

AN INTRODUCTION TO SURVIVAL ANALYSIS

Linda M. Quinn

Abstract: There are several procedures in SAS® that perform various aspects of survival analysis, LIFETEST, LIFEREG, and PHREG. This talk will introduce the topic of survival analysis including definitions of survival distributions and censoring. The LIFETEST and PHREG procedures require similar structures for your dataset prior to analysis. Finally, a brief introduction and comparison of the three procedures will be shown. PROC PHREG will be discussed in the most detail. This talk is suitable to users will a basic or intermediate understanding of survival analysis who want to be able to use the SAS system to perform analyses.

Keywords: survival analysis, censoring, proportional hazards, PHREG, LIFETEST.

Introduction: Survival analysis is an analytic method to study the occurrence of events. While the classic survival analysis is used to study mortality, survival analyses can be applied to a wide variety of disciplines and problems. Survival analysis goes by many other names, depending on the discipline of application: reliability analysis and failure time analysis in engineering, event history analysis in sociology, and duration or transition analysis in economics.

In the application of survival analysis, the components are a change from one category to another and when the change occurred. The two variables that must be collected then are the variable that is changing status (alive to dead, success to failure) and the time until such an event occurs. For example, how many days until death

occurred (change is from alive to dead); the number of hours until the product fails (change is from functional to non-functional); or the number of months until success is achieved (change is from failure to success). As seen in the last example, the outcome in survival analysis is not always a bad outcome.

Example: The following example will be used throughout this paper for illustration. A randomized study following all patients with a certain medical condition was conducted to determine if a new treatment modality will cure this medical condition more quickly than the standard treatment modality. In this case, all patients will be cured. The question of interest is whether the time until cure is better using the new treatment. All patients are either ill or cured and once they are cured, there are no relapses. We start counting days from the day they are enrolled and started on a treatment regimen. Our dataset, TDATA, will contain the status of each patient at the last known time point of the study:

STATUS = 1 if cured
0 if ill;

the number of days until cured (or the number of days studied if not cured), DAYS; the study group they are in, GROUP (E=Experimental and S=Standard), and a patient ID number, ID.

A partial listing of the dataset shows:

ID	GROUP	DAYS	STATUS
1	S	5	1
2	S	18	0
3	S	17	1
4	S	12	1
5	S	13	1
6	S	19	1
7	S	12	1
8	S	13	1
9	S	21	1

ID	GROUP	DAYS	STATUS
10	S	23	0
11	S	25	1
12	S	14	1
13	E	30	1
14	E	2	0
15	E	10	1
16	E	4	1
17	E	11	1
18	E	8	1
19	E	3	0
20	E	4	1
21	E	4	1
22	E	7	1
23	E	9	1
24	E	1	1

Patient 13 was on the experimental therapy for 30 days until cured, patient 14 was on the experimental therapy for 2 days and died. Patients 1 and 2 were on the standard therapy for 5 and 18 days respectively. Patient 1 was cured and patient 2 was still ill. As it turned out patient 2 was followed for 18 days and was not cured when the study ended. Patient 14 was studied for 2 days and died from causes unrelated to the study. The last known status was at 2 days and was still ill.

Additional variables collected include aspirin usage at the start of the study and gender.

Options and Assumptions:

Censoring - Censoring occurs when there is no information available about the status of the study unit either before the origin of the study or at the conclusion of the study. In our example, patients 2 and 14 will be considered right censored. They are right censored because we do not know the time that they were actually cured. We only know that at the last point that we have information on, the patient is still ill. If we throw these subjects out of the analyses because they did not achieve the desired endpoint, we are throwing out information. We know that it was at least 18 or 2 days before these patients were cured. Right censoring occurs for

a variety of reasons: death, machine failure, lost to follow, or end of study. Right censoring is the most common in biomedical applications.

Left censoring occurs when you do not know the starting point. If we were interested in knowing how long since the onset of this medical condition before cure, we would have to know the starting point of this condition. Many subjects cannot remember the exact date of onset. In that case the patient will be left censored since we know that at the time of inclusion in the study, they had the condition. For the remainder of this presentation, we will only consider right censoring.

Time dependent covariates - with survival analyses another issue to consider is time dependent covariates. In traditional regression analyses, covariates usually do not change over time, but in survival analysis it is possible for covariates to change over the course of the study. While gender and race may not change, medication status, hospitalizations, marital status may all change over the course of data collection. In this case we may want to consider if they are taking aspirin concurrently with the therapy.

Survival distributions - an initial assumption on the form of the survival distribution is needed. There are three different ways of describing the probability distribution: cumulative distribution function, probability density function and the hazard function. The hazard function is the more popular way of describing the distribution.

Some examples of the hazard models for consideration include exponential and Weibull.

Data structure - For the simpler case of no time dependent covariates, the data structure is shown previously in the example dataset. The more complex case will can be found in SAS manuals.

PROC LIFETEST:

Survival curves are useful for preliminary examinations of the data, for calculating median survival time, or the one year mortality rates, and for evaluating the fit of regression models. The simplest analytic test is to determine if two or more groups have the same survival distribution.

PROC LIFETEST produces estimates of survivor functions using either of two methods (Kaplan-Meier or life table); graphs the survivor function; provided three methods for testing for differences between groups; and can test for significance of quantitative covariates.

Let's start by looking at the basic survivor function for both groups combined.

```
proc lifetest data=tdata plots=(s);
  time days*status(0);
run;
```

The *plots=(s)* option is used to graph the survivor function. Other graphs include the hazard function and the cumulative distribution function. The *graphics* option could be used when a high resolution graphics version is desired (over character resolution style). The *time* statement gives the time variable - in our case, DAYS. If there were no censoring, the statement would simply be:

```
time days;
```

Since there is right censoring, the censoring variable, STATUS, is used along with the values which should be considered censored. Since STATUS is

coded that 0 is ill, 0 is the value to censor. If we had coded STATUS as 0=did not complete the study; 1=lost to followup; 2=death; 3=cured, then the statement would look like:

```
time days*status(0,1,2);
```

In biomedical applications, Kaplan-Meier estimation is the most widely used method and is the default estimator by default in LIFETEST. To explicitly request it, use the *method=KM* as an option on the *proc lifetest* statement. Details on the life table method can be found in the SAS manuals.

But wait..... we aren't really interested in both groups combined. We want to see if the two groups, standard and experimental therapy, are different. We can do this by adding a *strata* statement.

```
proc lifetest data=tdata plots=(s)
  graphics;
  time days*status(0);
  strata group;
run;
```

Now, you will see both curves on the same graph. But can you tell if the curves are really different from just looking at them? Maybe or maybe not. Luckily, LIFETEST calculates the log-rank test and the Wilcoxon test to test the null hypothesis of no strata differences. Moreover, LIFETEST gives you these automatically whenever you use a *strata* statement.

Okay, now we are set to look at an abbreviated output (See OUTPUT #1).

Some things to point out: With LIFETEST, you will get a listing of every event period in the dataset for each strata. This can be extremely long if the dataset is large or if you have a large number of strata. The columns, in

respective order, in this listing gives you the survival rate, the failure rate, the standard error, the number that have failed and the number left. Censored observations are starred. So we start out at time 0 with a 100% survival rate and a 0% failure rate. Failure rate plus success rate will always add to 100%. Initially the number left will be the total number of subjects in the study.

Remember that in our example, "survival" rate means patients still having the medical condition and "failure" rate is our cure rate. The number failed will be the number cured and the number left are the number left to be cured.

The summary statistics will be useful to described the median time to cure as well as the mean time to cure. When the last observation is censored, the estimate of the mean is biased.

Next is the graphics. Now you can see how the two curves look in relation to each other.

Finally, are the curves really different? The log-rank test shows that the curves are really different since the p-value is very small.

Now what if we were interested in other covariates and their influence on survival? We can test whether quantitative variables are associated with survival time using a *test* statement. However, you can test the same hypothesis within PROC LIFEREG and PHREG. In most cases, this is the next step anyway. LIFETEST does not give parameter estimates and the other procedures do. The testing of covariates is sequential in that if multiple covariates are used on the test statement, the test is for each covariate

controlling for those before it (but not after it) in the listing.

LIFETEST is a good first step in survival analysis. It can be used to describe the form of the survival curve and as an analytic tool to examine differences in survival across groups. However, it is inadequate for parameter estimation. For more advanced modeling and hypothesis testing, LIFEREG and PHREG should be used.

Lifetest can also estimate confidence intervals around survival probabilities using the *outsurv=* option on the PROC statement. *Outserv=* also creates an output dataset with the estimated survival. Alternatively, the option *alphaqt=* option requests confidence intervals for each quartile.

PROC LIFEREG: Although PHREG has become more widely known as the survival PROC, there are features in LIFEREG that are not currently available in PHREG. PROC LIFEREG can be used for right and left censoring while PHREG can only handle left censoring. PHREG only works with the survivor function, but LIFEREG can test the hazard function. PROC LIFEREG will produce smaller standard errors if the shape of the survival distribution is known. LIFEREG creates indicator variables to represent categorical variables, whereas PHREG requires the creation of indicator variables in a DATA step.

PROC LIFEREG's disadvantage is that it does not allow for time dependent covariates while PROC PHREG does.

For the purposes of this presentation, the details of PROC LIFEREG will be skipped in favor of PROC PHREG.

PROC PHREG:

The PHREG procedure uses a Cox proportional hazards model to perform regression analysis of survival data. Cox's model does not require a particular probability distribution for survival times and is therefore semiparametric and considered more robust. It is not however the cure-all for survival analysis. There are times when a more rigorous parametric method is preferable. While PHREG is deficient in its graphic capability, it is very versatile in its analytic capability.

Statistical details of proportional hazard modeling will be kept to a minimum and the focus will be the implementation and interpretation of PHREG.

Let's start by duplicating what was done in LIFETEST. We are interested in testing whether there is a difference in survival between the two treatment therapies.

```
proc phreg data=tdata;
  model days*status(0) = group;
run;
```

The statements look remarkably similar. There are no graphics options to worry about in PHREG so the PROC statement is very simple. Instead of a time statement, we will use a model statement (similar to PROC REG). The dependent variable is the same `days*status(0)` and now we can list independent variables, such as `group`. The problem now is that `group` needs to be an indicator variable, not a categorical variable (with values E and S). Create a new variable, in a DATA step, called NEWTX, with a value of 1 for `group=E` and a value of 0 for `group=S`.

```
proc phreg data=tdata;
  model days*status(0) = newtx;
run;
```

Now let's look at the output (See OUTPUT #2).

The output for the basic PHREG run is much more condensed than LIFETEST. It starts by summarizing which dataset and variables are being used. The default method for handling ties is Breslow. Other methods can be specified as an option on the *proc* statement. Then there is a summary of the censored events. Next is a test of the global hypothesis, is the model significant or not. The null hypothesis is that all coefficients are zero. Three tests are given.

Finally the parameter estimate with a corresponding significance test and risk ratio is given. The risk ratio is just e^{β} . This is most easily interpreted for indicator variables. The risk ratio is the ratio of the estimated hazard for subjects with a value of 1 to the estimated hazard for those with a value of 0, controlling for all other covariates. So for our example, NEWTX has a risk ratio of 2.864 which can be interpreted that subjects on the experimental therapy have an estimated hazard of being cured is 2.864 that of those receiving the standard therapy. This indicates that not only is therapy a significant variable, but that patients on the new therapy are cured faster. For quantitative covariates, such as age, risk ratio refers to a change of one unit (or one year).

But PHREG is much more powerful than this example shows. We can now examine other covariates by adding them to the *model* statement. The analysis would then control for all potential covariates. For example, let's include aspirin use at start of therapy and gender. Again, let's look at OUTPUT #2.

As can be seen by the significance tests, ASPIRIN is not significant and MALE is. Notice that the parameter estimate for MALE is negative and the risk ratio is less than one. This indicates that the survival curve is longer for males than females. Remember that in our case, we want to cure them as soon as possible so we want a survival function that dies out quickly. Notice that the risk ratio for NEWTX has increased! This means that the treatment is even more therapeutic once you control for gender and aspirin use.

What else can PHREG do? Phreg can provide stepwise model selection, allows time dependent covariates, test linear hypotheses about the regression coefficients, create an output dataset, and create a dataset that contains the survivor function estimates at the event times.

It is not surprising that Cox regression is the preferred method for survival data analysis. It has no assumptions about the distribution of survival times and it handles continuous and discrete times, time dependent covariates, left censoring. The major drawback is testing the hazard function.

Notes and Comments:

I cannot possibly cover all the details of survival analysis possible using the SAS System. I have skipped over the details of selecting hazard functions and the statistical rigor behind the analyses. Please refer to the appropriate SAS manual for more details. Other procedures can be used as well: LOGISTIC, PROBIT, and GENMOD.

SAS Online Samples: There are many samples of code, datasets, and macros for survival analysis available. SAS

online samples enables you to download sample programs and datasets from many of the SAS books by using either anonymous FTP, SASDOC-L, or the World Wide Web. Details are available from SAS.

References:

Allison, Paul D., *Survival Analysis Using the SAS System: A Practical Guide*, Cary, NC: SAS Institute Inc., 1995.

SAS Institute, Inc., *SAS Technical Report P-217, SAS/STAT Software: The PHREG Procedure, Version 6*, Cary, NC; SAS Institute Inc., 1991.

SAS Institute, Inc., *SAS/STAT Software: Changes and Enhancements through Release 6.11*, Cary, NC; SAS Institute Inc., 1996.

Product Information:

SAS is a registered trademark of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Author Contact:

Linda Quinn
Independent Statistical Consultant
1286 Gary Blvd.
Brunswick, OH 44212-2912
Phone/Fax: 330-225-4184
e-mail: lmq2@po.cwru.edu

The LIFETEST Procedure

Product-Limit Survival Estimates

GROUP = E

DAYS	Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.0000	1.0000	0	0	0	12
1.0000	0.9167	0.0833	0.0798	1	11
2.0000*	.	.	.	1	10
3.0000*	.	.	.	1	9
4.0000	.	.	.	2	8
4.0000	.	.	.	3	7
4.0000	0.6111	0.3889	0.1535	4	6
7.0000	0.5093	0.4907	0.1582	5	5
8.0000	0.4074	0.5926	0.1559	6	4
9.0000	0.3056	0.6944	0.1465	7	3
10.0000	0.2037	0.7963	0.1283	8	2
11.0000	0.1019	0.8981	0.0964	9	1
30.0000	0	1.0000	0	10	0

* Censored Observation

Summary Statistics for Time Variable DAYS

Quantile	Point Estimate	95% Confidence Interval [Lower, Upper)	
75%	10.0000	7.0000	30.0000
50%	8.0000	4.0000	10.0000
25%	4.0000	4.0000	8.0000
Mean	8.9444	Standard Error	2.5589

OUTPUT #1 (LIFETEST)

The LIFETEST Procedure

Product-Limit Survival Estimates

GROUP = S

DAYS	Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.0000	1.0000	0	0	0	12
5.0000	0.9167	0.0833	0.0798	1	11
12.0000	.	.	.	2	10
12.0000	0.7500	0.2500	0.1250	3	9
13.0000	.	.	.	4	8
13.0000	0.5833	0.4167	0.1423	5	7
14.0000	0.5000	0.5000	0.1443	6	6
17.0000	0.4167	0.5833	0.1423	7	5
18.0000*	.	.	.	7	4
19.0000	0.3125	0.6875	0.1398	8	3
21.0000	0.2083	0.7917	0.1262	9	2
23.0000*	.	.	.	9	1
25.0000	0	1.0000	0	10	0

* Censored Observation

Summary Statistics for Time Variable DAYS

Quantile	Point Estimate	95% Confidence Interval [Lower, Upper]	
75%	21.0000	14.0000	25.0000
50%	15.5000	13.0000	21.0000
25%	12.5000	12.0000	17.0000
Mean	16.5417	Standard Error	1.8128

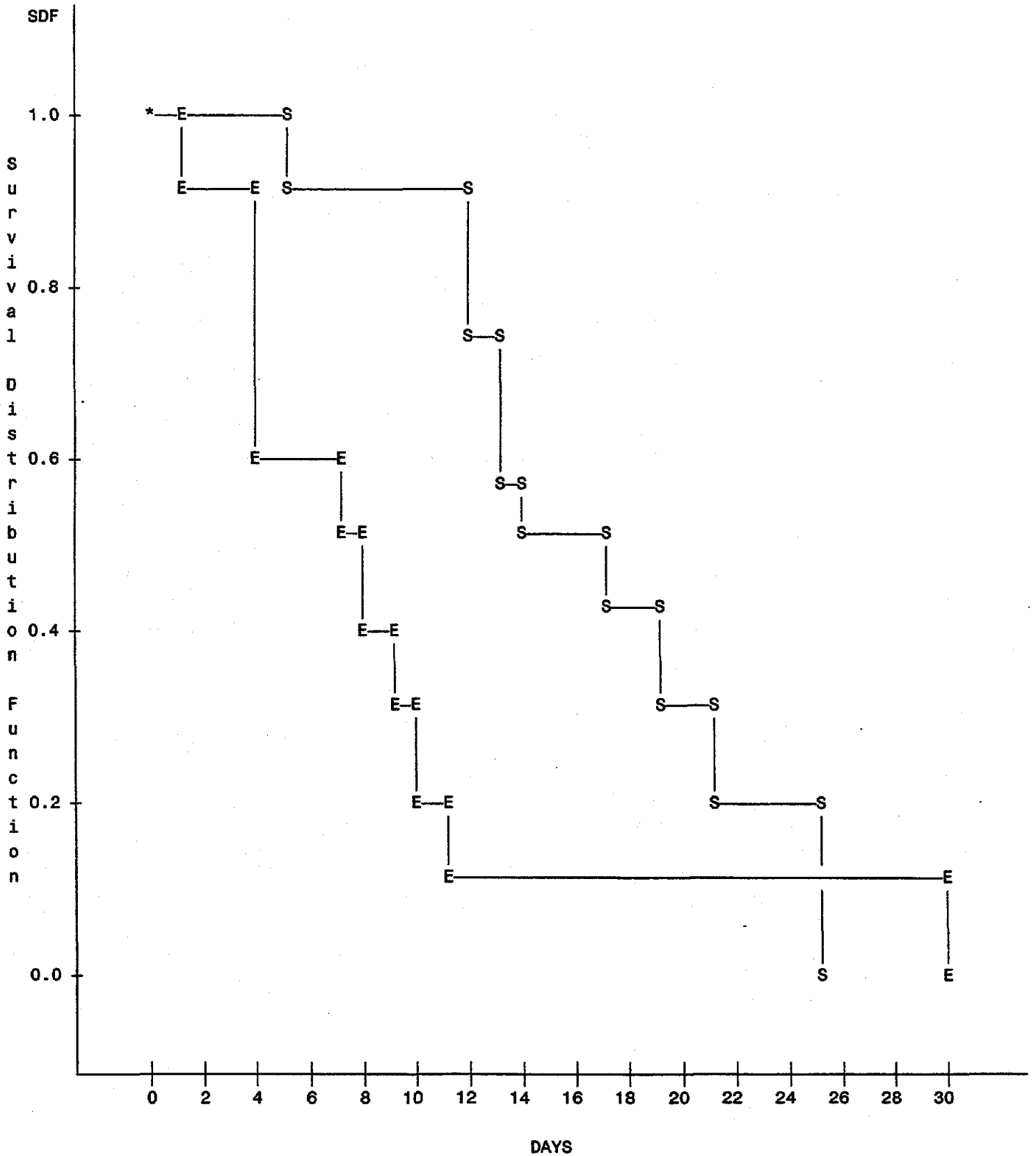
Summary of the Number of Censored and Uncensored Values

GROUP	Total	Failed	Censored	%Censored
E	12	10	2	16.6667
S	12	10	2	16.6667
Total	24	20	4	16.6667

OUTPUT #1 (LIFETEST)

The LIFETEST Procedure

Survival Function Estimates



OUTPUT #1 (LIFETEST)

The LIFETEST Procedure

Testing Homogeneity of Survival Curves over Strata
Time Variable DAYS

Rank Statistics

GROUP	Log-Rank	Wilcoxon
E	3.9962	88.000
S	-3.9962	-88.000

Covariance Matrix for the Log-Rank Statistics

GROUP	E	S
E	3.19898	-3.19898
S	-3.19898	3.19898

Covariance Matrix for the Wilcoxon Statistics

GROUP	E	S
E	784.600	-784.600
S	-784.600	784.600

Test of Equality over Strata

Test	Chi-Square	DF	Pr >
			Chi-Square
Log-Rank	4.9921	1	0.0255
Wilcoxon	9.8700	1	0.0017
-2Log(LR)	2.5718	1	0.1088

The PHREG Procedure

Data Set: WORK.TDATA
 Dependent Variable: DAYS
 Censoring Variable: STATUS
 Censoring Value(s): 0
 Ties Handling: BRESLOW

Summary of the Number of
 Event and Censored Values

Total	Event	Censored	Percent Censored
24	20	4	16.67

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	92.018	87.665	4.353 with 1 DF (p=0.0370)
Score	.	.	4.830 with 1 DF (p=0.0280)
Wald	.	.	4.532 with 1 DF (p=0.0333)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
NEWTX	1	1.052296	0.49430	4.53208	0.0333	2.864

The PHREG Procedure

Data Set: WORK.TDATA
 Dependent Variable: DAYS
 Censoring Variable: STATUS
 Censoring Value(s): 0
 Ties Handling: BRESLOW

Summary of the Number of
 Event and Censored Values

Total	Event	Censored	Percent Censored
24	20	4	16.67

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	92.018	81.410	10.608 with 3 DF (p=0.0140)
Score	.	.	10.293 with 3 DF (p=0.0162)
Wald	.	.	8.836 with 3 DF (p=0.0315)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
NEWTX	1	1.654288	0.63062	6.88153	0.0087	5.229
ASPIRIN	1	-0.293290	0.52641	0.31042	0.5774	0.746
MALE	1	-1.267995	0.54041	5.50531	0.0190	0.281