Received: 4 August 2020 | Revised: 5 July 2021 | Accepted: 14 July 2021

DOI: 10.1002/sim.9149

RESEARCH ARTICLE

Statistics in Medicine WILEY

A comparison of semiparametric approaches to evaluate composite endpoints in heart failure trials

Gerrit Toenges[1] | **Tobias Mütze**[2] | **Antje Jahn-Eimermacher**[3]

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

[2]Statistical Methodology, Novartis Pharma AG, Basel, Switzerland

[3]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

In heart failure (HF) trials efficacy is usually assessed by a composite endpoint including cardiovascular death (CVD) and heart failure hospitalizations (HFHs), which has traditionally been evaluated with a time-to-first-event analysis based on a Cox model. As a considerable fraction of events is ignored that way, methods for recurrent events were suggested, among others the semiparametric proportional rates models by Lin, Wei, Yang, and Ying (LWYY model) and Mao and Lin (Mao-Lin model). In our work we apply least false parameter theory to explain the behavior of the composite treatment effect estimates resulting from the Cox model, the LWYY model, and the Mao-Lin model in clinically relevant scenarios parameterized through joint frailty models. These account for both different treatment effects on the two outcomes (CVD, HFHs) and the positive correlation between their risk rates. For the important setting of beneficial outcome-specific treatment effects we show that the correlation results in composite treatment effect estimates, which are decreasing with trial duration. The estimate from the Cox model is affected more by the attenuation than the estimates from the recurrent event models, which both demonstrate very similar behavior. Since the Mao-Lin model turns out to be less sensitive to harmful effects on mortality, we conclude that, among the three investigated approaches, the LWYY model is the most appropriate one for the composite endpoint in HF trials. Our investigations are motivated and compared with empirical results from the PARADIGM-HF trial (ClinicalTrials.gov identifier: NCT01035255), a large multicenter trial including 8399 chronic HF patients.

KEYWORDS

composite endpoint, heart failure, joint frailty model, least false parameter, recurrent events

1 | INTRODUCTION

Clinical trials in chronic diseases have to deal with multiple types of disease-related events that can be divided into fatal and nonfatal ones. Nonfatal events can occur repeatedly and serve as a surrogate for the patients' morbidity.[1,2] Throughout this article we will focus on heart failure (HF) trials. In patients suffering from HF disease-related mortality strongly decreased under standard-of-care treatment over the past decades, calling for even larger required sample sizes if efficacy

was purely assessed in terms of mortality.[3] It is for that reason, among others, that nowadays the primary analysis is based on a composite endpoint including both heart failure hospitalizations (HFHs) and cardiovascular death (CVD). Composite endpoints in general have advantages and disadvantages, but in HF trials they are widely accepted by regulators and scientists as long as a time-to-first-event analysis based on a Cox model is used to assess efficacy.[4,5] As the name suggests, here only each patient's first event (be it death or hospitalization) is incorporated in the effect measure. This approach raised criticism, since approximately 40% to 50% of the observed HFHs and CVDs in major HF trials are ignored in the primary analysis.[6,7] One example that we will later on use as a case study is the PARADIGM-HF trial,[8] a large multicenter trial including a total of 8399 patients suffering from chronic HF with reduced ejection fraction. In the PARADIGM-HF trial, 2031 of 3181 events (63.8%) were first events and as such included in the primary analysis.[8,9] In contrast to the traditional time-to-first-event approach, models for recurrent events can incorporate all events and are therefore discussed in the scientific community.[7,10,11]

Recurrent event approaches are more complex and might make the interpretation of estimates more difficult, especially in the presence of potential model violations. Further methodological difficulties arise if patients die while being hospitalized for HF, that is, if CVD and HFH event times coincide. In addition, skeptics worry that the estimate for the composite might be dominated by the HFH component and be driven by a small proportion of patients with a disproportionately large number of HFH events. Such clustering pattern is provoked by unexplained heterogeneity in the patient-specific risk rates and has been empirically identified as an important issue in HF trials.[10,12] However, it is precisely the between-patient heterogeneity which might favor the use of recurrent event methods over the time-to-first-event approach: It is well known that the estimate of the Cox model is biased toward the null hypothesis in the presence of unexplained heterogeneity.[13,14] Such selection bias is less of an issue for recurrent event analysis, because here patients are not removed from the risk set after their first hospitalization. This is empirically supported by the fact that in some reanalyses of major HF trials recurrent event methods were shown to yield stronger effect estimates than a time-to-first-event analysis.[9-11,15] Hence proponents justifiably argue that recurrent event models yield estimates that are more accurate and, as all events are considered in the effect measure, are better capturing the effect on the disease burden. Moreover, recurrent event approaches have been proven to be more powerful than a time-to-first-event analysis in most, albeit not all, scenarios investigated so far.[10]

Originally recurrent event methodology only covered settings, where the observation of repeat events like HFHs within a subject could exclusively be stopped by censoring. Censoring does per definition only prevent the observation but not the occurrence of further recurrent events, implying that these models consider the recurrent event process to continue after the end of the observation time.[16] A prominent example is the semiparametric proportional rates model by Lin, Wei, Yang, and Ying[17] that we will refer to as LWYY model in the following. Although actually not being designed for that purpose, the LWYY model is considered as an alternative to the time-to-first-event analysis for the composite endpoint in HF trials.[18] By handling CVD technically both as last recurrent event and as censoring event, the LWYY estimate targets the effect on the composite event rate conditional on being alive.[11] In the past two decades numerous recurrent event models arose that specifically acknowledge the existence of a terminal event like CVD (eg, joint frailty models,[19] marginal models[20]). In contrast to censoring, the terminal event does not only prevent the observation but also the occurrence of further recurrent events. Different rate or mean functions can be modeled in this framework.[21,22] Recently a semiparametric proportional rates model for the composite endpoint was suggested by Mao and Lin[23] that we will refer to as Mao-Lin model in the following. To our knowledge this model is the only recurrent event approach that was specifically designed for the composite of recurrent and terminal events. It targets the effect on the composite event rate without conditioning on being alive and is in that sense marginal.[11] Let us illustrate the conceptual differences between the LWYY and the Mao-Lin approach for the composite with a simplified example: A patient is hospitalized 1 year after study inclusion and dies 2 years after study inclusion. Initially a follow-up of 3 years was scheduled. Here the LWYY rate yields (2 events)/(2 years), whereas the Mao-Lin rate is given by (2 events)/(3 years). While both approaches coincide in the number of counted events (nominator of the rate), they differ with respect to the "at-risk" time (denominator of the rate). In contrast, the rate targeted by the time-to-first-event analysis is given by (1 event)/(1 year), thus differing from the recurrent event approaches both in terms of counted events and "at-risk" time.

The recent qualification opinion of the European Medicines Agency (EMA) on recurrent event endpoints[24] has shown that regulators and researchers seem to be open-minded to new approaches. However, a better understanding of differences between estimates obtained from time-to-first-event analyses and recurrent event methods is still needed to establish recurrent event methodology in future HF trials. Whereas potential explanations for differences between time-to-first-event and single recurrent event methods have already been discussed,[10,25] explanations on differences among recurrent event methods are missing up to now in the context of HF trials. In particular, the Mao-Lin model has

not been considered so far—probably because the model is quite new and its computation is not straightforward in common statistical software. Moreover, consequences of a correlation between the composite's components (HFHs, CVD), as observed in HF trials,[12] have not yet been investigated in the context of composite treatment effect estimates. These issues are addressed in this article, which has two main objectives: First, we aim to develop the methodological foundations that explain the behavior of the LWYY, the Mao-Lin and the traditional Cox model estimates in clinically relevant scenarios. Second, we intend to identify strengths and weaknesses of each approach in order to recommend the most appropriate one for HF trials.

To explain differences between the composite treatment effect estimates, first a realistic parameterization of the data-generating process is needed. In particular, different component-specific effects on the two outcomes (HFHs, CVD) and a positive correlation between the component-specific risk rates need to be considered. A joint frailty model[19] accounts for these issues and was therefore considered as a suitable model for the data-generating process in the context of HF trials.[12] Here the correlation between the two outcomes is modeled by a shared frailty term, representing unknown risk factors (ie, unexplained heterogeneity) that are acting both on the HFH and the CVD rate. A joint frailty model is very useful to estimate the component-specific treatment effects and would thus be suited for the primary analysis, if the treatment effects on HFHs and CVD were specified as co-primary endpoints or as part of a hierarchical testing procedure. However, a joint frailty model cannot provide an estimate of the composite treatment effect, which generally defines the primary endpoint in HF trials. For the latter purpose, approaches like the Cox model, the LWYY model or the Mao-Lin model are thus of larger interest.

In our investigations we characterize the composite treatment effect estimates of these three analysis models in joint frailty scenarios that are motivated by the PARADIGM-HF trial. Therefore we apply both simulations and least false parameter theory. The term "least false parameter" refers to the estimate that is asymptotically obtained by some time-to-event estimator under a particular data distribution that violates the estimator's modeling assumptions (eg, nonproportionality of rates). This limit value is minimizing a certain distance measure between the true probability distribution of the data and the probability distributions underlying the analysis model. Given a particular estimator and the parameterization of the true data-generating process, the least false parameter can be derived numerically, that is, without simulation. First analytical results in that field go back to Struthers and Kalbfleisch[26] and Hjort,[27] who investigated the consequences of misspecifications in univariate survival models. Lin et al[17] later on contributed findings for misspecified recurrent event approaches, albeit the authors did not explicitly use the notion "least false parameter". These results have recently been used to investigate the consequences of model misspecifications (eg, through missing covariates or informative drop-outs) for a recurrent event endpoint,[28-30] but have not yet been applied to the composite endpoint of recurrent and terminal events.

The article is organized as follows: First we will give a brief outline of the PARADIGM-HF trial in Section 2. Then we will introduce models and notations (Section 3) and show how least false parameter theory enables to analytically derive composite treatment effect estimates relying on the different analysis models (Cox, LWYY, Mao-Lin) in data-generating processes that are parameterized through joint frailty models (Section 4). After that we present results obtained for the PARADIGM-HF data (Section 5) and use them to identify clinically relevant joint frailty scenarios, for which the three analysis models will be characterized in terms of estimates, standard errors and rejection probabilities (Section 6). Finally, in Section 7 we conclude with a discussion of our new results.

2 | CLINICAL EXAMPLE PART I: THE PARADIGM-HF TRIAL

The PARADIGM-HF trial (ClinicalTrials.gov Identifier: NCT01035255) was a multicenter, randomized, double-blind, parallel group, active-controlled study that evaluated the efficacy and safety of sacubitril/valsartan compared with enalapril on morbidity and mortality in patients with chronic heart failure and reduced ejection fraction.[8] 8399 patients were randomized and followed for a median of 27 months. The primary endpoint was the time to the first occurrence of the composite endpoint of CVD or HFH. In total, 3181 events (1251 CVDs, 1930 HFHs) were observed during the trial. Thereof 2031 events (836 CVDs, 1195 HFHs) were first events and as such considered in the primary analysis. Table 1 gives a detailed overview on the patient and event numbers for the two treatments. The trial was stopped early due to the clear efficacy of sacubitril/valsartan. The estimated time-to-first-event rate ratio (Cox model) comparing sacubitril/valsartan with enalapril as reference was 0.798 with a 95%-confidence interval of $[0.731,\ 0.871]$.[8]

While the primary endpoint did not include nonfirst events, there has been interest in comparing the effect of sacubitril/valsartan compared with enalapril with methods for recurrent events. Mogensen et al[9] published post hoc analyses

TABLE 1 Patient and event numbers for the two treatment regimens sacubitril/valsartan and enalapril in the PARADIGM-HF trial (inspired by Mogensen et al[9])

	HFHs		Composite (HFHs + CVD)	
	Sacubitril/valsartan	**Enalapril**	**Sacubitril/valsartan**	**Enalapril**
Patients, *n*	4187 (100%)	4212 (100%)	4187 (100%)	4212 (100%)
Events per patient				
0	3650 (87.2%)	3554 (84.4%)	3273 (78.2%)	3095 (73.5%)
1	367 (8.8%)	418 (9.9%)	629 (15.0%)	751 (17.8%)
2	110 (2.6%)	143 (3.4%)	186 (4.4%)	204 (4.8%)
3	33 (0.8%)	53 (1.3%)	59 (1.4%)	96 (2.3%)
4	9 (0.2%)	22 (0.5%)	17 (0.4%)	34 (0.8%)
5	5 (0.1%)	16 (0.4%)	7 (0.2%)	17 (0.4%)
6	3 (0.1%)	2 (0.0%)	5 (0.1%)	11 (0.3%)
7	6 (0.1%)	1 (0.0%)	4 (0.1%)	1 (0.0%)
8	1 (0.0%)	0 (0.0%)	4 (0.1%)	0 (0.0%)
9	2 (0.0%)	2 (0.0%)	0 (0.0%)	0 (0.0%)
10	0 (0.0%)	0 (0.0%)	2 (0.0%)	2 (0.0%)
11	0 (0.0%)	1 (0.0%)	0 (0.0%)	1 (0.0%)
18	1 (0.0%)	0 (0.0%)	1 (0.0%)	0 (0.0%)
At least one event	537 (12.8%)	658 (15.6%)	914 (21.8%)	1117 (26.5%)
At least two events	170 (4.1%)	240 (5.7%)	285 (6.8%)	366 (8.7%)
Total events	851	1079	1409	1772
Total follow-up years	9308	9235	9308	9235
Total CVDs			558 (13.3%)	693 (16.5%)
Total all-cause deaths			711 (17.0%)	835 (19.8%)

on the composite of CVD and recurrent HFHs with common recurrent event models such as the LWYY model. Essentially they showed that the treatment effect estimates and confidence intervals from these models are very close to those obtained from the Cox model for the time-to-first-event outcome. In particular, these results do not indicate a gain of statistical efficiency, as generally presumed for recurrent event methods. This may partially be explained by the large drug discontinuation rate ($> 40\%$) after first HFH in the trial.[10] It is worth to note that the Mao-Lin model was not represented in the recurrent event methods considered by Mogensen et al.[9]

3 | MODELS AND NOTATIONS

In this section first general notations will be introduced. Based on this, differences between the targeted rate functions and estimators of the Cox model, the LWYY model and the Mao-Lin model for the composite endpoint will be elaborated.

3.1 | General setting

First we introduce D_i as the terminal event (death) time, M_i as the number of recurrent events (hospitalizations) in the time interval $[0, D_i]$ and $0 < T_{i1} < T_{i2} < \cdots < T_{iM_i} \leq T_{i0} = \infty$ as the corresponding recurrent event times for patient i. In case of $M_i = 0$, we additionally define $T_{i1} = \infty$. A sample of n patients with independent and identically distributed between-patient data is assumed. The event times of the ith patient can be represented by a bivariate counting process

with components

$$N_{i1}(t) = \sum_{k=0}^{M_i} I(T_{ik} \leq t) \quad \text{and} \quad N_{i2}(t) = I(D_i \leq t) \tag{1}$$

that are counting the events of the respective outcome (recurrent events or terminal event) over time. Here $I(\cdot)$ denotes the indicator function. The composite counting process is given by $N_i(t) = N_{i1}(t) + N_{i2}(t)$.

With t^- being an infinitesimally smaller time than t and $\mathrm{d}t \to 0$ being a time increment that approaches zero it follows that $\mathrm{d}N(t) = (N((t + \mathrm{d}t)^-) - N(t^-))$ is the number of events occurring in the time interval $[t, t + \mathrm{d}t)$ for a univariate counting process $N(t)$ of interest. Let further $X_i = (X_{i1}, \ldots, X_{ip})'$ be a p-dimensional vector of external, time-independent covariates and $H_i(t) = \{N_{i1}(s), N_{i2}(s),\ 0 \leq s < t\}$ the event history of both counting processes for the single outcomes up to time t. In addition we denote C_i as a right-censoring time that prevents the observation but not the occurrence of further events, implying that patients are only followed up to time $V_i = D_i \wedge C_i$. Here "$\wedge$" denotes the minimum operator. It is important to notice that technically the censoring time can be larger than the death time.

3.2 | Semiparametric models for the composite endpoint

In order to estimate an overall treatment effect for the composite, often a time-to-first-event analysis based on a Cox model[31] is carried out. Beyond that, semiparametric recurrent event methods - incorporating all events - may be applied. As representatives for the latter, we will focus on the LWYY model[17] (in its repurposed form for the composite of recurrent and terminal events) and the Mao-Lin model.[23] All the aforementioned models assume the censoring time C_i to be noninformative and impose a proportionality assumption of the form $r_i(t|X_i) = r_0(t)\exp(\beta' X_i)$ on certain model-specific composite event rates that are given by

$$\begin{aligned} r_i^{Cox}(t|X_i)\ \mathrm{d}t &= E[\mathrm{d}N_i(t)|X_i, D_i \geq t, T_{i1} \geq t], \\ r_i^{LWYY}(t|X_i)\ \mathrm{d}t &= E[\mathrm{d}N_i(t)|X_i, D_i \geq t], \\ r_i^{Mao-Lin}(t|X_i)\ \mathrm{d}t &= E[\mathrm{d}N_i(t)|X_i]. \end{aligned} \tag{2}$$

Here $\beta = (\beta_1, \ldots, \beta_p)'$ denotes the parameter vector for the p covariates contained in the vector X_i and $r_0(t)$ is an unspecified baseline rate function. We have two remarks regarding the models: First, our model formulations only include external, time-independent covariates. This is sufficient for our purpose, as we will later focus on a randomized clinical trial setting with the treatment indicator being the only covariate. However, it should be mentioned that all these models could also accommodate time-dependent covariates. Second, Mao and Lin actually present inference for a more general endpoint that is defined as a weighted composite process $N_i(t) = a_1 N_{i1}(t) + a_2 N_{i2}(t)$ with arbitrary weights a_1 and a_2 reflecting the severity of the respective outcome. As already introduced before, in our article we will focus on the composite counting process $N_i(t) = N_{i1}(t) + N_{i2}(t)$, that is, an equal weighting with $a_1 = a_2 = 1$.

The composite event rate considered in the LWYY model differs from that of the Mao-Lin model in being defined conditional on being alive. For that reason the respective rates are often denoted as conditional (LWYY) and marginal (Mao-Lin). However, in the presence of a frailty variable (as will be the case later on) these terms may cause confusion, as none of the rates is defined conditional on a frailty term meaning that each rate is marginal with respect to the frailty. Therefore we will avoid that linguistic distinction between conditional and marginal rates and refer to the respective rates as the Cox, the LWYY and the Mao-Lin rate subsequently.

Since all models considered here are semiparametric, parameter estimates are obtained by solving the equation $U(\beta) = 0$ with

$$U(\beta) = \sum_{i=1}^{n} \int_0^{\infty} Y_i(t) \left[X_i - \frac{\sum_{j=1}^{n} Y_j(t) X_j \exp(\beta' X_j)}{\sum_{j=1}^{n} Y_j(t) \exp(\beta' X_j)} \right] \mathrm{d}N_i(t) \tag{3}$$

being the model-specific estimating function. Here $Y_i(t)$ denotes an appropriate at-risk indicator. The latter is given as $Y_i(t) = I(D_i \geq t, T_{i1} \geq t, C_i \geq t)$ in the Cox model and as $Y_i(t) = I(D_i \geq t, C_i \geq t)$ in the LWYY model. The resulting estimators will be denoted as $\widehat{\beta}^{Cox}$ and $\widehat{\beta}^{LWYY}$. Following the same logic as for the Cox model and for the LWYY model, the

at-risk indicator in the Mao-Lin model ought to be $Y_i(t) = I(C_i \geq t)$. This would of course result in an unbiased estimator under the modeling assumptions of Mao and Lin. However, such estimator is in general not applicable in practice: Since nonadministrative censoring has to be anticipated, the censoring status of patients dropping out due to death is generally unknown afterward. For that reason Mao and Lin make use of the inverse probability of censoring weighting (IPCW) technique and replace the at-risk indicator in the estimating function (3) by the estimate $\widehat{w}_i(t)$ of an appropriate weighting function $w_i(t)$, which is given by

$$w_i(t) = \frac{I(C_i \geq D_i \wedge t) S_{C_i}(t|X_i)}{S_{C_i}(D_i \wedge C_i \wedge t|X_i)} = \begin{cases} 1 & \text{if } t \leq D_i \leq C_i \quad \text{or} \quad t \leq C_i \leq D_i \\ \dfrac{S_{C_i}(t|X_i)}{S_{C_i}(D_i|X_i)} & \text{if } D_i < t \leq C_i \quad \text{or} \quad D_i \leq C_i < t \\ 0 & \text{if } C_i < t \leq D_i \quad \text{or} \quad C_i < D_i \leq t. \end{cases} \tag{4}$$

Here $S_{C_i}(t|X_i)$ denotes the survival function of the censoring distribution conditional on covariates. Replacing $S_{C_i}(t|X_i)$ by an appropriate estimate $\hat{S}_{C_i}(t|X_i)$, which may be either be the Kaplan-Meier estimator (in case of X_i discrete) or the Breslow estimator (relying on a Cox model), results in the practically applicable weighting function estimate $\widehat{w}_i(t)$. The estimator defined by solving $U(\beta) = 0$ with the estimating function given as

$$U(\beta) = \sum_{i=1}^{n} \int_0^{\infty} \widehat{w}_i(t) \left[X_i - \frac{\sum_{j=1}^{n} \widehat{w}_j(t) X_j \exp(\beta' X_j)}{\sum_{j=1}^{n} \widehat{w}_j(t) \exp(\beta' X_j)} \right] \mathrm{d}N_i(t) \tag{5}$$

will be referred to as the Mao-Lin estimator $\widehat{\beta}^{Mao-Lin}$ and was shown to be consistent for the unknown parameters in the respective model.[23]

The Mao-Lin approach differs from that of LWYY with respect to the risk sets that are defined as

$$\mathcal{R}^{Mao-Lin}(t) = \sum_{i=1}^{n} \widehat{w}_i(t) \quad \text{and} \quad \mathcal{R}^{LWYY}(t) = \sum_{i=1}^{n} I(D_i \geq t, C_i \geq t) \tag{6}$$

at time point t, respectively. In the Mao-Lin estimator a patient is fully counted to the risk set (weight $\widehat{w}_i(t) = 1$) up to the end of his follow-up at time $V_i = D_i \wedge C_i$. If the follow-up is terminated due to censoring, a patient is not considered (weight $\widehat{w}_i(t) = 0$) in the risk set for later timepoints. These properties comply with the LWYY estimator. But in opposite to the latter, patients dropping out due to death afterward still partially contribute to the risk set of later timepoints in the Mao-Lin estimator. The "after-death weight" at time point t is given by the estimated conditional probability not yet to be censored at time t given the known fact that censoring did not occur up to death, that is, $\widehat{w}_i(t) = \hat{P}(C_i > t|X_i, C_i > D_i) = \hat{S}_{C_i}(t|X_i)/\hat{S}_{C_i}(D_i|X_i)$. Hence the risk sets of the Mao-Lin and the LWYY approach will be very similar for early timepoints and increasingly deviate from each other with time due to mortality, with Mao-Lin risk sets being generally larger than those of LWYY.

In the approaches of LWYY and Mao-Lin, estimates of the respective cumulative rate functions

$$\begin{aligned} R_i^{LWYY}(t|X_i) &= \int_0^t r_i^{LWYY}(s|X_i)\, \mathrm{d}s \\ R_i^{Mao-Lin}(t|X_i) &= \int_0^t r_i^{Mao-Lin}(s|X_i)\, \mathrm{d}s \end{aligned} \tag{7}$$

may serve as a graphical illustration of the data. Thereby $R_i^{LWYY}(t|X_i) \geq R_i^{Mao-Lin}(t|X_i)$ holds for all t. The cumulative Mao-Lin rate is the mean composite event number conditional on covariates, that is, $R_i^{Mao-Lin}(t|X_i) = E[N_i(t)|X_i]$, and thus has a clear clinical interpretation. In addition, in the Mao-Lin model proportionality of covariate effects on the rate level directly translates to the level of the mean function. For that reason it is also referred to as proportional means model. In contrast, the cumulative LWYY rate does in general not coincide with the mean recurrent event number of survivors conditional on covariates, that is, with $E[N_i(t)|X_i, D_i \geq t]$. Hence $R_i^{LWYY}(t|X_i)$ lacks interpretability. Nonparametric and semiparametric estimates for the cumulative LWYY rate are derived by the Nelson-Aalen and the Breslow methods, respectively.[17,32] The corresponding estimates for the cumulative Mao-Lin rate are given in analogous manner, just with

a different definition of the risk sets (as already discussed), and we refer to Mao and Lin[23] as well as Ghosh and Lin[33] for details.

4 | DERIVATION OF ESTIMATES FOR THE COMPOSITE ENDPOINT WITH CORRELATED COMPONENTS

In this section we introduce the methodological foundations that will later on be used to explain the behavior of estimates for the composite endpoint in the presence of a correlation between its components induced by common unobserved risk factors. A positive correlation between the outcomes HFH and CVD is very plausible from the clinical perspective and has already been verified in major HF trials.[12] In that context the term "correlation" means, broadly speaking, that patients with a high (low) HFH rate also have a high (low) CVD rate. Consequently patients with a dense sequence of HFHs during their follow-up tend to drop out earlier (due to CVD) than patients with fewer HFHs. To investigate the consequences of such correlation for the composite endpoint estimates obtained from an analysis with the Cox model, the LWYY model and the Mao-Lin model, we adopt the joint frailty model[19] as the data-generating process and derive the asymptotic estimates, called least false parameters, in that setting. The joint frailty model accounts for both the correlation between the HFH and CVD components and for separate, component-specific treatment effects. It has already been applied in some post-hoc analyses of HF trials.[9,12,15] Following we will first introduce the joint frailty model and show thereafter, how least false parameter theory allows to derive asymptotic composite treatment effect estimates resulting from an analysis with the Cox model, the LWYY model and the Mao-Lin model in data-generating scenarios that are parameterized through the joint frailty model.

4.1 | Joint frailty model

Conditional on being alive, the patient-specific rates for recurrent and terminal events might not only depend on the observed covariates, but also on unknown risk factors affecting both outcomes. Such unexplained heterogeneity can be modeled by a positive valued random variable Z_i (called frailty) acting multiplicatively on the outcome-specific event rates. Within the joint frailty model,[19] these rates are defined and modeled as

$$\begin{aligned}\lambda_{i1}(t|H_i(t), X_i, Z_i)\ \mathrm{d}t &= E[\mathrm{d}N_{i1}(t)|H_i(t), X_i, Z_i, D_i \geq t] = Z_i \lambda_{10}(t) \exp(\beta_1' X_i)\ \mathrm{d}t,\\ \lambda_{i2}(t|H_i(t), X_i, Z_i)\ \mathrm{d}t &= E[\mathrm{d}N_{i2}(t)|H_i(t), X_i, Z_i, D_i \geq t] = Z_i^{\gamma} \lambda_{20}(t) \exp(\beta_2' X_i)\ \mathrm{d}t.\end{aligned} \tag{8}$$

Here $\beta_1 = (\beta_{11}, \ldots, \beta_{1p})'$ and $\beta_2 = (\beta_{21}, \ldots, \beta_{2p})'$ are the parameter vectors of the p covariates contained in X_i. Henceforth we will denote the target rates of the joint frailty model as (outcome-specific) "conditional rates", as they are—in contrast to the composite event rates previously presented in Section 3—defined conditional on a frailty variable. For the frailty Z_i, commonly a gamma or a lognormal distribution with variance $Var[Z_i] = \theta$ and, in order to avoid identifiability problems, with a fixed mean $E[Z_i] = 1$ is adopted. Furthermore, the frailty is modeled as being independent of covariates X_i and of the censoring time C_i. Censoring is additionally assumed to be noninformative. The association parameter $\gamma \in \mathbb{R}$ is relaxing the assumption of a common shared frailty for recurrent and terminal events. Depending on the specification of the conditional baseline rates $\lambda_{10}(t)$ and $\lambda_{20}(t)$, semiparametric,[19] quasi-semiparametric,[34] and parametric[35,36] techniques were suggested for parameter estimation.

4.2 | Estimates for the composite endpoint in joint frailty scenarios

Henceforth we will assume the joint frailty model to be the data-generating process. Later on it will be shown that this generally implies a violation of the proportional-rates assumption within all the aforementioned semiparametric analysis models for the composite endpoint. Therefore covariate effects will in general be time-dependent on the level of the respective composite event rates. This is raising the question, how the parameters in the joint frailty model translate into the composite effect estimated by the Cox model, the LWYY model and the Mao-Lin model. Thorough asymptotic theory for estimation in misspecified Cox models has been derived by Struthers and Kalbfleisch[26] and Hjort.[27] Lin et al[17] generalized these results by showing that each estimator being defined through an estimating function in the form of

Formula (3) will converge to a certain least false parameter (vector), which can be identified as the solution of $g(\beta) = 0$ with

$$g(\beta) = \int_0^\infty E\left[Y_i(t)\left\{X_i - \frac{E[Y_i(t)X_i \exp(\beta' X_i)]}{E[Y_i(t)\exp(\beta' X_i)]}\right\} \mathrm{d}N_i(t)\right]. \tag{9}$$

Here $E[\cdot]$ denotes an expectation taken with respect to the at-risk process $Y_i(t)$, the covariates X_i and the counting process $N_i(t)$. Importantly, this result applies to arbitrary dependence structures between $Y_i(t)$, X_i and $N_i(t)$. It can thus be used to study the behavior of any estimator relying on an estimating function of type (3) under arbitrary misspecifications—misspecifications related to joint frailty scenarios are only one specific application. For a single binary covariate $X_i \sim Bin(1, p)$, we can show that Formula (9) can be written as the least false parameter function

$$g(\beta) = \int_0^\infty \frac{\bar{y}_i^{(0)}(t)\bar{y}_i^{(1)}(t)}{\bar{y}_i^{(0)}(t) + \bar{y}_i^{(1)}(t)\exp(\beta)} \left[E[\mathrm{d}N_i(t)|X_i = 1, Y_i(t) = 1] - E[\mathrm{d}N_i(t)|X_i = 0, Y_i(t) = 1]\exp(\beta)\right]. \tag{10}$$

Here $\bar{y}_i^{(k)}(t) = P(X_i = k, Y_i(t) = 1)$ denotes the probability that a patient belongs to treatment group $k \in \{0, 1\}$ and is still at risk at time t. A derivation of that finding is given in Appendix A. We will use this finding to derive the least false parameters (ie, the asymptotic composite treatment effect estimates) resulting from an analysis with the Cox model, the LWYY model and the Mao-Lin model in joint frailty scenarios with a single binary covariate, reflecting the situation of a two-armed randomized controlled trial. These least false parameters are defined as the asymptotic limit of the respective estimators, that is,

$$\begin{aligned} \widehat{\beta}^{Cox} &\xrightarrow{n\to\infty} \bar{\beta}^{Cox}, \\ \widehat{\beta}^{LWYY} &\xrightarrow{n\to\infty} \bar{\beta}^{LWYY}, \\ \widehat{\beta}^{Mao-Lin} &\xrightarrow{n\to\infty} \bar{\beta}^{Mao-Lin}. \end{aligned} \tag{11}$$

First the appropriate, model-specific at-risk indicators have to be plugged in Formula (10). This is straightforward for the Cox model and the LWYY model, but needs some explanation in case of the Mao-Lin model: Even though the estimating function of the Mao-Lin model is defined with a weighting function estimate $\widehat{w}_i(t)$ instead of an at-risk indicator $Y_i(t)$, the corresponding least false parameter can be identified by choosing $Y_i(t) = I(C_i \geq t)$ in Formula (10). This is due to the fact that

$$E[\widehat{w}_i(t)|X_i] = E[w_i(t)|X_i] = E[I(C_i \geq t)|X_i] = S_{C_i}(t|X_i) \tag{12}$$

(see Mao and Lin[23]). As a consequence, the (nonapplicable) estimator defined through the at-risk indicator $Y_i(t) = I(C_i \geq t)$ and the (applicable) estimator of Mao and Lin defined by means of the weighting function estimate $\widehat{w}_i(t)$ converge toward the same limit.

Due to independent censoring, the rate $E[\mathrm{d}N_i(t)|X_i, Y_i(t) = 1]$ appearing in the least false parameter function (10) actually coincides with the respective model-specific rate defined in (2):

$$E[\mathrm{d}N_i(t)|X_i, Y_i(t) = 1] = \begin{cases} E[\mathrm{d}N_i(t)|X_i, D_i \geq t, T_{i1} \geq t] = r_i^{Cox}(t|X_i)\,\mathrm{d}t, & \text{if } Y_i(t) = I(D_i \geq t, T_{i1} \geq t, C_i \geq t) \\ E[\mathrm{d}N_i(t)|X_i, D_i \geq t] = r_i^{LWYY}(t|X_i)\,\mathrm{d}t, & \text{if } Y_i(t) = I(D_i \geq t, C_i \geq t) \\ E[\mathrm{d}N_i(t)|X_i] = r_i^{Mao-Lin}(t|X_i)\,\mathrm{d}t, & \text{if } Y_i(t) = I(C_i \geq t) \end{cases}. \tag{13}$$

Next we calculate the components of Formula (10) for each composite treatment effect estimator in the joint frailty setting. Table 2 gives an overview on the results. Detailed derivations, including explicit formulas for survival functions and conditional expectations appearing in Table 2, are outsourced to Appendix B. Essentially all these formulas show that the target rates of the three analysis models for the composite endpoint are in general nonproportional in joint frailty scenarios.

TABLE 2 Components to be specified within the function $g(\beta)$, given in Formula (10), in order to derive least false parameters of the Cox, LWYY, and Mao-Lin estimator for the composite endpoint in joint frailty scenarios

Model	$\bar{y}_i^{(k)}(t)$	$E\left[dN_i(t)\middle\vert X_i = k, Y_i(t) = 1\right]/dt$
Cox	$S_{D_i \wedge T_{i1}}(t\vert X_i = k)$	$E\left[Z_i\middle\vert X_i = k, D_i \geq t, T_{i1} \geq t\right] \lambda_{10}(t)\exp(\beta_1 k)$
	$\times S_{C_i}(t\vert X_i = k)$	$+ E\left[Z_i^{\gamma}\middle\vert X_i = k, D_i \geq t, T_{i1} \geq t\right] \lambda_{20}(t)\exp(\beta_2 k)$
LWYY	$S_{D_i}(t\vert X_i = k)$	$E\left[Z_i\middle\vert X_i = k, D_i \geq t\right] \lambda_{10}(t)\exp(\beta_1 k)$
	$\times S_{C_i}(t\vert X_i = k)$	$+ E\left[Z_i^{\gamma}\middle\vert X_i = k, D_i \geq t\right] \lambda_{20}(t)\exp(\beta_2 k)$
Mao-Lin	$S_{C_i}(t\vert X_i = k)$	$S_{D_i}(t\vert X_i = k)E\left[Z_i\middle\vert X_i = k, D_i \geq t\right] \lambda_{10}(t)\exp(\beta_1 k)$
		$+ S_{D_i}(t\vert X_i = k)E\left[Z_i^{\gamma}\middle\vert X_i = k, D_i \geq t\right] \lambda_{20}(t)\exp(\beta_2 k)$

TABLE 3 Fit of the Cox model, the LWYY model and the Mao-Lin model for the composite endpoint (HFHs + CVD) and fit of a parametric joint gamma frailty model with constant conditional baseline rates (annual time scale) for the component-specific outcomes in the PARADIGM-HF trial

Model	Endpoint	Parameter	Estimate (95% CI)
Cox	Composite (time to first HFH/CVD)	$\exp(\beta)$	0.798 [0.731, 0.871]
LWYY	Composite (recurrent HFHs + CVD)	$\exp(\beta)$	0.789 [0.711, 0.876]
Mao-Lin	Composite (recurrent HFHs + CVD)	$\exp(\beta)$	0.805 [0.729, 0.889]
Joint frailty	Recurrent HFHs	$\exp(\beta_1)$	0.755 [0.657, 0.868]
	CVD	$\exp(\beta_2)$	0.791 [0.693, 0.902]
	—	γ	0.632 [0.556, 0.707]
	—	θ	5.200 [4.748, 5.652]
	—	$\lambda_{10}(t)$	0.158 [0.143, 0.173]
	—	$\lambda_{20}(t)$	0.136 [0.123, 0.150]

5 | CLINICAL EXAMPLE PART II: THE PARADIGM-HF TRIAL

In this section, we compare the composite endpoint estimates from the Cox model, the LWYY model and the Mao-Lin model obtained for the PARADIGM-HF data. In addition, we present the estimates of a parametric joint gamma frailty model with constant conditional baseline rates. For the sake of simplicity, we did not adjust for covariates or region, in contrast to the analyses published by Mogensen et al.[9] Results of the respective model fits are summarized in Table 3.

All methods for the composite endpoint yield very similar estimates for the treatment effect, close to a rate ratio of 0.8. The largest effect size results from the LWYY model, the smallest one from the Mao-Lin model—however, the differences are minimal. Moreover, all approaches nearly coincide in the length of the confidence interval. The joint frailty model yields a conditional rate ratio estimate of 0.791 for CVD and of 0.755 for recurrent HFHs. Due to the larger number of composite events, each of the outcome-specific treatment effects estimated by the joint frailty model has a greater uncertainty than the treatment effects estimated by the three semiparametric models for the composite endpoint. The relatively large frailty variance estimate indicates the correlation between HFHs and CVD in the trial. In addition, the estimate of the association parameter shows that unexplained heterogeneity is larger with respect to recurrent HFHs than to CVD. This is in line with other HF trials where the joint frailty model has been applied.[12]

Figure 1 shows semiparametric estimates of the cumulative LWYY and Mao-Lin rates for the PARADIGM-HF data. We clearly see that both models demonstrate the superiority of sacubitril/valsartan, but target different rates, as introduced in Section 3. In particular, the cumulative rate in the Mao-Lin model can be interpreted as an estimate for the expected number of composite events up to a certain time point, while the cumulative rate in the LWYY model is in general not interpretable unless HFHs and CVD are uncorrelated (see Cook and Lawless[16]). In contrast to the LWYY model, in the Mao-Lin model an observed CVD does not lead to a patient's exclusion from the risk set. Hence the cumulative Mao-Lin rate is smaller than the cumulative LWYY rate. However, relative treatment effects resulting from the LWYY model and the Mao-Lin model are quite similar, as seen in Table 3.

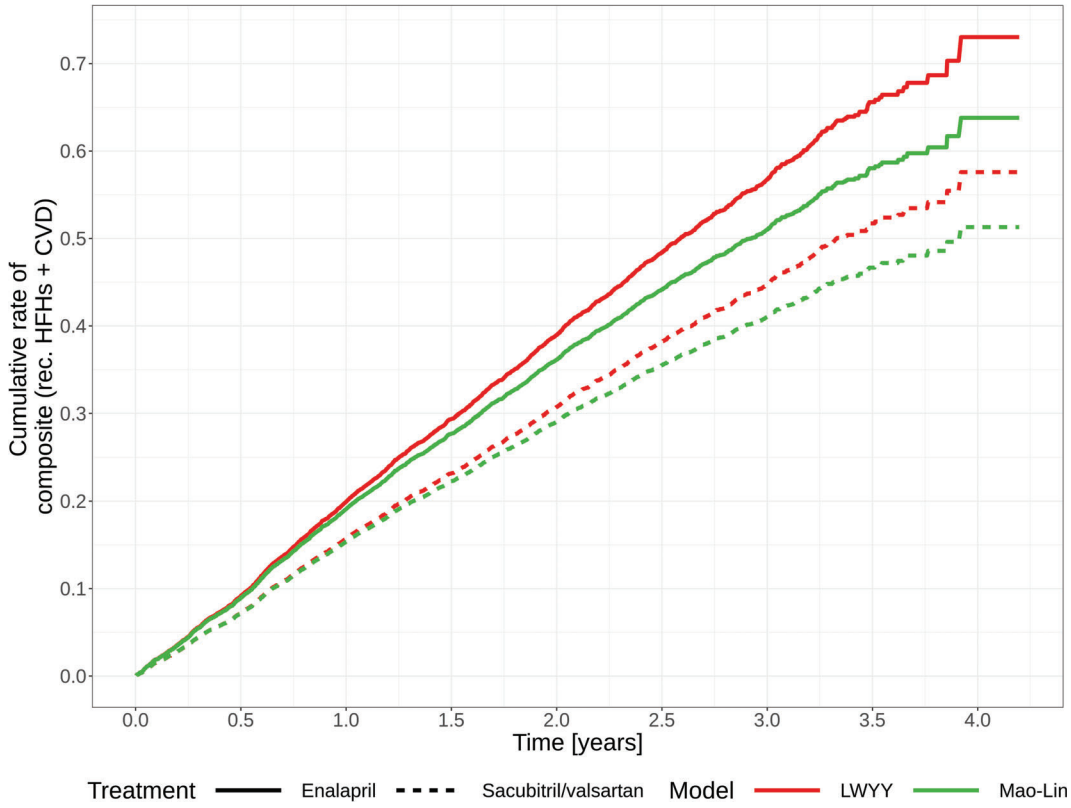

FIGURE 1 Semiparametric estimates of the cumulative LWYY and Mao-Lin rates in the PARADIGM-HF trial, both for the treatment (sacubitril/valsartan) and the control (enalapril) group [Colour figure can be viewed at wileyonlinelibrary.com]

6 | ANALYSIS OF ESTIMATES FOR THE COMPOSITE ENDPOINT WITH CORRELATED COMPONENTS

In the following we will investigate the behavior of the Cox, the LWYY and the Mao-Lin estimator for a binary, one-to-one randomized covariate $X_i \sim Bin(1, 0.5)$ in different joint frailty scenarios with a gamma-distributed frailty term. Asymptotic results are derived numerically by using the formulas presented in Section 4 and finite sample properties are studied with Monte Carlo simulations. In the description of our findings we throughout avoid the term "bias", since that would require to identify a single "true" effect on the composite, which is not defined in the joint frailty scenarios. All analyses in this section are performed in R software[37] (version 3.6.1). The code can be obtained from the corresponding author upon request.

6.1 | Numerical results (Figures 2 and 3)

Given a certain parameterization of a joint gamma frailty model, Table 2 allows to derive the time course of the composite rate ratios targeted by the Cox, the LWYY, and the Mao-Lin model. For a certain censoring scheme the asymptotic rate ratio estimates can then be derived by solving the least false parameter equation $g(\beta) = 0$ (see Formula (10)). In this subsection we present analytical findings for selected joint frailty parameterizations and censoring schemes.

6.1.1 | Consequences of varying the correlation and the trial duration (Figure 2)

Objectives:

We aim to evaluate how the time course of the composite rate ratios targeted by the three considered analysis models is affected by the degree of correlation between the outcomes (ie, the frailty variance θ). In addition, we investigate how

the asymptotic rate ratio estimate of the respective model depends on the interplay between the time course of the target rate ratio and the trial duration (ie, the censoring scheme).

Methods:

We focus on two joint frailty scenarios, one with no correlation ($\theta = 0$) and one with existing correlation between the outcomes ($\theta = 5.2$, as observed in PARADIGM-HF). The remaining model parameters correspond to the joint frailty estimates of the PARADIGM-HF data and are kept fixed in both scenarios (see Table 3: $\gamma = 0.63$, $\lambda_{10}(t) = 0.158$, $\lambda_{20}(t) = 0.136$, $\exp(\beta_1) = 0.755$, $\exp(\beta_2) = 0.791$). Using these parameterizations, we derive the composite rate ratios $r_i(t|X_i = 1)/r_i(t|X_i = 0)$ within the time interval $[0, 2.5]$ years by using the terms given in Table 2 (column 3) as nominator and denominator. For each time point t within that interval, we additionally investigate the composite rate ratio estimate $\exp(\overline{\beta})$ arising from a mixed censoring scheme with administrative censoring at time point t (probability $1 - 0.08t$) and uniformly distributed random censoring on $[0, t]$ (probability $0.08t$). Therefore we derive the least false parameter $\overline{\beta}$ of the respective estimator as the solution of $g(\beta) = 0$ (see Formula (10)). As an example, for $t = 2.5$ years the estimate resulting from a trial with 80% administrative censoring at $t = 2.5$ (ie, the trial duration) and 20% random censoring on $[0, 2.5]$ is computed. This would yield a median follow-up duration of 2.25 years (27 months, as in PARADIGM-HF), if the follow-up could only be terminated by censoring.

Results:

The time course of composite rate ratios and rate ratio estimates for both joint frailty scenarios is shown in Figure 2. In the absence of correlation ($\theta = 0$) the rates targeted by the Cox model and the LWYY model coincide. Moreover, these rates are constant (and thus of course proportional). The constancy of the rates does not hold in general, but applies here due to the constant outcome-specific, conditional baseline rates within the joint frailty settings considered here. Of course, the proportionality of the rates implies constant rate ratios over time. If the outcomes are correlated ($\theta = 5.2$), the rate ratios targeted by the Cox and LWYY approach are increasing and move toward the null hypothesis over time. This is the result of a selection process that can, exemplarily for the time-to-first-event rate (Cox), be explained as follows: Both in the treatment and in the control group, the time-to-first-event rates are decreasing over time as the frailest patients experience their first event early in time and healthier patients with lower conditional rates remain at risk longer. Due to the treatment's overall protective effect ($\exp(\beta_1) < 1$ and $\exp(\beta_2) < 1$), this selection proceeds faster in the control group, leading to a growing imbalance of frailty over time between the two cohorts at risk. Hence the decrease in the rate proceeds faster under control compared with treatment, resulting in an increase of the rate ratio over time. A similar explanatory pattern applies to the LWYY approach, with the important difference that patients remain at risk after a hospitalization. This has two implications in the scenario with correlated outcomes: First, the nonproportionality of the rates is exclusively induced by the protective treatment effect $\exp(\beta_2) < 1$ on mortality and not, as in the Cox approach, both by $\exp(\beta_1) < 1$ and $\exp(\beta_2) < 1$. Second, frail patients drop out faster in the time-to-first-event approach, leading to a faster evolution of the imbalance in frailty between the cohorts at risk. This manifests in a stronger increase of the Cox rate ratio compared with the LWYY rate ratio.

The Mao-Lin approach takes a special role, as the rate ratio is not only increasing in the presence of correlation but even in the scenario with uncorrelated outcomes. This is due to the fact that patients remain in the risk set up to censoring and are not removed in case of preceding death. First, we will have a look at the mechanisms in the absence of correlation: Due to death, event numbers (nominator of the rate) decrease over time. But concurrently the number of patients at risk (denominator of the rate) remains stable, since patients are still counted to the risk set after death. As a result, the Mao-Lin rate is decreasing over time in both groups. Of course, prolonged survival in the treatment group ($\exp(\beta_2) < 1$) comes along with a longer susceptibility-period for recurrent events. Consequently the Mao-Lin event rate in the treatment group is decreasing slower than in the control group, explaining the increase of the rate ratio over time. If the outcomes are correlated, that increase is more pronounced.

Time-dependent rate ratios raise the question what is actually estimated by the estimators derived by the Cox analysis, the LWYY analysis and the Mao-Lin analysis. An exact, but rather technical answer is given by the least false parameter theory presented in Section 4. Reduced to its essentials, the central message of that theory is: The rate ratio estimate after study time t is a weighted average of the rate ratio function on $[0, t]$, with weights determined by the censoring scheme. In particular, the censoring scheme includes the duration of the trial. This is confirmed by the rate ratio estimates shown in Figure 2. In case of uncorrelated outcomes, the estimates resulting from the time-to-first-event and the LWYY approach are not affected by the censoring scheme. As already mentioned, this can be explained by the constant conditional rates in our setting. However, if the outcomes are correlated, the Cox and LWYY estimates are sensitive to the censoring scheme. The same applies to the Mao-Lin estimates—even in the absence of correlation. All estimates that

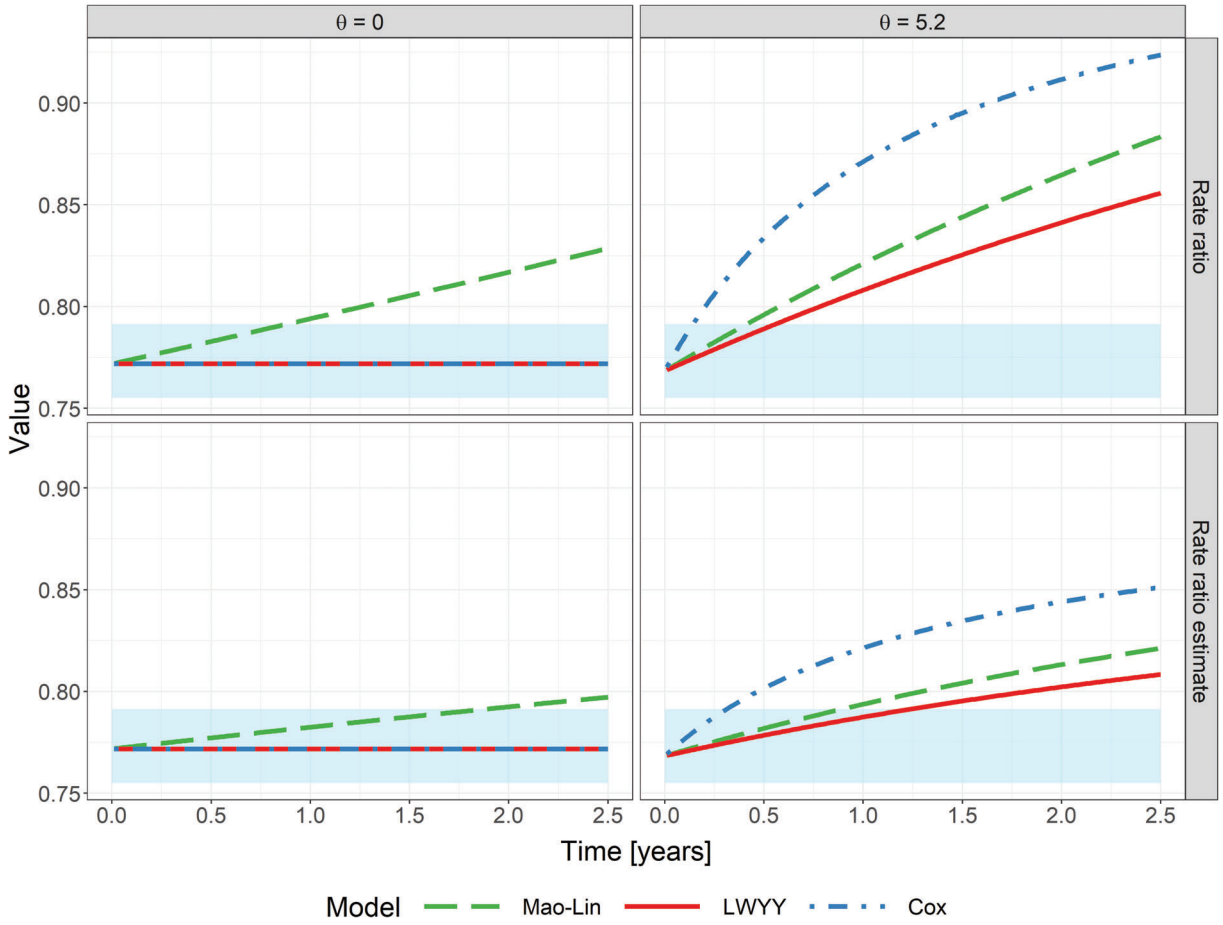

FIGURE 2 Rate ratios and asymptotic rate ratio estimates for the composite endpoint in joint gamma frailty models with frailty variances $\theta = 0$ and $\theta = 5.2$. Motivated by the PARADIGM-HF results, the remaining joint frailty parameters are fixed as $\exp(\beta_1) = 0.755$, $\exp(\beta_2) = 0.791$, $\lambda_{10}(t) = 0.158$, $\lambda_{20}(t) = 0.136$, and $\gamma = 0.63$. The shaded area shows the interval between the conditional rate ratios $\exp(\beta_1)$ (recurrent events) and $\exp(\beta_2)$ (terminal event). The respective composite rate ratio and rate ratio estimate at time point t is given by $r_i(t|X_i = 1)/r_i(t|X_i = 0)$ and $\exp(\overline{\beta})$, where $\overline{\beta}$ is the least false parameter obtained in a trial with the following censoring scheme: Administrative censoring at time point t (probability $1 - 0.08t$) and uniformly distributed random censoring on $[0, t]$ (probability $0.08t$). For example, for $t = 2.5$ the censoring scheme is given by 80% administrative censoring at $t = 2.5$ and 20% random censoring on $[0, 2.5]$ [Colour figure can be viewed at wileyonlinelibrary.com]

are affected by the censoring scheme shift toward the null hypothesis with increasing trial durations, that is, treatment effect estimates become attenuated the longer the trial lasts. In particular, if the trial duration increases, the estimates are not even located in the interval $[\exp(\beta_1) = 0.755, \exp(\beta_2) = 0.791]$ determined by the component-specific conditional rate ratios, wherein an estimate for the composite naturally ought to be expected. The time-to-first-event approach shows the strongest and the LWYY approach the least attenuation with time. The censoring scheme underlying the estimates shown for the largest time point ($t = 2.5$) comes closest to that one of the PARADIGM-HF trial. Here the expected composite rate ratio estimates in the presence of correlation amount to 0.851 for the Cox model, 0.821 for the Mao-Lin model and 0.808 for the LWYY model. How deviations from the actually observed rate ratio estimates for the PARADIGM-HF data (see Table 3) may be explained, will be discussed in Section 7.

6.1.2 | Consequences of varying the correlation and the component-specific treatment effects (Figure 3)

Objectives:

We aim to investigate how the asymptotic rate ratio estimates for the composite endpoint depend on the interplay between the degree of correlation (ie, the frailty variance θ) and the outcome-specific, conditional rate ratios $\exp(\beta_1)$ and $\exp(\beta_2)$.

Methods:

We evaluate joint frailty scenarios with all parameter combinations of $\theta \in \{0, 5.2\}$, $\exp(\beta_1) \in \{0.755, 1, 1.324\}$ and $\exp(\beta_2) \in [0.5, 1.5]$. The remaining model parameters correspond to the joint frailty estimates of the PARADIGM-HF data and are kept fixed in all scenarios (see Table 3: $\gamma = 0.63$, $\lambda_{10}(t) = 0.158$, $\lambda_{20}(t) = 0.136$). Likewise the censoring scheme is kept fixed in all scenarios, namely, as a mixture of administrative censoring at time point $t = 2.5$ years (probability 0.8) and uniformly distributed random censoring on $[0, 2.5]$ (probability 0.2), coming close to the censoring pattern in PARADIGM-HF. For all parameterizations, we investigate the composite rate ratio estimate $\exp(\overline{\beta})$ by deriving the least false parameter $\overline{\beta}$ of the respective estimator as the solution of $g(\beta) = 0$ (see Formula (10)).

Results:

Our findings on the behavior of the estimates are illustrated in Figure 3. First, we start with the identification of parameter settings where estimates coincide:

- In case of no effects on both outcomes ($\exp(\beta_1) = \exp(\beta_2) = 1$) all three approaches consistently yield a composite rate ratio estimate of $\exp(\overline{\beta}) = 1$, no matter if the outcomes are correlated or not. Although we generally try to avoid the term "bias", in that situation it is well justified to call the estimate "unbiased". None of the analysis models is misspecified in that situation, because the model-specific rates are equal between the treatment and control group and thus fulfill the proportionality assumption.
- The estimates of the Mao-Lin and the LWYY approach coincide and are not affected by a correlation of the outcomes if treatment has no effect on mortality ($\exp(\beta_2) = 1$). In both approaches, differences in mortality between the groups are a prerequisite for nonproportional rates. In case of equal mortality the rates targeted by both approaches coincide and are proportional—and the rate ratio is not affected by the degree of correlation between the outcomes.

In addition, the estimates of the Cox and LWYY approach coincide under uncorrelated outcomes ($\theta = 0$). However, this is again due to the constant conditional baseline rates in our joint frailty scenarios, which are leading to proportional rates for the composite. Equality of the Cox and LWYY estimates under uncorrelated outcomes does not hold for time-dependent conditional baseline rates.

Building on the findings regarding the time course of rate ratios described in Section 6.1.1, it should be stressed again: Proportional rates (ie, time-constant rate ratios) imply that the estimate is not affected by the censoring scheme, in particular by the trial duration. In our experience, most realistic scenarios in HF trials are given by settings with at least not harmful component-specific effects ($\exp(\beta_1) \leq 1$ and $\exp(\beta_2) \leq 1$) and with a correlation between the outcomes ($\theta > 0$). In these scenarios the rate ratio function of the time-to-first-event approach is only constant if treatment has no effect on both components ($\exp(\beta_1) = \exp(\beta_2) = 1$). In the approaches of Mao-Lin and LWYY constant rate ratios are only given if treatment has no effect on mortality ($\exp(\beta_2) = 1$). In all other scenarios considered as clinically relevant, rate ratios are time-dependent and their estimates will attenuate with trial duration.

Let us now put the emphasis on differences between the recurrent event approaches (LWYY, Mao-Lin). Both estimates are equal in case of $\exp(\beta_2) = 1$, but in scenarios with $\exp(\beta_2) \neq 1$ the Mao-Lin estimate is always shifted in the direction that favors the treatment regimen with higher mortality risks. As an example, we will have a look at scenarios with $\theta = 0$ and $\exp(\beta_1) = 0.755$ in Figure 3:

- In case of higher mortality risks in the control group ($\exp(\beta_2) = 0.791$):

$$\exp\left(\overline{\beta}^{Mao-Lin}\right) = 0.797 > 0.772 = \exp\left(\overline{\beta}^{LWYY}\right).$$

- In case of higher mortality risks in the treatment group ($\exp(\beta_2) = 1.264$):

$$\exp\left(\overline{\beta}^{Mao-Lin}\right) = 0.951 < 0.990 = \exp\left(\overline{\beta}^{LWYY}\right).$$

Both cases have in common that the more life-threatening treatment regimen appears in a better light with the Mao-Lin model than with the LWYY model. In other words: The Mao-Lin approach does neither appreciate a protective nor penalize a harmful treatment effect on mortality as much as the LWYY approach. In particular, a harmful treatment effect on mortality will rather be masked by the Mao-Lin approach. In the presence of correlation this behavior remains, but is less pronounced. Here estimates obtained from these approaches are quite similar in all the studied scenarios.

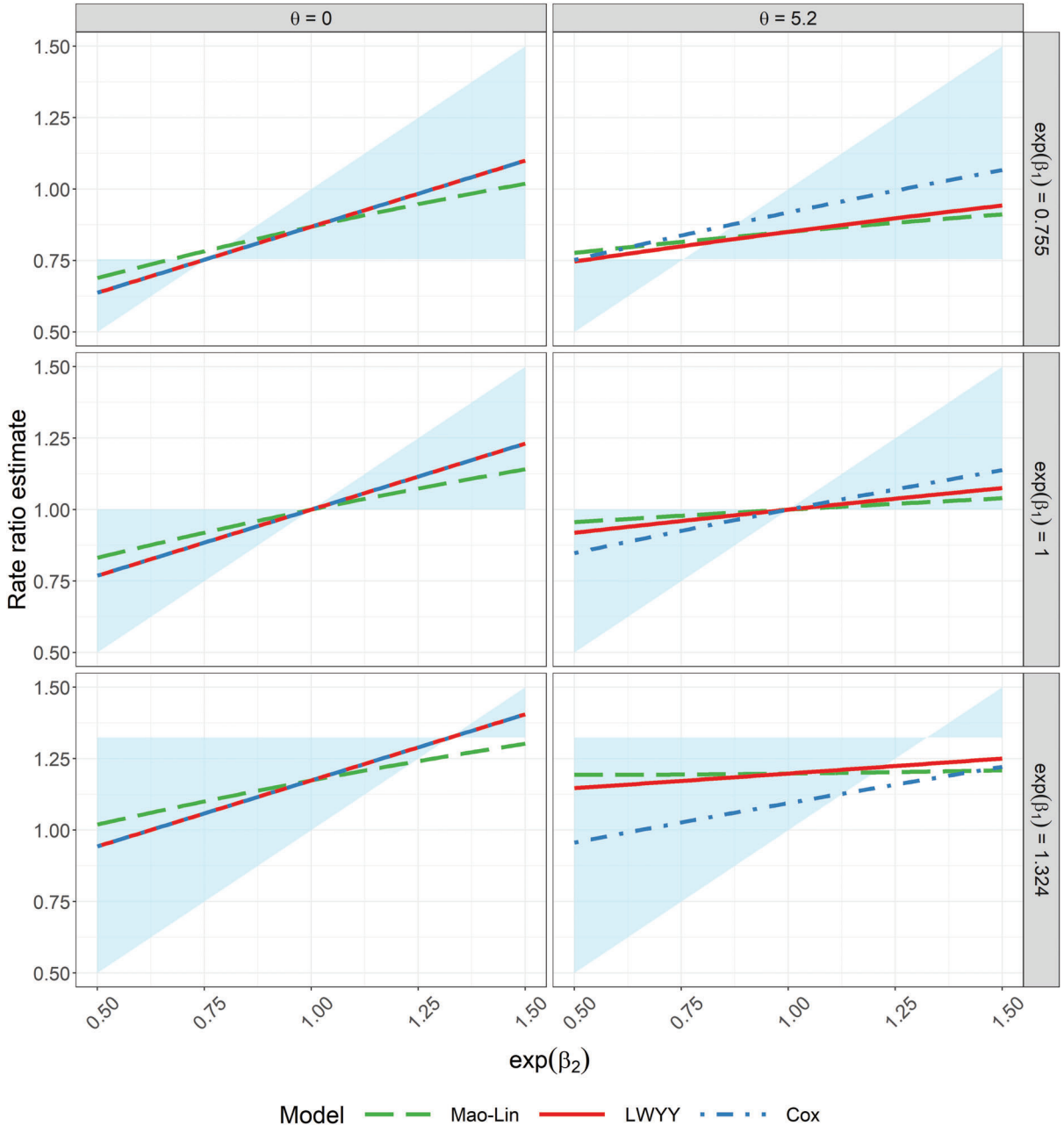

FIGURE 3 Asymptotic rate ratio estimates $\exp(\bar{\beta})$ for the composite endpoint in joint gamma frailty models with different frailty variances θ and conditional rate ratios $\exp(\beta_1)$ (recurrent events) and $\exp(\beta_2)$ (terminal event). The intervals between $\exp(\beta_1)$ and $\exp(\beta_2)$ in the respective scenario are represented by the shaded area. Motivated by the PARADIGM-HF results, the remaining joint frailty parameters are fixed as $\lambda_{10}(t) = 0.158$, $\lambda_{20}(t) = 0.136$ and $\gamma = 0.63$. The censoring scheme is given by administrative censoring at time point $t = 2.5$ years (probability 0.8) and uniformly distributed random censoring on $[0, 2.5]$ (probability 0.2) [Colour figure can be viewed at wileyonlinelibrary.com]

The definition of at-risk time within the Mao-Lin model is raising the question, whether a life-threatening treatment might even appear favorable just because early death is preventing further hospitalizations. To address that issue, we have a closer look at scenarios with no treatment effect on recurrent events ($\exp(\beta_1) = 1$) and a detrimental treatment effect on terminal events ($\exp(\beta_2) > 1$). Figure 3 suggests that the Mao-Lin approach still evaluates the treatment as harmful $\left(\exp\left(\bar{\beta}^{Mao-Lin}\right) > 1\right)$ in all studied settings, albeit showing the weakest estimate among the three methods. However, in all these settings conditional baseline rates are kept fixed as $\lambda_{10}(t) = 0.158$ for recurrent and $\lambda_{20}(t) = 0.136$ for terminal events (motivated by PARADIGM-HF). In Appendix C (Figure C1) we additionally investigate the consequences of

varying the conditional baseline rates in scenarios with $\exp(\beta_1) = 1$ and $\exp(\beta_2) = 1.264$. Indeed, if the recurrent event rate is sufficiently large, the amount of hospitalization events prevented by early death outweighs the amount of additional mortality events under treatment and thus the Mao-Lin approach favors the detrimental treatment $\left(\exp\left(\overline{\beta}^{Mao-Lin}\right) < 1\right)$, especially (but not only) in the presence of a correlation between the composite's components. Such an unintended inversion of the composite treatment effect estimate likewise applies to the LWYY approach, but in a slightly milder form and only in case of correlation.

6.2 | Simulation results (Table 4)

So far asymptotic, analytical results on rate ratio estimates were presented. In this subsection we investigate finite-sample properties of the Cox, the LWYY, and the Mao-Lin estimator for selected scenarios by a simulation study.

Objectives:

We aim to investigate finite-sample properties of the three estimators for the composite endpoint. In particular, we compare rate ratio estimates and the standard errors of the corresponding parameters. Moreover, rejection probabilities of the model-specific, 5% significance level tests on $H_0 : \exp(\beta) = 1$ are evaluated, where β denotes the parameter for the composite treatment effect in the respective analysis model.

Methods:

We evaluate a full factorial set of joint frailty scenarios with all parameter combinations of $\theta \in \{0, 5.2\}$, $\exp(\beta_1) \in \{0.755, 1, 1.324\}$ and $\exp(\beta_2) \in \{0.791, 1, 1.264\}$. The remaining model parameters correspond to the joint frailty estimates of the PARADIGM-HF data and are kept fixed in all scenarios (see Table 3: $\gamma = 0.63$, $\lambda_{10}(t) = 0.158$, $\lambda_{20}(t) = 0.136$). Likewise the censoring scheme is kept fixed in all scenarios, namely as a mixture of administrative censoring at time point $t = 2.5$ years (probability 0.8) and uniformly distributed random censoring on $[0, 2.5]$ (probability 0.2), coming close to the censoring pattern in PARADIGM-HF.

For the simulation we use the `simreccomp()` function of the R-package simrec[38] (version 1.0.0) to generate joint frailty datasets. For each scenario 10 000 datasets with each 1000 patients are generated, that is, 500 per group. Each of these datasets is evaluated (a) with the Cox model and the LWYY model using the `coxph()` function of the R-package survival[39] (version 3.1-7) and (b) with the Mao-Lin model using the `CompoML()` function whose R-Code is available on Lu Mao's website.[40] Within the fitting process for the Mao-Lin model we use the Cox model approach to estimate the inverse probability of censoring weights.

In all model fits we extract the parameter estimate, its standard error and the p-value of the Wald test on $H_0 : \exp(\beta) = 1$, where β denotes the parameter for the composite treatment effect in the respective analysis model. For the recurrent event methods (LWYY, Mao-Lin), model-based standard error estimates are used. These are per se robust. For the time-to-first-event approach, we use the robust sandwich estimate and not the model-based estimate for the standard error. Based on these data we calculate the following measures of evaluation for each model in each scenario:

- $\exp(\widehat{\beta})$: Empirical mean of the rate ratio estimates
- *SE*: Empirical standard error of $\widehat{\beta}$
- *SEE*: Empirical mean of the standard error estimates of $\widehat{\beta}$
- rp: Rejection probability, that is, the fraction of tests with $p < 0.05$

We additionally derive the least false parameter $\overline{\beta}$ of the respective estimator as the solution of $g(\beta) = 0$ (see Formula (10)) in order to compare the finite-sample rate ratio estimate $\exp(\widehat{\beta})$ with the asymptotically valid rate ratio estimate $\exp(\overline{\beta})$.

Results:

Table 4 is summarizing the results of the simulation study. The average composite rate ratio estimates $\exp(\widehat{\beta})$ of the simulation coincide with the estimates $\exp(\overline{\beta})$ that have been derived numerically based on least false parameter theory. Hence, asymptotic results already apply well for the studied sample size of 1000 patients. Larger sample sizes would not affect the estimates, but only reduce their standard errors. Since each of the considered standard error estimators is robust (ie, accounts for heterogeneity among the risk rates), the empirical standard error of the parameter estimates (*SE*) always

TABLE 4 Simulation results (10 000 simulated datasets, each with $n = 1000$ patients) comparing rate ratio estimates for the composite endpoint in joint gamma frailty models with different frailty variances θ and conditional rate ratios $\exp(\beta_1)$ (recurrent events) and $\exp(\beta_2)$ (terminal event)

Sim. setup		Mao-Lin					LWYY					Cox				
$\exp(\beta_1)$	$\exp(\beta_2)$	$\exp(\widehat{\beta})$	$\exp(\overline{\beta})$	*SE*	*SEE*	rp	$\exp(\widehat{\beta})$	$\exp(\overline{\beta})$	*SE*	*SEE*	rp	$\exp(\widehat{\beta})$	$\exp(\overline{\beta})$	*SE*	*SEE*	rp
(a) $\theta = 0$																
0.755	0.791	0.796	0.797	0.079	0.079	0.825	0.771	0.772	0.090	0.089	0.831	0.771	0.772	0.097	0.096	0.766
0.755	1.000	0.868	0.868	0.076	0.076	0.463	0.868	0.868	0.088	0.087	0.370	0.868	0.868	0.095	0.094	0.326
0.755	1.264	0.951	0.951	0.073	0.073	0.106	0.990	0.990	0.086	0.085	0.054	0.990	0.990	0.093	0.092	0.054
1.000	0.791	0.933	0.933	0.076	0.076	0.151	0.903	0.903	0.086	0.086	0.225	0.903	0.903	0.094	0.094	0.196
1.000	1.000	1.000	1.000	0.074	0.073	0.049	1.000	1.000	0.085	0.084	0.052	0.999	1.000	0.093	0.092	0.049
1.000	1.264	1.078	1.078	0.071	0.071	0.186	1.122	1.122	0.083	0.083	0.288	1.122	1.122	0.091	0.090	0.251
1.324	0.791	1.113	1.113	0.073	0.073	0.314	1.078	1.078	0.083	0.082	0.148	1.078	1.078	0.091	0.091	0.135
1.324	1.000	1.175	1.174	0.071	0.071	0.625	1.174	1.174	0.082	0.081	0.509	1.175	1.174	0.090	0.089	0.430
1.324	1.264	1.246	1.245	0.070	0.069	0.894	1.297	1.296	0.081	0.080	0.903	1.297	1.296	0.088	0.088	0.841
(b) $\theta = 5.2$																
0.755	0.791	0.822	0.821	0.145	0.144	0.280	0.809	0.808	0.151	0.150	0.294	0.850	0.851	0.132	0.131	0.236
0.755	1.000	0.851	0.850	0.141	0.140	0.215	0.851	0.850	0.148	0.147	0.200	0.918	0.918	0.130	0.129	0.106
0.755	1.264	0.885	0.884	0.137	0.136	0.148	0.902	0.900	0.145	0.144	0.116	0.998	0.999	0.128	0.127	0.051
1.000	0.791	0.981	0.982	0.145	0.143	0.054	0.966	0.967	0.151	0.149	0.060	0.939	0.938	0.130	0.128	0.079
1.000	1.000	0.999	1.000	0.141	0.140	0.051	0.999	1.000	0.148	0.146	0.051	1.001	1.000	0.128	0.127	0.053
1.000	1.264	1.021	1.021	0.137	0.136	0.052	1.040	1.040	0.145	0.144	0.056	1.076	1.074	0.126	0.125	0.088
1.324	0.791	1.196	1.195	0.144	0.143	0.242	1.177	1.177	0.150	0.149	0.194	1.038	1.038	0.125	0.126	0.054
1.324	1.000	1.200	1.198	0.141	0.140	0.253	1.199	1.198	0.147	0.146	0.235	1.094	1.095	0.124	0.124	0.109
1.324	1.264	1.206	1.204	0.138	0.137	0.275	1.227	1.226	0.145	0.144	0.300	1.163	1.163	0.123	0.123	0.228

Note: Motivated by the PARADIGM-HF results, the remaining joint frailty parameters are fixed as $\lambda_{10}(t) = 0.158$, $\lambda_{20}(t) = 0.136$ and $\gamma = 0.63$. The censoring scheme is given by administrative censoring at time point $t = 2.5$ years (probability 0.8) and uniformly distributed random censoring on $[0, 2.5]$ (probability 0.2). Here $\exp(\widehat{\beta})$ is the average rate ratio estimate obtained in the simulation, $\exp(\overline{\beta})$ is rate ratio estimate obtained numerically by least false parameter theory, *SE* is the empirical standard error of $\widehat{\beta}$, *SEE* is the empirical mean of the (robust) standard error estimates of $\widehat{\beta}$, and rp is the rejection probability of the Wald test on $H_0 : \exp(\beta) = 1$.

coincides with the empirical mean of the standard error estimates (*SEE*). Apparently, standard errors are increased by a correlation between the outcomes. Regarding the size of the standard errors, the following grading applies to all scenarios:

- In case of uncorrelated outcomes ($\theta = 0$): $SE(\widehat{\beta}^{Mao-Lin}) < SE(\widehat{\beta}^{LWYY}) < SE(\widehat{\beta}^{Cox})$
- In case of correlated outcomes ($\theta = 5.2$): $SE(\widehat{\beta}^{Cox}) < SE(\widehat{\beta}^{Mao-Lin}) < SE(\widehat{\beta}^{LWYY})$

The Mao-Lin approach has throughout slightly smaller standard errors than the LWYY approach, no matter if correlation is present between the outcomes. In addition, the time-to-first-event approach has the largest standard errors under uncorrelated outcomes, but the smallest standard errors in the presence of correlation. This is an important finding given the common opinion that recurrent event methods per se yield smaller standard errors than the time-to-first-event approach.

Finally we will have a look at the rejection probabilities (rp) of Wald tests (using the mentioned standard errors) for the null hypothesis $H_0 : \exp(\beta) = 1$. All approaches keep the type I error in case of no effects on both outcomes ($\exp(\beta_1) = \exp(\beta_2) = 1$), because no model is misspecified in that situation. In all other scenarios, rejection probabilities

result from the complex interplay of the estimate location and the standard error size. As there is no obvious pattern, we do not go into the details of each scenario here.

7 | DISCUSSION

In clinical heart failure trials efficacy is generally evaluated by using a composite endpoint consisting of HFHs and CVD.[1,2,6] In this article we systematically investigated the differences between estimates resulting from the traditional time-to-first-event analysis based on the Cox model[31] and two recurrent event analysis models for the composite, the prominent LWYY model[17] and the recently proposed Mao-Lin model.[23] Recurrent event models account for all HFHs and CVDs during the follow-up and are thus considered as an alternative to the time-to-first-event analysis.[7,10,11] For the comparison of these approaches we investigated clinically relevant scenarios, parameterized by joint frailty models[19] that account both for different component-specific treatment effects on HFHs and CVD and for the positive correlation between these two outcomes. Published comparisons between the time-to-first-event analysis and recurrent event methods in the heart failure context did so far neither investigate how the component-specific treatment effects translate into composite treatment effect estimates nor how these estimates are affected by the correlation between the two event processes (see, eg, Claggett et al,[10] Rauch et al[25]). Our article addresses these issues and thus contributes to a proper interpretation of different analysis approaches in clinical HF trials. This is what has been claimed by regulators as a prerequisite in order to seriously take recurrent event methods into consideration for the primary analysis.[7] We demonstrated that the composite endpoint estimates resulting from both recurrent event analysis models are quite similar and less adversely affected by a correlation between fatal and nonfatal event processes than the time-to-first-event analysis in realistic settings. Concurrently the Mao-Lin analysis was shown to be most insensitive to harmful effects on mortality among the three investigated approaches. In summary, our results clearly favor the LWYY approach to evaluate composite endpoints in future HF trials.

First, we showed that the proportionality assumption for the event rates imposed by the aforementioned analysis models for the composite endpoint is generally misspecified in the presence of correlated outcomes corresponding to a joint frailty model, that is, rate ratios are generally time-dependent. As a new methodological contribution, we demonstrated how least false parameter theory[17,26,28] can be used as a powerful tool to derive the respective asymptotic estimates in such scenarios numerically. We illustrated that the composite treatment effect estimates of the misspecified analysis models strongly depend on the censoring scheme, in particular on the trial duration. In clinically relevant scenarios, characterized by correlated outcomes and at least not harmful treatment effects on both components, all estimates are attenuated and increasingly shift toward the null hypothesis with trial duration. The estimate of the time-to-first-event analysis is by far most affected by that attenuation, whereas both recurrent event methods are more robust to the underlying misspecification. This result is clearly supporting the use of recurrent event methods in HF trials. It should be noted that the strong attenuation of the time-to-first-event estimate can only in parts be accounted to the correlation between HFHs and CVD itself. The driving force is rather the degree of unexplained heterogeneity in the patient-specific risk rates for each component, especially among HFHs, as indicated by results of Claggett et al.[10] However, in the joint frailty parameterization the actual correlation is not clearly separable from the variances of each component's unexplained heterogeneity. A clear separation of these quantities in the parameterization (eg, through using a bivariate frailty) would be desirable, but has only once been suggested in the literature, probably due to computational issues during model fitting.[36] To our knowledge the work of Tawiah et al[41] is the only one that makes use of a bivariate frailty with a certain correlation in the joint frailty modeling context—however, a published software implementation is missing.

Compared with the LWYY estimate, the Mao-Lin estimate is always slightly shifted to the direction that favors the more life-threatening treatment and is thus less sensitive to harmful effects on mortality. A harmful effect on mortality leads to more CVDs, but concurrently to less HFHs and is thus ambiguous with respect to the composite. In the Mao-Lin approach the avoidance of HFHs through death is less negatively graded than in the LWYY approach. In some (albeit clinically not very plausible) scenarios it might even happen that the Mao-Lin estimate is in favor of a treatment that has a harmful effect on mortality while being neutral regarding hospitalizations. Remarkably, such unintended inversion of the composite treatment effect estimate might also occur with the LWYY approach. However, in clinically relevant settings both recurrent event methods yield reasonable estimates and their differences are small. In their article, Mao and Lin obtained larger differences between analyses with the LWYY model and the Mao-Lin model in some simulated scenarios.[23] This can be explained as follows: First, in their investigated settings mortality risks during follow-up were assumed to be up to five times higher in the control group than in the treatment group, which is far off clinical reality in HF trials. Second, the outcomes were considered to be uncorrelated (at least in settings with treatment effect on mortality),

which might likewise not well reflect the disease process in heart failure. Among the two studied recurrent event methods, we recommend the LWYY approach, first and foremost because it is more sensitive to harmful effects on mortality. Second, there has been done quite some research related to the LWYY model on methods relevant for clinical trials such as sample size planning/reestimation[42,43] and group sequential designs.[44] All this research is missing for the Mao-Lin model. Third, the Mao-Lin model is not implemented in standard software and R-functions are only available on the author's website.[40] In favor of the Mao-Lin model it should be mentioned that higher weights could be assigned to terminal events to become more sensitive to harmful effects on mortality. The consequences of weighting events for the composite treatment effect estimate should therefore be further investigated, but goes beyond the scope of this article.

Our analytical findings are supported by empirical results from the PARADIGM-HF trial, a large trial on the efficacy of sacubitril/valsartan in patients with chronic HF and reduced ejection fraction.[8] Here each of the component-specific treatment effect estimates (obtained by a joint frailty model) is larger than the composite effect estimates obtained by the Cox model and the Mao-Lin model. The LWYY estimate for the composite endpoint shows the same tendency and nearly coincides with the weakest component-specific effect estimate, that is, that for CVD. This dilution of the composite treatment effect estimates in the presence of correlated components is completely in line with our analytical findings for joint frailty scenarios. However, the observed composite effect estimates in the PARADIGM-HF trial do not precisely match with those that would have been expected in a joint frailty scenario parameterized through the joint frailty estimates on the trial data. This may be explained by some violations of the joint frailty modeling assumptions: First, the conditional rates for each outcome may depend on the HFH history. As an example, it may be plausible that they are increasing with each HFH. Second, treatment effects on both outcomes may decrease over time (eg, due to drug discontinuation after first HFH, as often occurred in PARADIGM-HF), that is, the proportional rates assumption may already be violated on the patient's level. Both scenarios will manifest in an overestimation of the true frailty variance. The large estimate of the frailty variance for the PARADIGM-HF data may indicate that the degree of correlation between the outcomes is probably overestimated. This is supported by a recent publication from Balan and Putter,[45] who raised the important question, how well time-dependent treatment effects may be distinguished from unexplained heterogeneity in clustered survival data. A main result was that the frailty variance estimate is generally inflated in case of time-dependent treatment effects. The extent of overestimation is decreasing with the cluster sizes, that is, with the number of events per patient in our case. However, the authors did not specifically consider the joint frailty model in their investigations and it will be future work to characterize its estimates in such situations. Instead of a joint frailty parametrization for the data-generating process simulations could also be performed by bootstrapping data of historical trials. This has recently been shown to be a promising approach to generate realistic data of complex time-to-event processes.[46]

ACKNOWLEDGEMENTS

The authors thank the reviewers, the associate editor, and the editor for constructive comments which helped them to improve the manuscript. This research project was funded by the Deutsche Forschungsgemeinschaft (DFG project no. JA 1821/4-1).

CONFLICT OF INTEREST

The authors have declared no conflict of interest.

DATA AVAILABILITY STATEMENT

The data from the PARADIGM-HF trial (ClinicalTrials.gov Identifier: NCT01035255) that support the findings of this work are available from Novartis. Restrictions apply to the availability of these data, which were used under license for this article. Data are available at https://www.clinicalstudydatarequest.com(CSDR) with the permission of Novartis and the Independent Review Panel of CSDR.

ORCID

Gerrit Toenges https://orcid.org/0000-0002-9745-9437
Tobias Mütze https://orcid.org/0000-0002-4111-1941

REFERENCES

1. Hicks KA, Mahaffey KW, Mehran R, et al. 2017 cardiovascular and stroke endpoint definitions for clinical trials. *J Am Coll Cardiol*. 2018;71(9):1021-1034. https://doi.org/10.1016/j.jacc.2017.12.048
2. Zannad F, Garcia AA, Anker SD, et al. Clinical outcome endpoints in heart failure trials: a European society of cardiology heart failure association consensus document. *Eur J Heart Fail*. 2013;15(10):1082-1094. https://doi.org/10.1093/eurjhf/hft095

3. Packer M. Development and evolution of a hierarchical clinical composite end point for the evaluation of drugs and devices for acute and chronic heart failure. *Circulation*. 2016;134(21):1664-1678. https://doi.org/10.1161/CIRCULATIONAHA.116.023538
4. EMA's Committee for Medicinal Products for Human Use (CHMP) Guideline on multiplicity issues in clinical trials; 2017. https://www.ema.europa.eu/en/documents/scientific-guideline/draft-guide%line-multiplicity-issues-clinical-trials_en.pdf. Accessed June 23, 2020.
5. FDA's Office of Translational Sciences in the Center for Drug Evaluation and Research. Multiple Endpoints in Clinical Trials - Guidance for Industry; 2017. https://www.fda.gov/media/102657/download. Accessed June 23, 2020.
6. Anker SD, McMurray JJV. Time to move on from 'time-to-first': should all events be included in the analysis of clinical trials? *Eur Heart J*. 2012;33(22):2764-2765. https://doi.org/10.1093/eurheartj/ehs277
7. Anker SD, Schroeder S, Atar D, et al. Traditional and new composite endpoints in heart failure clinical trials: facilitating comprehensive efficacy assessments and improving trial efficiency. *Eur J Heart Fail*. 2016;18(5):482-489. https://doi.org/10.1002/ejhf.516
8. McMurray JJ, Packer M, Desai AS, et al. Angiotensin-Neprilysin inhibition versus enalapril in heart failure. *N Engl J Med*. 2014;371(11):993-1004. https://doi.org/10.1056/NEJMoa1409077
9. Mogensen UM, Gong J, Jhund PS, et al. Effect of Sacubitril/Valsartan on recurrent events in the prospective comparison of ARNI with ACEI to determine impact on global mortality and morbidity in heart failure trial (PARADIGM-HF). *Eur J Heart Fail*. 2018;20(4):760-768. https://doi.org/10.1002/ejhf.1139
10. Claggett B, Pocock S, Wei L, Pfeffer MA, McMurray JJ, Solomon SD. Comparison of time-to-first event and recurrent-event methods in randomized clinical trials. *Circulation*. 2018;138(6):570-577. https://doi.org/10.1161/CIRCULATIONAHA.117.033065
11. Akacha M, Binkowitz B, Claggett B, Hung HMJ, Mueller-Velten G, Stockbridge N. Assessing treatment effects that capture disease burden in serious chronic diseases. *Ther Innov Regul Sci*. 2019;53(3):387-397. https://doi.org/10.1177/2168479018784912
12. Rogers JK, Yaroshinsky A, Pocock SJ, Stokar D, Pogoda J. Analysis of recurrent events with an associated informative dropout time: application of the joint frailty model. *Stat Med*. 2016;35(13):2195-2205. https://doi.org/10.1002/sim.6853
13. Henderson R, Oman P. Effect of frailty on marginal regression estimates in survival analysis. *J Royal Stat Soc Ser B (Stat Methodol)*. 1999;61(2):367-379. https://doi.org/10.1111/1467-9868.00182
14. Aalen OO, Cook RJ, Røysland K. Does Cox analysis of a randomized survival study yield a causal treatment effect? *Lifetime Data Anal*. 2015;21(4):579-593. https://doi.org/10.1007/s10985-015-9335-y
15. Rogers JK, Pocock SJ, McMurray JJ, et al. Analysing recurrent hospitalizations in heart failure: a review of statistical methodology, with application to CHARM-preserved. *Eur J Heart Fail*. 2014;16(1):33-40. https://doi.org/10.1002/ejhf.29
16. Cook RJ, Lawless J. *The Statistical Analysis of Recurrent Events*. New York, NY: Springer; 2007.
17. Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric regression for the mean and rate functions of recurrent events. *J Royal Stat Soc Ser B (Stat Methodol)*. 2000;62(4):711-730. https://doi.org/10.1111/1467-9868.00259
18. Solomon SD, McMurray JJ, Anand IS, et al. Angiotensin-Neprilysin inhibition in heart failure with preserved ejection fraction. *N Engl J Med*. 2019;381(17):1609-1620. https://doi.org/10.1056/nejmoa1908655
19. Liu L, Wolfe RA, Huang X. Shared frailty models for recurrent events and a terminal event. *Biometrics*. 2004;60(3):747-756. https://doi.org/10.1111/j.0006-341X.2004.00225.x
20. Ghosh D, Lin D. Semiparametric analysis of recurrent events data. *Biometrics*. 2003;59(December):877-885.
21. Charles-Nelson A, Katsahian S, Schramm C. How to analyze and interpret recurrent events data in the presence of a terminal event: an application on readmission after colorectal cancer surgery. *Stat Med*. 2019;38(18):sim.8168. https://doi.org/10.1002/sim.8168
22. Luo X, Wang MC, Huang CY. A comparison of various rate functions of a recurrent event process in the presence of a terminal event. *Stat Methods Med Res*. 2010;19(2):167-182. https://doi.org/10.1177/0962280208090220
23. Mao L, Lin DY. Semiparametric regression for the weighted composite endpoint of recurrent and terminal events. *Biostatistics*. 2016;17(2):390-403. https://doi.org/10.1093/biostatistics/kxv050
24. EMA's Committee for Medicinal Products for Human Use (CHMP) Treatment effect measures when using recurrent event endpoints: qualification opinion of clinically interpretable treatment effect measures based on recurrent event endpoints that allow for efficient statistical analyses; 2020. https://www.ema.europa.eu/en/human-regulatory/research-development/scie%ntific-advice-protocol-assistance/novel-methodologies-biomarkers/opinions-lett%ers-support-qualification-novel-methodologies-medicine-development#treatment-e%ffect-measures-when-using-recurrent-event-endpoints-section. Accessed June 23, 2020.
25. Rauch G, Kieser M, Binder H, Bayes-Genis A, Jahn-Eimermacher A. Time-to-first-event versus recurrent-event analysis: points to consider for selecting a meaningful analysis strategy in clinical trials with composite endpoints. *Clin Res Cardiol*. 2018;107(5):437-443. https://doi.org/10.1007/s00392-018-1205-7
26. Struthers CA, Kalbfleisch JD. Misspecified proportional hazard models. *Biometrika*. 1986;73(2):363-369. https://doi.org/10.2307/2336212
27. Hjort NL. On inference in parametric survival data models. *Int Stat Rev / Revue Internationale de Statistique*. 1992;60(3):355-387. https://doi.org/10.2307/1403683
28. Zhong Y, Cook RJ. The effect of omitted covariates in marginal and partially conditional recurrent event analyses. *Lifetime Data Anal*. 2019;25(2):280-300. https://doi.org/10.1007/s10985-018-9430-y
29. Toenges G, Jahn-Eimermacher A. Marginal hazard ratio estimates in joint frailty models for heart failure trials. *Biom J*. 2019;61:1385-1401. https://doi.org/10.1002/bimj.201800133
30. Zhong Y, Cook RJ. Semiparametric recurrent event vs time-to-first-event analyses in randomized trials: estimands and model misspecification. *Stat Med*. 2021;40(16):3823-3842. https://doi.org/10.1002/sim.9002
31. Cox DR. Regression models and life-tables. *J Royal Stat Soc Ser B (Methodol)*. 1972;34(2):187-220.

32. Peña EA, Strawderman RL, Hollander M. Nonparametric estimation with recurrent event data. *J Am Stat Assoc*. 2001;96(456):1299-1315. https://doi.org/10.1198/016214501753381922
33. Ghosh D, Lin DY. Nonparametric analysis of recurrent events and death. *Biometrics*. 2000;56(2):554-562. https://doi.org/10.1111/j.0006-341X.2000.00554.x
34. Rondeau V, Mathoulin-Pelissier S, Jacqmin-Gadda H, Brouste V, Soubeyran P. Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events. *Biostatistics*. 2006;8(4):708-721. https://doi.org/10.1093/biostatistics/kxl043
35. Liu L, Huang X. The use of Gaussian quadrature for estimation in frailty proportional hazards models. *Stat Med*. 2008;27(14):2665-2683. https://doi.org/10.1002/sim.3077
36. Toenges G, Jahn-Eimermacher A. Computational issues in fitting joint frailty models for recurrent events with an associated terminal event. *Comput Methods Prog Biomed*. 2020;188:e105259. https://doi.org/10.1016/j.cmpb.2019.105259
37. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2019 https://www.R-project.org/
38. Ingel K, Jahn-Eimermacher A, Preussler S, Marini F. simrec: simulation of recurrent event data for non-constant baseline hazard. R package version 1.0.0; 2020. https://CRAN.R-project.org/package=simrec.
39. Therneau T. survival: a package for survival analysis in R. R package version 3.1-7; 2019. https://CRAN.R-project.org/package=survival.
40. Mao L. R-code for semi-parametric proportional rate regression for the weighted composite endpoint of recurrent and terminal events; 2016. https://sites.google.com/view/lmaowisc/software/compoml. Accessed July 17, 2020.
41. Tawiah R, GJ ML, Ng SK. A bivariate joint frailty model with mixture framework for survival analysis of recurrent events with dependent censoring and cure fraction. *Biometr Methodol*. 2019;2020:1-14. https://doi.org/10.1111/biom.13202
42. Ingel K, Jahn-Eimermacher A. Sample-size calculation and reestimation for a semiparametric analysis of recurrent event data taking robust standard errors into account. *Biom J*. 2014;56(4):631-648. https://doi.org/10.1002/bimj.201300090
43. Tang Y. Sample size for comparing negative binomial rates in noninferiority and equivalence trials with unequal follow-up times. *J Biopharm Stat*. 2018;28(3):475-491. https://doi.org/10.1080/10543406.2017.1333998
44. Mütze T, Glimm E, Schmidli H, Friede T. Group sequential designs with robust semiparametric recurrent event models. *Stat Methods Med Res*. 2019;28(8):2385-2403. https://doi.org/10.1177/0962280218780538
45. Balan TA, Putter H. Nonproportional hazards and unobserved heterogeneity in clustered survival data: when can we tell the difference? *Stat Med*. 2019;38(18):sim.8171. https://doi.org/10.1002/sim.8171
46. Bluhmki T, Putter H, Allignol A, Beyersmann J. Bootstrapping complex time-to-event data without individual patient data, with a view toward time-dependent exposures. *Stat Med*. 2019;38(20):3747-3763. https://doi.org/10.1002/sim.8177

How to cite this article: Toenges G, Mütze T, Jahn-Eimermacher A. A comparison of semiparametric approaches to evaluate composite endpoints in heart failure trials. *Statistics in Medicine*. 2021;40(26):5702-5724. https://doi.org/10.1002/sim.9149

APPENDIX A. SIMPLIFICATION OF THE LEAST FALSE PARAMETER FUNCTION IN CASE OF A SINGLE BINARY COVARIATE

First we recall that the general form of the least false parameter function is given by

$$g(\beta) = \int_0^\infty E\left[Y_i(t)\left\{X_i - \frac{E[Y_i(t)X_i \exp(\beta' X_i)]}{E[Y_i(t)\exp(\beta' X_i)]}\right\} \mathrm{d}N_i(t)\right]. \tag{A1}$$

Here $E[\cdot]$ denotes an expectation taken with respect to the at-risk process $Y_i(t)$, the covariates X_i and the counting process $N_i(t)$. We will henceforth simplify that function for $X_i \sim Bin(1, p)$ being binary. Therefore, we first dissolve the inner expectations:

$$\begin{aligned} E[Y_i(t)X_i \exp(\beta' X_i)] &= \sum_{a=0}^{1}\sum_{b=0}^{1} P(X_i = a, Y_i(t) = b)\cdot ba\exp(\beta a) \\ &= P(X_i = 1, Y_i(t) = 1)\cdot \exp(\beta), \end{aligned} \tag{A2}$$

$$\begin{aligned} E[Y_i(t)\exp(\beta' X_i)] &= \sum_{a=0}^{1}\sum_{b=0}^{1} P(X_i = a, Y_i(t) = b)\cdot b\exp(\beta a) \\ &= P(X_i = 0, Y_i(t) = 1) + P(X_i = 1, Y_i(t) = 1)\cdot \exp(\beta). \end{aligned} \tag{A3}$$

Now let us denote $\bar{y}_i^{(k)}(t) = P(X_i = k, Y_i(t) = 1)$ and dissolve the whole integrand of the least false parameter function:

$$
\begin{aligned}
& E\left[Y_i(t)\left\{X_i - \frac{\bar{y}_i^{(1)}(t)\exp(\beta)}{\bar{y}_i^{(0)}(t) + \bar{y}_i^{(1)}(t)\exp(\beta)}\right\} \mathrm{d}N_i(t)\right] \\
&= E[Y_i(t)X_i \mathrm{d}N_i(t)] - \frac{\bar{y}_i^{(1)}(t)\exp(\beta)}{\bar{y}_i^{(0)}(t) + \bar{y}_i^{(1)}(t)\exp(\beta)} E[Y_i(t)\mathrm{d}N_i(t)] \\
&= \left[\sum_{a=0}^{1}\sum_{b=0}^{1} P(X_i = a, Y_i(t) = b) \cdot baE[\mathrm{d}N_i(t)|X_i = a, Y_i(t) = b]\right] \\
&\quad - \frac{\bar{y}_i^{(1)}(t)\exp(\beta)}{\bar{y}_i^{(0)}(t) + \bar{y}_i^{(1)}(t)\exp(\beta)} \left[\sum_{a=0}^{1}\sum_{b=0}^{1} P(X_i = a, Y_i(t) = b) \cdot bE[\mathrm{d}N_i(t)|X_i = a, Y_i(t) = b]\right] \\
&= \bar{y}_i^{(1)}(t)E[\mathrm{d}N_i(t)|X_i = 1, Y_i(t) = 1] \\
&\quad - \frac{\bar{y}_i^{(1)}(t)\exp(\beta)}{\bar{y}_i^{(0)}(t) + \bar{y}_i^{(1)}(t)\exp(\beta)} [\bar{y}_i^{(0)}(t)E[\mathrm{d}N_i(t)|X_i = 0, Y_i(t) = 1] + \bar{y}_i^{(1)}(t)E[\mathrm{d}N_i(t)|X_i = 1, Y_i(t) = 1]] \\
&= \frac{\bar{y}_i^{(0)}(t)\bar{y}_i^{(1)}(t)}{\bar{y}_i^{(0)}(t) + \bar{y}_i^{(1)}(t)\exp(\beta)} \left[E[\mathrm{d}N_i(t)|X_i = 1, Y_i(t) = 1] - E[\mathrm{d}N_i(t)|X_i = 0, Y_i(t) = 1]\exp(\beta)\right]. \qquad \text{(A4)}
\end{aligned}
$$

In summary, the least false parameter function in case of a single binary covariate is given by

$$
g(\beta) = \int_0^\infty \frac{\bar{y}_i^{(0)}(t)\bar{y}_i^{(1)}(t)}{\bar{y}_i^{(0)}(t) + \bar{y}_i^{(1)}(t)\exp(\beta)} \left[E[\mathrm{d}N_i(t)|X_i = 1, Y_i(t) = 1] - E[\mathrm{d}N_i(t)|X_i = 0, Y_i(t) = 1]\exp(\beta)\right]. \qquad \text{(A5)}
$$

APPENDIX B. COMPONENTS OF THE LEAST FALSE PARAMETER FUNCTION IN A JOINT FRAILTY SETTING

First, we derive the rate functions targeted in the Cox model, the LWYY model and the Mao-Lin model:

Cox rate: $r_i^{Cox}(t|X_i)\ \mathrm{d}t = E[\mathrm{d}N_i(t)|X_i, Y_i(t) = 1]$ with $Y_i(t) = I(D_i \geq t, T_{i1} \geq t, C_i \geq t)$

$$
\begin{aligned}
E[\mathrm{d}N_i(t)|X_i, Y_i(t) = 1] &= E[\mathrm{d}N_i(t)|X_i, D_i \geq t, T_{i1} \geq t, C_i \geq t] \\
&= E[\mathrm{d}N_i(t)|X_i, D_i \geq t, T_{i1} \geq t] \\
&= E_{Z_i}[E[\mathrm{d}N_i(t)|X_i, Z_i, D_i \geq t, T_{i1} \geq t]|X_i, D_i \geq t, T_{i1} \geq t] \\
&= E_{Z_i}[E[\mathrm{d}N_{i1}(t)|X_i, Z_i, D_i \geq t, T_{i1} \geq t]|X_i, D_i \geq t, T_{i1} \geq t] \\
&\quad + E_{Z_i}[E[\mathrm{d}N_{i2}(t)|X_i, Z_i, D_i \geq t, T_{i1} \geq t]|X_i, D_i \geq t, T_{i1} \geq t] \\
&= E[Z_i|X_i, D_i \geq t, T_{i1} \geq t]\lambda_{10}(t)\exp(\beta_1 X_i)\ \mathrm{d}t + E[Z_i^\gamma|X_i, D_i \geq t, T_{i1} \geq t]\lambda_{20}(t)\exp(\beta_2 X_i)\ \mathrm{d}t. \qquad \text{(B1)}
\end{aligned}
$$

LWYY rate: $r_i^{LWYY}(t|X_i)\ \mathrm{d}t = E[\mathrm{d}N_i(t)|X_i, Y_i(t) = 1]$ with $Y_i(t) = I(D_i \geq t, C_i \geq t)$

$$
\begin{aligned}
E[\mathrm{d}N_i(t)|X_i, Y_i(t) = 1] &= E[\mathrm{d}N_i(t)|X_i, D_i \geq t, C_i \geq t] \\
&= E[\mathrm{d}N_i(t)|X_i, D_i \geq t] \\
&= E_{Z_i}[E[\mathrm{d}N_i(t)|X_i, Z_i, D_i \geq t]|X_i, D_i \geq t] \\
&= E_{Z_i}[E[\mathrm{d}N_{i1}(t)|X_i, Z_i, D_i \geq t]|X_i, D_i \geq t] + E_{Z_i}[E[\mathrm{d}N_{i2}(t)|X_i, Z_i, D_i \geq t]|X_i, D_i \geq t] \\
&= E[Z_i|X_i, D_i \geq t]\lambda_{10}(t)\exp(\beta_1 X_i)\ \mathrm{d}t + E[Z_i^\gamma|X_i, D_i \geq t]\lambda_{20}(t)\exp(\beta_2 X_i)\ \mathrm{d}t. \qquad \text{(B2)}
\end{aligned}
$$

Mao-Lin rate: $r_i^{Mao-Lin}(t|X_i)\ \mathrm{d}t = E[\mathrm{d}N_i(t)|X_i, Y_i(t) = 1]$ with $Y_i(t) = I(C_i \geq t)$

$$
E[\mathrm{d}N_i(t)|X_i, Y_i(t) = 1] = E[\mathrm{d}N_i(t)|X_i, C_i \geq t]
$$

$$
\begin{aligned}
&= E[\mathrm{d}N_i(t)|X_i] \\
&= S_{D_i}(t|X_i)E[\mathrm{d}N_i(t)|X_i, D_i \geq t] \\
&= S_{D_i}(t|X_i)E_{Z_i}[E[\mathrm{d}N_i(t)|X_i, Z_i, D_i \geq t]|X_i, D_i \geq t] \\
&= S_{D_i}(t|X_i)E_{Z_i}[E[\mathrm{d}N_{i1}(t)|X_i, Z_i, D_i \geq t]|X_i, D_i \geq t] \\
&\quad + S_{D_i}(t|X_i)E_{Z_i}[E[\mathrm{d}N_{i2}(t)|X_i, Z_i, D_i \geq t]|X_i, D_i \geq t] \\
&= S_{D_i}(t|X_i)E[Z_i|X_i, D_i \geq t]\lambda_{10}(t)\exp(\beta_1 X_i)\,\mathrm{d}t \\
&\quad + S_{D_i}(t|X_i)E[Z_i^\gamma|X_i, D_i \geq t]\lambda_{20}(t)\exp(\beta_2 X_i)\,\mathrm{d}t.
\end{aligned} \tag{B3}
$$

Next, we derive the conditional expectations appearing in the previously derived rate functions. Within the derivations, $\Lambda_{10}(t) = \int_0^t \lambda_{10}(s)\,\mathrm{d}s$ and $\Lambda_{20}(t) = \int_0^t \lambda_{20}(s)\,\mathrm{d}s$ denote the cumulative conditional baseline rates of the joint frailty model and $f_A(\cdot)$ denotes the probability density function of a random variable or random vector A.

Conditional expectations $E[Z_i|X_i, D_i \geq t, T_{i1} \geq t]$ and $E[Z_i^\gamma|X_i, D_i \geq t, T_{i1} \geq t]$

$$
\begin{aligned}
E[Z_i|X_i, D_i \geq t, T_{i1} \geq t] &= \int_0^\infty z f_{Z_i|X_i, D_i \geq t, T_{i1} \geq t}(z)\,\mathrm{d}z \\
&= \frac{1}{S_{D_i \wedge T_{i1}}(t|X_i)} \int_0^\infty \int_t^\infty z f_{Z_i, D_i \wedge T_{i1}|X_i}(z, s)\,\mathrm{d}s\,\mathrm{d}z \\
&= \frac{1}{E_{Z_i}[S_{D_i \wedge T_{i1}}(t|X_i, Z_i)]} \int_0^\infty \int_t^\infty z f_{D_i \wedge T_{i1}|X_i, Z_i = z}(s) f_{Z_i}(z)\,\mathrm{d}s\,\mathrm{d}z \\
&= \frac{1}{\int_0^\infty S_{D_i \wedge T_{i1}}(t|X_i, Z_i = z) f_{Z_i}(z)\,\mathrm{d}z} \int_0^\infty z S_{D_i \wedge T_{i1}}(t|X_i, Z_i = z) f_{Z_i}(z)\,\mathrm{d}z \\
&= \frac{\int_0^\infty z \exp(-z\exp(\beta_1 X_i)\Lambda_{10}(t) - z^\gamma \exp(\beta_2 X_i)\Lambda_{20}(t)) f_{Z_i}(z)\,\mathrm{d}z}{\int_0^\infty \exp(-z\exp(\beta_1 X_i)\Lambda_{10}(t) - z^\gamma \exp(\beta_2 X_i)\Lambda_{20}(t)) f_{Z_i}(z)\,\mathrm{d}z}.
\end{aligned} \tag{B4}
$$

Similarly the conditional expectation $E[Z_i^\gamma|X_i, D_i \geq t, T_{i1} \geq t]$ can be shown to be

$$
E[Z_i^\gamma|X_i, D_i \geq t, T_{i1} \geq t] = \frac{\int_0^\infty z^\gamma \exp(-z\exp(\beta_1 X_i)\Lambda_{10}(t) - z^\gamma \exp(\beta_2 X_i)\Lambda_{20}(t)) f_{Z_i}(z)\,\mathrm{d}z}{\int_0^\infty \exp(-z\exp(\beta_1 X_i)\Lambda_{10}(t) - z^\gamma \exp(\beta_2 X_i)\Lambda_{20}(t)) f_{Z_i}(z)\,\mathrm{d}z}. \tag{B5}
$$

Conditional expectations $E[Z_i|X_i, D_i \geq t]$ and $E[Z_i^\gamma|X_i, D_i \geq t]$

$$
\begin{aligned}
E[Z_i|X_i, D_i \geq t] &= \int_0^\infty z f_{Z_i|X_i, D_i \geq t}(z)\,\mathrm{d}z \\
&= \frac{1}{S_{D_i}(t|X_i)} \int_0^\infty \int_t^\infty z f_{Z_i, D_i|X_i}(z, s)\,\mathrm{d}s\,\mathrm{d}z \\
&= \frac{1}{E_{Z_i}[S_{D_i}(t|X_i, Z_i)]} \int_0^\infty \int_t^\infty z f_{D_i|X_i, Z_i = z}(s) f_{Z_i}(z)\,\mathrm{d}s\,\mathrm{d}z \\
&= \frac{1}{\int_0^\infty S_{D_i}(t|X_i, Z_i = z) f_{Z_i}(z)\,\mathrm{d}z} \int_0^\infty z S_{D_i}(t|X_i, Z_i = z) f_{Z_i}(z)\,\mathrm{d}z \\
&= \frac{\int_0^\infty z \exp(-z^\gamma \exp(\beta_2 X_i)\Lambda_{20}(t)) f_{Z_i}(z)\,\mathrm{d}z}{\int_0^\infty \exp(-z^\gamma \exp(\beta_2 X_i)\Lambda_{20}(t)) f_{Z_i}(z)\,\mathrm{d}z}.
\end{aligned} \tag{B6}
$$

Similarly the conditional expectation $E[Z_i^\gamma|X_i, D_i \geq t]$ can be shown to be

$$
E[Z_i^\gamma|X_i, D_i \geq t] = \frac{\int_0^\infty z^\gamma \exp(-z^\gamma \exp(\beta_2 X_i)\Lambda_{20}(t)) f_{Z_i}(z)\,\mathrm{d}z}{\int_0^\infty \exp(-z^\gamma \exp(\beta_2 X_i)\Lambda_{20}(t)) f_{Z_i}(z)\,\mathrm{d}z}. \tag{B7}
$$

Survival functions $S_{D_i \wedge T_{i1}}(t|X_i)$ and $S_{D_i}(t|X_i)$

Finally we resolve the survival functions appearing in the least false parameter function $g(\beta)$. Their derivations are already contained in the above calculations for the conditional expectations. These functions are given by

$$S_{D_i \wedge T_{i1}}(t|X_i) = \int_0^\infty \exp(-z\exp(\beta_1 X_i)\Lambda_{10}(t) - z^\gamma \exp(\beta_2 X_i)\Lambda_{20}(t)) f_{Z_i}(z)\, dz,$$
$$S_{D_i}(t|X_i) = \int_0^\infty \exp(-z^\gamma \exp(\beta_2 X_i)\Lambda_{20}(t)) f_{Z_i}(z)\, dz. \tag{B8}$$

APPENDIX C. CONSEQUENCES OF VARYING THE CORRELATION AND THE COMPONENT-SPECIFIC CONDITIONAL BASELINE RATES (FIGURE C1)

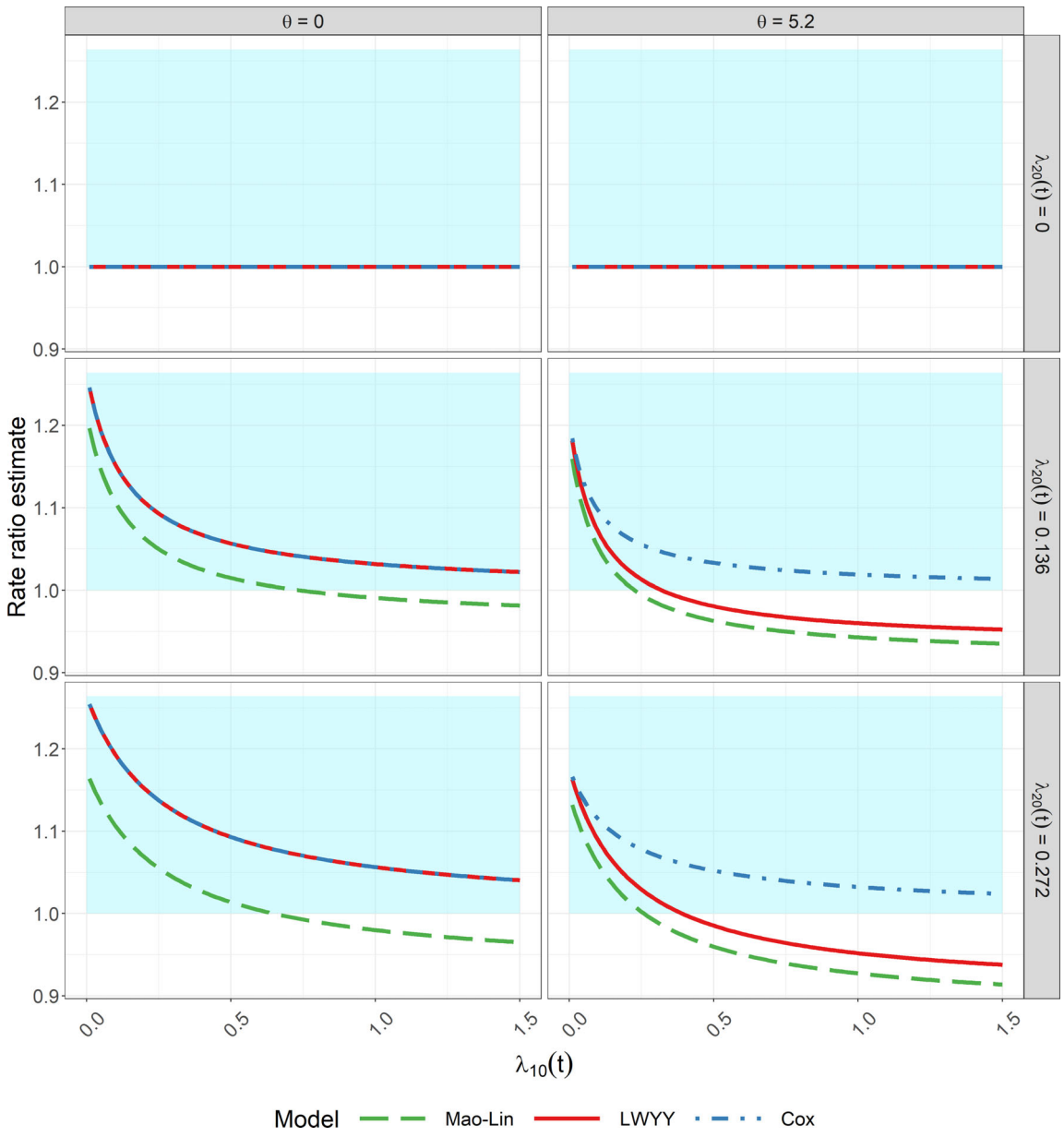

FIGURE C1 Asymptotic rate ratio estimates $\exp(\bar{\beta})$ for the composite endpoint in joint gamma frailty models with different frailty variances θ and different time-constant, conditional baseline rates $\lambda_{10}(t)$ (recurrent events), and $\lambda_{20}(t)$ (terminal event). The remaining joint frailty parameters are fixed as $\exp(\beta_1) = 1$, $\exp(\beta_2) = 1.264$, and $\gamma = 0.63$. The interval between $\exp(\beta_1)$ and $\exp(\beta_2)$ is represented by the shaded area. The censoring scheme is given by administrative censoring at time point $t = 2.5$ years (probability 0.8) and uniformly distributed random censoring on $[0, 2.5]$ (probability 0.2) [Colour figure can be viewed at wileyonlinelibrary.com]