STATISTICS IN MEDICINE
*Statist. Med.* **18**, 695–706 (1999)

# ESTIMATION OF FAILURE PROBABILITIES IN THE PRESENCE OF COMPETING RISKS: NEW REPRESENTATIONS OF OLD ESTIMATORS

TED A. GOOLEY[1,2,3]*, WENDY LEISENRING[1,2,3], JOHN CROWLEY[2,3] AND BARRY E. STORER[1,2,3]

[1] *Fred Hutchinson Cancer Research Center, Clinical Research Division, Seattle, WA 98109-1024, U.S.A.*
[2] *Fred Hutchinson Cancer Research Center, Public Health Sciences Division, Seattle, WA 98109-1024, U.S.A.*
[3] *Department of Biostatistics, University of Washington, U.S.A.*

## SUMMARY

A topic that has received attention in both the statistical and medical literature is the estimation of the probability of failure for endpoints that are subject to competing risks. Despite this, it is not uncommon to see the complement of the Kaplan–Meier estimate used in this setting and interpreted as the probability of failure. If one desires an estimate that can be interpreted in this way, however, the cumulative incidence estimate is the appropriate tool to use in such situations. We believe the more commonly seen representations of the Kaplan–Meier estimate and the cumulative incidence estimate do not lend themselves to easy explanation and understanding of this interpretation. We present, therefore, a representation of each estimate in a manner not ordinarily seen, each representation utilizing the concept of censored observations being 'redistributed to the right.' We feel these allow a more intuitive understanding of each estimate and therefore an appreciation of why the Kaplan–Meier method is inappropriate for estimation purposes in the presence of competing risks, while the cumulative incidence estimate is appropriate. Copyright © 1999 John Wiley & Sons, Ltd.

## 1. INTRODUCTION

The analysis of many time-to-event endpoints in various scientific disciplines, particularly in the medical sciences, is complicated by the presence of competing risks. We shall define a competing risk as an event whose occurrence either precludes the occurrence of another event under examination or fundamentally alters the probability of occurrence of this other event. As an example, when relapse of leukaemia is the outcome of interest among patients with this disease, death without relapse is a competing-risk event. Other examples include cause-specific death due to prostate cancer, where deaths from other causes (for example old age) are competing risks for death due to the index cancer.

* Correspondence to: Ted A. Gooley, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N., D5-290, P.O. Box 19024, Seattle, WA 98109-1024, U.S.A. E-mail: tgooley@fhcrc.org

Contract/grant sponsor: National Institutes of Health
Contract/grant number: CA18029, CA 38296 and HL 36444

CCC 0277–6715/99/060695–12$17.50

*Received February 1998*
*Accepted May 1998*

Methods for estimating the probability of failure for events that are subject to competing risks are not new,[1–9] and some work has been done in the area of inference in this setting.[10,11] Despite this, it is still quite common to see inappropriate methods used to estimate such probabilities for endpoints that suffer from competing risks. In particular, it is not unusual to see researchers utilize the complement of a Kaplan–Meier estimate (1 − KM) to represent the probability of occurrence of a specified endpoint, even if the appropriate data contain competing-risk events. We feel the common misuse of this methodology for estimation purposes stems from a lack of thorough understanding among clinical researchers of the assumptions required to obtain interpretable Kaplan–Meier estimates. A lack of knowledge concerning the mechanics of calculating these estimates also likely contributes to the misuse. The purpose of this article, therefore, is to review two estimators of the probability of failure commonly used (1 − KM and the cumulative incidence estimate, referred to as CI) with time-to-event endpoints that suffer from competing risks and to represent each of these mathematically in a manner not commonly seen. Based on our collaborative experiences with clinical investigators, we feel these alternative representations provide a useful description of why 1−KM is inappropriate and not interpretable when used in the presence of competing risks. We also believe these representations lead to a more intuitive form of CI than that ordinarily seen.

The term cumulative incidence has been used for various purposes. Similarly, as noted by Cheng *et al.*,[12] the cumulative incidence function has been referred to as cause-specific risk,[4] the crude incidence curve[13] and the cause-specific failure probability.[7] Our reference to this term will be consistent with that defined by Kalbfleisch and Prentice.[2]

Section 2 demonstrates the dependence of 1−KM and CI on the hazard of failure of each failure type for endpoints that are subject to competing risks. Section 3 reviews common representations of the Kaplan–Meier and cumulative incidence estimates and proceeds to present 1−KM and CI in the alternate manner alluded to above. Section 4 contains two examples, each taken from a clinical trial, and Section 5 presents conclusions and recommendations.

## 2. DEPENDENCE OF ESTIMATES ON HAZARDS OF FAILURE

Throughout this manuscript we shall assume the existence of two failure types; one corresponding to the event of interest and the other corresponding to one or more competing risks. If more than two types of competing risks exist, these will be lumped together and considered as a single group for ease of discussion.

Fundamental to the analysis of any time-to-event endpoint is the hazard of failure. Recall that the hazard is fundamentally a conditional quantity, that is, the hazard at time $t$ is conditional upon patients being at risk of failure at time $t$. Because of this, the hazard can be a useful measure when competing risks are present. However, if one is interested in estimating the probability of failure among *all* patients under study, the estimate should be consistent with the simple ratio given by the number of failures divided by the number of patients under study. If competing risks are present, then hazards for each of two types of failure exist. Therefore, the number of failures from the competing risk will influence the number of failures from the cause of interest, and consequently the estimate of the probability of failure from this cause. Failures from the competing risk reduce the number of patients at risk of failure from the cause of interest.

Recall that 1−KM is a function of the hazard of failure type 1 and does not depend on the hazard of failure type 2. The result of this is that 1−KM is not interpretable as the probability of failure type 1 when competing risks are present. CI, however, is a function of the hazard of failure

 

type 1 *and* the hazard of failure type 2. The consequence of this is that CI is interpretable as an estimate of the probability of failure type 1 when competing risks are present (or absent), and therefore is the appropriate tool to use for estimating the probability of failure. The examples below demonstrate that 1−KM fails to yield an unbiased and interpretable estimate of this probability when competing risks are present, but CI does. This is a result of the fact that CI is a function of the hazards of both failure types, while 1−KM depends solely on the hazard of failure from the cause of interest. These dependencies are well recognized and are shown mathematically in the next section.

In order to illustrate some of these concepts, we calculate below 1−KM and CI for each of three simulated data sets. The data were generated by assuming the existence of two possible types of failure. A time to each failure type was generated for each of 5000 simulated patients by assuming that time without failure for each type followed an exponential model and failure types were independent. In other words, $S_1(t) = e^{-\lambda_1(t)t}$ and $S_2(t) = e^{-\lambda_2(t)t}$, where $S_1(t)$ and $S_2(t)$ represent survival time without failure type 1 and type 2, respectively, and $\lambda_1(t)$ and $\lambda_2(t)$ represent the hazard of failure type 1 and type 2, respectively. For the three data sets, $\lambda_1(t) = 0{\cdot}25$ for each, and $\lambda_2(t) = 0$, 0·10 and 0·99, respectively. Without loss of generality and for ease of discussion, the unit of time will be taken to be years. Time to failure for each simulated patient was generated only up to 2 years so that each patient had failed from either type or survived without failure to 2 years, that is, follow-up was complete so that no observations were censored. If a patient suffered both failure types according to the above models, such a patient was assumed to have failed from the type that occurred earliest. The estimates 1−KM and CI of the probability of failure type 1 were calculated for each data set.

Since $\lambda_2(t) = 0$ for the first data set, there were no competing-risk events. There were 2001 failures of type 1 by 2 years and the remaining 2999 patients survived without failure to 2 years. Since there were no competing-risk events, 1−KM and CI are identical and interpretable as an estimate of the probability of failure type 1, that is, the estimate at 2 years is $2001/5000 = 0{\cdot}40$ (Figure 1).

For the second data set $\lambda_2(t) = 0{\cdot}10$ and competing-risk events therefore occurred. Even though $\lambda_1(t) = 0{\cdot}25$ for each of the first two data sets, fewer type 1 failures occurred in the second set because the competing-risk events reduced the number of patients at risk of failure type 1. In particular, there were 1794 type 1 failures by 2 years in this data set. The estimate of the probability of failure type 1 by this time should therefore be $1794/5000 = 0{\cdot}36$, exactly the value of CI. The estimate 1−KM, however, is 0·39 at 2 years (Figure 2). Since $\lambda_1(t) = 0{\cdot}25$ for each of the first two data sets, the values of 1−KM, which depend *only* on the hazard of failure type 1, are very similar for each data set, the difference being due only to the random variation that exists between the data sets.

For the third data set, $\lambda_2(t) = 0{\cdot}99$ and the data generated thus yielded a relatively large number of competing-risk events. This resulted in even fewer type 1 failures as compared to the first two data sets. In particular, only 913 type 1 failures occurred by 2 years. The estimate of the probability of failure type 1 by 2 years should therefore be $913/5000 = 0{\cdot}18$. Since CI depends on $\lambda_1(t)$ and $\lambda_2(t)$, CI yields this value. However, since 1−KM depends only on $\lambda_1(t)$, $1-KM = 0{\cdot}39$ at 2 years (Figure 3), equal to the values, within random error, that result from each of the first two data sets but clearly biased as an estimate of the probability of failure type 1.

The estimate 1−KM, as reported by others,[7–9] can only be interpreted as a hypothetical probability that assumes the probability of failure from the cause of interest would not change if competing risks were removed. If estimation of failure probabilities is the goal, CI is the

 

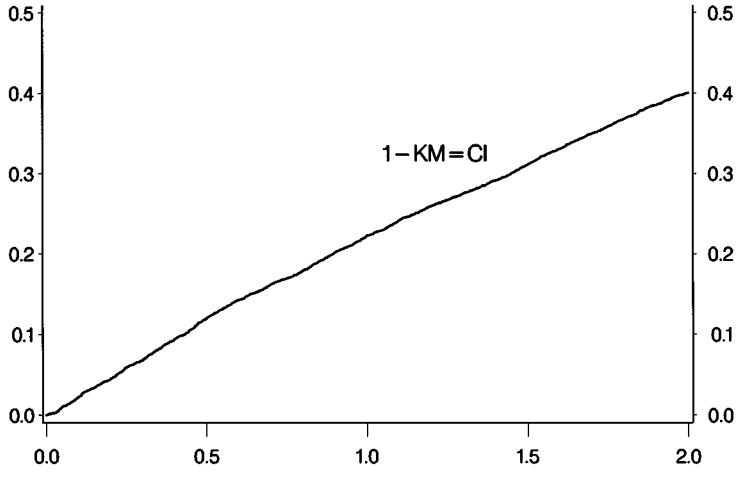


Figure 1. The complement of the KM estimate and the CI estimate of relapse for a simulated data set consisting of 5000 patients. Each observation is generated by assuming a hazard of relapse of 0·25 and a hazard of death without relapse of 0 percentage points per year.

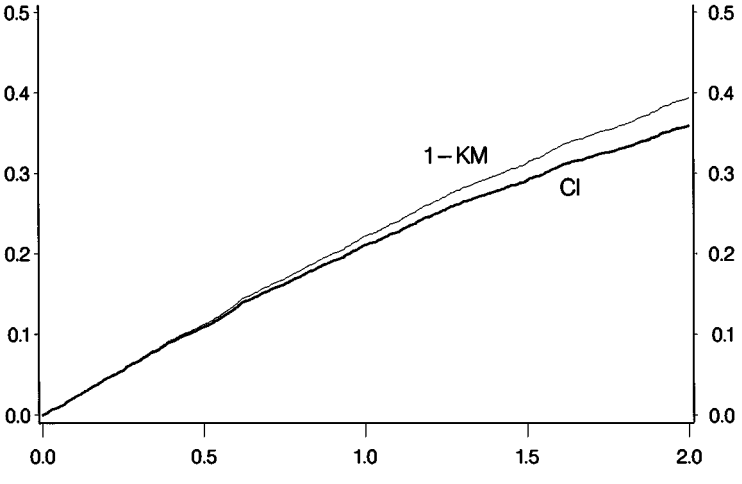


Figure 2. The complement of the KM estimate and the CI estimate of relapse for a simulated data set consisting of 5000 patients. Each observation is generated by assuming a hazard of relapse of 0·25 and a hazard of death without relapse of 0·10 percentage points per year

appropriate quantity to use. Despite this, 1 − KM continues to be used for these purposes. We feel the primary reason for this misuse is a fundamental misunderstanding among clinical researchers of the assumptions required for interpretable Kaplan–Meier estimates, coupled with a lack of thorough comprehension of how CI is computed. The censoring of observations that are failures from a competing risk is what causes 1 − KM to be non-interpretable and a biased estimate of probability of failure. Moreover, the way each method deals with such observations leads to the

 

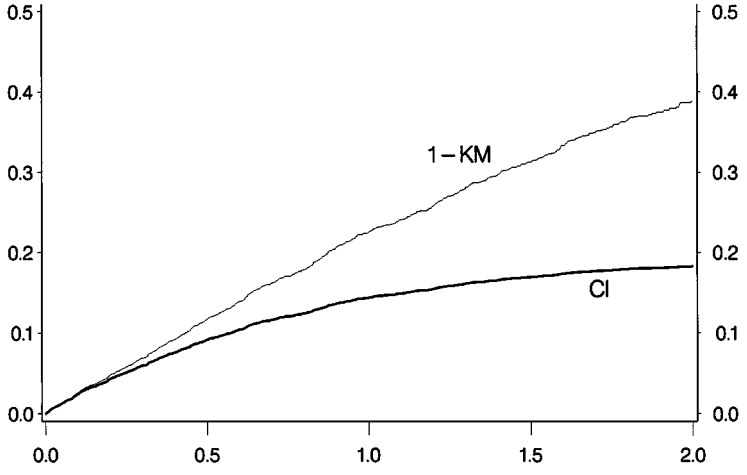


Figure 3. The complement of the KM estimate and the CI estimate of relapse for a simulated data set consisting of 5000 patients. Each observation is generated by assuming a hazard of relapse of 0·25 and a hazard of death without relapse of 0·99 percentage points per year

difference between CI and 1−KM. A thorough understanding of how censoring impacts the computation of CI and 1−KM therefore seems necessary to gain a full understanding and appreciation of the differences between these estimates. The common expressions for 1−KM and CI shown below do not, we feel, provide a clear understanding of this impact, however. In the next section, we present an alternate representation of 1−KM and CI from that commonly seen. We feel these representations provide a more intuitive feel for how censored observations are handled mathematically and therefore lead to a fuller understanding of the settings in which 1−KM and CI provide appropriate estimates of the probability of failure.

## 3. ALTERNATIVE REPRESENTATIONS OF 1−KM AND CI

While the usual formulations of 1−KM and CI, as reviewed below, are explicit functions of the estimates of the hazards, we shall use representations that are not. These representations are motivated by the work of Efron,[14] and we feel they more clearly demonstrate when and why the estimates differ. In particular, we have had success in using these formulae and the associated interpretations to clearly explain to clinical investigators the principles behind estimation of failure probabilities, both in the presence and absence of competing risks. It will be helpful to first introduce some notation.

We shall assume that $n$ patients are under study, that each patient will experience one and only one of three distinct outcomes, and that the times at which these outcomes occur can be ordered such that $t_1 \leqslant t_2 \leqslant ... \leqslant t_n$. The three potential outcomes for a patient at time $t_j$ consist of the following:

(i) patient fails from the event of interest at time $t_j$;
(ii) patient fails from a competing risk at time $t_j$;
(iii) patient has not failed from either cause but has follow-up only to time $t_j$.

 

A patient who has the third outcome is censored at time $t_j$ due to lack of follow-up beyond this time. Note that this outcome is distinct from the second, the occurrence of which removes a patient from risk of failure from the cause of interest.
Define the following:

$n \equiv$ the number of patients under study, that is, the number of patients initially at risk;
$e_j \equiv$ the number of patients who fail from the event of interest at time $t_j$;
$r_j \equiv$ the number of patients who fail from a competing risk at time $t_j$;
$c_j \equiv$ the number of patients who are censored at time $t_j$; and
$n_j \equiv$ the number of patients who are known to be at risk of failure *beyond* time $t_j$.

Then

$$n_j = n - \sum_{k=1}^{j} (e_k + r_k + c_k).$$

Using the above notation, the Kaplan–Meier estimate of survival without failure type 1 ($\mathrm{KM}_1(t)$) is commonly expressed as

$$\mathrm{KM}_1(t) = \prod_{j=1}^{s} \left(1 - \frac{e_j}{n_{j-1}}\right) \tag{1}$$

where $s$ is the largest $j$ such that $t_j < t$, $n_0 = n$ and $e_j/n_{j-1}$ is an estimate of the hazard of failure type 1 at time $t_j$, as shown in Kalbfleisch and Prentice.[2] Note that equation (1) is obtained by censoring patients who fail from the competing risk and therefore does not depend on the hazard of failure type 2. An expression similar to equation (1) for the Kaplan–Meier estimate of survival without failure type 2 ($\mathrm{KM}_2(t)$) is given by

$$\mathrm{KM}_2(t) = \prod_{j=1}^{s} \left(1 - \frac{r_j}{n_{j-1}}\right). \tag{2}$$

Note that patients who suffer from failure type 1 are censored in the derivation that leads to equation (2) so that $\mathrm{KM}_2(t)$ does not depend on the hazard of failure type 1, and $r_j/n_{j-1}$ is the estimate of the hazard of failure type 2. An overall Kaplan–Meier survivor function, $\mathrm{KM}_{12}(t)$, can be obtained by considering failures of any kind as events and represents the probability of surviving all causes of failure by time $t$. It can be shown that $\mathrm{KM}_{12}(t) = \mathrm{KM}_1(t)\mathrm{KM}_2(t)$, and CI is given by

$$\mathrm{CI}(t) = \sum_{j=1}^{s} \frac{e_j}{n_{j-1}} \mathrm{KM}_{12}(t_j) \tag{3}$$

as shown by Kalbfleisch and Prentice.[2] Given the formulation of $\mathrm{KM}_{12}(t)$, one can see the dependence of CI on the estimates of the hazard of each failure type.

Recall that 1−KM and CI are both intended to be marginal estimates of the probability of failure (where 1−KM henceforth corresponds to $1-\mathrm{KM}_1$ above). Each should therefore change if and only if a patient fails from the event of interest, and each time such a failure occurs the estimate should increase by a specified amount. In other words, the estimate at time $t$ can be

 

alternatively represented as

$$I(t) = \sum_{j=1}^{s} J(t_j) e_j \tag{4}$$

where $J(t_j)$ represents the 'jump' in the estimate at time $t_j$, the amount the estimate changes due to failures at this time, and $t_s$ is the largest $t_j \leqslant t$.

The amounts $J(t_j)$, $j = 1, \ldots, n$ will differ depending on whether one is calculating CI or $1-\mathrm{KM}$. These amounts will also change throughout time due to their dependence on the frequency and timing of censored observations and failures from the competing risk. We shall denote by $J_{\mathrm{CI}}(t_j)$ and $J_{\mathrm{KM}}(t_j)$ the values for $J(t_j)$ for CI and $1-\mathrm{KM}$, respectively. We show below how $(J_{\mathrm{CI}}(t_1), \ldots, J_{\mathrm{CI}}(t_n))$ and $(J_{\mathrm{KM}}(t_1), \ldots, J_{\mathrm{KM}}(t_n))$ are determined and how they depend on censored observations and failures from the competing risk.

### 3.1. Cumulative incidence estimator

If one assumes that all patients are equally likely to fail from the event of interest, then as long as all patients have complete follow-up, the estimate of the probability of failure from the cause of interest is increased by $1/n$ each time such a failure occurs. In other words, $J_{\mathrm{CI}}(t_j) = 1/n$ for all $t_j$ where a failure occurs.

Suppose, however, that not all patients have complete follow-up. Our alternative representation utilizes the concept of redistributing to the right as outlined in Efron.[14] Under this perspective, the potential contribution to the probability of failure from the event of interest for censored patients is evenly redistributed among all patients known to be at risk of failure beyond this time, and in our previous notation

$$\begin{aligned} J_{\mathrm{CI}}(t_j) &= J_{\mathrm{CI}}(t_{j-1}) + J_{\mathrm{CI}}(t_{j-1})\left(\frac{c_j}{n_j}\right) \\ &= J_{\mathrm{CI}}(t_{j-1})\left(1 + \frac{c_j}{n_j}\right) \quad \text{for } j = 2, \ldots, n \end{aligned} \tag{5}$$

where $J_{\mathrm{CI}}(t_1) = 1/n$.

Inclusion of equation (5) into equation (4) provides an alternative representation of $\mathrm{CI}(t)$. The contribution to the estimate of the probability of failure from the cause of interest due to failures that occur after patients are censored is increased over the contribution from previous failures. The increase is equal to the potential contribution from the censored patient(s) that is redistributed among patients known to be at risk of failure beyond the time that censoring occurred. Note that if a patient fails from a competing risk, the potential contribution to the estimate for this patient becomes zero, as failure from the event of interest is no longer possible. Hence, patients who fail from a competing risk are treated differently from patients who are censored due to lack of follow-up.

### 3.2. Complement of the Kaplan–Meier estimator

When calculating $1-\mathrm{KM}$ in the presence or absence of competing risks, patients who are censored due to lack of follow-up are handled exactly as they are in computing CI. When calculating $1-\mathrm{KM}$ in the presence of competing risks, however, patients who fail from a competing risk are also treated as censored at the time of failure. The potential contribution to the probability of failure from the event of interest for such patients is therefore redistributed among

 

the patients known to be at risk beyond this time, and the appropriate term in equation (4) can be expressed as

$$J_{\mathrm{KM}}(t_j) = J_{\mathrm{KM}}(t_{j-1}) + J_{\mathrm{KM}}(t_{j-1})\left(\frac{c_j + r_j}{n_j}\right)$$

$$= J_{\mathrm{KM}}(t_{j-1})\left(1 + \frac{c_j}{n_j} + \frac{r_j}{n_j}\right) \quad \text{for } j = 2, \ldots, n \tag{6}$$

where $J_{\mathrm{KM}}(t_1) = 1/n$.

Censoring patients who fail from a competing risk is not an appropriate action to take when estimation of the probability of failure is the goal as this implicitly assumes that failure from the cause of interest is still possible beyond the time at which the censoring occurred. If a patient fails from a competing risk, failure from the cause of interest is no longer possible and the potential contribution to the estimate from this patient should become zero. The method of calculation used to obtain 1 − KM therefore has the effect of inflating the estimate over what it should be so that 1 − KM does not yield an unbiased estimate of the probability of failure from the cause of interest.

As can be seen, equations (5) and (6) differ only in the way that the terms $J(t_j)$ of equation (4) are defined, the difference being due to the way failures from a competing risk are handled. The terms $J_{\mathrm{CI}}(t_j)$ and $J_{\mathrm{KM}}(t_j)$ are identical up to the point at which the first competing-risk event occurs. At the time of the first failure from the event of interest that follows a competing-risk event, however, $J_{\mathrm{KM}}(t_j)$ is larger than $J_{\mathrm{CI}}(t_j)$ by the amount $J(t_j)(r_j/n_j)$, where $J(t_j) = J_{\mathrm{KM}}(t_j) = J_{\mathrm{CI}}(t_j)$. Owing to the iterative nature of equations (5) and (6), the difference between 1 − KM and CI grows larger as more competing-risk events precede occurrences of the event of interest. Thus 1 − KM $\geqslant$ CI, with equality holding only up to the time of the first failure from the cause of interest that follows the first competing-risk event. If no failures from a competing risk are encountered, $r_j = 0$ in equation (6) for all $j$, and CI and 1 − KM are identical.

## 4. EXAMPLES FROM REAL DATA

### 4.1. Example 1

The first example is taken from a phase III Southwest Oncology Group (SWOG) study comparing conventional treatment (surgery and post-operative radiotherapy) with an experimental treatment (induction chemotherapy followed by surgery and post-operative radiotherapy) for patients with advanced stage resectable squamous cell carcinoma of the head and neck.[15] The objectives of this trial were to compare response rates, treatment failure, survival and pattern of treatment failure between the two treatment arms. We shall use data from this clinical trial to estimate the probability of disease progression by calculating 1 − KM and CI. For these purposes, we shall use only patients who were treated on the experimental arm.

Among 175 patients entered into the study, 17 were ruled ineligible. Of the 158 eligible patients, 76 were randomized to receive the conventional treatment and 82 were randomized to receive the experimental treatment. The chemotherapy in the experimental arm consisted of a combination of cisplatin, vincristine, methotrexate and bleomycin. Of the 82 eligible patients randomized to receive the experimental treatment, 38 had disease progression and 34 died without disease progression. Therefore 38 of 82 patients failed from the event of interest (disease progression) while 34 of 82 patients failed from the competing risk of death without progression. Ten of 82

 

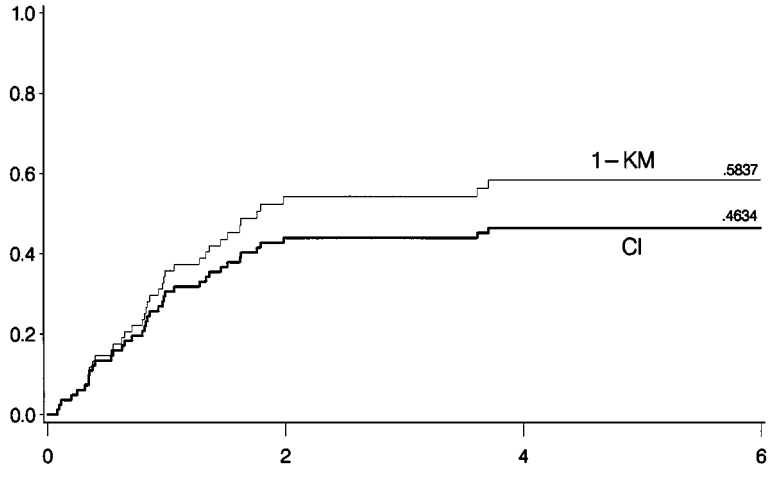


Figure 4. The complement of the KM estimate and the CI estimate of disease progression among 82 patients with head and neck cancer randomized to receive an experimental treatment

patients were alive without progression at last follow-up and are therefore censored. The shortest follow-up among these 10 patients is just under 9 years. All cases of progression occurred prior to this earliest censored observation so that all patients have complete follow-up through this time. The only natural estimate of the probability of progression by this time is therefore $38/82 = 0{\cdot}46$, which is precisely the value of CI. On the other hand, the value of $1-\mathrm{KM}$ at this time is 0·58 (Figure 4), the difference being due to the patients who failed from the competing risk of death without progression.

### 4.2. Example 2

The second example comes from the setting of bone marrow transplantation. A potential complication among patients who receive a bone marrow transplant is known as chronic graft-versus-host disease (CGVHD), and the probability of this complication is often an outcome of interest in clinical trials. We shall take the occurrence of CGVHD to be the failure of interest for this example. Competing risks for the occurrence of CGVHD are death without CGVHD and relapse without CGVHD. Relapse is considered to be a competing risk because patients who experience a relapse of their disease are often withdrawn from immunosuppressive therapy, where this therapy is given primarily as prophylaxis for development of CGVHD. Relapse therefore fundamentally alters the probability of occurrence of CGVHD and for this reason is regarded as a competing risk. By definition, CGVHD cannot occur until day 80 post-transplant. The data used for this example come from a phase II clinical trial whose primary endpoint was the occurrence of acute GvHD (AGVHD, a related complication that occurs prior to day 80 post-transplant).[16] A secondary endpoint of this trial was the occurrence of CGVHD, and we shall calculate CI and $1-\mathrm{KM}$ from these data for this endpoint.

Forty-three patients were enrolled on this phase II trial and given a combination of FK506 and a short course of methotrexate (MTX) for the prevention of AGVHD. Eighteen of the 43 relapsed or died before day 100 and therefore failed from a competing risk for CGVHD. Twelve of the

 

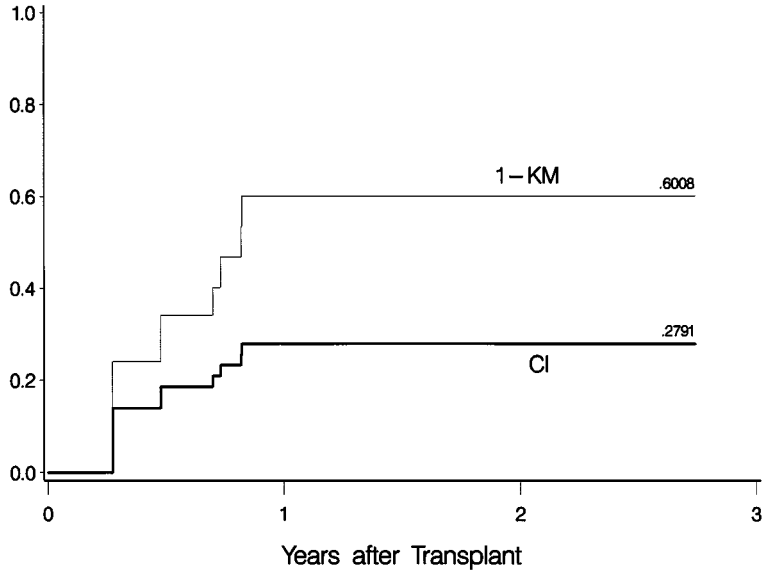


Figure 5. The complement of the KM estimate and the CI estimate of chronic GVHD among 43 patients receiving FK506 for GVHD prophylaxis

remaining 25 developed CGVHD, and all patients had complete follow-up to day 638. All cases of CGVHD occurred prior to day 638, so the estimate of the probability of CGVHD by day 638 should be $12/43 = 0{\cdot}28$, the value of CI as shown in Figure 5. The value of $1-\mathrm{KM}$ at this time, however, is 0·60, a biased estimate of the desired probability.

Another quantity that may be of interest to estimate in this setting involves a type of conditional estimate. Suppose that one restricts the population of patients to be those who survived without relapse to day 100, that is, those who are capable of developing CGVHD beyond this time. Conditioning on this event, it is interesting to note that $\mathrm{CI} = 0{\cdot}48$ (12/25), yet $1-\mathrm{KM}$ remains unchanged, a result that is not intuitive. In general, if $k$ failures from a competing risk occur by time $t_0$ among $n$ patients initially at risk and no failures from the event of interest occur by this time, $1-\mathrm{KM}$ among the population of $n$ patients will be identical to $1-\mathrm{KM}$ calculated among the population of $n-k$ patients. This can be seen from the following. Suppose $n$ subjects are at risk of failure at time $t=0$. Suppose further that $k$ patients fail from a competing risk before time $t_0$ so that $n-k$ patients are at risk of failure from the event of interest beyond $t_0$. If one calculates $1-\mathrm{KM}$ among all $n$ patients, the potential contribution of $1/n$ to the estimate of failure for each of the $k$ patients failing from a competing risk before time $t_0$ is redistributed among the $n-k$ patients who remain at risk beyond this time, that is, the contribution becomes

$$\frac{1}{n}+\frac{1}{n}\left(\frac{k}{n-k}\right)=\frac{1}{n-k}.$$

However, this is precisely the potential contribution to the estimate of failure of each patient in the reduced population of $n-k$ patients. Therefore, $1-\mathrm{KM}$ is the same for each population, a result that is non-intuitive and further demonstration that $1-\mathrm{KM}$ is not interpretable as an estimate of the probability of failure in the presence of competing risks.

 

## 5. DISCUSSION

Previous work has shown that 1−KM and CI, each of which are commonly used to estimate the probability of failure, can result in different estimates when competing risks are present.[7,9,17–19] Despite this, 1−KM continues to be misused for these purposes.[20] We have therefore proposed an alternative representation of each estimate that, in our experience, is more intuitive to clinicians and leads to a better understanding of the appropriateness of CI and 1−KM in addition to the difference between these estimates. The wide availability of statistical software packages that are capable of calculating KM estimates but which do not calculate CI estimates also undoubtedly contributes to the frequent misuse of 1−KM. Although we have not seen the CI estimate offered commercially in any software packages, calculation of CI is not computationally difficult as shown above and programs that accomplish this are straightforward to write.

Determining whether to use 1−KM or CI is unambiguous and CI should always be used if an estimate of the probability of failure of a specific type is desired. (This statement can be made without loss of generality since the two estimates are identical if competing risks are not present.) When analysing and presenting data where competing risks occur, however, it is important to describe the probability of failure from not only the cause of interest but also from failures due to competing risks. The focus of this article was to demonstrate that the methods discussed above lead to different estimates in the presence of competing risks and to provide an alternative representation of CI and 1−KM from that commonly seen. None the less, we feel it crucial to emphasize that presenting only results that describe the probability of failure from the cause of interest falls short of what may need to be examined in order to fully appreciate factors that affect the outcome. One approach to dealing with this problem is to present an estimate of the time to the minimum of the different types of failure. A discussion of related topics is contained elsewhere.[7–8]

We have focused purely on the estimation of the probability of failure in this paper, but the problem of comparing, say, treatment groups with respect to outcome will also be of interest when two or more groups are being analysed. How such comparisons are made and exactly what is compared can be complicated issues and have not been addressed here. Such topics are discussed in previous work.[10–12] We note only that if one is interested in evaluating the effect of a covariate, such as treatment, on the hazard of failure from the cause of interest, the use of CI estimates and tests related to them may be misleading if the treatment also affects the hazard of the competing risk. In such situations, the logrank test is appropriate for inference since it is a function solely of the hazard of failure from the cause of interest and failures from the competing risk can therefore be censored. If one is instead interested in comparing the actual probability of failure between two groups, tests that utilize CI estimates, that is, those that depend on the hazards of each type of failure, need to be used.[11–12] In either case, one must be careful to understand the relationship of treatment to the various causes of failure in order to draw appropriate conclusions.

## REFERENCES

1. Aalen, O. 'Nonparametric estimation of partial transition probabilities in multiple decrement models', *Annals of Statistics*, **6**, 534–545 (1978).
2. Kalbfleisch, J. D. and Prentice, R. L. *The Statistical Analysis of Failure Time Data*, John Wiley, New York, 1980.

 

3. Prentice, R. L. and Kalbfleisch, J. D. The analysis of failure times in the presence of competing risks, *Biometrics*, **34**, 541–554 (1978).
4. Benichou, J. and Gail, M. H. 'Estimates of absolute cause-specific risk in cohort studies', *Biometrics*, **46**, 813–826 (1990).
5. Dinse, G. E. and Larson, M. G. 'A note on semi-Markov models for partially censored data', *Biometrika* **73**, 379–386 (1986).
6. Lawless, J. F. *Statistical Models and Methods for Lifetime Data*, Wiley, New York, 1982.
7. Gaynor, J. J., Feuer, E. J., Tan, C. C., Wu, D. H, Little, C. R., Straus, D. J., Clarkson, B. D. and Brennan, M. F. 'On the use of cause-specific failure and conditional failure probabilities: Examples from clinical oncology data', *Journal of the American Statistical Association*, **88**, 400–409 (1993).
8. Pepe, M. S. and Mori, M. 'Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data?', *Statistics in Medicine*, **12**, 737–751 (1993).
9. Pepe, M. S., Longton, G., Pettinger, M., Mori, M., Fisher, L. D. and Storb, R. 'Summarizing data on survival, relapse, and chronic graft-versus-host disease after bone marrow transplantation: Motivation for and description of new methods', *British Journal of Haematology*, **83**, 602–607 (1993).
10. Pepe, M. S. 'Inference for events with dependent risks in multiple endpoint studies', *Journal of the American Statistical Association*, **86**, 770–778 (1991).
11. Gray, R. J. 'A class of k-sample tests for comparing the cumulative incidence of a competing risk', *Annals of Statistics*, **16**, 1141–1154 (1988).
12. Cheng, S. C., Fine, J. P. and Wei, L. J. 'Prediction of cumulative incidence function under the proportional hazards model', *Biometrics*, **54**, 219–228 (1998).
13. Korn, E. L. and Dorey, F. J. 'Application of crude incidence curves, *Statistics in Medicine*, **11**, 813–829 (1992).
14. Efron, B. 'The two sample problem with censored data', in *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics*, *IV*, Prentice-Hall, New York, 1967, pp. 831–853.
15. Schuller, D. E., Metch, B., Stein, D. W., Mattox, D. and McCracken, J. D. 'Preoperative chemotherapy in advanced resectable head and neck cancer: Final report of the Southwest Oncology Group', *Laryngoscope*, **98**, 1205–1211 (1988).
16. Nash, R. A., Pineiro, L. A., Storb, R., Deeg, H. J., Fitzsimmons, W. E. Furlong, T, Hansen, J. A., Gooley, T., Maher, R. M., Martin, P., McSweeney, P. A., Sullivan, K. M., Anasetti, C. and Fay, J. W. 'FK506 in combination with methotrexate for the prevention of graft-versus-host disease after marrow transplantation from matched unrelated donors', *Blood*, **88**, 3634–3641 (1996).
17. McGiffin, D. C., Naftel, D. C., Kirklin, J. K., Morrow, W. R., Towbin, J, Shaddy, R, Alejos, J., Rossi, A. and Pediatric Heart Transplant Study Group. 'Predicting outcome after listing for heart transplantation in children: Comparison of Kaplan-Meier and parametric competing risk analysis', *Pediatrics*, **16,** 713–722 (1997).
18. Caplan, R. J., Pajak, T. F. and Cox, J. D. 'Analysis of the probability and risk of cause-specific failure', *International Journal of Radiation Oncology Biology and Physics*, **29**, 1183–1186 (1994).
19. Gelman, R., Gelber, R., Henderson, C., Coleman, C. N. and Harris, J. R. 'Improved methodology for analyzing local and distant recurrence', *Journal of Clinical Oncology*, **8**, 548–555 (1990).
20. Niland, J. C., Gebhardt, J. A., Lee, J. and Forman, S. J. 'Study design, statistical analyses, and results reporting in the bone marrow transplantation literature', *Biology of Blood and Marrow Transplantation*, **1**, 47–53 (1995).

 