Survival Analysis of the Veterans' Administration Lung Cancer Study Using Proportional Hazards Models

Kenneth Broadhead

Department of Statistics, University of California, Davis

Abstract

In this paper, we analyze the well-known Veteran's Administration Lung cancer study data set using proportional hazards models. There are two primary goals in our analysis. First, we aim evaluate the efficacy of the new treatment for lung cancer in comparison to the standard treatment. Secondly, we evaluate the efficacy of using proportional hazards models to analyze this data. Our results indicate that the treatment is not more effective than the standard treatment for lung cancer. Additionally, the use of proportional hazards models in analyzing this data set is problematic. We suggest that additional methods for survival data should be applied to investigate this data to supplement the conclusions reached in this study.

Keywords: Cox PH model, Stratified Cox Model, Lung Cancer

Introduction

In the late 1970's-early 1980's, the US Veteran's Administration undertook the study of an experimental chemotherapy treatment for lung cancer. In total, 137 patients were randomized to one of two treatments: either the standard treatment at the time, or a new chemotherapy treatment. Various covariates were measured prior to randomization, and included: indicators for prior treatment received, age of the patient in years, time (in months) between diagnosis and the start of the study, Karnofsky performance score (an overall patient condition), and the histological type of the patient's tumor. Time to death, since treatment began, was recorded. The study has a very low censoring rate, with only 9 leaving the study before death. The data is itself very clean, with no missing values or apparent errors. The data is freely available as part of the R 'survival' package, and was first published in by Kalbfleisch and Prentice in *The Statistical Analysis of Failure Time Data (1980)*.

The available covariates are all potentially relevant in discerning the effectiveness of the new treatment. Prior treatment received, and current clinical indicators (cell type and Karnofsky score) are important variables to adjust for when assessing the potential impact and effeteness of a new medical treatment. Therefore, to adjust for these covariates in assessing treatment efficacy, we include all available covariates in our models unless it is necessary to discard them. Additionally, this report does not attempt to assess covariate importance; as a result, no classical or automatic model selection is performed.

Methods

Models and Notation

In proportional hazards (PH) models, the effect of covariates on the hazard function are modeled. Importantly, the effect of covariates on the hazard function is assumed to be multiplicative; this is the PH assumption. The general model is: $h(t|\boldsymbol{Z}) = h_0(t)f(\boldsymbol{Z})$, where $h(t|\boldsymbol{Z})$ is the hazard function given a set of covariates, $h_0(t)$ is an (unspecified) baseline hazard function, $\boldsymbol{Z}$ is a vector of covariates, and f is some known function describing the strength of covariates multiplicative effect on bassline hazard. In the Cox Proportional Hazard model, the function f is taken to be the exponential function, leading to the model $h(t|\boldsymbol{Z}) = h_0(t)\exp\{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\}$. The parameters $\boldsymbol{\beta}$ in the Cox PH model are usually estimated by the method of maximum (partial) likelihood. There are several ways to construct the partial likelihood function for a cox model; here we provide a simple derivation of the (partial) likelihood in the absence of ties using individual likelihoods.

Let there be D distinct observed death times, $t_1, \ldots t_D$. Denote by $R(t_i)$ the risk set at time t_i: the set of individuals who have yet to die just prior to time t_i. Let $\boldsymbol{Z}_j$ be the set of covariates for the j^{th} individual, and let $\boldsymbol{Z}_{(i)}$ be the covariate set for the individual with death time t_i.

We have for a given individual at time t_i , the probability of death at time t_i, given there is only one death at time t_i is equal to:

$$L_j = \frac{P(\text{the individual dies at } t_i | \text{one death at } t_i)}{P(\text{one death at } t_i | \text{ survival to } t_i)}$$

$$= \frac{h\left(t_i \middle| \boldsymbol{Z}_{(i)}\right)}{\sum_{j \in R(t_i)} h\left(t_j \middle| \boldsymbol{Z}_j\right)}$$

$$= \frac{h_0(t_i)\exp\left\{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{(\mathbf{i})}\right\}}{\sum_{j \in R(t_i)} h_0\left(t_j\right)\exp\left\{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{\mathbf{j}}\right\}}$$

$$= \frac{\exp\left\{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{(\mathbf{i})}\right\}}{\sum_{j \in R(t_i)} \exp\left\{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{\mathbf{j}}\right\}}$$

The partial likelihood is then given by the product of each individual likelihood above for all observed death times D:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{D} L_i = \prod_{i=1}^{D} \frac{\exp\{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{(\mathbf{i})}\}}{\sum_{j \in R(t_i)} \exp\{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{\mathbf{j}}\}}$$

An extension of this basic model, the stratified Cox PH model, is also used in this study. The stratified Cox model is often employed to correct for variables that violate the PH assumption. The stratified Cox model assumes the covariates effects are same across each stratum, but allows for a different baseline within each. The model is specified as:

$$h(t|\boldsymbol{Z}) = h_{og}(t) \exp\{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}\}$$

Here, g runs from $1, \ldots, s$, where s is the total number of strata. The log partial likelihood is given by

$$LL(\boldsymbol{\beta}) = \sum_{j=1}^{s} LL_j(\boldsymbol{\beta}).$$

Here, $LL_j(\boldsymbol{\beta})$ is the log partial likelihood (under the unstratified Cox Model) using only the data from the jth stratum. A Likelihood Ratio Test can be employed to test the assumption of equal covariate effects across stratum, as the stratified Cox model is nested within model assuming unequal covariate effects between stratum.

For all analysis, modeling, and estimation in this report, we use SAS. Code is provided in the appendices.

Statistical Analysis

In this study of the Veterans' Administration lung cancer data, Cox PH models are used to assess the efficacy of the experimental chemotherapy treatment. After initial investigation of

the data, it is discovered that the PH assumption is violated for a number of covariates, namely age, Karnofsky score, and cell type. To remedy this violation, and with sample size considerations in mind, age, cell type, and Karnofsky are dichotomized, and a stratified Cox PH model is employed with 8 strata total. However, additional problems arise. The sample size within each stratum is small enough to cause convergence issues in the estimation of coefficients for at least one stratum. Additionally, further tests reveal that the covariate effect is not the same across all stratum.

As a final attempt to remedy the situation, we note that cell type and Karnofsky score appear somewhat associated with one another. This allows dropping the Karnofsky score as a variable from the model without total loss of information. This reduces the number of strata down to 4, and allows for stable estimation of coefficients; tests for interactions also reveal that the covariate effects are identical within each stratum. This model is then employed to examine the efficacy of the experimental lung cancer treatment.

Sensitivity analysis was performed wherein the cell type variable was dropped, and Karnofsky score was retained. The same results were attained, and the same conclusions reached. All statistical tests in this report are conducted at the $\alpha = 0.05$ level.

Results

Descriptive statistics

Basic descriptive statistics are provided in Table 1 in the appendix. We note that covariates have a very similar value between treatment groups for Karnofsky score, age, diagnosis time, and proportion of patients who received prior treatment. There is a potential difference between treatment groups with regard to cell types, with the standard treatment group having much more small cell type patients. Additionally, the standard treatment group had a

mean survival time of approximately 124 days, while experimental treatment group had moderately larger mean survival time of 142. However, the median survival time for the standard treatment group was 103 days, while the median survival time for the experimental treatment group was only 52 days.

This odd behavior is easily explained by examining the Kaplan-Meier curves for the two treatment groups (Figure 1 in the Appendix). The survival curves cross around 175 days, indicating a violation of the PH assumption of standard Cox regression models.

Finally, table 2 shows a potential association between Cell Type and Karnofsky Score. Good Karnofsky scores (here defined to be above the median of 60) are somewhat associated with more 'large' and 'squamous' cell types, while poor Karnofsky scores appear to be somewhat associated with more 'small' and 'adeno' type cells. This provides some justification for the use of only one of these variables in the analysis below.

Modelling

The first model constructed to assess treatment efficacy was a standard Cox PH model utilizing all available covariates. As suggested by the crossing survival curves in Figure 1, analysis of Schoenfeld residuals suggest that the PH assumption is violated for the variables *age, Karnofsky score,* and *cell type* (Se Figures 2 to 3). Individual significance tests using time varying covariates confirm the PH violation for each variable (see table 3). Cell type was dichotomized to small or not small for the purposes of this test (see below). Martingale residuals suggest that a linear form is appropriate for all variables.

As a remedy, we attempt to stratify on these variables. Karnofsky score is dichotomized into good and poor values, with good being defined as greater than the overall median score of 60. Age is additionally dichotomized into young and old, with old being defined as the greater

than the overall median age of 50. Finally, due to the somewhat small sample size, cell type is grouped into "small cell" and "not small cell" categories (a relatively common distinction made in medicine). We then have a total of eight strata, and use them to fit a stratified Cox model. Multiple problems present themselves at once. Namely, the sample sizes within each stratum are, in some cases, too small to yield stable parameter estimates. Additionally (see table 4), the likelihood ratio test suggests that treatment effects are not the same within each stratum. While this may simply be an artifact of the poor parameter estimates, an additional remedy is still required.

As a final remedy, we remove Karnofsky score from the model entirely. The moderate association with cell type prevents complete loss of information. The final model is a stratified Cox model stratified on cell type and age, for a total of 4 stratum, and including diagnosis time and prior treatment indicator as covariates. The assumption of equal treatment effects within stratum appears to be satisfied, see table 4 for formal test results. A graphical inspection of Cox Snell residuals to assess overall model fit suggests an overall good fit (see figure 5). Sensitivity analysis was conducted by stratifying on age and Karnofsky score, and dropping cell type. The results can also be seen in tables 4 and 5. Similar results were obtained in both cases, and identical conclusions were drawn from both models.

Efficacy of treatment

We assess the efficacy of the experimental treatment using the final stratified cox model described above. A local test of significance for treatment efficacy adjusting for other treatments returned a p-value of $p = 0.3994 > 0.05$. See table 5. We therefore conclude on the basis of this data and analysis, that there is no evidence to suggest that the experimental treatment is more effective than the standard treatment. The complications arising from the use of PH models for

the analysis above may leave an analysist wanting for a more satisfactory analysis, or uneasy at the conclusions reached. To this end, we suggest potential remedies in the conclusion section below.

Conclusion

In this study, we used Cox proportional hazards models to assess the efficacy of an experimental lung cancer treatment using data from a study conducted by the US Veteran's Administration. Our analysis indicates that this new treatment is no more effective than the standard treatment. While this may seem at first glance to be surprising, lung cancer has been historically one of the most difficult cancer types to treat. In addition, this data, being from the late 1970's to early 1980's, likely represents the very beginning stages of the development of modern lung cancer treatment. It should therefore be no great surprise that this treatment was one of many that likely failed.

However, the use of proportional hazards models in this study was fraught with complications; due to a small to moderate sample size, the results obtained from the use of proportional hazards models are strained. It is suggested that additional analysis be conducted on this data, using models that do not make use of the proportional hazard assumption; namely accelerated failure time models (AFT), the most common alternative to proportional hazards models. The absence of the PH assumption in AFT models greatly improves the potential for more reliable analysis of this data, wherein the violation of the PH assumption made analysis using PH models rather complicated. Additionally, AFT models enjoy the added flexibility of allowing covariate effects to change over time.

References

Klein, J. P., & Moeschberger, M. L. (2005). *Survival analysis: Techniques for censored and truncated data* (2nd ed.). New York, NY: Springer.

Appendix I Tables and Figures

Table 1 Summary Statistics by Treatment Group.

	Standard Treatment ($N = 69$)		Experimental Treatment ($N = 68$)	
	Mean	SD	Mean	SD
Age	57.5	10.8	57.9	10.3
Karnofsky Score	59.2	18.7	57.9	21.4
Diagnosis Time	8.7	8.8	8.9	12.2
	N	%	N	%
Cell Type	-	-	-	-
Small	30	43.5	18	26.5
Adeno	9	13.0	18	26.5
Squamous	15	21.7	20	29.4
Large	15	21.7	12	17.7
Prior Treatment	21	0.30	19	0.28
	Mean (SE)	Median	Mean (SE)	Median
Survival Time	123.928 (14.961)	103.000	142.061 (27.023)	52.500

Table 2 Contingency Table for Cell type and Karnofsky Score

	Karno (< 60)	Karno (≥ 60)	Total
Adeno	12 (10.2)	15 (16.8)	27
Large	6 (10.2)	21 (16.8)	27
Small	22 (18.2)	26 (29.8)	48
Squamous	12 (13.3)	23 (21.7)	35
Total	52	85	137

*Expected counts assuming independence shown in parentheses.

Table 3 Individual Tests for Violation of PH Assumption Using Time Varying Coefficients

Test	**Chi Square**	**p-value**
CellSmall	51.8284	<.0001*
CellSmall * log(t)	44.7641	<.0001*
Karno	128.6461	<.0001*
Karno * log(t)	163.9676	<.0001*
Age	130.6841	<.0001*
Age * log(t)	145.7508	<.0001*

* Indicates Significance at the $\alpha = 0.05$ level

Table 4 Likelihood Ratio Tests for Unequal Covariate Effects within Stratified Cox Models

MODEL	**LRT**	**Df**	**p-value**
Cox Model (8 Strata)	33.404	21	0.0419*
Cox Model (4 Strata) Without Karnofsky	8.924	9	0.4397
Cox Model (4 Strata) Without Cell type	8.567	9	0.4782

***** Indicates Significance at the $\alpha = 0.05$ level

Table 5 Final model fit

Variable	**HR**	***p***	**95% Wald CI**
	Model without Karnofsky Score		
Treatment	1.184	0.3994	0.571, 1.250
Diagnosis Time	1.009	0.3199	0.991, 1.028
Prior Therapy	0.849	0.4725	0.544, 1.326
	Sensitivity Analysis		
	Model without Cell Type		
Treatment	0.992	0.9683	0.692, 1.467
Diagnosis Time	1.004	0.6730	0.986, 1.022
Prior Therapy	0.868	0.5233	0.562, 1.340

Standard treatment and no prior therapy are treated as the baseline.

Figure 1 Kaplan-Meier Survival Estimates

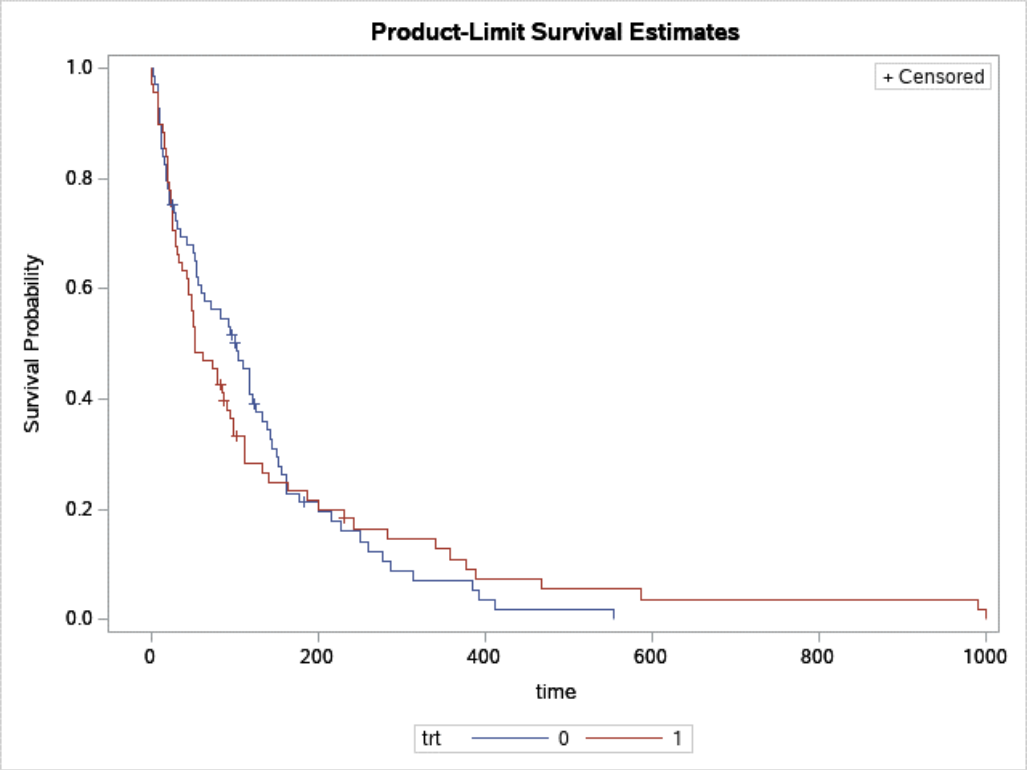


Figure 2 Smoothed Schoenfeld Residuals for Karnofsky Score

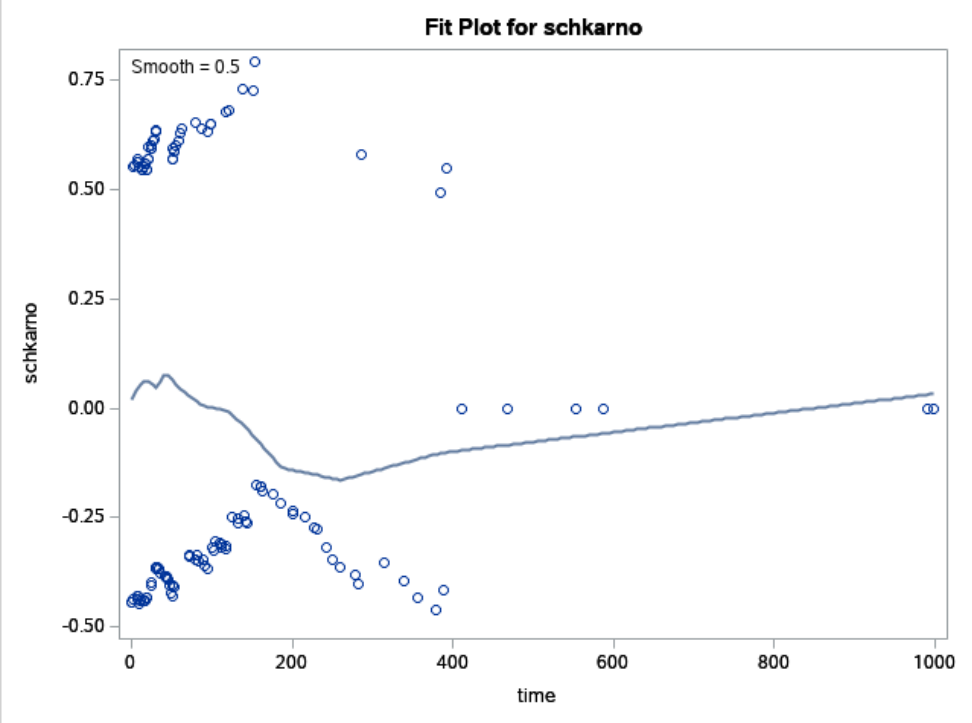


Figure 3 Smoothed Schoenfeld Residuals for Cell Type

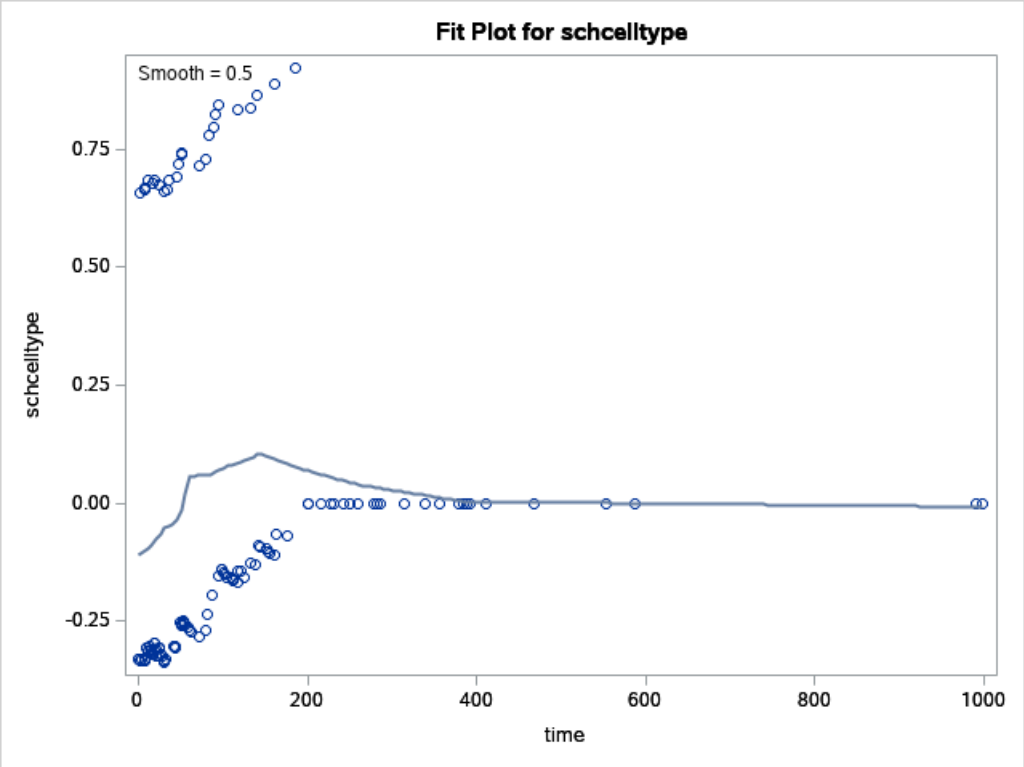


Figure 4 Smoothed Schoenfeld Residuals for Age

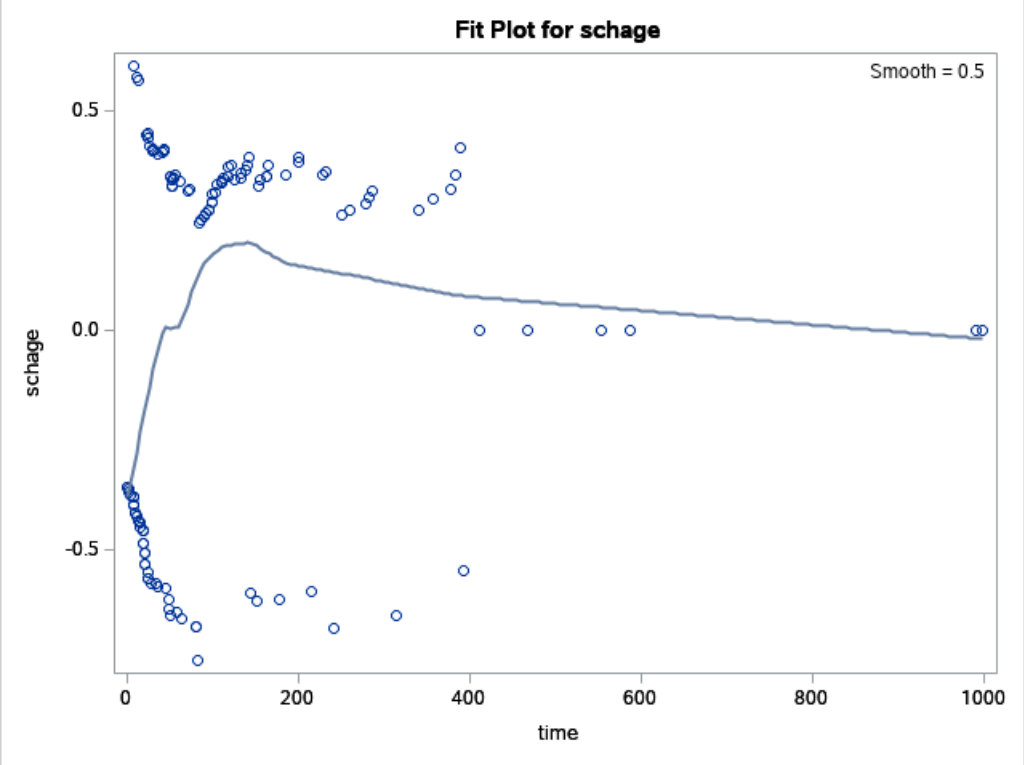
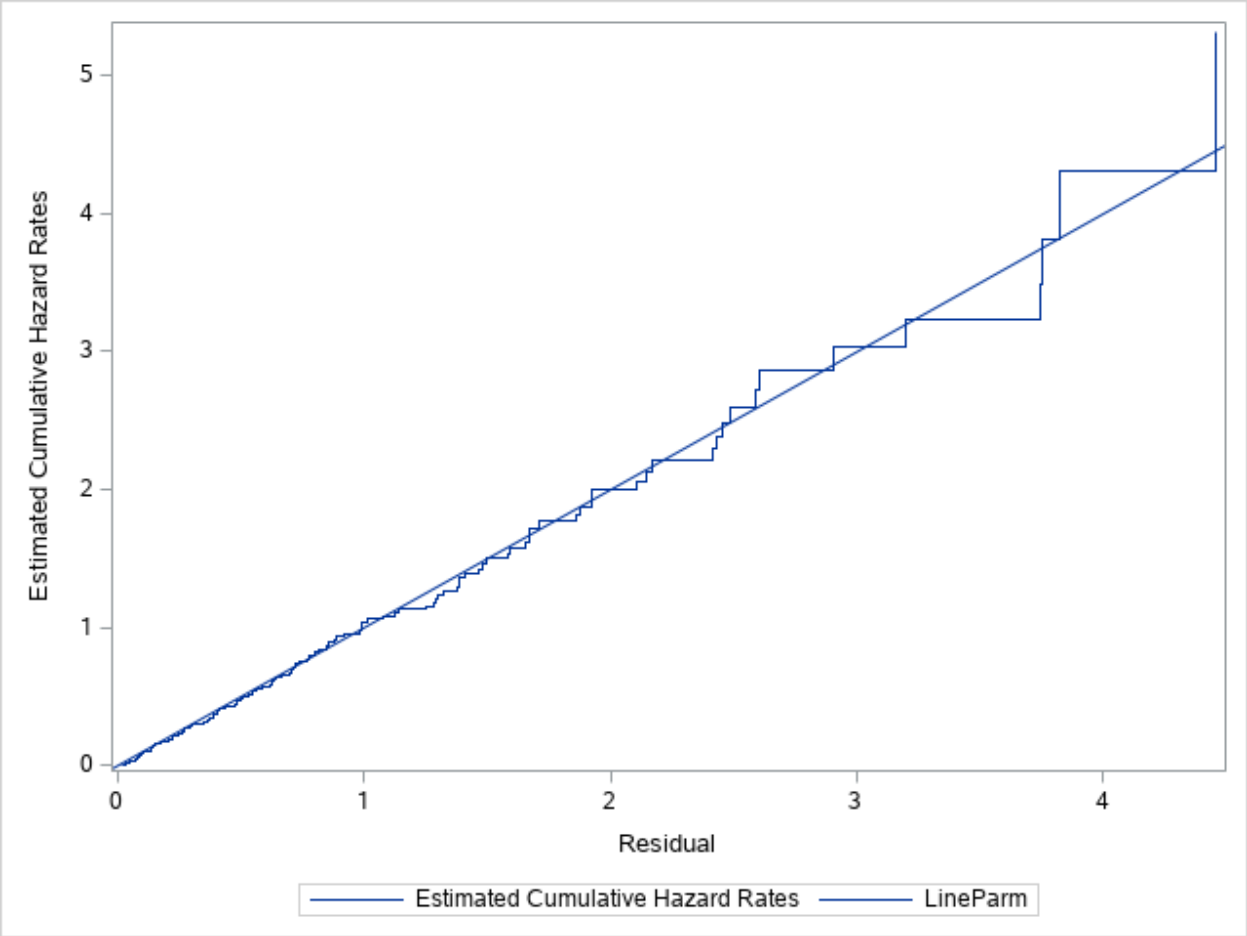


Figure 5 Snell residuals for the final Stratified Cox Model



Appendix II SAS Code

```
proc import datafile='/folders/myfolders/proj1/veteran.csv'
out=veteran dbms=csv replace;
datarow=2;
getnames=yes;
guessingrows=100;
run;
/*trt=1 is standard treatment; trt=2 is new treatment*/

data veteran;
set veteran;
if trt=2 then trt=1; else trt=0;
if prior=10 then prior=1; else prior=0;
run;
/*recode treatment: 0 is standard; 1 is experimental*/
/*recode prior treatment: 0 is no prior treatment; 1 is prior treatment */

proc means data=veteran
      nmiss max min mean std median;
      class trt;
run;

PROC SORT Data=veteran Out=veteran_sort;
BY trt;
RUN;
proc freq data=veteran_sort;
table celltype;
by trt;
run;

data veteran;
set veteran;
where trt in (0,1);
      if celltype='large' then celllarge=1; else celllarge=0;
      if celltype='smallcell' then cellsmall=1; else cellsmall=0;
      if celltype='adeno' then celladeno=1; else celladeno=0;
      if celltype='squamous' then cellsquamous=1; else cellsquamous=0;
trtt = trt*log(time);
karnot = karno*log(time);
diagtimet = diagtime*log(time);
aget = age*log(time);
priort = prior*log(time);
cellsmallt=cellsmall*log(time);
celllarget=celllarge*log(time);
celladenot=celladeno*log(time);
cellsquamoust=cellsquamous*log(time);
if (age>50) then ageold=1; else ageold=0;
if (karno>=60) then karnogood=1; else karnogood=0;
run;
```

```
proc lifetest data=veteran method=KM nelson plots=hazard(kernel=E);
time time*status(0);
strata trt;
run;

PROC SORT Data=veteran Out=veteran_sort1;
BY karnogood;
RUN;

proc freq data=veteran_sort1;
table celltype;
by karnogood;
run;

proc phreg data=veteran plots=survival;
class trt(ref="0") celltype(ref='large');
model time*status(0) = trt celltype karnogood diagtime age prior;
output RESMART=mgale out=diagnostics ressch=schtrt schcelltype schkarno schdiagtime
schage schprior;
run;

PROC LOESS data=diagnostics;
model mgale =karno diagtime age /smooth=(0.5);
run;

proc loess data=diagnostics;
model schtrt schcelltype schkarno schdiagtime schage schprior = time/smooth=(0.5);
run;

/*Time varying coefficent tests*/
proc phreg data=veteran;
class trt(ref='0') cellsmall(ref='0');
model time*status(0) = trt cellsmall diagtime age prior karno cellsmallt;
run;

proc phreg data=veteran;
class trt(ref='0') cellsmall(ref='0');
model time*status(0) = trt cellsmall diagtime age prior karno karnot;
run;

proc phreg data=veteran;
class trt(ref='0') cellsmall(ref='0');
model time*status(0) = trt cellsmall diagtime age prior karno aget;
run;

/*Stratafied Model with 8 strata*/
proc phreg data=veteran;
class trt(ref='0');
model time*status(0) = trt  diagtime prior;
STRATA cellsmall karnogood ageold;
run;
```

```
/*8 Separate STRATA*/
proc phreg data=veteran;
class trt(ref='0');
where cellsmall=0 and karnogood=0 and ageold=0;
model time*status(0) = trt diagtime prior;
run;

proc phreg data=veteran;
class trt(ref='0');
where cellsmall=0 and karnogood=0 and ageold=1;
model time*status(0) = trt diagtime prior;
run;

proc phreg data=veteran;
class trt(ref='0');
where cellsmall=0 and karnogood=1 and ageold=0;
model time*status(0) = trt diagtime prior;
run;

proc phreg data=veteran;
class trt(ref='0');
where cellsmall=0 and karnogood=1 and ageold=1;
model time*status(0) = trt diagtime prior;
run;

proc phreg data=veteran;
class trt(ref='0');
where cellsmall=1 and karnogood=0 and ageold=0;
model time*status(0) = trt diagtime prior;
run;

proc phreg data=veteran;
class trt(ref='0');
where cellsmall=1 and karnogood=0 and ageold=1;
model time*status(0) = trt diagtime prior;
run;

proc phreg data=veteran;
class trt(ref='0');
where cellsmall=1 and karnogood=1 and ageold=0;
model time*status(0) = trt diagtime prior;
run;

proc phreg data=veteran;
class trt(ref='0');
where cellsmall=1 and karnogood=1 and ageold=1;
model time*status(0) = trt diagtime prior;
run;

/*Final Stratified model, dropping Karno score, 4 total strata*/
```

```
proc phreg data=veteran;
class trt(ref='0');
model time*status(0) = trt diagtime prior;
STRATA cellsmall ageold;
output out = diagnostics LOGSURV = logsurv1 /method = ch;
hazardratio trt / CL=WALD;
hazardratio diagtime / CL=WALD;
hazardratio prior / CL=Wald;
run;
data diagnostics;
set diagnostics;
snell = -logsurv1;
cons = 1;
run;

proc phreg data = diagnostics;
model  snell*status(0) = cons;
output out = cox_snell_plot logsurv = logsurv2 /method = ch;
run;

data cox_snell_plot;
set cox_snell_plot;
cumhaz = - logsurv2;
run;

proc sort data = cox_snell_plot;
by snell;
run;

proc sgplot data = cox_snell_plot;
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;

/*4 Strata, dropping Karno completely*/
proc phreg data=veteran;
class trt(ref='0');
where cellsmall=0 and ageold=0;
model time*status(0) = trt diagtime prior;
run;

proc phreg data=veteran;
class trt(ref='0');
where cellsmall=0 and ageold=1;
model time*status(0) = trt diagtime prior;
run;

proc phreg data=veteran;
class trt(ref='0');
```

```
where cellsmall=1 and ageold=0;
model time*status(0) = trt diagtime prior;
run;

proc phreg data=veteran;
class trt(ref='0');
where cellsmall=1 and ageold=1;
model time*status(0) = trt diagtime prior;
run;

/*--------------------Sensitivity Analysis----------------------*/
/*Stratified model dropping cell type instead*/
proc phreg data=veteran;
class trt(ref='0');
model time*status(0) = trt diagtime prior;
STRATA karnogood ageold;
hazardratio trt / CL=WALD;
hazardratio diagtime / CL=WALD;
hazardratio prior / CL=Wald;
run;

/*4 Strata, dropping Karno completely*/
proc phreg data=veteran;
class trt(ref='0');
where karnogood=0 and ageold=0;
model time*status(0) = trt diagtime prior;
run;

proc phreg data=veteran;
class trt(ref='0');
where karnogood=0 and ageold=1;
model time*status(0) = trt diagtime prior;
run;

proc phreg data=veteran;
class trt(ref='0');
where karnogood=1 and ageold=0;
model time*status(0) = trt diagtime prior;
run;

proc phreg data=veteran;
class trt(ref='0');
where karnogood=1 and ageold=1;
model time*status(0) = trt diagtime prior;
run;
```