Survival Analysis I: The Kaplan-Meier Method

Vianda S. Stel[a] Friedo W. Dekker[a, b] Giovanni Tripepi[c] Carmine Zoccali[c]
Kitty J. Jager[a]

[a]ERA-EDTA Registry, Department of Medical Informatics, Academic Medical Center, University of Amsterdam, Amsterdam, and [b]Department of Clinical Epidemiology, Leiden University Medical Centre, Leiden, The Netherlands; [c]CNR-IBIM Clinical Epidemiology and Pathophysiology of Renal Diseases and Hypertension, Renal and Transplantation Unit, Ospedali Riuniti, Reggio Calabria, Italy

Key Words

Epidemiology · Survival analysis · Kaplan-Meier method · Nephrology

Abstract

The Kaplan-Meier (KM) method is used to analyze 'time-to-event' data. The outcome in KM analysis often includes all-cause mortality, but could also include other outcomes such as the occurrence of a cardiovascular event. The purpose of this article is to explain the basic concepts of the KM method, to provide some guidance regarding the presentation of the KM results and to discuss some important limitations of this method. To do this, we use a clinical example derived from the nephrology literature.

Copyright © 2011 S. Karger AG, Basel

Introduction

The Kaplan-Meier (KM) method is a popular method to analyze 'time-to-event' data. Within nephrology, the outcome variable in survival analyses is often all-cause mortality. For example, the aim of the study of Kovesdy et al. [1] was to examine the association of activated vitamin D treatment and predialysis all-cause mortality in patients with chronic kidney disease. In this study, the investigators used time from the start of the study to the date of death. However, the KM method can also be used for the occurrence of other events such as fracture in renal transplant recipients [2]. Investigators may even make use of composite endpoints. For example, in a study on the effects of intensive glucose lowering in type 2 diabetes, the outcome was a composite of nonfatal myocardial infarction, nonfatal stroke or death from cardiovascular causes, whichever came first [3]. The purpose of this article is to explain the basic concepts of the KM method, to provide some guidance regarding the presentation of the KM results and to discuss several limitations of this method.

Clinical Example

To explain the KM method, we make use of a clinical example that is based on a study by Tsakiris et al. [4]. One of the aims of this study was to examine all-cause mortality in patients starting renal replacement therapy (RRT)
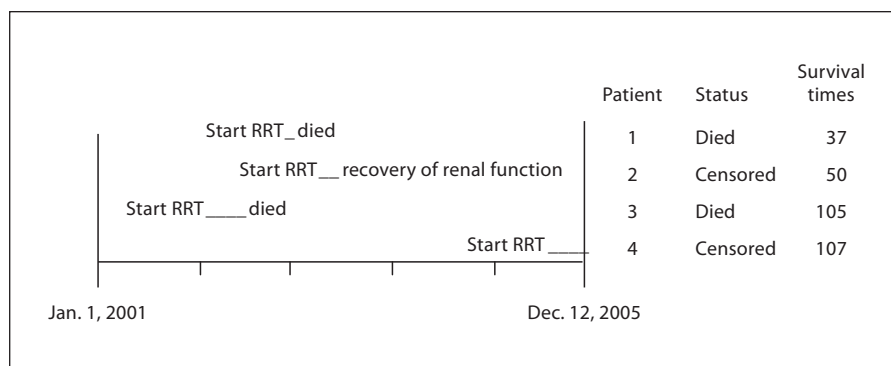
Vianda S. Stel, PhD
ERA-EDTA Registry, Academic Medical Center, Department of Medical Informatics
J1b 114.1, PO Box 22700
NL–1100 DE Amsterdam (The Netherlands)
Tel. +31 20 566 7637, E-Mail v.s.stel@amc.uva.nl

Fig. 1. Survival times of 4 non-MM patients at risk of death on RRT. The starting point was the start of RRT, and the end of the follow-up period was December 31, 2005. The event studied was death. Follow-up time was censored at loss to follow-up, recovery of renal function, 365 days after start of dialysis or at the end of the follow-up time on December 31, 2005 (whichever came first).

	Patient	Status	Survival times
Start RRT_died	1	Died	37
Start RRT__ recovery of renal function	2	Censored	50
Start RRT____died	3	Died	105
Start RRT ____	4	Censored	107

Jan. 1, 2001   Dec. 12, 2005

for end-stage renal disease due to multiple myeloma or light chain deposit disease. Data on 159,637 patients derived from renal registries sending individual data to the European Renal Association-European Dialysis and Transplant Association (ERA-EDTA) Registry for at least 12 years in the period 1986–2005 were used for the study. Patients with ERA-EDTA primary renal disease code 82 indicating multiple myeloma or light chain deposit disease were referred to as MM patients. Data from MM patients were compared with those from all others, referred to as non-MM patients.

For the current article, we used a random sample of 20 MM and 20 non-MM patients from this study who started RRT in 2001–2005 and compared their 1-year survival. In the analyses, the time of observation started at the initiation of RRT, the end of the follow-up period was 365 days after the start of RRT or December 31, 2005 (whichever came first), and the event studied was death (table 1).

Censored Observations

In some MM and non-MM patients death did not occur within the maximum observation period of 1 year. For these patients the survival time was censored. Table 1 shows the reasons for censoring used for the KM analyses in this study. Additionally, figure 1 shows 4 of the 20 non-MM patients from the start of RRT to death or censoring. From the time of recovery of renal function, patient 2 was no longer at risk of death on RRT. Furthermore, patient 4 was still alive at the end of the study. We did not know if patients 2 and 4 died later. We only knew that they did not die on RRT within the observation period of the study, and for this reason their survival times were censored [5].

Table 1. Overview of events and censoring as defined for the Kaplan-Meier analyses for the clinical example

Survival type	Start of observation	Event	Censoring
Patient survival on RRT	Start of RRT	Death	365 days after start of RRT, or end of follow-up (31/12/2005), whichever came first Loss to follow-up Recovery of renal function

There are several assumptions for censoring in survival analyses [6]. A first assumption when using the KM method is that at any time the patients who are censored should have the same survival prospects as those who are still being followed in the study. In the clinical example, however, patients were censored at the time of recovery of renal function and these patients may be healthier and, thus, may have better survival prospects than the patients still in the study. This means that although censoring at recovery of renal function is needed, this assumption on censoring may not be completely fulfilled. However, as only 1.3% of the study participants recovered renal function, this violation will induce only a very limited bias in the survival probability on RRT estimates, which might be somewhat underestimated [7].

Please note that 1 patient's survival time was censored because he experienced a 'competing' event that makes further follow-up impossible (a patient who recovers from renal function cannot die on RRT anymore) [5, 7]. In this situation, less biased survival probability estimates can be provided by using so-called competing risk models [7], but explanation of these models are beyond the scope of this paper. Another important assumption

for censoring is that the survival probabilities should be the same for patients who were recruited early and patients who were recruited late during the study period. This assumption was tested by comparing the survival curves for patients who were recruited early and those who were recruited late. The survival curves were similar, meaning that the assumption was fulfilled.

Calculations for KM Survival Table and Curve

A simple way to calculate the probability of 1-year mortality in non-MM patients on RRT would be to divide the number of deaths in this first year of RRT by the total number of patients at the initiation of RRT. This way would be correct in the hypothetical situation in which all patients could be followed for the entire first year on RRT; in other words, if in none of these patients' survival time needed to be censored.

In our clinical example including 20 non-MM patients, 3 patients died in the first year on RRT. The simple way of calculating the probability of 1-year mortality would therefore be $(3/20) \times 100 = 15\%$; however, we will show in our clinical example why this simple approach is incorrect.

Not all patients could be followed for a full year because they only recently started dialysis or recovered renal function. When having censored observations, as in patient 2 and 4 (fig. 1) during a follow-up time shorter than 1 year, the risk of dying within 1 year as calculated by the ratio will underestimate the real probability of death during the first year. This is because persons may have died after censoring but still within the period of 1 year after the start of RRT [8, 9]. The simple ratio is therefore inadequate.

This problem of underestimation of mortality will decrease or even disappear when calculating the risk of dying within a shorter period, such as 1 day. Within our clinical example, however, we are not interested in the risk of dying within 1 day only, though we are interested in the risk of dying at each of the 365 days within the follow-up period. The KM method can calculate the risk of dying at each specific day within the 1 year of follow-up, conditionally on having survived until that specific day. To this end, the KM method first calculates the probability of surviving a specific day, which is multiplied by the probability of having survived until that day (resulting in the cumulative survival at a certain day). Then, the probability of dying (cumulative mortality) is 1 minus the probability of survival (cumulative survival).

Table 2 shows the calculations on the data from the clinical example using the KM method (i.e. calculating the cumulative survival and cumulative mortality). First, the time in days was ordered separately for the non-MM and MM patients, starting with patients with the shortest follow-up time. It is then possible to calculate the cumulative survival (and cumulative mortality) at a certain day. In the following we describe the calculation of survival probability using the data of non-MM patients up to day 105 of follow-up.

Day 0
In this study 20 patients were at risk. All patients were alive; the proportion surviving on this day and the cumulative survival (i.e. proportion of all cases that were still alive at this day) were both 1. The cumulative mortality was 0.

Day 1 (not shown in table 2)
The cumulative survival on day 1 was the same as that on day 0, because on day 1 no patients died or were censored. Therefore, table 2 only includes calculations at days when events or censoring took place.

Day 37
Of the 20 patients who were at risk of dying, 1 died on this day. This means that the proportion surviving on this day was $19/20 = 0.9500$. The cumulative survival, which is the probability of surviving this day (i.e. day 37) multiplied by the probability of having survived the previous period, was $0.95 \times 1 = 0.9500$ (95.0%). The cumulative mortality up to this day was 0.05 (5.0%).

Day 50
There were 19 persons at risk of dying on this specific day (because 1 had already died at day 37). One patient's renal function recovered and survival time for this patient was censored on this day. As no persons died on this day, the proportion surviving on this day was 1 and the cumulative survival (95.0%) and cumulative mortality (5.0%) did not change. This shows that by censoring, one is able to reduce the number of patients at risk without affecting the cumulative survival.

Day 105
The number of patients at risk of dying on this day was 18 (because the survival time of 1 person was censored on day 50). At day 105, 1 person died and the proportion surviving on this day was $17/18 = 0.9444$. To calculate the cumulative survival up to day 105, this proportion was

Table 2. Cumulative survival table in 20 non-MM patients and 20 MM patients

Time in days	Number at risk	Deaths	Withdrawn alive (censored)	Proportion surviving on this day	Cumulative survival[1]	Cumulative mortality (1-cumulative survival)
Non-MM patients						
0	20	0	0	1	1	0
37	20	1	0	19/20 = 0.9500	0.9500	0.05
50	19	0	1	1	0.9500	0.05
105	18	1	0	17/18 = 0.9444	0.8972	0.1028
107	17	0	1	1	0.8972	0.1028
205	16	1	0	15/16 = 0.9375	0.8411	0.1589
365	15	0	15	1	0.8411	0.1589
MM patients						
0	20	0	0	1	1	0
23	20	1	0	19/20 = 0.9500	0.9500	0.05
25	19	1	0	18/19 = 0.9474	0.9000	0.1
48	18	1	0	17/18 = 0.9444	0.8500	0.15
51	17	0	1	1	0.8500	0.15
87	16	1	0	15/16 = 0.9375	0.7969	0.2031
98	15	1	0	14/15 = 0.9333	0.7438	0.2562
107	14	1	0	13/14 = 0.9286	0.6906	0.3094
138	13	1	0	12/13 = 0.9231	0.6375	0.3625
162	12	1	0	11/12 = 0.9167	0.5844	0.4156
188	11	1	0	10/11 = 0.9091	0.5313	0.4687
196	10	1	0	9/10 = 0.900	0.4781	0.5219
330	9	1	0	8/9 = 0.8889	0.4250	0.575
331	8	1	0	7/8 = 0.8750	0.3719	0.6281
345	7	1	0	6/7 = 0.8571	0.3188	0.6812
365	6	0	6	1	0.3188	0.6812

[1] Cumulative survival is calculated as the proportion surviving on this day × cumulative survival over the previous period.

multiplied by the 0.9500 cumulative survival of the previous step, resulting in a cumulative survival of 0.8972 (89.7%) and a cumulative mortality of 0.1028 (10.3%).

Reporting of Survival Results

Survival Plots

When analyzing survival data, the survival curves calculated with the KM method should always be plotted [10]. Using the cumulative survival table, it is possible to make a survival plot for the MM and non-MM patients (fig. 2a). A survival curve always goes in 'steps', as the cumulative survival remains the same until the day another person experiences the event. Furthermore, within survival curves the censored observations are indicated by a vertical dash. It is also possible to present a plot showing the increasing cumulative mortality (fig. 2b). It should be

noted that although the cumulative survival and mortality plots contain the same information, the visual perception could be quite different when using different scales on the y-axis [11].

Pocock et al. [11] provided recommendations for the presentation of survival plots. One of these recommendations is that the number of patients still at risk should be listed below the time axis and that the plots should only be extended through the period of follow-up achieved by a reasonable proportion (10–20%) of participants [fig. 2a, b; at the end of the first year there were still 15 patients at risk in the non-MM patients (75%) and 6 patients in the MM patients (30%)]. Another recommendation is that plots should include some measure of uncertainty when comparing the survival of two or more groups. For KM curves this could be the standard error or 95% confidence interval at regular time points to represent each plot's uncertainty. Please note that this does not directly display
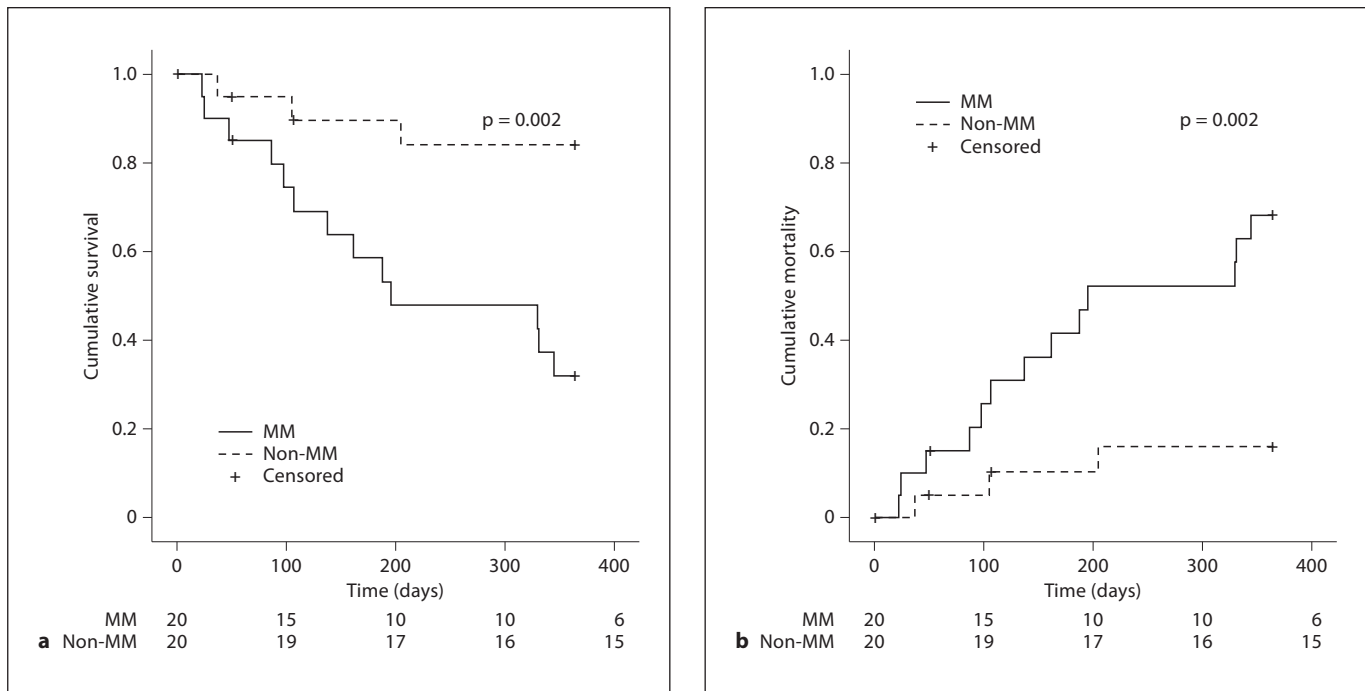
Fig. 2. Survival plots for MM and non-MM patients: cumulative survival (**a**) and cumulative mortality (**b**). The numbers below the figures denote the number of patients 'at risk' in each group.

the uncertainty of the treatment effect which is usually of primary interest. To present overall estimate of treatment effect, hazard ratios may be presented with confidence intervals using Cox regression analyses.

Summarizing Survival

Using the cumulative survival table one could present, for example, the 1-year cumulative mortality probabilities (68.1% in MM patients and 16.9% in non-MM patients) or 1-year cumulative survival probabilities (32.9% in MM patients and 84.1% in non-MM patients). Another way of summarizing survival is presenting the median survival, which is the time point at which the cumulative survival drops below 50%. In the MM group, the median survival was 196 days. In the non-MM group, however, it was not possible to present the median survival, as in this group the cumulative survival was more than 50% after 1 year of follow-up. It was neither possible to present the 70 or 75% survival in the non-MM group. Only by an analysis of a longer follow-up time would it have been possible to estimate any of these for this patient group. A disadvantage of reporting survival probabilities and the median survival is that it provides a comparison at one arbitrary point in time [10].

Log-Rank Test

The log-rank test compares the survival of two groups, taking the whole follow-up period into account [10]. The log-rank test, however, is only a test of statistical significance (providing a p value); it cannot provide an estimate of the size of the difference between the groups or any confidence interval [10]. Within our clinical example, the log-rank test provided a p value of 0.002 indicating that the difference in survival between the MM and non-MM group was statistically significant.

What We Can and Cannot Conclude from These Results

The results of the KM analyses suggested that the unadjusted 1-year mortality probability was higher in MM patients (68.1%) than in non-MM patients (16.9%), and that according to the log-rank test this difference was highly statistically significant (p = 0.002).

However, in order to answer a research question, additional analyses may be needed. A limitation of the KM method is that the log-rank test is purely a significance test and cannot provide an estimate of the size of the dif-

ference between the groups and its related confidence interval. Another limitation of the KM method is that it only provides unadjusted mortality (and survival) probabilities. However, to make a fair comparison between the MM and non-MM groups, it may be needed to adjust for potential confounders, like age and sex. The next article in this series will explain the basic concepts of Cox regression analyses which would be able to take into account these additional analyses.

Acknowledgement

References

1 Kovesdy CP, Ahmadzadeh S, Anderson JE, Kalantar-Zadeh K: Association of activated vitamin D treatment and mortality in chronic kidney disease. Arch Intern Med 2008;168: 397–403.
2 Nikkel LE, Hollenbeak CS, Fox EJ, et al: Risk of fractures after renal transplantation in the United States. Transplantation 2009;87: 1846–1851.
3 Gerstein HC, Miller ME, Byington RP, et al: Effects of intensive glucose lowering in type 2 diabetes. N Engl J Med 2008;358:2545–2559.

4 Tsakiris DJ, Stel VS, Finne P, et al: Incidence and outcome of patients starting renal replacement therapy for end-stage renal disease due to multiple myeloma or light-chain deposit disease: an ERA-EDTA Registry study. Nephrol Dial Transplant 2010;25: 1200–1206.
5 Altman DG, Bland JM: Time to event (survival) data. BMJ 1998;317:468–469.
6 Bland JM, Altman DG: Survival probabilities (the Kaplan-Meier method). BMJ 1998; 317:1572.
7 Verduijn M, Grootendorst DC, Dekker FW, et al: The analysis of competing events like cause-specific mortality – beware of the Kaplan-Meier method. Nephrol Dial Transplant 2010;26:56–61.

8 Noordzij M, Dekker FW, Zoccali C, Jager KJ: Measures of disease frequency: prevalence and incidence. Nephron Clin Pract 2010; 115:c17–c20.
9 Jager KJ, Zoccali C, Kramar R, Dekker FW: Measuring disease occurrence. Kidney Int 2007;72:412–415.
10 Bland JM, Altman DG: The logrank test. BMJ 2004;328:1073.
11 Pocock SJ, Clayton TC, Altman DG: Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls. Lancet 2002;359:1686–1689.