Simulation and Survival Analysis of Breast Cancer

Liela Meng*

Department of Biostatistics, University of California, Davis, ONE SHIELDS AVENUE, Davis, 95616, California, United States. Tel.: 5309336513

Abstract

Background: Breast cancer (BC) is the most leading cause of cancer among females worldwide. E baskhshi's article uses various models to analyze breast cancer patients' survival and identify risk factors.

Objectives: This study simulates 500 observations based on the Cox proportional hazard model. The primary hypothesis of survival analysis for the simulated data set was to investigate whether patients who received Modified Radical Mastectomy (MRM) had a lower risk of death from BC than patients who underwent Breast-Conserving Surgery (BCS).

Results: Tumor size, metastasis are significant factors for both the original data and simulated data. Though slightly different regarding whether age is a significant factor, all estimated coefficients were similar to the original ones. The new model build based on simulated data only has a significant factor, tumor size.

Conclusions: Simulated data was able to capture the original data's characteristics, which indicates the robustness of the CoxPH model. Both CoxPH models, with all variables, and the new model do not indicate a significant effect of surgery type for BC patients' survival.

Key words: simulation, survival analysis, breast cancer, Cox PH model

1 Introduction

The study described in the article is conducted from 2010 to 2015. Female with a confirmed diagnosis of BC who underwent Modified Radical Mastectomy (MRM) or Breast-Conserving Surgery (BCS) from March 2010 to 2014 were enrolled. All patients were followed-up to March 2015. 113 (22.6%) of patients died at the end of the study [1].

The event of interest is death from breast cancer. The recorded time is the on study time in months from the date of the BC treatment. Thus, the study has a generalized type I right censoring mechanism.

*E-mail: mzcmeng@ucdavis.edu

2 Methods

2.1 Notation

$h(t) = h_0 \exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9)$

X_1 represents the kind of surgery, which equals 0 when the patients had radical mastectomy surgery. X_2, X_3 represent age and tumor size respectively. X_4equals 0 if metastasis is no. X_5, X_6, X_7 are indicator variables of variable stage corresponding to stage II, III, IV, respectively. X_8 equals 0 if the pathology is IDC. X_9 equals 0 if family history is no.

The likelihood function would be $L = \prod_{i \in D} f_i(t_i) \prod_{i \in R} S_i(C_r)$ where $S(t)$ is the survival function and $f(t)$ is the probability density function. D is the set of event times, R is the set of right-censored observations. Since in this study, the time to death follows an exponential distribution with hazard rate λ , the likelihood function would be $L = \prod_{i=1}^{D} \lambda \exp(-\lambda t_i) \prod_{i=1}^{R} \exp(-\lambda t_i)$.

2.2 Simulation process

1) Generate categorical variables randomly with a probability corresponding to the percent of each level presented in the article.
2) Generate two continuous variables following a normal distribution with known mean and standard deviation (SD).
3) Generate event time following exponential distribution where the rate of the distribution is expressed as the product of baseline hazard rate and covariates effect. The covariate effect is assumed additive in the log-hazard scale.
4) Generate censoring time following uniform distribution with intervals from 0 to maximum follow-up time.
5) Generate time which is the smaller one between event time in step 3 and censoring time in step 4.
6) Generate a censoring indicator which is one when time in step 5 is the censoring time.
7) Select 500 observations out of the simulated dataset while maintaining 80% of chosen observations are censored to be consistent with the heavy censoring characteristic mentioned in the article.

Formulas used to generate simulated data:

1) Surgery type, metastasis : $X_1 \sim Bernoulli(0.73);\ X_4 \sim B(0.77);$;
Stage: $X_5 \sim B(0.155),\ X_6 \sim B(0.633),\ X_7 \sim B(0.174);$
Pathologh, family history: $X_8 \sim B(0.12);\ X_9 \sim B(0.91)$

2) Age, tumor size: $X_2 \sim N(50.39, 11.19);\ X_3 \sim N(3.6, 2.37)$;

3) Event time: $Time \sim \exp(\lambda_0 \exp(0.339X_1 + 0.047X_2 + 0.24X_3 + 2.47X_4 - 0.207X_5 - 0.12X_6 - 0.226X_7 + 0.183X_8 + 0.464X_9)),\ \lambda_0 = 0.0001$

4) Censoring time: $C \sim Unif(0, 60)$;

5) Time: $T = min(Time, C)$

6) Censoring indicator: $censor = 1,\ if\ T = C;\ else\ censor = 0$;

2.3 Statistical analysis

Descriptive statistical analysis was carried out to explore the characteristics of patients using mean$\pm$ SD for continuous variables, and frequency for categorical variables.

Initially, the Kaplan-Meier (KM) estimator was used to compare with the given KM curve. A CoxPH model with all variables included in the model would be build to compare coefficient estimation between original data and simulated data.

To investigate the difference between the two surgery types, the Kaplan-Meier (KM) estimator was used to quantify the mean survival time. Smoothed hazard rates were plotted to graphically detect survival time differences for two kinds of surgery. Hypothesis testings were also conducted to quantify the difference in survival time between two surgeries.

Then, the proportional Cox regression model was used to analyze the survival time adjusting for other covariates. Using forward procedure and AIC, local test p-value as criteria, a CoxPH model was built. Local hypothesis testings were used to check whether any interaction terms should be included in the model. Later, residuals plots were drawn to decide the proper form of each covariate included in the model, and check the overall fitness of the final model. Most importantly, the proportional hazard assumption was assessed. The stratified CoxPH model was contrasted to address the violation of the PH assumption problem. Lastly, a likelihood ratio test was used to assess whether each stratum shares the same regression coefficient.

A significant level of 0.05 was used. All statistical analyses were conducted using SAS University Edition.

3 Results

3.1 Comparison

Covariates that have a significant p-value in the frailty model at level 0.05 are included in the data generating procedure. Those covariables are age, tumor size, family history, stage, metastasis, pathology. Besides, the treatment variable, kind of surgery, was also included for further analysis (Table 1).

Table 1 shows the parameter estimation of the original paper and simulated data. All estimations have the same direction and similar magnitude. As for significant variables, age was no longer significant but tumor size and metastasis remained significant at level 0.05.

Kaplan-Meier method was used to generate an estimated survival curve. Figures 1 and 2 show similar results. The mean survival time was 45.7 months.

Figure 5 is the smoothed hazard rate for all simulated observations, the rate decreases from 30 months to 50 months, then increases rapidly.

3.2 Surgery type effect investigation based on simulated data

Of the 205 patients, 120 (24%) of them had breast saving surgery, and 380 (76%) had radical mastectomy surgery. In breast saving patients, 32 (26.7%) were dead due to BC

TABLE 1
Characteristics of Patients with BC and Their Association with Survival Time

Risk Factors	Levels	No. (%)	Estimate [a]	P value [a]	Estimate (simulated)	P value (simulated)
Age	Mean±SD	50.39±11.19	0.047	<0.001*	0.044	0.163
Tumor Size	Mean± SD	3.6±2.37	0.24	0.001*	0.165	0.006*
Family History	No	455 (91.2)				
	Yes	44 (8.8)	0.464	0.13	0.198	0.555
Stage	I	77 (15.5)				
	II	316 (63.3)	-0.207	0.575	-0.10	0.706
	III	87 (17.4)	-0.12	0.809	-0.316	0.414
	IV	19 (3.8)	-0.226	0.776	-0.427	0.567
Metastasis	NO	386 (77.4)				
	Yes	113 (22.6)	2.47	<0.001*	2.358	<0.001*
Pathology	IDC	60 (12)				
	DCIS	439 (88)	0.183	0.728	0.308	0.349
Kind of surgery	Radical mastectomy	365 (73.1)				
	Breast saving	134 (26.9)	0.339	0.193	0.322	0.148

Abbreviations: DCIS, ductal carcinoma in situ; IDC, invasive ductal carcinoma.
[a]: Estimate in multivariate survival analysis with CoxPH model presented in the original article
*: indicated statistically significant at level 0.05

with a mean survival time of 33.3 months. In radical mastectomy patients, 68 (17.9%) were dead due to BC with a mean survival time of 46.9 months.

Figures 3 and 4 shows the KM survival curve and smoothed hazard rates. Based on those two plots, one can see radical mastectomy constantly has a better survival experience than breast saving.

Using the forward method by adding one variable at each time, choosing minimal AIC and P value<0.05 as inclusion criteria, variables included in the model are: surgery type (always included), metastasis, and tumor size. Besides, since age has been shown as a significant factor in the former article's Cox PH model, age was also included in the model (final model AIC =1025.86). No interaction terms showed significance in local hypothesis tests, thus the model does not include any interaction terms.

Martingale residuals plots (Figures 6 and 7) indicated no need to categorize age or tumor size. Log-log survival plot of surgery types (Figure 8) seems parallel, besides, surgery effect is the question we are interested in, thus no stratification would be made for surgery type. However, log-log survival plots of metastasis (Figure 9) show stratified CoxPH model of metastasis should be constructed since the two curves are not paral-

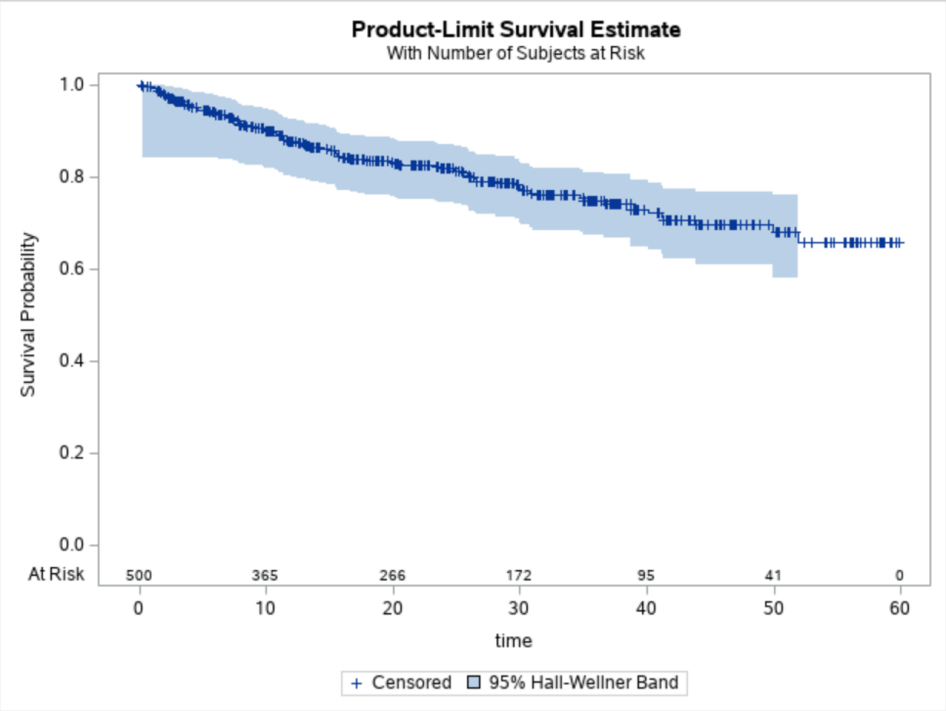

FIGURE 1: KM survival curve with simulated data

lel.Although Schoenfeld residuals (Figures 10 and 11) of tumor size and age seem a slight violation of the PH assumption, both curves vary around zero. Besides, sensitivity analysis comparing cox-nell residuals plot with adding tumor size* time and age*time separately into the stratified Cox PH model also suggested no need to add any of those two interaction terms (see Figures 12, 13, 14 and 15).

Partial likelihood function would be: $L_k(\beta) = \prod_{i=1}^{n_k} \frac{\exp(\beta^T Z_{k(i)})}{\sum_{j \in R_{kt(i)}} \exp(\beta^T Z_{kj})}$

where n_k is the number of subjects in stratum k. k=1 for metastasis equals no, and k=2 for metastasis equals yes. $\beta$is a three dimensions vector for age, tumor size and surgery type. $kt(i)$denote the event time for subject $i$at stratum k. $R_{kt(i)}$is the risk for stratum $k$set at time $t(i)$.

No interaction assumption for stratified Cox PH model was assessed via likelihood ratio test. $LRT = 895.15 - 388.079 - 504.709 = 2.362,\ P\_value = 0.5,\ df = 3$ which means each stratum of metastasis share the same regression coefficients.

New model would be:

$h_k(t) = h_{0k} \exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3),\ k = 1\ or\ 2$. where,

$h_{0,1}$for stratum metastasis equals no. X_1 represents the kind of surgery, which equals 0 when the patients conducted radical mastectomy surgery. X_2, X_3 represent age and tumor size respectively. The estimation of parameter betas is shown in Table 2

After adjusted for age, surgery type, on average, with one mm increase of tumor size, one has a 17% higher risk of dying from BC compared and I am 95% sure that the true

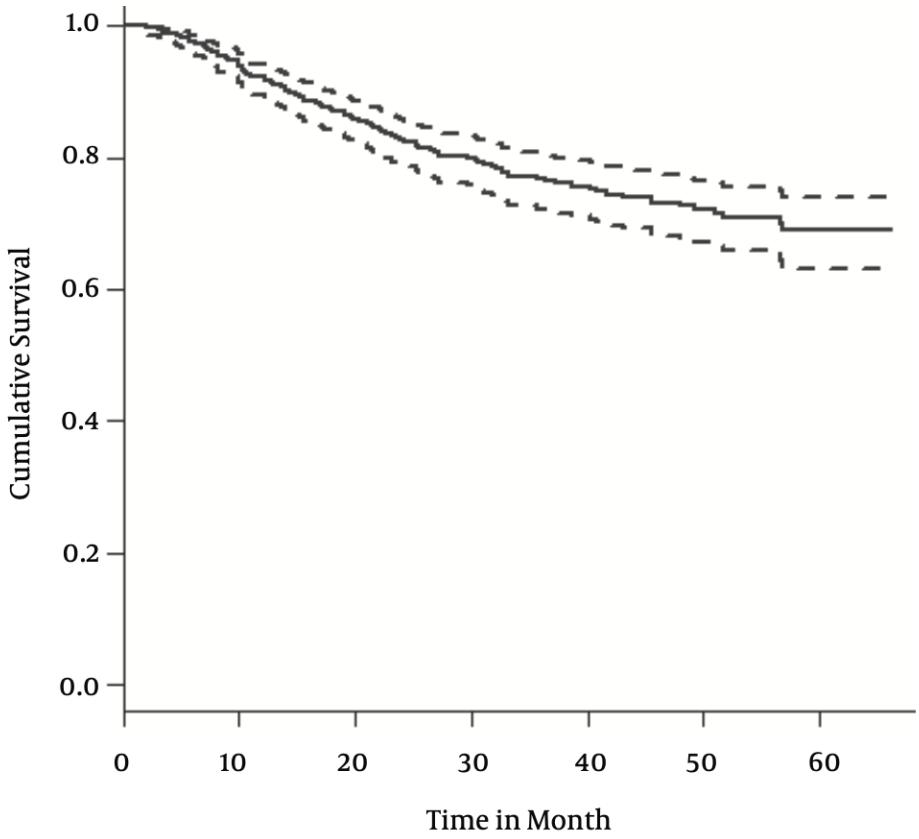

FIGURE 2: KM survival curve in the article

TABLE 2
Summary of stratified Cox PH model [a]

Parameter	Label	DF	Paramter Estimates	SE	Chi Square	P value	Hazard Ratio (95% CI)
Tumor size		1	0.154	0.06	6.58	0.01*	1.167 [1.037, 1.313]
Age		1	0.036	0.03	1.38	0.24	1.037 [0.976, 1.101]
Surgery	breast saving	1	0.33	0.22	2.40	0.12	1.397 [0.916, 2.133]

[a]: stratified on metastasis (yes or no)
*: Indicate statistically significant at level 0.05.

value is lying between 4%-31%.

Two-sample log-rank, Wilcoxon, and fleming(1,0) test of surgery type all showed a significant difference between two types (p-value=0.03). Cox PH model, after adjusting for age, tumor size, metastasis, indicated no significant effect of surgery type on survival experience for BC patients.

4 Discussion

Full model (including all variables in the model) build with simulated data have similar parameter estimations to ones presented in the original article. But age was no longer a significant factor for the survival of BC. Since the final was stratified on metastasis, the effect of metastasis could be not investigated.

For assumptions of the Cox PH model, the simulation process can guarantee independent identically distributed observations and non-informative censoring. Constant hazard ratio over time assumption was adjusted for metastasis by stratification. While age and tumor size might do not hold the PH assumption, no appropriate remedy has been discovered and taken. Besides, based on cox-snell residuals plots, the model, simply stratifies metastasis, seems fits data well overall. Similarly, the PH assumption also might not hold for surgery type. However, since the primary question is regarding surgery type, the stratification method would not work. As for the no interaction assumption for the stratified CoxPH model, the likelihood ratio test indicateds it holds.

5 Conclusions

The simulation was based on Cox proportional hazard model with exponential distribution works well for a given, highly censored, data. The primary hypothesis regarding surgery type affects the survival experience of breast cancer patients could not be verified. Further study should investigate whether age and tumor size are time-dependent variables and how to adjust for those two if the answer is yes. Besides, adding interaction of surgery type with a function of time, which might help the non-proportional hazard issue, requires further investigation.

A Appendix - SAS output

B Appendix - SAS code

```
**********
Import data
**********;
LIBNAME project2 "/folders/myfolders/BST222/project2";
    proc import datafile="/folders/myfolders/BST222/project2/data0.csv"
out=project2.data0 dbms=csv;
```

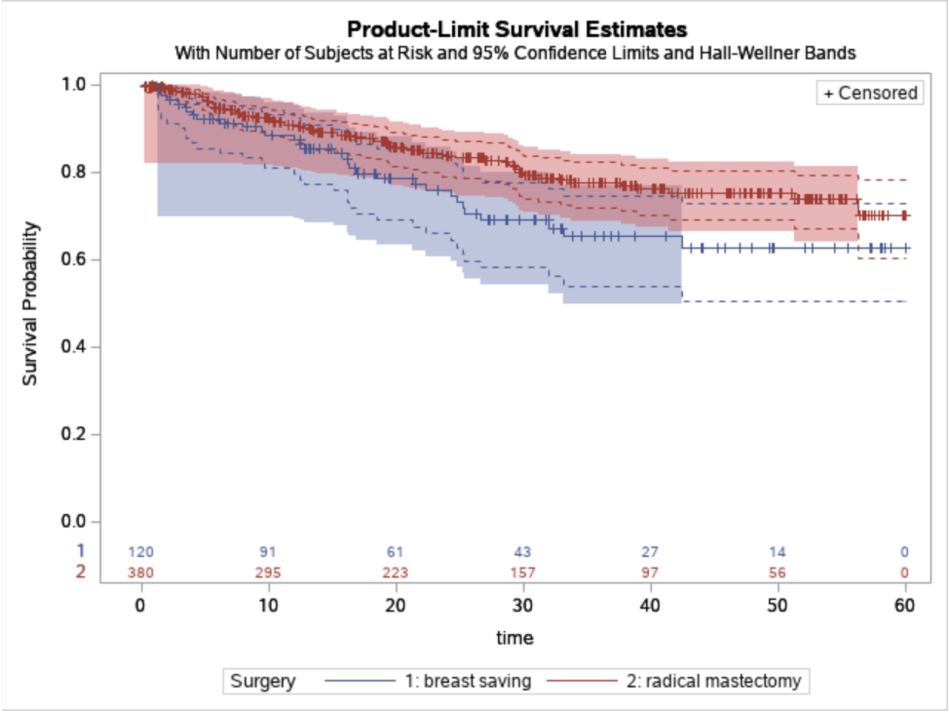

FIGURE 3: KM survival function for two surgery types

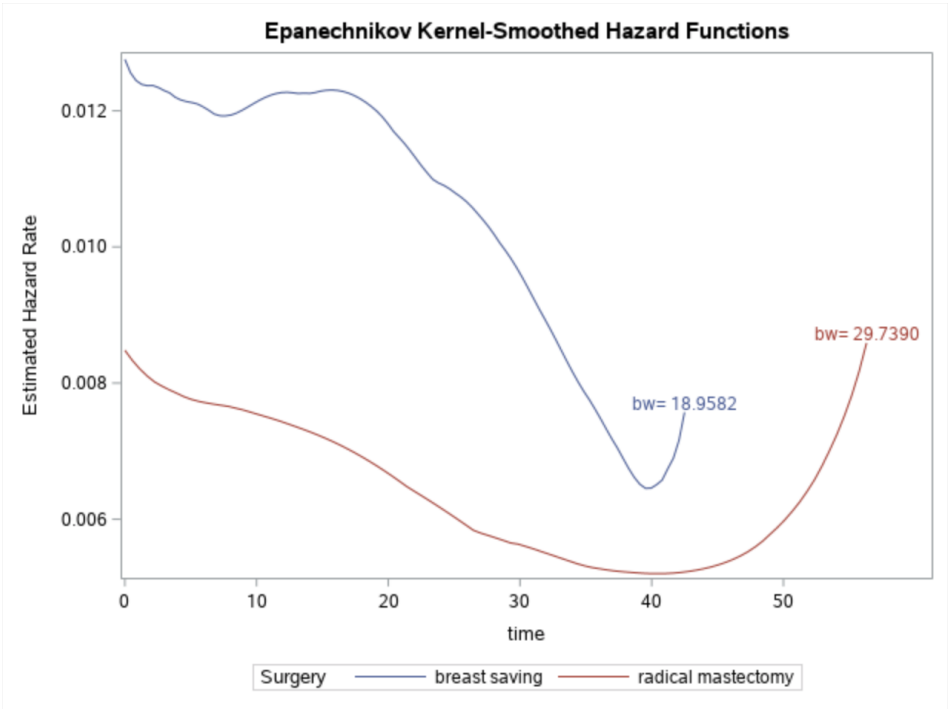

FIGURE 4: Smoothed hazard rates for two surgery types

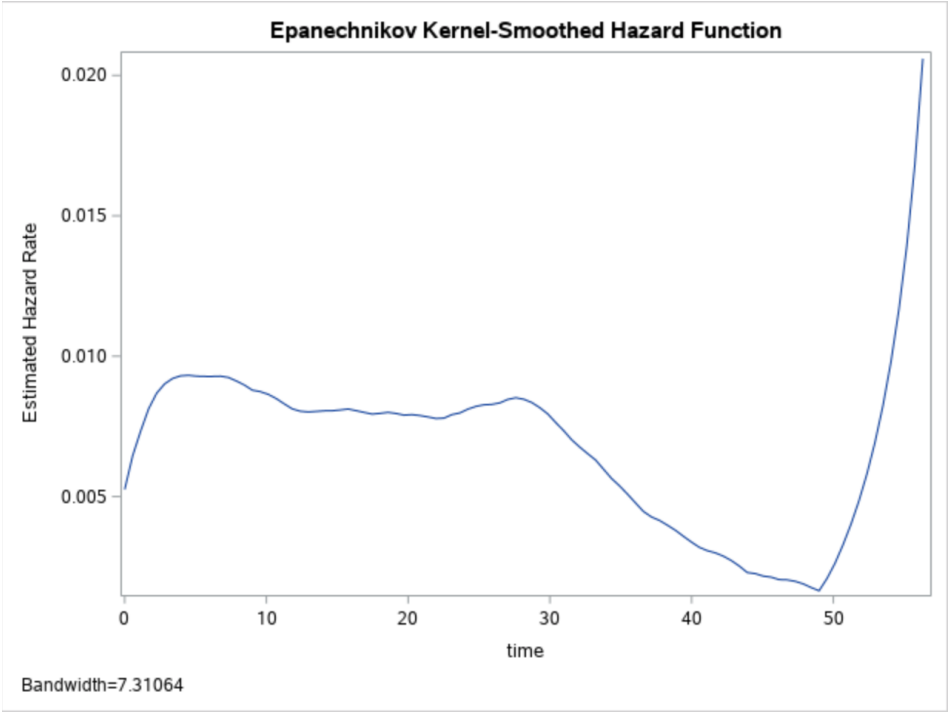

FIGURE 5: Smoothed hazard rate

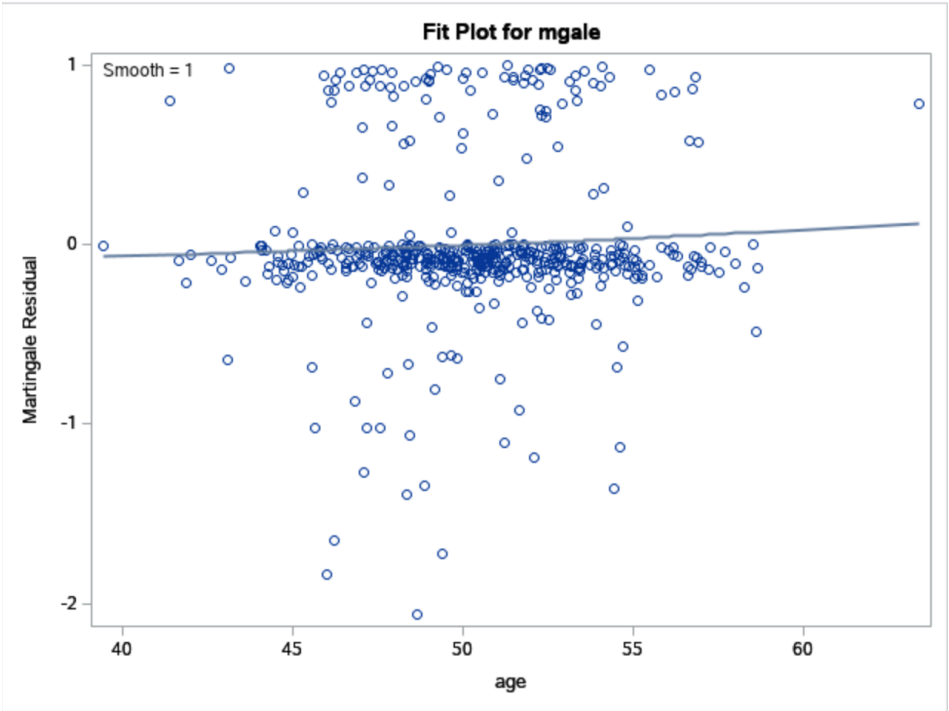

FIGURE 6: Martingale residuals plot of age

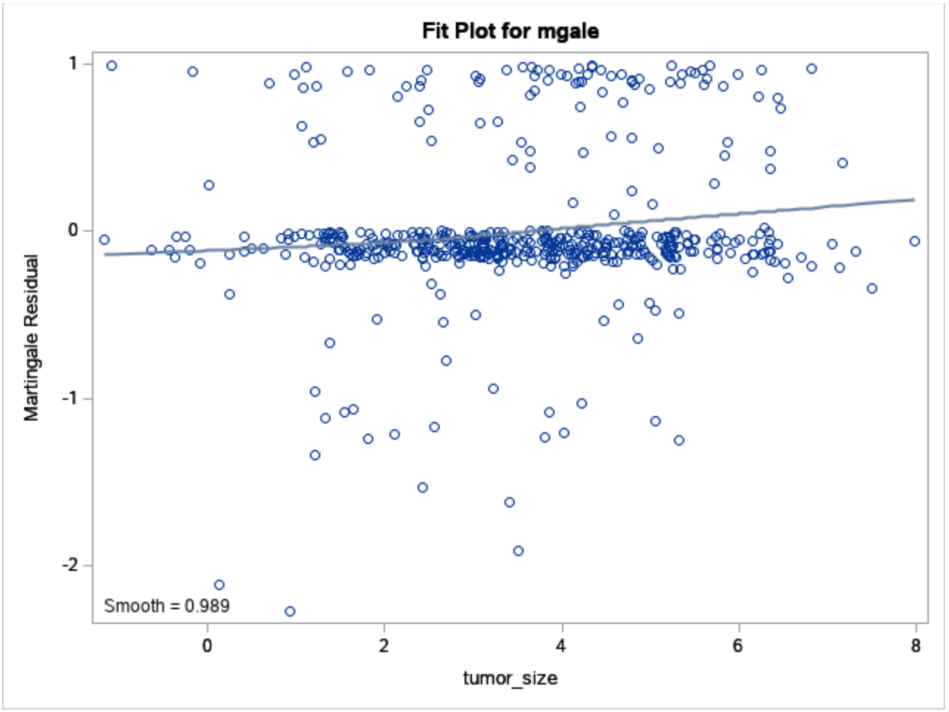

FIGURE 7: Martingale residuals of tumor size

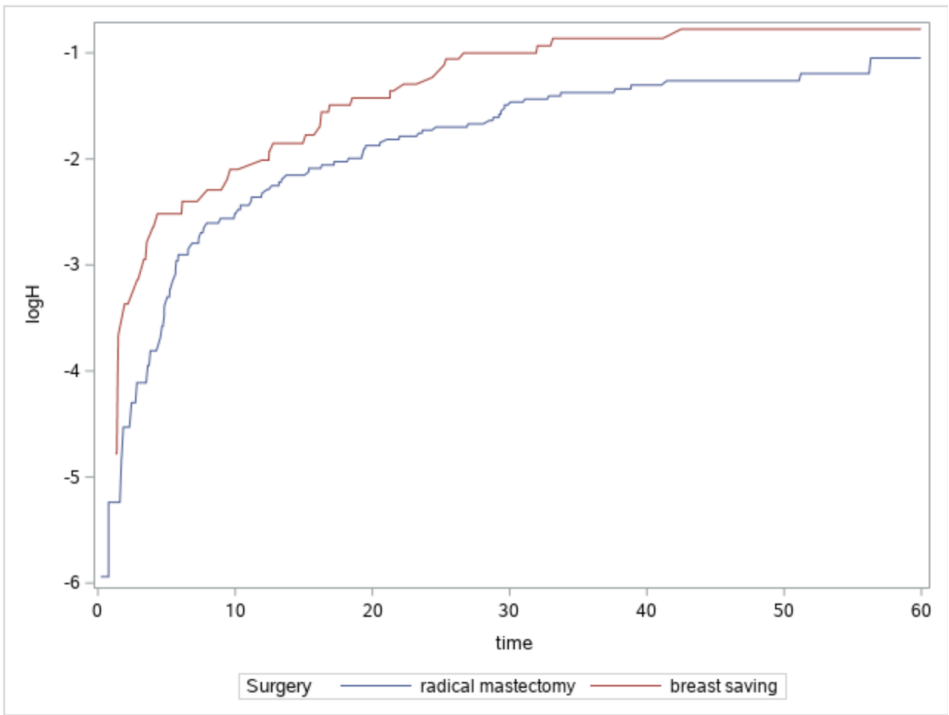

FIGURE 8: Log-log survival plot for surgery type

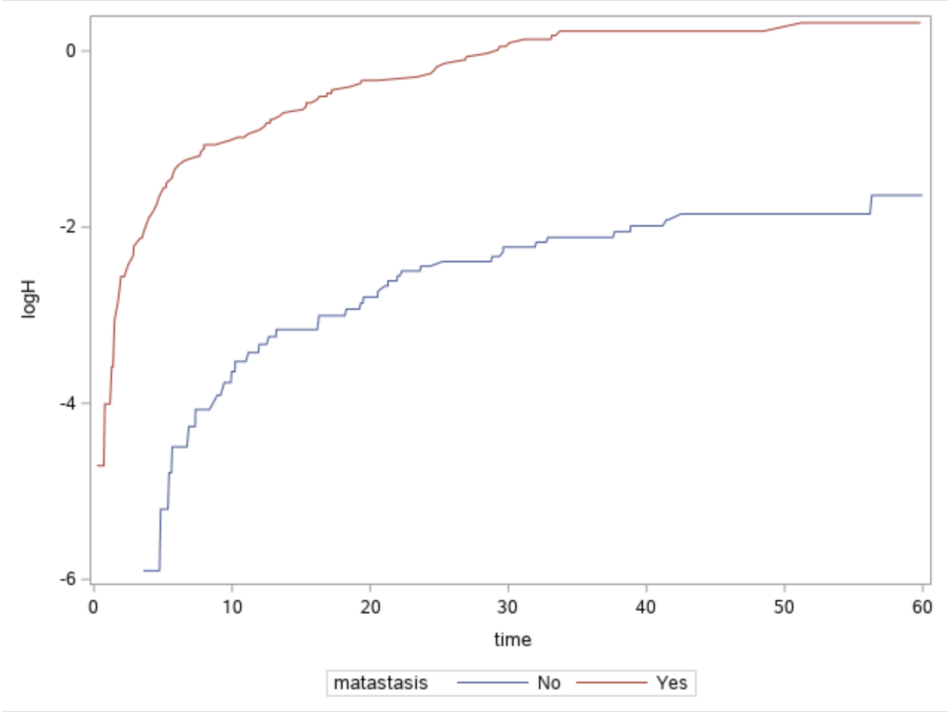

FIGURE 9: Log-log survival plot for metastasis

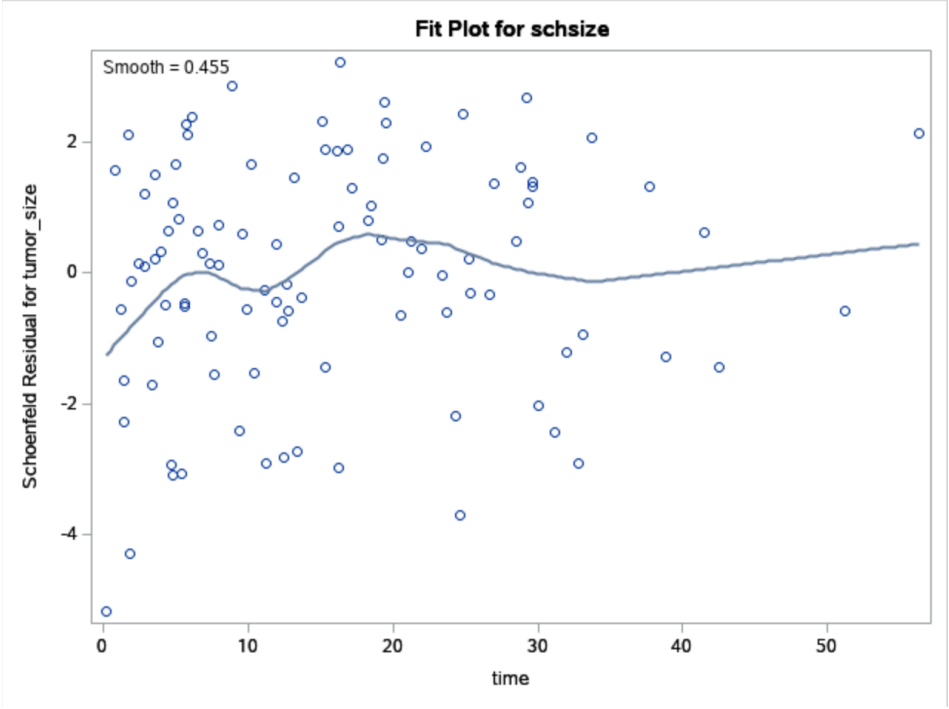

FIGURE 10: Schoenfeld residuals plot for tumor size

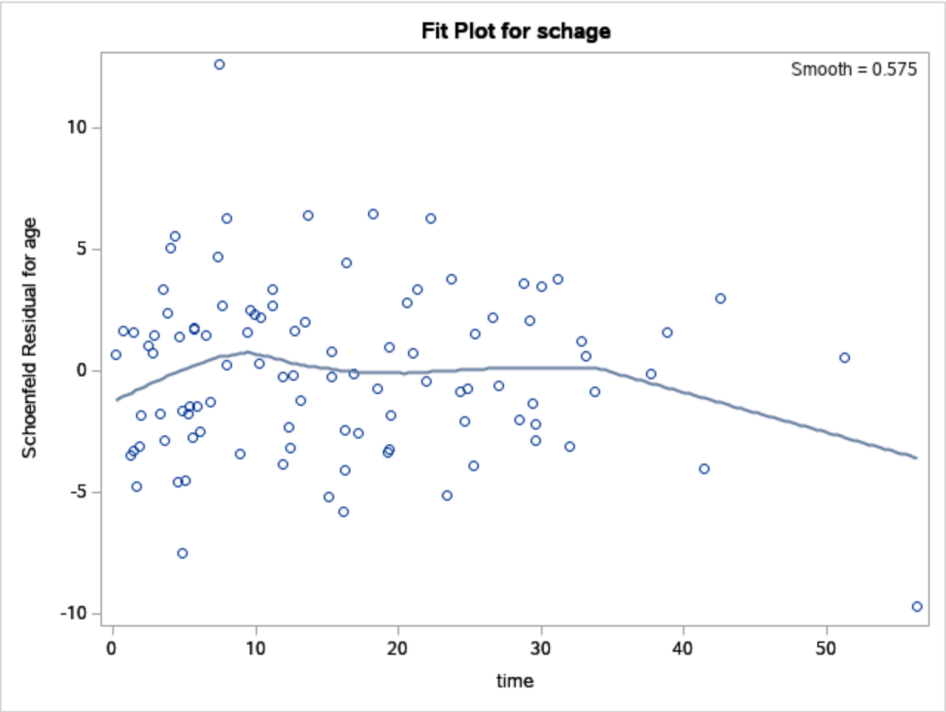

FIGURE 11: Schoenfeld residuals plot for tumor size age

```
getnames=yes;
run;
    data project2.data;
set project2.data0;
if stage=1 then do; stage1=1; stage2=0; stage3=0; stage4=0; end;
else if stage=2 then do; stage1=0; stage2=1; stage3=0; stage4=0; end;
else if stage=3 then do; stage1=0; stage2=0; stage3=1; stage4=0; end;
else if stage=4 then do; stage1=0; stage2=0; stage3=0; stage4=1; end;
format F_history F_history. Stage Stage. Metastasis Metastasis. Pathology Pathology.
Surgery Surgery.;
run;
    Proc format;
Value F_history
0 = 'No'
1 = 'Yes';
Value Stage
1 = 'I'
2 = 'II'
3 = 'III'
4 = 'IV';
Value Metastasis
```

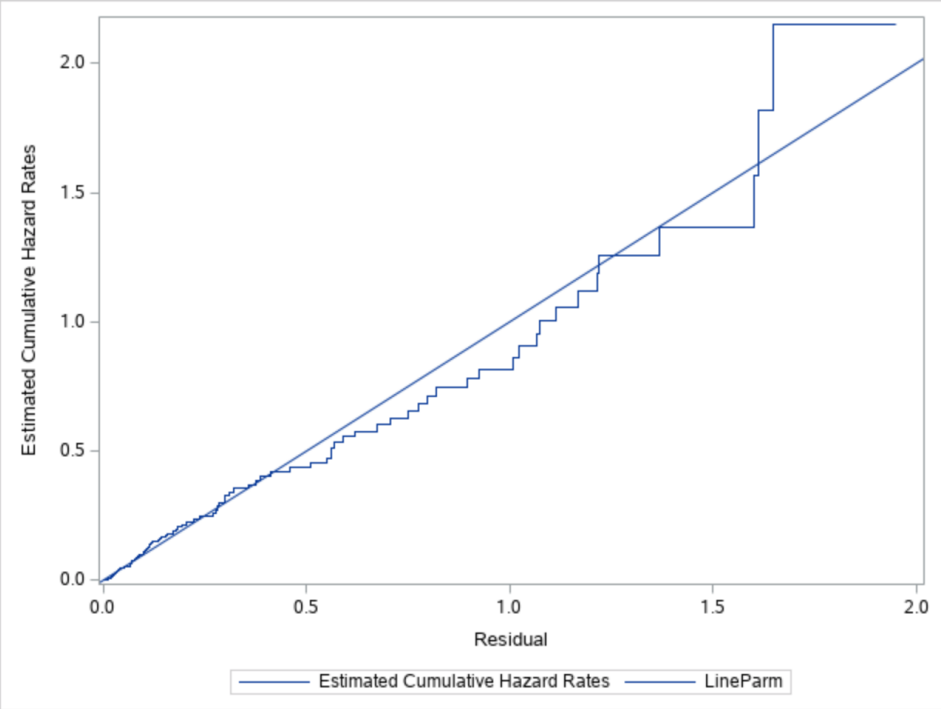

FIGURE 12: Cox-snell residuals plot with four variables (surgery, metastasis, tumor size, age) in the model

0 = 'No'
1 = 'Yes';
Value Pathology
0 = 'IDC'
1 = 'DCIS';
Value Surgery
0 = 'radical mastectomy'
1 = 'breast saving';
Run;
****
Analysis
**********;
proc lifetest data=project2.data plots=survival(atrisk cb);
time time*censored(1);
run;
* table 1;
proc freq data=project2.data;
tables surgery*(metastasis stage pathology f_history)/chisq nocol ;
format f_history f_history. stage stage. pathology pathology. metastasis metastasis. surgery surgery.;

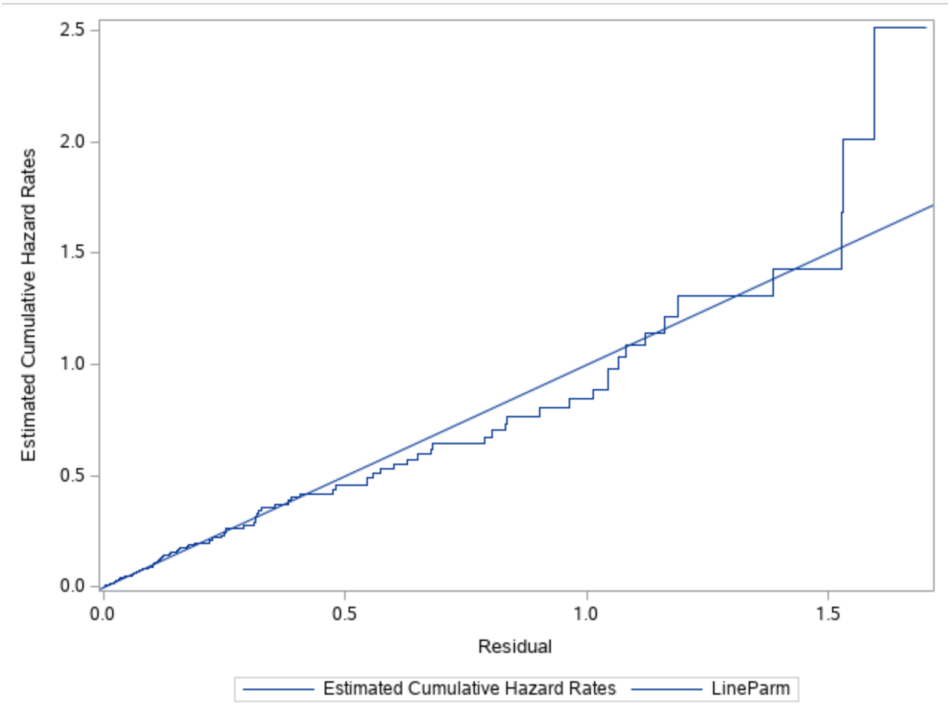

FIGURE 13: Cox-snell residuals plot with three variables (surgery, tumor size, age) in the stratified model (metastasis)

```
run;
proc sort data=project2.data;
by surgery;
run;
proc univariate data=project2.data;
var age tumor_size;
by surgery;
format surgery surgery.;
run;
proc ttest data=project2.data;
var age tumor_size;
class surgery;
format surgery surgery.;
run;
**********
non parametric
**********;
* KM;
proc lifetest data=project2.data plots=survival(atrisk cb cl);
time time*censored(1);
```

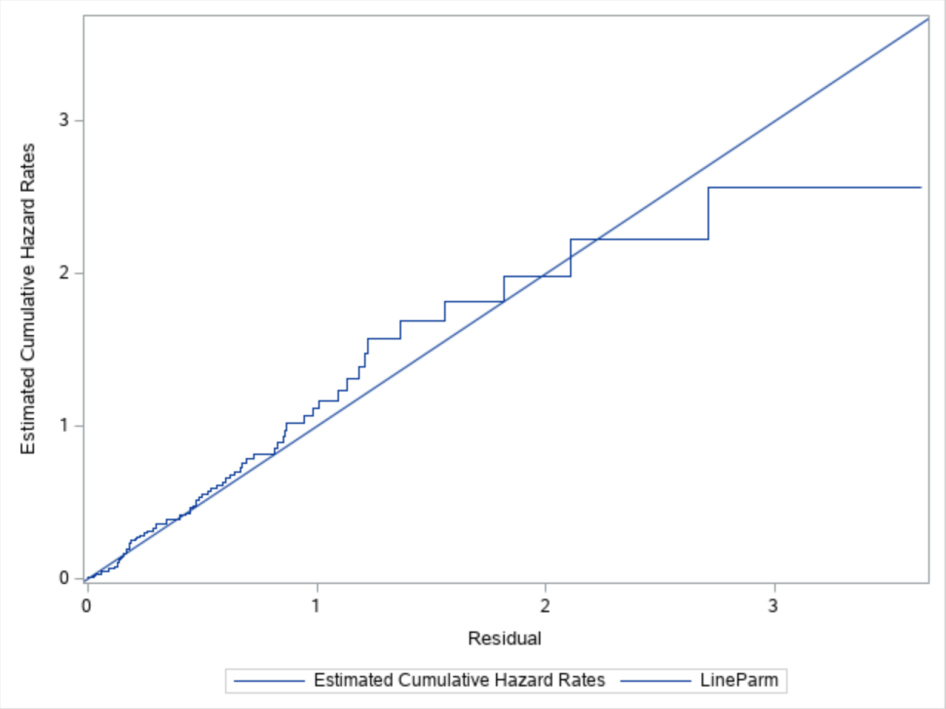

FIGURE 14: Cox-snell residuals plot with three variables (surgery, tumor size, age, tumor size*time) in the stratified model (metastasis)

```
format surgery surgery.;
run;
* hazard rate;
proc lifetest data=project2.data plots=hazard notable;
time time*censored(1);
format surgery surgery.;
run;
**********
surgery type
**********;
* KM;
proc lifetest data=project2.data plots=survival(atrisk cb cl);
time time*censored(1);
strata surgery;
format surgery surgery.;
run;
* hazard rate;
proc lifetest data=project2.data plots=hazard notable;
time time*censored(1);
strata surgery;
```

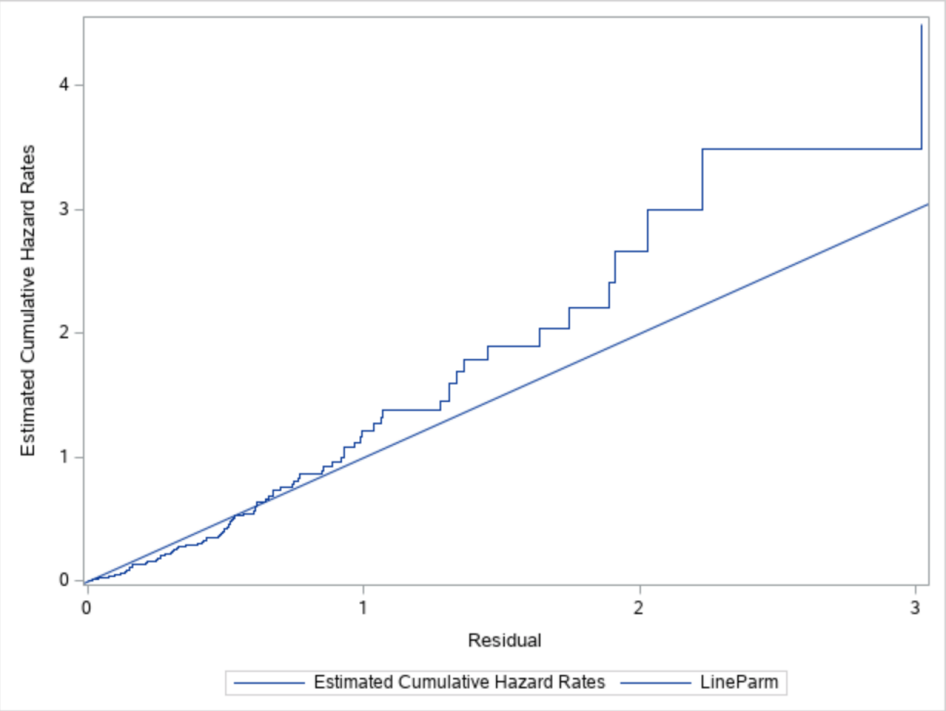

FIGURE 15: Cox-snell residuals plot with three variables (surgery, tumor size, age, age*time) in the stratified model (metastasis)

```
format surgery surgery.;
run;
* two sample test;
proc lifetest data=project2.data notable;
time time*censored(1);
strata surgery/test=(wilcoxon fleming(1,0)logrank);
format surgery surgery.;
run;
****************
model building
****************;
proc phreg data=project2.data;
class surgery;
model time*censored(1)= surgery age;
format surgery surgery. ;
run;
proc phreg data=project2.data;
class surgery;
model time*censored(1)= surgery tumor_size;
format surgery surgery. ;
```

```
run;
proc phreg data=project2.data;
class surgery f_history;
model time*censored(1)= surgery f_history;
format surgery surgery. f_history f_history.;
run;
proc phreg data=project2.data;
class surgery stage;
model time*censored(1)= surgery stage;
format surgery surgery. stage stage.;
run;
proc phreg data=project2.data;
class surgery metastasis;
model time*censored(1)= surgery metastasis;
format surgery surgery. metastasis metastasis.;
run;
proc phreg data=project2.data;
class surgery pathology;
model time*censored(1)= surgery pathology;
format surgery surgery. pathology pathology.;
run;
* +metastasis;
proc phreg data=project2.data;
class surgery metastasis;
model time*censored(1)= surgery metastasis age;
format surgery surgery. metastasis metastasis.;
run;
proc phreg data=project2.data;
class surgery;
model time*censored(1)= surgery metastasis tumor_size;
format surgery surgery. ;
run;
proc phreg data=project2.data;
class surgery f_history;
model time*censored(1)= surgery metastasis f_history;
format surgery surgery. f_history f_history.;
run;
proc phreg data=project2.data;
class surgery stage;
model time*censored(1)= surgery metastasis stage;
format surgery surgery. stage stage.;
run;
proc phreg data=project2.data;
class surgery pathology;
```

```
model time*censored(1)= surgery metastasis pathology;
format surgery surgery. pathology pathology.;
run;
* + metastasis;
proc phreg data=project2.data;
class surgery metastasis;
model time*censored(1)= surgery metastasis tumor_size age;
format surgery surgery. metastasis metastasis.;
run;
proc phreg data=project2.data;
class surgery f_history;
model time*censored(1)= surgery metastasis tumor_size f_history;
format surgery surgery. f_history f_history.;
run;
proc phreg data=project2.data;
class surgery stage;
model time*censored(1)= surgery metastasis tumor_size stage;
format surgery surgery. stage stage.;
run;
proc phreg data=project2.data;
class surgery pathology;
model time*censored(1)= surgery metastasis tumor_size pathology;
format surgery surgery. pathology pathology.;
run;
*final model;
proc phreg data=project2.data;
class surgery metastasis;
model time*censored(1)= surgery metastasis tumor_size age;
format surgery surgery. metastasis metastasis.;
run;
************************
Local test- interaction
************************;
*surgery;
proc phreg data=project2.data;
class surgery;
model time*censored(1)= surgery age age*surgery;
contrast "beta3=0" surgery*age 1 /test(wald lr score);
format surgery surgery. ;
run;
proc phreg data=project2.data;
class surgery;
model time*censored(1)= surgery tumor_size tumor_size*surgery;
contrast "beta3=0" surgery*tumor_size 1 /test(wald lr score);
```

```
format surgery surgery. ;
run;
proc phreg data=project2.data;
class surgery metastasis;
model time*censored(1)= surgery metastasis metastasis*surgery;
contrast "beta3=0" surgery*metastasis 1 /test(wald lr score);
format surgery surgery. ;
run;
*metastasis;
proc phreg data=project2.data;
class metastasis;
model time*censored(1)= tumor_size metastasis metastasis*tumor_size;
contrast "beta3=0" tumor_size*metastasis 1 /test(wald lr score);
run;
proc phreg data=project2.data;
class metastasis;
model time*censored(1)= age metastasis metastasis*age;
contrast "beta3=0" age*metastasis 1 /test(wald lr score);
format surgery surgery. ;
run;
*age;
proc phreg data=project2.data;
model time*censored(1)= tumor_size age age*tumor_size;
contrast "beta3=0" tumor_size*age 1 /test(wald lr score);
run;
**********
Cox PH
**********;
proc phreg data=project2.data;
class metastasis surgery(ref="radical mastectomy");
model time*censored(1)= surgery metastasis tumor_size age;
format Metastasis Metastasis. Surgery Surgery.;
run;
******* martingale;
* age;
proc phreg data=project2.data;
class metastasis surgery(ref="radical mastectomy");
model time*censored(1)= surgery metastasis tumor_size;
format Metastasis Metastasis. Surgery Surgery.;
output out=plot2_1 RESMART = mgale ;
run;
proc loess data=plot2_1;
model mgale = age / direct;
run;
```

```
* tumor size;
proc phreg data=project2.data;
class metastasis surgery(ref="radical mastectomy");
model time*censored(1)= surgery metastasis age;
format Metastasis Metastasis. Surgery Surgery.;
output out=plot2_1 RESMART = mgale ;
run;
proc loess data=plot2_1;
model mgale = tumor_size / direct;
run;
************************
PH assumption
************************;
*****************age;
* shoenfeld;
proc phreg data = project2.data;
class metastasis surgery(ref="radical mastectomy");
model time*censored(1)= age surgery metastasis tumor_size ;
format Metastasis Metastasis. Surgery Surgery.;
output out = schoen ressch= schage ;
run;
proc loess data=schoen;
model schage=time;
run;
* test;
proc phreg data = project2.data;
class metastasis surgery(ref="radical mastectomy");
model time*censored(1)= age surgery metastasis tumor_size age*time;
format Metastasis Metastasis. Surgery Surgery.;
run;
*****************tumorsize;
* shoenfeld;
proc phreg data = project2.data;
class metastasis surgery(ref="radical mastectomy");
model time*censored(1)= tumor_size age surgery metastasis;
format Metastasis Metastasis. Surgery Surgery.;
output out = schoen ressch= schsize ;
run;
proc loess data=schoen;
model schsize=time;
run;
*test;
proc phreg data = project2.data;
class metastasis surgery(ref="radical mastectomy");
```

```
model time*censored(1)= tumor_size age surgery metastasis tumor_size*time ;
format Metastasis Metastasis. Surgery Surgery.;
run;
********** surgery;
* log(H) vs time;
data new ;
set project2.data ;
cons = 1;
run;
proc phreg data = new;
class surgery/param=ref;
model time*censored(1) =cons/rl ;
strata surgery;
output out = base logsurv = ls /method = ch;
run;
data base;
set base ;
logH = log (-ls);
if surgery= 0 then logH1 = logH;
else if surgery= 1 then logH2 = logH;
proc sort; by surgery time ;
proc print; var surgery time logH logH1 logH2 ;
run;
proc sgplot data =base;
where logH ne .;
series x=time y=logH /group=surgery ;
format surgery surgery.;
run;
********** matastasis;
* log(H) vs time;
data new ;
set project2.data ;
cons = 1;
run;
proc phreg data = new;
class metastasis/param=ref;
model time*censored(1) =cons/rl ;
strata metastasis;
output out = base logsurv = ls /method = ch;
run;
data base;
set base ;
logH = log (-ls);
if metastasis= 0 then logH1 = logH;
```

```
else if metastasis= 1 then logH2 = logH;
proc sort; by metastasis time ;
proc print; var metastasis time logH logH1 logH2 ;
run;
proc sgplot data =base;
where logH ne .;
series x=time y=logH /group=metastasis ;
format metastasis metastasis.;
run;
*********snell;
proc phreg data=project2.data;
class metastasis surgery(ref="radical mastectomy");
model time*censored(1)= tumor_size age surgery metastasis;
format Metastasis Metastasis. Surgery Surgery.;
output out=plot1_1 logsurv = logsurv1/method=ch;
run;
data plot1_1;
set plot1_1;
snell = -logsurv1;
cons = 1;
run;
proc phreg data = plot1_1;
model snell*censored(1) = cons;
output out = plot1_2 logsurv = logsurv2 /method = ch;
run;
data plot1_2;
set plot1_2;
cumhaz = - logsurv2;
run;
proc sort data = plot1_2;
by snell;
run;
proc sgplot data = plot1_2;
step y=cumhaz x=snell /MARKERFILLATTRS=(color="red");
lineparm x=0 y=0 slope=1; /** intercept, slope **/
label cumhaz = "Estimated Cumulative Hazard Rates";
label snell = "Residual";
run;
*********snell;
proc phreg data=project2.data;
class metastasis surgery(ref="radical mastectomy");
model time*censored(1)= tumor_size age surgery;
strata metastasis;
format Metastasis Metastasis. Surgery Surgery.;
```

```
output out=plot1_1 logsurv = logsurv1/method=ch;
run;
data plot1_1;
set plot1_1;
snell = -logsurv1;
cons = 1;
run;
proc phreg data = plot1_1;
model snell*censored(1) = cons;
output out = plot1_2 logsurv = logsurv2 /method = ch;
run;
data plot1_2;
set plot1_2;
cumhaz = - logsurv2;
run;
proc sort data = plot1_2;
by snell;
run;
proc sgplot data = plot1_2;
step y=cumhaz x=snell /MARKERFILLATTRS=(color="red");
lineparm x=0 y=0 slope=1; /** intercept, slope **/
label cumhaz = "Estimated Cumulative Hazard Rates";
label snell = "Residual";
run;
************************
no interacton assumption
************************;
proc phreg data=project2.data;
class metastasis surgery(ref="radical mastectomy");
model time*censored(1)= tumor_size age surgery /rl;
strata metastasis;
format Metastasis Metastasis. Surgery Surgery.;
run;
proc phreg data=project2.data;
where metastasis=0;
class metastasis surgery(ref="radical mastectomy");
model time*censored(1)= tumor_size age surgery ;
format Metastasis Metastasis. Surgery Surgery.;
run;
proc phreg data=project2.data;
where metastasis=1;
class metastasis surgery(ref="radical mastectomy");
model time*censored(1)= tumor_size age surgery ;
format Metastasis Metastasis. Surgery Surgery.;
```

run;

C Appendix - R Simulation code

```
library(MASS)
    sim_cox<- function(N,lambda0, beta, censor.right)
{
# N = Total sample size
# beta = PH coefficients
# lambda0 = rate parameter of the exponential distribution for baseline
    # randomization to treatment or control
X1 <- sample(x=c(0, 1), size=N, replace=TRUE, prob=c(0.73, 0.27))
    # generate continuous covariates, mutually indepedent
X = mvrnorm(N,mu=c(50.39,3.6),Sigma=matrix(c(11.19,0,0,2.37),2,2))
X2=X[,1]
X3=X[,2]
    # generate categorical covariates
X4 <- sample(x=c(0, 1), size=N, replace=TRUE, prob=c(0.77, 0.23))
X5 <- sample(x=c(1, 2, 3, 4), size=N, replace=TRUE, prob=c(0.155, 0.633, 0.174, 0.038))

X8 <- sample(x=c(0, 1), size=N, replace=TRUE, prob=c(0.12, 0.88))
X9 <- sample(x=c(0, 1), size=N, replace=TRUE, prob=c(0.91, 0.09))
    # generate underlying event time
T <- rexp(n=N, rate = lambda0*exp(beta[1]*X1+beta[2]*X2+beta[3]*X3+beta[4]*X4+beta[5]*(X5==2

#mean(X)
#rexp(n=N, rate=lambda0*exp(beta*A))
    # censoring times
ctime = runif(N, min=0, max=censor.right)
    # follow-up times and event indicators
time <- pmin(T, ctime, censor.right)
censor <- as.numeric(T>ctime | T>censor.right)
    # data set
data.frame(id=1:N,
Surgery=X1,
age=X2,
tumor_size=X3,
metastasis=X4,
Stage=X5,
Pathology=X8,
F_history=X9,
time=time,
censored=censor)
```

```
}
mydata=sim_cox(N=100000, lambda0=0.0001, beta=c(0.339, 0.047, 0.24, 2.47, -0.207,
-0.12, -0.226, 0.183, 0.464), censor.right=60)

censored <- mydata %>% filter(censored==1) %>% sample_n(400)
notcensored <- mydata %>% filter(censored==0) %>% sample_n(100)
data<-rbind(censored, notcensored)
write.csv(data,
file="/Users/mengzichun/Dropbox/My Mac (Liela MacBook Pro)/Desktop/R/BST222/222
project 2/data.csv")
```

References

[1] Bakhshi, E., Sheikhaliyan, A., Atashgar, K., Kooshesh, M., and Biglarian, A. (2017) Survival Analysing of the Breast Cancer Patients Using Cure Model. *Iranian Red Crescent Medical Journal*, **19** (7), 19–19, doi:10.5812/ircmj.55575. URL https://dx.doi.org/10.5812/ircmj.55575.

[2] Bakhshi, E., Sheikhaliyan, A., Atashgar, K., Kooshesh, M., and Biglarian, A. (2017) Survival Analysing of the Breast Cancer Patients Using Cure Model. *Iranian Red Crescent Medical Journal*, **19** (7), doi:10.5812/ircmj.55575. /Users/mengzichun/Zotero/storage/ZZQP78SM/Bakhshi et al. - 2017 - Survival Analysing of the Breast Cancer Patients U.pdf.

[3] Bender, R., Augustin, T., and Blettner, M. (2005) Generating survival times to simulate Cox proportional hazards models: GENERATING SURVIVAL TIMES. *Statistics in Medicine*, **24** (11), 1713–1723, doi:10.1002/sim.2059. /Users/mengzichun/Zotero/storage/GQBF5T3T/Bender et al. - 2005 - Generating survival times to simulate Cox proporti.pdf.