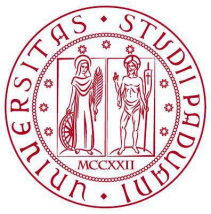


UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Statistiche
SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE STATISTICHE
CICLO XXV

# Frailty multi-state models for the analysis of survival data from multicenter clinical trials

**Direttore della Scuola:** Ch.mo Prof. Monica Chiogna

**Supervisore:** Ch.mo Prof. Monica Chiogna

**Co-supervisore:** Ch.mo Prof. Catherine Legrand

**Dottorando:** Federico Rotolo

31 gennaio 2013

**Abstract**

Proportional hazards models are among the most popular regression models in survival analysis. Multi-state models generalise them in the sense of jointly considering different types of events along with their interrelations, whereas frailty models introduce random effects to account for unobserved risk factors, possibly shared by groups of subjects. The integration of frailty and multi-state methodology is interesting to control for unobserved heterogeneity in presence of complex event history structures, particularly appealing in multicenter clinical trials applications.

In the present thesis we propose the incorporation of nested frailties in the transition-specific hazard function; then, we develop and evaluate both parametric and semi-parametric inference. Simulation studies, performed thanks to an innovative method for generating dependent multi-state survival data, show that parametric inference is correct but extremely imprecise, whilst semiparametric methods are very competitive to evaluate the effect of covariates.

Two case studies are presented, relative to cancer multicenter clinical trials. The multi-state nature of the models allows to study the treatment effect taking into account intermediate events, while the presence of frailties reduces the attenuation effect due to clustering.

Finally, we present two new software tools, one to fit parametric frailty models with up to twenty possible combinations of baseline and frailty distributions, and one implementing semiparametric inference for multilevel frailty models, essential to fit the new nested frailty multi-state models.

**Résumé**

Les modèles à risques proportionnels sont parmi les modèles de régression les plus célèbres de l'analyse de survie. Les modèles multi-états constituent une généralisation de ceux-ci qui prend simultanément en considération différents types d'événements et leurs interrelations ; les modèles de type frailty, quant à eux, utilisent des effets aléatoires pour tenir compte de facteurs de risque qui ne sont pas observés, éventuellement partagés par des groupes de sujets. L'intégration des méthodologies multi-états et frailty peut s'avérer très intéressante afin de contrôler l'hétérogénéité non observée en présence de structures complexes d'événements, ce qui est particulièrement attractif dans les applications cliniques d'études multicentriques.

Dans cette thèse on propose d'intégrer des frailties imbriqués dans la fonction de risque transition-spécifique ; on développe et on évalue des méthodes d'inférence paramétrique et semi-paramétrique. Par le biais d'une étude de simulation, effectuée grâce à une méthode innovante pour générer des données multi-états dépendantes, on montre que l'inférence paramétrique est correcte mais extrêmement imprécise, alors que les méthodes semi-paramétriques sont très compétitives pour évaluer l'effet des covariables.

Deux cas d'étude sont présentés concernant des études cliniques multicentriques en oncologie. La nature multi-états de ces modèles permet d'étudier l'effet du traitement en tenant compte des événements intermédiaires. La présence des frailties réduit l'effet d'atténuation en tenant compte de la corrélation due au regroupement.

Enfin, on présente deux nouveaux outils informatiques, le premier pour estimer des modèles frailty paramétriques avec jusqu'à vingt combinaisons possibles entre la distribution de la fonction baseline et celle des frailties ; le deuxième implémente des méthodes d'inférence semi-paramétrique pour des modèles frailty multiniveaux, ce qui est très utile pour estimer les nouveaux modèles multi-états avec frailties imbriqués.

**Riassunto**

I modelli a rischi proporzionali sono tra i modelli di regressione più conosciuti ed utilizzati in analisi di sopravvivenza. In modelli multi-stato sono una loro generalizzazione che permette di considerare congiuntamente diversi tipi di eventi e le loro interrelazioni, mentre i modelli di tipo *frailty* introducono effetti casuali per tenere conto di fattori di rischio non osservati, eventualmente in comune tra soggetti appartenenti allo stesso gruppo. L'integrazione dei modelli multi-stato e dei modelli *frailty* è interessante al fine di controllare l'eterogeneità non osservata in presenza di strutture complesse di eventi, particolarmente interessante nel caso di studi clinici multicentro.

In questa tesi proponiamo di incorporare *frailty* annidati nella funzione di rischio transizione-specifica, quindi sviluppiamo e valutiamo metodi di inferenza sia parametrica che semiparametrica. Studi di simulazione, effettuati grazie a un metodo innovativo per generare dati di sopravvivenza multi-stato dipendenti, mostrano che l'inferenza parametrica è corretta ma estremamente imprecisa, mentre i metodi semiparametrici sono molto competitivi per valutare l'effetto delle covariate.

Due casi-studio relativi a studi clinici multicento in oncologia vengono quindi presentati. La natura multi-stato dei modelli permette di studiare l'effetto del trattamento tenendo conto degli eventi intermedi, mentre la presenza di *frailty* riduce l'effetto di attenuazione dovuto ai gruppi di pazienti.

Infine, presentiamo due nuovi strumenti software, uno per stimare modelli *frailty* parametrici con fino a venti possibili combinazioni di distribuzioni baseline e *frailty*, e un altro che implementa metodi di inferenza semiparametrica per modelli *frailty* multilivello, essenziali per stimare i nuovi modelli multi-stato con *frailty* annidati.

# Acknowledgements

I would like to first spend a few words to express my gratitude to many people who played a role in many respects during my PhD.

My first thought is for my parents and my sister, who constantly supported me with curiosity and enthusiasm; and without whose advice I wouldn't have even begun this adventure.

These years wouldn't have been so extraordinary without the great time passed in Louvain-la-Neuve. The guidance and teachings of Prof. Catherine Legrand not only offered me the possibility to grow up so much from a professional and personal point of view, but also opened new unexpected routes for my professional future. I am also grateful to Prof. Ingrid Van Keilegom for fruitfully supervising my research when little Adèle arrived and Prof. Legrand was absent for a while.

This unique experience wouldn't have been possible without Prof. Monica Chiogna, who has always showed trust in me, providing continuous and enthusiast backing which helped me to address any situation with more confidence and optimism. Together with her, appropriate acknowledgement is due to Dr. Andrea Callegaro for his key contribution in outlining the focus objectives of the research project of this thesis.

At my arrival in Belgium I have been surprised by the warm welcome I received by a number of extraordinary people, among which a special role played notably Marco and my officemates Anne, Habiba, Diane and many others at the ISBA. Even though I cannot mention all of them, I will keep a beautiful memory of my 'Belgian period' thanks to them.

Finally, I would like to express my particular thankfulness to Giulia for patiently persevering despite the prolonged distance. Our starting new life, converting dreams into projects, is the best recompense for such an effort.

---

The author thanks

> the **European Organization for Research and Treatment of Cancer** for permission to use the data from EORTC studies 30781, 30782, 30791, 30831, 30832, 30845, 30863

and

> **Sanofi–Aventis** for permission to use the data from the TROPIC clinical trial (#EFC6193)

for this research.

The contents of this publication and methods used are solely the responsibility of the author and do not necessarily represent the official views of either the EORTC or Sanofi–Aventis Headquarters.

---

# Contents

# List of Figures

# List of Tables

# Definitions and notation

Definitions are given for the nested frailty multi-state model (Ch. 5), with indexes

| | |
|---|---|
| $h = 1, \ldots, H$ | the cluster |
| $n_1, \ldots, n_H$ | the cluster sizes, with $n = \sum_{h=1}^{H} n_h$ the total number of subjects |
| $i = 1, \ldots, n_h$ | the subject within the cluster $h$ |
| $q = 1, \ldots, Q$ | the transition type |

For multi-state models and for shared and nested frailty models the indexes can be deduced straightforwardly and are given in Chapter 3.

---

| | |
|---|---|
| $\mathbb{1}(\cdot)$ | the indicator function which is 1 if its argument is true and 0 otherwise |
| $\mathsf{E}[\cdot]$ | the expected value |
| $\mathsf{V}[\cdot]$ | the variance |
| $\mathsf{Cov}[\cdot, \cdot]$ | the covariance |
| $t_{qhi}$ | the event time |
| $c_{qhi}$ | the right censoring time |
| $y_{qhi}$ | the observed event or censoring time: $\min(t_{qhi}, c_{qhi})$ |
| $\delta_{qhi}$ | the event/censoring indicator: $\mathbb{1}(t_{qhi} \leq c_{qhi})$ |
| $d_{qh}$ | the number of events of type $q$ in cluster $h$: $\sum_{i=1}^{n_h} \delta_{qhi}$ |
| $d_h$ | the total number of events of any type in cluster $h$: $\sum_{i=1}^{n_h} \sum_{q=1}^{Q} \delta_{qhi} = \sum_{q=1}^{Q} d_{qh}$ |
| $\tau_{qhi}$ | the left truncation time |
| $\tau$ | the Kendall's Tau |
| $\mathbf{x}_{qhi}$ | the covariates |
| $\boldsymbol{\beta}_q$ | the regression parameters for transition $q$ |
| $U_{qh}$ | the cluster–transition frailty term: $V_h W_{qh}$ |
| $V_h$ | the cluster level random term |
| $W_{qh}$ | the cluster–transition level random term |
| $f_U(\cdot; \boldsymbol{\theta})$ | the distribution of the frailty $U$ |
| $f_V(\cdot; \boldsymbol{\theta}_V)$ | the distribution of $V_h$, $\forall h$ |

| | |
|---|---|
| $f_W(\cdot;\boldsymbol{\theta}_q)$ | the distribution of $W_q$ |
| $\boldsymbol{\theta}$ | the frailty parameters $(\theta_V, \theta_1, \ldots, \theta_Q)^\top$ |
| $\theta_V$ | the frailty parameter for $V_h$, $\forall h$ |
| $\theta_q$ | the frailty parameter for $W_q$ |
| $\boldsymbol{\xi}_q$ | the baseline hazard parameters for transition $q$ |
| $\boldsymbol{\zeta}$ | the vector of all the parameters: $\left(\boldsymbol{\xi}^\top, \boldsymbol{\theta}^\top, \boldsymbol{\beta}^\top\right)^\top$ |
| $\lambda_{q0}(\cdot;\boldsymbol{\xi}_q)$ | the baseline hazard function for transition $q$, sometimes simply $\lambda_{q0}(\cdot)$ |
| $\lambda_{qhi}(\cdot \mid u_{qh})$ | the conditional hazard (5.1), also $\lambda_{qhi}(\cdot \mid v_h, w_{qh})$ |
| $\Lambda_{q0}(\cdot;\boldsymbol{\xi}_q)$ | the cumulative baseline hazard function for transition $q$, sometimes simply $\Lambda_{q0}(\cdot)$ |
| $\Gamma(\cdot)$ | the gamma function |
| $\mathcal{L}(\cdot)$ | the Laplace transform |
| $L(\cdot)$ | the full likelihood function |
| $L_{\mathsf{C}}(\cdot)$ | the conditional likelihood function |
| $L_{\mathsf{M}}(\cdot)$ | the marginal likelihood function |
| $L_{\mathsf{PM}}(\cdot)$ | the penalised marginal likelihood function |
| $L_{\mathsf{P}}(\cdot)$ | the partial likelihood function |
| $L_{\mathsf{PP}}(\cdot)$ | the penalised partial likelihood function |
| $\ell_{\cdots}(\cdot)$ | the $\cdots$ loglikelihood: $\log L_{\cdots}(\cdot)$ |
| Gam | the gamma distribution |
| PS | the positive stable distribution |
| IG | the inverse Gaussian distribution |
| PVF | the power variance function distribution |
| CP | the compound Poisson distribution |
| LN | the lognormal distribution |
| $\mathbf{C}(\cdot)$ | the copula function |
| $\vartheta$ | the copula dependence parameter |
| $\mathcal{S}$ | the states set |
| $\mathcal{T}$ | the transitions set |
| $\mathcal{C}(\cdot)$ | the children set of a state |
| $\#(\cdot)$ | the size of a set |
| $\mathcal{Q}$ | a set of transition types |
| $\Omega_{\cdot,\cdot}$ | polynomial functions as defined by Equations 3.102 |

Table 1 – Definitions and notation.

# Chapter 1

# Introduction

## 1.1 Overview

Survival analysis techniques have an important role in biostatistics, as they allow an evaluation of the effect of different treatments and risk factors on the natural course of a disease. Survival data, also known as time-to-event or duration data, are measures of the time elapsed since an origin event until an event of interest. Their peculiarity relies in the fact that the times of those events which have not occurred yet are not completely missing data, but partial data: even though their true values are unknown, they are necessary greater than the value at the present moment. This partial information on the lower bound of missing values is called right censoring and is practically unavoidable for duration data.

The proportional hazards model, popularised by Cox (1972), is certainly one of the most widespread, used and studied regression models for time-to-event data. Frailty models and multi-state models are two broad families of survival models, extending proportional hazards models in two different directions to deal with distinct issues.

The problem of heterogeneity due to unobserved risk factors was first addressed by Clayton (1978) and Vaupel *et al.* (1979). They proposed to account for it by means of random effects, giving origin to a vast literature on what is now known as the frailty model or mixed proportional hazards model. If these effects are subject-specific, unobserved heterogeneity stands for overdispersion and the model is called univariate frailty model (Wienke, 2010, Ch. 3). On the contrary, in the case they are shared by groups of subjects, a clustering effect is there, i.e. observations belonging to the same group are dependent. This is the case of the so-called shared frailty models (Ch. 7–9 of Hougaard, 2000; Duchateau and Janssen, 2008).

On the other hand, multi-state models, widely investigated by Hougaard (2000, Ch. 5–6), Andersen and Keiding (2002) and Putter *et al.* (2007), are an appreciated tool to study the whole event history of subjects. They provide a framework to jointly model the hazards of many types of event, the occurrence of each can have an impact on the risk of the others. Most of the times, the Markov assumption is made, which allows estimation and prediction to be feasible in a very general and simple way (Hougaard, 1999a). This assumption seems to be a sensible approximation in many cases and it can be easily relaxed, giving the semi-Markov (or Markov extended) models.

The integration of frailty and multi-state models can provide powerful survival models to study

the risk of many interrelated events while accounting for dependence between grouped subjects. Many practical situations can be thought of in which such integration is of interest. The main problem motivating our research consists in answering clinical questions arising from multicenter cancer trials, while taking into account both the dependence between subjects and that between events. Typically, the focus is on the effect of the therapy on many different endpoints: death, local relapses, progression and distant metastases. The times to these events can be of different importance according to the context, but usually each one plays a role in changing the risk of the other ones. At the same time, patients recruited in the same hospital will arguably share some unobserved risk factors due to underlying features of the center, the population, the country and so on.

## 1.2 Main contributions of the thesis

The work presented in this thesis contributes to research in survival analysis from different points of view: modelling methodology and applications, simulation techniques, software availability.

Up to now, there existed no a general method in multi-state research for simulating data according to a given scenario. In this respect, the first major contribution of the present work is a simulation model for multi-state data (Rotolo *et al.*, 2012a). Dependence can be added between time variables of grouped subjects, to study the effect of clustering. One of the main assets of the proposed method is a numerical procedure, minimising a criterion function, which permits to choose simulation parameters in order to mimic information from real data. Moreover, the simulation method is able to introduce, thanks to copulas, dependence between times of different transitions while fixing the marginal distributions according to a given scenario. This is a useful tool to study, for instance, the robustness of (frailty) multi-state models with respect to departures from the Markov assumption.

The main contribution to modelling methodology consists in proposing multi-state models with two nested frailties, one to account for the global effect of unobserved group-specific factors, and a second one to deal with their effect on each event of interest. This model turns out to be a generalisation of what has been proposed so far in literature, i.e. models based on either shared or independent frailties. So, the nested frailty multi-state model is more flexible and requires less assumptions. It is very suited to describe the two-level dependence which links the times to different events within subjects which can share common unobserved risk factors.

Two estimation approaches have been developed and investigated in the thesis: a parametric and a semiparametric approach. First, fully parametric inference, based on maximum marginal likelihood, is considered. In this context double integration is needed to obtain the marginal likelihood; explicit integration is not possible in that case, and asymptotic or numerical approximation is needed. Simulation studies put in light that this approach is not much robust for small datasets, notably if few events are observed. Then, a semiparametric estimation approach, based on maximum penalised partial likelihood, is proposed and investigated (Rotolo and Legrand, 2012; Rotolo *et al.*, 2012b). We employed in this new context the EM-PL estimation algorithm, proposed by Horny (2009) for multilevel frailty models; this procedure alternates the EM and the PPL estimation methods at the two clustering levels. Two examples of applications are provided for phase III multicenter clinical trials: one to investigate the effect of further intravescical treatment after transurethral resection

in bladder cancer patients, and one to compare the efficacy of cabazitaxel versus mitoxantrone for metastatic progressive prostate cancer.

The third contribution of the present work is the implementation of frailty models in `R` (R Development Core Team, 2012) as ready-usable package. The `parfm` package (Rotolo and Munda, 2012) represents an effort to provide a user-friendly tool to fit parametric frailty models, with a wide range of baseline hazard functions and frailty distributions (Munda *et al.*, 2012). The importance of `parfm` is due to the fact that, even though parametric frailty models are largely used in literature, there existed no software offering a unified framework to deal with them. Finally, in a semiparametric context, we implemented in the `mlfm` package (Rotolo and Horny, 2012) the EM-PL estimation method proposed by Horny (2009) for multilevel frailty models. This can be used under mild conditions to fit in a semiparametric way the multi-state model with nested frailties.

## 1.3 Structure of the thesis

In Chapter 2 we introduce the motivating problem from a clinical point of view and its statistical implications. Then, we review the main literature on frailty models and multi-state models in Chapter 3. The simulation procedure for clustered multi-state data is provided in Chapter 4, whereas in Chapter 5 we propose the incorporation of correlated frailties into multi-state models, with the estimation methodology and we show a simulation study. Two case studies can be found in Chapter 6, with applications in multicenter bladder and prostate cancer clinical trials. Finally, Chapter 7 proposes two new software packages contributing to the research in frailty and multi-state research.

# Chapter 2

# The problem

## 2.1 The clinical motivating problem

Multicenter clinical trials are studies conducted collaboratively by many hospitals, according to a unique protocol. Even though they require a great organisational effort, they are more and more common in clinical research, and specifically in oncology, since they allow to reach the needed sample size in a shorter time (Fleiss, 1982; Knatterud *et al.*, 1998; Senn, 2007). This is a crucial asset when the effect to detect is expected to be small or when the endpoint of interest is relatively rare. In these cases, the time required for a single centre to collect a sufficient amount of data could be too long for the study to be useful in practice.

Another advantage of multicenter clinical trials is that the patients are sampled from a broader, and arguably less specific, population: they will come from a larger geographical area, have more heterogeneous genetics, lifestyles, nutritional habits and so on. This is supposed to make the results more generalisable to a wider population (Buyse *et al.*, 1984); nevertheless, this approach is sometimes criticized as most of the times centres are not randomly selected, (Yamaguchi and Ohashi, 1999).

Although the focus of many clinical trials is usually on the time to an event of interest, the study of the entire event-history of the subjects yields a sharper comprehension of the medical phenomenon (Andersen and Keiding, 2002; Aalen *et al.*, 2008). Event-history analysis allows to detect the effect of the occurrence of intermediate events on the risk of the following ones, as well as to separate the net effect of risk factors on the hazard of many competing events. The joint study of many events is of first importance in cancer research, the times to relapse, to progression or to metastases being of interest both *per se* and in relation to the time to death.

Multicenter clinical trials with different endpoints could benefit from the collaborative recruitment while taking advantage of the availability of a very complete picture of the evolution of the disease. Then, the study of event-histories of cancer patients in multicenter trials can offer, in reasonably short times, a deep insight into the dynamics of the disease.

## 2.2 The statistical challenge

If multicenter trials are very useful for recruiting a higher number of patients in shorter times, they give origin to the problem of clustering (Hougaard, 1995; Rodrìguez, 2005). Indeed, usually not all the important risk factors can be measured and it is reasonable to assume that at least some of them are shared by patients treated in the same hospital. Factors linked to the population (genetics and biology, behaviours, diet habits, lifestyles, etc.) add up to others due to socio-geography (environment, climate, healthcare systems and policies, etc.) and to the efficacy and quality of the whole patient care in each hospital. These shared latent factors make the level of risk vary across the groups, inducing dependence between observations belonging to the same cluster.

In survival analysis there exists two main approaches to deal with clustering (Glidden and Vittinghoff, 2004). Marginal modelling (Wei *et al.*, 1989; Lee *et al.*, 1992; Lin, 1994) aims at estimating the average effect of covariates in a given population, whereas shared frailty models (Duchateau and Janssen, 2008; Wienke, 2010) their conditional effect, given the reference risk level of the subpopulation to which a subject belongs. Frailty models are essentially proportional hazard models with a random effect shared by all the subjects in each group. Frailty models are getting more and more popular as they allow to model survival data by reducing inconsistency of regression parameter estimators in the case of common unobserved risk factors, and are naturally suited for multicenter trials (Ha *et al.*, 2011). In addition, they are a valuable tool to investigate the source and the type of dependence between clustered observations. Finally, they allow to correct the selection effect (Vaupel and Yashin, 1985b): if subjects have different baseline risk levels, the ones with higher risk will leave the study earlier and the risk at late times will be estimated based mostly on those with lower risk, so it will be underestimated (see Sec. 3.3, Fig. 3.4).

Though, frailty models are conceived for the time to a single event, while the interest in many contexts is in the joint study of the times to different endpoints and in the relations between them. Multi-state models (Hougaard, 1999b; Andersen and Keiding, 2002; Putter *et al.*, 2007) serve this aim, extending survival models for stochastic processes with more than the two classical states: “alive” and “dead”. They are particularly suited for cancer trials, where the occurrence of local relapses, progression or distant metastases is of strong interest, as well as the death time. On the other hand, inference for multi-state models is based on the assumption of independent and identically distributed observations, which does not hold in presence of clustering.

The incorporation of frailties into multi-state models is a primary aim of present research in survival analysis. In recent years, some first publications addressing this problem have appeared (e. g. Bhattacharyya and Klein, 2005; Yen *et al.*, 2010; Ma *et al.*, 2010; Liquet *et al.*, 2012), proposing some partial solutions. The main weakness of the present solutions is the dependence structure, as the frailties have been assumed to be shared by times to any transition or to be shared only by times to each transition, but independent across event types. These assumptions are done to keep the models simple and manageable, but they seem to be quite restrictive with respect to their meaning in biological terms. The use of correlated frailties, a midway between shared and independent, would be the most reasonable description of the reality. A possible way to obtain them is represented by the use of nested multiplicative frailties (Sastry, 1997; Rondeau *et al.*, 2006). The applicability of this solution to a multi-state framework raises more problems due to the link between different transitions within the same subject. The presence of several distinct events also

increases the parameter space size and requires the presence of sufficient information to recover the features of each transition.

# Chapter 3

# Background

## 3.1 Survival analysis

Survival analysis is an important field of statistics dealing with time-to-event, or duration, data which receives much attention in the literature: in addition to continuously appearing papers, many outstanding authors dedicated entire textbooks to this subject, exploiting different approaches and at levels of accessibility (Andersen *et al.*, 1993; Hougaard, 2000; Therneau and Grambsch, 2000; Kalbfleisch and Prentice, 2002; Klein and Moeschberger, 2003; Collett, 2003; Aalen *et al.*, 2008). Durations are typically measured as the amount of time elapsed since a so-called origin event until an event of interest. The first and most natural example is the time since birth to death, which gave origin to the name "survival data". Many examples exist in several research fields such as demography, industry and socio-economic sciences. We will concentrate on applications in clinical research, with particular attention to oncology. Then, typical endpoints are the occurrence of local relapse, progression, remission, distant metastasis or death since the diagnosis, the randomisation, the end of the therapy, surgery, etcetera.

Such kind of data are described by non-negative random variables. We will consider only the case of continuous measures; the problem of discrete times is usually tackled by interval censoring techniques, assuming that what is recorded is in fact a discrete measure of a continuous value. The peculiarity of duration data is the fact that unobserved times are not completely missing data, but partial data: if at a given moment $t$ the event of interest has not occurred yet, then the value of $T$ is missing but we know for sure that $T > t$. This partial information on the lower bound of missing values is called right censoring and is practically unavoidable for duration data.

Throughout this thesis, we will consider data collected on a sample of $n$ subjects; for the generic $i$-th one, the couple of random variables $(Y_i, \delta_i)$ is observed, with $Y_i = \min(T_i, C_i)$ the minimum between the event time $T_i$ and the censoring time $C_i$, and $\delta_i = \mathbb{1}(T_i \leq C_i)$ the event/censoring indicator which is 1 if the event time is observed and 0 if it is censored.

Another aspect of the same problem is the updating, or conditioning, process: as the data are collected along time, more and more information is available. This is not only the observation of new events, for which the time values become known, but also for those subjects which are still at risk for the event of interest: the knowledge that an amount of time has passed and the event has not occurred yet is not a totally missing data, but a valuable information to take into account.

From a mathematical point of view, the updating process means that, instead of considering the probability density function $f(t)$ of the time variable $T$, it is worth focusing on its hazard function

$$h(t) = \lim_{\Delta t \searrow 0} \frac{\mathsf{P}(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}, \tag{3.1}$$

which is the instantaneous probability for the event to occur, conditional on not having occurred yet. $S(t) = 1 - F(t)$ is called survival function, since it is the probability of not having experienced the event of interest yet, i.e. of having survived until time $t$ in a time-to-death framework. According to the context, it can be useful to express the relation between hazard function and survival function as one of the following:

$$h(t) = -\frac{\mathrm{d}}{\mathrm{d}t} \log S(t), \tag{3.2}$$

$$S(t) = \exp\big\{ -\Lambda(t) \big\}, \tag{3.3}$$

where $\Lambda(t) = \int_0^t \lambda(s)\mathrm{d}s$ is the cumulative hazard function.

### 3.1.1 Proportional hazard models

Two main approaches exist to model the impact of a set of covariates $\mathbf{x} = (x_1, \ldots, x_p)^\top$ on the risk of the event of interest. On one hand , accelerated failure time models (Wei, 1992) are based on a linear relation with the logarithm of the event times. On the other hand, the proportional hazard model, the popularity of which is mainly due to Cox (1972), is undoubtedly the most known, used, widespread and studied regression model for survival data. This model relies on the assumption that the hazard, given explanatory variables, is proportional to a given baseline hazard function $\lambda_0(t)$:

$$\lambda(t; \mathbf{x}) = \lambda_0(t) \exp\big\{\boldsymbol{\beta}^\top \mathbf{x}\big\}, \tag{3.4}$$

where $\lambda_0(t)$ is the hazard in the case $\mathbf{x} = (0, \ldots, 0)^\top$.

The ratio between the hazard of two subjects is

$$\frac{h(t; \mathbf{x}_i)}{h(t; \mathbf{x}_{i'})} = \exp\big\{\boldsymbol{\beta}^\top (\mathbf{x}_i - \mathbf{x}_{i'})\big\} \tag{3.5}$$

which is constant over time; so the hazard functions of different subjects are proportional to each other. As a consequence, each coefficient $\beta_j$ is the log-hazard ratio for two subjects for which the $x_j$ is the only covariate which differs, and the difference is 1. Indeed, in this case, we have

$$\frac{h(t; \mathbf{x}_i)}{h(t; \mathbf{x}_{i'})} = \exp\big\{\beta_j \left(x_{ij} - (x_{ij} - 1)\right)\big\} = \exp\{\beta_j\}. \tag{3.6}$$

The main contribution given by Cox (1972) is the semiparametric estimation approach for the regression parameters. The baseline hazard function can be estimated non-parametrically and profiled out of the likelihood. Regression parameters $\boldsymbol{\beta}$ can then be estimated via maximisation of the partial likelihood

$$L_{\mathsf{P}}(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left( \frac{\exp\big\{\boldsymbol{\beta}^\top \mathbf{x}_i\big\}}{\sum_{i' \in R(y_i)} \exp\big\{\boldsymbol{\beta}^\top \mathbf{x}_{i'}\big\}} \right)^{\delta_i}, \tag{3.7}$$

where $R(y) = \{i' \mid y_{i'} \geq y\}$ is the risk set, i.e. the set of subjects still at risk, at time $y$.

## 3.2 Multi-state models

In survival analysis, the time of an event of interest is usually assumed to be either observed or censored in an uninformative way. Most often, censoring is due to different causes, some of which are independent of the event of interest, while some others are not. Furthermore, some events can occur before the one of interest, changing its risk. Finally, more than one event can be of interest and the occurrence of one of them can preclude or be necessary for the occurrence of others.

Such complex sequences of events, called event-histories, can be thought of as sequences of transitions between some states. Then, a graphical representation is possible in terms of states (nodes) and admissible transitions (arrows).

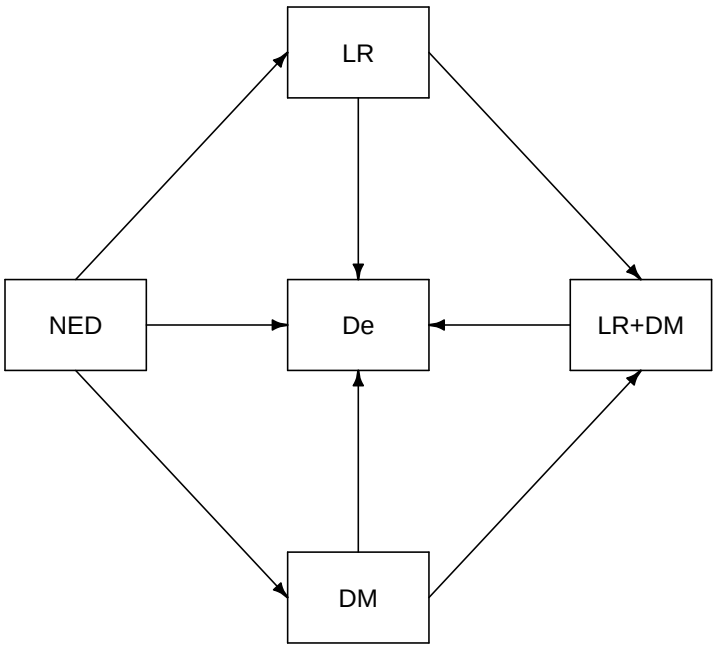



Figure 3.1 – Cancer model. NED: No Evidence of new Disease, De: Dead, LR: Local Relapse, DM: Distant Metastases. Transitions which represent analogous events are represented by arrows with the same line type.

Figure 3.1 shows an example, which is typical of cancer studies; the main endpoint is the time to death, but the local relapse (LR) and/or distant metastases (DM) can occur before. It is interesting to study the three risks jointly and the effect of the occurrence of intermediate adverse events on the risk of death.

Some typical state structures (Fig. 3.2) are reviewed by (Hougaard, 2000, Ch. 5): recurrent events, competing risks, alternating states, disability model, mortality model, bivariate model.

The problem of recurrent events (Cook and Lawless, 2002; Rondeau *et al.*, 2010) can mainly be dealt with frailty models (Sec. 3.3). Methods which are specific for competing risks have been developed, notably since the works by Gray (1988) and Fine and Gray (1999). A complete review

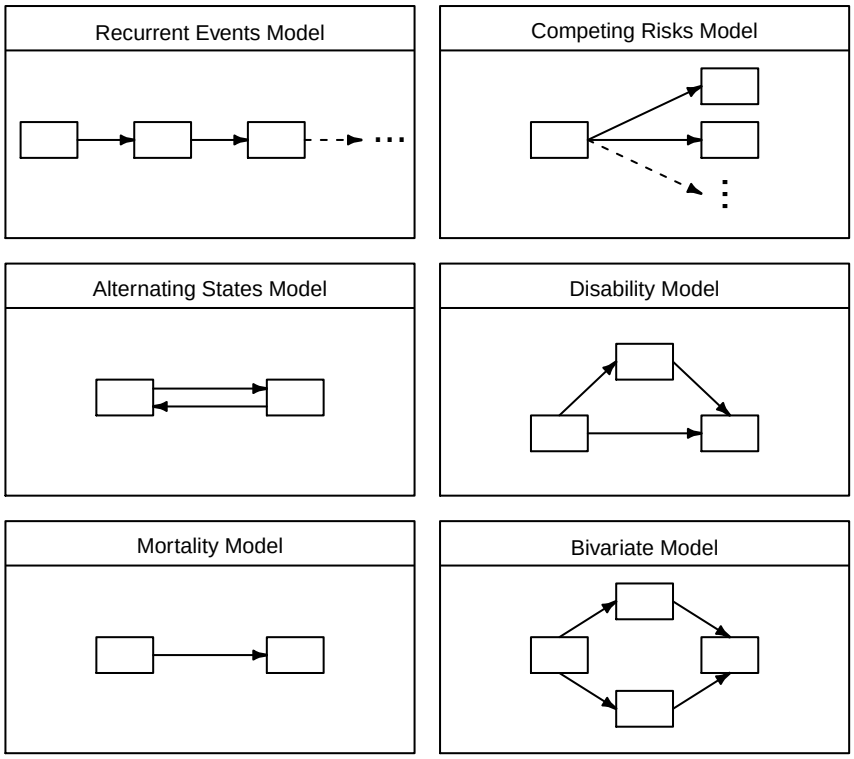



Figure 3.2 – Common multi-state structures (Hougaard, 2000, Ch. 5).

is the book by Pintilie (2006). The mortality model setting corresponds to the classical survival problem, which is largely debated in the literature.

Multi-state models represent the so called life-history (or event-history) approach to longitudinal data modelling (Andersen and Keiding, 2002), characterized by different types of possible events and event-related risk dependence.

In the context of multi-state models, the possible events are several and the risk is different for each one. The risk of entering the state $l$ at time $t$, given all the history $\mathcal{H}_t$ until time $t$ is the **transition-specific hazard**

$$\lambda_l(t \mid \mathcal{H}_{t^-}) = \lim_{\Delta t \searrow 0} \frac{\mathsf{P}\big[\mathcal{S}(t+\Delta t) = l \mid \mathcal{H}_{t^-}\big]}{\Delta t}, \tag{3.8}$$

with $\mathcal{S}(t)$ the state at time $t$ and $\mathcal{H}_{t^-} = \big\{\mathcal{S}(t'), t' \in [0,t)\big\}$.

For each subject $i = 1, \ldots, n$, the observed data are the last time at which the subject is observed, $C_i$, and a set of couples $(t_{i,j}, s_{i,j})$, $j = 1, \ldots, J_i$, concerning the $J_i$ observed transitions at times $t_{i,j}$ to states $s_{i,j}$. Then, each subject $i$ contributes to the likelihood by the hazard $\lambda_{s_{i,j}}(t_{i,j} \mid \mathcal{H}_{t_{i,j}^-})$ at transition times, and by the survival function $\exp\Big(-\int_{t_{i,j-1}}^{t_{i,j}} \lambda_{\bar{s}_{i,j}}(u \mid \mathcal{H}_{u^-})\mathrm{d}u\Big)$, with $\lambda_{\bar{s}_{i,j}}(t \mid \mathcal{H}_{t^-}) = \sum_{s \neq s_{i,j}} \lambda_s(t \mid \mathcal{H}_{t^-})$, for the sojourns periods. Hence, the likelihood function

is in general (Hougaard, 2000, Sec. 4)

$$L = \prod_{i=1}^{n} \left\{ \prod_{j=1}^{J_i} \left[ \lambda_{s_{i,j}}(t_{i,j} \mid \mathcal{H}_{t_{i,j}^-}) \exp\left( - \int_{t_{i,j-1}}^{t_{i,j}} \lambda_{\overline{s}_{i,j}}(u \mid \mathcal{H}_{u^-}) \mathrm{d}u \right) \right] \right.$$
$$\left. \exp\left( - \int_{t_{i,J_i}}^{C_i} \lambda_{\overline{s}_{i,j}}(u \mid \mathcal{H}_{u^-}) \mathrm{d}u \right) \right\}, \quad (3.9)$$

with $t_{i,0} = 0, \forall i,$. This likelihood is the product of the hazard functions at observed transition times and the survival functions on the sojourn periods, here included the one between the last observed transition and the last observation time $C_i$.

Often, the focus of interest in medical research is the biological understanding of the influence of risk factors and past history on the hazards. The clinical practice, on its side, needs forecasting tools based on probabilities of future events, given all the information available until the present moment. The transition probability at time $t$ into state $l$, given information available until time $u$, is defined as

$$P_l(u,t) = \mathsf{P}\big[\mathcal{S}(t) = l \mid \mathcal{H}_u\big]. \quad (3.10)$$

If $u \geq t$, then $P_l(u,t)$ is either 0 or 1 because the past is known. Thus, it only makes sense to compute transition probabilities for $u < t$.

A general analysis approach is not available for multi-state survival data; nevertheless, mild assumptions allow to hugely simplify the problem and to obtain very general and powerful analysis tools. In the following we concentrate on these particular cases, reviewed in details by Hougaard (2000, Ch. 5–6), notably on Markov models. First, we need some definitions.

**Definition 1** (**Parents and children of a state**)**.** For each state $s$, (i) the set of its parents is the set of states from which a direct transition into $s$ is possible and (ii) the set of its children is the set of states to which a direct transition from $s$ exists.

**Definition 2** (**Initial, transient and absorbing states**)**.** A state is called (i) initial if its parents set is empty, (ii) absorbing (or final) if its children set is empty, (iii) transient (or intermediate) if both its parents and children sets are non-empty. States which are both initial and absorbing are necessarily isolated, so of no interest.

**Progressive models.** A multi-state model is progressive if the size of the parents set of each state is at most 1. In other words only one possible incoming transition is possible for each state. In this case, the knowledge on the current state also implies the knowledge about the whole previous path, i.e. the ordered list of the states visited in the past. Among the examples in Figure 3.2, the recurrent events, the competing risks and the mortality models are progressive.

The interest in progressive multi-state models is that the transition probabilities can be expressed in terms of multiple integrals instead of differential equations. Indeed, the probability (3.10) for a

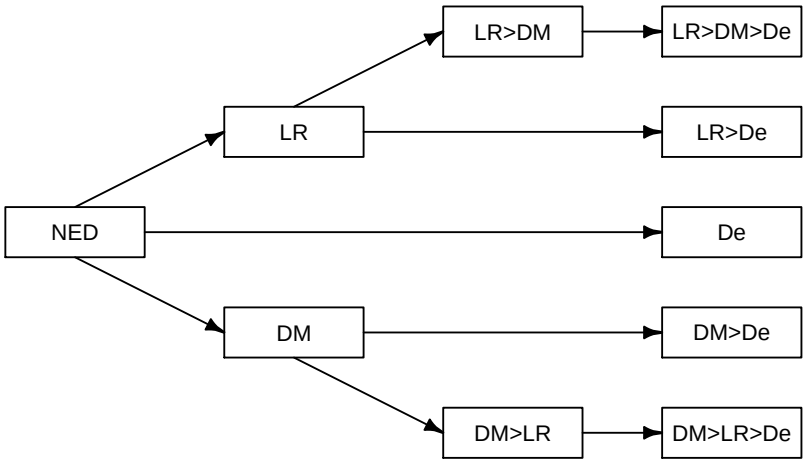



Figure 3.3 – Progressive form of cancer model (Fig. 3.1). NED: No Evidence of Disease, De: Dead, LR: Local Relapse, DM: Distant Metastes, LR>DM: Local Relapse then Distant Metastases , LR>De: Local Relapse then Dead, LR>DM>De: Local Relapse then Distant Metastases then Dead, DM>LR: Distant Metastases then Local Relapse, DM>De: Distant Metastases then Dead, DM>LR>De: Distant Metastases then Local Relapse then Dead.

progressive model is

$$P_l(u,t) = \int_u^t \int_{u_1}^t \cdots \int_{u_{k-1}}^t \prod_{j=1}^{k} \left[ \lambda_{l_j}(u_j \mid \mathcal{H}_{u_j^-}) \exp\left\{ -\int_{u_{j-1}}^{u_j} \lambda_{\bar{l}_j}(v \mid \mathcal{H}_{v^-}) \mathrm{d}v \right\} \right] \exp\left\{ -\int_{u_k}^{t} \lambda_{\bar{l}_j}(v \mid \mathcal{H}_{v^-}) \mathrm{d}v \right\} \mathrm{d}u_k \ldots \mathrm{d}u_1, \quad (3.11)$$

with $l_1, \ldots, l_{k-1}$ the states in the path between $\mathcal{S}(u)$ and $l$, $k-1$ their number, $u_0 = u$ and $l_k = l$ (Putter *et al.*, 2006, Sec. 4.5).

In many cases a non-progressive model can be transformed into a progressive one by splitting the states with more than one parent. For instance, the non-progressive cancer model in Figure 3.1 can be made progressive as shown in Figure 3.3. It is clear that, in order to obtain a progressive form, the number of states and of possible transitions rapidly grows. For complex multi-state structure this is unfeasible in practice, or not convenient at least.

**Markov, semi-Markov and Markov extension models.** According to the general formulation of the transition-specific hazard (3.8), the risk of leaving the present state depends on the past history $\mathcal{H}_{t^-}$. Under Markov assumption

$$\lambda_l(t \mid \mathcal{H}_{t^-}) = \lambda_l\left(t \mid \mathcal{S}(t^-)\right). \quad (3.12)$$

This means that, conditionally on the present state, the past has no influence on the risk. As a consequence, the hazards of transitions between each couple of states can be defined much simplier as

$$\lambda_{ml}(t) = \lambda_l \left( t \mid \mathcal{S}(t^-) = m \right). \tag{3.13}$$

Semi-Markov models are similar to Markov ones because the risk depends on past history through the present state and time. The difference is that semi-Markov models take into account, too, the time since last transition, that is since entry into the present state. This difference can also be linked to the choice of time scale: in the so called "clock-reset approach" (Putter *et al.*, 2007) or "gap time representation" (Duchateau and Janssen, 2008), the time is reset to 0 at each transition; on the other hand in the "clock-forward approach" or "calendar time representation" the time origin is only one and left censoring is used to separate the at-risk from the not-at-risk periods for each subject.

Markov-extension models relax the Markov and semi-Markov assumptions in several possible ways. The most common relaxation consists in allowing hazards to depend also on the time to previous events or on which states have been visited before. This information can be incorporated by means of time-dependent covariates.

Under the Markov assumption, the likelihood function (3.9) can be factorized into

$$L = \prod_m \prod_l \prod_{i=1}^{n} \prod_{j=1}^{J_i} \left[ \left(\lambda_{ml}(t_{i,j})\right)^{\delta_{ml}(i,j)} \exp\left( -\delta_m(i,j) \int_{t_{i,j-1}}^{t_{i,j}} \lambda_{ml}(u)\mathrm{d}u \right) \right], \tag{3.14}$$

with $\delta_{ml}(i,j)$ the indicator for the $j$-th transition of subject $i$ being from state $m$ to state $l$ and $\delta_m(i,j) = \sum_{l \neq m} \delta_{ml}(i,j)$ the indicator for the $j$-th transition of subject $i$ starting from state $m$. Analogous but more complex expressions are attainable for semi-Markov and Markov-extension models, with appropriate expressions of hazard functions.

The hazard of each transition can then be modelled separately if each one has different parameters. Constraints can be imposed in order to account for similarities of some events and to simplify the problem. In the example in Figure 3.1, for instance, proportionality of baseline hazards and common regression parameters can be assumed for analogous transitions like all those going into state De. In these cases, the transitions with common baseline and/or common regression parameters must be considered together in estimation procedures, as illustrated in details by de Wreede *et al.* (2011). In particular, common baseline hazards are treated as a unique transition type while dummy indicators are created to distinguish transitions and model their hazard ratios. Common regression parameters are modelled by a stratified model, with strata corresponding to transition types.

A counting process formulation of the likelihood quantities (Andersen *et al.*, 1993; de Wreede *et al.*, 2010) is sometimes convenient. In this representation, transitions are recorded as changes of the value of a stochastic process.

Let us denote by $q \in \{1, \ldots, Q\}$ all the ordered couples $(m, l)$ of states between which a direct transition is possible. Let then $\mathsf{N}_{qi}(t)$ be the counting process for transitions of type $q$ for subject $i \in \{1, \ldots, n\}$. Analogously, $\mathsf{Y}_{qi}(t)$ is defined as the at-risk stochastic process for transitions of type $q$ for subject $i$, being 1 if the subject is at risk of event $q$ at time $t^-$, 0 otherwise. Let further

$\mathsf{N}_q(t) = \sum_{i=1}^n \mathsf{N}_{qi}(t)$ and $\mathsf{Y}_q(t) = \sum_{i=1}^n \mathsf{Y}_{qi}(t)$ be the aggregated processes for transitions of type $q$ for the whole sample. Then, the likelihood function (3.14) can be expressed as

$$L = \prod_{q=1}^{Q} \prod_{i=1}^{n} \exp\left\{ \int_0^\infty \log \lambda_q(t) \mathrm{d}\mathsf{N}_{qi}(t) - \int_0^\infty \lambda_q(t) \mathsf{Y}_{qi}(t) \mathrm{d}t \right\} \tag{3.15}$$

(Kalbfleisch and Prentice, 2002, Eq. 8.30).

### 3.2.1 Inference

Summarising details by Andersen *et al.* (1993, Sec. VII.2) and de Wreede *et al.* (2010, Sec. 2.3.2), we consider a proportional hazards regression model for the transition-specific hazard

$$\lambda_{qi}(t; \mathbf{x}_i) = \lambda_{q0}(t) \exp\left\{ \boldsymbol{\beta}^\top \mathbf{x}_{qi}(t) \right\}, \tag{3.16}$$

with $q = 1, \ldots, Q$ the transition type, $i = 1, \ldots, n$ the subject, $\lambda_{q0}(t)$ the baseline hazard for transitions of type $q$, $\boldsymbol{\beta}$ the vector of the stacked transition-specific coefficient vectors $\boldsymbol{\beta}_q$, and $\mathbf{x}_{qi}(t)$ a vector of transition-specific covariates, possibly time-dependent, derived from $\mathbf{x}_i$.

The way $\mathbf{x}_{qi}(t)$ is obtained from $\mathbf{x}_i$ is linked to the data transformation from wide to long format (Putter, 2011; de Wreede *et al.*, 2011). This transformation allows, for instance, to include different covariates for different transitions or to leave the same covariate have different coefficients in different transitions; further, it is needed to include informations concerning past transitions. Finally, proportionality of hazards of different transitions can be specified by using the same baseline and by inserting a dummy variable into $\mathbf{x}_{qi}(t)$. A small explicative example is shown in Appendix B.1.

The full loglikelihood of the model (3.16) is (App. A.1)

$$\ell\big(\boldsymbol{\beta}, \boldsymbol{\lambda}_0(\cdot)\big) = \sum_{q=1}^{Q} \sum_{i=1}^{n} \left\{ \int_0^\infty \log \lambda_{qi}(t) \mathrm{d}\mathsf{N}_{qi}(t) - \int_0^\infty \lambda_{qi}(t) \mathsf{Y}_{qi}(t) \mathrm{d}t \right\}, \tag{3.17}$$

with $\boldsymbol{\lambda}_0(\cdot) = (\lambda_{10}(t), \ldots, \lambda_{Q0}(t))^\top$ the vector of the baseline hazard functions.

If a parametric form $\lambda_{q0}(\cdot; \boldsymbol{\xi}_q)$ is assumed for the baseline hazards $\boldsymbol{\lambda}_0(t)$, the full loglikelihood (3.17) can be maximised with respect to all the parameters $\left(\boldsymbol{\xi}^\top, \boldsymbol{\beta}^\top\right)^\top$.

In a semiparametric framework, instead, the regression parameters can be estimated like in the case of a Cox model stratified by transition type $q$. The partial loglikelihood for the model (3.16) is

$$\ell_{\mathsf{P}}(\boldsymbol{\beta}) = \sum_{q=1}^{Q} \sum_{i=1}^{n} \int_0^\infty \left[ \boldsymbol{\beta}^\top \mathbf{x}_{qi}(t) - \log \mathsf{Y}_q^{(0)}(\boldsymbol{\beta}, t) \right] \mathrm{d}\mathsf{N}_{qi}(t), \tag{3.18}$$

with

$$\mathsf{Y}_q^{(0)}(\boldsymbol{\beta}, t) = \sum_{i=1}^{n} \exp\left\{ \boldsymbol{\beta}^\top \mathbf{x}_{qi}(t) \right\} \mathsf{Y}_{qi}(t) \tag{3.19}$$

the weighted risk set.

The Fisher information matrix is

$$\mathcal{I}(\boldsymbol{\beta}) = \sum_{q=0}^{Q} \int_0^{\infty} \mathsf{V}_q(\boldsymbol{\beta}, t) \mathrm{d}\mathsf{N}_q(t), \tag{3.20}$$

with

$$\mathsf{V}_q(\boldsymbol{\beta}, t) = \frac{\mathsf{Y}_q^{(2)}(\boldsymbol{\beta}, t)}{\mathsf{Y}_q^{(0)}(\boldsymbol{\beta}, t)} - \mathsf{E}_q(\boldsymbol{\beta}, t)\mathsf{E}_q^{\top}(\boldsymbol{\beta}, t), \tag{3.21}$$

$$\mathsf{E}_q(\boldsymbol{\beta}, t) = \frac{\mathsf{Y}_q^{(1)}(\boldsymbol{\beta}, t)}{\mathsf{Y}_q^{(0)}(\boldsymbol{\beta}, t)}, \tag{3.22}$$

$$\mathsf{Y}_q^{(1)}(\boldsymbol{\beta}, t) = \sum_{i=1}^{n} \mathbf{x}_{qi}(t) \exp\left\{\boldsymbol{\beta}^{\top}\mathbf{x}_{qi}(t)\right\} \mathsf{Y}_{qi}(t), \tag{3.23}$$

$$\mathsf{Y}_q^{(2)}(\boldsymbol{\beta}, t) = \sum_{i=1}^{n} \mathbf{x}_{qi}(t)\mathbf{x}_{qi}^{\top}(t) \exp\left\{\boldsymbol{\beta}^{\top}\mathbf{x}_{qi}(t)\right\} \mathsf{Y}_{qi}(t). \tag{3.24}$$

**Nonparametric estimation of baseline hazards.** The Nelson–Aalen estimator (Nelson, 1969; Aalen, 1976) gives an estimate of the hazard function for independent and identically distributed observations, without accounting for covariates effect. The estimator of the transition-specific hazard is

$$\hat{\lambda}_q(t) = \frac{\mathrm{d}\mathsf{N}_q(t)}{\mathsf{Y}_q(t)}, \tag{3.25}$$

with $\mathrm{d}\mathsf{N}_q(t)$ the number of transitions of type $q$ at time $t$. Two (pointwise) variance estimators exist: the Aalen estimator

$$\hat{\mathsf{V}}(\hat{\lambda}_q(t)) = \frac{\mathrm{d}\mathsf{N}_q(t)}{\mathsf{Y}_q^2(t)} = \frac{\hat{\lambda}_q(t)}{\mathsf{Y}_q(t)} \tag{3.26}$$

and the Greenwood (1926) estimator

$$\hat{\mathsf{V}}(\hat{\lambda}_q(t)) = \frac{\mathrm{d}\mathsf{N}_q(t)(\mathsf{Y}_q(t) - \mathrm{d}\mathsf{N}_q(t))}{\mathsf{Y}_q^3(t)} = \frac{\hat{\lambda}_q(t)(1 - \hat{\lambda}_q(t))}{\mathsf{Y}_q(t)}. \tag{3.27}$$

In a regression context, the Nelson–Aalen estimator is replaced by a weighted version, the Breslow (1972) estimator for the baseline hazard functions

$$\hat{\lambda}_{q0}(t; \hat{\boldsymbol{\beta}}) = \frac{\mathrm{d}\mathsf{N}_q(t)}{\mathsf{Y}_q^{(0)}(\hat{\boldsymbol{\beta}}, t)}, \tag{3.28}$$

with $\mathsf{Y}_q^{(0)}(\hat{\boldsymbol{\beta}}, t)$ computed by plugging into (3.19) the regression parameters estimates $\hat{\boldsymbol{\beta}}$ obtained by semiparametric inference.

Once the estimates of both the regression parameters $\hat{\boldsymbol{\beta}}$ and the baseline hazards $\hat{\lambda}_{q0}(t; \hat{\boldsymbol{\beta}})$ are obtained, the cumulative hazards for a given patient with covariates $\mathbf{x}^*$ are estimated as

$$\hat{\Lambda}_q(t; \hat{\boldsymbol{\beta}}, \mathbf{x}^*) = \exp(\hat{\boldsymbol{\beta}}^{\top}\mathbf{x}_q^*) \int_0^t \hat{\lambda}_{q0}(u; \hat{\boldsymbol{\beta}}) \mathrm{d}\mathsf{N}_q(u), \qquad q = 1, \ldots, Q, \tag{3.29}$$

with $\mathbf{x}_q^*$ the expanded covariates for transition $q$. They can be collected into the square matrix $\hat{\Lambda}(t;\hat{\boldsymbol{\beta}};\mathbf{x})$, with lines corresponding to origin states and columns to arrival states. Diagonal elements are minus the sum of the other elements in the line.

### 3.2.2 Transition probabilities

Under Markov assumption, the transition probabilities (3.10) are

$$P_{ml}(u,t) = \mathsf{P}\big[\mathcal{S}(t) = l \mid S(u) = m\big] \tag{3.30}$$

and they are the solution of a set of differential equations (Andersen *et al.*, 1993, p. 93)

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{P}(u,t) = \boldsymbol{G}(t)^\top \boldsymbol{P}(u,t), \tag{3.31}$$

with $\boldsymbol{P}(u,t) = \{P_{ml}(u,t)\}_{m,l}$ and $\boldsymbol{G}(t)$ a matrix with off diagonal elements $G_{ml}(t) = \lambda_{ml}(t)$ and diagonal elements $G_{mm}(t) = -\sum_{l\neq m}\lambda_{ml}(t)$.

Even though the equations

$$\frac{\mathrm{d}}{\mathrm{d}t}P_{ml}(u,t) = \sum_k P_{mk}(u,t)\lambda_{kl}(t) \tag{3.32}$$

in (3.31) are not solvable due to the non-constancy over time of the matrix $G(t)$, the transition probabilities are directly attainable, under Markov assumption, by the hazards as product integral (Gill, 2001)

$$\boldsymbol{P}(u,t) = \boldsymbol{\Pi}_u^t\big(\boldsymbol{I} + \mathrm{d}\boldsymbol{\Lambda}(v)\big), \tag{3.33}$$

with $\boldsymbol{I}$ the identity matrix and $\boldsymbol{\Lambda}(v)$ the matrix with off-diagonal elements $\Lambda_{ml}(v) = \int_0^v \lambda_{ml}(s)\mathrm{d}s$ and diagonal elements $\Lambda_{mm}(v) = -\sum_{l\neq m}\Lambda_{ml}(v)$ (Andersen *et al.*, 1993). The corresponding Aalen-Johansen-type estimator based on the estimated cumulative hazards matrix is

$$\hat{\boldsymbol{P}}(s,t;\mathbf{x}) = \boldsymbol{\Pi}_s^t\left(\boldsymbol{I} + \mathrm{d}\hat{\boldsymbol{\Lambda}}(u;\hat{\boldsymbol{\beta}},\mathbf{x})\right). \tag{3.34}$$

The transition probabilities can be used for two types of prediction: *forward* and *fixed horizon* (Putter *et al.*, 2006, Sec. 4.5; de Wreede *et al.*, 2010, Sec. 2.3.3). In the former case, the time of prediction $s$ is fixed and the time for prediction $t$ varies, i.e. at a given time $s$ the probabilities of possible future events are evaluated for varying time horizons $t$. In the latter case, once the time horizon $t$ is fixed, prediction is made from several timepoints $s$, i.e. the interest is in making prediction for a chosen time $t$, assuming that something has or has not happened at different moments $s$.

In both cases, pointwise confidence intervals can be built by means of Aalen-type and Greenwood-type estimators of the variance-covariance matrix of $\hat{\boldsymbol{P}}(s,t;\mathbf{x})$. de Wreede *et al.* (2010) give both the direct and the recursive expressions for these estimators, in Equations 17–22. The authors also provide a bootstrap method for the computation of cumulative hazards and transition probabilities for semi-Markov and Markov-extension models.

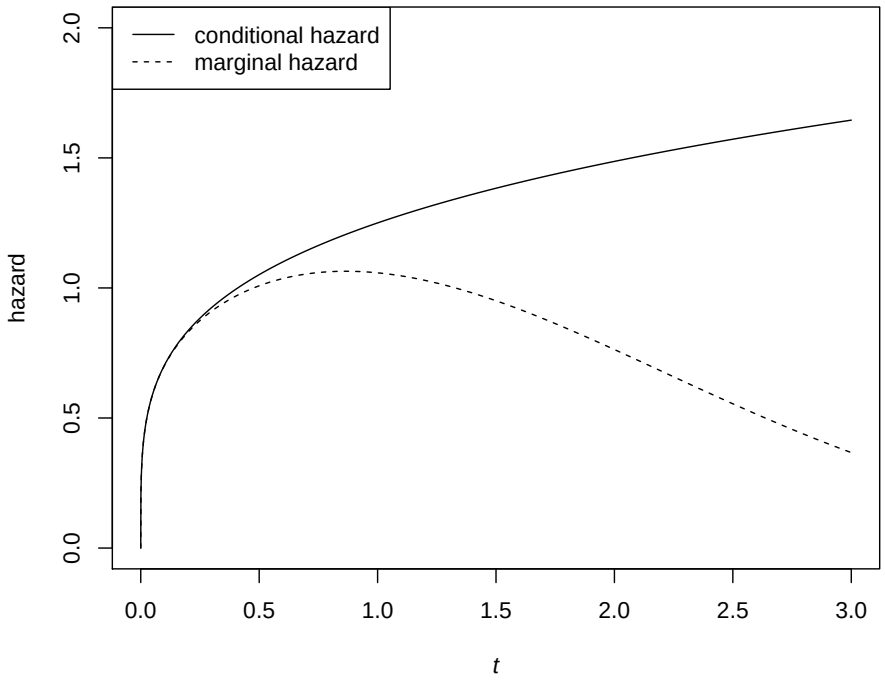



Figure 3.4 – Selection effect. The solid line is the conditional hazard, given the unobserved risk factors, then without the effect of unobserved heterogeneity. The dashed line is the marginal hazard, affected by the selection at late times.

## 3.3 Frailty models

Inference for proportional hazard models is usually based on the assumption of independent and identically distributed samples. If it does not hold, the estimation of regression coefficients is not consistent in general (Andersen *et al.*, 1999, Duchateau and Janssen, 2008, Sec. 3.4.2; Hougaard, 2000, Sec. 13.2). There exist, indeed, many situations in which it is more sensible to expect that subjects present different levels of risk, even conditionally on observed covariates. This is mainly due to the fact that not all the relevant risk factors are measured. Unobserved risk factors can be subject-specific or common to groups like families, communities, geographical areas and so on. The most important consequence of heterogeneous baseline hazards is the selection effect: on average the highest-risk subjects fail earlier, thus the hazard at late times is estimated on the basis of the lowest-risk ones (Vaupel and Yashin, 1985b). Figure 3.4 shows an example of the selection effect which results in a spurious reduction of the marginal risk at late times, if unobserved risk factors are not taken into account.

Moving from ideas similar to those of generalised linear mixed models (GLMM, McCulloch *et al.*, 2008), frailty models extend proportional hazard models by means of random effects to deal with differences in baselines. In a GLMM perspective, an unobserved random effect $w$ can be added on the log-hazard scale as

$$\log\left\{\lambda(t;\mathbf{x} \mid w)\right\} = \log\left\{\lambda_0(t)\right\} + w + \boldsymbol{\beta}^\top \mathbf{x}, \tag{3.35}$$

with $\lambda_0(t)$ the baseline hazard function, and $\mathbf{x}$ the vector of covariates. Usually, the frailty model

is rather expressed on the hazard scale, so that (3.35) is

$$\lambda(t; \mathbf{x} \mid u) = \lambda_0(t) u \exp\left\{\boldsymbol{\beta}^\top \mathbf{x}\right\}, \tag{3.36}$$

with $u = e^w$. Vaupel *et al.* (1979) first named it frailty term because it is a multiplicative factor which increases or decreases the risk level at all times, describing the different predisposition to the event of interest. Its distribution $f_U(u)$ determines many features of the unobserved heterogeneity effect. In Section 3.3.4 we present many possibilities and we discuss their main characteristics.

Note that, despite this is an extension of the proportional hazard model, hazards are not marginally proportional, at least in general. According to model (3.36), the ratio between the hazards of two subjects is

$$\frac{\lambda_i(t \mid u_i)}{\lambda_j(t \mid u_j)} = \frac{u_i \exp\left\{\boldsymbol{\beta}^\top \mathbf{x}_i\right\}}{u_j \exp\left\{\boldsymbol{\beta}^\top \mathbf{x}_j\right\}} \neq \frac{\exp\left\{\boldsymbol{\beta}^\top \mathbf{x}_i\right\}}{\exp\left\{\boldsymbol{\beta}^\top \mathbf{x}_j\right\}}, \tag{3.37}$$

except in the case of $u_i = u_j$. Therefore, the hazards are proportional conditionally on the frailty terms, but not in general. Then, the regression coefficients $\boldsymbol{\beta}$ can no longer be directly interpreted in terms of hazard ratios; their values represent the log-hazard ratios only conditionally on the frailty, i.e. between two subjects sharing frailty term $u$.

**Univariate frailty models.** It may happen that important risk factors are not included in the set of measured covariates $\mathbf{x}$. Thus, the baseline hazard varies across subjects, so time variables of different patients are not identically distributed, even conditionally on observed covariates. This causes the *overdispersion* phenomenon, that is an increased variability of the parameter estimates, caused by the uncertainty due to missing information in addition to variability of data.

The univariate frailty model, treated in details by Wienke (2010, Ch. 3), embodies the effect of all the relevant unobserved risk factors in a subject-specific frailty. The conditional hazard function (3.36) is then

$$\lambda_i(t \mid u_i) = \lambda_0(t) u_i \exp\left\{\boldsymbol{\beta}^\top \mathbf{x}_i\right\}, \tag{3.38}$$

with $\lambda_i(t \mid u_i)$ the hazard function of the $i$-th subject, $\lambda_0(t)$ the baseline hazard, $\mathbf{x}_i$ the vector of covariates and $u_i$ its frailty.

In the context of univariate frailty models, the regression coefficients still represent log-hazard ratios conditional on the frailty value, but the fact that each subject has a different frailty term does not allow any direct one-to-one comparison between subjects.

**Shared frailty models.** The other important cause of violation of the independence assumption is the so-called *clustering*: the presence of groups of subjects which have common unobserved risk factors. A common example is given by multicenter clinical trials: patients in the same hospital have arguably more similar risk levels than patients in different centres (Duchateau *et al.*, 2002; Ha *et al.*, 2011). Other examples of clustered data are repeated measures on the same patient or recurrent events (Duchateau *et al.*, 2003; Rondeau, 2010), paired organs from the same organism (Hougaard, 1995; Xue and Ding, 1999), patients who are relatives, etc.

Because of clustering, the baseline risks of different clusters are different, which induces dependence between their survival times. Common solutions to this problem (O'Quigley and Stare,

2002; Glidden and Vittinghoff, 2004) are the stratified model and the fixed effects model. By means of stratification, the baseline risk of each group is modelled separately, either parametrically or not. This approach is very flexible but not all information is used and interpretation is restricted to participating centers. This yields very inaccurate estimates in case of small size groups. The fixed effects model, on the contrary, requires the assumption of proportionality of baselines of different clusters but is easily readable in terms of hazard ratios. Fixed effects models have the big disadvantage of requiring as many parameters as the number of groups minus one.

Shared frailty models account for different risk levels of clusters by means of group-specific random effects. Only few parameters — typically only one — are needed and estimation is based on all the observations. Most interesting, the estimate of the variability of the random effect has a strong interpretation in terms of heterogeneity: it provides a measure of how much groups are different.

The name shared frailty models is due the fact that subjects in the same cluster share the same frailty factor. These are the most used and studied frailty models in literature (Hougaard, 2000; Duchateau and Janssen, 2008).

The conditional hazard (3.36) for the shared frailty model is

$$\lambda_{hi}(t \mid u_h) = \lambda_0(t) u_h \exp\left\{\boldsymbol{\beta}^\top \mathbf{x}_{hi}\right\} \tag{3.39}$$

for subject $i = 1, \ldots, n_h$ in group $h = 1, \ldots, H$, with $\lambda_0(t)$ the baseline hazard function, $\mathbf{x}_{hi}$ the vector of covariates and $u_h$ the frailty of the cluster $h$.

In shared frailty models, the time variables $T_{hi}$ are independent conditionally on the frailty values, that is for fixed realisations of the $U_h$'s random variables. Then, the regression coefficients $\boldsymbol{\beta}$ are the log-hazard ratios between two patients belonging to the same group.

Finally, note that in the case that all groups are of unit size, the shared frailty model (3.39) reduces to a univariate frailty model (3.38). Therefore, from a technical point of view, the univariate model is only a particular case of the shared frailty model. Nevertheless the interpretation is quite different in terms of unobserved information: the frailty variance is interpreted as the strength of dependence in case of shared information, whereas it is an index of overdispersion in the univariate case.

**Correlated frailty models.** A more general class of models is that of correlated frailty models (Wienke, 2010, Ch. 5; Duchateau and Janssen, 2008, Sec.'s 6.2 and 7.2), that account for correlation between observations in the same cluster through distinct but dependent frailty terms. This makes the dependence more flexible and allows to consider more complex dependence structures. Each subject has his own frailty $U_{hi}$ and independence across clusters is still assumed

$$U_{hi} \perp\!\!\!\perp U_{h'i'} \qquad \Leftarrow \qquad h \neq h', \tag{3.40}$$

while the dependence within the vector of frailties within each group $\mathbf{U}_h = (U_{h1}, \ldots, U_{hn_h})$ is modelled by assuming a joint distribution according to the nature of the phenomenon. These models are more suited for accounting for both common and individual risks, but no general estimation method exists for any choice of the joint frailties distribution. Ad hoc solutions exist for specific situations (Wienke, 2010, Ch. 5).

Shared frailty models are a particular case of correlated frailty models, with frailties of grouped subjects which are totally dependent:

$$U_{hi} = U_{hi'}, \qquad \forall i, i' \in \{1, \ldots, n_h\}. \tag{3.41}$$

On their turn, correlated frailty models are a particular case of the multivariate frailty models (Hougaard, 2000, Ch. 10), accounting in a more general way for correlation between grouped observations.

**Nested frailty models.** A very interesting way of obtaining correlated frailties is by means of nested frailties (Wienke, 2010, Sec. 7.3): the advantages include both straightforward interpretability and analytical convenience.

The underlying idea is that the frailties of subjects in the same group depend on unobserved factors, some of which are shared and some are subject-specific. Consequently, as the level of risk is the combination of a group and a subject factor, the time variables are distinct but positively dependent.

First publications on this subject were oriented towards additive frailties (Vaupel and Yashin, 1985a; Parner, 1998; Petersen, 1998; Hougaard, 2000, Sec. 10.5) but, since Sastry's paper in 1997, the multiplicative approach has become largely predominant (Yau, 2001; Rondeau *et al.*, 2006; Horny, 2009; Shih and Lu, 2009) as it allows factorisation, useful in marginalising the conditional likelihood.

If we consider subjects clustered in groups and subgroups, then the conditional hazard (3.36) becomes

$$\lambda(t \mid v_h, w_{hj}) = \lambda_0(t) v_h w_{hj} \exp\left\{\boldsymbol{\beta}^\top \mathbf{x}_{hji}\right\}, \tag{3.42}$$

for subject $i = 1, \ldots, n_{hj}$ in subgroup $j = 1, \ldots, J_h$ of group $h = 1, \ldots, H$, with $\mathbf{x}_{hi}$ the vector of covariates, $v_h$ the group frailty and $w_{hj}$ the subgroup frailty.

Let $U_{hj} = V_h W_{hj}$ be the (multiplicative) group-subgroup frailty and assume that the frailties $V_1, \ldots, V_H, W_{11}, \ldots, W_{HJ_H}$ are mutually independent. Then, we have independence across clusters,

$$U_{hj} \perp\!\!\!\perp U_{h'j'} \qquad \Leftarrow \qquad h \neq h', \tag{3.43}$$

and positive dependence across subclusters,

$$\begin{aligned} \mathsf{Cov}(U_{hj}, U_{hj'}) &= \mathsf{E}(V_h^2 W_{hj} W_{hj'}) - \mathsf{E}(V_h W_{hj}) \mathsf{E}(V_h W_{hj'}) \\ &= \mathsf{V}(V_h) \mathsf{E}(W_{hj}) \mathsf{E}(W_{hj'}) > 0. \end{aligned} \tag{3.44}$$

### 3.3.1 Notation

**Definition 3 (Conditional hazard).** Consider a generic frailty model. As mentioned above (Eq. 3.36), we define the conditional hazard for a subject with covariate vector $\mathbf{x}$ as

$$\lambda(t \mid u) = \lambda_0(t) u \exp\left\{\boldsymbol{\beta}^\top \mathbf{x}\right\}. \tag{3.45}$$

It is called conditional hazard as it is conditional on the value $u$ of the frailty term $U$.

**Definition 4** (**Conditional survival function**)**.** The survival function associated to the hazard (3.45) is

$$S(t \mid u) = \exp\big\{ - \Lambda(t \mid u)\big\}, \tag{3.46}$$

with

$$\begin{aligned}\Lambda(t \mid u) &= \int_0^t \lambda(s \mid u)\mathrm{d}s \\ &= u\Lambda_0(t)\exp\left\{\boldsymbol{\beta}^\top \mathbf{x}\right\}\end{aligned} \tag{3.47}$$

the conditional cumulative hazard and where $\Lambda_0(t) = \int_0^t \lambda_0(s)\mathrm{d}s$ is the cumulative baseline hazard.

**Definition 5** (**Population survival function**)**.** The marginal survival function, called population survival function, is

$$\begin{aligned}S(t) &= \int_0^\infty S(t \mid u) f_U(u)\mathrm{d}u \\ &= \mathcal{L}\Big(\Lambda_0(t)\exp\left\{\boldsymbol{\beta}^\top \mathbf{x}\right\}\Big),\end{aligned} \tag{3.48}$$

with $f_U(u)$ the frailty distribution and $\mathcal{L}(s) = \mathsf{E}\Big[\exp\big\{ - Us\big\}\Big]$ its Laplace transform. The population survival function is marginalised with respect to the frailty distribution, so it is somehow representative of the entire population, given covariates values $\mathbf{x}$. In addition, the frailty values are unknown and unobservable, so Equation 3.48 is particularly useful for estimation (see Sec. 3.3.3).

**Definition 6** (**Joint conditional survival function**)**.** Thanks to conditional independence, the joint conditional survival function of all the subjects within a cluster $h$ is the product of the subject-specific ones:

$$\begin{aligned}S_h(\mathbf{t} \mid u) &= \prod_{i=1}^{n_h} S_{hi}(t_i \mid u) \\ &= \exp\left\{-u\sum_{i=1}^{n_h}\Lambda_0(t_i)\exp\left\{\boldsymbol{\beta}^\top \mathbf{x}_{hi}\right\}\right\},\end{aligned} \tag{3.49}$$

with $\mathbf{t} = (t_1, \ldots, t_{n_h})^\top$ and $n_h$ the dimension of cluster $h$.

**Definition 7** (**Joint survival function**)**.** As the frailties are not observable, the distribution marginalised with respect to the frailty distribution is used for inference procedures: the joint survival function of subjects in cluster $h$ is

$$\begin{aligned}S_h(\mathbf{t}) &= \int_0^\infty S_h(\mathbf{t} \mid u) f_U(u)\mathrm{d}u \\ &= \mathcal{L}\left(\sum_{i=1}^{n_h}\Lambda_0(t_i)\exp\left\{\boldsymbol{\beta}^\top \mathbf{x}_{hi}\right\}\right).\end{aligned} \tag{3.50}$$

**Definition 8** (**Joint density**)**.** It is easy to derive from Equation 3.50 the joint density for cluster

$h$

$$f_h(\mathbf{t}) = (-1)^{n_h} \left( \prod_{i=1}^{n_h} \lambda_0(t_i) \exp\left\{ \boldsymbol{\beta}^\top \mathbf{x}_{hi} \right\} \right) \mathcal{L}^{(n_h)} \left( \sum_{i=1}^{n_h} \Lambda_0(t_i) \exp\left\{ \boldsymbol{\beta}^\top \mathbf{x}_{hi} \right\} \right), \tag{3.51}$$

with

$$\mathcal{L}^{(k)}(s) = \frac{\mathrm{d}^k}{\mathrm{d}s^k} \mathcal{L}(s) = (-1)^k \mathsf{E}\Big[ U^k \exp(-Us) \Big]. \tag{3.52}$$

All the quantities defined in this section are summarised in Table 3.1, together with their probability interpretation.

### 3.3.2 Likelihood

**Shared frailties.** In the shared frailty model (3.39) the frailties $u_h$ are unobservable realisations of random variables which are assumed to be independent and identically distributed, with a common frailty distribution:

$$U_h \overset{iid}{\sim} f_U(u; \theta), \qquad h = 1, \ldots, H. \tag{3.53}$$

Many different distributions have been proposed in the literature (Hougaard, 2000, Ch. 7; Duchateau and Janssen, 2008, Ch. 4; Wienke, 2010, Ch.'s 3–4) with different consequences on the dependence structure within clusters (cf. Sec. 3.3.4). Due to identifiability issues, the mean of the frailty distribution is usually fixed to 1, so that the baseline hazard $\lambda_0(t)$ represents the risk level with average frailty. Thus, in most cases, the remaining parameter is the frailty variance, which is a measure of the heterogeneity between clusters and, at the same time, of dependence within clusters.

Thanks to assumptions (3.53), the full likelihood can be obtained as a product over clusters:

$$L(\boldsymbol{\zeta}) = \prod_{h=1}^{H} L_{\mathsf{C},h}(\boldsymbol{\xi}, \boldsymbol{\beta} \mid u_h) f_U(u_h; \theta) \tag{3.54}$$

with $\boldsymbol{\zeta} = \left( \boldsymbol{\xi}^\top, \theta, \boldsymbol{\beta}^\top \right)^\top$ and where $\boldsymbol{\xi}$ is either the baseline hazard parameters or the baseline hazard function itself, and $\theta$ the frailty distribution parameter. $L_{\mathsf{C},h}(\boldsymbol{\xi}, \boldsymbol{\beta} \mid u_h)$ is the contribution of cluster $h$ to the conditional likelihood, given the frailty value,

$$\begin{aligned} L_{\mathsf{C}}(\boldsymbol{\xi}, \boldsymbol{\beta} \mid \mathbf{u}) &= \prod_{h=1}^{H} L_{\mathsf{C},h}(\boldsymbol{\xi}, \boldsymbol{\beta} \mid u_h) \\ &= \prod_{h=1}^{H} \prod_{i=1}^{n_h} \Big( \lambda_{hi}(y_{hi} \mid u_h) \Big)^{\delta_{hi}} S_{hi}(y_{hi} \mid u_h), \end{aligned} \tag{3.55}$$

with $\mathbf{u} = (u_1, \ldots, u_H)^\top$ the vector of the frailties.

As the frailties are not observable, the likelihood (3.54) is not evaluable. Therefore, in some

| (**Eq.**) | **Name** | **Symbol** | **Probabilistic expression** |
|---|---|---|---|
| (3.45) | **Conditional hazard** | $\lambda(t \mid u)$ | $\lim_{\Delta t \searrow 0} \mathsf{P}\left\{T \in [t, t+\Delta t) \mid T \geq t, U = u\right\} / \Delta t$ |
| (3.46) | **Conditional survival function** | $S(t \mid u)$ | $\mathsf{P}\left\{T \geq t \mid U = u\right\}$ |
| (3.48) | **Population survival function** | $S(t)$ | $\mathsf{P}\left\{T \geq t\right\}$ |
| (3.49) | **Joint conditional survival function** | $S_h(\mathbf{t} \mid u)$ | $\mathsf{P}\left\{T_{h1} \geq t_1, \ldots, T_{hn_h} \geq t_{n_h} \mid U_h = u\right\}$ |
| (3.50) | **Joint survival function** | $S_h(\mathbf{t})$ | $\mathsf{P}\left\{T_{h1} \geq t_1, \ldots, T_{hn_h} \geq t_{n_h}\right\}$ |
| (3.51) | **Joint density** | $f_h(\mathbf{t})$ | $\lim_{\Delta t \searrow 0} \mathsf{P}\left\{\cap_{i=1}^{n_h} T_{hi} \in [t_i, t_i + \Delta t)\right\} / \Delta t$ |

Table 3.1 – Probabilistic expressions of quantities defined in Section 3.3.1.

cases it is marginalised with respect to $\mathbf{U}$: the so-obtained marginal likelihood is

$$\begin{aligned} L_{\mathsf{M}}(\boldsymbol{\zeta}) &= \prod_{h=1}^{H} L_{\mathsf{M},h}(\boldsymbol{\zeta}) \\ &= \prod_{h=1}^{H} \left\{ \int_0^{\infty} L_{\mathsf{C},h}(\boldsymbol{\xi}, \boldsymbol{\beta} \mid u_h) f_U(u_h; \theta) \mathrm{d}u_h \right\} \\ &= \prod_{h=1}^{H} \left\{ \left[ \prod_{i=1}^{n_h} \left( \lambda_0(y_{hi}) \exp\left\{ \boldsymbol{\beta}^\top \mathbf{x}_{hi} \right\} \right)^{\delta_{hi}} \right] \right. \\ &\qquad \left. (-1)^{d_h} \mathcal{L}^{(d_h)} \left( \sum_{i=1}^{n_h} \Lambda_0(y_{hi}) \exp\left\{ \boldsymbol{\beta}^\top \mathbf{x}_{hi} \right\} \right) \right\}, \end{aligned} \tag{3.56}$$

with $d_h = \sum_{i=1}^{n_h} \delta_{hi}$ the number of events in cluster $h$.

In other cases, the baseline hazard is estimated non parametrically and profiled out of the conditional likelihood (3.55). This profiled likelihood is called partial likelihood and is given by

$$L_{\mathsf{P}}(\boldsymbol{\beta} \mid \mathbf{u}) = \prod_{h=1}^{H} \prod_{i=1}^{n_h} \left\{ \frac{u_h \exp\left\{ \boldsymbol{\beta}^\top \mathbf{x}_{hi} \right\}}{\sum_{(h',i') \in R(y_{hi})} u_{h'} \exp\{ \boldsymbol{\beta}^\top \mathbf{x}_{h'i'} \}} \right\}^{\delta_{hi}}, \tag{3.57}$$

with $R(y) = \left\{ (h', i') \mid y_{h'i'} \geq y \right\}$ the risk set at time $y$.

Of course, the likelihoods (3.54)–(3.57) for the shared frailty model (3.39) are still valid for the univariate frailty model (3.38) whenever $n_h = 1$, $\forall h = 1, \ldots, H$.

**Nested frailties.** The nested frailty model (3.42) is more general and likelihood expressions get more complex. Usual distributive assumptions on the frailties are

$$\begin{array}{rcll} V_h & \overset{iid}{\sim} & f_V(v; \theta_V), & h = 1, \ldots, H, \\ W_{hj} & \overset{iid}{\sim} & f_W(w; \theta_W), & h = 1, \ldots, H, \quad j = 1, \ldots, J_h, \\ V_h & \perp\!\!\!\perp & W_{hj}, & \forall (h, j). \end{array} \tag{3.58}$$

In the common case that the distributions $f_v(v)$ and $f_W(w)$ have unit mean and variances $\theta_V$ and $\theta_W$ respectively, then the correlated frailties $U_{hj} = V_h W_{hj}$ have unit mean and variance

$$\mathsf{V}(U_{hj}) = (\theta_V + 1)(\theta_W + 1) - 1 \tag{3.59}$$

(App. A.2). Figure 3.5 shows the contour plot for the variance on frailties $U_{hj}$ as a function of $\theta_V$ and $\theta_W$.

Similarly to shared frailty models, the full likelihood for model (3.42) is

$$L(\boldsymbol{\zeta}) = \prod_{h=1}^{H} \left\{ f_V(v_h; \theta_V) \prod_{j=1}^{J_h} L_{\mathsf{C},hj}(\boldsymbol{\xi}, \boldsymbol{\beta} \mid v_h, w_{hj}) f_W(w_{qh}; \theta_W) \right\} \tag{3.60}$$

with $\boldsymbol{\zeta} = \left( \boldsymbol{\xi}^\top, \boldsymbol{\theta}^\top, \boldsymbol{\beta}^\top \right)^\top$, $\boldsymbol{\theta} = (\theta_V, \theta_W)^\top$.

$L_{\mathsf{C},hj}(\boldsymbol{\xi}, \boldsymbol{\beta} \mid v_h, w_{hj})$ is the contribution of subgroup $j$ in group $h$ to the conditional likelihood,

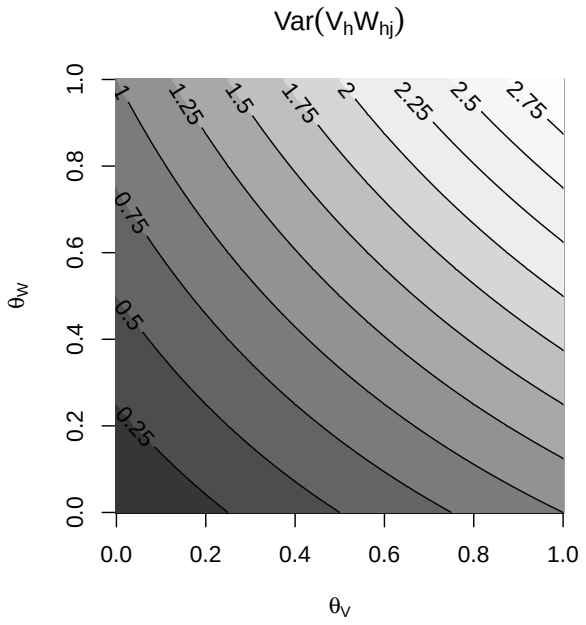



Figure 3.5 – Variance of nested frailties as function of the frailty variances (Eq. 3.59).

given the frailty values,

$$\begin{aligned} L_{\mathsf{C}}(\boldsymbol{\xi}, \boldsymbol{\beta} \mid \mathbf{v}, \mathbf{w}) &= \prod_{h=1}^{H} \prod_{j=1}^{J_h} L_{\mathsf{C},hj}(\boldsymbol{\xi}, \boldsymbol{\beta} \mid v_h, w_{hj}) \\ &= \prod_{h=1}^{H} \prod_{j=1}^{J_h} \prod_{i=1}^{n_{hj}} \Big(\lambda_{hji}(y_{hji} \mid u_h, w_{hj})\Big)^{\delta_{hji}} S_{hji}(y_{hji} \mid u_h, w_{hj}), \end{aligned} \tag{3.61}$$

with $\mathbf{v} = (v_1, \ldots, v_H)$ and $\mathbf{w} = (w_{11}, \ldots, w_{HJ_H})$.

Once the frailties $\mathbf{V}$ and $\mathbf{W}$ have been integrated out, the marginal likelihood is

$$\begin{aligned} L_{\mathsf{M}}(\boldsymbol{\zeta}) &= \prod_{h=1}^{H} L_{\mathsf{M},h}(\boldsymbol{\xi}, \boldsymbol{\beta}) \\ &= \prod_{h=1}^{H} \Bigg\{ \left[ \prod_{j=1}^{J_h} \prod_{i=1}^{n_h} \Big(\lambda_0(y_{hi}) \exp\left\{\boldsymbol{\beta}^\top \mathbf{x}_{hi}\right\}\Big)^{\delta_{hi}} \right] \\ &\qquad \int_0^{\infty} v_h^{d_h} (-1)^{d_h} \prod_{j=1}^{J_h} \mathcal{L}_W^{(d_{hj})} \left( \sum_{i=1}^{n_{hj}} \Lambda_0(y_{hji}) \exp\left\{\boldsymbol{\beta}^\top \mathbf{x}_{hji}\right\} \right) f_V(v_h; \theta_V) \mathrm{d}v_h \Bigg\}, \end{aligned} \tag{3.62}$$

with $d_{hj} = \sum_{i=1}^{n_{hj}} \delta_{hji}$, $d_h = \sum_{j=1}^{J_h} d_{hj} = \sum_{j=1}^{J_h} \sum_{i=1}^{n_{hj}} \delta_{hji}$, and $\mathcal{L}_W(\cdot)^{(k)}$ the $k$-th derivative of the Laplace transform of the distribution of $W_{hj}$.

The partial likelihood, obtained by profiling the baseline hazard out of the conditional likelihood

| **Distribution** | **Hazard function** $\lambda_0(t;\boldsymbol{\xi})$ | **Parameters** $\boldsymbol{\xi}$ |
|---|---|---|
| **exponential** | $\lambda$ | $\lambda > 0$ |
| **Weibull** | $\lambda\rho t^{\rho-1}$ | $\lambda, \rho > 0$ |
| **Gompertz** | $\lambda e^{\gamma t}$ | $\gamma, \lambda > 0$ |
| **lognormal** | $\dfrac{\exp\left\{-\frac{1}{2\gamma}(\log t-\mu)^2\right\}}{\left(t\sqrt{2\pi\gamma}\right)\left(1-\phi\left(\frac{\log t-\mu}{\sqrt{\gamma}}\right)\right)}$ | $\mu \in \mathbb{R}, \gamma > 0$ |
| **loglogistic** | $\dfrac{e^{\alpha}\kappa t^{\kappa-1}}{1+e^{\alpha}t^{\kappa}}$ | $\alpha \in \mathbb{R}, \kappa > 0$ |

Table 3.2 – Parametric models for the baseline hazard function.

(3.61), is

$$L_{\mathsf{P}}(\boldsymbol{\beta} \mid \mathbf{v}, \mathbf{w}) = \prod_{h=1}^{H}\prod_{j=1}^{J_h}\prod_{i=1}^{n_{hj}}\left\{\frac{v_h w_{hj}\exp\left\{\boldsymbol{\beta}^\top \mathbf{x}_{hji}\right\}}{\sum_{(h',j',i')\in R(y_{hji})} v_{h'} w_{h'j'}\exp\{\boldsymbol{\beta}^\top \mathbf{x}_{h'j'i'}\}}\right\}^{\delta_{hji}}, \tag{3.63}$$

with $R(y) = \left\{(h',j',i') \mid y_{h'j'i'} \geq y\right\}$ the risk set at time $y$.

### 3.3.3 Inference

As in the case of proportional hazard models, two main estimation approaches, the parametric and the semiparametric, are available for frailty models. The two approaches differ in the way they deal with baseline risks and on the likelihood function maximised for estimation.

#### Parametric approach

The conceptually simplest approach for parameter estimation consists is making distributional assumptions on the form of the baseline hazard function:

$$\lambda_0(t) \in \left\{\lambda_0(t;\boldsymbol{\xi}), \quad \boldsymbol{\xi} \in \boldsymbol{\Xi} \subseteq \mathbb{R}^k\right\}, \tag{3.64}$$

with $k$ the dimension of the vector $\boldsymbol{\xi}$. The most common distributions for baseline hazards are exponential, Weibull, Gompertz, lognormal and loglogistic; their hazard functions are shown in Table 3.2.

For these parametric frailty models (Duchateau and Janssen, 2008, Ch. 2), the marginal likelihood (Eq. 3.56 or Eq. 3.62) is completely specified and can be maximised to estimate the full set of parameters $\boldsymbol{\zeta} = \left(\boldsymbol{\xi}^\top, \boldsymbol{\theta}^\top, \boldsymbol{\beta}^\top\right)^\top$.

An alternative to the fully parametric approach is the maximum penalised likelihood estimation proposed by Rondeau *et al.* (2003) and Rondeau and Gonzalez (2005) for gamma frailty models. The authors propose to approximate the baseline hazard by cubic M-splines, by a compromise between goodness of fit and smoothness of the approximation. Parameters are then estimated by maximising the penalised marginal loglikelihood

$$\ell_{\mathsf{PM}}(\lambda_0(\cdot), \boldsymbol{\beta}, \boldsymbol{\theta}) = \ell_{\mathsf{M}}(\lambda_0(\cdot), \boldsymbol{\beta}, \boldsymbol{\theta}) - \kappa \int_0^\infty \lambda_0''(t)\mathrm{d}t, \tag{3.65}$$

with $\ell_{\mathsf{M}}(\lambda_0(\cdot), \boldsymbol{\beta}, \boldsymbol{\theta}) = \log L_{\mathsf{M}}(\lambda_0(\cdot), \boldsymbol{\beta}, \boldsymbol{\theta})$ the log-marginal likelihood, $\kappa > 0$ a positive smoothing parameter and $\lambda_0''(t)$ the second derivative of the baseline hazard, measuring its roughness.

Strictly speaking, these are parametric models, since splines are parametric functions. Nevertheless, no probability distribution lies behind and their flexibility is much more similar to semiparametric models; for these reasons, they are considered sort of a midway between the parametric and semiparametric ones. Hirsch and Wienke (2011), indeed, classify them as "quasi-semiparametric".

### Semiparametric approach

Many situations exist in which the research interest is mainly in evaluating how covariates, e.g. treatment, change the risk, regardless of the baseline level. Furthermore, it can happen that the available knowledge of the phenomenon under study is not sufficient to make meaningful parametric assumptions on the baseline hazard. In these situations, the baseline risk is left unspecified, so the conditional likelihood (Eq. 3.55 or Eq. 3.61) is not completely defined. Extending ideas by Cox (1972), this likelihood is profiled by estimating the baseline hazard non parametrically (Duchateau and Janssen, 2008, Ch. 5).

If the profiled version of the conditional likelihood (Eq. 3.55 or Eq. 3.61) is the partial likelihood (Eq. 3.57 or Eq. 3.63), the profiled version of the full (log)likelihood (Eq. 3.54 or Eq. 3.60) is the penalised partial loglikelihood

$$\ell_{\mathsf{PP}}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \ell_{\mathsf{P}}(\boldsymbol{\beta} \mid \mathbf{u}) + \log f_{\mathbf{U}}(\mathbf{u}; \boldsymbol{\theta}), \tag{3.66}$$

where $\mathbf{U}$ (and its realisation $\mathbf{u}$) has to be intended as $\left(\mathbf{V}^\top, \mathbf{W}^\top\right)^\top$ when working with nested frailties.

Different estimation approaches based on partial likelihoods have been developed, notably for lognormal frailties (Cortiñas Abrahantes *et al.*, 2007), but the two most general and used estimation approaches are the expectation-maximization (EM) and the maximum penalised partial likelihood (PPL) algorithms (Duchateau *et al.*, 2002; Duchateau and Janssen, 2008, Sec.'s 5.1–5.2).

**EM algorithm.** The expectation-maximisation algorithm (Dempster *et al.*, 1977) was first introduced in survival analysis by Klein (1992) and then Nielsen *et al.* (1992) proposed a valuable modification to speed it up. Dempster *et al.* (1977) showed that even though the maximisation of the 'observed data' loglikelihood $\ell_{\mathsf{M}}\big(\boldsymbol{\zeta}; (\mathbf{y}, \boldsymbol{\delta})\big)$ is problematic because of integration, the same estimates can be obtained by maximising the conditional expectation of 'augmented data' loglikelihood

$\ell\big(\boldsymbol{\zeta}; (\mathbf{y}, \boldsymbol{\delta}), \mathbf{u}\big)$, given the observed data $(\mathbf{y}, \boldsymbol{\delta})$. More specifically, at iteration step $k$ this expectation is

$$\begin{aligned} Q\left(\zeta \mid \zeta^{(k-1)}\right) &= \int \ell\big(\boldsymbol{\zeta}; (\mathbf{y}, \boldsymbol{\delta}), \mathbf{u}\big) f_{\mathbf{U}}\left(\mathbf{u}; \boldsymbol{\zeta}^{(k-1)} \mid (\mathbf{y}, \boldsymbol{\delta})\right) \mathrm{d}\mathbf{u} \\ &= \int \ell_{\mathsf{C}}\big(\boldsymbol{\beta}, \lambda_0(\cdot); (\mathbf{y}, \boldsymbol{\delta}) \mid \mathbf{u}\big) f_{\mathbf{U}}\left(\mathbf{u}; \boldsymbol{\zeta}^{(k-1)} \mid (\mathbf{y}, \boldsymbol{\delta})\right) \mathrm{d}\mathbf{u} \\ &\quad + \int \log f_{\mathbf{U}}\left(\mathbf{u}; \boldsymbol{\theta}\right) f_{\mathbf{U}}\left(\mathbf{u}; \boldsymbol{\zeta}^{(k-1)} \mid (\mathbf{y}, \boldsymbol{\delta})\right) \mathrm{d}\mathbf{u}, \end{aligned} \tag{3.67}$$

given provisional estimates of the parameters $\boldsymbol{\zeta}^{(k-1)}$. Nielsen *et al.* (1992) suggested to maximise (3.67) only with respect to the regression parameters $\boldsymbol{\beta}$, while estimating $\boldsymbol{\theta}$ in an outer loop, for instance via golden section search. Thus, the function to maximise is only the first line of (3.67), with

$$\begin{aligned} \ell_{\mathsf{C}}\big(\boldsymbol{\beta}, \lambda_0(\cdot); (\mathbf{y}, \boldsymbol{\delta}) \mid \mathbf{u}\big) = \sum_{h=1}^{H} \sum_{i=1}^{n_h} \Bigg\{ \delta_{hi} \Big[ \log \lambda_0(y_{hi}) + \cancel{\log u_h} + \boldsymbol{\beta}^\top \mathbf{x}_{hi} \Big] \\ - \Lambda_0(y_{hi}) u_h \exp\left\{\boldsymbol{\beta}^\top \mathbf{x}_{hi}\right\} \Bigg\}, \end{aligned} \tag{3.68}$$

where terms $\log u_h$ can be dropped as they are additive constants with respect to parameters. Then, estimation is performed by iterating until convergence the two following steps.

E: In the expectation step, for provisional estimates of the parameters $\boldsymbol{\zeta}^{(k-1)}$, the conditional expectation of Equation 3.68,

$$\mathsf{E}\Big[\ell_{\mathsf{C}}\left(\boldsymbol{\beta}, \lambda_0(\cdot); (\mathbf{y}, \boldsymbol{\delta}) \mid \mathbf{u}\right) \mid (\mathbf{y}, \boldsymbol{\delta}), \boldsymbol{\zeta}^{(k-1)}\Big], \tag{3.69}$$

has to be computed. Thanks to the enhancement provided by Nielsen *et al.* (1992), one actually needs only the predictions

$$\begin{aligned} \tilde{u}_h^{(k)} =& \mathsf{E}\left[U_h \mid (\mathbf{y}, \boldsymbol{\delta}), \boldsymbol{\zeta}^{(k-1)}\right] \\ =& -\frac{\mathcal{L}^{(d_i+1)}\left(\sum_{i=1}^{n_h} \Lambda_0^{(k-1)}(y_{hi}) \exp\left\{{\boldsymbol{\beta}^{(k-1)}}^\top \mathbf{x}_{hi}\right\}\right)}{\mathcal{L}^{(d_i)}\left(\sum_{i=1}^{n_h} \Lambda_0^{(k-1)}(y_{hi}) \exp\left\{{\boldsymbol{\beta}^{(k-1)}}^\top \mathbf{x}_{hi}\right\}\right)}, \end{aligned} \tag{3.70}$$

where $\mathcal{L}^{(j)}(\cdot)$, $j \in \mathbb{N}$, is the $j$-th derivative of the Laplace transform of the frailty distribution and with $\Lambda_0^{(k-1)}(y_{hi}) = \sum_{y_{h'i'} \leq y_{hi}} \hat{\lambda}_0^{(k-1)}(y_{h'i'})$, where $\hat{\lambda}_0^{(k-1)}(\cdot)$ is the non-parametric Nelson (1969) and Aalen (1976) estimator. Duchateau *et al.* (2002) show details for gamma frailties, Cortiñas Abrahantes *et al.* (2007) for lognormal ones, Wang *et al.* (1995) for the positive stable case, while Munda *et al.* (2012) provide, as an aside, expressions for the inverse Gaussian distribution. A detailed overview of these results is available in Section 3.3.4.

M: In the maximisation step, for provisional predictions of the frailties $\tilde{u}_h^{(k)}$, new regression parameter estimates are looked for as

$$\boldsymbol{\beta}^{(k)} = \underset{\boldsymbol{\beta}}{\operatorname{argmax}}\, Q\left(\boldsymbol{\beta} \mid \boldsymbol{\zeta}^{(k-1)}\right), \tag{3.71}$$

which is equivalent to maximise

$$\sum_{h=1}^{H}\sum_{i=1}^{n_h}\left\{\delta_{hi}\Big[\log\lambda_0(y_{hi})+\log\tilde{u}_h^{(k)}+\boldsymbol{\beta}^\top\mathbf{x}_{hi}\Big]-\Lambda_0(y_{hi})\tilde{u}_h^{(k)}\exp\left\{\boldsymbol{\beta}^\top\mathbf{x}_{hi}\right\}\right\}, \tag{3.72}$$

as the second term of $Q\left(\zeta \mid \zeta^{(k-1)}\right)$ is constant with respect to $\boldsymbol{\beta}$. Equation 3.72 corresponds to the conditional loglikelihood of a Cox model with offsets $\tilde{u}_h^{(k)}$; therefore, estimation of the regression parameters can be done by maximisation of the Cox's partial loglikelihood

$$\sum_{h=1}^{H}\sum_{i=1}^{n_h}\delta_{hi}\left\{\log\tilde{u}_h^{(k)}+\boldsymbol{\beta}^\top\mathbf{x}_{hi}-\log\left[\sum_{(h',i')\in R(y_{hi})}\tilde{u}_{h'}^{(k)}\exp\left\{\boldsymbol{\beta}^\top\mathbf{x}_{h'i'}\right\}\right]\right\} \tag{3.73}$$

with offsets $\tilde{u}_h^{(k)}$, and $R(y_{hi})=\left\{(h',i') \mid y_{h',i'}\geq y_{hi}\right\}$ the risk set at time $y_{hi}$.

**PPL algorithm.** McGilchrist and Aisbett (1991), McGilchrist (1993) and McGilchrist (1994) first proposed to use best linear unbiased predictors (BLUP) and restricted maximum likelihood (REML) estimators for lognormal frailty models. Laplace approximation of the marginal (posterior) distribution has been suggested in a Bayesian context by Ducrocq and Casella (1996) and in a frequentist one by Ripatti and Palmgren (2000), extending results for the generalised linear mixed models with Gaussian random effects by Breslow and Clayton (1993). Finally, Therneau *et al.* (2003) proved that in the case of gamma frailties — which is largely the most common — the PPL algorithm leads to the same estimates as the EM algorithm and Therneau (2012b) implemented it in the very popular `R` package `survival`.

Let us consider for the moment the random effects $w_h=\log u_h$, then the penalised partial loglikelihood

$$\ell_{\mathsf{PP}}(\boldsymbol{\beta},\boldsymbol{\theta})=\ell_{\mathsf{P}}(\boldsymbol{\beta} \mid \mathbf{w})+\ell_{\text{pen}}(\boldsymbol{\theta};\mathbf{w}) \tag{3.74}$$

is the sum of the (conditional) partial loglikelihood

$$\ell_{\mathsf{P}}(\boldsymbol{\beta} \mid \mathbf{w})=\sum_{h=1}^{H}\sum_{i=1}^{n_h}\delta_{hi}\left\{\log w_h+\boldsymbol{\beta}^\top\mathbf{x}_{hi}-\log\left[\sum_{(h',i')\in R(y_{hi})}u_{h'}\exp\left\{\boldsymbol{\beta}^\top\mathbf{x}_{h'i'}\right\}\right]\right\} \tag{3.75}$$

and the logarithm of the distribution of the random effects

$$\ell_{\text{pen}}(\boldsymbol{\theta};\mathbf{w})=-\sum_{h=1}^{H}\log f_W(w_h;\boldsymbol{\theta}) \tag{3.76}$$

which is considered as a penalty term: if the value of $w_h$ is far away from its mean, then $\log f_W(w_h;\boldsymbol{\theta})$ will be very small and $\ell_{\text{pen}}(\boldsymbol{\theta};\mathbf{w})$ very big, lowering the likelihood value.

Maximisation of (3.65) is performed by iteration over an outer and an inner loop, as follows.

IN: At outer step $l$, a Newton–Raphson iterative procedure indexed by $k$ is used to maximise $\ell_{\mathsf{PP}}(\boldsymbol{\beta},\boldsymbol{\theta})$ for a given provisional estimate of the frailty parameter $\boldsymbol{\theta}^{(k-1)}$. The BLUPs for $\boldsymbol{\beta}$

and $\mathbf{w}$ at step $k$ are

$$\begin{bmatrix} \boldsymbol{\beta}^{(l,k)} \\ \mathbf{w}^{(l,k)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}^{(l,k-1)} \\ \mathbf{w}^{(l,k-1)} \end{bmatrix} - \mathbf{V} \begin{bmatrix} 0 \\ (\boldsymbol{\theta}^{(l)})^{-1}\mathbf{w}^{(l,k-1)} \end{bmatrix} + \mathbf{V} \begin{bmatrix} \mathbf{X} & \mathbf{Z} \end{bmatrix}^{\top} \frac{\mathrm{d}\ell_{\mathsf{P}}(\boldsymbol{\beta} \mid \mathbf{w})}{\mathrm{d}\boldsymbol{\eta}}, \tag{3.77}$$

with $\boldsymbol{\eta}$ the vector of linear predictors with generic element $w_h + \boldsymbol{\beta}^{\top}\mathbf{x}_{hi}$, $\mathbf{Z}$ the design matrix of the random effects, and where

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} \tag{3.78}$$

is the inverse of the square $(p + H)$-dimensional matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{X}^{\top} \\ \mathbf{Z}^{\top} \end{bmatrix} \left( \frac{-\mathrm{d}^2\ell_{\mathsf{P}}(\boldsymbol{\beta} \mid \mathbf{w})}{\mathrm{d}\boldsymbol{\eta}\ \mathrm{d}\boldsymbol{\eta}^{\top}} \right) \begin{bmatrix} \mathbf{X} & \mathbf{Z} \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{p\times p} & \mathbf{0}_{p\times H} \\ \mathbf{0}_{H\times p} & (\boldsymbol{\theta}^{(l)})^{-1}\mathbf{I}_H \end{bmatrix}, \tag{3.79}$$

where $\mathbf{0}_{\cdot\times\cdot}$ are null matrices, $\mathbf{I}_H$ is the identity matrix, $H$ the number of clusters and $p$ the length of $\boldsymbol{\beta}$. The asymptotic variance of the regression parameters (McGilchrist and Aisbett, 1991) is $\hat{\mathsf{V}}(\boldsymbol{\beta}) = \mathbf{V}_{11}$.

OUT: Once the inner loop has converged to estimates $\boldsymbol{\beta}^{(l)}$ and $\mathbf{w}^{(l)}$, the frailty parameter is estimated via REML in the outer loop, indexed by $l$, as

$$\boldsymbol{\theta}^{(l)} = \frac{\sum_{h=1}^{H} \left(w_h^{(l)}\right)^2}{H - \mathrm{trace}(\mathbf{V}_{22})/\boldsymbol{\theta}^{(l-1)}}. \tag{3.80}$$

The asymptotic variance (McGilchrist, 1993) is

$$\hat{\mathsf{V}}(\boldsymbol{\theta}) = \frac{2\boldsymbol{\theta}^2}{H - 2\mathrm{trace}(\mathbf{V}_{22})/\boldsymbol{\theta}^{(l-1)} + \mathrm{trace}(\mathbf{V}_{22}^2)/\boldsymbol{\theta}^{(l-1)}}. \tag{3.81}$$

### 3.3.4 Frailty distributions

The use of random frailties in proportional hazards models can serve mainly two purposes: it allows to estimate more precisely the regression parameters and it is a good means of studying the dependence between time variables in each cluster. Some recent studies suggest that regression parameters estimation is improved by frailty models, without strong importance of the frailty distribution family. On the contrary, the frailty distribution $f_U(u)$ is very important in determining the characteristics of the dependence structure. As a consequence, if investigating this dependence is the main interest of the study, the choice of the parametric form of $f_U(u)$ is crucial. The form of the frailty distribution is decisive, too, in determining the analytical tractability of the marginal likelihood (3.56) and of the penalized partial likelihood (3.74) functions.

As discussed at the beginning of this Section 3.3, the selection effect is one of the most interesting aspects of the survival data: as time goes, high-risk individuals tend to experience the event first, thereby causing the selection of the more robust ones. To study this selection effect, it is useful to consider the updated (or conditional) frailty distribution for a given cluster (Duchateau and Janssen, 2008, Section 4.1.3), i.e. the distribution of the frailty term $U_h$, taking into account the

information collected until a given time $t \geq 0$.

**Updated frailty distribution.** Let $\mathcal{D}(t) = \Big\{\big(\breve{y}_{hi}(t), \delta_{hi}(t)\big); h = 1, \ldots, H, i = 1, \ldots, n_h\Big\}$ be the data collected until time $t$, with $\breve{y}_{hi}(t) = \min(y_{hi}, t)$ the event or censoring time at $t$ and $\delta_{hi}(t) = \delta_{hi}\mathbb{1}\big(y_{hi} \leq t\big)$ the event/censoring indicator. In the following, the likelihood quantities have to be intended as computed at values $\breve{y}$'s instead of usual $y$'s. Then, the updated, or conditional, frailty distribution of $U_h$ at time $t$ is

$$\begin{aligned} f_{U_h}\big(u; \boldsymbol{\xi} \mid \mathcal{D}(t)\big) =& \frac{\mathsf{P}[\mathcal{D}(t); \boldsymbol{\beta}, \boldsymbol{\zeta} \mid U_h = u]}{\mathsf{P}[\mathcal{D}(t); \boldsymbol{\xi}]} f_U(u; \boldsymbol{\theta}) \\ =& \frac{L_{\mathsf{C},h}(\boldsymbol{\beta}, \boldsymbol{\zeta}; u)}{L_{\mathsf{M},h}(\boldsymbol{\xi})} f_U(u; \boldsymbol{\theta}) \\ =& \frac{\exp\Big\{ - u \sum_{i=1}^{n_h} \Lambda_0\big(\breve{y}_{hi}(t)\big) \exp\big\{\boldsymbol{\beta}^\top \mathbf{x}_{hi}\big\}\Big\}}{(-1)^{d_h(t)} \mathcal{L}^{(d_h(t))}\Big( \sum_{i=1}^{n_h} \Lambda_0\big(\breve{y}_{hi}(t)\big) \exp\big\{\boldsymbol{\beta}^\top \mathbf{x}_{hi}\big\}\Big)} u^{d_h(t)} f_U(u; \boldsymbol{\theta}), \end{aligned} \tag{3.82}$$

$u > 0$, with $d_h(t) = \sum_{i=1}^{n_h} \delta_{hi}(t)$.

For ease of notation, in the following, we drop the dependence on parameters whenever this does not induce confusion and we denote just by $\mathcal{L}^{(d_t+j)}$ the derivatives of the Laplace transform $\mathcal{L}^{(d_h(t)+j)}\Big( \sum_{i=1}^{n_h} \Lambda_0\big(\breve{y}_{hi}(t)\big) \exp\big\{\boldsymbol{\beta}^\top \mathbf{x}_{hi}\big\}\Big)$, with $d_t = d_h(t)$.

Note that the updated frailty distribution is in general different from the marginal frailty distribution $f_U(u) = f_{U_h}\big(u \mid \mathcal{D}(0)\big)$ and that the updated distributions $f_{U_h}\big(u \mid \mathcal{D}(t)\big)$ are different to each other as different information is available for each cluster.

To understand how the updating process impacts on clustering, we now consider the updated mean, variance and skewness as functions of the time.

**Updated mean.** The updated mean of the frailty $U_h$ at time $t$ is

$$\begin{aligned} \mathsf{E}\big[U_h \mid \mathcal{D}(t)\big] =& \int_0^\infty u f_{U_h}\big(u \mid \mathcal{D}(t)\big) \mathrm{d}u \\ =& \frac{\int_0^\infty \exp\Big\{ - u \sum_{i=1}^{n_h} \Lambda_0\big(\breve{y}_{hi}(t)\big) \exp\big\{\boldsymbol{\beta}^\top \mathbf{x}_{hi}\big\}\Big\} u^{d_t+1} f_U(u) \mathrm{d}u}{(-1)^{d_t} \mathcal{L}^{(d_t)}} \\ =& - \frac{\mathcal{L}^{(d_t+1)}}{\mathcal{L}^{(d_t)}}. \end{aligned} \tag{3.83}$$

**Updated Variance.** The variance of the frailty term, when accounting for data collected until time $t$, is

$$\begin{aligned} \mathsf{V}\big[U_h \mid \mathcal{D}(t)\big] =& \int_0^\infty u^2 f_{U_h}\big(u \mid \mathcal{D}(t)\big) \mathrm{d}u - \mathsf{E}\big[U_h \mid \mathcal{D}(t)\big]^2 \\ =& \frac{\mathcal{L}^{(d_t)} \mathcal{L}^{(d_t+2)} - \big[\mathcal{L}^{(d_t+1)}\big]^2}{\big[\mathcal{L}^{(d_t)}\big]^2} > 0, \end{aligned} \tag{3.84}$$

positive by definition of variance.

**Updated Skewness.** The frailty variable has conditional skewness (Proof A.3)

$$\mathbb{A}\big[U_h \mid \mathcal{D}(t)\big] = \frac{3\mathcal{L}^{(d_t)}\mathcal{L}^{(d_t+1)}\mathcal{L}^{(d_t+2)} - \big[\mathcal{L}^{(d_t)}\big]^2\mathcal{L}^{(d_t+3)} - 2\big[\mathcal{L}^{(d_t+1)}\big]^3}{\Big(\mathcal{L}^{(d_t)}\mathcal{L}^{(d_t+2)} - \big[\mathcal{L}^{(d_t+1)}\big]^2\Big)^{3/2}}. \tag{3.85}$$

**Properties.** The definition of the derivatives of the Laplace transform (3.52) implies that

$$\mathsf{E}\big[U_h \mid \mathcal{D}(t)\big] > 0, \tag{3.86}$$

that is coherent with the fact that the frailty term can assume only positive values marginally, so a fortiori conditionally.

Furthermore, if the updated mean is considered as function of the time $t$, its derivative is

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{E}\big[U_h \mid \mathcal{D}(t)\big] = &- \frac{\mathcal{L}^{(d_t)}\mathcal{L}^{(d_t+2)} - \big[\mathcal{L}^{(d_t+1)}\big]^2}{\big[\mathcal{L}^{(d_t)}\big]^2} \\ &\frac{\mathrm{d}}{\mathrm{d}t}\left[\sum_{i=1}^{n_h}\Lambda_0\big(\breve{y}_{hi}(t)\big)\exp\big\{\boldsymbol{\beta}^\top\mathbf{x}_{hi}\big\}\right] \\ = &- \mathsf{V}\big[U_h \mid T > t\big]\left[\sum_{i=1}^{n_h}\mathbb{1}(y_{hi}\geq t)\lambda_0\big(\breve{y}_{hi}(t)\big)\exp\big\{\boldsymbol{\beta}^\top\mathbf{x}_{hi}\big\}\right] \leq 0.\end{aligned} \tag{3.87}$$

This means that, as time passes and new information is available, the (conditional) expected value of the frailty decreases for those subjects who are still at risk.

As time passes and the event of interest does not occur, also the conditional variance varies. One can show that the conditional variance decreases if and only if the conditional frailty density is positively skewed for any $t > 0$ (Proof A.4). Now the most frequently used frailty distributions are illustrated. For each one, we give the distribution, the Laplace transform and its derivatives — useful for computing the marginal likelihood — and the Kendall's Tau (Kendall, 1938; Duchateau and Janssen, 2008, Sec. 4.1.4). This is a measure of dependence that can be computed in general as

$$\tau = 4\int_0^\infty s\mathcal{L}(s)\mathcal{L}(s)^{(2)}\mathrm{d}s - 1, \tag{3.88}$$

originally developed as a concordance measure for bivariate clusters, defined as

$$\tau = \mathsf{E}\Big[\mathrm{sign}\big((T_{h1} - T_{k1})(T_{h2} - T_{k2})\big)\Big], \tag{3.89}$$

for any two clusters $h$ and $k$, with

$$\mathrm{sign}(x) = \begin{cases} -1, & x < 0, \\ 0, & x = 0, \\ 1, & x > 0. \end{cases} \tag{3.90}$$

In addition, we provide the updated distribution is given, too. The closure property, i.e. the fact that the updated frailty distribution belongs to the same family of the marginal distribution,

is appealing in some contexts: for instance Economou and Caroni (2005) developed diagnostic plots to test for the appropriateness of frailty distributions that benefit from the closure property. See Appendix B.2 for a summary table.

### Gamma

The one-parameter gamma distribution (Gam*), with density

$$f_U(u;\theta) = \frac{u^{1/\theta-1}\exp(-u/\theta)}{\Gamma(1/\theta)\theta^\theta}, \tag{3.91}$$

with $\theta > 0$, is the most popular frailty distribution as long as most computations can be done analytically (unconditional survival, cumulative hazard and hazard function). This is due to the simplicity of its Laplace transform

$$\mathcal{L}(s) = (1+\theta s)^{-1/\theta} \tag{3.92}$$

and of its derivatives

$$\mathcal{L}^{(k)}(s) = (-1)^k (1+\theta s)^{-k} \left[\prod_{l=0}^{k-1}(1+l\theta)\right] \mathcal{L}(s). \tag{3.93}$$

The gamma frailty typically models late dependence (Duchateau and Janssen, 2008, Fig. 4.8) and, despite there is no biological reason to use it, it is far the most used one due to its mathematical properties.

The updated distribution comes out to be a gamma distribution (Gam) with shape parameter $1/\theta$ and scale parameter $\left(\theta + \sum_{i=1}^{n_h} \Lambda_{hi}(\breve{y}(t))\right)^{-1}$.

The Kendall's $\tau$, measuring the overall dependence, becomes

$$\tau = \theta/(\theta+2) \in [0,1). \tag{3.94}$$

### Inverse Gaussian

The one-parameter inverse Gaussian (IG*) distribution, has density

$$f_U(u;\theta) = (2\theta\pi)^{-1/2} u^{-3/2} \exp\left\{-\frac{(u-1)^2}{2u\theta}\right\}, \tag{3.95}$$

with $\theta > 0$ and it is a special case of the PVF* distribution (see below, pg. 37), when $\nu = 1/2$.

The Laplace transform for the IG* distribution is

$$\mathcal{L}(s) = \exp\left[\left\{1-(1+2\theta s)^{1/2}\right\}\Big/\theta\right] \tag{3.96}$$

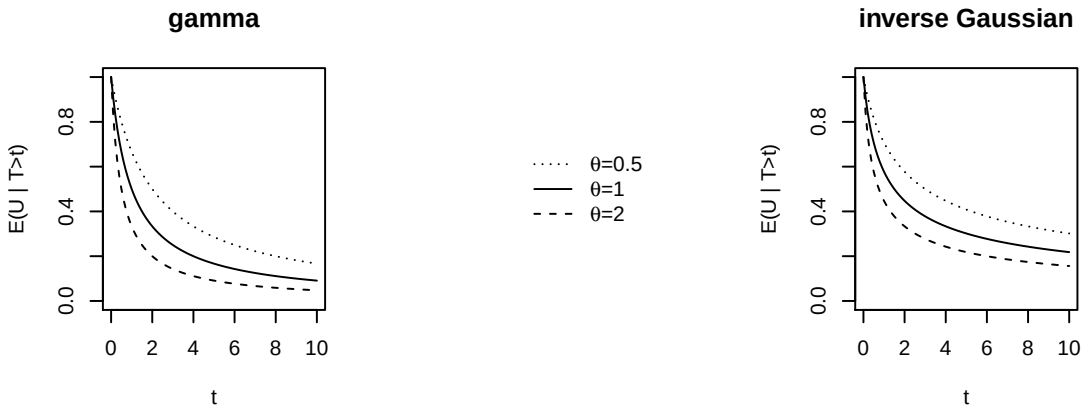



Figure 3.6 – Updated frailty mean for gamma and inverse Gaussian distributions. Univariate case with baseline hazard $\lambda_0(t) = 1$ and no covariates.

with derivatives

$$\mathcal{L}^{(k)}(s) = (-1)^k \left(2\theta s + 1\right)^{-\frac{k}{2}} \frac{K_{k-(1/2)}\left(\sqrt{2\theta^{-1}(s+\frac{1}{2\theta})}\right)}{K_{1/2}\left(\sqrt{2\theta^{-1}(s+\frac{1}{2\theta})}\right)} \mathcal{L}(s), \tag{3.97}$$

where $K(\cdot)$ is the modified Bessel function of the second kind (Hougaard, 2000, Section A.4.2)

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

The updated distribution $f_U(u \mid \mathcal{D}(t))$ is a PVF with location parameter $(1 + 2\theta\Lambda_{hi}(\breve{y}(t)))^{-1/2}$, scale parameter $\theta(1+2\theta\Lambda_{hi}(\breve{y}(t)))^{-1/2}$ and shape parameter $1/2$. It belongs, too, to the Generalized Inverse Gaussian family (Hougaard, 2000, pg. 243 and Appendix A.3.6).

Figure 3.6 shows an example of the behaviour of the updated frailty mean for the gamma and the inverse Gaussian distributions. The univariate case is considered, without covariates and with constant baseline hazard.

The Kendall's $\tau$ is

$$\tau = \frac{1}{2} - \frac{1}{\theta} + 2\frac{e^{2/\theta}}{\theta^2} \int_{2/\theta}^{\infty} u^{-1} e^{-u} \mathrm{d}u, \tag{3.98}$$

which implies that $\tau \in (0, 1/2)$ (Duchateau and Janssen, 2008, pg. 161).

The IG* frailty depicts a type of dependence which is, for a fixed Kendall's $\tau$, between the early (PS*) and the late (Gam*) ones (Duchateau and Janssen, 2008, Fig. 4.17).

### Positive stable

The one-parameter positive stable (PS*) distribution has density

$$f_U(u;\theta) = -\frac{1}{\pi u} \sum_{k=1}^{\infty} \frac{\Gamma(k\theta+1)}{k!} \left(-u^{-\theta}\right)^k \sin(\theta k \pi), \tag{3.99}$$

with $\theta \in (0,1)$. It has a very simple Laplace transform

$$\mathcal{L}(s) = \exp\left(-s^{\theta}\right) \tag{3.100}$$

with derivatives

$$\mathcal{L}^{(k)}(s) = (-1)^k \left(\theta s^{\theta-1}\right)^k \left[\sum_{m=0}^{k-1} \Omega_{k,m} s^{-m\theta}\right] \mathcal{L}(s), \tag{3.101}$$

where the $\Omega_{k,m}$'s are polynomials of degree $m$, given recursively by

$$\begin{aligned} \Omega_{k,0} &= 1, \\ \Omega_{k,m} &= \Omega_{k-1,m} + \Omega_{k-1,m-1}\left\{\frac{k-1}{\theta} - (k-m)\right\}, \qquad m = 1, \ldots, k-2, \\ \Omega_{k,k-1} &= \theta^{1-k}\frac{\Gamma(k-\theta)}{\Gamma(1-\theta)}. \end{aligned} \tag{3.102}$$

The PS* distribution has infinite mean and therefore undetermined variance; this implies that the heterogeneity parameter is independent of the covariates (Hougaard, 1986; Duchateau and Janssen, 2008, Sec. 4.4.1) and that the updated distribution $f_U(u \mid \mathcal{D}(t))$ belongs to the PVF family. An important feature of PS* frailties is that population hazards are still proportional.

The Kendall's $\tau$ is

$$\tau = 1 - \theta. \tag{3.103}$$

The PS frailty is best suited for modelling early dependence (Duchateau and Janssen, 2008, Fig. 4.25).

### Power variance function

The two-parameter power variance Function (PVF*) distribution has density

$$\begin{aligned} f_U(u;\theta,\nu) = &\exp\left\{-\frac{\nu}{\theta}\left(u + \frac{1}{\nu-1}\right)\right\}\frac{1}{\pi u} \\ &\times \sum_{k=1}^{\infty} \frac{(\nu/\theta)^{k\nu} u^{k(\nu-1)} \Gamma\{1-k(\nu-1)\}\sin\{\pi k(\nu-1)\}}{k!(\nu-1)^k}, \end{aligned} \tag{3.104}$$

with $\theta > 0$ and $\nu \in (0,1]$. It is a very important family, as it includes the inverse Gaussian ($\nu = 1/2$, $\mu = 1$) model as particular case, and the one-parameter gamma is its natural extension at the boundary of its parametric space.

Furthermore, the CP* (see below) and PS* models have important links to it, which will be showed later on.

The Laplace transform of the PVF* distribution is

$$\mathcal{L}(s) = \exp\left[\frac{\nu}{\theta(1-\nu)}\left\{1 - \left(1 + \frac{\theta s}{\nu}\right)^{1-\nu}\right\}\right]. \tag{3.105}$$

The updated distribution is a three-parameter power variance function distribution (PVF) with location parameter $(1 + \theta\Lambda_{hi}(\breve{y}(t))/\nu)^{-\nu}$, scale parameter $\theta(1 + \theta\Lambda_{hi}(\breve{y}(t))/\nu)^{\nu-1}$ and shape

parameter $\nu$.

The Kendall's $\tau$, measuring the overall dependence (Duchateau and Janssen, 2008, Sec. 4.5.5), becomes

$$\tau = \nu - 2\frac{\nu}{\theta} + \frac{4\nu^2}{\theta^2(1-\nu)} \exp\left(2\frac{\nu}{\theta(1-\nu)}\right) \int_1^\infty t^{-\nu/(1-\nu)} \exp\left(-2\frac{\nu t}{\theta(1-\nu)}\right) \mathrm{d}t. \tag{3.106}$$

### Compound Poisson

The two-parameter compound Poisson (CP*, Aalen, 1992) is a mixture distribution with

$$\begin{cases} \mathsf{P}[U=0;\theta,\nu] = \exp\left\{\dfrac{-\nu}{\theta(\nu-1)}\right\} \\ f_U(u;\theta,\nu) = \exp\left\{-\dfrac{\nu}{\theta}\left(u+\dfrac{1}{\nu-1}\right)\right\}\dfrac{1}{\pi u} \\ \qquad\qquad \times \displaystyle\sum_{k=1}^{\infty} \frac{(\nu/\theta)^{k\nu} u^{k(\nu-1)} \Gamma\{1-k(\nu-1)\}\sin\{\pi k(\nu-1)\}}{k!(\nu-1)^k} \end{cases} \tag{3.107}$$

with $\theta > 0$ and $\nu > 1$. Its Laplace transform is

$$\mathcal{L}(s) = \exp\left[\frac{\nu}{\theta(1-\nu)}\left\{1-\left(1+\frac{\theta s}{\nu}\right)^{1-\nu}\right\}\right]. \tag{3.108}$$

The updated distribution comes out to be a three-parameter compound Poisson distribution (CP) with location parameter $(1+\theta\Lambda_{hi}(\breve{y}(t))/\nu)^{-\nu}$, scale parameter $\theta(1+\theta\Lambda_{hi}(\breve{y}(t))/\nu)^{\nu-1}$ and shape parameter $\nu$.

The CP* frailty gives non-null probability to the event that the risk is constantly 0, i.e. that a cluster is totally out of risk. This is useful, for instance, in the case that clusters are patients and we suppose that they can be totally cured with some probability.

### Lognormal

The one-parameter lognormal (LN) distribution, with density

$$f_U(u;\sigma) = \frac{1}{u\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(\log u)^2}{2\sigma^2}\right\}, \tag{3.109}$$

with $\sigma > 0$, is used to link the frailty models to the widespread GLMMs, which use zero-mean Normal errors.

In this case, the mean is $\mathsf{E}(U) = e^{\sigma^2/2}$, the variance $\mathsf{V}(U) = \exp\{2\sigma^2\} - \exp\{\sigma^2\}$ but the Laplace transform is not available in an explicit form. Consequently, both the updated distribution and the Kendall's $\tau$ are not straightforwardly attainable.

### Other distributions

Some other possibilities exist, but they are rarely used. In Sec. 3.9–3.11 of Wienke (2010) a brief review and some references on the quadratic hazard, the log-Student $t$, and the Lévy distribution

can be found.

Table 3.3 resumes the probability density functions and the associated Laplace transforms for all the frailty distributions reviewed in this Section 3.3.4. Figure 3.7 shows the relations linking the main frailty families; note that the three-parameter PVF family can be obtained by left truncation of a PS* distribution (Hougaard, 2000, Sec. 7.5).

| **Distribution** | **Density** $f_U(u)$ | **Laplace Transform** $\mathcal{L}(s)$ |
|---|---|---|
| **Gam*** | $\frac{u^{1/\theta-1}e^{-u/\theta}}{\Gamma(1/\theta)\theta^{1/\theta}}$ | $(1+\theta s)^{-1/\theta}$ |
| **PS*** | $-\frac{1}{\pi u}\sum_{k=1}^{\infty}\frac{\Gamma(k\theta+1)}{k!}\left(-u^{-\theta}\right)^k\sin(\theta k\pi)$ | $\exp\left(-s^{\theta}\right)$ |
| **IG*** | $(2\theta\pi)^{-1/2}u^{-3/2}e^{-\frac{(u-1)^2}{2u\theta}}$ | $\exp\left[\left\{1-(1+2\theta s)^{1/2}\right\}/\theta\right]$ |
| **CP*** | $\frac{e^{-\frac{\nu}{\theta}\left(u+\frac{1}{\nu-1}\right)}}{\pi u}\times\sum_{k=1}^{\infty}\frac{(\nu/\theta)^{k\nu}u^{k(\nu-1)}\Gamma\{1-k(\nu-1)\}\sin\{\pi k(\nu-1)\}}{k!(\nu-1)^k}$ and $\mathsf{P}[U=0]=e^{\frac{-\nu}{\theta(\nu-1)}}$ | $\exp\left[\frac{\nu}{\theta(1-\nu)}\left\{1-\left(1+\frac{\theta s}{\nu}\right)^{1-\nu}\right\}\right]$ |
| **PVF*** | $\frac{e^{-\frac{\nu}{\theta}\left(u+\frac{1}{\nu-1}\right)}}{\pi u}\times\sum_{k=1}^{\infty}\frac{(\nu/\theta)^{k\nu}u^{k(\nu-1)}\Gamma\{1-k(\nu-1)\}\sin\{\pi k(\nu-1)\}}{k!(\nu-1)^k}$ | $\exp\left[\frac{\nu}{\theta(1-\nu)}\left\{1-\left(1+\frac{\theta s}{\nu}\right)^{1-\nu}\right\}\right]$ |
| **LN*** | $\frac{1}{u\sqrt{2\pi\gamma}}e^{-\frac{(\log u)^2}{2\gamma}}$ | — |

Table 3.3 – Frailty distributions, with density and Laplace transform. Gam*: gamma, PS*: Positive Stable, IG*: Inverse Gaussian, CP*: compound Poisson, PVF*: Power variance Function, LN*: lognormal.

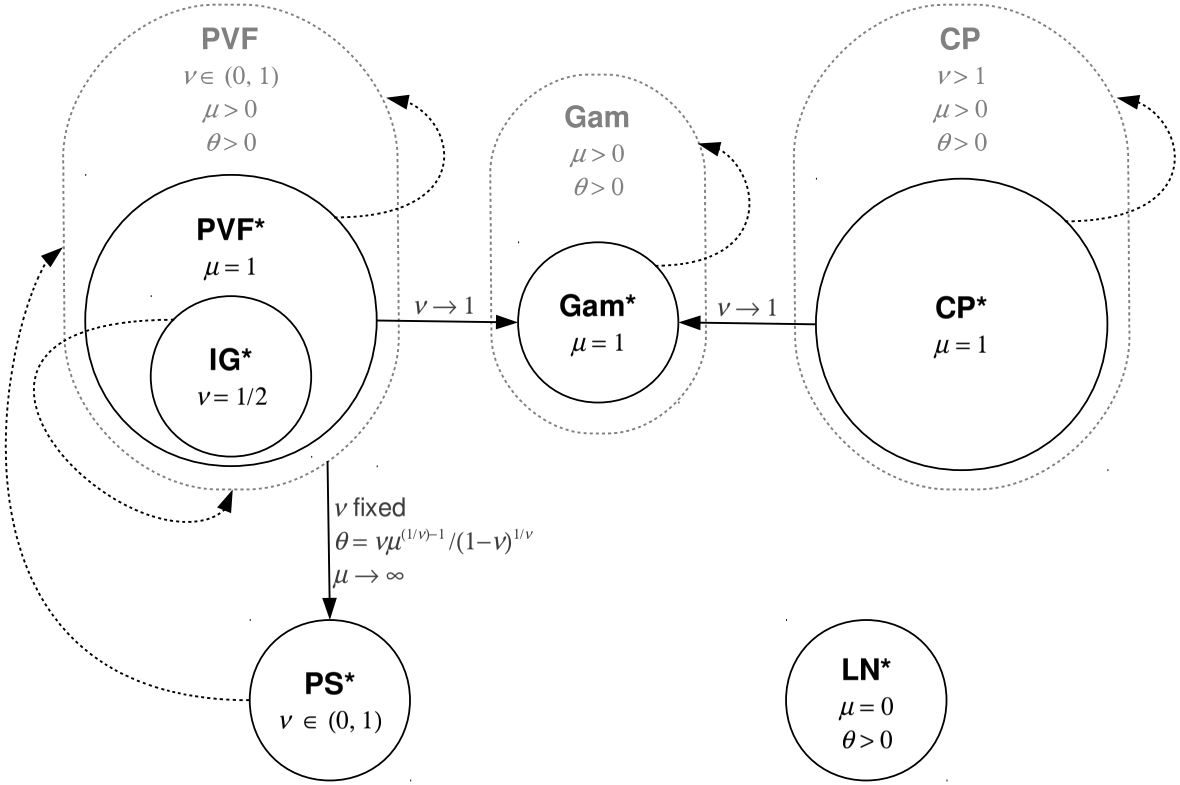



Figure 3.7 – Families of distributions for frailty terms. PVF: Power variance Function, Gam: gamma, CP: compound Poisson, IG: Inverse Gaussian, PS: Positive Stable, LN: lognormal. *: distribution with the constraint(s) needed by frailty terms. Grey dotted sets are general families; black solid sets are families used as frailty distributions. Solid arrows denote extensions by limits on parameters values. Dotted arrows denote the family of the updated distribution.

# Chapter 4

# Simulation of clustered multi-state data

Simulation studies are a powerful means to evaluate the performance of analysis methods (Burton *et al.*, 2006) and they are particularly useful when developing new models. To this aim, generating survival data from proportional hazards models is quite simple (Bender *et al.*, 2005). Simulated data from frailty models can be obtained by first generating the frailty terms from their distribution and then by using a Cox model, conditionally on them (Cortiñas Abrahantes *et al.*, 2007, Sec. 5). In the context of competing risks, Beyersmann *et al.* (2009) developed a method based on the cause-specific hazards (Gray, 1988), which they claim could be extended to multi-state models. It is very appealing, but in general the probabilities of competing events and any location measure of time variables cannot be analytically expressed in terms of the simulation parameters. So there is no direct way to choose these latter in order to obtain data with given features.

Simulation of multi-state data raises more problems and most of the examples in the literature are based on real data (Putter *et al.*, 2007; Commenges, 1999; de Wreede *et al.*, 2010, e.g.). In some cases data are simulated by assuming independence between times of different transitions of the same subject (Meira-Machado *et al.*, 2006, e.g.). Farlie-Gumbel-Morgenstern copula models (Farlie, 1960) are suited for bivariate models (de Uña-Álvarez and Meira-Machado, 2008; Amorim *et al.*, 2011) but, if extended to $K > 2$ events, they give times which are $(K-1)$-wise independent. Ad hoc solutions can be obtained for simple structures (Van Keilegom *et al.*, 2011, Sec. 4, e.g.), but with a lack of generality.

We propose (Rotolo *et al.*, 2012a) a general simulation procedure for clustered multi-state survival data based on a copula model for each group of competing events. Thanks to it, any parametric form can be chosen for the marginal distributions, whereas dependence is induced by the copula. This structure allows to introduce and tune many features; the most important ones are (i) the dependence between times of competing events, (ii) the dependence between times of successive events, (iii) the dependence between times of clustered subjects and (iv) the event-specific covariate effects. Frailties and covariates are inserted in a proportional-hazards way, as assumed by most parametric regression models. Both random and fixed censoring can be added, too.

This procedure to simulate data can be used with any user-defined values of the parameters.

However, we also propose an additional numerical procedure that can be used to find appropriate values of parameters to simulate data according to given requirements. The researcher can fix parameters to obtain chosen target values for clinically meaningful quantities such as median times and probabilities of censoring and of competing events.

## 4.1 Simulation model

Consider a general acyclic multi-state structure, with $N$ states and $Q$ possible transitions between them; let $\mathcal{S} = \{s_1, \ldots, s_N\}$ be the set of the states and $\mathcal{T} = \{T_1, \ldots, T_Q\}$ be the set of the transition time variables. As stated in Section 3.2, for each state $s_i$, the set of its children $\mathcal{C}(s_i) \subset \mathcal{S}$ is defined as the set of the states to which a direct transition from $s_i$ is possible. As an example, consider the multi-state structure in Figure 4.1, corresponding to the possible event history of patients in a cancer study. The set of states is $\mathcal{S} = \{\mathsf{NED}, \mathsf{LR}, \mathsf{DM}, \mathsf{De}\}$, corresponding to no evidence of new disease, occurrence of local relapse, occurrence of distant metastases and death, respectively. The children sets of the four states are $\mathcal{C}(\mathsf{NED}) = \{\mathsf{LR}, \mathsf{DM}, \mathsf{De}\}$, $\mathcal{C}(\mathsf{LR}) = \{\mathsf{De}\}$, $\mathcal{C}(\mathsf{DM}) = \{\mathsf{De}\}$ and $\mathcal{C}(\mathsf{De}) = \emptyset$. We only consider the occurrence of one intermediate adverse event ($\mathsf{LR}$ or $\mathsf{DM}$) as, once one of them has occurred, the occurrence of the second does not further worsen prognosis appreciably.

For each continuous transition time variable $T_q$, let $F_q(t)$ be an arbitrarily chosen distribution, with $f_q(t)$ the corresponding density function. The associated survival function is then $S_q(t) = 1 - F_q(t)$. The joint survival function of a given set of transition times $\{T_q\}_{q\in\mathcal{Q}} \subseteq \mathcal{T}$ is denoted by $S_{\mathcal{Q}}(\mathbf{t}_{\mathcal{Q}}) = \mathsf{P}\big[\cap_{q\in\mathcal{Q}}\,(T_q > t_q)\big]$, with $\mathbf{t}_{\mathcal{Q}} = (t_q, q \in \mathcal{Q})$. Except under independence, this joint survival function is not uniquely determined by the marginal ones. To combine them together into a joint survival function, a copula (Nelsen, 2006) can be used. Copulas are multivariate distribution functions with uniform margins, that are used to express multivariate distribution functions, or survival functions, in terms of their one-dimensional margins.

We define a *competing risks (or events) block* as the set of the transitions into all the children of a state. In the example, three competing events blocks are present: $\{T_1, T_2, T_3\}$, $\{T_4\}$ and $\{T_5\}$. The two latter ones are degenerate competing risks blocks, as they both contain only one element. The proposed simulation method adopts a copula model for each competing risks block.

**First transitions.** Consider first $\{T_q\}_{q\in\mathcal{Q}_0}$, the competing risks block of transitions from the starting state. A copula is used to combine the marginal survival functions into the joint survival function

$$S_{\mathcal{Q}_0}(\mathbf{t}_{\mathcal{Q}_0}) = \mathbf{C}_\vartheta\big(S_q(t_q), q \in \mathcal{Q}_0\big), \qquad \mathbf{t}_{\mathcal{Q}_0} = (t_q, q \in \mathcal{Q}_0) \tag{4.1}$$

with $\mathbf{C}_\vartheta(\cdot)$ the copula and $\vartheta$ the dependence parameter.

Within the block $\mathcal{Q}_0$, the joint survival function of the first $k$ times (the order has no importance) is $S_{\mathcal{Q}_0}\left(t_{(1)}, \ldots, t_{(k)}, 0, \ldots 0\right)$, with $(q)$ the order indices. The conditional survival function of the $k$-th time, given the $k-1$ previous ones, is

$$S_{(k)|(1),\ldots,(k-1)}\left(t_{(k)} \mid t_{(1)}, \ldots, t_{(k-1)}\right) = \frac{\frac{\partial^{k-1}}{\partial t_{(1)}\cdots\partial t_{(k-1)}} S_{\mathcal{Q}_0}\left(t_{(1)}, \ldots, t_{(k)}, 0, \ldots, 0\right)}{\frac{\partial^{k-1}}{\partial t_{(1)}\cdots\partial t_{(k-1)}} S_{\mathcal{Q}_0}\left(t_{(1)}, \ldots, t_{(k-1)}, 0, \ldots, 0\right)} \tag{4.2}$$

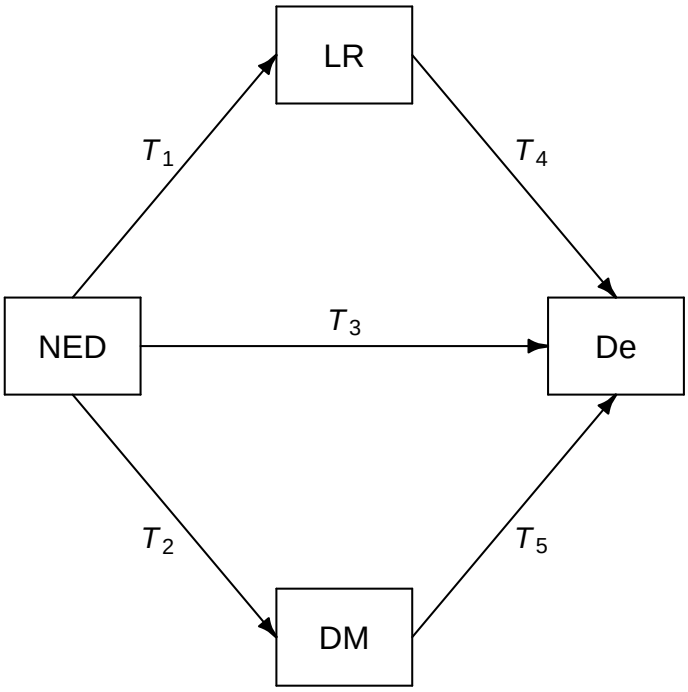



Figure 4.1 – Reduced cancer model. NED: No Evidence of new Disease, LR: Local Relapse, DM: Distant Metastases, De: Dead.

for $k = 2, \ldots, \#(\mathcal{Q}_0)$ with $\#(\mathcal{Q}_0)$ the size of $\mathcal{Q}_0$.

Without loss of generality, consider for ease of notation and of presentation the example of Figure 4.1, where $\mathcal{Q}_0 = \{1, 2, 3\}$, with a Clayton copula (Clayton, 1978). In this case (4.1) is

$$S_{\mathcal{Q}_0}(\mathbf{t}_{\mathcal{Q}_0}) = \left(1 + \sum_{q \in \mathcal{Q}_0} \left[S_q(t_q)^{-\vartheta} - 1\right]\right)^{-1/\vartheta}, \qquad \mathcal{Q}_0 = (1, 2, 3), \tag{4.3}$$

with $\vartheta > 0$ and one can easily show by induction (App. A.5) that the derivatives are of the form

$$\frac{\partial^k}{\partial t_{(1)} \cdots \partial t_{(k)}} S_{\mathcal{Q}_0}(\mathbf{t}_{\mathcal{Q}_0}) = (-1)^k \prod_{h=1}^{k} \left[(1 + (h-1)\vartheta) \left(S_h(t_h)\right)^{-\vartheta - 1} f_h(t_h)\right] \left(S_{\mathcal{Q}_0}(\mathbf{t}_{\mathcal{Q}_0})\right)^{1 + k\vartheta}. \tag{4.4}$$

Therefore (4.2) becomes

$$S_{(k)|(1),\ldots,(k-1)}\left(t_{(k)} \mid t_{(1)}, \ldots, t_{(k-1)}\right) = \left(1 + \frac{S_k(t_k)^{-\vartheta} - 1}{1 + \sum_{q=1}^{k-1}(S_q(t_q)^{-\vartheta} - 1)}\right)^{1 - k - \frac{1}{\vartheta}} \tag{4.5}$$

for $k = 2, 3$.

The survival functions $S_1(t_1)$, $S_{2|1}(t_2 \mid t_1)$ and $S_{3|1,2}(t_3 \mid t_1, t_2)$ can be used in sequence to

simulate the times in $\mathcal{Q}_0$ from their joint survival function (4.3).

**Second and following transitions.** Once a subject has moved to another state, another copula model can be adopted for the incoming transition and all the competing events from the new present state. In our example, we use again the copula (4.3) for $\{1\} \cup \mathcal{Q}_1$, with $\mathcal{Q}_1 = \{4\}$, and for $\{2\} \cup \mathcal{Q}_2$, with $\mathcal{Q}_2 = \{5\}$. Thus we have $S_{\{1,4\}}(\mathbf{t}_{\{1,4\}}) = \left(S_1(t_1)^{-\vartheta} + S_4(t_4)^{-\vartheta} - 1\right)^{-1/\vartheta}$ and $S_{\{2,5\}}(\mathbf{t}_{\{2,5\}}) = \left(S_2(t_2)^{-\vartheta} + S_5(t_5)^{-\vartheta} - 1\right)^{-1/\vartheta}$, which gives

$$S_{4|1}(t_4 \mid t_1) = \left[1 + \left(\frac{S_1(t_1)}{S_4(t_4)}\right)^{\vartheta} - S_1(t_1)^{\vartheta}\right]^{-1/\vartheta - 1} \tag{4.6}$$

and

$$S_{5|2}(t_5 \mid t_2) = \left[1 + \left(\frac{S_2(t_2)}{S_5(t_5)}\right)^{\vartheta} - S_2(t_2)^{\vartheta}\right]^{-1/\vartheta - 1}. \tag{4.7}$$

If the multi-state structure had other further transitions, the same approach should be replicated.

**Time scale.** Time values originated by $S_{q'|q}(t \mid t_q)$ range from 0 to $\infty$ (Fig. 4.2).

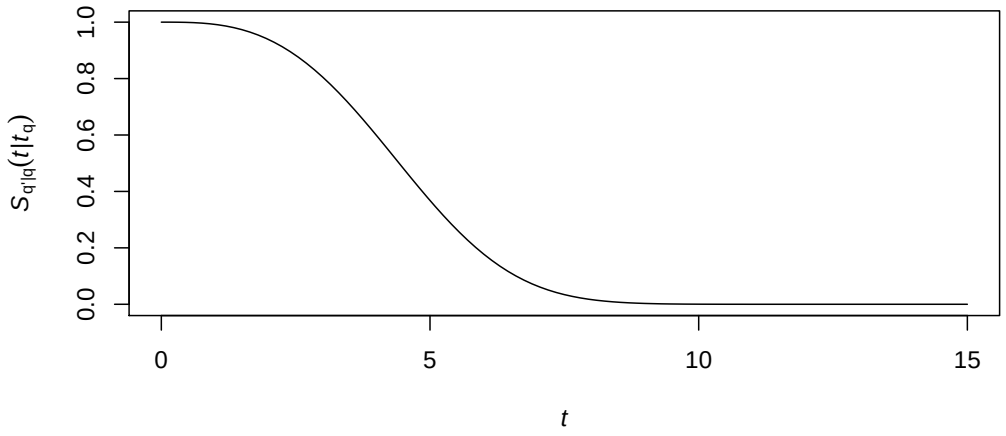



Figure 4.2 – Survival function used for clock-reset simulations.

As time until $t_q$ has already elapsed, the simulation of the additional time from $S_{q'|q}(t)$ corresponds to the clock-reset approach (Putter *et al.*, 2007, Sec. 4.2.2) and must be added to $t_q$.

In the case the *'clock forward'* approach is more appropriate, the truncated survival function

$$\begin{aligned} S_{q'|q}(t \mid t_q; T_{q'} > t_q) &= \mathbb{P}(T_{q'} > t \mid T_q = t_q, T_{q'} > t_q) \\ &= \frac{\mathbb{P}(T_{q'} > t, T_{q'} > t_q \mid T_q = t_q)}{\mathbb{P}(T_{q'} > t_q \mid T_q = t_q)} \\ &= \frac{S_{q'|q}(t \mid t_q)}{[S_{q'|q}(t_{q'} \mid t_q)]}, \qquad t > t_q, \end{aligned} \tag{4.8}$$

(Fig. 4.3) must be used.

The values of $T_{q'}$, which are necessary greater than $t_q$, do not have to be added to those of the first transition.

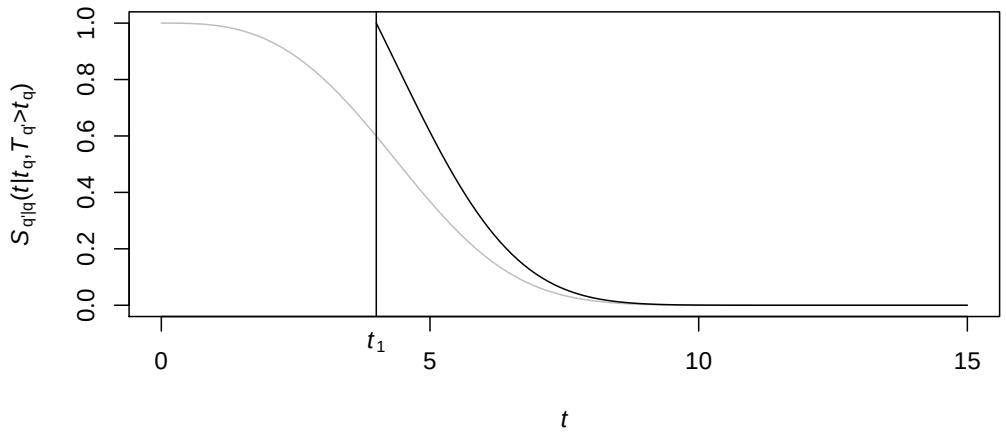



Figure 4.3 – Survival function used for clock-forward simulations.

### 4.1.1 Simulation Algorithm

We illustrate below the algorithm which implements the simulation model for the considered example. The adaptation to a different setup is trivial. The expression $S_q^{-1}(\cdot)$ is used to denote the inverse of a marginal survival function $S_q(\cdot)$. The variables $\mathcal{U}_q$ are assumed to be i.i.d. uniform on $[0, 1]$.

1 ▶ Generate a value for $T_1$ from its marginal survival function

$$T_1 = S_1^{-1}(\mathcal{U}_1). \tag{4.9}$$

2 ▶ Conditionally on $T_1 = t_1$, generate $T_2$ from (4.5) with $k = 2$:

$$\begin{aligned} T_2 \mid t_1 &= S_{2|1}^{-1}(\mathcal{U}_2 \mid t_1) \\ &= S_2^{-1}\left(\left\{\left[\mathcal{U}_2^{-\frac{\vartheta}{1+\vartheta}} - 1\right] S_1(t_1)^{-\vartheta} + 1\right\}^{-1/\vartheta}\right). \end{aligned} \tag{4.10}$$

3 ▶ Conditionally on $T_1 = t_1, T_2 = t_2$, generate $T_3$ from (4.5) with $k = 3$:

$$\begin{aligned} T_3 \mid t_1, t_2 &= S_{3|12}^{-1}(\mathcal{U}_3 \mid t_1, t_2) \\ &= S_3^{-1}\left(\left\{\left[\mathcal{U}_3^{-\frac{\vartheta}{1+2\vartheta}} - 1\right]\left[S_1(t_1)^{-\vartheta} + S_2(t_2)^{-\vartheta} - 1\right] + 1\right\}^{-1/\vartheta}\right). \end{aligned} \tag{4.11}$$

4 ▶ Generate a censoring time $T_{C_0}$ from an arbitrary distribution $F_{C_0}(\cdot)$ and compute $Y_0 = \min(T_{C_0}, T_1, T_2, T_3)$ which is the observed time of censoring or transition from state NED; it will also give the information on the possible arrival state.

5 ▶ If transition into state LR or DM is observed, generate $T_4$ or $T_5$ respectively, from (4.6) or

(4.7):

$$
\begin{aligned}
T_4 \mid (y_0 = t_1) &= S_{4|1}^{-1}(\mathcal{U}_4 \mid t_1) \\
&= S_4^{-1}\left(\left\{\left[\mathcal{U}_4^{-\frac{\vartheta}{1+\vartheta}} - 1\right] S_1(t_1)^{-\vartheta} + 1\right\}^{-1/\vartheta}\right), \qquad (4.12) \\
T_5 \mid (y_0 = t_2) &= S_{5|2}^{-1}(\mathcal{U}_5 \mid t_2) \\
&= S_5^{-1}\left(\left\{\left[\mathcal{U}_5^{-\frac{\vartheta}{1+\vartheta}} - 1\right] S_2(t_2)^{-\vartheta} + 1\right\}^{-1/\vartheta}\right), \qquad (4.13)
\end{aligned}
$$

a censoring time $T_{C_1}$ or $T_{C_2}$ from an arbitrary distribution $F_{C_1}(\cdot)$ or $F_{C_2}(\cdot)$ and compute $Y_1 = \min(T_{C_1}, T_4)$ or $Y_2 = \min(T_{C_2}, T_5)$. Note that if the *'clock forward'* approach is adopted, it is sufficient to use $T_{q'} \mid t_q = S_{q'|q}^{-1}(\mathcal{U}_{q'} S_{q'|q}(t_q \mid t_q) \mid t_q)$ instead of $T_{q'} \mid t_q = S_{q'|q}^{-1}(\mathcal{U}_{q'} \mid t_q)$.

### 4.1.2 Frailty terms and covariates

The proposed model can easily accommodate the effect of clustering and of simulated covariates, in a proportional-hazards way.

Let $\boldsymbol{U} = (U_1, \ldots, U_Q)^\top$ be the vector of the frailty terms for the $Q$ transitions. Let $F_{\boldsymbol{U}}(\boldsymbol{u}; \boldsymbol{\theta})$ be its multivariate distribution. Let finally $\mathbf{X} \sim F_{\mathbf{X}}(\mathbf{x})$ be the (possibly multidimensional) covariate. For each time variable $T_q$ a parametric form is chosen for the baseline hazard $\lambda_{q0}(t)$ corresponding to $\mathbf{x} = 0$ and $u_q = 1$; the associated baseline survival function is $S_{q0}(t) = \exp\left\{-\int_0^t \lambda_{q0}(u)\mathrm{d}u\right\}$. The conditional cumulative hazard, given the values of $U_q$ and $\mathbf{X}$, is $\Lambda_q(t \mid u_q, \mathbf{x}) = u_q \exp\left\{\beta_q^\top \mathbf{x}\right\} \int_0^t \lambda_0(u)\mathrm{d}u$, with $\beta_q$ the transition-specific vector of the regression coefficients. Then, conditional on $(U_q = u_q, \mathbf{X} = \mathbf{x})$, the marginal survival function of each $T_q$ is

$$
S_q(t \mid u_q, \mathbf{x}) = \exp\left\{-\Lambda_q(t \mid u_q, \mathbf{x})\right\} = S_{q0}(t)^{u_q \exp\left\{\beta_q^\top \mathbf{x}\right\}}. \qquad (4.14)
$$

Note that, unlike in the frailty literature, in this context "marginal" has to be understood with respect to the set of all the transition times and not to the frailty term.

The conditional distributions (4.2) are unchanged, except that in the final expressions (like (4.5)–(4.7) of the example) each $S_q(t_q)$ is replaced by $S_q(t_q \mid u_q, \mathbf{x})$. As a consequence, in the case of clustering and/or covariates, the algorithm in Section 4.1.1 can be used conditionally on previously simulated $\mathbf{X}$ and $\boldsymbol{U}$ without need of modifications.

## 4.2 Tuning simulation parameters

When the researcher designs a simulation study, he typically wants to reproduce some precise situations. In the context of multi-state data, what one would like to control are the probabilities of competing events and some location measure of transition times. The median would be a natural choice, but it cannot be estimated in the case that too many observations are censored. For this reason we use the median of uncensored times, always observable, but any other location measure can be used without consequences on the rest of the procedure.

The quantities to control depend on the choice of $\Pi$, the set of the parameters of the marginal

distributions of transition and censoring times. Since no general way exists to express them as a function of the target values, a trial-and-error procedure is needed. Let $p_q$ and $m_q$ $(q = 1, \ldots, Q)$ be the target values of the probabilities of competing events and of the median of uncensored times of each transition. Let $\hat{p}_q(\Pi)$ and $\hat{m}_q(\Pi)$ be their observed values in a dataset simulated with parameters $\Pi$. The censoring probability is not considered, as it directly follows from the other probabilities as one minus their sum.

The parameters in $\Pi$ must be chosen in such a way that the ratios $p_q/\hat{p}_q(\Pi)$ and $m_q/\hat{m}_q(\Pi)$ $(q = 1, \ldots, Q)$ are as close as possible to 1, that is equivalent to require that the quantities $\left[\log\left(p_q/\hat{p}_q(\Pi)\right)\right]^2$ and $\left[\log\left(m_q/\hat{m}_q(\Pi)\right)\right]^2$ are as small as possible. This transformation has the good property that, for any arbitrary $k$, the values of the parameters such that $\hat{p}_q(\Pi) = p_q \times k$ and $\hat{p}_q(\Pi) = p_q/k$ (and analogously with $\hat{m}_q(\Pi)$ and $m_q$) give the same contribution $(\log k)^2$, i.e. they are considered as deviations of the same amplitude from the target. Therefore we propose the criterion function

$$\Upsilon(\Pi) = \sum_{q=1}^{Q} \left\{ \left[\log \frac{p_q}{\hat{p}_q(\Pi)}\right]^2 + \left[\log \frac{m_q}{\hat{m}_q(\Pi)}\right]^2 \right\} \tag{4.15}$$

to be minimized over the space of $\Pi$. Note that the criterion function (4.15) is the sum of non-negative terms, and each term is a function of only the parameters of one competing events block:

$$\begin{aligned} \Upsilon(\Pi) &= \sum_{\mathcal{Q}_k} \sum_{q \in \mathcal{Q}_k} \left\{ \left[\log \frac{p_q}{\hat{p}_q(\Pi_{\mathcal{Q}_k})}\right]^2 + \left[\log \frac{m_q}{\hat{m}_q(\Pi_{\mathcal{Q}_k})}\right]^2 \right\} \\ &= \sum_{\mathcal{Q}_k} \Upsilon_{\mathcal{Q}_k}(\Pi_{\mathcal{Q}_k}), \end{aligned} \tag{4.16}$$

with $\mathcal{Q}_k$ the blocks. In the example of Figure 4.1 these are $\mathcal{Q}_0 = \{1, 2, 3\}$, $\mathcal{Q}_1 = \{4\}$ and $\mathcal{Q}_2 = \{5\}$. Thanks to decomposition (4.16), the optimization of $\Upsilon$ over $\Pi$ can be done by first minimizing $\Upsilon_{\mathcal{Q}_0}$ over $\Pi_{\mathcal{Q}_0}$ and then the following ones, conditionally on parameters chosen for the previous transitions. In the example of Section 4.3, for instance, the minimization over the space of $\Pi$, of size 13, can be split into subproblems of sizes 7, 3 and 3, which are much simpler.

The criterion function $\Upsilon$ uses empirical estimates $\hat{p}_q(\Pi)$'s and $\hat{m}_q(\Pi)$'s based on a random dataset simulated with the present candidate parameter values. Then, the minimization procedure deals with a function that is not deterministic. However, this is not really a problem as the interest is not in finding its exact minimum but only reasonable values for simulating data sufficiently similar to requirements.

## 4.3 Case study

We consider a dataset from a multicenter study concerning patients treated with radiotherapy for head and neck cancer in five Italian hospitals (Grillo Ruggieri *et al.*, 2005). The data contain the times to the occurrence, after the beginning of the therapy, of LR, DM and De. The multi-state structure is shown in Figure 4.4, together with the observed frequencies. It is clear that, even though the structure of these data is very interesting, the number of patients, 44, is a strong limitation. Nevertheless the information provided by these data can be valuable for correctly building a

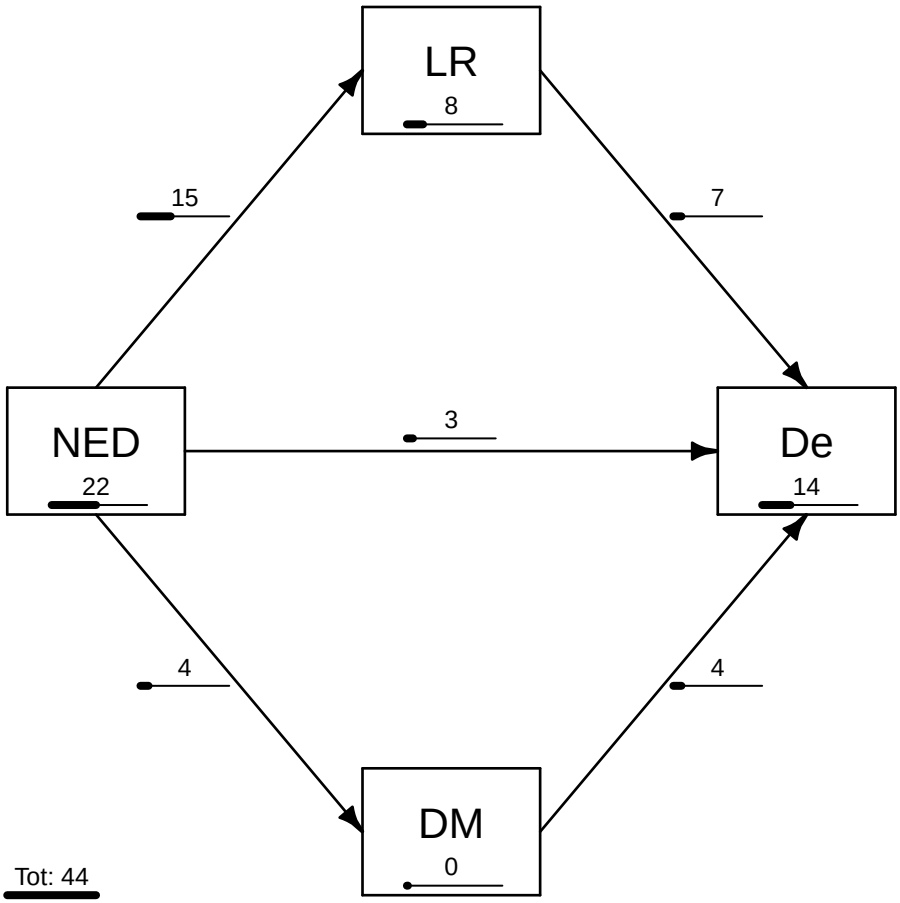



Figure 4.4 – States and transitions structure of real data. NED: No Evidence of new Disease, LR: Local Relapse, DM: Distant Metastases, De: Dead. The numbers on the arrows are the frequencies of each transition; those in the boxes are the numbers of patients ending the study in each state.

simulation study. Observed frequencies of competing events and median times of uncensored transitions can serve as benchmark to generate realistic data.

We want to generate a dataset of size 3000, with dependence between patients belonging to the same hospital. The number of hospitals is fixed to 40, whereas their dimensions are randomly generated. A one-parameter gamma frailty at the level of the hospital is used, with dispersion parameter $\theta = 0.5$, corresponding to Kendall's $\tau = 0.2$, i.e. a moderately high dependence.

We choose Weibull marginals $\mathsf{Wei}(\lambda_q, \rho_q)$, each with its own shape $\rho_q$ and location $\lambda_q$ parameters. This assumption, which also includes the exponential case, is very common for parametric proportional hazards models and has the main advantage that the expressions of the conditional distributions (4.2) are particularly simple. Indeed, the expressions for variables (4.9)–(4.11) simplify to

$$\begin{aligned} T_1 &= \left[-\log(\mathcal{U}_1)/\lambda_1\right]^{1/\rho_1}, \\ T_2|t_1 &= \left(\log\left\{\left(\mathcal{U}_2^{-1/2}-1\right)\exp\{\lambda_1 t_1^{\rho_1}\}+1\right\}\lambda_2\right)^{1/\rho_2}, \\ T_3|t_1,t_2 &= \left(\log\left\{\left(\mathcal{U}_3^{-1/3}-1\right)\left(\exp\{\lambda_1 t_1^{\rho_1}\}+\exp\{\lambda_2 t_2^{\rho_2}\}-1\right)+1\right\}\lambda_3\right)^{1/\rho_3}, \end{aligned} \tag{4.17}$$

and those for variables (4.12)–(4.13) are analogous to (4.17). Clustering is simulated by means of an exponential variable for each competing events block $\mathsf{Exp}(\lambda_{C_k})$.

Two covariates are simulated: a dichotomous one, $\texttt{Treat} \sim \mathsf{Bin}(0.5)$, and a continuous one, $\texttt{Age} \sim \mathsf{N}(60, 7)$. Regression coefficients are fixed to $\log(0.3)$ ($q = 1$), 0 ($q = 2$) and $\log(1.2)$ ($q = 3, 4, 5$) for `Treat` and $\log(0.8)/10$ ($q = 1$), $\log(0.9)/10$ ($q = 2$) and $\log(1.2)/10$ ($q = 3, 4, 5$) for `Age`. These values represent a treatment, as can be the case of radiotherapy, which reduces by 70% the hazard of LR, increases by 20% the risk of De and does not affect that of DM. Concerning the age, 10 more years of age reduce by 20% and 10% the hazard of LR and DM respectively and increase by 20% the risk of De.

Based on the real dataset, we want to simulate data with the following features. Starting from state NED: 50% of probability of censoring, 34% of LR occurrence, 9% of DM occurrence and 7% of dying; median uncensored times for LR, DM and De of 6, 10 and 3 months respectively. Starting from state LR: 53% of probability of censoring and 47% of dying; median uncensored De times of 3.25 additional months. Starting from state DM: 5% of probability of censoring and 95% of dying; median uncensored De times of 0.5 additional month. In the real data, the patients experiencing DM are only 4 and all of them die during the study. Nevertheless, a patient is not necessarily prevented from having the De time censored after DM, so the frequency 0 for transition DM $\to$ De is only accidental and its risk has to be considered strictly positive.

**Results.** The minimization of $\Upsilon$ should be done over the space of

$$\Pi = \{\lambda_{C_k}\}_{k=1,2,3} \cup \{\lambda_q, \rho_q, q \in \mathcal{Q}_k\}_{k=1,2,3}, \tag{4.18}$$

of size 13, but it can be decomposed into subproblems concerning $\Upsilon_{\mathcal{Q}_0}$, $\Upsilon_{\mathcal{Q}_1}$ and $\Upsilon_{\mathcal{Q}_2}$ of sizes 7, 3 and 3 respectively, which are much more affordable. First, the parameters of the transitions times $\{T_q\}_{q\in\mathcal{Q}_0}$, $\mathcal{Q}_0 = \{1, 2, 3\}$, are chosen by minimizing $\Upsilon_{\mathcal{Q}_0}$, conditionally on the frailties and the covariates. Each substep of the tuning procedure is iterated 10000 times with datasets of size 10000.

| | $T_1$ | $T_2$ | $T_3$ | $C_0$ | $T_4$ | $C_1$ | $T_5$ | $C_2$ |
|---|---|---|---|---|---|---|---|---|
| **Location** $\lambda$ | 0.276 | 0.019 | 0.013 | 0.031 | 0.029 | 0.099 | 0.192 | 0.039 |
| **Shape** $\rho$ | 0.851 | 1.076 | 0.569 | — | 1.078 | — | 1.000 | — |

Table 4.1 – Values of simulation parameters chosen by the tuning procedure.

| | | $p_i$ | | | | $m_i$ | | |
|---|---|---|---|---|---|---|---|---|
| | | LR | DM | De | C | LR | DM | De |
| **Target** | | 0.34 | 0.09 | 0.07 | 0.50 | 6.00 | 10.00 | 3.00 |
| **Simulated** size 3000 | 2.5% | 0.26 | 0.09 | 0.07 | 0.44 | 5.31 | 7.45 | 1.93 |
| | 50% | 0.31 | 0.11 | 0.08 | 0.50 | 6.32 | 9.19 | 2.67 |
| | 97.5% | 0.35 | 0.12 | 0.10 | 0.57 | 7.53 | 11.26 | 3.69 |
| **Simulated** size 300 | 2.5% | 0.24 | 0.07 | 0.05 | 0.42 | 4.21 | 5.07 | 0.95 |
| | 50% | 0.31 | 0.11 | 0.08 | 0.50 | 6.32 | 9.19 | 2.69 |
| | 97.5% | 0.37 | 0.15 | 0.12 | 0.59 | 9.28 | 15.65 | 6.26 |

Table 4.2 – Values of the median of the uncensored times and of the probabilities of competing events. First competing events block $\mathcal{Q}_0 = \{T_1, T_2, T_3\}$. Target values and 2.5th, 50th and 97.5th percentiles of values observed in 10000 datasets of sizes 3000 and 300.

In our case, the procedure takes several hours. Then, for simulated patients passed to LR or DM, the two following degenerate blocks $\mathcal{Q}_1 = \{T_4\}$ and $\mathcal{Q}_2 = \{T_5\}$ are considered for optimization of $\Upsilon_{\mathcal{Q}_1}$ and $\Upsilon_{\mathcal{Q}_2}$, respectively. This tuning procedure takes several hours, but has of course to be run only once for each setting. Table 4.1 shows the chosen values for the parameters.

To provide a comparison between target values and the ones observed in datasets simulated with chosen parameters, we show them together with boxplots containing 95% of the values within 10000 datasets (Fig. 4.5, Tab. 4.2–4.4). We consider two scenarios, one with datasets of size 3000 and one of size 300.

Initial values of the parameters were fixed to 1; further simulations (not presented here), with initial values ranging from $\exp(-2) \approx 0.14$ to $\exp(2) \approx 7.39$, showed that these results are quite

| | | $p_i$ | | $m_i$ |
|---|---|---|---|---|
| | | De | C | De |
| **Target** | | 0.47 | 0.53 | 3.25 |
| **Simulated** size 3000 | 2.5% | 0.41 | 0.50 | 3.04 |
| | 50% | 0.45 | 0.55 | 3.57 |
| | 97.5% | 0.50 | 0.59 | 4.21 |
| **Simulated** size 300 | 2.5% | 0.35 | 0.44 | 2.22 |
| | 50% | 0.45 | 0.55 | 3.57 |
| | 97.5% | 0.56 | 0.65 | 5.46 |

Table 4.3 – Values of the median of the uncensored times and of the probabilities of competing events. Degenerate competing events block $\mathcal{Q}_1 = \{T_4\}$. Target values and 2.5th, 50th and 97.5th percentiles of values observed in 10000 datasets of sizes 3000 and 300.

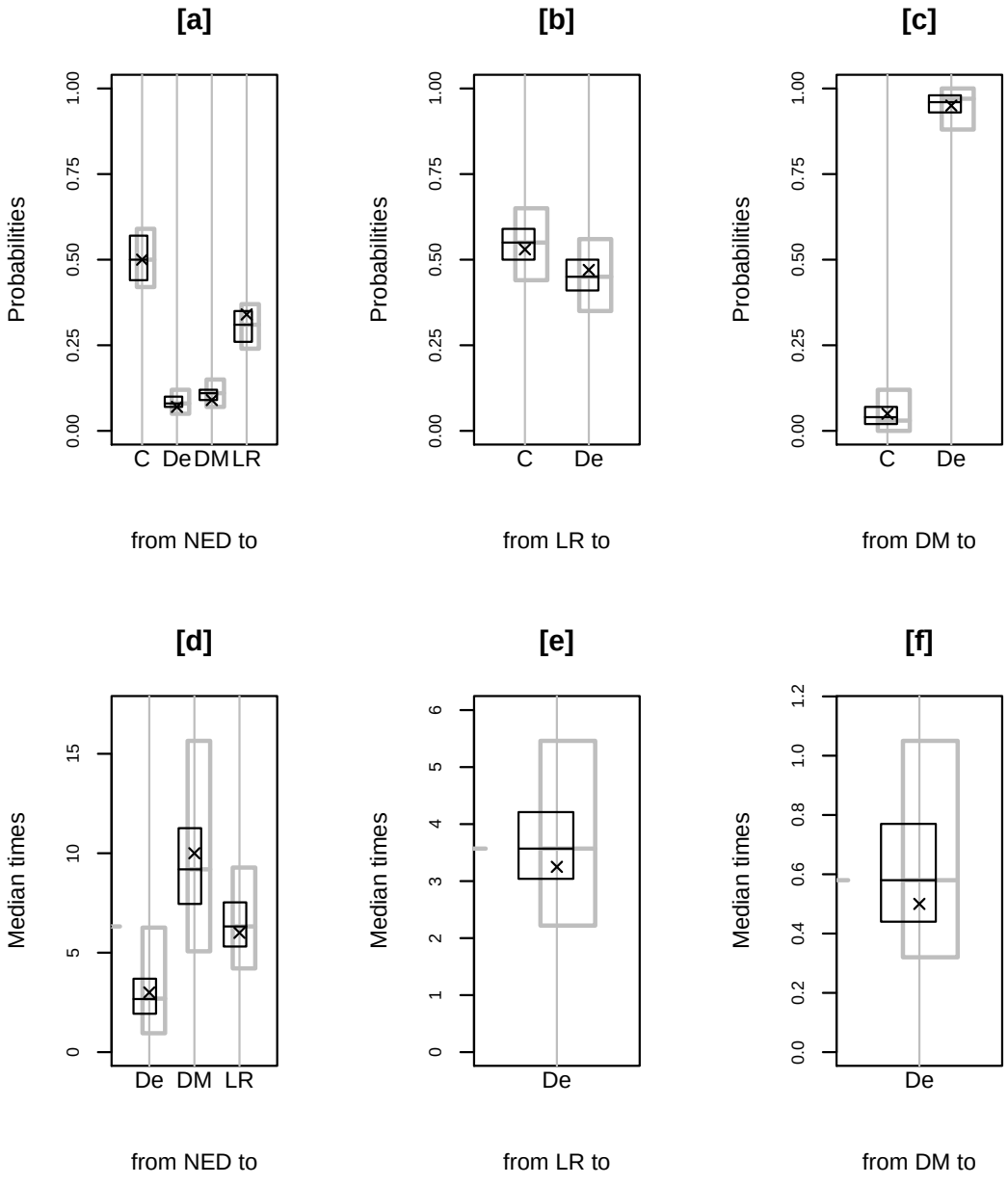



Figure 4.5 – Probabilities of competing events ([a]–[c]) and median uncensored times ([d]–[f]) for transitions from state NED ([a], [d]), LR ([b], [e]) and DM ([c], [f]). Comparison between target values ($\times$) and boxplots containing the 95% of the data over $10^4$ simulations. Black boxplots are referred to datasets of size 3000, grey ones to datasets of size 300.

| | | $p_i$ | | $m_i$ |
|---|---|---|---|---|
| | | De | C | De |
| **Target** | | 0.95 | 0.05 | 0.50 |
| **Simulated** size 3000 | 2.5% | 0.93 | 0.02 | 0.44 |
| | 50% | 0.96 | 0.04 | 0.58 |
| | 97.5% | 0.98 | 0.07 | 0.77 |
| **Simulated** size 300 | 2.5% | 0.88 | 0.00 | 0.32 |
| | 50% | 0.97 | 0.03 | 0.58 |
| | 97.5% | 1.00 | 0.12 | 1.05 |

Table 4.4 – Values of the median of the uncensored times and of the probabilities of competing events. Degenerate competing events block $\mathcal{Q}_2 = \{T_5\}$. Target values and 2.5th, 50th and 97.5th percentiles of values observed in 10000 datasets of sizes 3000 and 300.

robust with respect to the initial values of the parameters.

# Chapter 5

# Frailty multi-state models

The inclusion of frailties into multi-state models can provide complex survival models accounting for dependence between grouped subjects as well as between times to events of different types within the same group. For this reason, despite the challenge is quite complex, some works addressing this problem have begun to appear in recent years in applied statistics, while investigation of theoretical aspects is still moving its first steps.

Bhattacharyya and Klein (2005), for instance, considered progressive multi-state models with exponential baselines. They introduced frailties correlated within subjects, obtained by summing independent gamma random variables as suggested by Yashin *et al.* (1995) for correlated frailty models.

Putter and van Houwelingen (2011) critically discussed some aspects of more general multi-state models with dependent frailties within subjects, again. These models account for association between transition intensities of the same subject, while they consider event times of different subjects as independent. So, these models are more in the spirit of univariate frailty models, each subject having a different risk level due to his own unobserved factors. In such a context the bigger the frailty variances, the higher the heterogeneity between subjects and the dependence between event times of different types for each subject. No clustering effect can be accounted for by this approach.

On the other hand, Yen *et al.* (2010) studied models with gamma or binomial frailties, independent between transitions but shared by grouped subjects. Their application to large bowel adenoma–carcinoma natural history is focused on transition probabilities. In the same year, Ma *et al.* (2010) presented in a working paper a similar model with gamma frailties, for time-use pattern analysis, explicitly formulated in terms of transition-specific hazards. Both the models assume frailties which are independent for different transitions, but shared by groups of subjects: these assumptions are more in the spirit of shared frailty models, as they can account for clustering, even though not for dependence between transitions.

Both of these approaches — the shared-like and univariate-like frailty multi-state models — are used with quasi-semiparametric inference and prediction by Liquet *et al.* (2012), who show an application to intensive care units.

These two approaches move from two assumptions which can hold, but which seem quite strong in general: the independence between times of different subjects for the univariate-like models and

the independence between times of different transitions for the shared-like ones. Here, we propose the use of correlated frailties which can account for both between-subjects and between-transitions dependence. This approach extends and includes as limit cases the two above-mentioned ones.

First, the parametric estimation approach based on the maximization of the marginal likelihood is presented in Section 5.2.1. Quasi-semiparametric estimation, based on maximum penalized likelihood, raises computational problems which are briefly discussed. Then, in Section 5.2.2 we propose and discuss semiparametric estimation based on the EM-PL algorithm (Horny, 2009) for the maximisation of the penalized partial likelihood.

## 5.1 Nested frailty multi-state models

The addition of frailties into Markov multi-state models can account for heterogeneity across groups in the risk levels of different kinds of transition. In such cases, it is sensible to assume that, due to unobserved common risk factors, the risk for subjects varies across groups. Furthermore, we expect that variations for different transition types within the same group are different but somehow dependent. Indeed, if patients in a given group perform better or worse than the global population, they will do so with respect to all types of events, on average.

To model this kind of dependence, we borrow ideas coming from correlated frailty models and we exploit them in a multi-state framework. We define the frailty multi-state model by means of the **conditional transition-specific hazard**

$$\lambda_{qhi}(t \mid u_{qh}) = \lambda_{q0}(t) u_{qh} \exp\left\{\boldsymbol{\beta}_q^\top \mathbf{x}_{qhi}\right\}, \tag{5.1}$$

for transition of type $q = 1, \ldots, Q$ of subject $i = 1, \ldots, n_h$ in group $h = 1, \ldots, H$. In Equation 5.1, $\lambda_{q0}(t)$ is the transition-specific baseline hazard, $\boldsymbol{\beta}_q$ the vector of transition-specific regression parameters, and $\mathbf{x}_{qhi}$ the vector of transition-specific covariates.

In order to obtain frailties which are different but correlated across transitions in the same cluster, the use of multiplicative nested frailties (Sastry, 1997) is chosen, thanks to its analytical flexibility and straightforward interpretation. Hence, the frailty term $u_{qh}$ for transitions of type $q$ of subjects in cluster $h$ is assumed to be the unobservable realisation of the random variable

$$U_{qh} = V_h W_{qh}, \tag{5.2}$$

with

$$\begin{array}{rcll} V_h & \overset{iid}{\sim} & f_V(v;\theta_V), & h = 1, \ldots, H, \\ W_{qh} & \overset{iid}{\sim} & f_W(w;\theta_q), & q = 1, \ldots, Q, \quad h = 1, \ldots, H, \\ V_h & \perp\!\!\!\perp & W_{qh}, & \forall (h,q). \end{array} \tag{5.3}$$

Then, the model given by equations (5.1)–(5.3) has a random variable, $V_h$, which accounts for the global effect of the group and another one, $W_{qh}$, for that of the interaction between the transition type and the group.

Note that we speak of "nested frailties" just to be consistent with the terminology of the methodology used to obtain correlated frailties. In the original publications, the second-level frailties $W_{qh}$ are associated to a subsample of unities (families) contained in the first-level ones

(communities), to which are referred the terms $V_h$. The analogy with nested frailty multi-state models is essentially technical: the interactions $W_{qh}$ are a random sample just because the hospitals are a random sample, but the transitions within each hospitals are not randomly selected from a larger set.

In the context of all adverse (or all favourable) events — as in the example of the cancer model in Figure 3.1 — it is natural to assume positive dependence between cluster–transition frailties $U_{qh}$ within the same cluster, and independence across clusters. For frailties associated to different groups we have that

$$U_{qh} \perp\!\!\!\perp U_{q'h'}, \qquad h \neq h', \tag{5.4}$$

trivially follows from assumptions (5.3). On the other hand, if we consider the Pearson correlation coefficient as measure of (linear) dependence, then we have that within a given cluster

$$\mathsf{Cor}(U_{qh}, U_{q'h'}) \in [0, 1], \qquad h = h' \tag{5.5}$$

(App. A.6).

The limit cases of these assumptions are two interesting particular cases. When $V_h \equiv 1, \forall h$, the model reduces to that with $\mathsf{Cor}(U_{qh}, U_{q'h}) = 0$, similar to those by Yen *et al.* (2010) and Ma *et al.* (2010), where $U_{qh} \perp\!\!\!\perp U_{q'h}$. This assumption of independent frailties within clusters seems to be too strong in general because it ignores the interrelation of different events influenced by the same latent factors.

On the contrary, shared frailties are obtained if the $W_{qh}$ are degenerated random variables $W_{qh} \equiv 1$, $\forall q$, giving $U_{qh} = U_{q'h} = V_h$, i.e. $\mathsf{Cor}(U_{qh}, U_{q'h}) = 1$. The univariate version of this model has been considered by Putter and van Houwelingen (2011) to study the effect of subject-specific unobserved risk factors on multi-state models inference and prediction. Nevertheless assuming that the effect of unobserved risk factors is exactly the same on the risk of different types of events seems to be a little bit too rigid and correlated frailties are a reasonable relaxation of this constraint.

The use of correlated frailties can flexibly model situations which are between, and have as limit cases, the models with shared and independent frailties. Moreover, the use of nested frailties allows to obtain positive dependence, while dealing with independent random variables, which permits easier computations for the following inference methods.

## 5.2 Inference

**Notation.** We consider data collected from $n$ subjects grouped into $H$ clusters of sizes $n_1, \ldots, n_H$, with $\sum_{h=1}^{H} n_h = n$. For each subject, $Q$ possible types of transitions are possible, according to the multi-state structure. The type-$q$ transition time $t_{qhi}$ for subject $i$ in group $h$ can be censored at time $c_{qhi}$. So, the observable data are the event or censoring times $y_{qhi} = \min(t_{qhi}, c_{qhi})$ and the event/censoring indicators $\delta_{qhi} = \mathbb{1}(t_{qhi} \leq c_{qhi})$. As a result of the multi-state nature of the problem, subjects are at risk of each type of event only since the time at which they enter the associated starting state. These are recorded as left-truncation times $\tau_{qhi}$, corresponding to the time of the previous transition. If a subject has never been at risk for a given transition type, the left truncation time is set to infinity.

The frailty distributions $f_V(\cdot; \theta_V)$ and $f_W(\cdot; \theta_q)$ in (5.3) are assumed to be one-parameter

densities on the positive real line, with expectation fixed to 1 for identifiability reasons. The vector of all the frailty parameters is denoted by $\boldsymbol{\theta} = (\theta_V, \theta_1, \ldots, \theta_Q)^\top$. Analogously, the stacked vector of all the regression parameters is denoted by $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_Q^\top\right)^\top$, the vector of the baseline hazard functions by $\boldsymbol{\lambda}_0(\cdot) = \left(\lambda_{10}(\cdot), \ldots, \lambda_{Q0}(\cdot)\right)^\top$ and that of their parameters by $\boldsymbol{\xi} = \left(\boldsymbol{\xi}_1^\top, \ldots, \boldsymbol{\xi}_Q^\top\right)^\top$. Finally, let $\Lambda_{q0}(t) = \int_0^t \lambda_{q0}(s)\mathrm{d}s$ be the cumulative baseline hazard function for transitions of type $q$.

**Likelihood.** Thanks to the conditional independence given by assumptions (5.3), the conditional likelihood for the nested frailty multi-state model (5.1)–(5.2) can be obtained as the product over the groups and over the subjects of the contributions of only the transitions for which the subject has ever been at risk, i.e. those with finite left truncation time $\tau_{qhi} < \infty$:

$$\begin{aligned}
L_{\mathsf{C}}(\boldsymbol{\xi}, \boldsymbol{\beta}) &= \prod_{h=1}^{H} L_{\mathsf{C},h}(\boldsymbol{\xi}, \boldsymbol{\beta}) \\
&= \prod_{h=1}^{H} \prod_{i=1}^{n_h} \prod_{q=1}^{Q} \left[ \left\{ \lambda_{qhi}(y_{qhi} \mid v_h w_{qh}) \right\}^{\delta_{qhi}} \frac{S(y_{qhi} \mid v_h w_{qh})}{S(\tau_{qhi} \mid v_h w_{qh})} \right]^{\mathbb{1}(\tau_{qhi} < \infty)} \\
&= \prod_{h=1}^{H} \prod_{i=1}^{n_h} \prod_{q=1}^{Q} \left[ \left\{ v_h w_{qh} \lambda_{q0}(y_{qhi}) e^{\boldsymbol{\beta}_q^\top \mathbf{x}_{qhi}} \right\}^{\delta_{qhi}} \right. \\
&\qquad \left. \exp\left\{ -v_h w_{qh} \left( \Lambda_{q0}(y_{qhi}) - \Lambda_{q0}(\tau_{qhi}) \right) e^{\boldsymbol{\beta}_q^\top \mathbf{x}_{qhi}} \right\} \right]^{\mathbb{1}(\tau_{qhi} < \infty)} . \qquad (5.6)
\end{aligned}$$

The conditional likelihood cannot be maximised because it is function of unobservable quantities, the frailty terms. In analogy to most common practices in frailty modelling, we propose two estimation approaches to tackle this problem: a fully parametric one which maximises the marginal likelihood and a semiparametric one based on maximum pensalised partial likelihood.

### 5.2.1 Parametric inference

In a parametric context, the unobserved frailties can be integrated out of the conditional likelihood (5.6), so that the marginal likelihood of the observed data is obtained. This is a sort of "averaged" likelihood over the frailty distribution, this last incorporating all the information linked to unobserved quantities:

$$\begin{aligned}
L_{\mathsf{M}}(\boldsymbol{\zeta}) = \int_{\mathbb{R}_+^{H(Q+1)}} & L_{\mathsf{C}}(\boldsymbol{\beta}, \boldsymbol{\xi}) f_{\boldsymbol{V},\boldsymbol{W}}\Big(\boldsymbol{v}, \boldsymbol{w}; \boldsymbol{\zeta} \mid \{Y_{qhi} > \tau_{qhi}\}_{qhi}\Big) \\
& \mathrm{d}w_{11} \cdots \mathrm{d}w_{Q1} \ \cdots \ \mathrm{d}w_{1H} \cdots \mathrm{d}w_{QH} \ \mathrm{d}v_1 \cdots \mathrm{d}v_H . \qquad (5.7)
\end{aligned}$$

Thanks to independence between frailties given by assumptions (5.3), marginalising the conditional likelihood with respect to the frailties can be done cluster by cluster $\left(L_{\mathsf{M}}(\boldsymbol{\zeta}) = \prod_{h=1}^{H} L_{\mathsf{M},h}(\boldsymbol{\zeta})\right)$,

and the contribution of each cluster is

$$L_{\mathsf{M},h}(\boldsymbol{\zeta}) = \int\limits_{\mathbb{R}_+^{Q+1}} L_{\mathsf{C},h}(\boldsymbol{\beta},\boldsymbol{\xi}) f_{V_h,\boldsymbol{W}_h}\Big(v_h,\boldsymbol{w}_h;\boldsymbol{\zeta} \mid \{Y_{qhi} > \tau_{qhi}\}_{qi}\Big)\mathrm{d}w_{1h}\cdots\mathrm{d}w_{Qh}\ \mathrm{d}v_h, \tag{5.8}$$

with $\boldsymbol{W}_h = (W_{1h},\ldots,W_{Qh})^\top$ the vector of the cluster-by-transition interactions for cluster $h$, and $\boldsymbol{w}_h$ the vector of their realisations. The joint conditional frailty distribution used in this integration is

$$\begin{aligned} f_{V_h,\boldsymbol{W}_h}\Big(v_h,\boldsymbol{w}_h;\boldsymbol{\zeta} \mid \{Y_{qhi} > \tau_{qhi}\}_{qi}\Big) = \\ \prod_{qi|\tau_{qhi}<\infty} \exp\Big\{-v_h w_{qh}\Lambda_{q0}(\tau_{qhi})e^{\boldsymbol{\beta}_q^\top \mathbf{x}_{qhi}}\Big\} f_{V_h,\boldsymbol{W}_h}\Big(v_h,\boldsymbol{w}_h;\boldsymbol{\theta}\Big) \\ \Bigg[\int\limits_{\mathbb{R}_+} \prod_{qi|\tau_{qhi}<\infty} \exp\Big\{-v_h w_{qh}\Lambda_{q0}(\tau_{qhi})e^{\boldsymbol{\beta}_q^\top \mathbf{x}_{qhi}}\Big\} \\ f_{V_h,\boldsymbol{W}_h}\Big(v_h,\boldsymbol{w}_h;\boldsymbol{\theta}\Big)\mathrm{d}w_{1h}\cdots\mathrm{d}w_{Qh}\ \mathrm{d}v_h\Bigg]^{-1} \end{aligned} \tag{5.9}$$

(see App. A.7), with

$$f_{V_h,\boldsymbol{W}_h}\Big(v_h,\boldsymbol{w}_h;\boldsymbol{\theta}\Big) = f_V(v_h;\theta_V)\prod_{q=1}^{Q} f_W(w_{qh};\theta_q) \tag{5.10}$$

the joint marginal frailty distribution, because of independence assumptions (5.3).

Hence, by Equations 5.6, 5.8 and following, the marginal loglikelihood can be expressed as

$$\begin{aligned} \ell_{\mathsf{M}}(\boldsymbol{\zeta}) = \sum_{h=1}^{H} \Bigg\{ & \sum_{qi|\tau_{qhi}<\infty} \delta_{qhi}\Big\{\log\lambda_{q0}(y_{qhi}) + \boldsymbol{\beta}_q^\top \mathbf{x}_{qhi}\Big\} \\ & + \log\int_0^\infty v_h^{d_h} \prod_{q\in\mathcal{Q}_h} (-1)^{d_{qh}} \mathcal{L}_q^{(d_{qh})}\left(v_h \sum_{i|\tau_{qhi}<\infty} \Lambda_{q0}(y_{qhi})e^{\boldsymbol{\beta}_q^\top \mathbf{x}_{qhi}}\right) f_V(v_h;\theta_V)\mathrm{d}v_h \\ & - \log\int_0^\infty \prod_{q\in\mathcal{Q}_h} \mathcal{L}_q\left(v_h \sum_{i|\tau_{qhi}<\infty} \Lambda_{q0}(\tau_{qhi})e^{\boldsymbol{\beta}_q^\top \mathbf{x}_{qhi}}\right) f_V(v_h;\theta_V)\mathrm{d}v_h \Bigg\} \end{aligned} \tag{5.11}$$

(App. A.8), with $d_{qh} = \sum_{i=1}^{n_h}\delta_{qhi}$ the number of events of type $q$ in cluster $h$, $d_h = \sum_{i=1}^{n_h}\sum_{q=1}^{Q}\delta_{qhi}$ the number of all events in cluster $h$, $\mathcal{L}_q^{(k)}(s)$ the $k$-th derivative with respect to $s$ of the Laplace transform (Eq. 3.52) and where $\mathcal{Q}_h = \Big\{q \mid \sum_{i=1}^{n_h}\mathbb{1}(\tau_{qhi}<\infty) > 0\Big\}$ is the set of transitions for which there exists in group $h$ at least one subject who has ever been at risk.

This loglikelihood can accommodate any frailty distribution with derivatives of the Laplace

transform available in an explicit form (cf. Sec. 3.3.4). If we make the standard choice of unit mean gamma distributions in (5.3), then the marginal loglikelihood (5.11) is

$$\begin{aligned}\ell_{\mathsf{M}}(\boldsymbol{\zeta}) = \sum_{h=1}^{H} \Bigg\{ & \sum_{qi|\tau_{qhi}<\infty} \delta_{qhi} \Big\{ \log \lambda_{q0}(y_{qhi}) + \boldsymbol{\beta}_q^\top \mathbf{x}_{qhi} \Big\} \\ & + \log \int_0^\infty \frac{v_h^{1/\theta_V + d_h - 1} \exp\left(-v_h/\theta_V\right)}{\prod_{q\in\mathcal{Q}_h} \left(1 + \theta_q v_h \sum_{i|\tau_{qhi}<\infty} \Lambda_{q0}(y_{qhi}) e^{\boldsymbol{\beta}_q^\top \mathbf{x}_{qhi}}\right)^{1/\theta_q + d_{qh}}} \mathrm{d}v_h \\ & - \log \int_0^\infty \frac{v_h^{1/\theta_V - 1} \exp\left(-v_h/\theta_V\right)}{\prod_{q\in\mathcal{Q}_h} \left(1 + \theta_q v_h \sum_{i|\tau_{qhi}<\infty} \Lambda_{q0}(\tau_{qhi}) e^{\boldsymbol{\beta}_q^\top \mathbf{x}_{qhi}}\right)^{1/\theta_q}} \mathrm{d}v_h \\ & + \sum_{q\in\mathcal{Q}_h} \left[ \mathbb{1}(d_{qh} > 1) \sum_{l=0}^{d_{qh}-1} \log(1 + l\theta_q) \right] \Bigg\} \end{aligned} \tag{5.12}$$

(App. A.9).

More generally, provided that an appropriate frailty distribution has been fixed, the marginal loglikelihood (5.11) can be maximised in a fully parametric way. To this aim a parametric form is assumed for the $Q$ baseline hazard functions so that the marginal loglikelihood $\ell_{\mathsf{M}}(\boldsymbol{\zeta})$ can be maximised over the parameters vector $\boldsymbol{\zeta} = \left(\boldsymbol{\xi}^\top, \boldsymbol{\beta}^\top, \boldsymbol{\theta}^\top\right)^\top$. The most common distributions for the baseline hazard are shown in Table 3.2: exponential, Weibull, Gompertz, lognormal and loglogistic.

In order to obtain the marginal loglikelihood (5.11), integration at two levels is needed. If integration with respect to the interactions $w_{qh}$ can be done analytically and is expressed through the derivatives of the Laplace transform of their distribution, this is no longer possible for the group frailties $v_h$. Another approach is thus needed to integrate them out, based on a numerical or analytical approximation. In our study, we use a saddlepoint approximation (Goutis and Casella, 1999) to compute integrals in Equation 5.12 (see App. A.10).

It is worth noting that the marginal loglikelihood with gamma distributed frailties (5.12) is analogous to that given by Rondeau *et al.* (2006, Eq. 2) in the context of nested frailty models for left-truncated (and right-censored) data, but with the addition of stratification. In their paper, the authors approximate the baseline hazard by means of cubic M-splines; this is equivalent to leave the baseline hazard unspecified and to use a measure of its roughness as a penalisation when maximising the marginal loglikelihood (see Sec. 3.3.3, pg. 29). This approach is much more flexible than the fully parametric one, but splines demand a considerable number of parameters. In addition, our model has stratification, i.e. a different baseline function for each transition type. The use of one spline for each of them would increase the number of parameters excessively, making estimation prohibitive.

From a practical point of view, the `frailtypack` package (Gonzalez *et al.*, 2012, v. 2.2-23) can fit nested frailty models in `R` via maximum penalised marginal likelihood. Parametric inference is possible, too, with Weibull baselines. It can manage stratification, but only with two levels, so it is not straightforwardly usable for multi-state models, to date. Nevertheless, authors are working to

raise up to three or five the accepted number of strata in future releases of the package. For the examples shown in this thesis, ad hoc `R` code has been produced.

### 5.2.2 Semiparametric inference

In a semiparametric context, employing the ideas by Cox (1972), the conditional likelihood can be profiled out with respect to the baseline hazards, obtaining the (conditional) partial likelihood. The conditional likelihood (5.6) can be seen as the full likelihood of a multi-state model where the logarithm of the frailties act as offsets. Then, profiling out the baseline hazards comes out to be the same as for multi-state models (Andersen *et al.*, 1993; de Wreede *et al.*, 2010) with log-frailties as offsets. Then, conditionally on the frailties, the partial loglikelihood is

$$\ell_{\mathsf{P}}(\boldsymbol{\beta}) = \sum_{h=1}^{H} \left\{ d_h \log v_h + \sum_{q=1}^{Q} \left[ d_{qh} \log w_{qh} + \sum_{i=1}^{n_h} \left( \delta_{qhi} \boldsymbol{\beta}_q^\top \mathbf{x}_{qhi} - \log \sum_{h'i' \in R_q(y_{qhi})} v_{h'} w_{h'i'} \exp\left\{ \boldsymbol{\beta}_q^\top \mathbf{x}_{qh'i'} \right\} \right) \right] \right\}, \quad (5.13)$$

with $R_q(y) = \left\{ (h', i') \mid y \in [\tau_{qh'i'}, y_{qh'i'}] \right\}$ the risk set for transitions of type $q$ at time $y$ (App. A.11). Consequently, the profiled version of the full loglikelihood $\ell(\boldsymbol{\zeta}) = \ell_{\mathsf{C}}(\boldsymbol{\xi}, \boldsymbol{\beta}) + \log f_{\mathbf{V},\boldsymbol{W}}(\mathbf{v}, \boldsymbol{w}; \boldsymbol{\theta})$ is

$$\ell_{\mathsf{PP}}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \ell_{\mathsf{P}}(\boldsymbol{\beta}) + \log f_{\mathbf{V},\boldsymbol{W}}(\mathbf{v}, \boldsymbol{w}; \boldsymbol{\theta}), \quad (5.14)$$

which we call penalised partial loglikelihood by analogy to the PPL approach for semiparametric frailty models (Sec. 3.3.3). Under the constraint that all the cluster-by-transition interactions $W_{qh}$ have the same variance ($\theta_W := \theta_1 = \cdots = \theta_Q$), the penalised partial likelihood (5.14) can be maximised thanks to the EM-PL algorithm proposed by Horny (2009) for multilevel frailty models (Rotolo and Legrand, 2012). The algorithm is shown in details in Section 7.2 and, for the nested frailty multi-state model, it is the following.

Expectation step

$\tilde{v}_h = \max_{v_h} \ell_{\mathsf{PP}}(\hat{\boldsymbol{\beta}}, \theta_V)$ with offsets $\log \tilde{\boldsymbol{w}}$

$\tilde{w}_{qh} = \max_{w_{qh}} \ell_{\mathsf{PP}}(\hat{\boldsymbol{\beta}}, \theta_W)$ with offsets $\log \tilde{\boldsymbol{v}}$

Maximisation step

$\hat{\boldsymbol{\beta}} = \max_{\boldsymbol{\beta}} \ell_{\mathsf{P}}(\boldsymbol{\beta}; \tilde{\boldsymbol{v}}, \tilde{\boldsymbol{w}})$

$\hat{\boldsymbol{\theta}} = \max_{\boldsymbol{\theta}} \log f_{\boldsymbol{V},\boldsymbol{W}}(\tilde{\boldsymbol{v}}, \tilde{\boldsymbol{w}}; \boldsymbol{\theta})$

The algorithm iterates until convergence between an E-step and an M-step. First, at the expectation step (the name is held for analogy with the classical EM algorithm) previsions for the frailties are obtained via BLUPs: for each of the two frailty variables, a semiparametric frailty model is fitted via maximum PPL, with offsets given by the log-frailties of the other variable and the linear predictors with fixed provisional $\hat{\boldsymbol{\beta}}$. Then, estimates for regression and frailty parameters are updated, by maximisation of the two factors of the penalised partial likelihood, given present guesses for the frailty terms.

As a PPL estimation method is used at the E-step, only lognormal and gamma distributions can be used for the frailties (Sec. 3.3.3).

## 5.3 Simulation study

We present here a simulation study to investigate how the incorporation of nested frailties into multi-state models can improve parameter estimation. Simulation of clustered multi-state data is done thanks to the model presented in Chapter 4.

The multi-state structure of Figure 4.1 is chosen and three main scenarios (Tab. 5.1) are reproduced, with different number of clusters of different sizes. In scenario **A** only 10 groups of size

| **Scenario** | **A** | **B** | **C** |
|---|---|---|---|
| **Number of clusters** | 10 | 20 | 40 |
| **Number of subjects per cluster** | 20 | 50 | 125 |
| **Total number of subjects** | 200 | 1000 | 5000 |

Table 5.1 – Simulation scenarios.

20 are considered. Scenario **B** depicts a more favourable framework, with 1000 subjects clustered into 20 groups of size 50. Finally, scenario **C** considers 40 groups of 125 subjects. For each scenario, simulations are replicated 500 times. The data, expressed in months, are simulated as coming from a study lasting fifteen years, five of which of recruiting: censoring times are simulated as uniform random variables between 120 months ($15 - 5 = 10$ years) and 180 months (15 years). The effect of a binary treatment is added, with strong effect on local relapse (LR) occurrence, mild effect on distant metastases (DM) hazard, no effect on direct death (De), and strong negative effect on death after adverse events (Tab. 5.2). These values are not necessarily all close to the reality, but allow to cover different situations: null coefficients and small and big effects in both directions.

| **Transition** | **HR** | $\beta_{\text{Treat}}$ |
|---|---|---|
| NED→LR | 0.25 | -1.39 |
| NED→DM | 0.9 | -0.11 |
| NED→De | 1 | 0 |
| LR→De | 3 | 1.10 |
| DM→De | 5 | 1.61 |

Table 5.2 – Hazard ratios (HR) and regression parameters ($\beta_{\text{Treat}} = \log \text{HR}$) chosen to simulate the treatment effect on different events.

Clustering is generated by means of two one-parameter gamma frailties: one ($V_h$) for the global hospital effect with small variance and one ($W_{qh}$) for hospital–by-transition interaction with greater heterogeneity, equal for all the transition types:

$$\begin{aligned} \theta_V &= 0.5, \\ \theta_W := \theta_1 = \ldots = \theta_5 &= 1. \end{aligned} \tag{5.15}$$

The total variability of the frailties (see Eq. 3.59) is then

$$\mathsf{V}[U_{qh}] = (\theta_V + 1)(\theta_W + 1) - 1 = 2. \tag{5.16}$$

Weibull baselines hazards are used with parameters (Tab. 5.3) chosen in such a way that realistic proportions of competing events are obtained. See, for instance, those for Scenario **C** in Table 5.4.

| $(q)$ | **Transition** | $\lambda_q$ | $\rho_q$ |
|---|---|---|---|
| (1) | NED→LR | $3.47\times10^{-2}$ | 0.8 |
| (2) | NED→DM | $1.00\times10^{-3}$ | 1.3 |
| (3) | NED→De | $3.33\times10^{-4}$ | 1.4 |
| (4) | LR→De | $6.67\times10^{-3}$ | 1.4 |
| (5) | DM→De | $3.33\times10^{-2}$ | 1.7 |

Table 5.3 – Baseline hazards parameters chosen for simulations.

| | **to** | | | |
|---|---|---|---|---|
| **from** | LR | DM | De | Cens |
| NED | 0.30–0.51 | 0.15–0.31 | 0.08–0.14 | 0.17–0.32 |
| LR | ——— | ——— | 0.76–0.96 | 0.04–0.24 |
| DM | ——— | ——— | 0.95–1.00 | 0.00–0.05 |

Table 5.4 – Relative frequencies (by line, i.e. by starting state) of transitions into possible arrival state under Markov assumption, Scenario **C**. Minima and maxima over 500 replications.

In order to investigate the robustness of models with respect to the Markov assumption, simulations are replicated adding further dependence between event times of each subject, by means of a Clayton copula (see Sec. 4.1). So, in addition to the situation of no further dependence ($\vartheta = 0$), simulations are performed with two degrees of dependence: $\vartheta = 0.5$ (Kendall's $\tau = 0.2$) and $\vartheta = 1.5$ (Kendall's $\tau \simeq 0.43$).

### 5.3.1 Parametric models

As the implementation of parametric models is computationally much more demanding than in the semiparametric approach, we consider in this Section only two families of models:

**MSM**: a multi-state model without any frailty term,
**NFM**: our multi-state model with nested frailties.

Due to long computation times, we compare these models only in the case that the Markov assumption is true ($\vartheta = 0$). Moreover, results in this Section 5.3.1 correspond to simulation parameters which are slightly different from those presented above, but substantially similar, and which correspond to earlier choices ($\beta_5 = \beta_4 = \beta_3 = 0$ and $\theta_W = 1.5$). We opted for also positive betas ($\beta_5$ and $\beta_4$) and smaller heterogeneity ($\theta_W$) in a second moment, but we decided not to replicate simulations for the parametric models with these small variations because of demand of computational resources and little difference in parameters.

Tables in Appendix B.3 (Sec. B.3.1) report the detailed results of the study, comparing the mean and the standard error of the parameter estimates, together with the mean estimated standard error. Figures 5.1 and 5.2 compare the true simulation parameters to boxplots containing the 95% of the estimates over the 500 replications, and their medians.

These results clearly show that regression parameters are correctly estimated, on average, but their variability across datasets is in general largely higher than in the case of multi-state models without any frailty. This fact suggests that the Laplace approximation employed to approximate integrals in the marginal loglikelihood (5.12) is not that precise, even though correct. Indeed it is supposed to be a valid approximation whenever $d_h$ is big, which is not always the case.

As for the frailty parameters (Fig. 5.2), the estimates are so widespread that the value given by the analysis of a single dataset cannot be really trusted, notably that for the group $\hat{\theta}_V$. Nevertheless, the comparison of the three scenarios suggests that if we could consider more and larger clusters—from a computational point of view and in real clinical trials—the estimation of frailty parameters would probably converge to acceptable values, even though more slowly than for regression parameters.

Concerning the baseline hazards, the shape parameters ($\rho_q$) are estimated with very small bias and great precision, even for small datasets (scenario **A**). The scale parameters ($\lambda_q$), instead, are often overestimated by nested frailty multi-state models and those models without frailties perform at best in this regard. Shared frailty models encounter serious computational problems with large datasets.

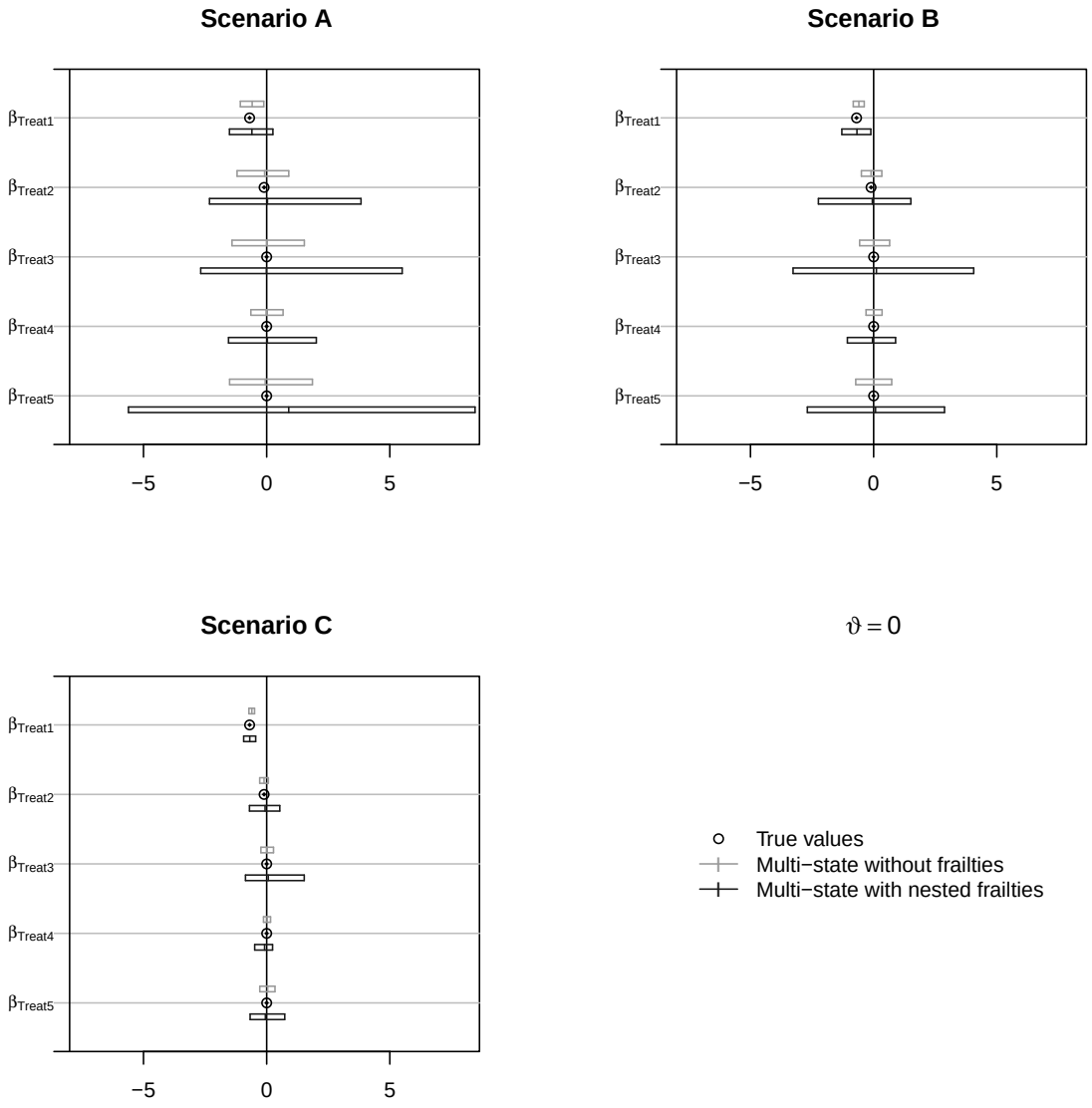



Figure 5.1 – Results of simulations for parametric models. Regression parameters ($\boldsymbol{\beta}$). Boxplots containing 95% of the values over 500 replications.

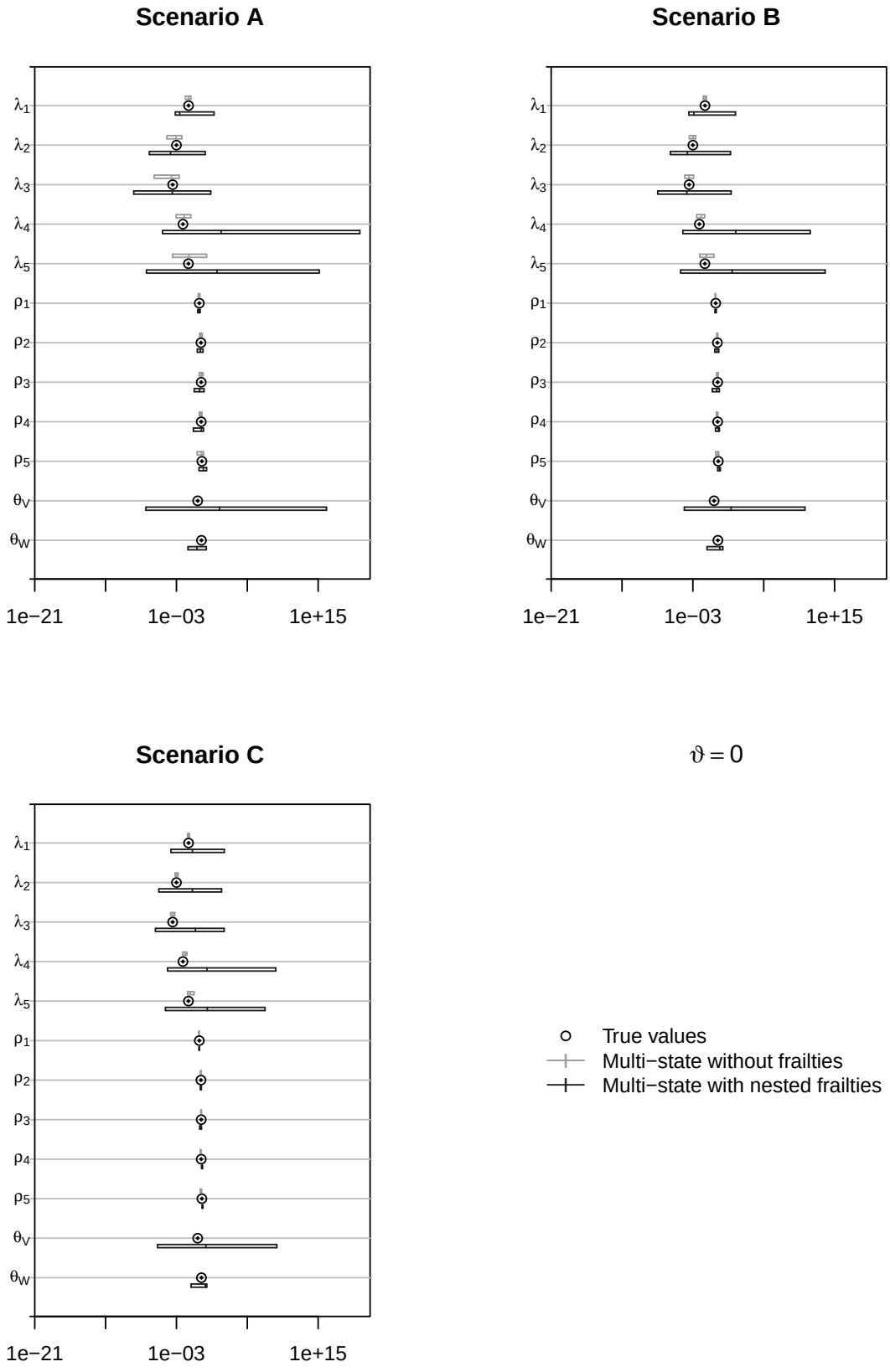



Figure 5.2 – Results of simulations for parametric models. Baseline ($\boldsymbol{\xi}$) and frailty ($\boldsymbol{\theta}$) parameters, on a $\log_{10}$ scale. Boxplots containing 95% of the values over 500 replications.

### 5.3.2 Semiparametric models

For each combination of scenarios (**A**, **B** and **C**) and extra-Markov dependence ($\vartheta \in \{0,\ 0.5,\ 1.5\}$), we evaluate the performance of the semiparametric nested frailty multi-state model by comparing three families of models:

**MSM**: a multi-state model without any frailty term,
**SFM**: a multi-state model with a shared frailty accounting for hospital effect,
**NFM**: our multi-state model with nested frailties.

The models are fitted on the 500 simulated datasets for each of the nine scenario–dependence combinations.

An example of the results is given in Table 5.5 for the most robust situation, i.e. with the largest number and size of groups (scenario **C**) and no departure from the Markov assumption ($\vartheta = 0$). The complete results are provided in Appendix B.3 (Sec. B.3.2). For each estimated parameter we provide the mean of the estimates (MEAN), empirical standard error (eSE), the mean estimated standard error (mSE). Figures 5.3 ($\vartheta = 0$), 5.4 ($\vartheta = 0.5$) and 5.5 ($\vartheta = 1.5$) show box-plots of the estimates over the 500 replications.

The first evidence is that in scenario **A** all the models perform quite poorly, whereas results are acceptable in scenario **B** and are really satisfactory in scenario **C** (note the different axis scale of the first scenario as compared to the two others).

A very good general result is that the violation of the Markov assumption does not affect results dramatically, most of them being comparable in all scenarios. This has a strong impact on multi-state modelling, as the Markov assumption is usually done just for mathematical convenience; indeed, many estimation and prevision methods are much simpler or exist only for Markov models (see Sec. 3.2).

Now, let us consider the estimation of regression parameters, which is typically the main interest in a clinical study. Multi-state models (**MSM**), which ignore the clustering, underestimate the treatment effect in all situations; this is coherent with previous results in survival analysis literature (Bretagnolle and Huber-Carol, 1988; Andersen *et al.*, 1999), showing that uncontrolled heterogeneity yields an attenuation of the covariates effect. The addition of nested frailties (**NFM**) proves to be able to reduce both the bias and the variability of regression parameter estimates, notably in the case of big datasets (Scenario **C**).

The coverage probabilities of the confidence intervals for regression parameters, obtained over the 500 replications, are showed in Figure 5.6. The well-known problem of underestimation of the variability of the estimators is reflected by the undercoverage from which all the considered methods suffer. Under the Markov assumption, the nested frailty model is uniformly the best performing one, with coverage probabilities very close to the nominal value. Nevertheless, this model seems to suffer more than the others from departures from the Markov assumption. A surprising result is given by the performance of all models and with all values of $\vartheta$ worse in scenario **C** than in scenario **A**, with intermediate behaviour in scenario **B**. A possible explanation might be related to underestimation of standard errors.

Estimation of the frailty parameters is more problematic. In general, the model with shared frailties is able to recover the cluster heterogeneity $\theta_V$ properly, but it cannot account for the

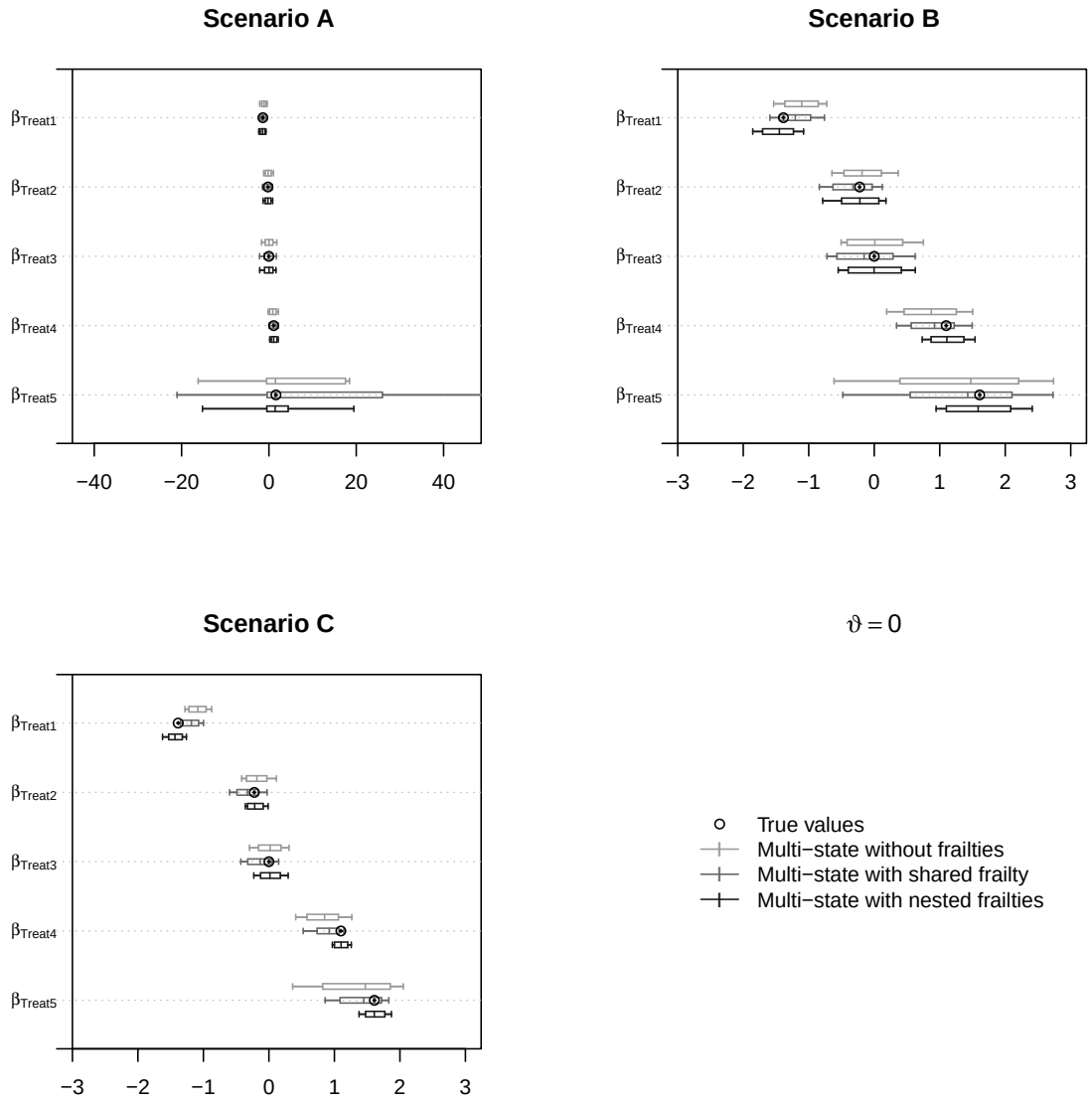



Figure 5.3 – Results of simulations for semiparametric models, with $\vartheta = 0$. Each box contains the 95% of the values over 500 replications; the vertical line within it is the median; the extreme ticks are the minimum and the maximum.

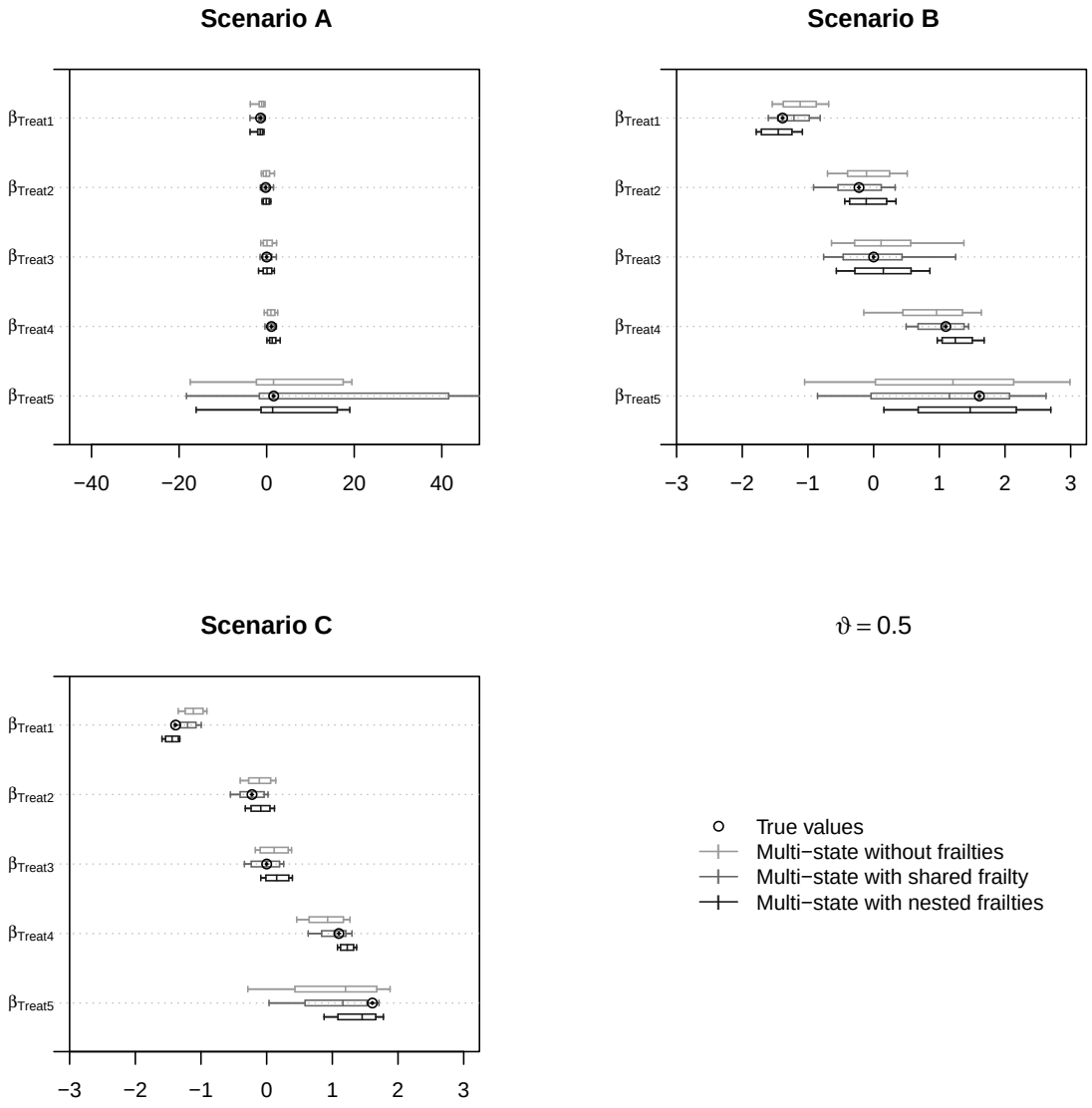



Figure 5.4 – Results of simulations for semiparametric models, with $\vartheta = 0.5$. Each box contains the 95% of the values over 500 replications; the vertical line within it is the median; the extreme ticks are the minimum and the maximum.

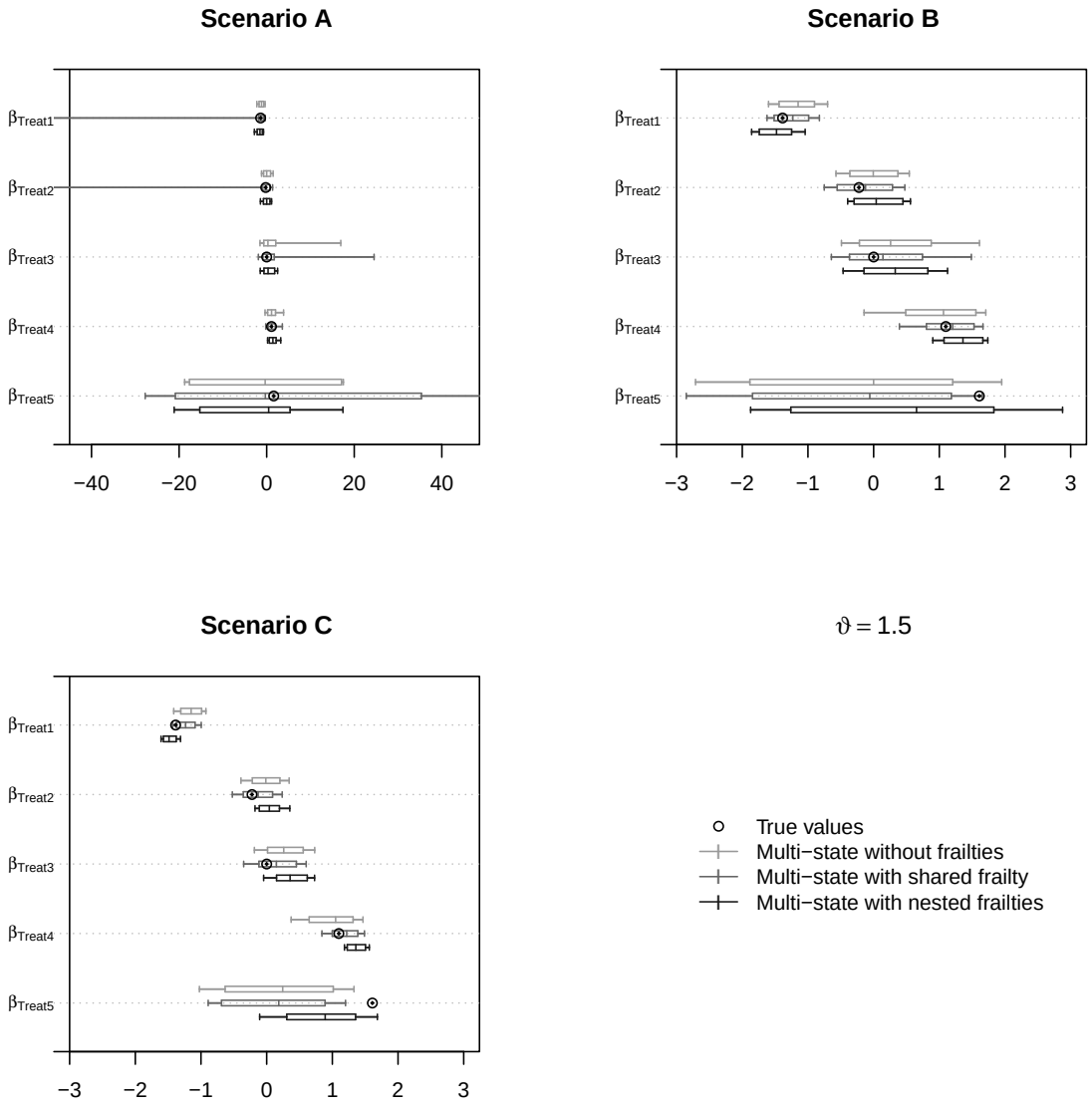



Figure 5.5 – Results of simulations for semiparametric models, with $\vartheta = 1.5$. Each box contains the 95% of the values over 500 replications; the vertical line within it is the median; the extreme ticks are the minimum and the maximum.

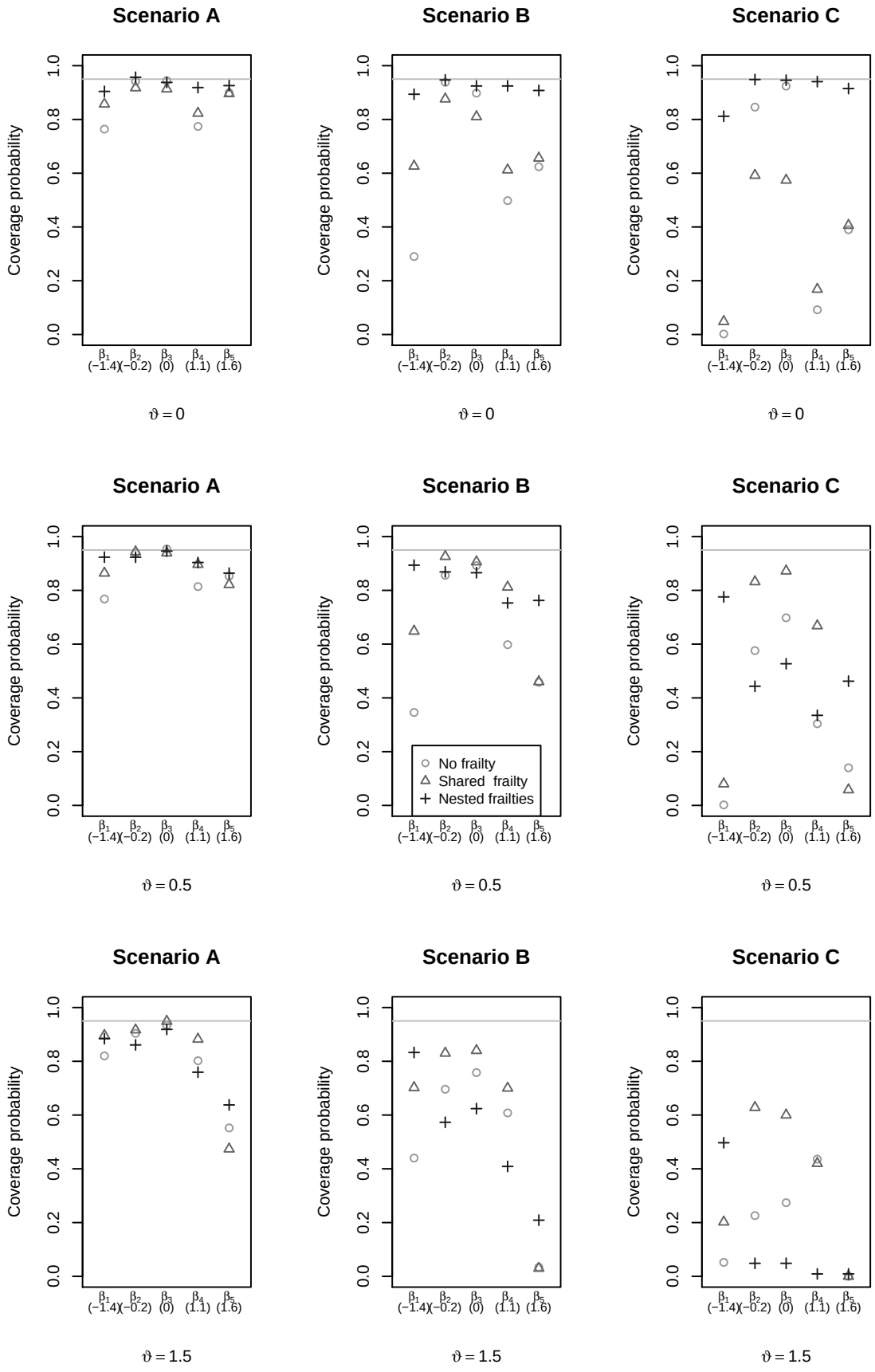



Figure 5.6 – Results of simulations for semiparametric models. Coverage probabilities for regression parameters. The horizontal line is the nominal value of the confidence intervals: 0.95.

Scenario **C** and $\vartheta = 0$

| | **True value** | **MSM** MEAN (eSE \| mSE) | **SFM** MEAN (eSE \| mSE) | **NFM** MEAN (eSE \| mSE) |
|---|---|---|---|---|
| $\beta_1$ | -1.39 | -1.08 (0.07 \| 0.05) | -1.19 (0.06 \| 0.05) | -1.43 (0.05 \| 0.05) |
| $\beta_2$ | -0.22 | -0.18 (0.08 \| 0.06) | -0.33 (0.09 \| 0.06) | -0.21 (0.06 \| 0.06) |
| $\beta_3$ | 0.00 | 0.01 (0.09 \| 0.08) | -0.14 (0.09 \| 0.08) | 0.02 (0.08 \| 0.08) |
| $\beta_4$ | 1.10 | 0.84 (0.12 \| 0.05) | 0.91 (0.09 \| 0.05) | 1.10 (0.05 \| 0.05) |
| $\beta_5$ | 1.61 | 1.44 (0.26 \| 0.07) | 1.44 (0.16 \| 0.07) | 1.61 (0.08 \| 0.07) |

| | **True value** | **SFM** MEAN ( eSE ) | **NFM** MEAN ( eSE ) |
|---|---|---|---|
| $\theta_V$ | 0.50 | 9.38e-01 (4.73e-02) | 3.22e-02 (9.39e-02) |
| $\theta_W$ | 1.00 | ——— (———) | 3.71e+00 (2.24e-01) |

Table 5.5 – Results of simulations for scenario **C** and $\vartheta = 0$. **MSM**: multi-state model without frailties. **SFM**: multi-state model with shared frailties. **NFM**: multi-state model with nested frailties. MEAN, eSE and mSE: mean estimates, empirical standard errors and mean estimated standard errors over 500 repetitions.

differences between transitions. On the other hand, the nested frailty model can account for both of them but with heavy and systematic bias: the cluster effect variance is underestimated, whereas that of the cluster-by-transition interactions is underestimated, whatever the context. In Appendix B.3.2 we show box-plots for the heterogeneity parameter estimates for this simulation study and for a second one with much higher heterogeneity ($\theta_V = \theta_W = 2$, i.e. $\mathsf{V}[U] = 8$). Results are analogous to the present setting, meaning that apparently the size of the $\theta$'s is not the reason for them to be recovered poorly. This is an evidence of an identification problem for the frailty parameters: the inferential process can account for unobserved heterogeneity to correctly estimate regression parameters, but it is not able to properly discriminate the contribution of each clustering level. This result should induce the researcher to be very sceptic in using these models to investigate unobserved heterogeneity and to study its sources, but to be very confident on their results to evaluate the effect of risk factors.

# Chapter 6

# Case studies

## 6.1 A pooled database of seven phase III bladder cancer clinical trials

One of the most common urological malignancies is bladder cancer, about 70-80% of which are superficial (stage Ta-T1). Standard treatment typically consists of transurethral resection (TUR) conducted with the aim of removing all tumours. However, a high proportion of patients experience recurrence or progression to muscle invasive disease, even after complete resection, because of residual tumour cells. Therefore, randomised phase III trials have been conducted over the last decades to investigate the use of prophylactic treatment following TUR. The objective of such treatment is to both remove residual, unresectable lesions and to prevent recurrence after complete resection.

In this study, we considered the data of 2649 eligible bladder cancer patients randomised by 63 European centers in seven consecutive phase III randomised clinical trials conducted by the Genito-Urinary Group of the European Organization for Research and Treatment of Cancer (EORTC trials 30781, 30782, 30791, 30831, 30832, 30845 and 30863; Newling *et al.*, 1995; Bouffioux *et al.*, 1992; Kurth *et al.*, 1984; Bouffioux *et al.*, 1995; Witjes *et al.*, 1998; Oosterlinck *et al.*, 1993). Among all these patients, the 2596 ones with stage Ta-T1 cancer were considered, approximately half with primary bladder cancer and half with recurrent disease. Other eleven patients were excluded for missing information about the time to some of the endpoints of interest. Eventually, 62 others were dropped as they had some important risk factors missing: 2523 patients are then used for our analyses. Within the context of these trials, patients in each of these participating centers were treated with or without further intravesical treatment after TUR. In total, 1368 patients (54.2%) received it, while 1155 patients (45.8%) did not.

The patients, 509 females and 2014 males aged between 12 and 92 years (median: 65 years), were recruited from 63 hospitals in 11 countries (Fig. 6.1).

Patients were followed up starting from the date of randomisation and the times to the first bladder recurrence (Rec), the first progression (Prog, increase to stage T2 or higher disease in the bladder), and to the date of death (De) were recorded. When censoring occurs, the most recent information is available at the date of the last follow-up cytoscopy.

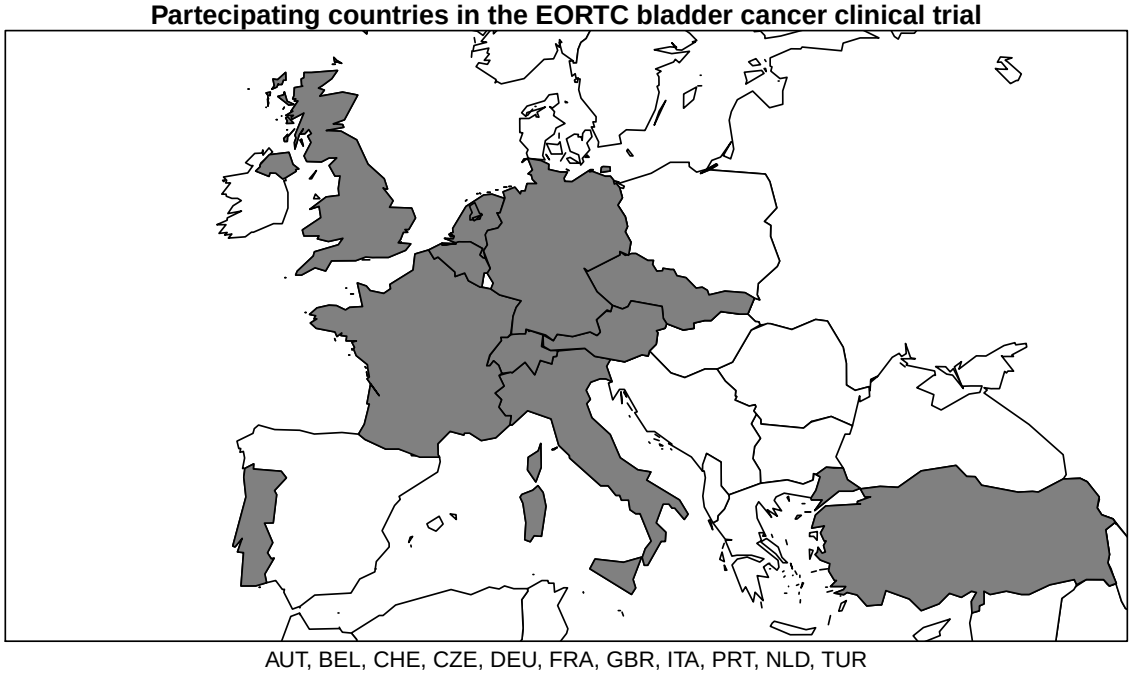



Figure 6.1 – Countries (with ISO 3166-1 alpha-3 codes) involved in the EORTC bladder cancer study: Austria, Belgium, Czech Republic, France, Germany, Italy, Portugal, Switzerland, Netherlands, Turkey, United Kingdom.

A total of 1179 (46.7%) recurrences were recorded with overall median time of 2.78 years, with significantly longer times in the intravesical treatment group (HR: 0.87, 95% CI: 0.77–0.97, p-value: 0.015). Fewer progressions, 276 (10.9%), were observed, with no significant difference between the two treatment groups (p-value: 0.701). During the study, 829 (32.9%) subjects died, with a median survival time of 10.8 years. Overall survival was significantly worse in the intravesical treatment group (HR: 1.15, 95% CI: 1.001–1.320, p-value: 0.048).

As concerns the multi-state nature of the problem, patients were followed since the randomisation moment and they could die during the study; during this period, they could have a recurrence or a progression or both of them. In 227 out the 231 cases of both adverse events, progression occurred not before a recurrence; hence we have chosen to consider as inadmissible the occurrence of a recurrence after a progression. Figure 6.2 shows the multi-state structure which reflects this framework. Finally, recurrence and progression were recorded at the same time for 64 patients. In these cases, the less relevant event — the recurrence — has been dropped; the number of remaining recurrences is then 1116.

### 6.1.1 Models

We consider and compare five semiparametric models for the bladder cancer data, fitted according to the multi-state structure in Figure 6.2.

MS: a simple multi-state model

ST: a multi-state model, stratified by center

SF1: a multi-state model with shared gamma frailties for the center effect

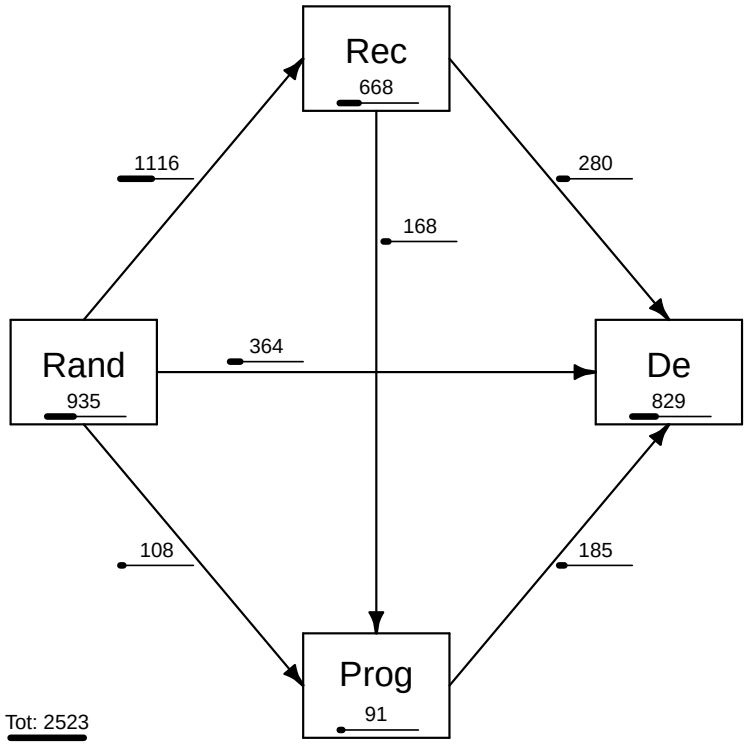



Figure 6.2 – States and transitions for the bladder cancer data. Rand: Randomisation, De: Dead, Rec: Recurrence, Prog: Progression. The number of observed transitions are showed on the arrows, whereas the number of patients ending the study in each state are in the boxes.

| Effect of | on Rec | on Prog | on De |
|---|---|---|---|
| **Tumour size** | • | • | ○ |
| **Number of tumours** | • | • | ○ |
| **Recurrent disease** | • | • | ○ |
| **Age** | ○ | • | • |
| **Sex** | ○ | ○ | • |

Table 6.1 – Bladder cancer study. Covariates included in the models, based on previous model selection for partial endopoints. A black bullet means that the covariate (row) is included for the hazard of the event (column); an empty circle means that the effect is not included.

SF2: a multi-state model with shared gamma frailties for the center-by-transition interaction effect

NF: a multi-state model with nested gamma frailties, one for the center effect and one for the center-by-transition interaction

A model with fixed effects for the center is not feasible, as this would require $(63 - 1) \times 6 = 186$ parameters just to keep under control the group effect.

Preliminary investigation on models MS, ST, SF1 and SF2 led to select some important factors to take into account. Namely, we decided to account for the effect on the recurrence and progression risks of the tumour size, of the number of tumours and of having a recurrent disease; the effect of the age on the risks of death and of progression; the effect of the sex on the risk of death (Tab. 6.1).

The effect of the treatment is estimated separately for each of the six transitions, as it is the focus of interest and as it is possible thanks to a sufficient number of observed events and at-risk subjects.

According to the results shown in Figure 6.3 and presented in details in Appendix B.4, the treatment has a protective effect against recurrences in all models, with an estimated reduction of risk ranging from 13.7% (NF) to 22.6% (MS). No model detects a statistically significant effect of further intravesical treatment after TUR on the rate of progression, neither directly nor after a recurrence.

The significant increase of death risk previously detected by the Cox model on the single endpoint, can be investigated in more details now. No effect of the treatment is present on overall survival without any intermediate event, whereas an increased risk is there for patients having a bladder cancer recurrence: HR estimate of the nested frailty model is 1.47, with p-value for the Wald test less than 0.002 against the hypothesis of no effect. This phenomenon, even though puzzling at a first glance, can be explained by the fact that patients with a recurrence are weaker and can suffer from undergoing further therapy, with benefits compensated and overtaken by the side effects. On the other hand, the most seriously ill patients, i.e. having a progression, show a beneficial effect of further treatment on survival (HR: 0.73, 95% CI: 0.54–0.99) but with borderline significance (p-value: 0.040).

As an aside, it results that the risk of death is about 70% higher for men than for women and that ten more years of age imply more than the double of risk.

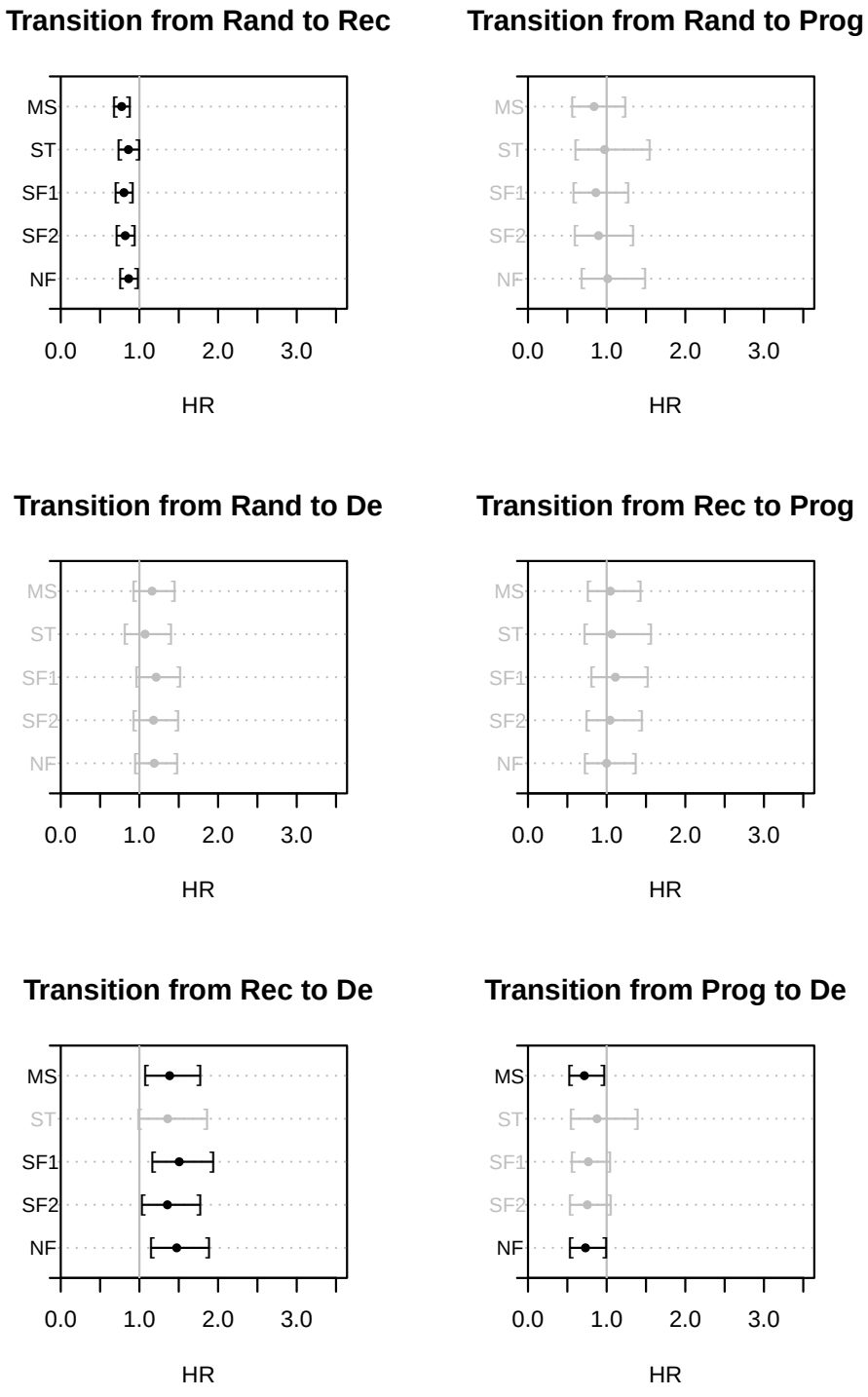



Figure 6.3 – Bladder cancer study. Hazard ratios, with approximate 95% confidence intervals for the treatment effect on the hazards of the six transitions. MS: multi-state model; ST: multi-state model stratified by hospitals; SF1: shared-frailty multi-state model for hospital; SF2: shared-frailty multi-state model for center-by-transition; NF: nested-frailty multi-state model. The black-coloured estimates are significantly different from 1, the grey ones are not.

## 6.2 A randomised trial for metastatic castration-resistant prostate cancer progressing after treatment

Prostate cancer is one of the first causes of death in men in occidental countries. Metastatic prostate cancers are treated via androgen suppression therapy, which improves survival but does not inhibit the disease progression. Docetaxel in combination with prednisone is administered for metastatic castration-resistant prostate cancer but, in case of still resistant diseases, no therapy has been approved by the US Food and Drug Administration. Mitoxantrone is used to improve patients' quality of life, but without survival benefit. The Sanofi–Aventis phase III clinical trial TROPIC (#EFC6193) showed that cabazitaxel plus prednisone improved survival of patients as compared to mitoxantrone plus prednisone, with an overall reduction of about 30% of death risk (de Bono *et al.*, 2010). Figure 6.4 resumes the disease and therapy history until randomisation.

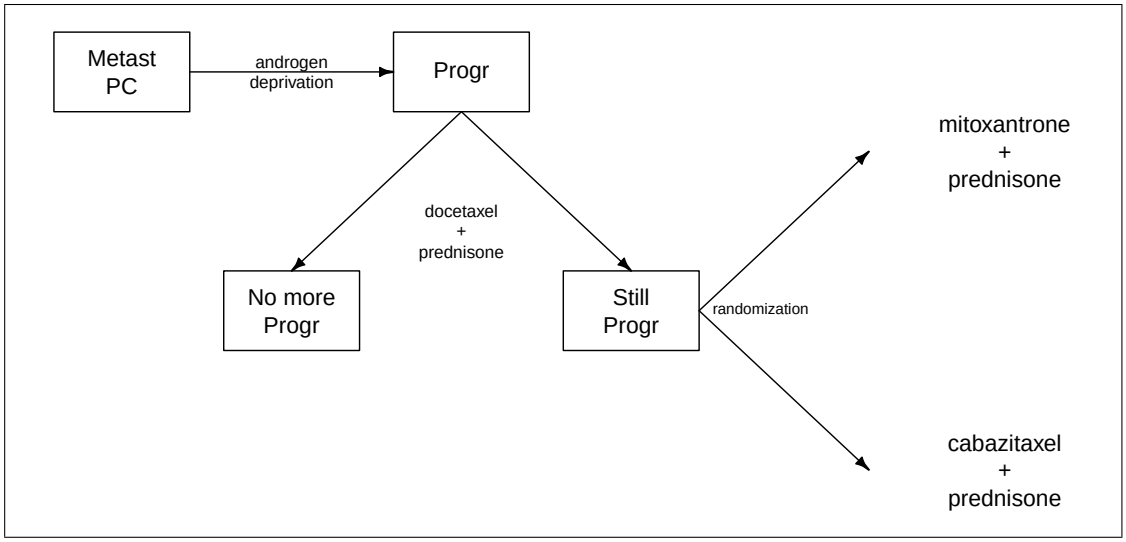



Figure 6.4 – Prostate cancer study. Pre-randomisation disease and therapy history.

The open-label trial was performed collaboratively by 132 hospitals in 26 countries (Fig. 6.5). Data are available on 755 men aged between 46 and 92 years (median: 67 years), randomly assigned to the cabazitaxel (378 pt.'s) or the mitoxantrone (377 pt.'s) treatment arms.

The survival time since randomisation is the primary endpoint, but also the time to tumour progression and Prostate-Specific Antigen (PSA) progression are recorded. A total of 513 (67.95%) deaths were recorded, with overall survival time of 13.93 months and significantly longer times of the experimental treatment arm (HR: 0.69, 95% CI: 0.58–0.82, p-value $< 0.001$). If progression-free survival is considered, 731 (96.82%) events are observed, with overall median time of 2.07 months; again, a strongly significant treatment effect is detected, with hazard ratio 0.75 (95% CI: 0.64–0.86, p-value $< 0.001$). Moreover, 504 (68.2%) PSA progressions (median time of 4.83 months) and 348 (46.09%) tumour progressions (median time of 7.23 months) were recorded. The effect of the treatment is also significant for the time to both PSA and tumour progression, with hazard ratios equal to 0.75 (95% CI: 0.63–0.89, p-value: 0.001) and 0.63 (95% CI: 0.51–0.78, p-value $< 0.001$), respectively. Figure 6.6 provides the survival curves estimated via the Kaplan–Meier method for the four endpoints, showing the general improvement of patients performances for the experimental

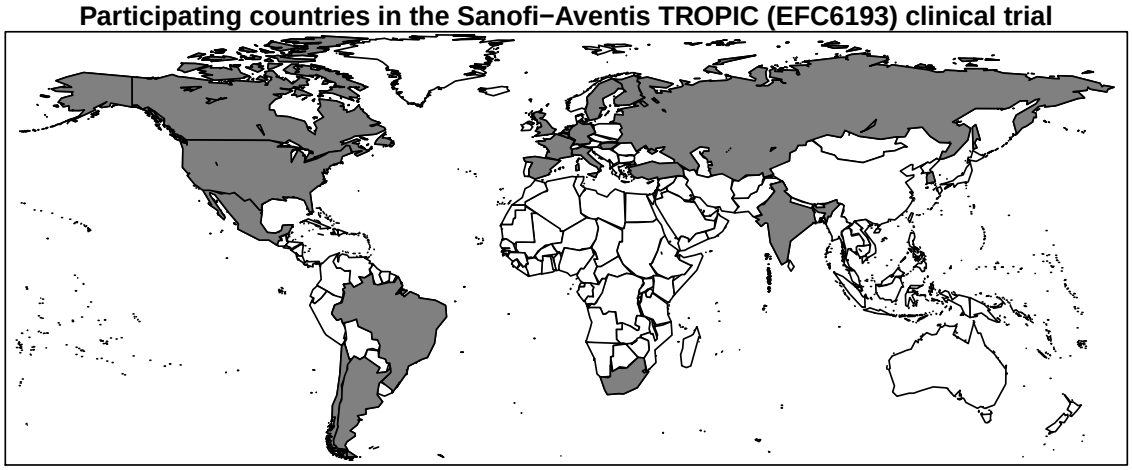



Figure 6.5 – Countries (with ISO 3166-1 alpha-3 codes) involved in the Sanofi–Aventis prostate cancer study: Argentine, Belgium, Brazil, Canada, Chile, Czech Republic, Denmark, Finland, France, Germany, Hungary, India, Italy, Republic of Korea, Mexico, Netherlands, Russia, Singapore, Slovakia, South Africa, Spain, Sweden, Taiwan, Turkey, United Kingdom, United States of America.

therapy arm.

The primary objective of the clinical trial was the evaluation of the effect of the therapy, cabazitaxel plus prednisone, as compared to mitoxantrone plus prednisone, on overall survival (de Bono *et al.*, 2010). Frailty multi-state models can provide a deeper insight in the comprehension of the disease and of the action of cabazitaxel, mainly in two ways. First, the use of frailties can account for dependence due to unobserved risk factors, shared by patients recruited in the same country. Secondly, the use of a multi-state structure can dynamically account for the effect of intermediate events, namely PSA progression and tumour progression, on the risk of death.

To this aim, we will consider three families of models

- two separate models for overall survival, one with intermediate PSA progression and one with tumour progression
- a model for progression-free survival, with intermediate PSA progression
- a model for overall survival, with intermediate events PSA progression and tumour progression

and, for each of them, we compare

MS: a multi-state model without any frailty,

ST: a multi-state model, stratified on the country,

SF1: a multi-state model with shared frailties for the country,

SF2: a multi-state model with shared frailties for the country-by-transition interaction,

NF: a multi-state model with nested frailties, accounting for both the country effect and the country-by-transition interaction,

fitted via semiparametric inference. In this trial patients are recruited from a very wide geographical area, so that country differences are supposed to be explicative enough of differences in risks. For this reason clustering in considered at country rather than at hospital level.

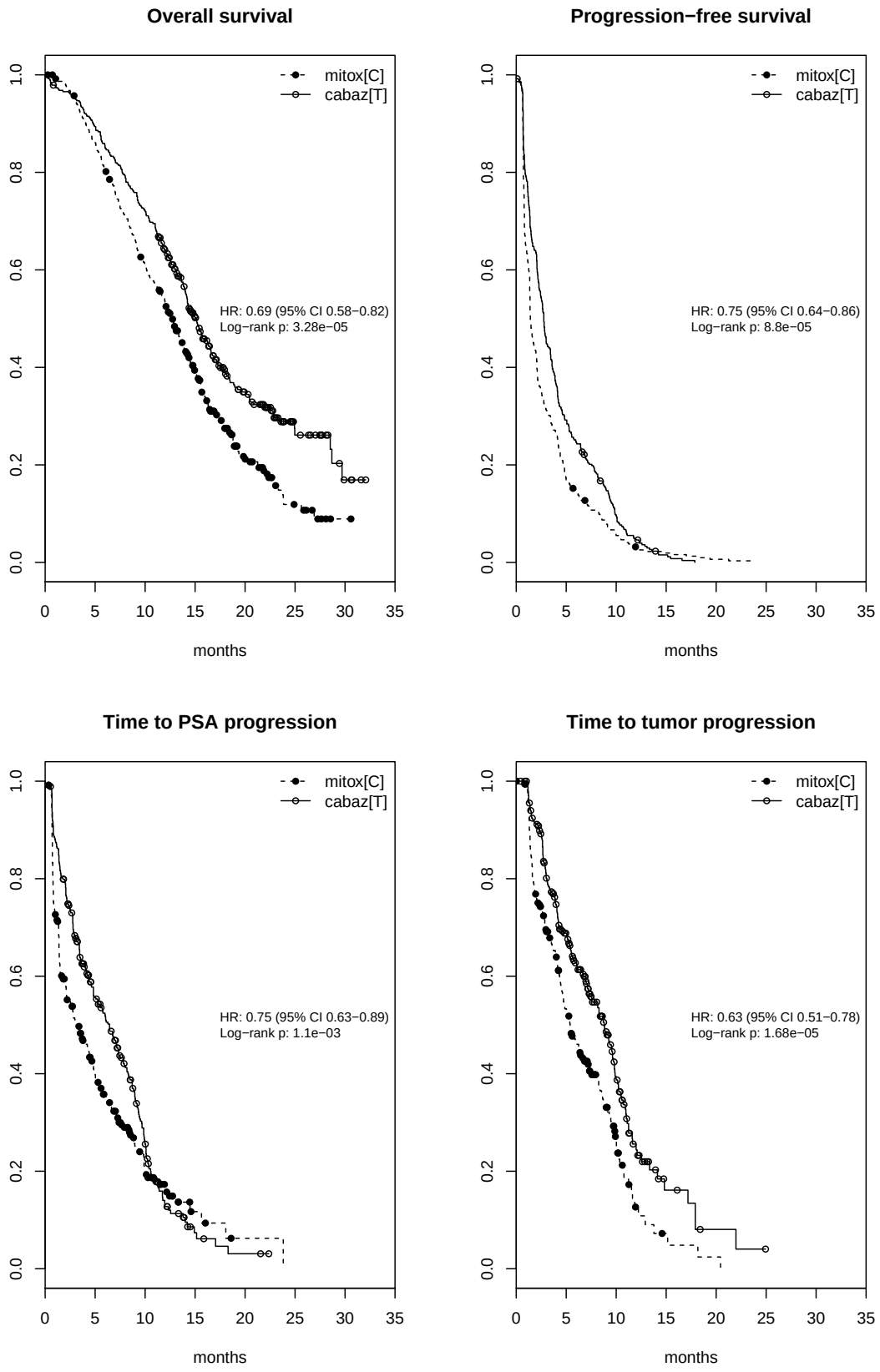



Figure 6.6 – Prostate cancer study. Kaplan–Meier curves for time to PSA progression, tumour progression, overall survival and progression-free survival, by treatment arm.

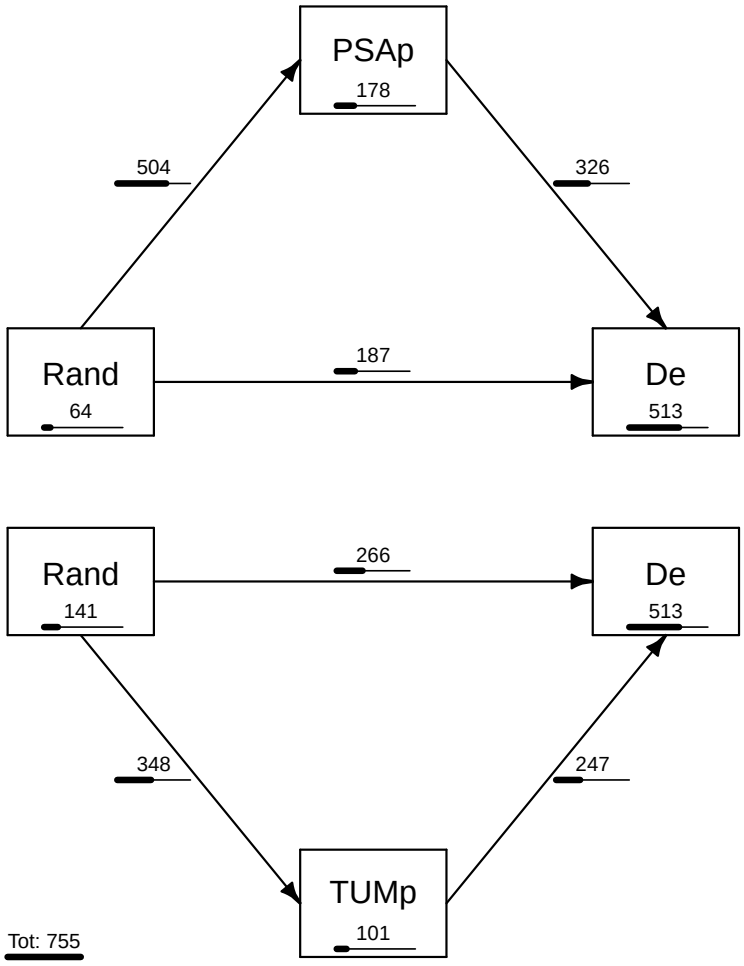



Figure 6.7 – States and transitions for the prostate cancer study. Separate models for overall survival with PSA and tumour progressions. Rand: Randomisation, De: Dead, PSAp: PSA progression, TUMp: Tumour progression. The number of observed transitions are showed on the arrows, whereas the number of patients ending the study in each state are in the boxes.

### 6.2.1 Separate models for overall survival

Let us first consider the effect of disease progression on the death risk. Progression was measured at two levels: tumour progression was assessed by Response Evaluation Criteria in Solid Tumour (RECST, Therasse *et al.*, 2000) with at least one visceral or soft-tissue metastasis, whereas diseases were classified as PSA progressive if an increase of at least 25% of nadir PSA was observed in non-respondent patients with at least 5 $\mu g/L$ PSA increase and for responders with an increase of at least 50% of nadir PSA. This latter criterion is expected to be able to detect progressive diseases earlier, whilst the former one at a more advanced stage.

Two separate models are fitted for overall survival, one with intermediate PSA progression and one with tumour progression (Fig. 6.7). More progressions are measured by PSA increase (504) than by RECST (348), since the first method detects progressive diseases earlier, when progression

is not physically visible, yet. Indeed, estimated median event time are 4.83 months for PSAp, 7.23 months for TUMp.

Results from the model with PSA progression (Fig. 6.8 , App. B.5.1) show that patients treated with cabazitaxel plus prednisone have a significantly reduced risk of progression (HR: 0.79, 95% CI: 0.67–0.95) and of death, both before progression (HR: 0.63, 95% CI: 0.46–0.86) and after it (HR: 0.70, 95% CI: 0.56–0.88). The occurrence of PSA progression does not turn out to have significant impact on death risk (p-value: 0.211).

If the progression is assessed by the RECST (Fig. 6.9, App. B.5.1), the new treatment has a significantly protective action against death, both before progression (HR: 0.66, 95% CI: 0.51–0.85) and after it (HR: 0.67, 95% CI: 0.52–0.87). Notice that the size of the hazard reduction is of the same magnitude as in the model with PSAp, about 30-40%.

The reduction of the progression hazard due to the new treatment (HR: 0.82, 95% CI: 0.67–1.02) is not significant but borderline (p-value: 0.071).

### 6.2.2 Models for progression-free survival

Then, we consider a model for progression-free survival, instead of overall survival, where the endpoint is either death or tumour progression, whilst PSA progression is a possible intermediate event (Fig. 6.10).

On the basis of these analyses (Fig. 6.11, App. B.5.2), the new treatment has a significant beneficial effect on the PSA progression (HR: 0.69, 95% CI: 0.57–0.83), coherently to results of the overall survival model with intermediate PSA progression. Consistently with models in Section 6.2.1, patients who undergone cabazitaxel therapy show a significant reduction of the hazard of dying or having tumour progression (HR: 0.71, 95% CI: 0.56–0.91). No treatment effect on progression-free survival is observed after having already had a PSA progression (p-value: 0.405). The occurrence of a PSA progression seems to increase the risk of death or tumour progression (HR: 1.17, 95% CI: 0.93–1.49) but this effect is not statistically significant (p-value: 0.187).

### 6.2.3 Global models for overall survival

Finally, a model for overall survival is considered with possible intermediate PSA and tumour progressions, and where a tumour progression is possible after PSA progression (Fig. 6.12). 447 patients had PSA progression as first event and 159 of them died, then. Tumour progression was observed in 348 men, 185 of which after a PSA progression; 247 of them died successively during the trial. Death without any clinical evidence of progression occurred for 107 patients. No event was recorded for 38 men.

The effect of the new therapy is estimated separately for each of the six transitions, by means of five variables treat.1, ..., treat.6. Three different baseline hazards are assumed for PSA progression, tumour progression, and death. The baseline hazard functions of the three transitions into the state De are assumed to be proportional to each other, with hazard ratios estimated by means of two artificially time-dependent indicator variables (PSAp.OS and TUMp.OS) recording whether each type of progression has been already observed or not. Analogously, proportionality is assumed for the hazards of the two transitions into TUMp, with log-proportionality estimated thanks to the time-dependent indicator PSAp.TUMp. The results, provided in details in Appendix B.5.3 and

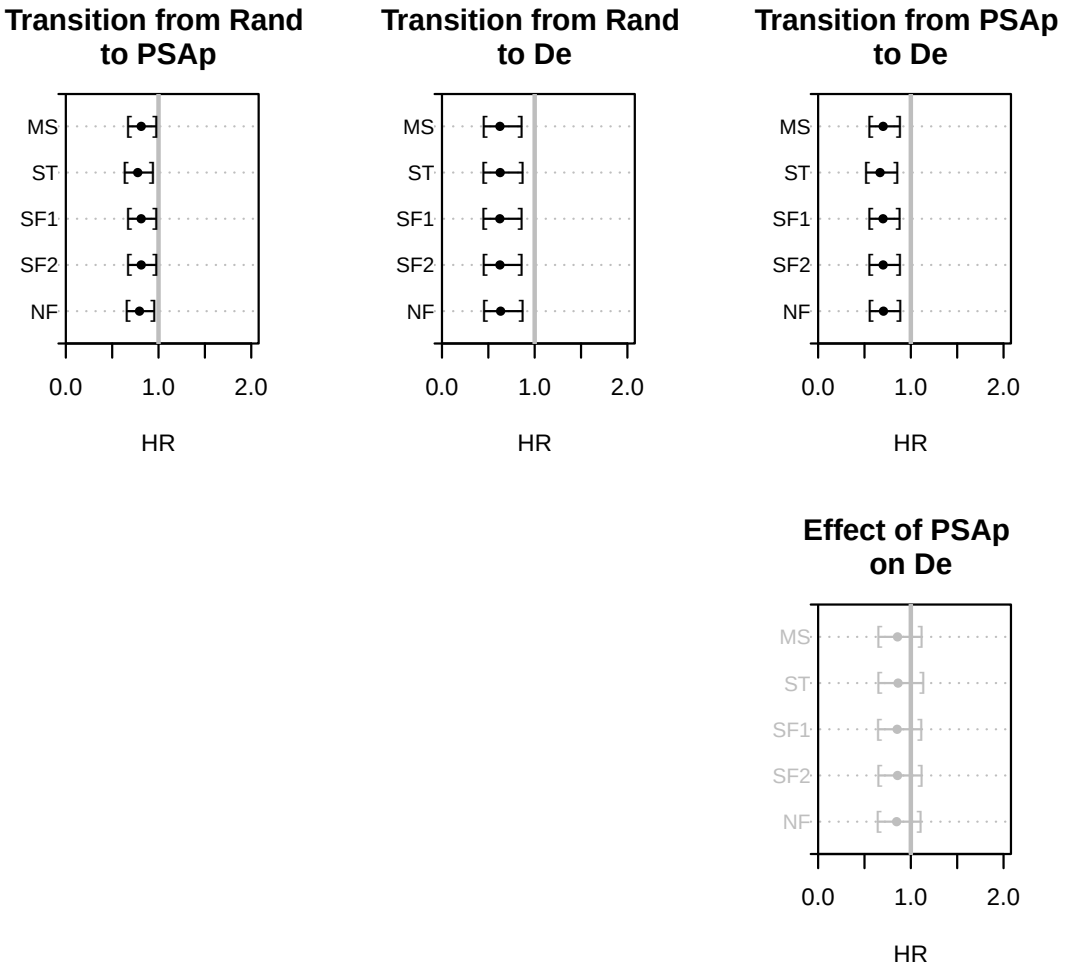



Figure 6.8 – Prostate cancer study. Models for overall survival with intermediate PSA progression. Estimated hazard, with approximate 95% confidence intervals. The first three graphs refer to the effect of the experimental treatment on the risk of each transition; the last one refers to the increase of death risk due to the occurrence of PSA progression (PSAp). The black-coloured estimates are significantly different from 1, the grey ones are not.

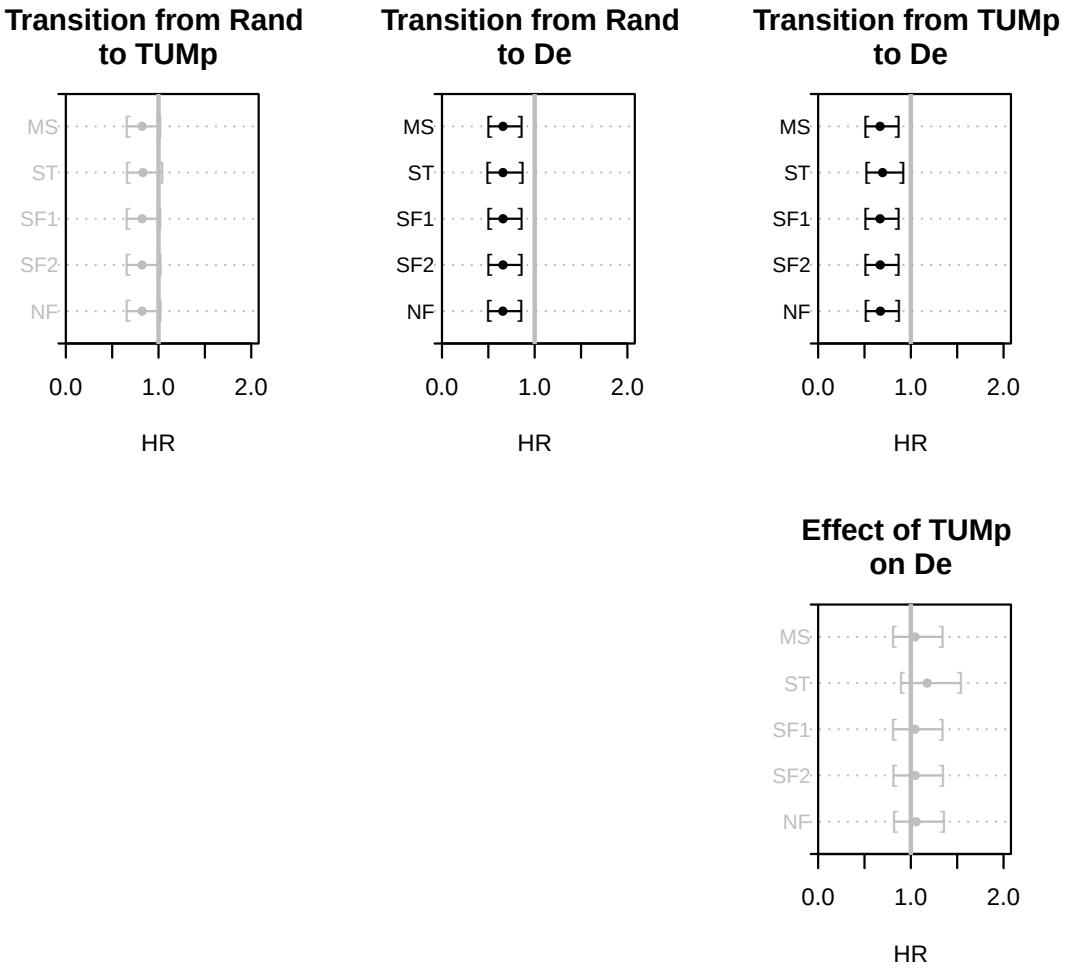



Figure 6.9 – Prostate cancer study. Models for overall survival with intermediate tumour progression. Estimated hazard, with approximate 95% confidence intervals. The first three graphs refer to the effect of the experimental treatment on the risk of each transition; the last one refers to the increase of death risk due to the occurrence of tumour progression (TUMp). The black-coloured estimates are significantly different from 1, the grey ones are not.

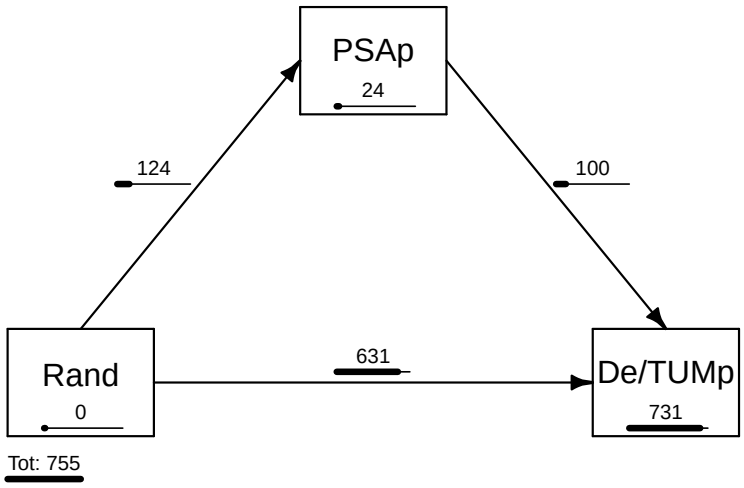



Figure 6.10 – States and transitions for the prostate cancer study. Model for progression-free survival with PSA progression. Rand: Randomisation, PSAp: PSA progression, De/TUMp: Dead or Tumour progression. The number of observed transitions are showed on the arrows, whereas the number of patients ending the study in each state are in the boxes.

resumed in Figure 6.13, show that cabazitaxel plus prednisone strongly reduces the risk of PSA progression (HR: 0.68, 95% CI: 0.57–0.83) with respect to mitoxantrone plus prednisone. If all the models suggest a reduction of tumour progression, too, only after controlling for unobserved heterogeneity at both levels (NF model) the reduction turns out to be significant (HR: 0.73, 95% CI: 0.53–0.99, p-value: 0.042).

The multi-state nature of the models is valuable in putting in evidence that the benefit of cabazitaxel on survival performance that de Bono *et al.* (2010) found — see also the top-left plot in Fig. 6.6 — is different according to intermediate events. The effect is strong and significant for patients without progression (HR: 0.60, 95% CI: 0.39–0.93) and for those having tumour progression (HR: 0.73, 95% CI: 0.57–0.94). On the other hand, the hazard reduction is less important and not significant after a PSA progression (HR: 0.79, 95% CI: 0.57–1.10, p-value: 0.160), revealing that the effect detected by the first models in Section 6.2.1 is spurious and is mainly due to patients with successive tumour progressions, ignored by those models.

The occurrence of a PSA progression strongly increases the death hazard (HR: 1.83, 95% CI: 1.27–2.65, p-value: 0.001) and the tumour progression hazard (HR: 2.01, 95% CI: 1.40–2.89, p-value $< 0.001$). Finally, as one can expect, the occurrence of a tumour progression increases the death risk even more than PSA progression (HR: 2.27, 95% CI: 1.60–3.22, p-value: $< 0.001$).

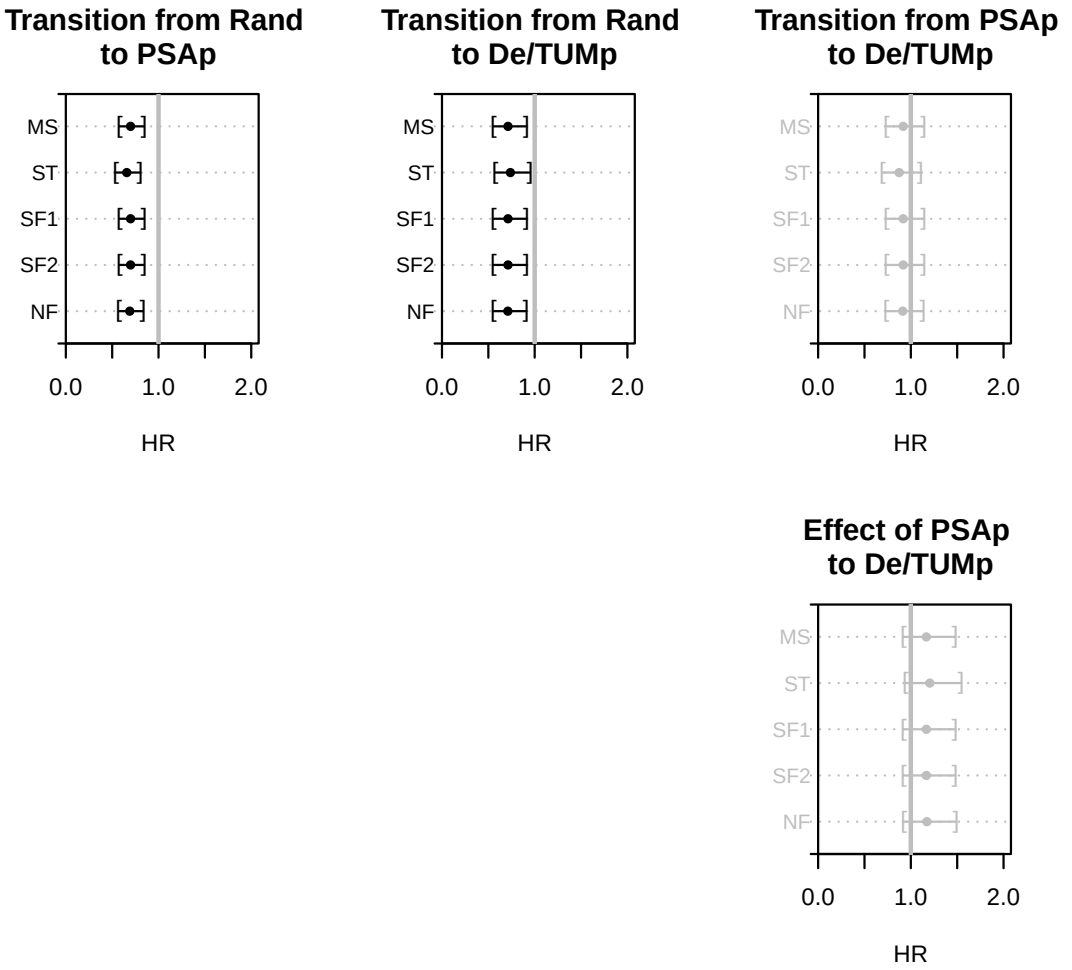



Figure 6.11 – Prostate cancer study. Models for overall survival with intermediate PSA progression. Estimated hazard, with approximate 95% confidence intervals. The first three graphs refer to the effect of the experimental treatment on the risk of each transition; the last one refers to the increase of death risk due to the occurrence of PSA progression (PSAp). The black-coloured estimates are significantly different from 1, the grey ones are not.

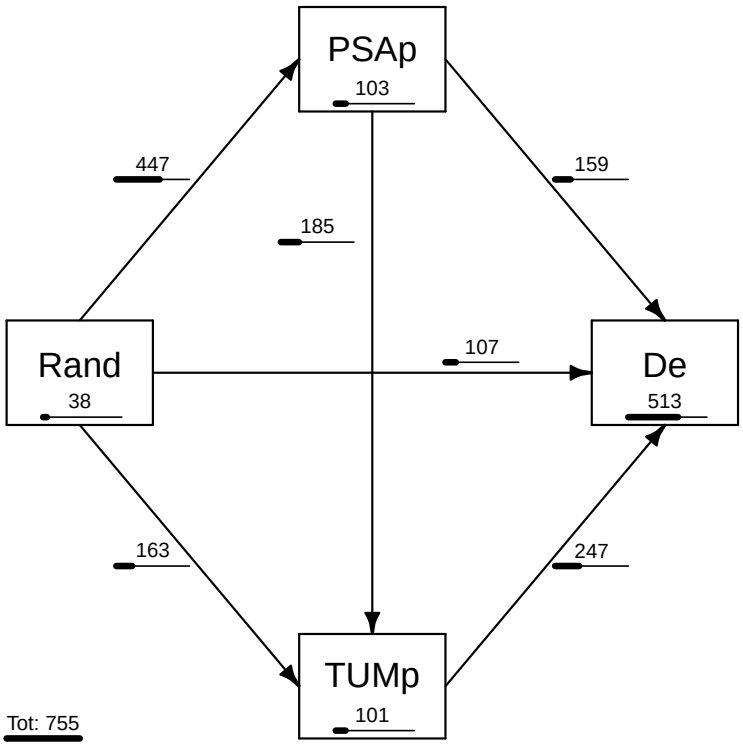



Figure 6.12 – States and transitions for the prostate cancer study. Rand: Randomisation, De: Dead, PSAp: PSA progression, TUMp: Tumour progression. The number of observed transitions are showed on the arrows, whereas the number of patients ending the study in each state are in the boxes.

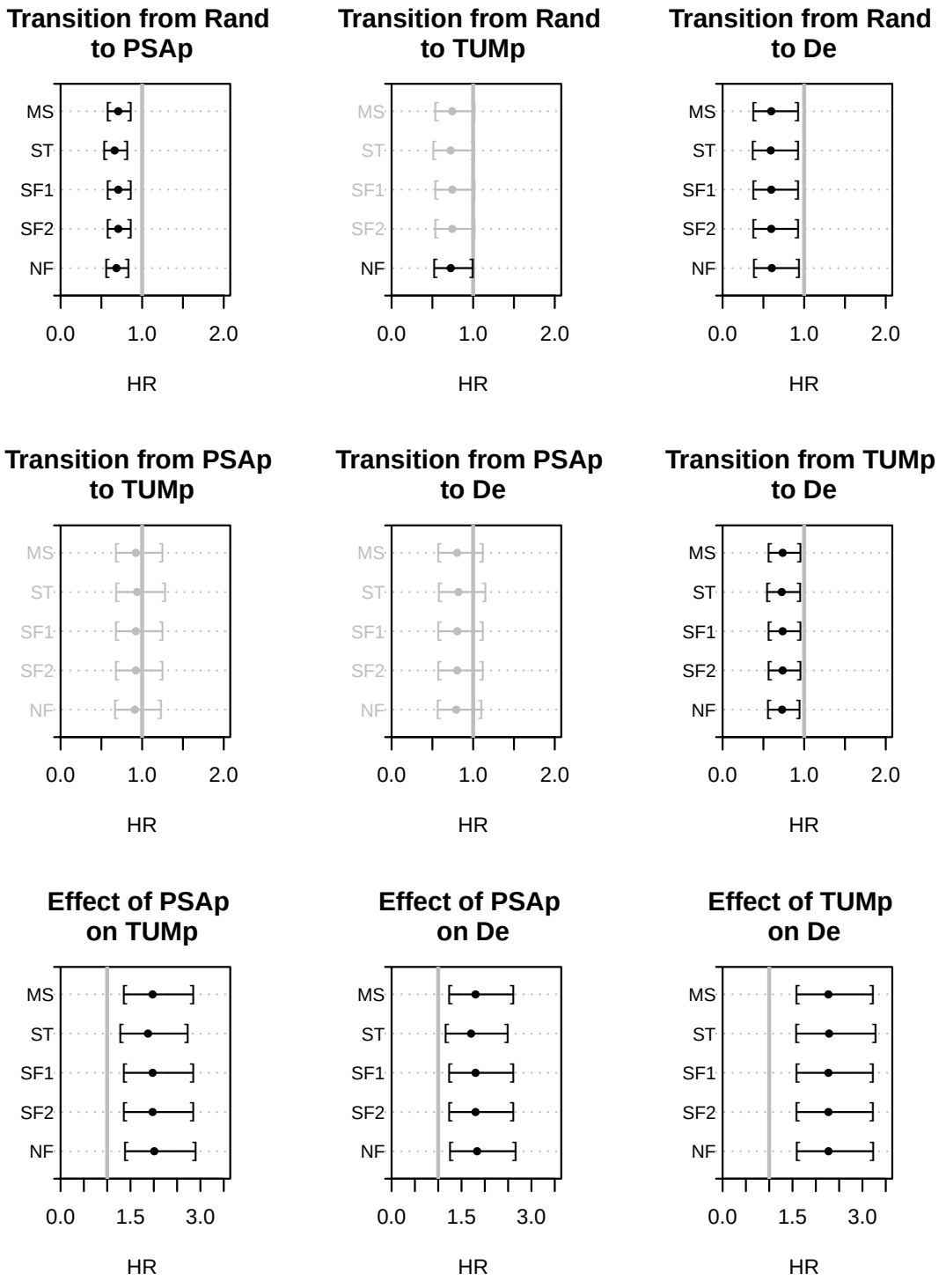



Figure 6.13 – Prostate cancer study. Estimated hazard ratios, with approximate 95% confidence intervals. Global models for overall survival. The first six graphs refer to the effect of the experimental treatment on the risk of each transition; the last three refer to the increase of the risk of of TUMp or death, due to the occurrence of PSAp and TUMp. The black-coloured estimates are significantly different from 1, the grey ones are not.

# Chapter 7

# Software

Many statistical software packages are available nowadays which range from more user-friendly graphical interfaces to powerful programming languages. Among them, SAS (SAS Institute Inc., 2011) and `R` (R Development Core Team, 2012) are the two most popular ones; the former one is mostly used in the industry as offers well-reviewed routines, is very fast and can deal with very large datasets. On the other hand, the latter one is particularly appreciated by the academic statistical scientific community, as the packages system allows everybody to contribute by providing ready-to-use functions implementing the most recent and advanced statistical tools.

`R` is a statistical programming language for high level computation and graphics; it is derived by `S` and offers interfaces to other languages, such as `C` or `FORTRAN`. The `R` language looks similar to `C`, but it is actually a functional programming language, allowing computing on the language.

Many packages exist for survival analysis and the `survival` one by Therneau (2012b) is certainly a *sine qua non*, offering a bunch of tools for descriptive statistics, two-sample tests, parametric accelerated failure models, Cox models, etc. It also allows to fit semiparametric frailty models with gamma, lognormal and log-t frailties. Two other packages, `coxme` by Therneau (2012a) and `phmm` by Donohue and Xu (2012), fit semiparametric lognormal models with other estimation algorithms than the PPL. Parametric proportional hazards models can be fitted by means of the `eha` (Broström, 2012) package. Finally, the `frailtypack` package Gonzalez *et al.* (2012) fit gamma frailty models with the baseline hazard function approximated by cubic M-splines.

Multi-state models can be fitted as fixed effects proportional hazard models (`survival` and `eha` packages, for instance), once data are recoded into long format (de Wreede *et al.*, 2011; Putter, 2011). Putter *et al.* (2012) provide in the `mstate` package a set of valuable functions for such data preparation process.

In the following Sections 7.1–7.2 we present two new `R` packages, `parfm` (Rotolo and Munda, 2012) and `mlfm` (Rotolo and Horny, 2012), that we developed to fit parametric frailty models (Munda *et al.*, 2012) and multilevel frailty models for our research and which are now publicly available.

## 7.1 `parfm` package: parametric frailty models in `R`

Slowly but surely, a variety of estimation procedures for frailty models becomes available in standard statistical software. In `R`, the `coxph()` function from the `survival` package (Therneau, 2012b) handles the semiparametric model with gamma and lognormal frailties. Important options supported by `coxph()` and its output are described in details by Therneau and Grambsch (2000, Chapter 9). Recently, the `frailtypack` package (Gonzalez *et al.*, 2012) by Rondeau and Gonzalez (2005) and Rondeau *et al.* (2012) has been updated and it stands now for gamma frailty models with a quasi-semiparametric estimation but also with a parametric approach using the Weibull baseline hazard. Other `R` packages include `coxme` (Therneau, 2012a) and `phmm` (Donohue and Xu, 2012). These two perform semiparametric estimation in the lognormal frailty model. SAS also deals with the lognormal distribution. On the one hand, `proc phreg` can now fit the semiparametric lognormal frailty model. On the other hand, `proc nlmixed` deals with the parametric version by using Gaussian quadrature to approach the marginal likelihood; see, e.g., Duchateau and Janssen (2008, Example 4.16). In the parametric setting, STATA (StataCorp, 2011) provides some flexibility. The `streg` command (Gutierrez, 2002) is able to perform maximum likelihood estimation with various choices of baselines: exponential, Weibull, Gompertz, lognormal, loglogistic, and generalised gamma. Take notice, however, that STATA fits the accelerated failure time model. Still, with exponential or Weibull baselines, both the proportional hazards and the accelerated failure time representations are allowed. As for the frailty distribution, the gamma and the inverse Gaussian are the only two that are supported. On a side note, Bayesian analyses can be conducted in WinBUGS (Spiegelhalter *et al.*, 2003); see, e.g., Duchateau and Janssen (2008, Example 6.4). For a deeper overview of what supports what, and for a comparison of some of the aforementioned functions, see Hirsch and Wienke (2011).

The new `parfm` package (Rotolo and Munda, 2012) fits the gamma, the inverse Gaussian, the positive stable and the lognormal proportional hazards frailty models with either exponential, Weibull, Gompertz, lognormal, or loglogistic baseline. The `parfm` package is flexible and easy to use. Parameter estimation is done by maximising the marginal log-likelihood; the `optim()` function is employed, and its `method` option is passed to `parfm()` (with `method="BFGS"` by default). If not specified in the `inip` option, initial values for all but the heterogeneity parameter are obtained by fitting an unadjusted (i.e., without frailty) parametric proportional hazards model using the `phreg()` function from the `eha` package (Broström, 2012). The initial heterogeneity parameter can also be specified by the user via the `iniFpar` option; otherwise it is set to 1 when frailties follow a gamma, lognormal or an inverse Gaussian distribution, or to 1/2 when they follow the positive stable distribution.

Additionally, when `frailty="none"`, `parfm()` fits the unadjusted parametric proportional hazards model, similar to `survreg()` (from the `survival` package) or to `phreg()`. However, `survreg()` returns the parameter estimates in the log-linear model and `phreg()` uses yet another parametrisation (see the documentation). Often, the user has then to transform back the parameters and to employ the delta method in order to get estimates for the standard errors. The `parfm()` function directly uses the proportional hazards representation.

Nonetheless, `parfm` might reach its limits when at least one $d_h$, the number of events in the $h$-th cluster, $h \in \{1, \ldots, H\}$, is very large. First, consider the positive stable distribution and observe

that, for a fixed value of $m \in \{1, \ldots, k-1\}$, $\Omega_{k,m}$ rapidly grows as $k$ increases (Eq. 3.102). At the extreme, some of them might exceed the largest representable number in R. These are then stored as Inf. This, in turn, prevents the marginal log-likelihood (7.1) to be evaluated and hence maximised. As a side note, also the SAS macro ps_frail that implements the EM algorithm to fit the semi-parametric positive stable frailty model has analogous difficulties when the number of events is large or even moderate. The following ad-hoc solution is implemented in parfm: in order to keep the polynomials $\Omega_{k,m}$'s reasonably small, they are divided by some factor $10^K$ which does not change the marginal log-likelihood except for an additive constant, equal to $s \times K \times \log(10)$. The value of $K$ is specified via the correct option (default is correct=0, i.e., no correction) and parfm() returns the re-adjusted log-likelihood value. That solution serves the purpose for moderately large values of $d_h$ (say up to about 200 events per cluster according to our experience, but it depends on the data, on the other parameters, and on the hardware characteristics). With the inverse Gaussian distribution, the Bessel function $K_{k-1/2}(z)$ in Equation 7.4 raises the same problem. Indeed, it explodes when $z$ is small relative to $k$ (see Fig. 7.1). Currently, that distribution should, therefore, preferably be avoided when there are very large values of $d_h$ (say above 200 events per cluster according to our experience, but, again, it depends on the data, on the other parameters and on the hardware characteristics). Moreover, $K_{k-1/2}(z)$ rapidly goes to zero as $z$ increases. So, in case of very small apparent heterogeneity, $\theta \to 0$ which implies $z \to \infty$, $K_{k-1/2}(z)$ might be stored as 0 in R and hence $\log(K_{k-1/2}(z))$ cannot be computed. However, as this problem occurs in the case of very small heterogeneity, this would rather suggest to fit the model with frailty="none". When frailties are gamma or lognormal distributed, which is by far the most

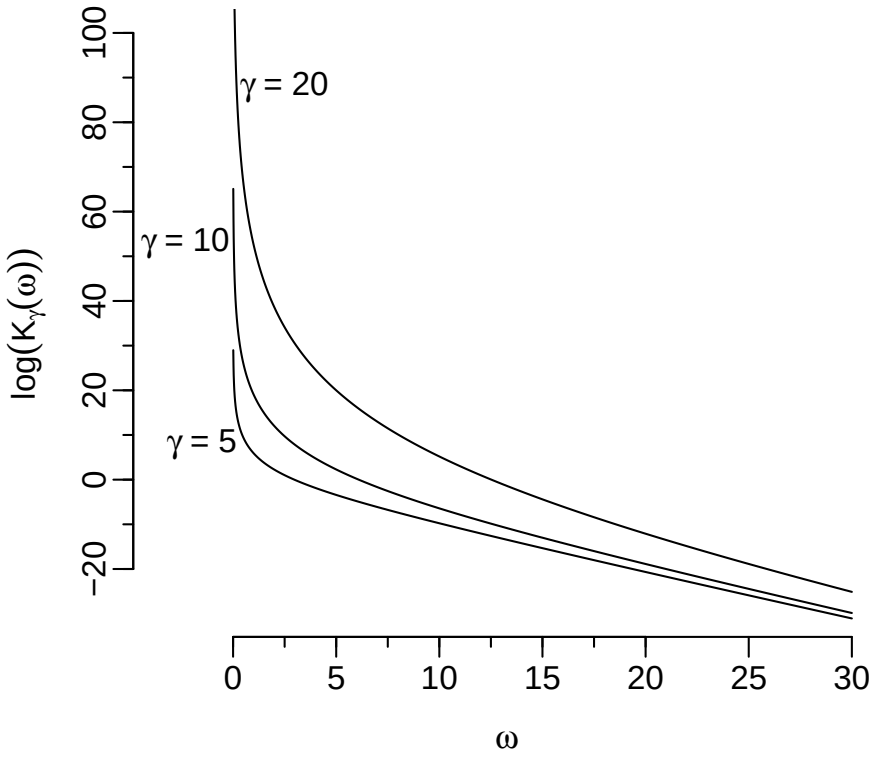



Figure 7.1 – The logarithm of the Bessel function, $\log(K_\gamma(\omega))$, versus $\omega$ for different values of $\gamma$.

popular assumption in common practice, the quantities involved in Equations 7.2 and 7.5 do not raise any worry. In practice, even when dealing with datasets with huge numbers of events per cluster, there is no real risk of exceeding the range of floating-point numbers.

### 7.1.1 Model estimation

From a modelling point of view, the multivariate model includes the univariate. We then refer to the shared frailty model (3.39) with any of the baseline hazard functions in Table 3.2.

Even though various frailty distributions exist (Sec. 3.3.4), we shall focus hereinafter on the gamma, the positive stable, the inverse Gaussian, and the lognormal. In all of these four, a single heterogeneity parameter indexes the degree of dependence.

For right-censored clustered survival data, the observation for subject $i \in \{1, \ldots, n_h\}$ from cluster $h \in \{1, \ldots, H\}$ is the couple $(y_{hi}, \delta_{hi})$, where $y_{hi} = \min(t_{hi}, c_{hi})$ is the minimum between the survival time $t_{hi}$ and the censoring time $c_{hi}$, and where $\delta_{hi} = I(t_{hi} \leq c_{hi})$ is the event indicator. Covariate information may also have been collected; in this case the observed data are the triplets $(y_{hi}, \delta_{hi}, \boldsymbol{x}_{hi})$, where $\boldsymbol{x}_{hi}$ denote the vector of covariates for the $hi$-th observation. Further, if left-truncation is also present, truncation times $\tau_{hi}$ are gathered in the vector $\boldsymbol{\tau}$.

Under assumptions of non-informative right-censoring and of independence between the censoring time and the survival time random variables, given the covariate information, the marginal (log)likelihood (3.56) can be written as (van den Berg and Drepper, 2012)

$$\begin{aligned}\ell_{\mathsf{M}}(\boldsymbol{\xi}) = \sum_{h=1}^{H} \Bigg\{ & \left[ \sum_{i=1}^{n_h} \delta_{hi} \left( \log(\lambda_0(y_{hi})) + \boldsymbol{\beta}^\top \boldsymbol{x}_{hi} \right) \right] \\ & + \log \left[ (-1)^{d_h} \mathcal{L}^{(d_h)} \left( \sum_{i=1}^{n_h} \Lambda_0(y_{hi}) \exp(\boldsymbol{\beta}^\top \boldsymbol{x}_{hi}) \right) \right] \\ & - \log \left[ \mathcal{L} \left( \sum_{i=1}^{n_h} \Lambda_0(\tau_{hi}) \exp(\boldsymbol{\beta}^\top \boldsymbol{x}_{hi}) \right) \right] \Bigg\}. \end{aligned} \tag{7.1}$$

Estimates of $\boldsymbol{\xi}$ are obtained by maximising the marginal loglikelihood (7.1); this can easily be done if one is able to compute higher order derivatives $\mathcal{L}^{(k)}(\cdot)$ of the Laplace transform up to $k = \max\{d_1, \ldots, d_H\}$. Symbolic differentiation might be performed in `R`, but is impractical here, mainly because this is very time consuming. Therefore, explicit formulas are rather desirable. Further, they will be used in the calculation of predictions as shown below.

The expression $\log\left((-1)^k \mathcal{L}^{(k)}(s)\right)$ in Equation 7.1 depends on the parametric form of the frailty distribution. For the gamma distribution it is

$$-\left(k + \frac{1}{\theta}\right) \log(1 + \theta s) + \sum_{l=0}^{k-1} \log(1 + l\theta), \tag{7.2}$$

for the positive stable

$$k \left( \log(1 - \theta) - \theta \log(s) \right) + \log \left[ \sum_{m=0}^{k-1} \Omega_{k,m} s^{-m(1-\theta)} \right] - s^{1-\theta}, \tag{7.3}$$

and for inverse Gaussian (cf. Sec. 3.3.4)

$$-\frac{k}{2}\log(2\theta s+1)+\log\left(K_{k-(1/2)}(z)\right)-\left[\frac{1}{2}\left(\log\left(\frac{\pi}{2z}\right)\right)-z\right]+\frac{1}{\theta}\left(1-\sqrt{1+2\theta s}\right), \tag{7.4}$$

with $z=\sqrt{2\theta^{-1}(s+\frac{1}{2\theta})}$. In the case of lognormal frailties, the derivatives of the Laplace transform do not exist in an explicit form, but the quantities $\log\left((-1)^k\mathcal{L}^{(k)}(s)\right)$ are computed as

$$k\tilde{w}-\exp(\tilde{w})s-\frac{\tilde{w}^2}{2\gamma}-\frac{1}{2}\log\Big(\gamma\exp(\tilde{w})s+1\Big), \tag{7.5}$$

with $\tilde{w}=\max_w g(w;k,s,\gamma)$, where $g(w;k,s,\theta)=kw-\exp(w)s-w^2/(2\gamma)$ (App. A.12).

**Prediction.** Besides parameter estimates, prediction of frailties are sometimes desirable. As an aside, they are needed at each expectation step of the EM algorithm that fits the semi-parametric frailty model.

The frailty term $u_h$ can be predicted by $\tilde{u}_h=\mathsf{E}\left(U\mid \boldsymbol{x}_h,\boldsymbol{y}_h,\boldsymbol{\tau}_h;\hat{\boldsymbol{\xi}}\right)$, with $\boldsymbol{x}_h$, $\boldsymbol{y}_h$, and $\boldsymbol{\tau}_h$ the data and the truncation times of the $h$-th cluster. This conditional expectation can be achieved as

$$\mathsf{E}\left(U\mid \boldsymbol{x}_h,\boldsymbol{y}_h,\boldsymbol{\tau}_h;\hat{\boldsymbol{\xi}}\right)=-\frac{\mathcal{L}^{(d_h+1)}\Big(\sum_{i=1}^{n_h}\Lambda_0(y_{hi})\exp\left\{\boldsymbol{\beta}^\top\boldsymbol{x}_{hi}\right\}\Big)}{\mathcal{L}^{(d_h)}\Big(\sum_{i=1}^{n_h}\Lambda_0(y_{hi})\exp\left\{\boldsymbol{\beta}^\top\boldsymbol{x}_{hi}\right\}\Big)}, \tag{7.6}$$

which can be seen from Appendix A.13, together with

$$\mathsf{E}\big[U^k\exp(-Us)\big]=(-1)^k\mathcal{L}^{(k)}(s). \tag{7.7}$$

### 7.1.2 Case study

We illustrate the `parfm` package with the very well-known `kidney` dataset that contains the recurrence times to kidney infection for 38 patients using portable dialysis equipment (McGilchrist and Aisbett, 1991).

```
R> R.Version()[["version.string"]]
[1] "R version 2.15.2 (2012-10-26)"

R> library("parfm")
R> packageDescription("parfm", fields="Version")

[1] "2.5.2"
```

The dataset is available in `parfm` via the command `data("kidney")` and it looks like the following:

```
R> head(kidney)

id time status age sex disease frail
1  1    8      1  28   1   Other   2.3
2  1   16      1  28   1   Other   2.3
```

```
3  2   23      1  48   2      GN   1.9
4  2   13      0  48   2      GN   1.9
5  3   22      1  32   1   Other   1.2
6  3   28      1  32   1   Other   1.2
```

Each observation corresponds to a kidney, the variable `id` being the patient's code. The time from insertion of the catheter to infection or censoring is stored in `time` while `status` is 1 when infection has occurred and 0 for censored observations (catheters may be removed for reasons other than infection). Three covariates are available: `age`, the age of the patient in years, `sex`, being `1` for males and `2` for females, and `disease`, the disease type (`GN`, `AN`, `PKD` or `Other`). Finally `frail` is the frailty prediction from the original paper which fits a semi-parametric lognormal frailty model.

First and foremost, `sex` is recoded as a 0–1 indicator for ease of interpretation:

```
R> kidney$sex <- kidney$sex - 1
```

The hazard of infection will be modelled as a function of the patient's age and sex. Clearly, kidneys from the same patient cannot be considered independent. Therefore, the use of a shared frailty model is advisable, with clusters of size 2 corresponding to patients.

The `parfm()` function must have the following inputs. `formula`: a formula with an object of class `Surv` on the left-hand side; `cluster`: the cluster variable's name; `data`: the dataset; `dist`: the baseline hazard, either `exponential`, `weibull`, `gompertz`, `lognormal` or `loglogistic`; `frailty`: the frailty distribution, either `none`, `gamma`, `possta` or `ingau`.

**Model estimation.** The model with exponential baseline hazard and gamma frailty distribution is first fitted.

```
R> mod <- parfm(Surv(time, status) ~ sex + age, cluster="id",
+               data=kidney,
+               dist="exponential", frailty="gamma")
R> mod

Execution time: 1.15 second(s)

Frailty distribution: Gamma
Baseline hazard distribution: Exponential
Loglikelihood: -333.248

       ESTIMATE SE    p-val
theta   0.301   0.157
lambda  0.025   0.015
sex    -1.485   0.398 0.000 ***
age     0.005   0.011 0.663
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Kendall's Tau: 0.131
```

Standard errors are computed as the square roots of the diagonal elements of the observed information matrix. According to this model, `sex` has a significant impact on the hazard of infection while it is not affected by `age`. Conditional on the patient's frailty and on the age, the hazard of infection for a female at any time $t$ is estimated to be $\exp(-1.485) \approx 0.227$ times that of a male, with Wald confidence interval

```
> ci.parfm(mod, level=0.05)["sex", ]

low    up
0.104 0.495
```

Estimated hazard ratios can be plotted with their 95% confidence interval, simply as `plot(mod)`, with black graphs for HRs significantly different from 1 and grey for those which are not (Fig. 7.2). As for the heterogeneity parameter, it is estimated to be 0.301 which corresponds to a Kendall's tau equal to 0.131.

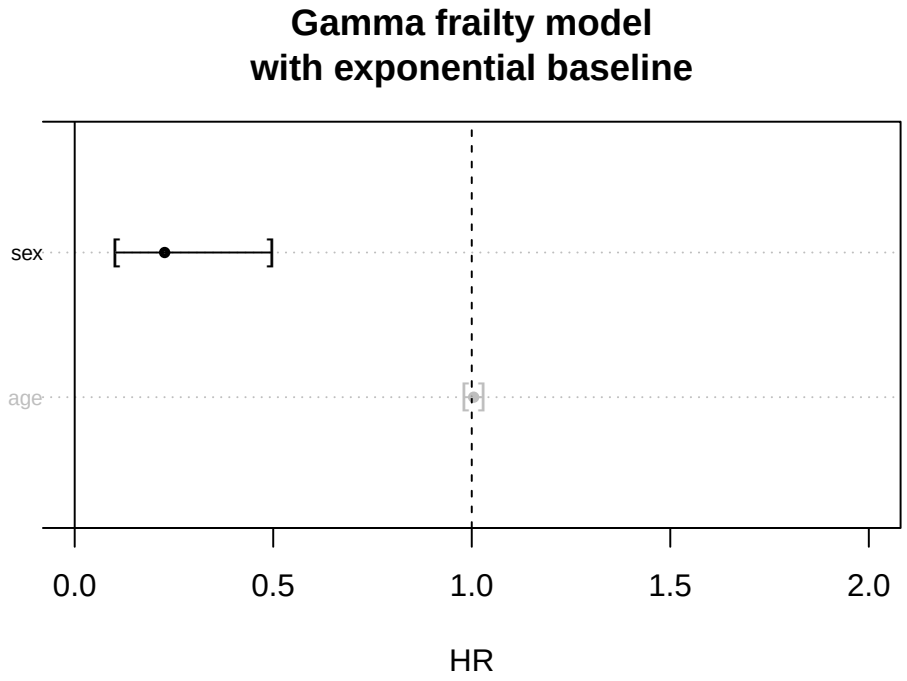



Figure 7.2 – Estimated hazard ratios can be plotted with 95% confidence interval for the kidney dataset. Parametric gamma–exponential shared frailty model.

**Frailty prediction.** Prediction of frailties can be obtained via the `predict()` function, with the parametric frailty model object as unique argument. For instance, the predictions for the gamma–exponential model, `mod`, are obtained via the command

```
R> u <- predict(mod)
```

which returns an object of class `predict.parfm`. These predictions can easily be plotted (Figure 7.3) with the command `plot(u, sort="i")`.

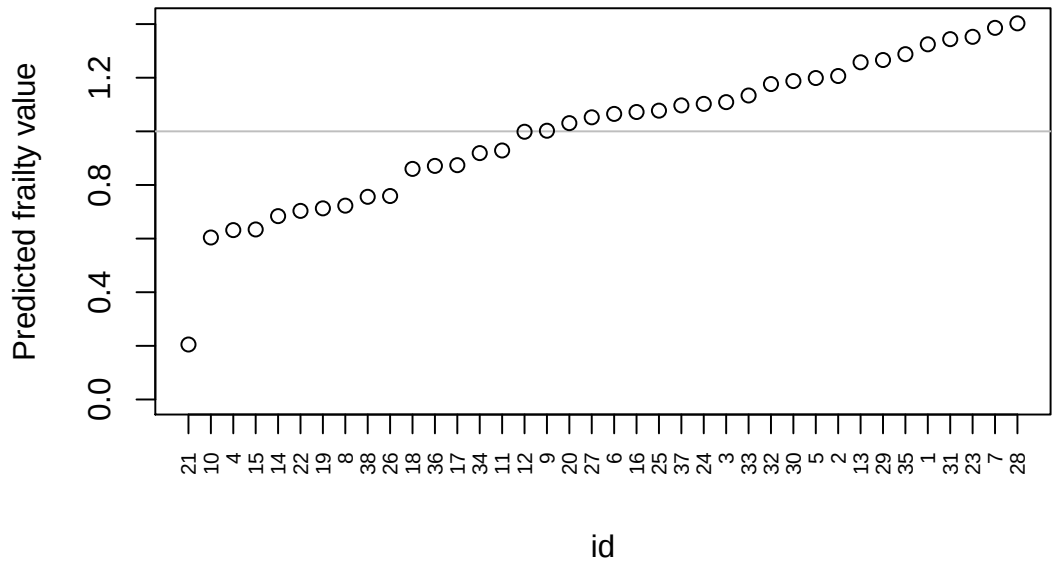



Figure 7.3 – Prediction of frailties for the kidney dataset. Parametric gamma–exponential shared frailty model.

**Comparison of different models.** In some circumstances, it might be useful to easily obtain AIC and BIC values for a series of candidate models. This can be done using the `select.parfm()` function. Its use is similar to that of the `parfm()` function, but the `dist` and `frailty` values are vectors that contain all the alternatives to try.

```
R> kidney.parfm <-
+    select.parfm(Surv(time, status) ~ sex + age,
+                 cluster="id", data=kidney,
+                 dist=c("exponential", "weibull", "gompertz",
+                        "loglogistic", "lognormal"),
+                 frailty=c("gamma", "ingau",
+                           "possta", "lognormal"))
R> kidney.parfm

             gamma   ingau  possta  lognor
exponential 674.496 675.699 682.264 675.212
weibull     674.376 676.627 682.315 675.726
gompertz    676.496 677.699 684.264 677.212
```

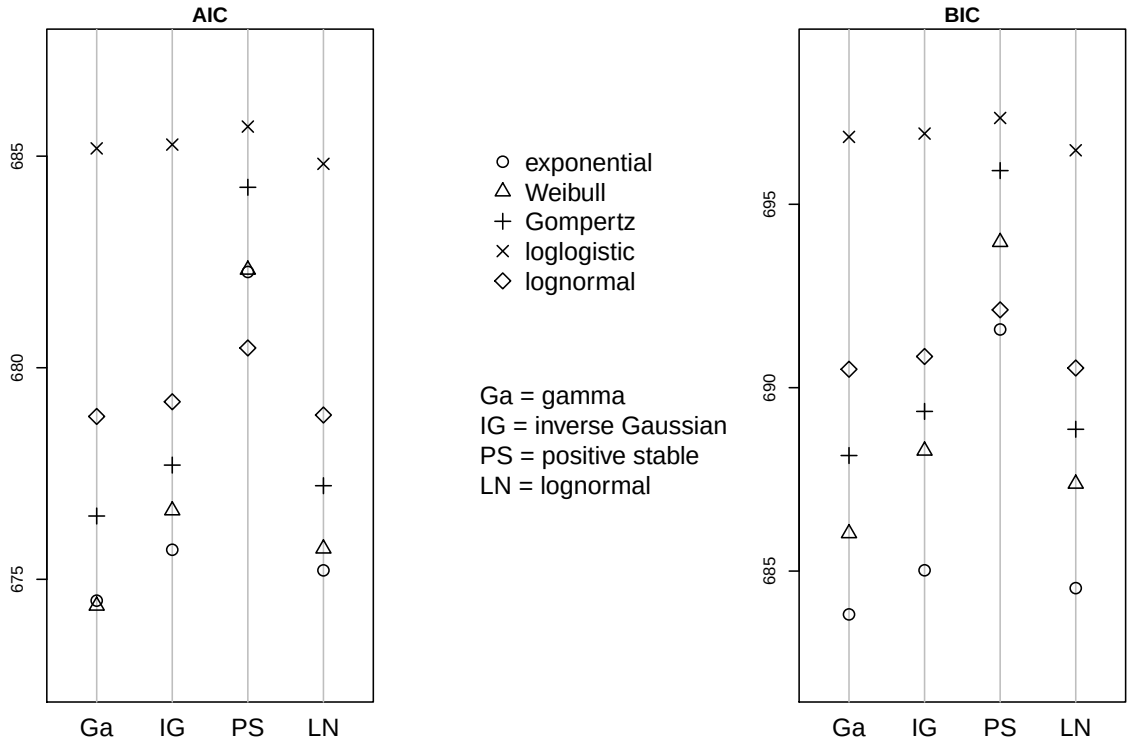



Figure 7.4 – AIC and BIC values of `parfm` models for the kidney dataset.

```
loglogistic 685.184 685.274 685.699 684.818
lognormal   678.849 679.196 680.467 678.882


BIC:
              gamma   ingau  possta  lognor
exponential 683.819 685.022 691.587 684.535
weibull     686.029 688.281 693.969 687.379
gompertz    688.150 689.353 695.918 688.866
loglogistic 696.837 696.927 697.353 696.472
lognormal   690.502 690.850 692.121 690.536
```

The results can be plotted (Figure 7.4) via the command `plot(kidney.parfm)`. In this particular example, the exponential baseline seems to be a good candidate.

As a comparison, the model with inverse Gaussian distributed frailties is fitted by changing the `frailty` argument into `'ingau'`.

```
R> parfm(Surv(time, status) ~ sex + age, cluster="id",
+        data=kidney, dist="exponential", frailty="ingau")

Execution time: 1.15 second(s)
```

```
Frailty distribution: Inverse Gaussian
Baseline hazard distribution: Exponential
Loglikelihood: -333.85

       ESTIMATE SE    p-val
theta   0.375   0.259
lambda  0.022   0.013
sex    -1.310   0.373 0.000 ***
age     0.004   0.011 0.694
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Kendall's Tau: 0.125
```

In this case, the conclusions drawn from the previous two models are essentially analogous.

Consider now the model with the positive stable frailty distribution. In this example, it converges to a solution which is not valid ($\nu = 0$, where $\nu$ denotes $\theta$ for the positive stable model in `parfm`) with the default settings.

```
R> parfm(Surv(time, status) ~ sex + age, cluster="id",
+        data=kidney, dist="exponential", frailty="possta")

Execution time: 1.16 second(s)

Frailty distribution: Positive Stable
Baseline hazard distribution: Exponential
Loglikelihood: -337.132

       ESTIMATE SE p-val
nu      0.000
lambda  0.012
sex    -0.885
age     0.004
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Kendall's Tau: 0
Warning message:
In parfm(Surv(time, status) ~ sex + age, cluster = "id",
  data = kidney,  :
  Error in solve.default(res$hessian) :
  Lapack routine dgesv: system is exactly singular
```

The default initial value for $\nu$ is $1/2$ in the case of positive stable frailties; it can be changed by means of the `iniFpar` option in `parfm()`. Let us try with $\nu = 0.25$.

```
R> parfm(Surv(time, status) ~ sex + age, cluster="id",
+        data=kidney, dist="exponential", frailty="possta",
+        iniFpar=0.25)

Execution time: 1.71 second(s)

Frailty distribution: Positive Stable
Baseline hazard distribution: Exponential
Loglikelihood: -336.182

       ESTIMATE SE    p-val
nu      0.112   0.084
lambda  0.014   0.008
sex    -0.951   0.348 0.006 **
age     0.004   0.011 0.698
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Kendall's Tau: 0.112
```

The problem might also be fixed by changing the optimisation method (see `optim()`). By default it is set to `'BFGS'`, but it can be changed through the `method` option.

```
R> parfm(Surv(time, status) ~ sex + age, cluster="id",
+        data=kidney, dist="exponential", frailty="possta",
+        method="Nelder-Mead")

Execution time: 1.51 second(s)

Frailty distribution: Positive Stable
Baseline hazard distribution: Exponential
Loglikelihood: -336.182

       ESTIMATE SE    p-val
nu      0.112   0.084
lambda  0.014   0.008
sex    -0.951   0.348 0.006 **
age     0.004   0.011 0.694
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Kendall's Tau: 0.112
```

In this example, the results obtained by changing the optimisation method are the same as those obtained by changing the initial value of $\nu$. When convergence problems occur, using different

starting values and/or different optimisation methods is generally sufficient to find the global maximum of the marginal likelihood function.

Notice that, as results can be slightly different with different initial parameter values, this could influence models comparison based on AIC and BIC values. In general, once the most interesting models are selected, finer comparisons should be done with several different starting values for parameters.

Finally, we provide a comparison with the semi-parametric model. As an example, we fit the semi-parametric model with gamma frailties via the `coxph()` function.

```
R> coxph(Surv(time, status) ~ sex + age +
+        frailty(id, distribution="gamma", eps=1e-11),
+        outer.max=15, data=kidney)

          coef     se(coef) se2    Chisq DF   p
sex       -1.58323 0.4594   0.3515 11.88  1.0 0.00057
age        0.00522 0.0119   0.0088  0.19  1.0 0.66000
frailty                            22.96 12.9 0.04100

      Variance of random effect= 0.408   I-likelihood = -181.6
```

Estimates of regression parameters are quite similar to those of the exponential–gamma model, while the frailty variance is sensibly different, arguably because of the difference in how the baseline hazard is treated.

## 7.2 `mlfm` package: multilevel frailty models in `R`

Frailty models account for unobserved risk factors, common to all subjects belonging to groups, such as hospitals, families, countries. Sometimes, several grouping criteria play a role in generating unobserved heterogeneity. For instance, in multicenter clinical trials one could suspect that, due to different national health care systems, clustering at country level should be taken into account in addition to that at hospital level. In this case, one grouping criterion — the hospital — is nested within the other one — the country. In other situations, different criteria can cross each other, without a hierarchy; in meta-analysis of multicenter studies, there can be the need of accounting for clustering generated by both the trial and the hospital. In this case, each study involves many centers, but each center can enter several studies.

Multilevel frailty models are proportional hazard models with many random effects, accounting for different grouping criteria. Two levels models are considered in discrete time by Manda and Meyer (2005), whereas Yau (2001) and Sastry (1997) studied two nested random effects with log-normal and gamma mixing distributions, respectively. Horny (2009) proposes the EM-PL procedure, alternating an EM and a PPL algorithm, to fit semiparametric proportional hazard models with any number $K$ of grouping criteria, which is valid for any frailty distribution admitting a Laplace transform.

### 7.2.1 Model estimation

The conditional hazard of a general multilevel frailty model is

$$\lambda\left(t \mid u^{(1)}, \ldots, u^{(K)}\right) = \lambda_0(t)\left(\prod_{k=1}^{K} u^{(k)}\right)\exp\left\{\boldsymbol{\beta}^\top \mathbf{x}\right\}, \tag{7.8}$$

where $K$ is the total number of grouping criteria and $u^{(k)}$ is the frailty term for criterion $k$.

**The algorithm.** The EM-PL algorithm that Horny (2009) proposed for model (7.8) is implemented for gamma frailties in the new `mlfm` package (Rotolo and Horny, 2012) as follows.

**Initialisation:**

- regression parameters $\boldsymbol{\beta}$ are initialized to estimates of a Cox model without frailties;
- for each level $k = 1, \ldots, K$, the frailty variance $\theta_k$ is estimated via a semiparametric shared frailty model, with only one clustering level. Previsions for frailties $u_h^{(k)}$ are stored as initial values as well;

**EM iteration:**

E: for each level $k = 1, \ldots, K$, previsions $\tilde{u}_h^{(k)}$ for frailties are obtained by a semiparametric frailty model, fitted via PPL algorithm, with a frailty for level $k$; frailties for other levels are fixed to their provisional values and introduced in the model as (log)offsets;

M: – regression parameters $\boldsymbol{\beta}$ are estimated via a Cox model with previsions of frailties as (log)offsets;

– frailty parameters $\boldsymbol{\theta}$ are estimated by maximising the frailty distribution as evaluated in the provisional values $\tilde{u}_h^{(k)}$.

**Standard errors.** As the EM-PL algorithm does not provide an estimate of the information matrix, Horny (2009) proposes to recover it using Louis' (1982) methodology. Nevertheless, this approach is difficult to implement in a general way. The well-known `coxph()` function, (`survival` package, Therneau, 2012b), estimates standard errors for the shared frailty model by fixing the frailty parameter. In the case of multilevel frailties, BLUPs cannot be obtained for the model with all the frailties at the same time, so the `mlfm()` function computes standard errors by fixing the frailties themselves at their predicted values. Since this approach leads to slightly underestimated standard errors (Therneau and Grambsch, 2000, Sec. 9.5.2), Massonnet (2008, Ch. 3) proposed and studied in details bootstrap methods for the shared frailty model. An analogous approach for the multilevel frailty model is not straightforward, and it would be an interesting subject for further research.

To give an idea of the effect of this estimation bias, we consider a small simulation example. Five hundred datasets of size 2000 are simulated with baseline hazard $\mathsf{Wei}(\lambda = 4, \rho = 1.2)$ and four frailty terms, with total variance 1.2. The effect of two covariates is added: a dichotomous treatment $\mathsf{Treat} \sim \mathsf{B}(0.5)$ with regression parameter $\beta_{\mathsf{Treat}} = 0.5$, and the patient age $\mathsf{Age} \sim \mathsf{N}(\mu = 60, \sigma = 4)$ with $\beta_{\mathsf{Age}} = \log(1.2)/10 \simeq 0.02$. Censoring is added at fixed time 1.5.

Results shown in Table 7.1 confirm that the standard errors are underestimated: $\widehat{\mathsf{SE}(\hat{\beta})} < \mathsf{ESE}(\hat{\beta})$.

**No censoring**

| | True $\beta$ | Mean $\hat{\beta}$ | Mean $\widehat{\mathsf{SE}}(\hat{\beta})$ | Emp. SE$(\hat{\beta})$ |
|---|---|---|---|---|
| **Treat** | 0.5000 | 0.5049 | 0.0454 | 0.0515 |
| **Age** | 0.0182 | 0.0181 | 0.0056 | 0.0064 |

**30% of censoring**

| | True $\beta$ | Mean $\hat{\beta}$ | Mean $\widehat{\mathsf{SE}}(\hat{\beta})$ | Emp. SE$(\hat{\beta})$ |
|---|---|---|---|---|
| **Treat** | 0.5000 | 0.4963 | 0.0591 | 0.0631 |
| **Age** | 0.0182 | 0.0176 | 0.0073 | 0.0083 |

**60% of censoring**

| | True $\beta$ | Mean $\hat{\beta}$ | Mean $\widehat{\mathsf{SE}}(\hat{\beta})$ | Emp. SE$(\hat{\beta})$ |
|---|---|---|---|---|
| **Treat** | 0.5000 | 0.5014 | 0.0503 | 0.0547 |
| **Age** | 0.0182 | 0.0178 | 0.0063 | 0.0069 |

Table 7.1 – Evaluation of performance of SE estimation in `mlfm()`. Mean estimated regression parameters as compared to true values, and average standard errors as compared to empirical standard errors. 500 simulations of datasets of size 2000 with 0%, 30% and 60% of censored data.

| | Nominal | No cens. | 30% cens. | 60% cens. |
|---|---|---|---|---|
| **Treat** | 95% | 92.0% | 95.2% | 93.0% |
| **Age** | | 92.4% | 91.2% | 92.4% |

Table 7.2 – Evaluation of performance of SE estimation in `mlfm()`. Coverage probabilities. 500 simulations of datasets of size 2000 with 0%, 30% and 60% of censored data.

Nevertheless, the size of the error is very small and can be accepted, alongside the awareness of the issue.

The underestimated standard errors give origin to confidence intervals whose coverage probabilities at nominal level of 95% are shown in Table 7.2. Under the assumption that the coverage probability is correct (0.95), the approximated 95% confidence interval for its estimator over 500 replications is $[93.4\%, 96.6\%]$. Therefore, underestimation of SEs is significant in all but one case. For this latter, there is no evidence of departure from the nominal probability.

### 7.2.2 Case study

The bladder cancer dataset presented in Section 6.1 contains data from seven different multicenter clinical trials (EORTC trials 30781, 30782, 30791, 30831, 30832, 30845, 30863) recruiting 2523 patients from 63 hospitals.

Hospitals recruited 1 to 219 patients, with median size of 30 patients. Studies consist of 266 to 450 observations, with a median of 373 patients.

Each study is a multicenter trial, so collects information from many hospitals. Conversely, many of the hospitals participated to several trials. As an example, we show here the number of patients for some hospitals, grouped by trial.

```
> head(table(bladder$center, bladder$studyN))

         30781 30782 30791 30831 30832 30845 30863
  A 0903     0     0    20    13     0     6     2
  B 0101     0     0     0     0     0     0     1
  B 0104     0    31     0    41     0    12    17
  B 0107     0     0     0     0     0     0    17
  B 0108     0    28     5    71     0     0    11
  B 0109     0    28     0     5     0     0     0
```

```
> tail(table(bladder$center, bladder$studyN))

          30781 30782 30791 30831 30832 30845 30863
  UK 0615    15     0     0     0     0     0     0
  UK 0621    11     0     0     0     0     0     0
  UK 0623    22     0     0     0     0     0     0
  UK 0629     0     0     0     0     0     2     8
  UK 0630     3     0     0     0     0     0     0
  UK 0649     0     0     0    13     0     0     0
```

On the basis of preliminary analyses, not shown here, and biological considerations three covariates other than treatment (`treat`) are retained in the model: an indicator of the tumour size being more than 1 cm (`tSize`), an indicator of the number of tumours being more than 1 (`tNum`), and an indicator for recurrent diseases (`tStat`). Two simple frailty models accounting separately for the center and the study effect are first fitted.

```
> coxph(Surv(tRec / 12, Rec) ~ treat + tSize + tNum + tStat +
+         frailty(center), eps=1e-10, data=bladder)

                coef   se(coef) se2    Chisq  DF   p
treatTRUE       -0.194 0.0666   0.0642   8.48  1.0 3.6e-03
tSize(1, 9] cm   0.255 0.0651   0.0643  15.38  1.0 8.8e-05
tNum>1           0.571 0.0666   0.0659  73.54  1.0 0.0e+00
tStatRecurrent   0.367 0.0729   0.0713  25.32  1.0 4.9e-07
frailty(center)                        111.65 33.7 2.8e-10

Variance of random effect= 0.114   I-likelihood = -8484.2
```

```
> coxph(Surv(tRec / 12, Rec) ~ treat + tSize + tNum + tStat +
+         frailty(studyN), eps=1e-10, data=bladder)

                coef    se(coef) se2    Chisq DF   p
treatTRUE       -0.0249 0.1900   0.1077  0.02 1.00 9.0e-01
```

```
tSize(1, 9] cm   0.2507 0.0624   0.0624 16.13 1.00 5.9e-05
tNum>1           0.5764 0.0698   0.0689 68.25 1.00 1.1e-16
tStatRecurrent   0.4218 0.0761   0.0752 30.75 1.00 2.9e-08
frailty(studyN)                         21.85 5.79 1.1e-03

Variance of random effect= 0.095   I-likelihood = -8496.9
```

These two models give similar results, except for the treatment effect. It is significantly effective according to the first model, whereas no evidence of benefit is present when only the study heterogeneity is accounted for. To take into account both sources of heterogeneity simultaneously, a model (`Model_1`) with two frailties is fitted for time to death, one for the trial id (`studyN`) and another for the cluster id (`center`).

```
> library(mlfm)
> packageDescription("mlfm")$Version
[1] "1.2-2"

> Model_1 <- mlfm(Surv(tRec / 12, Rec) ~ treat + tSize + tNum + tStat +
+                   frailty(center) + frailty(studyN),
+                 maxit=100, eps=1e-10, data=bladder)

Execution time: 18.24 secs

> Model_1
                  coef      se         p
treatTRUE       -0.157  0.0596  8.65e-03
tSize(1, 9] cm   0.258  0.0626  3.68e-05
tNum>1           0.567  0.0642  9.99e-19
tStatRecurrent   0.396  0.0673  3.89e-09

Iterations: 41 EM, 1195 outer, 3160 Newton-Raphson
     Variances of random effects
      center  0.0909
      studyN  0.0117
Total frailty variance: 0.104
```

On the basis of this model, multiple tumours have the strongest impact on the death hazard, with an increase which can range from 55% to 100% at a 5% level.

```
> signif(t(mapply(function(be, se) {
+   exp(be + qnorm(.025) * se * c("lower95%ci" = 1, HR = 0, "upper95%ci" = -1))
+ }, be = Model_1$beta, se = Model_1$ses)), digits = 3)

               lower95%ci    HR upper95%ci
treatTRUE           0.761 0.855      0.961
```

```
tSize(1, 9] cm      1.150 1.290      1.460
tNum>1              1.550 1.760      2.000
tStatRecurrent      1.300 1.490      1.700
```

Recurrent diseases and big tumours also play an analogous role, but at a lower degree. On the other side, the treatment has a significantly beneficial effect with an estimated reduction of almost 15% of death risk.

The total unobserved dependence is estimated as a total frailty variance of 0.104, corresponding to $\tau = 0.049$. This unobserved heterogeneity is mainly due to differences between hospital, as the frailty variance for center effect is about 7.8 times that for trial effect. It is interesting to compare this model with the one with only the treatment as covariate.

```
> Model_2 <- mlfm(Surv(tRec / 12, Rec) ~ treat +
+   frailty(center) + frailty(studyN),
+                 maxit=100, eps=1e-10, data=bladder)

Execution time: 15.56 secs


> signif(rbind(Model1 = Model_1$theta,
+              Model2 = Model_2$theta), digits = 2)

       center studyN
Model1  0.091  0.012
Model2  0.100  0.033
```

The frailty variance for center effect is substantially unchanged, whereas the one for the trial effect is almost three times the one for `Model_1`. This suggests that the populations of the different studies are quite different in terms of cancer features — size, number and state — and this has an important impact on mortality. On the contrary, the populations of different hospitals seem to be quite homogeneous and the small residual unobserved heterogeneity probably depends on other factors which are completely unavailable in the dataset at hand.

# Chapter 8

# Conclusions

Even though multicenter clinical trials are quite common in cancer research, usually the potential of jointly studying event histories remains largely unexploited and the problem of unobserved heterogeneity in survival data is often ignored. Nevertheless, interest in statistical methods for controlling dependence due to interrelated events and to clustering is continuously increasing. Multi-state and frailty models have been receiving more and more attention by both methodological and applied research, and a growing literature on these subjects is rapidly emerging.

The main objectives of this thesis were to explore the problem of between-events and between-subjects dependence in survival data, to provide useful tools to deal with them, and to propose a solution to take advantage of both multi-state and frailty methodologies jointly.

First, a simulation procedure has been proposed to generate multi-state survival data, possibly with unobserved heterogeneity and extra-Markov dependence. A tuning algorithm for simulation parameters allows to generate real-like data even for non-trivial multi-state structures and in presence of different risk factors. Simulation studies, missing in multi-state literature up to now, have been provided in this thesis, showing two main results: multi-state models are very robust with respect to the Markov assumption and, like simple proportional hazards models, they suffer from the attenuation effect of regression parameters in the case of unobserved heterogeneity.

Two new `R` packages for frailty modelling have been presented. The first one (`parfm`) offers a unified framework for parametric frailty models. They are quite common in literature, but the new package is the first one providing inference for a wide range of parametric families for both the baseline hazard function (five) and the frailty distribution (four), including the very popular gamma and lognormal. The package offers a number of flexible tools for inference, prediction, results plotting and models comparison. The second package (`mlfm`) implements the EM-PL estimation method for semiparametric multilevel frailty models, which is an essential tool to fit multi-state models with nested frailties, studied in this thesis. The estimation algorithm has been previously proposed and discussed in literature, but no publicly available software implemented it for users.

The problem of incorporating frailties into multi-state models is of primary interest both from a methodological and an applied point of view; though, research on this topic is just moving its first steps. This is still a budding research field, open to broad-range possibilities and presenting many unexplored issues. We have reviewed the (quite limited) literature concerning possible solutions responding to the need of accounting for unobserved heterogeneity in multi-state survival data

and we have proposed a model including frailties correlated between events of different types within the same cluster. The use of a nested hierarchy of random effects allows to obtain positive dependence, while keeping the structure of the hazard simple enough to make inference feasible. We have considered two estimation approaches, a parametric and a semiparametric one, and we have investigated their performances via simulation studies. Parametric methods turned out to reduce estimators bias, but approximation of integrals needed to marginalise the likelihood was not precise at all, so that the variability of the estimates was too large. Semiparametric inference, on the contrary, performed well and proved to be a competitive candidate to consistently estimate regression parameters, which is usually the main focus of interest in clinical trials comparing different treatments. Estimation of heterogeneity parameters is the weak point of these models, as the use of two multiplicative frailty terms seems to generate a sort of identification issue: multi-state models with nested frailties can catch and control for unobserved heterogeneity at the two levels, but they encounter serious problems in assigning to each level the correct portion of variability. Hence, the researcher should be aware that such models are a good tool to consistently estimate the effect of covariates in presence of clustering, but he should avoid to employ them to investigate the possible sources of dependence and to ascribe the influence of such sources either to the group effect or to its interaction with the transition type.

Two applications, to a bladder and a prostate cancer multicenter clinical trials, have been presented, in which nested frailty multi-state models provide new and more detailed results with respect to previous publications. The use of a multi-state structure has allowed to study the treatment effect on death, taking into account and evaluating the effect of intermediate events. At the same time, the presence of frailties has allowed to reduce the attenuation effect due to clustering, typical of collaborative studies.

# References


Aalen OO (1976). "Nonparametric inference in connection with multiple decrement models." *Scandinavian Journal of Statistics*, **3**(1), 15–27. [◀ p. 17, 30]

Aalen OO (1992). "Modelling heterogeneity in survival analysis by the compound Poisson distribution." *The Annals of Applied Probability*, **2**(4), 951–972. doi:10.1214/aoap/1177005583. [◀ p. 38]

Aalen OO, Borgan, Gjessing HK (2008). *Survival and event history analysis: a process point of view*. Springer Verlag. doi:10.1007/978-0-387-68560-1. [◀ p. 5, 9]

Amorim AP, de Uña-Alvarez J, Meira-Machado L (2011). "Presmoothing the transition probabilities in the illness–death model." *Statistics and Probability Letters*, **81**, 797–806. doi:10.1016/j.spl.2011.02.017. [◀ p. 43]

Andersen PK, Borgan, Gill RD, Keiding N (1993). *Statistical Models Based on Counting Processes*. Springer Series in Statistics, 2nd edition. Springer Verlag. doi:10.1007/978-1-4612-4348-9. [◀ p. 9, 15, 16, 18, 61]

Andersen PK, Keiding N (2002). "Multi-state models for event history analysis." *Statistical Methods In Medical Research*, **11**, 91–115. doi:10.1191/0962280202SM276ra. [◀ p. 1, 5, 6, 12]

Andersen PK, Klein JP, Zhang MJ (1999). "Testing for centre effects in multi-centre survival studies: a Monte Carlo comparison of fixed and random effects tests." *Statistics in Medicine*, **18**(12), 1489–1500. doi:10.1002/(SICI)1097-0258(19990630)18:12%3C1489::AID-SIM140%3E3.0.CO;2-%23. [◀ p. 19, 67]

Bender R, Augustin T, Blettner M (2005). "Generating survival times to simulate Cox proportional hazards models." *Statistics in Medicine*, **24**(11), 1713–23. doi:10.1002/sim.2059. [◀ p. 43]

Beyersmann J, Latouche A, Buchholz A, Schumacher M (2009). "Simulating competing risks data in survival analysis." *Statistics in Medicine*, **28**(6), 956–71. doi:10.1002/sim.3516. [◀ p. 43]

Bhattacharyya M, Klein JP (2005). "A random effects model for multistate survival analysis with application to bone marrow transplants." *Mathematical Biosciences*, **194**(1), 37–48. doi:10.1016/j.mbs.2004.07.005. [◀ p. 6, 55]

Bouffioux CH, Denis L, Oosterlinck W, Viggiano G, Vergison B, Keuppens F, De Pauw M, Silvester R, Cheuvart B (1992). "Adjuvant chemotherapy of recurrent superficial transitional cell

carcinoma: results of a European Organization for Research on Treatment of Cancer randomized trial comparing intravesical instillation of thiotepa, doxorubicin and cisplatin." *The Journal of Urology*, **148**(2). [◂ p. 73]

Bouffioux CH, Kurth KH, Bono A, Oosterlinck W, Kruger CB, De Pauw M, Sylvester R (1995). "Intravesical adjuvant chemotherapy for superficial transitional cell bladder carcinoma: results of 2 European Organization for Research and Treatment of Cancer randomized trials with mitomycin C and doxorubicin comparing early versus delayed instillations and short-term versus long-term treatment. European Organization for Research and Treatment of Cancer Genitourinary Group." *The Journal of Urology*, **153**(3 Pt 2), 934. [◂ p. 73]

Breslow NE (1972). "Discussion on Professor Cox's paper." *Journal of the Royal Statistical Society. Series B (Methodological)*, **34**, 216–217. [◂ p. 17]

Breslow NE, Clayton DG (1993). "Approximate inference in generalized linear mixed models." *Journal of the American Statistical Association*. doi:10.2307/2290687. [◂ p. 31]

Bretagnolle J, Huber-Carol C (1988). "Effects of Omitting Covariates in Cox's Model for Survival Data." *Scandinavian Journal of Statistics*, **15**(2). [◂ p. 67]

Broström G (2012). "`eha`: Event History Analysis." `R` package version 2.2-0. [◂ p. 89, 90]

Burton A, Altman DG, Royston P, Holder RL (2006). "The design of simulation studies in medical statistics." *Statistics in Medicine*, **25**(24), 4279–4292. doi:10.1002/sim.2673. [◂ p. 43]

Buyse ME, Staquet MJ, Sylvester RJ (1984). *Cancer clinical trials: methods and practice*. Oxford University Press, USA. [◂ p. 5]

Clayton DG (1978). "A Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence." *Biometrika*, **65**(1), 141–151. [◂ p. 1, 45]

Collett D (2003). *Modelling survival data in medical research*. Chapman & Hall/CRC Texts in Statistical Science. CRC press. [◂ p. 9]

Commenges D (1999). "Multi-state models in epidemiology." *Lifetime Data Analysis*, **5**(4), 315–327. doi:10.1191/0962280202SM276ra. [◂ p. 43]

Cook RJ, Lawless JF (2002). "Analysis of repeated events." *Statistical Methods In Medical Research*, **11**, 141–166. doi:10.1191/0962280202sm278ra. [◂ p. 11]

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**(8), 3913–3930. doi:10.1016/j.csda.2006.03.009. [◂ p. 29, 30, 43]

Cox DR (1972). "Regression models and life-tables." *Journal of the Royal Statistical Society. Series B (Methodological)*, **34**(2), 187–220. [◂ p. 1, 10, 29, 61]

de Bono JS, Oudard S, Ozguroglu M, Hansen S, Machiels JP, Kocak I, Gravis G, Bodrogi I, Mackenzie MJ, Shen L, Roessner M, Gupta S, Sartor AO (2010). "Prednisone plus cabazitaxel or mitoxantrone for metastatic castration-resistant prostate cancer progressing after docetaxel treatment: a randomised open-label trial." *The Lancet*, **376**(9747), 1147–1154. ISSN 01406736. doi:10.1016/S0140-6736(10)61389-X. [◀ p. 78, 79, 85]

de Uña-Álvarez J, Meira-Machado LF (2008). "A simple estimator of the bivariate distribution function for censored gap times." *Statistics & Probability Letters*, **78**(15), 2440–2445. doi:10.1016/j.spl.2008.02.031. [◀ p. 43]

de Wreede LC, Fiocco M, Putter H (2010). "The `mstate` package for estimation and prediction in non- and semi-parametric multi-state and competing risks models." *Computer Methods and Programs in Biomedicine*, **99**(3), 261–74. doi:10.1016/j.cmpb.2010.01.001. [◀ p. 15, 16, 18, 43, 61]

de Wreede LC, Fiocco M, Putter H (2011). "`mstate`: An `R` Package for the Analysis of Competing Risks and Multi-State Models." *Journal Of Statistical Software*, **38**(7), 1–30. [◀ p. 15, 16, 89]

Dempster AP, Laird NM, Rubin DB (1977). "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the Royal Statistical Society. Series B (Methodological)*, **39**(1), 1–38. [◀ p. 29]

Donohue M, Xu R (2012). "`phmm`: Proportional Hazards Mixed-Effects Model." `R` package version 0.7-4. [◀ p. 89, 90]

Duchateau L, Janssen P (2008). *The Frailty Model*. Springer Verlag, New York, NY. ISBN 978-0-387-72834-6. doi:10.1007/978-0-387-72835-3. [◀ p. 1, 6, 15, 19, 21, 24, 28, 29, 32, 34, 35, 36, 37, 38, 90]

Duchateau L, Janssen P, Kezic I, Fortpied C (2003). "Evolution of recurrent asthma event rate over time in frailty models." *Journal Of The Royal Statistical Society. Series C (Applied Statistics)*, **52**(3), 355–363. doi:10.1111/1467-9876.00409. [◀ p. 20]

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**(3), 603–620. doi:10.1016/S0167-9473(02)00057-9. [◀ p. 20, 29, 30]

Ducrocq V, Casella G (1996). "A Bayesian analysis of mixed survival models." *Genetics Selection Evolution*, **28**, 505–529. [◀ p. 31]

Economou P, Caroni C (2005). "Graphical tests for the assumption of gamma and inverse Gaussian frailty distributions." *Lifetime Data Analysis*, **11**(4), 565–582. [◀ p. 35]

Farlie DJG (1960). "The performance of some correlation coefficients for a general bivariate distribution." *Biometrika*, **47**(3/4), 307–323. [◀ p. 43]

Fine JP, Gray RJ (1999). "A proportional hazards model for the subdistribution of a competing risk." *Journal of the American Statistical Association*, **94**(446), 496–509. [◀ p. 11]

Fleiss JL (1982). "Multicentre clinical trials: Bradford Hill's contributions and some subsequent developments." *Statistics in Medicine*, **1**(4), 353–359. doi:10.1002/sim.4780010413. [◀ p. 5]

Gill RD (2001). "Product integration." *Encyclopedia of Biostatistics*, **5**, 3530–3534. [◀ p. 18]

Glidden DV, Vittinghoff E (2004). "Modelling clustered survival data from multicentre clinical trials." *Statistics in Medicine*, **23**(3), 369–388. doi:10.1002/sim.1599. [◀ p. 6, 21]

Gonzalez JR, Rondeau V, Mazroui Y, Diakite A (2012). "`frailtypack`: General Frailty Models Using a Semi-Parametrical Penalized Likelihood Estimation or a Parametrical Estimation." `R` package version 2.2-27. [◀ p. 60, 89, 90]

Goutis C, Casella G (1999). "Explaining the Saddlepoint Approximation." *The American Statistician*, **53**(3), 216–224. ISSN 00031305. doi:10.2307/2686100. [◀ p. 60, 130]

Gray RJ (1988). "A class of K-sample tests for comparing the cumulative incidence of a competing risk." *The Annals of Statistics*, **16**(3), 1141–1154. [◀ p. 11, 43]

Greenwood M (1926). "The natural duration of cancer." *Reports on Public Health and Medical Subjects*, **33**(1-26), 1–26. [◀ p. 17]

Grillo Ruggieri F, Pace MP, Bunkeila F, Cartei F, Panizza BM, Fabbietti L, Moroni G, Cammelli S, Api P, Giorgetti C, Barbieri E (2005). "Subcutaneous amifostine in head and neck cancer radiotherapy." *I supplementi di Tumori*, **4**(1). [◀ p. 49]

Gutierrez RG (2002). "Parametric frailty and shared frailty survival models." *Stata Journal*, **2**(1), 22–44. [◀ p. 90]

Ha ID, Sylvester R, Legrand C, MacKenzie G (2011). "Frailty modelling for survival data from multicentre clinical trials." *Statist. Med.*, **30**(17), 2144–2159. ISSN 1097-0258. doi:10.1002/sim.4250. [◀ p. 6, 20]

Hirsch K, Wienke A (2011). "Software for semiparametric shared gamma and log-normal frailty models: An overview." *Computer Methods and Programs in Biomedicine*, **107**(3). doi:10.1016/j.cmpb.2011.05.004. [◀ p. 29, 90]

Horny G (2009). "Inference in mixed proportional hazard models with K random effects." *Statistical Papers*, **50**(3), 481–499. doi:10.1007/s00362-007-0087-y. [◀ p. 2, 3, 22, 56, 61, 100, 101]

Hougaard P (1986). "Survival models for heterogeneous populations derived from stable distributions." *Biometrika*, **73**(2), 387–396. [◀ p. 37]

Hougaard P (1995). "Frailty models for survival data." *Lifetime Data Analysis*, **1**(3), 255–273. doi:10.1007/BF00985760. [◀ p. 6, 20]

Hougaard P (1999a). "Fundamentals of survival data." *Biometrics*, **55**(1), 13–22. [◀ p. 1]

Hougaard P (1999b). "Multi-state models: a review." *Lifetime Data Analysis*, **5**(3), 239–264. [◀ p. 6]

Hougaard P (2000). *Analysis of multivariate survival data*. Springer Verlag. [◀ p. 1, 9, 11, 12, 13, 19, 21, 22, 24, 36, 39]

Kalbfleisch JD, Prentice RL (2002). *The Statistical Analysis of Failure Time Data*. Wiley Series in Probability and Statistics, second edition. John Wiley & Sons. ISBN 9780471363576. doi:10.1002/9781118032985. [◀ p. 9, 16]

Kendall MG (1938). "A new measure of rank correlation." *Biometrika*, **30**(1/2), 81–93. [◄ p. 34]

Klein JP (1992). "Semiparametric estimation of random effects using the Cox model based on the EM algorithm." *Biometrics*, **48**(3), 795–806. doi:10.2307/2532345. [◄ p. 29]

Klein JP, Moeschberger ML (2003). *Survival analysis: techniques for censored and truncated data*. Second edition. Springer Verlag. doi:10.1007/b97377. [◄ p. 9]

Knatterud GL, Rockhold FW, George SL, Barton FB, Davis CE, Fairweather WR, Honohan T, Mowery R, O'Neill R (1998). "Guidelines for quality assurance in multicenter trials: a position paper." *Controlled Clinical Trials*, **19**(5), 477–493. [◄ p. 5]

Kurth KH, Schröder FH, Tunn U, Ay R, Pavone-Macaluso M, Debruyne F, De Pauw M, Dalesio O, Ten Kate F (1984). "Adjuvant chemotherapy of superficial transitional cell bladder carcinoma: preliminary results of a European organization for research on treatment of cancer. Randomized trial comparing doxorubicin hydrochloride, ethoglucid and transurethral resection alone." *The Journal of Urology*, **132**(2), 258. [◄ p. 73]

Lee EW, Wei LJ, Amato DA (1992). *Cox-type regression analysis for large numbers of small groups of correlated failure time observations*, volume 211, pp. 237–247. Kluwer Academic Publishers Dordrecht. [◄ p. 6]

Lin DY (1994). "Cox regression analysis of multivariate failure time data: The marginal approach." *Statist. Med.*, **13**(21), 2233–2247. doi:10.1002/sim.4780132105. [◄ p. 6]

Liquet B, Timsit JF, Rondeau V (2012). "Investigating hospital heterogeneity with a multi-state frailty model: application to nosocomial pneumonia disease in intensive care units." *BMC Medical Research Methodology*, **12**(1), 79. doi:10.1186/1471-2288-12-79. [◄ p. 6, 55]

Louis TA (1982). "Finding the Observed Information Matrix when Using the EM Algorithm." *Journal of the Royal Statistical Society. Series B*, **44**(2), 226–233. ISSN 00359246. doi:10.2307/2345828. [◄ p. 101]

Ma TY, Joly I, Raux C (2010). "A shared frailty semi-parametric markov renewal model for travel and activity time-use pattern analysis." Working Paper. [◄ p. 6, 55, 57]

Manda S, Meyer R (2005). "Age at first marriage in Malawi: a Bayesian multilevel analysis using a discrete time-to-event model." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **168**(2), 439–455. ISSN 0964-1998. doi:10.1111/j.1467-985X.2005.00357.x. [◄ p. 100]

Massonnet G (2008). *Contributions to frailty and copula modelling with applications to clinical trials and dairy cows data*. Ph.D. thesis, Universiteit Hasselt. [◄ p. 101]

McCulloch CE, Searle SR, Neuhaus JM (2008). *Generalized, Linear, and Mixed Models*. Second edition. Wiley, New York. ISBN 978-0-470-07371-1. [◄ p. 19]

McGilchrist CA (1993). "REML estimation for survival models with frailty." *Biometrics*. doi:10.2307/2532615. [◄ p. 31, 32]

McGilchrist CA (1994). "Estimation in Generalized Mixed Models." *Journal of the Royal Statistical Society. Series B (Methodological)*, **56**(1). [◀ p. 31]

McGilchrist CA, Aisbett CW (1991). "Regression with frailty in survival analysis." *Biometrics*, **47**, 461–466. [◀ p. 31, 32, 93]

Meira-Machado L, De Uña Alvarez J, Cadarso-Suárez C (2006). "Nonparametric estimation of transition probabilities in a non-Markov illness–death model." *Lifetime Data Analysis*, **12**(3), 325–344. doi:10.1007/s10985-006-9009-x. [◀ p. 43]

Munda M, Rotolo F, Legrand C (2012). "`parfm`: Parametric Frailty Models in `R`." *Journal of Statistical Software*, **51**(11). [◀ p. 3, 30, 89]

Nelsen RB (2006). *An Introduction to Copulas*. Second edition. Springer Verlag. [◀ p. 44]

Nelson W (1969). "Hazard plotting for incomplete failure data." *Journal of Quality Technology*, **1**, 27–52. [◀ p. 17, 30]

Newling DWW, Robinson MRG, Smith PH, Byar D, Lockwood R, Stevens I, De Pauw M, Sylvester R (1995). "Tryptophan metabolites, pyridoxine (vitamin B6) and their influence on the recurrence rate of superficial bladder cancer: results of a prospective, randomised phase III study performed by the EORTC GU Group." *European Urology*, **27**(2). [◀ p. 73]

Nielsen GG, Gill RD, Andersen PK, Sørensen TIA (1992). "A Counting Process Approach to Maximum Likelihood Estimation in Frailty Models." *Scandinavian Journal of Statistics*, **19**(1). [◀ p. 29, 30]

Oosterlinck W, Kurth KH, Schröder F, Bultinck J, Hammond B, Sylvester R (1993). "A prospective European Organization for Research and Treatment of Cancer Genitourinary Group randomized trial comparing transurethral resection followed by a single intravesical instillation of epirubicin or water in single stage Ta, T1 papillary carcinoma of the bladder." *The Journal of Urology*, **149**(4), 749. [◀ p. 73]

O'Quigley J, Stare J (2002). "Proportional hazards models with frailties and random effects." *Statistics in Medicine*, **21**(21), 3219–3233. doi:10.1002/sim.1259. [◀ p. 20]

Parner E (1998). "Asymptotic theory for the correlated gamma-frailty model." *The Annals of Statistics*, **26**(1). doi:10.1214/aos/1030563982. [◀ p. 22]

Petersen JH (1998). "An additive frailty model for correlated life times." *Biometrics*. doi:10.2307/3109771. [◀ p. 22]

Pintilie M (2006). *Competing risks: a practical perspective*. Wiley. [◀ p. 12]

Putter H (2011). "Tutorial in biostatistics: Competing risks and multi-state models, Analyses using the `mstate` package." [◀ p. 16, 89, 133]

Putter H, de Wreede L, Fiocco M (2012). "`mstate`: Data preparation, estimation and prediction in multi-state models." `R` package version 0.2.6. [◀ p. 89]

Putter H, Fiocco M, Geskus RB (2007). "Tutorial in biostatistics: competing risks and multi-state models." *Statistics in Medicine*, **26**(11), 2389–430. doi:10.1002/sim.2712. [◀ p. 1, 6, 15, 43, 46]

Putter H, van der Hage J, de Bock GH, Elgalta R, van de Velde CJH (2006). "Estimation and Prediction in a Multi-State Model for Breast Cancer." *Biometrical Journal*, **48**(3), 366–380. doi:10.1002/bimj.200510218. [◀ p. 14, 18]

Putter H, van Houwelingen HC (2011). "Frailties in multi-state models: Are they identifiable? Do we need them?" *Statistical Methods in Medical Research*. doi:10.1177/0962280211424665. [◀ p. 55, 57]

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. [◀ p. 3, 89]

Ripatti S, Palmgren J (2000). "Estimation of multivariate frailty models using penalized partial likelihood." *Biometrics*, **56**(4). doi:10.1111/j.0006-341X.2000.01016.x. [◀ p. 31]

Rodrìguez G (2005). "Unobserved heterogeneity." Handouts for 'Pop509: Survival Analysis' course at Princeton University. [◀ p. 6]

Rondeau V (2010). "Statistical models for recurrent events and death: Application to cancer events." *Mathematical and Computer Modelling*, **52**(7-8), 949–955. doi:10.1016/j.mcm.2010.02.002. [◀ p. 20]

Rondeau V, Commenges D, Joly P (2003). "Maximum penalized likelihood estimation in a gamma-frailty model." *Lifetime Data Analysis*, **9**(2), 139–153. [◀ p. 29]

Rondeau V, Filleul L, Joly P (2006). "Nested frailty models using maximum penalized likelihood estimation." *Statistics in Medicine*, **25**(23), 4036–4052. doi:10.1002/sim.2510. [◀ p. 6, 22, 60]

Rondeau V, Gonzalez JR (2005). "`frailtypack`: A computer program for the analysis of correlated failure time data using penalized likelihood estimation." *Computer Methods and Programs in Biomedicine*, **80**(2), 154–164. doi:10.1016/j.cmpb.2005.06.010. [◀ p. 29, 90]

Rondeau V, Mathoulin-Pélissier S, Tanneau L, Sasco A, 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**(1), 697. doi:10.1186/1471-2407-10-697. [◀ p. 11]

Rondeau V, Mazroui Y, Gonzalez JR (2012). "`frailtypack`: An `R` Package for the Analysis of Correlated Survival Data with Frailty Models Using Penalized Likelihood Estimation." *Journal of Statistical Software*, **47**(4). [◀ p. 90]

Rotolo F, Horny G (2012). "`mlfm`: Semiparametric multilevel frailty models, estimated via EMPL method." `R` package version 1.2.3. [◀ p. 3, 89, 101]

Rotolo F, Legrand C (2012). "Frailty Multi-State Models based on Maximum Penalized Partial Likelihood." In CLEUP (ed.), *Proceedings of the 46th Scientific Meeting of the Italian Statistical Society*. ISSN 978 88 6129 882 8. [◀ p. 2, 61]

Rotolo F, Legrand C, Van Keilegom I (2012a). "A simulation procedure based on copulas to generate clustered multi-state survival data." *Computer Methods and Programs in Biomedicine*. doi:10.1016/j.cmpb.2012.09.003. [◀ p. 2, 43]

Rotolo F, Munda M (2012). "`parfm`: Parametric Frailty Models." `R` package version 2.5.2. [◀ p. 3, 89, 90]

Rotolo F, Rondeau V, Legrand C (2012b). "Incorporation of nested frailties into semiparametric multi-state models." Working paper. [◀ p. 2]

SAS Institute Inc (2011). *SAS/STAT Software, Version 9.3*. Cary, NC. [◀ p. 89]

Sastry N (1997). "A Nested Frailty Model for Survival Data, With an Application to the Study of Child Survival in Northeast Brazil." *Journal of the American Statistical Association*, **92**(438), 426–435. [◀ p. 6, 22, 56, 100]

Senn S (2007). *Statistical issues in drug development*. Wiley-Interscience. doi:10.1002/9780470723586. [◀ p. 5]

Shih JH, Lu SE (2009). "Semiparametric estimation of a nested random effects model for the analysis of multi-level clustered failure time data." *Computational Statistics & Data Analysis*, **53**(11), 3864–3871. doi:10.1016/j.csda.2009.04.014. [◀ p. 22]

Spiegelhalter D, Thomas A, Best N, Lunn D (2003). "WinBUGS User Manual." [◀ p. 90]

StataCorp (2011). *STATA 12 Base Reference Manual*. College Station, TX. [◀ p. 90]

Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, Verweij J, Van Glabbeke M, van Oosterom AT, Christian MC, Gwyther SG (2000). "New Guidelines to Evaluate the Response to Treatment in Solid Tumors." *Journal of the National Cancer Institute*, **92**(3), 205–216. ISSN 0027-8874. doi:10.1093/jnci/92.3.205. [◀ p. 81]

Therneau TM (2012a). "`coxme`: Mixed Effects Cox Models." `R` package version 2.2-3. [◀ p. 89, 90]

Therneau TM (2012b). "`survival`: Survival Analysis, Including Penalised Likelihood." `R` package version 2.37-2. [◀ p. 31, 89, 90, 101]

Therneau TM, Grambsch PM (2000). *Modeling survival data: extending the Cox model*. Springer Verlag. [◀ p. 9, 90, 101]

Therneau TM, Grambsch PM, Pankratz VS (2003). "Penalized survival models and frailty." *Journal of Computational and Graphical Statistics*, **12**(1), 156–175. doi:10.1198/1061860031365. [◀ p. 31]

van den Berg GJ, Drepper B (2012). "Inference for Shared-Frailty Survival Models with Left-Truncated Data." *Technical report*, University of Mannheim, Department of Economics. [◀ p. 92]

Van Keilegom I, de Uña-Alvarez J, Meira-Machado L (2011). "Nonparametric location-scale models for censored successive survival times." *Journal of Statistical Planning and Inference*, **141**(3), 1118–1131. doi:10.1016/j.jspi.2010.09.010. [◀ p. 43]

Vaupel JW, Manton KG, Stallard E (1979). "The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality." *Demography*, **16**(3), 439–454. [◄ p. 1, 20]

Vaupel JW, Yashin AI (1985a). "The Deviant Dynamics of Death in Heterogeneous Populations." *Sociological Methodology*, **15**. doi:10.2307/270850. [◄ p. 22]

Vaupel JW, Yashin AI (1985b). "Heterogeneity's Ruses: Some Surprising Effects of Selection on Population Dynamics." *The American Statistician*, **39**(3), 176–185. doi:10.1080/00031305.1985.10479424. [◄ p. 6, 19]

Wang ST, Klein JP, Moeschberger ML (1995). "Semi-parametric estimation of covariate effects using the positive stable frailty model." *Applied Stochastic Models and Data Analysis*, **11**(2), 121–133. doi:10.1002/asm.3150110203. [◄ p. 30]

Wei LJ (1992). "The accelerated failure time model: A useful alternative to the Cox regression model in survival analysis." *Statistics in Medicine*, **11**(14-15), 1871–1879. doi:10.1002/sim.4780111409. [◄ p. 10]

Wei LJ, Lin DY, Weissfeld L (1989). "Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions." *Journal of the American Statistical Association*, **84**(408), 1065–1073. doi:10.1080/01621459.1989.10478873. [◄ p. 6]

Wienke A (2010). *Frailty Models in Survival Analysis*. Chapman & Hall/CRC biostatistics series. Taylor and Francis. doi:10.1201/9781420073911. [◄ p. 1, 6, 20, 21, 22, 24, 38]

Witjes JA, Meijden AP, Collette L, Sylvester R, Debruyne FMJ, van Aubel A, Witjes WPJ (1998). "Long-term follow-up of an EORTC randomized prospective trial comparing intravesical bacille calmette-guérinRIVM and mitomycin c in superficial bladder cancer." *Urology*, **52**(3), 403–410. ISSN 00904295. doi:10.1016/S0090-4295(98)00212-X. [◄ p. 73]

Xue X, Ding Y (1999). "Assessing heterogeneity and correlation of paired failure times with the bivariate frailty model." *Statistics in Medicine*, **18**(8), 907–918. doi:10.1002/(SICI)1097-0258(19990430)18:8%3C907::AID-SIM82%3E3.0.CO;2-X. [◄ p. 20]

Yamaguchi T, Ohashi Y (1999). "Investigating centre effects in a multi-centre clinical trial of superficial bladder cancer." *Statistics in Medicine*, **18**(15), 1961–1971. doi:10.1002/(SICI)1097-0258(19990815)18:15%3C1961::AID-SIM170%3E3.0.CO;2-3. [◄ p. 5]

Yashin AI, Vaupel JW, Iachine IA (1995). "Correlated individual frailty: an advantageous approach to survival analysis of bivariate data." *Mathematical Population Studies*, **5**(2), 145–159. [◄ p. 55]

Yau KKW (2001). "Multilevel Models for Survival Analysis with Random Effects." *Biometrics*, **57**(1), 96–102. doi:10.1111/j.0006-341X.2001.00096.x. [◄ p. 22, 100]

Yen AMF, Chen THH, Duffy SW, Chen CD (2010). "Incorporating frailty in a multi-state model: application to disease natural history modelling of adenoma-carcinoma in the large bowel." *Statistical Methods in Medical Research*, **19**(5), 529–546. doi:10.1177/0962280209359862. [◄ p. 6, 55, 57]

# Index

# Appendix A

# Mathematical results

## A.1 Full loglikelihood of Markov multi-state models

*Proof.* Let $\tau_{qi}$ and $t_{qi}$ be the left truncation time and event/censoring time, respectively, for the $qi$-th subject. Then, we have

$$\mathsf{Y}_{qi}(t) = \begin{cases} 1, & t \in [\tau_{qi}, t_{qi}) \\ 0, & t \notin [\tau_{qi}, t_{qi}), \end{cases} \tag{A.1}$$

so

$$\mathrm{d}\mathsf{Y}_{qi}(t) = \begin{cases} 1, & t = \tau_{qi}, \\ -1 & t = t_{qi}, \\ 0, & \text{otherwise.} \end{cases} \tag{A.2}$$

By omitting the explicit dependence on the baseline hazard functions $\lambda_{q0}(\cdot)$, we have that

$$\begin{aligned} -\int_0^\infty \log S_{qi}(t)\mathrm{d}\mathsf{Y}_{qi}(t) &= \int_0^\infty \Lambda_{qi}(t)\mathrm{d}\mathsf{Y}_{qi}(t) \\ &= \Lambda_{qi}(\tau_{qi}) \times (1) + \Lambda_{qi}(t_{qi}) \times (-1) \\ &= \Lambda_{qi}(\tau_{qi}) - \Lambda_{qi}(t_{qi}) \\ &= -\left[\int_0^{t_{qi}} \lambda_{qi}(t)\mathrm{d}t - \int_0^{\tau_{qi}} \lambda_{qi}(t)\mathrm{d}t\right] \\ &= -\int_{\tau_{qi}}^{t_{qi}} \lambda_{qi}(t)\mathrm{d}t \\ &= -\int_0^\infty \lambda_{qi}(t)\mathsf{Y}_{qi}(t)\mathrm{d}t. \end{aligned} \tag{A.3}$$

Therefore, the full likelihood (3.17) is

$$\begin{aligned}\ell\big(\boldsymbol{\beta}, \boldsymbol{\lambda}_{q0}(\cdot)\big) &= \sum_{q=1}^{Q}\sum_{i=1}^{n}\left\{\int_0^\infty \log \lambda_{qi}(t)\mathrm{d}\mathsf{N}_{qi}(t) - \int_0^\infty \log S_{qi}(t)\mathrm{d}\mathsf{Y}_{qi}(t)\right\} \\ &= \sum_{q=1}^{Q}\sum_{i=1}^{n}\left\{\int_0^\infty \log \lambda_{qi}(t)\mathrm{d}\mathsf{N}_{qi}(t) - \int_0^\infty \lambda_{qi}(t)\mathsf{Y}_{qi}(t)\mathrm{d}t\right\}. \end{aligned} \tag{A.4}$$

□

## A.2 Variance of nested frailties

*Proof.* For the nested frailty model (3.58), the variance of the frailties is

$$\begin{aligned}\mathsf{V}(U_{hj}) &= \mathsf{E}\left[V_h^2 W_{hj}^2\right] - \mathsf{E}[V_h W_{hj}]^2 \\ &= \mathsf{E}\left[V_h^2\right]\mathsf{E}\left[W_{hj}^2\right] - \mathsf{E}[V_h]^2\mathsf{E}[W_{hj}]^2 \\ &= \big(\mathsf{V}\left[V_h\right] + \mathsf{E}[V_h]^2\big)\big(\mathsf{V}\left[W_{hj}\right] + \mathsf{E}[W_{hj}]^2\big) - \mathsf{E}[V_h]^2\mathsf{E}[W_{hj}]^2 \\ &= (\theta_V + 1)(\theta_W + 1) - 1. \end{aligned} \tag{A.5}$$

□

## A.3 Updated skewness of the frailty term

*Proof.* Consider the random variable $U_h \mid \mathcal{D}(t)$. Its first three raw moments $\mu_i' = \mathsf{E}[U_h^i \mid \mathcal{D}(t)]$ are

$$\mu_1' = -\frac{\mathcal{L}^{(d+1)}}{\mathcal{L}^{(d)}} = \mathsf{E}[U_h \mid \mathcal{D}(t)] \qquad (\Rightarrow \text{Eq. 3.83}), \tag{A.6}$$

$$\mu_2' = \frac{\mathcal{L}^{(d+2)}}{\mathcal{L}^{(d)}}, \tag{A.7}$$

$$\mu_3' = -\frac{\mathcal{L}^{(d+3)}}{\mathcal{L}^{(d)}}. \tag{A.8}$$

The associated central moments $\mu_i = \mathsf{E}[(U_h - \mu_1')^i \mid \mathcal{D}(t)]$ are

$$\mu_1 = 0, \tag{A.9}$$

$$\begin{aligned}\mu_2 &= \mu_2' - {\mu_1'}^2 \\ &= \frac{\mathcal{L}^{(d)}\mathcal{L}^{(d+2)} - [\mathcal{L}^{(d+1)}]^2}{[\mathcal{L}^{(d)}]^2} = \mathsf{V}[U_h \mid \mathcal{D}(t)] \qquad (\Rightarrow \text{Eq. 3.84}), \end{aligned} \tag{A.10}$$

$$\begin{aligned}\mu_3 &= 2{\mu_1'}^3 - 3\mu_1'\mu_2' + \mu_3' \\ &= \frac{3\mathcal{L}^{(d)}\mathcal{L}^{(d+1)}\mathcal{L}^{(d+2)} - [\mathcal{L}^{(d)}]^2\mathcal{L}^{(d+3)} - 2[\mathcal{L}^{(d+1)}]^3}{[\mathcal{L}^{(d)}]^3}. \end{aligned} \tag{A.11}$$

The skewness of $U_h \mid \mathcal{D}(t)$, which is its standardised central moment $\gamma_i = \mathsf{E}\left[\left(\frac{U_h - \mu_1'}{\mu_2^{1/2}}\right)^i \mid \mathcal{D}(t)\right]$ of order three, is

$$\begin{aligned}\gamma_3 &= \frac{3\mathcal{L}^{(d)}\mathcal{L}^{(d+1)}\mathcal{L}^{(d+2)} - [\mathcal{L}^{(d)}]^2\mathcal{L}^{(d+3)} - 2[\mathcal{L}^{(d+1)}]^3}{\left(\mathcal{L}^{(d)}\mathcal{L}^{(d+2)} - [\mathcal{L}^{(d+1)}]^2\right)^{3/2}} \\ &= \mathbb{A}[U_h \mid \mathcal{D}(t)] \qquad (\Rightarrow \text{Eq. 3.85}).\end{aligned} \tag{A.12}$$

□

## A.4 Derivative of the conditional frailty variance

*Proof.* Consider the derivative of the conditional variance (3.84)

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{V}[U_h|T>t] =& \left\{\left[\mathcal{L}^{(d_t)}\right]^2\left(\mathcal{L}^{(d_t)}\mathcal{L}^{(d_t+3)} - \mathcal{L}^{(d_t+1)}\mathcal{L}^{(d_t+2)}\right)\right. \\ & \left. -2\mathcal{L}^{(d_t)}\mathcal{L}^{(d_t+1)}\left(\mathcal{L}^{(d_t)}\mathcal{L}^{(d_t+2)} - \left[\mathcal{L}^{(d_t+1)}\right]^2\right)\right\}\left[\mathcal{L}^{(d_t)}\right]^{-4} \\ & \frac{\mathrm{d}}{\mathrm{d}t}\left[\sum_{i=1}^{n_h}\Lambda_0\big(\breve{y}_{hi}(t)\big)\exp\left\{\boldsymbol{\beta}^\top\mathbf{x}_{hi}\right\}\right] \\ =& \frac{\left[\mathcal{L}^{(d_t)}\right]^2\mathcal{L}^{(d_t+3)} - 3\mathcal{L}^{(d_t)}\mathcal{L}^{(d_t+1)}\mathcal{L}^{(d_t+2)} + 2\left[\mathcal{L}^{(d_t+1)}\right]^3}{\left[\mathcal{L}^{(d_t)}\right]^3} \\ & \left[\sum_{i=1}^{n_h}\mathbb{1}(y_{hi}\geq t)\lambda_0\big(\breve{y}_{hi}(t)\big)\exp\left\{\boldsymbol{\beta}^\top\mathbf{x}_{hi}\right\}\right] \\ =& -\mathsf{V}[U_h|T>t]^{\frac{3}{2}}\mathbb{A}[U_h|T>t] \\ & \left[\sum_{i=1}^{n_h}\mathbb{1}(y_{hi}\geq t)\lambda_0\big(\breve{y}_{hi}(t)\big)\exp\left\{\boldsymbol{\beta}^\top\mathbf{x}_{hi}\right\}\right],\end{aligned} \tag{A.13}$$

which is negative if and only if $\mathbb{A}[U|T>t] > 0$. □

## A.5 Derivatives of the joint survival function for a Clayton copula model

If

$$\begin{aligned}&\frac{\partial^k}{\partial t_{(1)}\cdots\partial t_{(k)}}S_{\mathcal{Q}_0}(\mathbf{t}_{\mathcal{Q}_0}) = \\ &\qquad (-1)^k\prod_{h=1}^{k}\left[(1+(h-1)\vartheta)\left(S_h(t_h)\right)^{-\vartheta-1}f_h(t_h)\right]\left(S_{\mathcal{Q}_0}(\mathbf{t}_{\mathcal{Q}_0})\right)^{1+k\vartheta},\end{aligned} \tag{A.14}$$

holds, hence

$$\begin{aligned}
\frac{\partial^{k+1}}{\partial t_{(1)}\cdots\partial t_{(k+1)}} S_{\mathcal{Q}_0}(\mathbf{t}_{\mathcal{Q}_0}) &= \frac{\partial}{\partial t_{(k+1)}}\left[\frac{\partial^{k}}{\partial t_{(1)}\cdots\partial t_{(k)}} S_{\mathcal{Q}_0}(\mathbf{t}_{\mathcal{Q}_0})\right] \\
&= (-1)^k \prod_{h=1}^{k}\left[(1+(h-1)\vartheta)\left(S_h(t_h)\right)^{-\vartheta-1} f_h(t_h)\right] \\
&\qquad (-1)(1+k\theta)\left(S_{k+1}(t_{k+1})\right)^{-\vartheta-1} f_{k+1}(t_{k+1})\left(S_{\mathcal{Q}_0}(\mathbf{t}_{\mathcal{Q}_0})\right)^{1+(k+1)\vartheta} \\
&= (-1)^{k+1}\prod_{h=1}^{k+1}\left[(1+(h-1)\vartheta)\left(S_h(t_h)\right)^{-\vartheta-1} f_h(t_h)\right]\left(S_{\mathcal{Q}_0}(\mathbf{t}_{\mathcal{Q}_0})\right)^{1+(k+1)\vartheta}
\end{aligned} \tag{A.15}$$

which is the same as (A.14) computed in $k+1$ instead of $k$.

Now, let consider the first derivative

$$\frac{\partial}{\partial t_{(1)}} S_{\mathcal{Q}_0}(\mathbf{t}_{\mathcal{Q}_0}) = -\left[\left(S_1(t_1)\right)^{-\vartheta-1} f_1(t_1)\right]\left(S_{\mathcal{Q}_0}(\mathbf{t}_{\mathcal{Q}_0})\right)^{1+\vartheta}, \tag{A.16}$$

which is (A.14) computed in $k=1$.

Finally, since (A.14)⇒(A.15) and (A.14) holds for $k=1$ (Eq. A.16), we can conclude by induction that Equation A.14 holds for any $k \in \mathbb{N}$.

## A.6 Covariance of nested frailties

*Proof.* If $h = h'$, then the following holds:

$$\begin{aligned}
\mathsf{Cov}(U_{qh}, U_{q'h}) &= \mathsf{E}(U_{qh}U_{q'h}) - \mathsf{E}(U_{qh})\mathsf{E}(U_{q'h}) \\
&= \mathsf{E}(V_h W_{qh} V_h W_{q'h}) - \mathsf{E}(V_h W_{qh})\mathsf{E}(V_h W_{q'h}) \\
&= \mathsf{E}(V_h^2)\mathsf{E}(W_{qh})\mathsf{E}(W_{q'h}) - \mathsf{E}(V_h)^2\mathsf{E}(W_{qh})\mathsf{E}(W_{q'h}) \\
&= \mathsf{V}(V_h)\left[\mathsf{E}(W_{qh})\mathsf{E}(W_{q'h})\right] > 0.
\end{aligned} \tag{A.17}$$

Therefore, as $\mathsf{Cor}(U_{qh}, U_{q'h}) = \mathsf{Cov}(U_{qh}, U_{q'h})/(\mathsf{V}(U_{qh})\mathsf{V}(U_{q'h}))$ and the variances at the denominator are necessarily positive, then $\mathsf{Cor}(U_{qh}, U_{q'h}) > 0$. □

## A.7 Joint conditional frailty distribution of nested frailties

*Proof.* Let us consider the joint density function of the frailties of a given cluster $h$, conditionally on left truncation times. This is

$$\begin{aligned} & f_{V_h,\boldsymbol{W}_h}\Big(v_h,\boldsymbol{w}_h;\boldsymbol{\zeta} \mid \{Y_{qhi} > \tau_{qhi}\}_{qi}\Big) \\ &\quad = \frac{\mathsf{P}\Big(\cap_{qi}\{Y_{qhi} > \tau_{qhi}\};\boldsymbol{\beta},\boldsymbol{\xi} \mid V_h = v_h, \{W_{qh} = w_{qh}\}_q\Big)}{\mathsf{P}\Big(\cap_{qi}\{Y_{qhi} > \tau_{qhi}\};\boldsymbol{\zeta}\Big)} \\ &\qquad \times \frac{\partial^{Q+1}}{\partial v_h \partial w_{1h}\cdots\partial w_{Qh}} \mathsf{P}\Big(V_h \le v_h, \{W_{qh} \le w_{qh}\}_q;\boldsymbol{\theta}\Big) \\ &\quad = \frac{S_h\Big(\{\tau_{qhi}\}_{qi};\boldsymbol{\beta},\boldsymbol{\xi} \mid V_h = v_h, \{W_{qh} = w_{qh}\}_q\Big)}{S_h\Big(\{\tau_{qhi}\}_{qi};\boldsymbol{\zeta}\Big)} \\ &\qquad \times f_{V_h,\boldsymbol{W}_h}\Big(v_h,\boldsymbol{w}_h;\boldsymbol{\theta}\Big). \end{aligned} \tag{A.18}$$

The numerator and the denominator are, respectively, the joint *conditional* survival function

$$\begin{aligned} S_h\Big(\{\tau_{qhi}\}_{qi};\boldsymbol{\beta},\boldsymbol{\xi} \mid V_h = v_h, \{W_{qh} = w_{qh}\}_q\Big) = \\ \prod_{qi \mid \tau_{qhi} < \infty} \exp\Big\{-v_h w_{qh} \Lambda_{q0}(\tau_{qhi}) e^{\boldsymbol{\beta}_q^\top \mathbf{x}_{qhi}}\Big\}, \end{aligned} \tag{A.19}$$

and the joint *marginal* survival function

$$\begin{aligned} S_h\Big(\{\tau_{qhi}\}_{qi};\boldsymbol{\zeta}\Big) = \int\limits_{\mathbb{R}_+^{Q+1}} \prod_{qi \mid \tau_{qhi} < \infty} \exp\Big\{-v_h w_{qh} \Lambda_{q0}(\tau_{qhi}) e^{\boldsymbol{\beta}_q^\top \mathbf{x}_{qhi}}\Big\} \\ f_{V_h,\boldsymbol{W}_h}\Big(v_h,\boldsymbol{w}_h;\boldsymbol{\theta}\Big)\mathrm{d}w_{1h}\cdots\mathrm{d}w_{Qh}\ \mathrm{d}v_h \end{aligned} \tag{A.20}$$

of left truncation times in cluster $h$. Let define as $\mathcal{Q}_h = \Big\{q \mid \sum_{i=1}^{n_h} \mathbb{1}(\tau_{qhi} < \infty) > 0\Big\}$ the set of transitions for which there exists in group $h$ at least one subject who has ever been at risk. It immediately follows that the joint conditional frailty distribution of nested frailties in a cluster is given by Equation 5.9. □

## A.8 Marginal likelihood of nested frailties multi-state models

*Proof.* Under assumptions (5.3), the joint marginal survival function (A.20) is

$$
\begin{aligned}
&S_h\Big(\{\tau_{qhi}\}_{qi};\boldsymbol{\zeta}\Big)\\
&\quad=\int_{\mathbb{R}_+}\prod_{q\in\mathcal{Q}_h}\left[\int_{\mathbb{R}_+}\exp\left\{-v_h w_{qh}\sum_{i|\tau_{qhi}<\infty}\Lambda_{q0}(\tau_{qhi})e^{\boldsymbol{\beta}_q^\top\mathbf{x}_{qhi}}\right\}f_W\Big(w_{qh};\theta_q\Big)\mathrm{d}w_{qh}\right]f_V\Big(v_h;\theta_V\Big)\mathrm{d}v_h\\
&\quad=\int_{\mathbb{R}_+}\prod_{q\in\mathcal{Q}_h}\mathcal{L}_q\left(v_h\sum_{i|\tau_{qhi}<\infty}\Lambda_{q0}(\tau_{qhi})e^{\boldsymbol{\beta}_q^\top\mathbf{x}_{qhi}}\right)f_V\Big(v_h;\theta_V\Big)\mathrm{d}v_h,
\end{aligned}
\tag{A.21}
$$

with $\mathcal{L}_q(s)$ the Laplace transform of $f_W(\cdot;\theta_q)$. Then, the contribution (5.8) of a cluster to the marginal likelihood is

$$
\begin{aligned}
L_{\mathsf{M},h}(\boldsymbol{\zeta})=\int_{\mathbb{R}_+}\prod_{q\in\mathcal{Q}_h}\Bigg\{\int_{\mathbb{R}_+}\prod_{i|\tau_{qhi}<\infty}\Bigg[&\Big\{v_h w_{qh}\lambda_{q0}(y_{qhi})e^{\boldsymbol{\beta}_q^\top\mathbf{x}_{qhi}}\Big\}^{\delta_{qhi}}\\
&\frac{\exp\Big\{-v_h w_{qh}\Lambda_{q0}(y_{qhi})e^{\boldsymbol{\beta}_q^\top\mathbf{x}_{qhi}}\Big\}}{\exp\Big\{-v_h w_{qh}\Lambda_{q0}(\tau_{qhi})e^{\boldsymbol{\beta}_q^\top\mathbf{x}_{qhi}}\Big\}}\Bigg]\\
&\frac{\prod_{i|\tau_{qhi}<\infty}\exp\Big\{-v_h w_{qh}\Lambda_{q0}(\tau_{qhi})e^{\boldsymbol{\beta}_q^\top\mathbf{x}_{qhi}}\Big\}}{\int_{\mathbb{R}_+}\prod_{q\in\mathcal{Q}_h}\mathcal{L}_q\left(v_h\sum_{i|\tau_{qhi}<\infty}\Lambda_{q0}(\tau_{qhi})e^{\boldsymbol{\beta}_q^\top\mathbf{x}_{qhi}}\right)f_V\Big(v_h;\theta_V\Big)\mathrm{d}v_h}\\
&f_W(w_{qh};\theta_q\Big)\mathrm{d}w_{qh}\Bigg\}f_V\Big(v_h;\theta_V\Big)\mathrm{d}v_h
\end{aligned}
\tag{A.22}
$$

which simplifies to

$$
\begin{aligned}
L_{\mathsf{M},h}(\boldsymbol{\zeta})=&\prod_{qi|\tau_{qhi}<\infty}\Big\{\lambda_{q0}(y_{qhi})e^{\boldsymbol{\beta}_q^\top\mathbf{x}_{qhi}}\Big\}^{\delta_{qhi}}\\
&\times\frac{\int_{\mathbb{R}_+}v_h^{d_h}\prod_{q\in\mathcal{Q}_h}(-1)^{d_{qh}}\mathcal{L}_q^{(d_{qh})}\left(v_h\sum_{i|\tau_{qhi}<\infty}\Lambda_{q0}(y_{qhi})e^{\boldsymbol{\beta}_q^\top\mathbf{x}_{qhi}}\right)f_V\Big(v_h;\theta_V\Big)\mathrm{d}v_h}{\int_{\mathbb{R}_+}\prod_{q\in\mathcal{Q}_h}\mathcal{L}_q\left(v_h\sum_{i|\tau_{qhi}<\infty}\Lambda_{q0}(\tau_{qhi})e^{\boldsymbol{\beta}_q^\top\mathbf{x}_{qhi}}\right)f_V\Big(v_h;\theta_V\Big)\mathrm{d}v_h},
\end{aligned}
\tag{A.23}
$$

with $\mathcal{L}_q^{(k)}(s)$ the $k$-th derivative with respect to $s$ of the Laplace transform of $f_W(\cdot;\theta_q)$. Hence, the marginal loglikelihood of the nested frailty multi-state model (5.1)–(5.3) is given by Equation 5.11.

□

## A.9 Marginal likelihood of gamma nested frailties multi-state models

*Proof.* The marginal loglikelihood (5.11) depends on the frailty distributions through two quantities: the distribution of the $V_h$'s and the derivatives of the Laplace transform of the distribution of the $W_{qh}$'s. In the case of unit mean gamma distributions, we have

$$f_V(v_h;\theta_V) = \frac{v_h^{1/\theta_V-1}\exp(-v_h/\theta_V)}{\Gamma(1/\theta_V)\theta_V^{\theta_V}} \tag{A.24}$$

and

$$\mathcal{L}_q^{(d_{qh})}\left(v_h\sum_{i|\tau_{qhi}<\infty}\Lambda_{q0}(y_{qhi})e^{\boldsymbol{\beta}_q^\top \mathbf{x}_{qhi}}\right) = (-1)^{d_{qh}}\left[\prod_{l=0}^{d_{qh}-1}(1+l\theta_q)\right]^{\mathbb{1}(d_{qh}>1)} \left(1+\theta_q v_h\sum_{i|\tau_{qhi}<\infty}\Lambda_{q0}(y_{qhi})e^{\boldsymbol{\beta}_q^\top \mathbf{x}_{qhi}}\right)^{-d_{qh}-1/\theta_q}. \tag{A.25}$$

Hence, the second and third lines in Equation 5.11 get respectively

$$\log\left[\prod_{q\in\mathcal{Q}_h}\prod_{l=0}^{d_{qh}-1}(1+l\theta_q)\right] - \log\left[\Gamma(1/\theta_V)\theta_V^{\theta_V}\right] + \log\int_0^\infty v_h^{1/\theta_V+d_h-1}\prod_{q\in\mathcal{Q}_h}\left(1+\theta_q v_h\sum_{i|\tau_{qhi}<\infty}\Lambda_{q0}(y_{qhi})e^{\boldsymbol{\beta}_q^\top \mathbf{x}_{qhi}}\right)^{-d_{qh}-1/\theta_q} \exp(-v_h/\theta_V)\mathrm{d}v_h \tag{A.26}$$

and

$$\log\left[\Gamma(1/\theta_V)\theta_V^{\theta_V}\right] - \log\int_0^\infty v_h^{1/\theta_V-1}\prod_{q\in\mathcal{Q}_h}\left(1+\theta_q v_h\sum_{i|\tau_{qhi}<\infty}\Lambda_{q0}(y_{qhi})e^{\boldsymbol{\beta}_q^\top \mathbf{x}_{qhi}}\right)^{-1/\theta_q} \exp(-v_h/\theta_V)\mathrm{d}v_h \tag{A.27}$$

which immediately gives that the marginal loglikelihood is (5.12). □

## A.10 Saddlepoint approximation of integrals in nested frailty multi-state models

In Equation 5.12, we need to compute log-integrals of the form

$$\log \mathcal{I}_h = \log \int_0^\infty \frac{v_h^{1/\theta_V + d_h - 1} \exp\left(-v_h/\theta_V\right)}{\prod_{q \in \mathcal{Q}_h} \left(1 + \theta_q v_h \mathbf{\Lambda}_{qh}\right)^{1/\theta_q + d_{qh}}} \mathrm{d}v_h, \tag{A.28}$$

where the notation

$$\mathbf{\Lambda}_{qh} = \sum_{i | \tau_{qhi} < \infty} \Lambda_{q0}(y_{qhi}) e^{\boldsymbol{\beta}_q^\top \mathbf{x}_{qhi}}$$

is used for ease of presentation.

The log-integral (A.28) can be expressed as

$$\log \mathcal{I}_h = \log \int_0^\infty \exp\Big\{g(v_h)\Big\} \mathrm{d}v_h, \tag{A.29}$$

with

$$\begin{aligned} g(v_h) =& \log v_h \left(\frac{1}{\theta_V} + d_h - 1\right) - \frac{v_h}{\theta_V} \\ & - \sum_{q \in \mathcal{Q}_h} \left(\frac{1}{\theta_q} + d_{qh}\right) \log\left(1 + \theta_q v_h \mathbf{\Lambda}_{qh}\right), \end{aligned} \tag{A.30}$$

$$\begin{aligned} g'(v_h) =& \frac{1}{v_h} \left(\frac{1}{\theta_V} + d_h - 1\right) - \frac{1}{\theta_V} \\ & - \sum_{q \in \mathcal{Q}_h} \left(\frac{1}{\theta_q} + d_{qh}\right) \frac{1}{v_h + (\theta_q \mathbf{\Lambda}_{qh})^{-1}}, \end{aligned} \tag{A.31}$$

$$\begin{aligned} g''(v_h) =& -\frac{1}{v_h^2} \left(\frac{1}{\theta_V} + d_h - 1\right) \\ & + \sum_{q \in \mathcal{Q}_h} \left(\frac{1}{\theta_q} + d_{qh}\right) \frac{1}{\left(v_h + (\theta_q \mathbf{\Lambda}_{qh})^{-1}\right)^2}. \end{aligned} \tag{A.32}$$

We call $\hat{v} = \max_{(0,\infty)} g(v_h)$, found by numerical optimisation; then, the integral $\mathcal{I}_h$ can be approximated via the saddlepoint method, giving

$$\begin{aligned} \log \mathcal{I}_h \approx& g(\hat{v}) + \frac{1}{2} \log\left(\frac{2\pi}{-g''(\hat{v})}\right) \\ =& \log \hat{v} \left(\frac{1}{\theta_V} + d_h - 1\right) - \frac{\hat{v}}{\theta_V} - \sum_{q \in \mathcal{Q}_h} \left(\frac{1}{\theta_q} + d_{qh}\right) \log\left(1 + \theta_q \hat{v} \mathbf{\Lambda}_{qh}\right) \\ & + \frac{1}{2} \Bigg[ \log 2\pi - \log\Bigg( \frac{1}{\hat{v}^2} \left(\frac{1}{\theta_V} + d_h - 1\right) \\ & - \sum_{q \in \mathcal{Q}_h} \left(\frac{1}{\theta_q} + d_{qh}\right) \frac{1}{\left(\hat{v} + (\theta_q \mathbf{\Lambda}_{qh})^{-1}\right)^2} \Bigg) \Bigg]. \end{aligned} \tag{A.33}$$

Analogously, for left censoring times, the log-integral

$$\log \mathcal{J} = \log \int_0^\infty \frac{v_h^{1/\theta_V - 1} \exp\left(-v_h/\theta_V\right)}{\prod_{q \in \mathcal{Q}_h} \left(1 + \theta_q v_h \mathbf{\Lambda}^\star_{qh}\right)^{1/\theta_q}} \mathrm{d}v_h, \tag{A.34}$$

with

$$\mathbf{\Lambda}^\star_{qh} = \sum_{i \mid \tau_{qhi} < \infty} \Lambda_{q0}(\tau_{qhi}) e^{\boldsymbol{\beta}_q^\top \mathbf{x}_{qhi}}$$

can be approximated via the saddlepoint method as

$$\begin{aligned} \log \mathcal{J}_h \approx \log \hat{v}^\star \left(\frac{1}{\theta_V} - 1\right) - \frac{\hat{v}^\star}{\theta_V} - \sum_{q \in \mathcal{Q}_h} \frac{1}{\theta_q} \log\left(1 + \theta_q \hat{v}^\star \mathbf{\Lambda}_{qh}\right) \\ + \frac{1}{2}\Bigg[\log 2\pi - \log\Bigg(\frac{1}{\left(\hat{v}^\star\right)^2}\left(\frac{1}{\theta_V} - 1\right) \\ - \sum_{q \in \mathcal{Q}_h} \frac{1}{\theta_q} \frac{1}{\left(\hat{v}^\star + \left(\theta_q \mathbf{\Lambda}_{qh}\right)^{-1}\right)^2}\Bigg)\Bigg] \end{aligned} \tag{A.35}$$

with

$$\hat{v}^\star = \max_{(0,\infty)} \left\{ -\frac{1}{v_h^2}\left(\frac{1}{\theta_V} - 1\right) + \sum_{q \in \mathcal{Q}_h} \frac{1}{\theta_q} \frac{1}{\left(v_h + \left(\theta_q \mathbf{\Lambda}_{qh}\right)^{-1}\right)^2} \right\}. \tag{A.36}$$

## A.11 Partial likelihood of nested frailty multi-state models

*Proof.* In the transition-specific conditional hazard (Eq. 5.1–5.2) the log-frailties can be considered as fixed offsets:

$$\lambda_{qhi}(t \mid v_h, w_{qh}) = \lambda_{q0}(t) \exp\left\{\log v_h + \log w_{qh} + \boldsymbol{\beta}_q^\top \mathbf{x}_{qhi}\right\}. \tag{A.37}$$

This corresponds to the transition-specific hazard of a fixed-effects multi-state model, with the addition of log-frailties in the linear predictors. Hence, the profile version of the conditional likelihood (5.6) reduces to the partial likelihood of a multi-state model with offsets:

$$\begin{aligned} L_{\mathrm{P}}(\boldsymbol{\beta}) &= \prod_{h=1}^{H} \prod_{i=1}^{n_h} \prod_{q=1}^{Q} \left( \frac{\exp\left\{\log v_h + \log w_{qh} + \boldsymbol{\beta}_q^\top \mathbf{x}_{qhi}\right\}}{\sum_{h'i' \in R_q(y_{qhi})} \exp\left\{\log v_{h'} + \log w_{qh'} + \boldsymbol{\beta}_q^\top \mathbf{x}_{qh'i'}\right\}} \right)^{\delta_{qhi}} \\ &= \prod_{h=1}^{H} \left\{ v_h^{d_h} \prod_{q=1}^{Q} \left[ w_{qh}^{d_{qh}} \prod_{i=1}^{n_h} \left( \frac{\exp\left\{\boldsymbol{\beta}_q^\top \mathbf{x}_{qhi}\right\}}{\sum_{h'i' \in R_q(y_{qhi})} v_{h'} w_{qh'} \exp\left\{\boldsymbol{\beta}_q^\top \mathbf{x}_{qh'i'}\right\}} \right)^{\delta_{qhi}} \right] \right\}, \end{aligned} \tag{A.38}$$

with $R_q(y) = \left\{(h', i') \mid y \in [\tau_{qh'i'}, y_{qh'i'}]\right\}$ the risk set for transitions of type $q$ at time $y$, which takes into account the information given by the left truncation times $\tau_{qhi}$. Therefore, the conditional partial loglikelihood of the nested frailty multi-state model (5.1)–(5.2) is given by Equation 5.13. □

## A.12 Saddlepoint approximation for lognormal frailties

*Proof.* If the frailties follow a lognormal distribution with variance $\theta = e^{2\gamma} - e^{\gamma}$ (see Sec. 3.3.4, pg. 3.3.4), then we have in Equation 7.1 that

$$\begin{aligned}(-1)^k \mathcal{L}^{(k)}(s) &= \int_0^\infty u^k \exp -us f_U(u;\gamma)\mathrm{d}u \\ &= \int_0^\infty u^k \frac{1}{\sqrt{2\pi\gamma}} \exp\left\{-us - \frac{(\log u)^2}{2\gamma}\right\} \frac{\mathrm{d}u}{u} \\ &= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi\gamma}} \exp\left\{kw - \exp(w)s - \frac{w^2}{2\gamma}\right\} \mathrm{d}w, \end{aligned} \tag{A.39}$$

with $w = \log u$. This integral cannot be computed analytically, but saddlepoint approximation (Goutis and Casella, 1999) can be used with

$$g(w;s,k,\gamma) = kw - \exp(w)s - \frac{w^2}{2\gamma}, \tag{A.40}$$

$$g'(w;s,k,\gamma) = k - \exp(w)s - \frac{w}{\gamma}, \tag{A.41}$$

$$g''(w;s,k,\gamma) = -\exp(w)s - \frac{1}{\gamma}\cdot \tag{A.42}$$

Note that the second derivative (A.42) is always negative, so the maximum point $\tilde{w}$ of $g(w;s,k,\gamma)$ can be numerically looked for, without problems of local maxima.

Then, the logarithm of the quantity (A.39) can be approximated as

$$\begin{aligned}\log\Big((-1)^k \mathcal{L}^{(k)}(s)\Big) =& \log\left(\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi\gamma}} \exp\left\{g(w;s,k,\gamma)\right\} \mathrm{d}w\right) \\ \simeq& g(\tilde{w};s,k,\gamma) \\ &+ \log\left(\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi\gamma}} \exp\left\{\frac{(w-\tilde{w})^2}{2} g''(w;s,k,\gamma)\right\} \mathrm{d}w\right) \\ =& g(\tilde{w};s,k,\gamma) - \frac{1}{2}\log\Big(-\gamma g''(\tilde{w};s,k,\gamma)\Big) \\ =& k\tilde{w} - \exp(\tilde{w})s - \frac{\tilde{w}^2}{2\gamma} - \frac{1}{2}\log\Big(\gamma\exp(\tilde{w})s + 1\Big). \end{aligned} \tag{A.43}$$

□

## A.13 Conditional expectation of frailty terms

For ease of notation, let $\Lambda_{i\cdot,c}(\boldsymbol{y}_h)$ denote $\sum_{i=1}^{n_h} \Lambda_0(y_{hi}) \exp\left\{\boldsymbol{\beta}^\top \boldsymbol{x}_{hi}\right\}$. For any frailty distribution $f(u_h;\theta)$ and for any $\alpha \in \mathbb{N}$, we have

$$\mathsf{E}\left(U^{\alpha};\hat{\boldsymbol{\xi}} \mid \boldsymbol{x}_h,\boldsymbol{y}_h,\boldsymbol{\tau}_h\right) = \int_0^\infty u_h^{\alpha} f_{U_h}\left(u_h;\hat{\boldsymbol{\xi}} \mid \boldsymbol{x}_h,\boldsymbol{y}_h,\boldsymbol{\tau}_h\right)\mathrm{d}u_h \tag{A.44}$$

$$= \int_0^\infty u_h^{\alpha} \frac{L_{\mathsf{C},h}\left(\boldsymbol{\zeta},\boldsymbol{\beta} \mid u_h\right) f_{U_h}\left(u_h;\theta \mid \boldsymbol{\tau}_h\right)}{L_{\mathsf{M},h}\left(\boldsymbol{\zeta},\boldsymbol{\beta},\theta\right)}\mathrm{d}u_h, \tag{A.45}$$

with

$$\begin{aligned}
L_{\mathsf{C},h}\left(\boldsymbol{\zeta},\boldsymbol{\beta} \mid u_h\right) &= \left[\prod_{i=1}^{n_h}\left(\lambda_0(y_{hi})u_h \exp\left\{\boldsymbol{\beta}^\top \boldsymbol{x}_{hi}\right\}\right)^{\delta_{hi}}\right] \\
&\quad \times \exp\left\{-u_h\Lambda_{i\cdot,c}(\boldsymbol{y}_h)\right\}\exp\left\{u_h\Lambda_{i\cdot,c}(\boldsymbol{\tau}_h)\right\}, \\
f_{U_h}\left(u_h;\theta \mid \boldsymbol{\tau}_h\right) &= \frac{\exp\left(-u_h\Lambda_{i\cdot,c}(\boldsymbol{\tau}_h)\right)}{\mathcal{L}\left(\Lambda_{i\cdot,c}(\boldsymbol{\tau}_h)\right)} f_U(u_i;\theta), \\
L_{\mathsf{M},h}\left(\boldsymbol{\zeta},\boldsymbol{\beta},\theta\right) &= \int_0^\infty L_{\mathsf{C},h}\left(\boldsymbol{\zeta},\boldsymbol{\beta} \mid u_h\right) f_{U_h}\left(u_h;\theta \mid \boldsymbol{\tau}_h\right)\mathrm{d}u_h.
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathsf{E}\left(U^{\alpha};\hat{\boldsymbol{\xi}} \mid \boldsymbol{x}_h,\boldsymbol{y}_h,\boldsymbol{\tau}_h\right) &= \frac{\displaystyle\int_0^\infty u_h^{d_h+\alpha}\exp\left\{-u_h\Lambda_{i\cdot,c}(\boldsymbol{y}_h)\right\} f_{U_h}(u_h;\theta)\mathrm{d}u_h}{\displaystyle\int_0^\infty u_h^{d_h}\exp\left\{-u_h\Lambda_{i\cdot,c}(\boldsymbol{y}_h)\right\} f_{U_h}(u_h;\theta)\mathrm{d}u_h} \\
&= \frac{\mathsf{E}\left[U^{d_h+\alpha}\exp\left\{-U\Lambda_{i\cdot,c}(\boldsymbol{y}_h)\right\}\right]}{\mathsf{E}\left[U^{d_h}\exp\left\{-U\Lambda_{i\cdot,c}(\boldsymbol{y}_h)\right\}\right]}.
\end{aligned}$$

□

# Appendix B

# Further material

## B.1 Wide to long format for multi-state data

The transformation from wide to long format is crucial for multi-state modelling, as well as for frailty multi-state models. We give here a small example with the following data

```
> cdata
  Hospital Gender Age LR.time LR.status DM.time DM.status De.time De.status
1  Ferrara      M  59       8         0       8         0       8         0
2  Bologna      F  57       7         0       7         1      14         1
3   Ancona      M  53       2         1       2         0      17         0
```

referred to the multi-state structure in Figure 4.1. There are three patients, two males and a female, coming from three hospitals. For each of them the times of the possible events are reported in the variables `event.time`, whereas the variables `event.status` are 0 if the time is censored or 1 if observed.

Thanks to the `mstate` library, the transition matrix is built, with the possible transitions, denoted by progressive natural numbers.

```
> library(mstate)
> tmat <- transMat(list( 2:4, 4, 4, c()),
+                  c("AliveNED", "LR", "DM", "De"))
> tmat
          to
from       AliveNED LR DM De
  AliveNED       NA  1  2  3
  LR             NA NA NA  4
  DM             NA NA NA  5
  De             NA NA NA NA
```

The data `cdata` are transformed into long format data `cdata.ms` by means of the `msprep()` function (Putter, 2011).

```
> cdata.ms <- msprep(time   = c(NA, paste(c("LR", "DM", "De"), "time",
```

```
+                                        sep=".")),
+                    status = c(NA, paste(c("LR", "DM", "De"), "status",
+                                        sep=".")),
+                    keep   = c("Hospital", "Gender", "Age",
+                               "LR.time", "DM.time"),
+                    trans  = tmat,
+                    data   = cdata)
>
> cdata.ms
An object of class 'msdata'

Data:
   id from to trans Tstart Tstop time status Hospital Gender Age LR.time DM.time
1   1    1  2     1      0     8    8      0  Ferrara      M  59       8       8
2   1    1  3     2      0     8    8      0  Ferrara      M  59       8       8
3   1    1  4     3      0     8    8      0  Ferrara      M  59       8       8
4   2    1  2     1      0     7    7      0  Bologna      F  57       7       7
5   2    1  3     2      0     7    7      1  Bologna      F  57       7       7
6   2    1  4     3      0     7    7      0  Bologna      F  57       7       7
7   2    3  4     5      7    14    7      1  Bologna      F  57       7       7
8   3    1  2     1      0     2    2      1   Ancona      M  53       2       2
9   3    1  3     2      0     2    2      0   Ancona      M  53       2       2
10  3    1  4     3      0     2    2      0   Ancona      M  53       2       2
11  3    2  4     4      2    17   15      0   Ancona      M  53       2       2
```

Each observation has been split into one line for each transition for which the patient has been at risk, with

`id` the number of the original record in the dataset `cdata`,

`trans` the transition number as referred to in the `tmat` matrix,

`from` the associated starting state,

`to` the associated arrival state,

`Tstart` the left-truncation time,

`Tstop` the event or censoring times,

`time` the gap between them,

`status` the event–censoring indicator.

The times of the intermediate events (`LR.time` and `DM.time`) have been kept in order to possibly use them as covariates for transitions 4 and 5 in Markov-extended models.

If we assume that a given covariate has different effects for different events, transition-specific variables can be easily derived thanks to the `expand.covs()` function. To create, for instance, such dummy variables for `age` we type

```
> expand.covs(cdata.ms, "Age")
   id from to trans   ...   Age   ...   Age.1 Age.2 Age.3 Age.4 Age.5
1   1    1  2     1   ...    59   ...      59     0     0     0     0
```

```
2   1    1  3     2   ...    59   ...       0    59     0     0     0
3   1    1  4     3   ...    59   ...       0     0    59     0     0
4   2    1  2     1   ...    57   ...      57     0     0     0     0
5   2    1  3     2   ...    57   ...       0    57     0     0     0
6   2    1  4     3   ...    57   ...       0     0    57     0     0
7   2    3  4     5   ...    57   ...       0     0     0     0    57
8   3    1  2     1   ...    53   ...      53     0     0     0     0
9   3    1  3     2   ...    53   ...       0    53     0     0     0
10  3    1  4     3   ...    53   ...       0     0    53     0     0
11  3    2  4     4   ...    53   ...       0     0     0    53     0
```

We obtain one more variable for each possible transition. Each dummy variable has the age value only for lines corresponding to the associated transition; it is 0 otherwise. When used in regression models, their regression coefficients will be interpretable as the ratio between the transition-specific hazards of two patients with one year of difference.

In addition, the effect of the covariate can be estimated only for some of the transitions. If, for example, we assume that the age has influence only on death risks (`AliveNED`→`De`, `LR`→`De` and `DM`→`De`), the regression model will be like

```
> coxph(Surv(Tstart, Tstop, status)~Age.3 + Age.4 + Age.5 +
+       strata(trans), data=cdata.ms)
```

Finally, the baselines of different transitions can be modelled as proportional to each other. For example, for the three transitions to death we can do

```
> cdata.ms <- msprep(time   = c(NA, paste(c("LR", "DM", "De"),
+                                  "time",   sep=".")),
+                    status = c(NA, paste(c("LR", "DM", "De"),
+                                  "status", sep=".")),
+                    keep   = c("Hospital", "Gender", "Age",
+                                  "LR.time", "DM.time",
+                                  #### for proportionality ####
+                                  "LR.status", "DM.status"),
+                                  ##############################
+                    trans  = tmat,
+                    data   = cdata)
> cdata.ms$strata <- cdata.ms$trans
> cdata.ms[cdata.ms$trans %in% c(3, 4, 5), "strata"] <- 3
> cdata.ms
An object of class 'msdata'

Data:
   id from to trans        ...   strata  LR.status  DM.status
1   1    1  2     1        ...        1          0          0
2   1    1  3     2        ...        2          0          0
```

```
3   1    1  4     3          ...          3          0          0
4   2    1  2     1          ...          1          0          1
5   2    1  3     2          ...          2          0          1
6   2    1  4     3          ...          3          0          1
7   2    3  4     5          ...          3          0          1
8   3    1  2     1          ...          1          1          0
9   3    1  3     2          ...          2          1          0
10  3    1  4     3          ...          3          1          0
11  3    2  4     4          ...          3          1          0
> coxph(Surv(Tstart, Tstop, status)~LR.status + DM.status +
+       strata(strata), data=cdata.ms)
```

In such a model, a common baseline is estimated for transitions 3, 4 and 5, and the coefficients of the dummy regressors `LR.status` and `DM.status` are the log-ratios between their risks.

## B.2 Updated frailty distributions

| | **Frailty distribution** |
|---|---|
| **Marginal** | **Updated** |
| $f_U(u)$ | $f_{U_h}(u \mid T_h > t)$ |
| $\mathsf{Gam}^*(\theta)$ | $\mathsf{Gam}\left(\dfrac{1}{1+\theta\boldsymbol{\Lambda}(t)}, \theta\right)$ |
| $\mathsf{IG}^*(\theta)$ | $\mathsf{PVF}\left(\left(1+2\theta\boldsymbol{\Lambda}(t)\right)^{-\frac{1}{2}}, \theta\left(1+2\theta\boldsymbol{\Lambda}(t)\right)^{-\frac{1}{2}}, \dfrac{1}{2}\right)$ |
| $\mathsf{PS}^*(\nu)$ | $\mathsf{PVF}\left((1-\nu)\left(\boldsymbol{\Lambda}(t)\right)^{-\nu}, \dfrac{\nu}{1-\nu}\left(\boldsymbol{\Lambda}(t)\right)^{\nu-1}, \nu\right)$ |
| $\mathsf{PVF}^*(\theta,\nu)$ | $\mathsf{PVF}\left(\left(1+\dfrac{\theta\boldsymbol{\Lambda}(t)}{\nu}\right)^{-\nu}, \theta\left(1+\dfrac{\theta\boldsymbol{\Lambda}(t)}{\nu}\right)^{\nu-1}, \nu\right)$ |
| $\mathsf{CP}^*(\theta,\nu)$ | $\mathsf{CP}\left(\left(1+\dfrac{\theta\boldsymbol{\Lambda}(t)}{\nu}\right)^{-\nu}, \theta\left(1+\dfrac{\theta\boldsymbol{\Lambda}(t)}{\nu}\right)^{\nu-1}, \nu\right)$ |

Table B.1 – Updated frailty distributions in the univariate case, given that the event of interest has not occurred at present time $t$. $\boldsymbol{\Lambda}(t) = \sum_{i=1}^{n_h} \Lambda_{hi}(\breve{y}(t))$.

## B.3 Detailed results of the simulation study on nested frailties multi-state models

Detailed results are here provided for the simulation study discussed in Section 5.3.
Three families of models are considered and they are denoted as follows:
**MSM**: multi-state model without frailties
**SFM**: multi-state model with shared frailties
**NFM**: multi-state model with nested frailties
The following tables show

| | |
|---|---|
| MEAN | the mean of the parameter estimates |
| eSE | the empirical standard error, i.e. the standard error of the parameter |
| mSE | the mean standard error, i.e. the mean of the estimated standard error of the parameter estimates |

over 200 repetitions.

### B.3.1 Parametric models

Scenario **A** and $\vartheta = 0$

| | **True** | **MSM** | | **NFM** | |
|---|---|---|---|---|---|
| | **value** | MEAN (eSE | mSE) | MEAN (eSE | mSE) |
| $\lambda_1$ | 0.03 | 3.69e-02 (1.44e-02 | 2.75e+02) | 4.19e+00 (2.03e+01 | 5.75e+01) |
| $\lambda_2$ | 0.00 | 1.28e-03 (1.30e-03 | 8.15e+00) | 4.86e+00 (6.00e+01 | 3.92e+00) |
| $\lambda_3$ | 0.00 | 4.87e-04 (6.91e-04 | 4.07e-03) | 2.20e+02 (2.91e+03 | 4.09e+00) |
| $\lambda_4$ | 0.01 | 1.63e-02 (1.87e-02 | 3.95e+01) | 2.57e+43 (3.41e+44 | 1.54e+22) |
| $\lambda_5$ | 0.03 | 6.14e-01 (2.68e+00 | 5.66e+02) | 2.01e+25 (2.59e+26 | 6.89e+24) |
| $\rho_1$ | 0.80 | 7.41e-01 (9.13e-02 | 8.44e-02) | 7.78e-01 (1.24e-01 | 8.47e-02) |
| $\rho_2$ | 1.30 | 1.35e+00 (2.85e-01 | 2.85e-01) | 1.17e+00 (4.97e-01 | 2.48e-01) |
| $\rho_3$ | 1.40 | 1.51e+00 (4.24e-01 | 4.45e-01) | 1.13e+00 (1.07e+00 | 3.22e-01) |
| $\rho_4$ | 1.40 | 1.25e+00 (2.76e-01 | 2.77e-01) | 1.46e+00 (6.46e-01 | 2.78e-01) |
| $\rho_5$ | 1.70 | 1.54e+00 (5.95e-01 | 4.92e-01) | 3.11e+00 (3.03e+00 | 8.09e-01) |
| $\beta_1$ | -0.69 | -5.86e-01 (2.38e-01 | 2.56e-01) | -6.18e-01 (4.44e-01 | 2.61e-01) |
| $\beta_2$ | -0.11 | -1.38e-01 (5.64e-01 | 7.73e-09) | 7.63e-02 (1.50e+00 | 7.02e-01) |
| $\beta_3$ | 0.00 | -6.62e-02 (2.83e+00 | 4.38e-04) | 3.12e-01 (2.06e+00 | 1.25e+00) |
| $\beta_4$ | 0.00 | 2.16e-03 (3.44e-01 | 2.85e-09) | 6.53e-02 (8.74e-01 | 4.15e-01) |
| $\beta_5$ | 0.00 | -4.22e-02 (9.15e-01 | 2.35e-07) | 1.05e+00 (3.66e+00 | 1.47e+02) |
| $\theta_V$ | 0.50 | ( | ) | 3.22e+17 (4.19e+18 | 1.39e+21) |
| $\theta_W$ | 1.50 | ( | ) | 1.49e+00 (1.94e+00 | 9.92e-02) |

Scenario **B** and $\vartheta = 0$

| | **True** | **MSM** | | **NFM** | |
|---|---|---|---|---|---|
| | **value** | MEAN (eSE | mSE) | MEAN (eSE | mSE) |
| $\lambda_1$ | 0.03 | 3.64e-02 (9.18e-03 | 5.39e-03) | 2.76e+01 (1.91e+02 | 2.23e+02) |
| $\lambda_2$ | 0.00 | 1.10e-03 (5.02e-04 | 4.81e-04) | 1.61e+01 (1.34e+02 | 9.31e+01) |
| $\lambda_3$ | 0.00 | 4.32e-04 (3.31e-04 | 2.62e-04) | 6.63e+01 (6.36e+02 | 1.63e+02) |
| $\lambda_4$ | 0.01 | 1.33e-02 (7.74e-03 | 6.34e-03) | 4.73e+16 (5.40e+17 | 4.95e+15) |
| $\lambda_5$ | 0.03 | 1.11e-01 (2.57e-01 | 8.14e-02) | 4.42e+18 (6.22e+19 | 3.94e+18) |
| $\rho_1$ | 0.80 | 7.38e-01 (4.06e-02 | 3.68e-02) | 7.92e-01 (7.31e-02 | 3.84e-02) |
| $\rho_2$ | 1.30 | 1.27e+00 (1.16e-01 | 1.20e-01) | 1.17e+00 (3.36e-01 | 1.09e-01) |
| $\rho_3$ | 1.40 | 1.38e+00 (1.77e-01 | 1.74e-01) | 1.13e+00 (5.14e-01 | 1.44e-01) |
| $\rho_4$ | 1.40 | 1.19e+00 (1.41e-01 | 1.19e-01) | 1.73e+00 (3.93e-01 | 1.24e-01) |
| $\rho_5$ | 1.70 | 1.38e+00 (2.87e-01 | 1.83e-01) | 2.26e+00 (4.55e-01 | 2.14e-01) |
| $\beta_1$ | -0.69 | -5.99e-01 (1.19e-01 | 1.12e-01) | -6.69e-01 (3.10e-01 | 1.14e-01) |
| $\beta_2$ | -0.11 | -9.78e-02 (2.22e-01 | 3.15e-09) | -1.09e-01 (8.25e-01 | 2.33e-01) |
| $\beta_3$ | 0.00 | 9.23e-03 (3.33e-01 | 4.03e-09) | 2.21e-01 (1.57e+00 | 4.21e-01) |
| $\beta_4$ | 0.00 | 3.46e-03 (1.77e-01 | 1.51e-09) | -3.80e-02 (4.50e-01 | 1.67e-01) |
| $\beta_5$ | 0.00 | 1.56e-02 (3.45e-01 | 2.39e-09) | 7.78e-02 (1.23e+00 | 3.01e-01) |
| $\theta_V$ | 0.50 | ( | ) | 5.05e+21 (7.15e+22 | 2.04e+14) |
| $\theta_W$ | 1.50 | ( | ) | 2.54e+00 (2.10e+00 | 2.48e-02) |

Scenario **C** and $\vartheta = 0$

| | **True** | **MSM** | | **NFM** | |
|---|---|---|---|---|---|
| | **value** | MEAN (eSE | mSE) | MEAN (eSE | mSE) |
| $\lambda_1$ | 0.03 | 3.55e-02 (5.23e-03 | 2.37e-03) | 5.31e+02 (4.66e+03 | 1.76e+03) |
| $\lambda_2$ | 0.00 | 1.09e-03 (2.58e-04 | 2.16e-04) | 5.28e+03 (7.41e+04 | 7.00e+02) |
| $\lambda_3$ | 0.00 | 3.73e-04 (1.19e-04 | 1.08e-04) | 6.30e+03 (7.24e+04 | 7.24e+03) |
| $\lambda_4$ | 0.01 | 1.29e-02 (4.35e-03 | 2.75e-03) | 2.98e+20 (4.21e+21 | 5.22e+20) |
| $\lambda_5$ | 0.03 | 7.08e-02 (3.98e-02 | 2.17e-02) | 1.57e+17 (2.21e+18 | 1.73e+20) |
| $\rho_1$ | 0.80 | 7.36e-01 (2.10e-02 | 1.65e-02) | 7.96e-01 (3.15e-02 | 1.73e-02) |
| $\rho_2$ | 1.30 | 1.27e+00 (5.14e-02 | 5.20e-02) | 1.25e+00 (1.36e-01 | 5.09e-02) |
| $\rho_3$ | 1.40 | 1.37e+00 (7.42e-02 | 7.67e-02) | 1.25e+00 (2.29e-01 | 6.99e-02) |
| $\rho_4$ | 1.40 | 1.17e+00 (7.49e-02 | 5.24e-02) | 1.87e+00 (1.43e-01 | 5.41e-02) |
| $\rho_5$ | 1.70 | 1.34e+00 (1.56e-01 | 7.77e-02) | 2.10e+00 (1.75e-01 | 8.23e-02) |
| $\beta_1$ | -0.69 | -6.04e-01 (5.51e-02 | 5.02e-02) | -6.84e-01 (1.27e-01 | 5.05e-02) |
| $\beta_2$ | -0.11 | -1.04e-01 (8.85e-02 | 2.05e-09) | -7.88e-02 (2.82e-01 | 9.42e-02) |
| $\beta_3$ | 0.00 | 1.23e-02 (1.30e-01 | 2.78e-09) | 9.52e-02 (6.12e-01 | 1.34e-01) |
| $\beta_4$ | 0.00 | 1.02e-02 (7.07e-02 | 1.04e-09) | -9.19e-02 (2.10e-01 | 7.13e-02) |
| $\beta_5$ | 0.00 | 2.12e-02 (1.62e-01 | 1.04e-09) | -4.76e-02 (4.14e-01 | 1.07e-01) |
| $\theta_V$ | 0.50 | ( | ) | 3.19e+16 (4.50e+17 | 8.34e+16) |
| $\theta_W$ | 1.50 | ( | ) | 4.50e+00 (1.87e+00 | 1.83e-02) |

### B.3.2 Semiparametric models

Scenario **A** and $\vartheta = 0$

| | **True value** | **MSM** MEAN (eSE \| mSE) | **SFM** MEAN (eSE \| mSE) | **NFM** MEAN (eSE \| mSE) |
|---|---|---|---|---|
| $\beta_1$ | -1.39 | -1.10 (0.26 \| 0.25) | -1.21 (0.26 \| 0.25) | -1.45 (0.29 \| 0.25) |
| $\beta_2$ | -0.22 | -0.19 (0.33 \| 0.32) | -0.33 (0.35 \| 0.33) | -0.23 (0.33 \| 0.32) |
| $\beta_3$ | 0.00 | 0.02 (0.47 \| 0.43) | -0.13 (0.48 \| 0.43) | 0.01 (0.48 \| 0.43) |
| $\beta_4$ | 1.10 | 0.91 (0.39 \| 0.27) | 0.92 (0.36 \| 0.28) | 1.11 (0.32 \| 0.29) |
| $\beta_5$ | 1.61 | 3.44 (5.72 \| 281.35) | 3.76 (14.70 \| 1244.54) | 1.74 (2.53 \| 42.92) |

| | **True value** | **SFM** MEAN ( eSE ) | **NFM** MEAN ( eSE ) |
|---|---|---|---|
| $\theta_V$ | 0.50 | 3.30e-01 (1.94e-01) | 2.53e-09 (2.42e-10) |
| $\theta_W$ | 1.00 | ——— (———) | 3.41e+00 (1.07e+00) |

Scenario **B** and $\vartheta = 0$

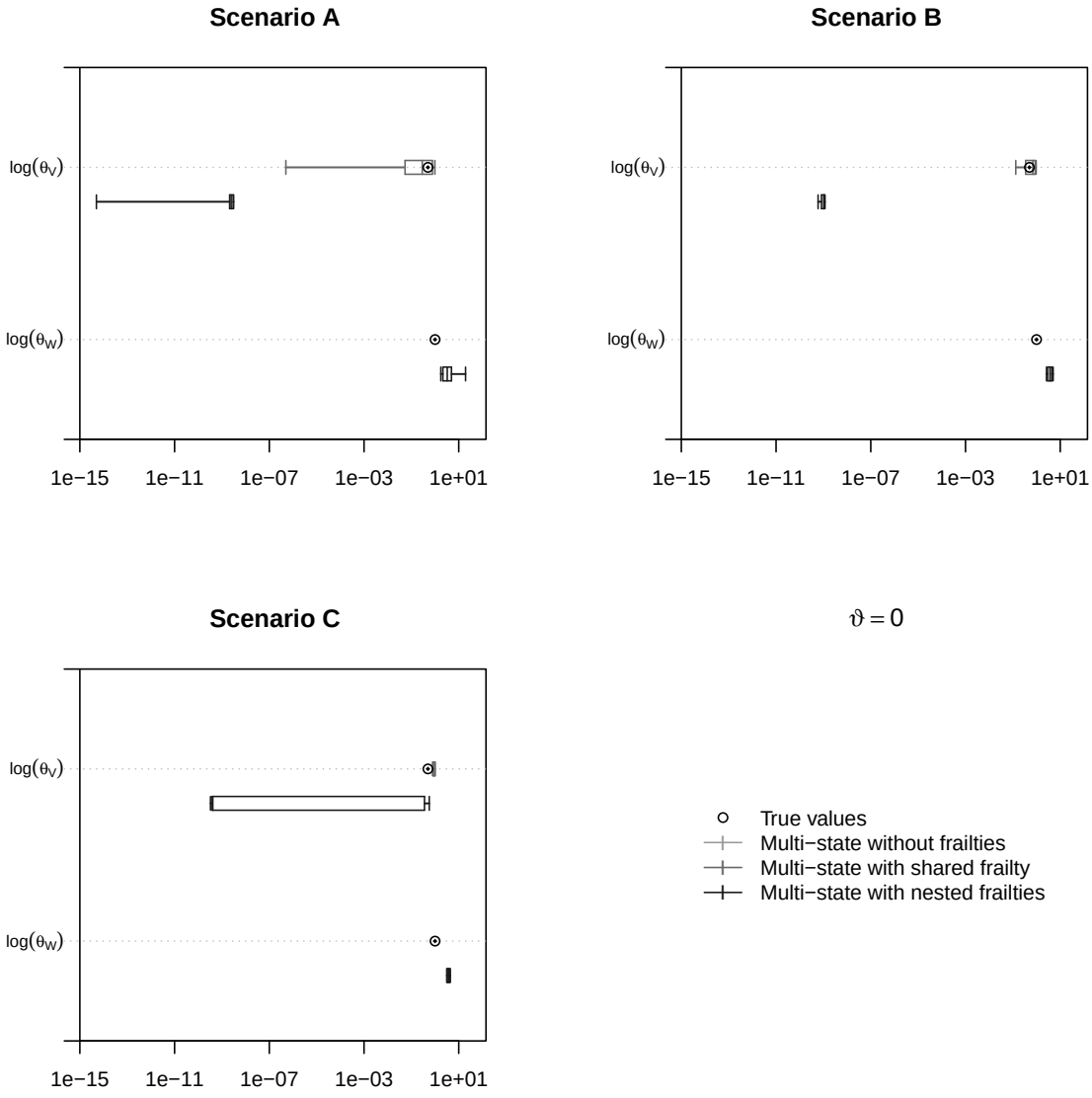
| | **True value** | **MSM** MEAN (eSE \| mSE) | **SFM** MEAN (eSE \| mSE) | **NFM** MEAN (eSE \| mSE) |
|---|---|---|---|---|
| $\beta_1$ | -1.39 | -1.11 (0.13 \| 0.11) | -1.21 (0.12 \| 0.11) | -1.45 (0.13 \| 0.11) |
| $\beta_2$ | -0.22 | -0.18 (0.15 \| 0.14) | -0.33 (0.15 \| 0.14) | -0.22 (0.14 \| 0.14) |
| $\beta_3$ | 0.00 | 0.01 (0.21 \| 0.18) | -0.15 (0.22 \| 0.18) | -0.01 (0.20 \| 0.18) |
| $\beta_4$ | 1.10 | 0.86 (0.21 \| 0.12) | 0.92 (0.17 \| 0.12) | 1.11 (0.13 \| 0.12) |
| $\beta_5$ | 1.61 | 1.43 (0.46 \| 0.19) | 1.41 (0.39 \| 0.20) | 1.58 (0.25 \| 0.22) |

| | **True value** | **SFM** MEAN ( eSE ) | **NFM** MEAN ( eSE ) |
|---|---|---|---|
| $\theta_V$ | 0.50 | 6.69e-01 (1.38e-01) | 1.02e-09 (7.40e-11) |
| $\theta_W$ | 1.00 | ——— (———) | 3.69e+00 (3.79e-01) |

Scenario **C** and $\vartheta = 0$

| | **True value** | **MSM** MEAN (eSE \| mSE) | **SFM** MEAN (eSE \| mSE) | **NFM** MEAN (eSE \| mSE) |
|---|---|---|---|---|
| $\beta_1$ | -1.39 | -1.08 (0.07 \| 0.05) | -1.19 (0.06 \| 0.05) | -1.43 (0.05 \| 0.05) |
| $\beta_2$ | -0.22 | -0.18 (0.08 \| 0.06) | -0.33 (0.09 \| 0.06) | -0.21 (0.06 \| 0.06) |
| $\beta_3$ | 0.00 | 0.01 (0.09 \| 0.08) | -0.14 (0.09 \| 0.08) | 0.02 (0.08 \| 0.08) |
| $\beta_4$ | 1.10 | 0.84 (0.12 \| 0.05) | 0.91 (0.09 \| 0.05) | 1.10 (0.05 \| 0.05) |
| $\beta_5$ | 1.61 | 1.44 (0.26 \| 0.07) | 1.44 (0.16 \| 0.07) | 1.61 (0.08 \| 0.07) |

| | True value | SFM MEAN ( eSE ) | NFM MEAN ( eSE ) |
|---|---|---|---|
| $\theta_V$ | 0.50 | 9.38e-01 (4.73e-02) | 3.22e-02 (9.39e-02) |
| $\theta_W$ | 1.00 | ——— (———) | 3.71e+00 (2.24e-01) |



Scenario **A** and $\vartheta = 0.5$

| | **True value** | **MSM** MEAN (eSE \| mSE) | **SFM** MEAN (eSE \| mSE) | **NFM** MEAN (eSE \| mSE) |
|---|---|---|---|---|
| $\beta_1$ | -1.39 | -1.13 (0.30 \| 0.25) | -1.23 (0.29 \| 0.26) | -1.47 (0.31 \| 0.26) |
| $\beta_2$ | -0.22 | -0.09 (0.38 \| 0.35) | -0.21 (0.39 \| 0.35) | -0.10 (0.35 \| 0.33) |
| $\beta_3$ | 0.00 | 0.11 (0.52 \| 0.46) | -0.02 (0.53 \| 0.46) | 0.10 (0.52 \| 0.45) |
| $\beta_4$ | 1.10 | 1.00 (0.42 \| 0.28) | 1.05 (0.36 \| 0.29) | 1.28 (0.36 \| 0.31) |
| $\beta_5$ | 1.61 | 4.49 (7.38 \| 487.79) | 5.10 (11.49 \| 2688.83) | 1.96 (3.63 \| 108.67) |

| | **True value** | **SFM** MEAN ( eSE ) | **NFM** MEAN ( eSE ) |
|---|---|---|---|
| $\theta_V$ | 0.50 | 3.26e-01 (1.95e-01) | 2.59e-09 (3.09e-10) |
| $\theta_W$ | 1.00 | ——— (———) | 3.83e+00 (1.31e+00) |

Scenario **B** and $\vartheta = 0.5$

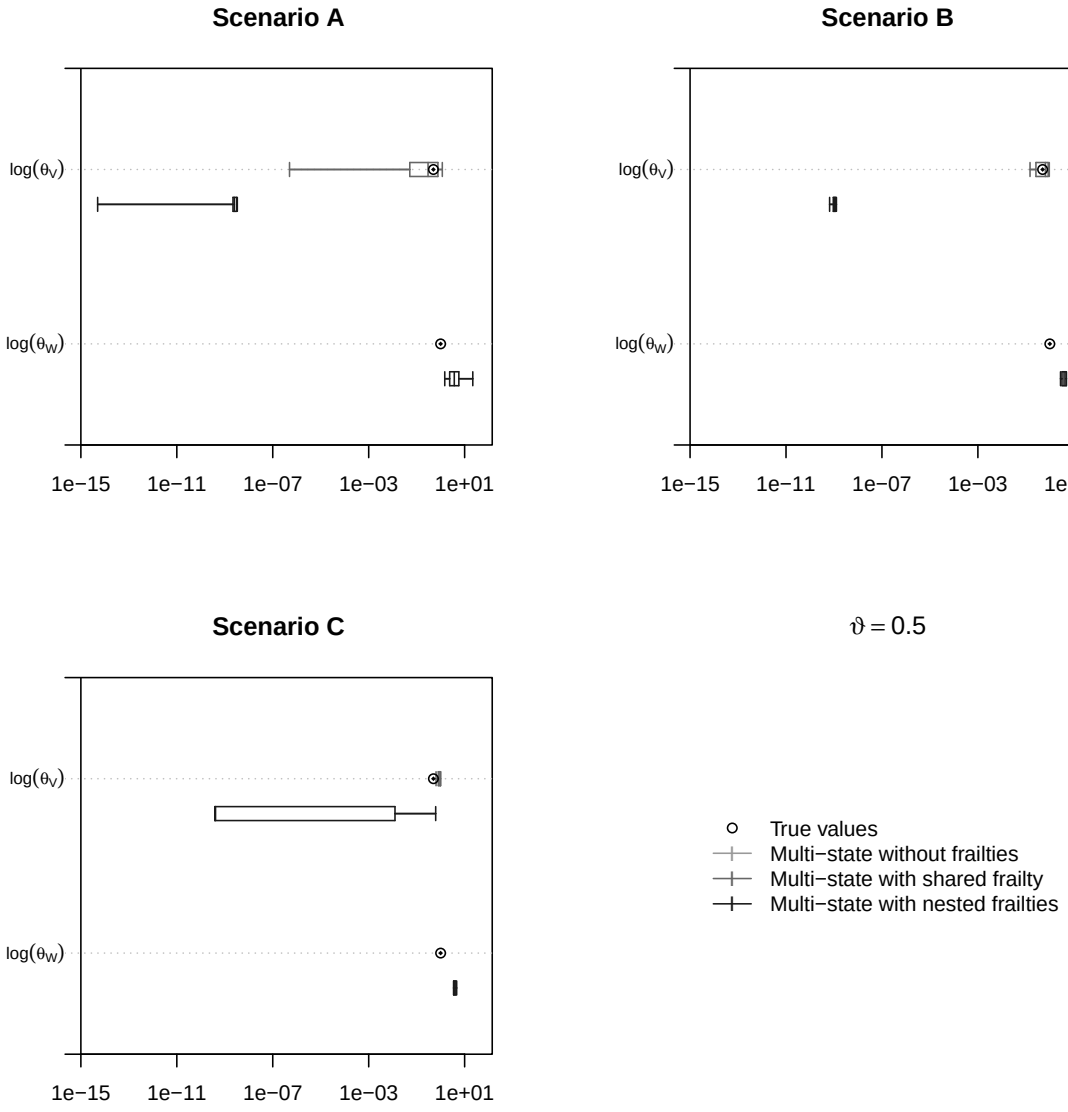
| | **True value** | **MSM** MEAN (eSE \| mSE) | **SFM** MEAN (eSE \| mSE) | **NFM** MEAN (eSE \| mSE) |
|---|---|---|---|---|
| $\beta_1$ | -1.39 | -1.12 (0.13 \| 0.11) | -1.21 (0.12 \| 0.11) | -1.46 (0.12 \| 0.11) |
| $\beta_2$ | -0.22 | -0.10 (0.17 \| 0.15) | -0.23 (0.17 \| 0.15) | -0.10 (0.15 \| 0.15) |
| $\beta_3$ | 0.00 | 0.12 (0.23 \| 0.19) | -0.01 (0.24 \| 0.19) | 0.15 (0.21 \| 0.19) |
| $\beta_4$ | 1.10 | 0.95 (0.23 \| 0.12) | 1.05 (0.18 \| 0.12) | 1.25 (0.13 \| 0.12) |
| $\beta_5$ | 1.61 | 1.17 (0.55 \| 0.21) | 1.11 (0.52 \| 0.22) | 1.44 (0.38 \| 0.24) |

| | **True value** | **SFM** MEAN ( eSE ) | **NFM** MEAN ( eSE ) |
|---|---|---|---|
| $\theta_V$ | 0.50 | 6.31e-01 (1.48e-01) | 1.05e-09 (6.51e-11) |
| $\theta_W$ | 1.00 | ——— (———) | 3.94e+00 (3.83e-01) |

Scenario **C** and $\vartheta = 0.5$

| | **True value** | **MSM** MEAN (eSE \| mSE) | **SFM** MEAN (eSE \| mSE) | **NFM** MEAN (eSE \| mSE) |
|---|---|---|---|---|
| $\beta_1$ | -1.39 | -1.11 (0.07 \| 0.05) | -1.20 (0.06 \| 0.05) | -1.44 (0.05 \| 0.05) |
| $\beta_2$ | -0.22 | -0.11 (0.09 \| 0.06) | -0.23 (0.09 \| 0.07) | -0.09 (0.07 \| 0.06) |
| $\beta_3$ | 0.00 | 0.11 (0.10 \| 0.08) | -0.01 (0.11 \| 0.08) | 0.16 (0.09 \| 0.08) |
| $\beta_4$ | 1.10 | 0.92 (0.13 \| 0.05) | 1.04 (0.10 \| 0.05) | 1.23 (0.05 \| 0.05) |
| $\beta_5$ | 1.61 | 1.16 (0.32 \| 0.07) | 1.13 (0.24 \| 0.08) | 1.43 (0.14 \| 0.08) |

| | True value | SFM MEAN ( eSE ) | NFM MEAN ( eSE ) |
|---|---|---|---|
| $\theta_V$ | 0.50 | 9.14e-01 (5.66e-02) | 6.53e-03 (5.09e-02) |
| $\theta_W$ | 1.00 | ——— (———) | 3.91e+00 (2.00e-01) |



Scenario **A** and $\vartheta = 1.5$

| | **True value** | **MSM** MEAN (eSE \| mSE) | **SFM** MEAN (eSE \| mSE) | **NFM** MEAN (eSE \| mSE) |
|---|---|---|---|---|
| $\beta_1$ | -1.39 | -1.18 (0.28 \| 0.25) | -1.48 (4.54 \| 0.26) | -1.54 (0.31 \| 0.26) |
| $\beta_2$ | -0.22 | 0.02 (0.40 \| 0.37) | -3246.71 (72013.21 \| 0.37) | 0.03 (0.39 \| 0.35) |
| $\beta_3$ | 0.00 | 0.56 (1.98 \| 28.08) | 0.34 (1.80 \| 225.87) | 0.39 (0.60 \| 0.53) |
| $\beta_4$ | 1.10 | 1.13 (0.48 \| 0.29) | 1.18 (0.40 \| 0.31) | 1.43 (0.39 \| 0.33) |
| $\beta_5$ | 1.61 | -1.19 (8.55 \| 515.42) | 5.75 (118.98 \| 2419.26) | -0.06 (4.43 \| 162.43) |

| | **True value** | **SFM** MEAN ( eSE ) | **NFM** MEAN ( eSE ) |
|---|---|---|---|
| $\theta_V$ | 0.50 | 3.18e-01 (1.99e-01) | 2.59e-09 (3.81e-10) |
| $\theta_W$ | 1.00 | ——— (———) | 5.16e+00 (3.88e+00) |

Scenario **B** and $\vartheta = 1.5$

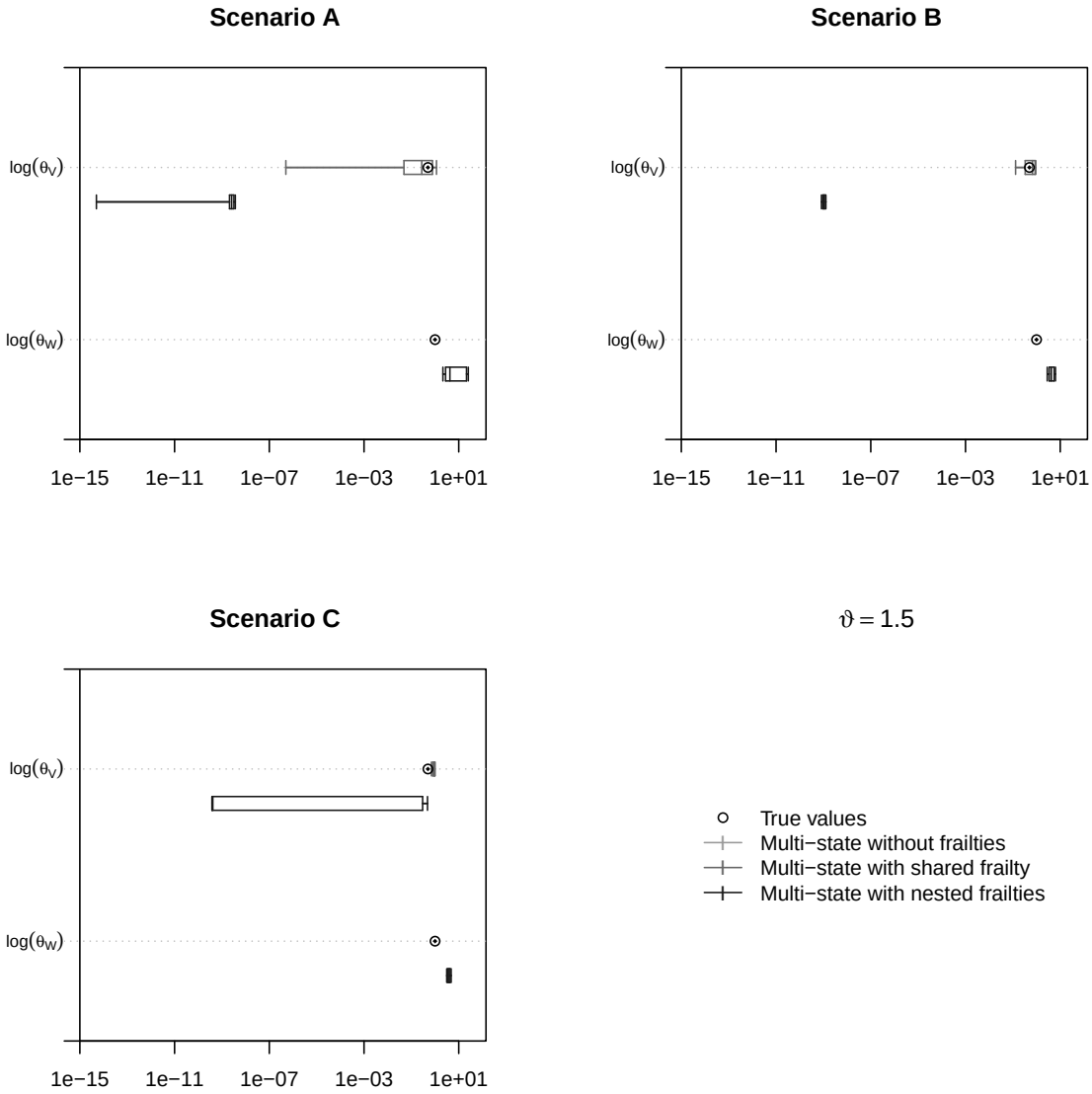
| | **True value** | **MSM** MEAN (eSE \| mSE) | **SFM** MEAN (eSE \| mSE) | **NFM** MEAN (eSE \| mSE) |
|---|---|---|---|---|
| $\beta_1$ | -1.39 | -1.15 (0.14 \| 0.11) | -1.24 (0.13 \| 0.11) | -1.48 (0.12 \| 0.11) |
| $\beta_2$ | -0.22 | -0.01 (0.19 \| 0.16) | -0.12 (0.20 \| 0.16) | 0.05 (0.18 \| 0.15) |
| $\beta_3$ | 0.00 | 0.27 (0.27 \| 0.22) | 0.15 (0.28 \| 0.22) | 0.34 (0.23 \| 0.21) |
| $\beta_4$ | 1.10 | 1.06 (0.28 \| 0.12) | 1.20 (0.19 \| 0.12) | 1.36 (0.15 \| 0.12) |
| $\beta_5$ | 1.61 | -0.06 (0.74 \| 0.24) | -0.13 (0.74 \| 0.25) | 0.60 (0.75 \| 0.26) |

| | **True value** | **SFM** MEAN ( eSE ) | **NFM** MEAN ( eSE ) |
|---|---|---|---|
| $\theta_V$ | 0.50 | 6.42e-01 (1.41e-01) | 1.05e-09 (6.62e-11) |
| $\theta_W$ | 1.00 | ——— (———) | 4.25e+00 (5.09e-01) |

Scenario **C** and $\vartheta = 1.5$

| | **True value** | **MSM** MEAN (eSE \| mSE) | **SFM** MEAN (eSE \| mSE) | **NFM** MEAN (eSE \| mSE) |
|---|---|---|---|---|
| $\beta_1$ | -1.39 | -1.15 (0.08 \| 0.05) | -1.23 (0.07 \| 0.05) | -1.48 (0.05 \| 0.05) |
| $\beta_2$ | -0.22 | -0.02 (0.11 \| 0.07) | -0.13 (0.11 \| 0.07) | 0.04 (0.08 \| 0.07) |
| $\beta_3$ | 0.00 | 0.26 (0.14 \| 0.09) | 0.15 (0.15 \| 0.09) | 0.36 (0.12 \| 0.09) |
| $\beta_4$ | 1.10 | 1.04 (0.17 \| 0.05) | 1.21 (0.10 \| 0.05) | 1.36 (0.07 \| 0.05) |
| $\beta_5$ | 1.61 | 0.23 (0.41 \| 0.09) | 0.18 (0.38 \| 0.09) | 0.88 (0.29 \| 0.09) |

| | True value | SFM MEAN ( eSE ) | NFM MEAN ( eSE ) |
|---|---|---|---|
| $\theta_V$ | 0.50 | 9.01e-01 (5.33e-02) | 1.92e-02 (7.82e-02) |
| $\theta_W$ | 1.00 | ——— (———) | 3.91e+00 (2.43e-01) |

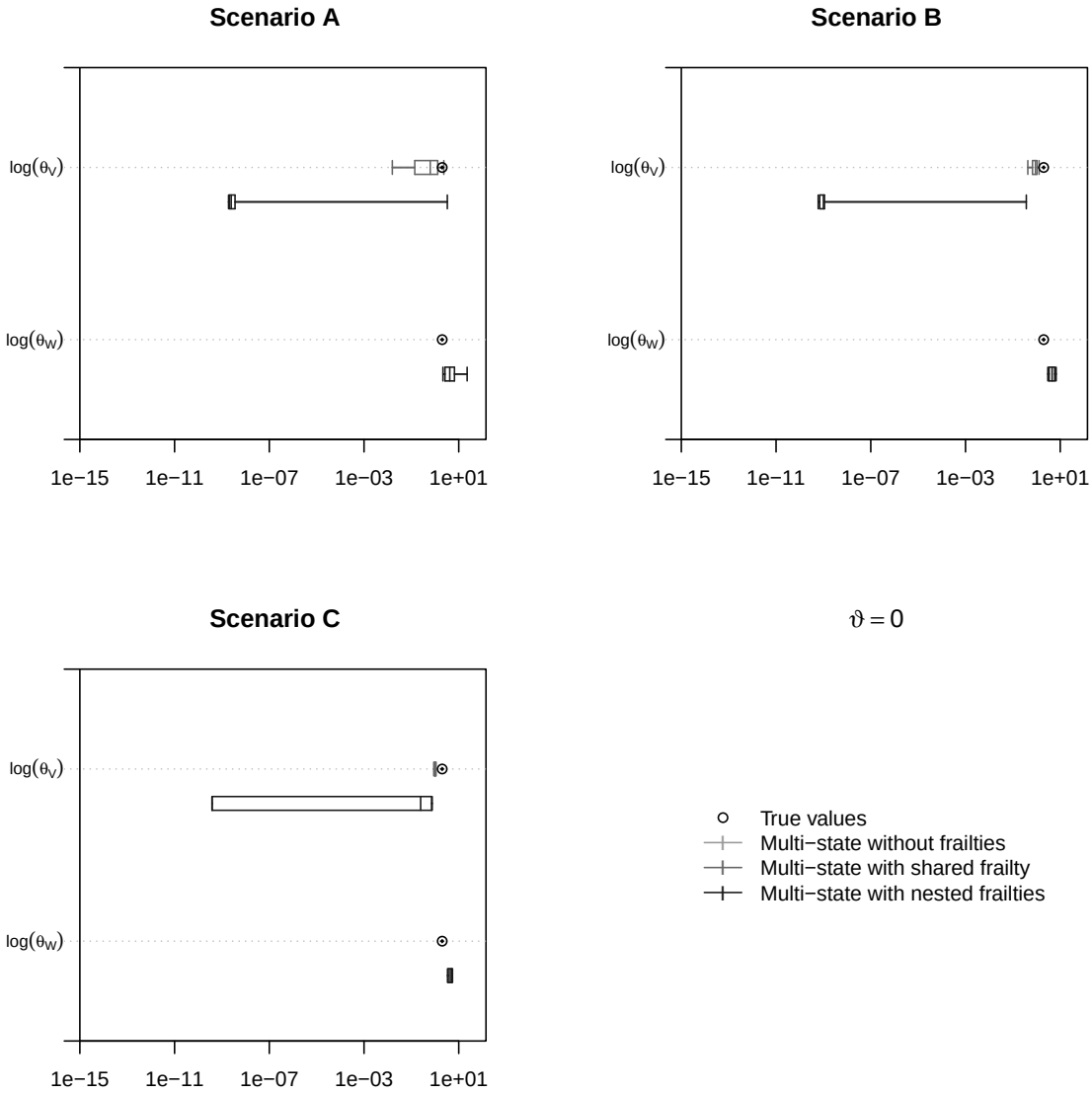


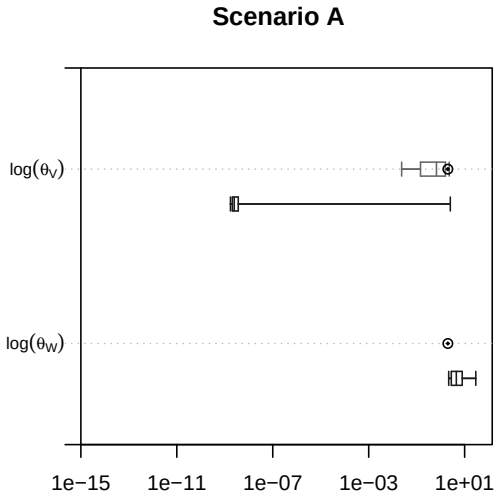
### Strong heterogeneity

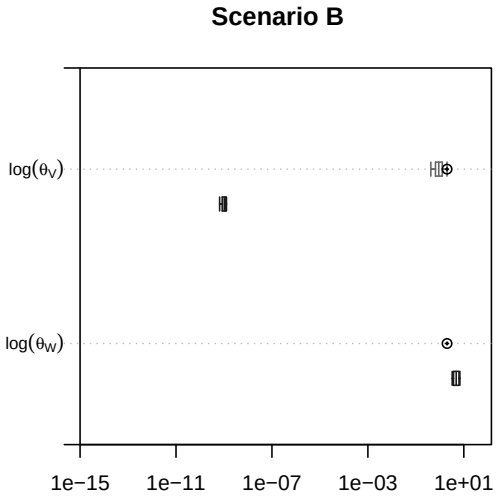
The following graphs show the results of a simulation study analogous to that in Section 5.3, expect for the heterogeneity parameters, now set to

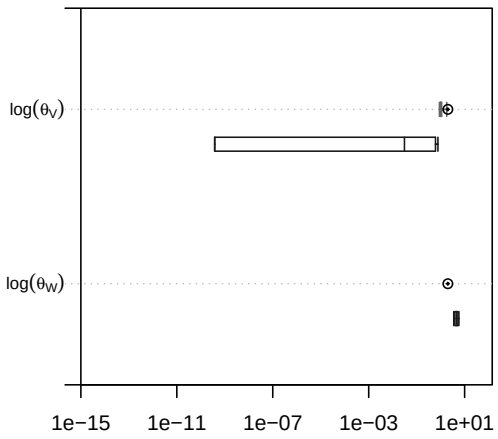
$$\begin{aligned} \theta_V &= 2, \\ \theta_W := \theta_1 = \ldots = \theta_5 &= 2. \end{aligned} \tag{B.1}$$

These values give a total frailty variance of $\mathsf{V}[U_{qh}] = 8$, that is much bigger than the previous study and is much bigger than one can except in real studies. This serves the only purpose of showing that small heterogeneity is not the reason for the poor estimation of the frailty parameters.



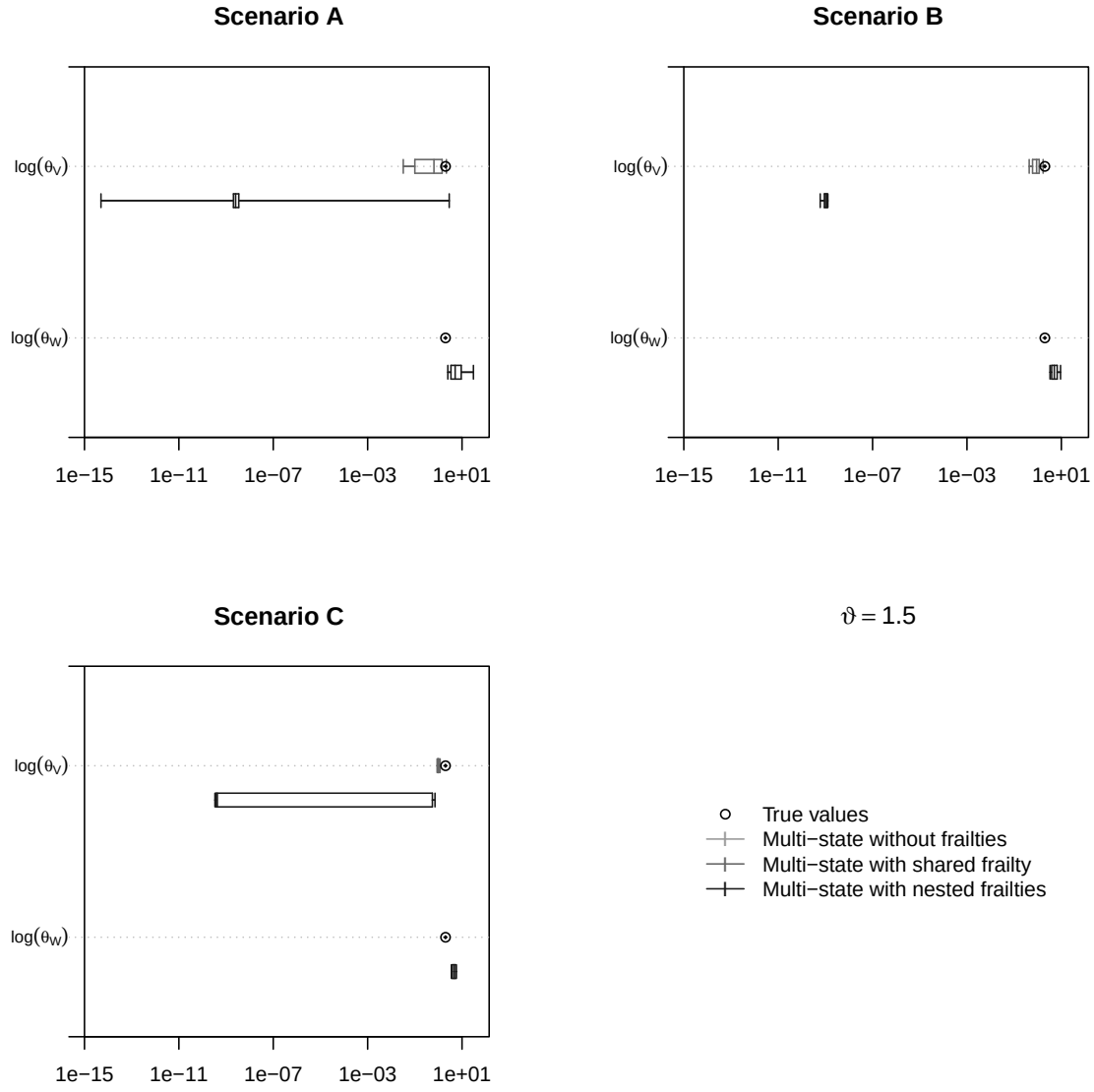



$\vartheta = 0.5$

True values
Multi−state without frailties
Multi−state with shared frailty
Multi−state with nested frailties




## B.4 Detailed results for the bladder cancer study

Note: **treat.1**, . . . , **treat.6** are the regression coefficients for treatment effect on the six transitions, with transitions numbered as

1 from Rand to Rec
2 from Rand to Prog
3 from Rand to De
4 from Rec to Prog
5 from Rec to De
6 from Prog to De

Furthermore, **BigT** is a binary variable for big against small tumours; **RecT** a binary variable for recurrent against non-recurrent tumours; **AGEdev** a continuous variable for age, in decades,

so that the regression parameter estimates the log-hazard ratio for a 10-year difference in age; **SEXMale** a binary variable for males against females. Their effect is evaluated on the risk of all transitions to a recurrent state (**.Rec**), all transitions to a progressive state (**.Prog**), all transitions to death (**.De**), or transitions to progression separately from randomisation (**.RandProg**) or from recurrent state (**.RecProg**).

**MS - Simple multi-state model**

| | estim | HR | se(estim) | p-val |
|---|---|---|---|---|
| **treat.1** | −0.2561 | 0.7741 | 0.0607 | < 0.001 |
| **treat.2** | −0.1742 | 0.8401 | 0.1954 | 0.373 |
| **treat.3** | 0.1488 | 1.1605 | 0.1108 | 0.179 |
| **treat.4** | 0.0458 | 1.0469 | 0.1574 | 0.771 |
| **treat.5** | 0.3250 | 1.3840 | 0.1248 | 0.009 |
| **treat.6** | −0.3343 | 0.7159 | 0.1514 | 0.027 |
| **BigT.Rec** | 0.1588 | 1.1721 | 0.0643 | 0.013 |
| **BigT.Prog** | 0.5544 | 1.7408 | 0.1247 | < 0.001 |
| **ManyT.Rec** | 0.5225 | 1.6863 | 0.0652 | < 0.001 |
| **ManyT.Prog** | 0.6021 | 1.8259 | 0.1295 | < 0.001 |
| **RecT.Rec** | 0.3145 | 1.3696 | 0.0684 | < 0.001 |
| **RecT.RandProg** | 0.8570 | 2.3560 | 0.2092 | < 0.001 |
| **AGEdec.De** | 0.7545 | 2.1266 | 0.0426 | < 0.001 |
| **AGEdec.RecProg** | 0.2656 | 1.3042 | 0.0816 | 0.001 |
| **SEXMale.De** | 0.4969 | 1.6437 | 0.0969 | < 0.001 |

**ST - Stratified multi-state model**

| | estim | HR | se(estim) | p-val |
|---|---|---|---|---|
| **treat.1** | −0.1512 | 0.8597 | 0.0729 | 0.038 |
| **treat.2** | −0.0270 | 0.9734 | 0.2351 | 0.909 |
| **treat.3** | 0.0688 | 1.0712 | 0.1356 | 0.612 |
| **treat.4** | 0.0626 | 1.0646 | 0.1939 | 0.747 |
| **treat.5** | 0.3070 | 1.3593 | 0.1581 | 0.052 |
| **treat.6** | −0.1309 | 0.8773 | 0.2344 | 0.577 |
| **BigT.Rec** | 0.1840 | 1.2020 | 0.0691 | 0.008 |
| **BigT.Prog** | 0.5299 | 1.6988 | 0.1362 | < 0.001 |
| **ManyT.Rec** | 0.5517 | 1.7362 | 0.0705 | < 0.001 |
| **ManyT.Prog** | 0.6078 | 1.8363 | 0.1422 | < 0.001 |
| **RecT.Rec** | 0.3047 | 1.3563 | 0.0775 | < 0.001 |
| **RecT.RandProg** | 0.5979 | 1.8184 | 0.2400 | 0.013 |
| **AGEdec.De** | 0.8044 | 2.2354 | 0.0492 | < 0.001 |
| **AGEdec.RecProg** | 0.2681 | 1.3075 | 0.0893 | 0.003 |
| **SEXMale.De** | 0.6206 | 1.8600 | 0.1125 | < 0.001 |

**SF1 - Shared-frailty multi-state model (hospital)**

| | estim | HR | se(estim) | p-val |
|---|---|---|---|---|
| **treat.1** | −0.2190 | 0.8034 | 0.0640 | $< 0.001$ |
| **treat.2** | −0.1478 | 0.8626 | 0.1969 | 0.453 |
| **treat.3** | 0.1925 | 1.2122 | 0.1127 | 0.088 |
| **treat.4** | 0.1054 | 1.1112 | 0.1593 | 0.508 |
| **treat.5** | 0.4095 | 1.5061 | 0.1276 | 0.001 |
| **treat.6** | −0.2650 | 0.7672 | 0.1549 | 0.087 |
| **BigT.Rec** | 0.1641 | 1.1783 | 0.0654 | 0.012 |
| **BigT.Prog** | 0.5498 | 1.7329 | 0.1251 | $< 0.001$ |
| **ManyT.Rec** | 0.5318 | 1.7019 | 0.0667 | $< 0.001$ |
| **ManyT.Prog** | 0.6097 | 1.8398 | 0.1308 | $< 0.001$ |
| **RecT.Rec** | 0.3231 | 1.3815 | 0.0711 | $< 0.001$ |
| **RecT.RandProg** | 0.8616 | 2.3670 | 0.2105 | $< 0.001$ |
| **AGEdec.De** | 0.7628 | 2.1442 | 0.0430 | $< 0.001$ |
| **AGEdec.RecProg** | 0.2698 | 1.3097 | 0.0828 | 0.001 |
| **SEXMale.De** | 0.5454 | 1.7253 | 0.0979 | $< 0.001$ |
| $\theta_V$ | 0.049 | | | |

**SF2 - Shared-frailty multi-state model (hospital×transition)**

| | estim | HR | se(estim) | p-val |
|---|---|---|---|---|
| **treat.1** | −0.1988 | 0.8197 | 0.0676 | 0.003 |
| **treat.2** | −0.1086 | 0.8971 | 0.2018 | 0.591 |
| **treat.3** | 0.1650 | 1.1794 | 0.1181 | 0.162 |
| **treat.4** | 0.0423 | 1.0432 | 0.1655 | 0.798 |
| **treat.5** | 0.3043 | 1.3556 | 0.1354 | 0.025 |
| **treat.6** | −0.2817 | 0.7545 | 0.1677 | 0.093 |
| **BigT.Rec** | 0.1879 | 1.2067 | 0.0670 | 0.005 |
| **BigT.Prog** | 0.5482 | 1.7301 | 0.1264 | $< 0.001$ |
| **ManyT.Rec** | 0.5370 | 1.7109 | 0.0682 | $< 0.001$ |
| **ManyT.Prog** | 0.6134 | 1.8467 | 0.1319 | $< 0.001$ |
| **RecT.Rec** | 0.3347 | 1.3975 | 0.0744 | $< 0.001$ |
| **RecT.RandProg** | 0.8045 | 2.2355 | 0.2148 | $< 0.001$ |
| **AGEdec.De** | 0.7705 | 2.1609 | 0.0439 | $< 0.001$ |
| **AGEdec.RecProg** | 0.2763 | 1.3182 | 0.0830 | $< 0.001$ |
| **SEXMale.De** | 0.5376 | 1.7119 | 0.0995 | $< 0.001$ |
| $\theta_W$ | 0.110 | | | |

**NF - Nested-frailty multi-state model**

| | estim | HR | se(estim) | p-val |
|---|---|---|---|---|
| **treat.1** | −0.1469 | 0.8634 | 0.06124 | 0.016 |
| **treat.2** | 0.0128 | 1.013 | 0.1946 | 0.948 |
| **treat.3** | 0.1726 | 1.188 | 0.1106 | 0.119 |
| **treat.4** | −0.0004 | 0.9996 | 0.1582 | 0.998 |
| **treat.5** | 0.3881 | 1.474 | 0.1243 | 0.002 |
| **treat.6** | −0.3143 | 0.7303 | 0.1537 | 0.041 |
| **BigT.Rec** | 0.2100 | 1.234 | 0.06478 | 0.001 |
| **BigT.Prog** | 0.5393 | 1.715 | 0.1244 | $< 0.001$ |
| **ManyT.Rec** | 0.5405 | 1.717 | 0.06609 | $< 0.001$ |
| **ManyT.Prog** | 0.6199 | 1.859 | 0.1298 | $< 0.001$ |
| **RecT.Rec** | 0.3358 | 1.399 | 0.0696 | $< 0.001$ |
| **RecT.RandProg** | 0.7024 | 2.019 | 0.209 | $< 0.001$ |
| **AGEdec.De** | 0.8045 | 2.236 | 0.04294 | $< 0.001$ |
| **AGEdec.RecProg** | 0.2871 | 1.333 | 0.08538 | $< 0.001$ |
| **SEXMale.De** | 0.6164 | 1.852 | 0.09725 | $< 0.001$ |
| $\theta_V$ | $4.192 \times 10^{-10}$ | | | |
| $\theta_W$ | 4.629 | | | |

## B.5 Detailed results for the prostate cancer study

### B.5.1 Separate models for overall survival

#### Models with intermediate PSA progression

Note: treat1, ..., treat3 are the regression coefficients for treatment effect on the three transitions, with transitions numbered as

1 from Rand to PSAp
2 from Rand to De
3 from PSAp to De

**Multi-state model**

| | estim | HR | se(estim) | p-val |
|---|---|---|---|---|
| **treat.1** | −0.21 | 0.81 | 0.09 | 0.023 |
| **treat.2** | −0.47 | 0.63 | 0.16 | 0.003 |
| **treat.3** | −0.36 | 0.70 | 0.11 | 0.002 |
| **PSAp.OS** | −0.16 | 0.86 | 0.13 | 0.245 |

**Stratified multi-state model**

| | estim | HR | se(estim) | p-val |
|---|---|---|---|---|
| **treat.1** | −0.25 | 0.78 | 0.09 | 0.007 |
| **treat.2** | −0.47 | 0.63 | 0.16 | 0.004 |
| **treat.3** | −0.40 | 0.67 | 0.12 | $< 0.001$ |
| **PSAp.OS** | −0.15 | 0.86 | 0.14 | 0.282 |

**Shared frailty (for country) multi-state model**

| | estim | HR | se(estim) | p-val |
|---|---|---|---|---|
| **treat.1** | −0.21 | 0.81 | 0.09 | 0.022 |
| **treat.2** | −0.47 | 0.62 | 0.16 | 0.003 |
| **treat.3** | −0.36 | 0.70 | 0.11 | 0.002 |
| **PSAp.OS** | −0.16 | 0.85 | 0.13 | 0.232 |
| $\theta_V$ | 0.004 | | | |

**Shared frailty (for transition-by-country) multi-state model**

| | estim | HR | se(estim) | p-val |
|---|---|---|---|---|
| **treat.1** | −0.21 | 0.81 | 0.09 | 0.023 |
| **treat.2** | −0.47 | 0.63 | 0.16 | 0.003 |
| **treat.3** | −0.36 | 0.70 | 0.11 | 0.002 |
| **PSAp.OS** | −0.16 | 0.86 | 0.13 | 0.245 |
| $\theta_W$ | $5 \times 10^{-7}$ | | | |

**Nested frailty multi-state model**

| | estim | HR | se(estim) | p-val |
|---|---|---|---|---|
| **treat.1** | −0.23 | 0.79 | 0.09 | 0.011 |
| **treat.2** | −0.46 | 0.63 | 0.16 | 0.004 |
| **treat.3** | −0.35 | 0.70 | 0.11 | 0.002 |
| **PSAp.OS** | −0.17 | 0.85 | 0.13 | 0.211 |
| $\theta_V$ | $5 \times 10^{-7}$ | | | |
| $\theta_W$ | 0.16 | | | |

### Models with intermediate tumour progression

Note: treat1, . . . , treat3 are the regression coefficients for treatment effect on the three transitions, with transitions numbered as

1 from Rand to TUMp

2 from Rand to De

3 from TUMp to De

**Multi-state model**

| | estim | HR | se(estim) | p-val |
|---|---|---|---|---|
| **treat.1** | −0.20 | 0.82 | 0.11 | 0.069 |
| **treat.2** | −0.42 | 0.66 | 0.13 | 0.002 |
| **treat.3** | −0.40 | 0.67 | 0.13 | 0.002 |
| **TUMp.OS** | 0.04 | 1.05 | 0.13 | 0.721 |

**Stratified multi-state model**

| | estim | HR | se(estim) | p-val |
|---|---|---|---|---|
| **treat.1** | −0.18 | 0.83 | 0.11 | 0.099 |
| **treat.2** | −0.42 | 0.66 | 0.14 | 0.003 |
| **treat.3** | −0.36 | 0.70 | 0.14 | 0.009 |
| **TUMp.OS** | 0.16 | 1.18 | 0.14 | 0.231 |

**Shared frailty (for country) multi-state model**

| | estim | HR | se(estim) | p-val |
|---|---|---|---|---|
| **treat.1** | −0.20 | 0.82 | 0.11 | 0.069 |
| **treat.2** | −0.42 | 0.66 | 0.13 | 0.002 |
| **treat.3** | −0.40 | 0.67 | 0.13 | 0.002 |
| **TUMp.OS** | 0.04 | 1.05 | 0.13 | 0.721 |
| $\theta_V$ | $5 \times 10^{-7}$ | | | |

**Shared frailty (for transition-by-country) multi-state model**

| | estim | HR | se(estim) | p-val |
|---|---|---|---|---|
| **treat.1** | −0.20 | 0.82 | 0.11 | 0.069 |
| **treat.2** | −0.42 | 0.66 | 0.13 | 0.002 |
| **treat.3** | −0.40 | 0.67 | 0.13 | 0.002 |
| **TUMp.OS** | 0.05 | 1.05 | 0.13 | 0.711 |
| $\theta_W$ | 0.003 | | | |

**Nested frailty multi-state model**

| | estim | HR | se(estim) | p-val |
|---|---|---|---|---|
| **treat.1** | −0.19 | 0.82 | 0.11 | 0.071 |
| **treat.2** | −0.42 | 0.66 | 0.13 | 0.001 |
| **treat.3** | −0.40 | 0.67 | 0.13 | 0.002 |
| **TUMp.OS** | 0.06 | 1.06 | 0.13 | 0.655 |
| $\theta_V$ | $4 \times 10^{-10}$ | | | |
| $\theta_W$ | 0.02 | | | |

### B.5.2 Model for progression-free survival

Note: treat1, ..., treat3 are the regression coefficients for treatment effect on the three transitions, with transitions numbered as

1 from Rand to PSAp
2 from Rand to De/TUMp
3 from PSAp to De/TUMp

**Multi-state model**

| | estim | HR | se(estim) | p-val |
|---|---|---|---|---|
| **treat.1** | −0.36 | 0.70 | 0.10 | $< 0.001$ |
| **treat.2** | −0.34 | 0.71 | 0.13 | 0.007 |
| **treat.3** | −0.09 | 0.92 | 0.11 | 0.437 |
| **PSAp.PFS** | 0.15 | 1.17 | 0.12 | 0.202 |

**Stratified multi-state model**

| | estim | HR | se(estim) | p-val |
|---|---|---|---|---|
| **treat.1** | −0.42 | 0.66 | 0.10 | $< 0.001$ |
| **treat.2** | −0.30 | 0.74 | 0.13 | 0.019 |
| **treat.3** | −0.13 | 0.87 | 0.12 | 0.256 |
| **PSAp.PFS** | 0.19 | 1.20 | 0.13 | 0.140 |

**Shared frailty (for country) multi-state model**

| | estim | HR | se(estim) | p-val |
|---|---|---|---|---|
| **treat.1** | −0.36 | 0.70 | 0.10 | $< 0.001$ |
| **treat.2** | −0.34 | 0.71 | 0.13 | 0.007 |
| **treat.3** | −0.09 | 0.92 | 0.11 | 0.437 |
| **PSAp.PFS** | 0.15 | 1.17 | 0.12 | 0.202 |
| $\theta_V$ | $2 \times 10^{-5}$ | | | |

**Shared frailty (for transition-by-country) multi-state model**

| | estim | HR | se(estim) | p-val |
|---|---|---|---|---|
| **treat.1** | −0.36 | 0.70 | 0.10 | $< 0.001$ |
| **treat.2** | −0.34 | 0.71 | 0.13 | 0.007 |
| **treat.3** | −0.09 | 0.92 | 0.11 | 0.437 |
| **PSAp.PFS** | 0.15 | 1.17 | 0.12 | 0.202 |
| $\theta_W$ | $5 \times 10^{-7}$ | | | |

**Nested frailty multi-state model**

| | estim | HR | se(estim) | p-val |
|---|---|---|---|---|
| **treat.1** | −0.37 | 0.69 | 0.10 | $< 0.001$ |
| **treat.2** | −0.34 | 0.71 | 0.13 | 0.006 |
| **treat.3** | −0.09 | 0.91 | 0.11 | 0.405 |
| **PSAp.PFS** | 0.16 | 1.17 | 0.12 | 0.187 |
| $\theta_V$ | $3 \times 10^{-10}$ | | | |
| $\theta_W$ | 0.060 | | | |

### B.5.3 Global models for overall survival

Note: treat1, ..., treat5 are the regression coefficients for treatment effect on the five transitions, with transitions numbered as

1 from Rand to PSAp
2 from Rand to TUMp
3 from Rand to De
4 from PSAp to TUMp
5 from PSAp to De
6 from TUMp to De

**Multi-state model**

| | estim | HR | se(estim) | p-val |
|---|---|---|---|---|
| **treat.1** | −0.35 | 0.71 | 0.10 | $< 0.001$ |
| **treat.2** | −0.30 | 0.74 | 0.16 | 0.061 |
| **treat.3** | −0.52 | 0.60 | 0.22 | 0.020 |
| **treat.4** | −0.08 | 0.92 | 0.15 | 0.599 |
| **treat.5** | −0.22 | 0.80 | 0.17 | 0.187 |
| **treat.6** | −0.31 | 0.74 | 0.13 | 0.018 |
| **PSAp.TUMp** | 0.68 | 1.97 | 0.18 | $< 0.001$ |
| **PSAp.OS** | 0.59 | 1.80 | 0.19 | 0.002 |
| **TUMp.OS** | 0.82 | 2.27 | 0.18 | $< 0.001$ |

**Stratified multi-state model**

| | estim | HR | se(estim) | p-val |
|---|---|---|---|---|
| **treat.1** | −0.41 | 0.66 | 0.10 | $< 0.001$ |
| **treat.2** | −0.32 | 0.72 | 0.17 | 0.055 |
| **treat.3** | −0.53 | 0.59 | 0.23 | 0.020 |
| **treat.4** | −0.06 | 0.94 | 0.16 | 0.681 |
| **treat.5** | −0.20 | 0.82 | 0.17 | 0.242 |
| **treat.6** | −0.32 | 0.73 | 0.14 | 0.019 |
| **PSAp.TUMp** | 0.63 | 1.88 | 0.19 | 0.001 |
| **PSAp.OS** | 0.53 | 1.71 | 0.19 | 0.005 |
| **TUMp.OS** | 0.83 | 2.28 | 0.18 | $< 0.001$ |

**Shared frailty (for country) multi-state model**

| | estim | HR | se(estim) | p-val |
|---|---|---|---|---|
| **treat.1** | −0.35 | 0.71 | 0.10 | $< 0.001$ |
| **treat.2** | −0.30 | 0.74 | 0.16 | 0.061 |
| **treat.3** | −0.52 | 0.60 | 0.22 | 0.020 |
| **treat.4** | −0.08 | 0.92 | 0.15 | 0.599 |
| **treat.5** | −0.22 | 0.80 | 0.17 | 0.187 |
| **treat.6** | −0.31 | 0.74 | 0.13 | 0.018 |
| **PSAp.TUMp** | 0.68 | 1.97 | 0.18 | $< 0.001$ |
| **PSAp.OS** | 0.59 | 1.80 | 0.19 | 0.002 |
| **TUMp.OS** | 0.82 | 2.27 | 0.18 | $< 0.001$ |
| $\theta_V$ | $3 \times 10^{-5}$ | | | |

**Shared frailty (for transition-by-country) multi-state model**

| | estim | HR | se(estim) | p-val |
|---|---|---|---|---|
| **treat.1** | −0.35 | 0.71 | 0.10 | $< 0.001$ |
| **treat.2** | −0.30 | 0.74 | 0.16 | 0.061 |
| **treat.3** | −0.52 | 0.60 | 0.22 | 0.020 |
| **treat.4** | −0.08 | 0.92 | 0.15 | 0.599 |
| **treat.5** | −0.22 | 0.80 | 0.17 | 0.187 |
| **treat.6** | −0.31 | 0.74 | 0.13 | 0.018 |
| **PSAp.TUMp** | 0.68 | 1.97 | 0.18 | $< 0.001$ |
| **PSAp.OS** | 0.59 | 1.80 | 0.19 | 0.002 |
| **TUMp.OS** | 0.82 | 2.27 | 0.18 | $< 0.001$ |
| $\theta_W$ | $5 \times 10^{-7}$ | | | |

| Nested frailty multi-state model | | | | |
|---|---|---|---|---|
| | **estim** | **HR** | **se(estim)** | **p-val** |
| **treat.1** | $-0.38$ | 0.68 | 0.10 | $< 0.001$ |
| **treat.2** | $-0.32$ | 0.73 | 0.16 | 0.042 |
| **treat.3** | $-0.51$ | 0.60 | 0.22 | 0.023 |
| **treat.4** | $-0.09$ | 0.91 | 0.15 | 0.533 |
| **treat.5** | $-0.23$ | 0.79 | 0.17 | 0.159 |
| **treat.6** | $-0.32$ | 0.73 | 0.13 | 0.015 |
| **PSAp.TUMp** | 0.70 | 2.01 | 0.18 | $< 0.001$ |
| **PSAp.OS** | 0.61 | 1.83 | 0.19 | 0.001 |
| **TUMp.OS** | 0.82 | 2.27 | 0.18 | $< 0.001$ |
| $\theta_V$ | $5 \times 10^{-7}$ | | | |
| $\theta_W$ | 0.35 | | | |

# Federico Rotolo

CURRICULUM VITAE

## Personal Details

Date of Birth: August 24, 1985
Place of Birth: Genova, Italy
Nationality: Italian

## Contact Information

Via Fereggiano 119/22
16144 Genova. Italy.

Tel. +39 340 80 42 558
e-mail: federico_rotolo@libero.it

## Research interests

- Biostatistics
- Bioinformatics
- Survival analysis
- Clinical trials
- Cancer research

## Education

*October 2007 – July 2009*
**Master (***laurea specialistica***) degree in Statistics**.
University of Padova, Faculty of Statistics
Title of dissertation: "Normalization for single-channel microarray data – A comparative study"
Supervisor: Prof. C. Romualdi
Final mark: 110/110 with honours

*October 2004 – July 2007*
**Bachelor degree (***laurea triennale***) in Maths and Statistics**.
University of Genova, Faculty of Maths, Physics and Natural Sciences
Title of dissertation: "Biostatistician and data manager role in oncological radiotherapy"
Supervisor: Prof. F. Grillo Ruggieri, MD
Final mark: 110/110 with honours.

*October 2004 – July 2007*
**Bachelor degree (***licence professionnelle***) in Statistics and Computer Science**.
Université Nice Sophia Antipolis, Institut Universitiare de Techonologie de Menton
Final mark: Bien 15.35/20.

## Visiting periods

*January 2011 – October 2012*
Institut de Statistique, Biostatistique et Sciences Actuarielles,
Université catholique de Louvain,
Louvain-la-Neuve, Belgium.
Supervisor: Prof. C. Legrand

## Work experience

*April 2007 – July 2007* and *December 2012*
**G. Galliera Hospital**, Genova (Italy). Department of Radiotherapy
Intern.

*Since January 2013*
**G. Roussy Cancer Institute**, Villejuif (France).
Biostatistician.

## Awards and Scholarship

*2010 – 2012*
Italian Ministry of University and Scientific Research: Ph. D. scholarship.

## Computer skills

- R: excellent knowledge
- LaTeX: excellent knowledge
- SAS: Certified Base Programmer
- Sweave
- Unix
- SQL
- SPSS

## Language skills

English: fluent French: fluent Italian: native

## Publications

### Articles in journals

Rotolo, F., Legrand, C., and Van Keilegom, I. (2012). Simulation of clustered multi-state survival data based on a copula model. *Computer Methods and Programs in Biomedicine*. DOI: 10.1016/j.cmpb.2012.09.003

Munda, M., Rotolo, F., and Legrand, C. (2012). parfm: Parametric frailty models in R. *Journal of Statistical Software* **51**(11). URL: http://www.jstatsoft.org/v51/i11

### Working papers

Rotolo, F., Rondeau, V., and Legrand, C. (2012). Incorporation of nested frailties into semiparametric multi-state models.

**Conference proceedings**

Rotolo, F. and Legrand, C. (2012). Frailty Multi-State Models based on Maximum Penalized Partial Likelihood, *Proceedings of the 46th Scientific Meeting of the Italian Statistical Society* . URL: `http://meetings.sis-statistica.org/index.php/sm/sm2012/paper/view/1948`

**`R` packages**

Rotolo, F. and Munda, M. (2012). `parfm`: Parametric frailty models.
URL: `http://cran.r-project.org/web/packages/parfm/`.

Rotolo, F. and Horny, G. (2012). `mlfm`: Multilevel frailty models.
URL: `http://r-forge.r-project.org/projects/mlfm/`.

## Conference presentations

Munda, M., Rotolo, F., Legrand, C. (2012). `parfm`: parametric frailty models in `R`. (poster) *Conference on Quantitative methods in statistics, biostatistics and actuarial sciences*, Leicester, United Kingdom, Apr 2–3, 2012
and *Conference on Quantitative methods in statistics, biostatistics and actuarial sciences*, Louvain-la-Neuve, Belgium, May 30–Jun 1, 2012.

Rotolo, F., Legrand, C., Van Keilegom, I. (2011). A copula-based simulation method for clustered multi-state survival data. (poster) *XIX Annual Meeting of the Belgian Statistical Society*, Hasselt, Belgium, Oct 12–14, 2011.

Rotolo, F., Legrand, C. (2012). Frailty multi-state models using maximum penalised partial likelihood estimation. (contributed) *XLVI Scientific Meeting of the Italian Statistical Society*, Rome, Italy, June 19, 2012.

Rotolo, F., Legrand, C. (2012). Incorporation of nested frailties into multi-state models, with an application to event-history analysis for multicenter clinical trials. (contributed) *6th PhD Day of the French-speaking Belgian Universities*, Louvain-la-Neuve, Belgium, September 14, 2012.

## References

**Prof. Monica Chiogna**
University of Padova
Via C. Battisti 241
35121 Padova (Italy)
Phone: +39 049 82 74 183,
e-mail: monica@stat.unipd.it

**Prof. Catherine Legrand**
Université catholique de Louvain
Voie du Roman Pays, 20
1348 Louvain-la-Neuve (Belgium)
Phone: +32 10 478 798,
e-mail: catherine.legrand@uclouvain.be