

Beyond the shared frailty model

Goodness-of-fit, decreasing heterogeneity and spatial dependence

A thesis submitted in partial fulfillment of the requirements for the degree of Docteur en Sciences, Université catholique de Louvain, September 2014

MARCO MUNDA

Thesis committee:

Catherine Legrand, supervisor (Université catholique de Louvain)
Ingrid Van Keilegom, President (Université catholique de Louvain)
Philippe Lambert (University of Liège, Université catholique de Louvain)
Paul Janssen (Hasselt University)
Luc Duchateau (Ghent University)
Hein Putter (Leiden University Medical Center)
Virginie Rondeau (Université Bordeaux Segalen)

“ Statisticians, like artists, have the bad habit of falling in love with their models. ”

George E. P. Box

Acknowledgements

This thesis was conducted at the Institute of Statistics, Biostatistics and Actuarial Sciences (ISBA) at the Université catholique de Louvain, which I am leaving after four years. I would like to take this opportunity to thank a number of people.

I thank my supervisor, Catherine Legrand. Catherine, thank you for the trust you placed in me, for the freedom you gave me, and for making this a pleasant experience.

From the start I benefited greatly from discussions and collaborations with Paul Janssen and Luc Duchateau. Paul, thank you for your advice and all the guidance you provided along the way. On many occasions we discussed time-varying frailties, but the support you showed me has always been constant. Luc, thank you for your enthusiasm and your continuous encouragement. You always had a kind word for me! Thank you also for entrusting me with the “malaria project”.

I thank the jury members for their interest and their suggestions during my private defence. Ingrid, Hein, I will remember that you always greeted me with a smile. Virginie, thank you for inviting me to join the Institute of Public Health, Epidemiology and Development at the University of Bordeaux and for the warm welcome during my visit.

My thanks are extended to the administrative staff of ISBA, with a special thank you to Nancy Guillaume. Nancy, thank you for helping me whenever I needed. I also thank the SMCS, the statistical consulting service of ISBA, which gave me the opportunity to get my hands dirty with real data.

Finally, I thank Anne, the best colleague ever! I also thank Adrien, Alain, Alexandra, Alina, Anna, Aurélie, Benjamin, Catherine, Cédric, Céline, Eduardo, Federico, Habiba, Jennifer, Maïlis, Mathieu, Michał, Nathalie, Nathan, Nicolas, Sylvie, Vincent, and Yassin with whom I loved spending the past four years.

Table of contents

1	Introduction	1
1.1	Outline	1
1.2	Elements of survival analysis	5
1.3	Frailty model	14
Part I		17
	The frailty distribution	
2	A unified framework for fitting the frailty model with different frailty distributions	19
2.1	Conditional, complete data, and marginal likelihoods .	20
2.2	Generic form of the marginal likelihood	21
2.3	Parametric framework	22
2.4	Derivative formulas	22
2.5	<code>parfm</code>	27
	Appendix	30
2.A	The $\Omega_{q,m}(\cdot)$'s polynomials	31
2.B	Derivative formulas in R	32
3	A diagnostic plot for the frailty distribution in the shared frailty model	35
3.1	Introduction	36
3.2	Dependence in bivariate survival data	36
3.3	Quantile dependence	41
3.4	Examples	48
3.5	Simulations	50
3.6	Discussion	53
	Appendix	55

3.A	Explicit formulas	57
3.B	Kendall's τ versus θ	59
3.C	Model-free and model-based estimators of $\lambda_\ell(q)$	61
Part II		63
	Heterogeneity in multicentre trials	
4	Adjusting for centre heterogeneity in multicentre clinical trials with a time-to-event outcome	65
4.1	Modelling clustered time-to-event data	66
4.2	Comparison	68
4.3	Robustness against frailty misspecification	76
4.4	Conclusions	77
	Appendix	81
4.A	Sample codes	81
5	Testing for decreasing heterogeneity between hospitals in time to death from chronic myeloid leukemia	83
5.1	Literature review and outline	84
5.2	Time-varying frailty model	85
5.3	Example	91
5.4	Simulation study	95
5.5	Discussion	104
	Appendix	104
5.A	Model (5.4) in matrix notation	109
5.B	Generation of time-varying random effects	110
Part III		113
	Spatial dependence	
6	A first step towards modelling spatial dependence in time to malaria data using the frailty model methodology	115
6.1	Literature review and outline	116

6.2	Spatial frailty model	118
6.3	Spatial dependence	121
6.4	Malaria data analysis	124
6.5	Future development	126
	Appendix	128
6.A	Laplace approximation of the marginal likelihood . . .	129
6.B	Semi-variograms of T and $\log(T)$ for the case $M = 1$.	130
Closing _____		133
Appendix _____		143
1	Frailty distributions	145
1.1	Gamma	145
1.2	Inverse Gaussian	147
1.3	Positive stable	149
1.4	Log-normal	151
1.5	Power variance function	153
1.6	Compound Poisson	154
2	Generation of survival times to simulate the frailty model	155
2.1	The inverse probability method	155
2.2	Examples	156
2.3	Sample code in R	157
3	Bootstrap in the frailty model	159
3.1	Model-based bootstrap	159
References _____		161

Introduction

1.1 --- Outline

In survival analysis, the outcome of interest (response variable) is a time-to-event endpoint, also called survival endpoint, e.g. time to death in cancer clinical trials. The central problem in the analysis of time-to-event endpoints is the presence of censoring. In its most common form (right censoring), censoring occurs when the follow-up is interrupted prior to the event's occurrence. In that case, only a lower boundary is known for the event time. Section 1.2 provides a general, practical overview of survival analysis. Topics covered include terminology, special features of survival data, working assumptions, and standard methods of analysis to deal with censored data (non-parametric Kaplan-Meier survival curve, parametric Weibull regression, semi-parametric Cox regression). Further details can be found in textbooks, e.g. Collett (2003).

Standard methods in survival analysis require independent event times, given the covariate information. In practice, many studies involve clusters. Examples of clusters include families, geographical areas, centres participating in a clinical trial, etc. Within a cluster, data are typically dependent. This thesis focuses on the frailty model, introduced in Section 1.3, to account for the dependence in clustered survival data. In the frailty model framework, the within-cluster association is taken into account by means of a cluster-specific factor, the frailty term. Typical for this model is that the frailty term is treated as a random effect.

Random effects for survival data were first considered in Beard (1959) and Vaupel et al. (1979) to improve the fit of mortality models at advanced ages. In these early papers, the frailty term acts at the individual level as an unobservable factor in the mortality model and indicates that “frail” people have an increased risk of death. The distribution of the individual-specific frailty term provides a way to model unob-

served/unexplained variation in susceptibility to death (heterogeneity) in the population.

Applications of the frailty model to clustered data were first discussed in Clayton (1978) for studies of familial aggregation of disease. Due to, for example, genetic and environmental factors, susceptibility to disease varies from family to family. Hence, variability in outcome among relatives tends to be lower than variability in outcome between non-relatives. In statistical terms, this translates into (i) unobserved heterogeneity between families, and into (ii) association among observations from the same family. Frailty models with a cluster-specific frailty term, also called shared frailty models, have been developed over the last three decades to deal with this type of data.

For an introduction to the most important aspects of the frailty model methodology, as well as an extensive list of references to major papers and recent developments in the area, see Duchateau & Janssen (2008) and Wienke (2010). Nice applications illustrating the usefulness of the frailty model in practice include Michiels et al. (2005), Legrand et al. (2006), and Rondeau et al. (2010).

This thesis is organised around three research themes of current interest in frailty modelling. The content of each research theme is outlined below.

Part I – The frailty distribution

Different distributions have been proposed to model the frailty term. To mimic the normal random effect distribution from the linear mixed model, the log-normal distribution can be used. Compared to the log-normal, the gamma distribution has the advantage of mathematical tractability; it is the most common distribution in practice. The inverse Gaussian and the positive stable have proven useful to model alternative dependence structures in the data. The power variance function distribution contains the inverse Gaussian as a special case, as well as the gamma and the positive stable as limiting cases. Therefore, the family of power variance function distributions might be useful in a goodness-of-fit context. The compound Poisson distribution has the particularity of having a point mass at zero. The compound Poisson distribution thus provides a way to model a cure fraction in the population. The general characteristics of these frailty distributions are studied in detail in Hougaard (2000, Chapter 7) and in Duchateau & Janssen (2008,

Chapter 4). In the thesis appendix (Appendix 1), we collect their main characteristics for ease of reference.

In practice, mainly the gamma and the log-normal distributions are used to model the frailty term. The application of other frailty distributions is hindered by the lack of software. To fill this gap, we have developed the R library `parfm`, which stands for “parametric frailty models”. To fit the frailty model in the `parfm` framework, it suffices to be able to compute higher-order derivatives of the Laplace transform of the frailty term. Currently, `parfm` supports the gamma, the inverse Gaussian, the positive stable, and the log-normal frailty distributions. Additional frailty distributions, including the power variance function and the compound Poisson, may be added in the future. Chapter 2, which extends the results published in Munda et al. (2012), outlines the methodology of `parfm`.

Implicit in the frailty model approach to adjust for dependence in clustered survival data is the assumption that the frailty term varies randomly among clusters according to the specified frailty distribution. For multicentre clinical trials, where the dependence is usually not of direct interest (nuisance), we will argue that inference about the treatment effect is robust against misspecification of the frailty distribution (cf. Chapter 4). In contrast, the dependence is the interesting aspect of some studies, e.g. genetic studies in twins (Hougaard et al., 1992) or treatment outcome studies (Legrand et al., 2002, 2006). Typical for such studies is that the dependence structure is dictated by the choice of frailty distribution (Anderson et al., 1992; Hougaard, 1995). Research on diagnostic techniques to assess the frailty distribution assumption is sparse and such techniques are rarely used in practice. Chapter 3 introduces a new diagnostic plot for the frailty distribution in the shared frailty model. The content of this chapter is submitted for publication (Munda & Legrand, 2014b).

Part II – Heterogeneity in multicentre trials

An important field of applications of frailty models concerns the analysis of multicentre randomised clinical trials. In such a trial, patients who meet the eligibility criteria are allocated at random to one of two (or more) treatment groups and followed up for the endpoint of interest. In a multicentre clinical trial, randomisation is usually stratified by centre, and centres may be quite different from one another (differences in centre

type, differences in specialty, differences in standard practice patterns, differences in patient management, differences in catchment area, etc.). Multicentre clinical trials are therefore subject to centre heterogeneity (clustering). The shared frailty model then provides a convenient tool for the analysis of multicentre clinical trials with a time-to-event endpoint.

The primary analysis of multicentre randomised clinical trials with a time-to-event endpoint is commonly based on the Cox model. The ICH E9 guidelines for the statistical analysis of clinical trials (Lewis, 1999) emphasise that “a method of adjustment” for centre heterogeneity should be used in the primary analysis. The cases of continuous (Chu et al., 2011; Kahan & Morris, 2013) and binary (Kahan, 2014) outcomes have been discussed previously. Mixed effects models can be used advantageously in those settings. A similar investigation for time-to-event outcomes is undertaken in Glidden & Vittinghoff (2004). The latter paper is the starting point for further discussion in Chapter 4. The central question addressed is whether the frailty approach is the method to be recommended, considering the fact that frailty misspecification is the rule, rather than the exception. The material of that chapter is published in Munda & Legrand (2014a).

Beyond the main analysis of the treatment effect, multicentre clinical trial data can be used to identify the centre characteristics or practice patterns that lead to variation in outcome between centres. This type of investigation, which primarily aims at improving the quality of patient care, is termed “treatment outcome research”. The use of the frailty model for treatment outcome research is illustrated in Legrand et al. (2006). Power considerations, in terms of number of centres and number of patients per centre, are discussed in Duchateau et al. (2002). A limitation of the frailty model in this context is that the frailty term is assumed to be constant over time, implying that centre differences persist throughout follow-up. Chapter 5 relaxes the time-constant heterogeneity assumption and proposes a new frailty model with a time-varying frailty term. The time-varying frailty model can be used to determine the extent of clustering over time. The particular application of Chapter 5 concerns a cancer clinical trial on chronic myeloid leukemia where heterogeneity between centres is shown to decrease after bone marrow transplantation. The material of that chapter is submitted for publication (Munda et al., 2014).

Part III – Spatial dependence

The dependence structure in the shared frailty model is in essence similar to the compound symmetry structure in the linear mixed effects model. That is, any two subjects within a cluster exhibit the same degree of association. While this is appropriate for clinical trial patients clustered within centres, other situations require more complex dependence structures. The malaria study conducted around the Gilgel-Gibe hydroelectric dam reservoir in Ethiopia (Yewhalaw et al., 2010, 2013) is one such study. The risk of malaria is more similar for children living in close proximity to each other than for children living far away from each other. In other words, dependence is a function of the distance between children (spatial dependence). The malaria data has already been analysed to assess the effect of the Gilgel-Gibe dam on malaria incidence by means of Poisson and frailty models using village as random effect/frailty term (Yewhalaw et al., 2013; Getachew et al., 2013). In Chapter 6, we provide new insights into these data by presenting a first step towards modelling dependence as a function of distance using a spatial frailty model. The work in this chapter is ongoing.

1.2 --- Elements of survival analysis

This section introduces some basic and general concepts in survival analysis used in the subsequent chapters. Further details can be found in textbooks, e.g. Collett (2003) and Klein & Moeschberger (2003), among many others. Section 1.3 introduces the concept of shared frailty, the main topic of this thesis.

To set the scene, consider the survival data in Table 1.1. The data show the time (in weeks) from remission to relapse, hereinafter referred to as “survival time”, in two groups of acute leukemia patients initially treated with corticosteroids. A total of 42 patients, in complete or partial corticosteroid-induced remission, were randomly assigned at remission to one of two maintenance therapies: 6-MP, an immunosuppressive drug, or placebo. Patients were paired by remission status (complete or partial), a fact that we ignore for the moment. Details on the study design can be found in Gehan & Freireich (2011).

Table 1.1: The remission duration data. The first two columns contain the patient identification number and the pair identification number. The third column gives the time (in weeks) from remission to relapse (“survival time”) and the fourth column is the event indicator (1 = relapse, 0 = censored). The fifth column specifies the treatment group (1 = 6-MP, 0 = placebo).

id	pair	time	status	group
1	1	1	1	0
2	1	10	1	1
3	2	22	1	0
4	2	7	1	1
5	3	3	1	0
6	3	32	0	1
7	4	12	1	0
8	4	23	1	1
9	5	8	1	0
10	5	22	1	1
11	6	17	1	0
12	6	6	1	1
13	7	2	1	0
14	7	16	1	1
15	8	11	1	0
16	8	34	0	1
17	9	8	1	0
18	9	32	0	1
19	10	12	1	0
20	10	25	0	1
21	11	2	1	0
22	11	11	0	1
23	12	5	1	0
24	12	20	0	1
25	13	4	1	0
26	13	19	0	1
27	14	15	1	0
28	14	6	1	1
29	15	8	1	0
30	15	17	0	1
31	16	23	1	0

32	16	35	0	1
33	17	5	1	0
34	17	6	1	1
35	18	11	1	0
36	18	13	1	1
37	19	4	1	0
38	19	9	0	1
39	20	1	1	0
40	20	6	0	1
41	21	8	1	0
42	21	10	0	1

1.2.1 Notations and terminology

The remission duration data is subject to right censoring. The survival time random variable, i.e. time from randomisation ($\approx$ time from remission) to relapse is denoted by T . The censoring time, i.e. time from randomisation to last contact in remission (study completion in this case), is denoted by C . Therefore, the random variables that we observe are $Y = \min(T, C)$ and $\Delta = I(T \leq C)$, i.e. the follow-up time and the event (relapse from remission) indicator. The realisations of (Y, Δ) for the remission duration data are given in columns 3 and 4 of Table 1.1. For example, patient 2 went out of remission at 10 weeks whereas observation 42 is censored at the same time. Though not directly observed, interest lies in T , the actual survival time. Besides the probability density function $f(\cdot)$ and the cumulative distribution function $F(\cdot)$, common ways of characterising T are:

- the survival function

$$S(t) = \Pr(T > t)$$

giving the probability that a patient continues to be in remission (“survives”) at time t ;

- the hazard function

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

giving the instantaneous rate at which relapses occur in time for patients still in remission;

- the cumulative hazard function

$$H(t) = \int_0^t h(v) \, dv$$

giving the total amount of risk accumulated from the start to the present (t).

We have the following relationships between $f(t)$, $F(t)$, $S(t)$, $h(t)$, and $H(t)$:

$$S(t) = 1 - F(t)$$

$$S(t) = -\frac{df}{dt}(t)$$

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

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

The density and survival functions of C will be denoted by $g(\cdot)$ and $G(\cdot)$, respectively.

1.2.2 Censoring

Coping with censored data requires working assumptions about the mechanism by which censoring occurs. A number of right censoring mechanisms, as well as other types of data incompleteness (e.g., interval censoring and truncation) are discussed in Lawless (2002, Chapter 2). The random right censorship model outlined above for the remission duration data (cf. Section 1.2.1) is adequate for most purposes. Further, it is typically required that the censoring mechanism is independent and non-informative. Under these conditions, a simple expression is obtained for the likelihood (see Section 1.2.3 below).

Independent censoring

Independent censoring is the requirement that the hazard rate of an at-risk subject coincides with the hazard rate in the surviving population, i.e.

$$\begin{aligned} & \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T < t + \Delta t \mid T \geq t, C \geq t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} \end{aligned}$$

This requirement essentially means that the uncensored subjects under follow-up must be representative of the surviving population; a condition that is satisfied when censoring occurs independently of the survival time (e.g., censoring due to calendar termination of the study). If there are covariates, then the independent censoring assumption is made conditional on the covariate information.

Non-informative censoring

If the censoring time distribution provides no information about the survival time distribution, then the censoring mechanism is said to be non-informative (for inference about the parameter(s) of interest).

1.2.3 Survival likelihood

With independent censoring, the joint distribution of T and C can be factored into the product of the marginals. Therefore,

- $(y, 1)$ contributes to the likelihood with the factor $f(y)G(y)$;
- $(y, 0)$ contributes to the likelihood with the factor $S(y)g(y)$.

Further, if the censoring time distribution does not contain information on the parameter(s) of the survival time distribution (non-informative censoring), then $g(y)$ and $G(y)$ can be dropped from the above contributions. In that case,

- the relevant information contained in an event data is that the event occurred at the observed time;
- the relevant information contained in a censored data is that the event time exceeds the censoring time.

Under these conditions, the survival likelihood for a sample of size N is

$$\begin{aligned} L &\propto \prod_{j=1}^N f(y_j)^{\delta_j} S(y_j)^{1-\delta_j} \\ &= \prod_{j=1}^N h(y_j)^{\delta_j} S(y_j) \end{aligned}$$

Note: independence assumption

As L is factored into the product of the individual data contributions, independence between observations is also required.

1.2.4 Kaplan-Meier survival curve

The Kaplan-Meier estimator is a non-parametric estimator of $S(t)$. Using non-parametric maximum likelihood estimation ideas, the Kaplan-Meier estimator treats T as a discrete random variable with probability mass at the observed event times only (Kalbfleisch & Prentice, 2002, Section 1.4.1). Accordingly, the Kaplan-Meier survival curve is a (right continuous) step function with jumps at the event times. When there is no censoring, the Kaplan-Meier survival curve reduces to the empirical survival function.

The survival function at $\tilde{y}_{(j)}$, the j^{th} ordered event time, equals

$$\begin{aligned} S(\tilde{y}_{(j)}) &= \Pr(T > \tilde{y}_{(j)}) \\ &= \Pr(T > \tilde{y}_{(j)} \mid T > \tilde{y}_{(j-1)}) S(\tilde{y}_{(j-1)}) \\ &= \prod_{\ell=1}^j \Pr(T > \tilde{y}_{(\ell)} \mid T > \tilde{y}_{(\ell-1)}) \end{aligned}$$

with $\tilde{y}_{(0)} := 0$. Under the assumption of independent censoring, the n_ℓ subjects at risk for the event at $\tilde{y}_{(\ell)}$ are representative of the surviving population at $\tilde{y}_{(\ell-1)}$ (since T is treated as discrete); cf. Section 1.2.2. Therefore, the conditional probability in the above formula can be estimated by the proportion of survivors in the sample at risk, i.e. $(n_\ell - d_\ell)/n_\ell$. Hence, by piecewise constant interpolation, the Kaplan-Meier estimator takes the form

$$\hat{S}(t) = \prod_{\ell: \tilde{y}_{(\ell)} \leq t} \frac{n_\ell - d_\ell}{n_\ell}$$

For a fixed t , the Kaplan-Meier estimator $\hat{S}(t)$ has an asymptotic normal distribution with mean $S(t)$. The variance of $\hat{S}(t)$ can be estimated by Greenwood's formula,

$$\widehat{\text{Var}}(\hat{S}(t)) = (\hat{S}(t))^2 \sum_{\ell: \tilde{y}_{(\ell)} \leq t} \frac{d_\ell}{n_\ell(n_\ell - d_\ell)}$$

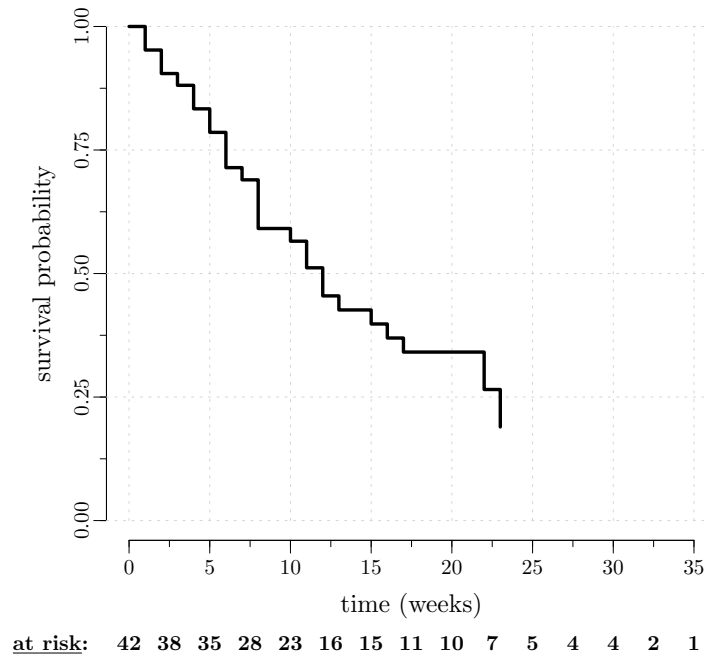


Figure 1.1: Kaplan-Meier survival curve in the remission duration data, with number of subjects at risk.

A plot of the Kaplan-Meier survival curve in the remission duration data is shown in Figure 1.1. The Kaplan-Meier survival curve crosses the 50% line at 12 weeks, the non-parametric estimate of the median survival time (i.e., the time beyond which the proportion in remission in the study population equals 0.5). The value at the largest relapse time, $\hat{S}(23) = 0.189$, is the proportion in remission at the end of the study. When different from zero (i.e. when the last observation is a censored data), the Kaplan-Meier survival curve is undefined from that point on.

1.2.5 Weibull proportional hazards model

A popular distribution to model the survival time T is the Weibull distribution, with the following characteristics:

$$\begin{aligned} f(t) &= \lambda \rho t^{\rho-1} \exp(-\lambda t^\rho) \\ F(t) &= 1 - \exp(-\lambda t^\rho) \\ S(t) &= \exp(-\lambda t^\rho) \\ h(t) &= \lambda \rho t^{\rho-1} \\ H(t) &= \lambda t^\rho \end{aligned}$$

for $\rho > 0$ and $\lambda > 0$.

A number of Weibull hazard functions are plotted in Figure 1.2. The Weibull hazard rate decreases, increases, or remains constant over time for $\rho < 1$, $\rho > 1$, or $\rho = 1$ (exponential distribution), respectively. That is, ρ is a shape parameter. In contrast, the parameter λ is a scale parameter taking the form of a multiplier in the hazard function.

Note: R syntax

The relation with the parametrisation used by `dweibull()` in R is as follows

$$\text{shape} = \rho \quad \text{and} \quad \text{scale} = \left(\frac{1}{\lambda}\right)^{1/\rho}$$

Weibull hazard rates with the same shape but different scales are proportional. Covariates that act on the Weibull scale parameter thus result in a proportional hazards model. The Weibull proportional hazards model for the remission duration data can be written as

$$h(t) = \begin{cases} \lambda \rho t^{\rho-1} & \text{in the placebo group} \\ (\lambda \exp(\beta)) \rho t^{\rho-1} & \text{in the 6-MP group} \end{cases}$$

with $\exp(\beta)$ a summary measure of the effect of 6-MP therapy on the relapse hazard rate (the exponential transformation ensures positivity without any parameter constraint).

By maximum likelihood estimation (cf. Section 1.2.3), we find $\hat{\rho} = 1.366$ (se = 0.201), $\hat{\lambda} = 0.046$ (se = 0.026) and $\hat{\beta} = -1.731$ (se = 0.413). The fact that $\hat{\beta}$ is significantly less than 0 (95% CI: $[-2.540, -0.921]$)

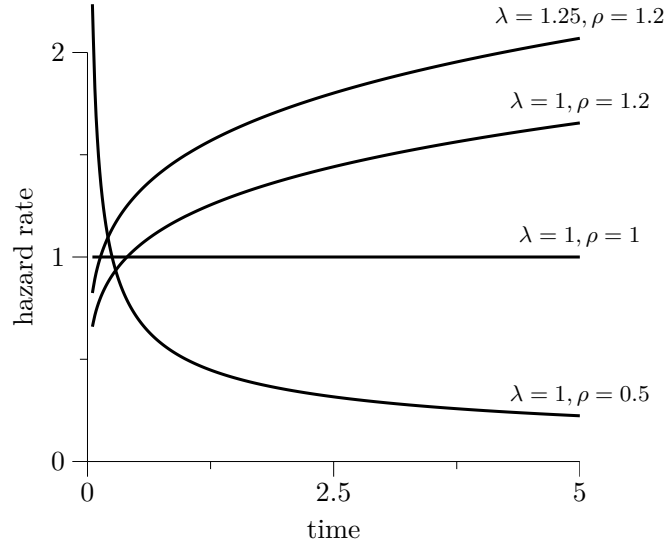


Figure 1.2: Weibull hazard functions.

indicates that 6-MP maintenance therapy reduces the risk of a relapse, and hence prolongs the duration of remission, in the study population. The estimated hazard ratio is $\exp(\hat{\beta}) = 0.177$ (95% CI: [0.079, 0.398]).

1.2.6 Cox model

Like the Weibull proportional hazards model (cf. Section 1.2.5), the Cox model specifies the way the explanatory variables act on the hazard rate, but, in contrast to the Weibull model, lets the time dependency of the hazard rate unspecified,

$$h(t) = h_0(t) \exp(\mathbf{x}'\boldsymbol{\beta})$$

with $h_0(\cdot)$ a non-specified baseline hazard function, $\mathbf{x} = (x_1 \dots x_p)'$ the vector of explanatory variables, and $\boldsymbol{\beta} = (\beta_1 \dots \beta_p)'$ the vector of regression parameters. Owing to its semi-parametric nature, the Cox model has become routine in survival analysis. A nice and concise review can be found in Katz & Hauck (1993).

The ratio of the hazard functions for two subjects with different

covariate information, say $\mathbf{x}_{j_1}$ and $\mathbf{x}_{j_2}$, is

$$\frac{h_{j_1}(t)}{h_{j_2}(t)} = \exp\left((\mathbf{x}_{j_1} - \mathbf{x}_{j_2})'\boldsymbol{\beta}\right)$$

A one-unit change in one of the explanatory variables, while all other are kept fixed, results in a proportional change in the hazard function (proportional hazards assumption). The parameter β_k in the Cox model is thus interpreted as a conditional log hazard ratio. In essence, the proportionality assumption is the requirement that each β_k be constant over time.

As $h_0(\cdot)$ is left unspecified, we cannot make use of the survival likelihood from Section 1.2.3 to fit the Cox model. An estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ can be obtained by maximising a partial likelihood instead. Assuming no ties in the event times, the partial likelihood of the Cox model is given by (see, e.g., Klein & Moeschberger, 2003, Section 8.3)

$$L_p(\boldsymbol{\beta}; \mathbf{z}) = \prod_{j=1}^N \left(\frac{\exp(\mathbf{x}'_{j_j}\boldsymbol{\beta})}{\sum_{\ell \in R(y_j)} \exp(\mathbf{x}'_{\ell}\boldsymbol{\beta})} \right)^{\delta_j}$$

with $R(y_j)$ the risk set at time y_j containing all subjects still under observation just prior to y_j . The partial likelihood of the Cox model can be derived as a profile likelihood obtained by maximising, for fixed $\boldsymbol{\beta}$, the survival likelihood with respect to the discretised version of the baseline hazard function (Klein & Moeschberger, 2003, Section 8.3). In the case of ties, approximations of the partial likelihood (e.g., Breslow, Efron) can be used (Klein & Moeschberger, 2003, Section 8.4). Approximate standard errors of the maximum partial likelihood estimates are given by the square roots of the diagonal entries of the negative inverse Hessian matrix of $\ell_p(\cdot; \mathbf{z}) := \log(L_p(\cdot; \mathbf{z}))$ evaluated at the maximum. Even though $L_p(\cdot; \mathbf{z})$ is not a genuine likelihood, it has been shown that consistency and asymptotic normality properties for the estimator of $\boldsymbol{\beta}$ are preserved (Gill, 1984).

In the remission duration data, we find (using the Breslow approximation to handle ties) $\exp(\hat{\boldsymbol{\beta}}) = 0.221$ (95% CI: [0.099, 0.493]).

1.3

Frailty model

In the remission duration data (Table 1.1), the response variable (time from remission to relapse) is observed for matched pairs of patients.

Because of the matching, there is likely to be dependence in the data. Matched patients tend to be more alike than unmatched patients, thus indicating positive association within pairs. This kind of data is referred to as “clustered data”. In the remission duration data, clusters are of size 2. Multicentre clinical trial data, where patients are clustered within centres, typically involve larger and varying cluster sizes. The shared frailty model has been introduced to cope with clustered survival data.

For the j^{th} subject ($j = 1, \dots, n_i$) of the i^{th} cluster ($i = 1, \dots, s$), the (shared) frailty model is defined as

$$h_{ij}(t) = h_0(t)u_i \exp(\mathbf{x}'_{ij}\boldsymbol{\beta})$$

where u_i denotes the multiplicative effect of cluster i . Thus, u_i represents the deviation of cluster i from the overall baseline risk. In the frailty model, the cluster effect is random. That is, the u_i 's are treated as the actual (unobserved) values of a random variable U , the frailty term, and $h_{ij}(\cdot)$ is interpreted as a conditional hazard given $U = u_i$.

The frailty term can be interpreted as an unobserved covariate common to, or shared by, individuals in a cluster. As U takes the same value for each subject in a cluster, the frailty term generates (positive) association between event times in a cluster. The frailty model is based on a conditional independence assumption. Given the frailty and covariates, observations are assumed to be independent. That is, if the common risk (frailty) term was known, observations would have been independent.

Another way to look at the frailty model is in terms of heterogeneity. Cluster heterogeneity, i.e. variation in susceptibility to the event from cluster to cluster, reflects different levels of risk across clusters and is modelled by the random variation of the frailty term. To put it differently, the presence of cluster heterogeneity indicates the presence of unknown risk factors varying from cluster to cluster. No cluster heterogeneity (U degenerated) implies independent data given the covariates, and vice versa.

To model the frailty term, a distribution, called the frailty distribution, is postulated. In principle, any distribution on the positive half-line can play the role of the frailty distribution. To keep the mathematics tractable though, distributions with simple Laplace transforms are preferred. The Laplace transform of the frailty term, i.e. $\mathcal{L}(x) = E(\exp(-Ux))$, characterises the frailty distribution uniquely

(provided that it exists) and plays an important role in frailty modelling. In the thesis appendix (Appendix 1), we describe, under the parametrisation that we shall use throughout the thesis[†], the main characteristics of six important frailty distributions: the gamma (Gam), the inverse Gaussian (IG), the positive stable (PS), the log-normal (LN), the power variance function (PVF), and the compound Poisson (CP). The thesis appendix only collects useful formulas for ease of reference. The general characteristics of these frailty distributions are studied in detail in Hougaard (2000, Chapter 7) and in Duchateau & Janssen (2008, Chapter 4). In practice, the gamma distribution is most often used to model the frailty term.

The concept of shared frailty was introduced in Clayton (1978). Suggestions for introductory reading on this topic are Liang et al. (1995) and Hougaard (1995). A recent overview of frailty modelling is given in Govindarajulu et al. (2011). The frailty model methodology is thoroughly presented in textbooks; see Hougaard (2000, Chapters 7–11), Duchateau & Janssen (2008), and Wienke (2010).

The latent nature of the frailty term makes it more difficult to fit frailty models. Inference for frailty models is discussed in Chapter 2. In practice, the most common way to fit the (gamma) frailty model is by means of `coxph()` in R (part of the `survival` library). For a detailed description of the proper use of `coxph()` for frailty models, see Therneau & Grambsch (2000, Chapter 9).

In the remission duration data, we find (using `coxph()` with the gamma frailty distribution and the Breslow approximation to handle ties) $\exp(\hat{\beta}) = 0.221$ (95% CI: [0.099, 0.493]). In this case, the frailty model produces the same result as the Cox model. This suggests that the frailty term does not contribute to the model. Further indication is provided by the estimated variance of the frailty term, $\hat{\theta} = 4.77 \times 10^{-8}$, which is virtually zero.

— **Note:** $h_{ij}(\cdot) := h(\cdot | \mathbf{x}_{ij}, u_i)$ —

The notation $h_{ij}(\cdot)$ is used throughout the thesis as shorthand for $h(\cdot | \mathbf{x}_{ij}, u_i)$, the conditional hazard for subject j in cluster i .

[†]The densities are already rescaled to ensure identifiability of the parameters in the frailty model (Hougaard, 2000, page 221), similar to the zero-mean constraint for the random effects in the linear mixed model.

Part I

The frailty distribution

A unified framework for fitting the frailty model with different frailty distributions

There are at least three things that need to be put into consideration when it comes to choosing a method of estimation to fit the frailty model. The first one is a recurrent question in survival analysis: shall we specify a parametric form for the baseline hazard (parametric approach) or shall we leave it unspecified (semi-parametric approach)? A comparison of parametric and semi-parametric survival models can be found in Nardi & Schemper (2003). This chapter focuses on the parametric approach. However, the results of this chapter are useful as well in the implementation of the EM algorithm within the semi-parametric setting. The second point concerns the choice of the frailty distribution. In this chapter, we show that many frailty distributions, including those introduced in the thesis appendix (Appendix 1), can be handled in a uniform way. Third, from a purely pragmatic point of view, software availability weighs in the balance. The method of this chapter, based on the maximum likelihood principle, only requires a numerical optimisation procedure. We have implemented the method in the R library `parfm`.

In Section 2.1, different likelihoods for frailty models are defined. The parametric frailty model is fitted based on the marginal likelihood. In Section 2.2, it is shown that a generic form of the marginal likelihood can be written by means of the derivatives of the Laplace transform of the frailty term. The parametric approach is outlined in Section 2.3. Explicit derivative formulas are given in Section 2.4. Section 2.5 illustrates the use of `parfm`.

For a detailed review of model estimation techniques in the semi-parametric setting, see Cortiñas Abrahantes et al. (2007). An overview of the available software in that setting is provided in Hirsch & Wienke

(2012).

The content of this chapter extends the results published in Munda et al. (2012).

2.1

Conditional, complete data, and marginal likelihoods

Observed data

Let $\mathbf{z}_i$ be the vector that contains the relevant information from cluster i ($i = 1, \dots, s$), i.e.

$$\mathbf{z}_i = \left\{ (y_{ij}, \delta_{ij}, \mathbf{x}_{ij}) \mid j = 1, \dots, n_i \right\}$$

where y_{ij} is the time to event or censoring, whichever comes first, δ_{ij} indicates whether an observation corresponds to an event ($\delta_{ij} = 1$) or is censored ($\delta_{ij} = 0$), and $\mathbf{x}_{ij}$ is the vector with the measured covariates for subject j of cluster i . The frailty of cluster i , u_i , is not observed.

Conditional likelihood

Given the latent data information $\mathbf{u} = (u_1 \dots u_s)'$ and the covariate information, event times are treated as independent (conditional independence assumption). The conditional likelihood of the data $\mathbf{z} = \{\mathbf{z}_i \mid i = 1, \dots, s\}$ is thus written as (cf. Section 1.2.3)

$$\begin{aligned} L_{\text{cond}}(h_0(\cdot), \boldsymbol{\beta}; \mathbf{z} \mid \mathbf{u}) &= \prod_{i=1}^s \prod_{j=1}^{n_i} \left(h_{ij}(y_{ij}) \right)^{\delta_{ij}} S_{ij}(y_{ij}) \\ &= \prod_{i=1}^s \prod_{j=1}^{n_i} \left(h_0(y_{ij}) u_i \exp(\mathbf{x}_{ij}' \boldsymbol{\beta}) \right)^{\delta_{ij}} \exp \left(-H_0(y_{ij}) u_i \exp(\mathbf{x}_{ij}' \boldsymbol{\beta}) \right) \end{aligned}$$

Complete data likelihood

The complete data likelihood treats the frailties as if they were observed. It therefore follows from the joint likelihood of $\mathbf{z}$ and $\mathbf{u}$,

$$L_{\text{full}}(h_0(\cdot), \boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{z}, \mathbf{u}) = L_{\text{cond}}(h_0(\cdot), \boldsymbol{\beta}; \mathbf{z} \mid \mathbf{u}) \times f(\mathbf{u})$$

with $f(\cdot) := f(\cdot; \boldsymbol{\theta})$ the frailty density with parameter vector $\boldsymbol{\theta}$. If the u_i 's are independent and identically distributed, which we assume in the following, then $f(\mathbf{u})$ can be written as $\prod_{i=1}^s f(u_i)$.

Marginal likelihood

To arrive at a likelihood not depending on the unobservables, the u_i 's have to be integrated out to form the marginal likelihood (also called the observed likelihood),

$$L_{\text{marg}}(h_0(\cdot), \boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{z}) = \int_0^\infty \cdots \int_0^\infty L_{\text{full}}(h_0(\cdot), \boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{z}, \mathbf{u}) du_1 \cdots du_s \quad (2.1)$$

2.2 Generic form of the marginal likelihood

It is convenient to rewrite the marginal likelihood (2.1) in a generic form in terms of $\mathcal{L}(\cdot)$, the Laplace transform of the frailty term U , i.e. in terms of

$$\mathcal{L}(x) = \mathbb{E}(\exp(-Ux)) = \int_0^\infty \exp(-ux) f(u) du$$

Let $D_i = \sum_{j=1}^{n_i} \delta_{ij}$ denote the number of events in cluster i , and $\mathcal{L}^{(q)}(\cdot)$ be the q^{th} derivative of $\mathcal{L}(\cdot)$, $q \geq 0$, i.e.

$$\mathcal{L}^{(q)}(x) = (-1)^q \int_0^\infty u^q \exp(-ux) f(u) du$$

We have

$$\begin{aligned} L_{\text{marg}}(h_0(\cdot), \boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{z}) &= \prod_{i=1}^s \left[\left(\prod_{j=1}^{n_i} \left(h_0(y_{ij}) \exp(\mathbf{x}'_{ij} \boldsymbol{\beta}) \right)^{\delta_{ij}} \right) \right. \\ &\quad \left. \times (-1)^{D_i} \mathcal{L}^{(D_i)} \left(\sum_{j=1}^{n_i} H_0(y_{ij}) \exp(\mathbf{x}'_{ij} \boldsymbol{\beta}) \right) \right] \end{aligned}$$

Taking the logarithm, the marginal log-likelihood is obtained

$$\begin{aligned} \ell_{\text{marg}}(h_0(\cdot), \beta, \theta; \mathbf{z}) = \sum_{i=1}^s \left[\left(\sum_{j=1}^{n_i} \delta_{ij} \left(\log(h_0(y_{ij})) + \mathbf{x}'_{ij} \beta \right) \right) \right. \\ \left. + \log \left((-1)^{D_i} \mathcal{L}^{(D_i)} \left(\sum_{j=1}^{n_i} H_0(y_{ij}) \exp(\mathbf{x}'_{ij} \beta) \right) \right) \right] \end{aligned}$$

2.3

Parametric framework

In the parametric framework, we adopt a baseline hazard function $h_0(\cdot)$ that is completely specified, except for a number of unknown parameters that are estimated together with β and θ . Common choices for $h_0(\cdot)$ are collected in Table 2.1. This results in a fully parametric marginal log-likelihood that can be maximised by means of an optimisation routine (e.g., a Newton-type algorithm). This task is most easily done when $\mathcal{L}^{(q)}(\cdot)$ exists in closed form. This is the case for the gamma, the inverse Gaussian, the positive stable, the power variance function, and the compound Poisson frailty distributions (cf. Sections 2.4.1–2.4.5). The Laplace transform of a log-normal frailty term, on the other hand, does not exist in closed form. In that case, we can approximate the marginal log-likelihood by means of the Laplace approximation of integrals (cf. Section 2.4.6).

2.4

Derivative formulas

2.4.1 Gamma frailties

If $U \sim \text{Gam}(\theta)$, then the Laplace transform is

$$\mathcal{L}(x) = (1 + \theta x)^{-1/\theta}, \quad x \geq 0$$

and, for $q \geq 1$, it is easy to see that

$$\mathcal{L}^{(q)}(x) = (-1)^q (1 + \theta x)^{-q} \left[\prod_{\ell=0}^{q-1} (1 + \ell\theta) \right] \mathcal{L}(x)$$

Table 2.1: Parametric baseline hazard and cumulative baseline hazard functions for selected event time distributions.

distribution	param.	$h_0(t)$	$H_0(t)$
exponential	$\lambda > 0$	λ	λt
Weibull	$\lambda > 0$ $\rho > 0$	$\lambda \rho t^{\rho-1}$	λt^ρ
Gompertz	$\lambda > 0$ $\gamma > 0$	$\lambda \exp(\gamma t)$	$\frac{\lambda}{\gamma}(\exp(\gamma t) - 1)$
log-logistic	$\alpha \in \mathbb{R}$ $\kappa > 0$	$\frac{\exp(\alpha) \kappa t^{\kappa-1}}{1 + \exp(\alpha) t^\kappa}$	$\log(1 + \exp(\alpha) t^\kappa)$
log-normal [†]	$\mu \in \mathbb{R}$ $\sigma > 0$	$\frac{\phi(z(t; \mu, \sigma))}{\sigma t [1 - \Phi(z(t; \mu, \sigma))]}$	$-\log [1 - \Phi(z(t; \mu, \sigma))]$

[†] $z(t; \mu, \sigma) := (\log(t) - \mu)/\sigma$; $\phi(\cdot)$ and $\Phi(\cdot)$ denote the density and the cumulative distribution functions of a standard normal random variable.

2.4.2 Inverse Gaussian frailties

If $U \sim \text{IG}(\theta)$, then the Laplace transform is

$$\mathcal{L}(x) = \exp \left\{ \frac{1}{\theta} \left(1 - \sqrt{1 + 2\theta x} \right) \right\}, \quad x \geq 0$$

and, for $q \geq 1$, we have (Munda et al., 2012)

$$\mathcal{L}^{(q)}(x) = (-1)^q (2\theta x + 1)^{-q/2} \frac{K_{q-(1/2)} \left(\sqrt{2\theta^{-1}(x + 1/2\theta)} \right)}{K_{1/2} \left(\sqrt{2\theta^{-1}(x + 1/2\theta)} \right)} \mathcal{L}(x)$$

with $K_\gamma(\cdot)$ the modified Bessel function of the second kind,

$$K_\gamma(\omega) = \frac{1}{2} \int_0^\infty t^{\gamma-1} \exp \left\{ -\frac{\omega}{2} \left(t + \frac{1}{t} \right) \right\} dt, \quad \gamma \in \mathbb{R}, \omega > 0$$

Alternatively, as $\text{IG}(\theta) = \text{PVF}(1, \theta, 1/2)$ (cf. Section 1.5 of the thesis appendix), the formula of $\mathcal{L}^{(q)}(x)$ from Section 2.4.4 can be used.

2.4.3 Positive stable frailties

If $U \sim \text{PS}(\theta)$, then the Laplace transform is

$$\mathcal{L}(x) = \exp \left(-x^{1-\theta} \right), \quad x \geq 0$$

and, for $q \geq 1$, it is found in Wang et al. (1995) that (cf. Lemma 3.1 in that paper, which we write under a different parametrisation)

$$\mathcal{L}^{(q)}(x) = (-1)^q \left((1 - \theta)x^{-\theta} \right)^q \left[\sum_{\ell=0}^{q-1} \Omega_{q,q-\ell}(\theta) (1 - \theta)^{-\ell} x^{-\ell(1-\theta)} \right] \mathcal{L}(x)$$

where the $\Omega_{q,m}(\theta)$'s are polynomials of degree $q-m$ in θ given recursively by (cf. Appendix 2.A)

$$\Omega_{q,m}(\theta) = \begin{cases} \frac{\Gamma(q - (1 - \theta))}{\Gamma(\theta)} & \text{if } m = 1 \\ \Omega_{q-1,m-1}(\theta) + \Omega_{q-1,m}(\theta)[(q-1) - m(1-\theta)] & \text{if } m = 2, \dots, q-1 \\ 1 & \text{if } m = q \end{cases}$$

2.4.4 Power variance function frailties

If $U \sim \text{PVF}(\mu, \theta, \nu)$, then the Laplace transform is

$$\mathcal{L}(x) = \exp \left\{ \frac{\nu}{\theta(1-\nu)} \left[1 - \left(1 + \frac{\theta\mu x}{\nu} \right)^{1-\nu} \right] \right\}, \quad x \geq 0$$

and, for $q \geq 1$, it is found in Hougaard (2000, Section A.3.4) that

$$\mathcal{L}^{(q)}(x) = (-1)^q \left[\sum_{\ell=1}^q \Omega_{q,\ell}(\nu) \mu^{\ell(1-\nu)} \left(\frac{\nu}{\theta} \right)^{\ell\nu} \left(\frac{\nu}{\theta\mu} + x \right)^{\ell(1-\nu)-q} \right] \mathcal{L}(x)$$

where the $\Omega_{q,m}(\nu)$'s are the same polynomials as above, evaluated at ν .

2.4.5 Compound Poisson frailties[†]

If $U \sim \text{CP}(\mu, \theta, \nu)$, then the Laplace transform has the same analytical form as for the PVF distribution (cf. Section 1.6 of the thesis appendix); therefore so are the derivatives.

[†]We mean the compound Poisson distribution generated by gamma random variables.

2.4.6 Log-normal frailties

If $U \sim \text{LN}(\theta)$, then the Laplace transform does not exist in closed form. Consequently

$$\begin{aligned}\mathcal{L}^{(q)}(x) &= (-1)^q \int_0^\infty u^q \exp(-ux) f(u) \, du \\ &= (-1)^q \frac{1}{\sqrt{2\pi\theta}} \int_0^\infty u^q \exp(-ux) \frac{1}{u} \exp\left(-\frac{1}{2\theta}(\log(u))^2\right) \, du\end{aligned}$$

needs to be approximated ($x \geq 0$). By using the change of variable $w = \log(u)$, we have

$$\begin{aligned}\mathcal{L}^{(q)}(x) &= (-1)^q \frac{1}{\sqrt{2\pi\theta}} \int_{-\infty}^\infty (\exp(w))^q \exp(-\exp(w)x) \exp\left(-\frac{w^2}{2\theta}\right) \, dw \\ &= (-1)^q \frac{1}{\sqrt{2\pi\theta}} \int_{-\infty}^\infty \exp\left\{qw - \exp(w)x - \frac{w^2}{2\theta}\right\} \, dw\end{aligned}$$

We now show how to approximate this by means of the Laplace approximation of integrals.

Let

$$\begin{aligned}g(w; x, \theta) &:= -qw + \exp(w)x + \frac{w^2}{2\theta} \\ g^{(1)}(w; x, \theta) &:= \frac{dg}{dw}(w; x, \theta) = -q + \exp(w)x + \frac{w}{\theta} \\ g^{(2)}(w; x, \theta) &:= \frac{d^2g}{dw^2}(w; x, \theta) = \exp(w)x + \frac{1}{\theta} > 0\end{aligned}$$

The approximation consists of replacing $g(\cdot)$ by the first three terms of its Taylor series expansion around some $\tilde{w}$,

$$g(w; x, \theta) \approx g(\tilde{w}; x, \theta) + (w - \tilde{w})g^{(1)}(\tilde{w}; x, \theta) + \frac{(w - \tilde{w})^2}{2}g^{(2)}(\tilde{w}; x, \theta)$$

The value of $\tilde{w}$ is chosen such that $g^{(1)}(\tilde{w}; x, \theta) = 0$, so that $\mathcal{L}^{(q)}(x)$ can be approximated by

$$\begin{aligned}\mathcal{L}^{(q)}(x) &\approx (-1)^q \frac{1}{\sqrt{2\pi\theta}} \exp\{-g(\tilde{w}; x, \theta)\} \\ &\quad \times \int_{-\infty}^\infty \exp\left\{-\frac{(w - \tilde{w})^2}{2}g^{(2)}(\tilde{w}; x, \theta)\right\} \, dw \\ &= (-1)^q \frac{1}{\sqrt{\theta}} \exp\{-g(\tilde{w}; x, \theta)\} [g^{(2)}(\tilde{w}; x, \theta)]^{-1/2}\end{aligned}$$

where the last line follows by recognising the kernel of a normal density with mean $\tilde{w}$ and variance $1/g^{(2)}(\tilde{w}; x, \theta)$. This is known as the Laplace approximation. The underlying idea is that the main contribution to the integral comes from where $g(\cdot)$ is close to its minimum. We refer to Goutis & Casella (1999) for further motivation and explanation of this kind of approximation.

Note: prediction of the frailty term

The explicit formulas for $\mathcal{L}^{(q)}(\cdot)$ are also useful to predict the value taken by the frailty term in a particular cluster. Indeed, the conditional expectation of the frailty term, given the observed data from cluster i and the parameters, can be written as

$$\begin{aligned} & \mathbb{E}(U \mid \mathbf{z}_i; h_0(\cdot), \boldsymbol{\beta}, \boldsymbol{\theta}) \\ &= \frac{\mathbb{E}\left(U^{D_i+1} \exp\left(-U \sum_{j=1}^{n_i} H_0(y_{ij}) \exp(\mathbf{x}'_{ij}\boldsymbol{\beta})\right)\right)}{\mathbb{E}\left(U^{D_i} \exp\left(-U \sum_{j=1}^{n_i} H_0(y_{ij}) \exp(\mathbf{x}'_{ij}\boldsymbol{\beta})\right)\right)} \\ &= -\frac{\mathcal{L}^{(D_i+1)}\left(\sum_{j=1}^{n_i} H_0(y_{ij}) \exp(\mathbf{x}'_{ij}\boldsymbol{\beta})\right)}{\mathcal{L}^{(D_i)}\left(\sum_{j=1}^{n_i} H_0(y_{ij}) \exp(\mathbf{x}'_{ij}\boldsymbol{\beta})\right)} \end{aligned}$$

which follows from applying Bayes' formula to $f(u \mid \mathbf{z}_i; h_0(\cdot), \boldsymbol{\beta}, \boldsymbol{\theta})$. In particular, predictions of the frailties are needed in the E-step of the EM algorithm (Nielsen et al., 1992).

The EM algorithm makes use of the complete data likelihood (cf. Section 2.1) to fit the frailty model in the semi-parametric setting. The algorithm iterates between an expectation step (or E-step) and a maximisation step (or M-step). In the E-step, the unobserved frailties are replaced by their conditional expectations given the observed data and the current parameter estimates. In the M-step, new parameter estimates are found by maximising the complete data likelihood, using the conditional expectations as offset terms.

2.5

parfm[†]

The R function `parfm()`, part of the `parfm` library, builds the marginal log-likelihood and calls `optim()`, part of the `stats` library, to perform the optimisation. The basic usage of `parfm()` (version 2.5.6) is as follows:

Usage

```
parfm(formula, cluster, dist, frailty, data)
```

Arguments

<code>formula</code>	a survival formula object
<code>cluster</code>	character string indicating the cluster variable
<code>dist</code>	character string indicating the baseline event time distribution; one of "exponential", "weibull", "gompertz", "lognormal", "loglogistic"
<code>frailty</code>	character string indicating the frailty distribution; one of "gamma", "ingau", "possta", "lognormal"
<code>data</code>	data frame containing the variables named in "formula" and "cluster"

To illustrate `parfm`, we consider a litter-matched experiment studying the effect of a drug on the time until the appearance of a tumour in rats (Mantel et al., 1977). Three (female) rats were chosen from each of 50 litters and followed for tumour incidence. Death from other causes was considered as a censoring event (73%). In each litter, one rat was selected at random and given the drug while the other two rats serve as controls. In R, the data can be loaded using

```
> data(rats, package="survival")
> rats$time <- rats$time * 0.0328549 # days to months
> head(rats, n=10)
  litter rx    time status
1      1  1 3.318345      0
2      1  0 1.609890      1
3      1  0 3.416910      0
4      2  1 3.416910      0
5      2  0 3.351200      0
6      2  0 3.416910      0
7      3  1 3.416910      0
8      3  0 3.416910      0
9      3  0 3.416910      0
10     4  1 2.529827      0
```

[†]Implementation of `parfm` is joint work with Federico Rotolo.

Below are the results of running `parfm()` with different frailty distributions on the rats data, using the Weibull distribution to model the baseline hazard.

Gamma frailties

```
> parfm(formula=Surv(time, status) ~ rx,
+       cluster="litter", dist="weibull", frailty="gamma",
+       data=rats)

Frailty distribution: gamma
Baseline hazard distribution: Weibull
Loglikelihood: -104.846
```

	ESTIMATE	SE	p-val
theta	0.489	0.469	
rho	3.929	0.569	
lambda	0.002	0.002	
rx	0.907	0.322	0.005 **

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Inverse Gaussian frailties

```
> parfm(formula=Surv(time, status) ~ rx,
+       cluster="litter", dist="weibull", frailty="ingau",
+       data=rats)

Frailty distribution: inverse Gaussian
Baseline hazard distribution: Weibull
Loglikelihood: -104.916
```

	ESTIMATE	SE	p-val
theta	0.541	0.647	
rho	3.931	0.572	
lambda	0.002	0.002	
rx	0.911	0.323	0.005 **

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Positive stable frailties

```
> # 'iniFpar=' sets the initial value of the frailty parameter (nu)
> parfm(formula=Surv(time, status) ~ rx,
+       cluster="litter", dist="weibull", frailty="possta",
+       data=rats, iniFpar=0.4)

Frailty distribution: positive stable
Baseline hazard distribution: Weibull
Loglikelihood: -104.947
```

	ESTIMATE	SE	p-val
--	----------	----	-------

```

nu      0.094      0.095
rho     4.103      0.627
lambda  0.002      0.001
rx       0.944      0.327 0.004 **

```

```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Log-normal frailties

```

> parfm(formula=Surv(time, status) ~ rx,
+       cluster="litter", dist="weibull", frailty="lognormal",
+       data=rats)

```

```

Frailty distribution: lognormal
Baseline hazard distribution: Weibull
Loglikelihood: -104.599

```

	ESTIMATE	SE	p-val
sigma2	0.575	0.489	
rho	3.963	0.576	
lambda	0.002	0.001	
rx	0.916	0.325	0.005 **

```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

All models lead to a significant effect of the drug. Additionally, the treatment effect appears not to depend on the frailty distribution assumption, anticipating the properties of robustness discussed in Chapter 4.

In Klein & Moeschberger (2003, Section 13.3, Example 1.13) the results are given for the semi-parametric gamma frailty model fitted with the EM algorithm: $\hat{\beta} = 0.904$ (se = 0.323) and $\hat{\theta} = 0.472$ (se = 0.462). These results are very close to the results obtained for the Weibull gamma frailty model fitted with **parfm**.

Alternatively, the Weibull gamma frailty model can be fitted by means of the **frailtypack** library (Rondeau et al., 2012):

frailtypack fit

```

> library(frailtypack)
> frailtyPenal(Surv(time, status) ~ rx + cluster(litter),
+             Frailty=TRUE, hazard="Weibull", data=rats)

```

```

Be patient. The program is computing ...

```

```

The program took 0.77 seconds

```

```

Call:

```

```

frailtyPenal(formula = Surv(time, status) ~ rx + cluster(litter),
             data = rats, Frailty = TRUE, hazard = "Weibull")

```

Shared Gamma Frailty model parameter estimates
using a Parametrical approach for the hazard function

```
coef exp(coef) SE coef (H)      z      p
rx 0.90751    2.47814    0.322339 2.81539 0.0048718
```

Frailty parameter, Theta: 0.488854 (SE (H): 0.469033)

marginal log-likelihood = -241.47
AIC = Aikaike information Criterion = 1.63648

The expression of the Aikaike Criterion is:

```
'AIC = (1/n)[np - l(.)]'
```

Scale for the weibull hazard function is : 140.91
Shape for the weibull hazard function is : 3.93

The expression of the Weibull hazard function is:

```
'lambda(t) = (shape.(t^(shape-1)))/(scale^shape)'
```

The expression of the Weibull survival function is:

```
'S(t) = exp[- (t/scale)^shape]'
```

```
n= 150
n events= 40  n groups= 50
number of iterations: 12
```

Appendix 2.A

The $\Omega_{q,m}(\cdot)$'s polynomials

The $\Omega_{q,m}(\cdot)$'s that appear in $\mathcal{L}^{(q)}(\cdot)$ for the positive stable, the power variance function, and the compound Poisson frailty distributions are polynomials of degree $q - m$ in α given recursively by

$$\Omega_{q,m}(\alpha) = \begin{cases} \frac{\Gamma(q - (1 - \alpha))}{\Gamma(\alpha)} & \text{if } m = 1 \\ \Omega_{q-1,m-1}(\alpha) + \Omega_{q-1,m}(\alpha)[(q-1) - m(1 - \alpha)] & \text{if } m = 2, \dots, q-1 \\ 1 & \text{if } m = q \end{cases}$$

for $\alpha \in (0, 1)$.

The first few polynomials are

	$m = 1$	$m = 2$	$m = 3$	$m = 4$
$q = 1$	1			
$q = 2$	α	1		
$q = 3$	$\alpha(\alpha + 1)$	3α	1	
$q = 4$	$\alpha(\alpha + 1)(\alpha + 2)$	$7\alpha^2 + 4\alpha$	6α	1

An R function to calculate the $\Omega_{q,m}(\alpha)$'s for $q = 1, \dots, Q$ is

```
Omega <- function(Q, alpha)
{ # Q = order of the derivative > 0, alpha in (0, 1)
  Omega <- matrix(NA, nrow=Q, ncol=Q, dimnames=list(q=1:Q, m=1:Q))
  diag(Omega) <- 1
  if(Q < 2)
    return(Omega)
  Omega[q=2:Q, m=1] <- cumprod(alpha + 0:(Q - 2))
  if(Q < 3)
    return(Omega)
  for(m in 2:(Q - 1))
    for(q in (m + 1):Q)
      Omega[q, m] <- Omega[q - 1, m - 1] +
        Omega[q - 1, m] * ((q - 1) - m * (1 - alpha))
  return(Omega)
}
```

For example,

```

> Omega(Q=4, alpha=0.1)
      m
q      1      2      3      4
1 1.000    NA    NA    NA
2 0.100 1.00    NA    NA
3 0.110 0.30 1.0    NA
4 0.231 0.47 0.6    1

```

Appendix 2.B

Derivative formulas in R

Gamma frailties

```

Lq_Gam <- function(q, theta, x)
{ # q >= 0, theta > 0, x >= 0
  if(q == 0){
    (1 + theta * x)^(-1 / theta)
  } else{
    (-1)^q * (1 + theta * x)^(-q) * prod(1 + 0:(q - 1) * theta) *
      Lq_Gam(q=0, theta=theta, x=x)
  }
}

```

Inverse Gaussian frailties

```

Lq_IG <- function(q, theta, x)
{ # q >= 0, theta > 0, x >= 0
  if(q == 0){
    exp((1 - sqrt(1 + 2 * theta * x)) / theta)
  } else{
    z <- theta^(-0.5) * sqrt(2 * x + 1 / theta)
    (-1)^q * (2 * theta * x + 1)^(-q / 2) *
      besselK(x=z, nu=q - 0.5) / (sqrt(0.5 * pi / z) * exp(-z)) *
      Lq_IG(q=0, theta=theta, x=x)
  }
}

```

Positive stable frailties

```

Lq_PS <- function(q, theta, Omega, x)
{ # q >= 0, theta in (0, 1), x >= 0
  if(q == 0){
    exp(-x^(1 - theta))
  } else{
    Sum <- 0
    for(l in 0:(q - 1)){
      Sum <- Sum +
        Omega[q, q - l] * (1 - theta)^(-l) * x^(-l * (1 - theta))
    }
  }
}

```

```

    }
    (-1)^q * ((1 - theta) * x^(-theta))^q * Sum *
    Lq_PS(q=0, theta=theta, x=x)
  }
}

```

Power variance function & compound Poisson frailties

```

Lq_PVF <- function(q, mu, theta, nu, Omega, x)
{ # q >= 0, mu > 0, theta > 0, nu in (0, 1) (PVF) or nu > 1 (CP),
  # x >= 0
  if(q == 0){
    exp(nu / (theta * (1 - nu)) *
        (1 - (1 + theta * mu * x / nu)^(1 - nu)))
  } else{
    Sum <- 0
    for(l in 1:q){
      Sum <- Sum +
        Omega[q, l] * mu^(l * (1 - nu)) * (nu / theta)^(l * nu) *
        (nu / (theta * mu) + x)^(l * (1 - nu) - q)
    }
    (-1)^q * Sum * Lq_PVF(q=0, mu=mu, theta=theta, nu=nu, x=x)
  }
}

```

Log-normal frailties

```

Lq_LN <- function(q, theta, x)
{ # q >= 0, theta > 0, x >= 0
  g <- function(w, q, x, theta)
    -q * w + exp(w) * x + 0.5 * w^2 / theta

  options(warn=-1)
  wTilde <- nlm(f=g, p=0, q=q, theta=theta, x=x)$estimate
  options(warn=0)

  (-1)^q * theta^(-0.5) * exp(-g(w=wTilde, q=q, x=x, theta=theta)) *
    (exp(wTilde) * x + theta^(-1))^(-0.5)
}

```

A diagnostic plot for the frailty distribution in the shared frailty model

When modelling dependence between event times, not only the degree of global dependence is of interest, but also the way the local (at a given time) dependence changes over time, i.e. the dependence structure. In particular, the times at which the dependence is high often receive special attention (see, e.g., Nan et al., 2006).

In the frailty model framework, the dependence structure is dictated by the frailty distribution (Anderson et al., 1992; Hougaard, 1995). Hence, the frailty distribution needs to be carefully chosen to correctly model the dependence structure in the data. As the frailties are unobserved, though, specifying the frailty distribution is a difficult issue. This chapter introduces a new diagnostic plot to guide the choice of the frailty distribution.

Section 3.1 provides references to key papers in the area and outlines a general diagnostic framework based on dependence measures. In Section 3.2, we review common ways to measure dependence in survival data and we discuss the patterns of dependence induced by different frailty distributions. From the discussion of Section 3.2, it will become clear that the probability mass that a frailty distribution puts in the tails is a key feature as it drives the dependence structure. To capture the behaviour in the tails, the proposed diagnostic plot is based on the quantile dependence function. Quantile dependence is introduced in Section 3.3 where we explain how to obtain the model-based and the non-parametric estimates that are used to construct the diagnostic plot. The method is illustrated in Section 3.4 and is assessed with simulations in Section 3.5. Concluding remarks are given in Section 3.6.

The content of this chapter is submitted for publication (Munda & Legrand, 2014b).

3.1

Introduction

As a matter of software availability, the gamma distribution is often used in practice to model the frailty term. Diagnostic checks specific to the gamma frailty distribution include Shih & Louis (1995), Cui & Sun (2004), and Geerdens et al. (2013). Only a few diagnostic procedures are available for other frailty distributions. The graphical tool developed in Economou & Caroni (2008) is valid if, for all t , the distribution of the frailty term among the “surviving clusters” at time t belongs to the same family as the original frailty distribution (closure property). In Oakes (1989), the cross ratio function, a measure of bivariate dependence, is used as a diagnostic tool. The method can in principle be used with any frailty distribution, but does not explicitly account for censoring. Some technical improvements of the method are proposed in Viswanathan & Manatunga (2001) and in Chen & Bandeen-Roche (2005). Another, more computationally demanding, diagnostic tool based on the cross ratio function is studied in Glidden (2007).

In the last four papers cited above, the basic idea is that although frailties are unobservable, the dependence structure that they impose on the data can be observed. Further, the dependence structure is dictated by the frailty distribution (Anderson et al., 1992; Hougaard, 1995). It follows that a graphical comparison of the observed dependence structure with selected model-based structures can be used to reveal the frailty distribution that best describes the pattern of dependence in the data. That is, the problem of assessing the frailty distribution assumption can be linked to the problem of measuring dependence in clustered survival data. Common ways to measure dependence in clustered survival data are reviewed in the next section.

3.2

Dependence in bivariate survival data

Dependence measures have been developed mainly for bivariate data. In this section, we review a number of coefficients that evaluate dependence between two event times, (T_1, T_2) .

3.2.1 Overall dependence

Kendall's τ is a rank-based dependence coefficient which ranges from -1 to 1 , with $\tau = 0$ under independence. To define Kendall's τ , we need an independent copy (i.e. with the same distribution) of (T_1, T_2) , say (T'_1, T'_2) . Kendall's τ is the probability that (T_1, T_2) and (T'_1, T'_2) are concordant minus the probability that this pair is discordant,

$$\tau = \Pr\left((T_1 - T'_1)(T_2 - T'_2) > 0\right) - \Pr\left((T_1 - T'_1)(T_2 - T'_2) < 0\right)$$

Thus, Kendall's τ expresses the probability that the order of the first coordinates is the same as the order of the second coordinates ("concordance", i.e. large values tend to occur with large values and small values tend to occur with small values) minus the probability that the order differs between the coordinates ("discordance"). Kendall's τ is scale-invariant, i.e. it remains unchanged under monotonic transformations of the random variables. It is worth noting that interpretation of Kendall's τ requires a pair of clusters and that observations within a cluster must have an ordering ($(T_1 - T'_1)(T_2 - T'_2)$ is not the same as $(T_1 - T'_2)(T_2 - T'_1)$).

An alternative to Kendall's τ , which can be interpreted with one cluster only, is given by the median concordance coefficient, also known as Blomqvist's β ,

$$\beta = \Pr\left((T_1 - \tilde{t}_1)(T_2 - \tilde{t}_2) > 0\right) - \Pr\left((T_1 - \tilde{t}_1)(T_2 - \tilde{t}_2) < 0\right)$$

with $\tilde{t}_1$ and $\tilde{t}_2$ the medians of T_1 and T_2 . Like Kendall's τ , Blomqvist's β lies between -1 and 1 , equals 0 under independence, and is invariant under strictly increasing transformations of T_1 and T_2 . Blomqvist's β often provides an accurate approximation to Kendall's τ (Nelsen, 2006, Section 5.1.4).

In the frailty model framework, dependence within a cluster is generated by the frailty term. Kendall's τ and Blomqvist's β can be written in terms of the Laplace transform of the frailty term as (Hougaard, 2000, Section 7.2.5)

$$\tau = 4 \int_0^\infty x \mathcal{L}(x) \mathcal{L}^{(2)}(x) dx - 1$$

with $\mathcal{L}^{(2)}(x)$ the second derivative of $\mathcal{L}(x) = \mathbb{E}(\exp(-Ux))$, and

$$\beta = 4\mathcal{L}\{2\mathcal{L}^{-1}(1/2)\} - 1$$

Appendix 3.A provides explicit forms of τ and β for a number of frailty distributions.

Note: $\tau = 0$ and $\beta = 0$ do not imply independence

Kendall's τ and Blomqvist's β only capture monotonic association. Under independence we have $\tau = 0$ and $\beta = 0$, but $\tau = 0$ and $\beta = 0$ do not imply independence in general.

Counterexample

Let X and Y be two continuous random variables with copula

$$C(u, v) = uv + \theta u(u-1)(2u-1)v(v-1)(2v-1)$$

for some parameter θ . In order for $C(\cdot, \cdot)$ to be 2-increasing (and hence a copula; cf. Nelsen (2006, Definition 2.2.2, page 10)), we restrict $\theta \in [-9, 9(\sqrt{3}-1)]$. Independence between X and Y corresponds to $\theta = 0$.

On the one hand, Kendall's τ between X and Y is (Nelsen, 2006, Theorem 5.1.3, page 161)

$$\tau = 4 \int_0^1 \int_0^1 C(u, v) \frac{\partial^2 C}{\partial u \partial v}(u, v) du dv - 1 = 0$$

On the other hand, Blomqvist's β between X and Y is (Nelsen, 2006, Section 5.1.4, Formula 5.1.27)

$$\beta = 4C\left(\frac{1}{2}, \frac{1}{2}\right) - 1 = 0$$

That is, $\tau = 0$ and $\beta = 0$ even though $\theta \neq 0$.

3.2.2 Local dependence

On average, frail subjects (those with a large value of U) experience the event first. This selection process, to a large extent governed by the frailty distribution, leads to changes in the dependence structure over time. Kendall's τ and Blomqvist's β are overall measures of dependence. Therefore, Kendall's τ and Blomqvist's β cannot detect changes in the dependence structure over time. To address this question, a local measure (i.e., as a function of time) is needed.

The cross ratio function is a local measure of dependence defined as (Clayton, 1978)

$$\zeta(t_1, t_2) = \frac{h_1(t_1 | T_2 = t_2)}{h_1(t_1 | T_2 > t_2)}$$

Thus, $\zeta(t_1, t_2)$ compares the hazard rate of member 1 at time t_1 given that member 2 failed at time t_2 to the hazard rate of member 1 at t_1 given that member 2 survived beyond t_2 . Unity indicates no association.

In the frailty model framework, $\zeta(t_1, t_2)$ depends on t_1 and t_2 only through the joint survival function,

$$\zeta(t_1, t_2) = \zeta^*(S(t_1, t_2))$$

where $\zeta^*(\cdot)$ determines the frailty distribution uniquely, up to a scale factor (Oakes, 1989). Explicit forms of $\zeta^*(\cdot)$ are given in Appendix 3.A. For the gamma frailty distribution, $\zeta^*(\cdot)$ is a constant. For both the inverse Gaussian and the positive stable frailty distributions, $\zeta^*(v)$ decreases to 1 as $v \rightarrow 0$.

The cross ratio as a diagnostic tool

The diagnostic tool proposed in Oakes (1989) compares graphically a non-parametric (model-free) estimate of the cross ratio to model-based estimates. In particular, substantial deviation of the model-free estimate from the constant line provides evidence against the gamma frailty distribution.

Model-based estimates of the cross ratio follow from estimates of the frailty parameter obtained by fitting the frailty model with different frailty distributions (cf. the formulas in Appendix 3.A).

The model-free estimate of the cross ratio can be based on an alternative, asymptotic representation. Given (T_{i1}, T_{i2}) and (T_{k1}, T_{k2}) , two randomly selected clusters (among $\binom{s}{2}$ possible pairs), let

$$\gamma(r) = \frac{\Pr\left((T_{i1} - T_{k1})(T_{i2} - T_{k2}) > 0 \mid R_{ik} = r\right)}{\Pr\left((T_{i1} - T_{k1})(T_{i2} - T_{k2}) < 0 \mid R_{ik} = r\right)}$$

with

$$R_{ik} = \#\left\{\ell \in \{1, \dots, s\} : T_{\ell 1} \geq \min(T_{i1}, T_{k1}), T_{\ell 2} \geq \min(T_{i2}, T_{k2})\right\}$$

If $r/s \rightarrow v$ as $s \rightarrow \infty$ and $r \rightarrow \infty$, then it can be shown that $\gamma(r) \rightarrow \zeta^*(v)$ (Oakes, 1989). Accordingly, a non-parametric estimate of $\gamma(r)$ can be used as a substitute for a non-parametric estimate of $\zeta^*(v)$. The non-parametric estimate of $\gamma(r)$ proposed in Oakes (1989) follows by counting the number of concordances and discordances among the pairs of clusters such that $R_{ik} = r$. More details can be found in Duchateau & Janssen (2008, Section 4.2.6). Some technical improvements of the method are proposed in Viswanathan & Manatunga (2001) and in Chen & Bandeen-Roche (2005).

Note: limitations of the method

There are problems inherent to the use of $\gamma(r)$ as a substitute for $\zeta^*(v)$:

- To determine the concordance/discordance status of a pair, observations within a cluster must have an ordering.
- In the presence of censoring, there are cases where the concordance/discordance status cannot be ascertained. In Oakes (1989), this problem is circumvented by assuming complete data (no censoring).

This calls for a new non-parametric estimator of the cross ratio that does not require the use of $\gamma(\cdot)$, or that can at least deal with censoring (see, e.g., Lakhal et al. (2009)).

3.2.3 Association patterns (early/late)

Another time-dependent association measure, with a simple probability interpretation, is given by (Anderson et al., 1992)

$$\psi(t_1, t_2) = \frac{S(t_1, t_2)}{S_1(t_1) S_2(t_2)} = \frac{\Pr(T_1 > t_1 | T_2 > t_2)}{\Pr(T_1 > t_1)}$$

with $S(\cdot, \cdot)$ the joint survival function of (T_1, T_2) , and $S_1(\cdot)$ and $S_2(\cdot)$ the marginal survival functions of T_1 and T_2 . Large values of $\psi(\cdot, \cdot)$ indicate positive association while unity indicates no association.

In Figure 3.1, contour plots of $\psi(\cdot, \cdot)$ depict the evolution of the dependence graphically for the gamma and the positive stable frailty distributions (Duchateau & Janssen, 2008, Chapter 4). The gamma frailty

distribution is characterised by late dependence (contour lines are very close together at late times) while the positive stable frailty distribution is characterised by early dependence (contour lines indicating positive dependence appear early). As the next section will formalise, this can be explained by the behaviour of the frailty distribution in the tails. On the one hand, the positive stable frailty distribution has a heavy right tail (cf. Figure 3.2), leading to strong dependence initially. On the other hand, the gamma frailty distribution has a lot of probability mass skewed to the left (cf. Figure 3.2), leading to strong dependence at late times.

We conclude that the tails of the frailty distribution play an important role in the way the dependence changes over time.

3.3 Quantile dependence

To capture the behaviour in the tails, we introduce an additional local measure of dependence, namely quantile dependence.

3.3.1 Definitions

Quantile dependence is a measure of local association between two random variables, here (T_1, T_2) . The lower quantile dependence coefficient $\lambda_\ell(q)$ is defined for $q \in (0, 1)$ as

$$\lambda_\ell(q) = \Pr \left(T_1 \leq F_1^{-1}(q) \mid T_2 \leq F_2^{-1}(q) \right)$$

with $F_1^{-1}(\cdot)$ and $F_2^{-1}(\cdot)$ the quantile functions of T_1 and T_2 . Thus, $\lambda_\ell(q)$ is the conditional probability that member 1 fails before time $F_1^{-1}(q)$, given that member 2 failed before time $F_2^{-1}(q)$. Under independence, $\lambda_\ell(\cdot)$ is the identity function. If we let q approach zero, then we obtain the lower tail dependence coefficient,

$$\Lambda_\ell = \lim_{q \rightarrow 0} \lambda_\ell(q)$$

The upper counterparts, $\lambda_u(\cdot)$ and Λ_u , are similarly defined,

$$\lambda_u(q) = \Pr \left(T_1 > F_1^{-1}(q) \mid T_2 > F_2^{-1}(q) \right)$$

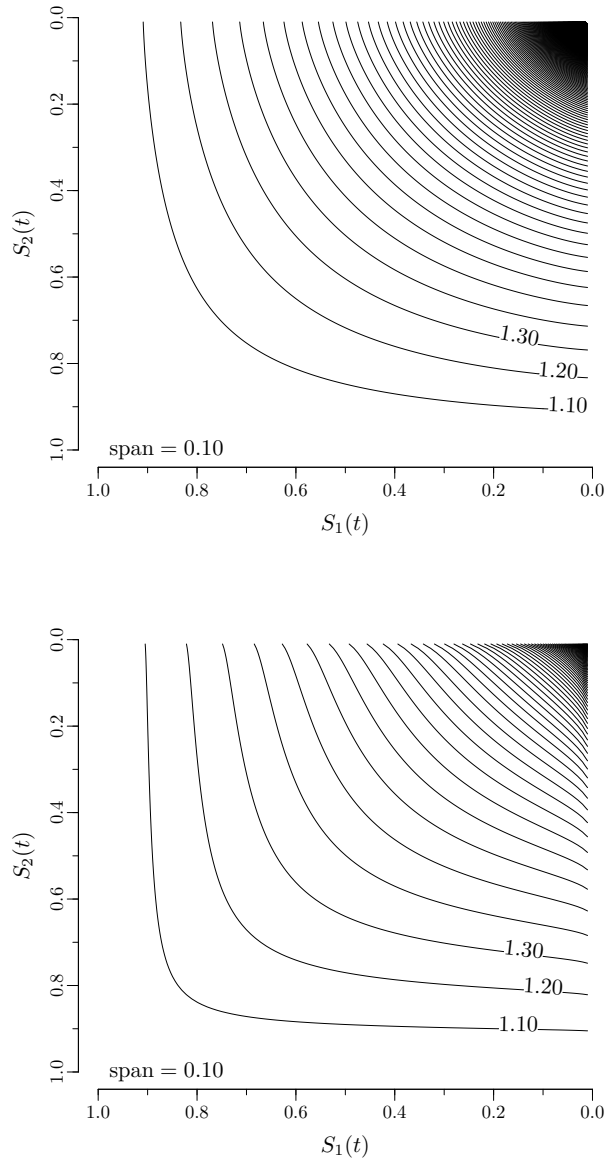


Figure 3.1: Contour plots of $\psi(t_1, t_2)$ for the gamma frailty distribution (upper panel) and for the positive stable frailty distribution (lower panel), with frailty parameter chosen so that $\tau = 0.4$.

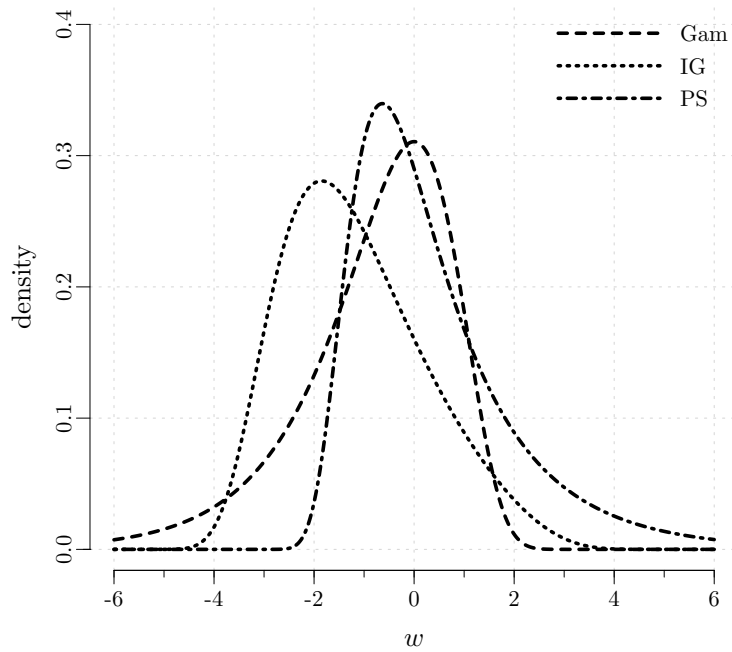


Figure 3.2: Density of $W = \log(U)$, for U following the gamma (Gam), the inverse Gaussian (IG), and the positive stable (PS) frailty distributions, with frailty parameter chosen so that $\tau = 0.4$. From the log densities, it can be seen that the gamma distribution has a large left tail, that the positive stable has a large right tail, and that the inverse Gaussian takes a position in between.

and

$$\Lambda_u = \lim_{q \rightarrow 1} \lambda_u(q)$$

A distribution is said to display lower tail dependence if $\Lambda_\ell > 0$ and upper tail dependence if $\Lambda_u > 0$. Tail independence means that extreme events occur independently in both tails ($\Lambda_\ell = 0$ and $\Lambda_u = 0$).

In terms of the joint survival function, $\lambda_\ell(q)$ and $\lambda_u(q)$ are written as

$$\lambda_\ell(q) = \frac{S(F_1^{-1}(q), F_2^{-1}(q)) + 2q - 1}{q} \quad (3.1)$$

and

$$\lambda_u(q) = \frac{S(F_1^{-1}(q), F_2^{-1}(q))}{1 - q} \quad (3.2)$$

In the frailty model framework, the joint survival function has the following Laplace transform representation (Oakes, 1989)

$$S(t_1, t_2) = \mathcal{L} \left\{ \mathcal{L}^{-1}(S_1(t_1)) + \mathcal{L}^{-1}(S_2(t_2)) \right\}$$

By plugging this representation of $S(t_1, t_2)$ into (3.1) and (3.2), we obtain

$$\lambda_\ell(q) = \frac{\mathcal{L} \left\{ 2\mathcal{L}^{-1}(1 - q) \right\} + 2q - 1}{q} \quad (3.3)$$

and

$$\lambda_u(q) = \frac{\mathcal{L} \left\{ 2\mathcal{L}^{-1}(1 - q) \right\}}{1 - q} \quad (3.4)$$

3.3.2 Association patterns

In Figure 3.3, plots of $\lambda_\ell(q)$ and $\lambda_u(q)$ versus q depict the evolution of the dependence graphically for the gamma, the inverse Gaussian, and the positive stable frailty distributions. The explicit forms, obtained by substituting the corresponding Laplace transforms into (3.3) and (3.4), are given in Appendix 3.A. The gamma frailty distribution displays upper tail dependence ($\Lambda_\ell = 0$ and $\Lambda_u > 0$), the positive stable frailty distribution displays lower tail dependence ($\Lambda_\ell > 0$ and $\Lambda_u = 0$), and

the inverse Gaussian frailty distribution yields independence in both tails ($\Lambda_\ell = 0$ and $\Lambda_u = 0$).

As

$$\lambda_u(q) = \frac{q\lambda_\ell(q) - 2q + 1}{1 - q}$$

we now focus solely on $\lambda_\ell(q)$.

3.3.3 Model-based estimate

To obtain a model-based estimate $\hat{\lambda}_\ell(q)$ of $\lambda_\ell(q)$, we plug an estimate of the frailty parameter into the Laplace transform in (3.3). An estimate of the frailty parameter can be obtained by fitting the frailty model. In this chapter, we use the R library `parfm` as it is, to date, the only package that supports all the above-mentioned frailty distributions (cf. Chapter 2). The standard error of $\hat{\lambda}_\ell(q)$ can be derived from the standard error of the frailty parameter by using the delta method. Simulation results for the bias and standard error of $\hat{\lambda}_\ell(q)$ are shown in Appendix 3.C.

3.3.4 Model-free estimate

To obtain a model-free estimate $\tilde{\lambda}_\ell(q)$ of $\lambda_\ell(q)$, we plug a non-parametric estimate of the joint survival function in (3.1). There is, at present, no fully satisfactory general solution to this non-parametric estimation problem (Hougaard, 2000, Chapter 14). However, non-parametric estimation of the joint survival function becomes quite straightforward under mild, common, and often reasonable assumptions regarding the censoring mechanism (Wang & Wells, 1997).

Given a bivariate cluster of event times (T_1, T_2) , possibly right-censored by (C_1, C_2) , the random variables that we observe are $Y_1 = \min(T_1, C_1)$, $\Delta_1 = I(T_1 \leq C_1)$, $Y_2 = \min(T_2, C_2)$, and $\Delta_2 = I(T_2 \leq C_2)$. We allow for the presence of a binary covariate that indicates position in a cluster (e.g., treatment arm). Finally, we denote by $S(\cdot, \cdot)$ and $G(\cdot, \cdot)$ the joint survival functions of (T_1, T_2) and (C_1, C_2) , respectively.

Under the common hypothesis of independence between the event times and the censoring times, we have

$$S(t_1, t_2) = \frac{\Pr(Y_1 > t_1, Y_2 > t_2)}{G(t_1, t_2)} \quad (3.5)$$

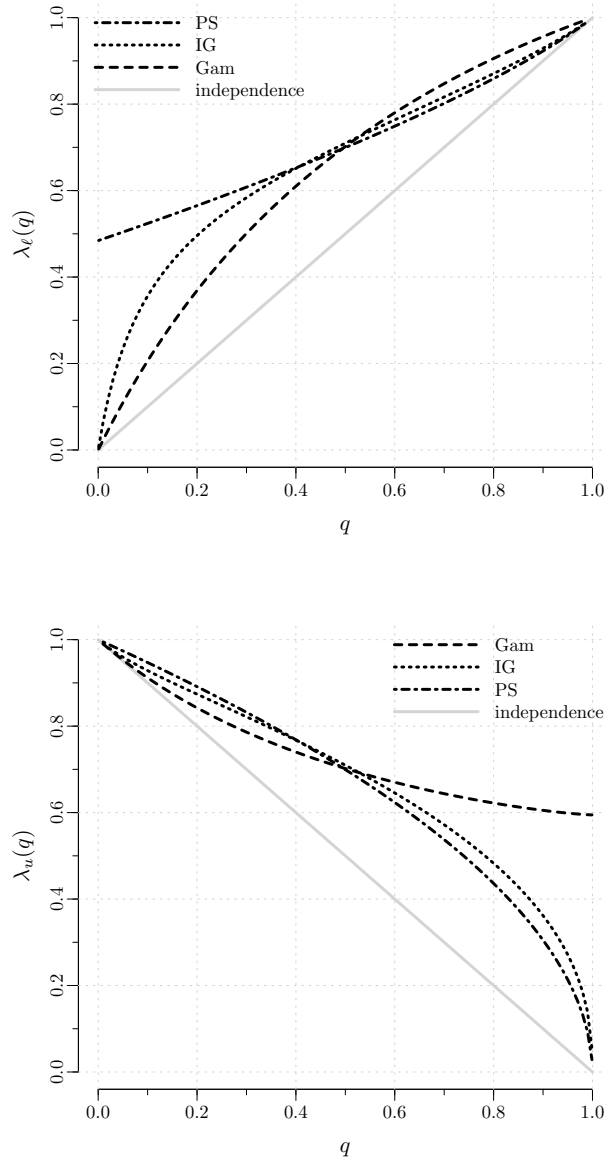


Figure 3.3: Lower and upper quantile dependence coefficients for the gamma (Gam), the inverse Gaussian (IG), and the positive stable (PS) frailty distributions, with frailty parameter chosen so that $\tau = 0.4$.

The numerator, i.e. the joint survival function of the observables (Y_1, Y_2) , can be estimated by the empirical survival function. Estimation of the denominator, i.e. the joint survival function of the censoring times, is similar to the estimation of $S(t_1, t_2)$, but can be simplified if the relationship between C_1 and C_2 is of simple form, as often the case in practice. Common scenarios include:

- independent censoring ($C_1 \perp\!\!\!\perp C_2$):

$$G(t_1, t_2) = \Pr(C_1 > t_1) \Pr(C_2 > t_2)$$

- univariate censoring ($C_1 = C_2 = C$):

$$G(t_1, t_2) = \Pr\left(C > \max(t_1, t_2)\right)$$

In these cases, estimation of $G(\cdot, \cdot)$ reduces to a univariate estimation problem (Kaplan-Meier method).

In (3.5), we need to take $t_j = \hat{F}_j^{-1}(q)$, with $\hat{F}_j^{-1}(\cdot)$ the Kaplan-Meier quantile function of T_j ($j = 1, 2$). In the absence of covariates, paired observations are exchangeable and we can use an overall Kaplan-Meier quantile function.

The standard error of $\tilde{\lambda}_\ell(q)$ can be obtained by means of the non-parametric bootstrap for clustered survival data (Therneau & Grambsch, 2000, page 249; Ren et al., 2010).

Simulation results for the bias and standard error of $\tilde{\lambda}_\ell(q)$ are shown in Appendix 3.C.

3.3.5 Quantile dependence as a diagnostic tool

We propose to use the quantile dependence function as a diagnostic tool for the frailty distribution in the shared frailty model. By comparing a model-based estimate with the model-free estimate over a grid of q -values, we obtain a graphical assessment of the frailty distribution.

The model-free estimator from Section 3.3.4 requires bivariate data, i.e. clusters of size 2. However, clustered survival data often involve larger and varying cluster sizes. As an ad-hoc solution, we propose to obtain $\tilde{\lambda}_\ell(q)$ as

$$\tilde{\lambda}_\ell(q) = \frac{1}{B} \sum_{b=1}^B \tilde{\lambda}_\ell^{(b)}(q)$$

for some $B \gg 0$, with $\tilde{\lambda}_\ell^{(b)}(q)$ the non-parametric estimate of $\lambda_\ell(q)$ based on a bivariate data set obtained by selecting two observations at random in each cluster (respecting the ordering, if any).

3.4 Examples

Example 1

The standard technique to assess when a fracture has healed in dogs is based on radiography. To evaluate the use of ultrasonography as an alternative technique, each of 106 dogs was evaluated for time to fracture healing with the two techniques (Duchateau & Janssen, 2008, Example 1.2). Since each dog serves as its own control, there is likely to be dependence in the data. Each cluster is of size 2 and can be ordered according to the binary covariate (radiography versus ultrasound). Further, there is no censoring. Under these conditions, the diagnostic plot based on the cross ratio discussed in Section 3.2.2 can be used. We also use the proposed graphical tool based on quantile dependence. The two diagnostic plots are shown in Figure 3.4. The model-based estimates are obtained by using a log-logistic baseline hazard function in the frailty model. It appears from both diagnostic plots that the dependence structure is well captured by the positive stable frailty distribution.

Example 2

As a second illustrative example, we consider data on 100 dairy cows which were followed up for infection (mastitis) in one or more udder quarters (Duchateau & Janssen, 2008, Example 1.4). Observations are at the udder quarter level. Each cluster is thus of size 4. About 20% of the observations are censored (no infection occurred during follow-up), with censoring at the cow level. The diagnostic plot based on the cross ratio function discussed in Section 3.2.2 does not apply to this particular example as (i) observations in a cluster are exchangeable (clusters cannot be ordered), and (ii) some of the observations are censored. The proposed diagnostic plot based on quantile dependence is shown in Figure 3.5. The model-based estimates are obtained by using a Weibull baseline hazard function in the frailty model. It appears that the positive

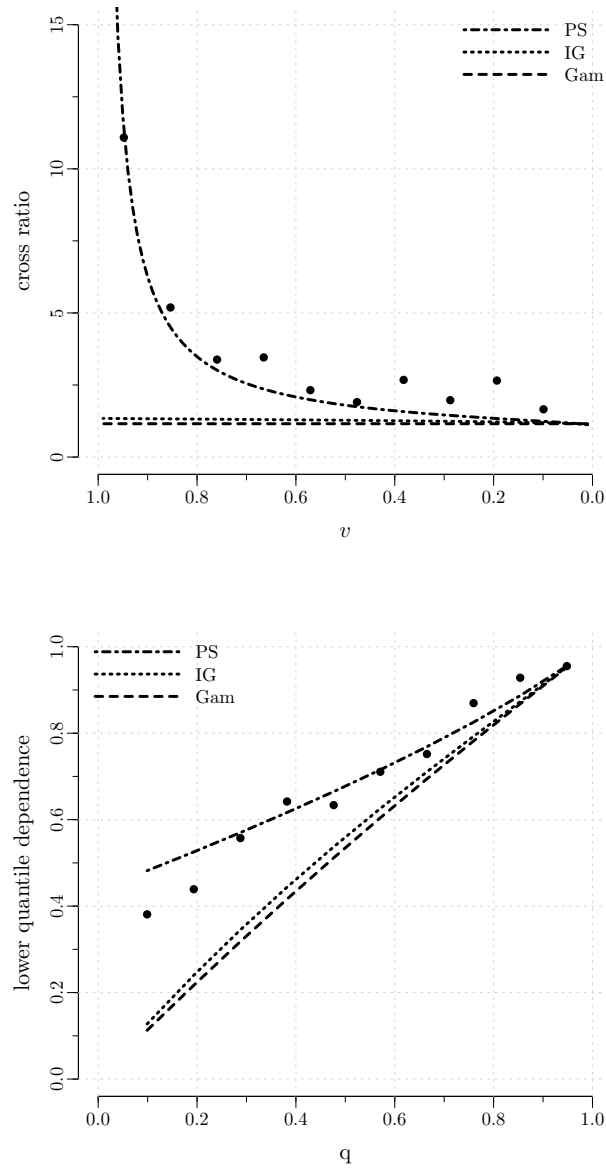


Figure 3.4: Diagnostic plots for Example 1. The dots represent the model-free estimates at discrete points.

stable frailty distribution is suitable to model the dependence structure in these data.

The gamma distribution has been rejected for these data by a formal goodness-of-fit test in Geerdens et al. (2013). To confirm this conclusion using the proposed diagnostic plot, the data can be resampled under the hypothesis that the frailty term is gamma distributed; cf. the thesis appendix (Appendix 3). The upper panel of Figure 3.6 shows 500 bootstrap model-based estimates of $\lambda_\ell(\cdot)$ obtained from the gamma frailty model together with the model-free estimate. The model-free estimate deviates substantially from the bootstrap estimates. The gamma frailty distribution assumption is thus called into question (thus confirming the result in Geerdens et al. (2013)). In the lower panel of Figure 3.6, the data are resampled under the hypothesis that the frailty term follows a positive stable distribution and the bootstrap model-based estimates of $\lambda_\ell(\cdot)$ are obtained from the positive stable frailty model. It is clear from the lower panel of Figure 3.6 that the positive stable distribution is a better choice than the gamma to model the frailty term.

3.5 Simulations

In this section, we use simulations to study the performance of the proposed diagnostic tool with respect to the number of clusters/observations per cluster $((s, n) = (150, 4), (75, 8), (50, 12), (30, 20), \text{ and } (20, 30))$ and amount of censoring (20% and 40%).

Event times are generated from the frailty model with a Weibull baseline hazard ($h_0(t) = \lambda \rho t^{\rho-1}$, with $\lambda = 0.7$ and $\rho = 1.5$) and a binary covariate dividing each cluster into two balanced groups ($\beta = -0.4$); cf. the thesis appendix (Appendix 2). The frailty term follows either a gamma (Gam), an inverse Gaussian (IG), or a positive stable (PS) distribution. The frailty parameter is chosen so that Kendall's τ equals 0.4.

We run 1000 simulations. For each simulation, we compute the L_2 -distance between the model-free estimate of $\lambda_\ell(\cdot)$ and the three model-based estimates (Gam, IG, and PS). Table 3.1 reports the selection percentage for each frailty distribution, i.e. the proportion of times that each frailty distribution is selected by the diagnostic plot (based on the smallest distance value). Specific comments are as follows:

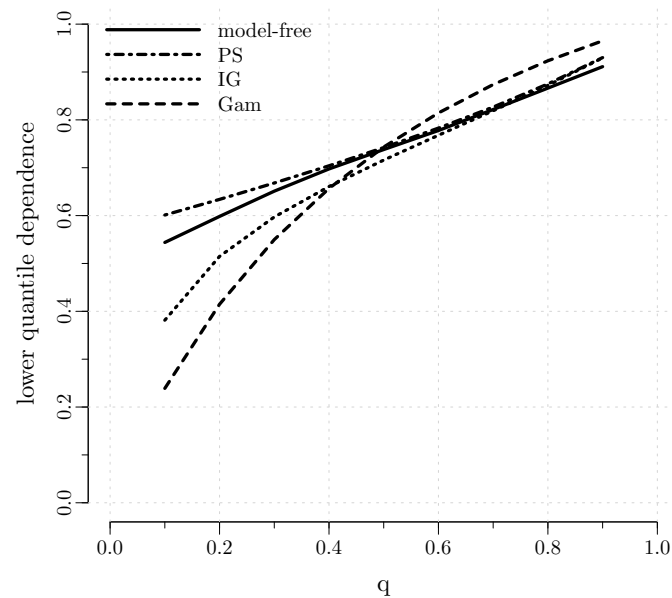


Figure 3.5: Diagnostic plot for Example 2. In this plot, the discrete model-free estimate has been smoothed by means of locally-weighted polynomial regression (cf. the `lowess()` function in R).

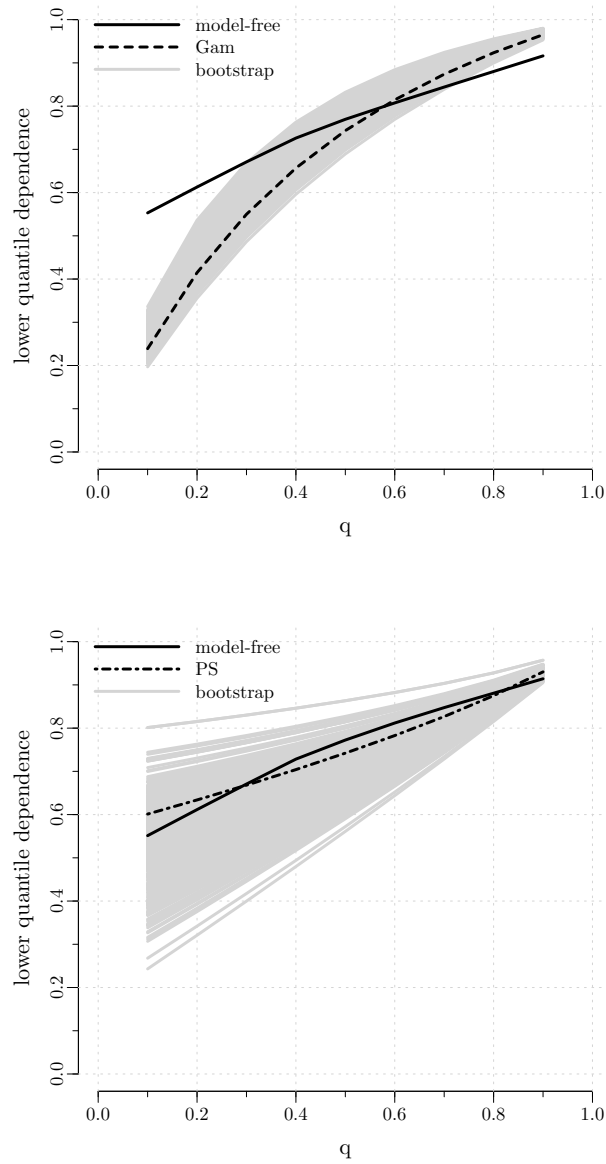


Figure 3.6: Diagnostic plot for Example 2 supplemented with 500 bootstrap model-based estimates obtained under the gamma (upper panel) or the positive stable (lower panel) frailty distribution assumption.

- The standard assumption for the frailty term is that it follows a gamma distribution. Therefore, the most common form of misspecification is that of using the gamma distribution while the actual frailty distribution is not gamma. This type of misspecification is detected by the proposed diagnostic plot.
- Distinction between the incorrect inverse Gaussian and the correct frailty distribution may be difficult in some settings. Indeed, there is a tendency for the inverse Gaussian model-based estimate of $\lambda_\ell(\cdot)$ to get close to the correct model-based estimate. It is noteworthy that an inverse Gaussian frailty term cannot model high levels of dependence (Kendall's $\tau < \frac{1}{2}$; cf. Appendix 3.A).
- Increasing the number of clusters appears to be more beneficial than increasing the number of observations per cluster. However, moderately large cluster sizes are needed when censoring limits the amount of information within a cluster.

3.6

Discussion

Diagnostic plots have proven to be useful in survival analysis, e.g. to check the assumption of proportional hazards in the Cox model. Diagnostic plots are generally not intended to replace formal tests, but are certainly useful companions to guide the model selection.

In the shared frailty model, different frailty distributions induce different dependence structures in the data. Based on this observation, we have constructed a diagnostic plot for the frailty distribution by comparing an empirical to a model-based estimate of the quantile dependence coefficient. The main motivation for using quantile dependence as a diagnostic tool was to capture the behaviour of the frailty distribution in the tails, a key feature that drives the way dependence changes over time.

Model-based estimates of the quantile dependence coefficient can be obtained by fitting the frailty model with different frailty distributions (cf. Section 3.3.3). The empirical estimate is developed for bivariate data (cluster of size 2), with possibly a binary covariate indicating position in a cluster (cf. Section 3.3.4). If additional covariate information is available, the empirical estimate can be obtained based on a binary

Table 3.1: Simulation results – selection percentages for each frailty distribution.

censoring	true: Gam			true: IG			true: PS		
	Gam	IG	PS	Gam	IG	PS	Gam	IG	PS
$(s, n) = (150, 4)$									
20%	85	15	0	1	86	13	0	19	81
40%	72	28	0	12	71	17	1	23	76
$(s, n) = (75, 8)$									
20%	94	6	0	7	87	6	2	30	68
40%	73	27	0	11	80	9	4	26	70
$(s, n) = (50, 12)$									
20%	71	29	0	1	69	30	0	28	72
40%	47	53	0	2	72	26	0	22	78
$(s, n) = (30, 20)$									
20%	57	43	0	0	59	41	0	33	67
40%	26	74	0	0	59	41	0	18	82
$(s, n) = (20, 30)$									
20%	41	58	1	0	53	47	0	31	69
40%	14	85	1	0	56	44	0	21	79

prognostic index combining the information on several risk factors into a single value indicative of the prognosis of the subject (“low-risk group” versus “high-risk group”). Larger and varying cluster sizes can be handled by using pairwise comparisons (cf. Section 3.3.5).

The proposed diagnostic plot is suitable for any frailty distribution for which the Laplace transform and its inverse can be evaluated (cf. Equation 3.3).

Overall, the proposed diagnostic plot provides a simple, readily available, and broadly applicable tool that can be used on its own or that can serve as a supplement to the few existing goodness-of-fit tests.

More discussion is provided at the end the thesis (closing part).

Appendix 3.A

Explicit formulas

Gamma frailty distribution ($\theta > 0$):

$$f(u) = \frac{(1/\theta)^{1/\theta} u^{(1/\theta)-1} \exp(-(1/\theta)u)}{\Gamma(1/\theta)}$$

$$\mathcal{L}(x) = (1 + \theta x)^{-1/\theta}$$

$$\mathcal{L}^{-1}(x) = \frac{1}{\theta} (x^{-\theta} - 1)$$

$$\tau = \frac{\theta}{\theta + 2}$$

$$\beta = 4 \left(2^{\theta+1} - 1 \right)^{-1/\theta} - 1$$

$$\zeta^*(v) = \theta + 1$$

$$\lambda_\ell(q) = \frac{\left(2(1-q)^{-\theta} - 1 \right)^{-1/\theta} + 2q - 1}{q}$$

$$\lambda_u(q) = \frac{\left(2(1-q)^{-\theta} - 1 \right)^{-1/\theta}}{1 - q}$$

$$\Lambda_\ell = 0$$

$$\Lambda_u = 2^{-1/\theta}$$

Inverse Gaussian frailty distribution ($\theta > 0$):

$$f(u) = \frac{1}{\sqrt{2\pi\theta}} u^{-3/2} \exp\left(-\frac{1}{2\theta u} (u-1)^2\right)$$

$$\mathcal{L}(x) = \exp\left\{\frac{1}{\theta} \left(1 - \sqrt{1 + 2\theta x}\right)\right\}$$

$$\mathcal{L}^{-1}(x) = \frac{\log(x)}{2} (\theta \log(x) - 2)$$

$$\tau = \frac{1}{2} - \frac{1}{\theta} + \frac{2}{\theta^2} \exp\left(\frac{2}{\theta}\right) \int_{2/\theta}^{\infty} u^{-1} \exp(-u) \, du < \frac{1}{2}$$

$$\beta = 4 \exp \left\{ \frac{1}{\theta} \left(1 - \sqrt{2(\theta \log(2) + 1)^2 - 1} \right) \right\} - 1 < 2^{(2-\sqrt{2})} - 1$$

$$\zeta^*(v) = 1 + \frac{1}{(1/\theta) - \log(v)}$$

$$\lambda_\ell(q) = \frac{\mathcal{L}\left\{ \log(1-q) \left(\theta \log(1-q) - 2 \right) \right\} + 2q - 1}{q}$$

$$\lambda_u(q) = \frac{\mathcal{L}\left\{ \log(1-q) \left(\theta \log(1-q) - 2 \right) \right\}}{1-q}$$

$$\Lambda_\ell = 0$$

$$\Lambda_u = 0$$

Positive stable frailty distribution ($0 < \theta < 1$):

$$f(u) = -\frac{1}{\pi u} \sum_{k=1}^{\infty} \left\{ \frac{\Gamma(k(1-\theta) + 1)}{k!} \sin((1-\theta)k\pi) \left(-u^{\theta-1}\right)^k \right\}$$

$$\mathcal{L}(x) = \exp\left(-x^{1-\theta}\right)$$

$$\mathcal{L}^{-1}(x) = \left(-\log(x)\right)^{1/(1-\theta)}$$

$$\tau = \theta$$

$$\beta = 4 \exp \left\{ -2^{1-\theta} \log(2) \right\} - 1$$

$$\zeta^*(v) = 1 - \frac{\theta}{(1-\theta) \log(v)}$$

$$\lambda_\ell(q) = \frac{\exp \left\{ 2^{1-\theta} \log(1-q) \right\} + 2q - 1}{q}$$

$$\lambda_u(q) = \frac{\exp \left\{ 2^{1-\theta} \log(1-q) \right\}}{1-q}$$

$$\Lambda_\ell = 2(1 - 2^{-\theta})$$

$$\Lambda_u = 0$$

Appendix 3.B

Kendall's τ versus θ

In the frailty model framework, Kendall's τ can be computed as (cf. Section 3.2.1)

$$\tau = 4 \int_0^\infty x \mathcal{L}(x) \mathcal{L}^{(2)}(x) dx - 1$$

which involves the frailty parameter θ through $\mathcal{L}(x)$. Kendall's τ can therefore be used to interpret the value of θ .

In Table 3.2, we give the values of θ that correspond to a sequence of values of τ for the gamma, the inverse Gaussian, the positive stable, and the log-normal frailty distributions. For the inverse Gaussian and the log-normal, these values are obtained by means of a numerical method rather than an explicit formula.

Inverse Gaussian frailties

The integral that appears in the formula of Kendall's τ of an IG frailty term (cf. Appendix 3.A) is an incomplete gamma function. This integral is also known as the exponential integral (Abramowitz & Stegun, 1972, Chapter 5). The exponential integral, and therefore Kendall's τ , has no solution in terms of elementary functions. Column 3 of Table 3.2 has therefore been obtained by means of numerical integration (cf. `gamma_inc()` in R, part of the `gsl` library).

Table 3.2: Kendall's τ versus θ .

	τ								
	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
Gam	0.22	0.50	0.86	1.33	2.00	3.00	4.67	8.00	18.0
IG [†]	0.28	0.81	2.03	6.20					
PS	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
LN	0.23	0.56	1.06	1.95	3.59	6.69	13.0	27.9	74.7

[†] The Kendall's τ of an IG frailty term is always lower than $1/2$.

Log-normal frailties

There is no explicit evaluation of the Laplace transform for log-normal frailties. Consequently, there is no explicit formula for Kendall's τ . In Section 2.4.6, we show how to approximate, for a given $x \geq 0$, the Laplace transform of a log-normal frailty term and its derivatives. This makes it possible to evaluate Kendall's τ using adaptive quadrature (cf. `integrate()` in R).

Appendix 3.C

Model-free and model-based estimators of $\lambda_\ell(q)$

Table 3.3: Simulation results – Estimates of $\lambda_\ell(q)$ and their empirical and estimated standard errors (SEs) in 1000 simulated data sets (cf. Section 3.5) based on the model-free (cf. Section 3.3.4) and model-based (cf. Section 3.3.3) estimators. Each simulated data set consists of 100 clusters of size 2 and has approximately 20% of censored observations.

q	true value	mean est.	emp. SE	mean SE
<i>frailty dist: Gam – estimator: model-free</i>				
0.10	0.206	0.256	0.153	0.147
0.25	0.438	0.440	0.106	0.095
0.50	0.702	0.704	0.063	0.057
0.75	0.877	0.875	0.044	0.042
0.90	0.956	0.941	0.039	0.034
<i>frailty dist: Gam – estimator: model-based Gam</i>				
0.10	0.206	0.207	0.021	0.021
0.25	0.438	0.438	0.032	0.032
0.50	0.702	0.701	0.029	0.028
0.75	0.877	0.877	0.016	0.016
0.90	0.956	0.956	0.007	0.007
<i>frailty dist: IG – estimator: model-based IG</i>				
0.10	0.356	0.356	0.074	0.073
0.25	0.543	0.537	0.051	0.051
0.50	0.710	0.705	0.022	0.022
0.75	0.843	0.842	0.007	0.007
0.90	0.929	0.929	0.002	0.002
<i>frailty dist: PS – estimator: model-based PS</i>				
0.10	0.524	0.524	0.056	0.055
0.25	0.586	0.586	0.046	0.046
0.50	0.699	0.700	0.030	0.030
0.75	0.830	0.830	0.014	0.014
0.90	0.923	0.923	0.005	0.005

Part II

Heterogeneity in multicentre trials

Adjusting for centre heterogeneity in multicentre clinical trials with a time-to-event outcome

Clinical trials are conducted at multiple centres for two main reasons: to accrue the required number of patients in a relatively short period of time and to broaden the scope of the results. Considerable efforts are made to standardise the way the trial is conducted in each centre according to the study protocol. The fact remains, however, that patients from different centres usually have different prognosis due to differences between centres (differences in disease diagnosis, differences in referral patterns, differences in indications for background therapies, etc.). In statistical terms, variability in outcome between patients within an individual centre tends to be lower than variability in outcome between patients at different centres (centre heterogeneity). For more reading on that and related issues, see Localio et al. (2001).

The International Conference on Harmonisation (ICH) guidance document E9 “Statistical Principles for Clinical Trials” (Lewis, 1999) clearly states that “*The main treatment effect may be investigated first using a model which allows for centre differences, but does not include a term for the treatment-to-centre interaction.*” For multicentre clinical trials with a time-to-event endpoint, however, recommendations on how to adjust for centre heterogeneity are limited. Inclusion of fixed effects and stratification are common methods to adjust for risk factors. Frailty models are not yet widely used in practice. Robustness properties against frailty misspecification have to be studied to build confidence in the frailty model.

In the setting of multicentre clinical trials, the fixed effects, stratified, and frailty approaches to estimating the treatment effect are reviewed and contrasted in Glidden & Vittinghoff (2004). Based on that paper, and along the lines of Chu et al. (2011), Kahan & Morris (2013), and

Kahan (2014) where the cases of continuous and binary outcomes are discussed, this chapter aims at providing guidelines for the practising statistician in the pharmaceutical industry. Using simulations, the performance of the frailty modelling approach over competing methods is illustrated. Special attention is paid to the problem of frailty misspecification. The central question addressed in this chapter is whether the frailty approach is the method to be recommended, considering the fact that frailty misspecification is the rule, rather than the exception.

In Section 4.1, we review the basics of the fixed effects, stratified, and frailty approaches. Section 4.2 highlights the limitations of ignoring centre heterogeneity as well as the pros and the cons of the aforementioned modelling strategies to adjust for centre heterogeneity. To illustrate this discussion, we use numerical results from a simulation study. We further investigate the performance of the frailty model when the frailty distribution is misspecified in Section 4.3. Section 4.4 summarises the conclusions and presents our recommendations.

The material presented in this chapter is published in Munda & Legrand (2014a).

4.1 --- Modelling clustered time-to-event data

We start with non-clustered time-to-event data for which the observed information consists of

$$\mathbf{z} = \left\{ (y_j, \delta_j, \mathbf{x}_j) \mid j = 1, \dots, N \right\}$$

where $y_j = \min(t_j, c_j)$ is the time to event or censoring, whichever comes first, $\delta_j = I(t_j \leq c_j)$ indicates whether an observation corresponds to an event ($\delta_j = 1$) or is censored ($\delta_j = 0$), and $\mathbf{x}_j$ is a vector of covariates. We make the standard assumption of independent and non-informative censoring (cf. Section 1.2.2).

Let $h_j(t)$ denote the hazard rate of subject j at time t . The (unadjusted) Cox model is written as

$$h_j(t) = h_0(t) \exp(\mathbf{x}_j' \boldsymbol{\beta}) \quad (4.1)$$

with $h_0(\cdot)$ a non-specified baseline hazard function and $\boldsymbol{\beta} = (\beta_1 \dots \beta_p)'$ a vector of fixed effects parameters. See Section 1.2.6 for more details.

Model (4.1) requires independent (homogeneous) data up to measured covariates. In multicentre clinical trial data, however, there is likely to be heterogeneity across centres. To account for centre heterogeneity, centre effects must somehow be included in the statistical model used for the analysis. Below, the fixed effects, stratified, and frailty approaches to estimating the treatment effect in the presence of centre heterogeneity are described in a nutshell. For more details, see Glidden & Vittinghoff (2004).

The fixed effects Cox model

Centre effects can enter model (4.1) as additional fixed effects parameters,

$$h_{ij}(t) = h_0(t) \exp(\mathbf{c}'_i \boldsymbol{\alpha} + \mathbf{x}'_{ij} \boldsymbol{\beta}) \quad (4.2)$$

where we now use two indices, $i \in \{1, \dots, s\}$ for the s centres and $j \in \{1, \dots, n_i\}$ for the n_i patients in centre i , to reflect the hierarchical structure of the data (the vector of observations $\mathbf{z}$ is changed accordingly). In model (4.2), $\boldsymbol{\alpha} = (\alpha_1 \dots \alpha_{s-1})'$ contains the fixed centre effects, and $\mathbf{c}_i$ denotes the vector with a 1 in the i^{th} position and 0's elsewhere ($i = 1, \dots, s-1$). The last centre does not need an indicator because an observation is known to belong to that centre when $\mathbf{c}_i = (0 \dots 0)'$. If we had included an additional indicator for the last centre, then the model would have been overparametrised. Choosing one particular centre as reference is consistent with the interpretation of $h_0(\cdot)$ as being the hazard rate of subjects with covariate values all equal to 0. However, this choice is arbitrary, and any centre can play the role of the reference centre.

The stratified Cox model

Instead of entering the centre variable as additional fixed effects parameters, the baseline hazard can be stratified on that variable to indicate that different subpopulations are exposed to different baseline risks, i.e.

$$h_{ij}(t) = h_{0i}(t) \exp(\mathbf{x}'_{ij} \boldsymbol{\beta}) \quad (4.3)$$

where $h_{01}(\cdot), \dots, h_{0s}(\cdot)$ are unspecified and unrelated baseline hazard functions. The partial likelihood approach (cf. Section 1.2.6) is readily adapted by multiplying the partial likelihoods specific to each stratum.

The use of model (4.3) to adjust for centre heterogeneity is recommended in O’Quigley & Stare (2002).

The frailty model

Participating centres may also be viewed as one possible sample from a broader population of centres. In that case, centre i has a random effect, called frailty and denoted by u_i , on the hazard rate. The frailty term reflects different levels of risk across centres. The (shared) frailty model is defined as

$$h_{ij}(t) = h_0(t)u_i \exp(\mathbf{x}'_{ij}\boldsymbol{\beta}) \quad (4.4)$$

The u_i ’s are the actual values of a random variable U with probability density $f(\cdot)$, called the frailty distribution. The (one-parameter) gamma distribution (cf. Section 1.1 of the thesis appendix) is the most commonly used. In that case, $E(U) = 1$ and $\text{Var}(U) = \theta$. The frailty parameter determines the degree of heterogeneity between centres.

Software

We use `coxph()` in R (part of the `survival` library) to fit models (4.1)–(4.4). For a detailed description of the proper use of `coxph()` for frailty models, see Therneau & Grambsch (2000, Chapter 9). Generic sample codes are provided in Appendix 4.A.

4.2 Comparison

In this section, we discuss the strengths and the weaknesses of models (4.2)–(4.4) to adjust for centre heterogeneity. This discussion is illustrated with simulations. Because the unadjusted model (4.1) is often used in practice, we consider this model as well.

4.2.1 Simulation setting

We consider two opposite situations with 6 centres of size 48 ($N = 6 \times 48$) and 48 centres of size 6 ($N = 6 \times 48$) as well as an intermediate situation with 8 centres of size 18 plus 24 centres of size 6 ($N = 8 \times 18 + 24 \times 6$),

thus keeping the total sample size fixed at $N = 288$. We mimic a 1:1 (respectively 2:1) allocation ratio in each centre by selecting $N/2$ (respectively $2N/3$) patients for the treatment arm ($x = 1$) and the remaining $N/2$ (respectively $N/3$) patients for the control arm ($x = 0$).

The event time for each patient is generated from model (4.4) using a Weibull baseline hazard ($h_0(t) = \lambda \rho t^{\rho-1}$) and a gamma frailty term with variance θ ; cf. the thesis appendix (Appendix 2). We take $\lambda = 0.7$, $\rho = 1.5$, $\theta = 0.5$ (Kendall's $\tau = 0.20$), and $\beta = \log(2/3) \approx -0.405$ or $\beta = 0$. The between-centre heterogeneity induced by this parameter setting is shown in Figure 4.1 by the spread in the median time to event from centre to centre (Duchateau & Janssen, 2005). The censoring time for each patient is generated from an exponential distribution with rate parameter chosen so that 30% of the observations are censored. Additional simulation results (varying the censoring rate and Kendall's τ) are available at <http://onlinelibrary.wiley.com/doi/10.1002/pst.1612/supinfo>.

For each setting, we fit models (4.1)–(4.4) to $K = 10000$ simulated data sets. For model (4.4), we use the correctly specified gamma frailty distribution (the impact of misspecification is addressed separately in Section 4.3). We report

- $\overline{\text{HR}}$: the average estimate of the hazard ratio, i.e. $\frac{1}{K} \sum_k \exp(\hat{\beta}_k)$;
- SD: the standard deviation of the $\hat{\beta}_k$'s;
- CI cov: the empirical coverage of the asymptotic 95% confidence interval based on the normal approximation, i.e. the proportion of such confidence intervals that cover the true value of β ;
- power/size: the empirical rejection rate for the null hypothesis of no treatment effect ($H_0: \beta = 0$)—under H_0 , it equals 1 minus the empirical coverage probability.

Note: power

Note that by disregarding the clustering under this parameter setting, one will expect to have approximately a 80% chance of declaring a hazard ratio of $HR = 2/3$ to be significant at the 5% level. Indeed, power calculation for the Cox model can be done by means of the following formula (see, e.g., Collett, 2003, Section 10.2.1)

$$\text{power} \approx \Phi \left(z_{\alpha/2} - \log(HR) \sqrt{d \pi (1 - \pi)} \right)$$

with

- $\Phi(\cdot)$: the standard normal cumulative distribution function;
- $z_{\alpha/2}$: the $\alpha/2$ quantile of the standard normal distribution;
- d : the expected number of events, i.e. total sample size $\times$ (1 – censoring rate);
- π : the proportion of individuals allocated to the control group.

Therefore, the power to detect a hazard ratio of $HR = 2/3$ at the 5% significance level with a sample size of 288 subjects and a censoring rate of 30% is of 0.82 under a 1:1 allocation ratio, and of 0.77 under a 2:1 allocation ratio.

Note: assessment of coverage

Let π_c be the true coverage probability and X the number of times the confidence interval covers β out of K replications; then $X \sim \text{Bin}(K, \pi_c)$. The empirical estimator $p_c = X/K$ has an asymptotic normal distribution with mean π_c and variance $\pi_c(1 - \pi_c)/K$ so that the width of its 95% confidence interval is approximately $2\sqrt{\pi_c(1 - \pi_c)/K}$, which is bounded from above by $\sqrt{1/K}$. With $K = 10000$, the width of that confidence interval is therefore at most equal to 0.01. Hence, empirical coverage probabilities below 0.945 correspond to under-coverage and empirical coverage probabilities above 0.955 correspond to over-coverage (Burton et al., 2006; Demirtas, 2007).

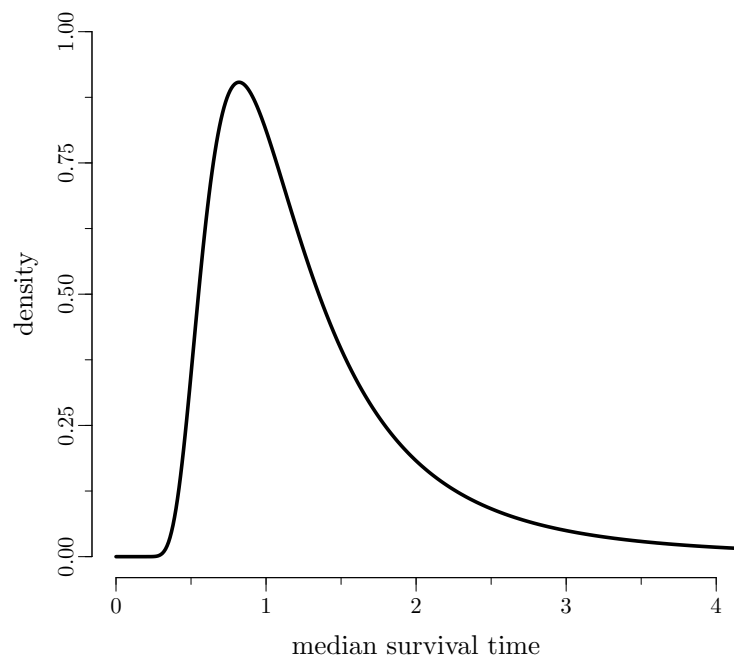


Figure 4.1: Density function of the median time to event from centre to centre (control group) in the setting of the simulation study.

4.2.2 Results and guidelines

The results are displayed in Table 4.1 ($\beta = \log(2/3) \approx -0.405$) and in Table 4.2 ($\beta = 0$).

The unadjusted Cox model

Model (4.1) makes no attempt to account for clustering. This alters the way the treatment effect ($\text{HR} = \exp(\beta)$) has to be interpreted. Indeed, HR has different meanings in model (4.1) (marginal model) and in models (4.2)–(4.4) (conditional models). In model (4.1), HR compares the hazard rates of two subjects, one treated and one untreated, randomly drawn from the population under study, regardless of where they come from (population-averaged interpretation). On the other hand, in conditional models (and in particular in model (4.4) used to generate the data), HR compares the hazard rates of two subjects, one treated and one untreated, randomly drawn from the same centre (centre-specific interpretation). Therefore, in our simulations, the unadjusted model estimates a quantity that is different from the target. In Table 4.1, we observe that the population-averaged effect is attenuated compared with the centre-specific effect. Under the null hypothesis of no treatment effect (Table 4.2), HR is well estimated (as there is no room for attenuation), but it can be seen from the type I error rate that ignoring the clustering leads to results that are too conservative. For more general results regarding the omission of important risk factors from non-linear regression models, see Hauck et al. (1998).

The fixed effects Cox model

Model (4.2) requires maximisation over a $(p + s - 1)$ -parameter space, with p the number of parameters in β (here, $p = 1$). This is numerically challenging whenever the number of centres, s , is large relative to the total sample size. The fixed effects approach therefore performs poorly for $s = 8 + 24$ and for $s = 48$. It produces estimates that are biased away from the true β , and the coverage of the confidence interval (respectively the type I error rate) is below 95% (respectively above 5%). Regarding multicentre clinical trials, the fixed effects approach further shows additional limitations. (i) It implicitly assumes that the centres participating in the trial are by themselves of interest. Inference is to be made for those centres only, and conclusions are thus restricted in scope. (ii) It

Table 4.1: Simulation results for the treatment effect, with $\beta = \log(2/3) \approx -0.405$, under correct specification of the frailty distribution.

sample size	statistic	model [†]			
		(4.1)	(4.2)	(4.3)	(4.4)
1:1					
6 × 48	$\overline{\text{HR}}$	0.731	0.668	0.673	0.674
	SD	0.133	0.148	0.149	0.145
	CI cov	0.929	0.948	0.954	0.953
	power	0.617	0.809	0.784	0.798
8 × 18 +24 × 6	$\overline{\text{HR}}$	0.742	0.645	0.674	0.674
	SD	0.127	0.173	0.168	0.153
	CI cov	0.918	0.910	0.944	0.944
	power	0.582	0.814	0.701	0.772
48 × 6	$\overline{\text{HR}}$	0.743	0.627	0.675	0.675
	SD	0.123	0.184	0.173	0.150
	CI cov	0.925	0.890	0.954	0.955
	power	0.572	0.825	0.656	0.764
2:1					
6 × 48	$\overline{\text{HR}}$	0.730	0.668	0.673	0.674
	SD	0.141	0.158	0.159	0.154
	CI cov	0.939	0.945	0.948	0.950
	power	0.580	0.768	0.736	0.755
8 × 18 +24 × 6	$\overline{\text{HR}}$	0.740	0.644	0.674	0.674
	SD	0.134	0.177	0.172	0.158
	CI cov	0.935	0.917	0.954	0.952
	power	0.545	0.780	0.665	0.732
48 × 6	$\overline{\text{HR}}$	0.742	0.628	0.675	0.675
	SD	0.132	0.195	0.182	0.160
	CI cov	0.933	0.893	0.951	0.950
	power	0.535	0.781	0.620	0.719

[†] model (4.1): unadjusted Cox model; model (4.2): fixed effects Cox model; model (4.3): stratified Cox model; model (4.4): semi-parametric gamma frailty model.

Table 4.2: Simulation results for the treatment effect, with $\beta = \log(1) = 0$, under correct specification of the frailty distribution.

sample size	statistic	model [†]			
		(4.1)	(4.2)	(4.3)	(4.4)
1:1					
6 × 48	$\overline{\text{HR}}$	1.007	1.009	1.010	1.009
	SD	0.128	0.149	0.151	0.146
	size	0.031	0.055	0.052	0.052
8 × 18 +24 × 6	$\overline{\text{HR}}$	1.010	1.016	1.015	1.013
	SD	0.124	0.170	0.164	0.150
	size	0.024	0.074	0.050	0.047
48 × 6	$\overline{\text{HR}}$	1.011	1.020	1.017	1.014
	SD	0.123	0.187	0.173	0.152
	size	0.022	0.090	0.049	0.051
2:1					
6 × 48	$\overline{\text{HR}}$	1.008	1.011	1.012	1.011
	SD	0.134	0.157	0.158	0.154
	size	0.028	0.053	0.049	0.048
8 × 18 +24 × 6	$\overline{\text{HR}}$	1.007	1.014	1.013	1.010
	SD	0.129	0.178	0.171	0.157
	size	0.022	0.074	0.050	0.048
48 × 6	$\overline{\text{HR}}$	1.006	1.017	1.016	1.011
	SD	0.130	0.198	0.183	0.162
	size	0.024	0.094	0.050	0.053

[†] model (4.1): unadjusted Cox model; model (4.2): fixed effects Cox model; model (4.3): stratified Cox model; model (4.4): semi-parametric gamma frailty model.

provides neither a summary measure of heterogeneity between centres nor a convenient framework to test for the presence of centre effects (Andersen et al., 1999). (iii) It might be of interest to assess whether a covariate explains heterogeneity in outcome between centres (Legrand et al., 2006). It is, however, unfeasible in this model to include a covariate whose values only change at the centre level. (iv) Precision in centre effects estimates is dependent upon the centre size. Interpretation can therefore be misleading. A related problem is that the centre effects estimates (and their interpretation) depend on the choice of the reference centre, which is generally arbitrary.

The stratified Cox model

Model (4.3) performs well, with good point estimates and good coverage probabilities. However, no between-centre comparisons are made by the stratified approach (information about β only comes from within-centre comparisons). In particular, centres where subjects have the same covariate values and centres with no event do not contribute any information. Therefore, the stratified approach does not make optimal use of all the information at hand. This explains why both the standard deviation inflates and the power deteriorates when the centre size decreases. Besides, similar to the fixed effects approach, (i) interpretation of the treatment effect is restricted to participating centres, (ii) no heterogeneity measure is returned, and (iii) centre-specific covariates cannot be investigated because no between-centre comparisons are made by the stratified approach.

The frailty model

Model (4.4) shows good performances in every investigated setting with virtually no bias and good coverage probabilities. Unlike the stratified model, the frailty model also makes use of between-centre comparisons to gather information on the treatment effect. This explains why both the standard deviation is smaller and the power is better for the frailty model than for the stratified model. Compared to the fixed effects approach, the number of parameters in the frailty model does not increase with the number of centres. The frailty modelling approach further provides a rich framework for the analysis of multicentre clinical trials. (i) Because of their random nature, the actual values of the frailty term

(i.e. the centre effects for those centres participating in the trial) are not of intrinsic interest, and the conclusions of the study are intended to be generalised more broadly to all hospitals represented by the sample at hand. (ii) The variance of the gamma frailty distribution, θ , is a key parameter that determines the degree of heterogeneity between centres. To help interpretation, this parameter can further be translated into clinically relevant quantities like the spread in the median time to event (as we did in Figure 4.1) or in the 5-year survival rate from centre to centre (Duchateau & Janssen, 2005). Alternatively, the θ parameter can be transformed into the Kendall's τ that measures the degree of association between outcomes within the same centre (cf. Section 3.2.1). For gamma frailties, Kendall's τ is $\tau = \theta/(\theta + 2)$. (iii) Considering the u_i 's as random effects parameters also makes it possible to study whether the inclusion of a centre-specific covariate explains/reduces heterogeneity between centres (Legrand et al., 2006).

4.3

Robustness against frailty misspecification

Different distributions can be used to model the frailty term. Diagnostic checks to assess the frailty distribution are not yet widely available (particularly in software), and research is still needed in this area (cf. Chapter 3). In the meantime, it is important to investigate robustness properties against frailty misspecification via simulations.

The most common assumption is that the frailties have a gamma distribution. Therefore, the most common form of misspecification is that of using the gamma distribution while the frailties actually follow another distribution. Alternative distributions that have received interest to model the frailty term include the inverse Gaussian (IG), the log-normal (LN), and the positive stable (PS).

To observe the impact of misspecifying the frailty distribution on the inference for the treatment effect (and more generally for the fixed effects parameters included in the model), we simulate data from model (4.4) (cf. Section 4.2.1) using the IG, LN, and PS distributions to generate the frailties, and we fit the misspecified gamma frailty model. For each frailty distribution, the heterogeneity parameter is chosen to yield a Kendall's τ of $\tau = 0.20$, as earlier.

By comparing the results obtained under misspecification with those

obtained under correct specification (Table 4.3 and Table 4.4), it appears that inferences on the fixed effect parameter β are robust against misspecification of the frailty distribution. In particular, the frailty approach performs better than the competing stratified approach in terms of power in either misspecified situation.

4.4

Conclusions

In clinical trials with a time-to-event outcome, the primary analysis is commonly based on model (4.1) with a single covariate for the treatment effect, or on the equivalent log-rank test. When the trial is conducted at multiple centres, the treatment effect obtained from model (4.1) has a population-averaged (marginal) interpretation, the effect being averaged over all centres. Model (4.1) leads to a consistent estimate of the population treatment effect, but the standard error is not consistent and a robust estimator that copes with the clustering should be used (cf. the `cluster()` argument in `coxph()`); see Glidden & Vittinghoff (2004) and Duchateau & Janssen (2008, Section 3.4). In contrast, the treatment effect obtained from models (4.2)–(4.4) has a centre-specific (conditional) interpretation. In the context of multicentre clinical trials, the centre-specific treatment effect interpretation is particularly relevant as it compares “like-for-like”.

Important conclusions from our simulation study are as follows:

- The population-averaged effect is attenuated compared with the centre-specific effect;
- Ignoring the clustering leads to results that are too conservative;
- The centre-specific treatment effect is usually biased when it is estimated from the fixed effects Cox model (4.2);
- In some settings, power is lost when fitting the stratified Cox model (4.3) compared with the frailty model (4.4);
- Inferences on the centre-specific treatment effect obtained from the frailty model (4.4) are robust against misspecification of the frailty distribution in many settings.

In the light of these results, we recommend to use the frailty model, which is now readily available in standard software (e.g. R and SAS), to adjust for centre heterogeneity in multicentre clinical trials with a time-to-event outcome.

Table 4.3: Simulation results for the treatment effect, with $\beta = \log(2/3) \approx -0.405$, under both correct (gamma) and incorrect specifications of the frailty distribution (allocation ratio: 1:1).

sample size	statistic	true frailty distribution											
		Gam				IG				LN			
		(4.3)	(4.4)	(4.3)	(4.4)	(4.3)	(4.4)	(4.3)	(4.4)	(4.3)	(4.4)	(4.3)	(4.4)
6×48	$\overline{\text{HR}}$	0.673	0.674	0.672	0.673	0.672	0.673	0.672	0.673	0.673	0.674	0.674	0.675
	SD	0.149	0.145	0.151	0.146	0.151	0.147	0.151	0.147	0.150	0.150	0.150	0.146
	CI cov	0.954	0.953	0.947	0.948	0.952	0.953	0.952	0.953	0.952	0.952	0.952	0.950
	power	0.784	0.798	0.784	0.806	0.778	0.793	0.778	0.793	0.778	0.778	0.778	0.793
8×18 $+24 \times 6$	$\overline{\text{HR}}$	0.674	0.674	0.673	0.674	0.673	0.674	0.673	0.674	0.674	0.675	0.675	0.676
	SD	0.168	0.153	0.166	0.152	0.165	0.151	0.165	0.151	0.167	0.167	0.151	0.151
	CI cov	0.944	0.944	0.950	0.948	0.951	0.950	0.951	0.950	0.952	0.952	0.950	0.950
	power	0.701	0.772	0.710	0.774	0.707	0.778	0.707	0.778	0.695	0.695	0.765	0.765
48×6	$\overline{\text{HR}}$	0.675	0.675	0.674	0.675	0.674	0.675	0.674	0.676	0.675	0.675	0.676	0.676
	SD	0.173	0.150	0.172	0.152	0.176	0.154	0.176	0.154	0.174	0.174	0.152	0.152
	CI cov	0.954	0.955	0.954	0.950	0.950	0.948	0.950	0.948	0.951	0.951	0.950	0.950
	power	0.656	0.764	0.660	0.758	0.654	0.756	0.654	0.756	0.654	0.654	0.757	0.757

[†] model (4.3): stratified Cox model; model (4.4): semi-parametric gamma frailty model.

Table 4.4: Simulation results for the treatment effect, with $\beta = \log(1) = 0$, under both correct (gamma) and incorrect specifications, of the frailty distribution (allocation ratio: 1:1).

		true frailty distribution											
		Gam			model [†]						PS		
					IG			LN					
sample size	statistic	(4.3)	(4.4)	(4.3)	(4.3)	(4.4)	(4.4)	(4.3)	(4.3)	(4.4)	(4.3)	(4.3)	(4.4)
6×48	$\overline{\text{HR}}$	1.010	1.009	1.010	1.010	1.009	1.012	1.012	1.012	1.011	1.010	1.010	1.010
	SD	0.151	0.146	0.150	0.150	0.145	0.149	0.149	0.149	0.144	0.148	0.148	0.144
	size	0.052	0.052	0.054	0.054	0.050	0.048	0.048	0.048	0.044	0.051	0.051	0.050
8×18 $+24 \times 6$	$\overline{\text{HR}}$	1.015	1.013	1.012	1.012	1.010	1.010	1.018	1.018	1.015	1.014	1.014	1.012
	SD	0.164	0.150	0.166	0.166	0.151	0.165	0.165	0.165	0.150	0.162	0.162	0.149
	size	0.050	0.047	0.055	0.055	0.052	0.054	0.054	0.054	0.049	0.049	0.049	0.048
48×6	$\overline{\text{HR}}$	1.017	1.014	1.015	1.015	1.012	1.012	1.015	1.015	1.011	1.014	1.014	1.012
	SD	0.173	0.152	0.172	0.172	0.152	0.172	0.172	0.172	0.153	0.170	0.170	0.152
	size	0.049	0.051	0.049	0.049	0.051	0.047	0.047	0.047	0.051	0.045	0.045	0.049

[†] model (4.3): stratified Cox model; model (4.4): semi-parametric gamma frailty model.

Appendix 4.A

Sample codes

Models (4.1)–(4.4) can be fitted in R by means of `coxph()` (part of the `survival` library) or in SAS by means of `proc phreg`. In the generic sample codes below, `data` has the following columns:

- **cluster**: cluster identification number;
- **time**: minimum between the actual event time and the censoring time;
- **status**: 1 if the observation is an event, 0 if it is right-censored;
- **x**: treatment group indicator (0 or 1).

R codes

```
# unadjusted Cox model
coxph(Surv(time, status) ~ x, data=data)

# fixed effects Cox model
coxph(Surv(time, status) ~ x + factor(cluster), data=data)

# stratified Cox model
coxph(Surv(time, status) ~ x + strata(cluster), data=data)

# semi-parametric gamma frailty model
# !!! 'sparse=(nclass > 5)' !!!
# cf. Therneau & Grambsch, 2000, Section 9.7
coxph(Surv(time, status) ~ x +
      frailty.gamma(x=cluster, eps=1e-11),
      outer.max=50, data=data)
```

SAS codes

```
/* unadjusted Cox model */
proc phreg data=data;
  class x(ref="0");
  model time*status(0) = x / ties=efron;
run;

/* fixed effects Cox model */
proc phreg data=data;
  class x(ref="0") cluster(ref="1");
  model time*status(0) = x cluster / ties=efron;
run;

/* stratified Cox model */
```

```
proc phreg data=data;
    class x(ref="0");
    model time*status(0) = x / ties=efron;
    strata cluster;
run;

/* semi-parametric log-normal frailty model */
/* !!! The gamma frailty distribution is not yet available !!! */
/* A likelihood reformulation method for non-normal random effects,
   with sample codes in proc nlmixed, is described in Liu & Yu
   (2008) */
proc phreg data=data;
    class x(ref="0") cluster;
    model time*status(0) = x / ties=efron;
    random cluster / method=REML;
run;
```

Testing for decreasing heterogeneity between hospitals in time to death from chronic myeloid leukemia

In the standard frailty model, the frailty term is assumed to be constant over time. For some particular studies, this assumption is known to be questionable. For example, the cancer clinical trial on chronic myeloid leukemia (CML) analysed in Wintrebert et al. (2004) is one such study. Patients first receive bone marrow transplantation, whence they are at risk of death due to transplant-related causes. Within that period, substantial heterogeneity between transplant centres can be foreseen. Thereafter, heterogeneity is expected to decrease.

In this chapter, we relax the time-constant heterogeneity assumption and consider frailty models with a time-varying frailty term. Instead of working with hazard models, we rather model the log cumulative hazard function, making use of the mixed model framework, and introduce a time-varying random effect at that level.

Section 5.1 presents a literature review and lays out the framework of this chapter. In Section 5.2, we introduce the time-varying frailty model and we show how to estimate the fixed effects, how to predict the time-varying random effect, and how to test for time-varying heterogeneity using the linear mixed model methodology. The proposed method is illustrated in Section 5.3 with the CML data mentioned above. In Section 5.4, we run simulations to assess the performance of the method. Concluding remarks are given in Section 5.5.

This chapter is submitted for publication (Munda et al., 2014).

5.1

Literature review and outline

Much of the work on time-varying frailties has concerned recurrent event time data. In Yau & McGilchrist (1998), the log frailty term is allowed to vary according to a first-order autoregressive (AR(1)) process to describe serial dependence between recurrent event times. A generalisation of the frailty model estimation of McGilchrist (1993), which is based on (restricted) maximum likelihood, is presented. The model is illustrated with recurrent infection data on chronic granulomatous disease patients. A related approach for the proportional odds model is developed in Lam et al. (2002). The model of Yau & McGilchrist (1998) is applied within the Bayesian framework in Manda & Meyer (2005). In Wang et al. (2007), the model is augmented with an additional (time-constant) shared frailty term to accommodate situations involving both clusters and recurrent events. This multilevel model is illustrated with data on recurrent urinary tract infections among elderly women residing in aged-care institutions. In Yue & Chan (1997), the frailty term is assumed to have a two-parameter gamma distribution, with parameters updated at each recurrence according to the past information. In essence, the updated scheme in Yue & Chan (1997) modifies the AR(1) process for the case of gamma random variables. This specification leads to a closed-form expression for the marginal likelihood. This model is further studied in Fong et al. (2001). The Bayesian model in Pennell & Dunson (2006) allows the frailty term to vary between time intervals, using Dirichlet priors to allow for uncertainty in the gamma frailty distribution.

For univariate survival data (no cluster), time-varying frailty models based on Lévy processes are studied in Gjessing et al. (2003).

In the context of multicentre clinical trials, differences between centres, e.g. in terms of practice patterns or patient management, may have an important influence on study outcomes at the beginning of follow-up, but less so afterwards, especially when patients have recovered sufficiently to leave the hospital. To model time-varying centre effects on survival after bone marrow transplantation in CML patients, a shared frailty model with a piecewise constant frailty term (Paik et al., 1994) is proposed in Wintrebert et al. (2004). To be specific, the frailty of cluster i in interval k is taken to be the sum of independent two-parameter gamma random variables, $U_{ik} = U_i + E_{ik}$. The cluster-specific frailty

component U_i models heterogeneity between clusters. Within cluster i , the independent interval-specific frailty components (the E_{ik} 's) allow the overall frailty term to vary over time. To simplify the model estimation, which is based on maximum likelihood, two alternatives are proposed. The first one sets $U_{ik} = U_i^{\gamma_k}$, with U_i log-normal and γ_k an unknown interval-specific parameter. The second one assumes that the log frailty term is a linear function of time with random centre-specific intercepts and slopes, $\log(U_i(t)) = a_i + b_i t$ where $(a_i, b_i)'$ follows a bivariate normal distribution. Application to the CML data indicates that heterogeneity between transplant centres decreases over time.

In this chapter, we propose an alternative model to study time-varying centre heterogeneity in large-scale multicentre clinical trials. Following Massonnet et al. (2008), Teodorescu et al. (2010), and López-de-Ullibarri et al. (2012), survival models can be based on the logarithm of the cumulative hazard. The key idea is that, on the logarithmic scale, the cumulative hazard function has a linear model structure. Random effects can then be introduced at that level, leading to a linear mixed effects model. Provided that enough data information is available, complex data structures can be studied in that framework. Furthermore, linear mixed effects models can be handled by standard software, e.g. `proc mixed` in SAS.

To model decreasing (or, more generally, time-varying) heterogeneity between clusters on a time-to-event endpoint, we therefore propose to model the logarithm of the estimated cumulative hazard as a linear mixed effects model with a time-varying random effect.

5.2

Time-varying frailty model

The focus on this chapter is on heterogeneity in large-scale multicentre clinical trial data. We consider the following data structure. Observations are partitioned into s clusters and each cluster is divided into two groups by a dichotomous covariate x (e.g., a treatment indicator or a prognostic biomarker whose value is used to partition the population into two risk groups). Let $h_{ij}(t)$ denote the hazard rate at time t for individual j of cluster i ($j = 1, \dots, n_i$; $i = 1, \dots, s$). The frailty model is defined as

$$h_{ij}(t) = h_0(t)u_i \exp(x_{ij}\beta)$$

$$= h_0(t) \exp(w_i + x_{ij}\beta) \quad (5.1)$$

where $h_0(\cdot)$ is a baseline hazard function, x_{ij} is the value of the covariate (0 or 1), β is the unknown fixed effect parameter, and $w_i = \log(u_i)$ is the log frailty, hereafter called random effect, of cluster i . An alternative formulation of model (5.1) in terms of the cumulative hazard function is

$$H_{ij}(t) = H_0(t) \exp(w_i + x_{ij}\beta)$$

Since we use the linear mixed model methodology, a convenient choice for the distribution of the w_i 's is the normal distribution with mean 0 and variance θ .

The extension that we propose is of form

$$H_{ij}(t) = H_0(t) \exp(w_i(t) + x_{ij}\beta) \quad (5.2a)$$

That is, the w_i 's are time-dependent at the cumulative hazard level.

Note: model constraint

In order for $H_{ij}(\cdot)$ in model (5.2a) to be a cumulative hazard function, we need to ensure that

1. $\lim_{t \rightarrow 0} H_{ij}(t) = 0$
2. $\lim_{t \rightarrow \infty} H_{ij}(t) = \infty$
3. $\frac{dH_{ij}(t)}{dt} \geq 0$ for all t

The first two conditions are satisfied. Condition 3 can be rewritten as

$$\frac{dH_0(t)}{dt} + H_0(t) \frac{dw_i(t)}{dt} \geq 0$$

On the one hand, if $w_i(\cdot)$ is constant or increasing, this condition is satisfied. On the other hand, if $w_i(\cdot)$ is decreasing, then condition 3 is that $\log(H_0(\cdot))$ increases faster than $w_i(\cdot)$ decreases.

The hazard function derived from model (5.2a) is

$$h_{ij}(t) = \left(h_0(t) + H_0(t) \frac{dw_i(t)}{dt} \right) \exp(w_i(t) + x_{ij}\beta)$$

Therefore, the way we extend the frailty model is different from extensions where the w_i 's are time-dependent in (5.1) (cf., e.g., Wintrebert

et al. (2004)). It is interesting to note that $\exp(\beta)$ still has a conditional hazard ratio interpretation.

Using a logarithmic transformation on both sides of (5.2a), we have

$$\log(H_{ij}(t)) = \log(H_0(t)) + w_i(t) + x_{ij}\beta \quad (5.2b)$$

Model (5.2b) provides the bridge between the survival model (5.2a) and the linear mixed effects model (see Section 5.2.1).

Note: hazard model versus cumulative hazard model

An alternative extension is

$$h_{ij}(t) = h_0(t) \exp(w_i(t) + x_{ij}\beta) \quad (5.3a)$$

i.e., the w_i 's are time-dependent at the hazard level. On the logarithmic scale, (5.3a) also has a linear model structure,

$$\log(h_{ij}(t)) = \log(h_0(t)) + w_i(t) + x_{ij}\beta \quad (5.3b)$$

To arrive at linear mixed effects models, we need “pseudo-responses” which will serve as responses in (5.2b) and (5.3b). In this chapter, the pseudo-responses are obtained non-parametrically. In order to make this as simple as possible, we work with (5.2b) rather than with (5.3b).

5.2.1 Towards the linear mixed effects model

Let T_{ij} be the event time of individual j from cluster i , possibly right-censored by the random variable C_{ij} ($j = 1, \dots, n_i$; $i = 1, \dots, s$). We make the standard assumptions that the event times and the censoring times are independent conditional on the frailty term and the covariate (independent censoring) and that the censoring distribution has no common parameter with the event time distribution (non-informative censoring). For each i and j , the random variables that we observe are $Y_{ij} = \min(T_{ij}, C_{ij})$ and $\Delta_{ij} = I(T_{ij} \leq C_{ij})$, together with the covariate $x_{ij} \in \{0, 1\}$. We denote by $H_i^{(k)}(\cdot)$ the cumulative hazard function common to all individuals of cluster i with $x_{ij} = k$ ($k = 0, 1$).

To arrive at a linear mixed effects model, we need “pseudo-responses” which will serve as responses in model (5.2b). In group k of cluster i , the

pseudo-responses $\hat{\phi}_{ik,\ell}$ are obtained by estimating $\phi_{ik,\ell} := \log(H_i^{(k)}(t_\ell))$ on a fixed grid of time points t_ℓ ; $\ell = 1, \dots, L$. The grid of time points can often be based on clinical or biological grounds. Practical considerations regarding the choice of the t_ℓ 's are further discussed for the CML data in Section 5.3.

Using the data from cluster i , $\{(y_{ij}, \delta_{ij}, x_{ij}) \mid j = 1, \dots, n_i\}$, we take

$$\hat{\phi}_{ik,\ell} = \log\left(\hat{H}_i^{(0)}(t_\ell) \exp(k\hat{\gamma}_i)\right)$$

with $\hat{H}_i^{(0)}(\cdot)$ an estimator of $H_i^{(0)}(\cdot)$ and $\hat{\gamma}_i$ the fixed effect parameter estimate obtained from the Cox model fitted to the data of cluster i . For $\hat{H}_i^{(0)}(\cdot)$, we propose a modification of the Breslow cumulative baseline hazard curve by combining this non-parametric estimator with a parametric fit to estimate $H_i^{(0)}(t_\ell)$ for t_ℓ below the first or above the last event time of cluster i .

— **Note: a hybrid estimator of $H_0(\cdot)$** —

Given a sample of independent survival data (up to a vector of measured covariates $\mathbf{x}$), the Breslow estimator of the cumulative hazard function is defined by

$$\hat{H}_{0B}(t) = \sum_{\tilde{y}_{(\ell)} \leq t} \frac{d_\ell}{\sum_{j \in R(\tilde{y}_{(\ell)})} \exp(\mathbf{x}'_j \hat{\beta})}$$

with $\tilde{y}_{(1)} < \dots < \tilde{y}_{(r)}$ the ordered distinct event times, d_ℓ the number of events at time $\tilde{y}_{(\ell)}$, and $R(\tilde{y}_{(\ell)})$ the set containing those individuals still under observation just prior to $\tilde{y}_{(\ell)}$. This estimator suffers from two limitations: (i) it returns zero below $\tilde{y}_{(1)}$ (regarding the logarithmic transformation, this is a problem), and (ii) it remains constant beyond $\tilde{y}_{(r)}$ (estimates are thus not reliable in the right tail). We therefore propose to complete $\hat{H}_{0B}(\cdot)$ using a parametric distribution to estimate the tails (Moeschberger & Klein, 1985)

$$\hat{H}_0(t) = \begin{cases} \hat{\lambda} t^{\hat{\rho}} & \text{if } t < \tilde{y}_{(1)} \\ \hat{H}_{0B}(t) & \text{if } \tilde{y}_{(1)} \leq t \leq \tilde{y}_{(r)} \\ \hat{H}_{0B}(\tilde{y}_{(r)}) + \hat{\lambda} (t^{\hat{\rho}} - \tilde{y}_{(r)}^{\hat{\rho}}) & \text{if } t > \tilde{y}_{(r)} \end{cases}$$

with $\hat{\lambda}$ and $\hat{\rho}$ obtained by fitting a Weibull distribution to the data where $\mathbf{x} = \mathbf{0}$ under the constraint that $\hat{\lambda} \tilde{y}_{(1)}^{\hat{\rho}} = \hat{H}_{0B}(\tilde{y}_{(1)})$.

In terms of the pseudo-responses, we have

$$\hat{\phi}_{ik,\ell} = \beta_{0,\ell} + w_{i,\ell} + k\beta + e_{ik,\ell} \quad (5.4)$$

with $\beta_{0,\ell} := \log(H_0(t_\ell))$, $w_{i,\ell} := w_i(t_\ell)$, and $e_{ik,\ell} := \hat{\phi}_{ik,\ell} - \phi_{ik,\ell}$.

Model (5.4) is a linear mixed effects model with the following assumptions:

- $\mathbf{w} = (w_{1,1} \dots w_{1,L} \dots \dots w_{s,1} \dots w_{s,L})' \sim N(\mathbf{0}, \mathbf{G})$
- $\mathbf{e} = (e_{10,1} \ e_{11,1} \dots e_{10,L} \ e_{11,L} \dots \dots e_{s0,1} \ e_{s1,1} \dots e_{s0,L} \ e_{s1,L})' \sim N(\mathbf{0}, \mathbf{R})$
- $\text{Cov}(\mathbf{w}, \mathbf{e}) = \mathbf{0}$

where the forms of the $\mathbf{G}$ and $\mathbf{R}$ matrices are specified in Section 5.2.2 and in Section 5.2.3, respectively. For more details on the model structure, we refer the reader to Appendix 5.A where we give the matrix formulation for model (5.4).

5.2.2 Random effects covariance structure

Random effects pertaining to different clusters are assumed to be independent so that

$$\text{Cov}(w_{i,\ell}, w_{i',\ell'}) = 0 \quad \text{if } i \neq i'$$

Consequently, the $\mathbf{G}$ matrix has a block diagonal structure, with s blocks of size $L \times L$. We further assume that the blocks are identical. The common block matrix specifies the covariance structure between the L random effects within the same cluster. Many different choices can be made (Littell et al., 2000). When the t_ℓ 's are equally spaced, a popular choice is the AR(1) structure, i.e.

$$\text{Cov}(w_{i,\ell}, w_{i,\ell'}) = \theta \rho^{|\ell-\ell'|} \quad (\theta > 0, 0 < \rho < 1)$$

In words, the between-cluster variability is the same at any point while association between pairs of random effects within the same cluster declines with increasing distance in time. However, the fact that the variance remains constant over time makes it impossible to assess whether the cluster-to-cluster variability is changing over time. For that purpose, an autoregressive structure with heterogeneous variances (ARH(1)) is more appropriate, i.e.

$$\text{Cov}(w_{i,\ell}, w_{i,\ell'}) = \sqrt{\theta_\ell} \sqrt{\theta_{\ell'}} \rho^{|\ell-\ell'|}$$

5.2.3 Residual covariance structure

In model (5.4), the error term $e_{ik,\ell}$ adjusts for the fact that we have substituted an estimate for $\phi_{ik,\ell}$. The $\mathbf{R} = \text{Var}(\mathbf{e}) = \text{Var}(\hat{\boldsymbol{\phi}} | \mathbf{w})$ matrix specifies the covariance structure of the error terms. By construction, this information cannot be recovered from the pseudo-data (the latter do not contain any replicate for given i , k , and ℓ). In addition to the pseudo-responses, therefore, the raw data also have to provide an estimate of $\mathbf{R}$. This can be done by bootstrap resampling.

The fact that $\hat{\phi}_{i0,\ell}$ and $\hat{\phi}_{i1,\ell}$ ($\ell = 1, \dots, L$) are constructed using only the data from cluster i results in a block diagonal structure for $\mathbf{R}$, with s blocks of size $2L \times 2L$. In addition, block i differs from block i' because clusters are generally not identical in terms of sample size and event rate. We obtain an estimate of $\text{Cov}(e_{ik,\ell}, e_{ik',\ell'})$ as follows. First, we draw B bootstrap samples from cluster i of the original data (Davison & Hinkley, 1997, Algorithm 7.2, page 351). For each bootstrap sample, we calculate the pseudo-responses $\hat{\phi}_{ik,\ell}^*$ and $\hat{\phi}_{ik',\ell'}^*$. An estimate of $\text{Cov}(e_{ik,\ell}, e_{ik',\ell'})$ then follows from the empirical covariance between the $\hat{\phi}_{ik,\ell}^*$'s and the $\hat{\phi}_{ik',\ell'}^*$'s.

5.2.4 Fitting the model

Model (5.4) can be fitted in SAS by means of `proc mixed` where the `parms` statement is used to fix the residual variance components.

sample code

```
proc mixed data=pseudoData method=reml scoring=8;
  class cluster timepoint x;
  model phi = timepoint x;
  random timepoint / subject=cluster type=arh(1);
  repeated / subject=cluster group=cluster type=un;
  parms / parmsdata=covPar hold=%eval(&length_covParG + 1) to
    &length_covPar;
run;

*'length_covPar' is the total number of variance components;
*'length_covParG' is the number of variance components in G;
```

5.2.5 Testing for decreasing heterogeneity

To see whether “a time-constant cluster-to-cluster variability” is a reasonable model assumption, we can test the null hypothesis

$$H_0: \theta_1 = \cdots = \theta_L$$

against the alternative

$$H_1: \exists \ell, \ell' \in \{1, \dots, L\} : \theta_\ell \neq \theta_{\ell'}$$

Under H_0 , each block of $\mathbf{G}$ has a homogeneous AR(1) structure. The AR(1) structure has 2 parameters (θ and ρ) as compared to $L+1$ parameters ($\theta_1, \dots, \theta_L$, and ρ) for the ARH(1) structure. Model comparison can be done via a (restricted) likelihood ratio test (Littell et al., 2006, Section A1.6).

For the CML study considered in Section 5.3, however, we are mostly interested in decreasing heterogeneity among clusters. In that case, we rather consider the ordered alternative

$$H_1: \theta_1 > \cdots > \theta_L$$

To account for the ordering specified under H_1 , we rely on the ordered heterogeneity family of tests (Rice & Gaines, 1994a,b). An ordered heterogeneity test can be used to convert almost any non-directional test into a directional one when a specific ordered test is not available or is thought to implement. The test statistic consists of combining a measure of evidence against H_0 with the independent ordering information specified under H_1 . In our case, the test statistic becomes $T_O = r_s(1 - p_{\text{LRT}})$, with p_{LRT} the p-value obtained from the (non-directional) likelihood ratio test, and r_s the Spearman’s rank correlation between the observed ranking of $\{\hat{\theta}_1, \dots, \hat{\theta}_L\}$ and the expected ranking under the alternative. The test statistic T_O becomes increasingly large as the data increasingly refute the null hypothesis in the direction of the alternative hypothesis (Rice & Gaines, 1994a). Critical values are tabulated in Rice & Gaines (1994b).

5.3 Example

We consider the CML data used in Wintrebert et al. (2004). To obtain good pseudo-responses, sufficient data information is required within

each cluster. In the CML data, many centres are quite small (number of patients per centre: min = 1, 1st quartile = 3, median = 10, mean = 19.43, 3rd quartile = 23.5, max = 280) and substantial data information is censored ($\approx 60\%$). We thus consider a subset of the data, leaving a sample of 1767 CML patients from 19 centres among which 706 deaths ($\approx 40\%$) were recorded (cf. Table 5.1).

We account for the effects of five known risk factors (patient age, disease stage, time interval from diagnosis to transplant, donor type, and donor-recipient sex combination). To avoid the “curse of dimensionality”, we incorporate the covariate information into the EBMT risk score (Gratwohl, 2012), a validated prognostic index ranging from 0 to 7 points, from which we identify a “low-risk group” (EBMT risk score = 0, 1, 2, 3) and a “high-risk group” (EBMT risk score = 4, 5, 6, 7). By fitting the standard frailty model (5.1) with the normal random effect distribution we find a hazard ratio of $\widehat{\text{HR}} = \exp(\hat{\beta}) = 2.140$ (95% CI: [1.827, 2.507]) and $\hat{\theta} = 0.025$.

To investigate whether the centre-to-centre variability decreases over follow-up time, we fit model (5.4). Pseudo-responses are calculated on a grid of $L = 5$ time points equally spaced between $t_1 = 1$ month and $t_5 = 9$ months, which is approximately the time it takes for a patient to recover and to produce normal blood cell levels.

Regarding the fixed effect obtained from model (5.4), we find $\widehat{\text{HR}} = \exp(\hat{\beta}) = 2.116$ (95% CI: [1.693, 2.643]), similar to what we have obtained with the standard frailty model.

Figure 5.1 displays the empirical best linear unbiased predictions (EBLUPs) of the random effects under the ARH(1) specification of the $\mathbf{G}$ matrix (cf. Section 5.2.2). The value of the variance at the different time points (i.e. the diagonal elements of each block of $\mathbf{G}$) are given at the bottom of Figure 5.1.

The likelihood ratio statistic (AR(1) versus ARH(1)) equals 17.655. As convergence to the limiting null chi-squared distribution is slow (this point is further discussed in the simulations below), we rather make use of the finite-sample distribution obtained by bootstrapping the data under H_0 (cf. Appendix 3). We find $p_{\text{LRT}} = 0.038$. The correlation coefficient between the observed ranking (5, 4, 3, 2, 1) and the expected ranking (5, 4, 3, 2, 1) equals $r_s = 1$. Thus, $T_O = 0.962$. For $L = 5$, the critical region C at the 5% significance level is $C = \{T_O > 0.509\}$. We therefore reject H_0 and we conclude that the data display declining

Table 5.1: Repartition of the patients in the CML data. In parentheses are the number of events.

centre	low-risk $x = 0$	high-risk $x = 1$	total
1	31 (8)	19 (7)	50 (15)
2	41 (13)	11 (3)	52 (16)
3	52 (12)	6 (6)	58 (18)
4	48 (15)	11 (4)	59 (19)
5	53 (9)	7 (12)	60 (21)
6	49 (17)	12 (6)	61 (23)
7	52 (13)	10 (11)	62 (24)
8	53 (16)	14 (8)	67 (24)
9	52 (16)	16 (9)	68 (25)
10	41 (23)	27 (3)	68 (26)
11	55 (20)	15 (8)	70 (28)
12	60 (20)	12 (8)	72 (28)
13	51 (21)	21 (9)	72 (30)
14	66 (16)	7 (16)	73 (32)
15	67 (27)	14 (9)	81 (36)
16	85 (29)	21 (10)	106 (39)
17	121 (42)	64 (41)	185 (83)
18	145 (61)	78 (32)	223 (93)
19	206 (82)	74 (44)	280 (126)
total	1328 (460)	439 (246)	1767 (706)

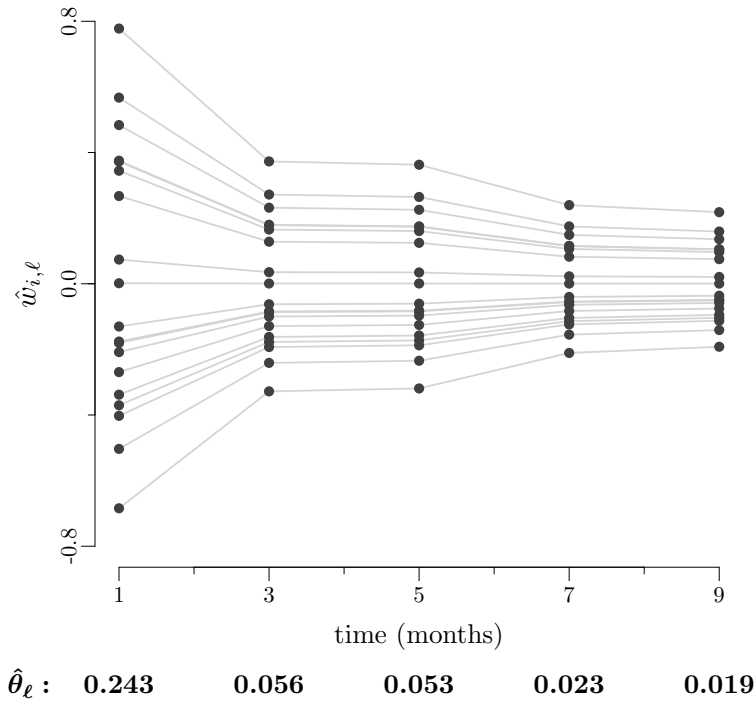


Figure 5.1: Empirical best linear unbiased predictions (EBLUP's) as a function of time in the CML study. The evolution of the variance over time is shown below the x -axis.

centre heterogeneity. Figure 5.1 indicates that there is heterogeneity between centres during the first few months following transplantation, but not much afterwards.

Different time points, varying in number and position, are considered in Figure 5.2 and Figure 5.3. It appears that the choice of the particular time points is not critical, provided that sufficient data information is available for each cluster during the considered time window. On the one hand, t_1 should not be taken too small to ensure enough variability in the pseudo-responses at t_1 . On the other hand, t_L should not be taken too large to avoid results driven by the particular distribution used to model the upper tail. A suitable choice of t_1 and t_L can be determined by

inspecting the Kaplan-Meier curve stratified by cluster (cf. Figure 5.4).

5.4 --- Simulation study

To provide further insight into the method, we have conducted a simulation study in the context of a large-scale multicentre clinical trial.

We consider 10 centres of size 200 or 400. Patients within a centre are randomly assigned to one of two treatment groups with an allocation ratio of 1:1. Three scenarios are examined:

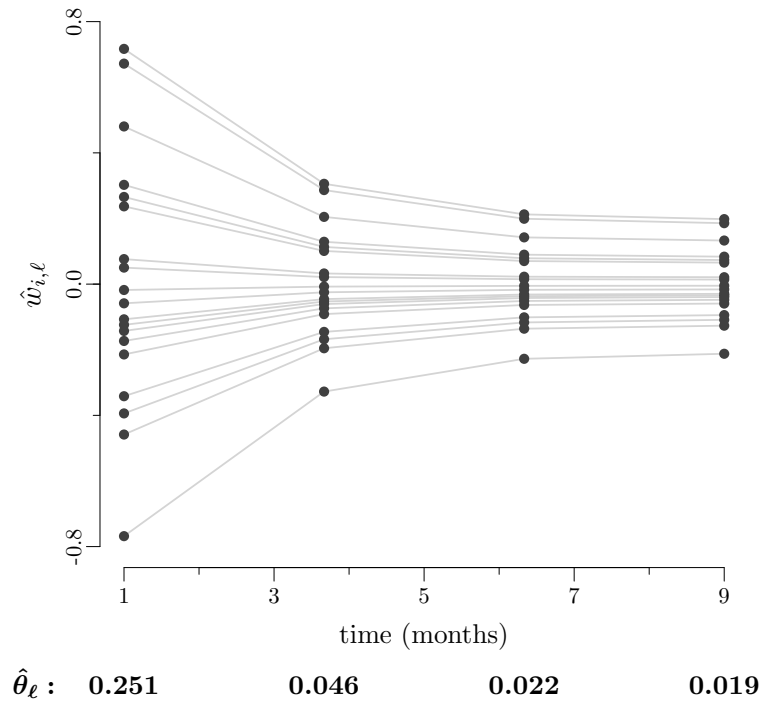
- Scenario 1: event times are generated from model (5.1), i.e. the w_i 's are time-constant;
- Scenario 2: event times are generated from model (5.2a), i.e. the w_i 's are time-dependent at the cumulative hazard level;
- Scenario 3: event times are generated from model (5.3a), i.e. the w_i 's are time-dependent at the hazard level.

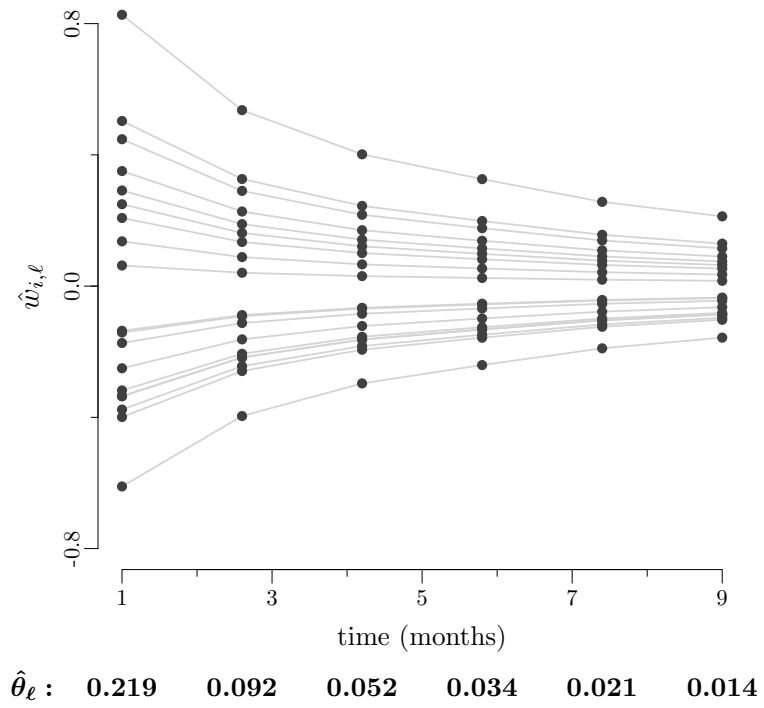
In all three scenarios, we take $\beta = \log(2) \approx 0.693$ and we use a Weibull distribution at baseline ($h_0(t) = \lambda \rho t^{\rho-1}$) with $\lambda = 0.005$ and $\rho = 2$. Censoring times are generated from an exponential distribution with rate parameter chosen to control the amount of censoring. We consider moderate censoring (20%) and high censoring (60%).

5.4.1 Generation of event times

To generate t_{ij} , the event time for patient j of centre i , we use the fact that $S_{ij}(T_{ij}) \sim U(0, 1)$ and we solve $S_{ij}(t_{ij}) = u$ for t_{ij} , with u a uniform variate. The forms of $S_{ij}(t)$ in the three scenarios are

- Scenario 1: $S_{ij}(t) = \exp \left\{ - \lambda t^\rho \exp(w_i + x_{ij}\beta) \right\}$
- Scenario 2: $S_{ij}(t) = \exp \left\{ - \lambda t^\rho \exp(w_i(t) + x_{ij}\beta) \right\}$
- Scenario 3: $S_{ij}(t) = \exp \left\{ - \int_0^t \lambda \rho v^{\rho-1} \exp(w_i(v) + x_{ij}\beta) dv \right\}$

Figure 5.2: Same as Figure 5.1 with $L = 4$.

Figure 5.3: Same as Figure 5.1 with $L = 6$.

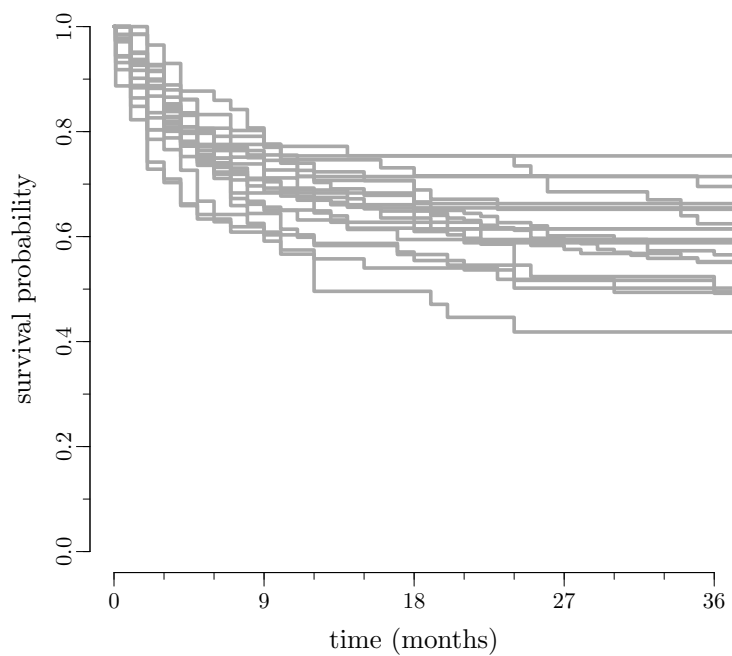


Figure 5.4: Kaplan-Meier curve (restricted to the first 3 years) stratified by transplant centre showing that most deaths occur during the first 9 months.

In scenario 1, w_i is randomly drawn from a $N(0, \theta)$ distribution, with $\theta = 0.56$ or $\theta = 1.06$. To help interpretation, the θ parameter can be transformed into the Kendall's τ that measures the degree of association between event times in a centre. For $\theta = 0.56$ and $\theta = 1.06$, this transformation gives $\tau \approx 0.20$ and $\tau \approx 0.30$, respectively (cf. Appendix 3.B).

In scenarios 2 and 3, the specific time-dependency in the random effects that we consider is depicted in Figure 5.5. The time-dependency dies out in time ($w_i(t) \rightarrow 0$ as $t \rightarrow \infty$). The actual value of the random effect in cluster i at time $t = 0$, $w_i(0)$, is sampled from a $N(0, \theta)$ distribution. We take $\theta = 0.56$ and $\theta = 1.06$, as above. Given the random start, the way the time-dependency dies out is deterministic. For details, we refer the reader to Appendix 5.B where we explain how Figure 5.5 is obtained.

5.4.2 Results

For each of 500 simulations, we have fitted model (5.4) with both the AR(1) and ARH(1) specifications of the $\mathbf{G}$ matrix. For the ARH(1) specification, we have recorded the EBLUP's of the random effects and the estimated variances over time. Figures 5.6–5.8 display averages at each time point for one particular parameter setting in the three scenarios (results are similar for the other parameter settings). The estimates of the fixed effect parameter β were also obtained (cf. Table 5.3).

The density of the likelihood ratio statistic is depicted in Figure 5.9 for increasing sample sizes. The likelihood ratio statistic converges to the limiting chi-squared distribution with $(L + 1) - 2$ degrees of freedom. Convergence is slow, though. Therefore, in practice, we recommend using the finite-sample distribution obtained by bootstrapping the data under the null hypothesis (cf. Appendix 3).

In the context of these simulations, to avoid prohibitively long simulation times, we have determined the null distribution of the test statistic T_O by simulations, using supplementary simulated data sets, rather than by bootstrap. The rejection rates for the ordered heterogeneity test are given in Table 5.2. Important conclusions are as follows:

- The rejection rate under the alternative hypothesis of scenarios 2 and 3, i.e. the power of the test, is generally reasonably high;

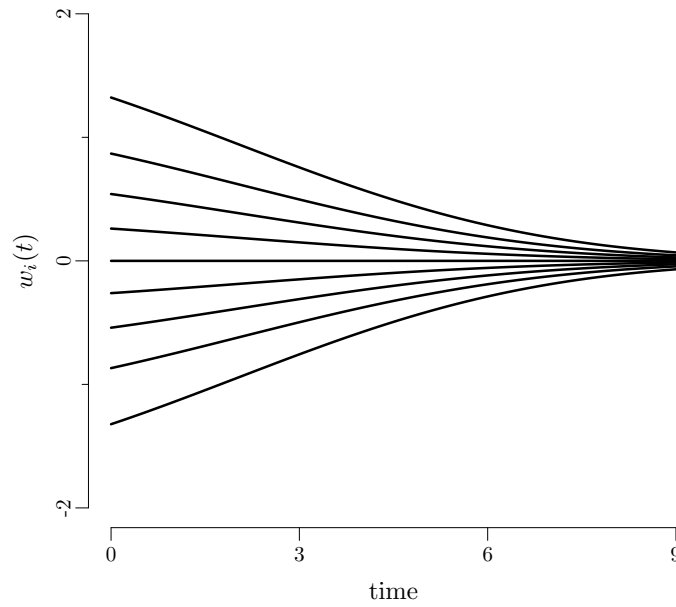


Figure 5.5: Declining heterogeneity between clusters as used in the simulation study. In this picture, the initial values correspond to the deciles of the $N(0, \theta)$ distribution, with $\theta = 1.06$. In the simulations, the initial values are randomly drawn from the $N(0, \theta)$ distribution, with $\theta = 0.56$ or $\theta = 1.06$.

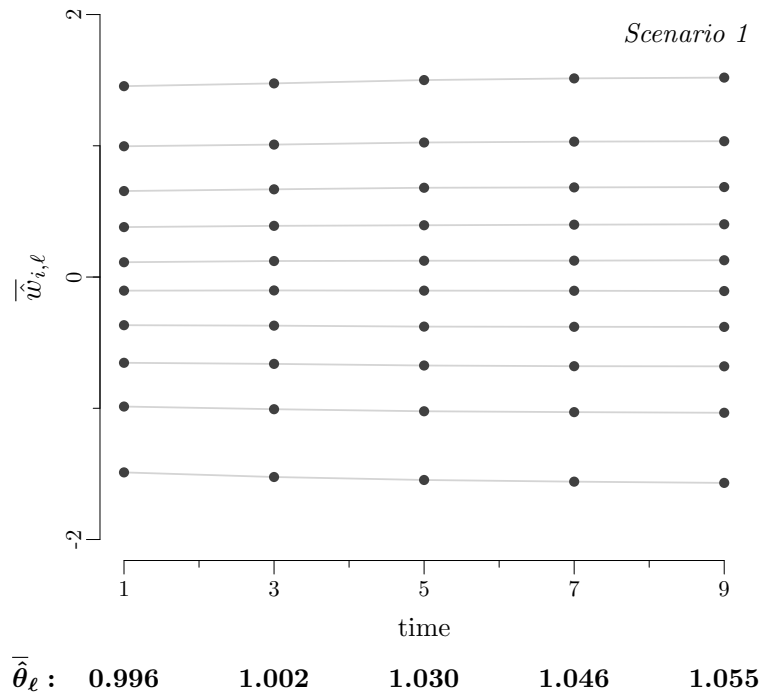


Figure 5.6: Simulation results in scenario 1 with cluster size = 400, censoring rate = 60%, and $\theta = 1.06$.

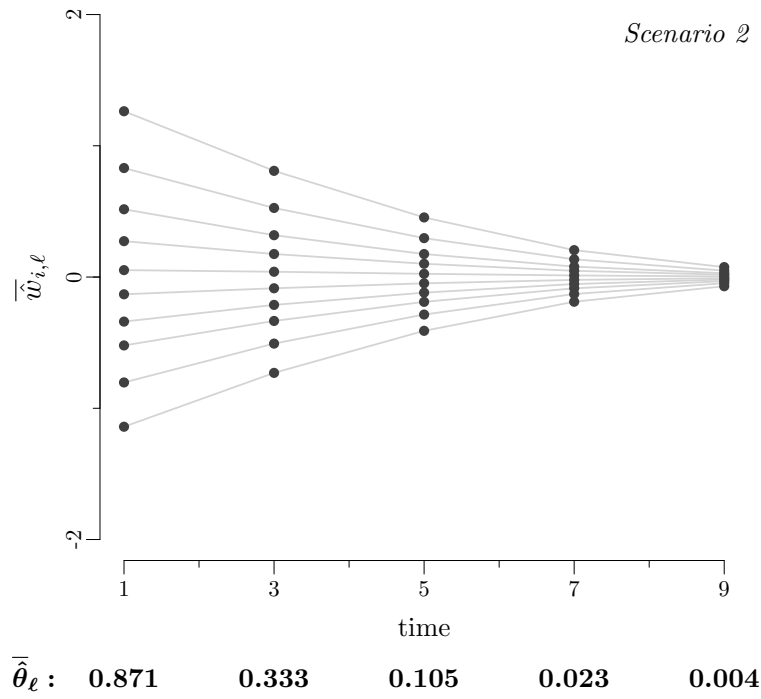


Figure 5.7: Simulation results in scenario 2 with cluster size = 400, censoring rate = 60%, and $\theta = 1.06$.

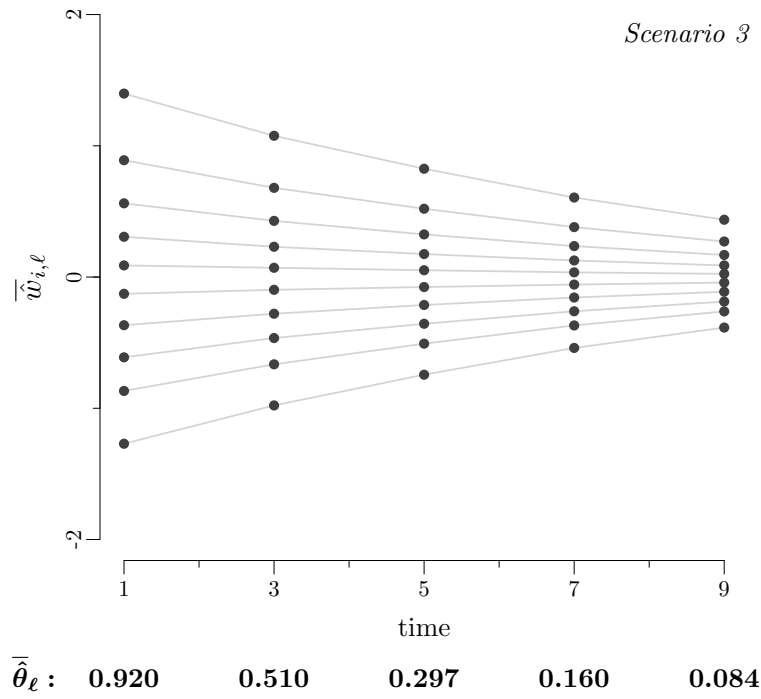


Figure 5.8: Simulation results in scenario 3 with cluster size = 400, censoring rate = 60%, and $\theta = 1.06$.

- The power decreases with decreasing data information (i.e. decreasing cluster size and/or increasing censoring);
- For θ (initial variance) decreasing, also the power is decreasing;
- The power remains reasonably high even under model misspecification (cf. scenario 3).

5.5

Discussion

We have proposed a method to test for decreasing heterogeneity in clustered survival data by extending the frailty model to a time-varying frailty model. Starting from the cumulative hazard representation, the method embeds the model into the linear mixed model world via a logarithmic transformation. Pseudo-responses for the linear mixed model are derived by estimating the log cumulative hazard on a grid of time points.

Within each cluster, sufficient data information is needed for accurate estimation of the pseudo-responses. Therefore, the method is intended for applications that involve big clusters, e.g. large-scale multicentre clinical trials. For the same reason, we recommend the use of a prognostic index to handle multiple covariates. For a continuous prognostic index, the Breslow estimator used to compute the pseudo-responses can be replaced by a kernel-based estimator (see López-de-Ullibarri et al., 2012).

We have illustrated the method with the same CML data as in Wintrebert et al. (2004) and confirmed the conclusion that heterogeneity between transplant centres declines over time. Of note, the way we extend the frailty model is different from extensions given in Wintrebert et al. (2004) where the random effects are made time-dependent in the hazard function. Compared to the latter, the proposed method is flexible in specifying the correlation structure between the random effects, avoids complex likelihood functions that are difficult to maximize, and offers the ability to use standard mixed model software.

More discussion is provided at the end the thesis (closing part).

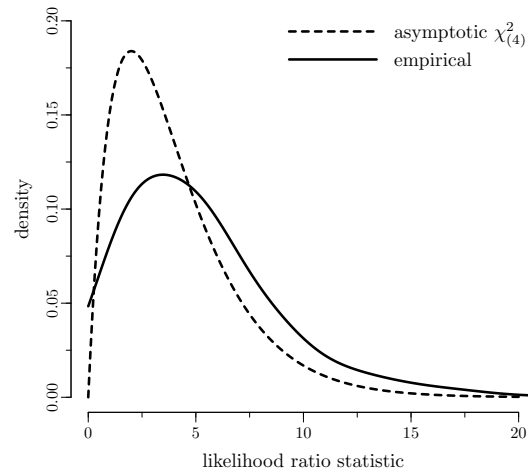
Table 5.2: Simulation results – Ordered heterogeneity test.

θ	cens.	size	rejection rate	
<i>scenario 1</i>				
0.56	20%	200	0.045	
		400	0.060	
	60%	200	0.065	
		400	0.070	
1.06	20%	200	0.032	
		400	0.065	
	60%	200	0.045	
		400	0.035	
<i>scenario 2</i>				
0.56	20%	200	0.630	
		400	0.813	
	60%	200	0.623	
		400	0.783	
1.06	20%	200	0.791	
		400	0.946	
	60%	200	0.776	
		400	0.907	
<i>scenario 3</i>				
0.56	20%	200	0.591	
		400	0.780	
	60%	200	0.570	
		400	0.765	
1.06	20%	200	0.754	
		400	0.953	
	60%	200	0.838	
		400	0.942	

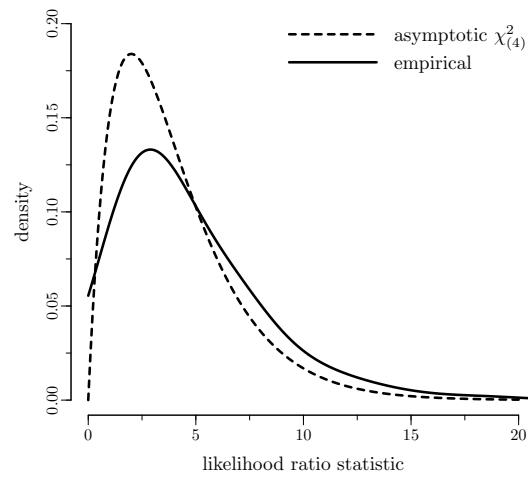
Table 5.3: Simulation results – Fixed effect parameter.

θ	cens.	size	$\hat{\beta}$ median	se($\hat{\beta}$) median	95% CI(β) coverage
<i>scenario 1</i>					
0.56	20%	200	0.686	0.053	0.954
		400	0.688	0.037	0.966
	60%	200	0.686	0.075	0.948
		400	0.688	0.052	0.950
1.06	20%	200	0.688	0.053	0.940
		400	0.691	0.037	0.946
	60%	200	0.680	0.075	0.964
		400	0.686	0.052	0.952
<i>scenario 2</i>					
0.56	20%	200	0.697	0.053	0.952
		400	0.693	0.037	0.940
	60%	200	0.682	0.075	0.944
		400	0.684	0.052	0.948
1.06	20%	200	0.686	0.053	0.974
		400	0.692	0.037	0.952
	60%	200	0.684	0.074	0.944
		400	0.689	0.052	0.948
<i>scenario 3</i>					
0.56	20%	200	0.694	0.053	0.944
		400	0.692	0.037	0.928
	60%	200	0.687	0.075	0.944
		400	0.687	0.052	0.942
1.06	20%	200	0.686	0.053	0.956
		400	0.693	0.037	0.956
	60%	200	0.676	0.074	0.966
		400	0.688	0.052	0.954

The true β equals $\log(2) \approx 0.693$.



(a) Cluster size = 200.



(b) Cluster size = 400.

Figure 5.9: Finite-sample distribution of the likelihood ratio statistic under H_0 (censoring rate = 60%; $\theta = 0.56$).

Appendix 5.A

Model (5.4) in matrix notation

In matrix form, the linear mixed effects model is written as

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{w} + \mathbf{e} \\ \mathbf{w} &\sim N(\mathbf{0}, \mathbf{G}) \\ \mathbf{e} &\sim N(\mathbf{0}, \mathbf{R}) \\ \text{Cov}(\mathbf{w}, \mathbf{e}) &= \mathbf{0} \end{aligned}$$

with $\mathbf{Y}$ the vector of responses, $\boldsymbol{\beta}$ and $\mathbf{w}$ the vectors of fixed and random effects parameters, $\mathbf{X}$ and $\mathbf{Z}$ the corresponding design matrices, and $\mathbf{e}$ the vector of random errors.

Below is the matrix form of model (5.4) for the particular case where $s = 3$ and $L = 2$:

$$\begin{pmatrix} \hat{\phi}_{10,1} \\ \hat{\phi}_{11,1} \\ \hat{\phi}_{10,2} \\ \hat{\phi}_{11,2} \\ \vdots \\ \hat{\phi}_{20,1} \\ \hat{\phi}_{21,1} \\ \hat{\phi}_{20,2} \\ \hat{\phi}_{21,2} \\ \vdots \\ \hat{\phi}_{30,1} \\ \hat{\phi}_{31,1} \\ \hat{\phi}_{30,2} \\ \hat{\phi}_{31,2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} \beta_{0,1} \\ \beta_{0,2} \\ \vdots \\ \beta \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} w_{1,1} \\ w_{1,2} \\ \vdots \\ w_{2,1} \\ w_{2,2} \\ \vdots \\ w_{3,1} \\ w_{3,2} \end{pmatrix} + \begin{pmatrix} e_{10,1} \\ e_{11,1} \\ e_{10,2} \\ e_{11,2} \\ \vdots \\ e_{20,1} \\ e_{21,1} \\ e_{20,2} \\ e_{21,2} \\ \vdots \\ e_{30,1} \\ e_{31,1} \\ e_{30,2} \\ e_{31,2} \end{pmatrix}$$

$$\underbrace{\begin{pmatrix} \hat{\phi}_{10,1} \\ \vdots \\ \hat{\phi}_{31,2} \end{pmatrix}}_{\mathbf{Y}} = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}}_{\mathbf{X}} \underbrace{\begin{pmatrix} \beta_{0,1} \\ \beta_{0,2} \\ \vdots \\ \beta \end{pmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}}_{\mathbf{Z}} \underbrace{\begin{pmatrix} w_{1,1} \\ w_{1,2} \\ \vdots \\ w_{2,1} \\ w_{2,2} \\ \vdots \\ w_{3,1} \\ w_{3,2} \end{pmatrix}}_{\mathbf{w}} + \underbrace{\begin{pmatrix} e_{10,1} \\ e_{11,1} \\ e_{10,2} \\ e_{11,2} \\ \vdots \\ e_{20,1} \\ e_{21,1} \\ e_{20,2} \\ e_{21,2} \\ \vdots \\ e_{30,1} \\ e_{31,1} \\ e_{30,2} \\ e_{31,2} \end{pmatrix}}_{\mathbf{e}}$$

In general, we have

$$\mathbf{X} = \begin{pmatrix} \mathbf{1}_s \otimes \mathbf{I}_L \otimes \mathbf{1}_2 & \vdots & \mathbf{1}_s \otimes \mathbf{1}_L \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} \end{pmatrix}$$

and

$$\mathbf{Z} = \mathbf{I}_s \otimes \mathbf{I}_L \otimes \mathbf{1}_2$$

with $\mathbf{I}_n$ the $n \times n$ identity matrix and $\mathbf{1}_n$ the vector of length n with all entries equal to 1.

Continuing the particular case where $s = 3$ and $L = 2$, $\mathbf{G} = \text{Var}(\mathbf{w})$ and $\mathbf{R} = \text{Var}(\mathbf{e}) = \text{Var}(\mathbf{Y} | \mathbf{w})$ are block diagonal matrices,

$$\mathbf{G} = \text{diag}(\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3) \quad \text{and} \quad \mathbf{R} = \text{diag}(\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3)$$

with $\mathbf{G}_i$ and $\mathbf{R}_i$ the symmetric matrices

$$\mathbf{G}_i = \begin{pmatrix} \text{Var}(w_{i,1}) & \text{Cov}(w_{i,1}, w_{i,2}) \\ & \text{Var}(w_{i,2}) \end{pmatrix}$$

and

$$\mathbf{R}_i = \begin{pmatrix} \text{Var}(e_{i0,1}) & \text{Cov}(e_{i0,1}, e_{i1,1}) & \text{Cov}(e_{i0,1}, e_{i0,2}) & \text{Cov}(e_{i0,1}, e_{i1,2}) \\ & \text{Var}(e_{i1,1}) & \text{Cov}(e_{i1,1}, e_{i0,2}) & \text{Cov}(e_{i1,1}, e_{i1,2}) \\ & & \text{Var}(e_{i0,2}) & \text{Cov}(e_{i0,2}, e_{i1,2}) \\ & & & \text{Var}(e_{i1,2}) \end{pmatrix}$$

The covariance matrix of $\mathbf{Y} = \hat{\boldsymbol{\phi}}$, given by $\mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$, is thus a block diagonal matrix with block i is given by

$$\text{Var} \begin{pmatrix} \hat{\phi}_{i0,1} \\ \hat{\phi}_{i1,1} \\ \hat{\phi}_{i0,2} \\ \hat{\phi}_{i1,2} \end{pmatrix} = \begin{pmatrix} \theta_1 & \theta_1 & \sqrt{\theta_1}\sqrt{\theta_2}\rho & \sqrt{\theta_1}\sqrt{\theta_2}\rho \\ \theta_1 & \theta_1 & \sqrt{\theta_1}\sqrt{\theta_2}\rho & \sqrt{\theta_1}\sqrt{\theta_2}\rho \\ \sqrt{\theta_1}\sqrt{\theta_2}\rho & \sqrt{\theta_1}\sqrt{\theta_2}\rho & \theta_2 & \theta_2 \\ \sqrt{\theta_1}\sqrt{\theta_2}\rho & \sqrt{\theta_1}\sqrt{\theta_2}\rho & \theta_2 & \theta_2 \end{pmatrix} + \mathbf{R}_i$$

Appendix 5.B

Generation of time-varying random effects

For data generation purposes, we assume that the differential equation

$$\frac{dw_i(t)}{dt} = k_{1i} \exp \left\{ - \left(\frac{t - k_2}{k_3} \right)^2 \right\}$$

$$w_i(0) = w_{i0}$$

describes the evolution in time of $w_i(t)$, where k_{1i} is a parameter that governs the increasing/decreasing rate, and where k_2, k_3 are two additional tuning constants related to “where variation takes place and how long it lasts”. The solution of the differential equation is given by

$$w_i(t) = w_{0i} + k_{1i}k_3\sqrt{\pi} \left[\Phi \left(\sqrt{2} \frac{t - k_2}{k_3} \right) + \Phi \left(\sqrt{2} \frac{k_2}{k_3} \right) - 1 \right]$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of a standard normal random variable. Taking

$$k_{1i} = -\frac{w_{0i}}{k_3\sqrt{\pi}\Phi \left(\sqrt{2} \frac{k_2}{k_3} \right)}$$

we obtain that the cluster-to-cluster variability dies out in time ($w_i(t) \rightarrow 0$ as $t \rightarrow \infty$).

We now need to ensure that, for this choice of $w_i(t)$,

$$H_{ij}(t) = H_0(t) \exp(w_i(t) + x_{ij}\beta)$$

does result in a cumulative hazard curve, i.e.,

1. $\lim_{t \rightarrow 0} H_{ij}(t) = 0$
2. $\lim_{t \rightarrow \infty} H_{ij}(t) = \infty$
3. $\frac{dH_{ij}(t)}{dt} \geq 0$ for all t

The first two conditions are satisfied. Assuming a Weibull baseline hazard ($h_0(t) = \lambda \rho t^{\rho-1}$; $\lambda > 0, \rho > 0$), condition 3 can be rewritten as

$$\lambda t^\rho \exp(w_i(t) + x_{ij}\beta) \left(\frac{\rho}{t} + \frac{dw_i(t)}{dt} \right) \geq 0$$

from which it follows that we must have

$$\frac{\rho}{t} \geq -\frac{dw_i(t)}{dt}$$

for all $t > 0$ and $i \in \{1, \dots, s\}$. If $w_i(t)$ is constant or increasing, then the constraint always holds. Otherwise, this condition is most easily interpreted from model (5.2b),

$$\log(H_{ij}(t)) = \log(H_0(t)) + w_i(t) + x_{ij}\beta$$

In that case, the constraint means that the rate at which $\log(H_0(t))$ (whose derivative is $\frac{\rho}{t}$) increases must be larger than the rate at which $w_i(t)$ decreases in order for $\log(H_{ij}(t))$, or equivalently for $H_{ij}(t)$, to be increasing.

Part III

Spatial dependence

A first step towards modelling spatial dependence in time to malaria data using the frailty model methodology

Malaria is found in tropical and subtropical regions where meteorological and environmental conditions (temperature, humidity, altitude, rainfall, etc.) are favourable to the development of the parasite and to the survival of the malaria vector mosquito. The presence of dams may further increase the incidence of malaria by providing breeding habitats for the malaria mosquitoes.

To assess the effect of dams on *Plasmodium falciparum* malaria incidence, a study was conducted in 16 villages surrounding the Gilgel-Gibe hydroelectric dam reservoir, South-West Ethiopia (Yewhalaw et al., 2010, 2013). It has been postulated that children living far away from the dam are less at risk for malaria as compared to children living nearby (Yewhalaw et al., 2009).

A shared frailty model with time-to-malaria as response, distance to the dam as main covariate, and village as frailty term has been used to assess this hypothesis (Yewhalaw et al., 2013; Getachew et al., 2013). In the shared frailty model, observations within a village have the same frailty. The village-specific frailty term aims at taking into account the dependence of observations from the same village. With a shared frailty term, however, any two children within a village exhibit the same degree of association. In addition, observations from different villages are assumed to be independent (between-village independence assumption). Thus, the spatial dependence, i.e. the tendency of observations taken in close proximity to exhibit positive association, is not explicitly taken into account by the shared frailty model. As a first step, we focus in this chapter on modelling spatial dependence between villages.

In the frailty model framework, spatial dependence between observations can be accounted for by modelling spatial dependence between

frailties (Diggle & Ribeiro Junior, 2007, Section 4.3.3). In this chapter, we adopt this strategy to explore spatial dependence in the malaria data.

A review of the relevant literature is presented in Section 6.1. The spatial frailty model is introduced for the malaria data in Section 6.2. Section 6.3 provides some insight into how spatial dependence between frailties induce spatial dependence between survival times. The malaria data is analysed in Section 6.4. Directions for future developments are proposed in Section 6.5.

The work in this chapter is ongoing.

6.1

Literature review and outline

Geostatistical methods are mainly developed under the assumption that the distribution of the outcome variable is normal (Diggle & Ribeiro Junior, 2007; Littell et al., 2006, Chapter 11). In the setting of survival analysis, these methods have received scant attention. A few Bayesian and frequentist approaches have been proposed to model spatial dependence in survival data using the frailty model methodology (see below). Alternatively, a normal transformation marginal model (no frailty), whereby the event time is transformed into a normally distributed random variable, is proposed in Li & Lin (2006).

Henderson et al. (2002) investigate spatial variation in leukemia incidence across districts in North-west England within the Bayesian framework. The model is based on a hierarchical specification. Individual frailties within district i have a gamma distribution with a district-specific mean μ_i . Spatial dependence is then induced by assuming that the μ_i 's follow a multivariate normal distribution with unit mean and a spatially-structured variance-covariance matrix. The model reduces to the univariate gamma frailty model in the case of no spatial dependence.

Banerjee et al. (2003) consider a cluster-specific random effect with a conditionally autoregressive prior. That prior distribution incorporates information about the adjacency of the clusters (the neighbourhood structure). The model is applied to infant mortality data in Minnesota counties. Further applications can be found in Banerjee & Carlin (2003), Diva et al. (2007), Ojiambo & Kang (2013), and Jin & Carlin (2005). The latter model relies on the position of the clusters (counties) relative to each other ("lattice approach"). In addition, Banerjee et al. (2003)

also use the exact geographical locations of the clusters (“geostatistical approach”) by means of a cluster-specific log frailty term following a multivariate normal distribution with variance-covariance matrix specified as a function of distance. The model reduces to the log-normal shared frailty model in the case of no spatial dependence. Both the lattice and the geostatistical approaches are also discussed within the framework of the proportional odds model in Banerjee & Dey (2005) with an application to breast cancer data in Iowa counties; see also Diva et al. (2008).

The frequentist counterpart of the geostatistical approach to model spatial dependence between cluster-specific frailty terms is undertaken in Li & Ryan (2002). The method relies either on Monte Carlo simulations or on the Laplace method to approximate the marginal likelihood and is applied to childhood asthma data in Boston counties.

Based on a somewhat similar approximation as in Li & Ryan (2002), a two-stage procedure iterating between a weighted estimating equation to estimate the regression coefficients and a variogram (cf. Section 6.3) to estimate the variance components is developed in Lin (2012) for a study of forest decline (tree mortality). It should be noted however that the use of variograms for formal inference is discouraged (Diggle & Ribeiro Junior, 2007, Chapter 5).

In this chapter, we consider the frequentist geostatistical approach using the Laplace method to approximate the marginal likelihood. The Laplace method for frailty models is implemented in R (`coxph()`, part of the `survival` library) and in SAS (`proc phreg`) for independent frailties (McGilchrist, 1993; Therneau et al., 2003). Relaxation of the independence assumption between frailties is addressed in Ripatti & Palmgren (2000). The method has been used in Pankratz et al. (2005) to model dependent survival times among family members in a large familial cohort study of breast cancer (within a family, frailties are different, but correlated, due to different degrees of genetic association) and is implemented in the `coxme()` function in R (part of the `coxme` library). Possible correlation structures in `coxme()`, however, are targeted to genetic applications. In this chapter, as a first step towards modelling spatial dependence in the malaria data, we model spatial dependence between villages using the maximum likelihood approach of Ripatti & Palmgren (2000), similar to Li & Ryan (2002). Extension to spatial dependence both between villages and between households within a village requires further work; cf. Section 6.5.

6.2

Spatial frailty model

To study the risk of malaria around the Gilgel-Gibe dam, a cohort of children living in villages within a 10 km radius from the dam reservoir were monitored at weekly intervals over a period of two years (July 2008 – June 2010) and times to malaria were collected. Figure 6.1 is a location map of the study area. Details of the study design have been previously published (Yewhalaw et al., 2010, 2013).

In the study area, six periods (three rainy seasons per year and two study years) can be identified and modelled by a piecewise constant baseline hazard; $h_0(t) = \lambda_m$ in period $[a_m, a_{m+1})$, $m = 1, \dots, M$. The household distance to the dam, calculated as the shortest distance to the dam shore, is the risk factor of interest. Following Getachew (2013, Chapter 4), we decompose the household distance to the dam into its between-village and within-village components. Further, we include a log frailty term acting at the village level. The model we consider is thus

$$h_{ij}(t) = \lambda_m \exp(\bar{x}_i \beta_b + d_{ij} \beta_w + w_i) \quad (6.1)$$

for $t \in [a_m, a_{m+1})$, where $h_{ij}(\cdot)$ is the malaria hazard rate for child j ($j = 1, \dots, n_i$) of village i ($i = 1, \dots, s$), $\bar{x}_i$ is the village averaged distance to the dam (with β_b the distance effect at the village level), and $d_{ij} = x_{ij} - \bar{x}_i$ is the deviation of the household distance from the village averaged distance (with β_w the distance effect within a village). The log frailty term of village i , hereafter called random effect, is denoted by w_i .

— **Note: piecewise constant $h_0(\cdot)$ and $H_0(\cdot)$** —

$$0 = a_1 \quad \left[\text{---} \right] \quad a_2 \quad a_3 \quad \dots \quad a_m \quad \left[\text{---} \right] \quad a_{m+1} \quad a_{M-1} \quad a_M \quad \left[\text{---} \right] \quad a_{M+1} = \infty$$

If $t \in [a_m, a_{m+1})$, then

$$h_0(t) = \lambda_m \quad \text{and} \quad H_0(t) = \left(\sum_{k=1}^{m-1} \lambda_k (a_{k+1} - a_k) \right) + \lambda_m (t - a_m)$$

where $\sum_{k=1}^{m-1} \lambda_k (a_{k+1} - a_k)$ is set to 0 if $m = 1$.

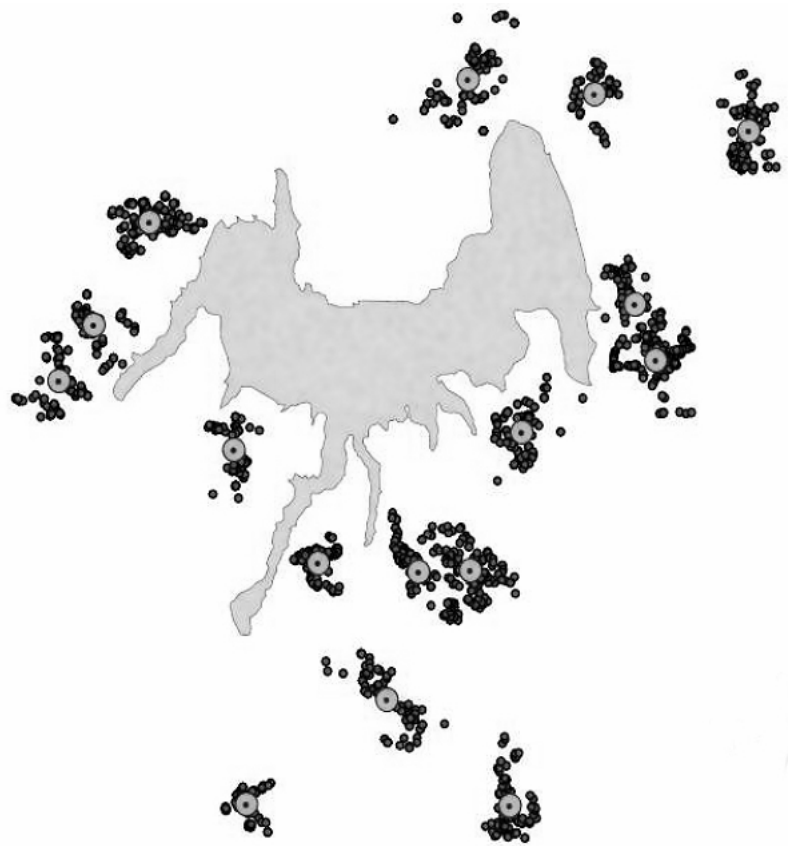


Figure 6.1: Location map of the study area showing the households (small circles) surrounding the Gilgel-Gibe hydroelectric dam reservoir. The big circles indicate the position of the village centres.

The random effect in (6.1) models the dependence among children within a village. In order to further model spatial dependence between villages, i.e. similarity in susceptibility to malaria in adjacent villages compared to distant villages, we model spatial dependence among random effects (Diggle & Ribeiro Junior, 2007, Section 4.3.3). We use the multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $\mathbf{D}(\boldsymbol{\theta})$. Typical in the spatial setting is that covariance is a function of distance. Accordingly, we assume that $\mathbf{D}(\boldsymbol{\theta})$ is an $s \times s$ matrix of form

$$[\mathbf{D}(\boldsymbol{\theta})]_{i_1 i_2} = \text{Cov}(w_{i_1}, w_{i_2}) = \theta_1 \rho(d_{i_1 i_2}; \boldsymbol{\theta}_2)$$

for some correlation function $\rho(\cdot; \boldsymbol{\theta}_2)$ of the distance between villages[†]. To put it differently, the vector of random effect parameters can be viewed as a realisation of a stationary Gaussian process $\{W(\mathbf{x}): \mathbf{x} \in R \subset \mathbb{R}^2\}$ with mean 0, variance θ_1 , and correlation function $\rho(d; \boldsymbol{\theta}_2) = \text{Cor}(W(\mathbf{x}), W(\mathbf{x}'))$, where $d = \|\mathbf{x} - \mathbf{x}'\|$ denotes the Euclidean distance between the village centre locations $\mathbf{x}$ and $\mathbf{x}'$. This type of process is said to be isotropic in the sense that $\rho(\cdot; \boldsymbol{\theta}_2)$ is symmetric with respect to direction (the correlation at any two locations depends solely on the separation between them).

For the malaria data, we focus on the Gaussian covariance structure

$$\rho(d; \boldsymbol{\theta}_2) = \exp \left\{ - (d/\theta_2)^2 \right\}, \quad \theta_2 > 0 \quad (6.2)$$

but other choices (see, e.g., Littell et al., 2006, Section 11.3.1) can be made (further work).

To fit the spatial frailty model (6.1), we can maximise the marginal likelihood. The marginal likelihood requires integration over the random effects. Since closed form integration is not possible in the case of normal random effects, an approximation is needed. Following Ripatti & Palmgren (2000), an analytical approximation of the marginal likelihood can be obtained by means of Laplacian integration (Goutis & Casella, 1999). The Laplace approximation of the marginal log-likelihood is given by

$$\begin{aligned} \ell_{\text{marg}}(\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{z}) \\ \approx -\frac{1}{2} \log \left(\det \left(\mathbf{D}(\boldsymbol{\theta}) \right) \right) - \frac{1}{2} \log \left(\det \left(K^{(2)}(\tilde{\mathbf{w}}) \right) \right) - K(\tilde{\mathbf{w}}) \end{aligned} \quad (6.3)$$

[†]In this chapter, the distance between two villages corresponds to the distance between the village centres.

with

$$\begin{aligned}
 K(\mathbf{w}) &:= K(\boldsymbol{\lambda}, \boldsymbol{\beta}, \mathbf{w}, \boldsymbol{\theta}; \mathbf{z}) \\
 &= - \left[\sum_{i=1}^s \sum_{j=1}^{n_i} \delta_{ij} \left(\log(h_0(y_{ij})) + \mathbf{x}'_{ij} \boldsymbol{\beta} + w_i \right) \right. \\
 &\quad \left. - H_0(y_{ij}) \exp(\mathbf{x}'_{ij} \boldsymbol{\beta} + w_i) \right] + \frac{1}{2} \mathbf{w}' \mathbf{D}^{-1}(\boldsymbol{\theta}) \mathbf{w} \quad (6.4)
 \end{aligned}$$

where $\tilde{\mathbf{w}} := \tilde{\mathbf{w}}(\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\theta})$ is such that $K^{(1)}(\tilde{\mathbf{w}}) := \frac{dK}{d\mathbf{w}}(\tilde{\mathbf{w}}) = \mathbf{0}$ and where $K^{(2)}(\cdot) := \frac{d^2 K}{d\mathbf{w} d\mathbf{w}'}(\cdot)$. Details are provided in Appendix 6.A. Maximisation of the (fully parametric) marginal log-likelihood (6.3) can be done numerically by means of an optimisation routine (e.g., a Newton-type algorithm).

6.3 Spatial dependence

In the spatial setting, the standard graphical tool to describe spatial dependence/variability is the semi-variogram. The semi-variogram is defined as one half the variance of the difference between two realisations of the underlying geostatistical process. For a stationary isotropic process Z , the semi-variogram is

$$\gamma_Z(d) = \frac{1}{2} \text{Var}(Z(\mathbf{x}) - Z(\mathbf{x}')) = \text{Var}(Z) - \text{Cov}(Z(\mathbf{x}), Z(\mathbf{x}'))$$

with $d = \|\mathbf{x} - \mathbf{x}'\|$.

A typical semi-variogram is shown in Figure 6.2. The value of the plateau in the semi-variogram corresponds to the variance of a realisation and is called the sill in the jargon of geostatistics. The distance at which the semi-variogram reaches the sill is called the range (if the sill is only reached asymptotically, then the practical range, defined as the distance at which the semi-variogram reaches 95% of the sill, is used in practice). For distances lower than the range, realisations exhibit less variability than expected due to spatial correlation. As the distance between two realisations increases, also the semi-variogram increases, indicating less spatial correlation. For distances greater than the range, realisations are no longer spatially correlated.

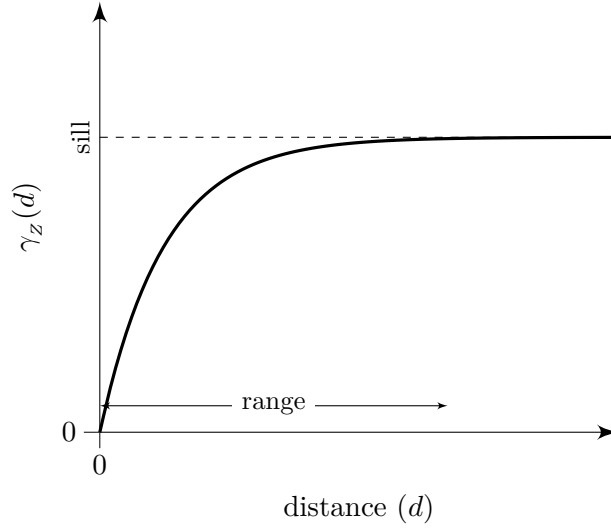


Figure 6.2: A typical semi-variogram with its main features (sill and range).

Given $\mathbf{x}$ and $\mathbf{x}'$ two village locations equally distant from the dam, the semi-variograms of the random effect (W) and frailty ($\exp(W)$) processes are

$$\gamma_W(d) = \theta_1 (1 - \rho(d; \theta_2))$$

and

$$\gamma_{\exp(W)}(d) = \exp(\theta_1) \left[\exp(\theta_1) - \exp(\theta_1 \rho(d; \theta_2)) \right]$$

with $d = \|\mathbf{x} - \mathbf{x}'\|$. See Figure (6.3) for plots of $\gamma_W(d)$ and $\gamma_{\exp(W)}(d)$ versus d in the case of the Gaussian correlation structure (6.2) for given values of θ_1 and θ_2 .

Observations from Figure (6.3) are as follows:

- The degree of unobserved heterogeneity between villages (sill) is controlled by the parameter θ_1 .
- The rate of spatial decay (range), i.e. the speed at which the semi-variogram reaches the sill, is governed by the parameter θ_2 . Higher values of θ_2 indicate larger ranges of spatial dependence.

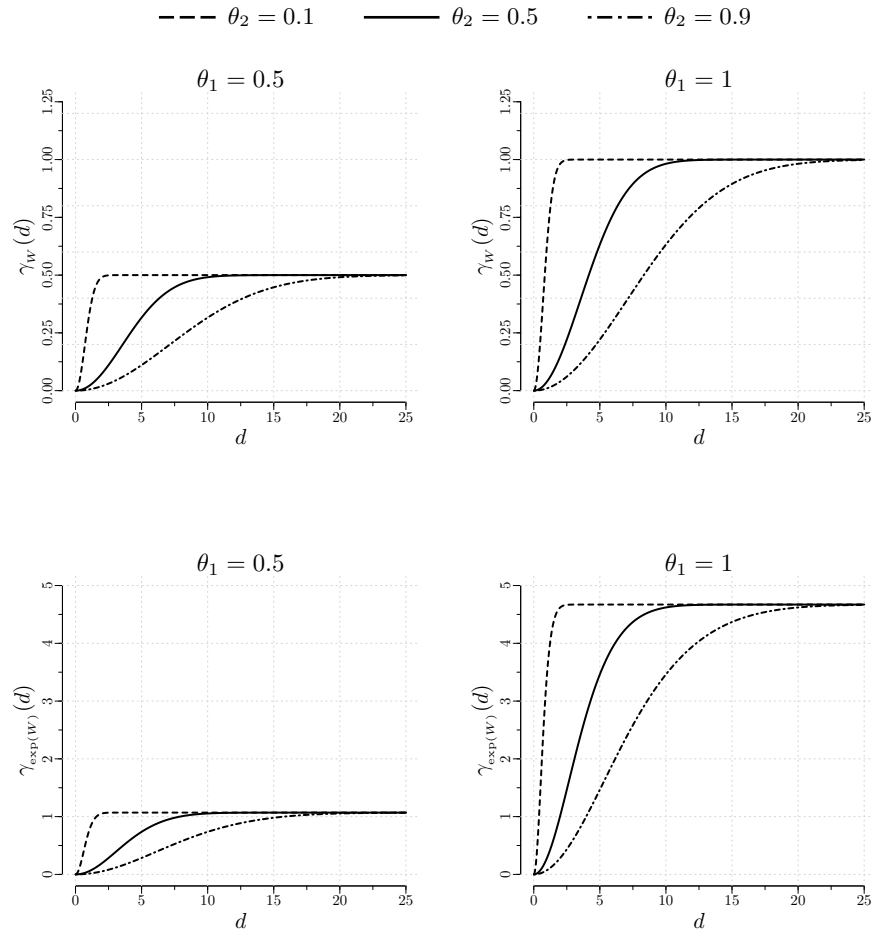


Figure 6.3: Semi-variograms of the random effect (W) and frailty ($\exp(W)$) processes in the case of the Gaussian correlation structure.

In order to provide some insight into how spatial correlation between random effects induces spatial dependence between survival times and log survival times, we assume a constant baseline hazard ($M = 1$). Given $\mathbf{x}$ and $\mathbf{x}'$ two village locations equally distant from the dam, the semi-variograms of the log survival time ($\log(T)$) and survival time (T) processes are (for details, see Appendix 6.B)

$$\begin{aligned}\gamma_{\log(T)}(d) &= \mathbb{E} \left[\text{Var}(\log(T) | W) \right] + \gamma_W(d) \\ &= \frac{\pi^2}{6} + \theta_1 (1 - \rho(d; \theta_2))\end{aligned}$$

and

$$\begin{aligned}\gamma_T(d) &= \mathbb{E} \left[\text{Var}(T | W) \right] + \left[\frac{1}{\lambda \exp(\mathbf{x}'\boldsymbol{\beta})} \right]^2 \gamma_{\exp(W)}(d) \\ &= \left[\frac{1}{\lambda \exp(\mathbf{x}'\boldsymbol{\beta})} \right]^2 \exp(\theta_1) \left\{ 2 \exp(\theta_1) - \exp(\theta_1 \rho(d; \theta_2)) \right\}\end{aligned}$$

with $d = \|\mathbf{x} - \mathbf{x}'\|$. In words:

- The semi-variogram of $\log(T)$ is the same as that of W , except for an intercept, the so-called nugget effect, here attributed to variability in outcomes measured at the same location;
- The semi-variogram of T is proportional to that of $\exp(W)$, plus the corresponding nugget effect.

Explicit from the previous formulas is that θ_2 in (6.2) retains its interpretation as a range parameter at the level of the survival times and log survival times.

6.4 Malaria data analysis

A total of 2039 children from 16 villages (between 123 and 130 children per village) participated to the malaria study. Children entered the study on the same day and were monitored at weekly intervals during a period of two years (July 2008 – June 2010). For each child diagnosed with malaria during the study period, the time to the first

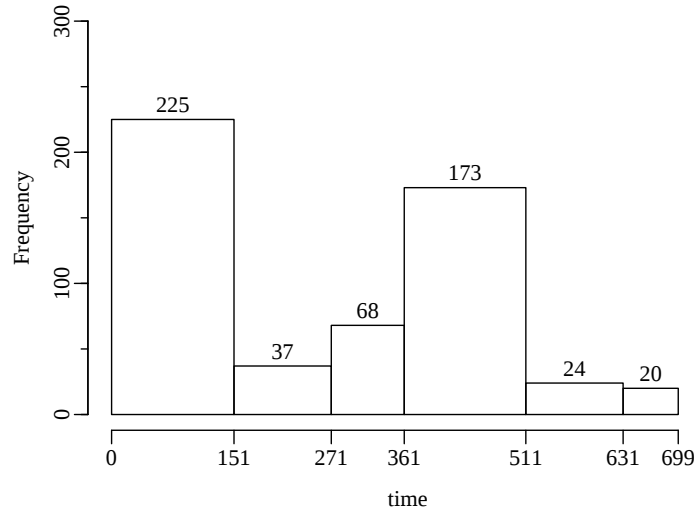


Figure 6.4: Number of cases by time period in the malaria data.

malaria episode (in days) was recorded. The household and village distances to the dam, as well as the exact household and village locations (latitude and longitude), are also available. The distance between two villages, calculated as the distance between the village centres, ranges from 1.475 km to 23.194 km (1st quartile = 7.306, median = 11.989, mean = 11.162, 3rd quartile = 14.379). Each of the two study years is divided into 3 rainy seasons, resulting in 6 intervals for the piecewise constant baseline hazard rate (Getachew, 2013, page 28): $[1, 151)$, $[151, 271)$, $[271, 361)$, $[361, 511)$, $[511, 631)$, $[631, 699)$. A total of 547 malaria cases ($\approx 27\%$) were reported (cf. Figure 6.4). The 1492 children ($\approx 73\%$) without malaria during the study period were censored at the last follow-up visit, which is the same for every child (698 days).

Results of fitting the shared frailty model and the spatial frailty model to the malaria data are shown in Table 6.1. The shared frailty model is obtained for $\theta_2 \rightarrow 0$ (in practice, θ_2 has been fixed to 10^{-9}). The likelihood ratio statistic ($LR = 6.33$) indicates that the spatial frailty model fits significantly better than the shared frailty model, but estimates and standard errors of the baseline hazard and fixed effects

Table 6.1: Shared frailty model and spatial frailty model fits to the malaria data.

parameter	shared		spatial	
	estimate	std. err.	estimate	std. err.
$\lambda_1 \times 10^4$	5.809	1.068	5.903	1.237
$\lambda_2 \times 10^4$	1.321	0.311	1.341	0.350
$\lambda_3 \times 10^4$	3.347	0.699	3.395	0.789
$\lambda_4 \times 10^4$	5.468	1.020	5.544	1.177
$\lambda_5 \times 10^4$	1.056	0.280	1.070	0.307
$\lambda_6 \times 10^4$	1.602	0.450	1.622	0.486
β_b	0.059	0.067	0.054	0.077
β_w	-0.122	0.075	-0.118	0.075
θ_1	0.314	0.128	0.407	0.304
θ_2			8.479	2.472
ℓ_{marg}	-4593.420		-4590.255	

parameters are similar in both models. In particular, no significant effect of the distance to the dam is found. Regarding spatial dependence, we find $\hat{\theta}_2 = 8.479$ (se = 2.472), corresponding to a practical range of approximately 15 km (cf. Figure 6.5).

6.5

Future development

This chapter provides a first step towards modelling spatial dependence in the malaria data by means of spatially correlated frailties. We have focused on modelling spatial dependence between villages, i.e. similarity in susceptibility to malaria in adjacent villages compared to distant villages. In the proposed model, any two children within a village exhibit the same degree of association (shared frailty). In principle, the method of this chapter can be extended to further model spatial dependence within a village. In that case, the frailty term acts at the household/child level rather than at the village level. Our preliminary attempts in that direction have revealed some technical difficulties. The main difficulty comes with solving the estimating equations for the vari-

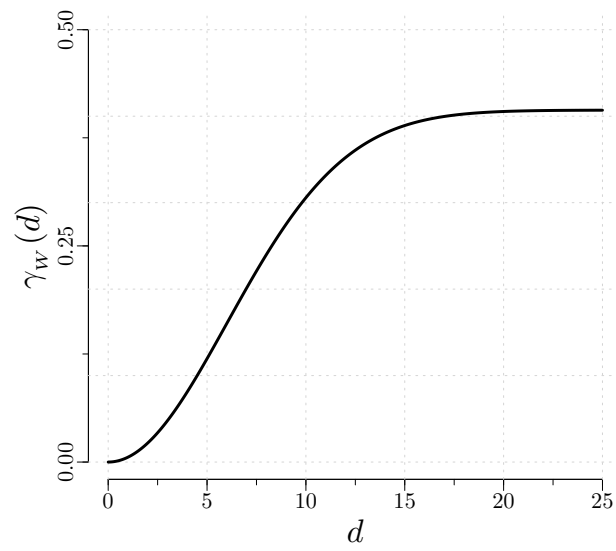


Figure 6.5: Estimate of the semi-variogram of the random effect process in the malaria data.

ance components,

$$\begin{aligned} \frac{\partial \ell_{\text{marg}}}{\partial \theta_q} = & -\frac{1}{2} \left[\text{tr} \left\{ \left[\mathbf{I}_{N \times N} - \left[K^{(2)} \right]^{-1} \mathbf{D}^{-1}(\boldsymbol{\theta}) \right] \frac{\partial \mathbf{D}(\boldsymbol{\theta})}{\partial \theta_q} \mathbf{D}^{-1}(\boldsymbol{\theta}) \right\} \right. \\ & \left. - \tilde{\mathbf{w}}' \mathbf{D}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{D}(\boldsymbol{\theta})}{\partial \theta_q} \mathbf{D}^{-1}(\boldsymbol{\theta}) \tilde{\mathbf{w}} \right] = 0 \end{aligned}$$

where $\mathbf{I}_{N \times N}$ denotes the identity matrix of dimension N , with N the total sample size. Contributing to the difficulty is the need to invert the $N \times N$ variance-covariance matrix of the random effects (in the malaria data, this matrix has dimension 2039×2039). In addition, the Laplace integration technique used to approximate the marginal likelihood has been reported to perform poorly with small cluster sizes (Ducrocq & Casella, 1996; see also Cortiñas Abrahantes & Burzykowski, 2005, and Rondeau et al., 2008). Further work is thus needed to model spatial dependence within a village. More discussion is provided at the end the thesis (closing part).

Appendix 6.A

Laplace approximation of the marginal likelihood

In the frequentist framework, an estimate of $(\lambda' \beta' \theta')'$, with λ the vector of parameters in $h_0(\cdot)$, β the vector of fixed effects parameters, and θ the vector of frailty parameters, is obtained by maximising the marginal likelihood of the observed data

$$\mathbf{z} = \{(y_{ij}, \delta_{ij}, \mathbf{x}_{ij}) \mid j = 1, \dots, n_i; i = 1, \dots, s\}$$

with y_{ij} the time to malaria or censoring, whichever comes first, δ_{ij} the event indicator identifying a case of malaria ($\delta_{ij} = 1$) or a right-censored data ($\delta_{ij} = 0$), and $\mathbf{x}_{ij}$ the vector of covariates for child j of village i . The random effect of village i , w_i , is not observed.

To obtain the marginal likelihood of the observed data, the latent data information $\mathbf{w} = (w_1 \dots w_s)'$ has to be integrated out from the joint density of $\mathbf{z}$ and $\mathbf{w}$,

$$\begin{aligned} L_{\text{marg}}(\lambda, \beta, \theta; \mathbf{z}) &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(\mathbf{z}, \mathbf{w}; \lambda, \beta, \theta) d\mathbf{w} \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(\mathbf{z} \mid \mathbf{w}; \lambda, \beta) f(\mathbf{w}; \theta) d\mathbf{w} \end{aligned}$$

with

$$\begin{aligned} f(\mathbf{z} \mid \mathbf{w}; \lambda, \beta) &= \prod_{i=1}^s \prod_{j=1}^{n_i} \left(h_0(y_{ij}) \exp(\mathbf{x}_{ij}' \beta + w_i) \right)^{\delta_{ij}} \\ &\quad \exp \left\{ -H_0(y_{ij}) \exp(\mathbf{x}_{ij}' \beta + w_i) \right\} \end{aligned}$$

and

$$f(\mathbf{w}; \theta) = (2\pi)^{-s/2} \det(\mathbf{D}(\theta))^{-1/2} \exp \left\{ -\frac{1}{2} \mathbf{w}' \mathbf{D}^{-1}(\theta) \mathbf{w} \right\}$$

There exists no closed form expression for L_{marg} when the random effects have a normal distribution. Following Ripatti & Palmgren (2000), an analytical approximation of the marginal likelihood can be obtained by means of Laplacian integration (Goutis & Casella, 1999).

Let $K(\mathbf{w}) := K(\lambda, \beta, \mathbf{w}, \theta; \mathbf{z})$ be such that

$$\begin{aligned} L_{\text{marg}}(\lambda, \beta, \theta; \mathbf{z}) &= (2\pi)^{-s/2} \det(\mathbf{D}(\theta))^{-1/2} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp \left\{ -K(\mathbf{w}) \right\} d\mathbf{w} \end{aligned}$$

The explicit form of $K(\mathbf{w})$ is given in (6.4). Using a Taylor series expansion of $K(\cdot)$ around $\tilde{\mathbf{w}}$, with $\tilde{\mathbf{w}}$ such that $\frac{dK}{d\mathbf{w}}(\tilde{\mathbf{w}}) = \mathbf{0}$, we have

$$K(\mathbf{w}) \approx K(\tilde{\mathbf{w}}) + \frac{1}{2}(\mathbf{w} - \tilde{\mathbf{w}})' K^{(2)}(\tilde{\mathbf{w}})(\mathbf{w} - \tilde{\mathbf{w}})$$

with $K^{(2)}(\cdot) := \frac{d^2 K}{d\mathbf{w} d\mathbf{w}'}(\cdot)$. An approximation of the marginal likelihood is thus given by

$$\begin{aligned} L_{\text{marg}}(\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{z}) &\approx (2\pi)^{-s/2} \det(\mathbf{D}(\boldsymbol{\theta}))^{-1/2} \exp\{-K(\tilde{\mathbf{w}})\} \\ &\quad \times \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}(\mathbf{w} - \tilde{\mathbf{w}})' K^{(2)}(\tilde{\mathbf{w}})(\mathbf{w} - \tilde{\mathbf{w}})\right\} d\mathbf{w} \\ &= \det(\mathbf{D}(\boldsymbol{\theta}))^{-1/2} \det(K^{(2)}(\tilde{\mathbf{w}}))^{-1/2} \exp\{-K(\tilde{\mathbf{w}})\} \end{aligned}$$

where the last equality follows by completing the integrand to recover the normal density with mean $\tilde{\mathbf{w}}$ and covariance matrix $[K^{(2)}(\tilde{\mathbf{w}})]^{-1}$.

Appendix 6.B

Semi-variograms of T and $\log(T)$ for the case $M = 1$

In this appendix, we work out the semi-variograms of T and $\log(T)$, i.e.

$$\gamma_T(d) = \frac{1}{2} \text{Var}(T(\mathbf{x}) - T(\mathbf{x}'))$$

and

$$\gamma_{\log(T)}(d) = \frac{1}{2} \text{Var}\left(\log(T(\mathbf{x})) - \log(T(\mathbf{x}'))\right)$$

in the case $M = 1$ (constant baseline hazard) by making use of the log-linear representation of model (6.1), where $d = \|\mathbf{x} - \mathbf{x}'\|$ denotes the Euclidean distance between the village centre locations $\mathbf{x}$ and $\mathbf{x}'$ equally distant from the dam.

Log-linear representation of model (6.1) in the case $M = 1$

The log-linear representation of model (6.1) when $M = 1$ is of form

$$\left(\log(T_{ij}) \mid W = w_i\right) = -\log(\lambda) - \mathbf{x}'_{ij}\boldsymbol{\beta} - w_i + E_{ij} \quad (6.5)$$

where E_{ij} is a random error term from a Gumbel distribution with density

$$f(e) = \exp \left\{ e - \exp(e) \right\}$$

The mean and the variance of E_{ij} are $E(E_{ij}) = -\gamma$ and $\text{Var}(E_{ij}) = \frac{\pi^2}{6}$, with $\gamma \approx 0.577$ the Euler's constant. Then, conditional on $W = w_i$, the survival time T_{ij} is given by

$$(T_{ij} | W = w_i) = \frac{1}{\lambda \exp(\mathbf{x}'_{ij}\boldsymbol{\beta})} \exp(-w_i) \exp(E_{ij}) \quad (6.6)$$

To see that the log-linear model (6.5) is equivalent to the hazard model (6.1) (when $M = 1$), we show that both representations lead to the same conditional survival function. Starting from (6.5), we have

$$\begin{aligned} \Pr(T_{ij} > t | W = w_i) &= \Pr(\log(T_{ij}) > \log(t) | W = w_i) \\ &= \Pr(E_{ij} > \log(t) + \log(\lambda) + \mathbf{x}'_{ij}\boldsymbol{\beta} + w_i) \\ &= \Pr(\exp(E_{ij}) > \lambda t \exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + w_i)) \\ &= \exp \left\{ -\lambda t \exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + w_i) \right\} \end{aligned}$$

where we have used the fact that $\exp(E_{ij}) \sim \text{Exp}(1)$.

Semi-variogram of $\log(T)$

Let $Z = \log(T)$. On the one hand, for two village locations $\mathbf{x}$ and $\mathbf{x}'$ equally distant from the dam, we have, using (6.5),

$$E(Z(\mathbf{x}) - Z(\mathbf{x}') | W) = W(\mathbf{x}') - W(\mathbf{x})$$

On the other hand,

$$\text{Var}(Z(\mathbf{x}) - Z(\mathbf{x}') | W) = 2 \text{Var}(Z | W)$$

Therefore, using the law of total variance, the semi-variogram of Z is

$$\begin{aligned} \gamma_Z(d) &= \frac{1}{2} \text{Var}(Z(\mathbf{x}) - Z(\mathbf{x}')) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2} \mathbb{E} \left[\text{Var} \left(Z(\mathbf{x}) - Z(\mathbf{x}') \mid W \right) \right] + \frac{1}{2} \text{Var} \left[\mathbb{E} \left(Z(\mathbf{x}) - Z(\mathbf{x}') \mid W \right) \right] \\
&= \mathbb{E} \left[\text{Var} \left(Z \mid W \right) \right] + \gamma_W(d)
\end{aligned}$$

with $d = \|\mathbf{x} - \mathbf{x}'\|$.

Note: law of total variance

$$\text{Var}(X) = \mathbb{E} \left[\text{Var}(X \mid Y) \right] + \text{Var} \left[\mathbb{E}(X \mid Y) \right]$$

Semi-variogram of T

Let $Z = T$. On the one hand, for two village locations $\mathbf{x}$ and $\mathbf{x}'$ equally distant from the dam, we have, using (6.6),

$$\mathbb{E} \left(Z(\mathbf{x}) - Z(\mathbf{x}') \mid W \right) = \frac{1}{\lambda \exp(\mathbf{x}'\boldsymbol{\beta})} \left[\exp(-W(\mathbf{x})) - \exp(-W(\mathbf{x}')) \right]$$

On the other hand,

$$\text{Var} \left(Z(\mathbf{x}) - Z(\mathbf{x}') \mid W \right) = 2 \text{Var} \left(Z \mid W \right)$$

Therefore, using the law of total variance, the semi-variogram of Z is

$$\begin{aligned}
&\gamma_Z(d) \\
&= \frac{1}{2} \text{Var} \left(Z(\mathbf{x}) - Z(\mathbf{x}') \right) \\
&= \frac{1}{2} \mathbb{E} \left[\text{Var} \left(Z(\mathbf{x}) - Z(\mathbf{x}') \mid W \right) \right] + \frac{1}{2} \text{Var} \left[\mathbb{E} \left(Z(\mathbf{x}) - Z(\mathbf{x}') \mid W \right) \right] \\
&= \mathbb{E} \left[\text{Var} \left(Z \mid W \right) \right] + \left[\frac{1}{\lambda \exp(\mathbf{x}'\boldsymbol{\beta})} \right]^2 \gamma_{\exp(W)}(d)
\end{aligned}$$

with $d = \|\mathbf{x} - \mathbf{x}'\|$.

Closing

Concluding discussion

The main focus of this thesis was on modelling the frailty term in the frailty model. Specific achievements include a unified framework for fitting the parametric frailty model with different frailty distributions and the accompanying R library `parfm` (Chapter 2), a new diagnostic plot to evaluate the frailty distribution assumption (Chapter 3), a simulation study to assess robustness of regression inference against frailty misspecification (Chapter 4), and a method to test for decreasing cluster heterogeneity in a new time-varying frailty model (Chapter 5). Also presented is a first step towards modelling spatial dependence in survival data by means of spatially correlated frailties (Chapter 6). The discussion below collects concluding remarks, summarises the findings, and suggests directions for further research.

Implementation of statistical methods, especially model estimation algorithms, in standard software is an essential step towards practical applications. It was not until the advent of computers that the paper by Kaplan & Meier (1958), one of the most cited papers in the entire field of science (Ryan & Woodall, 2005), began to be widely used among applied researchers (Garfield, 1989). With present technology, not providing the necessary tools to put theory into practice is like building a bridge to nowhere. After all, it is not surprise that number 2 on Donoho's list of causal factors for highly cited papers (Donoho, 2002) is "*Implement the method in software, place examples of the software's use in the paper, make the software of broad functionality, and give the software away for free.*" The R library `parfm` is a step in that direction. The `parfm` package presently supports four frailty distributions in the parametric frailty model. Although significant improvements are possible, feedback from users made it clear that `parfm`, or at least the `parfm` methodology, has filled a void in the realm of frailty models. Nonetheless, further developments are needed to make `parfm` of broad functionality and to open the door to new practical applications. In particular, the inclusion of the PVF frailty distribution should be quite straightforward (as demonstrated in Chapter 2) and might lead to simple goodness-of-fit likelihood ratio tests for the frailty distribution (cf. Section 1.5 of the thesis appendix). To ensure longevity, `parfm` could be merged with an existing, well-maintained package (e.g. `frailtypack`).

The `parfm` package has been useful as well in the development of a

diagnostic technique to evaluate the frailty distribution assumption. As seen in Chapter 4, frailty misspecification does not appear to affect inference for fixed effects. In contrast, the type of dependence that the frailty term generates between event times in a cluster is dictated by the choice of frailty distribution. This is evidenced by the fact that local measures of dependence (e.g. the cross ratio) take different forms depending on the frailty distribution assumption (cf. Appendix 3.A). Therefore, diagnostic techniques are particularly relevant to studies where the dependence structure is of special interest. The study on mastitis infection in dairy cows used to illustrate the methodology of Chapter 3 is one such example. The dependence structure between infection times of the four udder quarters of a cow provides insight into the risks of cross-contamination between infected and non-infected udder quarters. For this type of studies, diagnostic techniques to select the most appropriate frailty distribution are needed in order to correctly describe the dependence in the data. Much of the work towards this end focused on comparing an empirical to a model-based estimate of the cross ratio (Oakes, 1989; Viswanathan & Manatunga, 2001; Chen & Bandeen-Roche, 2005; Glidden, 2007). As an alternative, quantile dependence has been used in Chapter 3. The main motivation for using quantile dependence as a diagnostic tool was to capture the behaviour of the frailty distribution in the tails. To be specific, the heaviness of the tails of the frailty distribution has a major influence on whether dependence is mostly present between short, intermediate, or long event times, as is the case with positive stable, inverse Gaussian, and gamma frailties, respectively. Incidentally, quantile dependence is defined in terms of survival rather than in terms of hazard, making the non-parametric estimation somewhat easier. For the mastitis data, the non-parametric estimate suggests that early infections happen in the same cows (early dependence), a pattern consistent with the positive stable frailty distribution (cf. Figure 3.5). In essence, the proposed diagnostic tool parallels the QQ plot. The simulations conducted in Chapter 3 indicate satisfactory performance of the proposed graphic tool. Nonetheless, more experience is needed to demonstrate the value of the proposed method in real applications.

Alternatives to diagnostic plots and goodness-of-fit tests for model selection are information criteria, e.g. the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Modifications of AIC and BIC for use in survival analysis have been proposed (Liang et al., 2008; Volinsky & Raftery, 2000). For copula models (the relation-

ship between copula and frailty models is discussed in Goethals et al. (2008, 2012)), the use of AIC to select a copula in a list of candidates is investigated and promoted in Fang et al. (2014). For frailty models, and more generally for random effects models, the definitions of AIC and BIC are not unambiguous. What likelihood and penalty term should be used? This question has been addressed for the selection of random effects, first in the linear mixed model (Vaida & Blanchard, 2005; Liang & Zou, 2008; Greven & Kneib, 2010) and then in generalised linear mixed and frailty models (Donohue et al., 2011). For this purpose, it is argued that a modified version of the AIC, the conditional Akaike Information Criterion (cAIC), should be used. To the best of our knowledge, the use of cAIC, or of any other information criterion, for the selection of the random effect/frailty distribution has not been addressed in the literature. This may be an interesting subject for further research.

A second objective of this work was to ascertain the sensitivity of regression inference to the choice of frailty distribution. In a particular application, conclusions regarding the effect of frailty misspecification can be drawn by fitting the frailty model with different frailty distributions. For example, different frailty models for breast cancer recurrence data were applied in dos Santos et al. (1995). The results indicate that the effect of different frailty distributions on the parameter estimates and their standard errors was slight. Further insight can be gained from a simulation study. In a simulation setting, survival data are generated from the frailty model, with a known frailty distribution and known model parameters, and robustness properties are derived based on the parameters in the misspecified frailty model. A simulation study can be designed to mimic specific study characteristics (number of clusters, number of observations per cluster, amount of heterogeneity, effect size, event rate, etc.) and hence allows to evaluate the effect of frailty misspecification in a variety of settings. A general approach to generating survival data to simulate the frailty model is given in the thesis appendix (Appendix 2). In multicentre clinical trial data, there is increasing empirical evidence that inferences about the treatment effect are robust against frailty misspecification. The use of the frailty model could then be further promoted in that context (cf. Chapter 4, supplementing results from Glidden & Vittinghoff (2004)). It should be noted that the case of cluster-randomised trials (Murray et al., 2004), i.e. trials in which the centres rather than the individual patients are randomly allocated to intervention and control groups, has not been covered in Chapter 4

and might deserve a specific simulation study.

Up until that point, the frailty term was assumed to remain unchanged over time. Also of interest in some applications are frailty models with time-varying frailties. Time-varying frailties enable, e.g., to model changes in lifestyle (Gottard et al., 2012) or serial dependence between recurrent event times (cf. the references given in Section 5.1). In Chapter 5, time-varying frailties were used to investigate decreasing heterogeneity between bone marrow transplant centres in a study of chronic myeloid leukemia. We have approached that problem by using a transformation towards the linear mixed effects model (cf. Section 5.2). To be specific, we have modelled the logarithm of the cumulative hazard as a linear mixed effects model with a time-varying, random centre effect. The foundation of the proposed transformation model, for which the theoretical properties are studied in the case of time-varying fixed effects (no frailty) in Teodorescu et al. (2010), involves the non-parametric estimation of the pseudo-responses and subsequent use of linear mixed model techniques (Grigoletto & Akritas, 1999). In the leukemia data, we have found that heterogeneity declines over time, with almost no heterogeneity after roughly nine months from transplantation (cf. Figure 5.1). The statistical significance of the decrease has been assessed by means of a generic trend test procedure described in Rice & Gaines (1994a,b) (cf. Section 5.2.5). An alternative test would be one of the tests for the homogeneity of a set of variances against ordered alternatives studied in Fujino (1979). The ability to use standard software is a strength of the proposed method. However, this practical benefit comes at a cost in terms of applicability due to the preliminary estimation of the pseudo-responses which requires substantial information (number of events) within each cluster. For the leukemia data, small transplant centres have been discarded from the analysis. Further, continuous covariates represent an additional difficulty for the estimation of the pseudo-responses (kernel-based estimation; cf. López-de-Ullibarri et al. (2012)). For the leukemia data analysis, the covariate information has been summarised into a binary prognostic index (“low-risk” versus “high-risk”). Overall, I think that although the proposed method provides a practical way to get an insight on cluster heterogeneity and its time dependency in certain situations, e.g. a treatment outcome study with few centres of large sizes, a method with broader applicability is desirable.

The malaria data has been the source of inspiration for Chapter 6.

Specific for the malaria data is that dependence between event times is a function of the distance between subjects (spatial dependence). Our particular interest has been to determine the extent of spatial dependence in the malaria study area. In the last decade, this type of problem has prompted the development of geostatistical methods for survival data (cf. the references given in Section 6.1). A natural way to incorporate spatial dependence between event times is to use the multivariate log-normal frailty distribution with a spatially structured variance-covariance matrix (Diggle & Ribeiro Junior, 2007, Section 4.3.3). Within that framework, we have modelled, as a first step, spatial dependence at the cluster (village) level. This is essentially the approach taken in Li & Ryan (2002). Application to the malaria data indicates significant spatial dependence in the study area, with a range of spatial dependence estimated at 15 km. It must be noted, however, that some of the villages are rather large. The assumption that the frailty term takes the same value within a village, rather than different, spatially dependent values, is therefore called into question. Further, the village boundaries are quite vague and somewhat arbitrary. Extension to spatial dependence at the household level, with one random effect per household/child, entails a number of technical difficulties and requires further work. First, the practical implementation is more involved due to the large number of random effects. Contributing to the difficulty is the need to invert the variance-covariance matrix of the random effects in the marginal likelihood (6.3) and in its gradient. Further, it has been reported that the Laplace approximation used to obtain the marginal log-likelihood (6.3) (cf. Appendix 6.A) performs poorly for small cluster sizes (Ducrocq & Casella, 1996; see also Cortiñas Abrahantes & Burzykowski, 2005, and Rondeau et al., 2008). This suggests that another method is needed to fit the model with spatial dependence at the household level (see, e.g., Section 6.1). In addition, the household-specific random effect models unobserved heterogeneity between households (overdispersion), i.e. greater variability than expected (Hougaard, 1995; Wienke, 2010, Chapter 3), and the fact that overdispersion is needed for the presence of spatial dependence is a drawback of the model. In that respect, the accelerated failure time model with spatially dependent error terms might deserve some attention (see Klein et al. (1999) for one avenue to be pursued).

Future development

A further field of application of frailty models concerns the analysis of recurrent event times. Recurrent event analysis covers situations where the event of interest may occur multiple times per subject. Examples from medical research are chronic asthma attacks, repeated infections after surgery, cancer relapses, etc. Typical for recurrent event data is that outcomes from the same subject are dependent (recurrences are clustered within subjects). Further, for some diseases, recurrences are associated with an increased risk of death. Joint frailty models have been proposed to take into account the within-subject dependence as well as the dependent censoring of recurrent event times by death (Liu et al., 2004). A joint frailty model consists of a frailty model for each of the two endpoints (recurrence and death). On the one hand, the within-subject dependence is modelled by an individual-specific frailty term U in the hazard function of the recurrent event times. On the other hand, the dependence between the recurrent event times and the survival time is usually modelled by including U^γ in the hazard function of the survival time (thus linking the two models), with $\gamma \in \mathbb{R}$ an unknown parameter. Owing to the parameter γ , the common frailty term is allowed to have different effects on the two hazard rates. To fit the joint frailty model, an EM algorithm with a Metropolis-Hastings E-step is developed in Liu et al. (2004) and in Huang & Liu (2007). Alternatively, maximum penalised likelihood estimation is used in Rondeau et al. (2007). The latter approach is implemented in the R library `frailtypack` (Rondeau et al., 2012).

Application of joint frailty modelling of recurrent event times and time to death in cancer research allows to assess whether the treatment has an effect on the recurrent/death rate and to quantify the association between recurrence and death (Rondeau, 2010; Mazroui et al., 2012; see also Rondeau et al., 2011 and Taylor & Wang, 2002). In addition, joint frailty models have the potential to determine the extent to which the treatment effect on survival is mediated through the recurrences. Thus, joint frailty modelling may well have a role to play in evaluating the use of recurrences as a surrogate endpoint in cancer clinical trials (De Gruttola et al., 2001).

A further application of joint frailty models is to dynamically predict the risk of death for an individual patient, given the patient's history of

recurrences (Mauguen et al., 2013). Using ideas in Faucett et al. (2002), these predicted probabilities of death could then be used to recover part of the information that is lost due to censoring via multiple imputation. By treating the recurrences as auxiliary responses, efficiency of the treatment effect estimate on the primary response (survival) is expected to be gained.

Overall, it is proposed to evaluate whether recurrences are appropriate for use both as a surrogate endpoint for survival and as auxiliary responses to improve the efficiency of the treatment effect on survival.

Appendix

1

Frailty distributions

1.1 --- Gamma

notation & parameter space

$$U \sim \text{Gam}(\theta), \quad \theta > 0$$

density

$$f(u) = \frac{(1/\theta)^{1/\theta} u^{(1/\theta)-1} \exp(-(1/\theta)u)}{\Gamma(1/\theta)}$$

R syntax

```
dgamma(x=u, shape=1/theta, scale=theta)
[library: stats]
```

Laplace transform

$$\mathcal{L}(x) = (1 + \theta x)^{-1/\theta}$$

mean & variance

$$\text{E}(U) = 1 \quad \text{and} \quad \text{Var}(U) = \theta$$

Kendall's τ

$$\frac{\theta}{\theta + 2}$$

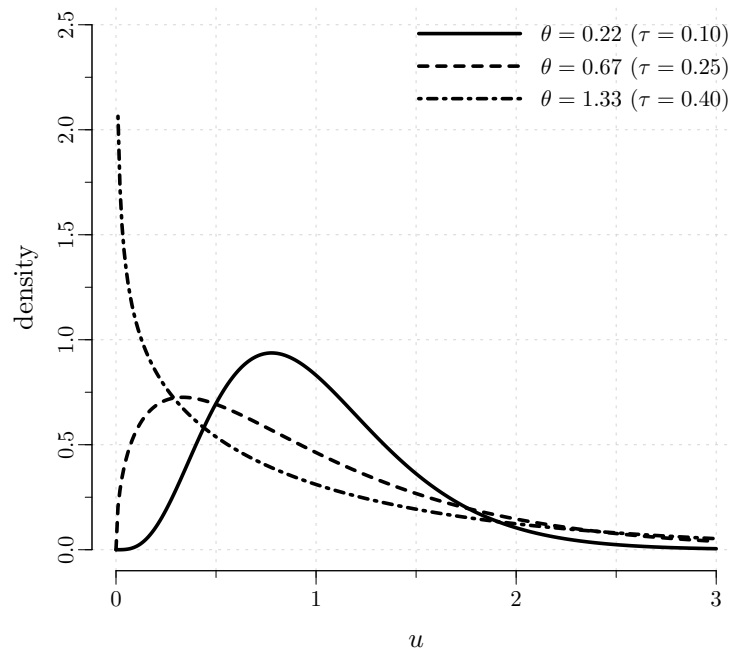


Figure 1.1: Gamma densities.

1.2

Inverse Gaussian

notation & parameter space

$$U \sim \text{IG}(\theta), \quad \theta > 0$$

density

$$f(u) = \frac{1}{\sqrt{2\pi\theta}} u^{-3/2} \exp\left(-\frac{1}{2\theta u} (u-1)^2\right)$$

R syntax

```
dinvgauss(x=u, mu=1, lambda=1/theta)
[library: statmod]
```

Laplace transform

$$\mathcal{L}(x) = \exp\left\{\frac{1}{\theta} \left(1 - \sqrt{1 + 2\theta x}\right)\right\}$$

mean & variance

$$\text{E}(U) = 1 \quad \text{and} \quad \text{Var}(U) = \theta$$

Kendall's τ

$$\tau = \frac{1}{2} - \frac{1}{\theta} + \frac{2}{\theta^2} \exp\left(\frac{2}{\theta}\right) \int_{2/\theta}^{\infty} \frac{\exp(-x)}{x} dx$$

Note: $\tau < 1/2$

Since

$$\int_{2/\theta}^{\infty} \frac{\exp(-x)}{x} dx < \frac{\theta}{2} \int_{2/\theta}^{\infty} \exp(-x) dx = \frac{\theta}{2} \exp\left(-\frac{2}{\theta}\right)$$

it follows that $\tau < 1/2$.

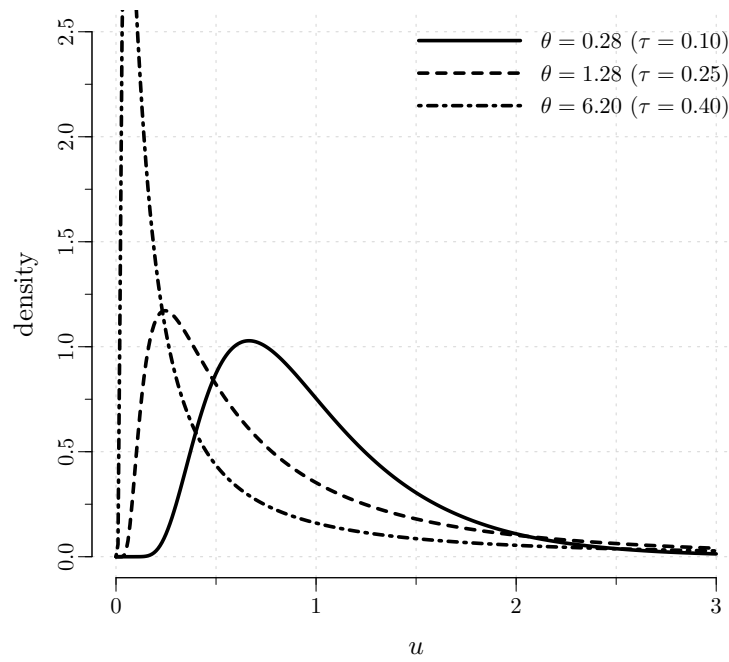


Figure 1.2: Inverse Gaussian densities.

1.3

Positive stable

notation & parameter space

$$U \sim \text{PS}(\theta), \quad \theta \in (0, 1)$$

density

$$f(u) = -\frac{1}{\pi u} \sum_{k=1}^{\infty} \left\{ \frac{\Gamma(k(1-\theta) + 1)}{k!} \sin((1-\theta)k\pi) (-u^{\theta-1})^k \right\}$$

R syntax[†]

```
dstable(x=u, alpha=1-theta, beta=1, gamma=g(theta), delta=0, pm=1)
[library: stabledist]
```

Laplace transform

$$\mathcal{L}(x) = \exp(-x^{1-\theta})$$

mean & variance

$$\text{E}(U) = \infty \quad \text{and} \quad \text{Var}(U) = \infty$$

Kendall's τ

$$\tau = \theta$$

— [†]**Note: the gamma argument of `dstable()`** —

$$\text{gamma} = \left| 1 - i \tan\left(\frac{\pi}{2}(1-\theta)\right) \right|^{-\frac{1}{1-\theta}} = \left[1 + \tan^2\left(\frac{\pi}{2}(1-\theta)\right) \right]^{-\frac{1}{2(1-\theta)}} \\ (\approx \theta \text{ in a neighbourhood of } 1/2)$$

```
g <- function(theta)
{
  i <- complex(real=0, imaginary=1)
  abs(1 - i * tan(pi / 2 * (1 - theta)))^(-1 / (1 - theta))
}
```

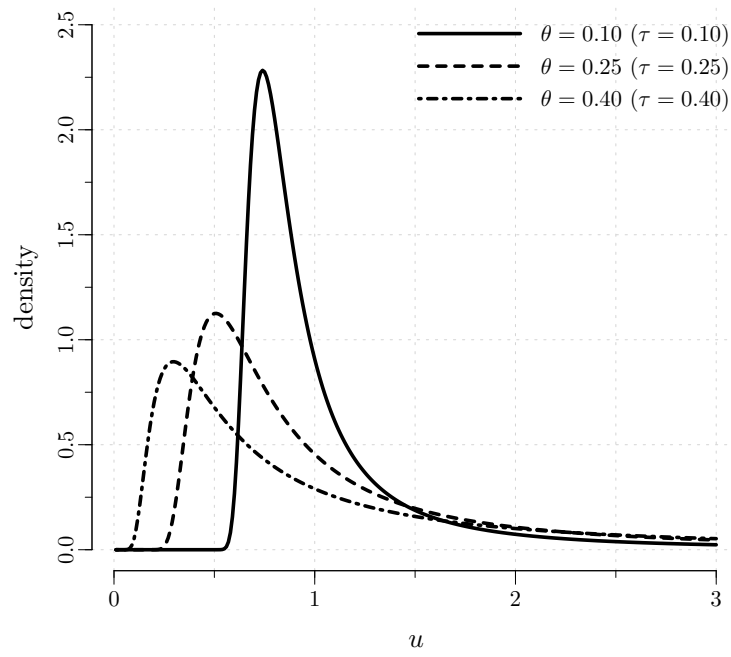


Figure 1.3: Positive stable densities.

1.4

Log-normal

notation & parameter space

$$U \sim \text{LN}(\theta), \quad \theta > 0$$

density

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

R syntax

```
dlnorm(x=u, meanlog=0, sdlog=sqrt(theta))
[library: stats]
```

Laplace transform

No explicit formula exists.

mean & variance

$$\text{E}(U) = \exp\left(\frac{\theta}{2}\right) \quad \text{and} \quad \text{Var}(U) = \exp(\theta)(\exp(\theta) - 1)$$

Kendall's τ

No explicit formula exists.

Note: relation with the normal distribution

Let $W \sim \text{N}(\mu, \theta)$. Then $U = \exp(W)$ has a two-parameter log-normal distribution with

$$\text{E}(U) = \exp\left(\mu + \frac{\theta}{2}\right) \quad \text{and} \quad \text{Var}(U) = \exp(2\mu + \theta)(\exp(\theta) - 1)$$

The standard assumption, adopted here, is to take $\text{E}(W) = \mu = 0$. Note in this case that $\text{E}(U) \neq 1$. Alternatively, one can take $\mu = -\theta/2$ so that $\text{E}(U) = 1$.

In terms of W , the frailty model is written as $h_{ij}(t) = h_0(t) \exp(w_i + \mathbf{x}'_{ij}\boldsymbol{\beta})$.

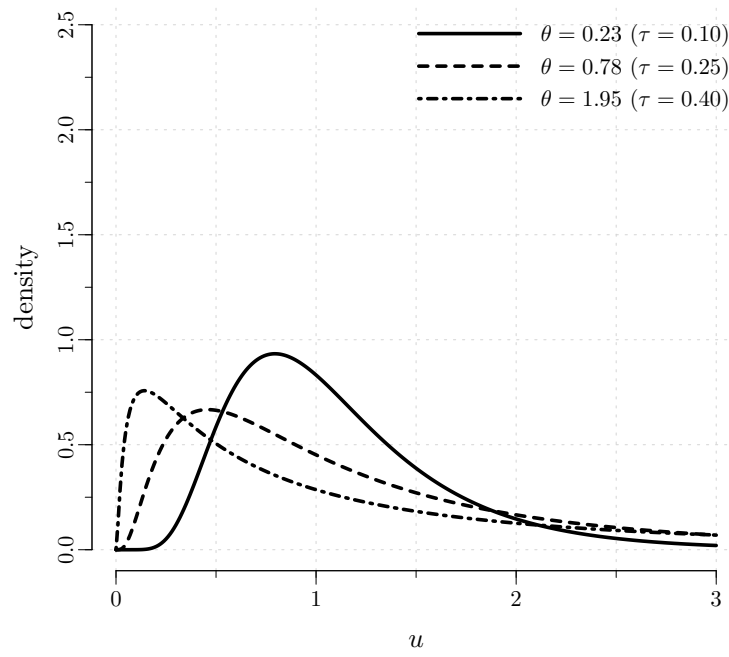


Figure 1.4: Log-normal densities.

1.5

Power variance function

The family of power variance function (PVF) distributions contains the inverse Gaussian as a particular case, as well as the gamma and the positive stable as limiting cases.

The Laplace transform of $U \sim \text{PVF}(\mu, \theta, \nu)$ is[†]

$$\mathcal{L}(x) = \exp \left\{ \frac{\nu}{\theta(1-\nu)} \left[1 - \left(1 + \frac{\theta\mu x}{\nu} \right)^{1-\nu} \right] \right\}$$

with $\mu > 0$, $\theta > 0$, and $\nu \in (0, 1)$.

The mean and the variance of U are

$$\mathbb{E}(U) = \mu \quad \text{and} \quad \text{Var}(U) = \theta\mu^2$$

The typical choice to ensure identifiability of the parameters in the frailty model is $\mu = 1$.

For $\mu = 1$ and $\nu \rightarrow 1$, the gamma frailty distribution $\text{Gam}(\theta)$ is obtained. For $\mu = 1$ and $\nu = 1/2$, the inverse Gaussian frailty distribution $\text{IG}(\theta)$ is obtained. To obtain the positive stable frailty distribution $\text{PS}(\nu)$, we need to take

$$\theta = \frac{\nu}{(1-\nu)^{1/\nu}} \mu^{(1/\nu)-1}$$

and to let μ go to infinity. For details, see Duchateau & Janssen (2008, Section 4.5.1).

[†]The relation with the parametrisation used in Hougaard (2000, Section A.3.4) is as follows:

$$\alpha_{\text{H}} = 1 - \nu \quad \theta_{\text{H}} = \frac{\nu}{\theta\mu} \quad \delta_{\text{H}} = \mu^{1-\nu} \left(\frac{\nu}{\theta} \right)^{\nu}$$

where the subscript “H” stands for “Hougaard”.

1.6 --- Compound Poisson

A compound Poisson (CP) random variable can be constructed as the sum of a Poisson distributed number k of independent random variables,

$$U = \begin{cases} X_1 + \cdots + X_k & \text{if } k > 0 \\ 0 & \text{if } k = 0 \end{cases}$$

As $\Pr(U = 0)$ is positive, the compound Poisson frailty distribution provides a way to model a cure fraction in the population (Price & Manatunga, 2001; Moger et al., 2004).

Interestingly, when the X_i 's have a two-parameter gamma distribution, the compound Poisson distribution has the same Laplace transform as the PVF, except that $\nu > 1$ (Duchateau & Janssen, 2008, Section 4.6.1).

2

Generation of survival times to simulate the frailty model

Together with graphs, Monte Carlo simulations are among the most helpful tools to develop or confirm intuition by repeating some experiment using (pseudo-)random samples[†]. The essence of Monte Carlo simulations is that we know and control the data generating mechanism.

2.1 The inverse probability method

To generate clustered event times from the frailty model, we can use the inverse probability method (Bender et al., 2005): if V is uniform on $(0, 1)$ and if $S(\cdot | \mathbf{x}, u)$ is the conditional survival function derived from the frailty model, i.e.

$$S(t | \mathbf{x}, u) = \exp \left(-H_0(t) u \exp(\mathbf{x}'\boldsymbol{\beta}) \right)$$

then it is a fact that the random variable

$$T = S^{-1}(V | \mathbf{x}, u) = H_0^{-1} \left(-\frac{\log(V)}{u \exp(\mathbf{x}'\boldsymbol{\beta})} \right)$$

has survival function $S(\cdot | \mathbf{x}, u)$. This result is known as “the inverse probability integral transformation”. Therefore, to generate a survival time $T \sim S(\cdot | \mathbf{x}, u)$ given the covariate vector and the frailty term, it suffices to draw v from $V \sim \text{U}(0, 1)$ and to make the inverse transformation $t = S^{-1}(v | \mathbf{x}, u)$.

[†]For a nice illustration of how repeating the “Monty Hall game show problem” has convinced the most reluctant readers of the “Ask Marilyn” column in Parade Magazine, see the discussion at marilynvossavant.com/game-show-problem.

2.2

Examples

Example 1 [Weibull baseline hazard]

Let $h_0(t) = \lambda \rho t^{\rho-1}$ with shape $\rho > 0$ and scale $\lambda > 0$. Then $H_0(t) = \lambda t^\rho$ and $H_0^{-1}(t) = (\frac{t}{\lambda})^{1/\rho}$. Following the inverse probability method, a realisation of $T \sim S(\cdot | \mathbf{x}, u)$ is obtained by computing

$$t = \left(-\frac{\log(v)}{\lambda u \exp(\mathbf{x}'\boldsymbol{\beta})} \right)^{1/\rho}$$

with v a uniform variate on $(0, 1)$. Using results on transformations of random variables, one may notice that T has a conditional Weibull distribution (given $\mathbf{x}$ and u) with shape ρ and scale $\lambda u \exp(\mathbf{x}'\boldsymbol{\beta})$.

Example 2 [piecewise constant baseline hazard]

Let $h_0(t) = \lambda_m$ if $t \in [a_m, a_{m+1})$, with $0 = a_0 < a_1 < \dots < a_M < a_{M+1} = \infty$. Then, for $t \in [a_m, a_{m+1})$,

$$H_0(t) = \left(\sum_{k=1}^{m-1} \lambda_k (a_{k+1} - a_k) \right) + \lambda_m (t - a_m)$$

and

$$H_0^{-1}(t) = a_m + \frac{1}{\lambda_m} \left[t - \left(\sum_{k=1}^{m-1} \lambda_k (a_{k+1} - a_k) \right) \right]$$

where we set $\sum_{k=1}^{m-1} \lambda_k (a_{k+1} - a_k)$ to zero if $m = 1$. Following the inverse probability method, a realisation of $T \sim S(\cdot | \mathbf{x}, u)$ is obtained by computing

$$t = a_m - \frac{1}{\lambda_m} \left[\frac{\log(v)}{u \exp(\mathbf{x}'\boldsymbol{\beta})} + \left(\sum_{k=1}^{m-1} \lambda_k (a_{k+1} - a_k) \right) \right]$$

with v a uniform variate on $(0, 1)$. To determine m in the above formula, note that if $a_m \leq t < a_{m+1}$, then

$$\begin{aligned} \exp \left\{ -u \exp(\mathbf{x}'\boldsymbol{\beta}) \left(\sum_{k=1}^m \lambda_k (a_{k+1} - a_k) \right) \right\} &< v \\ &\leq \exp \left\{ -u \exp(\mathbf{x}'\boldsymbol{\beta}) \left(\sum_{k=1}^{m-1} \lambda_k (a_{k+1} - a_k) \right) \right\} \end{aligned}$$

2.3

Sample code in R

```
# baseline hazard: Weibull
# frailty distribution: gamma

# s = number of clusters
# n = (n_1 ... n_s) with n_i the nb of obs in cluster i
# lambda = scale parameter in h0()
# rho = shape parameter in h0()
# beta = fixed effect parameter
# theta = frailty parameter
# rateC = rate parameter of the exponential dist of C

simulWeibGam <- function(s, n, lambda, rho, beta, theta, rateC)
{
  # total number of observations
  N <- sum(n)

  # cluster identification number
  cluster <- factor(rep(1:s, times=n))

  # gamma frailties
  u <- rep(rgamma(n=s, shape=1/theta, scale=theta), times=n)

  # covariate --> N Bernoulli trials
  x <- sample(x=c(0, 1), size=N, replace=TRUE, prob=c(0.5, 0.5))

  # Weibull latent event times
  v <- runif(n=N)
  Tlat <- (- log(v) / (lambda * u * exp(x * beta)))^(1 / rho)

  # censoring times
  C <- rexp(n=N, rate=rateC)

  # follow-up times and event indicators
  time <- pmin(Tlat, C)
  status <- as.numeric(Tlat <= C)

  # data set
  data.frame(id=1:N,
             cluster=cluster,
             time=time,
             status=status,
             x=x)
}
```

Bootstrap in the frailty model

The broad aim of the bootstrap is to simulate the data generating mechanism in order to create replicate data sets. In its non-parametric version, the empirical distribution function is used to resample from the original data. Alternatively, the model-based bootstrap uses a fitted model. In the hypothesis testing framework, a model-based bootstrap can be used to determine the finite-sample null distribution of the test statistic by resampling the data under H_0 .

Bootstrap methods for non-clustered survival data are described in Davison & Hinkley (1997, Section 3.5 and Section 7.3). In the presence of clustering, a model-based resampling plan, based on the frailty model, is developed in Massonnet et al. (2006). Some details are given below. The non-parametric bootstrap for clustered survival data simply consists in randomly selecting clusters with replacement (Therneau & Grambsch, 2000, page 249; Ren et al., 2010).

3.1

Model-based bootstrap

To resample the event times, we need a model-based estimate of the conditional event time survival function. The conditional event time survival function derived from the frailty model is

$$\hat{S}_{ij}(t) = \exp \left\{ -\hat{H}_0(t) u_i \exp(\mathbf{x}'_{ij} \hat{\boldsymbol{\beta}}) \right\}$$

where $\hat{H}_0(t)$ and $\hat{\boldsymbol{\beta}}$ are the estimates obtained by fitting the frailty model to the original data. In the semi-parametric setting, we take the Breslow estimator for $\hat{H}_0(\cdot)$, i.e.

$$\hat{H}_0(t) = \sum_{\tilde{y}_{(\ell)} \leq t} \frac{d_\ell}{\sum_{i,j \in R(\tilde{y}_{(\ell)})} u_i \exp(\mathbf{x}'_{ij} \hat{\boldsymbol{\beta}})}$$

with $\tilde{y}_{(1)} < \dots < \tilde{y}_{(r)}$ the ordered distinct event times, d_ℓ the number of events at time $\tilde{y}_{(\ell)}$, and $R(\tilde{y}_{(\ell)})$ the risk set at $\tilde{y}_{(\ell)}$.

To resample the censoring times, we need an estimate of the censoring time survival function. An estimator of the censoring time survival function can be obtained via the Kaplan-Meier estimator (cf. Section 1.2.4) by interchanging the role of the event times and the censoring times.

Algorithm

For individual j of cluster i ($j = 1, \dots, n_i$; $i = 1, \dots, s$),

1. Sample u_i^* from the frailty distribution (where an estimate $\hat{\theta}$ of the frailty parameter is obtained by fitting the frailty model to the original data);
2. Generate t_{ij}^* from the model-based estimate of the conditional event time survival function (with $u_i = u_i^*$);
3. If $\delta_{ij} = 0$, then set $c_{ij}^* = y_{ij}$; otherwise, generate c_{ij}^* from the estimate of the censoring time survival function given that $C_{ij} > y_{ij}$, i.e. $\hat{G}(\cdot)/\hat{G}(y_{ij})$;
4. Set $y_{ij}^* = \min(t_{ij}^*, c_{ij}^*)$ and $\delta_{ij}^* = I(t_{ij}^* \leq c_{ij}^*)$.

References

- ABRAMOWITZ, M. & STEGUN, I. A., eds. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover Publications, 10th ed.
- ANDERSEN, P. K., KLEIN, J. P. & ZHANG, M.-J. (1999). Testing for centre effects in multi-centre survival studies: A Monte Carlo comparison of fixed and random effects tests. *Statistics in Medicine* **18**, 1489–1500.
- ANDERSON, J. E., LOUIS, T. A., HOLM, N. V. & HARVALD, B. (1992). Time-dependent association measures for bivariate survival distributions. *Journal of the American Statistical Association* **87**, 641–650.
- BANERJEE, S. & CARLIN, B. P. (2003). Semiparametric spatio-temporal frailty modeling. *Environmetrics* **14**, 523–535.
- BANERJEE, S. & DEY, D. K. (2005). Semiparametric proportional odds models for spatially correlated survival data. *Lifetime Data Analysis* **11**, 175–191.
- BANERJEE, S., WALL, M. M. & CARLIN, B. P. (2003). Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota. *Biostatistics* **4**, 123–142.
- BEARD, R. E. (1959). Note on some mathematical mortality models. In *The Lifespan of Animals*, G. E. W. Wolstenholme & M. O'Connor, eds. Little, Brown, Boston, pp. 302–311.
- BENDER, R., AUGUSTIN, T. & BLETTNER, M. (2005). Generating survival times to simulate Cox proportional hazards models. *Statistics in Medicine* **24**, 1713–1723.
- BURTON, A., ALTMAN, D. G., ROYSTON, P. & HOLDER, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine* **25**, 4279–4292.
- CHEN, M.-C. & BANDEEN-ROCHE, K. (2005). A diagnostic for association in bivariate survival models. *Lifetime Data Analysis* **11**, 245–264.
- CHU, R., THABANE, L., MA, J., HOLBROOK, A., PULLENAYEGUM, E. & DEVEREAUX, P. (2011). Comparing methods to estimate treatment effects on a continuous outcome in multicentre randomized controlled trials: A simulation study. *BMC Medical Research Methodology* **11**, 21.

- CLAYTON, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65**, 141–151.
- COLLETT, D. (2003). *Modelling Survival Data in Medical Research*. Boca Raton: Chapman & Hall/CRC, 2nd ed.
- CORTIÑAS ABRAHANTES, J. & BURZYKOWSKI, T. (2005). A version of the EM algorithm for proportional hazard model with random effects. *Biometrical Journal* **47**, 847–862.
- CORTIÑAS ABRAHANTES, J., LEGRAND, C., BURZYKOWSKI, T., JANSSEN, P., DUCROCQ, V. & DUCHATEAU, L. (2007). Comparison of different estimation procedures for proportional hazards model with random effects. *Computational Statistics & Data Analysis* **51**, 3913–3930.
- CUI, S. & SUN, Y. (2004). Checking for the gamma frailty distribution under the marginal proportional hazards frailty model. *Statistica Sinica* **14**, 249–267.
- DAVISON, A. C. & HINKLEY, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge University Press.
- DE GRUTTOLA, V. G., CLAX, P., DEMETS, D. L., DOWNING, G. J., ELLENBERG, S. S., FRIEDMAN, L., GAIL, M. H., PRENTICE, R., WITTES, J. & ZEGER, S. L. (2001). Considerations in the evaluation of surrogate endpoints in clinical trials: Summary of a National Institutes of Health Workshop. *Controlled Clinical Trials* **22**, 485–502.
- DEMIRTAS, H. (2007). The design of simulation studies in medical statistics by Andrea Burton, Douglas G. Altman, Patrick Royston and Roger L. Holder, *Statistics in Medicine* 2006; 25:4279–4292. *Statistics in Medicine* **26**, 3818–3821.
- DIGGLE, P. J. & RIBEIRO JUNIOR, P. J. (2007). *Model-based Geostatistics*. New York: Springer.
- DIVA, U., BANERJEE, S. & DEY, D. K. (2007). Modelling spatially correlated survival data for individuals with multiple cancers. *Statistical Modelling* **7**, 191–213.

- DIVA, U., DEY, D. K. & BANERJEE, S. (2008). Parametric models for spatially correlated survival data for individuals with multiple cancers. *Statistics in Medicine* **27**, 2127–2144.
- DONOHO, D. (2002). How to be a widely cited author in the mathematical sciences. *in-cites*. Available at <http://www.in-cites.com/scientists/DrDavidDonoho.html>.
- DONOHUE, M. C., OVERHOLSER, R., XU, R. & VAIDA, F. (2011). Conditional Akaike information under generalized linear and proportional hazards mixed models. *Biometrika* **98**, 685–700.
- DOS SANTOS, D. M., DAVIES, R. B. & FRANCIS, B. (1995). Nonparametric hazard versus nonparametric frailty distribution in modelling recurrence of breast cancer. *Journal of Statistical Planning and Inference* **47**, 111–127.
- DUCHATEAU, L. & JANSSEN, P. (2005). Understanding heterogeneity in generalized mixed and frailty models. *The American Statistician* **59**, 143–146.
- DUCHATEAU, L. & JANSSEN, P. (2008). *The Frailty Model*. New York: Springer.
- DUCHATEAU, L., JANSSEN, P., LINDSEY, P., LEGRAND, C., NGUTI, R. & SYLVESTER, R. (2002). The shared frailty model and the power for heterogeneity tests in multicenter trials. *Computational Statistics & Data Analysis* **40**, 603–620.
- DUCROCQ, V. & CASELLA, G. (1996). A Bayesian analysis of mixed survival models. *Genetics Selection Evolution* **28**, 505–529.
- ECONOMOU, P. & CARONI, C. (2008). Graphical tests for the frailty distribution in the shared frailty model. *Communications in Statistics - Simulation and Computation* **37**, 978–992.
- FANG, Y., MADSEN, L. & LIU, L. (2014). Comparison of two methods to check copula fitting. *IAENG International Journal of Applied Mathematics* **44**, 53–61.
- FAUCETT, C. L., SCHENKER, N. & TAYLOR, J. M. G. (2002). Survival analysis using auxiliary variables via multiple imputation, with application to AIDS clinical trial data. *Biometrics* **58**, 37–47.

- FONG, D. Y. T., LAM, K. F., LAWLESS, J. F. & LEE, Y. W. (2001). Dynamic random effects models for times between repeated events. *Lifetime Data Analysis* **7**, 345–362.
- FUJINO, Y. (1979). Tests for the homogeneity of a set of variances against ordered alternatives. *Biometrika* **66**, 133–139.
- GARFIELD, E. (1989). Delayed recognition in scientific discovery: citation frequency analysis aids the search for case histories. In *Essays of an Information Scientist: Creativity, Delayed Recognition, and other Essays*, vol. 12.
- GEERDENS, C., CLAESKENS, G. & JANSSEN, P. (2013). Goodness-of-fit tests for the frailty distribution in proportional hazards models with shared frailty. *Biostatistics* **14**, 433–446.
- GEHAN, E. A. & FREIREICH, E. J. (2011). The 6-MP versus placebo clinical trial in acute leukemia. *Clinical Trials* **8**, 288–297.
- GETACHEW, Y. (2013). Modeling the effect of distance from a hydro-electric dam on malaria incidence based on frailty and mixed Poisson regression models. PhD thesis. <https://biblio.ugent.be/record/4215574>.
- GETACHEW, Y., JANSSEN, P., YEWHALAW, D., SPEYBROECK, N. & DUCHATEAU, L. (2013). Coping with time and space in modelling malaria incidence: a comparison of survival and count regression models. *Statistics in Medicine* **32**, 3224–3233.
- GILL, R. D. (1984). Understanding Cox’s regression model: A martingale approach. *Journal of the American Statistical Association* **79**, 441–447.
- GJESSING, H. K., AALEN, O. O. & HJORT, N. L. (2003). Frailty models based on Lévy processes. *Advances in Applied Probability* **35**, 532–550.
- GLIDDEN, D. V. (2007). Pairwise dependence diagnostics for clustered failure-time data. *Biometrika* **94**, 371–385.
- GLIDDEN, D. V. & VITTINGHOFF, E. (2004). Modelling clustered survival data from multicentre clinical trials. *Statistics in Medicine* **23**, 369–388.

- GOETHALS, K., JANSSEN, P. & DUCHATEAU, L. (2008). Frailty models and copulas: similarities and differences. *Journal of Applied Statistics* **35**, 1071–1079.
- GOETHALS, K., JANSSEN, P. & DUCHATEAU, L. (2012). Frailties and copulas, not two of a kind. *Risk and Decision Analysis* **3**, 247–253.
- GOTTARD, A., MATTEI, A. & VIGNOLI, D. (2012). How education affects fertility in the presence of time-varying frailty component. Tech. rep., Dipartimento di Statistica "Giuseppe Parenti", Università degli Studi di Firenze. Available at <http://eprints.unifi.it/archive/00002433/>.
- GOUTIS, C. & CASELLA, G. (1999). Explaining the saddlepoint approximation. *The American Statistician* **53**, 216–224.
- GOVINDARAJULU, U. S., LIN, H., LUNETTA, K. L. & D'AGOSTINO, R. B. (2011). Frailty models: Applications to biomedical and genetic studies. *Statistics in Medicine* **30**, 2754–2764.
- GRATWOHL, A. (2012). The EBMT risk score. *Bone Marrow Transplantation* **47**, 749–756.
- GREVEN, S. & KNEIB, T. (2010). On the behaviour of marginal and conditional AIC in linear mixed models. *Biometrika* **97**, 773–789.
- GRIGOLETTO, M. & AKRITAS, M. G. (1999). Analysis of covariance with incomplete data via semiparametric model transformations. *Biometrics* **55**, 1177–1187.
- HAUCK, W. W., ANDERSON, S. & MARCUS, S. M. (1998). Should we adjust for covariates in nonlinear regression analyses of randomized trials? *Controlled Clinical Trials* **19**, 249–256.
- HENDERSON, R., SHIMAKURA, S. & GORST, D. (2002). Modeling spatial variation in leukemia survival data. *Journal of the American Statistical Association* **97**, 965–972.
- HIRSCH, K. & WIENKE, A. (2012). Software for semiparametric shared gamma and log-normal frailty models: An overview. *Computer Methods and Programs in Biomedicine* **107**, 582–597.
- HOUGAARD, P. (1995). Frailty models for survival data. *Lifetime Data Analysis* **1**, 255–273.

- HOUGAARD, P. (2000). *Analysis of Multivariate Survival Data*. New York: Springer.
- HOUGAARD, P., HARVALD, B. & HOLM, N. V. (1992). Measuring the similarities between the lifetimes of adult Danish twins born between 1881–1930. *Journal of the American Statistical Association* **87**, 17–24.
- HUANG, X. & LIU, L. (2007). A joint frailty model for survival and gap times between recurrent events. *Biometrics* **63**, 389–397.
- JIN, X. & CARLIN, B. P. (2005). Multivariate parametric spatiotemporal models for county level breast cancer survival data. *Lifetime Data Analysis* **11**, 5–27.
- KAHAN, B. C. (2014). Accounting for centre-effects in multicentre trials with a binary outcome - when, why, and how? *BMC Medical Research Methodology* **14**, 20.
- KAHAN, B. C. & MORRIS, T. P. (2013). Analysis of multicentre trials with continuous outcomes: when and how should we account for centre effects? *Statistics in Medicine* **32**, 1136–1149.
- KALBFLEISCH, J. D. & PRENTICE, R. L. (2002). *The Statistical Analysis of Failure Time Data*. New York: Wiley, 2nd ed.
- KAPLAN, E. L. & MEIER, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53**, 457–481.
- KATZ, M. H. & HAUCK, W. W. (1993). Proportional hazards (Cox) regression. *Journal of General Internal Medicine* **8**, 702–711.
- KLEIN, J. P. & MOESCHBERGER, M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer, 2nd ed.
- KLEIN, J. P., PELZ, C. & ZHANG, M.-J. (1999). Modeling random effects for censored data by a multivariate normal regression model. *Biometrics* **55**, 497–506.
- LAKHAL, L., RIVEST, L.-P. & BEAUDOIN, D. (2009). IPCW estimator for Kendall’s tau under bivariate censoring. *The International Journal of Biostatistics* **5**, 1–20.

-
- LAM, K. F., LEE, Y. W. & LEUNG, T. L. (2002). Modeling multivariate survival data by a semiparametric random effects proportional odds model. *Biometrics* **58**, 316–323.
- LAWLESS, J. F. (2002). *Statistical Models and Methods for Lifetime Data*. New York: Wiley, 2nd ed.
- LEGRAND, C., DUCHATEAU, L., SYLVESTER, R., JANSSEN, P., VAN DER HAGE, J. A., VAN DE VELDE, C. J. H. & THERASSE, P. (2006). Heterogeneity in disease free survival between centers: lessons learned from an EORTC breast cancer trial. *Clinical Trials* **3**, 10–18.
- LEGRAND, C., SYLVESTER, R., DUCHATEAU, L., JANSSEN, P. & THERASSE, P. (2002). Treatment outcome studies: pitfalls in current methods and practice. *European Journal of Cancer* **38**, 1173–1180.
- LEWIS, J. A. (1999). Statistical principles for clinical trials (ICH E9): an introductory note on an international guideline. *Statistics in Medicine* **18**, 1903–1942.
- LI, Y. & LIN, X. (2006). Semiparametric normal transformation models for spatially correlated survival data. *Journal of the American Statistical Association* **101**, 591–603.
- LI, Y. & RYAN, L. (2002). Modeling spatial survival data using semiparametric frailty models. *Biometrics* **58**, 287–297.
- LIANG, H., WU, H. & ZOU, G. (2008). A note on conditional AIC for linear mixed-effects models. *Biometrika* **95**, 773–778.
- LIANG, H. & ZOU, G. (2008). Improved AIC selection strategy for survival analysis. *Computational Statistics & Data Analysis* **52**, 2538–2548.
- LIANG, K.-Y., SELF, S. G., BANDEEN-ROCHE, K. J. & ZEGER, S. L. (1995). Some recent developments for regression analysis of multivariate failure time data. *Lifetime Data Analysis* **1**, 403–415.
- LIN, P.-S. (2012). Analysis of spatial frailty models by a weighted estimating equation. *Journal of Statistical Planning and Inference* **142**, 1436–1444.

-
- LITTELL, R. C., MILLIKEN, G. A., STROUP, W. W., WOLFINGER, R. D. & SCHABENBERGER, O. (2006). *SAS for Mixed Models*. SAS Institute, 2nd ed.
- LITTELL, R. C., PENDERGAST, J. & NATARAJAN, R. (2000). Modelling covariance structure in the analysis of repeated measures data. *Statistics in Medicine* **19**, 1793–1819.
- LIU, L., WOLFE, R. A. & HUANG, X. (2004). Shared frailty models for recurrent events and a terminal event. *Biometrics* **60**, 747–756.
- LIU, L. & YU, Z. (2008). A likelihood reformulation method in non-normal random effects models. *Statistics in Medicine* **27**, 3105–3124.
- LOCALIO, A. R., BERLIN, J. A., TEN HAVE, T. R. & KIMMEL, S. E. (2001). Adjustments for center in multicenter studies: An overview. *Annals of Internal Medicine* **135**, 112–123.
- LÓPEZ-DE-ULLIBARRI, I., JANSSEN, P. & CAO, R. (2012). Continuous covariate frailty models for censored and truncated clustered data. *Journal of Statistical Planning and Inference* **142**, 1864–1877.
- MANDA, S. O. M. & MEYER, R. (2005). Bayesian inference for recurrent events data using time-dependent frailty. *Statistics in Medicine* **24**, 1263–1274.
- MANTEL, N., BOHIDAR, N. R. & CIMINERA, J. L. (1977). Mantel-Haenszel analyses of litter-matched time-to-response data, with modifications for recovery of interlitter information. *Cancer Research* **37**, 3863–3868.
- MASSONNET, G., BURZYKOWSKI, T. & JANSSEN, P. (2006). Resampling plans for frailty models. *Communications in Statistics - Simulation and Computation* **35**, 497–514.
- MASSONNET, G., JANSSEN, P. & BURZYKOWSKI, T. (2008). Fitting conditional survival models to meta-analytic data by using a transformation toward mixed-effects models. *Biometrics* **64**, 834–842.
- MAUGUEN, A., RACHET, B., MATHOULIN-PÉLISSIER, S., MACGROGAN, G., LAURENT, A. & RONDEAU, V. (2013). Dynamic prediction of risk of death using history of cancer recurrences in joint frailty models. *Statistics in Medicine* **32**, 5366–5380.

- MAZROUI, Y., MATHOULIN-PÉLISSIER, S., SOUBEYRAN, P. & RONDEAU, V. (2012). General joint frailty model for recurrent event data with a dependent terminal event: Application to follicular lymphoma data. *Statistics in Medicine* **31**, 1162–1176.
- MCGILCHRIST, C. A. (1993). REML estimation for survival models with frailty. *Biometrics* **49**, 221–225.
- MICHIELS, S., BAUJAT, B., MAHÉ, C., SARGENT, D. J. & PIGNON, J.-P. (2005). Random effects survival models gave a better understanding of heterogeneity in individual patient data meta-analyses. *Journal of Clinical Epidemiology* **58**, 238–245.
- MOESCHBERGER, M. L. & KLEIN, J. P. (1985). A comparison of several methods of estimating the survival function when there is extreme right censoring. *Biometrics* **41**, 253–259.
- MOGER, T. A., AALEN, O. O., HALVORSEN, T. O., STORM, H. H. & TRETTLI, S. (2004). Frailty modelling of testicular cancer incidence using Scandinavian data. *Biostatistics* **5**, 1–14.
- MUNDA, M. & LEGRAND, C. (2014a). Adjusting for centre heterogeneity in multicentre clinical trials with a time-to-event outcome. *Pharmaceutical Statistics* **13**, 145–152.
- MUNDA, M. & LEGRAND, C. (2014b). A diagnostic plot for the frailty distribution in the shared frailty model. Submitted.
- MUNDA, M., LEGRAND, C., DUCHATEAU, L. & JANSSEN, P. (2014). Testing for decreasing heterogeneity in a new time-varying frailty model. Submitted.
- MUNDA, M., ROTOLO, F. & LEGRAND, C. (2012). parfm: Parametric frailty models in R. *Journal of Statistical Software* **51**, 1–20.
- MURRAY, D. M., VARNELL, S. P. & BLITSTEIN, J. L. (2004). Design and analysis of group-randomized trials: a review of recent methodological developments. *American Journal of Public Health* **94**, 423–432.
- NAN, B., LIN, X., LISABETH, L. D. & HARLOW, S. D. (2006). Piecewise constant cross-ratio estimation for association of age at a marker event and age at menopause. *Journal of the American Statistical Association* **101**, 65–77.

- NARDI, A. & SCHEMPER, M. (2003). Comparing Cox and parametric models in clinical studies. *Statistics in Medicine* **22**, 3597–3610.
- NELSEN, R. B. (2006). *An Introduction to Copulas*. New York: Springer, 2nd ed.
- NIELSEN, G. G., GILL, R. D., ANDERSEN, P. K. & SØRENSEN, T. I. A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics* **19**, 25–43.
- OAKES, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association* **84**, 487–493.
- OJAMBO, P. S. & KANG, E. L. (2013). Modeling spatial frailties in survival analysis of cucurbit downy mildew epidemics. *Phytopathology* **103**, 216–227.
- O’QUIGLEY, J. & STARE, J. (2002). Proportional hazards models with frailties and random effects. *Statistics in Medicine* **21**, 3219–3233.
- PAIK, M. C., TSAI, W.-Y. & OTTMAN, R. (1994). Multivariate survival analysis using piecewise gamma frailty. *Biometrics* **50**, 975–988.
- PANKRATZ, V. S., DE ANDRADE, M. & THERNEAU, T. M. (2005). Random-effects Cox proportional hazards model: General variance components methods for time-to-event data. *Genetic Epidemiology* **28**, 97–109.
- PENNELL, M. L. & DUNSON, D. B. (2006). Bayesian semiparametric dynamic frailty models for multiple event time data. *Biometrics* **62**, 1044–1052.
- PRICE, D. L. & MANATUNGA, A. K. (2001). Modelling survival data with a cured fraction using frailty models. *Statistics in Medicine* **20**, 1515–1527.
- REN, S., LAI, H., TONG, W., AMINZADEH, M., HOU, X. & LAI, S. (2010). Nonparametric bootstrapping for hierarchical data. *Journal of Applied Statistics* **37**, 1487–1498.
- RICE, W. R. & GAINES, S. D. (1994a). Extending nondirectional heterogeneity tests to evaluate simply ordered alternative hypotheses. *Proceedings of the National Academy of Sciences* **91**, 225–226.

-
- RICE, W. R. & GAINES, S. D. (1994b). The ordered-heterogeneity family of tests. *Biometrics* **50**, 746–752.
- RIPATTI, S. & PALMGREN, J. (2000). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics* **56**, 1016–1022.
- RONDEAU, V. (2010). Statistical models for recurrent events and death: Application to cancer events. *Mathematical and Computer Modelling* **52**, 949–955.
- RONDEAU, V., MARZROUI, Y. & GONZALEZ, J. R. (2012). frailty-pack: An R package for the analysis of correlated survival data with frailty models using penalized likelihood estimation or parametrical estimation. *Journal of Statistical Software* **47**, 1–28.
- RONDEAU, V., MATHOULIN-PÉLISSIER, S., JACQMIN-GADDA, H., BROUSTE, V. & SOUBEYRAN, P. (2007). Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events. *Biostatistics* **8**, 708–721.
- RONDEAU, V., MATHOULIN-PÉLISSIER, S., TANNEAU, L., SASCO, A. J., MACGROGAN, G. & DEBLED, M. (2010). Separate and combined analysis of successive dependent outcomes after breast-conservation surgery: recurrence, metastases, second cancer and death. *BMC Cancer* **10**, 697.
- RONDEAU, V., MICHIELS, S., LIQUET, B. & PIGNON, J.-P. (2008). Investigating trial and treatment heterogeneity in an individual patient data meta-analysis of survival data by means of the penalized maximum likelihood approach. *Statistics in Medicine* **27**, 1894–1910.
- RONDEAU, V., PIGNON, J.-P. & MICHIELS, S. (2011). A joint model for the dependence between clustered times to tumour progression and deaths: A meta-analysis of chemotherapy in head and neck cancer. *Statistical Methods in Medical Research* **0**, 1–19.
- RYAN, T. P. & WOODALL, W. H. (2005). The most-cited statistical papers. *Journal of Applied Statistics* **32**, 461–474.
- SHIH, J. & LOUIS, T. (1995). Assessing gamma frailty models for clustered failure time data. *Lifetime Data Analysis* **1**, 205–220.

- TAYLOR, J. M. G. & WANG, Y. (2002). Surrogate markers and joint models for longitudinal and survival data. *Controlled Clinical Trials* **23**, 626–634.
- TEODORESCU, B., VAN KEILEGOM, I. & CAO, R. (2010). Generalized time-dependent conditional linear models under left truncation and right censoring. *Annals of the Institute of Statistical Mathematics* **62**, 465–485.
- THERNEAU, T., GRAMBSCH, P. M. & PANKRATZ, V. S. (2003). Penalized survival models and frailty. *Journal of Computational and Graphical Statistics* **12**, 156–175.
- THERNEAU, T. M. & GRAMBSCH, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. New York: Springer.
- VAIDA, F. & BLANCHARD, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika* **92**, 351–370.
- VAUPEL, J. W., MANTON, K. G. & STALLARD, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16**, 439–454.
- VISWANATHAN, B. & MANATUNGA, A. K. (2001). Diagnostic plots for assessing the frailty distribution in multivariate survival data. *Lifetime Data Analysis* **7**, 143–155.
- VOLINSKY, C. T. & RAFTERY, A. E. (2000). Bayesian information criterion for censored survival models. *Biometrics* **56**, 256–262.
- WANG, K., YAU, K. K. W., LEE, A. H. & McLACHLAN, G. J. (2007). Multilevel survival modelling of recurrent urinary tract infections. *Computer Methods and Programs in Biomedicine* **87**, 225–229.
- WANG, S. T., KLEIN, J. P. & MOESCHBERGER, M. L. (1995). Semi-parametric estimation of covariate effects using the positive stable frailty model. *Applied Stochastic Models and Data Analysis* **11**, 121–133.
- WANG, W. & WELLS, M. T. (1997). Nonparametric estimators of the bivariate survival function under simplified censoring conditions. *Biometrika* **84**, 863–880.

- WIENKE, A. (2010). *Frailty Models in Survival Analysis*. Boca Raton, Florida: Chapman and Hall/CRC.
- WINTREBERT, C. M. A., PUTTER, H., ZWINDERMAN, A. H. & VAN HOUWELINGEN, H. C. (2004). Centre-effect on survival after bone marrow transplantation: Application of time-dependent frailty models. *Biometrical Journal* **46**, 512–525.
- YAU, K. K. W. & MCGILCHRIST, C. A. (1998). ML and REML estimation in survival analysis with time dependent correlated frailty. *Statistics in Medicine* **17**, 1201–1213.
- YEWHALAW, D., GETACHEW, Y., TUSHUNE, K., W/MICHAEL, K., KASSAHUN, W., DUCHATEAU, L. & SPEYBROECK, N. (2013). The effect of dams and seasons on malaria incidence and anopheles abundance in Ethiopia. *BMC Infectious Diseases* **13**, 161.
- YEWHALAW, D., KASSAHUN, W., WOLDEMICHAEL, K., TUSHUNE, K., SUDAKER, M., KABA, D., DUCHATEAU, L., VAN BORTEL, W. & SPEYBROECK, N. (2010). The influence of the Gilgel-Gibe hydroelectric dam in Ethiopia on caregivers' knowledge, perceptions and health-seeking behaviour towards childhood malaria. *Malaria Journal* **9**, 47.
- YEWHALAW, D., LEGESSE, W., VAN BORTEL, W., GEBRE-SELASSIE, S., KLOOS, H., DUCHATEAU, L. & SPEYBROECK, N. (2009). Malaria and water resource development: the case of Gilgel-Gibe hydroelectric dam in Ethiopia. *Malaria Journal* **8**, 21.
- YUE, H. & CHAN, K. S. (1997). A dynamic frailty model for multivariate survival data. *Biometrics* **53**, 785–793.