Kaplan-Meier plots

Dr Ari Samaranayaka
BSc, MPhil, PhD
Senior Research Fellow and Biostatistician
Centre for Biostatistics
Division of Health Sciences
University of Otago

Associate Professor Robin Turner
BSc, MBiostat, PhD
Unit Director
Centre for Biostatistics
Division of Health Sciences
University of Otago

Dr Claire Cameron
BSc(Hons), DipGrad, MSc, PhD
Senior Research Fellow and Biostatistician
Centre for Biostatistics
Division of Health Sciences
University of Otago

Survival analysis investigates the time until the occurrence of an event. Often, the event of interest is death, however, it can equally be time to any other event such as recovery from a condition, wait time for elective surgery, first recurrence of cancer after surgery, the intervals between successive births, or the time until equipment failure. Why is time-to-event analysis different from other analyses? Why can we not use other standard statistical tools like a t-test or least square regression to analyse it? First, most of those analyses rely on the normally distributed residuals assumption, which does not hold with time-to-event data because this data is always positive and sometimes highly skewed. For example, consider time to death after a high risk surgery; many patients may die shortly after surgery and those who do survive will then live for a long time. Second, it is common, when measuring time-to-event outcomes, for events to not happen during the study period for some individuals, which means some observations are 'censored'. Survival analysis using the Kaplan-Meier (K-M) survival estimator does not assume a specific distribution, therefore it is a nonparametric method. This means that the normality assumption and the assumption that all outcomes are observed are not required. Before explaining the K-M estimator, let us look at some terminology.

Imagine measuring the time to an event among a cohort of individuals. Not everyone will enter the study on the same date, so time zero for each individual is the day the person entered the study. The study period might end before all participants experience the event, also, some people might drop out during the study. We will use the following terminology:

Censored: the event has not occurred, or the subject was not under observation when the event occurred.

Interval censoring: rather than observing the exact time of event, we only observed that the event occurred between two known time points.

Followup period: the period during which the subject was under observation. Followup starts when the person enters study, and ends when either the event occurs or the study ends – whichever comes first. This period can be shorter than the study period if the event occurred during the study, or if the person leaves the study.

Survival function: the probability of survival up to a particular time point as a function of time. This is different to the instantaneous probability of survival as a function of time.

Hazard rate: the instantaneous rate of an event occurring. This is known as the failure rate, conditional failure rate, or hazard function. This rate has no upper bound, unlike a probability.

Hazard ratio: the ratio of hazard rates corresponding to two levels of an explanatory variable. For example, in a drug study, the ratio of hazard rates among treated and control populations is used as a measure of the effect of the treatment.

Presenting a survival function as a K-M plot is one way to describe a cohort's survival time graphically. The focus of this article is to describe K-M plots and in which circumstances they can be used, and thus, how to interpret them correctly.

Table 1 gives an hypothetical example of survival times in days in ascending order for two groups of people treated with two different procedures for the same condition. All 21 people in Group 1 and 11 of 20 people in Group 2 died during the followup period of 36 days. The other nine people in Group 2 were either lost to followup, or alive at the end of the study, therefore their survival times are censored. The question is, how do we compare the survival in these two groups?

Comparing the mean survival times in two groups (ignoring censoring), Group 1 (8.4 days) has about half the survival time of Group 2 (16.3 days). Alternatively, comparing the risk of dying, or the hazard, in two groups (again, ignoring censoring): the mean hazard in Group 1 (21 deaths in 177 days of followup or 0.119 deaths per day) is about 3.5 times that of Group 2 (11 deaths over 326 days of followup or 0.034 deaths per day). Neither of these methods are satisfactory because they ignore the censored observations.

The K-M curve compares instantaneous rates in the two groups. The K-M curve is defined as the probability of surviving a given length of time (treating time as many small intervals). There are three assumptions in this analysis: (1) at each time interval, censored individuals have the same survival prospects as those who continue to be followed during the interval; (2) survival probabilities are the same for those recruited earlier and later in the study; and (3) the events happen at the times specified, rather than between two time points.

The K-M estimate involves first computing probabilities of survival during each time interval as the number who survived over the period divided by the number at risk at the start of the period. The total probability of survival to the end of each time interval is calculated by multiplying the probability of survival for that interval with all the

Table 1: survival times in days for each person.

Group 1	Group 2
1	6
1	6
2	6
2	6+
3	7
4	9+
4	10
5	10+
5	12+
6	13
8	16
8	17+
8	18
9	20+
11	22
12	23
13	25+
15	32+
17	32
20	36+
23	

+ *indicates censoring.*

Kaplan-Meier plot for two groups

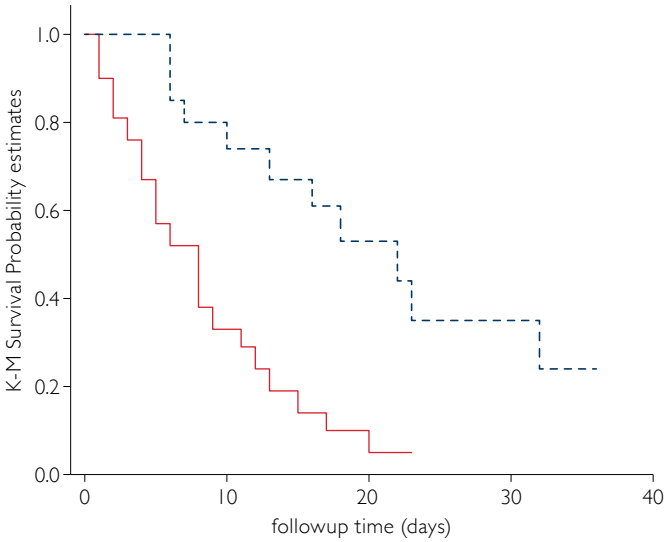

Figure 1: K-M survival plot comparatively describing survival in two groups.

probabilities for earlier time intervals. This calculation is shown in Table 2 for Group 2, the group with censored observations. The table begins at time zero (start of followup). The reason for this is to allow for the possibility of censoring before the earliest failure time.

Note that although 11 out of the 20 in Group 2 (55%) died over the 36 weeks (and 45% did not), the K-M estimate for the survival at 36 weeks is 24%. That is because the K-M estimator does not consider those who died or survived beyond their followup. The K-M survival plot displays the first and last columns of this table. Figure 1 shows the K-M plot for both groups.

Figure 1 shows that estimated survival is lower in Group 1 than in Group 2. The steeper slope shows that the rate of events is higher, i.e. events occurred faster. If we repeated the experiment, we would be unlikely to get the same two curves because there is uncertainty associated with these estimates. The logrank test is often used to decide if the observed difference between curves is expected if the two corresponding populations have the same survival rates.[1] In other words, the logrank test is used to test the hypothesis that there is no difference regarding survival among individuals in two groups. Another commonly used method to compare survival curves is the Cox proportional hazards model. This model allows for adjustment of potential confounding. More information on this method is given by Bewick and colleagues.[2]

References

1. Bland JM, Altman DG. The logrank test . BMJ. 2004 May 1;328(7447):1073.

2. Bewick V, Cheek L, Ball J. Statistics review 12: survival analysis. Crit Care. 2004 Oct;8(5):389–394.

Table 2: estimating K-M probabilities for Group 2 (the group with censored data).

Followup time period (days)	Number alive (ie, at risk) at start of the period	Number dead during the period	Number censored	Survival probability over the period	Probability of survival up to the end of the period
0	20	0	0	1.00	1.00
6	20	3	1	(20–3)/20=0.85	1.00*0.85=0.85
7	16	1	0	(16–1)16=0.94	0.85*0.94=0.80
9	15	0	1	(15–0)/15=1.00	0.80*1.00=0.80
10	14	1	1	(14–1)/14=0.93	0.80*0.93=0.74
12	12	0	1	1.00	0.74
13	11	1	0	0.91	0.67
16	10	1	0	0.90	0.61
17	9	0	1	1.00	0.61
18	8	1	0	0.88	0.53
20	7	0	1	1.00	0.53
22	6	1	0	0.83	0.44
23	5	1	0	0.80	0.35
25	4	0	1	1.00	0.35
32	3	1	1	0.67	0.24
36	1	0	1	1.00	0.24