

# Probation review report

*Alessandro Gasparini*

*2017-08-21*



# Contents

<b>Introduction</b>	<b>7</b>
<b>1 Terminology and notation</b>	<b>9</b>
<b>2 Survival models with random effects</b>	<b>13</b>
2.1 Univariate frailty models . . . . .	15
2.2 Shared frailty models . . . . .	17
2.3 Alternative formulation . . . . .	18
<b>3 Joint models for longitudinal and survival data</b>	<b>21</b>
3.1 Model formulation . . . . .	22
3.2 Estimation process . . . . .	24
<b>4 Computational challenges in survival models with random effects</b>	<b>27</b>
4.1 Numerical integration . . . . .	28
4.2 Numerical differentiation . . . . .	33
4.3 Numerical root finding . . . . .	34
<b>5 Simulating survival data</b>	<b>37</b>
<b>6 Simulation study: accuracy of Gaussian quadrature</b>	<b>41</b>
6.1 Aim . . . . .	41

6.2	Data-generating mechanisms . . . . .	42
6.3	Methods . . . . .	43
6.4	Estimands . . . . .	44
6.5	Performance measures . . . . .	44
6.6	Results . . . . .	45
6.7	Conclusions . . . . .	47
<b>7</b>	<b>Simulation study: impact of misspecification in survival models with shared frailty terms</b>	<b>49</b>
7.1	Aim . . . . .	50
7.2	Data-generating mechanisms . . . . .	51
7.3	Methods . . . . .	52
7.4	Estimands . . . . .	55
7.5	Performance measures . . . . .	56
7.6	Results . . . . .	56
7.7	Conclusions . . . . .	59
<b>8</b>	<b>Exploring results from simulation studies interactively</b>	<b>61</b>
<b>9</b>	<b>Informative visiting process</b>	<b>65</b>
<b>10</b>	<b>Future research developments</b>	<b>69</b>
<b>11</b>	<b>Personal development</b>	<b>73</b>
11.1	Supervisory meetings . . . . .	73
11.2	Training and courses . . . . .	73
11.3	Conferences . . . . .	76
<b>Appendix</b>		<b>76</b>
<b>A</b>	<b>Tables</b>	<b>77</b>

<i>CONTENTS</i>	5
<b>B Plots</b>	<b>93</b>
<b>C Slides</b>	<b>111</b>
C.1 2017 SAfJR Conference . . . . .	111
C.2 2017 SAM Conference and ISCB Conference . . . . .	119
C.3 Students' Day at the 2017 ISCB Conference . . . . .	126
<b>D Manuscript draft</b>	<b>135</b>
<b>Bibliography</b>	<b>147</b>



# Introduction

This report presents the work I have done during my first year as a PhD student at the Department of Health Sciences, University of Leicester, under the supervision of Dr. Michael Crowther and Prof. Keith Abrams.

I will begin by briefly introducing some terminology and notation in Chapter 1. Second, I will introduce survival models with random effects (e.g. frailties, in the simplest form) and joint models for longitudinal and time-to-event data in Chapters 2 and 3, respectively. Computational challenges that survival models with random effects and joint models pose are presented in Chapter 4. In Chapter 5, I will present a well established method for simulating survival data. I will then present the results of two simulation studies in Chapters 6 and 7; the first simulation study investigates the accuracy of quadrature methods when approximating analytically intractable terms, while the second simulation study investigates the impact of model misspecification in survival models with shared frailty terms. I will introduce an interactive tool I have been developing to aid the dissemination of results from simulation studies and motivated by the simulation studies of Chapter 6 and 7 in Chapter 8. Next, I will introduce the problem of informative visiting process in clinical research using healthcare consumption data in Chapter 9, and how I aim to evaluate and compare the different approaches that have been proposed and utilised in literature to tackle such problem in Chapter 10. Chapter 10 will also include further research goals for the second and third year of my PhD. Finally, I will summarise the training and personal development

activities I attended to during my first year in Chapter 11.

The text of this report is licensed under the [Creative Commons Attribution-ShareAlike 4.0 International License](#), while the underlying code is licensed under the [GPLv3](#). The report is written using R and the [bookdown](#) package ([Xie, 2016](#)), and can be accessed online at <https://ellessenne.github.io/prr/>.

# Chapter 1

## Terminology and notation

I will introduce in this Section some notation that I will be using throughout this report.

I denote the random variable for an individual's survival time with  $S$ ; since it denotes time,  $S$  can assume any non-negative value. The lower-case  $s$  represent a specific value of interest drawn from  $S$  for a given individual. In the case of right censoring, I denote with  $C$  the random variable representing censoring time, and  $c$  its realisation. The observed time is denoted with  $T = \min(S, C)$ , and its realisation is  $t$ . Finally, I denote with  $D = I(S \leq C)$  the random variable indicating either occurrence of the event of interest or censorship; analogously as before, its realisation is lower-case  $d$ .

Next, I will define two of the main quantities of interest in survival analysis, the *survival function* and the *hazard function*. They are both functions of the observed time  $t$  and are denoted by  $S(t)$  and  $h(t)$ , respectively.

The survival function is the complement of the cumulative distribution function of the observed time  $T$  and represent the probability that a given individual survives<sup>1</sup> longer than a

---

<sup>1</sup>I use the term *survives* loosely speaking, for conciseness - formally, I refer to *not experiencing the event of interest*.

specified time  $t$ :

$$S(t) = 1 - F_T(t) = 1 - P(T \leq t) = P(T > t)$$

$t$  ranges (theoretically) between 0 and infinity, hence the survival function can be plotted as a smooth, continuous function that tends to 0 as  $t$  goes to infinity. In practice, though, the survival function appears as a step function as (1) individuals can be observed at discrete times only and (2) not all individuals may experience the event before the end of the study.

The hazard function  $h(t)$  is the limit of the probability of the survival time  $T$  laying within an interval  $[t, t + \Delta(t))$  given that an individual survived up to time  $t$  divided by the length of the interval  $\Delta(t)$ , for  $\Delta(t)$  approaching zero:

$$h(t) = \lim_{\Delta(t) \rightarrow 0} \frac{P(t \leq T < t + \Delta(t) | T \geq t)}{\Delta(t)}$$

It represent the instantaneous potential (e.g. risk) for the event to occur within the interval  $[t, t + \Delta(t))$  (with  $\Delta(t) \rightarrow 0$ ), given that the individual survived up to time  $t$ . The hazard function is always non-negative, it can assume different shapes over time, and it has no upper bound.

The survival function and the hazard function are strictly related. In fact, there is a clearly defined mathematical relationship between them, and it is possible to derive the form of  $S(t)$  when knowing the form of  $h(t)$  (and vice versa). Formally:

$$S(t) = \exp \left[ - \int_0^t h(u) \, du \right]$$

$$h(t) = - \left[ \frac{dS(t)/dt}{S(t)} \right]$$

Finally, a third quantity of interest in survival analysis that is strictly related to the survival and hazard functions is the cumulative hazard function  $H(t)$ . The cumulative hazard

function represents the accumulation of hazard (e.g.  $h(t)$ ) over time, and can be defined as

$$H(t) = \int_0^t h(u) \, du;$$

it can conveniently be expressed in terms of survival function via the relationship  $H(t) = -\log S(t)$ , or alternatively with  $S(t) = \exp(-H(t))$ .



# Chapter 2

## Survival models with random effects

Random effects models are a kind of hierarchical model in which the data is assumed to have some sort of hierarchical structure: for instance, individual patients data clustered into families, cities, regions, and so on. It is also assumed that individuals are homogeneous within hierarchical unit, heterogeneous between different units. By comparison, fixed effects models do not take into account any hierarchy in the data. In biostatistics the terms *fixed effects* and *random effects* have a special meaning: they refer to the *population-average* and *subject-specific* effects, respectively, with the latter generally assumed to be unknown, unobserved variables.

Random effects models are generally used to analyse hierarchical data with a continuous, normally distributed outcome; such models are referred to as *linear mixed-effects models*, as they can incorporate both fixed and random effects, and generalise the linear regression model. When data consists in multiple observations for a given individual over time (and therefore the first level of clustering consists in the individual himself) the term *longitudinal data* is used. It is possible to encounter hierarchical data originating from a variety of distribution from the exponential family such as the Poisson, Gamma, and Binomial distribution. Linear mixed-effects models can be generalised to include such data, and these models

are generally referred to as *generalised linear mixed-effects models*. Practically speaking, is the same process of generalising linear models to generalised linear models. It is also possible to relax the normality assumption for continuous, hierarchical data and model the median (or any quantile, really) rather than the mean. Such models are called *linear quantile mixed-effects models*, and generalise the linear mixed-effects models as quantile regression is generalising the linear model. Survival data can present a hierarchical structure too; for instance, data could be clustered in geographical areas, institutions, or patients themselves. Meta-analysis of individual-patient data are a common example of survival data (when the outcome is time to event) with some hierarchical structure; another example is given by repeated-events data, such as infections or acute recurrent events, in which the first level of the hierarchical structure consists in the patient. A final example of survival data with biological cluster is given by twin data, in which siblings share some genetic factors. This heterogeneity structure often leads to violation of the implicit assumption that populations are homogeneous: sometimes it is impossible to include all relevant risk factors, or maybe such risk factors are not known at all. The result is unobserved heterogeneity. The simplest survival model with random effects is the *univariate frailty model*, in which a random effect - named frailty - is included in the model to account for the unobserved heterogeneity. The univariate frailty model can be generalised by allowing the frailty term to be shared between observations belonging to the same cluster of data. The resulting models are named *shared frailty model*. The frailty term generally acts multiplicatively on the baseline hazard, and it is modelled on the hazard scale; it is possible to alternatively formulate the model in terms of random effects rather than frailties, by including the frailty as an additive term on the log-hazard scale.

I will introduce the univariate frailty model in Section 2.1, and generalise it to allow shared frailty terms in Section 2.2. Finally, I will present the alternative formulation in terms of random effects in Section 2.3. A comprehensive treatment of frailty models in survival analysis is given in Hougaard (2000) and Wienke (2010).

## 2.1 Univariate frailty models

In those settings where risk factors are not measured, their relevance is unknown, or it is not known whether such risk factor exist at all or not, it is useful to consider two sources of variability in survival analysis: variability accounted for by observable risk factors included in the model and heterogeneity caused by unknown covariates. The unobserved heterogeneity is described by the frailty term, which is assumed to follow some distribution. Formally:

$$h(t|\alpha) = \alpha h_0(t),$$

where  $\alpha$  is a non-observed frailty effect and  $h_0(t)$  is the baseline hazard function. The random variable  $\alpha$ , the frailty term, is chosen to have a distribution  $f(\alpha)$  with expectation  $E(\alpha) = 1$  and variance  $V(\alpha) = \sigma^2$ .  $V(\alpha)$  is interpretable as a measure of heterogeneity across the population in baseline risk: as  $\sigma^2$  increases the values of  $\alpha$  are more dispersed, with greater heterogeneity in  $\alpha h_0(t)$ . Underlying assumptions are: the frailty is time independent, and it acts multiplicatively on the underlying baseline hazard function.

Introducing observed covariates into the model:

$$h(t|X, \alpha) = \alpha h_0(t) \exp(X\beta) = \alpha h(t|X),$$

with  $X$  and  $\beta$  covariates and regression coefficients, respectively. Given the relationship between hazard and survival function, it can be showed that the individual survival function conditional on the frailty is  $S(t|\alpha) = S(t)^{\alpha}$ . The population (i.e. marginal, or unconditional) survival function is obtained by integrating out the frailty from the conditional survival function:

$$S(t) = \int_0^{+\infty} [S(t)]^\alpha f(\alpha) d\alpha$$

The individual contribution to the likelihood (assuming no delayed entry) is conditional on

the unobserved frailty  $\alpha$

$$L_i = \prod_{i=1}^n (\alpha h_0(t_i) \exp(X_i \beta))^{d_i} \exp(-\alpha H_0(t_i) \exp(X_i \beta)),$$

with  $d_i$  event indicator variable,  $H_0(t_i)$  cumulative baseline hazard, and  $t_i$  observed survival time - all relative to the  $i$ -th individual.

Different choices for the frailty distribution are possible. Assigning a probability distribution implies that the frailty can be integrated out of the likelihood function. After this integration, the likelihood can be maximized in the usual way if its explicit form exists. Otherwise, more sophisticated approaches like numerical integration or Markov Chain Monte Carlo methods are required. The most often used frailty distributions are the gamma and the log-normal distribution; the positive stable and the inverse Gaussian distribution are also common.

Assuming that the frailty  $\alpha$  has a Gamma distribution is convenient: it has the appropriate range  $(0, \infty)$  and it is mathematically tractable. A Gamma distribution with parameters  $a$  and  $b$  has density

$$f(x) = \frac{x^{a-1} \exp(-x/b)}{\Gamma(a)b^a};$$

by choosing  $a = 1/\theta$  and  $b = \theta$  the resulting distribution has expectation 1 and finite variance  $\theta$ . In these settings, the model is analytically tractable: the population survival function has the form

$$S(t) = (1 - \theta \log(S(t)))^{-1/\theta};$$

the likelihood follows by substitution. Estimating such model becomes therefore straightforward, which likely contributed to the popularity of Gamma frailty models.

Together with the Gamma distribution, the log-normal distribution is the most commonly used frailty distribution, given its strong ties to random effect models; more on that in Section 2.3. Hence, assuming a log-normal distribution with a single parameter  $\theta > 0$  (for

comparison with the mathematically tractable Gamma frailty model) with density

$$f(x) = (2\pi\theta)^{-\frac{1}{2}}x^{-1} \exp\left(-\frac{(\log x)^2}{2\theta}\right),$$

the resulting model has a frailty whose expectation is finite. Nevertheless, this frailty distribution cannot be integrated out of the survival function analytically to obtain the population survival function or the likelihood.

## 2.2 Shared frailty models

Further generalising the model presented in Section 2.1, it is possible for the frailty effect  $\alpha$  to be shared between clusters of study subjects. Specifically, for the  $j$ -th observation in the  $i$ -th cluster:

$$h_{ij}(t|\alpha_i) = \alpha_i h(t|X_{ij}).$$

The conditional survival function is:

$$S_{ij}(t|\alpha_i) = S_{ij}(t)^{\alpha_i}.$$

In this setting, the cluster-specific contribution to the likelihood is obtained by calculating the cluster-specific likelihood conditional on the frailty, consequently integrating out the frailty itself:

$$L_i = \int_A L_i(\alpha_i) f(\alpha_i) d\alpha,$$

with  $f(\alpha)$  the distribution of the frailty,  $A$  its domain, and  $L_i(\alpha_i)$  the cluster-specific contribution to the likelihood, conditional on the frailty. The cluster-specific contribution to the likelihood is

$$L_i(\alpha_i) = \alpha_i^{D_i} \prod_{j=1}^{n_i} S_{ij}(t_{ij})^{\alpha_i} h_{ij}(t_{ij})^{d_{ij}},$$

with  $D_i = \sum_{j=1}^{n_i} d_{ij}$ . Analogously as before, analytical formulae can be obtained when  $\alpha_i$  follows a Gamma distribution:

$$L_i = \left[ \prod_{j=1}^{n_i} h_{ij}(t_{ij})^{d_{ij}} \right] \frac{\Gamma(1/\theta + D_i)}{\Gamma(1/\theta)} \left[ 1 - \theta \sum_{j=1}^{n_i} \log S_{ij}(t_{ij}) \right]^{-1/\theta - D_i};$$

further details in Gutierrez (2002). As in the univariate frailty model, assuming a log-normal distribution requires some numerical approximation to be performed, being the resulting model analytically intractable.

## 2.3 Alternative formulation

I mentioned in Section 2.1 that the log-normal distribution for the frailty term has strong ties to random-effects models. Recall the formulation for a log-normal shared frailty model:

$$h_{ij}(t|\alpha_i) = \alpha_i h(t|X_{ij}) = \alpha_i h_0(t) \exp(X_{ij}\beta),$$

with  $\alpha_i$  following a log-normal distribution. It is possible to formulate the same model on the log-hazard scale as

$$h_{ij}(t|\alpha_i) = h_0(t) \exp(X_{ij}\beta + \eta_i),$$

with  $\eta_i = \log \alpha_i$ .  $\eta_i$  results being normally distributed with parameters  $\mu$  and  $\sigma^2$  related to those of the log-normal distribution by the relationship

$$E(\alpha_i) = \exp(\mu + \sigma^2/2)$$

and

$$Var(\alpha_i) = \exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1)$$

By formulating the model on the log-hazard scale, the frailty term has a direct interpretation

as a random intercept in the model; that is, the heterogeneity is modelled by allowing the model intercept to vary between clusters. Consequently, it is possible to further extend this model by allowing random covariates effects, potentially ranging over multiple levels of clustering. Using the usual mixed-effects models notation:

$$h_{ij}(t|b_i) = h_0(t) \exp(X_{ij}\beta + Z_i b_i),$$

with  $X_{ij}$  representing the design matrix for the fixed effects  $\beta$  and  $Z_i$  representing the design matrix for the random effects  $b_i$ . Any distribution or functional form can be assumed for  $h_0(t)$  (Crowther et al., 2014), or it is possible to leave it unspecified altogether yielding a semiparametric Cox model with random effects (Ripatti and Palmgren, 2000; Therneau et al., 2003).



# Chapter 3

## Joint models for longitudinal and survival data

It is increasingly common for observational studies and trials to follow participants over time, recording abundant data on clinical features throughout the duration of the study. Moreover, routinely collected healthcare consumption data and population registries are being used more and more for research purpose, after being linked with other data sources. As a consequence, applied researchers often encounter longitudinally recorded covariates to account for when studying the clinical outcome of interest (e.g. time to event, that is what I will focus on). Researchers then face two options: (1) select only one of the multiple values per individual and analyse as such, ignoring part of the available data, or (2) take into account the potential dependency and association between the repeatedly measured covariates and the outcome interest. The latter is usually the most sensible choice, as the longitudinal data can contain important predictors or surrogates of the time to event outcome. A powerful tool to jointly model longitudinal and time-to-event data is given by joint models for longitudinal and time to event data, in which the longitudinal and survival processes are modelled jointly into a single model allowing to infer their association. The development of such models was

motivated by HIV/AIDS clinical trials, in which immune response was recorded over the duration of the trial and the association with survival was of interest. Seminal works on the topic are the papers by [Wulfsohn and Tsiatis \(1997\)](#), [Tsiatis and Davidian \(2004\)](#), [Henderson et al. \(2000\)](#), [Pawitan and Self \(1993\)](#); a more recent tractation of the topic is in [Ibrahim et al. \(2010\)](#), [Rizopoulos \(2012\)](#), [Gould et al. \(2015\)](#).

Previous attempts to tackle this problem consisted in (1) fitting a time-dependent Cox model ([Cox, 1972](#)) by splitting individual rows every time a new observation from the longitudinal covariate becomes available, and (2) by using two-stages methods in which the longitudinal and survival data were modelled separately ([Tsiatis et al., 1995](#)). Nevertheless, it has been showed that joint modeling increases efficiency and reduces bias ([Hogan and Laird, 1998](#)), while improving predictions at the same time ([Rizopoulos et al., 2014](#)).

In this Chapter I will focus on the basic joint model for longitudinal and survival data, with a single longitudinal process. I will present its formulation in [Section 3.1](#), and the estimation process in [Section 3.2](#). However, several extensions of the basic joint model presented in this Chapter have been proposed during the years, as the topic has received considerable attention. A review on the state of the art in joint models with a single longitudinal process is given by [Gould et al. \(2015\)](#). Further, the joint model has been extended to allow incorporating multiple longitudinal processes at one, measured intermittently and not necessarily at the same time or with the same association structure with the survival component; a recent review on the topic is given by [Hickey et al. \(2016\)](#).

## 3.1 Model formulation

A joint model for longitudinal and survival data consists of two components: a model for the longitudinal part (I will be assuming a single longitudinal trajectory from now on for simplicity) and a model for the survival part. These two components will then share a set of

parameters that will describe the association between the two processes. In literature, the dominant approach seems to be allowing the two components to share random effects; I will follow this approach.

Building on the notation from Section 1, let  $y_{ij} = \{y_{ij}(t_{ij}) \forall j = 1, \dots, n_i\}$  be the observed longitudinal response for the  $i^{\text{th}}$  subject, with  $y_{ij}(t_{ij})$  the observed response at time  $t_{ij}$  and  $n_i$  the number of longitudinal observations.

The longitudinal component of the joint model is modelled within the mixed-effects framework (Diggle et al., 2013), as longitudinal data is likely measured intermittently and with error. Therefore:

$$y_i(t) = m_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \sim N(0, \sigma^2)$$

and

$$m_i(t) = X_i(t)\beta + Z_i(t)b, \quad b \sim N(0, \Sigma)$$

with  $X_i(t)$  and  $Z_i(t)$  the time-dependent design matrices for the fixed and random effects, respectively.  $y_i(t)$  represents the observed longitudinal trajectory at time  $t$ , which could be decomposed into the true longitudinal trajectory  $m_i(t)$  plus the measurement error  $\epsilon_i(t)$ .

The survival component of the joint model is modelled using a proportional hazards time to event model, given the true unobserved longitudinal trajectory up to time  $t$ , i.e.  $M_i(t) = \{m_i(s) \forall 0 \leq s \leq t\}$ :

$$h(t|M_i(t)) = h_0(t) \exp(W\psi + \alpha m_i(t)),$$

where  $h_0(t)$  is the baseline hazard function and  $W$  is a vector of time-fixed covariates with their regression parameters  $\psi$ .  $\alpha$  is the association parameter that links the longitudinal component and the survival component of the joint model; it can be interpreted as the change in log-hazard ratio for a unit increase in the true longitudinal trajectory  $m_i(t)$ , at time  $t$ . This specific form of the association parameter is also known as the *current value* parametrisation; additional association structures are available, allowing for instance interactions, association

with the slope of the trajectory or its cumulative effect, and so on. Further details in Rizopoulos (2012).

The survival function follows as

$$S(t|M_i(t)) = \exp\left(-\int_0^t h_0(u) \exp(W\psi + \alpha m_i(u)) du\right)$$

Finally, regarding  $h_0(t)$ : the choice of the baseline hazard function follows the usual rationale. It can be left unspecified, therefore resulting in a Cox model for the survival component of the joint model, or it can be specified using a parametric distribution (e.g. a Weibull distribution) or some flexible alternative (Crowther et al., 2012). Nevertheless, Hsieh et al. (2006) showed that choosing the Cox model for the survival component yields standard errors that are underestimated; consequently, bootstrap is required to obtain correct standard errors in that situation.

## 3.2 Estimation process

Estimation of a joint model for longitudinal and survival data is a non-trivial task. The complexity of jointly modelling the longitudinal component and the survival component motivated using a two-stages procedure as mentioned in Section 3. With that approach, the longitudinal component is modelled and estimated separately; consequently, subject-specific predictions from the longitudinal model are produced and plugged into the survival model as time-varying covariates. Despite the simplicity of this approach, though, it has been showed that it produces substantial bias and poor coverage (Tsiatis and Davidian, 2001; Sweeting and Thompson, 2011). Therefore, an approach that models both processes jointly is required. in particular, two approaches are predominant: a full likelihood approach, and a Bayesian approach; both have appealing characteristics, but they share the feature of being computationally intensive.

Focusing on the full likelihood approach, it is possible to formulate the joint likelihood (Rizopoulos, 2012) for the overall parameter vector  $\theta = \{\theta_t, \theta_y, \theta_b\}$ , formed by the parameters of the survival component, the parameters of the longitudinal component, and the elements of the variance-covariance matrix of the random effects, respectively. The joint distribution of the survival time  $T_i$ , the event indicator  $d_i$ , and the longitudinal response  $y_i$ , conditional on the random effects  $b_i$ , can be expressed as:

$$P(T_i, d_i, y_i | b_i, \theta) = P(T_i, d_i | b_i, \theta) P(y_i | b_i, \theta),$$

with

$$P(y_i | b_i, \theta) = \prod_{j=1}^{n_i} P(y_i(t_{ij}) | b_i, \theta).$$

It follows that the contribution to the log-likelihood for the  $i^{\text{th}}$  patient is

$$\begin{aligned} \log L(\theta) &= \log \int_{-\infty}^{+\infty} P(T_i, d_i, y_i, b_i; \theta) db_i \\ &= \log \int_{-\infty}^{+\infty} P(T_i, d_i | b_i, \theta_t) \left[ \prod_{j=1}^{n_i} P(y_i(t_{ij}) | b_i, \theta_y) \right] P(b_i | \theta_b) db_i \end{aligned}$$

with  $P(T_i, d_i | b_i, \theta_t)$  the likelihood relative to the survival component of the model:

$$\begin{aligned} P(T_i, d_i | b_i, \theta_t) &= h_i(T_i | M_i(T_i), \theta_t)^{d_i} S_i(T_i | M_i(T_i), \theta_t) \\ &= [h_0(T_i) \exp(W\psi + \alpha m_i(T_i))]^{d_i} \exp \left[ - \int_0^{T_i} h_0(u) \exp(W\psi + \alpha m_i(u)) du \right], \end{aligned}$$

$P(y_i(t_{ij}) | b_i, \theta_y)$  the likelihood of the longitudinal process at time  $t_{ij}$ :

$$P(y_i(t_{ij}) | b_i, \theta_y) = (2\pi\sigma^2)^{-1/2} \exp \left[ - \frac{(y_i(t_{ij}) - m_i(t_{ij}))^2}{2\sigma^2} \right],$$

and  $P(b_i | \theta_b)$  the density of the random effects:

$$P(b_i | \theta_b) = (2\pi)^{-q_b/2} |\Sigma|^{-1/2} \exp \left[ - \frac{b_i^T \Sigma^{-1} b_i}{2} \right]$$

$q_b$  being the dimension of the random effects.

Historically, the predominant method for maximising the full joint likelihood has been the Expectation-Maximisation algorithm ([Dempster et al., 1977](#)); alternatively, it is possible to use general purpose optimisers to maximise the full joint likelihood via algorithms such as the Newton-Raphson algorithm. Nevertheless, significant computational challenges persist.

# Chapter 4

## Computational challenges in survival models with random effects

The models I presented in Chapter 2 and 3 present significant computational challenges during the estimation process. I showed how frailty models with a Gamma frailty are analytically tractable, as it is possible to obtain closed-form expressions for the marginal survival function and therefore the likelihood; conversely, including a log-normal frailty (or, correspondingly, random effects) in a survival model yields a survival function - and likelihood - that does not have a closed form. Recall the  $i^{\text{th}}$ -cluster-specific contribution to the likelihood for a shared frailty model:

$$L_i(\alpha_i) = \alpha_i^{D_i} \prod_{j=1}^{n_i} S_{ij}(t_{ij})^{\alpha_i} h_{ij}(t_{ij})^{d_{ij}}$$

The marginal survival function has the form

$$S_{ij}(t_{ij}) = \int_0^{+\infty} [S_{ij}(t_{ij})]_i^\alpha f(\alpha_i) d\alpha_i$$

with  $f(\alpha_i)$  a log-normal density function. This integral has no closed form, hence it is necessary to approximate it in order to obtain (1) marginal survival and (2) the likelihood.

Analogously, recall the joint likelihood in joint models for longitudinal and survival data:

$$\log L(\theta) = \log \int_{-\infty}^{+\infty} P(T_i, d_i, y_i, b_i; \theta) db_i.$$

Evaluating this likelihood requires evaluating an analytically intractable integral over a possibly multi-dimensional integral over the infinite domain; it is therefore necessary to use some method to approximate it numerically.

Methods for approximating intractable integrals form the majority of this Chapter, with more details in Section 4.1. I will also be introducing numerical methods for differentiating a function and for root-finding in Sections 4.2 and 4.3.

## 4.1 Numerical integration

The term *numerical integration* implies the approximation of the integral of a function; generally, it aims to use the minimum number of function evaluations possible as it tends to be numerically expensive. There is a variety of methods being proposed in literature to perform numerical integration; throughout this Section, I will focus on *quadrature rules*, i.e. any method that evaluates the function to be integrated at some points over the integration domain and combines the resulting values to obtain an approximation of the integral. Quadrature rules vary in complexity and accuracy, and generally accuracy improves as rules get more complex. Additionally, integration of functions in few dimensions is generally not too problematic; the task becomes more difficult when integrating over many dimensions as obtaining an acceptable level of accuracy often requires an unfeasible number of function evaluations.

### 4.1.1 Unidimensional functions

The simplest method to approximate the integral of a unidimensional function numerically is given by the *Riemann sum*. A Riemann sum is an approximation of the integral of a continuous function  $f(x)$  over an integration domain  $[a, b]$  by a finite sum, defined as:

$$\int_a^b f(x) dx \approx \sum_{i=1}^N f(x_i^*) \Delta(x_i),$$

with  $P = \{[x_0, x_1], [x_1, x_2], \dots, [x_{N-1}, x_N]\}$  a partition of  $[a, b]$  such that  $a = x_0 < x_1 < x_2 < \dots < x_{N-1} < x_N = b$ ,  $\Delta(x_i) = x_i - x_{i-1}$ , and  $x_i^* \in [x_{i-1}, x_i]$ .  $x_i^*$  can be defined in many ways: it could be the left extremity of  $\Delta(x_i)$ , the right extremity, the midpoint, or many more. In particular, when choosing  $x_i^*$  as the midpoint of the interval, I obtain the so called *midpoint rule*; it approximates the integral of a continuous function  $f(x)$  by the area under a set of  $N$  step functions, with the midpoint of each matching  $f$ :

$$\int_a^b f(x) dx \approx \frac{b-a}{N} \sum_{i=1}^N f(a + (i-0.5)(b-a)/N)$$

An alternative to the midpoint rule is given by the *trapezoidal rule*, which approximates the area under a continuous function  $f(x)$  as a trapezoid and then computes its area:

$$\int_a^b f(x) dx \approx (b-a) \left[ \frac{f(a) + f(b)}{2} \right]$$

it works best when partitioning the integration area into many subinterval, applying the trapezoidal rule to all of them, and then sum the results:

$$\int_a^b f(x) dx \approx \sum_{i=1}^N \frac{f(x_{k-1}) + f(x_k)}{2} \Delta(x_k),$$

with  $x_k$  a partition of  $[a, b]$  such that  $a = x_0 < x_1 < x_2 < \dots < x_{N-1} < x_N = b$  and  $\Delta(x_k) = x_k - x_{k-1}$  the length of the  $k^{\text{th}}$  subinterval.

Accuracy of the midpoint and trapezoidal rules depends on the number of steps (subintervals)  $N$  used to approximate the function, but so does complexity (computationally speaking). The only requirement for applying these rules is that one needs to be able to evaluate the function  $f(x)$  at a given point over its domain. If  $f(x)$  is cheap to evaluate, than the midpoint and trapezoidal rules may be just fine; otherwise, it would be better to move onto more complicated methods that yield more accurate results.

A first method that is only slightly more complicated but yields better results is the *Simpson's rule*. It works analogously to the midpoint and trapezoidal rule, but using a smooth quadratic interpolant which takes the same values as  $f(x)$  at the extremities of the integration interval  $[a, b]$  and at the midpoint  $m = (a + b)/2$ :

$$\int_a^b f(x) \, dx \approx \frac{b - a}{6} [f(a) + 4f((a + b)/2) + f(b)]$$

Analogously as the trapezoidal rule, it is possible to obtain greater accuracy by splitting the integration interval into many subintervals, applying the Simpson's rule to each subinterval, and sum the results.

Second, it is possible to show that by choosing carefully the points at which to evaluate  $f(x)$  and the weights assigned to each point it is possible to obtain an exact approximation of the integral of any polynomial of degree  $2N - 1$  or less with  $N$  function evaluations (proof in [Monahan \(2011\)](#)). Let  $f(x)$  be a function of order  $2N - 1$  or less to integrate over a domain  $[a, b]$ ; let  $w(x)$  be a weight function. The quadrature formula is defined as:

$$\int_a^b f(x)w(x) \, dx = \sum_{i=1}^N w_i f(x_i)$$

Depending on the choice of the weighting function  $w(x)$ , different Gaussian quadrature rules can be obtained. When  $w(x) = 1$ , the associated polynomials are Legendre polynomials, the quadrature rule is then named *Gauss-Legendre* quadrature rule, and it allows integrating

over the interval  $[-1, 1]$ . The integration points are then obtained as the the  $N$  roots of the Legendre polynomials:  $x = \{x_1, x_2, \dots, x_N\}$ . When choosing the weight function  $\exp(-x)$  the associated polynomials are Laguerre polynomials, the quadrature rule is named *Gauss-Laguerre* quadrature rule, and the integration domain is  $[0, +\infty)$ . Finally, when choosing the weight function  $\exp(-x^2)$  the associated polynomials are Hermite polynomials, the quadrature rule is named *Gauss-Hermite* quadrature rule, and the integration domain is  $(-\infty, +\infty)$ . Interestingly, the Gauss-Hermite quadrature can be re-formulated using a normal density kernel with mean  $\mu$  and standard deviation  $\sigma$  as weighting function:

$$\int_{-\infty}^{+\infty} f(x)\phi(x|\mu, \sigma^2) dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} f(x) \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx$$

By applying the change of variable  $x = \mu + \sigma\sqrt{2}r$ , the integral to approximate becomes

$$\int_{-\infty}^{+\infty} f(x)\phi(x|\mu, \sigma^2) dx = \frac{\sqrt{2}\sigma}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} f(\mu + \sigma\sqrt{2}r) \exp(-r^2) dr,$$

which can then be approximated by the quadrature rule

$$\frac{\sqrt{2}\sigma}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} f(\mu + \sigma\sqrt{2}r) \exp(-r^2) dr \approx \sum_{i=1}^N f(\mu + \sigma\sqrt{2}x_i) \frac{w_i}{\sqrt{\pi}}.$$

That is, a quadrature rule based on the normal kernel as weight function with nodes  $\mu + \sigma\sqrt{2}x_i$  and weights  $w_i/\sqrt{\pi}$  ( $x_i$  and  $w_i$  being the nodes and weights of the corresponding  $N$ -points Gauss-Hermite quadrature rule based on the usual weighting function).

A slightly more complicated version of Gaussian quadrature is given by the *Gauss-Kronrod* quadrature formula. In the Gauss-Kronrod quadrature rule the evaluation points are chosen dynamically so that an accurate approximation can be computed by re-using the information produced by the computation of a less accurate approximation. In practice, integration points from previous iterations can be reused as part of the new set of points, whereas usual Gaussian quadrature would require recomputation of all abscissas at each iteration.

This is particularly important when some specified degree of accuracy is needed but the number of points needed to achieve this accuracy is not known ahead of time. Despite this, the quadrature rule is the same as before, i.e.  $\int_a^b f(x) dx \approx \sum_{i=1}^n w_i f(x_i)$ . Gauss-Kronrod quadrature rule is implemented in R as the `integrate()` function.

### 4.1.2 Multidimensional functions

Finally, all the methods I presented so far only apply to the integration of unidimensional functions. It is of course possible to extend quadrature rules to multidimensional settings, by recursively applying unidimensional quadrature rules. Say I want to approximate the integral of a bidimensional function  $f(x, y)$ ; the bidimensional Gaussian quadrature rule has the form:

$$\int_X \int_Y f(x, y) dx dy \approx \sum_j \sum_i w_j w_i f(x_j, y_i)$$

This can be extended to any number of dimensions  $d$ , but it gets very computationally expensive very quickly as a  $N$ -points rule requires  $N^d$  function evaluations.

A better option when the number of dimensions  $d$  to integrate over is high is given by *Monte Carlo* integration. Consider integrating a multidimensional function  $f(x)$  over some region  $\Omega$  of volume  $V(\Omega)$ :

$$I_\Omega = \int_\Omega f(x) dx = E[f(U)]V(\Omega),$$

with  $U \sim \text{uniform}$  over  $\Omega$ . Drawing  $N$  uniform random vectors  $u_i$  an estimator for  $I_\Omega$  is

$$\hat{I}_\Omega = \frac{V(\Omega)}{N} \sum_{i=1}^N f(u_i),$$

and this defines Monte Carlo integration. The variance of the estimated integral  $\hat{I}_\Omega$  follows, assuming the  $u_i$  are independent, as  $\text{var}(\hat{I}_\Omega) = \frac{V(\Omega)^2}{N^2} N \text{var}(f(u_i))$ . More details in [Monahan \(2011\)](#).

Luckily, both Gaussian quadrature and Monte Carlo integration can be tweaked to improve accuracy and convergence rates: two appealing options are, respectively, adaptive Gaussian quadrature and importance sampling. Adaptive Gaussian quadrature works best when using the Gauss-Hermite rule with the normal density kernel as weighting function; in a multivariate setting, using an iterative algorithm, it is possible to update the mean vector  $M$  and variance-covariance matrix  $\Sigma$  of the multivariate normal density at each step (e.g. using empirical Bayes estimates of  $M, \Sigma$ ) to better adapt the grid of quadrature points to the actual shape of the integral to approximate. Conversely, Monte Carlo integration works best when it is possible to draw a sample from the target distribution (i.e. the distribution of the integral to approximate); unfortunately, that is rarely the case in practice. The idea of importance sampling consists then in drawing a sample from a proposal distribution and then re-weight the estimated integral using importance weights to better adapt to the target distribution.

## 4.2 Numerical differentiation

Numerical differentiation is a series of algorithms to numerically estimate the derivative of a function. They tend to be computationally less demanding than numerical integration methods, but they are more sensitive to cancellation error.

The easiest method for approximating the derivative of a function is to use finite difference approximation. Say I want to estimate the first derivative of a function  $f(\cdot)$  at  $x$ ; the finite difference approximation of the derivative  $f'(x)$  is calculated as

$$f'(x) \approx \frac{f(x + h) - f(x)}{h},$$

for a small  $h$ . This formula is affected by both truncation error (as it derives from a truncated Taylor series expansion of  $f(x)$ ) and cancellation error (as a machine works with finite-

precision arithmetic). It is necessary to choose a value  $h$  that gives a good balance between the two errors: it can be showed that a good choice in most cases is  $h = \sqrt{\epsilon}$ , with  $\epsilon$  being the machine precision.

The formula I presented for finite difference approximation is also known as *forward differencing*; alternatively, it is possible to use methods such as *central differencing* ( $[f(x + h) - f(x - h)]/2h$ , more accurate but more computationally expensive) and *backward differencing* ( $[f(x) - f(x - h)]/h$ ). Other methods are the *complex method*, which requires the function to be able to handle complex values and it is extremely powerful but with limited applicability, and the *Richardson's extrapolation method*, which is more accurate but slower than finite differencing. All these methods are implemented in R in the `numDeriv` package, which sets the standard for numerical differentiation.

## 4.3 Numerical root finding

Root-finding algorithms are algorithms for finding the values  $x$  such that  $f(x) = 0$ , for a given continuous function  $f(\cdot)$ . Such values  $x$  are named roots (or zeros) of a function. Most root-finding algorithms are based on the intermediate value theorem, which states that if a continuous function has values of opposite sign at the end points of an interval then the function has at least one root in the interval.

For instance, the easiest root-finding method is the *bisection method*: let  $f(x)$  be a continuous function, for which one knows an interval  $[a, b]$  such that  $f(a)$  and  $f(b)$  have opposite sign. Let  $c = (a + b)/2$  be the midpoint the bisect the interval: now, either  $f(a)$  and  $f(c)$  or  $f(c)$  and  $f(b)$  have opposite sign, and one has in fact divided by two the size of the interval. One can iterate this method until the difference between the extremities of the interval is small enough (e.g.  $< 1 \times 10^{-8}$ ).

Another well established method is the *secant method*: it uses a succession of roots of secant

lines to approximate the root of a function  $f(x)$ . Starting with values  $x_0$  and  $x_1$ , a line is constructed between  $(x_0, f(x_0))$  and  $(x_1, f(x_1))$ :

$$y = \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_1) + f(x_1)$$

The root of this line is

$$x = x_1 - f(x_1) \frac{x_1 - x_0}{f(x_1) - f(x_0)}$$

Now, we set  $x_2 = x$  and we iterate this method until the difference between the extremities of the interval is small enough (e.g.  $< 1 \times 10^{-8}$ ).

The secant method is also known as a *linear interpolation* method; it is also possible to use higher order interpolation, specifically *quadratic interpolation*, to find the root of a function using the same rationale presented for the secant method. Specifically, starting with three starting values  $x_0, x_1, x_2$  and their function values  $f(x_0), f(x_1), f(x_2)$ , applying the Lagrange interpolation formula to interpolate the inverse of  $f(x)$  yields the equation

$$f^{-1}(y) = \frac{(y - f(x_1))(y - f(x_2))}{(f(x_0) - f(x_1))(f(x_0) - f(x_2))} x_0 + \frac{(y - f(x_0))(y - f(x_2))}{(f(x_1) - f(x_0))(f(x_1) - f(x_2))} x_1 + \frac{(y - f(x_0))(y - f(x_1))}{(f(x_2) - f(x_0))(f(x_2) - f(x_1))} x_2$$

Substituting  $y = 0$  in the above equation yields the recursion formula, to be iterated until a desired precision is reached.

Finally, a well-established and robust method is the *Brent-Dekker* method, implemented in R with the `uniroot()` function. It combines the three methods presented before, trying to use the secant or quadratic interpolation method first - as they tend to converge faster to a solution - but falling back to the bisection method if necessary, for its robustness properties. More details on the Brent-Dekker method in [Brent \(1973\)](#).



# Chapter 5

## Simulating survival data

In this Chapter I will present a flexible and efficient method to simulate survival data from a variety of parametric distributions, first introduced by [Bender et al. \(2005\)](#). Then, I will present an extension that allows simulating from a variety of complex distributions proposed by [Crowther and Lambert \(2013\)](#).

Let  $h(t) = h_0(t) \exp(X\beta)$  be the hazard function of a proportional hazards model, with  $h_0(t)$  baseline hazard function and  $X$  a matrix of covariates with regression coefficients  $\beta$ . Let  $H_0(t) = H_0(t) \exp(X\beta)$  be the corresponding cumulative hazard function, with  $H_0(t) = \int_0^t h_0(u) du$ . The survival function  $S(t)$  and cumulative distribution function  $F(t)$  follow naturally:  $S(t) = \exp(-H(t))$  and  $F(t) = 1 - S(t) = 1 - \exp(-H(t))$ .

[Bender et al. \(2005\)](#) showed that by letting

$$F(\tau) = u, \quad u \sim U(0, 1)$$

and denoting the simulated survival time with  $\tau$ , it is possible to derive  $\tau$  analytically by

inverting  $H_0(t)$  if  $h_0(\tau) > 0$ :

$$\tau = H_0^{-1}(-\log(u) \exp(X\beta))$$

The only requirement is for  $H_0(t)$  to be directly invertible, which happens to be the case when simulating from an exponential, Weibull, or Gompertz distribution for the baseline hazard  $h_0(t)$ . The algorithm for simulating  $m$  survival times is as follows:

1. draw a vector  $u$  of  $m$  observations from a  $U(0, 1)$  distribution;
2. simulate  $X$  (e.g. a binary treatment from a Bernoulli distribution) and fix  $\beta$ ;
3. the survival times can be obtained directly by applying the formula  $H_0^{-1}(-\log(u) \exp(X\beta))$ .

[Bender et al. \(2005\)](#) derived the closed-form version of  $H_0^{-1}(t)$  for the exponential, Weibull, and Gompertz distributions, presented in Table 5.1.

Table 5.1: Closed-form formulas for simulating survival data from an exponential, Weibull, or Gompertz distribution.

	Exponential	Weibull	Gompertz
Hazard function $h_0(t)$	$\lambda$	$\lambda p t^{p-1}$	$\exp(\gamma t)$
Cumulative hazard function $H_0(t)$	$\lambda t$	$\lambda t^p$	$(\lambda/\gamma)(\exp(\gamma t) - 1)$
Inverse cumulative hazard function $H_0^{-1}(t)$	$\lambda^{-1}t$	$(\lambda^{-1}t)^{1/p}$	$(1/\gamma) \log((\gamma/\lambda)t + 1)$
Survival time $\tau$	$-\frac{\log(u)}{\lambda \exp(X\beta)}$	$\left[-\frac{\log(u)}{\lambda \exp(X\beta)}\right]^{1/p}$	$(1/\gamma) \log \left[1 - \frac{\gamma \log(u)}{\lambda \exp(X\beta)}\right]$

The requirement requirement for  $H_0(t)$  to be directly invertible impedes the use of distributions other than the exponential, Weibull, or Gompertz - which could be appropriate in some setting but too restricting in others. [Crowther and Lambert \(2013\)](#) generalised this method in order to accommodate more complex distributions, even with turning points, under a

proportional hazards model. In brief, when the cumulative baseline hazard function is not invertible it is not possible to solve the equation  $F(\tau) = u$  for  $\tau$ ; assuming it is possible to write  $H_0(t)$  analytically - a broader assumption compared to assuming  $H_0(t)$  is invertible - it is possible to use root-finding methods to solve for  $\tau$  numerically (Section 4.3). Formally, the survival time  $\tau$  can be simulated as the root of the equation  $S(\tau) - u = 0$ .

[Crowther and Lambert \(2013\)](#) present an example of their method by simulating from a two-components mixture distribution ([McLachlan and McGiffin, 1994](#)), defined by additive components on the survival scale:

$$S_0(t) = \pi S_1(t) + (1 - \pi) S_2(t),$$

with  $\pi \in [0, 1]$  mixing parameter.  $S_i(t)$  can be any standard parametric distribution. Choosing two Weibull components for the mixture distribution, it can be showed that a proportional hazards model has the form

$$h(t) = \frac{\lambda_1 p_1 t^{p_1-1} \pi \exp(-\lambda_1 t^{p_1}) + \lambda_2 p_2 t^{p_2-1} (1 - \pi) \exp(-\lambda_2 t^{p_2})}{\pi \exp(-\lambda_1 t^{p_1}) + (1 - \pi) \exp(-\lambda_2 t^{p_2})} \exp(X\beta);$$

the survival function can be obtained directly from  $h(t)$  in closed-form, plugged into the equation  $S(\tau) - u = 0$ , and numerically solved for  $\tau$ .



# Chapter 6

## Simulation study: accuracy of Gaussian quadrature

In this Chapter I will present the first simulation study I run during my first year, on accuracy on Gaussian quadrature methods. I presented part of this work as an oral presentation at the 2017 Survival Analysis for Junior Researchers conference (more details in Chapter 11 and slides available in Appendix C.1). The Chapter is partitioned into seven Sections, and in each Section I will present an aspect of the simulation study: the aim of the study (Section 6.1), the data-generating mechanisms (Section 6.2), the methods that I will compare (Section 6.3) and the estimands of interest (Section 6.4), the performance measures I will use to compare the methods (Section 6.5), the results of the study (Section 6.6), and finally a brief discussion in Section 6.7.

### 6.1 Aim

The aim of this simulation study is two-fold. First, I want to assess the accuracy of numerical integration methods - Gaussian quadrature, specifically - in settings where it is possible to

obtain analytical formulae; analytical formulae will be used as a control method. Second, I aim to assess the accuracy of Gaussian quadrature when analytical formulae are not available and therefore quadrature is indeed required. In order to fulfill these two aims I will simulate clustered survival data with a frailty component shared between individuals belonging to the same cluster; the distribution I will assign to the frailty will determine whether analytical formulae are available or not. I will choose a Gamma distribution for answering the first aim, and a log-normal distribution for the frailty using the random intercept parametrisation for answering the second aim.

## 6.2 Data-generating mechanisms

I generated survival data from a Weibull distribution with shape parameter  $\lambda = 0.5$  and scale parameter  $p = 0.6$  using the method of [Bender et al. \(2005\)](#) as explained in Section 5, and applying administrative censoring at time  $t = 5$ ; I used the following parametrisation for the Weibull distribution:

$$h(t) = \lambda p t^{p-1}$$

I included a binary covariate (e.g. a treatment) simulated by drawing from a Bernoulli random variable with parameter  $\pi = 0.5$ , and a frailty term shared between individuals in a cluster by drawing first from a Gamma distribution with shape parameter  $1/\theta$  and scale parameter  $\theta$  (for identifiability purposes) and then by drawing from a normal distribution with mean  $\mu = 0$  and standard deviation  $\sigma = \sqrt{\theta}$ . I varied  $\theta$ :  $\theta = \{0.25, 0.75, 1.25\}$ . I also varied the regression coefficient (e.g. the log-treatment effect)  $\beta$  associated with the binary covariate:  $\beta = \{-0.50, 0.00, 0.50\}$ . I simulated data for six different sample sizes: 15 clusters of 30, 100, or 500 individuals each, 50 clusters of 30 or 100 individuals, and 1000 clusters of 2 individuals; sample size then varied between 450 and 7500 individuals. Finally, I used a fully factorial design combining different frailty variances, frailty distributions, treatment effects,

and sample sizes; it resulted in 3 times 2 times 3 times 6 = 108 different data-generating mechanisms, and for each of them I generated 1000 datasets.

The 54 simulated scenarios with a Gamma frailty will be used to answer the first aim of the simulation study, and the remaining 54 scenarios with a log-normal frailty will be used to answer the second aim.

## 6.3 Methods

I fitted a set of models for each simulated dataset under each data-generating mechanism. Specifically, for the data generated assuming a Gamma frailty, I compared the following models:

- a shared Gamma frailty model with a baseline Weibull hazard using the analytical formulation of the likelihood (method *AN*);
- a shared Gamma frailty model with a baseline Weibull hazard using the likelihood approximated numerically via Gaussian quadrature (specifically, a Gauss-Laguerre quadrature rule) with 15, 35, 75, and 105 nodes (methods *GQ15*, *GQ35*, *GQ75*, *GQ105*);
- a shared Gamma frailty model with a baseline Weibull hazard using the likelihood approximated numerically via Gauss-Kronrod quadrature (as implemented in R's `integrate()` function; method *IN*).

Then, for data generated assuming a log-normal frailty, I fitted a Weibull model with a random intercept using the likelihood approximated via Gauss-Hermite quadrature using 15, 35, 75, and 105 nodes.

## 6.4 Estimands

For each model fitted under each scenario, I compared:

1. the estimated parameters of the Weibull distribution, i.e.  $\hat{\lambda}$  and  $\hat{p}$ ;
2. the estimated log-treatment effect, i.e.  $\hat{\beta}$ ;
3. the estimated variance of the frailty term, i.e.  $\hat{\theta}$ .

## 6.5 Performance measures

I am interested first of all in the performance of the maximum likelihood estimation procedure; that is, how precise is the maximum likelihood estimator. I will assess this by computing bias for each estimand, defined as  $b = E(\hat{\beta}) - \beta$ .

Next, I am interested in coverage, i.e. the proportion of times the  $100 \times (1 - \alpha)\%$  confidence interval  $\hat{\beta} \pm Z_{1-\alpha/2} \times SE(\hat{\beta})$  includes the true value  $\beta$ . This allow to assess whether the empirical coverage rate approaches the nominal coverage rate ( $100 \times (1 - \alpha)\%$ ), to properly control the type I error rate for testing a null hypothesis of no effect.

Finally, I am interested in overall accuracy and therefore I will compute the mean squared error, defined as the sum of bias and variability:  $(\bar{\hat{\beta}} - \beta)^2 + (SE(\hat{\beta}))^2$ .

Summary measures for  $\lambda$ ,  $p$ , and  $\theta$  are computed on the log-scale. For bias and coverage I will further include Monte Carlo standard errors to quantify the uncertainty in estimating the performance measures (further details in [White \(2010\)](#)). I will also discuss convergence rates of the different methods included in the comparison.

## 6.6 Results

I selected a single scenarios for each aim of this simulation study (out of 54) to present, for conciseness. Specifically, I will present results for the setting of a small frailty variance (0.25) with a negative regression coefficient (-0.50). The full results can be explored online interactively by clicking [here](#).

### 6.6.1 Aim 1: comparison of Gaussian quadrature with analytical formulae

Bias, coverage, mean squared error, and convergence rates for scenario 1 are presented in Tables A.1, A.2, A.3, A.4 and Figures B.2, B.3, B.4, B.1. Convergence rates were generally good ( $> 90\%$ ) for most sample sizes; the method that showed the worst convergence rates was Gauss-Kronrod quadrature with 1000 clusters of 2 individuals each, where approximately 75% of replications converged. Bias, coverage, and overall accuracy were optimal for all methods and across all sample sizes for the scale parameter of the Weibull distribution  $p$  and the regression coefficient  $\beta$ ; conversely, the methods performed quite differently for the shape parameter  $\lambda$  and the frailty variance  $\theta$ . The shape parameter estimated using analytical formulae or Gauss-Kronrod quadrature was generally unbiased, with good coverage and accuracy; vice versa, using Gauss-Laguerre quadrature produced underestimated coefficients when using a small number of nodes and required at least 75 nodes to yield unbiased results. As the number of nodes increased, coverage and mean squared error improved considerably. Also, sample sizes with a higher number of clusters generally yielded better estimates for the shape parameter in terms of bias, coverage, and mean squared error. The frailty variance  $\theta$  was the parameter estimated with greatest variability in the results. Analytical formulae required a high number of clusters to produce unbiased results (50 or 1000), yielding underestimated coefficients otherwise. Gauss-Kronrod performed similarly to analytical formulae,

as did Gauss-Laguerre quadrature with a sufficiently high number of nodes. Coverage was generally good, above 90% (except Gauss-Laguerre with 15 nodes, where coverage fell to 60-70% in some settings), symptom of overestimated standard errors for the frailty variance; this inflation of the standard errors was reflected in the mean squared error, which was generally greater than the other estimated parameters for all methods under all sample sizes explored in this scenario.

### 6.6.2 Aim 2: accuracy when analytical formulae are not available

Bias, coverage, mean squared error, and convergence rates for scenario 1 are presented in Tables A.5, A.6, A.7, A.8 and Figures B.6, B.7, B.8, B.5. Convergence rates are good for all sample sizes ( $> 97\%$ ) except when assuming 15 clusters of 500 individuals, where convergence rates drop to approximately 50% for all methods included in this comparison. Bias is generally negligible for the parameters of the Weibull distribution  $\lambda$  and  $p$  and the regression coefficient  $\beta$ : between 0.0059 and 0.0193 for  $\lambda$ , between -0.0424 and -0.0332 for  $p$ , between 0.0040 and 0.0867 for  $\beta$ . Conversely, estimates for  $\sigma$  were negatively biased for a sample size of 15 clusters - 100 individuals, 1000 clusters - 2 individuals, 15 clusters - 30 individuals (between -0.3057 and -0.0854) and positively biased for a sample size of 15 clusters - 500 individuals (between 0.2427 and 0.4020). Bias was negligible for a sample size of 50 clusters - 30 individuals and 50 clusters - 100 individuals (between -0.0536 and -0.0095). Coverage of all estimated coefficients was poor ( $< 75\%$ ) for a sample size of 15 clusters - 500 individuals. For the regression coefficient  $\beta$  and the frailty variance  $\sigma$  coverage was good or superoptimal for the remaining sample sizes, with the exception of  $\sigma$  estimated using Gauss-Hermite quadrature with 15 nodes that resulted in slight undercoverage for sample sizes of 15 clusters - 100 individuals and 50 clusters - 100 individuals. The parameters of the Weibull distribution were generally undercovered ( $< 80\%$ ) across sample sizes, except  $\lambda$  with a sample size of 1000 clusters - 2 individuals and  $p$  with a sample size of 15 clusters -

30 individuals for which coverage was in the range 90-95%. Finally, mean squared error was low for  $\lambda$ ,  $p$ , and  $\beta$ , comparable for  $\sigma$  with a sample size of 50 clusters - 30 individuals and 50 clusters - 100 individuals, much higher for  $\sigma$  with all the remaining sample sizes (i.e. overall accuracy was lower in these settings).

## 6.7 Conclusions

I showed in the previous Section how Gaussian quadrature performs (1) compared to analytical formulae and (2) when it is not possible to obtain analytical formulae. Overall, Gaussian quadrature performs well with a sufficient number of quadrature nodes but the variability is great. The regression coefficient  $\beta$  is the most robust estimand across different scenarios, it is mostly unbiased (or with little bias) and with good coverage and accuracy (in terms of mean squared error). The frailty variance is the least robust estimand, with precision and accuracy greatly depending on many factors: among others, important ones seems to be the number of quadrature nodes and the number of clusters. The latter makes sense on a theoretical level: with more clusters it should be easier to estimate a properly the variance of the frailty. Accuracy and precision of the parameters of the Weibull baseline hazard also varies greatly. In conclusion, using a shared frailty model to do inference on a regression coefficient seems to be robust to the accuracy of numerical integration methods; nevertheless, if the principal research interest lays in relative risk estimates, using a parametric model may not be the best choice after all. A semiparametric Cox model - even with frailty terms if necessary - could be utilised instead. If the research objectives include absolute risk estimations, though, a parametric model is immediately more appealing. However, checking the convergence, precision, and accuracy of numerical integration by evaluating and comparing an increasing number of quadrature knots appears to be fundamental.



# Chapter 7

## Simulation study: impact of misspecification in survival models with shared frailty terms

In this Chapter, I will present the second simulation study I set up and run during my first year. It investigates the impact of model misspecification in survival models with shared frailty terms, and part of this work was presented in oral form at the 2017 Statistical Analysis of Multi-Outcome Data (SAM) Conference and at the 38<sup>th</sup> Annual Conference of the International Society for Clinical Biostatistics (more details in Chapter 11 and slides available in Appendix C.2). I am also currently writing up this project into a manuscript for submission to a journal; a current draft is attached as well in Appendix D.

This Chapter is arranged as follows. First, I introduce the aim of the simulation study in greater detail in Section 7.1. Then, in Section 7.2 I will introduce the data-generating mechanisms, in Section 7.3 I will describe the different models I included in the comparison, in Section 7.4 I will define the estimands of interest, in Section 7.5 I will present the performance measures used to compare the different models, in Section 7.6 I will present some results,

and finally I will conclude the Chapter in Section 7.7.

## 7.1 Aim

The de-facto standard method used in medical research when dealing with time to event data is the Cox proportional hazards model. It is best suited when relative risk estimates are the quantities of interest. However, often the focus is on absolute measures of risk: in that context, modelling the baseline hazard is necessary, and it can be achieved by using standard parametric survival models with a simple parametric distribution (such as the exponential, Weibull, or Gompertz distribution) or by using the flexible parametric modelling approach (Royston and Parmar, 2002) to better capture the shape of complex hazard functions. The latter approach requires choosing the number of degrees of freedom for the spline term used to approximate the baseline hazard: in practice, sensitivity analyses and information criteria (AIC, BIC) have been used to select the best model. Recently, Rutherford et al. (2015) showed via simulation studies that, assuming a sufficient number of degrees of freedom is used, the approximated hazard function given by restricted cubic splines fit well for a number of complex hazard shapes and the hazard ratios estimation is insensitive to the correct specification of the baseline hazard. Moreover, it is common to encounter clustered survival data where the overall study population can be divided into heterogeneous clusters of homogeneous observations; examples are given in Chapter 2. As a consequence, survival times of individuals within a cluster are likely to be correlated and need to be analysed as such by including a random effect, e.g. a frailty term.

Flexible parametric survival models are a robust alternative to standard parametric survival models when the shape of the hazard function is complex; using a sufficient number of degrees of freedom, e.g. 2 or more, the spline-based approach is able to capture the underlying shape of the hazard function with minimal bias. AIC and BIC can guide the choice of the

best fitting model, but they tend to agree to within 1 or 2 degrees of freedom in practice (Rutherford et al., 2015). Analogously, the impact of the choice of a particular parametric frailty distribution on the regression coefficients is minimal (Pickles and Crouchley, 1995). Conversely, little is known about the impact of misspecifying the baseline hazard in survival models with frailty terms.

My aim with this work is to assess the impact of misspecifying the baseline hazard or the frailty distribution on the estimated regression coefficients, frailty variance, and absolute, marginal risk measures such as the integrated difference of survival curves and the survival difference at given time points. I will simulate data under a variety of data-generating mechanisms, and then compare a set of models that include the Cox model with frailties, fully parametric survival models with frailty, models with flexible baseline hazard, and models with flexible baseline hazard and a penalty for the complexity of the spline term.

## 7.2 Data-generating mechanisms

I simulate data under five different baseline hazard functions using the approaches presented in Chapter 5: Exponential, Weibull, Gompertz, and a two different two-components mixture Weibull-Weibull with turning points. In practice, I choose the following hazard functions: exponential with scale  $\lambda = 0.3$ , Weibull with scale  $\lambda = 0.5$  and shape  $p = 0.6$ , Gompertz with scale  $\lambda = 0.1$  and shape  $\gamma = 0.5$ , two-components mixture Weibull with scale parameters  $\lambda_1 = 0.5$ ,  $\lambda_2 = 0.3$ , shape parameters  $p_1 = 2.5$  and  $p_2 = 1.3$ , and mixing parameter  $\pi = 0.8$ , and two-components mixture Weibull-Weibull with scale parameters  $\lambda_1 = \lambda_2 = 1.0$ , shape parameters  $p_1 = 1.5$  and  $p_2 = 0.5$ , and mixing parameter  $\pi = 0.5$  (Figure 7.1). Then, for all possible baseline hazard function, I generated clustered data assuming 15 clusters of (30, 100) individuals each, 50 clusters of (30, 100) individuals each, or 1000 clusters of 2 individuals each. I included a binary treatment variable  $X \sim Bin(1, 0.5)$  with associated

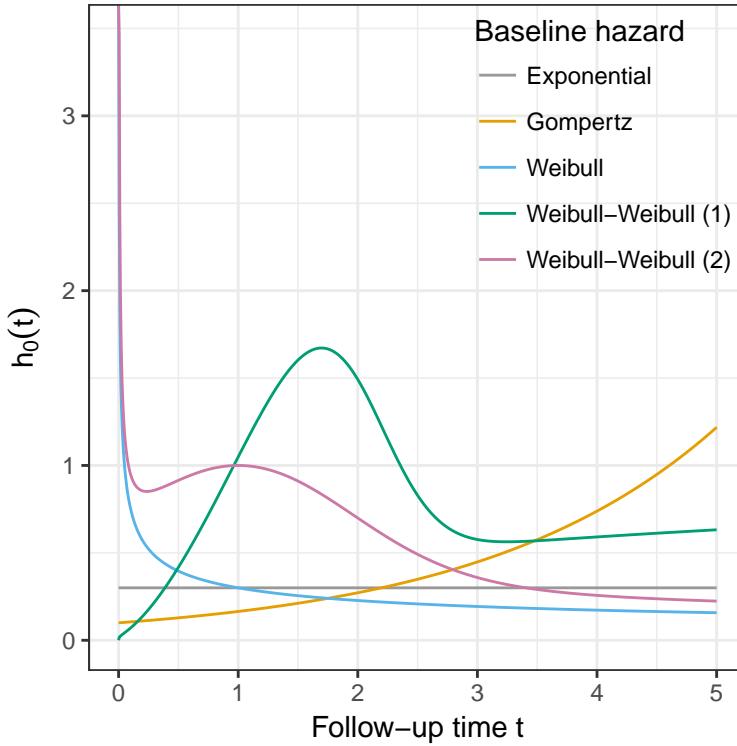


Figure 7.1: Baseline hazard functions chosen for this simulation study.

log-hazard ratio of  $-0.5$  and cluster-specific frailty terms  $\alpha_i$  following either a Gamma or a log-normal distribution with variance  $\theta$  ( $\theta = \{0.25, 0.75, 1.25\}$ , Figure 7.2). Finally, I generated an event indicator variable  $d$  by applying administrative censoring at 5 years. The true marginal survival functions corresponding to these simulated settings are depicted in Figure 7.3.

I applied a fully factorial design: this resulted in 150 simulation scenarios, 5 sample sizes  $\times$  5 baseline hazards  $\times$  2 frailty distributions  $\times$  3 true frailty variances.

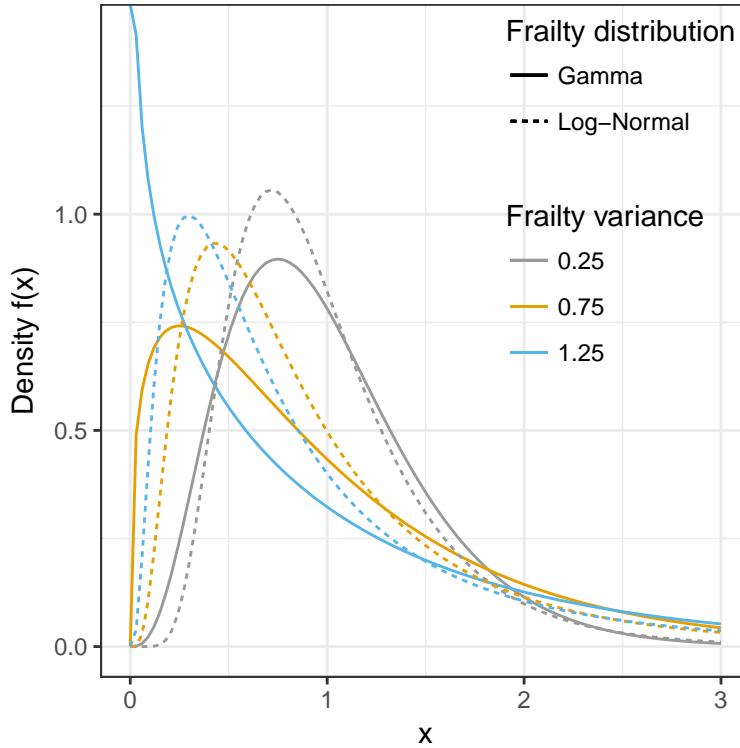


Figure 7.2: Frailty distributions chosen for this simulation study.

## 7.3 Methods

In this Section I will introduce the models I fitted in this simulation study. First, I fit a Cox model with a shared frailty term:

$$h_{ij}(t|\alpha_i) = \alpha_i h_0(t_{ij}) \exp(X_{ij}\beta),$$

with  $h_0(\cdot)$  left unspecified. The Cox model with a shared Gamma frailty is implemented in the R package `frailtyEM`, while the Cox model with a shared log-normal frailty is implemented in the R package `coxme`. Since the `coxme` package does not return a standard error for the estimated frailty variance by default, I used bootstrap with 1000 replications to estimate it; I resampled clusters of individuals rather than individuals to preserve the correlation within cluster. Then, I fitted fully parametric survival models with a shared frailty term, using the same model formulation of the Cox model but specifying the baseline hazard

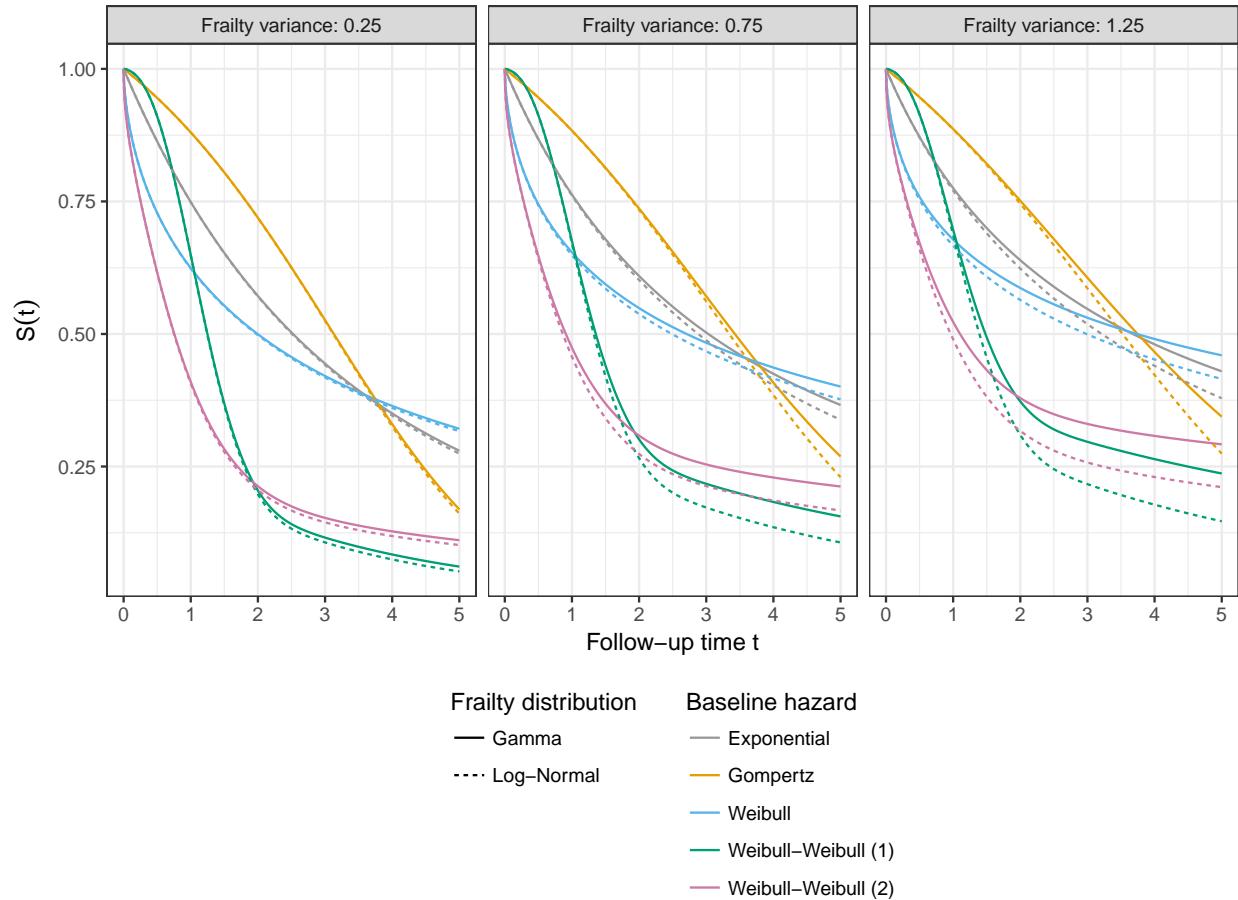


Figure 7.3: Marginal survival depending on baseline hazard and frailty distribution chosen for this simulation study.

function. I fitted six models, for each combination of baseline hazard (Exponential, Weibull, or Gompertz) and frailty distribution (Gamma, log-normal). These models are implemented in the R package `parfm`. Finally, I fit flexible Royston-Parmar models with a shared frailty term, either Gamma or log-normal:

$$\log H(t_{ij}|\alpha_i) = s(z; \gamma) + X\beta + \log(\alpha_i),$$

with  $s(\cdot)$  a restricted cubic spline function of log-time that smooths the logarithm of the baseline cumulative hazard  $H_0(\cdot)$ . This model requires choosing the number of degrees of freedom of the spline term, hence I varied between 3, 5, 7, and 9 degrees of freedom. I also fitted the same model using penalised likelihood (Liu et al., 2016), applying a penalty to the likelihood to avoid overfitting and sparing myself from having to choose the number of degrees of freedom for the spline term. These flexible parametric models are implemented in the R package `rstpm2`.

## 7.4 Estimands

The first estimand of interest is the regression coefficient  $\beta$  associated with the simulated binary treatment. This coefficient can be interpreted as log-treatment effect, and it would be interesting to see if and how misspecification of the baseline hazard or frailty distribution affects relative risk estimates. Second, I am interested in comparing estimates of the frailty variance obtained from each model - a quantity often used to quantify dependence between clustered observations. Finally, I am going to compare two measures of absolute risk: marginal survival difference at time  $t$ , defined as  $S(t)_{\text{diff}} = S(t|x=1) - S(t|x=0)$ , and integrated marginal survival difference, defined as  $iS_{\text{diff}} = iS(x=1) - iS(x=0)$ . The former is obtained by fixing the time  $t$  (I am using 1, 2, 3, and 4 years), and then integrating out the frailty term from the conditional survival function as explained in Chapter 2. Conversely,

the latter requires integrating the marginal survival function for both treatment groups and then computing their difference. I estimate it as follows:

1. I estimate marginal survival for a treatment group at 1000 equally spaced values of follow-up time  $t$ ;
2. I fit an interpolating natural spline over the 1000 estimates from step (1) using the `splinefun` R function;
3. I integrate the resulting spline function using Gauss-Kronrod quadrature as implemented in the `integrate` function;
4. I compute the difference of the integrated marginal survival for the two treatment groups.

The integral of a survival function can be interpreted as life expectancy; hence, the quantity I am computing can be interpreted as the difference in 5-years life expectancy between treated and untreated individuals.

## 7.5 Performance measures

Performance measures of interest are bias, coverage, and mean squared error - as in the previous simulation study of Chapter 6 (more details on each summary measure in Section 6.5). I will also include Monte Carlo standard errors for bias and coverage to quantify uncertainty in estimating such performance measures.

## 7.6 Results

Among the 150 different data-generating mechanisms I simulated data from, I chose (for conciseness) to present results only for the settings of 15 clusters of 100 individuals each,

with a frailty variance of 0.25. I will also exclude Royston-Parmar models with 3 or 7 degrees of freedom from this comparison, again for conciseness. Additional results can be explored interactively by clicking [here](#).

Convergence rates for the models included in this comparison are presented in Table A.9 and plot B.9. They varied considerably between different baseline hazards and frailty distributions, and were generally greater when the true baseline hazard was simple (exponential, Weibull, or Gompertz) with some exception, specifically the Cox model with a Gamma frailty and the Gompertz model with either frailty and a misspecified baseline hazard. With the first Weibull-Weibull baseline hazard convergence was optimal for all models except the Cox model with a Gamma frailty; with the second Weibull-Weibull baseline hazard, convergence rates dropped for all models except the Cox model with a log-normal frailty and parametric models with an exponential or Weibull baseline hazard and either frailty distributions.

Bias, coverage, and mean squared error of the estimated regression coefficient are presented in Tables A.10, A.11, A.12 and Figures B.10, B.11, B.12. With a true exponential baseline hazard, all models produced unbiased estimates under all scenarios; with a true Weibull baseline hazard, all models performed well except the parametric models with an exponential or Gompertz baseline hazard, which yielded underestimated regression coefficients (approximately -0.09 on the log-hazard rate scale). Analogously, with a true Gompertz baseline hazard the parametric Gompertz model, the flexible parametric models, and the Cox models performed well with unbiased estimates; the parametric exponential and Weibull models yielded overestimated results (approximately 0.13). With the first Weibull-Weibull baseline hazard, the flexible parametric models and the Cox model performed well; conversely, the parametric models yielded overestimated results (exponential and Gompertz, approximately -0.07) or underestimated results (Weibull, approximately 0.10). Similarly, with the second Weibull-Weibull the flexible parametric models and the Cox model returned unbiased estimates; the Weibull model returned unbiased results too. The exponential and Gompertz

parametric models, on the other side, return underestimated results (approximately -0.11). Coverage followed a similar pattern; when the model yielded unbiased results, coverage was optimal at approximately 95%. Conversely, when the estimates were biased and the parametric distribution was misspecified or failed to capture a complex hazard shape, coverage dropped up to 35% (with the exponential model performing worst). Mean squared error of the estimated coefficients was the smallest when the model was well specified, or when using the Cox model or Royston-Parmar models.

Bias, coverage, and mean squared error of the estimated frailty variance are presented in Tables A.13, A.14, A.15 and Figures B.13, B.14, B.15. With a true exponential baseline hazard, all models yielded slightly biased results irrespectively of the frailty distribution: models with a well specified frailty distribution yielded slightly negatively biased results (-0.03 to -0.01; the Cox model with a Gamma frailty performed worse with a negative bias of -0.09). When assuming a Gamma frailty in place of a log-normal frailty, negative bias was a somewhat greater (around -0.05, with the Cox model once again performing worse with a negative bias of -0.11); when assuming a log-normal frailty in place of a Gamma frailty, results were slightly positively biased (approximately 0.01). With more complex true baseline hazard functions, the flexible parametric models performed the best with performance similar to the exponential setting; conversely, fully parametric models performed worse when the baseline hazard was misspecified (with both negative and positive bias depending on the setting, up to -0.15 and 0.10). With a complex baseline hazard, negative and positive bias for the fully parametric models further increased up to -0.15 and 0.15, approximately. The Cox model with a Gamma frailty performed the worst, severely underestimating the frailty variance (up to -0.18, approximatley). Coverage was generally suboptimal, in the range 70-90%, with a few exceptions; the fully parametric models sometimes showed good coverage, symptom of overestimated standard errors (given that they returned biased estimates). Mean squared errors reflected the pattern observed for bias, with the flexible parametric models performing better than the other models across the range of frailty ditributions and baseline

hazards examined; the parametric models performed similarly when well specified, slightly worse otherwise. The Cox model with a log-normal frailty performed similarly to the Royston-Parmar models, while the Cox model with a Gamma frailty performed worse, especially with a complex baseline hazard (where it performed even worse than fully parametric models).

Finally, bias and mean squared error of the estimated difference in 5-years life expectancy are presented in Tables A.16, A.17 and Figures B.16, B.17. With a true Gamma frailty, the flexible parametric models perform well with estimates of the difference in 5-years life expectancy that are practically unbiased; the parametric models performed well when well specified, returned slightly biased results otherwise (both negative and positive bias, up to -0.04 and 0.08 respectively - an absolute difference of 0.5 to 1.0 months in terms of time). With a true log-normal frailty distribution, the Royston-Parmar models produced slightly overestimated results (0.01 to 0.05), while the remaining models performed similarly to the setting with a true Gamma distribution. Bias slightly increased with a complex baseline hazard when using parametric models, up to 0.12 (i.e. approximately 1.5 months). Mean squared errors showed a similar pattern, with all models performing comparably with a true exponential or Gompertz baseline hazard, and the flexible parametric models performing best otherwise (compared to misspecified models). The Cox model generally performed similarly to slightly worse than the flexible parametric models.

## 7.7 Conclusions

I showed that estimates of regression coefficients, frailty variance, and difference in expectation of life are relatively unsensitive to misspecification of the frailty distribution of the model. Conversely, misspecifying the baseline hazard has serious consequences as it impacts both relative and absolute measures of risk, and heterogeneity measures. This seems to be particularly important with respect to the regression coefficients, as bias on the log-hazard

ratio scale of up to 0.13 corresponds to a difference of approximately 15% on the relative risk scale, a clinically meaningful difference. All models seemed to produced biased estimates of the frailty variance, which may be due to the small number of cluster examined here; exploring additional scenarios will provide a greater insight on the topic. The bias in the difference of 5-years expectation of life seems to be less clinically relevant (bias up to 1.5 months), but it is something to bear in mind nonetheless. The fully parametric models perform well (as expected) when well specified, but relatively simple hazard forms may be too restrictive and unrealistic in practice; conversely, flexible parametric models showed robustness to all different shapes of the baseline hazards and generally performed best, even compared to the Cox model. Further to that, this robustness seemed to be independent on the number of knots for modelling the baseline hazard and on the estimation method (full or penalised likelihood).

# Chapter 8

## Exploring results from simulation studies interactively

The simulation studies I presented in Chapters 6 and 7 presented multiple challenges, one of them being how to present the results effectively given the amount of simulated scenarios (108 and 150 scenarios for simulation 1 and 2, respectively). Each scenario would require producing a variety of tables and plot for bias, coverage, and any other summary statistics I may be interested in: the amount of tables and plots would grow dramatically to an unsustainable number. A straightforward (and often used in practice) option could be selecting a handful of scenarios to present, limiting the number of tables and plots to what is believed to be most interesting. However, a reader may find other scenarios more interesting, or would like to compare different factors or even deep down more into the results; presenting only a subset of results may result limiting to some extent then. Therefore, facing this problem myself, I set out to develop an interactive tool to aid exploration and dissemination of results from simulation studies. Part of this work was presented in oral form at the Students' Day of the 38<sup>th</sup> Conference of the International Society for Clinical Biostatistics; slides are attached in Appendix C.3.

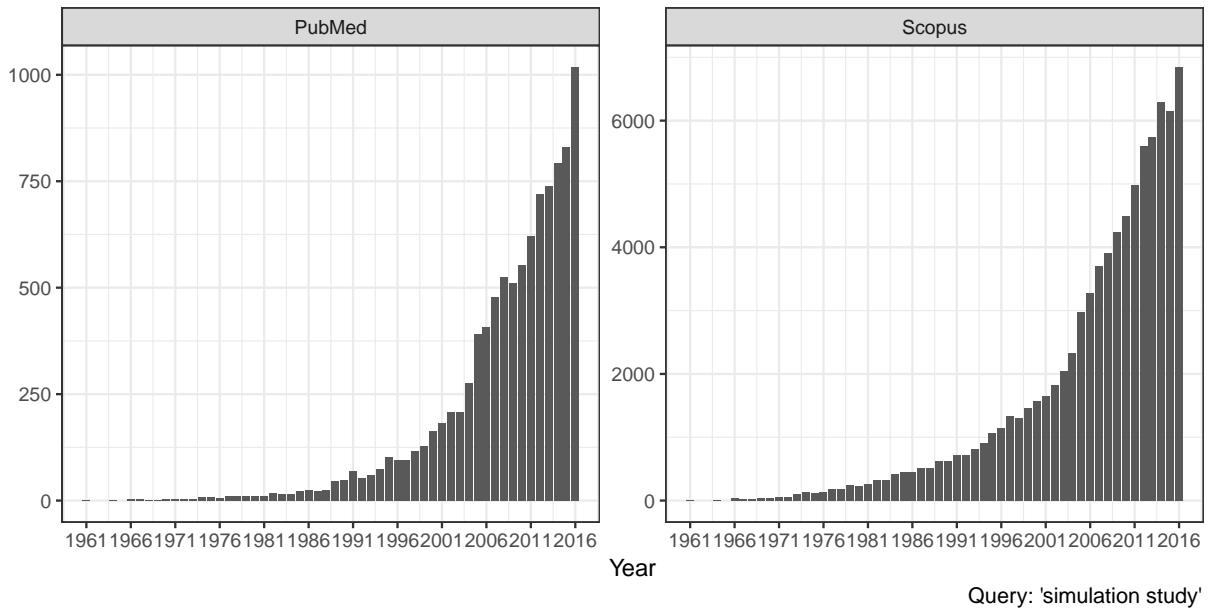


Figure 8.1: Numer of results querying 'simulation study' on Pubmed and Scopus.

Simulation studies represent a powerful tool with a multiplicity of aims: among others, evaluating new or existing statistical methods, comparing them, assessing the impact of modelling assumption violations, and helping with the understanding of statistical concepts. The increased availability of powerful computational tools (both personal and high-performance cluster computers) to the average researcher surely contributed to the rise of simulation studies in current literature. Searching on PubMed and Scopus with the query “simulation study” it is indeed possible to appreciate the greater use of this tool (Figure 8.1). Additionally, increased computational capabilities allow researchers to simulate an ever-growing number of scenarios, exploring multiple data-generating mechanisms, factors, and methods at once - making reporting results a non-trivial task.

It is necessary to bear in mind that dissemination of results plays a focal role in simulation studies:

1. it can drive practitioners and applied statisticians to methods that have been shown to perform well in their practical settings (e.g.: small sample size, high proportion of missing values);

2. it can guide researchers to develop new methods in a promising direction;
3. it can provide insights into less established methods.

As a consequence, several design and reporting guidelines emerged, often tailor-made to a specific research area (e.g. health technology assessment, medical statistics, social sciences). Despite that, challenges still persist and further research is needed into methods to help reporting of results.

To bridge the gap between the number of scenarios a researcher can simulate from and dissemination of results, I developed a tool for exploring results interactively. The tool is developed using R and shiny (<https://shiny.rstudio.com/>), a web application framework for R that allows creating interactive web applications in a straightforward way.

I named the interactive tool *SiReX*, acronym for *Simulation Results eXplorer*. It requires the researcher to upload a dataset in a standardised, tidy format (observations are in rows, variables are in columns) containing results from a simulation study. Then, it computes performance measures such as bias, coverage probability, Monte Carlo errors, and empirical standard errors automatically. Finally, it presents results and performance summaries both in tabular and graphical fashion (via bar plots and lolly plots) and allows the reader to vary simulation parameters and choose estimands of interest for further investigations.

A typical workflow with SiReX would consist in the following steps:

1. Upload a dataset with results from a simulation study in a tidy format compatible with the tool;
2. Summary statistics are computed automatically;
3. Factors identifying different data-generating mechanisms are identified automatically and drop-down menus are populated appropriately;
4. Now, it is possible to select and change data-generating mechanisms: summary tables

and plots are updated automatically;

5. Exporting summary statistics, tables, and plots for later use is supported.

A current demo of the tool is available at <https://ag475.shinyapps.io/sirex-demo/>, using an example dataset from a simulation study on multiple imputation.

# Chapter 9

## Informative visiting process

Healthcare consumption data is being used increasingly often in medical research, a successful example being the CArdiovascular disease research using LInked Bespoke studies and Electronic health Records (CALIBER) programme ([Denaxas et al., 2012](#)). CALIBER was constructed by extracting and linking electronic health records from primary care, hospital care, and nationwide registries (including information such as social deprivation and living status), and it consists in more than 10 million adults with 400 million person-years of follow-up; this vast amount of data allows researchers to answer more relevant and detailed clinical questions, but poses new methodological challenges.

First and foremost, in observational, healthcare consumption data observation times are likely to be correlated with the underlying disease severity. For instance, individuals tend to have irregular observation times as patients with more severe conditions (or showing early symptoms of a disease) tend to visit their GP or go to the hospital more often than those with milder conditions (and no symptoms). Their worse disease status is also likely to be reflected in worse biomarkers being recorded as such visits, causing abnormal values of such biomarkers to be overrepresented and normal values to be underrepresented. Additionally, for diseases with a high mortality rate, a terminal event that truncates observation of the

longitudinal process is likely to be informative in the sense that it likely correlates with disease severity. That is, dropout is likely to be informative as the tendency to dropout after the occurrence of a terminal event is related to the current level of the longitudinally recorded biomarker.

Traditional methods used to analyse longitudinal data rely on the assumption that the underlying mechanisms that controls the observation time is independent of disease severity; however, as I mentioned before, that is unlikely with healthcare consumption data. It can be showed that failing to account for informative dropout in a longitudinal study can yields biased estimates of the model parameters (Wu and Carroll, 1988). Analogously, an explicit assumption when jointly modelling longitudinal and survival data (using the framework presented in Chapter 3) is that the timing and number of measurements for each subject should be non-informative, i.e. it does not associate with the survival part of the model. However, it is currently unknown whether violations of this assumption lead to invalid inference or not in the context of joint models.

The topic of informative observation times and informative censoring has been often object of many recent investigations. Liu et al. (2008) developed a model that analyses repeated measures by taking into account informative observation times and an informative terminal event at the same time; they proposed a joint model formed by three submodels, a frailty model for the observation times, a mixed effects model for the longitudinal process, and a proportional hazards model for the terminal event. Ghosh et al. (2011) proposed a joint longitudinal and survival model that handles multiple changepoints in the longitudinal profile by including random spline coefficients; the survival part of the models handles an informative terminal event by using a semiparametric Cox model. Han et al. (2014) proposed a model for the joint distribution of the longitudinal process, the observation process, and the dropout process that uses, respectively, a semiparametric linear regression model for the longitudinal data and two accelerated failure time models for the observation and

dropout process; their model is semiparametric in the sense that it leaves the distributional form and the dependent structure unspecified. [Lesperance et al. \(2015\)](#) developed a joint model within the multi-state framework that handles examination times correlated with disease progression; they link transition intensities of a Markov model with a log-linear mixed model governing observation times through shared random effects. However, they do not integrate out the shared random effects from the model as their target of inference is the transition intensities conditional on the random effects. Analogously, there has been quite a lot of developments in the multi-state framework to handle informative visiting times and/or dropout. [Sweeting et al. \(2010\)](#) developed a model (similar to a hidden Markov model) in the setting of a response variable irregularly and infrequently observed by conditioning on regularly collected auxiliary data. [Lange et al. \(2015\)](#) generalise the work of [Sweeting et al. \(2010\)](#) to better fit the setting of observational data rather than clinical trial data with informative missingness or observation times by modelling the disease process with a Markov model and the observation process with a Poisson process that depends on the underlying disease status. [Lawless and Nazeri Rad \(2015\)](#) consider the effect of intermittent, irregular observation times on the estimation of Markov models; they show that it is not possible to estimate transition intensities in bi-directional Markov models with good precision when the gap between observation times is too big. They also show how the correlation between visit times and observed disease status can bias model assessment procedures, and propose an inverse intensity weighted estimation procedure for state prevalence. In brief, this approach consists in weighting each individual by their probability of being observed (or measured) at a given point in time; they discuss and develop this further in ([Nazeri Rad and Lawless, 2017](#)). Finally, [Li and Su \(2017\)](#) proposed a joint model for a longitudinal outcome and semi-competing risk data such as study dropout and death; they model the longitudinal process using a mixed model, and the semicompeting risks using two separate probit models.

In conclusion, joint models for longitudinal and survival data can handle effectively informative dropout in longitudinal study by modelling the longitudinal trajectory and the dropout

process jointly. However, it is not clear whether presence of an informative visiting process affects inference from joint models. Further investigating this topic will form a good part of work planned for the second year of my PhD, as I will discuss in Chapter 10. This project will have important practical implications, as it aims to provide guidance to practitioners and applied researchers using joint models for longitudinal and survival data with their observational data.

# Chapter 10

## Future research developments

In this Chapter I will present my plans for research during the second year of my PhD. I also included a Gantt chart in Figure 10.1.

First of all, I plan to conclude the simulation study on the impact of model misspecification in survival models with frailty terms. Simulations for some scenarios are currently still running on the University high-performance computing (HPC) facilities at the time of writing, and should finish soon. Consequently, I will summarise relevant results and I will finish writing up the project into a paper that will be submitted to a journal for publication by the end of the year. Current potential target journals are the Biometrical Journal and Statistics in Medicine.

At the same time, I will be planning the next project on modelling the visiting process and investigating the impact of an informative visiting process on inference using longitudinal data originating from healthcare records. This project aims to shed light on how an informative visiting process affects the analysis of longitudinal data that are intermittently and irregularly measured and recorded, in order to provide practical advice to applied researchers. In practice, I will be simulating complex survival data along with one or more longitudinal biomarkers under a variety of biologically plausible data-generating mechanisms; for instance,

I will vary:

1. the underlying risk of event, in terms of magnitude and shape of the baseline hazard;
2. the number and frequency of longitudinal measurements;
3. features of the longitudinal process such as functional form over time;
4. the strength of the informative observation process, i.e. the magnitude of the association between the underlying disease process and the observation process;
5. the shape of the association between the observation process and the disease process, that is, the parametrisation that links the two processes (e.g. current value parametrisation, intercept and slope, cumulative effect, etc.).

Then, I will compare different analytical approaches proposed in literature to tackle the problem, starting with simple approaches such as including the number of preceding measurements in the model as a proxy of disease severity moving onto more complex methods such as those introduced in Section 9. Some methods will be directly applicable using existing software, while others will require developing ad-hoc software if an existing implementation is not available in standard statistical software. Throughout this project we will be collaborating with Dr. Jessica Barrett from the MRC Biostatistic Unit in Cambridge to discuss factors that may affect the results of this study and what methods to include and compare. Once the planning phase is completed, I will code the simulation study and run it using the HPC facilities of the University. Finally, I plan to write up the project into a manuscript approximately at the end of my second year.

During my second year I will also work on an applied project in the area of cardiovascular epidemiology using joint models for multiple longitudinal biomarkers and survival data using CALIBER data ([Denaxas et al., 2012](#)). CALIBER includes a wide variety of biomarkers such as systolic and diastolic blood pressure, body-mass index, high- and low-density cholesterol, and so on. These biomarkers are likely to be correlated as they change over time, and they

may improve cardiovascular risk prediction and clear up the complex relationship between changes in the biomarkers and the risk of adverse events. Specifically, I will select a cohort of individuals with stable angina and type 2 diabetes (for whom regular monitoring of blood pressure is recommended) and evaluate the association between multiple, longitudinally measured biomarkers and the risk of adverse coronary events, fatal and non-fatal.

The applied project using CALIBER data will inform further projects, such as studying and developing discrimination and calibration tools to use with multiple longitudinal biomarkers. The importance of discrimination and classification tools is great, as it is fundamental to be able to discern whether the addition of longitudinal biomarkers improves predictions or not. This project will have wide reaching consequences, as it will be providing guidance in the use of joint models for longitudinal and survival data with multiple longitudinal biomarkers when the aim is prediction. I aim to work on this during the final year of my PhD.

Finally, an ongoing task will be the continuous development and expansion of the interactive tool for exploring results from simulation studies. Potential developments will include:

- polishing the underlying engine used to compute summary statistics;
- including more plots;
- allowing custom faceted plots and tables comparing multiple factors at once.

I aim to produce a polished version of SiReX to publish on-line to a wider audience as soon as possible.

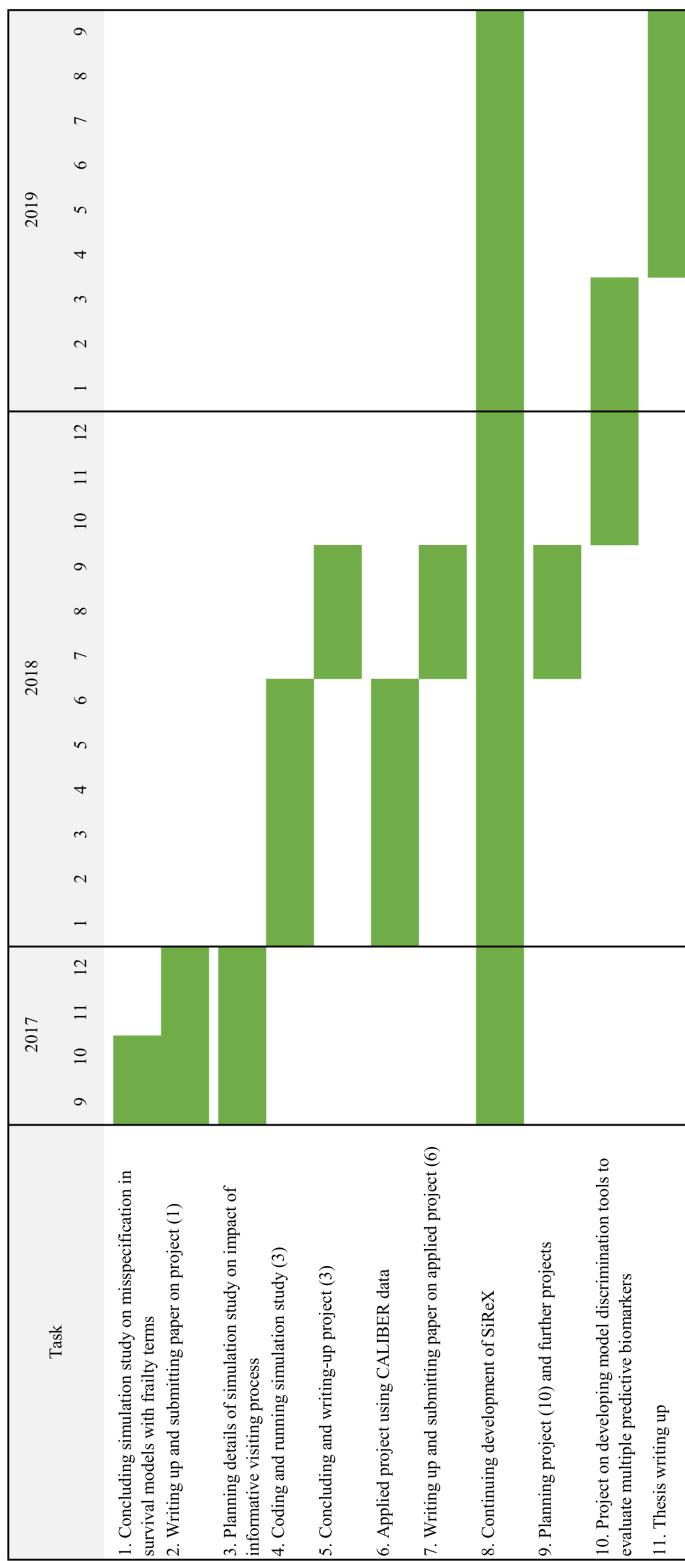


Figure 10.1: Gantt chart for current and future projects.

# **Chapter 11**

## **Personal development**

In this chapter I will introduce and briefly discuss the personal development activities I carried out during the first year of my PhD; specifically, I will present the supervisory meetings, training courses, and conferences I attended.

### **11.1 Supervisory meetings**

I have been having frequent meetings with my supervisors, formally and informally. Formal supervisory meetings, recorded on PROSE (<https://prose.le.ac.uk>), have been held on average every other week, with summaries of what was discussed produced and shared between us. A comprehensive list is available on PROSE. Additionally, we held more frequent informal meetings to discuss developments and more urgent matters, whenever it was necessary.

### **11.2 Training and courses**

I have attended a wide variety of courses during my first year, both externally and internally to the University of Leicester. The external courses I attended are:

- *Efficient R Programming*, on November 8<sup>th</sup> 2016, organised by the Royal Statistical Society in London. The instructor was Dr. Colin Gillespie, from the University of Newcastle, United Kingdom, and Jumping Rivers. The course covered how to program efficiently with R; in particular, it covered common pitfalls when writing R code, code profiling, interfacing with C++, and parallel programming. General hints and tips were provided.
- *Introduction to causal inference*, on April 25<sup>th</sup> and 26<sup>th</sup> 2017, organised by the Biostatistics Research Group at the University of Leicester and delivered by Dr. Arvid Sjölander from Karolinska Institutet, Stockholm, Sweden. The course provided foundational concepts of causal inference such as the difference between association and causation, the counterfactual framework, exchangeability, directed acyclic graphs, methods for estimating a causal effect, etc. Additionally, it provided an introduction to more advanced methods such as instrumental variables and Mendelian randomisation.
- *Using simulation studies to evaluate statistical methods*, on May 22<sup>nd</sup> 2017, organised by University College London. The course was delivered by Dr. Tim Morris, Prof. Ian White and Dr. Michael Crowther, and it covered the rationale for using simulation studies, important concepts to keep in mind when planning a simulation study, computational tools, estimates of uncertainty, and tools for improving reporting and dissemination.
- Workshop on *Joint modelling of longitudinal and time-to-event data with R*, on July 5<sup>th</sup>, 2017, organised by the Department of Biostatistics of the University of Liverpool. The course was delivered by Dr. Graeme Hickey, and provided an introduction to joint models of longitudinal and survival data, including extensions to incorporate competing risks and multiple longitudinal processes. The course included a practical session using R.

I have attended a few courses within the University and not offered on PROSE; specifically,

I attended a course on *Time series analysis with R* (November 10<sup>th</sup>, 2016), a course on *Data visualisation* (November 15<sup>th</sup>, 2016), and a course on *High performance computing at Leicester* (February 8<sup>th</sup>, 2017). The latter was particularly important, as it allowed me to take advantage of the high-performance computing facilities offered by the University more efficiently. I also attended the *Preparing to teach in higher education* workshop, strand A (July 24<sup>th</sup> and 27<sup>th</sup> 2017).

Finally, I attended the following PROSE training sessions to develop personal and communication skills in research settings:

- *Planning your literature search*, October 21<sup>st</sup> 2016;
- *Conducting your literature search*, October 25<sup>th</sup> 2016 ;
- *Assertiveness*, November 14<sup>th</sup> 2016;
- *Introduction to critical thinking*, December 15<sup>th</sup> 2016;
- *Presentations A: Fundamentals of an effective presentation*, January 30<sup>th</sup> 2017;
- *Communication in research and other work settings*, January 31<sup>st</sup> 2017;
- *Enhancing your digital profile*, February 2<sup>nd</sup> 2017;
- *Saying it with your abstract*, February 10<sup>th</sup> 2017;
- *Designing a poster*, February 27<sup>th</sup> 2017;
- *Leadership in research and other work environments*, February 28<sup>th</sup> 2017;
- *Preparing for the probation review (Physical natural and medical sciences)*, May 30<sup>th</sup> 2017.

### 11.3 Conferences

I attended a number of conferences during this year, in which I delivered the following oral presentations:

- Survival Analysis for Junior Researchers conference, held in Leicester, UK, on April 5<sup>th</sup> and 6<sup>th</sup> 2017. I delivered a talk titled *Direct likelihood maximisation using numerical quadrature to approximate intractable terms*;
- Statistical Analysis of Multi-Outcome Data (SAM) conference, held in Liverpool, UK, on July 3<sup>rd</sup> and 4<sup>th</sup> 2017. I delivered a talk titled *Impact of model misspecification in survival models with frailties*;
- Annual Conference of the International Society for Clinical Biostatistics conference, held in Vigo, Spain, on July 9<sup>th</sup> to July 13<sup>th</sup> 2017. I delivered two talks: a titled *Impact of model misspecification in survival models with frailties* during one of the contributed sessions on survival analysis, and a talk titled *Exploring results from simulation studies interactively* during the Students' Day organised on July 13<sup>th</sup>.

Additionally, I delivered an oral presentation on previous research work external to my PhD project during the 54<sup>th</sup> ERA-EDTA Congress held in Madrid, Spain, between June 3<sup>rd</sup> and June 6<sup>th</sup>. The ERA-EDTA Congress is the main conference in the field of Nephrology in Europe, with approximately 10,000 participants in 2017. I delivered my presentation, titled *Inappropriate prescription of nephrotoxic drugs to individuals with chronic kidney disease*, to an audience of clinicians, epidemiologists, clinical researchers, and other stakeholders.

# **Appendix A**

## **Tables**

Table A.1: Bias, comparison with analytical formulae, scenario with a small frailty variance and a negative regression coefficient.

Sample size	Method	Lambda	P	Beta	Theta
1000c. of 2i.	AF	0.002 (0.002)	0.001 (0.001)	-0.003 (0.002)	-0.017 (0.008)
1000c. of 2i.	IN	0.000 (0.002)	0.000 (0.001)	-0.003 (0.002)	-0.032 (0.008)
1000c. of 2i.	GQ15	0.011 (0.002)	0.005 (0.001)	-0.007 (0.002)	0.086 (0.003)
1000c. of 2i.	GQ35	0.002 (0.002)	0.001 (0.001)	-0.003 (0.002)	-0.011 (0.008)
1000c. of 2i.	GQ75	0.002 (0.002)	0.001 (0.001)	-0.003 (0.002)	-0.017 (0.008)
1000c. of 2i.	GQ105	0.002 (0.002)	0.001 (0.001)	-0.003 (0.002)	-0.017 (0.008)
15c. of 100i.	AF	-0.008 (0.004)	0.002 (0.001)	-0.001 (0.002)	-0.146 (0.013)
15c. of 100i.	IN	-0.007 (0.004)	0.002 (0.001)	-0.001 (0.002)	-0.146 (0.013)
15c. of 100i.	GQ15	-0.165 (0.008)	0.003 (0.001)	-0.002 (0.002)	0.223 (0.013)
15c. of 100i.	GQ35	-0.086 (0.006)	0.002 (0.001)	-0.001 (0.002)	-0.028 (0.013)
15c. of 100i.	GQ75	-0.040 (0.005)	0.001 (0.001)	-0.001 (0.002)	-0.114 (0.013)
15c. of 100i.	GQ105	-0.008 (0.005)	0.002 (0.001)	-0.001 (0.002)	-0.130 (0.013)
15c. of 30i.	AF	-0.010 (0.005)	0.002 (0.002)	-0.007 (0.004)	-0.215 (0.017)
15c. of 30i.	IN	-0.008 (0.005)	0.002 (0.002)	-0.008 (0.004)	-0.215 (0.017)
15c. of 30i.	GQ15	-0.053 (0.007)	0.003 (0.002)	-0.007 (0.004)	-0.025 (0.014)
15c. of 30i.	GQ35	-0.003 (0.005)	0.002 (0.002)	-0.008 (0.004)	-0.194 (0.021)
15c. of 30i.	GQ75	-0.010 (0.005)	0.002 (0.002)	-0.007 (0.004)	-0.216 (0.017)
15c. of 30i.	GQ105	-0.010 (0.005)	0.002 (0.002)	-0.007 (0.004)	-0.216 (0.017)
15c. of 500i.	AF	-0.011 (0.004)	0.000 (0.000)	0.001 (0.001)	-0.142 (0.012)
15c. of 500i.	IN	-0.049 (0.005)	0.000 (0.000)	0.001 (0.001)	-0.067 (0.013)
15c. of 500i.	GQ15	-0.225 (0.009)	0.001 (0.000)	0.001 (0.001)	0.317 (0.015)
15c. of 500i.	GQ35	-0.145 (0.006)	0.001 (0.000)	0.001 (0.001)	0.073 (0.013)
15c. of 500i.	GQ75	-0.109 (0.005)	0.000 (0.000)	0.001 (0.001)	-0.033 (0.012)
15c. of 500i.	GQ105	-0.101 (0.005)	0.000 (0.000)	0.001 (0.001)	-0.057 (0.012)
50c. of 100i.	AF	0.000 (0.002)	0.000 (0.001)	-0.002 (0.001)	-0.044 (0.007)
50c. of 100i.	IN	0.001 (0.002)	0.000 (0.001)	-0.002 (0.001)	-0.045 (0.007)
50c. of 100i.	GQ15	-0.207 (0.006)	0.001 (0.001)	-0.003 (0.001)	0.316 (0.009)
50c. of 100i.	GQ35	-0.088 (0.004)	0.000 (0.001)	-0.002 (0.001)	0.062 (0.007)
50c. of 100i.	GQ75	-0.020 (0.003)	0.000 (0.001)	-0.002 (0.001)	-0.025 (0.007)
50c. of 100i.	GQ105	-0.002 (0.003)	0.000 (0.001)	-0.002 (0.001)	-0.036 (0.007)
50c. of 30i.	AF	-0.002 (0.003)	0.000 (0.001)	-0.002 (0.002)	-0.046 (0.008)
50c. of 30i.	IN	-0.002 (0.003)	0.000 (0.001)	-0.004 (0.002)	-0.043 (0.008)
50c. of 30i.	GQ15	-0.045 (0.004)	0.001 (0.001)	-0.003 (0.002)	0.082 (0.007)
50c. of 30i.	GQ35	-0.002 (0.003)	0.000 (0.001)	-0.002 (0.002)	-0.038 (0.008)
50c. of 30i.	GQ75	-0.002 (0.003)	0.000 (0.001)	-0.002 (0.002)	-0.046 (0.008)
50c. of 30i.	GQ105	-0.002 (0.003)	0.000 (0.001)	-0.002 (0.002)	-0.046 (0.008)

Table A.2: Coverage, comparison with analytical formulae, scenario with a small frailty variance and a negative regression coefficient.

Sample size	Method	Lambda	P	Beta	Theta
1000c. of 2i.	AF	94.20 (0.74)	95.50 (0.66)	95.00 (0.69)	96.50 (0.58)
1000c. of 2i.	IN	93.42 (0.78)	95.70 (0.64)	93.02 (0.81)	77.45 (1.32)
1000c. of 2i.	GQ15	93.60 (0.77)	95.40 (0.66)	95.10 (0.68)	98.70 (0.36)
1000c. of 2i.	GQ35	94.29 (0.73)	95.10 (0.68)	95.30 (0.67)	95.90 (0.63)
1000c. of 2i.	GQ75	94.20 (0.74)	95.50 (0.66)	95.00 (0.69)	96.50 (0.58)
1000c. of 2i.	GQ105	94.20 (0.74)	95.50 (0.66)	95.00 (0.69)	96.50 (0.58)
15c. of 100i.	AF	92.40 (0.84)	95.50 (0.66)	94.70 (0.71)	92.60 (0.83)
15c. of 100i.	IN	90.65 (0.92)	95.48 (0.66)	94.45 (0.72)	92.39 (0.84)
15c. of 100i.	GQ15	27.80 (1.42)	94.60 (0.71)	94.90 (0.70)	86.70 (1.07)
15c. of 100i.	GQ35	44.04 (1.57)	94.79 (0.70)	94.89 (0.70)	93.99 (0.75)
15c. of 100i.	GQ75	66.00 (1.50)	95.10 (0.68)	95.10 (0.68)	93.00 (0.81)
15c. of 100i.	GQ105	76.20 (1.35)	95.50 (0.66)	94.80 (0.70)	92.50 (0.83)
15c. of 30i.	AF	92.70 (0.82)	95.00 (0.69)	94.70 (0.71)	94.50 (0.72)
15c. of 30i.	IN	92.21 (0.85)	94.91 (0.69)	94.08 (0.75)	93.15 (0.80)
15c. of 30i.	GQ15	68.10 (1.47)	94.70 (0.71)	94.70 (0.71)	93.10 (0.80)
15c. of 30i.	GQ35	88.91 (0.99)	94.61 (0.71)	94.20 (0.74)	95.23 (0.67)
15c. of 30i.	GQ75	92.50 (0.83)	95.00 (0.69)	94.70 (0.71)	94.10 (0.75)
15c. of 30i.	GQ105	92.70 (0.82)	95.00 (0.69)	94.60 (0.71)	94.60 (0.71)
15c. of 500i.	AF	92.50 (0.83)	95.40 (0.66)	95.70 (0.64)	91.60 (0.88)
15c. of 500i.	IN	31.74 (1.47)	95.02 (0.69)	95.63 (0.65)	91.25 (0.89)
15c. of 500i.	GQ15	11.00 (0.99)	92.50 (0.83)	94.90 (0.70)	80.40 (1.26)
15c. of 500i.	GQ35	14.20 (1.10)	93.80 (0.76)	95.70 (0.64)	91.50 (0.88)
15c. of 500i.	GQ75	18.40 (1.23)	94.80 (0.70)	95.90 (0.63)	93.40 (0.79)
15c. of 500i.	GQ105	22.00 (1.31)	95.20 (0.68)	95.40 (0.66)	93.20 (0.80)
50c. of 100i.	AF	95.40 (0.66)	94.90 (0.70)	94.30 (0.73)	93.30 (0.79)
50c. of 100i.	IN	93.07 (0.80)	94.56 (0.72)	93.82 (0.76)	91.68 (0.87)
50c. of 100i.	GQ15	17.20 (1.19)	94.30 (0.73)	94.70 (0.71)	60.60 (1.55)
50c. of 100i.	GQ35	37.80 (1.53)	94.60 (0.71)	94.30 (0.73)	91.50 (0.88)
50c. of 100i.	GQ75	69.80 (1.45)	94.80 (0.70)	94.50 (0.72)	93.70 (0.77)
50c. of 100i.	GQ105	82.90 (1.19)	94.90 (0.70)	94.20 (0.74)	93.60 (0.77)
50c. of 30i.	AF	93.80 (0.76)	95.30 (0.67)	95.30 (0.67)	95.40 (0.66)
50c. of 30i.	IN	92.65 (0.83)	95.28 (0.67)	94.73 (0.71)	93.30 (0.79)
50c. of 30i.	GQ15	71.90 (1.42)	95.00 (0.69)	95.10 (0.68)	92.60 (0.83)
50c. of 30i.	GQ35	92.00 (0.86)	95.10 (0.68)	95.10 (0.68)	93.90 (0.76)
50c. of 30i.	GQ75	93.80 (0.76)	95.40 (0.66)	95.20 (0.68)	95.40 (0.66)
50c. of 30i.	GQ105	93.70 (0.77)	95.30 (0.67)	95.30 (0.67)	95.40 (0.66)

Table A.3: Mean squared error, comparison with analytical formulae, scenario with a small frailty variance and a negative regression coefficient.

Sample size	Method	Lambda	P	Beta	Theta
1000c. of 2i.	AF	0.0026	0.0008	0.0042	0.0670
1000c. of 2i.	IN	0.0024	0.0008	0.0041	0.0635
1000c. of 2i.	GQ15	0.0024	0.0007	0.0042	0.0159
1000c. of 2i.	GQ35	0.0025	0.0008	0.0042	0.0571
1000c. of 2i.	GQ75	0.0026	0.0008	0.0042	0.0668
1000c. of 2i.	GQ105	0.0026	0.0008	0.0042	0.0669
15c. of 100i.	AF	0.0190	0.0009	0.0045	0.1886
15c. of 100i.	IN	0.0193	0.0009	0.0046	0.1880
15c. of 100i.	GQ15	0.0867	0.0009	0.0046	0.2116
15c. of 100i.	GQ35	0.0432	0.0009	0.0046	0.1621
15c. of 100i.	GQ75	0.0274	0.0009	0.0045	0.1816
15c. of 100i.	GQ105	0.0242	0.0009	0.0045	0.1877
15c. of 30i.	AF	0.0262	0.0029	0.0157	0.3341
15c. of 30i.	IN	0.0259	0.0029	0.0158	0.3353
15c. of 30i.	GQ15	0.0517	0.0029	0.0161	0.2038
15c. of 30i.	GQ35	0.0283	0.0029	0.0159	0.4801
15c. of 30i.	GQ75	0.0263	0.0029	0.0157	0.3365
15c. of 30i.	GQ105	0.0262	0.0029	0.0157	0.3414
15c. of 500i.	AF	0.0169	0.0002	0.0009	0.1691
15c. of 500i.	IN	0.0284	0.0002	0.0009	0.1835
15c. of 500i.	GQ15	0.1315	0.0002	0.0009	0.3364
15c. of 500i.	GQ35	0.0599	0.0002	0.0009	0.1638
15c. of 500i.	GQ75	0.0394	0.0002	0.0009	0.1486
15c. of 500i.	GQ105	0.0345	0.0002	0.0009	0.1477
50c. of 100i.	AF	0.0053	0.0003	0.0014	0.0487
50c. of 100i.	IN	0.0054	0.0003	0.0014	0.0488
50c. of 100i.	GQ15	0.0816	0.0003	0.0014	0.1827
50c. of 100i.	GQ35	0.0242	0.0003	0.0014	0.0555
50c. of 100i.	GQ75	0.0095	0.0003	0.0014	0.0480
50c. of 100i.	GQ105	0.0076	0.0003	0.0014	0.0491
50c. of 30i.	AF	0.0073	0.0009	0.0047	0.0628
50c. of 30i.	IN	0.0074	0.0009	0.0047	0.0629
50c. of 30i.	GQ15	0.0180	0.0009	0.0048	0.0560
50c. of 30i.	GQ35	0.0082	0.0009	0.0047	0.0647
50c. of 30i.	GQ75	0.0073	0.0009	0.0047	0.0627
50c. of 30i.	GQ105	0.0073	0.0009	0.0047	0.0628

Table A.4: Convergence rates, comparison with analytical formulae, scenario with a small frailty variance and a negative regression coefficient.

Sample size	AF	IN	GQ15	GQ35	GQ75	GQ105
1000c. of 2i.	100.0%	74.5%	100.0%	99.9%	100.0%	100.0%
15c. of 100i.	100.0%	97.3%	100.0%	99.9%	100.0%	100.0%
15c. of 30i.	100.0%	96.3%	100.0%	96.5%	100.0%	100.0%
15c. of 500i.	100.0%	98.3%	100.0%	100.0%	100.0%	100.0%
50c. of 100i.	100.0%	93.8%	100.0%	100.0%	100.0%	100.0%
50c. of 30i.	100.0%	91.1%	100.0%	100.0%	100.0%	100.0%

Table A.5: Bias, comparison without analytical formulae, scenario with a small frailty variance and a negative regression coefficient.

Sample size	Method	Lambda	P	Beta	Sigma
1000c. of 2i.	GQ15	0.0187 (0.0012)	-0.0415 (0.0008)	0.0149 (0.0019)	-0.3010 (0.0250)
1000c. of 2i.	GQ35	0.0188 (0.0012)	-0.0415 (0.0008)	0.0144 (0.0019)	-0.2842 (0.0232)
1000c. of 2i.	GQ75	0.0188 (0.0012)	-0.0415 (0.0008)	0.0146 (0.0019)	-0.3057 (0.0254)
1000c. of 2i.	GQ105	0.0187 (0.0012)	-0.0415 (0.0008)	0.0147 (0.0019)	-0.2946 (0.0243)
15c. of 100i.	GQ15	0.0102 (0.0040)	-0.0332 (0.0010)	0.0552 (0.0040)	0.0282 (0.0113)
15c. of 100i.	GQ35	0.0102 (0.0040)	-0.0333 (0.0010)	0.0173 (0.0024)	-0.0854 (0.0085)
15c. of 100i.	GQ75	0.0102 (0.0040)	-0.0333 (0.0010)	0.0088 (0.0022)	-0.1015 (0.0075)
15c. of 100i.	GQ105	0.0102 (0.0040)	-0.0333 (0.0010)	0.0088 (0.0022)	-0.1015 (0.0075)
15c. of 30i.	GQ15	0.0193 (0.0045)	-0.0348 (0.0017)	0.0047 (0.0039)	-0.2510 (0.0272)
15c. of 30i.	GQ35	0.0192 (0.0045)	-0.0348 (0.0017)	0.0041 (0.0039)	-0.2455 (0.0256)
15c. of 30i.	GQ75	0.0191 (0.0045)	-0.0348 (0.0017)	0.0041 (0.0039)	-0.2593 (0.0275)
15c. of 30i.	GQ105	0.0195 (0.0045)	-0.0347 (0.0017)	0.0040 (0.0039)	-0.2522 (0.0265)
15c. of 500i.	GQ15	0.0113 (0.0037)	-0.0411 (0.0006)	0.0867 (0.0050)	0.2427 (0.0104)
15c. of 500i.	GQ35	0.0157 (0.0038)	-0.0423 (0.0006)	0.0860 (0.0040)	0.4011 (0.0091)
15c. of 500i.	GQ75	0.0078 (0.0036)	-0.0405 (0.0005)	0.0707 (0.0033)	0.4020 (0.0112)
15c. of 500i.	GQ105	0.0059 (0.0036)	-0.0410 (0.0006)	0.0600 (0.0028)	0.3119 (0.0125)
50c. of 100i.	GQ15	0.0135 (0.0022)	-0.0365 (0.0005)	0.0244 (0.0020)	-0.0095 (0.0049)
50c. of 100i.	GQ35	0.0135 (0.0022)	-0.0365 (0.0005)	0.0102 (0.0012)	-0.0350 (0.0038)
50c. of 100i.	GQ75	0.0135 (0.0022)	-0.0365 (0.0005)	0.0105 (0.0012)	-0.0350 (0.0037)
50c. of 100i.	GQ105	0.0135 (0.0022)	-0.0365 (0.0005)	0.0105 (0.0012)	-0.0350 (0.0037)
50c. of 30i.	GQ15	0.0154 (0.0025)	-0.0367 (0.0009)	0.0079 (0.0021)	-0.0535 (0.0052)
50c. of 30i.	GQ35	0.0154 (0.0025)	-0.0367 (0.0009)	0.0079 (0.0021)	-0.0536 (0.0052)
50c. of 30i.	GQ75	0.0154 (0.0025)	-0.0367 (0.0009)	0.0079 (0.0021)	-0.0536 (0.0052)
50c. of 30i.	GQ105	0.0154 (0.0025)	-0.0367 (0.0009)	0.0079 (0.0021)	-0.0536 (0.0052)

Table A.6: Coverage, comparison without analytical formulae, scenario with a small frailty variance and a negative regression coefficient.

Sample size	Method	Lambda	P	Beta	Sigma
1000c. of 2i.	GQ15	92.5662 (0.8295)	64.8676 (1.5096)	94.5010 (0.7209)	99.0835 (0.3013)
1000c. of 2i.	GQ35	92.4413 (0.8359)	64.7600 (1.5107)	94.6885 (0.7092)	99.0807 (0.3018)
1000c. of 2i.	GQ75	92.4720 (0.8343)	64.6999 (1.5113)	94.6083 (0.7142)	99.0844 (0.3012)
1000c. of 2i.	GQ105	92.5586 (0.8299)	64.8318 (1.5100)	94.5973 (0.7149)	99.0826 (0.3015)
15c. of 100i.	GQ15	52.1000 (1.5797)	76.4000 (1.3428)	86.0000 (1.0973)	70.2000 (1.4464)
15c. of 100i.	GQ35	52.1000 (1.5797)	76.5000 (1.3408)	98.8000 (0.3443)	88.9000 (0.9934)
15c. of 100i.	GQ75	52.1000 (1.5797)	76.4000 (1.3428)	99.9000 (0.0999)	94.0000 (0.7510)
15c. of 100i.	GQ105	52.1000 (1.5797)	76.4000 (1.3428)	99.9000 (0.0999)	94.1000 (0.7451)
15c. of 30i.	GQ15	75.4016 (1.3619)	90.0602 (0.9461)	99.0964 (0.2992)	98.3936 (0.3976)
15c. of 30i.	GQ35	75.3769 (1.3624)	90.2513 (0.9380)	99.0955 (0.2994)	99.1960 (0.2824)
15c. of 30i.	GQ75	75.4263 (1.3614)	90.0702 (0.9457)	99.0973 (0.2991)	99.2979 (0.2640)
15c. of 30i.	GQ105	75.5020 (1.3600)	90.0602 (0.9461)	99.0964 (0.2992)	99.2972 (0.2642)
15c. of 500i.	GQ15	30.0813 (1.4503)	21.3415 (1.2956)	27.6423 (1.4143)	23.1707 (1.3342)
15c. of 500i.	GQ35	27.8826 (1.4180)	17.8197 (1.2101)	37.1069 (1.5277)	15.7233 (1.1511)
15c. of 500i.	GQ75	30.1053 (1.4506)	18.9474 (1.2392)	55.5789 (1.5713)	16.6316 (1.1775)
15c. of 500i.	GQ105	29.1028 (1.4364)	18.3807 (1.2248)	67.8337 (1.4771)	32.1663 (1.4771)
50c. of 100i.	GQ15	51.9000 (1.5800)	37.8000 (1.5333)	94.8000 (0.7021)	85.6000 (1.1102)
50c. of 100i.	GQ35	51.9000 (1.5800)	37.5000 (1.5309)	100.0000 (0.0000)	95.0000 (0.6892)
50c. of 100i.	GQ75	51.9000 (1.5800)	37.5000 (1.5309)	100.0000 (0.0000)	94.9000 (0.6957)
50c. of 100i.	GQ105	51.9000 (1.5800)	37.5000 (1.5309)	100.0000 (0.0000)	94.9000 (0.6957)
50c. of 30i.	GQ15	73.9000 (1.3888)	74.1000 (1.3853)	99.4000 (0.2442)	97.2000 (0.5217)
50c. of 30i.	GQ35	73.9000 (1.3888)	74.1000 (1.3853)	99.4000 (0.2442)	97.2000 (0.5217)
50c. of 30i.	GQ75	73.9000 (1.3888)	74.1000 (1.3853)	99.4000 (0.2442)	97.2000 (0.5217)
50c. of 30i.	GQ105	73.9000 (1.3888)	74.1000 (1.3853)	99.4000 (0.2442)	97.2000 (0.5217)

Table A.7: Mean squared error, comparison without analytical formulae, scenario with a small frailty variance and a negative regression coefficient.

Sample size	Method	Lambda	P	Beta	Sigma
1000c. of 2i.	GQ15	0.0019	0.0024	0.0040	0.7180
1000c. of 2i.	GQ35	0.0019	0.0024	0.0039	0.6198
1000c. of 2i.	GQ75	0.0019	0.0024	0.0039	0.7394
1000c. of 2i.	GQ105	0.0019	0.0024	0.0039	0.6752
15c. of 100i.	GQ15	0.0164	0.0021	0.0189	0.1285
15c. of 100i.	GQ35	0.0164	0.0021	0.0063	0.0790
15c. of 100i.	GQ75	0.0164	0.0022	0.0048	0.0665
15c. of 100i.	GQ105	0.0164	0.0022	0.0048	0.0665
15c. of 30i.	GQ15	0.0208	0.0041	0.0152	0.8029
15c. of 30i.	GQ35	0.0208	0.0040	0.0149	0.7164
15c. of 30i.	GQ75	0.0208	0.0041	0.0149	0.8228
15c. of 30i.	GQ105	0.0207	0.0041	0.0149	0.7636
15c. of 500i.	GQ15	0.0141	0.0020	0.0326	0.1666
15c. of 500i.	GQ35	0.0146	0.0021	0.0234	0.2438
15c. of 500i.	GQ75	0.0131	0.0019	0.0160	0.2881
15c. of 500i.	GQ105	0.0129	0.0020	0.0113	0.2542
50c. of 100i.	GQ15	0.0048	0.0016	0.0046	0.0237
50c. of 100i.	GQ35	0.0048	0.0016	0.0015	0.0153
50c. of 100i.	GQ75	0.0048	0.0016	0.0015	0.0153
50c. of 100i.	GQ105	0.0048	0.0016	0.0015	0.0153
50c. of 30i.	GQ15	0.0066	0.0022	0.0043	0.0301
50c. of 30i.	GQ35	0.0066	0.0022	0.0043	0.0300
50c. of 30i.	GQ75	0.0066	0.0022	0.0043	0.0300
50c. of 30i.	GQ105	0.0066	0.0022	0.0043	0.0300

Table A.8: Convergence rates, comparison without analytical formulae, scenario with a small frailty variance and a negative regression coefficient.

Sample size	GQ15	GQ35	GQ75	GQ105
1000c. of 2i.	98.2%	97.9%	98.3%	98.1%
15c. of 100i.	100.0%	100.0%	100.0%	100.0%
15c. of 30i.	99.6%	99.5%	99.7%	99.6%
15c. of 500i.	49.2%	47.7%	47.5%	45.7%
50c. of 100i.	100.0%	100.0%	100.0%	100.0%
50c. of 30i.	100.0%	100.0%	100.0%	100.0%

Table A.9: Convergence rates, simulation study on model misspecification in survival models with shared frailty terms, scenario with 15 clusters of 100 individuals each and a small frailty variance.

True frailty	Model frailty	Model baseline	Exponential	Weibull	Gompertz	Weibull-Weibull (1)	Weibull-Weibull (2)
Gamma	Gamma	Cox	34.8%	47.0%	10.5%	1.5%	4.5%
Gamma	Gamma	Exp	100.0%	100.0%	100.0%	100.0%	100.0%
Gamma	Gamma	Wei	100.0%	100.0%	100.0%	100.0%	97.9%
Gamma	Gamma	Gom	57.8%	37.6%	100.0%	96.8%	39.1%
Gamma	Gamma	RP(5)	99.8%	65.5%	99.8%	99.5%	22.0%
Gamma	Gamma	RP(9)	99.8%	65.5%	100.0%	99.2%	22.0%
Gamma	Gamma	RP(P)	99.4%	65.5%	99.9%	97.6%	22.0%
Log-Normal	Gamma	Cox	44.9%	61.4%	17.7%	4.1%	10.8%
Log-Normal	Gamma	Exp	100.0%	100.0%	100.0%	100.0%	100.0%
Log-Normal	Gamma	Wei	100.0%	100.0%	100.0%	100.0%	97.8%
Log-Normal	Gamma	Gom	59.3%	37.7%	100.0%	99.6%	38.4%
Log-Normal	Gamma	RP(5)	99.9%	65.8%	99.8%	99.6%	23.4%
Log-Normal	Gamma	RP(9)	99.9%	65.8%	99.7%	99.0%	23.4%
Log-Normal	Gamma	RP(P)	99.9%	65.8%	99.9%	98.4%	23.4%
Gamma	Log-Normal	Cox	100.0%	100.0%	100.0%	100.0%	100.0%
Gamma	Log-Normal	Exp	99.8%	99.9%	100.0%	100.0%	99.9%
Gamma	Log-Normal	Wei	99.8%	99.9%	99.9%	99.7%	97.7%
Gamma	Log-Normal	Gom	59.1%	38.2%	99.8%	95.0%	38.2%
Gamma	Log-Normal	RP(5)	99.8%	65.5%	100.0%	100.0%	22.0%
Gamma	Log-Normal	RP(9)	99.8%	65.5%	100.0%	100.0%	22.0%
Gamma	Log-Normal	RP(P)	99.8%	65.5%	100.0%	100.0%	22.0%
Log-Normal	Log-Normal	Cox	100.0%	100.0%	100.0%	100.0%	100.0%
Log-Normal	Log-Normal	Exp	99.9%	99.9%	100.0%	100.0%	99.6%
Log-Normal	Log-Normal	Wei	99.7%	99.9%	99.9%	99.7%	97.7%
Log-Normal	Log-Normal	Gom	60.9%	38.4%	99.8%	99.5%	39.6%
Log-Normal	Log-Normal	RP(5)	99.9%	65.8%	99.8%	99.9%	23.4%
Log-Normal	Log-Normal	RP(9)	99.9%	65.8%	99.8%	99.9%	23.4%
Log-Normal	Log-Normal	RP(P)	99.9%	65.8%	99.9%	100.0%	23.4%

Table A.10: Bias with Monte Carlo standard error of estimated regression coefficient, simulation study on model misspecification in survival models with shared frailty terms, scenario with 15 clusters of 100 individuals each and a small frailty variance.

True frailty	Model frailty	Model baseline	Exponential	Weibull	Gompertz	Weibull-Weibull (1)	Weibull-Weibull (2)
Gamma	Gamma	Cox	0.001 (0.002)	0.007 (0.002)	0.009 (0.002)	0.013 (0.002)	-0.007 (0.002)
Gamma	Gamma	Exp	-0.002 (0.002)	-0.083 (0.003)	0.133 (0.002)	0.102 (0.002)	-0.116 (0.002)
Gamma	Gamma	Wei	-0.002 (0.002)	0.002 (0.002)	0.034 (0.002)	-0.067 (0.002)	0.000 (0.002)
Gamma	Gamma	Gom	-0.006 (0.002)	-0.086 (0.003)	-0.002 (0.002)	0.028 (0.002)	-0.114 (0.002)
Gamma	Gamma	RP(5)	-0.002 (0.002)	-0.001 (0.002)	-0.002 (0.002)	-0.002 (0.002)	-0.001 (0.002)
Gamma	Gamma	RP(9)	-0.002 (0.002)	-0.001 (0.002)	-0.002 (0.002)	-0.001 (0.002)	-0.002 (0.002)
Gamma	Gamma	RP(P)	-0.002 (0.002)	-0.001 (0.002)	0.003 (0.002)	0.000 (0.002)	-0.000 (0.002)
Log-Normal	Gamma	Cox	0.004 (0.002)	-0.002 (0.002)	0.000 (0.002)	-0.006 (0.002)	0.002 (0.002)
Log-Normal	Gamma	Exp	0.001 (0.002)	-0.085 (0.002)	0.136 (0.001)	0.100 (0.002)	-0.115 (0.002)
Log-Normal	Gamma	Wei	0.001 (0.002)	-0.002 (0.002)	0.038 (0.002)	-0.066 (0.002)	0.001 (0.002)
Log-Normal	Gamma	Gom	0.001 (0.002)	-0.085 (0.002)	0.003 (0.002)	0.022 (0.002)	-0.115 (0.002)
Log-Normal	Gamma	RP(5)	0.002 (0.002)	-0.002 (0.002)	0.003 (0.002)	0.001 (0.002)	0.008 (0.002)
Log-Normal	Gamma	RP(9)	0.002 (0.002)	-0.002 (0.002)	0.003 (0.002)	0.002 (0.002)	0.007 (0.002)
Log-Normal	Gamma	RP(P)	0.002 (0.002)	-0.002 (0.002)	0.008 (0.002)	0.003 (0.002)	0.009 (0.002)
Gamma	Log-Normal	Cox	-0.002 (0.002)	0.001 (0.002)	-0.002 (0.002)	-0.001 (0.002)	0.001 (0.002)
Gamma	Log-Normal	Exp	-0.002 (0.002)	-0.083 (0.003)	0.133 (0.002)	0.102 (0.002)	-0.116 (0.002)
Gamma	Log-Normal	Wei	-0.002 (0.002)	0.001 (0.002)	0.034 (0.002)	-0.067 (0.002)	0.000 (0.002)
Gamma	Log-Normal	Gom	-0.005 (0.002)	-0.084 (0.003)	-0.002 (0.002)	0.027 (0.002)	-0.115 (0.002)
Gamma	Log-Normal	RP(5)	-0.002 (0.002)	-0.001 (0.002)	-0.002 (0.002)	-0.002 (0.002)	-0.001 (0.002)
Gamma	Log-Normal	RP(9)	-0.002 (0.002)	-0.001 (0.002)	-0.002 (0.002)	-0.001 (0.002)	-0.002 (0.002)
Gamma	Log-Normal	RP(P)	-0.002 (0.002)	-0.001 (0.002)	0.003 (0.002)	0.000 (0.002)	-0.000 (0.002)
Log-Normal	Log-Normal	Cox	0.001 (0.002)	-0.002 (0.002)	0.003 (0.002)	0.002 (0.002)	0.002 (0.002)
Log-Normal	Log-Normal	Exp	0.001 (0.002)	-0.086 (0.002)	0.136 (0.001)	0.100 (0.002)	-0.115 (0.002)
Log-Normal	Log-Normal	Wei	0.001 (0.002)	-0.002 (0.002)	0.038 (0.002)	-0.066 (0.002)	0.001 (0.002)
Log-Normal	Log-Normal	Gom	0.001 (0.002)	-0.084 (0.002)	0.003 (0.002)	0.022 (0.002)	-0.114 (0.002)
Log-Normal	Log-Normal	RP(5)	0.001 (0.002)	-0.002 (0.002)	0.003 (0.002)	0.001 (0.002)	0.008 (0.002)
Log-Normal	Log-Normal	RP(9)	0.001 (0.002)	-0.002 (0.002)	0.003 (0.002)	0.002 (0.002)	0.007 (0.002)
Log-Normal	Log-Normal	RP(P)	0.001 (0.002)	-0.002 (0.002)	0.007 (0.002)	0.003 (0.002)	0.009 (0.002)

Table A.11: Coverage with Monte Carlo standard error of estimated regression coefficient, simulation study on model misspecification in survival models with shared frailty terms, scenario with 15 clusters of 100 individuals each and a small frailty variance.

True frailty	Model frailty	Model baseline	Exponential	Weibull	Gompertz	Weibull-Weibull (1)	Weibull-Weibull (2)
Gamma	Gamma	Cox	93.68 (0.77)	96.17 (0.61)	93.33 (0.79)	93.33 (0.79)	93.33 (0.79)
Gamma	Gamma	Exp	95.20 (0.68)	72.60 (1.41)	35.30 (1.51)	53.80 (1.58)	49.20 (1.58)
Gamma	Gamma	Wei	95.20 (0.68)	95.50 (0.66)	91.70 (0.87)	72.80 (1.41)	94.59 (0.72)
Gamma	Gamma	Gom	95.67 (0.64)	70.21 (1.45)	94.80 (0.70)	89.98 (0.95)	49.87 (1.58)
Gamma	Gamma	RP(5)	95.09 (0.68)	95.27 (0.67)	94.89 (0.70)	94.27 (0.73)	93.64 (0.77)
Gamma	Gamma	RP(9)	95.09 (0.68)	95.27 (0.67)	95.00 (0.69)	94.35 (0.73)	93.64 (0.77)
Gamma	Gamma	RP(P)	95.17 (0.68)	95.42 (0.66)	94.59 (0.72)	94.36 (0.73)	93.64 (0.77)
Log-Normal	Gamma	Cox	95.55 (0.65)	94.79 (0.70)	97.18 (0.52)	92.68 (0.82)	92.59 (0.83)
Log-Normal	Gamma	Exp	95.30 (0.67)	74.80 (1.37)	35.40 (1.51)	53.30 (1.58)	47.50 (1.58)
Log-Normal	Gamma	Wei	95.50 (0.66)	95.00 (0.69)	91.50 (0.88)	74.10 (1.39)	94.68 (0.71)
Log-Normal	Gamma	Gom	95.28 (0.67)	76.13 (1.35)	95.40 (0.66)	93.17 (0.80)	49.22 (1.58)
Log-Normal	Gamma	RP(5)	95.30 (0.67)	96.20 (0.60)	95.49 (0.66)	95.38 (0.66)	95.73 (0.64)
Log-Normal	Gamma	RP(9)	95.30 (0.67)	96.35 (0.59)	95.39 (0.66)	95.05 (0.69)	95.73 (0.64)
Log-Normal	Gamma	RP(P)	95.60 (0.65)	96.20 (0.60)	95.60 (0.65)	94.92 (0.69)	95.30 (0.67)
Gamma	Log-Normal	Cox	95.20 (0.68)	95.30 (0.67)	95.00 (0.69)	94.30 (0.73)	94.70 (0.71)
Gamma	Log-Normal	Exp	95.19 (0.68)	72.37 (1.41)	35.30 (1.51)	53.80 (1.58)	48.95 (1.58)
Gamma	Log-Normal	Wei	95.19 (0.68)	95.50 (0.66)	91.69 (0.87)	73.22 (1.40)	94.68 (0.71)
Gamma	Log-Normal	Gom	95.94 (0.62)	71.47 (1.43)	94.99 (0.69)	90.42 (0.93)	49.48 (1.58)
Gamma	Log-Normal	RP(5)	95.09 (0.68)	95.27 (0.67)	94.90 (0.70)	94.20 (0.74)	93.64 (0.77)
Gamma	Log-Normal	RP(9)	95.09 (0.68)	95.27 (0.67)	94.90 (0.70)	94.30 (0.73)	93.64 (0.77)
Gamma	Log-Normal	RP(P)	95.09 (0.68)	95.42 (0.66)	94.60 (0.71)	94.50 (0.72)	93.64 (0.77)
Log-Normal	Log-Normal	Cox	95.20 (0.68)	95.40 (0.66)	95.30 (0.67)	95.10 (0.68)	94.80 (0.70)
Log-Normal	Log-Normal	Exp	95.30 (0.67)	74.87 (1.37)	35.40 (1.51)	53.20 (1.58)	47.59 (1.58)
Log-Normal	Log-Normal	Wei	95.49 (0.66)	95.10 (0.68)	91.59 (0.88)	73.92 (1.39)	94.78 (0.70)
Log-Normal	Log-Normal	Gom	94.75 (0.71)	76.04 (1.35)	95.39 (0.66)	93.17 (0.80)	48.99 (1.58)
Log-Normal	Log-Normal	RP(5)	95.20 (0.68)	96.50 (0.58)	95.49 (0.66)	95.40 (0.66)	95.73 (0.64)
Log-Normal	Log-Normal	RP(9)	95.30 (0.67)	96.35 (0.59)	95.39 (0.66)	95.10 (0.68)	95.73 (0.64)
Log-Normal	Log-Normal	RP(P)	95.40 (0.66)	96.05 (0.62)	95.70 (0.64)	94.90 (0.70)	95.73 (0.64)

Table A.12: Mean squared error of estimated regression coefficient, simulation study on model misspecification in survival models with shared frailty terms, scenario with 15 clusters of 100 individuals each and a small frailty variance.

True frailty	Model frailty	Model baseline	Exponential	Weibull	Gompertz	Weibull-Weibull (1)	Weibull-Weibull (2)
Gamma	Gamma	Cox	0.005	0.005	0.004	0.005	0.004
Gamma	Gamma	Exp	0.004	0.013	0.020	0.013	0.019
Gamma	Gamma	Wei	0.004	0.005	0.005	0.009	0.003
Gamma	Gamma	Gom	0.004	0.014	0.004	0.004	0.018
Gamma	Gamma	RP(5)	0.004	0.004	0.004	0.003	0.004
Gamma	Gamma	RP(9)	0.004	0.004	0.004	0.003	0.004
Gamma	Gamma	RP(P)	0.004	0.004	0.004	0.003	0.004
Log-Normal	Gamma	Cox	0.004	0.005	0.004	0.003	0.003
Log-Normal	Gamma	Exp	0.004	0.013	0.020	0.012	0.018
Log-Normal	Gamma	Wei	0.004	0.005	0.005	0.009	0.003
Log-Normal	Gamma	Gom	0.004	0.013	0.004	0.004	0.019
Log-Normal	Gamma	RP(5)	0.004	0.004	0.004	0.003	0.003
Log-Normal	Gamma	RP(9)	0.004	0.004	0.004	0.003	0.003
Log-Normal	Gamma	RP(P)	0.004	0.004	0.004	0.003	0.003
Gamma	Log-Normal	Cox	0.004	0.005	0.004	0.003	0.003
Gamma	Log-Normal	Exp	0.004	0.014	0.020	0.013	0.019
Gamma	Log-Normal	Wei	0.004	0.005	0.005	0.009	0.003
Gamma	Log-Normal	Gom	0.004	0.013	0.004	0.004	0.018
Gamma	Log-Normal	RP(5)	0.004	0.004	0.004	0.003	0.004
Gamma	Log-Normal	RP(9)	0.004	0.004	0.004	0.003	0.004
Gamma	Log-Normal	RP(P)	0.004	0.004	0.004	0.003	0.004
Log-Normal	Log-Normal	Cox	0.004	0.005	0.004	0.003	0.003
Log-Normal	Log-Normal	Exp	0.004	0.014	0.020	0.012	0.018
Log-Normal	Log-Normal	Wei	0.004	0.005	0.005	0.009	0.003
Log-Normal	Log-Normal	Gom	0.004	0.013	0.004	0.004	0.018
Log-Normal	Log-Normal	RP(5)	0.004	0.004	0.004	0.003	0.003
Log-Normal	Log-Normal	RP(9)	0.004	0.004	0.004	0.003	0.003
Log-Normal	Log-Normal	RP(P)	0.004	0.004	0.004	0.003	0.003

Table A.13: Bias with Monte Carlo standard error of estimated frailty variance, simulation study on model misspecification in survival models with shared frailty terms, scenario with 15 clusters of 100 individuals each and a small frailty variance.

True frailty	Model frailty	Model baseline	Exponential	Weibull	Gompertz	Weibull-Weibull (1)	Weibull-Weibull (2)
Gamma	Gamma	Cox	-0.096 (0.002)	-0.087 (0.002)	-0.144 (0.001)	-0.177 (0.001)	-0.164 (0.001)
Gamma	Gamma	Exp	-0.014 (0.003)	0.056 (0.004)	-0.116 (0.002)	-0.090 (0.002)	0.091 (0.004)
Gamma	Gamma	Wei	-0.014 (0.003)	-0.020 (0.003)	-0.043 (0.003)	0.049 (0.003)	-0.015 (0.003)
Gamma	Gamma	Gom	-0.018 (0.003)	0.063 (0.004)	-0.013 (0.003)	-0.042 (0.002)	0.089 (0.004)
Gamma	Gamma	RP(5)	-0.014 (0.003)	-0.020 (0.003)	-0.013 (0.003)	-0.011 (0.003)	-0.024 (0.003)
Gamma	Gamma	RP(9)	-0.014 (0.003)	-0.020 (0.003)	-0.013 (0.003)	-0.012 (0.003)	-0.023 (0.003)
Gamma	Gamma	RP(P)	-0.015 (0.003)	-0.020 (0.003)	-0.018 (0.003)	-0.016 (0.003)	-0.024 (0.003)
Log-Normal	Gamma	Cox	-0.113 (0.001)	-0.102 (0.001)	-0.148 (0.001)	-0.182 (0.001)	-0.168 (0.001)
Log-Normal	Gamma	Exp	-0.048 (0.003)	0.016 (0.004)	-0.147 (0.001)	-0.130 (0.001)	0.047 (0.004)
Log-Normal	Gamma	Wei	-0.048 (0.003)	-0.057 (0.003)	-0.081 (0.002)	-0.009 (0.003)	-0.056 (0.002)
Log-Normal	Gamma	Gom	-0.054 (0.003)	0.022 (0.004)	-0.053 (0.003)	-0.080 (0.002)	0.040 (0.004)
Log-Normal	Gamma	RP(5)	-0.048 (0.003)	-0.058 (0.003)	-0.053 (0.003)	-0.057 (0.002)	-0.065 (0.002)
Log-Normal	Gamma	RP(9)	-0.048 (0.003)	-0.058 (0.003)	-0.053 (0.003)	-0.057 (0.002)	-0.065 (0.002)
Log-Normal	Gamma	RP(P)	-0.048 (0.003)	-0.058 (0.003)	-0.057 (0.002)	-0.059 (0.002)	-0.066 (0.002)
Gamma	Log-Normal	Cox	0.037 (0.004)	0.027 (0.004)	0.038 (0.004)	0.039 (0.004)	0.033 (0.004)
Gamma	Log-Normal	Exp	0.016 (0.004)	0.091 (0.005)	-0.101 (0.003)	-0.067 (0.003)	0.144 (0.005)
Gamma	Log-Normal	Wei	0.016 (0.004)	0.006 (0.004)	-0.018 (0.003)	0.104 (0.005)	0.019 (0.004)
Gamma	Log-Normal	Gom	0.014 (0.004)	0.089 (0.005)	0.017 (0.004)	-0.009 (0.003)	0.138 (0.005)
Gamma	Log-Normal	RP(5)	0.016 (0.004)	0.007 (0.004)	0.017 (0.004)	0.019 (0.004)	0.007 (0.004)
Gamma	Log-Normal	RP(9)	0.016 (0.004)	0.007 (0.004)	0.017 (0.004)	0.018 (0.004)	0.007 (0.004)
Gamma	Log-Normal	RP(P)	0.016 (0.004)	0.007 (0.004)	0.012 (0.004)	0.017 (0.004)	0.006 (0.004)
Log-Normal	Log-Normal	Cox	-0.023 (0.003)	-0.032 (0.003)	-0.028 (0.003)	-0.032 (0.003)	-0.029 (0.003)
Log-Normal	Log-Normal	Exp	-0.039 (0.003)	0.027 (0.004)	-0.142 (0.002)	-0.121 (0.002)	0.066 (0.004)
Log-Normal	Log-Normal	Wei	-0.039 (0.003)	-0.048 (0.003)	-0.073 (0.002)	0.014 (0.003)	-0.043 (0.003)
Log-Normal	Log-Normal	Gom	-0.049 (0.003)	0.031 (0.004)	-0.043 (0.003)	-0.064 (0.002)	0.066 (0.004)
Log-Normal	Log-Normal	RP(5)	-0.039 (0.003)	-0.049 (0.003)	-0.043 (0.003)	-0.047 (0.002)	-0.056 (0.002)
Log-Normal	Log-Normal	RP(9)	-0.039 (0.003)	-0.049 (0.003)	-0.043 (0.003)	-0.047 (0.002)	-0.056 (0.002)
Log-Normal	Log-Normal	RP(P)	-0.039 (0.003)	-0.048 (0.003)	-0.047 (0.003)	-0.048 (0.002)	-0.057 (0.002)

Table A.14: Coverage with Monte Carlo standard error of estimated frailty variance, simulation study on model misspecification in survival models with shared frailty terms, scenario with 15 clusters of 100 individuals each and a small frailty variance.

True frailty	Model frailty	Model baseline	Exponential	Weibull	Gompertz	Weibull-Weibull (1)	Weibull-Weibull (2)
Gamma	Gamma	Cox	59.48 (1.55)	64.26 (1.52)	5.71 (0.73)	0.00 (0.00)	2.22 (0.47)
Gamma	Gamma	Exp	85.70 (1.11)	94.50 (0.72)	38.10 (1.54)	54.90 (1.57)	96.80 (0.56)
Gamma	Gamma	Wei	86.10 (1.09)	83.30 (1.18)	77.20 (1.33)	95.80 (0.63)	85.60 (1.11)
Gamma	Gamma	Gom	82.35 (1.21)	96.28 (0.60)	85.70 (1.11)	83.06 (1.19)	96.42 (0.59)
Gamma	Gamma	RP(5)	85.97 (1.10)	83.36 (1.18)	85.77 (1.10)	86.83 (1.07)	80.45 (1.25)
Gamma	Gamma	RP(9)	85.87 (1.10)	83.36 (1.18)	85.80 (1.10)	86.90 (1.07)	81.36 (1.23)
Gamma	Gamma	RP(P)	86.02 (1.10)	83.36 (1.18)	85.09 (1.13)	86.37 (1.08)	80.45 (1.25)
Log-Normal	Gamma	Cox	47.66 (1.58)	55.37 (1.57)	6.78 (0.79)	0.00 (0.00)	0.00 (0.00)
Log-Normal	Gamma	Exp	76.80 (1.33)	89.00 (0.99)	17.80 (1.21)	28.30 (1.42)	93.50 (0.78)
Log-Normal	Gamma	Wei	77.10 (1.33)	72.60 (1.41)	61.70 (1.54)	90.20 (0.94)	73.62 (1.39)
Log-Normal	Gamma	Gom	74.20 (1.38)	90.98 (0.91)	74.70 (1.37)	66.47 (1.49)	91.41 (0.89)
Log-Normal	Gamma	RP(5)	76.48 (1.34)	72.04 (1.42)	75.05 (1.37)	73.69 (1.39)	71.37 (1.43)
Log-Normal	Gamma	RP(9)	76.78 (1.34)	72.04 (1.42)	74.92 (1.37)	73.54 (1.40)	71.37 (1.43)
Log-Normal	Gamma	RP(P)	76.98 (1.33)	72.04 (1.42)	73.17 (1.40)	72.36 (1.41)	71.37 (1.43)
Gamma	Log-Normal	Cox	84.50 (1.14)	82.20 (1.21)	83.30 (1.18)	84.40 (1.15)	82.50 (1.20)
Gamma	Log-Normal	Exp	87.78 (1.04)	96.00 (0.62)	45.70 (1.58)	63.30 (1.52)	96.80 (0.56)
Gamma	Log-Normal	Wei	87.88 (1.03)	85.89 (1.10)	80.88 (1.24)	96.99 (0.54)	88.23 (1.02)
Gamma	Log-Normal	Gom	86.29 (1.09)	96.60 (0.57)	88.08 (1.02)	86.84 (1.07)	97.38 (0.50)
Gamma	Log-Normal	RP(5)	87.68 (1.04)	85.95 (1.10)	88.10 (1.02)	89.60 (0.97)	85.91 (1.10)
Gamma	Log-Normal	RP(9)	87.68 (1.04)	85.95 (1.10)	88.20 (1.02)	89.50 (0.97)	85.91 (1.10)
Gamma	Log-Normal	RP(P)	87.78 (1.04)	85.95 (1.10)	87.50 (1.05)	89.20 (0.98)	85.91 (1.10)
Log-Normal	Log-Normal	Cox	76.70 (1.34)	73.80 (1.39)	76.00 (1.35)	74.00 (1.39)	72.60 (1.41)
Log-Normal	Log-Normal	Exp	80.18 (1.26)	91.09 (0.90)	21.70 (1.30)	36.50 (1.52)	94.38 (0.73)
Log-Normal	Log-Normal	Wei	79.84 (1.27)	77.48 (1.32)	66.87 (1.49)	92.08 (0.85)	77.89 (1.31)
Log-Normal	Log-Normal	Gom	77.34 (1.32)	92.97 (0.81)	77.56 (1.32)	71.36 (1.43)	92.42 (0.84)
Log-Normal	Log-Normal	RP(5)	80.08 (1.26)	76.44 (1.34)	77.66 (1.32)	77.58 (1.32)	77.35 (1.32)
Log-Normal	Log-Normal	RP(9)	79.98 (1.27)	76.44 (1.34)	77.66 (1.32)	77.38 (1.32)	77.35 (1.32)
Log-Normal	Log-Normal	RP(P)	79.88 (1.27)	76.60 (1.34)	76.78 (1.34)	77.20 (1.33)	77.35 (1.32)

Table A.15: Mean squared error of estimated frailty variance, simulation study on model misspecification in survival models with shared frailty terms, scenario with 15 clusters of 100 individuals each and a small frailty variance.

True frailty	Model frailty	Model baseline	Exponential	Weibull	Gompertz	Weibull-Weibull (1)	Weibull-Weibull (2)
Gamma	Gamma	Cox	0.012	0.010	0.021	0.032	0.027
Gamma	Gamma	Exp	0.009	0.017	0.017	0.013	0.023
Gamma	Gamma	Wei	0.009	0.009	0.009	0.013	0.008
Gamma	Gamma	Gom	0.009	0.017	0.009	0.006	0.022
Gamma	Gamma	RP(5)	0.009	0.009	0.009	0.008	0.008
Gamma	Gamma	RP(9)	0.009	0.009	0.009	0.008	0.008
Gamma	Gamma	RP(P)	0.009	0.009	0.009	0.008	0.008
Log-Normal	Gamma	Cox	0.014	0.012	0.023	0.033	0.028
Log-Normal	Gamma	Exp	0.009	0.014	0.024	0.019	0.015
Log-Normal	Gamma	Wei	0.009	0.010	0.011	0.007	0.009
Log-Normal	Gamma	Gom	0.009	0.015	0.009	0.010	0.015
Log-Normal	Gamma	RP(5)	0.009	0.010	0.009	0.009	0.009
Log-Normal	Gamma	RP(9)	0.009	0.010	0.009	0.009	0.009
Log-Normal	Gamma	RP(P)	0.009	0.010	0.009	0.009	0.009
Gamma	Log-Normal	Cox	0.018	0.017	0.019	0.016	0.017
Gamma	Log-Normal	Exp	0.014	0.030	0.017	0.012	0.048
Gamma	Log-Normal	Wei	0.014	0.014	0.012	0.032	0.015
Gamma	Log-Normal	Gom	0.014	0.030	0.015	0.009	0.045
Gamma	Log-Normal	RP(5)	0.014	0.014	0.015	0.013	0.012
Gamma	Log-Normal	RP(9)	0.014	0.014	0.015	0.013	0.012
Gamma	Log-Normal	RP(P)	0.014	0.014	0.015	0.013	0.012
Log-Normal	Log-Normal	Cox	0.009	0.009	0.009	0.008	0.009
Log-Normal	Log-Normal	Exp	0.009	0.015	0.023	0.017	0.020
Log-Normal	Log-Normal	Wei	0.009	0.009	0.011	0.010	0.009
Log-Normal	Log-Normal	Gom	0.009	0.015	0.009	0.009	0.022
Log-Normal	Log-Normal	RP(5)	0.009	0.010	0.009	0.008	0.009
Log-Normal	Log-Normal	RP(9)	0.009	0.010	0.009	0.008	0.009
Log-Normal	Log-Normal	RP(P)	0.009	0.010	0.009	0.008	0.009

Table A.16: Bias with Monte Carlo standard error of difference in 5-years life expectancy, simulation study on model misspecification in survival models with shared frailty terms, scenario with 15 clusters of 100 individuals each and a small frailty variance.

True frailty	Model frailty	Model baseline	Exponential	Weibull	Gompertz	Weibull-Weibull (1)	Weibull-Weibull (2)
Gamma	Gamma	Cox	0.008 (0.003)	-0.000 (0.003)	0.003 (0.002)	0.009 (0.003)	0.053 (0.003)
Gamma	Gamma	Exp	0.001 (0.003)	0.024 (0.003)	-0.038 (0.002)	-0.008 (0.002)	0.072 (0.003)
Gamma	Gamma	Wei	0.001 (0.003)	-0.001 (0.003)	-0.023 (0.002)	0.061 (0.002)	0.019 (0.003)
Gamma	Gamma	Gom	0.005 (0.003)	0.030 (0.003)	0.001 (0.002)	0.037 (0.002)	0.069 (0.003)
Gamma	Gamma	RP(5)	0.001 (0.003)	0.000 (0.003)	0.000 (0.002)	-0.005 (0.002)	0.013 (0.003)
Gamma	Gamma	RP(9)	0.001 (0.003)	0.000 (0.003)	0.001 (0.002)	-0.002 (0.002)	0.012 (0.003)
Gamma	Gamma	RP(P)	0.001 (0.003)	0.001 (0.003)	-0.003 (0.002)	-0.001 (0.002)	0.014 (0.003)
Log-Normal	Gamma	Cox	0.001 (0.003)	0.008 (0.003)	0.006 (0.002)	0.032 (0.002)	0.033 (0.003)
Log-Normal	Gamma	Exp	-0.000 (0.003)	0.028 (0.003)	-0.040 (0.002)	-0.009 (0.002)	0.064 (0.003)
Log-Normal	Gamma	Wei	-0.000 (0.003)	0.006 (0.003)	-0.025 (0.002)	0.060 (0.002)	0.013 (0.003)
Log-Normal	Gamma	Gom	-0.000 (0.003)	0.027 (0.003)	-0.002 (0.002)	0.040 (0.002)	0.066 (0.003)
Log-Normal	Gamma	RP(5)	-0.000 (0.003)	0.005 (0.003)	-0.002 (0.002)	-0.011 (0.002)	-0.009 (0.002)
Log-Normal	Gamma	RP(9)	-0.000 (0.003)	0.005 (0.003)	-0.002 (0.002)	-0.009 (0.002)	-0.010 (0.002)
Log-Normal	Gamma	RP(P)	-0.000 (0.003)	0.005 (0.003)	-0.006 (0.002)	-0.007 (0.002)	-0.008 (0.002)
Gamma	Log-Normal	Cox	-0.012 (0.003)	-0.016 (0.003)	-0.004 (0.002)	0.066 (0.002)	0.045 (0.003)
Gamma	Log-Normal	Exp	-0.022 (0.003)	-0.004 (0.003)	-0.048 (0.002)	0.003 (0.002)	0.118 (0.003)
Gamma	Log-Normal	Wei	-0.022 (0.003)	-0.025 (0.003)	-0.035 (0.002)	0.094 (0.002)	0.042 (0.003)
Gamma	Log-Normal	Gom	-0.020 (0.003)	-0.000 (0.004)	-0.016 (0.002)	0.056 (0.002)	0.116 (0.003)
Gamma	Log-Normal	RP(5)	0.014 (0.003)	0.016 (0.003)	0.011 (0.002)	0.017 (0.002)	0.045 (0.003)
Gamma	Log-Normal	RP(9)	0.014 (0.003)	0.016 (0.003)	0.011 (0.002)	0.020 (0.002)	0.044 (0.003)
Gamma	Log-Normal	RP(P)	0.014 (0.003)	0.016 (0.003)	0.007 (0.002)	0.021 (0.002)	0.046 (0.003)
Log-Normal	Log-Normal	Cox	-0.009 (0.003)	-0.006 (0.003)	-0.005 (0.002)	0.053 (0.002)	0.036 (0.003)
Log-Normal	Log-Normal	Exp	-0.017 (0.003)	0.008 (0.003)	-0.047 (0.002)	0.001 (0.002)	0.109 (0.003)
Log-Normal	Log-Normal	Wei	-0.017 (0.003)	-0.012 (0.003)	-0.034 (0.002)	0.088 (0.002)	0.035 (0.003)
Log-Normal	Log-Normal	Gom	-0.016 (0.003)	0.005 (0.003)	-0.015 (0.002)	0.056 (0.002)	0.106 (0.003)
Log-Normal	Log-Normal	RP(5)	0.013 (0.003)	0.020 (0.003)	0.007 (0.002)	0.006 (0.002)	0.019 (0.003)
Log-Normal	Log-Normal	RP(9)	0.013 (0.003)	0.020 (0.003)	0.007 (0.002)	0.009 (0.002)	0.018 (0.003)
Log-Normal	Log-Normal	RP(P)	0.013 (0.003)	0.020 (0.003)	0.004 (0.002)	0.010 (0.002)	0.020 (0.003)

Table A.17: Mean squared error of difference in 5-years life expectancy, simulation study on model misspecification in survival models with shared frailty terms, scenario with 15 clusters of 100 individuals each and a small frailty variance.

True frailty	Model frailty	Model baseline	Exponential	Weibull	Gompertz	Weibull-Weibull (1)	Weibull-Weibull (2)
Gamma	Gamma	Cox	0.008	0.010	0.005	0.009	0.010
Gamma	Gamma	Exp	0.007	0.012	0.005	0.005	0.014
Gamma	Gamma	Wei	0.007	0.010	0.004	0.010	0.008
Gamma	Gamma	Gom	0.007	0.013	0.004	0.007	0.014
Gamma	Gamma	RP(5)	0.007	0.009	0.005	0.005	0.009
Gamma	Gamma	RP(9)	0.007	0.009	0.004	0.005	0.009
Gamma	Gamma	RP(P)	0.007	0.009	0.004	0.005	0.009
Log-Normal	Gamma	Cox	0.008	0.010	0.004	0.006	0.008
Log-Normal	Gamma	Exp	0.007	0.011	0.005	0.005	0.014
Log-Normal	Gamma	Wei	0.007	0.009	0.005	0.010	0.008
Log-Normal	Gamma	Gom	0.007	0.011	0.004	0.007	0.014
Log-Normal	Gamma	RP(5)	0.007	0.009	0.004	0.005	0.006
Log-Normal	Gamma	RP(9)	0.007	0.009	0.004	0.005	0.006
Log-Normal	Gamma	RP(P)	0.007	0.009	0.004	0.005	0.006
Gamma	Log-Normal	Cox	0.007	0.010	0.004	0.009	0.009
Gamma	Log-Normal	Exp	0.008	0.012	0.006	0.005	0.023
Gamma	Log-Normal	Wei	0.008	0.011	0.005	0.015	0.009
Gamma	Log-Normal	Gom	0.008	0.013	0.005	0.009	0.022
Gamma	Log-Normal	RP(5)	0.008	0.010	0.005	0.006	0.011
Gamma	Log-Normal	RP(9)	0.008	0.010	0.005	0.006	0.011
Gamma	Log-Normal	RP(P)	0.008	0.010	0.005	0.006	0.011
Log-Normal	Log-Normal	Cox	0.007	0.009	0.004	0.008	0.008
Log-Normal	Log-Normal	Exp	0.007	0.011	0.006	0.005	0.021
Log-Normal	Log-Normal	Wei	0.007	0.010	0.005	0.014	0.009
Log-Normal	Log-Normal	Gom	0.008	0.010	0.005	0.009	0.020
Log-Normal	Log-Normal	RP(5)	0.008	0.009	0.005	0.005	0.007
Log-Normal	Log-Normal	RP(9)	0.008	0.009	0.005	0.005	0.007
Log-Normal	Log-Normal	RP(P)	0.008	0.009	0.004	0.005	0.007

## **Appendix B**

### **Plots**

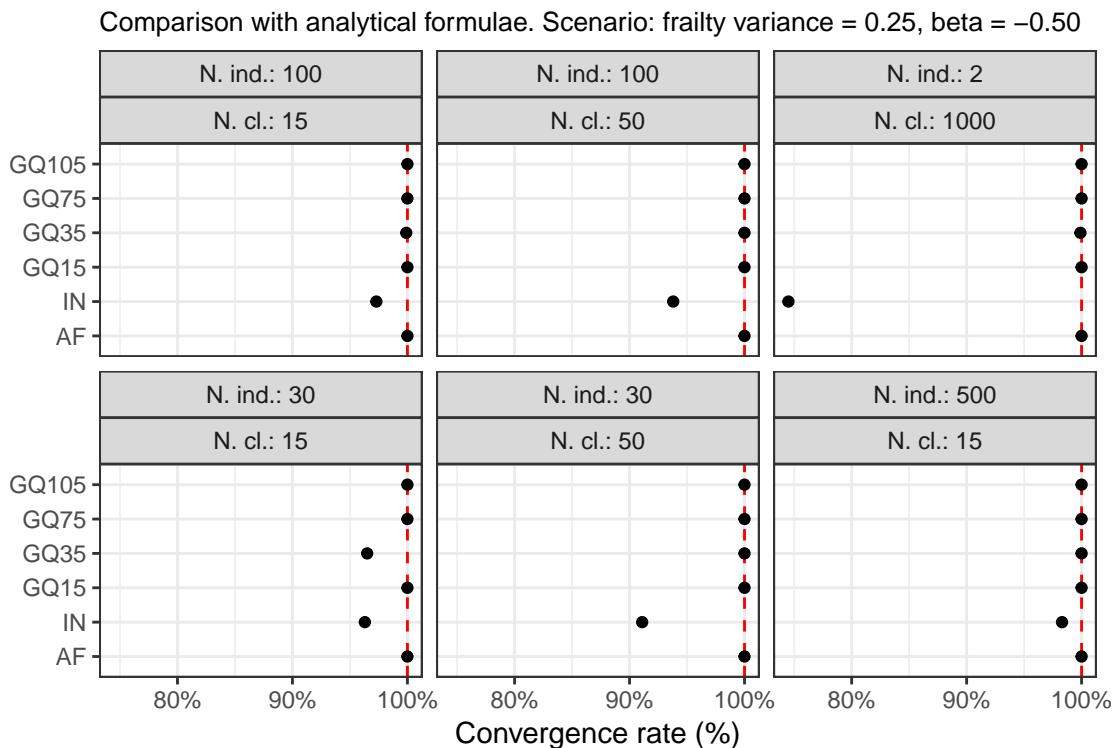


Figure B.1: Convergence rates, comparison with analytical formulae, scenario with a small frailty variance and a negative regression coefficient.

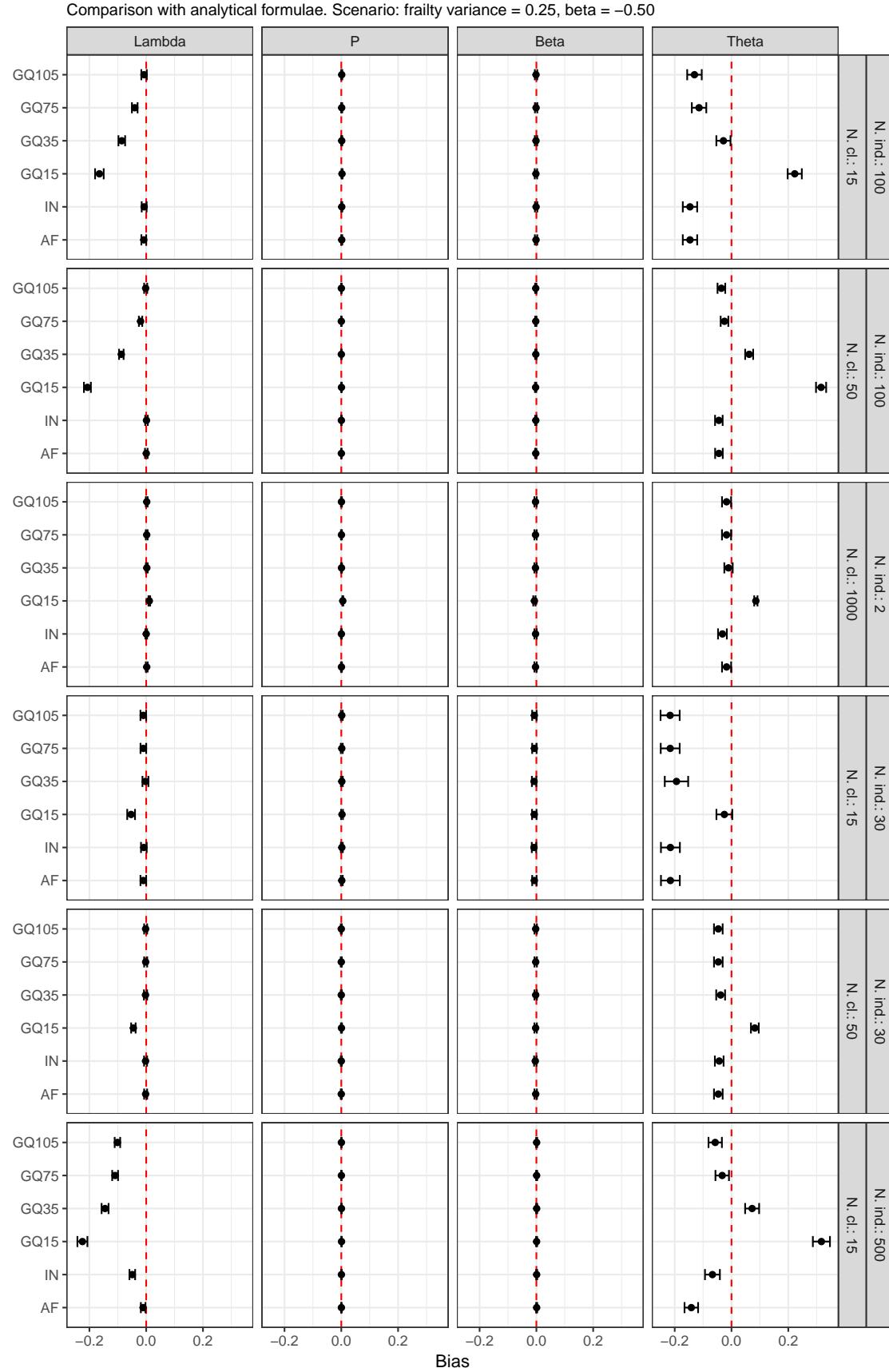


Figure B.2: Bias, comparison with analytical formulae, scenario with a small frailty variance and a negative regression coefficient.

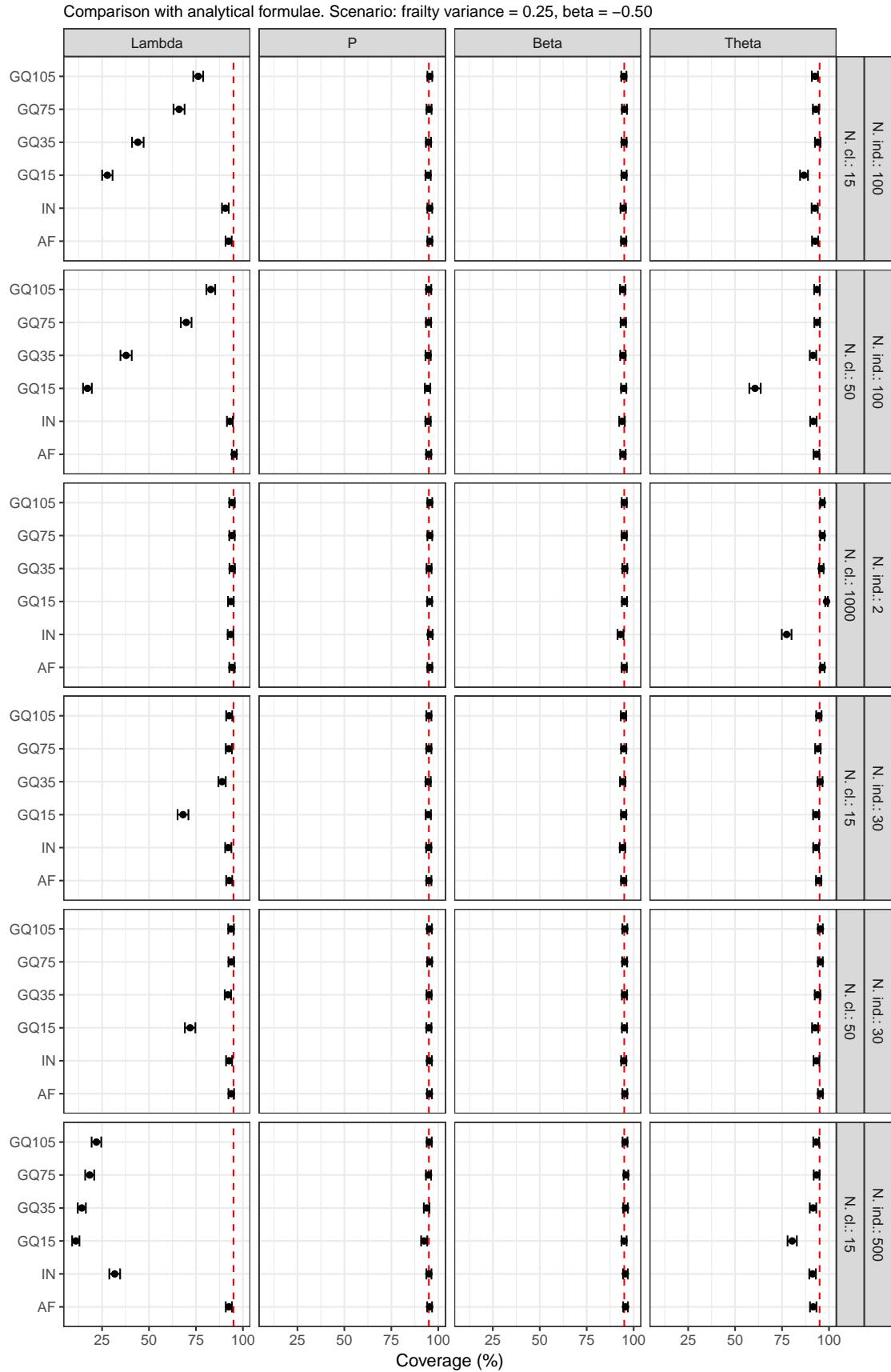


Figure B.3: Coverage, comparison with analytical formulae, scenario with a small frailty variance and a negative regression coefficient.

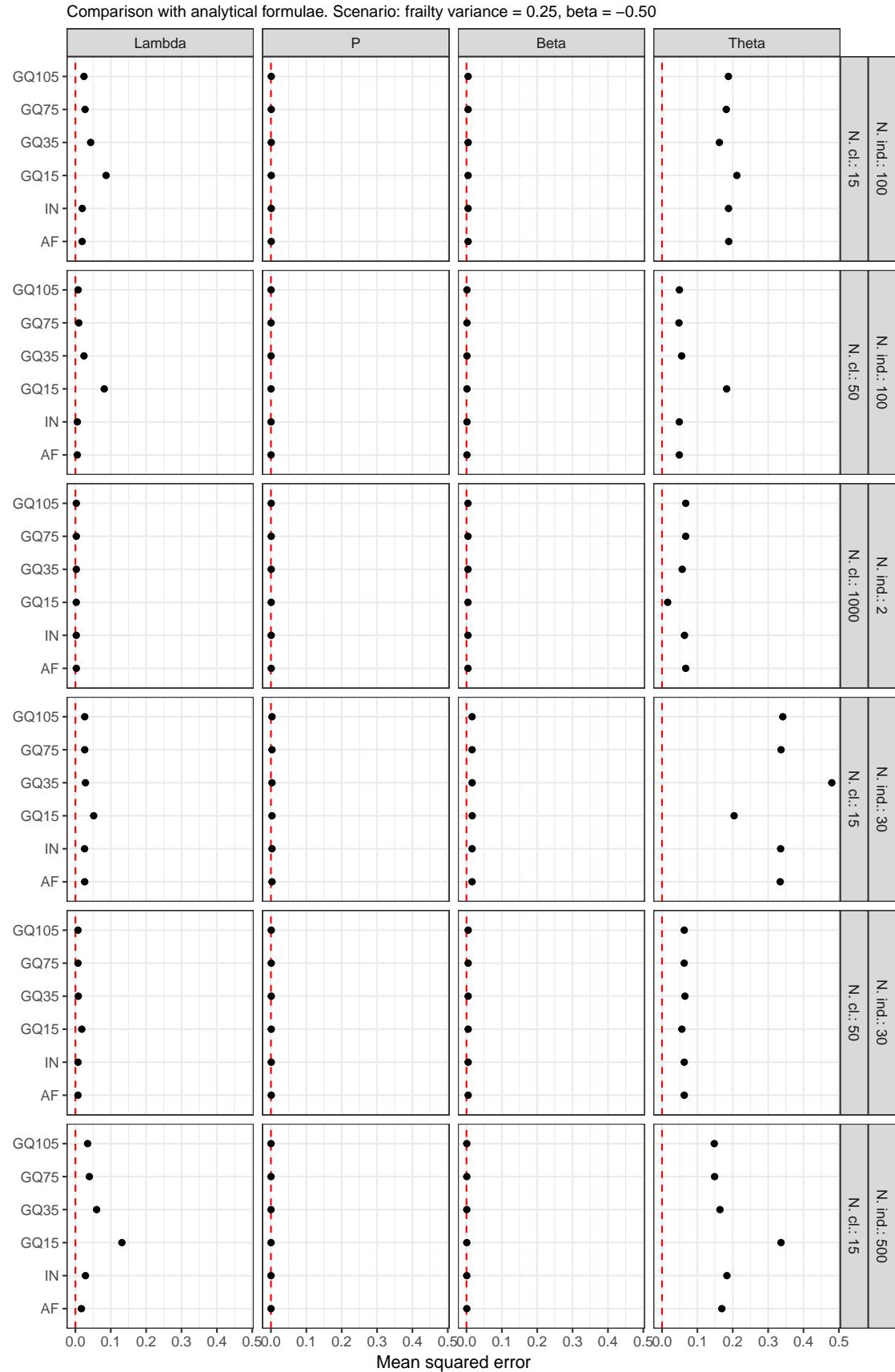


Figure B.4: Mean squared error, comparison with analytical formulae, scenario with a small frailty variance and a negative regression coefficient.

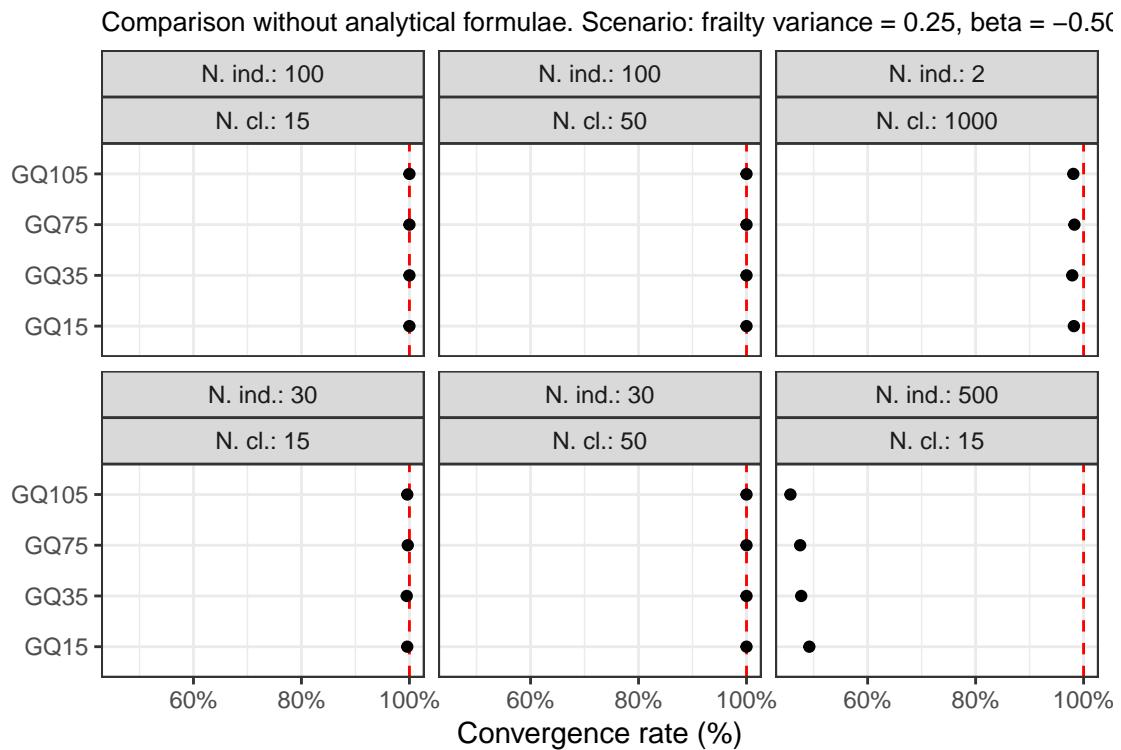


Figure B.5: Convergence rates, comparison without analytical formulae, scenario with a small frailty variance and a negative regression coefficient.

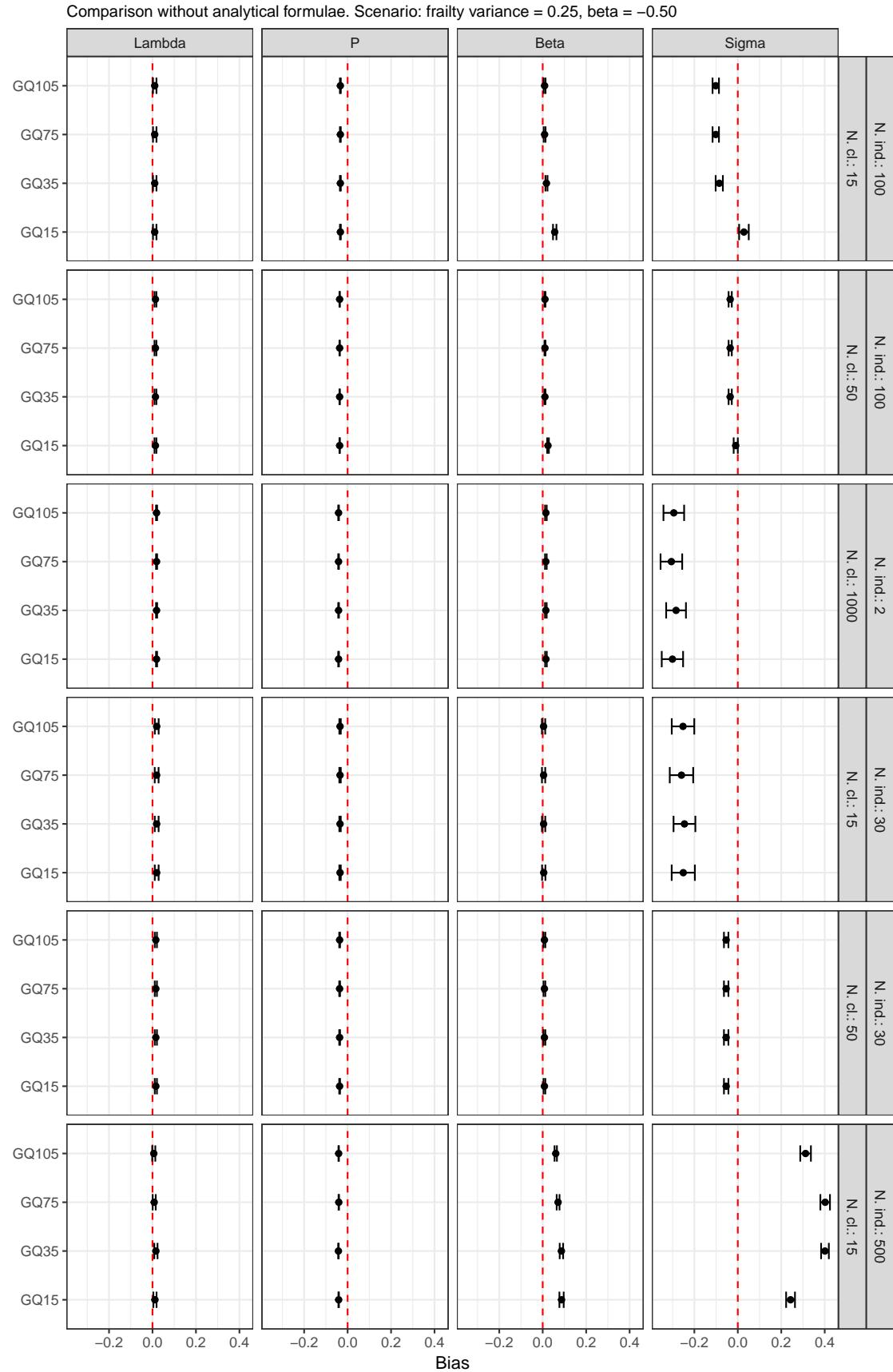


Figure B.6: Bias, comparison without analytical formulae, scenario with a small frailty variance and a negative regression coefficient.

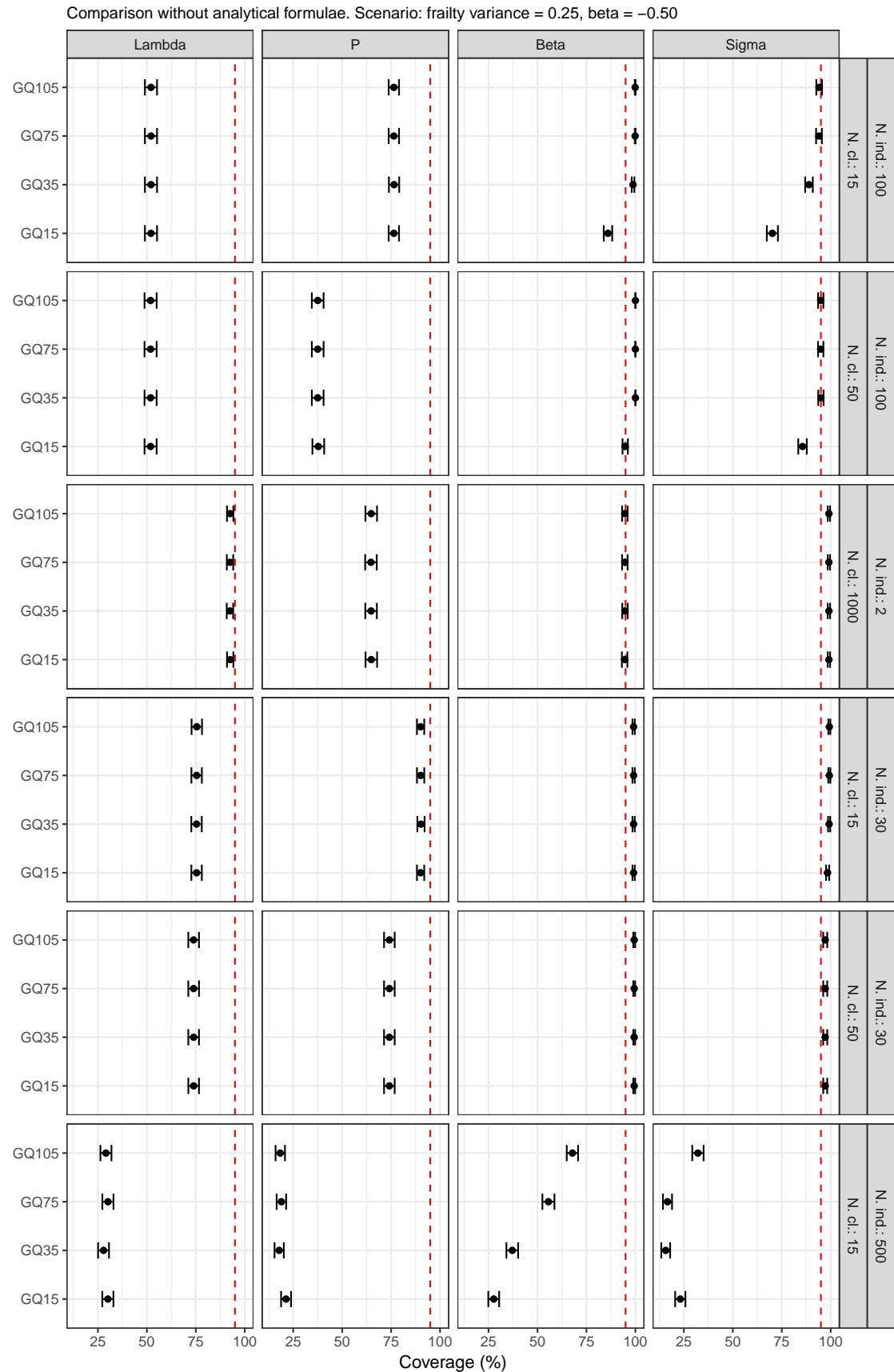


Figure B.7: Coverage, comparison without analytical formulae, scenario with a small frailty variance and a negative regression coefficient.

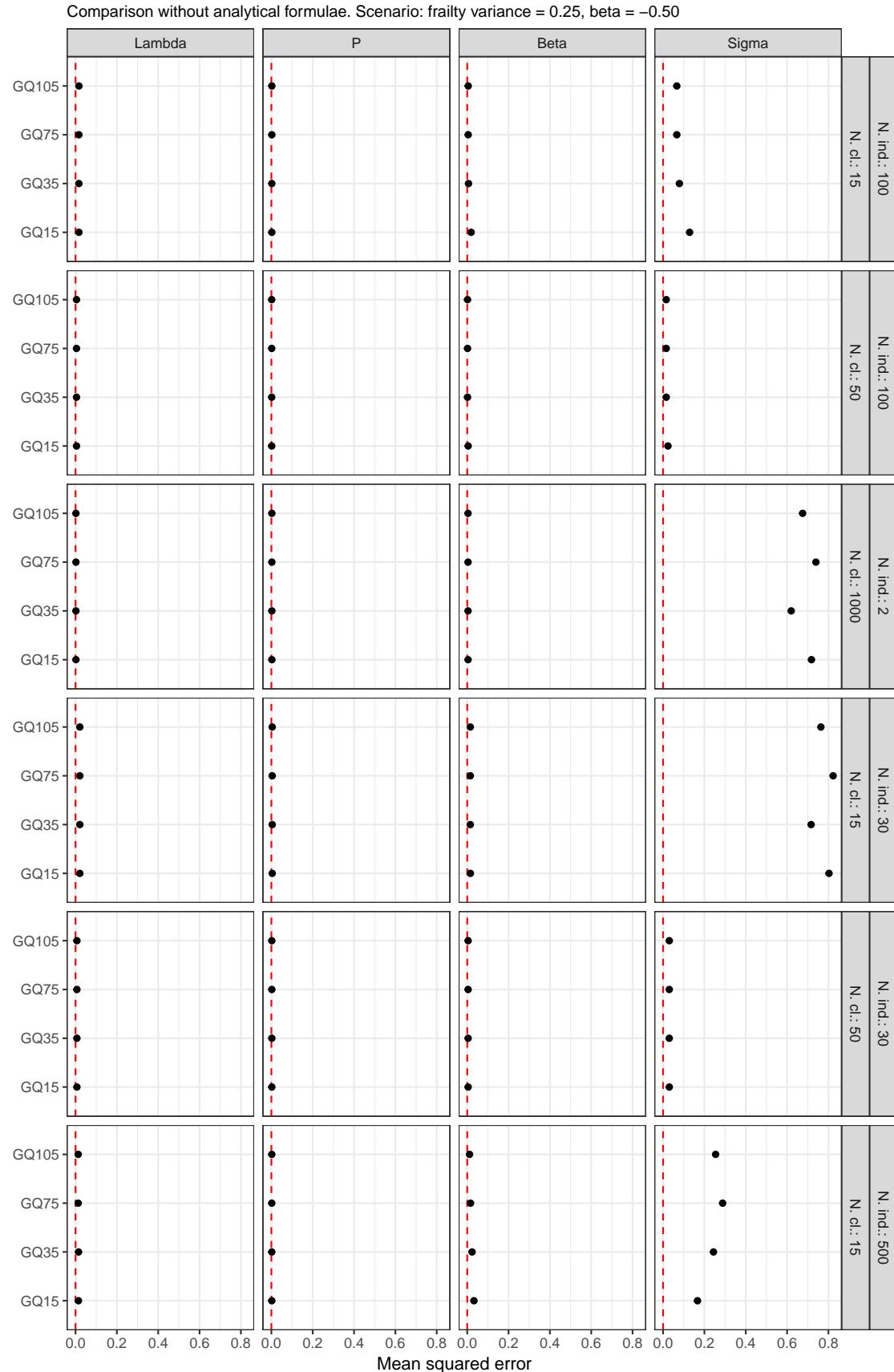


Figure B.8: Mean squared error, comparison without analytical formulae, scenario with a small frailty variance and a negative regression coefficient.

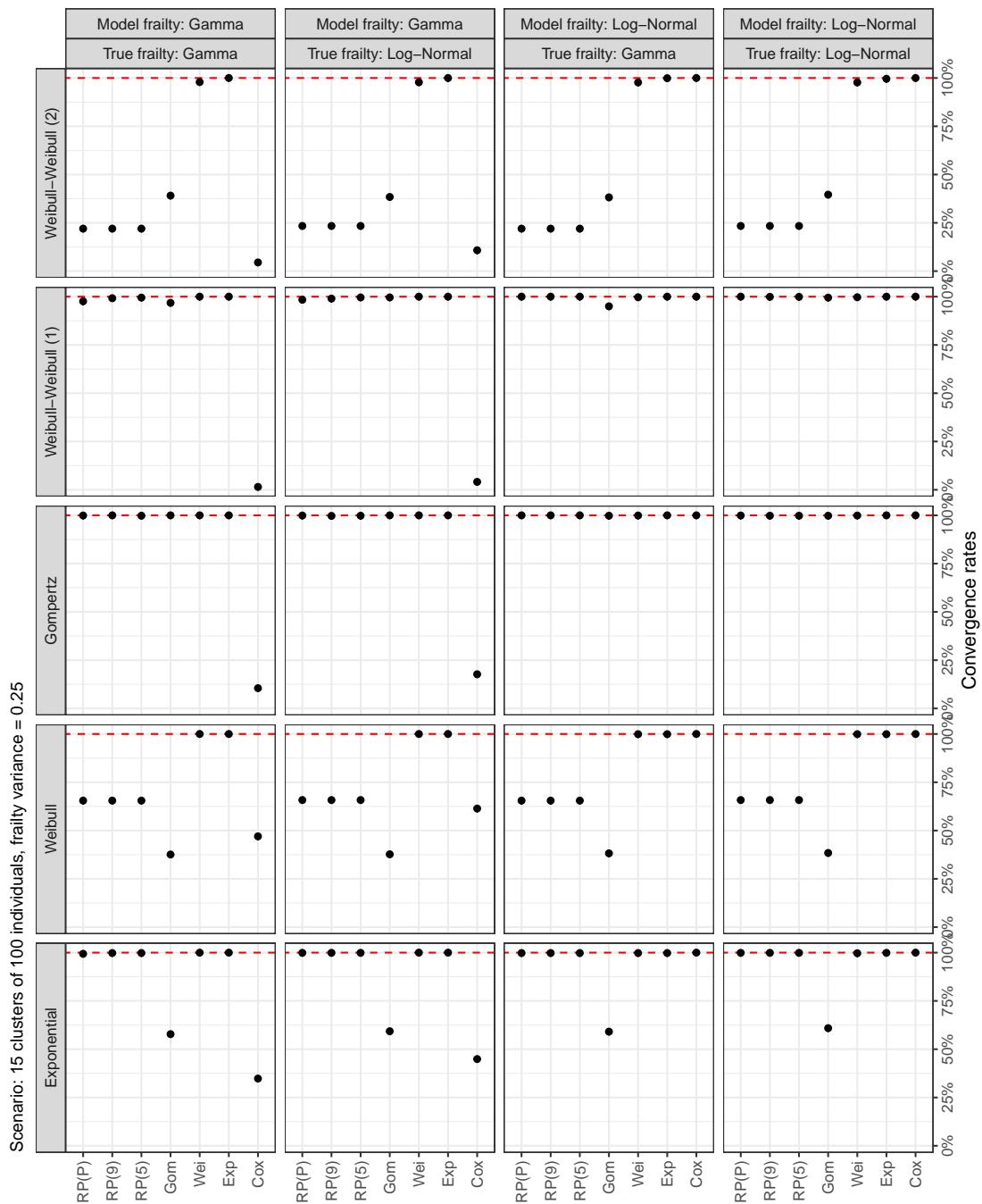


Figure B.9: Convergence rates, simulation study on model misspecification in survival models with shared frailty terms, scenario with 15 clusters of 100 individuals each and a small frailty variance.

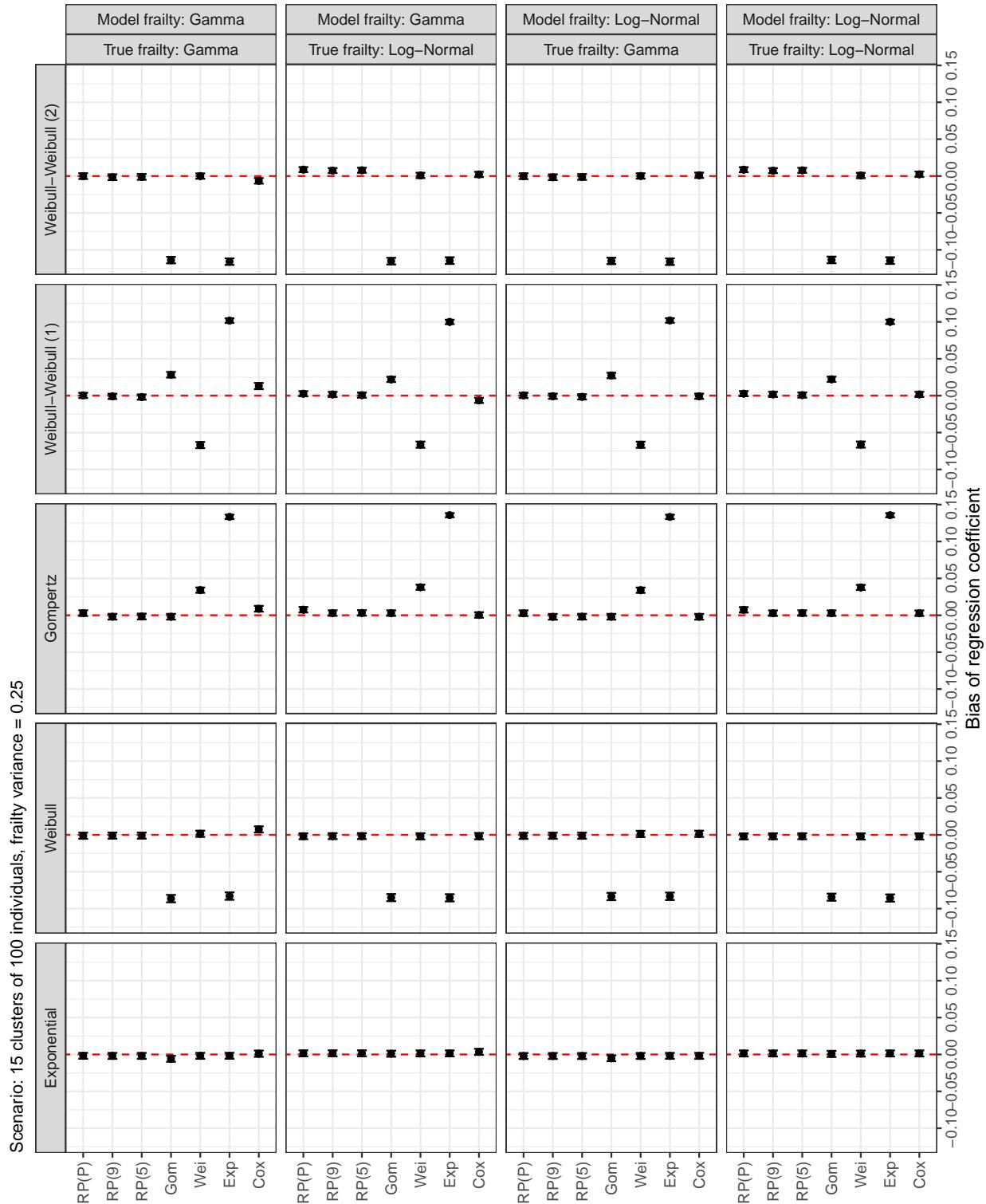


Figure B.10: Bias of estimated regression coefficient, simulation study on model misspecification in survival models with shared frailty terms, scenario with 15 clusters of 100 individuals each and a small frailty variance.

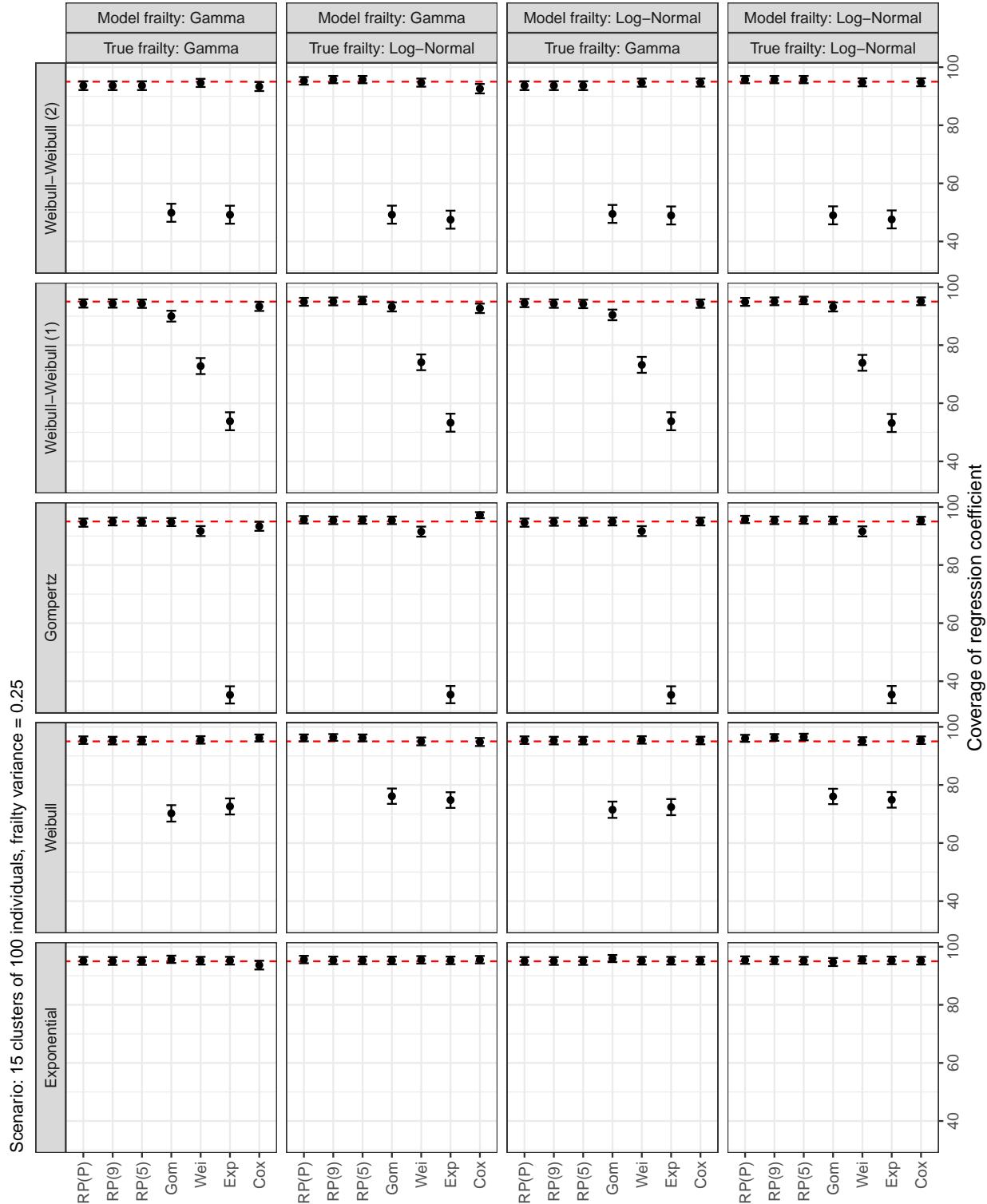


Figure B.11: Coverage of estimated regression coefficient, simulation study on model misspecification in survival models with shared frailty terms, scenario with 15 clusters of 100 individuals each and a small frailty variance.

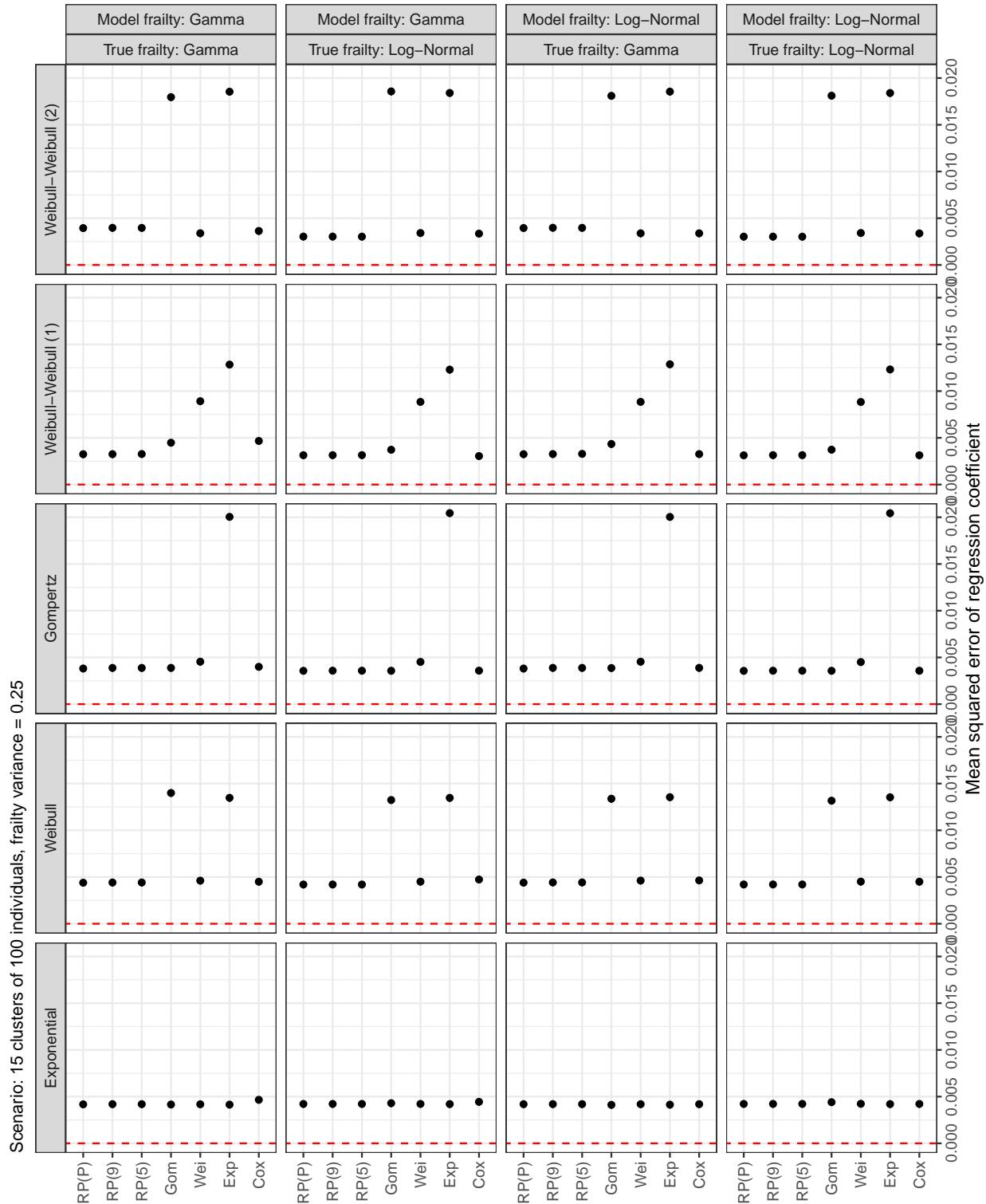


Figure B.12: Mean squared error of estimated regression coefficient, simulation study on model misspecification in survival models with shared frailty terms, scenario with 15 clusters of 100 individuals each and a small frailty variance.

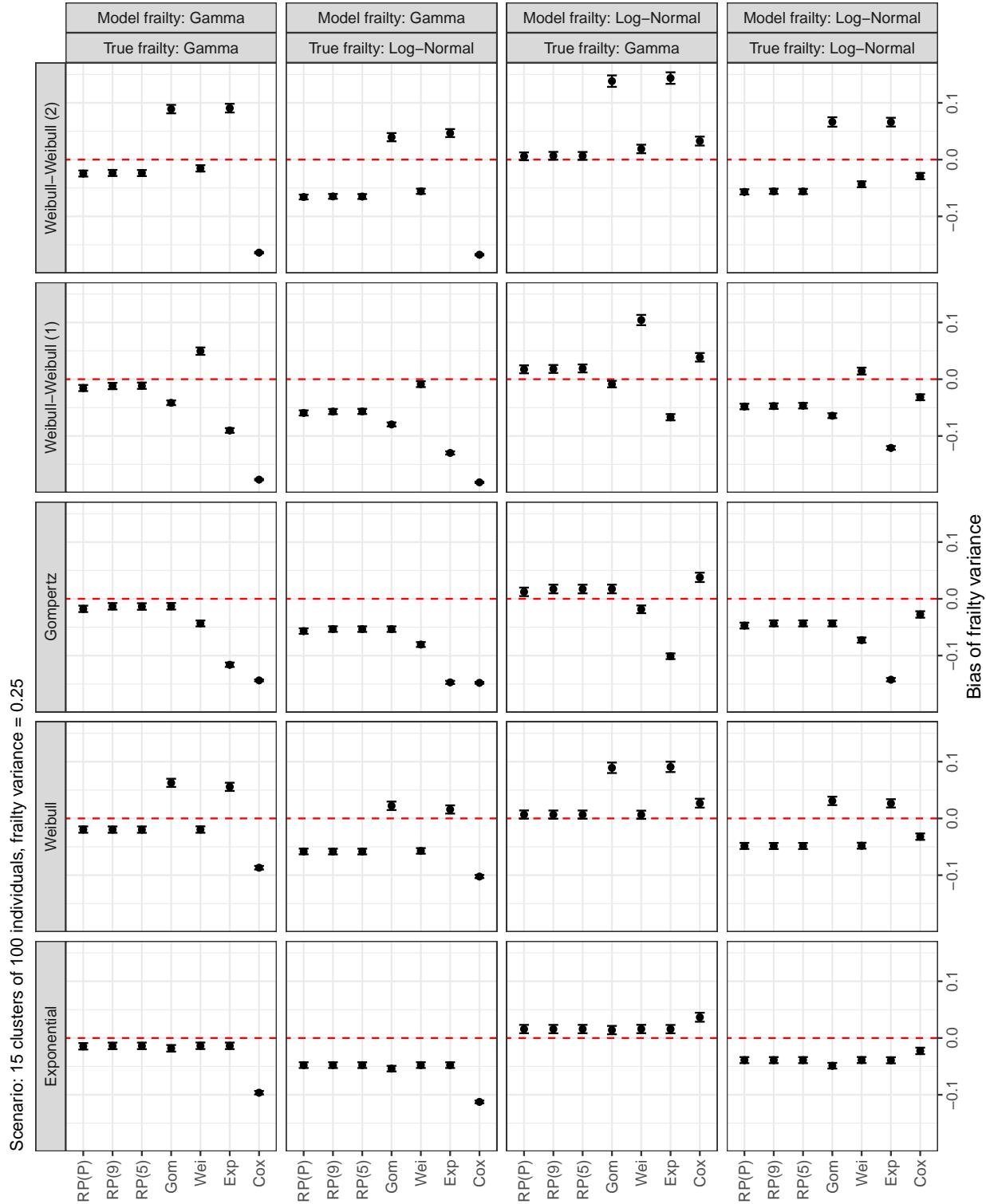


Figure B.13: Bias of estimated frailty variance, simulation study on model misspecification in survival models with shared frailty terms, scenario with 15 clusters of 100 individuals each and a small frailty variance.

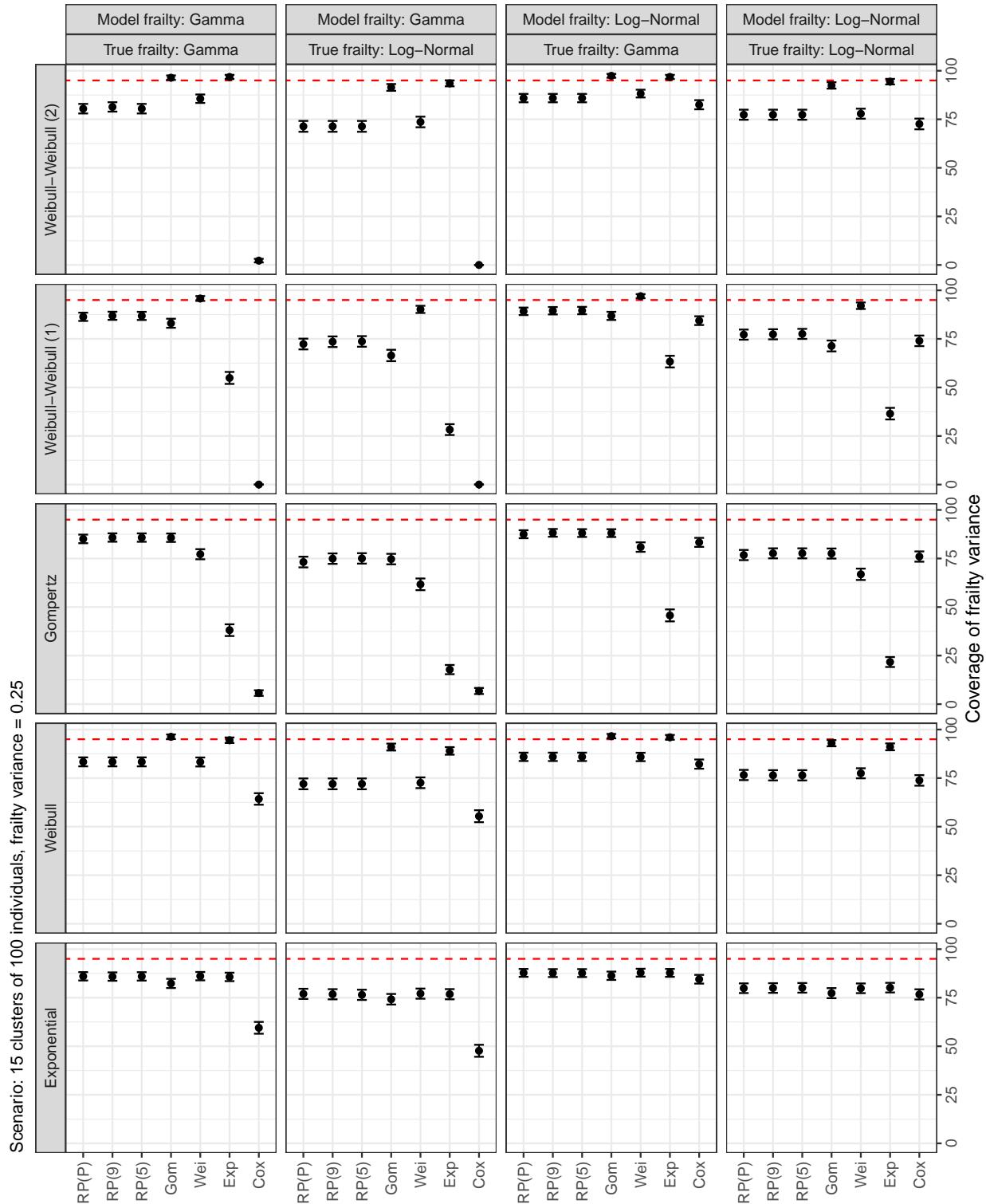


Figure B.14: Coverage of estimated frailty variance, simulation study on model misspecification in survival models with shared frailty terms, scenario with 15 clusters of 100 individuals each and a small frailty variance.

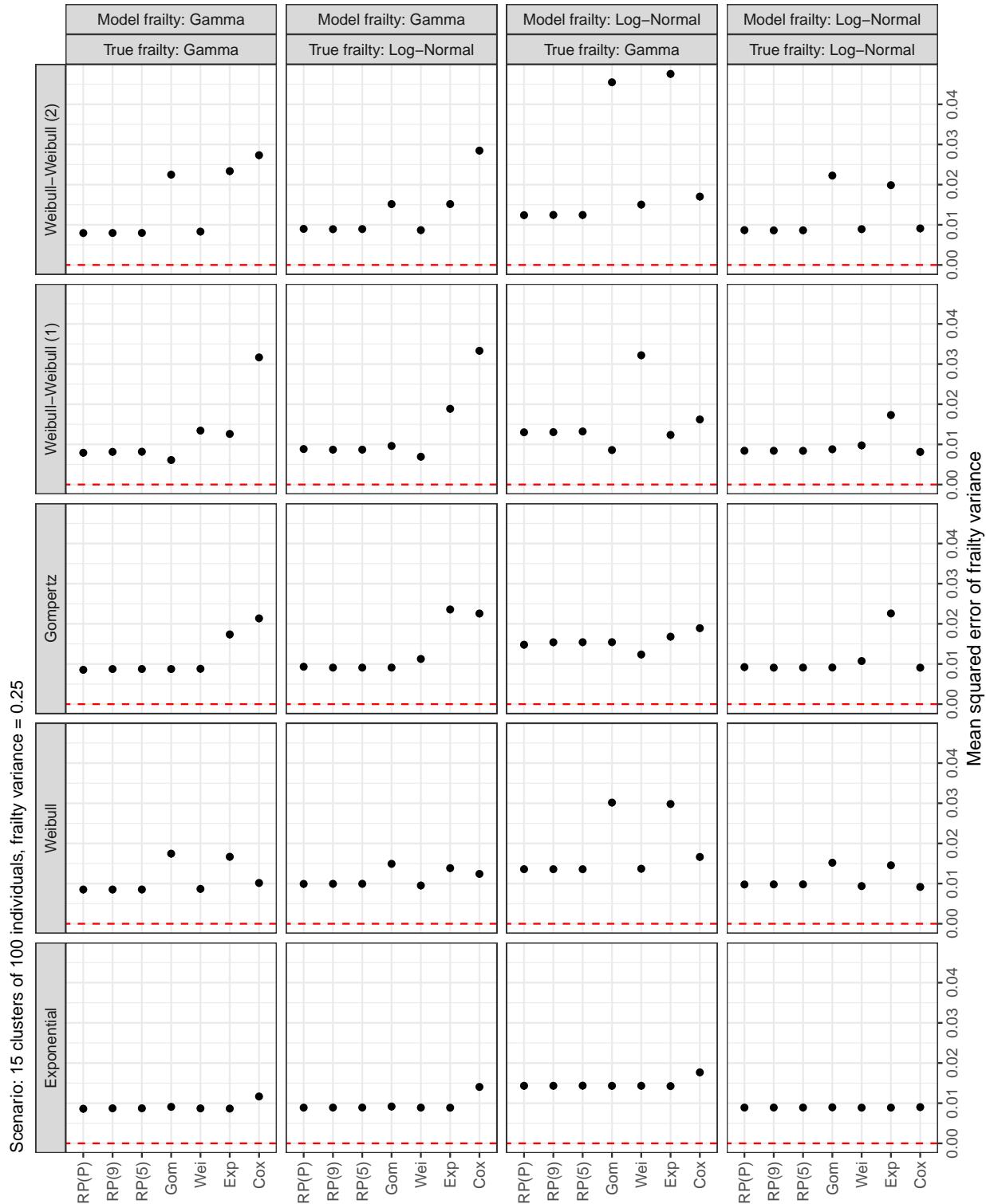


Figure B.15: Mean squared error of estimated frailty variance, simulation study on model misspecification in survival models with shared frailty terms, scenario with 15 clusters of 100 individuals each and a small frailty variance.

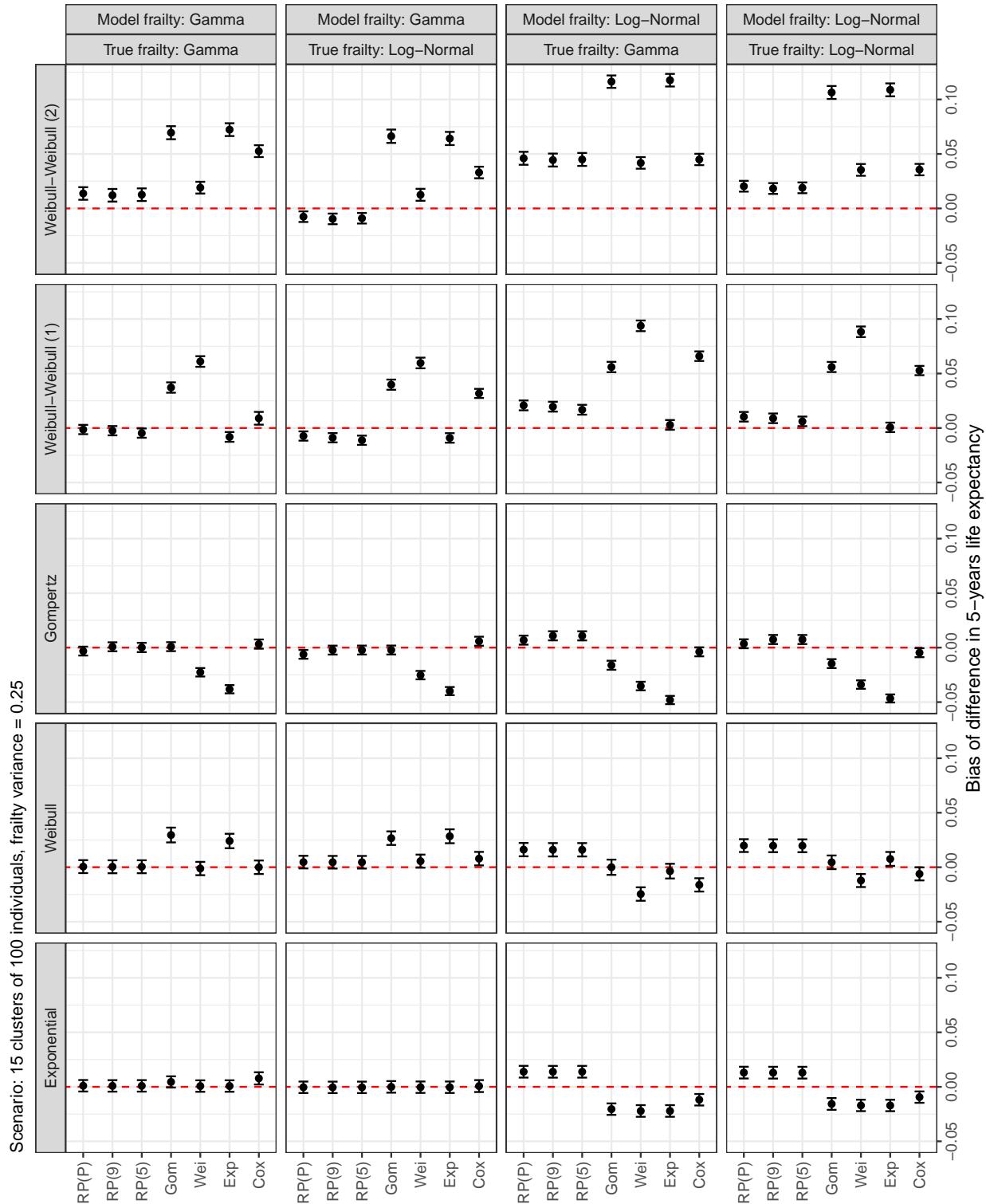


Figure B.16: Bias of estimated difference in 5-years life expectancy, simulation study on model misspecification in survival models with shared frailty terms, scenario with 15 clusters of 100 individuals each and a small frailty variance.

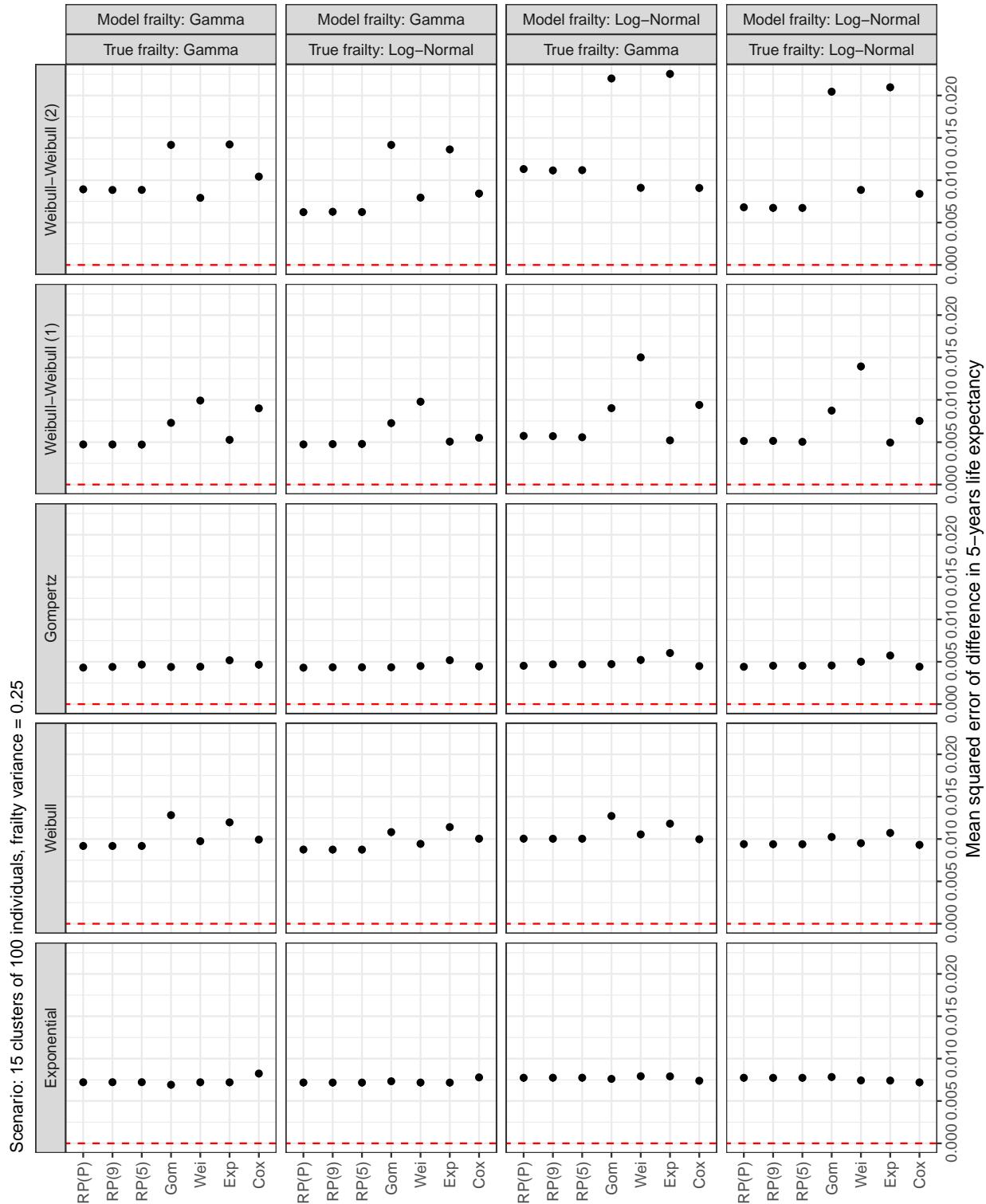


Figure B.17: Mean squared error of estimated difference in 5-years life expectancy, simulation study on model misspecification in survival models with shared frailty terms, scenario with 15 clusters of 100 individuals each and a small frailty variance.

# Appendix C

## Slides

### C.1 2017 SAfJR Conference

Direct likelihood maximisation using  
numerical quadrature to approximate  
intractable terms

Survival Analysis for Junior Researchers, Leicester,

UK

April 5<sup>th</sup>, 2017

Alessandro Gasparini<sup>1</sup>, Keith R Abrams<sup>1</sup>, Michael J Crowther<sup>1</sup>

## About me

- Currently a first-year PhD student at the Department of Health Sciences, University of Leicester
- Previous education: BSc in Statistics and Computing Technologies from University of Padua, Italy, and MSc in Biostatistics and Experimental Statistics from University of Milano-Bicocca, Italy
- PhD project:
  - joint modelling of longitudinal and survival data
  - modelling the visiting process
  - joint modelling of multiple biomarkers and their association with survival
  - model discrimination tools to evaluate multiple predictive biomarkers
  - application to health records data and cardiovascular epidemiology

2 / 17

## "... to approximate intractable terms"

- Case study 1: data on recurrent events, e.g. infections, cancer relapse, ...
- Model the within-patient correlation by assuming that it is the result of a latent patient-level effect, i.e. a frailty term
- Survival model with a shared frailty term:

$$h_{ij} = h_0(t) \exp(\beta^T Z_{ij} + v_i),$$

$$L_i = \int_{-\infty}^{+\infty} \prod_{j=1}^{n_i} [h_{ij}(x_{ij})]^{\Delta_{ij}} \exp\left[-\int_0^{x_{ij}} h_{ij}(t) dt\right] f_\theta(v_i) dv_i$$

- We need to choose a distribution  $f_\theta(v_i)$  for the frailty, and integrate it out

3 / 17

## "... to approximate intractable terms"

- Case study 2: data on repeated measurements of a biomarkers and survival
- Model the association between the biomarker and survival
- Joint model for longitudinal and survival data:

$$h(t|M_i(t), w_i) = h_0(t) \exp(\gamma^T w_i + \alpha m_i(t)),$$

$$L_i = \int_{-\infty}^{+\infty} P(T_i, d_i | b_i; \theta_t) \left[ \prod_{j=1}^{n_i} P(y_i(t_{ij}) | b_i; \theta_y) \right] P(b_i; \theta_{b_i}) db_i$$

- We need to integrate out the shared random effects

4 / 17

## "... using numerical quadrature ..."

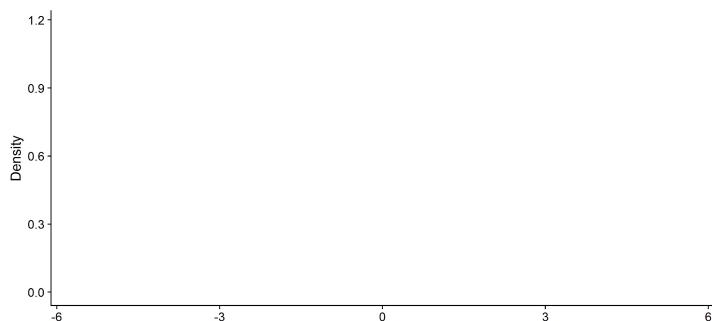
- Quadrature rule: approximation of the integral of a function, usually stated as a weighted sum of function values at specified points within the domain of integration
- n-points Gaussian quadrature rule:

$$\int_X f(x) dx = \sum_{i=1}^n w_i f(x_i)$$

- Goal: reach a given level of precision with the fewest possible function evaluations (the cost of computing n nodes is  $O(n^2)$  operations)

*But... what n do we pick?*

5 / 17



6 / 17

## "Direct likelihood maximisation

..."

- The expectation-maximisation [EM] algorithm:
  - E step: calculate  $E[l(\theta | X, Z)]$  given  $\theta^t$*
  - M step: find  $\text{argmax}(E[l(\theta | X, Z)]) = \theta^{t+1}$*
- Bayesian approach, using Markov Chain Monte Carlo [MCMC] techniques:
  - Choose prior distributions for the model parameters*
  - Derive posterior distributions for the model parameters*
- Direct likelihood maximisation:
  - Likelihood can be easily evaluated*
  - Many general purpose optimisers are readily available*

7 / 17

## Simulation studies

### Simulation N. 1

- *Aim:* evaluate the accuracy of Gaussian quadrature methods in settings where we do not need to use it, as we can derive analytical formulae

- Parametric survival model with shared Gamma frailty:

$$h_{ij}(t_{ij}|\alpha_i) = \alpha_i h_{ij}(t_{ij}) = \alpha_i p \lambda t_{ij}^{p-1} \exp(X_{ij}\beta)$$

- The unconditional contribution to the likelihood is:

$$L_i = \int_0^{+\infty} \alpha_i^{D_i} \prod_{j=1}^{n_i} \left[ S_{ij}(t_{ij})^{\alpha_i} (h_{ij}(t_{ij}))^{d_{ij}} \right] g(\alpha_i) d\alpha_i$$

8 / 17

## Simulation scenarios

- Weibull baseline hazard with shape  $p = 0.5$ , scale  $\lambda = 1$ , and shared Gamma-distributed frailty term
- 1,000 simulations per scenario
- number of clusters: {25, 50, 100, 200}, number of individuals per cluster: {25, 50, 100, 250, 500, 1000}
- treatment effect: {-0.50, 0.00, 0.50}
- variance of the frailty ( $\theta$ ): {0.25, 0.50, 1.00}

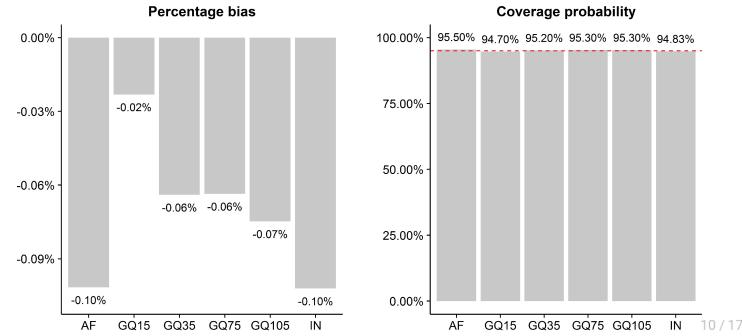
We compare estimates from:

1. model using analytical formulae
2. model using Gauss-Laguerre quadrature with {15, 35, 75, 105} nodes
3. model using Gauss-Kronrod quadrature, as implemented in the base-R `integrate()` function

9 / 17

## Results

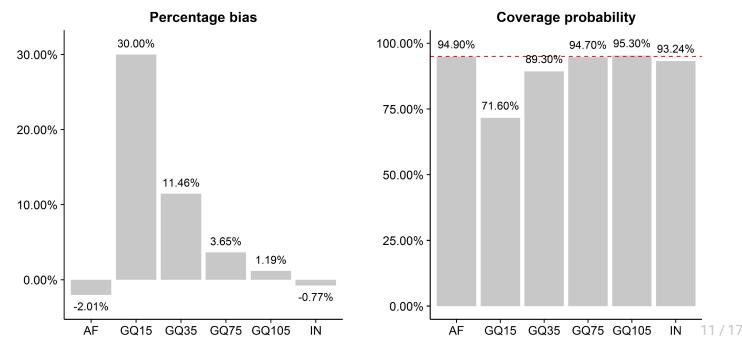
Scenario: 50 clusters of 250 individuals each, negative treatment effect (log-HR: -0.50), medium frailty variance (0.50).  
 Parameter of interest: treatment effect.



10 / 17

## Results

Scenario: 50 clusters of 250 individuals each, negative treatment effect (log-HR: -0.50), medium frailty variance (0.50).  
 Parameter of interest: frailty variance  $\theta$ .



11 / 17

## Simulation N. 2

- Aim: evaluate direct likelihood maximisation using quadrature in settings where it is not possible to derive analytical formulae

- Parametric survival model with a random treatment effect:

$$h_{ij}(t_{ij}|b_i) = p\lambda t_{ij}^{p-1} \exp[X_{ij}(\beta + b_i)]$$

- Cluster-specific contribution to the likelihood:

$$L_i = \int_{-\infty}^{+\infty} \left[ \prod_{j=1}^{n_i} h_{ij}(t_{ij})^{d_{ij}} S_{ij}(t_{ij}) \right] p(b_i) db_i$$

## Simulation scenarios

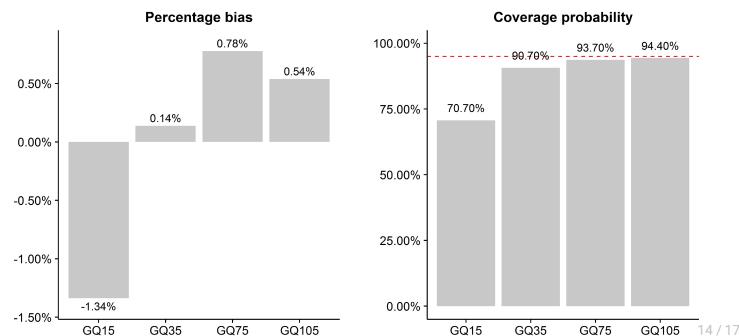
- Weibull baseline hazard with shape  $p = 1.5$ , scale  $\lambda = 3$ , and a random treatment effect
- 1,000 simulations per scenario
- number of clusters: {25, 50, 100, 200}, number of individuals per cluster: {25, 50, 100, 250, 500, 1000}
- treatment effect: {-0.50, 0.00, 0.50}
- standard deviation of the random effect ( $\sigma$ ): {0.25, 0.50, 1.00}

We compare estimates from models using Gauss-Hermite quadrature with {15, 35, 75, 105} nodes.

13 / 17

## Results

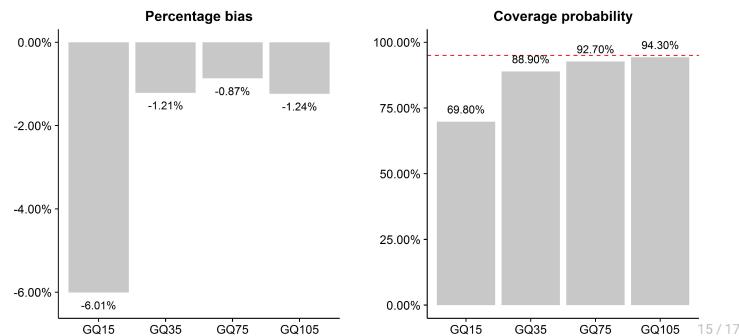
Scenario: 100 clusters of 50 individuals each, positive treatment effect (log-HR: 0.50), high random effect standard error (1.00). Parameter of interest: treatment effect.



14 / 17

## Results

Scenario: 100 clusters of 50 individuals each, positive treatment effect (log-HR: 0.50), high random effect standard error (1.00). Parameter of interest: random effect standard error  $\sigma$ .



15 / 17

## An example using R

- Step 1: get the quadrature nodes locations and weights:

```
library(pracma)
gl_rule = gaussLaguerre(35)
```

16 / 17

- Step 1: get the quadrature nodes locations and weights

- Step 2: code the negative log-likelihood function:

```
nloglik <- function(pars, data) {
  # loglik_i = ...
  loglik = sum(loglik_i)
  return(-loglik)
}
```

16 / 17

- Step 1: get the quadrature nodes locations and weights
  - Step 2: code the negative log-likelihood function
  - Step 3: minimise the negative log-likelihood:
- ```
optim(fn = nloglik, par = start)
```
- A plethora of general-purpose optimisers available in R: `nlm()`, `margLevAlg()`, `bobyqa()`, `nloptr()`, ...

16 / 17

- Step 1: get the quadrature nodes locations and weights
- Step 2: code the negative log-likelihood function
- Step 3: minimise the negative log-likelihood
- Step 4:



16 / 17

## Discussion

1. Gaussian quadrature works well, even in settings where it is not needed
2. Gaussian quadrature works well in settings where analytical formulae are not available
3. It is important to assess whether the number of quadrature nodes is appropriate
4. Direct likelihood maximisation is straightforward to implement
5. Future work:
  - extending to adaptive quadrature methods and other numerical integration methods (Monte-Carlo integration, importance sampling, ...)
  - exploring impact of model misspecification
  - developing an interactive tool for exploring simulations results

17 / 17

## References

- MacDonald IL, *Numerical maximisation of likelihood: a neglected alternative to EM?*. 2014, International Statistical Review, 82(2):296-308;
- Liu Q, and Pierce DA, *A note on Gauss-Hermite quadrature*. 1994, Biometrika, 81(3):624-629;
- Gautschi W, *Construction of Gauss-Christoffel quadrature formulas*. 1968, Mathematics of Computation, 22:251-270;
- Robert PC and Casella G, *Introducing Monte Carlo methods with R*. 2010, Springer-Verlag, New York;
- Crowther MJ, Look MP, and Riley RD, *Multilevel mixed effects parametric survival models using adaptive Gauss-Hermite quadrature with application to recurrent events and individual participant data meta-analysis*. 2014, Statistics in Medicine, 33(22):3844-3858;
- R Code and slides on my Github page: <https://github.com/ellessenne/SAFJR17>
- E-mail me at [ag475@leicester.ac.uk](mailto:ag475@leicester.ac.uk)

## C.2 2017 SAM Conference and ISCB Conference



UNIVERSITY OF  
LEICESTER

### Impact of model misspecification in survival models with frailties

Alessandro Gasparini<sup>1</sup>   Keith R Abrams<sup>1</sup>   Michael J Crowther<sup>1</sup>

<sup>1</sup> Department of Health Sciences, University of Leicester, Leicester, United Kingdom

38<sup>th</sup> Annual Conference of the International Society for Clinical Biostatistics

## Motivation

Survival data is commonly analysed by using parametric survival models or the Cox model. Nevertheless:

1. Subjects may be exposed to different baseline risk levels
2. Subjects may be clustered (clinical trials, geographical clusters, paired organs, twin studies, ...)
3. Subjects may experience repeated events (infections, cancer recurrence, ...)

An elegant and increasingly popular approach: including in the model a multiplicative random effect that allows accounting for this unobserved heterogeneity (i.e. a *frailty*).

Further details in Hougaard (2000) and Wienke (2010).

## Survival models with shared frailty

For the  $j^{\text{th}}$  individual in the  $i^{\text{th}}$  cluster:

$$h_{ij}(t) = h_0(t) \exp(X_{ij}\beta) u_i \quad (1)$$

$$h_{ij}(t) = h_0(t) \exp(X_{ij}\beta + w_i) \quad (2)$$

In a parametric world, we need to choose:

1. baseline hazard  $h_0(\cdot)$ : exponential, Weibull, Gompertz, flexible spline-based, ...
2. distribution of the frailty  $u_i$  (or  $w_i$ ): Gamma, log-Normal, positive stable, ...

## Misspecification

What we know:

1. The choice of the baseline hazard is often data-driven, using information criteria such as AIC and BIC
2. Relative risk estimates are insensitive to the correct specification of the baseline hazard (Rutherford, 2015)
3. Flexible parametric models (Royston, 2002) are robust to the choice of degrees of freedom for the spline function, assuming a sufficient number of degrees of freedom it is used (Rutherford, 2015)
4. The choice of frailty distribution has little impact on the estimation and testing of regression coefficients (Pickles, 1995)

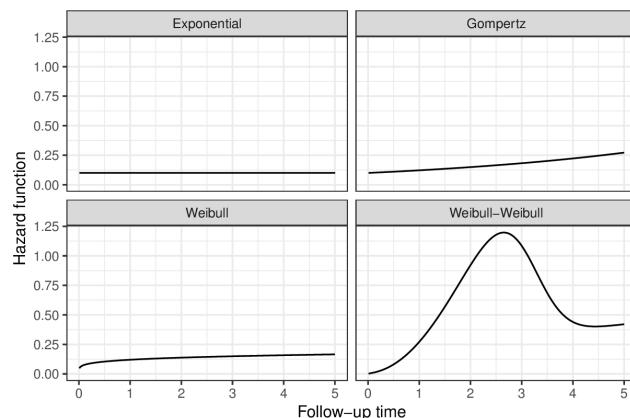
## A simulation study (1)

- ▶ Aim: assessing the impact of misspecifying the baseline hazard or the frailty distribution in a wide range of clinically and biologically plausible scenarios
- ▶ Data-generating mechanisms:
  - exponential baseline hazard
  - Weibull baseline hazard
  - Gompertz baseline hazard
  - mixture Weibull baseline hazard

ag475@leicester.ac.uk

4 of 19

## Data-generating baseline hazard functions



ag475@leicester.ac.uk

5 of 19

## A simulation study (2)

- ▶ Data-generating mechanisms:
  - Gamma and log-Normal frailty distribution
  - number of clusters (15, 50) and number of individuals per cluster (30, 100)
  - frailty variance (0.25, 0.50, 1.00)
  - log-treatment effect of -0.50
- ▶ Methods:
  - exponential, Weibull, Gompertz parametric survival models
  - Royston-Parmar model with 3 to 9 degrees of freedom
  - Royston-Parmar model using penalised likelihood
  - each model with Gamma or log-Normal frailty

ag475@leicester.ac.uk

6 of 19

### A simulation study (3)

- ▶ Estimands:
  - log-treatment effect
  - frailty variance
- ▶ Performance measures:
  - bias and percentage bias
  - coverage

ag475@leicester.ac.uk

7 of 19

## Results

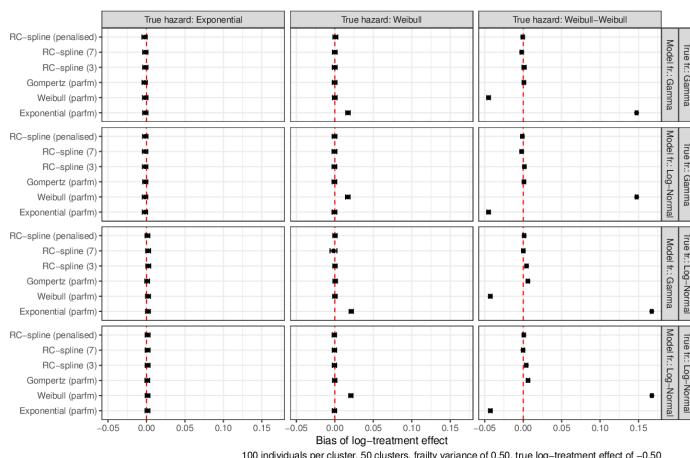
Fully factorial design: 96 simulated scenarios. We present:

- ▶ 50 clusters of 100 individuals each, frailty variance of 0.50
- ▶ 50 clusters of 30 individuals each, mixture Weibull baseline hazard

ag475@leicester.ac.uk

8 of 19

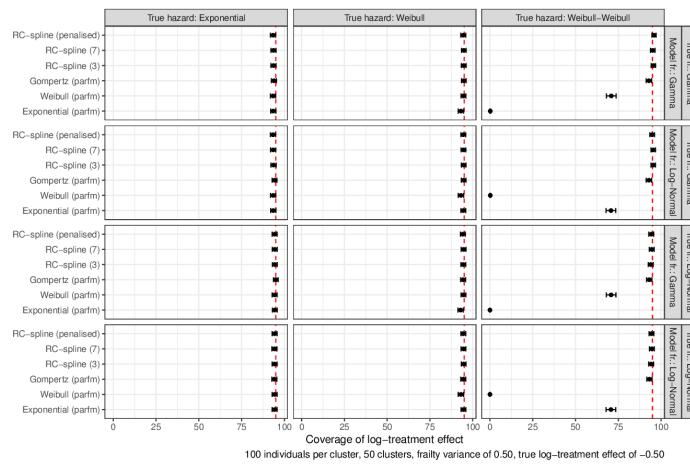
### Results: (1) bias of treatment effect



ag475@leicester.ac.uk

9 of 19

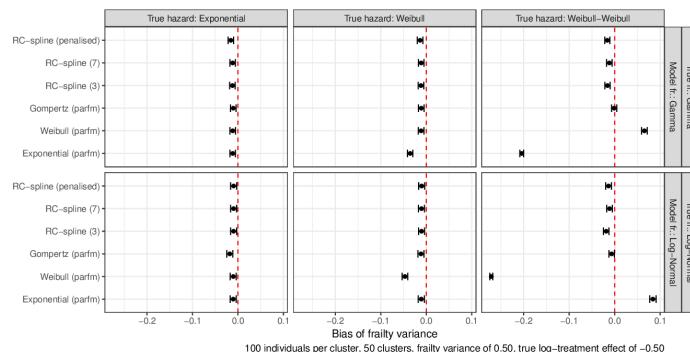
## Results: (i) coverage of treatment effect



ag475@leicester.ac.uk

10 of 19

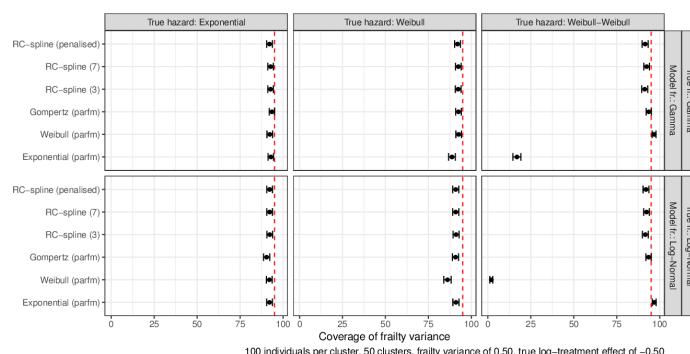
## Results: (i) bias of frailty variance



ag475@leicester.ac.uk

11 of 19

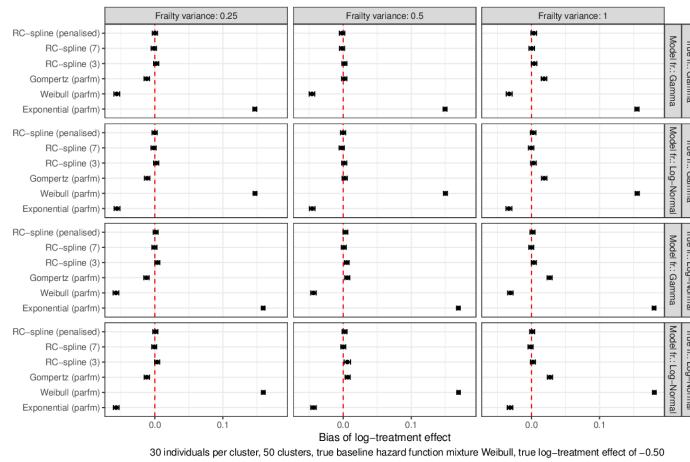
## Results: (i) coverage of frailty variance



ag475@leicester.ac.uk

12 of 19

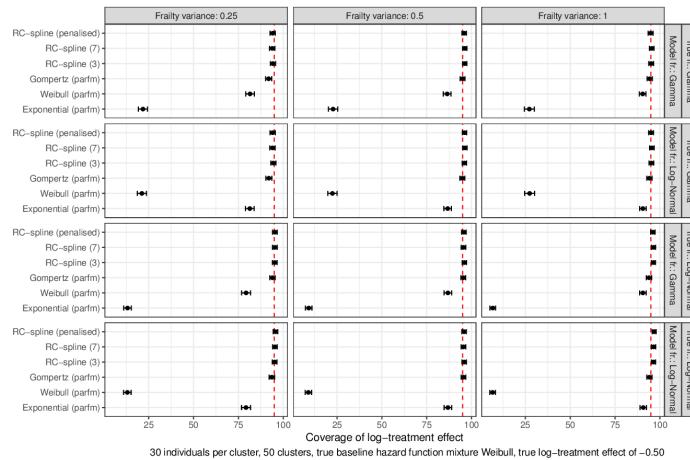
## Results: (2) bias of treatment effect



ag475@leicester.ac.uk

13 of 19

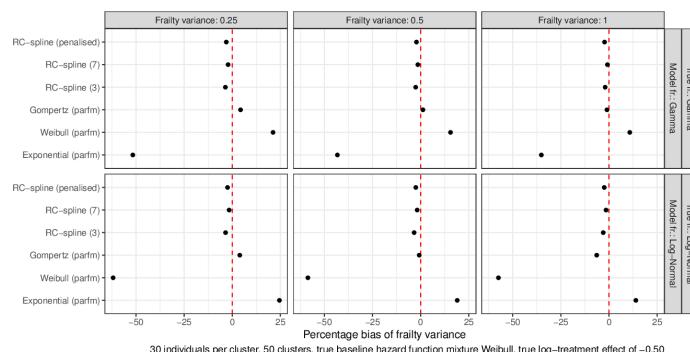
## Results: (2) coverage of treatment effect



ag475@leicester.ac.uk

14 of 19

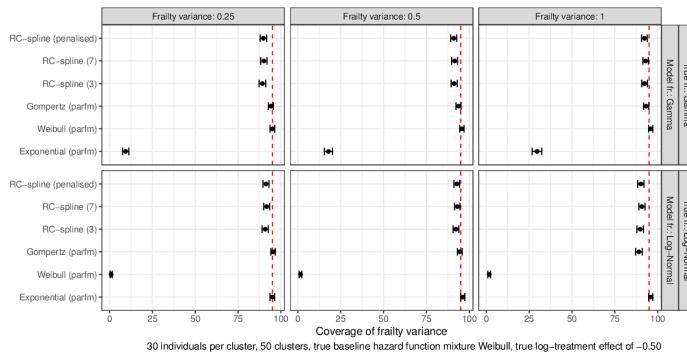
## Results: (2) percentage bias of frailty variance



ag475@leicester.ac.uk

15 of 19

## Results: (2) coverage of frailty variance



ag475@leicester.ac.uk

16 of 19

## Conclusions

- ▶ Misspecification of the baseline hazard can yield markedly biased regression coefficients, irrespectively of the frailty distribution
- ▶ Misspecification of the baseline hazard can also yield biased estimates of the frailty variance, even when the frailty distribution is well specified
- ▶ Misspecification of the frailty distribution has a negligible impact on bias of regression coefficients
- ▶ Flexible parametric models tend to be quite robust to model misspecification, using both full and penalised likelihood estimation procedures
- ▶ Further simulations will provide greater insight on the topic, especially on absolute risk predictions

ag475@leicester.ac.uk

17 of 19

## Next steps

1. Adding a simulation scenario with 1,000 clusters of 2 observations each: twin data
2. Adding marginal survival as estimand: ease of obtaining absolute risk predictions is one of the advantages of parametric models
3. Adding further comparisons with available software: shared frailty models with M-splines on the hazard scale estimated using penalised likelihood (R package *frailtypack*), ...

ag475@leicester.ac.uk

18 of 19

## References

- ▶ Hougaard P (2000). *Analysis of multivariate survival data*, Springer, New York
- ▶ Wienke A (2010). *Frailty models in survival analysis*, Chapman and Hall / CRC
- ▶ Rutherford MJ, Crowther MJ and Lambert PC (2015). *The use of restricted cubic splines to approximate complex hazard functions in the analysis of time-to-event data: a simulation study*, Journal of Statistical Computation and Simulation, 85(4):777-793
- ▶ Royston P and Parmar MK (2002). *Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects*, Statistics in Medicine, 21(15):2175-2197
- ▶ Pickles A and Crouchley R (1995). *A comparison of frailty models for multivariate survival data*, Statistics in Medicine, 14(13):1447-1461

ag475@leicester.ac.uk

19 of 19

## C.3 Students' Day at the 2017 ISCB Conference



UNIVERSITY OF  
LEICESTER

Exploring results from simulation studies  
interactively

A Gasparini<sup>1</sup>   IR White<sup>2</sup>   T Morris<sup>2</sup>   MJ Crowther<sup>1</sup>

<sup>1</sup> Department of Health Sciences, University of Leicester, Leicester, United Kingdom

<sup>2</sup> MRC Clinical Trials Unit at UCL, London, United Kingdom

Students' Day, 38<sup>th</sup> Annual Conference of the International Society for  
Clinical Biostatistics

## Key messages and questions

Key messages:

- ▶ simulation studies are being increasingly used
- ▶ dissemination of results is key
- ▶ interactive tools can effectively supplement reporting of simulation studies

Questions:

- ▶ What is your experience with presenting results from simulation studies?

ag475@leicester.ac.uk

1 of 21

## About me

- ▶ Currently: first-year PhD student at the University of Leicester, Leicester, United Kingdom
- ▶ Previously: BSc in Statistics and Computing Technologies from University of Padua, Italy; MSc in Biostatistics and Experimental Statistics from University of Milano-Bicocca, Italy
- ▶ PhD project:
  1. joint modelling of longitudinal and survival data
  2. survival models with random effects
  3. application to health records data and cardiovascular epidemiology

ag475@leicester.ac.uk

2 of 21

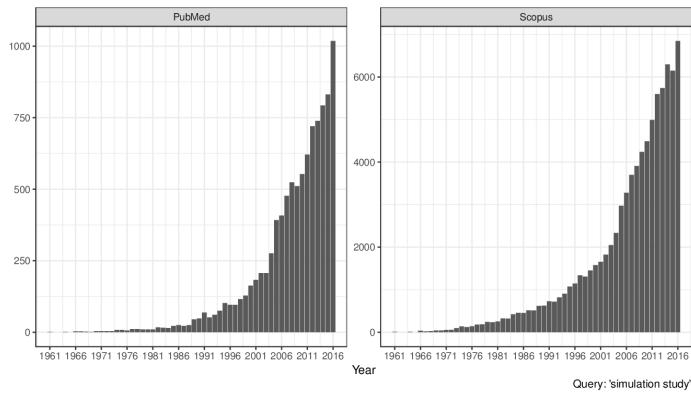
## Simulation studies are useful...

- ▶ Evaluating new statistical methods
- ▶ Evaluate large sample approximations
- ▶ Comparing the performance of different methods/models
- ▶ Assessing the impact of violating assumptions
- ▶ You name it!

ag475@leicester.ac.uk

3 of 21

...and common!



ag475@leicester.ac.uk

4 of 21

## Dissemination is key

1. Can drive practitioners and applied statisticians to methods that have been shown to perform well in their practical settings
2. Can guide researchers to develop new methods in a promising direction
3. Can provide insights into less established methods

ag475@leicester.ac.uk

5 of 21

## With great power comes great responsibility

- ▶ Increased availability of powerful computational tools surely contributed to the rise in popularity of simulation studies
- ▶ Adding multiple data-generating mechanisms or methods to compare is cheaper than ever (computationally speaking)
- ▶ Things can get out of control quickly<sup>1</sup>:

|                                      | Min | Max                |
|--------------------------------------|-----|--------------------|
| Number of methods evaluated          | 1   | 18                 |
| Number of estimands                  | 1   | 32                 |
| Number of factors varied across DGMs | 1   | 324                |
| Number of DGMs                       | 1   | $6 \times 10^{11}$ |

<sup>1</sup> Source: simulation studies published by Statistics in Medicine in 2015, unpublished data from the course *Using simulation studies to evaluate statistical methods* (White IR, Morris T and Crowther MJ)

ag475@leicester.ac.uk

6 of 21

## My experience

- ▶ Simulation study on the impact of misspecification in survival models with shared frailties
- ▶ Fully factorial design
- ▶ A priori factors that may affect the results:
  1. baseline hazard function
  2. sample size (number of clusters, number of individuals per cluster)
  3. variance of the frailty term
  4. distribution of the frailty term
- ▶ Massive number of simulation scenarios to summarise: how?

## ADMEP framework

A framework for harmonising reporting of simulation studies<sup>2</sup>:

- ▶ Aim(s)
- ▶ Data-generating mechanism(s)
- ▶ Method(s)
- ▶ Estimand(s)
- ▶ Performance measure(s)

Think of what you want to learn, and how: focusing on these aspect beforehand will make designing and reporting simulation studies easier.

<sup>2</sup>White IR, Morris T and Crowther MJ, unpublished

## Enter SiReX

- ▶ Interactive tools can supplement the ADMEP framework very effectively
- ▶ SiReX [saɪə(r)-ɛks], *Simulation Results eXplorer*
- ▶ Developed using R and Shiny
- ▶ Workflow:
  1. Upload your results
  2. Summary statistics are computed automatically
  3. Select a DGM and summary tables and plots are updated automatically
  4. Export summary statistics, tables, and plots for later use

## Demo

<https://goo.gl/lGU2Xc>



ag475@leicester.ac.uk

10 of 21

### Key messages and questions

#### Key messages:

- ▶ simulation studies are being increasingly used
- ▶ dissemination of results is key
- ▶ interactive tools can effectively supplement reporting of simulation studies

#### Questions:

- ▶ What is your experience with presenting results from simulation studies?

ag475@leicester.ac.uk

11 of 21

### Demo: landing page

A screenshot of a web browser window titled "SiReX". The address bar shows the URL "https://ag475.shinyapps.io/sirex-demo/". The main content area displays a dark sidebar with navigation links: "Introduction", "Data", "Explore results", and "Source code". To the right of the sidebar, there is a light blue tab with the text "Tab with some introductory stuff". The browser interface includes standard controls like back, forward, and search.

ag475@leicester.ac.uk

12 of 21

### Demo: load data

Load a tidy dataset with all results from the simulations study here:

Load data View data

Upload your .csv file

Browse... No file selected

Load demo data

Demo data is a simulated study consisting of different ways to handle missing covariates when fitting a Cox model (Little and Rubin 1986). One thousand datasets were simulated, each containing normally distributed covariates  $x_1$ ,  $x_2$ , and  $x_3$  and a time-dependent outcome. Both covariates had 40% of their values deleted independently of all other variables so the data became missing completely at random (Little and Rubin 2002). Each simulated dataset was analyzed in three ways. A Cox model was fit to the complete cases (CC). Then two methods of multiple imputation using chained equations (van Buuren, Boshuizen, and Knook 1999) were used. The MI\_LOGIT method multiply imputes the missing values of  $x_1$  and  $x_2$  with the outcome included as logit( $d$ ), where  $t$  is the survival time and  $d$  is the event indicator. The MI\_T method is the same except that logit( $t$ ) is replaced by  $t$  in the imputation model. The results are stored in long format, with variable 'dataset' identifying the simulated dataset number, string variable 'method' identifying the method used, variable 'b' holding the point estimate, and variable 'se' holding the SE.

ag475@leicester.ac.uk

13 of 21

### Demo: landing page

| method   | b          | se         | dataset |
|----------|------------|------------|---------|
| CC       | 0.7087682  | 0.1465100  | 1       |
| MI_T     | 0.6841483  | 0.1255043  | 1       |
| MI_LOGIT | 0.71247195 | 0.1410814  | 1       |
| CC       | 0.3485008  | 0.1599879  | 2       |
| MI_T     | 0.4060082  | 0.1409831  | 2       |
| MI_LOGIT | 0.4287003  | 0.1358549  | 2       |
| CC       | 0.6489075  | 0.1521568  | 3       |
| MI_T     | 0.5028701  | 0.13007181 | 3       |
| MI_LOGIT | 0.5640491  | 0.1168512  | 3       |
| CC       | 0.4320534  | 0.1262853  | 4       |
| MI_T     | 0.4673285  | 0.1177011  | 4       |
| MI_LOGIT | 0.4921503  | 0.1179779  | 4       |

ag475@leicester.ac.uk

14 of 21

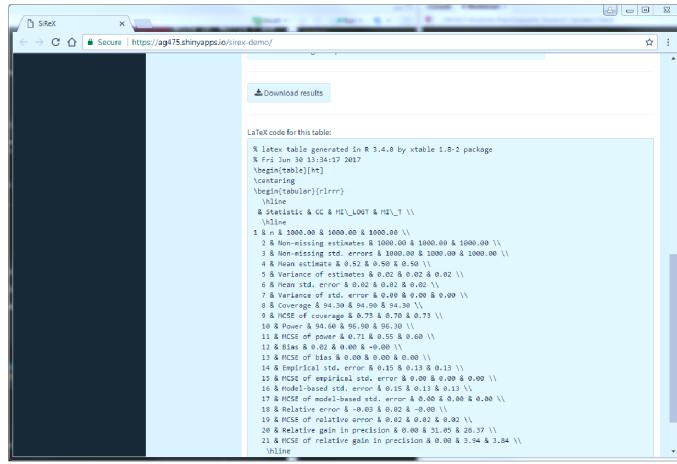
### Demo: table of summary results

| Statistic                      | CC          | MI_LOGIT    | MI_T        |
|--------------------------------|-------------|-------------|-------------|
| n                              | 1000.000000 | 1000.000000 | 1000.000000 |
| Non-missing estimates          | 1000.000000 | 1000.000000 | 1000.000000 |
| Non-missing std. errors        | 1000.000000 | 1000.000000 | 1000.000000 |
| Mean estimate                  | 0.516766    | 0.509023    | 0.498409    |
| Variance of estimates          | 0.022836    | 0.017426    | 0.018071    |
| Mean std. error                | 0.021437    | 0.018209    | 0.017912    |
| Variance of std. error         | 0.000024    | 0.000027    | 0.000025    |
| Coverage                       | 94.300000   | 94.900000   | 94.300000   |
| MCSE of coverage               | 0.733151    | 0.695694    | 0.733151    |
| Power                          | 94.000000   | 96.900000   | 96.300000   |
| MCSE of power                  | 0.174731    | 0.548078    | 0.599517    |
| Bias                           | 0.016766    | 0.000823    | -0.001191   |
| MCSE of bias                   | 0.004779    | 0.004174    | 0.004251    |
| Empirical std. error           | 0.151115    | 0.132006    | 0.134428    |
| MCSE of empirical std. error   | 0.003814    | 0.002933    | 0.000907    |
| Model-based std. error         | 0.147096    | 0.134941    | 0.133835    |
| MCSE of model-based std. error | 0.000527    | 0.000605    | 0.000588    |
| Relative error                 | 0.020594    | 0.022233    | -0.004412   |

ag475@leicester.ac.uk

15 of 21

## Demo: table of summary results



```

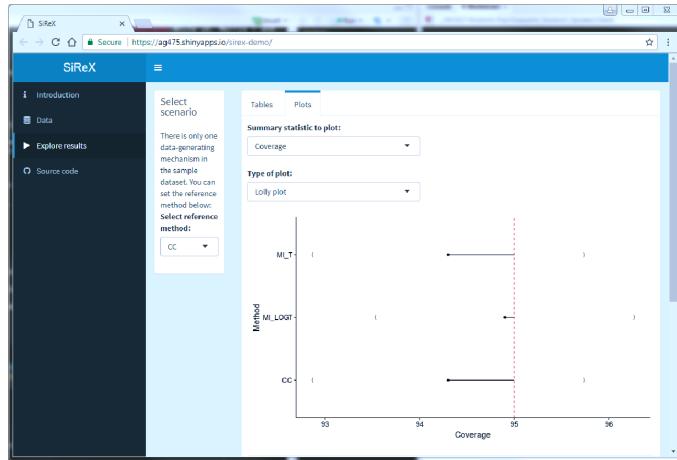

\begin{table}[ht]
\caption{Summary statistics}
\centering
\begin{tabular}{l}
\hline
Statistic & Value \\
\hline
1 & 1000.00 & 1000.00 & 1000.00 \\
2 & Non-missing estimator & 1000.00 & 1000.00 & 1000.00 \\
3 & Non-missing std. errors & 1000.00 & 1000.00 & 1000.00 \\
4 & Mean estimate & 0.92 & 0.90 & 0.89 \\
5 & Variance of estimate & 0.02 & 0.02 & 0.02 \\
6 & Mean std. error & 0.02 & 0.02 & 0.02 \\
7 & Variance of std. error & 0.00 & 0.00 & 0.00 \\
8 & Coverage & 94.39 & 95.00 & 94.39 \\
9 & Power & 0.72 & 0.73 & 0.71 \\
10 & Power & 94.40 & 96.90 & 95.30 \\
11 & MCSE of power & 0.72 & 0.55 & 0.40 \\
12 & Bias & 0.00 & 0.00 & 0.00 \\
13 & MCSE of bias & 0.00 & 0.00 & 0.00 \\
14 & Empirical std. error & 0.15 & 0.13 & 0.13 \\
15 & MCSE of empirical std. error & 0.00 & 0.00 & 0.00 \\
16 & MCSE of relative error & 0.15 & 0.13 & 0.13 \\
17 & MCSE of relative error & 0.00 & 0.00 & 0.00 \\
18 & Relative error & 0.03 & 0.02 & 0.00 \\
19 & MCSE of relative error & 0.02 & 0.02 & 0.02 \\
20 & Relative gain in precision & 0.00 & 51.95 & 26.37 \\
21 & MCSE of relative gain in precision & 0.00 & 5.54 & 3.84 \\
\hline
\end{tabular}


```

ag475@leicester.ac.uk

16 of 21

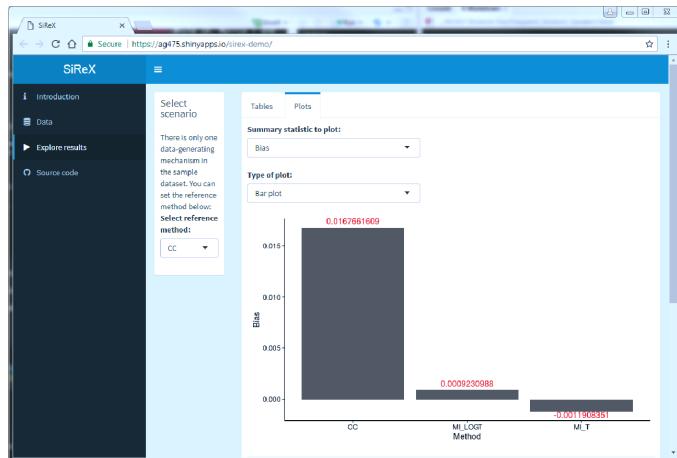
## Demo: plot of summary results



ag475@leicester.ac.uk

17 of 21

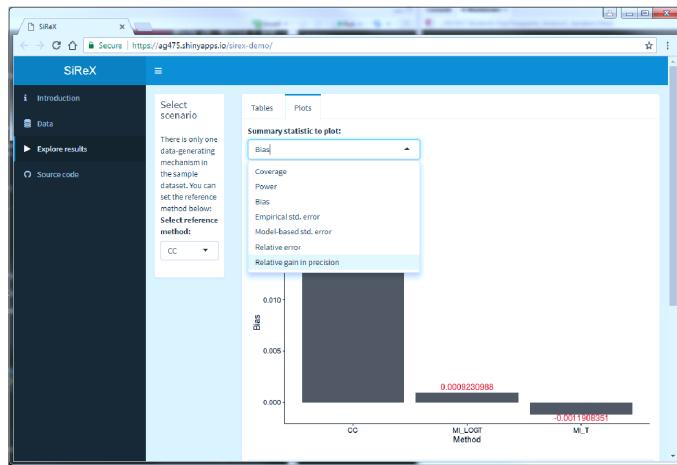
## Demo: plot of summary results



ag475@leicester.ac.uk

18 of 21

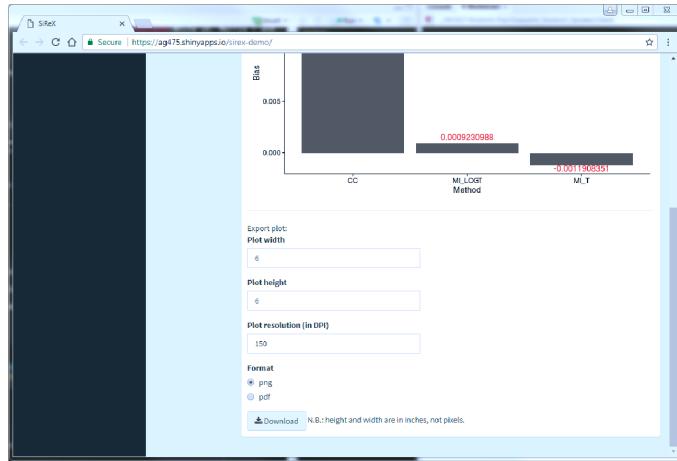
### Demo: plot of summary results



ag475@leicester.ac.uk

19 of 21

### Demo: plot of summary results



ag475@leicester.ac.uk

20 of 21

### Key messages and questions

#### Key messages:

- ▶ simulation studies are being increasingly used
- ▶ dissemination of results is key
- ▶ interactive tools can effectively supplement reporting of simulation studies

#### Questions:

- ▶ What is your experience with presenting results from simulation studies?

ag475@leicester.ac.uk

21 of 21



## Appendix D

### Manuscript draft

**Research Article****Statistics  
in Medicine**

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000

**The impact of model misspecification in frailty survival models****Alessandro Gasparini<sup>1,\*</sup>, Keith R. Abrams<sup>1</sup> and Michael J. Crowther<sup>1</sup>**

Survival models incorporating random terms to account for unmeasured heterogeneity are being used more and more in biostatistical research. Specifically, unmeasured covariates whose lack of inclusion in the model would lead to biased, inefficient results are commonly modelled by including a frailty term. Those frailty terms may be subject-specific or cluster/group-specific and are usually assumed to follow either a Gamma or log-Normal distribution. In the context of parametric frailty models, little is known about the impact of misspecifying the baseline hazard, or the frailty distribution. Therefore, we aim to quantify the impact of such misspecification in a wide variety of clinically plausible scenarios. We simulate survival data assuming various baseline hazard functions, including complex distributions with turning points, and furthermore assess the impact of sample size, magnitude of the frailty variance, and frailty distribution. We estimate the model coefficients via maximum likelihood, using numerical integration to handle intractable terms. Models compared include standard distributions, such as the exponential, Weibull and Gompertz, and spline-based approaches including the Royston-Parmar flexible parametric model and penalized splines on the log-hazard scale. Our simulations indicate the importance of assessing model fit with respect to the baseline hazard function, showing that bias can occur when it is misspecified. We illustrate our conclusions on a dataset from an international trial on treatment for acute ischaemic stroke.

Copyright © 2017 John Wiley &amp; Sons, Ltd.

**Keywords:** Survival; Frailty; Simulation study; Misspecification; Numerical integration; Royston-Parmar model;

**1. Introduction**

The standard, most common approach in medical research when dealing with time to event data consists in fitting a Cox proportional hazards model, where the baseline hazard is left unspecified and relative effect estimates are frequently reported as the main quantities of interest. However, often it is of interest to obtain absolute measures of risk: in that context, modelling the baseline hazard is necessary, and it can be achieved by using standard parametric survival models with a simple parametric distribution (such as the exponential, Weibull, or Gompertz distribution) or by using the flexible parametric modelling approach of Royston and Parmar [1] to better capture the shape of complex hazard functions. The latter approach requires choosing the number of degrees of freedom for the spline term used to approximate the baseline hazard: in practice, sensitivity analyses and information criteria (AIC, BIC) have been used to select the best model. Recently, Rutherford et al. [2] showed via simulation studies that, assuming a sufficient number of degrees of freedom is used, the approximated hazard function given by restricted cubic splines fit well for a number of complex hazard shapes and the hazard ratios estimation is insensitive to the correct specification of the baseline hazard.

Moreover, it is common to encounter clustered survival data where the overall study population can be divided into heterogeneous clusters of homogeneous observations. Examples are: multi-centre clinical trial data, individual-patient data meta-analysis, observational data with geographical clusters, .... As a consequence, survival times of individuals within cluster are likely to be correlated and need to be analysed as such. Analogously, correlated data may emerge as

---

<sup>1</sup> Department of Health Sciences, University of Leicester, Leicester, United Kingdom

\*Correspondence to: [ag475@leicester.ac.uk](mailto:ag475@leicester.ac.uk)

a consequence of recurrent events, i.e. events that may occur repeatedly within the same study subject. Unfortunately, covariates that contribute to explaining the heterogeneity between clusters are often not measured, e.g. for economical reasons. Hence, the frailty approach aims to account for the unobserved heterogeneity by including a random effect that acts multiplicatively on the baseline hazard and can be shared within cluster.

Univariate frailty models have been first proposed by Vaupel *et al.* [3] and Lancaster [4], and further discussed by Hougaard [5, 6] with specific focus on the frailty distribution. The Gamma distribution is widely used, being mathematically very convenient; the inverse Gaussian distribution is also common. A main difference between the two is that a Gamma frailty yields a time-independent heterogeneity, while an inverse Gamma frailty yields a decreasing heterogeneity that makes the population more homogeneous as time goes by. Additionally, Hougaard presents a family of distributions with infinite mean, such as the reciprocal Gamma distribution and the positive stable distribution. It is possible to use log-normal frailty as well, but that leads to analytically intractable formulae that require numerical quadrature or stochastic integration. Hougaard also extended the univariate frailty approach to accommodate frailty terms shared within cluster [7], which results more attractive when considering repeated event-times that are conditionally independent given the frailty [8]. Rondeau *et al.* further extended the shared frailty model allowing for hierarchical clustering of the data by including two nested frailty terms [9], by allowing to study both heterogeneity across trials and treatment-by-trial heterogeneity via additive frailty models [10], and by jointly modelling recurrent events and a dependent terminal event to jointly study the evolution of the two processes or account for violations of the proportional hazard assumption [11–13]. More recently, Crowther *et al.* [14] generalised the frailty approach by allowing for the inclusion of any number of random effects in a parametric survival model; under that framework, a survival model with a shared log-normal frailty term can be seen as a survival model with a random intercept, shared between individuals that belong to the same cluster. It follows that, under a more general formulation, a mixed effects survival model can include not only a random intercept (in which case it is equivalent to a model with a log-normal frailty) but multiple random effects as well.

As mentioned before, flexible parametric survival models are a robust alternative to standard parametric survival models when the shape of the hazard function is complex; using a sufficient number of degrees of freedom, e.g. 2 or more, the spline-based approach is able to capture the underlying shape of the hazard function with minimal bias. AIC and BIC can guide the choice of the best fitting model, but they tend to agree to within 1 or 2 degrees of freedom in practice [2]. Analogously, the impact of the choice of a particular parametric frailty distribution on the regression coefficients is minimal [15]. Conversely, little is known about the impact of misspecifying the baseline hazard in survival models with frailty terms. We aim therefore with this work to assess the impact of misspecifying the baseline hazard or the distribution of the frailty on both the regression coefficients and the estimated frailty variance by comparing a set of models under different data generating scenarios. We compare standard parametric survival models with frailties, models with flexible baseline hazard, and models with flexible baseline hazard and a penalty for the complexity of the spline.

The remaining of this manuscript is laid out as follows. In Section 2, we introduce survival models with frailties. In Section 3, we present the simulation study and its aims, data-generating mechanisms, methods, estimands, and performance measures employed. In Section 4, we present the results of our simulations, and in Section 5 we compared the different models in real settings using a clinical dataset from a trial on treatment for patients with acute ischaemic stroke. Finally, we conclude the paper in Section 6 with a discussion.

### 1.1. Acute ischaemic stroke data

The International Stroke Trial (IST) is a randomised trial of up to 14 days of aspirin, subcutaneous heparin, both, or neither among 19,435 patients with acute ischaemic stroke [16, 17]. The aim of the trial was to obtain clinical evidence on efficacy and safety of aspirin and heparin in patients starting therapy as soon as possible after the acute stroke event. Half of the patients were allocated to heparin, and half to avoid heparin; using a factorial design, half of the patients were allocated to aspirin, and half to avoid aspirin. Consequently, approximately a quarter of patients were allocated to aspirin and heparin, aspirin only, heparin only, or neither. The primary outcomes were death within 14 days and death or dependency at 6 months. The trial included 467 hospital in 36 countries.

## 2. Survival models with unobserved heterogeneity

Methods in survival analysis of common use implicitly assume that populations are homogenous (e.g. that all the patients have the same risk of event). However, that assumption is often not satisfied: sometimes it is impossible to include all relevant risk factors, or maybe such risk factors are not known at all. The violation of the homogeneity assumption leads to unobserved heterogeneity. A survival model with a frailty term is therefore a survival model with a (potentially shared) random effect, named frailty, to account for unobserved heterogeneity. The frailty term acts multiplicatively on the baseline

# Statistics in Medicine

Gasparini *et al.* (2017)

hazard, it is assumed to follow some distribution, and represents a straightforward way of modelling dependence between event times. A comprehensive treatment of frailty models in survival analysis is given in Wienke [18].

A univariate frailty model has the general form

$$h(t|\alpha) = \alpha h_0(t),$$

where  $\alpha$  is a non-observed frailty effect and  $h_0(t)$  is the baseline hazard function. The random variable  $\alpha$ , the frailty term, is chosen to have a distribution with expectation  $E(Z) = 1$  and variance  $V(\alpha) = \sigma^2$ .  $V(\alpha)$  is interpretable as a measure of heterogeneity across the population in baseline risk: as  $\sigma^2$  increases the values of  $\alpha$  are more dispersed, with greater heterogeneity in  $\alpha h_0(t)$ . Introducing observed covariates into the model:

$$h(t|X, \alpha) = \alpha h_0(t) \exp(X\beta) = \alpha h(t|X),$$

with  $X$  and  $\beta$  covariates and regression coefficients, respectively.

Generalising the univariate frailty model to the allow clusters of study observations to share a frailty term, we obtain the model

$$h_{ij}(t|\alpha_i) = \alpha_i h(t|X_{ij}),$$

with  $i$  indexing the cluster and  $j$  indexing the cluster subject; more details on shared frailty models in [19]. From now on, we will be focusing on shared frailty models.

Different choices for the frailty distribution are possible. Assigning a probability distribution implies that the frailty can be integrated out of the likelihood function. After this integration, the likelihood can be maximized in the usual way if an explicit form of it exists. Otherwise, more sophisticated approaches like numerical integration or Markov Chain Monte Carlo methods need to be applied. The most often used frailty distributions are the gamma and the log-normal distribution. The positive stable and the inverse Gaussian distribution are also common.

Assuming that the frailty  $\alpha_i$  has a Gamma distribution is convenient: it has the appropriate range  $(0, \infty)$  and it is mathematically tractable. A Gamma distribution with parameters  $a$  and  $b$  has density

$$f(x) = \frac{x^{a-1} \exp(-x/b)}{\Gamma(a)b^a};$$

by choosing  $a = 1/\theta$  and  $b = \theta$  the resulting distribution has expectation 1 and finite variance  $\theta$ . In these settings, the frailty can be integrated out of the likelihood analytically. Estimating such model becomes therefore straightforward, which likely contributed to the popularity of Gamma frailty models.

Together with the Gamma distribution, the log-normal distribution is the most commonly used frailty distribution, given its strong ties to random effect models. Nevertheless, the distribution of a log-normal frailty cannot be integrated out of the likelihood analytically and therefore requires more complex estimation procedures involving numerical integration (taking the maximum likelihood approach), or Markov Chain Monte Carlo methods [20, 21].

### 3. Simulation study

#### 3.1. Aims

The aim of this simulation study consists in assessing the impact of misspecifying the baseline hazard or the frailty distribution in shared frailty models, under a set of different, clinically plausible data-generating mechanisms.

#### 3.2. Data-generating mechanisms

We simulated data under five different baseline hazard functions using the inversion method [?] and one of its extensions to accommodate complex distributions with turning points [22]. We chose the Exponential, Weibull, Gompertz hazard function, and a two different two-components Weibull-Weibull mixture distribution; specifically, we used a baseline exponential hazard with scale  $\lambda = 0.3$ , a Weibull with scale  $\lambda = 0.5$  and shape  $p = 0.6$ , a Gompertz with scale  $\lambda = 0.1$  and shape  $\gamma = 0.5$ , a two-components mixture Weibull with scale parameters  $\lambda_1 = 0.5$ ,  $\lambda_2 = 0.3$ , shape parameters  $p_1 = 2.5$  and  $p_2 = 1.3$ , and mixing parameter  $\pi = 0.8$ , and two-components mixture Weibull-Weibull with scale parameters  $\lambda_1 = \lambda_2 = 1.0$ , shape parameters  $p_1 = 1.5$  and  $p_2 = 0.5$ , and mixing parameter  $\pi = 0.5$  (Figure 1, panel A). Then, for each baseline hazard function, we generated clustered data assuming 15 clusters of (30, 100) individuals each, 50 clusters of (30, 100) individuals each, or 1000 clusters of 2 individuals each. We included a binary treatment variable  $X \sim Bin(1, 0.5)$  with associated log-hazard ratio of  $-0.5$  and cluster-specific frailty terms  $\alpha_i$  following either a Gamma or a log-normal distribution with variance  $\theta$ , with  $\theta \in \{0.25, 0.75, 1.25\}$ . Finally, we generated an event indicator variable  $d$  by applying administrative censoring at 5 years. The true marginal survival functions corresponding to these simulated settings are depicted in Figure 1, panel B.

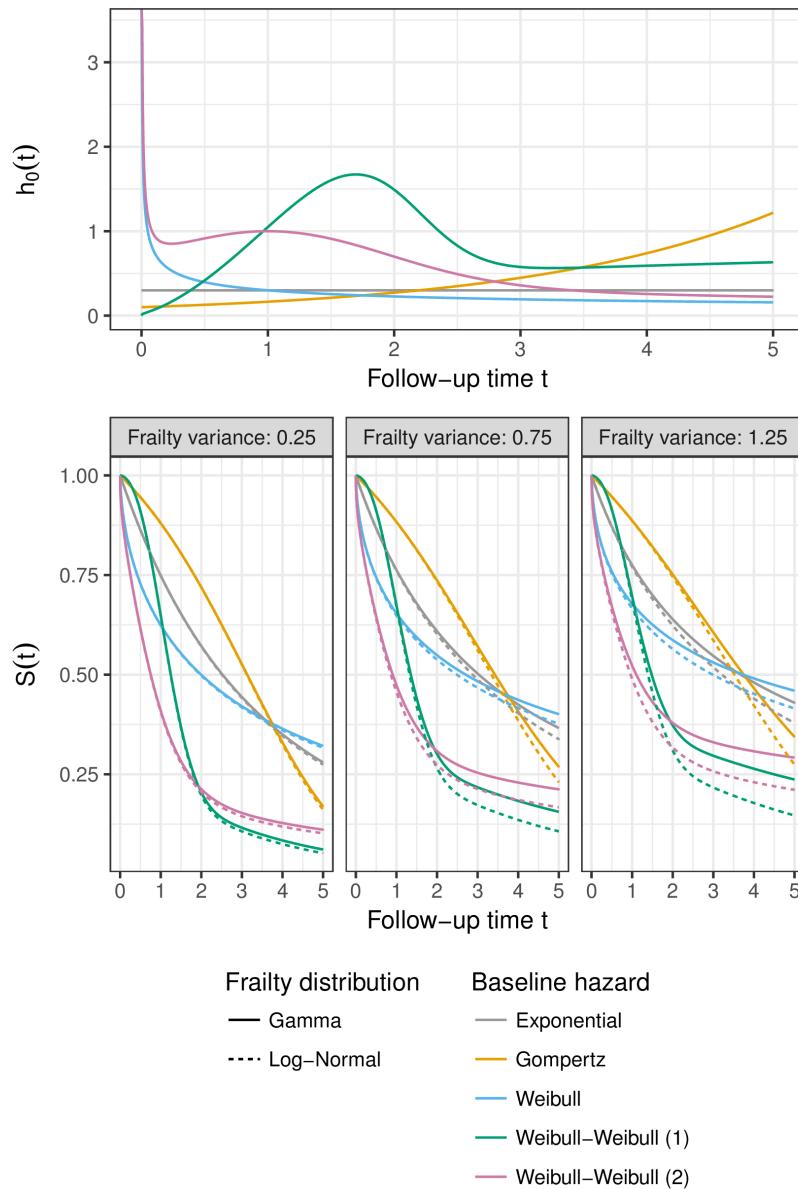


Figure 1. Baseline hazard functions chosen for this simulation study (panel A) and relative survival functions (panel B).

### 3.3. Methods

The different models we compared with this simulation study are introduced in this Section and summarised in Table 1. In particular, we fit a semiparametric Cox model with a shared frailty term (Section 3.3.1), a fully parametric survival model

# Statistics in Medicine

Gasparini *et al.* (2017)

with a shared frailty term (Section 3.3.2), a flexible parametric survival model with a shared frailty term and estimated using full likelihood and a varying number of degrees of freedom for the spline term (Section 3.3.3) or using penalised likelihood (Section 3.3.4). We estimate all models using available software, to provide insights into methods that are generally available to applied researchers.

**Table 1.** Models compared in this simulation study

| Model        | Baseline hazard                      | Fraity                         |
|--------------|--------------------------------------|--------------------------------|
| Model 1, 2   | Unspecified (Cox)                    | Gamma (1) and log-normal (2)   |
| Model 3, 6   | Exponential                          | Gamma (3) and log-normal (6)   |
| Model 4, 7   | Weibull                              | Gamma (4) and log-normal (7)   |
| Model 5, 8   | Gompertz                             | Gamma (5) and log-normal (8)   |
| Model 9, 13  | Flexible, 3 df                       | Gamma (9) and log-normal (13)  |
| Model 10, 14 | Flexible, 5 df                       | Gamma (10) and log-normal (14) |
| Model 11, 15 | Flexible, 7 df                       | Gamma (11) and log-normal (15) |
| Model 12, 16 | Flexible, 9 df                       | Gamma (12) and log-normal (16) |
| Model 17, 18 | Flexible, penalising the spline term | Gamma (17) and log-normal (18) |

**3.3.1. Semiparametric survival models with a shared frailty term** First, we estimate a semiparametric Cox model with a shared-frailty term. The model is formulated as

$$h_{ij}(t|\alpha_i) = \alpha_i h_0(t_{ij}) \exp(X_{ij}\beta), \quad (\text{models 1-2})$$

with  $\alpha_i$  following a Gamma or log-normal distribution, and  $h_0(t_{ij})$  left unspecified. The models are estimated using the R packages `coxme` and `frailtyEM`.

**3.3.2. Fully parametric survival models with a shared frailty term** Then, we estimate a shared-frailty model with a parametric baseline hazard. The model is formulated as

$$h_{ij}(t|\alpha_i) = \alpha_i h_0(t_{ij}) \exp(X_{ij}\beta), \quad (\text{models 3-8})$$

with  $\alpha_i$  following a Gamma or log-normal distribution, and  $h_0(t_{ij})$  following an exponential, Weibull, or Gompertz distribution. The models are estimated using the R package `parfm`.

**3.3.3. Flexible parametric survival model with a shared frailty term, estimated using full likelihood** This model is based on the flexible parametric survival model framework of Royston and Parmar [1], further generalised by Liu, Pawitan, and Clements [23]. The model is formulated on the log-cumulative hazard scale, and the baseline log-cumulative hazard is smoothed using a restricted cubic spline  $s(z; \gamma)$  with a given number of knots  $f \in \{3, 5, 7, 9\}$ :

$$\log H(t|x; \theta) = s(z; \gamma) + X\beta + \log(\alpha_i) \quad (\text{models 9-16})$$

The model parameters are estimated using the maximum likelihood method.

**3.3.4. Flexible parametric survival model with a shared frailty term, estimated using penalised likelihood** This model is formulated analogously as the models of Section 3.3.3, but applying a penalty to the likelihood function to prevent overfitting the data [23]. The penalised log-likelihood is computed as the logarithm of the likelihood, minus a penalty term:

$$\log pl_i(\theta|x_{ij}, t_{ij}, \psi) = \log l_i(\theta|x_{ij}, t_{ij}) - \frac{1}{2} \sum_{j=1}^m \psi_j P_j(\theta_j), \quad (\text{models 17-18})$$

where  $m$  is the number of penalised smoothers in the model,  $\psi_j$  is the smoothing parameter for the  $j$ -th smoother, and  $P_j(\theta_j)$  is the roughness penalty function. We used the default choice for the smoother, i.e. a spline-based smoother. Model estimation follows a two-steps procedure:

1. estimate initial values from (i) fitting a Cox regression model to predict the survival function and (ii) fitting a generalised additive model with the transformation of the predicted survival as outcome to obtain  $\hat{\theta}^{(0)}$ ;
2. estimate optimal smoothing parameters and regression parameters, starting from the initial values  $\hat{\theta}^{(0)}$  and by choosing the optimal  $\psi$  that minimises the likelihood-based cross-validation or the generalised Bayesian information criterion.

# Statistics in Medicine

Gasparini *et al.* (2017)

## 3.4. Estimands

The first estimand of interest is the regression coefficient  $\beta$  associated with the simulated binary treatment. This coefficient can be interpreted as log-treatment effect, and it would be interesting to see if and how misspecification of the baseline hazard or frailty distribution affects relative risk estimates. Second, we compare estimates of the frailty variance obtained from each model - a quantity often used to quantify dependence between clustered observations. Finally, we compare to compare two measures of absolute risk: marginal survival difference at time  $t$ , defined as  $S(t)_{\text{diff}} = S(t|x=1) - S(t|x=0)$ , and integrated marginal survival difference, defined as  $iS_{\text{diff}} = iS(x=1) - iS(x=0)$ . The former is obtained by fixing the time  $t$  (I am using 1, 2, 3, and 4 years), and then integrating out the frailty term from the conditional survival function. Conversely, the latter requires integrating the marginal survival function for both treatment groups and then computing their difference. It is estimated as follows:

1. Estimating marginal survival for a treatment group at 1000 equally spaced values of follow-up time  $t$ ;
2. Fitting an interpolating natural spline over the 1000 estimates from step (1);
3. Integrating the resulting spline function using Gauss-Kronrod quadrature;
4. Finally, computing the difference of the integrated marginal survival for the two treatment groups.

The integral of a survival function can be interpreted as life expectancy; hence, the quantity I am computing can be interpreted as the difference in 5-years life expectancy between treated and untreated individuals.

## 3.5. Performance measures

The first performance measure of interest is bias, representing the precision of an estimator and formally defined as  $E(\hat{\beta}) - \beta$ , with  $\hat{\beta}$  estimates of  $\beta$ . Second, we are interested in coverage, i.e. the proportion of times the  $100 \times (1 - \alpha)\%$  confidence interval  $\hat{\beta} \pm Z_{1-\alpha/2} \times SE(\hat{\beta})$  includes the true value  $\beta$ . This allow to assess whether the empirical coverage rate approaches the nominal coverage rate ( $100 \times (1 - \alpha)\%$ ), to properly control the type I error rate for testing a null hypothesis of no effect. Finally, we present as measure of overall accuracy and the mean squared error, defined as the sum of bias and variability:  $(\hat{\beta} - \beta)^2 + (SE(\hat{\beta}))^2$ .

We also report on convergence rates of each model, and we include Monte Carlo standard errors for bias and coverage to quantify the uncertainty in estimating such performance measures [24].

## 4. Results

## 5. Application to IST data

## 6. Discussion

## References

- [1] Royston P, Parmar MK. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine* 2002; **21**(15):2175–2197, doi:10.1002/sim.1203. URL <https://doi.org/10.1002/sim.1203>.
- [2] Rutherford MJ, Crowther MJ, Lambert PC. The use of restricted cubic splines to approximate complex hazard functions in the analysis of time-to-event data: a simulation study. *Journal of Statistical Computation and Simulation* oct 2015; **85**(4):777–793, doi:10.1080/00949655.2013.845890. URL <https://doi.org/10.1080/00949655.2013.845890>.
- [3] Vaupel JW, Manton KG, Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* 1979; **16**(3):439–454, doi:10.2307/2061224. URL <http://dx.doi.org/10.2307/2061224>.
- [4] Lancaster T. Econometric methods for the duration of unemployment. *Econometrica* jul 1979; **47**(4):939–956, doi:10.2307/1914140. URL <http://www.jstor.org/stable/1914140>.
- [5] Hougaard P. Life table methods for heterogeneous populations: distributions describing the heterogeneity. *Biometrika* 1984; **71**(1):75–83, doi:10.2307/2336399. URL <http://www.jstor.org/stable/2336399>.
- [6] Hougaard P. Survival models for heterogeneous populations derived from stable distributions. *Biometrika* 1986; **73**(2):387–396, doi:10.1093/biomet/73.2.387. URL <http://www.jstor.org/stable/2336215>.

# Statistics in Medicine

Gasparini *et al.* (2017)

- 
- [7] Hougaard P. A class of multivariate failure time distributions. *Biometrika* 1986; **73**(3):671–678, doi:10.1093/biomet/73.3.671. URL <https://doi.org/10.1093/biomet/73.3.671>.
- [8] Hougaard P. Frailty models for survival data. *Lifetime Data Analysis* 1995; **1**(3):255–273, doi:10.1007/BF00985760. URL <http://dx.doi.org/10.1007/BF00985760>.
- [9] Rondeau V, Filleul L, Joly P. Nested frailty models using maximum penalized likelihood estimation. *Statistics in Medicine* 2006; **25**(23):4036–4052, doi:10.1002/sim.2510. URL <https://doi.org/10.1002/sim.2510/abstract>.
- [10] Rondeau V, Michiels S, Liquet B, Pignon J. Investigating trial and treatment heterogeneity in an individual patient data meta-analysis of survival data by means of the penalized maximum likelihood approach. *Statistics in Medicine* 2008; **27**(11):1894–1910, doi:10.1002/sim.3161. URL <https://doi.org/10.1002/sim.3161>.
- [11] Rondeau V, Mathoulin-Pelissier S, Jacqmin-Gadda H, Brouste V, Soubeyran P. Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events. *Biostatistics* dec 2006; **8**(4):708–721, doi:10.1093/biostatistics/kx1043. URL <https://doi.org/10.1093/biostatistics/kx1043>.
- [12] Rondeau V, Pignon J, Michiels S. A joint model for the dependence between clustered times to tumour progression and deaths: A meta-analysis of chemotherapy in head and neck cancer. *Statistical Methods in Medical Research* dec 2015; **24**(6):711–729, doi:10.1177/0962280211425578. URL <https://doi.org/10.1177/0962280211425578>.
- [13] Mazroui Y, Mathoulin-Pelissier S, Soubeyran P, Rondeau V. General joint frailty model for recurrent event data with a dependent terminal event: Application to follicular lymphoma data. *Statistics in Medicine* feb 2012; **31**(11–12):1162–1176, doi:10.1002/sim.4479. URL <https://doi.org/10.1002/sim.4479>.
- [14] Crowther MJ, Look MP, Riley RD. Multilevel mixed effects parametric survival models using adaptive gauss-hermite quadrature with application to recurrent events and individual participant data meta-analysis. *Statistics in Medicine* may 2014; **33**(22):3844–3858, doi:10.1002/sim.6191. URL <http://dx.doi.org/10.1002/sim.6191>.
- [15] Pickles A, Crouchley R. A comparison of frailty models for multivariate survival data. *Statistics in Medicine* 1995; **14**(13):1447–1461, doi:10.1002/sim.4780141305. URL <https://doi.org/10.1002/sim.4780141305>.
- [16] International Stroke Trial Collaborative Group. The international stroke trial (ist): a randomised trial of aspirin, subcutaneous heparin, both, or neither among 19435 patients with acute ischaemic stroke. *The Lancet* 1997; **349**(9065):1569–1581, doi:10.1016/S0140-6736(97)04011-7. URL [http://dx.doi.org/10.1016/S0140-6736\(97\)04011-7](http://dx.doi.org/10.1016/S0140-6736(97)04011-7).
- [17] Sandercock P, Niewada M, Czlonkowska A. International stroke trial database (version 2). *Technical Report*, University of Edinburgh, Department of Clinical Neuroscience 2011, doi:10.7488/ds/104. URL <http://dx.doi.org/10.7488/ds/104>.
- [18] Wienke A. *Frailty models in survival analysis*. Chapman and Hall/CRC, 2010, doi:10.1201/9781420073911. URL <http://dx.doi.org/10.1201/9781420073911>.
- [19] Gutierrez RG. Parametric frailty and shared frailty survival models. *The Stata Journal* 2002; **2**(1):22 – 44.
- [20] Clayton DG. A Monte Carlo method for Bayesian inference in frailty models. *Biometrics* jun 1991; **47**(2):467, doi:10.2307/2532139. URL <https://doi.org/10.2307/2532139>.
- [21] Sinha D, Maiti T, Ibrahim JG, Ouyang B. Current methods for recurrent events data with dependent termination. *Journal of the American Statistical Association* jun 2008; **103**(482):866–878, doi:10.1198/016214508000000201. URL <https://doi.org/10.1198/016214508000000201>.
- [22] Crowther MJ, Lambert PC. Simulating biologically plausible complex survival data. *Statistics in Medicine* 2013; **32**(23):4118–4134, doi:10.1002/sim.5823. URL <https://doi.org/10.1002/sim.5823>.
- [23] Liu X, Pawitan Y, Clements M. Parametric and penalized generalized survival models. *Statistical Methods in Medical Research* sep 2016; doi:10.1177/0962280216664760. URL <https://doi.org/10.1177/0962280216664760>.
- [24] White IR. simsum: Analyses of simulation studies including monte carlo error. *The Stata Journal* 2010; **10**(3):369–385.

# Bibliography

- Bender, R., Augustin, T., and Blettner, M. (2005). Generating survival times to simulate cox proportional hazards models. *Statistics in Medicine*, 24:1713–1723.
- Brent, R. P. (1973). *Algorithms for minimization without derivatives*. Prentice-Hall.
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 34(2):187–220.
- Crowther, M. J., Abrams, K. R., and Lambert, P. C. (2012). Flexible parametric joint modelling of longitudinal and survival data. *Statistics in Medicine*, 31(30):4456–4471.
- Crowther, M. J. and Lambert, P. C. (2013). Simulating biologically plausible complex survival data. *Statistics in Medicine*, 32(23):4118–4134.
- Crowther, M. J., Look, M. P., and Riley, R. D. (2014). Multilevel mixed effects parametric survival models using adaptive gauss-hermite quadrature with application to recurrent events and individual participant data meta-analysis. *Statistics in Medicine*, 33(22):3844–3858.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 39(1):1–38.
- Denaxas, S. C., George, J., Herrett, E., Shah, A. D., Kalra, D., Hingorani, A. D., Kivimaki, M., Timmis, A. D., Smeeth, L., and Hemingway, H. (2012). Data resource profile:

- cardiovascular disease research using linked bespoke studies and electronic health records (CALIBER). *International Journal of Epidemiology*, 41:1625–1638.
- Diggle, P. J., Heagerty, P., Liang, K., and Zeger, S. L. (2013). *Analysis of longitudinal data*. Oxford Statistical Science Series. OUP Oxford, 2 edition.
- Ghosh, P., Ghosh, K., and Tiwari, R. C. (2011). Joint modeling of longitudinal data and informative dropout time in the presence of multiple changepoints. *Statistics in Medicine*, 30:611–626.
- Gould, A. L., Boye, M. E., Crowther, M. J., Ibrahim, J. G., Quartey, G., Micallef, S., and Bois, F. Y. (2015). Joint modeling of survival and longitudinal non-survival data: current methods and issues. report of the DIA Bayesian joint modeling working group. *Statistics in Medicine*, 34(14):2181–2195.
- Gutierrez, R. G. (2002). Parametric frailty and shared frailty survival models. *The Stata Journal*, 2(1):22 – 44.
- Han, M., Song, X., and Sun, L. (2014). Joint modeling of longitudinal data with informative observation times and dropouts. *Statistica Sinica*, 24:1487–1504.
- Henderson, R., Diggle, P., and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(4):465–480.
- Hickey, G. L., Philipson, P., Jorgensen, A., and Kolamunnage-Dona, R. (2016). Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC Medical Research Methodology*, 16(1).
- Hogan, J. W. and Laird, N. M. (1998). Increasing efficiency from censored survival data by using random effects to model longitudinal covariates. *Statistical Methods in Medical Research*, 7(1):28–48.
- Hougaard, P. (2000). *Analysis of multivariate survival data*. Springer New York.

- Hsieh, F., Tseng, Y., and Wang, J. (2006). Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics*, 62:1037–1043.
- Ibrahim, J. G., Chu, H., and Chen, L. M. (2010). Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology*, 28(16):2796–2801.
- Lange, J. M., Hubbard, R. A., Inoue, L. Y., and Minin, V. N. (2015). A joint model for multistate disease processes and random informative observation times, with applications to electronic medical records data. *Biometrics*, 71:90–101.
- Lawless, J. F. and Nazeri Rad, N. (2015). Estimation and assessment of markov multistate models with intermittent observations on individuals. *Lifetime Data Analysis*, 21:160–179.
- Lesperance, M. L., Sabelnykova, V., Nathoo, F. S., Lau, F., and Downing, M. G. (2015). A joint model for interval-censored functional decline trajectories under informative observation. *Statistics in Medicine*, 34:3929–3948.
- Li, Q. and Su, L. (2017). Accommodating informative dropout and death: a joint modelling approach for longitudinal and semicompeting risks data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*.
- Liu, L., Huang, X., and O’Quigley, J. (2008). Analysis of longitudinal data in presence of informative observational times and a dependent terminal event, with application to medical cost data. *Biometrics*, 64:950–958.
- Liu, X., Pawitan, Y., and Clements, M. (2016). Parametric and penalized generalized survival models. *Statistical Methods in Medical Research*.
- McLachlan, G. and McGiffin, D. (1994). On the role of finite mixture models in survival analysis. *Statistical Methods in Medical Research*, 3(3):221–226.
- Monahan, J. F. (2011). *Numerical methods of statistics*. Statistical and Probabilistic Mathematics. Cambridge University Press, 2 edition.

- Nazeri Rad, N. and Lawless, J. F. (2017). Estimation of state occupancy probabilities in multistate models with dependent intermittent observation, with application to hiv viral rebounds. *Statistics in Medicine*, 36:1256–1271.
- Pawitan, Y. and Self, S. (1993). Modeling disease marker processes in aids. *Journal of the American Statistical Association*, 88(423):719–726.
- Pickles, A. and Crouchley, R. (1995). A comparison of frailty models for multivariate survival data. *Statistics in Medicine*, 14(13):1447–1461.
- Ripatti, S. and Palmgren, J. (2000). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, 56(4):1016–1022.
- Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data: with applications in R*. Biostatistics. Chapman & Hall / CRC.
- Rizopoulos, D., Hatfield, L. A., Carlin, B. P., and Takkenberg, J. J. (2014). Combining dynamic predictions from joint models for longitudinal and time-to-event data using bayesian model averaging. *Journal of the American Statistical Association*, 109(508):1385–1397.
- Royston, P. and Parmar, M. K. (2002). Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine*, 21(15):2175–2197.
- Rutherford, M. J., Crowther, M. J., and Lambert, P. C. (2015). The use of restricted cubic splines to approximate complex hazard functions in the analysis of time-to-event data: a simulation study. *Journal of Statistical Computation and Simulation*, 85(4):777–793.
- Sweeting, M. J., Farewell, V. T., and De Angelis, D. (2010). Multi-state Markov models for disease progression in the presence of informative examination times: an application to hepatitis C. *Statistics in Medicine*, 29:1161–1174.

- 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.
- Therneau, T. M., Grambsch, P. M., and Pankratz, V. S. (2003). Penalized survival models and frailty. *Journal of Computational and Graphical Statistics*, 12(1):156–175.
- 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.
- Tsiatis, A. A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, 14:809–834.
- Tsiatis, A. A., Degruyter, 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.
- White, I. R. (2010). simsum: Analyses of simulation studies including monte carlo error. *The Stata Journal*, 10(3):369–385.
- Wienke, A. (2010). *Frailty models in survival analysis*. Chapman & Hall / CRC.
- Wu, M. C. and Carroll, R. J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, 44:175–188.
- Wulfsohn, M. S. and Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, 53(1):330–339.
- Xie, Y. (2016). *bookdown: Authoring books and technical documents with R Markdown*. The R Series. Chapman & Hall / CRC.