Joint Inference for Competing Risks Survival Data

Gang Li & Qing Yang

To cite this article: Gang Li & Qing Yang (2016) Joint Inference for Competing Risks Survival Data, Journal of the American Statistical Association, 111:515, 1289-1300, DOI: 10.1080/01621459.2015.1093942

To link to this article: http://dx.doi.org/10.1080/01621459.2015.1093942

View supplementary material

Accepted author version posted online: 21 Oct 2015.
Published online: 18 Oct 2016.

Submit your article to this journal

Article views: 247

View Crossmark data
Joint Inference for Competing Risks Survival Data

Gang Li\textsuperscript{a} and Qing Yang\textsuperscript{b}

\textsuperscript{a}Department of Biostatistics, University of California Los Angeles, Los Angeles, CA, USA; \textsuperscript{b}School of Nursing, Duke University, Durham, NC, USA

\textbf{ABSTRACT}
This article 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 (i) they describe different characteristics of a given type of failure, (ii) they do not uniquely determine each other, and (iii) 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.

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), and Putter, Fiocco, and Geskus (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, Kochar, and McKeague 1994; Sun and Tiwari 1995; Lam 1998; Luo and Turnbull 1999). Tiwari, Kulasekera, and Park (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, 2001), Klein and Andersen (2005), and Gerds, Scheike, and Andersen (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), and Haller, Schmidt, and Ulm (2012).

In this article, 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

\[
\lambda_1(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t, D = 1 | T \geq t) / \Delta t}{P(t < T \leq t + \Delta t, D = 1 | T > t) / \Delta t}, \quad t > 0,
\]

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 \( F_1(t) = P(T \leq t, D = 1), t > 0 \), the cumulative incidence rate of Type 1 failure by time \( t \), which can be uniquely characterized by the following subdistribution hazard:

\[
\lambda_1(t) = \lim_{dt \downarrow 0} \frac{P(t \leq T < t + dt, D = 1 | T = t)}{dt} = t \cup (T < t \cap D \neq 1) / dt = -d \log (1 - F_1(t)) / dt.
\]

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 subdistribution hazards model for the regression problem.

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 \( \{(X_{ik}, \delta_{ik}) = (1, \ldots, n_k)\} \), where \( X_{ik} = \min(T_{ik}, C_{ik}) \) and \( \delta_{ik} = D_{ik}I(T_{ik} \leq C_{ik}) \). Let \( S_k(t) = P(T_{ik} > t) \) and \( S^*_k(t) = P(C_{ik} > t) \). For group \( k (k = 1, 2) \), let \( \lambda_{1k}(t), F_{1k}(t), \) and \( \lambda_{12}(t) \) denote the cause-specific hazard function, the cumulative incidence function, and the subdistribution hazard function, respectively, for Type 1 failure. We develop nonparametric tests for the following null hypothesis,

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

where \( \tau \) is some prespecified 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).

2.1.1. 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

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

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

\[
U_{ik} = \int_{0}^{\tau} W_1(t)Y_k(t) \left\{ \frac{dN_{ik}(t)}{Y_k(t)} - \frac{dN_j(t)}{Y(t)} \right\},
\]

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 \( \tau \) 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
\[ 
\hat{\sigma}^{-2} = n^{-1} \int_0^\tau W^2(t) \frac{Y_1(t)Y_2(t)}{Y(t)} \frac{dN_1(t)}{Y(t)}.
\]

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

### 2.1.2. 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,
\[ H_0 : F_{11}(t) = F_{12}(t) \quad \text{for all } 0 < t < \tau. \]

The Gray (1988) nonparametric test statistic is defined as
\[ 
\tilde{U}_{1k} = \int_0^\tau \tilde{W}(t) R_k(t) \left\{ \frac{dN_{1k}(t)}{R_k(t)} - \frac{dN_1(t)}{R(t)} \right\}, \tag{6}
\]
where \( \tilde{W}(t) \) is a predictable weight function that converges in probability to some deterministic function \( \tilde{w}(t) \) as \( n \to \infty \), \( R_k(t) = I(t_k \geq t) Y_k(t) \tilde{G}_{ik}(t) \hat{S}_k(t) \) can be considered as an adjusted risk set size for group \( k \) at time \( t \), \( \tilde{G}_{ik}(t) \) is the left-hand limit of the Kaplan–Meier estimate of \( G_{ik}(t) = 1 - F_{ik}(t) \hat{S}_k(t) \) is the left-hand limit of the Kaplan–Meier estimate of \( S_k(t) \), \( t_k \) is some fixed time point satisfying \( \tilde{S}_k(t_k) \hat{S}_k(t_k) > 0 \), and \( R(t) \) represents the same quantity as \( R_k(t) \) using the pooled sample. Gray (1988) showed that under (5), \( n^{-1/2} \tilde{U}_{11} / \hat{\sigma} \) has a standard normal limiting distribution, where
\[ 
\hat{\sigma}^{-2} = \sum_{k=1}^2 \int_0^\tau \hat{a}_{1k}(t) \hat{h}_{1k}(t) \hat{h}^{-1}(t) dN_1(t) + \int_0^\tau \hat{b}_{2k}(t) \hat{h}^{-2}(t) dN_{2k}(t), \tag{7}
\]
with
\[
\hat{a}_{1k}(t) = \hat{d}_{1k}(t) + \hat{b}_{1k}(t), \\
\hat{b}_{1k}(t) = \left[ I(j = 1) - \hat{G}_1(t) \hat{S}_k(t) \right] \left[ \hat{c}_k(t_1) - \hat{c}_k(t) \right], \\
\hat{c}_k(t) = \int_0^t \hat{a}_{1k}(u) \tilde{G}_1(u) (u - 1) \hat{h}^{-1}(u) dN_1(u), \\
\hat{d}_{1k}(t) = n^{-1} \int_0^t \tilde{W}(t) R_1(t) I(t = j) \frac{dN_{1k}(t)}{R_k(t)} \times \left[ I(k = 1) - \hat{h}_k(t) \hat{h}(t) \right] / \tilde{G}_1(t), \\
\hat{h}_k(t) = I(t \leq t_k) n^{-1} \tilde{Y}_k(t) / \tilde{S}_k(t), \\
\hat{h}(t) = I(t \leq \max(t_1, t_2)) n^{-1} \tilde{Y}(t) / \tilde{S}(t), \\
\tilde{G}_1(t) = 1 - \tilde{F}_1(t) = 1 - n^{-1} \int_0^t \hat{h}^{-1}(u) dN_1(u). \tag{8}
\]

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

Examples of the weight functions in the abovementioned tests have been discussed by a number of authors (Gehan 1965; Breslow 1970; Peto and Peto 1972; Kalbfleisch 1980; Gray 1988).

### 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 \( \tilde{U}_{11} \).

**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 \( 0 \) and variance-covariance matrix \( \Sigma^{(1)} = (\sigma^{(1)}_{ij}) \) as \( n \to \infty \), where \( \Sigma^{(1)} \) is defined in (A.1) and (A.4) of Appendix A.1. Furthermore, \( \sigma^{(1)}_{11} \) and \( \sigma^{(1)}_{22} \) are consistently estimated by (4) and (7), and the covariance \( \sigma^{(1)}_{12} \) is consistently estimated by
\[ 
\hat{\sigma}^{(1)}_{12} = n^{-1} \left\{ \int_0^\tau W_1(t) Y_2(t) \frac{\hat{Y}_1(t)}{Y(t)} \hat{V}_1(t) + \int_0^\tau W_1(t) \frac{Y_1(t)}{Y(t)} \hat{V}_1(t) \hat{h}_1^{-1}(t) \right\} Y_1(t) d\hat{A}_{11}(t) + n^{-1} \left\{ \int_0^\tau W_1(t) Y_1(t) \frac{\hat{V}_1(t)}{Y(t)} \hat{V}_1(t) \right\} Y_1(t) d\hat{A}_{11}(t) + \hat{c}_2(t) \int_0^\tau W_1(t) Y_1(t) \frac{\hat{V}_1(t)}{Y(t)} \hat{V}_1(t) \hat{h}_2^{-1}(t) \right\} Y_2(t) d\hat{A}_{12}(t). \tag{9}
\]

where \( \hat{A}_{ik}(t) = \int_0^t Y_k(t') dN_{ik}(t') \), \( \hat{V}_j(t) = \left[ \hat{d}_{jk}(t) - \hat{E}_{jk}(t) \hat{c}_k(t) \right] \hat{h}_k^{-1}(t) \hat{h}_k^{-1}(t), \hat{E}_{jk}(t) = I(j = 1) - \hat{G}_{jk}(t) \hat{S}_k(t), \) and other quantities are defined in (8).

### 2.2.1. Chi-Square Joint Test for (1)

Define
\[ 
X^2 = n^{-1/2} \left( U_{11}, \tilde{U}_{11} \right) \frac{\Sigma^{(1)-1}}{\hat{\sigma}^{(1)}} \left( U_{11}, \tilde{U}_{11} \right). 
\]

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

Reject (1) at level \( \alpha \) if \( X^2 > \chi^2_2(\alpha) \), where \( \chi^2_2(\alpha) \) is the upper \( 1 - \alpha \) percentile of the standard \( \chi^2_2 \) 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.

### 2.2.2. Maximum Joint Test for (1)

Define
\[ 
T^* = \max(|Z_{11}|, |\tilde{Z}_{11}|),
\]
where \( Z_{11} = n^{-1/2} U_{11} / \sqrt{\sigma^{(1)}_{11}} \) and \( \tilde{Z}_{11} = n^{-1/2} \tilde{U}_{11} / \sqrt{\sigma^{(1)}_{22}} \). 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 \( \hat{\rho} = \frac{\hat{\sigma}_1}{\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}} \). 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 \( T^*_N \) be the upper 100(1 - \( \alpha \))th sample quantile of \( T^*_1, \ldots, T^*_N \). Reject the null hypothesis (1) at level \( \alpha \) if \( T^* > T^*_N \).

**Remark 1.** It is straightforward to modify the maximum joint test procedure to test one-sided alternative(s) based on \( T^* \) or \( T^* \). Let \( T^*_N \) be the upper 100(1 - \( \alpha \))th sample quantile of \( T^*_1, \ldots, T^*_N \). Reject the null hypothesis (1) at level \( \alpha \) if \( T^* > T^*_N \).

**Remark 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

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

where \( \tau \) is some pre-specified fixed time. Similar to Theorem 1, it can be shown that under the null hypothesis (10),

\[
\sum_{k=1}^{K} \hat{N}_k(t) = \sum_{k=1}^{K} \hat{N}_k(t) \quad \text{has an asymptotic bivariate normal distribution with mean } \mathbf{0} \quad \text{and variance-covariance matrix } \Sigma^*, \quad \text{where } \Sigma^* \text{ is defined as the limit of the variance-covariance matrix of } \hat{V}_k \quad \text{and can be consistently estimated as follows.}
\]

From Kulthimal and Gasbarra (2002), we have

\[
\sum_{k=1}^{K} \hat{N}_k(t) = \sum_{k=1}^{K} \hat{N}_k(t) \quad \text{has an asymptotic bivariate normal distribution with mean } \mathbf{0} \quad \text{and variance-covariance matrix } \Sigma^*, \quad \text{where } \Sigma^* \text{ is defined as the limit of the variance-covariance matrix of } \hat{V}_k \quad \text{and can be consistently estimated as follows.}
\]

Consider the following null hypotheses

\[
H_0 : \lambda_{11}(t) = \lambda_{12}(t) = \cdots = \lambda_{1K}(t) = \lambda_{21}(t) = \lambda_{22}(t) \quad \text{for all } 0 < t < \tau. \quad (11)
\]

Let

\[
U_k = \int_0^\tau \frac{dN_k(t)}{Y_k(t)} - \frac{dN_{1k}(t)}{Y_{1k}(t)}, \quad (12)
\]

be the weighted log-rank test statistic for \( H_0 : \lambda_{11}(t) = \lambda_{22}(t) \) for all \( 0 < t < \tau \), where \( N_k(t) = \sum_{j=1}^{K} N_{jk}(t), \quad N_{1k}(t) = \sum_{j=1}^{K} N_{1jk}(t), \) and \( Y(t) \) is a predictable weight function that converges in probability to some deterministic function \( w(t) \) as \( n \to \infty \). Let \( U_{11} \) and \( U_{12} \) be defined by (3) and (12). Then, \( n^{-1/2} U_{11} \) and \( U_{12} \) has an asymptotic bivariate normal distribution with mean \( \mathbf{0} \) and variance-covariance matrix \( \Sigma^2 = (\sigma^2_{ij}) \). Furthermore, \( \Sigma^2 \) is consistently estimated by

\[
\hat{\Sigma}^2 = (\hat{\sigma}^2_{ij}), \quad \text{where } \hat{\sigma}^2_{11}(t) = \int_0^\tau W^1_1(t) \frac{Y(t)Y(t)}{Y_{1k}(t)Y_{1k}(t)} dN_{1k}(t), \quad \text{and } \hat{\sigma}^2_{12}(t) = \int_0^\tau W^1_1(t) W^1_2(t) \frac{dN_{1k}(t)}{Y_{1k}(t)} \frac{dN_{2k}(t)}{Y_{2k}(t)}, \quad \text{and } \hat{\sigma}^2_{22}(t) = \int_0^\tau W^1_2(t) W^1_2(t) \frac{dN_{2k}(t)}{Y_{2k}(t)} \frac{dN_{2k}(t)}{Y_{2k}(t)}. \quad (13)
\]

These results allow one to construct a chi-square joint test and a maximum joint test for (11) similar to those for (1) in the previous section.

**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 \( \lambda_k(t) \) denote the other (Type 2) cause-specific hazard function and the all-cause hazard function, respectively.

\[
2.3.1. \quad \text{Two-Sample Joint Tests for Cause-Specific Hazards and All-Cause Hazard}
\]

Consider the following null hypotheses

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

Let

\[
U_k = \int_0^\tau W(t) Y_k(t) \left\{ \frac{dN_k(t)}{Y_k(t)} - \frac{dN_{1k}(t)}{Y_{1k}(t)} \right\}, \quad (12)
\]

be the weighted log-rank test statistic for \( H_0 : \lambda_{11}(t) = \lambda_{22}(t) \) for all \( 0 < t < \tau \), where \( N_k(t) = \sum_{j=1}^{K} N_{jk}(t), \quad N_{1k}(t) = \sum_{j=1}^{K} N_{1jk}(t), \) and \( Y(t) \) is a predictable weight function that converges in probability to some deterministic function \( w(t) \) as \( n \to \infty \). Let \( U_{11} \) and \( U_{12} \) be defined by (3) and (12). Then, \( n^{-1/2} U_{11} \) and \( U_{12} \) has an asymptotic bivariate normal distribution with mean \( \mathbf{0} \) and variance-covariance matrix \( \Sigma^2 = (\sigma^2_{ij}) \). Furthermore, \( \Sigma^2 \) is consistently estimated by

\[
\hat{\Sigma}^2 = (\hat{\sigma}^2_{ij}), \quad \text{where } \hat{\sigma}^2_{11}(t) = n^{-1} \int_0^\tau W^1_1(t) \frac{Y(t)Y(t)}{Y_{1k}(t)Y_{1k}(t)} dN_{1k}(t), \quad \text{and } \hat{\sigma}^2_{12}(t) = n^{-1} \int_0^\tau W^1_1(t) W^1_2(t) \frac{dN_{1k}(t)}{Y_{1k}(t)} \frac{dN_{2k}(t)}{Y_{2k}(t)}, \quad \text{and } \hat{\sigma}^2_{22}(t) = n^{-1} \int_0^\tau W^1_2(t) W^1_2(t) \frac{dN_{2k}(t)}{Y_{2k}(t)} \frac{dN_{2k}(t)}{Y_{2k}(t)}. \quad (13)
\]

These results allow one to construct a chi-square joint test and a maximum joint test for (11) similar to those for (1) in the previous section.

**Remark 3.** It can be shown that for group \( k \), the three pairs of functions \( (\lambda_{1k}(t), F_{1k}(t)), (\lambda_{1k}(t), \lambda_k(t)), \) and \( (\lambda_{1k}(t), \lambda_{2k}(t)) \) uniquely determine each other and that each pair uniquely determines the joint distribution of \( (X_{1k}, \delta_{1k}) \). 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 3.1. 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 \((\lambda_{1k}(\cdot), \lambda_{k}(\cdot))\), 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 \((X_i, \delta_i, Z_i)\), 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 a discrete random variable taking values on \(1, 2\) with \(D_i = j\) indicating that type \(j\) failure is observed, and \(Z_i\) 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 \(Z_i\), and \(pr(C_i \geq t) = G(t)\).

Let \(\lambda_1(t|z)\) and \(\tilde{\lambda}_1(t|z)\) be the conditional cause-specific hazard function and the conditional subdistribution hazard function for Type 1 failure for an individual with covariate \(z\). Assume the proportional cause-specific hazards model (Prentice et al. 1978)

\[
\lambda_1(t|z) = \lambda_{10}(t) \exp(\beta_1^T Z^{(1)}(t)),
\]

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

\[
\tilde{\lambda}_1(t|z) = \tilde{\lambda}_{10}(t) \exp(\gamma_1^T Z^{(2)}(t)),
\]

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 \(Z^{(1)}(t)\) and \(Z^{(2)}(t)\) are functions of the original covariates \(Z\) and \(t\) that allow time \(\times\) covariates interactions. Prentice et al. (1978) showed that inference for \(\beta_1\) 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 \(\gamma_1\).

Below we develop joint inference for \(\beta_1\) and \(\gamma_1\). Specifically, we consider the following joint null hypothesis

\[
H_0 : A_1^T \beta_1 = d_1 \text{ and } A_2^T \gamma_1 = d_2,
\]

where \(A_1\) and \(A_2\) are constant matrices, and \(d_1\) and \(d_2\) are constant column vectors.

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

\[
U_1(\beta_1) = \sum_{i=1}^{n} \int_0^\infty \left[ Z_i^{(1)}(t) - \tilde{Z}^{(1)}(\beta_1, t) \right] dN_i(t),
\]

and

\[
\tilde{U}_1(\gamma_1) = \sum_{i=1}^{n} \int_0^\infty \left[ Z_i^{(2)}(t) - \tilde{Z}^{(2)}(\gamma_1, t) \right] \omega_i(t) dN_i(t),
\]

be the score functions for \(\beta_1\) and \(\gamma_1\) under models (15) and (16), respectively, where

\[
\tilde{Z}^{(1)}(\beta_1, t) = \frac{\sum_{i=1}^{n} Y_i(t) Z_i^{(1)}(t) \exp(\beta_1^T Z_i^{(1)}(t))}{\sum_{i=1}^{n} Y_i(t) \exp(\beta_1^T Z_i^{(1)}(t))},
\]

\[
Y_i(t) = I(X_i \geq t),
\]

and

\[
\tilde{Z}^{(2)}(\gamma_1, t) = \frac{\sum_{i=1}^{n} \omega_i(t) \tilde{Y}_i(t) Z_i^{(2)}(t) \exp(\gamma_1^T Z_i^{(2)}(t))}{\sum_{i=1}^{n} \omega_i(t) \tilde{Y}_i(t) \exp(\gamma_1^T Z_i^{(2)}(t))},
\]

\[
\tilde{N}_i(t) = I(T_i \leq t, D_i = 1), \quad \tilde{Y}_i(t) = 1 - \tilde{N}_i(t), \quad \omega_i(t) = I(C_i \geq T_i \land D_i), \quad \tilde{G}(t) = G(t)/\tilde{G}(X_i \land t), \quad \text{and } \tilde{G}\text{'s are the Kaplan and Meier (1958) estimate of the survival function } G\text{' of the censoring variable } C. \text{ Note that } \tilde{N}_i(t) \text{ is different from } N_i(t) \text{ and may not be observed if the subject is censored, but } \omega_i(t) \tilde{N}_i(t) \text{ can always be computed. Let } \hat{\beta}_1 \text{ and } \hat{\gamma}_1 \text{ be the solutions of the score equations } U_1(\beta_1) = 0 \text{ and } \tilde{U}_1(\gamma_1) = 0, \text{ respectively.}

Theorem 2. Under similar regularity conditions to Andersen et al. (1982) and Fine and Gray (1999), we have

\[
n^{1/2} \left( \frac{\hat{\beta}_1 - \beta_1}{\hat{\gamma}_1 - \gamma_1} \right) \xrightarrow{d} (0, \Sigma^{(1)}), \quad \text{as } n \to \infty,
\]

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

\[
\hat{\Sigma}^{(1)} = \begin{pmatrix}
\hat{\Omega}^{(1)}_{pp} & \hat{\Omega}^{(1)}_{pq} & \hat{\Omega}^{(1)}_{qq} \\
\hat{\Omega}^{(1)}_{qp} & \hat{\Omega}^{(1)}_{qq} & \hat{\Omega}^{(1)}_{pq} \\
\hat{\Omega}^{(1)}_{qpp} & \hat{\Omega}^{(1)}_{qp} & \hat{\Omega}^{(1)}_{qq}
\end{pmatrix},
\]

(20)
where

\[
\hat{\Omega}_{1(p)}^{(1)} = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\infty} \left[ \sum_{i=1}^{n} \omega(t) Y_i(t) Z_i^{(1)}(t)^{\otimes 2} \exp(\hat{\beta}_1 T_{1i}^{(1)}(t)) \right] \frac{1}{\sum_{i=1}^{n} Y_i(t) \exp(\hat{\beta}_1 T_{1i}^{(1)}(t))} dt \tilde{N}_i(t),
\]

\[
\hat{\Omega}_{2(p)}^{(1)} = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\infty} \left[ \sum_{i=1}^{n} \omega(t) Y_i(t) Z_i^{(2)}(t)^{\otimes 2} \exp(\hat{\beta}_1 T_{2i}^{(2)}(t)) \right] \frac{1}{\sum_{i=1}^{n} Y_i(t) \exp(\hat{\beta}_1 T_{2i}^{(2)}(t))} dt \tilde{N}_i(t),
\]

\[
\hat{\Omega}_{3(p)}^{(1)} = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\infty} \left[ \sum_{i=1}^{n} \omega(t) Y_i(t) Z_i^{(3)}(t)^{\otimes 2} \exp(\hat{\beta}_1 T_{3i}^{(3)}(t)) \right] \frac{1}{\sum_{i=1}^{n} Y_i(t) \exp(\hat{\beta}_1 T_{3i}^{(3)}(t))} dt \tilde{N}_i(t),
\]

\[
\hat{\Omega}_{4(p)}^{(1)} = \frac{1}{n} \sum_{i=1}^{n} \left\{ \int_{0}^{\infty} \left( Z_i^{(1)}(t) - \tilde{Z}_i^{(1)}(\hat{\beta}_1, t) \right) dt \tilde{N}_i(t) \right\} \times \left( \tilde{N}_i(t) - Y_i(t) \exp(\hat{\beta}_1 T_{1i}^{(1)}(t)) \right) * \tilde{\eta}_i + \hat{\phi}_i \right\} + \sum_{i=1}^{n} \int_{0}^{\infty} \left( Z_i^{(2)}(t) - \tilde{Z}_i^{(2)}(\hat{\beta}_1, t) \right) dt \tilde{N}_i(t),
\]

\[
\hat{\Omega}_{5(p)}^{(1)} = \frac{1}{n} \sum_{i=1}^{n} \left\{ \int_{0}^{\infty} \left( Z_i^{(3)}(t) - \tilde{Z}_i^{(3)}(\hat{\beta}_1, t) \right) dt \tilde{N}_i(t) \right\} \times \left( \tilde{N}_i(t) - Y_i(t) \exp(\hat{\beta}_1 T_{1i}^{(1)}(t)) \right) * \tilde{\eta}_i + \hat{\phi}_i \right\} + \sum_{i=1}^{n} \int_{0}^{\infty} \left( Z_i^{(2)}(t) - \tilde{Z}_i^{(2)}(\hat{\beta}_1, t) \right) dt \tilde{N}_i(t),
\]

\[
\hat{\eta}_i = \int_{0}^{\infty} \left[ Z_i^{(1)}(t) - \tilde{Z}_i^{(1)}(\hat{\beta}_1, t) \right] \omega(t) d\tilde{M}_i(t),
\]

\[
\hat{M}_i(t) = \tilde{N}_i(t) - \int_{0}^{t} Y_i(u) \exp(\hat{\beta}_1 T_{1i}^{(1)}(u)) du d\tilde{\Lambda}_1(t),
\]

\[
\hat{\Lambda}_1(t) = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{t} \left\{ n \sum_{i=1}^{n} Y_i(u) \exp(\hat{\beta}_1 T_{1i}^{(1)}(u)) \right\}^{-1} \times \omega_i(u) du \tilde{N}_i(t),
\]

\[
\hat{\phi}_i = \int_{0}^{\infty} \dfrac{\tilde{q}(t)}{\tilde{\pi}(t)} d\tilde{M}_i(t),
\]

\[
\hat{M}_i(t) = I(X_i \leq t, \delta_i = 1) - \int_{0}^{t} I(X_i \geq u) d\hat{\Lambda}^c(u),
\]

\[
\hat{\Lambda}^c(t) = \int_{0}^{t} \sum_{i=1}^{n} I(X_i \leq u, \delta_i = 0) \sum_{i=1}^{n} I(X_i \geq u) du,
\]

\[
\hat{q}(t) = -n^{-1} \sum_{i=1}^{n} \int_{0}^{\infty} \left\{ Z_i^{(2)}(s) - \tilde{Z}_i^{(2)}(\hat{\beta}_1, s) \right\} \times I(s \geq t) \omega_i(s) d\tilde{M}_i(s),
\]

\[
\hat{\pi}(t) = n^{-1} \sum_{i=1}^{n} I(X_i \geq t),
\]

Corollary 1. Let \( \hat{\xi}_n = n^{1/2}(A_1 \hat{\beta}_1 - d_1) \) and \( \eta_n = n^{1/2}(A_2 \hat{\beta}_1 - d_2) \). Then, under the null hypothesis (17), we have

\[
\left( \begin{array}{c} \hat{\xi}_n \\ \eta_n \end{array} \right) \overset{N}{\rightarrow} \left( \begin{array}{c} 0 \\ 0 \end{array} \right), \quad \text{as } n \rightarrow \infty,
\]

where

\[
V = \left( \begin{array}{cc} A_1 & 0 \\ 0 & A_2 \end{array} \right) \Sigma^{(1)} \left( \begin{array}{cc} A_1^T & 0 \\ 0 & A_2^T \end{array} \right),
\]

Define the following Wald-type test statistic

\[
\hat{X}_W^2 = \left( \begin{array}{c} \hat{\xi}_n \\ \eta_n \end{array} \right)^T \hat{V}^{-1} \left( \begin{array}{c} \hat{\xi}_n \\ \eta_n \end{array} \right),
\]

where \( \hat{V} \) is a consistent estimate of \( V \) obtained by replacing \( \Sigma^{(1)} \) with \( \hat{\Sigma}^{(1)} \) in (22). It follows immediately from Corollary 1 that under (17), \( \hat{X}_W^2 \) has an asymptotic chi-squared distribution with \( p_{d1} + p_{d2} \) degrees of freedom, where \( p_{d1} \) and \( p_{d2} \) are the dimensions of \( d_1 \) and \( d_2 \), respectively. This leads to the following chi-square joint test for (17):

\[
\text{Reject (17) at level } \alpha \text{ if } \hat{X}_W^2 > \chi_{p_{d1} + p_{d2}}^2(\alpha),
\]

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

3.2. Joint Regression Analysis of Other Quantities

Besides analyzing \( \lambda_1(t|Z) \) and \( \hat{\lambda}_1(t|Z) \) jointly, it is sometimes also useful to consider other related quantities as discussed in Section 2.3 (Remark 3).

3.2.1. 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:

\[
\lambda(t|Z) = \lambda_0(t) \exp(\beta^T Z^{(3)}(t)),
\]

where \( \lambda(t|Z) \) denote the conditional all-cause hazard function given \( Z, \lambda_0(t) \) is an unknown baseline hazard, and \( Z^{(3)}(t) \) are functions of the original covariates \( Z \) and \( t \) that allow time \( \times \) covariates interactions. Below we derive joint inference for \( \beta_1 \) and \( \beta \).

Let

\[
U(\hat{\beta}) = \sum_{i=1}^{n} \int_{0}^{\infty} \left[ Z_i^{(1)}(t) - \tilde{Z}_i^{(1)}(\hat{\beta}, t) \right] dN_i(t),
\]

be the score function for \( \beta \) under model (23), where

\[
\tilde{Z}_i^{(3)}(\beta, t) = \sum_{i=1}^{n} Y_i(t) Z_i^{(3)}(t) \exp(\beta^T Z_i^{(3)}(t))
\]

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

Theorem 3. Under some regularity conditions, as \( n \rightarrow \infty \),

\[
n^{1/2} \left( \begin{array}{c} \hat{\beta}_1 - \beta_1 \\ \hat{\beta} - \beta \end{array} \right) \overset{N}{\rightarrow} \left( \begin{array}{c} 0 \\ 0 \end{array} \right),
\]
where \( \Sigma^{(2)} \) is defined by (A.13) in Appendix A.1. Furthermore, \( \Sigma^{(2)} \) can be consistently estimated by

\[
\hat{\Sigma}^{(2)} = \begin{pmatrix}
\hat{\Omega}^{(2)-(1)} & \hat{\Omega}^{(2)-(1)} & \hat{\Omega}^{(2)-(1)} \\
\hat{\Omega}^{(2)-(1)} & \hat{\Omega}^{(2)-(1)} & \hat{\Omega}^{(2)-(1)} \\
\hat{\Omega}^{(2)-(1)} & \hat{\Omega}^{(2)-(1)} & \hat{\Omega}^{(2)-(1)}
\end{pmatrix}
\]

(25)

where

\[
\hat{\Omega}^{(2)}_{(pp)} = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\infty} \left[ \sum_{i=1}^{n} Y_i(t) Z_i(t)^2 \exp(\hat{\beta}_1 Z_i(t)) \right] dN_{01}(t),
\]

\[
\hat{\Omega}^{(2)}_{(pq)} = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\infty} \left[ \sum_{i=1}^{n} \left( Z_i(t) - \hat{Z}_1(\hat{\beta}_1, t) \right) Y_i(t) \right] dN_{10}(t),
\]

\[
\hat{\Omega}^{(2)}_{(qq)} = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\infty} \left[ \sum_{i=1}^{n} \left( \hat{Z}_1(\hat{\beta}_1, t) \right) Y_i(t) \right] dN_{20}(t),
\]

Remark 4. In addition to being easy to interpret, the PH models for the cause-specific hazard and the all-cause hazard only require that the censoring time be conditionally independent of the survival time given the observed covariates, which is weaker than the completely censoring at random assumption needed by the proportional subdistribution hazards model.

Remark 5 (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, Oosting, and Hart 1984; Moreau, O’Quigley, and Mesbah 1985; Arjas 1988; Beyersmann et al. 2007; Latouche et al. 2007; Grambauer, Schumacher, and Beyersmann 2010; Andersen et al. 2012; Haller, Schmidt, and Ulm 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 to model (15) and (16) to allow time × covariates interactions. To check if (15) and (16) hold simultaneously, one needs to verify that for any \( z \), \( \Lambda_2(t|z) = \Lambda_1(t|z) - \Lambda_1(t|z) \log \lambda_1(t|z) - \log \tilde{\lambda}_1(t|z) \) is nondecreasing and satisfies \( \Lambda_2(0|z) = 0 \). In other words, the above defined \( \Lambda_2(t|z) \) 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) 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 \( \lambda_{k1} = \lambda_{k2} = 0.4 \). Figure 1(a) and 1(c) correspond 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 \( \lambda_{k1} = \lambda_{k2} = 0.4 \). Figure 1(b) and 1(d) correspond 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 \( \lambda_{k1} = \lambda_{k2} = 0.4 \).
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 Sections 4: (a) null case under $H_0$, (b) smaller group difference in CSH and larger group difference in CIF, (c) larger group difference in CSH and smaller group difference in CIF, and (d) similar group effects on CSH and CIF.

Figure 1(d), where $\lambda_{10}(t) = 0.05 \times I(0 \leq t < 1) + 0.1 \times I(t \geq 1)$, $\lambda^*_1(t) = 0.05 e^{-t} + 0.1 (1 - e^{-t})$, 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 Appendix A.3 (Figure A.5).

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 (Figure 1(b) and 1(c)), the chi-square joint test is observed to be most powerful with substantially 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 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 is imperative to include time by covariate interactions in the joint model. As an illustration, we consider the following...
joint model:

\[
\lambda_1(t|Z) = \lambda_{10}(t) \exp(\gamma^T Z \ast I(t < \tau_0) + \beta^T Z \ast I(t \geq \tau_0)),
\]

\[
\tilde{\lambda}_1(t|Z) = \tilde{\lambda}_{10}(t) \exp(\gamma^T Z \ast I(t < \tau_0) + \tilde{\beta}^T Z \ast I(t \geq \tau_0)),
\]

where \( \lambda_{10}(t) = a(t(0 \leq t < \tau_0) + bI(t \geq \tau_0) \), \( \tilde{\lambda}_{10}(t) = \frac{a(t(0 \leq t < \tau_0) + bI(t \geq \tau_0)}{1-e^{-(b+c)t}} \), \( Z = (Z_1, Z_2) \), with \( Z_1, Z_2 \) being binary variables, \( \gamma = (\gamma_1, \gamma_2) \), \( \beta = (\beta_1, \beta_2) \), and \( \tau_0 \) is some prespecified constant. Note that under model (28), the conditional cumulative cause-specific hazard function for cause 2 given \( Z = z \) is

\[
\Lambda_2(t|z) = \Lambda_1(t|z) - \left( \frac{\log \lambda_1(t|z) - \log \lambda_0(t|z)}{\log \lambda_0(t|z)} \right).
\]

For \( \Lambda_2(t|z) \) to be a proper conditional cumulative cause-specific hazard function, it must satisfy

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

which imply some constraints on the parameters in model (28). For simplicity, we further assume \( \tilde{\gamma} = \tilde{\beta} \) for our simulation.

In this case, it can be shown that \( \Lambda_2(t|z) \) is a proper cumulative cause-specific hazard function if the following constraints hold: (i) \( a = c \leq b \), (ii) \( e^{\tilde{\beta}^T z} < \frac{1-a}{c} \), (iii) \( e^{\tilde{\beta}^T z} < 1/a(1 - e^{-\tau_0}) \), and (iv) \( \tilde{\gamma} = \gamma \). We then generated competing risks data from \( \lambda_1(t|z) \) and \( \lambda_2(t|z) \) using the method of Beyersmann et al. (2009).

Figure 2 displays the simulated power curves of the three two-sided joint tests described in Sections 3.1 for the following local hypothesis regarding the effects of \( Z_1 \) on the CSH and the CIF hazard after time \( \tau_0 \):

\[
H_0 : \beta_1 = 0 \quad \text{and} \quad \gamma_1 = 0.
\]

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 \( a = 0.05 \), \( b = 0.1 \), \( \beta_2 = -0.2 \), \( \gamma_2 = -0.1 \), \( y = \tilde{\gamma} \), 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, 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 datasets. 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.

Figure 3(a) and 3(b) depicts the cumulative cause-specific hazard functions and the cumulative incidence functions, respectively, for time to second malignancy for the older (≥30) and younger (<30) groups. There appears to be a higher cause-specific hazard for the older patients since the slope of their cumulative cause-specific hazard is noticeably bigger (Figure 3(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 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

<table>
<thead>
<tr>
<th>Test</th>
<th>CSH</th>
<th>CIF</th>
<th>Bonferroni</th>
<th>( \chi^2 )</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-value</td>
<td>0.037</td>
<td>0.770</td>
<td>0.016</td>
<td>0.012</td>
<td>0.012</td>
</tr>
<tr>
<td>Test</td>
<td>CSH</td>
<td>ACH</td>
<td>Bonferroni</td>
<td>( \chi^2 )</td>
<td>Max</td>
</tr>
<tr>
<td>p-value</td>
<td>0.037</td>
<td>5.2E − 8</td>
<td>1.0E − 7</td>
<td>3.4E − 7</td>
<td>3.0E − 8</td>
</tr>
<tr>
<td>Test</td>
<td>CSH</td>
<td>OCH</td>
<td>Bonferroni</td>
<td>( \chi^2 )</td>
<td>Max</td>
</tr>
<tr>
<td>p-value</td>
<td>0.037</td>
<td>4.7E − 7</td>
<td>9.4E − 7</td>
<td>3.5E − 7</td>
<td>8.0E − 7</td>
</tr>
</tbody>
</table>

NOTE: \( \chi^2 \) and Max are abbreviations for the chi-square joint test and the maximum joint test described in Section 2.2.
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 did not experience any of the two events at the end of the follow-up. 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 antitumor 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 hemoglobin 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 hemoglobin 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 nontrivial since the proportional cause-specific hazards model and the proportional subdistributional 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 subdistributional hazards and discussed how to check if the two models hold simultaneously in Section 4.

### Supplementary Materials

Appendix: Proofs for the theorems and additional simulation results.

### Acknowledgments

The authors thank the co-editor, the associate editor, and the two anonymous referees for their valuable comments that helped improve this article significantly.

### Funding

Gang Li’s work was partially supported by NIH grant 5P30CA-16042 and NIH grant 8UL1TR000124.