Paper B2B10

Lost in Translation: A Statistician's (Basic) Perspective of PROC LIFETEST

Alexandra Buck, Rho®, Inc. Chapel Hill, NC

ABSTRACT

Survival analysis is a main-stay in clinical trials. It is the bread and butter of oncology trials. It can also be the most frustrating part of the statistical programmer's day. There can be a huge disconnect between the language of a biostatistician and a statistical programmer. The phrases "Survival Analysis", "Time to Event Analysis", and "PROC LIFETEST" can mean the same thing, or very different things. In this article, I provide a basic understanding of the reasoning and method behind PROC LIFETEST as well as a "statistician translator". Using example data and example code, I compare what the programmer sees to what the statistician would describe. Output such as .lst summaries, data sets, and plots gives the opportunity to explore what PROC LIFETEST can do and how to interpret the results. Tips for avoiding errors, choosing the best method, and utilizing the output data sets should make your next use of PROC LIFETEST more enjoyable and rewarding.

STATISTICIAN TRANSLATION

There is a famous quote attributed to Mark Twain: "There are three kinds of lies: lies, damned lies, and statistics." that statisticians hate to admit holds true in certain cases. We are magicians with a secret language. We increased our student loan debt and spent years learning the language, so we want to use it wherever possible. The problem occurs when statisticians want to use their language with programmers that are unfamiliar with the terms. This issue is compounded when we throw out these terms without explanation. Here are a few of the most-used terms with their basic definitions and information:

- Event: What we are calculating the time to. Also referred to as the 'event of interest'. This can include such things as death, recurrence of disease/symptoms, or complete recovery.
- Censor: What we use for subjects that do not experience an event. We cannot follow subjects indefinitely, but still need to be able to use them in the analysis. The most common form of censoring is right censoring. Think of a timeline, where there is a line marking the end of the study follow-up period. If the event ever did occur it would happen to the right of that line. With the CENSOR variable, 1 is commonly used to indicate that an event occurred and 0 indicates that the event did not occur (the subject was censored).
- Failure: Used to indicate subjects that experienced the event. This comes from the traditional use of survival analysis - analyzing death. If a subject experienced death, you would consider that subject and the treatment a failure in avoiding death.
- Survival: All of this is part of 'Survival Analysis', another name for time-to-event analysis. This is often used to calculate, model, and derive inference about the time to death. Hence the term 'survival' may be used even when we are analyzing recurrence or recovery.
- Survival Function: $S(t)=Pr[T>t]$. It is the probability that the actual time of event (if we could follow subjects indefinitely) is greater than the current time that we are evaluating.
- LIFETEST: The name of the SAS® procedure used. This procedure produces survival functions (estimates on how subjects are experiencing events), plots, and other pieces of survival analysis.

DATA

There are three required variables for input data sets: the subject variable, censor variable, and time to event variable. The subject variable can take any value to differentiate subjects. You should only have one observation per subject. If this is not the case, you will not have reliable results. The censor variable will denote whether the subject experienced the event or not. The time to event variable will take the value of the time the event occurred or the last observed time for censored subjects. This variable can be in any unit desired, whether hours, days, weeks, months, or years. Grouping variables such as gender or treatment are often included when the object of the analysis is to determine the effect of group differences in survival.

The appendix section Input Data set includes the data set that will be used to create output. SUBJID is our subject variable, TRT is our treatment grouping variable, and DUR (in days) is our time to event variable. The censor variable STATUS is the indicator of an event, where 1 indicates that death occurred and 0 indicates that it did not.

CODE

Now that there is data to analyze we can call on the LIFETEST procedure. This example code by no means illustrates the full capabilities of PROC LIFETEST. The terms to note are:

- METHOD: Specifies the method of estimation used to create the survival function. The Kaplan Meier (KM) method is the default and most common. This method is best when event times are measured with precision, especially if the number of observations is small. The other option is the Life-Table (LT) method and is better for large data sets or when measurement of event times is crude. For example if a subject has a visit every 6 months and a change occurred, you can only be sure of the window of the occurrence and not the exact time.
- STRATA: Denotes the variable to use when comparing survival functions.
- TIME: Assigns the time and censor variables.
    - Time variable * Censor variable (X)
    - (X) sets the value of the censor variable that indicates that no event occurred within the study.
- OUTSURV: Names the output data set that will contain survival estimates and confidence limits
- PLOTS: Specifies the output plot. The most common is the Kaplan-Meier plot (=s).
- ODS GRAPHICS: Produces more stylized plots.

With this information we can build our PROC LIFETEST call.

```
ODS GRAPHICS ON;
PROC LIFETEST DATA=tte METHOD=KM;
       OUTSURV=ttesurv PLOTS=s;
       STRATA TRT;
       TIME DUR*STATUS(0);
RUN;
ODS GRAPHICS OFF;
```

OUTPUT

PROC LIFETEST will produce a lot of helpful information. The hard part is deciphering the output. We will be breaking the information into sections and exploring the meaning behind the labels, values, and SAS notes. Refer to the Appendix for the output produced from the example data and code. The first section, Output #1, is the Product-Limit Survival Estimates. This section is output for both values of STRATUM. The first column is the time variable DUR, where all values of DUR in that stratum are displayed in order. There is an extra observation for time 0.00 that is always included in the output even if no event or censoring occurred at that time. If a time value is for a censored subject an asterisk (*) will be displayed to the right of the time.

The second column is the survival estimate at each time. As mentioned previously, survival can be defined as the probability that the actual time to event is greater than the current time. When a subject is censored, the survival estimate takes a missing value in the output. People are often confused by this output because when discussing survival, we treat the estimate as constant throughout censored times, not fading in and out of existence.

For example, in treatment 1 the survival estimate at 3 days would still be 1.0000 because no event had occurred. This estimate would not drop until 7 days when the first event occurred, lowering the survival estimate to 0.9091.

The third column (failure) is simply (1-survival), or the probability of the event occurring by that time point. Column four shows the survival standard error, which can used to create confidence intervals. SAS will also produce a reliable confidence interval in our output data set.

The fifth column, labeled Number Failed, will only increase when a subject experiences the event. The sixth column, Number Left, will start with the total number of subjects in that stratum and will decrease with each observation. A subject that is censored cannot be considered as part of the treatment group after the time of their last observation.

The next section denoted Output #2, Summary Statistics for Time Variable DUR, is where you can acquire the median survival time within the stratum. Percent indicates the quartile which can be then translated into survival values. 25 percent marks at the estimated time point where survival would drop to 0.75. The Point Estimate column displays that time point value if it available (if survival never drops below 0.25 then the 75 percent value cannot be estimated).

The median time to survival, or 50 percent is calculated as:

$$\hat{t}_{50} = \begin{Bmatrix} \min(t : \hat{S}(t) < 0.5, \text{ if } \hat{S}(t) \neq 0.5 \\ (1/2)\{t_j + t_{j+1}\}, \text{ if } \hat{S}(t) = 0.5 \end{Bmatrix}$$

If you do not have an event time where the survival estimate exactly equals 0.5 you would use the smallest observed survival time where the survival estimate is less than 0.5, which in the case of stratum 1 is 589 days. If there is a time where the survival estimate is exactly 0.5 you would take the average of the time where the estimate equals 0.50 and the following time.

Confidence intervals are also produced for the quartile estimates. Note that for the 75th percentile the upper limit was not computed. This is due to the right-censoring in the data and not a programmatic error. This right-censoring also caused SAS to generate a NOTE stating that the mean and standard error of survival time were underestimated. The actual event times of the censored subjects could be far into the future which would pull our mean survival time higher.

The Product-Limit Survival Estimates, Quartile Estimates and Mean/Standard Error Estimates are also produced for the second stratum, TRT=2. The comparison now begins. First, the Summary of the Number of Censored and Uncensored Values section in Output #3 gives basic frequencies of number of subjects, number of failures (experienced event), number of censored, and the percent censored. Note that having more censored subjects does not necessarily mean that the treatment was better. If all the subjects were censored before the end of study due to drop outs or loss of follow-up this is more of an indication that there was an issue with follow-up methods or possibly study design than an issue with the effectiveness of one treatment over another.

The Test of Homogeneity of Survival Curves for DUR over Strata section is usually of the most interest to clinicians. Output #5 is a partial illustration of this section. The Test of Equality over Strata section is where the p-values of interest lie. The output contains the default test. Other tests can be added using the /TEST() option in the STRATA statement. Our output does not give enough evidence to conclude a significant difference between survival curves at the alpha=0.05 level using any of the default tests.

Each default test has its own strengths. The Log-Rank test is better at detecting later differences in the survival curves, while the Wilcoxon test is better at detecting earlier differences in the survival curves. Results from either test should be considered with caution if the survival curves cross. Because the Likelihood ratio test [-2Log(LR) in the output] assumes that the data in the stratum are exponentially distributed its use is limited.

OUTPUT DAT ASET

The output data set created with the OUTSURV= option can be very useful, but can also cause confusion. This data will reflect what was in the Survival Estimates output, collapsed on DUR. If multiple subjects had the same time value they would be represented as two observations in Survival Estimates and one observation in the OUTSURV data. The most confusing part of this data set is the Censoring Flag variable. This is not the censor variable created in the input data set. In fact, given how we defined the censor variable, it is the opposite of the censor variable in terms of what the values represent. It takes 0 for subjects with an event and 1 for subjects without an event. Keep this in mind

when using the data set for graphing or other output. The OUTSURV data set contains the 95% confidence limits produced by SAS. The limits can be changed using the ALPHA= option in the PROC LIFETEST statement.

OUTPUT PLOTS

One of the quickest ways to analyze the data is to review the Kaplan Meier plot produced. Below is the default plot along with the plot produced when ODS GRAPHICS ON is initialized before the procedure. A higher survival probability at each time is preferred when the event is something undesirable, such as death. Treatment 2 stays above Treatment 1 throughout most of the plot. This resulted in relatively low p-values. However Treatment 1 has its last subject censored when the survival estimate is 0.1515 while Treatment 2 has its last subject experiencing the event, leading to the drop at the far right of the plot where the survival estimate is 0.

There are many ways to plot the survival data. There is the default output which is basic and gives the information needed. The addition of ODS GRAPHICS ON to the procedure creates a plot that is stylized and visually appealing. With both graphs there is no flexibility with how the information is displayed. If you need control of the axes, colors, labels, etc. you will have to use the OUTSURV data and your preferred plotting method.

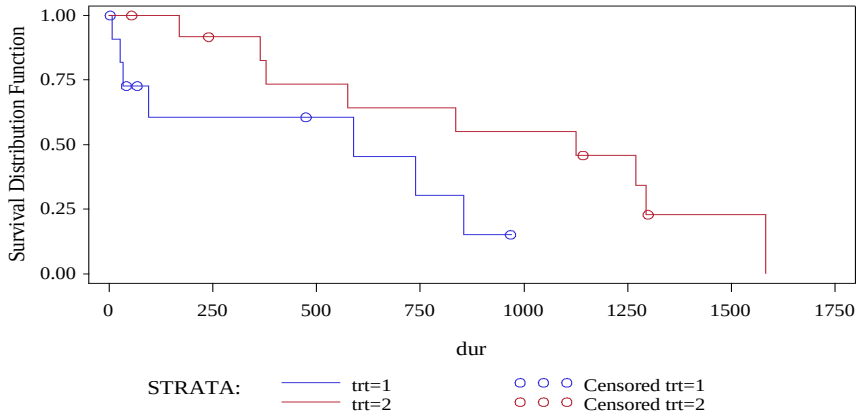

Figure 1. Default Survival Plot from LIFETEST

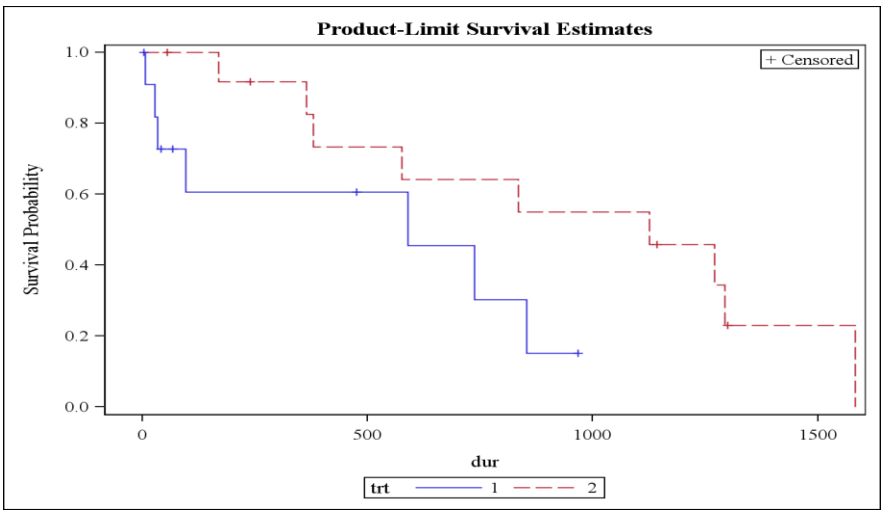

Figure 2. Survival Plot from LIFETEST with ODS GRAPHICS ON

TIPS/TRICKS

This section is simply a gathering of hard-learned lessons. Hopefully this will save you some time and headaches.

- Check your data
  - Multiple observations per subject will be counted twice in LIFETEST. LIFETEST is for one occurrence-only events. For example, death.
  - 0 is not always your censored value. Whatever value you decide to set in the data set can be your censored value
- Be sure of the method (KM or LT) you want to use
- You can display the survival probability with confidence limits at a specified cutoff. For example if you had survival estimates up to 300 days, but were only interested in survival at 150 days you could use the output data set already produced by LIFETEST utilizing the following code:
  - ```
    DATA DAY150;
          SET TTESURV;
          WHERE DUR<=150;
          BY TRT DUR;
          IF LAST.TRT;
    RUN;
    ```
  - This will give one observation per treatment group containing the forced 'final' survival estimate with 95% CI.
- If you want to display the survival function at certain dictated time points (such a 10 day periods) for easier reading you can use the TIMELIST option in the PROC LIFETEST statement
- Be aware that missing confident limit values (LCL or ULC) can come from the data and censoring. It is not necessarily the result of a coding error.
- If a group has no events, SAS will not be able to produce estimates for that group. Any output or analyzes created should not be considered reliable.

TRADEMARK NOTICE

REFERENCES

Allison, Paul D. 2010.Survival Analysis Using SAS®. Cary, NC: SAS Institute Inc.

CONTACT INFORMATION

Any questions or comments regarding the paper may be directed to:

Alexandra Buck
Rho, Inc.
6330 Quadrangle Drive
Chapel Hill, NC 27514
Phone: (919) 595-6326
Email: Alexandra_Buck@rhoworld.com
    ```

APPENDIX

Input Dataset:

SUBJID	TRT	DUR	STATUS
1	1	7	1
2	2	240	0
3	1	589	1
4	2	364	1
5	1	27	1
6	2	1300	0
7	2	169	1
8	1	854	1
9	2	55	0
10	1	96	1
11	2	1583	1
12	1	738	1
13	1	3	0
14	2	1126	1
15	1	968	0
16	2	576	1
17	1	34	1
18	2	835	1
19	2	1270	1
20	1	1143	0
21	1	475	0
22	2	68	0
23	2	379	1
24	1	42	0
25	2	1294	1

Output #1:

Stratum 1: TRT = 1

Product-Limit Survival Estimates

DUR	Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.00	1.0000	0	0	0	12
3.00*	.	.	.	0	11
7.00	0.9091	0.0909	0.0867	1	10
27.00	0.8182	0.1818	0.1163	2	9
34.00	0.7273	0.2727	0.1343	3	8
42.00*	.	.	.	3	7
68.00*	.	.	.	3	6
96.00	0.6061	0.3939	0.1574	4	5
475.00*	.	.	.	4	4
589.00	0.4545	0.5455	0.1765	5	3
738.00	0.3030	0.6970	0.1707	6	2
854.00	0.1515	0.8485	0.1370	7	1
968.00*	0.1515	0.8485	.	7	0

NOTE: The marked survival times are censored observations.

Output #2:

Stratum 1: TRT = 1

Summary Statistics for Time Variable DUR

Quartile Estimates

Percent	Point Estimate	Transform	95% Confidence Interval [Lower	Upper)
75	854.00	LOGLOG	589.00	.
50	589.00	LOGLOG	27.00	854.00
25	34.00	LOGLOG	7.00	589.00

Output #3:

Stratum 1: TRT = 1

Mean	Standard Error
477.67	125.62

NOTE: The mean survival time and its standard error were underestimated because the largest observation was censored and the estimation was restricted to the largest event time.

Output #4:

Summary of the Number of Censored and Uncensored Values

Stratum	TRT	Total	Failed	Censored	Percent Censored
1	1	12	7	5	41.67
2	2	13	9	4	30.77
Total		25	16	9	36.00

Output #5 (Partial):

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	3.2074	1	0.0733
Wilcoxon	3.4530	1	0.0631
-2Log(LR)	1.9573	1	0.1618