

RESEARCH ARTICLE

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

Qing Yang¹ | Wing K. Fung² | Gang Li³ 

¹School of Nursing, Duke University, Durham, NC, USA

²Department of Statistics and Actuarial Science, University of Hong Kong, Pokfulam Road, Pokfulam, Hong Kong

³Department of Biostatistics, University of California, Los Angeles, Los Angeles, CA, USA

Correspondence

Gang Li, Department of Biostatistics, University of California, Los Angeles, Los Angeles, CA, USA.
Email: vli@ucla.edu

Funding information

National Institute of Health, Grant/Award Number: P30 CA-16042, UL1TR000124-02 and P01AT003960

This article considers sample size determination for jointly testing a cause-specific hazard and the all-cause hazard for competing risks data. The cause-specific hazard and the all-cause hazard jointly characterize important study end points such as the disease-specific survival and overall survival, which are commonly used as coprimary end points in clinical trials. Specifically, we derive sample size calculation methods for 2-group comparisons based on an asymptotic chi-square joint test and a maximum joint test of the aforementioned quantities, taking into account censoring due to lost to follow-up as well as staggered entry and administrative censoring. We illustrate the application of the proposed methods using the Die Deutsche Diabetes Dialyse Studies clinical trial. An R package “*powerCompRisk*” has been developed and made available at the CRAN R library.

KEYWORDS

competing risks, joint test, power analysis, sample size, 2-sample test

1 | INTRODUCTION

In a clinical trial of a new treatment targeting a specific disease, the disease-specific survival (DSS), defined as time to death due to the disease, and overall survival (OS), defined as time to death due to all causes, are often used as coprimary end points for studying how the treatment works on the disease and on a patient's OS, respectively. In such applications, simultaneous inference on DSS and OS can be facilitated using a competing risks model under which death due to other causes is a competing risk that is potentially correlated to DSS.¹⁻⁴ For convenience, we will refer death due to the disease of interest as cause 1 failure and death due to other causes as cause 2 failure. Then, the treatment effects on DSS and OS are characterized jointly by the cause 1 cause-specific hazard (CSH_1), the instantaneous risk of failure due to cause 1, and the all-cause hazard (ACH), the instantaneous risk of failure due to all causes. Among other methods, Li and Yang⁴ developed simultaneous inference procedures for joint analysis of CSH_1 and ACH and showed that their joint tests can be substantially more powerful than applying the simple Bonferroni adjustment with equally split alpha levels between the 2 end points. In this paper, we develop a power analysis tool for simultaneous group comparison of both CSH_1 and ACH based on the joint test procedures of Li and Yang⁴ to aid the design and planning of a clinical trial with competing risks data.

There is an extensive literature on sample size calculation for time-to-event data. For a right censored outcome with no competing risks, Schoenfeld^{5,6} proposed a sample size calculation formula for 2-sample comparison under uniform patient entry and administrative censoring. Lachin and Foulkes⁷ extended the formula to more complex situations, allowing for truncated exponential patient entry, loss to follow-up, noncompliance, and stratified analysis. Yatemán and Skene⁸

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 the previous literatures⁹⁻¹⁴ and the references therein. In the presence of competing risks, current sample size calculation methods are based on either a CSH alone^{15,16} or a subdistribution hazard alone.^{17,18} However, sample size calculation methods for joint 2-sample comparison of both a CSH and the ACH are not available.

The primary goal of the paper is to derive power analysis methods for simultaneous 2-sample testing of both CSH₁ and ACH for competing risks data. Specifically, in Section 2, we introduce a chi-square joint test and a maximum joint test for 2-group comparison of both CSH₁ and ACH, which were proposed in Li and Yang,⁴ and derive their asymptotic properties under contiguous local alternatives. The asymptotic results are then used to develop an approximate sample size determination procedure for each of the joint test. Both 2-sided and 1-sided tests are considered. Our sample size calculation formulas will account for 2 different types of censoring, independent random censoring due to lost to follow-up and administrative censoring caused by staggered entry and end of the study. In Section 3, we present simulations to compare the sample sizes required by the 2 tests and evaluate their achieved power performance. In Section 4, we use the Die Deutsche Diabetes Dialyse Studies (4-D) clinical trial to provide a step-by-step demonstration of how to implement our methods in practice. Further remarks are given in Section 5. Technical proofs are deferred in the Appendix. An R package “powerCompRisk” implementing the proposed sample size calculation methods is available at the CRAN R library.¹⁹

2 | SAMPLE SIZE CALCULATION FOR JOINT TESTS OF CSH₁ AND ACH

2.1 | Joint tests of CSH₁ and ACH

For the reader's convenience, we first review the chi-square joint test and the maximum joint test of CSH₁ and ACH that were proposed in Li and Yang.⁴

Suppose that there are 2 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, $i = 1, \dots, n_k, k = 1, 2$. 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 , $\{(T_{ik}, D_{ik}, C_{ik}), i = 1, \dots, n_k\}$ are independent and identically distributed and that the censoring time C_{ik} is independent of the failure time T_{ik} . Assume further that the 2 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 $\{(X_{ik}, \delta_{ik}), i = 1, \dots, n_k\}$, 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 2 causes of failure and that cause 1 failure is of primary interest.

Consider the following joint hypothesis of CSH₁ and ACH:

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

where

$$\lambda_{jk}(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T_{ik} \leq t + \Delta t, D_{ik} = j | T_{ik} \geq t)}{\Delta t}$$

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

The chi-square joint test statistic of Li and Yang⁴ for (1) is defined as

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

where

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

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

$N_{jk}(t) = \sum_{i=1}^{n_k} I(X_{ki} \leq t, D_{ki} = j)$ is the number of observed failures due to cause j in group k by time $t, Y_k(t) = \sum_{i=1}^{n_k} I\{X_{ki} \geq t\}$ is the number of patients in group k who are at risk just prior to time $t, N_{\cdot j}(t) = \sum_{k=1}^2 N_{jk}(t)$ and $Y_{\cdot}(t) = \sum_{k=1}^2 Y_k(t), \hat{\Sigma}$ is the estimated variance-covariance matrix of $n^{-1/2}(U_{11}, U_{\cdot 1}), \tau$ is the smallest t such that $Y_1(t)Y_2(t) = 0$, and $W_1(t)$ and $W_{\cdot}(t)$ are 2 predictable weight functions that converge in probability to some deterministic functions $w_1(t)$ and $w_{\cdot}(t)$ as $n \rightarrow \infty$.

It has been shown by Li and Yang⁴ 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 (A9), are estimated variances of $n^{-1/2}U_{11}$ and $n^{-1/2}U_{\cdot 1}$, respectively. Li and Yang⁴ 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 .

Both the chi-square joint test and the maximum joint test aim to test the same joint null hypothesis (1) based on the same bivariate test statistic $(U_{11}, U_{\cdot 1})$. However, they are based on different distance measures from $(0,0)$. The chi-square joint test statistic is a Mahalanobis distance whereas the maximum joint test statistic is a maximum type distance. Consequently, they are sensitive to different alternative hypotheses as illustrated in the third paragraph of Section 3. Another distinction between the chi-square joint test and the maximum joint test is that the former applies only to 2-sided alternatives whereas the latter can be adapted to 1-sided alternatives.

In the next 2 sections, we will present sample size calculation procedures for the chi-square joint test and the maximum joint test. For the ease of interpretation and practical use, we will consider the simple case with $W_1(t) = W_{\cdot}(t) = 1$ and assume the following commonly used proportional hazards alternatives

$$H_1 : \begin{aligned} \lambda_{11}^{(n)}(t) &= e^{\gamma_1^*/2} \lambda_{10}(t), & \lambda_{12}^{(n)}(t) &= e^{-\gamma_1^*/2} \lambda_{10}(t), \\ \lambda_{\cdot 1}^{(n)}(t) &= e^{\gamma^*/2} \lambda_{\cdot 0}(t), & \lambda_{\cdot 2}^{(n)}(t) &= e^{-\gamma^*/2} \lambda_{\cdot 0}(t), \end{aligned} \quad (\text{either } \gamma_1^* \neq 0 \text{ or } \gamma^* \neq 0), \tag{4}$$

where γ_1^* is the log(CSH₁ ratio), γ^* is the log(ACH ratio), and $\lambda_{10}(t)$ and $\lambda_{\cdot 0}(t)$ are unspecified baseline CSH and ACH, respectively.

2.2 | Sample size calculation for the chi-square joint test

2.2.1 | Required number of cause 1 failures

We first consider sample size calculation for the chi-square joint test. To derive the sample size, we consider a sequence of local alternatives to the null hypothesis, under which γ_1^* and γ^* are of the order $O(n^{-1/2})$. It follows from part (a) of Theorem 1 in Appendix A.1 that under the local alternatives, \mathbf{X}_n^2 has an approximate noncentral chi-square distribution with 2 degrees of freedom and noncentrality parameter

$$\xi = D_1 a_1 a_2 \frac{\gamma_1^{*2} - 2\gamma_1^* \gamma^* + \gamma^{*2} / R}{1 - R}, \tag{5}$$

where $D_1 = n \times P_{10}(\tau)$ is the total number of cause 1 failures by time τ in the pooled sample across treatment groups and $R = P_{10}(\tau) / P_{\cdot 0}(\tau)$ is the ratio of the cumulative incidence of failure due to cause 1 to cumulative incidence for failure due to all causes by time τ , which are approximately equal to the corresponding quantities in the pooled sample across treatment groups for large n under the local alternatives. Detailed information on derivation for Equation 5 is presented in Appendix A.2.

Therefore, for a given type 1 error rate α and a power $1 - \beta$, the *required number of cause 1 failures* in the pooled sample across treatment groups is approximately

$$D_1 = \frac{\xi(1-R)}{a_1 a_2 (\gamma_1^{*2} - 2\gamma_1^* \gamma^* + \gamma^{*2} / R)}, \tag{6}$$

where ξ is solved from the following equation

$$1 - \beta = P(\text{Reject } H_0 | H_1) = P(\mathbf{X}_n^2 > \chi_{2,\alpha}^2 | \mathbf{X}_n^2 \sim \chi_2^2(\xi)),$$

or

$$1 - \beta = 1 - P_{\chi_2^2}(\chi_{2,\alpha}^2; 2, \xi), \tag{7}$$

with $P_{\chi_2^2}(x; k, \xi)$ being the noncentral chi-square cumulative distribution function with 2 degrees of freedom and noncentrality parameter ξ .

In summary, the *required number of cause 1 failures* is determined by (6) by 3 parameters: (1) the CSH₁ ratio, $\exp(\gamma_1^*)$, (2) the ACH ratio, $\exp(\gamma^*)$, and (3) R , the ratio of cumulative incidence for failure due to cause 1 to cumulative incidence for failure due to any cause in the pooled sample across treatment groups. As illustrated later, these parameters can be

obtained by specifying the cumulative incidence rate at a prespecified time for cause 1 failure and for any-cause failure in the control and experimental groups under the constant CSH assumption for each cause.

2.2.2 | Required number of patients

We next 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 τ 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

$$N = D_1 / \{a_1 \times Q_{11} + a_2 \times Q_{12}\}, \quad (8)$$

where for $k = 1, 2$,

$$Q_{1k} = P(T_{ik} < \tau - T_0, T_{ik} < C_{ik}) = \int_0^r f_r(z) \int_0^{\tau-z} f_{1k}(t) S_c(t) dt dz. \quad (9)$$

In particular, if one further assumes constant CSH in each group k , uniform stagger entry over $[0, r]$, and exponential lost to follow-up with constant hazard λ_c , then for $k = 1, 2$,

$$Q_{1k} = \frac{\lambda_{1k}}{\lambda_{.k} + \lambda_c} \left[1 - \frac{\exp[-(\lambda_{.k} + \lambda_c)(\tau - r)] - \exp[-(\lambda_{.k} + \lambda_c)\tau]}{(\lambda_{.k} + \lambda_c)r} \right], \quad (10)$$

where

$$\begin{aligned} \lambda_{12} &= \exp(-\gamma_1^*) \times \lambda_{11}, \\ \lambda_{.1} &= \exp\left(\frac{\gamma_1^* - \gamma_1^*}{2}\right) \times \lambda_{11} / R, \\ \lambda_{.2} &= \exp(-\gamma_1^*) \times \lambda_{.1}. \end{aligned} \quad (11)$$

In summary, the *total number of patients* needed to enroll in a study can be obtained from (8), (10), and (11) by specifying the following quantities: (1) the length of accrual time r , (2) the maximum follow-up time τ , (3) the patient proportion a_1 for group 1, (4) the hazard rate λ_c for loss to follow-up, and (5) the CSH for cause 1 in group k λ_{1k} , $k = 1, 2$.

2.3 | Sample size calculation for the maximum joint test

We now present an algorithm to calculate the required number of cause 1 failures and the required number of patients based on the maximum joint test defined in Equation 3.

2.3.1 | Required number of cause 1 failures

Under the same local alternatives as in the previous section, it can be shown from part (b) of Theorem 1 in Appendix A.1 that $(Z_{11}, Z_{.1})$ have approximately the following bivariate normal distribution

$$N \left(\left(\gamma_1^* \sqrt{a_1 a_2 D_1}, \gamma_1^* \sqrt{a_1 a_2 D_1 / R} \right)^T, (1, 1, \sqrt{R}) \right), \quad (12)$$

where $D_1 = n \times P_{10}(\tau)$ is the total number of cause 1 failures by time τ in the pooled sample across treatment groups and $R = P_{10}(\tau) / P_{.0}(\tau)$ is the ratio of the cumulative incidence of failure due to cause 1 to cumulative incidence of failure due to all causes by time τ .

Hence, given a type 1 error rate α and a type 2 error rate β , the error equations

$$\alpha = P(\text{Reject } H_0 | H_0) = P(\mathbf{M}_n > C_\alpha | H_0) \text{ and } \beta = P(\text{Accept } H_0 | H_a) = P(\mathbf{M}_n \leq C_\alpha | H_a)$$

can be rewritten as

$$\int_{-C_\alpha}^{C_\alpha} \int_{-C_\alpha}^{C_\alpha} f(x, y; 0, 0, \sqrt{R}) dx dy = 1 - \alpha \quad (13)$$

and

$$\int_{-C_\alpha}^{C_\alpha} \int_{-C_\alpha}^{C_\alpha} f(x, y; \gamma_1^* \sqrt{a_1 a_2 D_1}, \gamma_1^* \sqrt{a_1 a_2 D_1 / R}, \sqrt{R}) dx dy = \beta, \tag{14}$$

where C_α is the critical value of M_n and $f(x, y; \mu_1, \mu_2, \rho)$ denotes the bivariate normal density function with mean (μ_1, μ_2) , variances $(1,1)$, and correlation ρ .

In practice, one will first need to solve for C_α from (13) and then solve for D_1 from (14) using a numerical root finding routine. In summary, to obtain the *required total number of cause 1 failures* D_1 for the maximum joint test as a root of (14), one needs to input the following 3 parameters into Equations 13 and 14: (1) the CSH₁ ratio, $\exp(\gamma_1^*)$, (2) the ACH ratio, $\exp(\gamma^*)$, and (3) R , the ratio of cumulative incidence of failure due to cause 1 to cumulative incidence of failure due to all causes in the pooled sample across treatment groups.

2.3.2 | Required number of patients

The required number of patients for the maximum joint test is calculated exactly the same way as for the chi-square joint test as outlined in Section 2.2.2. In particular, if one assumes uniform staggered entry and constant hazards for all causes and loss to follow-up, then the *total number of patients* needed to enroll in a study is obtained from (8), (10), and (11) by specifying the following quantities: (1) the length of accrual time r , (2) the maximum follow-up time τ , (3) the patient proportion a_1 for group 1, (4) the hazard rate λ_c for loss to follow-up, and (5) the CSH for cause 1 in group k $\lambda_{1k}, k = 1, 2$.

Finally, we have implemented the above proposed procedures in an R package “*powerCompRisk*” with a detailed documentation and examples.¹⁹

3 | SIMULATION STUDIES

We present 3 simulations to illustrate the operating characteristics of the proposed methods. Competing risks data are generated by using the CSH-driven method²⁰ that requires only specification of the CSH for each type of failure. We assume constant CSH in each group k , uniform stagger entry over $[0, r]$, and exponential lost to follow-up with constant hazard λ_c in all simulations.

In the first simulation, we compare the required sample sizes between the chi-square joint test and the maximum joint test under different effect size scenarios by varying the CSH₁ ratio and the ACH ratio, where in Table 1, a hazard ratio of 1.2, 1.4, or 1.7 represents a small, medium, or large effect size. For each scenario, the implied CSH₂ ratio for the competing event as well as the cumulative incidence functions (by cause and group) is provided in Appendix B. We assume equal number of patients in the 2 groups ($a_1 = a_2 = 0.5$), a maximum follow-up time of $\tau = 10$, 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 $\lambda_c = \frac{R_c \cdot \lambda_{a1} + \lambda_{a2}}{1 - R_c}$. We set $\lambda_{11} = 0.3$ and $R = 0.8$. With $\alpha = 0.05$ and power $1 - \beta = 0.80$, Table 1 summarizes

TABLE 1 Required sample sizes for the chi-square joint test and the maximum joint test with nominal significance level $\alpha = 0.05$ and power $1 - \beta = 0.8$

CSH ₁ Ratio	ACH Ratio	Number of cause 1 failures (D_1)		Number of patients (N)	
		Chi-square	Max	Chi-square	Max
1.2	1.2	928	794	1266	1082
1.2	1.4	150	248	204	338
1.2	1.7	42	100	56	136
1.4	1.2	242	308	332	422
1.4	1.4	274	234	378	324
1.4	1.7	72	100	102	140
1.7	1.2	60	124	84	172
1.7	1.4	118	124	164	174
1.7	1.7	110	94	156	134

Abbreviations: ACH, all-cause hazard; CSH₁, cause 1 cause-specific hazard. Note: The cause-specific hazard for cause 1 in group 1 is $\lambda_{11} = 0.3$, the ratio of cumulative incidence of failure due to cause 1 to all causes is $R = 0.8$, the length of recruitment period is $r = 1$, the length of the follow-up time is $\tau = 10$, and the censoring rate due to loss of follow-up is $R_c = 0.05$.

the required number of cause 1 failures and the required total number of patients for each of the 3 methods for various combinations of the CSH₁ ratio $\exp(\gamma_1^*)$ and the ACH ratio $\exp(\gamma^*)$. For each scenario, cells with the smallest number of failures or patients between the 3 methods are highlighted in grey.

It is observed from Table 1 that the chi-square joint test tends to require a fewer number of failures and patients among the 3 methods when the CSH₁ and ACH ratios are different. For instance, when the CSH₁ ratio is 1.2 and the ACH 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. On the other hand, when the CSH₁ and ACH ratios are similar, the maximum joint test is observed to produce more 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 τ , and length of the accrual period r affect the sample size. Table 3 presents some simulation results under the scenario with $\exp(\gamma_1^*) = 1.4$ and $\exp(\gamma^*) = 1.2$ from Table 1. It is seen that the 3 parameters, r , τ , and R_c , have no effect on the required number of cause 1 failures, but can impact the total number of required subjects 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 (τ) is shorter, or the accrual period (r) is longer. In particular, when the follow-up time is shorter, the length of accrual period has a bigger impact on the sample size. For instance, for the chi-square test with attrition rate fixed at $R_c = 5\%$ in Table 3, increasing the length of accrual period from $r = 1$ to $r = 2$ would result in $396 - 346 = 50$ more patients when $\tau = 8$ as compared with $354 - 332 = 22$ more patients when $\tau = 10$.

TABLE 2 Observed power for the chi-square joint test and the maximum joint test with nominal significance level $\alpha = 0.05$ and power $1 - \beta = 0.8$ under the settings of Table 1

CSH ₁ Ratio	ACH Ratio	Chi-square		Maximum	
		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

Abbreviations: ACH, all-cause hazard; CSH₁, cause 1 cause-specific hazard.

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 (τ), and length of accrual period (r) ($\lambda_{11} = 0.3$, $R = 0.8$, $\exp(\gamma_1^*) = 1.4$, $\exp(\gamma^*) = 1.2$)

$R_c, \%$	τ	r	Number of Cause 1 failures (D_1)		Number of patients (N)	
			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

4 | A REAL DATA EXAMPLE

We illustrate how to implement our method step-by-step using the 4-D trial.²¹ The 4-D trial is a randomized, double-blinded, placebo-controlled trial to assess the efficacy of antihyperlipidemic treatment with atorvastatin, in reducing occurrence of nonfatal myocardial infarction and cardiovascular mortality. There are 3 competing risks: nonfatal 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 nonfatal myocardial infarction or death due to cardiovascular disease and cause 2 failure to be death due to other causes. Schulgen et al¹⁶ illustrated nicely how to perform a power analysis for comparing the CSH₁ between the atorvastatin and placebo groups. Here we demonstrate how to redesign this trial on the basis of a joint test of CSH₁ and ACH.

We first calculate the *required number of cause 1 failures* (nonfatal myocardial infection or cardiovascular death). As described in Section 2.2.1, the following quantities need to be specified: (1) CSH₁ ratio $\exp(\gamma_1^*)$, (2) ACH ratio $\exp(\gamma^*)$, and (3) R , the ratio of cumulative incidence of failure due to cause 1 failure to all causes.

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,¹⁶ we use information from a perspective cohort study from 1985 to 1994 in Germany.²² It was reported that the 4-year any-cause mortality rate²² was about 70%. 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 nonfatal 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 nonfatal myocardial infarction or death due to cardiovascular disease, $P_{11}(4) = P_{c1}(4) + P_{n1}(4) = 0.52$ and $P_{.1}(4) = P_{c1}(4) + P_{o1}(4) + P_{n1}(4) = 0.8$.

To propose appropriate effect sizes, γ_1^* and γ^* , one needs to specify the anticipated 4-year cause 1 cumulative incidence $P_{12}(4)$ and any-cause cumulative incidence $P_{.2}(4)$ in group 2 (atorvastatin). Schulgen et al¹⁶ 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 CSH for each type of failure, then the above information can be converted to obtain all the required input parameters $\lambda_{11} = 0.26$, $\lambda_{12} = 0.18$, $\lambda_{.1} = 0.4$, and $\lambda_{.2} = 0.30$ by using equations $P_{jk} = \frac{\lambda_{jk}}{\lambda_k} (1 - e^{-\lambda_k t})$ and (11). Consequently, the expected CSH₁ ratio and ACH ratio are set as $\exp(\gamma_1^*) = \frac{\lambda_{11}}{\lambda_{12}} = 1.44$ and $\exp(\gamma^*) = \frac{\lambda_{.1}}{\lambda_{.2}} = 1.33$, respectively. Furthermore, ratio of cumulative incidence for a cause 1 failure versus an all-cause failure in the pooled sample across treatment groups is approximated by $R = (\frac{0.26}{0.4} + \frac{0.42}{0.7})/2 = 0.625$.

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 and maximum joint test, under different combinations of attrition rate (R_c) due to lost to follow-up, maximum follow-up time (τ), and length of the accrual period (r)

$R_c, \%$	τ	r	Chi-square joint test		Maximum joint test	
			D_1	N	D_1	N
5	8	1	290	418	252	362
5	8	1.5	290	486	252	420
5	8	2	290	650	252	562
5	10	1	290	400	252	348
5	10	1.5	290	430	252	372
5	10	2	290	488	252	422
10	8	1	290	436	252	378
10	8	1.5	290	496	252	430
10	8	2	290	632	252	548
10	10	1	290	420	252	364
10	10	1.5	290	446	252	386
10	10	2	290	494	252	428

We set type 1 error rate $\alpha = 0.05$, type 2 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 (nonfatal 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 in this example, the design based on the maximum joint test would be more efficient with fewer 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 CSH₁ and ACH ratios are similar.

5 | DISCUSSION

Joint inference on a CSH and the ACH is often required for efficacy analysis of a clinical trial with competing risks data. We have developed power analysis tools for the design and planning of a clinical trial with competing risks for simultaneous 2-treatment comparisons of a CSH and the ACH based on either a chi-square joint test or a maximum joint test. Our simulations suggest that the chi-square joint test tends to require fewer events and patients than the maximum joint test when the effect sizes for CSH₁ and ACH are different, whereas the maximum joint test tends to be more efficient when the effect sizes for CSH₁ and ACH are similar. In practice, it would be helpful to perform power analyses for both tests as illustrated in Section 4 with the 4-D trial.

Our method assumes the common proportional hazards alternatives that facilitates an easily implementable and interpretable sample size calculation method. We also incorporate random censoring due to lost to follow-up, staggered patient entry, and administrative censoring. In practice, one may encounter more complicated circumstances such as nonproportional hazards, treatment crossover, and more than 2 treatment arms, which may impact the required sample size as illustrated by Barthel et al.²⁴ There are also situations where there are several CSHs of importance.²⁵ Although it is possible to extend our methods to the abovementioned complex settings along the lines of this paper and the previous literatures,^{4,24,25} the resulting formulas and their interpretations will be much more complicated. Further investigations are warranted in future research.

ACKNOWLEDGEMENT

The research of Gang Li was partly supported by National Institute of Health grants P30 CA-16042, UL1TR000124-02, and P01AT003960.

ORCID

Gang Li  <http://orcid.org/0000-0002-4753-9420>

REFERENCES

1. Gray RJ. A class of K-sample tests for comparing the cumulative incidence of a competing risk. *The Ann Stat*. 1988;16(3):1141-1154.
2. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94(446):496-509.
3. Latouche A, Allignol A, Beyersmann J, Labopin M, Fine JP. A competing risks analysis should report results on all cause-specific hazards and cumulative incidence functions. *J Clin Epidemiol*. 2013;66(6):648-653.
4. Li G, Yang Q. Joint inference for competing risks survival data. *J Am Stat Assoc*. 2015;111(515):1289-1300.
5. Schoenfeld D. The asymptotic properties of nonparametric-tests for comparing survival distributions. *Biometrika*. 1981;68(1):316-319.
6. Schoenfeld DA. Sample-size formula for the proportional-hazards regression-model. *Biometrics*. 1983;39(2):499-503.
7. Lachin JM, Foulkes MA. Evaluation of sample size and power for analyses of survival with allowance for nonuniform patient entry, losses to follow-up, noncompliance, and stratification. *Biometrics*. 1986;42(3):507-519.
8. Yatemana NA, Skene AM. Sample sizes for proportional hazards survival studies with arbitrary patient entry and loss to follow-up distributions. *Stat Med*. 1992;11(8):1103-1113.
9. Lachin JM. Introduction to sample size determination and power analysis for clinical trials. *Controlled Clin Trials*. 1981;2(2):93-113.
10. Lakatos E. Sample-size determination in clinical-trials with time-dependent rates of losses and noncompliance. *Controlled Clin Trials*. 1986;7(3):189-199.
11. Lakatos E. Sample sizes based on the log-rank statistic in complex clinical trials. *Biometrics*. 1988;44(1):229-241.
12. Lakatos E, Lan KK. A comparison of sample size methods for the logrank statistic. *Stat Med*. 1992;11(2):179-191.
13. Schmoor C, Sauerbrei W, Schumacher M. Sample size considerations for the evaluation of prognostic factors in survival analysis. *Stat Med*. 2000;19(4):441-452.

14. Eng KH, Kosorok MR. A sample size formula for the supremum log-rank statistic. *Biometrics*. 2005;61(1):86-91.
15. Pintilie M. Dealing with competing risks: testing covariates and calculating sample size. *Stat Med*. 2002;21:3317-3324.
16. Schulgen G, Olschewski M, Krane V, Wanner C, Ruf G, Schumacher M. Sample sizes for clinical trials with time-to-event endpoints and competing risks. *Contemp Clin Trials*. 2005;26(3):386-396.
17. Latouche A, Porcher R, Chevret S. Sample size formula for proportional hazards modelling of competing risks. *Stat Med*. 2004;23(21):3263-3274.
18. Latouche A, Porcher R. Sample size calculations in the presence of competing risks. *Stat Med*. 2007;26(30):5370-5380.
19. Yang Q, Fung WK, Kawaguchi E, Li G. powerCompRisk: Power analysis tool for joint testing hazards with competing risks data. *R package version 0.1.0* (<https://cran.r-project.org/web/packages/powerCompRisk/index.html>). 2017.
20. Beyersmann J, Latouche A, Buchholz A, Schumacher M. Simulating competing risks data in survival analysis. *Stat Med*. 2009;28(6):956-971.
21. Wanner C, Krane V, Ruf G, März W, Ritz E. Rationale and design of a trial improving outcome of type 2 diabetics on hemodialysis. *Kidney Int*. 1999;56:S222-S226.
22. Koch M, Kutkuhn B, Trenkwalder E, et al. Apolipoprotein B, fibrinogen, HDL cholesterol, and apolipoprotein (a) phenotypes predict coronary artery disease in hemodialysis patients. *J Am Soc Nephrol*. 1997;8(12):1889-1898.
23. Tschöpe W, Koch M, Thomas B, Ritz E. Serum lipids predict cardiac death in diabetic patients on maintenance hemodialysis. *Nephron*. 2008;64(3):354-358.
24. Barthel FM-S, Babiker A, Royston P, Parmar MKB. Evaluation of sample size and power for multi-arm survival trials allowing for non-uniform accrual, non-proportional hazards, loss to follow-up and cross-over. *Stat Med*. 2006;25(15):2521-2542.
25. Luo X, Turnbull BW. Comparing two treatments with multiple competing risks endpoints. *Statistica Sin*. 1999;9(4):985-998.
26. Gill RD. Censoring and stochastic integrals. *Statistica Neerlandica*. 1980;34(2):124-124.
27. Kosorok MR, Lin C-Y. The versatility of function-indexed weighted log-rank statistics. *J Am Stat Assoc*. 1999;94(445):320-332.
28. Fleming TR, Harrington DP. *Counting Processes and Survival Analysis*, Vol. 8. New York City, US: Wiley Online Library; 1991.

How to cite this article: Yang Q, Fung WK, Li G. Sample size determination for jointly testing a cause-specific hazard and the all-cause hazard in the presence of competing risks. *Statistics in Medicine*. 2017;1-13. <https://doi.org/10.1002/sim.7590>

APPENDIX A: GENERAL ASYMPTOTIC THEORY UNDER LOCAL ALTERNATIVES AND TECHNICAL DERIVATIONS

A.1 | General asymptotic theory under local alternatives

We establish the general asymptotic theory for the 2 joint tests described in Section 2.1 under some contiguous proportional hazards alternatives for cause 1 cause-specific hazard (CSH₁) and all-cause hazard (ACH), which provides a stepping stone for developing the sample size determination methods. The contiguous proportional hazards alternatives for CSH₁ and ACH can be formulated as

$$H_1 : \lambda_{11}^{(n)}(t) = e^{\gamma_1 \phi_1(t)/(2n^{1/2})} \lambda_{10}(t), \quad \lambda_{12}^{(n)}(t) = e^{-\gamma_1 \phi_1(t)/(2n^{1/2})} \lambda_{10}(t), \quad (A1)$$

$$\lambda_{-1}^{(n)}(t) = e^{\gamma \phi(t)/(2n^{1/2})} \lambda_{-0}(t), \quad \lambda_{-2}^{(n)}(t) = e^{-\gamma \phi(t)/(2n^{1/2})} \lambda_{-0}(t),$$

where either $\gamma_1 \neq 0$ or $\gamma \neq 0$, $\phi_1(t)$ and $\phi(t)$ are prespecified and possibly time-varying functions and $\lambda_{10}(t)$ and $\lambda_{-0}(t)$ are unspecified baseline CSH and ACH, respectively. Gill²⁶ 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 $w(t) = \phi(t)$.^{14,27}

Theorem 1.

- (a) Under the sequence of local alternatives (A1), as $n \rightarrow \infty$, \mathbf{X}_n^2 has an approximate noncentral chi-square distribution with 2 degrees of freedom and noncentrality parameter

$$\xi = \mu^T \Sigma^{-1} \mu, \tag{A2}$$

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

$$\begin{aligned} \mu_1 &= a_1 a_2 \gamma_1 \int_0^\tau \phi_1^2(t) dP_{10}(t), \\ \mu_2 &= a_1 a_2 \gamma \int_0^\tau \phi^2(t) dP_{\cdot 0}(t), \\ \sigma_{11} &= a_1 a_2 \int_0^\tau \phi_1^2(t) dP_{10}(t), \\ \sigma_{22} &= a_1 a_2 \int_0^\tau \phi^2(t) dP_{\cdot 0}(t), \\ \sigma_{12} = \sigma_{21} &= a_1 a_2 \int_0^\tau \phi_1(t) \phi(t) dP_{10}(t), \end{aligned} \tag{A3}$$

$P_{10}(t) = \int_0^t y_0(u) d\Lambda_{10}(u)$ is the cumulative incidence of failure due to cause 1 by time t , $P_{\cdot 0}(t) = \int_0^t y_0(u) d\Lambda_{\cdot 0}(u)$ is the cumulative incidence due to all cause 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 (A1), as $n \rightarrow \infty$, M_n converges in distribution to a random variable $M^* \equiv \max(|Z_{11}^*|, |Z_{\cdot 1}^*|)$, where $(Z_{11}^*, Z_{\cdot 1}^*)$ has a bivariate normal distribution

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

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 CSH functions of $\lambda_{1k}^{(n)}(t)$ and $\lambda_{\cdot k}^{(n)}(t)$, respectively. Then, we can rewrite $U_{11}(t)$ and $U_{\cdot 1}(t)$ as follows:

$$\begin{aligned} U_{11} &= \int_0^\tau W_1(t) \frac{Y_1(t)Y_2(t)}{Y(t)} \left\{ \frac{dM_{11}^{(n)}(t)}{Y_1(t)} - \frac{dM_{12}^{(n)}(t)}{Y_2(t)} \right\} \\ &+ \int_0^\tau W_1(t) \frac{Y_1(t)Y_2(t)}{Y(t)} (d\Lambda_{11}^{(n)}(t) - d\Lambda_{12}^{(n)}(t)) \end{aligned} \tag{A4}$$

and

$$\begin{aligned} U_{\cdot 1} &= \int_0^\tau W_{\cdot}(t) \frac{Y_1(t)Y_2(t)}{Y(t)} \left\{ \frac{dM_{\cdot 1}^{(n)}(t)}{Y_1(t)} - \frac{dM_{\cdot 2}^{(n)}(t)}{Y_2(t)} \right\} \\ &+ \int_0^\tau W_{\cdot}(t) \frac{Y_1(t)Y_2(t)}{Y(t)} (d\Lambda_{\cdot 1}^{(n)}(t) - d\Lambda_{\cdot 2}^{(n)}(t)). \end{aligned} \tag{A5}$$

It follows from the multivariate martingale central limit theorem (Fleming and Harrington²⁸, Theorem 5.3.5) that under the contiguous alternatives (A1), $n^{-1/2}(U_{11}, U_{\cdot 1})$ converges to (Z_1, Z_2) as $n \rightarrow \infty$, where (Z_1, Z_2) has a bivariate normal distribution with mean $\mu = (\mu_1, \mu_2)$ and variance-covariance matrix $\Sigma = (\sigma_{ij})$. Let $y_k(t)$ be the limiting value for $Y_k(t)/n_k$, $k = 1, 2$ when $n \rightarrow \infty$. Under the contiguous alternative and the assumption that the 2 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 the previous literatures,^{14,28} it can be shown that

$$\begin{aligned}
 \mu_1 &= \int_0^\tau \gamma_1 \phi_1^2(t) \frac{a_1 y_1(t) a_2 y_2(t)}{a_1 y_1(t) + a_2 y_2(t)} d\Lambda_{10}(t), \\
 &= \gamma_1 a_1 a_2 \int_0^\tau \phi_1^2(t) y_0(t) d\Lambda_{10}(t), \\
 &= \gamma_1 a_1 a_2 \int_0^\tau \phi_1^2(t) dP_{10}(t), \\
 \mu_2 &= \gamma_1 a_1 a_2 \int_0^\tau \phi^2(t) dP_{\cdot 0}(t), \\
 \sigma_{11} &= a_1 a_2 \int_0^\tau \phi_1^2(t) dP_{10}(t), \\
 \sigma_{22} &= a_1 a_2 \int_0^\tau \phi^2(t) dP_{\cdot 0}(t),
 \end{aligned}
 \tag{A6}$$

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 cumulative incidence for a cause 1 failure by time t , and $P_{\cdot 0}(t) = \int_0^t y_0(u) d\Lambda_{\cdot 0}(u)$ is the cumulative incidence for a failure due to all causes by time t . Furthermore, the covariance between $n^{-1/2}U_{11}$ and $n^{-1/2}U_{\cdot 1}$ under the contiguous alternative hypothesis is

$$\begin{aligned}
 &\langle n^{-1/2}U_{11}, n^{-1/2}U_{\cdot 1} \rangle \\
 &= n^{-1} \left\langle \int_0^\tau W_1(t) \frac{Y_1(t)Y_2(t)}{Y(t)} \left(\frac{dM_{11}^{(n)}(t)}{Y_1(t)} - \frac{dM_{12}^{(n)}(t)}{Y_2(t)} \right), \int_0^\tau W_{\cdot}(t) \frac{Y_1(t)Y_2(t)}{Y(t)} \left(\frac{dM_{\cdot 1}^{(n)}(t)}{Y_1(t)} - \frac{dM_{\cdot 2}^{(n)}(t)}{Y_2(t)} \right) \right\rangle \\
 &= n^{-1} \left\{ \int_0^\tau W_1(t)W_{\cdot}(t) \frac{Y_1^2(t)Y_2^2(t)}{Y(t)^2} \left(\frac{d \langle M_{11}^{(n)}(t), M_{\cdot 1}^{(n)}(t) \rangle}{Y_1(t)} + \frac{d \langle M_{12}^{(n)}(t), M_{\cdot 2}^{(n)}(t) \rangle}{Y_2(t)} \right) \right\} \\
 &= n^{-1} \left\{ \int_0^\tau W_1(t)W_{\cdot}(t) \frac{Y_1^2(t)Y_2^2(t)}{Y(t)^2} \left(\frac{d\Lambda_{11}^{(n)}(t)}{Y_1^2(t)} + \frac{d\Lambda_{12}^{(n)}(t)}{Y_2^2(t)} \right) \right\} \\
 &= n^{-1} \left\{ \int_0^\tau W_1(t)W_{\cdot}(t) \frac{Y_1^2(t)Y_2^2(t)}{Y(t)^2} \left(\frac{e^{\gamma_1 \phi_1(t)/(2\sqrt{n})} d\Lambda_{10}(t)}{Y_1(t)} + \frac{e^{-\gamma_1 \phi_1(t)/(2\sqrt{n})} d\Lambda_{10}(t)}{Y_2(t)} \right) \right\} \\
 &\approx n^{-1} \int_0^\tau W_1(t)W_{\cdot}(t) \frac{Y_1(t)Y_2(t)}{Y(t)} d\Lambda_{10}(t) \\
 &+ n^{-1} \int_0^\tau W_1(t)W_{\cdot}(t) \frac{Y_1(t)Y_2(t)(Y_1(t) - Y_2(t))}{Y^2(t)} \left(\gamma_1 \phi_1(t) / \sqrt{n} d\Lambda_{10}(t) \right),
 \end{aligned}
 \tag{A7}$$

which converges in probability to

$$\begin{aligned}
 \sigma_{12} &= \int_0^\tau \phi_1(t)\phi_{\cdot}(t) \frac{a_1 a_2 y_1(t) y_2(t)}{a_1 y_1(t) + a_2 y_2(t)} d\Lambda_{10}(t) \\
 &= \int_0^\tau \phi_1(t)\phi_{\cdot}(t) a_1 a_2 y_0(t) d\Lambda_{10}(t),
 \end{aligned}
 \tag{A8}$$

as $n \rightarrow \infty$.

Therefore, under the contiguous alternative hypothesis (A1), \mathbf{X}_n^2 has an asymptotic chi-square distribution with 2 degrees of freedom and noncentrality 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 $Z_{\cdot 1}^{(n)} = n^{-1/2}U_{\cdot 1}/\sqrt{\hat{\sigma}_{22}}$, where

$$\begin{aligned}
 \hat{\sigma}_{11} &= n^{-1} \frac{a_1 + a_2}{a_1 a_2} \int_0^\tau W_1^2(t) \frac{Y_1(t)Y_2(t)}{Y_1(t) + Y_2(t)} \left(\frac{dN_{11}(t) + dN_{12}(t)}{Y_1(t) + Y_2(t)} \right), \\
 \hat{\sigma}_{22} &= n^{-1} \frac{a_1 + a_2}{a_1 a_2} \int_0^\tau W_{\cdot}^2(t) \frac{Y_1(t)Y_2(t)}{Y_1(t) + Y_2(t)} \left(\frac{dN_{\cdot 1}(t) + dN_{\cdot 2}(t)}{Y_1(t) + Y_2(t)} \right)
 \end{aligned}
 \tag{A9}$$

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 (A1), $(Z_{11}^{(n)}, Z_{\cdot 1}^{(n)})$ converges to a bivariate normal random vector $(Z_{11}^*, Z_{\cdot 1}^*)$ 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 $M^* = \max(|Z_{11}^*|, |Z_{\cdot 1}^*|)$. □

A.2 | Derivation of Equation 6, formula for calculating number of failure due to cause 1 for chi-square test

The asymptotic theory in the above section A.1 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,¹⁴ a power analysis is usually based on a fixed alternative rather than a contiguous alternative. However, for given fixed alternatives γ_1^* and γ^* , the log(CSH₁ ratio) and the log(ACH ratio), the asymptotic results of Theorem 1 justify an approximate power calculation by setting $\gamma_1 = n^{1/2}\gamma_1^*$ and $\gamma = n^{1/2}\gamma^*$.

After plugging all the quantities from (A3) to (A2) and replace the local alternatives by fixed alternatives, we can get

$$\xi = n\boldsymbol{\mu}^{*T}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^*, \quad (\text{A10})$$

where $\boldsymbol{\mu}^* = (\mu_1^*, \mu_2^*)$, with

$$\begin{aligned} \mu_1^* &= \gamma_1^* a_1 a_2 P_{11} \\ \mu_2^* &= \gamma^* a_1 a_2 P_{.1}. \end{aligned} \quad (\text{A11})$$

After some algebraic manipulation, one can show that

$$\begin{aligned} \xi &= \frac{a_1 a_2 [\gamma_1^{*2} n P_{10} - 2\gamma_1^* \gamma^* n P_{10} + \gamma^{*2} n P_{.0}^*]}{1 - n P_{10} / n P_{.0}} \\ &= \frac{a_1 a_2 [\gamma_1^{*2} D_1 - 2\gamma_1^* \gamma^* D_1 + \gamma^{*2} D_1 / R]}{1 - R} \\ &= D_1 a_1 a_2 \frac{\gamma_1^{*2} - 2\gamma_1^* \gamma^* + \gamma^{*2} / R}{1 - R}, \end{aligned} \quad (\text{A12})$$

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 ratio of the cumulative incidence of failure due to cause 1 to cumulative incidence for failure due to all causes by time τ , which are approximately equal to the corresponding quantities in the pooled sample across treatment groups for large n under the contiguous alternative hypothesis (A1).

Then Equation 6, the final number of failure due to cause 1 can be easily solved from (5).

APPENDIX B: SUPPLEMENTARY FIGURES

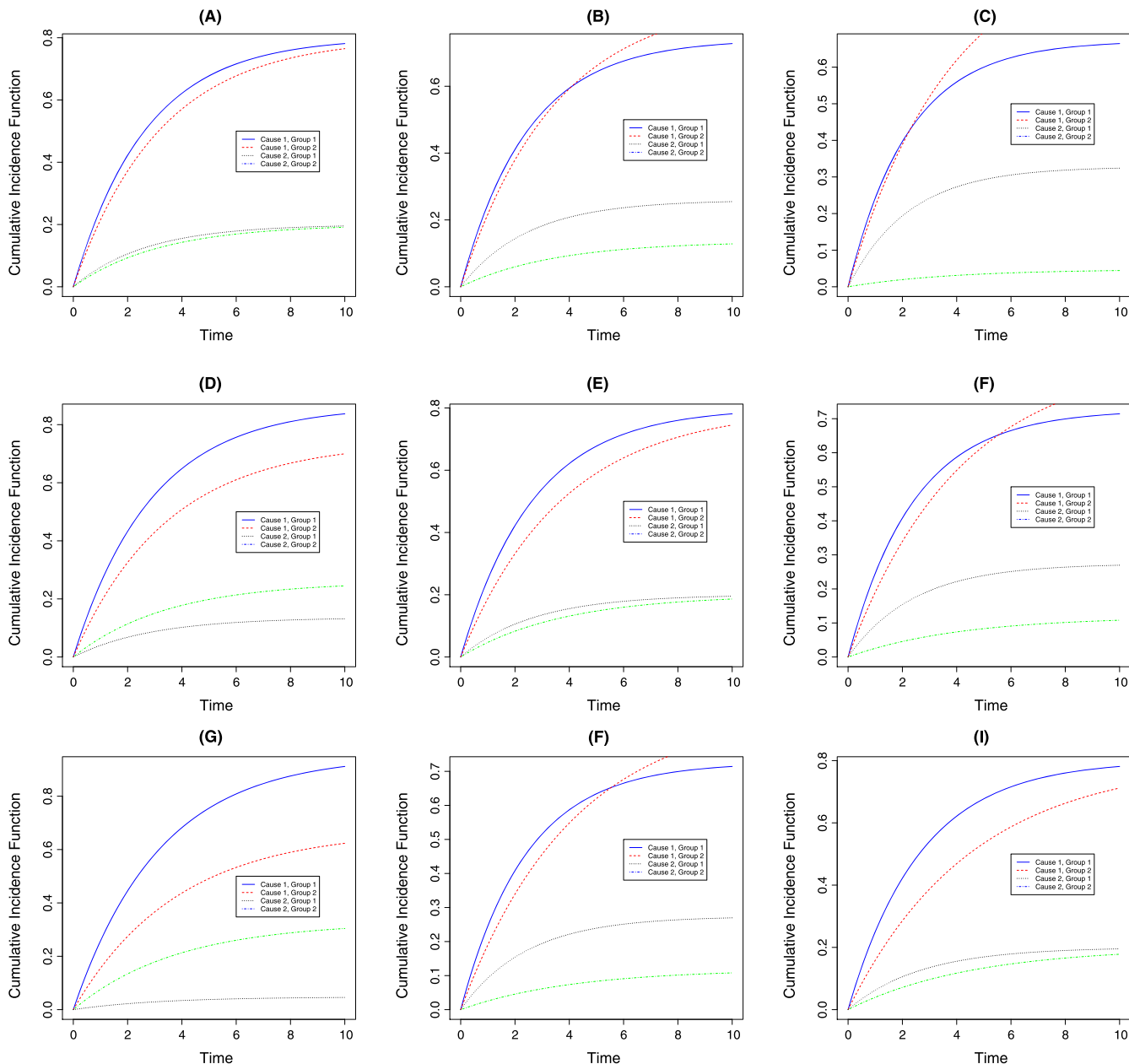


FIGURE B1 Cumulative incidence functions under 9 scenarios as described in Section 4. A, CSH_1 ratio = 1.2, ACH ratio = 1.2, CSH_2 ratio = 1.2. B, CSH_1 ratio = 1.2, ACH ratio = 1.4, CSH_2 ratio = 2.7. C, CSH_1 ratio = 1.2, ACH ratio = 1.7, CSH_2 ratio = 11.7. D, CSH_1 ratio = 1.4, ACH ratio = 1.2, CSH_2 ratio = 0.63. E, CSH_1 ratio = 1.4, ACH ratio = 1.4, CSH_2 ratio = 1.4. F, CSH_1 ratio = 1.4, ACH ratio = 1.7, CSH_2 ratio = 3.9. G, CSH_1 ratio = 1.7, ACH ratio = 1.2, CSH_2 ratio = 0.17. H, CSH_1 ratio = 1.7, ACH ratio = 1.4, CSH_2 ratio = 0.6. I, CSH_1 ratio = 1.7, ACH ratio = 1.7, CSH_2 ratio = 1.7 [Colour figure can be viewed at wileyonlinelibrary.com]