

Time to Event ... (Kaplan-Meier Curve, Log Rank Test, SAS, Spotfire, Shiny R)

Michaela Mertes, F. Hoffmann-La Roche, Ltd., Basel, Switzerland

ABSTRACT

Time-to-onset of an event will be analyzed using an un-stratified log-rank test. With the help of an example dataset the calculation will be explained step by step. This will be supported with output and graphs from SAS 9.2®. Further example outputs added from other graphic providers (TIBCO Spotfire® and Shiny R).

INTRODUCTION

This paper gives the reader an introduction into the statistic's formulas to re-calculate manually in an easy way. With support of a real data example the plot is shown with a SAS output, Spotfire and Shiny R.

Kaplan-Meier Derivation

Kaplan-Meier estimator is a product of the form:

$$\hat{S}(t) = \prod_{t_i < t} \frac{n_i - d_i}{n_i}$$

n_i = Number of subjects at risk in interval

d_i = Number of subjects with an event in interval

t_i = Time point

Group A: Kaplan-Meier Estimator



10 little JELLY BABIES went to do a job

1st one got a call on day one, after that he dropped

9 little jelly babies went to do a job

2nd caught a flu, did quit at day two.

8 little jelly babies went to do a job

3rd one did roar with laughter already on day four

7 little jelly babies went to do a job

4th one did it finalize at day **five**

6 little jelly babies went to do a job

5th one sadly went to heaven at day seven

5 little jelly babies went to do a job

6th one solved without a break at day **eight**

4 little jelly babies went to do a job

7th and 8th get it done together at day **nine**

2 little jelly babies went to do a job

9th on day **ten** and 10th finished on day **eleven**.

-> on day 10, 10 subjects are at risk
without an event ($n_{10} = 10$, $d_{10} = 0$)

-> on day 5, 7 subjects are at risk
with **one** subject with an **event** ($n_5 = 7$, $d_5 = 1$)

-> on day 9, 4 subjects are at risk
with **two** subjects with an **event** ($n_9 = 4$, $d_9 = 2$)

Manually recalculated table 1 is using the formula of Kaplan-Meier estimator results from the data of Group A.

t_i	n_i	d_i	$\frac{n_i - d_i}{n_i}$	$\hat{S}(t)$
0				1
1	10	0	$\frac{10-0}{10} = 1$	1
5	7	1	$\frac{7-1}{7} = \frac{6}{7} = 0,8571$	$0,8571$
8	5	1	$\frac{5-1}{5} = \frac{4}{5} = 0,8$	$\frac{6}{7} \cdot \frac{4}{5} = 0,6857$
9	4	2	$\frac{4-2}{4} = \frac{2}{4} = 0,5$	$\frac{6}{7} \cdot \frac{4}{5} \cdot \frac{2}{4} = 0,3429$
10	2	1	$\frac{2-1}{2} = \frac{1}{2} = 0,5$	$\frac{6}{7} \cdot \frac{4}{5} \cdot \frac{2}{4} \cdot \frac{1}{2} = 0,1714$
11	1	1	$\frac{1-1}{1} = 0$	0

Table 1

At day 9 there are 4 subjects at risk and 2 subjects did have an event. The risk to have an event in this time interval is 50%. The risk to have no event up to this time is 34.29%.

Kaplan-Meier Curve

The results from Group A calculated in Table 1 is drawn into a coordinate system manually into the curve shown in Figure 1.

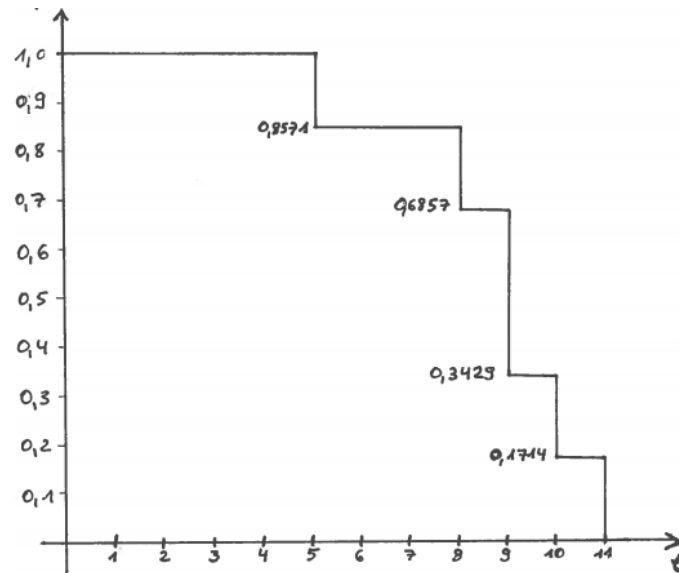


Figure 1

Group B: Kaplan-Meier Estimator with SAS

For the group B data are specified in a SAS program. First part of the 'data a' step is the information of Group A (arm='A'). The second part is specified with arm='B' for group B.

```
data a;
  usubjid='1- 1'; cnsr=0; time= 5; arm='A'; output;
  usubjid='1- 2'; cnsr=1; time= 7; arm='A'; output;
  usubjid='1- 3'; cnsr=0; time= 9; arm='A'; output;
  usubjid='1- 4'; cnsr=1; time= 1; arm='A'; output;
  usubjid='1- 5'; cnsr=0; time=10; arm='A'; output;
  usubjid='1- 6'; cnsr=1; time= 2; arm='A'; output;
  usubjid='1- 7'; cnsr=0; time= 9; arm='A'; output;
  usubjid='1- 8'; cnsr=0; time=11; arm='A'; output;
  usubjid='1- 9'; cnsr=0; time= 8; arm='A'; output;
  usubjid='1-10'; cnsr=1; time= 4; arm='A'; output;

  usubjid='2- 1'; cnsr=1; time=11; arm='B'; output;
  usubjid='2- 2'; cnsr=0; time=11; arm='B'; output;
  usubjid='2- 3'; cnsr=1; time= 9; arm='B'; output;
  usubjid='2- 4'; cnsr=0; time=11; arm='B'; output;
  usubjid='2- 5'; cnsr=1; time=10; arm='B'; output;
  usubjid='2- 6'; cnsr=0; time=12; arm='B'; output;
  usubjid='2- 7'; cnsr=0; time= 9; arm='B'; output;
  usubjid='2- 8'; cnsr=1; time= 9; arm='B'; output;
  usubjid='2- 9'; cnsr=0; time= 9; arm='B'; output;
  usubjid='2-10'; cnsr=1; time=13; arm='B'; output;
run;

proc lifetest data=a method=km alpha=0.05;
  time time * cnsr(1);
  strata arm /diff=control('B') test= (logrank) adjust=bon;
run;
```

The SAS output for Group B after running the program is the following:

The LIFETEST Procedure

Stratum 2: ARM = B

Product-Limit Survival Estimates

TIME	Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.0000	1.0000	0	0	0	10
9.0000	0.8000	0.2000	0.1265	1	9
9.0000*				2	8
9.0000*				2	7
9.0000*				2	6
10.0000*				2	5
11.0000	0.4800	0.5200	0.1910	3	4
11.0000				4	3
11.0000*				4	2
12.0000	0.2400	0.7600	0.1947	5	1
13.0000*				5	0

NOTE: The marked survival times are censored observations.

At day 9 there are 10 subjects at risk and 2 subjects did have an event, 2 subjects are censored.

Kaplan-Meier Curve

Figure 2 shows both groups A and B. In addition the p-value is specified provided by STREAM (Standard Reporting and Analysis Modules) Tool generated with SAS.

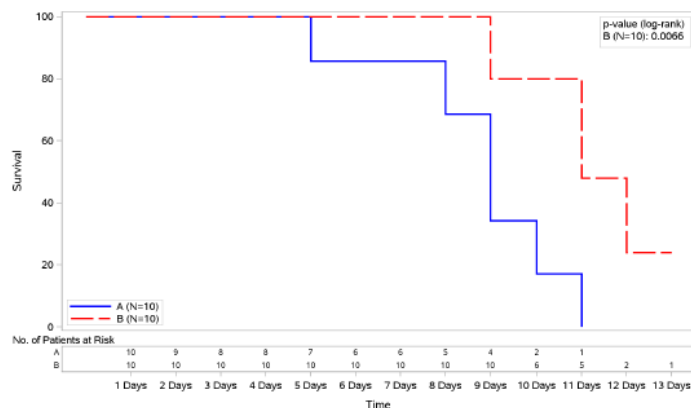


Figure 2

Log-Rank Test Calculation

Within a Microsoft Excel 2010® sheet Chi-square is being calculated manually to provide an insight how SAS or other systems do calculate the statistics:

examplelogrank

	A	B	C	D	E	F	G	H	I	J
1		d(Ai)	d(Bi)	d(i)	n(Ai)	n(Bi)	n(i)	e(Ai)	v(Ai)	e(Bi)
2								$\frac{d(i) \cdot n(Ai)}{n(i)}$	$\frac{d(i) \cdot n(Ai) \cdot n(Bi) \cdot (n(i) - d(i))}{n(i) \cdot n(i) \cdot (n(i) - 1)}$	$\frac{d(i) \cdot n(Bi)}{n(i)}$
3	time	Events A	Events B	Events (A or B)	At risk A	At risk B	At risk Total	expected events in A	Variance A	expected events in B
4	1	0	0	0	10	10	20	0	0	0
5	2	0	0	0	9	10	19	0	0	0
6	4	0	0	0	8	10	18	0	0	0
7	5	1	0	1	7	10	17	0.411764706	0.242214533	0.588235294
8	7	0	0	0	6	10	16	0	0	0
9	8	1	0	1	5	10	15	0.333333333	0.222222222	0.666666667
10	9	2	2	4	4	10	14	1.142857143	0.627943485	2.857142857
11	10	1	0	1	2	6	8	0.25	0.1875	0.75
12	11	1	2	3	1	5	6	0.5	0.25	2.5
13	12	0	1	1	0	2	2	0	0	1
14	Sum	6	5				Sum	2.637955182	1.52988024	8.362044818
15										
16							LogRank	3.362044818		-3.36204482
17							Chi-square		7.388385745	

At day 9 there are 2 events in Group A and in Group B, which is a total of 4 events. At day 9 there are only 4 patients at risk in Group A and still 10 patients at risk in Group B, a total of 14 patients.

To expect an event in A: $\frac{d(i) \cdot n(Ai)}{n(i)}$

è **Log-Rank** is the difference between the total number of Events in A and the sum of all expected events in A:

$$[6 - 2.637955182 = 3.362044818]$$

The variance is calculated: $\frac{d(i) * n(A_i) * n(B_i) * [n(i)-d(i)]}{n(i) * n(i) * [n(i)-1]}$

Chi-square: (Result of Log-Rank)² / Sum of variance = $3.362044818^2 / 1.52988024 = 7.388385745$

Output generated with SAS program code (proc lifetest) is as following:

Rank Statistics			
ARM	Log-Rank		
A	3.3620		
B	-3.3620		

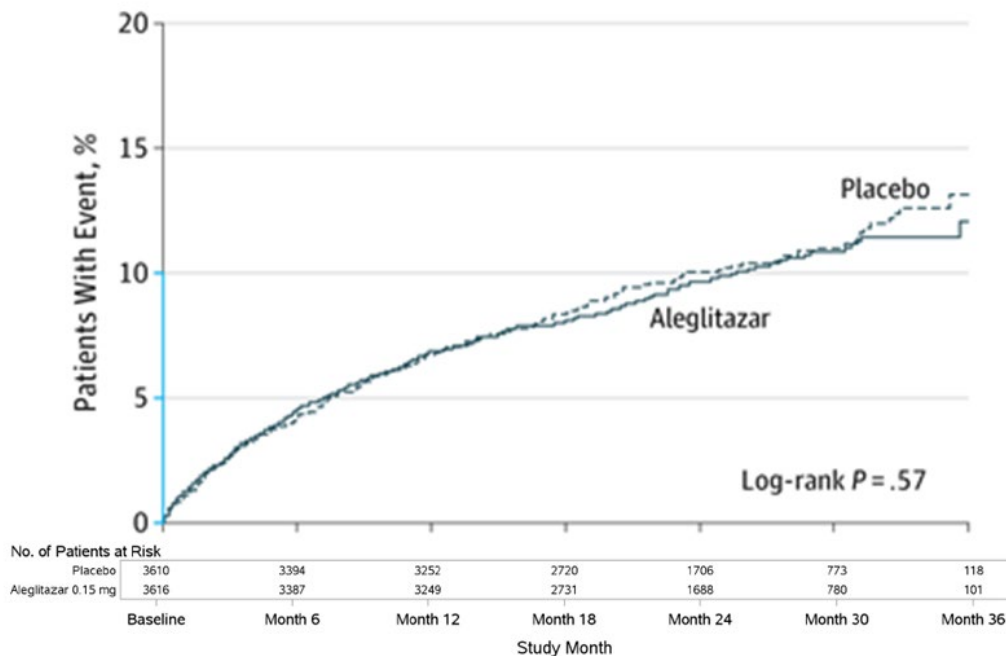
Covariance Matrix for the Log-Rank Statistics			
ARM	A	B	
A	1.52988	-1.52988	
B	-1.52988	1.52988	

Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	7.3884	1	0.0066

At the top of the Rank Statistics the Log-Rank test values are printed. The variance is shown below and according to the Chi-Square value the p-value is given: 0.0066. This p-value is also shown in Figure 2 where the Kaplan-Meier Curve was generated with STREAM.

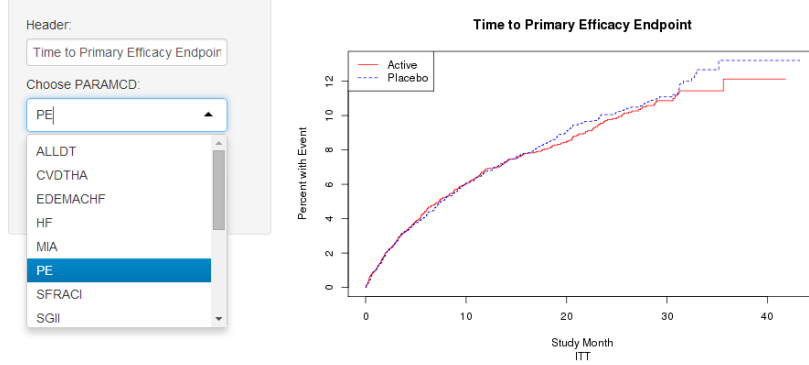
Real Data Example

effgpa_IT Time to first event in the composite primary efficacy endpoint (CEC-confirmed events only)
Protocol: BC22140
Population : ITT



Among patients with type 2 diabetes and recent ACS, use of aleglitazar compared with placebo did not reduce the risk of cardiovascular ischemic outcomes. These findings do not support the use of aleglitazar in this setting with a goal of reducing cardiovascular risk.⁽¹⁾

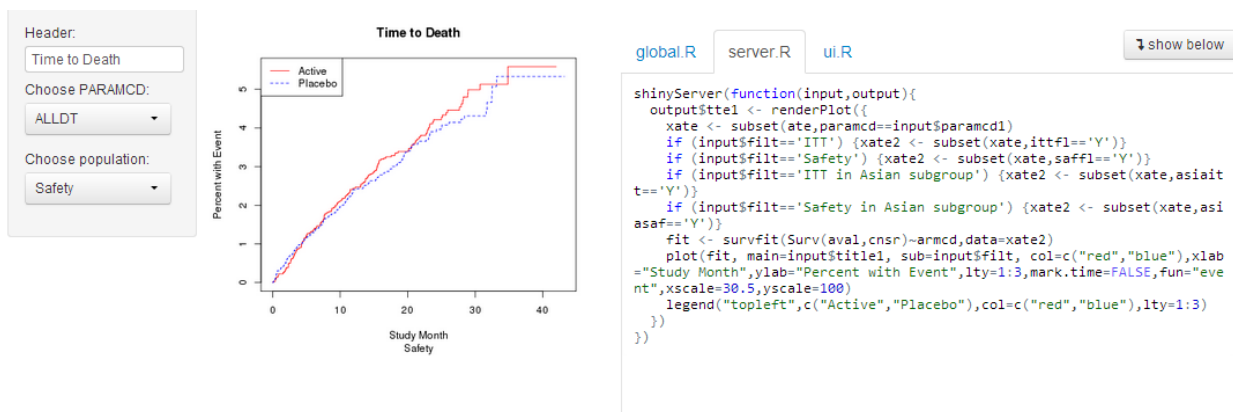
Real data example reproduced with Shiny R



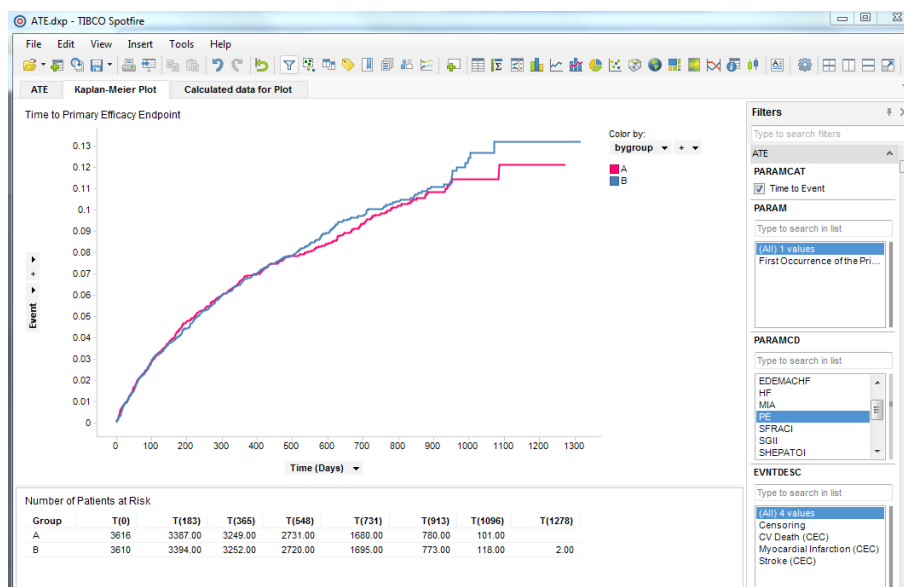
```
global.R server.R ui.R show with app

shinyUI(fluidPage(
  sidebarLayout(
    sidebarPanel(
      textInput("title1", "Header:", value="Time to Death"),
      selectizeInput("paramcd1", "Choose PARAMCD:", paramcd1vs, firstparamcd),
      selectizeInput("filt", "Choose population:", choices=list("ITT", "Safety", "ITT in Asian subgroup"), "Safety in Asian subgroup"), select="disp"
    ),
    mainPanel(
      plotOutput("t1e1")
    )
  )
))
```

The user can enter an endpoint and also the population (in ui.R). The next example using the same application shows Time to Death for the safety population.



Real data example reproduced with TIBCO Spotfire



PhUSE 2014

The plot created with Spotfire is based on a data function using an R script – Open Source R. Running this function on the time to event analysis dataset (ATE) new data tables for the plot are created. The Line Chart, which is an option in Spotfire, is using one of this created dataset.

The user can choose a different endpoint with clicking on a different filter on the panel on the right side for PARAMCD.

CONCLUSION

Time to event analysis and Kaplan-Meier Curves are manually calculated. It is easy to follow the SAS output created with proc lifetest. Creating the plot using STREAM is according to our standards.

Creating the plot using Shiny R needs insight in a new programming language. Set-up once, it can be reproduced and shared with others. My example gives the user the possibility to re-create different plots to different end-points and populations. If you want other filter options to be used, Shiny R has to be set-up differently.

Using Spotfire you need more knowledge to set-up the R script in a way, Spotfire can use the data for plotting. You can use more easily other data within the dataset for filtering the data. This doesn't need to be specified further as it has to be in Shiny R. But you don't have a program code to reproduce the plots as you have in SAS or R.

REFERENCES

http://en.wikipedia.org/wiki/Kaplan%E2%80%93Meier_estimator

Texts in Statistical Science: Modelling Survival Data in Medical Research, Second Edition, David Collett, 2003 by Chapman & Hall/CRC CRC Press LLC

⁽¹⁾<http://jama.jamanetwork.com/article.aspx?articleid=1854356#Conclusions>

ACKNOWLEDGMENTS

For producing the plot with Spotfire, I like to thank Jana Stoilova. I am using her script.

CONTACT INFORMATION

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

Michaela Mertes

F. Hoffmann-La Roche, Ltd.

Email: michaela.mertes@roche.com

Brand and product names are trademarks of their respective companies