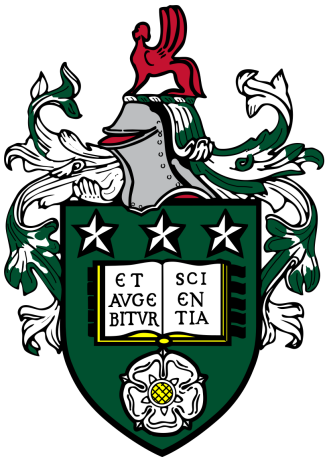
Advances in shared frailty models with application to twin data



Annah Mwikali Muli

School of Mathematics

University of Leeds

A thesis submitted in accordance with the requirements for the degree of

Doctor of Philosophy

March 2023

Acknowledgements

I would like to thank IMforFUTURE EU Marie Skłodowska-Curie Innovative Training Network grant agreement number 721815 for three years of funding and School of Mathematics University of Leeds for ten months extension. I would also like to acknowledge the training offered in the IMforFUTURE network.

Many thanks to my supervisors Prof. Jeanine Houwing-Duistermaat and Dr. Arief Gusnanto. Without your excellent guidance, constant encouragement, and endless support, this thesis would not have been possible. I would also like to acknowledge Dr. Mar Rodriguez-Girondo for her fruitful collaboration and supervision during my one year secondment to Leiden University Medical Center. Thanks also to the wonderful friends that I met here, in Leeds. Finally, I am grateful to my family, you have supported me every step of the way.

TwinsUK is funded by the Wellcome Trust, Medical Research Council, European Union, Chronic Disease Research Foundation (CDRF), Zoe Global Ltd and the National Institute for Health Research (NIHR)- funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London.

Abstract

This thesis focuses on advanced statistical methods for the analysis of correlated survival times with a focus on investigation of twin data. The methods we will discuss were motivated by the TwinsUK study to investigate the relationships between fracture incidence with covariates. To model the correlation between twins we use shared frailty models. However, there is a potential danger of bias in the estimation if the frailty distribution is misspecified. Frailties are often assumed to follow a gamma distribution. To safeguard us from the impact of the misspecification of this distribution, we consider flexible baseline hazards, for instance B-splines in addition to a parametric baseline hazard. We apply this methodology to the TwinsUK cohort to predict the probability of experiencing a fracture in the next five or ten years, given their bone mineral densities (BMD) and their health status. The models with parametric and more flexible baseline hazards yield very close results in estimating survival probabilities and thus a choice of parametric baseline hazard is generally preferred. We find that bone mineral density is a statistically significant predictor in the model whereas health status is not.

We then-via simulation studies-assess the consequences of frailty distribution misspecification of estimation of parameters and survival probabilities. When the Weibull baseline hazard is used, in most cases the scale parameter corrected for the wrong frailty. However, for some extreme cases it appears that the scale parameter cannot adjust and thus parameters and survival probabilities are affected. However, using a flexible function for the baseline hazard improves estimation of parameters as well as survival probabilities in the presence of frailty distribution misspecification.

Often age is preferred underlying time scale. However, participants have different ages at entry hence using age result in delayed entry. An additional challenge is clustering in the twins data. In this thesis we will develop methods to estimate models for the relationship between a time-varying covariate and age to an event while adjusting for delayed entry. Four approaches for modeling time varying covariates namely, last observation carried forward, risk set regression calibration, ordinary regression calibration, and joint modelling approaches will be adapted to the situation of clustered data with delayed entry.

Contents

List of Figures

List of Tables

Chapter 1

General introduction

1.1 Background

Time-to-event data are common in many disciplines including, but not limited to, medical research. The time-to-event is often regarded as the survival time. The Cox proportional hazard model is often used to investigate the effect of covariates on the hazards of an event such as disease onset, death, or fracture. The Cox proportional hazard model has two assumptions: first, the effect of the covariates on survival is proportional and, second, the population is homogeneous given the covariates. To verify the first assumption, some statistical tests have been developed (Grambsch and Therneau, 1994; Ng'andu, 1997). However, the second assumption is more difficult to verify, although we can still address that. For example, if we have correlated observations such as in twins, unobserved shared effects can be modelled by a frailty as introduced by Vaupel et al. (1979). This thesis will focus on the shared frailty model. The idea is to introduce an additional parameter to the hazard rate to model the random frailties.

The shared frailty model involves the conditional specification of the hazard as if the frailty would be observed. To derive the marginal likelihood for estimation of parameters we have to specify a distribution for frailties which has a positive support, mean of one and a positive variance. The gamma distribution has commonly been considered for the frailties because of mathematical convenience since it produces a more tractable marginal likelihood function for the parameters after integration. Some theoretical results in semi-parametric gamma frailty models, in particular, estimators and asymptotic properties including when

we have left truncation have been shown by Murphy (1994, 1995). However, other frailty distributions have been proposed in the literature (Section 2.2.1 to Section 2.2.4).

The work in this thesis is motivated by the estimation of covariates effect and survival probabilities of the TwinsUK study. We are particularly interested in survival probabilities because hazard rates are hard to interpret. We consider data from the TwinsUK Study in which we have fracture information, BMD and glycans measurements. The TwinsUK Study is a well-known longitudinal study that is ideal for study of the process of ageing and age related diseases (https://twinsuk.ac.uk). Specifically, we are interested to estimate the probability of experiencing a fracture in the next five or ten years, given their bone mineral densities (BMD) and their health status based on glycans. For the purpose of our analysis, we consider the difference between the glycan age (as biological age) and the chronological age. We term this difference as the *frailty index* (not to be confused with frailties as random effects in survival models). *Frailty index* will represent a participants health status. A positive *frailty index* indicates that the individual experiences faster biological ageing than their chronological age (unhealthy) and the other way around with a negative *frailty index*.

To achieve the objective in estimating the probability of fracture in the next time period we specify the baseline hazard and consider to model the hazard of fracture with a gamma distribution frailties for twin pairs. Unfortunately, it is hard to identify which frailty distribution fits real data best. This is because in practice, the underlying frailty distribution is unknown and thus the wrong frailty distribution might be chosen. However, because both baseline hazard and frailty act multiplicatively on the hazard, we suggest that a flexible baseline may compensate for a misspecification in the frailty distribution. We therefore fit both parametric and flexible baseline hazards to the data. We also investigate how specifying a frailty distribution may affect estimates of covariate effect and survival probabilities assuming both parametric and flexible baseline hazards via simulations. Four frailty distributions namely; gamma, inverse Gaussian, lognormal, and mixture normal are considered for generating the data. For each of the frailty model, a gamma/inverse Gaussian frailty model is then fitted to the data.

We then consider age as time scale especially because it is often used as an underlying time scale in ageing studies. Using age as time scale also does not make the assump-

tion that there is a linear relationship between age and hazard of event (Cologne et al., 2012). However, participants have different ages at entry. In the data analysis participants' inclusion criterion is dependent on a participant's survivorship at a given enrolment period. A participant was only included in the study if they are elderly (aged beyond 50). Such a selection criterion results in left truncated observations as a result of delayed entry. Delayed entry results in non-random samples because inclusion in study depends on the survivorship beyond truncation time i.e., truncation is dependent on survival of each observation. In theory frailer individuals experience the event first and hence it implies that individuals with higher frailty values are less likely to be observed after truncation. Consequently the mean of frailty distributions becomes smaller and smaller as the frail individuals die and less frail individuals remain and so does the variance also become smaller with time. Due to the selection criteria we are likely to be left with small frailty values. This selection scheme is thus complex. If this sampling scheme is not well addressed the assumed frailty distribution might not represent the observed values which consequently may affect estimation of regression coefficients, hazard rate and survival probabilities.

When we have delayed entry and use age as the time scale then an updated frailty distribution is used which assumes that the conditional density of the observed units within a cluster is averaged over the conditional distribution given the entry times (Jensen et al., 2004; Kvist et al., 2010; Rondeau et al., 2003; Van den Berg and Drepper, 2016). Although this method may still give biased estimates when not all clusters are fully observed (Rodríguez-Girondo et al., 2018).

Our work makes contributions in both statistical methodology and applications. In particular we make the following contributions:

- Via simulations we investigate how specifying a wrong frailty distribution may affect estimates of time independent covariate effect and survival probabilities assuming both parametric and flexible baseline hazard. We wrote an R code for fitting frailty model using B-splines for baseline hazard.
- Existing methodology can deal with either delayed entry or time varying covariates separately. We develop methodology for dealing with delayed entry and time varying covariates simultaneously in clustered data when using age as a time scale.

- Via simulations we investigate how different approaches model time dependent covariates generated with or without measurement error while adjusting for delayed entry.
- With the proposed approaches in Chapter 7, we were able to assess the effect of time varying covariate BMD on age specific fracture incidence for the TwinsUK study.

1.2 Structure of the thesis

This thesis has eight chapters. This section provides an overview of how the chapters are organized. The current is Chapter 1 which is a general introduction of the thesis.

Chapter 2 will detail time independent covariate model mathematical formulation and assumptions for Cox and shared frailty model including commonly used frailty distributions. Chapter 3 extends the Cox model to deal with time varying covariates. Four existing approaches are discussed: last observation carried forward, risk-set regression calibration, ordinary regression calibration and joint modelling. The regression calibration approach attempts to model the time-varying covariate process using a linear mixed model and then, in a second stage, plug the predicted response into the survival model.

Chapter 4 describes the TwinsUK study. We have data on fracture incidence, BMD and glycomics for the twins. We would like to estimate the survival probabilities for the occurrence of a fracture in the next time period given a participant's BMD and health status based on glycans using time to follow-up as underlying time scale (addressed in Chapter 5). The analysis does not consider glycans directly as a predictor. However, we consider the difference between the glycan age (as biological age) and the chronological age. We term this as the *frailty index*. We would also like to estimate the effect of BMD as a time dependent covariate on fracture incidence using age as time scale (addressed in Chapter 7).

In Chapter 5 we apply shared frailty model to the TwinsUK cohort to predict the probability of experiencing a fracture in the next time period (e.g. five or ten years) given their bone mineral densities (BMD) and their *frailty index*. The models with parametric and flexible baseline hazards yield very close results in estimating survival probabilities. We

find that bone mineral density is a significant predictor in the model whereas *frailty index* is not. Low BMD leads to a larger probability of fracture.

Chapter 5 on analysis of twins data motivates us to also investigate if the parametric and flexible baseline hazards will always yield close survival probabilities estimates for different frailty distributions. This leads us to investigate the impact of frailty misspecification on the estimation of parameters and survival probabilities as shown in Chapter 6. This is the first study to investigate the effect of frailty distribution misspecification on the estimation of survival probabilities. The simulation results show that the parametric baseline hazard (Weibull) scale parameter corrects for the wrong distribution. This suggests that using a flexible baseline hazard could correct for frailty misspecification. We investigate this claim using two flexible approaches for the baseline hazard. The first one uses a plug-in estimator for the baseline hazard while the second approach uses splines for the baseline hazard. Results show that flexible baseline hazards appears to improve estimation of population survival probabilities when a wrong frailty distribution is used. Therefore the choice of baseline hazard plays a role in frailty misspecification correction.

Chapter 7 extends models for time dependent covariates to adjust for delayed entry in twins. This is motivated by the analysis of the TwinsUK data done in Chapter 5. Since we have multiple observations per subject we consider BMD as a time varying covariate. The challenge faced will be to adjust for delayed entry because age at time of fracture is used as time scale. We adapt the approaches for modeling time varying covariates for singletons to modeling time varying covariates of clustered data with delayed entry. Via simulations we investigate how the last observation carried forward and regression calibration approach fit covariates generated with and without measurement error. We then model effect of BMD as a time varying covariate on fracture using age as time scale while taking into account delayed entry.

Chapter 8 gives our conclusions, future work and possible extensions of the thesis.

Chapter 2

Modelling time independent covariates

Time-to-event is also called survival time, failure time or event time and is modelled as a random variable. Some examples include time to disease onset, time to fracture time to death, time to the appearance of a tumor or recurrence of a disease e.t.c Various survival analysis methods exist in the literature to handle time-to-event data. These methods can be broadly classified into parametric methods (for the case where we make an assumption on the distribution of survival time) and non-parametric (does not make any assumption on the survival time distribution) (Cox and Oakes, 1984; Kalbfleisch and Prentice, 2011; Klein and Moeschberger, 2003)

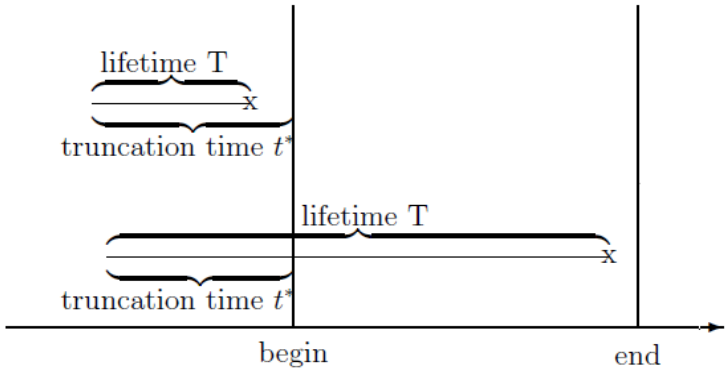
A unique feature of survival data is that typically not all individuals experience the event (eg, death) by the end of the study period, and hence actual survival times is unknown. This phenomenon is referred to as censoring and has to be accounted for in the analysis to allow for valid inferences. Censoring can broadly be classified to left censoring, right censoring and interval censoring. Left censoring occurs when a subject is known to have had the event before the start of the observation, but the exact time of the event is unknown. Similarly, interval censoring is where it is only known that the event occurred between two time points, but the exact time is unknown. Right censoring occurs when the time-to-event cannot be determined because either the study ended and the event hasn't occurred yet or such subject was lost to follow-up at any time during the study. If censoring is independent of the event process, then it is non-informative, otherwise it is informative. Throughout the thesis the censoring process is considered to be non-informative. Let T_i^*

denote the time elapsed for subject i to experience the event of interest, and let C_i denote the censoring time for subject i. We define T_i as the minimum of the true event time, T_i^* , and the censoring time, C_i, so $T_i = \min(T_i^*; C_i)$. δ_i is the event indicator defined as follows:

$$\delta_i = \begin{cases} 1 & \text{if event observed (i.e. } T_i^* \leq C_i) \\ 0 & \text{if censored (i.e. } T_i^* > C_i) \end{cases}$$

Another unique feature of survival data is referred to as truncation which is related to the sampling mechanism. Right truncation occurs when only individuals who have experienced the event of interest are observable. Left truncation occurs when subjects come under observation only if their failure times exceed some given time t_0. It is only because they did not fail before t_0 that we even know about their existence. Therefore, only individuals with survival time greater than t_0 are observed; see Figure 2.1.

Figure 2.1: Left truncation



Chapter 7 of this thesis will also entail a discussion on left truncation.

Relation between some basic survival quantities

Ways of characterizing the distribution of the time-to-event random variable T include:

$h(t)$: Probability of experiencing an event in the next time instance given that you have survived up to time t.

$H(t)$: Cumulative hazard function

$f(t)$: Unconditional probability density function of an event occuring at time t.

$F(t)$: Cumulative distribution function

$S(t)$: Probability of surviving beyond time t.

The above quantities have the following relations

$$\begin{aligned}
f(t) &= \lim_{dt\to 0} \frac{\Pr\left(t \le T < t+dt\right)}{dt} = \frac{dF(t)}{dt} \\
F(t) &= \Pr\left(T \le t\right) = \int_{s=0}^{t} f(s)ds \\
h(t) &= \lim_{dt\to 0} \frac{\Pr\left(t \le T < t+dt | T \ge t\right)}{dt} = \frac{f(t)}{S(t)} = -\frac{d\ln[S(t)]}{dt} \\
H(t) &= \int_{s=0}^{t} h(s)ds = -\ln[S(t)] \\
S(t) &= \Pr\left(T > t\right) = \exp(-H(t)) = 1 - F(t)
\end{aligned}$$

In order to explore the relationship between survival of an individual and time independent variables, statistical modelling can be used. The most popular regression model is Cox proportional hazard model. This model will be described in the next session. It can also be referred to as a relative risk model.

2.1 Cox proportional hazards model

Let T denote a non-negative random variable representing survival time, with probability density function $f(t)$ and cumulative distribution function $F(t)$. Let $h_i\left(t|\boldsymbol{x}_i\right)$ be the hazard rate at time t for i^{th} individual with risk covariate $\boldsymbol{x}_i$ $(1 \times p$ vector), $\boldsymbol{\beta} = \left(\beta_1, \beta_2, ..., \beta_p\right)^\top$ be a parameter vector, p is the number of predictors, and $\boldsymbol{x}_i\boldsymbol{\beta}$ is a linear predictor or risk score. $h_0\left(t\right)$ be the baseline (the hazard function of the outcome occurring for subjects with linear predictor zero.).

The Cox proportional regression model is given by (Cox, 1972)

$$h_i\left(t|\boldsymbol{x}_i\right) = h_0\left(t\right)\exp\left(\boldsymbol{x}_i\boldsymbol{\beta}\right) = h_0\left(t\right)\exp\left(\sum_{k=1}^{p}\beta_k x_{ik}\right). \tag{2.1}$$

This model describes the effect of the covariates on the hazard of the occurrence of the event.

$S(t|\boldsymbol{x}_i)$ is the survival function of an individual conditional to the risk covariate $\boldsymbol{x}_i$ and is

given by

$$S(t|\boldsymbol{x}_i) \quad = \quad \exp\left[-H_0(t)\exp\left(\boldsymbol{x}_i\boldsymbol{\beta}\right)\right].$$

The cumulative density function of the event times under the Cox proportional hazards model is:

$$F(t|x) \quad = \quad 1-\exp\left[-H_0(t)\exp\left(\boldsymbol{x}_i\boldsymbol{\beta}\right)\right].$$

2.1.1 Estimation of $\boldsymbol{\beta}$

The parameters ($\boldsymbol{\beta}$) in the Cox model can be estimated by using partial maximum likelihood estimation. Let $t_1 < t_2 < \ldots < t_n$ denote ordered distinct event times. Let $x_{(i)k}$ be the k^{th} covariate associated with individual whose failure time is t_i. Let δ_i be the censoring indicator. Define the risk set at time t_i to be $R(t_i)$ which is the set of all individuals who are still under study at a time just prior to t_i

The probability that an individual dies at time t_i with covariates $\boldsymbol{x}_{(i)}$, given one of the individuals in $R(t_i)$ dies at this time, is given by

$$\begin{aligned} P[\text{individual dies at } t_i \mid \text{one death at } t_i] &= \frac{P[\text{individual dies at } t_i \mid \text{survival to } t_i]}{P[\text{one death at } t_i \mid \text{survival to } t_i]} \\ &= \frac{h(t_i \mid \boldsymbol{x}_{(i)})}{\sum_{j\in R(t_i)} h(t_i \mid \boldsymbol{x}_j)} \\ &= \frac{h_0(t_i)\exp\left(\sum_{k=1}^{p}\beta_k x_{(i)k}\right)}{\sum_{j\in R(t_i)} h_0(t_i)\exp\left(\sum_{k=1}^{p}\beta_k x_{jk}\right)} \\ &= \frac{\exp\left(\sum_{k=1}^{p}\beta_k x_{(i)k}\right)}{\sum_{j\in R(t_i)}\exp\left(\sum_{k=1}^{p}\beta_k x_{jk}\right)}. \end{aligned} \tag{2.2}$$

The partial likelihood for distinct event time censored data is given by

$$L_p\left(\boldsymbol{\beta}\right) \quad = \quad \prod_{i=1}^{n}\left\{\frac{\exp\left(\sum_{k=1}^{p}\beta_k x_{(i)k}\right)}{\sum_{j\in R(t_i)}\exp\left(\sum_{k=1}^{p}\beta_k x_{jk}\right)}\right\}^{\delta_i}. \tag{2.3}$$

The corresponding partial log likelihood $\ell_p(\boldsymbol{\beta}) = \ln\left[L_p\left(\boldsymbol{\beta}\right)\right]$ is

$$\begin{aligned}
\ell_p(\boldsymbol{\beta}) &= \ln\left[\prod_{i=1}^{n}\left\{\frac{\exp\left(\sum_{k=1}^{p}\beta_k x_{(i)k}\right)}{\sum_{j\in R(t_i)}\exp\left(\sum_{k=1}^{p}\beta_k x_{jk}\right)}\right\}^{\delta_i}\right] \\
&= \sum_{i=1}^{n}\delta_i\ \ln\left[\frac{\exp\left(\sum_{k=1}^{p}\beta_k x_{(i)k}\right)}{\sum_{j\in R(t_i)}\exp\left(\sum_{k=1}^{p}\beta_k x_{jk}\right)}\right] \\
&= \sum_{i=1}^{n}\delta_i\left\{\ln\left[\exp\left(\sum_{k=1}^{p}\beta_k x_{(i)k}\right)\right] - \ln\left[\sum_{j\in R(t_i)}\exp\left(\sum_{k=1}^{p}\beta_k x_{jk}\right)\right]\right\} \\
&= \sum_{i=1}^{n}\delta_i\left\{\sum_{k=1}^{p}\beta_k x_{(i)k} - \ln\sum_{j\in R(t_i)}\exp\left(\sum_{k=1}^{p}\beta_k x_{jk}\right)\right\}. \qquad (2.4)
\end{aligned}$$

Numerical methods can be used to maximize this partial log-likelihood function e.g by Newton-Raphson procedure.

2.1.1.1 The Newton-Raphson procedure in matrix form

The partial likelihood for distinct event time censored data is given by

$$L_p\left(\boldsymbol{\beta}\right) = \prod_{i=1}^{n}\left\{\frac{\exp\left(\boldsymbol{x}_i\boldsymbol{\beta}\right)}{\sum_{j\in R(t_i)}\exp\left(\boldsymbol{x}_j\boldsymbol{\beta}\right)}\right\}^{\delta_i}, \qquad (2.5)$$

where δ_i is an event indicator.

Let $w_i = \exp\left(\boldsymbol{x}_i\boldsymbol{\beta}\right)$, $Y_j(t_i)$ be at risk indicator and $W_j = \sum_{j\in R(t_i)}\exp\left(\boldsymbol{x}_j\boldsymbol{\beta}\right) = \sum_{j=1}^{n}Y_j(t_i)\exp\left(\boldsymbol{x}_j\boldsymbol{\beta}\right)$ then equation (2.5) becomes

$$L_p\left(\boldsymbol{\beta}\right) = \prod_{i=1}^{n}\left\{\frac{w_i}{W_j}\right\}^{\delta_i},$$

The corresponding partial log-likelihood will be

$$\ell_p\left(\boldsymbol{\beta}\right) = \sum_{i=1}^{n}\delta_i\log w_i - \sum_{i=1}^{n}\delta_i\log W_i,$$

Let $\eta_i = \log w_i = \boldsymbol{x}_i\boldsymbol{\beta}$, $P_{n\times n} = ((\pi_{ij}))$and $\pi_{ki} = Y_i(t_k)\frac{w_k}{W_i}$ then,

$$\begin{aligned}
\frac{\partial\ell_p}{\partial\eta_k} &= \delta_k - \sum_i \pi_{ki}\delta_i, \\
U(\boldsymbol{\eta}) &= \boldsymbol{\delta} - \boldsymbol{P\delta}, \\
U(\boldsymbol{\beta}) &= \boldsymbol{X}^T(\boldsymbol{\delta} - \boldsymbol{P\delta}).
\end{aligned}$$

For the second derivative

$$\begin{aligned}
\frac{\partial^2 \ell_p}{\partial \eta_k^2} &= -\sum_i \delta_i \pi_{ki}(1-\pi_{ki}) \quad \text{Diagonal elements} \\
\frac{\partial^2 \ell_p}{\partial \eta_k \partial \eta_j} &= \sum_i \delta_i \pi_{ki} \pi_{ji} \quad \text{Off-diagonal elements} \\
I(\beta) &= \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X}.
\end{aligned}$$

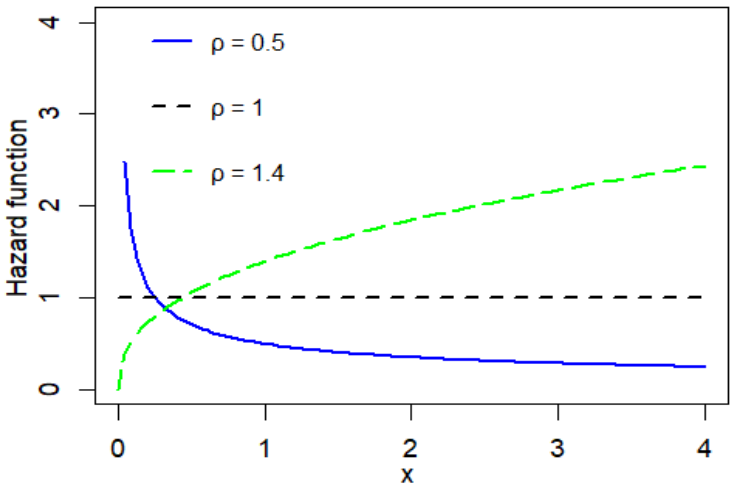
$\boldsymbol{W}$ matrix with terms given above (diagonal elements and off-diagonal elements). Therefore the Newton Raphson update will be

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{m+1} &= \hat{\boldsymbol{\beta}}_m + I^{-1}(\hat{\boldsymbol{\beta}}_m) U(\hat{\boldsymbol{\beta}}_m). \\
&= \hat{\boldsymbol{\beta}}_m + \left(\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T (\boldsymbol{\delta} - \boldsymbol{P}\boldsymbol{\delta})
\end{aligned} \tag{2.6}$$

Parametric proportional hazards model

A parametric form for the baseline hazard function in equation (2.1) leads to a parametric proportional hazards model. Some common distribution for the baseline hazards include the Weibull and exponential. The Weibull baseline hazard is characterized by two parameters i.e. shape parameter ρ and scale parameter λ. It is monotone increasing when $\rho > 1$, monotone decreasing for $\rho < 1$, and constant for $\rho = 1$ (note that the Weibull distribution reduces to exponential for $\rho = 1$).

Figure 2.2: Hazard function of Weibull for $\lambda = 1$ and $\rho = 0.5, 1, 1.4$.



Let $\boldsymbol{\xi}$ contain the parameters of the baseline hazard e.g for a Weibull hazard $\boldsymbol{\xi} = (\lambda, \rho)$.

The full likelihood is given by

$$\begin{aligned} L(\boldsymbol{\xi}, \boldsymbol{\beta}) &= \prod_{i=1}^{n} \left[h_i(t|\boldsymbol{x}_i)\right]^{\delta_i} S_i(t|\boldsymbol{x}_i) \\ &= \prod_{i=1}^{n} \left[h_0\left(t\right) \exp\left(\boldsymbol{x}_i\boldsymbol{\beta}\right)\right]^{\delta_i} \exp\left[-H_0(t) \exp\left(\boldsymbol{x}_i\boldsymbol{\beta}\right)\right]. \end{aligned} \tag{2.7}$$

The log-likelihood function is

$$\ell(\boldsymbol{\xi}, \boldsymbol{\beta}) = \sum_{i=1}^{n} \delta_i \left[\boldsymbol{x}_i\boldsymbol{\beta} + \log h_0(t)\right] - \sum_{i=1}^{n} H_0(t) \exp(\boldsymbol{x}_i\boldsymbol{\beta}). \tag{2.8}$$

The assumption in the proportional hazard model is that the hazard function for each individual is proportional to the baseline hazard, $h_0(t)$. This implies the hazard ratio is fully determined by the covariate vector. However, there may be unobserved covariates that cause this assumption to be violated hence the need for frailty models as discussed in section 2.2.

2.2 Shared frailty model

The models introduced above (section 2.2) make an homogeneity assumption, i.e., conditioned on covariates, all the individuals in the study population are assumed to have the same risk of experiencing the event of interest (death, remission,relapse, etc.). However, in practice, this homogeneity assumption may not often hold as most study samples are heterogeneous in nature (comprising of individuals who have different risks). When we have clustered data we may estimate a part of the heterogeneity when individuals within a cluster share relevant unobserved covariates.

Clustered survival data are encountered in many disciplines including human and veterinary medicine, biology, epidemiology, public health and demography. Example of studies with clusters: family members, twins, multi-centre clinical trial, parent-child. In such clustered settings, independence between the survival times cannot be assumed.

To model dependence of clustered event times is through the introduction of a cluster-specific random effect - the frailty. Vaupel et al. (1979) introduced the term frailty in order to account for individual unobserved heterogeneity in ageing studies. Here the frailty can be estimated from deviations of the model assumptions such as proportional hazards.

In a clustered setting frailty was introduced by Clayton (1978) (without referring it to "frailty") and extensively studied in Hougaard (2000). In particular, frailty model was used by Hougaard et al. (1992) to measure the similarities between the lifetimes of twins as they are correlated and thus have unobserved shared effects. In general, individuals in a cluster are assumed to share the same frailty, which is why this model is called shared frailty model. The survival times are assumed to be conditional independent with respect to the shared (common) frailty.

A frailty model is a multiplicative hazard model consisting of the random effects (the frailty), a baseline hazard function and a term modelling the influence of observed covariates. The advantage of frailty model over the Cox model is that the hazard ratio incorporates the unobserved (random) effects with the observed (fixed) effects.

Model Description

Suppose that there are G groups of individuals with n_i individuals in the i^{th} group, $i = 1, 2, ..., G$ and $j = 1, 2, ..., n_i$. Let v_i be the cluster frailty, $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_p)^T$ be a parameter vector, $\boldsymbol{x}_{ij}$ is an $1 \times p$ vector, $h_0(t)$ be the baseline (the hazard function of the outcome occurring for subjects with risk vector and frailty equal zero). Let δ_{ij} is the censoring indicator. For the proportional hazards model, the hazard of death at time t for the j^{th} individual in the i^{th} group conditional on v_i , is then given by

$$h_{ij}(t|\boldsymbol{x}_{ij}, v_i) = h_0(t)\, v_i \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})$$

$S_{ij}(t|\boldsymbol{x}_{ij}, v_i)$ is the survival function of an individual conditional on the frailty v_i and risk covariate vector $\boldsymbol{x}_{ij}$ and is given by

$$\begin{aligned} S_{ij}(t|\boldsymbol{x}_{ij}, v_i) &= \exp(-Hij(t|\boldsymbol{x}_{ij}, v_i)) \\ &= \exp\left(\int_0^t -h_0(u)\, v_i \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})\, du\right) \\ &= \exp\left[-H_0(t) v_i \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})\right]. \end{aligned}$$

The marginal survivor function will be

$$\begin{aligned} S_p(t) &= \mathbb{E}\left[S(t|\boldsymbol{x}, v)\right] \\ &= \mathbb{E}\left[\exp(-H_0(t)\exp(\boldsymbol{x}\boldsymbol{\beta})\, v)\right] \\ &= \int_v \exp(-vH_0(t)\exp(\boldsymbol{x}\boldsymbol{\beta}))\, g(v) dv \end{aligned}$$

This is referred to as marginal survivor function (Hougaard, 2000) because it is the observed population survivor function after v has been integrated out. Therefore to derive this function we have to specify a distribution for v say $g(v)$ which has a positive support, mean of 1 and variance θ. The variance θ determines the degree of heterogeneity (variability) in the population. Individuals with $v_i > 1$ are at a higher risk of experiencing the event (for negative events) and vice versa.

Integrating out the frailty reduces the problem to estimating the frailty variance. Therefore the goal is to estimate the frailty variance θ.

Let $\boldsymbol{\xi}$ contain the parameters of the baseline e.g for exponential baseline $\boldsymbol{\xi} = \lambda$ and for a Weibull hazard $\boldsymbol{\xi} = (\lambda, \rho)$. The conditional likelihood for the i^{th} cluster is given by

$$\begin{aligned} L_i(\boldsymbol{\xi}, \boldsymbol{\beta}|v_i) &= \prod_{j=1}^{n_i} \left[h_{ij}(t|\boldsymbol{x}_{ij}, v_i)\right]^{\delta_{ij}} S_{ij}(t|\boldsymbol{x}_{ij}, v_i) \\ &= \prod_{j=1}^{n_i} \left[h_0\left(t\right) v_i \exp\left(\boldsymbol{x}_{ij}\boldsymbol{\beta}\right)\right]^{\delta_{ij}} \exp\left[-H_0(t) \exp\left(\boldsymbol{x}_{ij}\boldsymbol{\beta}\right) v_i\right]. \end{aligned} \tag{2.9}$$

Therefore, the marginal likelihood function of the i^{th} cluster will be

$$L_i(\boldsymbol{\xi}, \theta, \boldsymbol{\beta}) = \int_0^\infty \prod_{j=1}^{n_i} \left[h_0\left(t\right) v \exp\left(\boldsymbol{x}_{ij}\boldsymbol{\beta}\right)\right]^{\delta_{ij}} \exp\left[-H_0(t) \exp\left(\boldsymbol{x}_{ij}\boldsymbol{\beta}\right) v\right] g(v) dv, \tag{2.10}$$

where $g(v)$ is the probability density function of the frailties. If parametric form for $h_0(t)$ is assumed, then maximum likelihood estimates can be obtained by maximizing the marginal likelihood function (Duchateau and Janssen, 2008).

Some common frailty distributions include gamma, log-normal frailty and inverse Gaussian frailty and will be described in section 2.2.1 to section 2.2.4.

2.2.1 Gamma frailty

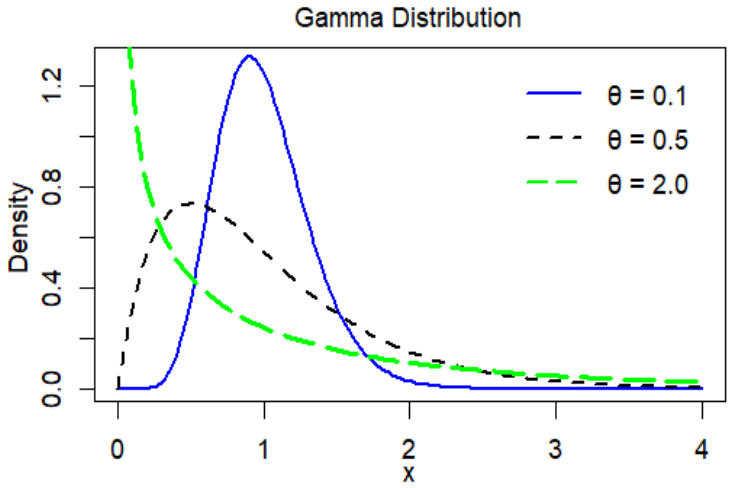
The gamma distribution has been widely applied as a frailty distribution (Congdon, 1995; dos Santos et al., 1995; Hougaard, 2000; Vaupel et al., 1979). It has a closed form likelihood function that can be readily maximized. It is easy to also derive the closed form expressions of unconditional survival, cumulative density and hazard function based on a computational and analytical point of view. This is the main reason why this distribu-

tion has been most commonly used. There is no biological reason that makes the gamma distribution most used as compared to other frailty distributions.

If V follows a gamma distribution with mean 1 and variance θ then V has the density function

$$g(v) = \frac{\left(\frac{1}{\theta}\right)^{1/\theta} v^{1/\theta-1} \exp\left(-v/\theta\right)}{\Gamma\frac{1}{\theta}} \qquad v > 0,\ \theta > 0.$$

Figure 2.3: Probability density functions of different gamma distributions with mean 1 and variances $\theta = 0.1,\ 0.5$ & 2.



The marginal likelihood function of the i^{th} cluster will be

$$\begin{aligned} L_i(\boldsymbol{\xi}, \theta, \boldsymbol{\beta}) &= \int_0^\infty \left\{ \prod_{j=1}^{n_i} \left[h_0\left(t\right) v e^{(\boldsymbol{x}_{ij}\boldsymbol{\beta})}\right]^{\delta_{ij}} e^{[-H_0(t) v \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})]} \right\} \frac{v^{\frac{1}{\theta}-1} \exp\left(-\frac{v}{\theta}\right)}{\theta^{\frac{1}{\theta}}\Gamma\frac{1}{\theta}} dv, \qquad (2.11) \\ &= \left\{ \prod_{j=1}^{n_i} \left[h_0\left(t\right) e^{(\boldsymbol{x}_{ij}\boldsymbol{\beta})}\right]^{\delta_{ij}} \right\} \frac{1}{\theta^{\frac{1}{\theta}}\Gamma\frac{1}{\theta}} \int_0^\infty v^{d_i} e^{\left[-v\sum_{j=1}^{n_i} H_0(t)\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})\right]} v^{\frac{1}{\theta}-1} \exp\left(-\frac{v}{\theta}\right) dv, \end{aligned}$$

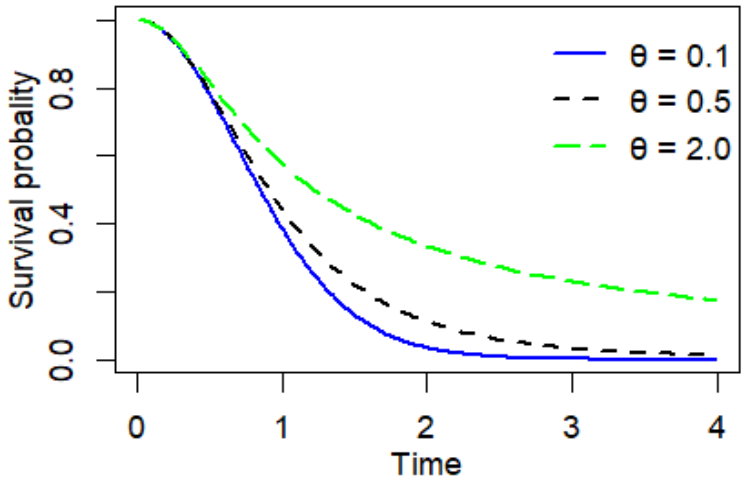
where $d_i = \sum_{j=1}^{n_i} \delta_{ij}$. The marginal loglikelihood function for all the clusters is (Duchateau and Janssen, 2008; Rodríguez-Girondo et al., 2018)

$$\begin{aligned} l\left(\boldsymbol{\xi}, \theta, \beta\right) &= \sum_{i=1}^{G} \log(L_i(\boldsymbol{\xi}, \theta, \boldsymbol{\beta})) \\ &= \sum_{i=1}^{G} \left[d_i \log(\theta) - \log\left(\Gamma\left(1\ /\theta\right)\right) + \log\left(\Gamma\left(1/\theta + d_i\right)\right) - \right. \\ &\quad \left. \left(\frac{1}{\theta} + d_i\right) \log\left(1 + \theta \sum_{j=1}^{n_i} H_0(t) e^{\boldsymbol{x}_{ij}\boldsymbol{\beta}}\right) + \sum_{j=1}^{n_i} \delta_{ij} \left[\boldsymbol{x}_{ij}\boldsymbol{\beta} + \log\left(h_0(t)\right)\right] \right] \qquad (2.12) \end{aligned}$$

The population survival function is given by (Duchateau and Janssen, 2008)

$$S_p(t) \quad = \quad (1 + \theta H_0(t) \exp(\boldsymbol{x}\boldsymbol{\beta}))^{-\frac{1}{\theta}} . \tag{2.13}$$

Figure 2.4: Population survival function i.e. Equation (2.13) for different gamma distributions with mean 1 and variances $\theta = 0.1,\ 0.5$ & 2 assuming no covariate ($\boldsymbol{x} = \mathbf{0}$) and a Weibull baseline hazard with parameters $\lambda = 1$ and $\rho = 2$.



Asymptotic properties of the non-parametric maximum likelihood estimates in the shared gamma frailty model are well established. Murphy shows consistency (Murphy, 1994) and asymptotic normality (Murphy, 1995) in the shared gamma frailty model without covariates.

2.2.2 Lognormal frailty

The density function when V follows a lognormal distribution is

$$g(v) \quad = \quad \frac{1}{v\sqrt{2\pi\sigma^2}} \exp\left(\frac{-[\log(v) - \mu]^2}{2\sigma^2}\right) \qquad v > 0, \quad \sigma^2 > 0.$$

The mean and variance of V are given by

$$\begin{aligned} \mathrm{E}(V) \quad &= \quad \exp\left(\mu + \sigma^2/2\right) \\ \mathrm{Var}(V) \quad &= \quad \theta \quad = \quad \left[\exp(\sigma^2) - 1\right] \exp\left(2 + \sigma^2\right) \end{aligned}$$

The mostly considered model in literature as $W \sim N(\mu, \sigma^2)$, such that $W = \log(V)$. The

hazard model specification is therefore:

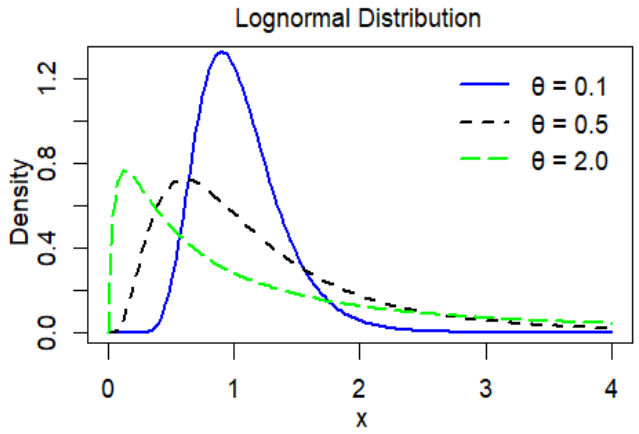
$$h_{ij}\left(t|\boldsymbol{x}_{ij}, w_i\right) = h_0\left(t\right)\exp\left(\boldsymbol{x}_{ij}\boldsymbol{\beta} + w_i\right)$$

In such a transformation to have unity mean and variance θ for V then

$$W \sim N\left(-\frac{\log(1+\theta)}{2}, \log(1+\theta)\right).$$

Let $\mu' = -\frac{\log(1+\theta)}{2}$ and $\sigma' = \sqrt{\log(1+\theta)}$.

Figure 2.5: Probability density functions of different log-normal distributions with mean 1 and and variances $\theta = 0.1,\ 0.5$ & 2.



The marginal likelihood function of the i^{th} cluster will be

$$\begin{aligned} L_i\left(\boldsymbol{\xi}, \theta, \boldsymbol{\beta}\right) &= \int_0^\infty \left\{\prod_{j=1}^{n_i}\left[h_0(t)\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta}+w)\right]^{\delta_{ij}} e^{-H_0(t)\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta}+w)}\right\}\frac{1}{\sigma'\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left[\frac{w-\mu'}{\sigma'}\right]^2\right)dw \\ &= \left\{\prod_{j=1}^{n_i}\left[h_0(t)\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})\right]^{\delta_{ij}}\right\}\frac{1}{\sigma'\sqrt{2\pi}} \\ &\quad \times \int_0^\infty \exp(w)^{d_i} e^{-\left(\sum_{j=1}^{n_i} H_0(t)\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta}+w)\right)}\exp\left(-\frac{1}{2}\left[\frac{w-\mu'}{\sigma'}\right]^2\right)dw \end{aligned}$$

Taking log, we obtain

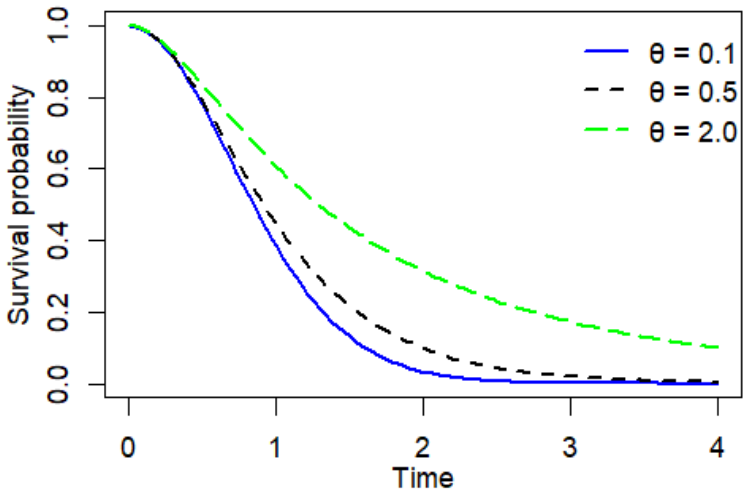
$$\begin{aligned} l_i\left(\boldsymbol{\xi}, \theta, \beta\right) &= \left\{\sum_{j=1}^{n_i}\delta_{ij}\log\left[h_0(t) + \boldsymbol{x}_{ij}\boldsymbol{\beta}\right]\right\} - \log\left(\sigma'\right) - \log(\sqrt{2\pi}) \\ &\quad + \log\left\{\int_0^\infty \exp(w)^{d_i} e^{-\left(\sum_{j=1}^{n_i} H_0(t)\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta}+w)\right)}\exp\left(-\frac{1}{2}\left[\frac{w-\mu'}{\sigma'}\right]^2\right)dw\right\} \end{aligned}$$

Therefore, marginal loglikelihood function for all the clusters will be

$$l\left(\boldsymbol{\xi}, \theta, \beta\right) = \sum_{i=1}^{G}\ell_i\left(\boldsymbol{\xi}, \theta, \boldsymbol{\beta}\right) \tag{2.14}$$

No explicit form of the marginal likelihood exists and thus to obtain the parameter estimates $\boldsymbol{\xi}, \theta, \boldsymbol{\beta}$ numerical integration is required to maximize the marginal loglikelihood functions Equation (2.14) above.

Figure 2.6: Population survival function for different lognormal distributions with mean 1 and variances $\theta = 0.1,\ 0.5$ & 2 assuming no covariate ($\boldsymbol{x} = \mathbf{0}$) and a Weibull baseline hazard with parameters $\lambda = 1$ and $\rho = 2$.



Most used in literature is $W \sim N(0, \sigma^2)$ (This is what has been used in R package *parfm* (Munda et al., 2012)).

2.2.3 Inverse Gaussian frailty

The inverse Gaussian (inverse normal) distribution was introduced as an alternative to the gamma distribution by Hougaard (1984).

If V follows a inverse Gaussian distribution with parameters μ and θ_0 then V has the density function

$$g(v) = \left[\frac{1}{2\pi\theta_0}\right]^{-1/2} v^{-3/2} \exp\left(-\frac{(v-\mu)^2}{2v\theta_0\mu^2}\right) \qquad v > 0,\ \theta_0 > 0.$$

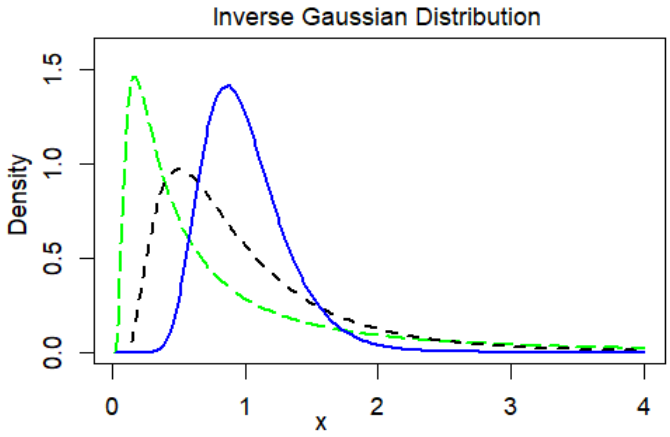
The mean and variance of V are given by

$$\begin{aligned} \mathrm{E}(V) &= \mu \\ \mathrm{Var}(V) &= \theta = \mu^3\theta_0 \end{aligned}$$

For mean= 1 and variance of θ then

$$g(v) = \left[\frac{1}{2\pi\theta}\right]^{-1/2} v^{-3/2} \exp\left(-\frac{(v-1)^2}{2v\theta}\right)$$

Figure 2.7: Probability density functions of different inverse Gaussian distributions with mean 1 and variances $\theta = 0.1,\ 0.5$ & 2.



The marginal likelihood function of the i^{th} cluster will be

$$\begin{aligned} L_i\left(\boldsymbol{\xi}, \theta, \beta\right) &= \int_0^{\infty} \prod_{j=1}^{n_i} \left\{h_0(t) v \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})\right\}^{\delta_{ij}} e^{-H_0(t) v \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})} \frac{1}{\sqrt{2\pi\theta}} \frac{1}{v^{3/2}} \exp\left(-\frac{(v-1)^2}{2v\theta}\right) dv \\ &= \prod_{j=1}^{n_i} \left\{h_0(t) \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})\right\}^{\delta_{ij}} \frac{1}{\sqrt{2\pi\theta}} \int_0^{\infty} v^{d_i - \frac{3}{2}} e^{-v\left\{\sum_{j=1}^{n_i} H_0(t) \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})\right\}} e^{-\frac{(v-1)^2}{2v\theta}} dv. \end{aligned}$$

Taking log, we obtain

$$\begin{aligned} l_i\left(\boldsymbol{\xi}, \theta, \beta\right) &= \sum_{j=1}^{n_i} \delta_{ij} \left\{\log\left(h_0(t)\right) + \boldsymbol{x}_{ij}\boldsymbol{\beta}\right\} - \log\left(\sqrt{\theta}\right) - \log(\sqrt{2\pi}) \\ &\quad + \log\left\{\int_0^{\infty} v^{d_i - \frac{3}{2}} e^{-v\left\{\sum_{j=1}^{n_i} H_0(t) \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})\right\}} e^{-\frac{(v-1)^2}{2v\theta}} dv\right\}. \end{aligned} \tag{2.15}$$

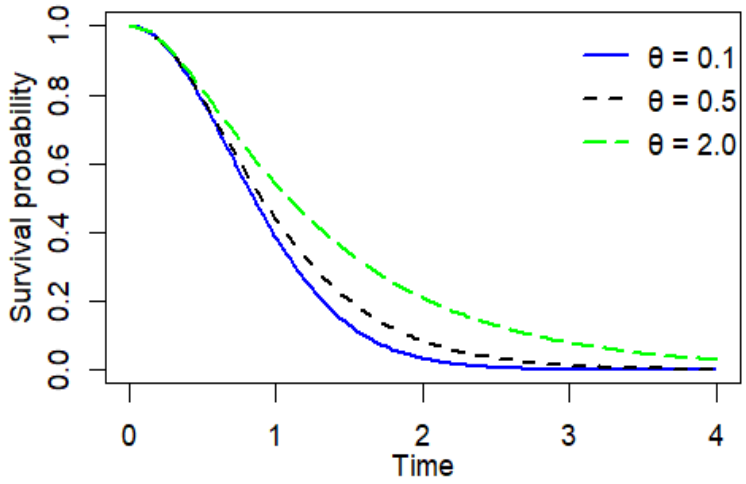
The marginal loglikelihood function for all the clusters is

$$l\left(\boldsymbol{\xi}, \theta, \boldsymbol{\beta}\right) = \sum_{i=1}^{G} l_i\left(\boldsymbol{\xi}, \theta, \boldsymbol{\beta}\right) \tag{2.16}$$

The population survival function is given by (Duchateau and Janssen, 2008)

$$S_p(t) = \exp\left(\frac{1 - \sqrt{1 + 2\theta H_0(t) \exp(\boldsymbol{x}\boldsymbol{\beta})}}{\theta}\right). \tag{2.17}$$

Figure 2.8: Population survival function i.e Equation (2.17) for different inverse Gaussian distributions with mean 1 and variances $\theta = 0.1,\ 0.5$ & 2 assuming no covariate ($\boldsymbol{x} = \mathbf{0}$) and a Weibull baseline hazard with parameters $\lambda = 1$ and $\rho = 2$.



2.2.4 Truncated Gamma frailty

Consider a random variable V following Gamma distribution $g(v;\alpha,\beta)$ with parameters α (shape) and β (rate). Suppose $a < V < b$, then, the truncated gamma distribution will be (Okasha and Alqanoo, 2014)

$$\begin{aligned} g(v;\alpha,\beta|a<V<b) &= \frac{g(v;\alpha,\beta)}{\int_a^b g(v;\alpha,\beta)dv} \\ &= \frac{\beta^\alpha v^{\alpha-1}e^{-v\beta}}{\Gamma(\alpha,a\beta)-\Gamma(\alpha,b\beta)}, \end{aligned} \tag{2.18}$$

where

$$\Gamma(p,q) = \int_q^\infty y^{p-1}e^{-y}dy.$$

The rth moment of the truncated Gamma distribution is given by (Okasha and Alqanoo, 2014)

$$E(V^r) = \frac{\Gamma(\alpha+r,a\beta)-\Gamma(\alpha+r,b\beta)}{\beta^r\left[\Gamma(\alpha,a\beta)-\Gamma(\alpha,b\beta)\right]}. \tag{2.19}$$

So, the mean and variance of a left truncated Gamma distribution ($a < V < \infty$) will be given by

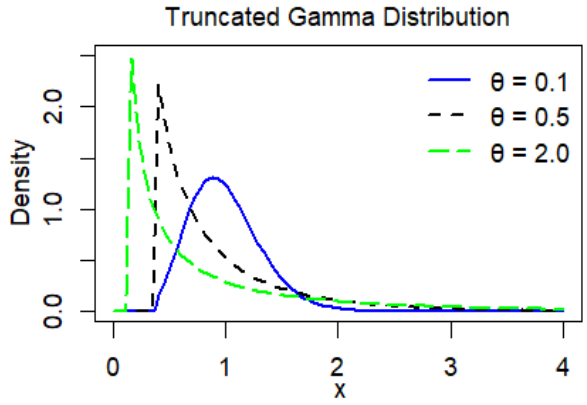
$$E(V) = \frac{\Gamma(\alpha+1, a\beta)}{\beta\Gamma(\alpha, a\beta)}, \tag{2.20}$$

$$E\left(V^2\right) = \frac{\Gamma(\alpha+2, a\beta)}{\beta^2\Gamma(\alpha, a\beta)},$$

$$\text{Var}(V) = E\left(V^2\right) - [E(V)]^2. \tag{2.21}$$

In order to satisfy the identifiability assumption of a frailty distribution, the mean is set to unity and the variance is set to a preset value of θ, these two equations are solved numerically to obtain α' and β', where α' and β' denote the values of α and β that satisfy the constraints (i.e. Equation (2.20) $= 1$, and Equation (2.21)$= \theta$).

Figure 2.9: Probability density functions of different left truncated gamma distributions with mean 1 and $a = 0.4$ for $\theta = 0.1, 0.5$ while $a = 0.15$ for $\theta = 2$.

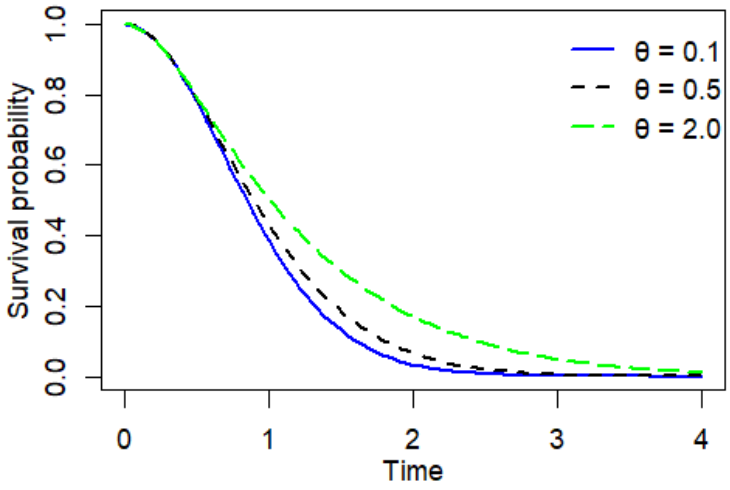


The choice for a above is user-defined. The population survival function given left truncated gamma frailty is

$$S_p(t) := \int_a^\infty \exp\left(-vH_0(t)\exp(\boldsymbol{x\beta})\right) g(v;\alpha,\beta|a<V<\infty)dv$$

$$= \int_a^\infty \exp\left(-vH_0(t)\exp(\boldsymbol{x\beta})\right) \frac{\beta^\alpha v^{\alpha-1}e^{-v\beta}}{\Gamma(\alpha, a\beta)}dv. \tag{2.22}$$

Figure 2.10: Population survival function i.e Equation (2.22) corresponding to left truncated distributions given above (Figure 2.9) assuming no covariate ($\boldsymbol{x} = \mathbf{0}$) and a Weibull baseline hazard with parameters $\lambda = 1$ and $\rho = 2$.



Note that different frailty distributions yield different survival probabilities and this is more evident for larger values of θ.

This section assumes time independent covariates. Incorporating time-dependent covariates in to the model is discussed in the next chapter (Chapter 3)).

2.2.5 Generating survival time

We generate survival time for use in simulation studies to assess the performance of current or novel statistical models. T (defined in section 2.2) cannot be generated directly but can be generated by inverting the cumulative hazard function as follows. Let Y be a random variable with distribution F, then $U = F(Y)$ follows a uniform distribution on interval 0 to 1 i.e $U \sim \text{Uni}[0,1]$. It follows also that $(1-U) \sim \text{Uni}[0,1]$. Let T be the survival time of the Cox model then it follows that

$$\begin{aligned} U &= \exp\left[-H_0(t)v_i \exp\left(\boldsymbol{x}_i\boldsymbol{\beta}\right)\right] \sim \text{Uni}[0,1] \\ \log(U) &= -H_0(t)v_i \exp\left(\boldsymbol{x}_{ij}\boldsymbol{\beta}\right) \\ -\log(U)v_i \exp\left(-\boldsymbol{x}_{ij}\boldsymbol{\beta}\right) &= H_0(t). \end{aligned}$$

If $h_0(t) > 0$ for all t, then H_0 can be inverted and the survival time T can be generated by (Bender et al., 2005).

$$T = H_0^{-1}\left[-\log(U)v_i \exp\left(-\boldsymbol{x}_{ij}\boldsymbol{\beta}\right)\right], \tag{2.23}$$

where $U \sim U(0,1)$ i.e the standard uniform distribution v_i follows distribution $g(v_i)$ with mean of 1 and variance θ..

Commonly used distributions for baseline hazard are exponential, the Weibull, and the Gompertz distributions with their characterization given in Table 2.1 below. This proce-

Table 2.1: Characterization of the exponential, Weibull, and Gompertz distributions in simulating survival time.

Characteristic	Exponential distribution	Weibull distribution	Gompertz distribution
Parameter	Scale $\lambda > 0$	Scale $\lambda > 0$ Shape $\rho > 0$	Scale $\lambda > 0$ Shape $-\infty < \eta < \infty$
Hazard function	$h_0(t) = \lambda$	$h_0(t) = \lambda\rho t^{\rho-1}$	$h_0(t) = \lambda \exp(\eta t)$
Cumulative hazard	$H_0(t) = \lambda t$	$H_0(t) = \lambda t^{\rho}$	$H_0(t) = \frac{\lambda}{\eta}(\exp(\eta t) - 1)$
Inverse cumhaz	$H_0^{-1}(t) = \frac{1}{\lambda}t$	$H_0^{-1}(t) = \left(\frac{1}{\lambda}t\right)^{1/\rho}$	$H_0^{-1}(t) = \frac{1}{\eta}\log\left(\frac{\eta}{\lambda}t + 1\right)$
Survival time	$T = -\frac{\log(U)}{\lambda v_i \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})}$	$T = \left(-\frac{\log(U)}{\lambda v_i \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})}\right)^{1/\rho}$	$T = \frac{1}{\eta}\log\left(1 - \frac{\eta \log(U)}{\lambda v_i \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})}\right)$

dure will be used for data generation in the simulations performed in chapter 6.

2.3 Software

Table 2.2: Overview of existing R packages for shared frailty model

Function	Package	Frailty distribution	Baseline hazard distribution	Left-truncation/ Time varying covariates
`parfm`	`parfm`	Log-normal, Positive stable, Inverse Gaussian, Gamma	Exponential, Weibull, inverse Weibull, Gompertz, lognormal,log-skew-normal, and loglogistic	Left truncation
`frailtyPenal`	`frailtypack`	Gamma and Log-normal	Piecewise constant, Cubic M-splines, Weibull, splines	Fits left truncation or time varying separately.
`coxph`	`survival`	Gamma, Log-normal, log-t	Non-parametric	Time varying covariates
`coxme`	`coxme`	Normal	Non-parametric	Time varying covariates
`frailtyEM`	`frailtyEM`	Gamma, Inverse Gaussian, Positive stable, PVF-family	Breslow-type non-parametric baseline hazard	Left truncation
`fitfrail`	`frailtySurv`	Gamma, Power variance function, Log-normal, and Inverse Gaussian frailty model	Breslow-type non-parametric baseline hazard	
`frailtyHL`	`frailtyHL`	Gamma and lognormal	Non-parametric	
`phmm`	`phmm`	Log-normal	Non-parametric	
`survBayes`	`survBayes`	Gamma and Log-normal	Non-parametric	
`sscox`	`gss`	Log-normal	Non-parametric	

Chapter 3

Modelling time varying covariates

Time varying covariate (also called time-dependent covariate) occurs when a given covariate changes over time during the period under study. In practice, many studies that generate time-to-event data, record other variables whose value vary in time and it is often of interest to investigate the relationship between such variables and the time-to-event outcome. Time varying covariates, in particular, provide us with important information on how changes in a subject's history effect survival.

Time varying covariates can be categorical or continuous. An example of a binary variable is transplantation in the Stanford heart transplant (Crowley and Hu, 1977). Patients are admitted to a waiting list, the time to event is time from admittance to the waiting list until death, which may be subject to censoring, and the interest is in the effect of a heart transplant on survival. The time varying covariate heart transplant is initially equal to 0, and attains the value 1 when the patient is given a heart transplant. If the patient never receives a heart transplant, the value remains 0 throughout his/her follow-up. In this scenario, the time varying covariate is fully observed. Example of a continuous time varying covariate include BMI when investigating it's effect on mortality (Kovesdy et al., 2007) and CD4+ T-cell counts on the occurrence of AIDS or death for HIV-infected patients (Tsiatis et al., 1995). In these cases, one wants to make use of the longitudinal covariate information that is available.

Time varying covariates can be either external and internal (Collett, 2015; Kalbfleisch and Prentice, 2011). It is classified as internal when the path is affected by survival status. An internal covariate is typically the output of a stochastic process generated by an individual

under study and observed only as long as the subject survives and uncensored. Examples would include the blood pressures and CD4 counts measured over the course of the study. An external covariate does not necessarily require the subject's survival for their existence, they remain measurable and their distributions unchanged after the occurrence of the event. Examples of external covariates are BMD as a predictor of fractures, or a measure of airborne pollution as a predictor of the frequency of asthma attacks. For internal time-varying covariates other alternatives like joint models may be considered.

Suppose that there are n individuals, $i = 1, 2, ..., n$. Let T_i be the random variable representing time, δ_i is the event indicator, and $\boldsymbol{Y}_i = \{Y_{il}(t_{il}); l = 1, \ldots, n_i\}$ be a time varying covariate. Let α be the time varying covariate effect. Let $\boldsymbol{\xi}$ be the parameters of the baseline hazard $h_0(t)$ and θ be the frailty variance. The hazard at time t for the i^{th} individual is given by

$$h_i(t|\boldsymbol{Y}_i) = h_0(t) \exp(\alpha y_i(t)).$$

The risk parameter α represents the effect on the hazard of a unit difference in the covariate at time zero or at any time after entry under the assumption that the effect of the covariate is time invariant.

In this approach the hazard at time t is assumed to depend on the current value at time t of the time varying covariate $y_i(t)$, through the product of the baseline hazard and $\exp(\alpha y_i(t))$. This approach yields valid inference if the value of the time varying covariate is known for all subjects at all event time points without error, and the regression model is correctly specified. However, in most cases the time varying covariate is measured only intermittently. Since time varying observations are only available at the time of measurements, a simple and frequently used estimator is to impute the last observed covariate value (Therneau et al., 2017) as discussed in section 3.1. Other methods of modelling a time varying covariate discussed in this chapter are two stage approaches as discussed in section 3.2 and joint models as discussed in section 3.3.

3.1 Last observation carried forward

Suppose the varying covariate $y_i(t)$ takes the value y_{il} within a given time interval $(t_{il}, t_{i(l+1)})$ at time points $l = 0, ..., (n_i - 1)$. The contribution of the i^{th} individual within a given time interval $(t_{il}, t_{i(l+1)})$ is given by (Klein and Moeschberger, 2003)

$f(t_{i(l+1)}|y_{il})^{\delta_{il}} S(t_{i(l+1)}|y_{il})^{1-\delta_{il}} / S(t_{il}|y_{il})$.

i.e for an individual to contribute to the likelihood within the interval $(t_{il}, t_{i(l+1)})$ an individual needs to have survived until time t_{il}.

Therefore the full likelihood is given by

$$\begin{aligned}
L &= \prod_{i=1}^{n} \prod_{l=0}^{n_i-1} \frac{f(t_{i(l+1)}|y_{il})^{\delta_{il}} S(t_{i(l+1)}|y_{il})^{1-\delta_{il}}}{S(t_{il}|y_{il},)} \\
&= \prod_{i=1}^{n} \prod_{l=0}^{n_i-1} \frac{h(t_{i(l+1)}|y_{il})^{\delta_{il}} S(t_{i(l+1)}|y_{il})^{\delta_{il}} S(t_{i(l+1)}|y_{il})^{1-\delta_{il}}}{S(t_{il}|y_{il})} \\
&= \prod_{i=1}^{n} \prod_{l=0}^{n_i-1} \frac{h(t_{i(l+1)}|y_{il})^{\delta_{il}} S(t_{i(l+1)}|y_{il})}{S(t_{il}|y_{il})} \\
&= \prod_{i=1}^{n} \prod_{l=0}^{n_i-1} \frac{h(t_{i(l+1)}|y_{il})^{\delta_{il}} \exp(-H(t_{i(l+1)}|y_{il}))}{\exp(-H(t_{il}|y_{il}))} \\
&= \prod_{i=1}^{n} \prod_{l=0}^{n_i-1} h(t_{i(l+1)}|y_{il})^{\delta_{il}} \exp(-[H(t_{i(l+1)}|y_{il}) - H(t_{il}|y_{il})]) \qquad (3.1) \\
&= \prod_{i=1}^{n} \prod_{l=0}^{n_i-1} [h_0(t_{i(l+1)}) \exp(\alpha y_{il})]^{\delta_{il}} \exp(-[H_0(t_{i(l+1)}) - H_0(t_{il})] e^{\alpha y_{il}}) \qquad (3.2)
\end{aligned}$$

The log-likelihood function is

$$\ell(\boldsymbol{\xi}, \alpha) = \sum_{i=1}^{n} \sum_{l=0}^{n_i-1} \delta_{il} \left[\alpha y_{il} + \log h_0(t_{i(l+1)})\right] - \sum_{i=1}^{n_i} \sum_{l=0}^{n_i-1} [H_0(t_{i(l+1)}) - H_0(t_{il})] \exp(\alpha y_{il}) \quad (3.3)$$

The maximum likelihood estimates of $\boldsymbol{\xi}, \boldsymbol{\gamma}$ are obtained by maximizing the loglikelihood function given in equation (3.3).

This method does not account for measurement error in the covariate value. This causes the estimated relative risk parameter to be biased toward the null (i.e underestimated), with the extent of this bias directly proportional to the amount of measurement error in the observed covariate (Prentice, 1982; Raboud et al., 1993).

This carrying forward of the longitudinal measurements between examination times and assuming that a longitudinal covariate does not change from the time of last measurement is a further limitation of this model, especially if measurements are sparse or taken a considerable time before failure. The nearest preceding marker value may not be biologically plausible if the covariates change a lot over time, such as CD4 counts for HIV patients. To weaken the assumption of longitudinal covariate remaining constant between measurements two stage approach methods were introduced and will be described in next section (Section 3.2).

3.2 A two stage approach

A two-stage or regression calibration approach models the time-varying covariate process using a linear mixed model in the first stage to obtain the MLE of the fixed-effect parameters and best linear unbiased predictors (BLUPs) of the random effects. Then in a second stage, plugs the predicted covariate into the survival model (Bycott and Taylor, 1998; Dafni and Tsiatis, 1998; Self and Pawitan, 1992; Sweeting and Thompson, 2011; Tsiatis et al., 1995; Ye et al., 2008)

Since the predicted covariate value can be evaluated continuously throughout time, the data set can be split into fine time-intervals, so that the assumption of constant longitudinal measurements between examination times is weakened (Sweeting and Thompson, 2011). Note that we split data into fine-time intervals when we consider a parametric baseline hazard but for a Cox model, the value of the covariate is only required at the event times, and thus there is no need to split the data further.

We consider two different two regression calibration methods: risk-set regression calibration and the ordinary regression calibration.

3.2.1 Risk set regression calibration (RRC)

Stage I

Multiple mixed-effects models are fit to the dataset, one for each of the unique event times in the dataset. For an event time T_r only individuals still at risk i.e $\{i : T_i \geq T_r\}$ are

included in estimation in the mixed model, and only the covariate measurements of these individuals taken before the event time are used $Y_i(T_r) = \{Y_i(t_{il}) : t_{il} < T_r\}$. The predicted value of the covariate for individuals still at risk at event time T_r is then computed as follows: $Y_i^*(T_r) = \mathbb{E}\{Y_i(T_r)|Y_i(t_{il}) : t_{il} < T_r\}, \hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_b^2\}$

3.2.1.1 Linear mixed model specification

In this thesis we consider the following random intercept model

$$y_i(t \mid b_i) = \beta_0 + b_i + \beta_1 t + \varepsilon_i(t)$$

$b_i \sim N(0, \sigma_b^2)$, where b_i individual specific random effect

$\boldsymbol{\varepsilon_i} \sim N(0, \sigma_\varepsilon^2)$, and $\boldsymbol{\varepsilon_i} \perp \boldsymbol{\varepsilon_{i'}}$, $\forall i \neq i'$.

β_0, β_1 : the regression coefficients.They represent the population intercept and slope respectively.

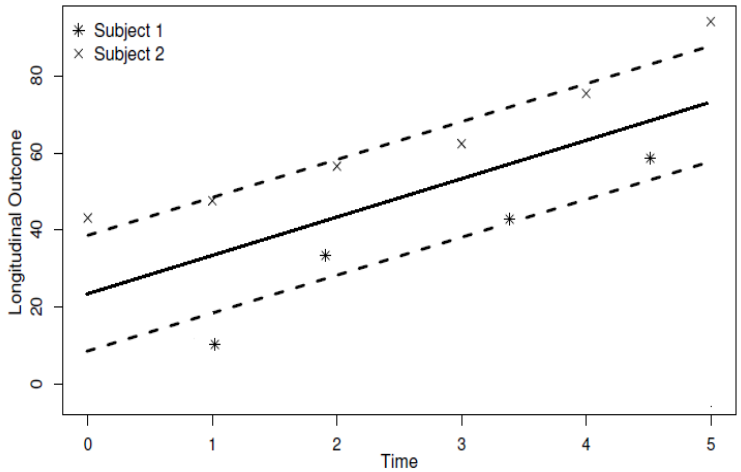
$y_{il} \perp y_{il'} | b_i$

$y_i \perp y_{i'}$ $\forall i \neq i'$

b_i and $\boldsymbol{\varepsilon_i}$ are independent of each other

The random effects in the linear mixed model the correlation from repeated measures of the outcome variable across time.

Figure 3.1: The points represent hypothetical longitudinal responses of two subjects in a longitudinal study. The dashed lines represent the subject-specific longitudinal evolutions. The solid line represents the population-averaged evolution.



The population mean trajectory is represented by $\beta_0 + \beta_1 t$. The random effects b_i are subject-specific deviations with respect to the population intercept.

This is a two-level linear mixed model and could be represented in matrix form as follows:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Zb} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{X}$ is the $m \times 2$ design matrix for the fixed effects and $\boldsymbol{Z}$ is $m \times 1$ design matrix associated to the random effects (b_i); m is the total number of repeated measurements for all subjects. Let $\boldsymbol{\Sigma_b} = \text{var}(\boldsymbol{b}) = \sigma_b^2 \mathbb{1}_n$, $\text{var}(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2 \mathbb{1}_m$ and $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 & \beta_1 \end{bmatrix}$. $\mathbb{1}_m$ is the identity matrix.

3.2.1.2 Linear Mixed Model Estimation

In this model, the fixed effects are directly estimated from the data. In contrast, the random variables b_i are not estimated directly, but rather the parameter of variance σ_b^2. The complete set of parameters to estimate in this model are $\boldsymbol{\Phi} = (\beta_0, \beta_1, \sigma_b^2, \sigma_\varepsilon^2)$. Maximum likelihood estimates of $\boldsymbol{\Phi}$ are obtained by maximizing the log-likelihood function.

$$\ell(\boldsymbol{\Phi}) = \sum_{i=1}^{n} \log\Big\{ \int_b f(\mathbf{y}_i \mid b_i) f(b_i) \mathrm{d}b_i \Big\}, \tag{3.4}$$

where

$$f\{y_{il} \mid b_i\} = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\Big(- \frac{[y_{il} - (\beta_0 + b_i + \beta_1 t_{il})]^2}{2\sigma_\varepsilon^2} \Big)$$

$$f(\mathbf{y}_i \mid b_i) = \prod_{l=1}^{n_i} f(y_{il} \mid b_i) = (\sqrt{2\pi\sigma_\varepsilon^2})^{-n_i/2} \exp\Big(- \frac{\sum_{l=1}^{n_i} [y_{il} - (\beta_0 + b_i + \beta_1 t_{il})]^2}{2\sigma_\varepsilon^2} \Big)$$

$$f(b_i) = \frac{1}{\sqrt{2\pi\sigma_b^2}} \exp\Big(- \frac{[b_i]^2}{2\sigma_b^2} \Big)$$

Estimation of $\boldsymbol{\Phi}$ is done iteratively by splitting into the parameters of the fixed effect (β_0, β_1), and the variance parameters $(\sigma_b^2, \sigma_\varepsilon^2)$.

A consequence of the normal distribution and independence assumptions of the random effects and the measurement error vectors is that, the m response vectors are conditionally independent and normally distributed given the random effects i.e $\boldsymbol{Y} \sim \mathcal{N}\left(\boldsymbol{X\beta}, \boldsymbol{Z\Sigma_b Z}^\top + \sigma_\varepsilon^2 \mathbb{1}_m\right)$ while $\boldsymbol{Y}|\boldsymbol{b} \sim \mathcal{N}\left(\boldsymbol{X\beta} + \boldsymbol{Zb}, \sigma_\varepsilon^2 \mathbb{1}_m\right)$.

Let $\boldsymbol{\Sigma_y} = \text{var}(\boldsymbol{Y}) = \boldsymbol{Z}\boldsymbol{\Sigma_b}\boldsymbol{Z}^\top + \sigma_\varepsilon^2 \mathbb{1}_n$.

If $\boldsymbol{\Sigma_y}$ is assumed to be known, then the maximum likelihood estimator of the fixed-effects $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^\top \boldsymbol{\Sigma_y^{-1}} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{\Sigma_y^{-1}} \boldsymbol{Y} \tag{3.5}$$

When $\boldsymbol{\Sigma_y}$ is not known then $\boldsymbol{\Sigma_y}$ in Equation (3.5) is replaced with $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{y}}$. To obtain this estimate $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{y}}$, we employ the maximum likelihood method, and maximize the log-likelihood $\ell(\sigma_b^2, \sigma_\varepsilon^2)$ for a given value of $\boldsymbol{\beta}$. An estimate of the variance parameters can be obtained by replacing $\hat{\boldsymbol{\beta}}$ in the log-likelihood function given in Equation (3.4) and maximizing $\ell(\sigma_b^2, \sigma_\varepsilon^2 | \boldsymbol{\beta} = \hat{\boldsymbol{\beta}})$ then iterative procedures like Newton–Raphson can be used to obtain $(\sigma_b^2, \sigma_\varepsilon^2)$ estimates.

By asymptotic maximum likelihood theory and under certain regularity conditions the maximum likelihood estimate of $\boldsymbol{\Sigma_y}$ will be asymptotically unbiased. However, in cases of small samples, the maximum likelihood estimate of $\boldsymbol{\Sigma_y}$ will be biased because it does not take into account the fact that $\boldsymbol{\beta}$ is estimated from the data as well. To address this problem the theory of restricted maximum likelihood (REML) estimation was developed (Harville, 1977; Patterson and Thompson, 1971). The main idea behind REML estimation is to separate the part of the data used in the estimation of $\boldsymbol{\Sigma_y}$ from the part used for the estimation of $\boldsymbol{\beta}$. Rather than maximizing the log-likelihood given in Equation (3.5), REML maximizes the modified log-likelihood function given by

$$\ell_{\text{REML}}(\sigma_b^2, \sigma_\varepsilon^2) = \ell(\sigma_b^2, \sigma_\varepsilon^2 | \boldsymbol{\beta} = \hat{\boldsymbol{\beta}}) - \frac{1}{2} \log |\boldsymbol{X}^\top \boldsymbol{\Sigma_y^{-1}} \boldsymbol{X}| \tag{3.6}$$

where $|\boldsymbol{X}^\top \boldsymbol{\Sigma_y^{-1}} \boldsymbol{X}|$ denotes the determinant of the square matrix $\boldsymbol{X}^\top \boldsymbol{\Sigma_y^{-1}} \boldsymbol{X}$.

The estimate $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{y}}$ obtained by maximizing $\ell_{\text{REML}}(\sigma_b^2, \sigma_\varepsilon^2)$ corrects for the fact that $\boldsymbol{\beta}$ has also been estimated.

The random effects can be predicted by obtaining the expectation of the posterior distribution of the random effects given the observed data (Fitzmaurice et al., 2012; Laird and Ware, 1982).

$$\begin{aligned} \mathbb{E}(\boldsymbol{b}|\boldsymbol{Y}) &= \mathbb{E}(\boldsymbol{b}) + \text{cov}(\boldsymbol{b}; \boldsymbol{Y}) \, \text{var}(\boldsymbol{Y})^{-1} [\boldsymbol{Y} - \mathbb{E}\boldsymbol{Y}] \\ &= \boldsymbol{0} + \boldsymbol{\Sigma_b} \boldsymbol{Z}^\top {\boldsymbol{\Sigma_y}}^{-1} [\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}] \end{aligned} \tag{3.7}$$

The parameters $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma_y}$ are replaced by their estimators in Equation (3.7) to obtain the empirical Bayes estimator, $\hat{\boldsymbol{b}} = \boldsymbol{\Sigma_b Z}^\top \hat{\boldsymbol{\Sigma}}_{\boldsymbol{y}}^{-1}[\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}]$

Stage II

Let $Y_i(t) = Y_i^*(t) + \varepsilon_i(t)$. From the non-differential measurement error assumption (Carroll et al., 2006) we have $h_i\left(t|\boldsymbol{Y}_i^*, \boldsymbol{Y}_i\right) = h_i\left(t|\boldsymbol{Y}_i^*\right)$. Non-differential measurement error is an error that is independent of outcome status (Fosgate, 2006).

The hazard at time t, based on the history of the observed covariate measurements $\boldsymbol{Y}_i$ upto time t can be expressed as the conditional expectation as follows (Prentice, 1982)

$$\begin{aligned} h_i\left(t|\boldsymbol{Y}_i\right) &= \mathbb{E}\{h_i\left(t|\boldsymbol{Y}_i^*, \boldsymbol{Y}_i, t_i \geq t\right)\} \\ &= \mathbb{E}\{h_i\left(t|\boldsymbol{Y}_i^*, t_i \geq t\right)\} \text{ from the nondifferential measurement error assumption} \end{aligned}$$

As a result the induced hazard function based on observed covariate measurements is given by:

$$h_i\left(t|\boldsymbol{Y}_i\right) = h_0(t)\mathbb{E}(\exp[\alpha Y_i^*(t)]|\boldsymbol{Y}_i, t_i \geq t) \tag{3.8}$$

Yu et al. (2018) and Dafni and Tsiatis (1998) propose to approximate this expectation by a first-order approximation i.e the regression calibration approximation. This approximation holds when predicted values are less variable.

Let $g(Y) = \mathbb{E}(Y_i^*(t)|\boldsymbol{Y}_i, t_i \geq t)$ and $\mathbb{E}Y = \mu$ then by a first-order approximation we have

$$g(Y) \approx g(\mu) + (Y - \mu)g'(\mu) \tag{3.9}$$

$$\mathbb{E}[g(Y)] = \mathbb{E}[g(\mu)] + \mathbb{E}(Y - \mu)g'(\mu) \qquad \text{but } \mathbb{E}(Y - \mu)g'(\mu) = 0 \text{ since } g'(\mu) \text{ is a constant}$$

$$\mathbb{E}[g(Y)] = g(\mu)$$

Therefore

$$\mathbb{E}(\exp[\alpha Y_i^*(t)]|\boldsymbol{Y}_i, t_i \geq t) \approx \exp[\alpha\mathbb{E}(Y_i(t)|\boldsymbol{Y}_i, t_i \geq t)] \tag{3.10}$$

Replacing this approximation (Equation (3.10)) into Equation (3.8) yields;

$$\begin{aligned} h_i\left(t|\boldsymbol{Y}_i\right) &= h_0(t)\mathbb{E}\{\exp[\alpha Y_i^*(t)]|\boldsymbol{Y}_i, t_i \geq t\} \\ &\approx h_0(t)\exp(\alpha\mathbb{E}[Y_i(t)|\boldsymbol{Y}_i, t_i \geq t]) \\ &= h_0(t)\exp(\alpha Y_i^*(t)) \end{aligned} \tag{3.11}$$

Note that $Y_i^*(t) = \mathbb{E}[Y_i(t)|\boldsymbol{Y}_i, t_i \geq t]$ is what was obtained in stage I.

The last observation carried forward full likelihood approach (equation (3.2) of section 3.1) is again used but now with covariate predictions i.e.

$$L(\boldsymbol{\xi}, \alpha) = \prod_{i=1}^{n} \prod_{l=0}^{n_i-1} [h_0(t_{i(l+1)}) \exp(\alpha y_{il}^*)]^{\delta_{il}} \exp\{-[H_0(t_{i(l+1)}) - H_0(t_{il})] \exp(\alpha y_{il}^*)\}$$

The log-likelihood function is then

$$\ell(\boldsymbol{\xi}, \alpha) = \sum_{i=1}^{n} \sum_{l=0}^{n_i-1} \delta_{il} \left[\alpha y_{il}^* + \log h_0(t_{i(l+1)})\right] - \sum_{i=1}^{n} \sum_{l=0}^{n_i-1} [H_0(t_{i(l+1)}) - H_0(t_{il})] \exp(\alpha y_{il}^*) \quad (3.12)$$

The maximum likelihood estimates of $\boldsymbol{\xi}, \boldsymbol{\gamma}$ are estimated by maximizing the loglikelihood function given in equation (3.12).

3.2.2 Ordinary regression calibration (ORC)

Ordinary regression calibration as proposed by Tsiatis et al. (1995) fits a single mixed-effects model using all individuals and data in the first stage. The predicted value of the covariate at any time t for individuals are computed as follows: $Y_i^*(t) = \mathbb{E}\{Y_i(t)|Y_i(t_{il}), \hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_b^2, \hat{\sigma}_u^2\}$ Stage II is obtained similarly to the RRC procedure discussed in section 3.2.1.

In the RRC procedure, $Y_i^*(t)$ are obtained through extrapolation in time in each risk set, whereas in the ORC procedure interpolation is used to predict these covariate values for most of the time points (i.e. ORC procedure uses future data to calculate $Y_i^*(t)$ for each t). According to Ye et al. (2008) when withdrawal from the risk set is related to the covariate, subjects at higher risk tend to have fewer covariate measurements. When the failure of the event strongly depends on the longitudinal covariate of using the ORC can be biased due to the informative dropout, and the corresponding $Y_i^*(t)$ would be biased. In contrast, the RRC estimator can accommodate this change of risk set over time, and the best linear unbiased predictors of the random effects estimates using the RRC procedure tend to have less bias than those in the ORC procedure. However, in a finite sample situation, the risk-set regression calibration procedure is subject to a decreasing number of subjects in risk set over time. Hence for large t, RRC-predicted values may be more variable thus may have a large error.

The two stage approaches do not incorporate the uncertainty of estimation in the first stage into the second stage, possibly leading to under-estimation of the standard errors (Sweeting and Thompson, 2011). Furthermore, the form of the best linear unbiased predictors of the random effects in stage I depends critically on the validity of normally distributed random effects and error terms, an assumption which becomes less satisfactory as time increases and subjects suffer informative drop-out (Tsiatis and Davidian, 2004). Two stage approaches have less bias as compared to last observation carried forward but may still have biased estimators when the measurement error is large (Tsiatis and Davidian, 2001).

3.3 Joint model approach

To circumvent the limitation of two stage approaches not incorporating the uncertainty of estimation of the first stage into the second stage, a joint model was introduced. In general the joint model consists of a survival submodel and a longitudinal submodel (Crowther et al., 2013; Henderson et al., 2000; Rizopoulos, 2008; Wulfsohn and Tsiatis, 1997). Let y_{il} denote the value of the l^{th} repeated measure of the longitudinal outcome for subject i taken at time point $t_{il},\ i = 1, \ldots, n_i,\ l = 1, \ldots, n$. Hence the vector of observed repeated measures of the longitudinal outcome for subject i consists of $\mathbf{y}_i = \{y_{il}; l = 1, \ldots, n_i\}$. Let the longitudinal model have random intercepts only and assume a link function of current value i.e $\beta_0 + b_i + \beta_1 t$. Considered in this chapter and throughout the thesis is a random intercept model. Let b_i be the individual specific random effect.

The standard joint model for a longitudinal and a time-to-event outcome is as follows:

$$\underset{\text{(Longitudinal)}}{y_i(t \mid b_i)} = \beta_0 + b_i + \beta_1 t + \varepsilon_i(t) \tag{3.13}$$

$$\underset{\text{(Survival)}}{h_i(t \mid b_i)} = h_0(t) \exp\{\alpha(\beta_0 + b_i + \beta_1 t)\} \tag{3.14}$$

We assume that measurement error and random effects are independent, $\varepsilon_i(t) \perp b_i$, and that the repeated measures and the time-to-event outcomes are conditionally independent given the random effects, $\mathbf{y}_i \perp \{T_i, \delta_i\} \mid \mathbf{b}_i$.

The following need to be estimated in the joint model formulation:

$$\boldsymbol{\Phi} = \left(\boldsymbol{\xi}, \boldsymbol{\beta}, \alpha, \sigma_\varepsilon^2, \sigma_b^2\right).$$

where

$\boldsymbol{\xi}$: the baseline hazard parameters

$\boldsymbol{\beta}$: Linear mixed model regression coefficients

α : the regression coefficient for association of time-varying covariate and hazard model

σ_ε^2 : is a constant variance for the error term.

σ_b^2 : is variance of the random effects.

Suppose that data, $\mathcal{D} = \{t_i, \delta_i, \mathbf{y}_i;\ i = 1, \ldots, n\}$, of both longitudinal and time-to-event outcomes are collected on subjects $i = 1, \ldots, n$. Maximum likelihood estimates of $\boldsymbol{\Phi}$ are obtained by maximizing the log-likelihood function of the joint distribution of the longitudinal and the time-to-event outcomes, $\{\mathbf{y}_i, t_i, \delta_i\}$:

$$
\begin{aligned}
\ell(\boldsymbol{\Phi}|\mathcal{D}) &= \sum_{i=1}^{n} \log\left(\int_b f(\mathbf{y}_i \mid b_i) f(t_i, \delta_i \mid b_i) f(b_i) \mathrm{d}b_i\right), \text{ where} \qquad (3.15)\\
f(\mathbf{y}_i \mid b_i) &= \prod_{l=1}^{n_i} f(y_{il} \mid b_i) = (\sqrt{2\pi\sigma_\varepsilon^2})^{-n_i/2} \exp\Big(-\frac{\sum_{l=1}^{n_i}[y_{il} - (\beta_0 + b_i + \beta_1 t_{il})]^2}{2\sigma_\varepsilon^2}\Big)\\
f(b_i) &= \frac{1}{\sqrt{2\pi\sigma_b^2}} \exp\Big(-\frac{[b_i]^2}{2\sigma_b^2}\Big)\\
f(t_i, \delta_i \mid b_i) &= \left[h_i\left(t_i \mid Y_i(t_i)\right)\right]^{\delta_i} S_i\left(t_i \mid Y_i(t_i)\right),\\
h_i\left(t_i \mid \boldsymbol{Y}_i\right) &= h_0(t_i) \exp\{\alpha(\beta_0 + b_i + \beta_1 t_i)\}\\
S_i\left(t_i \mid \boldsymbol{Y}_i\right) &= \int_0^{t_i} h_i\left(t \mid \boldsymbol{Y}_i\right) \mathrm{d}t.\\
h_0(t) &= \lambda\rho t^{\rho-1}
\end{aligned}
$$

The hazard rate $h_i\left(t_i \mid \boldsymbol{Y}_i\right)$ depends on the current covariate value $\beta_0 + b_i + \beta_1 t_i$ whereas the survival function $S_i\left(t_i \mid \boldsymbol{Y}_i\right)$ depends on knowing the whole trajectory of the longitudinal outcome up to time t_i. Numerical methods can be used to maximize this loglikelihood function e.g by Newton-Raphson procedure. Note that the integrals of the random effect require numerical methods (with the Gauss-Hermite quadrature technique the most commonly used).

A joint model is appropriate when interest lies in association between a time varying covariate measured with error in a survival analysis. The main limitation of this approach is that it is computationally intensive (Furgal et al., 2019).

Chapter 4

TwinsUK data

4.1 Data processing

The TwinsUK study is a longitudinal cohort of both monozygotic and dizygotic twins for research on multiple diseases and conditions. The cohort was first set up 1992 to investigate the incidence of osteoporosis and other rheumatologic diseases in monozygotic twins. The cohort has twins aged between eighteen and one hundred. The longitudinal nature of TwinsUK makes it an ideal cohort to study the process of ageing and age related diseases. The data can be obtained through `https://twinsuk.ac.uk` and is predominantly female.

Ageing is a complex process of accumulation of molecular, cellular, and organ damage, leading to loss of function and increased vulnerability to disease. Ageing leads to increased bone loss leading to low bone density and increasing the probability of developing fractures. The increasing rate of fracture occurrence in the elderly population has become a major public health problem worldwide and has resulted to increased morbidity, decreased quality of life, increased risks of dependence in daily living activities, and increased health care budgets (Marques et al., 2015; Wiklund et al., 2016). Furthermore, in addition to losing their independent mobility, fracture patients may be subject to a lot of psychological, physical, economic, and social stress during the perioperative period and may end up developing other complications such as deep venous thrombosis, pulmonary embolus, wound problem, and urinary retention (McLaughlin et al., 2006). It is estimated that the annual hip fracture incidence will be at least 4.5 million by the year 2050 from just 1.6 million in the year 2000 primarily as a result of an increased ageing population (Gullberg

et al., 1997).

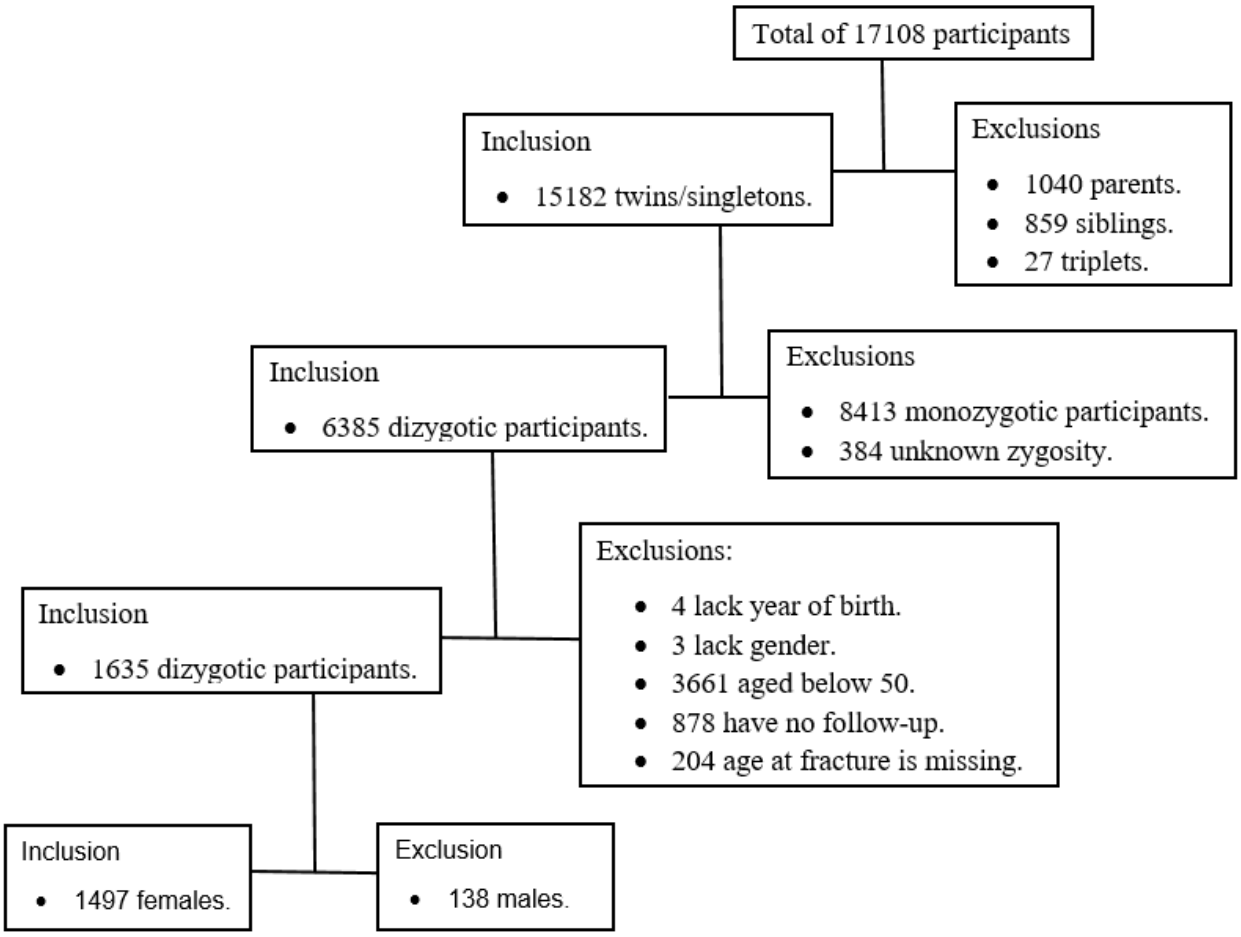
Our event of interest from this cohort is time to fracture for elderly twin pairs (aged 50 and above). The choice of age 50 and above is because a fracture is more likely to be due to ageing/osteoporosis and not other factors such as an accident. Previous studies that have considered fractures above age 50 include Agrawal and Sharma (2013); Baddoura et al. (2001); Lee and Khir (2007); Siris et al. (2006); Svedbom et al. (2014).

We used R for data processing and the code is available in Github (Data Processing, Muli (2022)). The raw data has a total of 17,108 individuals. We will consider only the dizygotic twins/singletons in this dissertation. We had a total of 50 variables (responses based on questionnaires) in different excel sheets with different visit dates that provided information on fractures.

The next step was to follow up each individual across the 50 variables to obtain their first visit and last visit dates and the age at which fracture occurred during the study. Here, the first visit was taken to be the first year of response after a twin turned 50 years of age. We let age-at-entry to study be defined by the age at the first visit. The age at end of study is defined as the age at the time of the occurrence of a fracture or the last visit after 50 years of age, whichever is earlier. If an individual had not reported any fracture at the last visit, the individual's time is considered censored at that last visit. We ignored any fracture below age 50. For individuals with multiple fractures above 50, we reported the first fracture as our age at fracture and ignored any subsequent fractures. We omit twins who were aged below 50 at the time of last visit.

Based on this criteria of being included in the study the inclusion/exclusion criteria is as show in the flowchart in Figure 4.1.

Figure 4.1: Flowchart for inclusion in study.



We consider female only because we have few males. 1497 is our sample size before inclusion of covariates. Note that only the year of birth (without day/month) and the age at fracture in years was provided and so the Kaplan Meier plot events occur at the end of the year.

4.2 Covariates

4.2.1 Time independent covariates

The most important risk factor for a fracture is bone mineral density (BMD) as demonstrated in multiple epidemiologic studies (Cummings et al., 1993; Group and O'Neill, 2002; Marshall et al., 1996; Miller et al., 2002). For our cohort, we have access to BMD measured

through dual-energy X-ray absorptiometry (DEXA). We shall therefore consider BMD as a covariate in this study. In particular, we use hip BMD because hip fractures has widely been studied in literature (Farzi et al., 2022; Marshall et al., 1996; Miller et al., 2002) . We assume the covariate value is missing at random for 112 participants who lack any BMD measurements taken after 50 years and omit them from study.

The other covariate we shall consider is health status which will be based on glycans. Glycans are one of the four primary components of the cell (alongside DNA, proteins and lipids). Glycans are important for maintenance of tissue structure, porosity, and integrity. We have access to glycans measured with Ultra-Performance Liquid Chromatography (UPLC) technology. Accuracy of this high-throughput glycomic methods is highly affected by complicated experimental procedure. The UPLC glycans were globally normalized to account for multiplicative error, log transformed, batch/plate normalised by ComBat based on methodology given in Menni et al. (2018). Normalization helps to make the measurements comparable. Since normalized glycans are right skewed, log transformation is performed on the data before batch correction is done. 303 participants lack glycan measurements taken after 50 years and are thus omitted. Therefore while considering BMD and health status as covariates the pre-processed data contains 1082 female dizygotic participants (411 twin pairs and 260 singletons). For this we let baseline BMD and glycomics be the first individual specific BMD measurements taken after age of 50.

Glycans have been shown to change with age, so can be used to predict biological age (Dall'Olio et al., 2013; Krištić et al., 2014; Vanhooren et al., 2007, 2010). Glycan age is a function of three immunoglobulin G (IgG) glycans (Krištić et al., 2014). The three UPLC glycans are a nongalactosylated (GP6) and two digalactosylated glycans (GP14 and GP15). We will use these UPLC glycans to compute Glycan age.

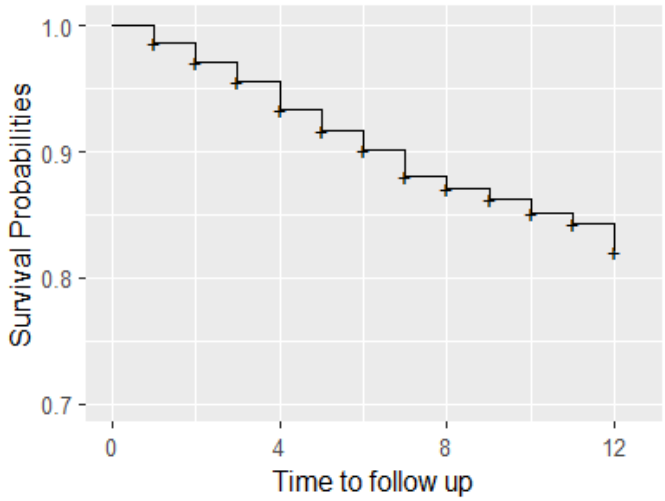
We do not consider glycans directly as a predictor. For the purpose of our analysis, we consider the difference between the glycan age (as biological age) and the chronological age. We term this as the *frailty index* (not to be confused with frailties as random effects in survival models). More details on *frailty index* are in Section 5.2. This *frailty index* is what will represent an individuals health status. A positive *frailty index* indicates that the individual experiences faster biological ageing than their chronological age (i.e. they are unhealthy) and the other way around with a negative *frailty index*.

To compute glycan age for our dataset, we fitted a linear model with age as the dependent variable and GP6,GP6^2, GP14 and GP15 as fixed independent variables (Krištić et al., 2014). The following formula was obtained

$$\text{Glycan age} \quad = \quad 59.20 + 2.10 \times \text{GP6} - 0.09 \times \text{GP6}^2 - 0.94 \times \text{GP14} - 1.12 \times \text{GP15}. \quad (4.1)$$

Some individuals had multiple fractures after entry in the study. About 18% (194 individuals) experienced a fracture within the study period of 20 years. Figure 4.2 below shows there is decrease in survival probabilities with increasing time to follow up. Therefore, there is increase in risk of fracture with increasing time to follow up.

Figure 4.2: Kaplan Meier for fractures using time to follow up as time scale.



4.2.2 Time varying covariate

Since we have multiple BMD observations for subjects as shown in Table 4.1 we would like to use this data to model BMD as a time-varying covariate. Also, based on Figure 4.3 BMD appears to change slightly with age. Let BMD at entry to be individual specific BMD measurements taken at entry while for individuals lacking BMD at entry age we carry forward the previous BMD taken after 50 years to be the BMD at entry.

We omit 357 participants who lack BMD measurement to carry forward to be BMD at entry (they have no BMD measurement between 50 and age at entry to study). After data processing this leaves us with 1028 female dizygotic participants (262 singletons and

383 twin pairs). This will be the dataset used to address question 2 of interest in Section 4.4. Table 4.1 below shows number of BMD measurements for the 1028 individuals when they are elderly. Figure 4.5 shows there is increase in risk of fracture with increasing age. Different subjects join the study at different entry ages as shown in Figure 4.4.

Table 4.1: BMD measured on different visits (from age 50 to end of study).

No. of BMD Measurement visits	Number of people
1	332
2	367
3	219
4	102
5	8

Figure 4.3: Spaghetti plot for 50 participants (a subset of the data) using BMD measurements taken after age 50.

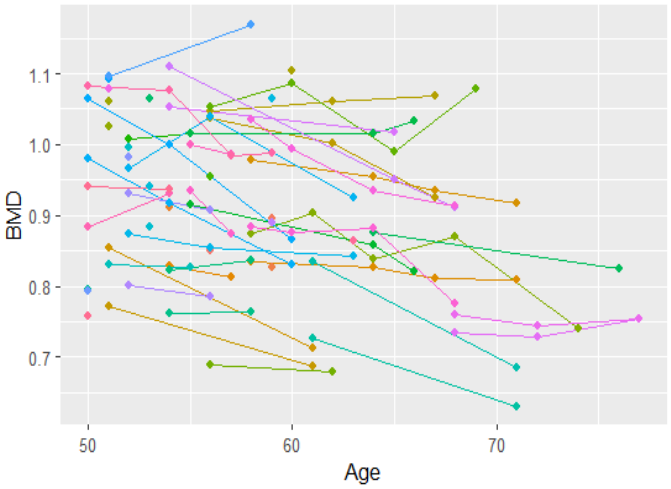


Figure 4.4: The following histogram shows the distribution of age at entry for the 1028 subjects.

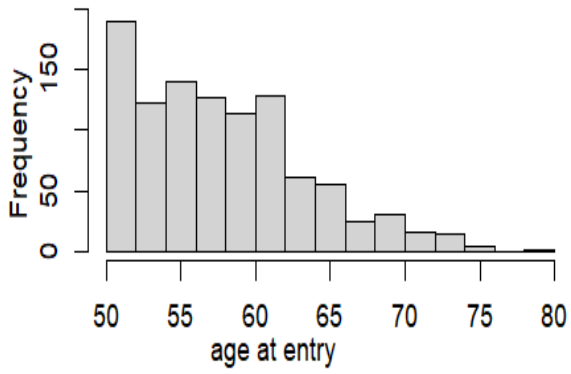
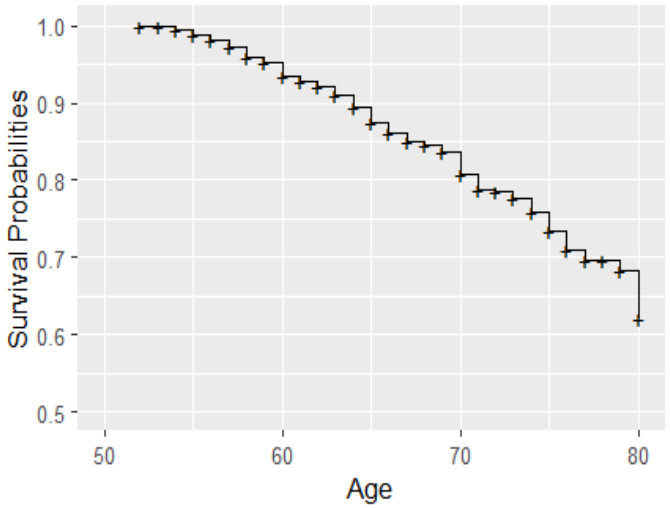


Figure 4.5: Kaplan Meier for fractures using age as time scale.



4.3 Time scale

Possible approaches for the time scale include time-to-follow and age at time of fracture. Time-to-follow up is defined as time a participant is in study i.e difference between age at entry and age at the end of study. When time-to-follow up is used as the time scale, age at entry is included as a covariate. The follow-up time approach makes the assumption that the relative change in rate of experiencing event is the same for a single unit increase in age in any follow-up interval. This may be unrealistic sometimes; for instance in chronic diseases the effect of age is often absent in the young and strong in the old (Cologne et al., 2012). Using time to follow up brings a challenge in interpretation. Age is a commonly used time scale in ageing studies that are analysed using the proportional hazards model (Cologne et al., 2012). We therefore consider to use age in order to be have easier comparison with other studies. When age is used as time scale in shared frailty model we need to take into account the delayed entry age as participants do not join at origin but rather at different ages. More details on delayed entry are in Chapter 7.

4.4 Questions of interest

(1) To estimate the survival probabilities for the occurrence of a fracture in the next time period given a participant's BMD and health status (i.e *frailty index*) using time to

follow-up as underlying time scale.

(2) To estimate the effect of BMD as a time dependent covariate on fracture incidence. Since we have multiple BMD observations for subjects we would like to use this data to construct a model to fit BMD as a time-varying covariate. Using BMD as time-varying covariate brings a challenge as the covariate measurements are taken on a few time-points within the study and not necessarily at event/censoring time-point. The question is therefore what to fill in for a persons covariate value at event ages. We will use age at time of event as time scale (instead of time to follow-up as time-scale) for better interpretation of the results. Using age as time scale means that an individual is only included in the study if their age at event/censoring is greater than age at which they join the study. This leads to the presence of left-truncated survival times due to delayed entry. If not carefully modelled, this will lead to selection bias, since more frail individuals tend to experience the event at younger age and thus less likely to be included in the study (Rodríguez-Girondo et al., 2018). For this case the frailty distribution needs to be updated with a new definition of the frailty distribution among survivors at the times of delayed entry. This will be the first study to incorporate time varying covariates and delayed entry in shared frailty models while using age as time-scale.

Note that the sample sizes considered in the above research questions differs. In the first question of interest two covariates are included in the study (i.e. BMD and Glycans) while in the second question only one covariate is considered(BMD).

Chapter 5

Analysis of the TwinsUK data

5.1 Introduction

This chapter is motivated by the the TwinsUK cohort described in Chapter 4. The cohort provides details on fracture incidence, BMD and glycans measurements. To model the effect of covariates on the hazards of an event Cox proportional hazard model is often used. For the Cox proportional hazard model, the effect of a covariate on the outcome is reported in a hazard rate. However, hazard rates are hard to interpret (Hernán, 2010). For an individual with a specific covariate profile the probability of the occurrence of the event in the next time period is more easily interpretable. In this chapter, we are specifically interested in estimating the probability of experiencing a fracture in the next five or ten years, given a participants bone mineral density (BMD) and health status (represented by *frailty index* and obtained from difference between glycan age and calendar age).

For a frailty model, the survival probability given a covariate profile is a function of frailty variance, the baseline hazard and the parameters modelling the effect of the covariates. Here, the frailty represents unobservable random effects which are shared by the twins and influence fracture incidence. In this chapter to model the fracture incidences, we shall consider proportional hazard models with gamma distributed random effects. We use the gamma frailty distribution in our analysis because of its mathematical convenience. Given a gamma frailty distribution the marginal survival probability exists in closed form. The marginal survival probability will be the probability of no fracture up to time t given the covariates. To compute the above marginal survival probability we need to specify the

baseline hazard. We will consider a parametric and flexible baseline hazard as will be described in section 5.4. Because both baseline hazard and frailty act multiplicatively on the hazard, we suggest that a flexible baseline may compensate for a misspecification in the frailty distribution in case a wrong frailty distribution is used.

To address the first question of interest described in Section 4.4 we consider BMD and health status as time independent risk factors for a fracture while using time to follow-up time-scale.

5.2 Materials

From TwinsUK we have follow up data from 932 females (313 twin pairs and 306 singletons) older than 50 years of age. The data was after removing those without glycan information and any BMD measurement taken after they were age 50. Note that this is slightly different dataset to what was described in Section 4.2.1 due some differences during data pre-processing (Section 4.2.1 has 150 more individuals). I used excel previously in performing the pre-processing which brings about challenge of reproducibility (some of the steps were manual while others were using a command). We defined follow up time as age at entry to age at the end of study. The entry age was defined as the age at the first visit after a twin turned 50 years of age. The average age at entry was 57.21 with sd 5.54. The age at end of study was defined as the age at the time of the occurrence of a fracture or the last visit after 50 years of age, whichever was first. If an individual had not reported any fracture at the last visit, the individual was censored at that last visit. For individuals with multiple fractures above 50, we reported the follow up time to the first fracture and ignored any subsequent fractures. About 16% (150 individuals) experienced a fracture within the study period of 20 years. We will model twins only data and both singletons with twins. We consider age at entry, BMD and *frailty index* as covariates. The data are described in table 5.1.

In addition to modelling BMD as a continuous variable, we also considered it as a categorical variable (low BMD, medium BMD, and high BMD) to enable estimation of the survival probability for different BMD categories. Here, low BMD was defined as a BMD value below the 15th percentile, medium BMD was defined as a BMD value between the

15th and 85th percentile, and high BMD was defined as a BMD value above the 85th percentile. The 15 percentile of our data was approximately equal to the 15th percentile of Finkelstein et al. (2008) who analysed BMD of a multiethnic cohort of menopausal women in America. Summary of BMD categories is given in Table 5.2. The measurement times varied across the twins. In Figure 5.1 the time between fracture occurrence or end of follow and BMD measurement is depicted. Twenty-two subjects appeared to have their BMD measured after the censoring/fracture age. For censored subjects, this means that the update on the occurrence of a fracture was prior to the BMD measurement. Our model assumes that the BMD does not change over the studied follow up time (mean=0.92, min= 0.55 max=1.38). To investigate this assumption, we used data on 822 individuals who had two BMD measurements. It appeared that BMD varies slightly with age (see Figure 5.3).

Analogously to BMD also the time between fracture occurrence or end of study and the first glycan measurements after entry differed between subjects (see Figure 5.2). For 67 subjects the glycans were measured after the censoring/fracture age. Analogously to BMD, our models assume that the *frailty index* which is the difference between glycan and calendar age will not change over the considered follow up time. To verify this assumption we used data of 40 subjects with two glycan measurements. It appeared that the *frailty index* only slightly changes with age (see Figure 5.4).

Table 5.1: Descriptives of age at entry (years), glycanage (years), and BMD (g/cm^2).

Covariate	N	Mean $\pm$ Std. deviation	Min	Max
Glycanage	932	60.42 ± 2.85	50.49	68.51
Frailty index	932	0.00 ± 6.00	-20.28	15.89
BMD	932	0.92 ± 0.12	0.55	1.38
Low BMD	140(15%)	0.75 ± 0.05	0.55	0.80
Medium BMD	652(70%)	0.92 ± 0.07	0.80	1.06
High BMD	140(15%)	1.12 ± 0.06	1.06	1.38
Age at entry	932	57.21 ± 5.54	50	80
Duration of follow up	932	9.33 ± 5.32	1	20

Figure 5.1: The time between fracture occurance and BMD measurements

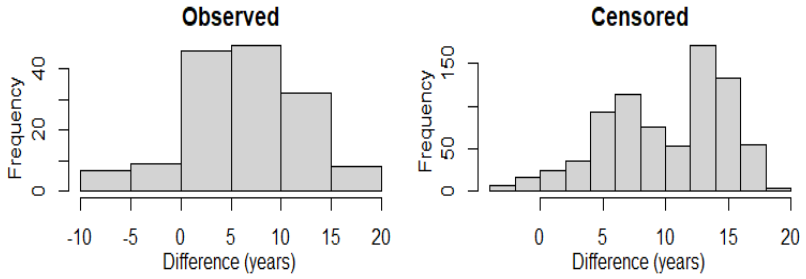


Figure 5.2: The time between fracture occurrence and glycan measurements.

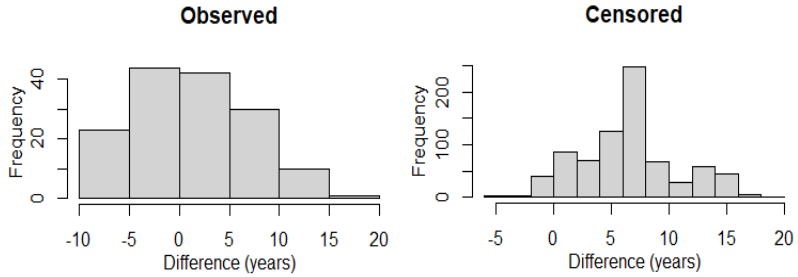


Figure 5.3: Scatter plot BMD at different visits. The straight line represents the identity line.

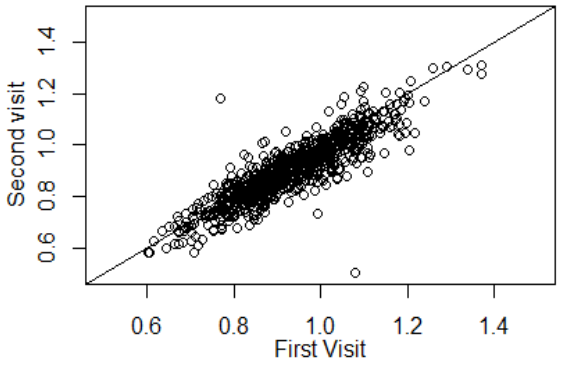
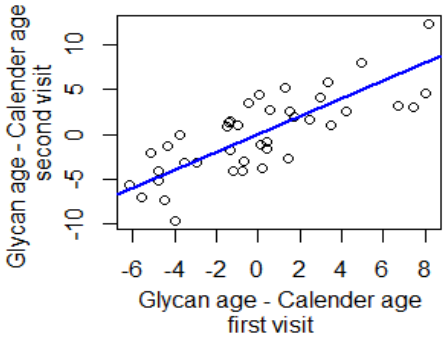


Figure 5.4: Scatter plot of glycan age − Calendar age at different visits. The straight line represents the identity line.



We also arbitrarily categorised *frailty index* into unhealthy, healthy, and very healthy subgroups in order to fit an interaction between BMD and *frailty index*. The unhealthy group comprises of individuals with higher glycan age than their corresponding calendar age by more than 2 years (i.e *frailty index* > 2). For the healthy group, the difference between glycan age and calendar age is 2 years or below (i.e $-2 \leq$ *frailty index* ≤ 2). The very healthy group includes individuals with lower glycan age than calendar age by more than 2 years (i.e *frailty index* < -2). Summary of *frailty index* categories is given in Table 5.3.

Number of fractures in different categories is as follows:

Table 5.2: Categorized by BMD

Category	No. of individuals	No. of observed fractures	% of fracture
LowBMD	92	26	0.28
MediumBMD	428	66	0.15
HighBMD	92	10	0.11

Table 5.3: Categorized by *frailty index*

Category	No. of individuals	No. of observed fractures	% of fracture
Very healthy	203	45	0.22
Healthy	152	28	0.18
Unhealthy	257	29	0.11
Total	612	102	0.17

Table 5.4: Categorized by health status interaction with BMD

		LowBMD	MediumBMD	HighBMD
Very healthy	No. of people	43	134	26
	No. of observed fractures	14	24	7
	% of fractures	0.33	0.18	0.27
Healthy	No. of people	20	107	25
	No. of observed fractures	6	20	2
	% of fractures	0.30	0.19	0.08
Unhealthy	No. of people	29	187	41
	No. of observed fractures	6	22	1
	% of fractures	0.21	0.12	0.02

5.3 Shared frailty model

Our goal is to estimate survival probabilities of fracture given covariates such as BMD and *frailty index* from twin data. To achieve this we use a shared frailty model. Suppose that there are G groups with n_i individuals in the i^{th} group, $i = 1, 2, ..., G$. Here, $n_i = 2$ for the case of twin pairs and $n_i = 1$ for the case of singletons. Let T_{ij} be the random variable representing time to fracture.

The hazard at time t for the j^{th} individual in the i^{th} group is then given by

$$h_{ij}\left(t|\boldsymbol{x}_{ij}, v_i\right) = h_0\left(t\right) v_i \exp\left(\boldsymbol{x}_{ij}\boldsymbol{\beta}\right).$$

We consider the gamma distribution to model the frailty in our analysis because of its mathematical convenience. The marginal loglikelihood function for all the clusters assuming gamma frailty distribution is given in Equation (2.12):

$$\begin{aligned} l\left(\boldsymbol{\xi}, \theta, \beta\right) = & \sum_{i=1}^{G} \left[d_i \log(\theta) - \log\left(\Gamma\left(1/\theta\right)\right) + \log\left(\Gamma\left(1/\theta + d_i\right)\right) - \right. \\ & \left. \left(\frac{1}{\theta} + d_i\right) \log\left(1 + \theta \sum_{j=1}^{n_i} H_0(t) e^{\boldsymbol{x}_{ij}\boldsymbol{\beta}}\right) + \sum_{j=1}^{n_i} \delta_{ij} \left[\boldsymbol{x}_{ij}\boldsymbol{\beta} + \log\left(h_0(t)\right)\right]\right] \end{aligned} \quad (5.1)$$

where $d_i = \sum_{j=1}^{n_i} \delta_{ij}$. The parameters of interest $\boldsymbol{\xi}, \theta, \boldsymbol{\beta}$ are estimated by maximizing Equation 5.1. Given gamma frailty distribution the survival probability exists in closed form and is given in Equation (2.13) i.e.

$$S_p(t|\boldsymbol{x}_{ij}) = \left[1 + \theta H_0(t) \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})\right]^{-1/\theta}. \quad (5.2)$$

This will be the probability of no fracture up to time t given the covariates.

5.4 Baseline hazard

To compute the marginal survival probability given in equation (5.2) above we need to specify the baseline hazard. We will consider both parametric and flexible baseline hazards. Because both baseline hazard and frailty act multiplicatively on the hazard, we suggest that a flexible baseline may compensate for a misspecification in the frailty distribution in case the wrong frailty distribution is chosen.

5.4.1 Parametric baseline

For this we consider the Weibull baseline hazard with shape parameter ρ and scale parameter λ:

$$h_0(t) = \lambda\rho t^{\rho-1}. \tag{5.3}$$

The Weibull baseline hazard is computationally attractive and can take different shapes depending on the value of the shape parameter ρ. The scale parameter λ of the baseline hazard acts multiplicatively on the hazard. Because both the frailty variable and λ act multiplicatively on the hazard, the scale parameter may adjust for frailty misspecification. For a more flexible alternative to the Weibull, we also consider non-parametric baseline hazard.

5.4.2 Flexible baseline hazard

We consider two options here: splines as described in section 5.4.2.1 and plug-in estimator for baseline hazard as described in section 5.4.2.2. Splines is a function defined piecewise by polynomials. It involves breaking time interval into sections and fit polynomials to the data in each one of these sections. The plug-in estimator for the cumulative hazard function is given by a step function with jumps at the ordered observed failure times.

5.4.2.1 Splines

We consider using B-splines for the baseline hazard. B-splines are known to be optimally stable (Peña, 1997; Perperoglou et al., 2019). Optimal stability implies that the numerical errors are not amplified when evaluating spline approximations.

$$h_0(t) = \sum_{\iota=0}^{q} \gamma_\iota B_{\iota,\kappa}(t), \qquad q \geq \kappa - 1. \tag{5.4}$$

where $\{\gamma_\iota \ \iota = 0, 1, 2, \cdots, q\}$ are the control points. The control points are like the data used for interpolation; they will determine the shape of the resulting B-spline curve and the curve may not intersect with these points. κ is the order of the polynomial segments of the B-spline curve (order κ means that the curve is made up of piecewise polynomial segments of degree $\kappa - 1$). The degree of a polynomial corresponds with the highest

coefficient that is non-zero. $B(.)$ denotes the B-spline basis functions described by the order κ and by a non decreasing sequence of real numbers $\{\tau_\iota : \iota = 0, \cdots, q+\kappa\}$ called the "knot sequence". Knots are points along a B-spline curve where the curve sections meet and connect. The basis functions are described as follows

$$B_{\iota,1}(t) = \begin{cases} 1 & \text{for } t \in [\tau_\iota, \tau_{\iota+1}), \\ 0 & \text{otherwise.} \end{cases}$$

and if $\kappa > 1$

$$B_{\iota,\kappa}(t) = \frac{t-\tau_\iota}{\tau_{\iota+\kappa-1}-\tau_\iota} B_{\iota,\kappa-1}(t) + \frac{\tau_{\iota+\kappa}-t}{\tau_{\iota+\kappa}-\tau_{\iota+1}} B_{\iota+1,\kappa-1}(t).$$

The above basis functions equations have the following properties

- Positivity: $B_{\iota,\kappa}(t) > 0$, for $\tau_\iota < t < \tau_{\iota+k}$.
- Local support: $B_{\iota,\kappa}(t) = 0$, for $\tau_0 \leq t \leq \tau_\iota$ and $\tau_{\iota+k} \leq t \leq \tau_{q+\kappa}$.
- Partition of unity: $\sum_{\iota=0}^{q} B_{\iota,\kappa}(t) = 1$, for $t \in [\tau_0, \tau_q]$.
- Continuity: $B_{\iota,\kappa}(t)$ has $\mathcal{C}^{\kappa-2}$ continuity at each simple knot. For cubic spline $B_{\iota,\kappa}(t)$ has $\mathcal{C}^2$ continuity. This ensures that curve segments are joined at ends and tangent vectors point in same direction.

Conditional on the knots, the $B_{\iota,\kappa}(t)$ are known functions of t and the γ_i are the parameters to be estimated. When maximizing the likelihood we obtain estimates of $(\gamma_0, \gamma_1, \cdots, \gamma_q)$ that maximize the marginal loglikelihod for a chosen number of knots and order of polynomial. In the simulations we chose to arbitrarily use 7 knots and cubic splines. We also tried fewer knots and more knots and the output was the same.

5.4.2.2 Plug-in estimator

This involves using a plug-in estimator for the cumulative hazard function given by a step function with jumps at the ordered observed failure times t_f, $f = 1, ..., F$ defined by (Gorfine et al., 2006; Zucker et al., 2008)

$$\Delta\hat{H}_0(t_f) = \frac{d_f}{\sum_{i=1}^{n} \psi_i \left(\beta, \theta, \hat{H}_0(t), t_f - 1\right) \sum_{j=1}^{m_i} Y_{ij}\left(t_f\right) e^{\mathbf{x}_{ij}\boldsymbol{\beta}}} \tag{5.5}$$

where d_f is the number of failures at time t_f,

$$\psi_i\left(\boldsymbol{\beta}, \theta, H_0(t), t\right) = \frac{\mathcal{L}^{D_i+1}\left\{\sum_{j=1}^{n_i} H_0(t_{ij}) e^{\mathbf{x}_{ij}\boldsymbol{\beta}}\right\}}{\mathcal{L}^{D_i}\left\{\sum_{j=1}^{n_i} H_0(t_{ij}) e^{\mathbf{x}_{ij}\boldsymbol{\beta}}\right\}}. \tag{5.6}$$

$\psi_i\left(\boldsymbol{\beta}, \theta, H_0(t), t\right)$ denotes the conditional expectation of the frailty term, given the observed data from cluster i and the parameters and D_i denotes the number of uncensored observations of cluster i, $\mathcal{L}\{.\}$ denotes the laplace transform of the term in the curly bracket, n_i represents cluster sizes, $\mathbf{x}_{ij}$ denotes the covariate, and $Y_{ij} = I(t_{ij} \geq t)$ with I denoting the indicator function.

In summary, the estimation procedure of the loglikelihood approach consists of the following steps (Gorfine et al., 2006) :

- Step 1. Use standard Cox regression software to obtain initial estimates of the regression coefficient $\boldsymbol{\beta}$, and set the initial value of frailty variance θ to be its value under within-cluster independence or under very weak dependency.
- Step 2. Use the current values of $\boldsymbol{\beta}$ and θ to estimate the cumulative baseline hazard $H_0(t)$ based on the estimation procedure defined by Equation 5.5.
- Step 3. Using the current value of $\hat{H}_0(t)$, estimate $\boldsymbol{\beta}$ and θ by maximizing the log likelihood function.
- Step 4. Iterate between Steps 2 and 3 until convergence.

The above algorithm is implemented in the `fitfrail()` function of `frailtySurv` R package which implements semi-parametric estimators for a variety of frailty distributions including gamma frailty distribution.

Software implementation

To fit the models we used R. The models with Weibull baseline hazard can be fitted by using the `parfm` package. The `parfm` fits a gamma frailty model using the maximum marginal likelihood estimation. The `frailtySurv` package fits semi-parametric shared gamma frailty model using a non-parametric plug-in estimator for the baseline hazard and involves optimization of the pseudo full likelihood. For the models where the baselines

are modelled with splines we developed our own code. The code requires the user to specify the observed time to event, censoring indicator, cluster identifier, and the matrix of covariates. It uses the `nlminb` function to optimize the maximum marginal likelihood. The `nlminb` function did not converge in some models resulting into singular convergence. In these cases, one cannot obtain the corresponding standard error of the variables. To solve this convergence problem we used a relative tolerance value of rel.tol=1e-6 which did not yield different results for parameter estimates as compared to the default relative tolerance value in the `nlminb` function. For all models, we wrote R functions to estimate marginal survival probabilities from the output of model fitting functions. This code is also available on Github (frailtySurvSplines).

5.5 Data analysis

We aim to estimate survival probabilities of fracture using a shared frailty model. As covariates, we consider, age at entry, BMD and *frailty index*. Based on Figure 5.3, BMD on second visit was slightly lower than BMD on first visit. However, for this analysis we only consider BMD at first visit. We also use *frailty index* at first visit as it appears not change over time as shown in Figure 5.4. In particular we model BMD as a continuous variable (model 1) and as a categorical variable (model 4), *frailty index* as a continuous variable (model 2) and as a categorical variable (model 5), both BMD and *frailty index* as continuous variables (model 3) and their categories interactions (model 6). We fit the categories (models 4,5 and 6) in order to make curves of the survival probabilities for different categories. Age at entry appeared to have no effect in the models and thus we did not consider it as covariate for fracture. For instance, considering both singletons and twin pairs data and using Weibull baseline hazard to fit model 1 to include age at entry as a covariate we have

	Variable	Estimate(s.e)
Model 1.	BMD	$-2.808(0.715)$
	Age at entry (θ)	0.013(0.015)
	Frailty variance (θ)	0.360(0.395)

We therefore exclude age at entry as covariate and fit all the models using (a) twin pairs only dataset(parameter estimates given in Table 5.5) and (b) both singletons and twin

pairs data (parameter estimates given in Table 5.6).

Table 5.5: Comparison of parameter estimates for the three specifications of baseline hazard assuming gamma frailty distribution using hip BMD for twin pairs only. The bold typeface indicates cases where use of plug-in estimator approach had convergence problems with estimating frailty variance.

		Weibull	Splines	Plug-in
	Variable	Estimate(s.e)	Estimate(s.e)	Estimate(s.e)
Model 1.	BMD	−3.164(0.885)	−3.153(0.898)	−3.302(1.176)
	Frailty variance (θ)	0.285(0.373)	0.321(0.383)	0.310(0.397)
Model 2.	*frailty index*	−0.037(0.018)	−0.037(0.018)	−0.038(0.022)
	Frailty variance (θ)	0.395(0.399)	0.448(0.404)	**1.000(0.849)**
Model 3.	BMD	−2.936(0.880)	−2.930(0.898)	−2.948(2.832)
	frailty index	−0.027(0.017)	−0.026(0.017)	−0.028(0.025)
	Frailty variance (θ)	0.192(0.355)	0.243(0.364)	**1.000(1.024)**
Model 4.	Low BMD	0.265(0.272)	0.283(0.277)	0.313(0.268)
	Medium BMD	1	1	1
	High BMD	−1.024(0.433)	−1.004(0.435)	−1.015(0.448)
	Frailty variance (θ)	0.348(0.400)	0.393(0.413)	0.371(0.452)
Model 5.	Unhealthy	−0.103(0.272)	−0.120(0.284)	−0.171(0.313)
	Healthy	1	1	1
	Very healthy	0.154(0.252)	0.146(0.263)	0.247(0.272)
	Frailty variance (θ)	0.486(0.422)	0.532(0.425)	0.467(0.439)
Model 6.	Veryhealthy⋆LowBMD	0.071(0.379)	0.063(0.383)	−0.008(0.393)
	Unhealthy⋆LowBMD	0.147(0.491)	0.176(0.495)	0.088(0.467)
	Healthy/Medium BMD	1	1	1
	Veryhealthy⋆HighBMD	−1.091(0.732)	−1.052(0.735)	−1.135(0.718)
	Unhealthy⋆HighBMD	−1.664(1.017)	−1.653(1.019)	−1.774(1.051)
	Frailty variance (θ)	0.465(0.423)	0.509(0.432)	0.405(0.434)

frailty index = glycan age − calendar age.

Table 5.6: Comparison of parameter estimates for the different specifications of baseline hazard assuming gamma frailty distribution using hip BMD for both singletons and twin pairs. The bold typeface indicates cases where use of plug-in estimator approach had convergence problems with estimating frailty variance.

		Weibull	Splines	Plug-in
	Variable	Estimate (s.e)	Estimate (s.e)	Estimate (s.e)
Model 1.	BMD	−2.918 (0.726)	−2.901 (0.730)	−2.927(0.939)
	Frailty variance (θ)	0.351 (0.390)	0.320 (0.377)	0.276(0.386)
Model 2.	*frailty index*	−0.021 (0.015)	−0.021 (0.015)	−0.024(0.0166)
	Frailty variance (θ)	0.539 (0.438)	0.529 (0.425)	**1.000(0.784)**
Model 3.	BMD	−2.816 (0.732)	−2.788 (0.737)	−2.775(2.107)
	frailty index	−0.011 (0.014)	−0.011 (0.014)	−0.014(0.018)
	Frailty variance (θ)	0.312 (0.384)	0.300 (0.375)	**1.000(0.928)**
Model 4.	Low BMD	0.339 (0.219)	0.356 (0.221)	0.356(0.217)
	Medium BMD	1	1	1
	High BMD	−0.663 (0.314)	−0.637 (0.314)	−0.644(0.322)
	Frailty variance (θ)	0.402 (0.411)	0.397 (0.404)	0.326(0.424)
Model 5.	Unhealthy	−0.026(0.227)	−0.042(0.234)	−0.094(0.251)
	Healthy	1	1	1
	Very healthy	0.110(0.216)	0.104(0.222)	0.194(0.230)
	Frailty variance (θ)	0.599(0.451)	0.586(0.439)	0.506(0.450)
Model 6.	Veryhealthy⋆LowBMD	0.218(0.307)	0.215(0.307)	0.133(0.322)
	Unhealthy⋆LowBMD	0.181(0.394)	0.201(0.394)	0.121(0.356)
	Healthy/Medium BMD	1	1	1
	Veryhealthy⋆HighBMD	−1.010(0.603)	−0.971(0.602)	−1.046(0.585)
	Unhealthy⋆HighBMD	−0.472(0.474)	−0.457(0.474)	−0.585(0.488)
	Frailty variance (θ)	0.552(0.446)	0.535(0.434)	0.422(0.434)

frailty index = glycan age − calendar age.

In Table 5.6 and Table 5.5, the parameter estimates for the fitted models are given. It appears that the models with Weibull baseline and the models with non-parametric baseline hazards yield similar estimates of the parameters and their standard errors. However is some scenarios, the `frailtySurv` approach (plug-in baseline hazard) had convergence problems especially with estimating frailty variance as shown model 2 and model 3 of Table 5.6 and Table 5.5. In model 1, BMD appears to be significantly associated with fracture incidence (a covariate effect of -2.918, p$-$value of < 0.001 for Weibull baseline considering whole dataset and a covariate effect of -3.164, p$-$value of < 0.001 for Weibull baseline considering twin pairs only dataset). A lower BMD value increases the risk for a fracture.

In model 2 of Table 5.5, *frailty index* is significantly associated with fracture incidence (a covariate effects of -0.037, p$-$value of 0.04). However, when we consider both twin pairs and singletons dataset *frailty index* becomes not significantly associated with fracture

incidence (a covariate effects of -0.021, p$-$value of 0.16).

Model 3 of both Table 5.5 and Table 5.6 shows that the parameter representing the effect of *frailty index* on fracture incidence is even smaller when controlling for BMD. All of the categories in model 4, model 5 and model 6 of both Table 5.5 and Table 5.6 are not significantly associated with fracture incidence with exception of high BMD category. High BMD effect is -0.663, p$-$value of 0.03 for model 4 of Table 5.6 and High BMD effect is -1.024, p$-$value of 0.02 for model 4 of Table 5.5.

To check the fit of the models, we tested the null hypothesis of proportional hazards assumption using scaled Schoenfeld residuals Grambsch and Therneau (1994). For all residuals models the null hypothesis could not be rejected. Models fitted with Weibull baseline hazard and with flexible baseline hazards yield same results of residuals because they had same estimates of parameters and their standard errors. Therefore we report residuals based on Weibull baseline hazard. For residuals of model 3, based on p$-$values the test is not statistically significant for both BMD (p$-$value 0.71) and *frailty index* (p$-$value 0.54) and the global test (p$-$value 0.86) is also not statistically significant. By graphical inspection of the scaled Schoenfeld residuals figure 5.5 shows that there is zero slope with time for both BMD and *frailty index* as covariates. For residuals of model 4 with categorical BMD as covariate, based on p$-$values the test is not statistically significant for low BMD (p$-$value 0.45) and high BMD (p$-$value 0.67) and the global test (p$-$value 1) is also not statistically significant. For model 5 with categorical *frailty index* as covariate, based on p$-$values the test is not statistically significant for unhealthy (p$-$value 0.80) and very healthy (p$-$value 0.56) and the global test (p$-$value 1) is also not statistically significant. Therefore, we can assume the proportional hazards for all the fitted univariate and multivariate models.

Figure 5.5: Graph of the scaled Schoenfeld residuals for model 3 against the time assuming Weibull baseline hazard for both singletons and twin pairs.

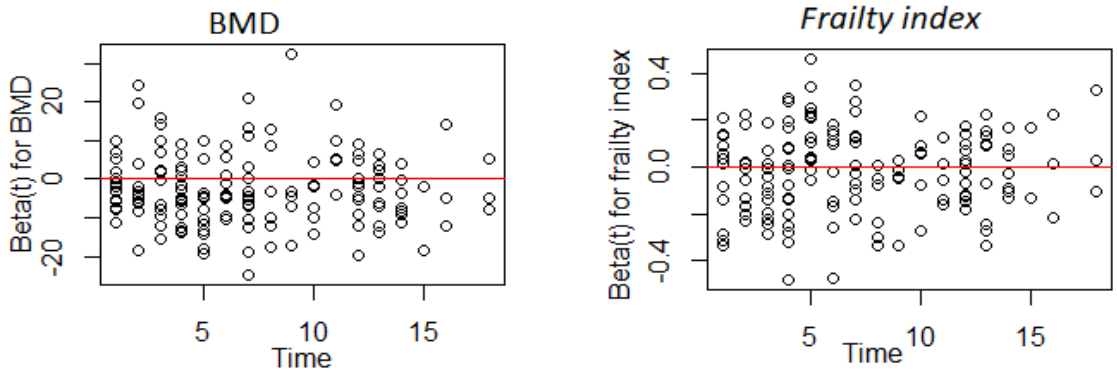


Figure 5.6: Graph of the scaled Schoenfeld residuals for model 4 against the time assuming Weibull baseline hazard for both singletons and twin pairs.

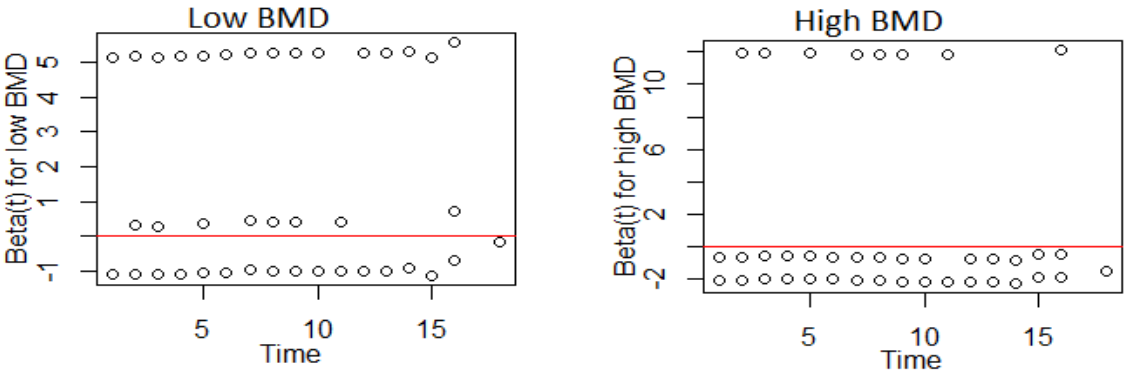
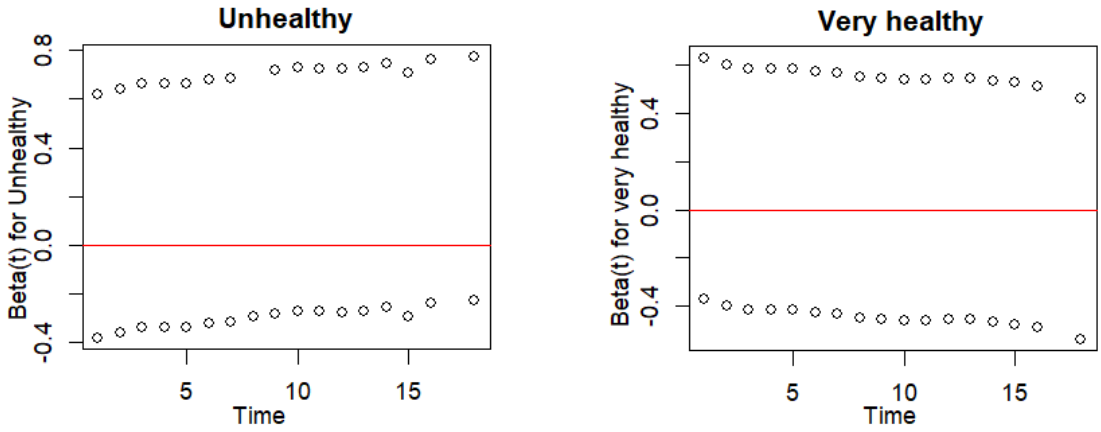


Figure 5.7: Graph of the scaled Schoenfeld residuals for model 5 against the time assuming Weibull baseline hazard for both singletons and twin pairs.



Since in the multivariate models, *frailty index* is not significant, we proceed with estimating survival probabilities as function of BMD. To facilitate this computation, we categorize BMD in three categories.

Figure 5.8: Comparison of the three specifications of baseline hazards in estimation of survival probabilities for model 4 of Table 5.6 for both singletons and twin pairs.

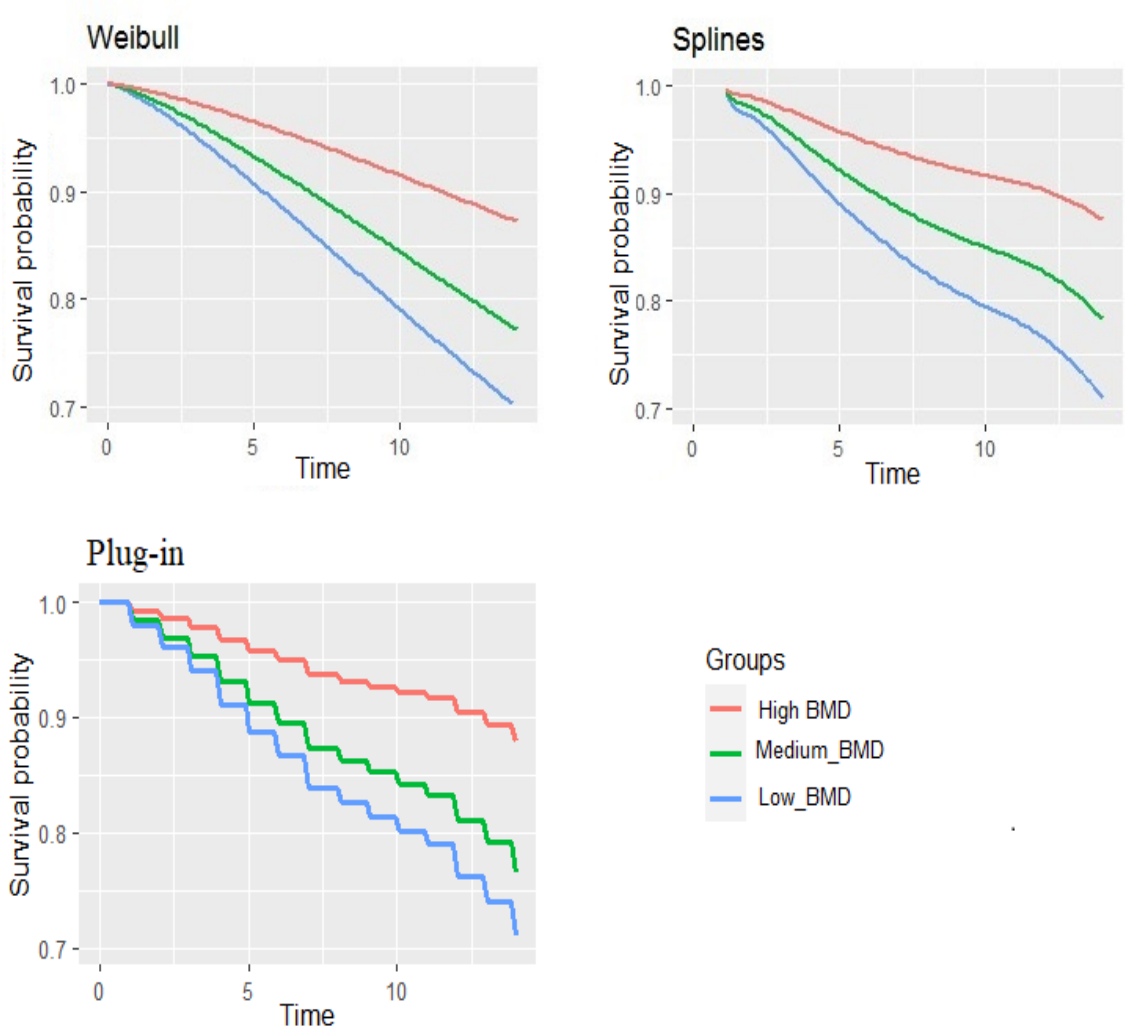
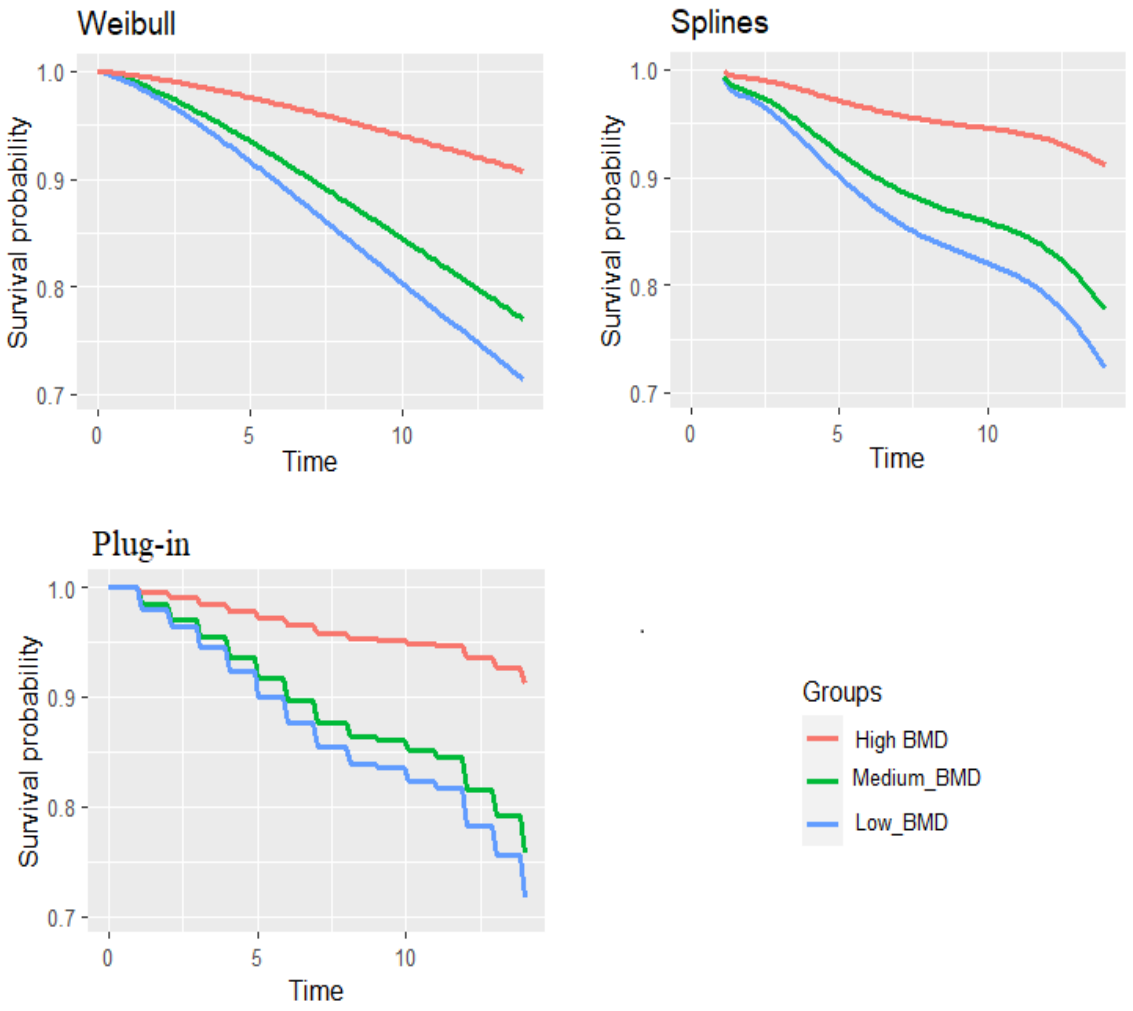


Figure 5.9: Comparison of the three specifications of baseline hazards in estimation of survival probabilities for model 4 of Table 5.5 for only twin pairs data.



In Figure 5.8 and Figure 5.9, the survival probabilities for subjects entering the study as function of time is given for the three categories of BMD. Regardless of the BMD category, the probability of fracture increases with follow up time. Subjects with low BMD have higher probability of fracture compared to the other groups at any given time. For instance, in the model with both singletons and twin pairs (Figure 5.8), assuming a Weibull baseline hazard the survival probability for the low BMD group is about 79% after 10 years of follow-up, while this probability is 84% for the medium BMD group and 92% for the high BMD group. Note that the curves for the Weibull, splines, and plug-in baseline hazards are very similar, which suggests that the Weibull hazards fit the data well.

5.6 Discussion

We estimated the survival probabilities for the occurance of a fracture in the next time period given a person's BMD value from twin data. A proportional hazard model with a gamma frailty is fitted. To estimate the survival probability we have to specify the baseline hazard. We considered a Weibull baseline and a non-parametric baseline hazard. BMD appeared to have a significant effect on fracture risks which is not a new finding as it has also been shown in literature (Cummings et al., 1993; Group and O'Neill, 2002; Marshall et al., 1996; Miller et al., 2002). The subjects with low BMD have a higher probability of fracture followed by the medium BMD group then the high BMD group. We also consdired the covariate *frailty index*. This covariate appeared not to have a significant effect on fracture risks.

In our analysis we considered gamma frailty distribution because it is the most commonly used distribution. However a miss-specified frailty distribution may result in biased estimates of the survival probabilities. Therefore we considered a model with a flexible baseline hazard as well as a parametric model. Both model gave the same estimates for the parameters representing the effect of the covariates on the survival and similar survival probabilities. We therefore conclude that the gamma distributed frailty and the weibull model probably fits the data well.

In this chapter we used follow up time as time scale. The advantage is that for females around 50 years this gives the probability of a fracture as function of the number of years in the future. Alternatively age as time scale could be used. Then survival probabilities for a fracture in the next five years can be computed for different ages. However by using age as underlying variable need to take into account delayed entry of the twins as will be discussed in Chapter 7. This brings a challenge in that the distance to the entry time varies between subjects, hence the assumption of non changing covariates becomes more critical.

A shortcoming of our model is that it assumes that the *frailty index* and BMD do not change over the studied period. Moreover we assume that the values of these covariates are not subject to measurement error. *Frailty index* was not significant probably because glycan age was estimated from the same dataset which is not correct and also it could be

frailty index plays a role in later ages while we only have it at baseline here. Since we have multiple BMD observations for subjects, a model to fit BMD as a time-varying covariate has been constructed in Chapter 7.

The analysis of twins data motivated us to also investigate if the parametric and non-parametric baseline hazards will always yield close survival probabilities estimates for different frailty distributions. This motivates the next chapter which is on investigating the impact of frailty misspecification on estimation of survival probabilities.

5.7 Conclusion

The parametric and flexible baseline hazards with gamma frailty yield similar results of parameter estimates and survival probabilities. BMD is a significant risk factor of fracture incidence whereas *frailty index* is not for the dataset of both singletons and twin pairs. Although *frailty index* is not a significant predictor, a person with a high value of *frailty index* (unhealthy person) has a smaller probability of developing a fracture. Also, low BMD value leads to higher probability of fracture.

Chapter 6

The role of baseline hazard in frailty misspecification correction

The analysis of twins data in Chapter 5 yielded similar survival probabilities estimates on fitting gamma frailty model with both parametric and flexible baseline hazards. We would therefore like to investigate if the parametric and flexible baseline hazards will always yield similar survival probabilities estimates for different frailty distributions. The choice of gamma distribution in Chapter 5 was for mathematical convenience since it produces a more tractable marginal likelihood function for the parameters after integration. However, in practice, the underlying frailty distribution is unknown and thus the gamma distributed frailty assumption might be wrong. This leads us to investigate the impact of using the wrong frailty distribution on estimation of regression coefficients, baseline hazard parameters and survival probabilities. This question has been partly addressed in literature.

Omori and Johnson (1993) investigated the influence of ignoring random effects on the unconditional hazard rate, survival function and other measures of dependence. The conclusion is that when the existence of random effects is ignored, the hazard rate will be underestimated regardless of heterogeneity distribution and the survival times are stochastically larger than expected. The paper only considers ignoring frailty and does not show the impact of misspecifying the frailty distribution.

Gasparini et al. (2019) investigated (using a simulation based study) the effect of misspecification of baseline hazard and frailty distribution in survival models with shared frailty

on estimation of regression coefficients as a measure of relative risk and on estimation of loss in life expectancy (the loss in life expectancy is defined as the difference in expectancy between exposed and non-exposed individuals). Data are generated to follow a gamma distribution, a log-normal distribution and a mixture normal distribution then a gamma frailty and log-normal frailty are fitted to the data. When a wrong frailty distribution is fitted we may obtain biased estimate of loss in life expectancy (largest bias is 16% bias). Hence misspecifying the frailty distribution may have an effect on estimation of loss in life expectancy (LLE). This investigation is the closest to our research question. Gasparini et al. (2019) investigates misspecification effect on estimation of LLE as a measure of absolute risk, our project however will be interested on investigating the effect of misspecifying the frailty distribution on population survival probabilities as a measure of absolute risk. Also, the paper does not report on effect of frailty mispecification on hazard parameters. Note that Gasparini et al. (2019) generates the lognormal and mixture normal frailties in the linear predictor additive with mean zero but models a gamma distribution which affects the hazard in a multiplicative manner assuming a mean of one. However the mean of such a model will not result in an additive zero mean in the linear predictor.

To show the effect of misspecifying frailty distribution on hazard ratio in order to build confidence in using frailty models, Munda and Legrand (2014) investigated robustness properties of parameters against frailty misspecification via simulations. Munda and Legrand (2014) generated frailty using inverse Gaussian, lognormal and positive stable then fitted a semi-parametric gamma distribution. The results show that inferences on the covariate effect are robust against misspecification of the frailty distribution and are under-estimated (biased) when the standard Cox model without frailty is assumed. Abiodun (2008); Gasparini et al. (2019); Glidden and Vittinghoff (2004); Henderson and Oman (1999) have also investigated and their finding is that the covariate effect is robust to misspecification of frailty distribution. Glidden and Vittinghoff (2004) generated data using inverse Gaussian and positive stable frailty distribution then fitted a gamma frailty distribution. If we compute the relative bias of covariate effect (which the paper didn't compute as they reported mean covariate effect), there is a noticeable negative relative bias of about 10% when the true underlying frailty distribution is positive stable.

Hsu et al. (2007) investigated the effect of frailty distribution misspecification on the

population averaged hazard function estimators when the marginalized hazard function obeys the Cox proportional hazard model for the case of case-control family data. The covariate effect and marginal hazard functions appear to be robust to frailty distribution misspecification.

Jiang et al. (2021) investigated the impact of misspecification of the frailty distribution on frailty prediction by simulation studies. The frailty term being predicted here is is the sum of the fixed effect and the cluster-level random effect. Six frailty distributions namely compound Poisson, inverse Gaussian, lognormal, positive stable, and the power variance function were considered. Jiang et al. (2021) restricted their attention onto the gamma distribution as the assumed frailty distribution, mostly because it is the most commonly used distribution in practice and its closed form of the marginal likelihood under the Cox proportional hazard shared frailty models can be obtained. The simulation results in Jiang et al. (2021) suggest that using a gamma frailty distribution in the Cox proportional hazard shared frailty models can produce robust frailty prediction even when the gamma frailty is a misspecified frailty model. However, when the underlying true frailty distribution is an extreme distribution such the positive stable frailty but a working gamma frailty distribution is employed, the prediction on frailty terms could be biased.

This chapter is organized as follows: section 6.1 is on simulation study to investigate the performances of the frailty model when the frailty distribution is misspecified while section 6.2 provides a discussion of the results.

6.1 Simulation Study

6.1.1 Aims

In this section, we perform simulations to investigate the effect of frailty misspecification on the estimation of regression coefficient β, baseline hazard parameters and the estimation of survival probability $S_p(t)$. Mutual dependence between lifespans of individuals in a cluster(twins in this case) is introduced. Data are generated using four frailty distributions; gamma, truncated gamma, lognormal , and mixture normal distribution. We then fit a gamma frailty and inverse Gaussian to the simulated data. The data is simulated to mimic real life scenarios.

6.1.2 Estimands

The estimands of interest include regression coefficient β, baseline hazard parameters, and population survival probability $S_p(t)$.

- Regression coefficient β - this is the log-covariate effect conditional on the value of frailty term.
- Baseline hazard parameters when a parametric form of the baseline hazard is considered. In this investigation we use the Weibull distribution with scale parameter λ and shape parameter ρ.
- Population survival probability $S_p(t)$ - this gives an interplay between parameter estimates of frailty variance, baseline hazard, and regression coefficient and is given by

$$S_p(t) \quad := \quad \int_v \exp\left[-vH_0(t)\exp(\boldsymbol{x}\boldsymbol{\beta})\right] g(v)dv.$$

6.1.3 Performance Measures

The misspecification effect on the estimation parameters was measured via the metric of relative bias, standard deviation and mean square error (MSE). The relative bias is used to quantify how far the estimator targets the true value on average relative to true value.

Let m represents the number of Monte Carlo trials and θ in this subsection represents any parameter in general (and not necessarily frailty variance). If $\hat{\theta}$ is the estimator of θ then,

$$\begin{aligned}
\text{Bias} &= \mathbb{E}\left(\hat{\theta}\right) - \theta \\
\text{Relative bias} &= \frac{\mathbb{E}\left(\hat{\theta}\right) - \theta}{\theta} \\
\text{MSE} &= \mathbb{E}\left(\hat{\theta} - \theta\right)^2 \\
\text{sd} &= \sqrt{\frac{\sum_{i=1}^{m}\left(\hat{\theta}_i - \mathbb{E}\left(\hat{\theta}\right)\right)^2}{m-1}}
\end{aligned}$$

A negative relative bias implies an underestimation of the estimand whereas a positive relative bias implies overestimation of the estimand.

6.1.4 Simulation design

We simulate data from the following theoretical model

$$h_{ij}(t) = h_0(t)\, v_i \exp(x_{ij}\beta) \quad j = 1, ..., n_i \quad i = 1, ..., G.$$

If $h_0(t) > 0$ for all t and assuming that the baseline hazard follows a Weibull distribution with scale parameter λ and shape parameter ρ then simulated survival time can be generated by (Bender et al., 2005)

$$T = \left(-\frac{\log(u)}{\lambda v_i \exp(x_{ij}\beta)} \right)^{1/\rho} \tag{6.1}$$

where $u \sim U(0,1)$ i.e the standard uniform distribution and v_i follows a distribution with mean 1 and variance θ.

The corresponding frailty model with a Weibull baseline hazard is given by

$$h_{ij}(t) = \lambda\rho t^{\rho-1} v_i \exp(x_{ij}\beta) \quad j = 1, ..., n_i \quad i = 1, ..., G.$$

Note that when $\rho = 1$ we have a exponential baseline as a special case. The simulation here in done for twin pairs (fixed cluster size $n_i = 2$) followed from birth and assuming no delayed entry.

The shared frailty v_i is simulated to follow a distribution with mean 1 and variance θ. Four frailties distributions are considered in data simulation; gamma, truncated gamma, lognormal, and mixture normal frailty distribution. Gamma distribution is most commonly used frailty distribution (discussed in section 2.2.1). For the lognormal frailty (discussed in section 2.2.2), we let $\mu = -\theta/2$ to achieve E(v) = 1. We consider left truncated gamma distribution (discussed in section 2.2.4) with the following sets of values: $a = 0.4, \eta_1 \approx 9.565, \eta_2 \approx 9.625$ for $\theta = 0.1$; $a = 0.4, \eta_1 \approx 0.0419, \eta_2 \approx 0.932$ for $\theta = 0.5$; and $a = 0.15, \eta_1 \approx 1.876 \times 10^{-5}, \eta_2 \approx 0.363$ for $\theta = 2$. Note that the choice for a was arbitrarily and user-defined. For the mixture normal frailty (discussed in Gasparini et al.

(2019)) we assume $g = 2$ hidden groups in each cluster (e.g an unmeasured binary covariate). Let π_g be the proportion in the groups and $\sum_g \pi_g = 1$. The hazard of the j^{th} individual in cluster i with g^{th} hidden group will be

$$h_{ij}(t) = h_0(t) \exp\left(\boldsymbol{x}_{ij}\boldsymbol{\beta} + \sum_g \pi_g H_g\right) \quad j = 1, ..., n_i \quad i = 1, ..., M \text{ and } g = 1, 2.$$

where $\sum_g \pi_g H_g$ follows a mixture normal distribution with mixing probabilities π_g and $H_g \sim N\left(\mu_g, \sigma_g^2\right)$. For this simulation study we let $\pi_1 = \pi_2 = 0.5$, $\sigma_1^2 = \sigma_2^2 = \theta$ (Gasparini et al., 2019), and $\mu = \{-6\theta, \log(2\exp(-\theta/2) - \exp(-6\theta))\}$.

The following settings are fixed in simulation of survival time: a binary covariate x with $\mathrm{P}(x = 1) = 0.5$, covariate effect β fixed to be 1.5, censoring is assumed to follow $U(0, 4)$, and the Weibull baseline hazard parameters are fixed to $\lambda = 1$ and $\rho = 2$.

We vary frailty variance θ ($\theta = 0.1, 0.5$ and 2) and number of clusters G ($G = 100, 200, 400$ and 800) when generating data for every frailty distribution used to generate the data. The settings for β, θ and G are based on Rodríguez-Girondo et al. (2018) We then fit the simulated data using both gamma frailty and inverse Gaussian frailty model and under three baseline hazards (B-splines, plug-in estimator and Weibull discussed in section 5.4).

The simulated data will consist of generated survival times, censoring indicator, cluster identifier and a covariate. 1000 Monte Carlo trial are generated.

To fit the simulated data with a parametric baseline (Weibull) hazard we use `parfm` R package by Munda et al. (2012). The package allows one to fit proportional hazards frailty models with gamma frailty distribution and the inverse Gaussian frailty. To analyse the simulated data with the first flexible baseline hazard (i.e plug-in estimator for cumulative hazard function given by a step function as discussed in section 5.4.2.2) we use the R package `frailtySurv`. Lastly, we fit B-splines for baseline hazard using a code which we developed (code is available in Github frailtySurvSplines).

6.1.5 Simulation Results

In all the simulated scenarios results are presented for relative bias, standard deviation, and mean square error across the 1000 Monte Carlo trials of regression coefficient (regression coefficient) β, baseline hazard parameters λ and ρ, and the survival probability $S_p(t)$.

Two values of t ($t = 0.3$ and $t = 1$) have been considered when computing the population survival probability. $t = 0.3$ is considered in order to be able to capture more individuals with high frailty.

Only graphical representation of results on relative bias for β, θ and $S_p(1)$ are provided in this section. The actual values of relative bias, standard deviation and mean square error are provided in the Appendix A.

Figure 6.1: Relative bias as a function of number of clusters. True: gamma ($\theta = 0.1, 0.5, 2$) with Weibull baseline ($\lambda = 1$) then fitted gamma and inverse Gaussian distributions with different baseline hazards.

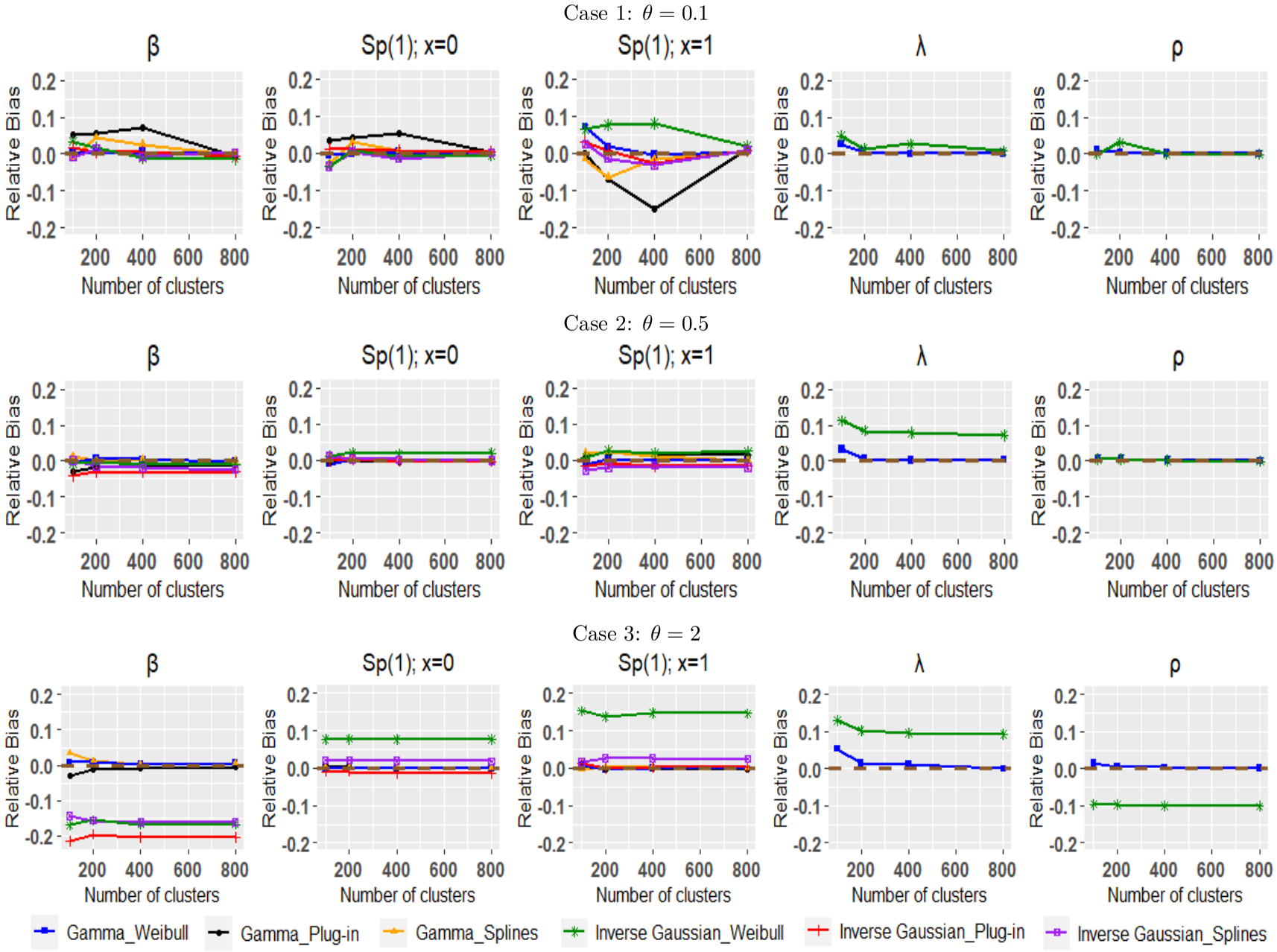


Figure 6.2: Relative bias as a function of number of clusters. True: truncated gamma ($\theta = 0.1, 0.5, 2$) with Weibull baseline ($\lambda = 1$) then fitted gamma and inverse Gaussian distributions with different baseline hazards.

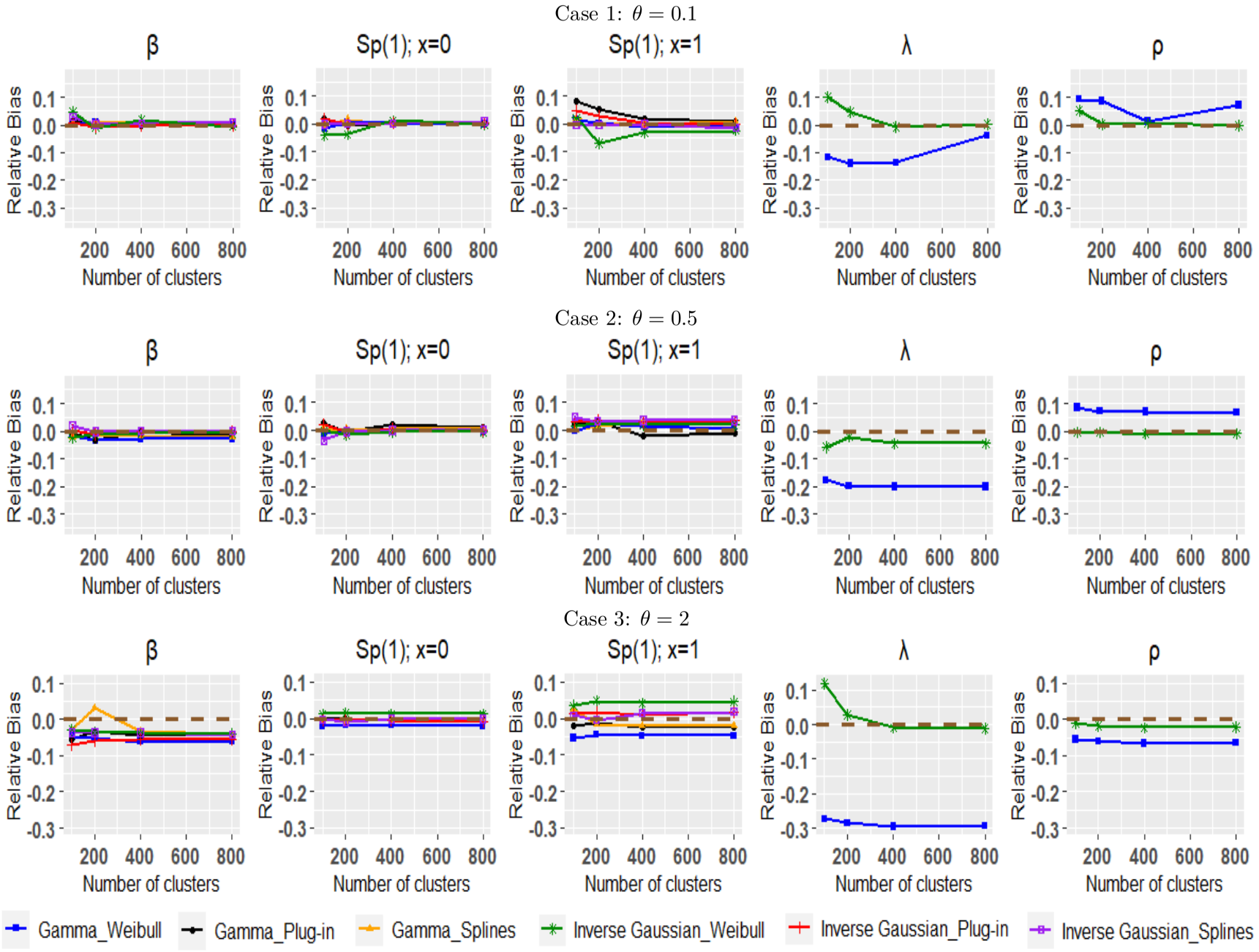


Figure 6.3: Relative bias as a function of number of clusters. True: lognormal ($\theta = 0.1, 0.5, 2$) with Weibull baseline ($\lambda = 1$) then fitted gamma and inverse Gaussian distributions with different baseline hazards.

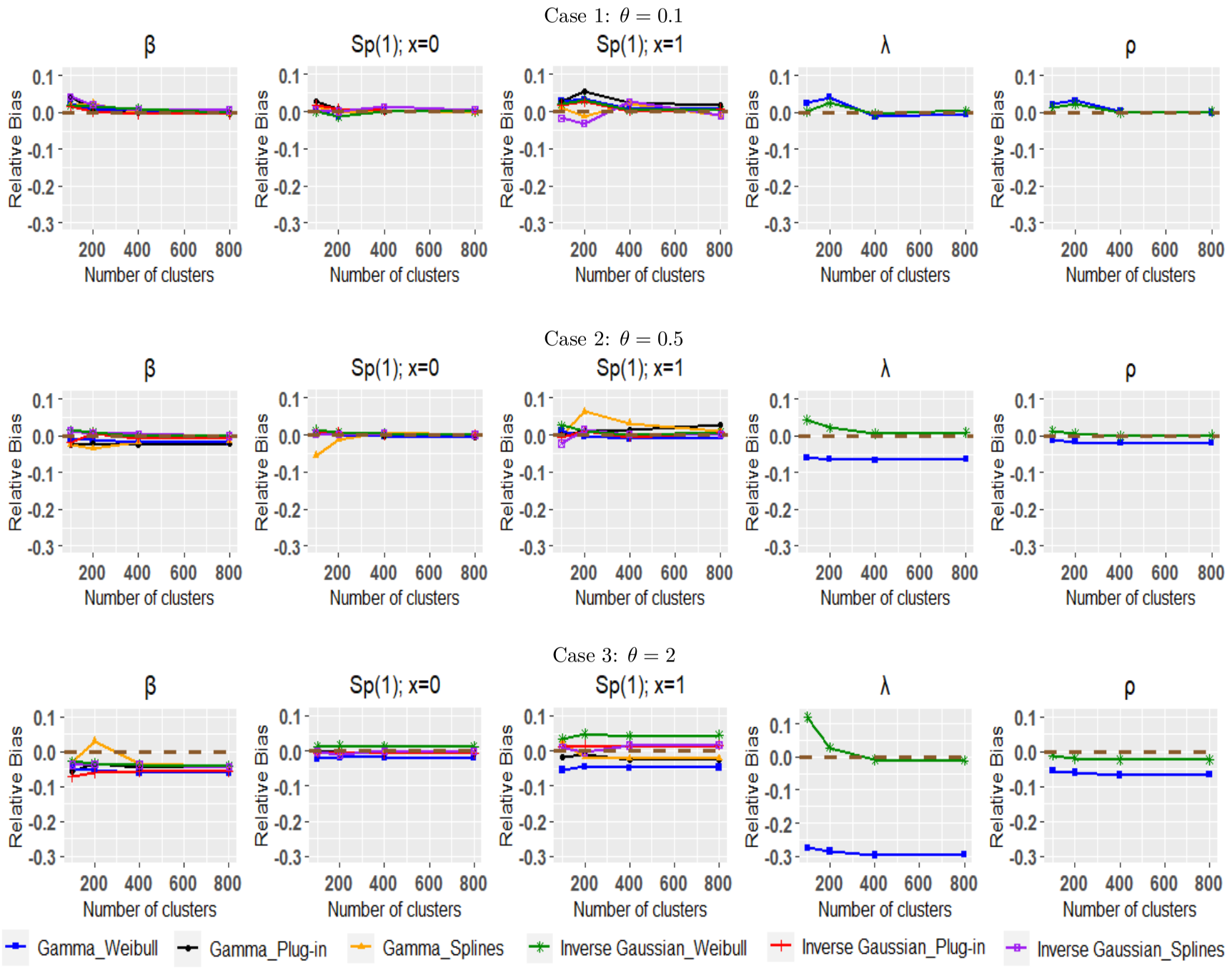
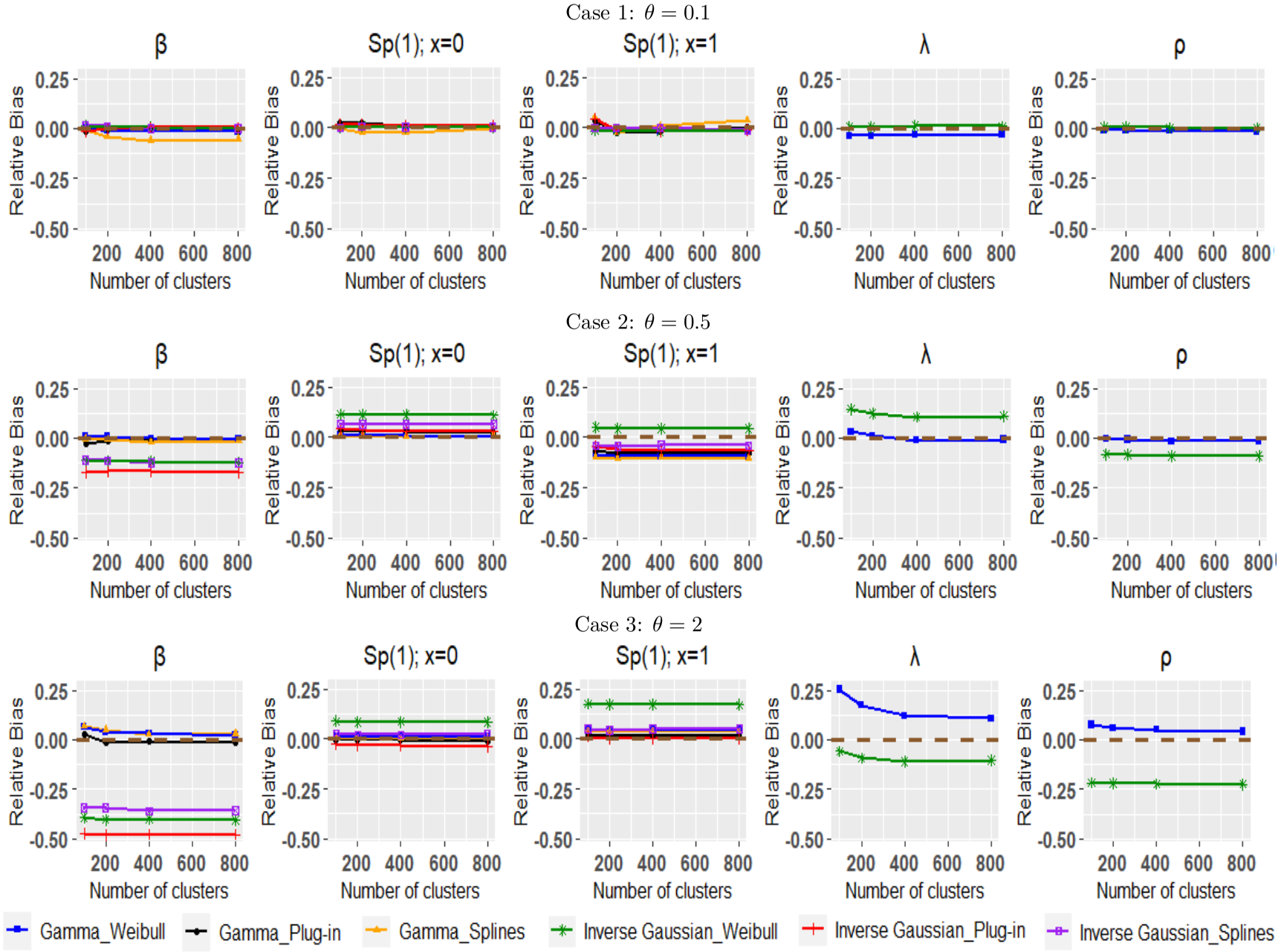


Figure 6.4: Relative bias as a function of number of clusters. True: mixture normal ($\theta = 0.1, 0.5, 2$) with Weibull baseline ($\lambda = 1$) then fitted gamma and inverse Gaussian distributions with different baseline hazards.



In this study we consider an absolute bias greater than 5% as biased estimation. We will discuss the results per every distribution used to generate the data.

For the first case we simulated data using a gamma frailty and then fitted gamma and inverse Gaussian frailty models with three baseline hazards(Weibull, plug-in and splines). From the plots in Figure 6.1, we observe that the regression coefficient β, is well estimated by true frailty model for all θ and for all G. The Weibull baseline hazard parameters λ and ρ, and the population survival probability $S_p(t)$ are all well estimated by true frailty model. For the wrong choice of an inverse Gaussian frailty, the regression coefficient β is underestimated with largest relative bias upto 20% at $G = 100$, $\theta = 2$ irregardless of the baseline hazard used. Similarly, the Weibull baseline hazard shape parameter ρ is underestimated with bias of about 10% at $G = 800$, $\theta = 2$. The Weibull baseline hazard scale parameter λ is overestimated. with a bias of about 13% at $G = 100$, $\theta = 2$. This overestimation in λ corrects for underestimation in β in this scenario. The population survival probability at time $t = 1$ is poorly estimated. Use of a more flexible baseline does not reduce the bias in estimators of covariate effect β when an inverse Gaussian frailty model is fitted although using splines performs better.

In the second case we simulated data using a truncated gamma frailty and then fitted gamma and inverse Gaussian frailty models with three baseline hazards(Weibull, plug-in and splines). We observe that the covariate effect β, and the population survival probability $S_p(t)$ appear to be less sensitive to the frailty misspecification as shown in Figure 6.2. (the largest bias is about 6% in estimators of β at $\theta = 0.2$, $G = 200$ and the largest bias is 8% in estimators of $S_p(1)$ at $\theta = 0.1$, $G = 200$). The Weibull baseline hazard shape parameter ρ seems to be sensitive to frailty misspecification (the largest bias is about 9% at $\theta = 0.1$, $G = 100$). The Weibull baseline hazard scale parameter λ seems to be affected by the wrong choice of frailty distribution. In particular, fitting the gamma misspecified frailty model has a bias of about 35% (at $\theta = 2$, $G = 800$) whereas the inverse Gaussian has a bias of about 16% (at $\theta = 2$, $G = 100$). λ seems to correct for bias in β estimation in this scenario. Use of a more flexible baseline(nonparametric) reduces the bias in estimators of covariate effect β.

In the third case we simulated data using a unity mean lognormal frailty and then fitted gamma and inverse Gaussian frailty models with three baseline hazards(Weibull, plug-in

and splines). Based on the plots in Figure 6.3, we observe that the estimation of the covariate effect β is less sensitive to the choice of gamma and inverse Gaussian frailty models (largest bias $\approx 6\%$). The Weibull baseline scale parameter λ is affected by both gamma and inverse Gaussian wrong frailty distributions. In particular, when $\theta = 2$ gamma yields a bias of about 30% for all considered G, inverse Gaussian has a bias of about 12% for $G = 100$. The Weibull baseline hazard shape parameter ρ is is less sensitive to any of the considered wrong frailty distributions. The population survival probability seems to be insensitive to the wrong choice of frailty distributions. Use of a more flexible baseline(plug-in estimator or splines) reduces the bias in estimators of covariate effect β.

In the fourth case we simulated data using a mixture normal frailty and then fitted gamma and inverse Gaussian frailty models with three baseline hazards(Weibull, plug-in and splines). From the plots in Figure 6.4, we observe that the estimation of the covariate effect β is affected by the inverse Gaussian wrong frailty with a bias of about 40% (at $\theta = 2$ for all considered G) but it is less affected by the wrong choice of gamma frailty (here, the largest bias is 6% at $\theta = 2$, $G = 100$). The population survival probability at $t = 1$ i.e. $S_p(1)$ appear to be sensitive to the frailty misspecification (the largest bias is 17% at $\theta = 2$, for all G and $x = 1$ when Inverse Gaussian model is fitted). We observe that the estimation of the Weibull baseline hazard scale parameter λ is affected by the wrong choice of gamma frailty (here, the largest bias is 25% at $\theta = 2$, $G = 100$) but is less affected by fitting inverse Gaussian frailty (largest relative bias of 14% at $\theta = 0.5$, $G = 100$, $G = 100$). The estimation of the Weibull baseline hazard shape parameter ρ is affected by fitting inverse Gaussian frailty with a bias of about 22% (at $\theta = 2$ for all considered G) but it is much less affected by the wrong choice of gamma frailty (here, the largest bias is about 8% at $\theta = 0.5$, $G = 100$). Use of a more flexible baseline does not reduce the bias in estimators of covariate effect β when inverse Gaussian frailty model is fitted although using splines performs better.

The number of clusters does seems not to have an effect on bias of estimation of parameters which means that misspecification effect cannot be compensated by using more clusters. We also observe that as the frailty variance increases, the estimates of the baseline hazard and the regression coefficient move further away from their true values. Note that increase in variance increases the dissimilarity in shapes between the true and wrong frailty distri-

butions. Different frailty distributions yield different survival probabilities. We have more bias when the true frailty distribution is more dissimilar to fitted distribution.

Overall, the results suggest that the estimation of regression coefficient β, the baseline hazard scale parameter λ and baseline hazard shape parameter ρ may be affected by frailty misspecification with λ being most affected. The population survival probability $S_p(t)$ is also affected by wrong choice of frailty distributions especially for small probabilities and large frailty variance. However, use of a flexible baseline hazard may compensate for a misspecified frailty distribution.

6.2 Discussion

The effect of frailty misspecification on survival probability $S_p(t)$ has not been investigated in the literature. It is also often concluded that regression coefficient β is robust with regard to use of wrong frailty. In this study, we have conducted extensive simulations to investigate the effect of frailty misspecification on the estimation accuracy of model parameters of interest namely regression coefficient β, baseline hazard scale parameter λ and baseline hazard shape parameter ρ, and survival probability $S_p(t)$.

Four frailty distributions namely gamma, truncated gamma, lognormal, and mixture normal are considered for generating the data. For each of the frailty model, a gamma/inverse Gaussian frailty model is then fitted to the data. The percentage (%) threshold of misspecification is an open question subject to debate, in this study we considered an absolute bias greater than 5% as biased estimation but one could consider other values depending on application.

From the simulations results in Section 6.1.5 β is affected by use of a wrong frailty especially with increasing the amount of heterogeneity in sample. This is contrary to the existing literature that estimation of regression coefficient is robust with respect to frailty misspecification which has been reported by previous studies including Gasparini et al. (2019), Munda and Legrand (2014), Abiodun (2008). Gasparini et al. (2019) considers lower values of frailty variance (largest is $\theta = 1.25$) but we consider high values of frailty variance (largest is $\theta = 2$) and does not include inverse Gaussian as a wrong choice of frailty distribution which we fit in our simulations. Gasparini et al. (2019) considers clus-

ters of size two (which is similar to our simulations) and clusters of size 150. Munda and Legrand (2014) does not fit a parametric gamma frailty model hence its a possibility that fitting a semi-parametric gamma distribution (baseline is not specified) corrected for any bias in the regression coefficient. This paper also considered smaller values of frailty variance (largest is $\theta = 1.33$) and larger cluster sizes (clusters of size 6 and 48). Lastly, Abiodun (2008) considered a larger cluster size of 10 and gives results based on only one replicate while in our simulations perform simulations based on 1000 replicates and assume a cluster size of size 2.

We observe from our simulation results that in the case of small frailty variance using a wrong frailty distribution still allows us to measure the fixed regression coefficient without underestimation or overestimation of the regression coefficient. However, this robustness does not seem to hold in several cases at larger frailty variances. Increase in variance increases the dissimilarity in shapes between the true and wrong frailty distributions. When the Weibull baseline hazard is used, in most the scale parameter λ corrected for the wrong frailty. However, for some extreme cases it appears that λ cannot compensate and thus the baseline hazard shape parameter ρ and regression coefficient β and survival probabilities are affected. Using a flexible function for the baseline hazard improves estimation of parameters as well as survival probabilities in presence of frailty distribution misspecification.

This study focused only on fitting gamma frailty and inverse Gaussian frailty distributions. This is a limitation as there is a wider variety of frailty distributions which can be considered for fitting (e.g. positive stable, log-normal, power variance function, etc). Also considered here is shared frailty model which assumes that the frailty value is same for individuals within a cluster. One could also examine the effect of misspecification in the case of correlated frailty or shared frailty with time varying covariates. The simulation study involves clusters of size 2 limiting generalization to cluster of size two like twins data. One could consider clusters of large sizes (> 2). One could also incorporate different levels of censoring to examine if censoring could have an effect on frailty misspecification.

In practical situations, the parametric assumption of the frailty distribution may not hold when one assumes a wrong frailty distribution. Most parametric models are sensitive to model assumptions and often lead to misleading results if some of the underlying model

assumptions are violated. The correct choice of frailty distribution is important especially when the frailty variance is large because the degree of bias is expected to increase with amount of heterogeneity in the data.

Chapter 7

Time dependent covariate model adjusting for delayed entry

This chapter is motivated by the TwinsUK study discussed in Chapter 5 where the aim was to predict what happens in the next time interval given the current values of covariates. The aim of analysis in this chapter is to model the relationship between BMD and fractures. Since BMD changes over time and we have longitudinal BMD information, in this chapter we will consider BMD as a time varying covariate and use age as underlying time scale to model the relationship between BMD and fractures. The challenge faced will be to adjust for delayed entry because individuals enter at various ages. We therefore propose to develop methodology that deals with delayed entry as well as time varying covariates simultaneously for clustered data. Previous methodological work that attempted to address these include; studies which fit a joint model while adjusting for delayed entry for singletons (Crowther et al., 2016; Schluchter and Piccorelli, 2019) and studies for clustered data without delayed entry (Brilleman et al., 2019; Chen, 2016; Elmi et al., 2018; Ratcliffe et al., 2004).

This chapter is organized as follows: Section 7.1 is on shared frailty model for time independent covariates incorporating delayed entry, Section 7.2 is on shared frailty model with time varying covariates, and Section 7.3 is on shared frailty model for time varying covariates and delayed entry.

7.1 Shared frailty model incorporating delayed entry and time independent covariate

Suppose that there are G groups with n_i individuals in the i^{th} group, $i = 1, 2, ..., G$ and $j = 1, ..., n_i$. Let T_{ij} be the random variable representing survival time given in age scale and T_{ij0} be the random variable representing the age at enrolment to the study. An individual is included in the study if $t_{ij} > t_{ij0}$. Let $\boldsymbol{x}_{ij}$ be a row vector of individual specific time independent covariates. Let $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_p)^T$ be a parameter vector for the time independent covariates. Let $\boldsymbol{\xi}$ be the parameters of the baseline hazard $h_0(t)$, θ be the frailty variance and δ_{ij} be the event indicator.

The hazard at time t for the j^{th} individual in the i^{th} group conditional on v_i is given by

$$h_{ij}(t|\boldsymbol{x}_{ij}, v_i) = h_0(t) v_i \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta}).$$

The contribution of individual j in the i^{th} cluster in time interval (t_{ij0}, t_{ij}) conditional on v_i is given by $f(t_{ij}|\boldsymbol{x}_{ij}, v_i)^{\delta_{ij}} S(t_{ij}|\boldsymbol{x}_{ij}, v_i)^{1-\delta_{ij}} / S(t_{ij0}|\boldsymbol{x}_{ij}, v_i)$. i.e for an individual to contribute to the likelihood within the interval (t_{ij0}, t_{ij}) an individual needs to have survived beyond time t_{ij0}. The likelihood for cluster i conditional on v_i is given by

$$\begin{aligned}
L_i &= f(t_{i1}, ..., t_{in_i}|t_{i1} > t_{i10}, ..., t_{in_i} > t_{in_i0}, v_i, \boldsymbol{x}_{ij}) \\
&= \prod_{j=1}^{n_i} \frac{f(t_{ij}|\boldsymbol{x}_{ij}, v_i)^{\delta_{ij}} S(t_{ij}|\boldsymbol{x}_{ij}, v_i)^{1-\delta_{ij}}}{S(t_{ij0}|\boldsymbol{x}_{ij}, v_i)} \\
&= \prod_{j=1}^{n_i} \frac{h(t_{ij}|\boldsymbol{x}_{ij}, v_i)^{\delta_{ij}} S(t_{ij}|\boldsymbol{x}_{ij}, v_i)^{\delta_{ij}} S(t_{ij}|\boldsymbol{x}_{ij}, v_i)^{1-\delta_{ij}}}{S(t_{ij0}|\boldsymbol{x}_{ij}, v_i)} \\
&= \prod_{j=1}^{n_i} \frac{h(t_{ij}|\boldsymbol{x}_{ij}, v_i)^{\delta_{ij}} S(t_{ij}|\boldsymbol{x}_{ij}, v_i)}{S(t_{ij0}|\boldsymbol{x}_{ij}, v_i)} \\
&= \prod_{j=1}^{n_i} \frac{h(t_{ij}|\boldsymbol{x}_{ij}, v_i)^{\delta_{ij}} \exp(-H(t_{ij}|\boldsymbol{x}_{ij}, v_i))}{\exp(-H(t_{ij0}|\boldsymbol{x}_{ij}, v_i))} \\
&= \prod_{j=1}^{n_i} h(t_{ij}|\boldsymbol{x}_{ij}, v_i)^{\delta_{ij}} \exp(-[H(t_{ij}|\boldsymbol{x}_{ij}, v_i) - H(t_{ij0}|\boldsymbol{x}_{ij}, v_i)]) \\
&= \prod_{j=1}^{n_i} [h_0(t_{ij}) v_i \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})]^{\delta_{ij}} \exp(-[H_0(t_{ij}) - H_0(t_{ij0})] v_i \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta}))
\end{aligned}$$

The marginal likelihood function is obtained by integrating out v_i from the likelihood function. Since age has been chosen as the basic timescale, the frailty distribution needs

to be updated with a new definition of the frailty distribution among survivors at the times of delayed entry. Specifically, when we have delayed entry there is informative selection of individuals. Frail individuals (higher v) have a higher hazard hence may experience event first, therefore are less likely to be observed. As a result, the frailty distribution among survivors at a given age t differs from the original one given by $g(v)$. The frailty distribution $g(v)$ corresponds to the distribution of the frailty values at origin time of the study. Specifically, the mean of the frailty distribution at a given age becomes smaller as the stronger individuals remain (i.e mean of $\mathbb{E}V \neq 1$ but rather $\mathbb{E}\ (V|T > t) \leq 1$). At the same time, the variance also becomes smaller since the remaining individuals at risk are more alike (i.e. $\text{var}(V|T > t) \leq \theta$). Since we do not have all subjects of the birth cohort, but only the ones surviving we need to use an updated frailty distribution. The frailty distribution at the left-truncation times is $g(v|t_{i1} > t_{i10}, ..., t_{in_i} > t_{in_i0})$ (Jensen et al., 2004; Rodríguez-Girondo et al., 2018; Van den Berg and Drepper, 2016). This updated frailty distribution allows the n_i members of a cluster i to enter at different entry times. The correction factor for delayed entry can be seen as a weights and these are simultaneously estimated with the model parameters. By doing so, the correct weights to estimate the parameters of the model are obtained.

The marginal likelihood for cluster i is

$$\begin{aligned} L_i &= \mathbb{E}_v[f(t_{i1}, ..., t_{in_i}|t_{i1} > t_{i10}, ..., t_{in_i} > t_{in_i0}, \boldsymbol{X}_i, v_i)] \\ &= \int_v \prod_{j=1}^{n_i} [h_0(t_{ij}) v \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})]^{\delta_{ij}} \exp\{-[H_0(t_{ij}) - H_0(t_{ij0})] v \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})\} \times \\ &\qquad dG(v|t_{i1} > t_{i10}, ..., t_{in_i} > t_{in_i0}) \end{aligned}$$

For instance, suppose v follows gamma distribution then:

$$\begin{aligned} g(v|t_{i1} > t_{i10}, ..., t_{in_i} > t_{in_i0}) &= \frac{g(v) S_i(t_{ij0}|v)}{S_i(t_{ij0})} \\ &= \frac{P(t_{i1} \geq t_{i10}, ..., t_{in_i} \geq t_{in_i0}, v)}{P(t_{i1} \geq t_{i10}, ...t_{in_i} \geq t_{in_i0})} \\ &= \frac{\exp(-v \sum_{j=1}^{n_i} H_0(t_{ij0}) \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})) \frac{v^{1/\theta - 1} \exp(-v/\theta)}{\theta^{1/\theta} \Gamma_\theta}}{\int_v \exp(-v \sum_{j=1}^{n_i} H_0(t_{ij0}) \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})) \frac{v^{1/\theta - 1} \exp(-v/\theta)}{\theta^{1/\theta} \Gamma_\theta} dv} \\ &= \frac{\exp(-v \sum_{j=1}^{n_i} H_0(t_{ij0}) \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})) v^{1/\theta - 1} \exp(-v/\theta)}{\int_v \exp(-v \sum_{j=1}^{n_i} H_0(t_{ij0}) \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})) v^{1/\theta - 1} \exp(-v/\theta)\, dv} \end{aligned}$$

$$= \frac{v^{1/\theta-1}\exp\left(-v\left[\frac{1}{\theta}+\sum_{j=1}^{n_i}H_0(t_{ij0})\exp\left(\boldsymbol{x}_{ij}\boldsymbol{\beta}\right)\right]\right)}{\int_v v^{1/\theta-1}\exp\left(-v\left[\frac{1}{\theta}+\sum_{j=1}^{n_i}H_0(t_{ij0})\exp\left(\boldsymbol{x}_{ij}\boldsymbol{\beta}\right)\right]\right)dv}$$

$$= \frac{v^{1/\theta-1}\exp\left(-v\left[\frac{1}{\theta}+\sum_{j=1}^{n_i}H_0(t_{ij0})\exp\left(\boldsymbol{x}_{ij}\boldsymbol{\beta}\right)\right]\right)}{\frac{\Gamma_\theta}{\left[\frac{1}{\theta}+\sum_{j=1}^{n_i}H_0(t_{ij0})\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})\right]^{\frac{1}{\theta}}}}$$

Since $\int_0^\infty e^{-vB}v^{\kappa-1}dv = \frac{\Gamma(\kappa)}{B^\kappa}$ the denominator of equation above simplifies and thus

$$g(v|t_{i1} > t_{i10}, ..., t_{in_i} > t_{in_i0}) = \frac{\left[\frac{1}{\theta}+\sum_{j=1}^{n_i}H_0(t_{ij0})\exp\left(\boldsymbol{x}_{ij}\boldsymbol{\beta}\right)\right]^{\frac{1}{\theta}}}{\Gamma_\theta}v^{1/\theta-1} \times \exp\left(-v\left[\frac{1}{\theta}+\sum_{j=1}^{n_i}H_0(t_{ij0})e^{\boldsymbol{x}_{ij}\boldsymbol{\beta}}\right]\right) \quad (7.1)$$

where $\Gamma_\theta = \Gamma\left(\frac{1}{\theta}\right)$. Thus if v follows gamma distribution then at the delayed entry times $v \sim \Gamma(\frac{1}{\theta}, \frac{1}{\theta}+\sum_{j=1}^{n_i}H_0(t_{0ij})\exp\left(\boldsymbol{x}_{ij}\boldsymbol{\beta}\right))$ as shown above in equation (7.1). Note that this updated method assumes fully observed clusters i.e all n_i entered study (Rodríguez-Girondo et al., 2018).

Substituting $g(v|t_{i1} > t_{i10}, ..., t_{in_i} > t_{in_i0})$ obtained in Equation (7.1) gives the following marginal likelihood for cluster i

$$\begin{aligned}
L_i &= \int_v \left[\prod_{j=1}^{n_i}[h_0(t_{ij})v\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})]^{\delta_{ij}}\exp\{-[H_0(t_{ij})-H_0(t_{ij0})]v\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})\}\right] \\
&\quad \times dG(v|t_{i1} > t_{i10}, ..., t_{in_i} > t_{in_i0}) \\
&= \int_v \left[\prod_{j=1}^{n_i}[h_0(t_{ij})v\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})]^{\delta_{ij}}\exp\{-[H_0(t_{ij})-H_0(t_{ij0})]v\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})\}\right] \\
&\quad \times \frac{\left[\frac{1}{\theta}+\sum_{j=1}^{n_i}H_0(t_{ij0})\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})\right]^{\frac{1}{\theta}}}{\Gamma_\theta}v^{1/\theta-1}\exp\left(-v\left[\frac{1}{\theta}+\sum_{j=1}^{n_i}H_0(t_{ij0})\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})\right]\right)dv \\
&= \prod_{j=1}^{n_i}[h_0(t_{ij})\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})]^{\delta_{ij}}\left(\frac{1}{\theta}\right)^{\frac{1}{\theta}}\frac{\left[1+\theta\sum_{j=1}^{n_i}H_0(t_{ij0})\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})\right]^{\frac{1}{\theta}}}{\Gamma_\theta} \\
&\quad \times \int_v v^{d_i+\frac{1}{\theta}-1}\exp\left(-v\left[\frac{1}{\theta}+\sum_{j=1}^{n_i}H_0(t_{ij0})\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})+\sum_{j=1}^{n_i}[H_0(t_{ij})-H_0(t_{ij0})]\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})\right]\right)dv \\
&= \prod_{j=1}^{n_i}[h_0(t_{ij})\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})]^{\delta_{ij}}\left(\frac{1}{\theta}\right)^{\frac{1}{\theta}}\frac{\left[1+\theta\sum_{j=1}^{n_i}H_0(t_{ij0})\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})\right]^{\frac{1}{\theta}}}{\Gamma_\theta} \\
&\quad \times \int_v v^{d_i+\frac{1}{\theta}-1}\exp\left(-v\left[\frac{1}{\theta}+\sum_{j=1}^{n_i}H_0(t_{ij})\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})\right]\right)dv
\end{aligned}$$

$$= \prod_{j=1}^{n_i} [h_0(t_{ij}) \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})]^{\delta_{ij}} \frac{\left[1+\theta\sum_{j=1}^{n_i} H_0(t_{ij0})\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})\right]^{\frac{1}{\theta}}}{\Gamma_\theta} \frac{\Gamma_{\theta'}}{\left(\frac{1}{\theta}\right)^{d_i}\left[1+\theta\sum_{j=1}^{n_i} H_0(t_{ij})\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})\right]}$$

where $\Gamma_{\theta'} = \Gamma\left(d_i + \frac{1}{\theta}\right)$ and $d_i = \sum_{j=1}^{n_i} \delta_{ij}$.

Taking the log we obtain the marginal loglikelihood function for cluster i as follows:

$$\begin{aligned} l_i(\boldsymbol{\xi}, \theta, \boldsymbol{\beta}) &= d_i \log(\theta) - \log(\Gamma_\theta) + \log\left(\Gamma_{\theta'}\right) \\ &\quad - \left(\frac{1}{\theta} + d_i\right) \log\left(1 + \theta \sum_{j=1}^{n_i} H_0(t_{ij}) \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})\right) \\ &\quad + \sum_{j=1}^{n_i} \delta_{ij} \left[\boldsymbol{x}_{ij}\boldsymbol{\beta} + \log(h_0(t_{ij}))\right] + \frac{1}{\theta} \log\left(1 + \theta \sum_{j=1}^{n_i} H_0(t_{ij0}) \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})\right) . \end{aligned} \quad (7.2)$$

The marginal loglikelihood function for all the clusters is

$$l(\boldsymbol{\xi}, \theta, \boldsymbol{\beta}) = \sum_{i=1}^{G} l_i(\boldsymbol{\xi}, \theta, \boldsymbol{\beta}) \quad (7.3)$$

The maximum likelihood estimates $\boldsymbol{\xi}, \theta, \boldsymbol{\beta}$ are obtained by maximizing the loglikelihood function given in equation (7.3). Numerical methods can be used to maximize this loglikelihood function e.g by Newton-Raphson procedure.

7.2 Shared frailty model incorporating time varying covariates

Let $\boldsymbol{Y}_{ij} = \{Y_{ij}(t_{ijl}); l = 1, \ldots, n_{ij}\}$ be a vector of length n_{ij} representing individual specific time varying covariate measurements. Let α be a time varying covariate effect. Let $\boldsymbol{\xi}$ be the parameters of the baseline hazard $h_0(t)$ and θ be the frailty variance. The hazard at time t for the j^{th} individual in the i^{th} group conditional on v_i is given by

$$h_{ij}(t|\boldsymbol{Y}_{ij}, v_i) = h_0(t) v_i \exp(\alpha y_{ij}(t)) .$$

The risk parameter α represents the effect on the hazard of a unit difference in the covariate at time zero or at any time after entry.

Methods of modelling a time dependent covariate discussed here include last observation carried forward in Section 7.2.1, two stage approaches in Section 7.2.2 and joint models in Section 7.2.3.

7.2.1 Last observation carried forward (LOCF)

Suppose the time varying covariate $y_{ij}(t)$ takes the value y_{ijl} within a given time interval $(t_{ijl}, t_{ij(l+1)})$ at timepoints $l = 0, ..., (n_{ij}-1)$. The contribution of an individual j in the i^{th} cluster within a given time interval $(t_{ijl}, t_{ij(l+1)})$ in presense of right censoring conditional on v_i is given by (Klein and Moeschberger, 2003)

$$f(t_{ij(l+1)}|y_{ijl}, v_i)^{\delta_{ijl}} S(t_{ij(l+1)}|y_{ijl}, v_i)^{1-\delta_{ijl}} / S(t_{ijl}|y_{ijl}, v_i).$$

Note that for an individual to contribute to the likelihood within the interval $(t_{ijl}, t_{ij(l+1)})$ an individual needs to have survived until beyond time t_{ijl}.

Therefore the likelihood for cluster i conditional on v_i will be

$$\begin{aligned}
L_i &= \prod_{j=1}^{n_i} \prod_{l=0}^{n_{ij}-1} \frac{f(t_{ij(l+1)}|y_{ijl}, v_i)^{\delta_{ijl}} S(t_{ij(l+1)}|y_{ijl}, v_i)^{1-\delta_{ijl}}}{S(t_{ijl}|y_{ijl}, v_i)} \\
&= \prod_{j=1}^{n_i} \prod_{l=0}^{n_{ij}-1} \frac{h(t_{ij(l+1)}|y_{ijl}, v_i)^{\delta_{ijl}} S(t_{ij(l+1)}|y_{ijl}, v_i)^{\delta_{ijl}} S(t_{ij(l+1)}|y_{ijl}, v_i)^{1-\delta_{ijl}}}{S(t_{ijl}|y_{ijl}, v_i)} \\
&= \prod_{j=1}^{n_i} \prod_{l=0}^{n_{ij}-1} \frac{h(t_{ij(l+1)}|y_{ijl}, v_i)^{\delta_{ijl}} S(t_{ij(l+1)}|y_{ijl}, v_i)}{S(t_{ijl}|y_{ijl}, v_i)} \\
&= \prod_{j=1}^{n_i} \prod_{l=0}^{n_{ij}-1} \frac{h(t_{ij(l+1)}|y_{ijl}, v_i)^{\delta_{ijl}} \exp(-H(t_{ij(l+1)}|y_{ijl}, v_i))}{\exp(-H(t_{ijl}|y_{ijl}, v_i))} \\
&= \prod_{j=1}^{n_i} \prod_{l=0}^{n_{ij}-1} h(t_{ij(l+1)}|y_{ijl}, v_i)^{\delta_{ijl}} \exp(-[H(t_{ij(l+1)}|y_{ijl}, v_i) - H(t_{ijl}|y_{ijl}, v_i)]) \\
&= \prod_{j=1}^{n_i} \prod_{l=0}^{n_{ij}-1} [h_0(t_{ij(l+1)}) v_i \exp(\alpha y_{ijl})]^{\delta_{ijl}} \exp(-[H_0(t_{ij(l+1)}) - H_0(t_{ijl})] v_i e^{\alpha y_{ijl}}) \quad (7.4)
\end{aligned}$$

The marginal likelihood for cluster i will be

$$\begin{aligned}
L_i &= \int_v \prod_{j=1}^{n_i} \prod_{l=0}^{n_{ij}-1} [h_0(t_{ij(l+1)}) v \exp(\alpha y_{ijl})]^{\delta_{ijl}} \\
&\quad \times \exp(-[H_0(t_{ij(l+1)}) - H_0(t_{ijl})] v \exp(\alpha y_{ijl})) dg(v) \quad (7.5)
\end{aligned}$$

When a gamma frailty distribution is considered the the marginal likelihood is as follows:

$$
\begin{aligned}
L_i &= \int_v \left[\prod_{j=1}^{n_i} \prod_{l=0}^{n_{ij}-1} [h_0(t_{ij(l+1)}) v \exp(\alpha y_{ijl})]^{\delta_{ijl}} \right. \\
&\qquad \left. \times \exp(-[H_0(t_{ij(l+1)}) - H_0(t_{ijl})] v \exp(\alpha y_{ijl})) \right] \frac{v^{1/\theta - 1} \exp(-v/\theta)}{\theta^{\frac{1}{\theta}} \Gamma_\theta} dv \\
&= \prod_{j=1}^{n_i} \prod_{l=0}^{n_{ij}-1} [h_0(t_{ijl}) \exp(y_{ijl}\boldsymbol{\beta})]^{\delta_{ijl}} \left(\frac{1}{\theta}\right)^{\frac{1}{\theta}} \frac{1}{\Gamma_\theta} \\
&\qquad \times \int_v v^{d_i + \frac{1}{\theta} - 1} \exp\left(-v \left[\frac{1}{\theta} + \sum_{j=1}^{n_i} \sum_{l=0}^{n_{ij}-1} [H_0(t_{ij(l+1)}) - H_0(t_{ijl})] \exp(\alpha y_{ijl}) \right] \right) dv \\
&= \prod_{j=1}^{n_i} \prod_{l=0}^{n_{ij}-1} [h_0(t_{ijl}) \exp(\alpha y_{ijl})]^{\delta_{ijl}} \frac{1}{\Gamma_\theta} \\
&\qquad \times \frac{\Gamma_{\theta'}}{\left(\frac{1}{\theta}\right)^{d_i} \left(1 + \theta \sum_{j=1}^{n_i} \sum_{l=0}^{n_{ij}-1} [H_0(t_{ij(l+1)}) - H_0(t_{ijl})] \exp(\alpha y_{ijl}) \right)^{d_i + \frac{1}{\theta}}}
\end{aligned}
$$

where $d_i = \sum_{j=1}^{n_i} \sum_{l=0}^{n_{ij}-1} \delta_{ijl}$.

Taking the log we obtain the marginal loglikelihood function for cluster i as follows:

$$
\begin{aligned}
l_i(\boldsymbol{\xi}, \theta, \alpha) &= d_i \log(\theta) - \log(\Gamma_\theta) + \log\left(\Gamma_{\theta'}\right) \\
&\quad + \sum_{j=1}^{n_i} \sum_{l=0}^{n_{ij}-1} \delta_{ijl} \left[y_{ijl}\boldsymbol{\beta} + \log\left(h_0(t_{ij(l+1)}) \right) \right] \\
&\quad - \left(\frac{1}{\theta} + d_i \right) \log\left(1 + \theta \left[\sum_{j=1}^{n_i} \sum_{l=0}^{n_{ij}-1} [H_0(t_{ij(l+1)}) - H_0(t_{ijl})] \exp(y_{ijl}\boldsymbol{\beta}) \right] \right).
\end{aligned}
$$

The marginal loglikelihood function for all the clusters is

$$
l(\boldsymbol{\xi}, \theta, \alpha) = \sum_{i=1}^{G} l_i(\boldsymbol{\xi}, \theta, \alpha) \tag{7.6}
$$

The maximum likelihood estimates $\boldsymbol{\xi}, \theta, \alpha$ are estimated by maximizing the loglikelihood function given in equation (7.6). Numerical methods are used to maximize this loglikelihood function e.g by Newton-Raphson procedure.

7.2.2 A two stage approach (TSA)

This section is an extension of Section 3.2 on singletons to considering clustered data. We consider the two different two regression calibration methods: risk-set regression calibration and the ordinary regression calibration.

7.2.2.1 Risk set regression calibration (RRC)

Stage I

For each of the unique event times in the dataset we fit a mixed-effects model. For an event time T_r only individuals still at risk i.e $\{ij : T_{ij} \geq T_r\}$ are included in the mixed model, and only the covariate measurements of these individuals taken before the event time are used i.e. $\boldsymbol{Y}_{ij(T_r)} = \{Y_{ij}(t_{ijl}) : t_{ijl} < T_r\}$. We then compute the predicted value of the covariate for individuals still at risk at event time T_r as follows: $Y^*_{ij}(T_r) = \mathbb{E}\{Y_{ij}(T_r)|Y_{ij}(t_{ijl}) : t_{ijl} < T_r\}, \hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2_b, \hat{\sigma}^2_u\}$

Linear mixed model specification

In this thesis we consider the following model

$$y_{ij}(t \mid b_{ij}, u_i) = \beta_0 + b_{ij} + u_i + \beta_1 t + \varepsilon_{ij}(t) \tag{7.7}$$

$b_{ij} \sim N(0, \sigma^2_b)$, where b_{ij} represents the individual specific random effect

$u_i \sim N(0, \sigma^2_u)$, where u_i represents the cluster specific random effect

$\varepsilon_{ijl} \sim N(0, \sigma^2_\varepsilon)$,

β_0, β_1 : the regression coefficients.They represent the population intercept and slope respectively.

b_{ij}, u_i and $\boldsymbol{\varepsilon}_{ij}$ are independent of each other

This is a three-level linear mixed model and could be represented in matrix form as follows:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Zd} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{X}$ is $m \times 2$ design matrix for the fixed effects and $\boldsymbol{Z}$ is $m \times (n+G)$ design matrix associated to the random effects (both b_{ij} and u_i). Here $n = \sum_{i=1}^{G}\sum_{j=1}^{n_i} n_{ij}$ is the total number of subjects and $m = \sum_{i=1}^{G}\sum_{j=1}^{n_i}\sum_{l=0}^{n_{ij}-1} n_{ijl}$ is the total number of responses for all subjects.

Illustration of model representation: assume groups of size two i.e. ($n_i = 2$). Suppose we have $G = 3$ groups and two measurements taken per individual (i.e $n_{ij} = 2$) then the model would be:

$$\boldsymbol{Y} = \begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ \dots \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \\ \dots \\ y_{311} \\ y_{312} \\ y_{321} \\ y_{322} \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} 1 & t_{111} \\ 1 & t_{112} \\ 1 & t_{121} \\ 1 & t_{122} \\ \dots & \dots \\ 1 & t_{211} \\ 1 & t_{212} \\ 1 & t_{221} \\ 1 & t_{222} \\ \dots & \dots \\ 1 & t_{311} \\ 1 & t_{312} \\ 1 & t_{321} \\ 1 & t_{322} \end{bmatrix}, \quad \boldsymbol{Z} = \left[\begin{array}{cccccc:ccc} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ \multicolumn{9}{c}{\dots} \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ \multicolumn{9}{c}{\dots} \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right], \quad \boldsymbol{d} = \begin{bmatrix} b_{11} \\ b_{12} \\ b_{21} \\ b_{22} \\ b_{31} \\ b_{32} \\ \dots \\ u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

Let $\boldsymbol{\Sigma_d} = \text{var}(\boldsymbol{d}) = \text{var}\begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{u} \end{bmatrix} = \begin{bmatrix} \sigma_b^2 \mathbb{1}_n & \boldsymbol{0} \\ \boldsymbol{0} & \sigma_u^2 \mathbb{1}_G \end{bmatrix}$, $\text{var}(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2 \mathbb{1}_m$ and $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 & \beta_1 \end{bmatrix}$.

Linear Mixed Model Estimation

Maximum likelihood estimates of $\mathbf{\Phi} = (\beta_0, \beta_1, \sigma_b^2, \sigma_u^2, \sigma_\varepsilon^2)$ are obtained by maximizing the log-likelihood function.

$$\ell(\mathbf{\Phi}) = \sum_{i=1}^{G} \log \int_u \Big\{ \prod_{j=1}^{n_i} \int_b f(\mathbf{y}_{ij} \mid b_{ij}, u_i) f(b_{ij}) \mathrm{d}b_{ij} \Big\} f(u_i) \mathrm{d}u_i, \tag{7.8a}$$

where

$$f\{y_{ijl} \mid b_{ij}\} = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\Big(- \frac{[y_{ijl} - (\beta_0 + b_{ij} + u_i + \beta_1 t_{ijl})]^2}{2\sigma_\varepsilon^2} \Big)$$

$$f(\mathbf{y}_{ij} \mid b_{ij}, u_i) = \prod_{l=1}^{n_{ij}} f(y_{ijl} \mid b_{ij}, u_i) = (\sqrt{2\pi\sigma_\varepsilon^2})^{-n_{ij}/2} \exp\Big(- \frac{\sum_{l=1}^{n_{ij}} [y_{ijl} - (\beta_0 + b_{ij} + u_i + \beta_1 t_{ijl})]^2}{2\sigma_\varepsilon^2} \Big)$$

$$f(b_{ij}) = \frac{1}{\sqrt{2\pi\sigma_b^2}} \exp\Big(- \frac{[b_{ij}]^2}{2\sigma_b^2} \Big)$$

$$f(u_i) = \frac{1}{\sqrt{2\pi\sigma_u^2}} \exp\Big(- \frac{[u_i]^2}{2\sigma_u^2} \Big)$$

Estimation of $\mathbf{\Phi}$ is done iteratively by splitting into the parameters of the fixed effect (β_0, β_1), and the variance parameters $(\sigma_b^2, \sigma_u^2, \sigma_\varepsilon^2)$.

Note that $\boldsymbol{Y} \sim \mathcal{N}\left(\boldsymbol{X\beta}, \boldsymbol{Z\Sigma_d Z}^\top + \sigma_\varepsilon^2 \mathbb{1}_m\right)$ while $\boldsymbol{Y}|\boldsymbol{d} \sim \mathcal{N}\left(\boldsymbol{X\beta} + \boldsymbol{Zd}, \sigma_\varepsilon^2 \mathbb{1}_m\right)$.

Let $\boldsymbol{\Sigma_y} = \mathrm{var}(\boldsymbol{Y}) = \boldsymbol{Z\Sigma_d Z}^\top + \sigma_\varepsilon^2 \mathbb{1}_m$.

The maximum likelihood estimator of the fixed-effects $\boldsymbol{\beta}$ and estimators of the variance parameters $(\sigma_b^2, \sigma_u^2, \sigma_\varepsilon^2)$ are obtained similarly to Section 3.2.1.2.

The random effects are predicted by obtaining the expectation of the posterior distribution of the random effects given the observed data (Fitzmaurice et al., 2012; Laird and Ware, 1982).

$$\begin{aligned} \mathbb{E}(\boldsymbol{d}|\boldsymbol{Y}) &= \mathbb{E}(\boldsymbol{d}) + \mathrm{cov}(\boldsymbol{d}; \boldsymbol{Y})\, \mathrm{var}(\boldsymbol{Y})^{-1} [\boldsymbol{Y} - \mathbb{E}\boldsymbol{Y}] \\ &= \mathbf{0} + \boldsymbol{\Sigma_d Z}^\top \Sigma_d Z^\top \boldsymbol{\Sigma_y}^{-1} [\boldsymbol{Y} - \boldsymbol{X\beta}] \end{aligned} \tag{7.9}$$

The parameters $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma_y}$ are replaced by their estimators in equation (7.9) to obtain the empirical Bayes estimator, $\hat{\boldsymbol{d}} = \boldsymbol{\Sigma}_d Z^\top \hat{\boldsymbol{\Sigma}_{\boldsymbol{y}}}^{-1} [\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}]$

Stage II

Let $Y_{ij}(t) = Y^*_{ij}(t)+\varepsilon_{ij}(t)$ where $Y^*_{ij}(t) = \mathbb{E}[Y_{ij}(t)|\boldsymbol{Y}_{ij(t)}]$. Similar to Equation (**??**) while assuming non-differential measurement error assumption the induced hazard function based on observed covariate measurements is given by:

$$h_{ij}\left(t|\boldsymbol{Y}_{ij}, v_i\right) = h_0(t)v_i\mathbb{E}(e^{\alpha Y^*_{ij}(t)}|\boldsymbol{Y}_{ij}, t_{ij} \geq t) \tag{7.10}$$

Similar to Yu et al. (2018) and Dafni and Tsiatis (1998) we propose to approximate this expectation by a first-order approximation i.e the regression calibration approximation.

$$\begin{aligned} h_{ij}\left(t|\boldsymbol{Y}_{ij}, v_i\right) &= h_0(t)v_i\mathbb{E}[e^{\alpha Y^*_{ij}(t)}|\boldsymbol{Y}_{ij(t)}, t_{ij} \geq t] \\ &\approx h_0(t)v_i e^{\alpha\mathbb{E}[Y_{ij}(t)|\boldsymbol{Y}_{ij(t)}, t_{ij}\geq t]} \\ &= h_0(t)v_i e^{\alpha Y^*_{ij}(t)} \end{aligned} \tag{7.11}$$

Note that $Y^*_{ij}(t) = \mathbb{E}[Y_{ij}(t)|\boldsymbol{Y}_{ij(t)}, t_{ij} \geq t]$ which was obtained in stage I.

The corresponding likelihood for cluster i conditional on v_i will therefore be

$$L_i(\boldsymbol{\xi}, \theta, \alpha) = \prod_{j=1}^{n_i}\prod_{l=0}^{n_{ij}-1}[h_0(t_{ij(l+1)})v_i\exp\left(\alpha y^*_{ijl}\right)]^{\delta_{ijl}}\exp\left(-[H_0(t_{ij(l+1)}) - H_0(t_{ijl})]v_i\exp\left(\alpha y^*_{ijl}\right)\right)$$

For an individual j in the i^{th} cluster we can therefore arrange the data obtained from Stage I in start stop format with intervals $(t_{ijl}, t_{ij(l+1)})$.

The marginal log-likelihood function for cluster i will be:

$$\begin{aligned} l_i\left(\boldsymbol{\xi}, \theta, \alpha\right) &= d_i\log(\theta) - \log\left(\Gamma_\theta\right) + \log\left(\Gamma_{\theta'}\right) \\ &\quad + \sum_{j=1}^{n_i}\sum_{l=0}^{n_{ij}-1}\delta_{ijl}\left[\alpha y^*_{ijl} + \log\left(h_0(t_{ij(l+1)})\right)\right] \\ &\quad - \left(\frac{1}{\theta} + d_i\right)\log\left(1 + \theta\left[\sum_{j=1}^{n_i}\sum_{l=0}^{n_{ij}-1}[H_0(t_{ij(l+1)}) - H_0(t_{ijl})]\exp(\alpha y^*_{ijl})\right]\right). \end{aligned}$$

where $d_i = \sum_{j=1}^{n_i}\sum_{l=0}^{n_{ij}-1}\delta_{ijl}$ and $\Gamma\left(1/\theta\right) = \int_0^\infty e^{-t}t^{1/\theta-1}$.

The marginal loglikelihood function for all the clusters is

$$l\left(\boldsymbol{\xi}, \theta, \alpha\right) = \sum_{i=1}^{G} l_i\left(\boldsymbol{\xi}, \theta, \alpha\right) \tag{7.12}$$

Numerical methods are used to maximize this loglikelihood function e.g by Newton-Raphson procedure.

7.2.2.2 Ordinary regression calibration (ORC)

We propose to extend ordinary regression calibration for (Tsiatis et al., 1995) which fits a single mixed-effects model using singletons and all data to the case of shared frailty model. In this approach the estimated random effects for an individual stay constant over follow-up.

Stage I

We fit one mixed-effects models using all available covariate measurements. We then compute the predicted value of the covariate for individuals as follows:

$$Y_{ij}^*(t) = \mathbb{E}\{Y_{ij}(t)|\boldsymbol{Y}, \hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_b^2, \hat{\sigma}_u^2\}.$$

Stage II

Stage II proceeds similarly to the RRC procedure discussed in section 7.2.2.1 where the prediction $Y_{ij}^*(t) = \mathbb{E}\{Y_{ij}(t)|\boldsymbol{Y}, \hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_b^2, \hat{\sigma}_u^2\}$ are now used.

7.2.3 Joint model approach

Let the longitudinal model have random intercepts only (as described in Equation (7.7)) and assume a link function of current value i.e $\beta_0 + b_{ij} + u_i + \beta_1 t$. This function of random effects will model correlation between longitudinal and the survival outcomes.

The standard joint model for a longitudinal and a time-to-event outcome will be

$$\begin{aligned} y_{ij}(t \mid u_i) &= \beta_0 + b_{ij} + u_i + \beta_1 t + \varepsilon_{ij}(t) \\ h_{ij}(t \mid b_{ij}, u_i, v_i) &= h_0(t) v_i \exp\left\{\alpha(\beta_0 + b_{ij} + u_i + \beta_1 t)\right\} \quad i = 1, 2, ..., G \text{ and } j = 1, ..., n_i. \end{aligned}$$

α : the regression coefficient of the time-varying covariate

$y_{ijl} \perp t_{ij} | b_{ij}, u_i$

Suppose that data, $\mathcal{D} = \{t_{ij}, \delta_{ij}, \mathbf{y}_{ij}\}$, of both longitudinal and time-to-event outcomes are collected on subjects $j = 1, \ldots, n_i$ and $i = 1, \ldots, G$.

The estimation of $\boldsymbol{\Phi} = (\boldsymbol{\xi}, \theta, \alpha, \sigma_u^2, \sigma_\varepsilon^2)$ is based on maximum likelihood principles by maximizing the log-likelihood function of the joint distribution of the longitudinal and the

time-to-event outcomes, $\{\mathbf{y}_{ij}, t_{ij}, \delta_{ij}\}$. ξ contains the parameters of the baseline hazard $h_0(t)$.

$$\ell(\mathbf{\Phi}|\mathcal{D}) = \sum_{i=1}^{G} \log\Bigg(\int_u \int_v \Big[\prod_{j=1}^{n_i} \int_b f(\mathbf{y}_{ij} \mid b_{ij}, u_i) f(t_{ij}, \delta_{ij} \mid b_{ij}, u_i, v) f(b_{ij}) \mathrm{d}b_{ij} \Big] g(v_i) f(u_i) \mathrm{d}v_i \mathrm{d}u_i \Bigg), \tag{7.13}$$

where

$$\begin{aligned}
f\{y_{ijl} \mid u_i\} &= \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\Big(- \frac{[y_{ijl} - (\beta_0 + b_{ij} + u_i + \beta_1 t_{ijl})]^2}{2\sigma_\varepsilon^2} \Big) \\
f(\mathbf{y}_{ij} \mid u_i) &= \prod_{l=1}^{n_{ij}} f(y_{ijl} \mid u_i) = (2\pi\sigma_\varepsilon^2)^{-n_{ij}/2} \exp\Big(- \frac{\sum_{l=1}^{n_{ij}} [y_{ijl} - (\beta_0 + b_{ij} + u_i + \beta_1 t_{ijl})]^2}{2\sigma_\varepsilon^2} \Big) \\
f(b_{ij}) &= \frac{1}{\sqrt{2\pi\sigma_b^2}} \exp\Big(- \frac{[b_{ij}]^2}{2\sigma_b^2} \Big) \\
f(u_i) &= \frac{1}{\sqrt{2\pi\sigma_u^2}} \exp\Big(- \frac{[u_i]^2}{2\sigma_u^2} \Big) \\
f(t_{ij}, \delta_{ij} \mid u_i, v_i) &= \left[h_{ij}\left(t_{ij} \mid \boldsymbol{Y}_{ij}, v_i\right) \right]^{\delta_{ij}} S_{ij}\left(t_{ij} \mid \boldsymbol{Y}_{ij}, v_i\right), \\
h_{ij}\left(t_{ij} \mid \boldsymbol{Y}_{ij}\right), v_i) &= h_0(t_{ij}) v_i \exp\left\{\alpha(\beta_0 + b_{ij} + u_i + \beta_1 t_{ij})\right\}, \\
S_{ij}\left(t_{ij} \mid \boldsymbol{Y}_{ij}, v_i\right) &= \exp\left(- \int_0^{t_{ij}} h_{ij}\left(s \mid \boldsymbol{Y}_{ij}, vi\right) \mathrm{d}s. \right) \\
h_0(t) &= \lambda\rho t^{\rho-1}
\end{aligned}$$

Numerical methods can be used to maximize this loglikelihood function e.g by Newton-Raphson procedure. Note that the integrals of the random effects require numerical methods (with the Gauss-Hermite quadrature technique the most commonly used).

7.3 Shared frailty model incorporating time varying covariates and delayed entry

Section 7.1 models delayed entry while Section 7.2 models time varying covariates. This section will simultaneously model time varying covariates and delayed entry. Let T_{ij0} be random variable representing the age at enrollment to study. An individual is included in the study if and only if $t_{ij} > t_{ij0}$.

The hazard at time t for the j^{th} individual in the i^{th} group conditional on v_i is given by

$$h_{ij}\left(t|\boldsymbol{Y}_{ij}, v_i\right) \quad = \quad h_0(t) v_i \exp\left(\alpha y_{ij}(t)\right).$$

Methods of modelling time dependent covariates simultaneously with delayed entry discussed in this chapter are last observation carried forward as will be discussed in 7.3.1 , two stage approaches as will be discussed in section 7.3.2 & section 7.3.3 and joint models as will be discussed in section 7.3.4.

7.3.1 Last observation carried forward (LOCF)

When we have time varying covariates then the likelihood for cluster i conditional on v_i as shown in Equation (7.4) of section 7.2.1 is

$$L_i = \prod_{j=1}^{n_i}\prod_{l=0}^{n_{ij}-1}[h_0(t_{ij(l+1)})v_i\exp(\alpha y_{ijl})]^{\delta_{ijl}}\exp(-[H_0(t_{ij(l+1)})-H_0(t_{ijl})]v_i e^{\alpha y_{ijl}}) \quad (7.14)$$

However, because of delayed entry let the first time interval given by $(t_{ij(l=0)}, t_{ij(l=1)})$ be represented by (t_{ij0}, t_{ij1}) and let the time varying covariate measurement at this interval be y_{ij0}.

The updated frailty distribution in this case is similar to Equation (7.1) by replacing $\boldsymbol{x}_{ij}\boldsymbol{\beta}$ with αy_{ij0}. Therefore, if v follows gamma distribution then at delayed entry times $v \sim \Gamma(\frac{1}{\theta}, \frac{1}{\theta}+\sum_{j=1}^{n_i} H_{ij}(t_{ij0})e^{\alpha y_{ij0}})$.

Therefore the marginal likelihood for cluster i will be

$$\begin{aligned}
L_i &= \int_v \left[\prod_{j=1}^{n_i}\prod_{l=0}^{n_{ij}-1}[h_0(t_{ij(l+1)})v\exp(\alpha y_{ijl})]^{\delta_{ijl}}\exp(-[H_0(t_{ij(l+1)})-H_0(t_{ijl})]v\exp(\alpha y_{ijl}))\right] \\
&\quad \times dG(v|t_{i1} > t_{i10}, ..., t_{in_i} > t_{in_i 0}) \quad (7.15)\\
&= \int_v \left[\prod_{j=1}^{n_i}\prod_{l=0}^{n_{ij}-1}[h_0(t_{ij(l+1)})v\exp(\alpha y_{ijl})]^{\delta_{ijl}}\exp(-[H_0(t_{ij(l+1)})-H_0(t_{ijl})]v\exp(\alpha y_{ijl}))\right] \\
&\quad \times \frac{\left[\frac{1}{\theta}+\sum_{j=1}^{n_i} H_{ij}(t_{ij0})e^{\alpha y_{ij0}}\right]^{\frac{1}{\theta}}}{\Gamma_\theta} v^{1/\theta-1}\exp\left(-v\left[\frac{1}{\theta}+\sum_{j=1}^{n_i} H_{ij}(t_{ij0})e^{\alpha y_{ij0}}\right]\right)dv \\
&= \left\{\prod_{j=1}^{n_i}\prod_{l=0}^{n_{ij}-1}[h_0(t_{ij(l+1)})\exp(\alpha y_{ijl})]^{\delta_{ijl}}\right\}\frac{\left[\frac{1}{\theta}+\sum_{j=1}^{n_i} H_{ij}(t_{ij0})e^{\alpha y_{ij0}}\right]^{\frac{1}{\theta}}}{\Gamma_\theta} \\
&\quad \times \int_v v^{d_i+1/\theta-1}\exp\left(-v\sum_{j=1}^{n_i}\sum_{l=0}^{n_{ij}-1}[H_0(t_{ij(l+1)})-H_0(t_{ijl})]\exp(\alpha y_{ijl})\right) \\
&\quad \times \exp\left(-v\left[\frac{1}{\theta}+\sum_{j=1}^{n_i} H_{ij}(t_{ij0})e^{\alpha y_{ij0}}\right]\right)dv
\end{aligned}$$

$$
\begin{aligned}
&= \left\{\prod_{j=1}^{n_i}\prod_{l=0}^{n_{ij}-1}\left[h_0(t_{ij(l+1)})\exp\left(\alpha y_{ijl}\right)\right]^{\delta_{ijl}}\right\}\frac{\left[\frac{1}{\theta}+\sum_{j=1}^{n_i}H_{ij}(t_{ij0})e^{\alpha y_{ij0}}\right]^{\frac{1}{\theta}}}{\Gamma_\theta}\\
&\quad\times\int_v v^{d_i+\frac{1}{\theta}-1}\exp\left(-v\left[\frac{1}{\theta}+\sum_{j=1}^{n_i}H_{ij}(t_{ij0})e^{\alpha y_{ij0}}+\sum_{j=1}^{n_i}\sum_{l=0}^{n_{ij}-1}\left[H_0(t_{ij(l+1)})-H_0(t_{ijl})\right]\exp\left(\alpha y_{ijl}\right)\right]\right)dv\\
&= \left\{\prod_{j=1}^{n_i}\prod_{l=0}^{n_{ij}-1}\left[h_0(t_{ij(l+1)})\exp\left(\alpha y_{ijl}\right)\right]^{\delta_{ijl}}\right\}\frac{\left[1+\theta\sum_{j=1}^{n_i}H_{ij}(t_{ij0})e^{\alpha y_{ij0}}\right]^{\frac{1}{\theta}}}{\Gamma_\theta}\\
&\quad\times\frac{\Gamma(d_i+1/\theta)}{\left(\frac{1}{\theta}\right)^{d_i}\left[1+\theta\sum_{j=1}^{n_i}H_{ij}(t_{ij0})e^{\alpha y_{ij0}}+\theta\sum_{j=1}^{n_i}\sum_{l=0}^{n_{ij}-1}\left[H_0(t_{ij(l+1)})-H_0(t_{ijl})\right]\exp\left(\alpha y_{ijl}\right)\right]^{d_i+\frac{1}{\theta}}}
\end{aligned}
\tag{7.16}
$$

Taking the log we obtain the marginal loglikelihood function for cluster i as follows:

$$
\begin{aligned}
l_i\left(\boldsymbol{\xi},\theta,\alpha\right) &= d_i\log(\theta)-\log\left(\Gamma_\theta\right)+\log\left(\Gamma_{\theta'}\right)\\
&\quad+\sum_{j=1}^{n_i}\sum_{l=0}^{n_{ij}-1}\delta_{ijl}\left[\alpha y_{ijl}+\log\left(h_0(t_{ij(l+1)})\right)\right]+\frac{1}{\theta}\log\left(1+\theta\sum_{j=1}^{n_i}H_{ij}(t_{ij0})e^{\alpha y_{ij0}}\right)\\
&\quad-\left(\frac{1}{\theta}+d_i\right)\log\left(1+\theta\sum_{j=1}^{n_i}H_{ij}(t_{ij0})e^{\alpha y_{ij0}}+\theta\sum_{j=1}^{n_i}\sum_{l=0}^{n_{ij}-1}\left[H_0(t_{ij(l+1)})-H_0(t_{ijl})\right]\exp\left(\alpha y_{ijl}\right)\right).
\end{aligned}
\tag{7.17}
$$

where $d_i=\sum_{j=1}^{n_i}\sum_{l=0}^{n_{ij}-1}\delta_{ijl}$ and $\Gamma\left(1/\theta\right)=\int_0^\infty e^{-t}t^{1/\theta-1}$.

The marginal loglikelihood function for all the clusters is

$$
l\left(\boldsymbol{\xi},\theta,\alpha\right) = \sum_{i=1}^{G}l_i\left(\boldsymbol{\xi},\theta,\alpha\right) \tag{7.18}
$$

The maximum likelihood estimates $\boldsymbol{\xi},\theta,\alpha$ are estimated by maximizing the loglikelihood function given in equation (7.18).

7.3.2 Risk set regression calibration (RRC)

Stage I

The difference of this case as compared to the case of no delayed entry (Section 7.2.2.1) is that we only observe individuals with $T_{ij}\geq T_{ij0}$ i.e. to be at risk at time T we require $\{T_{ij}\geq T\geq T_{ij0}\}$. Fit multiple mixed-effects models, one for each of the unique event time in the dataset. Compute the predicted value of the covariate for individuals still

at risk at event time T_r as follows: $Y^*_{ij}(T_r) = \mathbb{E}\{Y_{ij}(T_r)|Y_{ij}(t_{ijl}) : t_{ijl} < T_r,\ T_{ij} \geq T_r \geq T_{ij0}\}, \hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2_b, \hat{\sigma}^2_u\}$

Stage II

The hazard at time t, based on the history of the observed covariate measurements $\boldsymbol{Y}_{ij}$ up to time t can be expressed as the conditional expectation as follows (Prentice, 1982)

$$\begin{aligned} h_{ij}\left(t|\boldsymbol{Y}_{ij}\right) &= \mathbb{E}\left\{h_{ij}\left(t|\boldsymbol{Y}^*_{ij}, \boldsymbol{Y}_{ij}, t_{ij} \geq t \geq t_{ij0}\right)\right\} \\ &= \mathbb{E}\left\{h_{ij}\left(t|\boldsymbol{Y}^*_{ij}, t_{ij} \geq t \geq t_{ij0}\right)\right\} \text{ from the nondifferential measurement error assumption} \end{aligned}$$

The induced hazard function based on observed covariate measurements is:

$$h_{ij}\left(t|\boldsymbol{Y}_{ij}, v_i\right) = h_0(t)v_i\mathbb{E}(e^{\alpha Y^*_{ij}(t)}|\boldsymbol{Y}_{ij}, t_{ij} \geq t \geq t_{ij0}) \tag{7.19}$$

Similar to Section 7.2.2.1 we propose to approximate this expectation by a first-order approximation

$$\begin{aligned} h_{ij}\left(t|\boldsymbol{Y}_{ij}, v_i\right) &= h_0(t)v_i\mathbb{E}[e^{\alpha Y^*_{ij}(t)}|\boldsymbol{Y}_{ij}, t_{ij} \geq t \geq t_{ij0}] \\ &\approx h_0(t)v_i e^{\alpha\mathbb{E}[Y_{ij}(t)|\boldsymbol{Y}_{ij}, t_{ij} \geq t \geq t_{ij0}]} \\ &= h_0(t)v_i e^{\alpha Y^*_{ij}(t)} \end{aligned} \tag{7.20}$$

$Y^*_{ij}(t) = \mathbb{E}[Y_{ij}(t)|\boldsymbol{Y}_{ij(t)}, t_{ij} \geq t \geq t_{ij0}]$ which was obtained in stage I.

The corresponding likelihood for cluster i conditional on v_i will therefore be

$$L_i(\boldsymbol{\xi}, \theta, \alpha) = \prod_{j=1}^{n_i}\prod_{l=0}^{n_{ij}-1}[h_0(t_{ij(l+1)})v_i \exp\left(\alpha y^*_{ijl}\right)]^{\delta_{ijl}} \exp\left(-[H_0(t_{ij(l+1)}) - H_0(t_{ijl})]v_i \exp\left(\alpha y^*_{ijl}\right)\right)$$

For an individual j in the i^{th} cluster we can therefore arrange the data obtained from stage I in start stop format with intervals $(t_{ijl}, t_{ij(l+1)})$.

The marginal loglikelihood function for cluster i as follows (similar to LOCF likelihood obtained in equation (7.17)):

$$\begin{aligned} l_i\left(\boldsymbol{\xi}, \theta, \alpha\right) &= d_i \log(\theta) - \log\left(\Gamma_\theta\right) + \log\left(\Gamma_{\theta'}\right) \\ &+ \sum_{j=1}^{n_i}\sum_{l=0}^{n_{ij}-1}\delta_{ijl}\left[\alpha y^*_{ijl} + \log\left(h_0(t_{ij(l+1)})\right)\right] + \frac{1}{\theta}\log\left(1 + \theta\sum_{j=1}^{n_i}H_{ij}(t_{ij0})e^{\alpha y^*_{ij0}}\right) \\ &- \left(\frac{1}{\theta} + d_i\right)\log\left(1 + \theta\sum_{j=1}^{n_i}H_{ij}(t_{ij0})e^{\alpha y^*_{ij0}} + \theta\sum_{j=1}^{n_i}\sum_{l=0}^{n_{ij}-1}[H_0(t_{ij(l+1)}) - H_0(t_{ijl})]\exp\left(\alpha y^*_{ijl}\right)\right). \end{aligned} \tag{7.21}$$

The marginal loglikelihood function for all the clusters is

$$l\left(\boldsymbol{\xi},\theta,\alpha\right) \quad = \quad \sum_{i=1}^{G} l_i\left(\boldsymbol{\xi},\theta,\alpha\right) \tag{7.22}$$

The maximum likelihood estimates $\boldsymbol{\xi},\theta,\alpha$ are estimated by maximizing the loglikelihood function given in equation (7.18). Numerical methods can be used to maximize this loglikelihood function e.g by Newton-Raphson procedure.

7.3.3 Ordinary regression calibration (ORC)

We propose to extend ordinary regression calibration discussed in section 7.2.2.2 to simultaneously model time varying covariate while adjusting for delayed entry. Stage I involves fitting one mixed-effects models using the whole dataset which comprises of measurements of only individuals with $T_{ij} \geq T_{ij0}$. The predicted value of the covariate for individuals are then computed as: $Y_{ij}^{*}(t) = \mathbb{E}\{Y_{ij}(t)|Y_{ij}(t_{ijl}, T_{ij} \geq T_{ij0}), \hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_b^2, \hat{\sigma}_u^2\}$. $Y_{ij}^{*}(t)$ can be calculated for either unique event times for non-parametric baseline or fine grid for the parametric baseline. Stage II proceeds similarly to the RRC procedure discussed in section 7.3.2 but using ORC predictions in the survival model.

7.3.4 Joint model approach

This is an extension of joint model described in Section 7.2.3 but now only individuals with $T_{ij} \geq T_{ij0}$ are observed. The standard joint model for a longitudinal and a time-to-event outcome will be

$$\begin{aligned} y_{ij}(t \mid u_i) &= \beta_0 + b_{ij} + u_i + \beta_1 t + \varepsilon_{ij}(t) \\ h_{ij}(t \mid b_{ij}, u_i, v_i) &= h_0(t) v_i \exp\left\{\alpha(\beta_0 + b_{ij} + u_i + \beta_1 t)\right\}. \end{aligned}$$

To obtain the marginal likelihood it involves integrating out all random effects. For this section we make a modification to reduce the complexity of the joint model by plugging in an estimate of one of the two random effects and hence integrating just over one random effect. It will thus be a modified two stage approach, where we first fit a mixed effect longitudinal model to predict the individual random intercept b_{ij} and assume it does not

change over time. Let the longitudinal model have random intercepts only and assume a link function of current value i.e $\hat{\beta}_0 + \hat{b}_{ij} + u_i + \hat{\beta}_1 t$.

The standard joint model for a longitudinal and a time-to-event outcome will therefore be

$$\begin{aligned} y_{ij}(t \mid u_i) &= \hat{\beta}_0 + \hat{b}_{ij} + u_i + \hat{\beta}_1 t + \varepsilon_{ij}(t) \\ h_{ij}(t \mid b_{ij}, u_i, v_i) &= h_0(t) v_i \exp\{\alpha(\hat{\beta}_0 + \hat{b}_{ij} + u_i + \hat{\beta}_1 t)\}. \end{aligned}$$

Suppose that data, $\mathcal{D} = \{t_{ij}, t_{ij0}, \delta_{ij}, \mathbf{y}_{ij}\}$, of both longitudinal and time-to-event outcomes are collected on subjects $j = 1, \ldots, n_i$ and $i = 1, \ldots, G$. t_{ij0} denotes age at which individual enters the study. An individual included in the study only if $t_{ij} > t_{ij0}$

The estimation of $\mathbf{\Phi} = (\boldsymbol{\xi}, \alpha, \sigma_u^2, \sigma_\varepsilon^2)$ is based on maximum likelihood principles by maximizing the log-likelihood function of the joint distribution of the longitudinal and the time-to-event outcomes, $\{\mathbf{y}_{ij}, t_{ij}, \delta_{ij}\}$ conditional on an individual being in the study:

$$\ell(\mathbf{\Phi}|\mathcal{D}) = \sum_{i=1}^{G} \log\left(\int_u \int_v \Big[\prod_{j=1}^{n_i} f(\mathbf{y}_{ij} \mid u_i) f(t_{ij}, \delta_{ij} \mid u_i, v) \Big] g(v|t_{ij} > t_{ij0}) f(u_i) \mathrm{d}v \mathrm{d}u_i \right),$$

where

$$\begin{aligned} f\{y_{ijl} \mid u_i\} &= \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\Big(- \frac{[y_{ijl} - (\hat{\beta}_0 + \hat{b}_{ij} + u_i + \hat{\beta}_1 t_{ijl})]^2}{2\sigma_\varepsilon^2} \Big) \\ f(\mathbf{y}_{ij} \mid u_i) &= \prod_{l=1}^{n_{ij}} f(y_{ijl} \mid u_i) = (2\pi\sigma_\varepsilon^2)^{-n_{ij}/2} \exp\Big(- \frac{\sum_{l=1}^{n_{ij}} [y_{ijl} - (\hat{\beta}_0 + \hat{b}_{ij} + u_i + \hat{\beta}_1 t_{ijl})]^2}{2\sigma_\varepsilon^2} \Big) \\ f(u_i) &= \frac{1}{\sqrt{2\pi\sigma_u^2}} \exp\Big(- \frac{u_i^2}{2\sigma_u^2} \Big) \\ f(t_{ij}, \delta_{ij} \mid u_i, v_i) &= \frac{[h_{ij}(t_{ij} \mid \boldsymbol{Y}_{ij}), v_i)]^{\delta_{ij}} \, S_{ij}(t_{ij} \mid \boldsymbol{Y}_{ij}), v_i)}{S_{ij}(t_{ij0} \mid \boldsymbol{Y}_{ij}(t_{ij0}), v_i)} \\ &= [h_{ij}(t_{ij} \mid \boldsymbol{Y}_{ij}), v_i)]^{\delta_{ij}} \exp\Big(- [H_{ij}(t_{ij} \mid \boldsymbol{Y}_{ij}), v_i) - H_{ij}(t_{ij0} \mid \boldsymbol{Y}_{ij0}(t_{ij0}), v_i)] \Big) \\ &= [h_{ij}(t_{ij} \mid \boldsymbol{Y}_{ij}), v_i)]^{\delta_{ij}} \exp\left(- \int_{t_{ij0}}^{t_{ij}} h_{ij}(s \mid \boldsymbol{Y}_{ij}, vi) \, \mathrm{d}s. \right) \\ &= \Big[h_0(t_{ij}) v_i \exp\Big\{ \alpha(\hat{\beta}_0 + \hat{b}_{ij} + u_i + \hat{\beta}_1 t_{ij}) \Big\} \Big]^{\delta_{ij}} \\ &\qquad \times \exp\left(- \int_{t_{ij0}}^{t_{ij}} h_0(s) v_i \exp\Big\{ \alpha(\hat{\beta}_0 + \hat{b}_{ij} + u_i + \hat{\beta}_1 s) \Big\} \, \mathrm{d}s \right) \\ h_0(t) &= \lambda\rho t^{\rho-1} \end{aligned}$$

Updating the frailty distribution at the times of delayed entry will be as follows:

$$\begin{aligned}
g(v|t_{ij} > t_{ij0}) &= \frac{g(v)S(t_{ij0}|v)}{S(t_{ij0})} = \frac{\exp(-H(t_{ij0}|v))g(v)}{\exp(-H(t_{ij0}))} \\
&= \frac{P(t_{i11} \geq t_{i10}, ..., t_{in_i1} \geq t_{in_i0}, v)g(v)}{P(t_{i11} \geq t_{i10}, ...t_{in_i1} \geq t_{in_i0})} \\
&= \frac{\exp(-v\sum_{j=1}^{n_i}\int_0^{t_{ij0}} h_0(s)e^{\alpha(\hat{\beta}_0+\hat{b}_{ij}+u_i+\hat{\beta}_1 s)}\mathrm{d}s)\frac{v^{1/\theta-1}\exp(-v/\theta)}{\theta^{1/\theta}\Gamma_\theta}}{\int_v \exp(-v\sum_{j=1}^{n_i}\int_0^{t_{ij0}} h_0(s)e^{\alpha(\hat{\beta}_0+\hat{b}_{ij}+u_i+\hat{\beta}_1 s)}\mathrm{d}s)\frac{v^{1/\theta-1}\exp(-v/\theta)}{\theta^{1/\theta}\Gamma_\theta}dv} \\
&= \frac{\exp(-v\sum_{j=1}^{n_i}\int_0^{t_{ij0}} h_0(s)e^{\alpha(\hat{\beta}_0+\hat{b}_{ij}+u_i+\hat{\beta}_1 s)}\mathrm{d}s)v^{1/\theta-1}\exp(-v/\theta)}{\int_v \exp(-v\sum_{j=1}^{n_i}\int_0^{t_{ij0}} h_0(s)e^{\alpha(\hat{\beta}_0+\hat{b}_{ij}+u_i+\hat{\beta}_1 s)}\mathrm{d}s)v^{1/\theta-1}\exp(-v/\theta)\,dv} \\
&= \frac{v^{1/\theta-1}\exp\left(-v\left[\frac{1}{\theta}+\sum_{j=1}^{n_i}\int_0^{t_{ij0}} h_0(s)e^{\alpha(\hat{\beta}_0+\hat{b}_{ij}+u_i+\hat{\beta}_1 s)}\mathrm{d}s\right]\right)}{\int_v v^{1/\theta-1}\exp\left(-v\left[\frac{1}{\theta}+\sum_{j=1}^{n_i}\int_0^{t_{ij0}} h_0(s)e^{\alpha(\hat{\beta}_0+\hat{b}_{ij}+u_i+\hat{\beta}_1 s)}\mathrm{d}s\right]\right)dv} \\
&= \frac{v^{\frac{1}{\theta}-1}\exp\left(-v\left[\frac{1}{\theta}+\sum_{j=1}^{n_i}\int_0^{t_{ij0}} h_0(s)e^{\alpha(\hat{\beta}_0+\hat{b}_{ij}+u_i+\hat{\beta}_1 s)}\mathrm{d}s\right]\right)}{\frac{\Gamma_\theta}{\left[\frac{1}{\theta}+\sum_{j=1}^{n_i}\int_0^{t_{ij0}} h_0(s)e^{\alpha(\hat{\beta}_0+\hat{b}_{ij}+u_i+\hat{\beta}_1 s)}\mathrm{d}s\right]^{\frac{1}{\theta}}}}.
\end{aligned}$$

Therefore for joint model

$$\begin{aligned}
g(v|t_{i1} > t_{i10}, ..., t_{in_i} > t_{in_i0}) = &\frac{\left[\frac{1}{\theta}+\sum_{j=1}^{n_i}\int_0^{t_{ij0}} h_0(s)e^{\alpha(\hat{\beta}_0+\hat{b}_{ij}+u_i+\hat{\beta}_1 s)}\mathrm{d}s\right]^{\frac{1}{\theta}}}{\Gamma_\theta}v^{\frac{1}{\theta}-1} \\
&\times e^{-v\left[\frac{1}{\theta}+\sum_{j=1}^{n_i}\int_0^{t_{ij0}} h_0(s)e^{\alpha(\hat{\beta}_0+\hat{b}_{ij}+u_i+\hat{\beta}_1 s)}\mathrm{d}s\right]}
\end{aligned}$$

The marginal log-likelihood for all the clusters is:

$$\ell(\mathbf{\Phi}|\mathcal{D}) = \sum_{i=1}^{G} \log\left(\int_u \left(\sqrt{2\pi\sigma_\varepsilon^2}\right)^{-\sum_{j=1}^{n_i} n_{ij}/2} \exp\Big(- \frac{\sum_{j=1}^{n_i}\sum_{l=1}^{n_{ij}} [y_{ijl} - (\hat\beta_0 + \hat b_{ij} + u_i + \hat\beta_1 t_{ijl})]^2}{2\sigma_\varepsilon^2}\Big) \left[\int_v \prod_{j=1}^{n_i} f(t_{ij}, \delta_{ij} \mid u_i, v) g(v|t_{ij} > t_{ij0}) \mathrm{d}v \right] f(u_i)\mathrm{d}u_i \right),$$

where

$$\begin{aligned}
\int_v \prod_{j=1}^{n_i} f(t_{ij}, \delta_{ij} \mid u_i, v) g(v|t_{ij} > t_{ij0})\mathrm{d}v &= \int_v \prod_{j=1}^{n_i} \left[h_0(t_{ij}) v \exp\left\{ \alpha(\hat\beta_0 + \hat b_{ij} + u_i + \hat\beta_1 t_{ij}) \right\} \right]^{\delta_{ij}} \exp\left(- \int_{t_{ij0}}^{t_{ij}} h_0(s) v \exp\left\{ \alpha(\hat\beta_0 + \hat b_{ij} + u_i + \hat\beta_1 s) \right\} \mathrm{d}s \right) \\
&\quad \times \frac{\left[\frac{1}{\theta} + \sum_{j=1}^{n_i} \int_0^{t_{ij0}} h_0(s) \exp\left\{ \alpha(\hat\beta_0 + \hat b_{ij} + u_i + \hat\beta_1 s) \right\} \mathrm{d}s \right]^{\frac{1}{\theta}}}{\Gamma_\theta} v^{1/\theta - 1} \exp\left(-v \left[\frac{1}{\theta} + \sum_{j=1}^{n_i} \int_0^{t_{ij0}} h_0(s) \exp\left\{ \alpha(\hat\beta_0 + \hat b_{ij} + u_i + \hat\beta_1 s) \right\} \mathrm{d}s \right] \right) dv \\
&= \prod_{j=1}^{n_i} \left[h_0(t_{ij}) \exp\left\{ \alpha(\hat\beta_0 + \hat b_{ij} + u_i + \hat\beta_1 t_{ij}) \right\} \right]^{\delta_{ij}} \frac{\left[\frac{1}{\theta} + \sum_{j=1}^{n_i} \int_0^{t_{ij0}} h_0(s) \exp\left\{ \alpha(\hat\beta_0 + \hat b_{ij} + u_i + \hat\beta_1 s) \right\} \mathrm{d}s \right]^{\frac{1}{\theta}}}{\Gamma_\theta} \\
&\quad \times \int_v v^{d_i + 1/\theta - 1} \exp\left(-v \sum_{j=1}^{n_i} \int_{t_{ij0}}^{t_{ij}} h_0(s) \exp\left\{ \alpha(\hat\beta_0 + \hat b_{ij} + u_i + \hat\beta_1 s) \right\} \mathrm{d}s \right) \exp\left(-v \left[\frac{1}{\theta} + \sum_{j=1}^{n_i} \int_0^{t_{ij0}} h_0(s) \exp\left\{ \alpha(\hat\beta_0 + \hat b_{ij} + u_i + \hat\beta_1 s) \right\} \mathrm{d}s \right] \right) dv \\
&= \left\{ \prod_{j=1}^{n_i} \left[h_0(t_{ij}) \exp\left\{ \alpha(\hat\beta_0 + \hat b_{ij} + u_i + \hat\beta_1 t_{ij}) \right\} \right]^{\delta_{ij}} \right\} \frac{\left[\frac{1}{\theta} + \sum_{j=1}^{n_i} \int_0^{t_{ij0}} h_0(s) \exp\left\{ \alpha(\hat\beta_0 + \hat b_{ij} + u_i + \hat\beta_1 s) \right\} \mathrm{d}s \right]^{\frac{1}{\theta}}}{\Gamma_\theta} \\
&\quad \times \int_v v^{d_i + 1/\theta - 1} \exp\left(-v \left[\sum_{j=1}^{n_i} \int_{t_{ij0}}^{t_{ij}} h_0(s) \exp\left\{ \alpha(\hat\beta_0 + \hat b_{ij} + u_i + \hat\beta_1 s) \right\} \mathrm{d}s + \frac{1}{\theta} + \sum_{j=1}^{n_i} \int_0^{t_{ij0}} h_0(s) \exp\left\{ \alpha(\hat\beta_0 + \hat b_{ij} + u_i + \hat\beta_1 s) \right\} \mathrm{d}s \right] \right) dv \\
&= \left\{ \prod_{j=1}^{n_i} \left[h_0(t_{ij}) \exp\left\{ \alpha(\hat\beta_0 + \hat b_{ij} + u_i + \hat\beta_1 t_{ij}) \right\} \right]^{\delta_{ij}} \right\} \frac{\left[\frac{1}{\theta} + \sum_{j=1}^{n_i} \int_0^{t_{ij0}} h_0(s) \exp\left\{ \alpha(\hat\beta_0 + \hat b_{ij} + u_i + \hat\beta_1 s) \right\} \mathrm{d}s \right]^{\frac{1}{\theta}}}{\Gamma_\theta} \\
&\quad \frac{\Gamma(d_i + 1/\theta)}{\left[\frac{1}{\theta} + \sum_{j=1}^{n_i} \int_0^{t_{ij}} h_0(s) \exp\left\{ \alpha(\hat\beta_0 + \hat b_{ij} + u_i + \hat\beta_1 s) \right\} \mathrm{d}s \right]^{d_i + 1/\theta}} \qquad \text{since } \int_0^\infty e^{-vB} v^{\kappa - 1} = \frac{\Gamma\kappa}{B^\kappa}
\end{aligned}$$

Numerical methods can be used to maximize this loglikelihood function e.g by Newton-Raphson procedure.

7.4 Simulation study

We perform simulations for the following scenarios: dense grid and no measurement error (section 7.4.1) and sparse grid with/without measurement error (Section 7.4.2 and Section 7.4.3).

The aims of simulation study are to investigate:

- Influence of magnitude of measurement error.
- Influence of magnitude of time varying covariate effect.
- Influence of grid of measurements.

The estimands of interest include time varying covariate effect α, frailty variance θ, and Weibull baseline hazard parameters. For all the simulated scenarios, the estimated relative bias (reBias), standard deviation (SD) and mean square error (MSE) for estimation of parameters are reported.

Simulation Design

We simulate from a joint model of longitudinal and time to event data. Suppose that there are G groups with n_i individuals in the i^{th} group, $i = 1, 2, ..., G$ and $j = 1, ..., n_i$. Let $y_{ijl} = y_{ij}(t_{ijl})$ denote the response of subject j in cluster i at time $t_{ijl}, l = 1, \ldots, n_{ij}$.

We simulate data to follow the following theoretical model:

$$\begin{aligned} y_{ij}(t \mid b_{ij}, u_i) &= 1 + b_{ij} + u_i + 0.01t + \varepsilon_{ij}(t) \\ h_{ij}(t \mid b_{ij}, u_i, v_i) &= h_0(t) v_i \exp\left\{\alpha(1 + b_{ij} + u_i + 0.01t)\right\} \end{aligned} \tag{7.25}$$

$b_{ij} \sim N(0, 0.1^2)$, where b_{ij} individual specific random effect

$u_i \sim N(0, 0.1^2)$, where u_i cluster specific random effect

$\boldsymbol{\varepsilon}_{ij}(t) \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ where $\sigma_\varepsilon = 0, 0.1, 0.2, 0.3, 0.6$.

$\alpha = 1, 2, 3$.

$\beta_0 = 1,\ \beta_1 = 0.01$

Weibull baseline with shape parameter $\rho = 2$ and scale parameter $\lambda = 0.001$. Thus $h_0(t) = \lambda\rho t^{\rho-1} = 0.002t$.

In general the cumulative hazard function will be

$$H\left(t|\boldsymbol{Y}_{ij}, v_i\right) = \int_0^t \lambda\rho s^{\rho-1} v_i \exp\left(\alpha(\beta_0 + b_{ij} + u_i + \beta_1 s)\right) ds.$$

but

$$\begin{aligned} U &= S\left(t|\boldsymbol{Y}_{ij}(t), v_i\right) = \exp\left[-H\left(t|\boldsymbol{Y}_{ij}(t), v_i\right)\right] \sim \text{Uni}[0,1] \\ -\log(U) &= H\left(t|\boldsymbol{Y}_{ij}(t), v_i\right) \\ &= \int_0^t \lambda\rho s^{\rho-1} v_i \exp(\alpha(\beta_0 + b_{ij} + u_i + \beta_1 s)) ds. \end{aligned} \tag{7.26}$$

Given a realization of U, numerical integration is used to solve for t in equation 7.26.

If the Weibull baseline hazard shape parameter $\rho = 1$ (equivalent to exponential distribution) then t exists in closed format and is given by:

$$t = \frac{1}{\alpha\beta_1} \log\left(1 + \frac{\alpha\beta_1[-\log(u)]}{\lambda v_i \exp\left[\alpha(\beta_0 + b_{ij} + u_i)\right]}\right) \tag{7.27}$$

where v_i follows a gamma distribution with mean 1 and variance θ.

Algorithm

- Generate frailties v_i and random effects b_{ij}, u_i
- Generate u$\sim$ Uni(0, 1).
- Compute t_{ij} the time-to-event by solving Equation (7.26).
- Simulate individual specific entry time points t_{ij0} by some chosen mechanism. In our case we assume 0.5 probability of truncation and let $t_{ij0} \sim \text{Uni}(0, t_0)$.
- Generate follow-up times t_{ijl}, at points $l = 0, ..., n_{ij} - 1$ for an individual such that $t_{ij0} \leq t_{ijl}$ then compute $y_{ij}(t)$ at time points.

Assume random censoring to follow Uni$(0, 15)$. We report the relative bias (reBias), standard deviation (SD), and mean square error (MSE), coverage probabilities for estimation of parameters across 1000 Monte Carlo trials (100 Monte Carlo trials for the joint model). The data is generated using a joint model given in equation (7.25) above with gamma frailty (frailty variance $\theta = 0.5$). Let number of clusters be 2000. Inc.G represents the number of clusters included in the analysis after truncation and removal of singletons (i.e. we only consider fully observed clusters). For all models we fit a parametric Cox proportional hazards model (Weibull baseline) while adjusting for delayed entry.

Results

7.4.1 Scenario 1: Dense and no measurement error

For this scenario, the longitudinal measurements are taken with a regular gap/interval of 2 for all individuals i.e. at regular time points $0, 2, 4, 6, \cdots$ and no measurement error ($\sigma_\varepsilon = 0$). Table 7.1 shows the performance of the methods in estimating the parameters of the gamma shared frailty model in presence of delayed entry and time varying covariates for LOCF and 'naive' method (done using `frailtypack` R package which models time varying covariate without adjusting for delayed entry) .

Simulation Results

Table 7.1: Simulation to investigate the effect of the magnitude of the time varying covariate effect α. Naive is set up for modelling time-varying covariates without adjusting for delayed entry.

				Naive				LOCF			
α	Par.	Inc.G	Events	reBias	SD	MSE	CP	reBias	SD	MSE	CP
1	α	1637	578	−0.029	0.300	0.091	0.962	−0.022	0.301	0.091	0.965
	θ	1637	578	0.027	0.181	0.033	0.957	0.034	0.181	0.033	0.960
	λ	1637	578	0.085	0.000	0.000		0.081	0.000	0.000	
	ρ	1637	578	0.001	0.079	0.006		0.003	0.079	0.006	
2	α	1584	1135	−0.031	0.242	0.062	0.954	−0.012	0.246	0.061	0.948
	θ	1584	1135	−0.004	0.093	0.009	0.949	0.012	0.095	0.009	0.948
	λ	1584	1135	0.104	0.000	0.000		0.072	0.000	0.000	
	ρ	1584	1135	−0.001	0.062	0.004		0.003	0.062	0.004	
3	α	1465	1633	−0.054	0.228	0.078	0.896	−0.016	0.242	0.060	0.940
	θ	1465	1633	−0.032	0.062	0.004	0.939	−0.002	0.063	0.004	0.950
	λ	1465	1633	0.221	0.000	0.000		0.105	0.000	0.000	
	ρ	1465	1633	−0.009	0.052	0.003		0.003	0.052	0.003	

The LOCF is the standard method for scenario for the case of singletons, so should also be with twins. We therefore only compare LOCF to existing naive methodology. LOCF performs well in estimation of parameters.

7.4.2 Scenario 2: Sparse and no measurement error

For this scenario, all individuals have $\leq$ three measurements during the study (i.e $n_{ij} \leq 3$) and assume no measurement error ($\sigma_\varepsilon = 0$). The methods compared in this scenario are: LOCF approach discussed in section 7.3.1, RRC approach discussed in section 7.3.2 and ORC approach discussed in section 7.3.3.

Simulation Results

Table 7.2: Simulation to investigate the effect of the magnitude of the time varying covariate effect α.

				LOCF				RRC				ORC			
α	Par.	Inc.G	Events	reBias	SD	MSE	CP	reBias	SD	MSE	CP	reBias	SD	MSE	CP
1	α	1637	577	−0.306	0.342	0.211	0.805	−0.031	0.365	0.134	0.933	0.040	0.466	0.218	0.951
	θ	1637	577	−0.062	0.159	0.026	0.953	−0.018	0.178	0.032	0.957	−0.043	0.175	0.031	0.955
	λ	1637	577	0.439	0.001	0.000	0.000	0.331	0.001	0.000	0.000	0.256	0.001	0.000	0.000
	ρ	1637	577	0.009	0.084	0.007	0.000	−0.033	0.087	0.012	0.000	−0.032	0.089	0.012	0.000
2	α	1584	1136	−0.187	0.252	0.203	0.675	−0.021	0.296	0.089	0.950	−0.054	0.228	0.078	0.896
	θ	1584	1136	−0.079	0.087	0.009	0.915	−0.019	0.099	0.010	0.930	−0.032	0.062	0.004	0.939
	λ	1584	1136	0.500	0.000	0.000	0.000	0.275	0.000	0.000	0.000	0.221	0.000	0.000	0.000
	ρ	1584	1136	0.011	0.060	0.004	0.000	−0.030	0.064	0.008	0.000	−0.009	0.052	0.003	0.000
3	α	1464	1633	−0.140	0.231	0.229	0.565	−0.020	0.272	0.077	0.965	−0.037	0.361	0.143	0.929
	θ	1464	1633	−0.093	0.060	0.006	0.868	0.005	0.068	0.005	0.950	−0.066	0.066	0.005	0.929
	λ	1464	1633	0.595	0.000	0.000	0.000	0.252	0.000	0.000	0.000	0.390	0.001	0.000	0.000
	ρ	1464	1633	0.005	0.052	0.003	0.000	−0.027	0.055	0.006	0.000	−0.036	0.053	0.008	0.000

Based on Table 7.2, for LOCF increasing α leads to better estimation. This is as a result of the current data generation mechanism. Note that we hold all other settings constant and change the value of α during generation. All individuals have $\leq$ three measurements during the study (not taken at regular time-points). Generally for larger α while holding other settings constant results in lower values of generated survival time while smaller α results in larger values of generated survival time. Thus we have less sparse measurements for larger α. Both α and sparseness of measurements are changing in this case. This behaviour is also depicted in LOCF results in Table 7.3. Overall, the two stage approaches outperform the LOCF.

7.4.3 Sparse with measurement error.

The simulations performed in section 7.4.1 and 7.4.2 assume no covariate measurement error. We now extend to the scenario when you have covariate measurement error. Similar to Section 7.4.2 we consider all individuals to have $\leq$ three measurements during the study.

We perform simulations to investigate the effect of:

- The magnitude of measurement error σ^2_ε
- The magnitude of the time varying covariate effect α

The four proposed approaches are compared here: LOCF approach discussed in section 7.3.1, RRC approach discussed in section 7.3.2, ORC approach discussed in section 7.3.3, and joint modelling approach discussed in section 7.3.4.

Simulation Results

Table 7.3: Simulation to investigate the effect of the magnitude of the time varying covariate effect α. We let the $\sigma_\varepsilon = 0.1$

				Last observation carried forward				Risk set regression Calibration				Ordinary regression calibration				Joint model			
α	Par.	Inc.G	Events	reBias	SD	MSE	CP	reBias	SD	MSE	CP	reBias	SD	MSE	CP	reBias	SD	MSE	CP
1	α	1637	577	−0.542	0.255	0.359	0.455	−0.010	0.511	0.256	0.920	0.002	0.488	0.236	0.956	0.055	0.354	0.129	0.943
	θ	1637	577	−0.034	0.172	0.030	0.940	0.085	0.221	0.049	0.880	−0.076	0.189	0.037	0.917	0.020	0.181	0.033	0.885
2	α	1584	1136	−0.493	0.200	1.014	0.000	−0.009	0.345	0.118	0.921	−0.033	0.371	0.141	0.965	0.021	0.306	0.096	0.935
	θ	1584	1136	−0.077	0.092	0.010	0.910	0.014	0.103	0.011	0.978	−0.053	0.107	0.012	0.930	−0.002	0.092	0.009	0.946
3	α	1464	1633	−0.494	0.164	2.222	0.000	−0.035	0.306	0.104	0.939	−0.078	0.381	0.199	0.900	−0.009	0.316	0.096	0.875
	θ	1464	1633	−0.109	0.063	0.007	0.810	−0.006	0.067	0.004	0.945	−0.048	0.073	0.006	0.905	−0.029	0.060	0.004	0.920

Table 7.4: Simulation to investigate the effect of the magnitude of measurement error σ_ε^2. We let the $\alpha = 2$

				Last observation carried forward				Risk set regression Calibration				Ordinary regression calibration				Joint model			
σ_ε	Par.	Inc.G	Events	reBias	SD	MSE	CP	reBias	SD	MSE	CP	reBias	SD	MSE	CP	reBias	SD	MSE	CP
0	α	1584	1136	−0.187	0.252	0.203	0.675	−0.021	0.296	0.089	0.950	−0.054	0.228	0.078	0.896				
	θ	1584	1136	−0.079	0.087	0.009	0.915	−0.019	0.099	0.010	0.930	−0.032	0.062	0.004	0.939				
0.1	α	1584	1136	−0.493	0.200	1.014	0.000	−0.009	0.345	0.118	0.921	−0.033	0.371	0.141	0.965	0.021	0.306	0.096	0.935
	θ	1584	1136	−0.077	0.092	0.010	0.910	0.014	0.103	0.011	0.978	−0.053	0.107	0.012	0.930	−0.002	0.092	0.009	0.946
0.2	α	1584	1136	−0.765	0.132	2.359	0.000	−0.035	0.421	0.181	0.920	−0.044	0.466	0.224	0.955	0.041	0.400	0.165	0.892
	θ	1584	1136	−0.046	0.094	0.009	0.920	0.007	0.105	0.011	0.950	−0.034	0.098	0.010	0.950	−0.016	0.093	0.009	0.946
0.3	α	1584	1136	−0.876	0.094	3.081	0.000	−0.094	0.476	0.260	0.947	−0.082	0.588	0.371	0.915	0.059	0.497	0.247	0.909
	θ	1584	1136	−0.025	0.096	0.009	0.930	0.008	0.106	0.011	0.940	−0.012	0.100	0.010	0.960	−0.017	0.099	0.009	0.922
0.6	α	1584	1136	−0.965	0.048	3.730	0.000	−0.499	0.585	1.334	0.562	−0.097	1.270	1.642	0.915	0.052	0.735	0.526	0.894
	θ	1584	1136	−0.005	0.096	0.009	0.945	0.012	0.107	0.011	0.927	0.003	0.100	0.010	0.940	−0.010	0.091	0.008	0.835

Table 7.5: Estimation of other parameters : Effect of magnitude of covariate effect α

		Joint model		
α	Par.	reBias	SD	MSE
1	σ_u	−0.180	0.004	0.000
	λ	−0.018	0.000	0.000
	ρ	0.008	0.092	0.009
2	σ_u	−0.182	0.004	0.000
	λ	−0.008	0.000	0.000
	ρ	0.005	0.062	0.004
3	σ_u	−0.187	0.004	0.000
	λ	0.075	0.000	0.000
	ρ	0.004	0.058	0.003

Table 7.6: Estimation of other parameters : Effect of magnitude of measurement error σ_ε^2

		Joint model		
σ_ε	Par.	reBias	SD	MSE
0.1	σ_u	−0.182	0.004	0.000
	λ	−0.008	0.000	0.000
	ρ	0.005	0.062	0.004
0.2	σ_u	−0.184	0.007	0.000
	λ	−0.007	0.000	0.000
	ρ	0.003	0.062	0.004
0.3	σ_u	−0.193	0.013	0.001
	λ	−0.006	0.000	0.000
	ρ	0.004	0.065	0.004
0.6	σ_u	−0.150	0.025	0.001
	λ	0.103	0.001	0.000
	ρ	0.005	0.063	0.004

Tables 7.3 and Table 7.4 show the performance of the proposed methods (LOCF, RRC, ORC, and joint model) in estimating the parameters of the gamma shared frailty model in presence of delayed entry and time varying covariates. When using the LOCF approach the time-varying covariate effect α is estimated with bias in all scenarios (i.e. when varying either the magnitude of the time varying covariate effect α or the magnitude of measurement error σ_ε^2). From Table 7.3 the largest bias in LOCF estimator of α is about 54% when $\alpha = 1$ and from Table 7.4 the largest bias in LOCF estimator of α is about 96% when $\sigma_\varepsilon = 0.6$.

The two stage approaches yield similar results apart from when we have large magnitude of time varying covariate effect (α=3) and large magnitude of measurement error ($\sigma_\varepsilon = 0.6$). For large magnitude of the time varying covariate effect (α=3), RRC performs better than

ORC. From Table 7.3 the largest bias in RRC estimator of α is about 3% while largest bias in ORC estimator of α is about 8%. For large magnitude of measurement error ($\sigma_\varepsilon = 0.6$), ORC performs better than the RRC. From Table 7.4 the largest bias in RRC estimator of α is about 50% while the largest bias in ORC estimator of α is less than 10% .

The joint modelling approach performs well in estimation of the time-varying covariate effect α with varying either the magnitude of the time varying covariate effect α or the magnitude of measurement error σ_ε^2. From Table 7.3 the largest bias in joint model estimator of α is about 5% when $\alpha = 1$ while from Table 7.4 the largest bias in joint model estimator of α is about 6% for large σ_ε.

All the methods appear to estimate θ with low bias with varying either the magnitude of the time varying covariate effect α or the magnitude of measurement error σ_ε^2 (largest bias for LOCF estimator of θ is about 10% when $\alpha = 3$, largest bias for two stage approach estimators of θ is about 8% when $\alpha = 1$ and largest bias for joint model estimator of θ is about 3% when $\alpha = 3$ as shown in Table 7.3).

Overall the two stage approach and joint model yields less bias in estimation of parameters as compared to the last observation carried forward approach in all scenarios. Bias in estimation of α increases when measurement error increases with the exception of fitting a joint model. Both last observation carried forward and the two stages approach method results in consistent underestimation of the time varying covariate effect α with the last observation carried forward under-estimating more. The coverage probabilities especially for the time varying covariate effect α increases when two stage approach or joint model is used. It appears that for the two stage approach the ordinary regression calibration is sufficient in estimation of the time varying covariate effect, even for large measurement error as it yield notably low bias.

The joint model had convergence issues (about 10% of the simulated datasets do not yield standard error estimates of parameter α).

7.5 Data Analysis - BMD as a time varying covariate

In chapter 5 we performed analysis of the twinsUK data using BMD as a time fixed covariate to estimate the probability of a fracture in the next time period. The parametric (Weibull) and non-parametric (`frailtysurv`,B-splines) baseline hazards with gamma frailty yielded similar results for survival probabilities. BMD at baseline was also shown to be a significant risk factor of fracture incidence. The goal of these analysis is to address question 2 in Section 4.4 to model the relationship between BMD and fracture incidence. We would like to use the proposed approaches in this chapter (LOCF, ORC, RRC and joint model) to deal with delayed entry and model BMD as a time varying covariate as we have longitudinal BMD measurements for individuals in the study.

In this analysis we consider BMD measurements after age 50 for the 1028 individuals (383 twin pairs and 262 singletons). We perform analysis on the whole dataset (both singletons and twin pairs) and on twin pairs only. For the ordinary regression calibration we fit a single mixed effect model in the first stage using all available data to compute predicted covariate values at any age (age at entry/event times). For the risk set regression calibration, the data has 30 unique age at entry/event times (minimum age at entry=50, minimum event time = 52 years while maximum event time = 85 years). This is equivalent to intervals of gap 1. For age 51 and 52 we fit a random intercept at cluster-level (twin grouping) because individuals do not have repeated measurements at those time-points. For the other unique age at entry/event times, we fit random intercepts for both individual and twin pair levels. Next we compute the predicted value of the covariate for individuals still at risk at a specific event time, arrange data in start stop format then fit a frailty model.

Table 7.7: Parameter estimates assuming gamma frailty distribution and BMD as a time varying covariate in presence of delayed entry for only twin pairs. The number of observed events is 138. s.e represents the standard error.

	LOCF	RRC	ORC	Joint model
Variable	Estimate (s.e)	Estimate (s.e)	Estimate (s.e)	Estimate (s.e)
BMD	−2.507 (0.881)	−3.519 (1.220)	−3.569 (1.305)	−4.435 (1.491)
θ	0.280 (0.291)	0.425 (0.413)	0.277 (0.284)	0.374(0.285)
λ	<0.001	0.001	0.002	0.011
ρ	2.579	2.297	2.234	1.999
σ_u				0.047
σ_e				0.080
σ_b				0.084

Table 7.8: Parameter estimates assuming gamma frailty distribution and BMD as a time varying covariate in presence of delayed entry for both singletons and twin pairs. The number of observed events is 170. s.e represents the standard error.

	LOCF	RRC	ORC	Joint model
Variable	Estimate (s.e)	Estimate (s.e)	Estimate (s.e)	Estimate (s.e)
BMD	−2.900 (0.768)	−3.627 (1.040)	−3.809 (1.106)	−4.797 (1.278)
θ	0.123 (0.212)	0.285(0.327)	0.099 (0.201)	0.217 (0.195)
λ	0.003	0.003	0.049	0.215
ρ	1.944	2.072	1.528	1.415
σ_u				0.046
σ_e				0.077
σ_b				0.084

For all the approaches considered in Table 7.8 and Table 7.7, BMD appears to be significantly associated with fracture incidence (for instance last observation carried forward yields a covariate effect of -2.900, p-value of < 0.001 considering both twins and singletons dataset and yields a covariate effect of -2.507, p-value of < 0.001 considering twins only dataset). The two stage approaches and joint model yield larger BMD effect as compared to the last observation carried forward. There is a lower θ estimate for ordinary regression calibration as compared to risk set regression calibration which could be attributed to the difference in estimate of the Weibull scale parameter λ. The residual variance of the mixed model was quite small ($\sigma_e = 0.084$ for twins only dataset and $\sigma_e = 0.077$ for case of both singletons and twins dataset).

7.6 Discussion and Conclusion

This chapter has developed novel methodology for modeling time varying covariates and delayed entry in frailty models. Four approaches for modeling time varying covariates

namely, last observation carried forward, risk set regression calibration, ordinary regression calibration, and joint modelling approaches were adapted to the situation of clustered data with delayed entry. Based on simulations, without covariate measurement error, with a dense grid of measurements, and with observations not changing drastically in the observed intervals the last observation carried forward approach has been shown to well estimate parameters. However, with covariate measurement error the last observation carried forward approach becomes biased (it is the worst option as compared to two stage approaches and joint model). The naive method gives biased estimated. This is because it does not use correct estimates of the weights during parameter estimation.

Two stage approaches can estimate parameters well for less dense grid of measurements and low values of measurement error. For large measurement error ordinary regression calibration performs better than risk-set regression calibration however it uses future data. Risk-set regression calibration probably works less due to more error in estimation (less data). It appears that the ordinary regression calibration is sufficient in estimation of the time varying covariate effect for large measurement error. Based on the data analysis, the two stage approaches and joint model yield larger BMD effect as compared to the last observation carried forward. The R codes used in the simulation study and the data analysis are in Github (Time-Varying-Covariate).

In the data application, results of ordinary regression calibration and risk-set regression calibration are similar. This is probably because we have a large dataset or maybe the underlying covariate measurement error is relatively small. For a smaller dataset, the ordinary regression calibration or joint model would be preferred depending on the assumptions. There is some little differences in the parameter estimates when we consider twin pairs only versus when we consider both singletons and twin pairs. This differences could be due to randomness or due to having more data when we consider both singletons and twin pairs. Overall we have a larger BMD effect if we use BMD as a time-varying covariate as compared to BMD as a time fixed covariate. Therefore use of BMD as a time fixed covariate could potentially lead to underestimation of it's effect on fracture incidence.

An extension of simulation study in section 7.4.2 would be to investigate the magnitude of varying frailty variance θ. The joint model fitted here updates the frailty distribution only, an extension would be to update both frailty and twin effect from the longitudinal model.

This would be challenging as the updating the normal random effect distribution would increase the complexity of the marginal likelihood. This chapter considered one random effect (random intercept) in the longitudinal model. This could be extended to two random effects (random intercept and random slope) which would be more complicated to deal with multi-dimension in integration.

Chapter 8

Conclusion and future work

This chapter summarizes the main results of the thesis, provides future works and includes suggestions which could be improved further.

In summary: Chapter 1 gave a general introduction of the thesis, Chapter 2 gave a background of time independent covariate model while Chapter 3 gave a background of Cox model when dealing with with time varying covariates. Chapter 4 described the TwinsUK study on fracture incidence, BMD and health status. Chapter 5 applied shared frailty model to the TwinsUK cohort to predict the probability of experiencing a fracture in the next time period (e.g. five or ten years) given their bone mineral densities (BMD) and their *frailty index*. We considered models with both parametric and flexible baseline hazards. Chapter 6 is a simulation based study on investigating the impact of frailty misspecification on estimation of parameters and survival probabilities. This was the first study to investigate the effect of frailty misspecification on survival probabilities. In Chapter 7 novel methodology for modeling time varying covariates and delayed entry in frailty models using age as time scale was developed. In particular, four approaches for modeling time varying covariates for singletons namely, last observation carried forward, risk set regression calibration, ordinary regression calibration, and joint modelling approaches were adapted to the situation of clustered data with delayed entry. Via simulations we investigated how the last observation carried forward and regression calibration approach fit covariates generated with and without measurement error. We then model effect of BMD as a time varying covariate on fracture using age as time scale while taking into account delayed entry.

8.1 Summary of results

From this thesis we make the following conclusions:

1. From data analysis in Chapter 5 the gamma distributed frailty model with Weibull baseline probably fits the data well. This is because the parametric and flexible baseline hazards with gamma frailty yielded similar results of parameter estimates and survival probabilities. BMD is a significant risk factor of fracture incidence (which is in line with literature) whereas *frailty index* is not. Overall, low BMD value leads to higher probability of fracture.

2. From simulation results in Chapter 6 we observe that in the case of small frailty variance using a wrong frailty distribution still allows us to measure the fixed regression coefficient and survival probabilities without bias (underestimation or overestimation). However, this robustness does not seem to hold in several cases at larger frailty variances. The Weibull baseline hazard scale parameter corrected for the wrong frailty in most scenarios. However, for some extreme cases it appears that the scale parameter cannot adjust and thus parameters and survival probabilities are affected. Using a flexible function for the baseline hazard (plug-in estimator or splines) improves estimation of parameters as well as survival probabilities in the presence of frailty distribution misspecification.

3. From simulation results in Chapter 7 we observe that without covariate measurement error, with a dense grid of measurements, and with observations not changing drastically in the observed intervals the last observation carried forward approach has been shown to well estimate parameters but becomes the worst option when we have covariate measurement error. Two stage approaches can estimate parameters well for less dense grid of measurements and low values of measurement error. It appears that the ordinary regression calibration is sufficient in estimation of the time varying covariate effect for large measurement error as it performs notably better than risk-set regression calibration.

8.2 Publishable material

The publishable material from the thesis is as follows:

1. Title: ***Use of shared gamma frailty model in analysis of survival data in twins*** (Muli et al., 2021). This paper is already published. It is based on Chapter 5 and contains results on analysis of the TwinsUK cohort with an aim of estimating survival probabilities of fracture incidence in the next time period for elderly.

2. Title: ***The role of baseline hazard in frailty misspecification correction.*** This paper will be based on Chapter 6 and will contain results on the investigation of the impact of using the wrong frailty distribution on estimation of regression coefficients, baseline hazard parameters and survival probabilities.

3. Title: ***Modeling the effect of time varying covariate on survival subject to delayed entry in twins.*** This paper will be based on Chapter 7 and will develop novel methodology that deals with delayed entry as well as time varying covariates simultaneously for clustered data when using age a time scale. Results on application of developed methodology to the TwinsUK cohort will also be provided. Also included in this paper will be simulation results on how different approaches fit covariates generated with or without measurement error.

8.3 Improvement of the study and future work

In the joint model discussed in Section 7.3.4 we chose to estimate u_i and not b_{ij} because of mathematical convenience. Fixing b_{ij} leaves out only outer integrals in the likelihood shown in Equation (7.13). This could be improved to considering either a full joint estimation of all the parameters or fixing u_i and estimating b_{ij}. However, we hypothesize that this would be more computational complex. Also, sensitivity analysis is required.

In this dissertation we consider random intercept only for the longitudinal regression model. This could be extended to include a random slope. The challenge would be increasing random effects increases the complexity of the model because of increase dimension in integrating out random effects when computing marginal likelihood when a

joint model is fitted. This complexity could be dealt with by considering the two stage approaches. An extension of simulation study in Section 7.4.2 would be to investigate the magnitude of varying frailty variance θ. All simulations considered in this thesis (Chapter 6 and Chapter 7) involves clusters of size 2 limiting generalization to cluster of size two only like twins data. An extension would be to consider clusters of large sizes (> 2).

Glycan age changes with age but we do not have enough glycan information to be able to fit it as a time varying covariate. This is a limitation of the dataset and could be why *frailty index* was not significant. An extension for using glycomics would be to consider use of age specific profiles from other studies.

Analysis in both Chapter 5 and Chapter 7 considered modeling only dizygotic twin pairs, this could be extended to model both dizygotic and monozygotic twin pairs. This would bring about a challenge on assumptions to make to differentiate between dizygotic and monozygotic in the model. An approach to this would be to model with two variance components.

An extension of Chapter 5, Chapter 6 and Chapter 7 is updating the survival probability each time. This will allow to dynamically predict the risk of event for a subject, given the subject's history of covariates. A further extension would be to analyze fractures as a recurrent event. Recurrent event analysis covers situations where the event of interest may occur multiple times per subject. This increases the complexity of the model as it would require an additional random effect for the recurrent event. Competing risks could also be considered to model drop outs due to severe illness and death. In the dissertation the frailty term does not change over time or age. It may also be of interest in some applications to model time-varying frailties, for instance we could have time varying frailties to model changes in lifestyle (Gottard et al., 2012).

Appendix A

Simulation results

A.1 The role of baseline hazard in frailty misspecification correction

Relative bias (Bias), standard deviation (SD), and mean square error (MSE) for covariate effect β across 1000 Monte Carlo trials. The data are generated using gamma, truncated gamma, mixture normal, and lognormal frailty distributions and then Gamma frailty models (different baseline hazards) are fitted. The bold typeface indicates cases where the relative bias is greater than 5%.

			Gamma _Weibull			Gamma _Plug-in			Gamma _Splines		
	θ	G	Bias	SD	MSE	Bias	SD	MSE	Bias	SD	MSE
Gamma	0.1	100	0.005	0.209	0.043	**0.050**	0.230	0.059	−0.011	0.196	0.038
		200	0.003	0.137	0.019	**0.057**	0.168	0.035	0.044	0.171	0.033
		400	0.005	0.099	0.010	**0.073**	0.133	0.030	0.026	0.113	0.014
		800	0.000	0.072	0.005	−0.006	0.091	0.008	−0.001	0.081	0.006
	0.5	100	−0.001	0.235	0.055	−0.031	0.235	0.057	0.014	0.241	0.057
		200	0.005	0.178	0.032	−0.017	0.183	0.034	0.002	0.228	0.052
		400	0.004	0.115	0.013	−0.015	0.127	0.017	0.004	0.130	0.017
		800	−0.002	0.080	0.006	−0.010	0.019	0.008	−0.001	0.088	0.008
	2	100	0.008	0.308	0.095	−0.029	0.304	0.094	0.034	0.312	0.100
		200	0.008	0.217	0.047	−0.012	0.216	0.047	0.013	0.216	0.047
		400	0.003	0.147	0.022	−0.009	0.150	0.023	0.003	0.155	0.024
		800	0.005	0.103	0.011	−0.004	0.098	0.010	0.005	0.106	0.011
Truncated Gamma	0.1	100	0.002	0.210	0.044	0.016	0.208	0.044	0.004	0.154	0.023
		200	0.007	0.143	0.021	−0.004	0.147	0.022	0.008	0.159	0.025
		400	0.003	0.104	0.011	0.000	0.105	0.011	0.010	0.123	0.015
		800	0.000	0.072	0.005	−0.002	0.076	0.006	−0.003	0.074	0.006
	0.5	100	−0.016	0.225	0.051	−0.006	0.257	0.066	−0.021	0.239	0.057
		200	−0.034	0.153	0.026	−0.034	0.157	0.027	−0.011	0.172	0.030
		400	−0.029	0.103	0.013	−0.002	0.118	0.014	−0.018	0.181	0.033
		800	−0.027	0.073	0.007	−0.014	0.081	0.007	−0.019	0.088	0.009
	2	100	−0.044	0.253	0.068	−0.047	0.265	0.075	−0.041	0.255	0.069
		200	**−0.064**	0.172	0.039	−0.043	0.201	0.045	−0.048	0.179	0.037
		400	**−0.062**	0.119	0.023	−0.046	0.123	0.020	−0.047	0.128	0.021
		800	**−0.062**	0.081	0.015	−0.042	0.085	0.011	−0.048	0.083	0.012
Lognormal	0.1	100	0.025	0.237	0.051	0.043	0.241	0.062	−0.020	0.227	0.052
		200	0.011	0.122	0.015	0.007	0.147	0.022	0.025	0.195	0.039
		400	0.008	0.119	0.014	−0.002	0.110	0.012	0.001	0.107	0.011
		800	−0.002	0.072	0.005	−0.003	0.075	0.006	−0.004	0.078	0.006
	0.5	100	−0.007	0.231	0.053	−0.024	0.249	0.063	−0.024	0.385	0.142
		200	−0.011	0.157	0.025	−0.020	0.164	0.028	−0.034	0.235	0.058
		400	−0.018	0.116	0.014	−0.024	0.121	0.016	−0.018	0.181	0.033
		800	−0.019	0.078	0.007	−0.021	0.082	0.008	−0.015	0.085	0.008
	2	100	−0.048	0.278	0.083	−0.055	0.281	0.086	−0.030	0.291	0.087
		200	**−0.053**	0.184	0.040	−0.032	0.201	0.043	0.031	0.204	0.044
		400	**−0.061**	0.137	0.027	−0.044	0.142	0.025	−0.034	0.147	0.024
		800	**−0.060**	0.093	0.017	−0.041	0.097	0.013	−0.041	0.100	0.014
Mixture Normal	0.1	100	0.005	0.266	0.065	−0.014	0.234	0.055	−0.009	0.245	0.060
		200	−0.012	0.132	0.017	0.001	0.158	0.025	−0.044	0.352	0.118
		400	−0.011	0.107	0.012	−0.008	0.112	0.013	−0.059	0.166	0.034
		800	−0.013	0.078	0.006	−0.013	0.085	0.008	−0.053	0.180	0.038
	0.5	100	0.012	0.310	0.096	−0.026	0.307	0.096	−0.005	0.285	0.081
		200	0.007	0.218	0.048	−0.012	0.216	0.047	−0.008	0.210	0.044
		400	−0.003	0.155	0.024	−0.013	0.156	0.025	−0.021	0.141	0.021
		800	−0.007	0.108	0.012	−0.010	0.112	0.013	−0.016	0.103	0.011
	2	100	**0.063**	0.434	0.197	0.030	0.398	0.160	**0.066**	0.403	0.172
		200	0.038	0.300	0.093	−0.017	0.291	0.085	**0.051**	0.289	0.089
		400	0.032	0.208	0.046	−0.009	0.203	0.041	0.029	0.198	0.041
		800	0.020	0.145	0.022	−0.011	0.145	0.021	0.031	0.140	0.022

Relative bias (Bias), standard deviation (SD), and mean square error (MSE) for covariate effect β across 1000 Monte Carlo trials. The data are generated using gamma, truncated gamma, mixture normal, and lognormal frailty distributions and then inverse Gaussian frailty models (different baseline hazards) are fitted. The bold typeface indicates cases where the relative bias is greater than 5%.

			IG _Weibull			IG _Plug-in			IG _Splines		
	θ	G	Bias	SD	MSE	Bias	SD	MSE	Bias	SD	MSE
Gamma	0.1	100	0.034	0.175	0.028	0.016	0.222	0.050	−0.009	0.212	0.044
		200	0.016	0.162	0.025	0.004	0.153	0.023	0.014	0.165	0.028
		400	−0.011	0.122	0.015	0.007	0.112	0.013	−0.008	0.111	0.012
		800	−0.013	0.081	0.007	−0.008	0.092	0.008	0.004	0.079	0.006
	0.5	100	−0.008	0.239	0.057	−0.042	0.238	0.061	0.002	0.249	0.062
		200	−0.004	0.175	0.031	−0.029	0.176	0.033	−0.014	0.168	0.029
		400	−0.010	0.117	0.014	−0.032	0.119	0.016	−0.019	0.127	0.017
		800	−0.009	0.084	0.007	−0.031	0.085	0.009	−0.024	0.086	0.009
	2	100	**−0.170**	0.265	0.135	**−0.216**	0.250	0.168	**−0.144**	0.257	0.113
		200	**−0.156**	0.187	0.090	**−0.196**	0.182	0.120	**−0.157**	0.177	0.087
		400	**−0.167**	0.126	0.079	**−0.205**	0.124	0.110	**−0.161**	0.124	0.074
		800	**−0.167**	0.085	0.070	**−0.204**	0.086	0.101	**−0.161**	0.089	0.066
Truncated Gamma	0.1	100	0.047	0.209	0.038	0.009	0.219	0.048	0.029	0.229	0.054
		200	−0.011	0.127	0.016	−0.007	0.154	0.024	0.003	0.143	0.020
		400	0.018	0.143	0.020	−0.003	0.112	0.012	0.006	0.113	0.013
		800	−0.005	0.083	0.007	−0.002	0.079	0.006	0.008	0.077	0.006
	0.5	100	−0.024	0.216	0.045	0.005	0.244	0.060	0.019	0.265	0.070
		200	−0.012	0.155	0.024	−0.012	0.168	0.028	−0.001	0.170	0.029
		400	−0.008	0.114	0.013	−0.003	0.129	0.017	0.003	0.127	0.016
		800	−0.007	0.078	0.006	−0.002	0.088	0.008	0.000	0.085	0.007
	2	100	0.036	0.274	0.078	0.006	0.265	0.070	0.023	0.091	0.009
		200	0.012	0.182	0.033	0.012	0.187	0.035	0.025	0.193	0.038
		400	0.014	0.131	0.018	0.014	0.134	0.018	0.024	0.138	0.020
		800	0.014	0.088	0.008	0.018	0.091	0.009	0.023	0.091	0.009
Lognormal	0.1	100	0.020	0.257	0.059	0.018	0.216	0.047	0.043	0.201	0.044
		200	0.017	0.127	0.016	0.002	0.152	0.023	0.021	0.169	0.029
		400	0.010	0.121	0.014	−0.005	0.116	0.013	0.007	0.118	0.014
		800	0.001	0.073	0.005	−0.003	0.078	0.006	0.006	0.073	0.005
	0.5	100	0.015	0.236	0.056	−0.017	0.247	0.062	0.014	0.252	0.064
		200	0.011	0.163	0.027	0.007	0.166	0.028	0.008	0.184	0.034
		400	0.001	0.121	0.015	−0.006	0.128	0.016	0.005	0.125	0.016
		800	0.000	0.081	0.006	−0.004	0.085	0.007	−0.001	0.086	0.007
	2	100	−0.028	0.274	0.077	**−0.071**	0.264	0.081	−0.037	0.267	0.074
		200	−0.034	0.182	0.036	**−0.059**	0.178	0.039	−0.036	0.172	0.032
		400	−0.038	0.137	0.022	**−0.056**	0.135	0.025	−0.037	0.139	0.022
		800	−0.041	0.091	0.012	**−0.055**	0.092	0.015	−0.043	0.097	0.014
Mixture Normal	0.1	100	0.014	0.262	0.064	−0.014	0.245	0.060	0.015	0.244	0.060
		200	0.004	0.145	0.021	0.000	0.171	0.029	0.012	0.180	0.033
		400	0.005	0.112	0.013	0.004	0.121	0.015	−0.001	0.121	0.015
		800	0.002	0.080	0.006	0.004	0.089	0.008	0.001	0.088	0.008
	0.5	100	**−0.113**	0.257	0.095	**−0.174**	0.246	0.128	**−0.111**	0.262	0.097
		200	**−0.114**	0.188	0.065	**−0.168**	0.185	0.098	**−0.113**	0.194	0.066
		400	**−0.117**	0.132	0.048	**−0.167**	0.130	0.080	**−0.124**	0.134	0.053
		800	**−0.124**	0.093	0.043	**−0.171**	0.094	0.074	**−0.123**	0.100	0.044
	2	100	**−0.393**	0.289	0.431	**−0.473**	0.293	0.589	**−0.343**	0.312	0.362
		200	**−0.404**	0.209	0.411	**−0.480**	0.212	0.563	**−0.343**	0.215	0.310
		400	**−0.400**	0.143	0.381	**−0.477**	0.150	0.534	**−0.359**	0.156	0.314
		800	**−0.406**	0.100	0.381	**−0.478**	0.112	0.527	**−0.357**	0.113	0.299

Relative bias (Bias), standard deviation (SD), and mean square error (MSE) of population survival probability $S_p(0.3)$ without covariate ($x = 0$) across 1000 Monte Carlo trials. The data are generated using gamma, truncated gamma, and mixture normal frailty distributions and then gamma frailty models (different baseline hazards) are fitted.

	True			Gamma _Weibull			Gamma _Plug-in			Gamma _Splines		
	$S_p(0.3)$	θ	G	Bias	SD	MSE	Bias	SD	MSE	Bias	SD	MSE
Gamma	0.914	0.1	100	−0.001	0.017	0.000	0.004	0.019	0.000	−0.002	0.016	0.000
			200	0.000	0.011	0.000	0.006	0.013	0.000	0.003	0.014	0.000
			400	0.000	0.008	0.000	0.007	0.010	0.000	0.001	0.008	0.000
			800	0.000	0.006	0.000	0.000	0.007	0.000	0.000	0.007	0.000
	0.916	0.5	100	−0.002	0.017	0.000	−0.004	0.019	0.000	0.002	0.021	0.000
			200	0.000	0.013	0.000	−0.002	0.015	0.000	−0.006	0.031	0.001
			400	0.000	0.009	0.000	−0.001	0.010	0.000	0.000	0.010	0.000
			800	0.000	0.006	0.000	−0.001	0.008	0.000	0.000	0.007	0.000
	0.921	2	100	−0.001	0.020	0.000	−0.002	0.022	0.000	0.001	0.020	0.000
			200	0.000	0.013	0.000	−0.001	0.015	0.000	0.000	0.015	0.000
			400	0.000	0.010	0.000	−0.001	0.011	0.000	0.000	0.010	0.000
			800	0.000	0.007	0.000	0.000	0.008	0.000	0.000	0.007	0.000
Truncated Gamma	0.914	0.1	100	0.003	0.014	0.000	0.002	0.018	0.000	0.000	0.014	0.000
			200	0.002	0.012	0.000	−0.001	0.013	0.000	0.000	0.013	0.000
			400	0.001	0.008	0.000	0.000	0.009	0.000	0.000	0.010	0.000
			800	0.000	0.006	0.000	0.000	0.007	0.000	−0.001	0.007	0.000
	0.916	0.5	100	0.003	0.015	0.000	0.001	0.021	0.000	−0.001	0.019	0.000
			200	0.001	0.012	0.000	−0.001	0.013	0.000	0.000	0.013	0.000
			400	0.000	0.008	0.000	0.002	0.010	0.000	−0.001	0.018	0.000
			800	0.001	0.006	0.000	0.001	0.007	0.000	0.000	0.007	0.000
	0.913	2	100	0.003	0.018	0.000	0.002	0.022	0.000	0.000	0.020	0.000
			200	0.003	0.013	0.000	0.002	0.015	0.000	0.000	0.014	0.000
			400	0.002	0.009	0.000	0.002	0.010	0.000	0.001	0.010	0.000
			800	0.003	0.006	0.000	0.002	0.007	0.000	0.001	0.007	0.000
Lognormal	0.914	0.1	100	−0.001	0.024	0.001	0.003	0.019	0.000	−0.001	0.017	0.000
			200	0.001	0.011	0.000	0.001	0.013	0.000	0.002	0.014	0.000
			400	0.000	0.009	0.000	0.000	0.010	0.000	0.000	0.009	0.000
			800	0.000	0.006	0.000	0.000	0.007	0.000	0.000	0.006	0.000
	0.916	0.5	100	0.001	0.019	0.000	−0.002	0.020	0.000	−0.027	0.054	0.003
			200	0.001	0.012	0.000	0.000	0.014	0.000	−0.005	0.021	0.000
			400	0.000	0.009	0.000	−0.001	0.010	0.000	−0.001	0.018	0.000
			800	0.001	0.006	0.000	0.000	0.007	0.000	0.000	0.007	0.000
	0.927	2	100	0.004	0.017	0.000	0.002	0.019	0.000	0.003	0.019	0.000
			200	0.005	0.012	0.000	0.005	0.014	0.000	0.002	0.013	0.000
			400	0.004	0.009	0.000	0.004	0.010	0.000	0.003	0.010	0.000
			800	0.005	0.006	0.000	0.004	0.007	0.000	0.003	0.007	0.000
Mixture Normal	0.915	0.1	100	0.000	0.018	0.000	0.000	0.019	0.000	−0.003	0.020	0.000
			200	0.001	0.010	0.000	0.001	0.013	0.000	−0.009	0.025	0.001
			400	0.000	0.009	0.000	0.000	0.010	0.000	−0.027	0.052	0.003
			800	0.000	0.006	0.000	−0.001	0.007	0.000	−0.021	0.046	0.002
	0.921	0.5	100	0.000	0.019	0.000	−0.002	0.022	0.000	−0.003	0.020	0.000
			200	0.000	0.014	0.000	−0.001	0.015	0.000	−0.004	0.015	0.000
			400	−0.001	0.010	0.000	−0.001	0.011	0.000	−0.004	0.010	0.000
			800	−0.001	0.007	0.000	−0.001	0.008	0.000	−0.003	0.007	0.000
	0.949	2	100	−0.014	0.020	0.001	−0.015	0.021	0.001	−0.012	0.020	0.001
			200	−0.014	0.014	0.000	−0.015	0.015	0.000	−0.012	0.014	0.000
			400	−0.014	0.010	0.000	−0.015	0.011	0.001	−0.012	0.010	0.000
			800	−0.015	0.007	0.000	−0.016	0.008	0.000	−0.012	0.007	0.000

Relative bias (Bias), standard deviation (SD), and mean square error (MSE) of population survival probability $S_p(0.3)$ with covariate ($x = 1$) across 1000 Monte Carlo trials. The data are generated using gamma, truncated gamma, and mixture normal frailty distributions and then gamma frailty models(different baseline hazards) are fitted.

	True			Gamma _Weibull			Gamma _Plug-in			Gamma _Splines		
	$S_p(0.3)$	θ	G	Bias	SD	MSE	Bias	SD	MSE	Bias	SD	MSE
Gamma	0.673	0.1	100	−0.003	0.038	0.001	−0.003	0.045	0.002	0.002	0.041	0.002
			200	0.001	0.026	0.001	−0.005	0.032	0.001	−0.006	0.029	0.001
			400	0.000	0.018	0.000	−0.008	0.023	0.001	−0.008	0.022	0.001
			800	0.000	0.013	0.000	0.005	0.016	0.000	0.001	0.015	0.000
	0.693	0.5	100	−0.001	0.038	0.001	0.003	0.044	0.002	0.008	0.039	0.002
			200	0.003	0.028	0.001	0.006	0.032	0.001	−0.012	0.043	0.002
			400	0.001	0.019	0.000	0.003	0.023	0.001	−0.001	0.022	0.000
			800	0.000	0.014	0.000	0.001	0.015	0.000	−0.001	0.016	0.000
	0.744	2	100	0.002	0.038	0.001	0.007	0.042	0.002	−0.001	0.041	0.002
			200	−0.001	0.028	0.001	0.002	0.031	0.001	−0.001	0.029	0.001
			400	0.000	0.019	0.000	0.001	0.021	0.000	0.001	0.021	0.000
			800	0.000	0.013	0.000	0.000	0.015	0.000	−0.001	0.014	0.000
Truncated Gamma	0.673	0.1	100	−0.015	0.034	0.001	0.005	0.046	0.002	−0.001	0.027	0.001
			200	0.010	0.026	0.001	0.001	0.031	0.001	−0.001	0.030	0.001
			400	−0.002	0.015	0.000	0.001	0.022	0.000	−0.003	0.024	0.001
			800	0.001	0.013	0.000	0.000	0.015	0.000	−0.001	0.016	0.000
	0.690	0.5	100	−0.006	0.033	0.001	0.003	0.047	0.002	0.000	0.040	0.002
			200	0.005	0.027	0.001	0.002	0.031	0.001	−0.003	0.032	0.001
			400	0.004	0.019	0.000	−0.002	0.023	0.001	−0.002	0.031	0.001
			800	0.005	0.013	0.000	−0.002	0.016	0.000	0.000	0.015	0.000
	0.711	2	100	0.002	0.040	0.002	−0.001	0.047	0.002	−0.008	0.041	0.002
			200	0.004	0.029	0.001	−0.003	0.032	0.001	−0.008	0.032	0.001
			400	0.005	0.020	0.000	−0.004	0.023	0.001	−0.007	0.023	0.001
			800	0.006	0.015	0.000	−0.006	0.016	0.000	−0.008	0.015	0.000
Lognormal	0.673	0.1	100	−0.009	0.048	0.002	−0.003	0.043	0.002	−0.001	0.035	0.001
			200	0.002	0.019	0.000	0.002	0.030	0.001	−0.003	0.034	0.001
			400	−0.004	0.021	0.000	0.000	0.023	0.001	−0.002	0.020	0.000
			800	0.002	0.013	0.000	0.001	0.016	0.000	0.001	0.017	0.000
	0.696	0.5	100	0.006	0.034	0.001	0.000	0.044	0.002	0.065	0.072	0.007
			200	0.003	0.026	0.001	0.002	0.031	0.001	−0.006	0.030	0.001
			400	0.002	0.020	0.000	0.001	0.023	0.001	−0.002	0.031	0.001
			800	0.003	0.013	0.000	0.001	0.016	0.000	−0.001	0.015	0.000
	0.771	2	100	−0.002	0.038	0.001	−0.006	0.043	0.002	−0.011	0.043	0.002
			200	−0.002	0.027	0.001	−0.006	0.030	0.001	−0.014	0.031	0.001
			400	−0.001	0.020	0.000	−0.007	0.022	0.001	−0.013	0.021	0.001
			800	−0.001	0.014	0.000	−0.007	0.016	0.000	−0.011	0.015	0.000
Mixture Normal	0.684	0.1	100	−0.008	0.045	0.002	0.007	0.043	0.002	−0.002	0.047	0.002
			200	0.008	0.029	0.001	0.002	0.032	0.001	−0.007	0.046	0.002
			400	0.003	0.019	0.000	0.001	0.022	0.000	−0.055	0.090	0.009
			800	0.003	0.013	0.000	0.001	0.016	0.000	−0.037	0.064	0.005
	0.748	0.5	100	0.005	0.040	0.002	0.008	0.044	0.002	−0.007	0.039	0.002
			200	0.003	0.027	0.001	0.006	0.030	0.001	−0.004	0.029	0.001
			400	0.003	0.019	0.000	0.004	0.021	0.000	−0.001	0.021	0.000
			800	0.003	0.013	0.000	0.003	0.015	0.000	−0.002	0.014	0.000
	0.838	2	100	−0.005	0.033	0.001	−0.003	0.037	0.001	0.000	0.031	0.001
			200	−0.006	0.023	0.001	−0.005	0.026	0.001	−0.002	0.023	0.001
			400	−0.006	0.015	0.000	−0.007	0.017	0.000	−0.001	0.017	0.000
			800	−0.007	0.011	0.000	−0.008	0.012	0.000	−0.002	0.011	0.000

Relative bias (Bias), standard deviation (SD), and mean square error (MSE) of population survival probability $S_p(0.3)$ without covariate ($x = 0$) across 1000 Monte Carlo trials. The data are generated using gamma, truncated gamma, and mixture normal frailty distributions and then inverse Gaussian frailty models (different baseline hazards) are fitted.

	True			IG _Weibull			IG _Plug-in			IG _Splines		
	$S_p(0.3)$	θ	G	Bias	SD	MSE	Bias	SD	MSE	Bias	SD	MSE
Gamma	0.914	0.1	100	0.007	0.010	0.000	0.000	0.019	0.000	−0.006	0.018	0.000
			200	0.004	0.017	0.000	0.000	0.012	0.000	0.000	0.013	0.000
			400	−0.003	0.010	0.000	0.000	0.009	0.000	−0.002	0.009	0.000
			800	−0.001	0.006	0.000	0.000	0.007	0.000	0.001	0.007	0.000
	0.916	0.5	100	−0.006	0.018	0.000	−0.007	0.002	0.000	−0.003	0.020	0.000
			200	−0.004	0.013	0.000	−0.005	0.014	0.000	−0.004	0.014	0.000
			400	−0.005	0.009	0.000	−0.005	0.010	0.000	−0.004	0.010	0.000
			800	−0.005	0.006	0.000	−0.005	0.007	0.000	−0.004	0.007	0.000
	0.921	2	100	−0.016	0.019	0.001	−0.022	0.023	0.001	−0.013	0.020	0.001
			200	−0.015	0.012	0.000	−0.022	0.016	0.001	−0.014	0.014	0.000
			400	−0.016	0.009	0.000	−0.023	0.012	0.001	−0.014	0.010	0.000
			800	−0.016	0.006	0.000	−0.023	0.008	0.001	−0.014	0.008	0.000
Truncated Gamma	0.914	0.1	100	−0.008	0.013	0.000	0.001	0.019	0.000	0.001	0.020	0.000
			200	0.003	0.012	0.000	−0.001	0.013	0.000	0.000	0.013	0.000
			400	0.001	0.008	0.000	0.000	0.009	0.000	0.000	0.009	0.000
			800	0.001	0.007	0.000	0.000	0.007	0.000	0.000	0.006	0.000
	0.916	0.5	100	0.001	0.018	0.000	0.002	0.019	0.000	0.001	0.022	0.000
			200	0.001	0.011	0.000	0.000	0.014	0.000	0.001	0.012	0.000
			400	0.001	0.009	0.000	0.001	0.010	0.000	0.001	0.010	0.000
			800	0.001	0.006	0.000	0.001	0.007	0.000	0.001	0.007	0.000
	0.913	2	100	0.001	0.019	0.000	0.000	0.021	0.000	0.002	0.007	0.000
			200	0.002	0.013	0.000	0.002	0.015	0.000	0.001	0.014	0.000
			400	0.002	0.010	0.000	0.002	0.010	0.000	0.002	0.010	0.000
			800	0.002	0.007	0.000	0.002	0.007	0.000	0.002	0.007	0.000
Lognormal	0.914	0.1	100	−0.001	0.024	0.001	0.001	0.018	0.000	0.001	0.019	0.000
			200	0.001	0.011	0.000	0.000	0.013	0.000	0.003	0.012	0.000
			400	0.000	0.009	0.000	−0.001	0.010	0.000	0.001	0.009	0.000
			800	0.000	0.006	0.000	0.000	0.007	0.000	0.001	0.006	0.000
	0.916	0.5	100	0.000	0.019	0.000	−0.003	0.019	0.000	−0.002	0.020	0.000
			200	0.000	0.012	0.000	−0.005	0.014	0.000	−0.001	0.013	0.000
			400	−0.001	0.009	0.000	−0.001	0.010	0.000	0.000	0.010	0.000
			800	0.000	0.006	0.000	0.000	0.007	0.000	0.000	0.007	0.000
	0.927	2	100	−0.004	0.018	0.000	−0.005	0.020	0.000	−0.003	0.019	0.000
			200	−0.004	0.013	0.000	−0.004	0.014	0.000	−0.007	0.014	0.000
			400	−0.004	0.009	0.000	−0.004	0.010	0.000	−0.003	0.010	0.000
			800	−0.004	0.007	0.000	−0.005	0.008	0.000	−0.003	0.007	0.000
Mixture Normal	0.915	0.1	100	0.000	0.018	0.000	−0.002	0.190	0.000	−0.003	0.020	0.000
			200	0.001	0.011	0.000	0.000	0.014	0.000	−0.001	0.013	0.000
			400	0.000	0.009	0.000	0.000	0.010	0.000	−0.001	0.009	0.000
			800	0.000	0.006	0.000	0.000	0.007	0.000	0.000	0.007	0.000
	0.921	0.5	100	−0.006	0.017	0.000	−0.020	0.023	0.001	−0.010	0.020	0.000
			200	−0.007	0.012	0.000	−0.020	0.016	0.001	−0.009	0.015	0.000
			400	−0.007	0.009	0.000	−0.020	0.011	0.000	−0.009	0.011	0.000
			800	−0.008	0.006	0.000	−0.021	0.008	0.000	−0.010	0.008	0.000
	0.949	2	100	−0.014	0.014	0.000	−0.035	0.022	0.002	−0.019	0.018	0.001
			200	−0.015	0.010	0.000	−0.037	0.016	0.001	−0.021	0.014	0.001
			400	−0.015	0.007	0.000	−0.038	0.011	0.001	−0.020	0.010	0.000
			800	−0.016	0.005	0.000	−0.038	0.008	0.001	−0.020	0.007	0.000

Relative bias (Bias), standard deviation (SD), and mean square error (MSE) of population survival probability $S_p(0.3)$ with covariate ($x = 1$) across 1000 Monte Carlo trials. The data are generated using gamma, truncated gamma, and mixture normal frailty distributions and then inverse Gaussian frailty models (different baseline hazards) are fitted. The bold typeface indicates cases where the relative bias is greater than 5%.

	True			IG _Weibull			IG _Plug-in			IG _Splines		
	$S_p(0.3)$	θ	G	Bias	SD	MSE	Bias	SD	MSE	Bias	SD	MSE
Gamma	0.673	0.1	100	0.024	0.041	0.002	0.002	0.044	0.002	−0.010	0.041	0.002
			200	0.018	0.022	0.001	0.001	0.032	0.001	−0.003	0.029	0.001
			400	−0.001	0.015	0.000	0.000	0.023	0.001	−0.001	0.022	0.000
			800	0.004	0.011	0.000	0.006	0.015	0.000	0.004	0.015	0.000
	0.693	0.5	100	0.000	0.037	0.001	0.007	0.044	0.000	0.009	0.041	0.002
			200	0.003	0.027	0.001	0.010	0.031	0.001	0.006	0.030	0.001
			400	0.001	0.180	0.000	0.008	0.022	0.001	0.007	0.022	0.000
			800	0.000	0.013	0.000	0.007	0.015	0.000	0.007	0.015	0.000
	0.744	2	100	0.042	0.032	0.002	0.038	0.040	0.002	0.043	0.037	0.002
			200	0.039	0.023	0.001	0.033	0.029	0.001	0.043	0.026	0.002
			400	0.040	0.016	0.001	0.034	0.020	0.001	0.043	0.019	0.001
			800	0.040	0.011	0.001	0.033	0.014	0.001	0.042	0.014	0.001
Truncated Gamma	0.673	0.1	100	−0.026	0.024	0.001	0.006	0.046	0.002	−0.002	0.046	0.002
			200	−0.013	0.021	0.000	0.001	0.031	0.001	−0.003	0.029	0.001
			400	0.008	0.018	0.000	0.001	0.022	0.000	0.000	0.022	0.000
			800	−0.002	0.014	0.000	0.000	0.015	0.000	−0.002	0.014	0.000
	0.690	0.5	100	0.010	0.030	0.001	0.003	0.046	0.002	0.002	0.048	0.002
			200	−0.001	0.027	0.001	0.001	0.031	0.001	−0.001	0.032	0.001
			400	0.002	0.019	0.000	−0.001	0.023	0.001	−0.002	0.022	0.001
			800	0.001	0.013	0.000	−0.002	0.016	0.000	−0.003	0.015	0.000
	0.711	2	100	0.003	0.039	0.002	0.001	0.046	0.002	−0.005	0.015	0.000
			200	0.000	0.028	0.001	0.000	0.031	0.001	−0.004	0.030	0.001
			400	0.000	0.010	0.000	−0.001	0.022	0.000	−0.004	0.022	0.000
			800	−0.001	0.013	0.000	−0.003	0.015	0.000	−0.005	0.015	0.000
Lognormal	0.673	0.1	100	−0.009	0.047	0.002	−0.001	0.043	0.002	−0.011	0.045	0.002
			200	0.001	0.019	0.000	0.002	0.030	0.001	0.001	0.030	0.001
			400	−0.005	0.021	0.000	0.001	0.023	0.001	0.001	0.009	0.000
			800	0.001	0.013	0.000	0.002	0.016	0.000	0.000	0.015	0.000
	0.696	0.5	100	0.003	0.033	0.001	0.002	0.043	0.002	−0.004	0.043	0.002
			200	0.000	0.025	0.001	0.003	0.030	0.001	0.000	0.030	0.001
			400	−0.001	0.019	0.000	0.001	0.023	0.001	0.000	0.022	0.000
			800	0.000	0.013	0.000	0.002	0.016	0.000	0.001	0.015	0.000
	0.771	2	100	0.005	0.033	0.001	0.007	0.039	0.002	0.012	0.035	0.001
			200	0.005	0.023	0.001	0.008	0.028	0.001	−0.001	0.028	0.001
			400	0.004	0.017	0.000	0.007	0.020	0.000	0.006	0.019	0.000
			800	0.005	0.012	0.000	0.007	0.014	0.000	0.007	0.014	0.000
Mixture Normal	0.684	0.1	100	−0.011	0.045	0.002	0.009	0.043	0.002	−0.006	0.041	0.002
			200	0.006	0.029	0.001	0.004	0.031	0.001	−0.002	0.030	0.001
			400	0.001	0.019	0.000	0.002	0.022	0.001	0.001	0.020	0.000
			800	0.001	0.013	0.000	0.001	0.016	0.000	0.000	0.015	0.000
	0.748	0.5	100	**0.060**	0.030	0.003	0.040	0.040	0.002	**0.052**	0.036	0.003
			200	**0.058**	0.021	0.002	0.039	0.027	0.002	0.050	0.025	0.002
			400	**0.057**	0.015	0.002	0.038	0.019	0.001	**0.052**	0.019	0.002
			800	**0.057**	0.010	0.002	0.037	0.013	0.001	**0.051**	0.013	0.002
	0.838	2	100	**0.058**	0.023	0.003	0.031	0.034	0.002	0.045	0.030	0.002
			200	**0.057**	0.016	0.003	0.030	0.024	0.001	0.041	0.020	0.002
			400	**0.057**	0.011	0.002	0.029	0.017	0.001	0.045	0.015	0.002
			800	**0.056**	0.008	0.002	0.028	0.011	0.001	0.045	0.012	0.002

Relative bias (Bias), standard deviation (SD), and mean square error (MSE) of population survival probability $S_p(1)$ without covariate ($x = 0$) across 1000 Monte Carlo trials. The data are generated using gamma, truncated gamma, and mixture normal frailty distributions and then gamma frailty models (different baseline hazards) are fitted.

	True			Gamma _Weibull			Gamma _Plug-in			Gamma _Splines		
	$S_p(1)$	θ	G	Bias	SD	MSE	Bias	SD	MSE	Bias	SD	MSE
Gamma	0.386	0.1	100	−0.005	0.052	0.003	0.033	0.058	0.004	−0.028	0.052	0.003
			200	−0.001	0.030	0.001	0.043	0.042	0.002	0.029	0.044	0.002
			400	0.003	0.022	0.000	0.054	0.035	0.002	0.008	0.025	0.001
			800	−0.001	0.015	0.000	0.003	0.014	0.000	0.002	0.017	0.000
	0.444	0.5	100	−0.010	0.047	0.002	−0.004	0.054	0.003	0.016	0.056	0.003
			200	0.002	0.033	0.001	−0.001	0.037	0.001	0.003	0.041	0.002
			400	−0.002	0.023	0.001	−0.005	0.027	0.001	0.002	0.026	0.001
			800	0.000	0.017	0.000	−0.003	0.019	0.000	0.000	0.018	0.000
	0.577	2	100	0.000	0.046	0.002	0.005	0.050	0.003	0.003	0.050	0.002
			200	0.001	0.032	0.001	0.003	0.036	0.001	0.001	0.036	0.001
			400	0.000	0.023	0.001	0.000	0.026	0.001	0.001	0.025	0.001
			800	0.000	0.016	0.000	0.000	0.018	0.000	0.000	0.018	0.000
Truncated Gamma	0.386	0.1	100	−0.015	0.043	0.002	0.017	0.054	0.003	−0.003	0.033	0.001
			200	0.005	0.031	0.001	−0.004	0.038	0.001	0.016	0.034	0.001
			400	0.002	0.022	0.000	0.003	0.026	0.001	−0.001	0.024	0.001
			800	0.000	0.015	0.000	0.002	0.019	0.000	0.000	0.018	0.000
	0.431	0.5	100	−0.011	0.044	0.002	0.027	0.061	0.004	0.005	0.052	0.003
			200	−0.006	0.033	0.001	−0.005	0.039	0.002	0.003	0.038	0.001
			400	−0.002	0.023	0.001	0.021	0.033	0.001	0.006	0.030	0.001
			800	−0.004	0.016	0.000	0.013	0.023	0.001	0.006	0.019	0.000
	0.505	2	100	−0.011	0.050	0.003	0.013	0.057	0.003	0.002	0.055	0.003
			200	−0.023	0.036	0.001	−0.003	0.041	0.002	−0.003	0.039	0.001
			400	−0.016	0.024	0.001	0.000	0.028	0.001	0.001	0.027	0.001
			800	−0.018	0.017	0.000	−0.003	0.020	0.000	0.001	0.019	0.000
Lognormal	0.386	0.1	100	0.003	0.047	0.002	0.029	0.057	0.003	0.014	0.043	0.002
			200	−0.014	0.026	0.001	0.006	0.037	0.001	−0.002	0.033	0.001
			400	0.002	0.022	0.000	0.002	0.026	0.001	0.004	0.027	0.001
			800	−0.001	0.016	0.000	0.002	0.019	0.000	−0.003	0.018	0.000
	0.449	0.5	100	0.005	0.048	0.002	0.007	0.055	0.003	−0.055	0.055	0.003
			200	0.000	0.033	0.001	0.006	0.038	0.001	−0.013	0.038	0.001
			400	−0.004	0.024	0.001	−0.004	0.027	0.001	0.006	0.030	0.001
			800	−0.005	0.017	0.000	−0.005	0.019	0.000	0.002	0.018	0.000
	0.608	2	100	−0.021	0.049	0.003	−0.002	0.051	0.003	−0.007	0.053	0.003
			200	−0.019	0.035	0.001	−0.002	0.038	0.001	−0.007	0.038	0.001
			400	−0.019	0.025	0.001	−0.005	0.028	0.001	−0.007	0.027	0.001
			800	−0.020	0.018	0.000	−0.005	0.019	0.000	−0.006	0.019	0.000
Mixture Normal	0.418	0.1	100	−0.001	0.046	0.002	0.021	0.056	0.003	−0.004	0.054	0.003
			200	0.007	0.030	0.001	0.025	0.043	0.002	−0.026	0.032	0.001
			400	0.001	0.023	0.001	0.009	0.028	0.001	−0.023	0.022	0.001
			800	0.001	0.016	0.000	0.005	0.020	0.000	−0.006	0.021	0.000
	0.598	0.5	100	0.011	0.046	0.002	0.036	0.051	0.003	0.007	0.047	0.002
			200	0.007	0.033	0.001	0.029	0.037	0.002	0.005	0.033	0.001
			400	0.007	0.023	0.001	0.026	0.026	0.001	0.005	0.023	0.001
			800	0.005	0.016	0.000	0.023	0.018	0.001	0.007	0.016	0.000
	0.743	2	100	0.014	0.038	0.002	−0.004	0.043	0.002	0.005	0.040	0.002
			200	0.011	0.027	0.001	−0.008	0.031	0.001	0.003	0.028	0.001
			400	0.011	0.019	0.000	−0.007	0.021	0.000	0.003	0.020	0.000
			800	0.010	0.013	0.000	−0.009	0.015	0.000	0.004	0.014	0.000

Relative bias (Bias), standard deviation (SD), and mean square error (MSE) of population survival probability $S_p(1)$ with covariate ($x = 1$) across 1000 Monte Carlo trials. The data are generated using gamma, truncated gamma, and mixture normal frailty distributions and then gamma frailty models (different baseline hazards) are fitted. The bold typeface indicates cases where the relative bias is greater than 5%.

	True			Gamma _Weibull			Gamma _Plug-in			Gamma _Splines		
	$S_p(1)$	θ	G	Bias	SD	MSE	Bias	SD	MSE	Bias	SD	MSE
Gamma	0.025	0.1	100	**0.071**	0.039	0.002	0.000	0.014	0.000	−0.016	0.013	0.000
			200	0.017	0.009	0.000	**−0.069**	0.010	0.000	**−0.064**	0.007	0.000
			400	−0.004	0.006	0.000	**−0.151**	0.008	0.000	−0.017	0.005	0.000
			800	0.005	0.004	0.000	0.010	0.005	0.000	0.003	0.005	0.000
	0.095	0.5	100	−0.010	0.026	0.001	0.014	0.030	0.001	0.020	0.032	0.001
			200	0.004	0.018	0.000	0.024	0.020	0.000	0.024	0.020	0.000
			400	−0.001	0.013	0.000	0.016	0.014	0.000	0.010	0.014	0.000
			800	0.003	0.009	0.000	0.014	0.010	0.000	0.001	0.010	0.000
	0.317	2	100	0.006	0.043	0.002	0.007	0.047	0.002	−0.004	0.046	0.002
			200	−0.004	0.032	0.001	−0.005	0.035	0.001	0.003	0.033	0.001
			400	0.000	0.021	0.000	−0.001	0.023	0.001	0.004	0.024	0.001
			800	−0.002	0.015	0.000	−0.003	0.016	0.000	0.000	0.016	0.000
Truncated Gamma	0.024	0.1	100	0.013	0.011	0.000	**0.081**	0.012	0.000	0.019	0.008	0.000
			200	0.003	0.008	0.000	**0.053**	0.009	0.000	0.000	0.008	0.000
			400	−0.011	0.006	0.000	0.021	0.006	0.000	−0.003	0.007	0.000
			800	−0.001	0.004	0.000	0.011	0.005	0.000	0.007	0.005	0.000
	0.051	0.5	100	−0.003	0.018	0.000	0.024	0.021	0.000	0.002	0.021	0.000
			200	0.024	0.012	0.000	0.030	0.014	0.000	0.018	0.013	0.000
			400	0.018	0.009	0.000	−0.019	0.012	0.000	0.032	0.026	0.001
			800	0.008	0.006	0.000	−0.007	0.008	0.000	0.028	0.007	0.000
	0.149	2	100	−0.056	0.028	0.001	−0.007	0.034	0.001	−0.012	0.033	0.001
			200	−0.054	0.020	0.000	−0.007	0.025	0.001	−0.006	0.024	0.001
			400	−0.053	0.014	0.000	−0.013	0.017	0.000	−0.008	0.016	0.000
			800	−0.055	0.010	0.000	−0.013	0.012	0.000	−0.013	0.011	0.000
Lognormal	0.024	0.1	100	0.028	0.013	0.000	0.027	0.014	0.000	0.009	0.014	0.000
			200	0.033	0.007	0.000	**0.054**	0.009	0.000	−0.013	0.009	0.000
			400	0.008	0.005	0.000	0.023	0.006	0.000	0.022	0.007	0.000
			800	0.006	0.004	0.000	0.018	0.005	0.000	−0.011	0.004	0.000
	0.082	0.5	100	0.010	0.024	0.001	−0.001	0.026	0.001	−0.002	0.030	0.001
			200	−0.004	0.016	0.000	0.012	0.019	0.000	**0.063**	0.020	0.000
			400	−0.011	0.011	0.000	0.015	0.013	0.000	0.032	0.026	0.001
			800	−0.005	0.008	0.000	0.027	0.010	0.000	0.011	0.009	0.000
	0.292	2	100	**−0.054**	0.038	0.002	−0.019	0.044	0.002	0.024	0.044	0.002
			200	−0.045	0.028	0.001	−0.012	0.032	0.001	−0.016	0.032	0.001
			400	−0.048	0.020	0.001	−0.024	0.022	0.001	−0.021	0.021	0.000
			800	−0.047	0.013	0.000	−0.024	0.015	0.000	−0.021	0.015	0.000
Mixture Normal	0.053	0.1	100	−0.007	0.021	0.000	0.038	0.022	0.000	0.048	0.021	0.000
			200	−0.018	0.011	0.000	−0.022	0.016	0.000	−0.025	0.011	0.000
			400	−0.020	0.009	0.000	−0.016	0.011	0.000	0.007	0.009	0.000
			800	−0.013	0.006	0.000	0.004	0.007	0.000	0.036	0.008	0.000
	0.395	0.5	100	**−0.088**	0.042	0.003	**−0.066**	0.047	0.003	**−0.102**	0.042	0.003
			200	**−0.095**	0.028	0.002	**−0.077**	0.033	0.002	**−0.107**	0.031	0.003
			400	**−0.095**	0.021	0.002	**−0.079**	0.024	0.002	**−0.102**	0.021	0.002
			800	**−0.095**	0.014	0.002	**−0.079**	0.016	0.001	**−0.104**	0.015	0.002
	0.599	2	100	0.048	0.048	0.003	0.019	0.051	0.003	0.038	0.047	0.003
			200	0.045	0.033	0.002	0.017	0.035	0.001	0.036	0.034	0.002
			400	0.045	0.023	0.001	0.018	0.025	0.001	0.038	0.024	0.001
			800	0.043	0.015	0.001	0.018	0.017	0.000	0.037	0.017	0.001

Relative bias (Bias), standard deviation (SD), and mean square error (MSE) of population survival probability $S_p(1)$ without covariate ($x = 0$) across 1000 Monte Carlo trials. The data are generated using gamma, truncated gamma, and mixture normal frailty distributions and then inverse Gaussian frailty models (different baseline hazards) are fitted. The bold typeface indicates cases where the relative bias is greater than 5%.

	True			IG _Weibull			IG _Plug-in			IG _Splines		
	$S_p(1)$	θ	G	Bias	SD	MSE	Bias	SD	MSE	Bias	SD	MSE
Gamma	0.386	0.1	100	−0.033	0.045	0.002	0.011	0.054	0.003	−0.035	0.053	0.003
			200	0.009	0.030	0.001	0.010	0.037	0.001	0.005	0.036	0.001
			400	−0.008	0.025	0.001	0.008	0.027	0.001	−0.014	0.020	0.000
			800	−0.004	0.013	0.000	0.003	0.015	0.000	0.001	0.016	0.000
	0.444	0.5	100	0.011	0.045	0.002	−0.001	0.053	0.003	0.012	0.049	0.002
			200	0.021	0.032	0.001	0.003	0.036	0.001	0.003	0.036	0.001
			400	0.017	0.022	0.001	−0.003	0.026	0.001	0.002	0.026	0.001
			800	0.019	0.016	0.000	−0.003	0.018	0.000	−0.001	0.018	0.000
	0.577	2	100	**0.077**	0.043	0.004	−0.009	0.047	0.002	0.022	0.048	0.002
			200	**0.078**	0.030	0.003	−0.009	0.033	0.001	0.021	0.034	0.001
			400	**0.077**	0.022	0.002	−0.014	0.025	0.001	0.020	0.024	0.001
			800	**0.077**	0.015	0.002	−0.015	0.017	0.000	0.019	0.018	0.000
Truncated Gamma	0.386	0.1	100	−0.040	0.008	0.000	0.018	0.054	0.003	−0.001	0.057	0.003
			200	−0.036	0.030	0.001	−0.002	0.038	0.001	0.001	0.033	0.001
			400	0.012	0.019	0.000	0.004	0.026	0.001	0.003	0.025	0.001
			800	−0.003	0.018	0.000	0.002	0.019	0.000	0.009	0.018	0.000
	0.431	0.5	100	−0.001	0.037	0.001	0.022	0.055	0.003	−0.037	0.044	0.002
			200	−0.016	0.034	0.001	−0.003	0.039	0.001	−0.004	0.038	0.001
			400	−0.005	0.023	0.001	0.005	0.028	0.001	0.001	0.026	0.001
			800	−0.006	0.016	0.000	0.002	0.020	0.0008	0.002	0.018	0.000
	0.505	2	100	0.002	0.047	0.002	0.019	0.054	0.003	0.004	0.018	0.000
			200	−0.011	0.033	0.001	0.005	0.039	0.001	0.002	0.037	0.001
			400	−0.006	0.023	0.001	0.009	0.027	0.001	0.005	0.026	0.001
			800	−0.008	0.016	0.000	0.007	0.019	0.000	0.004	0.018	0.000
Lognormal	0.386	0.1	100	−0.001	0.047	0.002	0.017	0.051	0.003	0.005	0.048	0.002
			200	−0.014	0.026	0.001	0.007	0.037	0.001	0.000	0.035	0.001
			400	0.002	0.022	0.000	0.003	0.026	0.001	0.013	0.026	0.001
			800	−0.001	0.016	0.000	0.002	0.019	0.000	0.005	0.017	0.000
	0.449	0.5	100	0.013	0.045	0.002	0.005	0.054	0.003	0.004	0.056	0.003
			200	0.007	0.032	0.001	0.007	0.037	0.001	0.002	0.037	0.001
			400	0.003	0.023	0.001	0.001	0.027	0.001	0.005	0.026	0.001
			800	0.002	0.016	0.000	0.000	0.019	0.000	0.001	0.018	0.000
	0.608	2	100	0.012	0.043	0.002	−0.005	0.047	0.002	−0.003	0.046	0.002
			200	0.015	0.030	0.001	−0.005	0.034	0.001	−0.010	0.034	0.001
			400	0.012	0.022	0.001	−0.007	0.025	0.001	−0.001	0.025	0.001
			800	0.012	0.015	0.000	−0.008	0.017	0.000	−0.002	0.017	0.000
Mixture Normal	0.418	0.1	100	0.000	0.045	0.002	0.014	0.054	0.003	−0.001	0.054	0.003
			200	0.009	0.029	0.001	0.015	0.039	0.002	0.004	0.037	0.001
			400	0.003	0.022	0.000	0.008	0.026	0.001	0.002	0.025	0.001
			800	0.003	0.016	0.000	0.008	0.019	0.000	0.003	0.018	0.000
	0.598	0.5	100	**0.114**	0.041	0.006	0.041	0.049	0.003	**0.066**	0.048	0.004
			200	**0.111**	0.030	0.005	0.034	0.035	0.002	**0.064**	0.036	0.003
			400	**0.111**	0.021	0.005	0.032	0.029	0.001	**0.064**	0.025	0.002
			800	**0.108**	0.015	0.004	0.030	0.017	0.001	**0.064**	0.018	0.002
	0.743	2	100	**0.089**	0.032	0.005	−0.028	0.043	0.002	0.025	0.045	0.002
			200	**0.085**	0.023	0.005	−0.035	0.031	0.002	0.021	0.028	0.001
			400	**0.086**	0.017	0.004	−0.035	0.022	0.001	0.021	0.021	0.001
			800	**0.084**	0.012	0.004	−0.037	0.016	0.001	0.022	0.016	0.001

Relative bias (Bias), standard deviation (SD), and mean square error (MSE) of population survival probability $S_p(1)$ with covariate ($x = 1$) across 1000 Monte Carlo trials. The data are generated using gamma, truncated gamma, and mixture normal frailty distributions and then inverse Gaussian frailty models (different baseline hazards) are fitted. The bold typeface indicates cases where the relative bias is greater than 5%.

	True			IG _Weibull			IG _Plug-in			IG _Splines		
	$S_p(1)$	θ	G	Bias	SD	MSE	Bias	SD	MSE	Bias	SD	MSE
Gamma	0.025	0.1	100	**0.065**	0.012	0.000	0.031	0.014	0.000	0.025	0.014	0.000
			200	**0.076**	0.008	0.000	0.007	0.010	0.000	−0.014	0.009	0.000
			400	**0.080**	0.006	0.000	−0.029	0.007	0.000	−0.032	0.006	0.000
			800	0.019	0.003	0.000	0.008	0.005	0.000	0.009	0.005	0.000
	0.095	0.5	100	0.007	0.029	0.001	−0.017	0.030	0.001	−0.028	0.030	0.001
			200	0.026	0.021	0.000	−0.009	0.021	0.000	−0.019	0.020	0.000
			400	0.020	0.014	0.000	−0.017	0.015	0.000	−0.014	0.015	0.000
			800	0.025	0.010	0.000	−0.014	0.011	0.000	−0.020	0.010	0.000
	0.317	2	100	**0.153**	0.053	0.005	0.014	0.052	0.003	0.017	0.054	0.003
			200	**0.138**	0.040	0.004	0.000	0.038	0.001	0.026	0.038	0.002
			400	**0.147**	0.026	0.003	0.004	0.026	0.001	0.026	0.027	0.001
			800	**0.146**	0.018	0.002	0.003	0.018	0.000	0.024	0.019	0.000
Truncated Gamma	0.024	0.1	100	0.022	0.012	0.000	0.048	0.012	0.000	−0.004	0.012	0.000
			200	−0.068	0.006	0.000	0.027	0.009	0.000	−0.004	0.009	0.000
			400	−0.030	0.004	0.000	0.002	0.006	0.000	−0.001	0.006	0.000
			800	−0.023	0.003	0.000	0.001	0.005	0.000	−0.014	0.004	0.000
	0.051	0.5	100	0.017	0.019	0.000	0.029	0.020	0.000	0.048	0.018	0.000
			200	0.025	0.011	0.000	0.035	0.015	0.000	0.033	0.014	0.000
			400	0.019	0.009	0.000	0.030	0.010	0.000	0.036	0.010	0.000
			800	0.022	0.006	0.000	0.034	0.007	0.000	0.037	0.007	0.000
	0.149	2	100	−0.030	0.030	0.001	0.006	0.035	0.001	−0.004	0.012	0.000
			200	−0.023	0.022	0.000	0.003	0.025	0.001	−0.003	0.025	0.001
			400	−0.020	0.015	0.000	−0.001	0.018	0.000	−0.003	0.017	0.000
			800	−0.011	0.010	0.000	−0.002	0.012	0.000	−0.004	0.012	0.000
Lognormal	0.024	0.1	100	0.020	0.013	0.000	0.018	0.012	0.000	−0.017	0.014	0.000
			200	0.030	0.007	0.000	0.026	0.010	0.000	−0.032	0.009	0.000
			400	0.003	0.005	0.000	0.002	0.007	0.000	0.026	0.007	0.000
			800	0.005	0.004	0.000	0.008	0.005	0.000	−0.009	0.004	0.000
	0.082	0.5	100	0.026	0.026	0.001	−0.005	0.026	0.001	−0.023	0.028	0.001
			200	0.009	0.017	0.000	0.009	0.020	0.000	0.015	0.018	0.000
			400	0.001	0.012	0.000	−0.005	0.013	0.000	0.002	0.013	0.000
			800	0.008	0.008	0.000	0.006	0.009	0.000	0.003	0.009	0.000
	0.292	2	100	0.035	0.043	0.002	0.012	0.047	0.002	0.009	0.049	0.002
			200	0.048	0.032	0.001	0.017	0.034	0.001	−0.004	0.033	0.001
			400	0.042	0.022	0.001	0.011	0.024	0.001	0.015	0.025	0.001
			800	0.046	0.015	0.000	0.017	0.016	0.000	0.018	0.017	0.000
Mixture Normal	0.053	0.1	100	−0.016	0.022	0.000	0.043	0.021	0.000	0.000	0.022	0.000
			200	−0.015	0.011	0.000	−0.009	0.014	0.000	−0.003	0.015	0.000
			400	−0.019	0.009	0.000	−0.020	0.010	0.000	−0.001	0.010	0.000
			800	−0.012	0.006	0.000	−0.011	0.007	0.000	−0.014	0.007	0.000
	0.395	0.5	100	0.049	0.048	0.003	−0.045	0.054	0.003	−0.042	0.053	0.003
			200	0.043	0.034	0.001	**−0.058**	0.037	0.002	−0.046	0.039	0.002
			400	0.042	0.025	0.001	−0.065	0.027	0.001	−0.039	0.027	0.001
			800	0.043	0.016	0.001	−0.063	0.018	0.001	−0.042	0.019	0.001
	0.599	2	100	**0.175**	0.048	0.013	0.012	0.058	0.003	0.050	0.060	0.005
			200	**0.173**	0.033	0.012	0.004	0.041	0.002	0.043	0.040	0.002
			400	**0.172**	0.024	0.011	0.004	0.030	0.001	0.050	0.030	0.002
			800	**0.171**	0.016	0.011	0.003	0.020	0.000	0.050	0.023	0.001

Relative bias (Bias), standard deviation (SD), and mean square error (MSE) for baseline hazard scale parameter λ across 1000 Monte Carlo trials. The data are generated using gamma, truncated gamma, and mixture normal frailty distributions and then gamma and inverse Gaussian frailty models (Weibull baseline hazard) are fitted. The bold typeface indicates cases where the relative bias is greater than 5%.

			Gamma _Weibull			Inverse Gaussian _Weibull		
	θ	G	Bias	SD	MSE	Bias	SD	MSE
Gamma	0.1	100	0.026	0.283	0.081	0.048	0.151	0.021
		200	0.006	0.100	0.010	0.015	0.092	0.008
		400	0.000	0.072	0.005	0.027	0.081	0.007
		800	0.002	0.048	0.002	0.008	0.039	0.002
	0.5	100	0.032	0.196	0.039	**0.112**	0.248	0.074
		200	0.006	0.125	0.016	**0.084**	0.154	0.031
		400	0.001	0.088	0.008	**0.078**	0.108	0.018
		800	0.003	0.064	0.004	**0.072**	0.076	0.011
	2	100	**0.053**	0.300	0.093	**0.130**	0.313	0.115
		200	0.015	0.190	0.036	**0.103**	0.200	0.050
		400	0.011	0.134	0.018	**0.096**	0.141	0.079
		800	0.000	0.094	0.009	**0.093**	0.101	0.019
Truncated Gamma	0.1	100	**−0.117**	0.220	0.062	0.101	0.087	0.016
		200	**−0.140**	0.097	0.029	0.047	0.097	0.011
		400	**−0.136**	0.073	0.024	−0.008	0.051	0.003
		800	−0.038	0.052	0.022	0.003	0.055	0.003
	0.5	100	**−0.176**	0.186	0.066	**−0.056**	0.169	0.030
		200	**−0.199**	0.124	0.055	−0.019	0.133	0.018
		400	**−0.201**	0.098	0.051	−0.044	0.091	0.010
		800	**−0.201**	0.060	0.044	−0.044	0.063	0.006
	2	100	**−0.322**	0.250	0.166	**0.163**	0.430	0.211
		200	**−0.334**	0.158	0.136	**0.136**	0.260	0.086
		400	**−0.346**	0.107	0.131	**0.094**	0.167	0.037
		800	**−0.348**	0.091	0.126	**0.092**	0.116	0.022
Lognormal	0.1	100	0.026	0.150	0.020	0.003	0.143	0.018
		200	0.041	0.090	0.009	0.026	0.086	0.008
		400	−0.011	0.064	0.004	−0.004	0.065	0.004
		800	−0.004	0.051	0.003	0.005	0.055	0.003
	0.5	100	**−0.059**	0.184	0.037	0.044	0.253	0.066
		200	**−0.064**	0.115	0.017	0.022	0.157	0.025
		400	**−0.066**	0.079	0.011	0.007	0.100	0.010
		800	**−0.064**	0.058	0.007	0.010	0.076	0.006
	2	100	**−0.274**	0.183	0.109	0.119	0.697	0.499
		200	**−0.285**	0.128	0.098	0.028	0.276	0.077
		400	**−0.296**	0.084	0.095	−0.009	0.180	0.032
		800	**−0.295**	0.062	0.091	−0.011	0.121	0.015
Mixture Normal	0.1	100	−0.037	0.152	0.024	0.009	0.181	0.032
		200	−0.038	0.105	0.017	0.007	0.120	0.014
		400	−0.031	0.077	0.007	0.013	0.093	0.009
		800	−0.032	0.055	0.004	0.011	0.066	0.004
	0.5	100	0.031	0.328	0.108	**0.146**	0.328	0.129
		200	0.010	0.228	0.052	**0.123**	0.229	0.068
		400	−0.012	0.145	0.021	**0.105**	0.149	0.033
		800	−0.008	0.102	0.011	**0.111**	0.107	0.024
	2	100	**0.253**	0.644	0.478	**−0.058**	0.408	0.170
		200	**0.174**	0.406	0.195	**−0.092**	0.229	0.061
		400	**0.122**	0.239	0.072	**−0.110**	0.151	0.035
		800	**0.108**	0.165	0.039	**−0.103**	0.110	0.023

Relative bias (Bias), standard deviation (SD), and mean square error (MSE) for baseline hazard shape parameter ρ across 1000 Monte Carlo trials. The data are generated using gamma, truncated gamma, and mixture normal frailty distributions and then gamma and inverse Gaussian frailty models (Weibull baseline hazard) are fitted. The bold typeface indicates cases where the relative bias is greater than 5%.

			Gamma _Weibull			Inverse Gaussian _Weibull		
	θ	G	Bias	SD	MSE	Bias	SD	MSE
Gamma	0.1	100	0.009	0.155	0.024	−0.001	0.083	0.006
		200	0.004	0.106	0.011	0.031	0.171	0.030
		400	0.002	0.076	0.006	0.000	0.086	0.007
		800	0.000	0.054	0.003	−0.001	0.063	0.004
	0.5	100	0.007	0.169	0.029	0.006	0.170	0.029
		200	0.007	0.124	0.016	0.006	0.121	0.015
		400	0.001	0.080	0.006	0.001	0.079	0.006
		800	0.000	0.056	0.003	−0.001	0.057	0.003
	2	100	0.015	0.199	0.040	**−0.096**	0.136	0.056
		200	0.005	0.135	0.018	**−0.097**	0.095	0.047
		400	0.004	0.091	0.008	**−0.100**	0.066	0.044
		800	0.002	0.064	0.004	**−0.101**	0.046	0.043
Truncated Gamma	0.1	100	**0.092**	0.213	0.079	**0.051**	0.055	0.013
		200	**0.085**	0.143	0.049	0.003	0.090	0.008
		400	0.013	0.081	0.007	0.005	0.108	0.011
		800	**0.072**	0.074	0.026	−0.002	0.078	0.006
	0.5	100	**0.085**	0.238	0.085	−0.004	0.113	0.012
		200	**0.072**	0.161	0.046	−0.003	0.111	0.012
		400	**0.070**	0.112	0.033	−0.009	0.090	0.008
		800	**0.067**	0.078	0.024	−0.008	0.063	0.004
	2	100	**0.075**	0.266	0.093	0.021	0.204	0.043
		200	**0.058**	0.178	0.045	0.015	0.133	0.019
		400	**0.053**	0.122	0.026	0.009	0.097	0.010
		800	**0.051**	0.096	0.012	0.009	0.063	0.005
Lognormal	0.1	100	0.022	0.155	0.023	0.013	0.151	0.021
		200	0.032	0.102	0.014	0.023	0.102	0.012
		400	−0.003	0.086	0.007	0.000	0.089	0.008
		800	0.000	0.055	0.003	0.003	0.058	0.003
	0.5	100	−0.011	0.174	0.030	0.014	0.190	0.036
		200	−0.016	0.116	0.014	0.008	0.128	0.017
		400	−0.020	0.079	0.008	0.000	0.086	0.007
		800	−0.019	0.055	0.004	0.002	0.060	0.004
	2	100	**−0.055**	0.172	0.042	−0.010	0.173	0.030
		200	**−0.060**	0.123	0.030	−0.018	0.122	0.016
		400	**−0.067**	0.082	0.025	−0.023	0.086	0.009
		800	**−0.065**	0.058	0.021	−0.022	0.057	0.005
Mixture Normal	0.1	100	−0.010	0.192	0.036	0.007	0.200	0.039
		200	−0.009	0.097	0.010	0.008	0.112	0.013
		400	−0.011	0.081	0.007	0.004	0.089	0.008
		800	−0.013	0.056	0.004	0.002	0.061	0.004
	0.5	100	−0.003	0.201	0.040	**−0.081**	0.137	0.045
		200	−0.007	0.143	0.021	**−0.084**	0.094	0.037
		400	−0.016	0.098	0.011	**−0.089**	0.069	0.036
		800	−0.017	0.067	0.006	**−0.090**	0.047	0.034
	2	100	**0.075**	0.256	0.087	**−0.217**	0.118	0.203
		200	**0.059**	0.183	0.047	**−0.221**	0.082	0.203
		400	**0.051**	0.126	0.026	**−0.223**	0.059	0.202
		800	0.042	0.083	0.014	**−0.225**	0.040	0.204

References

Abiodun, A. (2008), 'Application of shared gamma and inverse-gaussian frailty models to cancer data', *Global Journal of Mathematical Sciences* **7**(1), 35–39. 63, 75, 76

Agrawal, N. K. and Sharma, B. (2013), 'Prevalence of osteoporosis in otherwise healthy Indian males aged 50 years and above', *Archives of Osteoporosis* **8**(1), 1–7. 37

Baddoura, R., Okais, J. and Awada, H. (2001), 'Incidence of fractures after the age of 50 years in the Lebanese population and implications in terms of osteoporosis', *Revue D'epidemiologie et de Sante Publique* **49**(1), 27–32. 37

Bender, R., Augustin, T. and Blettner, M. (2005), 'Generating survival times to simulate cox proportional hazards models', *Statistics in Medicine* **24**(11), 1713–1723. 23, 66

Brilleman, S. L., Crowther, M. J., Moreno-Betancur, M., Buros Novik, J., Dunyak, J., Al-Huniti, N., Fox, R., Hammerbacher, J. and Wolfe, R. (2019), 'Joint longitudinal and time-to-event models for multilevel hierarchical data', *Statistical Methods in Medical Research* **28**(12), 3502–3515. 78

Bycott, P. and Taylor, J. (1998), 'A comparison of smoothing techniques for cd4 data measured with error in a time-dependent cox proportional hazards model', *Statistics in Medicine* **17**(18), 2061–2077. 28

Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. M. (2006), *Measurement Error in Nonlinear Models: a Modern perspective*, Chapman and Hall/CRC. 32

Chen, J.-Y. (2016), Joint modeling of bivariate longitudinal and survival data in spouse pairs, PhD thesis, University of Pittsburgh. 78

Clayton, D. G. (1978), 'A model for association in bivariate life tables and its application

in epidemiological studies of familial tendency in chronic disease incidence', *Biometrika* **65**(1), 141–151. 13

Collett, D. (2015), *Modelling Survival Data in Medical Research*, CRC press. 25

Cologne, J., Hsu, W.-L., Abbott, R. D., Ohishi, W., Grant, E. J., Fujiwara, S. and Cullings, H. M. (2012), 'Proportional hazards regression in epidemiologic follow-up studies: an intuitive consideration of primary time scale', *Epidemiology* pp. 565–573. 3, 42

Congdon, P. (1995), 'Modelling frailty in area mortality', *Statistics in Medicine* **14**(17), 1859–1874. 14

Cox, D. and Oakes, D. (1984), 'Analysis of survival data (Chapman and Hail, New York)'. 6

Cox, D. R. (1972), 'Regression models and life-tables', *Journal of the Royal Statistical Society: Series B (Methodological)* **34**(2), 187–202. 8

Crowley, J. and Hu, M. (1977), 'Covariance analysis of heart transplant survival data', *Journal of the American Statistical Association* **72**(357), 27–36. 25

Crowther, M. J., Abrams, K. R. and Lambert, P. C. (2013), 'Joint modeling of longitudinal and survival data', *The Stata Journal* **13**(1), 165–184. 34

Crowther, M. J., Andersson, T. M.-L., Lambert, P. C., Abrams, K. R. and Humphreys, K. (2016), 'Joint modelling of longitudinal and survival data: incorporating delayed entry and an assessment of model misspecification', *Statistics in Medicine* **35**(7), 1193–1209. 78

Cummings, S. R., Browner, W., Black, D., Nevitt, M., Genant, H., Cauley, J., Ensrud, K., Scott, J. and Vogt, T. (1993), 'Bone density at various sites for prediction of hip fractures', *The Lancet* **341**(8837), 72–75. 38, 60

Dafni, U. G. and Tsiatis, A. A. (1998), 'Evaluating surrogate markers of clinical outcome when measured with error', *Biometrics* pp. 1445–1462. 28, 32, 88

Dall'Olio, F., Vanhooren, V., Chen, C. C., Slagboom, P. E., Wuhrer, M. and Franceschi, C. (2013), 'N-glycomic biomarkers of biological aging and longevity: a link with inflammaging', *Ageing Research Reviews* **12**(2), 685–698. 39

dos Santos, D. M., Davies, R. B. and Francis, B. (1995), 'Nonparametric hazard versus nonparametric frailty distribution in modelling recurrence of breast cancer', *Journal of Statistical Planning and Inference* **47**(1-2), 111–127. 14

Duchateau, L. and Janssen, P. (2008), *The Frailty Model*, Springer. 14, 15, 16, 19

Elmi, A. F., Grantz, K. L. and Albert, P. S. (2018), 'An approximate joint model for multiple paired longitudinal outcomes and time-to-event data', *Biometrics* **74**(3), 1112–1119. 78

Farzi, M., Pozo, J. M., McCloskey, E., Eastell, R., Harvey, N. C., Frangi, A. F. and Wilkinson, J. M. (2022), 'Quantitating age-related bmd textural variation from dxa region-free-analysis: A study of hip fracture prediction in three cohorts', *Journal of Bone and Mineral Research* **37**(9), 1679–1688. 39

Finkelstein, J. S., Brockwell, S. E., Mehta, V., Greendale, G. A., Sowers, M. R., Ettinger, B., Lo, J. C., Johnston, J. M., Cauley, J. A., Danielson, M. E. et al. (2008), 'Bone mineral density changes during the menopause transition in a multiethnic cohort of women', *The Journal of Clinical Endocrinology & Metabolism* **93**(3), 861–868. 46

Fitzmaurice, G. M., Laird, N. M. and Ware, J. H. (2012), *Applied Longitudinal Analysis*, Vol. 998, John Wiley & Sons. 31, 87

Fosgate, G. (2006), 'Non-differential measurement error does not always bias diagnostic likelihood ratios towards the null', *Emerging Themes in Epidemiology* **3**(1), 1–7. 32

Furgal, A. K., Sen, A. and Taylor, J. M. (2019), 'Review and comparison of computational approaches for joint longitudinal and time-to-event models', *International Statistical Review* **87**(2), 393–418. 35

Gasparini, A., Clements, M. S., Abrams, K. R. and Crowther, M. J. (2019), 'Impact of model misspecification in shared frailty survival models', *Statistics in Medicine* **38**(23), 4477–4502. 62, 63, 66, 67, 75

Glidden, D. V. and Vittinghoff, E. (2004), 'Modelling clustered survival data from multicentre clinical trials', *Statistics in Medicine* **23**(3), 369–388. 63

Gorfine, M., Zucker, D. M. and Hsu, L. (2006), 'Prospective survival analysis with a general semiparametric shared frailty model: A pseudo full likelihood approach', *Biometrika* **93**(3), 735–741. 51, 52

Gottard, A., Mattei, A. and Vignoli, D. (2012), How education affects fertility in the presence of time-varying frailty component, Technical report, Tech. rep., Dipartimento di Statistica" Giuseppe Parenti", Universitá degli 112

Grambsch, P. M. and Therneau, T. M. (1994), 'Proportional hazards tests and diagnostics based on weighted residuals', *Biometrika* **81**(3), 515–526. 1, 56

Group, E. P. O. S. E. and O'Neill, T. (2002), 'The relationship between bone density and incident vertebral fracture in men and women', *Journal of Bone and Mineral Research* **17**(12), 2214–2221. 38, 60

Gullberg, B., Johnell, O. and Kanis, J. (1997), 'World-wide projections for hip fracture', *Osteoporosis International* **7**(5), 407–413. 36

Harville, D. A. (1977), 'Maximum likelihood approaches to variance component estimation and to related problems', *Journal of the American statistical association* **72**(358), 320–338. 31

Henderson, R., Diggle, P. and Dobson, A. (2000), 'Joint modelling of longitudinal measurements and event time data', *Biostatistics* **1**(4), 465–480. 34

Henderson, R. and Oman, P. (1999), 'Effect of frailty on marginal regression estimates in survival analysis', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **61**(2), 367–379. 63

Hernán, M. A. (2010), 'The hazards of hazard ratios', *Epidemiology (Cambridge, Mass.)* **21**(1), 13. 44

Hougaard, P. (1984), 'Life table methods for heterogeneous populations: distributions describing the heterogeneity', *Biometrika* **71**(1), 75–83. 18

Hougaard, P. (2000), *Analysis of Multivariate Survival Data*, Vol. 564, Springer. 13, 14

Hougaard, P., Harvald, B. and Holm, N. V. (1992), 'Measuring the similarities between the lifetimes of adult danish twins born between 1881–1930', *Journal of the American Statistical Association* **87**(417), 17–24. 13

Hsu, L., Gorfine, M. and Malone, K. (2007), 'On robustness of marginal regression coefficient estimates and hazard functions in multivariate survival analysis of family data when the frailty distribution is mis-specified', *Statistics in Medicine* **26**(25), 4657–4678. 63

Jensen, H., Brookmeyer, R., Aaby, P. and Andersen, P. K. (2004), *Shared Frailty Model for Left-truncated Multivariate survival data*, University of Copenhagen. Department of Biostatistics. 3, 80

Jiang, X., Liu, W. and Zhang, B. (2021), 'A note on the prediction of frailties with misspecified shared frailty models', *Journal of Statistical Computation and Simulation* **91**(2), 219–241. 64

Kalbfleisch, J. D. and Prentice, R. L. (2011), *The Statistical Analysis of Failure Time Data*, John Wiley & Sons. 6, 25

Klein, J. P. and Moeschberger, M. L. (2003), *Survival Analysis: Techniques for Censored and Truncated Data*, Vol. 2, Springer. 6, 27, 83

Kovesdy, C. P., Anderson, J. E. and Kalantar-Zadeh, K. (2007), 'Paradoxical association between body mass index and mortality in men with CKD not yet on dialysis', *American Journal of Kidney Diseases* **49**(5), 581–591. 25

Krištić, J., Vučković, F., Menni, C., Klarić, L., Keser, T., Beceheli, I., Pučić-Baković, M., Novokmet, M., Mangino, M., Thaqi, K. et al. (2014), 'Glycans are a novel biomarker of chronological and biological ages', *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences* **69**(7), 779–789. 39, 40

Kvist, K., Andersen, P. K., Angst, J. and Kessing, L. V. (2010), 'Event dependent sampling of recurrent events', *Lifetime Data Analysis* **16**(4), 580–598. 3

Laird, N. M. and Ware, J. H. (1982), 'Random-effects models for longitudinal data', *Biometrics* pp. 963–974. 31, 87

Lee, J.-K. and Khir, A. S. (2007), 'The incidence of hip fracture in malaysians above 50 years of age: variation in different ethnic groups', *APLAR Journal of Rheumatology* **10**(4), 300–305. 37

Marques, A., Lourenço, Ó. and Da Silva, J. (2015), 'The burden of osteoporotic hip fractures in portugal: costs, health related quality of life and mortality', *Osteoporosis International* **26**(11), 2623–2630. 36

Marshall, D., Johnell, O. and Wedel, H. (1996), 'Meta-analysis of how well measures of bone mineral density predict occurrence of osteoporotic fractures', *BMJ* **312**(7041), 1254–1259. 38, 39, 60

McLaughlin, M. A., Orosz, G. M., Magaziner, J., Hannan, E. L., McGinn, T., Morrison, R. S., Hochman, T., Koval, K., Gilbert, M. and Siu, A. L. (2006), 'Preoperative status and risk of complications in patients with hip fracture', *Journal of General Internal Medicine* **21**(3), 219–225. 36

Menni, C., Gudelj, I., Macdonald-Dunlop, E., Mangino, M., Zierer, J., Bešić, E., Joshi, P. K., Trbojević-Akmačić, I., Chowienczyk, P. J., Spector, T. D. et al. (2018), 'Glycosylation profile of immunoglobulin G is cross-sectionally associated with cardiovascular disease risk score and subclinical atherosclerosis in two independent cohorts', *Circulation Research* **122**(11), 1555–1564. 39

Miller, P. D., Siris, E. S., Barrett-Connor, E., Faulkner, K. G., Wehren, L. E., Abbott, T. A., Chen, Y.-T., Berger, M. L., Santora, A. C. and Sherwood, L. M. (2002), 'Prediction of fracture risk in postmenopausal white women with peripheral bone densitometry: evidence from the national osteoporosis risk assessment', *Journal of Bone and Mineral Research* **17**(12), 2222–2230. 38, 39, 60

Muli, A. (2022).
URL: *https://github.com/Annah92/* 37

Muli, A., Houwing-Duistermaat, J. and Gusnanto, A. (2021), 'Use of shared gamma frailty model in analysis of survival data in twins', *Theoretical Biology Forum* pp. 45–58. 111

Munda, M. and Legrand, C. (2014), 'Adjusting for centre heterogeneity in multicentre clinical trials with a time–to–event outcome', *Pharmaceutical Statistics* **13**(2), 145–152. 63, 75, 76

Munda, M., Rotolo, F. and Legrand, C. (2012), 'parfm: Parametric frailty models in R', *Journal of Statistical Software* **51**, 1–20. 18, 67

Murphy, S. A. (1994), 'Consistency in a proportional hazards model incorporating a random effect', *The Annals of Statistics* **22**(2), 712–731. 2

Murphy, S. A. (1995), 'Asymptotic theory for the frailty model', *The Annals of Statistics* pp. 182–198. 2

Ng'andu, N. H. (1997), 'An empirical comparison of statistical tests for assessing the proportional hazards assumption of cox's model', *Statistics in Medicine* **16**(6), 611–626. 1

Okasha, M. K. and Alqanoo, I. (2014), 'Inference on the doubly truncated gamma distribution for lifetime data', *Int. J. Math. Stat. Invent* **2**, 1–17. 20

Omori, Y. and Johnson, R. A. (1993), 'The influence of random effects on the unconditional hazard rate and survival functions', *Biometrika* **80**(4), 910–914. 62

Patterson, H. D. and Thompson, R. (1971), 'Recovery of inter-block information when block sizes are unequal', *Biometrika* **58**(3), 545–554. 31

Peña, J. M. (1997), 'B-splines and optimal stability', *Mathematics of Computation* pp. 1555–1560. 50

Perperoglou, A., Sauerbrei, W., Abrahamowicz, M. and Schmid, M. (2019), 'A review of spline function procedures in r', *BMC medical Research Methodology* **19**(1), 1–16. 50

Prentice, R. L. (1982), 'Covariate measurement errors and parameter estimation in a failure time regression model', *Biometrika* **69**(2), 331–342. 27, 32, 93

Raboud, J., Reid, N., Coates, R. and Farewell, V. (1993), 'Estimating risks of progressing to AIDS when covariates are measured with error', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **156**(3), 393–406. 27

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), 892–899. 78

Rizopoulos, D. (2008), Joint modelling of longitudinal and survival data, PhD thesis, Catholic University of Leuven. 34

Rodríguez-Girondo, M., Deelen, J., Slagboom, E. P. and Houwing-Duistermaat, J. J. (2018), 'Survival analysis with delayed entry in selected families with application to

human longevity', *Statistical Methods in Medical Research* **27**(3), 933–954. 3, 15, 43, 67, 80, 81

Rondeau, V., Commenges, D. and Joly, P. (2003), 'Maximum penalized likelihood estimation in a gamma-frailty model', *Lifetime Data Analysis* **9**(2), 139–153. 3

Schluchter, M. D. and Piccorelli, A. V. (2019), 'Shared parameter models for joint analysis of longitudinal and survival data with left truncation due to delayed entry–applications to cystic fibrosis', *Statistical methods in medical research* **28**(5), 1489–1507. 78

Self, S. and Pawitan, Y. (1992), Modeling a marker of disease progression and onset of disease, *in* 'AIDS epidemiology', Springer, pp. 231–255. 28

Siris, E., Brenneman, S., Barrett-Connor, E., Miller, P., Sajjan, S., Berger, M. and Chen, Y.-T. (2006), 'The effect of age and bone mineral density on the absolute, excess, and relative risk of fracture in postmenopausal women aged 50–99: results from the national osteoporosis risk assessment (nora)', *Osteoporosis Iinternational* **17**(4), 565–574. 37

Svedbom, A., Ivergård, M., Hernlund, E., Rizzoli, R. and Kanis, J. A. (2014), 'Epidemiology and economic burden of osteoporosis in switzerland', *Archives of Osteoporosis* **9**(1), 1–8. 37

Sweeting, M. J. and Thompson, S. G. (2011), 'Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture', *Biometrical Journal* **53**(5), 750–763. 28, 34

Therneau, T., Crowson, C. and Atkinson, E. (2017), 'Using time dependent covariates and time dependent coefficients in the cox model', *Survival Vignettes* **2**, 3. 26

Tsiatis, A. A. and Davidian, M. (2001), 'A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error', *Biometrika* **88**(2), 447–458. 34

Tsiatis, A. A. and Davidian, M. (2004), 'Joint modeling of longitudinal and time-to-event data: an overview', *Statistica Sinica* pp. 809–834. 34

Tsiatis, A. A., Degruttola, V. and Wulfsohn, M. S. (1995), '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* **90**(429), 27–37. 25, 28, 33, 89

Van den Berg, G. J. and Drepper, B. (2016), 'Inference for shared-frailty survival models with left-truncated data', *Econometric Reviews* **35**(6), 1075–1098. 3, 80

Vanhooren, V., Desmyter, L., Liu, X.-E., Cardelli, M., Franceschi, C., Federico, A., Libert, C., Laroy, W., Dewaele, S., Contreras, R. et al. (2007), 'N-glycomic changes in serum proteins during human aging', *Rejuvenation Research* **10**(4), 521–531a. 39

Vanhooren, V., Dewaele, S., Libert, C., Engelborghs, S., De Deyn, P. P., Toussaint, O., Debacq-Chainiaux, F., Poulain, M., Glupczynski, Y., Franceschi, C. et al. (2010), 'Serum n-glycan profile shift during human ageing', *Experimental Gerontology* **45**(10), 738–743. 39

Vaupel, J. W., Manton, K. G. and Stallard, E. (1979), 'The impact of heterogeneity in individual frailty on the dynamics of mortality', *Demography* **16**(3), 439–454. 1, 12, 14

Wiklund, R., Toots, A., Conradsson, M., Olofsson, B., Holmberg, H., Rosendahl, E., Gustafson, Y. and Littbrand, H. (2016), 'Risk factors for hip fracture in very old people: a population-based study', *Osteoporosis International* **27**(3), 923–931. 36

Wulfsohn, M. S. and Tsiatis, A. A. (1997), 'A joint model for survival and longitudinal data measured with error', *Biometrics* pp. 330–339. 34

Ye, W., Lin, X. and Taylor, J. M. (2008), 'Semiparametric modeling of longitudinal measurements and time-to-event data–a two-stage regression calibration approach', *Biometrics* **64**(4), 1238–1246. 28, 33

Yu, T., Wu, L. and Gilbert, P. B. (2018), 'A joint model for mixed and truncated longitudinal data and survival data, with application to hiv vaccine studies', *Biostatistics* **19**(3), 374–390. 32, 88

Zucker, D. M., Gorfine, M. and Hsu, L. (2008), 'Pseudo-full likelihood estimation for prospective survival analysis with a general semiparametric shared frailty model: Asymptotic theory', *Journal of Statistical Planning and Inference* **138**(7), 1998–2016. 51