PharmaSUG China 2021 - Paper CC-026

Surviving Left Truncation: Number at Risk Calculation for a Specific Timepoint Using SAS

Rolland LUO and Lihua WU, Sanofi

ABSTRACT

As we know, PROC LIFETEST is a tool for producing survival function estimates and PROC PHREG can be used for this purpose as well. From existing publications, PROC PHREG is especially convenient in the case of left truncated data using the ENTRY= option to specify the left truncation time while PROC LIFETEST is not set up to handle this situation. Left-truncation occurs when individuals are not observed at the natural time origin of the phenomenon under study but come under observation at some known later time (called the left-truncation time). PROC PHREG for left truncated data produces the number at risk value only for timepoints with event occurred but not for all timepoints. While for study purpose, number at risk for a specific timepoint not only for event occurred point is needed. Naturally an expanded calculation based on PROC PHREG is needed. In order to meet such requirement, we will provide an example to show how to calculate number at risk for a specific timepoint based on the definition of number at risk and the nature of left-truncation and compare the accuracy between the result of PROC PHREG and our expanded calculation in this paper.

INTRODUCTION

Left-truncation occurs when individuals are not observed at the natural time origin of the phenomenon under study but come under observation at some known later time (called the left-truncation time). The risk set just prior to an event time does not include individuals whose left-truncation times exceed the given event time.

In this paper, the example vaccine study is a randomized, observer-blind, placebo-controlled, multi-center, Phase III trial in 10,278 subjects. Subjects will receive 3 vaccinations (Day 0, 6 months, 12 months), with an efficacy follow-up of 13 months after Dose 3 and follow-up for disease cases up to 60 months after Dose 3. Active phase of disease case detection will begin from Dose 1 and is expected to continue until 13 months after Dose 3. Hospital Phase will begin after the Active Phase. Subjects with a febrile illness and requiring hospitalization will be screened for disease. This phase will continue until consent to participate in the Surveillance Expansion period (SEP) is given or otherwise, until trial completion. For Surveillance Expansion period, subjects consenting to take part into the Surveillance Expansion period will be actively followed for disease cases detection similarly as in the Active Phase. Thus, surveillance will be designed to maximize the detection of symptomatic confirmed disease (hospitalized or not) in order to describe vaccine efficacy and safety in preventing symptomatic disease. Subjects having declined their participation will continue surveillance as in the Hospital Phase until trial completion (up to 60 months post-dose 3).

A basic understanding of survival analysis is assumed in this paper, as is familiarity with SAS/STAT® software. Number at risk calculation for simple survival function estimates using PROC LIFETEST will be mentioned, and number at risk calculation using PROC PHREG for left truncation data will be mentioned also. An expanded calculation of number at risk based on PROC PHREG will be illustrated in the final section.

NUMBER AT RISK CALCULATION

NUMBER AT RISK DEFINITION

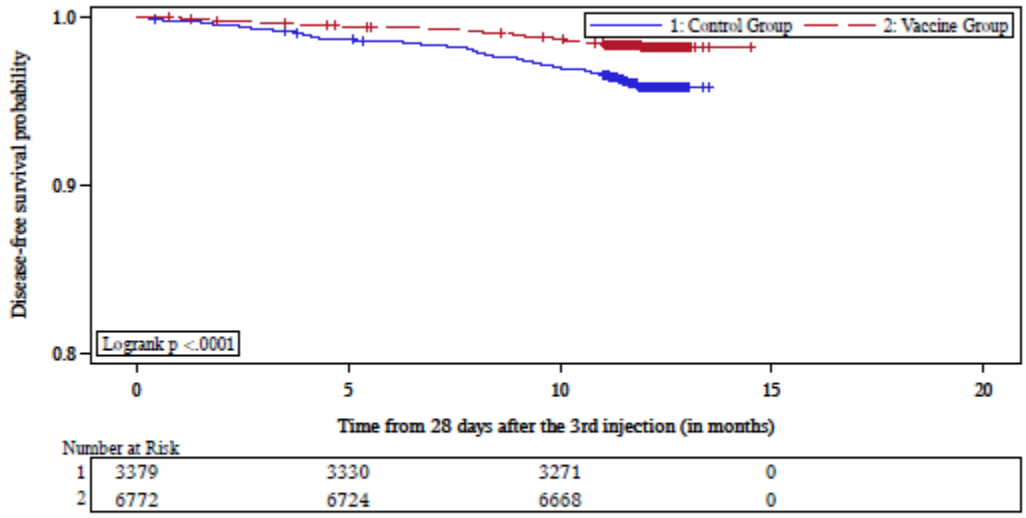
Number at risk is the number of subjects at risk immediately before the time point, t. Being "at risk" means that the subject has not had an event before time t, and is not censored before or at time t.

NUMBER AT RISK CALCULATION USING PROC LIFETEST

For simple survival function estimates, in the example study, we want to produce Figure 1 'Kaplan-Meier curve for symptomatic virologically-confirmed disease post-dose 3 due to any serotype - modified Full Analysis Set for Efficacy', all subjects were under observation at the beginning of the study. We can use below PROC LIFETEST code to produce the figure and the number at risk with ATRISK= option specified.

```
proc lifetest data=fortotal plots=s ( test atrisk =0 to 20 by 5)
    method=km;
    time pyr*censor(0);
    strata trtpn;
    format trtpn trtf.;
    ods output homtests=test SurvivalPlot=plot;
run;
```

Figure 1 : Kaplan-Meier curve for symptomatic virologically-confirmed disease post-dose 3 due to any serotype - modified Full Analysis Set for Efficacy



Cases: number of subjects with at least one symptomatic virologically-confirmed disease episode from 28 days post-injection 3 to the end of Active Phase. The p-value from the log-rank test comparing the 2 vaccine groups is presented.

Figure 1. PROC LIFETEST Simple Survival Function Estimates Figure

NUMBER AT RISK CALCULATION USING PROC PHREG

For left truncation data, in the example study, we want to produce Figure 2 'Kaplan-Meier curve for symptomatic virologically-confirmed disease during the SEP due to any serotype considering time from D0 - Full Analysis Set for SEP', all subjects were under observation at the beginning of the study but the time participating in SEP varied, and the figure considered time from D0 (i.e. beginning of study) with left truncation data of time at start of SEP. We can use below PROC PHREG code to produce the data the figure needed and the number at risk with ATRISK option specified.

```
proc phreg data=final atrisk;
    strata trtpn;
    model time*censor(0) = trtpn / entry=time_d0_sep ;
```

```
    output out=temp01 survival=survival atrisk=tatrisk;
run;
```

In above code, variable time is time from D0; time_d0_sep is the left-truncation time, from D0 to time at start of SEP; variable tatrisk in output data set Table 1 temp01 is the number at risk PROC PHREG produced.

USUBJID	TRTPN	censor	time	time_D0_SEP	tatrisk	Survival
0001	1	1	47.080082136	46.554414784	455	0.9978046112
0002	1	1	47.145790554	46.455852156	578	0.9960797981
0003	1	1	47.967145791	47.277207392	2303	0.995647378
0004	1	1	48	47.310061602	2336	0.9952212503
0005	1	1	48.197125257	47.244353183	2596	0.9948379569
0006	1	1	48.394250513	46.751540041	2750	0.9937532709
0007	1	1	48.394250513	47.572895277	2750	0.9937532709
0008	1	1	48.394250513	47.770020534	2750	0.9937532709
0009	1	1	48.492813142	47.540041068	2775	0.9933952261
0010	1	1	48.624229979	47.540041068	2808	0.9930415159
0011	1	1	48.887063655	47.540041068	2840	0.9926919149
0012	1	1	49.084188912	47.638603696	2845	0.9923430512
0013	1	1	49.215605749	47.967145791	2845	0.9916456913
0014	1	1	49.215605749	48.427104723	2845	0.9916456913
0015	1	1	49.248459959	47.342915811	2844	0.9912970727
0016	1	1	49.281314168	47.572895277	2843	0.9909484541
0017	1	1	49.314168378	47.277207392	2846	0.9906003254
0018	1	1	49.642710472	47.244353183	3104	0.9902812401
0019	1	1	49.938398357	47.112936345	3350	0.9899856778
0020	1	1	50.266940452	47.540041068	3746	0.989721435
0021	1	1	50.431211499	47.080082136	3837	0.9894635272
0022	1	1	50.464065708	47.835728953	3848	0.9892064235
0023	1	1	50.496919918	47.310061602	3858	0.9889500528
0024	1	1	50.924024641	47.474332649	3996	0.9887025987
0025	1	1	51.318275154	50.069815195	4078	0.9884601805
0026	1	1	51.351129363	48.427104723	4091	0.9882185918
0027	1	0	51.482546201	47.638603696	.	0.9882185918
0028	1	1	51.58110883	48.361396304	4168	0.9879815237
0029	1	1	52.13963039	46.390143737	4425	0.9877582763
0030	1	1	52.205338809	51.351129363	4481	0.9875378683
0031	1	1	52.468172485	48	4820	0.9873330062
0032	1	1	52.533880903	48.624229979	4904	0.9871316945
0033	1	1	52.960985626	47.310061602	5378	0.9869481616
0034	1	1	53.125256674	47.310061602	5560	0.9867706689
0035	1	1	53.420944559	51.351129363	5712	0.9865979301
0036	1	1	53.453798768	50.069815195	5717	0.9864253725
0037	1	1	53.552361396	47.605749487	5739	0.9862535064
0038	1	1	53.618069815	52.13963039	5755	0.986082148
0039	1	1	53.848049281	47.770020534	5786	0.9859117372

Table 1. PHREG Procedure Output Data Set temp01 First 39 Observations.

We can know from the output that PROC PHREG only produced the number at risk value for timepoints with event occurred, and PROC PHREG does not allow to specify timepoints for the number at risk calculation in the same way as PROC LIFETEST. Thus we could not tell from PROC PHREG the number at risk for a specific timepoint.

EXPANDED CALCULATION OF NUMBER AT RISK BASED ON PROC PHREG

In order to produce the number at risk for a specific timepoint and display as what shows in Figure 2, we think more about the data. We can know from the data that there were subjects come under observations from time to time (i.e. time_d0_sep variable shows) and there were also subjects censored or had event from time to time (i.e. time variable shows), which is a complex situation comparing to Figure 1 situation that all subjects were under observations at the beginning. When coming back to the definition of number at risk and combining it with the nature of left truncation, we figured out that, being "at risk" for left-truncation means that the subject has come under observation before time t and has not had an event before time t, in addition, the subject is not censored before or at time t. If we can calculate the number of subjects under observation before time t (n_obs) and number of subjects with an event or censored before time t (n_out), we can easily tell n_obs minus n_out is number at risk before time t.

We want to mention that below codes are just examples and may not be the best solution in SAS and it is above logic that matters rather than the programming codes. Based on the approach, we can do in below steps.

1. Create dummy time 0 to improve the look of the figure, and create censored variable for scatterplot.

```
/* include estimates at time 0 to improve the look of the figure for both
strata */
proc sort data=temp01 out=temp01_s nodupkey;
    by trtpn time censor tatrisk;
run;
data all;
    set temp01_s;
    by trtpn;
    output;
    if first.trtpn then do;
          time=0;
          survival=1;
          output;
    end;
run;
data all_01;
    set all;
    if censor=0 then censored=survival;
    if censor=1 then censored=.;
run;
```

2. Create dummy timepoints where we want to display the number at risk.

```
**add atrisk 0-75 by 75;
data dummy_nbrisk;
    do time=45 to 75 by 5;
          do trtpn=1 to 2;
                at_time=time;
                output;
          end;
    end;
run;
```

3. Calculate number of subjects entered in SEP before each time_d0_sep.

```
proc sort data = temp01 out=temp01_x;
    by trtpn time_d0_sep;
run;
data temp01_x_1;
    retain n_obs 0;
    set temp01_x;
    by trtpn time_d0_sep;
    if first.trtpn then n_obs=0;
    n_obs=n_obs+1;
    if last.time_d0_sep then output;
run;
```

4. Calculate number of subjects censored or had event before each time.

```
** calc number of subjects censored or had event before each time;
proc sort data = temp01 out=temp01_x_out;
    by trtpn time;
run;
data temp01_x_out1;
    retain n_out 0;
    set temp01_x_out;
    by trtpn time;
    if first.trtpn then n_out=0;
    n_out=n_out+1;
    if last.time then output;
run;
```

5. Merge n_obs and n_out back to dummy timepoints data set to know the n_obs and n_out before each dummy timepoint.

```
proc sql noprint;
    create table temp01_x_2 as
          select distinct a.*,max(b.n_obs) as n_obs,max(c.n_out) as n_out
          from dummy_nbrisk as a
          left join temp01_x_out1 as b
          on a.trtpn=b.trtpn and a.time>b.time
          left join temp01_x_1 as c
          on a.trtpn=c.trtpn and a.time>c.time_d0_sep
          group by a.trtpn,a.time
          order by a.trtpn, a.time
    ;
quit;
```

6. Number at risk (tatrisk in Table 2) is n_obs minus n_out.

```
data all_02;
    set all_01(drop=tatrisk) temp01_x_2;
    if missing(n_out) then n_out=0;
    if cmiss(n_out,n_obs)=0 then tatrisk= n_obs-n_out;
    if missing(n_obs) and ^missing(at_time) then tatrisk=0;
run;
```

TRTPN	censor	time	time_D0_SEP	survival	censored	at_time	n_obs	n_out	tatrisk
1	.	45	.	.	.	45	.	0	0
1	.	50	.	.	.	50	3451	19	3432
1	.	55	.	.	.	55	6136	46	6090
1	.	60	.	.	.	60	6325	123	6202
1	.	65	.	.	.	65	6327	214	6113
1	.	70	.	.	.	70	6327	249	6078
1	.	75	.	.	.	75	6327	6327	0
2	.	45	.	.	.	45	.	0	0
2	.	50	.	.	.	50	1707	7	1700
2	.	55	.	.	.	55	3037	36	3001
2	.	60	.	.	.	60	3137	82	3055
2	.	65	.	.	.	65	3138	133	3005
2	.	70	.	.	.	70	3138	165	2973
2	.	75	.	.	.	75	3138	3138	0

Table 2. Expanded Calculation of Number at Risk Based on Dummy Timepoints.

If we make the dummy timepoint data set same as the PROC PHREG output timepoints for number at risk, we can easily know what we calculated matches what PROC PHREG produced, see Table 3

TRTPN	time	at_time	tatrisk_phreg	n_obs	n_out	tatrisk_expand_calc
1	47.080082136	47.080082136	455	455	0	455
1	47.145790554	47.145790554	578	579	1	578
1	47.967145791	47.967145791	2303	2305	2	2303
1	48	48	2336	2339	3	2336
1	48.197125257	48.197125257	2596	2600	4	2596
1	48.394250513	48.394250513	2750	2755	5	2750
1	48.492813142	48.492813142	2775	2783	8	2775
1	48.624229979	48.624229979	2808	2817	9	2808
1	48.887063655	48.887063655	2840	2850	10	2840
1	49.084188912	49.084188912	2845	2856	11	2845
1	49.215605749	49.215605749	2845	2857	12	2845
1	49.248459959	49.248459959	2844	2858	14	2844
1	49.281314168	49.281314168	2843	2858	15	2843
1	49.314168378	49.314168378	2846	2862	16	2846
1	49.642710472	49.642710472	3104	3121	17	3104
1	49.938398357	49.938398357	3350	3368	18	3350
1	50.266940452	50.266940452	3746	3765	19	3746
1	50.431211499	50.431211499	3837	3857	20	3837
1	50.464065708	50.464065708	3848	3869	21	3848
1	50.496919918	50.496919918	3858	3880	22	3858
1	50.924024641	50.924024641	3996	4019	23	3996
1	51.318275154	51.318275154	4078	4102	24	4078
1	51.351129363	51.351129363	4091	4116	25	4091
1	51.58110883	51.58110883	4168	4195	27	4168
1	52.13963039	52.13963039	4425	4453	28	4425
1	52.205338809	52.205338809	4481	4510	29	4481
1	52.468172485	52.468172485	4820	4850	30	4820
1	52.533880903	52.533880903	4904	4935	31	4904
1	52.960985626	52.960985626	5378	5410	32	5378
1	53.125256674	53.125256674	5560	5593	33	5560
1	53.420944559	53.420944559	5712	5746	34	5712
1	53.453798768	53.453798768	5717	5752	35	5717
1	53.552361396	53.552361396	5739	5775	36	5739
1	53.618069815	53.618069815	5755	5792	37	5755
1	53.848049281	53.848049281	5786	5824	38	5786
1	53.9137577	53.9137577	5793	5832	39	5793
1	54.176591376	54.176591376	5830	5870	40	5830
1	54.209445585	54.209445585	5833	5874	41	5833
1	54.308008214	54.308008214	5852	5894	42	5852

Table 3. PHREG Procedure Produced Results (Variable tatrisk_phreg) Comparing to Expanded Calculation Results (Variable tatrisk_expand_calc) First 39 Observations.

Finally, according to what we calculated we can display the number at risk in Figure 2.

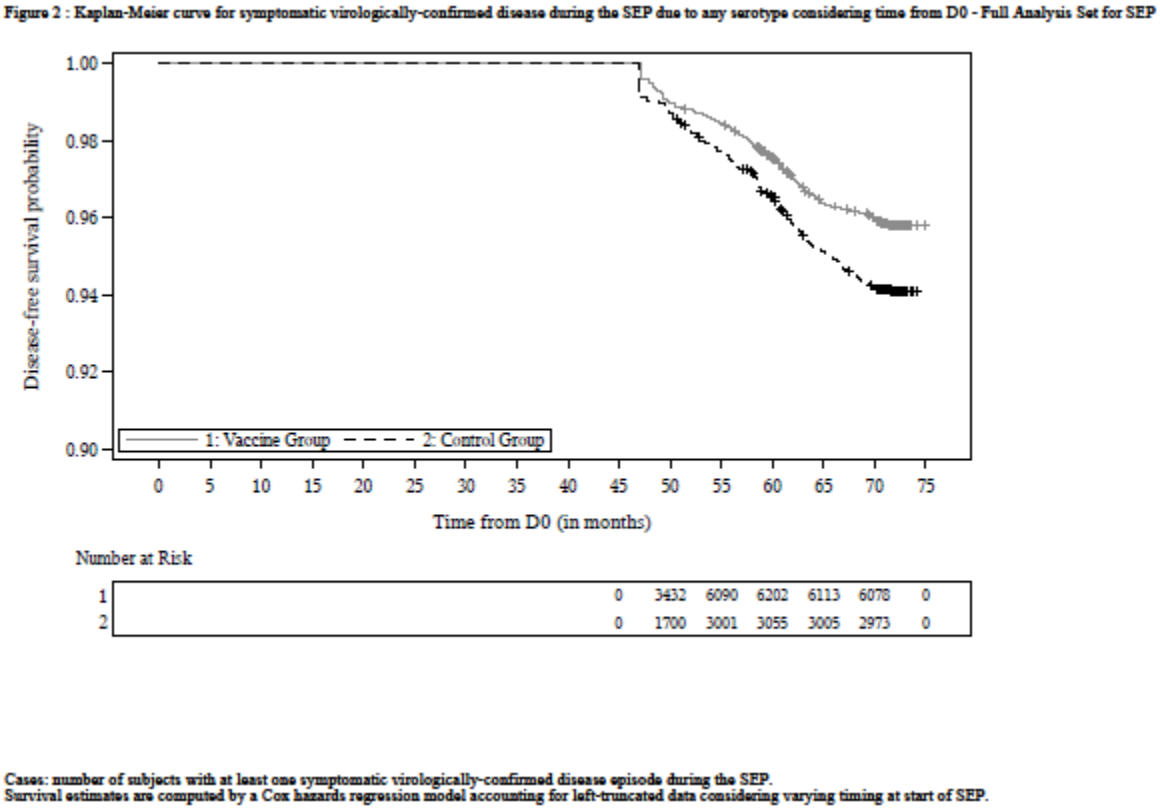


Figure 2. PROC PHREG Left Truncation Figure

CONCLUSION

Survival function estimates for left truncated data cannot be obtained from PROC LIFETEST but they are output using PROC PHREG with the ENTRY= option to specify the left truncation time. PROC PHREG for left truncated data produces the number at risk value only for timepoints with event occurred but not for all timepoints. As Left-truncation occurs when individuals are not observed at the natural time origin of the phenomenon under study but come under observation at some known later time (called the left-truncation time) and number at risk is the number of subjects at risk immediately before the time t, being "at risk" for left-truncation naturally means that the subject has come under observation before time t and has not had an event before time t, in addition, the subject is not censored before or at time t. According to the definition of number at risk and the nature of left-truncation, we can calculate the number of subjects under observation before time t (n_obs) and number of subjects with an event or censored before time t (n_out), thus number at risk before time t is n_obs minus n_out. The example shows that the number at risk PROC PHREG produced matches our expanded calculation based on PROC PHREG, furthermore our expanded calculation can provide the number at risk for a specific timepoint.

REFERENCES

Foreman AJ, Lai GP, Miller DP. Surviving Left Truncation Using PROC PHREG. In: Proceedings of the 2008 Western Users of SAS Software Conference; 2008 Nov 5-7; Universal City (CA), USA. Accessed July 12, 2021. Available from: https://www.lexjansen.com/wuss/2008/anl/anl03.pdf.

SAS Institute Inc. “The PHREG Procedure - Left-Truncation of Failure Times”. SAS Help Center. Accessed July 12, 2021.
https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.5/statug/statug_phreg_details13.htm

Wikipedia. “Survival analysis - section 1.2.2.” Accessed July 12, 2021. https://en.wikipedia.org/wiki/Survival_analysis#Life_table_for_the_aml_data

ACKNOWLEDGMENTS

I would like to acknowledge Lihua WU for the encouragement to me for submitting the paper and the valuable suggestions, detailed reviews and comments on the paper. Also, I want to thank Sanofi Pasteur colleagues’ great help during the internal review process. Last but not the least, thanks to all the colleagues in Sanofi for all the guidance and support provided.

CONTACT INFORMATION

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

Rolland LUO
Sanofi
Rolland.Luo@sanofi.com/549637376@qq.com

Lihua WU
Sanofi
Lihua1.Wu@sanofi.com