

Probation review report

Alessandro Gasparini

2017-08-20

Contents

Introduction	7
1 Introduction to survival analysis	9
1.1 Survival data	10
1.2 Censoring	10
1.3 Terminology and notation	12
1.4 Estimation of the survival function	15
1.5 Parametric survival models	17
1.6 The Cox proportional hazards model	29
1.7 Advances in survival analysis	35
2 Survival models with random effects	37
2.1 Univariate frailty models	39
2.2 Shared frailty models	41
2.3 Alternative formulation	42
3 Joint models for longitudinal and survival data	45
3.1 Model formulation	46
3.2 Estimation process	48
4 Computational challenges in survival models with random effects	51

4.1	Numerical integration	52
4.2	Other considerations	57
5	Simulating survival data	63
6	Simulation study: accuracy of Gaussian quadrature	67
6.1	Aim	67
6.2	Data-generating mechanisms	68
6.3	Methods	69
6.4	Estimands	70
6.5	Performance measures	71
6.6	Results	71
6.7	Conclusions	77
7	Simulation study: impact of misspecification in survival models with shared frailty terms	79
7.1	Aim	80
7.2	Data-generating mechanisms	81
7.3	Methods	83
7.4	Estimands	85
7.5	Performance measures	86
7.6	Results	87
7.7	Conclusions	90
8	Exploring results from simulation studies interactively	91
8.1	Rationale	92
8.2	SiReX	93
9	Informative visiting process	95

<i>CONTENTS</i>	5
10 Future research developments	99
11 Personal development	103
11.1 Supervisory meetings	103
11.2 Training and courses	103
11.3 Conferences	105
Appendix	106
A Tables	107
A.1 Simulation study 1	107
A.2 Simulation study 2	128
B Plots	137
B.1 Simulation study 1	137
B.2 Simulation study 2	162
C Slides	173
C.1 2017 SAfJR Conference	173
C.2 2017 SAM Conference and ISCB Conference	181
C.3 Students' Day at the ISCB Conference	188
D Manuscript	197
E R Session	199
Bibliography	207

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 the topic of survival analysis 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 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 the research goals for the second year of my PhD. Finally, I will briefly summarise the training and personal development activities I have participated to during

the first year of my PhD 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 bookdown (Xie, 2016), and can be accessed online at <https://ellessenne.github.io/prr/>.

Chapter 1

Introduction to survival analysis

Survival analysis is a branch of statistics in which the main outcome consists in the time until the occurrence of a given event. Time could be years, months, weeks, or any amount of calendar time or even age time; event could be death, disease occurrence or relapse, or any other experience of interest. Survival analysis is also known as reliability theory in engineering, duration analysis in economics, and event history analysis in sociology. A broad overview of survival analysis is given in [Kalbfleisch and Prentice \(2011\)](#) and in [Kleinbaum and Klein \(2012\)](#).

Some examples of time to event data are:

- disease remission in leukemia patients. In this study, leukemia patients are followed over several weeks to study how long they stay in remission status;
- heart disease occurrence. In this study, healthy subjects are followed over several years until occurrence of heart disease, or end of the study;
- renal failure. In this study, individuals with kidney disease are followed until renal failure, or end of the study;
- reliability of complex technical installations. For instance, studies assessing failure

rates of components such as bulbs and valves.

In this Chapter I will define survival data and its peculiarities in Section 1.1 and 1.2. Terminology and notation used throughout this report will be introduced in Section 1.3. I will introduce common non-parametric and parametric methods in survival analysis in Sections 1.4 and 1.5. I will introduce the widely used semi-parametric Cox model in Section 1.6. Finally, I will provide a brief overview on advances in survival analysis in Section 1.7.

1.1 Survival data

Survival data generally consists - as previously mentioned - in an event of interest and time until its occurrence. In the leukemia remission example, time to event would be how many weeks it takes before a given patient experiences disease relapse and the event would be whether the individual relapsed or not before the end of the study. Nevertheless, in certain situations we may have some information about the survival time but the actual survival time may be unknown. This problem is known as censoring and it is presented in Section 1.2.

1.2 Censoring

Censoring is a mechanism that causes survival times to be unobserved. There are many reasons why censoring may occur; among others:

1. a person does not experience the event before the end of the study;
2. a person drops out of the study before the occurrence of the event of interest;
3. a person experiences a competing event that impedes the occurrence of the event of interest (e.g.: death, when death is not the study outcome).

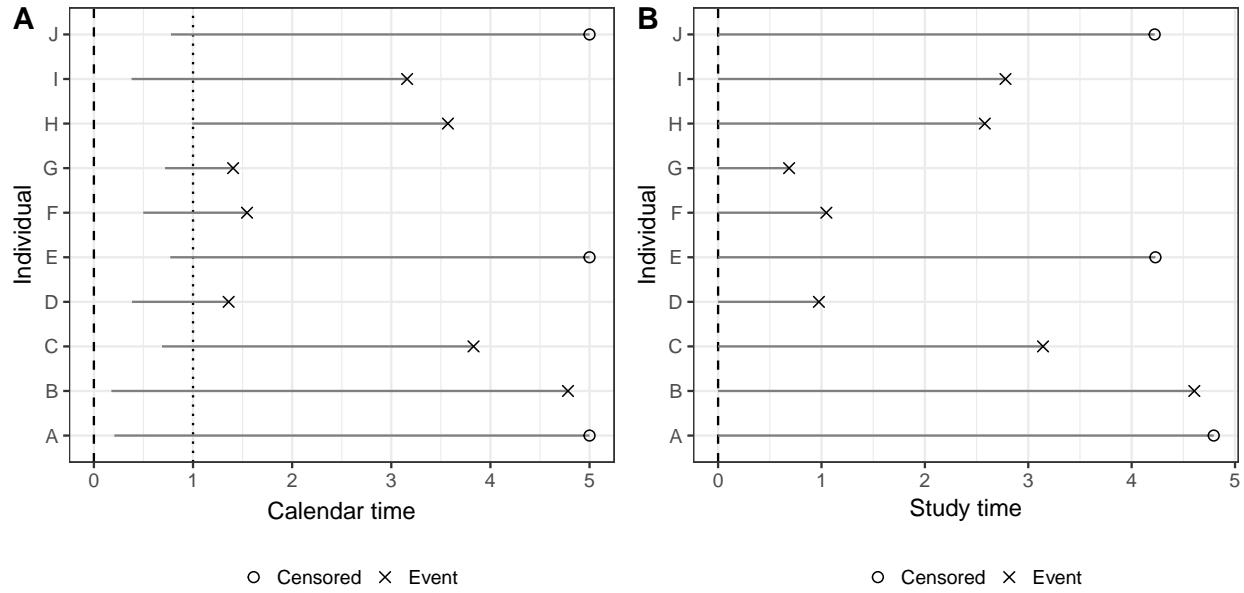


Figure 1.1: Simulated right censored survival data, plotted by their calendar time in panel A and by their study time in panel B.

I simulated survival data for illustration purposes: I assumed a clinical trial with 10 individuals enrolled during a recruitment window of 1 year, and followed for up to 5 years. Not all individuals experience the event of interest during the study period, and are therefore censored after five years from the start of the study. The observation time for each individual is depicted in Figure 1.1 with a solid dark grey line, a cross represents the occurrence of the study event, and a circle represents censoring. Individuals A, E, and J all have censored survival time: I know that they were still event-free at the end of follow-up, i.e. their real survival time is greater than the observed one, but the former is unknown. The simulated data is presented in Figure 1.1: in panel A, survival data is plotted against the calendar time; conversely, in panel B, survival data is plotted against the study time, e.g. each individual is assigned a *time zero* corresponding to their enrollment in the study, and survival time is counted from there.

This example represents a particular form of censoring: *right censoring*. The defining characteristic of right-censored data is that it is censored (or incomplete) at the right side of the

follow-up time, hence the true survival time is greater than the observed time. This example represents *administrative censoring* as well, as individuals are censored at the end of the study to artificially restrict follow-up time (e.g. for financial reasons).

It is also possible to encounter data that is *left censored* or *interval censored*. In the former case, the true survival time is shorter than the observed one, e.g. I know that the event occurred before the observation time, but I do not know when - imagine onset of a viral infection, which can be detected only at a visit time. In the latter, I know that the event occurred within a certain interval of time but I do not know when; using the same example of infection onset, if infection was detected at a visit date but the individual was known to be infection-free at the previous visit, the true infection onset time is unknown and the event time is said to be interval censored.

Finally, another important concept related to right censoring is that of *left truncation* (or *delayed entry*). Left truncation occurs when an individual enrolls in the study some time after the inclusion criteria are satisfied; individuals that die (or emigrate, ...) before the start of observation time will never enter the study, and inclusion time may differ between individuals. Data arising from such phenomenon is therefore said to be left truncated.

1.3 Terminology and notation

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 defined two 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 survives¹ longer than a 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. Figure 1.2 depicts this difference: in panel A I plotted a theoretical survival function, restricted to 15 years of follow-up for comparison purposes, while in panel B I plotted the survival function relative to the survival data simulated in Section 1.2. The former is a smooth function of time, and should we extend the x-axis to infinity the function would eventually reach zero. Conversely, the latter is a step function with steps at each event time, and should we extend the x-axis to infinity the function would remain flat after the last observed event.

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

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

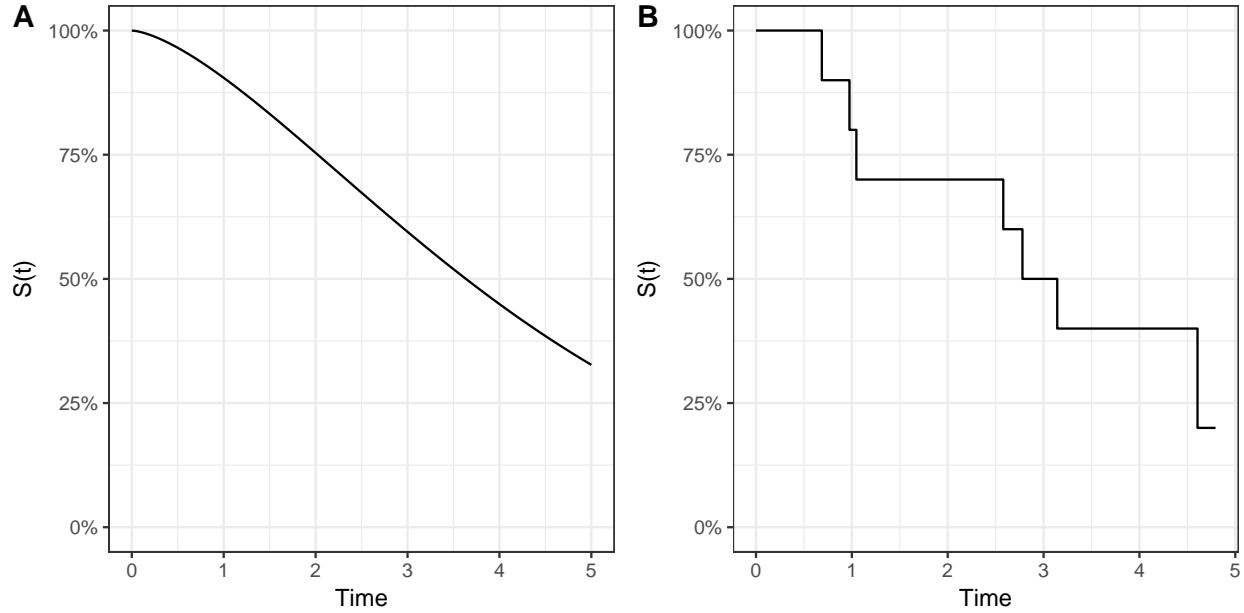


Figure 1.2: Theoretical survival function (A) and observed survival function for simulated data (B).

$[t, t + \Delta(t))$ (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. In Figure 1.3 I present a simple hazard function; it increases over time, which means that the instantaneous risk of event increases over time.

The survival function from Figure 1.2, panel A, and the hazard function from Figure 1.3 are strictly related. In fact, there is a clearly defined mathematical relationship between the survival and the hazard function: 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

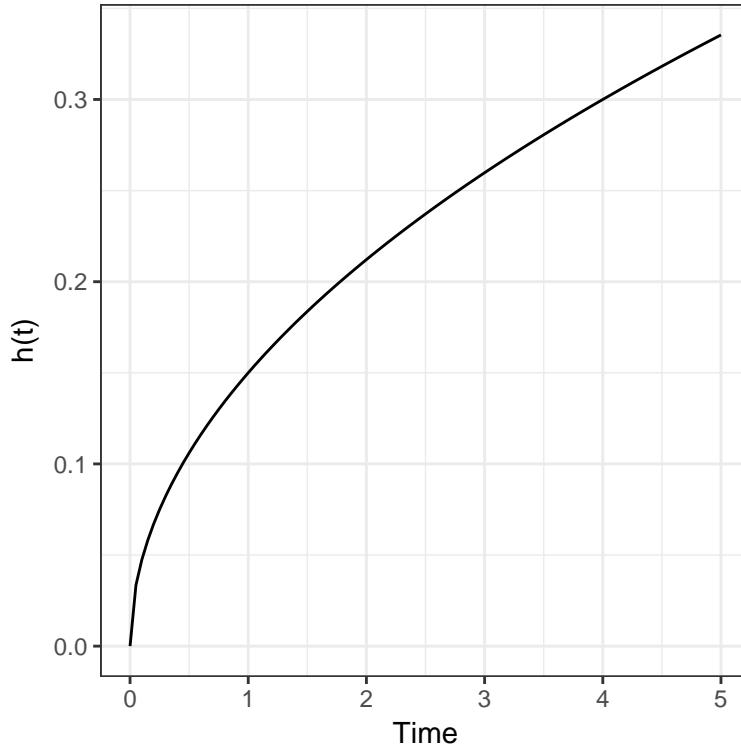


Figure 1.3: Example of hazard function.

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 $S(t) = \exp(-H(t))$.

1.4 Estimation of the survival function

The survival function presented in Figure 1.2, panel B, is a non-parametric estimate of the true survival function based on the data only. The estimator employed in this case is the Kaplan-Meier estimator of the survival function (Kaplan and Meier, 1958), with which the estimated survival probabilities are obtained using a product limit formula. The general

form for the Kaplan-Meier estimator at time $t_{(i)}$ is

$$\hat{S}(t_{(i)}) = \hat{S}(t_{(i-1)}) \times \hat{P}(T > t_{(i)} | T \geq t_{(i)}),$$

with $t_{(i)}$ being the i^{th} ordered failure time. The interpretation is straightforward: it is the product of the probability of surviving past the previous event-time ($\hat{S}(t_{(i-1)})$) times the conditional probability of surviving past the current time $t_{(i)}$ given survival to at least the current time ($\hat{P}(T > t_{(i)} | T \geq t_{(i)})$). The product limit formula is:

$$\hat{S}(t_{(i)}) = \prod_{j=1}^i \hat{P}(T > t_{(j)} | T \geq t_{(j)})$$

The conditional probability in the product limit formula can be estimated from the observed data as:

$$\hat{P}(T > t_{(i)} | T \geq t_{(i)}) = \frac{r_{(i)} - e_{(i)}}{r_{(i)}},$$

where $r_{(i)}$ and $e_{(i)}$ are the number of individuals at risk and the number of events at time $t_{(i)}$, respectively.

The Kaplan-Meier estimator can be computed using R and the function `survfit` from the `survival` package. An example using the simulated data from Section 1.2 (stored in a data frame named `data`):

```
# library(survival)

fit = survfit(Surv(time = t0, event = d) ~ 1, data = data)
summary(fit)

## Call: survfit(formula = Surv(time = t0, event = d) ~ 1, data = data)
##
##      time n.risk n.event survival std.err lower 95% CI upper 95% CI
##      0.687     10       1      0.9  0.0949      0.7320     1.000
```

##	0.975	9	1	0.8	0.1265	0.5868	1.000
##	1.047	8	1	0.7	0.1449	0.4665	1.000
##	2.580	7	1	0.6	0.1549	0.3617	0.995
##	2.780	6	1	0.5	0.1581	0.2690	0.929
##	3.143	5	1	0.4	0.1549	0.1872	0.855
##	4.606	2	1	0.2	0.1612	0.0412	0.971

By doing so, I obtain an estimate of the survival function (column `survival`) at each distinct failure time (column `time`). For instance, the survival probability at $t = 1.047$ is 0.700, with 95% confidence interval (0.467 - 1.000).

Finally, plotting the estimated survival curve I obtain Figure 1.4, which is exactly the same survival curve presented in panel B of Figure 1.2.

```
# library(ggfortify)

autoplot(fit, conf.int = FALSE, censor = FALSE) +
  theme_bw() +
  coord_cartesian(ylim = c(0, 1)) +
  labs(x = "Time", y = expression(hat(S)[KM](t)))
```

An alternative way of estimating the survival function is to use the *Nelson-Aalen* estimator for the cumulative hazard

$$\hat{H}(t) = \sum_{t_i < t} \frac{e_i}{r_i} = \sum_{t_i < t} \hat{h}_i,$$

and then use the relationship presented in Section 1.3 to obtain the survival function.

1.5 Parametric survival models

In applied settings it is often of interest to assess the association between observed covariates and the survival time of interest. For instance, it may be of interest to study whether a

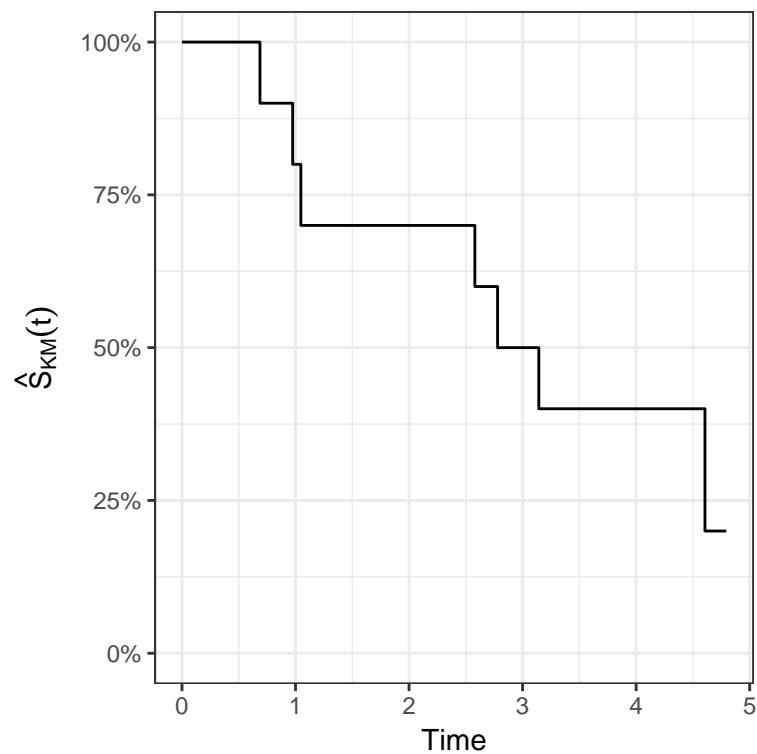


Figure 1.4: Estimated survival function using the Kaplan-Meier estimator on the simulated data.

treatment is effective in slowing disease relapse (e.g. relapse of leukemia), whether there are difference between genders or age categories. A common way of assessing the effect of covariates on a time to event outcome, while adjusting for potentially confounding factors at the same time, consists in using a regression model.

In the context of survival data, two models are commonly used: the *accelerated failure time* model (AFT), and the *proportional hazards* (PH) model. In the former, the natural logarithm of the observed survival time $\log t$ is expressed as a linear function of the covariates X :

$$\log t = X\beta + \epsilon,$$

with β a vector of regression coefficients and ϵ a vector of residual error terms. Assuming a parametric distribution for ϵ determines the regression model: log-normal, log-logistic, Weibull, etc. In the AFT model, a positive association of the covariates with survival time implies an increased expected time to event. In the PH model, the covariates have a multiplicative effect on the hazard function:

$$h(t; X) = h_0(t)g(X),$$

for some $h_0(t)$ and $g(X)$, with $g(\cdot)$ a non-negative function of the covariates. A popular choice for the latter is $g(X) = \exp(X\beta)$; conversely, it is possible to either leave the former unspecified, or assume a parametric distribution. The focus of this Section is on specifying a parametric distribution for $h_0(t)$, yielding the so-called *parametric survival regression models*; I will present commonly assumed parametric distributions in Section 1.5.1, the estimation procedure in Section 1.5.2, and an example using data from the International Stroke Trial (IST) (International Stroke Trial Collaborative Group, 1997; Sandercock et al., 2011) in Section 1.5.3. Leaving $h_0(t)$ unspecified yields the semi-parametric Cox model, that I will present in Section 1.6.

From now on I will focus on the proportional hazards formulation of the survival model.

1.5.1 Failure time distributions

I mentioned in Section 1 that the random variable representing the survival time is non-negative; hence, we can choose any non-negative distribution to assign to $h_0(t)$. Commonly used distribution are the Exponential, Weibull, log-Normal, and Gompertz distributions; other possible distributions are the inverse Weibull, the inverse Gamma, the positive stable, the log-skew-Normal, the log-logistic, and complex mixture distributions (such as the two components mixture Weibull distribution, [McLachlan and McGiffin \(1994\)](#)). Each distribution yields a different shape for the Survival and hazard functions; in particular, focusing on the distribution I will be utilising in the rest of this report:

- Exponential distribution:

- $h_0(t) = \lambda$
- $S(t) = \exp(-\lambda t)$
- $\lambda > 0$

- Weibull distribution:

- $h_0(t) = \lambda p t^{p-1}$
- $S(t) = \exp(-\lambda t^p)$
- $\lambda, p > 0$

- log-Normal distribution:

- $h_0(t) = \frac{\phi(\frac{\log t - \mu}{\sigma})}{\sigma t (1 - \Phi(\frac{\log t - \mu}{\sigma}))}$
- $S(t) = 1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right)$
- $\mu \in R; \sigma > 0$

- Gompertz distribution:

- $h_0(t) = \lambda \exp(\gamma t)$
- $S(t) = \exp\left[-\frac{\lambda}{\gamma}(\exp(\gamma t) - 1)\right]$

- $\lambda, \gamma > 0$
- Two components mixture Weibull distribution:
 - $h_0(t) = \frac{\lambda_1 p_1 t^{p_1-1} \pi \exp(-\lambda_1 t^{p_1}) + \lambda_2 p_2 t^{p_2-1} (1-\pi) \exp(-\lambda_2 t^{p_2})}{\pi \exp(-\lambda_1 t^{p_1}) + (1-\pi) \exp(-\lambda_2 t^{p_2})}$
 - $S(t) = \pi \exp(-\lambda_1 t^{p_1}) + (1 - \pi) \exp(-\lambda_2 t^{p_2})$
 - $\lambda_1, \lambda_2, p_1, p_2 > 0; \pi \in [0, 1]$

1.5.2 Estimation procedure

Assume n observations with the bivariate response (t_i, d_i) , with $i = 1, \dots, n$. For a given survival function $S(t)$ the density function is given by

$$f(t) = -\frac{dS(t)}{dt},$$

and the hazard function by

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

The parameters of the parametric proportional hazards survival model presented in Section 1.5 can be estimated via the maximum likelihood method. A subject that experiences the event of interest at time t_i contributes to the likelihood the density at time t_i , i.e. $f(t_i)$; conversely, a censored observation known to survive until time t_i contributes $S(t_i)$ to the likelihood. The individual contribution to the likelihood L_i can therefore be written as

$$L_i = h(t_i)^{d_i} S(t_i),$$

where d_i is the event indicator variable. The overall likelihood is the product of the individual contributions:

$$L = \prod_{i=1}^n L_i.$$

Taking the natural logarithm of the likelihood for ease of computation:

$$\begin{aligned}\log L &= \sum_{i=1}^n [d_i \log f_i(t_i) + (1 - d_i) \log S_i(t_i)] = \\ &= \sum_{i=1}^n [d_i \log h_i(t_i) + \log S_i(t_i)]\end{aligned}$$

Implicit in the above log-likelihood are the regression parameters β and the parameters of the parametric distribution of choice for $h_0(t)$.

The log-likelihood function $\log L$ has a closed-form; maximum likelihood estimates for β and the distribution parameters can hence be obtained by maximising $\log L$, e.g. using one of the many general purpose optimisers available in R (`optim`, `nlm`, ...).

1.5.3 Data analysis example

The International Stroke Trial (IST) was a large, prospective, randomised controlled trial conducted between 1991 and 1996. The aim of the trial was to assess whether early administration of aspirin, heparin, both or neither influenced clinical outcomes in patients with acute ischaemic stroke (International Stroke Trial Collaborative Group, 1997; Sandercock et al., 2011).

As illustration, I will evaluate the association between treatment with aspirin and/or heparin and survival after acute ischaemic stroke. I will start by reading the data, stored in the `ist.csv` file. This file is a subset of the full IST dataset containing information on age, gender, Country, treatment, and survival; further, individuals with missing values and individuals with a survival time of zero were dropped.

```
# library(readr)

ist = read_csv("data/ist.csv.gz",
  col_names = c("gender", "age", "rxasp", "rxhep", "country", "d", "t"),
```

```

col_types = "cicccii", skip = 1)

attr(ist, "spec") = NULL # removing "spec" attribute

# turn treatments into factors

ist$rxasp = factor(ist$rxasp, levels = c("N", "Y"))
ist$rxhep = factor(ist$rxhep, levels = c("N", "L", "H"))

```

I fit first a parametric survival model assuming a Weibull distribution for $h_0(t)$. The hazard function, including covariates and the imposing proportional hazards, has the form

$$h(t; X) = \lambda p t^{p-1} \exp(X\beta),$$

while the survival function has the form

$$S(t; X) = \exp(-\lambda t^p \exp(X\beta)).$$

X is the model design matrix, and β is the vector of regression coefficients. The log-likelihood has the form

$$\log L = \sum_{i=1}^n [d_i \log h_i(t_i) + \log S_i(t_i)]$$

First, I code a function with the model log-likelihood. The function depends on (1) the model parameters β , λ , and p (`pars` argument), (2) the model design matrix X (`X` argument), and (3) survival time t and event indicator d (`t` and `d` arguments):

```

ll = function(pars, X, t, d) {
  lambda = exp(pars[1])
  p = exp(pars[2])
  beta = pars[-(1:2)]
  log_hi = log(lambda) + log(p) + (p - 1) * log(t) + c(X %*% beta)
  log_lo = log(1 - lambda) + log(1 - p) + (p - 1) * log(1 - t) + c(X %*% beta)
  log_ll = d * log_hi + (1 - d) * log_lo
  sum(log_ll)
}

```

```

log_Si = -lambda * t ^ p * exp(c(X %*% beta))

ll = sum(d * log_hi + log_Si + log(t))

# + sum(log(t)) is the same adjustment that Stata

# does to remove the time units from log L

return(-ll)

}

```

The function `ll()` returns the negative log-likelihood as most optimisers minimise a target function (and so does `optim`); however, minimising the negative log-likelihood function is equivalent to maximising the log-likelihood.

I define the model matrix X for a model with aspirin treatment, heparin treatment, and their interactions. The first column is removed to avoid collinearity:

```
X = with(ist, model.matrix(t ~ rxasp * rxhep - 1))[, -1]
```

Next, I define the starting values for the optimisation routine. I choose the value 1 for the parameters of the Weibull distribution and the value 0 for the regression coefficients:

```

start = c(1, 1, rep(0, ncol(X)))

names(start) = c("lambda", "p", colnames(X))

```

The value of the log-likelihood function at the starting values is -92861140101.392. Finally, I use the robust-variance modification of the Marquardt algorithm, which is more efficient than Gauss-Newton-like algorithms when starting from points very far from the optimum (Marquardt, 1963; Commenges et al., 2006):

```

# library(marqLevAlg)

fit = marqLevAlg(b = start,

fn = function(x) ll(x, X = X, t = ist$t, d = ist$d))

## 

## Be patient. The program is computing ...

```

```
## The program took 7.27 seconds
```

Assess convergency:

```
fit$istop
```

```
## [1] 1
```

The convergence status indicator is equal to 1, hence the convergence criteria were satisfied. The log-likelihood at the maximum likelihood estimates is 61566.567. The optimising routine returns the upper triangle matrix of variance-covariance estimates at the stopping point, which can be used to obtain standard errors of the estimated coefficients:

```
fit$vcov = matrix(0,
  nrow = length(fit$b),
  ncol = length(fit$b))

fit$vcov[upper.tri(fit$vcov, diag = TRUE)] = fit$v
fit$vcov[lower.tri(fit$vcov)] = t(fit$vcov)[lower.tri(fit$vcov)]
```

Finally, I build a table of results:

```
res = data.frame(
  coef = fit$b,
  hr = exp(fit$b),
  se = sqrt(diag(fit$vcov)))
res$z = res$coef / res$se
res$p = 2 * pmin(pnorm(-abs(res$z)), 1 - pnorm(abs(res$z)))
kable(res,
  digits = 3,
  align = "rrrrr",
  booktabs = TRUE,
  col.names = c("Beta", "Hazard ratio", "SE (Beta)", "Z", "P > |Z|"),
```

Table 1.1: Results from a parametric Weibull model.

	Beta	Hazard ratio	SE (Beta)	Z	P > Z
lambda	-3.852	0.021	0.047	-82.665	0.000
p	-0.759	0.468	0.015	-52.050	0.000
rxaspY	-0.037	0.964	0.044	-0.850	0.395
rxhepL	0.085	1.088	0.052	1.640	0.101
rxhepH	0.044	1.045	0.052	0.843	0.399
rxaspY:rxhepL	-0.095	0.910	0.075	-1.269	0.204
rxaspY:rxhepH	0.037	1.038	0.074	0.503	0.615

```
linesep = """",
caption = "Results from a parametric Weibull model.")
```

I test the interaction term using the Wald test to assess whether combining aspirin and heparin alter their association with time to event. I use the Wald χ^2 test statistic as the sample size is big enough for it to be equivalent to its F counterpart:

```
# identify the interaction terms
idx = grep1(":", names(fit$b))

# compute the W statistic
W = t(fit$b[idx]) %*% solve(fit$vcov[idx, idx]) %*% fit$b[idx]

# produce the test
c(W = W, df = sum(idx), `p-value` = 1 - pchisq(W, sum(idx)))

##          W        df    p-value
## 2.6104650 2.0000000 0.2711095
```

The interaction terms seem to be not statistically significant. We can conclude the association of aspirin and heparin treatments with survival are not dependent on one another.

I can then re-fit the model excluding the interaction terms:

```

X = with(ist, model.matrix(t ~ rxasp + rxhep - 1))[, -1]

start = c(1, 1, rep(0, ncol(X)))

names(start) = c("lambda", "p", colnames(X))

re_fit = marqLevAlg(b = start,
fn = function(x) ll(x, X = X, t = ist$t, d = ist$d))

## 

## Be patient. The program is computing ...
## The program took 4.25 seconds

re_fit$istop

## [1] 1

```

The routine converged. I produce the variance-covariance matrix:

```

re_fit$vcov = matrix(0,
nrow = length(re_fit$b),
ncol = length(re_fit$b))

re_fit$vcov[upper.tri(re_fit$vcov, diag = TRUE)] = re_fit$v
re_fit$vcov[lower.tri(re_fit$vcov)] = t(re_fit$vcov)[lower.tri(re_fit$vcov)]

```

Finally, I build a new table of results:

```

re_res = data.frame(
coef = re_fit$b,
hr = exp(re_fit$b),
se = sqrt(diag(re_fit$vcov)))

re_res$z = re_res$coef / re_res$se
re_res$p = 2 * pmin(pnorm(-abs(re_res$z)), 1 - pnorm(abs(re_res$z)))

kable(re_res,
digits = 3,

```

Table 1.2: Results from a parametric Weibull model with no interactions.

	Beta	Hazard ratio	SE (Beta)	Z	P > Z
lambda	-3.845	0.021	0.044	-87.441	0.000
p	-0.759	0.468	0.015	-52.056	0.000
rxaspY	-0.051	0.950	0.030	-1.689	0.091
rxhepL	0.039	1.040	0.037	1.043	0.297
rxhepH	0.062	1.064	0.037	1.684	0.092

```


```

I now test the significance of the two coefficients related to heparin treatment jointly:

```
idx = grep("rxhep", names(re_fit$b))
W = t(re_fit$b[idx]) %*% solve(re_fit$vcov[idx, idx]) %*% re_fit$b[idx]
c(W = W, df = sum(idx), `p-value` = 1 - pchisq(W, sum(idx)))
##          W      df p-value
## 3.08001 2.00000 0.21438
```

Heparin treatment seems to be not statistically significantly associated with time to death in acute ischaemic stroke patients; the effect size is small, with a 4% and 7% increased risk for the L and H heparin treatment modalities versus no heparin treatment, respectively.

Finally, the treatment with aspirin is also not statistically significantly associated with the outcome; effect size is small as well, approximately a 5% risk reduction for aspirin treatment compared to no treatment with aspirin (Table 1.2).

This is a simple application of parametric survival models; a fully developed analysis should take further aspects into account, such as considering different hazard distributions. It is

possible to estimate various models and compare their fit to a specific distribution using information criteria such as the Akaike information criterion (AIC) and the Bayesian information criterion (BIC).

1.6 The Cox proportional hazards model

The parametric survival models of Section 1.5 could have both the accelerated failure time form and the proportional hazards form. Recall that the latter is formulated in terms of the hazard function:

$$h(t; X) = h_0(t) \exp(X\beta)$$

As I mentioned before, this model requires specifying the baseline hazard function $h_0(t)$ (e.g. using one of the parametric distributions of Section 1.5.1) and by leaving it unspecified I obtain the Cox proportional hazards model. Such model is also called *semi-parametric* as it is formed by a non-parametric component (the baseline hazard left unspecified) and a parametric component (the modelling assumption for the functional form of $g(\cdot)$, the usual $\exp(X\beta)$ in this case). The survival function for a Cox model can be written as:

$$S(t; X) = \exp \left[- \int_0^t h_0(u) \exp(X\beta) du \right]$$

The main problems when fitting a Cox model are related to estimation of the regression coefficients β and of the survival function $S(t)$. The former is tackled in Section 1.6.1, the latter in Section 1.6.2.

1.6.1 Estimation procedure

The main method for estimating the regression coefficients is the method of partial likelihood, proposed and discussed in detail in Cox (1972) and Cox (1975). In brief, the observed data are

assumed to have density function $f(t; \theta, \beta)$ in which β is the vector of regression coefficients of interest and θ can be considered a vector of nuisance parameters. In particular, θ represents the unspecified function $h_0(t)$. It can be shown that it is possible to factorise the density into two terms, one of which only depends on β : this term is called *partial likelihood*. Ignoring the term that depends on θ , and even if the partial likelihood is not directly interpretable as a likelihood in the ordinary sense, it can be used like an ordinary likelihood for estimation purposes as the usual asymptotic properties formulas and properties associated with the likelihood function and likelihood estimation apply. The partial likelihood applies directly to the relative risk model $h(t; X)$, assuming independent right censoring. The individual contribution to the likelihood has the form

$$L_i(\beta) = \frac{h(t_i; x_i) \Delta t_i}{\sum_{l \in R(t_i)} h(t_l; x_l) \Delta t_l},$$

and provides information on failures occurrence in the interval $[t_i, t_i + \Delta t_i]$; $R(t_i)$ is the risk set of individuals at risk of failing at time t_i^- , right before t_i . Under the relative risk model, the baseline hazard in $h(t; X)$ cancels out in the numerator and denominator; the product over i gives the partial likelihood for β :

$$L(\beta) = \prod_{i=1}^n \frac{\exp(x_i \beta)}{\sum_{l \in R(t_i)} \exp(x_l \beta)}.$$

The values of β that maximise the partial likelihood $\hat{\beta}$ can be obtained by using a Newton-Raphson-like algorithm; asymptotics are fully analogous to a parametric likelihood. A caveat of the partial likelihood method is that it assumes continuous failure times: in practice, that is unrealistic and there will be tied failure times (e.g. due to rounding). In that case, several methods have been proposed to adjust the partial likelihood in order to handle ties; see for instance [Peto \(1972\)](#), [Breslow \(1974\)](#), and [Efron \(1977\)](#)

1.6.2 Estimating the survival function

Consider deriving an estimator for the survival function from a Cox model: the form of $h_0(t)$ is unspecified, hence it is not possible to directly estimate the parameters of the distribution as in fully parametric survival models. Under a Cox model, the survival function has the form

$$S(t; X) = S_0(t)^{\exp(X\beta)}$$

The coefficients β are estimated using the penalised likelihood procedure, and the baseline survival function $S_0(t)$ is estimated by assuming that the baseline hazard function is constant between each pair of consecutive observed failure times. The resulting estimator, known as the Breslow estimator, estimates the cumulative baseline hazard function as

$$\hat{H}_0(t) = \sum_{t(i) \leq t} \frac{e(i)}{\sum_{l \in R(t(i))} \exp(x_l \hat{\beta})},$$

with $e(i)$ the number of events at time $t(i)$. The baseline survival function follows as

$$\hat{S}_0(t) = \exp[-\hat{H}_0(t)],$$

and the survival function as

$$\hat{S}(t; X) = \hat{S}_0(t)^{\exp(X\hat{\beta})}.$$

An alternative estimator based on approximating the baseline survival function as a step function and consequently solving k simultaneous equations has been proposed by [Kalbfleisch and Prentice \(2011\)](#), and is omitted here.

1.6.3 Model assumptions

The Cox model relies on two main assumptions. First, the assumption of non-informative censoring, e.g. the censoring process must be independent of any covariate, observed and not. Second, the proportional hazards assumption requires hazards to be proportional across time, e.g. the hazard ratios must be constant. There are several ways of testing the proportional hazards assumption, both analytical and graphical; see Chapter 4 of Kleinbaum and Klein (2012) for further details.

1.6.4 Data analysis example

In this section I re-analyse the IST data of Section 1.5.3 using a semi-parametric Cox model. I first read the dataset:

```
# library(readr)

ist = read_csv("data/ist.csv.gz",
  col_names = c("gender", "age", "rxasp", "rxhep", "country", "d", "t"),
  col_types = "cicccii", skip = 1)

attr(ist, "spec") = NULL # removing "spec" attribute

# turn treatments into factors

ist$rxasp = factor(ist$rxasp, levels = c("N", "Y"))
ist$rxhep = factor(ist$rxhep, levels = c("N", "L", "H"))
```

I fit the Cox model using the `coxph()` function from the `survival` package:

```
# library(survival)

fit = coxph(Surv(t, d) ~ rxasp * rxhep, data = ist)

summary(fit)
```

```

## Call:

## coxph(formula = Surv(t, d) ~ rxasp * rxhep, data = ist)

## n= 19378, number of events= 4315

##          coef exp(coef) se(coef)      z Pr(>|z|)

## rxaspY     -0.03710  0.96358  0.04356 -0.852   0.394
## rxhepL      0.08017  1.08347  0.05161  1.553   0.120
## rxhepH      0.04215  1.04305  0.05221  0.807   0.419
## rxaspY:rxhepL -0.09049  0.91348  0.07458 -1.213   0.225
## rxaspY:rxhepH  0.03595  1.03661  0.07415  0.485   0.628

##          exp(coef) exp(-coef) lower .95 upper .95

## rxaspY       0.9636    1.0378   0.8847   1.049
## rxhepL       1.0835    0.9230   0.9792   1.199
## rxhepH       1.0431    0.9587   0.9416   1.155
## rxaspY:rxhepL 0.9135    1.0947   0.7893   1.057
## rxaspY:rxhepH 1.0366    0.9647   0.8964   1.199

## Concordance= 0.513  (se = 0.004 )

## Rsquare= 0  (max possible= 0.987 )

## Likelihood ratio test= 7.98  on 5 df,  p=0.1574
## Wald test           = 8.02  on 5 df,  p=0.1554
## Score (logrank) test = 8.02  on 5 df,  p=0.1551

```

I test again the joint signficancy of the interaction terms using the Wald test:

```
idx = grep(": ", names(coef(fit)))
```

```

W = t(coef(fit)[idx]) %*% solve(vcov(fit)[idx, idx]) %*% coef(fit)[idx]
c(W = W, df = sum(idx), `p-value` = 1 - pchisq(W, sum(idx)))

##          W      df   p-value
## 2.3929136 2.0000000 0.3022633

```

Analogously as before, the interaction is not significantly different than zero. I re-fit the model without the interaction term:

```

# library(survival)

re_fit = coxph(Surv(t, d) ~ rxasp + rxhep, data = ist)
summary(re_fit)

## Call:
## coxph(formula = Surv(t, d) ~ rxasp + rxhep, data = ist)
##
##    n= 19378, number of events= 4315
##
##          coef  exp(coef)  se(coef)      z Pr(>|z|)
## rxaspY -0.05079  0.95048  0.03046 -1.668  0.0954 .
## rxhepL  0.03641  1.03708  0.03725  0.978  0.3283
## rxhepH  0.05991  1.06174  0.03707  1.616  0.1061
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef)  exp(-coef) lower .95 upper .95
## rxaspY     0.9505      1.0521    0.8954     1.009
## rxhepL     1.0371      0.9642    0.9641     1.116
## rxhepH     1.0617      0.9419    0.9873     1.142
##

```

```

## Concordance= 0.511  (se = 0.004 )

## Rsquare= 0  (max possible= 0.987 )

## Likelihood ratio test= 5.58  on 3 df,  p=0.1337

## Wald test          = 5.59  on 3 df,  p=0.1333

## Score (logrank) test = 5.59  on 3 df,  p=0.1332

```

Testing the significance of the heparin treatment using the Wald test:

```

idx = grep("rxhep", names(coef(fit)))

W = t(coef(fit)[idx]) %*% solve(vcov(fit)[idx, idx]) %*% coef(fit)[idx]

c(W = W, df = sum(idx), `p-value` = 1 - pchisq(W, sum(idx)))

##           W      df   p-value
## 2.4963250 2.0000000 0.2870317

```

Treatment with heparin is not statistically significantly associated with risk of death; besides that, the effect size of heparin treatment is small: approximately 4% and 6% risk increase for heparin treatment modalities L and H compared to no heparin treatment, respectively.

The treatment with aspirin is also barely significantly different than zero, assuming a significance level $\alpha = 0.10$, with a p-value of 0.0954. The effect size is comparable to the estimated effect size obtained with the Weibull model, approximately 5% risk reduction for treatment with aspirin compared to no aspirin treatment.

1.7 Advances in survival analysis

There are several extensions of the statistical methods presented in this Chapter: I will briefly introduce some of them in this Section, without going into great detail as that would be beyond the scope of this report.

The proportional hazards models have been extended to include time-dependent covariates

and time-dependent covariate effects; additionally, the Cox model has been extended to allow stratification by a given factor (details in [Kalbfleisch and Prentice \(2011\)](#) and [Kleinbaum and Klein \(2012\)](#)). Parametric survival models have been generalised to allow combinations of linear predictors and penalised smoothers for the effect of time and covariates, both in the proportional hazards and proportional odds framework ([Liu et al., 2016](#)). More generally, models have been developed to account for competing events and multi-state diseases, even with intermediate states, and for modelling a wide range of multivariate survival data ([Geskus, 2015](#); [Crowder, 2016](#)). Finally, the main advances that I will discuss further are models with random effects and joint models for longitudinal and survival data, in Chapters [2](#) and [3](#) respectively.

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: imagine having 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. In comparison, fixed effects models do not take into account any hierarchy in the data. In biostatistics, also, the terms *fixed effects* and *random effects* have a special meaning, referring 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 (e.g. hemoglobin levels, inflammation markers, ...); 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. Additionally, 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, the generalisation is analogous to 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, of course) 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. Another 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 rather than frailties 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 α is a non-observed frailty effect and $h_0(t)$ is the baseline hazard function. The random variable α , 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 σ^2 increases the values of α 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 β 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 α

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

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 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 α 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 θ . 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 straightfor-

ward, 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, and therefore requires more complex estimation procedures involving numerical integration (taking the maximum likelihood approach), mathematical approximations (such as the Laplace approximation), or Markov Chain Monte Carlo methods (Clayton, 1991; Sinha et al., 2008). Further details in Chapter 4.

2.2 Shared frailty models

Further generalising the model presented in Section 2.1, it is possible for the frailty effect α 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 α_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 briefly 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 α_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 $\eta_i = \log \alpha_i$. η_i results being normally distributed with parameters μ and σ^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 β 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 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 (and each other). 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 sensible choice, as the longitudinal data can be important predictors or surrogates of the time to event outcome. A powerful tool to jointly model longitudinal and time-to-event data is 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 (in terms of CD4 lymphocyte cells count) 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). Of course, 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 structure of the association with the survival outcome: a review on recent developments, software, and persisting issues 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 1.3, let $y_{ij} = \{y_{ij}(t_{ij}) \forall j = 1, \dots, n_i\}$ be the observed longitudinal response for the i^{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 ψ . α 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 distribution from Section [1.5.1](#)) 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^{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 full joint likelihood has been maximised using 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; I will discuss them further in Chapter 4.

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^{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 conclude with additional considerations on numerical methods in Section 4.2.

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^{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 μ and standard deviation σ 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 Ω of volume $V(\Omega)$:

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

with $U \sim \text{uniform}$ over Ω . Drawing N uniform random vectors u_i an estimator for I_Ω 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}_Ω 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 Σ of the multivariate normal density at each step (e.g. using empirical Bayes estimates of M, Σ) 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 Other considerations

4.2.1 Cancellation error, precision, and arithmetic over- and under-flow

One of the problems when doing calculations on a computer is *cancellation error* (or *round-off error*). That occurs as a side effect of performing finite-precision arithmetic, as computers can store numbers in memory using a finite number of digits. Cancellation error causes the number of significant digits in the result to be reduced unacceptably; when a sequence of calculations is performed, cancellation errors add up significantly, altering the final result. Cancellation error can be easily reproduced:

```
a <- 1e16
b <- 1e16 + pi
b - a

## [1] 4

pi

## [1] 3.141593
```

$b - a$ should be π , instead it is 4. Analogously, *arithmetic over- and under-flow* is a condition that happens when the result of a calculation is, respectively, bigger or smaller than the minimum or maximum value that a given machine can store in memory. On the laptop used to produce this report, the smallest (and largest) floating-point number that the machine can represent are:

```
.Machine$double.xmin

## [1] 2.225074e-308

.Machine$double.xmax

## [1] 1.797693e+308
```

Next, precision. Machines can only distinguish numbers that they can represent as different. For instance:

```
a = 1
b = (.Machine$double.eps ^ 2)
c = (.Machine$double.neg.eps ^ 2)
d = a + b
e = a - c

# The following equalities should be FALSE
```

```

a == d
## [1] TRUE

a == e
## [1] TRUE

# Check that b, c are not 0
b == 0
## [1] FALSE

c == 0
## [1] FALSE

```

In this case, `.Machine$double.eps` and `.Machine$double.neg.eps` are the smallest positive floating-point numbers x such that $1 + x \neq 1$ and $1 - x \neq 1$, respectively.

It is necessary to keep this potential problems in mind when doing numerical calculation using finite-precision arithmetic; for instance, a common situation where we may incur in arithmetic over- or under-flow is when maximising a likelihood. That is, the product of the individual contributions to the likelihood may be a number so large (or so small) that the computer cannot distinguish it from $\pm\infty$, or rounding error may seriously affect the results. This specific example is easy to fix by using the log-likelihood instead, as the sum of the logarithm of the individual contributions behaves much better; nevertheless, this problem may not be always evident nor as easy to solve.

4.2.2 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 ϵ 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.

4.2.3 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 β . 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 τ , it is possible to derive τ 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 β ;
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 $h_0(t)$	λ	$\lambda p t^{p-1}$	$\exp(\gamma t)$
Cumulative hazard function $H_0(t)$	λt	λt^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 τ	$-\frac{\log(u)}{\lambda \exp(X\beta)}$	$\left[-\frac{\log(u)}{\lambda \exp(X\beta)}\right]^{1/p}$	$(1/\gamma) \log \left[1 - \frac{\gamma \log(u)}{\lambda \exp(X\beta)}\right]$

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 τ ; 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 τ numerically (Section 4.2.3). Formally, the survival time τ 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 τ .

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, as I showed in Section 2.2 that in that setting the likelihood has a closed form; then, I will choose a log-normal distribution for the frailty using the random intercept parametrisation as explained in Section 2.3 for answering the second aim. The likelihood of this model doesn't have a closed form, hence I will need numerical integration to estimate the model.

6.2 Data-generating mechanisms

I generated survival data from a Weibull distribution with shape parameter $\lambda = 0.5$ and scale parameter $p = 0.6$ (Figure 6.1) 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 θ (for identifiability purposes) and then by drawing from a normal distribution with mean $\mu = 0$ and standard deviation $\sigma = \sqrt{\theta}$. I varied θ : $\theta = \{0.25, 0.75, 1.25\}$. I also varied the regression coefficient (e.g. the log-treatment effect) β associated with the binary covariate: $\beta = \{-0.50, 0.00, 0.50\}$. I simulated data for six different sample sizes:

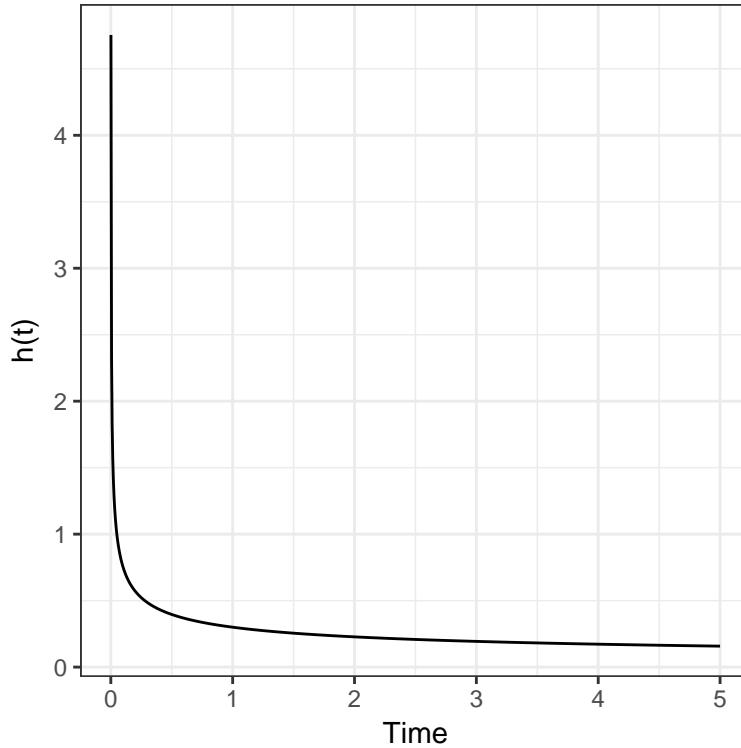


Figure 6.1: Weibull baseline hazard function used in Simulation 1 ($\lambda = 0.5, p = 0.6$)

15 clusters of 30, 100, or 500 individuals each, 50 clusters of 30 or 100 individuals, and 1000 clusters of 2 individuals. As a result of this, sample size 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.

I will present the results separately by frailty variance, as I will use the 54 scenarios with a Gamma frailty to answer the first aim and the remaining 54 scenarios with a log-normal frailty 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

First, I am interested 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 $b = E(\hat{\beta}) - \beta$. Additionally, I will compute relative bias (defined as $100 \times [E(\hat{\beta}) - \beta]/\beta$) for presentation purposes, as it will be useful to compare bias for estimands with different magnitude (and therefore bias may be greater in absolute value but smaller in relative value).

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 β . This allows 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 λ , p , and θ 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 (see [White \(2010\)](#) for further details). I will also discuss convergence rates of the different methods included in the comparison.

6.6 Results

I selected three scenarios for each aim of this simulation study (out of 54) to present, for conciseness. Specifically, I will present in this report:

1. small frailty variance (0.25) and negative regression coefficient (-0.50);
2. large frailty variance (1.25) and null regression coefficient (0.00);

3. 1000 clusters of 2 individuals each and positive 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

In this Section I will present results for each aforementioned scenario and the simulation comparing quadrature methods with analytical formulae, that is, fitting a shared Gamma frailty Weibull regression model.

First, 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 β ; conversely, the methods performed quite differently for the shape parameter λ and the frailty variance θ . 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 θ 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.

Next, bias, coverage, mean squared error, and convergence rates for scenario 2 are presented in Tables A.5, A.6, A.7, A.8 and Figures B.6, B.7, B.8, B.5. Analogously as in scenario 1, convergence rate was generally good with Gauss-Kronrod quadrature performing the worst and estimates of the scale parameter p and the regression coefficient β were unbiased with optimal coverage and low mean squared error. Bias for the shape parameter was substantial when using Gauss-Laguerre quadrature with 15 nodes, up to -0.5 on the log-scale; bias was reduced greatly by increasing the number of quadrature nodes, except when assuming 15 clusters of 500 individuals each where it remained substantial (approximately -0.35 on the log-scale) even when using 105 quadrature nodes. Coverage and mean squared error followed the same pattern; however, an exception was found in the case of 1000 clusters of 2 individuals: in that setting, λ was estimated properly with small bias and good coverage and mean squared error. The estimated frailty variance also followed again the pattern depicted in scenario 1: analytical formulae required 50 or 1000 clusters to yield unbiased results, Gauss-Kronrod quadrature performed similarly to analytical formulae, and Gauss-Laguerre quadrature generally yielded better results with a greater number of quadrature nodes. Coverage was also generally good, symptom once again of overestimated standard errors - except the setting of 1000 clusters of 2 individuals each in which Gauss-Laguerre quadrature with 15 or 35 nodes did not cover the true value at all; the situation improved with a greater number of nodes, up to a coverage of 68.3% with 105 nodes. Inflated standard errors for the estimated frailty variance yielded increased mean squared error in this scenario as well, with all the methods performing similarly except Gauss-Laguerre quadrature with

15 knots which performed the worst once again.

Finally, bias, coverage, percentage bias, mean squared error, and convergence rates for scenario 3 are presented in Tables A.9, A.10, A.11, A.12, A.13 and Figures B.10, B.11, B.12, B.13, B.9. Convergence rates were good as in the previous scenarios, with Gauss-Kronrod quadrature performing the worst again. The scale parameter p and the regression coefficient were estimated optimally once again, and further to that, in scenario 3 the shape parameter λ was also well estimated (in terms of bias, coverage, and overall accuracy) with the exception of methods using Gauss-Laguerre quadrature when the true frailty variance was large; in that setting, the shape parameter was slightly overestimated (3% to 8%), coverage was suboptimal (60-80%) and mean squared error was greatest. Nevertheless, performance improved when increasing the number of quadrature nodes to approach results obtained using analytical formulae and Gauss-Kronrod quadrature. Finally, performance measures for the estimated frailty variance were generally good with a true frailty variance small or medium; Gauss-Kronrod undercovered the true value (coverage of approximately 90%). Conversely, when the true frailty variance was large, only analytical formulae and Gauss-Kronrod quadrature performed well. Gauss-Laguerre quadrature yielded severely underestimated results, up to -140% with 15 nodes, with null or poor coverage and large mean squared error.

6.6.2 Aim 2: accuracy when analytical formulae are not available

In this section I will present results for the simulation comparing Gauss-Hermite quadrature with varying number of knots when fitting a model for which it is not possible to derive analytical formulae, specifically a Weibull regression model with a random intercept.

First, bias, coverage, mean squared error, and convergence rates for scenario 1 are presented in Tables A.14, A.15, A.16, A.17 and Figures B.15, B.16, B.17, B.14. 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 λ and p and the regression coefficient β : between 0.0059 and 0.0193 for λ , between -0.0424 and -0.0332 for p , between 0.0040 and 0.0867 for β . Conversely, estimates for σ 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 β and the frailty variance σ coverage was good or superoptimal for the remaining sample sizes, with the exception of σ 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 λ 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 λ , p , and β , comparable for σ with a sample size of 50 clusters - 30 individuals and 50 clusters - 100 individuals, much higher for σ with all the remaining sample sizes (i.e. overall accuracy was lower in these settings).

Next, bias, coverage, mean squared error, and convergence rates for scenario 2 are presented in Tables A.18, A.19, A.20, A.21 and Figures B.19, B.20, B.21, B.18. Analogously as before, coverage was generally good except for the sample size of 15 clusters - 100 individuals, where only 15% of replications successfully converged to a solution. Bias decreased for the regression coefficient β with increasing number of quadrature nodes; conversely, bias for the scale parameter p was not affected by the number of quadrature nodes and remained more or less constant across sample sizes at approximately -0.15. Bias for the shape parameter λ was also not affected by the quadrature nodes, with negative bias for the sample size of 15 clusters - 500 individuals (approximately -0.05) and positive bias elsewhere (0.0241 to 0.0622). σ was

consistently underestimated, with bias ranging between -0.3291 and -0.0821. The pattern for coverage was similar: coverage for β was good assuming at least 75 quadrature nodes were used. With a sample size of 1000 clusters - 2 individuals, 15 clusters - 30 individuals, 50 clusters - 30 individuals coverage for the regression coefficient was good irrespectively of the number of quadrature nodes, and for a sample size of 15 clusters - 500 individuals coverage was mediocre even when using 105 quadrature nodes. Coverage of λ and p was generally poor, between 0% and 64%. Coverage for σ was only slightly suboptimal with a sample size of 15 clusters - 30 individuals and 50 clusters - 30 individuals ($> 75\%$), poor to suboptimal otherwise (7% to 75%, with peaks above 80% when using 75 or 105 quadrature nodes). Bias and poor coverage were reflected in the mean squared error; interestingly, it seemed to generally decrease with increasing number of quadrature nodes (especially for the regression coefficient and the variance of the random intercept).

Finally, bias, coverage, mean squared error, and convergence rates for scenario 3 are presented in Tables A.22, A.23, A.25, A.26 and Figures B.23, B.24, B.26, B.22. Convergence rates were good, above 99% for each method included in the comparison and variance magnitude. Negative bias was present for λ , and β , between -15% and -2%; positive bias for p increased with greater variance magnitude, from 7% for a true $\sigma = 0.25$ to 35% for a true $\sigma = 1.25$. Estimates of σ with a small-medium frailty variance were positively biased (30 to 180%), negatively biased otherwise (-300%). Coverage was null to poor for parameters of the Weibull distribution and a medium-large frailty variance, less than 80%; with a small frailty variance, coverage was good for λ and poor for p . Coverage for the regression parameter β was good to suboptimal, decreasing to 75-80% with a large frailty variance. Coverage for σ was superoptimal for a small frailty variance, null to poor for a medium-large frailty variance. Mean squared error was greatest for σ , irrespectively of the magnitude of the frailty variance; conversely, mean squared error tended to increase as the frailty variance increased.

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 β 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 38th 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

I introduced in Section 1.6 the de-facto standard method used in medical research when dealing with time to event data: the Cox proportional hazards model. It is best suited when relative risk estimates are the 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 (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 of such data consist of multi-centre clinical trial data, individual-patient data meta-analysis, observational data with geographical clusters, As a consequence, survival times of individuals within a cluster are likely to be correlated and need to be analysed as such. Analogously, correlated data may emerge as 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 economic 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 a given cluster.

As I 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 (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. I aim therefore with this work to assess the impact of misspecifying the baseline hazard or the distribution of the frailty 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.

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

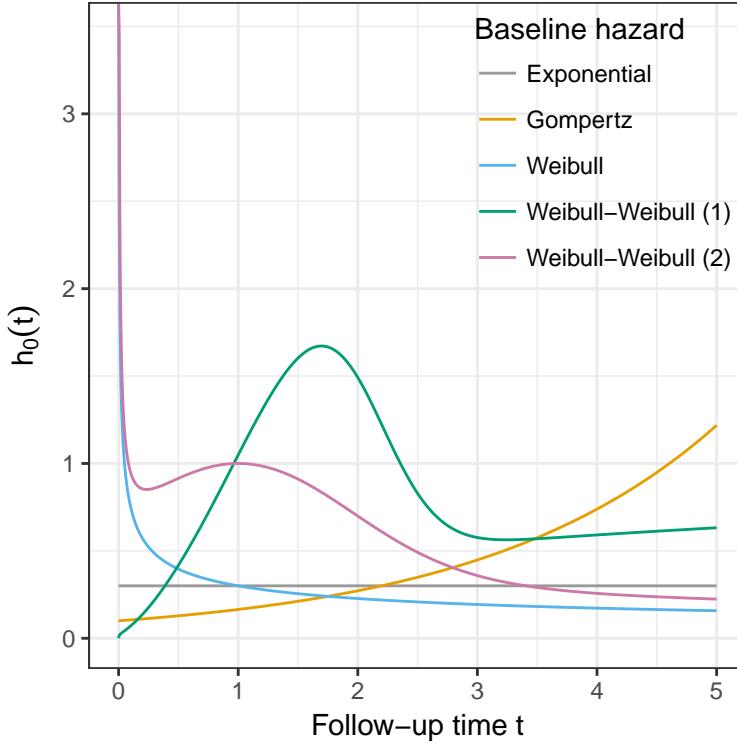


Figure 7.1: Baseline hazard functions chosen for this simulation study.

$\lambda_1 = 0.5$, $\lambda_2 = 0.3$, 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 log-hazard ratio of -0.5 and cluster-specific frailty terms α_i following either a Gamma or a log-normal distribution with variance θ ($\theta = \{0.25, 0.75, 1.25\}$, 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.

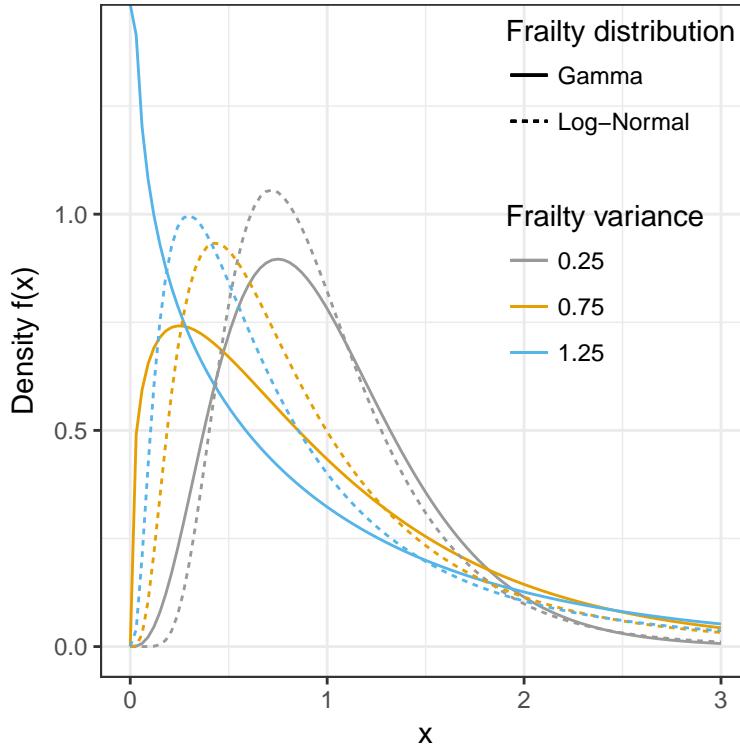


Figure 7.2: Frailty distributions chosen for this simulation study.

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, I computed a bootstrapped standard error using 1000 bootstrap replications; I resampled clusters of individuals rather than individuals to preserve the correlation between and within clusters. Then, I fitted fully parametric survival models with a shared frailty term, using the same model formulation of the Cox model but specifying

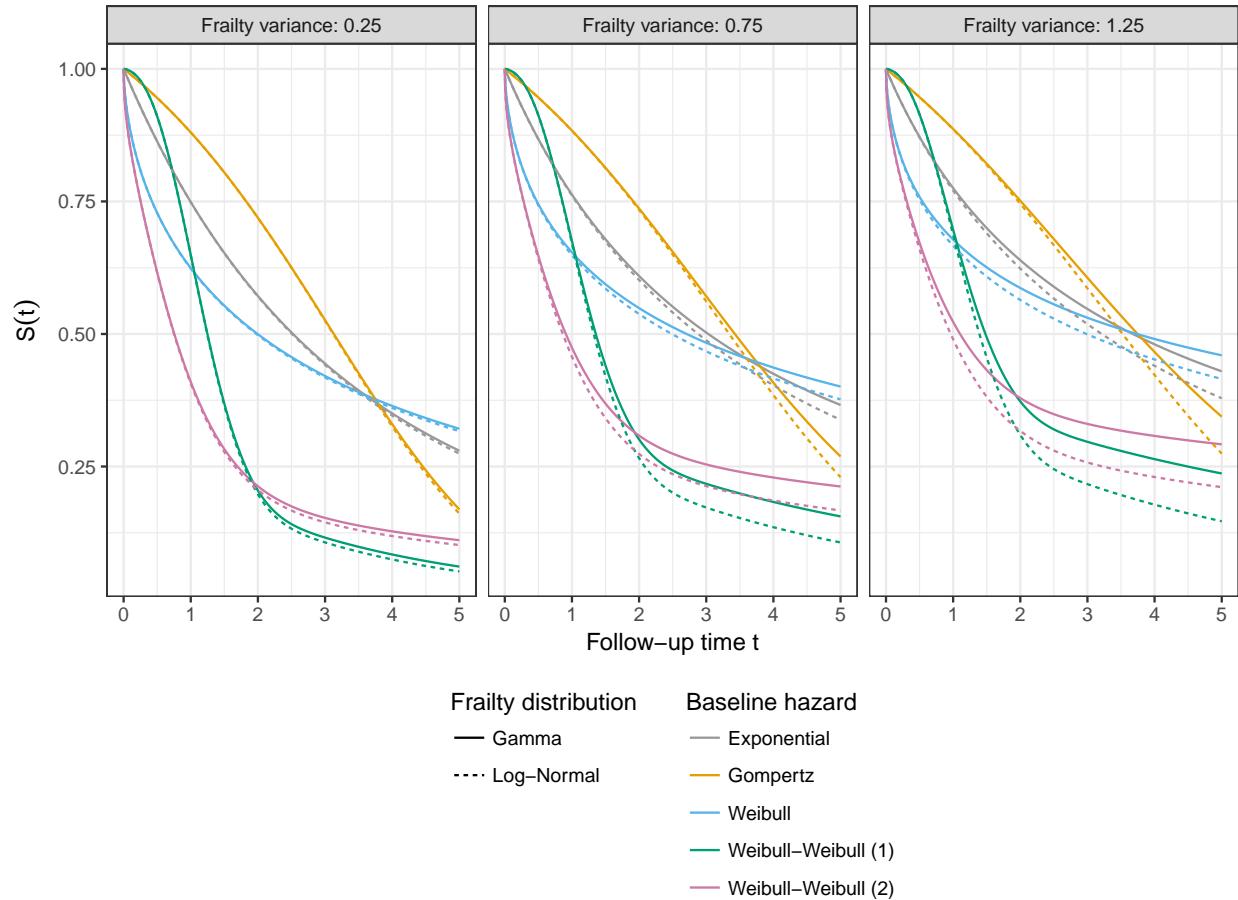


Figure 7.3: Marginal survival depending on baseline hazard and frailty distribution chosen for this simulation study.

the baseline hazard 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 in the regression coefficient β 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 presented in Chapter 2. Conversely,

the latter requires integrating the marginal survival function for both treatment groups and then computing their difference; I compute 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 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

Analogously with the previous simulation study of Chapter 6, the main performance measures of interest are bias, coverage, and mean squared error. A more in detail presentation of the characteristics of each summary measure is given in Section 6.5. I will be presenting percentage bias when relevant (i.e. when comparing estimates with different magnitude), and I will include Monte Carlo standard errors for bias and coverage to quantify uncertainty in estimating such performance measures.

7.6 Results

I mentioned in Section 7.2 that I simulated data under 150 different data-generating mechanisms. In this Section I will present results for a single scenario, for conciseness: 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. Nevertheless, the full results of this simulation study can be explored interactively by clicking [here](#).

Convergence rates for the models included in this comparison are presented in Table A.27 and plot B.27. Convergence rates varied considerably between different baseline hazards and frailty distributions, 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 when the baseline hazard was misspecified. With the first Weibull-Weibull baseline hazard convergence was optimal for all models except the Cox model with a Gamma frailty; finally, 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.28, A.29, A.30 and Figures B.28, B.29, B.30. 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.31, A.32, A.33 and Figures B.31, B.32, B.33. 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.34, A.35 and Figures B.34, B.35. 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 measures of heterogeneity such as the variance of the frailty. 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 - which is clinically meaningful. 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 (with bias up to 1.5 months over 5 years), but it is important to bear that 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. I will present further details on the rationale behind developing such tool in Section 8.1, and introduce the interactive tool in Section 8.2. I presented part of this work at the Students' Day of the 38th Conference of the International Society for

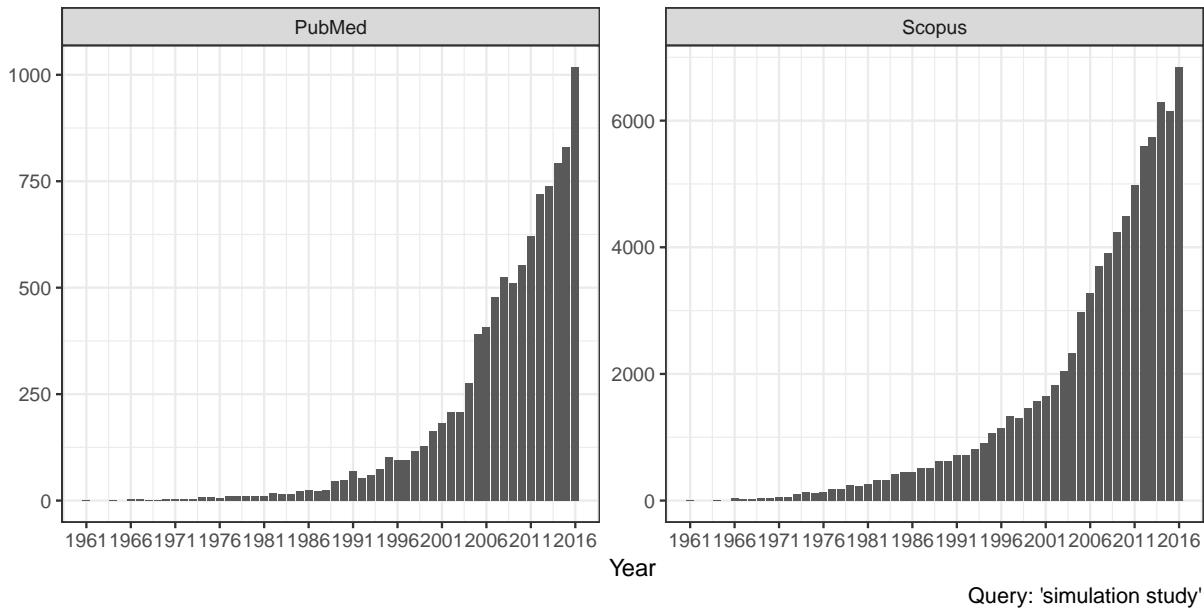


Figure 8.1: Numer of results querying 'simulation study' on Pubmed and Scopus.

Clinical Biostatistics; slides are attached in Appendix C.3.

8.1 Rationale

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 as aforementioned.

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.

8.2 SiReX

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, a web application framework for R that allows creating interactive web applications in a straightforward way (<https://shiny.rstudio.com/>).

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 when using 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 facilities at the time of writing, and should finish soon. Consequently, I will summarise the results in a meaningful way using tables and plots 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; 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, i.e. the parametrisation that links the two processes (e.g. the current value parametrisation, intercept and slope, cumulative effect, and so on);
6. eventually, sample size in terms of study subjects.

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 packages such as R and Stata. Throughout this project we will be collaborating with Dr. Jessica Barrett from the MRC Biostatistic Unit in Cambridge, and we will have extensive meetings 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 high-performance computing facilities of the University. Finally, I plan to write up the project and produce 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 classify whether additional biomarkers can improve predictions or not; this project will have wide reaching consequences by providing guidance in the use of joint models when developing risk predictions using multiple biomarkers. 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. For instance, 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.

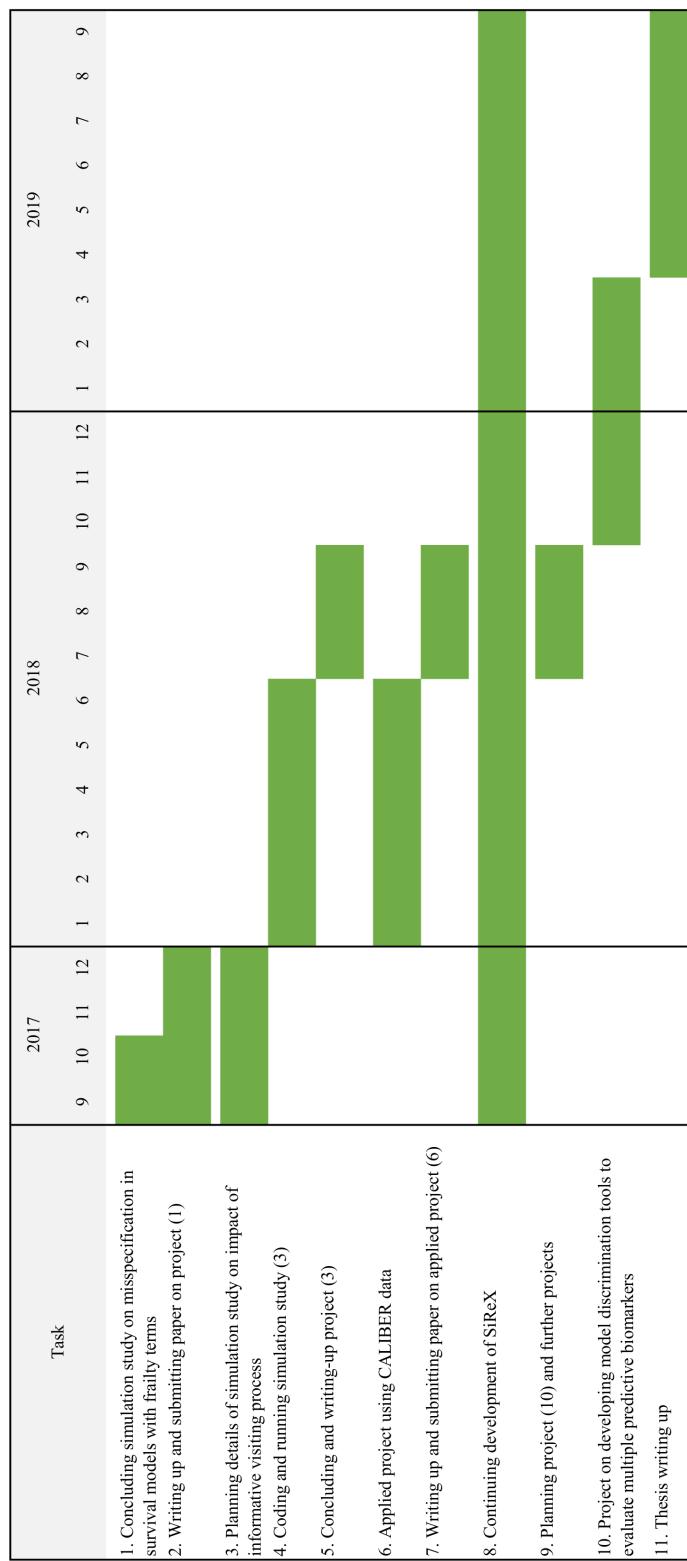


Figure 10.1: Gantt chart for current and future projects during my PhD.

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. In particular, 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 produced and shared between us. A comprehensive list is available on PROSE. Additionally, we held informal meetings to discuss developments and more urgent matters more often, whenever it was needed and without scheduling it.

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 8th 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, RCpp, and parallel programming. General hints and tips were provided.
- *Introduction to causal inference*, on April 25th and 26th 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 22nd 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 5th, 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 and 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 10th, 2016), a course on *Data visualisation* (November 15th, 2016), and a course on *High performance computing at*

Leicester (February 8th, 2017). The latter was particularly important, as it allowed me to make better use of the high-performance computing facilities offered by the University. I also attended the *Preparing to teach in higher education* workshop, strand A (July 24th and 27th 2017).

Additionally, I have attended the following PROSE training sessions to develop personal and communication skills in research settings. These are listed below:

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

11.3 Conferences

I have 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 5th and 6th 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 3rd and 4th 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 9th to July 13th 2017. I delivered two talks: a titled *Impact of model misspecification in survival models with frailties* during the main conference, and a talk titled *Exploring results from simulation studies interactively* during the Students' Day organised on July 13th.

Additionally, I delivered an oral presentation on previous work external to my PhD project during the 54th ERA-EDTA Congress held in Madrid, Spain, between June 3rd and June 6th. 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

A.1 Simulation study 1

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 with analytical formulae, scenario with a large frailty variance and a null regression coefficient.

Sample size	Method	Lambda	P	Beta	Theta
1000c. of 2i.	AF	-0.0013 (0.0021)	0.0025 (0.0009)	-0.0021 (0.0024)	-0.0008 (0.0032)
1000c. of 2i.	IN	0.0003 (0.0021)	0.0020 (0.0009)	-0.0018 (0.0024)	-0.0044 (0.0032)
1000c. of 2i.	GQ15	-0.0637 (0.0019)	-0.0376 (0.0008)	-0.0018 (0.0023)	-0.3201 (0.0010)
1000c. of 2i.	GQ35	-0.0493 (0.0019)	-0.0264 (0.0008)	-0.0018 (0.0023)	-0.2203 (0.0014)
1000c. of 2i.	GQ75	-0.0368 (0.0019)	-0.0183 (0.0009)	-0.0019 (0.0023)	-0.1550 (0.0017)
1000c. of 2i.	GQ105	-0.0319 (0.0019)	-0.0153 (0.0009)	-0.0019 (0.0023)	-0.1319 (0.0018)
15c. of 100i.	AF	-0.0586 (0.0088)	0.0014 (0.0010)	-0.0001 (0.0023)	-0.1229 (0.0113)
15c. of 100i.	IN	-0.0672 (0.0092)	0.0010 (0.0010)	0.0008 (0.0023)	-0.1375 (0.0114)
15c. of 100i.	GQ15	-0.4546 (0.0095)	-0.0027 (0.0010)	0.0001 (0.0022)	-0.0652 (0.0109)
15c. of 100i.	GQ35	-0.3520 (0.0089)	-0.0007 (0.0010)	-0.0000 (0.0023)	-0.0923 (0.0104)
15c. of 100i.	GQ75	-0.2557 (0.0082)	-0.0000 (0.0010)	0.0000 (0.0023)	-0.1170 (0.0106)
15c. of 100i.	GQ105	-0.1829 (0.0078)	0.0004 (0.0010)	-0.0001 (0.0023)	-0.1322 (0.0106)
15c. of 30i.	AF	-0.0407 (0.0099)	0.0020 (0.0018)	-0.0055 (0.0042)	-0.1121 (0.0121)
15c. of 30i.	IN	-0.0435 (0.0099)	0.0016 (0.0018)	-0.0059 (0.0042)	-0.1149 (0.0120)
15c. of 30i.	GQ15	-0.2776 (0.0097)	-0.0049 (0.0018)	-0.0058 (0.0042)	-0.1773 (0.0100)
15c. of 30i.	GQ35	-0.0915 (0.0098)	-0.0003 (0.0018)	-0.0056 (0.0042)	-0.1758 (0.0104)
15c. of 30i.	GQ75	-0.0465 (0.0101)	0.0013 (0.0018)	-0.0055 (0.0042)	-0.1550 (0.0108)
15c. of 30i.	GQ105	-0.0476 (0.0099)	0.0014 (0.0018)	-0.0055 (0.0042)	-0.1484 (0.0109)
15c. of 500i.	AF	-0.0369 (0.0093)	0.0007 (0.0004)	0.0010 (0.0011)	-0.1249 (0.0112)
15c. of 500i.	IN	-0.2891 (0.0081)	0.0002 (0.0004)	0.0010 (0.0011)	-0.1394 (0.0102)
15c. of 500i.	GQ15	-0.4658 (0.0101)	-0.0042 (0.0007)	0.0008 (0.0011)	-0.0547 (0.0104)
15c. of 500i.	GQ35	-0.4024 (0.0091)	-0.0011 (0.0005)	0.0008 (0.0010)	-0.0676 (0.0105)
15c. of 500i.	GQ75	-0.3619 (0.0084)	-0.0007 (0.0006)	0.0009 (0.0010)	-0.0733 (0.0108)
15c. of 500i.	GQ105	-0.3442 (0.0082)	-0.0002 (0.0006)	0.0010 (0.0011)	-0.0748 (0.0107)
50c. of 100i.	AF	-0.0146 (0.0050)	-0.0000 (0.0005)	-0.0016 (0.0012)	-0.0363 (0.0058)
50c. of 100i.	IN	-0.0376 (0.0050)	-0.0005 (0.0005)	-0.0006 (0.0012)	-0.0671 (0.0058)
50c. of 100i.	GQ15	-0.4998 (0.0074)	-0.0042 (0.0005)	-0.0016 (0.0013)	0.0609 (0.0072)
50c. of 100i.	GQ35	-0.3359 (0.0060)	-0.0024 (0.0005)	-0.0016 (0.0012)	-0.0120 (0.0056)
50c. of 100i.	GQ75	-0.1945 (0.0053)	-0.0014 (0.0005)	-0.0016 (0.0012)	-0.0455 (0.0053)
50c. of 100i.	GQ105	-0.1000 (0.0052)	-0.0008 (0.0005)	-0.0016 (0.0012)	-0.0560 (0.0053)
50c. of 30i.	AF	-0.0170 (0.0053)	0.0003 (0.0010)	-0.0000 (0.0022)	-0.0219 (0.0061)
50c. of 30i.	IN	-0.0126 (0.0053)	0.0007 (0.0009)	0.0007 (0.0022)	-0.0384 (0.0060)
50c. of 30i.	GQ15	-0.2482 (0.0064)	-0.0069 (0.0010)	0.0006 (0.0022)	-0.1152 (0.0050)
50c. of 30i.	GQ35	-0.0618 (0.0056)	-0.0019 (0.0010)	-0.0001 (0.0022)	-0.0958 (0.0050)
50c. of 30i.	GQ75	-0.0238 (0.0052)	-0.0004 (0.0010)	0.0000 (0.0022)	-0.0656 (0.0053)
50c. of 30i.	GQ105	-0.0224 (0.0052)	-0.0003 (0.0010)	0.0000 (0.0022)	-0.0570 (0.0054)

Table A.6: Coverage, comparison with analytical formulae, scenario with a large frailty variance and a null regression coefficient.

Sample size	Method	Lambda	P	Beta	Theta
1000c. of 2i.	AF	96.4000 (0.5891)	94.6000 (0.7147)	95.2000 (0.6760)	94.1000 (0.7451)
1000c. of 2i.	IN	92.6254 (0.8265)	93.9528 (0.7538)	94.2478 (0.7363)	87.3156 (1.0524)
1000c. of 2i.	GQ15	82.2000 (1.2096)	73.5000 (1.3956)	96.1000 (0.6122)	0.0000 (0.0000)
1000c. of 2i.	GQ35	88.3000 (1.0164)	84.6000 (1.1414)	96.0000 (0.6197)	1.7000 (0.4088)
1000c. of 2i.	GQ75	91.6000 (0.8772)	90.7000 (0.9184)	95.9000 (0.6270)	47.0000 (1.5783)
1000c. of 2i.	GQ105	92.7000 (0.8226)	92.5000 (0.8329)	95.9000 (0.6270)	68.3000 (1.4714)
15c. of 100i.	AF	94.2000 (0.7392)	95.9000 (0.6270)	95.0000 (0.6892)	92.6000 (0.8278)
15c. of 100i.	IN	74.1729 (1.3841)	95.4109 (0.6617)	93.8100 (0.7620)	90.7150 (0.9178)
15c. of 100i.	GQ15	11.1000 (0.9934)	93.7000 (0.7683)	95.2000 (0.6760)	94.0000 (0.7510)
15c. of 100i.	GQ35	17.0000 (1.1879)	95.6000 (0.6486)	94.9000 (0.6957)	94.9000 (0.6957)
15c. of 100i.	GQ75	36.9000 (1.5259)	95.6000 (0.6486)	95.1000 (0.6826)	93.9000 (0.7568)
15c. of 100i.	GQ105	51.5000 (1.5804)	95.9000 (0.6270)	94.9000 (0.6957)	93.5000 (0.7796)
15c. of 30i.	AF	92.8000 (0.8174)	96.1000 (0.6122)	96.0000 (0.6197)	93.0000 (0.8068)
15c. of 30i.	IN	91.8665 (0.8644)	96.0375 (0.6169)	95.6204 (0.6471)	91.8665 (0.8644)
15c. of 30i.	GQ15	46.7000 (1.5777)	96.0000 (0.6197)	95.6000 (0.6486)	94.4000 (0.7271)
15c. of 30i.	GQ35	78.0000 (1.3100)	96.1000 (0.6122)	96.0000 (0.6197)	94.0000 (0.7510)
15c. of 30i.	GQ75	88.8000 (0.9973)	95.9000 (0.6270)	96.0000 (0.6197)	93.7000 (0.7683)
15c. of 30i.	GQ105	92.0000 (0.8579)	96.1000 (0.6122)	96.0000 (0.6197)	93.8000 (0.7626)
15c. of 500i.	AF	92.3000 (0.8430)	94.4000 (0.7271)	94.1000 (0.7451)	92.9000 (0.8122)
15c. of 500i.	IN	12.9291 (1.0610)	95.0801 (0.6839)	92.9062 (0.8118)	92.1053 (0.8527)
15c. of 500i.	GQ15	3.8066 (0.6051)	86.1111 (1.0936)	94.1358 (0.7430)	93.9300 (0.7551)
15c. of 500i.	GQ35	5.1177 (0.6968)	90.9928 (0.9053)	94.7799 (0.7034)	94.6776 (0.7099)
15c. of 500i.	GQ75	7.3245 (0.8239)	92.6755 (0.8239)	94.6083 (0.7142)	93.8962 (0.7570)
15c. of 500i.	GQ105	8.9340 (0.9020)	93.2995 (0.7907)	94.2132 (0.7384)	94.2132 (0.7384)
50c. of 100i.	AF	95.2000 (0.6760)	95.7000 (0.6415)	94.4000 (0.7271)	93.8000 (0.7626)
50c. of 100i.	IN	71.3376 (1.4299)	94.2675 (0.7351)	87.5159 (1.0453)	84.4586 (1.1457)
50c. of 100i.	GQ15	2.7000 (0.5126)	94.0000 (0.7510)	94.0000 (0.7510)	85.4000 (1.1166)
50c. of 100i.	GQ35	9.5000 (0.9272)	94.4000 (0.7271)	94.9000 (0.6957)	94.9000 (0.6957)
50c. of 100i.	GQ75	31.3000 (1.4664)	95.1000 (0.6826)	94.8000 (0.7021)	95.7000 (0.6415)
50c. of 100i.	GQ105	54.3000 (1.5753)	95.6000 (0.6486)	94.5000 (0.7209)	95.0000 (0.6892)
50c. of 30i.	AF	94.6000 (0.7147)	96.0000 (0.6197)	95.6000 (0.6486)	95.1000 (0.6826)
50c. of 30i.	IN	89.7114 (0.9607)	95.6085 (0.6480)	91.8444 (0.8655)	91.3425 (0.8893)
50c. of 30i.	GQ15	33.6000 (1.4937)	94.6000 (0.7147)	96.4000 (0.5891)	95.1000 (0.6826)
50c. of 30i.	GQ35	81.4000 (1.2305)	95.8000 (0.6343)	95.9000 (0.6270)	95.8000 (0.6343)
50c. of 30i.	GQ75	93.7000 (0.7683)	95.8000 (0.6343)	95.7000 (0.6415)	96.2000 (0.6046)
50c. of 30i.	GQ105	94.2000 (0.7392)	95.8000 (0.6343)	95.7000 (0.6415)	96.2000 (0.6046)

Table A.7: Mean squared error, comparison with analytical formulae, scenario with a large frailty variance and a null regression coefficient.

Sample size	Method	Lambda	P	Beta	Theta
1000c. of 2i.	AF	0.0042	0.0009	0.0057	0.0100
1000c. of 2i.	IN	0.0043	0.0009	0.0056	0.0104
1000c. of 2i.	GQ15	0.0075	0.0021	0.0051	0.1035
1000c. of 2i.	GQ35	0.0060	0.0014	0.0052	0.0505
1000c. of 2i.	GQ75	0.0050	0.0011	0.0053	0.0270
1000c. of 2i.	GQ105	0.0047	0.0010	0.0054	0.0208
15c. of 100i.	AF	0.0806	0.0009	0.0052	0.1437
15c. of 100i.	IN	0.0888	0.0009	0.0051	0.1482
15c. of 100i.	GQ15	0.2960	0.0010	0.0051	0.1221
15c. of 100i.	GQ35	0.2031	0.0009	0.0051	0.1173
15c. of 100i.	GQ75	0.1321	0.0009	0.0051	0.1255
15c. of 100i.	GQ105	0.0948	0.0009	0.0051	0.1291
15c. of 30i.	AF	0.0994	0.0032	0.0180	0.1583
15c. of 30i.	IN	0.0995	0.0032	0.0180	0.1581
15c. of 30i.	GQ15	0.1716	0.0032	0.0178	0.1316
15c. of 30i.	GQ35	0.1052	0.0032	0.0179	0.1385
15c. of 30i.	GQ75	0.1038	0.0032	0.0179	0.1404
15c. of 30i.	GQ105	0.1000	0.0032	0.0180	0.1417
15c. of 500i.	AF	0.0882	0.0002	0.0011	0.1402
15c. of 500i.	IN	0.1485	0.0002	0.0011	0.1234
15c. of 500i.	GQ15	0.3182	0.0005	0.0011	0.1106
15c. of 500i.	GQ35	0.2453	0.0002	0.0011	0.1142
15c. of 500i.	GQ75	0.2012	0.0004	0.0011	0.1231
15c. of 500i.	GQ105	0.1859	0.0003	0.0011	0.1204
50c. of 100i.	AF	0.0249	0.0003	0.0015	0.0352
50c. of 100i.	IN	0.0263	0.0003	0.0015	0.0378
50c. of 100i.	GQ15	0.3048	0.0003	0.0016	0.0551
50c. of 100i.	GQ35	0.1491	0.0003	0.0015	0.0315
50c. of 100i.	GQ75	0.0661	0.0003	0.0015	0.0301
50c. of 100i.	GQ105	0.0375	0.0003	0.0015	0.0312
50c. of 30i.	AF	0.0279	0.0009	0.0049	0.0375
50c. of 30i.	IN	0.0279	0.0009	0.0048	0.0380
50c. of 30i.	GQ15	0.1023	0.0010	0.0048	0.0383
50c. of 30i.	GQ35	0.0351	0.0009	0.0049	0.0340
50c. of 30i.	GQ75	0.0281	0.0009	0.0049	0.0319
50c. of 30i.	GQ105	0.0279	0.0009	0.0049	0.0320

Table A.8: Convergence rates, comparison with analytical formulae, scenario with a large frailty variance and a null regression coefficient.

Sample size	AF	IN	GQ15	GQ35	GQ75	GQ105
1000c. of 2i.	100.0%	67.8%	100.0%	100.0%	100.0%	100.0%
15c. of 100i.	100.0%	93.7%	100.0%	100.0%	100.0%	100.0%
15c. of 30i.	100.0%	95.9%	100.0%	100.0%	100.0%	100.0%
15c. of 500i.	100.0%	87.4%	97.2%	97.7%	98.3%	98.5%
50c. of 100i.	100.0%	78.5%	100.0%	100.0%	100.0%	100.0%
50c. of 30i.	100.0%	79.7%	100.0%	100.0%	100.0%	100.0%

Table A.9: Bias, comparison without analytical formulae, scenario with 1000 clusters of 2 individuals and a positive regression coefficient.

fv	Method	Lambda	P	Beta	Theta
0.25	AF	-0.0016 (0.0015)	0.0002 (0.0008)	-0.0018 (0.0018)	-0.0342 (0.0070)
0.25	IN	0.0003 (0.0015)	0.0002 (0.0008)	-0.0027 (0.0018)	-0.0271 (0.0067)
0.25	GQ15	0.0069 (0.0014)	0.0073 (0.0007)	0.0035 (0.0018)	0.0723 (0.0016)
0.25	GQ35	-0.0015 (0.0015)	0.0004 (0.0008)	-0.0016 (0.0018)	-0.0296 (0.0067)
0.25	GQ75	-0.0016 (0.0015)	0.0002 (0.0008)	-0.0018 (0.0018)	-0.0342 (0.0070)
0.25	GQ105	-0.0016 (0.0015)	0.0002 (0.0008)	-0.0018 (0.0018)	-0.0342 (0.0070)
0.75	AF	0.0003 (0.0018)	0.0004 (0.0009)	0.0005 (0.0022)	-0.0032 (0.0035)
0.75	IN	-0.0008 (0.0018)	0.0005 (0.0009)	-0.0000 (0.0022)	-0.0039 (0.0035)
0.75	GQ15	0.0065 (0.0018)	0.0024 (0.0008)	0.0024 (0.0022)	-0.0001 (0.0017)
0.75	GQ35	0.0040 (0.0018)	0.0025 (0.0008)	0.0022 (0.0022)	0.0110 (0.0026)
0.75	GQ75	0.0023 (0.0018)	0.0017 (0.0008)	0.0015 (0.0022)	0.0065 (0.0031)
0.75	GQ105	0.0017 (0.0018)	0.0014 (0.0009)	0.0013 (0.0022)	0.0043 (0.0032)
1.25	AF	-0.0006 (0.0022)	0.0007 (0.0009)	-0.0004 (0.0024)	-0.0011 (0.0030)
1.25	IN	-0.0009 (0.0021)	0.0005 (0.0009)	0.0006 (0.0024)	0.0015 (0.0029)
1.25	GQ15	-0.0575 (0.0020)	-0.0426 (0.0008)	-0.0312 (0.0023)	-0.3082 (0.0010)
1.25	GQ35	-0.0437 (0.0020)	-0.0299 (0.0008)	-0.0226 (0.0023)	-0.2074 (0.0014)
1.25	GQ75	-0.0320 (0.0021)	-0.0209 (0.0008)	-0.0162 (0.0023)	-0.1436 (0.0017)
1.25	GQ105	-0.0276 (0.0021)	-0.0177 (0.0008)	-0.0139 (0.0024)	-0.1215 (0.0018)

Table A.10: Coverage, comparison with analytical formulae, scenario with 1000 clusters of 2 individuals and a positive regression coefficient.

Frailty variance	Method	Lambda	P	Beta	Theta
0.25	AF	94.9000 (0.6957)	93.5000 (0.7796)	96.0000 (0.6197)	95.7000 (0.6415)
0.25	IN	94.4730 (0.7226)	93.4447 (0.7827)	95.1157 (0.6816)	86.5039 (1.0805)
0.25	GQ15	95.5000 (0.6556)	93.6000 (0.7740)	96.2000 (0.6046)	99.8000 (0.1413)
0.25	GQ35	94.8898 (0.6964)	93.7876 (0.7633)	95.8918 (0.6276)	94.3888 (0.7278)
0.25	GQ75	94.9000 (0.6957)	93.5000 (0.7796)	96.0000 (0.6197)	95.7000 (0.6415)
0.25	GQ105	94.9000 (0.6957)	93.5000 (0.7796)	96.0000 (0.6197)	95.7000 (0.6415)
0.75	AF	94.6000 (0.7147)	94.5000 (0.7209)	94.7000 (0.7085)	94.7000 (0.7085)
0.75	IN	93.6198 (0.7729)	94.0104 (0.7504)	94.1406 (0.7427)	89.1927 (0.9818)
0.75	GQ15	95.0000 (0.6892)	95.4000 (0.6624)	94.6000 (0.7147)	99.4000 (0.2442)
0.75	GQ35	94.7000 (0.7085)	95.0000 (0.6892)	94.5000 (0.7209)	97.3000 (0.5126)
0.75	GQ75	94.5000 (0.7209)	94.9000 (0.6957)	94.6000 (0.7147)	95.7000 (0.6415)
0.75	GQ105	94.6000 (0.7147)	94.7000 (0.7085)	94.5000 (0.7209)	95.1000 (0.6826)
1.25	AF	94.4000 (0.7271)	95.0000 (0.6892)	95.1000 (0.6826)	94.3000 (0.7332)
1.25	IN	91.6064 (0.8769)	93.9219 (0.7556)	93.1983 (0.7962)	84.6599 (1.1396)
1.25	GQ15	81.7000 (1.2227)	61.2000 (1.5410)	92.5000 (0.8329)	0.0000 (0.0000)
1.25	GQ35	87.2000 (1.0565)	80.9000 (1.2431)	94.4000 (0.7271)	2.6000 (0.5032)
1.25	GQ75	91.0000 (0.9050)	90.1000 (0.9445)	95.2000 (0.6760)	49.6000 (1.5811)
1.25	GQ105	91.4000 (0.8866)	91.5000 (0.8819)	95.2000 (0.6760)	71.2000 (1.4320)

Table A.11: Percentage bias (%), comparison with analytical formulae, scenario with 1000 clusters of 2 individuals and a positive regression coefficient.

Frailty variance	Method	Lambda	P	Beta	Theta
0.25	AF	0.2294	-0.0426	-0.3674	2.4659
	IN	-0.0399	-0.0327	-0.5387	1.9523
	GQ15	-0.9919	-1.4226	0.7090	-5.2137
	GQ35	0.2175	-0.0801	-0.3289	2.1336
	GQ75	0.2298	-0.0421	-0.3678	2.4656
	GQ105	0.2297	-0.0423	-0.3677	2.4668
0.75	AF	-0.0404	-0.0798	0.1065	1.0954
	IN	0.1171	-0.0916	-0.0028	1.3386
	GQ15	-0.9319	-0.4739	0.4804	0.0287
	GQ35	-0.5823	-0.4846	0.4319	-3.8150
	GQ75	-0.3289	-0.3279	0.3001	-2.2747
	GQ105	-0.2506	-0.2665	0.2514	-1.4891
1.25	AF	0.0931	-0.1407	-0.0851	-0.4862
	IN	0.1267	-0.0883	0.1262	0.6778
	GQ15	8.2934	8.3381	-6.2492	-138.1108
	GQ35	6.2978	5.8510	-4.5265	-92.9554
	GQ75	4.6158	4.0956	-3.2494	-64.3743
	GQ105	3.9749	3.4630	-2.7833	-54.4579

Table A.12: Mean squared error, comparison with analytical formulae, scenario with 1000 clusters of 2 individuals and a positive regression coefficient.

Frailty variance	Method	Lambda	P	Beta	Theta
0.25	AF	0.0022	0.0007	0.0033	0.0503
	IN	0.0022	0.0007	0.0033	0.0451
	GQ15	0.0021	0.0006	0.0033	0.0077
	GQ35	0.0022	0.0007	0.0033	0.0452
	GQ75	0.0022	0.0007	0.0033	0.0502
	GQ105	0.0022	0.0007	0.0033	0.0503
0.75	AF	0.0032	0.0007	0.0048	0.0122
	IN	0.0032	0.0007	0.0049	0.0121
	GQ15	0.0031	0.0006	0.0048	0.0027
	GQ35	0.0032	0.0007	0.0048	0.0067
	GQ75	0.0032	0.0007	0.0048	0.0097
	GQ105	0.0032	0.0007	0.0048	0.0105
1.25	AF	0.0048	0.0008	0.0060	0.0088
	IN	0.0046	0.0008	0.0059	0.0085
	GQ15	0.0072	0.0024	0.0061	0.0961
	GQ35	0.0060	0.0015	0.0059	0.0450
	GQ75	0.0052	0.0011	0.0058	0.0235
	GQ105	0.0050	0.0010	0.0058	0.0180

Table A.13: Convergence rates, comparison with analytical formulae, scenario with 1000 clusters of 2 individuals and a positive regression coefficient.

Frailty variance	AF	IN	GQ15	GQ35	GQ75	GQ105
0.25	100.0%	77.8%	100.0%	99.8%	100.0%	100.0%
0.75	100.0%	76.8%	100.0%	100.0%	100.0%	100.0%
1.25	100.0%	69.1%	100.0%	100.0%	100.0%	100.0%

Table A.14: 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.15: 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.16: 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.17: 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.18: Bias, comparison without analytical formulae, scenario with a large frailty variance and a null regression coefficient.

Sample size	Method	Lambda	P	Beta	Sigma
1000c. of 2i.	GQ15	0.0622 (0.0015)	-0.1785 (0.0008)	0.0043 (0.0019)	-0.3291 (0.0032)
1000c. of 2i.	GQ35	0.0622 (0.0015)	-0.1785 (0.0008)	0.0043 (0.0019)	-0.3291 (0.0032)
1000c. of 2i.	GQ75	0.0622 (0.0015)	-0.1785 (0.0008)	0.0043 (0.0019)	-0.3291 (0.0032)
1000c. of 2i.	GQ105	0.0622 (0.0015)	-0.1785 (0.0008)	0.0043 (0.0019)	-0.3291 (0.0032)
15c. of 100i.	GQ15	0.0438 (0.0071)	-0.1252 (0.0016)	0.0506 (0.0064)	-0.1666 (0.0046)
15c. of 100i.	GQ35	0.0427 (0.0071)	-0.1230 (0.0016)	-0.0051 (0.0051)	-0.1106 (0.0043)
15c. of 100i.	GQ75	0.0431 (0.0071)	-0.1235 (0.0016)	-0.0299 (0.0041)	-0.1262 (0.0058)
15c. of 100i.	GQ105	0.0431 (0.0071)	-0.1238 (0.0016)	-0.0231 (0.0037)	-0.1356 (0.0060)
15c. of 30i.	GQ15	0.0241 (0.0077)	-0.1266 (0.0021)	-0.0382 (0.0072)	-0.1045 (0.0069)
15c. of 30i.	GQ35	0.0249 (0.0077)	-0.1282 (0.0021)	-0.0173 (0.0051)	-0.1319 (0.0072)
15c. of 30i.	GQ75	0.0252 (0.0077)	-0.1288 (0.0021)	0.0049 (0.0044)	-0.1442 (0.0068)
15c. of 30i.	GQ105	0.0252 (0.0077)	-0.1289 (0.0021)	0.0053 (0.0044)	-0.1457 (0.0067)
15c. of 500i.	GQ15	-0.0588 (0.0051)	-0.1209 (0.0012)	0.0878 (0.0049)	-0.2444 (0.0047)
15c. of 500i.	GQ35	-0.0544 (0.0051)	-0.1212 (0.0011)	0.0453 (0.0042)	-0.1481 (0.0036)
15c. of 500i.	GQ75	-0.0609 (0.0052)	-0.1179 (0.0011)	0.0199 (0.0032)	-0.1001 (0.0026)
15c. of 500i.	GQ105	-0.0438 (0.0051)	-0.1206 (0.0011)	0.0139 (0.0028)	-0.0982 (0.0023)
50c. of 100i.	GQ15	0.0436 (0.0038)	-0.1336 (0.0009)	0.0519 (0.0050)	-0.2172 (0.0032)
50c. of 100i.	GQ35	0.0423 (0.0038)	-0.1310 (0.0009)	0.0118 (0.0040)	-0.1119 (0.0027)
50c. of 100i.	GQ75	0.0422 (0.0038)	-0.1308 (0.0009)	-0.0157 (0.0028)	-0.0821 (0.0031)
50c. of 100i.	GQ105	0.0424 (0.0038)	-0.1310 (0.0009)	-0.0060 (0.0022)	-0.0830 (0.0033)
50c. of 30i.	GQ15	0.0349 (0.0041)	-0.1342 (0.0011)	0.0016 (0.0044)	-0.1006 (0.0036)
50c. of 30i.	GQ35	0.0349 (0.0041)	-0.1341 (0.0011)	0.0078 (0.0027)	-0.0891 (0.0037)
50c. of 30i.	GQ75	0.0350 (0.0041)	-0.1343 (0.0011)	0.0107 (0.0024)	-0.0939 (0.0035)
50c. of 30i.	GQ105	0.0350 (0.0041)	-0.1343 (0.0011)	0.0107 (0.0024)	-0.0940 (0.0035)

Table A.19: Coverage, comparison without analytical formulae, scenario with a large frailty variance and a null regression coefficient.

Sample size	Method	Lambda	P	Beta	Sigma
1000c. of 2i.	GQ15	64.4000 (1.5141)	0.0000 (0.0000)	95.8000 (0.6343)	6.9000 (0.8015)
1000c. of 2i.	GQ35	64.4000 (1.5141)	0.0000 (0.0000)	95.8000 (0.6343)	7.0000 (0.8068)
1000c. of 2i.	GQ75	64.4000 (1.5141)	0.0000 (0.0000)	95.8000 (0.6343)	7.0000 (0.8068)
1000c. of 2i.	GQ105	64.4000 (1.5141)	0.0000 (0.0000)	95.8000 (0.6343)	7.0000 (0.8068)
15c. of 100i.	GQ15	31.2000 (1.4651)	7.9000 (0.8530)	53.6000 (1.5770)	43.4000 (1.5673)
15c. of 100i.	GQ35	31.2000 (1.4651)	8.8000 (0.8959)	80.2000 (1.2601)	75.6000 (1.3582)
15c. of 100i.	GQ75	31.3313 (1.4668)	8.5085 (0.8823)	96.3964 (0.5894)	83.1832 (1.1827)
15c. of 100i.	GQ105	31.3000 (1.4664)	8.4000 (0.8772)	98.6000 (0.3715)	86.0000 (1.0973)
15c. of 30i.	GQ15	52.4000 (1.5793)	35.0000 (1.5083)	91.2000 (0.8959)	87.6000 (1.0422)
15c. of 30i.	GQ35	52.4000 (1.5793)	33.9000 (1.4969)	99.5000 (0.2230)	91.1000 (0.9004)
15c. of 30i.	GQ75	52.3000 (1.5795)	33.5000 (1.4926)	100.0000 (0.0000)	92.3000 (0.8430)
15c. of 30i.	GQ105	52.3000 (1.5795)	33.5000 (1.4926)	100.0000 (0.0000)	92.8000 (0.8174)
15c. of 500i.	GQ15	18.7919 (1.2353)	0.0000 (0.0000)	37.5839 (1.5316)	9.3960 (0.9227)
15c. of 500i.	GQ35	19.3548 (1.2493)	0.0000 (0.0000)	36.1290 (1.5191)	19.3548 (1.2493)
15c. of 500i.	GQ75	17.5325 (1.2024)	0.0000 (0.0000)	56.4935 (1.5677)	33.1169 (1.4883)
15c. of 500i.	GQ105	20.2454 (1.2707)	0.0000 (0.0000)	60.7362 (1.5443)	37.4233 (1.5303)
50c. of 100i.	GQ15	31.2312 (1.4655)	0.0000 (0.0000)	49.2492 (1.5810)	10.1101 (0.9533)
50c. of 100i.	GQ35	31.4000 (1.4677)	0.0000 (0.0000)	70.2000 (1.4464)	46.6000 (1.5775)
50c. of 100i.	GQ75	31.5000 (1.4689)	0.0000 (0.0000)	97.0000 (0.5394)	79.2000 (1.2835)
50c. of 100i.	GQ105	31.4314 (1.4681)	0.0000 (0.0000)	99.5996 (0.1997)	83.4835 (1.1742)
50c. of 30i.	GQ15	49.2000 (1.5809)	1.4000 (0.3715)	90.7000 (0.9184)	74.9000 (1.3711)
50c. of 30i.	GQ35	49.2000 (1.5809)	1.5000 (0.3844)	100.0000 (0.0000)	88.3000 (1.0164)
50c. of 30i.	GQ75	49.0000 (1.5808)	1.4000 (0.3715)	100.0000 (0.0000)	89.6000 (0.9653)
50c. of 30i.	GQ105	49.1000 (1.5809)	1.4000 (0.3715)	100.0000 (0.0000)	89.6000 (0.9653)

Table A.20: Mean squared error, comparison without analytical formulae, scenario with a large frailty variance and a null regression coefficient.

Sample size	Method	Lambda	P	Beta	Sigma
1000c. of 2i.	GQ15	0.0061	0.0325	0.0037	0.1187
1000c. of 2i.	GQ35	0.0061	0.0325	0.0037	0.1187
1000c. of 2i.	GQ75	0.0061	0.0325	0.0037	0.1187
1000c. of 2i.	GQ105	0.0061	0.0325	0.0037	0.1187
15c. of 100i.	GQ15	0.0524	0.0182	0.0434	0.0490
15c. of 100i.	GQ35	0.0524	0.0177	0.0259	0.0311
15c. of 100i.	GQ75	0.0524	0.0178	0.0179	0.0491
15c. of 100i.	GQ105	0.0523	0.0179	0.0139	0.0549
15c. of 30i.	GQ15	0.0595	0.0206	0.0527	0.0585
15c. of 30i.	GQ35	0.0594	0.0210	0.0265	0.0695
15c. of 30i.	GQ75	0.0594	0.0212	0.0194	0.0675
15c. of 30i.	GQ105	0.0594	0.0212	0.0194	0.0668
15c. of 500i.	GQ15	0.0299	0.0160	0.0319	0.0815
15c. of 500i.	GQ35	0.0291	0.0160	0.0193	0.0351
15c. of 500i.	GQ75	0.0310	0.0151	0.0104	0.0169
15c. of 500i.	GQ105	0.0282	0.0158	0.0080	0.0151
50c. of 100i.	GQ15	0.0163	0.0187	0.0273	0.0575
50c. of 100i.	GQ35	0.0162	0.0180	0.0162	0.0198
50c. of 100i.	GQ75	0.0162	0.0179	0.0083	0.0161
50c. of 100i.	GQ105	0.0162	0.0180	0.0048	0.0175
50c. of 30i.	GQ15	0.0178	0.0193	0.0193	0.0232
50c. of 30i.	GQ35	0.0178	0.0193	0.0075	0.0217
50c. of 30i.	GQ75	0.0178	0.0194	0.0060	0.0214
50c. of 30i.	GQ105	0.0178	0.0194	0.0060	0.0214

Table A.21: Convergence rates, comparison without analytical formulae, scenario with a large frailty variance and a null regression coefficient.

Sample size	GQ15	GQ35	GQ75	GQ105
1000c. of 2i.	100.0%	100.0%	100.0%	100.0%
15c. of 100i.	100.0%	100.0%	99.9%	100.0%
15c. of 30i.	100.0%	100.0%	100.0%	100.0%
15c. of 500i.	14.9%	15.5%	15.4%	16.3%
50c. of 100i.	99.9%	100.0%	100.0%	99.9%
50c. of 30i.	100.0%	100.0%	100.0%	100.0%

Table A.22: Bias, comparison without analytical formulae, scenario with 1000 clusters of 2 individuals and a positive regression coefficient.

fv	Method	Lambda	P	Beta	Sigma
0.25	GQ15	0.0156 (0.0013)	-0.0378 (0.0008)	-0.0128 (0.0017)	-0.2150 (0.0093)
0.25	GQ35	0.0156 (0.0013)	-0.0378 (0.0008)	-0.0128 (0.0017)	-0.2150 (0.0093)
0.25	GQ75	0.0156 (0.0013)	-0.0378 (0.0008)	-0.0128 (0.0017)	-0.2150 (0.0093)
0.25	GQ105	0.0156 (0.0013)	-0.0378 (0.0008)	-0.0127 (0.0017)	-0.2112 (0.0085)
0.75	GQ15	0.0429 (0.0014)	-0.1108 (0.0008)	-0.0431 (0.0019)	-0.2560 (0.0036)
0.75	GQ35	0.0429 (0.0014)	-0.1108 (0.0008)	-0.0431 (0.0019)	-0.2560 (0.0036)
0.75	GQ75	0.0429 (0.0014)	-0.1108 (0.0008)	-0.0431 (0.0019)	-0.2560 (0.0036)
0.75	GQ105	0.0429 (0.0014)	-0.1108 (0.0008)	-0.0431 (0.0019)	-0.2560 (0.0036)
1.25	GQ15	0.0652 (0.0015)	-0.1758 (0.0008)	-0.0749 (0.0019)	-0.3289 (0.0029)
1.25	GQ35	0.0652 (0.0015)	-0.1758 (0.0008)	-0.0749 (0.0019)	-0.3289 (0.0029)
1.25	GQ75	0.0652 (0.0015)	-0.1758 (0.0008)	-0.0749 (0.0019)	-0.3289 (0.0029)
1.25	GQ105	0.0652 (0.0015)	-0.1758 (0.0008)	-0.0749 (0.0019)	-0.3289 (0.0029)

Table A.23: Coverage, comparison without analytical formulae, scenario with 1000 clusters of 2 individuals and a positive regression coefficient.

Frailty variance	Method	Lambda	P	Beta	Sigma
0.25	GQ15	92.8929 (0.8125)	66.5666 (1.4918)	95.8959 (0.6273)	99.4995 (0.2232)
0.25	GQ35	92.8929 (0.8125)	66.5666 (1.4918)	95.8959 (0.6273)	99.4995 (0.2232)
0.25	GQ75	92.8929 (0.8125)	66.5666 (1.4918)	95.8959 (0.6273)	99.4995 (0.2232)
0.25	GQ105	92.8858 (0.8129)	66.6333 (1.4911)	95.8918 (0.6276)	99.4990 (0.2233)
0.75	GQ15	77.8000 (1.3142)	0.7000 (0.2636)	89.0000 (0.9894)	39.7000 (1.5472)
0.75	GQ35	77.8000 (1.3142)	0.7000 (0.2636)	89.0000 (0.9894)	39.7000 (1.5472)
0.75	GQ75	77.8000 (1.3142)	0.7000 (0.2636)	89.0000 (0.9894)	39.7000 (1.5472)
0.75	GQ105	77.8000 (1.3142)	0.7000 (0.2636)	89.0000 (0.9894)	39.7000 (1.5472)
1.25	GQ15	60.9000 (1.5431)	0.0000 (0.0000)	78.7000 (1.2947)	3.3000 (0.5649)
1.25	GQ35	60.9000 (1.5431)	0.0000 (0.0000)	78.7000 (1.2947)	3.3000 (0.5649)
1.25	GQ75	60.9000 (1.5431)	0.0000 (0.0000)	78.7000 (1.2947)	3.3000 (0.5649)
1.25	GQ105	60.9000 (1.5431)	0.0000 (0.0000)	78.7000 (1.2947)	3.3000 (0.5649)

Table A.24: Percentage bias (%), comparison without analytical formulae, scenario with 1000 clusters of 2 individuals and a positive regression coefficient.

Frailty variance	Method	Lambda	P	Beta	Sigma
0.25	GQ15	-2.2571	7.4087	-2.5529	31.0210
0.25	GQ35	-2.2571	7.4087	-2.5529	31.0219
0.25	GQ75	-2.2571	7.4087	-2.5529	31.0214
0.25	GQ105	-2.2498	7.4012	-2.5382	30.4701
0.75	GQ15	-6.1917	21.6854	-8.6277	177.9883
0.75	GQ35	-6.1917	21.6855	-8.6277	177.9884
0.75	GQ75	-6.1917	21.6855	-8.6277	177.9884
0.75	GQ105	-6.1917	21.6855	-8.6277	177.9884
1.25	GQ15	-9.4056	34.4137	-14.9841	-294.7880
1.25	GQ35	-9.4055	34.4134	-14.9838	-294.7815
1.25	GQ75	-9.4055	34.4134	-14.9838	-294.7815
1.25	GQ105	-9.4055	34.4134	-14.9838	-294.7815

Table A.25: Mean squared error, comparison without analytical formulae, scenario with 1000 clusters of 2 individuals and a positive regression coefficient.

Frailty variance	Method	Lambda	P	Beta	Sigma
0.25	GQ15	0.0019	0.0021	0.0029	0.1334
0.25	GQ35	0.0019	0.0021	0.0029	0.1335
0.25	GQ75	0.0019	0.0021	0.0029	0.1335
0.25	GQ105	0.0019	0.0021	0.0029	0.1173
0.75	GQ15	0.0039	0.0129	0.0054	0.0786
0.75	GQ35	0.0039	0.0129	0.0054	0.0786
0.75	GQ75	0.0039	0.0129	0.0054	0.0786
0.75	GQ105	0.0039	0.0129	0.0054	0.0786
1.25	GQ15	0.0065	0.0316	0.0094	0.1165
1.25	GQ35	0.0065	0.0316	0.0094	0.1165
1.25	GQ75	0.0065	0.0316	0.0094	0.1165
1.25	GQ105	0.0065	0.0316	0.0094	0.1165

Table A.26: Convergence rates, comparison without analytical formulae, scenario with 1000 clusters of 2 individuals and a positive regression coefficient.

Frailty variance	GQ15	GQ35	GQ75	GQ105
0.25	99.9%	99.9%	99.9%	99.8%
0.75	100.0%	100.0%	100.0%	100.0%
1.25	100.0%	100.0%	100.0%	100.0%

Table A.27: 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%

A.2 Simulation study 2

Table A.28: 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.29: 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.30: 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.31: 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.32: 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.33: 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.34: 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.35: 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

B.1 Simulation study 1

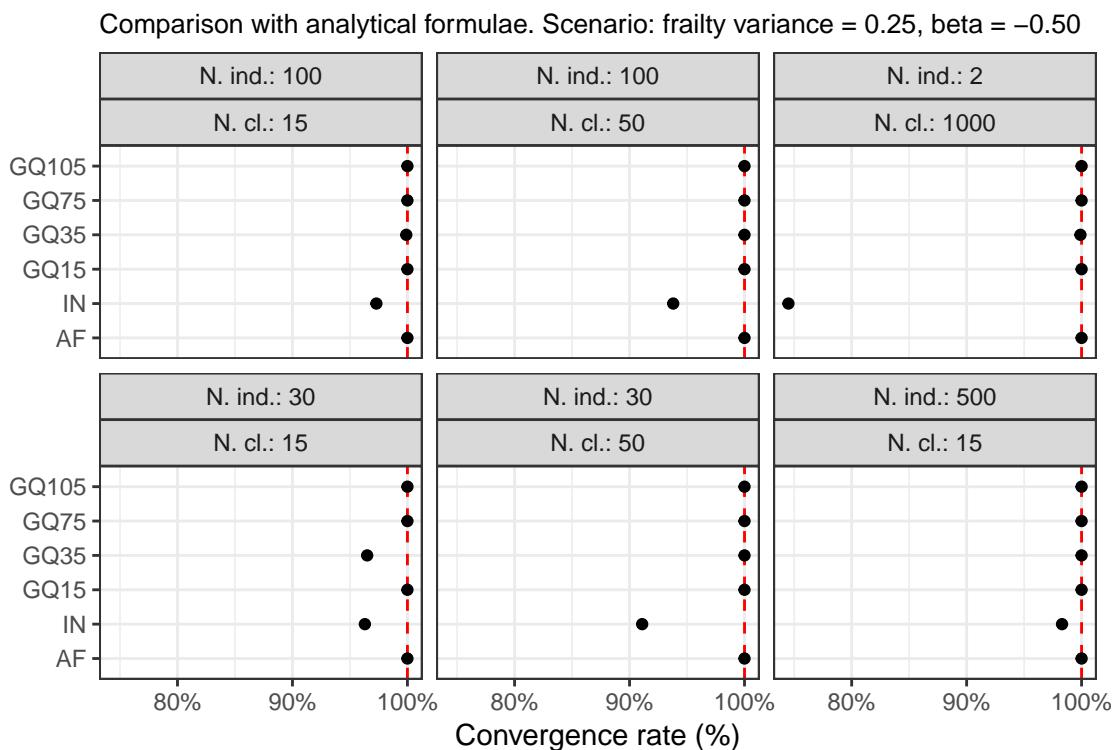


Figure B.1: Convergence rates, comparison with analytical formulae, scenario with a small frailty variance and a negative regression coefficient.

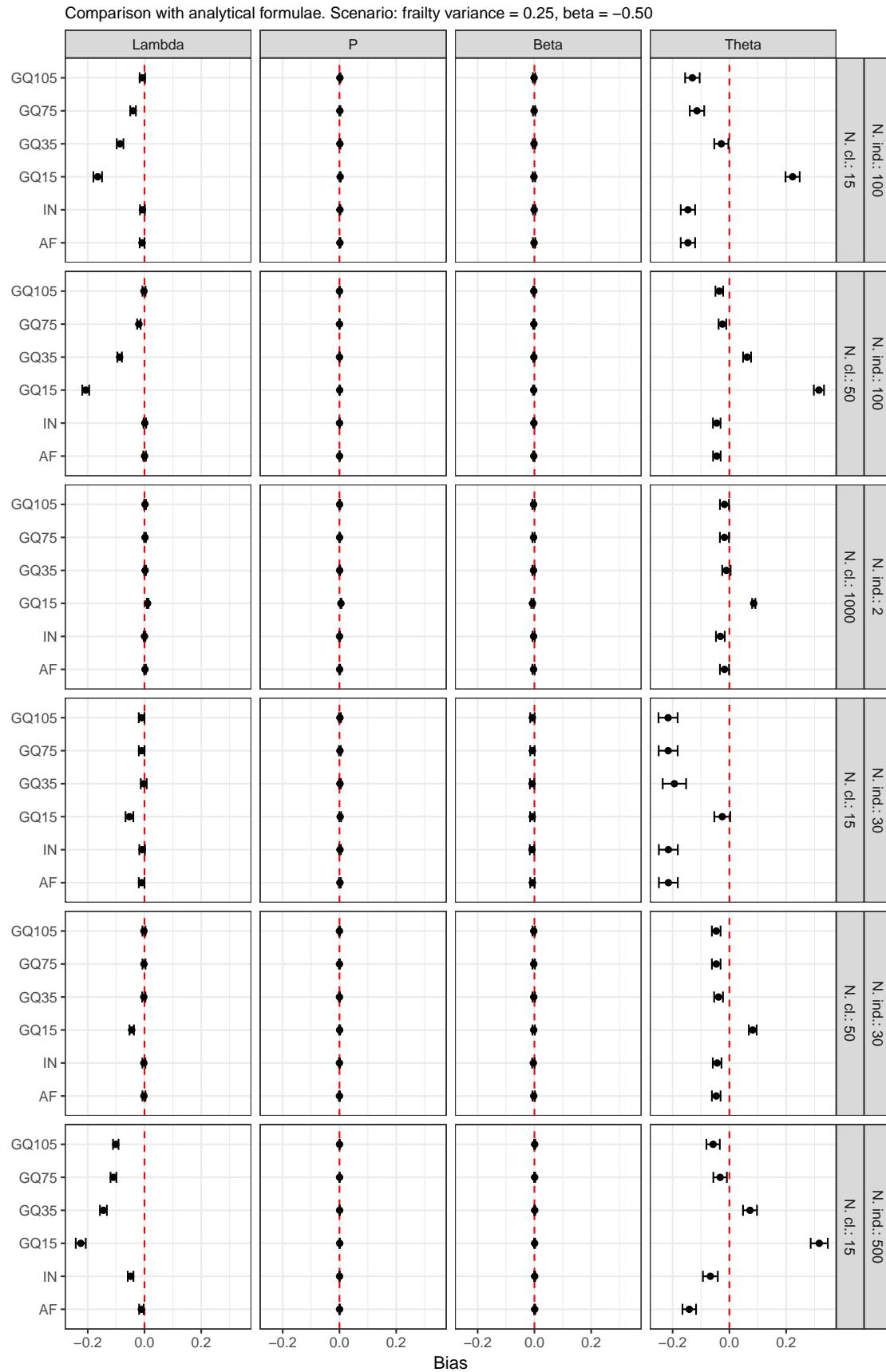


Figure B.2: Bias, comparison with analytical formulae, scenario with a small frailty variance and a negative regression coefficient.

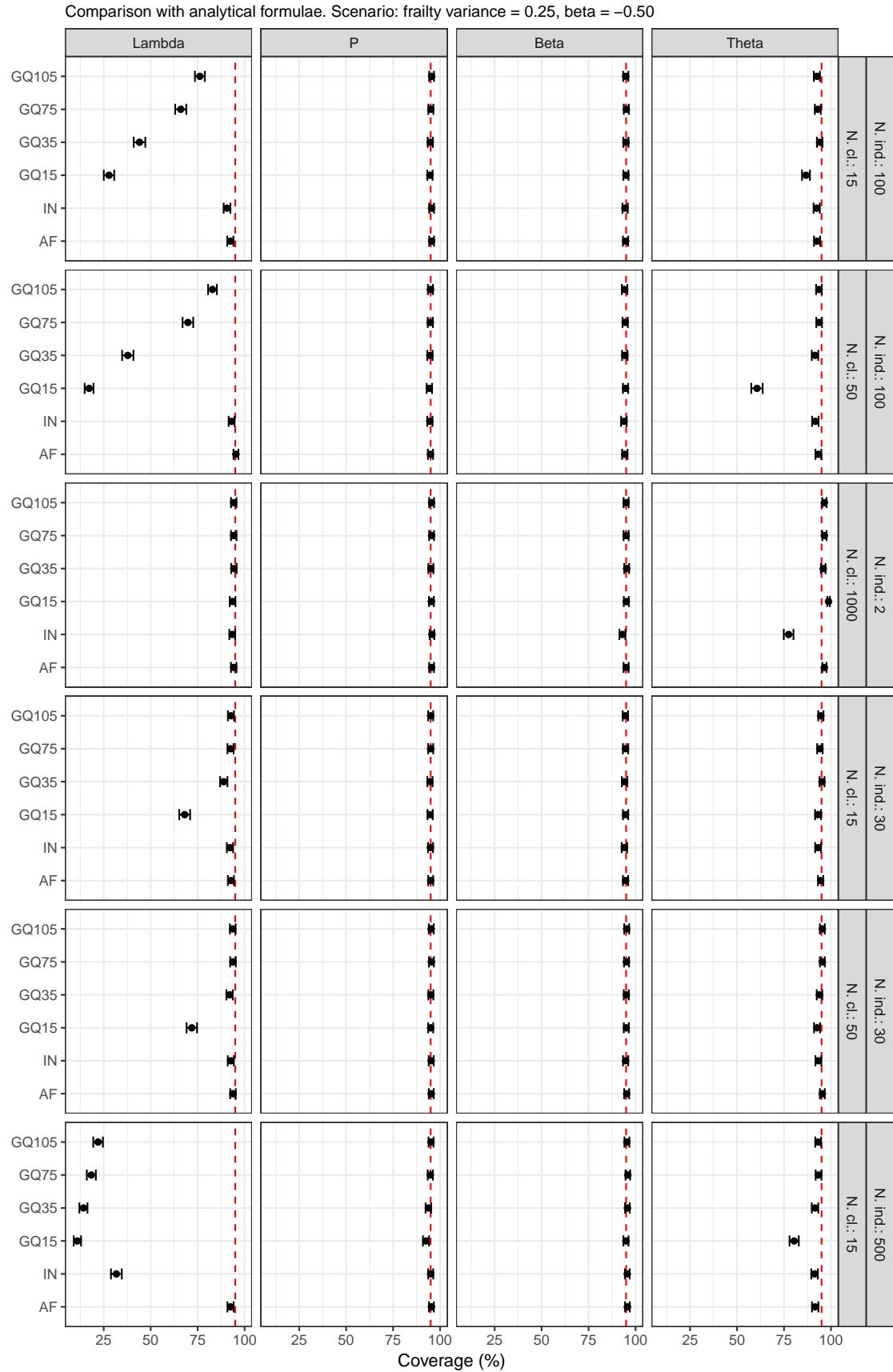


Figure B.3: Coverage, comparison with analytical formulae, scenario with a small frailty variance and a negative regression coefficient.

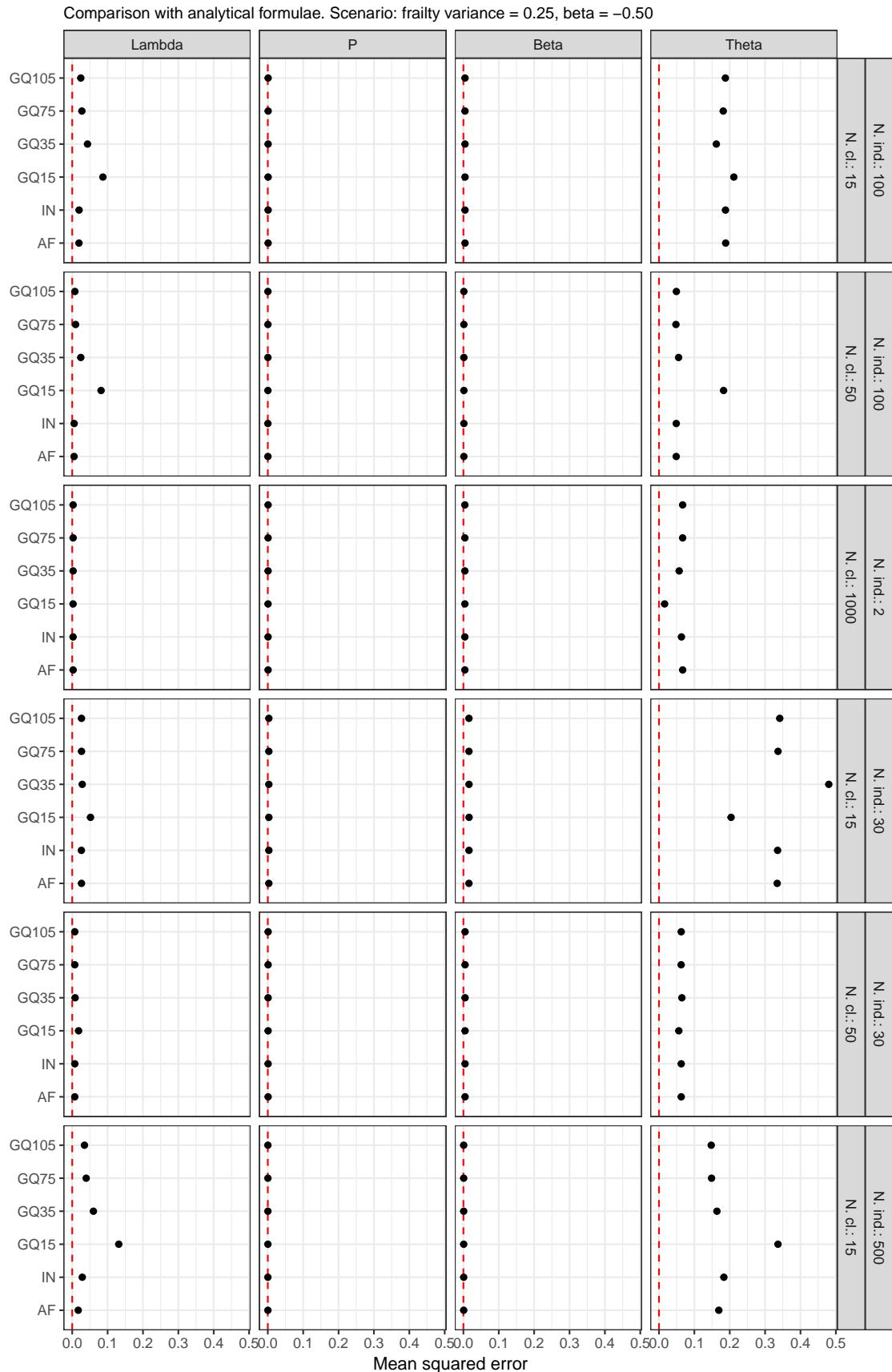


Figure B.4: Mean squared error, comparison with analytical formulae, scenario with a small frailty variance and a negative regression coefficient.

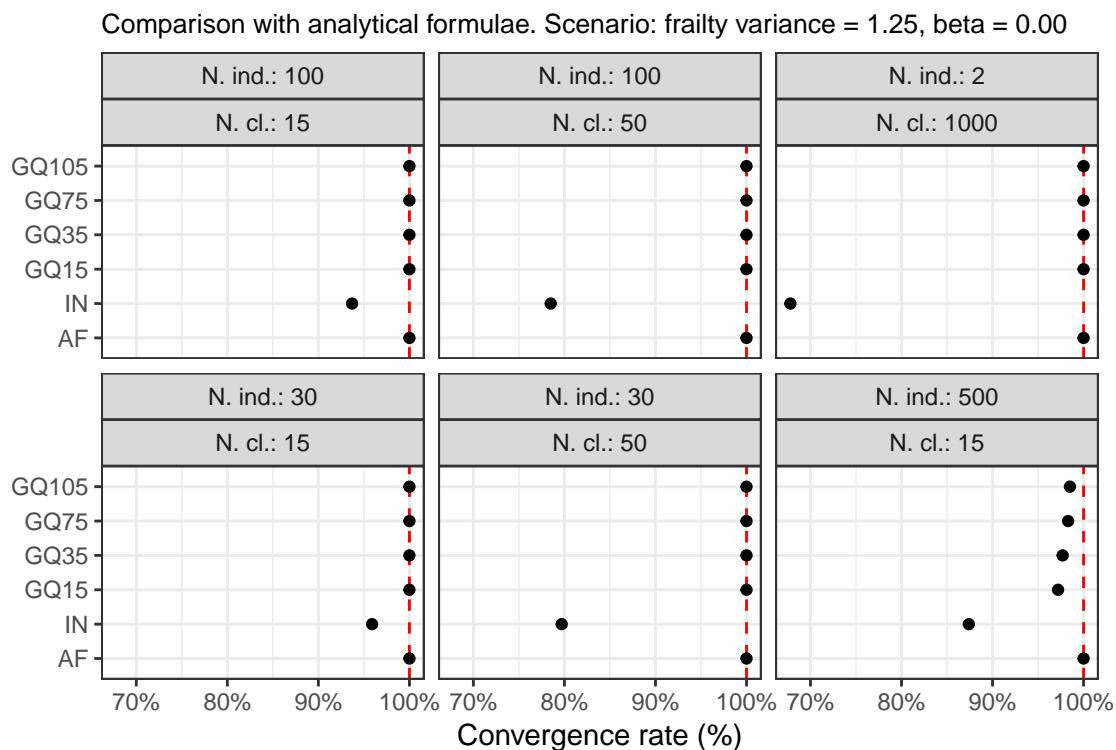


Figure B.5: Convergence rates, comparison with analytical formulae, scenario with a large frailty variance and a null regression coefficient.

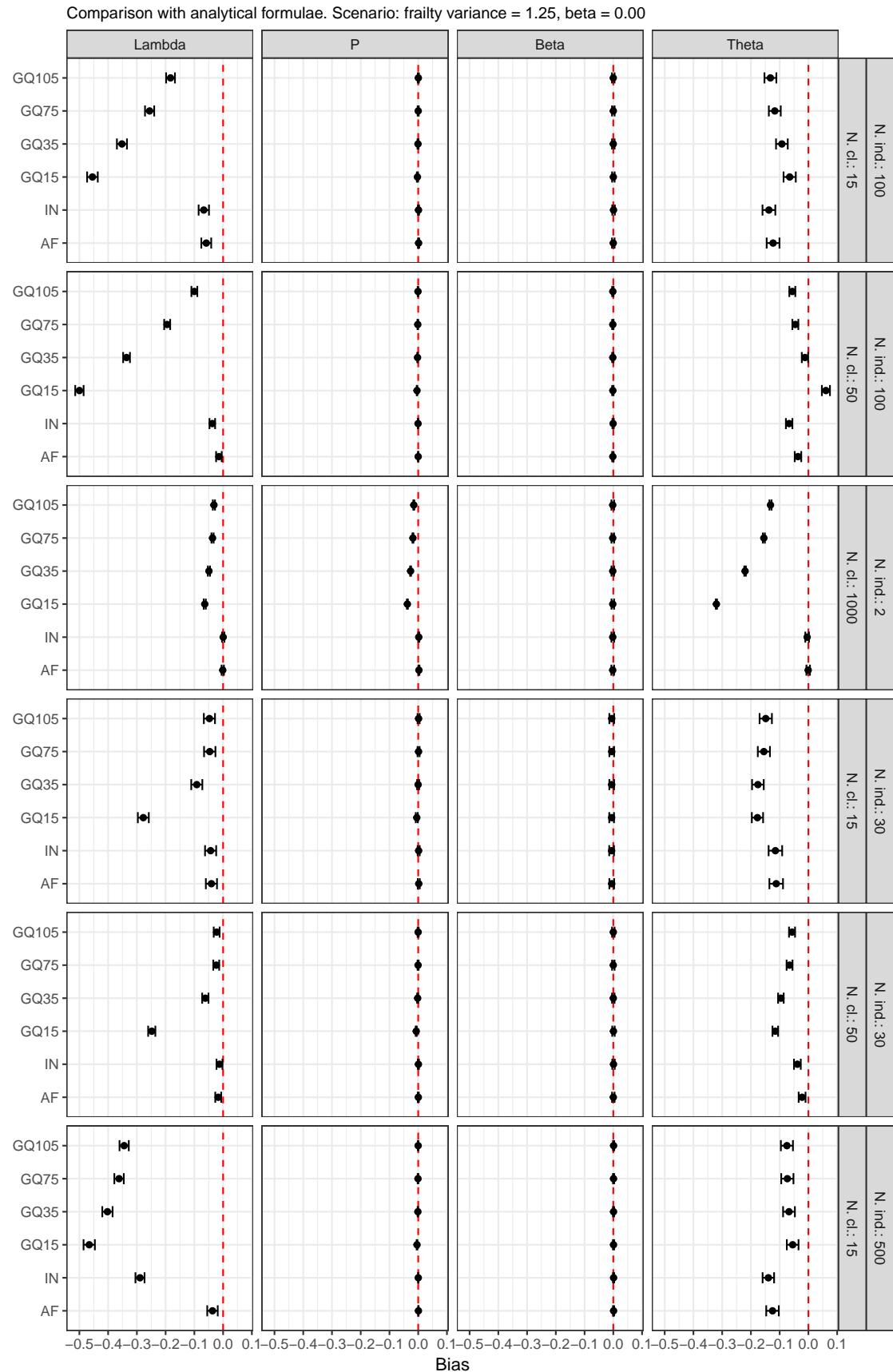


Figure B.6: Bias, comparison with analytical formulae, scenario with a large frailty variance and a null regression coefficient.

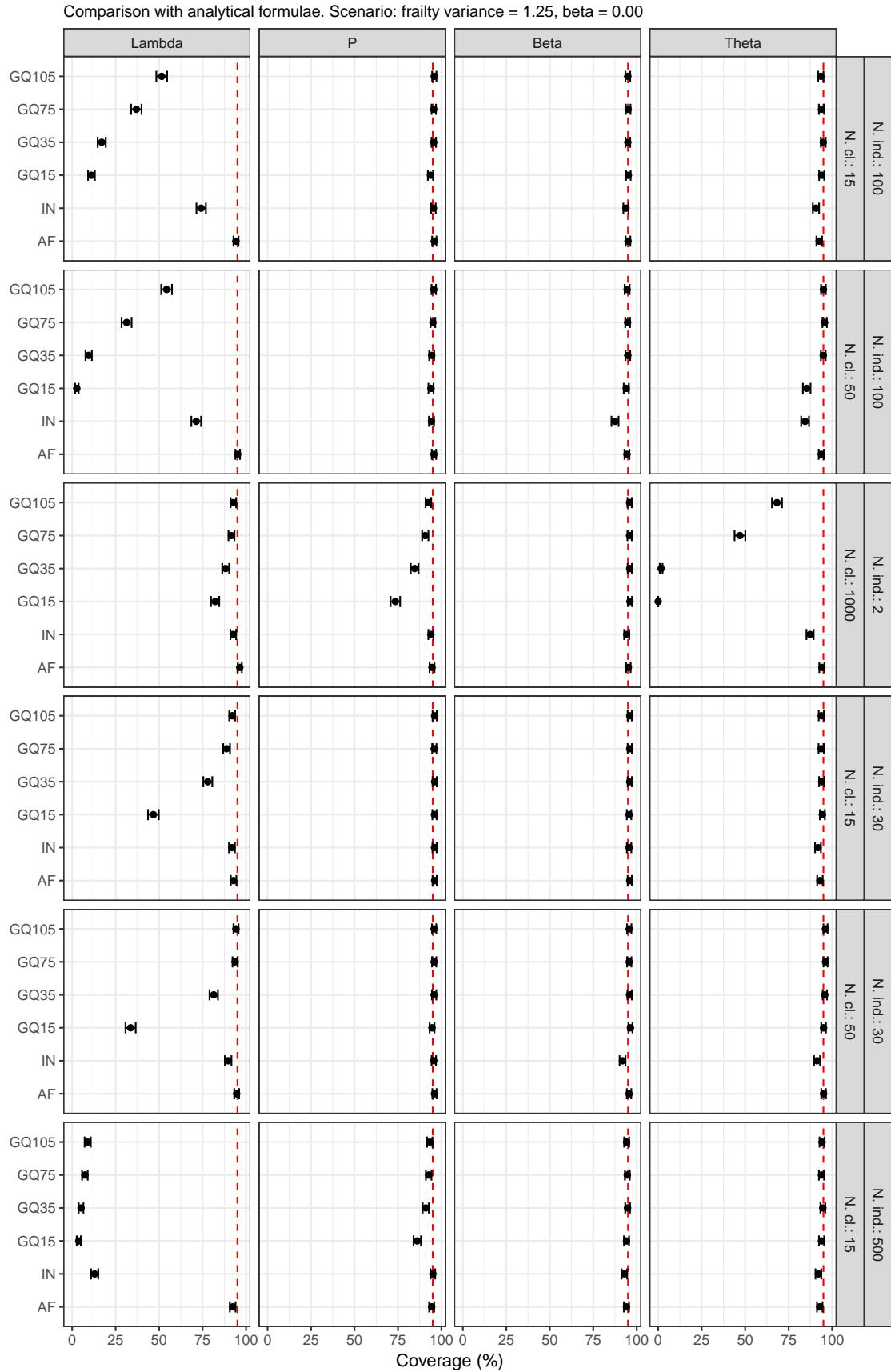


Figure B.7: Coverage, comparison with analytical formulae, scenario with a large frailty variance and a null regression coefficient.

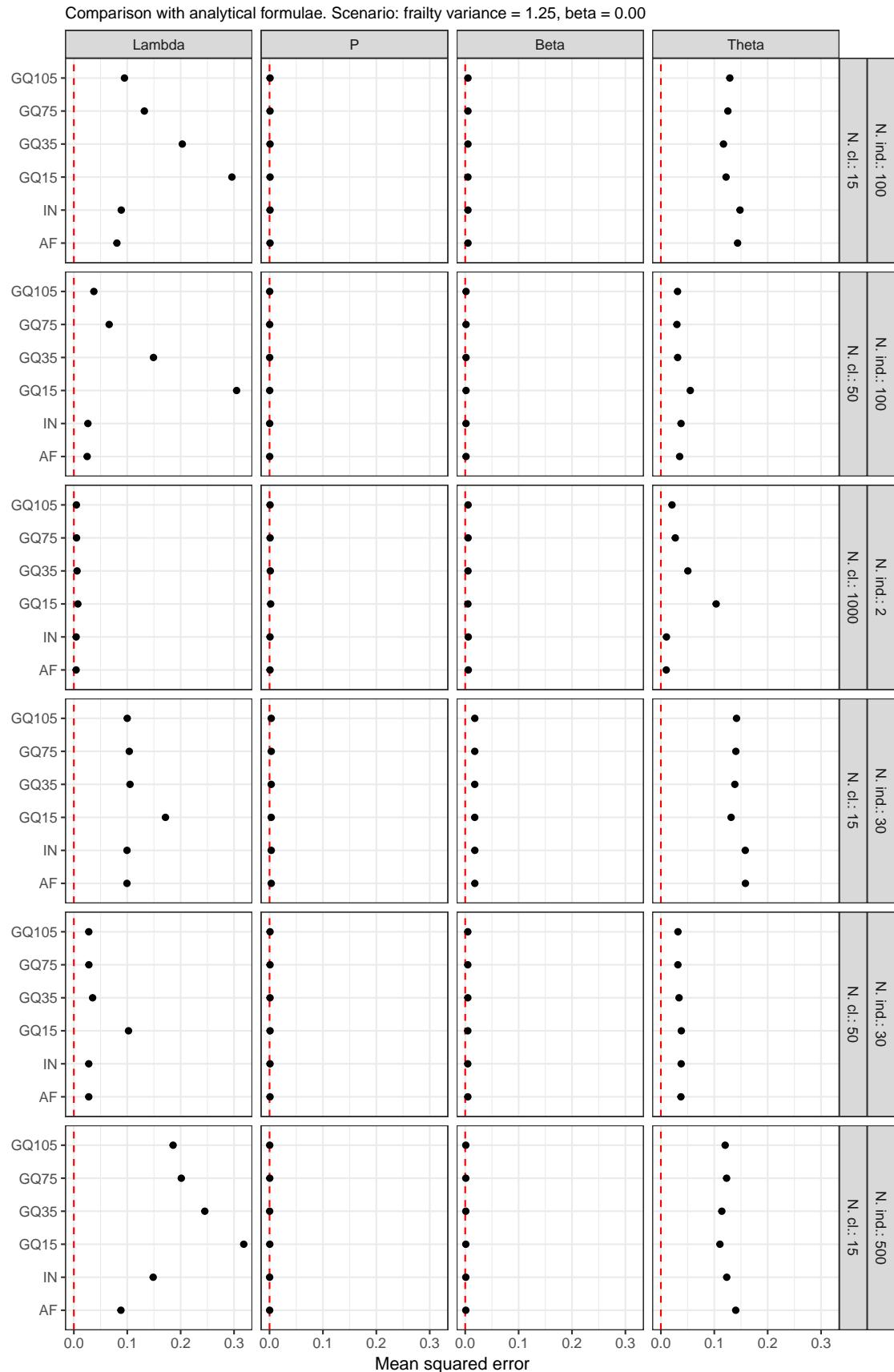


Figure B.8: Mean squared error, comparison with analytical formulae, scenario with a large frailty variance and a null regression coefficient.

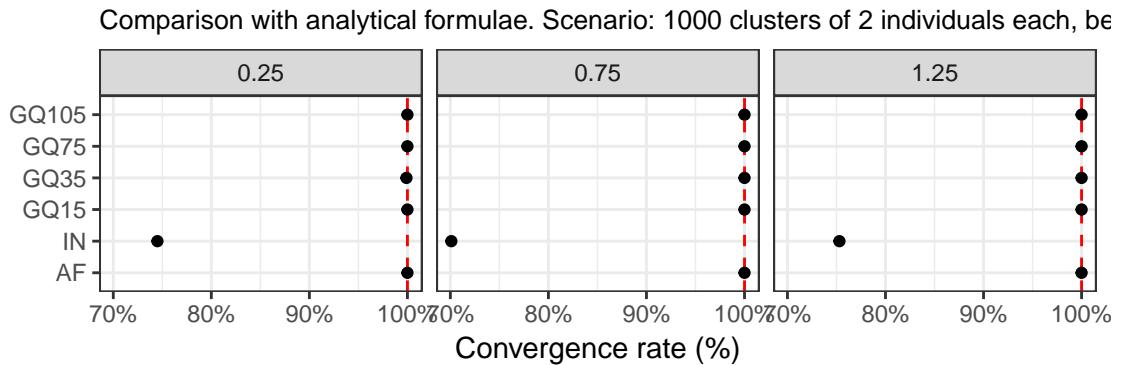


Figure B.9: Convergence rates, comparison with analytical formulae, scenario with 1000 clusters of 2 individuals each with a positive regression coefficient.

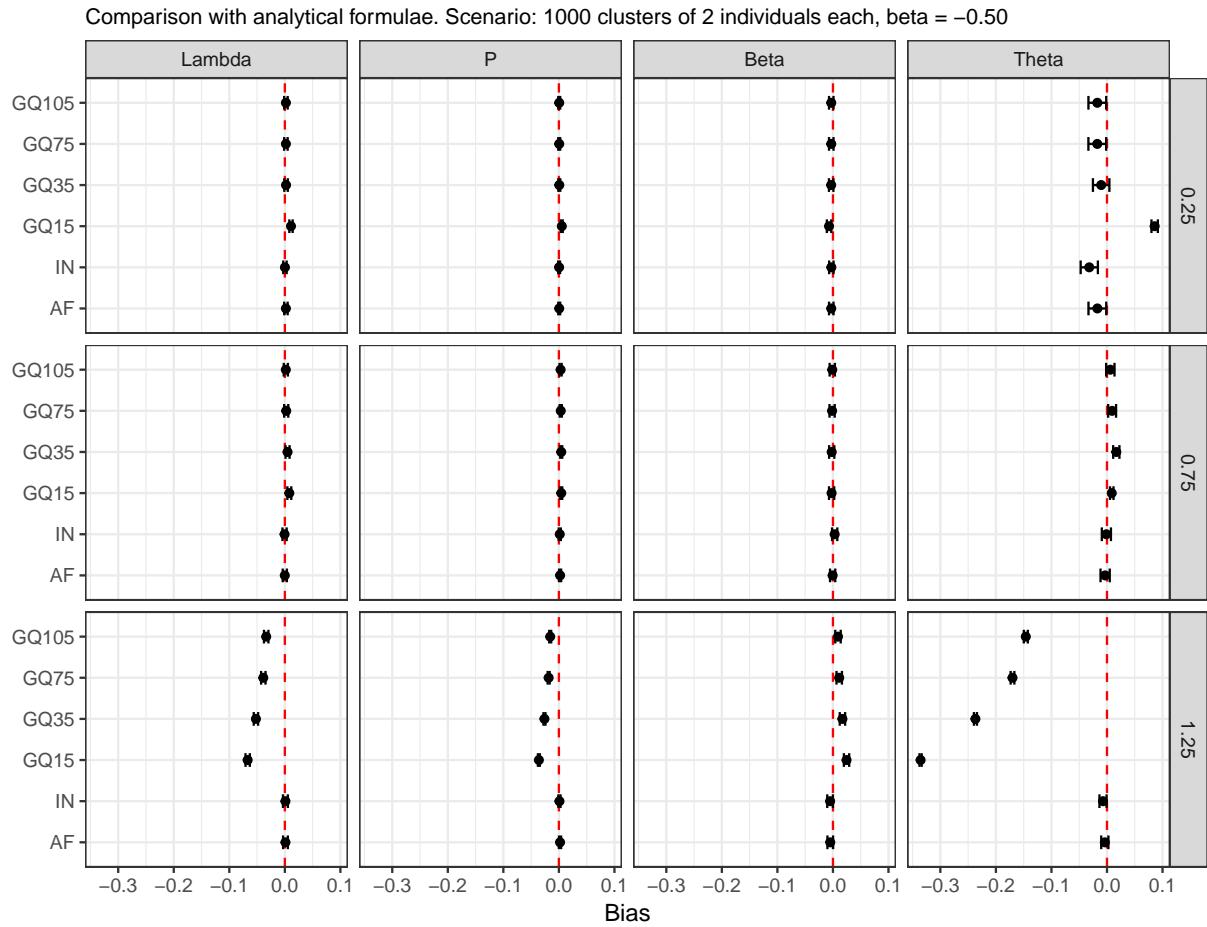


Figure B.10: Bias, comparison with analytical formulae, scenario with 1000 clusters of 2 individuals each with a positive regression coefficient.

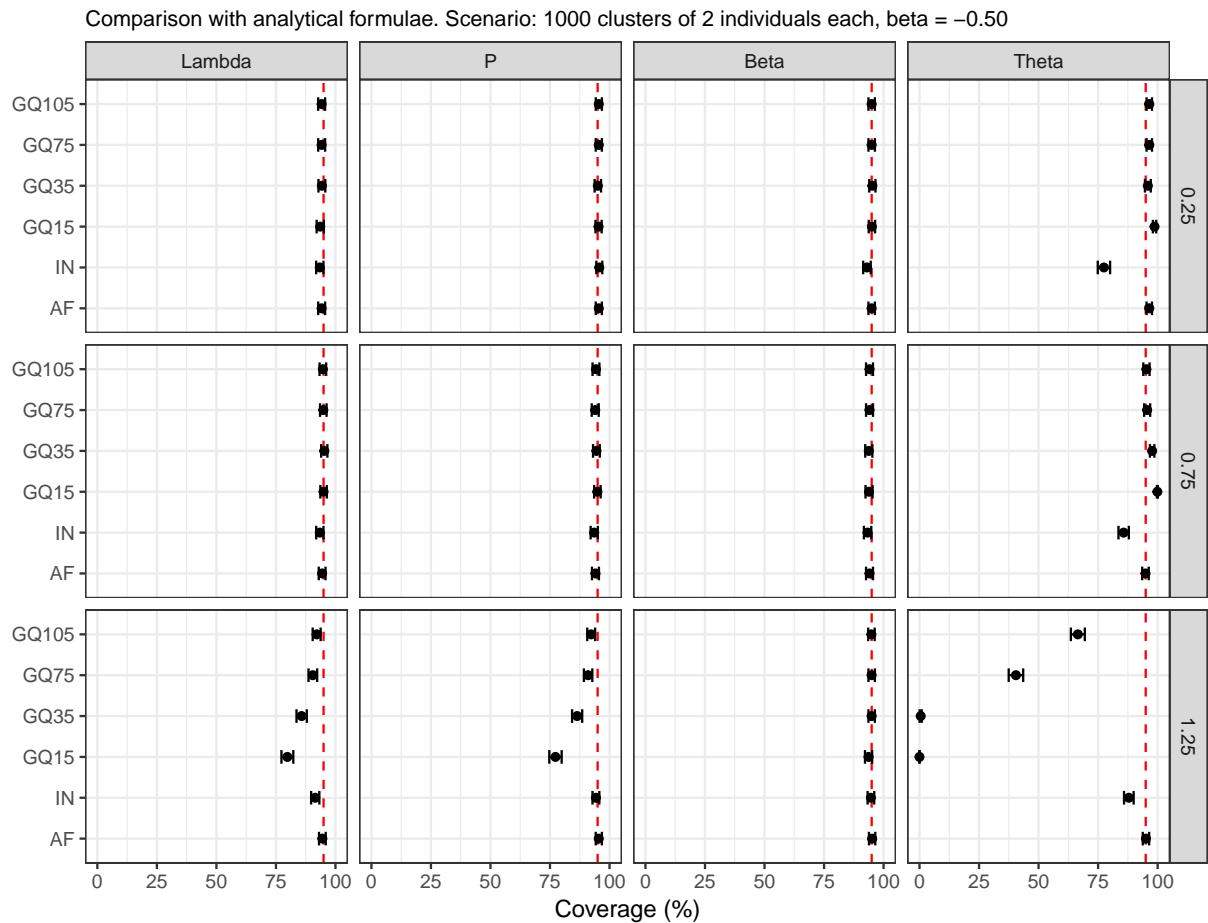


Figure B.11: Coverage, comparison with analytical formulae, scenario with 1000 clusters of 2 individuals each with a positive regression coefficient.

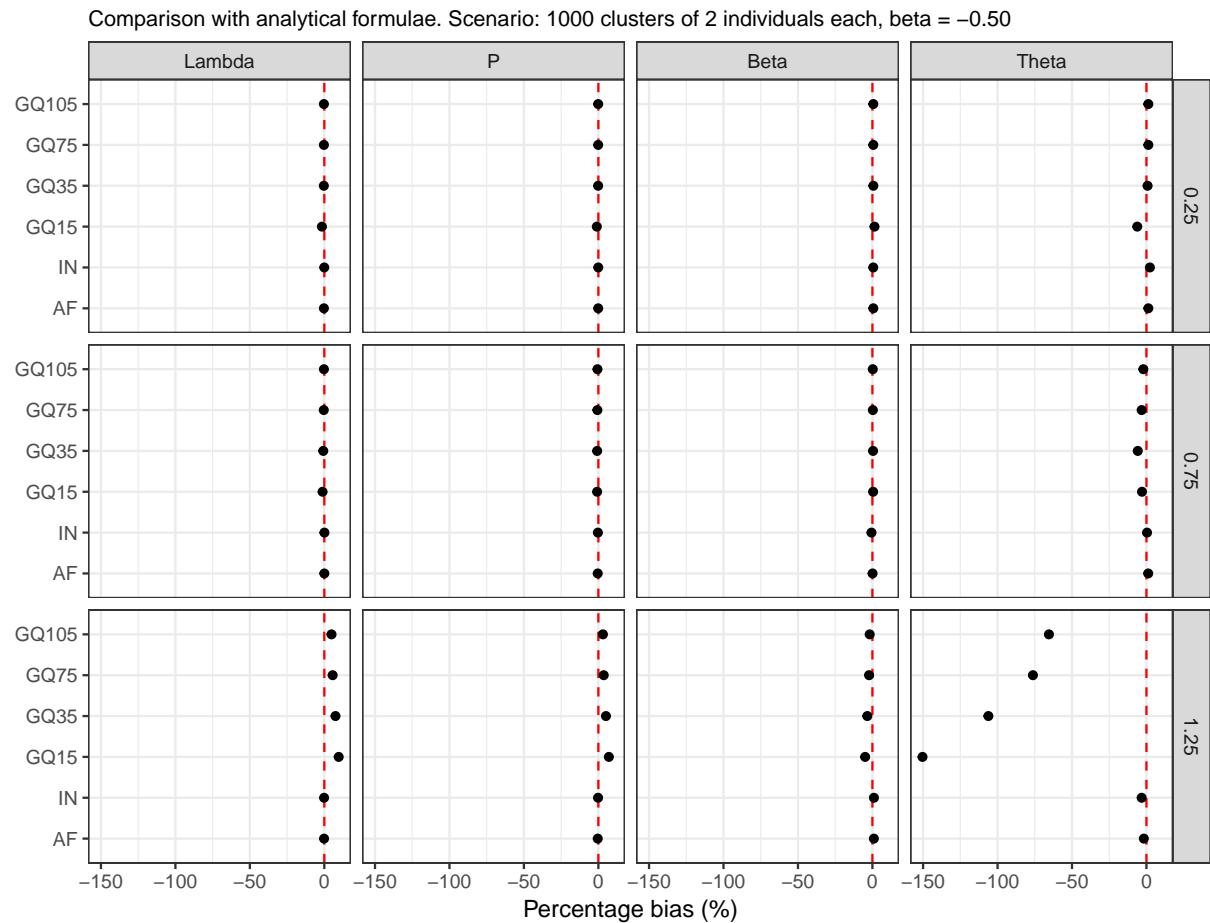


Figure B.12: Percentage bias, comparison with analytical formulae, scenario with 1000 clusters of 2 individuals each with a positive regression coefficient.

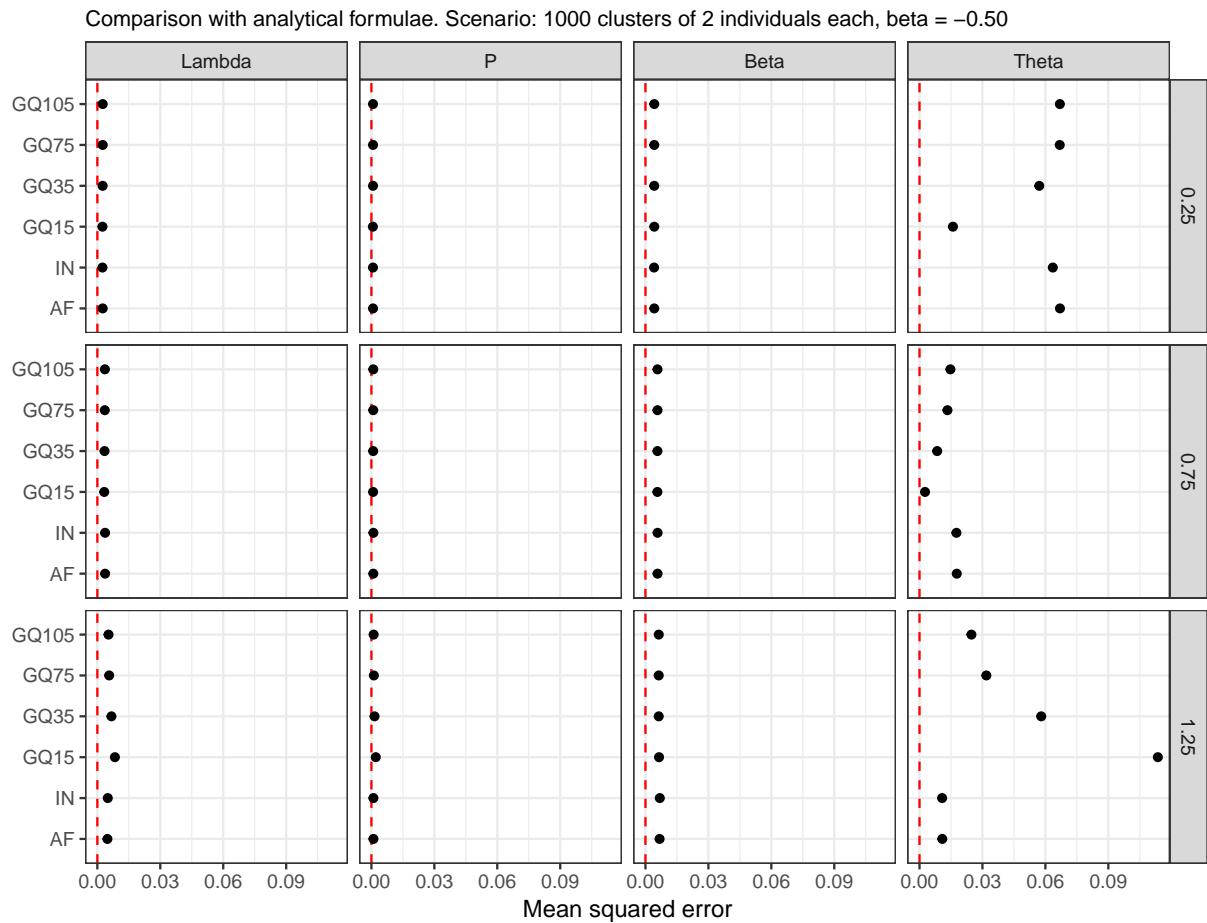


Figure B.13: Mean squared error, comparison with analytical formulae, scenario with 1000 clusters of 2 individuals each with a positive regression coefficient.

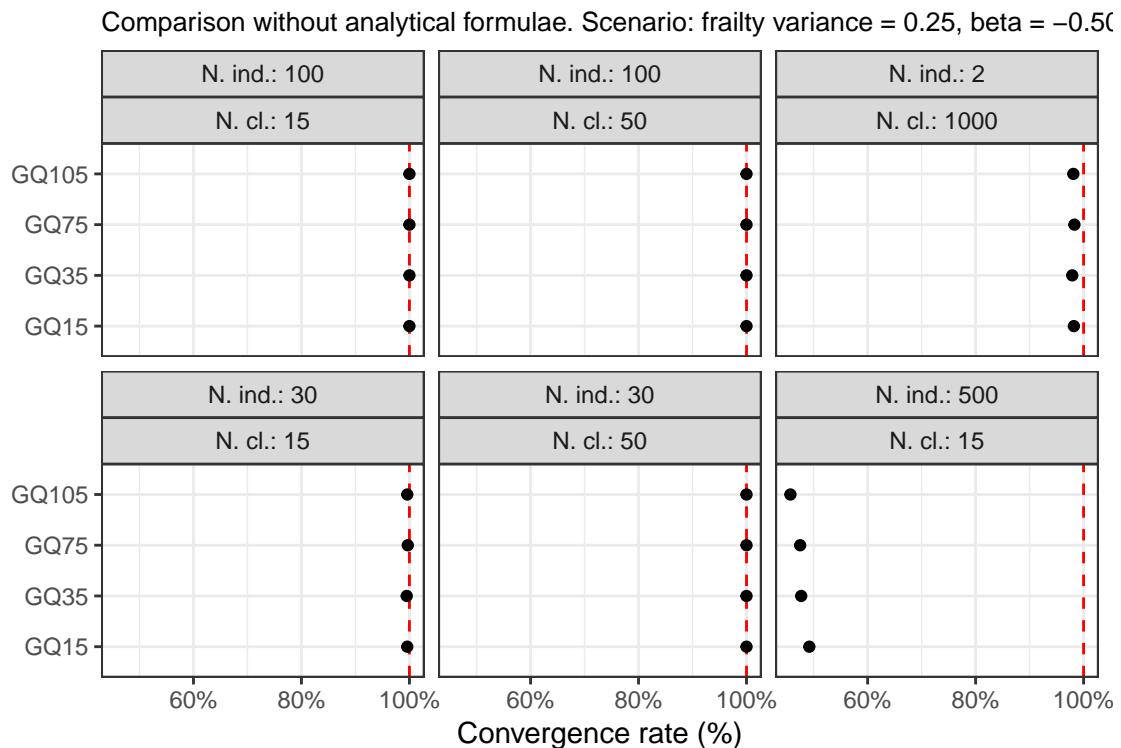


Figure B.14: Convergence rates, comparison without analytical formulae, scenario with a small frailty variance and a negative regression coefficient.

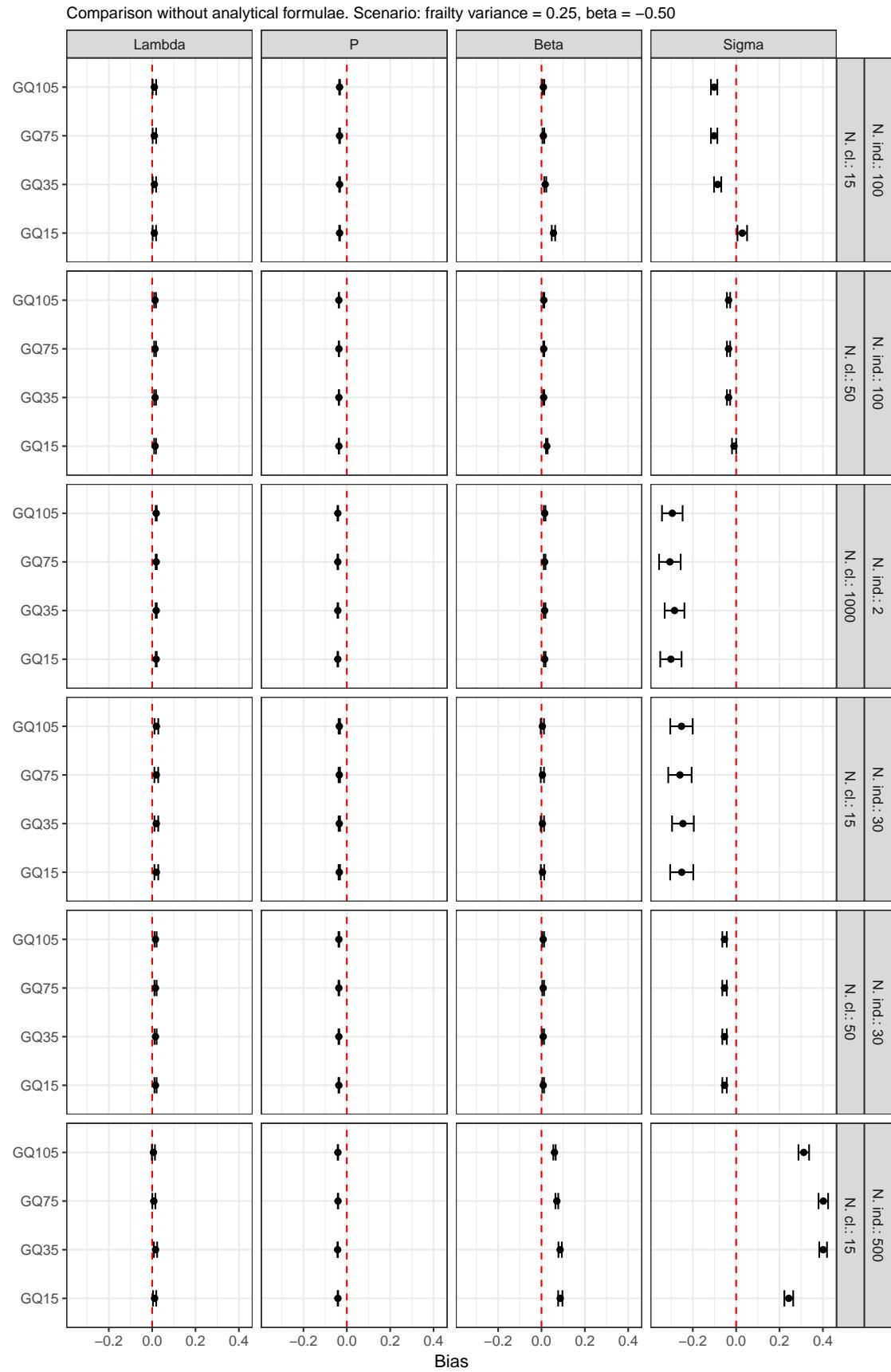


Figure B.15: Bias, comparison without analytical formulae, scenario with a small frailty variance and a negative regression coefficient.

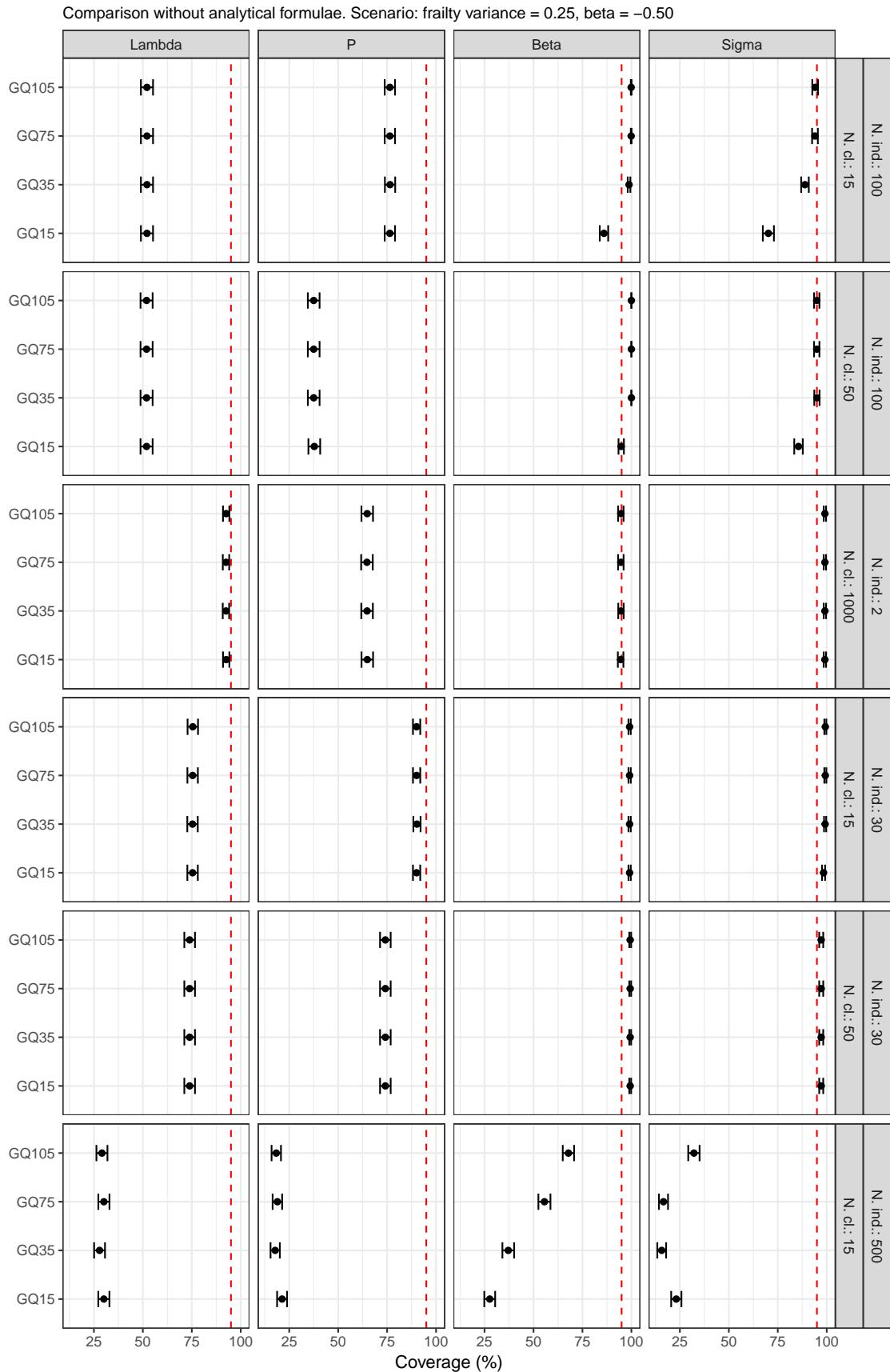


Figure B.16: Coverage, comparison without analytical formulae, scenario with a small frailty variance and a negative regression coefficient.

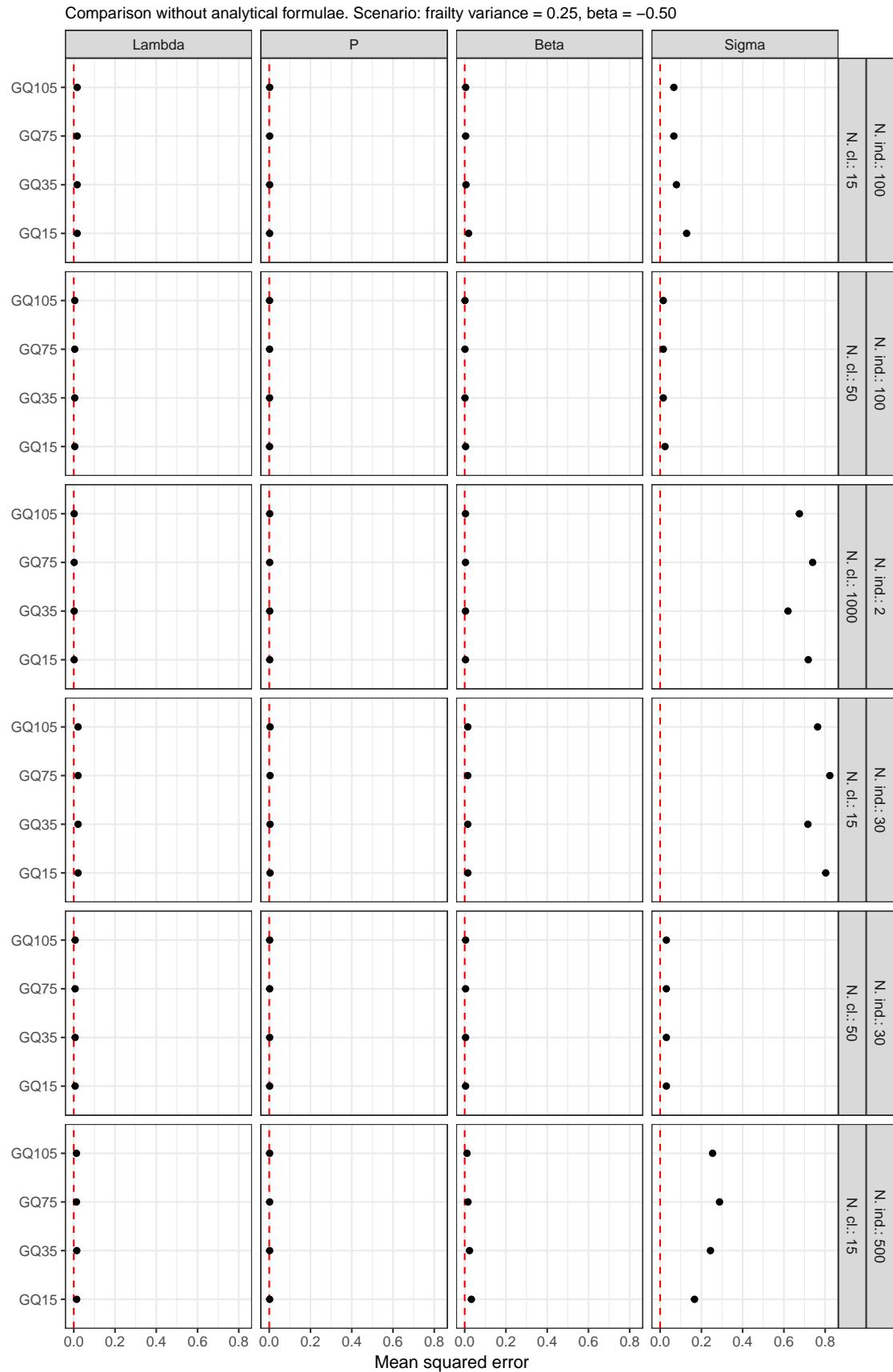


Figure B.17: Mean squared error, comparison without analytical formulae, scenario with a small frailty variance and a negative regression coefficient.

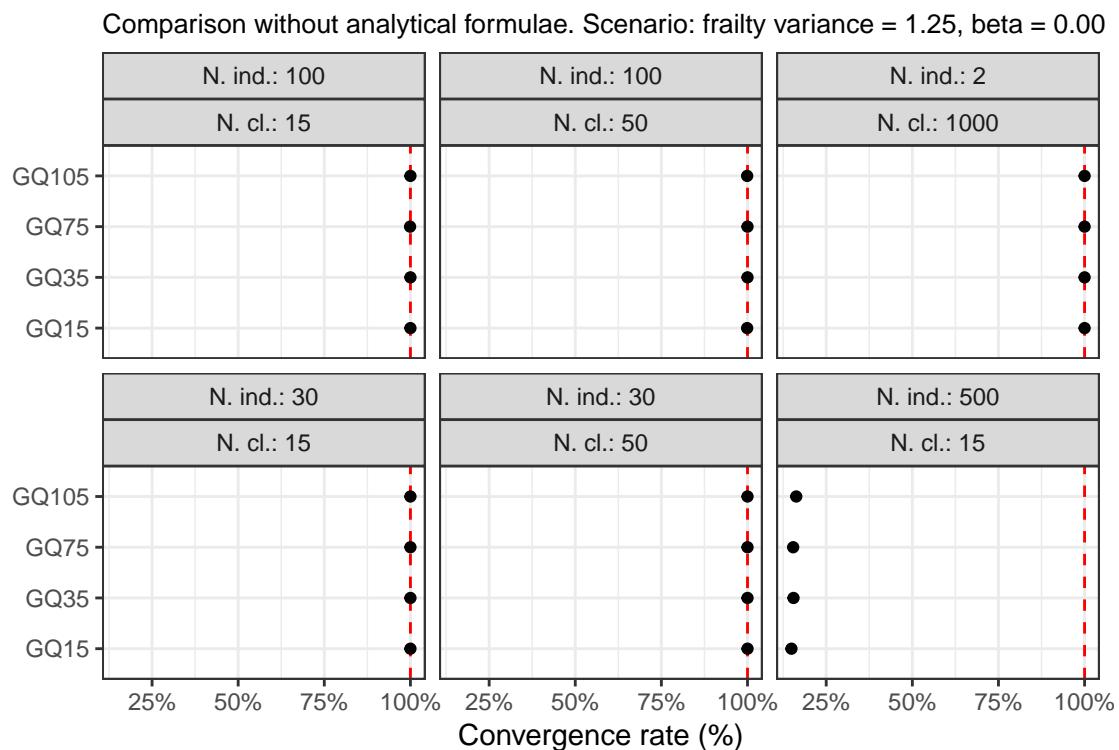


Figure B.18: Convergence rates, comparison without analytical formulae, scenario with a large frailty variance and a null regression coefficient.

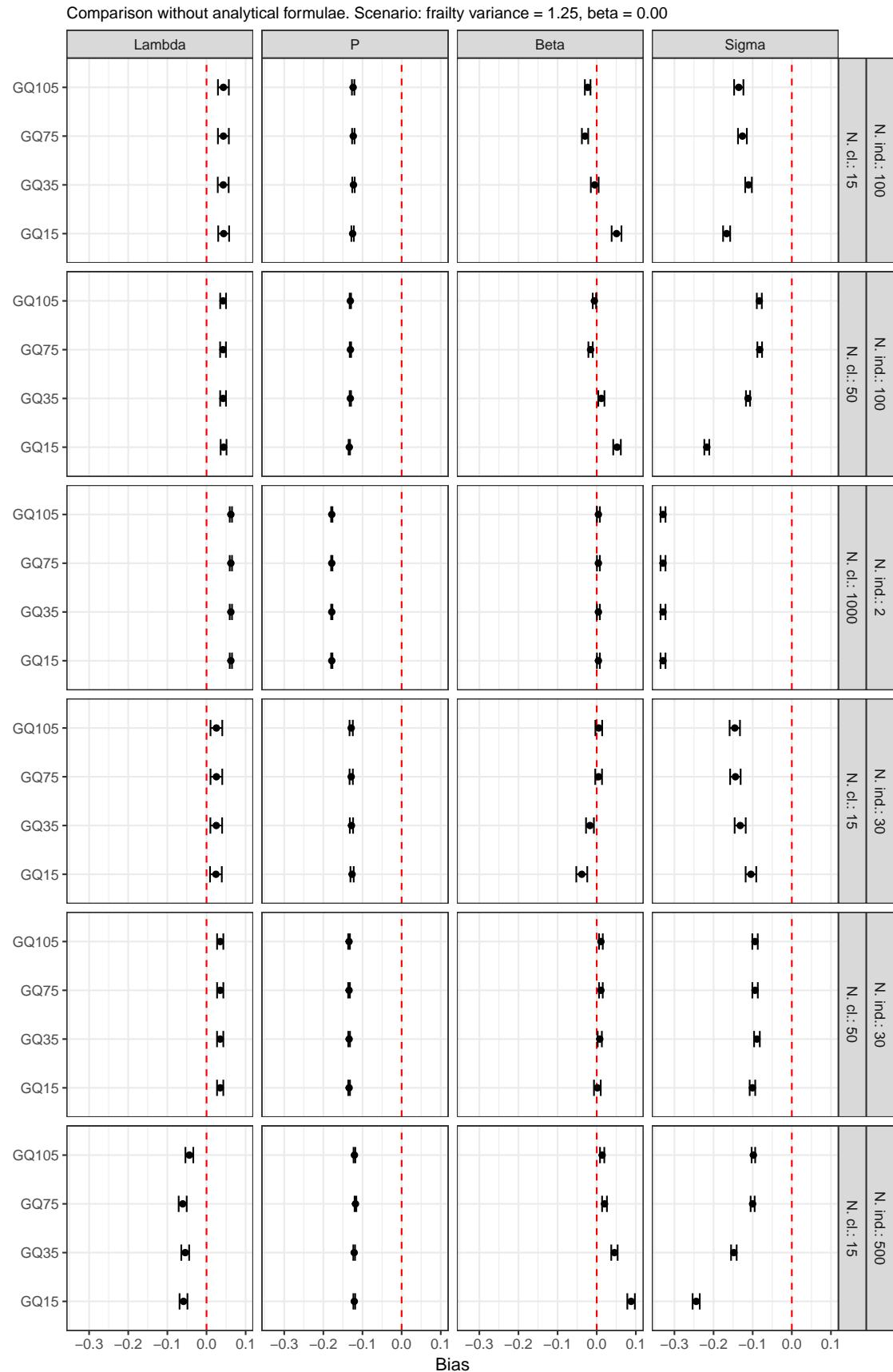


Figure B.19: Bias, comparison without analytical formulae, scenario with a large frailty variance and a null regression coefficient.

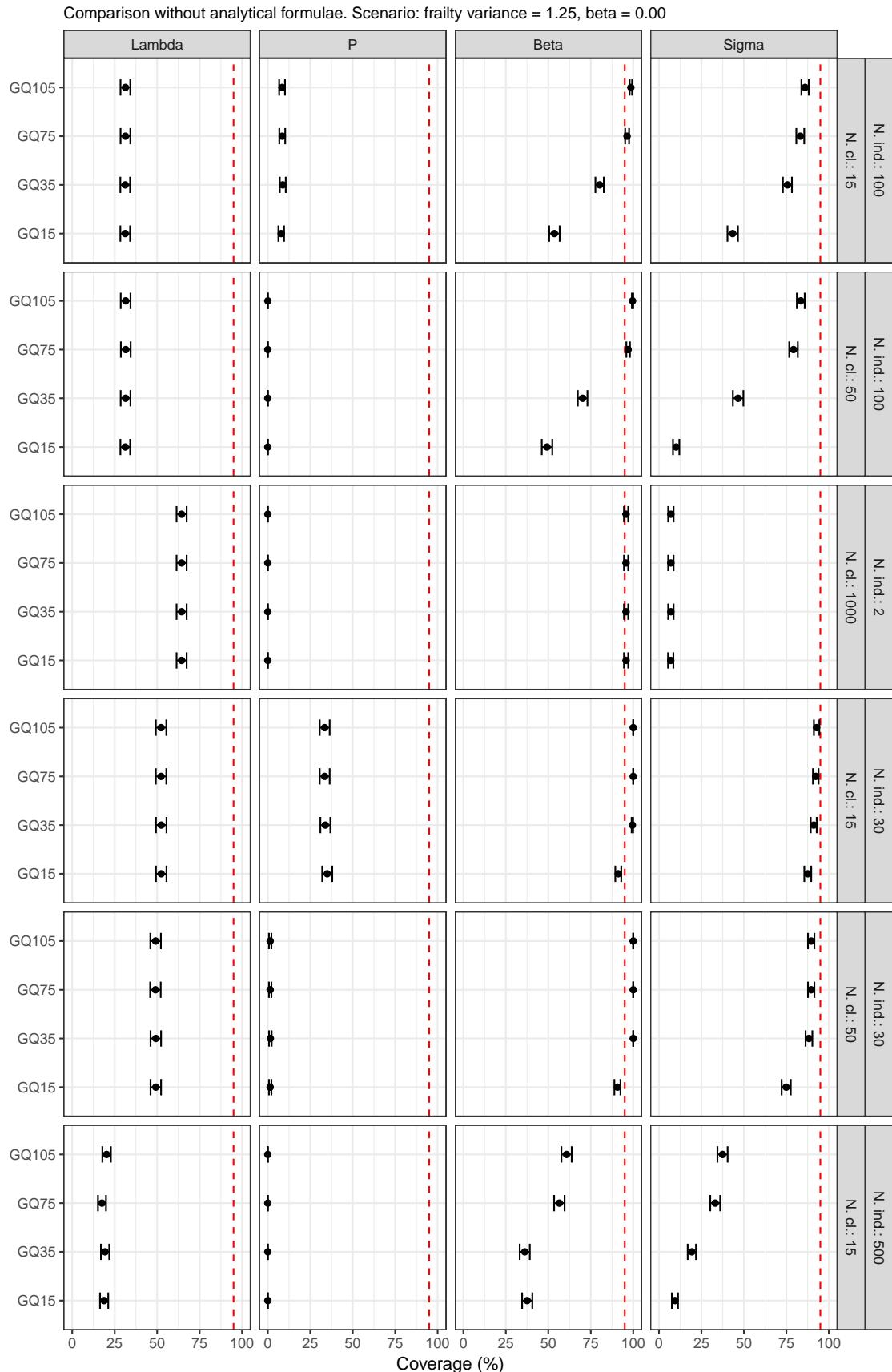


Figure B.20: Coverage, comparison without analytical formulae, scenario with a large frailty variance and a null regression coefficient.

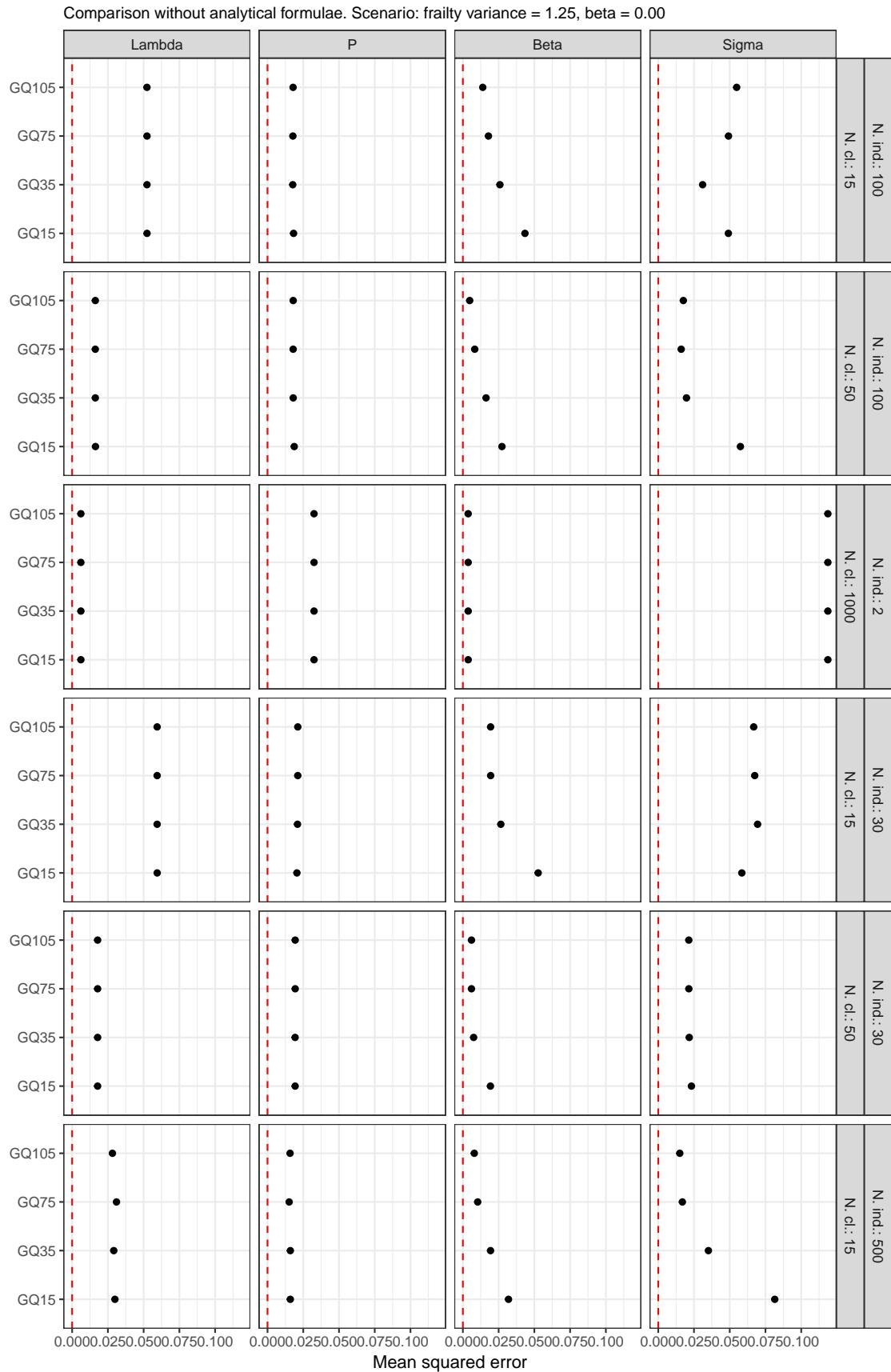


Figure B.21: Mean squared error, comparison without analytical formulae, scenario with a large frailty variance and a null regression coefficient.

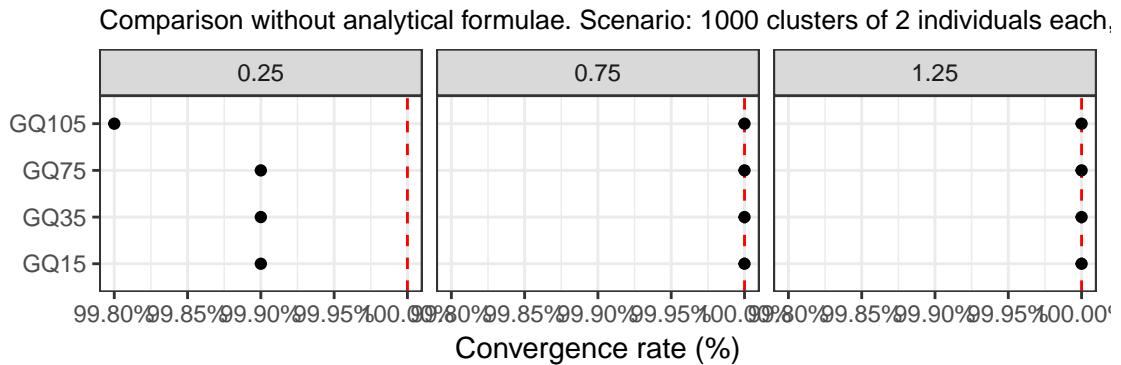


Figure B.22: Convergence rates, comparison without analytical formulae, scenario with 1000 clusters of 2 individuals each with a positive regression coefficient.

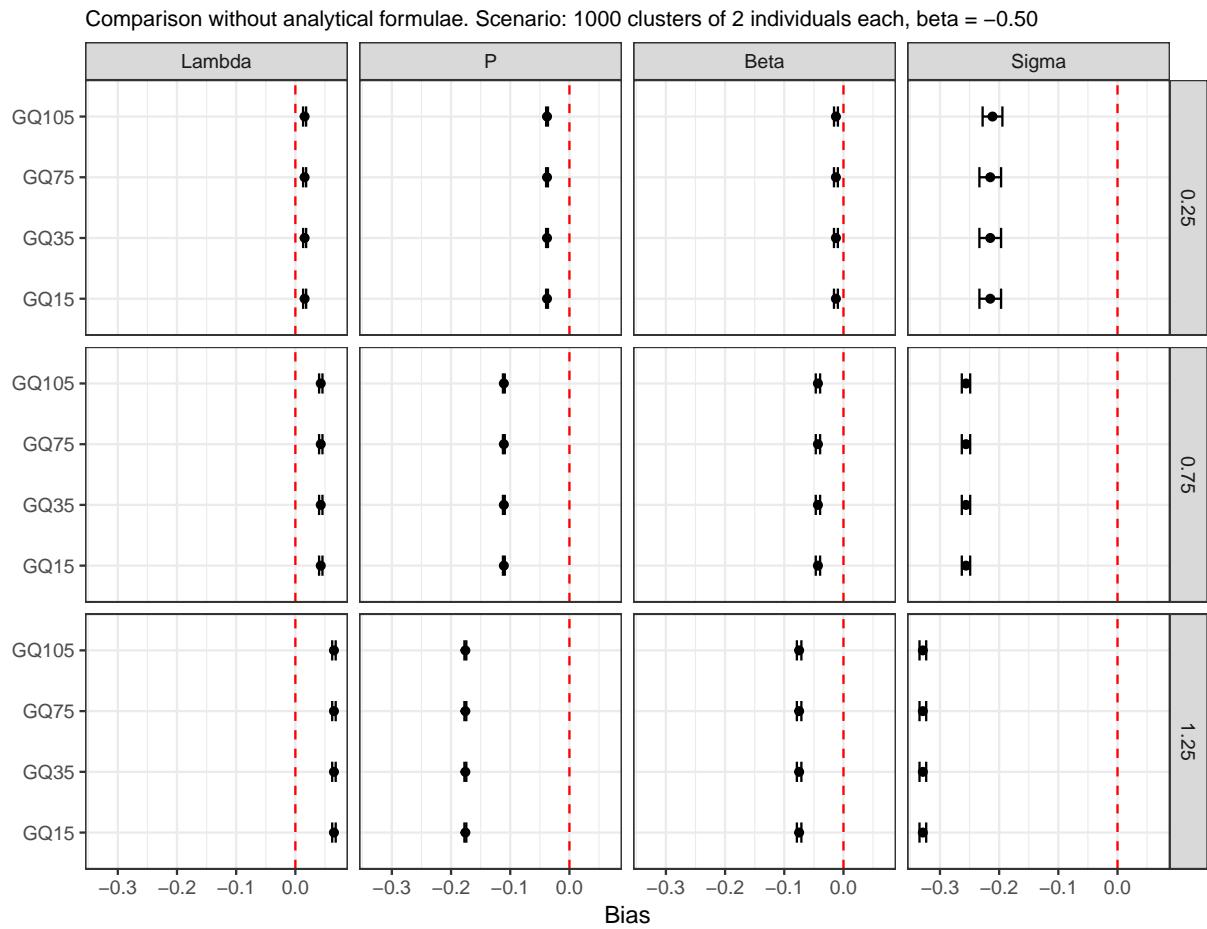


Figure B.23: Bias, comparison without analytical formulae, scenario with 1000 clusters of 2 individuals each with a positive regression coefficient.

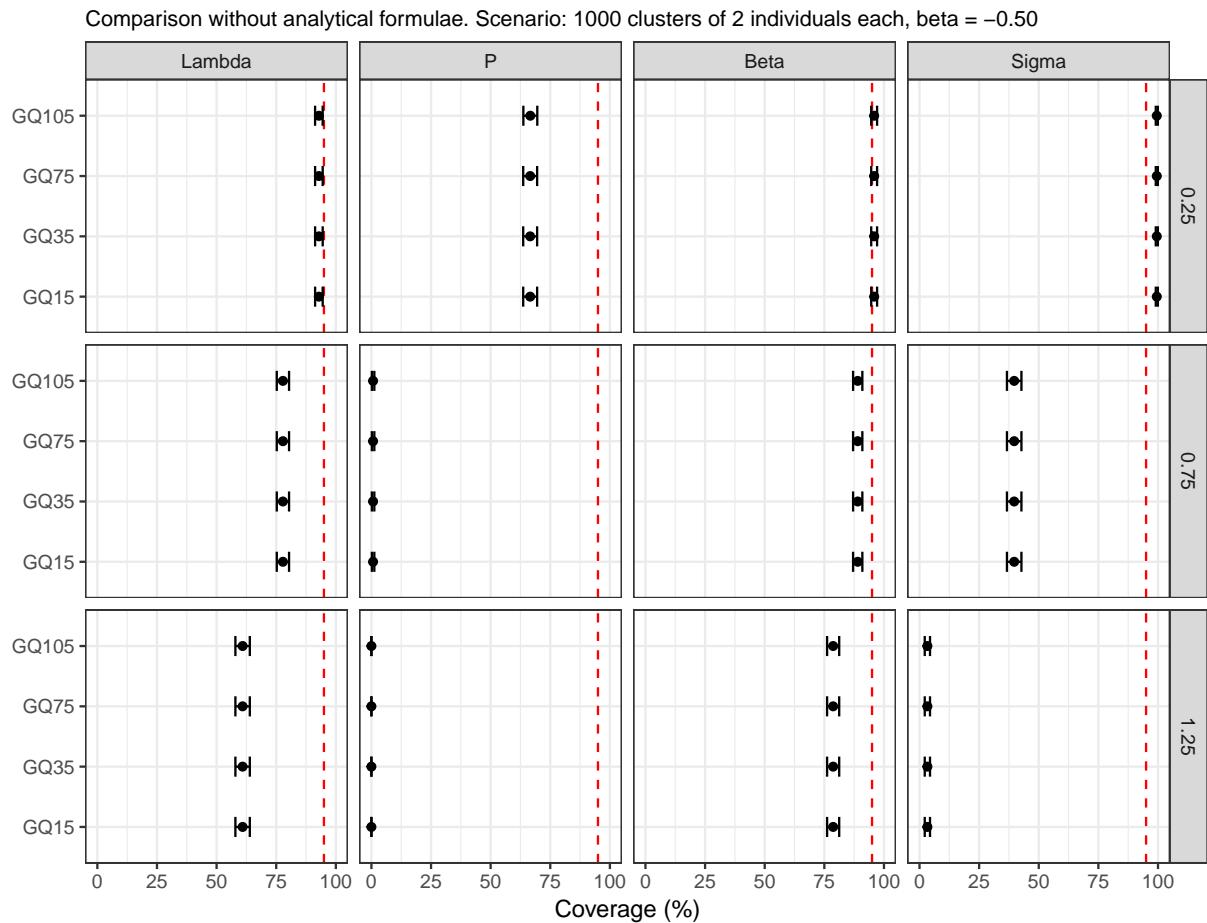


Figure B.24: Coverage, comparison without analytical formulae, scenario with 1000 clusters of 2 individuals each with a positive regression coefficient.

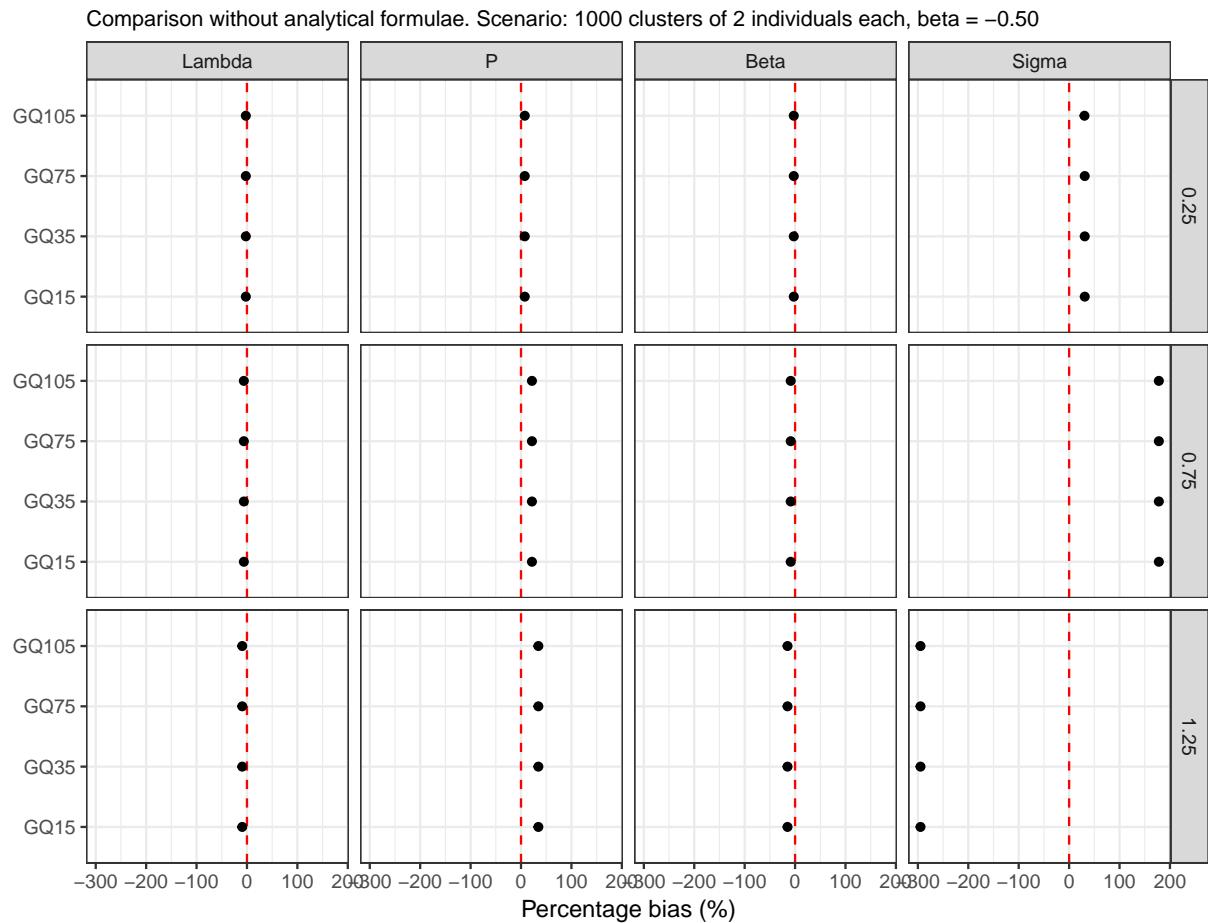


Figure B.25: Percentage bias, comparison without analytical formulae, scenario with 1000 clusters of 2 individuals each with a positive regression coefficient.

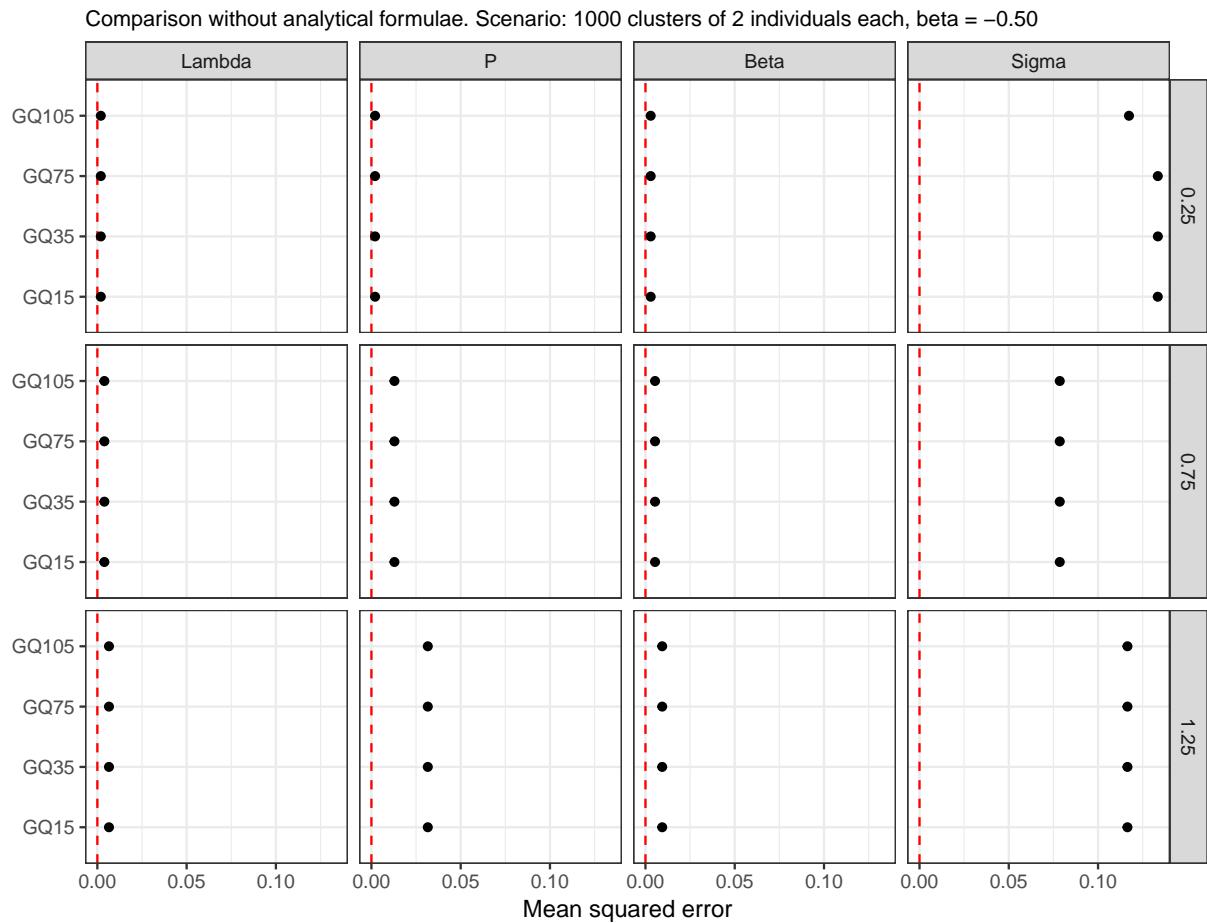


Figure B.26: Mean squared error, comparison without analytical formulae, scenario with 1000 clusters of 2 individuals each with a positive regression coefficient.

B.2 Simulation study 2

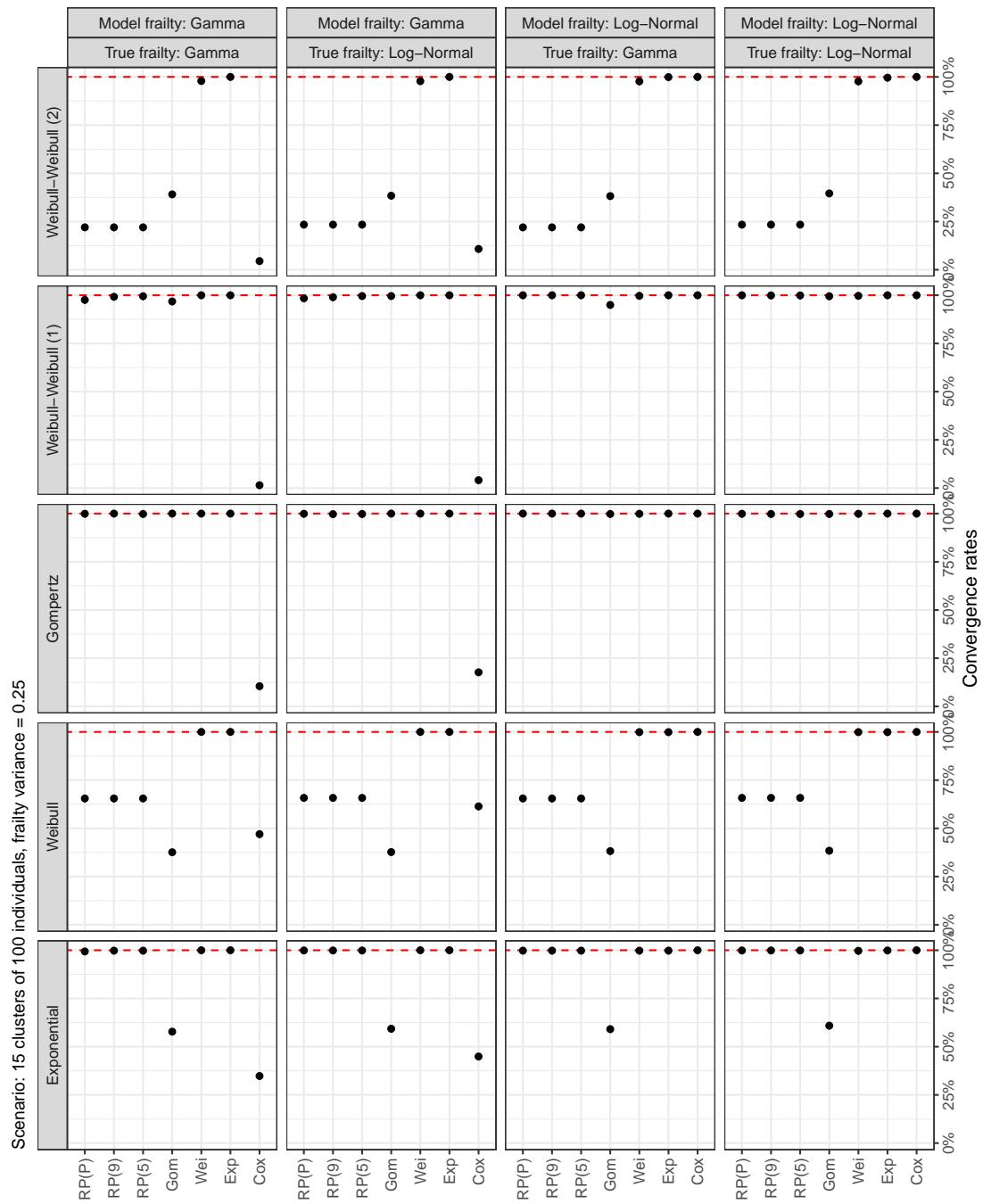


Figure B.27: Convergence rates, simulation study on model misspecification in survival models with shared frailty terms, scenario with 15 clusters of 100 individuals each and a small frailty variance.

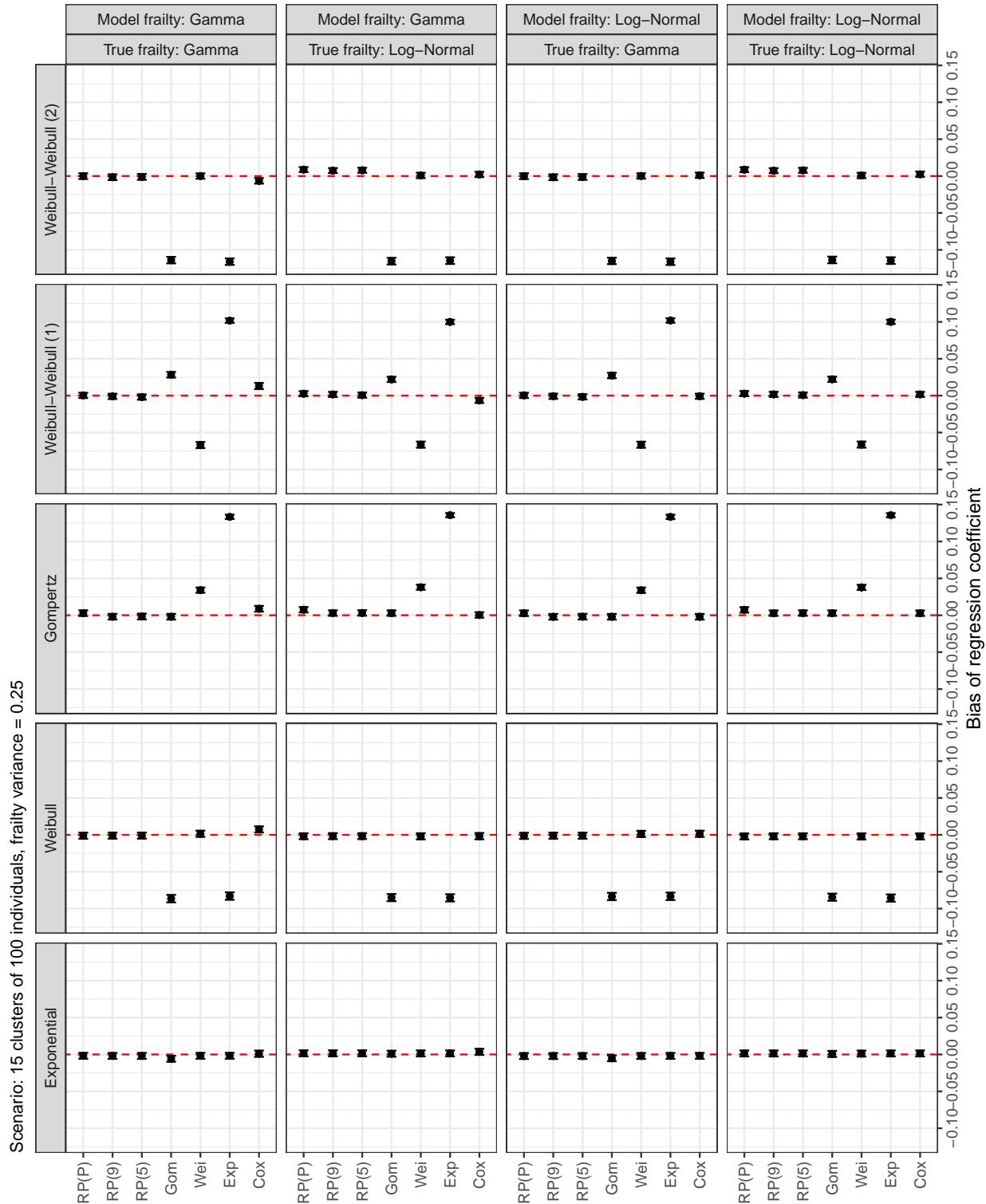


Figure B.28: Bias of estimated regression coefficient, simulation study on model misspecification in survival models with shared frailty terms, scenario with 15 clusters of 100 individuals each and a small frailty variance.

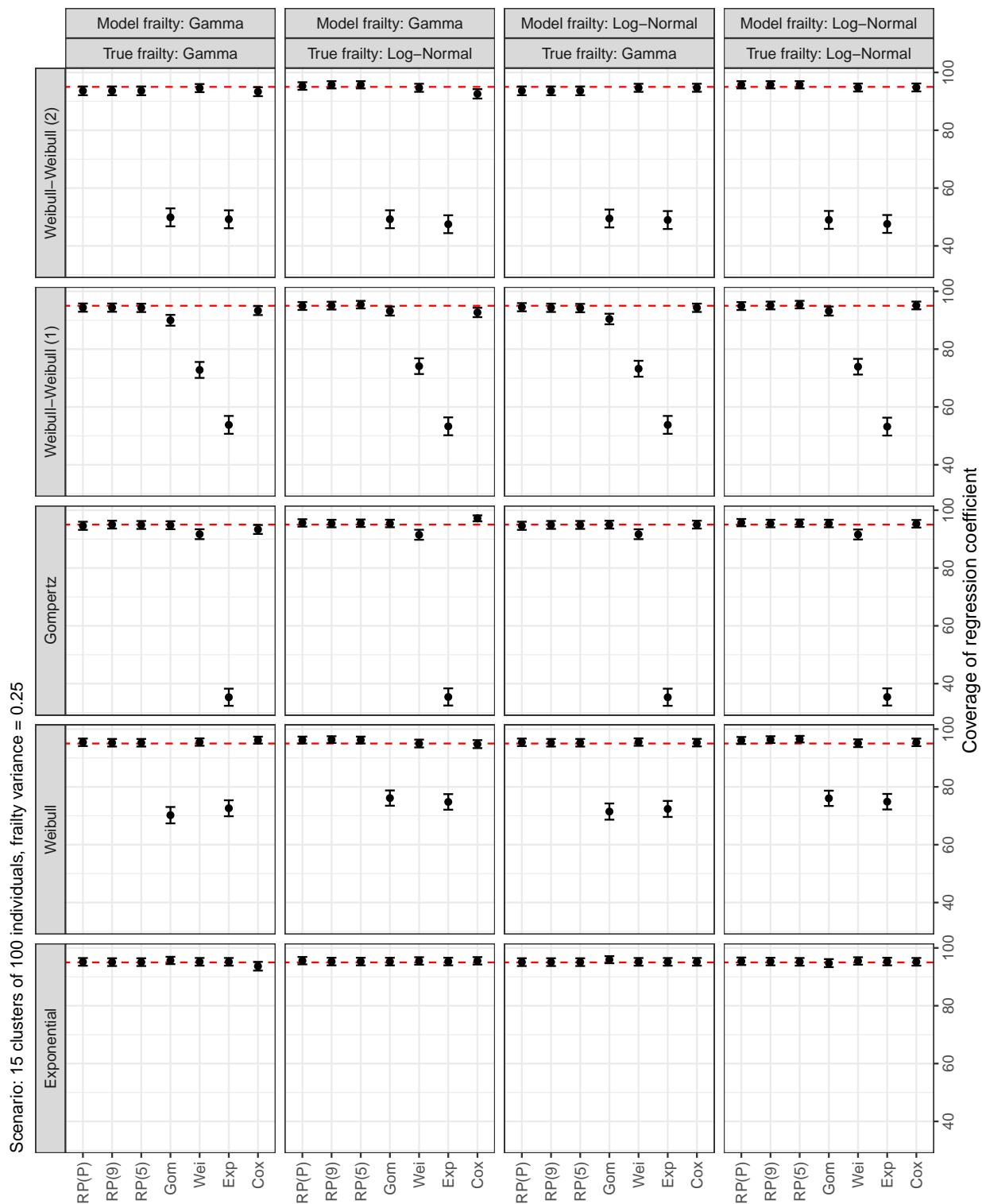


Figure B.29: Coverage of estimated regression coefficient, simulation study on model misspecification in survival models with shared frailty terms, scenario with 15 clusters of 100 individuals each and a small frailty variance.

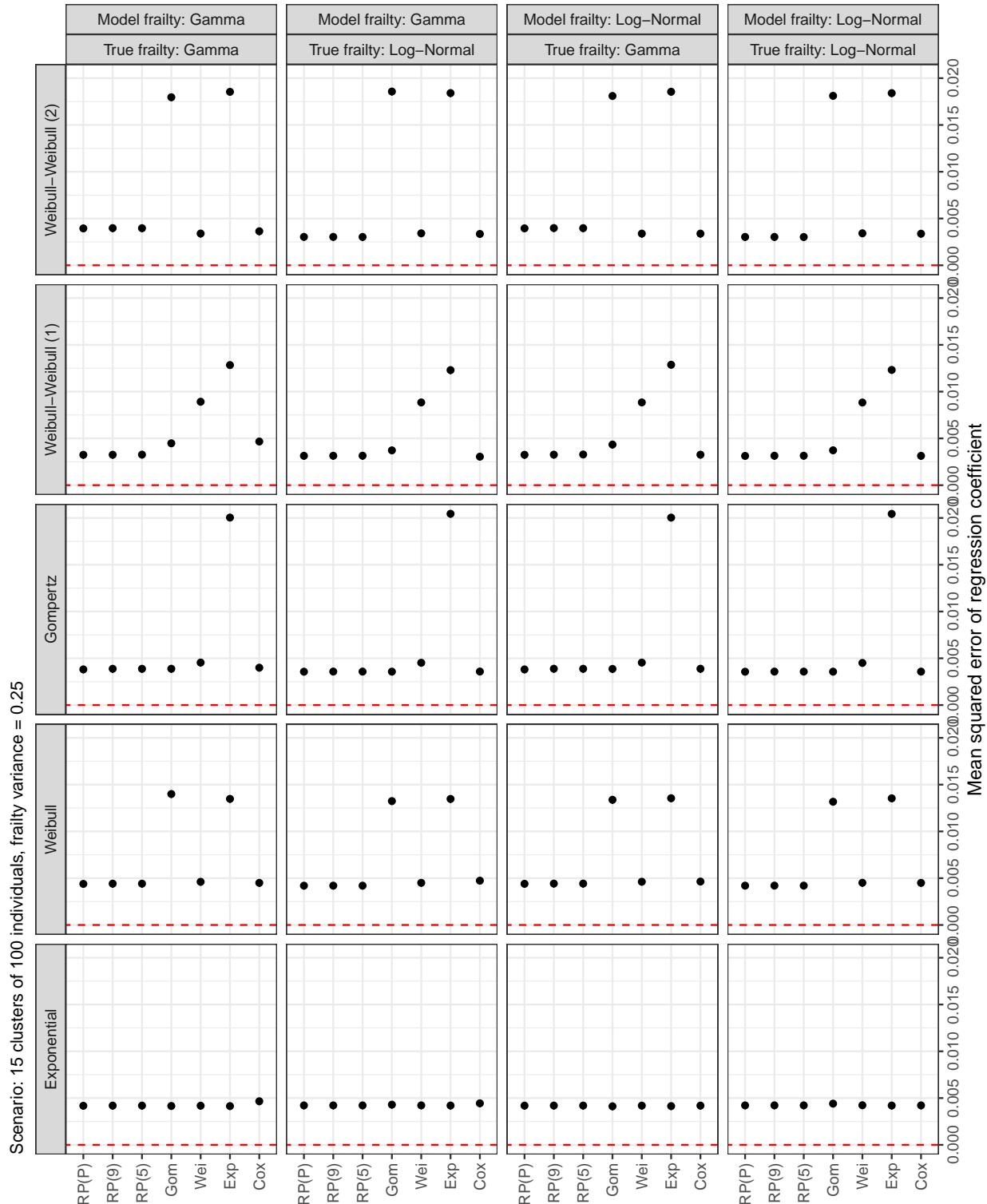


Figure B.30: Mean squared error of estimated regression coefficient, simulation study on model misspecification in survival models with shared frailty terms, scenario with 15 clusters of 100 individuals each and a small frailty variance.

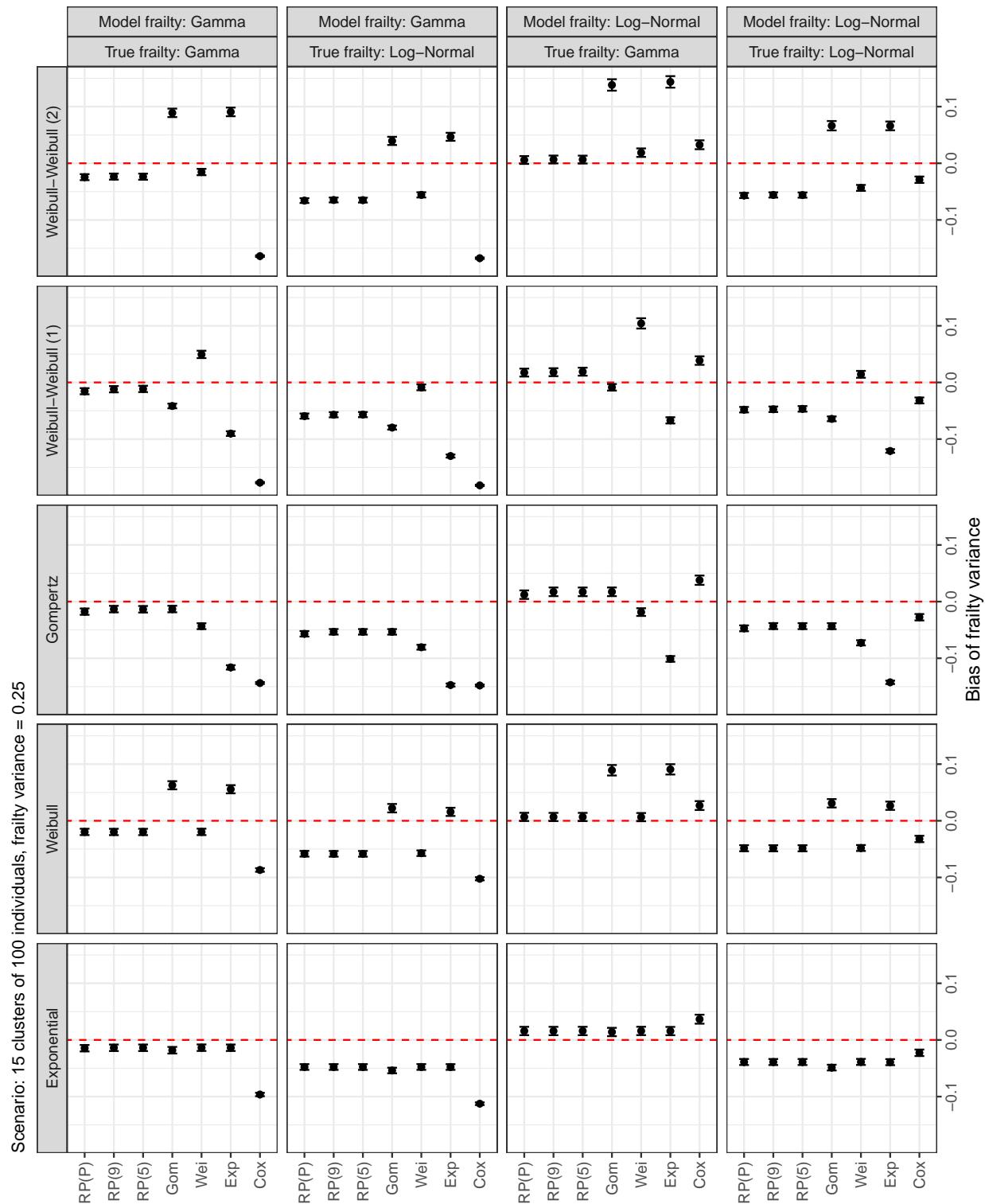


Figure B.31: Bias of estimated frailty variance, simulation study on model misspecification in survival models with shared frailty terms, scenario with 15 clusters of 100 individuals each and a small frailty variance.

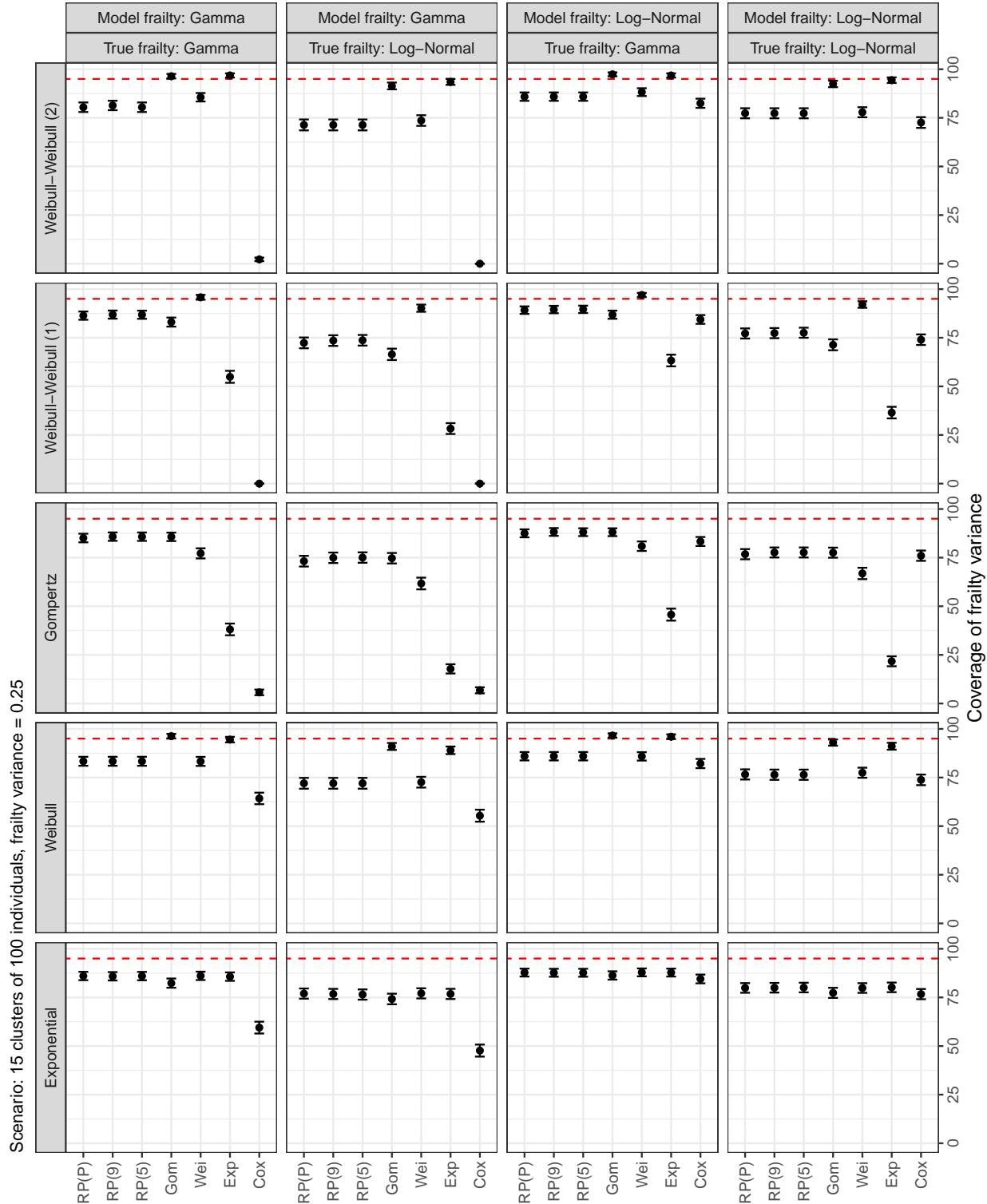


Figure B.32: Coverage of estimated frailty variance, simulation study on model misspecification in survival models with shared frailty terms, scenario with 15 clusters of 100 individuals each and a small frailty variance.

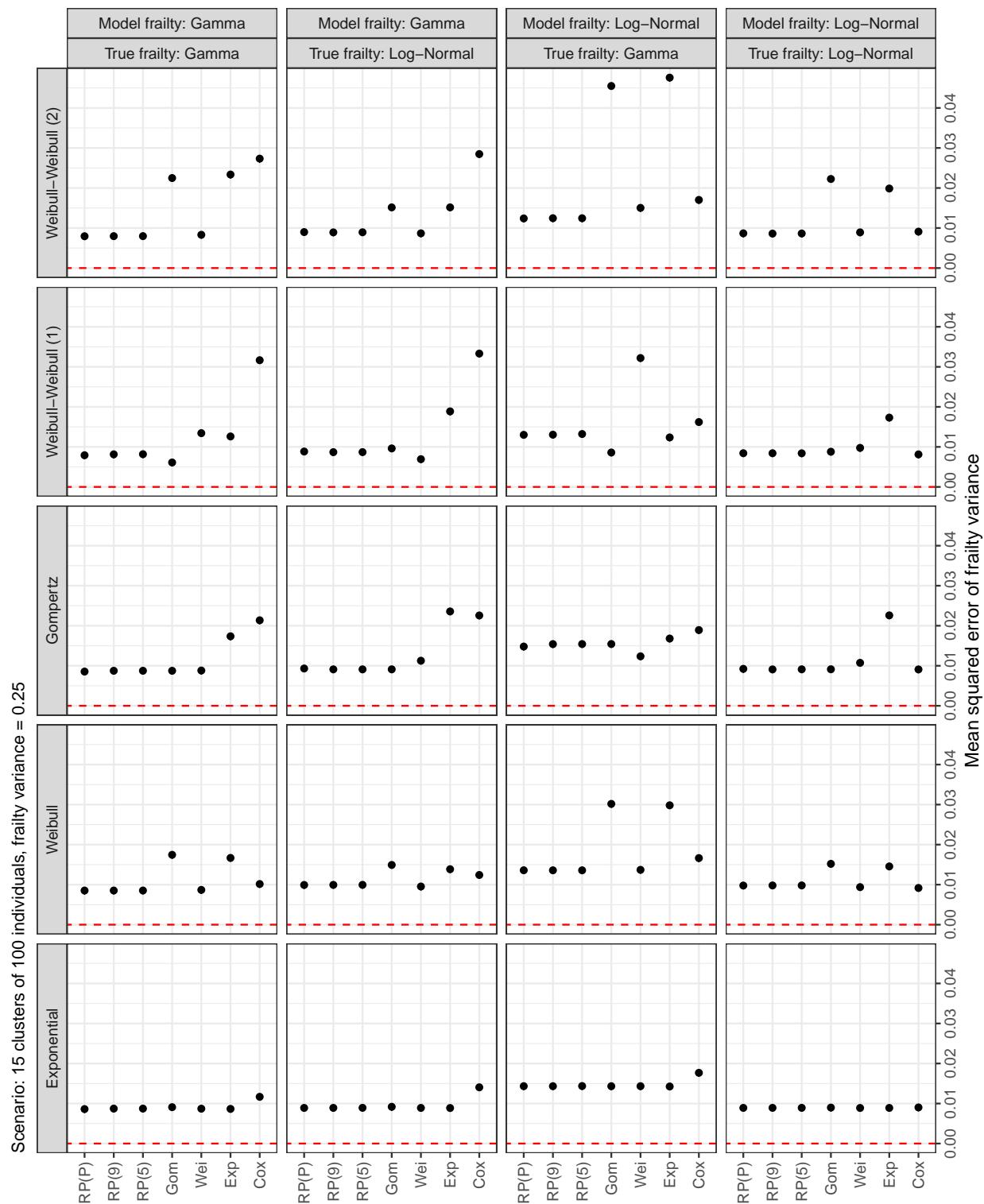


Figure B.33: Mean squared error of estimated frailty variance, simulation study on model misspecification in survival models with shared frailty terms, scenario with 15 clusters of 100 individuals each and a small frailty variance.

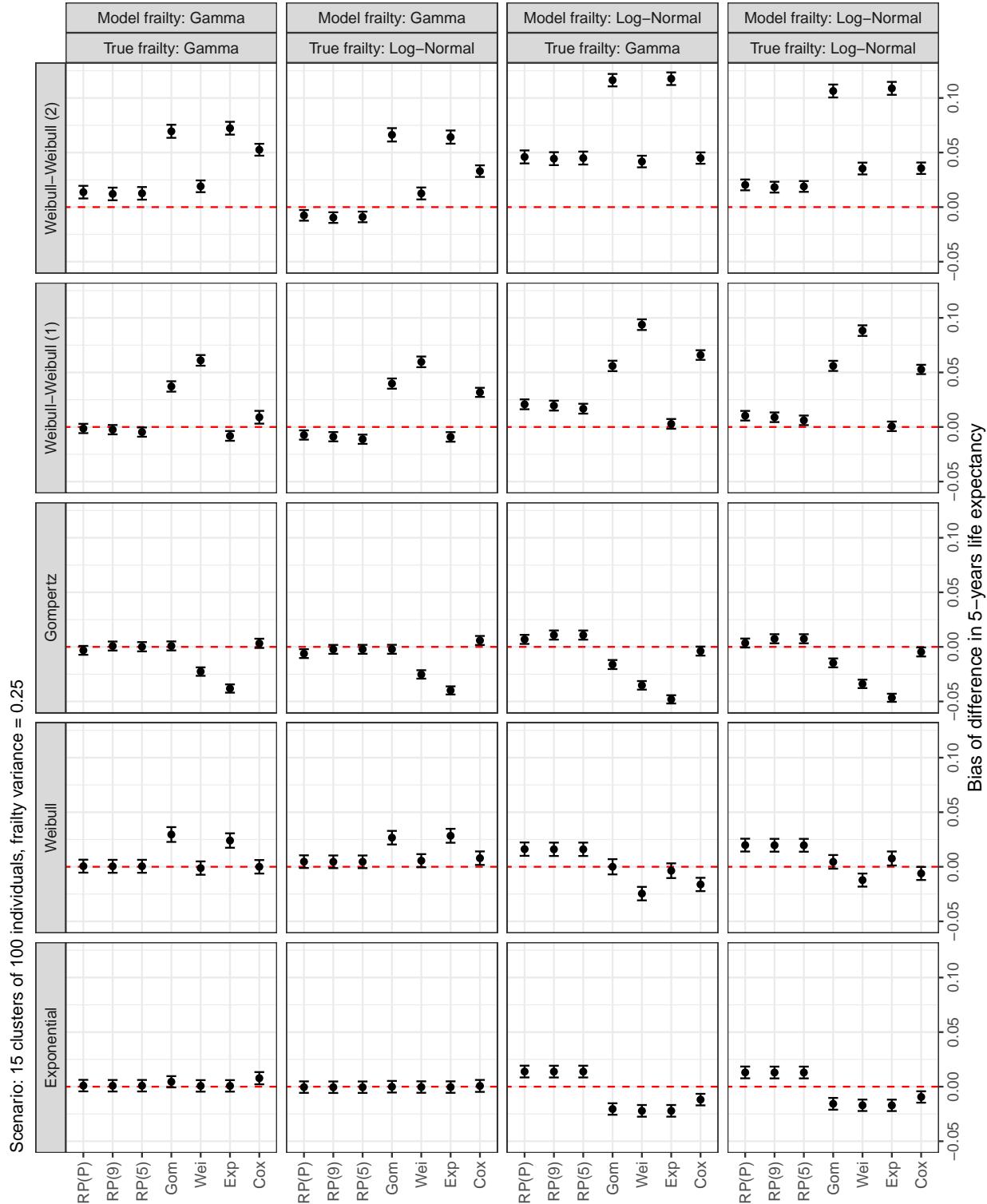


Figure B.34: Bias of estimated difference in 5-years life expectancy, simulation study on model misspecification in survival models with shared frailty terms, scenario with 15 clusters of 100 individuals each and a small frailty variance.

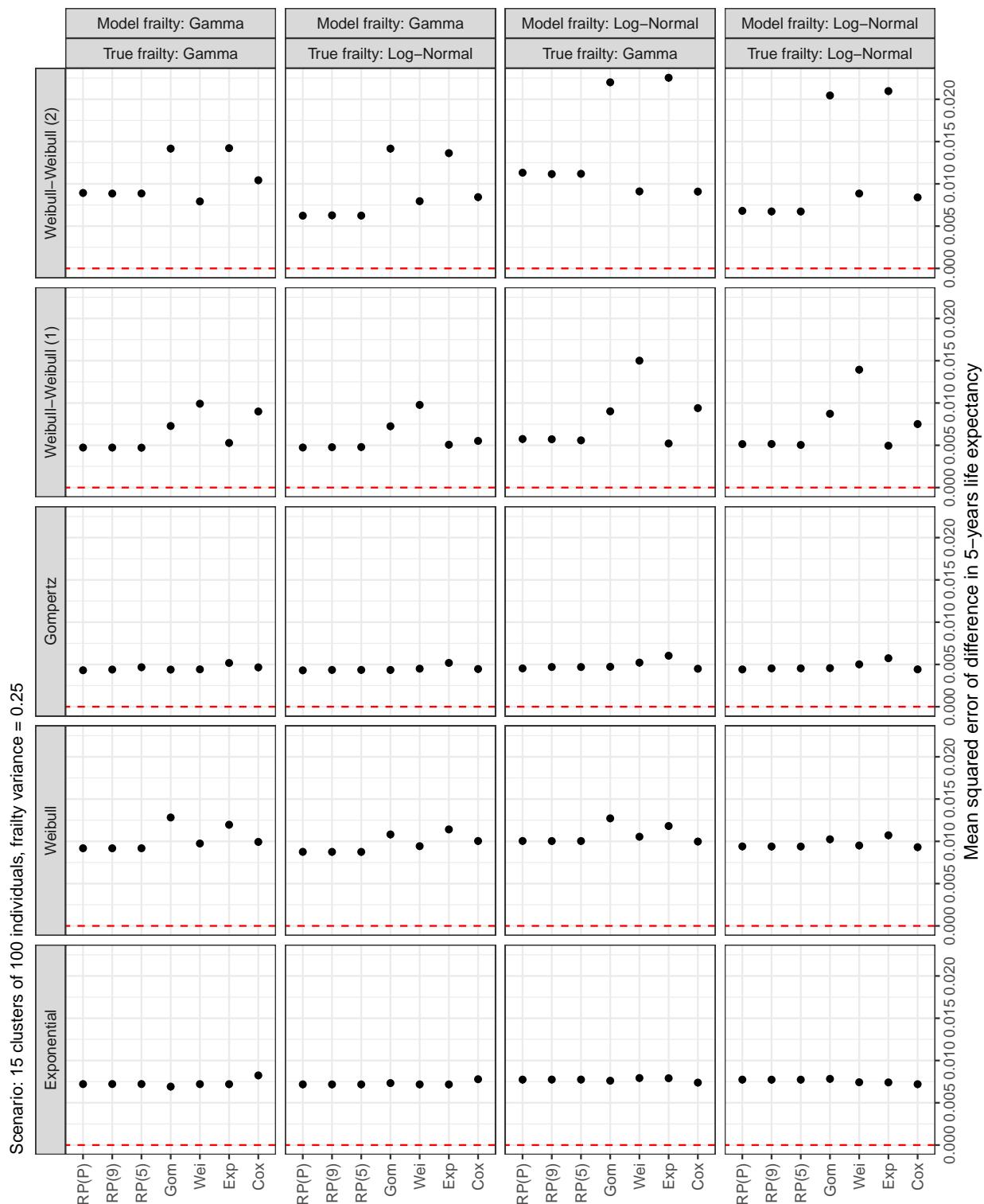


Figure B.35: Mean squared error of estimated difference in 5-years life expectancy, simulation study on model misspecification in survival models with shared frailty terms, scenario with 15 clusters of 100 individuals each and a small frailty variance.

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 5th, 2017

Alessandro Gasparini¹, Keith R Abrams¹, Michael J Crowther¹

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 ..."

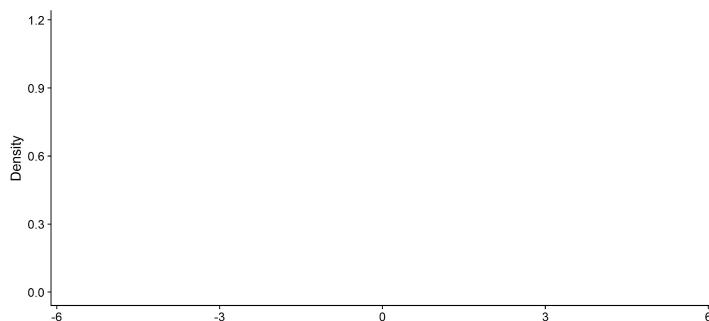
- Quadrature rule: approximation of the integral of a function, usually stated as a weighted sum of function values at specified points within the domain of integration
- n-points Gaussian quadrature rule:

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

- 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 θ^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 (θ): {0.25, 0.50, 1.00}

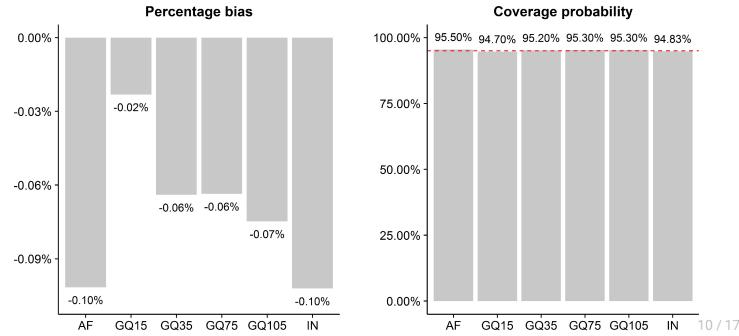
We compare estimates from:

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

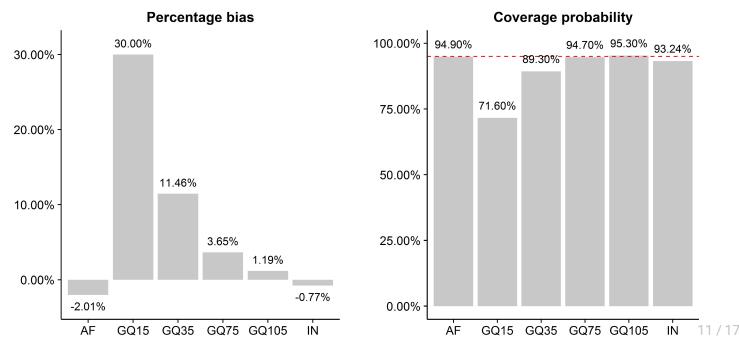
Scenario: 50 clusters of 250 individuals each, negative treatment effect (log-HR: -0.50), medium frailty variance (0.50).
 Parameter of interest: treatment effect.



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 θ .



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

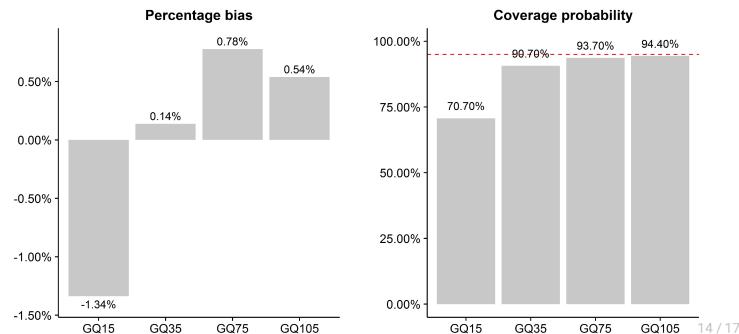
- Weibull baseline hazard with shape $p = 1.5$, scale $\lambda = 3$, and a random treatment effect
- 1,000 simulations per scenario
- number of clusters: {25, 50, 100, 200}, number of individuals per cluster: {25, 50, 100, 250, 500, 1000}
- treatment effect: {-0.50, 0.00, 0.50}
- standard deviation of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

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

13 / 17

Results

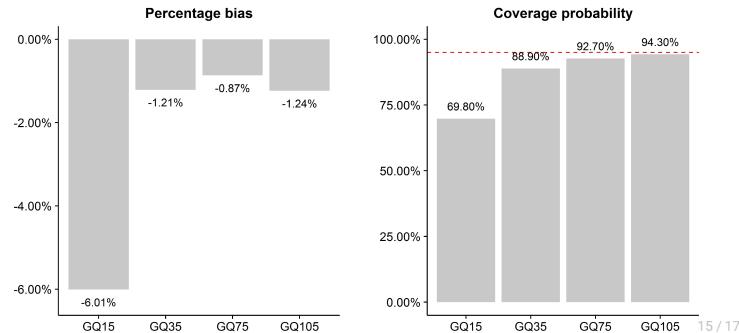
Scenario: 100 clusters of 50 individuals each, positive treatment effect (log-HR: 0.50), high random effect standard error (1.00). Parameter of interest: treatment effect.



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 σ .



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)

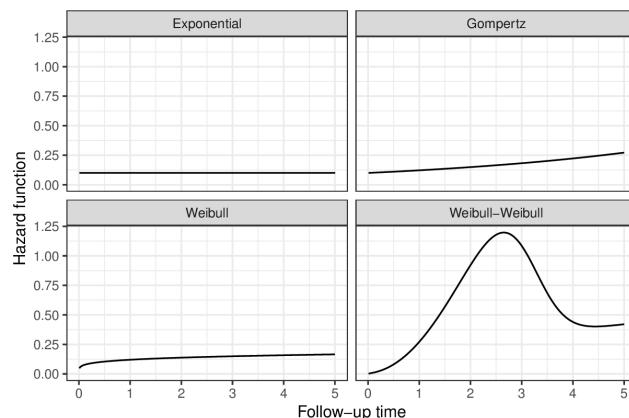
## A simulation study (1)

- ▶ 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

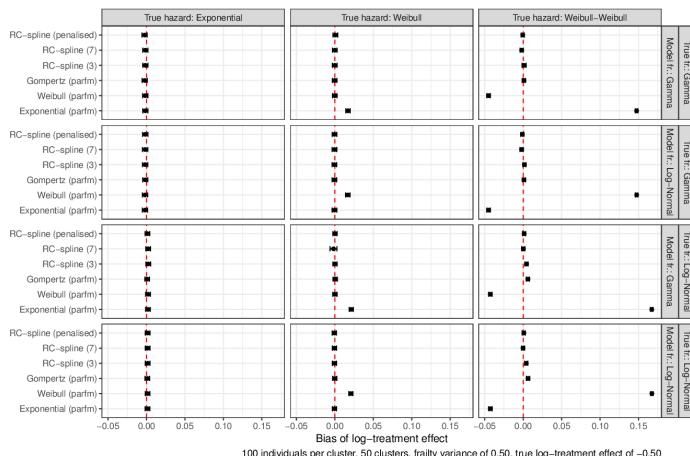
Fully factorial design: 96 simulated scenarios. We present:

- ▶ 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

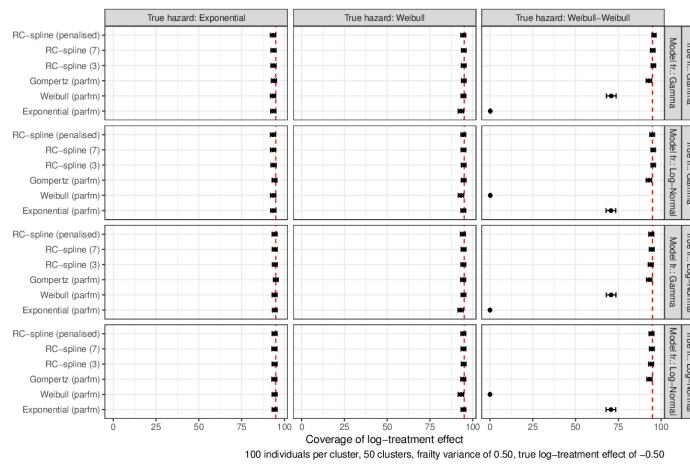
### Results: (1) bias of treatment effect



ag475@leicester.ac.uk

9 of 19

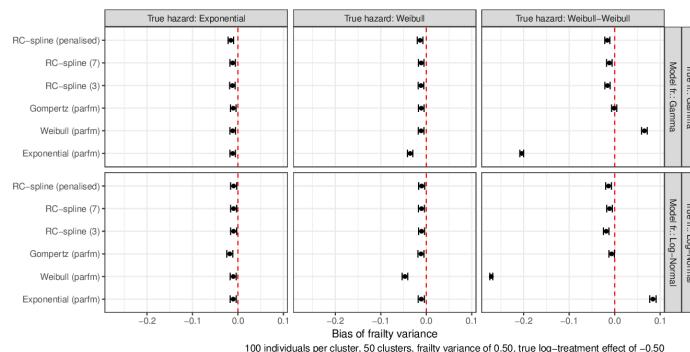
## Results: (i) coverage of treatment effect



ag475@leicester.ac.uk

10 of 19

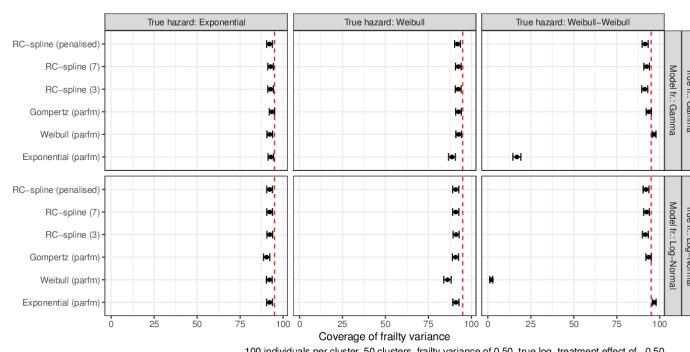
## Results: (i) bias of frailty variance



ag475@leicester.ac.uk

11 of 19

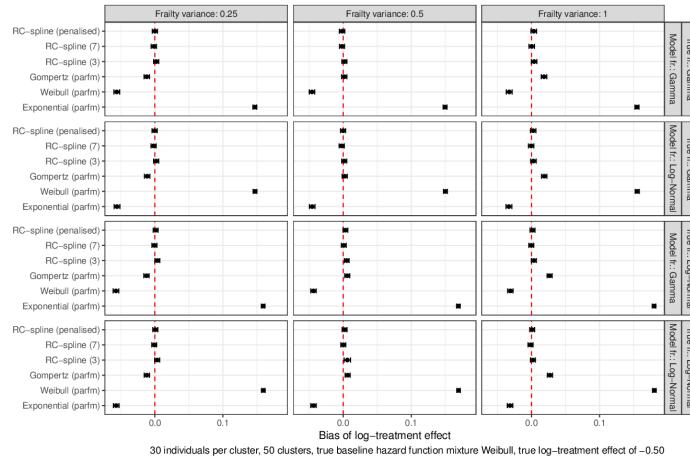
## Results: (i) coverage of frailty variance



ag475@leicester.ac.uk

12 of 19

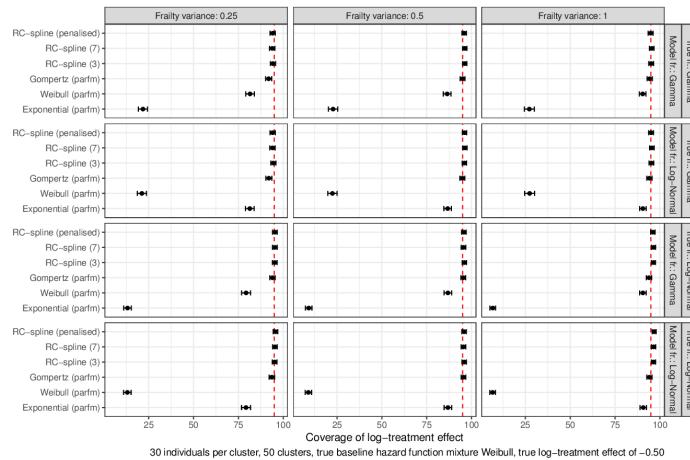
## Results: (2) bias of treatment effect



ag475@leicester.ac.uk

13 of 19

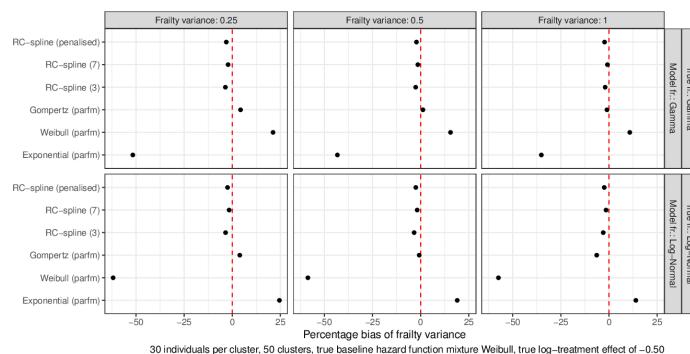
## Results: (2) coverage of treatment effect



ag475@leicester.ac.uk

14 of 19

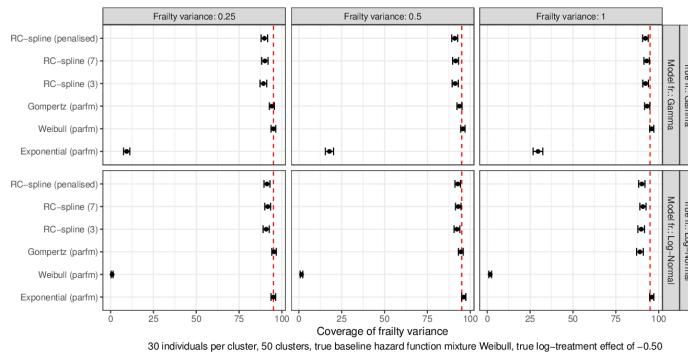
## Results: (2) percentage bias of frailty variance



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 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

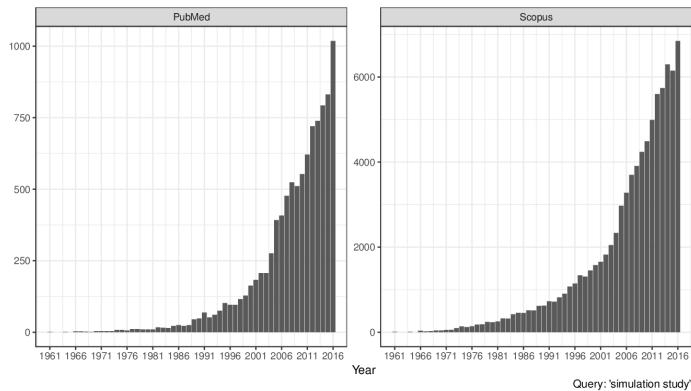
## Simulation studies are useful...

- ▶ 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' (selected), 'Data', 'Explore results', and 'Source code'. To the right of the sidebar is a light blue tab labeled '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: This study consisted 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 one 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( $d$ ) 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                 | 0.003624    | 0.002233    | -0.004412   |

ag475@leicester.ac.uk

15 of 21

## Demo: table of summary results

```

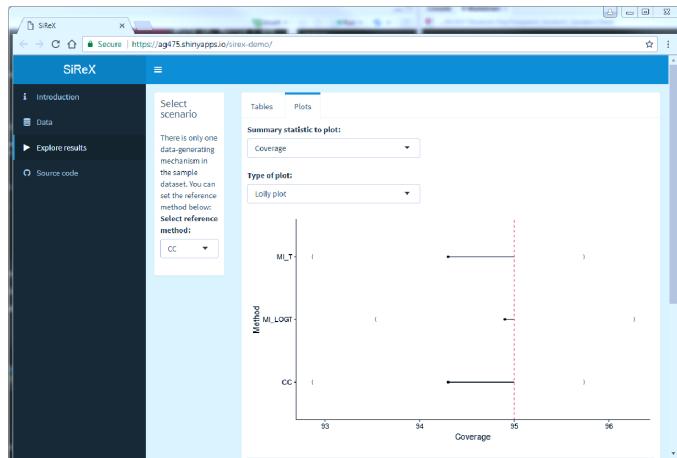
LatexCode for this table
%
% latex table generated in R 3.4.8 by xtable 1.8-2 package
% Sun Jun 30 13:34:37 2017
\begin{table}[ht]
\centering
\begin{array}{l}
\hline
\texttt{& statistic \& CC \& MI_LOOT \& MI_T \\\hline} \\
1 & 1000.00 \& 1000.00 \& 1000.00 \\\hline
2 & Non-missing estimator & 1000.00 \& 1000.00 \& 1000.00 \\\hline
3 & Non-missing std. errors & 1000.00 \& 1000.00 \& 1000.00 \\\hline
4 & Mean estimate & 0.92 \& 0.90 \& 0.89 \\\hline
5 & Variance of estimate & 0.02 \& 0.02 \& 0.02 \\\hline
6 & Mean std. error & 0.02 \& 0.02 \& 0.02 \\\hline
7 & Variance of std. error & 0.00 \& 0.00 \& 0.00 \\\hline
8 & Coverage & 94.39 \& 95.30 \& 94.39 \\\hline
9 & Power & 0.72 \& 0.73 \& 0.71 \\\hline
10 & Power & 94.40 \& 96.90 \& 95.30 \\\hline
11 & MCSE of power & 0.72 \& 0.55 \& 0.46 \\\hline
12 & Bias & 0.00 \& 0.00 \& 0.00 \\\hline
13 & MCSE of bias & 0.00 \& 0.00 \& 0.00 \\\hline
14 & Empirical std. error & 0.16 \& 0.13 \& 0.13 \\\hline
15 & MCSE of empirical std. error & 0.00 \& 0.00 \& 0.00 \\\hline
16 & MCSE of relative std. error & 0.15 \& 0.13 \& 0.13 \\\hline
17 & MCSE of relative std. error & 0.40 \& 0.40 \& 0.40 \\\hline
18 & Relative error & 0.03 \& 0.02 \& 0.00 \\\hline
19 & MCSE of relative error & 0.02 \& 0.02 \& 0.02 \\\hline
20 & Relative gain in precision & 0.00 \& 51.95 \& 26.37 \\\hline
21 & MCSE of relative gain in precision & 0.00 \& 5.54 \& 3.84 \\\hline
\end{array}

```

ag475@leicester.ac.uk

16 of 21

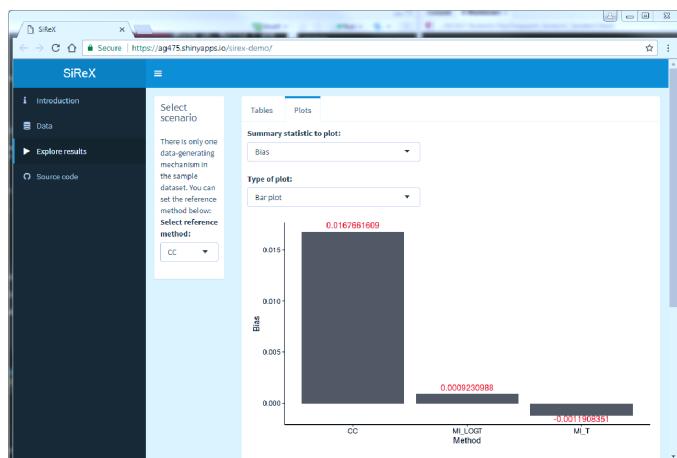
## Demo: plot of summary results



ag475@leicester.ac.uk

17 of 21

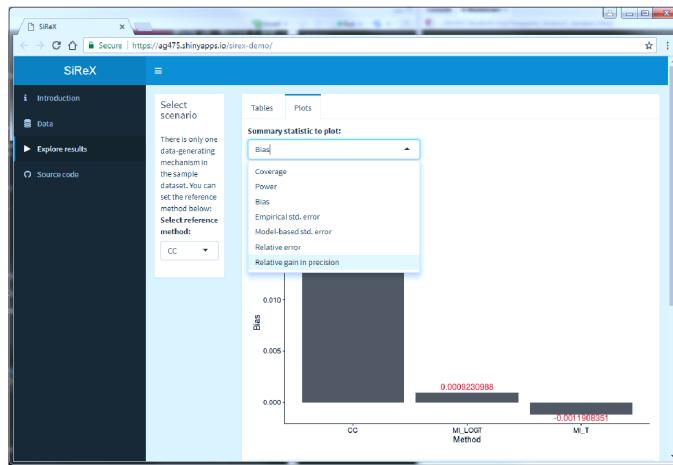
## Demo: plot of summary results



ag475@leicester.ac.uk

18 of 21

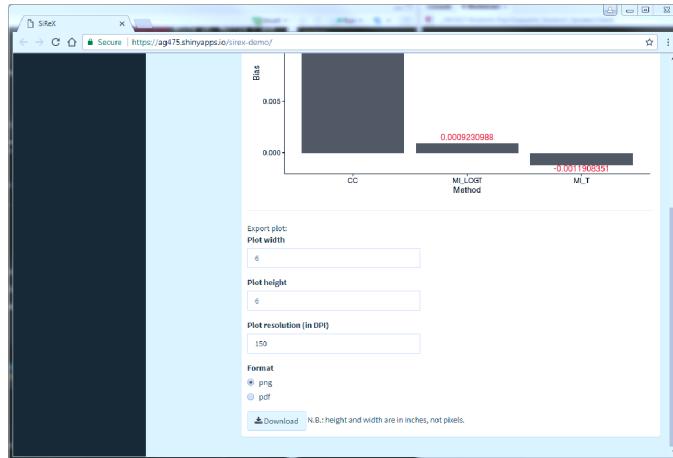
### Demo: plot of summary results



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



# Appendix E

## R Session

```
sessionInfo()

R version 3.4.1 (2017-06-30)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 7 x64 (build 7601) Service Pack 1
##
Matrix products: default
##
locale:
[1] LC_COLLATE=English_United Kingdom.1252
[2] LC_CTYPE=English_United Kingdom.1252
[3] LC_MONETARY=English_United Kingdom.1252
[4] LC_NUMERIC=C
[5] LC_TIME=English_United Kingdom.1252
##
attached base packages:
[1] stats graphics grDevices utils datasets methods base
```

```

other attached packages:
[1] bindrcpp_0.2 cowplot_0.8.0 kableExtra_0.4.0 tidyR_0.7.0
[5] scales_0.4.1 marqLevAlg_1.1 readr_1.1.1 ggfortify_0.4.1
[9] dplyr_0.7.2 ggpubr_0.1.4 magrittr_1.5 survival_2.41-3
[13] ggplot2_2.2.1 pacman_0.4.6 knitr_1.17 bookdown_0.4

loaded via a namespace (and not attached):
[1] Rcpp_0.12.12 highr_0.6 compiler_3.4.1
[4] plyr_1.8.4 bindr_0.1 tools_3.4.1
[7] digest_0.6.12 evaluate_0.10.1 tibble_1.3.3
[10] gtable_0.2.0 lattice_0.20-35 pkgconfig_2.0.1
[13] rlang_0.1.2 Matrix_1.2-11 rstudioapi_0.6
[16] yaml_2.1.14 gridExtra_2.2.1 xml2_1.1.1
[19] httr_1.3.0 stringr_1.2.0 htmlwidgets_0.9
[22] hms_0.3 tidyselect_0.1.1 rprojroot_1.2
[25] grid_3.4.1 glue_1.1.1 R6_2.2.2
[28] rmarkdown_1.6 reshape2_1.4.2 purrr_0.2.3
[31] backports_1.1.0 htmltools_0.3.6 splines_3.4.1
[34] rvest_0.3.2 assertthat_0.2.0 formattable_0.2.0.1
[37] colorspace_1.3-2 labeling_0.3 stringi_1.1.5
[40] lazyeval_0.2.0 munsell_0.4.3
```

# 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.
- Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics*, 30(1):89–99.
- Clayton, D. G. (1991). A Monte Carlo method for Bayesian inference in frailty models. *Biometrics*, 47(2):467.
- Commenges, D., Jacqmin-Gadda, H., Proust, C., and Guedj, J. (2006). A newton-like algorithm for likelihood maximization: the robust-variance scoring algorithm. *ArXiv Mathematics e-prints*.
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 34(2):187–220.
- Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62(2):269–276.
- Crowder, M. (2016). *Multivariate survival analysis and competing risks*. Text in Statistical Sciences. Chapman & Hall / CRC.
- 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.
- Efron, B. (1977). The efficiency of Cox’s likelihood function for censored data. *Journal of the American Statistical Association*, 72(359):557–565.
- Geskus, R. B. (2015). *Data analysis with competing risks and intermediate states*. The Biostatistics Series. Chapman & Hall / CRC.
- 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.
- International Stroke Trial Collaborative Group (1997). The international stroke trial (ist): a randomised trial of aspirin, subcutaneous heparin, both, or neither among 19 435 patients with acute ischaemic stroke. *The Lancet*, 349(9065):1569–1581.
- Kalbfleisch, J. D. and Prentice, R. L. (2011). *The statistical analysis of failure time data*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc, 2 edition.

- Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481.
- Kleinbaum, D. G. and Klein, M. (2012). *Survival analysis: A self-learning text*. Springer-Verlag New York, 3 edition.
- 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*.
- Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441.
- 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.
- Peto, R. (1972). Contribution to discussion of paper by D.R. Cox. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 34:205–207.
- 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.

- Sandercock, P., Niewada, M., and Czlonkowska, A. (2011). International stroke trial database (version 2). Technical report, University of Edinburgh, Department of Clinical Neuroscience.
- Sinha, D., Maiti, T., Ibrahim, J. G., and Ouyang, B. (2008). Current methods for recurrent events data with dependent termination. *Journal of the American Statistical Association*, 103(482):866–878.
- 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., Degruydtola, 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.