Biometrical Journal

RESEARCH PAPER

Marginal hazard ratio estimates in joint frailty models for heart failure trials

Gerrit Toenges[1] 🆔 | **Antje Jahn-Eimermacher[1,2]**

[1]Institute of Medical Biostatistics, Epidemiology and Informatics, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany

[2]Department of Mathematics and Natural Sciences, Darmstadt University of Applied Sciences, Darmstadt, Germany

Correspondence
Gerrit Toenges, Institute of Medical Biostatistics, Epidemiology and Informatics, University Medical Center of the Johannes Gutenberg University Mainz, Obere Zahlbacher Str. 69, 55131 Mainz, Germany.
Email: gtoenges@uni-mainz.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: JA 1821/4-1

Abstract

This work is motivated by clinical trials in chronic heart failure disease, where treatment has effects both on morbidity (assessed as recurrent non-fatal hospitalisations) and on mortality (assessed as cardiovascular death, CV death). Recently, a joint frailty proportional hazards model has been proposed for these kind of efficacy outcomes to account for a potential association between the risk rates for hospital admissions and CV death. However, more often clinical trial results are presented by treatment effect estimates that have been derived from marginal proportional hazards models, that is, a Cox model for mortality and an Andersen–Gill model for recurrent hospitalisations. We show how these marginal hazard ratios and their estimates depend on the association between the risk processes, when these are actually linked by shared or dependent frailty terms. First we derive the marginal hazard ratios as a function of time. Then, applying least false parameter theory, we show that the marginal hazard ratio estimate for the hospitalisation rate depends on study duration and on parameters of the underlying joint frailty model. In particular, we identify parameters, for example the treatment effect on mortality, that determine if the marginal hazard ratio estimate for hospitalisations is smaller, equal or larger than the conditional one. How this affects rejection probabilities is further investigated in simulation studies. Our findings can be used to interpret marginal hazard ratio estimates in heart failure trials and are illustrated by the results of the CHARM-Preserved trial (where CHARM is the 'Candesartan in Heart failure Assessment of Reduction in Mortality and morbidity' programme).

KEYWORDS
heart failure trials, joint frailty model, least false parameter, recurrent events, unexplained heterogeneity

1 | INTRODUCTION

Many clinical trials comprise a non-fatal event type that can occur repeatedly, as primary or secondary endpoint. The analysis of such recurrent event data is well established and there is a plenty of statistical models for treatment effect estimation (Cook & Lawless, 2007; Kelly & Lim, 2000). However, in some situations the occurrence of a competing, terminal event must be

considered, which precludes the occurrence of further recurrent events. Such kind of data particularly arise in clinical heart failure trials where treatment effects both on mortality (assessed as cardiovascular death, CV death) and morbidity (assessed as recurrent non-fatal hospitalisations) are to be evaluated (Zannad et al., 2013; Zanolla & Zardini, 2003). One example is the CHARM-Preserved trial (where CHARM is the 'Candesartan in Heart failure Assessment of Reduction in Mortality and morbidity' programme), a large multicentre trial in heart failure disease. Here treatment with the drug Candesartan did only affect the hospitalisation rate, but not mortality (Yusuf et al., 2003).

The CHARM-Preserved trial is only one of many heart failure trials, where a composite of heart failure hospitalisations (HFHs) and CV death was defined as the primary endpoint, with a time to first event analysis as primary analysis (Anker et al., 2016). This procedure is criticised for a long time, as only a part of the data is integrated in the analysis (Anker & McMurray, 2012). In order to capture the whole disease burden, alternative statistical approaches were suggested, for example to simultaneously consider death and hospitalisations as events in Poisson or negative binomial regression (Rogers et al., 2012; Rogers, Jhund, et al., 2014; Rogers, Pocock, et al., 2014). Hence, all available data are incorporated in the treatment effect estimate. However, every approach of aggregating all the available information into a single summary measure for the treatment effect does not differentiate between the treatment effects on mortality and morbidity (Ferreira-González et al., 2007). For this reason statistical models that allow to investigate both the effect on mortality and the effect on the hospitalisation rate, are of scientific interest.

Recently, a joint frailty regression model has been recommended to simultaneously evaluate treatment effects on HFHs and CV death whilst accounting for a potential association between the two processes (Rogers, Yaroshinsky, Pocock, Stokar, & Pogoda, 2016). Here a frailty term acts multiplicatively on both hazard rates, reflecting unexplained heterogeneity and inducing an association between the two event processes. In heart failure trials a positive association in the sense that, even conditional on covariates, a high hospitalisation risk comes along with a high mortality risk, seems plausible. Conditional on frailty, that is, on the subject's level, the joint frailty model implies a proportional hazards assumption for both risk processes (Huang & Liu, 2007; Liu, Wolfe, & Huang, 2004; Rondeau, Commenges, & Joly, 2003; Rondeau, Mathoulin-Pelissier, Jacqmin-Gadda, Brouste, & Soubeyran, 2007).

Even though the joint frailty model seems to reflect cardiovascular disease courses appropriately, in practice the effects on morbidity and mortality are still more often quantified by means of marginal proportional hazards models, that is, by the Cox model (Cox, 1972) for CV death and the Andersen–Gill model (Andersen & Gill, 1982) for the hospitalisation rate, respectively. Marginally (i.e. on the population's level), the proportional hazards assumption will be violated in the presence of a joint frailty term and treatment effect estimates relying on marginal proportional hazards models are expected to differ from those of conditional proportional hazards models (Aalen, Cook, & Røysland, 2015). Although results derived from joint frailty and marginal models have been contrasted for single case studies (Rogers, Jhund, et al., 2014; Rogers, Pocock, et al., 2014; Rogers, Yaroshinsky, et al., 2016), there exists no systematic investigation of the differences in effect estimates. However, this is strongly needed not only for a proper interpretation of trial results but also for a careful planning of a clinical trial as power depends on effect size. For this reason, we investigate analytically the properties of marginal model estimates in the situation where the data-generating process actually corresponds to a joint frailty model.

In our derivations we use, that a maximum partial likelihood estimator (MPLE) in a misspecified proportional hazards model converges with sample size to a least false parameter, that is, a weighted average of the marginal hazard ratios over time (Struthers & Kalbfleisch, 1986). This result has already been used to investigate the consequences of non-proportional event-specific hazards in multistate models (Grambauer, Schumacher, & Beyersmann, 2010) and of covariate omission (Bretagnolle & Huber-Carol, 1988; Schmoor & Schumacher, 1997) or functional misspecification of covariates (Gerds & Schumacher, 2001) in the Cox model. Furthermore, Henderson and Oman (1999) and Cécilia-Joseph, Auvert, Broët, and Moreau (2015) used Struther's results to study the effects of non-consideration of unobserved heterogeneity in univariate survival analysis. However, the theory of least false parameters has not yet been applied to investigate the consequences of an association between the risk rates for recurrent and terminal events on marginal treatment effect estimates and rejection probabilities for the recurrent event outcome. This paper fills that gap: our findings explain the relation between marginal and joint frailty estimates, which is a relevant and as yet unanswered question in heart failure trials. Thus, they can contribute to the ongoing debate on defining an appropriate statistical model for recurrent hospitalisations and CV death (Anker et al., 2016).

The paper is organised as follows. In Section 2, we will introduce to the general framework of modelling recurrent events in the presence of a terminating event. The marginal hazard ratios as a function of time are derived in Section 3. In Section 4, we analytically derive results on the asymptotic marginal treatment effect estimates. In Section 5, these findings are related to clinical trials, for example the CHARM-Preserved trial in chronic heart failure. Simulation studies for the investigation of finite sample properties of the marginal analysis including rejection probabilities are given in Section 6. Finally, in Section 7 we conclude with a discussion where clinical trial results are interpreted in the light of our new results.

2 | MODELS AND NOTATIONS

Let D be the terminal event (death) time, M be the number of recurrent events (hospitalisations) within $[0, D]$, $T_1 < T_2 < \cdots < T_M \leq T_0 = \infty$ the corresponding recurrent event times and C a censoring time. We are interested in the bivariate counting process $N(t) = (N_1(t), N_2(t))'$ with elements

$$N_1(t) = \sum_{k=0}^{M} I(T_k \leq t) \quad \text{and} \quad N_2(t) = I(D \leq t) \tag{1}$$

that are counting the events of the respective outcome over time. Let t^- denote a time that is infinitesimally smaller than t. For $\Delta > 0$, the increment $\Delta N_i(t) = (N_i(t + \Delta)^- - N_i(t^-))$ is the number of events occurring in the time interval $[t, t + \Delta)$ in the respective process ($i = 1, 2$). Let further X denote a p-dimensional vector representing known covariates and let $H(t) = \{X, N_1(t), N_2(t), 0 \leq s < t\}$ represent the observed process history up to time t.

The hazards of the two counting process components are defined as the instantaneous rates of occurrence for a recurrent and for a terminal event, respectively, among individuals still alive and conditional on the process history:

$$r_i(t|H(t)) = \lim_{\Delta \searrow 0} \frac{P\big(\Delta N_i(t) = 1 \mid H(t), D \geq t\big)}{\Delta} \quad (i = 1, 2). \tag{2}$$

Let $Y(t) = I(D \geq t)$ be the at-risk-indicator. Then $Y(t) r_i(t|H(t))$ ($i = 1, 2$) is the intensity of the respective counting process component.

The processes N_1 and N_2 are by definition dependent, as no recurrent events occur after death. But also beyond that, there might be an association between the event processes. To define that formally, consider the rate of the recurrent event process given the process history and the terminal event time $D = s$:

$$r_1^*(t|H(t), D = s) = \lim_{\Delta \searrow 0} \frac{P\big(\Delta N_1(t) = 1 \mid H(t), D = s\big)}{\Delta}. \tag{3}$$

Then $r_1^*(t|H(t), D = s) = 0$ for $t \geq s$. We consider the two processes as 'associated', if the rate $r_1^*(t|H(t), D = s)$ does, for $t < s$, depend on s. This means that a subject's time of death is informative for its prior recurrent event rate.

The use of frailty variables, which represent risk factors that are not covered by the known covariates, has been proposed for modelling such an association of the event processes (Huang & Liu, 2007; Liu et al., 2004; Rondeau et al., 2003, 2007). We will first introduce a correlated frailty model, which is very flexible regarding the form of association. Thereafter we will consider a joint frailty model, which is more restrictive with respect to the kind of dependency between the processes.

Instead of modelling the hazards $r_i(t|H(t))$ directly, frailty models are modelling conditional hazards. In general, we may think of positive random variables Z and W, without making any assumptions on their joint distribution. The frailties Z and W are assumed to be independent of the known covariates X. The conditional hazards are defined as follows:

$$\lambda_1(t|H(t), Z) = \lim_{\Delta \searrow 0} \frac{P\big(\Delta N_1(t) = 1 \mid Z, H(t), D \geq t\big)}{\Delta}$$

$$\lambda_2(t|H(t), W) = \lim_{\Delta \searrow 0} \frac{P\big(\Delta N_2(t) = 1 \mid W, H(t), D \geq t\big)}{\Delta}. \tag{4}$$

The conditional hazards within the correlated frailty model are modelled as

$$\lambda_1(t|H(t), Z) = \lambda_1(t|X, Z) = Z \lambda_{10}(t) \exp(\beta_1' X)$$

$$\lambda_2(t|H(t), W) = \lambda_2(t|X, W) = W \lambda_{20}(t) \exp(\beta_2' X). \tag{5}$$

In the correlated frailty model $\lambda_{10}(t)$, $\lambda_{20}(t)$ and β_1, β_2 are unspecified baseline hazard rates and regression coefficient vectors for the two event processes. The model contains a proportional hazards assumption on the subject's level (i.e. for the conditional hazards) and the censoring time C is assumed to be independent. The association between the two event processes is solely captured by the dependency between the two frailty variables – meaning that the process history (except for covariates) is excluded from the conditional hazards and that $N_1(t|X, Z)$ is a Poisson process. The processes N_1 and N_2 are associated on the marginal level, as shown in Appendix A.1.

Without pre-specification of the joint distribution of the frailties Z and W, model (5) is not identifiable. For this reason, further assumptions, for example $W = Z$ or $W = Z^\alpha$ with Z having a particular distribution were discussed (Huang & Liu, 2007; Liu et al., 2004; Rondeau et al., 2003, 2007). Conveniently, a gamma or a log-normal distribution with mean $E(Z) = 1$ and variance $Var(Z) = \theta$ is used for that purpose.

Below we will mainly focus on the case $W = Z^\alpha$. In line with other authors (see Huang & Liu, 2007; Liu et al., 2004; Rondeau et al., 2003, 2007) we will refer to that model as joint frailty model:

$$\lambda_1(t|X, Z) = Z \lambda_{10}(t) \exp(\beta_1' X)$$

$$\lambda_2(t|X, Z) = Z^\alpha \lambda_{20}(t) \exp(\beta_2' X). \tag{6}$$

Both the frailty variance θ and the α-parameter determine the degree of association between the two event processes. The frailty variance is additionally a measure for the correlation between subsequent recurrent event times.

3 | MARGINAL HAZARDS IN THE CORRELATED FRAILTY MODEL

As mentioned before, the hazard functions $r_1(t|H(t))$ and $r_2(t|H(t))$, which are defined in (2), are not modelled explicitly within the correlated frailty model. However, for a particular distribution of Z and W, they are implicitly given by

$$r_1(t|H(t)) = r_1(t|X) = E_Z\left(\lambda_1(t|X, Z)\right)$$

$$r_2(t|H(t)) = r_2(t|X) = E_W\left(\lambda_2(t|X, W)\right). \tag{7}$$

As we introduced $\lambda_1(t|X, Z)$ and $\lambda_2(t|X, W)$ as the conditional hazards, we will call $r_1(t|X)$ and $r_2(t|X)$ the marginal hazards from now on. If the two event processes are associated according to the correlated frailty model (5), the marginal hazards are in general not proportional anymore, as shown below for both endpoints. Due to selection effects, a time-constant treatment effect on the subject's level will become time-dependent on the population's (marginal) level. In the following, we will briefly derive the marginal hazards and their ratios as a function of time. Thereby, we will focus on the situation of a two-arm randomised controlled trial (RCT). Hence, we consider X to be a binary $Bin(1, p)$ distributed covariate indicating the treatment group. The results will afterwards be used to investigate how regression coefficient estimation is affected, if the estimation relies on (in general misspecified) marginal proportional hazards models.

For the derivation of the marginal hazards we will use the Laplace transform of the distribution of a random variable U, which is defined as

$$\mathcal{L}_U(s) = E(\exp(-sU)) = \int \exp(-su) f_U(u) \, du \tag{8}$$

with f_U being the probability density function of U. The first derivative of $\mathcal{L}_U$ is given by

$$\mathcal{L}_U'(s) = \frac{d}{ds} \int \exp(-su) f_U(u) \, du = \int -u \exp(-su) f_U(u) \, du. \tag{9}$$

If the two risk processes are associated according to the correlated frailty model (5), the marginal hazards are given by

$$r_1(t|X) = \lambda_{10}(t) \exp(\beta_1 X) \cdot E(Z|X, D \geq t) \tag{10}$$

$$r_2(t|X) = \lambda_{20}(t) \exp(\beta_2 X) \cdot E(W|X, D \geq t) \tag{11}$$

with conditional expectations being dissolved as

$$E(Z|X, D \geq t) = \frac{\int_0^\infty \int_0^\infty z \, \exp\left(-w \exp(\beta_2 X) \Lambda_{20}(t)\right) f_{Z,W}(z, w) \, dw \, dz}{\mathcal{L}_W\left(\exp(\beta_2 X) \Lambda_{20}(t)\right)} \tag{12}$$

$$E(W|X, D \geq t) = -\frac{\mathcal{L}_W'\left(\exp(\beta_2 X) \Lambda_{20}(t)\right)}{\mathcal{L}_W\left(\exp(\beta_2 X) \Lambda_{20}(t)\right)}. \tag{13}$$

Here $\Lambda_{20}(t) = \int_0^t \lambda_{20}(s)\,\mathrm{d}s$ denotes the cumulative baseline hazard for mortality. Details of these derivations are given in Appendix A.2. In the special case of $W = Z$ the numerator in (12) simplifies to $-\mathcal{L}'_W(\exp(\beta_2 X)\Lambda_{20}(t))$ so that (12) and (13) coincide.

Using (10)–(13), the marginal hazard ratios for the two endpoints are given by

$$\frac{r_1(t|X = 1)}{r_1(t|X = 0)} = \exp(\beta_1) \cdot \frac{E(Z|X = 1, D \geq t)}{E(Z|X = 0, D \geq t)}$$

$$= \exp(\beta_1) \tag{14}$$

$$\cdot \frac{\mathcal{L}_W(\Lambda_{20}(t)) \cdot \int_0^\infty \int_0^\infty z \, \exp\left(-w \exp(\beta_2)\Lambda_{20}(t)\right) f_{Z,W}(z, w)\,\mathrm{d}w\,\mathrm{d}z}{\mathcal{L}_W(\exp(\beta_2)\Lambda_{20}(t)) \cdot \int_0^\infty \int_0^\infty z \, \exp\left(-w\Lambda_{20}(t)\right) f_{Z,W}(z, w)\,\mathrm{d}w\,\mathrm{d}z}$$

$$\frac{r_2(t|X = 1)}{r_2(t|X = 0)} = \exp(\beta_2) \cdot \frac{E(W|X = 1, D \geq t)}{E(W|X = 0, D \geq t)}$$

$$= \exp(\beta_2) \tag{15}$$

$$\cdot \frac{\mathcal{L}_W(\Lambda_{20}(t)) \cdot \mathcal{L}'_W(\exp(\beta_2)\Lambda_{20}(t))}{\mathcal{L}_W(\exp(\beta_2)\Lambda_{20}(t)) \cdot \mathcal{L}'_W(\Lambda_{20}(t))}.$$

Both marginal hazard ratios can be dissected into the product of the conditional hazard ratio and a deviation factor. Again, in the special case of $W = Z$, the deviation factors in the hazard ratios of both endpoints coincide. The deviation factors of both endpoints depend on the frailty distributions, on the covariate-effect β_2 on the terminal event rate and on the cumulative baseline terminal event rate $\Lambda_{20}(t)$. As the latter is a function of time, the marginal hazard ratios are time-dependent in contrast to the time-constant conditional hazard ratios,

$$\frac{\lambda_1(t|X = 1, Z)}{\lambda_1(t|X = 0, Z)} = \exp(\beta_1) \quad \text{and} \quad \frac{\lambda_2(t|X = 1, W)}{\lambda_2(t|X = 0, W)} = \exp(\beta_2). \tag{16}$$

It is worth mentioning that the deviation between the marginal and the conditional hazard ratio over time is independent of the regression coefficient β_1 and the cumulative baseline hazard $\Lambda_{10}(t)$ that describe the recurrent event process. This holds for both endpoints (recurrent and terminal).

4 | ASYMPTOTICS OF MARGINAL HAZARD RATIO ESTIMATES

We will now derive asymptotic characteristics of marginal parameter estimates (log-hazard-ratio estimates) that erroneously rely on marginal proportional hazards models, when in fact the proportional hazards assumption holds on the conditional (subject's) level. Thus, instead of model (5), an Andersen–Gill model for the recurrent events and a Cox model for the terminal event is assumed. Both do not include the frailty variables Z and W. As before, we consider a two-arm RCT setting with X being binary. Under the Andersen–Gill and Cox modelling assumptions, regression coefficients are estimated by maximum partial likelihood estimators (MPLEs) (Andersen & Gill, 1982; Cox, 1972).

Let $T_{i1} < T_{i2} < \cdots < T_{iM_i} \leq T_{i0} = \infty$ denote the recurrent event times, C_i the right-censoring time, D_i the time of death and X_i the binary covariate for the ith of n subjects. The MPLEs relying on the marginal models are given by

$$\widehat{\beta}_1 = \underset{\beta \in \mathbb{R}}{\mathrm{argmax}}\, L_1(\beta) \quad \text{with} \quad L_1(\beta) = \prod_{i=1}^n \prod_{j\,:\,T_{ij} < C_i} \frac{\exp(\beta X_i)}{\sum_{k \in R(T_{ij})} \exp(\beta X_k)} \tag{17}$$

$$\widehat{\beta}_2 = \underset{\beta \in \mathbb{R}}{\mathrm{argmax}}\, L_2(\beta) \quad \text{with} \quad L_2(\beta) = \prod_{i\,:\,D_i < C_i} \frac{\exp(\beta X_i)}{\sum_{k \in R(D_i)} \exp(\beta X_k)}. \tag{18}$$

$R(t)$ is the set of subjects being at risk at time t, that is, $R(t) = \{k \mid (D_k \wedge C_k) \geq t\}$. In the Andersen–Gill estimator $\widehat{\beta}_1$, death is handled as independent censoring.

As shown by Struthers and Kalbfleisch (1986), the MPLE's $\widehat{\beta}_1$ and $\widehat{\beta}_2$ converge (in probability) to some least false parameters β_1^* and β_2^*, respectively. By applying Struther and Kalbfleisch's general results to our situation, the least false parameter β_i^* ($i = 1, 2$) can be identified as being the unique solution of the parameter integral equation $g_i(s) = 0$ with

$$g_i(s) = \int_0^{t_{max}} \frac{\bar{y}^{(0)}(t)\bar{y}^{(1)}(t)}{\bar{y}^{(0)}(t) + \bar{y}^{(1)}(t)\exp(s)} \left[r_i(t|X=1) - \exp(s)r_i(t|X=0) \right] dt . \tag{19}$$

Here t_{max} denotes the maximum of observation times. In addition, $\bar{y}^{(k)}(t)$ denotes the probability that a randomly selected subject belongs to treatment group $k \in \{0, 1\}$ and is still at risk at time t, that is,

$$\bar{y}^{(k)}(t) = P(X = k)S(t|X = k)P(C > t) . \tag{20}$$

In the following, we apply (19) to investigate the asymptotic behaviour of marginal estimates for various settings. Thereby we will focus on the Andersen–Gill estimator $\widehat{\beta}_1$, as the properties of the Cox estimator $\widehat{\beta}_2$ in the presence of unexplained heterogeneity are already well investigated by other authors (see Cécilia-Joseph et al., 2015; Henderson & Oman, 1999; Schmoor & Schumacher, 1997).

4.1 | No treatment effect on mortality ($\beta_2 = 0$)

First of all, we will make some weak assumptions regarding censoring and baseline hazards: we will assume that the baseline hazards $\lambda_{10}(t)$ and $\lambda_{20}(t)$ are continuous functions which implies that the marginal hazard $r_1(t|X = k)$ and the marginal survival function $S(t|X = k)$ are likewise continuous. By further assuming censoring to be continuous on $(0, t_{max})$, we obtain $\bar{y}^{(k)}(t)$ to be continuous on $(0, t_{max})$. After Cauchy's mean value theorem there exists a $t^* \in (0, t_{max})$ so that

$$0 = \left[r_1(t^*|X=1) - \exp(\beta_1^*)r_1(t^*|X=0) \right] \int_0^{t_{max}} \frac{\bar{y}^{(0)}(t)\bar{y}^{(1)}(t)}{\bar{y}^{(0)}(t) + \bar{y}^{(1)}(t)\exp(\beta_1^*)} dt . \tag{21}$$

As the integral is strictly positive, we obtain

$$\beta_1^* = \log\left(\frac{r_1(t^*|X=1)}{r_1(t^*|X=0)} \right) . \tag{22}$$

So the least false parameter β_1^* is in the value range of the logarithmised marginal hazard ratio over the interval $(0, t_{max})$.

If the treatment has no effect on mortality, that is, $\beta_2 = 0$, the marginal hazard ratio is constant over time and coincides with the conditional hazard ratio $\exp(\beta_1)$ as is evident from (14). Applying (22) shows that the Andersen–Gill estimator asymptotically coincides with the conditional treatment effect for recurrent events, that is,

$$\beta_1^* \overset{(22)}{=} \log\left(\frac{r_1(t^*|X=1)}{r_1(t^*|X=0)} \right) \overset{\beta_2=0}{=} \beta_1 \tag{23}$$

as the latter equation is true for all t according to (14). This is illustrated in Figures 2–4 and in Table 3 for various settings in a joint gamma frailty model. But the result generally holds in every kind of correlated frailty model. Even though the marginal hazards $r_1(t|X = 1)$ and $r_1(t|X = 0)$ do not correspond to the their conditional counterparts $\lambda_1(t|X = 1, Z)$ and $\lambda_1(t|X = 0, Z)$, their respective ratios are equal. If treatment has no effect on mortality, the selection due to unexplained heterogeneity affects the treatment and the control group in the same way, resulting in a still proportional time course of the marginal hazards. Hence this is in fact the only situation, where the proportional hazards assumption is not violated on the marginal level. However, it should be underlined, that the processes are associated even if $\beta_2 = 0$.

4.2 | Joint frailty model

The least false parameter β_1^* can be derived by numerically solving $g_1(s) = 0$ (see formula (19)) after specifying the correlated frailty model. To characterise the marginal treatment effect estimate for recurrent events more in detail, we will consider the
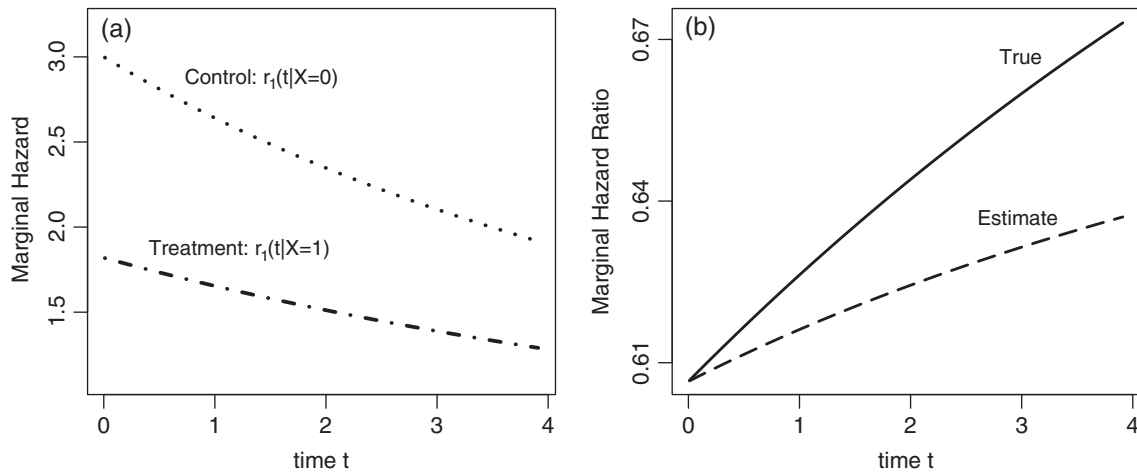
FIGURE 1 Marginal hazards and hazard ratios in a joint gamma frailty model with frailty variance $\theta = 1$, association parameter $\alpha = 0.8$, constant subject-specific hazards (Weibull scale parameters $\lambda_1 = 3$ and $\lambda_2 = 0.178$, Weibull shape parameters $\nu_1 = 1$ and $\nu_2 = 1$) and treatment being protective both for hospitalisations and for mortality ($\exp(\beta_1) = \exp(-0.5) = 0.607$, $\exp(\beta_2) = \exp(-0.3) = 0.741$). (a) Marginal hazards for recurrent events in the treatment and control group. (b) True: marginal hazard ratio for recurrent events. Estimate: hazard ratio estimate $\exp(\beta_1^*)$ resulting from a marginal Andersen–Gill analysis after administrative censoring at time point t

joint frailty model (i.e. $W = Z^\alpha$, see formula (6)) from now on. We will show results for a gamma-distributed frailty Z with mean $E(Z) = 1$ and variance $Var(Z) = \theta$. However, subsequent results do qualitatively not change if a log-normal distribution is adopted (results not shown). Furthermore, we will consider a 1:1 allocation ratio and administrative censoring after t_{max} time units. The subject-specific hazards are assumed to originate from Weibull distributions. Thereby the conditional baseline hazard for the ith endpoint (1 = recurrent, 2 = terminal) is given by $\lambda_{i0}(t) = \lambda_i \nu_i t^{\nu_i - 1}$, with λ_i being the scale parameter and ν_i being the shape parameter. The latter determines, if the hazard is decreasing ($\nu_i < 1$), constant ($\nu_i = 1$) or increasing ($\nu_i > 1$) over time.

Figure 1a shows the time course of the marginal recurrent event hazards for the situation where both processes are positively associated ($\alpha = 0.8$), meaning that patients with a high mortality risk also have a high risk for recurrent events and vice versa. We further consider that treatment is protective with regard to both endpoints ($\beta_1 < 0$, $\beta_2 < 0$) and that the conditional baseline hazards for both endpoints are constant over time. At the time point of randomisation ($t = 0$), treatment and control group do not differ with regard to the distribution of the frailty variables. Due to dropouts of frail patients, the marginal hazard is decreasing over time in both groups. But as selection effects differ between the groups ($\beta_2 < 0$), that decrease is non-proportional. Hence, the marginal hazard ratio corresponds to the conditional one only at time point 0 and is afterwards increasing over time, as shown as 'True' in Figure 1b. It is further demonstrated that the estimated marginal hazard ratio $\exp(\beta_1^*)$ at time point $t_{max} = t$ is a weighted average of the marginal hazard ratios over the interval $(0, t)$, shown as 'Estimate' in Figure 1b. For this reason, the marginal hazard ratio estimate depends on the length of follow-up.

In the following we focus on the evaluation of the estimate β_1^* after a fixed follow-up period of $t_{max} = 2$ time units. In each of Figures 2–4, the difference $(\beta_1^* - \beta_1)$ between the marginal treatment effect estimate (after two time units of follow-up) and the conditional treatment effect is shown in dependence of various parameters of the joint frailty model.

First, we will focus on situations where β_1^* and β_1 coincide: this is the case if either $\beta_2 = 0$ (Figures 2b, 3 and 4) or $\alpha = 0$ (Figure 4b) or $\theta = 0$ (Figures 3 and 4). It has already been shown analytically in Subsection 4.1 that $\beta_1^* = \beta_1$, if treatment has no effect on mortality ($\beta_2 = 0$). If $\alpha = 0$, the joint frailty model reduces to a model with inter-individual heterogeneity in recurrent events but without an association between recurrent and terminal event rates. In this situation, the Andersen–Gill estimator is consistent for the regression parameter (i.e. $\beta_1^* = \beta_1$), as shown by other authors (Lin, Wei, Yang, & Ying, 2000; Liu, 2014). There is likewise no association if $\theta = 0$ because then the submodel for recurrent events reduces to an Andersen–Gill model and of course β_1 is estimated consistently by the Andersen–Gill estimator in that situation.

Next, we will take a closer look at parameters that do not affect the difference $(\beta_1^* - \beta_1)$: the scale parameter λ_1 of the conditional recurrent event hazard, which mainly determines how many recurrent events are expected within the trial time, has this property. This can easily be derived analytically, as λ_1 can be extracted as a factor in (19) and thus does not affect the solution of $g_1(s) = 0$. The second parameter not influencing the difference $(\beta_1^* - \beta_1)$ is β_1 itself, as illustrated in Figure 2 (however, an analytical proof for that result is lacking). Accordingly, results in Figures 3 and 4 do not depend on β_1, which is for that reason no longer specified there.

$\cdots v_1 = 0.5, v_2 = 2$ $- - v_1 = 1, v_2 = 1$ $\cdot - \cdot v_1 = 2, v_2 = 0.5$
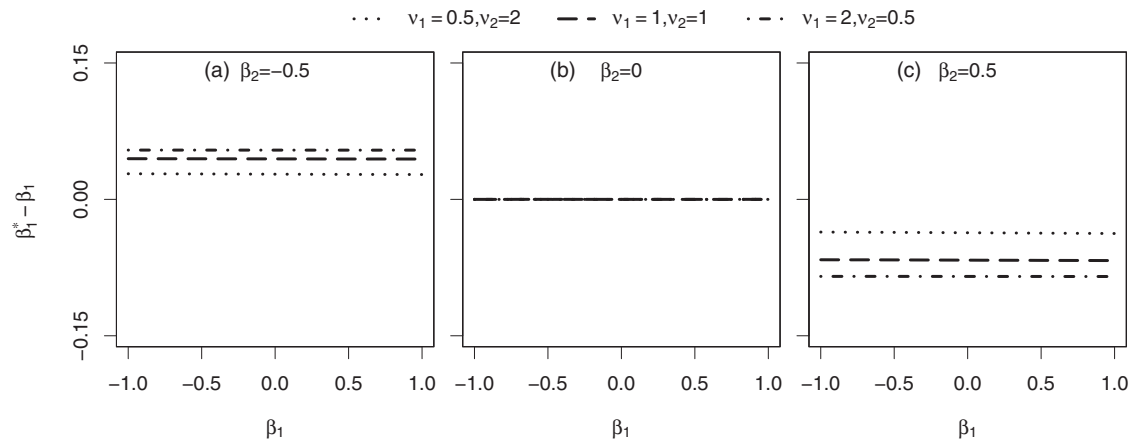
FIGURE 2   Asymptotic difference between the marginal treatment effect estimate β_1^* and the conditional treatment effect β_1 in a joint gamma frailty model with frailty variance $\theta = 1$ and association parameter $\alpha = 0.8$. *Note*: Subject-specific hazards originate from Weibull distributions (scale parameters $\lambda_1 = 3$, λ_2 and shape parameters v_1, v_2). The scale-parameter λ_2 is not shown here, but was in each scenario selected in such a way that the conditional survival probability at the end of follow-up is $S(2|X = 0, Z = 1) = 0.7$. Subjects are censored administratively after $t_{max} = 2$ time units

$\cdots \theta = 1$ $- - \theta = 0.5$ $\cdot - \cdot \theta = 0$
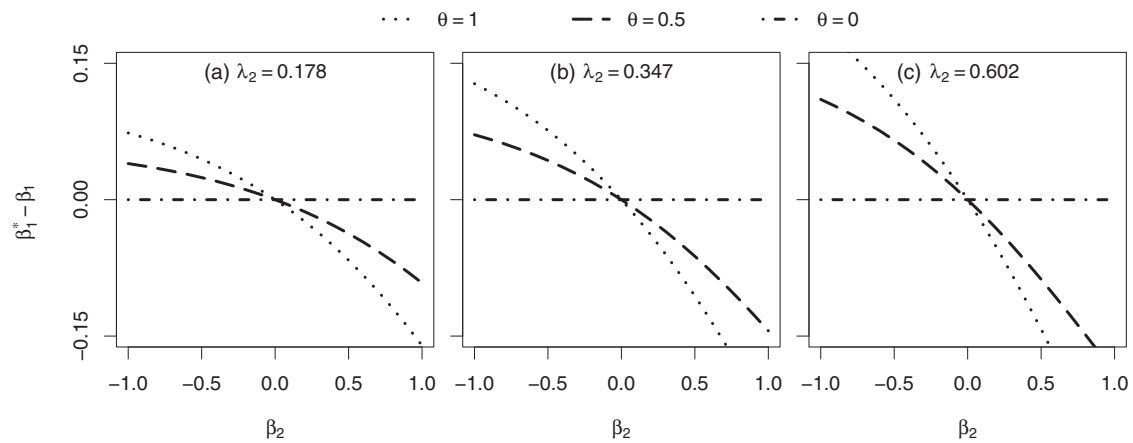
FIGURE 3   Asymptotic difference between the marginal treatment effect estimate β_1^* and the conditional treatment effect β_1 in a joint gamma frailty model with frailty variance θ and association parameter $\alpha = 0.8$. *Note*: Subject-specific hazards are constant (Weibull scale parameters $\lambda_1 = 3$ and λ_2, Weibull shape parameters $v_1 = 1$ and $v_2 = 1$), resulting in a conditional survival probability of $S(2|X = 0, Z = 1) = 0.7$ in (a), 0.5 in (b) and 0.3 in (c) at the end of follow-up. Subjects are censored administratively after $t_{max} = 2$ time units

$\cdots \theta = 1$ $- - \theta = 0.5$ $\cdot - \cdot \theta = 0$
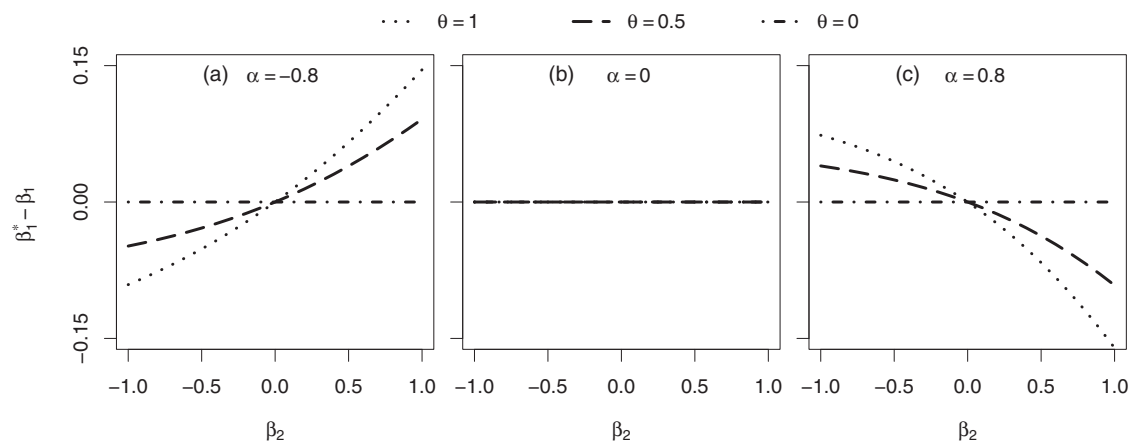
FIGURE 4   Asymptotic difference between the marginal treatment effect estimate β_1^* and the conditional treatment effect β_1 in a joint gamma frailty model with frailty variance θ and association parameter α. *Note*: Subject-specific hazards are constant (Weibull scale parameters $\lambda_1 = 3$ and $\lambda_2 = 0.178$, Weibull shape parameters $v_1 = 1$ and $v_2 = 1$), resulting in a conditional survival probability of $S(2|X = 0, Z = 1) = 0.7$ at the end of follow-up. Subjects are censored administratively after $t_{max} = 2$ time units

TABLE 1 Event numbers and (unadjusted) treatment effect estimates of the CHARM-Preserved trial according to Rogers, Pocock, et al. (2014) and Yusuf et al. (2003)

	Placebo	Candesartan	Hazard ratio (95% CI)	
			Marginal	Conditional
Number of patients	1,509	1,514		
Total follow-up years	4,374.03	4,424.62		
Number of CV deaths	170	170	0.99 (0.80, 1.22)	0.96 (0.73, 1.26)
Total number of heart failure hospitalisations	547	392	0.71 (0.57, 0.88)	0.69 (0.55, 0.85)

Marginal treatment effect estimates rely on the Cox model (CV death) or the Andersen–Gill model (heart failure hospitalisations). Conditional treatment effect estimates rely on a joint gamma frailty model.

Now we focus on parameters that determine the absolute difference $|\beta_1^* - \beta_1|$ between the marginal treatment effect estimate and the conditional treatment effect. An increase in the frailty variance θ (Figures 3 and 4) and an increase in the absolute value of the association parameter $|\alpha|$ (Figure 4) result in an increase of that absolute difference. In addition, increasing the mortality risk during follow-up leads to the same result (Figure 3). As illustrated in Figure 2, the absolute difference further depends on the ratio v_1/v_2 between the shapes of the baseline hazards. The larger the v_1/v_2 ratio, the less recurrent events can occur before death and the more the marginal treatment effect estimate deviates from the conditional treatment effect.

In particular, the treatment effect on mortality β_2 and the association parameter α determine, if the marginal treatment effect estimate β_1^* is smaller, equal or larger than the conditional treatment effect β_1, that is, they specify the sign of the difference $(\beta_1^* - \beta_1)$. This becomes apparent in Figure 4: if $\alpha > 0$, the difference is positive in case of $\beta_2 < 0$ and negative in case of $\beta_2 > 0$ (Figure 4c). For the special situation of $\alpha = 1$, an analytical proof for that finding is given in Appendix A.3. The effects turn around for $\alpha < 0$, that is, the difference is negative in case of $\beta_2 < 0$ and positive in case of $\beta_2 > 0$ (Figure 4a). The latter scenario is however less reasonable for heart failure trials. So let us have a look at particular situations that might be most realistic in heart failure trials: here a positive association ($\alpha > 0$) seems reasonable because patients with a high hospitalisation rate will probably die earlier than patients who are rarely in need for hospitalisation. In addition, most heart failure drugs either have a protective or negligible effect on mortality ($\beta_2 \leq 0$). So in these settings, the marginal treatment effect estimate is larger than the conditional treatment effect ($\beta_1^* \geq \beta_1$). Hence, a conditional (joint frailty) analysis of the data will result in an estimate that favours treatment more compared to a marginal (Andersen–Gill) analysis. To illustrate that, we will consider the joint frailty setting that is shown in Figure 4c ($\alpha = 0.8$, $\theta = 1$) in greater detail: if $\beta_2 = -0.5$, the difference of interest is given by $(\beta_1^* - \beta_1) = 0.045$. Hence in case of $\beta_1 = -0.3$ the marginal estimate is $\beta_1^* = -0.3 + 0.045 = -0.255$. As already emphasised before, the parameter β_1 itself does not affect the deviation of β_1^* from β_1. Hence in case of $\beta_1 = -0.02$, the marginal estimate is given by $\beta_1^* = -0.98 + 0.045 = 0.025$. This example shows that there may exist situations, in which even the direction does not match between the marginal estimate and the conditional treatment effect. Generally said, it may happen that $\text{sign}(\beta_1^*) \neq \text{sign}(\beta_1)$, corresponding to a misjudgement of the treatment effect direction for recurrent events.

5 | APPLICATIONS

The 'Candesartan in Heart failure Assessment of Reduction in Mortality and morbidity' (CHARM) programme comprised three independent, randomised clinical trials that evaluated the benefit of Candesartan as a supplementary therapy in patients with chronic heart failure. The CHARM-Preserved trial focussed on heart failure patients with preserved ejection fraction (defined as >40% left ventricular ejection fraction). A total of 3,023 patients met the inclusion criteria and were randomised to receive either Candesartan or Placebo in addition to the recommended standard therapy. The median follow-up was 36.6 months. Regarding the primary endpoint of the trial, a composite of CV death and HFHs that was evaluated by a time to first event analysis, no treatment benefit could be shown (Rogers, Pocock, et al., 2014; Yusuf et al., 2003).

Table 1 shows results on CHARM-Preserved that were reported in Rogers, Pocock, et al. (2014) and Yusuf et al. (2003). As apparent, CV death did not differ between the two groups. But patients in the Candesartan group had fewer HFHs compared to the Placebo group, which is reflected by a HFH hazard ratio of 0.71 in a marginal analysis (Andersen–Gill model) and of 0.69 in a conditional analysis with the joint gamma frailty model. We used our findings from Section 4 to derive the expected marginal hazard ratio estimates, thereby varying the frailty variance, and compare these with the observed estimates. For that, formula (19) is applied with parameters, which match the situation in CHARM-Preserved (see Table 2). Thereby we assumed an association parameter of $\alpha = 0.8$, as published joint frailty results from other heart failure trials (CHARM-Added,

TABLE 2 Expected marginal hazard ratio estimates in joint gamma frailty models that reflect the situation of the CHARM-Preserved trial (using constant baseline hazards with rates that match the observed event numbers and administrative censoring after $t_{max} = 3.05$ years follow-up)

| Parameters joint gamma frailty model						Expected marginal model estimates	
θ	α	$\exp(\beta_1)$	$\exp(\beta_2)$	$\lambda_{10}(t)$	$\lambda_{20}(t)$	$\exp(\beta_1^*)$	$\exp(\beta_2^*)$
0	0.8	0.69	0.96	0.125	0.039	0.6900	0.9600
1	0.8	0.69	0.96	0.125	0.039	0.6912	0.9613
2	0.8	0.69	0.96	0.125	0.039	0.6921	0.9624
3	0.8	0.69	0.96	0.125	0.039	0.6929	0.9633
4	0.8	0.69	0.96	0.125	0.039	0.6936	0.9641
5	0.8	0.69	0.96	0.125	0.039	0.6942	0.9648

CHARM-Alternative) suggest that heterogeneity might be greater with respect to HFHs than for mortality (Rogers et al., 2016). We can conclude that our results can well explain that there are only slight differences between marginal and conditional estimates in CHARM-Preserved: in the presence of a negligible treatment effect on the terminal event (CV death), the hazard ratio estimates for recurrent events (HFHs) from marginal and conditional models nearly coincide (see $\beta_2 = 0$ from Subsection 4.1).

As in CHARM-Preserved, also in many other heart failure trials the randomised treatment groups differ only slightly in their risk for CV death (Zheng et al., 2018). Therefore, marginal and conditional hazard ratio estimates are supposed to coincide in these trials. HFHs and CV death are also the outcomes of interest in type 2 diabetes studies. Opposite to heart failure trials, substantial treatment effects for CV death were shown here (Schnell, Rydén, Standl, & Ceriello, on behalf of the D&CVD EASD Study Group, 2017). For example, we consider the EMPA-REG trial (Zinman et al., 2015) that reported marginal hazard ratios of 0.65 for HFHs and 0.62 for CV death with a median observation time of 3.1 years. Note that the reported marginal HFH hazard ratio estimate is derived from a Cox analysis (time to first HFH) only and may therefore not exactly match to the (Andersen–Gill) estimate investigated in this paper. Following our findings, in this trial the conditional hazard ratio estimate for HFHs ought to be smaller than the marginal one, that is, $\exp(\beta_1) < \exp(\beta_1^*) = 0.65$, if the rates for HFHs and CV death are positively associated ($\alpha > 0$, $\theta > 0$, see Subsection 4.2). However, as conditional estimates are not published, an empirical verification cannot be provided at this point.

6 | SIMULATION STUDY

As differences between marginal and conditional hazard ratio estimates may also affect the rejection probability of the corresponding two-sided statistical Z-test (based on the Andersen–Gill estimate in combination with its robust standard error) for evaluation of the null hypothesis $H_0 : \beta_1 = 0$, we performed simulations to investigate this effect. Data were simulated according to a joint gamma frailty model with association parameter $\alpha = 0.8$. We consider administrative censoring after $t_{max} = 2$ time units. As before, the conditional baseline hazards for the two processes (1 = recurrent, 2 = terminal) originate from Weibull distributions with scale parameters λ_1, λ_2 and shape parameters ν_1, ν_2. In all scenarios, the cumulative baseline hazard for recurrent events at the end of follow-up is $\Lambda_{10}(2) = 6$, corresponding to six expected recurrent events during follow-up if no terminal event was present. This is achieved both in a setting with a constant baseline hazard ($\lambda_1 = 3$, $\nu_1 = 1$, see Figure 5 and Table 3) and a setting with a decreasing baseline hazard ($\lambda_1 = 4.24$, $\nu_1 = 0.5$, see Table 3) for recurrent events. Furthermore, we always consider a constant baseline mortality hazard ($\lambda_2 = 0.178$, $\nu_2 = 1$), corresponding to a probability of $S(2|X = 0, Z = 1) = 0.7$ to survive the follow-up.

In all scenarios shown in Figure 5 and in half of the scenarios shown in Table 3, no treatment effect on the recurrent event rate is present (i.e. $\beta_1 = 0$). Here the rejection probability of the test corresponds to its type I error probability. The test is exact if $\beta_2 = 0$ and anticonservative if $\beta_2 \neq 0$. In the latter case, the rejection probability under H_0 increases with sample size (Figure 5) because the Andersen–Gill estimator $\hat{\beta}_1$ converges in probability to $\beta_1^* \neq \beta_1 = 0$, while its robust standard error $SE(\hat{\beta}_1)$ converges in probability to 0. Besides β_2, the rejection probability under H_0 is of course also affected by all the additional joint frailty parameters that determine how much β_1^* deviates from $\beta_1 = 0$ (see Subsection 4.2). For example, it increases by increasing the frailty variance θ (Figure 5, Table 3) and by switching from a constant to a decreasing baseline hazard for recurrent events (Table 3). Table 3 further shows scenarios with an existing treatment effect on the recurrent event rate ($\beta_1 = -0.5$). Here the rejection probability of the test corresponds to its power. In these selected scenarios, the null hypothesis is rejected almost for
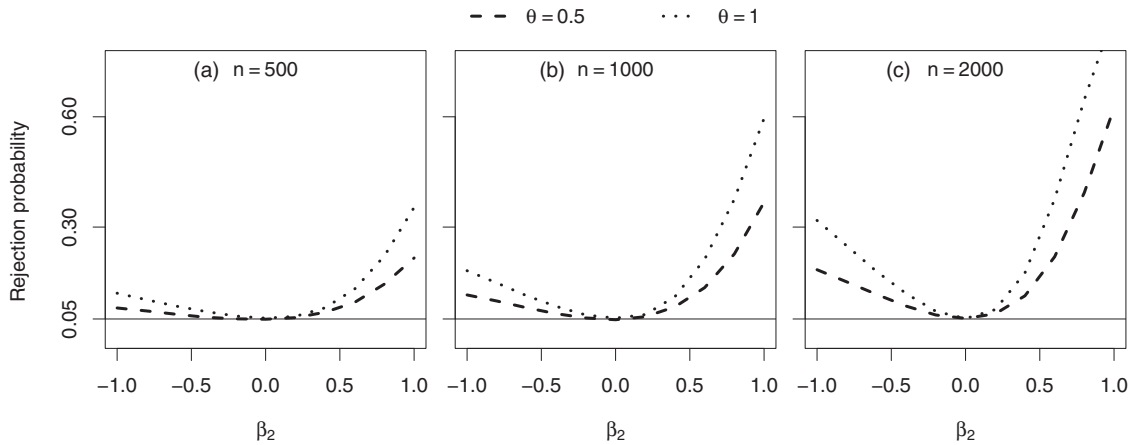
FIGURE 5 Simulation results (10,000 simulated datasets, each with n subjects): rejection probability of a two-sided Z-test (based on the Andersen–Gill estimate in combination with its robust standard error) for $H_0 : \beta_1 = 0$. *Note*: Data were simulated from a joint gamma frailty model with frailty variance θ, association parameter $\alpha = 0.8$ and treatment effect $\beta_1 = 0$. Subject-specific hazards are constant (Weibull scale parameters $\lambda_1 = 3$ and $\lambda_2 = 0.178$, Weibull shape parameters $\nu_1 = 1$ and $\nu_2 = 1$), resulting in a conditional survival probability of $S(2|X = 0, Z = 1) = 0.7$ at the end of follow-up. Subjects are censored administratively after $t_{max} = 2$ time units

certain. The reason for this is that both baseline recurrent event rate and treatment effect were chosen to be relatively large. However, power would be rather poor in scenarios where $\beta_1^* = 0$ but $\beta_1 \neq 0$.

Furthermore, the simulation results of Table 3 confirm the effects that were already observed in the numerical findings shown in Figures 2–4: the difference $(\hat{\beta}_1 - \beta_1)$ is zero in case of $\beta_2 = 0$, while being positive for $\beta_2 < 0$ and negative for $\beta_2 > 0$. These directions of deviation are caused by the positive association of $\alpha = 0.8$ (compare with Figure 4c). In addition, the magnitude of the difference $(\hat{\beta}_1 - \beta_1)$ is not affected by the treatment effect β_1 itself (compare with Figure 2), but is increasing with the frailty variance θ (compare with Figures 3 and 4). The difference is smaller in the scenario with a decreasing subject-specific recurrent event hazard because here a subject has on average more recurrent events before death may terminate its follow-up compared to the scenario with a constant recurrent event hazard (compare with Figure 2a).

Although this paper is primarily dealing with the recurrent event endpoint, Table 3 also shows estimates and rejection probabilities of a marginal analysis for the terminal event (using the Cox model). Of course, the difference $(\hat{\beta}_2 - \beta_2)$ is not affected by parameters that solely refer to recurrent events, for example the treatment effect β_1 and the shape of the hazard for recurrent events. But the difference is increasing with the frailty variance θ and strongly depends on β_2: although $\hat{\beta}_2 = \beta_2$ in case of $\beta_2 = 0$, $\hat{\beta}_2$ is always shifted towards zero (relative to β_2) in case of $\beta_2 \neq 0$. Importantly, the Cox analysis is keeping the type I error despite unobserved heterogeneity. These results are in line with those of other authors who already investigated the behaviour of Cox model estimation in the presence of unobserved heterogeneity (see Cécilia-Joseph et al., 2015; Henderson & Oman, 1999; Schmoor & Schumacher, 1997).

In each scenario shown in Table 3, we see slight differences between $\hat{\beta}_1$ (or $\hat{\beta}_2$), that is, the result obtained by simulation, and β_1^* (or β_2^*), that is, the result obtained by numerically solving $g_1(s) = 0$ (or $g_2(s) = 0$). These differences may be due to asymptotics and/or slight numerical imprecision.

7 | DISCUSSION

In the present paper, we derive the properties of marginal hazard ratio estimates for a recurrent event in the presence of an associated terminal event. Our findings are answering the up to now unresolved question if, why and how terminal events will affect marginal hazard ratio estimates for recurrent events. These results will contribute to a proper interpretation of marginal hazard ratios that are commonly presented from clinical trials. An important field of applications are clinical studies in heart failure, which have motivated our research: although the rate of hospitalisations due to heart failure disease will most probably be associated with the risk of CV death, the treatment effect on the hospitalisation rate is often quantified by the Andersen–Gill model – an approach that is assuming proportional hazards on the marginal level.

We have derived the least false parameter β_1^* that is consistently estimated by the Andersen–Gill estimator in a correlated frailty model by applying theory on misspecified proportional hazards models (Struthers & Kalbfleisch, 1986). Within the correlated

TABLE 3 Simulation results (10,000 simulated datasets, each with 1,000 subjects) for a marginal analysis of data from a joint gamma frailty model with association parameter $\alpha = 0.8$, frailty variance θ and treatment effects β_1 (recurrent events) and β_2 (mortality)

Simulation parameters			Recurrent events				Terminal event			
β_1	β_2	θ	$\hat{\beta}_1 - \beta_1$	$\beta_1^* - \beta_1$	$SE(\hat{\beta}_1)$	rp	$\hat{\beta}_2 - \beta_2$	$\beta_2^* - \beta_2$	$SE(\hat{\beta}_2)$	rp
(a) $\lambda_1 = 3$ and $v_1 = 1$										
0.0	−0.5	1.0	0.045	0.045	0.070	0.100	0.037	0.036	0.139	0.912
0.0	−0.5	2.0	0.076	0.077	0.094	0.130	0.063	0.063	0.147	0.850
0.0	0.0	1.0	0.000	0.000	0.071	0.052	0.002	0.000	0.125	0.053
0.0	0.0	2.0	−0.001	0.000	0.094	0.049	0.001	0.000	0.132	0.051
0.0	0.5	1.0	−0.066	−0.067	0.072	0.155	−0.053	−0.055	0.115	0.974
0.0	0.5	2.0	−0.111	−0.111	0.094	0.224	−0.089	−0.092	0.122	0.927
−0.5	−0.5	1.0	0.043	0.045	0.072	1.000	0.034	0.036	0.140	0.916
−0.5	−0.5	2.0	0.077	0.077	0.095	0.994	0.063	0.063	0.147	0.850
−0.5	0.0	1.0	−0.002	0.000	0.073	1.000	−0.001	0.000	0.125	0.048
−0.5	0.0	2.0	0.000	0.000	0.096	0.999	0.001	0.000	0.132	0.051
−0.5	0.5	1.0	−0.068	−0.067	0.074	1.000	−0.054	−0.055	0.115	0.976
−0.5	0.5	2.0	−0.111	−0.110	0.096	1.000	−0.091	−0.092	0.122	0.920
(b) $\lambda_1 = 4.24$ and $v_1 = 0.5$										
0.0	−0.5	1.0	0.029	0.029	0.069	0.073	0.037	0.036	0.139	0.912
0.0	−0.5	2.0	0.049	0.050	0.092	0.086	0.063	0.063	0.147	0.850
0.0	0.0	1.0	0.001	0.000	0.069	0.052	0.002	0.000	0.125	0.053
0.0	0.0	2.0	−0.001	0.000	0.091	0.048	0.001	0.000	0.132	0.051
0.0	0.5	1.0	−0.041	−0.043	0.069	0.097	−0.053	−0.055	0.115	0.974
0.0	0.5	2.0	−0.070	−0.070	0.091	0.121	−0.089	−0.092	0.122	0.927
−0.5	−0.5	1.0	0.028	0.029	0.070	1.000	0.034	0.036	0.140	0.916
−0.5	−0.5	2.0	0.049	0.050	0.093	0.998	0.063	0.063	0.147	0.850
−0.5	0.0	1.0	−0.001	0.000	0.071	1.000	−0.001	0.000	0.125	0.048
−0.5	0.0	2.0	0.000	0.000	0.093	1.000	0.001	0.000	0.132	0.051
−0.5	0.5	1.0	−0.043	−0.043	0.071	1.000	−0.054	−0.055	0.115	0.976
−0.5	0.5	2.0	−0.069	−0.070	0.093	1.000	−0.091	−0.092	0.122	0.920

Subject-specific hazards originate from Weibull distributions with scale parameters λ_1, λ_2 and shape parameters v_1, v_2. We consider constant subject-specific mortality hazards ($\lambda_2 = 0.178$, $v_2 = 1$) in combination both with (a) constant ($\lambda_1 = 3$, $v_1 = 1$) and (b) decreasing ($\lambda_1 = 4.24$, $v_1 = 0.5$) subject-specific recurrent event hazards. Subjects are censored administratively after $t_{max} = 2$ time units. The table shows the estimates ($\hat{\beta}_i$), the robust standard errors ($SE(\hat{\beta}_i)$) and rejection probabilities (rp) resulting from the simulation of a marginal model analysis. In addition, the asymptotically valid least false parameters (β_i^*, numerical calculation) are shown.

frailty model, an association between the recurrent and the terminal event process is induced by correlated random effects Z and W. It implies the assumption of proportional hazards with treatment effects β_1 (recurrent events) and β_2 (terminal event) on the subject's level. Moreover, it assumes the subject-specific counting process for the recurrent events to be of Poisson type. On the marginal level, the proportional hazards assumption does in general not hold anymore, leading to a misspecification of the Andersen–Gill model. We have shown that the least false parameter is implicitly defined by the true underlying data-generating process, that is, it depends on parameters of the correlated frailty model (treatment effects, baseline hazards, frailty distributions). In addition, the asymptotic estimate strongly depends on the censoring distribution including especially the length of follow-up. Using these derivations we proved that the Andersen–Gill estimator asymptotically coincides with the conditional (i.e. subject-specific) treatment effect, if the treatment does not affect mortality. Further results were obtained numerically for the case of a joint frailty model, where the frailties affecting the two processes have a deterministic dependency: here we have shown that the sign of the difference ($\beta_1^* - \beta_1$) is, amongst others, determined by the treatment effect on mortality β_2. Moreover, our results suggest that the size of the difference ($\beta_1^* - \beta_1$) is not affected by the treatment effect β_1 itself. In particular, a misjudgement of the treatment effect direction on recurrent events is possible, that is, sign(β_1^*) $\neq$ sign(β_1). In this paper, we focussed on the gamma frailty as it is most widely applied (Wienke, 2010). Equal conclusions are obtained when a log-normal frailty is applied. To investigate other frailty distributions, further numerical investigations or simulation studies could be performed.

Some authors use the term 'bias' for describing deviations of marginal model estimates from conditional ones. As both approaches have different purposes depending on the aim of the study, differences between marginal and conditional model estimates are not a matter of bias. Marginal estimates provide an information about the treatment effect on the population's level, whereas conditional models, which account for the association by using frailty terms, are intended for the assessment of a treatment effect on the subject's level.

As pointed out by Bretagnolle and Huber-Carol (1988), Schmoor and Schumacher (1997) and Aalen et al. (2015), omission of relevant covariates in a usual Cox analysis for a survival outcome leads to non-causal treatment effect estimates despite randomisation. In the present paper, we transfer these concepts to the analysis of recurrent events and show that the same issue arises in the assessment of a treatment effect on recurrent events if an associated terminal event is present. If the treatment affects mortality, a constant conditional hazard ratio (i.e. on the subject's level) translates to a time-dependent marginal hazard ratio (i.e. on the population's level), resulting in a time-dependent marginal hazard ratio estimate $\exp(\beta_1^*)$. However, if treatment does not affect mortality, unobserved risk factors that affect both endpoints, remain balanced over time between the survivors in the treatment and the control group. Hence, the marginal treatment effect estimate coincides with the conditional treatment effect estimate despite an association between the processes.

Recently, Rogers et al. (2016) recommended to evaluate future heart failure trials by using a joint frailty model. The authors performed simulations to study the behaviour of different marginal treatment effect estimates for the recurrent event endpoint, when informative dropouts according to a joint frailty model occur. Our findings extend these simulation results by providing numerical solutions for the Andersen–Gill estimates. Our analytical results are supported by clinical data of the CHARM-Preserved heart failure trial. Here the Andersen–Gill estimate does only slightly differ from the joint frailty estimate for the recurrent events because mortality seems to be unaffected by the Candesartan treatment (Rogers, Pocock, et al., 2014; Yusuf et al., 2003). Nowadays, lots of heart failure drug trials fail to show a treatment effect on mortality (Zheng et al., 2018), suggesting that marginal and conditional treatment effect estimates for the hospitalisation rate would coincide – provided the assumptions of a joint frailty model hold. Our results can also be applied for diabetes trials where cardiovascular outcomes such as HFHs and CV death are also of interest (Schnell et al., 2017). Furthermore, the findings presented here are relevant for the analysis of safety data in randomised clinical trials, where the treatment effect on the rate of recurrent, adverse events is of interest (Allignol, Beyersmann, & Schmoor, 2016; Hengelbrock, Gillhaus, Kloss, & Leverkus, 2016; Schmoor, Bender, Beyersmann, Kieser, & Schumacher, 2016). In that context, treatment discontinuation (e.g. due to death) represents the terminal event that is probably associated with the risk of adverse events.

In a recent article, Putter and van Houwelingen critically discussed the use of frailties in multistate models to account for possible associations between transition times. They have pointed out that deviations from the modelling assumptions can be absorbed by the frailty term in terms of estimating an increased frailty variance (Putter & van Houwelingen, 2015). Consequently, it can be difficult to disentangle (true) heterogeneity and violations of the modelling assumptions when applying the joint frailty model. Besides assuming proportional hazards on the subject's level, a further crucial assumption of this model is that the subject-specific hazards for both endpoints only depend on time, but not on the history of recurrent events. This is questionable and the model could be improved by allowing for event dependency on the subject's level. However, such complex models will result in identifiability problems, as it is difficult, if not impossible, to distinguish between event dependency and unexplained heterogeneity in the data.

As aforementioned, several potential pitfalls have to be considered when treatment effects are assessed by more and more complicated models. Therefore, simpler marginal models may be preferred and we need to be able to interpret their estimates, if the underlying data-generating processes are rather complex. We contribute to the interpretation of the Andersen–Gill estimate for the treatment effect on the recurrent event rate, if informative dropouts due to an associated terminal event are present. Our results complement previous articles reporting on treatment effect estimation in a misspecified Andersen–Gill model: Lin et al. (2000) and Liu (2014) showed that marginal and conditional treatment effects coincide, if unexplained heterogeneity, but no associated terminal event is present. Cheung, Xu, Tan, Cutts, and Milligan (2010) analysed asymptotic properties of the Andersen–Gill estimator in a model with unexplained heterogeneity and event dependency, again without considering an associated terminal event. To complete the picture, a straightforward next step would be to characterise the estimator's properties if the data-generating process contains both informative dropouts and event dependency, probably reflecting the most realistic scenario in heart failure trials.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ORCID

Gerrit Toenges (iD) https://orcid.org/0000-0002-9745-9437

REFERENCES

Aalen, O. O., Cook, R. J., & Røysland, K. (2015). Does Cox analysis of a randomized survival study yield a causal treatment effect? *Lifetime Data Analysis*, *21*(4), 579–593.

Allignol, A., Beyersmann, J., & Schmoor, C. (2016). Statistical issues in the analysis of adverse events in time-to-event data. *Pharmaceutical Statistics*, *15*(4), 297–305.

Andersen, P. K., & Gill, R. (1982). Cox's regression model for counting processes: A large sample study. *Annals of Statistics*, *10*(4), 1100–1120.

Anker, S. D., & McMurray, J. J. (2012). Time to move on from 'time-to-first': Should all events be included in the analysis of clinical trials? *European Heart Journal*, *33*(22), 2764–2765.

Anker, S. D., Schroeder, S., Atar, D., Bax, J. J., Ceconi, C., Cowie, M. R., … Ruschitzka, F. (2016). Traditional and new composite endpoints in heart failure clinical trials: Facilitating comprehensive efficacy assessments and improving trial efficiency. *European Journal of Heart Failure*, *18*(5), 482–489.

Bretagnolle, J., & Huber-Carol, C. (1988). Effects of omitting covariates in Cox's model for survival data. *Scandinavian Journal of Statistics*, *15*(2), 125–138.

Cécilia-Joseph, E., Auvert, B., Broët, P., & Moreau, T. (2015). Influence of trial duration on the bias of the estimated treatment effect in clinical trials when individual heterogeneity is ignored. *Biometrical Journal*, *57*(3), 371–383.

Cheung, Y. B., Xu, Y., Tan, S. H., Cutts, F., & Milligan, P. (2010). Estimation of intervention effects using first or multiple episodes in clinical trials: The Andersen–Gill model re-examined. *Statistics in Medicine*, *29*(3), 328–336.

Cook, R. J., & Lawless, J. F. (2007). *The statistical analysis of recurrent events*. New York: Springer Science & Business Media.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, *34*(2), 187–220.

Ferreira-González, I., Busse, J. W., Heels-Ansdell, D., Montori, V. M., Akl, E. A., Bryant, D. M., … Guyatt, G. H. (2007). Problems with use of composite end points in cardiovascular trials: Systematic review of randomised controlled trials. *British Medical Journal*, *334*(7597), 786.

Gerds, T. A., & Schumacher, M. (2001). On functional misspecification of covariates in the Cox regression model. *Biometrika*, *88*(2), 572–580.

Grambauer, N., Schumacher, M., & Beyersmann, J. (2010). Proportional subdistribution hazards modeling offers a summary analysis, even if misspecified. *Statistics in Medicine*, *29*(7–8), 875–884.

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

Hengelbrock, J., Gillhaus, J., Kloss, S., & Leverkus, F. (2016). Safety data from randomized controlled trials: Applying models for recurrent events. *Pharmaceutical Statistics*, *15*(4), 315–323.

Huang, X., & Liu, L. (2007). A joint frailty model for survival and gap times between recurrent events. *Biometrics*, *63*(2), 389–397.

Kelly, P. J., & Lim, L. L. (2000). Survival analysis for recurrent event data: An application to childhood infectious diseases. *Statistics in Medicine*, *19*(1), 13–33.

Lin, D. Y., Wei, L. J., Yang, I., & Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *62*, 711–730.

Liu, X. (2014). Survival models on unobserved heterogeneity and their applications in analyzing large-scale survey data. *Journal of Biometrics and Biostatistics*, *5*(2), 157–162.

Liu, L., Wolfe, R. A., & Huang, X. (2004). Shared frailty models for recurrent events and a terminal event. *Biometrics*, *60*(3), 747–756.

Putter, H., & van Houwelingen, H. C. (2015). Frailties in multi-state models: Are they identifiable? Do we need them? *Statistical Methods in Medical Research*, *24*(6), 675–692.

Rogers, J. K., Jhund, P. S., Perez, A. C., Böhm, M., Cleland, J. G., Gullestad, L., … Pocock, S. J. (2014). Effect of rosuvastatin on repeat heart failure hospitalizations: The CORONA Trial (Controlled Rosuvastatin Multinational Trial in Heart Failure). *JACC: Heart Failure*, *2*(3), 289–297.

Rogers, J. K., McMurray, J. J., Pocock, S. J., Zannad, F., Krum, H., van Veldhuisen, D. J., … Pitt, B. (2012). Eplerenone in patients with systolic heart failure and mild symptoms: Analysis of repeat hospitalizations. *Circulation*, *126*(19), 2317–2323.

Rogers, J. K., Pocock, S. J., McMurray, J. J., Granger, C. B., Michelson, E. L., Östergren, J., … Yusuf, S. (2014). Analysing recurrent hospitalizations in heart failure: A review of statistical methodology, with application to CHARM-Preserved. *European Journal of Heart Failure*, *16*(1), 33–40.

Rogers, J. K., Yaroshinsky, A., Pocock, S. J., Stokar, D., & Pogoda, J. (2016). Analysis of recurrent events with an associated informative dropout time: Application of the joint frailty model. *Statistics in Medicine*, *35*(13), 2195–2205.

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

Rondeau, V., Mathoulin-Pelissier, S., Jacqmin-Gadda, H., Brouste, V., & Soubeyran, P. (2007). Joint frailty models for recurring events and death using maximum penalized likelihood estimation: Application on cancer events. *Biostatistics*, *8*(4), 708–721.

Schmoor, C., Bender, R., Beyersmann, J., Kieser, M., & Schumacher, M. (2016). Adverse event development in clinical oncology trials. *Lancet Oncology*, *17*(7), e263–e264.

Schmoor, C., & Schumacher, M. (1997). Effects of covariate omission and categorization when analysing randomized trials with the Cox model. *Statistics in Medicine*, *16*(1–3), 225–237.

Schnell, O., Rydén, L., Standl, E., Ceriello, A., on behalf of the D&CVD EASD Study Group (2017). Updates on cardiovascular outcome trials in diabetes. *Cardiovascular Diabetology*, *16*(1), 128.

Struthers, C. A., & Kalbfleisch, J. D. (1986). Misspecified proportional hazard models. *Biometrika*, *73*(2), 363–369.

Wienke, A. (2010). *Frailty models in survival analysis*. Boca Raton, FL: Chapman and Hall/CRC Biostatistics Series.

Yusuf, S., Pfeffer, M. A., Swedberg, K., Granger, C. B., Held, P., McMurray, J. J. V., … Östergren, J. (2003). Effects of candesartan in patients with chronic heart failure and preserved left-ventricular ejection fraction: The CHARM-Preserved Trial. *Lancet*, *362*(9386), 777–781.

Zannad, F., Stein, K., Garcia, A. A., Anker, S. D., Armstrong, P. W., Calvo, G., … McMurray, J. J. V. (2013). Clinical outcome endpoints in heart failure trials: A European Society of Cardiology Heart Failure Association consensus document. *European Journal of Heart Failure*, *15*(10), 1082–1094.

Zanolla, L., & Zardini, P. (2003). Selection of endpoints for heart failure clinical trials. *European Journal of Heart Failure*, *5*(6), 717–723.

Zheng, S. L., Chan, F. T., Nabeebaccus, A. A., Shah, A. M., McDonagh, T., Okonko, D. O., & Ayis, S. (2018). Drug treatment effects on outcomes in heart failure with preserved ejection fraction: A systematic review and meta-analysis. *Heart*, *104*(5), 407–415.

Zinman, B., Wanner, C., Lachin, J. M., Fitchett, D., Bluhmki, E., Hantel, S., … Inzucchi, S. E. (2015). Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. *New England Journal of Medicine*, *373*(22), 2117–2128.

SUPPORTING INFORMATION

Additional supporting information including source code to reproduce the results may be found online in the Supporting Information section at the end of the article.

APPENDIX

A.1 Associated processes in the correlated frailty model

$$r_1^*(t|H(t), D = s) = r_1^*(t|X, D = s)$$

$$= \lim_{\Delta \searrow 0} \frac{\int_0^\infty P(\Delta N_1(t) = 1|X, Z = z, D = s) \cdot f_{Z|X,D=s}(z)\, dz}{\Delta}$$

$$= \begin{cases} 0, & t \geq s \\ \int_0^\infty \lambda_{10}(t) \exp(\beta_1' X) \cdot z \cdot f_{Z|X,D=s}(z)\, dz, & t < s \end{cases}$$

$$= \begin{cases} 0, & t \geq s \\ \lambda_{10}(t) \exp(\beta_1' X) E(Z|X, D = s), & t < s. \end{cases}$$

As Z and D are dependent (via W), the conditional expectation $E(Z|X, D = s)$ depends on s. Consequently the rate $r_1^*(t|H(t), D = s)$ does, for $t < s$, depend on s. This is considered as an association of the processes throughout the paper.

A.2   Marginal hazards in the correlated frailty model: Derivations

Derivation of marginal hazards in the correlated frailty model (for binary X) as given in (10) and (11):

$$r_1(t|X) = E_Z(\lambda_1(t|X, Z))$$

$$= \lim_{\Delta \searrow 0} \frac{\int_0^\infty P(\Delta N_1(t) = 1|X, Z = z, D \geq t) \cdot f_{Z|X,D\geq t}(z) \, dz}{\Delta}$$

$$= \int_0^\infty \lambda_{10}(t) \exp(\beta_1 X) \cdot z \cdot f_{Z|X,D\geq t}(z) \, dz$$

$$= \lambda_{10}(t) \exp(\beta_1 X) \cdot E(Z|X, D \geq t)$$

$$r_2(t|X) = E_W(\lambda_2(t|X, W))$$

$$= \lim_{\Delta \searrow 0} \frac{\int_0^\infty P(\Delta N_2(t) = 1|X, W = w, D \geq t) \cdot f_{W|X,D\geq t}(w) \, dw}{\Delta}$$

$$= \int_0^\infty \lambda_{20}(t) \exp(\beta_2 X) \cdot w \cdot f_{W|X,D\geq t}(w) \, dw$$

$$= \lambda_{20}(t) \exp(\beta_2 X) \cdot E(W|X, D \geq t).$$

Let $S(t|X)$ and $S(t|X, W = w)$ denote the marginal and the conditional survival function, respectively. Let further $\Lambda_{20}(t) = \int_0^t \lambda_{20}(s) \, ds$ denote the cumulative baseline hazard for mortality. Using these notations, we can now derive the conditional expectation of Z appearing in the marginal hazard formula for the recurrent events (10):

$$E(Z|X, D \geq t) = \int_0^\infty z f_{Z|X,D\geq t}(z) \, dz$$

$$= \frac{1}{S(t|X)} \int_0^\infty \int_0^\infty \int_t^\infty z f_{Z,W,D|X}(z, w, s) \, ds \, dw \, dz$$

$$= \frac{1}{E_W(S(t|X, W))} \int_0^\infty \int_0^\infty \int_t^\infty z f_{D|X,Z=z,W=w}(s) f_{Z,W}(z, w) \, ds \, dw \, dz$$

$$= \frac{1}{\int_0^\infty S(t|X, W = w) f_W(w) \, dw} \int_0^\infty \int_0^\infty z S(t|X, W = w) f_{Z,W}(z, w) \, dw \, dz$$

$$= \frac{\int_0^\infty \int_0^\infty z \, \exp\left(-w \exp(\beta_2 X)\Lambda_{20}(t)\right) f_{Z,W}(z, w) \, dw \, dz}{\int_0^\infty \exp\left(-w \exp(\beta_2' X)\Lambda_{20}(t)\right) f_W(w) \, dw}$$

$$= \frac{\int_0^\infty \int_0^\infty z \, \exp\left(-w \exp(\beta_2 X)\Lambda_{20}(t)\right) f_{Z,W}(z, w) \, dw \, dz}{\mathcal{L}_W\left(\exp(\beta_2 X)\Lambda_{20}(t)\right)}.$$

Here we used that the marginal survival function can be expressed in terms of the Laplace transform, that is $S(t|X) = \mathcal{L}_W(\exp(\beta_2 X)\Lambda_{20}(t))$. Analogously, we can derive the conditional expectation of W appearing in the marginal hazard formula for the terminal event (11):

$$E(W|X, D \geq t) = \int_0^\infty w f_{W|X,D\geq t}(w) \, dw$$

$$= \frac{1}{S(t|X)} \int_0^\infty \int_t^\infty w f_{W,D|X}(w, s) \, ds \, dw$$

$$= \frac{1}{\mathcal{L}_W\left(\exp(\beta_2 X)\Lambda_{20}(t)\right)} \int_0^\infty \int_t^\infty w f_{D|X,W=w}(s) f_W(w)\, ds\, dw$$

$$= \frac{1}{\mathcal{L}_W\left(\exp(\beta_2 X)\Lambda_{20}(t)\right)} \int_0^\infty w S(t|X, W = w) f_W(w)\, dw$$

$$= \frac{\int_0^\infty w \, \exp\left(-w \exp(\beta_2 X)\Lambda_{20}(t)\right) f_W(w)\, dw}{\mathcal{L}_W\left(\exp(\beta_2 X)\Lambda_{20}(t)\right)}$$

$$= -\frac{\mathcal{L}'_W\left(\exp(\beta_2 X)\Lambda_{20}(t)\right)}{\mathcal{L}_W\left(\exp(\beta_2 X)\Lambda_{20}(t)\right)}.$$

A.3 Direction of deviation for the marginal treatment effect estimate in a joint gamma frailty model with $\alpha = 1$

Here we provide a proof regarding the sign of the difference $(\beta_1^* - \beta_1)$ for the special situation of a joint gamma frailty model with association parameter $\alpha = 1$. As shown by other authors (see, e.g. Wienke, 2010), the Laplace transform of a gamma distribution (with mean 1 and variance θ) and its derivative are given by

$$\mathcal{L}_Z(s) = (1 + \theta s)^{-\frac{1}{\theta}} \quad \text{and} \quad \mathcal{L}'_Z(s) = -(1 + \theta s)^{-\frac{1}{\theta} - 1},$$

respectively. Using (14) for the special case of $W = Z$, the marginal hazard ratio for the recurrent event endpoint simplifies to

$$\frac{r_1(t|X = 1)}{r_1(t|X = 0)} = \exp(\beta_1)\frac{1 + \theta\Lambda_{20}(t)}{1 + \theta\exp(\beta_2)\Lambda_{20}(t)} \begin{cases} \leq \exp(\beta_1) \ \forall t & \Leftrightarrow \ \beta_2 > 0 \\ = \exp(\beta_1) \ \forall t & \Leftrightarrow \ \beta_2 = 0 \\ \geq \exp(\beta_1) \ \forall t & \Leftrightarrow \ \beta_2 < 0 \end{cases}.$$

The sign of the treatment effect on mortality determines, in which direction the marginal hazard ratio deviates from the conditional one for every time-point during follow-up. Applying (22) results in

$$\beta_1^* \leq \beta_1 \ \Leftrightarrow \ \beta_2 > 0$$

$$\beta_1^* = \beta_1 \ \Leftrightarrow \ \beta_2 = 0$$

$$\beta_1^* \geq \beta_1 \ \Leftrightarrow \ \beta_2 < 0.$$