Title

Joint Inference for Competing Risks Data

Authors
Li, Gang
Yang, Qing

Publication Date
2015-09-04

Joint Inference for Competing Risks Survival Data

Gang Li
Department of Biostatistics, University of California, Los Angeles
and
Qing Yang *
School of Nursing, Duke University

Abstract

This paper 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 time to 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 given 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 regression problem. We also discuss joint inference for other related functions. Our simulation shows that our 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 method using a Hodgkin disease data and a lymphoma data. Supplementary materials for this article are available online.

Keywords: Cause-Specific Hazard; Censoring; Cumulative Incidence; Cox's Model; Log-rank Test; Sub-distribution Hazard.

*Gang Li is Professor, Department of Biostatistics, University of California, Los Angeles, Los Angeles, CA 90095 (e-mail: vli@ulca.edu). Qing Yang is Assistant Professor, School of Nursing, Duke University, Durham, NC 27710 (e-mail: qing.yang@duke.edu). Gang Li's work was partially supported by NIH grant 5P30CA-16042 and NIH grant 8UL1TR000124. The authors thank the co-editor, the associate editor, and the two anonymous referees for their valuable comments that helped improve this article significantly.

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 outcome variable is the waiting time to become eligible for transplantation. In this case, death before becoming eligible for 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 Prentice *et al.* (1978); Pintilie (2006); Gichangi and Vach (2005); Putter *et al.* (2007), and the references therein. There is a broad literature on statistical methods for competing risks failure time 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 failure time data, Prentice *et al.* (1978), Lagakos (1978), Holt (1978), Cox and Oakes (1984, chap.9), Larson (1984), and Lunn and McNeil (1995) studied proportional cause-specific hazards models. Fine and Gray (1999) introduced a proportional subdistribution hazards model for the cumulative incidence function. 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 the cause specific hazard and the cumulative incidence function. Comprehensive survey of statistical methods for competing risks survival data and further references can be found in Beyersmann *et al.* (2007); Latouche *et al.* (2007); Haller *et al.* (2012).

In this paper we focus on the problem of assessing the effects of a variable (treatment or covariate) on the time to a particular type of failure. For convenience, we assume hereafter that there are only two types of failure, where type 1 represents the failure type of interest and type 2 includes all other competing risks. As discussed earlier, there are mainly two approaches to this problem based on either the cause-specific hazard function or the cumulative incidence function. The cause-specific hazard function for type 1 failure is defined as

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

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 multiple failure 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 effects of a variable on the cause-specific hazard $\lambda_1(t)$ by treating other types of failures as independent right censoring events. The cumulative incidence function is defined as $F_1(t) = P(T \leq t, D = 1), \quad t > 0$, the cumulative incidence rate of type 1 failure by time t, which can be uniquely characterized by the following sub-distribution hazard:

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

In particular, Gray (1988) developed a class of nonparametric tests to compare the cumulative incidence function of a given type of failure between different groups and Fine and Gray (1999) introduced a proportional sub-distribution hazards model for the regression problem.

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 effects of a variable on $\lambda_1(t)$ can be different from its effects 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, no formal joint inference procedure for these quantities is available in the literature. Although the Bonferroni method provides a straightforward solution, it can be severely under-powered as demonstrated later in Sections 4 and 5.

The primary purpose of this paper 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 $F_1(t)$, 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-type models for $\lambda_1(t)$ and $F_1(t)$(or $\tilde{\lambda}_1(t)$). We also discuss joint inference for other related quantities.

In section 2 we first review the weighted log-rank test for group comparison of $\lambda_1(t)$ and the Gray (1988) test for $F_1(t)$. Then, we develop joint test procedures for group comparisons of both $\lambda_1(t)$ and $F_1(t)$. 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. Section 3 develops joint regression analysis methods for $\lambda_1(t)$ and $F_1(t)$(or $\tilde{\lambda}_1(t)$) under Cox-type regression models. Section 4 presents some simulation results to evaluate the proposed methods and compare them with the Bonferroni method. In section 5, we illustrate our methods on a Hodgekin disease data and a lymphoma data. Section 6 gives some further remarks. The proofs for the theorems and additional simulation results are provided in the Appendix in the supplementary material.

2 Two-Sample Joint Tests for Competing Risks Data

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 failure 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 failure time data $\{(X_{ik}, \delta_{ik}), i = 1, \ldots, n_k\}$, where $X_{ik} = \min(T_{ik}, C_{ik})$ and $\delta_{ik} = D_{ik} I(T_{ik} \le C_{ik})$. 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,

$$H_0 : \lambda_{11}(t) = \lambda_{12}(t) \quad \text{and} \quad F_{11}(t) = F_{12}(t) \quad \text{for all } 0 < t < \tau, \tag{1}$$

where τ is some pre-specified fixed time.

2.1 Preliminaries

We first review the two-sample weighted log-rank test for the cause-specific hazard and the Gray (1988) two-sample test for the cumulative incidence for type 1 failure. These tests will be used as building blocks to develop joint tests for the hypothesis (1).

Two-Sample Tests for Cause-Specific Hazard

It is now well known that the standard (weighted) log-rank test (Peto and Peto, 1972; Andersen *et al.*, 1982) for right censored failure time data can be applied to test

$$H_0 : \lambda_{11}(t) = \lambda_{12}(t) \quad \text{for all } 0 < t < \tau, \tag{2}$$

by treating all other competing risks as independent right censoring (Tsiatis, 1975; Prentice *et al.*, 1978; Lindkvist and Belyaev, 1998). Specifically, let $N_{jk}(t) = \sum_{i=1}^{n_k} I(X_{ki} \le t, D_{ki} = j)$ be the counting process of the number of observed type j failures in group k by time t, and $Y_k(t) = \sum_{i=1}^{n_k} I\{X_{ki} \ge t\}$ 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) is defined as

$$U_{1k} = \int_0^\tau W_1(t) Y_k(t) \left\{ \frac{dN_{1k}(t)}{Y_k(t)} - \frac{dN_{1\cdot}(t)}{Y_\cdot(t)} \right\}, \tag{3}$$

where $W_1(t)$ is a predictable weight function that converges in probability to some deterministic function $w_1(t)$ as $n \to \infty$, 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 (2), $n^{-1/2}U_{11}/\hat{\sigma}$ has a standard normal limiting distribution where

$$\hat{\sigma}^2 = \int_0^\tau W_1^2(t) \frac{Y_1(t)Y_2(t)}{Y.(t)} \frac{dN_1.(t)}{Y.(t)}. \tag{4}$$

This leads to an asymptotic χ^2 test or a Z test for (2).

Two-Sample Tests for Cumulative Incidence Function

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

$$H_0 : F_{11}(t) = F_{12}(t) \quad \text{for all } 0 < t < \tau. \tag{5}$$

The Gray (1988) nonparametric test statistic is defined as

$$\tilde{U}_{1k} = \int_0^{\tau_k} \tilde{W}(t) R_k(t) \left\{ \frac{dN_{1k}(t)}{R_k(t)} - \frac{dN_1.(t)}{R.(t)} \right\}, \tag{6}$$

where $\tilde{W}(t)$ is a predictable weight function that converges in probability to some deterministic function $\tilde{w}(t)$ as $n \to \infty$, $R_k(t) = I(\tau_k \geq t)Y_k(t)\hat{G}_{1k}(t-)/\hat{S}_k(t-)$ can be considered as an adjusted risk set size for group k at time t, $\hat{G}_{jk}(t-)$ is the the left-hand limit of the Kaplan-Meier (1958) estimate of $G_{jk}(t) = 1 - F_{jk}(t)$, $\hat{S}_k(t-)$ is the left-hand limit of the Kaplan-Meier estimate of $S_k(t)$, τ_k is some fixed time point satisfying $S_k(\tau_k)S_k^c(\tau_k) > 0$, and $R.(t)$ represents the same quantity as $R_k(t)$ using the pooled sample. Gray (1988) showed that under (5), $n^{-1/2}\tilde{U}_{11}/\hat{\tilde{\sigma}}$ has a standard normal limiting distribution, where

$$\hat{\tilde{\sigma}}^2 = \sum_{k=1}^{2} n^{-1} \left\{ \int_0^{\tau_1} \hat{a}_k^2(t)\hat{h}_k^{-1}(t)\hat{h}_.^{-1}(t)dN_1.(t) + \int_0^{\tau_1} \hat{b}_{2k}^2(t)\hat{h}_k^{-2}(t)dN_{2k}(t) \right\}, \tag{7}$$

with

$$
\begin{aligned}
\hat{a}_k(t) &= \hat{d}_{1k}(t) + \hat{b}_{1k}(t), \\
\hat{b}_{jk}(t) &= \left[I(j=1) - \hat{G}_{1\cdot}(t)/\hat{S}_k(t)\right]\left[\hat{c}_k(\tau_1) - \hat{c}_k(t)\right], \\
\hat{c}_k(t) &= \int_0^t \hat{d}_{1k}(u)\hat{G}_{1\cdot}(u-)^{-1}\hat{h}_\cdot^{-1}(u)dN_{1\cdot}(u), \\
\hat{d}_{jk}(t) &= n^{-1}I(j=1)\tilde{W}(t)R_1(t)\left[I(k=1) - \hat{h}_k(t)/\hat{h}.(t)\right]/\hat{G}_{1\cdot}(t-), \qquad (8) \\
\hat{h}_k(t) &= I(t \le \tau_k)n^{-1}Y_k(t)/\hat{S}_k(t-), \\
\hat{h}.(t) &= I(t \le max(\tau_1,\tau_2))n^{-1}Y.(t)/\hat{S}.(t-), \\
\hat{G}_{1\cdot}(t) &= 1 - \hat{F}_{1\cdot}(t) = 1 - n^{-1}\int_0^t \hat{h}_\cdot^{-1}(u)dN_{1\cdot}(u).
\end{aligned}
$$

This gives an asymptotic χ^2 test for (5) based on $n^{-1}\tilde{U}_{11}^2/\hat{\tilde{\sigma}}^2$ or a Z test based on $n^{-1/2}\tilde{U}_{11}/\hat{\tilde{\sigma}}$.

Examples of the weight functions in the above test statistics have been discussed by a number of authors (Gray, 1988; Gehan, 1965; Breslow, 1970; Peto and Peto, 1972; Kalbfleisch, 1980). A nice survey of various weight functions and their applications can be found in Klein and Moeschberger (2003, chap 7.2).

2.2 Joint Two-Sample Tests for Cause-Specific Hazard and Cumulative Incidence Function

To test the joint null hypothesis (1), we first 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 (3) and (6). Under the null hypothesis (1), $n^{-1/2}(U_{11}, \tilde{U}_{11})$ has an asymptotically bivariate normal distribution with mean $\mathbf{0}$ and variance-coviance matrix $\mathbf{\Sigma}^{(1)} = (\sigma_{ij}^{(1)})$ as $n \to \infty$, where $\mathbf{\Sigma}^{(1)}$ is defined in (A.1) and (A.4) of Appendix A.1. Furthermore, $\sigma_{11}^{(1)}$ and $\sigma_{22}^{(1)}$ are consistently estimated by (4) and (7), and* *the covariance* $\sigma_{12}^{(1)}$ *is consistently estimated by*

$$\begin{aligned}\hat{\sigma}_{12}^{(1)} &= n^{-1}\left\{\int_0^\tau W_1(t)\frac{Y_2(t)}{Y_\cdot(t)}\hat{V}_{11}(t) + \hat{c}_1(\tau)\int_0^\tau W_1(t)\frac{Y_2(t)}{Y_\cdot(t)}\hat{E}_{11}(t)\hat{h}_1^{-1}(t)\right\} Y_1(t)d\hat{\Lambda}_{11}(t) \\ &+ n^{-1}\left\{\int_0^\tau W_1(t)\frac{Y_1(t)}{Y_\cdot(t)}\hat{V}_{12}(t) + \hat{c}_2(\tau)\int_0^\tau W_1(t)\frac{Y_1(t)}{Y_\cdot(t)}\hat{E}_{12}(t)\hat{h}_2^{-1}(t)\right\} Y_2(t)d\hat{\Lambda}_{12}(t),\end{aligned} \quad (9)$$

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 other quantities are defined in (8).*

Chi-square Joint Test for (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 (1), X^2 has an asymptotically chi-square distribution with 2 degrees of freedom. This leads to the following chi-square test for (1):

$$\text{Reject (1) at level } \alpha \text{ if } X^2 > \chi_2^2(\alpha),$$

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

Rejection of (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.

Maximum Joint Test for (1)

Define

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

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 (1) if the observed T^* 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 T^* 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 denote it by T_l^*. Let T_α be the upper $100(1-\alpha)$-th sample quantile of $T_1^*, \ldots, T_N^*$. Reject the null hypothesis (1) at level α if $T^* > T_\alpha$.

Remark 2.1: 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.

Remark 2.2: (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

$$H_0 : \lambda_{11}(t) = \cdots = \lambda_{1K}(t) \quad \text{and} \quad F_{11}(t) = \cdots = F_{1K}(t) \quad \text{for all } 0 < t < \tau, \tag{10}$$

where τ is some pre-specified fixed time. Similar to Theorem 1, it can be shown that under the null hypothesis (10), $V_n = n^{-1/2}(U_{11}, \cdots, U_{1K-1}, \tilde{U}_{11}, \cdots, \tilde{U}_{1K-1})$ has an asymptotic multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix Σ^*, where Σ^* is defined as the limit of the variance-covariance matrix of V_n and can be consistently estimated as follows. From Kulathinal and Gasbarra (2002), we have $\widehat{Cov}(n^{-1/2}U_{1k}, n^{-1/2}U_{1k'}) = -\int_0^\tau W_1^2(t)\frac{Y_k(t)Y_{k'}(t)}{Y.(t)}d\hat{\Lambda}_1(t)$, where $k, k' = 1, \cdots, K$. $\widehat{Cov}(n^{-1/2}\tilde{U}_{1k}, n^{-1/2}\tilde{U}_{1k'})$ is given by equation (2.10) on page 1146 of Gray (1988). Similar to the proof of Theorem 1,

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

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

$\hat{E}_{jk}(t) = I(j = 1) - \hat{G}_{1k}(t-)/\hat{S}_k(t-)$, and all other quantities are defined in (8). These results allow one to derive a chi-square test and a maximal test similar to the two-sample case.

2.3 Joint Two Sample Tests for Other Quantities

Joint tests can also be derived for other related quantities. For group k, let $\lambda_{2k}(t)$ and $\lambda_{\cdot k}(t)$ denote the other (type 2) cause-specific hazard function and the all-cause hazard function, respectively.

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

Consider the following null hypotheses

$$H_0 : \lambda_{11}(t) = \lambda_{12}(t) \quad \text{and} \quad \lambda_{\cdot 1}(t) = \lambda_{\cdot 2}(t) \quad \text{for all } 0 < t < \tau. \tag{11}$$

Let

$$U_{\cdot k} = \int_0^\tau W_{\cdot}(t) Y_k(t) \left\{ \frac{dN_{\cdot k}(t)}{Y_k(t)} - \frac{dN_{\cdot\cdot}(t)}{Y_{\cdot}(t)} \right\}, \tag{12}$$

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 that converges in probability to some deterministic function $w_{\cdot}(t)$ as $n \to \infty$. Let U_{11} and $U_{\cdot 1}$ be defined by (3) and (12). Then, $n^{-1/2}(U_{11}, U_{\cdot 1})$ has an asymptotic bivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $\mathbf{\Sigma}^{(2)} = (\sigma_{ij}^{(2)})$. Furthermore, $\mathbf{\Sigma}^{(2)}$ is consistently estimated by $\hat{\mathbf{\Sigma}}^{(2)} = (\hat{\sigma}_{ij}^{(2)})$ where $\hat{\sigma}_{11}^{(2)} = \int_0^\tau W_1^2(t) \frac{Y_1(t)Y_2(t)}{Y_1(t)+Y_2(t)} \frac{dN_{11}(t)}{Y_1(t)}$, $\hat{\sigma}_{22}^{(2)} = \int_0^\tau W_{\cdot}^2(t) \frac{Y_1(t)Y_2(t)}{Y_1(t)+Y_2(t)} \frac{dN_{\cdot 1}(t)}{Y_1(t)}$, and $\hat{\sigma}_{12}^{(2)} = \int_0^\tau W_1(t) W_{\cdot}(t) \frac{Y_1(t)Y_2(t)}{Y_1(t)+Y_2(t)} \frac{dN_{11}(t)}{Y_1(t)}$. These results allows one to construct a chi-square joint test and a maximum joint test for (11) similar to those for (1) in the previous section.

Two-Sample Joint Tests for Both Cause-Specific Hazards

Consider

$$H_0 : \lambda_{11}(t) = \lambda_{12}(t) \quad \text{and} \quad \lambda_{21}(t) = \lambda_{22}(t) \quad \text{for all } 0 < t < \tau. \tag{13}$$

Let

$$U_{2k} = \int_0^\tau W_2(t) Y_k(t) \left\{ \frac{dN_{2k}(t)}{Y_k(t)} - \frac{dN_{2\cdot}(t)}{Y_\cdot(t)} \right\}, \tag{14}$$

be the weighted log-rank test statistic for $H_0 : \lambda_{21}(t) = \lambda_{22}(t)$ for all $0 < t < \tau$, where $W_2(t)$ is a predictable weight function that 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 a chi-square joint test and a maximum joint test for (13) based on the joint distribution of the two test statistics. Joint test for (13) was also studied previously by Lindkvist and Belyaev (1998) and Kulathinal and Gasbarra (2002) among others. In particular, the K-sample chi-square test of Kulathinal and Gasbarra (2002, page 150) for the $(\lambda_{1k}, \lambda_{2k})$ pair with a special weight function $K^n_{kij}(t) = I(i = j)W_1(t)$ reduces to that based on U_{1k} and U_{2k}. We also note that the ideas of Kulathinal and Gasbarra (2002) could be extended to derive a test for the $(\lambda_{1k}, \lambda_{\cdot k})$ pair, although it was not explicitly developed in their paper.

Remark 2.3: It can be 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 joint distribution of (X_{ik}, δ_{ik}). This implies that the three null hypotheses (1), (11), and (13) are equivalent. On the other hand, their alternative hypotheses are different because the three pairs of functions characterize different features of competing risks data. Furthermore, a significant effect of a variable on one pair does not necessarily imply a significant effect on another pair, as illustrated later in Section 5.1. A practical question is which pair(s) should be used, especially when planning a study. The answer would depend on the specific research questions of a study. The cause-specific hazard and cumulative incidence pair, or $(\lambda_{1k}(\cdot), F_{1k}(\cdot))$, would be useful when studying the effects of a variable on a given type (type 1) failure since they directly characterize two distinct and easily interpretable features of type 1 failure. The cause-specific hazard and all-cause hazard pair, or $(\lambda_{1k}(\cdot), \lambda_{\cdot k}(\cdot))$, would be useful when the all-cause hazard describes a meaningful clinical outcome such as "overall survival" (death due to any disease) in a randomized clinical trial of a new treatment versus a standard treatment for a specific disease in which the disease-specific survival and overall survival are co-primary endpoints. Note that the all-cause hazard may not always describe a meaningful clinical outcome especially when the two types of failures are negatively correlated as exemplified in the kidney transplantation program example discussed in the beginning of Section 1. Finally, joint inference for both cause-specific hazards, or $(\lambda_{1k}(\cdot), \lambda_{2k}(\cdot))$, would useful when both types of failures are of interest to the study.

3 Joint Regression Analysis for Competing Risks Data

3.1 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 $1, 2$ 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 $pr(C_i \geq t) = G^c(t)$.

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)

$$\lambda_1(t|\mathbf{Z}) = \lambda_{10}(t)\exp(\boldsymbol{\beta}_1^T \mathbf{Z}^{(1)}(t)), \tag{15}$$

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

$$\tilde{\lambda}_1(t|\mathbf{Z}) = \tilde{\lambda}_{10}(t)\exp(\boldsymbol{\gamma}_1^T \mathbf{Z}^{(2)}(t)), \tag{16}$$

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 (15) 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 (16) 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

$$H_0 : A_1^T \boldsymbol{\beta}_1 = \mathbf{d}_1 \text{ and } A_2^T \boldsymbol{\gamma}_1 = \mathbf{d}_2, \tag{17}$$

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

$$\mathbf{U}_1(\boldsymbol{\beta}_1) = \sum_{i=1}^{n} \int_0^\infty \left\{ \mathbf{Z}_i^{(1)}(t) - \bar{\mathbf{Z}}^{(1)}(\boldsymbol{\beta}_1, t) \right\} dN_{i1}(t), \tag{18}$$

and

$$\tilde{\mathbf{U}}_1(\boldsymbol{\gamma}_1) = \sum_{i=1}^{n} \int_0^\infty \left\{ \mathbf{Z}_i^{(2)}(t) - \bar{\mathbf{Z}}^{(2)}(\boldsymbol{\gamma}_1, t) \right\} \omega_i(t) d\tilde{N}_{i1}(t), \tag{19}$$

be the score functions for $\boldsymbol{\beta}_1$ and $\boldsymbol{\gamma}_1$ under models (15) and (16), 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 2 *Under similar regularity conditions to Andersen* et al. *(1982) and Fine and Gray (1999), we have*

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

where $\boldsymbol{\Sigma}^{(1)}$ *is defined by (A.11) in the Appendix A.1. Furthermore,* $\boldsymbol{\Sigma}^{(1)}$ *can be consistently estimated by*

$$\hat{\boldsymbol{\Sigma}}^{(1)} = \begin{pmatrix} \hat{\boldsymbol{\Omega}}_{(pp)}^{(1)-1} & \hat{\boldsymbol{\Omega}}_{(pp)}^{(1)-1}\hat{\boldsymbol{\Omega}}_{(pq)}^{(1)}\hat{\boldsymbol{\Omega}}_{(qq)}^{(1)-1} \\ \hat{\boldsymbol{\Omega}}_{(qq)}^{(1)-1}\hat{\boldsymbol{\Omega}}_{(qp)}^{(1)}\hat{\boldsymbol{\Omega}}_{(pp)}^{(1)-1} & \hat{\boldsymbol{\Omega}}_{(qq)}^{(1)-1}\hat{\boldsymbol{\Omega}}_{(qq)}^{*(1)}\hat{\boldsymbol{\Omega}}_{(qq)}^{(1)-1} \end{pmatrix}, \tag{20}$$

where

$$\begin{aligned}
\hat{\boldsymbol{\Omega}}_{(pp)}^{(1)} &= \frac{1}{n}\sum_{i=1}^n \int_0^\infty \left[\frac{\sum_{l=1}^n Y_l(t)\mathbf{Z}_l^{(1)}(t)^{\otimes 2}\exp(\hat{\boldsymbol{\beta}}_1^T\mathbf{Z}_l^{(1)}(t))}{\sum_{l=1}^n Y_l(t)\exp(\hat{\boldsymbol{\beta}}_1^T\mathbf{Z}_l^{(1)}(t))} - \bar{\mathbf{Z}}^{(1)}(\hat{\boldsymbol{\beta}}_1, t)^{\otimes 2}\right] dN_{i1}(t), \\
\hat{\boldsymbol{\Omega}}_{(qq)}^{(1)} &= \frac{1}{n}\sum_{i=1}^n \left\{\frac{\sum_{l=1}^n \omega_l(t)\tilde{Y}_l(t)\mathbf{Z}_l^{(2)}(t)^{\otimes 2}\exp(\hat{\boldsymbol{\gamma}}_1^T\mathbf{Z}_l^{(2)}(t))}{\sum_{l=1}^n \omega_l(t)\tilde{Y}_l(t)\exp(\hat{\boldsymbol{\gamma}}_1^T\mathbf{Z}_l^{(2)}(t))} - \bar{\mathbf{Z}}^{(2)}(\hat{\boldsymbol{\gamma}}_1, t)^{\otimes 2}\right\} I(\delta_i = 1), \\
\hat{\boldsymbol{\Omega}}_{(pq)}^{(1)} &= \frac{1}{n}\sum_{i=1}^n \left\{\int_0^\infty \left(\mathbf{Z}_i^{(1)}(t) - \bar{\mathbf{Z}}^{(1)}(\hat{\boldsymbol{\beta}}_1, t)\right)\left(dN_{i1}(t) - Y_i(t)\exp(\hat{\boldsymbol{\beta}}_1^T\mathbf{Z}_i^{(1)}(t)d\hat{\Lambda}_{10}(t)\right) * \hat{\boldsymbol{\eta}}_i\right\} \\
&+ \frac{1}{n}\sum_{i=1}^n \left\{\int_0^\infty \left(\mathbf{Z}_i^{(1)}(t) - \bar{\mathbf{Z}}^{(1)}(\hat{\boldsymbol{\beta}}_1, t)\right)\left(dN_{i1}(t) - Y_i(t)\exp(\hat{\boldsymbol{\beta}}_1^T\mathbf{Z}_i^{(1)}(t))d\hat{\Lambda}_{10}(t)\right) * \hat{\boldsymbol{\phi}}_i\right\} \\
\hat{\boldsymbol{\Omega}}_{(qq)}^{*(1)} &= \frac{1}{n}\sum_{i=1}^n (\hat{\boldsymbol{\eta}}_i + \hat{\boldsymbol{\phi}}_i)^{\otimes 2},
\end{aligned} \tag{21}$$

with

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

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 (17), we have*

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

where

$$
\mathbf{V} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 \end{pmatrix} \boldsymbol{\Sigma}^{(1)} \begin{pmatrix} \mathbf{A}_1^T & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2^T \end{pmatrix}. \tag{22}
$$

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 (22). It follows immediately from Corollary 1 that under (17), 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 (17):

$$\text{Reject (17) at level } \alpha \text{ if } X_W^2 > \chi^2_{p_{d1}+p_{d2}}(\alpha),$$

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 Joint Regression Analysis of 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 as discussed in Section 2.3 (Remark 2.3).

Joint Regression Analysis of Cause-Specific Hazard and All-Cause Hazard
Assume that the proportional cause-specific hazards model (15) holds. In addition, assume the proportional all-cause hazards model:

$$\lambda(t|\mathbf{Z}) = \lambda_0(t)\exp(\boldsymbol{\beta}_.^T \mathbf{Z}^{(3)}(t)), \tag{23}$$

where $\lambda(t|\mathbf{Z}))$ denote the conditional all-cause hazard function given $\mathbf{Z}$, $\lambda_0(t)$ is an unknown baseline 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 $\boldsymbol{\beta}_.$.

Let

$$\mathbf{U}_.(\boldsymbol{\beta}_.) = \sum_{i=1}^{n} \int_0^\infty \left\{ \mathbf{Z}_i^{(3)}(t) - \bar{\mathbf{Z}}^{(3)}(\boldsymbol{\beta}_., t) \right\} dN_i(t), \tag{24}$$

be the score function for $\boldsymbol{\beta}_.$ under model (23), where $\bar{\mathbf{Z}}^{(3)}(\boldsymbol{\beta}_., t) = \frac{\sum_{l=1}^n Y_l(t)\mathbf{Z}_l^{(3)}(t)\exp(\boldsymbol{\beta}_.^T \mathbf{Z}_l^{(3)}(t))}{\sum_{l=1}^n Y_l(t)\exp(\boldsymbol{\beta}_.^T \mathbf{Z}_l^{(3)}(t))}$ and $N_i(t) = I(X_i \le t, \delta_i = 1)$. Let $\hat{\boldsymbol{\beta}}_.$ be the solution of the score equation $\mathbf{U}_.(\boldsymbol{\beta}_.) = 0$.

Theorem 3 *Under some regularity conditions, as $n \to \infty$,*

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

where $\mathbf{\Sigma}^{(2)}$ *is defined by (A.13) in the Appendix A.1. Furthermore,* $\mathbf{\Sigma}^{(2)}$ *can be consistently estimated by*

$$\hat{\mathbf{\Sigma}}^{(2)} = \begin{pmatrix} \hat{\mathbf{\Omega}}^{(2)-1}_{(pp)} & \hat{\mathbf{\Omega}}^{(2)-1}_{(pp)}\hat{\mathbf{\Omega}}^{(2)}_{(pq)}\hat{\mathbf{\Omega}}^{(2)-1}_{(qq)} \\ \hat{\mathbf{\Omega}}^{(2)-1}_{(qq)}\hat{\mathbf{\Omega}}^{(2)}_{(qp)}\hat{\mathbf{\Omega}}^{(2)-1}_{(pp)} & \hat{\mathbf{\Omega}}^{(2)-1}_{(qq)} \end{pmatrix}, \tag{25}$$

where

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

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.*

Theorem 3 enables one to draw joint inference for $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_.$ along the lines of the previous section.

Joint Regression Analysis of Both Cause-Specific Hazards

Assume the proportional cause-specific hazards model (15) for type 1 failure. In addition, assume the following proportional cause-specific hazards model for type 2 failure:

$$\lambda_2(t|\mathbf{Z}) = \lambda_{20}(t)\exp(\boldsymbol{\beta}_2^T \mathbf{Z}^{(4)}(t)), \tag{26}$$

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

$$\mathbf{U}_2(\boldsymbol{\beta}_2) = \sum_{i=1}^n \int_0^\infty \left\{\mathbf{Z}_i^{(4)}(t) - \bar{\mathbf{Z}}^{(4)}(\boldsymbol{\beta}_2, t)\right\} dN_{i2}(t), \tag{27}$$

be the score test statistic under model (26), 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 $\mathbf{U}_2(\boldsymbol{\beta}_2) = 0$. It can be shown that U_1 and

U_2 are asymptotically independent since $N_{i1}(t)$ and $N_{i2}(t)$ do not jump at the same time. Therefore one draw joint inference for $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ similar to the previous sections.

Remark 3.1. 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 3.2. (Model Checking) Model diagnostic techniques for the standard Cox (1972) proportional hazards model can be readily applied to assess model assumptions of the individual models (15), (23) and (26) (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 (16). Formal goodness-of-fit tests for (16) have been developed by Scheike and Zhang (2008). In addition to assessing goodness-of-fit of an individual model, it is also important to check if two individual models hold simultaneously. For example, it has been well recognized that the proportional hazards assumption for a time-independent covariate does not hold simultaneously for the cause-specific hazard and the cause-specific subdistribution hazard, and thus it is important for models (15) and (16) to allow time $\times$ covariates interactions. To check if (15) and (16) hold simultaneously, one needs to verify that for any $\mathbf{z}$, $\Lambda_2(t|\mathbf{z}) \equiv \tilde{\Lambda}_1(t|\mathbf{z}) - \Lambda_1(t|\mathbf{z}) + log\lambda_1(t|\mathbf{z}) - log\tilde{\lambda}_1(t|\mathbf{z})$ is nondecreasing and satisfies $\Lambda_2(0|\mathbf{z}) = 0$. In other words, the above defined $\Lambda_2(t|\mathbf{z})$ is a proper conditional cumulative cause-specific hazard function for type 2 failure. We provide an example of the joint model of (15) and (16) in Section 4(model (28)).

4 Simulations

We present some simulation results to illustrate the advantage of the proposed joint tests over the Bonferroni method. The weight function is set to be a constant 1 in all simulations.

The first simulation considers two-group comparison of type 1 failure with respect to both cause specific hazard (CSH) and cumulative incidence function (CIF). We assign equal number of patients in the two groups. Competing risks data are generated using Beyersmann *et al.* (2009)'s cause-specific hazard driven method that requires only specification of the cause-specific hazard for each type of failure.

Figure 1 depicts simulated rejection power of the two-sided chi-square joint test, maximum joint test and Bonferroni joint test for hypothesis (1) for various sample sizes per group under four scenarios. Figure 1(a) corresponds to a null case under H_0. Figure 1(b) corresponds to a scenario where there is a small group difference in CSH and a large group difference in CIF, whereas Figure 1(c) corresponds to an opposite situation. Figure 1(d) corresponds to a case where the group effects on CSH and CIF are similar. Specifically, in the first two scenarios, we assume constant cause specific hazard for both causes, with $\lambda_{11} = \lambda_{12} = 0.04, \lambda_{21} = \lambda_{22} = 0.01$ for Figure 1(a) and $\lambda_{11} = \lambda_{12} = 0.1, \lambda_{21} = 0.04, \lambda_{22} = 0.01$ for Figure 1(b), where λ_{jk} denotes the cause-specific hazard for type j failure in group k. In the last two scenarios, we assume $\lambda_1(t|\mathbf{Z}) = \lambda_{10}(t)\exp(\gamma Z * I(t < 1) + \beta Z * I(t \geq 1))$ and $\tilde{\lambda}_1(t|\mathbf{Z}) = \tilde{\lambda}_{10}(t)\exp(\gamma Z)$, with $\beta = 0.4, \gamma = 0.01$ for Figure 1(c) and $\beta = 0.5, \gamma = 0.5$ for Figure 1(d), where $\lambda_{10}(t) = 0.05 * I(0 \leq t < 1) + 0.1 * I(t \geq 1)$, $\tilde{\lambda}_{10}(t) = \frac{0.05e^{-t}}{1-0.05(1-e^{-t})}$, and Z is a binary group variable. The censoring rate is set to be 0.1 with an independent exponential censoring time in each scenario. The nominal significance level is 0.05. A graphical illustration of the CIF by groups under all four scenarios is presented in the Appendix A.3 (Figure A.5).

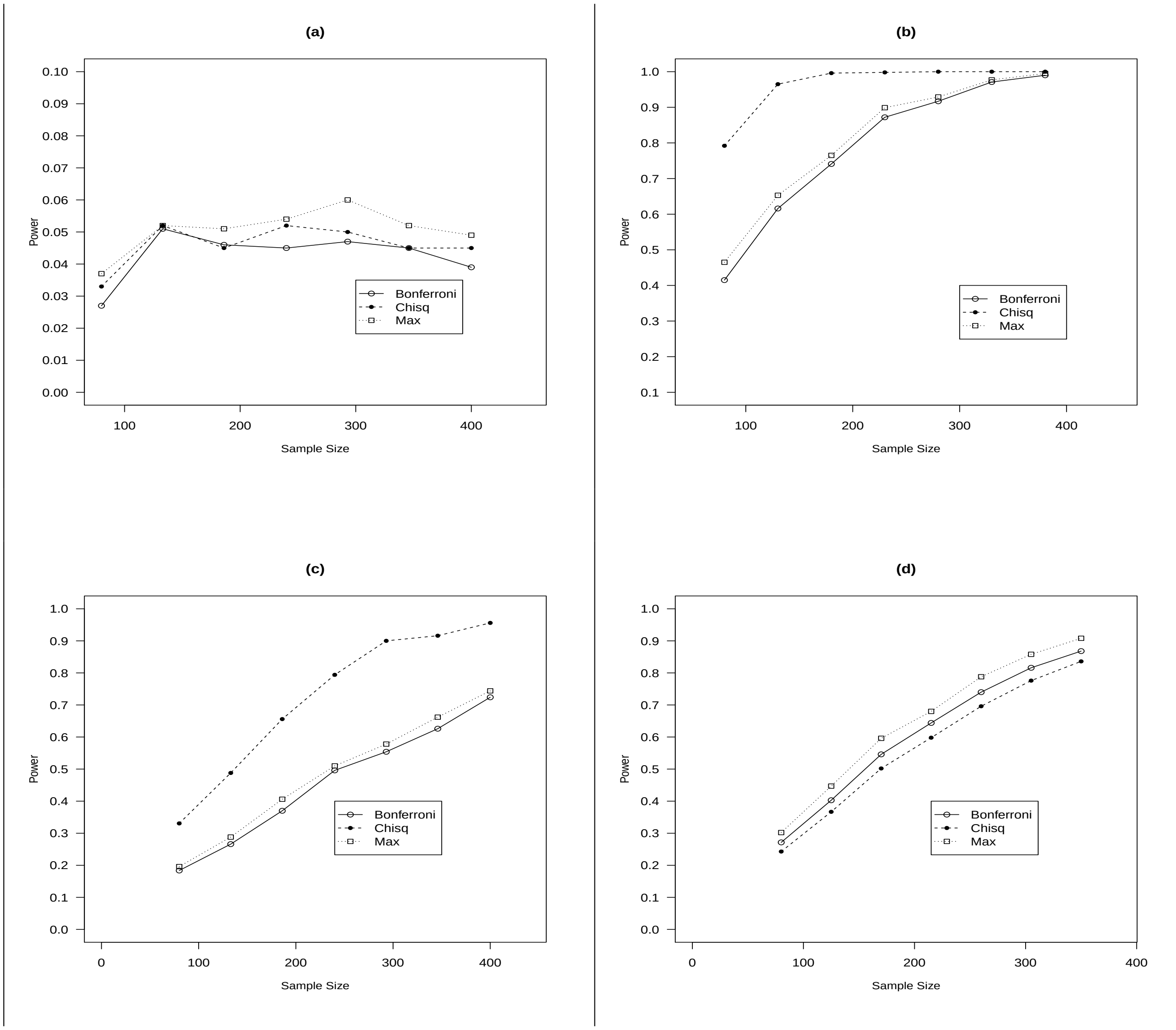

Figure 1: Simulated power of the two-sided chi-square joint test, maximum joint test and Bonferroni joint test for two-group type 1 failure comparison with respect to the CSH and CIF pair under four scenarios as described in Section 4: (a) null case under H_0, (b) smaller group difference in CSH and larger group difference in CIF, (b) larger group difference in CSH and smaller group difference in CIF, and (d) similar group effects on CSH and CIF.

It is seen from Figure 1(a) that the type I error rates for all three tests are well controlled around the 0.05 nominal level. In all three alternative cases ((b)-(d)), either the chi-square joint test, or the maximum joint test, or both are more powerful than the Bonferroni method. In the cases where the group effects on CSH and CIF are quite different (Figures 1(b) and 1(c)), the chi-square joint test is observed to be most powerful with subtaintially improved power. When the effect sizes for CSH and CIF are similar (Figure 1(d)), the maximum joint test outperforms the others. The improved power of the proposed joint tests has important implications for the design of clinical trials in the presence of competing risks. For example, to achieve 80% power under the second scenario (Figure 1(b)), it would require $n = 80$ patients for the chi-square joint test, about 200 patients for the maximum joint test, and more than 200 patients for the Bonferroni joint test.

We also conducted power comparisons for one-sided joint tests under the same four scenarios as in Figure 1. The results are presented in Figure A.1 in the Appendix A.2. The results are consistent with the two-sided case except that the maximum joint test has much more pronounced improvement over the chi-square joint test in the last scenario. We note that the chi-square joint test is constructed for a two-sided hypothesis, and thus can be underpowered when used as a one-sided test as shown in Figure A.1(d).

The second simulation study considers a joint regression model of CSH and CIF with respect to type 1 failure. It is well known that the proportional hazards assumption for a time-independent covariate usually does not hold simultaneously for the CSH and the CIF hazard (or subdistribution hazard), so it's imperative to include time by covariate interactions in the joint model. As a illustration, we consider the following joint model:

$$\begin{aligned} \lambda_1(t|\mathbf{Z}) &= \lambda_{10}(t)\exp(\boldsymbol{\gamma}^T\mathbf{Z} * I(t < \tau_0) + \boldsymbol{\beta}^T\mathbf{Z} * I(t \geq \tau_0)), \\ \tilde{\lambda}_1(t|\mathbf{Z}) &= \tilde{\lambda}_{10}(t)\exp(\tilde{\boldsymbol{\gamma}}^T\mathbf{Z} * I(t < \tau_0) + \tilde{\boldsymbol{\beta}}^T\mathbf{Z} * I(t \geq \tau_0)), \end{aligned} \tag{28}$$

where $\lambda_{10}(t) = aI(0 \leq t < \tau_0) + bI(t \geq \tau_0)$, $\tilde{\lambda}_{10}(t) = \frac{ce^{-t}}{1-c(1-e^{-t})}$, $\mathbf{Z} = (Z_1, Z_2)$, with Z_1, Z_2

being binary variables, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)$, $\boldsymbol{\beta} = (\beta_1, \beta_2)$, and τ_0 is some pre-specified constant. Note that under model (28), the conditional cumulative cause-specific hazard function for cause 2 given $\mathbf{Z} = \mathbf{z}$ is $\Lambda_2(t|\mathbf{z}) = \tilde{\Lambda}_1(t|\mathbf{z}) - \Lambda_1(t|\mathbf{z}) + log\lambda_1(t|\mathbf{z}) - log\tilde{\lambda}_1(t|\mathbf{z})$. For $\Lambda_2(t|\mathbf{z})$ to be a proper conditional cumulative cause-specific hazard function, it must satisfy

$$\Lambda_2(0|\mathbf{z}) = 0 \quad \text{and} \quad \lambda_2(t|\mathbf{z}) = \frac{\partial \Lambda_2(t|\mathbf{z})}{\partial t} \geq 0 \quad \text{for all } t \geq 0,$$

which imply some constraints on the parameters in model (28). For simplicity, we further assume $\tilde{\boldsymbol{\gamma}} = \tilde{\boldsymbol{\beta}}$ for our simulation. In this case, it can be shown that $\Lambda_2(t|\mathbf{z})$ is a proper cumulative cause-specific hazard function if the following constraints hold: 1) $a = c \leq b$, 2) $e^{\boldsymbol{\gamma}^T \mathbf{z}} < \frac{1-a}{c}$, 3) $e^{\boldsymbol{\beta}^T \mathbf{z}} < 1/a(1 - e^{-\tau_0})$, and 4) $\tilde{\gamma} = \gamma$. We then generated competing risks data from $\lambda_1(t|\mathbf{z})$ and $\lambda_2(t|\mathbf{z})$ using the method of Beyersmann *et al.* (2009).

Figure 2 displays the simulated power curves of the three two-sided joint tests described in Section 3.1 for the following local hypothesis regarding the effects of Z_1 on the CSH and the CIF hazard after time τ_0:

$$H_0 : \beta_1 = 0 \text{ and } \gamma_1 = 0. \tag{29}$$

We consider four scenarios: (a) the null case ($\beta_1 = 0, \gamma_1 = 0$); (b) smaller Z_1 effect on CSH and larger Z_1 effect on CIF ($\beta_1 = -0.1$,$\gamma_1 = -0.4$); (c) larger Z_1 effect on CSH and smaller Z_1 effect on CIF ($\beta_1 = -0.6$,$\gamma_1 = -0.2$); and (d) similar Z_1 effects on CSH and CIF ($\beta_1 = -0.5$,$\gamma_1 = -0.5$). In all four scenarios, we set $a = 0.05$, $b = 0.1$, $\beta_2 = -0.2$, $\gamma_2 = -0.1$, $\boldsymbol{\gamma} = \tilde{\boldsymbol{\gamma}}$, and $\tau_0 = 1$.

Figure 2 leads to similar conclusions to what we have observed for the two-group case in the first simulation study. In the supplementary material, we also present some simulations for the CSH and all-cause hazard (ACH) pair which have similar conclusions.

Finally, we conducted a small-scale simulation to compare the power of the three joint tests for (1), (11), and (13). When there is little group difference in a particular quantity,

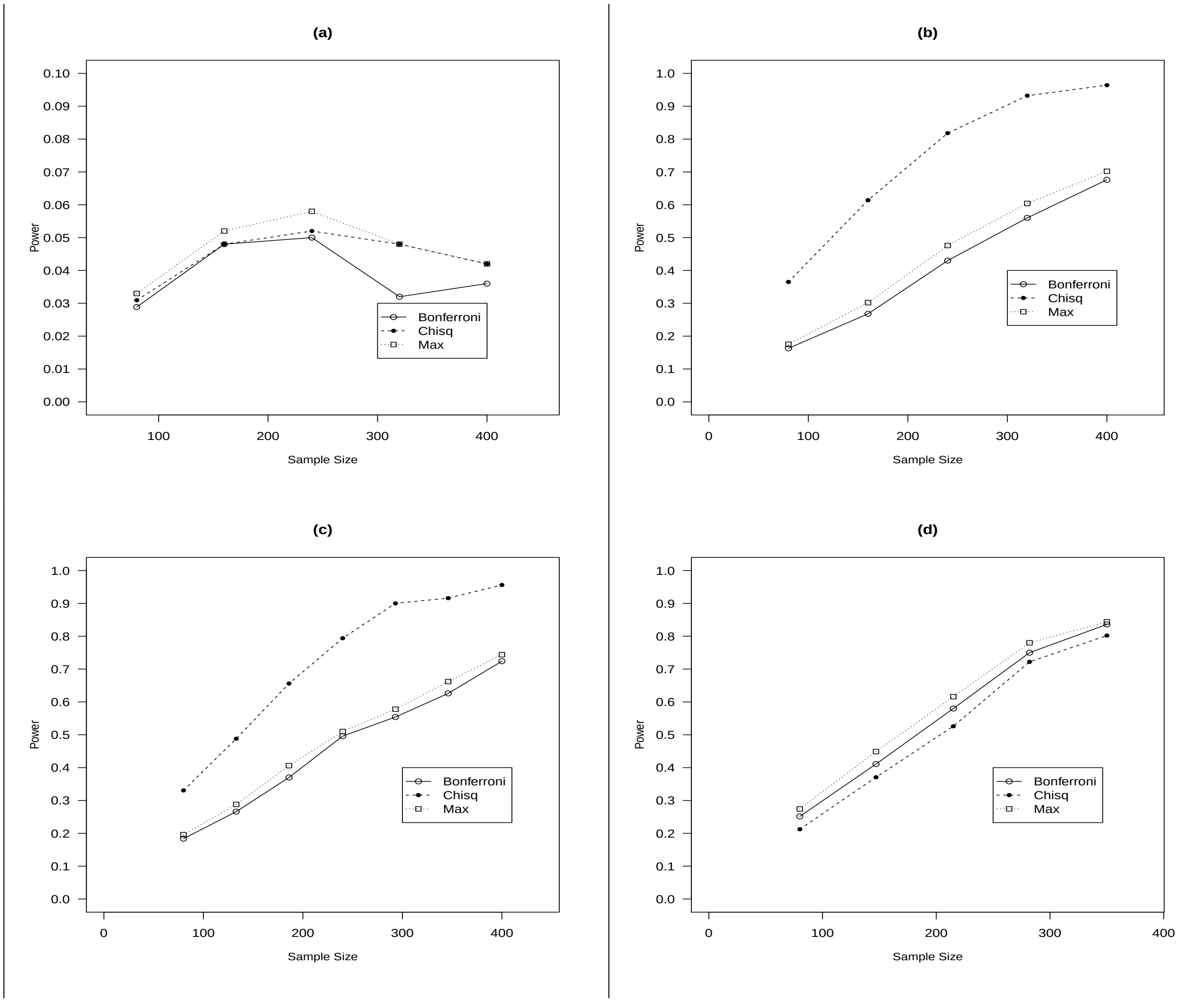

Figure 2: Simulated power of the two-sided chi-square joint test, maximum joint test and Bonferroni joint test of a local hypothesis (29) for a joint regression model (28) of CSH and CIF under four scenarios as described in Section 4: (a) null case, (b) smaller effects on CSH and larger effects on CIF, (b) larger effects on CSH and smaller effects on CIF, and (d) similar effects on CSH and CIF.

a test for a pair involving that quantity was observed to have lower power than those for other pairs. This is not surprising because a joint test for a specific pair is constructed to detect a group difference in the direction of that pair. The details are omitted.

5 Real Data Example

We illustrate our methods on two real data sets. In the first example we consider joint inference for time 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 the all-cause hazard (ACH) for time to progression or death (progression free survival or PFS) for follicular type lymphoma patients.

5.1 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.

Figures 3 (a) and (b) 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

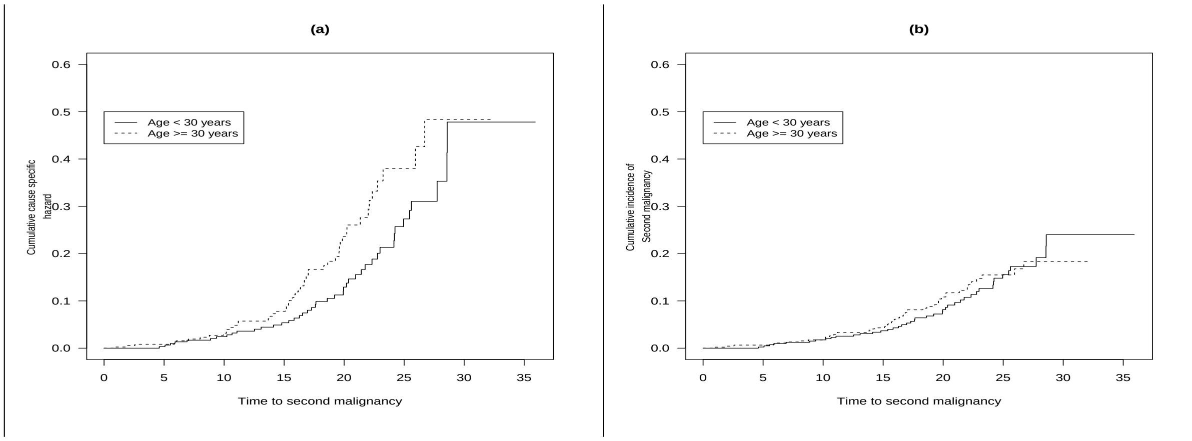

Figure 3: (a) Cumulative cause-specific hazard functions for time to second malignancy for older($\geq$ 30) and younger ($<$ 30) patients (log-rank test p-value=0.037). (b) Cumulative incidence functions for time to second malignancy for older($\geq$ 30) and younger ($<$ 30) patients (Gray's test p-value=0.770).

older patients since the slope of their cumulative cause-specific hazard is noticeably bigger (Figure 3(a)). However, the cumulative incidence functions for the two age groups are barely distinguishable (Figure 3(b)) . 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.

We performed the chi-square joint test and the maximum joint test for the null hypothesis that there is no difference in the cause-specific hazard (CSH) and the cumulative incidence (CIF) for time to second malignancy between older and younger patients. The p-values are presented in the first part of Table 1, along with the results of the individual tests and the Bonferroni's method. In contrast to the Bonferroni method, the two-sample

Table 1: Separate and Joint Test Results for Hodgkin Disease Example for Three Pairs of Quantities

	Separate Test		Joint Test		
Test	CSH	CIF	Bonferroni	χ^2	Max
p-value	0.037	0.770	0.074	0.020	0.050
Test	CSH	ACH	Bonferroni	χ^2	Max
p-value	0.037	$5.2E-8$	$1.0E-7$	$3.4E-7$	$3.0E-8$
Test	CSH	OCH	Bonferroni	χ^2	Max
p-value	0.037	$4.7E-7$	$9.4E-7$	$3.5E-7$	$8.0E-7$

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

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. The maximum joint test is also significant at level 0.05 (p-value =0.05). As illustrations, we also performed joint tests for (CSH, ACH) and for CSH with the other cause-specific hazard (OCH) (parts 2 and 3 of Table 1), which show that in addition to an elevated cause-specific hazard for time to second malignancy, the older patients also had a higher risk of dying from other life-threatening diseases without developing second malignancy. This explains why their observed cumulative incidence for time to second malignancy was not significantly different from the younger patients.

5.2 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 (15) and (23) because they correspond to two commonly used clinical endpoints, namely time to progression (TTP) and progression free survival (PFS), in oncology trials. Here TTP, defined as time to relapse or no treatment response, is an endpoint for the anti-tumor activity of a treatment, and PFS, defined as time to progression or death before progression, is an 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 A.6(a)) and the proportional cause-specific hazards model (Figure A.6(b)), which presented in Appendix A.3, indicate reasonable overall fit of both models. We conducted the 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 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 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 at 5% overall significance level, CMT group has a lower risk of TTP (cause-specific hazard) and a lower risk of PFS (ACH) as compared to the RT group adjusting for patient's baseline age, stage, and Haemoglobin

Table 2: Separate and Joint Test Results for Follicular Cell Lymphoma Study

	Separate Test		Joint Test		
Test	CSH	ACH	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.

level (hgb). Finally, the chi-square joint test has a relatively large p-value because it is actually a two-sided test that is not powered for a one-sided hypothesis, especially when the effect sizes for CSH and ACH are similar, which is consistent with our simulation results (Figure A.3(d)).

6 Discussion

We emphasize the importance of joint inference for the cause-specific hazard and the cumulative incidence because one quantity alone does not fully characterize the time to a particular type of failure in the presence of competing risks. As illustrated in our simulations and real data examples, the proposed chi-square joint test and maximum joint test can be much more powerful than the Bonferroni method. The increased power implies substantial saving in the number of patients required in a clinical trial. In a sequel, we will develop power analysis methods to determine the required sample size to test a group difference based on the developed joint tests. We also note that the chi-square joint test tends to be more powerful than the maximum joint test when the effects on the two quantities are very different and that the maximum joint test dominates the chi-square joint test when the effects on the two quantities are similar. In practice, we recommend that both joint tests be performed together with the separate tests for the individual quantities as illustrated in our real data example. The joint regression methods in Section 3 can be extended to beyond Cox's 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 sub-distribution hazard. Joint inference procedures for these models can be developed similarly. Finally, joint modeling of the cause-specific hazard and the cumulative incidence is non-trivial since the proportional cause-specific hazards model and the proportional sub-distributional hazards model are unlikely to hold simultaneously, especially for a time-independent covariate. However, this issue can be resolved by including time by covariate interactions in the regression models. In particular, we presented a joint model with piecewise proportional cause-specific hazards and piecewise proportional sub-distributional hazards and discussed how to check if the two models hold simultaneously in Section 4.

SUPPLEMENTARY MATERIAL

Appendix: Proofs for the theorems and additional simulation results. (pdf)

References

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

Andersen, P., Borgan, Ø., Gill, R., and Keiding, N. (1982). 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.

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

Andersen, P. K. (1982). Testing goodness of fit of cox's regression and life model. *Biometrics*, pages 67–77.

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

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

Beyersmann, J., Dettenkofer, M., Bertz, H., and Schumacher, M. (2007). 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.

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

Breslow, N. (1970). A generalized kruskal-wallis test for comparing k samples subject to unequal patterns of censorship. *Biometrika*, **57**(3), 579–594.

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

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

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

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

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

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

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

Gehan, E. A. (1965). A generalized wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika*, **52**(1-2), 203–223.

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

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

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

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

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

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

Kalbfleisch, J. (1980). and prentice, rl (1980): The statistical analysis of failure time data.

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

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

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

Klein, J. P. and Moeschberger, M. L. (2003). *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Prentice, R., Kalbfleisch, J., Peterson Jr, A., Flournoy, N., Farewell, V., and Breslow, N. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, **34**(4), 541–554.

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

Sancho, A., Ávila, A., Gavela, E., Beltrán, S., Fernández-Nájera, J., Molina, P., Crespo, J., and Pallardó, L. (2007). Effect of overweight on kidney transplantation outcome. In *Transplantation proceedings*, volume 39, pages 2202–2204. Orlando, FL: Grune & Stratton, 1969-.

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

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

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

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

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

Tiwari, R., Kulasekera, K., and Park, C. (2006). 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.

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

APPENDIX A. Supplementary Material

APPENDIX A.1. Proofs for the theorems in the manuscript

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 (3) as

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

and (6) 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

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

Under some regularity conditions, by using the 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 multivariate normal limiting 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), where

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

with

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

To obtain the covariance $\sigma_{12}^{(1)}$, we first note that

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

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

$$
\langle M_{jk}(t), M_{j'k'}(t) \rangle = \gamma_{jj'}\gamma_{kk'} \int_0^t Y_k(u)d\Lambda_{jk}(u), \tag{A.3}
$$

where $\gamma_{uv} = 1$ if $u = v$. After plugging (A.3) into (A.2), we have

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

which converges in probability to

$$\begin{aligned} \sigma_{12}^{(1)} = \;& \left[\int_0^\tau w_1(t) \tfrac{y_2(t)}{y.(t)} v_{11}(t) + c_1(\tau) \int_0^\tau w_1(t) \tfrac{y_2(t)}{y.(t)} e_{11}(t) h_1^{-1}(t) \right] y_1(t) d\Lambda_{11}(t) \\ & + \left[\int_0^\tau w_1(t) \tfrac{y_1(t)}{y.(t)} v_{12}(t) + c_2(\tau) \int_0^\tau w_1(t) \tfrac{y_1(t)}{y.(t)} e_{12}(t) h_2^{-1}(t) \right] y_2(t) d\Lambda_{12}(t), \end{aligned} \tag{A.4}$$

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 (9) by its consistent sample estimate. □

Proof for Theorem 2. First, we 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

$$\begin{aligned} n^{-1/2} \mathbf{U}_1(\boldsymbol{\beta}_1) &= n^{-1/2} \textstyle\sum_{i=1}^n \mathbf{U}_{i1}(\boldsymbol{\beta}_1) + o_p(1), \\ n^{-1/2} \tilde{\mathbf{U}}_1(\boldsymbol{\gamma}_1) &= n^{-1/2} \textstyle\sum_{i=1}^n (\boldsymbol{\eta}_i(\boldsymbol{\gamma}_1) + \boldsymbol{\phi}_i(\boldsymbol{\gamma}_1)) + o_p(1), \end{aligned} \tag{A.5}$$

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 $\tilde{\Lambda}_{10}(t) = \int_0^t \tilde{\lambda}_{10}(u)du$ being the baseline cause-specific cumulative hazard for cause 1, $\Lambda^c(t) = \int_0^t \lambda^c(u)du$ 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 (A.5) 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

$$\boldsymbol{\Omega}^{(1)} = \begin{pmatrix} \boldsymbol{\Omega}_{(pp)}^{(1)} & \boldsymbol{\Omega}_{(pq)}^{(1)} \\ \boldsymbol{\Omega}_{(qp)}^{(1)} & \boldsymbol{\Omega}_{(qq)}^{(1)*} \end{pmatrix}, \tag{A.6}$$

where

$$\boldsymbol{\Omega}_{(pp)}^{(1)} = \int_0^\infty \left[\frac{\lim_{n\to\infty} n^{-1} \sum_l Y_l(t) \mathbf{Z}_l^{(1)}(t)^{\otimes 2} \exp(\boldsymbol{\beta}_1^T \mathbf{Z}_l^{(1)}(t))}{\lim_{n\to\infty} n^{-1} \sum_l Y_l(t) \exp(\boldsymbol{\beta}_1^T \mathbf{Z}_l^{(1)}(t))} - \bar{\mathbf{z}}^{(1)}(\boldsymbol{\beta}_1, t)^{\otimes 2} \right] \lim n^{-1} \sum_{l=1}^n Y_l(t) \exp(\boldsymbol{\beta}_1^T \mathbf{Z}_l^{(1)}(t)) d\Lambda_{10}(t), \tag{A.7}$$

and

$$\boldsymbol{\Omega}_{qq}^{*(1)} = E\left\{ (\boldsymbol{\eta}_i(\boldsymbol{\gamma}_1) + \boldsymbol{\phi}_i(\boldsymbol{\gamma}_1)) (\boldsymbol{\eta}_i(\boldsymbol{\gamma}_1) + \boldsymbol{\phi}_i(\boldsymbol{\gamma}_1))^T \right\}. \tag{A.8}$$

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

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

which converges in probability to

$$
\begin{aligned}
\boldsymbol{\Omega}_{(pq)}^{(1)} \; = \; & E\int_0^\infty \left\{\mathbf{Z}_i^{(1)}(t) - \bar{\mathbf{z}}^{(1)}(\boldsymbol{\beta}_1, t)\right\}\left\{\mathbf{Z}_i^{(2)}(t) - \bar{\mathbf{z}}^{(2)}(\boldsymbol{\gamma}_1, t)\right\}\omega_i(t) d < M_{i1}, \tilde{M}_{i1} > (t) \\
+ \; & E\int_0^\infty \left\{\mathbf{Z}_i^{(1)} - \bar{\mathbf{z}}^{(1)}(\boldsymbol{\beta}_1, t)\right\}\frac{\mathbf{q}(t)}{\pi(t)} d < M_{i1}, M_i^c > (t).
\end{aligned} \tag{A.9}
$$

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

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

where

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

This, together with (A.6), implies that

$$
\boldsymbol{\Sigma}^{(1)} = \begin{pmatrix} \boldsymbol{\Omega}_{(pp)}^{(1)-1} & \boldsymbol{\Omega}_{(pp)}^{(1)-1}\boldsymbol{\Omega}_{(pq)}^{(1)}\boldsymbol{\Omega}_{(qq)}^{(1)-1} \\ \boldsymbol{\Omega}_{(qq)}^{(1)-1}\boldsymbol{\Omega}_{(qp)}^{(1)}\boldsymbol{\Omega}_{(pp)}^{(1)-1} & \boldsymbol{\Omega}_{(qq)}^{(1)-1}\boldsymbol{\Omega}_{(qq)}^{(1)*}\boldsymbol{\Omega}_{(qq)}^{(1)-1} \end{pmatrix}. \tag{A.11}
$$

A consistent estimator for $\boldsymbol{\Sigma}^{(1)}$ is obtained by replacing all unknown quantities with their respective sample estimates in (21) in section 3. $\square$

Proof for Theorem 3. Under the null hypothesis, it was shown by Fleming and Harrington (1991) that

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

where

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

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

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

The first part of $(\mathbf{U}_1(\boldsymbol{\beta}_1), \mathbf{U}.(\boldsymbol{\beta}.))^T, i = 1, 2, \ldots, n$ 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 (2) has a bivariate normal distribution with mean $\mathbf{0}$, and variance-covariance matrix $\boldsymbol{\Omega}^{(2)}$. Since $\bar{\mathbf{Z}}^{(1)}(\boldsymbol{\beta}_1, t)$ and $\bar{\mathbf{Z}}^{(3)}(\boldsymbol{\beta}., t)$ converge in probability to some deterministic process $\bar{\mathbf{z}}^{(1)}(\boldsymbol{\beta}_1, t)$ and $\bar{\mathbf{z}}^{(3)}(\boldsymbol{\beta}., t)$, respectively, we can prove the second part of the vector (2) 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 $n^{-1/2}(\mathbf{U}_1(\boldsymbol{\beta}_1), \mathbf{U}.(\boldsymbol{\beta}.))^T$ has a $p + q$ dimension multivariate normal limiting distribution with mean $\mathbf{0}$ and variance-covariance matrix

$$\begin{pmatrix} \boldsymbol{\Omega}_{(pp)}^{(2)} & \boldsymbol{\Omega}_{(pq)}^{(2)} \\ \boldsymbol{\Omega}_{(qp)}^{(2)} & \boldsymbol{\Omega}_{(qq)}^{(2)} \end{pmatrix}, \tag{A.12}$$

where $\boldsymbol{\Omega}_{(pp)}^{(2)}$ is defined in (A.7),

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

and the covariance

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

is the limit of

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

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

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

This, together with (A.12), implies that

$$\mathbf{\Sigma}^{(2)} = \begin{pmatrix} \mathbf{\Omega}^{(2)-1}_{(pp)} & \mathbf{\Omega}^{(2)-1}_{(pp)}\mathbf{\Omega}^{(2)}_{(pq)}\mathbf{\Omega}^{(2)-1}_{(qq)} \\ \mathbf{\Omega}^{(2)-1}_{(qq)}\mathbf{\Omega}^{(2)}_{qp}\mathbf{\Omega}^{(2)-1}_{(pp)} & \mathbf{\Omega}^{(2)-1}_{(qq)} \end{pmatrix}. \tag{A.13}$$

Finally, a consistent estimator for $\mathbf{\Sigma}^{(2)}$ is obtained by replacing all unknown quantities with their respective sample estimates in (A.13). □

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

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

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

$$
\begin{aligned}
& cov(n^{-1/2} U_{1k}, n^{-1/2} \tilde{U}_{1k'}) \\
= & \left(\int_0^\tau w_1(t) v_{1k'k}(t) + c_{k'k}(\tau) \int_0^\tau w_1(t) e_{1k}(t) \hat{h}_k^{-1}(t) \right) y_k(t) d\Lambda_{1k}(t) \\
& + \sum_{l=1}^K \left(\int_0^\tau w_1(t) \frac{y_k(t)}{y_{\cdot}(t)} v_{1k'l}(t) + c_{k'l}(\tau) \int_0^\tau w_1(t) \frac{y_k(t)}{y_{\cdot}(t)} e_{1l}(t) \hat{h}_l^{-1}(t) \right) y_l(t) d\Lambda_{1l}(t),
\end{aligned} \tag{A.14}
$$

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 (A.14) by its consistent sample estimate.

□

APPENDIX A.2. Additional Simulation Results

Simulation results for one-sided two-sample tests with respect to the CSH and CIF pair under the simulation setting of Figure 1

Under the simulation setting of Figure 1, we also conducted a simulation for the one-sided two-sample tests with respect to the CSH and CIF pair. The results are presented in Figure A.1 below.

Simulation results for two-sample comparisons with respect to the CSH and ACH pair

Below we present a simulation for the two-group comparison problem with respect to the CSH and all-cause hazard (ACH) pair. Let λ_{1k} and λ_k denote the CSH for type 1 failure and the ACH, respectively, for group k ($k = 1, 2$). Assume that in each group, both types of failures have constant cause-specific hazards and thus the all-cause hazard is also constant. The censoring rate is set to be 0.1 with an independent exponential censoring time. The nominal significance level is 0.05.

Figure A.2 below depicts the simulated rejection power curves of the two-sided chi-square joint test, maximum joint test, and Bonferroni joint test for (11) under four scenarios. Figure A.2(a) represents a null case where there is no difference with respect to type 1 failure between the two groups ($\lambda_{11} = \lambda_{12} = 0.04, \lambda_{1\cdot} = \lambda_{2\cdot} = 0.05$). Figure A.2(b) corresponds to a situation where the group difference in CSH is smaller than ACH ($\lambda_{11} = 0.6, \lambda_{12} = 0.61, \lambda_{2\cdot} = 0.7, \lambda_{2\cdot} = 0.81$). Figure A.2(c) corresponds to a situation where the group difference in CSH is bigger than ACH ($\lambda_{11} = 0.05, \lambda_{12} = 0.0625, \lambda_{2\cdot} = 0.058, \lambda_{2\cdot} = 0.17$).

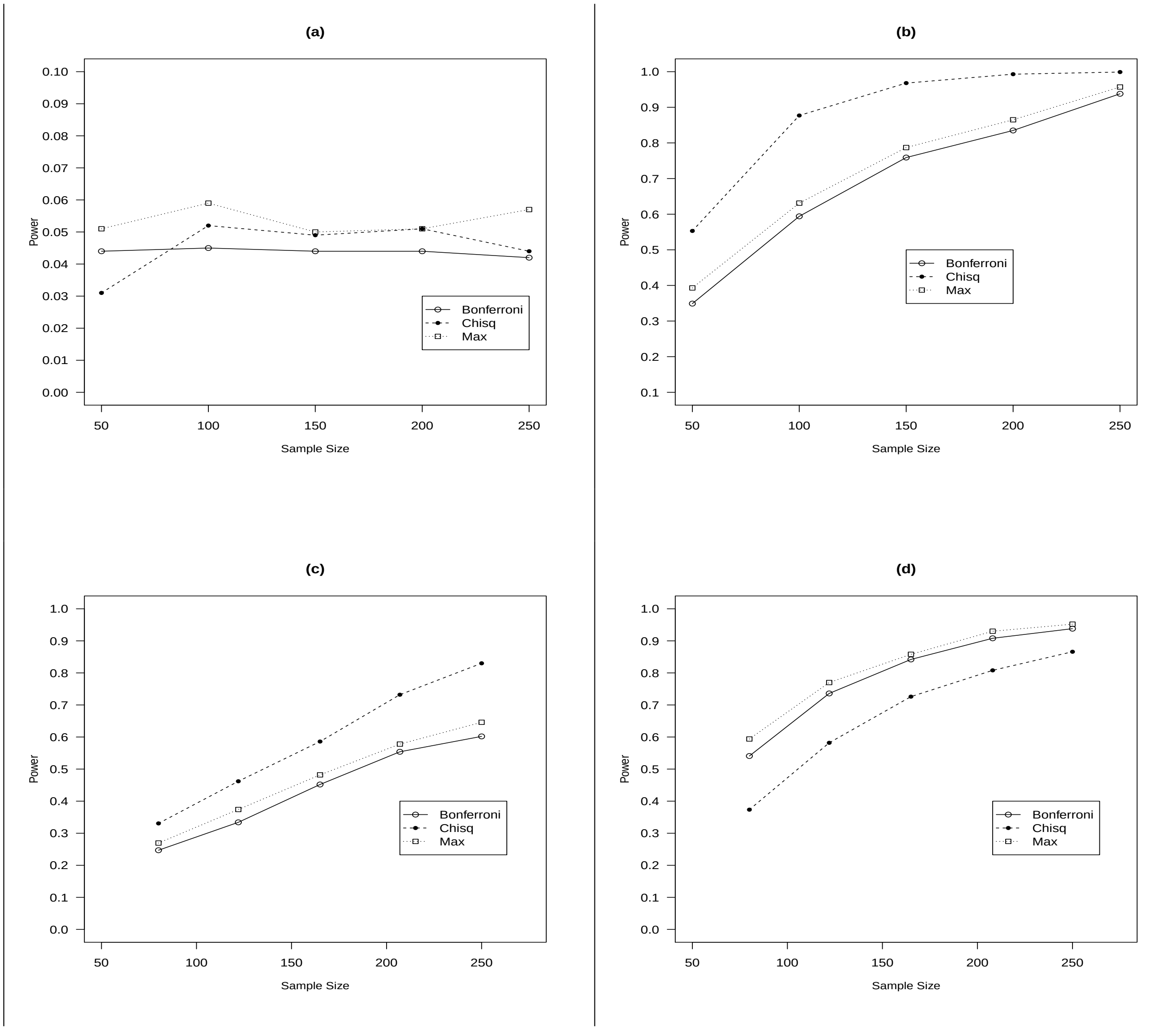

Figure A.1: Simulated power of the one-sided chi-square joint test, maximum joint test and Bonferroni joint test for two-group type 1 failure comparison with respect to the CSH and CIF pair under four scenarios as described in Section 4: (a) null case under H_0, (b) smaller group difference in CSH and larger group difference in CIF, (b) larger group difference in CSH and smaller group difference in CIF, and (d) similar group effects on CSH and CIF.

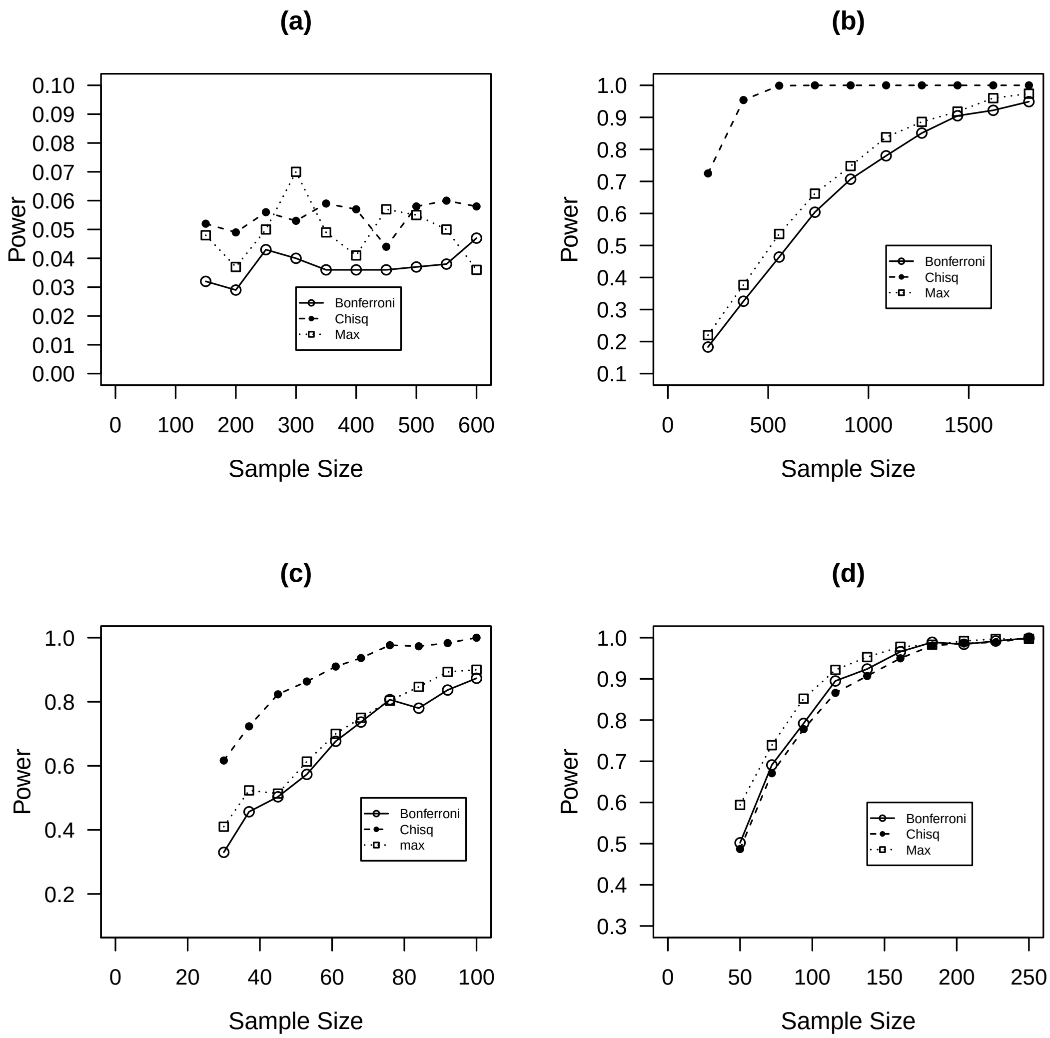

Figure A.2: Simulated power of the two-sided chi-square joint test, maximum joint test and Bonferroni joint test for two-group type 1 failure comparison with respect to the CSH and ACH pair under four scenarios: (a) null case under H_0, (b) smaller group difference in CSH and larger group difference in ACH, (b) larger group difference in CSH and smaller group difference in ACH, and (d) similar group effects on CSH and ACH.

Figure A.3 shows the power curves for the three tests under the same scenarios as Figure A.2, except that one-sided tests are performed.

It is seen from Figure A.2 and Figure A.3 that the results for the (CSH, ACH) pair is consistent with those for the (CSH, CIF) pair. The type I errors are well controlled at nominal level 0.05 for both two-sided and one-sided tests. When the group differences are different in the two quantities, the chi-square joint test is much more powerful than the Bonferroni test and the maximum test. When the group differences are similar in CSH and ACH, then the maximum joint test performs better especially for a one-sided test.

Simulation results for joint regression of CSH and ACH

This section presents some simulation results for the joint Cox model with respect to the CSH and ACH pair described in Section 3.2. Assume the following model:

$$\begin{aligned} \lambda_1(t|\mathbf{Z}) &= \lambda_{10}(t)\exp(\beta_{11}Z_1 + \beta_{21}Z_2) \\ \lambda(t|\mathbf{Z}) &= \lambda_0(t)\exp(\beta_{1\cdot}Z_1 + \beta_{2\cdot}Z_2), \end{aligned} \tag{A.15}$$

where Z_1 and Z_2 are binary variables. We are interested in testing the effects of Z_1 on both CSH and ACH, or $H_0 : \beta_{11} = 0 \quad \text{and} \quad \beta_{1\cdot} = 0$. We generated data under various different alternatives. Figure A.4 (a) represents the null case where $\beta_{11} = \beta_{1\cdot} = 0$. Figure A.4 (b) corresponds to a situation where the effect size of Z_1 for CSH is smaller than ACH ($H_a : \beta_{11} = 0 \quad \text{and} \quad \beta_{1\cdot} = -0.13$). Figure A.4(c) corresponds to a situation where the effect size of Z_1 for CSH is bigger than ACH ($H_a : \beta_{11} = -0.15 \quad \text{and} \quad \beta_{1\cdot} = 0$). Figure A.4(d) corresponds to a case when the effect sizes of Z_1 are similar for CSH and ACH ($H_a : \beta_{11} = -0.15 \quad \text{and} \quad \beta_{1\cdot} = -0.14$). For all the four scenarios, $\lambda_{10}(t) = 0.06$, $\lambda_0(t) = 0.08$, $\beta_{21} = 0.3$ and $\beta_{2\cdot} = 0.24$.

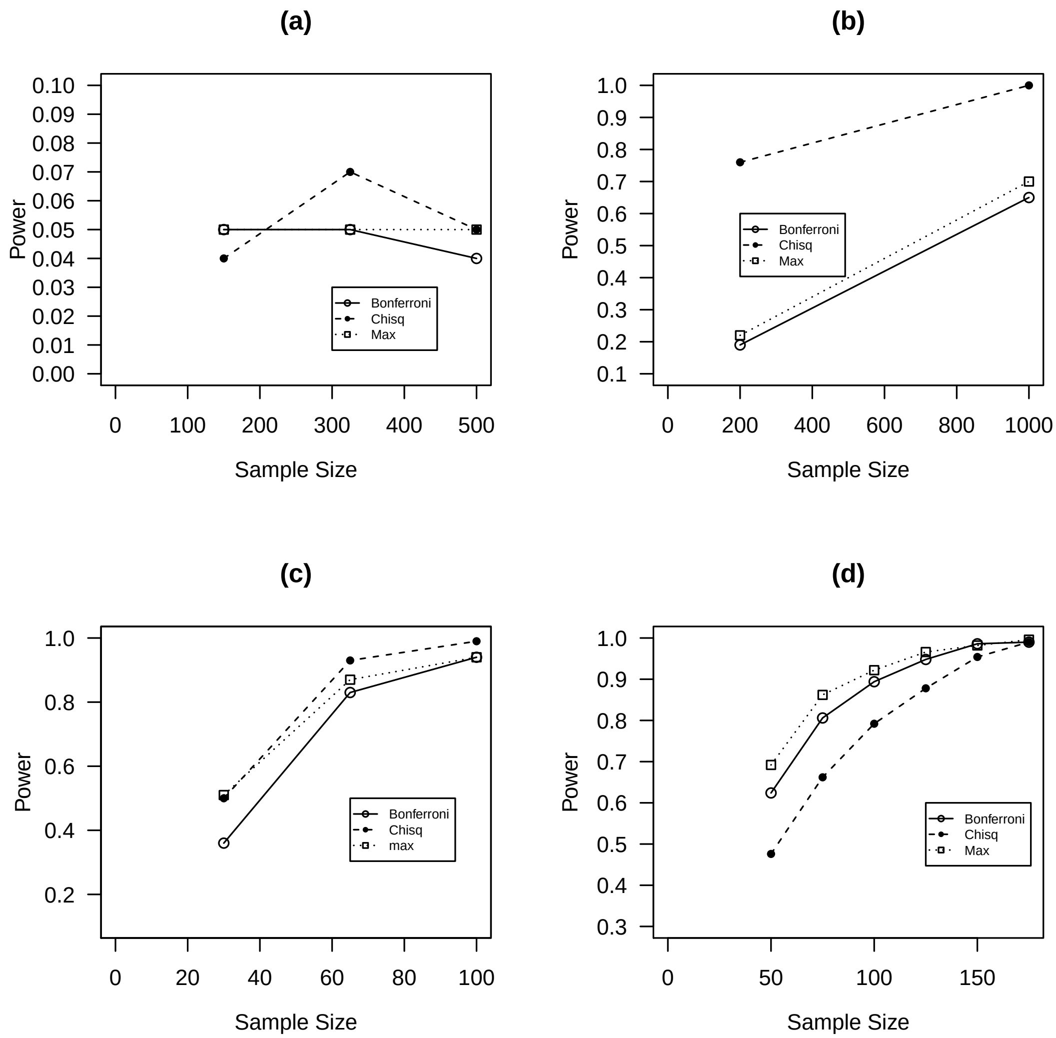

Figure A.3: Simulated power of the one-sided chi-square joint test, maximum joint test and Bonferroni joint test for two-group type 1 failure comparison with respect to the CSH and ACH pair under four scenarios: (a) null case under H_0, (b) smaller group difference in CSH and larger group difference in ACH, (b) larger group difference in CSH and smaller group difference in ACH, and (d) similar group effects on CSH and ACH.

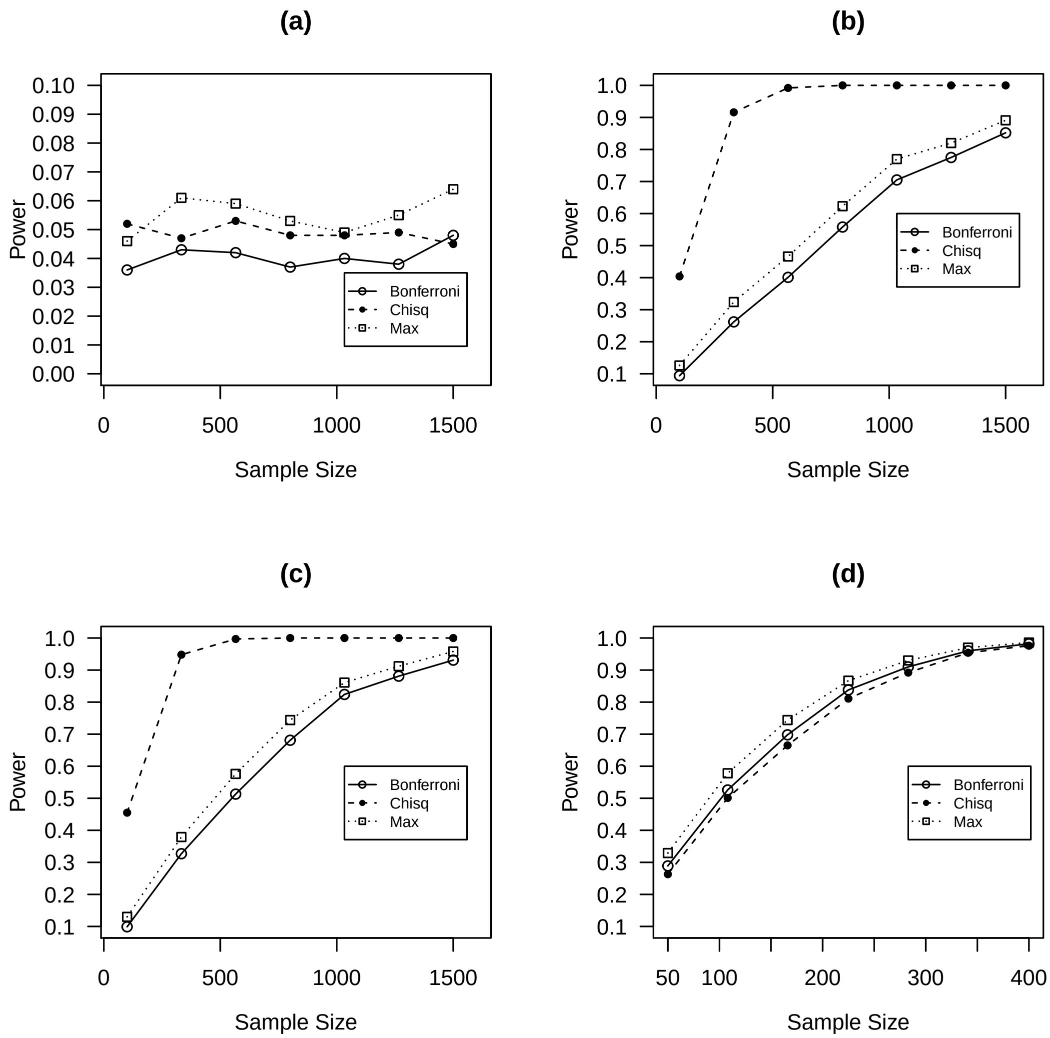

Figure A.4: Simulated power of the two-sided chi-square joint test, maximum joint test and Bonferroni joint test of a local hypothesis ($H_0 : \beta_{11} = 0 \quad \text{and} \quad \beta_{1\cdot} = 0.$) for a joint regression model (A.15) of CSH and ACH under four scenarios: (a) null case, (b) smaller effects on CSH and larger effects on ACH, (b) larger effects on CSH and smaller effects on ACH, and (d) similar effects on CSH and ACH.

Figure A.4 shows that the results are consistent with what we have observed for the CSH and CIF pair.

APPENDIX A.3. Additional Simulation Results

Graphical display of the cumulative incidence function $F_1(t)$ by group under the simulation setting of Figure 1

The cumulative incidence function $F_1(t)$ by group under the simulation setting of Figure 1 is illustrated in Figure A.5 below.

Cox-Snell residual plots for Follicular cell lymphoma study

For the Follicular cell lymphoma data, we constructed the Cox-Snell plot to check the overall fit of the Cox model for the all-cause hazard and the cause-specific hazard. The plots, along with pointwise 95% bootstrap confidence intervals are depicted in Figure A.6 below.

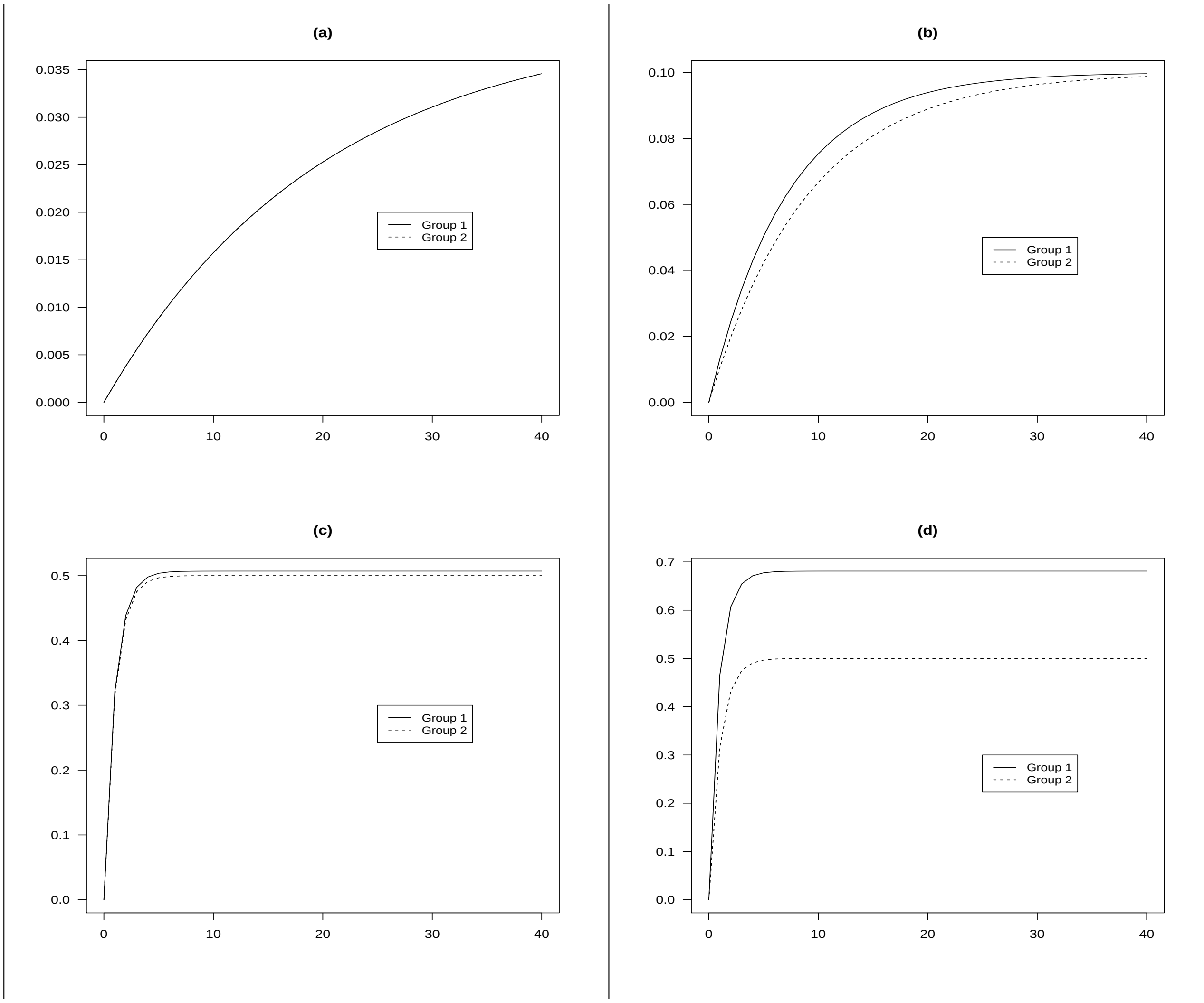

Figure A.5: Graphical illustration of the cumulative incidence function (CIF) by group under the simulation setting of Figure 1.

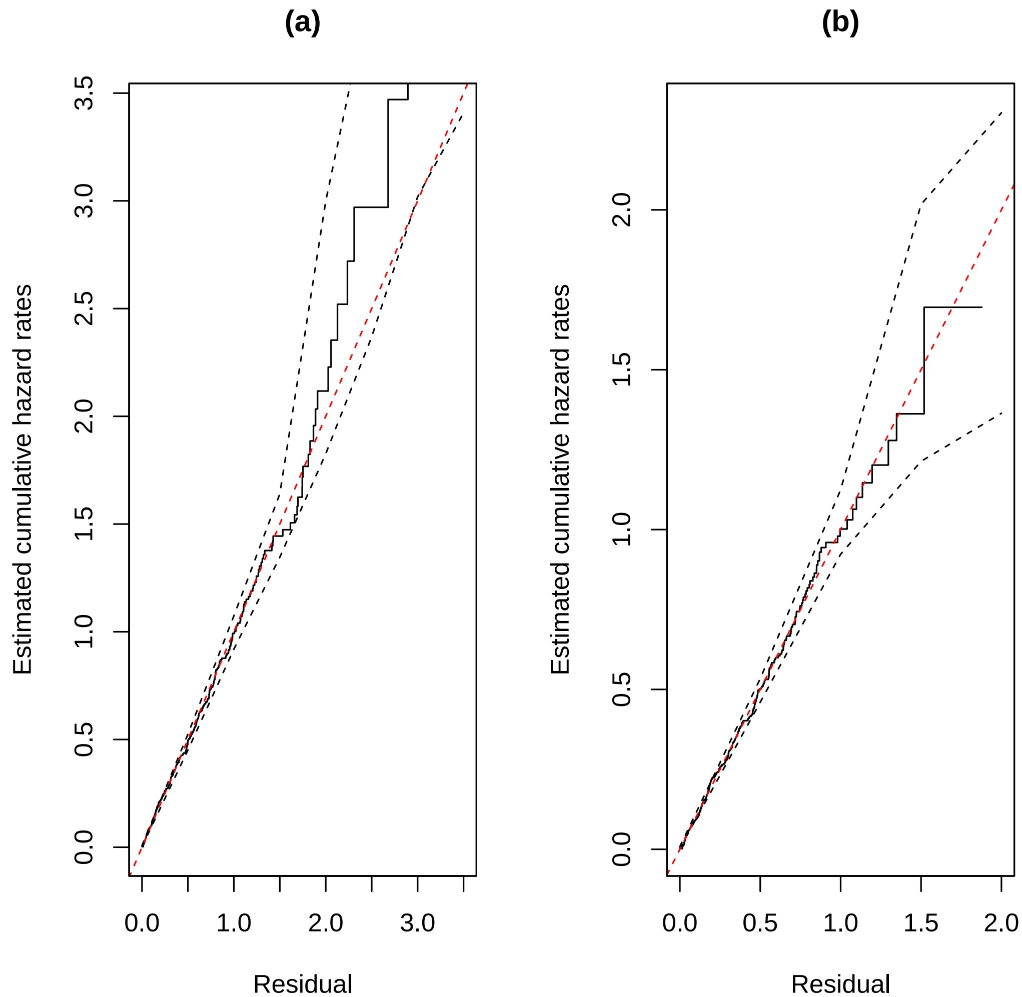

Figure A.6: The Cox-Snell residual plot (solid line) for the proportional all-cause hazards model (panel (a)) and the proportional cause-specific hazards model (panel(b)), with point-wise 95% bootstrap confidence intervals (dashed lines), and the 45 degree line (dotted lines) for the Follicular cell lymphoma data