Dealing with Non-Proportional Hazards in Survival Analysis

July 2013

Declaration

This is my original work and has never been presented for any academic award in any other learning institution.

Otieno Otieno C.

Sign: .....

Admission Number:I56/74041/2011

Date: .....

APPROVAL

This is to certify that the project titled "Dealing with Non-Proportional Hazards in

Survival Analysis" carried out by above named student has been read and approved for meeting part of the requirements and regulations governing the award of the Masters of Science in Statistics (Biomerty) degree of University of Nairobi, Kenya.

Supervisor:Dr.Nelson O.Owuor

Sign: .....

Date: .....

Dedication

This project is dedicated to my mother Y.Atieno Otieno and father G.Otieno Nyabinge for taking me to school, inspirations and sponsoring me all through for the fulfillment of this dream.

Acknowledgements

I would like to acknowledge with gratefulness the endeavor from my supervisor Dr. Nelson O.Owour for his guidance, patience, encouragement and overwhelming support provided to me throughout this work.

I am profoundly grateful to the entire University of Nairobi, especially the Mathematics Department staff for their support at whatever echelon by making me realize one of my dreams. It would be unfair without revealing members of our family for ensuring that my financial tribulations were sorted accordingly, for condusive environment, for peace of mind and peace of heart till the end of my work.

Finally, it is indeed my great pleasure to appreciate the support accorded to me by my teammates and fellow students during the study period.

Above all I thank the Almighty God for the perfect health and unrelenting safeguard all the time.

Abstract

For the last few decades,exceptional concentration has been given to the field of survival analysis for its techniques applied in different areas of research.One of the main assumptions regarding a number of survival analysis techniques is that of proportional hazards.If the hazards rates are found to cross,non parametric analyses such as Cox proportional hazards regression,Kaplan Meier and Log- rank test will either be rendered scarce or lose power.One way to suffiently analyze survival data with crossing hazard rate is to obtain inference using the Renyi test statistic and Stratified Cox Regression Model,interaction model with time dependent covariate.

Contents

Chapter 1

Introduction

Background

It is not common for clinical trials to present results on survival time as Kaplan-Meier survival curves that cross,indicating non-proportional hazards. A recent example was given in a pivotal trial in advanced non-small cell lung cancer (The 'IPASS study' [1]).Trials such as these present a hazard ratio and log-rank test for treatment comparison as this is their planned primary analysis. However,the validity of such analysis is questionable and has received published criticism. Therefore reviews have used the log-rank test with crossing curves and alternatives proposed.

The analysis of lifetimes is an important topic within biology and medicine in particular, but also in reliability analysis with engineering applications. Such data

are often highly nonnormally distributed, so that the use of standard linear models is problematic. Lifetime data are often censored: You do not know the exact lifetime, only that it is longer than a given value. For instance, in a cancer trial some people are lost to follow-up or simply live beyond the study period. It is an error to ignore the censoring in the statistical analysis, sometimes with extreme consequences. Consider, for instance, the case where a new treatment is introduced toward the end of the study period, so that nearly all the observed lifetimes will be cut short.

Essential concepts

Let X be the true lifetime and T a censoring time. What you observe is the minimum of X and T together with an indication of whether it is one or the other. T can be a random variable or a fixed time depending on context, but if it is random then it should generally be noninformative for the methods we describe here to be applicable. Sometimes "dead from other causes" is considered a censoring event for the mortality of a given disease, and in those cases it is particularly important to ensure that these other causes are unassociated with the disease state.The survival function S(t) measures the probability of being alive at a given time. It is really just 1 minus the cumulative distribution function for X,1 - F(t).The hazard function or force of mortality h(t) measures the (infinitesimal) risk of dying within a short interval of time t, given that the subject is alive at time t. If the lifetime distribution

has density f , then h(t) = f (t)/S(t).This is often considered a more fundamental quantity than (say) the mean or median of the survival distribution and used as a basis for modelling.

1.1 Statement of Problem

Analysis in a case of non-proportionality in Survival analysis

1.2 Objective of the study

1.To present an alternative way of working out non-proportional hazards in Survival analysis.

1.3 Justification/Significance of Study

The study offers a breakthrough in that new techniques of handling even complex situations are realised,therefore increasing the bank of knowledge in the field of Survival analysis.

Chapter 2

Literature Review

2.1 Non Proportionality

Nathan Mantel and David Cox,(1966),improvised a weighting function W(ti)=1 called the Log rank test.

Gehan and Wilcoxon,(1965),Generalized Mannwhitney-Wilcoxon test,

Breslow's(1970) Generalization of Kruskal-Wallis test suggested a weighting function W(ti)=Yi.

Tarone and Ware(1977) put forward a weighting function W(ti)=f(yi) where f=fixed function, with a choice of f(y)= y^1/2.

Peto Peto(1972) and Kalbfleisch and Prentice(1980) advanced Generalized Mannwhitney-

Wilcoxon test by proposing a weighting function W(ti)=$\hat{S}$(ti) where $\hat{S}$(ti) is given by

$$\hat{S}(t_i) = \prod_{t_i \leq t} (1 - \frac{d_i}{Y_{i+1}})$$

2.2 Other papers partaining Non proportional hazards

In the paper Cox analysis of survival data with non-proportional hazard functions by Michael Schemper (1992),The consequences of violated assumptions for Cox's proportional hazards model and current options to deal with non-proportionality in Cox's model are reviewed. An additional option for analysis is suggested, which produces weighted estimates of log hazard ratios, weighted at the time points where failures occur. The procedure amounts to generalizations of the tests by Breslow or Prentice for multiple covariates in the same manner that the proportional hazards model is a generalization of the log rank test by Mantel. Its advantages are representative estimates of average hazard ratios also for covariates with non-proportional and, in particular, converging hazard functions. The latter are often encountered in clinical applications. By means of an empirical study these average hazard ratios are

shown to be very close to exact calculations of average hazard ratios as defined by Kalbfleisch and Prentice. Two examples illustrate the advantages of the weighted estimation and of other strategies for analysis with the Cox model in the presence of non-proportional hazards. Furthermore, with respect to checking proportionality, it is demonstrated how misleading the frequently used log-minus-log plots can be and that the lesser known Arjas plots seem to perform quite well.

In nonproportional Hazards and Event History Analysis in International Relations by Janet M. Box-Steffensmeier and Dan Reiter Christopher Zorn,(Feb., 2003) illustrate how to relax the PH assumption, detailed description of the PH assumption in event history and the use of what we refer to here as "nonproportional hazards" (NPH) analysis and then demonstrate how NPH can be applied to international relations research.Specifically,they reanalyse data from previously published scholarship on postwar peace, civil wars, and alliances to show how using NPH permits improved estimation and the testing of more refined hypotheses and conclude by highlighting why NPH make sense substantively and emphasizing the methodological point that the assumption should be tested for all PH event history models.

Confidence intervals for the first crossing point of two hazard functions by Ming-Yen Cheng · Peihua Qiu · Xianming Tan · Dongsheng Tu ,(1 November 2009) argued that the phenomenon of crossing hazard rates is common in clinical trials with time

to event endpoints and that methods have been proposed for testing equality of hazard functions against a crossing hazards alternative. However, there has been relatively few approaches available in the literature for point or interval estimation of the crossing time point. The problem of constructing confidence intervals for the first crossing time point of two hazard functions is considered in this paper.After reviewing a recent procedure based on Cox proportional hazard modeling with Box-Cox transformation of the time to event, a nonparametric procedure using the kernel smoothing estimate of the hazard ratio is proposed. The proposed procedure and the one based on Cox proportional hazard modeling with Box-Cox transformation of the time to event are both evaluated by Monte–Carlo simulations and applied to two clinical trial datasets.

Regression Models and Non-proportional Hazards in the Analysis of Breast Cancer Survival by Sheila M. Gore, Stuart J. Pocock and Gillian R. Kerr,(January 1984) .A Western General breast cancer series of 3922 patients sets research methodology for survival data in practical perspective illustrates that the waning of covariate effects through time is an important phenomenon in medical applications. Non-monotone convergent hazard functions are associated with most clinical covariates in breast cancer,an unusual hazard pattern according to menopausal state is also reported. These features contraindicate the use of standard regression models for sur-

vival such as Weibull and proportional hazards. Inferences about covariate effects are compared under these and a log-logistic model which implies proportionality of the cumulative odds on death. Regression models are shown to be useful in exploratory analysis. In particular, a step-function proportional hazards model elucidates the time-dependent influence of initial covariates and leads to a more appropriate final model, but one whose virtues are balanced by computational difficulty.

Crossing Hazard Functions in Common Survival Models by Jiajia Zhang and Yingwei Peng ,confer that crossing hazard functions have extensive applications in modeling survival data . However, existing studies in the literature mainly focus on comparing crossed hazard functions and estimating the time at which the hazard functions cross, and there is little theoretical work on conditions under which hazard functions from a model will have a crossing.These paper investigate crossing status of hazard functions from the proportional hazards (PH) model, the accelerated hazard (AH) model, and the accelerated failure time (AFT) model. It then provides and proves conditions under which the hazard functions from the AH and the AFT models have no crossings or a single crossing. A few examples are also provided to demonstrate how the conditions can be used to determine crossing status of hazard functions from the three models.

Analysis of survival data with nonproportional hazard functions by Donald M.

Stablein, Walter H. Carter Jr., Joel W. Novak (June 1981), the log-rank test or the proportional hazard model is a valuable, widely accepted method for analyzing time-to-response data from comparative clinical trials. When the hazard ratio is constant in time, this procedure is optimal. Indiscriminate or unthinking use of this approach results in problems in the determination of treatment differences. For example, when the true survival curves intersect, the hazard ratio cannot be constant, i.e., the hazard functions are not proportional. It is shown that by considering time-by-treatment interactions we gain flexibility in describing the relationships among hazard functions. In this paper we demonstrate with the results of a clinical trial how available methodology can be used to permit tests for the appropriateness of the model and to enable infirmative analysis of such data.

Comparing two crossing hazard rates by Cox proportional hazards modelling by Liu K, Qiu P, Sheng J,(2007),motivated by a clinical trial of zinc nasal spray for the treatment of the common cold considered the problem of comparing two crossing hazard rates. A comprehensive review of the existing methods for dealing with the crossing hazard rates problem is provided. A new method, based on modelling the crossing hazard rates, is proposed and implemented under the Cox proportional hazards framework. The main advantage of the proposed method is the utilization of the Box-Cox transformation which covers a wide range of hazard crossing patterns. Sim-

ulation studies are conducted for comparing the performance of the existing methods and the proposed one, which show that the proposed method outperforms some of its peers in certain cases. Applications to a kidney dialysis patients data and the zinc nasal spray clinical trial data are discussed.

In Modelling Survival Data with Crossing Hazards by MacKenzie, Gilbert; Do Ha, II converse that despite the ubiquity of Cox's proportional hazards (PH) model it is being realised increasingly that not all survival data obey the PH assumption.In multi-factor studies the effect of one or more covariates may be noticeably non-PH. A clear signal is that of crossing hazards. A classical example is the well-known data set of the Gastrointestinal Tumor Study Group (GTSG)(1982), reporting the effects of chemotherapy and combined chemotherapy and radiotherapy on the survival times of gastric cancer patients.The question then arises as to how best to model these effects. Sometimes, in practice, non-PH covariates are ignored and they are analysed as being PH in a larger model, but the optimality of this expediency is unclear.An alternative approach is to adopt a model which can cope with non-PH and PH effects. The Generalised Time-Dependent Logistic family of survival models contains two non-PH parametric models which are potential competitors for Cox's model, namely, the GTDL model (MacKenzie, 1996) and the logistic accelerated life model, the LAL (Al-tawarah & MacKenzie, 2003). Recently, the family has been extended to incorporate

frailty (Blagojevic, MacKenzie & Ha, 2003) and to more general multivariate forms (Blagojevic & MacKenzie, 2007).In relation to tests and models developed specically for crossing hazards situations per se we refer the reader to Stablein & Koutrouvelis (1985),Aalen(1994), Hseish (2001) and Bagdonavicius et al (2005).

In the paper Significance tests of differences between two crossing survival curves for small samples by Tomasz Jurkiewicz and Ewa Wycinka,(2011) examined a small-sample characteristics of recently developed tests that compare survival at two or more cohorts,most popular are log-rank test and tests of Gehan,Tarone-Ware,Peto-Peto,Harrington-Fleming,Renyi-type.There were made a variety of simulations by means of Monte Carlo simulations.With the assumption that survival curve has Weibull distribution,there were taken into consideration different share of censored observations(randomly appeared due to uniform distribution) and the ability of these test to detect overall differnces between crossing survival curves.

Cox regression models with nonproportional hazards applied to lung cancer survival data by Nihal Ata and M.Tekin Sozer,(28:09:2007) acknowledge that Cox regression model is widely used for the analysis of treatment and prognostic effect with censored survival data,makes the assumption of constant hazard ratio.In the violation of this assumption,different methods should be used to deal with nonproportionality of hazards.In this study the stratified cox regression model and extended

cox regression model,which uses time dependent covariate terms with fixed functions of time are discussed.Results are illustrated by an analysis of lung cancer data in order to compare these methods with respect to Cox regression model in the presence of nonproportional hazards.

Keywords

- hazard function;log-rank test;survival data;time-by-treatment interactions;accelerated hazard model; accelerated failure time model; monotone hazard; U-shape hazard; bellshape hazard.Cox regression model,hazard ratio,Non-proportional hazards,Stratified Cox regression model,Extended Cox regression model,Time dependent-covariate,Lung cancer.

Chapter 3

Methodology

Several datasets used to reinforce different concepts are used but two of them are heavily employed,which are Stablein and Koutrouvelis(1985) and Survival data from a survey in KEMRI-Kilifi,Mombasa.A few forms of weighting functions and a detailed illustration on how Renyi type test works is also exemplified.For Stablein and Koutrouvelis(1985),Kaplan Meier technique is used to illustrate non-proportionality of hazard functions by crossing,whereas stratified Cox Regression Model is utilized for the reason that assumption of proportionality is violated as well,that is, effects of a given covariate(s) are changing over time.

Assumptions of Kaplan Meier:

$$\hat{S(t_j)} = \prod_{t=0}^{t_j} \frac{Y_t - d_t}{Y_t}$$

1.Probabilities for the event of interest should depend only on time after the initial event–they are assumed to be stable with respect to absolute time. That is, cases that enter the study at different times (for example, patients who begin treatment at different times) should behave similarly.

2.There should also be no systematic differences between censored and uncensored cases. If, for example, many of the censored cases are patients with more serious conditions, your results may be biased.

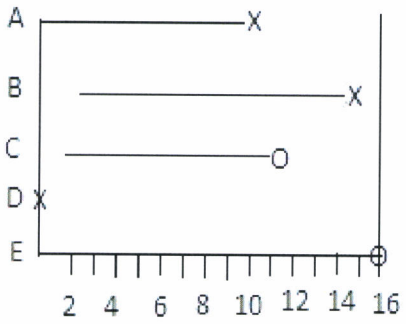
The stratified Cox Regression Model is given by

$$h_g(t,x) = h_{0g}(t)\exp[\beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p], g = 1, 2, ..., k$$

where $h_{0g}(t)$ is the baseline hazard function.To obtain estimates of the regression coefficients $\beta_1, \beta_2, ..., \beta_p$ a likelihood function L that is obtained by multiplying together the likelihood functions for each stratum is maximized.The software used for analysis is R.

3.1 Comparing two Survival Curves Kaplan-Meier Curves

An illustration of how Survival data is captured



X=Event Recorded

O=No event

	Data Structure	
Individual	Survival time	$\delta(t)$
A	10	1
B	14	1
C	11	0
D	0	1
E	16	0

Data=$\{10, 14, 11+, 0, 16+\}$ where $11+$,$16+$ are right censored

Quartiles of Survival analysis

Hazard Survival Function

Time,tj	Death,dj	Risk Set,Yj	Probability of dying,dj/Yj	Survival
t_1	d_1	Y_1	d_1/Y_1	$1-d_1/Y_1$
t_2	d_2	Y_2	d_2/Y_2	$1-d_2/Y_2$
	.		.	.
	.		.	.
	.		.	.
t_r	d_r	Y_r	d_r/Y_r	$1-d_r/Y_r$

where

$$d_j/Y_j = Hazard\ \ Rate$$

Actual Kaplan Meier estimate

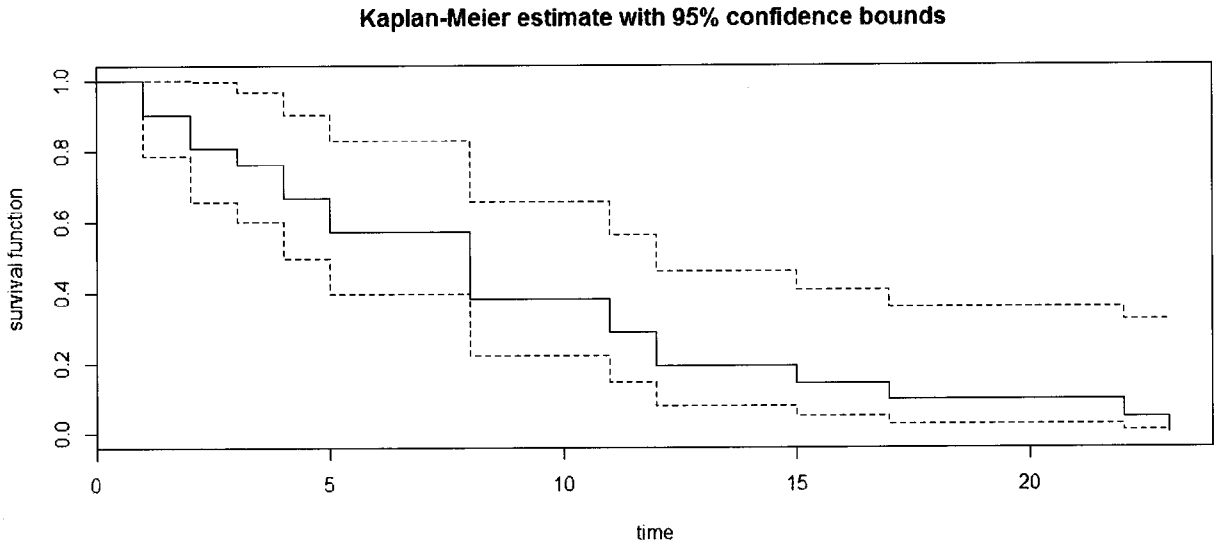
$$\hat{S(t_j)} = \prod_{t=0}^{t_j} \frac{Y_t - d_t}{Y_t}$$

3.2 Consider the data (Survival times) without Censoring

1,1,2,2,3,4,4,5,5,8,8,8,8,11,11,12,12,15,17,22,23

t_i	d_i	Y_i	$\frac{Y_i - d_i}{Y_i}$	$\hat{S(ti)} = \prod_{t=0}^{ti} \frac{Yi-di}{Yi}$
0	0	21	1.000	1.000
1	2	21	0.905	0.905
2	2	19	0.895	0.810
3	1	17	0.941	0.762
4	2	16	0.875	0.667
5	2	14	0.857	0.571
8	4	12	0.667	0.381
11	2	8	0.750	0.286
12	2	6	0.667	0.190
15	1	4	0.750	0.143
17	1	3	0.667	0.095
22	1	2	0.500	0.048
23	1	1	0.000	0.000

Kaplan Meier Plot of $\hat{S(t)}$

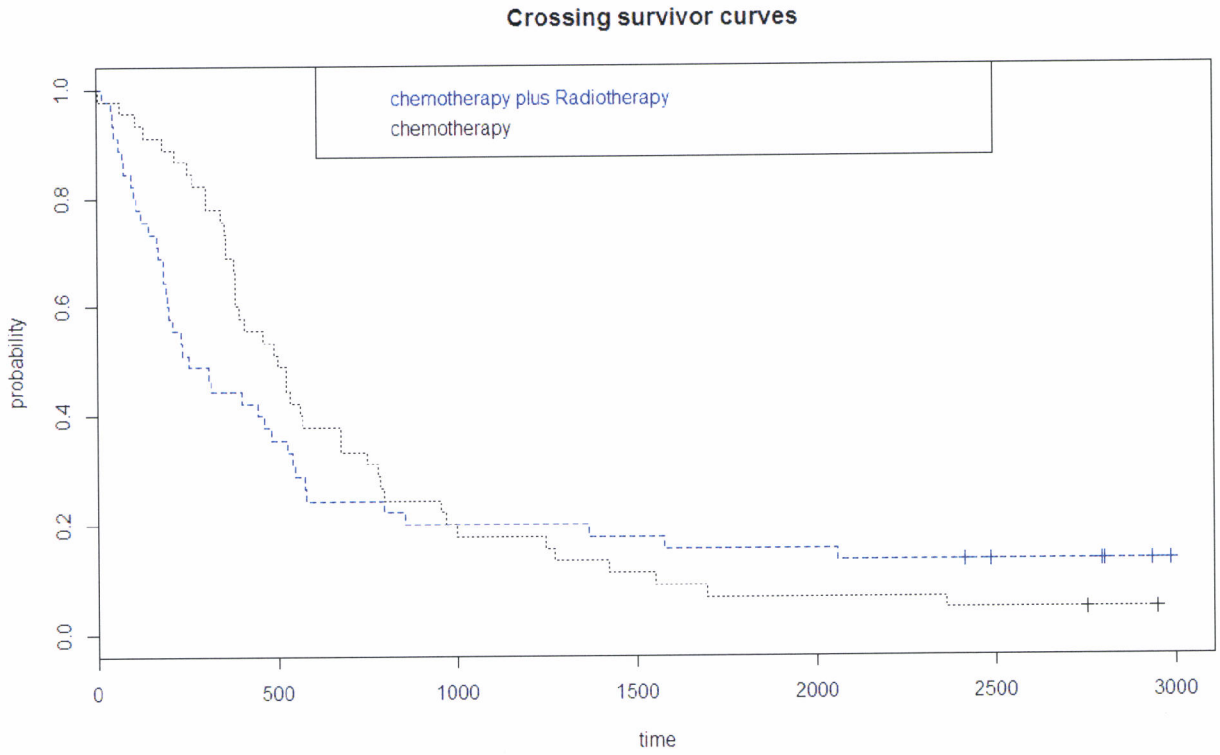


Property

-Kaplan Meier Curve if no censoring begins at one and end at zero.

Illustration

A clinical trial of chemotherapy against chemotherapy combined with radiotherapy in the treatment of locally unresectable gastric cancer was conducted by the Gastrointestinal Tumor Study Group(1982).In this trial,forty-five patients were randomized to each of the arms and followed for about eight years.The data,found in Stablein and Koutrouvelis(1985),is as follows.


Crossing survivor curves
chemotherapy plus Radiotherapy
chemotherapy
probability
0.0
0.2
0.4
0.6
0.8
1.0
0
500
1000
1500
2000
2500
3000
time

Test of hypothesis on two (or more) Kaplan Meier Curves H_0:$h_1(t) = h_2(t) = ... = h_k(t)$; $t \leq \tau$

H_1:at least one hazard is different

Assume:

Group1=$\{30, 35, 40\}$

Group2=$\{30+, 45, 50, 65\}$

Group3=$\{20+, 25+, 30, 35, 42\}$

$\tau = 40$ *i.e.*smallest of the largest actual time in all the k groups

D=65 (the largest actual death time in all groups) $t_1 < t_2 < ... < t_D$

j=1,2,...,k(groups)

i=1,2,...,D(actual death times)

$$d_i = \sum_{j=1}^{k} d_{ij}$$

(total deaths in all groups, at time t_i)

$$Y_i = \sum_{j=1}^{k} Y_{ij}$$

(total risk set at t_i,across the groups) and i =time and j =sample

Sample j

Time=t_i Risk Set=Y_{ij} Deaths= d_{ij}

If H_0 (null hypothesis) is true,then the estimate of expected hazard rate is d_i/Y_i,(at time t_i,over all groups) and d_{ij}/Y_{ij}, (in the jth sample).

The test is to compare the sample hazard to the expected hazard but weighted by some weight

$$w_j(t_i) = \begin{cases} 0 & \text{if} \quad Y_{ij} = 0 \\ \\ w_j(t_i) & \text{Otherwise} \end{cases}$$

Test statistic

$$Z_j(\tau) = \sum_{i=1}^{D} w_j(t_i) \left(\frac{d_{ij}}{Y_{ij}} - \frac{d_i}{Y_i} \right)$$

;j=1,2,...k if $Z_j(\tau) \to 0$ fail to reject H_0 and a Larger $Z_j(\tau)$ implies the two groups are different.

Note:Purpose of the weight function is to reduce the difference between sample hazard and expected hazard

$$\left(\frac{d_{ij}}{Y_{ij}} - \frac{d_i}{Y_i} \right)$$

3.3 The Weight Function

If the weight function $w_j(t_i) = 1$ then we are doing a Log Rank Test which is a naive test,commonly used weight function is

$w_j(t_i) = Y_{ij}w(t_i)$ where $w(t_i)$ is constant for all groups

implying that

$$Z_j(t_i) = \sum_{i=1}^{D} w(t_i)\left(d_{ij} - Y_{ij}\frac{d_i}{Y_i}\right), \forall j$$

At time *ti*

Groups	Number of deaths	Number of survivors	Risk set
Group1	d_{i1}	$Y_{i1}-d_{i1}$	Y_{i1}
Group2	d_{i2}	$Y_{i2}-d_{i2}$	Y_{i2}
Total	d_i	Y_i-d_i	Y_i

If we know the distribution of d_{i1} and d_{i2} then we know everything about the table.

$$Z_j(t_i) = \sum_{i=1}^{D} w(t_i)\left(d_{ij} - Y_{ij}\frac{d_i}{Y_i}\right), \forall j$$

d_{ij} =Observed values, $Y_{ij}\frac{d_i}{Y_i}$ =Expected values

Standardizing $Z_j(\tau)$

$$\frac{Z_j(\tau) - E[Z_j(\tau)]}{\sqrt{VarZ_j(\tau)}} \sim N(0,1) \quad or \quad \frac{(Z_j(\tau) - E)^2}{var(Z_j(\tau))} \sim \chi^2_{(1)}$$

$$Var[Z_j(\tau)] = \hat{\delta}_{jj} = \sum_{i=1}^{D} w(t_i)^2 \frac{Y_{ij}}{Y_i}\left(1 - \frac{Y_{ij}}{Y_i}\right)\left(\frac{Y_i - d_i}{Y_i - 1}\right) di$$

$$cov(Zj(\tau), Zg(\tau)) = \hat{\delta}_{jg} = -\sum_{i=1}^{D} w(t_i)^2 \frac{Y_{ij}}{Y_i}\frac{Y_{ig}}{Y_i}\left(\frac{Y_i - d_i}{Y_i - 1}\right) di \quad \text{for } g \neq j$$

where

$$\frac{Y_i - d_i}{Y_i - 1} = \begin{cases} 1 & \text{if no two individuals have a common event time} \\ \neq 1 & \text{if there are common event times i.e. } d_i > 1 \end{cases}$$

Note:For $di > 1$ scales up the estimate

Proof. The terms $\frac{Y_{ij}}{Y_i}\left(1 - \frac{Y_{ij}}{Y_i}\right) di$ and - $\frac{Y_{ij}}{Y_i}\frac{Y_{ig}}{Y_i} di$ are variance and covariance of a multinomial random variable with parameters d_i and $p_j = \frac{Y_{ij}}{Y_i}$ ∀j=1,2,...,k ■

In multinomial experiment $X = \{1, 2, 3, ..., k\}$ a set of outcomes.

Let Y_1 =number of 1's with Probability P_1

Y_2 = number of 2's with Probability P_2

.

.

.

Y_k = number of k's with Probability P_k

such that number of $Y_1, Y_2, ...Y_k$'s is a binomial experiment and that $Y_1 + Y_2 + ... + Y_k = n$

$\underline{Y}$ ~multinomial(n,$P_1, P_2, ..., P_k$)

$$Pr[Y_k = y_k|n, P(vector)] = \frac{n!}{Y_1!Y_2!...Y_k!} P_1^{Y_1} P_2^{Y_2} ... P_k^{Y_k}$$

Yj ~Binomial(n, P_j)

E(Yj)=nP_j and Var(Yj)=$nP_j(1 - P_j)$

Note:

$Y_{j*} + Y_j$ ~Binomial($n, P_{j*+}P_j$) ,the sum of two binomial distributions is binomial

Var($Y_{j*} + Y_j$)=$n(P_{j*} + P_j)(1 - (P_{j*+}P_j))$ implying that Var(Y_j)=$nP_j(1 - P_j)$ = $\frac{Y_{ij}}{Y_i}\left(1 - \frac{Y_{ij}}{Y_i}\right) d_i$

Besides:Var($Y_{j*} + Y_j$)=Var(Y_{j*})+Var(Y_j) + 2cov(Y_{j*}, Y_j)

Equating the two equations then cov(Y_{j*}, Y_j)=-$nP_{j*}P_j$

$$Z_j(\tau) = \sum_{i=1}^{D} w(t) \left(d_{ij} - Y_{ij} \frac{d_i}{Y_i} \right)$$

Note:Log rank case is $w(t) = 1$

$$Z_j(t_i) = \sum_{i=1}^{D} \left(d_{ij} - Y_{ij} \frac{d_i}{Y_i} \right) \ \forall j$$

Time	Group1		Group2		Combined	
t_i	Deaths	Risk set	Deaths	Risk set	Deaths	Risk set
t_1	d_{11}	Y_{11}	d_{12}	Y_{12}	d_1	Y_1
t_2	d_{21}	Y_{21}	d_{22}	Y_{22}	d_2	Y_2
.	.		.		.	
.	.		.		.	
.	.		.		.	
t_m	d_{m1}	Y_{m1}	d_{m2}	Y_{m2}	d_m	Y_m

j=1,2 i=1,2,...,m

We note that $Z_j(\tau)$ is computed for each group j=1,2,...,k.Define the vector $\mathbb{Z} = [Z_1(\tau), Z_2(\tau), ...Z_k(\tau)]^T$

Note:$\sum_{j=1}^{k} Z_j(\tau) = 0$ and test statistic is constructed from k-1 because they are linearly independent vectors.

Test statistics:

$\mathrm{X}^2 = \mathbb{Z}^T \sum^{-1} \mathbb{Z} \sim \chi^2_{(k-1)}$ and for equal dimensions $\mathbb{Z}^* = \left[Z_1(\tau), Z_2(\tau), ...Z_{(k-1)}(\tau)\right]^T$

$\sum = (k-1)(k-1)$ variance covariance matrix of $Zj(\tau)$'s=

$$\begin{pmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{12} & . & . & . & \hat{\sigma}_{1(k-1)} \\ \hat{\sigma}_{21} & \hat{\sigma}_{22} & . & . & . & \hat{\sigma}_{2(k-1)} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \hat{\sigma}_{(k-1)1} & \hat{\sigma}_{(k-1)2} & . & . & . & \hat{\sigma}_{(k-1)(k-1)} \end{pmatrix}$$

3.4 For $K = 2$ or $K \geq 2$ (Special Case)

We are testing the hypothesis

H_0:$h_1(t) = h_2(t)$

H_1:$h_1(t) > h_2(t)$ (one graph is above the other)

Reject H_0 if $\mathrm{X}^2_{special} \geq Z_\alpha$

$$X^2_{special} = \frac{\sum_{i=1}^{D} w(t_i)\left(d_{ij} - Y_{ij}\frac{d_i}{Y_i}\right)}{\sqrt{\sum_{i=1}^{D} w(t_i)^2 \frac{Y_{ij}}{Y_i}\left(1 - \frac{Y_{ij}}{Y_i}\right)\left(\frac{Y_i - d_i}{Y_i - 1}\right) d_i}}$$

Note few forms of weighting functions $w(t_i)$:

1.$w(t_i) = 1$ (Log rank test)

2.$w(t_i) = Y_i$ (Gehan-Wilcoxon) where Y_i =Risk set in all groups at time t_i

3.$w(t_i) = \hat{S}(t_i)\frac{Y_i}{Y_{i+1}}$ where $\hat{S}(t_i) = \prod_{t_i \leq t}(1 - \frac{d_i}{Y_{i+1}})$

3.5 Dealing with crossing hazards by Renyi-Type test

Renyi-test is an equivalent of Kolmogorov-Sirnov test of goodness of fit for two samples (Uncensored,normal data),but Renyi test for (Censored data).

Here we consider two samples of survival data,size n_1 and n_2 such that n=$n_1 + n_2$

Consider times $t_1 < t_2 < ... < t_D$

$$Y_i = Y_{i1} + Y_{i2}$$

$$d_i = d_{i1} + d_{i2}$$

i=time group=2

$\forall t_i$ death times,we compute

$$Z(t_i) = \sum_{t_k=0}^{t_i} w(t_k) \left(d_{k1} - Y_{k1} \frac{d_k}{Y_k} \right)$$

Assume the Log rank case

Time	Deaths			Risk sets			$Z(t_i)$
t_i	Group1	Group2	Total	Group1	Group2	Total	
t_1	d_{11}	d_{12}	d_1	Y_{11}	Y_{12}	Y_1	$d_{11} - Y_{11}\frac{d_1}{Y_1}$
t_2	d_{21}	d_{22}	d_2	Y_{21}	Y_{22}	Y_2	$d_{11} - Y_{11}\frac{d_1}{Y_1} + d_{21} - Y_{21}\frac{d_2}{Y_2}$
.	.			.			.
.	.			.			.
.	.			.			.
t_i	d_{i1}	d_{i2}	d_i	Y_{i1}	Y_{i2}	Y_i	$\sum_{t_k=0}^{t_i} \left[d_{k1} - Y_{k1} \left(\frac{dk}{Yk} \right) \right]$

Note:$Z(t_i)$'s are computed cummulatively till t_i and not till τ because of the difference in parts of the curves due to distinct variances at different times.

We also define a variance

$$\sigma^2(\tau) = \sum_{t_k \leq t} w(t_k)^2 \frac{Y_{k1}}{Y_k} \frac{Y_{k2}}{Y_k} \left(\frac{Y_k - d_k}{Y_k - 1} \right) d_k \qquad \text{where } \tau \text{ is}$$

the largest t_k such that Y_{k1} and $Y_{k2} > 0$.

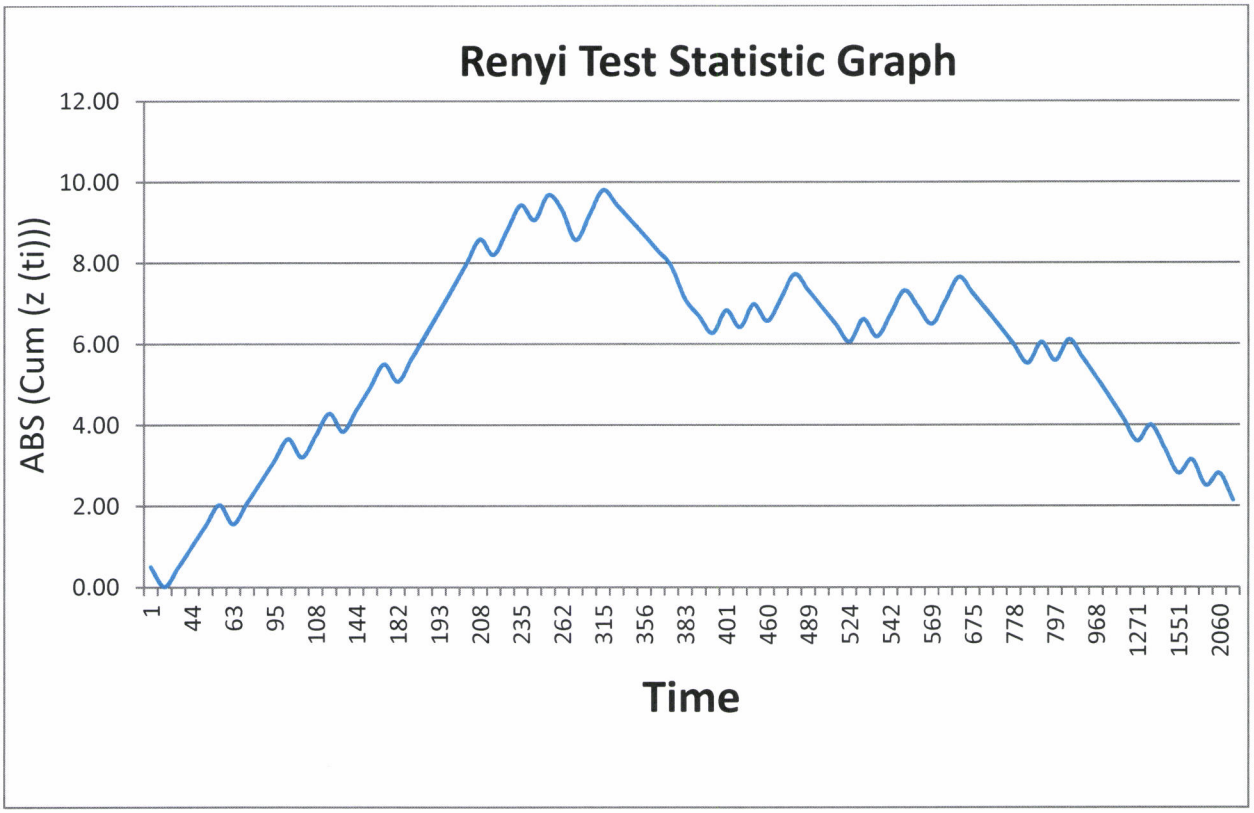
The test statistics is given by

$$Q = \frac{\sup\{|Z(t_i)|; t_i \leq \tau\}}{\sigma(\tau)}$$

The distribution of Q is approximated by the distribution of $\sup\{|B(x)|;0<x<1\}$ where B is a standard Brownian motion process.

Chapter 4

Data Analysis and Results



Shows the value of ABS (Cum (z (ti))), with the maximum at time t_i = 315 with a value of 9.80.

Data	Group	d_{1j}	d_{2j}	d_i	y_{ij}	y_{2i}	y_i	z(ti)	ABS(z(ti))	Cum(z(ti))	ABS(Cum(z(ti)))	$\sigma^2(\tau)$
1	a	1	0	1	45	45	90	0.50	0.50	0.50	0.50	0.25
17	b	0	1	1	44	45	89	-0.49	0.49	0.01	0.01	0.25
42	b	0	1	1	44	44	88	-0.50	0.50	-0.49	0.49	0.25
44	b	0	1	1	44	43	87	-0.51	0.51	-1.00	1.00	0.25
48	b	0	1	1	44	42	86	-0.51	0.51	-1.51	1.51	0.25
60	b	0	1	1	44	41	85	-0.52	0.52	-2.03	2.03	0.25
63	a	1	0	1	44	40	84	0.48	0.48	-1.55	1.55	0.25
72	b	0	1	1	43	40	83	-0.52	0.52	-2.07	2.07	0.25
74	b	0	1	1	43	39	82	-0.52	0.52	-2.60	2.60	0.25
95	b	0	1	1	43	38	81	-0.53	0.53	-3.13	3.13	0.25
103	b	0	1	1	43	37	80	-0.54	0.54	-3.66	3.66	0.25
105	a	1	0	1	43	36	79	0.46	0.46	-3.21	3.21	0.25
108	b	0	1	1	42	36	78	-0.54	0.54	-3.75	3.75	0.25
122	b	0	1	1	42	35	77	-0.55	0.55	-4.29	4.29	0.25
129	a	1	0	1	42	34	76	0.45	0.45	-3.84	3.84	0.25
144	b	0	1	1	41	34	75	-0.55	0.55	-4.39	4.39	0.25
167	b	0	1	1	41	33	74	-0.55	0.55	-4.95	4.95	0.25
170	b	0	1	1	41	32	73	-0.56	0.56	-5.51	5.51	0.25
182	a	1	0	1	41	31	72	0.43	0.43	-5.08	5.08	0.25
183	b	0	1	1	40	31	71	-0.56	0.56	-5.64	5.64	0.25
185	b	0	1	1	40	30	70	-0.57	0.57	-6.21	6.21	0.24
193	b	0	1	1	40	29	69	-0.58	0.58	-6.79	6.79	0.24
195	b	0	1	1	40	28	68	-0.59	0.59	-7.38	7.38	0.24
197	b	0	1	1	40	27	67	-0.60	0.60	-7.98	7.98	0.24
208	b	0	1	1	40	26	66	-0.61	0.61	-8.58	8.58	0.24

216	a	1	0	1	40	25	65	0.38	0.38	-8.20	8.20	0.24
234	b	0	1	1	39	25	64	-0.61	0.61	-8.81	8.81	0.24
235	b	0	1	1	39	24	63	-0.62	0.62	-9.43	9.43	0.24
250	a	1	0	1	39	23	62	0.37	0.37	-9.06	9.06	0.23
254	b	0	1	1	38	23	61	-0.62	0.62	-9.68	9.68	0.23
262	a	1	0	1	38	22	60	0.37	0.37	-9.31	9.31	0.23
301	a	2	0	2	37	22	59	0.75	0.75	-8.57	8.57	0.46
307	b	0	1	1	35	22	57	-0.61	0.61	-9.18	9.18	0.24
315	b	0	1	1	35	21	56	-0.63	0.63	-9.80	9.80	0.23
342	a	1	0	1	35	20	55	0.36	0.36	-9.44	9.44	0.23
354	a	1	0	1	34	20	54	0.37	0.37	-9.07	9.07	0.23
356	a	1	0	1	33	20	53	0.38	0.38	-8.69	8.69	0.23
358	a	1	0	1	32	20	52	0.38	0.38	-8.31	8.31	0.24
380	a	1	0	1	31	20	51	0.39	0.39	-7.92	7.92	0.24
383	a	2	0	2	30	20	50	0.80	0.80	-7.12	7.12	0.47
388	a	1	0	1	28	20	48	0.42	0.42	-6.70	6.70	0.24
394	a	1	0	1	27	20	47	0.43	0.43	-6.27	6.27	0.24
401	b	0	1	1	26	20	46	-0.57	0.57	-6.84	6.84	0.25
408	a	1	0	1	26	19	45	0.42	0.42	-6.42	6.42	0.24
445	b	0	1	1	25	19	44	-0.57	0.57	-6.99	6.99	0.25
460	a	1	0	1	25	18	43	0.42	0.42	-6.57	6.57	0.24
464	b	0	1	1	24	18	42	-0.57	0.57	-7.14	7.14	0.24
484	b	0	1	1	24	17	41	-0.59	0.59	-7.72	7.72	0.24
489	a	1	0	1	24	16	40	0.40	0.40	-7.32	7.32	0.24
499	a	1	0	1	23	16	39	0.41	0.41	-6.91	6.91	0.24
523	a	1	0	1	22	16	38	0.42	0.42	-6.49	6.49	0.24
524	a	1	0	1	21	16	37	0.43	0.43	-6.06	6.06	0.25
528	b	0	1	1	20	16	36	-0.56	0.56	-6.62	6.62	0.25
535	a	1	0	1	20	15	35	0.43	0.43	-6.19	6.19	0.24
542	b	0	1	1	19	15	34	-0.56	0.56	-6.75	6.75	0.25
547	b	0	1	1	19	14	33	-0.58	0.58	-7.32	7.32	0.24
562	a	1	0	1	19	13	32	0.41	0.41	-6.92	6.92	0.24
569	a	1	0	1	18	13	31	0.42	0.42	-6.50	6.50	0.24
577	b	0	1	1	17	13	30	-	0.57	-7.06	7.06	0.25

								0.57				
580	b	0	1	1	17	12	29	-0.59	0.59	-7.65	7.65	0.24
675	a	1	0	1	17	11	28	0.39	0.39	-7.26	7.26	0.24
676	a	1	0	1	16	11	27	0.41	0.41	-6.85	6.85	0.24
748	a	1	0	1	15	11	26	0.42	0.42	-6.43	6.43	0.24
778	a	1	0	1	14	11	25	0.44	0.44	-5.99	5.99	0.25
786	a	1	0	1	13	11	24	0.46	0.46	-5.53	5.53	0.25
795	b	0	1	1	12	11	23	-0.52	0.52	-6.05	6.05	0.25
797	a	1	0	1	12	10	22	0.45	0.45	-5.59	5.59	0.25
855	b	0	1	1	11	10	21	-0.52	0.52	-6.12	6.12	0.25
955	a	1	0	1	11	9	20	0.45	0.45	-5.67	5.67	0.25
968	a	1	0	1	10	9	19	0.47	0.47	-5.19	5.19	0.25
1000	a	1	0	1	9	9	18	0.50	0.50	-4.69	4.69	0.25
1245	a	1	0	1	8	9	17	0.53	0.53	-4.17	4.17	0.25
1271	a	1	0	1	7	9	16	0.56	0.56	-3.60	3.60	0.25
1366	b	0	1	1	6	9	15	-0.40	0.40	-4.00	4.00	0.24
1420	a	1	0	1	6	8	14	0.57	0.57	-3.43	3.43	0.24
1551	a	1	0	1	5	8	13	0.62	0.62	-2.82	2.82	0.24
1577	b	0	1	1	4	8	12	-0.33	0.33	-3.15	3.15	0.22
1694	a	1	0	1	4	7	11	0.64	0.64	-2.51	2.51	0.23
2060	b	0	1	1	3	7	10	-0.30	0.30	-2.81	2.81	0.21
2363	a	1	0	1	3	6	9	0.67	0.67	-2.15	2.15	0.22
2412*	b	0	0	0	2	6	8	0.00	0.00	-2.15	2.15	0.00
2486*	b	0	0	0	2	6	8	0.00	0.00	-2.15	2.15	0.00
2754*	a	0	0	0	2	6	8	0.00	0.00	-2.15	2.15	0.00
2796*	b	0	0	0	2	6	8	0.00	0.00	-2.15	2.15	0.00
2802*	b	0	0	0	2	6	8	0.00	0.00	-2.15	2.15	0.00
2934*	b	0	0	0	2	6	8	0.00	0.00	-2.15	2.15	0.00
2950*	a	0	0	0	2	6	8	0.00	0.00	-2.15	2.15	0.00
2988*	b	0	0	0	2	6	8	0.00	0.00	-2.15	2.15	0.00
Total											Sup=9.80	19.86

Q=ABS (Cum (z (ti))) /σ (τ) therefore Q=9.80 /4.46=2.20

So we find that the *p*-value of this test is 0.053 so the null hypothesis of no difference in survival rates between the two treatment groups is not rejected at the 5% level.

Performing a Log Rank Test

	N	Observed	Expected	(O-E) ^2/E	(O-E)^2/V
Group=1	45	43	45.1	0.102	0.232
Group=2	45	39	36.9	0.125	0.232

Chisq= 0.2 on 1 degrees of freedom, p= 0.63

Given a p-value=0.63 implies that we fail to reject a null hypothesis that there is no difference in survival rates between those two groups at 5% level of significance.

Comparison:

Renyi test p-value is 0.053 and Log rank test p-value is 0.63 evident that both of them show no difference in survival rates but Renyi test in marginally insignificant, its leaner compared to Log rank test. Therefore, Renyi test is a better when hazards are crossing indicative of non-proportionality.

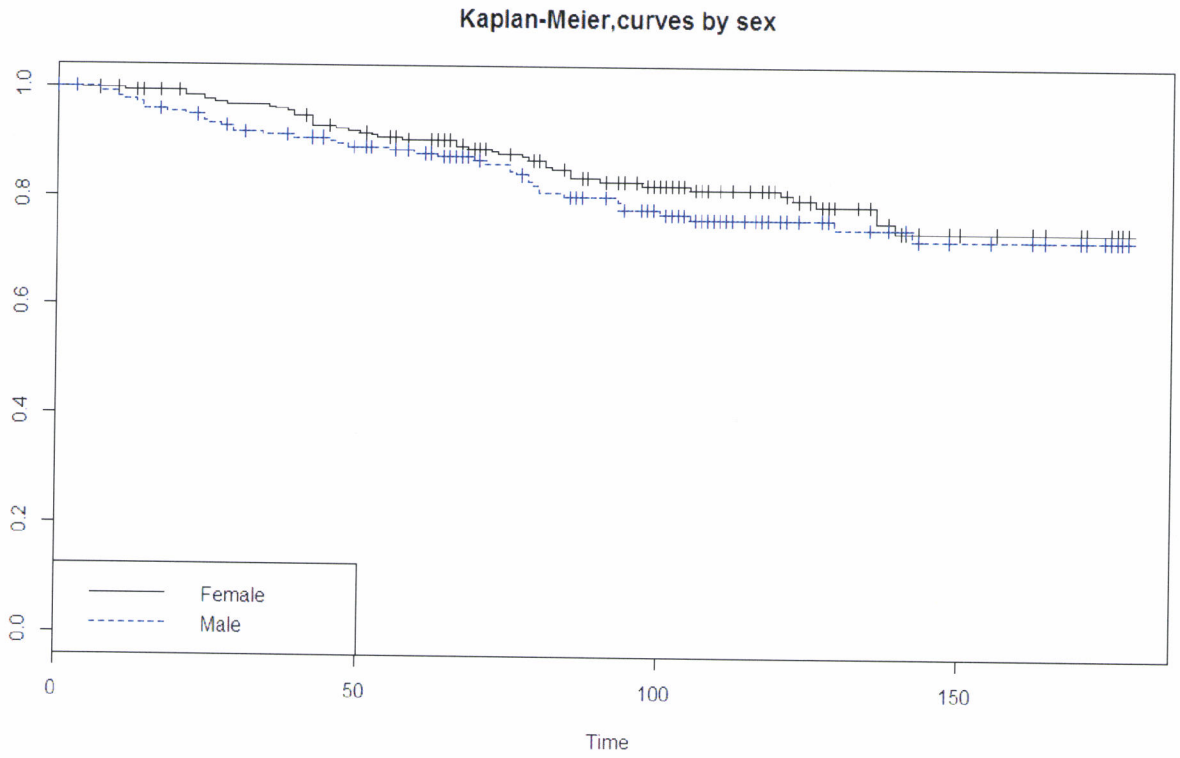


Figure 1.Kaplan Meier Curve for sex

4.3 Cox PH Results

Table 1.Shows a model with all the covariates.

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
adeno_prevyes	0.048204	1.049385	0.224330	0.215	0.829861
corona_prev	-0.496691	0.608541	0.226615	-2.192	0.028395 *
ageyrs	-0.031532	0.968960	0.009213	-3.422	0.000621 ***
hhsize	0.046236	1.047321	0.010891	4.245	2.18e-05 ***
sexmale	0.275478	1.317160	0.216188	1.274	0.202575
locationRoka	-0.188063	0.828563	0.518127	-0.363	0.716629
	exp(coef)	exp(-coef)	lower .95	upper .95	
adeno_prevyes	1.0494	0.9529	0.6761	1.6289	
corona_prev	0.6085	1.6433	0.3903	0.9488	
ageyrs	0.9690	1.0320	0.9516	0.9866	
hhsize	1.0473	0.9548	1.0252	1.0699	
sexmale	1.3172	0.7592	0.8622	2.0121	
locationRoka	0.8286	1.2069	0.3001	2.2875	
Concordance=0.665	(se = 0.032)				
Rsquare= 0.062(max = 0.872)					
Likelihood ratio test= 31.43,df 6	p=2.102e-05				
Wald test = 29.64 on 6 df	p=4.596e-05				
Score (logrank) test = 30.19,df 6	p=3.618e-05				

Table 2. Diagnosis of Cox Proportional Hazard to check for proportionality.

	rho	chisq	p
adeno_prevyes	0.0409	0.1730	0.67746
corona_prev	-0.3246	8.0875	0.00446
hhsize	-0.0388	0.2069	0.64920
sexmale	-0.1293	1.4721	0.22501
locationRoka	0.0338	0.1041	0.74693
ageyrs	0.0143	0.0242	0.87639
GLOBAL	NA	11.3015	0.07949

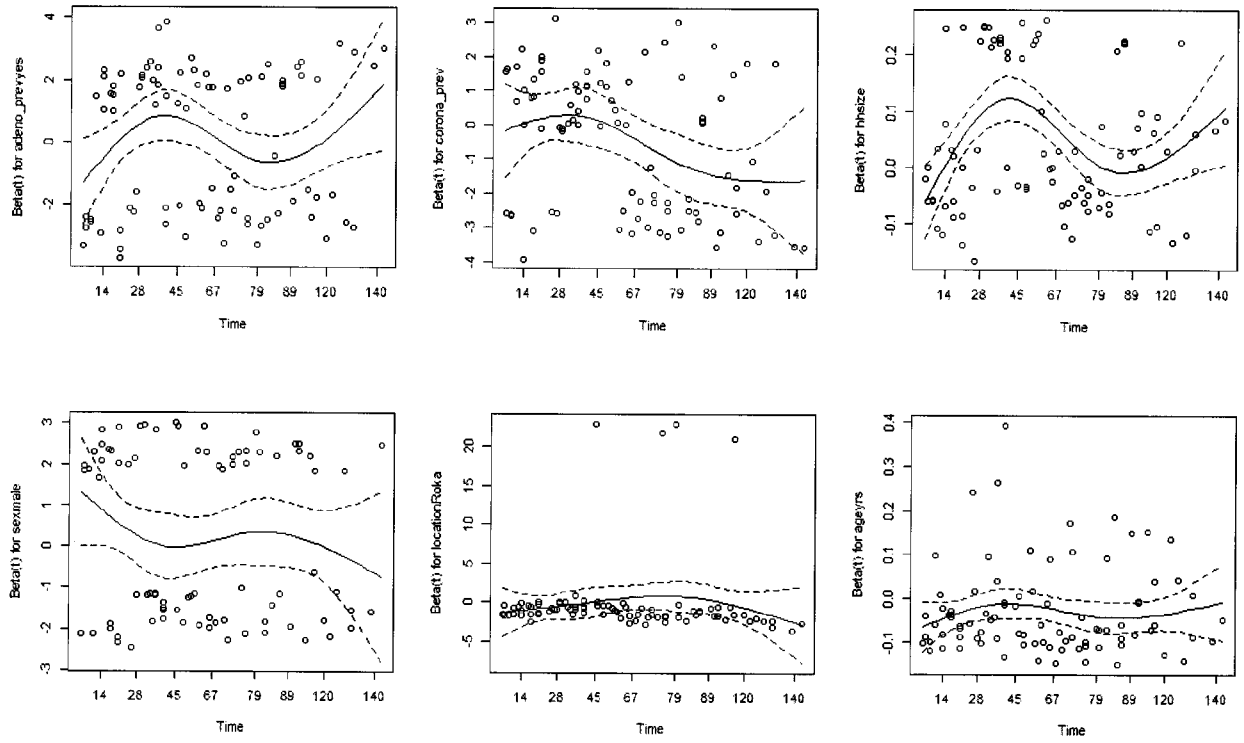


Figure 2. Plots of scaled Schoenfeld residuals against transformed time for each covariate in a model fit to the Survival data.

Table 3.Clearly shows that all the covariates are proportional.

	rho	chisq	p
adeno_prevyes	0.0381	0.1515	0.6971
corona_prev	-0.0983	0.9669	0.3255
hhsize	-0.0478	0.3213	0.5708
sexmale	-0.1392	1.7354	0.1877
locationRoka	0.0316	0.0912	0.7626
ageyrs	0.1362	1.7623	0.1843
corona_prev:ageyrs	-0.1402	2.3691	0.1238
GLOBAL	NA	13.8328	0.0542

Table 4.Summary of all covariates with interaction of corona_prev and ageyrs.

	coef	exp(coef)	Se(coef)	z	Pr(>|z|)
adeno_prevyes	0.05155	1.05291	0.22386	0.230	0.81787
corona_prev	-0.61789	0.53908	0.31395	-1.968	0.04905*
ageyrs	-0.03626	0.96439	0.01282	-2.828	0.00469**
hhsize	0.04666	1.04776	0.01097	4.252	2.12e-05***
sexmale	0.28421	1.3287	0.21691	1.310	0.19010
locationRoka	-0.18520	0.83094	0.51855	-0.357	0.72098
corona_prev:ageyrs	0.01005	1.01010	0.01810	0.555	0.57862
	exp(coef)	exp(-coef)	lower .95	upper .95	
adeno_prevyes	1.0529	0.9498	0.6789	1.6328	
corona_prev	0.5391	1.8550	0.2914	0.9974	
ageyrs	0.9644	1.0369	0.9405	0.9889	
hhsize	1.0478	0.9544	1.0255	1.0705	
sexmale	1.3287	0.7526	0.8686	2.0326	
locationRoka	0.8309	1.2035	0.3007	2.2959	
corona_prev:ageyrs	1.0101	0.9900	0.9749	1.0466	
Concordance=0.668	(se=0.032)				
Likelihood ratio test=31.73,df 7	P=4.552e-05				
Wald test=30.25,df 7	P=8.554e-05				
Score (logrank)test=31.17,df 7	P=5.772e-05				

Table 5.Output with the summary of all covariates with the stratification of corona_prev.

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
adeno_prevyes	0.007314	1.007341	0.224278	0.033	0.973984
ageyrs	-0.031801	0.968699	0.009201	-3.456	0.000548 ***
hhsize	0.044643	1.045655	0.010698	4.173	3.01e-05 ***
sexmale	0.274537	1.315921	0.216284	1.269	0.204322
locationRoka	-0.243524	0.783861	0.520596	-0.468	0.639942
	exp(coef)	exp(-coef)	lower .95	upper .95	
adeno_prevyes	1.0073	0.9927	0.6490	1.5634	
ageyrs	0.9687	1.0323	0.9514	0.9863	
hhsize	1.0457	0.9563	1.0240	1.0678	
sexmale	1.3159	0.7599	0.8612	2.0106	
locationRoka	0.7839	1.2757	0.2826	2.1746	
Concordance=0.67	(se = 0.044)				
Likelihood ratio test= 30.21,df 5	p=1.344e-05				
Wald test = 28.43 on 5 df	p=2.998e-05				
Score (logrank) test = 29.15,df5	p=2.164e-05				

Chapter 5

Conclusions and Recommendations

Variables Description;

Response variable;timetoinfect_rsva2 implies time to develop respiratory syncli-tial virus type A.

Independent covariates;adeno_prev implies previous infection with adeno virus.

corona_prev implies previous infection with corona virus.

ageyrs implies age in years.

hhsize implies household size or those living in a particular household.

sex(male/female).

Location(Roka/Matasango) implies residential area whether Roka or Matasango.

5.1 Interpretation of the outputs

Figure 1 Kaplan Meier curve for sex covariate shows that females have a higher survival rate as compared to males in the actual data.

Table 1 corona_prev,ageyrs and hhsize are statistically significant at 5 % level of significance in a model with all the covariates implying that they help us explain a scenario when one has our outcome variable(timetoinfect_rsva2).

Table 2 is the output for verifying proportionality,therefore strong evidence of non-proportional hazards for corona_prev variable while GLOBAL model for the test as a whole is not quite statistically significant.

Figure 2 Plots of scaled Schoenfeld residuals against transformed time for each covariate, with the broken lines representing a $\pm$ 2-standard-error band around the fit.A systematic departure from a horizontal line like Corona_Prev is indicative of non-proportional hazards.

Table 3 is the output after interacting corona_prev with a time dependent covariate (ageyrs)

which proportionalises it (corona_prev) as depicted by p-value 0.3255 at 5 % level of signicance.GLOBAL is almost at the threshold though not statistically significant.

Table 4 is the output with the summary of all covariates with the interaction term of corona_prev with ageyrs.

-Those who had adeno previously were 1.0529 times more likely to have the response variable (timetoinfect_rsva2) compared to those who did not have.

-Those who had corona previously were 0.5391 less likely to contract our varialbe of interest(timetoinfect_rsva2) compared to those who did not.

-As hhsize increases by one,those living in that household are 1.0478 times likely to have virus (timetoinfect_rsva2) or 4.78% likely to than not.

-As one increases age by one, he/she becomes less likely to develop the variable of interest(timetoinfect_rsva2) by 4%.

-Those who resided in Roka were 17% less likely to contract response variable compared to those who reside in Matasango.

-Male are more likely by 32.87% to develop (timetoinfect_rsva2) as compared to female who are the reference category.

Table 5 is the output of a stratified Cox model with corona_prev blocked for it is not proportional as illustrated from Table 2.

-Ageyrs and hhsize are statistically significant at 5% level.

5.2 Recommendation

Ignoring non-proportional hazards in analysis of survival can lead us to incorrect results, so one should first check the proportional hazards assumption. In medical

research, survival analysis is erupting as an area with many developments and I recommend that readers do more research work in this area of Survival analysis.

Chapter 6

Appendix

Brownian motion is the simplest stochastic process on a continuous domain, and it is a limit of both simpler (random walk) and more complicated stochastic processes. This universality is closely related to the universality of the normal distribution. Brownian motion is related to the random walk problem and it is generic in the sense that many different stochastic processes reduce to Brownian motion in suitable limits.

R Codes:

```
data=c(1,1,2,2,3,4,4,5,5,8,8,8,8,11,11,12,12,15,17,22,23)
length(data)
status=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1)
```

```
length(data)
surv.fn=Surv(data,status==1)
surv.fn
surv.fn1=survfit(Surv(data,status==1)~1)
surv.fn1
plot(surv.fn1)
/***Codes for chemotherapy plus radiotherapy curve ***/
chemplRad=c(17,42,44,48,60,72,74,95,103,108,122,144,167,170,183,185,193,195,197,208,234,235,
254,307,315,401,445,464,484,528,542,547,577,580,795,855,1366,1577,2060,2412,2486,2796,2802,
2934,2988)
h=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0)
surv.fn=Surv(chemplRad,h==1)
surv.fn=survfit(Surv(chemplRad,h==1)~1)
plot(surv.fn)
/***Codes for chemotherapy curve ***/
chem=c(1,63,105,129,182,216,250,262,301,301,342,354,356,358,380,383,383,383,394,408,460,489,
499,523,524,535,562,569,675,676,748,778,786,797,955,968,1000,1245,1271,1420,1551,1694,2363,
2754,2950)
h1=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0)
```

```
surv.fn=Surv(chem,h1==1)

surv.fn=survfit(Surv(chem,h1==1)~1)

plot(surv.fn)

/***Codes for crossing survival curves ***/

data=c(17,42,44,48,60,72,74,95,103,108,122,144,167,170,183,185,193,195,197,208,234,235,254,307,
315,401,445,464,484,528,542,547,577,580,795,855,1366,1577,2060,2412,2486,2796,2802,2934,2988,
1,63,105,129,182,216,250,262,301,301,342,354,356,358,380,383,383,383,394,408,460,489,499,523,
524,535,562,569,675,676,748,778,786,797,955,968,1000,1245,1271,1420,1551,1694,2363,2754,2950)

status=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,1,1,1,1
,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0)

group=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,
2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2)

my_data=data.frame(data,status,group)

my_data

fit=survfit(Surv(data,status==1)~group)

plot(fit,xlab='time',ylab='probability',main='Crossing survivor curves')

plot(fit,lty=2:3,col =c('blue','black'), xlab='time',ylab='probability',main='Crossing
survivor curves')

legend("top", legend = c("chemoplsRad", "chemo"),col=c('blue','black'),
```

```
    text.col = c('blue','black'))

    cox_all=coxph(formula = Surv(timetoinfect_rsva2, everrsva == "yes") ~adeno_prev
+ corona_prev
    + ageyrs+ hhsize + sex + location)
    qqt2=cox.zph(cox_all)
    qqt2

    coll1=coxph(formula=Surv(timetoinfect_rsva2,everrsva=="yes")~adeno_prev+
    corona_prev*ageyrs+hhsize+sex+location)
    t=cox.zph(coll1)
    t
    coxph(formula = Surv(timetoinfect_rsva2, everrsva == "yes") ~adeno_prev +
corona_prev * ageyrs
    + hhsize + sex + location)
    coxph(formula = Surv(timetoinfect_rsva2, everrsva == "yes") ~adeno_prev +
strata(corona_prev) +
    hhsize + sex + location + ageyrs)
```

REFERENCES

[1] Michael Schemper,(1992), "Cox Analysis of Survival Data with Non-Proportional Hazard Functions",Source: Journal of the Royal Statistical Society. Series D (The Statistician), Vol. 41, No. 4, pp. 455-465,Published by: Wiley for the Royal Statistical Society,Stable URL: http://www.jstor.org/stable/2349009 ,.Accessed: 11/05/2013 05:41

[2] Janet M. Box-Steffensmeier, Dan Reiter and Christopher Zorn,(Feb., 2003),"Nonproportional Hazards and Event History Analysis in International Relations", Source: The Journal of Conflict Resolution, Vol. 47, No. 1, Building a Science of WorldPolitics , pp. 33-53,Published by: Sage Publications, Inc.Stable URL: http://www.jstor.org/stable/3176181

[3] Ming-Yen Cheng,Peihua Qiu,Xianming Tan •and Dongsheng Tu,(2009),"Confidence intervals for the first crossing point of two hazard functions",Lifetime Data Anal 15:441–454 DOI 10.1007/s10985-009-9132-

6,Received: 3 February 2009 / Accepted: 14 October 2009 / Published online: 1 November 2009,© Springer Science+Business Media, LLC 2009

[4] Sheila M. Gore, Stuart J. Pocock and Gillian R. Kerr,(1984),"Regression Models and Non-Proportional Hazards in the Analysis of Breast Cancer Survival", Source: Journal of the Royal Statistical Society. Series C (Applied Statistics), Vol. 33, No. 2, pp. 176-195,Published by: Wiley for the Royal Statistical SocietyStable URL: http://www.jstor.org/stable/2347444 .Accessed: 11/05/2013 08:47

[5] Jiajia Zhang and Yingwei Peng,(2009 October 15),"Crossing Hazard Functions in Common Survival Models",Published in final edited form as:Stat Probab Lett. ; 79(20): 2124–2130. doi:10.1016/j.spl.2009.07.002.

[6] Donald M. Stablein, Walter H. Carter Jr and Joel W. Novak,(June 1981),"Analysis of survival data with nonproportional hazard functions",Controlled Clinical Trials, Volume 2, Issue 2, , Pages 149-159

[7] Liu K, Qiu P and Sheng J,(15 March 2006),"Comparing two crossing hazard rates by Cox proportional hazards modelling",Statist. Med. 2007; 26:375–391,Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sim.2544

[8] Gilbert,MacKenzie ,Do Ha and II,(2007),"Modelling Survival Data with Crossing Hazards",Publisher: IWSM, Publication type: info:eu-repo/semantics/conferenceObject,Stat Med. 2007 Jan 30;26(2):375-91.URI: http://hdl.handle.net/10344/2787

[9] Matthew Davis, M.S. and Sharon X. Xie, Ph.D,"Caution: Hazards Crossing! Using the Renyi Test Statistic in Survival Analysis",PharmaSUG2011 - Paper SP06

[10] Tomasz Jurkiewicz and Ewa Wycinka,(2011),"Significance tests of differences between two crossing survival curves for small samples",ACTA UNIVERSITATIS LODZIENSIS FOLIA OECONONMICA 255

[11] John.P.Klein,Melvin.L.Moeschberger,(2003)"Statistics for Biology and Health", Series Editors,K. Dietz, M. Gail, K. Krickeberg, J. Samet, A. Tsiatis

[12] Dr. Lutz Edler3,Prof. Dr. Christos P. Kitsos,Prof. Vilijandas B. Bagdonavičius and Dr. Mikhail S. Nikuline,(17 MAR 2006)."Analysis of Survival Data with Non-proportional Hazards and Crossings of Survival Functions",Published Online: DOI: 10.1002/0470857706.ch12;Copyright © 2005 John Wiley & Sons, Ltd,http://www.ats.ucla.edu/stat/examples/asa/test_proportionality.htm

[13] Nihal Ata and M.Tekin Sozer,(28:09:2007),"Cox regression models with nonproportional hazards applied to Lung cancer survival data",Hacettepe Journal of Mathematics and Statistics,Volume 36(2) (2007),157-167