases of left-truncation and right-censoring. These models are widely popular, especially in actuarial applications. The reasons for this popularity are that the effect of the variables on survival time is modelled explicitly, that full likelihood is used and that different shapes of the hazard are allowed.

In case of right-censoring the likelihood is given by (9) (20) (21) (63) (71):

$$L(\theta, \delta_i; t_i) = \prod_{i=1}^{n} \{f(t_i;\theta)\}^{\delta_i} \{S(t_i;\theta)\}^{1-\delta_i}, \tag{13}$$

or equivalently it can be also given by (138):

$$L(\theta, \delta_i; t_i) = \prod_{i=1}^{n} \{h(t_i;\theta)\}^{\delta_i} \{S(t_i;\theta)\}, \tag{14}$$

where n is the number of individuals, θ denotes all the parameters that need to be estimated, t_i is the follow-up time of the i^{th} individual and δ_i is the censoring indicator of the i^{rh} individual.

In case that left-truncation is present in the data, along with right-censoring, all probabilities are changed to conditional ones. To be specific, if the truncation time for the i^{th} individual is denoted by A_i, the probability density and survival function are replaced by $\frac{f(t_i;\theta)}{S(A_i;\theta)}$ and $\frac{S(t_i;\theta)}{S(A_i;\theta)}$ respectively (71).

Therefore, for left-truncation and right-censoring, the likelihood changes to (71):

$$L(\theta, \delta_i; t_i) = \prod_{i=1}^{n} (\frac{f(t_i;\theta)}{S(A_i;\theta)})^{\delta_i} (\frac{S(t_i;\theta)}{S(A_i;\theta)})^{1-\delta_i}. \tag{15}$$

The difference between the two likelihoods (13) and (15), is that the left-truncated and right-censored likelihood is divided by the survival functions of the truncation times.

1.2.5.1 Weibull Distribution Weibull and Stockholm (1951) (146) suggested the use of Weibull distribution. The hazard, cummulative hazard, survival and probability density function of the Weibull distribution are given by (71) (92):

$$h(t;\sigma,\gamma) = \gamma\sigma^{-1}(t\sigma^{-1})^{\gamma-1}, \tag{16}$$

$$H(t;\sigma,\gamma) = (t\sigma^{-1})^{\gamma}, \tag{17}$$

$$S(t;\sigma,\gamma) = \exp\left\{-(t\sigma^{-1})^{\gamma}\right\}, \tag{18}$$

and

$$f(t;\sigma,\gamma) = \gamma\sigma^{-1}(t\sigma^{-1})^{\gamma-1}\exp\left\{-(t\sigma^{-1})^{\gamma}\right\}, \tag{19}$$

respectively. $\sigma > 0$ and $\gamma > 0$ are the distribution parameters, scale and shape, respectively, and t is the follow-up time. Weibull is only defined for $t \geq 0$. When $\gamma = 1$, the Weibull distribution becomes Exponential distribution (87), which is a special case of it. The hazard function (16) becomes

$$h(t;\sigma) = \sigma^{-1}, \tag{20}$$

i.e. $h(t)$ is constant over time.

Weibull data can be modelled both parametrically, and semi-parametrically. This is due to the fact that the Weibull distribution has the property of the proportional hazards i.e.

the Weibull hazard function can take the form

$$h(t, x, \beta) = h_0(t) \exp\{x\beta\}, \tag{21}$$

where $h_0(t)$ is the baseline hazard, x denotes a vector of covariates and β the vector of regression coefficients, linking the covariates to the hazard function. Setting $\lambda = \sigma^{-\gamma}$ and $\lambda = \exp(\beta_0 + \beta x)$, (16) becomes $h(t, x, \beta) = \gamma t^{\gamma-1} \exp(\beta_0) \exp(\beta x)$, where $h_0(t) = \gamma t^{\gamma-1} \exp(\beta_0)$

The proportional hazards model leaves the baseline hazard undefined, but it assumes that the ratio of the hazards is constant over time i.e. does not depend on time (63). However, parametric methods have other advantages like making use of the assumption that the survival times come from a specific distribution (20) (71) (63), the effect of the variables on survival time is modelled explicitly, that full likelihood is used and that different shapes of the hazard are allowed.

1.2.5.2 Log-Logistic Distribution The hazard, cummulative hazard, survival and probability density functions of the Log-Logistic distribution are given by (13) (63) (71):

$$h(t; \lambda, p) = \lambda p t^{p-1} (1 + \lambda t^p)^{-1}, \tag{22}$$

$$H(t; \lambda, p) = \log[1 + \lambda t^p], \tag{23}$$

$$S(t; \lambda, p) = (1 + \lambda t^p)^{-1}, \tag{24}$$

and

$$f(t; \lambda, p) = (\lambda p t^{p-1})(1 + \lambda t^p)^{-2}, \tag{25}$$

respectively. $\lambda > 0$ and $p > 0$ are the distribution parameters, scale and shape, respectively. It is only defined for $t \geq 0$.

1.2.5.3 Log-Normal Distribution The hazard, cummulative hazard, survival and probability density function of the Log-Normal distribution are given by (40) (63) (71):

$$h(t;\mu,\sigma) = \exp\left\{-(\log\{t\}-\mu)^2 2^{-1}\sigma^{-2}\right\} t^{-1}\sigma^{-1}(2\pi)^{-\frac{1}{2}} \left\{1-\Phi[\frac{\log t-\mu}{\sigma}]\right\}^{-1}, \tag{26}$$

$$H(t;\mu,\sigma) = -\log\left\{1-\Phi[(\log t-\mu)\sigma^{-1}]\right\}, \tag{27}$$

$$S(t;\mu,\sigma) = 1-\Phi[(\log\{t\}-\mu)\sigma^{-1}], \tag{28}$$

and

$$f(t;\mu,\sigma) = \exp\left\{-(\log\{t\}-\mu)^2 2^{-1}\sigma^{-2}\right\}(t\sigma)^{-1}(2\pi)^{-\frac{1}{2}}, \tag{29}$$

respectively. $\mu \in \Re$ and $\sigma > 0$ are the distribution parameters, location and scale respectively. $\Phi(z)$ is the cummulative distribution function of the standard normal distribution (71).

1.2.5.4 Extreme Value Distribution The hazard, cummulative hazard, survival and probability density function of the Extreme distribution, are given by (81) (106) (148):

$$h(t;\mu,\sigma) = \sigma^{-1}\exp\left\{(t-\mu)\sigma^{-1}\right\}, \tag{30}$$

$$H(t;\mu,\sigma) = \exp\left\{(t-\mu)\sigma^{-1}\right\}, \tag{31}$$

$$S(t;\mu,\sigma) = \exp\left\{-\exp\left\{(t-\mu)\sigma^{-1}\right\}\right\} \tag{32}$$

and

$$f(t;\mu,\sigma) = \sigma^{-1}\exp\left\{(t-\mu)\sigma^{-1}-\exp\left\{(t-\mu)\sigma^{-1}\right\}\right\}, \tag{33}$$

respectively. $\mu \in \Re$ and $\sigma > 0$ are the distribution parameters, location and scale respectively. t is defined from $-\infty$ to ∞. We deal with log-Extreme Value, which is essentially Weibull, see above for the Weibull.

1.2.5.5 Logistic Distribution The hazard, cummulative hazard, survival and probability density functions of the Logistic distribution are given by (106) (148):

$$h(t;\mu,\sigma) = \sigma^{-1}\exp\left\{(t-\mu)\sigma^{-1}\right\}\left[1+\exp\left\{(t-\mu)\sigma^{-1}\right\}\right]^{-1}, \tag{34}$$

$$H(t;\mu,\sigma) = \log[1+\exp\left\{(t-\mu)\sigma^{-1}\right\}], \tag{35}$$

$$S(t;\mu,\sigma) = \left[1+\exp\left\{(t-\mu)\sigma^{-1}\right\}\right]^{-1} \tag{36}$$

and

$$f(t;\mu,\sigma) = \sigma^{-1}\exp\left\{(t-\mu)\sigma^{-1}\right\}\left[1+\exp\left\{(t-\mu)\sigma^{-1}\right\}\right]^{-2}, \tag{37}$$

respectively. $\mu \in \Re$ and $\sigma > 0$ are the distribution parameters, location and scale respectively. t is defined from $-\infty$ to ∞. We deal with Log-Logistic, which is only defined for $t \geq 0$.

Log-Logistic is a survival distribution, but Logistic is not.

1.2.5.6 Gompertz Distribution Benjamin Gompertz, a British actuary, introduced a law of mortality (48), which is nowadays called the Gompertz law of mortality (22) (86). He assumed the exponential increase of mortality with age. The hazard function and cummulative hazard function of the Gompertz law are (106) (107) (71):

$$h(t) = t\exp(at) \tag{38}$$

and

$$H(t) = -ta^{-1}[1-\exp(at)], \tag{39}$$

respectively. t the follow-up time and $a > 0$ are the distribution parameters. The survival function is given by:

$$S(t) = \exp\left(ta^{-1}[1-\exp(at)]\right). \tag{40}$$

1.2.5.7 Perks Distribution Perks (1932) (98) found empirically that the mortality could be approximated by a logistic curve. The hazard function, cummulative hazard function and survival function for the Perks mortality law are given by (106) (107):

$$h_x(t) = \frac{\exp\{a+bx\}}{1+\exp\{a+bx\}}, \tag{41}$$

$$H_x(t) = b^{-1}\log\left\{\frac{1+\exp\{a+b(x+t)\}}{1+\exp\{a+bx\}}\right\} \tag{42}$$

and

$$S_x(t) = \left\{\frac{1+\exp\{a+b(x+t)\}}{1+\exp\{a+bx\}}\right\}^{-b^{-1}}, \tag{43}$$

respectively, where $a, b \in \Re$.

1.2.5.8 Beard Distribution Beard (1959) (10) added to Perks distribution (1932) (98) a heterogeneity parameter ρ. The hazard, cummulative hazard and survival function of the Beard mortality law are given by:

$$h_x(t) = \frac{\exp\{a+bx\}}{1+\exp\{a+\rho+bx\}}, \tag{44}$$

$$H_x(t) = \exp\{-\rho\}b^{-1}\log\left\{\frac{1+\exp\{a+\rho+b(x+t)\}}{1+\exp\{a+\rho+bx\}}\right\} \tag{45}$$

and

$$S_x(t) = \left\{\frac{1+\exp\{a+\rho+b(x+t)\}}{1+\exp\{a+\rho+bx\}}\right\}^{-\exp\{-\rho\}b^{-1}}, \tag{46}$$

respectively, where $a, b \in \Re$.

1.2.5.9 Makeham Distribution Makeham (1859) (84) found that Gompertz's law could be improved by adding a constant term. The hazard, cummulative hazard and survival function are given by (106) (107):

$$h_x(t) = \exp\{\epsilon\} + \exp\{a+bx\}, \tag{47}$$

$$H_x(t) = t\exp\{\epsilon\} + b^{-1}(\exp\{bt\} - 1)\exp\{a + bx\} \tag{48}$$

and

$$S_x(t) = \exp\left\{-[t\exp\{\epsilon\} + b^{-1}(\exp\{bt\} - 1)\exp\{a + bx\}]\right\}, \tag{49}$$

where $a, b, \epsilon \in \Re$.

1.2.5.10 Makeham-Perks Distribution Makeham-Perks distribution was empirically derived by Perks (1932) (98). Using Richards (2008) (106) and Richards (2011) (107) notation, the hazard, cummulative hazard and survival function are given by:

$$h_x(t) = \frac{\exp\{\epsilon\} + \exp\{a + bx\}}{1 + \exp\{a + bx\}}, \tag{50}$$

$$H_x(t) = t\exp\{\epsilon\} + (1 - \exp\{\epsilon\})b^{-1}\log\left\{\frac{1 + \exp\{a + b(x + t)\}}{1 + \exp\{a + bx\}}\right\} \tag{51}$$

and

$$S_x(t) = \exp\{-t\exp\{\epsilon\}\}\left\{\frac{1 + \exp\{a + b(x + t)\}}{1 + \exp\{a + bx\}}\right\}^{(1-\exp\{\epsilon\})b^{-1}}, \tag{52}$$

where $a, b, \epsilon \in \Re$.

1.2.5.11 Makeham-Beard Distribution Makeham-Beard distribution was found empirically by Perks (1932) (98). Using Richards (2008) (106) and Richards (2011) (107) notation, the hazard, cummulative hazard and survival function are given by:

$$h_x(t) = \frac{\exp\{\epsilon\} + \exp\{a + bx\}}{1 + \exp\{a + \rho + bx\}}, \tag{53}$$

$$H_x(t) = t\exp\{\epsilon\} + (\exp\{-\rho\} - \exp\{\epsilon\})b^{-1}\log\left\{\frac{1 + \exp\{a + \rho + b(x + t)\}}{1 + \exp\{a + \rho + bx\}}\right\} \tag{54}$$

and

$$S_x(t) = \exp\{-t\exp\{\epsilon\}\}\left\{\frac{1 + \exp\{a + \rho + b(x + t)\}}{1 + \exp\{a + \rho + bx\}}\right\}^{(\exp\{-\rho\}-\exp\{\epsilon\})b^{-1}}, \tag{55}$$

where $a, b, \rho, \epsilon \in \Re$.

1.2.6 Maximum Likelihood Estimation

Let n be the number of subjects. If $t_1, t_2, ..., t_n$ is a set of independent observations with probability density function $f_i(t_i; \theta)$, where θ is a vector of J parameters, $i = 1, 2, ..., n$, the likelihood function is defined as (145):

$$L(\theta; t_1, t_2, ...t_n) = \prod_{i=1}^{n} f(t_i; \theta). \tag{56}$$

However, the log-likelihood function is easier to use due to the properties of the logarithms.

$$l(\theta; t_1, t_2, ..., t_n) = \log L(\theta; t_1, t_2, ...t_n). \tag{57}$$

In maximum likelihood estimation, the aim is to obtain the value of the estimates that maximize the log-likelihood. This is done by differentiating the log-likelihood with respect to each parameter, setting the derivative equal to zero and solving the resulting equation, in order to obtain the maximum likelihood estimate $\hat{\theta}$ (78):

$$\frac{\partial l(\hat{\theta}; t)}{\partial \theta_j} = 0, \quad j = 1, ..., J. \tag{58}$$

In the above equations $l(\theta; t)$ denotes the log-likelihood, t is the set of times $t = \{t_1, t_2, ..., t_n\}$ for the individuals, θ is the set of θ_j parameters, $\theta = \{\theta_1, \theta_2, ..., \theta_J\}$.

The Hessian matrix is given by (63):

$$H[l(\theta; y)] = \begin{bmatrix} \frac{\partial^2 l(\theta;t)}{\partial \theta_1^2} & \frac{\partial^2 l(\theta;t)}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 l(\theta;t)}{\partial \theta_1 \partial \theta_j} \\ \frac{\partial^2 l(\theta;t)}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 l(\theta;t)}{\partial \theta_2^2} & \cdots & \frac{\partial^2 l(\theta;t)}{\partial \theta_2 \partial \theta_j} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 l(\theta;t)}{\partial \theta_j \partial \theta_1} & \frac{\partial^2 l(\theta;t)}{\partial \theta_j \partial \theta_2} & \cdots & \frac{\partial^2 l(\theta;t)}{\partial \theta_j^2} \end{bmatrix}. \tag{59}$$

The information matrix is (31):

$$I(\hat{\theta}) = -E[H[l(\theta; y)]] \tag{60}$$

and the approximation of variance-covariance matrix is the inverse of the information matrix (8) (123):

$$\widehat{Var}(\hat{\theta}) = I^{-1}(\hat{\theta}). \tag{61}$$

1.2.7 Diagnostics for Survival Analysis Models

When a model is fitted to a data, certain assumption are made. Before making any conclusions from the model fitted, it is crucial to ensure that all the model assumptions made are valid. In case that assumption violations do occur, then the model may provide faulty conclusions. Thus, it is extremely important to perform the appropriate model diagnostics.

1.2.7.1 Cox-Snell Residuals The Cox-Snell residuals are used for assessing the overall fit of the model (127). They are defined as (73) (127):

$$\hat{c}_i = \hat{H}(t_i) = -\log \hat{S}(t_i), \tag{62}$$

where $\hat{S}$ is the estimated survival function. After evaluating the Cox-Snell residuals, they can define time and censoring indicator can define the events (63). Using them, Kaplan-Meier estimator is obtained and from that the estimated cummulative hazard function of the Cox-Snell residuals (71). If the latter is plotted against the Cox-Snell residuals, the overall fit can be seen. If the values obtained are approximately forming a straight line that passes through the origin and has gradient equal to one, then the model is considered to be adequate (73) (89).

1.2.7.2 Martingale Residuals The martingale residuals are a modification of the Cox-Snell residuals and are defined as (63):

$$\hat{M}_i = \delta_i - \hat{c}_i, \tag{63}$$

where δ_i is the event indicator (71). Large martingale values indicate a poor fit of the model for these points (41).

1.2.7.3 Deviance Residuals The deviance residuals are defined as (71) (32) (23)

$$D_i = (\hat{M}_i)\sqrt{(-2(\hat{M}_i + \delta_i \log(\delta_i - \hat{M}_i)))}, \tag{64}$$

$(\hat{M}_i)$ in (64) is the signum function which is defined as follows (55):

$$(x) = \begin{cases} -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ 1 & \text{if } x > 0. \end{cases} \tag{65}$$

The distribution of the deviance residuals is symmetric around zero (121) due to the logarithm in the definition and thus, they are easier to work with than the martingale residuals which can take values from $-\infty$ to $+1$ (71) (32) (23). Generally, the deviance residuals identify the individuals fitted inadequately by the model (121).

1.2.7.4 Score Residuals The score residuals identify which subjects are highly affecting the estimation of the model (121) and they are defined as (63):

$$\frac{\partial l(\hat{\theta}; t)}{\partial \theta_j} = \sum_{i=1}^{n} \widehat{L_{ij}}, \tag{66}$$

where t is the set of t_i values for $i = 1, 2, ..., n$, θ is the set of θ_j, where $j = 1, 2, ..., J$, parameters to be estimated. The vector of the score residuals of the i^{th} subject corresponding to θ parameters is denoted as $\hat{L}_i = (\hat{l}_{i1}, \hat{l}_{i2}, ..., \hat{l}_{i\theta_j})$ (63).

Each score equation that contributes to the sum in (66), is in fact a re-weighted Schoenfeld residual (62).

1.2.7.5 Scaled Score Residuals The scaled score residuals are (63):

$$\hat{L}'_i = \widehat{Var(\theta)}\hat{L}_i. \tag{67}$$

The scaled score residuals measure how influential an observation is, in terms of parameter estimation (scaled version of the score residuals (63)) (37).

1.2.7.6 Cook's Distance Cook's distance is the influence that a subject may have on the estimation of the coefficients (83) and it can be evaluated using the score and scaled score residuals (63):

$$ld_i = \hat{L}_i\hat{L}'_i. \tag{68}$$

1.3 Longitudinal Analysis

Longitudinal data consists of repeated observations (126) on the same individual. This data is typically highly unbalanced due to the difference in number and timing of observations for each individual (139). This is because the conditions for each measurement cannot be fully controlled.

1.3.1 Linear Mixed Models

Laird and Ware (1982) (72) suggested a linear model for longitudinal data which includes the within-person and the between-person variation.

Let n be the number of individuals and m_i denote the number of observations for the i^{th} individual, where $i = 1, 2, ..., n$. Let Y_i be a $m_i \times 1$ vector of responses for the i^{th} individual. Let a denote a $p \times 1$ vector of unknown population parameters, X_i denote a known $m_i \times p$ design matrix linking a to Y_i, b_i be a $k \times 1$ vector of unknown random individual effects and Z_i be a known $m_i \times k$ design matrix linking b_i to Y_i.

For each individual i the model states that,

$$Y_i = X_i a + Z_i b_i + e_i, \tag{69}$$

where the errors e_i are assumed to be independent and normally distributed $N(0, R_i)$ (88) i.e. they have mean zero and $m_i \times m_i$ covariance matrix R_i (139). The covariance is a positive-definite matrix that depends on i, via its dimension m_i.

The random effects $b_i \sim N(0, D)$, where D is a $k \times k$ positive-definite covariance matrix (139) and are assumed to be independent of each other and of the errors e_i. The population parameters a are treated as fixed effects.

Conditionally on random effects b_i, $Y_i \sim N(X_i a + Z_i b_i, R_i)$ (139).

Marginally, $y_i \sim N(X_i a, R_i + Z_i D Z_i^T)$ (139). The model can be simplified when the random effects b_i are independent i.e. $R_i = \sigma^2 I$, where I is a $m_i \times m_i$ identity matrix. Then, the

marginal density function of Y_i is (139):

$$f(y_i) = \int f(y_i|b_i)f(b_i)\,db_i\,, \tag{70}$$

where $f(y_i|b_i)$ and $f(b_i)$ are the density functions of Y_i conditional on b_i and b_i respectively.

In order to fit the model and obtain the model parameters, maximum likelihood or restricted maximum likelihood estimation can be used (139).

1.3.2 Missing Data Mechanisms

Little and Rubin (1987) (80) and Diggle and Kenward (1994) (33) classified the missingness mechanisms for longitudinal measurements into three categories. These are:

1. MCAR- Missing Completely At Random: this happens when the probability of missing does not depend on observed or unobserved measurements.

2. MAR- Missing At Random: appears when the probability of missing depends on observed measurements, but not on the unobserved ones.

3. MNAR- Missing Not At Random: arises when the probability of missing data depends on observed and unobserved measurements.

If the drop-out (missing data) is present in the data, then ignoring the relationship between the drop-out and the response, is inappropriate (33). Hogan and Laird (1997b) (60) concluded that when the missing observations are not missing at random, then the missing

data has to be modelled with the longitudinal data, so that the estimates will be valid and not biased.

1.4 Joint Analysis of Longitudinal and Survival Data

Longitudinal data include the set of repeated measurements, and survival data includes the set of times until the event of interest, for the same patients. These two data subsets were analyzed separately in the past, until it was realised that the longitudinal data can be associated with the survival data (88).

A biomarker is a biological characteristic (e.g. molecule) which is used to identify a pathological or physiological process (e.g. disease) (125). Biomarkers are often measured over time and referred to as longitudinal biomarkers. They may affect survival.

The simple version of longitudinal data for joint modelling is to have one longitudinal biomarker. There are however, more complicated scenarios where there are multiple longitudinal biomarkers. Models that can handle different types of failing are called competing risks models. These different types of failing can also be considered as informative censoring. Multiple longitudinal biomarkers and competing risks models are briefly discussed in Chapter 6.

For the simple joint modelling with only one biomarker and without any competing risks, the first naive approaches to estimate the parameters were the last value carried forward which is known to cause large bias, Prentice (1982) (101), and the two-stage procedure, Self and Pawitan (1992) (118). A modern alternative is the simultaneous parameter estimation for both processes, longitudinal and survival, which is now widely used. Different modelling

strategies are used depending on the primary interest of the study (126).

1.4.1 Two-Stage Approach

The two-stage approach consists of estimating the longitudinal component first and then substituting them to estimate the survival component. Tsiatis et al (1995) (135) and Albert and Shih (2010) (4) use this approach with some variations.

Tsiatis et al (1995) (135) estimate the longitudinal components using repeated measures random component model in the first stage. In the second stage, they estimate the parameters of a Cox proportional hazards model, using empirical Bayes estimates as in Dafni and Tsiatis (1994) (28) and Laird and Ware (1982) (72).

Albert and Shih (2010) (4) used the two-stage approach for discrete time-to-event data. First, they estimated the longitudinal components, then simulated the data from the longitudinal component and re-fitted the model. They related the longitudinal component to the survival time, using a probit model. The simulation is performed to minimize bias due to informative drop out.

The two stage approach however does not perform as well as the simultaneous parameter estimation (129).

1.4.2 Simultaneous Joint Modelling

The simultaneous parameter estimation, using a joint likelihood approach has become popular due to an increased accuracy of estimation (129). The joint modelling paradigm includes different modelling strategies such as pattern-mixture and selection models (126). These can be extended by incorporating random effects and they are called random pattern-mixture, random selection and an additional category called random effects models. There are some differences in the definitions for each type of model and some overlaps between them. For more details see Little (1995) (79), Hogan and Laird (1997b) (60), Sousa (2011) (126) and McCrink (2013) (88).

Simple, yet detailed explanation used by Sousa (2011) (126) and McCrink (2013) (88) for different joint modelling strategies is shown below. Let Y, T and B represent the longitudinal, time-to-event and random effects parts.

Selection models: $[Y, T, B] = [B][Y|B][T|Y]$.

Pattern-Mixture models: $[Y, T, B] = [B][T|B][Y|T]$.

Random effects models: $[Y, T, B] = [B][Y|B][T|B]$.

Due to a better performance of the simultaneous joint modelling compared to the other methods such as last value carried forward or two-stage approach, this is the class of methods developed in this thesis. The primary interest is random effects (or shared parameter) models.

1.4.3 Pattern Mixture Models

In pattern mixture models, the main interest is centered on the longitudinal component (126).

Wu and Bailey (1989) (149) were the first to consider a pattern-mixture model, under informative missingness. They used a linear random effects model for the longitudinal variable and the random effects were conditional on the event time.

Hogan and Laird (1997a) (59) described a mixture model for longitudinal and time-to-event data. They assumed that the data is missing at random (unlike Wu and Bailey (1989)), and that the individuals are subject to right-censoring. The longitudinal component is modelled using a linear mixed effects model and they made no parametric assumptions for the time-to-event distribution. Instead they estimate multinomial probabilities with incomplete data for the survival component, as suggested by Cox and Oakes (1984) (25).

1.4.4 Selection Models

In selection models, the primary interest is the time-to-event component with the longitudinal component most commonly modelled by a linear mixed effects model (126). The time-to-event component is modelled using either a Cox proportional hazards or a probit model.

Brown et al.(2005) (17) used a proportional hazards model for the survival component. However, they proposed a cubic B-spline, instead of a linear mixed effects model, to model the longitudinal data. Instead of making assumptions about the behaviour of the longi-

tudinal component over time, they use cubic B-splines for greater flexibility. They also extended the model to accomodate multiple longitudinal response variables.

1.4.5 Random Effects Models

In random effects models (also called shared parameter models) the two processes, longitudinal and time-to-event, are linked through shared random effects, i.e. they are conditionally independent given the random effects. (88) (126).

Wu and Carroll (1988) (150) suggested a linear random effects model for two-treatment group data when informative censoring is present. They assumed that missingness process and staggered entry are non-informative and independent of right-censoring. The probability of being right-censored is given by a probit model conditional on the random effects.

Follmann and Wu (1995) (45) suggested a shared parameter joint model, which is a generalization of the model suggested by Wu and Carroll (1988) (150). Suppose Y_i are the longitudinal measurements for the i^{th} subject and T_i the time-to-event for the i^{th} subject. These are linked with a set of shared random effects b_i, which follow a distribution with mean zero and distribution function $H(\cdot)$.

The longitudinal value conditional on the random effects follows a generalized linear model:

$$\psi[E(y_{ij}|b_i)] = aw_{ij} + b_i z_{ij}, \tag{71}$$

where $\psi(\cdot)$ is the link function, w_{ij} and z_{ij} are the vectors of covariates. Additionally, it is assumed that the time-to-drop-out conditional on the random effects also follows a generalized linear model.

The full joint shared parameter model is:

$$f(Y_i, T_i) = \int g_{Y_i}(Y_i|b_i) g_{T_i}(T_i|b_i)\, dH(b_i)\,, \tag{72}$$

where $g_{Y_i}(Y_i|b_i)$ and $g_{T_i}(T_i|b_i)$ are the conditional models and $H(b_i)$ is the distribution function of the b_i. Follmann and Wu (1995) showed that this model does not require the data to be MAR or MCAR.

DeGruttola and Tu (1994) (30) proposed an extension of the model suggested by Wu and Carroll (1988) (150) i.e. a model that has a more general random effect structure, rather than two random effects only. The longitudinal component is modelled using a more general linear random effects model and random effects follow a multivariate normal distribution. Censoring of survival time is assumed to be non-informative. Specifically, the probability of being censored is independent of the unobserved failure time. Additionally, missing observations are missing at random. The relationship between the longitudinal component and survival component is modelled via linear or nonlinear regression, e.g. Cox proportional hazards model.

Faucett and Thomas (1996) (42) and Wulfsohn and Tsiatis (1997) (151) introduced a random effects model with proportional hazards. The parameters are estimated by using MCMC method of Gibbs sampling by Faucett and Thomas (1996) (42), while Wulfsohn and Tsiatis (1997) (151) introduce an EM algorithm.

The two most common submodels used for shared parameter models are, for longitudinal component

$$Y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + \epsilon_{ij}, \tag{73}$$

where $W_i(t_{ij}) = \beta_{0i} + \beta_{1i} t_{ij}$ is the true unobserved longitudinal value for the i^{th} individual at the j^{th} time-point, while Y_{ij} are the observed longitudinal measurements. Errors $\epsilon_{ij} \sim N(0, \sigma_e^2)$ while β_{0i} and β_{1i} have a bivariate normal distribution with means μ_{β_0} and μ_{β_1}

and the covariance matrix $\Sigma = \begin{bmatrix} \sigma^2_{\beta_0} & \sigma_{\beta_0\beta_1} \\ \sigma_{\beta_0\beta_1} & \sigma^2_{\beta_1} \end{bmatrix}$, where $\sigma_{\beta_0\beta_1} = \rho\sigma_{\beta_0}\sigma_{\beta_1}$ with ρ being the correlation between the random effects. For survival component,

$$\lambda_i(t) = \lambda_0(t)e^{\alpha(\beta_{0i}+\beta_{1i}t_{ij})}, \tag{74}$$

where α is the association between the longitudinal and the survival component. Additionaly, Wulfsohn and Tsiatis (1997) (151) discuss the two-stage modelling, which was suggested in Tsiatis, De Gruttola and Wulfsohn (1995) (135). Two-stage modelling was performed by first fitting a model to longitudinal measurements and then estimating the Cox model by substituting the empirical Bayesian estimates by Laird and Ware (1982) (72). This approach does not take into account the survival information, while fitting the longitudinal model and therefore, it cannot be more efficient than performing a direct maximization.

Henderson et al (2000) (56) extended the model used by Faucett and Thomas (1996) (42) and Wulfsohn and Tsiatis (1997) (151). They do not use the same $W_i(t_{ij})$, instead they used

$$W_i^{'}(t_{ij}) = \gamma_1 U_{0i} + \gamma_2 U_{1i} t_{ij} + \gamma_3(U_{0i} + U_{1i}t_{ij}) + U_{2i}, \tag{75}$$

where $U_{2i} \sim N(0, \sigma^2_{U_2})$ is independent of (U_{0i}, U_{1i}), while the rest remain the same as in Wulfsohn and Tsiatis (1997) (151). γ_1 and γ_2 measure the association between $W_i^{'}(t_{ij})$ and $W_i(t_{ij})$ through intercept and slope, while U_{2i} models frailty. They also suggest a Monte Carlo method for evaluating the likelihood integral for the reduction of the variance.

This type of model has as special case the model considered by Laird and Ware (1982) (72). By extending that model, Henderson et al (2000) (56) consider more situations where an association between longitudinal and survival component may exist. Henderson et al. (2000) (56) assume that the time, in which the measurements are taken, is non-informative and that censoring is also non-informative.

Other shared parameter models considered in the literature had different characteristics. Wang and Taylor (2001) (144) use a longitudinal model for continuous data and a proportional hazards model, which includes the longitudinal biomarker (as a time-dependent variable) and other covariates for survival component. The longitudinal model contained fixed and random effects, independent measurement error and an integrated Ornstein-Uhlenbeck (IOU) stochastic process. They use Bayesian techniques in order to fit the model. Their simulation results indicate that the method gives quite accurate estimates, except for the parameter of the IOU, for which they assume that it is due to a very skewed distribution of the parameter.

1.4.6 Integral Approximation and Parameter Estimation

Maximum likelihood is among the first estimation techiniques used for the joint model (88) (120). Expectation-Maximization (EM) algorithm is another popular algorithm that is often implemented (29) (88). It treats the random effects as missing data, see Wulfsohn and Tsiatis (1997) (151).

Joint modelling likelihood includes integrals that cannot be solved analytically. The main difficulty in joint modelling is the numerical integration with respect to the random effects. Thus numerical approximations are widely used, such as Gauss-Hermite rule and its adaptations (88). This is used by Wulfsohn and Tsiatis (1997) (151), Henderson et al.(2000) (56), Song et al.(2002) (124) and Rizopoulos et al.(2008) (108). However, when the dimension of the random effects increases, the computational complexity of the Gaussian Quadrature increases exponentially. Rizopoulos (2012) (113) suggest to use a pseudo-adaptive Gauss-Hermite Quadrature rule. Specifically, this is achieved, by first fitting the mixed effects model for the longitudinal component and extracting information about the location and the scale of the posterior distribution of the random effects. There is an issue,

though, with the quadrature points' location with respect to the main mass of the integrand's location and integrand spread may differ from the spread of the weight function. In this case, even for large number of quadrature points, the approximation is poor and this is due to the fact that quadrature is not located where the most of the mass of the integrand is. Thus, by centering and scaling the integrand in order to make the weight-function proportional to the density function, fewer quadrature points are required.

The centering and scaling are achieved using the location of the mode $\hat{b_i}$ and the second order derivative matrix $\hat{H_i}$ for each subject. So, after fitting the linear mixed effects model, the empirical Bayes estimates of $\tilde{b_i}$ and $\tilde{H_i}^{-1}$ are extracted and used in the transformation $\tilde{r_t} = \tilde{b_i} + \sqrt{2}\tilde{B^{-1}}b_t$, where $\tilde{B_i}$ denotes the Cholesky factor of $\tilde{H_i}$ and b_t denote the absciccas. As the number of observations for each individual increases, it is sufficient to use the information from the mixed effects model, i.e. the maximum likelihood estimates for the joint model will be close to the maximum likelihood estimates of the linear mixed model. Thus, this procedure is only required to be implemented once, at the start of optimization, and not at every quadrature point like the adaptive Gauss-Hermite rule. So, there is no need for computational relocation at each iteration. Using simulations, Rizopoulos (2012) (113) found that this pseudo-adaptive quadrature rule performs well, even when the number of observations is quite small.

Laplace Approximations are also used to make the computational approach easier (88). Rizopoulos et al.(2009) (109) suggest a new computational approach which requires fewer repetitions than the often used Monte Carlo and Gaussian Quadrature. For these the amount of time needed for the integral evaluation increases as the dimensions of the integral increase. Thus, they use the Laplace approximation for integrals and develop an EM algorithm for the estimates. This is made under the assumption that censoring and visiting processes (for the longitudinal measurements) are non-informative i.e. that they are independent of the random effects, survival time and longitudinal measurement, similar to Tsiatis and Davidian (2004) (134).

Bayesian methods are also used to estimate parameters. Faucett and Thomas (1996) (42) used a Markov Chain Monte Carlo (MCMC), while other authors use MCMC algorithm with a Gibbs sampler, Faucett et al.(2002) (43), Xu and Zeger (2001) (152), Brown et al.(2005) (17), Guo and Carlin (2004) (51).

Hsieh, Tseng and Wang (2006) (64) suggest bootstrapping for the estimation of standard errors, while they notice that the EM estimates given in Wulfsohn and Tsiatis (1997) (151) are efficient and robust as long as the longitudinal data contains a lot of information (i.e. not have large measurement errors or are not too dispersed).

Sweeting and Thompson (2011) (129) have given a comparison of maximum likelihood and Bayesian estimation of the joint modelling and concluded that despite the computational advantages of Bayesian approaches, the need to choose prior distributions requires a sensitivity analysis.

Further estimation techniques include the conditional score approach proposed by Tsiatis and Davidian (2001) (133) and the Dimension Reduction Method by Bianconcini (2015) (12).

This thesis will work with Gauss-Hermite approximation and maximum likelihood estimation will be used because Bayesian approaches depend on a prior and require a sensitivity analysis, Sweeting and Thompson (2011) (129). Besides maximum likelihood, bootstrapping will also be used for the estimation of the parameters and their standard errors.

1.4.7 Accelerated Failure Time Models

Despite Cox proportional hazards model's popularity in conjuction with joint modelling, its proportionality assumption often fails to be satisfied. Thus, there is a need for other models that do not use this assumption. Parametric models make use of the assumption that the survival times come from a specific distribution (20) (71) (63).

Accelerated failure time models are the models that are linearly-related with the log survival time (63). In this section, the joint models that use accelerated failure time models for the survival sub-model are considered.

Pantazis et al. (2005) (95) suggested a bivariate linear mixed model for longitudinal measurements and a log-normal model for time-to-event. Additionally, they assume that the individual level random coefficients and survival residuals follow jointly a multivarite normal distribution with zero mean. Pantazis and Touloumi (2007) (94) found that the model suggested by Pantazis et al. (2005) (95) is quite robust, but the standard errors may be underestimated, especially under heavily skewed distributions.

Vonesh et al. (2006) (142) proposed a shared random parameter model for the joint distribution of longitudinal and time-to-event data. The submodels are generalized nonlinear mixed model for longitudinal and accelerated failure time model for the survival data.

Tseng, Hsieh and Wang (2005) (132) proposed a joint modelling approach that includes an accelerated failure time model for the survival submodel and a linear mixed model for the longitudinal submodel. This can be used when the proportionality assumption of the Cox model fails.

A general parametric survival family including the accelerated failure time models is considered in this thesis.

1.4.8 Left-Truncated Survival Data

The majority of researchers working either with survival data on its own, or in joint modelling, consider only right-censoring. Left-truncation is often neglected. However, if left-truncation is not taken into account in joint modelling then subjects with shorter survival times are excluded from the sample, and thus longitudinal measurements are sampled with bias. An exclusion is the paper by Su and Wang (2012) (128), who suggested an approach for joint modelling that overcomes the bias caused by left-truncation. They developed an approach to left-truncated and right-censored survival data with longitudinal covariates. For the longitudinal component, they use a linear mixed effects model, while for the survival part they use a proportional hazards model.

The subject i is enrolled into a study only if the survival time is greater or equal to the truncation time. They use the standard assumption that the survival time of the i^{th} subject, the truncation time of the i^{th} subject and the difference between the censoring time and truncation time for the i^{th} individual are conditionally independent given the covariates, which is equivalent to assuming conditional independence of survival time, truncation time and the difference between truncation and censoring time, given the random effects. Additionally, they assume that the survival time and the difference between truncation time and censoring time are independent of the random effects.

They found that the full likelihood can be simplified to a conditional likelihood. Left-truncation, though, makes score equations more complicated, and thus they use a modified likelihood in order to simplify the estimation.

This thesis will also use left-truncation since not taking it into account when fitting a model, causes bias.

1.5 Summary

Joint Modelling is the simultaneous modelling of longitudinal and survival data. There are two different factorizations for this type of modelling: pattern-mixture and selection models (more details in Chapter 2). For both factorizations random effects can be incorporated resulting in random pattern-mixture and random selection models. Additionally there is another category of models, random effects models (or shared parameter models) in which the longitudinal and survival submodels are independent conditional on the random effects. This thesis considers a shared parameter model.

The two processes, longitudinal and time-to-event, can be analyzed separately, but the association between them can be used to increase the efficiency of the joint analysis, aimed at the evaluation of the effects of longitudinal covariates on the occurrence of the events.

The longitudinal covariates are typically observed in a set of patient-specific time-points and can include missing values. Historically, the first, naive approaches were based on the last value carried forward method which is known to cause substantial biases, and the two-stage estimation procedure. A simultaneous parameter estimation for both processes, longitudinal and survival is the widely used modern alternative. The modelling strategies used in the joint modelling differ depending on the primary interest of the research. The thesis primary interest is the association of the longitudinal component to the survival time.

Up to date, all published joint modelling models consider joint models for right-censoring only, neglecting left-truncation. However, data often includes left-truncation which can cause bias to estimates of the parameters, since the survival of the individuals in a study is conditional on the survival up to the point of entry. This thesis considers not only shared parameter model for right-censoring case, but also for the case of left-truncation and right-censoring.

The majority of joint modelling literature considers Cox proportional hazards model for the time-to-event data. The proportional hazards model leaves the baseline hazard undefined, but it assumes that the hazard functions are multiplicatively related which means that their ratio is constant over time. Despite Cox proportional hazards model's popularity, its proportionality assumption often fails to be satisfied. Thus, there is a need for alternative models that do not use this assumption. Parametric methods assume that the survival times come from a specific distribution. The main distributions used are: Weibull, Log-Logistic, Log-Normal, Extreme Value, Logistic, Gompertz, Perks and Beard. These models are widely popular, especially in actuarial applications. The reasons for their popularity include the direct inference of the effect that the variables have on survival time, the use of maximum likelihood estimation and the ability to incorporate different shapes of the hazard.

Statistical inference in parametric modelling is mostly likelihood-based. The joint likelihood is easy to write out but the integrals that are part of it cannot be solved analytically. Approximations used include Gauss-Hermite approximation, Laplace approximation and the Dimension Reduction Method. This thesis uses a Gauss-Hermite approximation for the integral.

The estimation techniques include Bayesian methods based on Markov Chain Monte Carlo, or frequentist methods such as conditional score approach, maximum likelihood, and max-

imum likelihood with bootstrap estimation for standard errors. In this thesis maximum likelihood estimation, bootstrap, and maximum likelihood estimation with bootstrap for standard errors are used.

Reviews of joint modelling approaches for this field can be found in Tsiatis and Davidian (2004) (134), Sousa (2011)(126) and McCrink et al (2013) (88). Software available in R include the JM package (110) and the the joineR package, (99). The joint modelling software is also available in stata (27), matlab (105), winBUGS and SAS, using the NLMixed Procedure by (51).

Chapter 3 contains the main theoretical considerations of the thesis. It presents the joint modelling framework, the sub-models and the log-likelihood for right-censoring and left-truncation, describes the Gauss-Hermite and Laplace approximation that are applied to the likelihood integral, and the maximum likelihood and the bootstrap estimation of the parameters. Chapter 4 includes a simulation study for the joint modelling presented in Chapter 3. Chapter 5 describes the applications of the developed methods to two medical problems, prostate cancer and primary biliary cirrhosis (PBC) dataset from the JM package (110). Chapter 6 presents the future extensions of the joint modelling approach presented on Chapter 3.

2 Joint Modelling of Longitudinal and Survival Data

2.1 Introduction

The previous chapter described joint modelling methods for longitudinal and survival data, and also longitudinal and survival analysis separately. There was no research so far on left-truncation and right-censoring joint modelling using parametric survival models and this is what this chapter will describe.

First, the necessary notation is introduced, and then the linear mixed model that is used for the longitudinal component and a general class of models for the survival component are described. Then, the shared parameter likelihood for two cases, right-censoring and right-censoring and left-truncation, is obtained. Due to the difficulty in evaluating the likelihood integral analytically, integral approximations are applied and a Weibull application is described in detail. Finally, the methods of maximum likelihood and boostrap to obtain the estimates of the parameters are explained.

2.2 Notation

Let A_i, T_i^* and C_i denote the left-truncation time, true survival time and censoring time respectively, for the i^{th} individual, $i = 1, ..., n$. The observed survival time for the i^{th} individual is given by $T_i = \min(T_i^*, C_i)$. $\delta_i = I(T_i^* \leq C_i)$ is the censoring indicator (1-dead, 0-alive). Let $t_i = \{t_{ij}, j = 1, 2, ..., m_i\}$, be the set of m_i time-points, at which the i^{th} individual is observed, these could differ for each individual, and $Y_i(t_i) = \{Y_i(t_{ij}), j = 1, ..., m_i\}$ be a vector of the observed values of a longitudinal biomarker.

$X_i(t)$, $Z_i(t)$ and $Y_i(t)$ are matrix functions of time t. These are not observed at all time-points, but only at t_{ij}, $j = \{1, ..., m_i\}$. $X_i(t_i)$, $Z_i(t_i)$ are the matrices of the p fixed and q random time-dependent covariates, respectively, with $x_i^T(t)$, $z_i^T(t)$ being their row vectors at time t. Let U_i be the set of time-independent covariates for the i^{th} patient and u_i^T be a row vector at all times t. Let M_i^T be the vector of the s time-independent survival covariates, related only to survival.

2.3 Longitudinal Sub-Model

The subject-specific observed values of the longitudinal biomarker are

$$Y_i(t_i) = X_i^T(t_i)\beta + Z_i^T(t_i)b_i + U_i^T\lambda + e_i(t_i), \tag{76}$$

((109),(139)) where β is a vector denoting time-dependent fixed effects, λ a vector denoting a time-independent fixed effects and b_i a vector denoting the random effects for the i^{th} individual. $e_i(t_i)$ is a $m_i \times 1$ vector of the measurement errors, independent of all other variables. The errors are normally distributed $N(0, \sigma_e^2)$ ((109),(139)) and correlation between repeated measurements in longitudinal process is $(e_i(t_{ij}), e_i(t_{ij'})) = 0$ for $j \neq j'$, ((109),(139)).

Assume that there are p time-dependent fixed parameters. Then β is $p \times 1$ vector and $X_i^T(t_i)$ is a $m_i\times$ p design matrix for time-dependent fixed effects. Assuming there are q random effects, b_i is a $q \times 1$ vector, and $Z_i^T(t_i)$ is a $m_i \times q$ design matrix for random effects. The elements of $b_i \sim MVN(0, D)$ and D is a $q \times q$ variance-covariance matrix. If there are r time-independent fixed parameters, λ is a $r \times 1$ vector of time-independent fixed effects that links a $m_i \times r$ matrix, U_i^T, to the longitudinal observations $Y_i(t_i)$.

The conditional expectation of $Y_i(t_i)$ given b_i, is denoted by $W_i(t_i) = \{W_i(t_{ij}), j = 1, ..., m_i\}$.

It is given by

$$E[Y_i(t_i)|b_i] = W_i(t_i) = X_i^T(t_i)\beta + Z_i^T(t_i)b_i + U_i^T\lambda, \tag{77}$$

((109),(139)), and thus the longitudinal biomarker is given by

$$Y_i(t_i) = W_i(t_i) + e_i(t_i). \tag{78}$$

2.4 Survival Sub-Model

Let h(t) be a function of time. Two important cases are

$$h(t) = \begin{cases} \log(t), \text{if accelerated failure time model} \\ t, \text{for the rest of the parametric models.} \end{cases} \tag{79}$$

For the i^{th} individual $h(t)$ can be written as

$$h(t_i) = \mu_i(t_i) + \sigma \times \epsilon_i, \tag{80}$$

where σ is the scale parameter, ϵ_i are the errors from the appropriate survival distribution and $\mu_i(t)$ is given by

$$\mu_i(t) = \alpha W_i(t) + M_i^T\gamma, \tag{81}$$

((63),(90),(97)), where α is the univariate association parameter between the longitudinal and the survival sub-models, ((109),(113)). When $\alpha = 0$, there is no relation between the two processes ((113),(109)). In equation (145), γ is the regression coefficient which links the time-independent covariates, M_i^T. For s time-independent survival covariates, γ is $s \times 1$ vector of regression coefficients for $1 \times s$ vector of time-independent survival covariates, M_i^T.

Therefore, the joint model survival $h_i(t)$ for the i^{th} individual at an arbitrary time-point t is given by

$$h_i(t) = \alpha[x_i^T(t)\beta + z_i^T(t)b_i + u_i^T\lambda] + M_i^T\gamma + \sigma \times \epsilon_i. \tag{82}$$

2.5 Likelihood for Joint Modelling

Censoring and drop-out mechanisms are assumed to be non-informative, i.e. independent of the random effects, survival time and longitudinal measurements.

For a shared parameter model, the conditional independence assumption is needed ((113),(109)). The two processes, longitudinal and survival, are assumed to be independent given the random effects b_i, i.e. the only association between them is induced by the random effects ((113),(109)).

2.5.1 Joint Modelling Parameterization

Let Y and T denote the longitudinal and survival component of the joint modelling likelihood. To ease notation, let us not include any other quantities such as parameters.

The likelihood of observed response data is given by $L(Y,T) = \int L((Y,T)|b)L(b)db$. Y and T are both conditional on b.

Let us assume that b underlies both processes Y and T ie those two components are independent conditional on b and so the full likelihood of a shared-random effects model is given by $L(Y,T) = \int L(Y|b)L(T|b)L(b)db$.

2.5.2 Likelihood under Right-Censoring

The likelihood of the shared parameter model is defined by

$$L(Y_i, T_i, \delta_i; \theta) = \int L(Y_i|b_i; (\theta_y, \sigma_e^2)) L(T_i, \delta_i|b_i; (\theta_t, \sigma^2)) \phi_q(b_i; D) db_i, \tag{83}$$

where $\phi_q(\cdot; D)$ is the q-variate normal density with zero mean and the variance-covariance matrix D, and $L(\cdot)$ denote the appropriate likelihood functions. $\theta_y = (\beta, \lambda)$ and $\theta_t = (\theta_y, \gamma, \alpha)$.

Since the longitudinal measurements, $Y_i(t_{ij})$, $j = \{1, ..., m_i\}$, conditional on b_i are independent,

$$L(Y_i|b_i; (\theta_y, \sigma_e^2)) = \prod_{j=1}^{m_i} \phi\left(Y_i(t_{ij}); W_i(t_{ij}), \sigma_e^2\right), \tag{84}$$

where $\phi(\cdot, \mu, \sigma^2)$ is the univariate Gaussian density with the mean μ and the variance σ^2. Denote $f(\cdot)$ and $S(\cdot)$ the probability density function and the survival function, respectively, for the appropriate survival distribution. Then

$$L(T_i, \delta_i|b_i; (\theta_t, \sigma^2)) = \left[f(T_i|b_i; \theta_t, \sigma^2)\right]^{\delta_i} \left[S(T_i|b_i; \theta_t, \sigma^2)\right]^{1-\delta_i}.$$

The joint likelihood contribution for right-censoring is given by

$$\begin{aligned} L(Y_i, T_i, \delta_i; \theta) = \; & \prod_{i=1}^{n} \int_{-\infty}^{\infty} \prod_{j=1}^{m_i} \phi(Y_i(t_{ij}); W_i(t_{ij}), \sigma_e^2) \times \\ & \left[f(T_i|b_i; \theta_t, \sigma^2)\right]^{\delta_i} \left[S(T_i|b_i; \theta_t, \sigma^2)\right]^{1-\delta_i} \phi_q(b_i; D)\, db_i \end{aligned} \tag{85}$$

2.5.3 Likelihood under Left-Truncation and Right-Censoring

The same considerations apply for likelihood under left-truncation and right-censoring. The only difference is that the survival part includes truncation times.

Subjects are part of the study only if the survival time is greater or equal to the truncation, i.e. the truncation time is the starting point and survival time will now be $T_i' = T_i + A_i$ for the i^{th} subject, where A_i and T_i are the truncation and the un-adjusted survival time. Survival time, truncation time and difference between censoring and truncation time are conditionally independent given the covariates.

The likelihood of the shared parameter model is defined by

$$L(Y_i, T_i, A_i, \delta_i; \theta) = \int L(Y_i|b_i; (\theta_y, \sigma_e^2)) L(T_i, A_i, \delta_i|b_i; (\theta_t, \sigma^2)) \phi_q(b_i; D) db_i, \tag{86}$$

where $\phi_q(\cdot; D)$ is the q-variate normal density with zero mean and the variance-covariance matrix D, and $L(\cdot)$ denote the appropriate likelihood functions.

Denote $f(\cdot)$ and $S(\cdot)$ the probability density function and the survival function, respectively, for the appropriate survival distribution and $T_i' = T_i + A_i$. Then

$$L(T_i, A_i, \delta_i|b_i; (\theta_t, \sigma^2)) = \left[\frac{f(T_i'|b_i; \theta_t, \sigma^2)}{f(A_i|b_i; \theta_t, \sigma^2)}\right]^{\delta_i} \left[\frac{S(T_i'|b_i; \theta_t, \sigma^2)}{S(A_i|b_i; \theta_t, \sigma^2)}\right]^{1-\delta_i}.$$

The joint likelihood contribution for left-truncation and right-censoring is given by

$$\begin{aligned} L(Y_i, T_i, A_i, \delta_i; \theta) = \; & \prod_{i=1}^{n} \int_{-\infty}^{\infty} \prod_{j=1}^{m_i} \phi(Y_i(t_{ij}); W_i(t_{ij}), \sigma_e^2) \times \\ & \left[\frac{f(T_i'|b_i; \theta_t, \sigma^2)}{f(A_i|b_i; \theta_t, \sigma^2)}\right]^{\delta_i} \left[\frac{S(T_i'|b_i; \theta_t, \sigma^2)}{S(A_i|b_i; \theta_t, \sigma^2)}\right]^{1-\delta_i} \phi_q(b_i; D)\, db_i \end{aligned} \tag{87}$$

2.6 Approximations to Likelihood

The integrals for the likelihood of a shared parameter model (85) or (87) are complicated and it is difficult to evaluate them analytically. Thus, an integral approximation is needed.

Two possible approximations, Gauss-Hermite and Laplace approximation, are considered in this section. The Gauss-Hermite approximation was used in our R program described in the next Chapter.

2.6.1 Gauss-Hermite Approximation

For univariate integrals, the following relationship holds

$$\int_{-\infty}^{\infty} f(x)exp(-x^2)dx \approx \sum_{i=1}^{m} w_i f(x_i), \tag{88}$$

where the x_i are zeros of the m^{th} order Hermite polynomial and w_i are suitably corresponding weights (82).

In the case where the integrals are multi-dimensional it is easier to expand around a single mode. The variables should be transformed and centered at the mode and scaled by the inverse of the second derivatives matrix, in order to center the quadrature rule near the mode and to make variables' scaling similar (50) (113) (137).

Suppose that the integral has the form $\int_{R^L} q(\Theta) \exp\{h(\Theta)\}\ d\Theta$, where Θ is a L-dimensional. Assuming that $\hat{\Theta}$ is the mode of h, denote $H = \frac{-\partial^2 h(\hat{\Theta})}{\partial\Theta\partial\Theta'}$ and let $B'B = H$ be the Cholesky factorization of H (50) (113) (137).

The Gauss-Hermite weight function is proportional to a $N(0, 2^{-1}I)$ density, and thus a function that is proportional to that is preferred, in order to use a Gauss-Hermite rule. Considering Θ as random quantity with log-density given by $h(\Theta)$, the first order normal approximation to this distribution is the $N(\hat{\Theta}, H^{-1})$ distribution. Parameterizing Θ as $A = 2^{-1/2}B(\Theta - \hat{\Theta})$, the A variable has an approximate $N(0, 2^{-1}I)$ distribution (50) (113)

(137).

Therefore, the integral becomes $2^{L/2}|B|^{-1}\int \exp(-A^{'}A)f(A)dA$, where $f(A) = q(2^{1/2}B^{-1}A+\hat{\Theta})\exp[h(2^{1/2}B^{-1}A\hat{\Theta})+A^{'}A]$. The integral now can be approximated using a Q-point Gauss-Hermite rule by $2^{L/2}|B|^{-1}\sum_{q_1=1}^{Q}\omega_{q_L}f(\Xi_{q_l})$.

2.6.2 Gauss-Hermite Approximation to Likelihood Integral

The integrals in the likelihoods (85) and (87) have the form $\int g(b_i)\phi(b_i;D)db_i$. They are difficult to evaluate analytically, and a Gauss-Hermite approximation with p nodes can be used. The Gauss-Hermite weight function is proportional to $N(0,2^{-1}I)$ density, where I is the identity matrix. Therefore a scaling transformation $a(b_i) = 2^{-\frac{1}{2}}Bb_i$, where B is the Cholesky factorisation of D, is used.

The Gauss-Hermite approximation to the likelihood of the random shared model with right-censoring given by integral (85) is given by the weighted sum

$$\begin{aligned}\widetilde{L}(Y_i,T_i,\delta_i;\theta) = \; & 2^{\frac{qn}{2}}|B|^{-n}\prod_{i=1}^{n}\sum_{l_i=1}^{p}\omega_{l_i}\exp\{E_i'E_i\}\prod_{j=1}^{m_i}\phi(Y_i(t_{ij});\widehat{W}_i(t_{ij}),\sigma_e^2) \\ & \left[f(T_i|2^{1/2}B^{-1}E_i;\theta_t,\sigma^2)\right]^{\delta_i}\left[S(T_i|2^{1/2}B^{-1}E_i;\theta_t,\sigma^2)\right]^{1-\delta_i}\phi_q(2^{1/2}B^{-1}E_i;D),\end{aligned} \tag{89}$$

where E_i are the abscissas, ω are the weights, p is the number of nodes, and $\widehat{W}_i(t_{ij}) = X_i^T(t_{ij})\beta + Z_i^T(t_{ij})(2^{\frac{1}{2}}B^{-1}E_{il_i}) + U_i^T\lambda$.

The Gauss-Hermite approximation to the likelihood of the random shared model with

left-truncation and right-censoring given by integral (87) is given by the weighted sum

$$\begin{aligned}\widetilde{L}(Y_i, T_i, A_i, \delta_i; \theta) = \; & 2^{\frac{qn}{2}} |B|^{-n} \prod_{i=1}^{n} \sum_{l_i=1}^{p} \omega_{l_i} \exp\{E_i' E_i\} \prod_{j=1}^{m_i} \phi(Y_i(t_{ij}); \widehat{W}_i(t_{ij}), \sigma_e^2) \\ & \left[\frac{f(T_i'|2^{1/2}B^{-1}E_i;\theta_t,\sigma^2)}{f(A_i|2^{1/2}B^{-1}E_i;\theta_t,\sigma^2)}\right]^{\delta_i} \left[\frac{S(T_i'|2^{1/2}B^{-1}E_i;\theta_t,\sigma^2)}{S(A_i|2^{1/2}B^{-1}E_i;\theta_t,\sigma^2)}\right]^{1-\delta_i} \phi_q(2^{1/2}B^{-1}E_i; D),\end{aligned} \tag{90}$$

where E_i are the abscissas, ω are the weights, p is the number of nodes, and $\widehat{W}_i(t_{ij}) = X_i^T(t_{ij})\beta + Z_i^T(t_{ij})(2^{\frac{1}{2}}B^{-1}E_{il_i}) + U_i^T\lambda$.

2.6.3 Laplace Approximation

Suppose there is an integral $\int [\exp l(t)]\, dt$. The main goal is to approximate $l(t)$ by a Taylor series expansion about the maximum $l(t)$

$$l(t) \approx l(\hat{t}) + \frac{\partial l(t)^T}{\partial t}|_{t=\hat{t}}(t-\hat{t}) + \frac{1}{2}(t-\hat{t})^T l''(\hat{t})(t-\hat{t}), \tag{91}$$

where $l''(\hat{t})$ is the Hessian matrix evaluated at the maximum $l''(\hat{t}) = \frac{\partial^2 l(t)}{\partial t \partial t^T}|_{t=\hat{t}}$ (50) (137).

At the maximum of $l(t)$, the first derivative is zero. So, substituting $l(t)$ by its Taylor series expansion
$\int \exp[l(t)]\, dt \approx \int \exp[l(\hat{t}) + \frac{1}{2}(t-\hat{t})^T l''(\hat{t})(t-\hat{t})]\, dt = \exp[l(\hat{t})] \int \exp[\frac{1}{2}(t-\hat{t})^T l''(\hat{t})(t-\hat{t})]\, dt$.
Thus,

$$\int [\exp l(t)]\, dt \approx (2\pi)^{\frac{Q}{2}} \det\left\{-l''(\hat{t})\right\}^{-\frac{1}{2}} \exp[l(\hat{t})], \tag{92}$$

where Q is the dimension of t (50) (137).

2.6.4 Laplace Approximation to Likelihood Integral

Suppose there is an integral $\int [\exp h(b_i)]\, db_i$. The main goal is to approximate $h(b_i)$ by a Taylor series expansion about the maximum $h(b_i)$:

$$h(b_i) \approx h(\hat{b}_i) + \frac{\partial h(b_i)^T}{\partial b_i}|_{b=\hat{b}}(b_i - \hat{b}) + \frac{1}{2}(b_i - \hat{b})^T h''(\hat{b})(b_i - \hat{b}), \tag{93}$$

((50),(137)), where $h''(\hat{b})$ is the Hessian matrix evaluated at the maximum: $h''(\hat{b}) = \frac{\partial^2 h(b_i)}{\partial b_i \partial b_i^T}|_{b=\hat{b}}$. At the maximum of $h(b_i)$, the first derivative is zero. So, substituting $h(b_i)$ by its Taylor series expansion

$$\begin{aligned} \int \exp[h(b_i)]\, db_i \quad &\approx \int \exp[h(\hat{b}) + \tfrac{1}{2}(b_i - \hat{b})^T h''(\hat{b})(b_i - \hat{b})]\, db_i \\ &= \exp[h(\hat{b})] \int \exp[\tfrac{1}{2}(b_i - \hat{b})^T h''(\hat{b})(b_i - \hat{b})]\, db_i \\ &= (2\pi)^{\frac{q}{2}} \det\left\{-h''(\hat{b_i})\right\}^{-\frac{1}{2}} \exp[h(\hat{b})], \end{aligned} \tag{94}$$

((50),(137)), where q is the dimension of b.

Laplace Approximation to likelihoods (83) and (86) is applied by using $h(b_i) = \log b_i$ to change the form of the integrals.

The adjusted likelihood for (83) is

$$L(Y_i, T_i, \delta_i; \theta) = \int \exp\left[\log\left[L(Y_i|b_i; (\theta_y, \sigma_e^2))L(T_i, \delta_i|b_i; (\theta_t, \sigma^2))\phi_q(b_i; D)\right]\right] db_i, \tag{95}$$

Applying Laplace approximation to (95), the likelihood becomes

$$\widetilde{L}(Y_i, T_i, \delta_i; \theta) = (2\pi)^{\frac{nq}{2}} \hat{\Omega}^{-\frac{n}{2}} \prod_{i=1}^{n} \phi(Y_i(t_{ij}); \widehat{W}_i(t_{ij}), \sigma_e^2) L(T_i, \delta_i|\hat{b}; (\theta_t, \sigma^2))\phi_q(\hat{b}; D), \tag{96}$$

where $\hat{\Omega} = (-\hat{H})^{-1}$ is the asymptotic variance-covariance of $\hat{H}$ evaluated at mode $\hat{b}$.

Similarly, the adjusted likelihood for (86) is

$$L(Y_i, T_i, A_i, \delta_i; \theta) = \int \exp\left[\log\left[L(Y_i|b_i; (\theta_y, \sigma_e^2))L(T_i, A_i, \delta_i|b_i; (\theta_t, \sigma^2))\phi_q(b_i; D)\right]\right]db_i, \quad (97)$$

Applying Laplace approximation to (97), the approximate likelihood is

$$\widetilde{L}(Y_i, T_i, A_i, \delta_i; \theta) = (2\pi)^{\frac{nq}{2}}\hat{\Omega}^{-\frac{n}{2}}\prod_{i=1}^{n}\phi(Y_i(t_{ij}); \widehat{W}_i(t_{ij}), \sigma_e^2)L(T_i, A_i, \delta_i|\hat{b}; (\theta_t, \sigma^2))\phi_q(\hat{b}; D), \quad (98)$$

2.7 Applications to particular survival distributions

The main survival distributions that are used in practice are shown in Table 1 below. Three of the survival distributions, Gompertz, Beard and Perks can be obtained from the distributions in Table 1 by using transformations (see Appendices F, H and I for more details).

Table 1: Main Survival Distributions

Distributions	Probability Density Function	Cummulative Distribution Function	Survival Function	Hazard Function	Cummulative Hazard Function
Weibull	$\frac{\gamma}{n}(\frac{t}{n})^{\gamma-1}e^{-(\frac{t}{n})^{\gamma}}$	$1-e^{-(\frac{t}{n})^{\gamma}}$	$e^{-(\frac{t}{n})^{\gamma}}$	$\frac{\gamma}{n}(\frac{t}{n})^{\gamma-1}$	$(\frac{t}{n})^{\gamma}$
Log-Logistic	$\frac{\lambda p t^{p-1}}{(1+\lambda t^{p})^{2}}$	$\frac{\lambda t^{p}}{1+\lambda t^{p}}$	$\frac{1}{1+\lambda t^{p}}$	$\frac{\lambda p t^{p-1}}{1+\lambda t^{p}}$	$\log[1+\lambda t^{p}]$
Log-Normal	$\frac{e^{-\frac{(\log t-\mu)^{2}}{2\sigma^{2}}}}{t\sigma(2\pi)^{\frac{1}{2}}}$	$\Phi[\frac{\log t-\mu}{\sigma}]$	$1-\Phi[\frac{\log t-\mu}{\sigma}]$	$\frac{e^{-\frac{(\log t-\mu)^{2}}{2\sigma^{2}}}}{t\sigma(2\pi)^{\frac{1}{2}}}$ $\frac{}{1-\Phi[\frac{\log t-\mu}{\sigma}]}$	$-\log(1-\Phi[\frac{\log t-\mu}{\sigma}])$
Extreme Value	$\frac{1}{\sigma}e^{\frac{t-\mu}{\sigma}-e^{\frac{t-\mu}{\sigma}}}$	$1-e^{-e^{\frac{t-\mu}{\sigma}}}$	$e^{-e^{\frac{t-\mu}{\sigma}}}$	$\frac{1}{\sigma}e^{\frac{t-\mu}{\sigma}}$	$e^{\frac{t-\mu}{\sigma}}$
Logistic	$\frac{1}{\sigma}\frac{e^{\frac{t-\mu}{\sigma}}}{[1+e^{\frac{t-\mu}{\sigma}}]^{2}}$	$\frac{e^{\frac{t-\mu}{\sigma}}}{1+e^{\frac{t-\mu}{\sigma}}}$	$\frac{1}{1+e^{\frac{t-\mu}{\sigma}}}$	$\frac{1}{\sigma}\frac{e^{\frac{t-\mu}{\sigma}}}{[1+e^{\frac{t-\mu}{\sigma}}]}$	$\log[1+e^{\frac{t-\mu}{\sigma}}]$

2.7.1 Shared parameter model with Weibull survival distribution

As an example of a parametric shared parameter model, the Weibull survival distribution is considered. The longitudinal component is modelled with a linear mixed effects model. The survival component is modelled using Weibull distribution. Theory required in joint modelling distributions is provided in the Appendices (B)-(I).

2.7.1.1 From Weibull distribution to model As can be seen from Table 1 the survival function of Weibull distribution is given by

$$S(t;\sigma,\gamma) = \exp\left\{-(tn^{-1})^{\gamma}\right\}, \tag{99}$$

Write $\lambda = n^{-\gamma}$, thus

$$S(t;\lambda,\gamma) = \exp\left\{-\lambda t^{\gamma}\right\}. \tag{100}$$

Take the log transformation of time ($s = \log t$)

$$S(s;\lambda,\gamma) = \exp\left\{-\lambda \exp\left\{\gamma s\right\}\right\}. \tag{101}$$

Redefine parameters, $\gamma = \sigma^{-1}$ and $\lambda = \exp\left\{-\mu\sigma^{-1}\right\}$, then the random variable $s = \log T$

$$s = \log T = \mu + \sigma W, \tag{102}$$

where W is the Extreme Value distribution with probability density function

$$f_W(w) = \sigma^{-1} \exp\left\{w - \exp\left\{w\right\}\right\} \tag{103}$$

and survival function

$$S_W(w) = \exp\left\{-\exp\left\{w\right\}\right\}. \tag{104}$$

Substitute $W = \sigma^{-1}(s-\mu)$ to obtain the probability density and the survival function of log survival time. Therefore, the probability density function for s, for the i^{th} individual is:

$$f(s_i;\mu_i,\sigma) = \sigma^{-1} \exp\left\{\sigma^{-1}(s_i-\mu_i) - \exp\left\{\sigma^{-1}(s_i-\mu_i)\right\}\right\} \tag{105}$$

and the survival function for the i^{th} individual is:

$$S(s_i;\mu_i,\sigma) = \exp\left\{-\exp\left\{\sigma^{-1}(s_i-\mu_i)\right\}\right\}. \tag{106}$$

2.7.1.2 Joint Modelling with Weibull survival sub-model The log survival time for the i^{th} individual, following Weibull distribution is given by

$$\log T_i(t_i) = \mu_i(t_i) + \sigma \times \epsilon_i, \tag{107}$$

where ϵ_i comes from Extreme Value distribution, σ is the scale and $\mu_i(t_i)$ is given by

$$\mu_i(t_i) = \alpha W_i(t) + M_i^T \gamma, \tag{108}$$

where γ is the regression coefficient that links the time-independent covariates, M_i^T, related only to survival and $W_i(t)$ is the longitudinal component.

The probability density and survival function of Weibull model are given by

$$f(T_i|b_i;\theta,t) = \quad (t\exp[\mu_i(t_i)])^{\exp(1/\sigma-(t\exp[\mu_i(t_i)])^{\exp(1/\sigma)})} \tag{109}$$

and

$$S(T_i|b_i;\theta,t) = \exp[-(t\exp[\mu_i(t_i)])^{\exp(1/\sigma)}], \tag{110}$$

respectively ((15),(63),(90)). For the longitudinal submodel, the probability density functions of biomarker $Y_i(t_{ij})$ conditional on random effects, and that of random effects b_i are given by

$$p(Y_i|b_i;\theta) = (2\pi\sigma_e^2)^{-\frac{1}{2}}\exp[-(Y_i(t_{ij}) - W_i(t_{ij}))^2(2\sigma_e^2)^{-1}], \tag{111}$$

and

$$p(b_i;\theta) = (2\pi|D|)^{-\frac{1}{2}}\exp(-2^{-1}b_i^{'}D^{-1}b_i), \tag{112}$$

respectively, (139).

2.7.1.3 Joint Likelihood The Weibull joint likelihood contribution for right-censoring only is

$$\begin{aligned} L(Y_i,T_i,\delta_i;\theta) = \; & \textstyle\prod_{i=1}^n [\int_{-\infty}^{\infty}\prod_{j=1}^{m_i}[(2\pi\sigma_e^2)^{-\frac{1}{2}}\exp[-(Y_i(t_{ij}) - W_i(t_{ij}))^2(2\sigma_e^2)^{-1}]] \\ & \times(2\pi|D|)^{-\frac{1}{2}}\exp(-2^{-1}b_i^{'}D^{-1}b_i)((t_i\exp[\theta x])^{\exp(1/\sigma-(t_i\exp[\mu_i(t_i)])^{\exp(1/\sigma)})})^{\delta_i} \\ & \times(\exp[-(t_i\exp[\mu_i(t_i)])^{\exp(1/\sigma)}])^{1-\delta_i}\,db_i\,]. \end{aligned} \tag{113}$$

Applying Gauss-Hermite approximation, the likelihood becomes

$$\begin{aligned}\widetilde{L}(Y_i,T_i,\delta_i;\theta) = \; & 2^{\frac{qn}{2}}|B|^{-n}\prod_{i=1}^{n}\sum_{l_i=1}^{p}\omega_{l_i}\exp\{E_i'E_i\}\prod_{j=1}^{m_i}[(2\pi\sigma_e^2)^{-\frac{1}{2}} \\ & \times\exp[-(Y_i(t_{ij})-\hat{W}_i(t_{ij}))^2(2\sigma_e^2)^{-1}]] \\ & \times(2\pi|D|)^{-\frac{1}{2}}\exp(-2^{-1/2}B^{-1}E_iD^{-1}2^{1/2}B^{-1}E_i) \\ & \times((t_i\exp[\mu_i(t_i)])^{\exp(1/\sigma-(t_i\exp[\mu_i(t_i)])^{\exp(1/\sigma)})})^{\delta_i} \\ & \times(\exp[-(t_i\exp[\mu_i(t_i)])^{\exp(1/\sigma)}])^{1-\delta_i},\end{aligned} \tag{114}$$

where $\widehat{W}_i(t_{ij}) = X_i^T(t_{ij})\beta + Z_i^T(t_{ij})(2^{\frac{1}{2}}B^{-1}E_{il_i}) + U_i^T\lambda$ and $\hat{\mu}_i(t_i) = \alpha\hat{W}_i(t) + M_i^T\gamma$.
Applying Laplace approximation the likelihood becomes

$$\begin{aligned}\widetilde{L}(Y_i,T_i,\delta_i;\theta) = \; & (2\pi)^{\frac{nq}{2}}\hat{\Omega}^{-\frac{n}{2}}\prod_{i=1}^{n}\prod_{j=1}^{m_i}[(2\pi\sigma_e^2)^{-\frac{1}{2}}\exp[-(Y_i(t_{ij})-\hat{W}_i(t_{ij}))^2(2\sigma_e^2)^{-1}]] \\ & \times(2\pi|D|)^{-\frac{1}{2}}\exp(-2^{-1}\hat{b}_i'D^{-1}\hat{b}_i) \\ & \times((t_i\exp[\mu_i(t_i)])^{\exp(1/\sigma-(t_i\exp[\mu_i(t_i)])^{\exp(1/\sigma)})})^{\delta_i} \\ & \times(\exp[-(t_i\exp[\mu_i(t_i)])^{\exp(1/\sigma)}])^{1-\delta_i},\end{aligned} \tag{115}$$

where $\widehat{W}_i(t_{ij}) = X_i^T(t_{ij})\beta + Z_i^T(t_{ij})\hat{b} + U_i^T\lambda$ and $\hat{\mu}_i(t) = \alpha\hat{W}_i(t) + M_i^T\gamma$.

For the data with left-truncation and right-censoring, the likelihood is more complicated. In this case,

$$\begin{aligned}L(Y_i,T_i,A_i,\delta_i;\theta) = \; & \prod_{i=1}^{n}[\int_{-\infty}^{\infty}\prod_{j=1}^{m_i}[(2\pi\sigma_e^2)^{-\frac{1}{2}}\exp[-(Y_i(t_{ij})-W_i(t_{ij}))^2(2\sigma_e^2)^{-1}]] \\ & \times(2\pi|D|)^{-\frac{1}{2}}\exp(-2^{-1}b_i'D^{-1}b_i) \\ & \times(\tfrac{((t_i+A_i)\exp[\mu_i(t_i)])^{\exp(1/\sigma-((t_i+A_i)\exp[\mu_i(t_i)])^{\exp(1/\sigma)})}}{(t_i\exp[\mu_i(t_i)])^{\exp(1/\sigma-(t_i\exp[\mu_i(t_i)])^{\exp(1/\sigma)})}})^{\delta_i} \\ & \times(\tfrac{\exp[-((t_i+A_i)\exp[\mu_i(t_i)])^{\exp(1/\sigma)}]}{\exp[-(t_i\exp[\mu_i(t_i)])^{\exp(1/\sigma)}]})^{1-\delta_i}\,db_i\;].\end{aligned} \tag{116}$$

When Gauss-Hermite approximation is applied, the likelihood becomes

$$\begin{aligned}\widetilde{L}(Y_i,T_i,A_i,\delta_i;\theta) = \; & 2^{\frac{qn}{2}}|B|^{-n}\prod_{i=1}^{n}\sum_{l_i=1}^{p}\omega_{l_i}\exp\{E_i'E_i\}\prod_{j=1}^{m_i}[(2\pi\sigma_e^2)^{-\frac{1}{2}} \\ & \times\exp[-(Y_i(t_{ij})-\hat{W}_i(t_{ij}))^2(2\sigma_e^2)^{-1}]] \\ & \times(2\pi|D|)^{-\frac{1}{2}}\exp(-2^{-1/2}B^{-1}E_iD^{-1}2^{1/2}B^{-1}E_i) \\ & \times(\tfrac{((t_i+A_i)\exp[\mu_i(t_i)])^{\exp(1/\sigma-((t_i+A_i)\exp[\mu_i(t_i)])^{\exp(1/\sigma)})}}{(t_i\exp[\mu_i(t_i)])^{\exp(1/\sigma-(t_i\exp[\mu_i(t_i)])^{\exp(1/\sigma)})}})^{\delta_i} \\ & \times(\tfrac{\exp[-((t_i+A_i)\exp[\mu_i(t_i)])^{\exp(1/\sigma)}]}{\exp[-(t_i\exp[\mu_i(t_i)])^{\exp(1/\sigma)}]})^{1-\delta_i},\end{aligned} \tag{117}$$

where $\widehat{W}_i(t_{ij}) = X_i^T(t_{ij})\beta + Z_i^T(t_{ij})(2^{\frac{1}{2}}B^{-1}E_{il_i}) + U_i^T\lambda$ and $\hat{\mu}_i(t) = \alpha\hat{W}_i(t) + M_i^T\gamma$. When Laplace approximation is applied, the likelihood becomes

$$\begin{aligned}\widetilde{L}(Y_i, T_i, \delta_i; \theta) = \; & (2\pi)^{\frac{nq}{2}}\hat{\Omega}^{-\frac{n}{2}}\prod_{i=1}^{n}\prod_{j=1}^{m_i}[(2\pi\sigma_e^2)^{-\frac{1}{2}}\exp[-(Y_i(t_{ij}) - \hat{W}_i(t_{ij}))^2(2\sigma_e^2)^{-1}]] \\ & \times(2\pi|D|)^{-\frac{1}{2}}\exp(-2^{-1}\hat{b}_i'D^{-1}\hat{b}_i) \\ & \times\Big(\frac{((t_i+A_i)\exp[\mu_i(t_i)])^{\exp(1/\sigma-((t_i+A_i)\exp[\mu_i(t_i)])^{\exp(1/\sigma)})}}{(t_i\exp[\mu_i(t_i)])^{\exp(1/\sigma-(t_i\exp[\mu_i(t_i)])^{\exp(1/\sigma)})}}\Big)^{\delta_i} \\ & \times\Big(\frac{\exp[-((t_i+A_i)\exp[\mu_i(t_i)])^{\exp(1/\sigma)}]}{\exp[-(t_i\exp[\mu_i(t_i)])^{\exp(1/\sigma)}]}\Big)^{1-\delta_i},\end{aligned} \tag{118}$$

where $\widehat{W}_i(t_{ij}) = X_i^T(t_{ij})\beta + Z_i^T(t_{ij})\hat{b} + U_i^T\lambda$ and $\hat{\mu}_i(t) = \alpha\hat{W}_i(t) + M_i^T\gamma$.

2.8 Maximum Likelihood estimation of the parameters

The estimates of the parameters are obtained using maximum likelihood estimation. To achieve this, the log-likelihood is differentiated with respect to each parameter, the derivative is set equal to zero and the resulting equations are solved in respect to each parameter to obtain the maximum likelihood estimate $\hat{\theta}$, where $\theta = \{\theta_1, ..., \theta_K\}$ is the set of all K parameters, (63). The maximum likelihood equations are given by

$$\frac{\partial l(\hat{\theta}; t_i)}{\partial\theta_k} = 0. \tag{119}$$

In the above equation $l(\theta; t)$ denotes the log-likelihood, t is the set of time-points for the i^{th} individual, where $i = 1, 2, ..., n$ and θ_k is the k^{th} parameter to be estimated, $\theta = \{\theta_1, \theta_2, ..., \theta_K\}$.

The Hessian matrix is given by

$$H[l(\theta;y)] = \begin{bmatrix} \frac{\partial^2 l(\theta;t)}{\partial \theta_1^2} & \frac{\partial^2 l(\theta;t)}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 l(\theta;t)}{\partial \theta_1 \partial \theta_K} \\ \frac{\partial^2 l(\theta;t)}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 l(\theta;t)}{\partial \theta_2^2} & \cdots & \frac{\partial^2 l(\theta;t)}{\partial \theta_2 \partial \theta_K} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 l(\theta;t)}{\partial \theta_K \partial \theta_1} & \frac{\partial^2 l(\theta;t)}{\partial \theta_K \partial \theta_2} & \cdots & \frac{\partial^2 l(\theta;t)}{\partial \theta_K^2} \end{bmatrix}, \tag{120}$$

(63). The information matrix is

$$I(\hat{\theta}) = -E[H[l(\theta;y)]], \tag{121}$$

(31) and the variance-covariance matrix is approximated by the inverse of the information matrix

$$\widehat{Var(\hat{\theta})} = I^{-1}(\hat{\theta}), \tag{122}$$

((8),(123)). The square roots of the diagonal elements of this matrix are the estimates of the standard errors ($\hat{SE}(\hat{\theta})$) of each parameter, (63). The confidence intervals for the estimates $\hat{\theta}$ are obtained using $\hat{\theta} \pm z_{1-a/2}\hat{SE}(\hat{\theta})$, where $z_{1-a/2}$ is the upper $a/2$ percentile of the normal distribution, (63).

2.9 Bootstrap estimation of the parameters

Bootstrap is an approach to statistical inference by resampling the data, (46). The idea behind bootstrap is that the sample is to bootstrap samples what the population is to sample, (46). There are different forms of bootstrap, but here the nonparametric bootstrap is used, because it allows to estimate the sampling distribution empirically without making assumptions about the population, (46).

The algorithm for non-parametric resampling is as follows, (19):

1. Sample s observations randomly, with replacement, to create a bootstrap set Y_b^* out of the original sample y_{obs}.

2. Calculate the bootstrap version of the statistics $\psi_b^* = \psi(Y_b^*)$

3. Repeat the first two steps B times to obtain a boostrap estimate.

The estimates can be obtained by using

$$\widehat{\psi^*} = \sum_{b=1}^{B} \frac{\widehat{\psi^*}(b)}{B} \tag{123}$$

((34),(36)) and their standard errors are

$$\hat{se}_B = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}[\hat{\psi}^*(b) - \hat{\psi}^*]^2}, \tag{124}$$

((34),(36)).

These are used to estimate the parameters, the standard errors of the parameters and then the confidence intervals.

For bootstrap confidence intervals, there exist many different methods (46). This thesis is using the percentile method, which uses the empirical quantiles of $\hat{\psi}^*(b)$ to form the confidence interval and does not requite the evaluation of standard errors (46). Let $\hat{\psi}^*_{(1)}(b), \hat{\psi}^*_{(2)}(b), ..., \hat{\psi}^*_{(B)}(b)$ be the ordered bootstrap estimates, the lower limit of the confidence interval is $\hat{\psi}^*_{[(B+1)a/2]}$ and the upper limit is $\hat{\psi}^*_{[(B+1)(1-a/2)]}$. (46).

2.10 Summary

This chapter described a shared parameter model for longitudinal and survival data under left-truncation and right-censoring. A linear mixed effects model was described to model the longitudinal component and a general class of parametric survival models to model the survival component. Two approximations were explored for the likelihood integral of the model. The two methods, maximum likelihood and bootstrap are described for the estimation of the parameters. Even though both Gauss-Hermite and Laplace approximations are described, only Gauss-Hermite approximation is used in my R program which implements joint modelling.

The next chapter describes the simulations for the joint model with a Weibull survival submodel, as presented in this chapter and the R program written to implement this method and used for simulations.

3 Simulation Study

3.1 Introduction

In the previous section, a shared parameter model for longitudinal and time-to-event data, under left-truncation and right-censoring was introduced. As stated in the previous chapter, Gauss-Hermite approximation was applied to obtain the parameters.

In this chapter, the estimation quality for the proposed ML and the bootstrap methods for a Weibull survival sub-model is assessed. Estimates of the mean and the median bias of the estimators, the standard errors, and the actual coverage at 95% nominal level are calculated. The results from the *lme* (100), package for the longitudinal component, and from the *survreg* (130) and $aftreg$ (15) survival packages for the survival component are compared with those from the joint model, and the results from $jointModel$ (110) are also compared to assess the model fit in comparison with other joint models. For *survreg* only the first observation on the time-dependent covariate was used as it cannot handle time-dependent covariates. On the contrary, $aftreg$ can handle time-dependent covariates, but it cannot handle random effects. *lme* can only handle longitudinal data and not survival data. R documentation is provided for the programs that were created to fit the models and to simulate the data in the Appendix J.

3.2 Design

The joint model framework is a rather complex framework and thus, the simulations needed to be quite simple for the computers to handle. This was mainly the criterion for the

simulations described below.

The simulations are based on the Weibull survival sub-model for $n = 100$ and $n = 500$ patients, using 500 replications for each scenario. Ten Gauss-Hermite points were used for ML estimation, and 100 bootstrap samples for bootstrap-based estimation.

The left-truncation time t_{i0} is taken to be zero, and the first observation for each patient was made at 0. The subsequent observation times for each patient were generated from the Poisson process with intensity 3 or 23, right-censored at 1. This resulted in 1 to 4 and 1 to 24 observations per patient, respectively.

For simplicity, time-dependent covariate $X_i^T(t_{ij}) = t_{ij}$ and for each time-point t_{ij}, the biomarker values $Y_i(t_{ij})$ are simulated as

$$Y_i(t_{ij}) = t_{ij}\beta + b_i + \lambda + e_{ij}, \tag{125}$$

where the univariate random effects b_i are generated from $\mathcal{N}(0, D)$ and the errors $e_i(t_{ij})$ are independently generated from $\mathcal{N}(0, \sigma_e^2)$. The variances were chosen as $D = 0.1$ and $\sigma_e^2 = (0.1)^2$.

For survival, the time-independent covariates M_i^T include an intercept (taken to be zero) and a binary (0 or 1) factor g_i such as gender, with the values generated from the Bernoulli distribution with the probability of success 0.5.

Leemis et al.(1989) denote the cummulative link function as $\Psi(t) = \psi(F(T); \eta)$, where $F(T)$ is the covariate history and η is the set of covariate parameters, and generate survival time T using the algorithm $\Psi^{-1}(H_0^{-1}(-log(1-u)))$, for $u \sim Unif(0,1)$. $H_0^{-1}(t)$ is the inverse cummulative baseline hazard function of $H_0(t) = (t\sigma^{-1})^\gamma$. The model here is $log(T) = \alpha(t_{ij}\beta + b_i + \lambda) + \gamma_1 + \gamma_2 M_i^T + \sigma\epsilon$, where ϵ comes from the Extreme Value

distribution and thus the baseline distribution is Weibull.

So, $\Psi(T) = \int_0^T \exp[\alpha(t_{ij}\beta + b_i + \lambda) + \gamma_1 + \gamma_2 M_i^T + \sigma\epsilon]dt$
$= (\alpha\beta)^{-1} \exp[\alpha\beta](\exp[\alpha(T\beta + b_i + \lambda) + \gamma_1 + \gamma_2 M_i^T + \sigma\epsilon] - 1)$.
The inverse function is $\Psi^{-1}(z) = (\alpha\beta)^{-1} \log[1 + \exp[-(\alpha(\lambda + b_i) + \gamma_1 + \gamma_2 M_i^T)]\beta\alpha z]$.

The baseline hazard for Weibull distribution is $H_0(t) = (t\sigma^{-1})^\gamma$ and the inverse function is $H_0^{-1}(z) = \sigma z^\gamma$.

Generating Weibull survival time $T = \Psi^{-1}[H^{-1}(-\log(1-u))] = \Psi^{-1}[\sigma(-\log(1-u))^\gamma]$

Thus, the time T^* is simulated as

$$T^* = (\alpha\beta)^{-1} \log(1 + \exp(-(\alpha(\lambda + b_i) + \gamma_1 + \gamma_2 M_i^T))\alpha\beta\sigma(-\log(1-u))^\gamma, \tag{126}$$

where $u \sim Unif(0, 1)$.

We aimed at no more than 20% of the total data to be censored. First, we have used Bernoulli random number generator with probability of success 0.2 to select censored patients. For censored patients, the censoring times C_i are simulated from the uniform distribution $(0, T_i^*)$. Otherwise $C_i = \infty$.

The observed survival time is $T_i = \min(T_i^*, C_i)$.

Scenario 1 has only the longitudinal component i.e. $\alpha = 1$, $\beta = 1$, $\lambda = 1$, $\gamma_1 = 0$, $\gamma_2 = 0$. Scenario 2 includes both longitudinal and survival parts, with the parameters $\alpha = 1$, $\beta = 1$, $\lambda = 1$, $\gamma_1 = 0$, $\gamma_2 = 1$. These scenarios were chosen because the joint model needed to be assessed when one of its two parts (longitudinal or survival) was zero and when both of

them were not zero.

Simulations were comparatively slow and needed large memory. The time needed for the task to complete was increasing exponentially with the number of patients and the number of observations. Personal computers could not handle the simulations. Therefore, they were run on a high performance compute cluster (HPC) that is available at University of East Anglia. Many dedicated nodes were assigned for this task because of previous attempts failing on shared nodes. Even then, the scenario with 500 patients and up to 24 observations, took more than two months to complete. The scenario for 100 patients with up to 4 observations, took around 2 weeks.

3.3 Results

Results of our simulations are presented in Tables 2 to 9.
For *survreg*, most of the estimates are considerably biased. This is as expected because it cannot handle time-dependent covariates. However, it estimates the survival slope γ_2 in scenarios 1 quite accurately, and provides fair coverage for it. It is not clear why it performs much worse in scenario 1. It does not estimate the survival error variance $\log(\sigma)$ well in any of the scenarios.

For *aftreg*, survival slope γ_2 is estimated accurately in scenario 1, with a reasonable coverage. The rest of the estimates are biased in all scenarios. This is probably due to the fact that aftreg cannot handle random effects.

For *lme*, the point estimation of all longitudinal parameters is very accurate, including the variance components, but the coverage of β seems to drop when the number of patients

increases.

In the joint modelling, the analytical point estimators for all regression parameters are quite accurate in both scenarios with the exception of γ_1, β when the the number of patients is 500 and the number of observations=23 for both scenarios, α when the number of patients is 100 and the variance of the random effects. In bootstrap, all estimates are not good with the variances being the exception.

All analytical coverage is bad, with the exception of the coverages for survival intercept γ_1, survival slope γ_2 and shared parameter α. Coverage, however, from analytical estimates and bootstrap errors always includes the actual value. Using analytical point estimators with bootstrap standard errors results in reliable though mostly too conservative coverage for all parameters.

Coming to the estimation of the association parameter α, there is a clear difference between analytical and bootstrap estimation. Analytical estimation results in a really small bias for both scenarios when the number of patients is relatively small, while bootstrap estimation is always biased for both scenarios. To summarise, it appears that the safest option is to use analytic estimation with bootstrap standard errors.

JM provides unbiased estimates for both scenarios with really good coverage. The only exception is the survival intercept γ_1 which is biased in all scenarios, but still has a good coverage when the number of patients is 100 and a fair coverage when the number of patients is 500.

It is fair to say that JM program is indeed the 'gold standard' of joint modelling and definitely better from our program.

Table 2: Longitudinal slope β

β	No people	No obs	survreg	aftreg	lme	rizopoulos	bootstrap	analytical	survreg	aftreg	lme	rizopoulos	bootstrap	analytical	survreg	aftreg	lme	rizopoulos	bootstrap	analytical	analytical with se bootstrap
					mean(bias)						mean(se)							coverage			
Scenario 1	100	1to4	1.442	-1.726	-0.021	-0.003	-0.623	-0.010	0.430	0.562	0.057	0.057	1.102	0.002	0.026	0.102	0.934	0.950	0.249	0.057	0.992
	100	1to24	2.246	-1.808	-0.005	0.000	-0.600	-0.043	0.250	0.479	0.023	0.023	1.050	0.001	0.000	0.006	0.936	0.952	0.137	0.003	1.000
	500	1to4	1.371	-1.646	-0.022	-0.004	-0.410	-0.030	0.188	0.243	0.025	0.025	1.115	0.001	0.000	0.000	0.854	0.934	0.009	0.029	1.00
	500	1to24	2.186	-1.738	-0.006	-0.001	-0.413	-0.113	0.110	0.209	0.010	0.010	1.146	0.000	0.000	0.000	0.891	0.929	0.000	0.000	1.000
Scenario 2	100	1to4	1.869	-1.790	-0.028	-0.005	-0.635	0.006	0.622	0.741	0.076	0.075	1.090	0.003	0.056	0.322	0.956	0.958	0.317	0.033	0.997
	100	1to24	3.034	-1.867	-0.007	0.000	-0.588	-0.062	0.391	0.619	0.031	0.031	1.102	0.001	0.000	0.096	0.956	0.968	0.240	0.014	1.000
	500	1to4	1.802	-1.693	-0.026	-0.005	-0.438	-0.008	0.272	0.319	0.033	0.033	1.068	0.001	0.000	0.002	0.890	0.940	0.039	0.017	1.000
	500	1to24	2.974	-1.734	-0.007	-0.001	-0.403	-0.133	0.174	0.268	0.014	0.014	1.128	0.001	0.000	0.000	0.936	0.941	0.000	0.003	1.000

Table 3: Longitudinal intercept λ

λ	No people	No obs	survreg	aftreg	lme	rizopoulos	bootstrap	analytical	survreg	aftreg	lme	rizopoulos	bootstrap	analytical	survreg	aftreg	lme	rizopoulos	bootstrap	analytical	analytical with se bootstrap
					mean(bias)						mean(se)							coverage			
Scenario 1	100	1to4	-2.399	-1.847	0.003	0.002	-0.282	0.053	0.118	0.150	0.033	0.033	1.690	0.001	0.000	0.000	0.954	0.952	0.571	0.014	1.000
	100	1to24	-3.136	-1.744	-0.002	-0.002	-0.258	0.026	0.088	0.181	0.032	0.032	1.691	0.000	0.000	0.000	0.958	0.954	0.656	0.005	1.000
	500	1to4	-2.383	-1.849	0.000	-0.001	-0.098	0.019	0.053	0.067	0.015	0.015	1.649	0.000	0.000	0.000	0.938	0.934	0.780	0.011	1.000
	500	1to24	-3.120	-1.749	-0.001	-0.001	-0.067	-0.007	0.039	0.080	0.014	0.014	1.783	0.000	0.000	0.000	0.972	0.967	0.811	0.000	1.000
Scenario 2	100	1to4	-2.453	-1.843	-0.001	-0.002	-0.235	0.046	6.173	0.165	0.033	0.033	1.774	0.001	0.000	0.000	0.948	0.942	0.668	0.011	1.000
	100	1to24	-3.331	-1.730	-0.002	-0.003	-0.260	0.012	0.114	0.208	0.032	0.032	1.684	0.000	0.000	0.000	0.934	0.934	0.656	0.003	1.000
	500	1to4	-2.445	-1.846	0.000	-0.001	-0.079	0.012	0.061	0.074	0.015	0.015	1.637	0.000	0.000	0.000	0.962	0.960	0.792	0.009	1.000
	500	1to24	-3.317	-1.754	0.000	-0.001	-0.055	-0.017	0.174	0.092	0.014	0.014	1.740	0.000	0.000	0.000	0.948	0.948	0.773	0.003	1.000

Table 4: Survival intercept γ_1

survival intercept	No people	No obs	rizopoulos	bootstrap	analytical	rizopoulos	bootstrap	analytical	rizopoulos	bootstrap	analytical	analytical with se bootstrap
				mean(bias)			mean(se)			coverage		
Scenario1	100	1to4	0.371	-0.940	0.207	0.541	2.559	4.510	0.932	0.820	0.997	0.992
	100	1to24	0.377	-0.687	0.240	0.531	3.237	4.923	0.950	0.954	0.997	1.000
	500	1to4	0.303	-1.243	0.232	0.232	2.497	1.381	0.776	0.597	0.998	1.000
	500	1to24	0.311	-1.004	0.283	0.227	2.697	3.578	0.734	0.886	0.975	1.000
Scenario 2	100	1to4	0.344	-1.003	0.212	0.509	2.594	3.909	0.932	0.844	0.989	1.000
	100	1to24	0.330	-0.810	0.249	0.496	3.271	7.687	0.928	0.967	0.948	1.000
	500	1to4	0.312	-1.213	0.223	0.224	2.452	1.431	0.740	0.654	0.944	1.000
	500	1to24	0.320	-1.294	0.281	2.691	0.218	3.137	0.728	0.869	0.955	1.000

Table 5: Survival slope γ_2

γ	No people	No obs	survreg	aftreg	rizopoulos	bootstrap	analytical	survreg	aftreg	rizopoulos	bootstrap	analytical	survreg	aftreg	rizopoulos	bootstrap	analytical	analytical with se bootstrap
					mean(bias)					mean(se)					coverage			
Scenario 1	100	1to4	-0.003	-0.003	0.005	0.180	0.086	0.146	0.189	0.207	0.985	2.803	0.920	0.918	0.932	0.992	0.997	0.989
	100	1to24	-0.002	0.003	-0.010	-1.467	0.035	0.082	0.194	0.204	2.573	2.449	0.966	0.906	0.952	0.919	1.000	1.000
	500	1to4	0.004	0.001	0.000	0.155	0.047	0.065	0.084	0.091	0.727	0.860	0.940	0.948	0.952	0.996	0.993	1.000
	500	1to24	0.003	0.001	0.000	0.480	-0.019	0.036	0.086	0.091	1.390	1.762	0.959	0.947	0.957	0.958	0.989	1.000
Scenario 2	100	1to4	-1.505	-1.908	0.007	-1.341	0.072	0.167	1.287	0.214	1.348	3.186	0.000	0.000	0.948	0.562	0.994	0.992
	100	1to24	-1.092	-1.970	0.004	-0.667	0.083	0.109	0.222	0.213	2.383	9.540	0.000	0.000	0.940	0.721	0.986	1.000
	500	1to4	-1.493	-1.895	-0.001	-1.531	0.025	0.075	0.092	0.096	1.396	1.431	0.000	0.000	0.948	0.569	0.978	1.000
	500	1to24	-1.095	-1.947	0.003	-1.104	0.021	0.048	0.099	0.095	1.595	2.729	0.000	0.000	0.948	0.861	0.979	0.997

Table 6: Survival standard deviation $\log \sigma$

$log(\sigma)$	No people	No obs	survreg	aftreg	rizopoulos	bootstrap	analytical	survreg	aftreg	rizopoulos	bootstrap	analytical	survreg	aftreg	rizopoulos	bootstrap	analytical	analytical with se bootstrap
			mean(bias)					mean(se)					coverage					
Scenario 1	100	1to4	-0.546	-0.290	0.017	-2.003	-0.065	0.088	0.092	0.109	5.026	0.046	0.000	0.074	0.932	0.000	0.374	1.000
	100	1to24	-1.131	-0.262	0.009	-1.467	-0.029	0.093	0.094	0.107	3.627	3.557	0.000	0.174	0.928	0.005	0.306	1.000
	500	1to4	-0.531	-0.282	0.006	-2.542	-0.019	0.039	0.041	0.048	4.864	0.016	0.000	0.000	0.930	0.000	0.225	1.000
	500	1to24	-1.116	-0.257	0.002	-1.810	0.046	0.041	0.042	0.048	3.718	0.017	0.000	0.000	0.924	0.000	0.139	1.000
Scenario 2	100	1to4	-0.444	-0.257	0.026	-2.885	-0.050	0.088	0.090	0.099	5.915	0.183	0.006	0.178	0.914	0.000	0.565	1.000
	100	1to24	-0.899	-0.234	0.024	-1.746	-0.063	0.093	0.093	0.098	4.147	0.456	0.000	0.282	0.912	0.005	0.483	1.000
	500	1to4	-0.426	-0.242	0.005	-3.270	-0.015	0.039	0.040	0.044	6.138	0.082	0.000	0.000	0.906	0.000	0.303	1.000
	500	1to24	-0.879	-0.224	0.005	-2.423	0.023	0.041	0.042	0.044	4.825	0.144	0.000	0.000	0.906	0.000	0.208	1.000

Table 7: Shared parameter α

α	No people	No obs	rizopoulos	bootstrap	analytical	rizopoulos	bootstrap	analytical	rizopoulos	bootstrap	analytical	analytical with se bootstrap
			mean(bias)			mean(se)			coverage			
Scenario 1	100	1to4	0.063	-0.766	0.263	0.409	1.172	4.49	0.943	0.662	0.986	0.984
	100	1to24	0.074	-0.808	0.235	0.403	2.214	6.557	0.954	0.683	0.997	0.995
	500	1to4	0.004	-0.553	0.051	0.176	1.118	1.317	0.948	0.804	0.996	1.000
	500	1to24	0.009	-0.654	0.056	0.171	1.450	2.926	0.949	0.875	0.994	1.000
Scenario 2	100	1to4	0.026	-0.787	0.215	0.383	1.137	4.111	0.930	0.729	0.975	0.995
	100	1to24	0.032	-0.692	0.194	0.373	2.218	6.468	0.962	0.738	0.724	0.997
	500	1to4	0.007	-0.624	0.027	0.168	1.125	1.410	0.948	0.844	0.959	1.000
	500	1to24	0.014	-0.484	0.011	0.163	1.746	2.268	0.953	0.931	0.920	1.000

Table 8: Longitudinal standard deviation σ_e

σ_e	No people	No obs	lme	rizopoulos	bootstrap	analytical
			mean(bias)			
Scenario 1	100	1to4	0.000	-0.001	-0.096	-0.094
	100	1to24	0.000	0.000	-0.096	-0.094
	500	1to4	0.000	0.000	-0.095	-0.094
	500	1to24	0.000	0.000	-0.094	-0.094
Scenario 2	100	1to4	-0.001	-0.001	-0.095	-0.094
	100	1to24	0.000	0.000	-0.096	-0.094
	500	1to4	0.000	0.000	-0.095	-0.094
	500	1to24	0.000	0.000	-0.095	-0.094

Table 9: Variance of the longitudinal random effects D_{11}

D_{11}	No people	No obs	lme	rizopoulos	bootstrap	analytical
			mean(bias)			
Scenario 1	100	1to4	0.000	-0.001	0.054	0.530
	100	1to24	-0.001	-0.002	-0.004	0.543
	500	1to4	0.000	0.000	0.009	0.262
	500	1to24	0.000	0.000	0.001	0.287
Scenario 2	100	1to4	-0.001	-0.002	0.023	0.503
	100	1to24	0.000	-0.001	-0.010	0.450
	500	1to4	0.000	-0.001	0.035	0.250
	500	1to24	-0.001	-0.001	0.004	0.280

3.4 Summary

In the previous chapter, a shared parameter model for longitudinal and time-to-event data, under left-truncation and right-censoring was introduced. Gauss-Hermite approximation was applied for maximum likelihood estimation. This chapter described the simulation scenarios and the estimation quality of the parameters for the model described in the previous chapter, using a Weibull survival sub-model.

The estimation quality for the proposed maximum-likelihood and the bootstrap method were assessed in respect to the mean bias of the estimators, the standard errors, and the actual coverage at 95% nominal level for a Weibull survival sub-model. Median was not used as it provided no different results than the mean. Results were compared to the results from the *lme* package for the longitudinal component, and to the results from the $survreg$, $aftreg$ survival packages for the survival component and the joint modelling JM package.

The program implementing this method and providing two estimation procedures (analytical and bootstrap) was assessed by simulation using 10 nodes for Weibull survival distribution.

Overrall, the analytical procedure is quite good, though the variance components are not estimated with sufficient quality. This could be due to a fair number of nodes, but it should be noted that the increase in number of nodes increased the computing time exponentially. The most trustworthy method is to combine the analytic point estimation with bootstrap standard errors. Rizopoulos program, however, is more reliable in providing more accurate estimates.

The details enabling the application of the joint modelling theory to other parametric distributions can be found in the Appendices C-I. However, it should be noted that the quality of our procedure for other parametric models needs to be assessed by simulation.

The next chapter will provide two medical applications of the shared parameter model with Weibull survival distribution.

4 Applications

4.1 Introduction

In the previous chapter, the estimation quality for the proposed maximum-likelihood and the bootstrap methods in respect to the mean bias of the estimators, the standard errors, and the actual coverage at 95% nominal level for a Weibull survival sub-model were assessed and compared with the results from the *lme* package for the longitudinal component, and from the *survreg* and $aftreg$ survival packages for the survival component.

In this chapter the shared parameter model with the Weibull survival submodel presented in this thesis will be used to analyse two medical datasets, the Prostate Cancer data and Primary Biliary Cirrhosis data.

4.2 Prostate Cancer Data

4.2.1 Introduction

Prostate cancer (PCa) is the growth of cancerous cells in the prostate (26). The easiest clinical way to diagnose it is through PSA (Prostate-Specific Antigen) screening.

The human-specific prostate antigens were identified in 1960, but they could not be isolated (44). Later on, investigations on human semen led to the discovery of a unique antigen, similar to PSA (53) (54) and in 1980 it was found that PSA in prostate was similar to the

PSA in serum. From that point prostate cancer could be diagnosed by a serum test (96). Thus, PSA can be used as a longitudinal biomarker for prostate cancer and therefore it can be used as an example for the joint modelling methodology that is developed in this thesis. PSA can be used as a diagnostic biomarker for prostate cancer screening, and also as a prognostic biomarker, to ascertain the possible course of cancer and its treatment.

Prostate cancer screening has been controversial, however, and randomized trials gave conflicting results. Schroder et al (2009) (117) and Welch and Albertsen (2009) (147) found that PSA screening leads to overdiagnosis, but Schroder et al (2009) (117) also found that PSA screening decreased the death rate, while Andriole et al (2009) (7) and Sandblom et al (2011) found that for both male groups in their study, the screening group and the control group, the rate of death from prostate cancer did not differ significantly.

PSA value boundary for prostate cancer is even more controversial. Some researchers found that no PSA value boundary can be set and that new biomarkers need to be associated with prostate cancer (57) (61), while others concluded that limit values for PSA can be set, but only when there are no additional risk factors (116). In the presence of other risk factors, the PSA values cannot be considered as accurate diagnostic tools (116). Other researchers provide more specific PSA limits for different ages. Vickers et al. (140) found that 60-year-old males that have PSA less than 1 nm/ml might develop a prostate cancer, but it will not possibly be fatal and thus, no further screenings are required. Other authors categorized the PSA values into 3 different categories for males at the age of 60 (18). They concluded that males that have a PSA level more than 2 ng/mL, should undergo further screening, while those with PSA level less than 1 ng/mL do not need to do this. For PSA levels of 1-2 ng/mL it is not clear what the best strategy is (18).

4.2.2 Data Description

The data on PCa were provided by Bettencourt-Silva et al.(2011) (11). These data were collected from multiple hospital information systems in Norfolk and Norwich University Hospitals (administration, biochemistry, histopathology, radiotherapy, radiology, oncology) and were validated and complemented with data from the local cancer registry. Further information on data collection methods can be found in Bettencourt-Silva et al.(2011) (11).

The data collection covered the period from 6/1/2004 to 26/7/2014. The entry age is up to 75 years (inclusive). The initial number of patients was 1154. 7 patients that had had other cancers, before they were treated for prostate cancer were removed from the data, resulting in the total sample size of 1147.

We have grouped the treatments into the following four types on the intention to treat basis: Hormone therapy (H), Hormone and Radiotherapy (HR), Surgery (S) or Watchful Waiting (W).

A list of pre-existing comorbidities was compiled from hospital records. These were subdivided into the following three classes: Heart (CVA/CVD-Ischaemic heart and Cerebrovascular diseases), Respiratory (Chronic Lower Respiratory diseases and Infuenza and Pneumonia), and Prostate diseases (Inflammatory disease of Prostate, Prostate hyperplasia and other disorders of the Prostate). For more details see Table 11.

A comprehensive list of blood test results up to one year prior to prostate cancer treatment was compiled by Bettencourt-Silva et al.(2011) (11). The blood test results used in further survival modelling included: urea, white blood count (WBC), creatinine, haemoglobin (Hg), Mean Corpuscular Volume (MCV), Mean Corpuscular Haemoglobin (MCH) and

Sodium. See Table 12 for details.

Multivariate survival modelling included patients at cancer stage 2 with treatments H, HR, S and W; cancer stage 3 with treatments H, HR and S, and cancer stage 4 with treatment H. There were not enough patients and/or events for other stage/treatment combinations. For instance, there were only 9 deaths in 124 patients at cancer stage 1. We also excluded 26 patients at cancer stage 4, 34 patients at stage 2 and 34 patients at stage 3 who received 'atypical' treatments. These 218 patients in total were excluded from survival modelling.

The final dataset also includes cancer staging (2 to 4) which shows how far the cancer has spread. For patients with cancer stage 2 (Total: 662) there are 87 patients that take H treatment, 212 that take HR treatment, 272 that take S treatment and 91 that take W. For the patients that have cancer stage 3 (Total: 206), 31 receive H treatment, 39 are under HR treatment and 136 took S treatment. The 67 patients that have cancer stage 4 are under H treatment. See Table 10 for more details.

Table 10 also includes information on PSA measurements (these are right-skewed, and some are missing), and death (165 patients out of the 935 patients died i.e. 17.65%).

Most of the 935 remaining patients (91.97% of the patients) are also ascribed a Gleason score. This is based on a biopsy and it shows how aggresive the cancer is, i.e. how likely the tumour it is to grow and spread outside the prostate (67). For a patient, Primary and Secondary Gleason scores are recorded, but the medical professionals often use sum Gleason, i.e. if the Primary and the Secondary Gleason are 3, then the sum Gleason would be 6. Sum Gleason is divided into two categories, where the sum is "6 or 7" for 701 patients (74.97%), or "8 to 10" for 159 patients (17.01%). See Table 13 for more details.

PSA measurements are right-skewed with minimal missing percentage (Check Table 10).

Table 10: Characteristics of the patients with PCa by Cancer Stage and Treatment (PSA, age, follow-up time, No of Deaths)

				PSA							age						follow up time						Death		Death	from PCa
Cancer Stage	Treatment	No of patients	%**	%* missing	min	1st qu.	median	mean	3rd qu.	max	min	1st qu.	median	mean	3rd qu.	max	min	1st qu.	median	mean	3rd qu.	max	No of deaths	%*	No of deaths	%*
2	H	87	13.14	1.15	0.60	17.55	33.50	84.92	59.40	2383.00	53.70	66.95	70.40	68.95	72.55	75.00	182	973	1426	1526	2139	2857	35	40.23	9	10.34
	HR	212	32.02	0.47	0.90	8.40	12.80	15.70	20.05	72.70	50.70	63.10	68.25	67.03	71.10	75.00	135	1126	1448	1586	2052	2930	23	10.85	5	2.36
	S	272	41.09	6.25	0.10	5.70	7.40	10.04	10.20	212.60	41.30	60.77	65.20	64.43	69.23	75.00	47	817	1278	1405	2013	3458	30	11.03	5	1.84
	W	91	13.75	0	0.500	5.750	7.200	7.921	9.950	19.800	52.00	66.40	70.00	68.99	73.05	75.00	20	1074	1498	1598	2154	3407	15	16.48	1	1.10
Total		662																					103	15.56	20	3.02
3	H	31	15.05	0	2.60	18.60	45.80	90.55	113.10	660.10	59.70	69.30	71.70	70.41	72.75	74.80	64.0	602.5	1346.0	1397.0	2002.0	2960.0	15	48.39	5	16.13
	HR	39	18.93	5.13	4.00	13.00	16.70	19.35	19.90	111.10	54.20	64.75	69.70	67.86	71.55	74.50	430	1608	1869	1853	2304	2703	7	17.95	1	2.56
	S	136	66.02	0.74	2.600	5.900	8.200	9.425	12.200	25.000	45.70	60.88	64.95	64.21	68.32	75.00	53.0	696.8	1230.0	1292.0	1810.0	3156.0	7	5.15	2	1.47
Total		206																					29	14.08	8	3.88
4	H	67	100	1.49	0.6	33.4	103.2	601.7	413.3	8101.0	54.20	61.95	66.90	66.24	70.85	74.90	108	834	1220	1306	1730	3208	33	49.25	12	17.91
Total		67																					33	49.25	12	17.91
Grand Total		935																					165	17.65	40	4.28

$*$: % out of cancer stage by treatment

$**$: % out of cancer stage

Table 11: Comorbidities by Cancer Stage and Treatment

			Prostate				CVA/CVD				Respiratory			
Cancer Stage	Treatment	Total	After Treatment No of patients	Treatment %∗	Before Treatment No of patients	Treatment %∗	After Treatment No of patients	Treatment %∗	Before Treatment No of patients	Treatment % ∗	After Treatment No of patients	Treatment %∗	Before Treatment No of patients	Treatment %∗
2	H	87	10	11.49	6	6.90	30	34.48	18	20.69	18	20.69	3	3.45
	HR	212	38	17.92	26	12.26	40	18.87	21	9.91	28	13.21	8	3.77
	S	272	65	23.90	61	22.43	36	13.24	24	8.82	26	9.56	22	8.09
	W	91	16	17.58	13	14.29	22	24.18	16	17.58	16	17.58	9	9.89
Total		662	129	19.49	106	16.01	128	19.34	79	11.93	88	13.29	42	6.34
3	H	31	3	9.68	3	9.68	11	35.48	4	12.90	2	6.45	1	3.23
	HR	39	5	12.82	4	10.26	9	23.08	8	20.51	8	20.51	5	12.82
	S	136	14	10.29	14	10.29	19	13.97	12	8.82	16	11.76	11	8.09
Total		206	22	10.68	21	10.19	39	18.93	24	11.65	26	12.62	17	8.25
4	H	67	11	16.42	10	14.93	15	22.39	7	10.45	4	5.97	1	1.49
Total		67	11	16.42	10	14.93	15	22.39	7	10.45	4	5.97	1	1.49
Grand Total		935	162	17.33	137	14.65	182	19.47	110	11.76	118	12.62	60	6.42

∗ : % out of cancer stage by treatment

Table 12: Blood Test Results by Cancer Stage and Treatment

Blood	Treatment	H Total	%∗ missing	min	1st qu.	median	mean	3rd qu.	max	HR Total	%∗ missing	min	1st qu.	median	mean	3rd qu.	max	S Total	% ∗ missing	min	1st qu.	median	mean	3rd qu.	max	W Total	% ∗ missing	min	1st qu.	median	mean	3rd qu.	max
	Cancer Stage																															
Creatinine	2	87	24.14	59.00	87.25	97.50	101.90	112.80	236.00	212	17.45	60.00	84.00	91.00	94.67	101.50	214.00	272	1.10	58.00	83.00	91.00	95.05	102.00	216.00	91	23.08	64.00	83.00	91.00	95.99	102.00	223.00
	3	31	12.90	70.00	83.00	96.00	98.22	103.00	163.00	39	23.08	72.00	86.25	92.00	95.67	106.20	138.00	136	3.68	60.00	78.50	88.00	89.65	99.50	139.00								
	4	67	28.36	66.00	81.75	93.00	100.80	116.00	184.00																							
Urea	2	87	24.14	3.300	5.250	6.400	6.356	7.275	10.700	212	17.92	2.200	5.025	6.000	6.166	6.900	14.300	272	1.84	2.800	5.050	5.900	6.084	6.950	17.600	91	24.18	2.100	5.300	6.200	6.455	7.400	15.800
	3	31	9.68	3.300	4.850	5.750	6.061	6.800	11.200	39	17.95	3.700	5.050	6.000	5.891	6.500	9.300	136	1.47	2.400	4.900	5.700	5.833	6.600	9.900								
	4	67	28.36	3.200	5.275	6.800	7.050	7.900	12.900																							
Hg	2	87	28.74	8.80	13.60	14.50	14.38	15.30	18.00	212	31.13	9.20	14.10	14.80	14.71	15.60	18.40	272	3.68	7.70	14.00	14.80	14.53	15.40	17.50	91	38.46	11.00	13.68	14.60	14.36	15.02	16.90
	3	31	6.45	8.4	13.9	15.0	14.7	15.8	18.9	39	30.77	10.90	14.20	14.90	14.74	15.30	17.80	136	5.15	8.80	14.10	14.80	14.65	15.60	17.10								
	4	67	47.76	8.20	13.60	14.40	13.93	15.20	16.40																							
WBC	2	87	22.99	3.000	6.200	7.100	7.749	9.200	18.200	212	29.25	3.300	6.000	7.200	7.528	8.625	17.500	272	2.94	3.200	5.700	6.900	7.415	8.200	30.600	91	36.26	4.100	5.575	6.400	6.898	7.875	12.200
	3	31	9.68	4.900	5.900	7.200	7.575	8.825	11.300	39	28.21	3.800	6.250	7.050	7.736	9.100	13.200	136	2.21	4.000	5.900	7.000	7.399	8.400	15.900								
	4	67	44.78	4.500	6.600	8.200	8.111	9.200	13.100																							
MCV	2	87	41.38	81.00	88.00	90.00	90.51	93.00	106.00	212	37.26	79.00	88.00	90.00	90.92	94.00	107.00	272	9.93	62.00	87.00	89.00	89.21	92.00	100.00	91	45.05	77.00	87.00	89.50	89.68	93.00	97.00
	3	31	16.13	80.00	88.00	90.00	89.62	93.00	96.00	39	33.33	82.00	89.00	90.00	91.15	92.75	100.00	136	11.03	81.00	87.00	89.00	89.64	92.00	102.00								
	4	67	49.25	79.00	86.00	88.50	88.71	92.00	99.00																							
MCH	2	87	24.14	26.70	29.58	30.60	30.69	31.85	34.90	212	30.19	25.20	29.60	30.75	30.72	31.50	36.70	272	3.31	20.70	29.50	30.40	30.34	31.20	34.30	91	39.56	24.30	29.65	30.50	30.64	31.80	33.60
	3	31	12.90	26.30	29.50	30.40	30.23	31.30	33.20	39	28.21	27.00	30.42	30.90	30.95	31.50	33.70	136	4.41	26.90	29.40	30.35	30.43	31.28	34.50								
	4	67	41.79	26.50	28.90	30.10	29.91	30.90	33.70																							
Sodium	2	87	34.48	129.0	138.0	140.0	139.7	141.0	145.0	212	33.02	128	139	140	140	142	146	272	12.13	130.0	139.0	140.0	140.2	142.0	146.0	91	37.36	130.0	139.0	141.0	140.3	142.0	146.0
	3	31	22.58	134.0	139.8	141.0	141.2	143.2	149.0	39	43.59	137.0	140.0	141.0	141.4	142.8	146.0	136	11.03	131.0	139.0	140.0	139.9	141.0	145.0								
	4	67	38.81	129	137	139	139	141	145																							

∗ : % out of cancer stage by treatment

Table 13: PSA and Sum Gleason Scores by Cancer Stage and Treatment

				PSA				Sum Gleason			
				<10		≥ 10		6 or 7		8 to 10	
Cancer Stage	Treatment	No of patients	%**	No of patients	%*	No of patients	%*	No of patients	%*	No of patients	%*
2	H	87	13.14	11	12.64	75	86.21	37	42.53	43	49.43
	HR	212	32.02	74	34.91	137	64.62	163	76.89	41	19.34
	S	272	41.09	185	68.01	70	25.74	256	94.12	10	3.68
	W	91	13.75	68	74.73	23	25.27	78	85.71	1	1.1
Total		662		338	51.06	305	46.07	534	80.66	95	14.35
3	H	31	15.05	5	16.13	26	83.87	7	22.58	13	41.94
	HR	39	18.93	8	20.51	29	74.36	25	64.1	14	35.9
	S	136	66.02	86	63.24	49	36.03	127	93.38	8	5.88
Total		206		99	48.06	104	50.49	159	77.18	35	16.99
4	H	67	100	1	1.49	65	97.01	8	11.94	29	43.28
Grand Total		935		438	46.84	474	50.7	701	74.97	159	17.01

4.2.3 Methods for standard Survival Analysis of Prostate Cancer Data

In this and the next section we consider the values of predictors recorded up to a year before treatment. The reason behind that is because treatment may affect some of the blood value like PSA and because not all patients had blood measurements right before treatment.

Kaplan-Meier survival curves were used to estimate survival by treatment and by stage.

Cox proportional hazards regression was used to model the survival time of the cancer patients. Survival package (130) in R statistical software (103) was used for analysis.

We have considered the following list of possible predictors of survival: age at diagnosis, stage and Gleason score, treatment type on the intention to treat basis, pre-existing co-morbidities by class, PSA and various blood test results (see above). All these predictors were initially included in the model.

Backward elimination at 0.1 significance level was used to obtain the final model.

Proportional hazards assumption was tested by using a chi-square test of 4 different survival time transformations (identity, log, Kaplan-Meier and rank) with the scaled Schoenfeld residuals. The quality of the model was assessed by using the Akaike Information Criterion, Cook's distance (63), O' Quigleys R^2 (70) and Concordance (49). The values of these parameters for the final model are AIC=994.61, 0% of the data has Cook's distance larger than 1, O' Quigleys $R^2 = 0.63$and $Concordance = 0.75$.

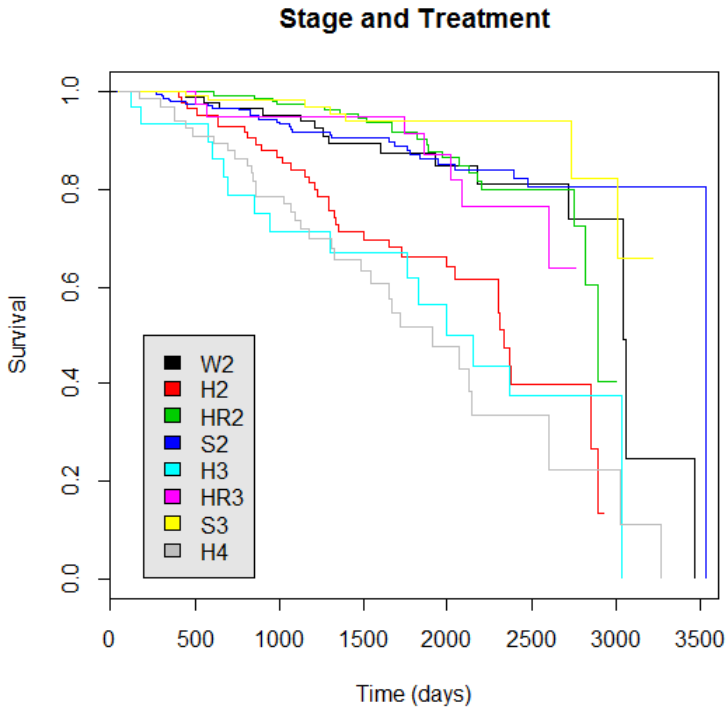


Figure 1: Kaplan-Meier Plot for Prostate Cancer Survival by Cancer Stage and Treatment

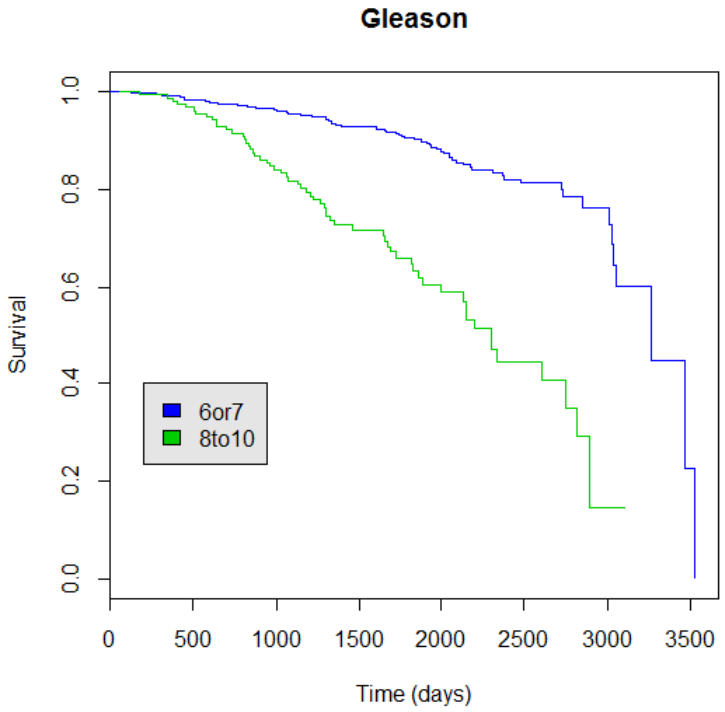


Figure 2: Kaplan-Meier Plot for Prostate Cancer Survival by Sum Gleason Scores

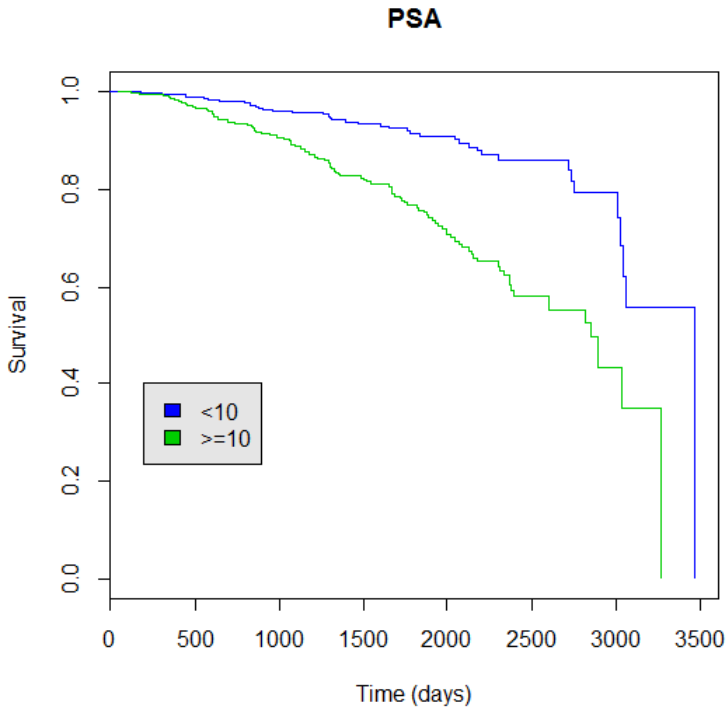


Figure 3: Kaplan-Meier Plot for Prostate Cancer Survival by PSA

Figure 1 shows the Kaplan-Meier Plot for Prostate Cancer Survival by Cancer Stage and Treatment. All groups include censored and dead people by the end of the study. Not all curves, however, drop to zero. When the survival curve drops to zero, it does not mean that all patients in the study died. The curve drops to zero when there is a death after the last censoring and it remains higher when there is no death after the last censoring. So, W2, S2, H3 and H4 groups drop to zero because there were deaths within remaining patients, after the last censoring. H2, HR2, HR3, S3 do not drop to zero because after the last censoring there were no deaths and there are still people alive that haven't being censored or died, making the probability of surviving higher than zero. It should be noted that when there are censored patients, the bottom point of the Kaplan-Meier survival curve is not equal to

the fraction of the patients that have survived. A patient, contributes to this fraction up until he is censored. After censoring the patient does not affect the calculations.

4.2.4 Results of the standard Survival Analysis of the Prostate Cancer Data

Kaplan-Meier plots for Prostate Cancer are given in Figures 1-3. They show that Hormone treatment at all stages seems to have the lowest survival compared with the other treatments, Gleason level 8 to 10 results in worse survival than 6 or 7 and this is also true for PSA larger or equal to 10 compared to PSA less than 10.

In the Cox modelling, the continuous PSA was found not to be significant and we used PSA categorised into two categories in further analysis.

The final Cox regression for survival of the prostate cancer patients is given below. The model includes age at diagnosis, Gleason score category, treatment and stage, PSA category, urea and WBC. The model is based on 666 patients with 96 events. 269 patients were excluded due to missing values in some of the predictors.

Comparing Hormonal treatment across stages we can see the deleterious effect of a higher cancer stage. Survival with hormonal treatment does not differ between stages 2 and 3, but both are significantly better than survival at stage 4 (the baseline), with hazard ratios of survival 2.11 and 2.15, respectively. Hormonal treatment results in the worst survival at all stages, and we shall treat this treatment as the baseline for further comparisons within each stage. At stage 2, Surgery provides 1.320 times better hazard of survival, the HR therapy provides 2.381 times better hazard of survival. The Watchful Waiting provides 1.161 times better hazard of survival. At stage 3, Surgery provides 2.400 times better

hazard of survival and HR provides 1.257 times better hazard of survival. Higher age of diagnosis results in worse survival, the hazard of death increasing 1.04 times per year of age. Combined Gleason score above 7 results in 2.5 times higher hazard of death. The PSA of 10 or above prior to diagnosis results in 1.74 times higher hazard of death. Urea and WBC count are also significant, resulting in 43.380 and 6.890 times higher hazards per unit, respectively.

Table 14: Cox Proportional Hazards Model for PCa

estimates	coef	se(coef)	p
age	0.039	0.023	0.085
treatment and stage H2	-0.746	0.375	0.046
treatment and stage H3	-0.764	0.479	0.111
treatment and stage HR2	-1.614	0.411	0.000
treatment and stage HR3	-0.992	0.492	0.044
treatment and stage S2	-1.024	0.410	0.012
treatment and stage S3	-1.639	0.531	0.002
treatment and stage W2	-0.896	0.538	0.096
PSA 10andmore	0.554	0.262	0.035
gleason 8to10	0.915	0.255	0.000
log(Urea)	1.328	0.392	0.001
log(White Blood Count)	0.655	0.392	0.095

4.2.5 Time-dependent and Joint modelling of Prostate Cancer Data

On average there were 23 longitudinal observations of PSA per patient (Total: 839 patients).

Initially a survival model with time-dependent covariates was fitted using aftreg program from eha R package (15). When a main-effects model (see below) was fitted the results

were similar with standard survival analysis. But PSA as factor was not signifficant and PSA continuous had a coefficient of -0.001 (see Table 15). The program failed to estimate numerous standard errors and therefore, the p-values were not available (for PSA inclusive). Blood tests and comorbidities were not signifficant.

Table 15: Time-Dependent Main-Effects aftreg Model for PCa

Covariate	Coef	se(Coef)	p
age	-0.074	NaN	NaN
gleason	-0.799	0.089	0.000
treatment and stage H2	0	(reference)	
treatment and stage H3	-0.122	0.252	0.628
treatment and stage H4	-0.344	0.229	0.133
treatment and stage HR2	0.276	0.200	0.168
treatment and stage HR3	0.020	0.238	0.933
treatment and stage S2	0.249	0.241	0.301
treatment and stage S3	0.231	0.281	0.411
treatment and stage W2	0.061	0.268	0.821
PSA	-0.001	NaN	NaN
log(scale)	13.854	NaN	NaN
log(shape)	0.359	NaN	NaN

When an interaction of PSA with stage and treatment and sum of Gleason scores was introduced, the results made no sense even though all terms were signifficant (see Table 16). One of the problems is that the higher the PSA was, the higher the log survival time was. To understand this in terms of the interactions, let two patients have a PSA value 5 and 10, with treatment and cancer stage S3 and sum Gleason scores 8 to 10 being the same. For the patient with PSA= 5, the log(survival time)= $5 \times (-0.012 + 0.011 + 0.247) + 0.159 - 0.603 = 0.786$, while for the patient with PSA= 10 the log(survival time)= $10 \times (-0.012 + 0.011 + 0.247) + 0.159 - 0.603 = 2.016$. This shows that the patient with higher PSA has a higher log(survival time), which contradicts established view. Another problem is that surgery in stage 2, S2 seems to have the worse log(survival

time) than W2, when in the Cox proportional hazards model it was the other way around.

Table 16: Model with Time-Dependent Covariates with interactions for PCa obtained by aftreg

Covariate	Coef	se(Coef)	p
age	-0.054	0.013	0.000
gleason	-0.603	0.140	0.000
treatment and stage H2	0	(reference)	
treatment and stage H3	-0.090	0.232	0.698
treatment and stage H4	-0.301	0.214	0.161
treatment and stage HR2	0.265	0.184	0.150
treatment and stage HR3	0.058	0.230	0.801
treatment and stage S2	0.192	0.212	0.365
treatment and stage S3	0.159	0.283	0.574
treatment and stage W2	-0.024	0.432	0.956
PSA	-0.012	0.001	0.000
gleason : PSA	0.011	0.001	0.000
treatment and stage H3: PSA	-0.001	0.000	0.000
treatment and stage H4: PSA	0.001	0.000	0.000
treatment and stage HR2: PSA	-0.012	0.002	0.000
treatment and stage HR3: PSA	-0.001	0.001	0.029
treatment and stage S2: PSA	-0.001	0.000	0.000
treatment and stage S3: PSA	0.247	0.353	0.485
treatment and stage W2: PSA	0.045	0.053	0.391
log(scale)	12.413	0.970	0.000
log(shape)	0.578	NaN	NaN

The analyses using the JM package (with a Cox proportional hazards sub-model and a Weibull sub-model in accelerated failure time form), have not converged. The joint model presented in Chapter 3 also gave odd results. The model found a positive relation between log(PSA) and log(survival time), i.e. increase of log(PSA) by 1, increases log(survival time) by 0.217, see Table 17. There were also an attempt to fit the ratio of the PSA value at each

time point over the first PSA measurement and different PSA transformations, but this led nowhere. A possible reason for the peculiar results of the time-dependent analysis and joint analysis is that PSA is not a good biomarker and indicator of the Prostate Cancer.

Table 17: Analytical Joint Model for PCa

analytical estimates	parameters	se	p
sigma.e	0.921	NA	NA
alpha	0.217	NA	NA
log.sigma	0.844	NA	NA
time	0.000	NA	NA
longitudinal intercept	3.932	NA	NA
gleason	-0.439	NA	NA
treatment and stage H2	-0.512	NA	NA
treatment and stage H3	-0.061	NA	NA
treatment and stage HR2	-0.131	NA	NA
treatment and stage HR3	1.005	NA	NA
treatment and stage S2	-0.300	NA	NA
treatment and stage S3	-0.079	NA	NA
treatment and stage W2	-0.063	NA	NA
survival intercept	-0.414	NA	NA
D11	0.283	NA	NA

Table 18: Bootstrap Joint Model for PCa

bootstrap estimates	parameters	se	p
sigma.e	0.918	0.001	0.000
alpha	0.090	0.022	0.000
log.sigma	0.842	0.003	0.000
time	0.001	0.000	0.000
longitudinal intercept	3.864	0.019	0.000
gleason	-0.554	0.019	0.000
treatment and stage H2	-0.507	0.012	0.000
treatment and stage H3	-0.027	0.022	0.220
treatment and stage HR2	-0.159	0.029	0.000
treatment and stage HR3	1.137	0.020	0.000
treatment and stage S2	-0.107	0.031	0.001
treatment and stage S3	0.030	0.023	0.178
treatment and stage W2	0.160	0.020	0.000
survival intercept	-0.176	0.020	0.000
D11	0.281	0.001	0.000

4.3 Primary biliary cirrhosis (PBC2) data

4.3.1 Data Explanation

Primary Biliary Cirrhosis (PBC) is an autoimmune disease of the liver, see (58) for more details. The Mayo Clinic conducted a follow-up study of 312 patients with the diagnosis of PBC from 1974 to 1984, (91) to compare a drug D-penicillamine (158 patients) to placebo (154 patients). The PBC dataset used here was taken from the JM package, (110). It contains 1945 total observations, i.e. 6.23 observations per patient on average over time, spanning from year 1 to year 11. The median follow-up time was 6.3 years. At each patient

visit the serum bilirubin (a biomarker) measurement (in mg/dl) was taken. Higher values of serum bilirubin are considered to be a strong indicator for the disease progression. Instead of working with serum bilirubin, we are using its log transformation due to its right-skewneness. The patient's age in years at registration (26 to 78 years) is considered to be the truncation time for our model with left-truncation and right-censoring. Time of each visit is counted from registration. The right censoring is due to liver transplantation or the end of the study. By the end of the study 140 patients were dead (69 and 71 in placebo and treatment arms, respectively), and 172 were censored out. See Tables 19 and 20 for more details.

Table 19: Survival Data for PBC

drug	years						age						serBilir						status2	
	Min	1st Qu.	Median	Mean	3rd Qu.	Max	Min	1st Qu.	Median	Mean	3rd Qu.	Max	Min	1st Qu.	Median	Mean	3rd Qu.	Max	0	1
placebo	0.1396	3.6040	6.4000	6.3960	8.8840	14.2200	30.57	41.43	48.11	48.58	55.81	74.53	0.300	0.725	1.300	3.649	3.600	28.000	85	69
D-penicil	0.1123	3.7740	6.2630	6.4260	8.8290	14.3100	26.28	42.98	51.93	51.42	58.91	78.44	0.300	0.800	1.400	2.795	3.200	20.000	87	71

Table 20: Longitudinal Data for PBC

drug	year					
	Min	1st Qu.	Median	Mean	3rd Qu.	Max
placebo	0.0000	0.5229	2.0290	3.0700	5.0150	14.1100
D-penicil	0.0000	0.5284	2.1100	3.2010	5.0610	13.9000

4.3.2 Methods

A Weibull-based analytical model with bootstrap errors for the case of left-truncation and right-censoring was fitted to the data. In this model, the log survival time is given by

$$\log T_i = \alpha[\beta t + b_i + \lambda] + \gamma_1 + M * \gamma_2 + \sigma \times \epsilon_i. \tag{127}$$

In the above model, the biomarker $Y(t)$ (log serum bilirubin) depends on the time of its measurement t (a time-dependent covariate), the patient id (as a random effect b_i) and includes an intercept λ. The survival submodel additionally includes the binary factor treatment denoted by M and taking on values 0 and 1, and an intercept γ_1. The association of log(serum bilirubin) with log(survival time) is denoted by α, and $\log \sigma$ is the scale.

We are mainly interested in the effect that serum bilirubin has on survival time and the changes in serum bilirubin over time, i.e. the association α of the longitudinal biomarker with the log survival time and β.

4.3.3 Results

The coefficients of the fitted model are given in Table 21. The results show that log(survival time) is associated with log(serum bilirubin) negatively, with the slope of $\alpha = -0.386$. Every additional year of the PBC progression decreases the survival time $\exp(0.386 \times 3.934) = 4.57$ times on average. When the patient is on D-penicillamine, the survival time increases $\exp(0.51) = 1.67$ times, compared to being on placebo.

Table 21: PBC Left-Truncated and Right-Censored Model

estimates	parameters	se	p	lower	upper
sigma.e	0.555				
alpha	-0.386	0.021	0.000	-0.428	-0.344
log.sigma	0.685	0.003	0.000	0.679	0.691
year(β)	3.934	0.023	0.000	3.888	3.979
intercept within longitudinal(λ)	-0.265	0.016	0.000	-0.297	-0.233
intercept within survival(γ_1)	-0.319	0.013	0.000	-0.346	-0.293
drug(γ_2)	0.510	0.012	0.000	0.488	0.533
D11	0.103				

4.3.4 Regression Diagnostics

There are no known regression diagnostics methods for parametric joint modelling. Even though this research was not about the regression diagnostics of this type of models, there is a need for developing such tools, and it will be explored in the future. A first thought and a preliminary work on this is to plot scaled score residuals, Cook's distance versus Martingale residuals and cummulative Kaplan-Meier versus Cox-Snell residuals. These were explained in Chapter 2 and they are used for assessing the fit of a parametric survival model, but a further investigation needs to be done for their use in parametric joint modelling.

For a plot of a cummulative Kaplan-Meier versus Cox-Snell residuals the closer the step function is to the line passing from point $(0, 0)$ with gradient 1, the better the fit is. In the PBC example, the joint model seems to be fitting the data really well, except perhaps near $(0, 0)$. See Figure 4. Figure 5 should provide an indication of poorly fitted influential observations. There are 7 patients that have larger absolute martingale values than the rest, but these values are not too large to conclude that they are poorly fitted. Overall, the Cook's distance seems to be well under 1, which indicates that no patient affects the estimates more than the other patients. In Figures 6 to 10, the scaled score residuals seem quite dispersed, but in fact they are not too far away from each other taking into account the scale of the graphs.

Overall, based on these 3 diagnostic methods the model seems to fit well to the data. However, it should be noted that, to the author's knowledge, there is no research indicating how well these 3 methods can perform for a parametric joint model. This is an area for a future research.

It should be noted that there were not any diagnostics for the joint model for PCa, because

it made no sense.

Figure 4: Cummulative Kaplan-Meier Vs Cox-Snell Residuals

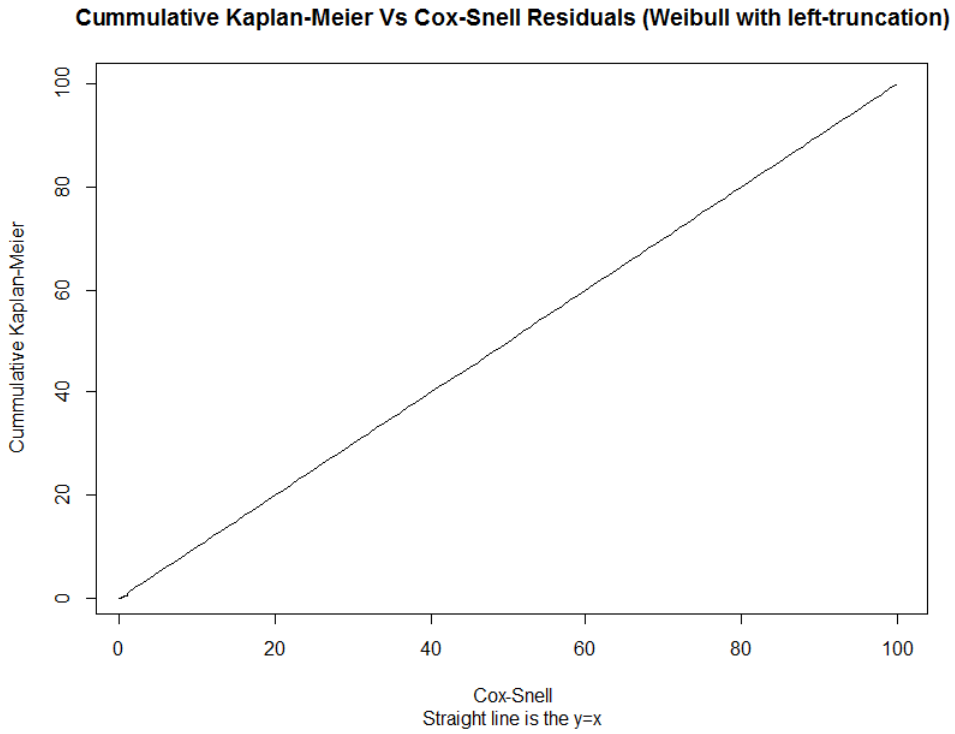


Figure 5: Cooks Distance Vs Martingale Residuals

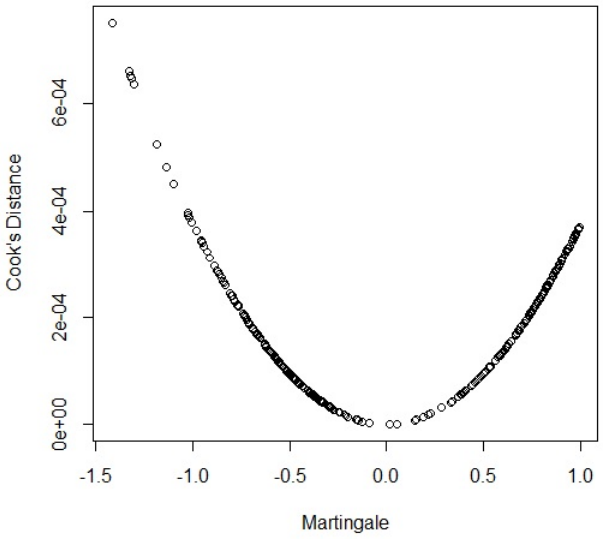


Figure 6: Scaled Score Residuals for association

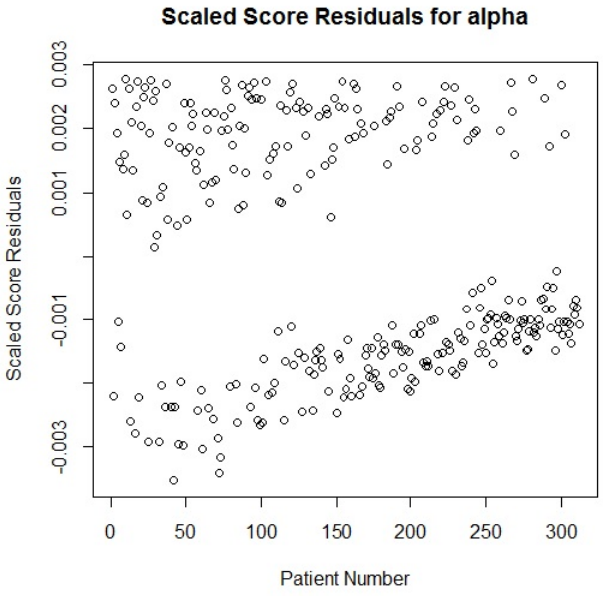


Figure 7: Scaled Score Residuals for drug

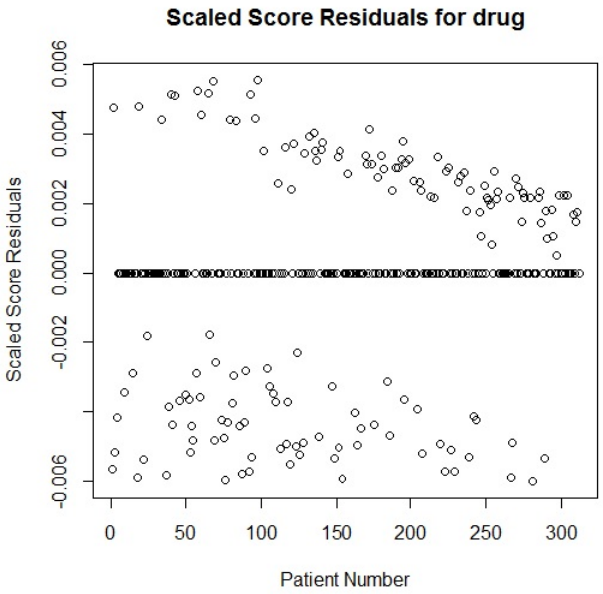


Figure 8: Scaled Score Residuals for log(sigma)

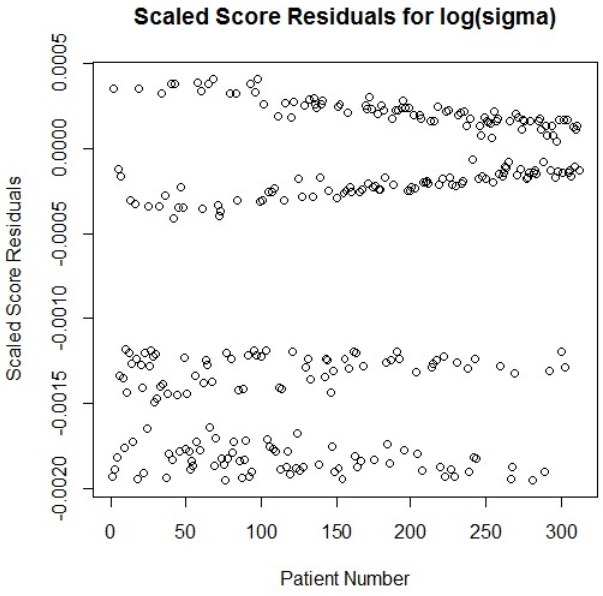


Figure 9: Scaled Score Residuals for longitudinal intercept

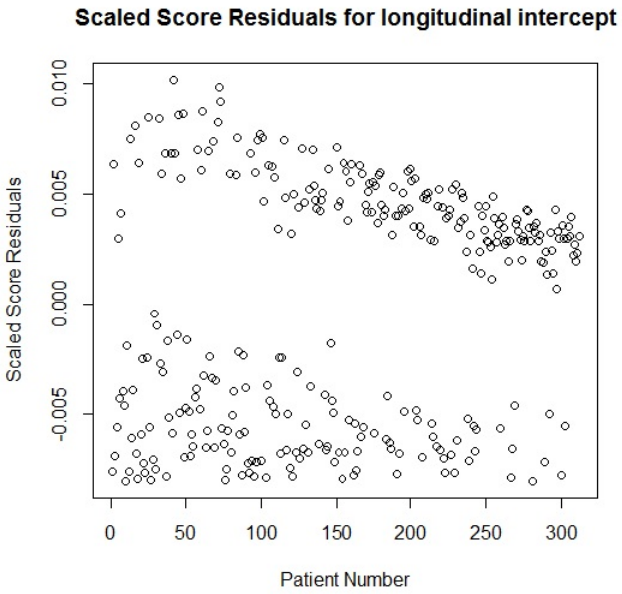
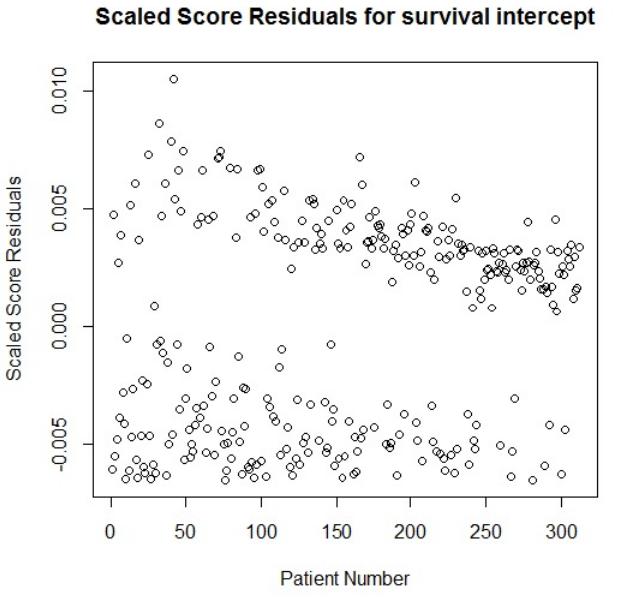


Figure 10: Scaled Score Residuals for survival intercept



4.4 Summary

In this chapter the shared parameter model with Weibull survival sub-model presented in Chapter 3, was applied to two datasets. For prostate cancer data, the joint model did not work well possibly due to the fact that PSA is not a reliable biomarker. For the primary biliary cirrhosis data, the joint model provided reasonable results explaining the data well. The regression diagnostics also indicated that the model fits well to the data, but further investigation needs to be done for the use of diagnostic methods on parametric joint modelling.

In the next chapter the future extensions of the joint model are presented.

5 Future Developments

The simple version of longitudinal data for joint modellling is to have one longitudinal biomarker. There are however, more complicated scenarios where there are multiple longitudinal biomarkers. Additionally, there may be different types of failing. Models that can handle different types of failing are called competing risks models. These different types of failing can also be considered as informative censoring.

Rizopoulos and Ghosh (2011) (114) suggested a joint model relating multiple longitudinal biomarkers to time-to-event data. They assumed that the baseline risk function is piecewise constant and they used a spline-based method for longitudinal outcomes.

Albert and Shih (2011) (5) extended the two-stage regression calibration approach, used in Albert and Shih (2010), for multiple longitudinal measurements and discrete time-to-event data. First, they fitted a longitudinal model for multiple longitudinal measurements and then this model was related to the survival model, while informative drop-out was dealt by regression calibration. They used a two-staged procedure. At first, they fit multivariate linear mixed models to the longitudinal data and then they estimate the time-to-event model by replacing the random effects with corresponding Bayes estimates. For standard errors the bootstrap procedure is used.

Elashoff, Li and Li (2007) (38) extended the work of Henderson et al (2000) (56). They suggested a joint model for repeated measurements and competing risks failure time data, which considers more than one failure types. The longitudinal components and the competing risks failure time data are linked together by latent random variables. They assume that each subject can experience one of g distinct failure types or can be right-censored. Their joint model consists of a linear mixed effects model for the longitudinal component, a

marginal probability of the i^{th} subject failing from k^{th} risk and a conditional cause-specific hazard model for the k^{th} risk.

In this chapter future extensions of the shared parameter model presented in chapter 3 are presented. These are a shared parameter model with multiple biomarkers and competing risks, under left-truncation and right-censoring.

5.1 Joint Modelling with Multiple Longitudinal Markers

5.1.1 Longitudinal Sub-Model

Suppose there are n individuals which we observe at m_i time points each. These could be different for each individual. Let β be a vector denoting time-dependent fixed effects and b_i be a vector denoting the random effects for the i^{th} individual, where $i = 1, 2, ..., n$. Let λ be a vector denoting the time-independent fixed effects. Suppose $Y_{ik}(t_{ijk})$ is the observed value of the k^{th} longitudinal biomarker for the i^{th} individual at the j^{th} time point, where $j = 1, 2, ..., m_i$, $k = 1, 2, ..., K$. All biomarkers may not have the same timepoints. $W_{ik}(t_{ijk})$ is the conditional expectation of the k^{th} longitudinal biomarker $Y_{ik}(t_{ijk})$ for the i^{th} individual at the j^{th} time-point, given b_{ik}. $b_i = b_{i1}, ..., b_{iK}$ are the random effects for the i^{th} individual for the k^{th} biomarker.

The observed values of the longitudinal biomarkers are

$$Y_{ik}(t_{ijk}) = X_{ik}^T(t_{ijk})\beta + Z_{ik}^T(t_{ijk})b_{ik} + U_{ik}^T\lambda + e_{ik}(t_{ijk}). \tag{128}$$

Let $Y_{ik} = \{Y_{i1}, Y_{i2}, ..., Y_{iK}\}$ be the set of the k^{th} longitudinal biomarkers for the i^{th} individual that includes all time points, to make notation easier. $e_{ik}(t_{ijk})$ is a $m_{ik} \times 1$

vector that denotes measurements errors. $e_{ik}(t_{ijk})$ is independent of all variables and has independent marginal distribution $N(0, \sigma_e^2)$. The correlation between repeated measurements, for the same k^{th} longitudinal biomarker, in longitudinal process is given by $cov(e_i(t_{ijk}), e_i(t_{ij'k})) = 0$ for $j \neq j'$. For each k^{th} longitudinal biomarker, suppose, there exist r time-independent fixed parameters and thus, λ is a $r \times 1$ vector of time-independent fixed effects that links U_i^T, a $m_i \times r$ matrix to the longitudinal observations $Y_i(t_{ij})$. If there are q random effects, then b_i is a $q \times 1$ vector, while $Z_i^T(t_{ij})$ is a $m_i \times q$ design matrix for random effects. $b_i \sim MVN(0, D)$ and D is a $q \times q$ variance-covariance matrix. Assuming that there are p time-dependent fixed parameters, β is $p \times 1$ vector and $X_i^T(t_{ij})$ is a $m_i \times p$ design matrix for time-dependent fixed effects.

Now, let

$$W_{ik}(t_{ijk}) = X_{ik}^T(t_{ijk})\beta + Z_{ik}^T(t_{ijk})b_{ik} + U_{ik}^T\lambda. \tag{129}$$

Thus, the model is given by

$$Y_{ik}(t_{ijk}) = W_{ik}(t_{ijk}) + e_{ik}(t_{ijk}). \tag{130}$$

5.1.2 Survival Sub-Model

Let T_i^* denote the true survival time for the i^{th} individual. Let C_i and A_i denote the censoring and truncation time for the i^{th} individual. The observed survival time for the i^{th} individual is given by $T_i = \min(T_i^*, C_i)$. $\delta_i = I(T_i^* \leq C_i)$ is the censoring indicator, which is equal to 0 if the individual is censored and equal to 1 if the survival time was observed.

Now, the longitudinal regression is related to the survival via α in equation (133), where $\alpha = \{\alpha_1, ..., \alpha_K\}$, with α_k denoting the relation of the k^{th} longitudinal biomarker to survival. If $\alpha_k = 0$, then this means that there is no relation between the survival time and the k^{th}

longitudinal biomarker.

Let h(t) be an arbitrary function of the survival time. Two important cases are:

$$h(t) = \begin{cases} \log(t), \text{for an accelerated failure time model} \\ t, \text{for the other parametric models.} \end{cases} \tag{131}$$

For the i^{th} individual the survival is modelled as

$$h(t_i) = \mu_i(t_i) + \sigma \times \epsilon_i, \tag{132}$$

where σ is the $\log(scale)$ and $\mu_i(t)$ is given by

$$\mu_i(t) = \sum \alpha_k W_{ik}(t) + M_i^T \gamma, \tag{133}$$

where γ is the regression coefficient that links the time-independent covariates, M_i^T, related only to survival.

Denote by $x_i^T(t_{t_{ijk}})$ and $z_i^T(t_{ijk})$ the rows of the design matrices $X^T(t_{ijk})$ and $Z^T(t_{ijk})$ from equation (128). These vectors for the k^{th} biomarker are observed at particular time points t_{ijk}. Let t_{ik}, be the set of all time-points for the i^{th} individual and for the k^{th} biomarker, i.e. $t_{ik} = \{t_{i1k}, t_{i2k}, t_{im_ik}\}$ with $j = 1, \cdots, m_i$. Let us extend this notation to an arbitrary time point t, to include the $p \times k$ matrix $x_{ik}^T(t)$, the $q \times k$ matrix $z_{ik}^T(t)$ and the k^{th} longitudinal biomarker $Y_{ik}(t_{ik})$.
Then, the joint model survival $h_i(t)$ for the i^{th} individual at an arbitrary (single) time-point t is given by

$$h_i(t) = \alpha_k[x_{ik}^T(t)\beta + z_{ik}^T(t)b_{ik} + u_{ik}^T\lambda] + M_i^T\gamma + \sigma \times \epsilon_i. \tag{134}$$

ϵ_i is a vector of $n \times 1$, σ and α are numeric. Again, it is assumed that censoring and drop-out mechanisms are non-informative. Specificaly, these two mechanisms are considered to be independent of the random effects, the survival time and the longitudinal measurements, (134).

The joint likelihood contribution is given by

$$L(Y_i, T_i, A_i, \delta_i; \theta) = \int L(Y_i|b_i; (\theta_y, \sigma_e^2))L(T_i, A_i, \delta_i|b_i; (\theta_t, \sigma^2))\phi_q(b_i; D)db_i, \tag{135}$$

where $L(Y_i|b_i; (\theta_y, \sigma_e^2)) = \prod_{k=1}^{K} \left\{ \prod_{j=1}^{m_i} \phi\left(Y_{ijk}(t_{ijk}); W_{ijk}(t_{ijk}), \sigma_e^2\right) \right\}$, (114).

The likelihood includes integral which is difficult to solve analytically, and thus an integral approximation is needed. Gauss-Hermite or Laplace approximation can be applied and a maximum likelihood can be used to estimate the parameters.

The next step is the development of a program implementing this methodology.

5.2 Competing Risks

In this section, let us assume that there are $G - 1$ distinct failure types.

5.2.1 Longitudinal Sub-model

Let $t_i = \{t_{ij}, j = 1, 2, ..., m_i\}$, be the set of m_i time-points, at which the i^{th} individual is observed, these could differ for each individual, and $Y_i(t_i) = \{Y_i(t_{ij}), j = 1, ..., m_i\}$ be a vector of the observed values of a longitudinal biomarker.
Consider the following mixed linear model for the subject-specific observed values of the longitudinal biomarker:

$$Y_i(t_i) = X_{1i}^T(t_i)\beta_1 + Z_{1i}^T(t_i)b_{1i} + U_{1i}^T\lambda_1 + e_i(t_i), \tag{136}$$

where β_1, λ_1 and b_{1i} are the vectors of p time-dependent fixed effects, r time-independent fixed effects, and q random effects for the i^{th} individual with the corresponding design matrices $X_{1i}^T(t_i)$, U_{1i}^T, and $Z_{1i}^T(t_i)$, respectively. The matrix U_{1i}^T has m_{1i} identical rows u_{1i}^T. An m_{1i}-vector of the independent and normally distributed measurement errors $e_{1i}(t_{ij}) \sim \mathcal{N}(0, \sigma_{1e}^2)$, so that the correlation between repeated measurements in the longitudinal process is $(e_{1i}(t_{ij}), e_{1i}(t_{ij'})) = 0$ for $j \neq j'$. The random effects b_{1i} have multivariate normal distribution, so that

$$b_{1i} \sim MVN(0, D_1), \tag{137}$$

with a $q_1 \times q_1$ variance-covariance matrix D_1. The errors e_i and the random effects b_i are mutually independent.

Denote by $W_{1i}(t_i) = \{W_{1i}(t_{ij}), j = 1, ..., m_i\}$ the conditional expectation of the longitudinal biomarker $Y_i(t_i)$ given b_{1i}. Then

$$E[Y_i(t_i)|b_{1i}] = W_{1i}(t_i) = X_{1i}^T(t_i)\beta_1 + Z_{1i}^T(t_i)b_{1i} + U_{1i}^T\lambda_1, \tag{138}$$

and the longitudinal biomarker is given by

$$Y_i(t_i) = W_{1i}(t_i) + e_{1i}(t_i). \tag{139}$$

Therefore conditionally on b_i, the biomarker is normally distributed

$$Y_i(t_{ij})|b_{1i} \sim \mathcal{N}(W_{1i}(t_{ij}), \sigma_{1e}^2) \tag{140}$$

5.2.2 Marginal Failure Probability

The marginal probability of the i^{th} subject failing from k^{th} risk is given by:

$$\pi_g(X_{2i}, W_{ki}; \theta) = \frac{\exp\left\{X_{2i}^T(t_i)\beta_2 + Z_{2i}^T(t_i)b_{2i} + U_{2i}^T\lambda_2\right\})}{1 + \sum_{g=1}^{G-1} \exp\left\{X_{2i}^T(t_i)\beta_2 + Z_{2i}^T(t_i)b_{2i} + U_{2i}^T\lambda_2\right\}}, \tag{141}$$

for $g = 1, ..., G-1$, ((38),(39),(74),(65)).

β_2, λ_2 and b_{2i} are the vectors of p_2 time-dependent fixed effects, r_2 time-independent fixed effects, and q_2 random effects for the i^{th} individual with the corresponding design matrices $X_{2i}^T(t_i)$, U_{2i}^T, and $Z_{2i}^T(t_i)$, respectively. The matrix U_{2i}^T has m_{2i} identical rows u_{2i}^T. The random effects b_{2i} have multivariate normal distribution, so that

$$b_{2i} \sim MVN(0, D_2), \tag{142}$$

with a $q_2 \times q_2$ variance-covariance matrix D_2. All covariates here are denoted by $X_{2i}^T(t_i)$, U_{2i}^T, and $Z_{2i}^T(t_i)$ because they may or may not differ than their corresponding $X_{1i}^T(t_i)$, U_{1i}^T, and $Z_{1i}^T(t_i)$.

5.2.3 Cause-Specific Survival Sub-Model

For a specific survival component, let A_i, T_i^* and C_i denote the left-truncation time, true survival time and censoring time respectively, for the i^{th} individual, $i = 1, ..., n$. The observed survival time is given by $T_i = \min(T_i^*, C_i)$. Denote by $\delta_i = I(T_i^* \leq C_i)$ the censoring indicator (1-dead, 0-alive). Let M_i^T be the vector of the s time-independent survival covariates, related only to survival.
Let h(t) be an arbitrary function of the survival time. Two important cases are:

$$h(t) = \begin{cases} \log(t), \text{for an accelerated failure time model} \\ t, \text{for the other parametric models.} \end{cases} \tag{143}$$

For the i^{th} individual the survival is modelled as

$$h(t_i) = \mu_i(t_i) + \sigma \times \epsilon_i, \tag{144}$$

where σ is the scale parameter and $\epsilon_i \sim G(t)$ is an error term. The mean survival $\mu_i(t)$ is given by

$$\mu_i(t) = \alpha_1 W_{1i}(t) + M_i^T \gamma, \tag{145}$$

where α is the association between the longitudinal and the survival sub-models, γ is the s-vector of the regression coefficients, and M_i^T is the design matrix of s time-independent survival covariates. When $\alpha = 0$, there is no relation between the longitudinal and the survival processes.

Denote by $x_{1i}^T(t_{t_ij})$ and $z_{1i}^T(t_{ij})$ the rows of the design matrices $X_1^T(t_i)$ and $Z_1^T(t_i)$ from the equation (136). These vector functions are observed at particular time points $t_{ij},\ j = 1, \cdots, m_i$ but we extend this notation to an arbitrary time point t, to include the p-vector $x_{1i}^T(t)$, the q-vector $z_{1i}^T(t)$ and the longitudinal biomarker $Y_i(t)$.
Then, the joint model survival $h_i(t)$ for the i^{th} individual at an arbitrary (single) time-point t is given by

$$h_i(t) = \alpha[x_{1i}^T(t)\beta_1 + z_{1i}^T(t)b_{1i} + u_{1i}^T\lambda_1] + M_i^T\gamma + \sigma \times \epsilon_i. \tag{146}$$

5.2.4 Joint Likelihood for Competing Risks

The joint likelihood for competing risks is given by:

$$L(Y_i, T_i, A_i, \delta_i; \theta) = \int L(Y_i|b_i; (\theta_y, \sigma_e^2))L(T_i, A_i, \delta_i|b_i; (\theta_t, \sigma^2))\phi_q(b_i; D)db_i, \tag{147}$$

where

$$L(T_i, A_i, \delta_i|b_i; (\theta_t, \sigma^2)) = \prod_{k=1}^{g} \left\{ f(t)\pi_k(X_{2i}, W_{ki}; a))^{\delta_i} (S(t\pi_k(X_{2i}, W_{ki}; a))^{1-\delta_i} \right\}, \tag{148}$$

((38),(39),(65),(74)), where $f(t)$ and $S(t)$ are the appropriate probability density and survival functions for the survival distribution under right-censoring or left-truncation and

right-censoring.

This likelihood also includes integral which is difficult to solve analytically, and thus an integral approximation is needed. Gauss-Hermite or Laplace approximation can be applied and a maximum likelihood can be used to estimate the parameters.

A program to implement this methodology is yet to be written.

5.3 Summary

In this chapter the future extensions of the shared parameter model developed in this thesis are presented. These are a shared parameter model with multiple biomarkers and competing risks, under left-truncation and right-censoring. These two, multiple biomarkers and competing risks, are more complicated versions of the simple joint model presented in this thesis. First of all, specific likelihoods for the appropriate survival distribution need to be constructed. These likelihoods include integrals which cannot be solved analytically, thus Gauss-Hermite or Laplace approximations can be applied. An R-program will be developed for these two methods, multiple biomarkers and competing risks, and will be tested using simulations.

It needs to be noted that there is no known methods for assessing the fit of a parametric shared parameter model (even though there is some work for a shared parameter model using a Cox proportional hazards survival sub-model), and this is also an issue that needs to be investigated. First thoughts on this matter are to use the plot of a cumulative hazard function of the Cox-Snell residuals against the Cox-Snell for assessing the overall fit of the model, plot the scaled score residuals to identify how influential an observation is and

plot the Cook's distance to identify influential subjects in respect to the estimation of the parameters. These were discussed on Chapter 2 for assessing a parametric survival model's fit, but may or may not be suitable for a parametric shared parameter model.

References

[1] Aalen O. O., Borgan O. and Gjessing H. K., *Survival and Event History Analysis: A process point of view*, Springer, 2008

[2] Aitchison J., Brown J. A. C., *The lognormal distribution: with special reference to its uses in economics*, Cambridge Univ. Press, 1963

[3] Aitken M., Anderson D., Francis B. and Hinde J., *Statistical modelling in GLIM*, Oxford University Oxford University press, 1989, pg. 283-285

[4] Albert P. S. and Shih J. H., *On Estimating the Relationship between Longitudinal Measurements and Time-to-Event Data Using a Simple Two-Stage Procedure*, Biometrics. 2010 Sep;66(3):983-7

[5] Albert P. S. and Shih J. H., *An Approach For Jointly Modeling Multivariate Longitudinal Measurements and Discrete Time-To-Event Data*, Ann Appl Stat. 2010 September 1; 4(3): 1517-532

[6] Allison P. D., *Survival Analysis Using SAS: A Practical Guide, Second Edition*, SAS Institute

[7] Andriole GL, Crawford ED, Grubb RL 3rd, Buys SS, Chia D, Church TR, Fouad MN, Gelmann EP, Kvale PA, Reding DJ, Weissfeld JL, Yokochi LA, O'Brien B, Clapp JD, Rathmell JM, Riley TL, Hayes RB, Kramer BS, Izmirlian G, Miller AB, Pinsky PF, Prorok PC, Gohagan JK, Berg CD; PLCO Project Team, *Mortality results from a randomized prostate-cancer screening trial*, N Engl J Med. 2009 Mar 26;360(13):1310-9

[8] Arndt C., *Information Measures: Information and Its Description in Science and Engineering*, Springer-Verlag Berlin Heidelberg, 2004

[9] Balakrishman N., *Methods and Applications of Statistics in the Life and Health Sciences*, John Wiley and Sons, 2010

[10] Beard R. E., *Note on Some Mathematical Mortality Models*, In G. E. W. Wolstenholme and M. O' Connor (Eds.), The lifespan of animals, pp. 302-311. Boston, MA: Little Brown (1959).

[11] Bettencourt-Silva, J, De La Iglesia, B, Donell, S and Rayward-Smith, V (2011) *On creating a patient-centric database from multiple Hospital Information Systems in a National Health Service secondary care setting.* Methods of Information in Medicine. pp. 6730-6737

[12] Bianconcini, S., Cagnone S., and Rizopoulos D., *Approximate likelihood inference in generalized linear latent variable models based on integral dimension reduction.* arXiv preprint arXiv:1503.01249 (2015)

[13] Box-Steffensmeier J. M. and Jones B. S., *Event History Modeling: a guide for social scientists*, Cambridge University Press, 2004

[14] Breslow N. E., *Covariance Analysis of Censored Survival Data*, Biometrics (1974), 30: 89-100

[15] Brostrm, G. (2013). *eha: Event History Analysis.* R package version 2.3-1.

[16] Brown E. R. and Ibrahim J. G., *A Bayesian Semiparametric Joint Hierarchical Model for Longitudinal and Survival Data*, Biometrics 59, 221-228, June 2003

[17] Brown E. R., Ibrahim J. G. and DeGruttola V., *A flexible B-spline Model for Multiple Longitudinal Biomarkers and Survival*, Biometrics 61, 64-73, March 2005

[18] Carlsson S., Assel M., Sjoberg D., Ulmert D., Hugosson J., Lilja H. and Vickers A.,*Influence of blood prostate specific antigen levels at age* 60 *on benefits and harms of prostate cancer screening: population based cohort study*, BMJ, 28 March 2014

[19] Carpenter J. and Bithell J., *Bootstrap Confidence Intervals: when, which, what? A practical guide for Medical Statisticians*, Statistics in Medicine, 2000; 19:1141-1164

[20] Collett D., *Modelling Survival Data in Medical Research*, Boca Raton, FL: CRC Press, 2003

[21] Collett D., *Modelling Survival Data in Medical Research*, Taylor and Francis e-Library, 2009

[22] Conn P. M., *Handbook of Models for Human Aging*, Elsevier Academic Press, 2006

[23] Cook R. D. and Weisberg S., *Applied Regression Including Computing and Graphics*, John Wiley and Sons, 1999

[24] Cox D. R., *Regression Models and Life Tables (with discussion)*, Journal of Royal Statistical Society: Series B (1972), 34: 187-220

[25] Cox, D. and Oakes, D. *Analysis of Survival Data*. Chapman Hall, London, 1984

[26] Cramer S. D., Alcamo I. E., Heymann D. L. *Prostate Cancer*, Infobase publishing, 2007

[27] Crowther, M. J. (2012). *Stjm11: Stata module to fit shared parameter joint models of longitudinal and survival data.* Statistical Software Components.

[28] Dafni U. G. and Tsiatis A. A., *A Method for Evaluating Surrogate Markers When Measured with Error Using the Cox Model*, Drug Information Journal (July 1994) 28: 667-690

[29] Dempster, A.P., Laird, N.M. Rublin, D.B.,*Maximum likelihood from incomplete data via the EM algorithm.* J. R. Stat. Soc. B Methodol. (1977), 39(1), 138.

[30] DeGruttola V. and Tu X. M., *Modelling Progression of CD4-Lymphocyte Count and its Relationship to Survival Time*, Biometrics, Vol. 50, No. 4 (Dec., 1994), pp. 1003-1014

[31] Demidenko E., *Mixed models: theory and applications*, John Wiley and Sons, 2004

[32] Denuit M., Marchal X., Pitrebois S. and Walhin J., *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems*, John Wiley and Sons, 2007

[33] Diggle P. and Kenward M. G., *Informative Drop-Out in Longitudinal Data Analysis*, Applied Statistics, Volume 43, Issue 1(1994), 49-93

[34] Efron B. and Tibshirani R., *Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy*, Statistical Science, 1986

[35] Efron B., *The Efficiency of Cox's Likelihood Function for Censored Data*, Journal of American Statistical Association (1977), 72: 557-565

[36] Efron B. and Tibshirani R. J., *An introduction to the bootstrap* Chapman & Hall, 1993

[37] E. A., Agalliu I., Thurston S. W., Coull B. A., Checkoway H., *Smoothing in occupational cohort studies: an illustration based on penalised splines*, Occup. Environ. Med. 2004; 61:854-860

[38] Elashoff R. M., Li G. and Li N., *An approach to joint analysis of longitudinal measurements and competing risks failure time data*, Statistics in Medicine, 2007; 26:28132835

[39] Elashoff R. M., Li G. and Li N., *A Joint Models for Longitudinal Measurements and Survival Data in the Presence of Multiple Failure Types*, Biometrics, 2008 September

[40] Evans M., Hastings N. and Peacock B., *Statistical Distributions*, Wiley and Sons, 2000

[41] Everitt B. and Pickles A., *Statistical aspects of the design and analysis of clinical trials*, Imperial College Press, 1999

[42] Faucett C. L. and Thomas D. C., *Simultaneously Modelling Censored Survival Data and Repeatedly Measured Covariates: A Gibbs Sampling Approach*, Statistics in Medicine, Vol. 15, 1663-1685(1996)

[43] Faucett, C.L., Schenker, N. Taylor, J.M.G. *Survival analysis using auxiliary variables via multiple imputation, with application to AIDS clinical trial data.* Biometrics, 58(1), 3747, 2002

[44] Flocks RH, Urich VC, Patel CA, Opitz JM., *Studies on the antigenic properties of prostatic tissue.* I. J Urol 1960; 84: 13443

[45] Follmann D. and Wu M., *An Approximate Generalized Linear Model with Random Effects for Informative Missing Data*, Biometrics, Vol. 51, No. 1 (Mar., 1995), pp. 151-168

[46] Fox J., *Applied Regression Analysis and Generalized Linear Models*, Second Edition, SAGE Publications, Inc., 2008

[47] German R. M. and Park S. J., *Handbook of Mathematical Relations in Particulate Materials Processing: ceramics, powder metals, cermets, carbides, hard materials, and minerals*, John Wiley and Sons, 2008

[48] Gompertz B., *The nature of the function expressive of the law of human mortality*, Philosophical Transactions of the Royal Society 115 (1825), 513-855.

[49] Gonen M. and Heller G., *Concordance probability and discriminatory power in propotional hazards regression*, Biometrika, 2005, 95, 4, pp. 965-970

[50] Gray R., *Advanced Statistical Computing*, Harvard Course Notes BIO 248 cd, 2003

[51] Guo, X. Carlin, B.P. *Separate and joint modeling of longitudinal and event time data using standard computer packages.* Am. Stat., 58(1), 1624, 2004

[52] Hans C., *Modeling the redundant signals effect by specifying the hazard function*, Springer New York, 1988

[53] Hara M, Inoue T, Koyanagi Y et al., *Preparation and immunoelectrophoretic assessment of antisera to human seminal plasma.* Nippon Hoigaku Zasshi 1966; 20:356

[54] Hara M, Koyanagi Y, Inoue T et al. *Some physico-chemical characteristics of "γ-seminoprotein", an antigenic component specific for human seminal plasma. Forensic immunological study of body fluids and secretion. VII.* Nihon Hoigaku Zasshi 1971;25: 3224

[55] Harris J. and Stocker H., *Handbook of mathematics and computational science*, Springer-Verlag New York, 1998

[56] Henderson R., Diggle P. and Dobson A., *Joint Modelling of Longitudinal measurements and event time data*, Biostatistics (2000), 1, 4, pp. 465-480

[57] Hernandez J, Thompson IM., *Prostate-specific antigen: a review of the validation of the most commonly used cancer biomarker.* Cancer 2004;101:894-904.

[58] Hirschfield, G. M. and Gershwin, M. E. (2013). *The immunobiology and pathophysiology of primary biliary cirrhosis.* Annual review of pathology: Mechanisms of disease, 8:303330.

[59] Hogan J. W. and Laird N. M., *Mixture Models for the Joint Distribution of Repeated Measures and Event Times*, Statistics in Medicine, Vol. 16, 239-257(1997)

[60] Hogan J. W. and Laird N. M., *Model-Based Approaches to Analysing Incomplete Longitudinal and Failure Time Data*, Statistics in Medicine, Vol. 16, 259-272(1997)

[61] Holmstrom B, Johansson M, Bergh A, Stenman UH, Hallmans G, Stattin P. *Prostate specific antigen for early detection of prostate cancer: longitudinal study.* BMJ 2009;339:b3537

[62] Hosmer D. W., Lemeshow S. and May S., *Appied Survival Analysis: Regression Modeling of Time to Event Data*, John Wiley and Sons, 1999

[63] Hosmer D. W., Lemeshow S. and May S., *Applied Survival Analysis: Regression Modeling of Time-to-Event Data*, John Wiley and Sons, Second Edition, 2008

[64] Hsieh F., Tseng Y. and Wang J., *Joint Modeling of Survival and Longitudinal Data: Likelihood Approach Revisited*, Biometrics, 2006 Dec;62(4):1037-43.

[65] Huang X., Gang L. and Elashoff R. M., *A joint model of longitudinal and competing risks survival data with heterogeneous random effects and outlying longitudinal measurements*, Stat Interface. 2010 ; 3(2): 185-95

[66] Huang X., Li G., Elashoff R. M. and Pan J., *A general joint model for longitudinal measurements and competing risks survival data with heterogeneous random effects*, Lifetime Data Anal (2011) 17:80100

[67] Jamnicky L., Nam R., The Canadian Guide to Prostate Cancer, Second Edition, SCRIPT Medical Press Inc., 2013

[68] Kalbfleisch J. and Prentice R., *The Statistical Analysis of Failure Time Data*, New York: John Wiley and Sons, 1980

[69] Kaplan E. and Meier P., *Nonparametric estimation from incomplete observations*, Journal of the American Statistical Association. 53 (282): 457-481, 1958

[70] Kent J. T. and O Quigley J., *Measures of dependence for censored survival data*, Biometrika, 1988, 75, 3, pp 525-34

[71] Klein P. J. and Moeschberger L. M., *Survival Analysis- Techniques for Censored and Truncated Data*, Springer-Verlag, New York, 1997

[72] Laird N. M. and Ware J. H., *Random-Effects for Longitudinal Data*, Biometrics, Vol. 38, No. 4 (Dec., 1982), pp. 963-974

[73] Lee E. T. and Wang J. W., *Statistical methods for survival data analysis*, Wiley and Sons, 2003

[74] Li N., Elashoff R. M. and Li G., *Robust Joint Modeling of Longitudinal Measurements and Competing Risks Failure Time Data*, Biometrical Journal 51 (2009) 1, 19-0

[75] Li N., Elashoff R. M., Li G. and Tseng C., *Joint analysis of Bivariate Longitudinal Ordinal Outcomes and Competing Risks Survival Times with Nonparametric Distributions for Random Effects*, John Wiley and Sons Ltd, 2012

[76] Lin H., McCulloch C. E., Turnbull B. W., Slate E. H. and Clark L. C., *A Latent Class Mixed Model for Analysing Biomarker Trajectories with Irregularly Scheduled Observation*, Statistics in Medicine 2000; 19: 1303-1318

[77] Lin H., McCulloch E. and Rosenheck R. A., *Latent Pattern Mixture Models for Informative Intermittent Missing Data in Longitudinal Studies*, Biometrics 60, 295-305, June 2004

[78] Lindsey J. K., *Parametric Statistical Inference*, Oxford University Press (1996)

[79] Little R. J. A., *Modelling the Drop-Out Mechanism in Repeated-Measures Studies*, Journal of the American Statistical Association, Vol. 90, No. 431 (Sep. 1995), pp. 1112-1121

[80] Little R. J. A. and Rubin D. B., *Statistical Analysis with Missing Data*, New York: Wiley (1987)

[81] Liu X., *Survival Analysis: Models and Applications*, John Wiley and Sons (2012)

[82] Liu Q. and Pierce D. A., *A note on Gauss-Hermite quadrature*, Biometrika (1994), 81, 3, pp. 624-9

[83] Maindonald J. H. and Braun J., *Data analysis and graphics using R: an example-based approach*, Cambridge University Press (2003)

[84] Makeham W. M., *On the law of mortality and the construction of annuity tables*, Journal of the Institute of Actuaries 8 (1859), 301-310.

[85] Marshall A. W. and Olkin I., *Life distributions: structure of nonparametric, semiparametric and parametric families*, Springer (2007)

[86] Masoro E. J. and Austad S., *Handbook of the Biology of Aging*, Elsevier Academic Press, 6^{th} edition (2006)

[87] McCool J., *Using the Weibull Distribution: reliability, modeling, and inference*, John Wiley and Sons (2012)

[88] Lisa M. McCrink, Adele H. Marshall and Karen J. Cairns, *Advances in Joint Modelling: A Review of Recent Developments with Application to the Survival of End Stage Renal Disease Patients*, International Statistical Review (2013), 81, 2, 249269

[89] Mills M., *Introducing Survival and Event History Analysis*, Sage Publications (2011)

[90] Klein P. J. and Moeschberger L. M., *Survival Analysis- Techniques for Censored and Truncated Data*, Springer-Verlag, New York, 1997

[91] Murtaugh, P. A., Dickson, E. R., Van Dam, G. M., Malinchoc, M., Grambsch, P. M., Langworthy, A. L., and Gips, C. H. (1994).*Primary biliary cirrhosis: Prediction of short-term survival based on repeated patient visits.* Hepatology, 20(1):126134.

[92] Murthy D. N. P., Xie M. and Jiang R., *Weibull Models*, Wiley and Sons (2004)

[93] Nelson W., *Applied Life Data Analysis*, Wiley and Sons (2004)

[94] Pantazis N. and Touloumi G., *Robustness of a parametric model for informatively censored bivariate longitudinal data under misspecification of its distributional assumptions: A simulation study*, Statistics In Medicine (2007); 26:5473-485

[95] Pantazis N., Touloumi G., Walker A.S. and Babiker A.G., *Bivariate modelling of longitudinal measurements of two human immunodeficiency type 1 disease progression markers in the presence of informative drop-outs*, Appl. Statist. (2005), 54, Part 2, pp. 405-23

[96] Papsidero LD, Wang MC, Valenzuela LA, Murphy GP, Chu TM. *A prostate antigen in sera of prostatic cancer patients.* Cancer Res (1980); 40: 242832

[97] Pericleous P., *Parametric Survival Models for Actuarial Applications*, MSc Thesis UEA , Unpublished, 2012

[98] Perks W., *On Some Experiments in the Graduation of Mortality Statistics*, Journal of the Institute of Actuaries (1886-1994) Vol. 63, No. 1 (March 1932), pp. 12-57

[99] Philipson, P., Diggle, P., Sousa, I., Kolamunnage-Dona, R., Williamson, P., and Henderson, R. (2012). *joiner: Joint modelling of repeated measurements and time-to-event data.*

[100] Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team (2013). *nlme: Linear and Nonlinear Mixed Effects Models.* R package version 3.1-113.

[101] Prentice R., *Covariate measurement errors and parameter estimates in a failure time regression model*, Biometrika (1982) 69,331-42,

[102] Proust-Lima C., Joly P., Dartigues J. and Jacqmin-Gadda H., *Joint modelling of multivariate longitudinal outcomes and a time-to-event: A nonlinear latent class approach*, Computational Statistics and Data Analysis 53 (2009) 1142-1154

[103] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna Austria, 2014

[104] Rao, A. R., Motiwala, H. G. and Karim, O. M.A., *The discovery of prostate-specific antigen.* BJU International (2008), 101: 510

[105] Ratcliffe, S. J., Guo, W., and Ten Have, T. R. (2004). *Joint modeling of longitudinal and survival data via a common frailty.* Biometrics, 60(4):892899.

[106] Richards S. J., *Applying Survival Models to Pensioner Mortality Data*, Institute of Actuaries, February 2008

[107] Richards S. J., *A handbook of parametric survival models for actuarial use*, Scandinavian Actuarial Journal, May 2011

[108] Rizopoulos, D., Verbeke, G. and Molenberghs, G. *Shared parameter models under random effects misspecification.*, Biometrika (2008), 95(1), 6374.

[109] Rizopoulos D., Verbeke G. and Lesaffre E., *Fully exponential Laplace approximations for the joint modelling of survival and longitudinal data*, J. R. Statist. Soc. B (2009), 71, Part 3, pp. 637-54

[110] Rizopoulos, D. et al. (2010). *JM: An R package for the joint modelling of longitudinal and time-to-event data.* Journal of Statistical Software, 35(9):1-33.

[111] Rizopoulos D., *Dynamic Predictions and Prospective Accuracy in Joint Models for Longitudinal and Time-To-Event Data*, Biometrics, 67, 819-829, September 2011

[112] Rizopoulos D. and Ghosh P., *A Bayesian Semiparametric Multivariate Joint Model for Multiple Longitudinal Outcomes and a Time-To-Event*, Statistics in Medicine, February 2011

[113] Rizopoulos D., *Fast Fitting of Joint Models for Longitudinal and Event Time Data Using a Pseudo-Adaptive Gaussian Quadrature rule*, Computational Statistics and Data Analysis, 56(2012), 491-501

[114] Rizopoulos D. and Ghosh P., *A Bayesian Semiparametric Multivariate Joint Model for Multiple Longitudinal Outcomes and a Time-To-Event*, Statistics in Medicine, February 2011

[115] Sandblom G, Varenhorst E, Rosell J, Lfman O, Carlsson P., *Randomised prostate cancer screening trial: 20 year follow-up*, BMJ. 2011 Mar 31;342:d1539

[116] Schrder F, Kattan MW. *The comparability of models for predicting the risk of a positive prostate biopsy with prostate-specific antigen alone: a systematic review.* Eur Urol 2008;54:274-90.

[117] Schrder FH, Hugosson J, Roobol MJ, Tammela TL, Ciatto S, Nelen V et al. *Screening and prostate-cancer mortality in a randomized European study.* N Engl J Med 2009;360:1320-8.

[118] Self, S. Pawitan, Y. *Modeling a marker of disease progression and onset of disease. In AIDS Epidemiology: Methodological Issues*, Eds. Jewell, N.P., Dietz, K. Farewell, V.T. Boston: Birkhuser, 1992

[119] Sensabaugh GF. *Isolation and characterization of a semen-specific protein from human seminal plasma: a potential new marker for semen identification.* J Forensic Sci 1978; 23:10615

[120] Schluchter,M.D. *Methods for the analysis of informatively censored longitudinal data.* Stat. Med., 11(1415), 18611870, 1992

[121] Singer J. D. and Willett J. B., *Applied Longitudinal Data Analysis: modeling change and event occurrence*, Oxford University Press, 2003

[122] Sokoll LJ, Chan DW. *Prostate-specific antigen. Its discovery and biochemical characteristics.* Urol Clin North Am 24: 2539, 1997

[123] Song Z., Chen Y., Sastry C. R., Tas N. C., *Optimal Observation for Cyber-physical Systems: A Fisher-information-matrix-based Approach*, Springer-Verlag London Limited, 2009

[124] Song X., Davidian M. and Tsiatis A. A., *A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data*, Biometrics, Volume 58, Issue 4, pages 742-53, December 2002

[125] Soloviev M., Andrn P., Shaw C., *Peptidomics: Methods and Applications*, John Wiley Sons, 21 Dec 2007

[126] Sousa I., *A Reviw On Joint Modelling Of Longitudinal Measurements And Time-To-Event*, REVSTAT Statistical Journal Volume 9, Number 1, 5781, March 2011

[127] Stepanova M. and Thomas L., *Survival Analysis Methods For Personal Loan Data*, Operations Research, Vol. 50, No. 2, pp. 277-28, (Mar. - Apr., 2002)

[128] Yu-Ru Su and Jane-Ling Wang *Modeling left-truncated and right-censored survival data with longitudinal covariates* The Annals of Statistics, Volume 40, Number 3, 1465-1488, 2012

[129] Sweeting M. J. and Thompson S. G., *Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture*, Biometrical Journal 53, 5, 750-63, 2011

[130] Therneau, T. M. (2013). *A Package for Survival Analysis in S.* R package version 2.37-4.

[131] Touloumi G., Pocock S. J., Babiker A. G. and Darbyshire J. H., *Estimation and Comparison of Rates of Change in Longitudinal Studies with Informative Drop-Outs*, Statistics in Medicine, 18, 1215-1233, 1999

[132] Tseng Y., Hsieh F. and Wang J., *Joint Modelling of Accelereated Failure Time and Longitudinal Data*, Biometrika, 92, 3, pp. 587-603, 2005

[133] Tsiatis, A.A. Davidian,M. *A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error.* Biometrika, 88(2), 447458, 2001

[134] Tsiatis A. A and Davidian M., *Joint Modeling Of Longitudinal and Time-To-Event Data: An Overview*, Statistica Sinica 14, 809-834, 2004

[135] Tsiatis A. A., DeGruttola V. and Wulfsohn M. S., *Modeling the Relationship of Survival to Longitudinal Data Measured with Error. Applications to Survival and CD4 Counts in Patients with AIDS*, Journal of the American Statistical Association, Vol. 90, No. 429, pp. 27-37, Mar., 1995

[136] Tsonaka R., Verbeke G. and Lesaffre E., *A Semi-Parametric Shared Parameter Model to Handle Nonmonotone Nonignorable Missingness*, Biometrics 65, 81-7, March 2009

[137] Tuerlinckx F., Rijmen F., Verbeke G. and De Boeck P., *Statistical Inference in generalized linear mixed models: A review*, British Journal of Mathematical and Statistical Psychology (2006), 59, 225-255

[138] Venables W. N. and Ripley B. D., *Modern Applied Statistics with S*, Springer Science and Business Media, 2002

[139] Verbeke G. and Molenberghs G., *Linear Mixed Models for Longitudinal Data*, Springer Verlag New York, LLC, 2009

[140] Vickers AJ, Cronin AM, Bjork T, Manjer J, Nilsson PM, Dahlin A, et al. *Prostate specific antigen concentration at age 60 and death or metastasis from prostate cancer: case-control study.* BMJ 341:c4521, 2010

[141] Vittinghoff E., Glidden D. V., Shiboski S. C. and McCulloch C. E., *Regression Methods in Biostatistics: Linear, Logistic, Survival and Repeated Measures model*, Springer, 2005

[142] Vonesh E. F., Greene T. and Schluchter M. D., *Shared parameter models for the joint analysis of longitudinal data and event times*, Statistics In Medicine, Statist. Med. 25:143-63, 2006

[143] Wang MC, Valenzuela LA, Murphy GP, Chu TM. *Purification of a human prostate specific antigen 1979.* J Urol 167:9604, 2002

[144] Wang Y. and Taylor J. M. G., *Jointly Modeling Longitudinal and Event Time Data with application to Acquired Immunodeficiency Syndrome*, Journal of the American Statistical Association, Vol. 96, No. 455, pp. 895-905, Sep., 2001

[145] Weerahandi S., *Exact statistical methods for data analysis*, Springer-Verlag New York, 1995

[146] Weibull W. and Stockholm S., *A Statistical Distribution Function of Wide Applicability*, J. Appl. Mech. 18:293-7, 1951

[147] Welch HG, Albertsen PC. *Prostate cancer diagnosis and treatment after the introduction of prostate-specific antigen screening: 1986-2005.* J Natl Cancer Inst 101:1325-9, 2009

[148] Willekens F., *Forecasting Mortality in Developed Countries, Insights from a Statistical, Demographic and Epidemiological Perspective, Gompertz in Context: the Gompertz and Related Distributions,* pg. 105-123, Springer Netherlands, 2001

[149] Wu, M.C. and Bailey, K., *Estimation and comparison of changes in the presence of informative right censoring:conditional linear model*, Biometrics, 45, 939955, 1989

[150] Wu M. C. and Carroll R. J., *Estimation and Comparison of Changes in the Presence of Informative Right-Censoring by Modeling the Censoring Process*, Biometrics, Vol. 44, No. 1, pp. 175-188, Mar., 1988

[151] Wulfsohn M. S. and Tsiatis A. A., *A Joint Model for Survival and Longitudinal Data Measured with Error*, Biometrics, Vol. 53, No. 1, pp. 330-339, Mar., 1997

[152] Xu, J. and Zeger, S.L. *The evaluation of multiple surrogate endpoints*. Biometrics, 57(1), 8187., 2001

[153] Zill D. G. and Cullen M. R., *Advanced Engineering Mathematics*, Jones and Bartlett Publishers, 2006

A Proof of fundamental survival analysis relations

The cumulative distribution function is given by

$$F(t) = P(T \leq t), \tag{A.1}$$

and the survival function is

$$S(t) = P(T > t). \tag{A.2}$$

From (A.1) and (A.2):

$$S(t) = 1 - F(t). \tag{A.3}$$

The probability density function is related to the probability distribution by

$$f(t) = \frac{d}{dt}(F(t)). \tag{A.4}$$

Differentiating the survival function,

$$S^{'}(t) = \frac{d}{dt}(S(t)) = \frac{d}{dt}(1 - F(t)) = -f(t). \tag{A.5}$$

The hazard function is given by:

$$h(t) = \lim_{\Delta t \to 0} \frac{P[t \leq T < t + \Delta t | T \geq t]}{\Delta t}. \tag{A.6}$$

$P[t \leq T < t + \Delta t | T \geq t]$ is the probability that t is in the interval (t+Δt)$i.e.$
P[t $\leq T < t + \Delta t | T \geq t] = f(t)$, while $P[T \geq t]$ is the survival function (A.2). Thus,

$$h(t) = \frac{f(t)}{S(t)}. \tag{A.7}$$

The cumulative hazard function is the sum of all individual hazards over time. It can be expressed as an area under the curve $h(u)$. So,

$$\begin{aligned} H(t) &= \int_0^t h(u)du \\ &= \int_0^t \overbrace{\frac{f(u)}{S(u)}}^{A.7} du \\ &= \int_0^t \overbrace{\frac{-S^{'}(u)}{S(u)}}^{A.5} du \\ &= \int_0^t \overbrace{\frac{d}{du}(-\log(S(u)))}^{\frac{d}{du}(\log(S(u)))=\frac{S^{'}(u)}{S(u)}} \\ &= -\log(S(t)) \end{aligned} \tag{A.8}$$

B Transformations from distributions to models

B.1 Transformation from Log-Logistic distribution to Log-Logistic Model

The survival function for the Log-Logistic distribution is given by:

$$S(t;\lambda,p) = (1+\lambda t^p)^{-1}. \tag{B.1}$$

Taking the log transformation of time ($y = \log t$) and redefining the parameters $p = \sigma^{-1}$ and $\lambda = \exp\{-\sigma^{-1}\mu\}$:

$$S(y;\lambda,p) = (1+\lambda \exp\{py\})^{-1} \tag{B.2}$$

and Y follows a linear model:

$$Y = \log T = \mu + \sigma W, \tag{B.3}$$

where W is the random variable from the standard logistic distribution with probability density function:

$$f_W(w) = \exp\{w\}\,\sigma^{-1}(1+\exp\{w\})^{-2} \tag{B.4}$$

and survival function:

$$S_W(w) = (1+\exp\{w\})^{-1}. \tag{B.5}$$

Thus, the probability density function of Y for the i^{th} individual is:

$$f(y_i;\mu_i,\sigma) = \exp\left\{\sigma^{-1}(y_i-\mu_i)\right\}\sigma^{-1}(1+\exp\left\{\sigma^{-1}(y_i-\mu_i)\right\})^{-2} \tag{B.6}$$

and the survival function for the i^{th} individual is:

$$S(y_i;\mu_i,\sigma) = [1+\exp\left\{\sigma^{-1}(y_i-\mu_i)\right\}]^{-1}. \tag{B.7}$$

C Application of Joint Modelling with Log-Logistic Sub-model

The longitudinal component is modelled with a linear mixed effects model and the survival component is modelled using Log-Logistic distribution.

The log survival time for the i^{th} individual, following Log-Logistic distribution is given by

$$\log T_i(t_i) = \mu_i(t_i) + \sigma \times \epsilon_i, \tag{C.1}$$

where ϵ_i comes from Logistic distribution, σ is the scale and $\mu_i(t_i)$ is given by

$$\mu_i(t) = \alpha W_i(t) + M_i^T \gamma, \tag{C.2}$$

where γ is the regression coefficient that links the time-independent covariates, M_i^T, related only to survival and $W_i(t)$ is the longitudinal component.

The probability density and survival function of Log-Logistic model are given by (B.6) and (B.7) respectively. For the longitudinal parts, the probability density functions of $Y_i(t_{ij})$ conditional on random effects b_i and of random effects are given by (111) and (112) respectively.

C.1 Likelihood for Application of Joint Modelling with Log-Logistic Sub-Model

C.1.1 Right-Censoring Likelihood for Application of Joint Modelling with Log-Logistic Sub-Model

The Log-Logistic joint likelihood contribution for right-censoring only is

$$\begin{aligned} L(Y_i, T_i, \delta_i; \theta) = \; & \prod_{i=1}^{n} [\int_{-\infty}^{\infty} \prod_{j=1}^{m_i} [(2\pi\sigma_e^2)^{-\frac{1}{2}} \exp[-(Y_i(t_{ij}) - W_i(t_{ij}))^2 (2\sigma_e^2)^{-1}]] \\ & \times (2\pi|D|)^{-\frac{1}{2}} \exp(-2^{-1} b_i' D^{-1} b_i) \\ & \times (\exp\{\sigma^{-1}(\log T_i(t_i) - \mu_i(t_i))\} \sigma^{-1} (1 + \exp\{\sigma^{-1}(\log T_i(t_i) - \mu_i(t_i))\})^{-2})^{\delta_i} \\ & \times ([1 + \exp\{\sigma^{-1}(\log T_i(t_i) - \mu_i(t_i))\}]^{-1})^{1-\delta_i} \, db_i \,]. \end{aligned} \tag{C.3}$$

When Gauss-Hermite approximation is applied the likelihood becomes

$$\begin{aligned} \widetilde{L}(Y_i, T_i, \delta_i; \theta) = \; & 2^{\frac{qn}{2}} |B|^{-n} \prod_{i=1}^{n} \sum_{l_i=1}^{p} \omega_{l_i} \exp\{E_i' E_i\} \prod_{j=1}^{m_i} [(2\pi\sigma_e^2)^{-\frac{1}{2}} \\ & \times \exp[-(Y_i(t_{ij}) - \hat{W}_i(t_{ij}))^2 (2\sigma_e^2)^{-1}]] \\ & \times (2\pi|D|)^{-\frac{1}{2}} \exp(-2^{-1/2} B^{-1} E_i D^{-1} 2^{1/2} B^{-1} E_i) \\ & \times (\exp\{\sigma^{-1}(\log T_i(t_i) - \hat{\mu}_i(t_i))\} \sigma^{-1} (1 + \exp\{\sigma^{-1}(\log T_i(t_i) - \hat{\mu}_i(t_i))\})^{-2})^{\delta_i} \\ & \times ([1 + \exp\{\sigma^{-1}(\log T_i(t_i) - \hat{\mu}_i(t_i))\}]^{-1})^{1-\delta_i}, \end{aligned} \tag{C.4}$$

where $\widehat{W}_i(t_{ij}) = X_i^T(t_{ij})\beta + Z_i^T(t_{ij})(2^{\frac{1}{2}} B^{-1} E_{il_i}) + U_i^T \lambda$ and $\hat{\mu}_i(t_i) = \alpha \hat{W}_i(t) + M_i^T \gamma$. When Laplace approximation is applied the likelihood becomes

$$\begin{aligned} \widetilde{L}(Y_i, T_i, \delta_i; \theta) = \; & (2\pi)^{\frac{nq}{2}} \hat{\Omega}^{-\frac{n}{2}} \prod_{i=1}^{n} \prod_{j=1}^{m_i} [(2\pi\sigma_e^2)^{-\frac{1}{2}} \exp[-(Y_i(t_{ij}) - \hat{W}_i(t_{ij}))^2 (2\sigma_e^2)^{-1}]] \\ & \times (2\pi|D|)^{-\frac{1}{2}} \exp(-2^{-1} \hat{b}_i' D^{-1} \hat{b}_i) \\ & \times (\exp\{\sigma^{-1}(\log T_i(t_i) - \hat{\mu}_i(t_i))\} \sigma^{-1} (1 + \exp\{\sigma^{-1}(\log T_i(t_i) - \hat{\mu}_i(t_i))\})^{-2})^{\delta_i} \\ & \times ([1 + \exp\{\sigma^{-1}(\log T_i(t_i) - \hat{\mu}_i(t_i))\}]^{-1})^{1-\delta_i}, \end{aligned} \tag{C.5}$$

where $\widehat{W}_i(t_{ij}) = X_i^T(t_{ij})\beta + Z_i^T(t_{ij})\hat{b} + U_i^T \lambda$ and $\hat{\mu}_i(t_i) = \alpha \hat{W}_i(t) + M_i^T \gamma$.

C.1.2 Left-Truncation and Right-Censoring Likelihood for Application of Joint Modelling with Log-Logistic Sub-Model

For the case of left-truncation and right-censoring, the likelihood becomes more complicated.

$$\begin{aligned} L(Y_i, T_i, A_i, \delta_i; \theta) = \; & \prod_{i=1}^{n} [\int_{-\infty}^{\infty} \prod_{j=1}^{m_i} [(2\pi\sigma_e^2)^{-\frac{1}{2}} \exp[-(Y_i(t_{ij}) - W_i(t_{ij}))^2 (2\sigma_e^2)^{-1}]] \\ & \times (2\pi|D|)^{-\frac{1}{2}} \exp(-2^{-1} b_i' D^{-1} b_i) \\ & \times ([A_i(\log(T_i(t_i)) + A_i)]^{\sigma^{-1}} [1 + \exp[\tfrac{\log A_i - \mu_i(t_i)}{\sigma}]]^2 \\ & \times [1 + \exp[\tfrac{\log (T_i(t_i) + A_i) - \mu_i(t_i)}{\sigma}]]^{-2})^{\delta_i} \\ & \times [([1 + \exp[\tfrac{\log A_i - \mu_i(t_i)}{\sigma}]][1 + \exp[\tfrac{\log (T_i(t_i) - A_i) - \mu_i(t_i)}{\sigma}]]^{-1}]^{1-\delta_i} \, db_i \;]. \end{aligned} \tag{C.6}$$

When Gauss-Hermite approximation is applied the likelihood becomes

$$\begin{aligned} \widetilde{L}(Y_i, T_i, A_i, \delta_i; \theta) = \; & 2^{\frac{qn}{2}} |B|^{-n} \prod_{i=1}^{n} \sum_{l_i=1}^{p} \omega_{l_i} \exp\{E_i' E_i\} \prod_{j=1}^{m_i} [(2\pi\sigma_e^2)^{-\frac{1}{2}} \\ & \times \exp[-(Y_i(t_{ij}) - \hat{W}_i(t_{ij}))^2 (2\sigma_e^2)^{-1}]] \\ & \times (2\pi|D|)^{-\frac{1}{2}} \exp(-2^{-1/2} B^{-1} E_i D^{-1} 2^{1/2} B^{-1} E_i) \\ & \times ([A_i(\log(T_i(t_i)) + A_i)]^{\sigma^{-1}} [1 + \exp[\tfrac{\log A_i - \hat{\mu}_i(t_i)}{\sigma}]]^2 \\ & \times [1 + \exp[\tfrac{\log (T_i(t_i) + A_i) - \hat{\mu}_i(t_i)}{\sigma}]]^{-2})^{\delta_i} \\ & \times [([1 + \exp[\tfrac{\log A_i - \hat{\mu}_i(t_i)}{\sigma}]][1 + \exp[\tfrac{\log (T_i(t_i) - A_i) - \hat{\mu}_i(t_i)}{\sigma}]]^{-1}]^{1-\delta_i}, \end{aligned} \tag{C.7}$$

where $\widehat{W}_i(t_{ij}) = X_i^T(t_{ij})\beta + Z_i^T(t_{ij})(2^{\frac{1}{2}} B^{-1} E_{il_i}) + U_i^T \lambda$ and $\hat{\mu}_i(t_i) = \alpha \hat{W}_i(t) + M_i^T \gamma$. When Laplace approximation is applied the likelihood becomes

$$\begin{aligned} \widetilde{L}(Y_i, T_i, \delta_i; \theta) = \; & (2\pi)^{\frac{nq}{2}} \hat{\Omega}^{-\frac{n}{2}} \prod_{i=1}^{n} \prod_{j=1}^{m_i} [(2\pi\sigma_e^2)^{-\frac{1}{2}} \exp[-(Y_i(t_{ij}) - \hat{W}_i(t_{ij}))^2 (2\sigma_e^2)^{-1}]] \\ & \times (2\pi|D|)^{-\frac{1}{2}} \exp(-2^{-1} \hat{b}_i' D^{-1} \hat{b}_i) \\ & \times ([A_i(\log(T_i(t_i)) + A_i)]^{\sigma^{-1}} [1 + \exp[\tfrac{\log A_i - \hat{\mu}_i(t_i)}{\sigma}]]^2 \\ & \times [1 + \exp[\tfrac{\log (T_i(t_i) + A_i) - \hat{\mu}_i(t_i)}{\sigma}]]^{-2})^{\delta_i} \\ & \times [([1 + \exp[\tfrac{\log A_i - \hat{\mu}_i(t_i)}{\sigma}]][1 + \exp[\tfrac{\log (T_i(t_i) - A_i) - \hat{\mu}_i(t_i)}{\sigma}]]^{-1}]^{1-\delta_i}, \end{aligned} \tag{C.8}$$

where $\widehat{W}_i(t_{ij}) = X_i^T(t_{ij})\beta + Z_i^T(t_{ij})\hat{b} + U_i^T\lambda$ and $\hat{\mu}_i(t_i) = \alpha\hat{W}_i(t) + M_i^T\gamma$.

D Application of Joint Modelling with Log-Normal Sub-Model

The longitudinal component is modelled with a linear mixed effects model and the survival component is modelled using Log-Normal distribution.

The log survival time for the i^{th} individual, following Log-Normal distribution is given by

$$\log T_i(t_i) = \mu_i(t_i) + \sigma \times \epsilon_i, \tag{D.1}$$

where ϵ_i comes from Normal distribution, σ is the scale and $\mu_i(t_i)$ is given by

$$\mu_i(t) = \alpha W_i(t) + M_i^T\gamma, \tag{D.2}$$

where γ is the regression coefficient that links the time-independent covariates, M_i^T, related only to survival and $W_i(t)$ is the longitudinal component.

The probability density and survival function of Log-Normal model are given in Table (1). For the longitudinal parts, the probability density functions of $Y_i(t_{ij})$ conditional on random effects b_i and of random effects are given by (111) and (112) respectively.

D.1 Likelihood for Application of Joint Modelling with Log-Normal Sub-Model

D.1.1 Right-Censoring Likelihood for Application of Joint Modelling with Log-Normal Sub-Model

The Log-Normal joint likelihood contribution for right-censoring only is

$$\begin{aligned} L(Y_i, T_i, \delta_i; \theta) = \; & \textstyle\prod_{i=1}^{n} [\int_{-\infty}^{\infty} \prod_{j=1}^{m_i} [(2\pi\sigma_e^2)^{-\frac{1}{2}} \exp[-(Y_i(t_{ij}) - W_i(t_{ij}))^2 (2\sigma_e^2)^{-1}]] \\ & \times (2\pi|D|)^{-\frac{1}{2}} \exp(-2^{-1} b_i' D^{-1} b_i) \\ & \times [\exp[\tfrac{-(T_i(t_i)\sigma(2\pi)^{1/2})^{-1}[\log T_i(t_i) - \mu_i(t_i)]^2}{2\sigma^2}]]^{\delta_i} \\ & \times [1 - \Phi(\tfrac{\log(T_i(t_i)) - \mu_i(t_i)}{\sigma})]^{1-\delta_i} \, db_i \;]. \end{aligned} \tag{D.3}$$

When Gauss-Hermite approximation is applied the likelihood becomes

$$\begin{aligned} \widetilde{L}(Y_i, T_i, \delta_i; \theta) = \; & 2^{\frac{qn}{2}} |B|^{-n} \textstyle\prod_{i=1}^{n} \sum_{l_i=1}^{p} \omega_{l_i} \exp\{E_i' E_i\} \prod_{j=1}^{m_i} [(2\pi\sigma_e^2)^{-\frac{1}{2}} \\ & \times \exp[-(Y_i(t_{ij}) - \hat{W}_i(t_{ij}))^2 (2\sigma_e^2)^{-1}]] \\ & \times (2\pi|D|)^{-\frac{1}{2}} \exp(-2^{-1/2} B^{-1} E_i D^{-1} 2^{1/2} B^{-1} E_i) \\ & \times [\exp[\tfrac{-(T_i(t_i)\sigma(2\pi)^{1/2})^{-1}[\log T_i(t_i) - \hat{\mu}_i(t_i)]^2}{2\sigma^2}]]^{\delta_i} \\ & \times [1 - \Phi(\tfrac{\log(T_i(t_i)) - \hat{\mu}_i(t_i)}{\sigma})]^{1-\delta_i}, \end{aligned} \tag{D.4}$$

where $\widehat{W}_i(t_{ij}) = X_i^T(t_{ij})\beta + Z_i^T(t_{ij})(2^{\frac{1}{2}} B^{-1} E_{il_i}) + U_i^T \lambda$ and $\hat{\mu}_i(t_i) = \alpha \hat{W}_i(t) + M_i^T \gamma$. When Laplace approximation is applied the likelihood becomes

$$\begin{aligned} \widetilde{L}(Y_i, T_i, \delta_i; \theta) = \; & (2\pi)^{\frac{nq}{2}} \hat{\Omega}^{-\frac{n}{2}} \textstyle\prod_{i=1}^{n} \prod_{j=1}^{m_i} [(2\pi\sigma_e^2)^{-\frac{1}{2}} \exp[-(Y_i(t_{ij}) - \hat{W}_i(t_{ij}))^2 (2\sigma_e^2)^{-1}]] \\ & \times (2\pi|D|)^{-\frac{1}{2}} \exp(-2^{-1} \hat{b}_i' D^{-1} \hat{b}_i) \\ & \times [\exp[\tfrac{-(T_i(t_i)\sigma(2\pi)^{1/2})^{-1}[\log T_i(t_i) - \hat{\mu}_i(t_i)]^2}{2\sigma^2}]]^{\delta_i} \\ & \times [1 - \Phi(\tfrac{\log(T_i(t_i)) - \hat{\mu}_i(t_i)}{\sigma})]^{1-\delta_i}, \end{aligned} \tag{D.5}$$

where $\widehat{W}_i(t_{ij}) = X_i^T(t_{ij})\beta + Z_i^T(t_{ij})\hat{b} + U_i^T \lambda$ and $\hat{\mu}_i(t_i) = \alpha \hat{W}_i(t) + M_i^T \gamma$.

D.1.2 Left-Truncation and Right-Censoring Likelihood for Application of Joint Modelling with Log-Normal Sub-Model

For left-truncation and right-censoring case, the likelihood becomes more complicated.

$$\begin{aligned} L(Y_i, T_i, A_i, \delta_i; \theta) = \; & \prod_{i=1}^{n} [\int_{-\infty}^{\infty} \prod_{j=1}^{m_i} [(2\pi\sigma_e^2)^{-\frac{1}{2}} \exp[-(Y_i(t_{ij}) - W_i(t_{ij}))^2 (2\sigma_e^2)^{-1}]] \\ & \times (2\pi|D|)^{-\frac{1}{2}} \exp(-2^{-1} b_i' D^{-1} b_i) \\ & \times [\exp[(\log A_i - \mu_i(t_i))^2 - (\log(A_i + T_i(t_i)) - \mu_i(t_i))^2]]^{\delta_i} \\ & \times [[1 - \Phi(\tfrac{\log(T_i(t_i)+A_i) - \mu_i(t_i)}{\sigma})][1 - \Phi(\tfrac{\log A_i - \mu_i(t_i)}{\sigma})]^{-1}]^{1-\delta_i} \, db_i \;]. \end{aligned} \tag{D.6}$$

When Gauss-Hermite approximation is applied the likelihood becomes

$$\begin{aligned} \widetilde{L}(Y_i, T_i, A_i, \delta_i; \theta) = \; & 2^{\frac{qn}{2}} |B|^{-n} \prod_{i=1}^{n} \sum_{l_i=1}^{p} \omega_{l_i} \exp\{E_i' E_i\} \prod_{j=1}^{m_i} [(2\pi\sigma_e^2)^{-\frac{1}{2}} \\ & \times \exp[-(Y_i(t_{ij}) - \hat{W}_i(t_{ij}))^2 (2\sigma_e^2)^{-1}]] \\ & \times (2\pi|D|)^{-\frac{1}{2}} \exp(-2^{-1/2} B^{-1} E_i D^{-1} 2^{1/2} B^{-1} E_i) \\ & \times [\exp[(\log A_i - \hat{\mu}_i(t_i))^2 - (\log(A_i + T_i(t_i)) - \hat{\mu}_i(t_i))^2]]^{\delta_i} \\ & \times [[1 - \Phi(\tfrac{\log(T_i(t_i)+A_i) - \hat{\mu}_i(t_i)}{\sigma})][1 - \Phi(\tfrac{\log A_i - \hat{\mu}_i(t_i)}{\sigma})]^{-1}]^{1-\delta_i}, \end{aligned} \tag{D.7}$$

where $\widehat{W}_i(t_{ij}) = X_i^T(t_{ij})\beta + Z_i^T(t_{ij})(2^{\frac{1}{2}} B^{-1} E_{il_i}) + U_i^T \lambda$ and $\hat{\mu}_i(t_i) = \alpha \hat{W}_i(t) + M_i^T \gamma$. When Laplace approximation is applied the likelihood becomes

$$\begin{aligned} \widetilde{L}(Y_i, T_i, \delta_i; \theta) = \; & (2\pi)^{\frac{nq}{2}} \hat{\Omega}^{-\frac{n}{2}} \prod_{i=1}^{n} \prod_{j=1}^{m_i} [(2\pi\sigma_e^2)^{-\frac{1}{2}} \exp[-(Y_i(t_{ij}) - \hat{W}_i(t_{ij}))^2 (2\sigma_e^2)^{-1}]] \\ & \times (2\pi|D|)^{-\frac{1}{2}} \exp(-2^{-1} \hat{b}_i' D^{-1} \hat{b}_i) \\ & \times [\exp[(\log A_i - \hat{\mu}_i(t_i))^2 - (\log(A_i + T_i(t_i)) - \hat{\mu}_i(t_i))^2]]^{\delta_i} \\ & \times [[1 - \Phi(\tfrac{\log(T_i(t_i)+A_i) - \hat{\mu}_i(t_i)}{\sigma})][1 - \Phi(\tfrac{\log A_i - \hat{\mu}_i(t_i)}{\sigma})]^{-1}]^{1-\delta_i}, \end{aligned} \tag{D.8}$$

where $\widehat{W}_i(t_{ij}) = X_i^T(t_{ij})\beta + Z_i^T(t_{ij})\hat{b} + U_i^T \lambda$ and $\hat{\mu}_i(t_i) = \alpha \hat{W}_i(t) + M_i^T \gamma$.

E Joint Modelling with Extreme Value for Survival Sub-model

If the Extreme Value distribution, is preferred to be used for the survival sub-model, it should be noted that when covariates are present, the log-likelihood, the probability density function and the survival function (along with other quantities) of the extreme value distribution have the same formulas as those for Weibull's distribution. The only difference is that there is no log tranformation of time ((63),(90), (97)).

F Joint Modelling with Gompertz Survival Sub-model

If Gompertz distribution is needed to fit the data, it can be achieved by using the Extreme Value model. Under the transformations $a = -\log\{\sigma\} - \mu\sigma^{-1}$ and $b = \sigma^{-1}$, the Gompertz hazard function, becomes the Extreme hazard function. Thus, by fitting the Extreme Value model, and using the above transformation, the Gompertz model is obtained ((3), (97),(106)).

G Joint Modelling with Logistic Survival Sub-model

If the Logistic distribution, is preferred to be used for the survival sub-model, it should be noted that when covariates are present, the log-likelihood, the probability density function and the survival function (along with other quantities) of the Logistic model have the same

formulas as those for Log-Logistic's distribution. The only difference is that there is no log tranformation of time ((63),(90),(97)).

H Joint Modelling with Beard Survival Submodel

If Beard model is needed to fit the data, it can be achieved by using the Logistic model. Under the transformations $a = -\log\{\sigma\} - \mu\sigma^{-1}$, $b = \sigma^{-1}$ and $\rho = \log\sigma$, the Beard hazard function becomes the Logistic hazard function. Thus, by fitting the Logistic model, and using the above transformation, the Beard model is obtained, ((3),(97),(106)).

I Joint Modelling with Perks Survival Submodel

If Perks model is needed to fit the data, it can be achieved by using the Beard (and thus the Logistic) model. Under the transformation $\rho = 0$, the Beard hazard function, becomes the Perks hazard function. Thus, by fitting the Beard model, and using the above transformation, the Perks model is obtained, ((3),(97),(106)).

J R documentation

J.1 Fitting Parametric Joint Models

Description

This function fits parametric joint models to right-censored data, and left-truncated and right-censored data for Weibull, Log-Logistic, Log-Normal, Extreme and Logistic survival distributions. Gompertz, Beard and Perks distributions can be obtained using the models mentioned before as shown in the Appendix F, H and I respectively.

It obtains the maximum likelihood estimates, by maximizing the log-likelihood, using Newton-Raphson algorithm i.e. $x_{n+1} = x_n - \frac{f^{'}(x)}{f(x)}$. The algorithm uses starting parameters obtained from $survreg()$, $aftreg()$ and $lme()$ functions.

Usage

```
fit.joint.model(dist="weibull",random=2,censored="right",
truncation="no",patient.in.longitudinal.form,
start.time.longitudinal,end.time.longitudinal,X1,Z1,U1,
censoring.longitudinal,M.longitudinal,
patient.in.survival.form,X2,Z2,U2,M,age=NULL,follow.up.time.for.survival,censorin
Y,nnodes,bootstrap=F,B)
```

Arguments

patient.in.longitudinal.form: vector containing the patient id in a longitudinal form

X1: vector or matrix that includes all time-depedenent variables in a longitudinal form

Z1: vector or matrix with the random effects in a longitudinal form. It can only take up to 2 random effects

U1: vector or matrix with time-independent variables in a longitudinal form

patient.in.survival.form: vector showing the patient id in a simple survival form

X2: vector or matrix that includes all time-dependent variable in a simple survival form

Z2: vector or matrix with the random effects in a survival form

U2: vector or matrix with time-independent variables in a survival form

M: vector or matrix with the survival covariates in a simple survival form

follow.up.time.for.survival: vector with the follow-up time for each patient in a simple survival form

start.time.longitudinal: vector with the starting time for each patient in a longitudinal

form for each new observation

end.time.longitudinal: vector with the ending time for each patient in a longitudinal form before the next observation

censoring.longitudinal: vector showing the status of the patient in a longitudinal form (1-died,0-alive)

M.longitudinal: vector or matrix with the survival covariates in a longitudinal form

censoring.survival: vector showing the status of the patient in a survival form (1-died,0-alive)

Y: vector showing the value of the longitudinal biomarker in a longitudinal form

random: 2 indicating 2 type of random effects, 1 indicating for 1 type of random effects

nnodes: the number of nodes to be used in Gauss-Hermite approximation

bootstrap: If TRUE, it performs a boostrap analysis with B number of replications

dist="weibull": One of the distributions Weibull, Log-Logistic, Log-Normal, Extreme, Logistic. Additionally, Gompertz, Perks, Beard distributions can be obtained by using the transformations mentioned in the Appendix.

censored="right": right-censoring only

truncation="no": "no" or "yes" (left-truncation only)

age=NULL: If truncation="yes", a vector with the starting ages of all individuals

Example

```
#taking a data that will be used from a library in R
>library(JM)
>Long=pbc2
>Surv=pbc2.id

>patient.in.longitudinal.form=Long$id
#patient id in longitudinal form

>patient.in.survival.form=Surv$id
#patient id in survival form

>X1=Long$year
#time in longitudinal form

>X2=Surv$year
#time in survival form

>Z1=Long$id
#random intercept in longitudinal form

>Z2=Surv$id
```

```
#random intercept in survival form

>U1=rep(1,length(X1))
#longitudinal intercept in longitudinal form

>U2=rep(1,length(X2))
#longitudinal intercept in survival form

>M=rep(1,length(X2))
#survival intercept in survival form

>M.longitudinal=rep(1,length(X1))
#survival intercept in survival form

>follow.up.time.for.survival=Surv$years
#follow up time in survival form

>start.time.longitudinal=Long$year
#time for each longitudinal measurement

>end.time.longitudinal=Long$years
#time before the next longitudinal measurements

>censoring.longitudinal=Long$status2
#censoring in longitudinal form

>censoring.survival=Surv$status2
#censoring in survival form
```

```
>Y=Long$serBilir
#longitudinal biomarker in longitudinal form

>nnodes=2
#nodes for Gauss-Hermite

>B=100
#bootstrap samples

>age=pbc2.id$age
#age used for truncation time in left-truncated model

#fitting a left-truncated and right censored Weibull model
> fit.joint.model(dist="weibull",random=2,censored="right",
truncation="left",patient.in.longitudinal.form,
start.time.longitudinal,end.time.longitudinal,X1,Z1,U1,
censoring.longitudinal,M.longitudinal,
patient.in.survival.form,X2,Z2,U2,M,age=age,follow.up.time.for.survival,censoring
Y,nnodes,bootstrap=F,B)
```

J.2 Simulations

Description

This function simulates Weibull shared parameter model with patient observations 1 to 4 or 1 to 24. The a Weibull shared parameter model is fitted using analytical and bootstrap

methods.

It obtains the maximum likelihood estimates, by maximizing the log-likelihood, using Newton-Raphson algorithm i.e. $x_{n+1} = x_n - \frac{f^{'}(x)}{f(x)}$. The algorithm uses starting parameters obtained from $survreg()$, $aftreg()$ and $lme()$ commands applied to the data.

Usage

```
autosimulation.generate(numberOfPatients,averageObservationCount,D,
alpha,beta,lambda,gamma,log.sigma,sigma.e,nnodes,B,
Simulations,Total.Simulations)
```

Arguments

numberOfPatients: the number of patients the user wants to simulate for the data

averageObservationCount: the number of observations the user wants to simulate. It takes 3 and 23 which simulates 1to4 and 1to24 observations for each patient.

nnodes: the number of nodes the user wants to use for the Gauss-Hermite approximation

B: the number of bootstrap analyses to perform

Simulations: the number of successful simulations that is required for the program to

stop

Total.Simulations: the number of the total simulations that need to be performed until the program will stop

The rest of the arguments D,alpha,beta,lambda,gamma,log.sigma, sigma.e would be D, α, β, λ, γ and $\log \sigma$ respectively, that form the shared parameter model $T_i^*(t_{ij}) = \exp[\alpha(t_{ij}\beta + b_i + \lambda) + \gamma_1 + \gamma_2 g_i + \sigma \times \epsilon_i]$, where the univariate random effects b_i are generated from $\mathcal{N}(0, D)$ and the errors $e_i(t_{ij})$ are independently generated from $\mathcal{N}(0, \sigma_e^2)$ and $\gamma = c(\gamma_1, \gamma_2)$.

Example

```
#Defining the actual values of the parameters

>D.1=0.1
#variance-covariance for the random effects
#constant

>alpha.1=1.0
#association between the longitudinal
#and survival procedures

>beta.1=1.0
#time-dependent longitudinal covariate

>lambda.1=1.0
#time-independent longitudinal covariate
```

```
>gamma.1=1.0
#survival covariate

>log.sigma.1=0.1
#log(scale) for the survival model

>sigma.e.1=0.1
#longitudinal variance for the longitudinal model

>nnodes.1=2
#Gauss-Hermite nodes to be used

>numberOfPatients.1=100
#number of patients to be simulated

>averageObservationCount.1=3
#average observations for the patients
#when it is 3, they have observation from 1 to 4

>B.1=100
#bootstrap samples

>Simulations.1=500
#total replications until the simulation are finished

>Total.Simulations.1=500
#total replications for the simulations
```

```
#begin the simulations program
>SKEVI1=autosimulation.generate(numberOfPatients.1,
averageObservationCount.1,D.1,alpha.1,beta.1,
lambda.1,gamma.1,log..sigma.1,sigma.e.1,nnodes.1,
B.1,Simulations.1,Total.Simulations.1)
```