Evaluating Health Outcomes in the Presence of Competing Risks

A Review of Statistical Methods and Clinical Applications

Ravi Varadhan, PhD,*‡ Carlos O. Weiss, MD, MHS,* Jodi B. Segal, MD, MPH,*† Albert W. Wu, MD, MPH,*† Daniel Scharfstein, ScD,‡ and Cynthia Boyd, MD, MPH*†

Background: An evaluation of the effect of a healthcare intervention (or an exposure) must consider multiple possible outcomes, including the primary outcome of interest and other outcomes such as adverse events or mortality. The determination of the likelihood of benefit from an intervention, in the presence of other competing outcomes, is a competing risks problem. Although statistical methods exist for quantifying the probability of benefit from an intervention while accounting for competing events, these methods have not been widely adopted by clinical researchers.

Objectives: (1) To demonstrate the importance of considering competing risks in the evaluation of treatment effectiveness, and (2) to review appropriate statistical methods, and recommend how they might be applied.

Research Design and Methods: We reviewed 3 statistical approaches for analyzing the competing risks problem: (*a*) cause-specific hazard (CSH), (*b*) cumulative incidence function (CIF), and (*c*) event-free survival (EFS). We compare these methods using a simulation study and a reanalysis of a randomized clinical trial.

Results: Simulation studies evaluating the statistical power to detect the effect of intervention under different scenarios showed that: (1) CSH approach is best for detecting the effect of an intervention if the intervention only affects either the primary outcome or the competing event; (2) EFS approach is best only when the intervention affects both primary and competing events in the same manner; and (3) CIF approach is best when the intervention affects both primary and competing events, but in opposite directions. Using data from a randomized controlled trial, we demonstrated that a comprehensive approach using all 3 approaches provided useful insights on the effect of an intervention on the relative and absolute risks of multiple competing outcomes.

Conclusions: CSH is the fundamental measure of outcome in competing risks problems. It is appropriate for evaluating treatment effects in the presence of competing events. Results of CSH analysis for primary and competing outcomes should always be reported even when EFS or CIF approaches are called for. EFS is appropriate for evaluating the composite effect of an intervention, only when combining different endpoints is clinically and biologically meaningful, and the treatment has similar effects on all event types. CIF is useful for evaluating the likelihood of benefit from an intervention over a meaningful period. CIF should be used for absolute risk calculations instead of the widely used complement of the Kaplan-Meier $(1 - \mathrm{KM})$ estimator.

Key Words: cause-specific hazards, event-free survival, cumulative incidence function, semicompeting risks, comparative effectiveness

(*Med Care* 2010;48: S96–S105)

In evaluating the effectiveness of an intervention, clinicians need to consider the evidence on the effect of that intervention on the outcome of interest. This is straightforward when the outcome is all-cause mortality, but more challenging when the outcome is a specific event, such as myocardial infarction or cause-specific mortality. Patients with multiple conditions may experience outcomes due to other illnesses that reduce their likelihood of deriving benefit from the intervention being considered. Determining the likelihood of benefit from an intervention for a specific outcome, in the presence of competing outcomes, is referred to as the competing risks problem.

This article is divided into 6 sections: (1) definition of the competing risks problem, with illustrations; (2) description of 3 methods for analyzing competing risks: causes-specific hazards (CSH), cumulative incidence function (CIF), and event-free survival (EFS) modeling; (3) a simulation study to demonstrate that a treatment can have varying impact on CSH, CIF, and EFS, and evaluation of the statistical power for testing and detecting the treatment effect, under different scenarios of how the treatment affects the different outcomes;

(4) demonstration using data from a randomized clinical trial of the potential dangers in ignoring competing risks by focusing on either the primary outcome or a composite outcome, and description of how to rigorously address competing risks by using all 3 methods; (5) brief discussion of statistical methods that attempt to draw inferences about the "true" biologic effect of interventions, highlighting the major challenges to doing this; (6) recommendations for appropriate methods according to study objectives. Throughout, we maintain that a rigorous evaluation of effectiveness should consider the effect of the intervention on the rates and probabilities of both primary and competing outcomes.

SECTION 1: DEFINITION OF THE COMPETING RISKS PROBLEM

Biomedical studies often evaluate the effect of a variable (eg, an intervention, exposure, or risk-factor) on the time to occurrence of an event. The time to occurrence may not be observed in all participants due to censoring from withdrawal or the end of study. When there is only one type of event, eg, all-cause mortality, widely-used statistical methods such as Kaplan-Meier curves, log-rank tests, or Cox proportional hazard models are appropriate. However, if there are different types of outcomes, this is a competing risks problem.

A classic competing risks situation occurs when the event types are mutually exclusive, that is, the occurrence of any one type of event precludes that of all other types of events (Fig. 1A). The classic competing risks problem is of major interest to demographers who analyze the patterns and inter-relationships of different causes of mortality.

A semicompeting risks situation arises when a particular type of event can preclude other types of events, but not vice versa (Fig. 1B), such as when a specific disease-related outcome is the primary event of interest but death acts as a competing risk. For example, an individual can have a stroke and survive, but die later from an unrelated cause. For that individual we can observe both time to stroke and time to death. However, if an individual dies without having had a stroke, we can only observe the time to death. This makes it difficult to draw inferences about, for example, the impact of an intervention for preventing stroke, because death prevents us from observing the complete process. Most problems in medicine are semi-competing risks problems. The distinction between the classic and semi-competing risks problems becomes important when the goal is to make inferences about unobservable quantities, for example, the "true" biologic effect of an intervention on stroke in the absence of competing events. This is known as the "net" treatment effect; whereas the treatment effects obtained based on observable variables alone are known as "crude" treatment effects. The methods we present are applicable in both classic and semi-competing risks situations for the estimation of the crude treatment effects. More advanced methods are required for the estimation of net treatment effects (see Section 5).

Illustrations of Competing Risks in the Literature

The competing risks problem dates to the 1760s when Daniel Bernoulli analyzed the effect of small-pox eradication
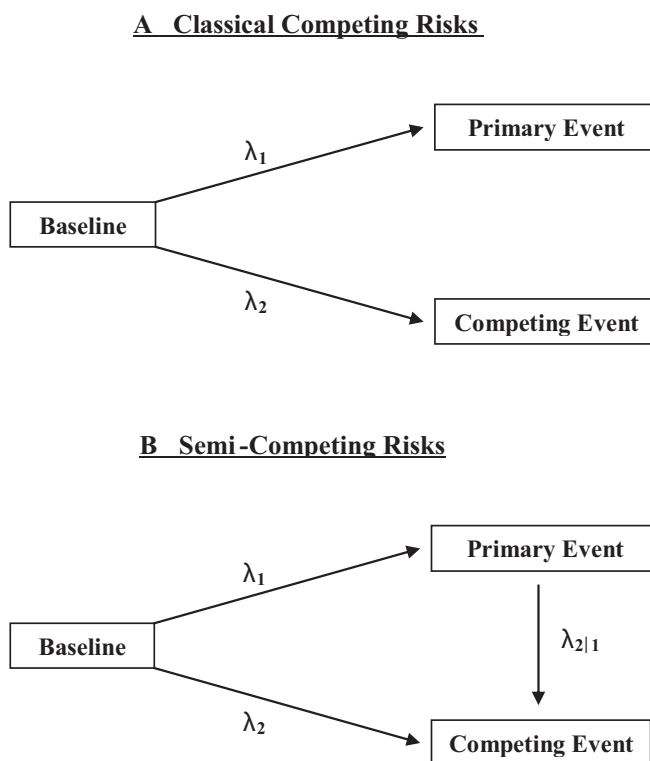

A   Classical Competing Risks

B   Semi-Competing Risks

FIGURE 1. A schematic illustrating (A) classic and (B) semi-competing risks. In (A) either event occurring first makes the other event impossible, whereas in (B) primary event is not possible if the competing event occurs first, but the competing event can still occur if the primary event occurs first. Note λs are event rates, for example, number of events per 1000 person-years.

on life-expectancy.[1] Since the seminal articles of Fix and Neyman, Tsiatis, and Prentice et al, there have been numerous developments in statistical methodology for modeling competing risks.[2–4] Standard methods of survival analysis appropriate for single failure-type problems, like the log-rank test, and the complement of the Kaplan-Meier survival estimate, cannot be used with competing risks, because their validity hinges on the unlikely assumption that the competing events are statistically independent.[5–8] Their use assumes a world in which all competing risks are absent. Many clinical researchers have failed to appreciate this and related subtleties, and have not adopted appropriate competing risks methods.

We present 3 examples to illustrate the situations in which competing risks may pose analytic challenges.

Example 1: An Evaluation of the Absolute Risk of an Outcome

A recent article by Lentine, et al aimed to identify predictors of cerebrovascular events associated with kidney transplantation.[9] They reported the cumulative incidence of cerebrovascular events in each comparison group, using the Kaplan-Meier (KM) method. This method estimates the "net" incidence of cerebrovascular events in the absence of competing causes of mortality, and assumes independence of

primary and competing events. Here death from noncerebro-vascular causes is the competing event. It is biologically implausible that the primary and competing events are independent. Furthermore, it is problematic to imagine the absence of competing causes of mortality. If viewed as estimating the "crude" incidence of stroke, Lentine's estimates are inflated because individuals who died of noncerebrovascular causes are treated as being at risk for stroke even after their death (see Section 2).

Example 2: An Evaluation of the Effect of an Intervention on Disease-Specific Outcomes

We consider the example of a trial of rosuvastatin in people with normal cholesterol levels and low-level inflammation.[10] The primary outcome was a composite variable of myocardial infarction, stroke, arterial revascularization, hospitalization for unstable angina, or death from cardiovascular causes.[11] The investigators reported EFS. As in all cases in which a disease-specific death is defined as the primary event, deaths from other causes create a classic competing risks problem. Competing risks were not addressed, although results of all-cause mortality analyses were reported.

Like most trial reports, this report conflates the relative risk reduction with evidence for a reduction in the absolute risk. When competing risks are present, a relative risk reduction for primary outcome does not necessarily translate into an absolute risk reduction, which also depends on the effect of the treatment on the competing outcome.

Example 3: An Evaluation of the Benefit and Harms of Screening, Over a Long Period

Walter and Covinsky have addressed competing risks in screening evaluation.[12] They present a framework for making individualized decisions based on the baseline risk of dying of a screen-detectable cancer, the length of time that must elapse for screening to yield a survival benefit, and the probability of dying from causes unrelated to the cancer before realizing the intended benefit of screening. An alternative framework has been proposed by Gail et al.[13] This framework weighs the risks and benefits of an intervention, based on the probabilities of the primary and competing events occurring within a meaningful period. This example illustrates that decisions regarding appropriateness of interventions in older adults should also consider competing risks of mortality.

SECTION 2: MAJOR ANALYTIC OPTIONS FOR ADDRESSING COMPETING RISKS PROBLEM

Notation

The observed data, for individual i, in a competing risks problem has the form: $(Y_i, \Delta_i, \Delta_I * \varepsilon_i, X_i)$, where Y_i denotes the observation time; (Δ_i is a binary variable indicating whether the individual was observed to have any one of the J competing events ($\Delta_i = 1$) or was censored ($\Delta_i = 0$); ε_i is a positive integer denoting which one of J competing events was experienced if $\Delta_i = 1$, and is unobserved if $\Delta_i = 0$; and X_i is a vector containing information on covariates (eg, age, gender, treatment indicator). We also note that $Y_i = T_i$, the

event time, when $\Delta_i = 1$, but $Y_i = C_i$, the censoring time, when $\Delta_i = 0$.

There are several analytic options available for addressing competing risks. Each is associated with a different estimand. An estimand is a theoretical property of the underlying population that is the target of inference. It is estimated by taking a finite sample from the population and then applying an estimation procedure to the sample. Some well-known estimands are the mean difference in linear regression and the odds ratio in logistic regression. The survival function of event time T, $S(t) = \Pr(T > t)$, is an estimand, which is typically estimated using the KM product-limit estimator. We focus on 3 estimands that are most widely used in competing risks problems (1): CSH (2) EFS, and (3) CIF.

These estimands are applied in 2 types of situations: (*a*) K-sample testing and (*b*) regression estimation. In K-sample testing, we test whether the true values of the estimands are different between two (K = 2) or more groups, like randomized study participants. In regression modeling, we estimate how the estimands are influenced by covariates.

Cause-Specific Hazard

CSH for type j is the instantaneous rate of experiencing the event of type j at time t, having not experienced any of the J competing events until time t (Table 1). This is denoted as $\lambda_j(t)$. The unit of λ_j is number of events per person-time unit, and hence its value depends on the time scale. There are as many CSH functions as there are types of events. We can also define the conditional CSH, $\lambda_j(t \mid X)$, as the CSH for the event of type j, at time t, for a covariate value X:

$$\lambda_j(t \mid X) = \lim_{h \to 0} \frac{1}{h} \Pr (t < T \le t + h, \varepsilon = j \mid T \ge t, X),$$

$$j = 1,2,\ldots, J \quad (1)$$

The CSH corresponds to the rate of type j event at time t, in the presence of all types of events. The presence of other events affects the "risk set," that is, people who survive up to time t. For example, if a person dies from a type 2 event (eg, cancer-caused death), then that person is no more at risk for the type 1 event (eg, stroke). Thus, the estimation of the CSH for type j event involves removing people who have experienced other types of competing events before time t from the "risk set" and assumes that only those who are not censored prior to time t as being at risk for experiencing the type j event. When X is a single categorical variable (eg, treatment indicator), we may evaluate whether $\lambda_j(t)$ differs across categories of X. When λ_j differs by a constant multiplicative factor across categories (the proportional hazards assumption), we can use the standard log-rank test to test for difference in hazards. When X is continuous or contains multiple covariates, we may use the Cox model to estimate how different variables affect the CSH.[14]

CSH modeling is useful for estimating the average relative risk reduction due to treatment in trials. It is impor-

TABLE 1. Commonly Used Analytic Approaches for Competing Risks Problems: Definition, Estimation Methods, and Recommendations on Usage

Estimand	Definition	Mathematical Definition	Problem Setting	Estimation Methods	When & How to Use
Cause-specific hazard	Instantaneous rate of experiencing the event of type j at time t, having not experienced any event until t	$\lambda_k(t) = \lim_{h \to 0} \dfrac{1}{h} \mathrm{Pr}$ $(t < Y \le t+h, \varepsilon = k \mid Y \ge t),$ $k = 1, 2, \ldots, K$	K-sample* Regression	Log-rank test[31] Cox Model for each $\lambda_k(t)$[14]	Fundamental measure of outcome in competing risks problems. To evaluate impact of risk-factor on the rate of events To estimate average tx effect in RCTs Should always report tx effects on the CSH for primary and competing outcomes
Event-free survival	Instantaneous rate of experiencing the first event, regardless of type, at time t, having been event-free until then	$\lambda(t) = \displaystyle\sum_{k=1}^{K} \lambda_k(t)$ $S(t) = \exp\left[-\displaystyle\int_0^t \lambda(u)\,du\right]$	K-sample Regression	Log-rank test Cox Model for $\lambda(t)$	To evaluate effect of tx or risk factor on overall survival Should be used only when tx has similar effects on different event types Should provide plots of $1 - $ EFS when reporting RCT results
Cumulative incidence function	Probability of experiencing an event of type j in the time interval (o, t)	$F_k(t) = \mathrm{Pr}(Y \le t, \varepsilon = k),$ $k = 1, 2, \ldots, K$ $F_k(t) = \displaystyle\int_0^t \lambda_k(u)\,S(u)\,du$	K-sample Regression	Gray's K-sample test[17] Indirect modeling[18] Direct modeling[19,21,22]	To evaluate impact of tx or risk factor on the absolute risk (probability) of events For absolute risk prediction modeling For risk/benefit analysis of an intervention To identify sub-groups likely to benefit from intervention Should provide plots of all CIFs when reporting results of RCTs

*K-sample testing refers to testing for equality of the estimand between K groups.
RCT indicates randomized controlled trial; tx, treatment; CSH, cause-specific hazard; EFS, event-free survival; CIF, cumulative incidence function.

tant to model the CSH for both the primary and competing outcomes to obtain a comprehensive assessment of the effects of the treatment.

Event-Free Survival

The EFS, S(t), is defined as the probability of not experiencing any event until time t.

$$S(t \mid X) = \mathrm{Pr}(T > t \mid X). \qquad (2)$$

EFS can be written in terms of the overall hazard function, $\lambda(t \mid X)$:

$$S(t \mid X) = \exp\left[-\int_0^t \lambda(u \mid X)\,du\right], \qquad (3)$$

where:

$$\lambda(t \mid X) = \lim_{h \to 0} \frac{1}{h}\mathrm{Pr}(t < T \le t+h \mid T \ge t, X) \qquad (4)$$

The overall hazard function is the sum of all the CSHs:

$$\lambda(t \mid X) = \sum_{j=1}^{J} \lambda_j(t \mid X). \qquad (5)$$

EFS is commonly used, because it simplifies the competing risks problem by transforming it into a univariate survival problem, where only the time to occurrence of the first event, whatever it may be, is considered. A well-known example is the composite outcome strategy in clinical trials. Frequently, endpoints such as the time to myocardial infarction, stroke, and all-cause mortality are combined as a single outcome. Therefore, standard techniques of univariate survival analysis including the log-rank test and Cox regression model can be used (when censoring is independent).

EFS may be used to estimate the average effect of the treatment on the overall survival in a trial. However, it should only be used when the treatment alters the risk of primary and competing outcomes in the same direction, in which case it has greater power to detect a treatment effect than the CSH approach.

Cumulative Incidence Function

CIF of event of type j, $F_j(t \mid X)$, is the probability of experiencing the event of type j in the time interval (0, t]. It is defined as:

$$F_j(t \mid X) = \Pr(T \leq t, \varepsilon = j \mid X), j = 1, 2, \ldots, J \qquad (6)$$

It is related to the CSH of the various types of events as follows:

$$F_j(t \mid X) = \int_o^t \lambda_j(u \mid X) \, S(u \mid X) \, du \qquad (7)$$

The CIF is also known as the absolute cause-specific risk[16] and the crude incidence function.[18] Although the CSH denotes the instantaneous rate of event occurrence, the CIF is the probability of the event's occurrence over a meaningfully long time interval. The CIF is a subdistribution function; it is a nondecreasing function of time with $F_j(t = \infty) = \Pr(\varepsilon = j) \leq 1$.

In Eq. 7, $\lambda_j(u \mid X)$ is the CSH (Eq.1) and $S(u \mid X)$ is the EFS (Eq. 3). Thus, Eq. 7 demonstrates the connection between the 3 estimands: CIF, CSH, and EFS. The CIF of type j event depends upon the CSH of all types of events through its dependence on the EFS, $S(u \mid X)$. Importantly, even if an intervention has no direct impact on λ_1, it can still influence F_1 by affecting the CSH of one of the competing outcomes. Conversely, it is possible for an intervention to have a significant impact on λ_1, but have no impact on F_1. This is an essential feature of the competing risks problem. Because of this dependence of $F_1(t \mid X)$ on all $\lambda_j(t \mid X)$, the testing for equality between K groups and regression estimation for F_1 is not straightforward. The hypothesis of equality of λ_1 between K groups is not equivalent to the hypothesis of equality of F_1. Consequently, the log-rank test is not applicable for testing the equality of CIFs. Gray proposed a test for testing the equality of the CIFs of a particular type of event for K different groups.[19]

Regression modeling for estimating the association between the CIF and covariates, X, is complicated because $F_j(t \mid X)$ depends on X in a complex manner via the individual CSH regression functions (Eq. 7). If the main goal is to estimate the CIF for a given set of covariates X, we can construct Cox models for each CSH and use the definition in Eq. (7). This is the "indirect" approach. It is valid provided the proportional hazards assumptions for the individual CSH models are reasonable.[18] If, however, the goal is not estimation of F_j, but estimation of covariate effect on F_j, the indirect approach is not suitable. Fine and Gray proposed a direct regression model assuming a proportional hazard form for the "hazard rate" of the subdistribution function $\dfrac{dF_j/dt}{1 - F_j}$. The covariates are treated as linear on $\log(-\log(1 - F_j(t \mid X)))$.[19] This model also allows for time-varying effects of covariates on the CIF. Since Fine and Gray,[19] several modeling techniques have been proposed for direct regression estimation of

the CIF.[20–22] These techniques can be implemented using R functions or SAS macros.

Because the CIF is a measure of the actual probability of events, it may be used to evaluate the effect of a treatment on the absolute risk of outcomes. It is better suited for risk-benefit calculations than the CSH. Fine and Gray's regression approach is useful for developing risk prediction models for the primary outcome (Table 1).

Now, we show why the 1 − KM estimator overestimates the CIF. The 1 − KM estimate of cumulative incidence of event type j is:

$$G_j(t \mid X) = \int_o^t \lambda_j(u \mid X) \, S_j(u \mid X) \, du \qquad (8)$$

where:

$$S_j(t \mid X) = \exp\left[-\int_o^t \lambda_j(u \mid X) \, du\right] \qquad (9)$$

Comparison of Eqns. (8) with (7) and (9) with (3) shows that the 1 − KM estimator sets the CSH of all event types, except that of type j, to zero. Consequently, $S_j(t \mid X) > S(t \mid X)$, and $G_j(t \mid X) > F_j(t \mid X)$. Thus, the 1 − KM estimator overestimates the cumulative incidence of event type j. This estimator treats those who die of causes other than j as being at risk for type j event, or equivalently, it means that 1 − KM estimates cumulative incidence for cause j in the hypothetical scenario where other causes of failure are absent.

SECTION 3: A SIMULATION STUDY OF THE 3 ANALYTIC METHODS FOR TWO-SAMPLE TESTING

Three different measures of outcome were defined in Section 2: CSH, EFS, and CIF. We illustrate the use of these estimands for detecting the effect of intervention on the outcomes, under different scenarios of how the intervention affects the different outcomes. Of the 3, CSH is fundamental. EFS and CIF are nonlinear functions of CSH (Eqs.1, 3, 5, and 7). The implication of this is that an intervention can have different effect on the 3 estimands. For example, an intervention may have a significant reduction in the CSH of primary outcome, but it may not affect the EFS or the CIF of primary outcome to a comparable extent.

We consider 2 types of events (J = 2): the primary event of clinical interest ($\varepsilon = 1$) and the competing event ($\varepsilon = 2$). We test whether there is a difference between 2 treatment groups (A = new treatment and B = control) in terms of the EFS, the CSH for the primary event, and the CIF for the primary event.

We use the log-rank test for testing equality of the CSH for the primary event between the groups A and B:
$$H_0: \lambda_1^A(t) = \lambda_1^B(t), \text{ for all } t \qquad (10)$$
We use Gray's K-sample test for testing the equality of CIF for the primary event between the groups A and B:
$$I_0: F_1^A(t) = F_1^B(t), \text{ for all } t \qquad (11)$$

TABLE 2. Probability of Rejecting the Null Hypotheses at Two-sided 0.05 Level (No Censoring; Sample Size Is 100 in Each Group; 10 000 Replications)

Scenarios	Hypothesis Tests		
	CSH (H_0)	CIF (I_0)	EFS (S_0)
(a) Treatment has no effect on primary event and			
1. No effect on competing event			
$\lambda_1^A = 1, \lambda_2^A = 1$	0.055	0.051	0.052
2. Decreases competing event rate			
$\lambda_1^A = 1, \lambda_2^A = 0.6$	0.053	0.25	0.34
3. Increases competing event rate			
$\lambda_1^A = 1, \lambda_2^A = 1.6$	0.050	0.24	0.45
(b) Treatment has no effect on competing event and			
4. Decreases primary event rate			
$\lambda_1^A = 0.6, \lambda_2^A = 1$	0.65	0.50	0.34
5. Increases primary event rate			
$\lambda_1^A = 1.6, \lambda_2^A = 1$	0.69	0.50	0.45
(c) Treatment affects both events and			
6. Decreases primary and competing event rates			
$\lambda_1^A = 0.6, \lambda_2^A = 0.6$	0.70	0.11	0.95
7. Increases primary and competing event rates			
$\lambda_1^A = 1.6, \lambda_2^A = 1.6$	0.63	0.10	0.91
8. Decreases primary and increases competing event rates			
$\lambda_1^A = 0.6, \lambda_2^A = 1.6$	0.58	0.89	0.10
9. Increases primary and decreases competing event rates			
$\lambda_1^A = 1.6, \lambda_2^A = 0.6$	0.73	0.91	0.10

Superscripts A and B refer to the intervention and control arms, respectively; subscripts 1 and 2 refer to primary outcome and competing risks, respectively. See Equations 8–10 for statements of hypotheses H_0, I_0, and S_0, respectively. We let $\lambda_1^B = \lambda_2^B = 1$.
CSH indicates cause-specific hazard model; CIF, cumulative incidence function model; EFS, event-free survival model.

We use the log-rank test for testing equality of EFS. This is equivalent to testing the overall hazard rate between the groups A and B:

S_0: $S^A(t) = S^B(t)$, which is equivalent to testing: $\lambda^A(t) = \lambda^B(t)$.   (12)

For our simulation, we generated 4 sets of uncorrelated random numbers from exponential distributions, representing failure times for the primary, and the competing events and the 2 treatment groups, A (treatment) and B (placebo). The rates for the 4 exponential distributions were chosen to reflect 9 possible effects that an intervention might have on the 2 events. (Table 2). The sample size was 100 in each group. For log-rank tests of H_0 and S_0, we used the survdiff function in the R package, survival.[24] For Gray's K-sample test of I_0, we used the cuminc function in the R package, cmprsk.[24] All the tests were set up to have a nominal type-I error rate of 0.05.

Table 2 (a) shows scenarios under which there is no treatment effect on the primary event. In scenario 1, where treatment has no effect on both the primary and competing events, all 3 tests reject the null hypothesis at the correct probability of 0.05. In scenarios 2 and 3, where only the competing events rate is affected by the treatment, hypothesis of equality of CSH (H_0) is rejected at the correct probability level of 0.05. Hypothesis of equality of CIF of primary event (test I_0) in the 2 groups is rejected at a probability greater than 0.05. This is because the treatment does affect F_1 by affecting λ_2, but not λ_1. We can, in fact, detect this treatment effect on the competing event rate λ_2, with greater statistical power, by per-forming the log-rank test: $\lambda_2^A(t) = \lambda_2^B(t)$, for all t. This underscores the need to perform testing on CSH for all causes.

Table 2 (b) shows scenarios (4 and 5) where the treatment only affects the primary event. These scenarios illustrate that the test of equality of CSH of primary event (test H_0) has the greatest power to detect a treatment effect. In Table 2 (c) the treatment affects both primary and competing events. When the treatment affects both event types in the same direction (scenarios 6 and 7), testing the equality of EFS (test S_0) is most powerful for detecting a treatment effect. However, there are many examples of treatments whose effects on competing events were not anticipated before large trials of sufficient duration were performed.[25] Therefore, results of EFS analysis should be reported along with the effects of treatment on the CSH of primary and competing events. When treatment affects both event types, but in the opposite directions (scenarios 8 and 9), the EFS based test has the least power to detect treatment impact. In this case, the test for equality of CIF of the primary event (test of I_0) has most power for detecting a treatment effect. This includes instances where the treatment reduces primary events (such a beneficial reduction in the incidence of bad outcomes) but increases competing events (such as a harmful increase in the incidence of competing events), a situation that is of paramount relevance to clinical trials. Results of CIF analysis should not be reported without also reporting the effects of treatment on the CSH of primary and competing events.

TABLE 3. Treatment Effect Estimates for 6-Year Event-Free Survival, Cause-Specific Hazards, and 6-Year Cumulative Incidence for Mortality From Different Causes

Coefficient	Event-Free Survival	Cause-Specific Hazard			Cumulative Incidence		
		Cancer	**CVD**	**Other**	**Cancer**	**CVD**	**Other**
Treatment	−0.20 (0.11)	−0.42 (0.17)*	0.19 (0.17)	−0.57 (0.28)*	−0.45 (0.17)*	0.30 (0.17)	−0.53 (0.28)*

*Individual significance at approximately the 5% level at least. Standard errors are given in parentheses.
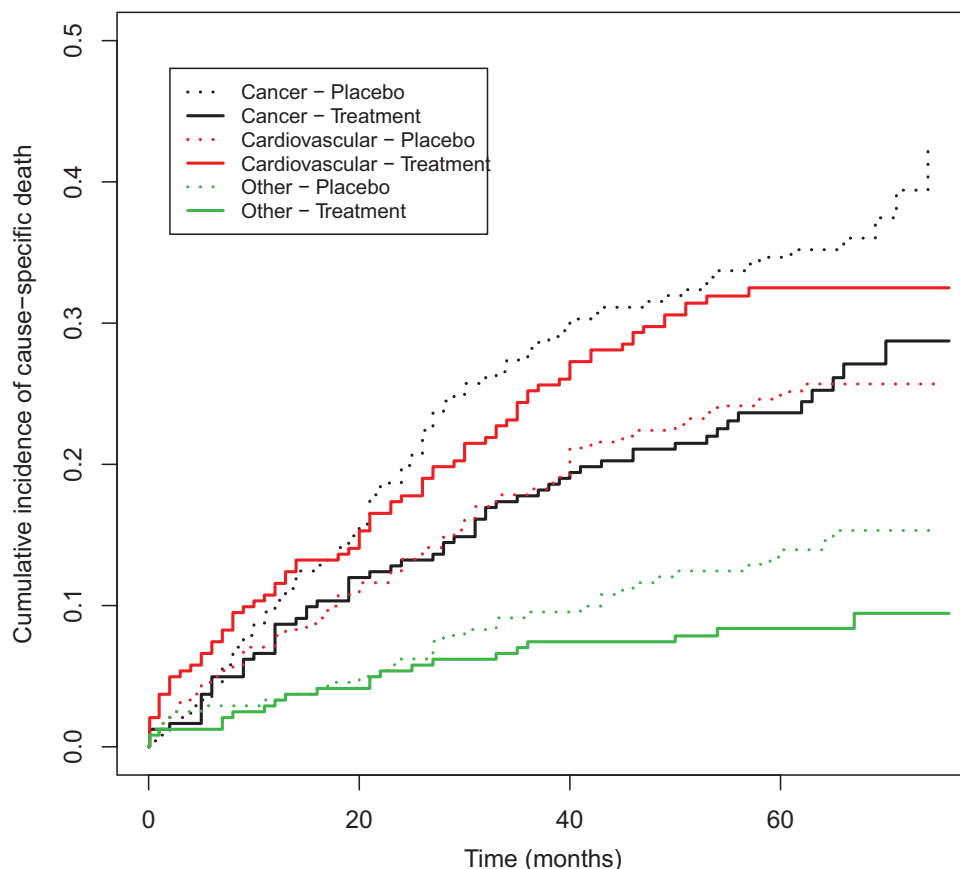

FIGURE 2. Cumulative incidence curves for prostate cancer, cardiovascular disease, and remaining causes of deaths in the treatment and placebo groups. The *P* values are for testing difference between the 2 groups using Gray's K-sample test (K = 2). The *P* values for comparing the CIFs of placebo and treatment groups for different causes of death are: $P <$ 0.01 for cancer, $P = 0.08$ for cardiovascular disease, and $P =$ 0.05 for other causes of death.

SECTION 4: REGRESSION MODELING OF COMPETING RISKS—APPLICATION TO AN RCT

We now use an example of a randomized controlled trial of the effect of diethylstilbestrol for treating prostate cancer to demonstrate a comprehensive approach to the evaluation of treatment effectiveness in the presence of competing risks.[28] The trial outcomes were published in 1980; the data were reanalyzed using CSH in 1986.[27] Participants were allocated to 1 of 4 regimens: placebo, 0.2, 1.0, and 5.0 mg of diethylstilbestrol daily. Following Green and Byar, we dichotomized the treatments and designated placebo and 0.2 mg as the placebo group, and 1.0 and 5.0 mg as the treated group. Death from prostate cancer was the primary outcome; deaths from cardiovascular events and other causes were competing events. We report results for the 483 patients with complete information on all relevant variables. Of these, 344 died, with 149 from prostate cancer, 139 from cardiovascular causes, and 56 from other causes.

Kay reported the results of separate Cox proportional hazards models for each cause of failure, where failures from the other causes were treated as censored observations[27]; we analyze these data by modeling the CIF, using the direct regression approach of Fine and Gray[19] (Table 3).

An analysis of only the primary outcome, prostate cancer, without also examining the treatment effect on competing risks, cardiovascular, and other deaths, would overestimate the benefits of the treatment. From a CSH analysis of all outcome types it is evident that the treatment is protective against prostate cancer and death from "other causes," but raises the risk of cardiovascular deaths (Table 3, first row). A CIF analysis leads to a similar inference. Figure 2 compares the CIFs. Gray's 2-sample test was conducted to determine whether the treatment affected the CIFs (*P* values are reported in figure caption).[17] The cumulative incidence plots reveal that the probability of cardiovascular death is high initially in the treatment group, so that the treatment has
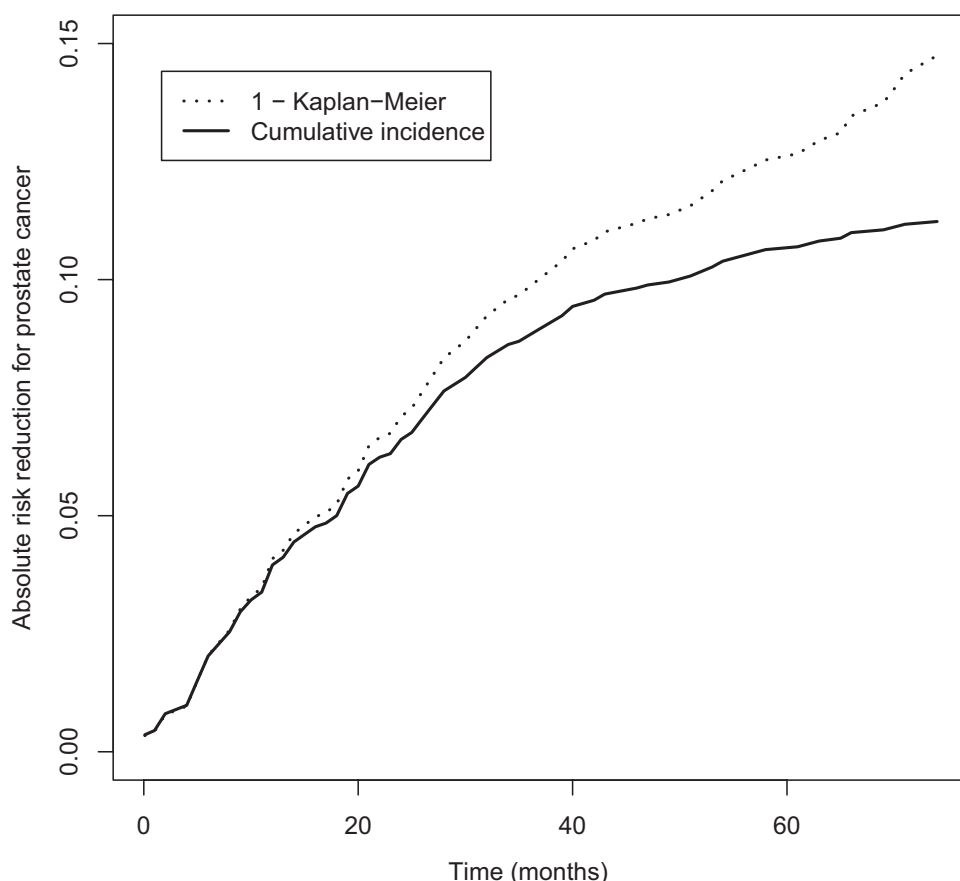
FIGURE 3. A comparison between the $1 - \mathrm{KM}$ and cumulative incidence function (CIF) estimates of the reduction in the absolute risk of prostate cancer due to treatment. The $1 - \mathrm{KM}$ estimator overestimates the reduction in absolute risk of prostate cancer due to treatment. The amount of overestimation increases monotonically with time.

increased cardiovascular death within 1 year of initiating the treatment.

An EFS analysis using all-cause mortality suggests that the treatment is beneficial since it reduced the rate of overall death. This is a dangerously misleading inference, because the data also show that the treatment is related to excess cardiovascular mortality. Furthermore, there can be no good justification for combining cancer and cardiovascular mortality. It is worth noting the lack of power to detect a significant treatment difference on all-cause mortality, because the treatment effect on overall death is diluted by the contrasting treatment effects on cancer plus other mortality and cardiovascular mortality. This is consistent with scenarios 8 and 9 in Table 2 (c), where the EFS based test S_0 had the least power for detecting a treatment effect (Figure 2).

To demonstrate the potential for bias when estimating treatment induced absolute risk reduction (ARR) without accounting for competing events, we evaluated the effectiveness of DES in reducing the absolute risk of prostate cancer. A popular but flawed approach is to estimate this as the difference between the $1 - \mathrm{KM}$ estimates of the cumulative incidences of prostate cancer in the treatment and control groups. The correct approach is to estimate the ARR as the difference between the CIFs of the 2 groups. We compare these 2 approaches in Figure 3, which shows that the $1 - \mathrm{KM}$ overestimates the ARR due to DES treatment. The positive bias of $1 - \mathrm{KM}$ is especially pronounced when the treatment

has a protective effect on the competing risks. In this example, DES increased the risk of cardiovascular death, during the initial 1-year period, but it was also modestly protective against the other competing event, "other" causes of death, during the later study period. Therefore, in the initial 1-year period we do not see any significant bias of $1 - \mathrm{KM}$ estimate of ARR for prostate cancer, but we see a positive bias after 1 year.

SECTION 5: STATISTICAL APPROACHES BASED ON UNOBSERVABLE VARIABLES

The 3 approaches discussed so far are aimed at estimating the effect of interventions on observable quantities: the CSH itself, and CIF and EFS, which are functions of CSH. These are called "crude" estimates of treatment effect (Section 1). The crude estimands are observable in the sense that they can be estimated using only observable information such as time-to-event and the type of event (under the assumption that censoring due to drop-out is completely at random). Furthermore, the modeling assumptions can be checked using observed data. It may be argued that the crude treatment effects do not represent the "true" biologic impacts of the intervention on outcomes, and that such evaluations should be based on the unobservable marginal distributions of the times to outcomes.[23,28,29] Treatment effects on the marginal failure time distributions are called "net" treatment effects. Although it

seems attractive, estimation of net effects is fraught with conceptual and methodological difficulties: (*a*) in what sense does the time to stroke exist for an individual who has died of breast cancer? and (*b*) how do we check assumptions about the dependency between the multivariate failure times when only one of them is observable on each person? Prentice et al provides a cogent critique of this approach.[4] Some recent advances have attempted to address these conceptual issues. Rubin, who called this problem "truncation by death," proposed that a valid comparison of 2 interventions, in terms of their effects on the primary outcome, can only be carried out in such persons who would not experience failure from any of the other competing outcomes under either intervention.[30] However, we cannot know who such people are from observed data. Further assumptions, which are usually untestable, are required to identify such people and to estimate the differences in survival times for the 2 interventions in that group. In summary, assessing the "true" biologic impact of interventions in a competing risks problem is challenging.

SECTION 6: SUMMARY AND RECOMMENDATIONS

We have illustrated the appropriateness of different estimands for addressing different study objectives (Table 1). CSH is the fundamental and most commonly used estimand in competing risks problems. It is appropriate for investigating the effect of an intervention on the rate of occurrence of an event, allowing for the presence of all types of outcome events. For example, the CSH approach might be useful to evaluate the average treatment effect of aspirin to prevent stroke in a population with hypertension, where mortality from other causes can occur. Results of CSH analysis for primary and competing outcomes should always be reported in a competing risks problem.

EFS analysis, frequently used for evaluating the composite effect of an intervention in randomized trials, is appropriate only when combining different endpoints is clinically and biologically meaningful, and when the treatment has similar effects (both in magnitude and sign) on the individual event types. For example, EFS might appropriately be used to study the effectiveness of antiretroviral therapy in patients with HIV/AIDS. The therapy is effective at both reducing opportunistic infections and mortality, which may be combined appropriately in a composite outcome. Results of EFS analysis should always be accompanied by the effects of treatment on the CSH of primary and competing events.

CIF is useful for evaluating the effect of an intervention on the probability of occurrence of an event of a particular type over a meaningful period of time. This is best for absolute risk calculations, risk prediction, risk/benefit analyses of a treatment, and identification of subgroups most likely to benefit from a treatment. CIF might be appropriate for investigating an intervention with both substantial benefits and risks like warfarin for the management of atrial fibrillation. Warfarin reduces the risk of stroke but also raises the risk of hemorrhage. Results of CIF analysis should always be reported alongside the effects of treatment on the CSH of primary and competing events.

The 1 − KM estimator should not be used for estimating the absolute risk when competing risks are present,

because (*a*) its validity is problematic, (*b*) it overestimates the absolute risk, and (*c*) it likely overestimates the ARR due to an intervention.

The relevance and usefulness of the evidence derived from the use of different estimands may depend on the stakeholder. The needs of the clinician may differ from those of a payor or a policymaker, and this may guide the investigator choosing an estimand. It may be informative for the investigator to explore all 3 estimands (CSH, CIF, and EFS) when analyzing data, although it may not be necessary to report all the results, especially if the study objectives clearly dictate the use of a particular estimand.

The technical challenges of implementing the competing risks methodology are diminishing. This methodology is becoming increasingly available in popular software so that investigators can readily implement the methodology for evaluating health outcomes in the presence of competing risks.

This discussion is also relevant to comparative effectiveness research where the goal is identifying effective interventions for use in usual care settings. In these settings, the patients are heterogeneous in their physiological and functional characteristics. This heterogeneity can impact treatment effectiveness by affecting the baseline risk of the primary outcome, treatment responsiveness, treatment induced harm, and the rate of competing events. Application of the methods described in this article may contribute to an improved understanding of risks and benefits of interventions under different settings.

REFERENCES

1. Seal HL. Studies in the history of probability and statistics. XXXV: multiple decrements or competing risks. *Biometrika*. 1977;64:429–439.
2. Fix E, Neyman J. A simple stochastic model of recovery, relapse, death and loss of patients. *Hum Biol*. 1951;23:205–241.
3. Tsiatis A. A nonidentifiability aspect of the problem of competing risks. *Proc Natl Acad Sci USA*. 1975;72:20–22.
4. Prentice RL, Kalbfleisch JD, Peterson AV Jr, et al. The analysis of failure times in the presence of competing risks. *Biometrics*. 1978;34: 541–554.
5. Gooley TA, Leisenring W, Crowley J, et al. Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Stat Med*. 1999;18:695–706.
6. Gaynor JJ. On the use of cause-specific failure and conditional failure probabilities. Examples from clinical oncology data. *J Am Stat Assoc*. 1993;88:400.
7. Pepe MS. Inference for events with dependent risks in multiple endpoint studies. *J Am Stat Assoc*. 1991;86:770.
8. Kay R, Schumacher M. Unbiased assessment of treatment effects on disease recurrence and survival in clinical trials. *Stat Med*. 1983;2:41–58.
9. Lentine KL, Rocca Rey LA, Kolli S, et al. Variations in the risk for cerebrovascular events after kidney transplant compared with experience on the waiting list and after graft failure. *Clin J Am Soc Nephrol*.

2008;3:1090–1101.

10. Ridker PM, Danielson E, Fonseca FA, et al. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. *N Engl J Med*. 2008;359:2195–2207.

11. Freemantle N, Calvert M, Wood J, et al. Composite outcomes in randomized trials: greater precision but with greater uncertainty? *JAMA*. 2003;289:2554–2559.

12. Walter LC, Covinsky KE. Cancer screening in elderly patients: a framework for individualized decision making. *JAMA*. 2001;285:2750–2756.

13. Gail MH, Constantino JP, Bryant J, et al. Weighing the risk and benefits of tamoxifen treatment for preventing breast cancer. *J Natl Cancer Inst*. 1999;91:1829–1846.

14. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. New York, NY: Wiley; 1980.

15. Benichou J, Gail MH. Estimates of absolute cause-specific risk in cohort studies. *Biometrics*. 1990;46:813–826.

16. Korn EL, Dorey FJ. Applications of crude incidence curves. *Stat Med*. 1992;11:813–829.

17. Gray RJ. Class of K-sample tests for comparing the cumulative incidence of a competing risk. *Ann Stat*.1988;16:1141.

18. Cheng SC, Fine JP, Wei LJ. Prediction of cumulative incidence function under the proportional hazards model. *Biometrics*. 1998;54:219–228.

19. Fine JP, Gray RJ. Proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94:496.

20. Fine JP. Regression modeling of competing crude failure probabilities. *Biostatistics*. 2001;2:85–97.

21. Scheike TH. Direct modelling of regression effects for transition probabilities in multistate models. *Scand J Stat*. 2007;34:17.

22. Klein JP, Andersen PK. Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics*. 2005;61:223–229.

23. Robins JM. An analytic method for randomized trials with informative censoring: part 1. *Lifetime Data Anal*. 1995;1:241–254.

24. R Development Care Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2009.

25. Lauer MS, Topol EJ. Clinical trials—-multiple treatments, multiple end points, and multiple lessons. *JAMA*. 2003;289:2575–2577.

26. Byar DP, Green SB. The choice of treatment for cancer patients based on covariate information. *Bull Cancer*. 1980;67:477–490.

27. Kay R. Treatment effects in competing-risks analysis of prostate cancer data. *Biometrics*. 1986;42:203–211.

28. Lin DY, Robins JM, Wei JY. Comparing two failure time distributions in the presence of dependent censoring. *Biometrika*. 1996;83:381–393.

29. Peng L, Fine JP. Regression modeling of semicompeting risks data. *Biometrics*. 2007;63:96–108.

30. Rubin DB. Causal inference without counterfactuals, by A.P. Dawid [comment]. *J Am Stat Assoc*. 2000;95:435–438.

31. Peto R, Peto J. Asymptotically efficient rank invariance test procedures. *J R Stat Soc*. 1972;135:185–206.