

Comparing Kaplan-Meier curves - what are the (SAS) options?

Rob Allis, Amgen Ltd., Uxbridge, UK

ABSTRACT

In survival analysis the log rank test is commonly used to compare the Kaplan-Meier curves of two treatments. Part of our role is to provide the SAS® code to perform the log rank test, but this is only part of the picture. Do you understand what assumptions are being made? Do you know when the log rank test might not be optimal? Are you aware of the other options for comparing Kaplan-Meier curves? Why has the statistician chosen the log rank test over another?

This paper reviews different statistical techniques for comparing Kaplan-Meier curves and gives answers to some of the how, when and why's which may not be immediately obvious from looking at the Statistical Analysis Plan.

INTRODUCTION

In oncology trials the event of interest, including but not limited to disease progression or death, may not occur for all subjects before the end of the study; subjects may be withdrawn for a variety of other reasons (censored event). The effect of other ancillary factors may also be judged to extend or decrease this time to the event of interest endpoint. All of this data can be taken into account to build an estimate of the survival probability. We can then use this to plot Kaplan Meier curves representing this survival function over time. Test statistics can also be formulated to compare two or more survival curves.

From a SAS programmer perspective, the PROC LIFETEST procedure can be used to create and provide tests to compare the survival curves for two different populations. The usual format of a SAS dataset for this analysis will comprise one observation per subject, a binary indicator variable (CENSOR) with a value of 1 indicating the time to the event of interest is complete or 0 indicating the time to the event was censored, a time to event (MONTHS), a treatment group (TRT) used to formulate a comparison and several covariates (SEX, AGE) which might also be considered to have an effect on survival.

This paper will set the scene by introducing the default output of PROC LIFETEST and then take the reader on a journey through the range of statistical tests used in the context of comparing survival curves.

DEFAULT PROC LIFETEST OUTPUT

Assuming that we have two treatment groups $k=1,2$ to which n_1, n_2 subjects are allocated, such that the total number of subjects is given by $N = n_1 + n_2$. The statistical survival methodology in PROC LIFETEST, invoked by the syntax below generates a table of survival probabilities for survival times t_1 to t_M (for each treatment group) such that because of ties (events occurring at the same time), $M \leq N$. The **STRATA** statement divides the data into two separate strata comprising the two treatment groups in this instance. The output in Table 1 details an example of the product-limit estimates for a hypothetical 1st treatment group (Stratum 1).

```
PROC LIFETEST data=OncData;  
  TIME months*censor(0);  
  STRATA trtgrp;  
RUN;
```

PhUSE 2009

Stratum 1 (TRT k=1)					
Time	Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.000	1.0000	0	0	0	20
171.000	0.9500	0.0500	0.0487	1	19
179.000	0.9000	0.1000	0.0671	2	18
217.000	0.8500	0.1500	0.0798	3	17
224.000*	.	.	.	3	16
225.000	0.7969	0.2031	0.0908	4	15
225.000	.	.	.	5	14
225.000	0.6906	0.3094	0.1053	6	13
256.000	.	.	.	7	12

Table 1: Product-Limit Survival Estimates for treatment group 1.

At time 0, all subjects n_1 (20) are alive so the probability of survival is 1.000.

At time 171, one event of interest occurs and the cumulative probability of survival from time 0 is $1 \cdot (r_j - 1) / r_j = 19/20 = 0.9500$ where 1 corresponds to the probability of survival at the previous time point(s) and r_j is the number at risk at time j .

At time 224, a censoring event (indicated with an asterisk) occurs, however this censored event does not alter the probability of survival however it does affect the risk set, decreasing the survival probability for future calculations.

At time 225, a tied event has occurred.

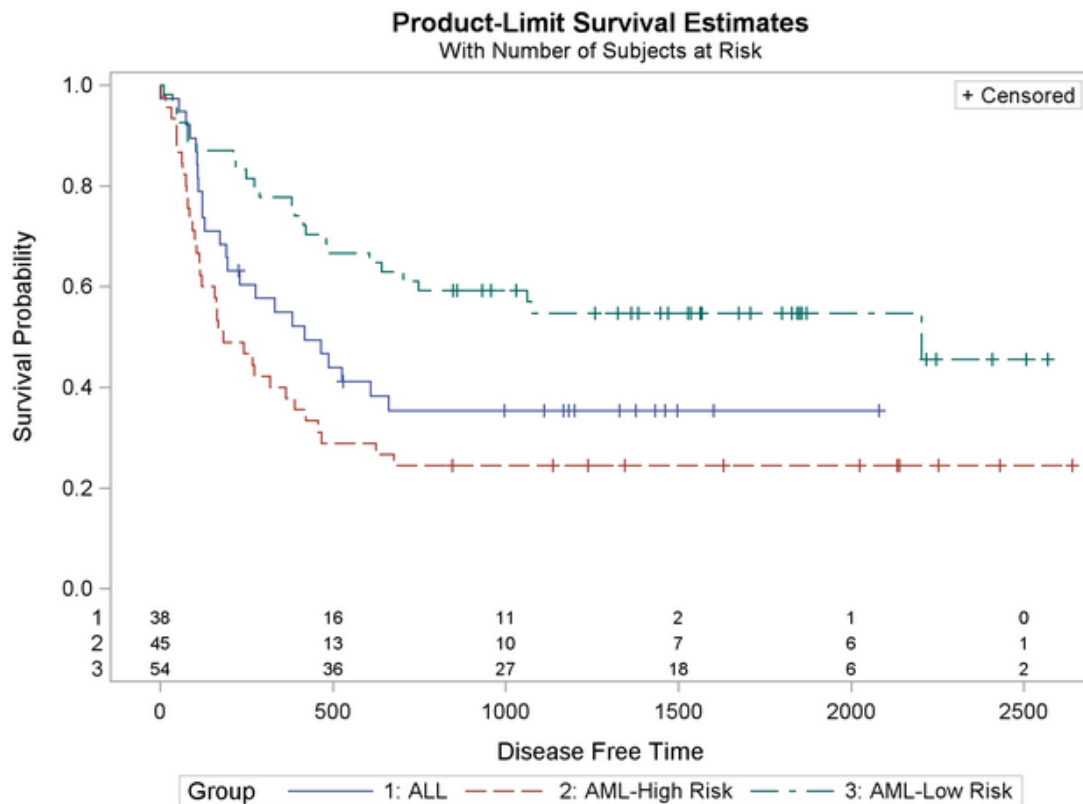


Figure 1: Example output from SAS online doc (v9.2) showing risk sets annotated via ODS GRAPHICS.

The Kaplan Meier graph, a plot of the survival distribution function over time can be generated directly from PROC LIFETEST with the **PLOTS = (s)** option. Several other plots are available and are discussed later. A more tailored graph can be obtained by extracting the survival probabilities from LIFETEST using the **OUTSURV=** option and using SAS GRAPH with the annotate procedure. This plateau and stepped plot is a non increasing function and documents the distribution of the survival probabilities over time. Each plateau represents the situation where the survival probability stays constant as time increases and it is common to see tick marks on the plot during the plateau representing subjects where time to an event is censored (suppressed using **NOHTICK** option). The stepped section represents a point at which a progression or death event has occurred.

The Greenwood's standard errors provided by PROC LIFETEST offer an insight into the precision of the estimates of survival. Since the Greenwood's formula requires large risk sets (asymptotic theory) when the risk set is low (censoring proportion less than 50%) this may make the estimates questionable and a review of the risk sets should be used to check this. This can be obtained from the PROC LIFETEST output and plotted on the graph via annotate or in SAS version 9.2 a table of risk sets can be plotted directly through the **ODS GRAPHICS PLOTS** statement. An alternative to the Greenwood's formula is Peto's formula which produces variance estimates that increase apropos to diminishing number of subjects at risk as apposed to just the death or progression events. The alternative Peto's formula is **not** currently an option within SAS.

To visualize the confidence interval of a survival probability at a single fixed time point on the Kaplan Meier curve Pointwise confidence limits can be plotted around the survival curve. The probability assumption of these being between 0 and 1 can fall down in certain circumstances however the **CONFTYPE=** option can be used to specify either the **log-log(default)**, **arcsine-square root**, **logit**, **log** or **linear** functions. These methods will not be discussed in this paper. Note SAS version 8 calculated the pointwise confidence intervals using a linear statistical model however in SAS version 9 this has changed to a log-log transformation. Interpretation of and conclusions drawn from the afore mentioned confidence interval should be limited to a particular time point, however when conclusions need to be made on a range of time points or the entire survival period, simultaneous confidence intervals with upper and lower bands can be used. The **SURVIVAL** statement with the **CONFBAND=** option and keyword **EP** – equal precision confidence bands (proportional to the pointwise confidence bounds), **HW** – Hall and Wellner confidence bands (not proportional to the pointwise confidence bounds) or **ALL** – both EP and HW can be used to specify these bands.

The PROC LIFETEST also outputs estimates of the 25th, 50th and 75th percentiles. The 50th percentile is the median and represents the time at which half the subjects on the trial have experienced the event of interest. Similarly the 25th and 75th percentiles occur when $\frac{1}{4}$ and $\frac{3}{4}$ of subjects have experienced the event. These statistics provide a useful summary of the rate at which events occur. Also estimated is the mean survival time which corresponds to the area under the Kaplan-Meier curve. If the largest observed time in the data is censored (plateau in the graph) the survival curve is not a closed area. However the **TIMELIN=time-limit** option can be used in this situation to calculate the area under the curve up to a certain time.

STATISTICAL COMPARISON

All test statistics that compare Kaplan Meier curves between two groups, weight the differences between the curves in different ways. For example the Log-rank test (**/TEST=(LOGRANK)**) weights differences that occur earlier and later in the curve equally. On the other hand the Wilcoxon (**/TEST=(WILCOXON)**) test weights earlier differences higher than later differences (in-fact by the number in the risk set). Along with the likelihood ratio test these tests are provided by default when the **STRATA** statement is used. Other, non-default tests (detailed in table 2) that can be specified as an option on the **STRATA** statement include the Tarone-Ware test (**/TEST=TARONE**) which uses a weight based on the square root of the number of subjects at risk. This means that weights attached to individual events are greater than the log-rank test and less than the Wilcoxon test. In comparison the Tarone-Ware test is always superior to the least powerful of the Log-rank or Wilcoxon test. The Peto-Peto test (**/TEST=PETO**) uses weights equal to the Kaplan-Meier estimate of the survival function. Similar to the Wilcoxon test, this provides greater weight to the early events, weights eventually diminishing as the survivor function declines. The extension of this is the Modified Peto-Peto test (**/TEST=MODPETO**) that also takes account of the number in the risk-set. The Fleming 'family' of tests allows for similar alternatives but these will not be discussed here. The likelihood ratio test is also calculated however this assumes an exponential distribution which is rarely applicable in a survival model and can be largely ignored.

TEST=(list)	Name of test	Weight
LOGRANK	Log-rank	$w = 1$
WILCOXON	Generalised Wilcoxon (also known as Gehan/Breslow)	$w = R$
TARONE	Tarone-Ware	$w = \sqrt{R}$
PETO	Peto-Peto (also known as Peto-Peto-Prentice test)	$w = \hat{S}(t)$
MODPETO	Modified Peto-Peto test	$w = \hat{S}(t) \frac{R}{(R+1)}$
FLEMING(ρ_1, ρ_2)	Fleming-Harrington G ρ family of tests $\rho_2 = 0$ - Flemming(ρ) with one argument then $\rho = 0$ - log rank test then $\rho = 1$ - very close to Peto-Peto test	
LR	Likelihood ratio test based on exponential model	
ALL	All the nonparametric tests above with $\rho_1=1$ and $\rho_2=0$ for the Fleming (...) test.	
The log-rank test and collection of weighted tests above is a chi-squared test with k-1 degrees of freedom, where k is the number of groups. $\chi^2_{k-1} = \sum \frac{w(d - E)^2}{E}$		k = Number of (treatment) groups w = Weight function d = Number of deaths E = Expected number of deaths R = Number of subjects at risk $\hat{S}(t)$ = Survival function

Table 2: Table of test statistics.

The log rank test is optimal and will have maximum power out of all the linear rank tests under the proportional hazards assumption and when the distribution of the censoring events are the same across the strata. Using the **PLOTS=(lls c)** option, this provides two plots – the first of which, a plot of log(-log(estimated Survival distribution function) versus log time confirms proportional hazards if the lines are parallel. The second provides a plot of censored observations by strata. The addition of ticked points on the Kaplan-Meier graph can also help to identify bias caused by different patterns of follow-up. In cases where the assumption of proportional hazards does not hold other tests may have greater power. However neither the log-rank, nor the weighted log rank tests are good at detecting differences when survival curves cross. As can be seen there are many different weighting systems used which each provide a different test and it is the role of the statistician to pre specify the correct test for the most likely effect of the treatment.

Where increasing doses of a drug within a treatment group are assumed to benefit survival (e.g. a dose response study) a trend test can be formulated in PROC LIFETEST to test for this directional dosing effect within treatment using the **TREND** statement. An ascending or descending ordering variable needs to be created to enable these tests to be created.

If covariates are known or suspected of influencing the survival the **GROUP=** along with the **STRATA** statement can be used to formulate linear rank statistics to test the effect of particular covariates on survival. In this instance the **GROUP=**variable defines the treatment group whilst the **STRATA** statement facilitates the creation of stratified tests of homogeneity adjusted for the covariate SEX. Note: using the BY trtgrp statement to define strata works differently to the strata statement and will not pool over the strata to perform either a test of association of survival time with covariates nor a test of homogeneity across treatment groups.

PhUSE 2009

```
PROC LIFETEST data=OncData;  
  TIME time*censor(0);  
  STRATA sex / GROUP = trtgrp;  
RUN;
```

The **TEST** statement can be used to test a list of (**continuous**) covariates for their association to/what they bring to the survival estimate. In the example above using the statement **STRATA trtgrp / TEST sex age**, rank statistics are computed to test for which covariate brings the largest increase to the joint survival statistic thus testing for association. If the STRATA statement was omitted no tests of homogeneity would be performed.

CONCLUSIONS

Whilst there are a whole host of different options available in PROC LIFETEST to facilitate the creation of Kaplan Meier curves and tests to facilitate comparisons between survival curves, there is an equally comparative number of assumptions that need to be acknowledged to fully appreciate what is produced is correct and conclusions valid.

When making a choice on these methods one must pay particular attention to among other things; the proportional hazards assumption, the proportion of censoring and when and where along the survival time frame it is occurring, the size of the sample under consideration and or the distribution of the subjects at risk. Once these are taken into account it is possible to make a more informed decision on the type of test that may be used to compare Kaplan Meier curves.

REFERENCES

1. SAS OnlineDoc , V9.1.3, V9.2, <http://support.sas.com/documentation>
2. SAS Survival Analysis Techniques for Medical Research, 2nd Edition. Alan B.Cantor
3. SAS Survival Analysis using SAS: A Practical Guide. Paul D. Allison
4. A Handbook of Statistical Analyses using SAS, 3rd Edition. Geoff Der and Brian S. Everitt.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Rob Allis
Amgen Ltd
1 Uxbridge Business Park
Sanderson Road
Uxbridge UB8 1DH
UK
Email: rallis@amgen.com Web: <http://www.amgen.com>