Practicum Report

The Comparison of log-normal with Cox Model Estimates for the Adjuvant Colon Cancer Endpoints (ACCENT) database

Nianping HU

Supervisor: Dr. Keyue Ding

Dr. Judy-Anne Chapman

Department of Community Health & Epidemiology

Queen's University

2010-09-24

Acknowledgments

I would like to express my gratitude to my supervisor, Dr. Keyue Ding, who gave me many great advices and provided at all levels of assistance for my Master program. I appreciate his vast knowledge and skill in mathematics and biostatistics. I would like to thank my co-supervisor, Dr. Judy-Anne Chapman, who helped me to analyze this ACCENT database, Dr. Chapman provided me the analysis direction, trouble-shooting, and technical assistance, and without her help I can't finish my practicum. I also acknowledge Dr. Bingshu Chen, who took time to answer my question from his busy schedule, he always explained very clear, and make the theory become easy to understand. I would like also to thank Weidong Kong, who gave me many helps for SAS and R programming.

I would like to thank NCIC CTG for offering me this space and computer for my practicum.

I would also like to thank my family for the support; I must acknowledge my wife, Manhong, without whose love and encouragement, I would not have finished this biostatistics program.

Table of contents

Introduction

Colorectal cancer is the 2nd leading cause of cancer death in Canada. For 2009 there are 22,000 new cases and 9,100 deaths.

The treatment strategy of colon cancer has improved at several decade years, from surgery only developed to surgery with adjuvant therapy combination; the adjuvant therapy has been a success treatment for colon cancer over last thirty years, it includes chemotherapy, Radiation therapy, Immunotherapy etc.

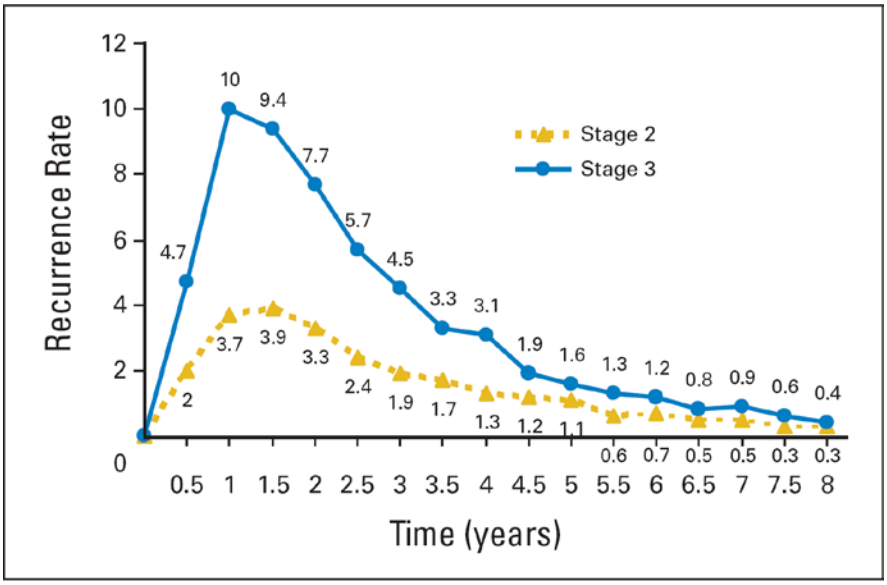
In colon cancer, chemotherapy is often applied after surgery (adjuvant), before surgery (neo-adjuvant), or as the primary therapy (palliative). Adjuvant chemotherapy is usually involved in the combination of infusional 5-fluorouracil (5-FU) with other chemotherapy regimens, Commonly first line used 5-FU, leucovorin (LV), and oxaliplatin (FOLFOX) with bevacizumab or infusional 5-FU, LV and irinotecan (FOLFIRI) with bevacizumab [(1)], if the colon cancer patients have KRAS wild type tumors, the chemotherapy can choose the same drug combinations with cetuximab [(2)], the 5-FU-base chemotherapy have been shown by clinical trials that can improves survival and/or reduces mortality rate. The adjuvant chemotherapy becomes a routine treatment for the colon cancer spreading to the lymph nodes.

Radiotherapy is not used commonly in colon cancer, as it could lead to side effect, such as radiation enteritis, and it is difficult to target colon cancer since the colon changes location often, it usually chooses for decreasing pain and palliation in colon cancer; Radiotherapy is more common for radiation to be used for rectal cancer,

Immunotherapy is a new approach for colon cancer, Bacillus Calmette-Guérin (BCG) [(3)] or interferon can treat colon cancer patients to enhance the host immune system, or cancer vaccine and other targeted drug can identify and attack specific cancer cells without harming normal cells [(4)].

Adjuvant Colon Cancer Endpoints (ACCENT) is a database of 18 international colon cancer clinical trials and 20,898 patients that is coordinated by Dr. Daniel Sargent at Mayo Clinic, clinical trial group in Queen's University is part of these international trials. Dr. Sargent used COX Model to analyze the ACCENT dataset, and found the 2 or 3 years Disease-Free-Survival (DFS) outcomes are retained an excellent concordance with 5 year Overall-Survival

(OS), but not for 1 year DFS. These findings suggest 2- or 3-years DFS may be an appropriate primary end point to be considered instead the longer 5-years OS in the future trials. This work was the basis for FDA approval of DFS as a surrogate for OS for colon cancer. Dr. Sargent, et al. found substantive evidence against the Cox assumption of proportional hazards, and the figure for stage 2 and stage 3 ACCENT database patients' survival function from Dr. Sargent's paper shown empirical log-normal shaped hazard functions, and this plot was pivotal in motivating this work [5].



Drs. Judy-Anne Chapman and Christopher J. O'Callaghan have approved use of the data to investigate the efforts of a new paradigm of parametric (log-normal) survival analysis on efficacy estimates, and to compare log-normal estimates with standard Cox model estimates.

Practicum objectives and goals

The log-normal parametric model is applied to analysis the ACCENT dataset, and compares with the standard COX model, an appropriate parametric model will be generated by comparison, which can assess this dataset more accurately and efficiently.

Training during the Practicum

We used the SAS to do the data analysis and most of the figures; and some of the figures were done with R.

In SAS program, we used **proc freq** to check the data quality, and identify the missing data from the variables, also, the subjects from groups can be counted, we can always double check if we picked right amount of subjects for later analysis. For residual analysis, I chose **proc lifereg** to check the COX-Snell residual and **proc gplot** was performed to make the residual graph. The homogeneity test was used to check the difference between all trials or treatments, and the appropriate treatments were selected and combined together to compare with other treatments, in this particular dataset, we first defined the Overall-Survival (OS), Time-To-Relapse (TTR) and Disease-Free-Survival (DFS), then **proc lifereg** was proceeded to analyze all trials and treatment for these three endpoints. We then applied a MACRO to stepwise select the multivariate and identify the significant variables for further lognormal and COX model fit. During the model fitting, we used the **proc lifereg** for lognormal model and **proc phreg** for COX model, and we practiced how to calculate the survival probability from estimated baseline survival function for various given subset of covariates.

We also trained to use SAS and R to make graph, such as residual plot, histogram figure and survival plot.

During the practicum, we learnt how to deal with the real data, practiced many SAS and R programming skills, this was a great opportunity for me to study in NCIC CTG, this work experience will benefit me in my professional field.

Main activities of the practicum

Procedure for analysis the ACCENT dataset

- **Data quality check**
- **Endpoint definitions**
- **Q-Q plot**
- **Univariate effects of therapy**

- **Homogeneity testing log-normal step-wise multivariate selection by SAS MACRO of Ryan Browne, COX step-wise multivariate selection by SAS**
- **Residual analysis**
- **log-normal and COX model Comparison**
- **Survivor Plots**

Data quality check

We checked the quality for ACCENT dataset, and found there are 3 data records with missing of gender, 4 with stage and 939 with missing value in arm. Additionally, there were patients missing endpoint data; the previous ACCENT database use of using maximal information for each endpoint was used here. We also followed the previous ACCENT database practice of truncating follow-up at 8 years for a patient who had not experienced a particular event since follow-up after that was not uniform across all trials. There were 2454 subjects allocated to receive only surgery, and 18444 subjects allocated to systemic chemotherapy combination. For the chemotherapy combination group, 4534 subjects were allocated by 5-FU combined with LV with or without LEV, and 13910 subjects other combination by other chemotherapy. (Details wasn't include)

Endpoint definitions

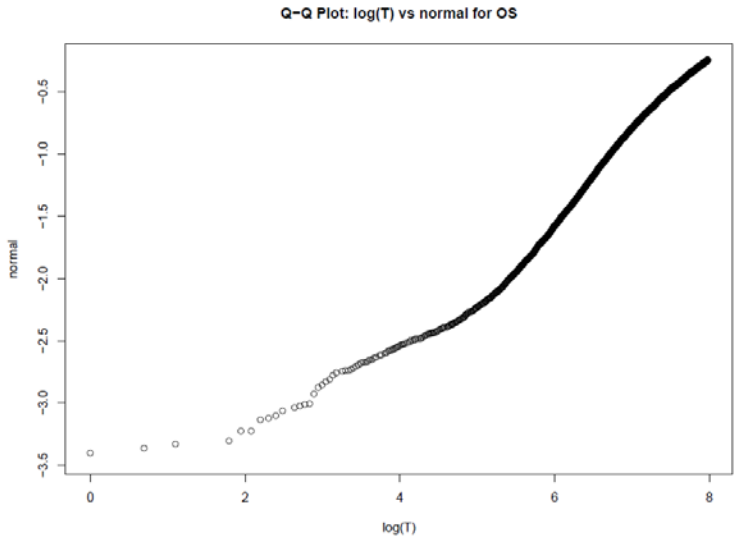
Overall-Survival (OS) defined as the time from randomization ("Registration") to the date of death (indicator factor named "FollowupStatus", with "FollowupDate").

Time-to-Relapse (TTR) defined as the time from study entry ("Registration") to the date of the first confirmed relapse (indicator factor named "Progression", with "Progression date"); for patients who died without recurrence, they were censored at time of death for time-to recurrence analyses.

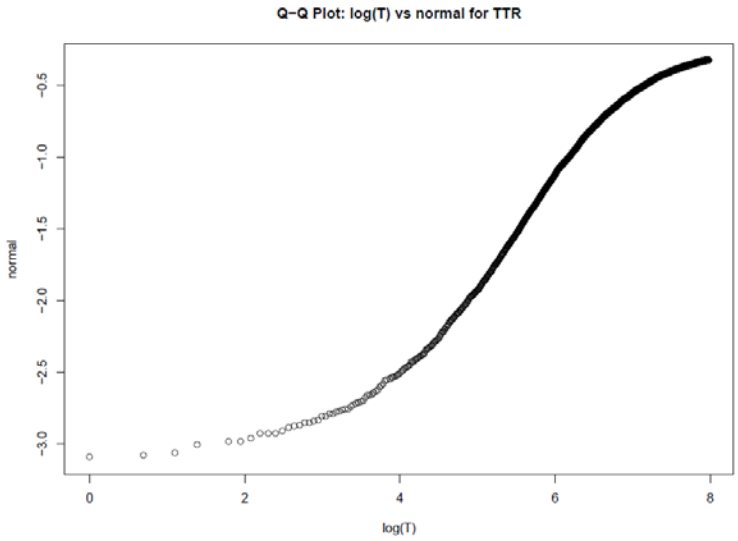
Disease-Free-Survival (DFS) defined as the time from randomization ("Registration") to the first event of either recurrent disease (indicator factor named "Progression", with "Progression date"), or if no recurrence (Progression), death (indicator factor named "FollowupStatus", with "FollowupDate").

Q-Q plot

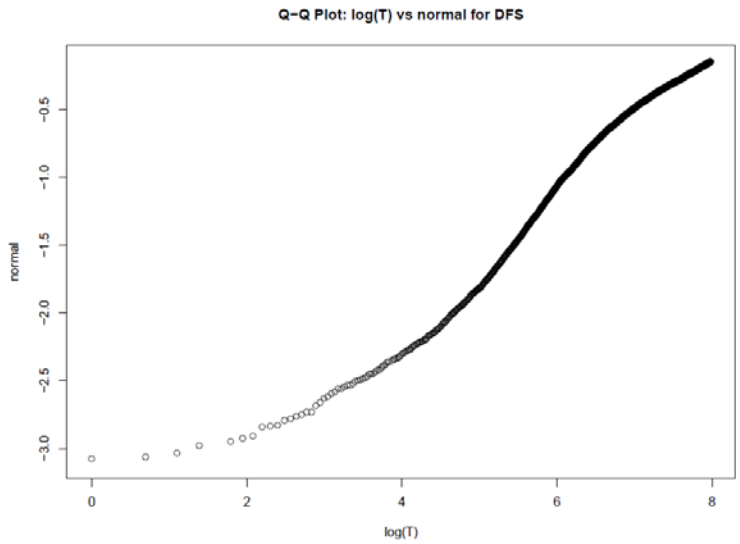
Q-Q plot for Overall-Survival



Q-Q plot for Time-To-Relapse



Q-Q plot for Disease-Free-Survival



Univariate effects of therapy

Summary of Investigative Results:

1. **Can control arms of surgery be combined?**

9 trials had **control arms of surgery**: by trial tests of beta being zero indicate similar survival across all trials, except for NCIC CTG patients

Test of homogeneity for these 9 arms:likelihood ratio criterion, -2logR= -2(Li - L) ~ chi-square with 8 df, where Li is maximum likelihood with 9 beta mle for control arms and L is maximum likelihood with single common beta mle.

➔ **p<0.001 for OS, TTR, DFS**

There was no single or group of common trials across 3 endpoints, which if removed from test, would lead to remaining trials having similar (non-significant) test criterion.i.e. Removing NCIC trial leads to OS p-value of 0.06, but p<0.01 for TTR, DFS.

➔ **Conclusion:** Maintain 9 surgery arms.

---

2. **Can any of the 23 treatment arms in 18 trials be combined?**
   a. **Test for differences with addition of LEV**

with likelihood ratio criterion -2logR= -2(Li - L) ~ chi-square with 1 df, where Li is maximum likelihood with beta mle of 2 trial treatment arms and L is maximum likelihood with single common beta mle for the 2.

5FU+175L-folinic acid+Placebo

vs 5FU+175L-folinic acidV+LEV

➔ **OS: p=0.04**

TTR: p=0.12

DFS: p=0.04

➔ **Conclusion**: Do not combine arms.

5FU+25L-folinic acid+LEV

vs 5FU+25L-folinic acid+Placebo

➔ **OS: p=0.63**

TTR: p=0.99

DFS: p=0.95

➔ **Conclusion:** Arms can be combined; tests below indicate a third arm which can be combined with these.

5FU+LV

vs 5FU+LV+LEV

➔ **OS: p=0.44**

TTR: p=0.35

DFS: p=0.85

➔ **Conclusion:** Arms can be combined.

---

b-1. **Test for differences by dose**

with likelihood ratio criterion -2logR= -2(Li - L) ~ chi-square with 1 df, where Li is maximum likelihood with beta mle of 2 trial treatment arms and L is maximum likelihood with single common beta mle for the 2.

5FU+175L-folinic acid+Placebo

vs 5FU+25L-folinic acid+Placebo

➔ **OS: p=0.82**

TTR: p=0.50

DFS: p=0.66

➔ **Conclusion:** Arms can be combined; tests below indicate a third arm which can be combined with these.

5FU+175L-folinic acidV+LEV

vs 5FU+25L-folinic acid+LEV

➔ **OS: p=0.16**

TTR: p=0.38

DFS: p=0.13

➔ **Conclusion:** Arms can be combined; however, the next test indicates a 3 arm collapse.

5FU+175L-folinic acid+Placebo

vs 5FU+25L-folinic acid+Placebo

vs 5FU+25L-folinic acid+LEV

➔ **OS: p=0.47 df=2**

TTR: p=0.44 df=2

DFS: p=0.59 df=2

➔ **Conclusion:** Arms can be combined.

5FU+HDLV

vs 5FU+LDLV

➔ **OS: p=0.36**

TTR: p=0.68

DFS: p=0.41

➔ **Conclusion:** Arms can be combined.

High Dose LEV+5FU/CF

vs Standard Dose LEV+5FU/CF

➔ **OS: p=0.23**

TTR: p=0.58

DFS: p=0.36

➔ **Conclusion:** Arms can be combined.

---

b-2. **Tests for difference by duration**

5FU+LEV x 1yr

vs 5FU+LEV x 6 mos

➔ **OS: p=0.06**

TTR: p=0.15

DFS: p=0.08

➔ **Conclusion:** Weak evidence for difference; do not combine arms.

5FU+LEV+LV x 1yr

vs 5FU+LEV+LV x 6 mos

➔ **OS: p=0.04**

TTR: p=0.25

DFS: p=0.19

➔ **Conclusion:** Do not combine arms.

---

c. **Test for differences between 5FU+LV versus everything else**

(After confirming that 5FU+LV vs 5FU+LV+Interferon should not be combined,

OS: p=0.37, TTR: p=0.01, DFS: p=0.11, and 5FU+LV+/-LEV could be combined, OS: p=0.44, TTR: p=0.35, DFS: p=0.85)

Li with beta of 5FU+LV and pooled all others, vs common beta for all trials for likelihood ration test of 1 df;

➔ **p<<0.0001 for OS, TTR, DFS.**

➔ **Overall Conclusions:** Surgery alone is no longer a therapy offered to current patients so patients who had only surgery will have separate multivariate analyses. For multivariate analyses with patients who received systemic therapies, therapy arms could be collapsed as indicated above; however, at this point as indicated on the next page's line-up of factors, we

have only looked at the dominant classification of previous standard therapy (5FU+LV+/-LEV) versus all others.

Log-normal step-wise multivariate selection by SAS MACRO of Dr. Ryan Browne

TTR, DFS, OS Step-wise forward log-normal likelihood where factor is entered if $p \leq 0.05$, by likelihood ratio criterion $\sim C^2_{(1)}$

Class variable control (v. below) is going to be in separate runs to Class variable treated (vi.).

I. For Patients allocated surgery alone, class variable control (surgery by trial) TTR, DFS, OS runs.

Factors to be considered:

i. registration year of patient
ii. Age
iii. Sex
iv. Stage
v. class variable control (surgery by trial) - no collapses are possible

II. For Patients allocated to adjuvant systemic therapy, class variable treated TTR, DFS, OS runs:

Factors to be considered:

i. registration year of patient
ii. Age
iii. Sex
iv. Stage
v. class variable treated with programmed collapses of following treatments
   a. 5FU + LV +/-LEV
   b. all others

Step-wise multivariate selection

The program consists of four MACROs: fitModel, updateResult, fitInactiveSets,

forwardSelect

fitModel (result, var, cvar): given the variables in cvar (class variables) and var (continous variables) a survival model is fitted and the loglik and number of parameters are output to the dataset result.

updateResult (old, new, newv, curv): Adds the variables newv and curv to the dataset new and then updates the dataset old with the data from new.

forwardFitting (result, actv, inactv, actcv, inactcv)

actv = active continuous variable set

inactv = inactive continuous variable set

actcv = active class variable set

inactcv = inactive class variable set

Fit a survival model use fitModel for every variable in inactv and inactcv while including with the variables in actv and inactv. then saves the output from fitModel for each model along with the variable name in the dataset result.

forwardSelect (out, data, actv, actcv, inactv, inactcv, maxiter=30, cutoff=0.05): performs forward selection on the variables sets inactv and inactcv given that the variables actv and actcv are already in the model. It continues until maxiter is hit or no more variables have pvalues from the asymptotic loglikelihood test (a chisquare test) are less than the cutoff It first calls fitModel with the variables from actv and actcv Then calls fitInactiveSets with the corresponding sets.

Summary of Multivariate Results

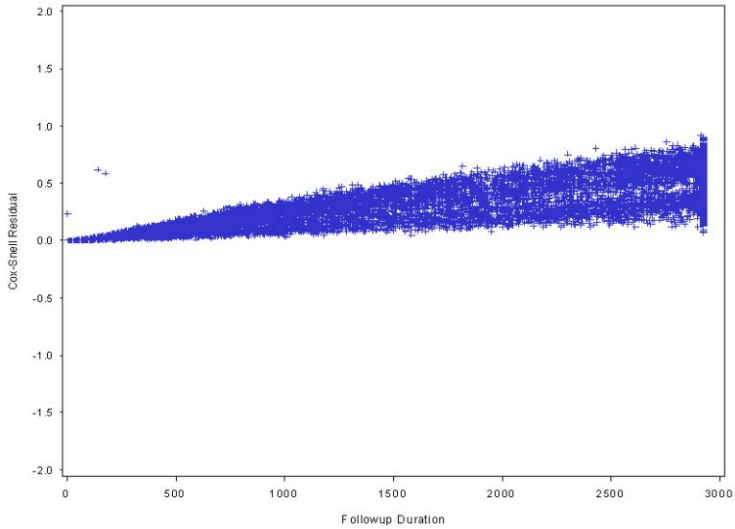
	All Patients			Patients with Systemic Treatment			Patients with Surgery Alone		
	Log-normal		Cox	Log-normal		Cox	Log-normal		Cox
	P-value[1]	P-value[2]	P-value[2]	P-value[1]	P-value[2]	P-value[2]	P-value[1]	P-value[2]	P-value[2]
OS Age	<0.0001,	<0.0001	<0.0001	<0.0001,	<0.0001	<0.0001	<0.0001,	<0.0001	<0.0001
Stage	<0.0001,	<0.0001	<0.0001	<0.0001,	<0.0001	<0.0001	<0.0001,	<0.0001	<0.0001
Treatment (5FU+LV+/-LEV, Other systemic, Surgery)	<0.0001,	<0.0001	<0.0001	<0.0001,	<0.0001	<0.0001	NA,	NA	NA
Gender	0.005,	<0.0001	<0.0001	0.01,	<0.0001	<0.0001	NA,	NA	NA
Registration year	0.04,	<0.0001	<0.0001	NA,	NA	NA	NA,	NA	NA
Male	(same p-values (Gender excluded)			(same p-values (Gender excluded)					
Female	(same p-values Age, Stage)			(same p-values Age, Stage)					
Treatment	0.001,	<0.0001		NA,	NA				
TTR Stage	<0.0001,	<0.0001	<0.0001	<0.0001,	<0.0001	<0.0001	<0.0001,	<0.0001	<0.0001
Treatment (5FU+LV+/-LEV, Other systemic, Surgery)	<0.0001,	<0.0001	<0.0001	<0.0001,	<0.0001	<0.0001	0.03,	-----	0.003
Male	(same p-values)			(same p-values)					
Female	(same p-values Stage)			(same p-values Stage)					
Treatment	0.004,	<0.0001		0.10,	NA				
DFS Age	<0.0001,	<0.0001	<0.0001	<0.0001,	<0.0001	<0.0001	0.02,	<0.0001	<0.0001
Stage	<0.0001,	<0.0001	<0.0001	<0.0001,	<0.0001	<0.0001	<0.0001,	<0.0001	<0.0001
Treatment (5FU+LV+/-LEV, Other systemic, Surgery)	<0.0001,	<0.0001	<0.0001	<0.0001,	<0.0001	<0.0001	NA,	NA	NA
Gender	0.04,	<0.0001	<0.0001	0.10,	NA	NA	NA,	NA	NA
Registration year	0.08,	NA	NA	NA,	NA	NA	NA,	NA	NA
Male	(same p-values, Age, Stage)			(same p-values, Gender excluded)					
Registration year	0.04,	<0.0001							
Female									
Registration year	<0.0001,	0.07		NA,	NA				
Stage	<0.0001,	<0.0001		<0.0001,	<0.0001				
Age	<0.0001,	<0.0001		<0.0001,	<0.0001				
Treatment (5FU+LV+/-LEV, Other systemic, Surgery)	<0.001,	<0.0001		0.02,	<0.0001				

[1]Likelihood ratio criterion p-value for step-wise entry into model.
[2]Wald p-value in final model.

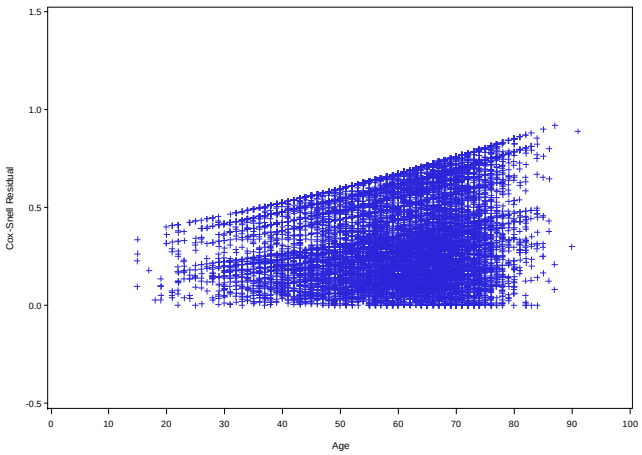
Please find the details for multivariate selection and lognormal, COX model fitting in Appendix part-2

Lognormal Model residual analysis

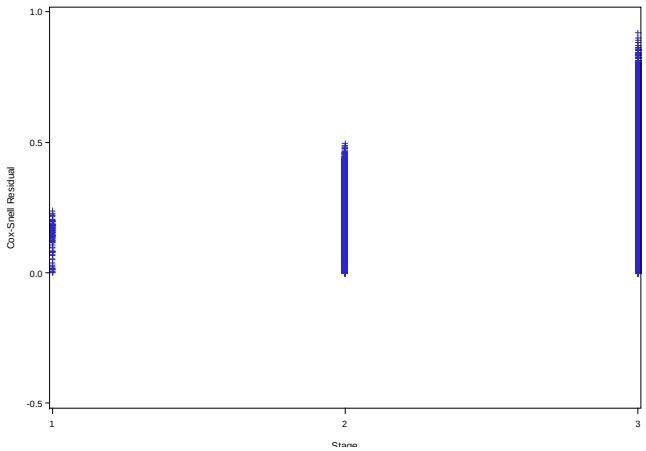
COX-Snell Residual vs followup time



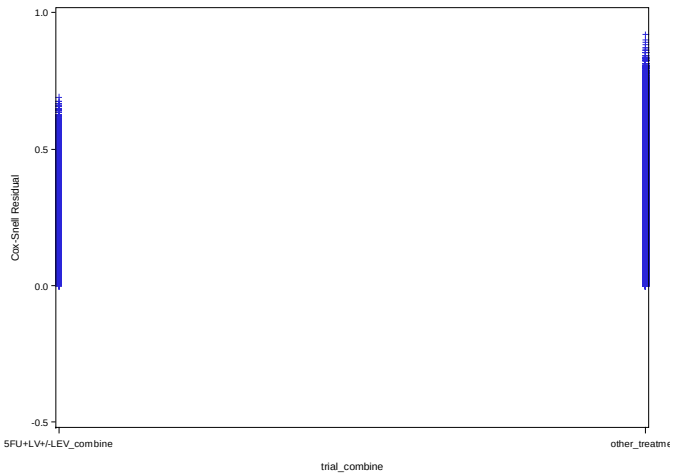
COX-Snell Residual vs Age



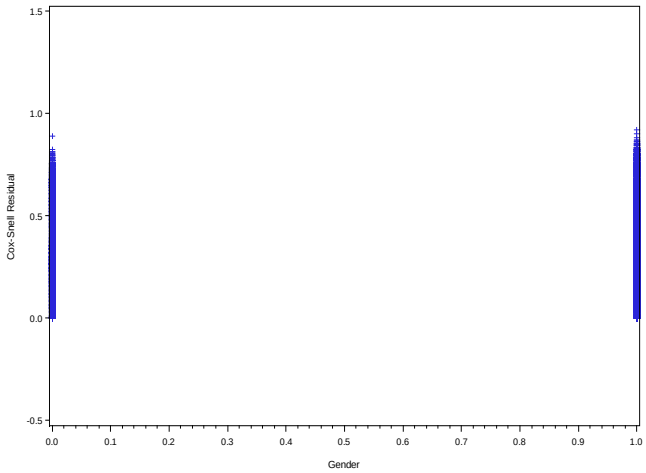
COX-Snell Residual vs Stage



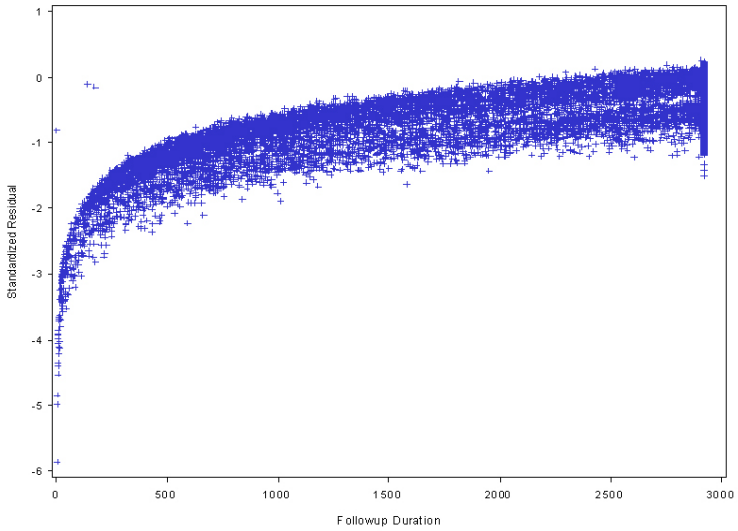
COX-Snell Residual vs Treatment



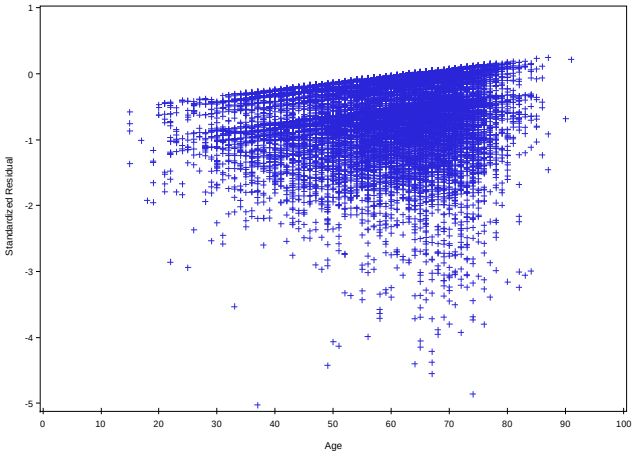
COX-Snell Residual vs Gender



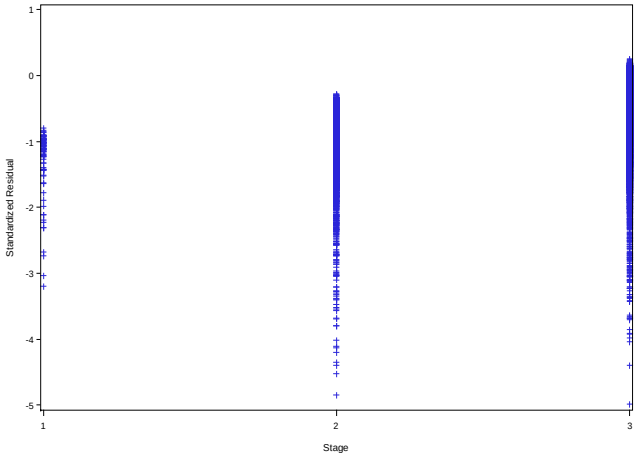
Standardized Residual vs followup time

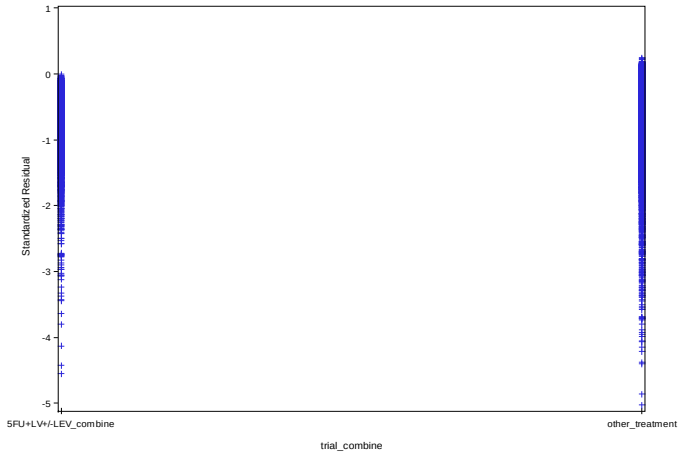


Standardized Residual vs Age

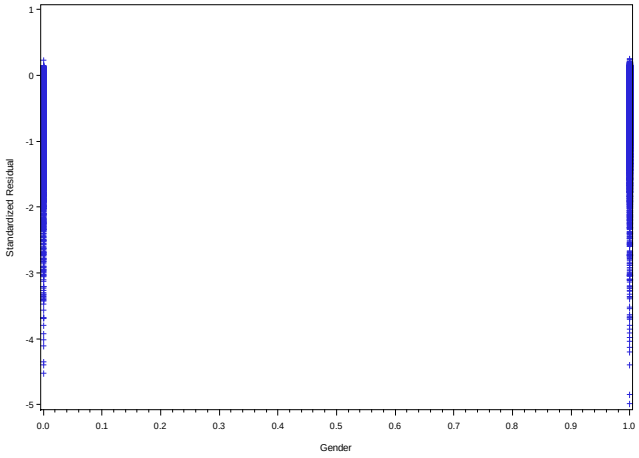


Standardized Residual vs Stage





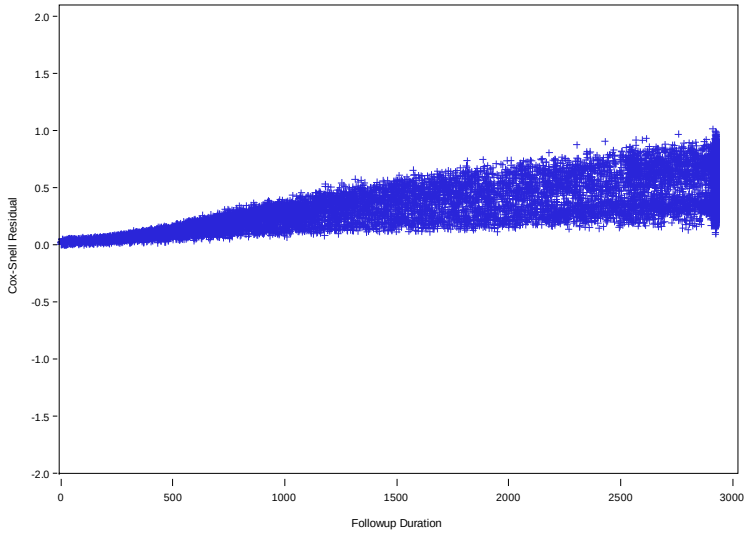
Standardized Residual vs Treatment



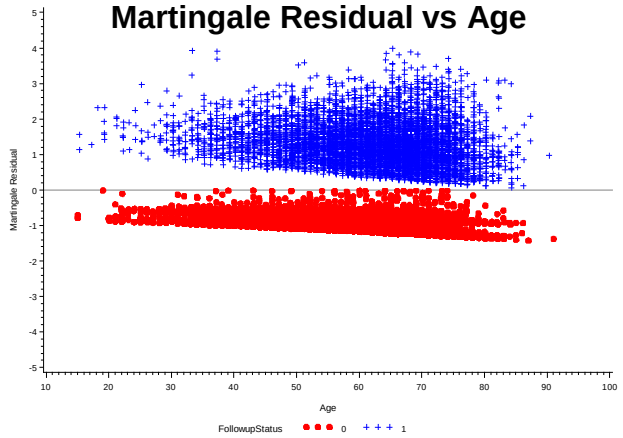
Standardized Residual vs Gender

COX model Residual analysis

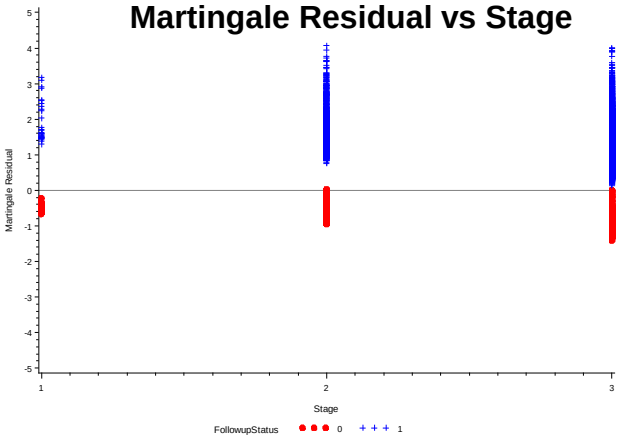
COX-Snell Residual vs followup time



Martingale Residual

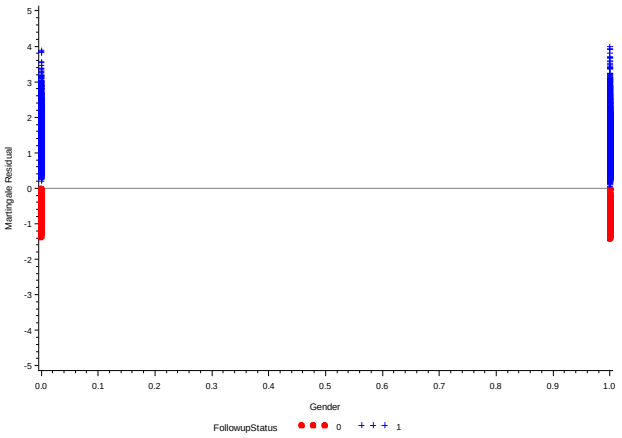


Martingale Residual vs Age



Martingale Residual vs Stage

Martingale Residual vs Treatment



Martingale Residual vs Gender

I also did the COX-Snell Residual vesus age, stage, treatment and gender for COX model; the figures are similar to the figures in lognormal model, so I didn't include the results here. The residuals analysis for TTR and DFS, Please see Appendix (part 3)

Comparison between log-normal and COX model

Survival function for Log-normal, where α is the intercept, β_i is estimates for individual variable, σ is the scale.

$$W=(y - \alpha - \sum \beta_i Z_i) / \sigma \sim N(0,1)$$

$$S_i(t) = 1 - \Phi (W)$$

Survival function for COX model

$$S_i(t|x_i) = S_0(t)^{Exp\{\Sigma x_i \beta_i\}}$$

For example, if we want calculate the survival probability for 60 yrs old, male, stage-II colon cancer patients, treated by 5-FU + LV +/- LEV, 3 yrs DFS for lognormal and COX model. We can calculate by following step.

For lognormal:

W = (log (365.25 * 3) – 11.231 - (- 0.0138)*60 - (- 0.8836) * 2 - 0.3475)/1.798 = -1.103507

S (3 yrs DFS) = 1 – NORMDIST (-1.103507) = 0.865096

For COX Model:

We first need get the baseline level $S_0(t)$

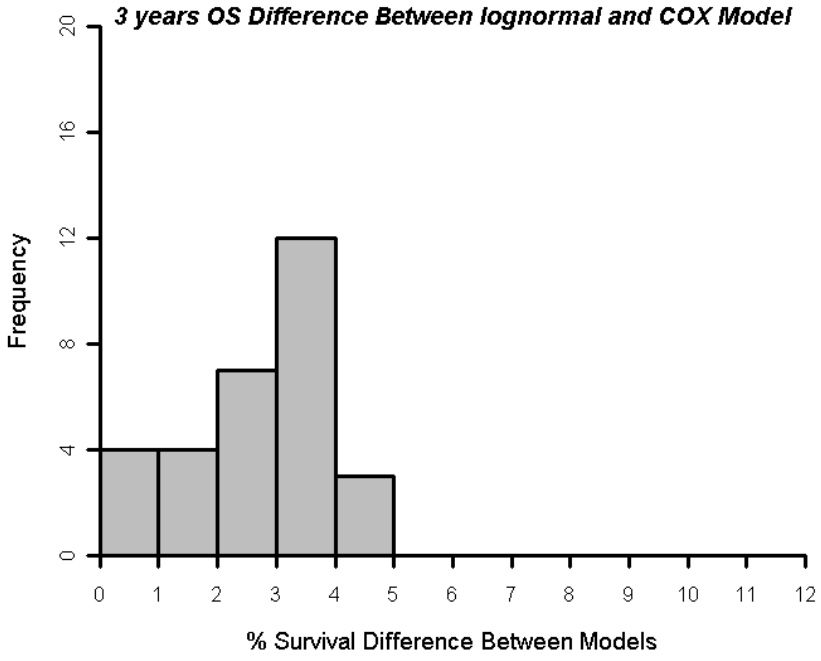
Male		Female	
Endpoint	$S_0(t)$	Endpoint	$S_0(t)$
3 yrs OS	0.9847	3 yrs OS	0.98808
3 yrs TTR	0.96624	3 yrs TTR	0.96692
3yrs DFS	0.96891	3yrs DFS	0.96832

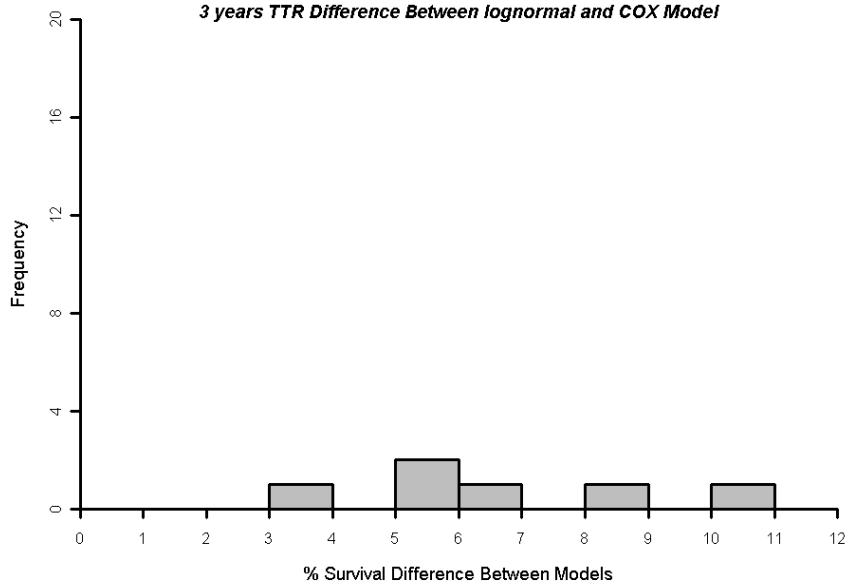
5 yrs OS	0.9751	5 yrs OS	0.98162
5 yrs TTR	0.95723	5 yrs TTR	0.95971
5yrs DFS	0.959	5yrs DFS	0.96033

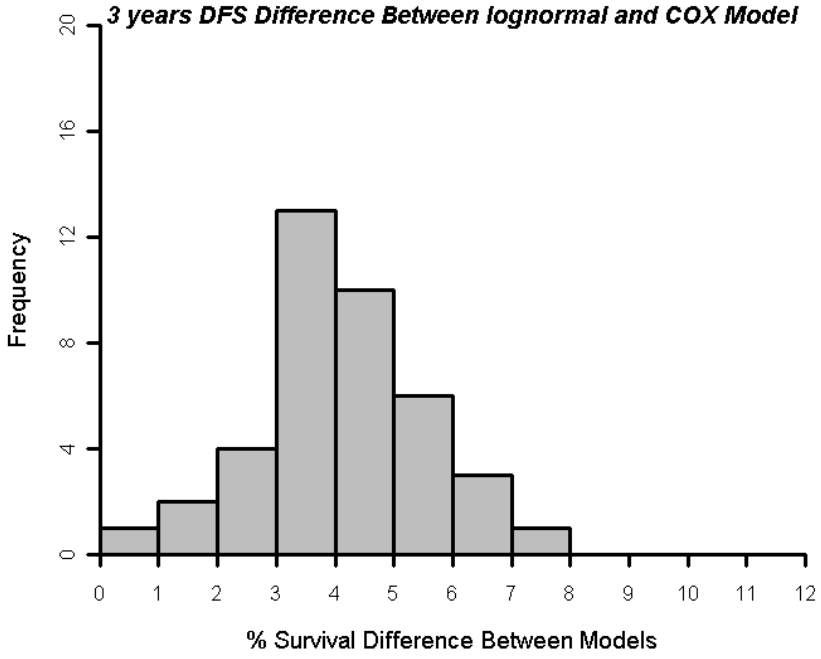
We can get the survival probability for COX model then.

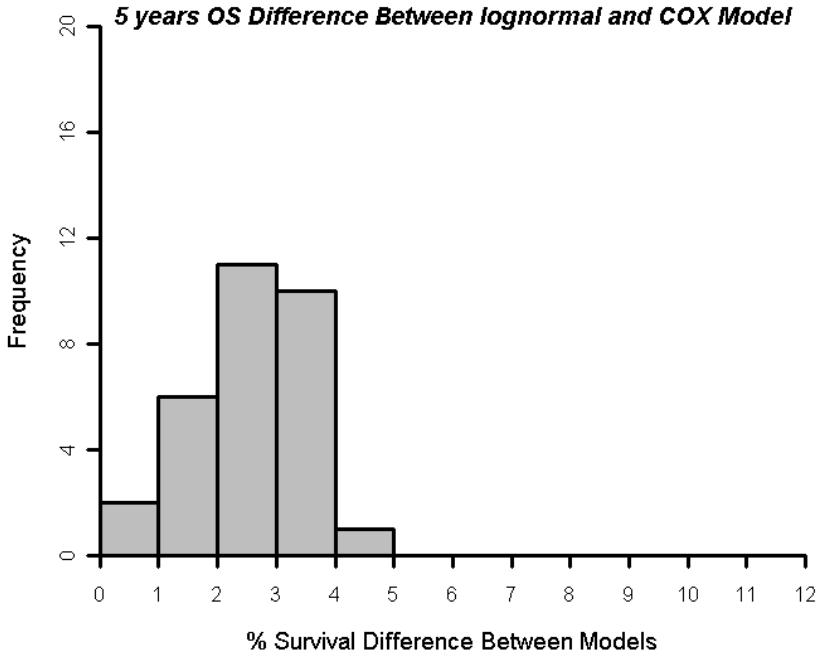
S (3 yrs DFS) = (0.96891) ** Exp (0.01249 * 60 + 0.65176 * 2 + (-0.25268)) = 0.82644

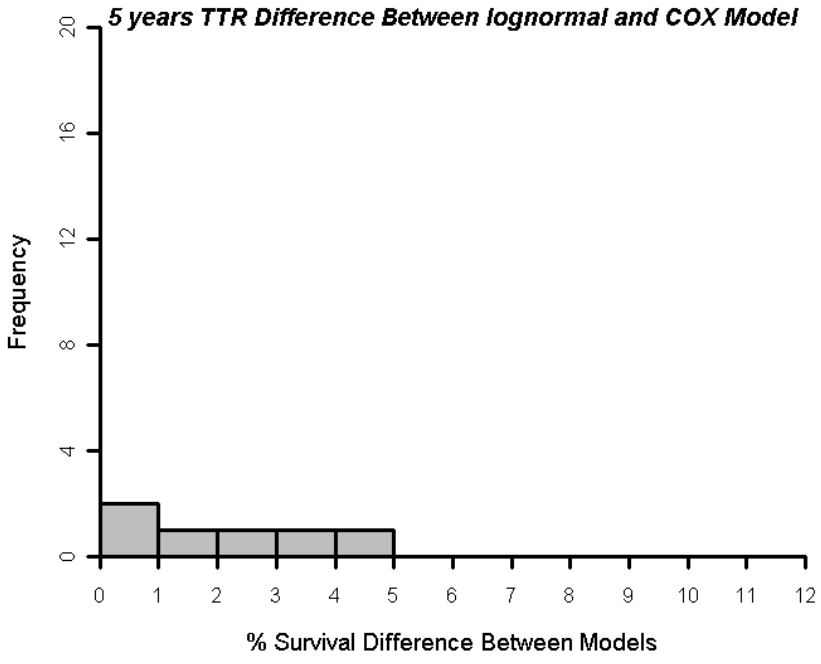
We further calculated the absolute difference for survival probability between lognormal and COX model for various given subset of covariates, and plotted the histogram as following graph. (Please see the details for survival probability for lognormal and COX model and absolute difference in Appendix Part 2).

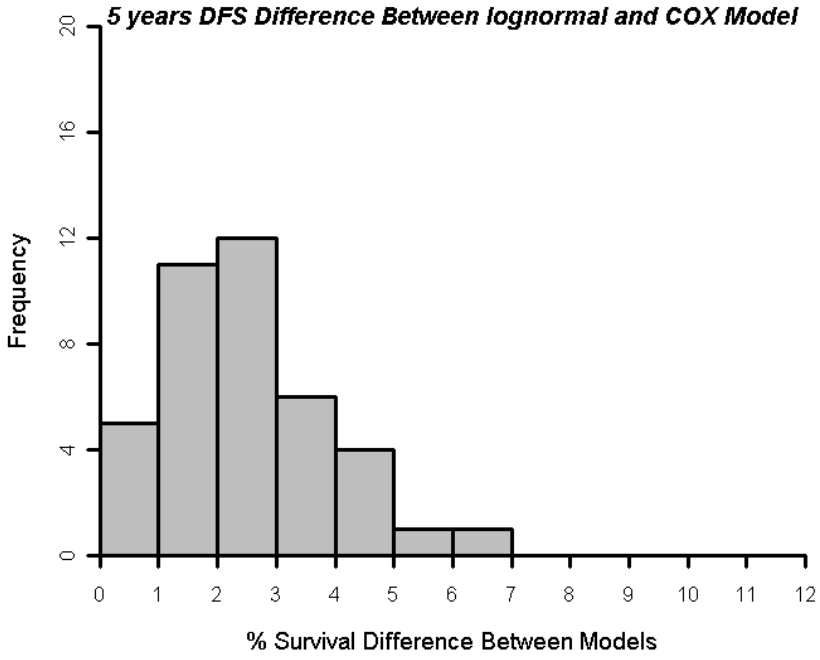



3 years TTR Difference Between lognormal and COX Model
Frequency
0
4
8
12
16
20
0 1 2 3 4 5 6 7 8 9 10 11 12
% Survival Difference Between Models


3 years DFS Difference Between lognormal and COX Model
Frequency
0
4
8
12
16
20
0 1 2 3 4 5 6 7 8 9 10 11 12
% Survival Difference Between Models


5 years OS Difference Between lognormal and COX Model
Frequency
0
4
8
12
16
20
0 1 2 3 4 5 6 7 8 9 10 11 12
% Survival Difference Between Models

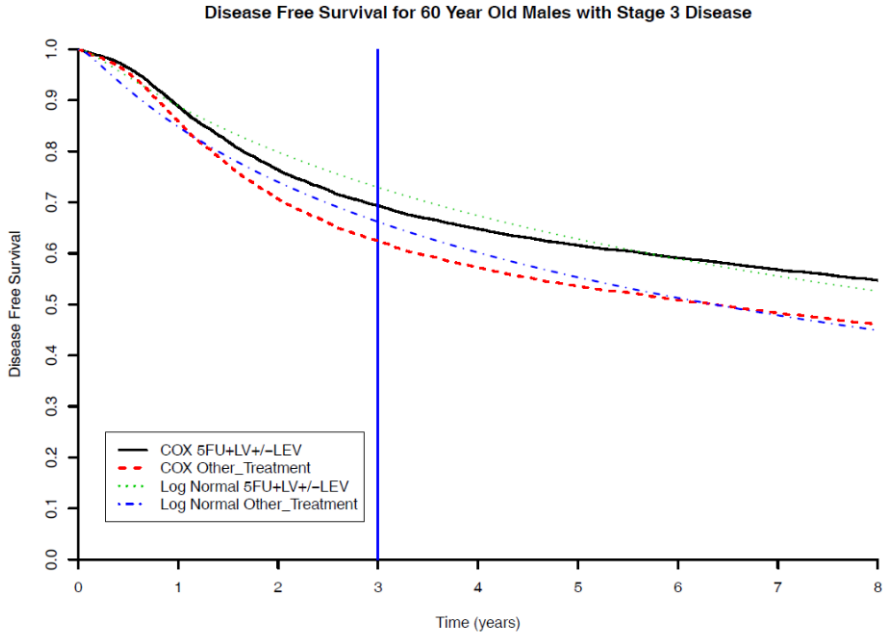

5 years TTR Difference Between lognormal and COX Model
Frequency
0
4
8
12
16
20
0 1 2 3 4 5 6 7 8 9 10 11 12
% Survival Difference Between Models



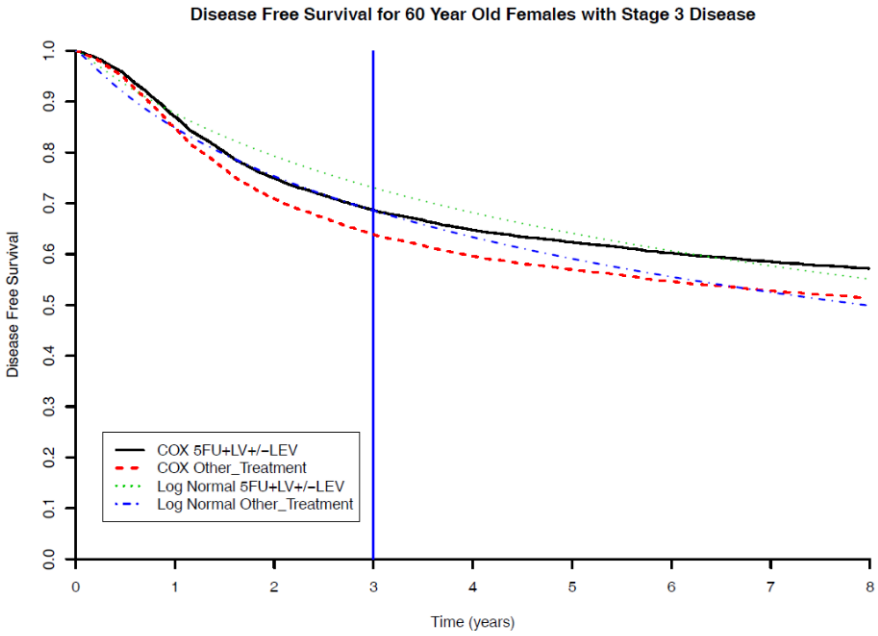
Survivor Plots

Our goal for this analysis is whether 2- or 3 yrs DFS has good concordance with 5 yrs OS, so we made the 3 yrs DFS and 5 yrs OS survival curve for 60 yrs, male or female, treated by 5-FU+LV+/-LEV or other-treatment by lognormal and COX model,

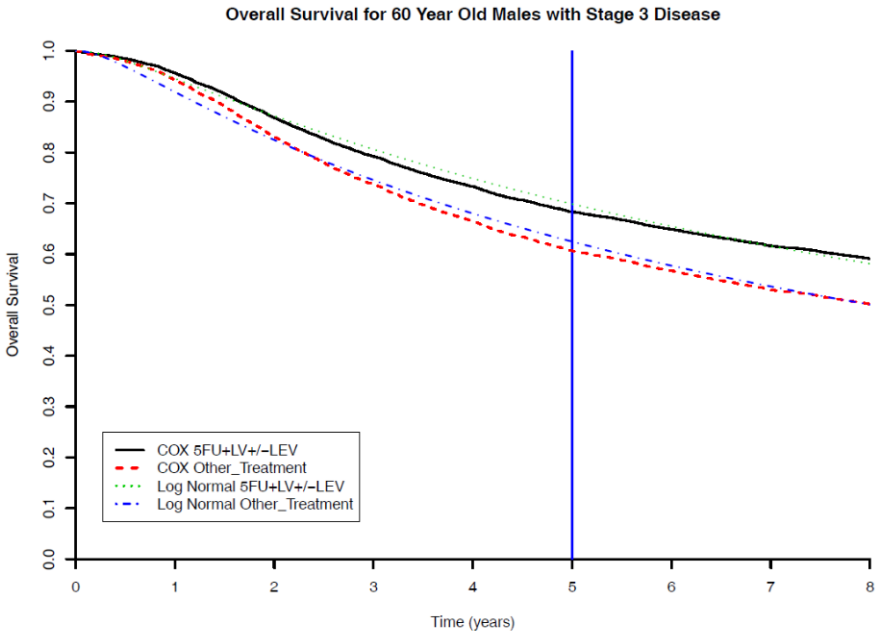
1. 3 yrs DFS for 60-years male (5-FU+LV+/-LEV vs other-treatment)



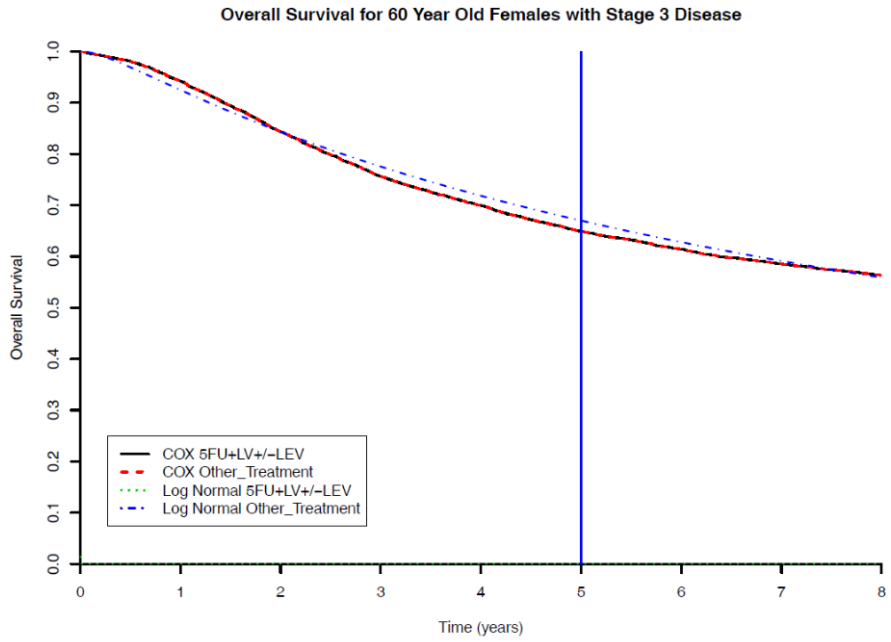
2. 3 yrs DFS for 60-years female (5-FU+LV+/-LEV vs other-treatment), (lognormal vs COX)



3. 5 yrs OS for 60-years male (5-FU+LV+/-LEV vs other-treatment), (lognormal vs COX)



4. 5 yrs OS for 60-years female (5-FU+LV+/-LEV vs other-treatment), (lognormal vs COX)



Treatment was not a significant factor (p<=0.05) for female 5yrs OS.

Comparison of 3 year DFS and 5 year OS by Model-type

The ACCENT group's FDA premise was that 3 year DFS can be used as a surrogate for 5 year OS (Sargent DJ, et al. JCO 2005; 23:8664-8670.). The FDA approved the use of the DFS surrogate.

Examples of model-specific differences are provided in Figures 1-4, survivor plots by model type, for a representative set of patient characteristics (Males or Females; 60 years old; stage 3 disease; receiving 5FU+LV+/-LEV or other systemic therapy).

Patients	3 year DFS Log-Normal	Cox	Absolute Difference	5 year OS Log-Normal	Cox	Absolute Difference
Males:						
5FU+LV+/-LEV	73.0%	69.3%	3.7%	69.9%	68.3%	1.5%
Other systemic	66.2%	62.4%	3.9%	62.5%	60.6%	1.8%

Females:						
5FU+LV+/-LEV	73.1%	68.7%	4.4%	66.9%	64.9%	2.1%
Other systemic	68.6%	64.0%	4.7%	66.9%[1]	64.9%[1]	2.1%[1]

[1]Treatment was not a significant factor (p<=0.05) for female OS.

Achievements

During the practicum, I managed the real large dataset under Dr. Chapman's indication, and followed every step from the rough data to generate the final results; I understand the procedures of biostatistics analysis, which is very important to create the thoughts of analysis. This is the major gain for me.

I was familiar with the clinic trail dataset, and learnt how to manage and analyze the large database, also I can generate new variables and dataset for statistics analysis. I grasped the essential skill for the large database.

When I dealed with this project, things are not going smoothly at beginning, I got lots of difficult, Dr. Chapman often teach me how to find and solve the problem, after this training, I can solve some problem by myself late on.

I practiced many SAS and R program; also processed MACRO in my project, I learnt lots of valuable computer and programming skills.

I have improved my organization, and learnt how to work as a team.

Discussion

It is very positive experience for me to practice in NCIC CTG, I met lots of greatly teachers, who understand and patience, they always provided me with statistical advices.

In this summer, my major project is ACCENT analysis; I finished the data quality check, residual analysis, and homogeneity test, step-wise multivariate selection, lognormal and COX model fitting, and comparison for this two models, I made some tables and graphs for this project, hopefully, I can contribute some results for publishing. I also analyzed Breast TMA

dataset from Breast Cancer databases, Department of Pathology and Molecular Medicine, Kingston General Hospital.

Meanwhile, I also got a chance worked with Wenbin Li as a team to do the case study for 2010 SSC annual meeting in Quebec city, and posted our result in the meeting.

The practicum was a great opportunity for me to get starting the first step in my future career Biostatistics, and what I learnt here will benefit me in my professional field.

Reference

1. Grothey A. Is there a third-line therapy for metastatic colorectal cancer? Semin Oncol 2006;33:S36–38.
2. Saltz LB, Meropol NJ, Loehrer PJ, Sr., Needle MN, Kopit J, Mayer RJ. Phase II trial of cetuximab in patients with refractory colorectal cancer that expresses the epidermal growth factor receptor. J Clin Oncol 2004;22:1201–8.
3. Yip D, Strickland AH. Immunomodulation therapy in colorectal carcinoma. Cancer Treat Rev. 2000 Jun;26(3):169-90.
4. Pastor F, Kolonias D, Induction of tumour immunity by targeted inhibition of nonsense-mediated mRNA decay. Nature.2010 May 13;465(7295):227-30.
5. Sargent DJ, Patiyil S, et al. ACCENT Group. End points for colon cancer adjuvant trials: observations and recommendations based on individual patient data from 20,898 patients enrolled onto 18 randomized trials from the ACCENT Group. J Clin Oncol. 2007 Oct 10;25(29):4569-74. Epub 2007 Sep 17.

Appendix

Part 1

1. log-normal Step-wise selection for OS, male,

1a. log-normal MACRO selected the variables (stage, age, combined trial, Registration)

OS 5FU+LV+/-LEV vs other treatment (no surgery). Male				
step	**variables**	**loglik**	**df**	**pvalue**
1		-10959.64	2	.
1	age	-10898.37	3	0
1	STAGE	-10745.91	3	0
1	trial_combine	-10906.85	3	0
1	Registration	-10959.48	3	0.99726
2	age	-10898.37	3	.
2	STAGE age	-10691.19	4	0
2	trial_combine age	-10856.5	4	0
2	Registration age	-10898.3	4	0.99999
3	age STAGE	-10691.19	4	.
3	trial_combine age STAGE	-10665.96	5	0
3	Registration age STAGE	-10690.15	5	0.9901
4	**age STAGE trial_combine**	-10665.96	5	.
4	Registration age STAGE trial_combine	-10662.77	6	0.84662

1b. Selected variables fit log-normal and COX model. Variables (stage, age, combined trial)

Fit log-normal (other treatment as Ref)		**Fit COX Model (other treatment as Ref)**	
Parameter	Estimate	Parameter	Estimate
Intercept	11.1520		
Age	-0.0160	Age	0.0156

Stage	-0.7361	Stage	0.68381
5FU+LV+/-LEV	0.3020	5-FU+LV+/-LEV	-0.27303
Scale	1.4880		

2. log-normal Step-wise selection for OS, female,

2a. log-normal MACRO selected the variables (stage, age, combined trial, Registration)

OS 5FU+LV+/-LEV vs other treatment (no surgery). Female				
step	**variables**	**loglik**	**df**	**pvalue**
1		-8778.81	2	.
1	age	-8735.29	3	0
1	STAGE	-8609.04	3	0
1	trial_combine	-8757.66	3	0
1	Registration	-8777.76	3	0.83554
2	age	-8735.29	3	.
2	STAGE age	-8570.52	4	0
2	trial_combine age	-8719.86	4	0.00007
2	Registration age	-8734.44	4	0.97392
3	**age STAGE**	-8570.52	4	.
3	trial_combine age STAGE	-8562.23	5	0.05556
3	Registration age STAGE	-8570.39	5	1

2b. Selected variables fit log-normal and COX model. Variables (stage, age)

Fit log-normal		**Fit COX Model**	
Parameter	Estimate	Parameter	Estimate
Intercept	11.7729		
Age	-0.0169	Age	0.01413
Stage	-0.8476	Stage	0.76720

Scale	1.6110		

3. log-normal Step-wise selection for TTR, male,

3a. log-normal MACRO selected the variables (stage, age, combined trial, Registration)

TTR 5FU+LV+/-LEV vs other treatment (no surgery). Male				
step	**variables**	**loglik**	**df**	**pvalue**
1		-10935.14	2	.
1	STAGE	-10671.61	3	0
1	trial_combine	-10896.67	3	0
1	Registration	-10927.87	3	0.01251
1	age	-10932.67	3	0.42304
2	STAGE	-10671.61	3	.
2	trial_combine STAGE	-10652.37	4	0
2	Registration STAGE	-10669.01	4	0.63606
2	age STAGE	-10670.8	4	0.97802
3	**STAGE trial_combine**	-10652.37	4	.
3	Registration STAGE trial_combine	-10651.49	5	0.99475
3	age STAGE trial_combine	-10652.16	5	0.99998

3b. Selected variables fit log-normal and COX model. Variables (stage, combined trial)

Fit log-normal (other treatment as Ref)		**Fit COX Model (other treatment as Ref)**	
Parameter	Estimate	Parameter	Estimate
Intercept	11.3847		
Stage	-1.1292	Stage	0.85616
5FU+LV+/-LEV	0.3471	5-FU+LV+/-LEV	-0.24972
Scale	1.91120		

4. log-normal Step-wise selection for TTR, female,

4a. log-normal MACRO selected the variables (stage, age, combined trial, Registration)

TTR 5FU+LV+/-LEV vs other treatment (no surgery). Female				
step	**variables**	**loglik**	**df**	**pvalue**
1		-9211.2	2	.
1	STAGE	-9023.86	3	0
1	trial_combine	-9197.78	3	0.00006
1	Registration	-9204.84	3	0.02614
1	age	-9210.76	3	0.97185
2	**STAGE**	-9023.86	3	.
2	trial_combine STAGE	-9017.85	4	0.09945
2	Registration STAGE	-9019.56	4	0.2822
2	age STAGE	-9023.86	4	1

4b. Selected variables fit log-normal and COX model. Variables (stage)

Fit log-normal (other treatment as Ref)		**Fit COX Model (other treatment as Ref)**	
Parameter	Estimate	Parameter	Estimate
Intercept	11.7896		
Stage	-1.1883	Stage	0.83614
Scale	2.0875		

5. log-normal Step-wise selection for DFS, male,

5a. log-normal MACRO selected the variables (stage, age, combined trial, Registration)

DFS 5FU+LV+/-LEV vs other treatment (no surgery). Male				
step	**variables**	**loglik**	**df**	**pvalue**
1		-12591.85	2	.
1	STAGE	-12369.66	3	0

1	trial_combine	-12542.3	3	0
1	age	-12555.4	3	0
1	Registration	-12590.26	3	0.67363
2	STAGE	-12369.66	3	.
2	age STAGE	-12338.63	4	0
2	trial_combine STAGE	-12339.54	4	0
2	Registration STAGE	-12369.66	4	1
3	STAGE age	-12338.63	4	.
3	trial_combine STAGE age	-12314.26	5	0
3	Registration STAGE age	-12338.62	5	1
4	**STAGE age trial_combine**	-12314.26	5	.
4	Registration STAGE age trial_combine	-12313.58	6	0.99976

5b. Selected variables fit log-normal and COX model. Variables (stage, age, combined trial)

Fit log-normal (other treatment as Ref)		**Fit COX Model (other treatment as Ref)**	
Parameter	Estimate	Parameter	Estimate
Intercept	11.2309		
Age	-0.0138	Age	0.01249
Stage	-0.8836	Stage	0.65176
5FU+LV+/-LEV	0.3475	5-FU+LV+/-LEV	-0.25268
Scale	1.7980		

6. log-normal Step-wise selection for DFS, female,

6a. log-normal MACRO selected the variables (stage, age, combined trial, Registration)

DFS 5FU+LV+/-LEV vs other treatment (no surgery). Female				
step	**variables**	**loglik**	**df**	**pvalue**

1		-10213.73	2	.
1	STAGE	-10042.59	3	0
1	trial_combine	-10193.38	3	0
1	age	-10196.85	3	0
1	Registration	-10210.59	3	0.27943
2	STAGE	-10042.59	3	.
2	age STAGE	-10029.43	4	0.00044
2	trial_combine STAGE	-10031.02	4	0.00161
2	Registration STAGE	-10041.05	4	0.87669
3	STAGE age	-10029.43	4	.
3	trial_combine STAGE age	-10020.1	5	0.02825
3	Registration STAGE age	-10028	5	0.9698
4	**STAGE age trial_combine**	-10020.1	5	.
4	Registration STAGE age trial_combine	-10019.51	6	0.99988

6b. Selected variables fit log-normal and COX model. Variables (stage, age, combined trial)

Fit log-normal (other treatment as Ref)		**Fit COX Model (other treatment as Ref)**	
Parameter	Estimate	Parameter	Estimate
Intercept	11.6397		
Age	-0.0109	Age	0.00922
Stage	-1.0040	Stage	0.69326
5FU+LV+/-LEV	0.2639	5-FU+LV+/LEV	-0.17692
Scale	2.0120		

Part 2

3yrs OS	
Covariate Value	**Survival Estimate**

Gender	Age	Stage	Treatment	lognormal	COX	Absolute Difference
male	30	2	5FU+LV+/LEV	0.953702	0.929411	0.024291
male	30	2	other-treatment	0.930417	0.908428	0.021989
male	30	3	5FU+LV+/LEV	0.882419	0.864938	0.017482
male	30	3	other-treatment	0.837494	0.826663	0.010831
male	40	2	5FU+LV+/LEV	0.942295	0.917901	0.024394
male	40	2	other-treatment	0.914871	0.893697	0.021174
male	40	3	5FU+LV+/LEV	0.859849	0.843835	0.016014
male	40	3	other-treatment	0.809671	0.800304	0.009366
male	50	2	5FU+LV+/LEV	0.928785	0.904612	0.024173
male	50	2	other-treatment	0.896858	0.876762	0.020096
male	50	3	5FU+LV+/LEV	0.834503	0.819794	0.014709
male	50	3	other-treatment	0.7791	0.770524	0.008576
male	60	2	5FU+LV+/LEV	0.912971	0.889305	0.023665
male	60	2	other-treatment	0.876225	0.85735	0.018875
male	60	3	5FU+LV+/LEV	0.806367	0.792528	0.013839
male	60	3	other-treatment	0.745896	0.737079	0.008817
male	70	2	5FU+LV+/LEV	0.89467	0.871722	0.022948
male	70	2	other-treatment	0.852864	0.83518	0.017684
male	70	3	5FU+LV+/LEV	0.775493	0.76177	0.013723
male	70	3	other-treatment	0.710246	0.699779	0.010467
female	30	2		0.94478	0.918326	0.026454
female	30	3		0.857709	0.832349	0.025361
female	40	2		0.93206	0.906527	0.025533
female	40	3		0.832773	0.809487	0.023286
female	50	2		0.917189	0.893126	0.024063
female	50	3		0.805182	0.783933	0.021249
female	60	2		0.899992	0.877935	0.022057

female	60	3		0.774987	0.755499	0.019489
female	70	2		0.880323	0.86076	0.019563
female	70	3		0.742305	0.724024	0.018281

Treatment was not a significant factor (p<=0.05) for female 3yrs OS.

3 yrs TTR						
Covariate Value				**Survival Estimate**		
Gender	**Age**	**Stage**	**Treatment**	**lognormal**	**COX**	**Absolute Difference**
male		2	5FU+LV+/LEV	0.902268	0.843355	0.058913
male		2	other-treatment	0.867139	0.80369	0.063449
male		3	5FU+LV+/LEV	0.759206	0.669678	0.089528
male		3	other-treatment	0.699212	0.597897	0.101315
female		2		0.876224	0.836605	0.039619
female		3		0.721421	0.662037	0.059384

Age was not a significant factor (p<=0.05) for male TTR

Age and Treatment was not a significant factor (p<=0.05) for female 3yrs TTR.

3yrs DFS						
Covariate Value				**Survival Estimate**		
Gender	**Age**	**Stage**	**Treatment**	**lognormal**	**COX**	**Absolute Difference**
male	30	2	5FU+LV+/LEV	0.908859	0.877376	0.031484
male	30	2	other-treatment	0.87296	0.845081	0.027879
male	30	3	5FU+LV+/LEV	0.800198	0.778135	0.022063
male	30	3	other-treatment	0.741849	0.724142	0.017708
male	40	2	5FU+LV+/LEV	0.895625	0.862161	0.033464
male	40	2	other-treatment	0.856277	0.826273	0.030004
male	40	3	5FU+LV+/LEV	0.778036	0.752467	0.025568
male	40	3	other-treatment	0.716443	0.693553	0.022889

male	50	2	5FU+LV+/LEV	0.881051	0.845232	0.03582
male	50	2	other-treatment	0.838176	0.805456	0.03272
male	50	3	5FU+LV+/LEV	0.754533	0.724389	0.030144
male	50	3	other-treatment	0.689895	0.660435	0.029461
male	60	2	5FU+LV+/LEV	0.865096	0.82644	0.038657
male	60	2	other-treatment	0.818651	0.782489	0.036163
male	60	3	5FU+LV+/LEV	0.729755	0.692994	0.036781
male	60	3	other-treatment	0.66232	0.6236542	0.0386658
male	70	2	5FU+LV+/LEV	0.847732	0.80564	0.042092
male	70	2	other-treatment	0.797714	0.757241	0.040473
male	70	3	5FU+LV+/LEV	0.703786	0.660725	0.043061
male	70	3	other-treatment	0.633843	0.586713	0.047131
female	30	2	5FU+LV+/LEV	0.899206	0.867056	0.03215
female	30	2	other-treatment	0.874077	0.843783	0.030294
female	30	3	5FU+LV+/LEV	0.781726	0.751406	0.03032
female	30	3	other-treatment	0.741142	0.711541	0.029602
female	40	2	5FU+LV+/LEV	0.88931	0.855379	0.033931
female	40	2	other-treatment	0.862518	0.830269	0.032249
female	40	3	5FU+LV+/LEV	0.765424	0.731268	0.034156
female	40	3	other-treatment	0.723308	0.688892	0.034416
female	50	2	5FU+LV+/LEV	0.878736	0.842772	0.035964
female	50	2	other-treatment	0.850255	0.815719	0.034536
female	50	3	5FU+LV+/LEV	0.748471	0.709834	0.038636
female	50	3	other-treatment	0.704892	0.664917	0.039974
female	60	2	5FU+LV+/LEV	0.867471	0.829181	0.038291
female	60	2	other-treatment	0.837283	0.800079	0.037204
female	60	3	5FU+LV+/LEV	0.730891	0.6870591	0.0438319
female	60	3	other-treatment	0.68593	0.638926	0.047004

female	70	2	5FU+LV+/LEV	0.855505	0.814548	0.040957
female	70	2	other-treatment	0.823602	0.783295	0.040306
female	70	3	5FU+LV+/LEV	0.712715	0.663006	0.04971
female	70	3	other-treatment	0.666465	0.61302	0.053445

5yrs OS						
Covariate Value				**Survival Estimate**		
Gender	**Age**	**Stage**	**Treatment**	**lognormal**	**COX**	**Absolute Difference**
male	30	2	5FU+LV+/LEV	0.909643	0.887088	0.022555
male	30	2	other-treatment	0.871939	0.854547	0.017392
male	30	3	5FU+LV+/LEV	0.800629	0.788617	0.012013
male	30	3	other-treatment	0.739211	0.73231	0.0069
male	40	2	5FU+LV+/LEV	0.890845	0.869179	0.021666
male	40	2	other-treatment	0.848043	0.831985	0.016058
male	40	3	5FU+LV+/LEV	0.769239	0.757372	0.011867
male	40	3	other-treatment	0.703117	0.694484	0.008633
male	50	2	5FU+LV+/LEV	0.869389	0.84868	0.020709
male	50	2	other-treatment	0.821358	0.806338	0.01502
male	50	3	5FU+LV+/LEV	0.735266	0.722378	0.012888
male	50	3	other-treatment	0.664895	0.652692	0.012203
male	60	2	5FU+LV+/LEV	0.845181	0.825304	0.019877
male	60	2	other-treatment	0.791898	0.777326	0.014572
male	60	3	5FU+LV+/LEV	0.698918	0.6834252	0.0154928
male	60	3	other-treatment	0.624884	0.6064485	0.0184355
male	70	2	5FU+LV+/LEV	0.818181	0.798765	0.019415
male	70	2	other-treatment	0.75975	0.744699	0.015051
male	70	3	5FU+LV+/LEV	0.660476	0.640599	0.019877
male	70	3	other-treatment	0.583481	0.557513	0.025968

female	30	2		0.899575	0.876484	0.023091
female	30	3		0.774275	0.752811	0.021463
female	40	2		0.879848	0.859121	0.020727
female	40	3		0.741538	0.721059	0.020479
female	50	2		0.857538	0.839548	0.01799
female	50	3		0.706501	0.686143	0.020357
female	60	2		0.832582	0.817557	0.015025
female	60	3		0.669411	0.6487889	0.0206221
female	70	2		0.804972	0.792941	0.012031
female	70	3		0.630577	0.606725	0.023852

Treatment was not a significant factor (p<=0.05) for female 5yrs OS.

5yrs TTR						
Covariate Value				**Survival Estimate**		
Gender	**Age**	**Stage**	**Treatment**	**lognormal**	**COX**	**Absolute Difference**
male		2	5FU+LV+/LEV	0.847861	0.826964	0.020897
male		2	other-treatment	0.801137	0.783708	0.017428
male		3	5FU+LV+/LEV	0.668752	0.639448	0.029304
male		3	other-treatment	0.600583	0.5635	0.037083
female		2		0.819012	0.803988	0.015025
female		3		0.633961	0.603886	0.030075

Age was not a significant factor (p<=0.05) for male 5yrs TTR

Age and Treatment was not a significant factor (p<=0.05) for female 5yrs TTR.

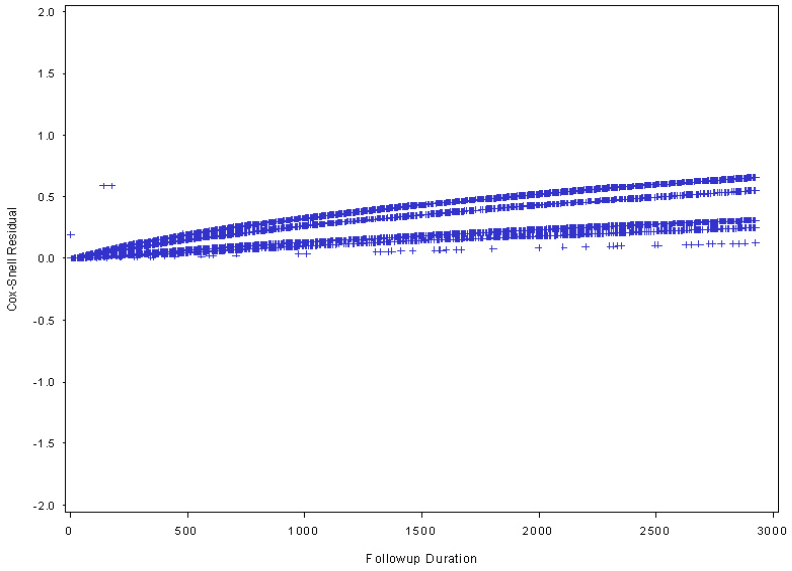
5yrs DFS						
Covariate Value				**Survival Estimate**		
Gender	**Age**	**Stage**	**Treatment**	**lognormal**	**COX**	**Absolute Difference**
male	30	2	5FU+LV+/LEV	0.853062	0.840682	0.012379

male	30	2	other-treatment	0.804108	0.799882	0.004226
male	30	3	5FU+LV+/LEV	0.711653	0.716932	-0.00528
male	30	3	other-treatment	0.642426	0.651699	-0.00927
male	40	2	5FU+LV+/LEV	0.834699	0.821399	0.0133
male	40	2	other-treatment	0.782197	0.776353	0.005844
male	40	3	5FU+LV+/LEV	0.684908	0.68573	-0.00082
male	40	3	other-treatment	0.613402	0.615435	-0.00203
male	50	2	5FU+LV+/LEV	0.814914	0.800071	0.014843
male	50	2	other-treatment	0.758935	0.750513	0.008422
male	50	3	5FU+LV+/LEV	0.657157	0.651995	0.005162
male	50	3	other-treatment	0.58373	0.576757	0.006973
male	60	2	5FU+LV+/LEV	0.793721	0.776561	0.01716
male	60	2	other-treatment	0.734385	0.722257	0.012128
male	60	3	5FU+LV+/LEV	0.628531	0.615752	0.012779
male	60	3	other-treatment	0.553573	0.535839	0.017735
male	70	2	5FU+LV+/LEV	0.771152	0.750741	0.020411
male	70	2	other-treatment	0.708627	0.691507	0.017119
male	70	3	5FU+LV+/LEV	0.599175	0.577093	0.022082
male	70	3	other-treatment	0.523103	0.492949	0.030154
female	30	2	5FU+LV+/LEV	0.846882	0.835735	0.011147
female	30	2	other-treatment	0.8138	0.807615	0.006185
female	30	3	5FU+LV+/LEV	0.699911	0.698008	0.001903
female	30	3	other-treatment	0.652834	0.651747	0.001086
female	40	2	5FU+LV+/LEV	0.833721	0.821601	0.01212
female	40	2	other-treatment	0.798932	0.791378	0.007554
female	40	3	5FU+LV+/LEV	0.680812	0.674558	0.006254
female	40	3	other-treatment	0.632623	0.62576	0.006863
female	50	2	5FU+LV+/LEV	0.819852	0.806398	0.013454

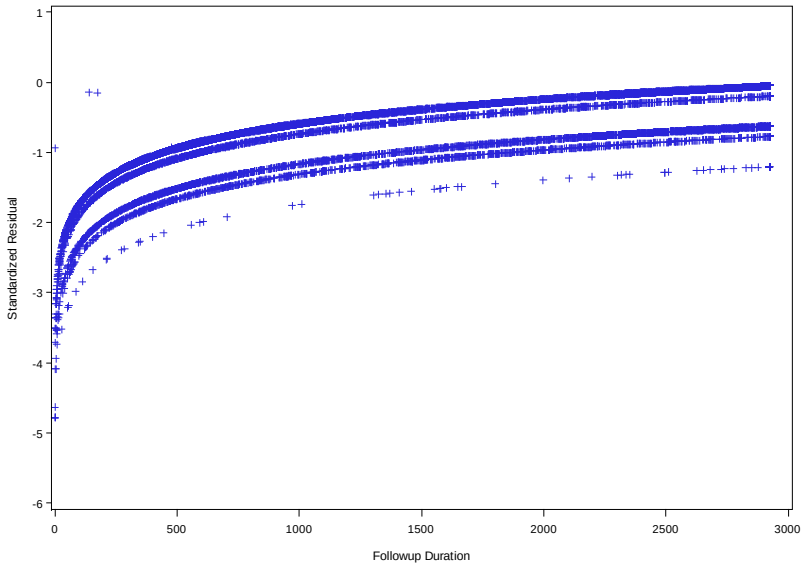
female	50	2	other-treatment	0.783374	0.773972	0.009401
female	50	3	5FU+LV+/LEV	0.66122	0.649782	0.011438
female	50	3	other-treatment	0.612037	0.598489	0.013548
female	60	2	5FU+LV+/LEV	0.805278	0.790073	0.015205
female	60	2	other-treatment	0.767141	0.755351	0.01179
female	60	3	5FU+LV+/LEV	0.641182	0.623694	0.017489
female	60	3	other-treatment	0.591132	0.569987	0.021144
female	70	2	5FU+LV+/LEV	0.790009	0.772575	0.017434
female	70	2	other-treatment	0.750254	0.735473	0.01478
female	70	3	5FU+LV+/LEV	0.620748	0.596325	0.024422
female	70	3	other-treatment	0.569964	0.540332	0.029632

Part 3

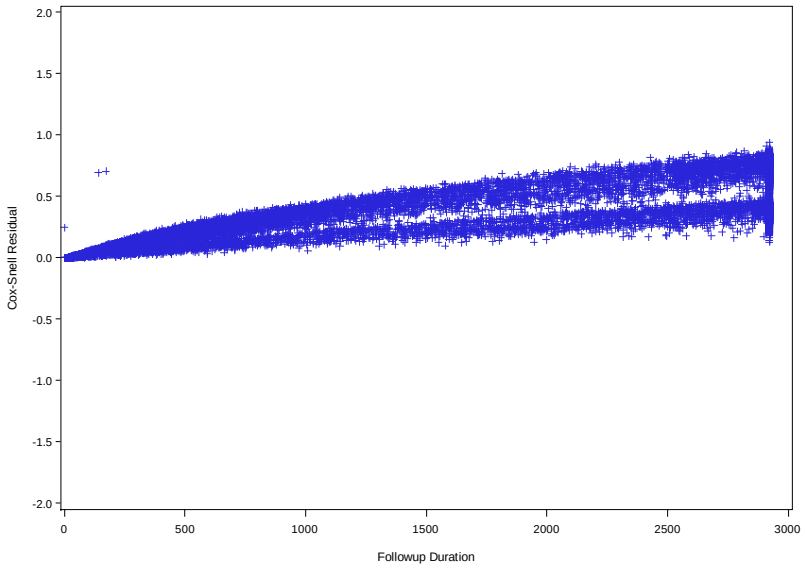
COX-Snell residual vs Progression time (lognormal) TTR



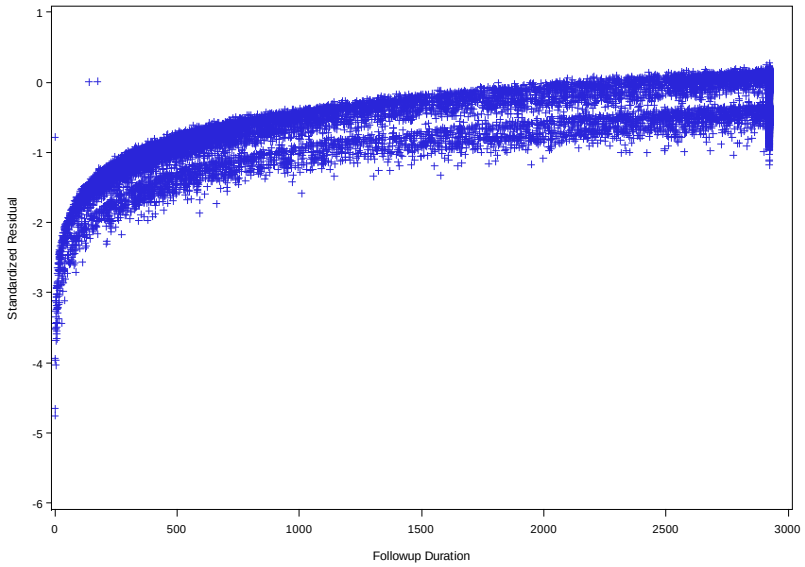
Standardized residual vs Progression time (lognormal) TTR



COX-Snell residual vs Disease-Free-Survival time (lognormal) DFS



Standardized residual vs Disease-Free-Survival time (lognormal) DFS



Part 4

SAS and R code.

MACRO code for lognormal multivariate selection

```
data accent_8y;
set accent;
Registration=YEAR(Registrationdate);
      if Pduration>2922 then do;
       Pduration=2922;
       ProgressionStatus=0;
         end;
run;
data accent_8yrs;
set accent_8y;
if fduration>2922 then do;
      fduration=2922;
        FollowupStatus=0;
        end;
run;

data accent_surgeryonly;
set accent_8yrs;
if drug='Surgery Only';
run;
data accent_surgeryonly_DFS;
set accent_surgeryonly;
      if ProgressionStatus=1 then do;
```

```
       DFStime=Pduration;
       DFSstatus=ProgressionStatus;
         end;
         else if ProgressionStatus=0 then do;
              if FollowupStatus=1 then do;
                 DFStime=fduration;
                 DFSstatus=FollowupStatus;
                         end;
                      else if ProgressionStatus=0 then do;
                 if FollowupStatus=0 then do;
                    DFStime=fduration;
                    DFSstatus=FollowupStatus;
                    end;
               end;
            end;
run;

data accent_surgeryonly_m;
set accent_surgeryonly;
where gender=1;
run;
data accent_surgeryonly_m_DFS;
set accent_surgeryonly_m;
      if ProgressionStatus=1 then do;
       DFStime=Pduration;
       DFSstatus=ProgressionStatus;
         end;
         else if ProgressionStatus=0 then do;
              if FollowupStatus=1 then do;
                 DFStime=fduration;
                 DFSstatus=FollowupStatus;
                         end;
                      else if ProgressionStatus=0 then do;
                 if FollowupStatus=0 then do;
                    DFStime=fduration;
                    DFSstatus=FollowupStatus;
                    end;
               end;
            end;
run;

data accent_surgeryonly_f;
set accent_surgeryonly;
where gender=0;
run;
data accent_surgeryonly_f_DFS;
set accent_surgeryonly_f;
      if ProgressionStatus=1 then do;
       DFStime=Pduration;
       DFSstatus=ProgressionStatus;
         end;
         else if ProgressionStatus=0 then do;
              if FollowupStatus=1 then do;
                 DFStime=fduration;
                 DFSstatus=FollowupStatus;
                         end;
```

```
                    else if ProgressionStatus=0 then do;
                 if FollowupStatus=0 then do;
                    DFStime=fduration;
                    DFSstatus=FollowupStatus;
                    end;
               end;
            end;
run;

data accent_TrialCombine;
set accent_8yrs;
length trial_combine $35;
if drug='5FU+LV' or drug='5FU+LV+LEV' then trial_combine='5FU+LV+/-LEV';
else trial_combine='other_treatment';
run;
data accent_TrialCombine_DFS;
set accent_TrialCombine;
      if ProgressionStatus=1 then do;
       DFStime=Pduration;
       DFSstatus=ProgressionStatus;
         end;
         else if ProgressionStatus=0 then do;
              if FollowupStatus=1 then do;
                 DFStime=fduration;
                 DFSstatus=FollowupStatus;
                         end;
                    else if ProgressionStatus=0 then do;
                 if FollowupStatus=0 then do;
                    DFStime=fduration;
                    DFSstatus=FollowupStatus;
                    end;
               end;
            end;
run;

data accent_TrialCombine_m;
set accent_TrialCombine;
if gender=1;
run;
data accent_TrialCombine_m_DFS;
set accent_TrialCombine_m;
      if ProgressionStatus=1 then do;
       DFStime=Pduration;
       DFSstatus=ProgressionStatus;
         end;
         else if ProgressionStatus=0 then do;
              if FollowupStatus=1 then do;
                 DFStime=fduration;
                 DFSstatus=FollowupStatus;
                         end;
                    else if ProgressionStatus=0 then do;
                 if FollowupStatus=0 then do;
                    DFStime=fduration;
                    DFSstatus=FollowupStatus;
                    end;
               end;
```

```
            end;
run;

data accent_TrialCombine_f;
set accent_TrialCombine;
if gender=0;
run;

data accent_TrialCombine_f_DFS;
set accent_TrialCombine_f;
      if ProgressionStatus=1 then do;
       DFStime=Pduration;
       DFSstatus=ProgressionStatus;
         end;
         else if ProgressionStatus=0 then do;
              if FollowupStatus=1 then do;
                 DFStime=fduration;
                 DFSstatus=FollowupStatus;
                         end;
                     else if ProgressionStatus=0 then do;
                 if FollowupStatus=0 then do;
                    DFStime=fduration;
                    DFSstatus=FollowupStatus;
                    end;
               end;
            end;
run;

data accent_TrialCombine_nosurgery;
set accent_8yrs;
length trial_combine $35;
if drug='5FU+LV' or drug='5FU+LV+LEV' then trial_combine='5FU+LV+/-LEV';
else if drug='Surgery Only' then delete;
else trial_combine='other_treatment';
run;
data accent_TrialCombine_nosgy_DFS;
set accent_TrialCombine_nosurgery;
      if ProgressionStatus=1 then do;
       DFStime=Pduration;
       DFSstatus=ProgressionStatus;
         end;
         else if ProgressionStatus=0 then do;
              if FollowupStatus=1 then do;
                 DFStime=fduration;
                 DFSstatus=FollowupStatus;
                         end;
                     else if ProgressionStatus=0 then do;
                 if FollowupStatus=0 then do;
                    DFStime=fduration;
                    DFSstatus=FollowupStatus;
                    end;
               end;
            end;
run;

data accent_TrialCombine_nosurgery_m;
```

```
set accent_TrialCombine_nosurgery;
if gender=1;
run;
data accent_TrialCombine_nosgy_m_DFS;
set accent_TrialCombine_nosurgery_m;
      if ProgressionStatus=1 then do;
       DFStime=Pduration;
       DFSstatus=ProgressionStatus;
         end;
         else if ProgressionStatus=0 then do;
              if FollowupStatus=1 then do;
                 DFStime=fduration;
                 DFSstatus=FollowupStatus;
                         end;
                     else if ProgressionStatus=0 then do;
                 if FollowupStatus=0 then do;
                    DFStime=fduration;
                    DFSstatus=FollowupStatus;
                    end;
               end;
            end;
run;

data accent_TrialCombine_nosurgery_f;
set accent_TrialCombine_nosurgery;
if gender=0;
run;
data accent_TrialCombine_nosgy_f_DFS;
set accent_TrialCombine_nosurgery_f;
      if ProgressionStatus=1 then do;
       DFStime=Pduration;
       DFSstatus=ProgressionStatus;
         end;
         else if ProgressionStatus=0 then do;
              if FollowupStatus=1 then do;
                 DFStime=fduration;
                 DFSstatus=FollowupStatus;
                         end;
                     else if ProgressionStatus=0 then do;
                 if FollowupStatus=0 then do;
                    DFStime=fduration;
                    DFSstatus=FollowupStatus;
                    end;
               end;
            end;
run;

%let dataset = accent_surgeryonly_f;
%let time    = pduration;
%let cen     = progressionstatus;
%let disttype= Lnormal;

%macro fitModel(result, var, cvar);
      ods listing close;
      proc lifereg data=&dataset;
            class &cvar;
```

```
            model &time*&cen(0)= &var &cvar / distribution = &disttype;
      ods output ModelInfo=minfo ParameterEstimates =pest;
      run;

      data pest;
            set pest;
            if parameter eq "Weibull Shape" or parameter eq "Weibull Scale" then
delete;
      run;

      proc means data=pest sum;
            var DF;
            ods output Summary=sumdf ;
      run;
      ods listing;

      data logval;
         set minfo;
         where label1 = 'Log Likelihood';
         loglik = nValue1;
         keep loglik;
      run;

      options mergenoby = nowarn;
      data &result;
            merge logval sumdf;
            df = df_sum;
            drop df_sum;
      run;
      options mergenoby = error;

      proc datasets nolist;
      delete pest sumdf logval minfo;
      run;

%mend fitModel;

%macro updateResult(old, new, newv, curv);
      data &new;
            set &new;
            inact = &newv;
            act = &curv;
      run;

      data &old;
         set &old &new;
      run;
%mend updateResult;

%macro forwardFitting(result, actv, inactv, actcv, inactcv);

  data &result;
      length act $100;
      length inact $20;
  run;
```

```
  %let k=1;
  %let var = %scan(&inactv, &k, " " );
  %do %while("&var" NE "");

      %fitModel(result=temp, var = &actv &var, cvar = &actcv);
      %updateResult(old=&result, new=temp, newv="&var", curv="&actv &actcv");

    %let k = %eval(&k + 1);
    %let var = %scan(&inactv, &k, " " );
  %end;

  %let k=1;
  %let var = %scan(&inactcv, &k, " ");
  %do %while("&var" NE "");

      %fitModel(result=temp, var = &actv, cvar = &actcv &var);
      %updateResult(old=&result, new=temp, newv="&var", curv="&actv &actcv");

    %let k = %eval(&k + 1);
    %let var = %scan(&inactcv, &k, " ");
  %end;

   data &result;
         set &result;
         if Loglik = . then delete;
         run;

   proc datasets nolist;
   delete temp;

%mend forwardFitting;
%macro forwardSelect(out, actv, actcv,inactv, inactcv, maxiter=30, cutoff=0.05);

  %let iter   = 1;
  %let curloglik = 0;
  %let curdf = 0;

  data &out;
      length act $100;
      length inact $20;
  run;

  %let numvar = %eval(%sysfunc(countw("&inactv", " ")) +
%sysfunc(countw("&inactcv", " "))) ;
  data _null_;
      if &numvar < &maxiter then
      call symput("maxiter",  %eval(&numvar+1) );
   run;

  %do %until( &iter > &maxiter );
      %fitModel(result=temp, var = &actv, cvar = &actcv );

      data temp;
            step = %eval(&iter);
            set temp;
            call symput("curloglik", trim(left(loglik)) );
```

```
            call symput(    "curdf", trim(left(df)) );
      run;

      %updateResult(old=&out, new=temp, newv="", curv="&actv &actcv");

      %forwardFitting(result=curstep,actv = &actv, inactv = &inactv, actcv=
&actcv, inactcv = &inactcv);
      data curstep;
            set curstep;
            chi_statistic = 2*(loglik-1*(&curloglik));
            pvalue = 1-probchi(chi_statistic, df + &curdf);
            drop chi_statistic;
            step = %eval(&iter);
      run;

      proc sort data=curstep;
            by pvalue;
      run;
      data &out;
            set &out curstep;
      run;

      data _null_;
      set curstep(firstobs=1 obs=1);

      if pvalue lt &cutoff then do;
            if index("&inactcv", trim(inact)) ne 0 then do;
                  call symput('actcv' , trim(left("&actcv" || " " || inact)) );
                  call symput("inactcv", trim(left(tranwrd("&inactcv",
trim(inact), " "))) );
                  end;
            else if index("&inactv", trim(inact)) ne 0 then do;
                  call symput('actv', trim(left("&actv" || " " || inact)) );
                  call symput('inactv', trim(left(tranwrd("&inactv", trim(inact),
" "))) );
                  end;
            else  do;
                  call execute('
                        %PUT "Variable is neither a class or contiuous";
                        ');
                  end;
            end;
      else do;
            call symput("iter", %eval( &maxiter) );
            end;

      run;
      proc datasets nolist;
      delete temp curstep;
      run;

      %let iter = %eval(&iter +1);
  %end;
      data &out;
            set &out;
            if df = . then delete;
```

```
            variables = trim(COMPBL( trim(inact) || " " || trim(act)));
            drop act inact;
      run;

      proc sql;
      create table temp as
      select step, variables, loglik, df, pvalue
      from &out;
      quit;

      data &out;
            set temp;
      run;

      proc datasets nolist;
      delete temp;
      run;
%mend forwardSelect;
%forwardSelect(out=result, actv=,actcv=,inactv=age STAGE Registration, inactcv=
trial, cutoff=0.05, maxiter=30);
```

R code for survivor plot

```
**************5yrs OS Male*********************************
accent<-read.csv('fu_m.csv', header=T)
fit.cox<-coxph(Surv(fduration,FollowupStatus)~Age+Stage+as.factor(trial_combine), data=accent)
sf.cox0<-survfit(fit.cox,newdata=data.frame(Age=60,Stage=3,  Gender=1, trial_combine=0))
sf.cox1<-survfit(fit.cox,newdata=data.frame(Age=60, Stage=3, Gender=1, trial_combine=1))
accent1<-read.csv('accent_subgroupos_m.csv', header=T)
fitall<-
survreg(Surv(fduration,FollowupStatus)~Age+Stage+as.factor(trial_combine),data=accent,dist='lognormal')
fitall
sig<-fitall$scale
sig
t<-1:2922
st.lognm1<-1-pnorm((log(t)-(11.45399880 )-(-0.01601429*60 )-( -0.73605471*3))/sig)
st.lognm2<-1-pnorm((log(t)-(11.45399880 )-(-0.01601429*60 )-( -0.73605471*3) -(-0.30202673))/sig)
postscript("fudfscox&lognormal0&1.ps",width=6,height=6,horizontal=F)
pdf(file="-5yrs OS Male-.pdf", width=10,height=7.5)
plot(sf.cox0, conf.int=F, main="Overall Survival for 60 Year Old Males with Stage 3 Disease", xlim=c(0,2922),
ylim=c(0,1), xaxs='i', yaxs='i',axes=F, xlab="Time (years)", ylab="Overall Survival",lwd=2)
axis(1,at=0:8*365.25, paste(0:8,sep=''),  cex.axis=1.0,lwd=2)
axis(2,at=0:10*0.10,cex.axis=1.0,lwd=2)
abline(v=5*365.25,col=4,lwd=2)
lines(sf.cox1, lty=2, col=2,lwd=2)
lines(st.lognm1, lty=3,col=3,lwd=2)
lines(st.lognm2, lty=4,col=4,lwd=2)
legend(100, 0.25, c("COX 5FU+LV+/-LEV", "COX Other_Treatment", "Log Normal 5FU+LV+/-LEV", "Log
Normal Other_Treatment"),lty=1:4, col=1:4, lwd=2)
dev.off()


**************5yrs OS Female*******************************
accent<-read.csv('fu_f.csv', header=T)
```

```
fit.cox<-coxph(Surv(fduration,FollowupStatus)~Age+Stage, data=accent)
sf.cox0<-survfit(fit.cox,newdata=data.frame(Age=60,Stage=3,Gender=0, trial_combine=0))
sf.cox1<-survfit(fit.cox,newdata=data.frame(Age=60, Stage=3,Gender=0, trial_combine=1))
fitall<-survreg(Surv(fduration,FollowupStatus)~Age+Stage,data=accent,dist='lognormal')
fitall
sig<-fitall$scale
sig
t<-1:2922
st.lognm1<-1-pnorm((log(t)-11.77293022 -(-0.01686959*60)-(-0.84755013*3))/sig)
st.lognm2<-1-pnorm((log(t)-11.77293022 -(-0.01686959*60)-(-0.84755013*3))/sig)
postscript("fudfscox&lognormal0&1.ps",width=6,height=6,horizontal=F)
pdf(file="-5yrs OS Female-.pdf", width=10,height=7.5)
plot(sf.cox0, conf.int=F, main="Overall Survival for 60 Year Old Females with Stage 3 Disease", xlim=c(0,2922),
ylim=c(0,1), xaxs='i', yaxs='i',axes=F, xlab="Time (years)", ylab="Overall Survival",lwd=2)
axis(1,at=0:8*365.25, paste(0:8,sep=''),  cex.axis=1.0,lwd=2)
axis(2,at=0:10*0.10,cex.axis=1.0,lwd=2)
abline(v=5*365.25,col=4,lwd=2)
lines(sf.cox1, lty=2, col=2,lwd=2)
lines(st.lognm1, lty=3,col=3,lwd=2)
lines(st.lognm2, lty=4,col=4,lwd=2)
legend(100, 0.25, c("COX 5FU+LV+/-LEV", "COX Other_Treatment", "Log Normal 5FU+LV+/-LEV", "Log
Normal Other_Treatment"),lty=1:4, col=1:4, lwd=2)
dev.off()

**********3yrs DFS Male*******************
accent<-read.csv('fu_male_dfs.csv', header=T)
fit.cox<-coxph(Surv(DFStime,DFSstatus)~Age+Stage+as.factor(trial_combine), data=accent)
sf.cox0<-survfit(fit.cox,newdata=data.frame(Age=60,Stage=3, Gender=1, trial_combine=0))
sf.cox1<-survfit(fit.cox,newdata=data.frame(Age=60, Stage=3, Gender=1, trial_combine=1))
fitall<-survreg(Surv(DFStime,DFSstatus)~Age+Stage+as.factor(trial_combine),data=accent,dist='lognormal')
fitall
sig<-fitall$scale
sig
t<-1:2922
st.lognm1<-1-pnorm((log(t)-11.57841063 -(-0.01384099 *60)-(-0.88357393 *3))/sig)
st.lognm2<-1-pnorm((log(t)-11.57841063 -(-0.01384099 *60)-(-0.88357393 *3)-(-0.34745100 ))/sig)
postscript("fudfscox&lognormal0&1.ps",width=6,height=6,horizontal=F)
pdf(file="-3yrs DFS Male-.pdf", width=10,height=7.5)
plot(sf.cox0, conf.int=F, main="Disease Free Survival for 60 Year Old Males with Stage 3 Disease",
xlim=c(0,2922), ylim=c(0,1), xaxs='i', yaxs='i',axes=F, xlab="Time (years)", ylab="Disease Free Survival",lwd=2)
axis(1,at=0:8*365.25, paste(0:8,sep=''),  cex.axis=1.0,lwd=2)
axis(2,at=0:10*0.10,cex.axis=1.0,lwd=2)
abline(v=3*365.25,col=4,lwd=2)
lines(sf.cox1, lty=2, col=2,lwd=2)
lines(st.lognm1, lty=3,col=3,lwd=2)
lines(st.lognm2, lty=4,col=4,lwd=2)
legend(100, 0.25, c("COX 5FU+LV+/-LEV", "COX Other_Treatment", "Log Normal 5FU+LV+/-LEV", "Log
Normal Other_Treatment"),lty=1:4, col=1:4, lwd=2)
dev.off()

**********3yrs DFS Female*******************
accent<-read.csv('fu_female_dfs.csv', header=T)
fit.cox<-coxph(Surv(DFStime,DFSstatus)~Age+Stage+as.factor(trial_combine), data=accent)
sf.cox0<-survfit(fit.cox,newdata=data.frame(Age=60,Stage=3, Gender=0, trial_combine=0))
sf.cox1<-survfit(fit.cox,newdata=data.frame(Age=60, Stage=3, Gender=0, trial_combine=1))
```

```
fitall<-survreg(Surv(DFStime,DFSstatus)~Age+Stage+as.factor(trial_combine),data=accent,dist='lognormal')
fitall
sig<-fitall$scale
sig
t<-1:2922
st.lognm1<-1-pnorm((log(t)-11.90355621  -( -0.01090634 *60)-( -1.00404195 *3))/sig)
st.lognm2<-1-pnorm((log(t)-11.90355621  -( -0.01090634 *60)-( -1.00404195 *3)-(-0.26386722 ))/sig)
postscript("fudfscox&lognormal0&1.ps",width=6,height=6,horizontal=F)
pdf(file="-3yrs DFS Female-.pdf", width=10,height=7.5)
plot(sf.cox0, conf.int=F, main="Disease Free Survival for 60 Year Old Females with Stage 3 Disease",
xlim=c(0,2922), ylim=c(0,1), xaxs='i', yaxs='i',axes=F, xlab="Time (years)", ylab="Disease Free Survival",lwd=2)
axis(1,at=0:8*365.25, paste(0:8,sep=''),  cex.axis=1.0,lwd=2)
axis(2,at=0:10*0.10,cex.axis=1.0,lwd=2)
abline(v=3*365.25,col=4,lwd=2)
lines(sf.cox1, lty=2, col=2,lwd=2)
lines(st.lognm1, lty=3,col=3,lwd=2)
lines(st.lognm2, lty=4,col=4,lwd=2)
legend(100, 0.25, c("COX 5FU+LV+/-LEV", "COX Other_Treatment", "Log Normal 5FU+LV+/-LEV", "Log
Normal Other_Treatment"),lty=1:4, col=1:4, lwd=2)
dev.off()
```