UCLA
Research Reports

Title
Professor

Permalink
https://escholarship.org/uc/item/9t20g0pr

Authors
Yang, Qing
Fung, Wing K
Li, Gang

Publication Date
2017-05-12

Sample size determination for jointly testing a cause-specific hazard and the any-cause hazard in the presence of competing risks

Qing Yang, Wing K. Fung, & Gang Li [*]

Abstract

This article considers sample size determination for jointly testing a cause-specific hazard and the any-cause hazard for competing risks data. The cause-specific hazard and the any-cause hazard jointly characterize important study endpoints such as the disease-specific survival and overall survival, which are commonly used as co-primary endpoints in clinical trials. Specifically, we derive sample size calculation methods for two-group comparisons based on an asymptotic chi-square joint test and a maximum joint test of the aforementioned quantities, taking into account of censoring due to lost to follow-up as well as staggered entry and administrative censoring. Our simulations demonstrate that the proposed methods can produce substantial sample size savings over the classical Bonferroni adjustment method and generally have satisfactory finite sample performance. We illustrate the application of the proposed methods using the 4-D (Die Deutsche Diabetes Dialyse Studie) clinical trial.

1 Introduction

It has been widely recognized in the statistics and clinical trials literature that in the presence of competing risks, the effects of a variable on the time to a particular type of failure, say cause-1 failure, is not completely characterized by the cause-1 cause-specific hazard (CSH_1) alone [6, 3, 14, 17]. An additional quantity, such as the cause-1 cumulative incidence function, the any-cause hazard (ACH) due to any cause, or the cause-specific hazard due to other causes, needs to be studied jointly and reported side-by-side with the cause-1 cause-specific hazard. Li and Yang [17] developed joint test procedures for each of the above three pairs of quantities for right-censored competing risks data, and showed that the joint tests can be substantially more powerful than the classical Bonferroni method. In this paper we develop a power analysis tool using joint test procedures for the planning of a clinical trial with completing risks data.

There is an extensive literature on sample size calculation for time-to-event data. For a time-to-event outcome with no competing risks, Schoenfeld [20, 21] proposed a sample size calculation formula for two-sample comparison under uniform patient entry and administrative censoring. Lachin and Foulkes [10] extended the formula to more complex situations, allowing for truncated exponential patient entry, loss to follow-up, noncompliance and stratified analysis. Yateman and Skene [25] used piecewise exponential distribution to approximate arbitrary patient entry pattern and loss to follow-up distribution. Further discussion of this topic can be found in [9, 11, 12, 13, 19, 2] and the references therein. In the presence of competing risks, current sample size calculation methods are based on either the cause-specific hazard alone [18, 22] or the cumulative incidence hazard alone [16, 15]. However, sample size calculation methods for jointly testing multiple quantities with competing risks data has yet to be developed.

The primary goal of the paper is to derive power analysis methods for competing risks data based on the joint tests of CSH_1 and ACH developed by Li and Yang [17]. Note that CSH_1 and ACH jointly characterize some commonly used endpoints in clinical trials. For instance, in a clinical trial of a new treatment targeting a specific disease, the disease-specific survival (DSS) and overall survival (OS) are often used as co-primary endpoints for how the treatment works on the disease and on a patient's overall survival, respectively. In this case, CSH_1 and ACH should be used together to study the treatment effects on DSS and OS jointly.

In Section 2, we review the joint tests of Li and Yang [17] for CSH_1 and ACH and derive their asymptotic properties under contiguous local alternatives. The asymptotic results are then used to develop approximate sample size determination procedures. Both two-sided and one-sided tests are considered. We also incorporate random censoring due to lost to follow up, staggered patient entry and administrative censoring. In Section 3 we present simulations to illustrate potential sample size savings of our methods in comparison of the classical Bonferroni adjustment method and evaluate their finite sample performance. In Section 4, we use the 4-D (Die Deutsche Diabetes Dialyse Studies) clinical trial to provide a step-by-step demonstration of how our methods are implemented and used in practice. Further remarks are given in Section 5. Technical proofs are deferred to the appendix. An R code implementing the proposed sample size calculation methods is available upon request.

2 Sample size calculation for joint tests of CSH_1 and ACH

2.1 Joint tests of CSH_1 and ACH

For the reader's convenience, we first review the joint tests of CSH_1 and ACH [17].

Suppose that there are two independent groups of subjects (1 for control and 2 for treatment). For subject i in group k, let T_{ik}, D_{ik}, and C_{ik} denote its failure time, failure type, and censoring time, respectively, $i = 1, \ldots, n_k$, $k = 1, 2$. Let a_1 and $a_2 = 1 - a_1$ be the patient allocation proportions for groups 1 and 2, respectively. Let $n = n_1 + n_2$ denote the total sample size. Assume that within group k, $\{(T_{ik}, D_{ik}, C_{ik}), i = 1, \ldots, n_k\}$ are independent and identically distributed and that the censoring time C_{ik} is independent of the failure time T_{ik}. Assume further that the two groups have the same censoring survival function $S_c(t) = P(C_{ik} > t)$. 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 $\delta_{ik} = D_{ik}I(T_{ik} \leq C_{ik})$. Denote by $S_k(t) = P(T_{ik} > t)$ the any-cause survival function for group k, $k = 1, 2$. For convenience, we assume hereafter that there are only two causes of failure and that cause-1 failure is of primary interest.

Consider the following joint hypothesis of CSH_1 and ACH:

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

where

$$\lambda_{jk}(t) = \lim_{\Delta t \to 0} \frac{P(t \le T_{ik} \le t + \Delta t, D_{ik} = j | T_{ik} \ge t)}{\Delta t}$$

is the cause-specific hazard for cause-j failure in group k ($j, k = 1, 2$), and $\lambda_{\cdot k}(t) = \lambda_{1k}(t) + \lambda_{2k}(t)$ is the any-cause hazard for group k, $k = 1, 2$.

The chi-square joint test statistic of Li and Yang [17] for (1) is defined as

$$\mathbf{X}_n^2 = n^{-1} (U_{11}, U_{\cdot 1}) \, \widehat{\boldsymbol{\Sigma}}^{-1} \begin{pmatrix} U_{11} \\ U_{\cdot 1} \end{pmatrix}, \tag{2}$$

where

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

$$U_{\cdot 1} = \int_0^\tau W_\cdot(t) \frac{Y_1(t) Y_2(t)}{Y_\cdot(t)} \left\{ \frac{dN_{\cdot 1}(t)}{Y_1(t)} - \frac{dN_{\cdot 2}(t)}{Y_2(t)} \right\},$$

$N_{jk}(t) = \sum_{i=1}^{n_k} I(X_{ki} \le t, D_{ki} = j)$ is the number of observed failures due to cause j in group k by time t, $Y_k(t) = \sum_{i=1}^{n_k} I\{X_{ki} \ge t\}$ is the number of patients in group k who are at risk just prior to time t, $N_{j\cdot}(t) = \sum_{k=1}^2 N_{jk}(t)$ and $Y_\cdot(t) = \sum_{k=1}^2 Y_k(t)$, $\widehat{\boldsymbol{\Sigma}}$ is the estimated variance-covariance matrix of $n^{-1/2}(U_{11}, U_{\cdot 1})$ [17], τ is the smallest t such that $Y_1(t) Y_2(t) = 0$, and $W_1(t)$ and $W_\cdot(t)$ are two predictable weight functions that converge in probability to some deterministic functions $w_1(t)$ and $w_\cdot(t)$ as $n \to \infty$. It has been shown by Li and Yang [17] that the limiting null distribution of $\mathbf{X}_n^2$ under H_0 is standard chi-square with 2 degrees of freedom. Thus, one rejects H_0 at level α if $\mathbf{X}_n^2 > \chi_{2,\alpha}^2$, where $\chi_{2,\alpha}^2$ is the upper α percentile of the standard χ_2^2 distribution.

The maximum joint test statistic for (1) is defined as

$$\mathbf{M_n} = \max(|Z_{11}^{(n)}|, |Z_{\cdot 1}^{(n)}|), \tag{3}$$

where $Z_{11}^{(n)} = n^{-1/2} U_{11} / \sqrt{\hat\sigma_{11}}$, $Z_{\cdot 1}^{(n)} = n^{-1/2} U_{\cdot 1} / \sqrt{\hat\sigma_{22}}$, and $\hat\sigma_{11}$ and $\hat\sigma_{22}$, defined in (24), are estimated variances of $n^{-1/2} U_{11}$ and $n^{-1/2} U_{\cdot 1}$, respectively. Li and Yang [17] showed that under the null hypothesis (1), $\mathbf{M}_n$ converges to a random variable $M = max(|Z_{11}|, |Z_{\cdot 1}|)$, where $Z_{11}, Z_{\cdot 1}$ have the bivariate normal distribution $N((0,0)^T, (1, 1, \rho))$. This allows one to obtain an approximate maximum test of H_0 based on $\mathbf{M_n}$.

2.2   General asymptotic theory under local alternatives

In this section, we establish the general asymptotic theory for the above joint tests under some contiguous proportional hazards alternatives for CSH_1 and ACH, which provides a step stone for developing the sample size determination methods in the following section. The contiguous proportional hazards alternatives for CSH_1 and ACH can be formulated as

$$H_1 : \quad \begin{aligned} \lambda_{11}^{(n)}(t) &= e^{\gamma_1\phi_1(t)/(2n^{1/2})}\lambda_{10}(t), \quad \lambda_{12}^{(n)}(t) = e^{-\gamma_1\phi_1(t)/(2n^{1/2})}\lambda_{10}(t), \\ \lambda_{\cdot 1}^{(n)}(t) &= e^{\gamma_\cdot\phi_\cdot(t)/(2n^{1/2})}\lambda_{\cdot 0}(t), \quad \lambda_{\cdot 2}^{(n)}(t) = e^{-\gamma_\cdot\phi_\cdot(t)/(2n^{1/2})}\lambda_{\cdot 0}(t), \end{aligned} \quad (4)$$

where either $\gamma_1 \neq 0$ or $\gamma_\cdot \neq 0$, $\phi_1(t)$ and $\phi_\cdot(t)$ are pre-specified and possibly time-varying functions, and $\lambda_{10}(t)$ and $\lambda_{\cdot 0}(t)$ are unspecified baseline cause-specific hazard and any-cause hazard, respectively. Gill [5] showed that a weighted log-rank test with a weight function converging to $\phi(t)$ gives the optimal power against the contiguous hazards alternative with a time-varying proportionality function $\phi(t)$. Therefore, we focus on weight functions $w_1(t) = \phi_1(t)$ and $w_\cdot(t) = \phi_\cdot(t)$ [2, 8].

Theorem 1   (a) *Under the sequence of local alternatives (4), as $n \to \infty$, $\mathbf{X}_n^2$ has an approximate non-central chi-square distribution with 2 degrees of freedom and non-centrality parameter*

$$\xi = \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}, \quad (5)$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2)$ and $\boldsymbol{\Sigma} = [\sigma_{ij}]$ are defined by

$$\begin{aligned} \mu_1 &= a_1 a_2 \gamma_1 \int_0^\tau \phi_1^2(t) dP_{10}(t), \\ \mu_2 &= a_1 a_2 \gamma_\cdot \int_0^\tau \phi_\cdot^2(t) dP_{\cdot 0}(t), \\ \sigma_{11} &= a_1 a_2 \int_0^\tau \phi_1^2(t) dP_{10}(t), \\ \sigma_{22} &= a_1 a_2 \int_0^\tau \phi_\cdot^2(t) dP_{\cdot 0}(t), \\ \sigma_{12} &= \sigma_{21} = a_1 a_2 \int_0^\tau \phi_1(t)\phi_\cdot(t) dP_{10}(t), \end{aligned} \quad (6)$$

$P_{10}(t) = \int_0^t y_0(u) d\Lambda_{10}(u)$ is the probability of observing an cause-1 failure by time t, $P_{\cdot 0}(t) = \int_0^t y_0(u) d\Lambda_{\cdot 0}(u)$ is the probability of observing an any-cause failure by time t, with $y_0(t) = S_0(t)S_c(t)$ and $S_0(t) = \exp\{-\int_0^t \lambda_{\cdot 0}(s) ds\}$.

(b) *Under the sequence of local alternatives (4), as $n \to \infty$, $\mathbf{M}_n$ converges in distribution to a random variable $M^* \equiv \max(|Z_{11}^*|, |Z_{\cdot 1}^*|)$, where*

$(Z_{11}^*, Z_{.1}^*)$ *has a bivariate normal distribution*

$$N\left(\left(\frac{\mu_1}{\sqrt{\sigma_{11}}}, \frac{\mu_2}{\sqrt{\sigma_{22}}}\right)^T, \left(1, 1, \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}\right)\right).$$

The proof of Theorem 1 is deferred to the appendix.

2.3 Sample size calculation for the chi-square joint test

The asymptotic theory in the above section enables one to develop sample size calculation methods for the joint tests. However, the resulting methods for time-varying alternatives would require estimation of multiple quantities that are difficult to interpret and thus inconvenient to use in practice. For the easy of interpretation and practical use, we will focus only on the simple case $\phi_1(t) = \phi_.(t) = 1$ from now on.

As pointed out by Eng and Kosorok [2], a power analysis is usually based on a fixed alternative rather than a contiguous alternative. However, for given fixed alternatives γ_1^* and $\gamma_.^*$, the log(cause-1 cause-specific hazard ratio) and the log(any-cause hazard ratio), the asymptotic results of Theorem 1 justify an approximate power calculation by setting $\gamma_1 = n^{1/2}\gamma_1^*$ and $\gamma_. = n^{1/2}\gamma_.^*$. Our sample size calculation formulas will account for two different types of censoring.: independent random censoring C_{ik} due to lost to follow-up and administrative censoring caused by staggered entry and end of the study.

2.3.1 Required number of cause-1 failures.

Under the above setting, it is easy to verify that (5) reduces to

$$\xi = D_1 a_1 a_2 \frac{\gamma_1^{\star 2} - 2\gamma_1^{\star}\gamma_.^{\star} + \gamma_.^{\star 2}/R}{1-R}, \tag{7}$$

where $D_1 = n \times P_{10}(\tau)$ is the total number of cause-1 failures by time τ and $R = P_{10}(\tau)/P_{.0}(\tau)$ is the relative risk of observing a cause-1 failure to an any-cause failure, which are approximately equal to the corresponding quantities in the pooled group for large n under the contiguous alternative hypothesis (4).

Therefore, for a given Type 1 error rate α and a power $1 - \beta$, the *required number of cause-1 failures* in the pooled group is approximately

$$D_1 = \frac{\xi(1-R)}{a_1 a_2 (\gamma_1^{\star 2} - 2\gamma_1^{\star}\gamma_.^{\star} + \gamma_.^{\star 2}/R)}, \tag{8}$$

6

where ξ is solved from the following equation

$$1 - \beta \;\; = \;\; P(\text{Reject } H_0|H_1) = P(\mathbf{X_n}^2 > \chi^2_{2,\alpha}|\mathbf{X_n}^2 \sim \chi^2_2(\xi)),$$

or

$$1 - \beta = 1 - P_{\chi^2_2}(\chi^2_{2,\alpha}; 2, \xi). \tag{9}$$

with $P_{\chi^2_2}(x; k, \xi)$ being the non-central chi-square cumulative distribution function with 2 degrees of freedom and non-centrality parameter ξ. In our R implementation, we use *pchisq()* to evaluate $P_{\chi^2_2}(x; k, \xi)$ and *uniroot()* to solve for ξ.

In summary, the *required number of cause-1 failures* is determined by (8) by three parameters: 1) the cause-1 cause-specific hazard ratio, $\exp(\gamma_1^\star)$, 2) the any-cause hazard ratio, $\exp(\gamma^\star)$, and 3) the relative risk R of observing a failure due to cause 1 to any cause in the pooled sample. As illustrated later, these parameters can be obtained by specifying the cumulative incidence rate at a pre-specified time for cause-1 failure and for any-cause failure in the control and experimental groups under the constant cause-specific hazard assumption for each cause.

2.3.2 Required number of patients.

In this section, we discuss how to determine the required number of patients for a trial with staggered entry, administrative censoring, and loss to follow-up. Let $f_{1k}(t)$ and $P_{1k}(t)$ be the density and cumulative incidence function for cause-1 failure in group k, respectively, $S_c(t)$ be the survival function of the independent censoring due to lost to follow-up, $f_r(t)$ be the density function of the entry time T_0, r be the length of the accrual period, and f be the total length of the study period. Let Q_{1k} denote the probability of observing a cause-1 failure in groups k by the end of the study. Then, the totally number of patients required is given by

$$N = D_1/\{a_1 \times Q_{11} + a_2 \times Q_{12}\}, \tag{10}$$

where for $k = 1, 2$,

$$Q_{1k} = P(T_{ik} < f - T_0, T_{ik} < C_{ik}) = \int_0^r f_r(z) \int_0^{f-z} f_{1k}(t)S_c(t)dtdz. \tag{11}$$

For example, if one assumes constant cause-specific hazard in each group k, uniform stagger entry over $[0, r]$, and exponential lost to follow-up with

7

constant hazard λ_c, then for $k = 1, 2$,

$$Q_{1k} = \frac{\lambda_{1k}}{\lambda_{.k} + \lambda_c} \left[1 - \frac{\exp\left[-(\lambda_{.k} + \lambda_c)(f - r)\right] - \exp\left[-(\lambda_{.k} + \lambda_c)f\right]}{(\lambda_{.k} + \lambda_c)r} \right], \quad (12)$$

where

$$
\begin{array}{rcl}
\lambda_{12} & = & exp(-\gamma_1^\star) \times \lambda_{11}, \\
\lambda_{.1} & = & exp(\frac{\gamma_.^\star - \gamma_1^\star}{2}) \times \lambda_{11}/R, \\
\lambda_{.2} & = & exp(-\gamma_.^\star) \times \lambda_{.1}.
\end{array}
\quad (13)
$$

In summary, the *total number of patients* needed to enroll in a study can be obtained from (10) and (11) by specifying the following quantities: 1) the length of accrual time r, 2) the maximum follow-up time f, 3) the patient proportion a_1 for group 1, 4) the hazard rate λ_c for loss to follow up, and 5) the cause specific hazard for cause 1 in group k $\lambda_{1k}, k = 1, 2$.

2.4 Sample size calculation for the maximum joint test

In this subsection, we present an algorithm to calculate the required number of cause-1 failures and the required number of patients based on the joint maximum joint test defined in (3).

Under the same assumptions of the previous section, it can be shown that $(Z_{11}, Z_{.1})$ have the following bivariate normal distribution

$$N\left(\left(\gamma_1^\star \sqrt{a_1 a_2 D_1}, \gamma_.^\star \sqrt{a_1 a_2 D_1/R} \right)^T, \left(1, 1, \sqrt{R} \right) \right), \quad (14)$$

where $R = P_{10}/P_{.0}$, the relative risk of failure due to cause 1 vs. any cause.

Given a Type I error rate α and a Type II error rate β, we have the following error equations:

$$\alpha = P(\text{Reject } H_0 | H_0) = P(\mathbf{M_n} > C_\alpha | H_0), \quad (15)$$

and

$$1 - \beta = P(\text{Reject } H_0 | H_a) = P(\mathbf{M_n} > C_\alpha | H_a), \quad (16)$$

where C_α is the critical value of the test. Let $f(x, y; \mu_1, \mu_2, \rho)$ be the bivariate normal density function with mean (μ_1, μ_2), variances $(1, 1)$, and correlation ρ. Then (15) and (16) can be rewritten as

$$\int_{-C_\alpha}^{C_\alpha} \int_{-C_\alpha}^{C_\alpha} f(x, y; 0, 0, \sqrt{R}) dx dy = \alpha, \quad (17)$$

8

and

$$\int_{-C_\alpha}^{C_\alpha} \int_{-C_\alpha}^{C_\alpha} f(x, y; \gamma_1^\star \sqrt{a_1 a_2 D_1}, \gamma_\cdot^\star \sqrt{a_1 a_2 D_1 / R}, \sqrt{R}) dx dy = \beta. \qquad (18)$$

We solve for C_α and D_1 sequentially from equations (17) and (18). In our R implementation, we use $pmvnorm()$ to evaluate a bivariate normal probability and $uniroot()$ to solve an equation.

Lastly, with staggered entry and lost to follow-up, the required number of patients in a trial is computed from (10) and (11) as in Section 2.3.2 .

3 Simulation studies

We present three simulations to illustrate the operating characteristics of the proposed methods. Competing risks data are generated by using Beyersmann et al. [1]'s cause-specific hazard driven method that requires only specification of the cause-specific hazard for each type of failure. We assume constant cause-specific hazard in each group k, uniform stagger entry over $[0, r]$, and exponential lost to follow-up with constant hazard λ_c in all simulations.

The first simulation compares the required sample sizes between the chi-square joint test, the maximum joint test, and the Bonferroni method under different effect size scenarios. Specifically, we consider three hazard ratios $1.7, 1.4, 1.2$, representing a large, medium, and small effect size respectively for CSH_1 and ACH. We assume equal number of patients in the two groups $(a_1 = a_2 = 0.5)$, a maximum follow-up time of $f = 10$, the length of accrual period $r = 1$, and the rate of random censoring (attrition) due to lost to follow-up $R_c = 5\%$. The hazard rate for lost to follow-up is calculated by $\lambda_c = \frac{R_c}{1-R_c} \frac{\lambda_{a1}+\lambda_{a2}}{2}$. We set $\lambda_{11} = 0.3$ and a relative risk $R = 0.8$. With $\alpha = 0.05$ and power $1 - \beta = 0.80$, Table 1 summarizes the required number of cause-1 failures and the required total number of patients for each of the three methods for various combinations of the cause-1 cause-specific hazard ratio $\exp(\gamma_1^\star)$ and the any-cause hazard ratio $\exp(\gamma_\cdot^\star)$. For each scenario, cells with the smallest number of failures or patients between the three methods are highlighted in grey.

As expected, the Bonferroni method is conservative, requiring the largest sample size in essentially all scenarios. The chi-square joint test tends to require the lowest number of failures and patients among the three methods when the CSH_1 and ACH hazard ratios are different. For instance, when

Table 1: Required sample sizes for the chi-square joint test, the maximum joint test and the Bonferroni method. ($\alpha = 0.05$, $1 - \beta = 0.8$, $\lambda_{11} = 0.3$ $R = 0.8$, $r = 1$, $f = 10$, and $R_c = 0.05$)

CSH_1	ACH	Number of Cause-1 Failures (D_1)			Number of Patients (N)		
Ratio	Ratio	Chi-Square	Max	Bonferrroni	Chi-Square	Max	Bonferrroni
1.2	1.2	928	794	916	1266	1082	1248
1.2	1.4	150	248	270	204	338	368
1.2	1.7	42	100	110	56	136	150
1.4	1.2	242	308	338	332	422	462
1.4	1.4	274	234	270	378	324	372
1.4	1.7	72	100	110	102	140	152
1.7	1.2	60	124	138	84	172	188
1.7	1.4	118	124	138	164	174	190
1.7	1.7	110	94	110	156	134	154

the CSH_1 hazard ratio is 1.2 and the ACH hazard ratio is 1.4, the chi-square joint test requires 150 cause-1 failures and 204 patients, which are substantially lower than the 248 failures and 338 patients required by the maximum joint test and the 270 failures and 368 patients by the Bonferroni method. On the other hand, when the CSH_1 and ACH hazard ratios are similar, the maximum joint test is observed to produce the most sample size savings.

In the second simulation we investigate the finite sample behavior of the proposed asymptotic sample size calculation methods by simulating their rejection powers under the scenarios considered in Table 1. The observed powers are reported in Table 2, which are close to the nominal power 0.80 across almost all cases considered.

In the third simulation, we explore how the attrition rate R_c due to lost to follow-up, maximum follow-up time f, and length of the accrual period r affect the sample size. Table 3 presents some simulation results under the scenario with $\exp \gamma_1^\star = 1.4$ and $\exp \gamma_\cdot^\star = 1.2$ from Table 1. It is seen that the three parameters, r, f and R_c, have no effect on the required number of cause-1 failures, but can impact the total number of required subjects

Table 2: Observed power for the chi-square joint test and the maximum joint test under the scenarios of Table 1. ($\alpha = 0.05$, $1-\beta = 0.8$, $\lambda_{11} = 0.3$ $R = 0.8$, $r = 1$, $f = 10$, and $R_c = 0.05$)

CSH_1	ACH	Chi-square		Maximum	
Ratio	Ratio	Sample Size (N)	Observed Power	Sample Size (N)	Observed Power
1.2	1.2	1266	0.80	1082	0.83
1.2	1.4	204	0.81	338	0.81
1.2	1.7	56	0.81	136	0.82
1.4	1.2	332	0.81	422	0.82
1.4	1.4	378	0.83	324	0.83
1.4	1.7	102	0.81	140	0.80
1.7	1.2	84	0.86	172	0.82
1.7	1.4	164	0.81	174	0.82
1.7	1.7	156	0.82	134	0.81

significantly. As expected, the required number of patients (N) increases when the attrition rate due to lost to follow-up is higher, the maximum follow-up time is shorter, or the accrual period is longer.

4   A real data example

We illustrate how to implement our method step-by-step using the 4D trial (Die Deutsche Disbetes Dialyse Studie) [24]. The 4D trial is a randomized, double-blinded, placebo-controlled trial to assess the efficacy of antihyperlipidemic treatment with atovastatin, in reducing occurrence of non-fatal myocardial infarction and cardiovascular mortality. There are three competing risks: non-fatal myocardial infarction (cause-n), death due to cardiovascular disease (cause-c), and death due to other causes (cause-o). As an illustration, we define the cause-1 failure of interest to be the composite event of either non-fatal myocardial infarction or death due to cardiovascular disease, and cause-2 failure to be death due to other causes. Schulgen et al. [22] illustrated nicely how to perform a power analysis for comparing the cause-1 cause-specific hazard (CSH_1) between the atovastatin and placeb groups.

Table 3: Required number of cause-1 failures (D_1) and total number of patients (N) for the chi-square joint test and the maximum joint test under different combinations of attrition rate (R_c) due to lost to follow-up, maximum follow-up time (f), and length of accrual period (r) ($\lambda_{11} = 0.3$, $R = 0.8$, $\exp(\gamma_1^\star) = 1.4$, $\exp(\gamma_\cdot^\star) = 1.2$)

			Number of Cause-1 Failures (D_1)		Number of Patients (N)	
R_c	f	r	Chi-square	Maximum	Chi-square	Maximum
5%	8	1	242	308	346	442
5%	8	1.5	242	308	396	506
5%	10	1	242	308	332	422
5%	10	1.5	242	308	354	452
10%	8	1	242	308	360	460
10%	8	1.5	242	308	406	518
10%	10	1	242	308	348	444
10%	10	1.5	242	308	366	468

Here we demonstrate how to re-design this trial based on a joint test of CSH_1 and ACH.

We first calculate the *required number of cause-1 failures* (non-fatal myocardial infection or cardiovascular death). As described in Section 2.3.1, the following quantities need to be specified: 1) cause-1 cause-specific hazard ratio $\exp(\gamma_1^\star)$, 2) any-cause hazard ratio $\exp(\gamma_\cdot^\star)$, and 3) relative risk R of observing a cause-1 failure to any cause.

Let $P_{c1}(t)$, $P_{o1}(t)$, and $P_{n1}(t)$ denote the probabilities of observing a cause-c, cause-o and cause-n failure, respectively, by time t in group 1. As in Schulgen et al. [22] , we use information from a perspective cohort study from 1985 to 1994 in Germany [7]. It was reported that the 4-year any-cause mortality rate was about 70% [7]. About 60% of the deaths were due to cardiovascular diseases [23]. This implies that $P_{c1}(4) = 0.7 * 0.6 = 0.42$ and $P_{o1}(4) = 0.7 * (1 - 0.6) = 0.21$. In addition, a 10% 4-year rate of non-fatal myocardial infarction was anticipated among diabetes patients, which means that $P_{n1}(4) = 0.10$. Since the primary outcome is time to either the occurrence of non-fatal myocardial infarction or death due to cardiovascular

disease, $P_{11}(4) = P_{c1}(4) + P_{n1}(4) = 0.52$, and $P_{\cdot1}(4) = P_{c1}(4) + P_{o1}(4) + P_{n1}(4) = 0.8$.

To propose appropriate effect sizes, $\gamma_1^\star$ and $\gamma_\cdot^\star$, one needs to specify the anticipated 4-year cause-1 cumulative incidence $P_{12}(4)$ and any-cause cumulative incidence $P_{\cdot2}(4)$ in group 2 (atovastatin). Schulgen et al. [22] assumed that the intervention is efficacious if it reduces the 4-year occurrence of the cause-1 failure from 52% to 42%. We assume further that a reduction of the 4-year any-cause incidence from 80% to 70% is clinically significant. Assuming a constant cause-specific hazard for each type of failure, then the above information can be converted to obtain all the required input parameters $\lambda_{11} = 0.26$, $\lambda_{12} = 0.18$, $\lambda_{\cdot1} = 0.4$ and $\lambda_{\cdot2} = 0.30$ by using equations $P_{jk} = \frac{\lambda_{jk}}{\lambda_{\cdot k}}(1 - e^{-\lambda_{\cdot k}t})$ and (13). Consequently, the expected CSH1 hazard ratio and ACH ratio are set as $\exp(\gamma_1^\star) = \frac{\lambda_{11}}{\lambda_{12}} = 1.44$ and $\exp(\gamma_\cdot^\star) = \frac{\lambda_{\cdot1}}{\lambda_{\cdot2}} = 1.33$, respectively. Furthermore, relative risk of a cause-1 failure versus an any cause failure in the pooled sample is approximated by $R = (\frac{0.26}{0.4} + \frac{0.42}{0.7})/2 = 0.625$.

We set type I error rate $\alpha = 0.05$, type II error rate $\beta = 0.2$, and equal patient allocation proportions $a_1 = a_2 = 0.5$. Table 4 gives the required number of cause-1 events (non-fatal myocardial infarction or death due to cardiovascular disease) and the total number of patients under different combinations of the attrition rate, maximum follow-up time, and accrual period. It is observed that the maximum joint test is most efficient design in this example with the fewest number of required cause-1 events and total number of patients. This is consistent with the observation in our simulation study (Table 1) that the maximum joint is more efficient when the cause-1 and any-cause hazard ratios are similar.

5   Discussion

Joint inference on multiple quantities is highly recommended for efficacy analysis of a clinical trial with competing risks data. The proposed method provides a power analysis tool for the design and planning of a clinical trial with competing risks based on some joint tests of the cause-specific hazard and the any-cause hazard. As shown in our simulations, the chi-square joint test generally leads to a more efficient design requiring fewer events and patients when the effect sizes for CSH_1 and ACH are different, whereas the maximum joint test tends to be more efficient when the effect sizes for CSH_1 and ACH are are similar. In practice we recommend that one perform power

Table 4: Required number of cause-1 failures (D_1) and number of patients (N) for the 4-D trial based on the chi-square joint test, maximum joint test, and Bonferroni joint test under different combinations of attrition Rate (R_c) due to lost to follow-up, maximum follow-up time (f), and length of the accrual period (r)

			Chi-Square Joint Test		Maximum Joint Test		Bonferroni Method	
R_c	f	r	D_1	N	D_1	N	D_1	N
5%	8	1	290	418	252	362	288	412
5%	8	1.5	290	486	252	420	288	478
5%	8	2	290	650	252	562	288	640
5%	10	1	290	400	252	348	288	396
5%	10	1.5	290	430	252	372	288	424
5%	10	2	290	488	252	422	288	482
10%	8	1	290	436	252	378	288	430
10%	8	1.5	290	496	252	430	288	488
10%	8	2	290	632	252	548	288	624
10%	10	1	290	420	252	364	288	414
10%	10	1.5	290	446	252	386	288	440
10%	10	2	290	494	252	428	288	488

analyses for both tests and then choose the most efficient design as illustrated on the 4-D trial in Section 4.

APPENDIX A.   Technical proofs

APPENDIX A.1.   Proof of Theorem 1

(a) Let $M_{1k}^{(n)}(\tau) = N_{1k}(\tau) - \int_0^\tau Y_k(t) d\Lambda_{1k}^{(n)}(t)$ and $M_{\cdot k}^{(n)}(\tau) = N_{\cdot k}(\tau) - \int_0^\tau Y_k(t) d\Lambda_{\cdot k}^{(n)}(t)$, where $\Lambda_{1k}^{(n)}(t)$ and $\Lambda_{\cdot k}^{(n)}(t)$ are the cumulative cause-specific hazard functions of $\lambda_{jk}^{(n)}(t)$ and $\lambda_{\cdot k}^{(n)}(t)$, respectively. Then, we can rewrite $U_{11}(t)$ and $U_{\cdot 1}(t)$ as follows:

$$
\begin{aligned}
U_{11} &= \int_0^\tau W_1(t) \tfrac{Y_1(t)Y_2(t)}{Y_\cdot(t)} \left\{ \tfrac{dM_{11}^{(n)}(t)}{Y_1(t)} - \tfrac{dM_{12}^{(n)}(t)}{Y_2(t)} \right\} \\
&+ \int_0^\tau W_1(t) \tfrac{Y_1(t)Y_2(t)}{Y_\cdot(t)} (d\Lambda_{11}^{(n)}(t) - d\Lambda_{12}^{(n)}(t)),
\end{aligned}
\tag{19}
$$

and

$$
\begin{aligned}
U_{\cdot 1} &= \int_0^\tau W_\cdot(t) \tfrac{Y_1(t)Y_2(t)}{Y_\cdot(t)} \left\{ \tfrac{dM_{\cdot 1}^{(n)}(t)}{Y_1(t)} - \tfrac{dM_{\cdot 2}^{(n)}(t)}{Y_2(t)} \right\} \\
&+ \int_0^\tau W_\cdot(t) \tfrac{Y_1(t)Y_2(t)}{Y_\cdot(t)} (d\Lambda_{\cdot 1}^{(n)}(t) - d\Lambda_{\cdot 2}^{(n)}(t)).
\end{aligned}
\tag{20}
$$

It follows from the multivariate martingale central limit theorem (Fleming and Harrington [4], Theorem 5.3.5) that under the contiguous alternatives (4), $n^{-1/2}(U_{11}, U_{\cdot 1})$ converges to (Z_1, Z_2) as $n \to \infty$, where (Z_1, Z_2) has a bivariate normal distribution with mean $\boldsymbol{\mu} = (\mu_1, \mu_2)$ and variance-covariance matrix $\boldsymbol{\Sigma} = (\sigma_{ij})$. Let $y_k(t)$ be the limiting value for $Y_k(t)/n_k, k = 1, 2$ when $n \to \infty$. Under the contiguous alternative and the assumption that the two groups have the same censoring distribution G, we have $y_1(t) = y_2(t) = y_0(t) = S_0(t)S_c(t)$, where $S_0(t) = \exp\{\int_0^t -\lambda_{\cdot 0}(s)ds\}$. Similar to Fleming and Harrington [4], Eng and Kosorok [2] , it can be shown that

$$
\begin{aligned}
\mu_1 &= \int_0^\tau \gamma_1 \phi_1^2(t) \tfrac{a_1 y_1(t) a_2 y_2(t)}{a_1 y_1(t) + a_2 y_2(t)} d\Lambda_{10}(t), \\
&= \gamma_1 a_1 a_2 \int_0^\tau \phi_1^2(t) y_0(t) d\Lambda_{10}(t), \\
&= \gamma_1 a_1 a_2 \int_0^\tau \phi_1^2(t) dP_{10}(t), \\
\mu_2 &= \gamma_\cdot a_1 a_2 \int_0^\tau \phi_\cdot^2(t) dP_{\cdot 0}(t), \\
\sigma_{11} &= a_1 a_2 \int_0^\tau \phi_1^2(t) dP_{10}(t), \\
\sigma_{22} &= a_1 a_2 \int_0^\tau \phi_\cdot^2(t) dP_{\cdot 0}(t),
\end{aligned}
\tag{21}
$$

where $\Lambda_{10}(t) = \int_0^t \lambda_{10}(u)du$, $\Lambda_{\cdot 0}(t) = \int_0^t \lambda_{\cdot 0}(u)du$, and $P_{10}(t) = \int_0^t y_0(u)d\Lambda_{10}(u)$ is the probability of observing a cause-1 failure by time t, and $P_{\cdot 0}(t) =$

$\int_0^t y_0(u)d\Lambda._0(u)$ is the probability of observing a failure due to any cause by time t. Furthermore, the covariance between $n^{-1/2}U_{11}$ and $n^{-1/2}U._1$ under the contiguous alternative hypothesis is

$$
\begin{aligned}
&\left\langle n^{-1/2}U_{11}, n^{-1/2}U._1 \right\rangle \\
&= n^{-1}\left\langle \int_0^\tau W_1(t)\frac{Y_1(t)Y_2(t)}{Y.(t)}\left(\frac{dM_{11}^n(t)}{Y_1(t)} - \frac{dM_{12}^{(n)}(t)}{Y_2(t)}\right), \int_0^\tau W.(t)\frac{Y_1(t)Y_2(t)}{Y.(t)}\left(\frac{dM._1^{(n)}(t)}{Y_1(t)} - \frac{dM._2^{(n)}(t)}{Y_2(t)}\right) \right\rangle \\
&= n^{-1}\left\{ \int_0^\tau W_1(t)W.(t)\frac{Y_1^2(t)Y_2^2(t)}{Y.(t)^2}\left(\frac{d<M_{11}^{(n)}(t),M._1^{(n)}(t)>}{Y_1(t)} + \frac{d<M_{12}^{(n)}(t),M._2^{(n)}(t)>}{Y_2(t)}\right) \right\} \\
&= n^{-1}\left\{ \int_0^\tau W_1(t)W.(t)\frac{Y_1^2(t)Y_2^2(t)}{Y.(t)^2}\left(\frac{d\Lambda_{11}^{(n)}(t)}{Y_1^2(t)} + \frac{d\Lambda_{12}^{(n)}(t)}{Y_2^2(t)}\right) \right\} \\
&= n^{-1}\left\{ \int_0^\tau W_1(t)W.(t)\frac{Y_1^2(t)Y_2^2(t)}{Y.(t)^2}\left(\frac{e^{\gamma_1\phi_1(t)/(2\sqrt{n})}d\Lambda_{10}(t)}{Y_1(t)} + \frac{e^{-\gamma_1\phi_1(t)/(2\sqrt{n})}d\Lambda_{10}(t)}{Y_2(t)}\right) \right\} \\
&\approx n^{-1}\int_0^\tau W_1(t)W.(t)\frac{Y_1(t)Y_2(t)}{Y.(t)}d\Lambda_{10}(t) \\
&\quad + n^{-1}\int_0^\tau W_1(t)W.(t)\frac{Y_1(t)Y_2(t)(Y_1(t)-Y_2(t))}{Y.^2(t)}\left(\gamma_1\phi_1(t)/\sqrt{n}d\Lambda_{10}(t)\right),
\end{aligned}
$$
(22)

which converges in probability to

$$
\begin{aligned}
\sigma_{12} &= \int_0^\tau \phi_1(t)\phi.(t)\frac{a_1a_2y_1(t)y_2(t)}{a_1y_1(t)+a_2y_2(t)}d\Lambda_{10}(t) \\
&= \int_0^\tau \phi_1(t)\phi.(t)a_1a_2y_0(t)d\Lambda_{10}(t),
\end{aligned}
$$
(23)

as $n \to \infty$.

Therefore, under the contiguous alternative hypothesis (4), $\mathbf{X}_n^2$ has an asymptotic chi-square distribution with 2 degrees of freedom and non-centrality parameter $\xi = \boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$. This proves part (a) of the theorem.

(b). Let $Z_{11}^{(n)} = n^{-1/2}U_{11}/\sqrt{\hat{\sigma}_{11}}$ and $Z._1^{(n)} = n^{-1/2}U._1/\sqrt{\hat{\sigma}_{22}}$, where

$$
\begin{aligned}
\hat{\sigma}_{11} &= n^{-1}\frac{a_1+a_2}{a_1a_2}\int_0^\tau W_1^2(t)\frac{Y_1(t)Y_2(t)}{Y_1(t)+Y_2(t)}\left(\frac{dN_{11}(t)+dN_{12}(t)}{Y_1(t)+Y_2(t)}\right), \\
\hat{\sigma}_{22} &= n^{-1}\frac{a_1+a_2}{a_1a_2}\int_0^\tau W.^2(t)\frac{Y_1(t)Y_2(t)}{Y_1(t)+Y_2(t)}\left(\frac{dN._1(t)+dN._2(t)}{Y_1(t)+Y_2(t)}\right)
\end{aligned}
$$
(24)

are consistent estimators of σ_{11} and σ_{22}, respectively. Again applying the martingale central limit theorem, it can be shown that under the contiguous alternatives (4), $(Z_{11}^{(n)}, Z._1^{(n)})$ converges to a bivariate normal random vector $(Z_{11}^\star, Z._1^\star)$ with mean $\left(\frac{\mu_1}{\sqrt{\sigma_{11}}}, \frac{\mu_2}{\sqrt{\sigma_{22}}}\right)$ and correlation $\frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$. By the continuous mapping theorem, we see that $\mathbf{M}_n$ converges to $M^\star = \max(|Z_{11}^\star|, |Z._1^\star|)$.
$\square$

APPENDIX A.2. Derivation of equation (5)

After plugging all the quantities from (6) to (5), we can get

$$
\begin{aligned}
\xi &= \frac{a_1 a_2 \left[\phi_1^{\star 2} n P_{10} - 2\phi_1^\star \phi_\cdot^\star n P_{10} + \phi_\cdot^{\star 2} n P_{\cdot 0}^\star \right]}{1 - n P_{10}/n P_{\cdot 0}} \\
&= \frac{a_1 a_2 \left[\phi_1^{\star 2} D_1 - 2\phi_1^\star \phi_\cdot^\star D_1 + \phi_\cdot^{\star 2} D_1/R \right]}{1 - R},
\end{aligned}
\tag{25}
$$

where $D_1 = n P_{10}$ and $D_\cdot = n P_{\cdot 0}$ are the number of failure due to cause 1 and due to any cause, respectively, R is the relative risk of failure due to cause 1 versus any cause, which is defined in Theorem 1. □

References

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

[2] Eng, K. H. and Kosorok, M. R. (2005). A sample size formula for the supremum log-rank statistic. *Biometrics* **61,** 86–91.

[3] 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,** 496–509.

[4] Fleming, T. and Harrington, D. (1991). *Counting processes and survival analysis*, volume 8. Wiley Online Library, New York City, US.

[5] Gill, R. D. (1980). Censoring and stochastic integrals. *Statistica Neerlandica* **34,** 124–124.

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

[7] Koch, M., Kutkuhn, B., Trenkwalder, E., Bach, D., Grabensee, B., Dieplinger, H., and Kronenberg, F. (1997). Apolipoprotein b, fibrinogen, hdl cholesterol, and apolipoprotein (a) phenotypes predict coronary artery disease in hemodialysis patients. *Journal of the American Society of Nephrology* **8,** 1889–1898.

[8] Kosorok, M. R. and Lin, C.-Y. (1999). The versatility of function-indexed weighted log-rank statistics. *Journal of the American Statistical Association* **94,** 320–332.

[9] Lachin, J. (1981). Introduction to sample size determination and power analysis for clinical trials. *Controlled Clinical Trials* **2,** 93–113.

[10] Lachin, J. and Foulkes, M. (1986). Evaluation of sample size and power for analyses of survival with allowance for nonuniform patient entry, losses to follow-up, noncompliance, and stratification. *Biometrics* **42,** 507–519.

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

[12] Lakatos, E. (1988). Sample sizes based on the log-rank statistic in complex clinical trials. *Biometrics* **44,** 229–241.

[13] Lakatos, E. and Lan, K. (1992). A comparison of sample size methods for the logrank statistic. *Statistics in medicine* **11,** 179–191.

[14] Latouche, A., Allignol, A., Beyersmann, J., Labopin, M., and Fine, J. P. (2013). A competing risks analysis should report results on all cause-specific hazards and cumulative incidence functions. *Journal of clinical epidemiology* **66,** 648–653.

[15] Latouche, A. and Porcher, R. (2007). Sample size calculations in the presence of competing risks. *Statistics in Medicine* **26,** 5370–5380.

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

[17] Li, G. and Yang, Q. (2015). Joint inference for competing risks survival data. *Journal of the American Statistical Association* **111,** 1289–1300.

[18] Pintilie, M. (2002). Dealing with competing risks: testing covariates and calculating sample size. *Statistics in Medicine* **21,** 3317–3324.

[19] Schmoor, C., Sauerbrei, W., and Schumacher, M. (2000). Sample size considerations for the evaluation of prognostic factors in survival analysis. *Statistics in medicine* **19,** 441–452.

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

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

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

[23] Tschöpe, W., Koch, M., Thomas, B., and Ritz, E. (2008). Serum lipids predict cardiac death in diabetic patients on maintenance hemodialysis. *Nephron* **64,** 354–358.

[24] Wanner, C., Krane, V., Ruf, G., März, W., and Ritz, E. (1999). Rationale and design of a trial improving outcome of type 2 diabetics on hemodialysis. *Kidney International* **56,** S222–S226.

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