University of California
Los Angeles

Joint Inference for Competing Risk Data

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Biostatistics

by

Qing Yang

2014

Abstract of the Dissertation

Joint Inference for Competing Risk Data

by

Qing Yang

Doctor of Philosophy in Biostatistics

University of California, Los Angeles, 2014

Professor Gang Li, Chair

This dissertation develops joint inferential methods for the cause specific hazard function and the cumulative incidence function of a specific type of failure to assess the effects of a variable on the type of failure of interest in the presence of competing risks. Joint inference for the two functions are needed in practice because 1) they describe different characteristics of a particular type of failure, 2) they do not uniquely determine each other, and 3) the effects of a variable on the two functions can be different and one often does not know which effects are to be expected. We study both the group comparison problem and the Cox's regression problem. We also develop joint inference for other equivalent pairs of functions. Our simulation shows that the derived joint tests can be considerably more powerful than the Bonferroni method, which has important practical implications to the analysis and design of clinical studies with competing risks data. We illustrate our methods using a Hodgkin disease data and a lymphoma data.

We also develop sample size calculation methods based on nonparametric two-sample joint tests of the cause-specific hazard and the all-cause hazard. A user friendly R-function is developed to implement the method. We illustrate the implementation of our method and the potential saving on the required sample size

over the Bonferroni method through simulations and the 4-D (Die Deutsche Diabetes Dialyse Studie) clinical trial designed to compare a lipid lowering treatment with placebo in type 2 diabetic patients on hemodialysis.

The dissertation of Qing Yang is approved.

Weng Kee Wong

Jiaoti Huang

Robert Elashoff

Ron Brookmeyer

Gang Li, Committee Chair

University of California, Los Angeles

2014

To my parents, my husband, Xinming An, my sons, Alex Muyang An, and Anthony Muan Yang

Table of Contents

List of Figures

List of Tables

Acknowledgments

First of all, I would like to express my deepest gratitude to my advisor, Dr. Gang Li, for his guidance and support throughout my graduate studies in UCLA. I cannot finish this dissertation without his patience, encouragement, and understanding. His wide knowledge and logical way of thinking provided me with an excellent atmosphere for doing research.

Also I would like to thank my dissertation committee, Dr. Ron Brookmeyer, Dr. Robert Elashoff, Dr. Jiaoti Huang, and Dr. Weng Kee Wong for their valuable suggestions and comments.

I am very grateful for the valuable opportunities of being a graduate student researcher under the guidance of Dr. Karabi Nandy in School of Nursing, who have been helping me in many ways throughout my research and personal life.

I also want to thank the faculty, staff, and students of the Division of Biostatistics, for their help and support during my six years of study. Special thanks go to Dr. Thomas Belin and Dr. William Cumberland, who always try to help students in many ways.

Id like to thank all my friends, their support gave me strength and confidence and made my stay in the US much more enjoyable.

Lastly, I would like to give my special thanks to my parents, Jianguo Yang, Guifen Wu, and my husband, Xinming An for their faith in me and their support and unconditional love to me for all these years.

Vita

2003	B.S. (Mathematics), Beijing Institute of Techonolgy, Beijing, China.
2006	M.S. (Statistics), Beijing Institute of Techonolgy, Beijing, China.
2008	M.S. (Biotatistics), University of California, Los Angeles

CHAPTER 1

Introduction

Competing risks failure time data arise commonly in clinical trials, reliability testing, and other fields. For instance, in a clinical trial, one may be interested in time to death due to a particular disease, but a patient can also die from other competing diseases that are potentially positively correlated with the disease of interest. Competing risks can also be negatively correlated with the event time of interest. For example, in a kidney transplantation program, patients who are ineligible for transplantation due to reasons, such as being overweight, are put on a waiting list until they become eligible (see, e.g., Sancho et al. (2007)). An important variable of interest is the waiting time to become eligible for transplantation. In this case, death before becoming eligible for the transplantation is a competing risk event that is potentially negatively correlated with the waiting time. More examples of competing risks failure time data can be found in Graunt (1899); Halley (1942); Prentice et al. (1978); Pintilie (2006); Gichangi and Vach (2005); Putter et al. (2007), and the references therein.

There is a broad literature on analysis of competing risks survival data. Group comparison of a specific type of failure has been studied using either the cause specific hazard (Prentice et al., 1978; Lindkvist and Belyaev, 1998; Kulathinal and Gasbarra, 2002) or the cumulative incidence (Gray, 1988; Pepe and Mori, 1993; Bajorunaite and Klein, 2007). Methods to compare failures across failure types have been developed with respect to either the cause specific hazard, or the

cumulative incidence, or both (Aly et al., 1994; Sun and Tiwari, 1995; Lam, 1998; Luo and Turnbull, 1999). Tiwari et al. (2006) proposed a test to check equality of cause specific hazards across all failure types and groups. For regression analysis of competing risks survival data, Prentice et al. (1978), Lagakos (1978), Holt (1978), Cox and Oakes (1984, chap.9), and Larson (1984); Lunn and McNeil (1995) studied proportional cause-specific hazards models. Fine and Gray (1999) introduced a proportional subdistribution hazards model for cumulative incidence regression. Fine (1999), Fine (2001), Klein and Andersen (2005), and Gerds et al. (2012) used transformation models to directly model the cumulative incidence function. Klein (2006) discussed additive models for both cause specific hazard and cumulative incidence function. Comprehensive survey of methods for competing risks data and further references can be found in Beyersmann et al. (2007); Latouche et al. (2007); Haller et al. (2012).

1.1 Joint Inference for Competing Risks Data

In the first part of the dissertation we focus on the problem of assessing the effects of a variable (treatment or covariate) on a particular type of failure. For convenience, we assume hereafter that there are only two types of failure, where type 1 failure is of interest and type 2 includes all other competing risks. As discussed earlier, there are mainly two approaches to this problem. One approach is based on the cause-specific hazard (CSH)

$$\lambda_1(t) = \lim_{dt \downarrow 0} P(t \leq T < t + dt, D = 1 | T \geq t)/dt, \quad t > 0$$

the instantaneous risk for type 1 failure at time t given that the subject is at risk just prior to t, where T is the continuous failure time with J possibly correlated types and D is the failure type. For example, Prentice et al. (1978) showed that

the standard Cox (1972, 1975) regression method can be used to study the effect of a variable on the cause-specific hazard $\lambda_1(t)$ by treating other types of failures as independent censoring. Another approach uses the cumulative incidence function (CIF),

$$F_1(t) = P(T \leq t, D = 1), \quad t > 0$$

the cumulative incidence rate of type 1 failure by time t, or equivalently, the sub-distribution hazard (SDH) of $F_1(t)$, which is defined by

$$\begin{aligned} \tilde{\lambda}_1(t) &= \lim_{dt \downarrow 0} P(t \leq T < t + dt, D = 1 | T \geq t \cup (T < t \cap D \neq 1))/dt \\ &= -d \log \{1 - F_1(t)\} / dt, \end{aligned}$$

In particular, Gray (1988) developed a class of nonparametric tests to compare the CIFs of a particular type of failure between different groups and Fine and Gray (1999) introduced a proportional subdistribution hazards model for regression problems.

Despite of the extensive literature on this topic, there are still confusions to practitioners as to which method should be used in practice when studying the effects of a variable on type 1 failure. We point out that joint inference for both $\lambda_1(t)$ and $F_1(t)$ should be made. First of all, these two quantities describe different characteristics of type 1 failure: $\lambda_1(t)$ represents the instantaneous type 1 failure rate at time t given survival to t, whereas $F_1(t)$ summarizes the prevalence or cumulative incidence of type 1 failure over the time interval $[0, t]$. Secondly, $\lambda_1(t)$ and $F_1(t)$ do not uniquely determine each other except when $J = 1$. It can be shown that $F_1(t) = \int_0^t S(u)\lambda_1(u)du$, where $S(u) = P(T > u)$ is the all-cause survival function. Thus $F_1(t)$ depends not only on $\lambda_1(t)$, but also on other cause-specific hazards through the all-cause survival function $S(t)$. Finally, the effect of a variable on $\lambda_1(t)$ can be different from its effect on $F_1(t)$ (Gray, 1988; Fine and Gray, 1999), and one often does not know which effects are to be expected. To the

best of our knowledge, however, no formal joint inference procedure is available for these quantities in the literature. Although Bonferroni's method provides a straightforward solution, we demonstrate later in Sections 3 and 4 that it can be severely under-powered since it does not account for potential correlation between the estimates of the two quantities.

The primary purpose of this first part of the dissertation is to develop joint inference procedures to assess the effects of a variable on $\lambda_1(t)$ and $F_1(t)$ simultaneously. We allow independent right censoring in addition to competing risks. We first consider the two-sample comparison problem with respect to both $\lambda_1(t)$ and $F_1(t)$. By establishing the asymptotic joint distribution of the weighted log-rank test statistic for $\lambda_1(t)$ and the Gray (1988) test statistic for $F_1(t)$ (or $\tilde{\lambda}_1(t)$), we derive two-sample joint tests for $\lambda_1(t)$ and $F_1(t)$. We then extend our method to a regression setting based on Cox's type models for $\lambda_1(t)$ and $F_1(t)$. We also remark that our approach can be extended to develop joint tests for other related quantities.

1.2 Sample Size Calculation for Joint Test of Competing Risks Data

To establish the efficacy of a treatment with sufficient power at a pre-specified Type I error level, one needs to determine the adequate number of patients to be enrolled in a randomized placebo-controlled clinical trial. There is an extensive literature on sample size calculation for a time-to-event outcome with independent right censoring. For example, a widely used sample size calculation formula was proposed by Schoenfeld (1981, 1983) who considered uniform patient entry and administrative censoring. Lachin and Foulkes (1986) extended the formula

to more complex clinical situations, allowing for truncated exponential patient entry, losses to follow-up, noncompliance and stratified analysis. Yateman and Skene (1992) used piecewise exponential distribution to approximate arbitrary patient entry pattern and losses to follow-up distribution. A nice review of sample size calculation and further references can be found in (Lachin, 1981; Lakatos, 1986; Lachin and Foulkes, 1986; Lakatos, 1988; Lakatos and Lan, 1992; Schmoor et al., 2000; Eng and Kosorok, 2005).

We will develop sample size calculation methods for jointly testing the cause specific hazard and the all cause hazard in a two sample comparison situation in the presence of competing risks and independent right censoring.

1.3 Research Outline

The organization of the dissertation is as follows. In Chapter 2, we will introduce some basic background information for competing risks data, sample size calculation for survival data and counting process theory we need to use in the following chapters. In Chapter 3, we develop joint test procedures for group comparisons of $\lambda_1(t)$ and $F_1(t)$ and joint regression analysis methods for $\lambda_1(t)$ and $\tilde{\lambda}_1(t)$ under Cox-type regression models. We also discuss joint tests for other equivalent pairs including $\lambda_1(t)$ with the all-cause hazard, and $\lambda_1(t)$ with the cause-specific hazard for other failure types. We presents some simulation results to evaluate the proposed methods and compare them with the Bonferroni method, and illustrate our methods on a Hodgekin disease data and a follicular lymphoma study. In Chapter 4, we develop a sample size calculation method for two sample nonparametric joint test with respect to cause specific hazard and all cause hazard as well as some simulation studies and a real data example. We present some discussion and future researches in chapter 5.

CHAPTER 2

Preliminaries

In this chapter, first we introduce two important quantities in competing risks data analysis and the related non-parametric test and semi-parametric Cox models. Then we summarize the sample size calculation methods proposed by Schoenfeld (1981, 1983) for survival data with independent right censoring. We also review basic counting process and martingale theories which will be useful for our derivations.

2.1 Competing Risks Models

Standard survival data considers the time span from some time origin to a major end point defined by the occurrence of a certain event of interest when a possible independent right censoring is involved. However, in practice, the more commonly seen situation is that there are more than one type of events which can potentially lead to subject's failure. Typically we can only observe one of the events in this case, which forms a competing risk situation. We usually are interested in only one type of event, and other possible events are called competing risks. For example, in cancer research, we are interested in time to death due to a specific type of cancer, so death due to other diseases, such as heart attack, patients' losses to follow-up and end of study are considered as competing risks. Under the competing risks model, we can have dependent competing risks or independent

competing risks. For the latter, we usually call it independent censoring, which means the competing risk or the censoring has nothing to do with the disease or other risks. In our example, end of study can be considered as independent censoring, while patient's losses to follow-up may fall in this category if we can assume patients' losses to follow-up were due to moving out of the area etc. For death due to other medical problems, it will be hard to justify that they are independent of the death due to a specific cancer. We can imagine that there are always some biological mechanism behind this. It's possible that people who died due to heart attack is more likely to experience death due to cancer since their health condition in general are not good.

2.1.1 Two Sample Comparison

Suppose that there are two independent groups of subjects. Let T_{ik}, D_{ik}, and C_{ik} denote the continuous failure time, the type of failure, and the censoring time, respectively, for subject i in group k, $i = 1, \ldots, n_k$, $k = 1, 2$. Assume that the triplets (T_{ik}, D_{ik}, C_{ik}) for different subjects within each group are independent and identically distributed and that the censoring time C_{ik} is independent of the survival time T_{ik}. The two groups are allowed to have different censoring distributions. For group k $(k = 1, 2)$, one observes a right censored competing risks survival data $\{(X_{ik}, \delta_{ik}), i = 1, \ldots, n_k\}$, where $X_{ik} = \min(T_{ik}, C_{ik})$ and $\delta_{ik} = D_{ik} I(T_{ik} \leq C_{ik})$. Let $S_k(t) = P(T_{ik} > t)$ and $S_k^c(t) = P(C_{ik} > t)$. For group k $(k = 1, 2)$, let $\lambda_{1k}(t)$, $F_{1k}(t)$, and $\tilde{\lambda}_{1k}(t)$ denote the cause-specific hazard function, the cumulative incidence function, and the sub-distribution hazard function, respectively, for type 1 failure.

2.1.1.1 Logrank Test for Cause Specific Hazard

We first review the two-sample weighted log-rank test for the cause-specific hazard for type 1 failure.

It is now well known that the standard (weighted) log-rank test (Peto and Peto, 1972; Andersen et al., 1982) for right censored survival data can be used to test the following null hypothesis

$$H_0 : \lambda_{11}(t) = \lambda_{12}(t) \quad \text{for all } t > 0, \tag{2.1}$$

by treating all other competing risks as independent right censoring (Tsiatis, 1975; Prentice et al., 1978; Lindkvist and Belyaev, 1998). Specifically, let $N_{jk}(t) = \sum_{i=1}^{n_k} I(X_{ki} \leq t, D_{ki} = j)$ be the counting process of the number of observed type j failure in group k by time t, and $Y_k(t) = \sum_{i=1}^{n_k} I\{X_{ki} \geq t\}$ be the at risk process indicating the number of subjects in group k who are at risk prior to time t, $k = 1, 2$. Let $N_{j\cdot}(t) = \sum_{k=1}^{2} N_{jk}(t)$ and $Y_{\cdot}(t) = \sum_{k=1}^{2} Y_k(t)$. The weighted log-rank test statistic for (2.1) is defined as

$$U_{1k} = \int_0^\tau W_1(t) Y_k(t) \left\{ \frac{dN_{1k}(t)}{Y_k(t)} - \frac{dN_{1\cdot}(t)}{Y_{\cdot}(t)} \right\}, \tag{2.2}$$

where $W_1(t)$ is a predictable weight function, which converges in probability to some deterministic function $w_1(t)$ when $n \to \infty$, and τ is the largest time at which all of the groups have at least one subject at risk. It can be shown that under the null hypothesis, $n^{-1/2} U_{11}/\hat{\sigma}$ has a standard normal limiting distribution where

$$\hat{\sigma}^2 = \int_0^\tau W_1^2(t) \frac{Y_1(t) Y_2(t)}{Y_{\cdot}(t)} \frac{dN_{1\cdot}(t)}{Y_{\cdot}(t)}. \tag{2.3}$$

This leads to an asymptotic χ_1^2-test or a Z test for (2.1).

2.1.1.2 Two-Sample Tests for Cumulative Incidence

Gray (1988) developed a class of K-sample nonparametric tests to compare the cumulative incidence between different groups. Consider the following null hypothesis

$$H_0 : F_{11}(t) = F_{12}(t) \quad \text{for all } t > 0. \tag{2.4}$$

or equivalently

$$H_0 : \tilde{\lambda}_{11}(t) = \tilde{\lambda}_{12}(t) \quad \text{for all } t > 0. \tag{2.5}$$

The Gray (1988) nonparametric test statistic is defined as

$$\tilde{U}_{1k} = \int_0^{\tau_k} \tilde{W}(t) R_k(t) \left\{ \frac{dN_{1k}(t)}{R_k(t)} - \frac{dN_{1\cdot}(t)}{R_\cdot(t)} \right\}, \tag{2.6}$$

where $\tilde{W}(t)$ is a predictable weight function, which converges in probability to some deterministic function $\tilde{w}(t)$ when $n \to \infty$, $R_k(t) = I(\tau_k \geq t) Y_k(t) \hat{G}_{jk}(t) / \hat{S}_k(t-)$ can be considered as an adjusted risk set size for group k at time t and τ_k is some fixed time point satisfying $S_k(\tau_k) S_k^c(\tau_k) > 0$. $R_\cdot(t)$ represents the same quantity in the pooled sample. Let $S_\cdot(t)$ and $F_{j\cdot}(t)$ denote the survival function and cumulative incidence function with respect to cause j for pooled sample in two groups, respectively. $\hat{G}_{jk}(t-)$ is the the left-hand limit of the Kaplan-Meier (1958) estimate of $G_{jk}(t) = 1 - F_{jk}(t)$, and $\hat{G}_{j\cdot}(t-)$ is the the left-hand limit of the Kaplan-Meier estimate of $G_{j\cdot}(t) = 1 - F_{j\cdot}(t)$. $\hat{S}_k(t-)$ and $\hat{S}_\cdot(t-)$ are the left-hand limit of the Kaplan-Meier estimate of $S_k(t)$ and $S_\cdot(t)$, respectively. Gray (1988) showed that $n^{-1/2} \tilde{U}_{11} / \hat{\tilde{\sigma}}$ has a standard normal limiting distribution, where

$$\hat{\tilde{\sigma}}^2 = \sum_{k=1}^{2} n^{-1} \left\{ \int_0^{\tau_1} \hat{a}_k^2(t) \hat{h}_k^{-1}(t) \hat{h}_\cdot^{-1}(t) dN_{1\cdot}(t) + \int_0^{\tau_1} \hat{b}_{2k}^2(t) \hat{h}_k^{-2}(t) dN_{2k}(t) \right\}, \tag{2.7}$$

with

$$
\begin{aligned}
\hat{a}_k(t) &= \hat{d}_{1k}(t) + \hat{b}_{1k}(t), \\
\hat{b}_{jk}(t) &= \left[I(j=1) - \hat{G}_{1\cdot}(t)/\hat{S}_k(t) \right] \left[\hat{c}_k(\tau_1) - \hat{c}_k(t) \right], \\
\hat{c}_k(t) &= \int_0^t \hat{d}_{1k}(u) \hat{G}_{1\cdot}(u-)^{-1} \hat{h}_{\cdot}^{-1}(u) dN_{1\cdot}(u), \\
\hat{d}_{jk}(t) &= n^{-1} I(j=1) \tilde{W}(t) R_1(t) \left[I(k=1) - \hat{h}_k(t)/\hat{h}_{\cdot}(t) \right] / \hat{G}_{1\cdot}(t-), \\
\hat{h}_k(t) &= I(t \le \tau_k) n^{-1} Y_k(t) / \hat{S}_k(t-), \\
\hat{h}_{\cdot}(t) &= I(t \le max(\tau_1, \tau_2)) n^{-1} Y_{\cdot}(t) / \hat{S}_{\cdot}(t-), \\
\hat{G}_{1\cdot}(t) &= 1 - \hat{F}_{1\cdot}(t) = 1 - n^{-1} \textstyle\int_0^t \hat{h}_{\cdot}^{-1}(u) dN_{1\cdot}(u).
\end{aligned}
\tag{2.8}
$$

This gives an asymptotic χ_1^2-test for (2.4) based on $n^{-1}\tilde{U}_{11}^2/\hat{\tilde{\sigma}}^2$ or a Z test based on $n^{-1/2}\tilde{U}_{11}/\hat{\tilde{\sigma}}$.

2.1.2 Proportional Hazard Model

Assume that one observes n independent and identically distributed triples $(X_i, \delta_i, \boldsymbol{Z}_i)$, where for subject i $(i = 1, \ldots, n)$, $X_i = \min\{T_i, C_i\}$, $\delta_i = D_i I(T_i \le C_i)$, T_i is the failure time of interest, C_i is a right censoring time, D_i is discrete random variable taking values on $1, \ldots, J$ with $D_i = j$ indicating that type j failure is observed, and $\mathbf{Z}_i$ is a vector of fixed or time-varying covariates that are observed on $[0, X_i]$. Assume C_i is independent of T_i, D_i and $\mathbf{Z}_i$, and $pr(C_i \ge t) = G^c(t)$.

2.1.2.1 Cox Regression for Cause Specific Hazard

When we evaluate cause specific hazard in a competing risks set, we are actually treating the competing risks other than the cause of interest, j, as independent censoring. This can be seen by writing out the partial likelihood based on the cause specific hazard with respect to each causes (Prentice et al., 1978) (Tsiatis, 1975). So we can use the standard Cox (1972, 1974) model to model effects of regression variables on a cause-specific hazard function by regarding all failures

from other causes as independently censored at their time of failure. Suppose the covariates we want to include in the model is a p dimension covariates vector $\mathbf{Z}(t)$, which include treatment indicator, time varying covariates chosen from $\boldsymbol{\mathcal{Z}}(t)$, and some possible interactions. Conditional on $\mathbf{Z}(t)$, the model can be formulated as

$$\lambda_j(t|\mathbf{Z}(t)) = \lambda_{j0}(t)\exp(\mathbf{Z}^T(t)\boldsymbol{\beta}_0), \tag{2.9}$$

where $\lambda_{j0}(t)$ is the baseline cause specific hazard for cause j. The p dimension score test statistics with respect to coefficient vector $\boldsymbol{\beta}$ is

$$\mathbf{S}_j^{(n)}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\int_0^{\infty}\left\{\mathbf{Z}_i(t) - \bar{\mathbf{Z}}(\boldsymbol{\beta},t)\right\}dN_{ij}(t), \tag{2.10}$$

where $\bar{\mathbf{Z}}(\boldsymbol{\beta},t) = \frac{\sum_l Y_l(t)\mathbf{Z}_l(t)\exp(\mathbf{Z}_l^T(t)\boldsymbol{\beta})}{\sum_l Y_l(t)\exp(\mathbf{Z}_l^T(t)\boldsymbol{\beta})}$, $Y_i(t) = I\{X_i \geq t\}$ and $N_{ij}(t) = I(X_i \leq t, D = j)$. It's been proved that under the null hypothesis, $n^{-1/2}\mathbf{S}_j^{(n)}(\boldsymbol{\beta})$ has a p dimension normal limiting distribution with mean $\mathbf{0}$ and variance-covariance matrix

$$\boldsymbol{\Omega}_{(pp)} = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\infty}\left[\mathbf{Z}_i(t) - \bar{\mathbf{Z}}(\boldsymbol{\beta},t)\right]^{\otimes 2}\exp(\mathbf{Z}_i^T(t)\boldsymbol{\beta})Y_i(t)\lambda_{j0}(t)dt, \tag{2.11}$$

which can be estimated by

$$\hat{\boldsymbol{\Omega}}_{(pp)} = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\infty}\left[\frac{\sum_l Y_l(t)\mathbf{Z}_l(t)^{\otimes 2}\exp(\mathbf{Z}_l^T(t)\hat{\boldsymbol{\beta}})}{\sum_l Y_l(t)\exp(\mathbf{Z}_l^T(t)\hat{\boldsymbol{\beta}})} - \bar{\mathbf{Z}}(\hat{\boldsymbol{\beta}},t)^{\otimes 2}\right]dN_{ij}(t), \tag{2.12}$$

where $\hat{\boldsymbol{\beta}}$ is the Maximum Likelihood Estimate of $\boldsymbol{\beta}$. Then $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \approx \boldsymbol{\Omega}_{(pp)}^{-1}n^{-1/2}S_j^{(n)}(\boldsymbol{\beta}_0)$. The details can be found in Prentice et al. (1978), Fleming and Harrington (1991, chap. 8) and Andersen et al. (1993, chap.7).

2.1.2.2 Fine & Gray's Proportional Hazard Model for Subdistribution Hazard

Cause-specific hazard is a widely used quantity in evaluating treatment effect with respect to one specific cause, however, the interpretation is restricted to

actual study conditions, and there is no implication that the same effect would be observed under a new setting. Moreover, the instantaneous risk of specific failure type is usually of less interest compare to the overall probability of the specific failure type, which can be formulated as the cumulative incidence function. The problem is there is no one-to-one relationship between the cause-specific hazard and the cumulative incidence function, and the effect of a covariate on the cause-specific hazard function of a particular failure type may be quite different from the effect of the covariate on the corresponding cumulative incidence function(Gray 1988, Pepe 1991). Thus, Fine and Gray (1999) proposed a new semiparametric proportional hazards model for the subdistribution which akin to proportional cause-specific hazards model.

The second hypothesis we are interested in testing is if cumulative incidence $F_j(t)$ for cause of interest, j, are the same in the two treatment groups. In our paper, we adapt the proportional hazard model for subdistribution after log(-log) transformation proposed by Fine and Gray (1999). Subdistribution hazard or cumulative incidence hazard is defined by

$$\begin{aligned} \tilde{\lambda}_j(t) &= \lim_{dt\downarrow 0} P(t \leq T < t+dt, D = j | T \geq t \cup (T < t \cap D \neq j))/dt \\ &= -d \log \{1 - F_j(t)\} /dt. \end{aligned} \tag{2.13}$$

$\tilde{\lambda}_j(t)$ is modeled by

$$\tilde{\lambda}_j(t|\tilde{\mathbf{Z}}(t)) = \tilde{\lambda}_{j0}(t) \exp(\tilde{\mathbf{Z}}^T(t)\boldsymbol{\gamma}_0), \tag{2.14}$$

where $\tilde{\lambda}_{j0}(t)$ is the baseline hazard for cause specific cumulative incidence hazard for cause j, and $\tilde{\mathbf{Z}}(t)$ is a q dimension vector of time varying covariates we chose from $\boldsymbol{\mathcal{Z}}(t)$, which might be different from $\mathbf{Z}(t)$ we use for model (3.18).

By using similar notations from Fine and Gray (1999), let us assume the censoring variable C is independent of $T, D, \tilde{\mathbf{Z}}$ with survival function $G^c(t) =$

$P(C \geq t)$. The q dimension score test statistics with respect to $\boldsymbol{\gamma}$ can be expressed as

$$\tilde{\mathbf{S}}_j^{(n)}(\boldsymbol{\gamma}) = \sum_{i=1}^{n} \int_0^{\infty} \left\{ \tilde{\mathbf{Z}}_i(t) - \bar{\tilde{\mathbf{Z}}}(\boldsymbol{\gamma}, t) \right\} \omega_i(t) d\tilde{N}_{ij}(t), \tag{2.15}$$

where $\bar{\tilde{\mathbf{Z}}}(\boldsymbol{\gamma}, t) = \frac{\sum_l \omega_l(t) \tilde{Y}_l(t) \tilde{\mathbf{Z}}_l \exp(\tilde{\mathbf{Z}}_l^T(t)\boldsymbol{\gamma})}{\sum_l \omega_l(t) \tilde{Y}_l(t) \exp(\tilde{\mathbf{Z}}_l^T \boldsymbol{\gamma})}$, $\tilde{N}_{ij}(t) = I(T_i \leq t, D = j)$, and $\tilde{Y}_i(t) = 1 - \tilde{N}_{ij}(t-)$. $\omega_i(t)$ is a time dependent weight constructed by adapting the inverse probability of censoring weighting techniques. It's calculated by $\omega_i(t) = I(C_i \geq T_i \wedge t)\hat{G}^c(t)/\hat{G}^c(X_i \wedge t)$, where $\hat{G}^c$ is a Kaplan-Meier estimator of the survival function of the censoring variable, C. Notice that $\tilde{N}_{ij}(t)$ is different from $N_{ij}(t)$ and may not be observed if the subject is censored, but $\omega(t)\tilde{N}_{ij}(t)$ is always computable.

It's proved that under the null hypothesis, $n^{-1/2}\tilde{\mathbf{S}}_j^{(n)}(\boldsymbol{\gamma})$ has a normal limiting distribution with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Omega}_{(qq)}$, which can be estimated by

$$\hat{\boldsymbol{\Omega}}_{(qq)} = n^{-1} \sum_{i=1}^{n} (\hat{\boldsymbol{\eta}}_i + \hat{\boldsymbol{\phi}}_i)^{\otimes 2}, \tag{2.16}$$

with

$$\hat{\boldsymbol{\eta}}_i = \int_0^{\infty} \left\{ \tilde{\mathbf{Z}}_i(t) - \bar{\tilde{\mathbf{Z}}}(\hat{\boldsymbol{\gamma}}, t) \right\} \omega_i(t) d\hat{\tilde{M}}_{ij}(t), \tag{2.17}$$

where

$$\hat{\tilde{M}}_{ij}(t) = \tilde{N}_{ij}(t) - \int_0^t (1 - I(T_i < u, D = j)) \exp(\tilde{\mathbf{Z}}_i(t)^T(u)\hat{\boldsymbol{\gamma}}) d\hat{\tilde{\Lambda}}_{j0}(u),$$

$$\hat{\tilde{\Lambda}}_{j0}(t) = \frac{1}{n} \sum_{i=1}^{n} \int_0^t \left\{ \sum_{l=1}^{n} Y_l(u) \exp(\tilde{\mathbf{Z}}_l^T(u)\hat{\boldsymbol{\gamma}}) \right\}^{-1} d\tilde{N}_{ij}(u),$$

and

$$\bar{\tilde{\mathbf{Z}}}(\hat{\boldsymbol{\gamma}}, t) = \frac{\sum_{l=1}^{n} \omega_l(t) \tilde{Y}_l(t) \tilde{\mathbf{Z}}_l \exp(\tilde{\mathbf{Z}}_l^T(t)\hat{\boldsymbol{\gamma}})}{\sum_{l=1}^{n} \omega_l(t) \tilde{Y}_l(t) \exp(\tilde{\mathbf{Z}}_l^T(t)\hat{\boldsymbol{\gamma}})},$$

and

$$\hat{\boldsymbol{\phi}}_i = \int_0^{\infty} \frac{\hat{\mathbf{q}}(t)}{\hat{\pi}(t)} d\hat{\tilde{M}}_{ic}(t), \tag{2.18}$$

where

$$\hat{\tilde{M}}_{ic}(t) = I(X_i \leq t, D_i = 0) - \int_0^t I(X_i \geq u) d\hat{\tilde{\Lambda}}_c(u),$$

$$\hat{\tilde{\Lambda}}_c(t) = \int_0^t \frac{\sum_i d\left\{I(X_i \leq u, D_i = 0)\right\}}{\sum_i I(X_i \geq u)},$$

$$\hat{\mathbf{q}}(t) = -n^{-1} \sum_{i=1}^{n} \int_0^{\infty} \left\{\tilde{\mathbf{Z}}_i - \bar{\tilde{\mathbf{Z}}}(\hat{\boldsymbol{\gamma}}, s)\right\} \omega_i(s) d\hat{\tilde{M}}_{ij}(s) I(s \geq t > X_i)$$

and

$$\hat{\pi}(t) = n^{-1} \sum_{i=1}^{n} I(X_i \geq t).$$

$\hat{\boldsymbol{\gamma}}$, the MLE for $\boldsymbol{\gamma}$ is consistent for $\boldsymbol{\gamma}_0$. $n^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) \approx \boldsymbol{\Sigma}_{(qq)}^{-1}(n^{-1/2}\tilde{S}_j^{(n)}(\boldsymbol{\gamma}))$, where $\boldsymbol{\Sigma}_{(qq)}$ can be estimated by $\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\sum_l \omega_l(t)\tilde{Y}_l(t)\tilde{\mathbf{Z}}_l^{\otimes 2}\exp(\tilde{\mathbf{Z}}_l^T(t)\hat{\boldsymbol{\gamma}})}{\sum_l \omega_l(t)\tilde{Y}_l(t)\exp(\tilde{\mathbf{Z}}_l^T\hat{\boldsymbol{\gamma}})} - \bar{\tilde{Z}}(\hat{\boldsymbol{\gamma}})^{\otimes 2}\right\} I(D_i = j)$. The detailed information can be found in Fine and Gray (1999). Suppose the null hypothesis we are interesting in testing is

$$H_0 : \tilde{\mathbf{C}}^T\boldsymbol{\gamma} = \tilde{\mathbf{C}}^T\boldsymbol{\gamma}_0$$

for some prespecified $\tilde{\mathbf{C}}$. The corresponding Wald test statistics is $n^{1/2}(\tilde{\mathbf{C}}^T\hat{\boldsymbol{\gamma}} - \tilde{\mathbf{C}}^T\boldsymbol{\gamma}_0)$, which has a normal limiting distribution with mean $\mathbf{0}$ and variance-covariance matrix $\tilde{\mathbf{C}}^T\hat{\boldsymbol{\Sigma}}_{(qq)}^{-1}\hat{\boldsymbol{\Omega}}_{(qq)}\hat{\boldsymbol{\Sigma}}_{(qq)}^{-1}\tilde{\mathbf{C}}$.

2.2 Sample Size Calculation for Competing Risks Data

2.2.1 Sample Size Calculation for Cox Model

Schoenfeld (1981, 1983) presents an explicit sample size calculation formula based on logrank test statistic and Cox model for survival data. It is a two steps procedure.

- Step 1: Given Type I error α, power $1 - \beta$, and hazard ratio ϕ, calculate

the number of failure (information) needed to accumulated by

$$D = \frac{(z_{\alpha/2} + z_{\beta})^2}{(\log \phi)^2 * a_1 * a_2}, \tag{2.19}$$

where a_k is the proportion of patient allocated in group $k, k = 1, 2$, and $Z_{1-\alpha/2}$ and $Z_{1-\beta}$ are the $1-\alpha/2$ and $1-\beta$ quantiles of the standard normal distribution.

Note that D is the number of events which is observed during the study. We can follow the patients until D failures are observed. However, this is not practical. We usually need to make decision about the study duration and total recruitment before the study is launched.

- Step 2: The actual number of patients needed to be enrolled in the study can be calculated by

$$N = D/(a_1 P_1 + a_2 P_2), \tag{2.20}$$

where a_k are the proportions of patients assigned to group $k, k = 1, 2$, and P_k is the probability a patient will experience event of interest in group k.

2.2.2 Sample Size Calculation for Fine and Gray's Model

Fine & Gray's proportional subdistribution hazard model has become a more popular tool to handle competing risks data. Latouche et al. (2004) proposed a similar two steps method to determine the number of patients needed to enroll in the study for the Cox Model. First, to detect subdistribution hazard ratio θ, the number of failures of interest is determined to achieve power of $1 - \beta$ at α nominal level as follows,

$$D = \frac{(Z_{1-\alpha/2} + Z_{\beta})^2}{(\log \theta)^2 p_1 (1 - p_1)}$$

This formula holds in both 'complete data' and 'censoring complete data' case. The total number of patients needed to be enrolled is calculated by

$$N = D/\psi,$$

where ψ is the proportion of failures of interest observed by time of analysis t. This formula looks very similar to the one under Cox model except the parameter used to specify the difference between two treatments is subdistribution hazard ratio instead of cause-specific hazard ratio. We can then use the same way to evaluate ψ as we did for Cox model.

2.3 Counting Processes and Martingales

When we derive the joint distribution of logrank type test statistics, we need to use counting process to express the test statistic, and then use martingale theory to derive the asymptotic distributions. In this section, we briefly review some basic concepts and properties with regards to these two useful tools.

2.3.1 Counting Processes

Given observed competing risks survival data $\{(X_i, \delta_i, Z_i), i = 1, \ldots, n\}$, the information we collect is called counting process

$$N_{ij}(x) = I(X_i \leq x, \delta_i = j),$$

and risk process

$$R(x) = I(X_i \geq x),$$

where i indicates the patient, j denotes the competing risk, and t is the calendar time.

$$n(x) = \sum_{i=1}^{n} I(X_i \geq x) = \sum_{i=1}^{n} R(x)$$

is counting the number of patients at risk at time x.

2.3.2 Martingales

Given a time point t, we know that the expected number of events at t is equal to the number of patients still at risk multiplied by the probability of failure at that time. The jump of $N_{ij}(t)$ at time t can be predicted by

$$E(dN_{ij}(t)|F_{t-}) = I(X_i \geq t)\lambda_j(t)dt, \tag{2.21}$$

where F_{t-} denotes the filtration by time t, which is interpreted as a collection of histories just prior to t. (2.21) is usually called intensity process of counting process $N_{ij}(t)$. It is the expectation of whether a failure due to cause j would happen in the next small interval dt given all the histories prior to t. The difference between expectation and the reality is measured by

$$M_{ij}(x) = N_{ij}(x) - \int_0^x \lambda_j(u)I(X_i \geq u)du.$$

Andersen et al. (1993) proved that M_{ij} is a local square integrable martingale with varitation process

$$\langle M_{ij} \rangle (t) = \int_0^t I(X_i \geq t)d\Lambda_{ij}(s).$$

Martingales has zero means and independent increments, which means $E[M_{ij}(t)] = 0$ and $E[M_{ij}(t)|F_s] = M_{ij}(s)$ for all $s < t$, where F_s contains all the histories up to time s. There are some important properties for martingale. We review one of them which will be useful in our later derivation. The details of the proof are in Andersen et al. (1993).

Lemma. If M is a finite variation local square integrable martingale, and H is a locally bounded predictable process, then $\int HdM$ is still a local square

integrable martingale with variation

$$\left\langle \int H dM \right\rangle = \int H^2 d \langle M \rangle .$$

CHAPTER 3

Joint Inference for Competing Risks Data

The purpose of this chapter is to develop joint inference procedures to assess the effects of a variable on $\lambda_1(t)$ and $F_1(t)$ simultaneously. We allow independent right censoring in addition to competing risks. We first consider the two-sample comparison problem with respect to both $\lambda_1(t)$ and $F_1(t)$. By establishing the asymptotic joint distribution of the weighted log-rank test statistic for $\lambda_1(t)$ and the Gray (1988) test statistic for $F_1(t)$ (or $\tilde{\lambda}_1(t)$), we derive two-sample joint tests for $\lambda_1(t)$ and $F_1(t)$. We then extend our results to other related quantities and to a regression setting based on Cox's type models for CSH and CIF.

3.1 Two Sample Joint Test for Competing Risks Data

3.1.1 Two Sample Joint Test for Cause-Specific Hazard and Cumulative Incidence

Suppose that there are two independent groups of subjects. Let T_{ik}, D_{ik}, and C_{ik} denote the continuous failure time, the type of failure, and the censoring time, respectively, for subject i in group k, $i = 1, \ldots, n_k$, $k = 1, 2$. Assume that the triplets (T_{ik}, D_{ik}, C_{ik}) for different subjects within each group are independent and identically distributed and that the censoring time C_{ik} is independent of the survival time T_{ik}. The two groups are allowed to have different censoring

distributions. For group k ($k = 1, 2$), one observes a right censored competing risks survival data $\{(X_{ik}, \delta_{ik}), i = 1, \ldots, n_k\}$, where $X_{ik} = \min(T_{ik}, C_{ik})$ and $\delta_{ik} = D_{ik} I(T_{ik} \leq C_{ik})$. Let $S_k(t) = P(T_{ik} > t)$ and $S_k^c(t) = P(C_{ik} > t)$. For group k ($k = 1, 2$), let $\lambda_{1k}(t)$, $F_{1k}(t)$, and $\tilde{\lambda}_{1k}(t)$ denote the cause-specific hazard function, the cumulative incidence function, and the sub-distribution hazard function, respectively, for type 1 failure. We develop nonparametric tests for the following null hypothesis

$$H_0 : \lambda_{11}(t) = \lambda_{12}(t) \quad \text{and} \quad F_{11}(t) = F_{12}(t) \quad \text{for all } t > 0, \tag{3.1}$$

or equivalently

$$H_0 : \lambda_{11}(t) = \lambda_{12}(t) \quad \text{and} \quad \tilde{\lambda}_{11}(t) = \tilde{\lambda}_{12}(t) \quad \text{for all } t > 0. \tag{3.2}$$

To test the joint null hypothesis (3.1), we establish the joint limiting distribution of U_{11} and $\tilde{U}_{11}$ below.

Theorem 1 *Let U_{11} and $\tilde{U}_{11}$ be defined by (2.2) and (2.6). Under the null hypothesis (3.1), $n^{-1/2}(U_{11}, \tilde{U}_{11})$ has an asymptotically bivariate normal distribution with mean $\mathbf{0}$ and variance-covriance matrix $\mathbf{\Sigma}^{(1)} = (\sigma_{ij}^{(1)})$, where $\mathbf{\Sigma}^{(1)}$ is defined in (3.4) and (3.7). Furthermore, $\sigma_{11}^{(1)}$ and $\sigma_{22}^{(1)}$ are consistently estimated by (2.3) and (2.7), and the covariance $\sigma_{12}^{(1)}$ is consistently estimated by*

$$\begin{aligned} \hat{\sigma}_{12}^{(1)} &= n^{-1} \left\{ \int_0^\tau W_1(t) \tfrac{Y_2(t)}{Y_\cdot(t)} \hat{V}_{11}(t) + \hat{c}_1^n(\tau) \int_0^\tau W_1(t) \tfrac{Y_2(t)}{Y_\cdot(t)} \hat{E}_{11}(t) \hat{h}_1^{-1}(t) \right\} Y_1(t) d\hat{\Lambda}_{11}(t) \\ &\quad + n^{-1} \left\{ \int_0^\tau W_1(t) \tfrac{Y_1(t)}{Y_\cdot(t)} \hat{V}_{12}(t) + \hat{c}_2^n(\tau) \int_0^\tau W_1(t) \tfrac{Y_1(t)}{Y_\cdot(t)} \hat{E}_{12}(t) \hat{h}_2^{-1}(t) \right\} Y_2(t) d\hat{\Lambda}_{12}(t), \end{aligned} \tag{3.3}$$

where $\hat{\Lambda}_{1k}(\tau) = \int_0^\tau Y_k^{-1}(t) dN_{1k}(t)$, $\hat{V}_{jk}(t) = \left[\hat{d}_{jk}(t) - \hat{E}_{jk}(t) \hat{c}_k(t) \right] \hat{h}_k^{-1}(t)$, $\hat{E}_{jk}(t) = I(j = 1) - \hat{G}_{1k}(t-)/\hat{S}_k(t-)$, and all other quantities are defined in (2.8).

Proof for Theorem 1. Let $M_{jk}(t) = N_{jk}(t) - \int_0^t Y_k(u) d\Lambda_{jk}(u)$, where $\Lambda_{jk}(t) = \int_0^t \lambda_{jk}(u) du$ is the cumulative cause-specific hazard for cause j in group k. Under

the null hypothesis, we can rewrite (2.2) as

$$n^{-1/2}U_{11} = \int_0^\tau W_1(t)\frac{Y_1(t)Y_2(t)}{Y_\cdot(t)}\left\{\frac{dM_{11}(t)}{Y_1(t)} - \frac{dM_{12}(t)}{Y_2(t)}\right\},$$

and (4.19) as

$$n^{-1/2}\tilde{U}_{11} = \sum_{j=1}^{2}\sum_{k=1}^{2}\left\{A_{jk}(\tau) + C_k(\tau)B_{jk}(\tau)\right\},$$

where

$$\begin{aligned}
A_{jk}(\tau) &= \int_0^\tau \left[D_{jk}(t) - E_{jk}(t)C_k(t)\right]\hat{h}_k^{-1}(t)n^{-1/2}dM_{jk}(t),\\
B_{jk}(\tau) &= \int_0^\tau E_{jk}(t)\hat{h}_k^{-1}(t)n^{-1/2}dM_{jk}(t),\\
C_k(\tau) &= \int_0^\tau n^{-1}\tilde{W}(t)R_1(t)\left[I(k=1) - R_k(t)/R_\cdot(t)\right]/\hat{G}_{1k}(t-)dF_{1k}(t),\\
D_{jk}(\tau) &= I(j=1)n^{-1}\tilde{W}(\tau)R_1(\tau)\left[I(k=1) - R_k(\tau)/R_\cdot(\tau)\right]/\hat{G}_{1k}(\tau-),\\
E_{jk}(\tau) &= I(j=1) - G_{1k}(\tau)/S_k(\tau).
\end{aligned}$$

Under usual regularity conditions, by using multivariate martingale central limiting theorem (Fleming and Harrington (1991), Theorem 5.3.5), we can prove that $n^{-1/2}(U_{11}, \tilde{U}_{11})^T$ has a asymptotic multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance $\mathbf{\Sigma}^{(1)} = (\sigma_{ij}^{(1)})$, where $\sigma_{11}^{(1)}$ and $\sigma_{22}^{(1)}$ are developed by Fleming and Harrington (1991); Gray (1988),

$$\begin{aligned}
\sigma_{11}^{(1)} &= \sigma^2 = \textstyle\int_0^\tau w_1^2(t)\frac{y_1(t)y_2(t)}{y_\cdot(t)}d\Lambda_{11}(t),\\
\sigma_{22}^{(1)} &= \tilde{\sigma}^2 = \textstyle\sum_{k=1}^{2} n^{-1}\left\{\int_0^{\tau_1} a_k^2(t)h_k^{-1}(t)h_\cdot^{-1}(t)dF_{1\cdot}(t) + \int_0^{\tau_1} b_{2k}^2(t)h_k^{-2}(t)dF_{2k}(t)\right\},
\end{aligned} \tag{3.4}$$

with

$$\begin{aligned}
\Lambda_{jk}(t) &= \int_0^t \lambda_{jk}(u)du, \\
a_k(t) &= d_{jk}(t) + b_{jk}(t), \\
b_{jk}(t) &= \left[I(j=1) - G_{1k}(t)/S_k(t)\right]\left[c_k(\tau_1) - c_k(t)\right], \\
c_k(t) &= \int_0^t d_{1k}(u)\tilde{\lambda}_{1k}(u)du, \\
d_{jk}(t) &= I(j=1)\tilde{W}(t)R_1(t)\left[I(k=1) - h_k(t)/h_\cdot(t)\right]/G_{1k}(t), \\
h_k(t) &= I(t \le \tau_k)y_k(t)/S_k(t), \\
h_\cdot(t) &= I(t \le max(\tau_1,\tau_2))(y_1(t) + y_2(t)/S_k(t), \\
y_k(t) &= p_k S_k(t) S_k^c(t), \\
p_k &= n_k/(n_1+n_2).
\end{aligned}$$

The covariance $\sigma_{12}^{(1)}$ between the two test statistics is

$$\begin{aligned}
& \left\langle n^{-1/2}U_{11}, n^{-1/2}\tilde{U}_{11}\right\rangle \\
= \;& n^{-1}\left\langle \int_0^\tau W_1(t)\tfrac{Y_1(t)Y_2(t)}{Y_\cdot(t)}\left\{\tfrac{dM_{11}(t)}{Y_1(t)} - \tfrac{dM_{12}(t)}{Y_2(t)}\right\}, \textstyle\sum_{k=1}^2\sum_{j=1}^2\left\{A_{jk}(\tau) + C_k(\tau)B_{jk}(\tau)\right\}\right\rangle \\
= \;& n^{-1}\left\langle \int_0^\tau W_1(t)\tfrac{Y_1(t)Y_2(t)}{Y_\cdot(t)}\left\{\tfrac{dM_{11}(t)}{Y_1(t)} - \tfrac{dM_{12}(t)}{Y_2(t)}\right\}, \right. \\
& \textstyle\int_0^t V_{11}(t)dM_{11}(t) + C_1(\tau)\int_0^\tau E_{11}(t)\hat{h}_1^{-1}(t)dM_{11}(t) \\
& \left. + \textstyle\int_0^\tau V_{12}(t)dM_{12}(t) + C_2(\tau)\int_0^\tau E_{12}(t)\hat{h}_2^{-1}(t)dM_{12}(t)\right\rangle \\
= \;& n^{-1}\left\{\int_0^\tau W_1(t)\tfrac{Y_2(t)}{Y_\cdot(t)}V_{11}(t) + C_1(\tau)\int_0^\tau W_1(t)\tfrac{Y_2(t)}{Y_\cdot(t)}E_{11}(t)\hat{h}_1^{-1}(t)\right\}d\langle M_{11}, M_{11}\rangle(t) \\
& + n^{-1}\left\{\int_0^\tau W_1(t)\tfrac{Y_1(t)}{Y_\cdot(t)}V_{12}(t) + C_2(\tau)\int_0^\tau W_1(t)\tfrac{Y_1(t)}{Y_\cdot(t)}E_{12}(t)\hat{h}_2^{-1}(t)\right\}d\langle M_{12}, M_{12}\rangle(t),
\end{aligned} \tag{3.5}$$

where $V_{jk}(t) = \left[D_{jk}(t) - E_{jk}(t)C_k(t)\right]\hat{h}_k^{-1}(t)$. Note that $M_{jk}(t)$ are orthogonal square intergrable martingales with predictable variation process

$$\langle M_{jk}(t), M_{j'k'}(t)\rangle = \gamma_{jj'}\gamma_{kk'}\int_0^t Y_k(u)d\Lambda_{jk}(u), \tag{3.6}$$

where $\gamma_{uv} = 1$ if $u = v$. After plugging (3.6) into (3.5), we have

$$\begin{aligned}
&\left\langle n^{-1/2}U_{11}, n^{-1/2}\tilde{U}_{11}\right\rangle \\
= \;& n^{-1}\left[\int_0^\tau W_1(t)\tfrac{Y_2(t)}{Y_\cdot(t)}V_{11}(t) + C_1(\tau)\int_0^\tau W_1(t)\tfrac{Y_2(t)}{Y_\cdot(t)}E_{11}(t)\hat{h}_1^{-1}(t)\right]Y_1(t)d\Lambda_{11}(t) \\
& + n^{-1}\left[\int_0^\tau W_1(t)\tfrac{Y_1(t)}{Y_\cdot(t)}V_{12}(t) + C_2(\tau)\int_0^\tau W_1(t)\tfrac{Y_1(t)}{Y_\cdot(t)}E_{12}(t)\hat{h}_2^{-1}(t)\right]Y_2(t)d\Lambda_{12}(t),
\end{aligned}$$

which converges in probability to

$$\begin{aligned}
\sigma_{12}^{(1)} = \;& \left[\int_0^\tau w_1(t)\tfrac{y_2(t)}{y_\cdot(t)}v_{11}(t) + c_1(\tau)\int_0^\tau w_1(t)\tfrac{y_2(t)}{y_\cdot(t)}e_{11}(t)h_1^{-1}(t)\right]y_1(t)d\Lambda_{11}(t) \\
& + \left[\int_0^\tau w_1(t)\tfrac{y_1(t)}{y_\cdot(t)}v_{12}(t) + c_2(\tau)\int_0^\tau w_1(t)\tfrac{y_1(t)}{y_\cdot(t)}e_{12}(t)h_2^{-1}(t)\right]y_2(t)d\Lambda_{12}(t),
\end{aligned} \tag{3.7}$$

where

$$\begin{aligned}
e_{jk}(t) &= I(j=1) - G_{1k}(t)/S_k(t) \\
v_{jk}(t) &= \left[d_{jk}(t) - e_{jk}(t)c_k(t)\right]h_k^{-1}(t).
\end{aligned}$$

Finally a consistent estimator of $\sigma_{12}^{(1)}$ is obtained by replacing each unknown quantity in (3.3) by its consistent sample estimate. □

3.1.1.1 Chi-square Joint Test for (3.1)

Define

$$X^2 = n^{-1}\left(U_{11}, \tilde{U}_{11}\right)\hat{\mathbf{\Sigma}}^{(1)(-1)}\begin{pmatrix} U_{11} \\ \tilde{U}_{11}\end{pmatrix}.$$

It follows from Theorem 1 that under (3.1), X^2 has an asymptotically chi-square distribution with 2 degrees of freedom. This leads to the following chi-square test for (3.1):

$$\text{Reject (3.1) at level } \alpha \text{ if } X^2 > \chi_2^2(\alpha),$$

where $\chi_2^2(\alpha)$ is the upper $1-\alpha$ percentile of the standard χ_2^2 distribution.

Rejection of (3.1) by the above chi-square test implies that there is a difference in either cause-specific hazard or cumulative incidence between the two groups.

However, it does not indicate which individual quantity has a difference. The following maximum test provides an alternative joint test that allows one to draw a conclusion on each individual quantity. It also allows one-sided test.

3.1.1.2 Maximum Joint Test for (3.1)

Define

$$M = \max(|Z_{11}|, |\tilde{Z}_{11}|)),$$

where $Z_{11} = n^{-1/2}U_{11}/\sqrt{\hat{\sigma}_{11}^{(1)}}$ and $\tilde{Z}_{11} = n^{-1/2}\tilde{U}_{11}/\sqrt{\hat{\sigma}_{22}^{(1)}}$. We would reject (3.1) if the observed M is large. It follows from Theorem 1 that for large samples, the distribution of $(Z_{11}, \tilde{Z}_{11})$ can be approximated by the bivariate normal distribution $N\left((0,0)^T, (1,1,\hat{\rho})\right)$, where $\hat{\rho} = \frac{\hat{\sigma}_{12}^{(1)}}{\sqrt{\hat{\sigma}_{11}^{(1)}}\sqrt{\hat{\sigma}_{22}^{(1)}}}$. Thus we can approximate the distribution of M using Monte Carlo simulation. Specifically, we generate N pairs of random variables from the bivariate normal distribution $N\left((0,0)^T, (1,1,\hat{\rho})\right)$. For the l-th generated pair, compute the maximum absolute value, and denoted it by M_l. Let T_α be the upper $100(1-\alpha)$-th sample quantile of $M_1, \ldots, M_N$. Reject the null hypothesis (3.1) at level α if $M > M_\alpha$.

Remark 3.1. It is straightforward to modify the maximum joint test procedure to test one-sided alternative(s) based on either $M = \max(Z_{11}, \tilde{Z}_{11})$, $M = \max(|Z_{11}|, \tilde{Z}_{11})$, or $M = \max(Z_{11}, |\tilde{Z}_{11}|)$ as deemed appropriate.

Remark 3.2. (K-Sample Joint Tests) The above two-sample joint tests can be easily extended to the K-sample problem ($K \geq 2$) for the following null hypothesis

$$H_0 : \lambda_{11}(t) = \cdots = \lambda_{1K}(t) \quad \text{and} \quad F_{11}(t) = \cdots = F_{1K}(t) \quad \text{for all } 0 < t < \tau, \tag{3.8}$$

where τ is some pre specified fixed time. Similar to Theorem 1, it can be shown that under the null hypothesis (3.8), $n^{-1/2}(U_{11}, \cdots, U_{1K-1}, \tilde{U}_{11}, \cdots, \tilde{U}_{1K-1})$ has an asymptotic multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix Σ^*, where the elements of Σ^*, can be obtained as the limit of the pairwise covariances of the test statistics. From Kulathinal and Gasbarra (2002), we have

$$\hat{Cov}(n^{-1/2}U_{1k}, n^{-1/2}U_{1k'}) = -\int_0^\tau W_1^2(t)\frac{Y_k(t)Y_{k'}(t)}{Y_\cdot(t)}d\Lambda_1(t),$$

where $k, k' = 1, \cdots, K$. $\hat{Cov}(n^{-1/2}\tilde{U}_{1k}, n^{-1/2}\tilde{U}_{1k'})$ is given by equation (2.10) on page 1146 of Gray (1988). Similar to the proof of Theorem 1, we can show that

$$\begin{aligned}
& \hat{Cov}(n^{-1/2}U_{1k}, n^{-1/2}\tilde{U}_{1k'}) \\
= & \left(\int_0^\tau W_1(t)\hat{V}_{1k'k}(t) + \hat{c}_{k'k}(\tau)\int_0^\tau W_1(t)\hat{E}_{1k}(t)\hat{h}_k^{-1}(t)\right) Y_k(t)d\hat{\Lambda}_{1k}(t) \\
+ & \sum_{l=1}^K \left(\int_0^\tau W_1(t)\tfrac{Y_k(t)}{Y_\cdot(t)}\hat{V}_{1k'l}(t) + \hat{c}_{k'l}(\tau)\int_0^\tau W_1(t)\tfrac{Y_k(t)}{Y_\cdot(t)}\hat{E}_{1l}(t)\hat{h}_l^{-1}(t)\right) Y_l(t)d\hat{\Lambda}_{1l}(t),
\end{aligned} \tag{3.9}$$

where

$$\begin{aligned}
\hat{\Lambda}_{1k}(\tau) &= \int_0^\tau Y_k^{-1}(t)dN_{1k}(t), \\
\hat{V}_{jkl}(t) &= \left[\hat{D}_{jkl}(t) - \hat{E}_{jl}(t)\hat{c}_{kl}(t)\right]\hat{h}_l^{-1}(t), \\
\hat{D}_{jkl} &= n^{-1}I(j=1)\tilde{W}(t)R_k(t)\left[I(k=l) - \hat{h}_l(t)/\hat{h}_\cdot(t)\right]/\hat{G}_{1\cdot}(t-), \\
\hat{c}_{kl}(t) &= n^{-1}\int_0^t \hat{d}_{1kl}(u)\hat{G}_{1\cdot}(u-)^{-1}\hat{h}_\cdot^{-1}(u)dN_{1\cdot}(u), \\
\hat{E}_{jk}(t) &= I(j=1) - \hat{G}_{1k}(t-)/\hat{S}_k(t-),
\end{aligned}$$

and all other quantities are defined in (2.8).

Derivation for Covariance between U_{1k} and $\tilde{U}_{1k'}$

$$
\begin{aligned}
& n^{-1/2} < (U_{1k}, \tilde{U}_{1k'}) > \\
= \; & < \int_0^\tau W_1(t) Y_K(t) \left(\tfrac{dN_{1k}(t)}{Y_k(t)} - \tfrac{dN_{1.}(t)}{Y.(t)} \right), \sum_{l=1}^K \sum_{j=1}^2 \left(A_{jk'l}(\tau) + c_{k'l}(\tau) B_{jl}(\tau) \right) > \\
= \; & < \int_0^\tau W_1(t) Y_k(t) \left(\tfrac{dM_{1k}(t)}{Y_k(t)} - \tfrac{dM_{1.}(t)}{Y.(t)} \right), \\
& \sum_{l=1}^K \int_0^\tau V_{1k'l}(t) dM_{1l}(t) + c_{k'l}(\tau) \int_0^\tau E_{1l}(t) \hat{h}_l^{-1}(t) dM_{1l}(t) > \\
= \; & < \int_0^\tau W_1(t) Y_k(t) \tfrac{dM_{1k}(t)}{Y_k(t)}, \int_0^\tau V_{1k'k}(t) dM_{1k}(t) + c_{k'k}(\tau) \int_0^\tau E_{1k}(t) \hat{h}_k^{-1}(t) dM_{1k}(t) > \\
+ \; & \int_0^\tau W_1(t) Y_k(t) \tfrac{\sum_{l=1}^K dM_{1l}(t)}{Y.(t)}, \sum_{l=1}^K \int_0^\tau V_{1k'l}(t) dM_{1l}(t) + c_{k'l}(\tau) \int_0^\tau E_{1l}(t) \hat{h}_l^{-1}(t) dM_{1l}(t) > \\
= \; & \left(\int_0^\tau W_1(t) V_{1k'k}(t) + c_{k'k}(\tau) \int_0^\tau W_1(t) E_{1k}(t) \hat{h}_k^{-1}(t) \right) d < M_{1k}(t), M_{1k}(t) > \\
+ \; & \sum_{l=1}^K \left(\int_0^\tau W_1(t) \tfrac{Y_k(t)}{Y.(t)} V_{1k'l}(t) + c_{k'l}(\tau) \int_0^\tau W_1(t) \tfrac{Y_k(t)}{Y.(t)} E_{1l}(t) \hat{h}_l^{-1}(t) \right) d < M_{1l}(t), M_{1l}(t) > \\
= \; & \left(\int_0^\tau W_1(t) V_{1k'k}(t) + c_{k'k}(\tau) \int_0^\tau W_1(t) E_{1k}(t) \hat{h}_k^{-1}(t) \right) Y_k(t) d\Lambda_{1k}(t) \\
+ \; & \sum_{l=1}^K \left(\int_0^\tau W_1(t) \tfrac{Y_k(t)}{Y.(t)} V_{1k'l}(t) + c_{k'l}(\tau) \int_0^\tau W_1(t) \tfrac{Y_k(t)}{Y.(t)} E_{1l}(t) \hat{h}_l^{-1}(t) \right) Y_l(t) d\Lambda_{1l}(t),
\end{aligned}
\tag{3.10}
$$

where $V_{jkl}(t) = \left[D_{jkl}(t) - E_{jl}(t) c_{kl}(t) \right] \hat{h}_l^{-1}(t)$ and all other quantities are defined in Gray (1988) on page 1153. $n^{-1/2} < (U_{1k}, \tilde{U}_{1k'}) >$ converges in probability to

$$
\begin{aligned}
& cov(n^{-1/2} U_{1k}, n^{-1/2} \tilde{U}_{1k'}) \\
= \; & \left(\int_0^\tau w_1(t) v_{1k'k}(t) + c_{k'k}(\tau) \int_0^\tau w_1(t) e_{1k}(t) \hat{h}_k^{-1}(t) \right) y_k(t) d\Lambda_{1k}(t), \\
& + \sum_{l=1}^K \left(\int_0^\tau w_1(t) \tfrac{y_k(t)}{y.(t)} v_{1k'l}(t) + c_{k'l}(\tau) \int_0^\tau w_1(t) \tfrac{y_k(t)}{y.(t)} e_{1l}(t) \hat{h}_l^{-1}(t) \right) y_l(t) d\Lambda_{1l}(t),
\end{aligned}
\tag{3.11}
$$

where

$$
\begin{aligned}
e_{jk}(t) &= I(j=1) - G_{1k}(t)/S_k(t) \\
v_{jkl}(t) &= \left[d_{jkl}(t) - e_{jl}(t) c_{kl}(t) \right] h_l^{-1}(t),
\end{aligned}
$$

and d_{jkl} is defined in Gray (1988) on page 1146. Finally a consistent estimator of $cov(n^{-1/2} U_{1k}, n^{-1/2} \tilde{U}_{1k'})$ is obtained by replacing each unknown quantity in (3.11) by its consistent sample estimate. $\square$

For Chisquare test, define

$$
X^2 = n^{-1} \left(U_{11}, \cdots, U_{1K-1}, \tilde{U}_{11}, \cdots, \tilde{U}_{1K-1} \right) \Sigma^{*-1} \left(U_{11}, \cdots, U_{1K-1}, \tilde{U}_{11}, \cdots, \tilde{U}_{1K-1} \right)^T .
$$

We can prove that that under (3.8), X^2 has an asymptotically chi-square distribution with 2(K-1) degrees of freedom. Reject (3.8) at level α if $X^2 > \chi^2_{2(K-1)}(\alpha)$, where $\chi^2_{2(K-1)}(\alpha)$ is the upper $1-\alpha$ percentile of the standard $\chi^2_{2(K-1)}$ distribution.

For maximum test, define

$$T^* = \max(|Z_{11}|, \cdots, |Z_{1K-1}|, |\tilde{Z}_{11}|, \cdots, |\tilde{Z}_{1K-1}|),$$

where Z_{1k} and $\tilde{Z}_{1k}$ are the standardized test statistics of U_{1k} and $\tilde{U}_{1k}$. The Test can be conducted by generating Monte Carlo sample which is similar to what did for Maximum joint test in two sample case.

Remark 3.3: It can be easily shown that for group k, the three pairs of functions $(\lambda_{1k}(\cdot), F_{1k}(\cdot))$, $(\lambda_{1k}(\cdot), \lambda_{\cdot k}(\cdot))$, and $(\lambda_{1k}(\cdot), \lambda_{2k}(\cdot))$ uniquely determine each other and that each pair uniquely determines the distribution of the observed pair (X_{ik}, δ_{ik}). A practical question is then which pair should one consider in a particular study when studying the effects on a variable on type 1 failure. Such a decision should be made based upon which pair is scientifically more relevant in the study although it is always sensible to use (CSH, CIF). For instance, in the kidney transplantation program example discussed in the beginning of the introduction section, death before becoming eligible for transplantation is a competing risk event for the waiting time to become eligible. The two competing events are negatively correlated. Suppose that one is interested in studying the effects of a weight loss intervention to improve the waiting time to become eligible for transplantation, then the (CSH, CIF) pair would be of primary interest. One would not want to study the (CSH, ACH) pair because ACH combines two negatively correlated competing events and does not have a meaningful interpretation. The (CSH, OSH) pair would also not be very interesting if the weight loss intervention is not expected to significantly impact the risk of death before becoming

eligible for transplantation. On the other hand, in a randomized confirmatory Phase III clinical trial of a new treatment versus a standard treatment for a specific disease, it might be preferable to use the disease-specific survival (effects on the target disease) and overall survival (effects on a patient's overall health), or equivalently (CSH, ACH), as co-primary endpoints, although the other two pairs are also meaningful. Finally, although the three null hypotheses (3.1), (3.2), and (3.12) are equivalent, their corresponding alternative hypotheses are not. Consequently, a joint test for a specific pair is powered to detect group differences in the direction of that pair. For example, when there are group differences in CSH, OCH, and CIF, but no group difference in ACH, the maximum joint test for (CSH, ACH) was observed to have poorer power than that for (CSH, OCH) or (CSH, CIF) in our limited simulation studies.

3.1.2 Two-Sample Joint Tests for Other Quantities

Some related quantities are also useful to study the effects of a variable on type 1 failure. For group k, let $\lambda_{\cdot k}(t)$ and $\lambda_{2k}(t)$ denote the all-cause hazard function and the cause-specific hazard function for type 2 failure, respectively. Then, it can be shown that the three pairs $(\lambda_{1k}(\cdot), F_{1k}(\cdot))$, $(\lambda_{1k}(\cdot), \lambda_{\cdot k}(\cdot))$, and $(\lambda_{1k}(\cdot), \lambda_{2k}(\cdot))$ uniquely determine each other. Therefore, the hypothesis (3.1) is equivalent to

$$H_0 : \lambda_{11}(t) = \lambda_{12}(t) \quad \text{and} \quad \lambda_{\cdot 1}(t) = \lambda_{\cdot 2}(t) \quad \text{for all } t > 0, \tag{3.12}$$

or

$$H_0 : \lambda_{11}(t) = \lambda_{12}(t) \quad \text{and} \quad \lambda_{21}(t) = \lambda_{22}(t) \quad \text{for all } t > 0. \tag{3.13}$$

3.1.2.1 Two-Sample Tests for Cause-Specific Hazard and All-Cause Hazard

Let

$$U_{\cdot k} = \int_0^\tau W_{\cdot}(t) Y_k(t) \left\{ \frac{dN_{\cdot k}(t)}{Y_k(t)} - \frac{dN_{\cdot\cdot}(t)}{Y_{\cdot}(t)} \right\}, \tag{3.14}$$

be the weighted log-rank test statistic for $H_0 : \lambda_{\cdot 1}(t) = \lambda_{\cdot 2}(t)$ for all $t > 0$, where $N_{\cdot k}(t) = \sum_{j=1}^{2} N_{jk}(t)$, $N_{\cdot\cdot}(t) = \sum_{k=1}^{2} \sum_{j=1}^{2} N_{jk}(t)$, and $W_{\cdot}(t)$ is a predictable weight function, which converges in probability to some deterministic function $w_{\cdot}(t)$ when $n \to \infty$.

Theorem 2 *Let U_{11} and $U_{\cdot 1}$ be defined by (2.2) and (3.14). Then, $n^{-1/2}(U_{11}, U_{\cdot 1})$ has an asymptotic bivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Sigma}^{(2)} = (\sigma_{ij}^{(2)})$ defined by (3.16). Furthermore, $\boldsymbol{\Sigma}^{(2)}$ is consistently estimated by $\hat{\boldsymbol{\Sigma}}^{(2)} = (\hat{\sigma}_{ij}^{(2)})$ where*

$$\begin{aligned} \hat{\sigma}_{11}^{(2)} &= \int_0^\tau W_1^2(t) \tfrac{Y_1(t)Y_2(t)}{Y_1(t)+Y_2(t)} \tfrac{dN_{11}(t)}{Y_1(t)}, \\ \hat{\sigma}_{22}^{(2)} &= \int_0^\tau W_{\cdot}^2(t) \tfrac{Y_1(t)Y_2(t)}{Y_1(t)+Y_2(t)} \tfrac{dN_{\cdot 1}(t)}{Y_1(t)}, \\ \hat{\sigma}_{12}^{(2)} &= \int_0^\tau W_1(t) W_{\cdot}(t) \tfrac{Y_1(t)Y_2(t)}{Y_1(t)+Y_2(t)} \tfrac{dN_{11}(t)}{Y_1(t)}. \end{aligned} \tag{3.15}$$

Proof for Theorem 2. Under usual regularity condition, by using multivariate martingale central limiting theorem (Fleming and Harrington (1991), Theorem 5.3.5), we can prove that $(n^{-1/2}U_{11}, n^{-1/2}U_{\cdot 1})^T$ has a bivariate normal limiting distribution with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Sigma}^{(2)} = (\sigma_{ij}^{(2)})$, where

$$\begin{aligned} \sigma_{11}^{(2)} &= \int_0^\tau w_1^2(t) \tfrac{y_1(t)y_2(t)}{y_1(t)+y_2(t)} d\Lambda_{11}(t), \\ \sigma_{22}^{(2)} &= \int_0^\tau w_{\cdot}^2(t) \tfrac{y_1(t)y_2(t)}{y_1(t)+y_2(t)} d\Lambda_{\cdot 1}(t), \\ \sigma_{12}^{(2)} &= \int_0^\tau w_1(t) w_{\cdot}(t) \tfrac{y_1(t)y_2(t)}{y_1(t)+y_2(t)} d\Lambda_{11}(t). \end{aligned} \tag{3.16}$$

Here $\Lambda_{\cdot 1}(t) = \int_0^t \lambda_{\cdot 1} dt$ is the all-cause cumulative hazard in group 1. Note that $\sigma_{12}^{(2)}$ is the limit of

$$\begin{aligned}
& \left\langle n^{-1/2} U_{11}, n^{-1/2} U_{\cdot 1} \right\rangle \\
= \; & n^{-1} \left\langle \int_0^\tau W_1(t) \frac{Y_1(t) Y_2(t)}{Y_\cdot(t)} \left\{ \frac{dM_{11}(t)}{Y_1(t)} - \frac{dM_{12}(t)}{Y_2(t)} \right\}, \int_0^\tau W_\cdot(t) \frac{Y_1(t) Y_2(t)}{Y_\cdot(t)} \left\{ \frac{dM_{\cdot 1}(t)}{Y_1(t)} - \frac{dM_{\cdot 2}(t)}{Y_2(t)} \right\} \right\rangle \\
= \; & n^{-1} \left\{ \int_0^t W_1(t) W_\cdot(t) \frac{Y_1^2(t) Y_2^2(t)}{Y_\cdot(t)^2} \left(\frac{d<M_{11}(t), M_{\cdot 1}(t)>}{Y_1^2(t)} + \frac{d<M_{12}(t), M_{\cdot 2}(t)>}{Y_2^2(t)} \right) \right\} \\
= \; & n^{-1} \left\{ \int_0^t W_1(t) W_\cdot(t) \frac{Y_1^2(t) Y_2^2(t)}{Y_\cdot(t)^2} \left(\frac{d\Lambda_{11}(t)}{Y_1(t)} + \frac{d\Lambda_{12}(t)}{Y_2(t)} \right) \right\},
\end{aligned}$$

which converges in probability to $\int_0^t w_1(t) w_\cdot(t) \frac{y_1(t) y_2(t)}{y_\cdot(t)} d\Lambda_{11}(t)$ Similarly, the diagonal elements are the asymptotic variances of $n^{-1/2} U_{11}$ and $n^{-1/2} U_{\cdot 1}$, respectively (Fleming and Harrington, 1991).

Finally, a consistent estimate of $\mathbf{\Sigma}^{(2)}$ is obtained by replacing each quantity in (3.16) with its consistent empirical estimate. $\square$

The asymptotic results in Theorem 2 allow one to construct a chi-square joint test and a maximum joint test for (3.12) along the lines of Section 3.1.1.1 and 3.1.1.2.

3.1.2.2 Two-Sample Joint Tests for Cause-Specific Hazard and Other-Cause Hazard

Let

$$U_{2k} = \int_0^\tau W_2(t) Y_k(t) \left\{ \frac{dN_{2k}(t)}{Y_k(t)} - \frac{dN_{2\cdot}(t)}{Y_\cdot(t)} \right\}, \tag{3.17}$$

be the weighted log-rank test statistic for $H_0 : \lambda_{21}(t) = \lambda_{22}(t)$ for all $t > 0$, where $W_2(t)$ is a proper weight function, which converges in probability to some deterministic function $w_2(t)$ as $n \to \infty$. It's well known that U_{1k} and U_{2k} are asymptotically independent (Prentice et al., 1978). Hence one can construct joint tests for (3.13) based on the joint distribution of the two test statistics.

3.2 Joint Regression Analysis for Competing Risks Data

3.2.1 Joint Regression Analysis of Cause-Specific Hazard and Cumulative Incidence

We now consider joint inference for the cause-specific hazard and the cumulative incidence hazard under a regression setting. Assume that one observes n independent and identically distributed triples $(X_i, \delta_i, \boldsymbol{Z}_i)$, where for subject i $(i = 1, \ldots, n)$, $X_i = \min\{T_i, C_i\}$, $\delta_i = D_i I(T_i \leq C_i)$, T_i is the failure time of interest, C_i is a right censoring time, D_i is discrete random variable taking values on $1, \ldots, J$ with $D_i = j$ indicating that type j failure is observed, and $\mathbf{Z}_i$ is a vector of fixed or time-varying covariates that are observed on $[0, X_i]$. Assume C_i is independent of T_i, D_i and $\mathbf{Z}_i$, and $pr(C_i \geq t) = G^c(t)$.

Again, for simplicity, we assume there are only two types of failures, and type 1 is the event of interest. Let $\lambda_1(t|\mathbf{z})$ and $\tilde{\lambda}_1(t|\mathbf{z})$ be the conditional cause-specific hazard function and the conditional subdistribution hazard function for type 1 failure for an individual with covariate $\mathbf{z}$. Assume the proportional cause-specific hazards model (Prentice et al., 1978)

$$\lambda_1(t|\mathbf{Z}) = \lambda_{10}(t) \exp(\boldsymbol{\beta}_1^T \mathbf{Z}^{(1)}(t)), \tag{3.18}$$

and the proportional subdistribution hazards model (Fine and Gray, 1999)

$$\tilde{\lambda}_1(t|\mathbf{Z}) = \tilde{\lambda}_{10}(t) \exp(\boldsymbol{\gamma}_1^T \mathbf{Z}^{(2)}(t)), \tag{3.19}$$

where $\lambda_{10}(t)$ and $\tilde{\lambda}_{10}(t)$ are unknown baseline cause-specific hazard and baseline subdistribution hazard for type 1 failure, respectively, and $\mathbf{Z}^{(1)}(t)$ and $\mathbf{Z}^{(2)}(t)$ are functions of the original covariates $\mathbf{Z}$ and t that allow time $\times$ covariates interactions. Prentice et al. (1978) showed that inference for $\boldsymbol{\beta}_1$ under the proportional cause-specific hazards model (3.18) can be made using the standard Cox (1972,

1975) partial likelihood method by regarding other types of failure as independent censoring. The proportional subdistribution hazards model (3.19) was introduced by Fine and Gray (1999) who developed large sample inference for $\boldsymbol{\gamma}_1$.

Below we develop joint inference for $\boldsymbol{\beta}_1$ and $\boldsymbol{\gamma}_1$. Specifically, we consider the following joint null hypothesis

$$H_0 : A_1^T \boldsymbol{\beta}_1 = \mathbf{d}_1 \text{ and } A_2^T \boldsymbol{\gamma}_1 = \mathbf{d}_2, \tag{3.20}$$

where A_1 and A_2 are constant matrices, and $\mathbf{d}_1$ and $\mathbf{d}_2$ are constant column vectors.

Following Prentice et al. (1978) and Fine and Gray (1999), let

$$\mathbf{U}_1(\boldsymbol{\beta}_1) = \sum_{i=1}^{n} \int_0^{\infty} \left\{ \mathbf{Z}_i^{(1)}(t) - \bar{\mathbf{Z}}^{(1)}(\boldsymbol{\beta}_1, t) \right\} dN_{i1}(t), \tag{3.21}$$

and

$$\tilde{\mathbf{U}}_1(\boldsymbol{\gamma}_1) = \sum_{i=1}^{n} \int_0^{\infty} \left\{ \mathbf{Z}_i^{(2)}(t) - \bar{\mathbf{Z}}^{(2)}(\boldsymbol{\gamma}_1, t) \right\} \omega_i(t) d\tilde{N}_{i1}(t), \tag{3.22}$$

be the score functions for $\boldsymbol{\beta}_1$ and $\boldsymbol{\gamma}_1$ under models (3.18) and (3.19), respectively, where $\bar{\mathbf{Z}}^{(1)}(\boldsymbol{\beta}_1, t) = \frac{\sum_{l=1}^{n} Y_l(t) \mathbf{Z}_l^{(1)}(t) \exp(\boldsymbol{\beta}_1^T \mathbf{Z}_l^{(1)}(t))}{\sum_{l=1}^{n} Y_l(t) \exp(\boldsymbol{\beta}_1^T \mathbf{Z}_l^{(1)}(t))}$, $Y_i(t) = I\{X_i \geq t\}$ and $N_{i1}(t) = I(X_i \leq t, D_i = 1)$, $\bar{\mathbf{Z}}^{(2)}(\boldsymbol{\gamma}_1, t) = \frac{\sum_{l=1}^{n} \omega_l(t) \tilde{Y}_l(t) \mathbf{Z}_l^{(2)} \exp(\boldsymbol{\gamma}_1^T \mathbf{Z}_l^{(2)}(t))}{\sum_{l=1}^{n} \omega_l(t) \tilde{Y}_l(t) \exp(\boldsymbol{\gamma}_1^T \mathbf{Z}_l^{(2)}(t))}$, $\tilde{N}_{i1}(t) = I(T_i \leq t, D_i = 1)$, $\tilde{Y}_i(t) = 1 - \tilde{N}_{i1}(t-)$, $\omega_i(t) = I(C_i \geq T_i \wedge t) \hat{G}^c(t) / \hat{G}^c(X_i \wedge t)$, and $\hat{G}^c$ is the Kaplan and Meier (1958) estimate of the survival function G^c of the censoring variable C. Note that $\tilde{N}_{i1}(t)$ is different from $N_{i1}(t)$ and may not be observed if the subject is censored, but $\omega_i(t) \tilde{N}_{i1}(t)$ can always be computed.

Let $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\gamma}}_1$ be the solutions of the score equations $\mathbf{U}_1(\boldsymbol{\beta}_1) = 0$ and $\tilde{\mathbf{U}}_1(\boldsymbol{\gamma}_1) = 0$, respectively.

Theorem 3 *Under similar regularity conditions to Andersen et al. (1982) and*

Fine and Gray (1999), we have

$$n^{1/2}\begin{pmatrix}\hat{\boldsymbol{\beta}}_1-\boldsymbol{\beta}_1\\ \hat{\boldsymbol{\gamma}}_1-\boldsymbol{\gamma}_1\end{pmatrix}\xrightarrow{d} N(\mathbf{0},\boldsymbol{\Sigma}^{(1)}),\quad \textit{as } n\to\infty,$$

where $\boldsymbol{\Sigma}^{(1)}$ *is defined by (3.26), (3.27), (3.28), and (3.29). Furthermore,* $\boldsymbol{\Sigma}^{(1)}$ *can be consistently estimated by*

$$\hat{\boldsymbol{\Sigma}}^{(1)}=\begin{pmatrix}\hat{\boldsymbol{\Omega}}^{(1)-1}_{(pp)} & \hat{\boldsymbol{\Omega}}^{(1)-1}_{(pp)}\hat{\boldsymbol{\Omega}}^{(1)}_{(pq)}\hat{\boldsymbol{\Omega}}^{(1)-1}_{(qq)}\\ \hat{\boldsymbol{\Omega}}^{(1)-1}_{(qq)}\hat{\boldsymbol{\Omega}}^{(1)}_{(qp)}\hat{\boldsymbol{\Omega}}^{(1)-1}_{(pp)} & \hat{\boldsymbol{\Omega}}^{(1)-1}_{(qq)}\hat{\boldsymbol{\Omega}}^{*(1)}_{(qq)}\hat{\boldsymbol{\Omega}}^{(1)-1}_{(qq)}\end{pmatrix},\tag{3.23}$$

where

$$\begin{aligned}
\hat{\boldsymbol{\Omega}}^{(1)}_{(pp)} &= \tfrac{1}{n}\textstyle\sum_{i=1}^n\int_0^\infty\left[\frac{\sum_{l=1}^n Y_l(t)\mathbf{Z}_l^{(1)}(t)^{\otimes 2}\exp(\hat{\boldsymbol{\beta}}_1^T\mathbf{Z}_l^{(1)}(t))}{\sum_{l=1}^n Y_l(t)\exp(\hat{\boldsymbol{\beta}}_1^T\mathbf{Z}_l^{(1)}(t))}-\bar{\mathbf{Z}}^{(1)}(\hat{\boldsymbol{\beta}}_1,t)^{\otimes 2}\right]dN_{i1}(t),\\
\hat{\boldsymbol{\Omega}}^{(1)}_{(qq)} &= \tfrac{1}{n}\textstyle\sum_{i=1}^n\left\{\frac{\sum_{l=1}^n \omega_l(t)\tilde{Y}_l(t)\mathbf{Z}_l^{(2)}(t)^{\otimes 2}\exp(\hat{\boldsymbol{\gamma}}_1^T\mathbf{Z}_l^{(2)}(t))}{\sum_{l=1}^n \omega_l(t)\tilde{Y}_l(t)\exp(\hat{\boldsymbol{\gamma}}_1^T\mathbf{Z}_l^{(2)}(t))}-\bar{\mathbf{Z}}^{(2)}(\hat{\boldsymbol{\gamma}}_1,t)^{\otimes 2}\right\}I(\delta_i=1),\\
\hat{\boldsymbol{\Omega}}^{(1)}_{(pq)} &= \tfrac{1}{n}\textstyle\sum_{i=1}^n\left\{\int_0^\infty\left(\mathbf{Z}_i^{(1)}(t)-\bar{\mathbf{Z}}^{(1)}(\hat{\boldsymbol{\beta}}_1,t)\right)\left(dN_{i1}(t)-Y_i(t)\exp(\hat{\boldsymbol{\beta}}_1^T\mathbf{Z}_i^{(1)}(t)d\hat{\Lambda}_{10}(t)\right)*\hat{\boldsymbol{\eta}}_i\right\}\\
&+ \tfrac{1}{n}\textstyle\sum_{i=1}^n\left\{\int_0^\infty\left(\mathbf{Z}_i^{(1)}(t)-\bar{\mathbf{Z}}^{(1)}(\hat{\boldsymbol{\beta}}_1,t)\right)\left(dN_{i1}(t)-Y_i(t)\exp(\hat{\boldsymbol{\beta}}_1^T\mathbf{Z}_i^{(1)}(t))d\hat{\Lambda}_{10}(t)\right)*\hat{\boldsymbol{\phi}}_i\right\}\\
\hat{\boldsymbol{\Omega}}^{*(1)}_{(qq)} &= \tfrac{1}{n}\textstyle\sum_{i=1}^n(\hat{\boldsymbol{\eta}}_i+\hat{\boldsymbol{\phi}}_i)^{\otimes 2},
\end{aligned}$$

with

$$\begin{aligned}
\hat{\boldsymbol{\eta}}_i &= \textstyle\int_0^\infty\left\{\mathbf{Z}_i^{(2)}(t)-\bar{\mathbf{Z}}^{(2)}(\hat{\boldsymbol{\gamma}}_1,t)\right\}\omega_i(t)d\hat{\tilde{M}}_{i1}(t),\\
\hat{\tilde{M}}_{i1}(t) &= \tilde{N}_{i1}(t)-\textstyle\int_0^t\tilde{Y}_i(u)\exp(\hat{\boldsymbol{\gamma}}_1^T\mathbf{Z}_i^{(2)}(u))d\hat{\tilde{\Lambda}}_{10}(u),\\
\hat{\tilde{\Lambda}}_{10}(t) &= \tfrac{1}{n}\textstyle\sum_{i=1}^n\int_0^t\left\{\sum_{l=1}^n\tilde{Y}_l(u)\exp(\hat{\boldsymbol{\gamma}}_1^T\mathbf{Z}_l^{(2)}(u))\right\}^{-1}\omega_i(u)d\tilde{N}_{i1}(u),\\
\hat{\boldsymbol{\phi}}_i &= \textstyle\int_0^\infty\frac{\hat{\mathbf{q}}(t)}{\hat{\pi}(t)}d\hat{M}_i^c(t),\\
\hat{M}_i^c(t) &= I(X_i\le t,\delta_i=0)-\textstyle\int_0^t I(X_i\ge u)d\hat{\Lambda}^c(u),\\
\hat{\Lambda}^c(t) &= \textstyle\int_0^t\frac{\sum_{i=1}^n d\{I(X_i\le u,\delta_i=0)\}}{\sum_{i=1}^n I(X_i\ge u)},\\
\hat{\mathbf{q}}(t) &= -n^{-1}\textstyle\sum_{i=1}^n\int_0^\infty\left\{\mathbf{Z}_i^{(2)}(s)-\bar{\mathbf{Z}}^{(2)}(\hat{\boldsymbol{\gamma}}_1,s)\right\}I(s\ge t>X_i)\omega_i(s)d\hat{\tilde{M}}_{i1}(s),\\
\hat{\pi}(t) &= n^{-1}\textstyle\sum_{i=1}^n I(X_i\ge t).
\end{aligned}$$

Proof for Theorem 3. First, we will derive the asymptotic joint distribution

of $n^{-1/2}(\mathbf{U}_1(\boldsymbol{\beta}_1), \tilde{\mathbf{U}}_1(\boldsymbol{\gamma}_1))^T$. It can be shown that

$$\begin{aligned} n^{-1/2}\mathbf{U}_1(\boldsymbol{\beta}_1) &= n^{-1/2}\textstyle\sum_{i=1}^n \mathbf{U}_{i1}(\boldsymbol{\beta}_1) + o_p(1), \\ n^{-1/2}\tilde{\mathbf{U}}_1(\boldsymbol{\gamma}_1) &= n^{-1/2}\textstyle\sum_{i=1}^n (\boldsymbol{\eta}_i(\boldsymbol{\gamma}_1) + \boldsymbol{\phi}_i(\boldsymbol{\gamma}_1)) + o_p(1), \end{aligned} \tag{3.24}$$

where

$$\mathbf{U}_{i1}(\boldsymbol{\beta}_1) = \int_0^\infty \left\{ \mathbf{Z}_i^{(1)}(t) - \bar{\mathbf{z}}^{(1)}(\boldsymbol{\beta}_1, t) \right\} dM_{i1}(t),$$

$$\boldsymbol{\eta}_i = \int_0^\infty \left\{ \mathbf{Z}_i^{(2)}(t) - \bar{\mathbf{z}}^{(2)}(\boldsymbol{\gamma}_1, t) \right\} w_i(t) d\tilde{M}_{i1}(t),$$

$$\boldsymbol{\phi}_i = \int_0^\infty \frac{\mathbf{q}(t)}{\pi(t)} dM_i^c(t),$$

$$\tilde{M}_{i1}(t) = \tilde{N}_{i1}(t) - \int_0^t \tilde{Y}_i(t) \exp(\boldsymbol{\gamma}_1^T \mathbf{Z}_i^{(2)}(u)) d\tilde{\Lambda}_{10}(u),$$

$$M_i^c(t) = I(X_i \le t, \delta_i = 0) - \int_0^t I(X_i \ge u) d\Lambda^c(u),$$

$$\mathbf{q}(t) = -n^{-1} \sum_{i=1}^n \int_0^\infty \left\{ \mathbf{Z}_i^{(2)}(u) - \bar{\mathbf{Z}}^{(2)}(\boldsymbol{\gamma}_1, u) \right\} w_i(u) d\tilde{M}_{i1}(u) I(u \ge t > X_i),$$

and

$$\pi(t) = n^{-1} \sum_{i=1}^n I(X_i \ge t),$$

with $\tilde{\Lambda}_{10}(t) = \int_0^t \tilde{\lambda}_{10}(u)du$ is the baseline cause-specific cumulative hazard for cause 1, $\Lambda^c(t) = \int_0^t \lambda^c(u)du$ is the cumulative hazard for censoring variable,

$$\bar{\mathbf{z}}^{(1)}(\boldsymbol{\beta}_1, t) = \frac{\lim_{n\to\infty} n^{-1} \sum_{l=1}^n Y_l(t) \mathbf{Z}_l^{(1)}(t) \exp(\boldsymbol{\beta}_1^T \mathbf{Z}_l^{(1)}(t))}{\lim_{n\to\infty} n^{-1} \sum_{l=1}^n Y_l(t) \exp(\boldsymbol{\beta}_1^T \mathbf{Z}_l^{(1)}(t))},$$

and

$$\bar{\mathbf{z}}^{(2)}(\boldsymbol{\gamma}_1, t) = \frac{\lim_{n\to\infty} \sum_{l=1}^n \omega_l(t) \tilde{Y}_l(t) \mathbf{Z}_l^{(2)}(t) \exp(\boldsymbol{\gamma}_1^T \mathbf{Z}_l^{(2)}(t))}{\lim_{n\to\infty} \sum_{l=1}^n \omega_l(t) \tilde{Y}_l(t) \exp(\boldsymbol{\gamma}_1^T \mathbf{Z}_l^{(2)}(t))}.$$

It follows from (3.24) and the multivariate central limit theorem that $n^{-1/2}(\mathbf{U}_1(\boldsymbol{\beta}_1), \tilde{\mathbf{U}}_1(\boldsymbol{\gamma}_1))^T$ has a $p+q$ multivariate normal limiting distribution with mean $\mathbf{0}$ and variance-covariance

$$\boldsymbol{\Omega}^{(1)} = \begin{pmatrix} \boldsymbol{\Omega}_{(pp)}^{(1)} & \boldsymbol{\Omega}_{(pq)}^{(1)} \\ \boldsymbol{\Omega}_{(qp)}^{(1)} & \boldsymbol{\Omega}_{(qq)}^{(1)} \end{pmatrix}, \tag{3.25}$$

where

$$\boldsymbol{\Omega}^{(1)}_{(pp)} = \int_0^\infty \left[\frac{\lim_{n\to\infty} n^{-1}\sum_l Y_l(t)\mathbf{Z}_l^{(1)}(t)^{\otimes 2}\exp(\boldsymbol{\beta}_1^T\mathbf{Z}_l^{(1)}(t))}{\lim_{n\to\infty} n^{-1}\sum_l Y_l(t)\exp(\boldsymbol{\beta}_1^T\mathbf{Z}_l^{(1)}(t))} - \bar{\mathbf{z}}^{(1)}(\boldsymbol{\beta}_1,t)^{\otimes 2}\right] Y_i(t)\exp(\boldsymbol{\beta}_1^T\mathbf{Z}_i^{(1)}(t))d\Lambda_{10}(t), \tag{3.26}$$

$$\boldsymbol{\Omega}^{(1)}_{(qq)} = \int_0^\infty \left\{\frac{\lim_{n\to\infty}\frac{1}{n}\sum_{l=1}^n \omega_l(t)\tilde{Y}_l(t)\mathbf{Z}_l^{(2)}(t)^{\otimes 2}\exp(\boldsymbol{\gamma}_1^T\mathbf{Z}_l^{(2)}(t))}{\lim_{n\to\infty}\frac{1}{n}\sum_{l=1}^n \omega_l(t)\tilde{Y}_l(t)\exp(\boldsymbol{\gamma}_1^T\mathbf{Z}_l^{(2)}(t))} - \bar{\mathbf{z}}^{(2)}(\boldsymbol{\gamma}_1,t)^{\otimes 2}\right\} \omega_i(t)\tilde{Y}_i(t)\exp(\boldsymbol{\gamma}_1^T\mathbf{Z}_i^{(2)}(t))d\tilde{\Lambda}_{10}(t), \tag{3.27}$$

and

$$\boldsymbol{\Omega}^{*(1)}_{qq} = E\left\{\left(\boldsymbol{\eta}_i(\boldsymbol{\gamma}_1)+\boldsymbol{\phi}_i(\boldsymbol{\gamma}_1)\right)\left(\boldsymbol{\eta}_i(\boldsymbol{\gamma}_1)+\boldsymbol{\phi}_i(\boldsymbol{\gamma}_1)\right)^T\right\}. \tag{3.28}$$

Note that variance-covariance matrix between the two score test statistics is obtained as the limit of

$$\begin{aligned}
&\left\langle n^{-1/2}\mathbf{U}_1(\boldsymbol{\beta}_1), n^{-1/2}\tilde{\mathbf{U}}_1(\boldsymbol{\gamma}_1)\right\rangle \\
=\ & n^{-1}\textstyle\sum_{i=1}^n \langle \mathbf{U}_{i1}(\boldsymbol{\beta}_1), \boldsymbol{\eta}_i(\boldsymbol{\gamma}_1)+\boldsymbol{\phi}_i(\boldsymbol{\gamma}_1)\rangle \\
=\ & n^{-1}\textstyle\sum_{i=1}^n \langle \mathbf{U}_{i1}(\boldsymbol{\beta}_1), \boldsymbol{\eta}_i(\boldsymbol{\gamma}_1)\rangle + n^{-1}\sum_{i=1}^n \langle \mathbf{U}_{i1}(\boldsymbol{\beta}_1), \boldsymbol{\phi}_i(\boldsymbol{\gamma}_1)\rangle \\
=\ & n^{-1}\textstyle\sum_{i=1}^n \int_0^\infty \left\{\mathbf{Z}_i^{(1)}(t)-\bar{\mathbf{z}}^{(1)}(\boldsymbol{\beta}_1,t)\right\}\left\{\mathbf{Z}_i^{(2)}(t)-\bar{\mathbf{z}}^{(2)}(\boldsymbol{\gamma}_1,t)\right\}\omega_i(t)d<M_{i1},\tilde{M}_{i1}>(t) \\
& +n^{-1}\textstyle\sum_{i=1}^n \int_0^\infty \left\{\mathbf{Z}_i^{(1)}-\bar{\mathbf{z}}^{(1)}(\boldsymbol{\beta}_1,t)\right\}\frac{\mathbf{q}(t)}{\pi(t)}d<M_{i1},M_i^c>(t),
\end{aligned}$$

which converges in probability to

$$\begin{aligned}
\boldsymbol{\Omega}^{(1)}_{(pq)} &= E\textstyle\int_0^\infty \left\{\mathbf{Z}_i^{(1)}(t)-\bar{\mathbf{z}}^{(1)}(\boldsymbol{\beta}_1,t)\right\}\left\{\mathbf{Z}_i^{(2)}(t)-\bar{\mathbf{z}}^{(2)}(\boldsymbol{\gamma}_1,t)\right\}\omega_i(t)d<M_{i1},\tilde{M}_{i1}>(t) \\
&+ E\textstyle\int_0^\infty \left\{\mathbf{Z}_i^{(1)}-\bar{\mathbf{z}}^{(1)}(\boldsymbol{\beta}_1,t)\right\}\frac{\mathbf{q}(t)}{\pi(t)}d<M_{i1},M_i^c>(t).
\end{aligned} \tag{3.29}$$

Let $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\gamma}}_1$ be solutions to $\mathbf{U}_1(\hat{\boldsymbol{\beta}}_1)=0$ and $\tilde{\mathbf{U}}_1(\hat{\boldsymbol{\gamma}}_1)=0$, respectively. Appying Taylor series expansion to $(\mathbf{U}_1(\hat{\boldsymbol{\beta}}_1),\tilde{\mathbf{U}}_1(\hat{\boldsymbol{\gamma}}_1))^T$ around $(\boldsymbol{\beta}_1,\boldsymbol{\gamma}_1)$, we have

$$n^{1/2}\begin{pmatrix}\hat{\boldsymbol{\beta}}_1-\boldsymbol{\beta}_1\\ \hat{\boldsymbol{\gamma}}_1-\boldsymbol{\gamma}_1\end{pmatrix} = \begin{pmatrix}\boldsymbol{\Omega}^{(1)-1}_{(pp)} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Omega}^{(1)-1}_{(qq)}\end{pmatrix}\begin{pmatrix}\mathbf{U}_1(\boldsymbol{\beta}_1)\\ \tilde{\mathbf{U}}_1(\boldsymbol{\gamma}_1)\end{pmatrix}+o_p(1).$$

This, together with (3.25), implies that

$$\boldsymbol{\Sigma}^{(1)} = \begin{pmatrix}\boldsymbol{\Omega}^{(1)-1}_{(pp)} & \boldsymbol{\Omega}^{(1)-1}_{(pp)}\boldsymbol{\Omega}^{(1)}_{(pq)}\boldsymbol{\Omega}^{(1)-1}_{(qq)}\\ \boldsymbol{\Omega}^{(1)-1}_{(qq)}\boldsymbol{\Omega}^{(1)}_{(qp)}\boldsymbol{\Omega}^{(1)-1}_{(pp)} & \boldsymbol{\Omega}^{(1)-1}_{(qq)}\end{pmatrix}. \tag{3.30}$$

A consistent estimator for $\boldsymbol{\Sigma}^{(1)}$ is obtained by replacing all unknown quantities with their respective sample estimates. $\square$

Corollary 1 *Let* $\boldsymbol{\xi}_n = n^{1/2}(\mathbf{A}_1\hat{\boldsymbol{\beta}}_1 - \mathbf{d}_1)$ *and* $\boldsymbol{\eta}_n = n^{1/2}(\mathbf{A}_2\hat{\boldsymbol{\gamma}}_1 - \mathbf{d}_2)$. *Then, under the null hypothesis (3.20), we have*

$$\begin{pmatrix} \boldsymbol{\xi}_n \\ \boldsymbol{\eta}_n \end{pmatrix} \xrightarrow{d} N(\mathbf{0}, \mathbf{V}), \quad \text{as } n \to \infty,$$

where

$$\mathbf{V} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 \end{pmatrix} \boldsymbol{\Sigma}^{(1)} \begin{pmatrix} \mathbf{A}_1^T & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2^T \end{pmatrix}. \tag{3.31}$$

Define the following Wald-type test statistic

$$X_W^2 = \left(\boldsymbol{\xi}_n^T, \boldsymbol{\eta}_n^T\right) \hat{\mathbf{V}}^{-1} \begin{pmatrix} \boldsymbol{\xi}_n \\ \boldsymbol{\eta}_n \end{pmatrix},$$

where $\hat{\mathbf{V}}$ is a consistent estimate of $\mathbf{V}$ obtained by replacing $\boldsymbol{\Sigma}^{(1)}$ with $\hat{\boldsymbol{\Sigma}}^{(1)}$ in (3.31). It follows immediately from Corollary 1 that under (3.20), X_W^2 has an asymptotic chi-squared distribution with $p_{d1} + p_{d2}$ degrees of freedom, where p_{d1} and p_{d2} are the dimensions of $\mathbf{d}_1$ and $\mathbf{d}_2$, respectively. This leads to the following chi-square joint test for (3.20):

$$\text{Reject (3.20) at level } \alpha \text{ if } X_W^2 > \chi^2_{p_{d1}+p_{d2}}(\alpha),$$

where $\chi^2_{p_{d1}+p_{d2}}(\alpha)$ is the upper $1-\alpha$ percentile of the standard $\chi^2_{p_{d1}+p_{d2}}$ distribution.

3.2.2 Joint Regression Analysis for Other Quantities

Besides analyzing $\lambda_1(t|\mathbf{Z})$ and $\tilde{\lambda}_1(t|\mathbf{Z})$ jointly, it is sometimes also useful to consider other related quantities. Let $\lambda(t|\mathbf{Z}))$ denote the all-cause hazard function.

Then, the three pairs $(\lambda_1(t|\mathbf{Z}), \tilde{\lambda}_1(t|\mathbf{Z}))$, $(\lambda_1(t|\mathbf{Z}), \lambda(t|\mathbf{Z}))$, and $(\lambda_1(t|\mathbf{Z}), \lambda_2(t|\mathbf{Z}))$ uniquely determine each other.

3.2.2.1 Joint Regression Analysis of Cause-Specific Hazard and All-Cause Hazard

Assume that the proportional cause-specific hazards model (3.18) holds. In addition, assume the proportional all-cause hazards model:

$$\lambda(t|\mathbf{Z}) = \lambda_0(t) \exp(\boldsymbol{\beta}_\cdot^T \mathbf{Z}^{(3)}(t)), \tag{3.32}$$

where $\lambda_0(t)$ is an unknown baseline all-cause hazard, and $\mathbf{Z}^{(3)}(t)$ are functions of the original covariates $\mathbf{Z}$ and t that allow time $\times$ covariates interactions. Below we derive joint inference for $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_\cdot$.

Let

$$\mathbf{U}_\cdot(\boldsymbol{\beta}_\cdot) = \sum_{i=1}^{n} \int_0^\infty \left\{ \mathbf{Z}_i^{(3)}(t) - \bar{\mathbf{Z}}^{(3)}(\boldsymbol{\beta}_\cdot, t) \right\} dN_i(t), \tag{3.33}$$

be the score function for $\boldsymbol{\beta}_\cdot$ under model (3.32), where

$$\bar{\mathbf{Z}}^{(3)}(\boldsymbol{\beta}_\cdot, t) = \frac{\sum_{l=1}^n Y_l(t) \mathbf{Z}_l^{(3)}(t) \exp(\boldsymbol{\beta}_\cdot^T \mathbf{Z}_l^{(3)}(t))}{\sum_{l=1}^n Y_l(t) \exp(\boldsymbol{\beta}_\cdot^T \mathbf{Z}_l^{(3)}(t))}$$

and $N_i(t) = I(X_i \leq t, \delta_i = 1)$. Let $\hat{\boldsymbol{\beta}}_\cdot$ be the solution of the score equation $\mathbf{U}_\cdot(\boldsymbol{\beta}_\cdot) = 0$.

Theorem 4 *Under some regularity conditions, as $n \to \infty$,*

$$n^{1/2} \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \\ \hat{\boldsymbol{\beta}}_\cdot - \boldsymbol{\beta}_\cdot \end{pmatrix} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}^{(2)}),$$

where $\boldsymbol{\Sigma}^{(2)}$ is defined by (3.37). Furthermore, $\boldsymbol{\Sigma}^{(2)}$ can be consistently estimated by

$$\hat{\boldsymbol{\Sigma}}^{(2)} = \begin{pmatrix} \hat{\boldsymbol{\Omega}}_{(pp)}^{(2)-1} & \hat{\boldsymbol{\Omega}}_{(pp)}^{(2)-1} \hat{\boldsymbol{\Omega}}_{(pq)}^{(2)} \hat{\boldsymbol{\Omega}}_{(qq)}^{(2)-1} \\ \hat{\boldsymbol{\Omega}}_{(qq)}^{(2)-1} \hat{\boldsymbol{\Omega}}_{(qp)}^{(2)} \hat{\boldsymbol{\Omega}}_{(pp)}^{(2)-1} & \hat{\boldsymbol{\Omega}}_{(qq)}^{(2)-1} \end{pmatrix}, \tag{3.34}$$

where

$$\begin{aligned}
\hat{\mathbf{\Omega}}^{(2)}_{(pp)} &= \tfrac{1}{n}\sum_{i=1}^{n}\int_0^{\infty}\left[\frac{\sum_{l=1}^{n}Y_l(t)\mathbf{Z}_l^{(1)}(t)^{\otimes 2}\exp(\hat{\boldsymbol{\beta}}_1^T\mathbf{Z}_l^{(1)}(t))}{\sum_{l=1}^{n}Y_l(t)\exp(\hat{\boldsymbol{\beta}}_1^T\mathbf{Z}_l^{(1)}(t))}-\bar{\mathbf{Z}}^{(1)}(\hat{\boldsymbol{\beta}}_1,t)^{\otimes 2}\right]dN_{i1}(t),\\
\hat{\mathbf{\Omega}}^{(2)}_{pq} &= \tfrac{1}{n}\sum_{i=1}^{n}\int_0^{\infty}\left(\mathbf{Z}_i^{(3)}(t)-\bar{\mathbf{Z}}^{(3)}(\hat{\boldsymbol{\beta}}_\cdot,t)\right)\left(\mathbf{Z}_i^{(1)}(t)-\bar{\mathbf{Z}}^{(1)}(\hat{\boldsymbol{\beta}}_1,t)\right)\\
&\quad Y_i(t)\exp(\hat{\boldsymbol{\beta}}_1^T\mathbf{Z}^{(1)}(t))d\hat{\Lambda}_{10}(t),\\
\hat{\mathbf{\Omega}}^{(2)}_{(qq)} &= \tfrac{1}{n}\sum_{i=1}^{n}\int_0^{\infty}\left[\frac{\sum_{l=1}^{n}Y_l(t)\mathbf{Z}_l^{(3)}(t)^{\otimes 2}\exp(\hat{\boldsymbol{\beta}}_\cdot^T\mathbf{Z}_l^{(3)}(t))}{\sum_{l=1}^{n}Y_l(t)\exp(\hat{\boldsymbol{\beta}}_\cdot^T\mathbf{Z}_l^{(3)}(t))}-\bar{\mathbf{Z}}^{(3)}(\hat{\boldsymbol{\beta}}_\cdot,t)^{\otimes 2}\right]dN_i(t),
\end{aligned}$$

with $\hat{\Lambda}_{10}(t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t\left\{\sum_{l=1}^{n}Y_l(u)(\hat{\boldsymbol{\beta}}_1^T\mathbf{Z}_i^{(1)}(u))\right\}^{-1}dN_{i1}(u)$ *is an estimator of the baseline cumulative cause-specific hazard for type 1 failure.*

Proof for Theorem 4. Under the null hypothesis, it was shown by (Fleming and Harrington, 1991) that

$$\begin{aligned}
\mathbf{U}_\cdot(\boldsymbol{\beta}_\cdot) &= \sum_{i=1}^{n}\int_0^{\tau}\left(\mathbf{Z}_i^{(3)}(t)-\bar{\mathbf{z}}^{(3)}(\boldsymbol{\beta}_\cdot,t)\right)dM_i(t)-n\int_0^{\tau}\left(\bar{\mathbf{Z}}^{(3)}(\boldsymbol{\beta}_\cdot,t)-\bar{\mathbf{z}}^{(3)}(\boldsymbol{\beta}_\cdot,t)\right)dM_i(t)\\
\mathbf{U}_1(\boldsymbol{\beta}_1) &= \sum_{i=1}^{n}\int_0^{\tau}\left(\mathbf{Z}_i^{(1)}(t)-\bar{\mathbf{z}}^{(1)}(\boldsymbol{\beta}_1,t)\right)dM_{i1}(t)-n\int_0^{\tau}\left(\bar{\mathbf{Z}}^{(1)}(\boldsymbol{\beta}_1,t)-\bar{\mathbf{z}}^{(1)}(\boldsymbol{\beta}_1,t)\right)dM_{i1}(t),
\end{aligned}$$

where

$$M_i(t) = N_{i\cdot}(t)-\int_0^t\lambda_0(u)Y_i(u)\exp(\boldsymbol{\beta}_\cdot^T\mathbf{Z}_i^{(3)}(u))du,$$

$$M_{i1}(t) = N_{i1}(t)-\int_0^t\lambda_{j0}(u)Y_i(u)\exp(\boldsymbol{\beta}_1^T\mathbf{Z}_i^{(1)}(u))du,$$

and $\bar{\mathbf{z}}^{(3)}(\boldsymbol{\beta}_\cdot,t) = \frac{\lim_{n\to\infty}n^{-1}\sum_{l=1}^{n}Y_l(t)\mathbf{Z}_l^{(3)}(t)\exp(\boldsymbol{\beta}_\cdot^T\mathbf{Z}_l^{(3)}(t))}{\lim_{n\to\infty}n^{-1}\sum_{l=1}^{n}Y_l(t)\exp(\boldsymbol{\beta}_\cdot^T\mathbf{Z}_l^{(3)}(t))}$.

The first part of $(\mathbf{U}_\cdot(\boldsymbol{\beta}_\cdot),\mathbf{U}_1(\boldsymbol{\beta}_1))^T, i = 1,2,\ldots,n$ can be viewed as a sum of independently identically distributed random vector. By using multivariate central limit theory, we can prove the first part of the vector (3) has a bivariate normal distribution with mean $\mathbf{0}$, and variance-covariance matrix $\mathbf{\Omega}^{(2)}$. Since $\bar{\mathbf{Z}}^{(3)}(\boldsymbol{\beta}_\cdot,t)$ and $\bar{\mathbf{Z}}^{(1)}(\boldsymbol{\beta}_1,t)$ converge in probability to some deterministic process $\bar{\mathbf{z}}^{(3)}(\boldsymbol{\beta}_\cdot,t)$ and $\bar{\mathbf{z}}^{(1)}(\boldsymbol{\beta}_1,t)$ respectively, we can prove the second part of the vector (3) converge in probability to zero by using the central limit theory for stochastic integrals with respect to counting process martingales. Then we can use Slusky

theorem to prove $n^{-1/2}(\mathbf{U}_\cdot(\boldsymbol{\beta}_\cdot), \mathbf{U}_1(\boldsymbol{\beta}_1))^T$ has a $p+q$ dimension multivariate normal limiting distribution with mean $\mathbf{0}$ and variance-covariance matrix

$$\begin{pmatrix} \boldsymbol{\Omega}^{(2)}_{(pp)} & \boldsymbol{\Omega}^{(2)}_{(pq)} \\ \boldsymbol{\Omega}^{(2)}_{(qp)} & \boldsymbol{\Omega}^{(2)}_{(qq)} \end{pmatrix}, \tag{3.35}$$

where $\boldsymbol{\Omega}^{(2)}_{(pp)} = \boldsymbol{\Omega}^{(1)}_{(pp)}$, which is defined in (3.26),

$$\begin{aligned} \boldsymbol{\Omega}^{(2)}_{(qq)} = \int_0^\infty & \left[\frac{\lim_{n\to\infty} n^{-1} \sum_{l=1}^n Y_l(t) \mathbf{Z}_l^{(3)}(t)^{\otimes 2} \exp(\boldsymbol{\beta}_\cdot^T \mathbf{Z}_l^{(3)}(t))}{\lim_{n\to\infty} n^{-1} \sum_{l=1}^n Y_l(t) \exp(\boldsymbol{\beta}_\cdot^T \mathbf{Z}_l^{(3)}(t))} - \bar{\mathbf{z}}^{(3)}(\boldsymbol{\beta}_\cdot, t)^{\otimes 2} \right] \\ & Y_i(t) \exp(\boldsymbol{\beta}_\cdot^T \mathbf{Z}_i^{(3)}(t)) d\Lambda_0(t). \end{aligned} \tag{3.36}$$

and the covariance

$$\boldsymbol{\Omega}^{(2)}_{(pq)} = E \int_0^\tau \left(\mathbf{Z}_i^{(3)}(t) - \bar{\mathbf{z}}^{(3)}(\boldsymbol{\beta}_\cdot, t) \right) \left(\mathbf{Z}_i^{(1)}(t) - \bar{\mathbf{z}}^{(1)}(\boldsymbol{\beta}_1, t) \right) \exp(\boldsymbol{\beta}_1^T \mathbf{Z}_i^{(1)}(t)) Y_i(t) d\Lambda_{10}(t)$$

is the limit of

$$\begin{aligned} & \left\langle n^{-1/2} \mathbf{U}_\cdot(\boldsymbol{\beta}_\cdot), n^{-1/2} \mathbf{U}_1(\boldsymbol{\beta}_1) \right\rangle \\ = \; & n^{-1} \sum_{i=1}^n \left\langle \int_0^\tau \left(\mathbf{Z}_i^{(3)}(t) - \bar{\mathbf{z}}^{(3)}(\boldsymbol{\beta}_\cdot, t) \right) dM_i(t), \int_0^\tau \left(\mathbf{Z}_i^{(1)}(t) - \bar{\mathbf{z}}^{(1)}(\boldsymbol{\beta}_1, t) \right) dM_{i1}(t) \right\rangle \\ = \; & n^{-1} \sum_{i=1}^n \int_0^\tau \left(\mathbf{Z}_i^{(3)}(t) - \bar{\mathbf{z}}^{(3)}(\boldsymbol{\beta}_\cdot, t) \right) \left(\mathbf{Z}_i^{(1)}(t) - \bar{\mathbf{z}}^{(1)}(\boldsymbol{\beta}_1, t) \right) d < M_i, M_{i1} > (t) \\ = \; & n^{-1} \sum_{i=1}^n \int_0^\tau \left(\mathbf{Z}_i^{(3)}(t) - \bar{\mathbf{z}}^{(3)}(\boldsymbol{\beta}_\cdot, t) \right) \left(\mathbf{Z}_i^{(1)}(t) - \bar{\mathbf{z}}^{(1)}(\boldsymbol{\beta}_1, t) \right) d < M_{i1+i2}, M_{i1} > (t) \\ = \; & n^{-1} \sum_{i=1}^n \int_0^\tau \left(\mathbf{Z}_i^{(3)}(t) - \bar{\mathbf{z}}^{(3)}(\boldsymbol{\beta}_\cdot, t) \right) \left(\mathbf{Z}_i^{(1)}(t) - \bar{\mathbf{z}}^{(1)}(\boldsymbol{\beta}_1, t) \right) d < M_{i1}, M_{i1} > (t) \\ = \; & n^{-1} \sum_{i=1}^n \int_0^\tau \left(\mathbf{Z}_i^{(3)}(t) - \bar{\mathbf{z}}^{(3)}(\boldsymbol{\beta}_\cdot, t) \right) \left(\mathbf{Z}_i^{(1)}(t) - \bar{\mathbf{z}}^{(1)}(\boldsymbol{\beta}_1, t) \right) \exp(\boldsymbol{\beta}_1^T \mathbf{Z}_i^{(1)}(t)) Y_i(t) d\Lambda_{10}(t). \end{aligned}$$

Applying the Taylor series expansion to $(\mathbf{U}_\cdot(\hat{\boldsymbol{\beta}}_\cdot), \mathbf{U}_1(\hat{\boldsymbol{\beta}}_1))^T$ around $(\boldsymbol{\beta}_\cdot, \boldsymbol{\beta}_1)$, we have

$$n^{1/2} \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \\ \hat{\boldsymbol{\beta}}_\cdot - \boldsymbol{\beta}_\cdot \end{pmatrix} \approx \begin{pmatrix} \boldsymbol{\Omega}^{(2)-1}_{(pp)} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}^{(2)-1}_{(qq)} \end{pmatrix} \begin{pmatrix} \mathbf{U}_\cdot(\boldsymbol{\beta}_\cdot) \\ \mathbf{U}_1(\boldsymbol{\beta}_1) \end{pmatrix}.$$

This, together with (3.35), implies that

$$\boldsymbol{\Sigma}^{(2)} = \begin{pmatrix} \boldsymbol{\Omega}^{(2)-1}_{(pp)} & \boldsymbol{\Omega}^{(2)-1}_{(pp)} \boldsymbol{\Omega}^{(2)}_{(pq)} \boldsymbol{\Omega}^{(2)-1}_{(qq)} \\ \boldsymbol{\Omega}^{(2)-1}_{(qq)} \boldsymbol{\Omega}^{(2)}_{qp} \boldsymbol{\Omega}^{(2)-1}_{(pp)} & \boldsymbol{\Omega}^{(2)-1}_{(qq)} \end{pmatrix}. \tag{3.37}$$

Finally, a consistent estimator for $\mathbf{\Sigma}^{(2)}$ is obtained by replacing all unknown quantities with their respective sample estimates in (3.37). $\square$

Theorem 4 enables one to draw joint inference for $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_\cdot$ along the lines of the previous section.

3.2.2.2 Joint Regression Analysis of Cause-Specific Hazard and Other-Cause Hazard

Assume the proportional cause-specific hazards model (3.18) for type 1 failure. In addition, assume the following proportional cause-specific hazards model for type 2 failure:

$$\lambda_2(t|\mathbf{Z}) = \lambda_{20}(t)\exp(\boldsymbol{\beta}_2^T\mathbf{Z}^{(4)}(t)), \tag{3.38}$$

where $\lambda_{20}(t)$ is a unknown baseline cause-specific hazard, and $\mathbf{Z}^{(4)}(t)$ are functions of the original covariates $\mathbf{Z}$ and t that allow time $\times$ covariates interactions.

Let

$$\mathbf{U}_2(\boldsymbol{\beta}_2) = \sum_{i=1}^{n}\int_0^\infty \left\{\mathbf{Z}_i^{(4)}(t) - \bar{\mathbf{Z}}^{(4)}(\boldsymbol{\beta}_2, t)\right\} dN_{i2}(t), \tag{3.39}$$

be the score test statistic under model (3.38), where

$$\bar{\mathbf{Z}}^{(4)}(\boldsymbol{\beta}_2, t) = \frac{\sum_{l=1}^n Y_l(t)\mathbf{Z}_l^{(4)}(t)\exp(\boldsymbol{\beta}_2^T\mathbf{Z}_l^{(4)}(t))}{\sum_{l=1}^n Y_l(t)\exp(\boldsymbol{\beta}_2^T\mathbf{Z}_l^{(4)}(t))}.$$

Let $\hat{\boldsymbol{\beta}}_2$ be the solution of the score equations $\mathbf{U}_2(\boldsymbol{\beta}_2) = 0$.

Theorem 5 *Assume the usual regularity conditions, as* $n \to \infty$,

$$n^{1/2}\begin{pmatrix}\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \\ \hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2\end{pmatrix} \xrightarrow{d} N(0, \mathbf{\Sigma}^{(3)}),$$

where $\mathbf{\Sigma}^{(3)}$ *is defined in the Appendix. Furthermore,* $\mathbf{\Sigma}^{(3)}$ *can be consistently*

estimated by

$$\hat{\boldsymbol{\Sigma}}^{(3)} = \begin{pmatrix} \hat{\boldsymbol{\Omega}}^{(3)-1}_{(pp)} & 0 \\ 0 & \hat{\boldsymbol{\Omega}}^{(3)-1}_{(qq)} \end{pmatrix}, \tag{3.40}$$

where

$$\begin{aligned} \hat{\boldsymbol{\Omega}}^{(3)}_{(pp)} &= \tfrac{1}{n}\textstyle\sum_{i=1}^{n}\int_0^{\infty}\left[\frac{\sum_{l=1}^{n} Y_l(t)\mathbf{Z}_l^{(1)}(t)^{\otimes 2}\exp(\hat{\boldsymbol{\beta}}_1^T\mathbf{Z}_l^{(1)}(t))}{\sum_{l=1}^{n} Y_l(t)\exp(\hat{\boldsymbol{\beta}}_1^T\mathbf{Z}_l^{(1)}(t))} - \bar{\mathbf{Z}}^{(1)}(\hat{\boldsymbol{\beta}}_1,t)^{\otimes 2}\right] dN_{i1}(t), \\ \hat{\boldsymbol{\Omega}}^{(3)}_{(qq)} &= \tfrac{1}{n}\textstyle\sum_{i=1}^{n}\int_0^{\infty}\left[\frac{\sum_{l=1}^{n} Y_l(t)\mathbf{Z}_l^{(4)}(t)^{\otimes 2}\exp(\hat{\boldsymbol{\beta}}_2^T\mathbf{Z}_l^{(4)}(t))}{\sum_{l=1}^{n} Y_l(t)\exp(\hat{\boldsymbol{\beta}}_2^T\mathbf{Z}_l^{(4)}(t))} - \bar{\mathbf{Z}}^{(4)}(\hat{\boldsymbol{\beta}}_2,t)^{\otimes 2}\right] dN_{i2}(t). \end{aligned}$$

Proof of theorem 5. Under null hypothesis, it can be shown by (Fleming and Harrington, 1991) that

$$\begin{aligned} \mathbf{U}_1(\boldsymbol{\beta}_1) &= \textstyle\sum_{i=1}^{n}\int_0^{\tau}\left(\mathbf{Z}_i^{(1)}(t) - \bar{\mathbf{z}}^{(1)}(\boldsymbol{\beta}_1,t)\right) dM_{i1}(t) - n\int_0^{\tau}\left(\bar{\mathbf{Z}}^{(1)}(\boldsymbol{\beta}_1,t) - \bar{\mathbf{z}}^{(1)}(\boldsymbol{\beta}_1,t)\right) dM_{i1}(t) \\ \mathbf{U}_2(\boldsymbol{\beta}_2) &= \textstyle\sum_{i=1}^{n}\int_0^{\tau}\left(\mathbf{Z}_i^{(4)}(t) - \bar{\mathbf{z}}^{(4)}(\boldsymbol{\beta}_2,t)\right) dM_{i2}(t) - n\int_0^{\tau}\left(\bar{\mathbf{Z}}^{(4)}(\boldsymbol{\beta}_2,t) - \bar{\mathbf{z}}^{(4)}(\boldsymbol{\beta}_2,t)\right) dM_{i2}(t), \end{aligned} \tag{3.41}$$

where $M_{i2}(t) = N_{i2}(t) - \int_0^t \lambda_{j0}(u)Y_i(u)\exp(\boldsymbol{\beta}_2^T\mathbf{Z}_i^{(4)}(u))du$, and

$$\bar{\mathbf{z}}^{(4)}(\boldsymbol{\beta}_2,t) = \frac{\lim_{n\to\infty} n^{-1}\sum_{l=1}^{n} Y_l(t)\mathbf{Z}_l^{(4)}(t)\exp(\boldsymbol{\beta}_2^T\mathbf{Z}_l^{(4)}(t))}{\lim_{n\to\infty} n^{-1}\sum_{l=1}^{n} Y_l(t)\exp(\boldsymbol{\beta}_2^T\mathbf{Z}_l^{(4)}(t))}.$$

The first part of $(\mathbf{U}_1(\boldsymbol{\beta}_1), \mathbf{U}_2(\boldsymbol{\beta}_2))^T, i = 1, 2, \ldots, n$ can be viewed as a sum of independently identically distributed random vector. By the multivariate central limit theorem, the first part of the vector (3.41) has a bivariate normal distribution with mean $\mathbf{0}$, and variance-covariance matrix $\boldsymbol{\Omega}^{(3)}$. Since $\bar{\mathbf{Z}}^{(1)}(\boldsymbol{\beta}_1,t)$ and $\bar{\mathbf{Z}}^{(4)}(\boldsymbol{\beta}_2,t)$ converge in probability to some deterministic process $\bar{\mathbf{z}}^{(1)}(\boldsymbol{\beta}_1,t)$ and $\bar{\mathbf{z}}^{(4)}(\boldsymbol{\beta}_2,t)$, respectively, it can be shown that the second part of the vector (3.41) converges in probability to zero. It then follows from the Slusky theorem that $n^{-1/2}(\mathbf{U}_1(\boldsymbol{\beta}_1), \mathbf{U}_2(\boldsymbol{\beta}_2))^T$ has a $p+q$ dimension multivariate normal limiting distribution with mean $\mathbf{0}$ and variance-covariance matrix

$$\begin{pmatrix} \boldsymbol{\Omega}^{(3)}_{(pp)} & 0 \\ 0 & \boldsymbol{\Omega}^{(3)}_{(qq)} \end{pmatrix}, \tag{3.42}$$

where $\boldsymbol{\Omega}^{(3)}_{(pp)}$ is defined in (3.26) and

$$\boldsymbol{\Omega}^{(3)}_{(qq)} = \int_0^\infty \left[\frac{\lim_{n\to\infty} n^{-1} \sum_{l=1}^n Y_l(t) \mathbf{Z}_l^{(4)}(t)^{\otimes 2} \exp(\boldsymbol{\beta}_2^T \mathbf{Z}_l^{(4)}(t))}{\lim_{n\to\infty} n^{-1} \sum_{l=1}^n Y_l(t) \exp(\boldsymbol{\beta}_2^T \mathbf{Z}_l^{(4)}(t))} - \bar{\mathbf{z}}^{(4)}(\boldsymbol{\beta}_2, t)^{\otimes 2} \right] Y_i(t) \exp(\boldsymbol{\beta}_2^T \mathbf{Z}_i^{(4)}(t)) d\Lambda_{02}(t). \tag{3.43}$$

The variance-covariance matrix between $(\mathbf{U}_1(\boldsymbol{\beta}_1), \mathbf{U}_2(\boldsymbol{\beta}_2))^T$ equals $\mathbf{0}$ because $< M_{i1}, M_{i2} > (t) = 0$ (Fleming and Harrington, 1991, Theorem 2.5.2.). Finally, a consistent estimator for $\boldsymbol{\Sigma}^{(3)}$ is obtained by replacing all unknown quantities with their respective sample estimates in (3.42). □

Remark 3.4: In addition to being easy to interpret, the PH models for the cause-specific hazard and the all-cause hazard only require that the censoring time be conditionally independent of the survival time given the observed covariates, which is weaker than the completely censoring at random assumption needed by the proportional subdistribution hazards model.

Remark 3.5: It is important in practice to routinely assess how well a regression model fits the data. It has been well recognized that the proportional hazard assumption for a time-independent covariate does not hold simultaneously for the cause-specific hazard and cause-specific subdistribution hazard, which is why we allow time × covariates interactions in models (3.18), (3.19), (3.32) and (3.38). Standard model building and diagnostics techniques for the standard Cox (1972) proportional hazards model can be readily applied to models (3.18), (3.32) and (3.38) (Schoenfeld, 1980, 1982; Lagakos, 1981; Andersen, 1982; Nagelkerke et al., 1984; Moreau et al., 1985; Arjas, 1988; Beyersmann et al., 2007; Latouche et al., 2007; Grambauer et al., 2010; Haller et al., 2012; Andersen et al., 2012). Graphical methods for these models can also be adapted for the proportional subdistribution hazards model (3.19). Formal goodness of fit tests for (3.19) are developed by (Scheike and Zhang, 2008).

3.3 Simulations

In this section, we present some simulation studies to illustrate the advantages of the proposed joint tests over the Bonferroni adjustment method that spends half of the significance level for each of the two individual tests. Specifically, we consider a two-sample model with $J = 2$ competing risks, and we focus on the problem of comparing type 1 failure time between the two groups. The competing risks data are generated using Beyersmann et al. (2009)'s cause-specific hazard driven method which requires only specification of the cause-specific hazard for each type of failure. The underlying model assumes that in each group, both types of failures have constant cause-specific hazards and thus the all-cause hazard is also constant. As discussed in Section 3.2.2, to study the group difference in type 1 failure, it is sufficient to test any one of the three equivalent hypotheses (3.1), (3.12), and (3.13). In this chapter, we did simulation studies with respect to both CSH and CIF pair, and CSH and ACH pair.

3.3.1 Simulations for Joint Two sample test with respect to CSH and CIF Pair

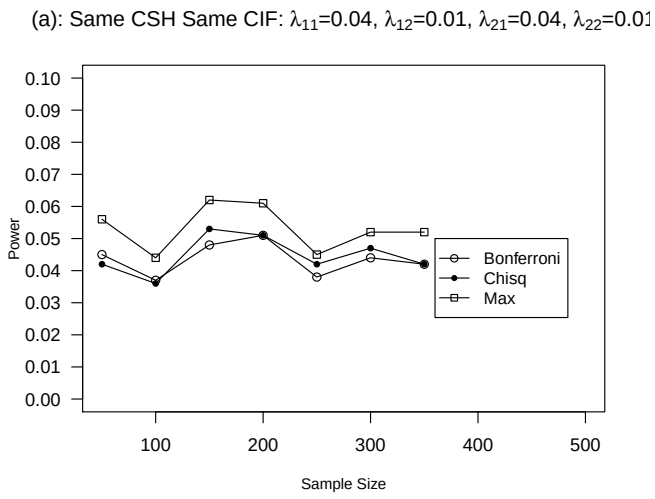
Figure 3.1 depicts the simulated rejection power of the chi-square joint test, the maximum joint test and the Bonferroni adjustment method for (3.1) for various sample sizes under three scenarios. Here λ_{jk} denotes the cause-specific hazard for type j in group k. The censoring rate is set to be 0.1 with an independent exponential censoring time. The nominal significance level is 0.05. Figure 3.1 represents a null case where there is no difference with respect to type 1 failure between the two groups (same CSH and CIF for cause 1 between two groups). Figure 3.1 (b) corresponds to a situation where there is no group difference in cause-specific hazard for type 1 failure, but there is a group difference in the

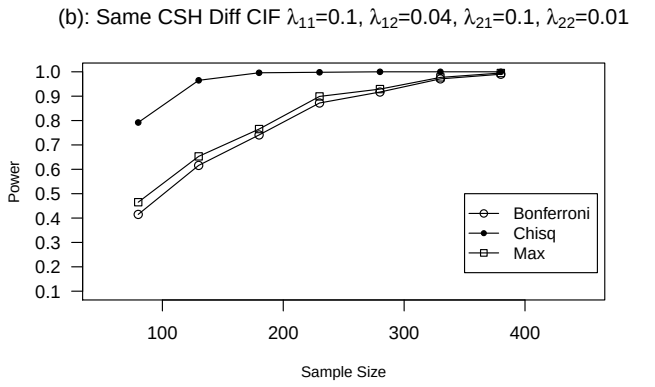
cumulative incidence function (or cumulative incidence hazard). Figure 3.1 (c) corresponds to a situation where there are group differences in both the cause-specific hazard and cumulative incidence for type 1 failure. Each simulation uses 1,000 Monte Carlo samples.

It is seen from Figure 3.1 (a) that the type I error rates for all three tests are well controlled around the 0.05 nominal level. In the case when there is a group difference in cumulative incidence hazard (Figure 3.1 (b)) or both the cause-specific hazard and cumulative incidence hazard for type 1 failure (Figure 3.1(c)), the chi-square joint test or the maximum join test can be much more powerful than the Bonferroni method. This has important implications for the design of a clinical trial in the presence of competing risks. For example, to achieve 80% power under the second scenario (Figure 3.1 (b)), it would require $n = 80$ patients in each group for the chi-square joint test, $n = 180$ patients for the joint maximum test, and at least $n = 200$ patients in each group for the Bonferroni adjustment method.

3.3.2 Simulations for Joint Two sample Test with respect to CSH and ACH Pair

Figure 3.2 depicts the simulated rejection power of the chi-square joint test, the maximum joint test, and the Bonferroni adjustment method for (3.12) for various sample sizes under four scenarios. Here λ_{1k} and λ_k denote the cause-specific hazard for type 1 failure and the all-cause hazard, respectively, for group k ($k = 1, 2$). The censoring rate is set to be 0.1 with an independent exponential censoring time. Patients were equally assigned to two groups. The nominal significance level is 0.05. Figure 3.2 (a) represents a null case where there is no difference with respect to type 1 failure between the two groups. Figure 3.2





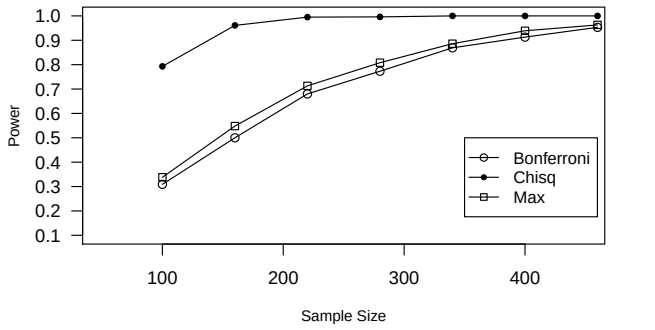


Figure 3.1: For CSH and CIF pair: simulated power for the chi-square test (solid line), the maximum test (dotted line), and the Bonferroni method (dashed line) under a null case (panel (a)) and two different alternatives (panels (b)-(c))

(b) corresponds to a situation where there is a group difference in cause-specific hazard for type 1 failure, but there is no group difference in the all-cause hazard. Figure 3.2 (c) corresponds to a situation where there is no group difference in cause-specific hazard for type 1 failure, but there is a group difference in the all-cause hazard. Both Figure 3.2 (d) and (e) corresponds to a situation where there are group differences in both the cause-specific hazard for type 1 failure, and the all-cause hazard. Each simulation uses 1,000 Monte Carlo samples.

It is seen from Figure 3.2 (a) that the type I error rates for all three tests are well controlled around the 0.05 nominal level. When there is a group difference in only one of the two quantities (Figure 3.2 (b) or (c)), the chi-square joint test or the maximum join test can be much more powerful than the Bonferroni method. When there are group differences in both the cause-specific hazard for type 1 failure and the all-cause hazard and CSH ratio, and ACH ratio are very different (Figure 3.2(d)), chi-square test will have much higher power than Bonferroni test. Only in the case when CSH ratio and ACH ratio are very similar (0.09/0.15=0.6, 0.11/0.17=0.65), the three methods have similar powers with the maximum test being slightly more powerful. This has important implications for the design of a clinical trial in the presence of competing risks. For example, to achieve 80% power under the second scenario (Figure 3.2 (b)), it would require around $n = 460$ patients for the chi-square joint test, around $n = 1900$ patients for the joint maximum test, and at least $n = 2200$ patients for the Bonferroni adjustment method.

3.4 Real Data Example

We illustrate our methods on two real data sets. In the first example we consider joint inference for the cause-specific hazard and cumulative incidence for time

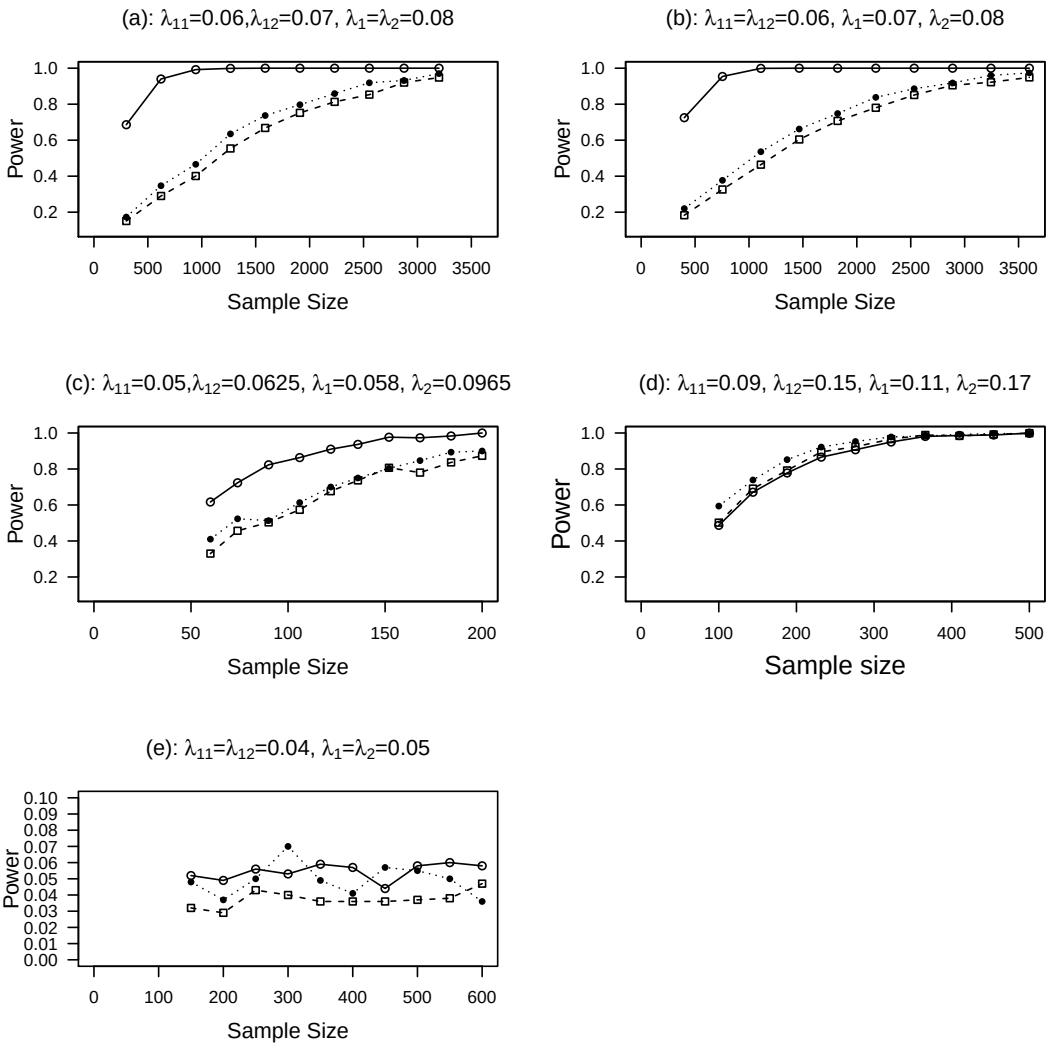


Figure 3.2: For CSH and ACH pair: Simulated power of the chi-square test (solid line), the maximum test (dotted line), and the Bonferroni method (dashed line) under a null case (panel (a)) and four different alternatives (panels (b)-(e))

to second malignancy in Hodgkin disease patients. In the second example, we perform joint analysis of the cause-specific hazard (CSH) for time to progression (TTP) and all-cause hazard for time to progression or death (progression free survival or PFS) for follicular type lymphoma patients.

3.4.1 Hodgkin Disease

The Hodgkin disease data was described in Pintilie (2006). It consists of 865 patients who were diagnosed with Hodgkin disease and received radio therapy in Princess Margaret Hospital between 1968 and 1986. Here we are interested in studying time to second malignancy after receiving radio therapy, which is an important variable for evaluating the side effects of radio therapy. Death without second malignancy is a competing risk. Among the 865 patients, 93 developed second malignancy, 386 were dead without the second malignancy, and 386 were right censored who did not experience any of the two events by the end of study. For illustration purpose, we investigate whether or not the risks of developing second malignancy were the same among older(≥ 30) and younger (< 30) patients.

Figure 3.3 and Figure 3.4 depict the cumulative cause-specific hazard functions and the cumulative incidence functions, respectively, for time to second malignancy for the older(≥ 30) and younger (< 30) groups. There appears to be a higher cause-specific hazard for the older patients since the slope of their cumulative cause-specific hazard is noticeably bigger (Figure 3.3). However, the cumulative incidence functions for the two age groups are barely distinguishable (Figure 3.4). The two-sample log-rank test for the cause-specific hazard for time to second malignancy yields a p-value=0.037. The Gray (1988) two-sample test for the cumulative incidence for time to second malignancy gives a p-value=0.770.

At 5% overall significant level, none of the individual tests is statistically significant at the Bonferroni adjusted level 0.05/2=0.025.

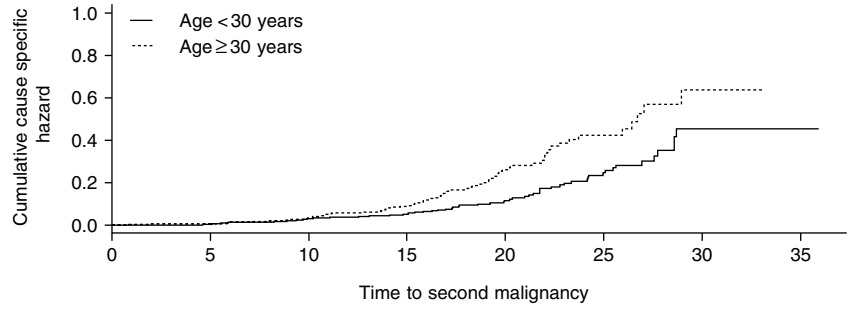


Figure 3.3: Cumulative cause-specific hazard functions for time to second malignancy for older($\geq$ 30) and younger ($<$ 30) patients. Log-rank test p-value=0.037.

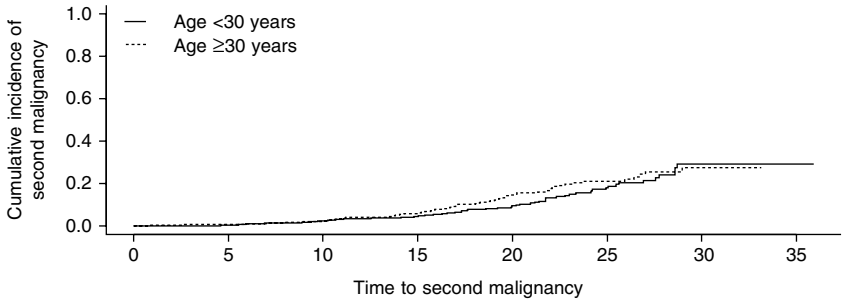


Figure 3.4: Cumulative incidence functions for time to second malignancy for older($\geq$ 30) and younger ($<$ 30) patients. Gray's test p-value=0.770.

We performed the chi-square joint test and the maximum test for the null hypothesis that there is no difference in the cause-specific hazard and the cumulative incidence for time to second malignancy between older and younger patients. The p-values are presented in Table 3.1, along with the results of the individual tests and the Bonferroni's method. In contrast to the Bonferroni method, the two-sample chi-square joint test for the cause-specific hazard and the cumulative incidence yields a p-value 0.02, which is highly significant at 5% significance level.

Table 3.1: Separate and Joint Test Results for Hodgkin Disease Example

	Separate Test		Joint Test		
Test	CSH	CIH	Bonferroni	χ^2	Max
p-value	0.037	0.770	0.074	0.020	0.050

NOTE: χ^2 and Max are abbreviations for the Chi-square test and the maximum test described in Section 2.2.

The maximum joint test is also significant at level 0.05 (p-value =0.05). Finally, in addition to an elevated cause-specific hazard for time to second malignancy, the older patients were also observed to have a higher risk of dying from other life-threatening diseases without developing second malignancy, which explains why their observed cumulative incidence was not significantly different from the younger patients. This is consistent with our simulation results in Section 4 that the chi-square test tends to be more powerful test when there is a difference in one quantity.

3.4.2 Follicular Cell Lymphoma Study

The follicular cell lymphoma study (Pintilie, 2006; Scheike and Zhang, 2011) consists of 541 early stage (I or II) follicular type lymphoma patients who were enrolled between 1967 and 1996 and treated with either radiation alone (RT) or with radiation and chemotherapy (CMT). There were 272 events due to disease (relapse or no treatment response), 76 competing risk events (death without relapse), and 193 censored individuals who didn't experience any of the two events at the end of the followup. As in Scheike and Zhang (2011), we test if the CMT group has a longer time to relapse or no treatment response than the RT group. Although one could study different pairs of quantities, we consider joint inference of the cause-specific hazard and the all-cause hazard based on models (3.18) and

Table 3.2: Separate and Joint Test Results for Follicular Cell Lymphoma Study

	Separate Test		Joint Test		
Test	CSH	CIH	Bonferroni	χ^2	Max
p-value	0.035	0.037	0.070	0.182	0.047

NOTE: χ^2 and Max are abbreviations for the Chi-square joint test and the maximum joint test.

(3.19) because they correspond to two commonly used endpoints, namely time to progression (TTP) and progression free survival (PFS), in oncology clinical trials. Here TTP, defined as time to relapse or no treatment response, is an important endpoint for the anti-tumor activity of an agent, and PFS, defined as time to progression or death before progression, is a common endpoint for the overall effects on a patient. In addition to a binary treatment variable (1 for RT and 0 for CMT), we adjust for patient's baseline age, stage, and Haemoglobin level (hgb) by including them as covariates in our models. The Cox-Snell residual plots for the proportional all-cause hazards model (Figure 3.5 (a)) and the proportional cause-specific hazards model (Figure 3.5 (b)) indicate reasonable overall fit of both models. We conducted chi-square joint test and the maximum joint test for the treatment variable and summarized the results along with Bonferroni adjustment method and the individual tests in Table 3.2. The maximum joint test (p-value= 0.047) is significant, whereas the chi-square joint test (p-value=0.182) and the Bonferroni method (p-value=0.07) are not significant at 5% significance level. The chi-square joint test has a relatively large p-value because it can only do two-sided test. Finally, the one-sided individual test statistics for CSH and ACH are 1.81 and 1.78, respectively, both exceeding 1.77, the cutoff value of the maximum test. Therefore we conclude that CMT group has a lower risk of TTP (cause-specific hazard) and a lower risk of PFS (ACH) as compared to the RT group at 5% overall significance level.

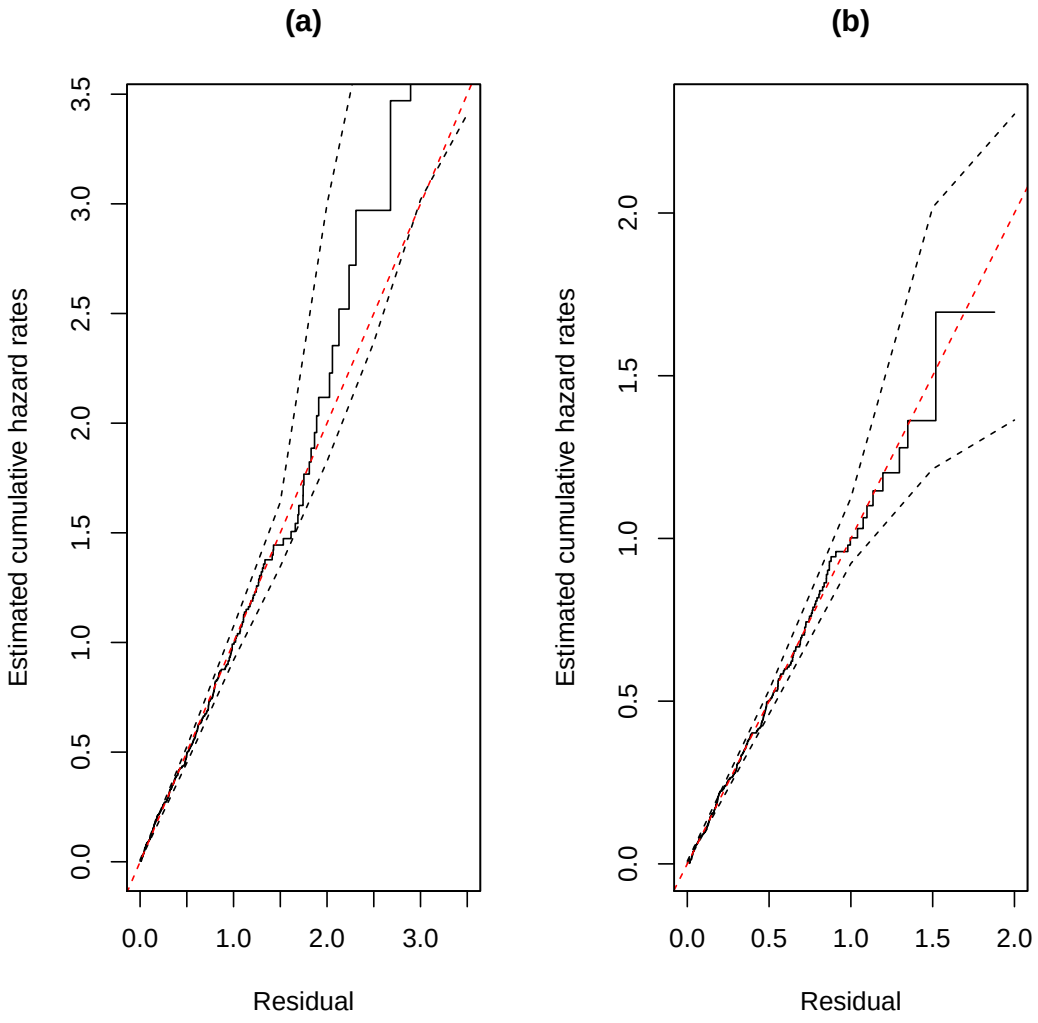


Figure 3.5: The Cox-Snell residual plots (solid lines) for all-cause hazard regression model (panel (a)) and cause-specific hazard model (panel(b)), with 95% bootstrap confidence intervals (dashed lines), and the 45 degree line (dotted lines)

CHAPTER 4

Sample Size Calculation for Joint Inference of Cause Specific Hazard and All Cause Hazard

The purpose of this chapter is to develop sample size calculation methods for jointly testing the cause specific hazard and the all cause hazard in a two sample comparison situation in the presence of competing risks and independent right censoring. The methods are based on the nonparametric two-sample joint tests developed in chapter 3. We also consider uniform patient entry and the administrative censoring.

It is worth noting that the cause-specific hazard and all-cause hazard correspond to two important surrogate endpoints, time to progression (TTP) and progression free survival (PFS), in oncology trials. For example, in cancer research, one is often interested in time to tumor progression while other correlated events, such as death without a tumor progression, can prevent one from seeing the event of interest. In this case, the TTP is a direct measure of the antitumor activity of the new drug or agent, whereas the PFS, time to either disease progression or time to death whichever happens first, is a commonly used surrogate endpoint in early phase oncology trials.

4.1 Joint Distribution for Two Individual Test Statistics under Alternative Hypothesis

In chapter 3, we developed the joint distribution for the two logrank test statistics (2.2) and (3.14) under the null hypothesis

$$H_0 : \lambda_{11}(t) = \lambda_{12}(t) \text{ and } \lambda_{\cdot 1}(t) = \lambda_{\cdot 2}(t). \tag{4.1}$$

Based on the joint distribution, we constructed a chi-square test statistics and maximum statistics to detect any group difference in the either CSH or ACH. In order to do sample size calculation, we also need to find the joint distribution of the two test statistics under the alternative hypothesis. Similar to other sample size studies (Schoenfeld, 1983; Schulgen et al., 2005), we assume local asymptotic alternative

$$H_a : \lambda_{11}(t)/\lambda_{12}(t) = 1 + \phi_1/\sqrt{n} \text{ or } \lambda_{\cdot 1}(t) = \lambda_{\cdot 2}(t) = 1 + \phi_\cdot/\sqrt{n}, \tag{4.2}$$

where ϕ_1 and $\phi_\cdot$ are bounded integrable function and have order of $O(1)$.

Theorem 6 *Let $U_{11}(\tau)$ and $U_{\cdot 1}(\tau)$ be defined by (2.2) and (3.14). The asymptotic joint distribution of $n^{-1/2}(U_{11}(\tau), U_{\cdot 1}(\tau))$ under a sequence of local alternatives (4.2) is an approximately bivariate normal distribution with mean $\boldsymbol{\mu} = (\mu_1, \mu_2)$ and variance-covriance matrix $\boldsymbol{\Sigma}$, where*

$$\mu_1 = \int_0^\tau \phi_1 \frac{y_1(t)y_2(t)}{y_1(t) + y_2(t)} d\Lambda_{11}(t), \tag{4.3}$$

$$\mu_2 = \int_0^\tau \phi_\cdot \frac{y_1(t)y_2(t)}{y_1(t) + y_2(t)} d\Lambda_{\cdot 1}(t), \tag{4.4}$$

and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11}\sigma_{12} \\ \\ \sigma_{21}\sigma_{22} \end{pmatrix}, \tag{4.5}$$

where

$$\sigma_{11} = \int_0^\tau \frac{y_1(t)y_2(t)}{y_1(t)+y_2(t)} d\Lambda_{11}(t), \tag{4.6}$$

$$\sigma_{22} = \int_0^\tau \frac{y_1(t)y_2(t)}{y_1(t)+y_2(t)} d\Lambda_{\cdot 1}(t), \tag{4.7}$$

and

$$\sigma_{12} = \sigma_{21} = \int_0^\tau \frac{y_1(t)y_2(t)}{y_1(t)+y_2(t)} d\Lambda_{11}(t), \tag{4.8}$$

with $\Lambda_{jk}(t) = \int_0^t \lambda_{jk}(u)du$ *and* $\Lambda_{\cdot k}(t) = \int_0^t \lambda_{\cdot k}(u)du$, *and* $y_k(t) = a_k S_k(t) S_c(t)$.

Proof of Theorem 6: Under usual regularity condition, by using multivariate martingale central limiting theorem (Fleming and Harrington (1991), Theorem 5.3.5), we can prove that $(n^{-1/2}U_{11}, n^{-1/2}U_{\cdot 1})^T$ has a bivariate normal limiting distribution with mean $\boldsymbol{\mu} = (\mu_1, \mu_2)$ and variance-covariance matrix $\boldsymbol{\Sigma} = (\sigma_{ij})$, where

$$\begin{aligned}
\mu_1 &= \int_0^\tau \phi_1 \tfrac{y_1(t)y_2(t)}{y_1(t)+y_2(t)} d\Lambda_{11}(t), \\
\mu_2 &= \int_0^\tau \phi_\cdot \tfrac{y_1(t)y_2(t)}{y_1(t)+y_2(t)} d\Lambda_{\cdot 1}(t), \\
\sigma_{11} &= \int_0^\tau \tfrac{y_1(t)y_2(t)}{y_1(t)+y_2(t)} d\Lambda_{11}(t), \\
\sigma_{22} &= \int_0^\tau \tfrac{y_1(t)y_2(t)}{y_1(t)+y_2(t)} d\Lambda_{\cdot 1}(t), \\
\sigma_{12} &= \int_0^\tau \tfrac{y_1(t)y_2(t)}{y_1(t)+y_2(t)} d\Lambda_{11}(t).
\end{aligned} \tag{4.9}$$

Here $\Lambda_{\cdot 1}(t) = \int_0^t \lambda_{\cdot 1} dt$ is the all-cause cumulative hazard in group 1. Note that σ_{12} is the limit of

$$\begin{aligned}
& \left\langle n^{-1/2}U_{11}, n^{-1/2}U_{\cdot 1} \right\rangle \\
= \; & n^{-1} \left\langle \int_0^\tau \tfrac{Y_1(t)Y_2(t)}{Y_\cdot(t)} \left\{ \tfrac{dM_{11}(t)}{Y_1(t)} - \tfrac{dM_{12}(t)}{Y_2(t)} \right\}, \int_0^\tau \tfrac{Y_1(t)Y_2(t)}{Y_\cdot(t)} \left\{ \tfrac{dM_{\cdot 1}(t)}{Y_1(t)} - \tfrac{dM_{\cdot 2}(t)}{Y_2(t)} \right\} \right\rangle \\
= \; & n^{-1} \left\{ \int_0^t \tfrac{Y_1^2(t)Y_2^2(t)}{Y_\cdot(t)^2} \left(\tfrac{d<M_{11}(t), M_{\cdot 1}(t)>}{Y_1^2(t)} + \tfrac{d<M_{12}(t), M_{\cdot 2}(t)>}{Y_2^2(t)} \right) \right\} \\
= \; & n^{-1} \left\{ \int_0^t \tfrac{Y_1^2(t)Y_2^2(t)}{Y_\cdot(t)^2} \left(\tfrac{d\Lambda_{11}(t)}{Y_1(t)} + \tfrac{d\Lambda_{12}(t)}{Y_2(t)} \right) \right\},
\end{aligned}$$

which converges in probability to $\int_0^\tau \frac{y_1(t)y_2(t)}{y_\cdot(t)} d\Lambda_{11}(t)$. Similarly, the diagonal elements are the asymptotic variances of $n^{-1/2}U_{11}$ and $n^{-1/2}U_{\cdot 1}$, respectively (Fleming and Harrington, 1991). $\square$

4.2 Sample Size Calculation for Chisquare Test Statistics

4.2.1 Non-central Chisquare Test Statistics under Alternative Hypothesis

Define

$$X^2 = n^{-1} \left(U_1(\tau), U_{\cdot 1}(\tau)\right) \mathbf{\Sigma}^{-1} \begin{pmatrix} U_{11}(\tau) \\ U_{\cdot 1}(\tau) \end{pmatrix}. \tag{4.10}$$

It follows from theorem (6) that X^2 has an approximate non-central chi-square distribution with 2 degree freedom and non-centrality parameter ξ, where

$$\xi = n\boldsymbol{\mu}^T \mathbf{\Sigma}^{-1} \boldsymbol{\mu}. \tag{4.11}$$

Analogous to Schoenfeld's method Schoenfeld (1981), we can also do sample size calculation in two steps, calculate the number of failures due to cause of interest first, and then get the actual number of patients needed to enroll in the study. In the next two subsections, we will show why the two step methods carries to the joint test situation and how to implement them in details.

4.2.2 The Required Number of Failure Due to Cause of Interest

In this subsection, we will calculate the number of failure due to cause 1 needed first for the chi-square test. We will first state a lemma to establish the relationship between ξ and the number of failure due to cause 1.

Lemma 1 *The number of failure due to cause* 1 *can be calculated by*

$$D_1 = \frac{\xi(1-R)}{a_1 a_2 \left[(\phi_1 - \phi_\cdot)^2 + \phi_\cdot^2 (1 - 1/R)\right]}, \tag{4.12}$$

where a_1, a_2 *are the sample proportions in group 1 and group 2, respectively, and* R *is the relative risk of failure due to cause 1 versus any cause in group* 1.

Proof of Lemma 1:

$$\begin{aligned}
\mu_1 &= \int_0^\tau \phi_1 \tfrac{y_1(t)y_2(t)}{y_1(t)+y_2(t)} d\Lambda_{11}(t) \\
&= \int_0^\tau \phi_1 \tfrac{a_1 S_1(t) S_c(t) a_2 S_2(t) S_c(t)}{a_1 S_1(t) S_c(t) + a_2 S_2(t) S_c(t)} d\Lambda_{11}(t) \\
&= \int_0^\tau \phi_1 a_1 a_2 \tfrac{S_2(t)}{a_1 S_1(t) + a_2 S_2(t)} S_1(t) S_c(t) d\Lambda_{11}(t) \\
&= \int_0^\tau \phi_1 a_1 a_2 \tfrac{e^{-\int_0^t \lambda_{\cdot 2}(u)du}}{a_1 e^{-\int_0^t \lambda_{\cdot 1}(u)du} + a_2 e^{-\int_0^t \lambda_{\cdot 2}(u)du}} S_1(t) S_c(t) d\Lambda_{11}(t) \\
&\approx \phi_1 a_1 a_2 P_{11}(t),
\end{aligned} \tag{4.13}$$

and by similar derivations, we can get

$$\begin{aligned}
\mu_2 &= \int_0^\tau \phi_\cdot \tfrac{y_1(t)y_2(t)}{y_1(t)+y_2(t)} d\Lambda_{\cdot 1}(t) = \phi_\cdot a_1 a_2 P_{\cdot 1}(t), \\
\sigma_{11} &= \int_0^\tau \tfrac{y_1(t)y_2(t)}{y_1(t)+y_2(t)} d\Lambda_{11}(t) = a_1 a_2 P_{11}(t), \\
\sigma_{22} &= \int_0^\tau \tfrac{y_1(t)y_2(t)}{y_1(t)+y_2(t)} d\Lambda_{\cdot 1}(t) = a_1 a_2 P_{\cdot 1}(t), \\
\sigma_{12} &= \int_0^\tau \tfrac{y_1(t)y_2(t)}{y_1(t)+y_2(t)} d\Lambda_{11}(t) = a_1 a_2 P_{11}(t),
\end{aligned} \tag{4.14}$$

where a_1, a_2 are the sample proportions in group 1 and group 2, respectively, $P_{11}(t) = \int_0^\tau S_1(t) S_c(t) d\Lambda_{11}(t)$ and $P_{\cdot 1}(t) = \int_0^\tau S_1(t) S_c(t) d\Lambda_{\cdot 1}(t)$ are the cumulative incidence in group 1 by time t due to cause 1 and all cause, respectively.

After plugging all the quantities from (4.13) and (4.14) to (4.11), we can get

$$\begin{aligned}
\xi &= \tfrac{a_1 a_2 \left[\phi_1^2 n P_{11}(t) - 2\phi_1 \phi_\cdot n P_{11}(t) + \phi_\cdot^2 n P_{\cdot 1}(t)\right]}{1 - n P_{11}(t)/n P_{\cdot 1}(t)} \\
&= \tfrac{a_1 a_2 \left[\phi_1^2 D_1(t) - 2\phi_1 \phi_\cdot D_1(t) + \phi_\cdot^2 D_\cdot(t)\right]}{1 - D_1(t)/D_\cdot(t)} \\
&= \tfrac{a_1 a_2 \left[\phi_1^2 D_1(t) - 2\phi_1 \phi_\cdot D_1(t) + \phi_\cdot^2 D_\cdot(t)/R(t)\right]}{1 - R(t)} \\
&= \tfrac{a_1 a_2 \left[(\phi_1 - \phi_\cdot)^2 D_1 + \phi_\cdot^2 (1/R(t) - 1) D_1\right]}{1 - R(t)},
\end{aligned} \tag{4.15}$$

where $R(t)$ is the relative risk of failure due to cause 1 versus all causes. We can solve for D_1 from (4.15). $\square$

Given type I error α and type II error β, we can establish the two error equations below.

$$\alpha = P(\text{Reject } H_0 | H_0) = P(X^2 > C | X^2 \sim \chi_2^2(0)), \tag{4.16}$$

$$1-\beta = P(\text{Reject } H_0|H_\alpha) = P(X^2 > C|X^2 \sim \chi_2^2(\xi)), \tag{4.17}$$

where C is the critical value for the test. We can solve for C from (4.21). After plugging C in (4.22), theoretically speaking, we can solve for ξ and in turn solve for D_1 by (4.12). However, (4.22) involves a non-central chi-square distribution, so it's hard to find an analytical formula for ξ. In this paper, we will use an algorithm (proposed by an unpublished dissertation by Nan Zhang) to solve the equations iteratively. The algorithm for solving D_1 is listed below.

1. In k^{th} iteration, let $D_1^{(k)} = k$ and $D_1^{(k)*} = k+1$;
2. Calculate $\xi^{(k)}$ and $\xi^{(k)*}$ by plugging in $D_1^{(k)}$ and $D_1^{(k)*}$ in (4.12) respectively;
3. Use (4.22) to get $C_{D_1^{(k)}}$ and $C_{D_1^{(k)*}}$;
4. If C falls between $C_{D_1^{(k)}}$ and $C_{D_1^{(k)*}}$, return sample size $D_1^{(k)}$. Otherwise, go back to Step (2).

To summarize, we need the following information to find the required number of failure due to cause of interest for the trial in order to have $1-\beta$ power to detect treatment effect at α level.

- Pre-specified type I error α and power $1-\beta$;
- The sample proportion allocation in group 1 and group 2, a_1 and a_2;
- Relative risk of failure due to cause 1 versus any causes in group 1, which is denoted by R.
- CSH ratio for cause 1 in group 1 versus group 2, ϕ_1;
- ACH ratio for group 1 versus group 2, $\phi_.$.

4.2.3 The Required Number of Patients

Similar to Schoenfeld (1983), the actual number of patients needed to enroll in the trial can be calculated by

$$N = D_1 / \left(a_1 * P_{11} + a_2 * P_{12} \right), \tag{4.18}$$

where P_{1k} is the probability of observing a failure due to cause 1 in group k, a_1 and a_2 ($a_1 + a_2 = 1$) are patients allocation in two groups.

In practice, when we do sample size calculation, we could make some assumptions to simply the required input partakers from practitioner. For example, we may assume both CSH and ACH are constant in control and treatment groups ($\lambda_{1k}(t) = \lambda_{1k},\ \lambda_{\cdot k}(t) = \lambda_{\cdot k}, k = 1, 2$). The censoring hazard is also assumed to be constant, denoted by λ_c. Patients enter study uniformly over accrual period $[0, t_1]$ and the follow-up time is t_2. Under these assumptions, according to Lachin and Foulkes (1986) Schulgen et al. (2005), the probability of observing a failure due to cause j by time $t_1 + t_2$ can be calculated by

$$P_{jk} = \frac{\lambda_{jk}}{\lambda_{\cdot k}} \left[1 - \frac{\exp(-\lambda_{\cdot k} t_2) - \exp(-\lambda_{\cdot k}(t_1 + t_2))}{\lambda_{\cdot k} t_1} \right], \tag{4.19}$$

where $\lambda_{\cdot k} = \lambda_{1k} + \lambda_{2k} + \lambda_c$, and the probability of observing a failure due to any cause by time $t_1 + t_2$ is

$$P_{\cdot k} = \frac{\lambda_{1k} + \lambda_{2k}}{\lambda_{\cdot k}} \left[1 - \frac{\exp(-\lambda_{\cdot k} t_2) - \exp\left(-\lambda_{\cdot k}(t_1 + t_2)\right)}{\lambda_{\cdot k} t_1} \right]. \tag{4.20}$$

In summary, to calculate the actual number of patients required for the trial, besides the information listed in last subsection, we also need to know the accrual time t_1, the maximum follow-up time t_2, cumulative incidence for cause 1 in both groups, P_{11} and P_{12} or equivalently λ_{11} and λ_{12}.

4.3 Sample Size Calculation for Maximum Test Statistics

Define

$$M = \max(|Z_{11}|, |Z_{\cdot 1}|)),$$

where $Z_{11} = U_{11}/\sqrt{\sigma_{11}}$ and $Z_{\cdot 1} = U_{\cdot 1}/\sqrt{\sigma}_{22}$. It follows from Theorem 1 that for large samples, the distribution of $n^{-1/2}(Z_{11}, Z_{\cdot 1})$ can be approximated by the bivariate normal distribution $N\left((0,0)^T, (1,1,\sqrt{R})\right)$ under null hypothesis (4.1), and $N\left(\left(\phi_1\sqrt{a_1a_2P_{11}(t)}, \phi_\cdot\sqrt{a_1a_2P_{11}(t)/R}\right)^T, \left(1,1,\sqrt{R}\right)\right)$ under alternative hypothesis (4.2).

Given type I error α and type II error β, we can establish the two error equations below,

$$\alpha = P(\text{Reject } H_0|H_0) = P(M > C^*|H_0), \tag{4.21}$$

and

$$1-\beta = P(\text{Reject } H_0|H_a) = P(M > C^*|H_a), \tag{4.22}$$

where C^* is the critical value for the test.

We can solve for C^* from equation (4.21). Since we can not find the exact distribution functions for M, we will need to proximate them by using Monte Carlo samples. Specifically, we generate N pairs of random variables from the bivariate normal distribution $N\left((0,0)^T, (1,1,\sqrt{R})\right)$. For the l-th generated pair, compute the maximum absolute value, and denoted it by $M_l, l = 1, ..., N$. We can approximate C^* by the upper $100(1-\alpha)$-th sample quantile of $M_1, \ldots, M_N$.

For a fixed alternative, we can prove that $(Z_{11}, Z_{\cdot 1})$ has a limiting distribution with mean $\left(\phi_1\sqrt{a_1a_2D_1}, \phi_\cdot\sqrt{a_1a_2D_1/R}\right)$. To solve equation (4.22) for sample size, we need to use a similar algorithm proposed in the previous section.

1. In k^{th} iteration, let $D_1^{(k)} = k$, and generate N pairs of random numbers $(W_i, V_i), i = 1, ..., N$ from bivariate normal distribution

$$N\left(\left(\phi_1\sqrt{a_1a_2D_1}, \phi_\cdot\sqrt{a_1a_2D_1/R}\right), (1, 1, \sqrt{R})\right).$$

2. Calculate the maximum absolute value within each pair and denote it by $M_l, l = 1, ..., N$. Find the first M_l which is bigger than C^* and denote the index by l^*.

3. If $1 - l^*/N > 1 - \beta$, then return $D_1 = D_1^{(k)}$. Otherwise go back to step 1.

Similar to Chisquare test, we can get the number of patients needed to be enrolled in the trial by (4.18) (4.19) and (4.20).

Remark 4.1. It is straightforward to modify the maximum joint test procedure to test one-sided alternative(s) based on either $T^* = \max(Z_{11}, \tilde{Z}_{11})$, $T^* = \max(|Z_{11}|, \tilde{Z}_{11})$, or $T^* = \max(Z_{11}, |\tilde{Z}_{11}|)$ as deemed appropriate.

4.4 Simulation Studies

A simulation study was conducted to illustrate the saving in sample sizes for using Chiquare test or maximum test compare to Bonferroni Test. The CSH for cause 1 in group 1, λ_{11}, was set to be 0.05. Simulation results with different values for λ_{11} are similar and available upon request. The relative risk of failure due to cause 1 versus all causes in group 1, R, is set to be 0.6 and 0.85 to indicate a middle and high relative risk. The censoring rate was set to be 0.1. Various combination of CSH ratio for cause 1 and ACH ratio were used to cover most of situations we can encounter in real analysis. In order to make sure $\lambda_{\cdot 2} > \lambda_{12}$, we need to choose the parameter combinations which meet $\phi_1 > R * \phi_\cdot$. The Monte Carlo sample we used to approximate distribution of T under alternative hypothesis (4.2), and

distributions of M under both null and alternative hypothesis (4.1) and (4.2) is set to be 5000. Let the maximum followup time to be $t = 200$. Type I error α was chose to be 0.05, to achieve a 80% power, both number of failure due to cause 1 and total number of patients required by all three methods are listed in in table 4.1.

As we can see in table 4.1, chisquare test requires less sample sizes than maximum test or Bonferroni method in almost all cases except when CSH ratio and ACH ratio is the same. The bigger the differences between CSH ratio and ACH ratio, the smaller the sample sizes required by chi-square test compare to maximum test or Bonferroni method. In the cases when difference in CSH ratio and ACH are the same, maximum test requires least sample size among all three methods.

4.5 Real Data Example

We illustrate our method by using a 4D trial (Die Deutsche Disbetes Dialyse Studie) (Wanner et al., 1999). This 4D trial is to evaluate the efficacy of antihyperlipidemic treatment with atovastatin, in reducing cardiovascular mortality and frequency of non-fatal myocardial infection. It is a randomized, double-blinded, placebo-controlled trial. The event of interest is the composite endpoint of death due to cardiovascular disease and non-fatal myocardial infarction, and the possible competing risks is death due to other cause. Schulgen et al. (2005) conducted sample size calculation based on the test of the cause-specific hazard(CSH) for event of interest (death due to cardiovascular disease or non-fatal myocardial infection). As discussed earlier in Li and Yang (2013), CSH is not enough to quantify competing risk data, so a joint test is necessary to compare the treatment with placebo with respect to time to event of interest. In our paper, we

Table 4.1: Sample Size Based on the Chi-square Test and the Bonferroni Test for the Joint Hypothesis of CSH and ACH ($\alpha = 0.05$, $1 - \beta = 0.8$)

R	CSH ratio	ACH ratio	Chi-square Test		Maximum Test		Bonferroni Test	
			Event	Patient	Event	Patient	Event	Patient
0.85	0.6	0.6	125	163	104	136	124	162
	0.6	0.7	125	150	133	160	146	175
	0.7	0.6	82	116	115	162	115	162
	0.7	0.7	257	334	217	282	255	330
	0.7	0.8	217	263	270	327	299	362
	0.8	0.6	44	67	109	163	108	162
	0.8	0.7	143	199	222	308	238	330
	0.8	0.8	657	852	540	700	650	841
	0.8	0.9	365	445	644	785	764	930
	0.9	0.6	27	43	110	173	103	162
	0.9	0.7	67	99	228	334	226	330
	0.9	0.8	255	350	577	792	613	841
	0.9	0.9	2950	3819	2020	2615	2912	3769
0.60	0.6	0.6	88	162	82	152	88	162
	0.6	0.7	141	239	122	207	146	247
	0.6	0.8	132	208	137	216	146	229
	0.6	0.9	89	131	140	206	146	214
	0.7	0.6	77	154	83	166	82	162
	0.7	0.7	181	333	164	302	180	330
	0.7	0.8	302	517	266	456	299	512
	0.7	0.9	218	350	283	454	299	480
	0.9	0.6	45	101	83	185	73	162
	0.9	0.7	104	216	169	351	160	330
	0.9	0.8	327	636	433	842	433	841
	0.9	0.9	2083	3820	1877	3442	2055	3769

Note: R represents relative risk of failure due to cause 1 versus any cause.

illustrate how to re-design the trial with respect to CSH and ACH. We will show that there is a big saving in sample size by using chi-square test or maximum test compare to Bonferroni method under most scenarios.

First, we calculate the number of cardiovascular death or non-fatal myocardial infection needed to detect the treatment difference. In Schulgen et al. (2005)'s paper, they used prior information from two pervious studies about diabetes patients (Koch et al., 1997; Tschöpe et al., 2008). Beside all the information needed in Schulgen et al. (2005), such as type I error α, power $1-\beta$, patients allocation a_1 and a_2, and the cause specific hazard ratio between treatment group and control with respect to event of interest (death due to cardiovascular disease or non-fatal myocardial infection) ϕ_1, we also need to know $\phi_.$, the all cause hazard ratio between the two groups , and R, the relative risk of event of interest to death due to any cause or non-fatal myocardial infection in control group.

Similar to Schulgen et al. (2005), we set $\alpha = 0.05$, $\beta = 0.1$, $a_1 = a_2 = 0.5$, and $\phi_1 = 0.73$. We also assume $R = 0.26/0.4$, where 0.26 is the CSH for cause of interest and 0.4 is ACH based on numbers used in table 1 in Schulgen et al. (2005)'s paper. For all cause hazard ratio, we assume three different scenarios (table 4.2), some beneficial treatment effect on competing risk ($\phi_. = 0.3/0.4$, scenario A), adverse treatment effect on competing risk ($\phi_. = 0.36/0.4$, scenario B), and no treatment effect on competing risk (death due to other causes) ($\phi_. = 0.33/0.4$, scenario C). We can plug in all the prior information to our R function and get the required number of event of interest for both chi-square test and maximum test.

In order to calculate the actual number of patients to be enrolled in the study, we need to calculate all the cause specific hazards based on each specific cumulative incidence Schulgen et al. (2005). Assume the accrual period $t_1 = 1.5$,

Table 4.2: Assumptions for Hazard Rates under Control and Experimental Treatment

	Scenario A		Scenario B		Scenario C	
	Control	treatment	Control	treatment	Control	treatment
Cause of interest	0.26	0.19	0.26	0.19	0.26	0.19
Competing Risk	0.14	**0.11**	0.14	**0.17**	0.14	**0.14**
All Causes	0.40	0.30	0.40	0.36	0.40	0.33

Note: Scenario A, B, and C correspond to the scenarios that there is beneficial, adverse, and no treatment effects on failure due to the competing risks, respectively.

a trial duration $t_2 = 4$, and equal allocation for the patients $a_1 = a_2 = 0.5$, we can calculate the probability of observing an event of interest and the number of patients enrolled based (4.18) and (4.19).

The results for both number of events of interest and total number of patients enrolled are summarized in table 4.3. We can see in all three scenarios, maximum test perform better than Bonferroni method, and Chisquare test works the best when if there is any beneficial or adverse effect from the treatment group to death due to other causes.

Table 4.3: Sample Sizes for Chi-square Joint Test, Maximum Test, and Bonferroni Joint Test

	Chi-square		Maximum		Bonferroni	
Scenario	EOI	Patient	EOI	Patient	EOI	Patient
A	155	333	158	339	163	349
B	358	712	460	915	505	1003
C	511	1030	462	931	505	1017

Note: EOI is number of event of interested required to detect treatment effect.

CHAPTER 5

Discussion and Future Work

We emphasize the need for joint inference of the cause-specific hazard and the cumulative incidence because one quantity alone does not fully capture the effects of a variable on a particular type of failure. The developed joint inference procedures can be considerably more powerful than the Bonferroni method as illustrated in our simulation and real data examples. This has important practical implications for the analysis and design of clinical trials with competing risks data. For example, the increased power of the joint tests implies that fewer patients would be required in a clinical trial. In a sequel, we develop power analysis methods to determine the required sample size to test a group difference based on the developed joint tests.

We also gave some key steps in developing joint tests for the K-sample ($K > 2$) problem in Chapter 3. The joint regression methods in Chapter 3 Section 2 can also be extended to beyond Coxs models. For example, the accelerated failure time models can be used to model the cause-specific hazard. Scheike and Zhang (2008) considered other regression models for the subdistribution hazard. Joint inference procedures for these models can be developed similarly.

As we discussed in Chapter 2, to study the group difference in type 1 failure, it is sufficient to test any one of the three equivalent hypotheses (3.1), (3.12), and (3.13). The three pairs of test statistics will have different powers because they test different alternatives. We did some simulation under various alternative

hypothesis. Generally speaking, when there is a group difference in CSH, but no group difference in ACH, then the maximum test and Bonferroni method for the CSH & ACH pair will perform the worst in comparison to chi-square test for CSH & ACH pair, and all the joint tests with respect to the other two pairs. If there is a group difference in CSH, but no group difference in OCH, the three joint test statistics with respect to pair CSH & OCH performs a little worse than the joint test statistics with respect to two other pairs. Since we can not predict where the true group difference is in terms of CSH, ACH or CIF, which pair to choose for testing the null hypothesis will depend on the research question we are truly interested in. For instance, in our Hodgkin disease example, if our interest is really the time to second malignancy, we should choose the CSH & CIF pair.

We have also done some simulation for CSH and ACH pair with respect to Cox model. For simplicity, only two covariates are included. One is the group variable which is our interest, and another is gender variable, which is a nuisance parameter. We draw power curves under different alternatives. Chisquare test has highest power nearly all circumstances, which Bonfferonni perform the worst. The conclusion is similar to what we have for the two sample comparison.

Currently we developed sample size calculation for two sample joint test of cause specific hazard and all cause hazard. In the future, we could also develop similar procedure for the CSH and CIF as well as in regression settings.

CHAPTER 6

Appendix

R functions for performing sample size calculation for chi-square test and maximum test proposed in this dissertation

```
#########################################################################
## Sample size for Maximum Test
## alpha,beta,lambda_11,RR,phi_1,phi_all,t1,t2,censorrate are required inputs:
## -- alpha:  type I error
## -- beta:  type II error
## -- lambda_11: Cause Specific Hazard for cause 1 in group 1
##-- phi1: 'log(phi_1)' is expected cause specific hazard ratio in alternative.
##-- phi_all:  'log(phi_all)' is expected all cause hazard ratio in alternative.
##-- RR:        Relative risk of cause of interest vs. all causes in group 1.
##-- t1: patient accrual time period
##-- t2:       maximum follow up period
##-- censorrate: censor rate
##-- a1: sample allocation proportion in group 1
#########################################################################
library(MASS)
```

```
Maximum_RR<-function(alpha,beta,lambda_11,RR,phi_1,phi_all,t1,t2,censorrate,a1)
{
  ### Input parameters ###
  alpha=alpha
  beta=beta
  censorrate=censorrate
  lambda_11=lambda_11
  RR=RR
  phi_1=phi_1
  phi_all=phi_all
  lambda_a1=lambda_11/RR
  lambda_12=lambda_11/phi_1
  lambda_a2=lambda_a1/phi_all
  a2=1-a1

  # parameter combination has to meet phi_1 > RR * phi_all
  if (lambda_a2< lambda_12) {
    print(NA)
    break
  }

  # same censoring distribution in two groups
  lambda_c=censorrate*(lambda_a1+lambda_a2)/2

  corr=matrix(c(1,sqrt(RR),sqrt(RR),1),nrow=2,ncol=2)
  nbootstrap=5000
```

```
#simulate null distribution
sample=mvrnorm(n = nbootstrap, c(0,0), corr, tol=1e-4)
max=matrix(NA,nrow=nbootstrap,ncol=1)
max_oneside=matrix(NA,nrow=nbootstrap,ncol=1)
for (j in 1:nbootstrap){
  max[j]=max(abs(sample[j,]))
  max_oneside[j]=max(sample[j,])
}
cutoff2= quantile(max, 1-alpha)

d_m=1 #initial value
for (i in 1: 5000){
  mu_1=-log(phi_1)*sqrt(a1*a2*d_m )      # calculate \mu_1
  mu_a=-log(phi_all)*sqrt(a1*a2*d_m/RR)     # calculate \mu_all
  sample2=mvrnorm(n = nbootstrap, c(mu_1,mu_a), corr, tol=1e-4)
  max=matrix(NA,nrow=nbootstrap,ncol=1)
  for (j in 1:nbootstrap){
    max[j]=max(abs(sample2[j,]))
  }
  index<-seq(1,nbootstrap)
  index.a<-index[sort(max)>cutoff2]
  if (1-index.a[1]/nbootstrap < 1-beta)
  {d_m=d_m+1}
  else {break}
  if (i==5000) print("reach maximum loop")
}
```

```
# Calculate the probability of observing event of interest
lambda_all1=lambda_a1+lambda_c
lambda_all2=lambda_a2+lambda_c

if (t1==0){
  P_11=lambda_11/lambda_all1*(1-(exp(-lambda_all1*t2)))
  P_12=lambda_12/lambda_all2*(1-(exp(-lambda_all2*t2)))

  P_a1=lambda_a1/lambda_all1*(1-(exp(-lambda_all1*t2)))
  P_a2=lambda_a2/lambda_all2*(1-(exp(-lambda_all2*t2)))
}
else {
  P_11=lambda_11/lambda_all1*(1-(exp(-lambda_all1*t2)
   -exp(-lambda_all1*(t1+t2)))/lambda_all1*t1)
  P_12=lambda_12/lambda_all2*(1-(exp(-lambda_all2*t2)
   -exp(-lambda_all2*(t1+t2)))/lambda_all2*t1)

  P_a1=lambda_a1/lambda_all1*(1-(exp(-lambda_all1*t2)
   -exp(-lambda_all1*(t1+t2)))/lambda_all1*t1)
  P_a2=lambda_a2/lambda_all2*(1-(exp(-lambda_all2*t2)
   -exp(-lambda_all2*(t1+t2)))/lambda_all2*t1)
}

P_1=a1*P_11+a2*P_12
P_a=a1*P_a1+a2*P_a2
```

```
  samplesize_m = d_m/P_1

  return(list(d_m=d_m, samplesize_m=samplesize_m))
}



#####################################################################
## Sample size for chi-square test
## alpha,beta,lambda_11,RR,phi_1,phi_all,t1,t2,censorrate are required inputs:
## -- alpha:  type I error
## -- beta:  type II error
## -- lambda_11: Cause Specific Hazard for cause 1 in group 1
##-- phi1: 'log(phi_1)' is expected cause specific hazard ratio in alternative.
##-- phi_all:  'log(phi_all)' is expected all cause hazard ratio in alternative.
##-- RR:        Relative risk of cause of interest vs. all causes in group 1.
##-- t1: patient accrual time period
##-- t2:       maximum follow up period
##-- censorrate: censor rate
##-- a1: sample allocation proportion in group 1
#####################################################################
library(MASS)

Chisquare_RR<-function(alpha,beta,lambda_11,RR,phi_1,phi_all,t1,t2,censorrate,a1
{

  ### Input parameters ###
```

```
 alpha=alpha
 beta=beta
 censorrate=censorrate
 lambda_11=lambda_11
 RR=RR
 phi_1=phi_1
 phi_all=phi_all
 lambda_a1=lambda_11/RR
 lambda_12=lambda_11/phi_1
 lambda_a2=lambda_a1/phi_all
 a2=1-a1

# parameter combination has to meet phi_1 > RR * phi_all
 if (lambda_a2< lambda_12) {
   print(NA)
   break
 }

 # same censoring distribution in two groups
 lambda_c=censorrate*(lambda_a1+lambda_a2)/2

 ##############################################################################
 ## Find D, number of event of interest, for chisquare test by iterations
 c = qchisq(1-alpha,2,0)
 d = 1
 noncentral1 = a1*a2*((log(phi_1)-log(phi_all))^2*d
  +log(phi_all)^2*d*(1/RR-1))/(1-RR)
```

```
noncentral2 = a1*a2*((log(phi_1)-log(phi_all))^2*(d+1)
 +log(phi_all)^2*(d+1)*(1/RR-1))/(1-RR)
while ( !(qchisq(beta,2,noncentral1) < c & qchisq(beta,2,noncentral2) >= c) )
{
  d = d+1
  noncentral1 = a1*a2*((log(phi_1)-log(phi_all))^2*d
   +log(phi_all)^2*d*(1/RR-1))/(1-RR)
  noncentral2 = a1*a2*((log(phi_1)-log(phi_all))^2*(d+1)
   +log(phi_all)^2*(d+1)*(1/RR-1))/(1-RR)
}
d_c=d

# Calculate the probability of observing event of interest
lambda_all1=lambda_a1+lambda_c
lambda_all2=lambda_a2+lambda_c

if (t1==0){
  P_11=lambda_11/lambda_all1*(1-(exp(-lambda_all1*t2)))
  P_12=lambda_12/lambda_all2*(1-(exp(-lambda_all2*t2)))

  P_a1=lambda_a1/lambda_all1*(1-(exp(-lambda_all1*t2)))
  P_a2=lambda_a2/lambda_all2*(1-(exp(-lambda_all2*t2)))
}
else {
  P_11=lambda_11/lambda_all1*(1-(exp(-lambda_all1*t2)
   -exp(-lambda_all1*(t1+t2)))/lambda_all1*t1)
  P_12=lambda_12/lambda_all2*(1-(exp(-lambda_all2*t2)
```

```
     -exp(-lambda_all2*(t1+t2)))/lambda_all2*t1)

    P_a1=lambda_a1/lambda_all1*(1-(exp(-lambda_all1*t2)
     -exp(-lambda_all1*(t1+t2)))/lambda_all1*t1)
    P_a2=lambda_a2/lambda_all2*(1-(exp(-lambda_all2*t2)
     -exp(-lambda_all2*(t1+t2)))/lambda_all2*t1)
  }

  P_1=a1*P_11+a2*P_12
  P_a=a1*P_a1+a2*P_a2

  samplesize_c = d_c/P_1

  return(list(d_c=d_c,samplesize_c=samplesize_c))
}

##################
### real data example ====== Diabetics
Chisquare_RR(alpha=0.05,beta=0.1,lambda_11=0.26,RR=0.26/0.4,phi_1=0.19/0.26,
phi_all=0.33/0.4,t1=1.5,t2=4,censorrate=0,a1=0.5)
Maximum_RR(alpha=0.05,beta=0.1,lambda_11=0.26,RR=0.26/0.4,phi_1=0.19/0.26,
phi_all=0.26/0.4,t1=1.5,t2=4,censorrate=0,a1=0.5)
```

Bibliography

E.E.A.A. Aly, S.C. Kochar, and I.W. McKeague. Some tests for comparing cumulative incidence functions and cause-specific hazard rates. *Journal of the American Statistical Association*, 89(427):994–999, 1994.

Per Kragh Andersen. Testing goodness of fit of cox's regression and life model. *Biometrics*, pages 67–77, 1982.

P.K. Andersen, Ø. Borgan, R. Gill, and N. Keiding. Linear nonparametric tests for comparison of counting processes, with applications to censored survival data, correspondent paper. *International Statistical Review/Revue Internationale de Statistique*, pages 219–244, 1982.

P.K. Andersen, Ø. Borgan, R. Gill, and N. Keiding. *Statistical models based on counting processes*. Springer Verlag, 1993.

P.K. Andersen, R.B. Geskus, T. de Witte, and H. Putter. Competing risks in epidemiology: possibilities and pitfalls. *International journal of epidemiology*, 41(3):861–870, 2012.

Elja Arjas. A graphical method for assessing goodness of fit in cox's proportional hazards model. *Journal of the American Statistical Association*, 83(401):204–212, 1988.

R. Bajorunaite and J.P. Klein. Two-sample tests of the equality of two cumulative incidence functions. *Computational statistics & data analysis*, 51(9):4269–4281, 2007.

J. Beyersmann, M. Dettenkofer, H. Bertz, and M. Schumacher. A competing risks analysis of bloodstream infection after stem-cell transplantation using

subdistribution hazards and cause-specific hazards. *Statistics in medicine*, 26 (30):5360–5369, 2007.

J. Beyersmann, A. Latouche, A. Buchholz, and M. Schumacher. Simulating competing risks data in survival analysis. *Statistics in Medicine*, 28(6):956–971, 2009.

David R Cox. Partial likelihood. *Biometrika*, 62(2):269–276, 1975.

D.R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220, 1972. ISSN 0035-9246.

D.R. Cox and D. Oakes. *Analysis of survival data*, volume 21. Chapman & Hall/CRC, 1984.

Kevin Hasegawa Eng and Michael R Kosorok. A sample size formula for the supremum log-rank statistic. *Biometrics*, 61(1):86–91, 2005.

J. P. Fine and R. J. Gray. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94(446): 496–509, 1999.

JP Fine. Analysing competing risks data with transformation models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(4):817–830, 1999.

J.P. Fine. Regression modeling of competing crude failure probabilities. *Biostatistics*, 2(1):85–97, 2001.

T.R. Fleming and D.P. Harrington. *Counting processes and survival analysis*, volume 8. Wiley Online Library, 1991.

T.A. Gerds, T.H. Scheike, and P.K. Andersen. Absolute risk regression for competing risks: interpretation, link functions, and prediction. *Statistics in Medicine*, 2012.

A. Gichangi and W. Vach. The analysis of competing risks data: A guided tour. *Statistics in Medicine*, 2005.

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

J. Graunt. 1662. natural and political observations mentioned in a following index and made upon the bills of mortality. *London: John Martyn. Reprinted in The Economic Writings of Sir William Petty, Together with the Observations upon the Bills of Mortality, More Probably by Captain John Graunt*, 1899.

R.J. Gray. A class of K-sample tests for comparing the cumulative incidence of a competing risk. *The Annals of Statistics*, 16(3):1141–1154, 1988. ISSN 0090-5364.

B. Haller, G. Schmidt, and K. Ulm. Applying competing risks regression models: an overview. *Lifetime Data Analysis*, pages 1–26, 2012.

E. Halley. An estimate of the degrees of the mortality of mankind, drawn from curious tables of the births and funerals at the city of breslaw; with an attempt to ascertain the price of annuities upon lives. by mr. e. halley, rss. *Philosophical Transactions*, 17(192-206):596–610, 1942.

JD Holt. Competing risk analyses with special reference to matched pair experiments. *Biometrika*, 65(1):159–165, 1978.

E.L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958. ISSN 0162-1459.

J.P. Klein. Modelling competing risks in cancer studies. *Statistics in medicine*, 25(6):1015–1034, 2006.

J.P. Klein and P.K. Andersen. Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics*, 61 (1):223–229, 2005.

Michael Koch, Bernd Kutkuhn, Evi Trenkwalder, Dieter Bach, Bernd Grabensee, Hans Dieplinger, and Florian Kronenberg. Apolipoprotein b, fibrinogen, hdl cholesterol, and apolipoprotein (a) phenotypes predict coronary artery disease in hemodialysis patients. *Journal of the American Society of Nephrology*, 8 (12):1889–1898, 1997.

SB Kulathinal and D. Gasbarra. Testing equality of cause-specific hazard rates corresponding to m competing risks among k groups. *Lifetime Data Analysis*, 8(2):147–161, 2002.

J.M. Lachin. Introduction to sample size determination and power analysis for clinical trials. *Controlled Clinical Trials*, 2(2):93–113, 1981. ISSN 0197-2456.

J.M. Lachin and M.A. Foulkes. Evaluation of sample size and power for analyses of survival with allowance for nonuniform patient entry, losses to follow-up, noncompliance, and stratification. *Biometrics*, 42(3):507–519, 1986. ISSN 0006-341X.

S.W. Lagakos. A covariate model for partially censored data subject to competing causes of failure. *Applied Statistics*, pages 235–241, 1978.

SW Lagakos. The graphical evaluation of explanatory variables in proportional hazard regression models. *Biometrika*, 68(1):93–98, 1981.

E. Lakatos. Sample-size determination in clinical-trials with time-dependent rates of losses and noncompliance. *Controlled Clinical Trials*, 7(3):189–199, 1986.

E. Lakatos. Sample sizes based on the log-rank statistic in complex clinical trials. *Biometrics*, 44(1):229–241, 1988. ISSN 0006-341X.

E. Lakatos and KK Lan. A comparison of sample size methods for the logrank statistic. *Statistics in medicine*, 11(2):179–191, 1992. ISSN 1097-0258.

KF Lam. A class of tests for the equality of k cause-specific hazard rates in a competing risks model. *Biometrika*, 85(1):179–188, 1998.

M.G. Larson. Covariate analysis of competing-risks data with log-linear models. *Biometrics*, pages 459–469, 1984.

A. Latouche, R. Porcher, and S. Chevret. Sample size formula for proportional hazards modelling of competing risks. *Statistics in Medicine*, 23(21):3263–3274, 2004.

A. Latouche, V. Boisson, S. Chevret, and R. Porcher. Misspecified regression model for the subdistribution hazard of a competing risk. *Statistics in Medicine*, 26(5):965–974, 2007.

G. Li and Q. Yang. Joint Inference for Competing Risks Data. *Journal of American Statistical Association*, 2013.

H. Lindkvist and Y. Belyaev. A class of non-parametric tests in the competing risks model for comparing two samples. *Scandinavian journal of statistics*, 25 (1):143–150, 1998.

M. Lunn and D. McNeil. Applying cox regression to competing risks. *Biometrics*, pages 524–532, 1995.

X. Luo and B.W. Turnbull. Comparing two treatments with multiple competing risks endpoints. *Statistica Sinica*, 9(4):985–998, 1999.

T Moreau, J O'quigley, and M Mesbah. A global goodness-of-fit statistic for the proportional hazards model. *Applied Statistics*, pages 212–218, 1985.

NJD Nagelkerke, J Oosting, and AAM Hart. A simple test for goodness of fit of cox's proportional hazards model. *Biometrics*, pages 483–486, 1984.

M.S. Pepe and M. Mori. Kaplanmeier, marginal or conditional probability curves in summarizing competing risks failure time data? *Statistics in medicine*, 12 (8):737–751, 1993.

R. Peto and J. Peto. Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society. Series A (General)*, 135(2):185–207, 1972. ISSN 0035-9238.

M. Pintilie. *Competing risks: a practical perspective*. John Wiley & Sons New York:, 2006.

RL Prentice, J.D. Kalbfleisch, A.V. Peterson Jr, N. Flournoy, VT Farewell, and NE Breslow. The analysis of failure times in the presence of competing risks. *Biometrics*, 34(4):541–554, 1978. ISSN 0006-341X.

H. Putter, M. Fiocco, and RB Geskus. Tutorial in biostatistics: competing risks and multi-state models. *Statistics in medicine*, 26(11):2389–2430, 2007.

A. Sancho, A. Ávila, E. Gavela, S. Beltrán, JE Fernández-Nájera, P. Molina,

JF Crespo, and LM Pallardó. Effect of overweight on kidney transplantation outcome. 39(7):2202–2204, 2007.

T.H. Scheike and M.J. Zhang. Flexible competing risks regression modeling and goodness-of-fit. *Lifetime data analysis*, 14(4):464–483, 2008.

T.H. Scheike and M.J. Zhang. Analyzing competing risk data using the r timereg package. *Journal of statistical software*, 38(2), 2011.

C. Schmoor, W. Sauerbrei, and M. Schumacher. Sample size considerations for the evaluation of prognostic factors in survival analysis. *Statistics in medicine*, 19(4):441–452, 2000. ISSN 1097-0258.

D. Schoenfeld. The asymptotic properties of nonparametric-tests for comparing survival distributions. *Biometrika*, 68(1):316–319, 1981.

D. A. Schoenfeld. Sample-size formula for the proportional-hazards regression-model. *Biometrics*, 39(2):499–503, 1983.

David Schoenfeld. Chi-squared goodness-of-fit tests for the proportional hazards regression model. *Biometrika*, 67(1):145–153, 1980.

David Schoenfeld. Partial residuals for the proportional hazards regression model. *Biometrika*, 69(1):239–241, 1982.

G. Schulgen, M. Olschewski, V. Krane, C. Wanner, G. Ruf, and M. Schumacher. Sample sizes for clinical trials with time-to-event endpoints and competing risks. *Contemporary clinical trials*, 26(3):386–396, 2005. ISSN 1551-7144.

Y. Sun and R.C. Tiwari. Comparing cause-specific hazard rates of a competing risks model with censored data. *Lecture Notes-Monograph Series*, pages 255–270, 1995.

R. Tiwari, KB Kulasekera, and C. Park. Nonparametric tests for cause specific hazard rates with censored data for competing risks among several groups. *Journal of statistical planning and inference*, 136(5):1718–1745, 2006.

W Tschöpe, Michael Koch, Bernhard Thomas, and Eberhard Ritz. Serum lipids predict cardiac death in diabetic patients on maintenance hemodialysis. *Nephron*, 64(3):354–358, 2008.

A. Tsiatis. A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences*, 72(1):20–22, 1975.

Christoph Wanner, Vera Krane, Günther Ruf, Winfried März, and Eberhard Ritz. Rationale and design of a trial improving outcome of type 2 diabetics on hemodialysis. *Kidney International*, 56:S222–S226, 1999.

N. A. Yateman and A. M. Skene. Sample sizes for proportional hazards survival studies with arbitrary patient entry and loss to follow-up distributions. *Statistics in Medicine*, 11(8):1103–1113, 1992.