The Effect of Hormone Therapy on Breast Cancer

Rong Duan

1 Introduction

Breast cancer is a global health concern and it is the primary cause of death among women worldwide and constitutes the most expensive malignancy to treat [1]. Some types of breast cancer are affected by hormones, like estrogen and progesterone. The breast cancer cells have receptors that attach to estrogen and progesterone, which helps the cancer cell grow [2]. Treatments that stop these hormones from attaching to these receptors are called hormone therapy. Hormone therapy is often used after surgery as adjutant therapy to help reduce the risk of cancer coming back, and sometimes it is started before surgery as neoadjuvant therapy [3].

The primary object of this study is to explore the effect of hormone therapy on breast cancer. The hormone therapy effect is compared with and without other covariates, to see whether the hormone therapy reaches a significant result. Besides, other covariates are analyzed for the influence of breast cancer. A Cox regression model is built to predict the hazard rate for a general population. This study aims to found the influential factors to breast cancer, thus some individuals with these characteristics can be identified earlier to prevent bad events happen.

2 Method

2.1 Study design and population

From 1984 to 1989, 720 patients with positive node breast cancer were randomized in the German Breast Cancer Study Group (GBSG) Trial, after their breast surgery, to receive either hormonal therapy or no hormonal therapy. Because menopausal status is highly correlated with breast cancer [5], at the beginning of the study, both pre-menopausal and postmenopausal patients were randomized into the two therapy group [4]. The patients are followed until either dead or the recurrence of breast cancer within 900 days. At beginning of the study, some features are recorded, that is age, menopausal status, tumor size, tumor grade, number of positive lymph nodes, progesterone receptors (fmol/l), estrogen receptors (fmol/l), whether receive hormonal therapy, recurrence-free survival time (days to first of recurrence, death or last follow-up), and status (whether it is recurrenced or dead). The event time is recurrence-free survival time, and event indicator is status. It retains the 686 patients with complete data for the prognostic variables.

2.2 Basic covariates exploration

The baseline hormone therapy effect without other covariates is analyzed with Kaplan-Meier Estimator for survival probability and The Nelson-Aalen estimator for Cumulative hazard function. The hazard rate of the therapy group is compared with the control group by the log-rank test. The null hypothesis of the log-rank test is the hazard rate of the two groups in each event time is the same. A small p-value can reject the null hypothesis and therefore the hazard rate is not same at least in one of the event

times. Hormone therapy is suspected to work differently in different tumor grades, and a stratified test is conducted to test whether hormone therapy works in each tumor grade group. Also, tumor grade is suspected for its effect on breast cancer. A trend test is conducted to test whether the higher grade, the higher risk.

Based on substantial research, menopausal status, progesterone receptors, and estrogen receptors have been found a high correlation-ship with breast cancer [6]. This trial randomized the pre-menopausal and postmenopausal women to attune its effect on breast cancer. Estrogens and progestogens have been found to act with proto-oncogenes and growth factors to affect the proliferation of breast cells. Thus these two types of hormones behave importantly in breast cancer.

2.3 Model building and data analysis

From the above covariates analysis, 4 covariates are included in the model because of the study object and scientific significance, that is progesterone receptors, estrogen receptors, menopause status, and hormonal therapy. Other 4 variables and two interaction terms, nodes tumor size interaction, and hormone tumor grade interaction, are the candidates for the model. The AIC and p-value are conferred as the model selection criteria. 0.05 P-value is considered significant based on some other breast cancer research standard [7–9]. The Assumption of the selected model is tested, which are time-independent Hazard ratio assumption and proportional hazard ratio assumption. Schoenfeld residuals and Log Log survival plots are applied for testing the assumptions. The PH assumption violated variables will be considered to use the Stratified Cox model as a remedy. The final model is explained with the hazard ratio of each covariate.

3 Results

3.1 3.1 Descriptive statistics of the Dataset

In this study, the final dataset including 686 patients with positive diagnostic nodes in the breast cancer analysis. The covariate proportions of relapse and censored case in the hormone treatment and control group are explored [Table 1]. Notably, the proportion of relapse events in a postmenopausal group with hormone therapy is smaller than the control group. Combined with the 34 missing data in the study duration, the aim of randomization for weakening the effect of menopause on breast cancer might not work. As denoted in the academic paper, the progesterone of this study shows that its volume in the censored group is around twice larger in the relapse group. Also, the tumor size and the number of nodes in relapse group are larger than the censored group no matter in the hormone therapy group or control group. These may indicate the high correlation with the relapse of breast cancer. Some further analysis is conducted below.

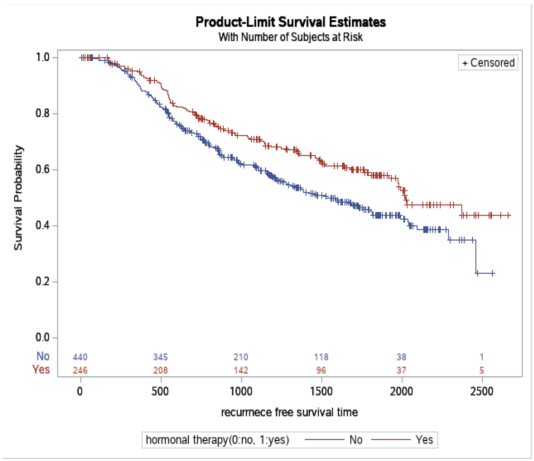
3.2 risk factors analysis

The effect of hormone therapy and placebo are compared by survival probability and hazard rate. Figure 1 shows the estimated survival probability of the hormone group is always higher than the placebo group.

Table 1: Baseline information of breast cancer patients

Factors	Hormone Therapy		Placebo	
	Relapse(N/%)	Cencered (N/%)	Relapse(N/%)	Censored (N/%)
Hormon	94(31.44%)	152(39.28%)	205(68.56%)	235(60.72%)
postmenopause	72(76.60%)	115(75.66%)	108(52.68%)	101(42.98%)
grade-1	6(6.38%)	27(17.76%)	12(5.85%)	36(15.32%)
grade-2	65(69.15%)	98(64.47%)	137(66.83%)	144(61.28%)
grade-3	23(24.47%)	27(17.76%)	56(27.32%)	55(23.40%)
Factors	Hormone Therapy		Placebo	
	Relapse(mean)	Cencered(mean)	Relapse(mean)	Censored (mean)
progesterone	54.89	167.21	77.69	123.20
estrogen	111.3	134.75	73.69	84.98
nodes	7.34	3.76	6.14	3.90
age(aobve 21)	14.5	14.7	9.4	8.8
size	32.39	26.58	31.03	28.40

The overall hazard rate of the placebo group is higher than the hormone therapy group, although the hazard rate crosses a little bit in the middle part. The log-rank test is utilized to test the equality hazard rate of the treatment and control group, and the result shows the two groups are statistically significant concerning the hazard rate(log test=8.5648, df=1, p-value=0.0034). Also, some suspicious influential factors (tumor grade and tumor size) are stratified to compare the hormone therapy effect with each level, respectively. The result of stratified test with tumor grade(p-value=0.0065) and size(p-value=0.0051) shows the hormone treatment effect is significant in each level. The trend test for tumor grade has a significant result (log-rank test=44.5342, p-value<0.0001), and it means the higher grade the higher the hazard rate[figure 2].



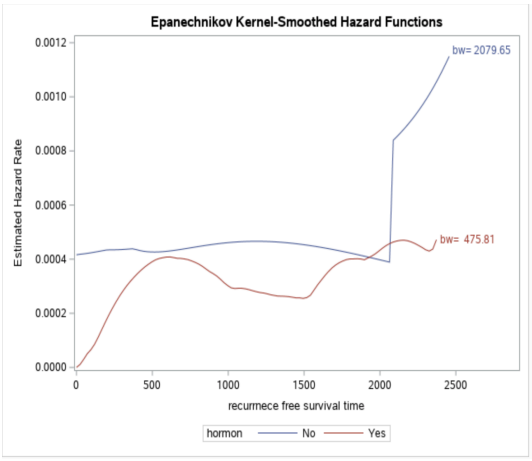


Figure 1: Survival probability and hazard rate in hormone therapy group

3.3 Model selection and diagnostic

Based on the formal analysis, 4 variables are included in the model automatically, that is menopausal status, progesterone receptor, estrogen receptor and hormonal therapy. With AIC criteria and 0.05 p-value, 2 more variables are added (number of positive nodes and tumor grade). The model expression

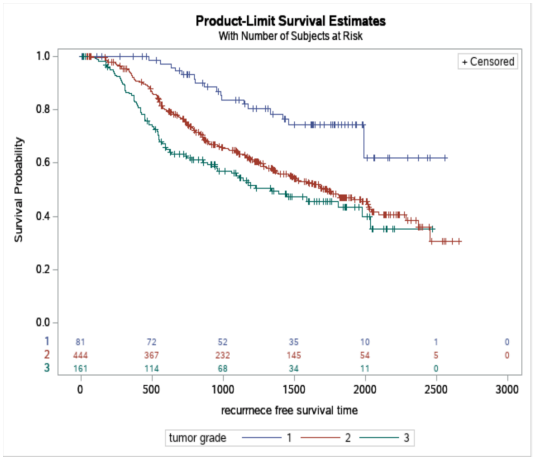


Figure 2: Trend test for tumor grade

and fit table is as follows:

Table 2: Cox PH model .

Covariate	DF	Parameter	SE	P-value	HR
hormone therapy	1	-0.3489	0.1290	0.0069	1.417
number of posi-nodes	1	0.0541	0.0068	<.0001	1.056
progesterone receptor	1	-0.0022	0.0006	0.0001	0.998
menopausal status	1	0.1141	0.1249	0.3604	1.121
estrogen receptors	1	0.00005	0.0004	0.9014	1.000
tumor grade 3	1	0.8086	0.2677	0.0025	2.245
tumer grade 2	1	0.6462	0.2490	0.0094	1.908

$h(t, \mathbf{Z}) = h_0(t) \exp(-0.341 \times hormone + 0.055 \times nodes - 0.002 \times progesteron + 0.126 \times menopausal + 0.00002 \times estroge + .297 \times grade)$

Before the explanation of the model, fitness and assumptions are conducted. By Cox-snell residual plot, the fitness of the model is checked[Figure 3]. It shows the overall fit is not satisfactory. The assumptions of the Cox model is checked by the Log Log survival plot and Schoenfeld residuals. The ph assumption of hormone variable is checked by the Log Log plot, it seems the assumption is satisfied with the three plots[Figure 4]. The PH assumption of other variables is checked by Schonefeld plots. All the assumptions of covariates are not severe violated[Figure 5] except tumer grade. Considering the PH assumption is not violated much, we still use this model as the final model for analysis.

3.4 Model explanation

The hazard ratio of placebo to hormone therapy is 1.417, which means the placebo will increase 41.7% hazard rate. One number of additional nodes will increase the hazard rate by 5.6%, and postmenopausal status increases the hazard rate by 12.1%. One progesterone receptor will decrease the hazard rate by 0.2%, while the estrogen receptor in this case seems to not influence the hazard rate. Tumor grade 3

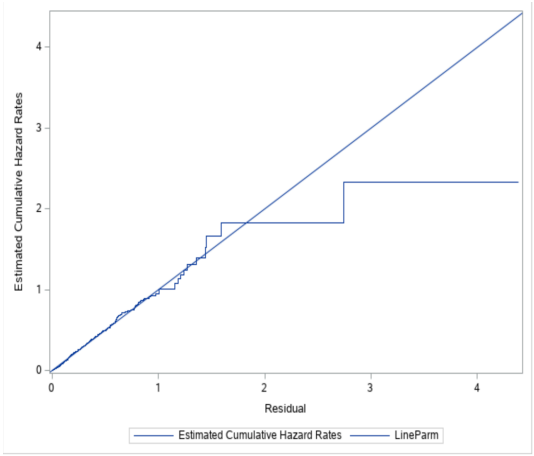


Figure 3: Cox-Snell residual plot

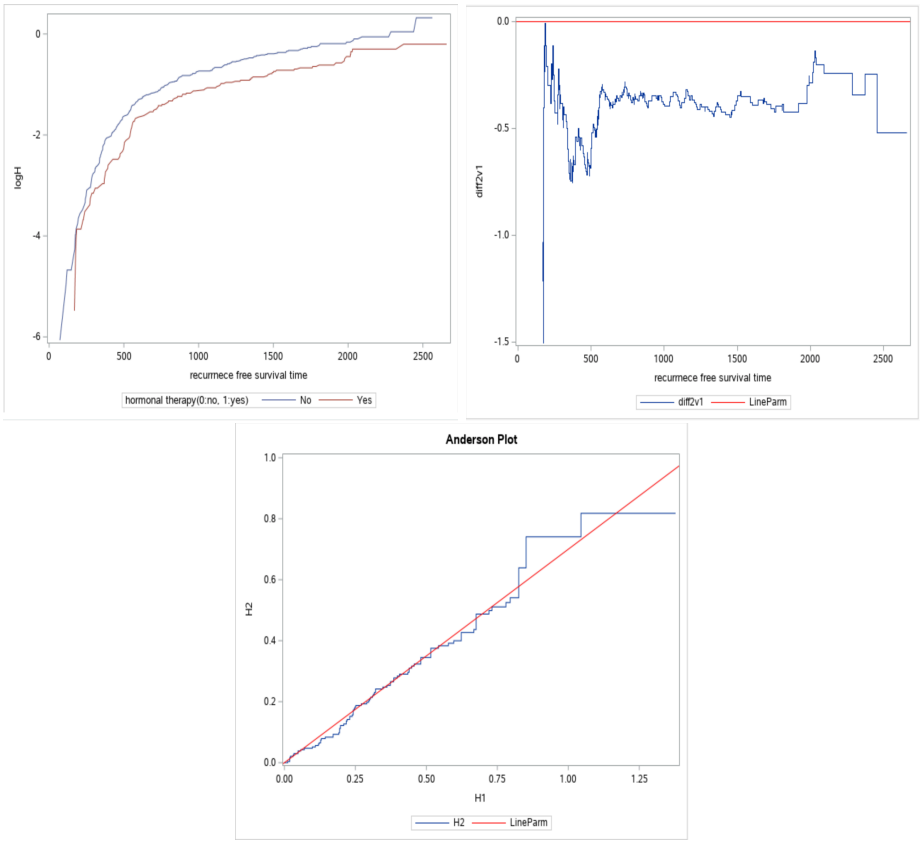


Figure 4: Log Log Survival plot

has the most hazard rate, and it increases the hazard rate by 124.5% compared with grade 1. Tumor grade 2 is also very risky and increases the hazard rate by 90.9% compared with grade 1.

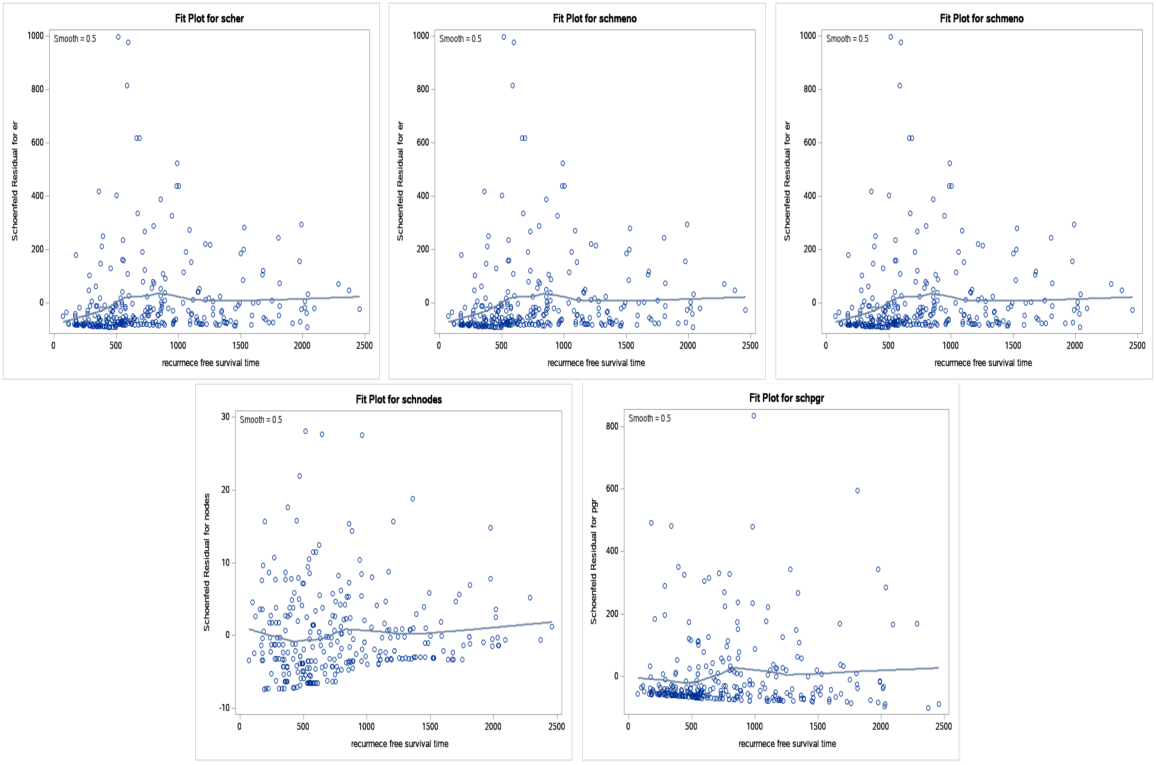


Figure 5: Schonefeld Residual Plot

4 Discussion

4.1 Importance of breast cancer

Breast cancer is a global health issue, and tons of studies have been conducted to find effective treatment and risk factors. Some hormone therapy is also a complex and conflicting issue to breast cancer risk [10]. The study found this type of hormone therapy did works better concerning breast cancer. 2 and 3 tumor grade highly increases the hazard rate and more number of positive nodes the higher risk. The progesterone receptor is found to lower the hazard rate, but the estrogen, which is highly correlated with breast cancer is not significant here. Menopausal status is randomized in a certain degree, and its effect is attuned in a great deal. Age is tumor size are not correlated with relapse risk.

4.2 Limitation of the study

There are some limitations to the analysis. First, 36 people are lost in the duration, and some important features might be in them, for example, most of them are in postmenopausal status. If the missing data is non-randomly missed, then it will lead to a biased analysis. Also, the estrogen is not significant in the result, and the reason of this non-significance needed some further exploration. The time range in this study is nearly 3 years, some side effects of the hormone therapy might be uncovered in this period. Although with the analysis the hormone therapy is effective for breast cancer, the safety and effect time need some further exploration.

5 Conclusion

By this research, hormone therapy to the breast cancer did works and lower the risk of cancer relapse and cancer related death. Also the tumor grade can indicate the hazard rate of breast cancer in a certain degree. Women in postmenopausal status with more number of positive nodes has a higher risk.

References

[1] Aarts, H., & Dijksterhuis, A. (1999). *How often did I do it? Experienced ease of retrieval and frequency estimates of past behavior.* Acta Psychologica, 103(3), 77–89.

[2] D'Abreo N, Hindenburg AA.*Sex hormone receptors in breast cancer.* Vitam Horm. 2013;93:99-133. doi: 10.1016/B978-0-12-416673-8.00001-0. PMID: 23810004.

[3] Apuri S.*Neoadjuvant and Adjuvant Therapies for Breast Cancer.* South Med J. 2017 Oct;110(10):638-642. doi: 10.14423/SMJ.0000000000000703. PMID: 28973704.

[4] Colleoni M, Litman HJ, Castiglione-Gertsch M, Sauerbrei W, Gelber RD, Bonetti M, Coates AS, Schumacher M, Bastert G, Rudenstam CM, Schmoor C, Lindtner J, Collins J, Thürlimann B, Holmberg SB, Crivellari D, Beyerle C, Neumann RL, Goldhirsch A; *International Breast Cancer Study Group; German Breast Cancer Study Group. Duration of adjuvant chemotherapy for breast cancer: a joint analysis of two randomised trials investigating three versus six courses of CMF.* Br J Cancer. 2002 Jun 5;86(11):1705-14. doi: 10.1038/sj.bjc.6600334. PMID: 12087454; PMCID: PMC2375405.

[5] Collaborative Group on Hormonal Factors in Breast Cancer. *Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies.* Lancet Oncol. 2012 Nov;13(11):1141-51. doi: 10.1016/S1470-2045(12)70425-4. Epub 2012 Oct 17. PMID: 23084519; PMCID: PMC3488186.

[6] Pike MC, Spicer DV, Dahmoush L, Press MF. *Estrogens, progestogens, normal breast cell proliferation, and breast cancer risk.* Epidemiol Rev. 1993;15(1):17-35. doi: 10.1093/oxfordjournals.epirev.a036102. PMID: 8405201.

[7] Huang CC, Chan SY, Lee WC, Chiang CJ, Lu TP, Cheng SH. *Development of a prediction model for breast cancer based on the national cancer registry in Taiwan.* Breast Cancer Res. 2019 Aug 13;21(1):92. doi: 10.1186/s13058-019-1172-6. PMID: 31409418; PMCID: PMC6691540.

[8] Kanstrup C, Teilum D, Rejnmark L, Bigaard JV, Eiken P, Kroman N, Tjønneland A, Mejdahl MK. *25-Hydroxyvitamin D at time of breast cancer diagnosis and breast cancer survival.* Breast Cancer Res Treat. 2020 Feb;179(3):699-708. doi: 10.1007/s10549-019-05486-4. Epub 2019 Nov 9. PMID: 31707511.

[9] Martinez-Ramos D, Simon-Monterde L, Queralt-Martin R, *Suelves-Piqueres C, Menor-Duran P, Escrig-Sos J. Breast cancer in octogenarian. Are we doing our best? A population-registry based study.* Breast. 2018 Apr;38:81-85. doi: 10.1016/j.breast.2017.12.007. Epub 2018 Jan 4. PMID: 29274475.

[10] Hodis HN, Sarrel PM. *Menopausal hormone therapy and breast cancer: what is the evidence from randomized trials?* Climacteric. 2018 Dec;21(6):521-528. doi: 10.1080/13697137.2018.1514008. Epub 2018 Oct 9. PMID: 30296850; PMCID: PMC6386596.

Appendix–SAS code

```
*Import data;
proc import datafile='/folders/myfolders/222-lab/project/BST222project1_data.csv'
out=cancer dbms=csv replace;
datarow=2;
getnames=yes;
run;

*Add lable;
data cancer(drop=var1);
set cancer;
label pid=patient identifier
age=patient age
meno=menopausal status (0:premenopausal, 1:postmenopausal)
size=tumor size
grade=tumor grade
nodes=number of positive lymph nodes
pgr=progesterone receptors (fmol/l)
er=estrogen receptors (fmol/l)
hormon=hormonal therapy(0:no, 1:yes)
rfstime=recurrnece free survival time
status=alive without recurrence(0),recurrence or death(1);
run;
*Convert the format of the dataset;
proc format;
value menof 0='pre' 1='post';
value gradef 1='1'  2='2' 3='3';
value hormonf 0='No' 1='Yes';
value node 0="number less than " 1="number greater than ";
run;
data cancer;
set cancer;
format meno menof. grade gradef. hormon hormonf.;
run;
proc freq data=cancer;
table status;
run;

*Summary;
proc sort data=cancer;by status hormon;run;
proc freq data=cancer;
by status hormon;
table  grade ;
run;
proc means data=cancer;
by status hormon;
var pgr nodes er age size;
run;
```

```
*Test equality over hormon group 8.5648^^I1^^Ip-value=0.0034;
proc lifetest data=cancer plots=(survival(atrisk=0 to 2700 by 500) hazard) notable;
time rfstime*status(0);
strata hormon/test=(logrank TARONE PETO MODPETO
FLEMING(0,1)    );
run;

* Tests for trend: tumer grade, test=44.5342, p-value<0.0001;
proc lifetest data=cancer plots=survival(atrisk=0 to 3000 by 500) notable;
time rfstime*status(0);
strata grade/trend test=(logrank TARONE PETO MODPETO
FLEMING(0,1) );
run;

* Stratefied test for hormon therapy in tumor grade groups 7.3958^^I1^^I0.0065;
proc lifetest data=cancer notable ;
time rfstime*status(0);
strata grade/group=hormon  test=(logrank TARONE PETO MODPETO
FLEMING(0,1)    );
run;

* Stratefied test for hormon therapy in size groups 7.8598^^I1^^I0.0051 ;
proc lifetest data=cancer notable ;
time rfstime*status(0);
strata size/group=hormon  test=(logrank TARONE PETO MODPETO
FLEMING(0,1)    );
run;


* create additional covariates;
data cancer;
   set cancer;
age=age-21;
h_g=hormon*grade;
label h_g="hormone tumer grade interaction";
n_s=nodes*size;
label n_s="Positive nodes and tumer size interaction";
run;



/*------------Model selection 1----------*/

* Create macro-variables with value covariate ;
proc sql noprint;
select name into : var1-:var6
from dictionary.columns
where LIBNAME = upcase("work")
and MEMNAME = upcase("cancer")
```

```
and upcase(name) ne upcase("rfstime")
and upcase(name) ne upcase("status")
and upcase(name) ne upcase("hormon")
and upcase(name) ne upcase("meno")
and upcase(name) ne upcase("pgr")
and upcase(name) ne upcase("er")
and upcase(name) ne upcase("pid");
quit;
%put &var1 &var2 &var3 &var4 &var5 &var6 ;
%macro Mselection1(var=);
proc phreg data=cancer;
model rfstime*status(0) = hormon pgr er meno &var;
&var: test &var=0;
ods output FitStatistics=%str(&var)aic;
ods output TestStmts = %str(&var)stat;
run;
%mend;

* Model selection funciton;
%macro MS1;
%DO I=1 %TO 6;
%Mselection1(var=&&var&I)
%END;
%mend MS1;
%MS1;

* Summarize the result by first run;
DATA aic_table_1 (drop=withoutcovariates);
set %str(&var1)aic %str(&var2)aic %str(&var3)aic %str(&var4)aic
%str(&var5)aic %str(&var6)aic ;
where criterion = "AIC";
cov=_n_;
run;
data stat_table_1(drop=status);
  set %str(&var1)stat %str(&var2)stat %str(&var3)stat
  %str(&var4)stat %str(&var5)stat %str(&var6)stat;
cov=_n_;
run;
proc sort data=aic_table_1;by cov;run;
proc sort data=stat_table_1;by cov;run;
data select_1;
merge aic_table_1 stat_table_1;
by cov;
run;
proc sort data=select_1;by ProbChiSq;
/*--------table------------*/
proc print data=select_1;
run;
* Nodes is selected;
```

```
/*------------Model selection 2----------*/

* Create macro-variables with value covariate ;
proc sql noprint;
select name into : var1-:var5
from dictionary.columns
where LIBNAME = upcase("work")
and MEMNAME = upcase("cancer")
and upcase(name) ne upcase("rfstime")
and upcase(name) ne upcase("status")
and upcase(name) ne upcase("hormon")
and upcase(name) ne upcase("pgr")
and upcase(name) ne upcase("meno")
and upcase(name) ne upcase("er")
and upcase(name) ne upcase("nodes")
and upcase(name) ne upcase("pid");
quit;
%put &var1 &var2 &var3 &var4 &var5 ;
* Fit Cox Model with 6 variables;
%macro Mselection2(var=);
proc phreg data=cancer;
model rfstime*status(0) = hormon pgr er meno nodes &var;
&var: test &var=0;
ods output FitStatistics=%str(&var)aic;
ods output TestStmts = %str(&var)stat;
run;
%mend;
* Model selection funciton;
%macro MS2;
%DO I=1 %TO 5;
%Mselection2(var=&&var&I)
%END;
%mend MS2;
%MS2;
* Summarize the result by first run;
DATA aic_table_2 (drop=withoutcovariates);
set %str(&var1)aic %str(&var2)aic %str(&var3)aic %str(&var4)aic
%str(&var5)aic ;
where criterion = "AIC";
cov=_n_;
run;
data stat_table_2(drop=status);
  set %str(&var1)stat %str(&var2)stat %str(&var3)stat %str(&var4)stat
  %str(&var5)stat ;
cov=_n_;
run;
proc sort data=aic_table_2;by cov;run;
proc sort data=stat_table_2;by cov;run;
data select_2;
merge aic_table_2 stat_table_2;
```

```
by cov;
run;
proc sort data=select_2;by ProbChiSq;
/*--------table------------*/
proc print data=select_2;
run;
*grade is selected;



/*------------Model selection 3----------*/

* Create macro-variables with value covariate ;
proc sql noprint;
select name into : var1-:var4
from dictionary.columns
where LIBNAME = upcase("work")
and MEMNAME = upcase("cancer")
and upcase(name) ne upcase("rfstime")
and upcase(name) ne upcase("status")
and upcase(name) ne upcase("hormon")
and upcase(name) ne upcase("pgr")
and upcase(name) ne upcase("meno")
and upcase(name) ne upcase("grade")
and upcase(name) ne upcase("er")
and upcase(name) ne upcase("nodes")
and upcase(name) ne upcase("pid");
quit;
%put &var1 &var2 &var3 &var4 ;
* Fit Cox Model with 6 variables;
%macro Mselection2(var=);
proc phreg data=cancer;
model rfstime*status(0) = hormon pgr er nodes grade meno &var;
&var: test &var=0;
ods output FitStatistics=%str(&var)aic;
ods output TestStmts = %str(&var)stat;
run;
%mend;
* Model selection funciton;
%macro MS2;
%DO I=1 %TO 4;
%Mselection2(var=&&var&I)
%END;
%mend MS2;
%MS2;
* Summarize the result by first run;
DATA aic_table_2 (drop=withoutcovariates);
set %str(&var1)aic %str(&var2)aic %str(&var3)aic %str(&var4)aic ;
where criterion = "AIC";
cov=_n_;
run;
```

```
data stat_table_2(drop=status);
  set %str(&var1)stat %str(&var2)stat %str(&var3)stat %str(&var4)stat ;
cov=_n_;
run;
proc sort data=aic_table_2;by cov;run;
proc sort data=stat_table_2;by cov;run;
data select_2;
merge aic_table_2 stat_table_2;
by cov;
run;
proc sort data=select_2;by ProbChiSq;
/*--------table------------*/
proc print data=select_2;
run;
*Nosignificant result;


/*-------- Model analysis ------------*/

*Fit PH model with 6 variables;
/*-----Plot-----*/
proc phreg data=cancer plots(overlay)=(survival);
class meno hormon grade(desc);
model rfstime*status(0) = hormon nodes pgr meno er grade;
run;

* Cox-snell;
proc phreg data = cancer;
  model rfstime*status(0) = hormon nodes pgr meno er grade ;
  output out = plot1_1 LOGSURV = logsurv1 /method = ch;  /*-logsurv is the cox-snell residual*/
run;

data plot1_1;
  set plot1_1;
  snell = -logsurv1;
  cons = 1;
run;
proc phreg data = plot1_1;
  model  snell*status(0) = 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;



* Graphical for PH assumption;
data cancer1;
  set cancer;
  cons = 1;
run;
* The baseline cumulative hazards are estimated using Breslows estimator for each stratu;
* plot log(h) function;
proc phreg data = cancer1 ;
class hormon/param=ref;
 model rfstime*status(0) = cons/rl ; * risk limit: provide CI for harzard ratio;
 strata hormon;
 output out = base_1 logsurv = ls /method = ch;
 run;

data base_1;
  set base_1 ;
    logH = log (-ls);
^^Iif hormon=0 then logH1 = logH;
    if hormon=1 then logH2 = logH;
    proc sort;by rfstime hormon ;
    proc print;var rfstime hormon logH logH1 logH2 ;
run;

proc sgplot data =base_1;
where logH ne .;
series x=rfstime y=logH /group=hormon ;
run;

*Plot Log difference function;
proc sort data = base_1;
 by rfstime;
run;
data base_2;
  set base_1;
  retain temp1 temp2;
   if logH1 ~= . then temp1 = logH1;
   if logH2 ~= . then temp2 = logH2;
   diff2v1 = temp2 - temp1;
   proc print;
   var  rfstime hormon temp1 temp2 diff2v1 ;
run;
proc sgplot data =base_2;
step x=rfstime y=diff2v1  ;
lineparm x=0 y=0 slope=0 /lineattrs=(color= red ); /** intercept, slope **/
run;
```

```
* Plot log(ht)/log(ht);
proc sort data = base_1;
by rfstime;
run;
data base_3;
  set base_1;
  retain H1 H2;
  *retain its value from one iteration of the DATA step to the next;
  if hormon=0 then H1 = -ls;
  if hormon=1 then H2 = -ls;
   proc print;
   var rfstime hormon ls H1 H2  ;
run;

proc sgplot data =base_3;
title "Anderson Plot";
series x=H1 y=H2 ;
lineparm x=0 y=0 slope=0.7/lineattrs=(color= red ); /** intercept, slope **/
run;

/*---------PH for hormone---------*/
* Shoenfeld residual for time with hormon;
proc phreg data = cancer1 plots=survival ;
class hormon/param=ref;
 model rfstime*status(0) =nodes pgr meno grade hormon er/rl ;
 output out=schoen
 ressch=schnodes schpgr schgrade schhon scher schmeno ;
run;

proc loess data = schoen plots=FITPLOT;
model schhon=rfstime / smooth=(0.5);
run;
* Shoenfeld residual for hormon time correlation test;
* p-value is 0.6552 means the ph assumption is not violated;
proc sort data=schoen;by rfstime;run;
data schoen;
set schoen;
retain temp 0;
temp=temp+1;
rank=temp;
proc print ;var hormon rank;
run;
proc corr data=schoen; var rank schhon;run;

/*-----------PH for nodes--------*/
* Shoenfeld residual for time with nodes;
* Both plot and p-value 0.2036 shows the nodes meet assumption;
proc loess data = schoen plots=FITPLOT;
model schnodes=rfstime / smooth=(0.5);
run;
```

```
proc corr data=schoen; var rank schnodes;run;

/*-----------PH for pgr--------*/
* Shoenfeld residual for time with pgr;
* Both plot and p-value 0.0723 shows the pgr meet ph assumption;
proc loess data = schoen plots=FITPLOT;
model schpgr=rfstime / smooth=(0.5);
run;
proc corr data=schoen; var rank schpgr;run;

/*-----------PH for er--------*/
* Shoenfeld residual for time with pgr;
* Both plot and p-value 0.0570 shows the pgr meet ph assumption;
proc loess data = schoen plots=FITPLOT;
model scher=rfstime / smooth=(0.5);
run;
proc corr data=schoen; var rank scher;run;

/*-----------PH for grade--------*/
* Shoenfeld residual
* Both plot and p-value 0.0076 shows the grade  does not meet ph assumption;
proc loess data = schoen plots=FITPLOT;
model schgrade=rfstime / smooth=(0.5);
run;
proc corr data=schoen; var rank schgrade;run;

/*-----------PH for meno--------*/
* Shoenfeld residual for time with meno;
* Both plot and p-value 0.0616 shows the grade  does not meet ph assumption;
proc loess data = schoen plots=FITPLOT;
model schmeno=rfstime / smooth=(0.5);
run;
proc corr data=schoen; var rank schmeno;run;
```