Aus dem Institut für Medizinische Biometrie, Epidemiologie und Informatik (IMBEI)
der Universitätsmedizin der Johannes Gutenberg-Universität Mainz

Addressing the competing risk of death in the analysis of recurrent hospitalisations in randomised clinical heart failure trials

Über die Analyse rekurrenter Hospitalisierungen unter Berücksichtigung des konkurrierenden Risikos Tod in randomisierten klinischen Herzinsuffizienz-Studien

Inauguraldissertation
zur Erlangung des Doktorgrades der
physiologischen Wissenschaften
der Universitätsmedizin
der Johannes Gutenberg-Universität Mainz

Vorgelegt von

Gerrit Toenges
aus Wuppertal

Mainz, 2020

Tag der Promotion: 08. Dezember 2020

Contents

Zusammenfassung

Diese Arbeit befasst sich mit statistisch-methodischen Aspekten bei der Schätzung des Behandlungseffekts auf die Hospitalisierungsrate in randomisierten klinischen Herzinsuffizienz-Studien. Konkret stellt sich dabei die Frage nach dem adäquaten Umgang mit der Mortalität als konkurrierendem Risiko. Hierzu werden in der Literatur verschiedene statistische Ansätze diskutiert. Deren Zusammenhänge und Software-Implementierungsmöglichkeiten werden erstmals im Rahmen der zwei publizierten Artikel dieser kumulativen Dissertation hergeleitet.

Chronische Herzinsuffizienz zählt zu den häufigsten Ursachen für vollstationäre Krankenhausbehandlungen weltweit. Herzinsuffizienz-bedingte Hospitalisierungen (HFH, *Heart Failure Hospitalisations*) stellen nicht nur für die betroffenen Patienten eine große Belastung dar, sondern sind auch aus rein wirtschaftlicher Sicht für das Gesundheitssystem ein enormer Kostenfaktor. Aus diesem Grund gehört der Effekt auf die HFH-Rate zu den wichtigen Endpunkten von randomisierten, klinischen Studien zu neuen Behandlungen in diesem Indikationsgebiet. Dabei handelt es sich in der Regel um zweiarmige, longitudinale Interventionsstudien, in denen eine Behandlungsgruppe mit der Kombination aus Standardmedikation und neuer Behandlung gegen eine Kontrollgruppe ausschließlich mit Standardmedikation verglichen wird. Bei der statistischen Schätzung des Behandlungseffekts auf die HFH-Rate sind in diesen Studien folgende Aspekte zu berücksichtigen:

- Patienten können während der Beobachtungszeit mehrfach hospitalisiert werden. Es handelt sich also um ein sogenanntes "rekurrentes Ereignis".
- Die Beobachtungszeiten der Patienten variieren aufgrund von
  - Zensierungen: Eine Zensierung terminiert definitionsgemäß nur die Beobachtung, jedoch nicht das Auftreten weiterer HFHs. Sie kommt entweder durch Lost to follow-up oder das Studienende zustande.
  - Mortalität als konkurrierendem, terminalem Ereignis: Ein konkurrierendes, terminales Ereignis terminiert definitionsgemäß nicht nur die Beobachtung, sondern auch das Auftreten weiterer HFHs. Insbesondere die Mortalität aufgrund

kardiovaskulärer Ursachen (CVD, *Cardiovascular Death*) ist ein wichtiger Grund für Drop-outs in Herzinsuffizienz-Studien.

Die Analyse von rekurrenten Ereignissen ist gut verstanden und es existieren einige statistische Modelle, die sich inzwischen als Standard etabliert haben. Zensierungen werden in diesen Standard-Modellen stets berücksichtigt, ein terminales, konkurrierendes Ereignis jedoch nicht explizit. Daher wird das terminale Ereignis Tod bei der Anwendung solcher Modelle technisch wie eine Zensierung gehandhabt. Das ist jedoch kritisch zu hinterfragen, da diese Modelle Zensierungen als "nicht informativ" für die Rate rekurrenter Ereignisse annehmen. Eine Assoziation zwischen den CVD- und HFH-Raten scheint jedoch klinisch plausibel und wurde auch in einigen post-hoc Auswertungen von Studien nachgewiesen. In diesem Zusammenhang ist der Begriff "Assoziation" so zu verstehen, dass für beide Endpunkte in der Studienpopulation eine gewisse Heterogenität in Bezug auf die Patienten-spezifischen Raten existiert und Patienten mit einer hohen (niedrigen) HFH-Rate auch gleichzeitig eine hohe (niedrige) CVD-Rate aufweisen. Folglich werden Hochrisiko-Patienten, die vieler Hospitalisierungen bedürfen, in der Studie aufgrund früher Drop-outs durch CVD im Schnitt wesentlich kürzer beobachtet als Niedrigrisiko-Patienten. CVD verletzt daher die Annahme, nicht informativ für die HFH-Rate zu sein.

Trotzdem ist es gängige Praxis, Standard-Modelle für rekurrente Ereignisse auf den HFH-Endpunkt anzuwenden und CVD dabei wie eine Zensierung zu behandeln. Häufig wird in diesem Kontext der Andersen-Gill Schätzer zur Quantifizierung des Behandlungseffekts verwendet. Eine methodische Alternative, die ein assoziiertes terminales Ereignis adäquater berücksichtigt, sich jedoch in der Praxis bislang noch nicht durchgesetzt hat, würde in der Anwendung eines Joint Frailty Modells bestehen. Hier werden die Endpunkte HFH und CVD gemeinsam modelliert, wodurch die Assoziation zwischen den Patienten-spezifischen HFH- und CVD-Raten explizit im Modell berücksichtigt werden kann. Insbesondere wird CVD bei diesem Ansatz somit nicht einfach wie eine Zensierung behandelt. Die Parameterschätzung in diesem Modell gestaltet sich dadurch jedoch deutlich komplexer und rechenintensiver. Verlässliche Software-Implementierungen sind bislang nicht verfügbar, was die Anwendung des Modells in klinischen Studien erschwert.

Daraus ergeben sich folgende Fragen, mit denen sich die beiden Artikel dieser kumulativen Dissertation befassen:

1. Wie unterscheiden sich die aus dem Andersen-Gill- und Joint Frailty-Ansatz resultierenden Effektschätzer für den HFH-Endpunkt, wenn eine Assoziation mit dem CVD-Endpunkt vorliegt? In welcher Richtung und Größenordnung weichen die Schätzer der beiden Methoden voneinander ab und von welchen Faktoren ist dies abhängig?

2. Welche Möglichkeiten zur Implementierung des Joint Frailty Modells gibt es, insbesondere für die im Umfeld klinischer Studien verwendete Software SAS®? Welche Implementierungsvariante garantiert die verlässlichsten Ergebnisse?

Der erste Artikel nimmt Bezug auf Fragestellung (1). Hier wurden Unterschiede zwischen den aus dem Andersen-Gill- und Joint Frailty-Ansatz resultierenden Behandlungseffekt-Schätzungen für den HFH-Endpunkt untersucht und erklärt. Es konnte gezeigt werden, dass der unterschiedliche Umgang mit dem assoziierten, terminalen Ereignis CVD in beiden Ansätzen zu teils deutlichen Unterschieden in den Schätzungen führt. Die Richtung des Unterschieds hängt im Wesentlichen davon ab, ob die Behandlung einen positiven oder negativen Effekt auf den CVD-Endpunkt hat. Für die Größenordnung der Abweichung sind die Stärke der Assoziation, die Stärke des Behandlungseffekts auf den CVD-Endpunkt sowie die Studiendauer ausschlaggebend. Desweiteren wurden in dem Artikel die mathematisch-theoretischen Hintergründe erarbeitet, die dieses Verhalten der Schätzer erklären und im Folgenden erläutert werden.

Ein zentrales Ergebnis dieser Arbeit ist, dass im Falle assoziierter Endpunkte zwei Typen von HFH-Raten und dementsprechend auch zwei Behandlungseffekte voneinander unterschieden werden müssen. Das Joint Frailty Modell misst den Behandlungseffekt auf Patienten-Ebene, der wahlweise als Patienten-spezifischer oder bedingter Effekt bezeichnet wird. Dieser gibt an, um welchen Faktor sich die heterogenen Patienten-spezifischen HFH-Raten durch die untersuchte Behandlung ändern. Um den Effekt in Form eines einzelnen Faktors ausdrücken zu können, wird im Joint Frailty Modell implizit angenommen, dass der Effekt zeitlich konstant und für alle Patienten gleich ist. Den heterogenen Patienten-spezifischen HFH-Raten steht die marginale HFH-Rate gegenüber. Der Wert der marginalen Rate zu einem bestimmten Zeitpunkt ist die

durchschnittliche Patienten-spezifische Rate derjenigen Patienten, die zu diesem Zeitpunkt noch unter Beobachtung sind, d.h. zuvor weder verstorben sind noch zensiert wurden. Es handelt sich also um eine Populations-bezogene Größe. Weiterhin konnte gezeigt werden, dass der Behandlungseffekt auf die marginale HFH-Rate unter Assoziation trotz eines Zeit-konstanten Effekts auf die Patienten-spezifischen HFH-Raten in der Regel Zeit-abhängig ist. Einzig in Szenarien mit gleicher CVD-Mortalität in beiden Gruppen, was gerade in Herzinsuffizienz-Studien durchaus vorkommt, stimmen der Patienten-spezifische und der marginale Effekt überein. Realistisch sind jedoch auch Szenarien, in denen die Behandlung protektiv auf die HFH- und CVD-Endpunkte wirkt. Während der Patienten-spezifische Effekt hier konstant ist, schwächt der marginale Effekt über die Zeit hinweg ab. Dies ist dadurch zu erklären, dass Hochrisiko-Patienten in der Kontrollgruppe wesentlich schneller aufgrund von CVD aus der Studienpopulation ausscheiden als Hochrisiko-Patienten in der Behandlungsgruppe. Da der durchschnittliche Gesundheitsstatus der beiden Gruppen im Laufe der Studie immer weiter auseinander divergiert, wird der Unterschied zwischen den marginalen Raten der beiden Gruppen über die Zeit hinweg immer geringer. Der Andersen-Gill Schätzer zielt auf den marginalen Behandlungseffekt ab, spezifiziert diesen aber (in der Regel fälschlicherweise) als zeitlich konstant. Hier konnte im Rahmen dieser Arbeit mathematisch belegt werden, dass die resultierende Schätzung dann einen zeitlich gemittelten marginalen Behandlungseffekt darstellt. Die Schätzung hängt somit stark von der Länge der Beobachtungszeit, also der Dauer der Studie ab. Weil dies die Vergleichbarkeit von Ergebnissen aus Studien unterschiedlicher Dauer im Grunde unmöglich macht, ist der Andersen-Gill Schätzer in der Konsequenz also eher ungeeignet zur Quantifizierung des Behandlungseffekts auf den HFH-Endpunkt.

Vor diesem Hintergrund scheint der Einsatz des Joint Frailty Modells zur Quantifizierung des Behandlungseffekts auf den HFH-Endpunkt sinnvoller. Die bisher in der publizierten Literatur zu beobachtende Zurückhaltung im Hinblick auf eine Joint Frailty Analyse ist vermutlich zum Teil dadurch zu erklären, dass es bislang an verlässlichen Software-Lösungen zur Schätzung dieses Modells mangelte. Dies betrifft insbesondere die Statistik-Software SAS, die im streng regulierten Umfeld klinischer Studien verwendet wird. In SAS existiert keine Prozedur, die eine direkte Anwendung des Joint Frailty Modells ohne eigene Programmierung und eine genaue Kennt-

nis der mathematischen Grundlagen erlaubt. Daher befasst sich der zweite Artikel dieser Dissertation mit Fragestellung (2), d.h. mit der Ausarbeitung und dem Vergleich möglicher Implementierungsvarianten in SAS. Dazu wurden Funktionen der SAS-Prozedur NLMIXED genutzt, welche die Maximierung einer vom Nutzer selbst zu spezifizierenden Likelihood-Funktion ermöglicht. Konkret muss die auf die Frailty-Variable bedingte log-Likelihood-Funktion des Modells in spezieller Form zerlegt und implementiert werden. Darüber hinaus ist die NLMIXED-Prozedur auf eine Lognormal-Verteilung für die Frailty-Variable beschränkt. Um davon abweichende Verteilungen (z.B. Gamma-Verteilung) im Modell zu ermöglichen und somit die NLMIXED-Einschränkung zu umgehen, bedarf es weiterer Modifikationen der bedingten log-Likelihood-Funktion. Hierfür wurden mit der PIT- (*Probability Integral Transformation*) und der LR-Methode (*Likelihood Reformulation*) zwei verschiedene Ansätze in der Literatur in allgemeiner Form beschrieben, jedoch bislang noch nie im Joint Frailty Kontext miteinander verglichen. Ein wichtiges Zwischenresultat dieser Arbeit besteht daher (a) in der Herleitung und geeigneten Zerlegung bedingter log-Likelihood-Funktionen für Joint Frailty Modelle und (b) in deren Umsetzung in Programm-Code. Zuletzt existieren in der NLMIXED-Prozedur noch zahlreiche numerische Optionen für die Likelihood-Berechnung und -Maximierung, deren Einfluss auf die Qualität der Schätzungen in der Literatur bislang noch nicht untersucht wurde.

Neben der mathematisch-theoretischen Ausarbeitung wurde im Rahmen des zweiten Artikels der Einfluss all dieser verschiedenen Implementierungsmöglichkeiten auf die Qualität der Schätzungen durch eine große Simulationsstudie untersucht. Insbesondere für Modelle mit Gamma-Frailty zeigten sich starke Unterschiede zwischen den verschiedenen Optionen bzw. Implementierungsvarianten. Daraus wurden Empfehlungen dafür abgeleitet, welche Implementierung die verlässlichsten, stabilsten Ergebnisse verspricht. In der Simulation wurden die Ergebnisse der verschiedenen SAS-Implementierungen auch mit denen eines zur Joint Frailty Analyse fähigen Paketes der Software R verglichen. Hier hat sich ergeben, dass man mit gut überlegten SAS-Implementierungen deutlich bessere Resultate erzielt als mit vorhandenen R-Funktionen. Um Anwendern die Schätzung des Modells in SAS zu erleichtern, wurden alle Implementierungsvarianten zu einem SAS-Makro zusammengefügt, welches samt Bedienungsanleitung als ergänzendes Online-Material zu dem Artikel zur Verfügung steht.

Insgesamt konnte mit dieser Arbeit gezeigt werden, dass der Andersen-Gill Schätzer für den HFH-Endpunkt unter einer Assoziation mit dem CVD-Endpunkt von der Studiendauer abhängt und sich deshalb als Analysemethode in Herzinsuffizienz-Studien disqualifiziert. Die Anwendung eines Joint Frailty Modells, welches die Assoziation zwischen den Endpunkten explizit berücksichtigt, wird im Kontext dieser Studien als sinnvoller erachtet. Hierzu wurden Software-Lösungen erarbeitet, die verlässliche Ergebnisse versprechen und somit den Weg für die Anwendung des Modells in der Praxis klinischer Studien ebnen könnten. Somit leistet diese Arbeit einen Beitrag zu der seit einigen Jahren geführten Diskussion, wie CVD bei der Analyse des HFH-Endpunktes adäquat behandelt werden sollte.

1 Background

Disease-related hospitalisations are a major outcome of randomised clinical trials in chronic heart failure. Assessing a treatment effect on that outcome is complicated by mortality, whose rate is likely associated with the hospitalisation rate. This raises important statistical issues that will be addressed in this cumulative PhD-thesis.

1.1 Chronic heart failure syndrome

Chronic heart failure (HF) is defined as the progressively ongoing incapacity of the heart muscle to pump sufficient amounts of blood to maintain a circulation that meets the body's needs. This weakened performance is caused by conditions that result in structural or functional changes of the left (left-sided HF) or right (right-sided HF) ventricle, or even both (biventricular HF). In general, people initially develop heart failure due to problems with the left ventricle, whose physiological consequences in turn also affect the right ventricle.

Depending on the etiology, left-sided heart failure can be subclassified in pathologies with preserved (HFpEF) and reduced ejection fraction (HFrEF). As an example, the heart reacts to long-term systemic hypertension with a hypertrophy of the left ventricle, which makes it stiff and unable to relax adequately. Hence the end-diastolic volume, that is, the blood volume when the left ventricle is maximally filled, is decreasing. This leads to a reduced amount of ejected blood per beat, as characteristic for heart failure. However, the ejection fraction, given by the percentage of the blood volume ejected from the left ventricle with each heartbeat divided by the end-diastolic volume, is normal ($> 50\%$). Such clinical pattern is referred to as heart failure with preserved ejection fraction. In contrast, heart failure with reduced ejection fraction refers to a scenario with a normal end-diastolic volume but a reduced ejection volume of the left ventricle. The reduced ejection volume is due to a weakening of the left-ventricular muscle, arising for example from a reduced blood flow in the coronary arteries (coronary artery disease) or from tissue damages e.g. from infections or myocardial infarctions. Right-sided HF generally manifests in a right-ventricular hypertrophy due to enduring

pulmonary hypertension, which itself is provoked by a left-sided HF in many patients.

Chronic heart failure is a syndrome with a wide range of symptoms and degrees of severity. Symptoms of left- and right-sided HF are caused by the congestion of the blood in the circulation upstram the impaired ventricle. Left-sided HF leads to a congestion in the pulmonary circulation, causing respiratory problems (dyspnoea), poor exercise performance and fatigue due to insufficient oxygen uptake. In serious cases fluid even leaks out from the blood to the lung (pulmonary edema), which exacerbates the patient's symptoms dramatically. Right-sided HF comes along with a blood congestion in the systemic circulation, resulting in water retention in the tissues. As a consequence patients suffer, amongst others, from fluid accumulations in the legs (peripheral edema) and in the abdominal cavity (ascites).

In most cases lifestyle modifications and pharmacotherapy of the HF-causes (e.g. hypertension) do not achieve a recovery of the full heart performance. Hence chronic heart failure can in general not be cured in the proper sense and needs lifelong management. For that reason the main goals of medication are to attenuate symptoms and to maintain clinical stability, that is, to prevent a progression of the disease. There is a broad range of drugs being used and the treatment schemes are individually adjusted to the severity and etiology of the disease. Usually patients need multiple medications targeting separate symptoms or causal factors. The most frequently applied pharmacological drug classes are vasodilatators (widening the blood vessels), beta blockers (slowing down the heart rate) and diuretics (increasing water excretion). In sum all these medications aim to improve the hemodynamic flow and to relieve the weakened heart (McMurray and Pfeffer, 2005; Bui *et al.*, 2011; Metra and Teerlink, 2017).

1.2 Efficacy endpoints in heart failure trials

Randomised clinical trials are a mandatory requirement for the regulatory approval of new drugs. In general, the addition of the new medication to standard of care needs to show an improved performance compared to standard of care plus placebo. However, faced to the multidimensional nature of the heart failure syndrome a unifying definition of "improved performance" is hard to find and there is an ongoing debate on

this matter (see e.g. Anker and McMurray, 2012; Zannad *et al.*, 2013; Anker *et al.*, 2016; Packer, 2016; Hicks *et al.*, 2018; Akacha *et al.*, 2019).

Trial designs and primary efficacy endpoints in that field of indication continuously underwent changes during the last decades. For a long time, edema weight or functional exercise performance at a certain timepoint after drug intake were used to evaluate efficacy. These short-term trials were replaced by large long-term trials with the advent of new drugs that significantly reduced mortality. Since then the primary endpoint switched to all-cause mortality in most trials for a long period of time. However, no drug can be expected to have a positive impact on all causes of death. Incorporating modes of death that are not affected by the treatment is diluting the treatment effect and thus calling for larger sample sizes. Hence the scientific communtity waived non-cardiovascular death (nCVD) and focused on cardiovascular death (CVD) as a more disease-related, cause-specific outcome to prove efficacy. But even that endpoint was not maintained for the following reason: Disease-related mortality fortunately stongly decreased under standard of care medication over the past decades. In addition, new drugs are only expected to show, if at all, low improval of cardiovascular mortality. This poses challenges for the (financial) feasability of trials, as these would require even greater sample sizes to be sufficiently powered for a proof of efficacy (Packer, 2016).

It is for that reason, amongst others, that nowadays the primary analysis rests on a composite endpoint including both recurrent heart failure associated hospitalisations (HFH) and CVD. Thereby heart failure hospitalisations serve as a surrogate for the patients' morbidity. The consideration of such non-fatal events in the primary endpoint is relevant both for patients and health care providers, as hospitalisations reflect the disease burden and at the same time present a crucial economic burden. Composite endpoints in general have advantages and disadvantages, but are widely accepted by regulators as long as a time-to-first-event analysis based on a Cox model is used to prove efficacy. Here only the patient's first event (be it CVD or HFH) is incorporated in the effect measure given by the hazard ratio. This approach raised criticism, as approximately 40%-50% of the observed events (CVD or HFH) in major heart failure trials are completely ignored in the primary analysis that way (Anker *et al.*, 2016; Akacha *et al.*, 2019). Various alternative statistical approaches that capture all primary endpoint

events were suggested, but due to a lack of consensus in the scientific community the time-to-first-event analysis is still the established standard.

1.3 Statistical issues in heart failure trials

A drawback of the composite endpoint is its insufficient interpretability. A protective effect on the composite outcome does not necessarily imply protective effects on both components. If a treatment is slightly increasing cardiovascular mortality but at the same time strongly decreasing the hospitalisation rate, its effect on the composite outcome is likely to be evaluated as beneficial. For safety reasons it is required to additionally report the separate (cause-specific) treatment effects on both outcomes (Sankoh *et al.*, 2017). The effect on the HFH-outcome is generally not evaluated by a time-to-first-HFH analysis with the Cox model, but with more advanced models for multivariate time-to-event data, that incorporate all HFHs (Rogers *et al.*, 2012, 2014a,b, 2016). However, effect estimation for the HFH-outcome is complicated by the competing risk of death, which needs particular statistical considerations covered in this thesis.

Censoring and competing risks The statistical setting can be illustrated by a multistate process as shown in Figure 1. Whenever one of the events CVD, nCVD or HFH occurs, the patient is switching its state. While being alive, the patient is at risk for all three outcomes at each point in time. HFH is an event that can occur repeatedly within one patient and therefore referred to as "recurrent event". Statistical models target the transition rates from one state to another. The observation of the patient's "journey" through that multistate scheme may be stopped for two reasons (Cook and Lawless, 2007, p. 47 ff.):

1. Competing risk of death: A patient may die due to cardiovascular or due to non-cardiovascular death. Both modes of death constitute so-called "competing risks" for the HFH-outcome, because they do not only prevent the observation but even the occurrence of further heart failure hopitalisations.

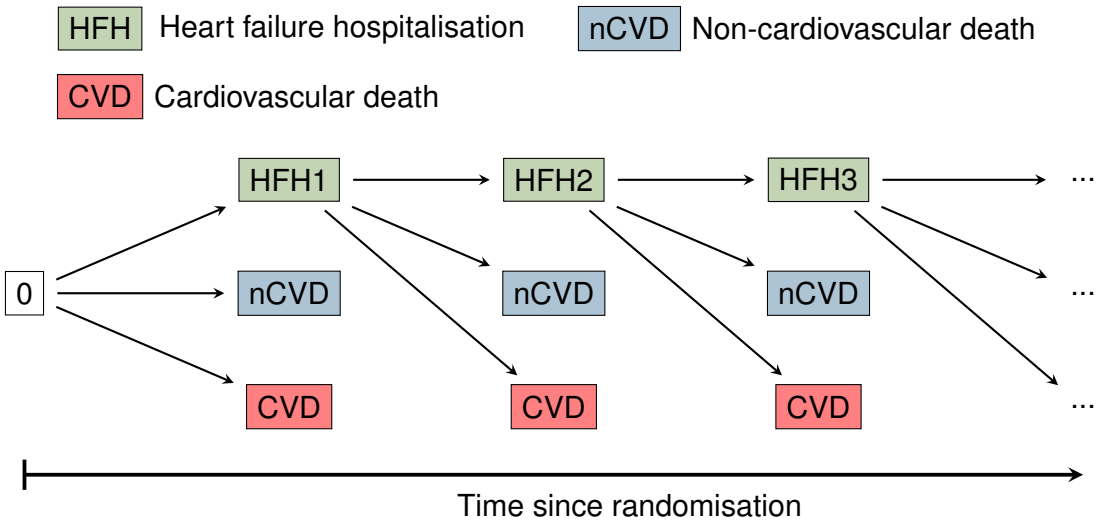

FIGURE 1: Potential event sequences in clinical heart failure trials, illustrated as a multistate process.

2. Censoring: Censoring refers to any circumstance that prevents the further observation but not the occurrence of the event of interest. This implies both administrative censoring due to study end and lost to follow-up due to study-withdrawal from the patient's side.

Figure 2 exemplarily shows the observed event profiles of three patients within a HF-trial. While patients B and C experience three and two HFHs before dying within the study period, patient A has three HFHs and is still alive at the end of the administered follow-up duration of three years.

A rather complex type of analysis for the HFH-outcome would be to consider different treatment effects for each transition, i.e. one for the transition to first HFH, one for the transition from first to second HFH and so on. Such an approach is enabled by multistate regression models (Cook and Lawless, 2018). However, it is questionable whether these transition-specific effects largely differ. In addition, a single summary effect on the HFH-outcome is more catchy and therefore requested in clinical trials. Recurrent event regression models are preferred, as they quantify the treatment effect on the HFH-outcome in terms of a single rate ratio. Prominent representatives of that model class are the Andersen-Gill model and its strongly related robust version, the LWYY model (Andersen and Gill, 1982; Lin *et al.*, 2000). They are natural extensions of

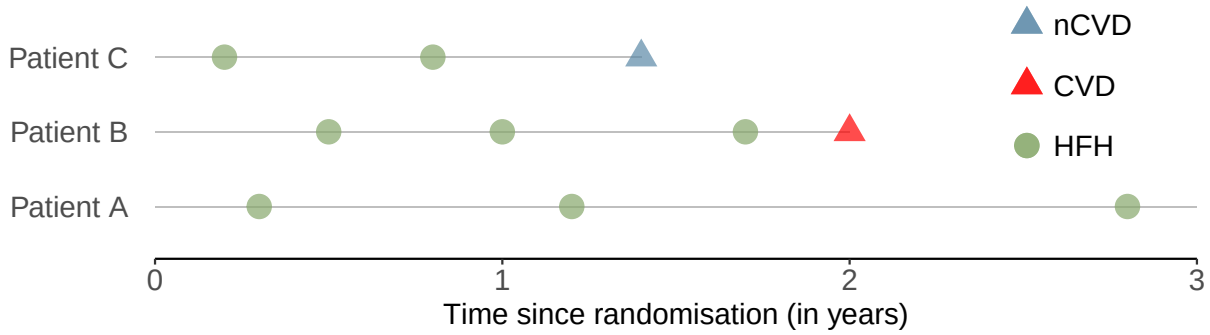

FIGURE 2: Observed event profiles of three exemplary patients within a heart failure trial.

the Cox model for recurrent events. Both analysis methods yield the same estimate for the treatment effect and only differ with regard to standard error calculation. For that reason, they will henceforth be summarized under the term "Andersen-Gill approach". An important assumption within the Andersen-Gill approach is, that a patient's recurrent event rate at a certain time point does not depend on the number and timing of previous recurrent events. In addition, that type of analysis does not specifically account for the competing risk of death. The latter is therefore technically handled as censoring.

Association of time-to-event outcomes The special feature of regression models for time-to-event data, including those for recurrent event outcomes like HFH, is their ability to deal with censored observations. A key modeling assumption in all time-to-event approaches is, that censoring is non-informative. This means, broadly speaking, that the censoring pattern in each subgroup defined by the modeled covariates is completely random and thus in particular not related to the outcome of interest (Emura *et al.*, 2019, p. 16 ff.). In a two-armed randomised clinical trial the binary treatment indicator is often chosen as the only covariate in the model (Ciolino *et al.*, 2019). Then the censoring pattern in both groups needs to be completely random to meet the model assumption. As outlined above, censoring in the proper sense does not comprise competing events. However, within regression models the occurrence of competing events can technically be handled like censoring, as long as the competing events are non-informative for the modeled time-to-event outcome. With regard to the HFH-outcome there is agreement that non-cardiovascular death meets that condition. On the con-

trary, it is questionable whether this assumption likewise applies to cardiovascular death. A positive association between the CVD- and HFH-rates seems plausible and has been verified in reanalyses of some major heart failure trials (Rogers *et al.*, 2012, 2014a,b, 2016). Patients with a high HFH-rate (and consequently a dense sequence of HFHs within their follow-up) tend to drop out early due to CVD – such pattern can be observed both in the treatment and in the control group. Handling CVD like censoring in the Andersen-Gill approach is thus likely violating a key modeling assumption.

How an association between the HFH- and CVD-outcomes manifests in the data is illustrated in Figure 3. It shows simulated data of a HF-trial with 500 patients per group, both for scenarios with and without an association between the HFH- and CVD-rates. In order to clearly visualize the differences, a planned follow-up period of three years is assigned to all patients and prior drop-outs are only possible through death during that time. Hence censoring is purely administrative. The treatment effects are equal in both scenarios, namely protective for HFH and CVD and neutral for the nCVD-outcome. Here the term "treatment effect" refers to the rate ratio between a treated and an untreated patient, who otherwise share the same pattern of known and unknown risk factors. Hence this quantity is comparing a high-risk (low-risk) patient in the treatment group with a high-risk (low-risk) patient from the control group. In the association-scenario the majority of HFH-events is clustered within a small fraction of patients, namely specifically in those who die early from CVD. Such a pattern is often observed in heart failure trials. In contrast, in the scenario with missing association HFH-events are randomly distributed and show no association with the CVD-outcome. It should be mentioned that the patient-specific HFH-rates have the same degree of between-patient heterogeneity as in the association-setting.

Association of time-to-event outcomes is closely related to the concept of unexplained heterogeneity. The latter refers to heterogeneity in the patient-specific event rates, that cannot be explained by covariates considered in the respective modeling approach (Aalen *et al.*, 2015b; Emura *et al.*, 2019, p. 20 ff.). In a randomised clinical HF-trial at first glance it seems reasonable to choose the binary treatment indicator as the only covariate within a model, because all remaining variables affecting the HFH-outcome (potential confounders) ought to be equally distributed among the two groups. Then even conditional on the covariate, that is, within each of both study arms, the patient-

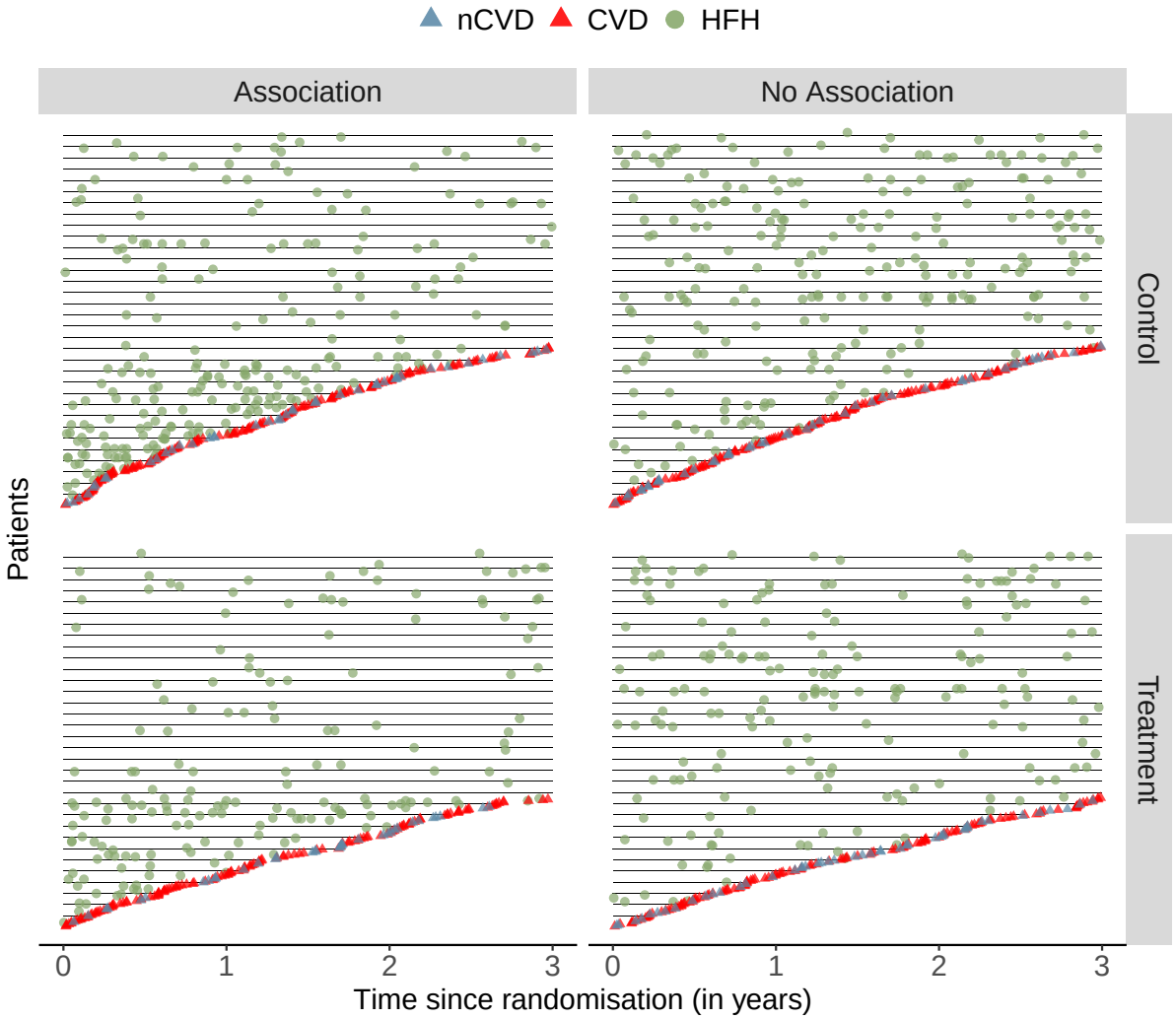

FIGURE 3: Simulated data of a 1:1 randomised, two-armed heart failure trial with 1000 patients (500 per group), each with a planned follow-up of three years. Data were generated both for a scenario with and for a scenario without an association between the HFH- and CVD-rates. Both scenarios have the same degree of heterogeneity among the patient-specific baseline-HFH-rates and the same patient-specific treatment effects (rate ratio HFH: 0.67, hazard ratio CVD: 0.74, hazard ratio nCVD: 1).

specific HFH-rates are most likely heterogeneous. Of course, a smoking, obese patient with diabetes is expected to have a higher HFH-rate than a patient without these risk factors, even if they are equally treated within the trial. Likewise the patient-specific rates for the competing risks CVD and nCVD are supposed to be heterogeneous within both groups. Two time-to-event outcomes are referred to as associated, if there is an association of their risk rates even conditional on modeled covariates. This ought to

apply for the HFH- and the CVD-outcomes if the treatment indicator is considered as the only covariate. In each study arm patients with both high and low HFH- and CVD-rates will participate and those with high (low) HFH-rates will also have a high (low) CVD-rate. Such positive association is induced by unmodeled (known or unknown, at least omitted) risk factors that are affecting the CVD- and HFH-rates in the same direction. There might likewise be some heterogeneity with regard to the nCVD outcome, but most probably without an association to the HFH-outcome.

Randomisation is an important step to achieve structural equality, meaning that potentially confounding variables are equally distributed in both study groups. This is generally thought to assure, that differences observed in the outcome of interest can causally be assigned to the treatment (Greenland, 1990). Thus, there should be no compelling need for confounder-adjustment within models in order to obtain an unbiased estimate for the causal effect. However, in longitudinal trials with time-to-event outcomes things are more complicated: Here structural equality of potentially confounding variables among patients at risk in both groups is generally just guaranteed for the time point of randomisation, but may get lost over time. Hence time-to-event approaches can yield a non-causal treatment effect estimate despite randomisation, if these variables are ommitted in the model (Henderson and Oman, 1999; Aalen *et al.*, 2015a). For univariate time-to-event analysis within a Cox modeling approach that phenomenon is well understood, but less for recurrent event data (HFH) in the presence of an associated terminal event (CVD).

In heart failure trials, structural imbalance of risk factors among patients at risk may arise due to the association of HFH and CVD. How this is evolving, is schematically illustrated in Figure 4. Here a binary risk factor is subdividing the patient-population into a group at low and a group at high risk for HFH. These are assumed to be represented with equal proportions (i.e. 0.5). Positive association among the HFH- and CVD-outcomes is induced, if that risk factor has a similar effect on the CVD-outcome, that is, if the group at high (low) risk for HFH is likewise at high (low) risk for CVD. In contrast, if the risk factor does not affect the CVD-outcome, association is missing. Let for the moment treatment be protective for the CVD-outcome. In both scenarios (existing or missing association) structural equality between the study arms prevails at the time point of randomisation. It persists over time under missing association,

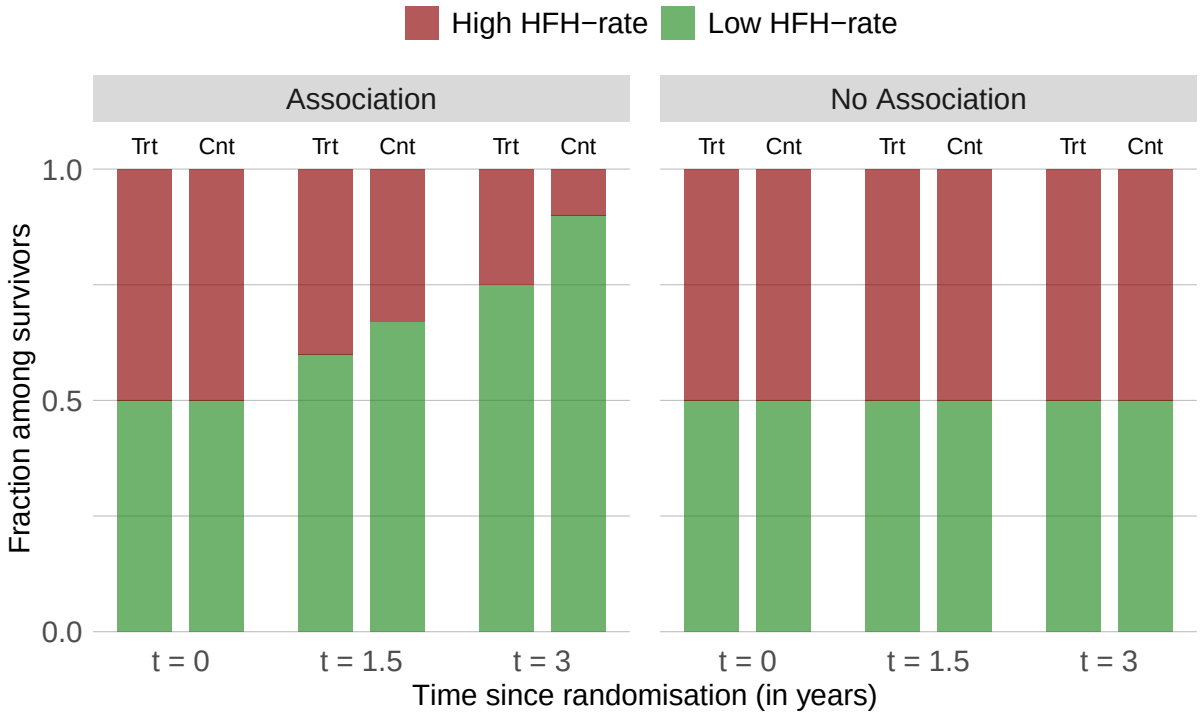

FIGURE 4: Schematic, time-dependent distribution of a binary risk factor among survivors in both study arms (Trt = Treatment, Cnt = Control), both for a scenario with and for a scenario without a (positive) association between the HFH- and CVD-rates. In either scenario treatment is assumed to be protective for the CVD-outcome.

but gets lost in case of association: Patients at high risk for HFH are likewise at high risk for CVD and thus die earlier, leading to a decreasing fraction of high risk patients among survivors within both groups over time. As treatment is protective for CVD, the drop-out of high risk patients proceeds faster in the control group than in the treatment group and the risk factor becomes increasingly imbalanced between the patients at risk over time. That growing structural inequality between the groups needs to be considered by adjusting for the risk factor within the analysis model for HFH, as it will otherwise yield a non-causal treatment effect estimate despite randomisation.

In general, the model for the HFH-outcome needs to be adjusted for all risk factors that are affecting both the HFH- and the CVD-outcome. However, it must be doubted if all these risk factors are known and even available for the analysis. For that reason the use of a joint frailty model was suggested to evaluate the treatment effect on HFHs in chronic heart failure trials (Rogers *et al.*, 2016). Frailty models for some time-to-

event outcome specifically consider heterogeneity in patient-specific event rates that is arising through the existence of unmodeled risk factors. This is achieved through a random effect (called frailty) within the model, which is acting multiplicatively on the event rates. The frailty can be thought of as a single latent variable representing the total effect of all unmodeled risk factors. Each patient is supposed to be frail to a different degree, which is inducing a continuous grading in the patient-specific rates. Frailty models account for that heterogeneity without requiring a specification of the patients' frailty-values in the analysis. The variance of the frailty variable is an important model parameter, whose estimate quantifies the degree of heterogeneity within the studied population (Wienke, 2007). The joint frailty model is an extension of classical frailty models for univariate time-to-event data and specifically targeted for the analysis of a recurrent event outcome (e.g. HFH) in the presence of an associated terminal event outcome (e.g. CVD). The term "joint" refers to the fact, that recurrent and terminal event rates are simultaneously considered within separate sub-models. These are connected by a shared frailty term affecting both outcome-specific rates. This way the joint frailty model accounts for the association between the outcomes. In particular, the terminal event is not just handled as censoring like in the Andersen-Gill approach. Within the joint frailty model the frailty variance is no longer only a measure for heterogeneity, but additionally or rather for the degree of association between recurrent and terminal events (Liu *et al.*, 2004). Although the joint frailty model appears to be a reasonable approach to properly account for the association with CVD in the analysis of the HFH-outcome, its application is so far limited to few post-hoc analyses of clinical HF-trials (Rogers *et al.*, 2012, 2014a,b, 2016). This may be due to the fact, that parameter estimation within that model is quite complex and reliable software-implementations are lacking up to now.

1.4 Scope of this thesis

The present cumulative PhD-thesis consists of two published methodological articles that are addressing important and up to now open issues in the analysis of recurrent HFHs in the presence of CVD as an associated, terminal event. In particular, the articles elaborate the consequences of misspecifications within the Andersen-Gill ap-

proach and derive reliable computational solutions for parameter estimation within joint frailty models. Thus the articles contribute to the ongoing discussion on an appropriate analysis model for the HFH-outcome in chronic heart failure trials. The main questions and scope underlying this research are briefly outlined below:

- **Article I: Marginal hazard ratio estimates in joint frailty models for heart failure trials** (Section 3)

  This article is answering the question, how the treatment effect estimate for the HFH-outcome is adversely affected, if the association between HFH and CVD is ignored in the analysis model. Specifically, the treatment effect estimated by the Andersen-Gill approach is characterized for scenarios with associated outcomes corresponding to joint frailty models. The Andersen-Gill approach is handling CVD like censoring and does hence erroneously not account for the association among HFH- and CVD-rates. It is shown, in which direction and to which extent the estimate deviates from the true patient-specific treatment effect and which factors determine that behavior. For the investigation of these issues least false parameter theory was applied as a methodological tool, that allows to numerically derive asymptotic estimates in misspecified time-to-event models. In addition, simulations were performed for finite sample properties.

  Toenges G, Jahn-Eimermacher A. Marginal hazard ratio estimates in joint frailty models for heart failure trials. *Biometrical Journal*. 2019; 61:1385 - 1401.
  https://doi.org/10.1002/bimj.201800133

- **Article II: Computational issues in fitting joint frailty models for recurrent events with an associated terminal event** (Section 4)

  In this article valid implementations for fitting joint frailty models in SAS® software are derived. Despite being the primarily used statistical software in the highly regulated field of clinical trials, fitting of joint frailty models is not straightforward in SAS and requires specific programming effort as well as an in-depth understanding of the underlying mathematical theory. Guidance on possible approaches to implement the model in SAS was missing up to now. In this article different implementation strategies are developed and their performance is com-

pared within a large simulation study. On this basis recommendations on an optimal software usage in that context are deduced. Results obtained in SAS are additionally contrasted to those from an existing package in R software. The article is supplemented by SAS macros along with manuals, that may encourage statistical practitioners to apply the joint frailty model in clinical trials.

Toenges G, Jahn-Eimermacher A. Computational issues in fitting joint frailty models for recurrent events with an associated terminal event. *Computer Methods and Programs in Biomedicine*. 2020; 188. https://doi.org/10.1016/j.cmpb.2019.105259

The thesis is structured as follows: First, the general statistical framework including in particular the definition of the joint frailty model and different types of recurrent event rates is introduced in section 2. Sections 3 and 4 consist of the two articles, each preceded by some preliminaries. Finally, in section 5 the thesis is concluded with a joint discussion of the results presented in the articles.

2 The statistical framework

This section is first introducing the relevant notations that are needed to formalize the statistical setting. Following the joint frailty model as well as different rate functions will be defined. Finally it will be shown how these rate functions dissolve in scenarios parameterized by the joint frailty model.

2.1 General setting

Let D_i be the terminal event (CVD) time , M_i the number of recurrent events (HFH) in the time interval $[0, D_i]$ and $0 < T_{i1} < T_{i2} < \ldots < T_{iM_i} \leq T_{i0} = \infty$ the corresponding recurrent event times for subject i ($i = 1, \ldots, n$ i.i.d.). These time-to-event data can be represented by a bivariate counting process $N_i(t) = (N_{i1}(t), N_{i2}(t))'$ with components

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

that are counting the events of the respective outcome (recurrent events or terminal event) over time. Here $I(\cdot)$ denotes the indicator function.

With t^- being an infinitesimally smaller time than t and $\mathrm{d}t \to 0$ being an infinitesimally small time increment it follows, that $\mathrm{d}N_{ij}(t) = (N_{ij}((t + \mathrm{d}t)^-) - N_{ij}(t^-))$ is the number of events occurring in the time interval $[t, t + \mathrm{d}t)$ in the outcome-specific process $N_{ij}(t)$ ($j = 1, 2$). Let further $X_i = (X_{i1}, \ldots, X_{ip})'$ be a p-dimensional vector of known covariates and $H_i(t) = \{X_i, N_{i1}(s), N_{i2}(s),\ 0 \leq s < t\}$ the history of the bivariate counting process up to time t. In addition C_i denotes a right-censoring time, implying that subjects are only followed up to time $V_i = \min(D_i, C_i)$.

The bivariate counting process is unambiguously characterized by its intensity process with respect to the history $H_i(t)$, denoted as

$$\Gamma_i(t|H_i(t)) = (\Gamma_{i1}(t|H_i(t)), \Gamma_{i2}(t|H_i(t)))'. \tag{2}$$

In particular, this intensity process specifies the whole dependence structure among recurrent events and the terminal event. Its components are defined as

$$\begin{aligned}\Gamma_{i1}(t|H_i(t))\mathrm{d}t &= E\left[\mathrm{d}N_{i1}(t)|H_i(t)\right] = I(D_i \geq t)E\left[\mathrm{d}N_{i1}(t)|H_i(t), D_i \geq t\right] \\ \Gamma_{i2}(t|H_i(t))\mathrm{d}t &= E\left[\mathrm{d}N_{i2}(t)|H_i(t)\right] = I(D_i \geq t)E\left[\mathrm{d}N_{i2}(t)|H_i(t), D_i \geq t\right].\end{aligned} \tag{3}$$

2.2 The joint frailty model

Conditional on being alive, the rates for recurrent and terminal events might not only depend on the process history (including known covariates), but also on unknown risk factors affecting both outcomes. Known covariates then cannot fully explain the whole range in the population's subject-specific risk rates, which is referred to as unexplained heterogeneity. The joint frailty model considers unexplained heterogeneity by the inclusion of a frailty term Z_i, a positive valued random variable, that is multiplicatively acting on the subject-specific rates for both outcomes (Liu *et al.*, 2004):

$$\begin{aligned}\lambda_{i1}(t|H_i(t), Z_i)\mathrm{d}t &= E\left[\mathrm{d}N_{i1}(t)|H_i(t), Z_i, D_i \geq t\right] = Z_i\lambda_{10}(t)\exp(\beta_1' X_i) \\ \lambda_{i2}(t|H_i(t), Z_i)\mathrm{d}t &= E\left[\mathrm{d}N_{i2}(t)|H_i(t), Z_i, D_i \geq t\right] = Z_i^{\gamma}\lambda_{20}(t)\exp(\beta_2' X_i)\end{aligned} \tag{4}$$

Here $\lambda_{ij}(t|H_i(t), Z_i)$ denotes the outcome-specific event rate conditional on being alive and on the (recurrent) event history including known and unknown covariates. The joint effect of the unknown covariates is modeled by the shared frailty term Z_i, which is inducing an association between the outcomes. Thereby the association-parameter γ is relaxing the assumption of a common shared frailty for recurrent and terminal events.

The parameter vectors $\beta_1 = (\beta_{11}, \ldots, \beta_{1p})'$ and $\beta_2 = (\beta_{21}, \ldots, \beta_{2p})'$ denote the regression coefficients of the p covariates X_i. For the frailty Z_i, commonly a gamma or a lognormal distribution with variance $Var\left[Z_i\right] = \theta$ and, in order to avoid identifiability problems, with a fixed mean $E\left[Z_i\right] = 1$ is adopted. Furthermore, the frailty is modeled as being independent of covariates X_i and of the censoring time C_i. The latter is additionally assumed to be non-informative. Depending on the specification of the conditional baseline hazards $\lambda_{10}(t)$ and $\lambda_{20}(t)$, semiparametric (Liu *et al.*, 2004), quasi-semiparametric (Rondeau *et al.*, 2007) and parametric (Liu and Huang, 2008)

techniques were suggested for parameter estimation.

Importantly, frailty only refers to unexplained heterogeneity, that is, it only covers the remaining heterogeneity that persists even conditional on the known covariates. In other words: Frailty is modeling the residual heterogeneity in subject-specific risk rates in the subgroups defined by known covariates. Each subject is unambiguously characterized by the combination of its (known) covariates and its (unknown) subject-specific frailty term. For that reason the rates targeted within the joint frailty model, that are defined conditional on unknown frailty, may also be termed as subject-specific rates or subject-specific hazards.

Within the joint frailty model, covariates have a proportional effect on the subject-specific rates. Another crucial assumption is, that the rates of both outcomes do not depend on the history of recurrent events on the subject's level (i.e. conditional on frailty). Hence a subject's risk rates at a certain time point are thought to be unaffected by the number of previous recurrent events.

2.3 Rate functions

In the context of recurrent events, in particular if they can be stopped by a terminal event, different rate functions may be targeted for modeling purposes. In the following, these rates will be first generally defined and second explicitly derived for joint frailty scenarios.

Definition of important rate functions As introduced before, the joint frailty model targets the rates

$$\begin{aligned} \lambda_{i1}(t|H_i(t), Z_i)\mathrm{d}t &= E\left[\mathrm{d}N_{i1}(t)|H_i(t), Z_i, D_i \geq t\right] \\ \lambda_{i2}(t|H_i(t), Z_i)\mathrm{d}t &= E\left[\mathrm{d}N_{i2}(t)|H_i(t), Z_i, D_i \geq t\right], \end{aligned} \tag{5}$$

that is, the outcome-specific rates conditional on being alive, on the (recurrent) event history and on both known and unknown covariates. These will be termed as "subject-specific rates/hazards" or "conditional rates/hazards". Of course, this is not a defini-

tion in the proper sense, as it already contains the joint frailty modeling assumption, that the effect of unknown covariates can be condensed to the scalar frailty term Z_i.

Another important class of outcome-specific rates is defined conditional on being alive, on the (recurrent) event history and on known covariates:

$$\begin{aligned} \bar{\lambda}_{i1}(t|H_i(t))\mathrm{d}t &= E\left[\mathrm{d}N_{i1}(t)|H_i(t), D_i \geq t\right] \\ \bar{\lambda}_{i2}(t|H_i(t))\mathrm{d}t &= E\left[\mathrm{d}N_{i2}(t)|H_i(t), D_i \geq t\right] \end{aligned} \tag{6}$$

Here the condition contains all the process information that is available at the respective time point, which does not imply unknown covariates (frailty). Due to their close relation to the intensities

$$\begin{aligned} \Gamma_{i1}(t|H_i(t)) &= I(D_i \geq t)\bar{\lambda}_{i1}(t|H_i(t)) \\ \Gamma_{i2}(t|H_i(t)) &= I(D_i \geq t)\bar{\lambda}_{i2}(t|H_i(t)) \end{aligned} \tag{7}$$

$\bar{\lambda}_{i1}$ and $\bar{\lambda}_{i2}$ will be termed as "intensity rates".

A third class of outcome-specific rates arises by only conditioning on being alive and on known covariates. Due to the missing incorporation of the recurrent event history and the frailty, these rates represent, for each time point, the average of the respective subject-specific rates among survivors. They will be referred to as "marginal rates/hazards":

$$\begin{aligned} r_{i1}(t|X_i)\mathrm{d}t &= E\left[\mathrm{d}N_{i1}(t)|X_i, D_i \geq t\right] \\ r_{i2}(t|X_i)\mathrm{d}t &= E\left[\mathrm{d}N_{i2}(t)|X_i, D_i \geq t\right] \end{aligned} \tag{8}$$

Rates in joint frailty scenarios Within the joint frailty model, the subject-specific rates $\lambda_{ij}(t|H_i(t), Z_i)$ are defined and modeled as shown in formula (4). Henceforth the joint frailty model is taken to be the true data-generating process. Although the intensity rates $\bar{\lambda}_{ij}(t|H_i(t))$ and the marginal rates $r_{ij}(t|X_i)$ are not explicitly defined within the joint frailty model, they are implicitly prescribed and will be dissolved in the following. Details on the mathematical derivations are given in appendices A and B.

The intensity rates are given by

$$\begin{aligned}\bar{\lambda}_{i1}(t|H_i(t)) &= E\left[Z_i|H_i(t), D_i \geq t\right]\lambda_{10}(t)\exp(\beta_1' X_i)\\ \bar{\lambda}_{i2}(t|H_i(t)) &= E\left[Z_i^{\gamma}|H_i(t), D_i \geq t\right]\lambda_{20}(t)\exp(\beta_2' X_i).\end{aligned} \tag{9}$$

Thereby the conditional expectations of the frailty dissolve as follows:

$$\begin{aligned}E\left[Z_i|H_i(t), D_i \geq t\right] &= \frac{\int_0^{\infty} z_i^{N_{i1}(t^-)+1}\exp(-\int_0^t z_i\lambda_{10}(u)\exp(\beta_1' X_i)\mathrm{d}u)}{\times\exp(-\int_0^t z_i^{\gamma}\lambda_{20}(u)\exp(\beta_2' X_i)\mathrm{d}u)f(z_i)\mathrm{d}z_i}\\ &\quad\frac{\int_0^{\infty} z_i^{N_{i1}(t^-)}\exp(-\int_0^t z_i\lambda_{10}(u)\exp(\beta_1' X_i)\mathrm{d}u)}{\times\exp(-\int_0^t z_i^{\gamma}\lambda_{20}(u)\exp(\beta_2' X_i)\mathrm{d}u)f(z_i)\mathrm{d}z_i}\\ E\left[Z_i^{\gamma}|H_i(t), D_i \geq t\right] &= \frac{\int_0^{\infty} z_i^{N_{i1}(t^-)+\gamma}\exp(-\int_0^t z_i\lambda_{10}(u)\exp(\beta_1' X_i)\mathrm{d}u)}{\times\exp(-\int_0^t z_i^{\gamma}\lambda_{20}(u)\exp(\beta_2' X_i)\mathrm{d}u)f(z_i)\mathrm{d}z_i}\\ &\quad\frac{\int_0^{\infty} z_i^{N_{i1}(t^-)}\exp(-\int_0^t z_i\lambda_{10}(u)\exp(\beta_1' X_i)\mathrm{d}u)}{\times\exp(-\int_0^t z_i^{\gamma}\lambda_{20}(u)\exp(\beta_2' X_i)\mathrm{d}u)f(z_i)\mathrm{d}z_i}\end{aligned} \tag{10}$$

Here $f(z_i)$ denotes the probability density function of the frailty variable Z_i. As evident, the intensity rates $\bar{\lambda}_{ij}(t|H_i(t))$ depend on the recurrent event history in terms of the number of previous recurrent events. This is reflecting the frailty-induced association of the two counting processes. However, the time points of previous recurrent events are irrelevant:

$$\bar{\lambda}_{ij}(t|H_i(t))\mathrm{d}t = E\left[\mathrm{d}N_{ij}(t)|H_i(t), D_i \geq t\right] = E\left[\mathrm{d}N_{ij}(t)|N_{i1}(t^-), X_i, D_i \geq t\right] \tag{11}$$

The marginal rates within the joint frailty are given by

$$\begin{aligned}r_{i1}(t|X_i) &= E\left[Z_i|X_i, D_i \geq t\right]\lambda_{10}(t)\exp(\beta_1' X_i)\\ r_{i2}(t|X_i) &= E\left[Z_i^{\gamma}|X_i, D_i \geq t\right]\lambda_{20}(t)\exp(\beta_2' X_i).\end{aligned} \tag{12}$$

Thereby the conditional expectations of the frailty dissolve as follows:

$$
\begin{aligned}
E\left[Z_i | X_i, D_i \geq t\right] &= \frac{\int_0^\infty z_i \exp(-\int_0^t z_i^\gamma \lambda_{20}(u) \exp(\beta_2' X_i) \mathrm{d}u) f(z_i) \mathrm{d}z_i}{\int_0^\infty \exp(-\int_0^t z_i^\gamma \lambda_{20}(u) \exp(\beta_2' X_i) \mathrm{d}u) f(z_i) \mathrm{d}z_i} \\
E\left[Z_i^\gamma | X_i, D_i \geq t\right] &= \frac{\int_0^\infty z_i^\gamma \exp(-\int_0^t z_i^\gamma \lambda_{20}(u) \exp(\beta_2' X_i) \mathrm{d}u) f(z_i) \mathrm{d}z_i}{\int_0^\infty \exp(-\int_0^t z_i^\gamma \lambda_{20}(u) \exp(\beta_2' X_i) \mathrm{d}u) f(z_i) \mathrm{d}z_i}
\end{aligned}
\tag{13}
$$

Whereas the conditional rates $\lambda_{ij}(t|H_i(t))$ are proportional with respect to the known covariates X_i, this property gets in general lost on the level of the intensity rates $\bar{\lambda}_{ij}(t|H_i(t))$ and marginal rates $r_{ij}(t|X_i)$. However, the Andersen-Gill approach is relying on proportionality of the marginal rate for the recurrent event outcome. How a violation of that modeling assumption is affecting the estimates, will be shown next.

3 Article I: Marginal hazard ratio estimates in joint frailty models for heart failure trials

Often the Andersen-Gill approach is used to evaluate the effect of covariates on a recurrent event outcome, e.g. to assess the effect of a treatment on the rate of recurrent hospitalisations in clinical heart failure trials (Rogers *et al.*, 2012, 2014a,b, 2016). Thereby both non-cardiovascular and cardiovascular death are handled as censoring. How an association of the CVD- and HFH-rates is affecting the treatment effect estimate resulting from that analysis is investigated in the first article.

3.1 Preliminaries

The Andersen-Gill estimator can be derived from two different semiparametric models for recurrent event rates, the Andersen-Gill model and the LWYY model. Both do not account for a terminal event like death in their original formulation. Thus subjects are thought to be permanently at risk for another recurrent event. However, inference in these models transfers to a scenario with a terminal event in a straightforward manner by conditioning the target-rate on being alive in the model formulation. Practically, death is then handled as censoring within the estimator. Following, the Andersen-Gill and the LWYY model will be introduced.

The Andersen-Gill model is a semiparametric, intensity-based model for a recurrent event outcome. It specifies the intensity rate of the recurrent event outcome as

$$\bar{\lambda}_{i1}(t|H_i(t)) = \bar{\lambda}_{10}(t)\exp(\bar{\beta}_1' X_i) \tag{14}$$

(Andersen and Gill, 1982). The intensity rate is assumed to be independent of the recurrent event history and covariates are affecting that rate proportionally. As follows from formulas (9) and (10), both assumptions are in general not met in joint frailty scenarios.

Lin *et al.* (2000) extended the Andersen-Gill model by targeting the marginal recurrent event rate. Their likewise semiparametric model is given by

$$r_{i1}(t|X_i) = r_{10}(t)\exp(\widetilde{\beta}_1' X_i)\ . \tag{15}$$

Here covariates are thought to have a proportional effect on the marginal rate, which does again, according to formulas (12) and (13), in general not comply with joint frailty scenarios. In contrast to the Andersen-Gill model, the LWYY model leaves the dependence structure among recurrent events completely unspecified and is thus less restrictive. The Andersen-Gill model is a special case of the LWYY model, that is, if the Andersen-Gill model holds true likewise the LWYY model applies. On the contrary, the inverse implication is in general not true.

As shown by Lin *et al.* (2000), both models result in exactly the same estimator for the unknown regression coefficients. Following, that estimator will be denoted as the "Andersen-Gill estimator" $\widehat{\beta}_1$. The only difference between the two approaches consists in the calculation of standard errors. Standard errors within the LWYY model are generally denoted as "robust" standard errors, as they are somewhat larger than those of the Andersen-Gill model.

Estimation in semiparametric time-to-event models is generally performed by maximizing a partial likelihood function, which is why corresponding estimators are summarized under the term "maximum partial likelihood estimators" (MPLE). The Andersen-Gill estimator is a MPLE, that is defined as the maximum of the following partial likelihood function:

$$\widehat{\beta}_1 = \underset{\beta\in\mathbb{R}^p}{\operatorname{argmax}}\, L_1(\beta) \quad \text{with} \quad L_1(\beta) = \prod_{i=1}^{n} \prod_{\substack{j \in \\ \{r\in\mathbb{N}|T_{ir}<C_i\}}} \frac{\exp(\beta' X_i)}{\sum_{k=1}^{n} Y_k(T_{ij})\exp(\beta' X_k)} \tag{16}$$

Here $Y_i(t) = I(D_i \geq t, C_i \geq t)$ denotes the at-risk indicator. Equivalently, the Andersen-Gill estimator may be defined as the solution of the estimating equation $U(\beta) = 0$ with

$$U(\beta) = \frac{\partial}{\partial\beta}\log(L_1(\beta))) = \sum_{i=1}^{n}\int_0^\infty Y_i(t)\left[X_i - \frac{\sum_{j=1}^{n} Y_j(t)X_j\exp(\beta' X_j)}{\sum_{j=1}^{n} Y_j(t)\exp(\beta' X_j)}\right] \mathrm{d}N_{i1}(t) \tag{17}$$

being the partial likelihood score function.

If the joint frailty model specifies the true data-generating process, the proportionality assumption contained both in the Andersen-Gill and in the LWYY model is generally violated. Covariate effects are in fact time-dependent on the level of the respective recurrent event rates targeted by these two models. However, the Andersen-Gill approach results in a scalar estimate for each covariate. Asymptotically, that estimate will be referred to as the "least false parameter".

Thorough asymptotic theory for estimation in misspecified Cox models has been derived by Struthers & Kalbfleisch (1986). Lin *et al.* (2000) generalized these results by deriving the asymptotic properties of MPLEs in models for recurrent event counting processes. Next, their main result will first be presented in its whole generality and afterwards be used to characterize the asymptotic behavior of the Andersen-Gill estimator in joint frailty scenarios.

Asymptotic properties of MPLEs (Lin *et al.*, 2000)

Let $M_i(t)$ be an arbitrary counting process and τ_i a stopping time that terminates the further observation of the process [Note: In general, the stopping time is either a censoring time or a terminal event time, as in both cases no further events will be observed]. Let $A_i(t) = I(\tau_i \geq t)$ denote the at-risk indicator and X_i be a vector of p known covariates. The observed data of subject i are given by $\{M_i(t), A_i(t), X_i | 0 \leq t \leq \tau_i\}$. Finally, an i.i.d. sample of n subjects is assumed.
In that situation the MPLE $\widehat{\beta}$, defined as the solution of the score equation

$$\sum_{i=1}^{n} \int_0^{\infty} A_i(t) \left[X_i - \frac{\sum_{j=1}^{n} A_j(t) X_j \exp(\beta' X_j)}{\sum_{j=1}^{n} A_j(t) \exp(\beta' X_j)} \right] \mathrm{d}M_i(t) = 0, \tag{18}$$

will almost surely converge to a least false parameter vector β^*, that is uniquely identified as the solution of the least false parameter equation $g(\beta) = 0$ with

$$g(\beta) = \int_0^{\infty} E \left[A_i(t) \left\{ X_i - \frac{E\left[A_i(t) X_i \exp(\beta' X_i)\right]}{E\left[A_i(t) \exp(\beta' X_i)\right]} \right\} \mathrm{d}M_i(t) \right]. \tag{19}$$

Here $E[\cdot]$ denotes an expectation taken with respect to the covariates X_i, to the at-risk process $A_i(t)$ and to the counting process $M_i(t)$.

This result directly applies to the Andersen-Gill estimator $\widehat{\beta}_1$ by choosing $A_i(t) = Y_i(t) = I(D_i \geq t, C_i \geq t)$ as at-risk indicator and $M_i(t) = N_{i1}(t)$ as counting process. Consequently the least false parameter vector, that will henceforth be denoted as β_1^*, is given by the solution of $g(\beta) = 0$ with

$$g(\beta) = \int_0^\infty E\left[Y_i(t)\left\{X_i - \frac{E\left[Y_i(t)X_i\exp(\beta' X_i)\right]}{E\left[Y_i(t)\exp(\beta' X_i)\right]}\right\}\mathrm{d}N_{i1}(t)\right]. \tag{20}$$

In order to investigate the situation of a two-armed randomised clinical trial, the focus is on a single binary covariate $X_i \sim Bin(1,p)$ from now on. In that situation formula (20) dissolves to the simplified least false parameter function

$$\begin{aligned} g(\beta) = \int_0^\infty \frac{\bar{y}_i^{(0)}(t)\bar{y}_i^{(1)}(t)}{\bar{y}_i^{(0)}(t) + \bar{y}_i^{(1)}(t)\exp(\beta)}\Big[&E\left[\mathrm{d}N_{i1}(t)|X_i = 1, Y_i(t) = 1\right] \\ &- E\left[\mathrm{d}N_{i1}(t)|X_i = 0, Y_i(t) = 1\right]\exp(\beta)\Big]. \end{aligned} \tag{21}$$

Here $\bar{y}_i^{(k)}(t) = P(X_i = k, Y_i(t) = 1)$ denotes the probability, that a subject 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 C. As the censoring time C_i is throughout assumed to be non-informative, the rate $E\left[\mathrm{d}N_{i1}(t)|X_i, Y_i(t) = 1\right]$ appearing in the least false parameter function (21) is actually the marginal recurrent event rate r_{i1} (see formula 8) targeted by the LWYY model (see formula 15):

$$E\left[\mathrm{d}N_{i1}(t)|X_i = k, Y_i(t) = 1\right] = E\left[\mathrm{d}N_{i1}(t)|X_i = k, D_i \geq t\right] = r_{i1}(t|X_i = k) \tag{22}$$

How the marginal recurrent event rate dissolves in joint frailty scenarios was already shown in formulas (12) and (13). If the LWYY model holds true, the least false parameter coincides with the proportionality-parameter given in the model formulation. However, if covariate effects on the marginal recurrent event rate are time-dependent (as generally the case in joint frailty scenarios), the least false parameter β_1^* represents a time-averaged effect on the marginal recurrent event rate. In the first paper of this thesis, the asymptotic behavior of the Andersen-Gill estimator in joint frailty scenarios

with a single binary covariate is investigated by numerically solving the least false parameter equation $g(\beta) = 0$. Please note, that the marginal recurrent event rate $r_{i1}(t|X_i)$ is erroneously described as being equal to the intensity rate $\bar{\lambda}_{i1}(t|H_i(t))$ within the article. This does not affect any further derivations, results or conclusions as these are consistently based on $r_{i1}(t|X_i)$ only.

3.2 Article

Toenges G, Jahn-Eimermacher A. Marginal hazard ratio estimates in joint frailty models for heart failure trials. *Biometrical Journal*. 2019; 61:1385 - 1401. https://doi.org/10.1002/bimj.201800133

Link to publication:

https://doi.org/10.1002/bimj.201800133

Link to publication:

https://doi.org/10.1002/bimj.201800133

Link to publication:

https://doi.org/10.1002/bimj.201800133

Link to publication:

https://doi.org/10.1002/bimj.201800133

Link to publication:

https://doi.org/10.1002/bimj.201800133

Link to publication:

https://doi.org/10.1002/bimj.201800133

Link to publication:

https://doi.org/10.1002/bimj.201800133

Link to publication:

https://doi.org/10.1002/bimj.201800133

Link to publication:

https://doi.org/10.1002/bimj.201800133

Link to publication:

https://doi.org/10.1002/bimj.201800133

Link to publication:

https://doi.org/10.1002/bimj.201800133

Link to publication:

https://doi.org/10.1002/bimj.201800133

Link to publication:

https://doi.org/10.1002/bimj.201800133

Link to publication:

https://doi.org/10.1002/bimj.201800133

Link to publication:

https://doi.org/10.1002/bimj.201800133

Link to publication:

https://doi.org/10.1002/bimj.201800133

Link to publication:

https://doi.org/10.1002/bimj.201800133

4 Article II: Computational issues in fitting joint frailty models for recurrent events with an associated terminal event

Although drop-outs due to CVD in heart failure trials are probably informative for the HFH-outcome and the joint frailty model seems to be a reasonable model for the disease process, examples of application remain rare in the literature and are often limited to journals with a focus on mathematical statistics. This might be due to the fact that user-friendly software-implementations and guidance on the choice of computational options is missing for statistical practitioners. This particularly applies to SAS®, which is the primarily used statistical software in the strongly regulated field of clinical trials. In the second article of this thesis different approaches of implementing joint frailty models in SAS software are derived and their performance is compared among each other, but also with existing functions in the statistical software R. SAS macros for the generation of joint frailty data and for fitting joint frailty models are provided as supplementary to that paper, along with manuals for a proper usage.

4.1 Preliminaries

The joint frailty term imposes some computational challenges for fitting joint frailty models. Several statistical methods for parameter estimation were suggested in the literature. These differ with respect to the specification of the conditional baseline hazards and can be classified into parametric, semiparametric and quasi-semiparametric approaches. Semiparametric models leave the baseline hazards completely unspecified. Here the Monte Carlo EM-algorithm was proposed for parameter estimation (Huang and Wolfe, 2002; Liu *et al.*, 2004). In parametric models baseline hazards are specified parametrically (for example as being constant). Then the maximization of the (marginal) likelihood function is the method of choice for model fitting (Liu and Huang, 2008). Quasi-semiparametric models are somewhat in between the parametric and semiparametric approach. Here splines are used to model the baseline hazards, which is, in a strict sense, also a parametric specification. However, for parameter es-

timation a penalized likelihood is used, which controls the trade-off between data-fit and smoothness (Rondeau *et al.*, 2007). Hence this approach needs to be differentiated from the purely parametric one.

A major drawback in the field of joint frailty models is the lack of reliable and user-friendly software-implementations. In particular, functions for a semiparametric model fit do not exist at all in any software. In the statistical software R, the frailtypack package provides functions that allow for parametric or quasi-semiparametric estimation (Rondeau *et al.*, 2012). In contrast, SAS software does not contain any procedure, that allows for a straightforward, direct model fit. However, in SAS users are enabled to computationally implement parametric joint frailty models on their own by using specific tools of the NLMIXED procedure. This requires considerable programming effort and a tricky decomposition of the model's likelihood. Implementation becomes even more complicated if another frailty distribution as lognormal (e.g. gamma) shall be applied, because NLMIXED only allows for normally distributed random effects. Specific programming strategies, referred to as "Likelihood Reformulation" (LR) and "Probability Integral Transformation" (PIT), are needed to circumvent that restriction in SAS (Nelson *et al.*, 2006; Liu and Yu, 2008). Finally, the NLMIXED procedure comes along with a vast abundance of numerical options for the estimation process. The second article of this thesis is addressing these issues first by first deriving valid SAS-implementations of parametric joint frailty models and second by comparing them within a large simulation study. This way the article provides guidance on optimal software usage to pave the way for an application of that model in clinical trials.

4.2 Article

Toenges G, Jahn-Eimermacher A. Computational issues in fitting joint frailty models for recurrent events with an associated terminal event. *Computer Methods and Programs in Biomedicine*. 2020; 188. https://doi.org/10.1016/j.cmpb.2019.105259

Link to publication:

https://doi.org/10.1016/j.cmpb.2019.105259

Link to publication:

https://doi.org/10.1016/j.cmpb.2019.105259

Link to publication:

https://doi.org/10.1016/j.cmpb.2019.105259

Link to publication:

https://doi.org/10.1016/j.cmpb.2019.105259

Link to publication:

https://doi.org/10.1016/j.cmpb.2019.105259

Link to publication:

https://doi.org/10.1016/j.cmpb.2019.105259

Link to publication:

https://doi.org/10.1016/j.cmpb.2019.105259

Link to publication:

https://doi.org/10.1016/j.cmpb.2019.105259

Link to publication:

https://doi.org/10.1016/j.cmpb.2019.105259

Link to publication:

https://doi.org/10.1016/j.cmpb.2019.105259

Link to publication:

https://doi.org/10.1016/j.cmpb.2019.105259

Link to publication:

https://doi.org/10.1016/j.cmpb.2019.105259

Link to publication:

https://doi.org/10.1016/j.cmpb.2019.105259

5 Discussion

Hospitalisations due to chronic heart failure (HFH) pose a huge burden not only for patients, but also for health care providers. Therefore they present a major outcome of randomised (normally two-armed) clinical trials performed in that indication area (Packer, 2016). In such trials, patients can be hospitalised multiple times and their follow-up durations vary due to death (CVD = cardiovascular death, nCVD = non-cardiovascular death) or censoring (i.e. lost to follow-up or end of study). Statistical methods for the analysis of recurrent events in the presence of censoring are well established. The Andersen-Gill approach (Andersen and Gill, 1982; Lin *et al.*, 2000) is a prominent example and frequently applied to assess the treatment effect on the HFH-outcome in clinical trials. Here death is technically handled as censoring, which is only a valid practice if death is non-informative for the HFH-rate. However, data of large clinical trials confirm that the patient-specific HFH- and CVD-rates are associated, meaning that patients at high risk for HFH are likewise at high risk for CVD and accordingly drop out earlier than those at low risk for both outcomes (Rogers *et al.*, 2012, 2014a,b, 2016). This is violating a key modeling assumption of the Andersen-Gill approach. In this thesis it is shown, how that violation can adversely affect statistical inference, in particular the treatment effect estimate. These results call for the need of more complex, but due to a lack of reliable software only rarely applied joint frailty models (Liu *et al.*, 2004; Rogers *et al.*, 2016). For that reason valid implementations of that model in SAS® software are derived and quality-checked within this thesis to pave the way for its application in the practice of clinical heart failure trials.

An association of the patients' HFH- and CVD-rates is induced by risk factors that exert effects of equal direction on both outcomes. As long as these risk factors are all known and available in the analysis, the association of the patients' risk rates can be fully explained and standard techniques like the Andersen-Gill approach, where CVD is technically handled as censoring, can be used to assess a treatment effect on the patient-specific HFH-rates. Despite randomisation the adjustment for all these shared risk factors within the regression model is necessary to measure the causal effect of treatment on the patients' level. However, it is unlikely that all these risk factors are known and available – there will always remain unexplained heterogeneity in the risk

rates for both outcomes. Moreover, most clinical trials only report unadjusted treatment effect estimates, demonstrating that even the known and available risk factors often remain unconsidered (Ciolino *et al.*, 2019). The joint frailty model specifically accounts for the residual association between the outcomes that cannot be explained by the modeled covariates. In particular, CVD is therein not just handled as censoring, but as a separate outcome that is jointly modeled with the HFH-outcome (Liu *et al.*, 2004).

In the first article of this thesis it is shown, that substantial differences between the true patient-specific effect on the HFH-outcome and the estimate resulting from the Andersen-Gill approach may arise in the presence of an unexplained association between the patients' HFH- and CVD-rates. In clinically relevant scenarios, characterized through protective patient-specific treatment effects on both outcomes, the estimate attenuates with the trial duration. As an example, two trials with median follow-up durations of 2 and 4 years could yield estimates of 0.84 and 0.89 in the Andersen-Gill approach – despite having an equal patient-specific effect of 0.8 for the HFH-outcome. Of course, this hampers the comparison of estimates resulting from trials with different lengths of follow-up. A further important result of this work is, that the Andersen-Gill approach only quantifies the patient-specific treatment effect, if cardiovascular mortality is equal in both study arms. This is empirically supported by results of the CHARM-Preserved trial, where treatment did not affect CVD and estimates resulting from the Andersen-Gill and joint frailty analysis coincide (Rogers *et al.*, 2014a; Yusuf *et al.*, 2003). Indeed a missing effect on CVD is not uncommon in trials on new heart failure treatments nowadays (Zheng *et al.*, 2018). However, a statistical analysis method that guarantees to yield the effect of interest (i.e. the patient-specific effect) only under this requirement and otherwise depends on the trial duration cannot be considered as an appropriate one.

This thesis also provides thorough mathematical derivations for the findings described above. First it is shown, that patient-specific HFH-rates need to be distinguished from the marginal HFH-rate. The marginal HFH-rate at a certain time point is given by the average of the individual HFH-rates of those patients being still under observation at that time point. The ratio of the marginal rates between the treatment and control group constitutes the treatment effect on the population's level, referred to as marginal treatment effect. Second it is demonstrated, that the marginal HFH-rates are in gen-

eral non-proportional, that is, the marginal treatment effect is time-dependent, if the patient-specific CVD- and HFH-rates are associated according to a joint frailty model. Hence a constant treatment effect on the patient's level goes in general hand in hand with a time-dependent treatment effect on the population's level in case of association between the HFH- and CVD-outcome. Both kinds of effects are equally true and may be targeted in clinical trials. The Andersen-Gill approach is targeting the treatment effect on the marginal HFH-rate, but erroneously specifies it as being constant over time. As a consequence, the resulting estimate represents a weighted average of the time-dependent marginal treatment effect function, with weights determined mainly by the censoring scheme. This particularly explains, that the estimate depends on the trial duration. This third main finding was obtained by using least false parameter theory, which allows to derive estimates of misspecified, semiparametric time-to-event models numerically (Struthers and Kalbfleisch, 1986; Lin *et al.*, 2000). That methodology turned out to be a powerful tool to study estimates of misspecified models and was, apart from this work, only recently applied in one additional article (Zhong and Cook, 2019) in the context of recurrent events so far. Up to now the consequences of unexplained between-patient heterogeneity were only well understood in the context of univariate survival analysis within the Cox model (Henderson and Oman, 1999; Aalen *et al.*, 2015a). This thesis reveals for the first time, that recurrent event analysis with the Andersen-Gill approach is similarly affected by an unexplained association between the recurrent (i.e. HFH) and a terminal (i.e. CVD) event outcome.

Faced to the previously described drawbacks of the Andersen-Gill approach it may be more reasonable to target the patient-specific treatment effect by using the joint frailty model for the analysis. However, application of that model is so far limited to few post-hoc analyses of clinical trials (e.g. Rogers *et al.*, 2012, 2014a,b, 2016) and was not yet used as the main analysis model to the author's knowledge. That caution may be due to the lack of reliable software implementations for parameter estimation within that model. This particularly applies to the statistical software SAS, whose usage is obligatory in the highly regulated field of clinical trials. Although the joint frailty model was first suggested in 2004 (Liu *et al.*, 2004), there exists no procedure that allows for a direct joint frailty model fit in SAS. Up to now only a single, hardly customizable and even incorrect SAS macro (Lu and Liu, 2008) for that purpose was available in the

literature.

For that reason in the second article solid implementations of the joint frailty model in SAS software are derived and made available to statistical practitioners in the form of a SAS macro. The elaborated implementations make use of tools from the SAS NLMIXED procedure (Kurada, 2016; SAS Institute Inc., 2015), that requires the user to provide a specific decomposition of the model's conditional log-likelihood by appropriate programming statements (McTernan and Blozis, 2013). As the NLMIXED procedure is restricted to a lognormal distribution for the frailty, additional programming strategies, referred to as "Probability Integral Transformation" (PIT, see Nelson *et al.*, 2006) and "Likelihood Reformulation" (LR, see Liu and Yu, 2008), need to be applied to fit a model with other frailty distributions. In a first step of this work, appropriate decompositions of the joint frailty model's conditional log-likelihood were derived both for a lognormal and a gamma distributed frailty. The actual estimation process within the NLMIXED procedure can be further customized, for example by specifying starting values for the parameters or by specifying the numerical technique to solve integrals within the marginal likelihood (e.g. choice of quadrature rule, quadrature points). The combination of all these aspects demonstrates the wealth of possible implementations to fit joint frailty models in SAS software. How these differ with respect to the reliability of the output, was investigated by means of a large simulation study as a second step within the article.

It could be shown, that the performance of an implementation is generally insensitive to the choice of starting values, but sensitive to the settings chosen for numerical integration, i.e. the quadrature rule and the number of quadrature points. Adaptive quadrature is superior to non-adaptive quadrature and the quality of the output is increasing with the number of quadrature points. This is in line with literature from the field of numerical analysis, which states that adaptive quadrature is more effective for "badly behaved" integrands than other quadrature approaches (Rice, 1975). Large differences became apparent with respect to the method to deal with a gamma-distributed frailty. The LR-method performs significantly worse than the PIT-method. Whereas both approaches yield unbiased treatment effect estimates, the remaining (nuisance) parameter estimates are strongly biased in case of the LR-method. This can be partially remedied by increasing the number of quadrature points. That finding runs con-

trary to that of Liu and Yu (2008) , who first proposed the LR method and showed an equal performance of the LR- und PIT-strategies in simulations. However, they compared the performance for simpler frailty models and did not evaluate joint frailty models. Finally the performance of the aforementioned SAS implementations was also contrasted to that of an existing package in R software (Rondeau *et al.*, 2012), which enables to fit joint frailty models but is much less customizable than the NLMIXED procedure of SAS. Here likewise biased nuisance parameter estimates could be identified as a big problem, both for a lognormal and gamma distributred frailty. In summary, a NLMIXED-implementation with the PIT method, adaptive quadrature and at least 10 quadrature points was shown to yield best results for a gamma frailty model. In case of a lognormal frailty model, the most reliable performance was reached by a NLMIXED-implementation with adaptive quadrature and at least 5 quadrature points.

As a main result with respect to these computational aspects it can be stated, that treatment effect estimates from joint frailty models are unbiased, regardless of which software and which implementation within the software is used. On the contrary, the quality of the output for the nuisance parameter estimates strongly depends on the chosen implementation. Nuisance parameters are usually not of primary interest in the analysis. However, it is of great importance to obtain likewise unbiased estimates for them, first for correct conclusions on the disease process and second for a proper sample size planning in clinical trials: The required sample size for a sufficiently powered joint frailty analysis has to be determined by simulations, whose input parameters generally rely on joint frailty estimates of pilot data. Hence biased estimates for the nuisance parameters might lead to wrong simulation parameters and, as a consequence, to wrong conclusions regarding the required sample size for the trial. The provided SAS macros and the recommendations on its best settings may help to overcome that issue.

The results of this thesis strongly motivate the use of the joint frailty model. It should however be mentioned, that even in that complex model some of the underlying assumptions could be too simplified and need to be questioned. As an example, the patient-specific HFH-rates are thought to be unaffected by the number of previous HFHs. It could be resonable to relax that assumption by including the number of previous HFHs as a time-dependent covariate into the model. In addition, a shared frailty

term for both processes (even though relaxed through an exponent-parameter) seems unsatisfactory. A bivariate frailty with a certain correlation would certainly better reflect the disease process. However, the second article has shown how sensitive fitting of joint frailty models is with respect to the computational settings. To the author's experience, even more complex models suffer from lacking convergence of the model fitting process. This calls for further improvements of numerical techniques and software solutions.

The methodological results of the first article along with the provided software solutions of the second article will contribute to the ongoing discussion on a proper analysis of recurrent hospitalisations in clinical heart failure trials (e.g. Rogers *et al.*, 2012, 2014a,b, 2016). Moreover, the findings of this thesis are likewise relevant for trials on other indications, that have to deal with similar statistical issues. For example, recurrent heart failure hospitalisations and cardiovascular death also constitute relevant endpoints in trials on type II diabetes (Zinman *et al.*, 2015). Another application could be given by the analysis of recurrent adverse events in safety data of clinical trials (Unkel *et al.*, 2019). Here treatment discontinuation might be considered to be an associated terminal event: As an example, in trials of chronic obstructive pulmonary disease (COPD) the rate of recurrent exacerbations is probably associated with the rate of drug-withdrawal (Keene *et al.*, 2008). Finally, results of this thesis may also be of relevance for the analysis of cancer recurrences in the presence of death as an associated terminal event (Rondeau *et al.*, 2007; Lafourcade *et al.*, 2018).

Appendix

A. Rates in joint frailty scenarios

A.1 Intensity rates

The intensity rate for the recurrent event outcome in a joint frailty model is given by

$$\begin{aligned}
&\bar{\lambda}_{i1}(t|H_i(t))\mathrm{d}t \\
&= E\left[\mathrm{d}N_{i1}(t)|H_i(t), D_i \geq t\right] \\
&= E\left[E\left[\mathrm{d}N_{i1}(t)|H_i(t), Z_i, D_i \geq t\right]|H_i(t), D_i \geq t\right] \\
&= E\left[Z_i\lambda_{10}(t)\exp(\beta_1' X_i)|H_i(t), D_i \geq t\right]\mathrm{d}t \\
&= E\left[Z_i|H_i(t), D_i \geq t\right]\lambda_{10}(t)\exp(\beta_1' X_i)\mathrm{d}t.
\end{aligned} \tag{23}$$

The corresponding intensity rate for the terminal event outcome can be similarly dissolved to

$$\bar{\lambda}_{i2}(t|H_i(t)) = E\left[Z_i^{\gamma}|H_i(t), D_i \geq t\right]\lambda_{20}(t)\exp(\beta_2' X_i). \tag{24}$$

A.2 Marginal rates

The marginal recurrent event rate in a joint frailty model is given by

$$\begin{aligned}
&r_{i1}(t|X_i)\mathrm{d}t \\
&= E\left[\mathrm{d}N_{i1}(t)|X_i, D_i \geq t\right] \\
&= E\left[E\left[\mathrm{d}N_{i1}(t)|X_i, Z_i, D_i \geq t\right]|X_i, D_i \geq t\right] \\
&= E\left[Z_i\lambda_{10}(t)\exp(\beta_1' X_i)|X_i, D_i \geq t\right]\mathrm{d}t \\
&= E\left[Z_i|X_i, D_i \geq t\right]\lambda_{10}(t)\exp(\beta_1' X_i)\mathrm{d}t.
\end{aligned} \tag{25}$$

The corresponding marginal terminal event rate can be similarly dissolved to

$$r_{i2}(t|X_i) = E\left[Z_i^{\gamma}|X_i, D_i \geq t\right] \lambda_{20}(t) \exp(\beta_2' X_i). \tag{26}$$

B. Conditional expectations of frailty in joint frailty scenarios

In the following derivations $f(\cdot|\cdot)$ will denote a (potentially conditional) probability density function of a certain random variable, random vector or counting process.

B.1 Conditional on the event history and being alive

$f(T_{i1}, \ldots, T_{iN_{i1}(t^-)}|X_i, Z_i, D_i \geq t)$ is the conditional density of the outcome

{exactly $N_{i1}(t^-)$ recurrent events in $[0, t)$, namely at time points $T_{i1}, \ldots, T_{iN_{i1}(t^-)}$},

given known covariates X_i, frailty Z_i and being alive at t^-. The counting process $N_{i1}(t|X_i, Z_i, D_i \geq t)$ is the subject-specific recurrent event process that will never be stopped due to death. This counting process is a Poisson process within the joint frailty model. According to Poisson process theory (Cook and Lawless, 2007, p. 47 ff.), the density is then given by

$$\begin{aligned} &f(T_{i1}, \ldots, T_{iN_{i1}(t^-)}|X_i, Z_i = z_i, D_i \geq t) \\ &= \left[\prod_{j=1}^{N_{i1}(t^-)} z_i \lambda_{10}(T_{ij}) \exp(\beta_1' X_i)\right] \exp(-\int_0^t z_i \lambda_{10}(u) \exp(\beta_1' X_i) \mathrm{d}u). \end{aligned} \tag{27}$$

In case of $N_{i1}(t^-) = 0$, the product-term $\left[\prod_{j=1}^{N_{i1}(t^-)} \ldots\right]$ is per definition 1. Using that preliminary result, the expectations of the frailty conditional on the event history and on being alive dissolve as follows:

$$
\begin{aligned}
& E\left[Z_i | H_i(t), D_i \geq t\right] \\
= & \int_0^\infty z_i f(z_i | H_i(t), D_i \geq t) \mathrm{d}z_i \\
= & \frac{\int_0^\infty z_i f(z_i, T_{i1}, \ldots, T_{iN_{i1}(t^-)} | X_i, D_i \geq t) \mathrm{d}z_i}{f(T_{i1}, \ldots, T_{iN_{i1}(t^-)} | X_i, D_i \geq t)} \\
= & \frac{\int_0^\infty z_i f(z_i, T_{i1}, \ldots, T_{iN_{i1}(t^-)} | X_i, D_i \geq t) \mathrm{d}z_i}{\int_0^\infty f(z_i, T_{i1}, \ldots, T_{iN_{i1}(t^-)} | X_i, D_i \geq t) \mathrm{d}z_i} \\
= & \frac{\int_0^\infty z_i f(T_{i1}, \ldots, T_{iN_{i1}(t^-)} | X_i, Z_i = z_i, D_i \geq t) f(z_i | X_i, D_i \geq t) \mathrm{d}z_i}{\int_0^\infty f(T_{i1}, \ldots, T_{iN_{i1}(t^-)} | X_i, Z_i = z_i, D_i \geq t) f(z_i | X_i, D_i \geq t) \mathrm{d}z_i} \\
= & \frac{P(D_i \geq t | X_i) \int_0^\infty \int_t^\infty z_i f(T_{i1}, \ldots, T_{iN_{i1}(t^-)} | X_i, Z_i = z_i, D_i \geq t) f(z_i, d_i | X_i) \mathrm{d}d_i \mathrm{d}z_i}{P(D_i \geq t | X_i) \int_0^\infty \int_t^\infty f(T_{i1}, \ldots, T_{iN_{i1}(t^-)} | X_i, Z_i = z_i, D_i \geq t) f(z_i, d_i | X_i) \mathrm{d}d_i \mathrm{d}z_i} \\
= & \frac{\int_0^\infty \int_t^\infty z_i f(T_{i1}, \ldots, T_{iN_{i1}(t^-)} | X_i, Z_i = z_i, D_i \geq t) f(d_i | X_i, Z_i = z_i) f(z_i | X_i) \mathrm{d}d_i \mathrm{d}z_i}{\int_0^\infty \int_t^\infty f(T_{i1}, \ldots, T_{iN_{i1}(t^-)} | X_i, Z_i = z_i, D_i \geq t) f(d_i | X_i, Z_i = z_i) f(z_i | X_i) \mathrm{d}d_i \mathrm{d}z_i} \\
= & \frac{\int_0^\infty z_i f(T_{i1}, \ldots, T_{iN_{i1}(t^-)} | X_i, Z_i = z_i, D_i \geq t) P(D_i \geq t | X_i, Z_i = z_i) f(z_i) \mathrm{d}z_i}{\int_0^\infty f(T_{i1}, \ldots, T_{iN_{i1}(t^-)} | X_i, Z_i = z_i, D_i \geq t) P(D_i \geq t | X_i, Z_i = z_i) f(z_i) \mathrm{d}z_i} \\
= & \frac{\begin{array}{l} \int_0^\infty z_i \left[\prod_{j=1}^{N_{i1}(t^-)} z_i \lambda_{10}(T_{ij}) \exp(\beta_1' X_i)\right] \exp(-\int_0^t z_i \lambda_{10}(u) \exp(\beta_1' X_i) \mathrm{d}u) \\ \quad \times \exp(-\int_0^t z_i^\gamma \lambda_{20}(u) \exp(\beta_2' X_i) \mathrm{d}u) f(z_i) \mathrm{d}z_i \end{array}}{\begin{array}{l} \int_0^\infty \left[\prod_{j=1}^{N_{i1}(t^-)} z_i \lambda_{10}(T_{ij}) \exp(\beta_1' X_i)\right] \exp(-\int_0^t z_i \lambda_{10}(u) \exp(\beta_1' X_i) \mathrm{d}u) \\ \quad \times \exp(-\int_0^t z_i^\gamma \lambda_{20}(u) \exp(\beta_2' X_i) \mathrm{d}u) f(z_i) \mathrm{d}z_i \end{array}} \\
= & \frac{\int_0^\infty z_i^{N_{i1}(t^-)+1} \exp(-\int_0^t z_i \lambda_{10}(u) \exp(\beta_1' X_i) \mathrm{d}u) \exp(-\int_0^t z_i^\gamma \lambda_{20}(u) \exp(\beta_2' X_i) \mathrm{d}u) f(z_i) \mathrm{d}z_i}{\int_0^\infty z_i^{N_{i1}(t^-)} \exp(-\int_0^t z_i \lambda_{10}(u) \exp(\beta_1' X_i) \mathrm{d}u) \exp(-\int_0^t z_i^\gamma \lambda_{20}(u) \exp(\beta_2' X_i) \mathrm{d}u) f(z_i) \mathrm{d}z_i}
\end{aligned}
\tag{28}
$$

Similarly the conditional expectation $E\left[Z_i^{\gamma}|H_i(t), D_i \geq t\right]$ can be shown to be

$$
\begin{aligned}
& E\left[Z_i^{\gamma}|H_i(t), D_i \geq t\right] \qquad (29)\\
= & \frac{\int_0^{\infty} z_i^{N_{i1}(t^-)+\gamma} \exp(-\int_0^t z_i \lambda_{10}(u) \exp(\beta_1' X_i) \mathrm{d}u) \exp(-\int_0^t z_i^{\gamma} \lambda_{20}(u) \exp(\beta_2' X_i) \mathrm{d}u) f(z_i) \mathrm{d}z_i}{\int_0^{\infty} z_i^{N_{i1}(t^-)} \exp(-\int_0^t z_i \lambda_{10}(u) \exp(\beta_1' X_i) \mathrm{d}u) \exp(-\int_0^t z_i^{\gamma} \lambda_{20}(u) \exp(\beta_2' X_i) \mathrm{d}u) f(z_i) \mathrm{d}z_i}.
\end{aligned}
$$

B.2 Conditional on the covariates and being alive

The expectation of the frailty conditional on known covariates and on being alive dissolves as follows:

$$
\begin{aligned}
& E\left[Z_i|X_i, D_i \geq t\right] \\
= & \int_0^{\infty} z_i f(z_i|X_i, D_i \geq t) \mathrm{d}z_i \\
= & \frac{\int_0^{\infty} \int_t^{\infty} z_i f(z_i, d_i|X_i) \mathrm{d}d_i \mathrm{d}z_i}{P(D_i \geq t|X_i)} \\
= & \frac{\int_0^{\infty} \int_t^{\infty} z_i f(d_i|X_i, Z_i = z_i) f(z_i|X_i) \mathrm{d}d_i \mathrm{d}z_i}{\int_0^{\infty} P(D_i \geq t|X_i, Z_i = z_i) \mathrm{d}z_i} \qquad (30)\\
= & \frac{\int_0^{\infty} z_i P(D_i \geq t|X_i, Z_i = z_i) f(z_i) \mathrm{d}z_i}{\int_0^{\infty} P(D_i \geq t|X_i, Z_i = z_i) f(z_i) \mathrm{d}z_i} \\
= & \frac{\int_0^{\infty} z_i \exp(-\int_0^t z_i^{\gamma} \lambda_{20}(u) \exp(\beta_2' X_i) \mathrm{d}u) f(z_i) \mathrm{d}z_i}{\int_0^{\infty} \exp(-\int_0^t z_i^{\gamma} \lambda_{20}(u) \exp(\beta_2' X_i) \mathrm{d}u) f(z_i) \mathrm{d}z_i}
\end{aligned}
$$

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

$$
E\left[Z_i^{\gamma}|X_i, D_i \geq t\right] = \frac{\int_0^{\infty} z_i^{\gamma} \exp(-\int_0^t z_i^{\gamma} \lambda_{20}(u) \exp(\beta_2' X_i) \mathrm{d}u) f(z_i) \mathrm{d}z_i}{\int_0^{\infty} \exp(-\int_0^t z_i^{\gamma} \lambda_{20}(u) \exp(\beta_2' X_i) \mathrm{d}u) f(z_i) \mathrm{d}z_i}. \qquad (31)
$$

C. Simplification of the least false parameter function in case of a single binary covariate

The general form of the least false parameter function is given by

$$g(\beta) = \int_0^\infty E\left[Y_i(t)\left\{X_i - \frac{E\left[Y_i(t)X_i\exp(\beta' X_i)\right]}{E\left[Y_i(t)\exp(\beta' X_i)\right]}\right\}\mathrm{d}N_{i1}(t)\right]. \tag{32}$$

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 $N_{i1}(t)$. That function will now be simplified for $X_i \sim Bin(1,p)$ being binary. Therefore, first the inner expectations are dissolved:

$$\begin{aligned} & E\left[Y_i(t)X_i\exp(\beta' X_i)\right] \\ =\ & \sum_{a=0}^{1}\sum_{b=0}^{1} P(X_i=a, Y_i(t)=b)\cdot ba\exp(\beta a) \\ =\ & P(X_i=1, Y_i(t)=1)\cdot\exp(\beta) \end{aligned} \tag{33}$$

$$\begin{aligned} & E\left[Y_i(t)\exp(\beta' X_i)\right] \\ =\ & \sum_{a=0}^{1}\sum_{b=0}^{1} P(X_i=a, Y_i(t)=b)\cdot b\exp(\beta a) \\ =\ & P(X_i=0, Y_i(t)=1) + P(X_i=1, Y_i(t)=1)\cdot\exp(\beta) \end{aligned} \tag{34}$$

Using the notation $\bar{y}_i^{(k)}(t) = P(X_i=k, Y_i(t)=1)$, following the whole integrand of the least false parameter function is dissolved:

$$\begin{aligned} & E\left[Y_i(t)\left\{X_i - \frac{\bar{y}_i^{(1)}(t)\exp(\beta)}{\bar{y}_i^{(0)}(t)+\bar{y}_i^{(1)}(t)\exp(\beta)}\right\}\mathrm{d}N_{i1}(t)\right] \\ =\ & E\left[Y_i(t)X_i\mathrm{d}N_{i1}(t)\right] - \frac{\bar{y}_i^{(1)}(t)\exp(\beta)}{\bar{y}_i^{(0)}(t)+\bar{y}_i^{(1)}(t)\exp(\beta)}E\left[Y_i(t)\mathrm{d}N_{i1}(t)\right] \\ =\ & \left[\sum_{a=0}^{1}\sum_{b=0}^{1} P(X_i=a, Y_i(t)=b)\cdot baE\left[\mathrm{d}N_{i1}(t)|X_i=a, Y_i(t)=b\right]\right] \end{aligned}$$

$$
\begin{aligned}
&- \frac{\bar{y}_i^{(1)}(t)\exp(\beta)}{\bar{y}_i^{(0)}(t)+\bar{y}_i^{(1)}(t)\exp(\beta)} \left[\sum_{a=0}^{1}\sum_{b=0}^{1} P(X_i=a, Y_i(t)=b)\cdot bE\left[\mathrm{d}N_{i1}(t)|X_i=a, Y_i(t)=b\right]\right] \\
&= \bar{y}_i^{(1)}(t)E\left[\mathrm{d}N_{i1}(t)|X_i=1, Y_i(t)=1\right] \qquad (35) \\
&- \frac{\bar{y}_i^{(1)}(t)\exp(\beta)}{\bar{y}_i^{(0)}(t)+\bar{y}_i^{(1)}(t)\exp(\beta)} \left[\bar{y}_i^{(0)}(t)E\left[\mathrm{d}N_{i1}(t)|X_i=0, Y_i(t)=1\right] + \bar{y}_i^{(1)}(t)E\left[\mathrm{d}N_{i1}(t)|X_i=1, Y_i(t)=1\right]\right] \\
&\quad \vdots \qquad\qquad\qquad \vdots \qquad\qquad\qquad \vdots \\
&= \frac{\bar{y}_i^{(0)}(t)\bar{y}_i^{(1)}(t)}{\bar{y}_i^{(0)}(t)+\bar{y}_i^{(1)}(t)\exp(\beta)} \left[E\left[\mathrm{d}N_{i1}(t)|X_i=1, Y_i(t)=1\right] - E\left[\mathrm{d}N_{i1}(t)|X_i=0, Y_i(t)=1\right]\exp(\beta)\right]
\end{aligned}
$$

The "⋮"-signs in the derivations above indicate, that these intermediate steps just contain simple arithmetic operations which is why they are not explicitly shown here. In summary, the least false parameter function in case of a single binary covariate is given by $g(\beta) =$

$$
\int_0^\infty \frac{\bar{y}_i^{(0)}(t)\bar{y}_i^{(1)}(t)}{\bar{y}_i^{(0)}(t)+\bar{y}_i^{(1)}(t)\exp(\beta)} \left[E\left[\mathrm{d}N_{i1}(t)|X_i=1, Y_i(t)=1\right] - E\left[\mathrm{d}N_{i1}(t)|X_i=0, Y_i(t)=1\right]\exp(\beta)\right]. \qquad (36)
$$

References

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

Aalen, O. O., Valberg, M., Grotmol, T., and Tretli, S. (2015b). Understanding variation in disease risk: the elusive concept of frailty. *International Journal of Epidemiology*, **44**(4), 1408–1421.

Akacha, M., Binkowitz, B., Claggett, B., Hung, H. M. J., Mueller-Velten, G., and Stockbridge, N. (2019). Assessing Treatment Effects That Capture Disease Burden in Serious Chronic Diseases. *Therapeutic Innovation & Regulatory Science*, **53**(3), 387–397.

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

Andersen, P. K. and Gill, R. D. (1982). Cox's Regression Model for Counting Processes: A Large Sample Study. *The Annals of Statistics*, **10**(4), 1100–1120.

Angus, J. E. (1994). The Probability Integral Transform and Related Results. *SIAM Review*, **36**(4), 652–654.

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

Anker, S. D., Schroeder, S., Atar, D., Bax, J. J., Ceconi, C., Cowie, M. R., Crisp, A., Dominjon, F., Ford, I., Ghofrani, H.-A., Gropper, S., Hindricks, G., Hlatky, M. A., Holcomb, R., Honarpour, N., Jukema, J. W., Kim, A. M., Kunz, M., Lefkowitz, M., Le Floch, C., Landmesser, U., McDonagh, T. A., McMurray, J. J., Merkely, B., Packer, M., Prasad, K., Revkin, J., Rosano, G. M., Somaratne, R., Stough, W. G., Voors, A. A., and Ruschitzka, F. (2016). Traditional and new composite endpoints in heart failure clinical trials: facilitating comprehensive efficacy assessments and improving trial efficiency. *European Journal of Heart Failure*, **18**(5), 482–489.

Bretagnolle, J. and Huber-Carol, C. (1988). Effects of Omitting Covariates in Cox's Model for Survival Data. *Scandinavian Journal of Statistics*, **15**(2), 125–138.

Bui, A. L., Horwich, T. B., and Fonarow, G. C. (2011). Epidemiology and risk profile of heart failure. *Nature Reviews Cardiology*, **8**(1), 30–41.

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

Cheung, Y. B., Xu, Y., Tan, S. H., Cutts, F., and Milligan, P. (2009). Estimation of intervention effects using first or multiple episodes in clinical trials: The Andersen-Gill model re-examined. *Statistics in Medicine*, **29**(3), 328–336.

Ciolino, J. D., Palac, H. L., Yang, A., Vaca, M., and Belli, H. M. (2019). Ideal vs. real: A systematic review on handling covariates in randomized controlled trials. *BMC Medical Research Methodology*, **19**, e136.

Cook, R. J. and Lawless, J. (2007). *The Statistical Analysis of Recurrent Events*. Springer, New York.

Cook, R. J. and Lawless, J. F. (2018). *Multistate Models for the Analysis of Life History Data*. CRC Press, Taylor & Francis Group, Boca Raton.

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

Dennis, Jr., J. E. and Moré, J. J. (1977). Quasi-Newton Methods, Motivation and Theory. *SIAM Review*, **19**(1), 46–89.

Emura, T., Matsui, S., and Rondeau, V. (2019). *Survival Analysis with Correlated Endpoints - Joint Frailty-Copula Models*. Springer Nature Singapore, Singapore.

Ferreira-González, I., Permanyer-Miralda, G., Domingo-Salvany, A., Busse, J. W., Heels-Ansdell, D., Montori, V. M., Akl, E. A., Bryant, D. M., Alonso-Coello, P., Alonso, J., Worster, A., Upadhye, S., Jaeschke, R., Schünemann, H. J., Pacheco-Huergo, V., Wu, P., Mills, E. J., and Guyatt, G. H. (2007). Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. *BMJ*, **334**, e786.

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

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

Greenland, S. (1990). Randomization, Statistics, and Causal Inference. *Epidemiology*, **1**(6), 421–429.

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

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

Hicks, K. A., Mahaffey, K. W., Mehran, R., Nissen, S. E., Wiviott, S. D., Dunn, B., Solomon, S. D., Marler, J. R., Teerlink, J. R., Farb, A., Morrow, D. A., Targum, S. L., Sila, C. A., Hai, M. T. T., Jaff, M. R., Joffe, H. V., Cutlip, D. E., Desai, A. S., Lewis, E. F., Gibson, C. M., Landray, M. J., Lincoff, A. M., White, C. J., Brooks, S. S., Rosenfield, K., Domanski, M. J., Lansky, A. J., McMurray, J. J., Tcheng, J. E., Steinhubl, S. R., Burton, P., Mauri, L., O'Connor, C. M., Pfeffer, M. A., Hung, H. J., Stockbridge, N. L., Chaitman, B. R., and Temple, R. J. (2018). 2017 Cardiovascular and Stroke Endpoint Definitions for Clinical Trials. *Circulation*, **137**(9), 961–972.

High, R. and Elrayes, W. (2017). Fitting Statistical Models with PROCs NLMIXED and MCMC. In *SAS Global Forum 2017 Conference*, pages 1–14, Cary, NC, USA. SAS Institute Inc.

Huang, X. and Liu, L. (2007). A Joint Frailty Model for Survival and Gap Times Between Recurrent Events. *Biometrics*, **63**(2), 389–397.

Huang, X. and Wolfe, R. A. (2002). A Frailty Model for Informative Censoring. *Biometrics*, **58**(3), 510–520.

Jahn-Eimermacher, A., Ingel, K., Ozga, A.-K., Preussler, S., and Binder, H. (2015). Simulating recurrent event data with hazard functions defined on a total time scale. *BMC Medical Research Methodology*, **15**, e16.

Keene, O. N., Calverley, P. M. A., Jones, P. W., Vestbo, J., and Anderson, J. A. (2008). Statistical analysis of exacerbation rates in COPD: TRISTAN and ISOLDE revisited. *European Respiratory Journal*, **32**(1), 17–24.

Kelly, P. J., Lim, L. L., Cain, L. E., and Cole, S. R. (2000). Survival analysis for recurrent event data: an application to childhood infectious diseases. *Statistics in Medicine*, **19**(1), 13–33.

Kurada, R. R. (2016). SAS4720-2016: Fitting Multilevel Hierarchical Mixed Models Using PROC NLMIXED. In *SAS Global Forum 2016 Conference*, pages 1–15, Cary, NC, USA. SAS Institute Inc.

Lafourcade, A., His, M., Baglietto, L., Boutron-Ruault, M.-C., Dossus, L., and Rondeau, V. (2018). Factors associated with breast cancer recurrences or mortality and dynamic predic-

tion of death using history of cancer recurrences: the French E3N cohort. *BMC Cancer*, **18**, e171.

Lin, D. Y., Wei, L. J., Yang, I., and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **62**(4), 711–730.

Liu, L. and Huang, X. (2008). The use of Gaussian quadrature for estimation in frailty proportional hazards models. *Statistics in Medicine*, **27**(14), 2665–2683.

Liu, L. and Yu, Z. (2008). A likelihood reformulation method in non-normal random effects models. *Statistics in Medicine*, **27**(16), 3105–3124.

Liu, L., Wolfe, R. A., and Huang, X. (2004). Shared Frailty Models for Recurrent Events and a Terminal Event. *Biometrics*, **60**(3), 747–756.

Liu, X. (2014). Survival Models on Unobserved Heterogeneity and their Applications in Analyzing Large-scale Survey Data. *Journal of Biometrics & Biostatistics*, **5**, e2.

Lu, L. and Liu, C. (2008). Analysis of Correlated Recurrent and Terminal Events Data in SAS. In *Proceedings of Northeastern SAS Users' Group*, pages 1–8, Cary, NC, USA. SAS Institute Inc.

Marquardt, D. (1963). An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM Journal on Applied Mathematics*, **11**(2), 431–441.

McMurray, J. J. V. and Pfeffer, M. A. (2005). Heart Failure. *Lancet*, **365**, 1877–1889.

McMurray, J. J. V., Östergren, J., Swedberg, K., Granger, C. B., Held, P., Michelson, E. L., Olofsson, B., Yusuf, S., and Pfeffer, M. A. (2003). Effects of candesartan in patients with chronic heart failure and reduced left-ventricular systolic function taking angiotensin-converting-enzyme inhibitors: the CHARM-Added trial. *Lancet*, **362**, 767–771.

McTernan, M. and Blozis, S. (2013). User-specified likelihood expressions using NLMIXED and the GENERAL statement. In *Western Users of SAS Software*, pages 1–4, Las Vegas, NV, USA. SAS Institute Inc.

Metra, M. and Teerlink, J. R. (2017). Heart failure. *Lancet*, **390**, 1981–1995.

Nelson, K. P., Lipsitz, S. R., Fitzmaurice, G. M., Ibrahim, J., Parzen, M., and Strawderman, R. (2006). Use of the Probability Integral Transformation to Fit Nonlinear Mixed-Effects Models with Nonnormal Random Effects. *Journal of Computational and Graphical Statistics*, **15**(1), 39–57.

Packer, M. (2016). Development and Evolution of a Hierarchical Clinical Composite End Point for the Evaluation of Drugs and Devices for Acute and Chronic Heart Failure. *Circulation*, **134**(21), 1664–1678.

Pinheiro, J. C., Bates, D. M., and Pinheiro, J. C. (1995). Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model. *Journal of Computational and Graphical Statistics*, **4**(1), 12–35.

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

Rice, J. R. (1975). A Metalgorithm for Adaptive Quadrature. *Journal of the ACM (JACM)*, **22**(1), 61–82.

Rizopoulos, D. (2012). *Joint Models for Longitudinal and Time-to-Event Data*. Chapman & Hall, Boca Raton.

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

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

Rogers, J. K., Jhund, P. S., Perez, A. C., Böhm, M., Cleland, J. G., Gullestad, L., Kjekshus, J., van Veldhuisen, D. J., Wikstrand, J., Wedel, H., McMurray, J. J. V., and Pocock, S. J. (2014b). Effect of rosuvastatin on repeat heart failure hospitalizations: The CORONA trial (controlled rosuvastatin multinational trial in heart failure). *JACC: Heart Failure*, **2**(3), 289–297.

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

Rondeau, V. and Gonzalez, J. R. (2005). frailtypack: A computer program for the analysis of correlated failure time data using penalized likelihood estimation. *Computer Methods and Programs in Biomedicine*, **80**(2), 154–164.

Rondeau, V., Commenges, D., and Joly, P. (2003). Maximum penalized likelihood estimation in a gamma-frailty model. *Lifetime data analysis*, **9**(2), 139–53.

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

Rondeau, V., Mazroui, Y., and Gonzalez, J. R. (2012). frailtypack : An R Package for the Analysis of Correlated Survival Data with Frailty Models Using Penalized Likelihood Estimation or Parametrical Estimation. *Journal of Statistical Software*, **47**(4).

Sankoh, A. J., Li, H., and D'Agostino, R. B. (2017). Composite and multicomponent end points in clinical trials. *Statistics in Medicine*, **36**(28), 4437–4440.

SAS Institute Inc. (2015). The NLMIXED Procedure Manual. Technical report, SAS Institute Inc., Cary, NC, USA.

Schmoor, C. and Schumacher, M. (1997). Effects of Covariate Omission and Categirization when Analysing Randomized Trials with the Cox Model. *Statistics in Medicine*, **16**(3), 225–237.

Schmoor, C., Bender, R., Beyersmann, J., Kieser, M., Schumacher, M., Thanarajasingam, G., Sloan, J. A., and Grothey, A. (2016). Adverse event development in clinical oncology trials - Comment and Authors' reply. *Lancet Oncology*, **17**(7), e264–e265.

Schnell, O., Rydén, L., Standl, E., and Ceriello, A. (2017). Updates on cardiovascular outcome trials in diabetes. *Cardiovascular Diabetology*, **16**, e128.

Stroup, W. W., Milliken, G. A., Claassen, E. A., and Wolfinger, R. D. (2018). *SAS for Mixed Models: Introduction and Basic Applications*. SAS Institute Inc., Cary, NC, USA.

Struthers, C. A. and Kalbfleisch, J. D. (1986). Misspecified Proportional Hazard Models. *Biometrika*, **73**(2), 363–369.

Toenges, G. and Jahn-Eimermacher, A. (2019). Marginal hazard ratio estimates in joint frailty models for heart failure trials. *Biometrical Journal*, **61**(6), 1385–1401.

Toenges, G. and Jahn-Eimermacher, A. (2020). Computational issues in fitting joint frailty models for recurrent events with an associated terminal event. *Computer Methods and Programs in Biomedicine*, **188**, e105259.

Unkel, S., Amiri, M., Benda, N., Beyersmann, J., Knoerzer, D., Kupas, K., Langer, F., Leverkus, F., Loos, A., Ose, C., Proctor, T., Schmoor, C., Schwenke, C., Skipka, G., Unnebrink, K., Voss,

F., and Friede, T. (2019). On estimands and the analysis of adverse events in the presence of varying follow-up times within the benefit assessment of therapies. *Pharmaceutical Statistics*, **18**(2), 166–183.

Wienke, A. (2007). *Frailty Models in Survival Analysis*. Chapman and Hall/CRC, Boca Raton.

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

Zannad, F., Garcia, A. A., Anker, S. D., Armstrong, P. W., Calvo, G., Cleland, J. G., Cohn, J. N., Dickstein, K., Domanski, M. J., Ekman, I., Filippatos, G. S., Gheorghiade, M., Hernandez, A. F., Jaarsma, T., Koglin, J., Konstam, M., Kupfer, S., Maggioni, A. P., Mebazaa, A., Metra, M., Nowack, C., Pieske, B., Piña, I. L., Pocock, S. J., Ponikowski, P., Rosano, G., Ruilope, L. M., Ruschitzka, F., Severin, T., Solomon, S., Stein, K., Stockbridge, N. L., Stough, W. G., Swedberg, K., Tavazzi, L., Voors, A. A., Wasserman, S. M., Woehrle, H., Zalewski, A., and McMurray, J. J. (2013). Clinical outcome endpoints in heart failure trials: a European Society of Cardiology Heart Failure Association consensus document. *European Journal of Heart Failure*, **15**(10), 1082–1094.

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

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

Zhong, Y. and Cook, R. J. (2019). The effect of omitted covariates in marginal and partially conditional recurrent event analyses. *Lifetime Data Analysis*, **25**(2), 280–300.

Zinman, B., Wanner, C., Lachin, J. M., Fitchett, D., Bluhmki, E., Hantel, S., Mattheus, M., Devins, T., Johansen, O. E., Woerle, H. J., Broedl, U. C., and Inzucchi, S. E. (2015). Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. *The New England Journal of Medicine*, **373**, 2117–2128.

Danksagung

– Aus Datenschutzgründen nicht dargestellt –

Lebenslauf

– Aus Datenschutzgründen nicht dargestellt –