Survival Analysis in SAS

Abigail Smith, PhD

Michigan SAS Users Group Conference
June 9, 2022

Arbor Research
Collaborative for Health

Outline

- Introduction
- Non-parametric models – Kaplan-Meier
- Semi-parametric models – Cox
- Checking assumptions
  - Functional form for continuous covariates
  - Proportional hazards
- Extensions of survival analysis

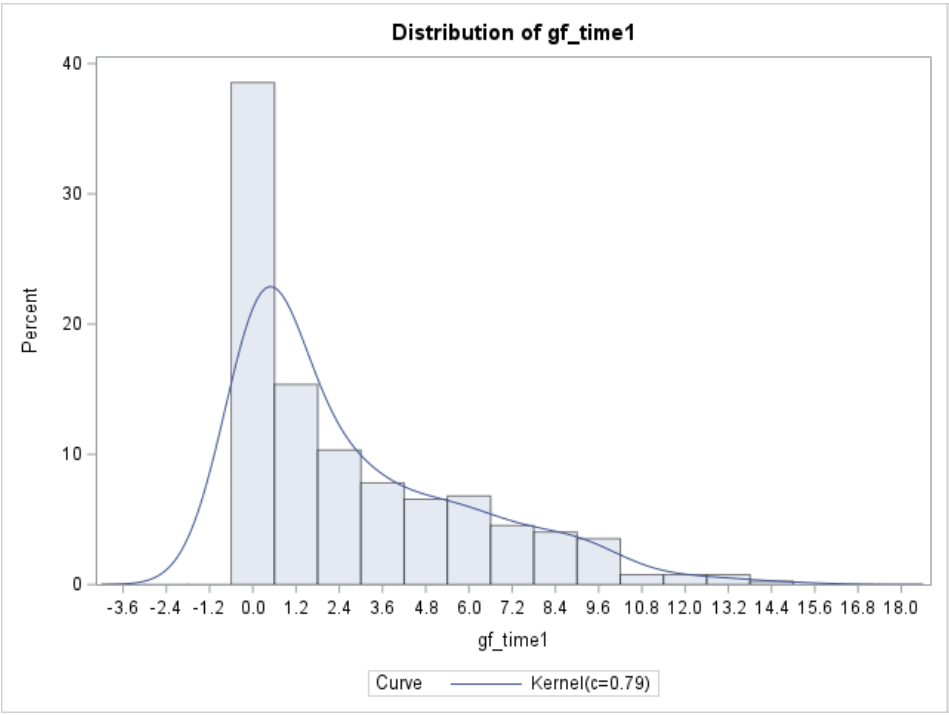
Introduction

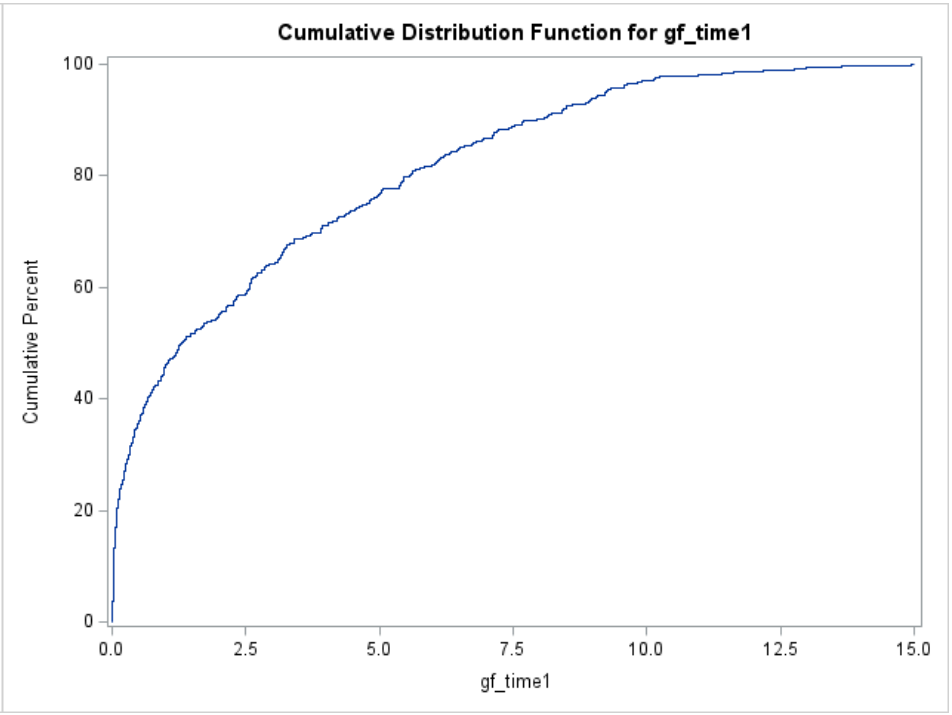
- A collection of methods used to model time-to-event data
- Linear regression?
  - OLS assumptions fail – survival times are not normally distributed
- Logistic regression?
  - OK in certain (limited) scenarios but censoring creates challenges

Data Structure

Participant ID	Time (Years)	Event (1=event, 0=censoring)	Covariate 1 (categorical)	Covariate 2 (continuous)
1	2.5	0	1	38
2	0.8	1	1	29
3	1.2	1	0	44
4	3	0	1	41
5	2.8	1	0	32
6	0.3	0	0	26
7	1.9	0	1	37
8	3.7	0	1	28
9	1	1	1	33
10	0.5	1	0	35

Distribution of Survival Times






```
proc univariate data=outcomes (where=(gf=1));
var gf_time1;
histogram gf_time1/kernel;
run;
```

```
proc univariate data=outcomes (where=(gf=1));
var gf_time1;
cdfplot gf_time1;
run;
```

Types of models

- Non-parametric
  - **Kaplan-Meier**
  - Nelson-Aalen
- Semi-parametric
  - **Cox**
- Parametric
  - Accelerated failure time models (exponential, Weibull, etc)

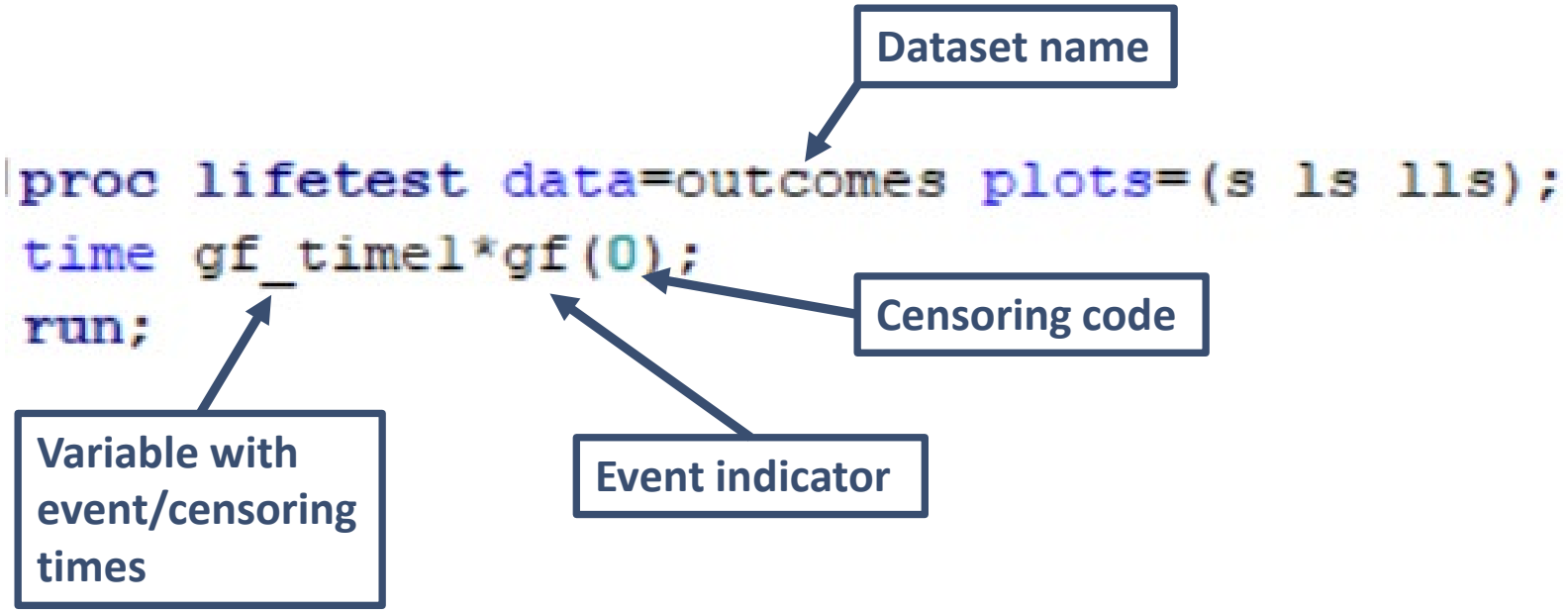
Kaplan-Meier

- Product limit estimation: calculates probability of event at each time and multiplies successive probabilities

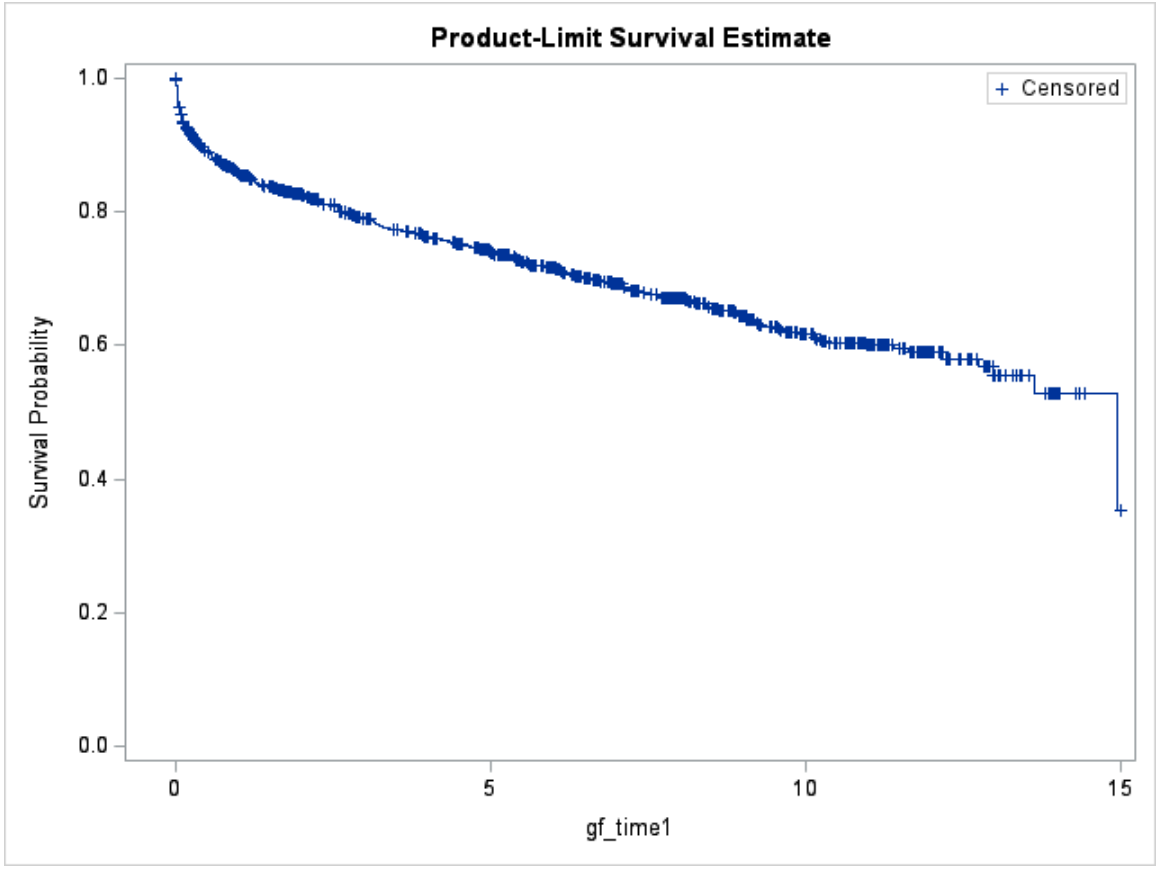
$$\hat{S}(t) = \prod_{i:t_i<t} \left(\frac{n_i - d_i}{n_i}\right)$$

- Probability of event for censored observations is distributed forward (i.e. to the right)
- **Assumes censoring is independent of event times**

Proc Lifetest



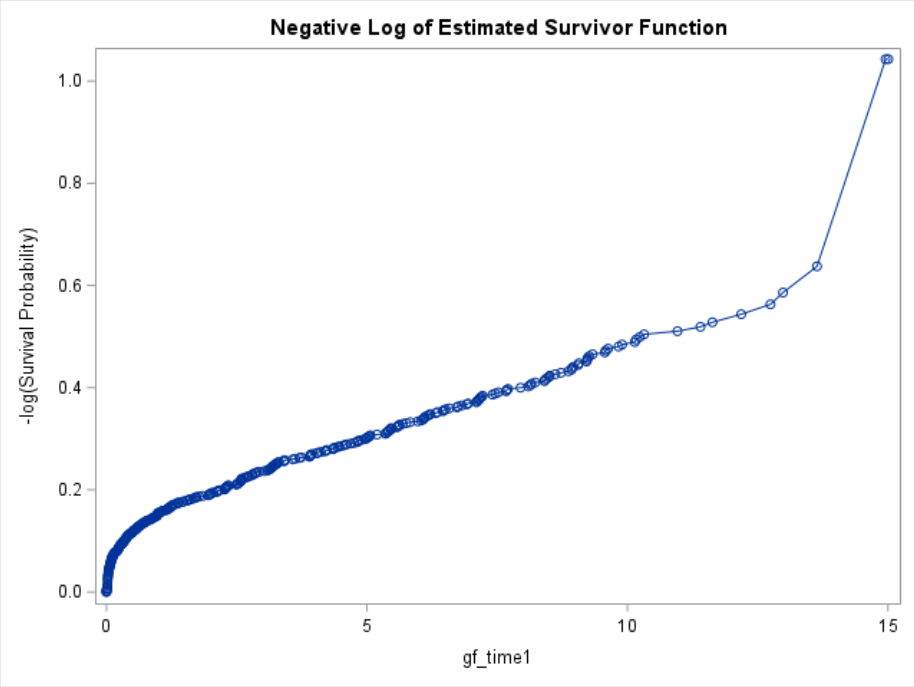
LIFETEST Output

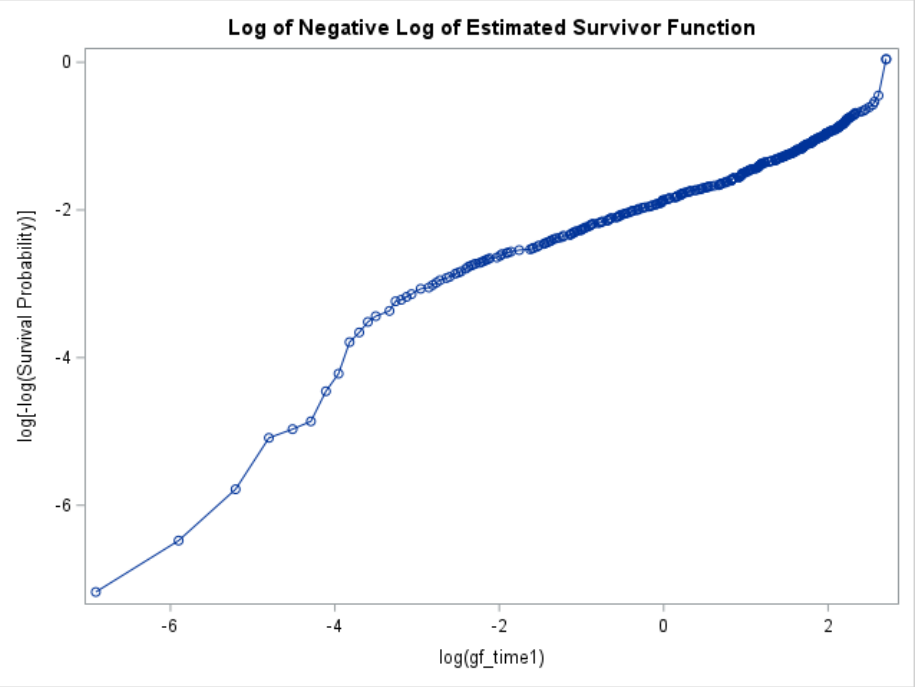


Summary of the Number of Censored and Uncensored Values			
Total	Failed	Censored	Percent Censored
1306	397	909	69.60

Quartile Estimates				
		95% Confidence Interval		
Percent	Point Estimate	Transform	[Lower	Upper)
75	.	LOGLOG	14.9541	.
50	14.9541	LOGLOG	13.6372	.
25	4.5804	LOGLOG	3.4086	5.5743

LIFETEST Plots



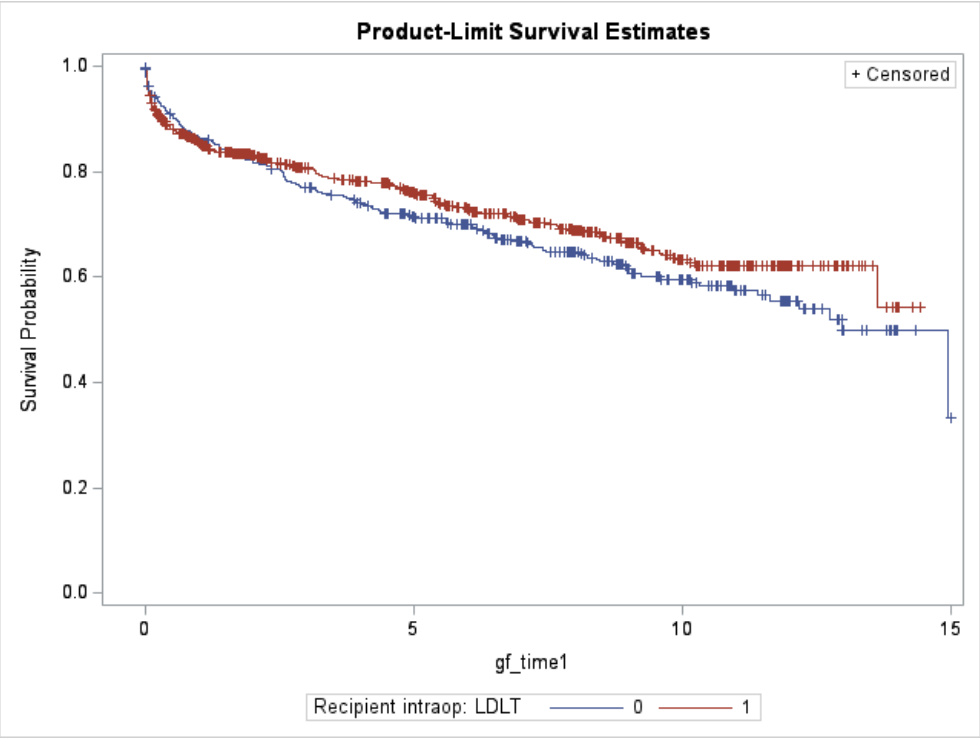


If exponential, should be approximately linear through the origin

If Weibull, should be approximately linear through the origin

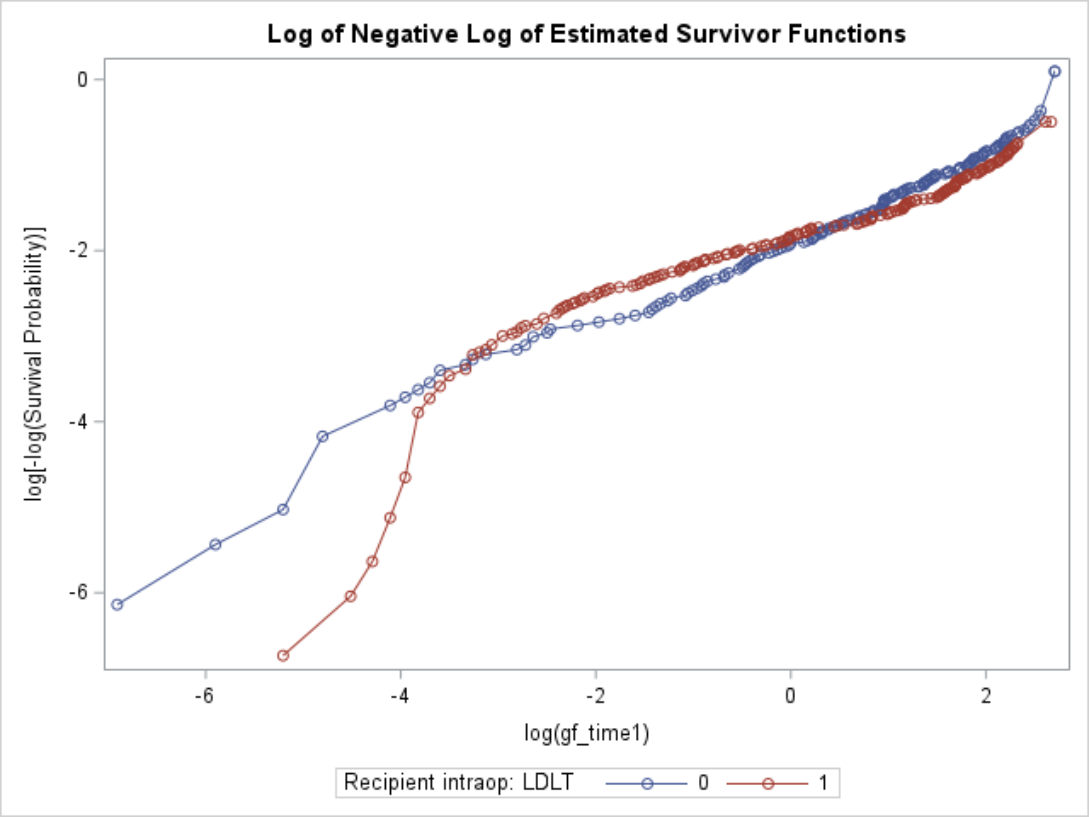
Stratification

```
proc lifetest data=outcomes plots=(s ls lls);
strata ldlt_rcx;
time gf_time1*gf(0);
run;
```



Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	2.1390	1	0.1436
Wilcoxon	0.8259	1	0.3635
-2Log(LR)	0.3585	1	0.5493

Checking Proportional Hazards



Lines should be parallel if proportional hazards assumption is valid

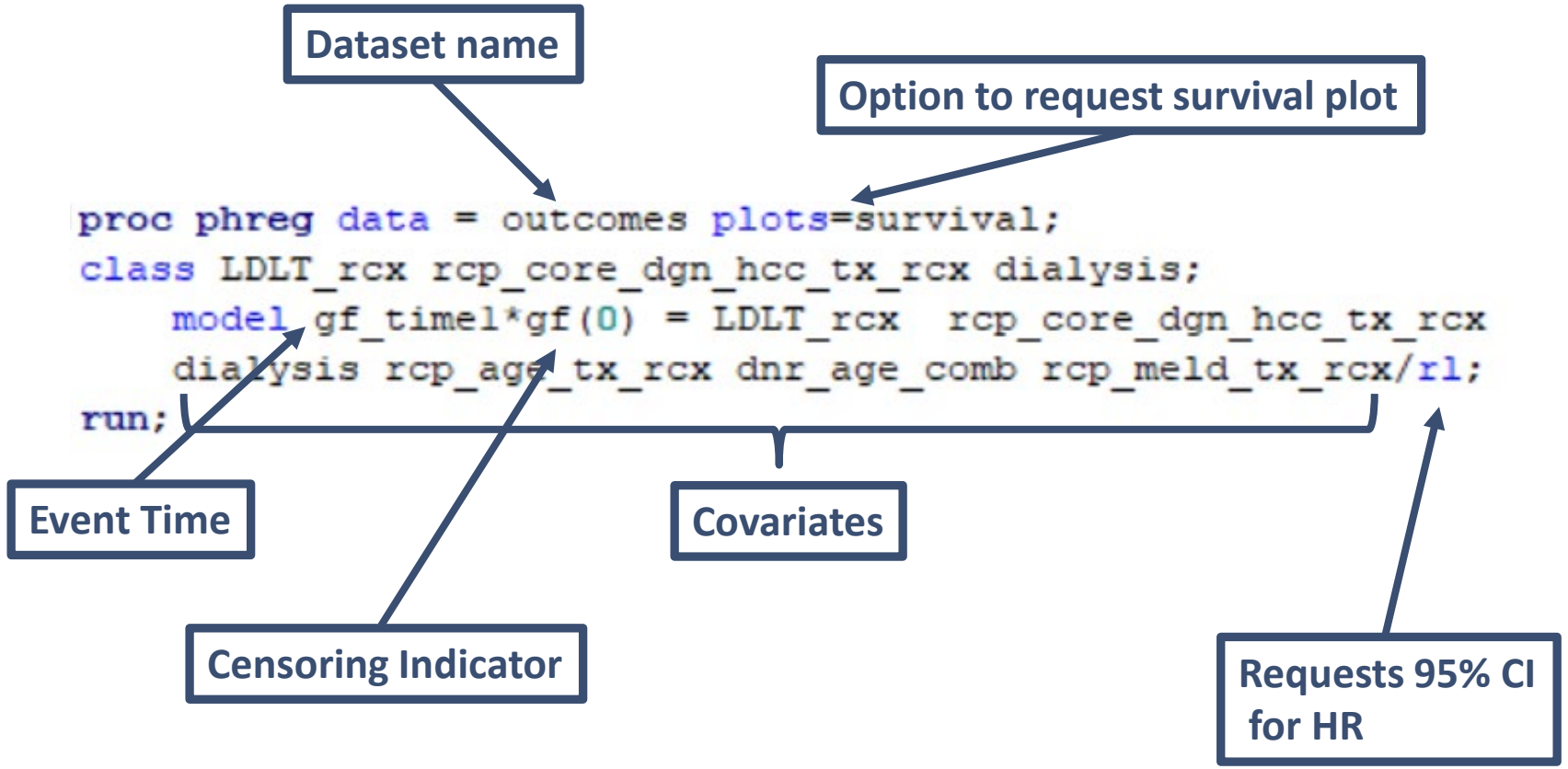
Cox Proportional Hazards Models

- Semi-parametric model – no assumptions about the shape of the baseline hazard

$$\lambda(t) = \lambda_0(t)e^{x^T\beta}$$

- Key assumptions:
  - Independence of survival times
  - Multiplicative relationship between covariates and the hazard
  - Constant hazard over time

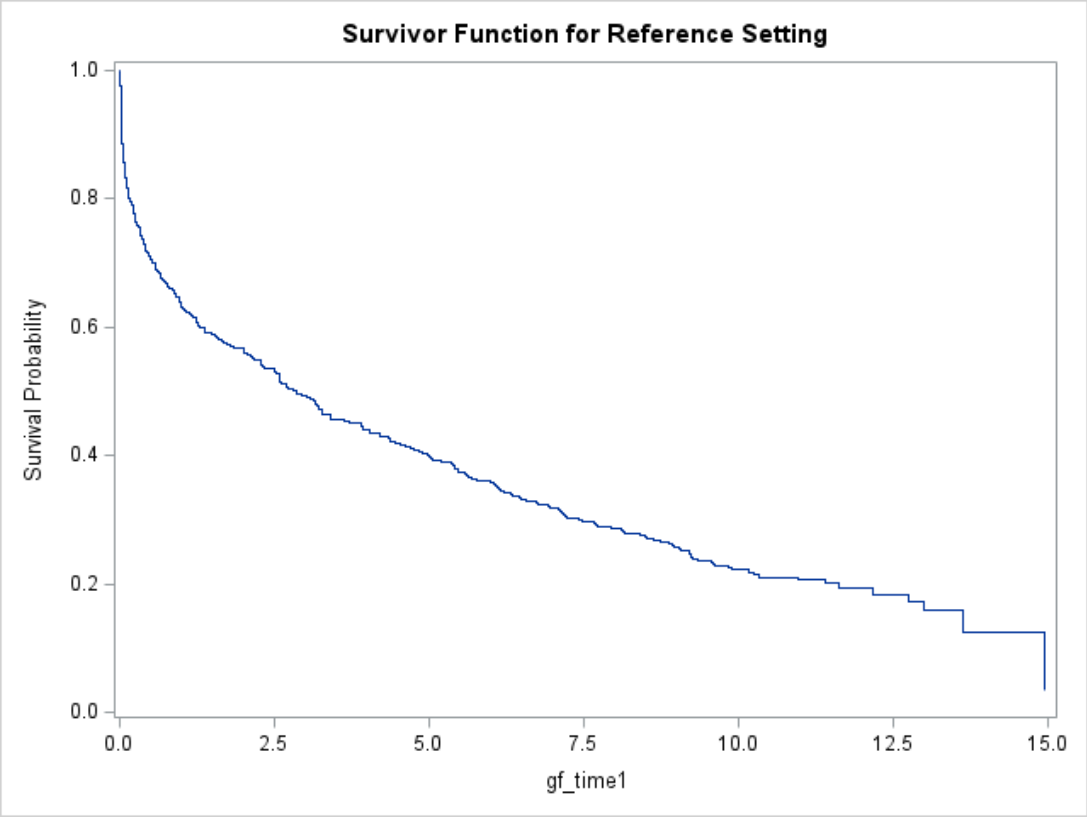
Proc Phreg


Dataset name
Option to request survival plot
proc phreg data = outcomes plots=survival;
class LDLT_rcx rcp_core_dgn_hcc_tx_rcx dialysis;
model gf_time1*gf(0) = LDLT_rcx rcp_core_dgn_hcc_tx_rcx
dialysis rcp_age_tx_rcx dnr_age_comb rcp_meld_tx_rcx/rl;
run;
Event Time
Covariates
Censoring Indicator
Requests 95% CI for HR

Phreg Output

Analysis of Maximum Likelihood Estimates									
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	Label
LDLT_rcx	0	1	-0.09779	0.11582	0.7128	0.3985	0.907	0.723 1.138	Recipient intraop: LDLT 0
rcp_core_dgn_hcc_tx_	0	1	-0.36686	0.13773	7.0950	0.0077	0.693	0.529 0.908	Recipient at enr/list/dnr eval/pretx assess/txp: any dgn: Hepatocellular Carcinoma (HCC) (merges with Core) 0
dialysis	0	1	-0.86353	0.26876	10.3235	0.0013	0.422	0.249 0.714	dialysis 0
rcp_age_tx_rcx		1	0.01480	0.00520	8.0941	0.0044	1.015	1.005 1.025	Recipient at transplant: age
dnr_age_comb		1	0.01601	0.00398	16.1924	<.0001	1.016	1.008 1.024	
rcp_meld_tx_rcx		1	0.01672	0.00786	4.5187	0.0335	1.017	1.001 1.033	Recipient at transplant: MELD at transplant (max 40)

Phreg Survival Curve



Reference Set of Covariates for Plotting					
rcp_age_tx_rcx	dnr_age_comb	rcp_meld_tx_rcx	LDLT_rcx	rcp_core_dgn_hcc_tx_rcx	dialysis
51.695523	37.942945	17.100649	1	1	1

Modifying Adjusted Survival Curves

Step 1: Create dataset with reference values

```
data covs;
input LDLT_rcx  rcp_core_dgn_hcc_tx_rcx
    dialysis rcp_age_tx_rcx dnr_age_comb rcp_meld_tx_rcx;
    datalines;
    0 0 0 50 35 17
    1 0 0 50 35 17
    ;
run;

proc phreg data = outcomes plots(overlay)=survival;
class LDLT_rcx rcp_core_dgn_hcc_tx_rcx dialysis;
    model gf_time1*gf(0) = LDLT_rcx  rcp_core_dgn_hcc_tx_rcx
    dialysis rcp_age_tx_rcx dnr_age_comb rcp_meld_tx_rcx/rl;
    baseline covariates=covs out=base/rowid=ldlt_rcx;
run;
```

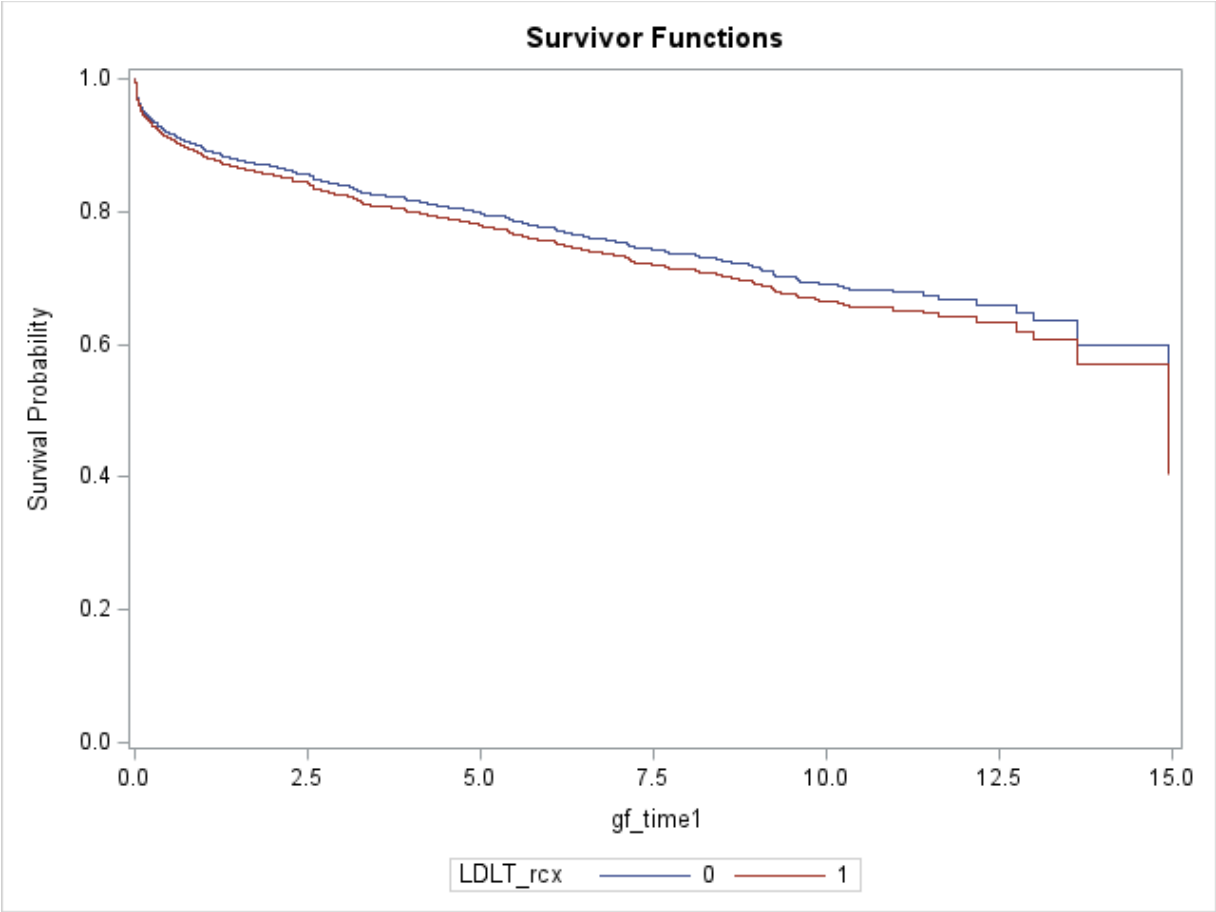
Requests stratified plot

Reference Dataset

Output Dataset

Stratification Variable

Modifying Adjusted Survival Curves



Interaction Terms

- Interactions are used to assess effect modification, i.e. does the effect of a variable on the outcome differ by levels of a second variable

```
proc phreg data = outcomes plots=survival;
class LDLT_rcx rcp_core_dgn_hcc_tx_rcx dialysis;
    model gf_time1*gf(0) = LDLT_rcx  rcp_core_dgn_hcc_tx_rcx
    dialysis rcp_age_tx_rcx|dnr_age_comb rcp_meld_tx_rcx/rl;
    hazardratio 'Recipient Age by Donor Age' rcp_age_tx_rcx/at(dnr_age_comb=(25 35 45 50)) units=10;
run;
```

Results

Analysis of Maximum Likelihood Estimates									
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	
LDLT_rcx	0	1	-0.07624	0.11612	0.4311	0.5114	0.927	0.738	1.163
rcp_core_dgn_hcc_tx_	0	1	-0.34856	0.13757	6.4193	0.0113	0.706	0.539	0.924
dialysis	0	1	-0.87717	0.26884	10.6456	0.0011	0.416	0.246	0.705
rcp_age_tx_rcx		1	0.04859	0.01728	7.9030	0.0049	.	.	.
dnr_age_comb		1	0.05992	0.02155	7.7339	0.0054	.	.	.
rcp_age_t*dnr_age_co		1	-0.0008244	0.0003975	4.3025	0.0381	.	.	.
rcp_meld_tx_rcx		1	0.01541	0.00786	3.8418	0.0500	1.016	1.000	1.031

Effect of Recipient Age Across Donor Ages: Hazard Ratios for Recipient at transplant: age			
Description	Point Estimate	95% Wald Confidence Limits	
rcp_age_tx_rcx Unit=10 At dnr_age_comb=25	1.323	1.123	1.558
rcp_age_tx_rcx Unit=10 At dnr_age_comb=35	1.218	1.087	1.365
rcp_age_tx_rcx Unit=10 At dnr_age_comb=45	1.122	1.009	1.246
rcp_age_tx_rcx Unit=10 At dnr_age_comb=50	1.076	0.953	1.216

Assumption Checking – Functional Form

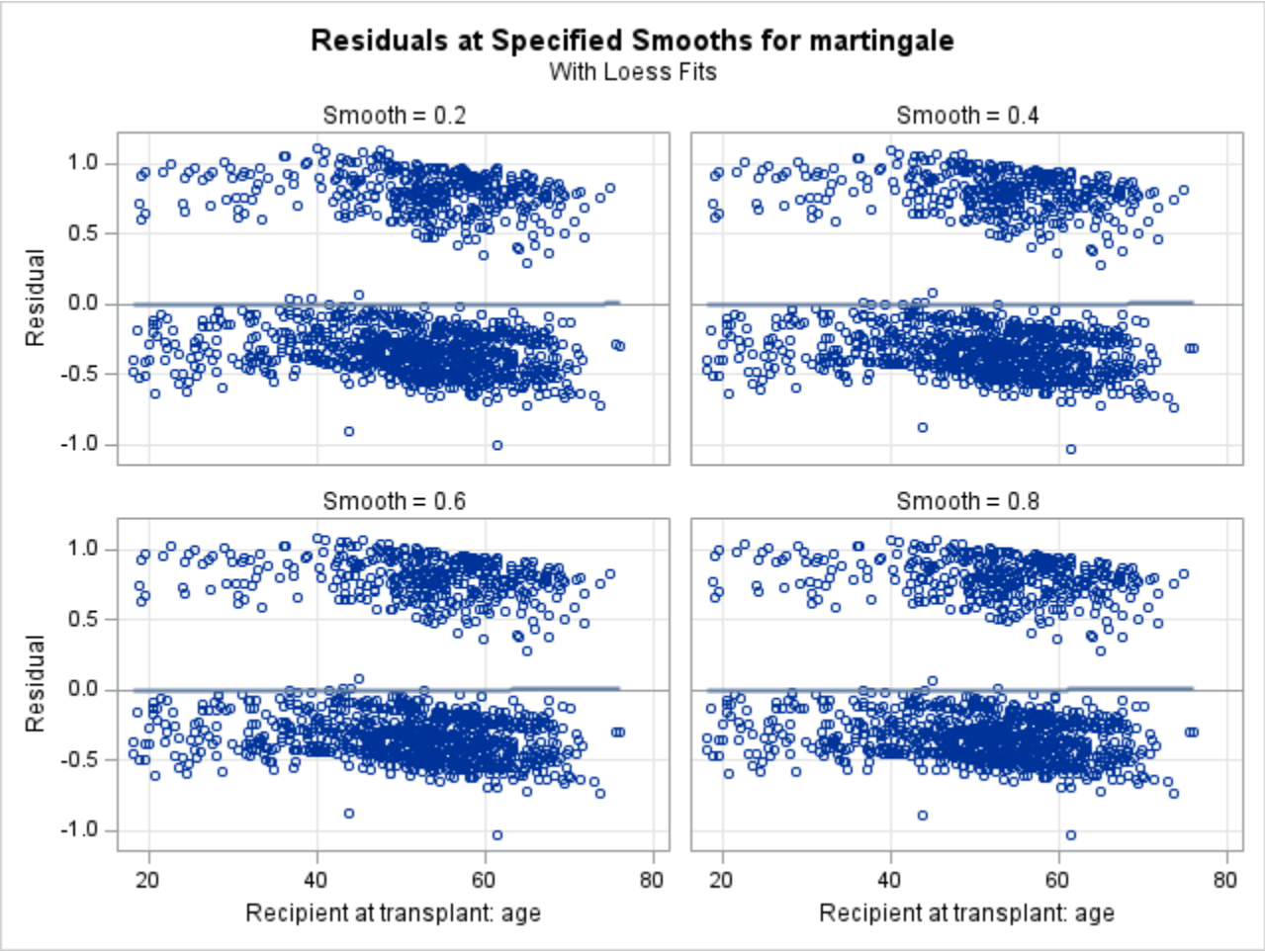
Method 1 – Martingale residuals

```
proc phreg data=outcomes;
model gf_time1*gf(0)=;
output out=residuals resmart=martingale;
run;
```

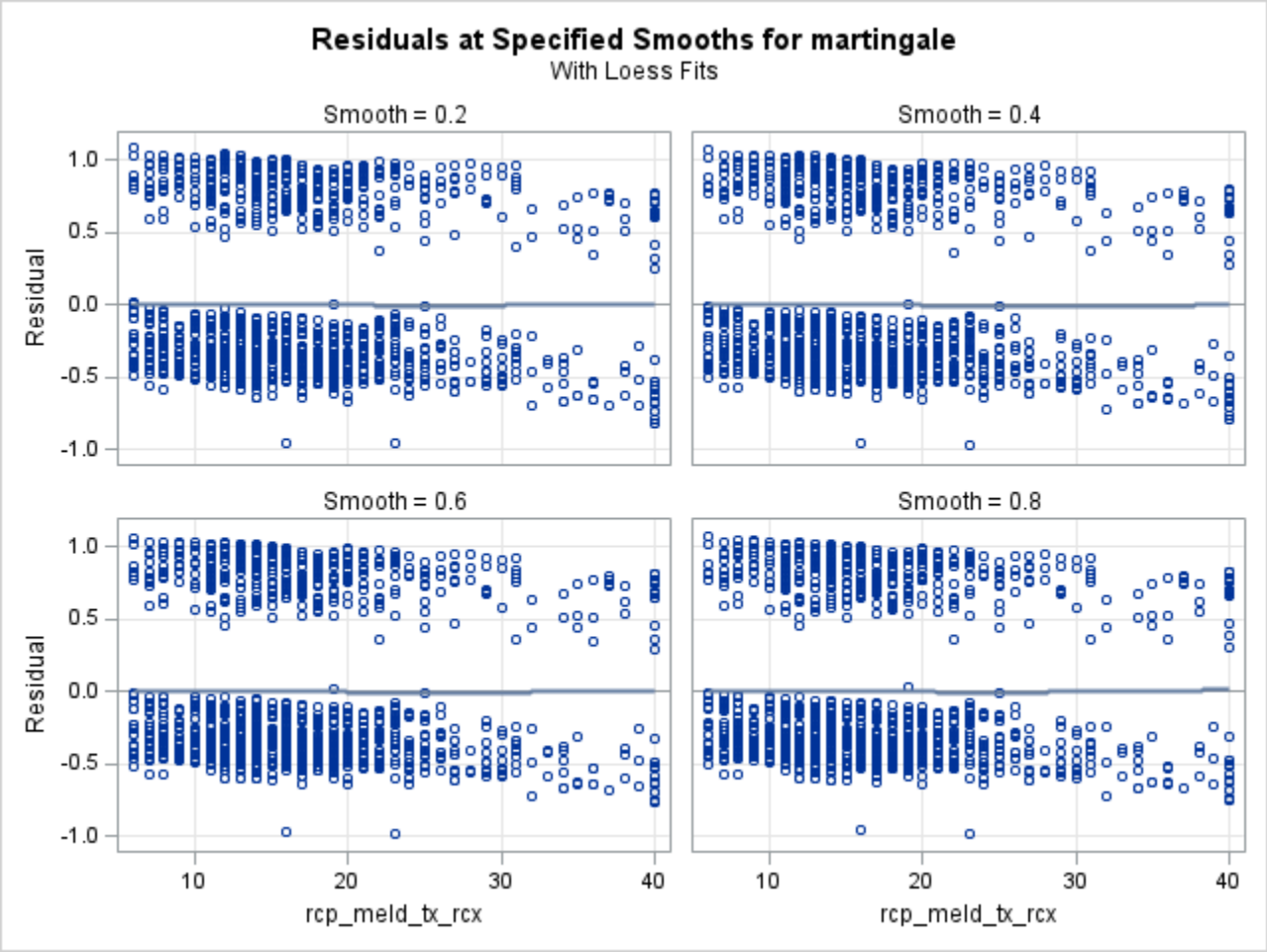
```
proc loess data = residuals plots=ResidualsBySmooth(smooth);
model martingale = rcp_age_tx_rcx / smooth=0.2 0.4 0.6 0.8;
run;
```

```
proc loess data = residuals plots=ResidualsBySmooth(smooth);
model martingale = rcp_meld_tx_rcx / smooth=0.2 0.4 0.6 0.8;
run;
```

Recipient Age



MELD


Residuals at Specified Smooths for martingale
With Loess Fits
Smooth = 0.2
Smooth = 0.4
Smooth = 0.6
Smooth = 0.8
Residual
1.0
0.5
0.0
-0.5
-1.0
10
20
30
40
rcp_meld_tx_rcx

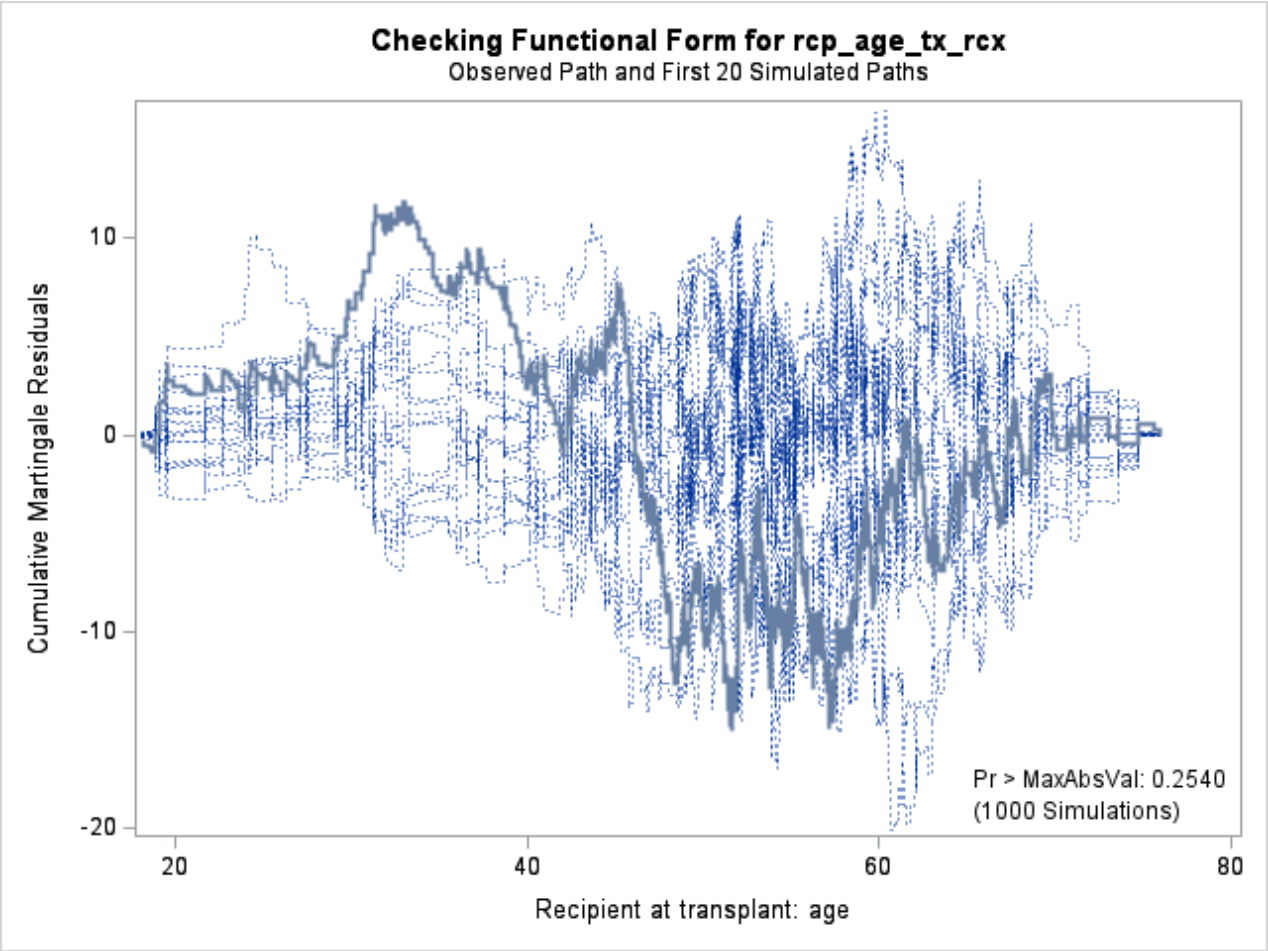
Assumption Checking – Functional Form

Method 1 – ASSESS statement

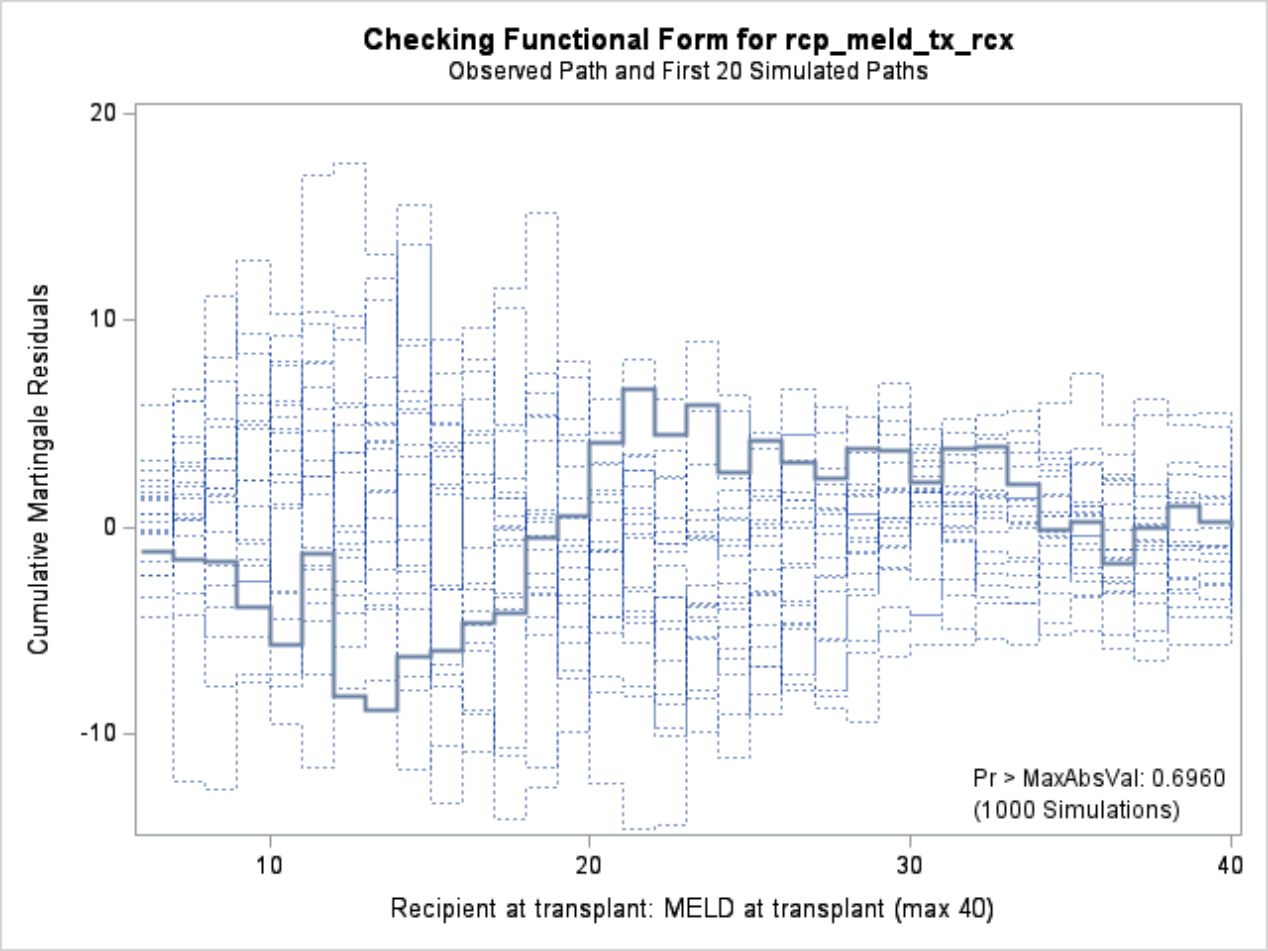
```
proc phreg data = outcomes ;
class LDLT_rcx rcp_core_dgn_hcc_tx_rcx dialysis/ref=first;
    model gf_time1*gf(0) = LDLT_rcx  rcp_core_dgn_hcc_tx_rcx
    dialysis rcp_age_tx_rcx dnr_age_comb rcp_meld_tx_rcx/rl;
    assess var=(rcp_age_tx_rcx dnr_age_comb rcp_meld_tx_rcx)/resample;
run;
```

Supremum Test for Functional Form				
Variable	Maximum Absolute Value	Replications	Seed	Pr > MaxAbsVal
rcp_age_tx_rcx	14.9069	1000	677662574	0.2540
dnr_age_comb	9.7880	1000	677662574	0.7750
rcp_meld_tx_rcx	8.8101	1000	677662574	0.6960

Recipient Age



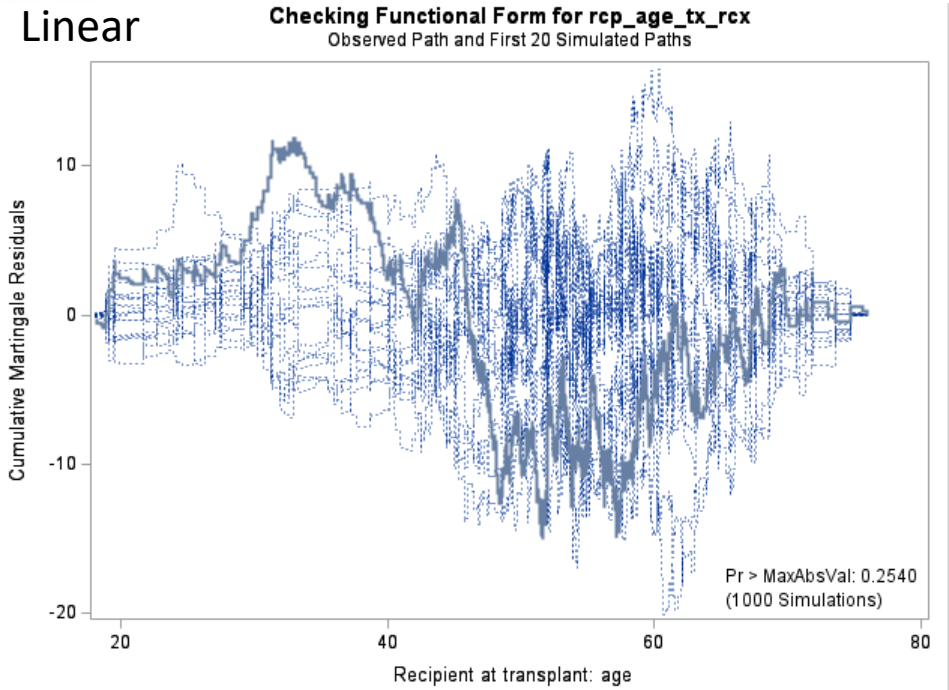
MELD


Checking Functional Form for rcp_meld_tx_rcx
Observed Path and First 20 Simulated Paths
Cumulative Martingale Residuals
20
10
0
-10
10
20
30
40
Recipient at transplant: MELD at transplant (max 40)
Pr > MaxAbsVal: 0.6960
(1000 Simulations)

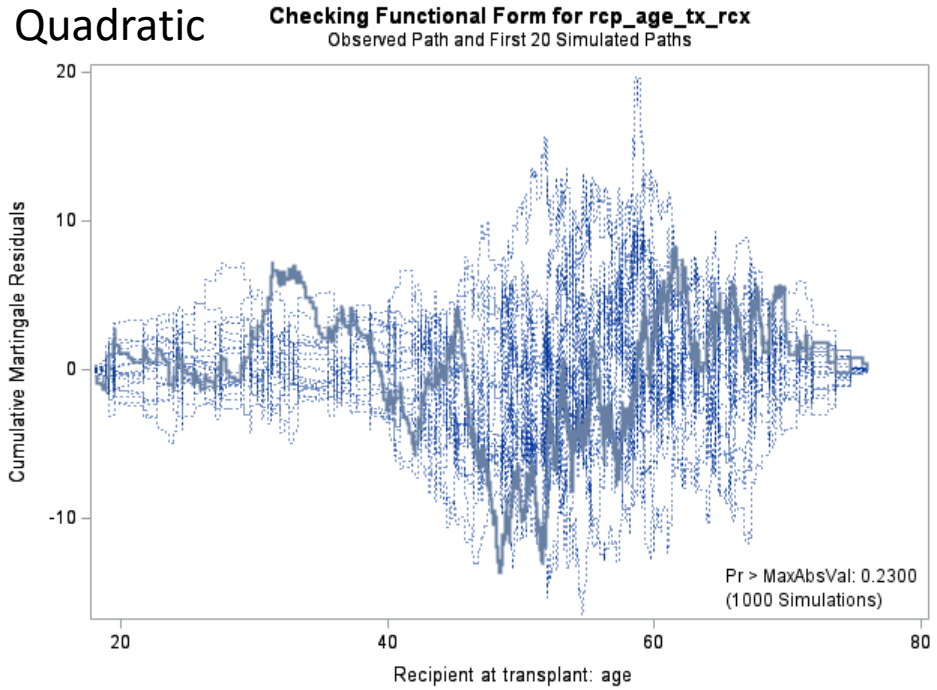
Recipient Age – Non-linear

```
proc phreg data = outcomes ;
class LDLT_rcx rcp_core_dgn_hcc_tx_rcx dialysis/ref=first;
    model gf_time1*gf(0) = LDLT_rcx  rcp_core_dgn_hcc_tx_rcx
    dialysis rcp_age_tx_rcx|rcp_age_tx_rcx dnr_age_comb rcp_meld_tx_rcx/rl;
    assess var=(rcp_age_tx_rcx dnr_age_comb rcp_meld_tx_rcx)/resample;
run;
```

Linear



Quadratic



Assumption Checking – Proportional Hazards

- Interactions with time
- Schoenfeld residuals
- Assess statement

Interactions with Time

```
proc phreg data = outcomes ;
    model gf_time1*gf(0) = LDLT_rcx LDLT_rcx_time rcp_core_dgn_hcc_tx_rcx
    dialysis rcp_age_tx_rcx dnr_age_comb  rcp_meld_tx_rcx/rl;
    LDLT_rcx_time=LDLT_rcx*gf_time1;
run;
```

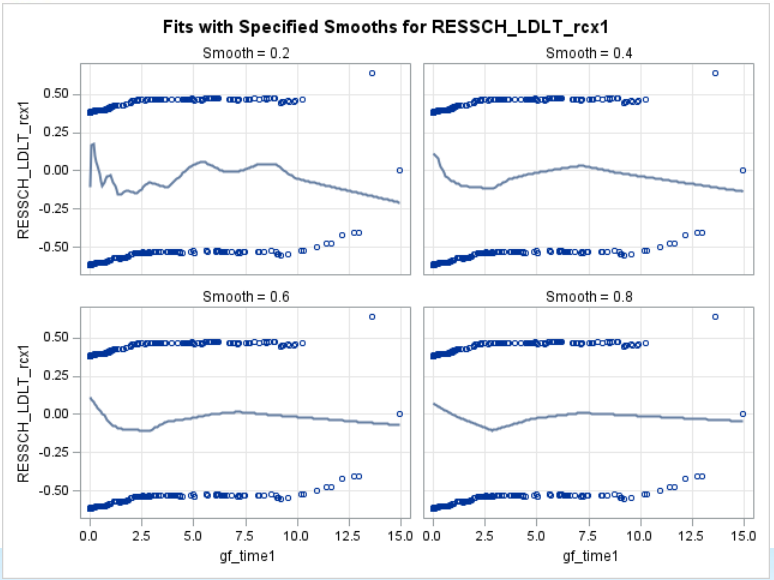
Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits		Label
LDLT_rcx	1	0.17486	0.15089	1.3428	0.2465	1.191	0.886	1.601	Recipient intraop: LDLT
LDLT_rcx_time	1	-0.02710	0.03377	0.6441	0.4222	0.973	0.911	1.040	
rcp_core_dgn_hcc_tx_	1	0.36670	0.13772	7.0901	0.0078	1.443	1.102	1.890	Recipient at enr/list/dnr eval/pretx assess/txp: any dgn: Hepatocellular Carcinoma (HCC) (merges with Core)
dialysis	1	0.87280	0.26899	10.5279	0.0012	2.394	1.413	4.055	
rcp_age_tx_rcx	1	0.01473	0.00520	8.0225	0.0046	1.015	1.005	1.025	Recipient at transplant: age
dnr_age_comb	1	0.01603	0.00398	16.2343	<.0001	1.016	1.008	1.024	
rcp_meld_tx_rcx	1	0.01659	0.00787	4.4429	0.0350	1.017	1.001	1.033	Recipient at transplant: MELD at transplant (max 40)

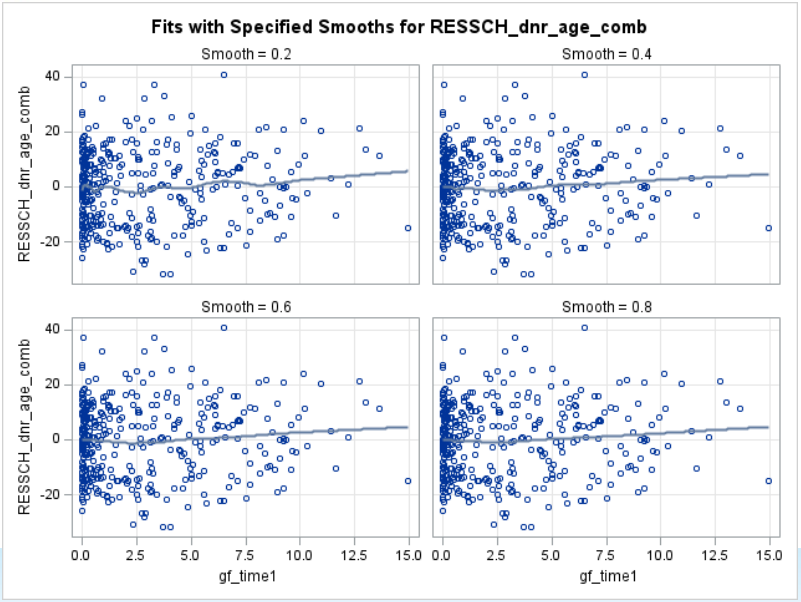
Schoenfeld Residuals

```
proc phreg data = outcomes ;
    class LDLT_rcx rcp_core_dgn_hcc_tx_rcx dialysis/ref=first;
    model gf_time1*gf(0) = LDLT_rcx rcp_core_dgn_hcc_tx_rcx
    dialysis rcp_age_tx_rcx dnr_age_comb  rcp_meld_tx_rcx/rl;
    output out=schoen ressch= _ALL_ ;
run;
```

```
proc loess data = schoen;
model ressch_LDLT_rcx1=gf_time1 / smooth=(0.2 0.4 0.6 0.8);
run;
```

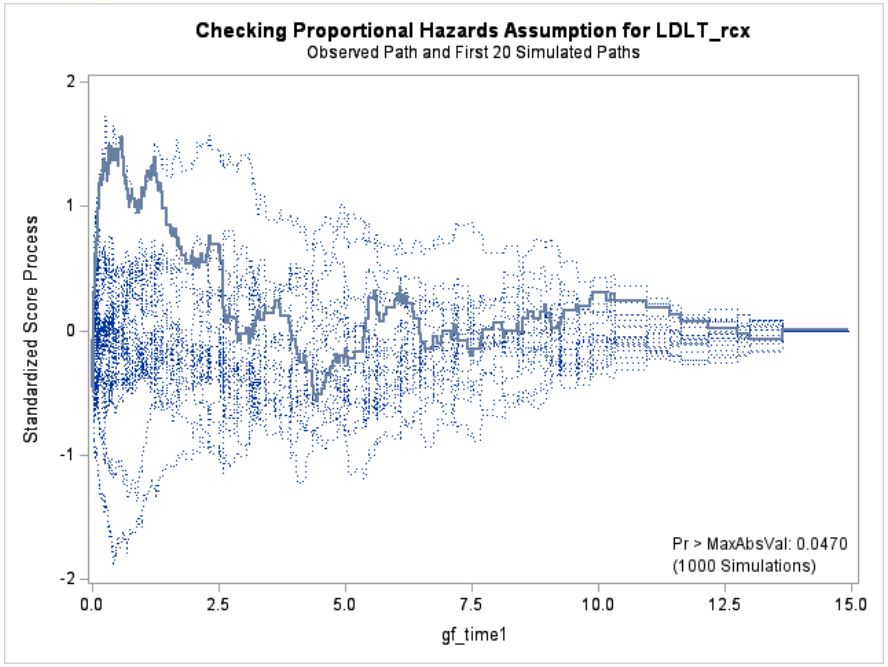


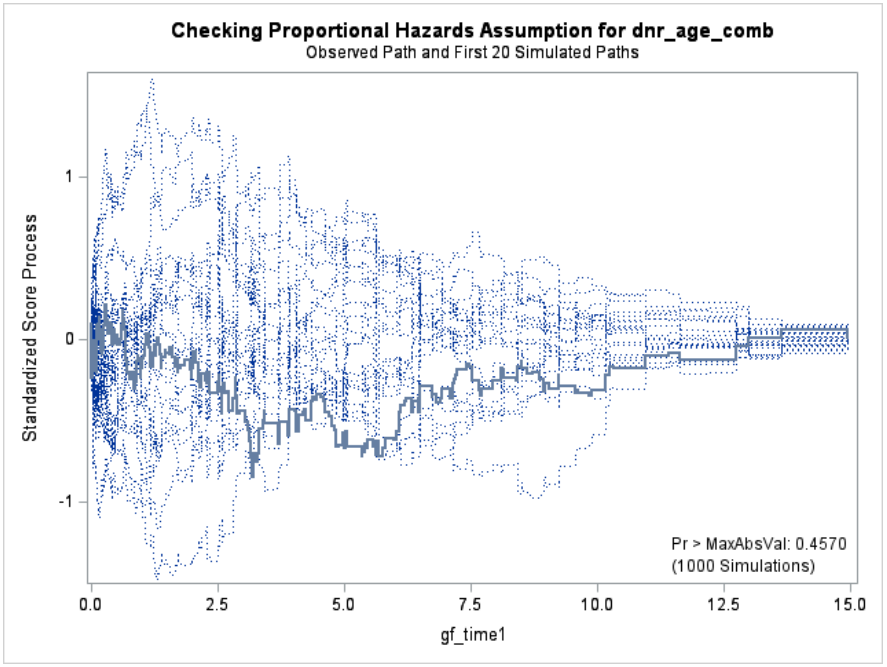

```
proc loess data = schoen;
model ressch_dnr_age_comb=gf_time1 / smooth=(0.2 0.4 0.6 0.8);
run;
```



Assess Statement

```
proc phreg data = outcomes ;
    model gf_time1*gf(0) = LDLT_rcx  rcp_core_dgn_hcc_tx_rcx
    dialysis rcp_age_tx_rcx dnr_age_comb rcp_meld_tx_rcx/rl;
    assess var=(LDLT_rcx  rcp_core_dgn_hcc_tx_rcx
    dialysis rcp_age_tx_rcx dnr_age_comb rcp_meld_tx_rcx) ph/resample;
run;
```





Dealing with non-proportional hazards

- Stratification
  - Cannot estimate hazard ratio directly but can assess interactions
- Interactions with time
  - Important to assess functional form of time
- Use shorter time intervals
  - PH may hold for certain intervals of time

Extensions of Survival Analysis

- Time-dependent covariates
  - E.g. time-varying lab values
- Recurrent events
  - E.g. infection, hospitalization

Thank you! Questions?

Abigail.Smith@ArborResearch.org