Cumulative Incidence in Competing Risks Data and Competing Risks Regression Analysis

Haesook T. Kim

Abstract   Competing risks occur commonly in medical research. For example, both treatment-related mortality and disease recurrence are important outcomes of interest and well-known competing risks in cancer research. In the analysis of competing risks data, methods of standard survival analysis such as the Kaplan-Meier method for estimation of cumulative incidence, the log-rank test for comparison of cumulative incidence curves, and the standard Cox model for the assessment of covariates lead to incorrect and biased results. In this article, we discuss competing risks data analysis which includes methods to calculate the cumulative incidence of an event of interest in the presence of competing risks, to compare cumulative incidence curves in the presence of competing risks, and to perform competing risks regression analysis. A hypothetical numeric example and real data are used to compare those three methods in the competing risks data analysis to their respective counterparts in the standard survival analysis. The source and magnitude of bias from the Kaplan-Meier estimate is also detailed.

Cumulative incidence of an event is often of interest in medical research and is frequently presented in medical articles. The graphical display of the cumulative incidence function (i.e., failure probabilities) over time is intuitive and appealing. However, methods for estimating cumulative incidence function must be clearly understood. The Kaplan-Meier (KM) method has been a widely used tool for estimating survival function and cumulative incidence function, a complementary of survival function (1). This method is conceptually easy to understand and easy to calculate. However, if there is more than one type of event (or failure), and if these events are dependent, KM estimates are biased. This bias arises because the KM method assumes that all events are independent, and thus, censors events other than the event of interest. For example, disease relapse is an event of interest in studies of allogeneic hematopoietic stem cell transplantation (HSCT), as is mortality related to complications of transplantation (transplant-related mortality or TRM). Relapse and TRM are not independent in this setting because these two events are likely related to immunologic effector mechanisms following HSCT, whereby efforts to reduce TRM may adversely affect the risk of relapse; moreover, patients who die from TRM cannot

be at further risk of relapse. Therefore, the KM method is inappropriate for estimating the cumulative incidence rate of relapse in the presence of TRM because it censors TRM.

In addition to estimating the cumulative incidence of an event, it is often of interest to determine whether there is a difference in the cumulative incidence rates among different groups. In standard survival analysis, this is done using the log-rank test to compare curves generated with the KM method (2). In the presence of competing risks, however, this is inappropriate, for the same reason given above. Instead, Gray investigated this issue and proposed a class of tests for comparing the cumulative incidence curves of a particular type of failure among different groups in the presence of competing risks (3).

Finally, when there is a difference in the cumulative incidence curves among different treatment groups, it is also important to determine whether this difference is solely due to the treatment or to the confounding factors, such as age or baseline disease stage. In standard survival analyses, this question is usually addressed by fitting a Cox proportional hazards model (4, 5). In fact, one may attempt to construct a cause-specific standard Cox model for a particular failure treating other competing risks censored. However, the effect of a covariate on an event from either a cause-specific (e.g., relapse) model or cause nonspecific (e.g., relapse and TRM combined) model may be very different from the effect of the covariate of the event (e.g., relapse) in the presence of competing risks (e.g., TRM). Fine and Gray (6) and Klein and Andersen (7) proposed a method for direct regression modeling of the effect of covariates on the cumulative incidence function for competing risks data. As in any other regression analysis, modeling cumulative incidence functions for competing risks can be used to identify potential prognostic factors for a particular failure in the presence of competing risks, or to assess a prognostic factor of interest after adjusting for other potential risk factors in the model.

Author's Affiliation: Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, Massachusetts

Requests for reprints: Haesook T. Kim, Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, 44 Binney Street, M218 Boston, MA 02115. Phone: 617-638-6547; Fax: 617-632-2444; E-mail: kim.haesook@jimmy.harvard.edu.

In this article, we will present the three aforementioned methods in the analysis of competing risks data in more detail. We will compare each method to its counterpart in standard survival analysis, and show why the latter are inadequate in the presence of competing risks. Using hypothetical numeric example and real data, we will demonstrate the use of these three methods, compare the results to the results obtained from standard survival analysis, and discuss the source and magnitude of bias that arises from standard methods.

Estimating Cumulative Incidence in the Presence of Competing Risks

KM method. Survival probability at a certain time is a conditional probability of surviving beyond that time, given that an individual has survived just prior to that time. This conditional probability can be estimated in a study as the number of patients who are alive or event-free without loss to follow-up at that time, divided by the number of patients who were alive just prior to that time. The KM estimate of survival probability is then the product of these conditional probabilities up until that time. Similarly, the probability of failure at a certain time is a conditional probability of having an event at that time, given that an individual has not had an event just prior to that time. This conditional probability can be estimated in a study as the probability of surviving just prior to that time multiplied by the number of patients with the event at that time, divided by the number of patients at risk. Then cumulative incidence of a failure is the sum of these conditional probabilities over time. Of note, "event" and "failure" are used interchangeably in the literature, and the event of interest could be death from any cause, relapse, treatment-related mortality, and stroke in cardiovascular disease. Individuals without an event are "censored"—meaning that their observed failure time is incomplete because these individuals have not failed yet, whereas the failure time for individuals with an event is fully observed.

Suppose that 10 patients enter a clinical trial at different time points (i.e., staggered entry), are followed until a specified time, and have independent and distinct failure times. Suppose that the event of interest is relapse. Patients who fail from other causes or who are still alive at the specified time point are censored. Patients who are lost to follow up during the study period are considered alive and relapse-free at the last time seen alive and thus censored. In order to calculate the KM estimate of cumulative relapse-free survival probability, the observed failure times for all patients need to be ordered from the smallest to the largest, irrespective of the censoring status, and each failure time is paired with the information of censoring status. For example, let t_0 denote the time zero or the time of study entry, t_1 denote the smallest observed failure time, and t_{10} denote the largest observed failure time to the event. Then the KM estimate of relapse-free survival probability, $S_{KM}^{rel}(t)$, is

$S_{KM}^{rel}(t_0) = 1$ (because everyone is alive at study entry)

$$S_{KM}^{rel}(t_1) = S_{KM}^{rel}(t_0) * \left[\frac{\text{no. of relapse} - \text{free patients at } t_1}{\text{no. of patients at risk just prior to } t_1} \right]$$

$$S_{KM}^{rel}(t_2) = S_{KM}^{rel}(t_1) * \left[\frac{\text{no. of relapse} - \text{free patients at } t_2}{\text{no. of patients at risk just prior to } t_2} \right]$$

$$S_{KM}^{rel}(t_3) = S_{KM}^{rel}(t_2) * \left[\frac{\text{no. of relapse} - \text{free patients at } t_3}{\text{no. of patients at risk just prior to } t_3} \right]$$

$$S_{KM}^{rel}(t_{10}) = S_{KM}^{rel}(t_9) * \left[\frac{\text{no. of relapse} - \text{free patients at } t_{10}}{\text{no. of patients at risk just prior to } t_{10}} \right]$$

The KM estimate of incidence of relapse at a specified time point is then the probability of relapse-free survival just prior to that time, multiplied by the number of relapses at that time, divided by the number of patients at risk (that is, alive, relapse-free, and not lost to follow-up) just prior to that time. Cumulative incidence is then a sum of these conditional probabilities over time. More specifically, the cumulative incidence using the KM method, denoted as CI_{KM}^{rel} is calculated as follows:

$CI_{KM}^{rel}(t_0) = 1 - S_{KM}^{rel}(t_0) = 0$ (because all subjects were alive and relapse-free at study entry)

$$CI_{KM}^{rel}(t_1) = 1 - S_{KM}^{rel}(t_1) = CI_{KM}^{rel}(t_0)$$
$$+ S_{KM}^{rel}(t_0) * \left[\frac{\text{no. of relapses at } t_1}{\text{no. of patients at risk just prior to } t_1} \right]$$

$$CI_{KM}^{rel}(t_2) = 1 - S_{KM}^{rel}(t_2) = CI_{KM}^{rel}(t_1)$$
$$+ S_{KM}^{rel}(t_1) * \left[\frac{\text{no. of relapses at } t_2}{\text{no. of patients at risk just prior to } t_2} \right]$$

$$CI_{KM}^{rel}(t_3) = 1 - S_{KM}^{rel}(t_3) = CI_{KM}^{rel}(t_2)$$
$$+ S_{KM}^{rel}(t_2) * \left[\frac{\text{no. of relapses at } t_3}{\text{no. of patients at risk just prior to } t_3} \right]$$

$$CI_{KM}^{rel}(t_{10}) = 1 - S_{KM}^{rel}(t_{10}) = CI_{KM}^{rel}(t_9)$$
$$+ S_{KM}^{rel}(t_9) * \left[\frac{\text{no. of relapses at } t_{10}}{\text{no. of patients at risk just prior to } t_{10}} \right]$$

The KM estimate of cumulative incidence function is simple and useful for a single end point such as relapse. However, if deaths due to other causes exclusive of relapse (i.e., TRM) are also of interest, the KM method does not capture the dependency of competing risks. For example, in allogeneic HSCT for patients with hematologic malignancies, both relapse and TRM are of equal importance to patients as well as physicians. As previously mentioned, efforts to modify the relapse rate through immune effector mechanisms may adversely affect TRM rates (vice versa is also true), and therefore, relapse and TRM are not independent events. If the KM method is used to estimate the cumulative incidence of relapse in the presence of TRM, patients dying of TRM are censored. However, unlike truly censored observations, patients who die of TRM cannot then relapse, and hence, their risk for relapse is 0. Therefore, the survival probability is overestimated and thus cumulative incidence of relapse is also overestimated (as shown in the example below) in the KM method.

Competing Risks Method

The cumulative incidence of relapse in the presence of TRM as a competing risk (CR) can be calculated similarly as in the KM method. The difference lies in the probability of an event-free survival just prior to a certain time. In the KM method, this probability is a relapse-free probability $[S_{KM}^{rel}(t)]$, whereas it is a relapse- and TRM-free probability in the CR method $[S_{KM}^{rel,trm}(t)]$. More specifically, the cumulative incidence of relapse in the presence of TRM as a competing risk, denoted as CI_{CR}^{rel}, is calculated as

$$CI_{CR}^{rel}(t_0) = 1$$

$$CI_{CR}^{rel}(t_1) = CI_{CR}^{rel}(t_0) + S_{KM}^{rel,\,trm}(t_0)$$
$$* \left[\frac{\text{no. of relapses at } t_1}{\text{no. of patients at risk just prior to } t_1} \right]$$

$$CI_{CR}^{rel}(t_2) = CI_{CR}^{rel}(t_1) + S_{KM}^{rel,\,trm}(t_1)$$
$$* \left[\frac{\text{no. of relapses at } t_2}{\text{no. of patients at risk just prior to } t_2} \right]$$

$$CI_{CR}^{rel}(t_3) = CI_{CR}^{rel}(t_2) + S_{KM}^{rel,\,trm}(t_2)$$
$$* \left[\frac{\text{no. of relapses at } t_3}{\text{no. of patients at risk just prior to } t_3} \right]$$

$$CI_{CR}^{rel}(t_{10}) = CI_{CR}^{rel}(t_9) + S_{KM}^{rel,\,trm}(t_9)$$
$$* \left[\frac{\text{no. of relapses at } t_{10}}{\text{no. of patients at risk just prior to } t_{10}} \right]$$

where $S_{KM}^{rel,TRM}(t)$ denotes the KM estimate of survival probability for the joint events of relapse and TRM. In the CR method, unlike the KM method, patients dying from TRM are counted as events when calculating the event-free survival, $S_{KM}^{rel,TRM}(t)$, and only patients who are truly alive are considered at risk for relapse. Because the relapse-free survival $[S_{KM}^{rel}(t)]$ in the KM method is always greater than or equal to the relapse- and TRM-free survival $[S_{KM}^{rel,TRM}(t)]$ in the CR method, the cumulative incidence using the KM method is always greater than or equal to the cumulative incidence using the CR method. The cumulative incidence of TRM in the presence of relapse as a competing risk can be calculated similarly. The magnitude of overestimation in the KM method depends on the incidence rate levels of competing events.

Example 1: a hypothetical numeric example. We shall show the computation of cumulative incidence function of relapse below using the KM and CR methods. Suppose that there are 10 patients with the ordered failed or censored times shown below

$$10(R), 20+, 35(R), 40(T), 50+, 55(R), 70(T), 71(T), 80(R), 90+$$

where + denotes a censored time, (R) denotes relapse, and (T) denotes death due to treatment-related complications (TRM). The detailed calculation of KM estimate for cumulative incidence of relapse (CI_{KM}^{rel} or KM CIR, for the purpose of simplicity) is presented in Table 1. At the time of study entry,

all patients were alive, and thus, the KM estimate of relapse-free survival (S_{KM}^{rel} or RFS, for the purpose of simplicity) at $t = 0$ was 1 and the KM CIR at $t = 0$ was 0. If one relapse occurs at $t = 10$, then the RFS at $t = 10$ is 0.9 and the KM CIR at $t = 10$ is 0.1 (10 patients were alive just prior to $t = 10$, but because one death occurred at $t = 10$, the RFS dropped to 0.9 at $t = 10$). Because the patient with the second smallest observed failure time was alive (thus censored), the RFS at $t = 20$ was 0.9 and the KM CIR at $t = 20$ was 0.1. The patient with the third smallest observed time died from relapse at $t = 35$, and thus the RFS at $t = 35$ was 0.79, i.e., the second patient who was alive and thus censored at $t = 20$ was excluded from the denominator of the number of patients at risk just prior to $t = 35$ because this patient's survival information beyond $t = 20$ is unknown. The KM CIR at $t = 35$ was thus 0.21. The RFS at $t = 20$ (0.9) was multiplied by the incidence rate of relapse at $t = 35$ because only those patients who were alive just prior to $t = 35$ were subject to relapse at $t = 35$. If the patient with the fourth smallest observed time, $t = 40$, dies from TRM, then the KM method treats this as no event, and thus censors the observation. Therefore, the RFS at $t = 40$ was 0.79, and the corresponding KM CIR was 0.21. Notably, the RFS and KM CIR at $t = 40$ was the same as the ones at $t = 35$, even though the calculation is slightly different. In fact, the RFS and CIR remain the same until there is another incidence of relapse at $t = 55$. Because of this, RFS and KM CIR are often calculated at failure times only. At $t = 55$, the RFS drops to 0.63, and the KM CIR increases to 0.37. At all time points, the sum of RFS and KM CIR was equal to 1 [i.e., $S_{KM}^{rel}(t) + CI_{KM}^{rel}(t) = 1$].

This can be contrasted with the detailed calculation of cumulative incidence function using the CR method presented in Table 1. In the CR method, death due to TRM is counted as death in the calculation of event-free survival (EFS) at each time point. At $t = 40$, the relapse- and TRM-free survival ($S_{KM}^{rel,TRM}$ or EFS) in the CR method is 0.68, whereas the RFS in the KM method is 0.79 in Table 1. However, this difference does not immediately affect the CIR because there is no incidence of relapse at $t = 40$. At $t = 55$, the cumulative incidence of relapse using the CR method (CI_{CR}^{rel} or CR CIR) is 0.35, which is lower than the KM CIR of 0.37. This is because the KM method treats one death due to TRM at $t = 40$ as censored, which results in overestimation of the RFS at $t = 50$ (0.79). At $t = 80$, the difference between KM CIR (0.69) and CR CIR (0.48) is more apparent (0.21). This is because the RFS (0.63) in the KM method is substantially overestimated, compared with the EFS (0.27) in the CR method at $t = 71$, and consequently, the KM CIR is also overestimated. Following similar steps, the cumulative incidence of TRM in the presence of relapse as a competing risk (CI_{CR}^{TRM} or CIT) is presented in Table 1. In the CR method, at all time points, the sum of EFS, CR CIR, and CR CIT is equal to 1 $[S_{KM}^{rel,TRM}(t) + CI_{CR}^{rel}(t) + CI_{CR}^{rel}(t) = 1]$. Figure 1 shows the cumulative incidence curves using the KM and CR methods in the presence of TRM as a competing risk. Two curves jump at the same time whenever relapse occurs, but the magnitude of each jump is smaller in the CR method.

Example 2a: myeloablative versus nonmyeloablative allogeneic HSCT for patients >50 years of age—a real data example. To illustrate the KM and CR methods with real data, we considered a typical competing risks data set of relapse and TRM presented in Alyea et al. (8) as an example. One hundred and fifty-two patients over the age of 50 who underwent

Table 1. Cumulative incidence of relapse and TRM

Patient no.	Follow-up time	Event	No. at risk	KM method			
				No. of relapses	No. of censored observations	RFS	KM CIR
1	10	Relapse	10	1	0	$1 \times (9/10) = 0.9$	$0 + 1 \times (1/10) = 0.1$
2	20+	Alive	9	0	1	$0.9 \times (9/9) = 0.9$	$0.1 + 0.9 \times (0/9) = 0.1$
3	35	Relapse	8	1	0	$0.9 \times (7/8) = 0.7875$	$0.1 + 0.9 \times (1/8) = 0.2125$
4	40	TRM	7	0	1	$0.7875 \times (7/7) = 0.7875$	$0.2125 + 0.7875 \times (0/7) = 0.2125$
5	50+	Alive	6	0	1	$0.7875 \times (6/6) = 0.7875$	$0.2125 + 0.7875 \times (0/6) = 0.2125$
6	55	Relapse	5	1	0	$0.7875 \times (4/5) = 0.63$	$0.2125 + 0.7875 \times (1/5) = 0.37$
7	70	TRM	4	0	1	$0.63 \times (4/4) = 0.63$	$0.37 + 0.63 \times (0/4) = 0.37$
8	71	TRM	3	0	1	$0.63 \times (3/3) = 0.63$	$0.37 + 0.63 \times (0/3) = 0.37$
9	80	Relapse	2	1	0	$0.63 \times (1/2) = 0.315$	$0.37 + 0.63 \times (1/2) = 0.685$
10	90	Alive	1	0	1	$0.315 \times (1/1) = 0.315$	$0.685 + 0.315 \times (0/1) = 0.685$

NOTE: +, censored time; RFS (S_{KM}^{rel}), KM estimate for relapse-free survival; KM CIR (CI_{KM}^{rel}), KM estimate of cumulative incidence of relapse; EFS ($S_{KM}^{rel,TRM}$), KM estimate for relapse- and TRM-free survival; CR CIR (S_{CR}^{rel}), CR estimate of cumulative incidence of relapse; CR CIT (CI_{CR}^{TRM}), CR estimate of cumulative incidence of TRM.

T cell–replete HLA-matched allogeneic transplantation from 1997 to 2002 at our institution were included. Of these 152 patients, 81 underwent myeloablative and 71 underwent nonmyeloablative transplantation. In the study, as in many HSCT studies, both relapse and TRM were equally important. Figure 2A compares the CIR curves between the KM and CR method for 81 patients who underwent myeloablative transplantation. The 4-year CIR was 30% using the CR method but 50% using the KM method. This rather large difference in the CIR suggests that there is a high incidence of a competing risk of TRM. In fact, approximately two thirds of the deaths in that study were due to TRM and not relapse-related. Figure 2B presents the cumulative incidence of TRM (CIT) using both methods. The 4-year CIT was 50% using the CR method and 59% using the KM method. This difference is smaller compared with the difference in CIR. This is because the competing risk of relapse is lower than the incidence of TRM after myeloablative allogeneic transplantation. As a consequence, if the KM method is used, one could falsely claim that the CIR is high (almost as high as CIT) after myeloablative allogeneic transplantation. As in our hypothetical example above (Fig. 1), the jumps in Fig. 2A and B occur at the same time in both methods because, irrespective of the method used, the occurrence time of an event does not change.

Comparison of Cumulative Incidence Curves in the Presence of Competing Risks

In addition to estimating the cumulative incidence of an event, comparing cumulative incidence curves among different treatment groups is useful when selecting the appropriate treatment for a particular patient. When there are no competing risks, a Mantel-Haenzel log-rank test (2) is used to compare KM cumulative incidence curves. This test is obtained by constructing a 2×2 table at each distinct failure time, comparing the failure rates between two groups, and then combining tables over time. In CR data, this test is inappropriate. Gray (3) proposed a class of tests for comparing the cumulative incidence curves of a particular type of failure among different

groups in the presence of competing risks. A detailed description of the method is beyond the scope of this review. Simply stated, the Gray test compares weighted averages of the hazards of cumulative incidence function using the cumulative incidence estimation equation $[CI_{CR}^{rel}(t)]$ described in the previous section and the test statistic has a χ^2 distribution.

Example 2b. Using the same data set presented in example 2a, Fig. 3 [Fig. 1 in Alyea et al. (8)] shows cumulative incidences of relapse and TRM using the CR method after myeloablative and nonmyeloablative transplantation. The 3-year cumulative incidence of TRM (CIT) was 50% after myeloablative transplantation, but was 32% after nonmyeloablative transplantation ($P = 0.01$). The CIT was calculated in the presence of relapse as a competing risk and this difference was tested using the Gray method. Similarly, the 3-year CIR was 30% after myeloablative transplantation and 46% after nonmyeloablative transplantation ($P = 0.052$). In comparison with the CR method, the 3-year CIR using the KM method was 50% after myeloablative and 61% after nonmyeloablative transplantation ($P = 0.35$). These results suggest that even though the cumulative incidence rates of combined events (relapse and TRM) are similar between two types of transplantation (80% versus 78%), the immunologic effects are different.
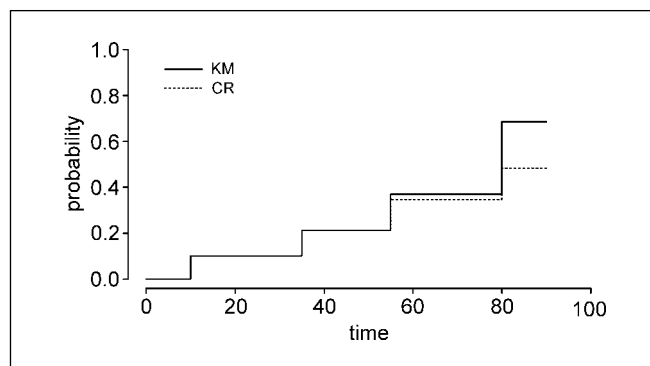
Fig. 1. Comparison of cumulative incidence of relapse between the KM and CR methods using a numerical example.

Table 1. Cumulative incidence of relapse and TRM (Cont'd)

| No. of relapses or TRM	No. of censored observations	EFS	CR CIR	CR CIT
			CR method	
1	0	1 × (9/10) = 0.9	0 + 1 × (1/10) = 0.1	0 + 1 × (0/10) = 0
0	1	0.9 × (9/9) = 0.9	0.1 + 0.9 × (0/9) = 0.1	0 + 0.9 × (0/9) = 0
1	0	0.9 × (7/8) = 0.7875	0.1 + 0.9 × (1/8) = 0.2125	0 + 0.9 × (0/8) = 0
1	0	0.7875 × (6/7) = 0.675	0.2125 + 0.7875 × (0/7) = 0.2125	0 + 0.7875 × (1/7) = 0.1125
0	1	0.675 × (6/6) = 0.675	0.2125 + 0.675 × (0/6) = 0.2125	0.1125 + 0.675 × (0/6) = 0.1125
1	0	0.675 × (4/5) = 0.54	0.2125 + 0.675 × (1/5) = 0.3475	0.1125 + 0.675 × (0/5) = 0.1125
1	0	0.54 × (3/4) = 0.405	0.3475 + 0.54 × (0/4) = 0.3475	0.1125 + 0.54 × (1/4) = 0.2475
1	0	0.405 × (2/3) = 0.27	0.3475 + 0.405 × (0/3) = 0.3475	0.2475 + 0.405 × (1/3) = 0.3825
1	0	0.27 × (1/2) = 0.135	0.3475 + 0.27 × (1/2) = 0.4825	0.3825 + 0.27 × (0/2) = 0.3825
0	1	0.135 × (0/1) = 0.135	0.4825 + 0.135 × (0/1) = 0.4825	0.3825 + 0.135 × (0/1) = 0.3825

CR Regression Analysis

In standard survival analyses, when testing a group (or treatment) difference, the usual steps are to present KM survival or cumulative incidence curves, test the difference of these survival curves using a log-rank test, and perform a proportional hazards regression analysis. This last step can be used to identify important potential prognostic factors, and/or to assess the prognostic significance of a given factor after adjusting for other risk factors. In a competing risks situation, the equivalent steps are to generate cumulative incidence curves using the CR method described above, test the difference between cumulative incidence curves using the Gray method, and perform a CR regression analysis. A standard Cox proportional hazards model analysis is not adequate in the presence of competing risks because the cause-specific Cox model treats competing risks of the event of interest as censored observations, and the cause-specific hazard function does not have a direct interpretation in terms of survival probability.

Because the simple relationship between a single end point and a single crude hazard does not hold in the presence of competing risks, Fine and Gray (6) and Klein and Andersen (7) proposed a direct regression modeling of the effect of covariates on the cumulative incidence function for CR data. These models distinguish between patients who are still alive and those who have already failed from competing causes and allow direct inference regarding the effects of covariates on the cumulative incidence function. The Fine and Gray method is based on proportional hazards model, whereas the Klein and Andersen method is based on the pseudo-values from a jackknife statistic from the cumulative incidence curve. When the two methods were compared in a real data example, results from both approaches were in close agreement (7).

Example 2c. Returning to our example of myeloablative versus nonmyeloablative allogeneic HSCT for patients >50 years of age, the cumulative incidence curves in Fig. 3 indicate that myeloablative transplantation is associated with an increased risk for TRM (P = 0.01) and nonmyeloablative transplantation is associated with an increased risk for relapse (P = 0.052). To investigate these differences further, we fitted regression models using the standard Cox and the Fine and Gray approach for relapse and TRM. In the Cox model, relapse and TRM are considered jointly in the outcome; in the Fine

and Gray model, they are considered individually. Both models used the type of transplantation (myeloablative versus nonmyeloablative) as a covariate, along with patient age, unrelated versus related donor, bone marrow versus peripheral blood progenitor cells, unfavorable versus favorable prognosis, second versus first transplant, FK506 versus cyclosporine-based acute graft-versus-host disease (GVHD) prophylaxis, and donor-recipient sex mismatch. The results are shown in Table 2. In both models, the type of transplantation was not a significant factor for outcome after adjusting for pretransplant characteristics. This can be explained by confounding. Of the 40 patients with myeloablative transplantation who died of
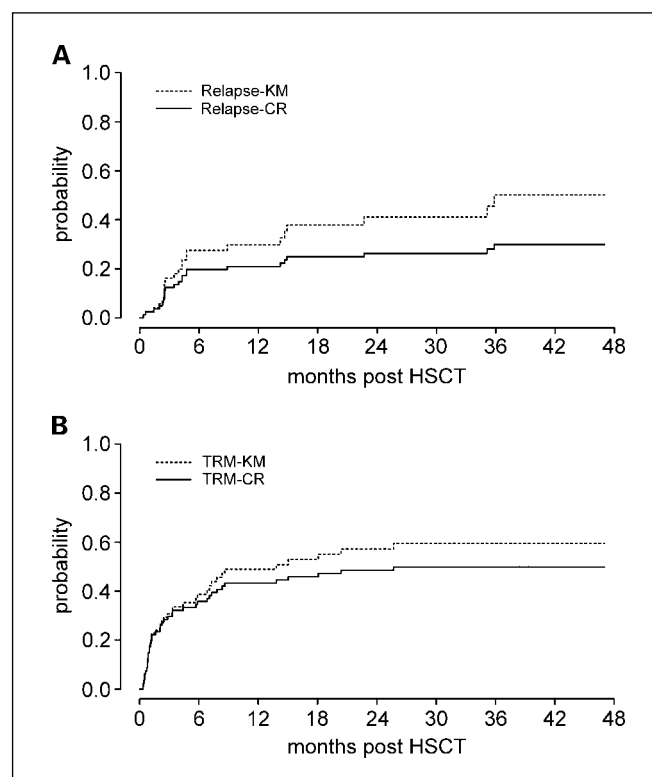
Fig. 2. Myeloablative versus nonmyeloablative allogeneic HSCT for patients >50 years of age. *A,* comparison of cumulative incidence of relapse between the KM and CR methods. *B,* comparison of cumulative incidence of TRM between the KM and CR methods.
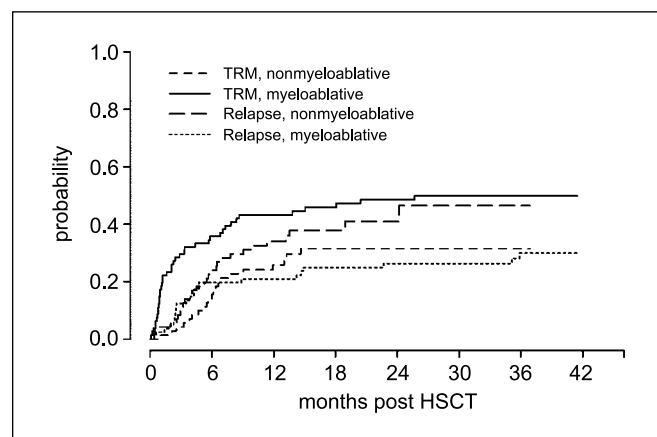
Fig. 3. Cumulative incidence of relapse and TRM using the CR method after myeloablative or nonmyeloablative transplantation for patients over the age of 50.

TRM, 35 received bone marrow stem cells. Of five patients with nonmyeloablative transplantation, four received bone marrow stem cells and died of TRM. Therefore, the difference in the cumulative incidence of TRM (Fig. 3) was confounded in part by the source of progenitor stem cells (i.e., bone marrow versus peripheral blood). This is apparent in the CR regression model for TRM in Table 2. In that model, the hazard ratio (HR) for TRM of bone marrow use was 2.24 ($P = 0.057$). In contrast, in the Cox model for relapse and TRM combined, the HR for bone marrow as compared with peripheral blood stem cell use was 1.13 ($P = 0.71$).

Similarly, in Fig. 3, it seems that nonmyeloablative transplantation is associated with an increased risk for relapse. However, this is also confounded by unfavorable risk status at the time of transplantation. There were 51 patients with relapse: 23 after myeloablative and 28 after nonmyeloablative transplantation. Of these 51 patients, 41 (24 nonmyeloablative, 17 myeloablative patients) had unfavorable risk characteristics at the time of transplantation. When unfavorable prognosis is controlled for in the multivariate model, nonmyeloablative transplantation is no longer associated with an increased risk for relapse (HR = 0.57, $P = 0.33$). Note that results from the CR regression analysis were not presented in the published article (8).

CR regression analysis is also useful to identify other prognostic factors, other than type of transplantation, for each type of failure. In Table 2, donor-recipient sex mismatched were associated with a decreased risk for relapse ($\beta = -0.66$, HR = 0.52, $P = 0.04$). Although this is indicated in the Cox model of relapse and TRM combined events in Table 2 ($\beta = -0.39$, HR = 0.68, $P = 0.053$), it is unknown from this model whether this is due to relapse or TRM prevention. In fact, the hazard of relapse and TRM for sex-mismatched patients compared with sex-matched patients point to the opposite direction: $\beta = -0.069$ with HR = 0.52 for relapse versus $\beta = 0.179$ with HR = 1.20 for TRM, even though the latter is not statistically significant. Again, the Cox model is not appropriate for identifying risk factors for cumulative incidence of a specific event in the presence of competing risks.

Discussion

Competing risks occur frequently in cancer research even though their presence may not always be recognized at the time of analysis. In many cancer treatments, efforts have been made to decrease disease recurrence but often at the cost of increased toxicity due to the treatment. Examples of competing risks in HSCT studies include: incidence of acute GVHD and 100-day death without development of acute GVHD, death due to disease and due to transplant-related complications, and relapse and chronic GVHD for long-term survivors. Similarly, in umbilical cord blood transplantation, a reduced incidence of acute GVHD with an increased incidence of infection-related morbidity and mortality have been reported (9). In breast cancer studies, competing risks include death due to local/regional recurrence of disease versus distant metastases and distant metastases versus non–cancer death in the study of radiation therapy (3).

When competing risks are present, there are three ways to analyze the data: (*a*) analysis of the event of interest ignoring CR, (*b*) analysis of joint events as a single end point, and (*c*) analysis of CR. The first approach is incorrect and will lead to erroneous results as shown in the previous sections. The magnitude of the error could be substantial if the incidence of CR is high, or could be minimal if the incidence of CR is low. However, one could not know the effect of competing risks a priori unless CR analysis is done. The second approach is correct, but is too limited to address various important

Table 2. Results from the myeloablative versus nonmyeloablative HSCT study

Variable	Relapse and TRM: Cox			Relapse: CRR			TRM: CRR		
	β	HR	*P*	β	HR	*P*	β	HR	*P*
Patient age	0.017	1.02	0.56	0.049	1.05	0.27	−0.030	0.97	0.48
Unrelated donor	0.047	1.05	0.83	−0.164	0.85	0.63	0.271	1.31	0.35
Bone marrow	0.118	1.13	0.71	−0.781	0.46	0.11	0.808	2.24	0.057
Nonmyeloablative	−0.428	0.65	0.25	−0.556	0.57	0.33	0.057	1.06	0.90
Unfavorable prognosis	0.426	1.53	0.06	0.891	2.44	0.02	0.255	1.29	0.38
Second transplant	0.258	1.29	0.40	0.089	1.09	0.83	0.305	1.36	0.46
aGVHD prophylaxis	0.027	1.03	0.94	0.080	1.08	0.87	−0.647	0.52	0.31
Sex mismatch	−0.385	0.68	0.053	−0.659	0.52	0.04	0.179	1.20	0.50

NOTE: β, parameter estimate for each variable; Cox, Cox proportional hazards model for relapse and TRM combined as a single event; CRR, competing risks regression model using the Fine and Gray method.

research questions. The combination of the second and third approaches is a comprehensive approach that addresses general as well as specific study questions. In the example of the myeloablative versus nonmyeloablative HSCT study, when the events of relapse and TRM are combined and analyzed as a single event as progression, there is no difference between the two types of transplantation [see Fig. 4 in Alyea et al. (8)]. This is because this type of analysis is limited to answer to only one question—whether the progression-free survival of nonmyeloablative HSCT is better than myeloablative HSCT. Although this approach is informative, this analysis is not sufficient to answer whether the CIR or CIT between myeloablative and non-myeloablative transplantations are different. The CR analysis of relapse and TRM is an appropriate method to answer these more advanced questions.

In the analysis of CR data, it is important to present both the results of the event of interest and the results of competing risks. In the example of the myeloablative versus nonmyeloablative HSCT study, presentation of CIR without CIT would be misleading because even though the CIR is higher in the nonmyeloablative transplantation compared with the myeloablative transplantation, the CIT is lower. Similarly, in a study of umbilical cord blood transplantation, comparison of cumulative incidence of acute GVHD between umbilical cord blood and other types of transplantation would be erroneous if the early infection rates are not considered simultaneously.

Fitting a CR regression model is also important. Just as in the standard survival analysis, analysis of competing risks is incomplete without CR regression analysis. CR regression analysis is used to identify risk factors for each competing risk. In example 2c, donor-recipient sex mismatch was associated with a decreased risk of relapse (HR = 0.52, P = 0.04), but not for TRM. Also, the effects of bone marrow stem cells on relapse (β = −0.78, HR = 0.46) and on TRM (β = 0.81, HR = 2.24) were contrasting, as shown in the coefficient estimate, β, and HR (even though these effects are not statistically significant) and this opposite effect of a covariate would not be detected by fitting a standard Cox model. This is because the standard Cox model is not designed to answer what risk factors contribute to relapse in the presence of the competing risk of TRM. Using a cause-specific Cox regression model is incorrect because it ignores competing risks and treats them as censored. Fitting a CR regression model is also important to confirm whether the difference seen in the cumulative incidence curves is true or confounded by other risk factors. This is illustrated in example 2c. The difference in the cumulative incidence of TRM was confounded by the bone marrow stem cell use. Similarly the difference in the cumulative incidence of relapse was confounded by unfavorable prognosis at stem cell infusion.

In summary, the important first step for the analysis of CR data is the recognition that competing risks are present. Following this, the analysis should include a calculation of cumulative incidence of an event of interest in the presence of competing risks, a proper test for cumulative incidence curves of an event, and CR regression analyses. Software packages for the KM method and Cox proportional hazards regression model are available in S-plus, SAS, and SPSS. R and S programs for the Gray test (3) and the Fine and Gray CR regression model (6) can be obtained from the web page of Robert Gray[1] or by contacting him at gray.robert@jimmy. harvard.edu. Future research should include sample size and/or power calculation in CR data.

Acknowledgments

---

[1] http://biowww.dfci.harvard.edu/~gray/.

References

1. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc 1958;53: 457–81.
2. Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemother Rep 1966;50: 163–70.
3. Gray RJ. A class of *K*-sample tests for comparing the cumulative incidence of a competing risk. Ann Stat 1988;16:1140–54.
4. Cox DR. Regression models and life-tables (with discussion). J R Stat Soc B 1972;34: 187–220.
5. Cox DR, Oakes D. Analysis of survival data. London: Chapman and Hall; 1984. p. 91–110.
6. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. J Am Stat Assoc 1999;94:496–509.
7. Klein JP, Andersen PK. Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. Biometrics 2005;61: 223–9.
8. Alyea EP, Kim HT, Ho V, et al. Comparative outcome of nonmyeloablative and myeloablative allogeneic hematopoietic cell transplantation for patients older than 50 years of age. Blood 2005;105:1810–4.
9. Saavedra S, Sanz GF, Jarque I, et al. Early infections in adult patients undergoing unrelated donor cord blood transplantation. Bone Marrow Transplant 2002; 30:937–43.