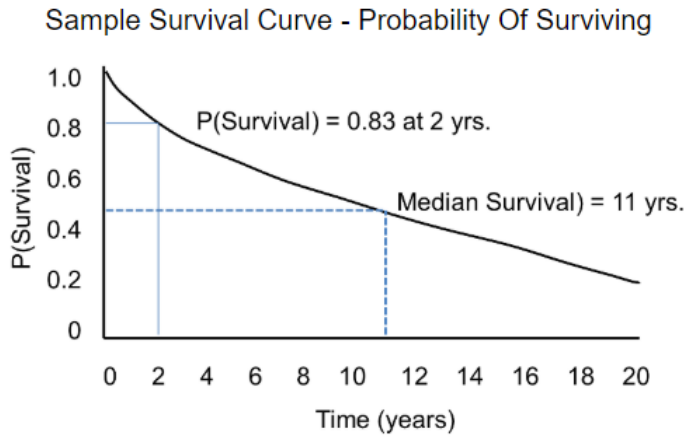
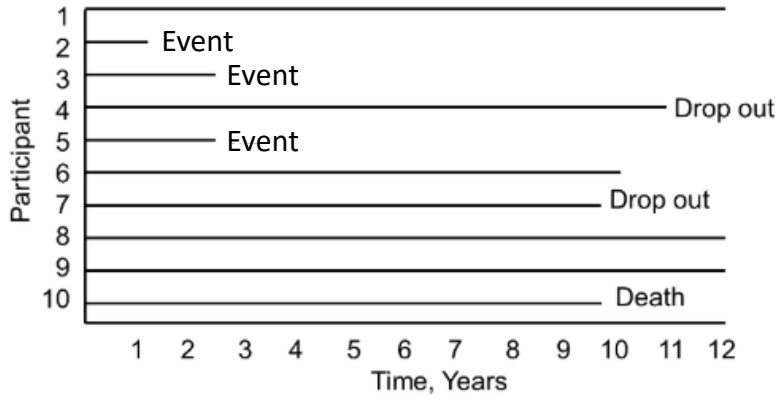
Survival Tips for Survival Analysis

XIAOTING (TING) WU, PH.D. MS. , IS A BIOSTATISTICIAN AT THE DEPARTMENT OF CARDIAC SURGERY AT UNIVERSITY OF MICHIGAN, US. SHE HAS SAS EXPERIENCE FOR MORE THAN 4 YEARS, AND IS ONE OF THE AWARDEES FOR NEW SAS PROFESSIONAL AWARD SAS GLOBAL FORUM 2019. TING HAS MANY YEAR OF EXPERIENCE IN DATA AND STATISTICAL ANALYSIS INCLUDING PREDICTION MODELS, SURVIVAL ANALYSIS, MIXED EFFECT MODELS AND CAUSAL INFERENCE. SHE HAS BROAD INTERESTS IN STATISTICAL METHODOLOGY, STATISTICAL CONSULTING, DATA VISUALIZATION AND SAS APPLICATIONS.

Introduction

- Survival analysis is a statistical technique that look at the probability of an event occurrence over time.





- Two pieces of information are needed in a survival analysis:

1) If the subjects have the event of interest (eg. an indicator status can be coded as 1=event; 0= no event (censoring);

2) The follow up time for each subject (specifically this is the time to events for those with events, and time to censoring for those without the events).

Introduction

Censoring and Truncation

- Censoring: Left, Right, Interval Censoring
- Assumption: censoring is independent and non-informative.
- Truncation: Left, Right

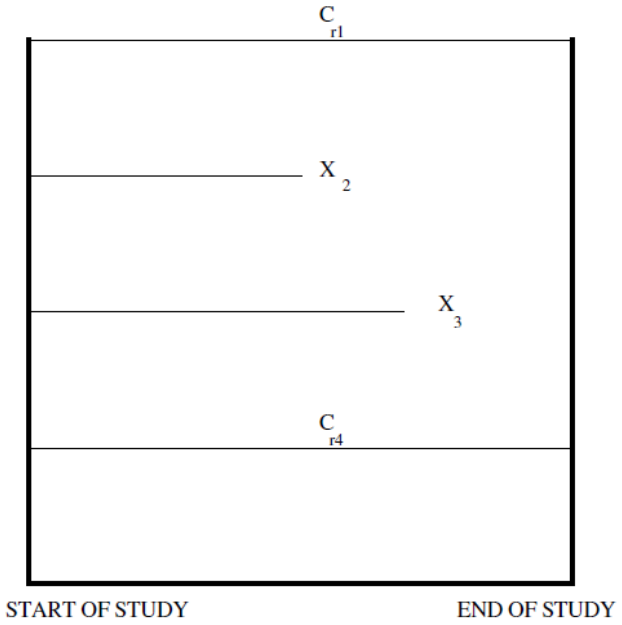


Figure 3.1 *Example of Type I censoring*

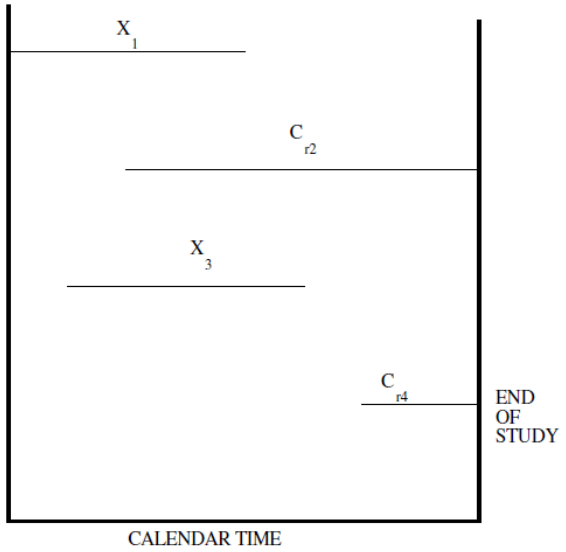


Figure 3.3 *Generalized Type I censoring when each individual has a different starting time*

Ref: John P. Klein 2013

Introduction-Basic functions in survival analysis

Assuming X is a continuous random variable with probability density function f(x) and cumulative distribution function $F(x) = Pr(X \leq x).$, giving the probability that the event has occurred by duration t.

1. Survival Function

$$S(x) = Pr(X > x) = \int_x^{\infty} f(t)\, dt. = 1 - F(x)$$

which is the probability that the event of interest has not occurred by duration t

2. Hazard Function

$$h(x) = \lim_{\Delta x \to 0} \frac{P[x \leq X < x + \Delta x \mid X \geq x]}{\Delta x}.$$

which is the instantaneous rate of occurrence of event conditional on subjects survived after time x

3. Cumulative hazard function

$$H(x) = \int_0^x h(u)\, du = -\ln[S(x)].$$

4. The relationship between survival function, probability density function and hazard function

$$h(x) = f(x)/S(x) = -d\ln[S(x)]/dx.$$

Ref: John P. Klein 2013

Introduction-Non-parametric Methods

- Use Kaplan Meier Method- Product-Limit estimator to estimate survival function

$$\hat{S}(t) = \begin{cases} 1 & \text{if } t < t_1, \\ \prod_{t_i \le t}[1 - \frac{d_i}{Y_i}], & \text{if } t_1 \le t \end{cases}$$

For values of *t* beyond the largest observation time this estimator is not well defined

Variance can be defined as $\hat{V}[\hat{S}(t)] = \hat{S}(t)^2 \sum_{t_i \le t} \frac{d_i}{Y_i(Y_i - d_i)}$.

- Use Nelson-Aalen estimator of cumulative hazard to estimate cumulative hazard

$$\tilde{H}(t) = \begin{cases} 0, & \text{if } t \le t_1, \\ \sum_{t_i \le t} \frac{d_i}{Y_i}, & \text{if } t_1 \le t. \end{cases}$$

Introduction-Semi-parametric model

- Cox (1972) Proportional Hazard Model is a semi-parametric model

Given Z as the covariates,

$$h(t \mid \mathbf{Z}) = h_0(t)\exp(\boldsymbol{\beta}^t\mathbf{Z}) = h_0(t)\exp\left(\sum_{k=1}^{p}\beta_k Z_k\right)$$

Proportional Hazard assumption:

$$\frac{h(t \mid \mathbf{Z})}{h(t \mid \mathbf{Z}^*)} = \frac{h_0(t)\exp\left[\sum_{k=1}^{p}\beta_k Z_k\right]}{h_0(t)\exp\left[\sum_{k=1}^{p}\beta_k Z_k^*\right]} = \exp\left[\sum_{k=1}^{p}\beta_k(Z_k - Z_k^*)\right]$$

If the proportional hazard assumption is violated for a variable, one approach is to stratify on this variable which fits a different baseline hazard function for each stratum.

Introduction- Parametric models

- An example of accelerated failure time (AFT) model with location and scale parameters:

$$\text{Exponential: } \log h(t) = \mu + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$

$$\text{Gompertz: } \log h(t) = \mu + \alpha t + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$

$$\text{Weibull: } \log h(t) = \mu + \alpha \log t + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$

- Parametric models and their distribution assumption:

Distribution	*Hazard Rate* $h(x)$	*Survival Function* $S(x)$	*Probability Density Function* $f(x)$	*Mean* $E(X)$
Exponential $\lambda > 0, x \geq 0$	λ	$\exp[-\lambda x]$	$\lambda \exp(-\lambda x)$	$\frac{1}{\lambda}$
Weibull $\alpha, \lambda > 0, x \geq 0$	$\alpha\lambda x^{\alpha-1}$	$\exp[-\lambda x^{\alpha}]$	$\alpha\lambda x^{\alpha-1}\exp(-\lambda x^{\alpha})$	$\frac{\Gamma(1+1/\alpha)}{\lambda^{1/\alpha}}$
Gamma $\beta, \lambda > 0, x \geq 0$	$\frac{f(x)}{S(x)}$	$1 - I(\lambda x, \beta)^*$	$\frac{\lambda^{\beta} x^{\beta-1}\exp(-\lambda x)}{\Gamma(\beta)}$	$\frac{\beta}{\lambda}$
Log normal $\sigma > 0, x \geq 0$	$\frac{f(x)}{S(x)}$	$1 - \Phi\left[\frac{\ln x - \mu}{\sigma}\right]$	$\frac{\exp\left[-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2\right]}{x(2\pi)^{1/2}\sigma}$	$\exp(\mu + 0.5\sigma^2)$

Ref: John P. Klein 2013

Introduction- SAS

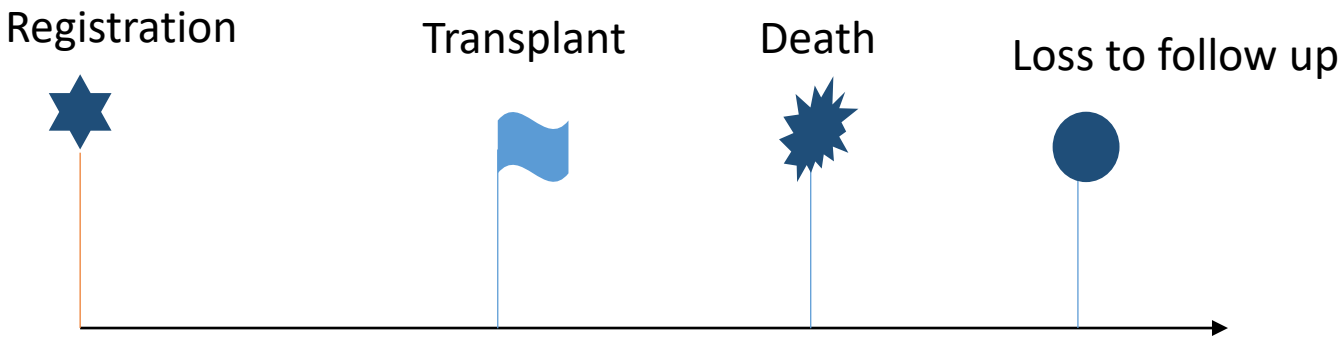
- SAS procedures for survival analysis

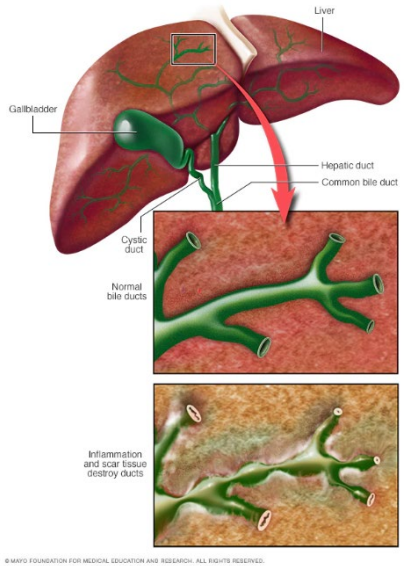
LIFETEST	is primarily designed for univariate analysis of the timing of events. It produces life tables and graphs of survival curves. Using several methods, this procedure tests whether survival curves are the same in two or more groups.
PHREG	uses Cox's partial likelihood method to estimate regression models with censored data. The model is less restrictive than the models in PROC LIFEREG, and the estimation method allows for time-dependent covariates.
LIFEREG	estimates regression models with censored, continuous-time data under several alternative distributional assumptions. PROC LIFEREG allows for several varieties of censoring, but it does not allow for time-dependent covariates.
RELIABILITY	provides tools for reliability and survival data analysis and for recurrent events data analysis; fit regression models including accelerated life test models, to combinations of censored data; construct probability and fitted life distribution of censored data.

Ref: Adapted from Paul Allison 2010

Data example

- The data set is described in Fleming and Harrington (1991). The study originates from Mayo Clinic trial (1974-1984). Primary Biliary Cirrhosis (PBC) is a rare but fatal chronic liver disease, which results in destruction of interlobular bile ducts.
- Event of interest : Death risk of PBC
- Time origin: Treatment start time
- Censoring events: Liver transplant, loss of follow up and end of study period.





Data example

Table 1. Variable specification of data example

Variables	Variable specification
id	case number
futime	number of days between registration and the earlier of death, transplant, or study analysis time in July, 1986
status	0=alive, 1=liver transplant, 2=dead
drug	1= D-penicillamine, 2=placebo
age	age in days
sex	0=male, 1=female
ascites	presence of ascites: 0=no 1=yes
hepato	presence of hepatomegaly 0=no 1=yes
spiders	presence of spiders 0=no 1=yes
edema	presence of edema: 0=no edema and no diuretic therapy for edema; .5 = edema present without diuretics, or edema resolved by diuretics; 1 = edema despite diuretic therapy
bili	serum bilirubin in mg/dl
chol	serum cholesterol in mg/dl
albumin	albumin in gm/dl
copper	urine copper in ug/day
alk_phos	alkaline phosphatase in U/liter
sgot	SGOT in U/ml
trig	triglicerides in mg/dl
platelet	platelets per cubic ml/1000
stage	histologic stage of disease; 1, 2, 3 or 4.

Get to know the data

```
proc freq data=pbc; tables drug;run;
proc means data=pbc n mean std median p25 p75 min max fw=6
maxdec=2;  var _numeric_; run;
```

Variable	N	Mean	Std Dev	Median	25th Pctl	75th Pctl	Minimum	Maximum
id	312	156.5	90.21	156.5	78.5	234.5	1	312
futime	312	2006.4	1123.3	1839.5	1191	2702.5	41	4556
status	312	0.86	0.96	0	0	2	0	2
drug	312	1.49	0.5	1	1	2	1	2
age	312	50.02	10.58	49.79	42.14	56.73	26.28	78.44
sex	312	0.88	0.32	1	1	1	0	1
ascites	312	0.08	0.27	0	0	0	0	1
hepato	312	0.51	0.5	1	0	1	0	1
spiders	312	0.29	0.45	0	0	1	0	1
edema	312	0.11	0.27	0	0	0	0	1
bili	312	3.26	4.53	1.35	0.8	3.45	0.3	28
chol	284	369.51	231.94	309.5	249	400	120	1775
albumin	312	3.52	0.42	3.55	3.31	3.8	1.96	4.64
copper	310	97.65	85.61	73	41	123	4	588
alk_phos	312	1982.7	2140.4	1259	867	1985	289	13862
sgot	312	122.56	56.7	114.7	80.6	151.9	26.35	457.25
trig	282	124.7	65.15	108	84	151	33	598
platelet	308	261.94	95.61	257	199.5	323	62	563
protime	312	10.73	1	10.6	10	11.1	9	17.1
stage	312	3.03	0.88	3	2	4	1	4

LIFETEST Procedure

- SAS LIFETEST procedure can be used to perform Kaplan-Meier analysis and describe nonparametric estimate of survival functions. An example of commonly used LIFETEST code is given below.

```
proc lifetest data=pbc outsurv=km_sur2 plots=survival( cl cb=all test  atrisk
STRATA=PANEL ) maxtime=4000  ;
   time futime*status(0,1);
   strata drug;
run;
```

- Obtain Kaplan Meier estimates

$$\hat{S}(t) = \begin{cases} 1 \ if \ t < t_1 \\ \prod_{t_i \le t} \left[1 - \frac{d_i}{Y_i}\right], if \ t_1 \le t \end{cases}$$

The probability of an individual surviving beyond time x (experiencing the event after time x), defined as S(x)=Pr (X >x).

	futime	Censoring Flag: 0=Failed 1=Censored	Survival Distribution Function Estimate	SDF Lower 95.00% Confidence Limit	SDF Upper 95.00% Confidence Limit
265	3336	1	0.5114055615	.	.
266	3358	0	0.5000409935	0.4204226209	0.5744468843
267	3388	1	0.5000409935	.	.
268	3395	0	0.4884121332	0.4074842323	0.5643988816
269	3422	1	0.4884121332	.	.
270	3428	0	0.4764996421	0.3943142214	0.5540592284
271	3445	0	0.464587151	0.3813107797	0.5435993402

LIFETEST Procedure

- Median survival time

Given x is a continuous random variable then the p_{th} quantile is found by solving $S(x^p)=1-p$. The median lifetime is the 50^{th} percentile $x^{0.5}$, the 75 percentile lifetime is the 75^{th} percentile $x^{0.75}$. ($S(x^{0.5}) = 1-0.5=0.5$; $S(x^{0.75}) = 1-0.75=0.25$).

In product limit estimate, we find the smallest time x^p for which the product limit estimator is less than or equal to 1-p. That is $\hat{X}_p = \inf\{t: \hat{S}(t) \leq 1-p\}$

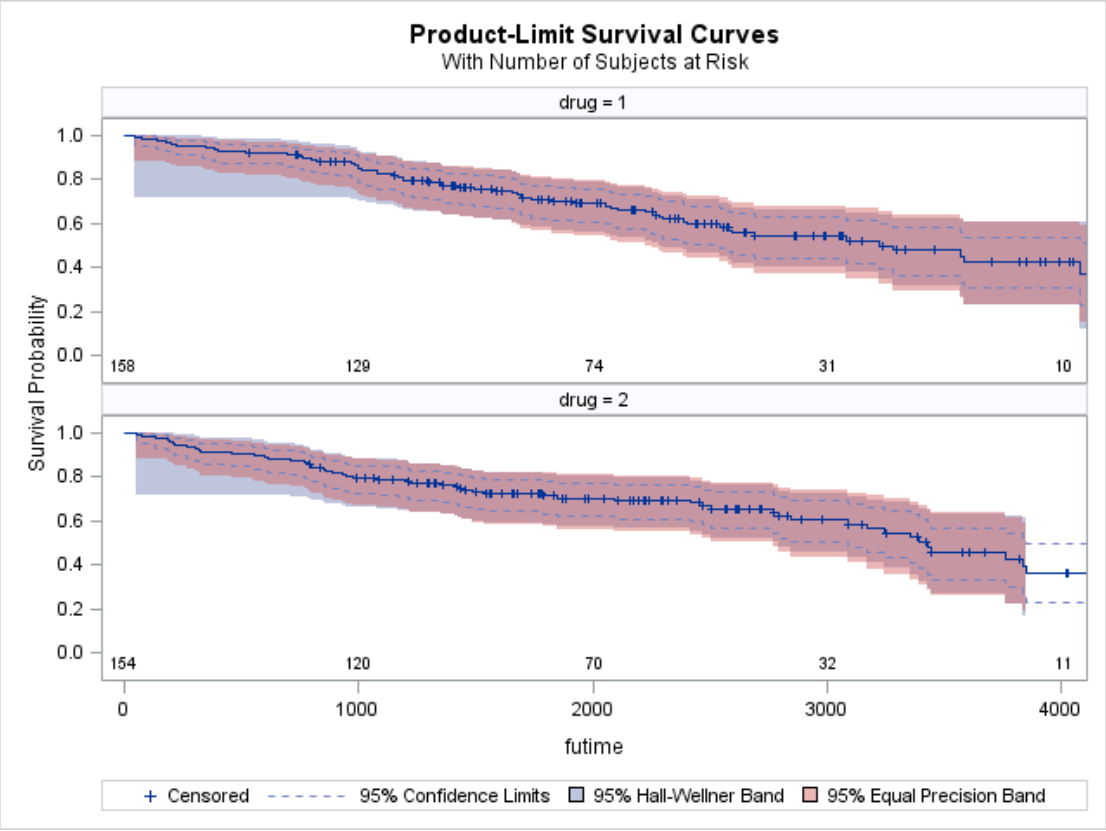
Quartile Estimates				
Percent	Point Estimate	95% Confidence Interval		
		Transform	[Lower	Upper)
75	.	LOGLOG	4191.00	.
50	3395.00	LOGLOG	3086.00	3839.00
25	1487.00	LOGLOG	1170.00	1925.00

futime	Censoring Flag: 0=Failed 1=Censored	Survival Distribution Function Estimate	SDF Lower 95.00% Confidence Limit	SDF Upper 95.00% Confidence Limit
3336	1	0.5114055615	.	.
3358	0	0.5000409935	0.4204226209	0.5744468843
3388	1	0.5000409935	.	.
3395	0	0.4884121332	0.4074842323	0.5643988816
3422	1	0.4884121332	.	.
3428	0	0.4764996421	0.3943142214	0.5540592284
3445	0	0.464587151	0.3813107797	0.5435993402
3445	1	0.464587151	.	.

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.

LIFETEST Procedure

```
proc lifetest data=pbc outsurv=km_sur2 plots=survival( cl cb=all test  atrisk  nocensor
STRATA=PANEL ) maxtime=4000  ;
   time futime*status(0,1);
   strata drug;
run;
```



- Option MAXTIME- truncate plots to avoid time points with small at risk number
- Option ATRISK – show number of subjects at risk at each time
- Option nocensored – suppress the censor symbols
- Option cl – show pointwise confidence limits for the survivor functions (you may also specify the type of confidence band (Hall-Wellner or Equal-Precision) using cb=hw or ep)
- Confidence band for survival function

1. Pointwise CI
2. Equal-Precision (EP) confidence band (log transform)
3. Hall-Wellner (HW) confidence band

Hall–Wellner bands are wider for small *t* and shorter for large *t* . Both bands are wider than the curves one obtains by using pointwise confidence intervals.

LIFETEST Procedure

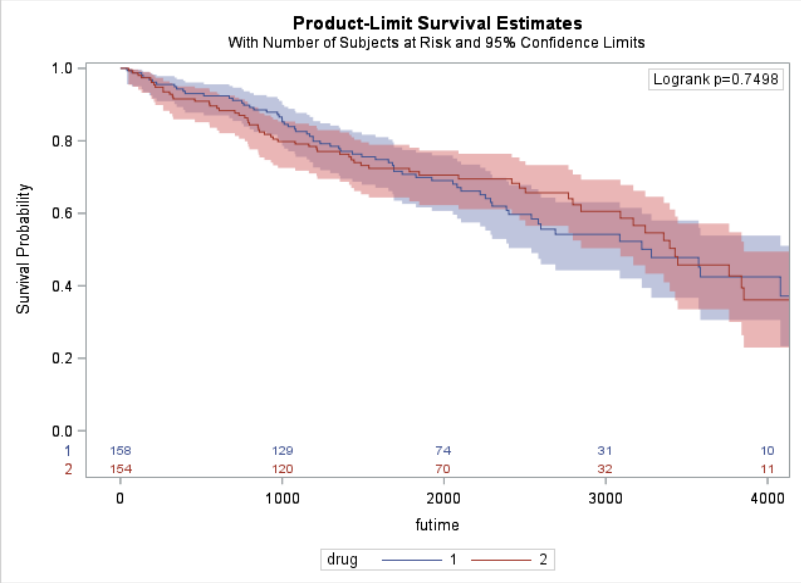


Table 3. Two groups survival comparison

Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	0.1017	1	0.7498
Wilcoxon	0.0018	1	0.9664
-2Log(LR)	0.0634	1	0.8013

- Log-rank test is standard method to test if the survival curve is different from two groups assuming the hazard is proportional constant over time. $Z_i = \frac{\sum_{j=1}^{J}(O_{i,j} - E_{i,j})}{\sqrt{\sum_{j=1}^{J} V_{i,j}}} \xrightarrow{d} \mathcal{N}(0,1)$
- Wilcoxon test places more weight on shorter survival times. When the two survival curves cross, it's possible that the hazard ratios change over the time. For example, DRUG 1 has higher survival probability in the earlier time, and DRUG 2 is higher in the later time. In this case, Wilcoxon test will be a better choice for the two group comparison.
- A likelihood ratio test which is assuming an exponential model is also included to compare the two group survivals

LIFETEST Procedure- Customize your plot

Step 1: Define the macro

Web at http://support.sas.com/documentation/onlinedoc/stat/ex_code/121/templft2.html. 1

Step 2: Call the macro

```
%SurvivalTemplateRestore    *make the macro;
```

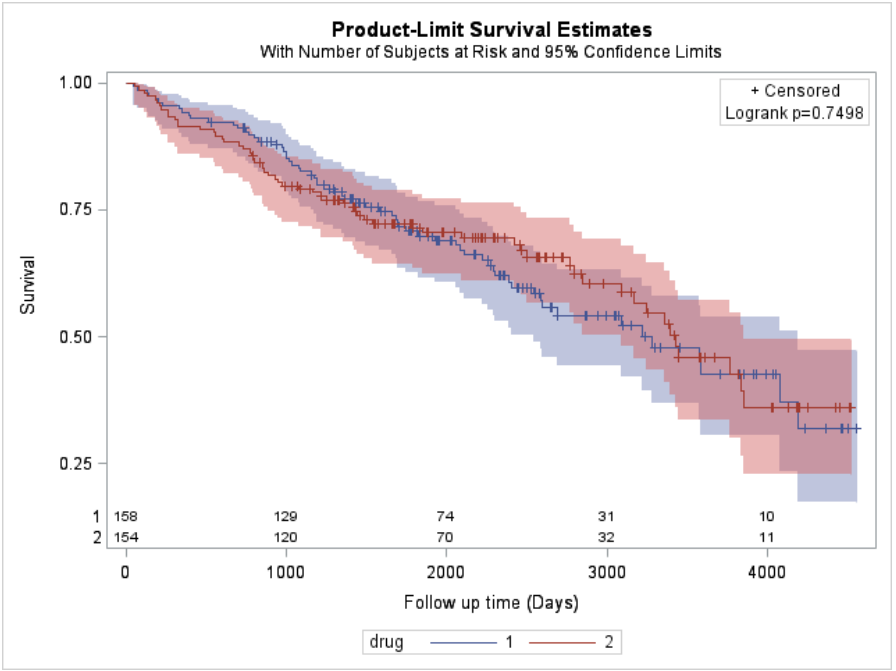
```
WARNING: SASUSER.TEMPLAT is not a template store! It will be ignored.
NOTE: There are no template stores with write access in the path.
      The default template path will be used for this operation.
ERROR: Template 'Stat.Lifetest.Graphics.ProductLimitSurvival' was unable to write to template store!
WARNING: SASUSER.TEMPLAT is not a template store! It will be ignored.
NOTE: There are no template stores with write access in the path.
      The default template path will be used for this operation.
```

Warning message solution

```
ods path (prepend) work.templat (update);
```

Step 3: Customize the plots and compile the template

```
%SurvivalTemplateRestore    *make the macro;
%let xOptions = label="Follow up time (Days)";    *change the
title;
%let yOptions = label="Survival" linearopts=(viewmin=0.2
viewmax=1 tickvaluelist=(0 .25 .5 .75 1));
%SurvivalTemplate    * compile the templates with the new title;
```



Step 4: Restore the default macro and the default templates

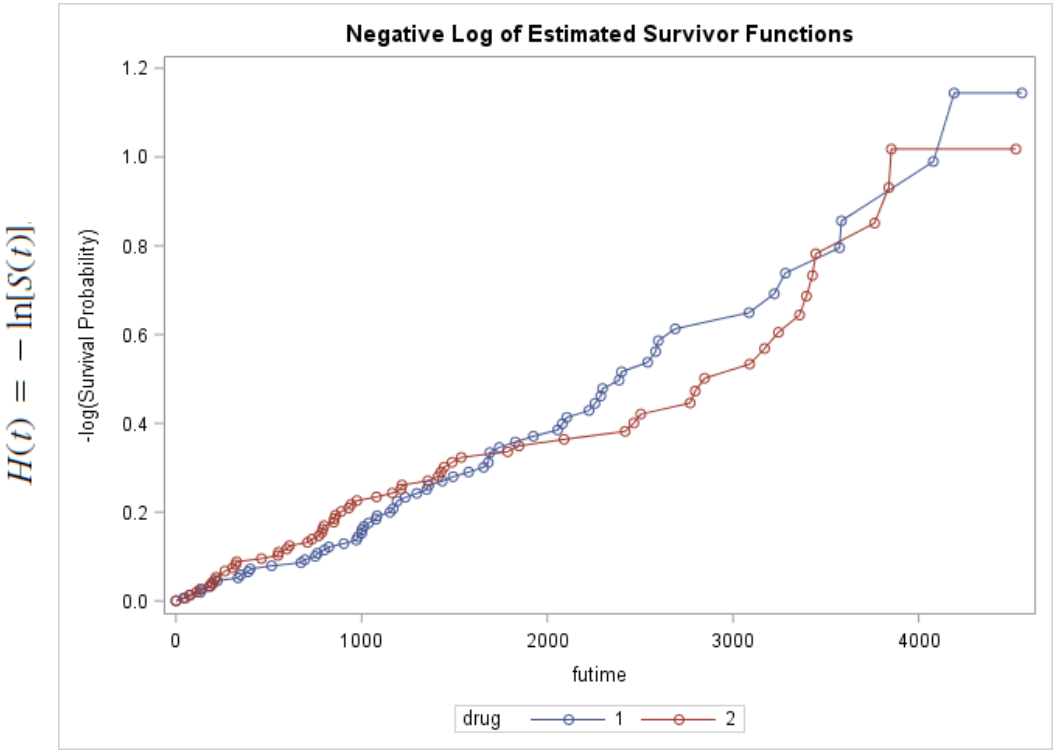
```
%SurvivalTemplateRestore    *restore the default macro;
proc template; /* Restore the default templates. */
delete Stat.Lifetest.Graphics.ProductLimitSurvival /
store=sasuser.templat;
delete Stat.Lifetest.Graphics.ProductLimitSurvival2 /
store=sasuser.templat;
run;
```

Ref: *Warren Kuhfeld and Ying So, SASGF 2013*

LIFETEST Procedure

- `Option plots=` `all` provides survival curves and cumulative hazard curve and log of negative log estimated survivor functions cuves.

Figure 2. Plot of Estimated Negative Log Survivor Functions



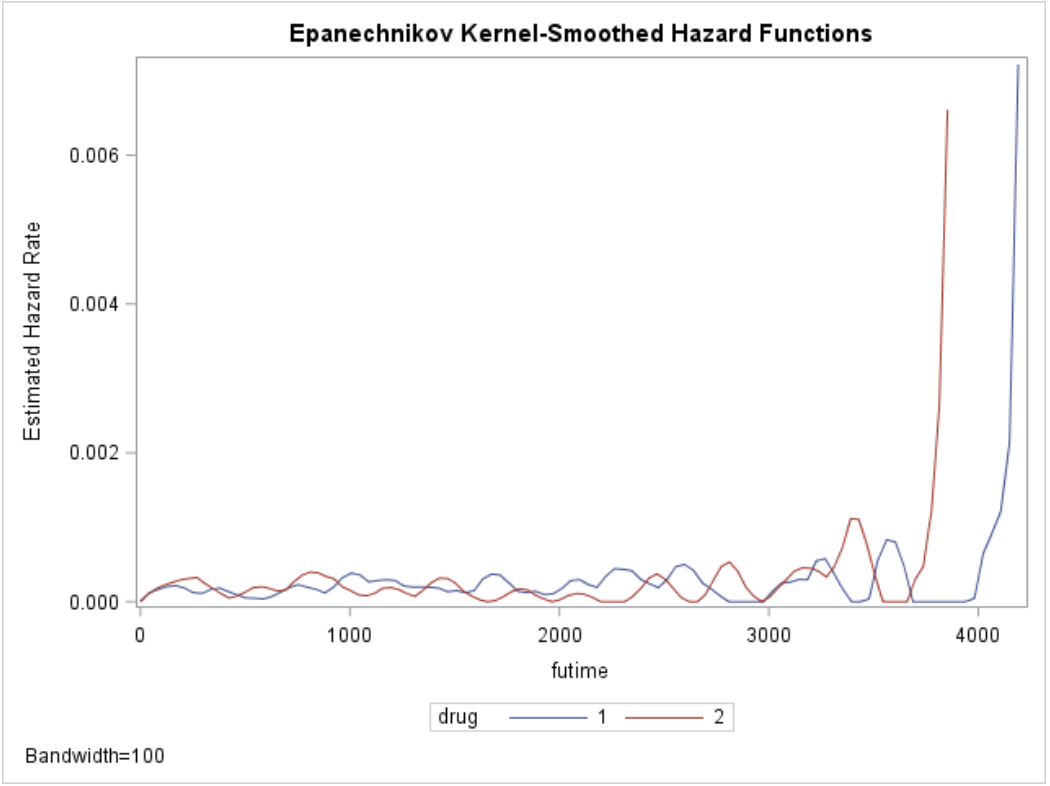
$H(t) = -\ln[S(t)]$

* Negative log survival function is the cumulative hazard function. This curve approximates a straight line through the origin, indicating an exponential model.

LIFETEST Procedure

- Provide a crude look at the hazard rate over time

```
proc lifetest data=pbc method=pl nelson plots(only)= hazard(kernel=e
bw=100);
    time futime*status(0,1);
    strata drug;
run;
```



PHREG procedure

- The PHREG procedure perform survival regression based on Cox proportional hazard model (Cox 1972). Cox model is a semiparametric model and assumes the hazard of explanatory variables are constant over time.

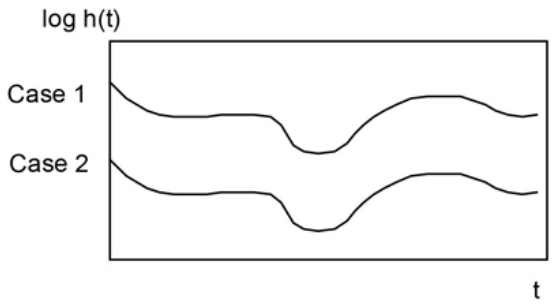
- The event time for each subject is defined by hazard function.

$$h_i\,(t) = h\,(t;\, z_i) = \; h_0\,(t)\exp(z_i'\beta)$$

where h_0(t) is a unspecified baseline hazard function, z_i is vector of explanatory variables, β is the model coefficients.

- Cox (PH) model assumption:
  1) regression coefficient β is constant over time
  2) linear combination of the covariates
  3) link function is exponential

Figure 5.1 *Parallel Log-Hazard Functions from Proportional Hazards Model*



Ref: Allison 2010

PHREG procedure

An example of commonly used PHREG codes are given below.

```
proc phreg data=pbc simple;
class drug (ref='2') /param=ref;
model futime*status(0,1)=drug;
hazardratio drug/diff=ref cl=wald;
run;
```

Table 4. Output of PHREG procedure

Summary of the Number of Event and Censored Values			
Total	Event	Censored	Percent Censored
312	125	187	59.94

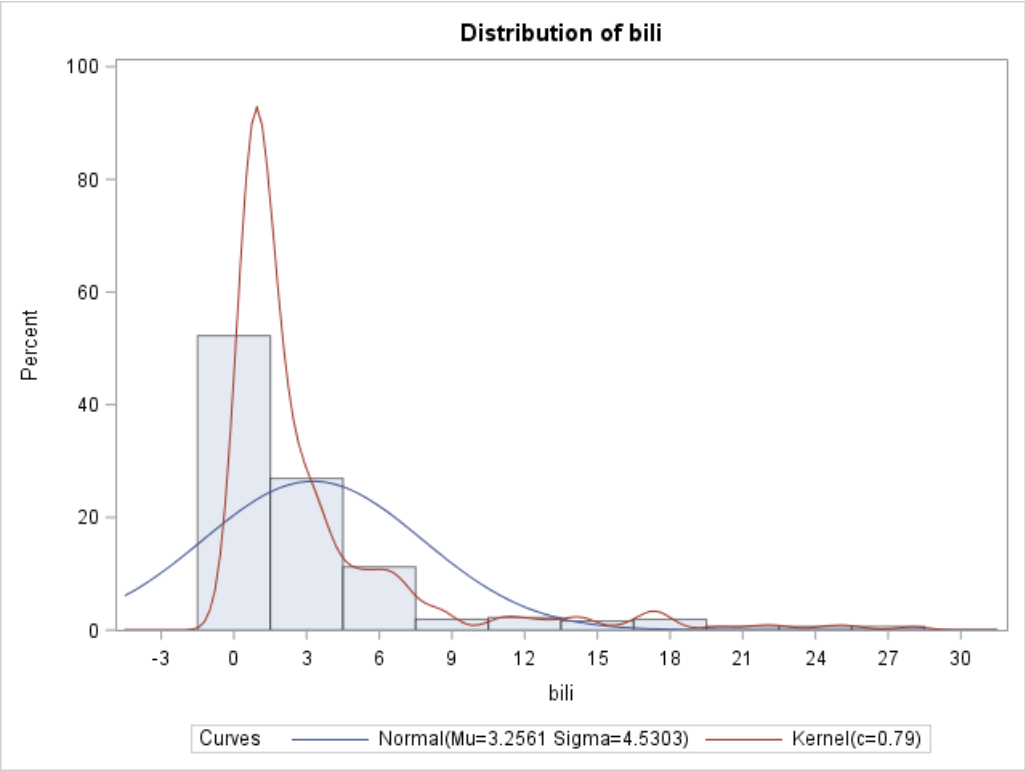
Analysis of Maximum Likelihood Estimates								
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
drug	1	1	0.05709	0.17916	0.1015	0.7500	1.059	drug 1

Hazard Ratios for drug			
Description	Point Estimate	95% Wald Confidence Limits	
drug 1 vs 2	1.059	0.745	1.504

PHREG procedure

- Check covariates distribution

```
proc univariate data=pbc; where drug ~=.;
var bili;
histogram bili/normal(mu=est sigma=est) kernel;
*histogram bili/lognormal(scale=est shape=est) kernel;
run;
```



PHREG procedure- Variable selection

```
* Consider a model with all the 'significant' variables from the univariable analyses.

proc phreg data=pbc2 ;
class drug (ref='Placebo') sex (ref='male') edema (ref='no') ascites(ref='no') hepato(ref='no')
spiders(ref='no') stage (ref='4')/param=ref;
model futime*status(0,1)= drug sex edema ascites hepato spiders stage age logbili logalbumin
logprotime platelet logsgot/ include=1 selection=stepwise slentry=.15 slstay=.20 details;
logbili=log(bili); logalbumin=log(albumin); logprotime=log(protime); logsgot=log(sgot);
format sex sex. drug drug. edema edema. ascites hepato spiders affirm.;

run;
```

- Variable selection can be performed using SELECTION option in the MODEL statement. Stepwise selection uses alternative steps of forward and backward selection.
- SLENTRY is the significance level specified for inclusion (entry).
- The covariate that is found to be *most significant* is selected for entry. SAS uses the score test instead of the LRT.
- SLSTAY governs if the variable should remain in the model. Set SLSTAY *higher* than SLENTRY.
- Some crucial variables can be forced to remain in model during selection process using option INCLUDE.

PHREG procedure

Table 5. Output of variable selection in a Cox model

Summary of Stepwise Selection							
Step	Effect		DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
	Entered	Removed					
1	logbili		1	2	155.9923		<.0001
2	logalbumin		1	3	33.3877		<.0001
3	age		1	4	17.6357		<.0001
4	logprotime		1	5	12.6612		0.0004
5	logsgot		1	6	3.7772		0.0520
6	edema		1	7	4.4604		0.0347
7	stage		3	8	5.4774		0.1400

PHREG procedure

Table 5. Output of variable selection in a Cox model

Analysis of Maximum Likelihood Estimates								
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
drug	D-pencl	1	-0.11018	0.18790	0.3439	0.5576	0.896	drug D-pencl
edema	yes	1	0.47382	0.23042	4.2287	0.0397	1.606	edema yes
stage	1	1	-1.70138	1.02752	2.7417	0.0978	0.182	stage 1
stage	2	1	-0.46428	0.31712	2.1434	0.1432	0.629	stage 2
stage	3	1	-0.28418	0.22643	1.5751	0.2095	0.753	stage 3
age		1	0.03620	0.00929	15.1993	<.0001	1.037	
logbili		1	0.73800	0.11515	41.0736	<.0001	2.092	
logalbumin		1	-2.62254	0.76130	11.8667	0.0006	0.073	
logprotime		1	3.52896	1.19089	8.7811	0.0030	34.089	
logsgot		1	0.59723	0.25200	5.6167	0.0178	1.817	

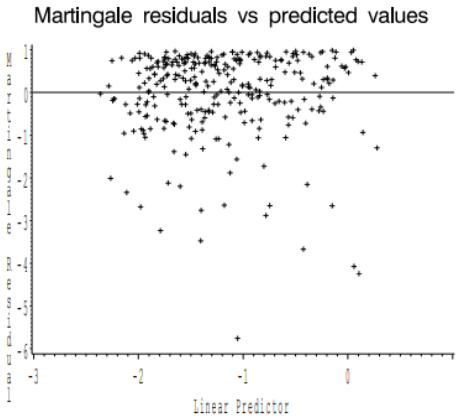
PHREG procedure- Check Model Specification

- In PHREG procedure, the ASSESS statement performs the graphical and numerical methods of Lin, Wei, and Ying for checking the assumed Cox regression model. The methods show cumulative sums of martingale residuals over follow-up times or covariates. Supremum test is also provided to compare the simulation process given the assumed model to the observed process. If the observed process is within the patterns of the simulated paths and the p values from supremum test is not significant, it indicates that the model specification is valid.

- Martingale residuals

$$\hat{M}_j = \delta_j - \hat{H}_o(T_j)\exp\left(\sum_{k=1}^{p} Z_{jk} b_k\right) = \delta_j - r_j, j = 1, \ldots, n.$$

Here, the residual Mi can be interpreted as the difference between the observed number of deaths (0 or 1) for subject i between time 0 and Xi , and the expected numbers based on the fitted model.



- Cumulative sums of martingale residuals

$$W_k(z) = \sum_{i=1}^{n} I(Z_{ik} \le z) M_i$$

Plotting this vs. z checking the functional form of covariate

$$W_r(r) = \sum_{i=1}^{n} I(\hat{\beta}'\mathbf{Z}_i \le r) M_i$$

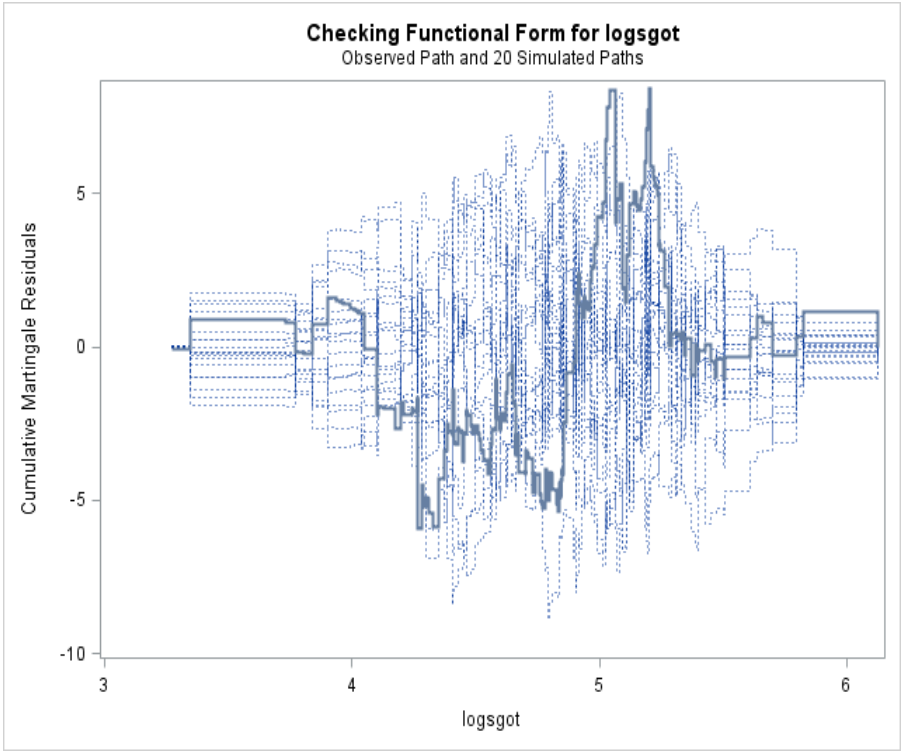
Plotting this vs. r checking the proportional hazard assumption of Cox model

PHREG procedure- Check Model Specification

- To test if the function forms of continuous variables are correctly specified, option VAR can be used in the ASSESS statement.

```
assess var=(age logbili logalbumin logprotime  logsgot);
```

Figure 3. Checking functional form of a continuous variable in Cox model

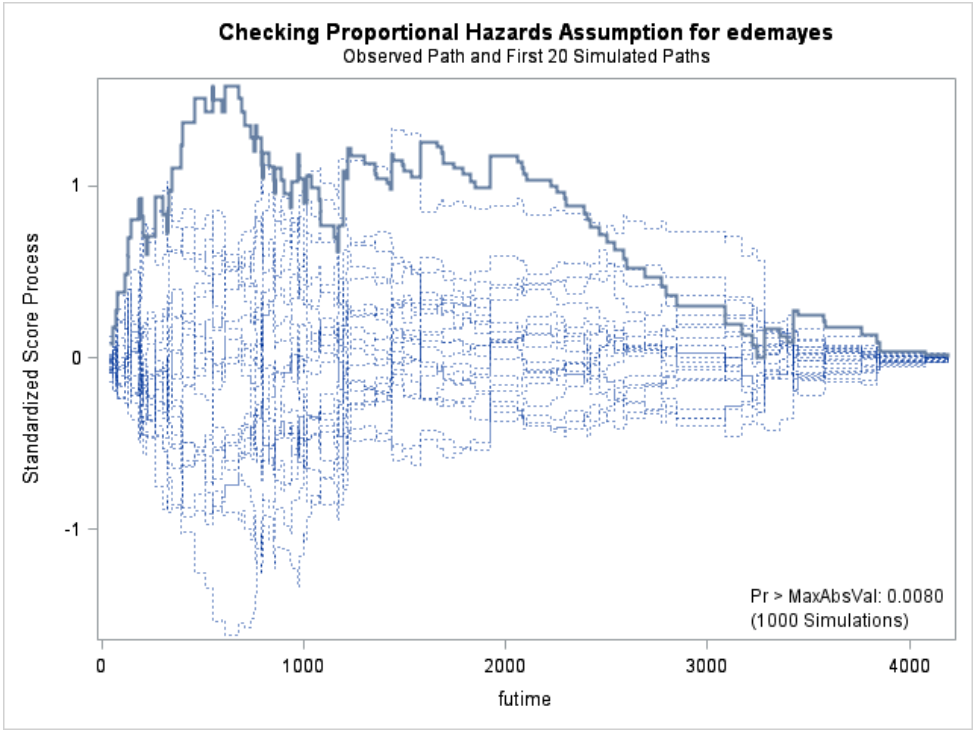


PHREG procedure- Check Model Specification

- To test the proportional hazard assumption, option PH can be used in the ASSESS statement.
  assess ph /resample=**1000** seed=**3538626** npath=**20**;

Table 6. Supremum Test for Proportional Hazards Assumption

Supremum Test for Proportional Hazards Assumption				
Variable	Maximum Absolute Value	Replications	Seed	Pr > MaxAbsVal
drugD_pencl	1.1200	1000	3538626	0.1640
edemayes	1.5784	1000	3538626	0.0080
stage1	0.5833	1000	3538626	0.3300
stage2	1.0363	1000	3538626	0.2880
stage3	1.3854	1000	3538626	0.0760
age	1.0090	1000	3538626	0.2150
logbili	1.2851	1000	3538626	0.1270
logalbumin	0.8449	1000	3538626	0.5210
logprotime	1.9162	1000	3538626	0.0030
logsgot	0.9886	1000	3538626	0.3380



PHREG procedure- Check Model Specification

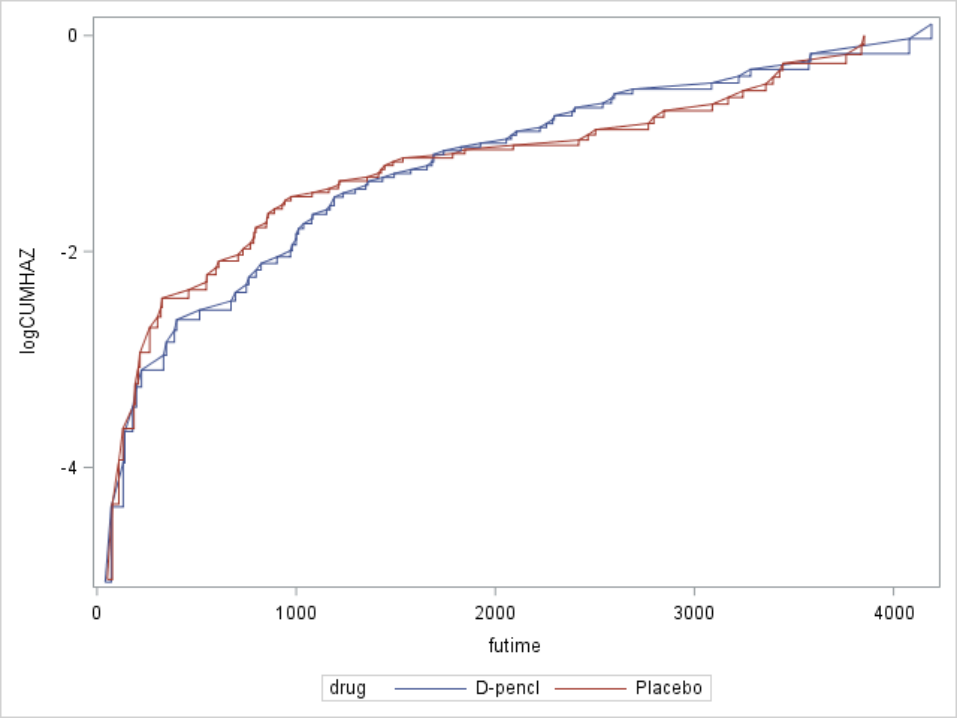
- Examining Cumulative Hazard plots for PH assumption
- If proportional hazards assumption holds, then the vertical distance between groups should be constant.

```
ods output productlimitestimates=pl;

proc lifetest data=pbc nelson;
   time futime*status(0,1);
   strata drug;
   format drug drug.;
run;

data pl2;
set pl(keep=drug futime Cumhaz censor
where=(censor=0));
logCUMHAZ=log(CUMhaz);
run;
*Use SGPLOT for plotting;

proc sgplot data=pl2;
series x=futime y=logcumhaz/group=drug;
step x=futime y=logcumhaz/group=drug;
format drug drug.;
run;
```



PHREG procedure- Stratified COX model

- Use stratified COX model to deal with variables with violating PH assumption. Stratified COX model assumes a different baseline hazard function for each stratum.

- SAS statement: `strata edema;`

$$\lambda(t|\mathbf{Z}(t)) = \lambda_0(t)\ \exp\{\boldsymbol{\beta}'\mathbf{Z}(t)\}$$

$\lambda_0(t)$ Is the baseline hazard common to all factors in the model

$$\lambda_k(t|\mathbf{Z}(t)) = \lambda_{0k}(t)\ \exp\{\boldsymbol{\beta}'\mathbf{Z}(t)\},$$

$\lambda_{0k}(t)$ is the baseline hazard for stratum k, k=1, …, K .

PHREG procedure- Prediction

For a patient with given condition, the survival probability over time can be calculated from the given model.

```
data var;
   format sex sex. drug drug. edema edema. ascites hepato spiders affirm.;
   input  drug  edema    stage age logbili logalbumin logprotime  logsgot label;
   datalines;
  1 0 1 50 0.6 1.3 2.4 4.7  1
  0 0 1 50 0.6 1.3 2.4 4.7  0
  ;


proc phreg data=pbc2 plots (cl overlay )=survival atrisk;
where drug ~=.;
class drug (ref='Placebo')  edema (ref='no') stage (ref='4')/param=ref;
model futime*status(0,1)=drug  edema stage age logbili logalbumin logprotime  logsgot;
logbili=log(bili); logalbumin=log(albumin); logprotime=log(protime); logsgot=log(sgot);
format sex sex. drug drug. edema edema. ascites hepato spiders affirm.;
strata edema;
baseline out=test1 covariates=var survival=a / group=drug rowid=label;run;
hazardratio drug/diff=ref cl=wald;
*hazardratio edema/diff=ref cl=wald;
hazardratio stage/diff=all cl=wald;
run;
```

PHREG procedure- Prediction

- Output of stratified COX model and hazard ratios

Summary of the Number of Event and Censored Values					
Stratum	edema	Total	Event	Censored	Percent Censored
1	no	263	89	174	66.16
2	yes	49	36	13	26.53
Total		312	125	187	59.94

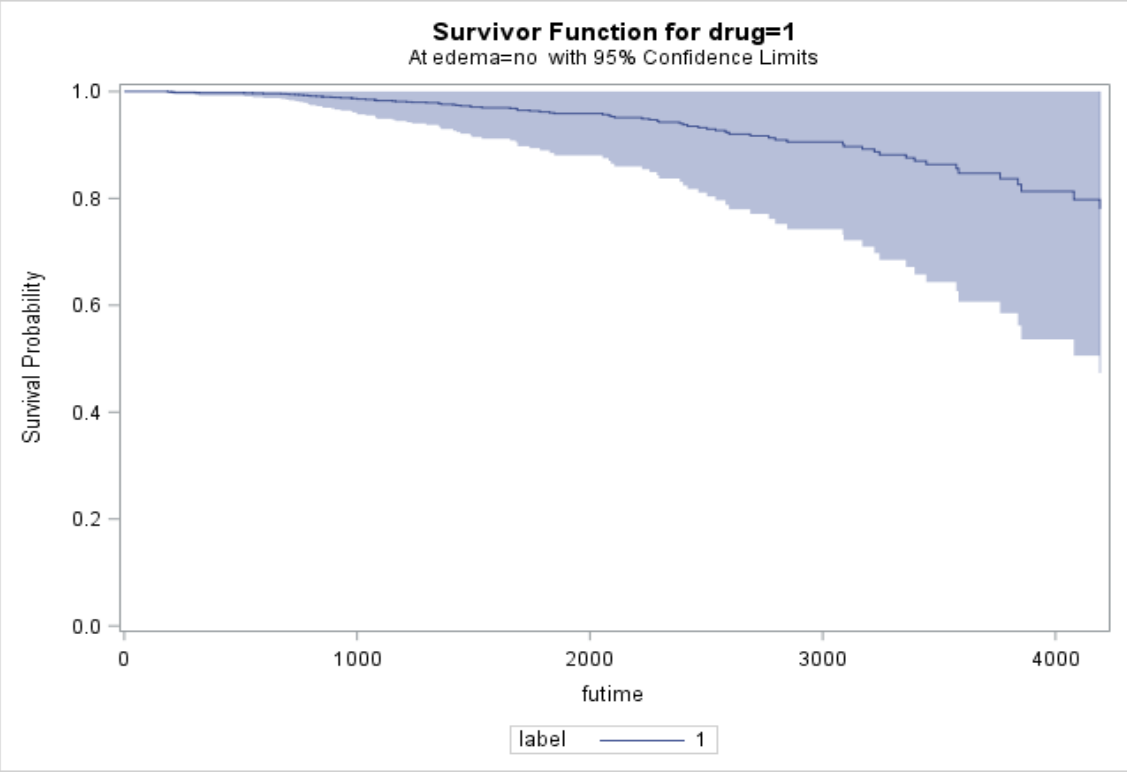
Hazard Ratios for drug			
Description	Point Estimate	95% Wald Confidence Limits	
drug D-pencl vs Placebo	0.936	0.648	1.352

Analysis of Maximum Likelihood Estimates								
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
drug	D-pencl	1	-0.06656	0.18779	0.1256	0.723	0.936	drug D-pencl
stage	1	1	-1.87263	1.03158	3.2953	0.0695	0.154	stage 1
stage	2	1	-0.47868	0.3193	2.2475	0.1338	0.62	stage 2
stage	3	1	-0.21147	0.22555	0.8791	0.3485	0.809	stage 3
age		1	0.03644	0.00938	15.1031	0.0001	1.037	
logbili		1	0.70303	0.11517	37.2605	<.0001	2.02	
logalbumin		1	-2.46088	0.76273	10.4098	0.0013	0.085	
logprotime		1	3.70524	1.21602	9.2843	0.0023	40.66	
logsgot		1	0.57008	0.24661	5.3437	0.0208	1.768	

Hazard Ratios for stage			
Description	Point Estimate	95% Wald Confidence Limits	
stage 1 vs 2	0.248	0.031	1.982
stage 1 vs 3	0.19	0.025	1.441
stage 1 vs 4	0.154	0.02	1.161
stage 2 vs 3	0.766	0.421	1.391
stage 2 vs 4	0.62	0.331	1.158
stage 3 vs 4	0.809	0.52	1.259

PHREG procedure- Prediction

Figure 4. Plot predicted survival curve from a Cox model



PHREG procedure- Time Dependent Variable

- Time dependent variable could be used to test proportional hazard assumption
- Generate the time dependent covariates by creating interactions of the predictors and a function of survival time and include in the model. If any of the time dependent covariates are significant then those predictors are not proportional.

```
proc phreg data=pbc2 ;
class drug (ref='Placebo')  edema (ref='no') stage (ref='4')/param=ref;
model futime*status(0,1)=drug  edema  edema*gt  stage age logbili logalbumin
logprotime  logsgot ;
logbili=log(bili); logalbumin=log(albumin); logprotime=log(protime);
logsgot=log(sgot);
gt= log(futime);
format sex sex. drug drug. edema edema. ascites hepato spiders affirm.;
run;
```

Analysis of Maximum Likelihood Estimates

Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
drug	D-pencl	1	-0.08461	0.18717	0.2043	0.6512	0.919	drug D-pencl
edema	yes	1	5.66722	1.79178	10.0040	0.0016	.	edema yes
gt*edema	yes	1	-0.77959	0.26643	8.5618	0.0034	.	edema yes * g

LIFEREG procedure

- Parametric modeling using exponential distribution:

```
proc lifereg data=pbc2;
class drug stage;
model  futime*status(0,1)=drug stage
age/ dist=gamma noshape1 shape1=1
noscale scale=1;
run;
```

$$\log T_i = \beta_0 + \beta_1 x_{\lambda} + \ldots + \beta_k x_{ik} + \sigma\varepsilon_i$$

Dist=*distribution-type*: specifies the distribution type assumed for the failure time

DISTRIBUTION=	NOLOG Specified?	Resulting Distribution
EXPONENTIAL	No	Exponential
EXPONENTIAL	Yes	One-parameter extreme value
GAMMA	No	Generalized gamma
GAMMA	Yes	Generalized gamma with untransformed responses
LOGISTIC	No	Logistic
LOGISTIC	Yes	Logistic (NOLOG has no effect)
LLOGISTIC	No	Log-logistic
LLOGISTIC	Yes	Logistic
LNORMAL	No	Lognormal
LNORMAL	Yes	Normal
NORMAL	No	Normal
NORMAL	Yes	Normal (NOLOG has no effect)
WEIBULL	No	Weibull
WEIBULL	Yes	Extreme value

LIFEREG procedure

- Output from LIEFREG procedure:

Model Information	
Data Set	WORK.PBC2
Dependent Variable	Log(futime)
Censoring Variable	status
Censoring Value(s)	0 1
Number of Observations	312
Noncensored Values	125
Right Censored Values	187
Left Censored Values	0
Interval Censored Values	0
Number of Parameters	6
Name of Distribution	Exponential
Log Likelihood	-295.1126529

Analysis of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	9.3633	0.4927	8.3976	10.3291	361.1	<.0001
drug	1	1	-0.0082	0.1824	-0.3658	0.3494	0	0.9641
drug	2	0	0	.	.	.	.	.
stage	1	1	2.7605	1.0102	0.7805	4.7405	7.47	0.0063
stage	2	1	1.3197	0.2806	0.7698	1.8696	22.13	<.0001
stage	3	1	0.7815	0.1993	0.3909	1.1721	15.38	<.0001
stage	4	0	0	.	.	.	.	.
age		1	-0.029	0.0088	-0.0463	-0.0117	10.77	0.001
Scale		0	1	0	1	1		
Weibull Shape		0	1	0	1	1		

- The signs of the coefficients tell us the direction of the relationship. $\beta > 0$ indicates longer time to death (lower hazard) compared to the reference group. $\beta < 0$ suggests shorter time (increased hazard) to event.
- Take exponential transformation of coefficients, it could be interpreted as the ratio of the expected (mean or median) survival times for the two groups. For example, given the same other covariates, for treatment effect on death time, there are 100 (exp (-0.0082)-1) = - 0.8 percent change in the expected survival time for drug 1 versus drug 2.

RELIABILITY procedure

Suppose we want estimates and 95% confidence intervals for the 25th, 50th and 75th percentiles for age 58.76 , stage 4 with the drug=1 treatment.

```
data covar;
input drug stage age @@;
datalines;
1 4 58.765229295
;
run;

data pbc_var;
set covar(in=one) pbc2;
if one then control=1;
else control=0;
run;

ods select ModObstats;
proc reliability data=pbc_var;
class drug stage  ;
distribution  EXPONENTIAL;
model futime*status(0,1)=drug stage age  /obstats(quantiles=.25 .50 .75  control=control);
run;
```

RELIABILITY procedure

- Output from RELIABILITY procedure:

Exponential Parameter Estimates					
Parameter		Estimate	Standard	Asymptotic Normal	
			Error	95% Confidence Limits	
				Lower	Upper
Intercept		9.3633	0.4927	8.3976	10.329
drug	1	-0.0082	0.1824	-0.3658	0.3494
drug	2	0	0	0	0
stage	1	2.7605	1.0102	0.7805	4.7405
stage	2	1.3197	0.2806	0.7698	1.8696
stage	3	0.7815	0.1993	0.3909	1.1721
stage	4	0	0	0	0
age		-0.029	0.0088	-0.0463	-0.0117
Shape		1	0	1	1

Observation Statistics									
futime	status	age	drug	stage	Prob	Pcntl	Stderr	Lower	Upper
.	.	58.765229	1	4	0.25	605.6902	92.07121	449.6339	815.9096
.	.	58.765229	1	4	0.5	1459.363	221.8383	1083.357	1965.87
.	.	58.765229	1	4	0.75	2918.725	443.6766	2166.715	3931.739

Summary

- Review fundamental survival analysis terms and principles
- Implementation SAS procedures: LIFETEST, PHREG, LIFEREG, RELIABILITY
- Showcase graphs describing survival probability over time
- Apply non-parametric, semi-parametric and parametric methods to estimate risk factors in time to event process

Ref: *http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Survival/BS704_Survival_print.html*

Reference

1. John P. Klein , M.L.M., *Survival Analysis: Techniques for Censored and Truncated Data.* 2005.
2. Allison, P.D., *Survival Analysis Using SAS: A Practical Guide.* Sas Inst., 2010.
3. Lin, D., Wei, L. J., and Ying, Z., *Checking the Cox Model with Cumulative Sums of Martingale-Based Residuals.* Biometrika, 1993. 80: p. 557–572.
4. Warren F. Kuhfeld and Ying So, SAS Institute Inc. *Creating and Customizing the Kaplan-Meier Survival Plot in PROC LIFETEST.* SAS Global Forum 2013

Name: Xiaoting Wu (Ting), PhD, MS

Organization: Department of Cardiac Surgery, University of Michigan

Work Phone: 734.936.7731

E-Mail: xiaotinw@med.umich.edu