ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/282208297

An Overview on the Complement of Kaplan-Meir Estimation and Cumulative Incidence Estimation in the Presence of Competing Risks_Simulation Approach

Article *in* International Scientific Research Journal · January 2015

DOI: 10.18483/IRJSci.69

CITATIONS
0

READS
445

1 author:

Chinnaiyan Ponnuraja
Indian Council of Medical Research
97 PUBLICATIONS **750** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Survival Models on Competing Risks Data View project

prohpylactic TB regimens in HIV View project

All content following this page was uploaded by Chinnaiyan Ponnuraja on 10 May 2016.

The user has requested enhancement of the downloaded file.

International Scientific Research Journal
Open Access Journal ~ www.irj.science ~ ISSN 2412-026X

An Overview on the Complement of Kaplan-Meir Estimation and Cumulative Incidence Estimation in the Presence of Competing Risks_Simulation Approach

Valarmathi Srinivasan[1], Babu C Lakshmanan[2], Chinnaiyan Ponnuraja[3]

[1]Department of Epid., the TN Dr MGR Medical University (Chennai), India
[2]Statistics, Cognizant Technology Solutions (Chennai), India
[3]Department of Statistics, NIRT, ICMR (Chennai), India

Abstract: Researchers are concerned with the methodological problems arising in the analysis of clinical trials when competing risks are present. A competing risk is defined as an event whose occurrence precludes or changes the probability of occurrence of a main event under examination. In this setting, the appropriate estimate of the probability of failure is described by the cumulative incidence. This function is not available in all statistical software packages except very few, the complement of Kaplan–Meier estimates are often used unsuitably instead of cumulative incidence function. When competing risks are present, the appropriate estimate of the failure probabilities is the cumulative incidence rather than the complement of Kaplan–Meier estimate. This paper compares these two methods of estimating cumulative probability of cause-specific events in the present of other competing events. The simulated data with three competing events is used to demonstrate the different estimates given by one minus Kaplan-Meier (1-KM) and cumulative incidence function. Also this paper evaluates the advantages and suitability of statistical methods using the cumulative incidence estimate over the complement of Kaplan Meier estimates (1-KM) method in clinical trial time to event competing data.

Keywords: Complement of Kaplan-Meier(1-KM), Cumulative Incidence, Cause-Specific Hazards, Competing Risks

INTRODUCTION

The standard preference of statistical methods in medical research for analysing time-to-event data are Kaplan Meier curves, the log rank test, hazard ratios and the Cox proportional hazard model. Time-to-event data arise in studies where we observe the time from particular starting point to a primary end point defined by the occurrence of a specific particular event of interest. However, in many situations, the primary endpoint consists of more than a few distinct events of interest and the ultimate failure being attributed to one event exclusively to the others, which defines competing risks circumstances. Even though the basic statistical methodology for analysing such competing risks data has been known for decades [Prentice, et al.(1978), Kalbfleisch and Prentice (1980)], there is still an uncertainty in the medical research especially in bio statistical community pertaining to how to approach this type of time to event data. The very big argument is that the usage between Kaplan-Meier and cumulative incidence function for group comparisons of survival data. The complement of Kaplan-Meier (1-KM) and Cumulative incidence are often used even though the prensence of competing risks and that they are belief that one and the other same. However, researchers often use the Kaplan Meier approach; that is the complement of Kaplan Meier (1-KM) to evaluate the survival probability of occurrence of a cause-specific endpoint, even if the appropriate data contain competing-risk events (Gooley, Leisenring et al. 1999). But it is not so. In the clinical literature says that it is still reasonably common to see this probability incorrectly estimated as the 1 - KM estimator (Gaynor et al., 1993). In Addition, this could over-estimation of the cumulative probability

This article is published under the terms of the Creative Commons Attribution License 4.0
Author(s) retain the copyright of this article. Publication rights with IRJ.Science.
Published at: http://www.irj.science/pub/issue/2015-06/
Article Number: 120150669; ISSN 2412-026X

Chinnaiyan Ponnuraja (Correspondence)
cponnuraja@gmail.com
+

of cause-specific failure. Majority of researchers, most of the printed articles, thesis and others are fall short to notice the reality. Quite a few articles reported that they were identical and several others reported in disparity especially in some situations. The failures from a competing event are treated as censored at the time of the event occurs while estimating the (1-KM). In this way, it assumes that the patients failing from a competing risk are no more or less likely to fail from the event of interest than the patients still at risk beyond this time. When the aim is to estimate the failure probabilities, this censoring is inappropriate because, after a competing event has occurred, failure of the event of interest is no longer possible. The subject of competing risk events and the estimation of cumulative incidence of an event of interest have been discussed by many authors. Gail (1975) reviewed the theoretical concepts and the estimation of cumulative incidence of an event using a collected works of models. Prentice et al (1978) discussed likelihood inference to examine the effect of prognostic factors on the event of interest in the presence of competing risk events. Pepe and Mori (1993) described various probability models for summarizing competing risk survival data. Jager et al. (2008) discussed elaborately with reference to Kaplan-Meier methods for survival data. Tai et al (2001) extended a method to estimate the cumulative incidence of a specific event based on an extension of the Cox proportional hazards regression model. They compare their estimates to the Kaplan–Meier estimate of cumulative incidence as well as the cumulative incidence accounting for competing risk. There are several other authors cautioned us to make use of these approaches appropriately. Gooley et al. (1999) remarkably demonstrate the difference between the cumulative incidence and the complement of a Kaplan–Meier estimate (1-KM). Klien et al. (2001) reviewed more in detail the difference between the Kaplan–Meier and cumulative incidence curves, focusing the interpretation and examine how they perform in the presence of competing risks of bone marrow transplant data. The imperative procedures were discussed elaborately on the usage and inevitability of cumulative incidence estimation through Kaplan-Meier estimation procedure in the presence of competing risk breast cancer data by Satagopan et al. (2004). In addition, a detailed explanation was given on how the complement of Kaplan-Meier is one and the same to the cumulative incidence estimation method with a briefed example. Southern et al. (2006) mentioned Kaplan-Meier estimation methods yielded misleading results in competing risk circumstances. Verduijn et al. (2011) suggested that the Kaplan–Meier method profoundly overestimates the cumulative mortality probabilities for each of the separate causes of death especially in cardiovascular mortality data.

METHODS

Two approaches for an example given in (Satagopan et al. (2004)), it assumes 100 breast cancer patients lived for at least 1 year following surgery and five patients died at the end of the first year and subsequently 10 more patients die in the next year. This could be viewed into two different approaches. The Kaplan-Meier approach says the estimated survival at first year is 95% and 89.5% for the second year. The overall survival probability for a specified time up to second year is 85%. The cu,ulative incidence function approach looked into this issue in another way. That is the cumulative incidence of an event at a specified time of two years is 15%. This is basically the converse of survival. In other words, the cumulative incidence of an event at a given time is one minus the overall survival probability at that time. Therefore, this proves the complement of Kaplan-Meier is equal to cumulative incidence. The dispute is that it is not reflecting the reality at all occasions, especially in the presence competing risks.

DATA SIMULATION

Coviello and Boggess (2004) produced stata commands for competing risks simulation using STATA. Simulations were done to investigate to move toward the goal of this paper. The dataset is being reproduced easily using stata command as stcompet. To this make an effort, two types of failures are assumed, and a time for each type of failure is generated for 10,000 subjects with a constant hazard being 0.25 for the first type of failure and 0.99 for the second type of failure. A subject is assumed to fail from the event that occurs early if it occurs before time equals to two units. In addition, this is being compared with a real data set for further confirmation.

```
.set obs 10000
obs was 0, now 10000
.set seed 1234
.gen t25 = -1/.25 * log(1-uniform())
.gen t99 = -1/.99 * log(1-uniform())
.gen time = min(t99,t25)
.gen byte fail = (t25<t99) + 2*(t25>=t99)
.replace fail = 0 if time>=2
(799 real changes made)
.replace time = 2 if time>2
(799 real changes made)
.stset time, f(fail==1) noshow
.sts gen KM = s
.gen Complement = 1- KM
.stcompet CumInc=ci, compet1(2)
```

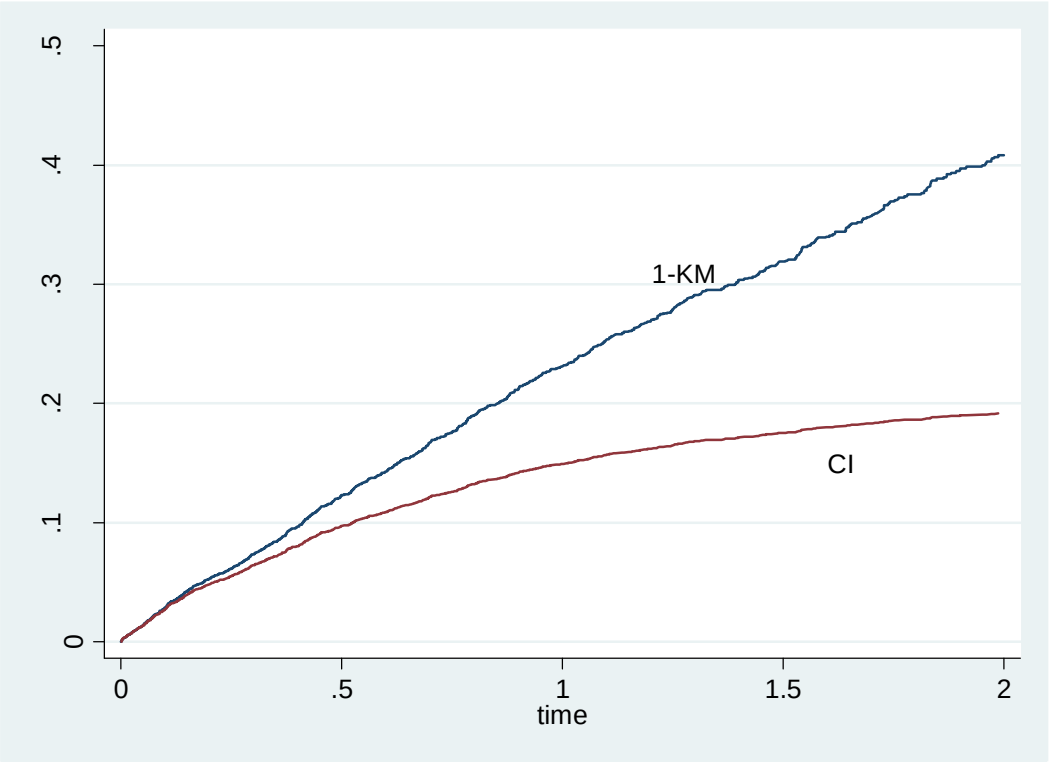
Table1: Simulated data using STATA command for 10000 observations with cumulative incidence and the complement of Kaplan-Meier(1-KM)

t25	t99	Time	fail	KM	Complement	CumInc
2.592848	1.703039	1.703039	2	0.65543232	0.3445677	0.7016
1.605753	0.4594849	0.4594849	2	0.8883607	0.1116393	0.3416
1.689981	0.37906	0.37906	2	0.90774236	0.0922576	0.2921
1.771015	0.9418474	0.9418474	2	0.79277793	0.2072221	0.5514
11.43155	0.2922089	0.2922089	2	0.92798877	0.0720112	0.2316
7.85666	3.760532	2	0	0.59522048	0.4047795	
2.803153	0.0987121	0.0987121	2	0.97311244	0.0268876	0.0885
0.3814599	0.4152758	0.3814599	1	0.90730919	0.0926908	0.0785
1.209467	0.6798491	0.6798491	2	0.84342303	0.156577	0.4517
0.5543839	0.0456071	0.0456071	2	0.9878636	0.0121364	0.0419
0.4237287	1.334055	0.4237287	1	0.89755741	0.1024426	0.0851
5.217102	3.439965	2	0	0.59522048	0.4047795	
9.209641	0.9679183	0.9679183	2	0.78811693	0.2118831	0.5597
0.3671193	1.584009	0.3671193	1	0.91161063	0.0883894	0.0755
1.527169	0.6594082	0.6594082	2	0.84923266	0.1507673	0.4424
1.536165	2.728147	1.536165	1	0.6858788	0.3141212	0.1717
4.064649	0.1344252	0.1344252	2	0.96216435	0.0378357	0.118
4.649208	1.724921	1.724921	2	0.65047198	0.349528	0.7051
0.3147765	0.5804275	0.3147765	1	0.92306492	0.0769351	0.0672
1.899437	0.2999618	0.2999618	2	0.92653758	0.0734624	0.2368
4.245422	1.062849	1.062849	2	0.76891417	0.2310858	0.5845
4.357155	0.621662	0.621662	2	0.85503038	0.1449696	0.4255
6.750236	0.4837573	0.4837573	2	0.88359533	0.1164047	0.3558
2.706083	0.3819796	0.3819796	2	0.90730919	0.0926908	0.2941
0.9547318	1.580237	0.9547318	1	0.78968294	0.2103171	0.141

The main event of interest for failure is the occurrence of the first type of failure and then the competing risk is specifying as the occurrence of the second type of failure. The cumulative incidence created, where the estimate of the cumulative incidence is recorded for both types of failure at each time when a corresponding failure occurs. To attain a plot for comparing 1-KM with the cumulative incidence of the first type of failure, a new variable "Complement" have been created to contain only the approximate pertaining to it.

.gen CI=CumInc if fail==1
(8127 missing values generated)
.twoway line Complement CI time, sort xlabel(0(.5)2) ylabel(0(.1).5) ytitle("Probability") legend(off) text(.31 1.2 "1-KM", place(e)) text(.15 1.6 "CI", place(e))

Figure 1: Complement of the Kaplan-Meier (1-KM) estimate and cumulative incidence of the first type of failure.



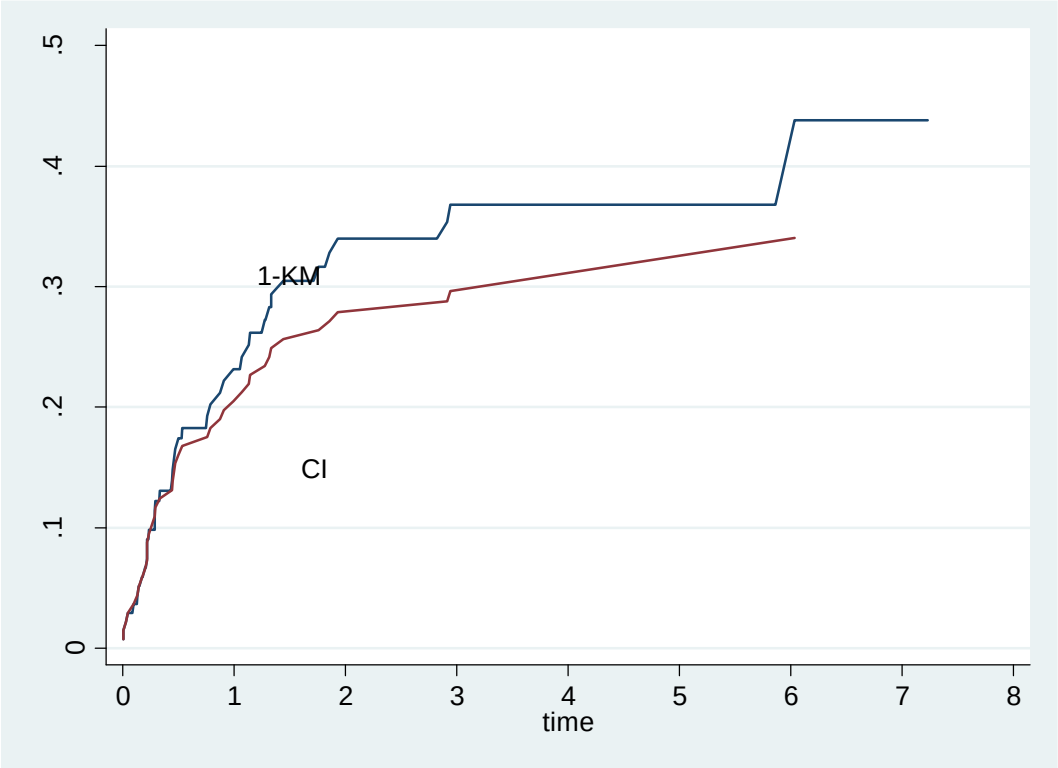
The figure 1 outlined the comparison between the complement of Kaplan-Meier (1-KM) with the cumulative incidence. In a competing risk setting, the complement of the Kaplan–Meier overestimates the true failure probability, whereas the cumulative incidence is the appropriate quantity to use.

Example from Real Data

A Dataset of 137 patients who underwent bone marrow transplant was used for the study from Klein and Moeschberger (1997). The Patients were categorized at the time of transplant into one of three risk categories: ALL (acute lymphoblastic leukemia) is coded as 1, AML (acute myelocytic leukemia)-Low Risk is coded as 2 and AML-High Risk is coded as 3.

The endpoint of interest is the disease-free survival time, which is the time to death or relapse or to the end of the study in days. In this data set, the variable Group represents the patient's risk category, the variable T represents the disease-free survival time, and the variable status is the censoring indicator, with the value 1 indicating an event time, value 2 indicating patients die before experiencing the event, the value 0 as a censored time. This data is used to estimate the probability of disease progression by calculating 1-KM and CI. 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 precisely the value of CI. On the other hand, the value of 1-KM at this time is (Figure 2), the difference being due to the patients who failed from the competing risk of death without progression

Figure 2 Complement of the Kaplan-Meier (1-KM) estimate and cumulative incidence of the first type of failure of Bone Marrow Transplant Data



DISCUSSION

The cumulative incidence of an event of interest, estimated by accounting for competing risk events is the probability of experiencing the event of interest by a given time and it is not experiencing a competing risk event by this time. Though, Mathematically the complement of Kaplan-Meier that is one minus Kaplan–Meier and cumulative incidence estimations are equivalent in producing results (Satagopan et al.(2004)) but it is entirely different in the case of competing risk . Noordzij et al. (2013) demonstrated that the Kaplan-Meier method overestimates the probabilities of both the event of interest and the competing event but they yield similar result when there are no competing risks. This paper imitates exactly the same version of Noordzij et al. (2013) and also elucidates that the complement of Kaplan-Meier estimates are not reciprocal of the cumulative incidence estimation method particularly when in the presence of competing risks. Perhaps these two methods may yield equivalent results as in giving their probabilities conceivably when there are no competing risks.

REFERENCES

1. Coviello, V. and M. Boggess. 2004. Cumulative incidence estimation in the presence of competing risks. The Stata Journal 4 : 103–112.
2. Gooley TA, Leisenring W, Crowley J, Storer B. Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. Stat Med 1999; 18: 695 – 706 http://dx.doi.org/10.1002/(sici)1097-0258(19990330)18:6%3C695::aid-sim60%3E3.0.co;2-o
3. Jager KJ, van Dijk PC, Zoccali C, et al. (2008). *The analysis of survival data: the Kaplan-Meier method. Kidney Int .74:560-565. http://dx.doi.org/10.1038/ki.2008.217*
4. Klein, JP. and Moeschberger, ML. (1997): *Survival Analysis, Techniques for Censored and Truncated Data Springer. http://dx.doi.org/10.1002/pst.135*
5. Noordzij M, Leffondré K, van Stralen KJ, Zoccali C, Dekker FW, Jager KJ (2013). When do we need competing risks, methods for survival analysis in Nephrology? Nephrol Dial Transplant. 28(11):2670-7. http://dx.doi.org/10.1093/ndt/gft355
6. Pepe MS, Mori M. (1993). Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Stat Med*; 12: 737-751 http://dx.doi.org/10.1002/sim.4780120803
7. Satagopan JM[1], Ben-Porat L, Berwick M, Robson M, Kutler D, Auerbach AD. (2004). A note on competing risks in survival data analysis. Br J Cancer. 4;91(7):1229-35. http://dx.doi.org/10.1038/sj.bjc.6602102
8. Southern DA, Faris PD, Brant R, et al (2006). Kaplan-Meier methods yielded misleading results in competing risk scenarios. *J Clin Epidemiol.* 59:1110-1114. http://dx.doi.org/10.1016/j.jclinepi.2006.07.002
9. Prentice R, Kalbfleisch J, et al. The analysis of failure times in the presence of competing risks. Biometrics 1978; 34:541–554. http://dx.doi.org/10.2307/2530374
10. Kalbfleisch J, Prentice R. The Statistical analysis of failure time data. Wiley, New York., 1980. http://dx.doi.org/10.2307/3315078

View publication stats