OPEN

Introduction to the Analysis of Survival Data in the Presence of Competing Risks

Peter C. Austin, PhD; Douglas S. Lee, MD, PhD; Jason P. Fine, PhD

Abstract—Competing risks occur frequently in the analysis of survival data. A competing risk is an event whose occurrence precludes the occurrence of the primary event of interest. In a study examining time to death attributable to cardiovascular causes, death attributable to noncardiovascular causes is a competing risk. When estimating the crude incidence of outcomes, analysts should use the cumulative incidence function, rather than the complement of the Kaplan-Meier survival function. The use of the Kaplan-Meier survival function results in estimates of incidence that are biased upward, regardless of whether the competing events are independent of one another. When fitting regression models in the presence of competing risks, researchers can choose from 2 different families of models: modeling the effect of covariates on the cause-specific hazard of the outcome or modeling the effect of covariates on the cumulative incidence function. The former allows one to estimate the effect of the covariates on the rate of occurrence of the outcome in those subjects who are currently event free. The latter allows one to estimate the effect of covariates on the absolute risk of the outcome over time. The former family of models may be better suited for addressing etiologic questions, whereas the latter model may be better suited for estimating a patient's clinical prognosis. We illustrate the application of these methods by examining cause-specific mortality in patients hospitalized with heart failure. Statistical software code in both R and SAS is provided. (*Circulation*. 2016;133:601–609. DOI: 10.1161/CIRCULATIONAHA.115.017719.)

Key Words: cumulative incidence function ■ data interpretation, statistical ■ incidence ■ models, statistical ■ proportional hazards models ■ risk assessment ■ survival analysis

Cardiovascular research often focuses on outcomes that are defined as the time to the occurrence of an outcome of interest. Common examples include time to death attributable to any cause, time to cause-specific death (eg, death attributable to cardiovascular causes), and time to the first of any major adverse cardiac event (MACE; eg, cardiovascular death or acute myocardial infarction). Less common examples include time to shock or appropriate therapy in patients undergoing implantable cardiac defibrillator implantation or time to heart transplant in patients on a transplant waiting list. We use the term survival data to refer to data in which the outcome variable denotes the time to the occurrence of the event of interest. Others refer to such data as time-to-event data or event history data. The Kaplan-Meier method for estimating survival functions and the Cox proportional hazards model for estimating the effects of covariates on the hazard of the occurrence of the event are commonly used statistical methods for the analysis of survival data. We refer the interested reader elsewhere for further background on these methods and others for the analysis of survival data.[1–7]

A distinctive feature of survival data is the concept of censoring. If the event of interest is death, then the time of the event is censored for those subjects who are still alive at the end of the study. This means that the statistical analysis must proceed without knowledge of when the subject will die. All that is known about the timing of their death is that it is after the time at which the study ended. More generally, in a follow-up study, subjects who drop out are censored, because they are typically lost to follow-up and the time of the occurrence of their event is unknown. The event time is unobserved for censored subjects: all that is known is that their event time occurred after the time at which they were censored. An implicit concept in the definition of censoring is that, if the study had been prolonged (or if subjects had not dropped out),

---

eventually the outcome of interest would have been observed to occur for all subjects. Conventional statistical methods for the analysis of survival data make the important assumption of independent or noninformative censoring.[1,3,6] This means that, at a given point in time, subjects who remain under follow-up have the same future risk for the occurrence of the event as those subjects no longer being followed (either because of censoring or study dropout), as if losses to follow-up were random and thus noninformative.

A competing risk is an event whose occurrence precludes the occurrence of the primary event of interest. For instance, in a study in which the primary outcome was time to death attributable to a cardiovascular cause, death attributable to a noncardiovascular cause serves as a competing event. A subject who dies of cancer is no longer at risk of death attributable to cardiovascular causes. Regardless of how long the duration of follow-up is extended, a subject will not be observed to die of cardiovascular causes once he or she has died of cancer. Competing risks can also be present in studies in which the different event types include nonfatal outcomes when one is interested in which event type occurred first. Thus, one may have a study with 3 types of events: diagnosis of heart disease, diagnosis of cancer, and death. Each type of event serves as a competing risk, because a diagnosis of cancer before a diagnosis of heart disease or of death precludes either of these latter 2 events from happening first.

Conventional statistical methods for the analysis of survival data assume that competing risks are absent. Two competing risks are said to be independent if information about a subject's risk of experiencing 1 type of event provides no information about the subject's risk of experiencing the other type of event. The methods that we describe in this tutorial apply equally to settings in which competing risks are independent of one another and to settings in which competing risks are not independent of one another. In biomedical applications, the biology often suggests at least some dependence between competing risks, which in many cases may be quite strong. Accordingly, independent competing risks may be relatively rare in biomedical applications.

When analyzing survival data in which competing risks are present, analysts frequently censor subjects when a competing event occurs. Thus, when the outcome is time to death attributable to cardiovascular causes, an analyst may consider a subject as censored once that subject dies of noncardiovascular causes. However, censoring subjects at the time of death attributable to noncardiovascular causes may be problematic. First, it may violate the assumption of noninformative censoring: it may be unreasonable to assume that subjects who died of noncardiovascular causes (and were thus treated as censored) can be represented by those subjects who remained alive and had not yet died of any cause. Second, even when the competing events are independent, censoring subjects at the time of the occurrence of a competing event may lead to incorrect conclusions because the event probability being estimated is interpreted as occurring in a setting where the censoring (eg, the competing events) does not occur. In the cardiovascular example described above, this corresponds to a setting where death from noncardiovascular causes is not a possibility. Although such probabilities may be of theoretical interest, they are of questionable relevance in many practical applications, and generally lead to overestimation of the cumulative incidence of an event in the presence of the competing events.[8–11]

The objective of this tutorial is to introduce readers to statistical methods for the analysis of survival data that account for competing risks. This article is structured as follows. In Statistical Methods for the Analysis of Survival Data in the Presence of Competing Risks, we introduce statistical concepts and methods for the analysis of survival data in the presence of competing events. In Case Study, we illustrate the application and interpretation of these methods by using a data set of subjects hospitalized with heart failure. Finally, in Discussion, we summarize our tutorial and place it in the context of the existing literature.

Statistical Methods for the Analysis of Survival Data in the Presence of Competing Risks

In this section we compare and contrast conventional statistical methods for survival analysis with statistical methods that account for competing risks.

Estimating Crude Incidence

We assume that there is a well-defined baseline time in the cohort and that T denotes the time from baseline time until the occurrence of the event of interest. In the absence of competing risks, the survival function, $S(t)$, describes the distribution of event times: $S(t) = \Pr(T > t)$. One minus the survival function (ie, the complement of the survival function), $F(t) = 1 - S(t) = \Pr(T \leq t)$ describes the incidence of the event over the duration of follow-up. Two key properties of the survival function are that $S(0) = 1$ (ie, at the beginning of the study, the event has not yet occurred for any subjects) and $\lim_{t \to \infty} S(t) = 0$ (ie, eventually the event of interest occurs for all subjects). In practice, the latter assumption may not be required, because the probability of the event over a restricted follow-up period may be <1.

Estimating the incidence of an event as a function of follow-up time provides important information on the absolute risk of an event. In the absence of competing risks, the Kaplan-Meier estimate of the survival function is frequently used for estimating the survival function. One minus the Kaplan-Meier estimate of the survival function provides an estimate of the cumulative incidence of events over time. In the case study that follows, we examine the incidence of cardiovascular death in patients hospitalized with heart failure. When the complement of the Kaplan-Meier function was used, the estimated incidence of cardiovascular death within 5 years of hospital admission was 43.0%. However, using the Kaplan-Meier estimate of the survival function to estimate the incidence function in the presence of competing risks generally results in upward biases in the estimation of the incidence function.[9,10,12] In particular, the sum of the Kaplan-Meier estimates of the incidence of each individual outcome will exceed the Kaplan-Meier estimate of the incidence of the composite outcome defined as any of the event types. Even when the competing events are independent, the Kaplan-Meier estimator yields biases in the probability of the event of interest. The

problem here is that the Kaplan-Meier estimator estimates the probability of the event of interest in the absence of competing risks, which is generally larger than that in the presence of competing risks. Furthermore, the hypothetical population in which competing risks do not exist may not be the population of greatest interest for clinical and/or policy making,[13] as in the cardiovascular setting where noncardiovascular death may be an important consideration.

The Cumulative Incidence Function (CIF), as distinct from $1 - S(t)$, allows for estimation of the incidence of the occurrence of an event while taking competing risk into account. This allows one to estimate incidence in a population where all competing events must be accounted for in clinical decision making. The cumulative incidence function for the kth cause is defined as: $\text{CIF}_k(t) = \Pr(T \leq t, D = k)$, where D is a variable denoting the type of event that occurred. A key point is that, in the competing risks setting, only 1 event type can occur, such that the occurrence of 1 event precludes the subsequent occurrence of other event types. The function $\text{CIF}_k(t)$ denotes the probability of experiencing the kth event before time t and before the occurrence of a different type of event. The CIF has the desirable property that the sum of the CIF estimates of the incidence of each of the individual outcomes will equal the CIF estimates of the incidence of the composite outcome consisting of all of the competing events. Unlike the survival function in the absence of competing risks, $\text{CIF}_k(t)$ will not necessarily approach unity as time becomes large, because of the occurrence of competing events that preclude the occurrence of events of type k. In the case study that follows, when using the CIF, the estimated incidence of cardiovascular death within 5 years of hospital admission was 36.8%. This estimate was 6.2% lower than the estimate obtained using the complement of the Kaplan-Meier function. This illustrates the upward bias that can be observed when naively using Kaplan-Meier estimate in the presence of competing risks.

Hazard Function Regression

A key concept in survival analysis is that of the hazard function. In the absence of competing risks, the hazard function is defined as

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{\text{Prob}(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}.$$

The hazard function, which is a function of time, describes the instantaneous rate of occurrence of the event of interest in subjects who are still at risk of the event. In a setting in which the outcome was all-cause mortality, the hazard function at a given point in time would describe the instantaneous rate of death in subjects who were alive at that point in time. The Cox proportional hazards regression model relates the hazard function to a set of covariates.[2] In the absence of competing events, the Cox proportional hazards regression model can be written as $\log(\lambda(t)) = \log(\lambda_0(t)) + \mathbf{X\beta}$, where $\lambda_0(t)$ denotes the baseline hazard function (ie, the hazard function for a subject whose covariates are all set equal to zero), $\mathbf{X}$ denotes a set of explanatory variables, and $\mathbf{\beta}$ denotes the associated regression parameters. The model can also be written in multiplicative format: $\lambda(t) = \lambda_0(t)\exp(\mathbf{X\beta})$. The Cox model relates the covariates to the hazard function of the outcome of interest

(and not directly to the survival times themselves). The covariates have a relative effect on the hazard function because of the use of the logarithmic transformation. The regression coefficients are interpreted as log-hazard ratios. The hazard ratio is equal to the exponential of the associated regression coefficient. The hazard ratio denotes the relative change in the hazard function associated with a 1-unit increase in the predictor variable. Although the regression coefficients from the Cox model describe the relative effect of the covariates on the hazard of the occurrence of the outcome, the following relationship also holds in the absence of competing risks: $S(t) = S_0(t)^{\exp(X\beta)}$, where $S(t)$ denotes the survival function for an individual whose set of covariates is equal to X, and $S_0(t)$ denotes the baseline survival function (ie, the survival function for a subject whose covariates are all equal to zero). Thus, the relative effect of a given covariate on the hazard of the outcome is equal to the relative effect of that covariate on the logarithm of the survival function. Therefore, in the absence of competing risks, making inferences about the effect of a covariate on the hazard function permits one to make equivalent inferences about the effect of that covariate on prognosis or survival. This direct correspondence between the hazard function and incidence in the absence of competing risks has allowed authors to be imprecise in their language when interpreting the fitted Cox regression model. Authors have been able to conclude that a given risk factor or variable increased the risk of an event, without specifying whether risk denoted the hazard of an event (ie, the rate of the occurrence of the event in those still at risk of the event) or the incidence of the event (ie, the probability of the occurrence of the event).

Competing risks implies that a subject can experience one of a set of different events or outcomes. In this case, 2 different types of hazard functions are of interest: the cause-specific hazard function and the subdistribution hazard function. The former function[9,10] is defined as

$$\lambda_k^{cs}(t) = \lim_{\Delta t \to 0} \frac{\text{Prob}(t \leq T < t + \Delta t, D = k \mid T \geq t)}{\Delta t}.$$

The cause-specific hazard function denotes the instantaneous rate of occurrence of the kth event in subjects who are currently event free (ie, in subjects who have not yet experienced any of the different types of events). If one were considering 2 types of events, death attributable to cardiovascular causes and death attributable to noncardiovascular causes, then the cause-specific hazard of cardiovascular death denotes the instantaneous rate of cardiovascular death in subjects who have not yet experienced either event (ie, in subjects who are still alive). The subdistribution hazard function, introduced by Fine and Gray,[9,10,14] is defined as

$$\lambda_k^{sd}(t) = \lim_{\Delta t \to 0} \frac{\text{Prob}(t < T \leq t + \Delta t, D = k \mid T > t \cup (T < t \cap K \neq k))}{\Delta t}.$$

It denotes the instantaneous risk of failure from the kth event in subjects who have not yet experienced an event of type k. Note that this risk set includes those who are currently event free as well as those who have previously experienced a competing event. This differs from the risk set for the cause-specific hazard function, which only includes those who are currently event free. Using the same example as above, the

subdistribution hazard of cardiovascular death denotes the instantaneous rate of cardiovascular death in subjects who are still alive (ie, who have not yet experienced either event) or who have previously died of noncardiovascular causes. There is a distinct cause-specific hazard function for each of the distinct types of events and a distinct subdistribution hazard function for each of the distinct types of events.

In settings in which competing risks are present, 2 different hazard regression models are available: modeling the cause-specific hazard and modeling the subdistribution hazard function. Both models account for competing risks, but do so by modeling the effect of covariates on different hazard functions. Consequently, each model has its unique interpretation. We refer to these 2 models as cause-specific hazard models and subdistribution hazard models. The second model has also been described as a CIF regression model. The latter name makes explicit the link between the subdistribution hazard and the effect on the incidence of an event. That is, one may directly predict the cumulative incidence for an event of interest using the usual relationship between the hazard and the incidence function under the proportional hazards model. Thus, the subdistribution hazard model allows one to estimate the effect of covariates on the cumulative incidence function for the event of interest.

Lau et al suggest that cause-specific hazard models are "better suited for studying the etiology of diseases, where the [subdistribution hazard model] has use in predicting an individual's risk."[9(p245)] Similarly, Koller and colleagues[15] suggested that subdistribution hazard-based methods are preferable when the focus is on estimating actual risks and prognosis. This echoes the previous distinction made between interpretation of the incidence rate and the hazard rate. Furthermore, the subdistribution hazard may be of greater interest if one is interested in the overall impact of covariates on the incidence of the outcome of interest, even when predictions of incidence are not of direct interest. These arguments suggest that subdistribution hazards models should be used for developing clinical prediction models and risk-scoring systems for survival outcomes, whereas cause-specific hazard models may be more appropriate for addressing epidemiological questions of etiology. The rationale for this suggestion is that the cause-specific hazard function denotes the instantaneous rate of the primary outcome in those subjects who are currently event free. Thus, a regression coefficient from a cause-specific hazard model can be interpreted as the relative effect of the corresponding covariate on the relative increase in the rate of the occurrence of the primary event in subjects who are currently event free. In contrast to this, clinical prediction models and risk-scoring systems are interested in estimating the absolute incidence of the event of interest. As the subdistribution hazard model allows one to model directly the effect of covariates on the incidence of the primary event after accounting for competing events, it lends itself naturally to risk prediction. The need to carefully consider which model is appropriate to address the research question is illustrated in the case study below, in which the effect of cancer is shown to have a different effect on the subdistribution hazard of cardiac death in comparison with its effect on the cause-specific hazard of cardiac death. Cancer was associated with a substantial decrease in the incidence of cardiac death (subdistribution hazard ratio, 0.82), whereas it had no association with the rate of cardiac death in subjects who were still alive (cause-specific hazard ratio, 0.96). An important point to grasp is that in the presence of competing risks, the simple 1-cause Cox model is inadequate. Instead, researchers must be aware of the different hazard functions that are available in the presence of competing risks and decide which one is best suited to their research objectives.

Statistical Software for Competing Risks Analyses

CIFs can be estimated in R using the cuminc function in the cmprsk package; in SAS, one can use the %CIF macro; Stata permits estimation of the CIF using the stcurve function. Note that in SAS/STAT 13.1, %CIF is an autocall macro, and thus does not need to be loaded manually by the analyst.

Cause-specific hazard models can be fit in any statistical software package that permits estimation of the conventional Cox proportional hazards model. One simply treats those subjects who experience a competing event as being censored at the time of the occurrence of the competing event. In R, one can use the coxph function in the survival package, in SAS, one can use PROC PHREG, and in Stata, one can use the stcox function.

Subdistribution hazard models can be fit in R by using the crr function in the cmprsk package. In SAS, PROC PHREG permits estimation of subdistribution hazard models through the use of the 'eventcode' option in the model statement (in SAS/STAT version 13.1). In Stata, the stcrreg function permits estimation of subdistribution hazard regression models.

In the case study below, we used the R (version 3.1.2) statistical programming language and the cmprsk package (version 2.2–6) for all of the statistical analyses. R code for estimating the CIFs, the subdistribution hazard models and the cause-specific hazard models is described in Appendix A in the online-only Data Supplement. SAS code for fitting these functions and models is described in Appendix B in the online-only Data Supplement.

Case Study

Data Sources

The Enhanced Feedback for Effective Cardiac Treatment (EFFECT) Study was designed to assess the effect of public reporting of hospital performance on the quality of care provided to patients with cardiovascular disease in Ontario, Canada.[16] We obtained detailed clinical data by retrospective chart review on patients hospitalized with heart failure (HF) between April 1, 1999 and March 31, 2001 (phase 1) and between April 1, 2004 and March 31, 2005 (phase 2) at 103 hospitals in Ontario, Canada. Trained cardiovascular nurse abstractors collected data on patient demographics, vital signs and physical examination at presentation, medical history, and results of laboratory tests. These data sets were linked by using unique, encoded identifiers and they were analyzed at the Institute for Clinical Evaluative Sciences.

We considered 11 baseline covariates, which make up the EFFECT-HF mortality prediction model: age, systolic blood pressure on admission, respiratory rate on admission, low sodium serum concentration (<136 mEq/L), low serum hemoglobin (<10.0 g/dL), serum urea nitrogen, presence

of cerebrovascular disease, presence of dementia, chronic obstructive pulmonary disease, hepatic cirrhosis, and cancer. The initial sample consisted of 18 284 patients hospitalized with HF. We excluded 255 subjects who were on dialysis for end-stage renal failure, because the EFFECT-HF mortality prediction model was not intended for use in these subjects. We excluded an additional 1792 subjects with missing data on continuous covariates in the EFFECT-HF mortality prediction model. This left 16 237 patients for analysis (8521 patients in phase 1 and 7716 patients in phase 2).

Subjects were linked by using an encoded version of the patient's Ontario health insurance number to the Vital Statistics database maintained by the Ontario Office of the Registrar General. This database contains information on date of death and cause of death for residents of Ontario. Each subject was followed for 5 years from the date of hospital admission for the occurrence of death. For those subjects who died within 5 years of discharge, the cause of death was noted in the Vital Statistics database. We categorized cause of death as cardiovascular versus noncardiovascular: 10 215 (63%) patients died during the 5 years of follow-up. Of these, 5970 (58%) died of cardiovascular causes, and 4245 (42%) died of noncardiovascular causes.

Descriptive Statistical Analyses

Descriptive statistics for patients in the study sample are reported in Table 1. We summarized continuous variables by using medians and the 25th and 75th percentiles, whereas dichotomous variables were summarized by using frequencies and percentages. Patients who died of cardiovascular causes tended to be older and to have lower systolic blood pressure, increased serum urea nitrogen, a higher prevalence of cerebrovascular diseases, a lower prevalence of cancer, chronic obstructive pulmonary disease, and low serum hemoglobin in comparison with patients who died of noncardiovascular causes.

Crude Incidence of Cardiovascular and Noncardiovascular Death

Cumulative incidences of cardiovascular and noncardiovascular death in the overall sample are described in Figure 1, along with the incidence of the composite outcome of all-cause mortality. The cumulative incidence of all-cause mortality is equal to the sum of the cumulative incidences of the 2 cause-specific mortalities. Although the cumulative incidence of cardiovascular death exceeded that of noncardiovascular death at each point in time, the incidence of noncardiovascular death was not negligible in this population. A figure similar to Figure 1 should be presented to estimate cumulative incidence in the presence of competing risk.[13]

In Figure 2, 3 additional curves have been added to the cumulative incidence functions described in Figure 1: estimates of the incidence of each of the 2 outcomes derived from the complement of the Kaplan-Meier estimate of the survival function, along with the sum of the incidence of each outcome derived from the 2 Kaplan-Meier survival functions. Two observations warrant merit. First, as anticipated, the Kaplan-Meier estimate of incidence of each of the 2 outcomes is larger than the corresponding estimate derived from the CIF. The overestimates of incidence when using the Kaplan-Meier estimates are moderately large for both cardiovascular death and noncardiovascular death. Second, the sum of the 2 Kaplan-Meier estimates of incidence is greater than the estimate of incidence of the composite outcome of all-cause mortality. The estimates described in Figure 2 illustrate the incorrect estimates of cumulative incidence that can arise when an analyst naïvely uses the Kaplan-Meier survival function to estimate cumulative incidence.

Hazard Models for Cardiovascular Death

We fit cause-specific and subdistribution hazard models for both cardiovascular death and noncardiovascular death. For each of the 2 causes of death, we regressed the hazard of death on the 11 covariates described above. The estimated hazard ratios, along with their associated confidence intervals, are reported in Table 2. It is important to remember that the 2 models have different interpretations when interpreting the regression coefficients from the cause-specific hazard models and the subdistribution hazard models.

Table 1. Baseline Characteristics of Patients in the EFFECT-HF Sample

Variable	EFFECT-HF Sample (N=16 237)	Cardiac Death (N=5970)	Noncardiac Death (N=4245)
Age	78 (70–84)	81 (74–87)	79 (73–85)
Respiratory rate at admission, breaths/min	24 (20–28)	24 (20–28)	24 (20–28)
Systolic blood pressure at admission, mm Hg	145 (124–169)	140 (120–162)	144 (124–168)
Serum urea nitrogen, mg/dL	24 (17–34)	27 (19–39)	25 (18–38)
Cancer	1889 (12)	631 (11)	764 (18)
Cirrhosis	120 (1)	44 (1)	46 (1)
Cerebrovascular disease	2827 (17)	1267 (21)	784 (18)
Dementia	1457 (9)	743 (12)	510 (12)
Chronic obstructive pulmonary disease	4066 (25)	1499 (25)	1316 (31)
Low serum hemoglobin (<10.0 g/dL)	2068 (113)	781 (13)	773 (18)
Low sodium serum concentration (<136 mEq/L)	3457 (21)	1419 (224)	998 (24)

Cells reporting patient characteristics contain either n (%) for dichotomous variables or median (25th percentile–75th percentile [Q1–Q3]) for continuous variables. EFFECT-HF indicates Enhanced Feedback for Effective Cardiac Treatment Heart Failure Study.
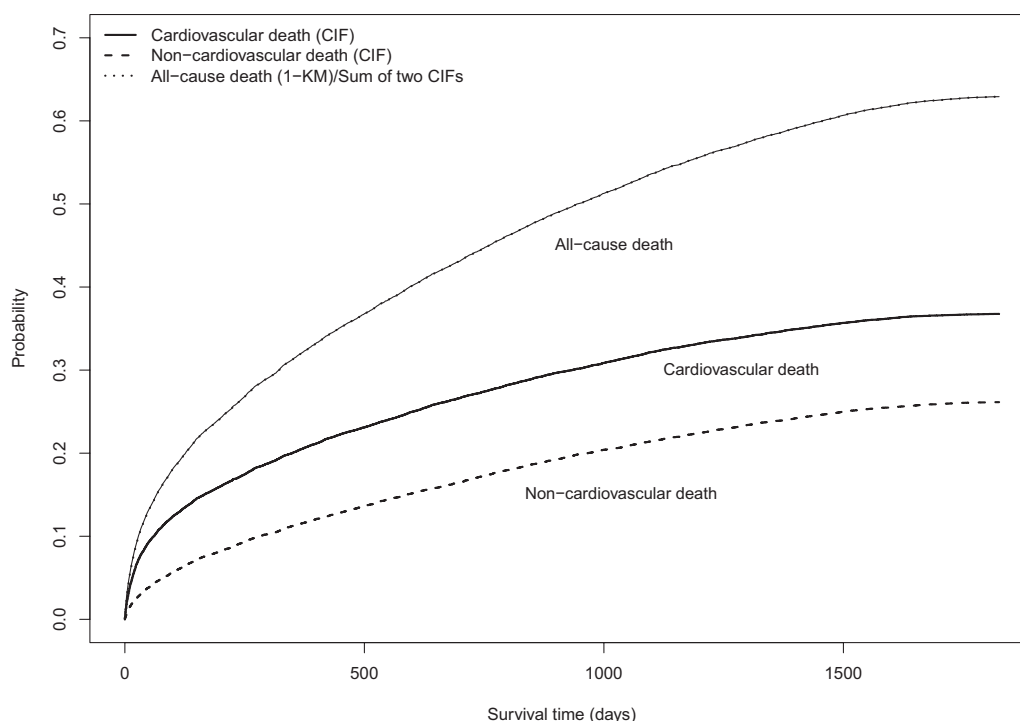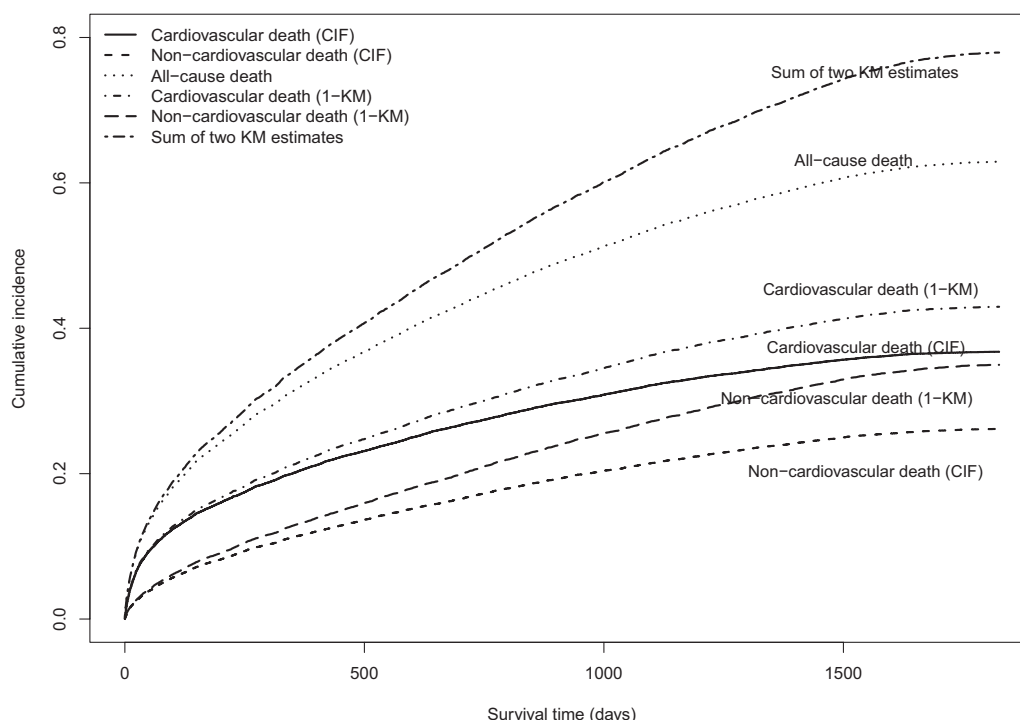
Figure 1. Cumulative incidence functions. CIF indicates cumulative incidence function; and KM, Kaplan–Meier.

A 10-year increase in age increased the relative incidence of cardiac death by 42%, whereas it increased the relative incidence of noncardiac death by 14%. Similarly, a 10-year increase in age increased the cause-specific hazard of cardiac death by 52%, whereas it increased the cause-specific hazard of noncardiac death by 31%. Thus, age had a more pronounced effect on both the incidence and cause-specific hazard of cardiac mortality than on noncardiac mortality. Furthermore, age had a more pronounced effect on the cause-specific hazard of a given outcome than it did on the incidence of the same outcome.

The presence of cancer had a small and nonsignificant effect on the cause-specific hazard of cardiac death, whereas

Figure 2. Cumulative incidence functions and Kaplan–Meier estimates. CIF indicates cumulative incidence function; and KM, Kaplan–Meier.

Table 2. Hazard Ratios (and 95% Confidence Intervals) From Cause-Specific and Subdistribution Hazard Models for Cardiac and Noncardiac Death

Variable	Subdistribution Hazard Model		Cause-Specific Hazard Model	
	Cardiac Death	Noncardiac Death	Cardiac Death	Noncardiac Death
Age (per 10-y increase in age)	1.42 (1.38–1.46)	1.14 (1.10–1.17)	1.52 (1.48–1.56)	1.31 (1.27–1.35)
Respiratory rate (per 10 breaths/min increase)	1.12 (1.08–1.17)	1.07 (1.02–1.12)	1.18 (1.14–1.23)	1.15 (1.09–1.20)
Systolic blood pressure (per 10 mm Hg increase)	0.93 (0.92–0.94)	1.00 (0.99–1.01)	0.91 (0.90–0.92)	0.96 (0.95–0.97)
Urea nitrogen	1.02 (1.02–1.02)	1.01 (1.01–1.01)	1.03 (1.02–1.03)	1.02 (1.02–1.02)
Cancer	0.82 (0.75–0.89)	1.85 (1.71–2.01)	0.96 (0.89–1.04)	1.85 (1.71–2.00)
Cirrhosis	1.12 (0.82–1.54)	1.49 (1.10–2.02)	1.20 (0.89–1.62)	1.66 (1.24–2.22)
Cerebrovascular disease	1.26 (1.18–1.35)	1.01 (0.93–1.09)	1.30 (1.22–1.38)	1.13 (1.04–1.22)
Dementia	1.22 (1.11–1.33)	1.35 (1.22–1.50)	1.48 (1.37–1.60)	1.74 (1.59–1.92)
COPD	1.03 (0.97–1.1)	1.42 (1.33–1.52)	1.14 (1.08–1.21)	1.53 (1.44–1.64)
Low hemoglobin	0.87 (0.80–0.94)	1.49 (1.37–1.62)	0.95 (0.88–1.02)	1.50 (1.39–1.63)
Low sodium concentration	1.14 (1.07–1.22)	1.13 (1.05–1.22)	1.18 (1.11–1.25)	1.19 (1.11–1.28)

Each cell contains the hazard ratio and associated 95% confidence interval for the given covariate and the given hazard model. COPD indicates chronic obstructive pulmonary disease.

it increased the cause-specific hazard of noncardiac death by 85%. This illustrates how the apparent reduction in the absolute risk of cardiac death from cancer may be explained via the effect of cancer on noncardiac death.[17]

In examining the estimated hazard ratios for the different outcomes and different types of hazard models, one notes that some variables have a qualitatively similar effect on the incidence of cardiac death as on the incidence of noncardiac death. However, other variables have a qualitatively different effect on the incidence of cardiac death in comparison with the effect on the incidence of noncardiac death. Furthermore, some variables have a qualitatively similar effect on the incidence of a given type of death as on the cause-specific hazard for the same type of death. As described elsewhere, the interpretation of these results requires careful consideration of the effects of variables on competing causes of death.[17] As an example, a strong and opposing effect of a variable on the cause-specific hazard of a competing event may lead to an indirect effect on the cumulative incidence of the event of interest. To be concrete, a strong prognostic factor for the cause-specific hazard for cardiovascular death might lead to an apparent decrease in the cumulative incidence for noncardiovascular death when such factor has no effect on the cause-specific hazard for noncardiovascular death. This indirect effect of the prognostic factor for cardiovascular death occurs because noncardiovascular death cannot occur in those who die of cardiovascular causes and hence have a decreased risk for that event.

Discussion

The analysis of survival data plays a key role in cardiovascular research. Competing risks are prevalent in much of cardiovascular research. Failure to account correctly for competing events can result in adverse consequences, including overestimation of the probability of the occurrence of the event and mis-estimation of the magnitude of relative effects of covariates on the incidence of the outcome. Koller et al[15] found that competing risks were present in a large majority of studies

published in a sample of high-impact journals. This suggests that it is crucial that investigators be aware of appropriate methods to account for competing risks when analyzing survival data. We have provided a brief, nontechnical, introduction to statistical methods to account for the presence of competing risks. We refer the interested reader to introductions and reviews of differing levels of statistical depth.[8–13,15,17–21] We summarize our recommendations in Table 3.

When estimating crude incidence of the outcome of interest, it is inappropriate to use the complement of the Kaplan-Meier survival function, because this will result in an overestimate of the incidence of the outcome of interest when competing risks are present. Instead, authors and analysts are encouraged to use the CIF. Our analyses illustrated the overestimates of the incidence of both cardiac death and noncardiac death that occurred when using the complement of the Kaplan-Meier survival curves to estimate incidence.

It is important to present results for all causes and for both cause-specific hazard functions and subdistribution hazard functions.[17] Such a practice enables a more complete understanding not only of the effects of prognostic factors, but also

Table 3. Recommendations for Analyzing Competing Risk Survival Data

- Cumulative incidence functions (CIFs) should be used to estimate the incidence of each of the different types of competing risks. Do not use the Kaplan-Meier estimate of the survival function for this purpose.
- Researchers need to decide whether the research objective is on addressing etiologic questions or on estimating incidence or predicting prognosis.
- Use the Fine-Gray subdistribution hazard model when the focus is on estimating incidence or predicting prognosis in the presence of competing risks.
- Use the cause-specific hazard model when the focus is on addressing etiologic questions.
- In some settings, both types of regression models should be estimated for each of the competing risks to permit a full understanding of the effect of covariates on the incidence and the rate of occurrence of each outcome.

of the absolute risks of the different outcomes in the study sample. This is critical for decision makers, who may need to consider the risks of all events when making clinical decisions. Someone with high risk of noncardiovascular mortality, but with a low risk of cardiovascular mortality may be treated differently than someone with low risk of noncardiovascular mortality and high risk of cardiovascular mortality. For instance, the latter patient may be an excellent candidate for early cardiovascular revascularization procedures, whereas the former patient may not be a suitable candidate for early revascularization.

There is a growing awareness of the impact of competing risks when developing prognostic models. Wolbers et al[20] compared the performance of a standard Cox survival model with that of a Fine-Gray model for predicting the incidence of coronary heart disease in older women. They found that the standard Cox model overestimated the 10-year risk of coronary heart disease in comparison with the estimate from the Fine-Gray model. They classified 18% of subjects as being high risk based on the conventional Cox model, whereas they classified only 8% of subjects as being high risk when using the Fine-Gray model. They attributed this discrepancy to the increased risk of death attributable to competing risks in this elderly population. This example serves as a constructive warning about the use of inappropriate statistical methods when estimating patient prognosis. We recommend that analysts use the Fine-Gray subdistribution hazard model when the focus is on estimating incidence or predicting prognosis in the presence of competing risks.

Given the availability of software, analyses of the cumulative incidence function have become increasingly popular and widely reported in recent years. Biases may occur in naïvely estimating the cumulative incidence of the event of interest using the Kaplan-Meier estimator, as well as in naïvely estimating the effects of covariates on the cumulative incidence function via a proportional hazards model for the cause-specific hazard function. The impact of incorrectly treating competing events as censoring events in these analyses has practical importance. In general, the greater the percentage of competing events, the greater the potential for bias in treating competing events as censoring events. Absolute percentages of competing events of >10% merit serious consideration, demanding careful attention to the scientific objectives of the analysis and the appropriate choice of end point and method of analysis.

In our case study, we showed that a variable can have an effect on the incidence of an outcome that differs from its effect on the cause-specific hazard of the outcome. Such an observation has been made by Lau et al[9] previously. This highlights the need to examine the effect of covariates on both the incidences and the cause-specific hazard functions of all event types to develop a complete understanding of the relationships between the covariates and the different competing events. Competing risks entail events that preclude the occurrence of the outcome of interest. In some cases, further clarity may be required when deciding on what constitutes a competing risk before embarking on the analysis. For example, if a subject develops 1 form of heart disease, can he or she subsequently develop a second form of heart disease, or are the 2 conditions mutually exclusive, thus precluding the later second disease?

Such clinical questions require resolution in the design phase before conducting the statistical analysis.

In summary, competing risks are prevalent in cardiovascular research. We encourage analysts to take full advantage of the range of statistical methods for the analysis of survival data that have been developed in the statistical literature. Investigators need to be cognizant of the presence of competing risks and their potential effect on statistical analyses. Researchers should select the appropriate method to address the study objectives and ensure that the analysis results are interpreted correctly.

Sources of Funding

Disclosures

None.

References

1. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data.* 2nd ed. New York, NY: John Wiley and Sons; 2002.
2. Cox D. Regression models and life tables (with discussion). *J Roy Stat Soc - Series B* 1972;34:187–220.
3. Cox D, Oakes D. Analysis of Survival Data. London, UK: Chapman & Hall; 1984.
4. Therneau TM, Grambsch PM. *Modeling Survival Data: Extending the Cox Model.* New York, NY: Springer-Verlag; 2000.
5. Klein JP, Moeschberger ML. Survival Analysis: Techniques for Censored and Truncated Data. New York, NY: Springer-Verlag; 1997.
6. Lawless JF. *Statistical Models and Methods for Lifetime Data.* New York, NY: John Wiley & Sons; 1982.
7. Rao SR, Schoenfeld DA. Survival methods. *Circulation.* 2007;115:109–113. doi: 10.1161/CIRCULATIONAHA.106.614859.
8. Berry SD, Ngo L, Samelson EJ, Kiel DP. Competing risk of death: an important consideration in studies of older adults. *J Am Geriatr Soc.* 2010;58:783–787. doi: 10.1111/j.1532-5415.2010.02767.x.
9. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. *Am J Epidemiol.* 2009;170:244–256. doi: 10.1093/aje/kwp107.
10. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. *Stat Med.* 2007;26:2389–2430. doi: 10.1002/sim.2712.
11. Satagopan JM, Ben-Porat L, Berwick M, Robson M, Kutler D, Auerbach AD. A note on competing risks in survival data analysis. *Br J Cancer.* 2004;91:1229–1235. doi: 10.1038/sj.bjc.6602102.
12. Varadhan R, Weiss CO, Segal JB, Wu AW, Scharfstein D, Boyd C. Evaluating health outcomes in the presence of competing risks: a review of statistical methods and clinical applications. *Med Care.* 2010;48(6 suppl):S96–S105. doi: 10.1097/MLR.0b013e3181d99107.
13. Pepe MS, Mori M. Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Stat Med.* 1993;12:737–751.
14. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc.* 1999;94:496–509.
15. Koller MT, Raatz H, Steyerberg EW, Wolbers M. Competing risks and the clinical community: irrelevance or ignorance? *Stat Med.* 2012;31:1089–1097. doi: 10.1002/sim.4384.

16. Tu JV, Donovan LR, Lee DS, Wang JT, Austin PC, Alter DA, Ko DT. Effectiveness of public report cards for improving the quality of cardiac care: the EFFECT study: a randomized trial. *JAMA*. 2009;302:2330–2337. doi: 10.1001/jama.2009.1731.

17. Latouche A, Allignol A, Beyersmann J, Labopin M, Fine JP. A competing risks analysis should report results on all cause-specific hazards and cumulative incidence functions. *J Clin Epidemiol*. 2013;66:648–653. doi: 10.1016/j.jclinepi.2012.09.017.

18. Pintilie M. Competing Risks: A Practical Perspective. West Sussex: John Wiley & Sons Ltd; 2006.

19. Szychowski JM, Roth DL, Clay OJ, Mittelman MS. Patient death as a censoring event or competing risk event in models of nursing home placement. *Stat Med*. 2010;29:371–381. doi: 10.1002/sim.3797.

20. Wolbers M, Koller MT, Witteman JC, Steyerberg EW. Prognostic models with competing risks: methods and application to coronary risk prediction. *Epidemiology*. 2009;20:555–561. doi: 10.1097/EDE.0b013e3181a39056.

21. Wolbers M, Koller MT, Stel VS, Schaer B, Jager KJ, Leffondré K, Heinze G. Competing risks analyses: objectives and approaches. *Eur Heart J*. 2014;35:2936–2941. doi: 10.1093/eurheartj/ehu131.