ELSEVIER

Journal of Clinical Epidemiology 59 (2006) 1110–1114

Journal of Clinical Epidemiology

Kaplan–Meier methods yielded misleading results in competing risk scenarios

Danielle A. Southern[a,b], Peter D. Faris[b], Rollin Brant[a], P. Diane Galbraith[b,c], Colleen M. Norris[d], Merril L. Knudtson[d], William A. Ghali[a,b,d,*], for the APPROACH Investigators

[a]*Department of Community Health Sciences, University of Calgary, Calgary, Alberta, Canada*
[b]*Centre for Health and Policy Studies, University of Calgary, Calgary, Alberta, Canada*
[c]*Calgary Health Region, Calgary, Alberta, Canada*
[d]*Department of Medicine, University of Calgary, Calgary, Alberta, Canada*

Accepted 11 July 2006

Abstract

Background and Objective: Time-to-event curves are routinely presented in the medical literature. The most widely used method is the Kaplan–Meier (K-M) method, but this analysis approach may not be appropriate when an analysis focuses on time-to-first event in scenarios where there are competing events. We compared K-M methods applying various censoring approaches with the lesser-known ''cumulative incidence competing risks'' (CICR) method in an analysis of competing events.

Methods: A registry containing data on 21,624 patients undergoing cardiac catheterization was analyzed. Time to coronary artery bypass grafting (CABG) was assessed in an analysis for which percutaneous coronary intervention and death were competing events. Time-to-CABG curves were calculated using the ''K-M censor all method,'' ''K-M censor death only method,'' ''K-M ignore all method,'' and the CICR method.

Results: One-year CABG rates calculated for the K-M ''censor all,'' ''censor death only,'' and ''ignore all'' methods were 28.8%, 22.8%, and 22.4%, respectively compared to the ''actual'' rate of 20.8%. For the CICR method, the corresponding 1-year rate was identical to the ''actual'' rate.

Conclusion: In situations with competing risks, and where an analysis focuses on first events, the CICR method is most appropriate, as K-M methods will tend to overestimate event rates.

Keywords: Competing risks; Survival; Time to event; Kaplan–Meier; K-M plots; Cumulative incidence

1. Introduction

Survival and time-to-event curves are widely used in the literature to present the probabilities of events over time. Survival curves present the probabilities of remaining event free and begin at 100% and then decrease over time. Time-to-event curves present the probabilities of having an event and start at 0% and then increase over time.

The most widely used method to generate time-to-event and survival curves is the Kaplan–Meier (K-M) method [1,2]. However, this analysis tool is generally meant to describe time to a single type of event.

When competing events are present, an alternative approach may be required. Consider for example a prospective study of patients undergoing cardiac catheterization to examine mortality, coronary artery bypass grafting (CABG), or percutaneous coronary intervention (PCI) as competing events after catheterization. If one wishes to focus on occurrence of CABG with PCI and death as competing events, the K-M method may not be appropriate because deaths preclude subsequent CABG, and also because death and need for CABG cannot be considered as statistically independent unless all deaths arise from completely unrelated conditions (e.g., passenger death in vehicular accidents). Similarly, if one wishes to simultaneously examine PCI, the fact that the occurrence of CABG makes subsequent PCI less likely leads to potential problems in applying the K-M method.

To avoid the difficulties discussed above, some authors have suggested using a cause-specific cumulative incidence

* Corresponding author. HSG239, 3330 Hospital Dr. NW, Calgary, Alberta, Canada T2N 4N1. Tel.: 403-210-9317; fax: 403-210-3818.
E-mail address: wghali@ucalgary.ca (W.A. Ghali).

0895-4356/06/$ – see front matter
doi: 10.1016/j.jclinepi.2006.07.002

function [1,3–9]. This method takes into account other events and does not make any assumptions about their independence. However, despite the existence of such a ''cumulative incidence competing risks'' (CICR) method [1], it is not used enough in the medical literature. A multitude of papers present data on competing risks, and either use a clearly suboptimal analysis method, or provide little detail on how the competing risks were analyzed. Two illustrative examples are described in detail in the discussion.

In this paper, we present conceptual descriptions of three separate approaches to using the K-M method in the context of competing risks and a different analysis method called the CICR method for determining survival curves and time-to-event curves. We then apply these four methods to data on competing events after cardiac catheterization. Access to computer code is also provided that will enable readers to replicate the analyses presented in this paper for use on other data [10]. Our global objective is to increase general awareness of these methods and to promote the transfer of the CICR method from relatively remote statistical references to more widespread use by medical researchers.

2. Methods

2.1. Data source

Our study used data from the Alberta Provincial Project for Outcome Assessment in Coronary Heart Disease (APPROACH), an inception cohort that includes all patients in Alberta, Canada undergoing cardiac catherization. Patients in the database are followed up longitudinally for assessment of the outcomes of death, CABG, PCI, quality of life, and health-care costs. Patient survival and time from enrollment catherization until death were ascertained through semiannual linkage to Alberta Vital Statistics records. For this study, we analyzed data for patients enrolled in APPROACH during the calendar years 1995 through 1998, with complete follow-up of patients to December 31, 1999. We used these data to analyze and to plot time-to-event curves for occurrences of CABG surgery as a first event within 1 year of catheterization using four different statistical methods. In this analysis scenario, PCI, CABG, and death represent competing events.

2.2. Censoring

Three of the four methods use K-M plots with varying degrees of ''censoring'' of events. A case is censored in survival analysis when information is no longer collected on that patient due to loss to follow-up. With censoring, observations are removed from the ''at-risk'' set and it is then assumed that the individual removed would have had the same risk of subsequent occurrence of the event of interest as do those who were not censored [5]. If the probability of loss to follow-up is independent of the event of interest, the resulting K-M curve will provide an unbiased estimate of the true survival curve for a population not subject to censoring [5]. This assumption is clearly violated when one uses censoring to account for a nonindependent competing event. Indeed, if the competing event of death leads to censoring, and the objective of the analysis is to describe the experience of the actual cohort, a censored individual would be assumed to still be at risk for an event after death, clearly a nonsensical situation [5].

The four methods used in this study to analyze time-to-CABG surgery curves are described below.

2.2.1. K-M censor all method

The K-M censor all approach is a widely used method for competing risk situations. Following this approach, the occurrence of a competing event in advance of the event of interest results in censoring. For a time-to-CABG analysis extending to 1 year after catheterization, individuals are censored if they undergo PCI, die, or are lost to follow-up before CABG. In the present example, there was no censoring due to loss to follow-up because all patients were followed up for at least 1 year after catheterization.

2.2.2. K-M censor death only method

The K-M censor death only approach is a modification of the K-M method for competing risk situations. In the present example of time-to-CABG analyses, individuals are censored only if they experience death before CABG or are lost to follow-up. Therefore, the competing events are CABG and death. PCI is entirely ignored in this analysis.

Both the K-M censor all and censor death only methods share a common deficiency in that the probability of death for a patient depends on whether or not the person receives CABG, a situation that leads to nonindependence of these endpoints in the population under study. This violates the assumption for the censoring mechanism because the probability of being censored cannot depend on whether or not the event occurs. Also, the practical relevance of considering a hypothetical population where death has been eliminated is open to question. Lastly, the K-M curve provides a nonsensical projection for the actual rate because this approach assumes that censored individuals (including those who have died) are still at risk for subsequent CABG. Thus, neither approach provides an accurate indication of typical time to CABG in the postcatheterization population.

2.2.3. K-M ignore all method

The K-M ignore all approach is a more extreme version of the censor death only method where only those truly lost to follow-up are considered to be censored, and all other events, including death, are ignored. Of concern, patients who die are regarded as still being at risk for CABG for the full 1-year follow-up period. It is clear that this will tend to produce upwardly biased rate estimates because

subjects who die contribute to the denominator of the time-to-event curve but cannot contribute to the numerator.

2.2.4. CICR method

The CICR approach, described in detail by Tai et al. [1], modifies the event-free probability estimate to avoid analysis assumptions relating to event rates that would have arisen had censored patients not died or undergone PCI. The method addresses the issue of the risk of CABG rather than the risk of CABG if no patient died or had PCI, as does the K-M method. If a number, c, of event types are to be considered, the CICR approach provides a method for decomposing the overall time-to-event curve, $I(t)$ into $I_1(t) + I_2(t) + \ldots I_c(t)$, where for a particular event type, j, $I_j(t)$ is an estimate for cumulative incidence for event j as a first event, allowing for competing events. By following this approach, competing events are taken into account with no assumptions of independence between events. Further explanation of the calculation of $I_j(t)$ can be found in the article by Tai et al. [1].

Application of the K-M method relies on the strong assumption of independence between event and censoring, whereas the CICR method takes into account the presence of these other events, regardless of independence. Furthermore, unlike the K-M method, the cumulative incidence method provides a breakdown of the expected distribution of patients into the possible endpoints, or states, at each point in time, such that the sum of individual event rates (including the ''no event rate'') will always be 100%. This contrasts with K-M methods, where this sum will exceed 100%.

We have made SAS Version 8.1 (Cary, NC) [11] code available on the Internet [10]. The programs provided include code for the CICR method adapted from that provided by Tai et al. [1], and code for the remaining three methods, described above. The programs are easy to use and can be applied to either the sample dataset provided on the Internet or to other databases.

3. Results

The study data included 21,624 patients who underwent cardiac catheterization between 1995 and 1998. Time to CABG was calculated and plotted using the four different methods described above. The percentage of patients having CABG as a first event at 1 year after catheterization was 20.8 (Table 1). Throughout the paper, we refer to this as the ''actual'' percentage undergoing CABG within 1 year and can accurately do so because there were no losses to follow-up in the 1-year period of interest. The corresponding percentages of patients undergoing CABG, PCI, and death at various time points in the first year after cardiac catheterization are shown in Table 1.

Table 1
Number (%) of patients experiencing an event at various time points in the first year

Time (days)	100	200	300	365
No event	12,026 (55.6)	10,749 (49.7)	10,167 (47.0)	9,958 (46.1)
CABG	3,132 (14.5)	3,986 (18.4)	4,401 (20.4)	4,502 (20.8)
PCI	6,122 (28.3)	6,421 (29.7)	6,515 (30.1)	6,548 (30.3)
Death	344 (1.6)	468 (2.2)	541 (2.5)	611 (2.8)

Fig. 1 presents the results of time-to-event analyses generated by each of the four analysis methods assessed. The K-M censor all approach will lead some to conclude that 28.8% of patients had CABG at 1 year, a considerable overestimate of the actual percentage (20.8%) of patients who received CABG. The K-M censor death only approach yields a percentage of 22.8 undergoing CABG surgery. This proportion is also higher than the true proportion of patients receiving CABG at 1 year. However, because patients are only censored for death, and because only 2.8% of patients died within the first year (Table 1), the curve is overestimated by a smaller margin. The K-M ignore all approach yields a percentage of 22.4, which is again high relative to the actual percentage having CABG as a first event. The CICR method produces a percentage of 20.8 at 1 year (Fig. 1), which accurately represents the true percentage of patients undergoing CABG as a first event within 1 year and also at the earlier time points, as reported in Table 1.

4. Discussion

Our paper continues a theme of recent methodological articles relating to survival methods in the Journal of Clinical Epidemiology [12–14]. The first important distinction to be made between the CICR approach and K-M based methods is that the various approaches have different estimation targets. The CICR approach can be thought of as providing either a projection for the actual rate in the cohort of study subjects or an estimate for the actual rate in the larger population from which the subjects are sampled. By contrast, K-M approaches are aimed at providing estimates relative to a population not subject to censoring (however defined). CICR estimates will always be less than or equal to K-M type estimates, and CICR estimates do not depend on independence assumptions and hence are more broadly applicable to a variety of survival scenarios than are K-M estimates.

In the CABG example, the exact equality of the CICR estimate and the actual portion of patients undergoing CABG at 1 year is due to the complete follow-up in the subject cohort. The much higher value for the K-M censor all method reflects the anticipated time-to-CABG experience of a hypothetical population where prior death and PCI have been eliminated as competing events. The notion of eliminating death is of course inherently artificial, and it is not likely that its occurrence could be construed as being independent of CABG. In our example, then, the K-M censor all method does not provide meaningful estimates.

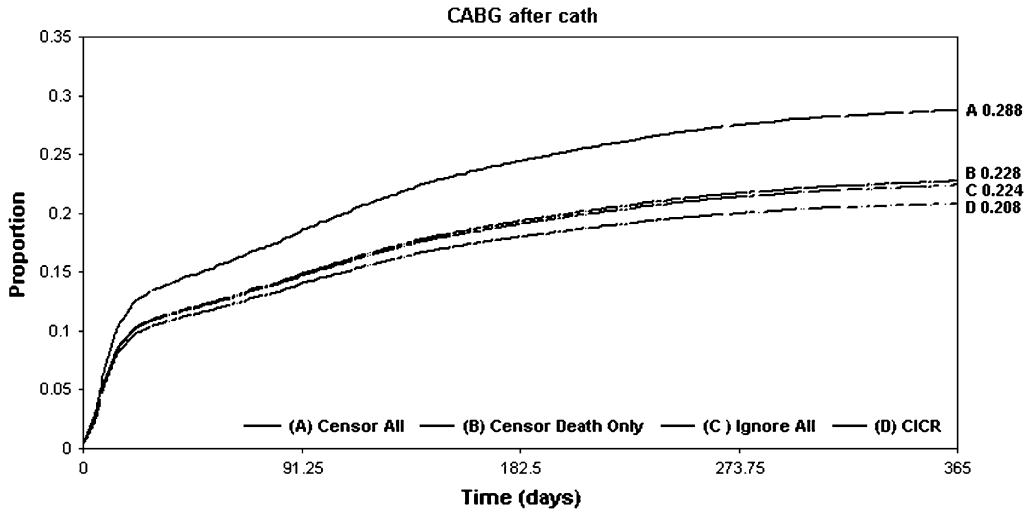

Fig. 1. Cumulative incidence estimates of first CABG for the K-M censor all, the K-M censor death only, the K-M ignore all, and the CICR methods.

The K-M censor death method provides a lower estimate than the censor all method. This arises because PCI is a negative predictor for subsequent CABG, because the occurrence of PCI tends to at least delay, if not prevent, subsequent CABG. This underlines the fact that PCI is not independent of CABG, and thus is inappropriate for handling as a censoring factor in the K-M method. Furthermore, the difficulties arising from treating deaths as censored observations in the K-M censor death only method are the same as for the censor all method.

Lastly, the ignore all method is subject to an upward bias relative to the CICR method in our example because it falsely identifies CABG occurring post-PCI as a first event, and also because of a countervailing negative bias arising from failure to account for deaths. In circumstances where there is appreciable mortality in the recruited cohort, the role of death as a competing event cannot simply be ignored. The viability of ignoring other events will depend on the real underlying aims of the investigators. In our example, a plausible case could be made for simply ignoring PCI if the only interest of the investigators is CABG utilization regardless of its timing.

In general, then, K-M methods are problematic in circumstances with competing risks. This is particularly true when the focus of analysis is on first events and when censoring is applied to competing events that occur with a high frequency (i.e., occurrence of PCI in a time-to-CABG analysis). The CICR method has conceptual advantages to the K-M methods applied here and provides more meaningful results in this example. Although we simply present the CABG analysis here, it is important to remember that competing event scenarios often indicate a need to report CICR curves for each and every competing event to present a complete picture of patient outcomes.

Although the CICR method described is not novel in the biomedical literature, its use in situations of competing risk does not appear to be widespread. We suspect that this relates, at least partially, to the fact that the existing descriptions of this methodology are somewhat ‘‘hidden’’ in methodological journals, and presented in a highly technical format that is foreign to many (or even most) medical readers. Our objective with this paper is to clearly demonstrate this important statistical issue in the pages of a high profile epidemiology journal that is widely read and circulated, in hopes that this demonstration will encourage the more widespread use of the CICR method when the K-M method is suboptimal. We also notably provide Web site access to statistical code that will help researchers implement this method more readily in future work.

The remarkable differences in the curves presented here are of great importance to researchers, journal editors, and also notably general readers of the medical literature, as many published papers in high profile medical journals either do not explicitly specify the type of censoring or simply overlook the problems associated with censoring in competing risk situations. Studies often use the K-M method, but most do not explain the method of censoring applied or use a ‘‘censor all’’ approach that can lead to a bias in the survival or time-to-event curves. In the data of the present analysis, there is a negative correlation between CABG and PCI leading to an upward bias of the time-to-event curves.

To further illustrate how misleading curves have arisen in the literature, we provide a more detailed description of two representative published examples, selected from a collection of *many* other published articles where this problem arises (that we have found without extensive searching). One study evaluated the risk of progressing to atrial fibrillation in 17,413 patients with atrial flutter [15]. A K-M plot of atrial fibrillation-free survival provides the impression that by 8 years, approximately 55% of patients progressed from atrial flutter to atrial fibrillation. However,

information presented elsewhere in the paper indicates that, in fact, only 6,599 (38%) of the atrial flutter patients progressed to atrial fibrillation at 8 years. It is apparent from the description of the related modeling that censoring on deaths is the likely cause of this discrepancy. Similarly, a study on the outcome of bypass surgery vs. PCI in diabetic patients [16] provides a K-M survival plot for freedom from repeat PCI. There is no clear description of how death as a competing risk was handled in the analysis, but the fact that the presented proportion of those remaining free from additional repeat PCI exceeds the overall survival indicates the need for caution in interpretation.

This general warning regarding K-M methods is qualified by the recognition that traditional K-M methods with censoring will not be misleading if the competing risks situation is averted by using composite endpoints such as "CABG or PCI or death." Also, competing risks are not an issue when the event that was studied is a major and final event such as death. In closing, we also qualify that our paper has focused on the general application of the CICR and K-M methods, and not on the estimation of variance and confidence intervals for these methods; for more on this, we refer readers to other statistical references [17–20].

Given the frequent occurrence of competing risk situations, researchers should either provide explicit descriptions of their censoring approach in K-M analyses or should consider using the CICR method of analysis. Journal editors and manuscript reviewers should be aware of the potential problems associated with K-M curves in such situations, and should encourage researchers to use optimal approaches to calculating curves. Most importantly, readers of the medical literature should be aware of this potential problem and should look for information on censoring methods. If such information is not available, readers will need to search for true rate data that are often presented elsewhere in the article to guide their interpretation of survival and time-to-event curves.

Acknowledgments

APPROACH is supported by donations from a variety of sponsors. Dr. Ghali is supported by a Health Scholar Award from the Alberta Heritage Foundation for Medical Research and by a Government of Canada Research Chair in Health Services Research.

References

[1] Tai BC, Machin D, White I, Gebski V. Competing risks analysis of patients with osteosarcoma: a comparison of four different approaches. Stat Med 2001;20(5):661–84.
[2] Pepe MS, Mori M. Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data? Stat Med 1993;12(8):737–51.
[3] Kalbfleisch D, Prentice RL. The statistical analysis of failure time data. New York: John Wiley & Sons, Inc.; 1980.
[4] Prentice RL, Kalbfleisch D, Peterson AV, Flournoy N, Farewell VT, Breslow NE. The analysis of failure times in the presence of competing risks. Biometrics 1978;34(4):541–54.
[5] Grunkemeier GL, Anderson RP, Miller DC, Starr A. Time-related analysis of nonfatal heart valve complications: cumulative incidence (actual) versus Kaplan-Meier (actuarial). Circulation 1997;96(9 Suppl). II-70–4.
[6] Grunkemeier GL, Anderson RP, Starr A. Actuarial and actual analysis of surgical results: empirical validation. Ann Thorac Surg 2001;71(6):1885–7.
[7] Grunkemeier GL, Wu Y. Interpretation of nonfatal events after cardiac surgery: actual versus actuarial reporting. J Thorac Cardiovasc Surg 2001;122:216–9.
[8] Miller CC III, Safi HJ, Winnerkvist A, Baldwin JC. Actual versus actuarial analysis for cardiac valve complications: the problem of competing risks. Curr Opin Cardiol 1999;14(2):79–83.
[9] Gooley TA, Leisenring W, Crowley J, Storer BE. Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. Stat Med 1999;18:695–706.
[10] Competing risks analysis programs. Accessed February 1, 2004. Available at http://www.ucalgary.ca/~dasouthe/CRcode.html.
[11] Statistical analysis system Version 8.1. Cary, NC: SAS Institute Inc.; 2000.
[12] Alberti C, Métivier F, Landais P, Thervet E, Legendre C, Chevret S. Improving estimates of event incidence over time in populations exposed to other events: applications to three large databases. J Clin Epidemiol 2003;56:536–45.
[13] Yan Y, Hoover DR, Moore RD, Xiong C. Multivariate estimation of cumulative incidence, prevalence, and morbidity time of a disease when death is likely. J Clin Epidemiol 2003;56:546–52.
[14] Yan Y, Moore RD, Hoover DR. Competing risk adjustment reduces overestimation of opportunistic infection rates in AIDS. J Clin Epidemiol 2000;53:817–22.
[15] Biblo LA, Yuan Z, Quan KJ, Mackall JA, Rimm AA. Risk of stroke in patients with atrial flutter. Am J Cardiol 2001;87:346–9. A9.
[16] Weintraub WS, Stein B, Kosinski A, Douglas JS Jr, Ghazzal ZM, Jones EL, et al. Outcome of coronary bypass surgery versus coronary angioplasty in diabetic patients with multivessel coronary artery disease. J Am Coll Cardiol 1998;31(1):10–9.
[17] Pepe MS. Inference for events with dependent risks in multiple endpoint studies. J Am Stat Assoc 1991;86:770–8.
[18] Gaynor JJ, Feuer EJ. On the use of cause-specific failure and conditional failure probabilities: examples from clinical oncology data. J Am Stat Assoc 1993;88:400–9.
[19] Lin D. Non-parametric inference for cumulative incidence functions in competing risks studies. Stat Med 1997;16:901–10.
[20] Betensky RA, Schoenfeld DA. Nonparametric estimation in a cure model with random cure times. Biometrics 2001;57:282–6.