

Survival Analysis: Left-Truncated Data

Introduction: The random variable of most interest in survival analysis is time-to-event. Often in biomedical studies, the event is death. It is because of this common application the field is termed *Survival Analysis*. Survival analysis methods can be applied to a wide range of data not just biomedical survival data. Other time-to-event data can include: time-to-relapse of a disease, length of stay in a hospital, duration of a strike, money paid by insurance company, time-to-employment (Klein), failure times of electrical components, etc. (Survival).

There is generally a relatively short window for the duration of any study where the event is observed to occur or not occur. When the duration of study is small, adjustments must be made to account for potential biases. Useful data will be excluded when data is censored but not accounted for, and biases can be introduced when data is truncated. Since censoring and truncation are often confused, a brief discussion on censoring with examples is helpful to more fully understand left-truncation.

There are three general types of censoring, right-censoring, left-censoring, and interval-censoring. The most common type of censoring encountered in survival analysis data is right censored (Survival). In our course, we adjusted our model for the herpes data to account for right censoring. It is called right censoring because the true unobserved event occurs after the censoring time. Left-censoring occurs when we cannot observe the time the event occurred, but we know it occurred before study began. For obvious reasons if the event is death, the data can't be left-censored. A good example is discussed in an ASA paper on survival analysis, "e.g. [a] study of age at which African children learn a task. Some already knew (left-censored), some learned during a study (exact), some had not yet learned by end of study (right-censored)." Interval-censoring is also discussed in *Survival Analysis: Introduction* (Survival) and a link to the book is provided.

Truncation is due to sampling bias that only those individuals whose lifetimes lie within a certain interval can be observed. To further discuss left-truncation we will use the main example. A useful insight and directly applicable to the main example that will be discussed comes from John P. Klein and Melvin L. Moeschberger's book *Survival Analysis: Techniques for Censored and Truncated Data*, "Age is often used as a covariate when it should be used as a left-truncation point. When age is used as a left-truncation point, it is unnecessary to use it as a covariate in the model" (Klein).

Motivating Example: Survival data on 26 psychiatric inpatients admitted to the University of Iowa hospitals during the years 1935–1948. This sample is part of a larger study of psychiatric inpatients discussed by Tsuang and Woolson (1977). Data for each patient consists of age at first admission to the hospital, sex, number of years of follow-up (years from admission to death or censoring) and patient status at the follow-up time. Ultimately we want to know if male psychiatric patients and female psychiatric patients are more likely to die than the general public. The data from the hospitals are displayed in the table below (Klein).

Obs	gender	age	time	death
1	1	51	1	1
2	1	58	1	1
3	1	55	2	1
4	1	28	22	1
5	0	21	30	0
6	0	19	28	1
7	1	25	32	1
8	1	48	11	1
9	1	47	14	1
10	1	25	36	0
11	1	31	31	0
12	0	24	33	0
13	0	25	33	0
14	1	30	37	0
15	1	33	35	0
16	0	36	25	1
17	0	30	31	0
18	0	41	22	1
19	1	43	26	1
20	1	45	24	1
21	1	35	35	0
22	0	29	34	0
23	0	35	30	0
24	0	32	35	1
25	1	36	40	1
26	0	32	39	0

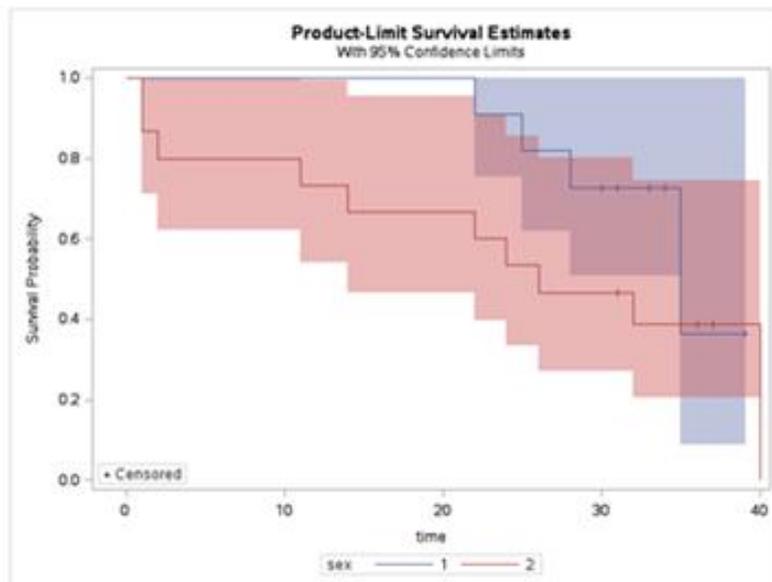
The way we coded the variable death, death(0) indicates censoring time. It is important to adjust for left-truncation in this data because patients entered into the study (admitted to the hospital) have not

died. This fact could bias the results, when results are compared to the general population of Iowa residents.

To begin, we show KM curves using PROC LIFETEST not accounting for left-truncation. We stratified by sex because many studies show women live longer than men so we want to compare by gender.

Figure 1 is obtained by the SAS code:

```
proc lifetest data=psych method=km plots=(survival(cl));
  strata gender;
  time time*death(0); /* time var is years from admission to death or
  censoringcensoring indicated when death=0 */
  survival out=out1 conftype=log;
run;
```



Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	1.6074	1	0.2049
Wilcoxon	2.8418	1	0.0918
-2Log(LR)	2.5358	1	0.1113

Figure 1

The test of equality over strata for Log-Rank yields p-value= 0.2049. PROC LIFETEST is the traditional SAS code to plot survival codes, but it is limiting in that it cannot accommodate left-truncated data. In order to account for left truncation, we need to utilize PROC PHREG.

We will begin by illustrating that essentially the same KM curves produced in Figure 1 can also be done in PROC PHREG which assumes the Cox proportional hazards model.

Figure 2 illustrates the same KM curve as produced in Figure 1 only without overlaying plots for male and female patients using the code:

```
proc phreg data=psych plots(c1 overlay=stratum)=survival;
  strata gender;
  model time*death(0)=;
  output out=psych01 survival=S;
run;
```

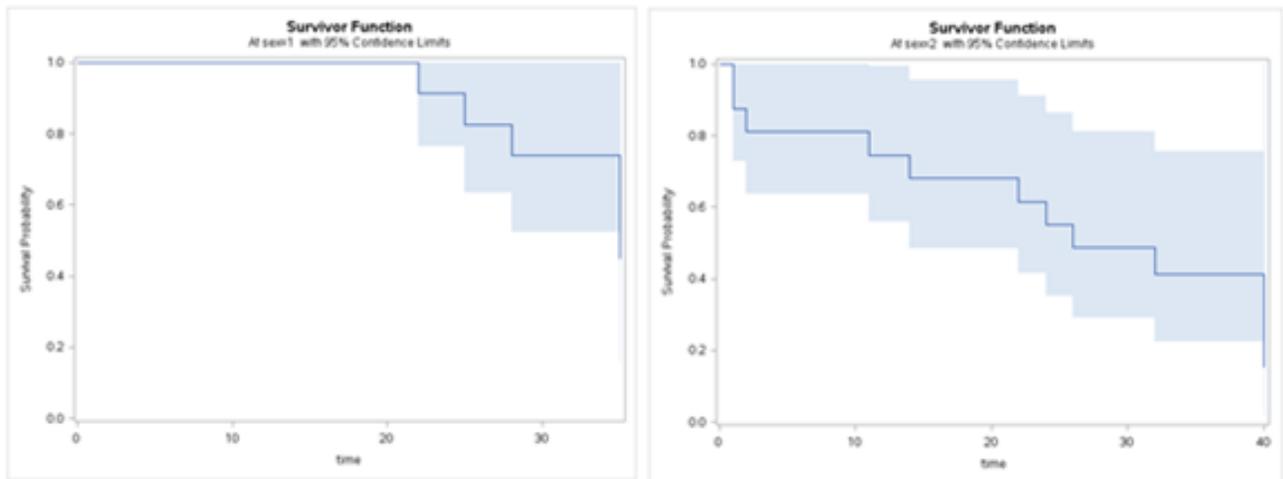


Figure 2

To obtain a p-value and proportional hazards estimates for comparison between strata (sex of patient) one can just delete the `strata sex;` statement. Figure 3 shows the results.

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
gender	1	0.74555	0.60560	1.5156	0.2183	2.108

Figure 3

We see a very similar p-value to that of the log-rank test in Figure 1 by using the PROC LIFETEST procedure.

Now accounting for left-truncation we essentially use the same code that produced Figure 2 with the addition of `/entry=age;` to the model statement. We must first define the left-truncation time using a data step.

The following code produced Figure 4:

```
data temp; set psych;
    agetime=age+time;
run;
proc phreg data=temp plots(cl overlay=stratum)=survival;
    strata gender;
    model agetime*death(0)=/entry=age;
    output out=psych02 survival=S;
run;
```

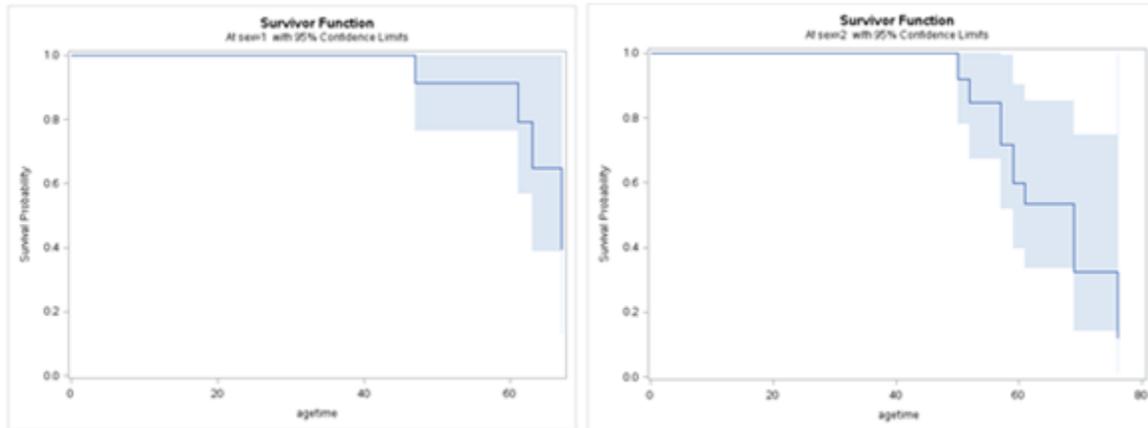


Figure 4

Figures 4 and Figures 2 can be used to illustrate the effect accounting for left-truncation has on our example.

The R statistical software can produce KM curves that account for left-truncation as well and can be done for this example using the following code:

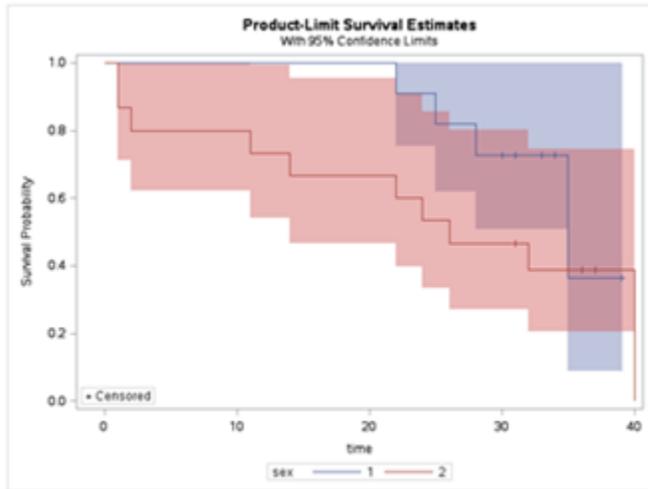
```
install.packages('survival')
install.packages('KMsurv')
```

```
library(survival)
library(KMsurv)
```

```
data(psych); attach(psych)
my.surv.object <- Surv(age, age+time, death)
my.surv.object
survfit(my.surv.object~sex)
my.fit <- survfit(my.surv.object~sex)
```

```
plot(my.fit, main="Kaplan-Meier estimate with 95% confidence bounds", xlab="time", ylab="survival function")
```

Figure 5 are the results obtain from the above R code which accounts for left-truncation. In a comparison with Figure 4 we can see Figure 5 is in effect the same plot only that plots for male and female patients are overlaid whereas in Figure 4 they are not. The Km Plot from Figure 1 and Figure 5 can be directly and easily compared to graphically display the difference not accounting for left-truncation and accounting for it.



KM Plot from Figure 1

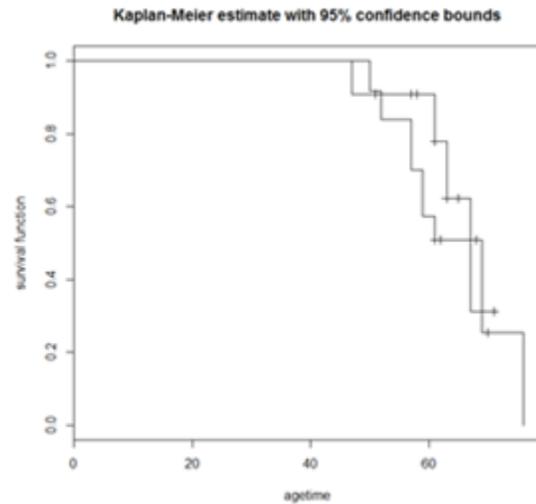


Figure 5

Figure 6 is the proportional hazards estimates and the significant values when left-truncation is taken into account.

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
gender	1	0.36153	0.61157	0.3495	0.5544	1.436

Figure 6

We can see that the p-value is less significant and also the Hazards Ratio is less than what was obtained when left-truncation was not taken into account in Figure 3. Figure 6 was obtained by the following code:

```
proc phreg data=temp plots(cl overlay=stratum)=survival;
  model agetime*death(0)= gender/entry=age;
  output out=psych02 survival=S;
run;
```

SAS Appendix:

```
data psych;
  input gender age time death; /*gender: 0=male, 1=female death: 0=alive,
  1=death*/
  cards;
  1 51 1 1
  1 58 1 1
  1 55 2 1
  .
  .
  .
  1 36 40 1
  0 32 39 0
  ;
run;

proc print data=psych;
run;

proc lifetest data=psych method=km plots=(survival(c1));
  strata gender;
  time time*death(0); /* time var is years from admission to death or
  censoring censoring indicated when death=0 */
  survival out=out1 conftype=log;
run;

proc phreg data=psych plots(c1 overlay=stratum)=survival;
  strata gender;
  model time*death(0)=;
  output out=psych01 survival=S;
run;

proc phreg data=psych plots(c1 overlay=stratum)=survival;
  model time*death(0)= gender;
  output out=psych01 survival=S;
run;

proc print data=psych01;
run;

data temp; set psych;
  agetime=age+time;
run;

proc phreg data=temp plots(c1 overlay=stratum)=survival;
  strata gender;
  model agetime*death(0)=/entry=age;
  output out=psych02 survival=S;
run;

proc phreg data=temp plots(c1 overlay=stratum)=survival;
  model agetime*death(0)= gender/entry=age;
  output out=psych02 survival=S;
run;
```

References:

Surviving Left Truncation Using PROC PHREG

Aimee J. Foreman, Ginny P. Lai, Dave P. Miller,
ICON Clinical Research, San Francisco, CA

<http://www.wuss.org/proceedings08/08WUSS%20Proceedings/papers/anl/anl03.pdf>

Survival Analysis in R

David Diez

<http://stat.ethz.ch/education/semesters/ss2011/seminar/diez.pdf>

The Basic of Survival Analysis

www.stat.nus.edu.sg/~stachenz%2FST3242Notes1.pdf&ei=sFqeUrm9EInTqgHEwoDQDg&usg=AFQjCNG9xBEUwNzBPhit2rxcdpehosjx4w&sig2=6fdDO0e2ey4kGSUzUVobrg

Survival Analysis: Introduction – American Statistical Association

www.amstat.org%2Fchapters%2Fnortheasternillinois%2Fpastevents%2Fpresentations%2Fsummer05_lb rahim_J.pdf&ei=TIqeUsrSF82grgH2q4AQ&usg=AFQjCNEdeHLLNgy6ISlyCSXdiurzHexzpg&sig2=RA5v5VTuLYRObayH1n8QhQ

Survival Analysis: Techniques for Censored and Truncated Data

John P. Klein and Melvin L. Moeschberger

[http://f3.tiera.ru/2/M_Mathematics/MV_Probability/MVsa_Statistics%20and%20applications/Klein%20J.,%20Moeschberger%20M.%20Survival%20analysis..%20Techniques%20for%20censored%20and%20truncated%20data%20\(2ed.,%20Springer,%202003\)\(ISBN%20038795399X\)\(542s\)_MVsa_.pdf](http://f3.tiera.ru/2/M_Mathematics/MV_Probability/MVsa_Statistics%20and%20applications/Klein%20J.,%20Moeschberger%20M.%20Survival%20analysis..%20Techniques%20for%20censored%20and%20truncated%20data%20(2ed.,%20Springer,%202003)(ISBN%20038795399X)(542s)_MVsa_.pdf)