Published in final edited form as:

Stat Med. 2013 March 15; 32(6): 899–913. doi:10.1002/sim.5597.

The use of group sequential designs with common competing risks tests

Brent R. Logan^{a,*} and Mei-Jie Zhang^a

^a Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, WI USA 53226-0509

Abstract

Clinical trials are often performed using a group sequential design in order to allow investigators to review the accumulating data sequentially and possibly terminate the trial early for efficacy or futility. Standard methods for comparing survival distributions have been shown under varying levels of generality to follow an independent increments structure. In the presence of competing risks, where the occurrence of one type of event precludes the occurrence of another type of event, researchers may be interested in inference on the cumulative incidence function, which describes the probability of experiencing a particular event by a given time. This manuscript shows that two commonly used tests for comparing cumulative incidence functions, a pointwise comparison at a single point, and Gray's test, also follow the independent increments structure when used in a group sequential setting. A simulation study confirms the theoretical derivations even for modest trial sample sizes. Two examples of clinical trials in hematopoietic cell transplantation are used to illustrate the techniques.

Keywords

cumulative incidence; independent increments; Gray's test; clinical trials

1. Introduction

Clinical trials are often performed using a group sequential design in order to allow investigators to review the accumulating data sequentially and possibly terminate the trial early for efficacy or futility. Group sequential methods require adjustment to the critical values of the test statistics based on the joint distribution of the test statistics at different calendar times. Often this joint distribution fits into the independent increments structure or canonical joint distribution [1], which facilitates calculation of the critical values. For clinical trials comparing survival endpoints, standard methods for comparing survival distributions have been shown under varying levels of generality to follow an independent increments structure, including the log-rank test [2], the Cox model score process [3, 4], the weighted log-rank test under certain weight conditions [5, 6], and a pointwise comparison of survival probabilities at a fixed time [7, 8].

In many clinical trial settings, patients can experience several different types of events; for example, in cardiology clinical trials, patients could experience nonfatal myocardial infarction, cardiovascular death, or death due to non-cardiovascular causes. These events are

typically competing risks, where the occurrence of one type of event precludes the occurrence of other types of events. Often in clinical trials multiple causes of failure are incorporated into a composite endpoint which serves as a primary endpoint, and this composite endpoint is then used for group sequential monitoring. However, in some clinical trials with competing risks, researchers may be interested in event specific inference. This is typically considered when the intervention is anticipated to only affect one of the failure causes of interest. Latouche and Porcher [9] and Schulgen et al. [10] provide some discussion and examples of the use of event specific inference in clinical trials with competing risks. Typically this analysis is done in one of two ways. Analysis of the cause specific hazard function compares the instantaneous failure rates of a particular cause between the groups often using the log-rank test or Cox model. In this case the methodology of sequential monitoring is handled by the already existing literature on monitoring survival endpoints reviewed above. However, analysis of cause specific hazard functions are not directly interpretable in terms of probabilities of failure from a particular cause. Therefore, researchers may be more interested instead in the effect of treatment on the cumulative incidence function, defined as the probability of failing from a particular cause of interest before time t [11]. This cumulative incidence function depends on the hazard rates for all of the causes, rather than on the cause of interest alone.

Analysis and monitoring of cumulative incidence functions are especially important in clinical trials where the competing events have opposite clinical interpretations, i.e. the event of interest is a positive outcome while the competing event is a negative outcome. For example, consider a study of an intervention to improve the speed of engraftment after an umbilical cord blood transplant. Probability of engraftment is typically described using the cumulative incidence function, treating death or second transplant prior to engraftment as a competing risk. In this setting, the event of interest (engraftment) is a positive outcome, while the competing event (death or second transplant) is a negative outcome, so it doesn't make sense to construct a composite endpoint. Monitoring the cumulative incidence of engraftment would be appropriate for capturing the effect of treatment on engraftment. Another example of this using discharge alive from an intensive care unit (ICU) and death in the ICU as competing events is discussed in [9] and [12]. Other settings where monitoring of cumulative incidences may be useful include monitoring of specific types of toxicities for Data Safety Monitoring Board (DSMB) reporting.

This manuscript shows that two commonly used tests of cumulative incidence functions, Gray's test [13] and a pointwise comparison of cumulative incidence, follow the independent increments structure when applied in a group sequential design. The paper is organized as follows. In Section 2 we lay out the notation, review the test statistics, and derive the joint distribution of the test statistics over calendar time. In Section 3 we show the results of a simulation study to investigate the type I error control of the tests of cumulative incidence when used in a group sequential setting. In Section 4 we illustrate the procedures with two examples; the first looks at designing a study of engraftment after umbilical cord blood transplantation described previously, and the second considers toxicity monitoring in a clinical trial. Finally, in Section 5 we summarize our conclusions.

2. Methods

2.1. Notation

Suppose that there are two groups with n_1 and n_2 patients. Without loss of generality we just consider $K = 2$ causes of failure, and will focus on monitoring the cumulative incidence for cause 1. The k th cause specific hazard function for the i th group is $\lambda_{ik}(t)$; the analogous cumulative hazard function for the k th cause is $\Lambda_{ik}(t) = \int_0^t \lambda_{ik}(u) du$. For the j th individual in

the i th group, define the study entry time (calendar time) as τ_{ij} , the censoring time as C_{ij} , the event time as T_{ij} , and the event type as $\epsilon_{ij} \in \{1, 2\}$. The observed data at calendar time s is $(X_{ij}(s), \delta_{ij}(s), \epsilon_{ij})$, where $X_{ij}(s) = \max(0, (T_{ij} \wedge C_{ij} \wedge (s - \tau_{ij})))$ is the event or censoring time available at calendar time s , $\delta_{ij}(s) = I(T_{ij} \leq C_{ij} \wedge (s - \tau_{ij}))$ is the event indicator and ϵ_{ij} denotes the failure type, observable if an event has occurred prior to $C_{ij} \wedge (s - \tau_{ij})$. Define the at risk indicator as $Y_{ij}(t, s) = I(X_{ij}(s) > t)$, and define $a_{ij}(u, s) = I(C_{ij} \wedge (s - \tau_{ij}) \geq u)$. We can define a counting process for event type k as

$$N_{ijk}(t, s) = I(X_{ij}(s) \leq t, \delta_{ij}(s) = 1, \epsilon_{ij} = k) = \int_0^t a_{ij}(u, s) d\tilde{N}_{ijk}(u),$$

where the counting process without censoring $\tilde{N}_{ijk}(t) = I(T_{ij} \leq t, \epsilon_{ij} = k)$ has corresponding Martingale

$$M_{ijk}(t) = I(T_{ij} \leq t, \epsilon_{ij} = k) - \int_0^t I(T_{ij} \geq w) d\Lambda_{ik}(w).$$

The number at risk at calendar time s and event time t for the i th group is given by

$$Y_i(t, s) = \sum_j Y_{ij}(t, s) = \sum_j a_{ij}(t, s) I(T_{ij} \geq t).$$

We are interested in group sequential inference on the cumulative incidence function for cause 1, defined as

$$F_{i1}(t) = \int_0^t S_i(u) d\Lambda_{i1}(u).$$

Here $S_i(t)$ is the event free survival function for group i where any type of event is counted.

At calendar time s and event time t we estimate the cumulative incidence as

$$\widehat{F}_{i1}(t, s) = \int_0^t \widehat{S}_i(u-, s) d\widehat{\Lambda}_{i1}(u, s),$$

where $\widehat{S}_i(t, s)$ is the Kaplan-Meier estimate of $S_i(t)$ at calendar time s given by

$$\widehat{S}_i(t, s) = \prod_{u \leq t} \left\{ 1 - \frac{\sum_{j,k} dN_{ijk}(u, s)}{Y_i(u, s)} \right\},$$

and $d\widehat{\Lambda}_{ik}(t, s)$ is the estimated cause k specific hazard function

$$d\widehat{\Lambda}_{ik}(t, s) = \sum_j dN_{ijk}(t, s) / Y_i(t, s).$$

To avoid tail instability issues, we restrict our attention to the region $0 \leq t \leq s \leq \tau$, and assume that

$$\liminf \frac{1}{n_i} E(Y_i(t, s)) = \inf y_i(t, s) > 0,$$

for $i = 1, 2$ and all t, s in this region.

2.2. Sequential design for a single cumulative incidence

First we consider sequential design for the estimated cumulative incidence from a single sample, where one repeatedly compares the estimated cumulative incidence at a particular event time to a prespecified cumulative incidence at that event time. This single sample setting might be useful for example when monitoring the incidence of a toxicity in the presence of other competing toxicities. Monitoring for a single sample is based on the statistic $D_i(t, s) = \sqrt{n_i}(\widehat{F}_{i1}(t, s) - F_{i1}(t))$ for cause 1. Note that in practice, one may fix the event time t to be a time point of interest, and use that same t for all sequential analyses. However, we derive the distribution of the process $D_i(t, s)$ over both t and s to allow flexibility in monitoring different cumulative incidences at different calendar times. This type of flexibility was discussed for the survival setting in [8].

We obtain a similar representation for $D_i(t, s)$ as was derived by [14] for the cumulative incidence in a fixed sample setting. First, $D_i(t, s)$ can be written as

$$D_i(t, s) = \sqrt{n_i} \int_0^t \widehat{S}_i(u-, s) d(\widehat{\Lambda}_{i1}(u, s) - \Lambda_{i1}(u)) + \sqrt{n_i} \int_0^t (\widehat{S}_i(u-, s) - S_i(u)) d\Lambda_{i1}(u) du.$$

Because of the relationship between $\widehat{S}_i(t, s)$ and the cumulative hazard for all causes $\widehat{\Lambda}_i(t, s)$, we can write this as

$$D_i(t, s) = \sqrt{n_i} \int_0^t \widehat{S}_i(u-, s) d(\widehat{\Lambda}_{i1}(u, s) - \Lambda_{i1}(u)) - \sqrt{n_i} \int_0^t S_i(u) \left[\int_0^u \frac{\widehat{S}_i(v-, s)}{S_i(v)} d(\widehat{\Lambda}_i(v, s) - \Lambda_i(v)) \right] d\Lambda_{i1}(u) + o_p(1),$$

where $\Lambda_i(t) = \Lambda_{i1}(t) + \Lambda_{i2}(t)$ is the cumulative hazard function of all causes for treatment group i . Using integration by parts,

$$\begin{aligned} D_i(t, s) &= \sqrt{n_i} \int_0^t \widehat{S}_i(u-, s) d(\widehat{\Lambda}_{i1}(u, s) - \Lambda_{i1}(u)) \\ &\quad - \sqrt{n_i} F_{i1}(t) \int_0^t \frac{\widehat{S}_i(u-, s)}{S_i(u)} d(\widehat{\Lambda}_i(u, s) - \Lambda_i(u)) \\ &\quad + \sqrt{n_i} \int_0^t F_{i1}(u) \frac{\widehat{S}_i(u-, s)}{S_i(u)} d(\widehat{\Lambda}_i(u, s) - \Lambda_i(u)) + o_p(1). \end{aligned}$$

Using the Martingale representations

$$\sqrt{n_i}(\widehat{\Lambda}_{ik}(t, s) - \Lambda_{ik}(t)) = \sqrt{n_i} \int_0^t \sum_j a_{ij}(u, s) dM_{ijk}(u) / Y_i(u, s),$$

and

$$\sqrt{n_i}(\widehat{\Lambda}_i(t, s) - \Lambda_i(t)) = \sqrt{n_i} \int_0^t \sum_j a_{ij}(u, s) (dM_{ij1}(u) + dM_{ij2}(u)) / Y_i(u, s),$$

and combining terms across the independent Martingales M_{ij1} and M_{ij2} , we can rewrite $D_i(t, s)$ as the Martingale representation,

$$\begin{aligned} D_i(t, s) &= \sqrt{n_i} \int_0^t \left[\widehat{S}_i(u-, s) - \frac{\widehat{S}_i(u-, s)}{S_i(u)} (F_{i1}(t) - F_{i1}(u)) \right] \frac{\sum_j a_{ij}(u, s) dM_{ij1}(u)}{Y_i(u, s)} \\ &\quad - \sqrt{n_i} \int_0^t \frac{\widehat{S}_i(u-, s)}{S_i(u)} [F_{i1}(t) - F_{i1}(u)] \frac{\sum_j a_{ij}(u, s) dM_{ij2}(u)}{Y_i(u, s)} + o_p(1). \end{aligned} \quad (1)$$

In Appendix A we show using this formulation and following the arguments of Gu and Lai [6], referred to as GL hereafter, that $D_i(t, s)$ converges weakly to a Gaussian process $\xi_i(t, s)$ with covariance function across calendar times s, s^* and event times t, t^* given by

$$\begin{aligned} E[\xi_i(t, s)\xi_i(t^*, s^*)] &= \int_0^{t \wedge t^*} [S_i(u) - F_{i1}(t) + F_{i1}(u)] [S_i(u) - F_{i1}(t^*) + F_{i1}(u)] \frac{\lambda_{i1}(u)}{y_i(u, s^*)} du \\ &\quad + \int_0^{t \wedge t^*} [F_{i1}(t) - F_{i1}(u)] [F_{i1}(t^*) - F_{i1}(u)] \frac{\lambda_{i2}(u)}{y_i(u, s^*)} du. \end{aligned}$$

An important practical application is when the event time is the same for each sequential analysis, so that the same time point on the cumulative incidence function is being monitored sequentially. In this case, $t = t^*$ in the covariance expression above, and the limiting covariance for s, s^* is

$$\text{Cov}(D_i(t, s), D_i(t, s^*)) = \int_0^t [S_i(u) - F_{i1}(t) + F_{i1}(u)]^2 \frac{\lambda_{i1}(u)}{y_i(u, s^*)} du + \int_0^t [F_{i1}(u) - F_{i1}(t)]^2 \frac{\lambda_{i2}(u)}{y_i(u, s^*)} du = \text{Var}(D_i(t, s^*)).$$

The standardized test statistic

$$Z_i(t, s) = D_i(t, s) / \sqrt{\text{Var}(D_i(t, s))} \quad (2)$$

then has covariance across calendar time $\text{Cov}(Z_i(t, s), Z_i(t, s^*)) = \sqrt{I_i(t, s) / I_i(t, s^*)}$, where $I_i(t, s) = n_i \text{Var}(D_i(t, s))^{-1}$ is the information on $D_i(t, s)$ accumulated by calendar time s . This follows the independent increments or canonical joint distribution structure [1]. The information can be estimated by

$$\widehat{I}_i(t, s) = \left\{ \int_0^t \left[\widehat{S}_i(u) - \widehat{F}_{i1}(t) + \widehat{F}_{i1}(u) \right]^2 \frac{\sum_j dN_{ij1}(u, s)}{Y_i^2(u, s)} du + \int_0^t \left[\widehat{F}_{i1}(u) - \widehat{F}_{i1}(t) \right]^2 \frac{\sum_j dN_{ij2}(u, s)}{Y_i^2(u, s)} du \right\}^{-1}.$$

These results show that the usual group sequential monitoring techniques based on an independent increments structure can be used to monitor this single sample cumulative incidence when the same time point on the cumulative incidence function is used at all calendar times. For example, using an error spending approach [15], one could derive the critical value for $Z_i(t, s)$ according to

$$P\left(Z_i(t, s_1) \leq c_1, \dots, Z_i(t, s_{p-1}) \leq c_{p-1}, Z_i(t, s_p) > c_p\right) = \pi_p,$$

for a sequence of calendar times $s_1, \dots, s_p$. Here the π_p are chosen so that $\sum_p \pi_p = \alpha$. They can either be prespecified [15] or chosen according to an error spending function [16], $\alpha(\cdot)$, defined on the information time scale such that $\alpha(0) = 0$ and $\alpha(1) = \alpha$. The independent increments structure facilitates the multivariate normal integration needed to evaluate these critical values. To implement the error spending function, the information time scale (constrained between 0 and 1) at time s is the information fraction

$$IF_i(s) = I_i(t, s) / I_i(t, s_{\max}),$$

for maximum calendar time $s_{\max}$.

2.3. Sequential Design for Two Samples

When comparing the cumulative incidence for two independent treatment groups, there are multiple tests that could be considered. In this section we discuss comparisons of the cumulative incidence at specified event time points, whereas in the next section we discuss comparisons of the entire cumulative incidence functions using Gray's test. To compare cumulative incidence at time t between two independent samples, we are interested in the statistic computed at calendar time s

$$D(t, s) = \sqrt{n_1 n_2 / n} \left[\left(\widehat{F}_{21}(t, s) - F_{21}(t) \right) - \left(\widehat{F}_{11}(t, s) - F_{11}(t) \right) \right] = \sqrt{\frac{n_1}{n}} D_2(t, s) - \sqrt{\frac{n_2}{n}} D_1(t, s).$$

We define $\rho_i = \lim_{n \rightarrow \infty} n_i / n$ and assume $0 < \rho_i < 1$. Under $H_0: F_{11}(t) = F_{21}(t)$, we can use the previous results to show that $D(t, s)$ converges weakly to a Gaussian process

$\xi(t, s) = \sqrt{\rho_1} \xi_2(t, s) - \sqrt{\rho_2} \xi_1(t, s)$ with covariance function

$$E[\xi(t, s) \xi(t^*, s^*)] = \sum_{i=1}^2 \rho_{3-i} E[\xi_i(t, s) \xi_i(t^*, s^*)],$$

for $s \leq s^*$. If we consider the common case where the event time t is the same for each sequential analysis, then the limiting covariance is

$$\text{Cov}(D(t, s), D(t, s^*)) = \sum_{i=1}^2 \rho_{3-i} \text{Cov}(D_i(t, s), D_i(t, s^*)) = \sum_{i=1}^2 \rho_{3-i} \text{Var}(D_i(t, s^*)) = \text{Var}(D(t, s^*)).$$

Therefore $D(t, s)$ when t is the same for each calendar time also follows the independent increments structure, and similar group sequential monitoring under the independent increments structure as was described for the single sample problem can be implemented for the two sample problem. Here the information at calendar time s is given by $I(t, s) = n(\rho_1 \rho_2) \text{Var}(D(t, s))^{-1}$.

2.4. Gray's test

Gray [13] proposed a test comparing the entire cumulative incidence functions through an integrated weighted difference in the so called subdistribution hazard functions. This test is an analog of the log-rank test which instead integrates the weighted difference in cause specific hazard functions. The subdistribution hazard for cause 1 in group i is given by

$$\gamma_{i1}(t) = \frac{dF_{i1}(t)}{1 - F_{i1}(t)}.$$

The subdistribution hazard is related to the cumulative incidence through the cumulative subdistribution hazard $\Gamma_{i1}(t) = \int_0^t \gamma_{i1}(u) du$ using the relationship

$$F_{i1}(t) = 1 - \exp\left\{-\int_0^t \gamma_{i1}(u) du\right\}.$$

An estimator for the cumulative subdistribution hazard at calendar time s is given by

$$\widehat{\Gamma}_{i1}(t, s) = \int_0^t \frac{\sum_j a_{ij}(u, s) d\tilde{N}_{ij1}(u)}{R_i(u, s)},$$

where $R_i(u, s) = Y_i(u, s) [1 - \widehat{F}_{i1}(u-, s)] / \widehat{S}_i(u-, s)$. Recall here that $a_{ij}(u, s)$ as defined in Section 2.1 is an indicator that at calendar time s , the patient hasn't been censored by event time u and entered the study at least u amount of time prior, so that the at risk indicator is $Y_{ij}(u, s) = a_{ij}(u, s) I(T_{ij} > u)$. Then the Gray's test statistic at calendar time s is an integrated weighted difference in estimated subdistribution hazards given by

$$G(t, s) = n^{-1/2} \int_0^t \frac{R_1(u, s) R_2(u, s)}{R(u, s)} \left[\frac{\sum_j a_{1j}(u, s) d\tilde{N}_{1j1}(u)}{R_1(u, s)} - \frac{\sum_j a_{2j}(u, s) d\tilde{N}_{2j1}(u)}{R_2(u, s)} \right],$$

where $R(u, s) = R_1(u, s) + R_2(u, s)$. This test can be more efficient than the pointwise comparison of cumulative incidence at a specified time discussed in the previous section, if the subdistribution hazard functions are proportional.

Under $H_0: F_{11}(t) = F_{21}(t) = F_1^0(t)$, $\forall t \in [0, \tau]$, we can write $G(t, s)$ as

$$G(t, s) = n^{-1/2} \int_0^t \frac{R_1(u, s) R_2(u, s)}{R(u, s)} \left[\frac{\sum_j a_{1j}(u, s) dM_{1j1}(u)}{R_1(u, s)} - \frac{\sum_j a_{2j}(u, s) dM_{2j1}(u)}{R_2(u, s)} \right] + o_p(1). \quad (3)$$

In appendix B we show that under H_0 , $G(t, s)$ converges weakly to a Gaussian process $\xi_G(t, s)$ with covariance function given by

$$E[\xi_G(t, s) \xi_G(t^*, s^*)] = \int_0^{t \wedge t^*} \frac{r_1(u, s) r_2(u, s)}{r(u, s)} \frac{dF_1^0(u)}{1 - F_1^0(u)},$$

for $s \leq s^*$, where

$$r_i(u, s) = \rho_i y_i(u, s) \frac{1 - F_{i1}(u)}{S_i(u)},$$

and

$$r.(u, s) = \sum_i r_i(u, s).$$

In practice, sequential analyses are often based on $G(s, s)$, which utilizes all the available data at a particular calendar time s , so that for $s = s^*$,

$$\text{Cov}(G(s, s), G(s^*, s^*)) = \int_0^s \frac{r_1(u, s) r_2(u, s)}{r.(u, s)} \frac{dF_1^0(u)}{1 - F_1^0(u)} = \text{Var}(G(s, s)).$$

Therefore, Gray's test monitored in this way also follows the independent increments structure, and so similar group sequential monitoring methods as described for single sample tests can be applied. Here the information is $I_G(s, s) = \text{Var}(G(s, s))$, which can be estimated by

$$\widehat{I}_G(s, s) = \int_0^s \frac{R_1(u, s) R_2(u, s)}{n R.(u, s)} \frac{d\widehat{F}_1^0(u, s)}{1 - \widehat{F}_1^0(u-, s)},$$

where $\widehat{F}_1^0(u, s)$ is the cumulative incidence estimate for cause 1 from the pooled samples. The maximum information is given by

$$I_{\max} = n \text{Var}(G(s_{\max}, s_{\max})),$$

where $s_{\max}$ is the maximum calendar time in the study. The information fraction is given by $IF_G(s) = I_G(s, s)/I_{\max}$.

For study planning purposes, an approximation to the maximum information and the information fraction is available. This approximation is analogous to the approximation to the variance of the log-rank test as the total number of deaths divided by 4 [17]. A similar approximation was derived for competing risks data in [18] for the Fine and Gray [19] model, which is a model for the subdistribution hazards in a regression setting, assuming a proportional subdistribution hazard function; see [9] also for a review. They show that the variance of the parameter estimate from the Fine and Gray model is $1/(\rho_1 \rho_2 e_1)$, where e_1 is the expected number of type 1 events. Here we derive a similar result for Gray's test. First note that the limiting probability of being at risk at calendar time s and event time u can be written as $y_i(u, s) = \lim_{n_j \rightarrow \infty} E(Y_i(u, s))/n_j = S_i(u)C(u, s)$, where $C(u, s)$ is the limit in probability of $\sum_j a_{ij}(u, s)/n_i$ and can be interpreted as the likelihood at calendar time s of not being censored by event time u and having entered the study at least u amount of time prior. We assume $C(u, s)$ is the same for each treatment group (equal censoring/accrual patterns). Then under H_0 , $r_i(u, s) = \rho_i (1 - F_1^0(u)) C(u, s)$, so that $r_i(u, s)/r.(u, s) = \rho_i$. Then we can write the limiting variance of Gray's test as

$$\text{Var}(G(s, s)) = \int_0^s \rho_1 \rho_2 r(u, s) \frac{dF_1^0(u)}{1 - F_1^0(u)} = \rho_1 \rho_2 \int_0^s C(u, s) dF_1^0(u).$$

The information at calendar time s is $I_G(s) = \rho_1 \rho_2 e_1(s)$, where

$$e_1(s) = n \int_0^s C(u, s) dF_1^0(u),$$

is the expected number of type 1 events observed by calendar time s . Therefore, similar to the log-rank test, the variance of Gray's test with equal allocation of sample size is roughly the total number of type 1 events divided by 4. This approximation can be used as the basis for designing a group sequential trial using Gray's test, since the maximum information for the study can be set according to the targeted number of type I events at the final calendar time based on an anticipated accrual and censoring pattern and anticipated cumulative incidence functions. Furthermore, the information fraction is $IF_G(s) = I_G(s, s)/I(s_{\max}, s_{\max}) = e_1(s)/e_1(s_{\max})$, so that sequential monitoring can be based on the ratio of the number of type 1 events at calendar time s to the number of type 1 events expected by the end of the study.

3. Simulation Studies

A simulation study was conducted to confirm the control of the type I error when Gray's test is used in a group sequential design setting. Competing risks data were generated from the following cumulative incidence function for cause 1 for both treatment groups under the null hypothesis:

$$F_{11}(t) = F_{21}(t) = F_1^0(t) = p(1 - e^{-t}).$$

The cumulative incidence function for cause 2 and group i was generated according to $F_{2i}(t) = (1 - p)(1 - e^{-\gamma_i t})$. Here p represents the proportion of cause 1 events. Data was simulated by generating a cause 1 indicator with probability p , and then using an inverse CDF technique to generate the event time conditional on the failure cause according to $F_{ik}(t)/F_{ik}(\infty)$. Values of $p = 0.25, 0.5, 0.75$ and pairs of γ values of $(\gamma_1, \gamma_2) = (1, 1), (1, 1.5), (1.5, 1.5)$ were used in the simulations to reflect both situations where the competing event had the same cumulative incidence function for both groups as well as a situation where the competing event had a different cumulative incidence function depending on the group. Sample curves are shown in Figure 1 for the cumulative incidence for cause 1 for each value of p ; note these curves apply regardless of the γ values for the competing event, and they are the same for both treatment groups under the null hypothesis.

A trial design with staggered uniform accrual was set up with accrual period A and total study time T . In the first set of simulations, no censoring other than administrative censoring due to staggered accrual was used. Two values of the pair (A, T) were used, $(A, T) = (1, 1.61)$ and $(2.3, 2.5)$, which resulted in approximately 2/3 of the total number of type I events being observable during the study period. These two scenarios had different relative weightings of accrual time vs. minimum follow-up time. For this particular model, the information fraction at calendar time a has a closed form given by

$$IF_G(s) = \frac{\min(s, A) - e^{-(s - \min(s, A))} + e^{-s}}{A - e^{-(T - A)} + e^{-T}},$$

which doesn't depend on p or γ_i . Therefore, the calendar times for interim analyses selected for equal information looks are $s = (0.635, 0.936, 1.22, 1.61)$ for $(A, T) = (1, 1.61)$ and $s = (1.037, 1.575, 2.043, 2.5)$ for $(A, T) = (2.3, 2.5)$. An O'Brien-Fleming type error spending function with $\alpha(IF_G(s)) = \min(\alpha[IF_G(s)]^3, \alpha)$ for information fraction $IF_G(s)$ was used with $\alpha = 0.05$. Cumulative type I error rates at each of the four interim analyses for sample sizes of 100 per group are shown in Table 1, based on 10,000 Monte Carlo simulations for each scenario. The theoretical value is also shown at the bottom of the table.

In a second set of simulations, the same scenarios were used except that an additional independent exponential censoring mechanism was added, resulting in approximately 10% of event times censored by this mechanism. Note that this independent censoring mechanism causes slight changes in the information fractions, so that one must use an error spending approach rather than assuming equal information increments. These simulation results are shown in Table 2.

The cumulative type I error rate is well controlled across all scenarios, supporting the theoretical results that Gray's test can be used in a group sequential setting by directly applying standard methods based on independent increments. Note in particular that the type I error rate is well-controlled even in settings where there are differences in the competing event incidence (i.e. cause 2).

4. Examples

4.1. Example 1

In this first example, we describe the application of the methods to the design of a new clinical trial of an intervention to improve platelet engraftment after umbilical cord blood transplantation. Data from the registry of the Center for International Blood and Marrow Transplant Research (CIBMTR) can be used to provide information on the anticipated platelet engraftment rates for the standard or control umbilical cord blood transplant group. Historical CIBMTR data is shown in Figure 2, indicating that the cumulative incidence of platelet engraftment by 6 months is expected to be 63% after a standard cord blood transplant. We would like to have 90% power to detect a 15% improvement in the 6 month platelet engraftment with the new intervention, using a 5% type I error rate. Assuming a proportional subdistribution hazards model, Latouche et al. [18] point out that the subdistribution hazard ratio is given by

$$\theta = \frac{\log [1 - F_{21}(t)]}{\log [1 - F_{11}(t)]},$$

so that the targeted value is $\theta = 1.52$. Following [18] the targeted number of platelet engraftment events is given by

$$e_1 = \frac{4(z_{\alpha/2} + z_\beta)^2}{(\log \theta)^2} = 240.$$

If an O'Brien-Fleming group sequential design is planned with 4 equally spaced interim analyses, the maximum information should be inflated by a factor of 1.022 to maintain the desired power, for a target of 245 platelet engraftment events. Other options for conducting interim analyses such as futility analyses would also need be accounted for in the targeted number of events, and the impact on power. A number of accrual and follow up strategies

can be used in the study design in order to target this number of type 1 events. One conservative strategy would be to target 245 platelet engraftment events occurring within 6 months of follow-up, use a minimum follow up of 6 months for all patients, and enroll $245 / (0.5 \times 0.63 + 0.5 \times 0.78) = 348$ total patients in order to get 245 platelet engraftment events under the alternative hypothesis. This assumes no additional censoring beyond administrative censoring, and would likely result in only a slight excess of total platelet engraftment events, since the platelet engraftment curves seem to flatten by 6 months. Interim analyses after equally spaced information increments could be conducted after 61, 122, 184, and 245 platelet engraftment events have occurred. Alternatively, error spending functions could be used to account for interim analyses at unequal information increments.

4.2. Example 2

The next example illustrates group sequential monitoring of the cumulative incidence of a particular toxicity for safety reporting. BMTCTN 0101 [20] was a randomized phase 3 clinical trial conducted by the Blood and Marrow Transplant Clinical Trials Network to compare voriconazole vs. fluconazole as a fungal prophylaxis agent during a hematopoietic cell transplant (HCT). One of the prespecified toxicity monitoring rules was based on comparing the cumulative incidence of renal failure by 100 days to the historical incidence of 10% separately in each treatment arm. In the HCT setting, there is a non-negligible risk of competing events for this toxicity, namely death without renal failure; by monitoring the cumulative incidence of renal failure we can focus on this potential drug related toxicity in the presence of the background mortality of the transplant procedure. For illustration, we will show the results of monitoring the renal failure incidence in the voriconazole arm using the one sample Z test in (2), with $t = 100$ days, a null hypothesis value of $F_{H1}(t) = 0.1$, and using a one-sided type I error rate of 5% similar to the trial protocol. The maximum information assuming all patients are evaluable for 100 day renal failure is $n[p(1-p)]$ where p is the cumulative incidence of renal failure by 100 days. Since this depends on the unknown p , we compute the information fraction using the current estimate $\hat{p}$ in the maximum information expression at each calendar time; this avoids the issue of potentially exceeding the maximum information before all patients are evaluable if the assumed p is misspecified. A total of 305 patients were enrolled to the voriconazole arm. We monitor every 3 months for 33 months, starting with month 6 of the study where sufficient enrollment and follow-up is available. Estimates of the cumulative incidence, standard errors, maximum information, and information fractions are given in Table 3 for each of the 10 evaluation calendar times.

Boundary values were computed using the power error spending function $\alpha(t) = \alpha t$, which is similar to a Pocock type shape. The Z test statistic and boundary are plotted in Figure 3, and the boundary was never crossed, illustrating that there was no evidence of excessive renal toxicity for this agent in this trial.

5. Discussion

We have shown that standard methods for analyzing competing risks data, including pointwise comparisons of cumulative incidence and Gray's test, can be used in a group sequential design setting using standard methods for test statistics with an independent increments structure. Although often clinical trial monitoring of multiple event types is implemented on a single composite primary endpoint, there are some settings where monitoring of the cumulative incidence function may be valuable. A composite endpoint reflects the overall impact of the treatment on both the event of interest and its competing event, and may be more appropriate when the treatment can impact both event types, especially when the competing event is an important clinical endpoint such as mortality. Treatment effects on the competing event may cause changes in the cumulative incidence

function of the event of interest, particularly if the competing event typically occurs sooner than the event of interest. However, comparisons of cumulative incidence functions provide a more direct measure of the impact of treatment on an event of interest, and they may be more powerful than comparisons of composite endpoint probabilities when the treatment is not anticipated to impact the competing event. Comparisons of cumulative incidence functions also provide direct inference on a clinically interpretable probability of observing the event of interest, whereas comparisons of the cause specific hazard rates do not correspond directly to changes in the event probability. Relative sample size requirements for inference on the cumulative incidence vs. inference on the cause specific hazards are considered in [9]. Cumulative incidence functions may be a key secondary endpoint in addition to a composite primary endpoint, for which it is important to account for group sequential testing. In some settings, as in the engraftment example, a composite endpoint doesn't make sense, since the event of interest is a positive outcome while the competing event is a negative outcome. Finally, clinical trials often include ongoing monitoring of toxicities to determine if there are safety concerns with the treatments being studied. Sequential monitoring of cumulative incidence functions can be a useful tool for monitoring targeted toxicities on an ongoing basis.

Acknowledgments

The author would like to thank the Blood and Marrow Transplant Clinical Trials Network for providing the renal toxicity monitoring data from the BMTCTN 0101 trial used in example 2 (BMT CTN is supported in part by grant # U01HL069294 from the National Heart, Lung, and Blood Institute). This research was partially supported by a grant (R01 CA54706-10) from the National Cancer Institute.

Appendix A

Note that $D_i(t, s)$ is asymptotically equivalent for fixed s to

$$\tilde{D}_i(t, s) = \sqrt{n_i} \int_0^t [S_i(u) - F_{i1}(t) + F_{i1}(u)] \frac{\sum_j a_{ij}(u, s) dM_{ij1}(u)}{Y_i(u, s)} - \sqrt{n_i} \int_0^t [F_{i1}(t) - F_{i1}(u)] \frac{\sum_j a_{ij}(u, s) dM_{ij2}(u)}{Y_i(u, s)},$$

since $\widehat{S}_i(u-, s)$ converges uniformly to $S_i(u)$. Then the covariance between the one sample pointwise test computed at times (t, s) and times (t^*, s^*) for $s \leq s^*$ is

$$\begin{aligned} \text{Cov}(\tilde{D}_i(t, s), \tilde{D}_i(t^*, s^*)) &= n_i \int_0^{t \wedge t^*} [S_i(u) - F_{i1}(t) + F_{i1}(u)] [S_i(u) - F_{i1}(t^*) + F_{i1}(u)] \frac{\sum_j a_{ij}(u, s) \langle dM_{ij1}(u) \rangle}{Y_i(u, s) Y_i(u, s^*)} + n_i \int_0^{t \wedge t^*} [F_{i1}(u) - F_{i1}(t)] [F_{i1}(u) - F_{i1}(t^*)] \\ &\rightarrow_p \int_0^{t \wedge t^*} [S_i(u) - F_{i1}(t) + F_{i1}(u)] [S_i(u) - F_{i1}(t^*) + F_{i1}(u)] \frac{\lambda_{i1}(u)}{y_i(u, s^*)} du + \int_0^{t \wedge t^*} [F_{i1}(t) - F_{i1}(u)] [F_{i1}(t^*) - F_{i1}(u)] \frac{\lambda_{i2}(u)}{y_i(u, s^*)} du \end{aligned}$$

Let

$$\eta_{n1}(u, s) = \frac{\widehat{S}_i(u-, s)}{Y_i(u, s)/n_i}, \eta_1(u, s) = \frac{S_i(u)}{y_i(u, s)},$$

$$\eta_{n2}(u, s) = \frac{\eta_{n1}(u, s)(F_{i1}(t) - F_{i1}(u))}{S_i(u)}, \eta_2(u, s) = \frac{F_{i1}(t) - F_{i1}(u)}{y_i(u, s)}.$$

Then (1) can be written as

$$n_i^{-1/2} \left[\int_0^t (\eta_{n1}(u, s) - \eta_{n2}(u, s)) \sum_j a_{ij}(u, s) dM_{ij1}(u) - \int_0^t \eta_{n2}(u, s) \sum_j a_{ij}(u, s) dM_{ij2}(u) \right].$$

To prove weak convergence of (1), following corollary 1 of GL [6], it is sufficient to show that $\eta_{nv}(u, s)$ converges uniformly to $\eta_v(u, s)$ for $v = 1, 2$, and that each has bounded total variation,

$$\sup_{0 \leq s \leq \tau} \left\{ V_{0 \leq u \leq s} \left[(Y_i(u, s) / n_i)^\alpha \eta_{nv}(u, s) \right] + V_{0 \leq u \leq s} \left[y_1^\alpha(u, s) \eta_v(u, s) \right] \right\} = O_p(1), v=1, 2$$

for $0 < \alpha < 1/2$. Uniform convergence follows from the uniform convergence of $\widehat{S}_i(u-, s)$ and $Y_i(u, s)/n_i$ to $S_i(u)$, and $y_i(u, s)$, respectively. To show bounded variation for $v = 1$, note that using lemma 3(ii) of GL,

$$V_{0 \leq u \leq s} \left[(Y_i(u, s) / n_i)^\alpha \eta_{n1}(u, s) \right] \leq \sup |Y_i(u, s) / n_i|^\alpha V_{0 \leq u \leq s} \left[\frac{\widehat{S}_i(u-, s)}{Y_i(u, s) / n_i} \right] + V_{0 \leq u \leq s} \left[(Y_i(u, s) / n_i)^\alpha \right] \sup \left| \frac{\widehat{S}_i(u-, s)}{Y_i(u, s) / n_i} \right|.$$

Since $Y_i(u, s)/n_i$ is bounded and nonincreasing, it has bounded variation. Also,

$$V_{0 \leq u \leq s} \left[\frac{\widehat{S}_i(u-, s)}{Y_i(u, s) / n_i} \right]$$

is bounded since for any partition P ,

$$n_i \sum_{t_k \in P} \left[\frac{\widehat{S}_i(t_{k+1}, s)}{Y_i(t_{k+1}, s)} - \frac{\widehat{S}_i(t_k, s)}{Y_i(t_k, s)} \right] = n_i \sum_{t_k \in C} \frac{\widehat{S}_i(t_k, s)}{(Y_i(t_k, s) - 1) Y_i(t_k, s)} = O_p(1),$$

where C is the set of censored times. It is straightforward to see that all other terms are also bounded in the limit. To show bounded variation for $v = 2$, note using lemma 3(ii) of GL that

$$\begin{aligned} V_{0 \leq u \leq s} \left[(Y_1(u, s) / n)^\alpha \eta_{n2}(u, s) \right] &\leq \sup \left| \frac{F_{i1}(t) - F_{i1}(u)}{S_i(u) - S_i(t)} \times \frac{S_i(u) - S_i(t)}{S_i(u)} \right| V_{0 \leq u \leq s} \left[(Y_1(u, s) / n)^\alpha \eta_{n1}(u, s) \right] \\ &+ \sup |Y_i(u, s) / n_i|^\alpha \eta_{n1}(u, s) V_{0 \leq u \leq s} \left[\frac{F_{i1}(t) - F_{i1}(u)}{S_i(u) - S_i(t)} \times \frac{S_i(u) - S_i(t)}{S_i(u)} \right]. \end{aligned}$$

Applying lemma 3(i) of GL to the last term since each component is nonincreasing and the numerator is bounded by the denominator, we can see that bounded variation holds. Similarly, it is straightforward to show that all other terms are bounded, so that the assumptions of corollary 1 of GL hold and $D_i(t, s)$ converges weakly to a Gaussian process.

Appendix B

The covariance between Gray's test computed at times (t, s) and times (t^*, s^*) under H_0 is

$$\begin{aligned}
\text{Cov}(G(t, s), G(t, s^*)) &= \int_0^{t \wedge t^*} \sum_{i=1}^2 \frac{R_{3-i}(u, s) R_{3-i}(u, s^*)}{n R_{\cdot}(u, s) R_{\cdot}(u, s^*)} \sum_j a_{ij}(u, s) a_{ij}(u, s^*) \langle dM_{ij1}(u) \rangle \\
&= \int_0^{t \wedge t^*} \sum_{i=1}^2 \frac{R_{3-i}(u, s) R_{3-i}(u, s^*)}{n R_{\cdot}(u, s) R_{\cdot}(u, s^*)} \lambda_{i1}(u) Y_i(u, s) du \\
&= \int_0^{t \wedge t^*} \frac{R_1(u, s) R_2(u, s)}{n R_{\cdot}(u, s)} \left\{ \frac{\lambda_{11}(u) \widehat{S}_1(u-, s) R_2(u, s^*)}{[1 - \widehat{F}_{11}(u-, s)] R_{\cdot}(u, s^*)} + \frac{\lambda_{21}(u) \widehat{S}_2(u-, s) R_1(u, s^*)}{[1 - \widehat{F}_{21}(u-, s)] R_{\cdot}(u, s^*)} \right\} du \\
&\rightarrow_p \int_0^{t \wedge t^*} \frac{r_1(u, s) r_2(u, s)}{r_{\cdot}(u, s)} \left\{ \frac{dF_1^0(u)}{[1 - F_1^0(u)]} \right\}
\end{aligned}$$

Proof of weak convergence and tightness is based on applying corollary 1 of GL as in appendix A. Defining $\eta_n(u, s) = R_2(u, s)/R_{\cdot}(u, s)$ and $\eta(u, s) = r_2(u, s)/r_{\cdot}(u, s)$, showing weak convergence of the first term in (3) requires that we show

$$\lim_{n \rightarrow \infty} \sup_{0 \leq u \leq s \leq \tau} |\eta_n(u, s) - \eta(u, s)| = 0,$$

and

$$\sup_{0 \leq s \leq \tau} \left\{ V_{0 \leq u \leq s} [(Y_1(u, s)/n)^\alpha \eta_n(u, s)] + V_{0 \leq u \leq s} [\rho_1^\alpha Y_1^\alpha(u, s) \eta(u, s)] \right\} = O_p(1), \quad (4)$$

for $0 < \alpha < 1/2$. Since

$$\begin{aligned}
R_i(u, s)/n &= \frac{Y_i(u, s)}{n \widehat{S}_i(u, s)} (F_{i1}(u) - \widehat{F}_{i1}(u-, s)) \\
&\quad + \frac{Y_i(u, s)(1 - F_{i1}(u))}{n} (\widehat{S}_i^{-1}(u-, s) - S_i^{-1}(u)) \\
&\quad + \frac{1 - F_{i1}(u)}{S_i(u)} (Y_i(u, s)/n - y_i(u, s) \rho_i),
\end{aligned}$$

and $\widehat{S}_i(u-, s)$, $\widehat{F}_{i1}(u-, s)$, and $Y_i(u, s)/n$ converge uniformly to $S_i(u)$, $F_{i1}(u)$, and $\rho_i y_i(u, s)$ respectively, then $R_i(u, s)/n$ converges uniformly to $r_i(u, s)$ and $\eta_n(u, s)$ converges uniformly to $\eta(u, s)$. To show bounded variation, consider the first term in (4). Applying Lemma 3(ii), we can write this as

$$\sup |(Y_i(u, s)/n)^\alpha| V_{0 \leq u \leq s} \eta_n(u, s) + \sup |\eta_n(u, s)| V_{0 \leq u \leq s} (Y_i(u, s)/n)^\alpha.$$

Since $Y_i(u, s)/n$ is bounded and nonincreasing, it has bounded variation. Note also that $\eta_n(u, s)$ can be written as a ratio of two bounded nonincreasing functions, where the denominator is greater than the numerator,

$$\eta_n(u, s) = \frac{n^{-1} Y_2(u, s) (1 - \widehat{F}_{21}(u-, s)) \widehat{S}_1(u-, s)}{n^{-1} Y_1(u, s) (1 - \widehat{F}_{11}(u-, s)) \widehat{S}_2(u-, s) + n^{-1} Y_2(u, s) (1 - \widehat{F}_{21}(u-, s)) \widehat{S}_1(u-, s)}.$$

Therefore, $\eta_n(u, s)$ is bounded and by applying Lemma 3(i) of GL we can see that it has bounded variation. Therefore, the bounded variation condition of Corollary 1 of GL is met, and weak convergence of

$$n^{-1/2} \int_0^t \eta_n(u, s) \sum_j a_{1j}(u, s) dM_{ij1}(u),$$

holds. Weak convergence of the other term in (3) can be shown in a similar way.

References

1. Jennison, C.; Turnbull, BW. Group sequential Methods with application to clinical trials. Chapman and Hall/CRC; Boca Raton: 2000.
2. Tsiatis AA. The asymptotic joint distribution of the efficient scores test for the proportional hazards model calculated over time. *Biometrika*. 1981; 68:311–315.
3. Tsiatis AA, Rosner GL, Titchler DL. Group sequential tests with censored survival data adjusting for covariates. *Biometrika*. 1985; 72:365–373.
4. Biliyas Y, Gu M, Ying Z. Towards a general asymptotic theory for Cox model with staggered entry. *Annals of Statistics*. 1996; 25:662–682.
5. Slud EV. Sequential linear rank tests for two-sample censored survival data. *Annals of Statistics*. 1984; 12:551–571.
6. Gu M, Lai KT. Weak convergence of time-sequential censored rank statistics with applications to sequential testing in clinical trials. *Annals of Statistics*. 1991; 19:1403–1433.
7. Jennison C, Turnbull BW. Repeated confidence intervals for the median survival time. *Biometrika*. 1985; 72:619–625.
8. Lin DY, Shen L, Ying Z, Breslow NE. Group sequential designs for monitoring survival probabilities. *Biometrics*. 1996; 52:1033–1041. [PubMed: 8805766]
9. Latouche A, Porcher R. Sample size calculations in the presence of competing risks. *Statistics in Medicine*. 2007; 26:5370–5380. [PubMed: 17955563]
10. 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. *Contemporary Clinical Trials*. 2005; 26:386–396. [PubMed: 15911472]
11. Kalbfleisch, JD.; Prentice, RL. The statistical analysis of failure time data. Wiley; New York: 1980.
12. Resche-Rigon M, Azoulay E, Chevret S. Evaluating mortality in intensive care units: contribution of competing risks analyses. *Critical Care*. 2006; 10:R5. [PubMed: 16356211]
13. Gray RJ. A class of k-sample tests for comparing the cumulative incidence of a competing risk. *Annals of Statistics*. 1988; 116:1141–1154.
14. Lin DY. Non-parametric inference for cumulative incidence functions in competing risks studies. *Statistics in Medicine*. 1997; 16:901–910. [PubMed: 9160487]
15. Slud EV, Wei LJ. Two-sample repeated significance tests based on the modified Wilcoxon statistic. *Journal of the American Statistical Association*. 1982; 77:862–868.
16. Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. *Biometrika*. 1983; 70:659–663.
17. Schoenfeld DA. Sample size formula for the proportional hazards regression model. *Biometrics*. 1983; 39:499–503. [PubMed: 6354290]
18. Latouche A, Porcher R, Chevret S. Sample size formula for proportional hazards modelling of competing risks. *Statistics in Medicine*. 2004; 23:3263–3274. [PubMed: 15490425]
19. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*. 1999; 94:496–509.
20. Wingard JR, Carter SL, Walsh TJ, Kurtzberg J, Small TN, Baden LR, Gersten ID, Mendizabal AM, Leather HL, Confer DL, Maziarz RT, Stadtmauer EA, Bolaos-Meade J, Brown J, DiPersio JF, Boeckh M, Marr KA. Randomized, double-blind trial of fluconazole versus voriconazole for prevention of invasive fungal infection after allogeneic hematopoietic cell transplantation. *Blood*. 2010; 116:5111–5118. [PubMed: 20826719]

Figure 1. Cumulative incidence functions in simulation study, for $p=0.25$ (lower line), $p=0.5$ (middle line), and $p=0.75$ (upper line).

Figure 2. Cumulative incidence of platelet engraftment after an umbilical cord blood transplant, from historical CIBMTR data.

Figure 3. Test statistic (thin line) and boundary (thick line) for monitoring the cumulative incidence of renal toxicity at 100 days.

Table 1

Cumulative type I error rates for Gray's test applied in a group sequential setting, with an O'Brien-Fleming type error spending function, $n = 100$ per group, and administrative censoring only.

Cumulative type I error by interim look						
(A, T)	p	(γ_1, γ_2)	1	2	3	4
(1,1,61)	0.25	(1,1)	0.000	0.005	0.020	0.048
		(1,1.5)	0.000	0.005	0.021	0.047
		(1.5,1.5)	0.000	0.005	0.021	0.048
<hr/>						
	0.5	(1,1)	0.000	0.005	0.020	0.052
		(1,1.5)	0.000	0.005	0.019	0.050
		(1.5,1.5)	0.000	0.005	0.020	0.052
<hr/>						
	0.75	(1,1)	0.001	0.006	0.020	0.048
		(1,1.5)	0.000	0.006	0.021	0.050
		(1.5,1.5)	0.001	0.006	0.020	0.048
<hr/>						
(2,3,2.5)	0.25	(1,1)	0.000	0.005	0.020	0.050
		(1,1.5)	0.000	0.004	0.018	0.046
		(1.5,1.5)	0.000	0.005	0.018	0.044
<hr/>						
	0.5	(1,1)	0.000	0.007	0.021	0.047
		(1,1.5)	0.000	0.005	0.019	0.047
		(1.5,1.5)	0.001	0.005	0.019	0.049
<hr/>						
	0.75	(1,1)	0.001	0.004	0.017	0.049
		(1,1.5)	0.001	0.006	0.021	0.048
		(1.5,1.5)	0.001	0.007	0.022	0.052
<hr/>						
$\alpha(\cdot)$			0.001	0.006	0.021	0.050

Table 2

Cumulative type I error rates for Gray's test applied in a group sequential setting, with an O'Brien-Fleming type error spending function, $n = 100$ per group, and an additional 10% nonadministrative censoring.

(A, T)	p	(γ_1, γ_2)	Cumulative type I error by interim look			
			1	2	3	4
(1,1,61)	0.25	(1,1)	0.000	0.005	0.021	0.047
		(1,1.5)	0.000	0.005	0.021	0.050
		(1.5,1.5)	0.001	0.004	0.019	0.050
	0.5	(1,1)	0.001	0.006	0.020	0.047
		(1,1.5)	0.001	0.007	0.022	0.048
		(1.5,1.5)	0.000	0.006	0.021	0.049
	0.75	(1,1)	0.001	0.006	0.022	0.048
		(1,1.5)	0.001	0.007	0.024	0.051
		(1.5,1.5)	0.001	0.004	0.020	0.048
(2,3,2.5)	0.25	(1,1)	0.000	0.006	0.020	0.051
		(1,1.5)	0.000	0.006	0.021	0.048
		(1.5,1.5)	0.000	0.004	0.0205	0.047
	0.5	(1,1)	0.001	0.007	0.022	0.047
		(1,1.5)	0.001	0.006	0.021	0.047
		(1.5,1.5)	0.000	0.005	0.020	0.050
	0.75	(1,1)	0.001	0.005	0.022	0.048
		(1,1.5)	0.001	0.006	0.022	0.050
		(1.5,1.5)	0.001	0.006	0.020	0.049

Table 3
Group sequential analyses of the cumulative incidence of renal toxicity at quarterly calendar times.

Month	6	9	12	15	18	21	24	27	30	33
Estimate	0.119	0.123	0.097	0.112	0.107	0.088	0.097	0.096	0.090	0.084
SE	0.068	0.041	0.031	0.029	0.025	0.021	0.020	0.019	0.017	0.016
Max Inf	2916	2836	3494	3075	3203	3806	3495	3527	3742	3978
IF	0.074	0.207	0.299	0.376	0.485	0.592	0.707	0.817	0.911	1.000