2017

Smooth tests of goodness of FIT for hazard Functions: an application to HIV retention data

Collins Ojwang' Odhiambo
Strathmore Institute of Mathematical Sciences
Strathmore University

SMOOTH TESTS OF GOODNESS OF FIT FOR HAZARD FUNCTIONS-AN APPLICATION TO HIV RETENTION DATA

Submitted in total fulfillment of the requirements for the Degree of Doctor of Philosophy in Applied Statistics at Strathmore University

Collins Ojwang' Odhiambo

May 2017

I certify that I have read this PhD thesis and that, in my opinion, it is fully adequate in scope and quality as a thesis for the degree of Doctor of Philosophy in Applied Statistics.

_____

(Professor Bernard Omolo) Principal Supervisor

I certify that I have read this PhD thesis and that, in my opinion, it is fully adequate in scope and quality as a thesis for the degree of Doctor of Philosophy in Applied Statistics.

_____

(Professor John Odhiambo)

School of Graduate Studies

_____

Declaration

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

Approval

The thesis of Collins Ojwang' Odhiambo was reviewed and approved by the following:

Professor Bernard Omolo
Principal Supervisor,
University of South Carolina-Upstate
Professor and Chair-Division of Mathematics and Computer Science.

Professor John Odhiambo,
Strathmore University,
Vice-chancellor -Strathmore University,

Ferdinard Othieno,
Strathmore University,

Dean -Strathmore Institute of Mathematical Sciences

Professor Ruth Kiraka,

Strathmore University,

Dean -School of Graduate Studies

Dedication

To my wife Sara W. Ojwang

Acknowledgments

Firstly, I would like to express my sincere gratitude to my principal supervisor Prof. Bernard Omolo for the continuous support of my Ph.D study, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study.

I would like to also thank co-supervisor Prof.John Odhiambo, for not only his insightful comments and encouragement, but also for the hard question which challenged me to widen my research from various perspectives.

My sincere thanks also goes to Prof. Vitalis Onyango-Otieno, and the entire Institute of Mathematical Sciences who provided me with the opportunity to join Strathmore University and work for my PhD.

Abstract

In this study, we apply the methodology of smooth tests of goodness-of-fit to hazard functions. The smooth test formulation applied here is an extension of Neyman's smooth test and is obtained by nesting the null hypothesis in a larger class of probability and hazard rate functions. The study revisits Neyman's smooth tests and its data-driven versions in the context of classical probability and survival analysis. Though several authors have theoretically looked at the development of Neyman's smooth tests, the main contribution of this study is modelling loss to follow-up in HIV retention. To the best of our knowledge, this issue has not been given its due share of coverage in the literature. We extend methods proposed by Rayner et al. (2009); Pena (1998a,b) and Kraus (2007a), to an HIV retention setting. The applications dealt with in this thesis also covers performance of other goodness-of-fit (GOF) approaches and compares them with that of smooth tests.

Three main methodological approaches are covered under the research methodology. Part I revisits smooth tests for various probability distributions and applies the test when assessing the fit for the two-parameter Weibull distribution to an HIV retention data under the complete and uncensored data scenario. Part II looks at the application of smooth test to Cox proportional hazards models. We assess the proportionality assumption in the two-sample problem in cancer survival studies. Part III covers recurrent event situation. We fit Block, Borges and Savits (BBS) minimal repair model to loss to follow-up (LFTU) data and assesses the performance of the smooth test in terms of power.

More specifically, Chapter 1 deals with background of GOF in classical probability and

survival distributions. The motivation for the study, overview of the smooth test of GOF and comparison with other GOF tests is also covered in this chapter. In Chapter 2, we provide a review of the literature. Chapter 3 details research methodology. We present analysis and results in Chapter 4. Chapter 5 discusses important findings using simulated and real data in the context of HIV retention and overall survival in cancer studies. Chapter 6 covers summary of the thesis, the limitations of the study and possible extensions of the smooth GOF to discrete probability cases.

All computations have been implemented in R and the scripts are briefly described in Appendix A. The chapters are self-contained in order to achieve our objective of covering the applications smooth tests of goodness-of-fit approach from distributions with non-censored data to extensions in recurrent events.

A major limitation of this study, is that, in clinical studies, particularly involving LTFU data, incomplete data is frequently encountered. Analysis of severity of data incompleteness is a subject of future research.

Keywords: Smooth Tests of Goodness of Fit; Probability Distribution; Goodness-of-Fit; Baseline Hazard Function; Recurrent Events Models; Simulations; Orthonormal Functions; Loss to Follow-up; Proportionality Assumption; HIV retention; Cancer

Abbreviation

AD Anderson-Darling

AG Andersen and Gill

AIDS Acquired Immune Deficiency Syndrome

ART Antiretroviral Therapy

BBS Block, Borges and Savits minimal repair model

BFGS Broyden-Fletcher-Goldfarb-Shanno algorithm

BP Brown-Proschan model

BPl Brown-Proschan model with log-linear failure intensity

BPp Brown-Proschan model with power-law failure intensity

cdf Cumulative Distribution Function

CPHM Cox Proportional Hazards Model

CVM Cramér-von Misses

DHR Decreasing Hazard Rate

DFR Decreasing Failure Rate

GOF Goodness-of-fit

GRA Geometric Reduction of Age

GT-UR Gap Time - Unrestricted

HAART Highly Active Antiretroviral Therapy

HIV Human Immunodeficiency Virus

HPP Homogeneous Poisson Process

IHR Increasing Hazard Rate

iid Independent and Identically Distributed

IFR Increasing Failure Rate

KS Kolmogorov-Smirnov

LLP Log-linear Process

LSE Least Squared Estimator

LTFU Lost To Follow-Up

LWA Lee, Wei and Amato

ME Moment Estimator

ML Maximum Likelihood

MLE Maximum Likelihood Estimator

NHPP Non Homogeneous Poisson Process

pdf Probability Density Function

PIT Probability Integral Transformation

PLMLE Partial Likelihood Maximum Likelihood Estimator

PLP Power-Law Process

PM Preventive Maintenance

PWP-CP Prentice, Williams and Peterson-Counting Process

PWP-GT Prentice, Williams and Peterson-Gap Time

RP Renewal Process

TT-R Total Time - Restricted

UBT Upside-down Bathtub shaped hazard rate

WLW Wei, Lin and Weissfeld

WHO World Health Organization

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

Applications of probability and survival models have increased rapidly in the last few decades. For specific practical examples in application of exponential distribution and survival distribution, Weibull distribution and generalized extreme value distribution, Gamma distribution, log-normal distribution, log-logistic distribution, Gompertz distribution and Gompertz-type hazard models, Hypergeometric distribution and other distributions see Liu (2012); Kalbfleisch and Prentice (2011) and Rayner et al. (2009). In particular, fitting both parametric and non-parametric models to a given data set and then using the results to make well informed decisions have gained popularity. There are many advantages of fitting probability and survival models to a given data set i.e survival models, enables prediction of the reliability of a component or a system. On the other hand, goodness-of-fit (GOF) tests measure compatibility of a random sample with a theoretical probability distribution function or hazard function. Constructing valid tests for statistical hypotheses is a critical statistical problem that has been studied for many years. The common approach to constructing GOF test statistics involve measuring the "distance" between data and the corresponding empirical probability distributions and hazard functions. Empirical GOF tests (i.e. Cramér-von

Mises, Kolmogorov-Smirnov and Anderson-Darling tests) are classical examples that use this approach. More generally, these distance-based tests, as well as graphical tests based on confidence intervals, usually belong to this class of empirical GOF (Langovoy, 2007).

GOF tests are based on test statistics that measure the distance between two distributions: a theoretical one, which characterizes the tested hypothesis H_0, and an empirical one computed from the given data set. The null hypothesis in this case is rejected when the test statistic is too large. The critical region is the set of values of the statistic for which H_0 is rejected. If the observed value of statistic belongs to the critical region, the conclusion of the test is the rejection of H_0. The tests can either be one-sided or two-sided. Details on the various GOF procedures have been covered extensively by Huber-Carol et al. (2012); Krit (2014) and Andersen (1982) among others.

Without GOF tests, one cannot objectively validate the fitness of a probability distribution or hazard function. This can lead to wrong decision-making if an invalid distribution is used. The question now is: what are best GOF tests and if the smooth test is one of the best tests then how is it applied in real data settings when assessing baseline hazard functions? A test is said to be smooth if the null hypothesis is embedded to form a class of smooth alternatives. This is achieved by nesting the null distribution or hazard rate function to form a larger class of alternatives (Neyman, 1937; Rayner et al., 2009; Pena, 1998b,a; Kraus, 2007a).

1.2 Motivation

1.2.1 Analysis of LTFU in HIV Retention

Patients receiving antiretroviral therapy (ART) are sometimes lost to follow-up (LTFU), which may result in discontinuation of treatment, drug toxicity and treatment failure due to poor adherence and drug resistance (Rachlis et al., 2014). LTFU accounts for an increased

risk of death of up to 40% for ART patients in sub-Saharan Africa (Berheto et al., 2014). Studies have shown that LTFU has negative impact on immunological benefit of ART and increases AIDS-related morbidity, mortality, and hospitalizations (Berheto et al., 2014). Individuals who miss visits in the first year of treatment have a higher mortality rate (Rachlis et al., 2014). Asiimwe et al. (2016) and Rasschaert et al. (2012) showed that retention of patients who are on ART treatment remains stable after 12 months of ART initiation, with LTFU being the main cause of attrition. Previous studies has also illustrated associations between frequent LTFU and more severe opportunistic illnesses (Haddow et al., 2003). Analysis of LTFU is therefore important and has been used in HIV care to monitor and improve programme effectiveness, using patient retention as a measure of quality of care (Sengayi et al., 2013).

The main objective in the analysis of LTFU data is to check retention of patients in care. In an HIV program, this is considered an important determinant of successful ART long-term outcomes. Patients who experience LFTU eventually get enrolled in other facilities with different regimen combinations, which are likely to compromise their immune system. Retaining patients for long allows provision of long term Highly Active Antiretroviral Therapy (HAART), tracking World Health Organization (WHO) staging, tracking immunosuppression profiles and evaluation of emergence of medication toxicities. More innovation is therefore required for further ART scale-up and improve retention in care. The study looks at the application of smooth tests of GOF to probability and survival models. The HIV retention data used in this study is a typical primary data set that has not been studied or published anywhere.

1.2.2 Two-Sample Problem in Cancer Survival Studies

Despite decades of research in cancer, the overall prognosis for cancer, recurrences and survival rates are still attracting a huge research interest. There are over 200 different types

of cancer known today (Siegel et al., 2015). Since these numbers are likely to grow, cancer research will particularly, be important in the fight against cancer. Most clinical trials are highly specialized with respect to the type of cancer and are beneficial to patients through advancing technologies and cancer treatment protocols. According to mortality data by the National Center for Health Statistics (NCHS) (Siegel et al., 2015), between 1930 to 2012, cancer was a major public health concern and was the second leading cause of death in the United States. Prostate, lung and bronchus, and colorectal cancers account for 44% of all cases in men, with prostate cancer alone accounting for 20% of new diagnoses. For women, the three most commonly diagnosed cancers are breast, lung and bronchus, and colorectum, representing 50% of all cases. Breast cancer alone is expected to account for 29% all new cancer diagnoses in women. NCHS estimates 1,600 deaths per day as a result of cancer in 2016. These four types of cancers account for 46% of all cancer deaths with more than one-quarter (27%) due to lung cancer. The largest geographic variation in cancer occurrence by far is for lung cancer, reflecting the large historical and continuing differences in smoking prevalence among states. Cancer in adolescents (aged 15 to 19 years) differ somewhat from those in children in terms of type and distribution. With these variations in mind, this study does not aim to provide an exhaustive performance of smooth tests for proportionality for all types of cancer in the two-sample problem, but instead it aims to statistically compare its performance in selected eight different practical settings. The analysis is focused on the most common and prevalent cancers. Our goal is to provide an overview of the performance of smooth tests to help in validation. We hope that the issues and features we comment on will result in higher overall standards and quality of oncological research by the survival analysis community, and limit the risk of using invalid models. This is specifically important when analysing overall survival in cancer studies. Authors who have analysed the eight different data sets (see chapter 4 section 4.2) did not validate the proportionality assumption when comparing the two groups.

1.3 Pearson's χ^2 test

The Karl Pearson chi-square procedure has been used over the years to compare observed data with expected data obtained according to a specific hypothesis. If $O_1, O_2, \cdots, O_n$ denote observations that are assumed to come from n non-overlapping classes that are expected to contain $E_1, E_2, \cdots, E_n$ observations, then the Pearson's χ^2 test statistic is defined broadly by Rayner et al. (2015) as

$$\chi_p^2 = \sum_{i=1}^{n} \frac{(0_i - E_i)^2}{E_i}. \tag{1.1}$$

If this is larger than $100\alpha\%$ point of the χ^2_{n-1} distribution, then the hypothesized expectation is rejected at the $100\alpha\%$ level of significance.

Karl Pearson first introduced a criterion in Pearson (1992) to examine whether the observed data support a given specification. He called it the chi-square goodness of-fit test, which motivated research in testing of hypotheses and estimation of unknown parameters. This eventually led to the development of statistics as a separate discipline. Pearson (1992) assumed that parameters of the probability model giving rise to the cell expectations were known, and showed that the asymptotic distribution of the statistic in equation (1.1) is the χ^2 distribution. There is a lot of literature on how the cells may best be constructed. (See McHugh (2013), Sürücü (2008), Rayner et al. (2009), Miller Jr and Quesenberry (1979) among others). The recommendation is that for the approximation to be adequate, each cell expectation should be at least five and this have been covered by multiple authors. The most critical assumption of the test involves estimating parameters. Pearson (1992) wrongly stated that estimating parameters makes no difference to the asymptotic null distribution of the test statistic. Furthermore, several authors have shown that if asymptotic distribution is estimated by maximum likelihood method using the ungrouped data then the Pearson's statistic no longer have an asymptotic chi-squared distribution. See McHugh

(2013), Sürücü (2008) and Rayner et al. (2009).

1.3.1 Deficiencies of Pearson χ^2 Test

Pearson's χ^2 test is essentially applicable for testing discrete data when there is no parameter to be estimated. According to Barton (1956) and Neyman (1937), the smooth test was developed to overcome deficiencies in Pearson χ^2 test. Further discussions on deficiencies of Karl Pearson χ^2 test have been described by Stigler et al. (2008).

1. With right-censored data, determining the exact values of observation O_j is not possible.

2. There is need to estimate O_j using product limit estimators which is not provided for in the Pearson χ^2 test.

3. It is difficult to examine the power or optimality properties of resulting Pearson generalization due to their deviations.

1.4 Comparing Goodness-of-fit Tests

1.4.1 GOF for Probability Distributions

Several statistical distributions are applicable in many areas (e.g. business and commerce, law, science, public health, medicine, engineering, etc.). As alluded earlier in section 1.1, GOF procedures are useful in assessing how well a proposed distribution model fits a given dataset. GOF tests have therefore, remained competitively useful in model validation and model selection. Authors who have looked at GOF tests for probability distributions with particular interest in finding a probability model that fits well for a given dataset, include Huber-Carol et al. (2012), Cousineau (2009), Rayner et al. (2009), Stuart et al. (1968), among others.

Let $F(x)$ be the unknown distribution function of the sample. $F(x)$ can either be assumed to be continuous or discrete. The GOF of a classical probability model is assessed by describing how well the model fits the observed data. Many tests for GOF evaluate the differences between the observed values, the actual data, and the expected values from the model. In this research the interest lies in evaluating smooth tests of GOF procedures in testing hypotheses on distribution functions.

Hypotheses testing can either be simple or composite. The simple hypothesis is of the form $H_0 : F(x) = F(x;\theta)$, where $F(x;\theta)$ is the probability distribution with which an observed sample is being tested to fit, and θ is some known parametric value. A composite hypothesis can be expressed as $H_0 : F(x) = F(x;\xi(\theta))$, where $\xi(\theta)$ is the domain of an unknown parameter θ. The unknown parameter θ is generally, estimated from the samples.

The normal distribution remains the most conventionally used distributions. Many commonly used method of statistical inference (e.g. t-test, computation of p-values etc.) assume the normal distribution. Other statistical methods exhibit optimum features when normality is assumed. However, normality may not be the case in reality. Failure of the distribution assumptions means failure of the model and conclusions based on the failure model may be invalid. This possibility of failure can objectively be assessed by a GOF test.
For a comprehensive review of the importance of GOF tests in classical probabilities, see Tiku (1986), Chambers et al. (1983), Lawless (2011), Hahn and Shapiro (1968) and Stuart et al. (1968).

1.4.2 GOF for Hazard-based Functions For Single Events

Under hazard-based functions, we can consider testing the assumed form of the underlying probability model in the presence of censored data. The goal is to assesses fitness of a particular model in the presence of a censored variable. This situation that is common with failure-time data. It is usually assumed that the hypothesized model follows the Cox

Proportional Hazard (CPH) model.

Let $T_1, \cdots, T_n$ be independent continuous failure times, with the distribution function of T_i given by $F(t;\lambda_i,\delta)$, where λ_i refers to vectors of known covariates, and $\gamma = (\beta,\delta)$ is a vector of unknown parameters. For the cumulative distribution $\overline{F} = F(t;\lambda_i,\delta)$, we write the survivor function $S = 1-\overline{F}$, and the density function, f_i. Assuming that the censoring times V_i have distribution functions C_i. Therefore, we denote $Y_i = \min(T_i,V_i)$ and the indicator $Z_i = I(T_i < V_i)$ are observed.

In order to assesses the CPH assumption, several methods have been proposed (e.g. Cox (1972), Cox (1975), Schoenfeld (1980), Schoenfeld (1982), Andersen (1982), Kalbfleisch and Prentice (2011) etc.). These methods consist either of graphical techniques designed so as to visualize departures from the proportional hazards (PH) assumption or of formal tests based on parameterizing the interaction between covariates and time. The CPH model assumes that the hazard ratio is constant when comparing two treatment groups.In this case the covariates are assumed to be time dependent (Nagelkerke et al., 1984).

The smooth tests of GOF utilized in this study are score tests obtained by nesting the null hypothesis in a larger class of hazard functions. The formulation of this nesting goes back to Neyman (1937) and have comprehensively been covered by several authors e.g. Pena (1998b), Pena (1998a), Kraus (2007b), David (1939), David (1939), Rayner et al. (2009), among others.

1.4.3 GOF for Hazard-based Functions For Recurrent Events

Modelling recurrent events can be approached in different ways. Here, we build the framework from intensity functions and counting processes. Thereafter, we will focus on gap-time.

Given a single recurrent event at time t, let $0 \leq T_1 \leq T_2 \leq \cdots$ denote the event times, where T_k is the time of the k^{th} event. The associated counting process is defined as

$\{N(t), 0 \leq t\}$, which denotes the cumulative number of event generated by the process. Here $N(t) = \sum_{k=1}^{\infty} I(T_k \leq t)$ is the number of events occurring over the time interval $[0,t]$. If $N(s,t) = N(t) - N(s)$ represents the number of events occurring over the intervals (s,t), t^- is infinitesimally smaller than t and t^+ is infinitesimally larger than t.

Models for recurrent events can be specified more generally by considering the probability distribution for the number of events in a short interval $[t, t+\triangle t]$, given the history of event occurrence before time t.

Let $\triangle N(t) = N(t+\triangle t^-) - N(t^-)$ denote the number of events in the interval $[t, t+\triangle t]$ and $H(t) = \{N(s) : 0 < s < t\}$ denote the history of the process at time t.

The intensity of the event is denoted as

$$\lambda(t|H(t)) = \lim_{\triangle t \to 0} \frac{Pr\{\triangle N(t) = 1|H(t)\}}{\triangle t}. \tag{1.2}$$

The intensity function is useful in modelling events processes. See Munk et al. (2011). The GOF problem is to test the hypothesis that $\lambda(t|H(t))$ follows some hypothesised form. Smooth tests for recurrent events have been discussed by Agustin and Peña (2005) and Agustin and Peña (2001).

1.5 Objectives

This study has three specific objectives, namely

1. Smooth test of GOF for the two-parameter Weibull distribution:

   To fit an HIV retention data to the two-parameter Weibull distribution and assess the fit using a smooth test of goodness-of-fit. The HIV retention data is a primary data from a typical HIV care setting in Kenya.

2. Smooth test of GOF for the proportionality assumption in the CPH models:

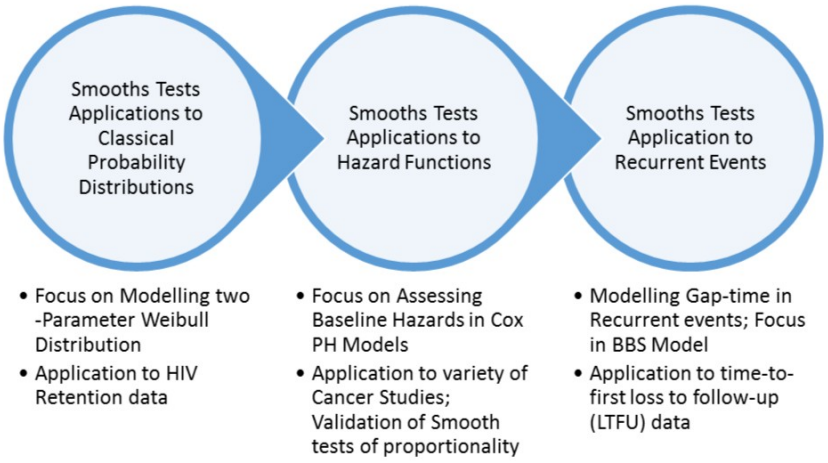
To validate the smooth test of goodness-of-fit for proportionality of the hazard function in the two sample problem in cancer survival studies. Eight real cancer datasets from different settings are assessed for the proportional hazard assumption with an interest of validating the assumption of proportionality in CPH models.

3. Smooth test of GOF based on the BBS modelling of time to first LTFU:
   To test the baseline hazard function of time-to-first loss to follow-up using smooth test of GOF. With a sample size 2,987 and 28% experienced first lost to follow-up during observation period, the interest is to fit BBS model to first LTFU data and test the fit using smooth test of GOF.

This study is organised in three phases i.e. phase one involves, fitting LFTU data to parametric models in order to reflect the randomness of LTFU. In the second phase, we validate the smooth test under variety of real cancer data sets. In phase three, we fits LTFU data to time-to first occurrence recurrent event and assesses the fit using smooth test.

The general framework is described on the next page:

Figure 1.1: General Framework and flow of work



Chapter 2

Literature Review

The importance of probability and hazard models cannot be understated. Models and their associated statistical analyses are crucial in the process of decision making in different areas, for example, medicine, psychology and biology, among others. As stated earlier, GOF tests are mainly concerned with assessing the validity of probability and hazard models. Assessing model fit (i.e., the discrepancy between a model and the data) is critical in applications, as inferences drawn on poorly fitting models may be misleading.

2.1 GOF for Complete Samples

Let $X_1, \cdots, X_n$ be i.i.d. random variables with density $f(x)$. We can test the simple hypothesis $H_0 : f(x) = f_0(x)$, where $f_0(x)$ is some specified density. Several tests exist in literature ranging from graphical and probability plots to non-graphical tests. Some of the conventionally used GOF tests for H_0 are the empirical GOF tests (KS test, CVM test and AD test). These tests have been discussed by many authors and a lot of work has been done on their empirical and asymptotic powers, efficiencies and other properties (Sürücü, 2008; Stuart et al., 1968; Chambers et al., 1983; Huber-Carol et al., 2012). Although the tests are omnibus, for moderate sample sizes only a few deviations from f_0 can be detected

by these tests with substantial frequency (Inglot et al., 1997; Miller Jr and Quesenberry, 1979; LaRiccia, 1991; Ledwina, 1994). Their results show how these tests distribute their asymptotic powers in the space of all alternatives. In particular, they show that there are only very few directions of deviations from $f_0(x)$ for which the tests are of reasonable asymptotic power. When testing for GOF, alternative hypotheses are treated as omnibus tests. An omnibus test is a test that is consistent against essentially all alternatives. See Neyman (1937); Rayner et al. (2009); Rayner and Best (1990); Bargal and Thomas (1983); LaRiccia (1991); Lemeshko et al. (2009) for details.

The approach here is based on modification of the test statistic to allow for nuisance parameters (Koziol, 1987). For hypotheses about a scalar parameter, this modification was studied further by Rayner et al. (2009). The advantage of the score statistic here is that the sample being considered can be used for computation of nuisance parameters using maximum likelihood estimator (MLE).

The smooth test of GOF is constructed to have good power against an alternative whose probability density function or hazard function departs "smoothly" from that specified by the null hypothesis (cf Rayner et al. (2009)). Neyman's concept of smooth test is that it should be constructed to be locally most powerful, unbiased and of size α for testing for uniformity against some order k alternative. Comparison of the power of smooth test have been elucidated by Inglot et al. (1994). Note, however, that the class of "heavy-tailed" alternatives is slightly larger for Neyman's test than for the chi-square test. Typical statements for choosing the orthonormal system in Neyman's test have substantially greater power than the chi-squared test for smooth alternatives (Sürücü, 2008) and to detect alternatives of particular interest, an orthonormal basis should be selected that gives a compact representation of those alternatives (Rayner et al., 2009).

2.2 GOF for Hazard Functions

The Cox (1972) model has become the most widely used statistical tool for analyzing censored failure time data due to its flexibility and versatility. The model specifies that the hazard function $h(t) = \lim_{d\to 0} Pr[T < t + d/T > t]$ for the failure time T of an entity with a p-vector of covariates Z has the form

$$h(t|Z(t)) = \lambda_0(t)\exp\{\boldsymbol{\beta}^t Z(t)\}, \tag{2.1}$$

where β is a p-vector of unknown regression coefficients, and $\lambda_0(t)$ is the baseline hazard function. Model (2.1) assumes that (i) all relevant covariates are included; (ii) the regression form of the hazard function on covariates is exponential; and (iii) the relationship between the baseline hazard function and the regression function of covariates is multiplicative. The violation of these assumptions may have adverse effects on the statistical inference. Furthermore, model miss-specification can lead to distortion of the size and reduction of the power of the partial likelihood score test (Lin, 1991; Lin et al., 1993). Various graphical techniques have been proposed to check the CHPM assumptions (e.g., Crowley and Hu (1977); Kay (1977); Cox (1975); Kalbfleisch and Prentice (2011); Andersen et al. (2012); Schoenfeld (1980); Arjas (1988)).

Non-graphical approaches for testing the CPH model have also been studied by several authors. In his paper, Cox (1972) proposed a way of model checking by introducing a "dummy" time-varying covariate. This method is restricted to testing against a specific alternative. Schoenfeld (1980) compared the observed and the expected numbers of deaths in the cells arising from a partition of the Cartesian product of the range of covariates and the times axis. Similar approaches were taken by O'Quigley and Pessione (1989). However, the partitions of time-axis and covariates are often arbitrary. In addition, different partitions might lead to conflicting results. Lin et al. (1993) proposed an omnibus test for the two-sample problem. Other procedures appeared in Andersen et al. (2012), Barlow and

Prentice (1988), Nagelkerke et al. (1984), O'Quigley and Pessione (1989). These tests are only applicable to specific problems.

Authors that have dealt with extensions of smooth tests of GOF include Pena (1998b), Pena (1998a), Andersen and Gill (1982), Janic-Wróblewska (2004), Agustin and Pena (1999), Arjas (1988) and Baltazar-Aban and Pena (1995). Excellent summaries of some of these procedures have also been given by Nagelkerke et al. (1984) and Koziol (1987). However, there discussion didn't consider wide range of real life applications. The authors considered at most two data sets. This study aims to fill this gap by assessing the performance of smooth test under wide range of different real data sets.

2.3 GOF for Recurrent events

Inference about GOF procedures in recurrent failure data have been discussed by Purohit (1994). Although semi-parametric models are popular in modelling hazard functions, parametric models often provide viable alternatives. In reliability theory, the exponential distribution, Weibull distributions and subsequently IFR (increasing failure rate) and DFR (decreasing failure rate) distributions brings an understanding of the practical application of semi-parametric models. An important aspect then is how to check validity of a specific model assumption. Some research have been done in this regard i.e. Chen et al. (2004).

Considering the proportional hazards model, the hazard function for individual i is written as $h_i(t) = h_0(t)\exp(\boldsymbol{\beta}^t \mathbf{X}_i)$, where $h_0(t)$ is the baseline hazard function, $\mathbf{X}_i$ is a (row) vector of covariates, and $\boldsymbol{\beta}$ is a vector of regression parameters. The two basic assumptions of this model are that the effect of $\mathbf{X}$ is linear and that $h_0(t)$ is constant over time. The latter being the actual proportional hazards assumption when time to even is assumed to be exponentially distributed. A number of GOF tests for the CPH model have already been proposed. See Cox (1972); Preedalikit et al. (2016); Hirose (2011) and Andersen (1982).

The fit of CPH model can also be assessed graphically using martingale residuals or partial residuals (Andersen et al., 2012; Xue et al., 2013; Aalen, 1978). In this case, the martingale residual is considered as the difference between the observed and the expected numbers of events for a given individual and a plot of martingale residuals against a covariate can therefore be used to assesses departures from the linearity assumption. Furthermore, the partial residual is the difference between the covariates of the individual that fails at a given time and the (weighted) mean of the covariates of all individuals at risk at that time. A plot of partial residuals for a given covariate may reveal a violation of the proportional hazards assumption.

The test based on Neyman smooth test is applicable even when the available data is incomplete because of censoring, hence are applicable to survival models. Important papers dealing with the proportional hazards model (Pena, 1998a,b; Kraus, 2007b; David, 1939; Adekpedjou et al., 2012; Agustin and Peña, 2001; Janic-Wróblewska, 2004; Kraus, 2007a) forms the basis of this study. Smooth GOF tests for the classical formulation have the appealing property of having good power over a wide range of alternatives compared to other GOF tests, to the extent that Rayner et al. (2009) implored practitioners to use a smooth test rather than other methods.

Although the theory of the original Neyman's smooth test was well known, there has been several extensions to cover different orthonormal systems. Most of these efforts are directed at completely specified hypotheses (simple case). For composite hypotheses, Barton (1956) considered statistic with unknown parameter. The statistics of Thomas and Pierce (1979) are based on the quadratic score statistic, have convenient distribution theory and appropriate optimality properties. The smooth test based on orthonormal functions are chosen from the class of more powerfully detect particular alternatives. Since it is always possible to transform any probability to uniformity and use one of the many systems orthonormal on the uniform distribution, using an orthonormal system makes it flexible (Rayner et al., 2009; Rayner and Best, 1990).

To the best of our knowledge no author have considered modelling LTFU; this forms the main contribution of this study.

2.4 HIV Retention

Disruption in care through missed scheduled visits undermines both social as well as clinical outcomes, including risk of virological failure (Rachlis et al., 2014; Asiimwe et al., 2016; Sengayi et al., 2013; Megerso et al., 2016; Wang et al., 2011; Ramadhani et al., 2007). Discontinuation of ART can lead to drug resistance, HIV-related illnesses and death. It has been shown in Rachlis et al. (2014) that individuals who miss visits in the first year of treatment have a higher mortality rate. Studies also show that retention of patients who are on ART treatment remains stable after 12 months of ART initiation, with LTFU being the main cause of attrition (Rasschaert et al., 2012). In resource-limited settings, its common to find patients dropping out of ART treatment. The dropouts are usually attributed to LTFU (Asiimwe et al., 2016). Due to significant drop-outs, patients may not realise the benefits of ART if they are LTFU. Previous studies have singled out associations between frequent LTFU and more severe opportunistic illnesses (Haddow et al., 2003). However, modelling LTFU data in either parametric or non-parametric models and assessing the fit using smooth test of GOF have not been done.

Patients who are actively receiving ART are particularly vulnerable to developing drug-resistant infection when virological failure occurs, which could potentially result in broad resistance to ART and transmission of drug-resistant viruses (Rachlis et al., 2014). Determining correct patterns of LTFU and exploring factors associated with them is therefore crucial in identifying the patients who are at-risk of LTFU. Further, analyses of time to LTFU is useful in informing development of evidence-based interventions that improve patient outcomes (Rachlis et al., 2014).

Chapter 3

Methodology

3.1 Smooth Tests of GOF for Probability Distributions

Smooth test of GOF was introduced by Neyman (1937). In his paper, Neyman only considered situations where there were no nuisance parameters (Neyman, 1937; Rayner et al., 2009; Rayner and Best, 1990)). He applied the probability integral transformation (PIT) to test uniformity. PIT is useful in transforming any distribution to uniformity and consequently enables one to use Legendre polynomials.

The components of the smooth test statistics are designed to detect mean, variance, skewness and kurtosis in that order (Rayner and Best, 1990). The test statistics are derived to be the sum of orthonormal sets. The orthornormal sets are chosen with the alternative that is most powerfully.

Rayner et al. (2009) and Koziol (1987) suggested smooth tests based on orthonormal functions as opposed to quadratic approach which were proposed by Thomas and Pierce (1979). Some of the advantages of orthonormal approach include:

1. They involves sum of squares and not quadratic forms therefore making them relatively easier to implement,

2. Numerical integration is not needed to specify constants in the test statistics,

3. The components are always identifiable with known moment-type statistics used in tests of fit,

4. Components are asymptotically independent.

Let $f(x)$ be a probability density function. The null probability density can be nested to form (Rayner et al., 2009, 2008; Rayner and Best, 1990)

$$C(\boldsymbol{\theta})\exp\left\{\sum_{i=1}^{k}\theta_i h_i(x)\right\}f(x), \tag{3.1}$$

where $\boldsymbol{\theta}=(\theta_1,\theta_2,\cdots,\theta_k)^t$ is a vector of real parameters, $C(\boldsymbol{\theta})$ is the normalizing constant that ensures the new functions integrates to 1, i.e.

$$\int_{-\infty}^{\infty} C(\boldsymbol{\theta})\exp\left\{\sum_{i=1}^{k}\theta_i h_i(x)\right\}f(x)dx=\int_{-\infty}^{\infty} g(x;\theta)=1, \tag{3.2}$$

and $h_i(x)$ is a set of orthonormal functions on $f(x)$ so that

$$\int_{-\infty}^{\infty} h_i(x)h_j(x)f(x)dx=\delta_{ij}, \tag{3.3}$$

where

$$\delta_{ij}=\begin{cases}1, & \text{if i is equal to j;}\\ 0, & \text{if i is not equal to j.}\end{cases} \tag{3.4}$$

Let $X_1,X_2,\cdots,X_n$ be a random sample from $f(x)$. We wish to test $H_0:\boldsymbol{\theta}=\mathbf{0}$ against $H_A:\boldsymbol{\theta}\neq\mathbf{0}$.

The score test can be derived as follows:

From equation 3.2, the likelihood function L becomes

$$L = \prod_{j=1}^{n} C(\boldsymbol{\theta}) \exp\left\{\sum_{i=1}^{k} \theta_i h_i(x_j)\right\} f(x_j), \tag{3.5}$$

yielding

$$L = (C(\boldsymbol{\theta}))^n \exp\left\{\sum_{j=1}^{n}\sum_{i=1}^{k} \theta_i h_i(x_j)\right\} \prod_{j=1}^{n} f(x_j). \tag{3.6}$$

The log-likelihood function becomes

$$l = \log L = n\log(C(\boldsymbol{\theta})) + \sum_{j=1}^{n}\sum_{i=1}^{k} \theta_i h_i(x_j) + \sum_{j=1}^{n} \log f(x_j) \tag{3.7}$$

In Rayner et al. (2009), theorem 4.2.1, the derivation and proof of partial derivatives of $\log C(\boldsymbol{\theta})$ satisfies

$$\frac{\partial \log C(\boldsymbol{\theta})}{\partial \theta_i} = \mathrm{E}_k[h_i(x)] \tag{3.8}$$

and

$$\frac{\partial^2 \log c(\boldsymbol{\theta})}{\partial \theta_i \theta_j} = -\mathrm{cov}_k[h_i(x)h_j(x)]. \tag{3.9}$$

The test statistic Ψ_k can therefore be derived by the formulae of generalised score test under null H_0

$$\Psi_k = U^T(\theta) I(\theta)^{-1} U(\theta), \tag{3.10}$$

where the score function is $U(\theta) = \frac{\partial \log L}{\partial \theta}$ and the Fisher's information is $I(\theta) = -E\left[\frac{\partial^2 \log L}{\partial \theta_i \theta_j}\right]$. Differentiating the log-likelihood function (equation 3.7) yields

$$\frac{\partial l}{\partial \theta} = n\mathrm{E}_k[h_i(x)] + \sum_i h_{i(x)} + 0 \tag{3.11}$$

and

$$\frac{\partial^2 l}{\partial\theta_i\theta_j} = -n[\mathrm{E}_k(h_i(x)h_j(x)) - \mathrm{E}_k(h_i(x))\mathrm{E}_k(h_j(x))] = -n\mathrm{cov}_k[h_i(x)h_j(x)]. \tag{3.12}$$

Considering that $\mathrm{E}_0[h_r(x)] = 0$ and by orthonormality property (see Rayner et al. (2009))

$$\mathrm{E}[h_i(x)h_j(x)] = \begin{cases} 1, & \text{for i equal to j;} \\ 0, & \text{for i not equal to j.} \end{cases} \tag{3.13}$$

The score test statistic is given as

$$\Psi_k^2 = \begin{pmatrix} 0 & \sum_i h_i(x) \end{pmatrix} \frac{1}{n} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ \sum_i h_i(x) \end{pmatrix}, \tag{3.14}$$

which can be rewritten as

$$\Psi_k^2 = \left[\sum_i h_i^k(x)\right] \frac{1}{n} \left[\sum_i^k h_i(x)\right] = \frac{1}{n}\left[\sum_i^k h_i(x)\right]^2 \tag{3.15}$$

Theorem 3.1.1 *Under null hypothesis, as the sample size n becomes large ($n \to \infty$), $S(\hat{\boldsymbol{\beta}}) \sim \chi_k^2$. Where the estimate, $\hat{\boldsymbol{\beta}}$ is a vector representing nuisance parameter estimate.*

Proof:

Consider a sequence of independent and identical orthonormal samples $h_r(X_1;\hat{\boldsymbol{\beta}}), h_r(X_2;\hat{\boldsymbol{\beta}}), \cdots, h_r(X_n;\hat{\boldsymbol{\beta}})$. By the orthonormality condition,

1. $h_0 = 0$,
2. $E_0[h_i(X;\hat{\boldsymbol{\beta}})] = 0$,
3. $E_0[h_i(X;\hat{\boldsymbol{\beta}})h_j(X;\hat{\boldsymbol{\beta}})] = \delta_{ij} = \begin{cases} 1, & \text{i} = \text{j}; \\ 0, & \text{i} \neq \text{j}. \end{cases}$

Define V_{i1} as

$$V_{i1} = \frac{h_i(X_1;\hat{\boldsymbol{\beta}}) - nE_0[h_i(X_1;\hat{\boldsymbol{\beta}})]}{\sqrt{n}\delta_{11}}. \tag{3.16}$$

The variable V_{i1} represents a standard score. Using the orthonormal conditions (i.e. $E_0[h_i(X;\hat{\boldsymbol{\beta}})] = 0$ and δ_{11}=1), the variable reduces to

$$V_{i1} = \frac{1}{\sqrt{n}} h_i(X_1;\hat{\boldsymbol{\beta}}). \tag{3.17}$$

Applying the Central Limit Theorem, for each identical and independently distributed $V_{i1}, V_{i2}, \cdots$ the sum of the standard score $\phi_i = V_{i1} + V_{i2} + \cdots$ tends to the standard normal distribution with mean 0 and variance 1 as the size becomes sufficiently large. That is

$$\phi_i = \lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{\sqrt{n}} h_i(X;\hat{\boldsymbol{\beta}}) \to N(0,1). \tag{3.18}$$

The distribution of the standard score υ_n converges to the standard normal distribution as $n \to \infty$ and the sum of squares of a normal variate is χ^2 distributed; so

$$\sum_{i=1}^{p} \upsilon_i^2 \sim \chi_p^2,$$

end of proof.

The υ_i are components of the S_k, and provide directional tests to compliment the omnibus tests based on the S_k,. Since q are elements of β which needs to be estimated, q are typical components of the υ_i. S_k, has the χ^2_{k-q} distribution under both the null and contiguous alternatives. S_{k-q} is usually redefined as a test statistic, in order to emphasize the number of useful components (Rayner and Best, 1990).

The smooth test of GOF, stands out a preferable procedure as it doesn't depend on the sample size n. Other GOF procedures are affected by the value of n where as smooth test only depends on order K.

3.1.1 Testing Uniformity

Application of PIT renders any GOF test, for continuous data, a test for uniformity when the null distribution is completely specified. Tests of uniformity are also applicable when one wishes to combine probabilities found as a result of significance tests on independent data sets relating to the same null hypothesis. More specialized uses of tests of uniformity are (Best and Rayner, 1985)

1. testing for normality on the basis of independent samples,
2. testing GOF for censored samples,
3. two sample tests of equality of circular populations,
4. testing material faults, road accidents, mine explosions etc

To define alternatives, we use Legendre polynomials i.e. $h_r(x)$ are the polynomials that are orthonormal on the uniform distribution $U(0,1)$. The first five polynomials (Rayner and Rayner, 2001) are given by

$h_0(x) = 1,$

$h_1(x) = \sqrt{3}(2x+1),$

$h_2(x) = \sqrt{5}(6x^2-6x+1),$

$h_3(x) = \sqrt{7}(20x^3-30x^2+12x-1),$

$h_4(x) = 3(70x^4-140x^3+90x^2-20x+1).$

The smooth test of order k for uniformity is based on the statistic

$$\Psi_k = \left[\sum_i h_i^k(x)\right]\frac{1}{n}\left[\sum_i^k h_i(x)\right] = \frac{1}{n}\left[\sum_i^k h_i(x)\right]^2 \tag{3.19}$$

3.1.2 Case of Composite Hypothesis

As discussed in chapter 1, the smooth test is derived as an extension of Neyman (1937) test. Here we extend our derivations to cover situations where we have composite hypothesis

(i.e. situations where we have nuisance parameters).

Let $f(x;\boldsymbol{\beta})$ be a probability density function, where $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_q)$ is a vector of nuisance parameters. The null probability density can be nested to form (Rayner et al., 2009)

$$C(\boldsymbol{\theta}, \boldsymbol{\beta}) \exp\left\{\sum_{i=1}^{k} \boldsymbol{\theta}_i h_i(x, \boldsymbol{\beta})\right\} f(x;\boldsymbol{\beta}) \tag{3.20}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_k)$ is a vector of real parameters, $C(\boldsymbol{\theta}, \boldsymbol{\beta})$ is the normalizing constant that ensures the new functions integrates to 1 i.e.

$$\int_{-\infty}^{\infty} C(\boldsymbol{\theta}, \boldsymbol{\beta}) \exp\left\{\sum_{i=1}^{k} \theta_i h_i(x, \boldsymbol{\beta})\right\} f(x;\boldsymbol{\beta}) dx = \int_{-\infty}^{\infty} \varphi_k(x;\boldsymbol{\theta}, \boldsymbol{\beta}) = 1 \tag{3.21}$$

and $h_i(x, \boldsymbol{\beta})$ is a set of orthonormal functions on $f(x;\boldsymbol{\beta})$ so that

$$\int_{-\infty}^{\infty} h_i(x, \boldsymbol{\beta}) h_j(x, \boldsymbol{\beta}) f(x;\boldsymbol{\beta}) dx = \delta_{ij}, \tag{3.22}$$

If $x_1, x_2, \cdots, x_n$ is a random sample from the distribution function $\varphi_k(x;\boldsymbol{\theta}, \boldsymbol{\beta})$, then testing $f(x;\boldsymbol{\beta})$ is the same as testing $H_0 : \boldsymbol{\theta} = \mathbf{0}$ vs $H_A : \boldsymbol{\theta} \neq \mathbf{0}$. The assumption here is that the derivatives of the log-likelihood function with respect to θ and β exist up to second order. The observed random sample $x_1, x_2, \cdots, x_n$ has its log-likelihood function defined as

$$\log L = n \log C(\boldsymbol{\theta}, \boldsymbol{\beta}) + \sum_{i=1}^{k} \sum_{j=1}^{n} \theta_i h_i(x_j, \boldsymbol{\beta}) + \sum_{j=1}^{n} \log f(x_j;\boldsymbol{\beta}). \tag{3.23}$$

The first and second order, partial derivatives of the log-likelihood function with respect to θ_r and β_u yields

$$\frac{\partial \log L}{\partial \theta_r} = n \frac{\log C(\boldsymbol{\theta}, \boldsymbol{\beta})}{\partial \theta_r} + \sum_{j=1}^{n} h_r(x_j, \boldsymbol{\beta}), \tag{3.24}$$

but Rayner et al. (2009) showed

$$\frac{\partial \log C(\boldsymbol{\theta},\boldsymbol{\beta})}{\partial \theta_r} = -\mathrm{E}_k[h_r] \tag{3.25}$$

and

$$\frac{\partial \log C(\boldsymbol{\theta},\boldsymbol{\beta})}{\partial \beta_u} = -\sum_{i=1}^{k} \theta_i \mathrm{E}_k\left[\partial \frac{h_i(x,\boldsymbol{\beta})}{\partial \beta_u}\right] - \mathrm{E}_k\left[\partial \frac{f(x,\boldsymbol{\beta})}{\partial \beta_u}\right]. \tag{3.26}$$

Therefore

$$\frac{\partial \log L}{\partial \theta_r} = n\frac{\log C(\boldsymbol{\theta},\boldsymbol{\beta})}{\partial \theta_r} + \sum_{j=1}^{n} h_r(x_j,\boldsymbol{\beta}) = \sum_{j=1}^{n}(h_r(x_j,\boldsymbol{\beta}) - \mathrm{E}_k(h_r(x,\boldsymbol{\beta}))) \tag{3.27}$$

and

$$\frac{\partial \log L}{\partial \beta_u} = n\frac{\log C(\boldsymbol{\theta},\boldsymbol{\beta})}{\partial \beta_u} + \sum_{i=1}^{k}\sum_{j=1}^{n} \theta_i \frac{\partial h_r(x_j,\boldsymbol{\beta})}{\partial \beta_u} + \sum_{j=1}^{n} \frac{\partial f(x_j,\boldsymbol{\beta})}{\partial \beta_u}. \tag{3.28}$$

Equation 3.28 can be represented as

$$\frac{\partial \log L}{\partial \beta_u} = \sum_{j=1}^{n}\left\{\frac{\partial \log f(x_j,\boldsymbol{\beta})}{\partial \beta_u} - \mathrm{E}_k\left[\frac{\partial \log f(x_j,\boldsymbol{\beta})}{\partial \beta_u}\right]\right\} + \sum_{i=1}^{k}\sum_{j=1}^{n}\theta_i\left\{\frac{\partial h_i(x_j,\boldsymbol{\beta})}{\partial \beta_u} - \mathrm{E}_k\left[\frac{\partial h_i(x_j,\boldsymbol{\beta})}{\partial \beta_u}\right]\right\}. \tag{3.29}$$

The second derivatives will yield

$$\frac{\partial^2 \log L}{\partial \theta_r \partial \theta_s} = n\frac{\partial^2 C(\boldsymbol{\theta},\boldsymbol{\beta})}{\partial \theta_r \partial \theta_s} = -n\mathrm{cov}_k(h_r(x,\boldsymbol{\beta}), h_s(x,\boldsymbol{\beta})), \tag{3.30}$$

$$\frac{\partial^2 \log L}{\partial \theta_r \partial \beta_u} = -n\mathrm{cov}_k\left[h_r(x,\boldsymbol{\beta}), \frac{\partial \log f(x_j,\boldsymbol{\beta})}{\partial \beta_u}\right] + \sum_{j=1}^{n}\left\{\frac{\partial h_i(x_j,\beta)}{\partial \beta_u} - \mathrm{E}_k\left[\frac{\partial h_i(x_j,\boldsymbol{\beta})}{\partial \beta_u}\right]\right\} - \sum_{i=1}^{k}\theta_i \mathrm{cov}_k\left[h_r(x,\boldsymbol{\beta}), \frac{\partial h_i(x_j,\boldsymbol{\beta})}{\partial \beta_u}\right]. \tag{3.31}$$

Given orthonormal condition for $\theta = 0$ (i.e. $E_0[h_r(X;\boldsymbol{\beta})h_s(X;\boldsymbol{\beta})] = n\delta_{rs}$), the partial

derivatives for the log-likelihood function generates the score function U_θ and the asymptotic covariance matrix Σ of U_θ as

$U_\theta = (h_r(X_1;\boldsymbol{\beta}) + h_r(X_2;\boldsymbol{\beta}) + \cdots + h_r(X_n;\boldsymbol{\beta}))$,

$\Sigma = \mathbf{I}_{\theta\theta} - \mathbf{I}_{\theta\beta}\mathbf{I}_{\beta\beta}\mathbf{I}_{\beta\theta}$,

where

$(\mathbf{I}_{\theta\theta})_{rs} = n\delta_{rs}$,

$(\mathbf{I}_{\theta\theta})_{ru} = n\text{cov}_0\left[h_r, \frac{\partial \log f}{\partial \beta_u}\right]$,

$(\mathbf{I}_{\theta\theta})_{uv} = n\text{cov}\left[\frac{\partial \log f}{\partial \beta_u}, \frac{\partial \log f}{\partial \beta_v}\right]$.

The score statistic therefore takes the form $S(\hat{\boldsymbol{\beta}}) = U_\theta^T \hat{\Sigma}^{-1} U_\theta$. Here the score function $U_\theta = U_\theta(\boldsymbol{\beta})$ has r^{th} element $(h_r(X_1;\boldsymbol{\beta}) + h_r(X_2;\boldsymbol{\beta}) + \cdots + h_r(X_n;\boldsymbol{\beta}))/\sqrt{n}$ and Σ is the asymptotic covariance matrix of U_θ.

But

$$\Sigma = \mathbf{I}_{\theta\theta} - \mathbf{I}_{\theta\beta}\mathbf{I}_{\beta\beta}^{-1}\mathbf{I}_{\beta\theta} = nM, \tag{3.32}$$

where

$$M = \mathbf{I}_k - cov_0\left[h, \frac{\partial \log f}{\boldsymbol{\beta}}\right]\left\{\text{var}_0\left[\frac{\partial \log f}{\partial \boldsymbol{\beta}}\right]\right\}^{-1}\text{cov}_0\left[\frac{\partial \log f}{\boldsymbol{\beta}}, h\right], \tag{3.33}$$

which essentially reduces to $M = \mathbf{I}_k$ and the score test takes the form $S(\hat{\boldsymbol{\beta}}_0)$. $\hat{\boldsymbol{\beta}}_0$ is the maximum likelihood estimator of $\boldsymbol{\beta}$ under the null hypothesis and

$$S(\hat{\boldsymbol{\beta}}) = U_\theta^T(\hat{\boldsymbol{\beta}}_0)\hat{M}^{-1}U_\theta(\hat{\boldsymbol{\beta}}_0) = \frac{1}{n}\sum_{j=1}^{p}\hat{V}_j^2, \tag{3.34}$$

where $\hat{V}_j = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} h_r(x_i;\hat{\boldsymbol{\beta}}_0)$. The score statistic for testing $H_0 : \theta = 0$ against $H_A : \theta \neq 0$ is denoted by $S_k(\hat{\beta})$. The choice of k depends on $\hat{\beta}$ through the model dependent modified Bayes information criterion (modBIC) given by

$$\text{modBIC}_k = \hat{S}_k - k\log n$$

in which, relative to BIC, twice the maximized log-likelihood has been replaced by the score statistic. We define k as the smallest order that maximizes modBIC_k i.e.

$$k = \min\{k \in \{1,2,\cdots,d\} \text{ and } \text{modBIC}_k \geq \text{modBIC}_r,\ r = 1,2,\cdots,d\}.$$

This is also referred to as the selection rule (Rayner et al., 2009).

3.1.3 Categorical data

In the same way, we can derive χ^2 as a score test for the cell probabilities $P_1, P_2, \cdots, P_m$. This is done by nesting the null probabilities in the smooth alternative cell probabilities. For $j = 1,2,\cdots,m$

$$\gamma_j = c(\theta)\exp\left\{\sum_{i=1}^{k}\theta_i h_{ij}(x)\right\}P_j, \tag{3.35}$$

so that testing P_j is equivalent to testing $H_0 : \theta = 0$ vs $H_A : \theta \neq 0$. We are dealing with completely specified categorical model. $\theta_i, i = 1,2,\cdots,k$ are real parameters and $c(\theta)$ is a normalizing constant that ensures $\gamma_1 + \gamma_2 + \cdots + \gamma_m = 1$.

Since the parameter space have $m-1$ dimension, $k \leq m-1$. Assuming all expectation exist, the partial derivatives yield (Rayner et al., 2009)

$$\frac{\partial \log \gamma_j}{\partial \theta_r} = h_{rj} - E_k[H_r], \tag{3.36}$$

and

$$\frac{\partial^2 \log \gamma_j}{\partial \theta_r \partial \theta_s} = \frac{\partial^2 \log C(\theta)}{\partial \theta_r \partial \theta_s}. \tag{3.37}$$

The smooth test for most common discrete probability distributions including Poisson, geometric, binomial and negative binomial together with their applications have been covered extensively by Rayner and Best (1990); Rayner et al. (2009); Best and Rayner (1985).

3.1.4 Smooth Tests for Continuous Distributions

Similarly, smooth tests of GOF for common continuous distributions e.g. Normal distribution, exponential distribution, Gamma distribution, Weibull distribution have been covered well in literature. See De Boeck et al. (2011); Rayner et al. (2009, 2008); Rayner and Best (1990); Rayner and Rayner (2001); Rayner and Best (1986); Best and Rayner (1985) for detailed derivation.

Here we generalise for Normal Distribution, Exponential Distribution and Weibull Distribution.

Normal Distribution

The Normal Distribution (i.e. X$\sim N(\mu,\sigma^2)$),is defined as

$$f(x;\mu,\sigma^2)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\},\ -\infty<x<\infty. \tag{3.38}$$

The smooth model for order k will be

$$\varphi_k(x;\theta,\mu,\sigma^2)=C(\theta;\mu,\sigma^2)\exp\left\{\sum_{i=1}^{k}\theta_i h_i(x;\mu,\sigma^2)\right\}\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}. \tag{3.39}$$

We define the set of normalized Hermite polynomials as (De Boeck et al., 2011; Rayner et al., 2009, 2008; Rayner and Best, 1990)

$$\int_{-\infty}^{\infty}H_r(z)H_s(z)\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{z^2}{2}\right\}dz=\delta_{rs} \tag{3.40}$$

Putting $X=\mu+\sigma Z$ in equation 3.39, shows the set of $h_r(x;\mu,\sigma)=H_r(\frac{x-\mu}{\sigma})$ are orthonormal with respect to $X's$.

The first four orthonormal functions will be

$h_0(x;\mu,\sigma^2)=1$,

$h_1(x;\mu,\sigma^2) = \frac{(x-\mu)}{\sigma}$,
$h_2(x;\mu,\sigma^2) = \frac{\left[\frac{(x-\mu)^2}{\sigma^2}-1\right]}{\sqrt{2}}$,
$h_3(x;\mu,\sigma^2) = \frac{\frac{(x-\mu)}{\sigma}\left[\frac{(x-\mu)^2}{\sigma^2}-1\right]}{\sqrt{6}}$.
The score statistic will be $S(\hat{\mu},\hat{\sigma}) = \sum_{r=3}^{k} \hat{U}_r^2$ in which $\hat{U}_r = \frac{1}{\sqrt{n}}\sum_{j=1}^{n} h_r(X_j;\hat{\mu},\hat{\sigma})$.

Exponential Distribution

We define the exponential distribution function as

$$f(x;\beta) = \beta \exp(-\beta x),\ x > 0. \tag{3.41}$$

The smooth model will be

$$\varphi_k(x;\theta,\beta) = C(\theta,\beta)\exp\{\sum_{i=1}^{k} \theta_i h_i(x;\beta)\}\beta\exp(-\beta x), \tag{3.42}$$

Suppose $L_r(Z)$ are the Laguerre polynomials, orthonormal with respect to $f(z;1)$. Then (De Boeck et al., 2011; Rayner et al., 2009, 2008; Rayner and Best, 1990)

$$L_r(Z) = \sum_{s=0}^{r} \binom{r}{s} \frac{(-Z)^s}{s!}. \tag{3.43}$$

By substituting $X = Z/\beta$ in the orthonormality conditions for $L_r(z)$, we find that $h_r(x;\beta) = L_r(\beta x), r = 0, 1, 2, \cdots$ are orthonormal. Since $h_1(x;\beta) = 1-\beta x$,

$$\frac{\partial \log f}{\partial \beta} = \frac{1}{\beta} - x = \frac{h_1(x;\beta)}{\beta}. \tag{3.44}$$

The first four orthonormal functions will be
$h_0(x;\beta) = 1$,
$h_1(x;\beta) = -x\beta$,

$h_2(x;\beta) = 1 - 2x\beta + \frac{x^2\beta^2}{2}$,
$h_3(x;\beta) = 1 - 6x\beta + \frac{3x^2\beta^2}{2} - \frac{x^3\beta^3}{6}$.

Weibull Distribution

A two-parameter Weibull Distribution is define as

$$f(x;\beta,\eta) = \frac{\beta}{\eta}\left[\frac{x}{\eta}\right]^{\beta-1} \exp\left\{-\left[\frac{x}{\eta}\right]^{\beta}\right\}, \tag{3.45}$$

where η is the scale parameter and β is the shape parameter. The orthonormal polynomials for Weibull Distribution for the first four orders are derived from Extreme value distribution because the distribution approaches Weibull distribution as the sample size n becomes large.

The first five orthonormal polynomials are given as (De Boeck et al., 2011; Rayner et al., 2009, 2008; Rayner and Best, 1990):

$h_0 = 1$,
$h_1 = \frac{6}{\pi}Z$,
$h_2 = \frac{Z^2 - \frac{12\zeta(3)Z}{\pi} - \frac{\pi^2}{6}}{\sqrt{\frac{11}{90}\pi^4 \frac{24}{\pi^2}\zeta(3)}}$,
$h_3 = 0.10605Z^3 - 0.49440Z^2 - 0.21942Z + 0.55831$
$h_4 = 0.02493Z^4 - 0.24168Z^3 + 0.26908Z^2 + 0.77691Z - 0.22588$,
where $Z = \frac{x-\mu}{\sigma-\gamma}$, γ is the Euler's constant approximated to be 0.57722, and $\zeta(3)$ involves the ζ function approximated to be 1.20206.

The score statistic is given as

$$S(\hat{\beta},\hat{\eta}) = \sum_{r=2}^{k} \hat{V}_r^2 = \hat{S}_{k-2}, \tag{3.46}$$

where $\hat{V}_r = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} h_j$. The first four values of score statistics is therefore $S_1 = 0$,
$S_2 = 0$,
$S_3 = \frac{\sqrt{b_1} - 1.139547}{\sqrt{20/n}}$,
$S_4 = \frac{b_2 - 7.55\sqrt{b_1} + 3.21}{\sqrt{219.72/n}}$,
where $\sqrt{b_1} = \frac{1}{n} \sum_{j=1}^{n} [(X - \bar{X})/S]^3$ and $b_2 = \frac{1}{n} \sum_{j=1}^{n} \{(X - \bar{X})/S\}^4$. The hypothesis being tested here is $H_0 : f(x; \boldsymbol{\beta}) =$ Weibull distribution(η, β) against a generalised Weibull. The test is equivalent to testing $H_0 : \boldsymbol{\theta} = \mathbf{0}$ against $H_A : \boldsymbol{\theta} \neq \mathbf{0}$. We reject the null hypothesis for large values of the test statistics (simple hypothesis H_0 (the Weibull distribution with parameters: given scale and shape) versus the hypothesis H_1 : a class of the Weibull distribution).

3.1.5 Empirical Goodness-of-fit Tests

Since we are dealing with complete data in this section, we employ other standard empirical goodness-of-fit methods for comparison. The convectional empirical goodness-of-fit tests considered here are AD (A^2), KS (D_n) and CVM (ω^2).

These tests are based on departure between the empirical distribution function $\mathbb{F}_n$ and theoretical distribution function $\mathbb{F}_0$ of the sampled data. The null hypothesis is rejected when the difference is too large with a conclusion that the sampled data doesn't come from the underlying distribution. For the case of a Weibull Distribution, we consider Extreme Value Distribution (Krit, 2014) and therefore apply the empirical cumulative distribution function of $\ln(X_i)$ instead of X_i (Rayner et al., 2008; Bargal and Thomas, 1983).

The measure of difference from the empirical cumulative distribution function of $\ln(X_i)$ is compared against estimated theoretical cumulative distribution function using Maximum Likelihood Estimates i.e. $\hat{\mathbb{F}}_0(x) = \mathbb{F}(x; \ln \hat{\eta}_n, \frac{1}{\hat{\beta}})$

Anderson-Darling test of goodness-of-fit:

$$A^2 = -n + \frac{1}{n}\sum_{i=1}^{n}[(2i-1) - 2n)\ln(1-\hat{U}_i^*) - (2i-1)\ln(\hat{U}_i^*)] \tag{3.47}$$

One-sample Kolmogorov-Smirnov test:

$$D_n = \sqrt{n}\sup|\mathbb{F}_n(x) - \hat{\mathbb{F}}_0(x)| \tag{3.48}$$

$$= \sqrt{n}\max[\max\{\frac{i}{n} - U_i^*\}, \max\{U_i^* - \frac{i-1}{n}\}] \tag{3.49}$$

Cramér-von Mises statistic (CM):

$$\omega^2 = \sum_{i=1}^{n}(\hat{U}_i^* - \frac{2i-1}{2n})^2 + \frac{1}{12n}, \tag{3.50}$$

where $\hat{U}_i^* = \hat{\mathbb{F}}_0(\ln X_i)$.

3.2 Smooth Tests for Hazard Functions

3.2.1 Survival Setting and Data Framework

In classical survival analysis, a collection of individuals are observed from some entry time until a particular event (e.g. death) happens. Often time, not all individuals experience the event of interest; for some, it is only known that the event had not happened at some specified time and in this case the observations of the time to the occurrence of the event is right-censored (Andersen et al., 2012; Fleming and Harrington, 2011).

Consider a sample of n (uncensored) continuous distributed survival times $X_1, X_2, \cdots, X_n$. For a survival function S, with hazard rate λ ; thus $\lambda = \frac{f}{1-F}$ where $F = 1 - S$ is the distribution function and f is the density of X_i.

The hazard rate λ completely, determines the distribution through the relationship Andersen et al. (2012),

$$S(t) = Pr\{X_i > t\} = \Pi_0^t[1-\lambda(s)ds] = \exp(-\int_0^t \lambda(s)ds), \tag{3.51}$$

where the product-integral $\Pi(1-\lambda)$ suggest the Kaplan-Meier estimator

$$\hat{S(t)} = \Pi_0^t(1-d\hat{A}).$$

We can also write

$$1 - dA(s) \approx \exp(-dA(s)).$$

We can also write λ by the heuristic

$$Pr\{X_i \in [t, t+dt] | X_i \geq t\} = \lambda(t)dt. \tag{3.52}$$

Consider the non-parametric estimation of the hazard rate or cumulative hazard rate

$$A(t) = \int_0^t \lambda(s)ds. \tag{3.53}$$

Typically, in survival analysis problems, complete observations $X_1, X_2, \cdots, X_n$ is not possible. The observations are therefore recorded as $(\tilde{X}_i, D_i)$, $i = 1, 2, \cdots, n$ where D_i is a *"censoring indicator"*

$$X_i = \tilde{X}_i \; if \; D_i = 1$$

$$X_i > \tilde{X}_i \; if \; D_i = 0$$

$\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_n$ are random times.

Different kinds of hazard models may be obtained by making different assumptions about the baseline hazard function. More particularly, semi-parametric regression models

for censored survival data have been widely discussed in the literature, starting with the important paper by Cox (1972). The hazard function $\lambda(t,\mathbf{x})$ is defined by

$$\lambda(t,\mathbf{x}) = \lim_{h\to 0}\frac{1}{h}Pr(Y \leq t+h|Y > t;\mathbf{x}), t > 0, \tag{3.54}$$

The cumulative hazard describes the accumulated risk until time t.

$$H(t) = \int_0^t h(u)du, \tag{3.55}$$

Other relationships among the four functions $S(t)$, $H(t)$, $f(t)$ and $h(t)$ are;

$$h(t) = \frac{\partial \log S(t)}{\partial t},$$

$$H(t) = -\log S(t),$$

$$S(t) = \exp\{-H(t)\},$$

$$h(t) = \frac{f(t)}{S(t)} = \frac{f(t)}{1-F(t)}.$$

The Estimators of $S(t)$ and $H(t)$ can be estimated using non-parametric approach i.e. Kaplan-Meier and using parametric assumptions (i.e. Exponential Distribution, Weibull Distribution, Gamma Distribution, Log-normal Distribution etc). Applications of parametric survival function can be found in several literatures (e.g. Cousineau (2009); Andersen et al. (2012); Baltazar-Aban and Pena (1995); Fleming and Harrington (2011)).

3.2.2 Smooth test for Cox Baseline Hazard function

For the CPH model, the GOF problem of testing whether the baseline hazard rate function equals a specified hazard function in the presence of incomplete data is of interest. The GOF tests are score tests obtained by nesting in a larger family of hazard rate functions

developed through smooth and possibly random transformations. The tests are score tests derived by reformulating Neyman's idea of smooth tests in terms of hazard functions (Pena, 1998b).

Score Test Process

The hazard function for the Cox proportional hazard model (Cox, 1972) for T at time t is defined by

$$\lambda(t|X(t)) = \lambda_0(t)\exp\{\beta^t X(t)\} \tag{3.56}$$

where $\lambda_0(\cdot)$ is the baseline hazard function, β is a vector of regression coefficients and transcript t denotes transpose of a vector.

Considering the counting process approach, let $N(t) = \{(N_1(t), N_2(t), \cdots, N_n(t)) : t \in T\}$ be a multivariate counting process in probability space (ω, F, P) where P is some collection of probability measures. Let this family be filtered by a filtration $F = F_t : t \in T$. The time index T may be $[0, \Upsilon]$, where $\Upsilon \leq \infty$. For the Cox proportional hazards model, which is a special case of the multiplicative intensity model (Aalen, 1978; Andersen et al., 2012; Pena, 1998b) the vector of F compensators of N is given by $A = \{(A_l(t), (A_l(t), \cdots, A_n(t)) : t \in T\}$ with

$$A_i(t) = \int_0^t Y_i(s)\lambda(s)\exp\{\beta^t X_i(s)\}ds, i = 1, 2, \cdots, n \tag{3.57}$$

where $Y = \{(Y_1(t), Y_2(t), \cdots, Y_n(t)) : t \in T\}$ is a vector of predictable processes, $\lambda(\cdot)$ is the baseline hazard rate function,β is the $qx1$ vector of regression coefficients, and $X_1(\cdot), X_2(\cdot), \cdots, X_n(\cdot)$ are $qx1$ vectors of predictable covariate processes.

The null hypothesis $H_0 : \lambda(\cdot) = \lambda_0(\cdot)$, where $\lambda_0(\cdot)$ is a completely specified hazard rate function. The associated (cumulative) hazard function is given by

$$\Lambda_0(\cdot) = \int_0^t \lambda_0(s)ds, \tag{3.58}$$

with $\Lambda_0(\cdot)$ strictly non-decreasing. The smooth class of alternatives of order k in which $\lambda_0(\cdot)$ is embedded is defined for the i^{th} component by

$$A_i = \{\lambda_i(\cdot;\theta,\beta) = \lambda_0(\cdot)\exp\{\theta^t\psi_i(\cdot;\beta)\} : \theta \in \Re^k\}, \tag{3.59}$$

where $k \in 1,2,\cdots$ and $\psi_i(\cdot;\beta) = (\psi_{i1}(\cdot;\beta),\psi_{i2}(\cdot;\beta),\cdots,\psi_{ik}(\cdot;\beta))^t$ is a vector of locally bounded predictable processes that are twice-differentiable with respect to β.
Note that the hypothesized baseline hazard rate function $\lambda_0(\cdot)$ is obtained by taking $\theta = 0$, so within this embedding, $H_0 : \lambda(\cdot) = \lambda_0(\cdot)$ is equivalent to $H_0^* : \theta = 0$. The compensator process of $N(\cdot)$ is $A(\cdot;\theta,\beta) = A_1(\cdot;\theta,\beta),A_2(\cdot;\theta,\beta),\cdots,A_n(\cdot;\theta,\beta)$ where

$$A_i(\cdot;\theta,\beta) = \int_0^t Y_i(s)\lambda_0(s)\exp\{\theta^t\psi_i(\cdot;\beta)\}\exp\{\beta^t X_i(s)\}ds. \tag{3.60}$$

The likelihood process is thus given by (Andersen et al., 2012)

$$L(t;\theta,\beta) = \prod_{s=0}^{t}\left\{\prod_{i=1}^{n}[A_i(ds;\theta,\beta)]^{\Delta N_i(s)}\right\} \times [1 - A.(ds;\theta,\beta)]^{1-\Delta N.(s)} \tag{3.61}$$

where $N. = \sum_{i=1}^n N_i(s)$, $A. = \sum_{i=1}^n A_i$, and Π denotes the product integral. Equation 3.61 can be rewritten as

$$L(t;\theta,\beta) = \left\{\prod_{s=0}^{t}\prod_{i=1}^{n}[A_i(ds;\theta,\beta)]^{\Delta N_i(s)}\right\} \times \exp\{-A.(t;\theta,\beta)\} \tag{3.62}$$

The log-likelihood process will therefore be

$$\log L(t;\theta,\beta) = \sum_{i=1}^{n}\int_0^t \log\{Y_i(s)\lambda_i(s;\theta,\beta) \times \exp\{\beta^t X_i(s)\}\}dN_i(s) - A.(t;\theta,\beta) \tag{3.63}$$

The score process associated with θ is therefore

$$U_\theta(t;\theta,\beta) = \sum_{i=1}^{n} \int_0^t \left[\frac{\partial}{\partial \theta} \log \lambda_i(s;\theta,\beta) \right] dM_i(s;\theta,\beta) \tag{3.64}$$

where $M_i(\cdot;\theta,\beta) = N_i(\cdot) - A_i(\cdot;\theta,\beta)$, $(i = 1,2,\cdots,n)$.
Under H_0^*, assuming β_0 is the true, $M(\cdot) = (M_1(\cdot), M_2(\cdot), \cdots, M_n(\cdot))$ with $M_i(\cdot) = M_i(\cdot;0,\beta_0)$ is a vector of local square-integrable orthogonal martingales with quadratic variation process (Pena, 1998b; Andersen et al., 2012)

$$\langle M \rangle(\cdot) = diag\{A(\cdot;0,\beta_O)\} = diag\left\{ \int_0^{\cdot} Y_i(s)\lambda_0(s)\exp\{\beta_0^t X_i(s)\} ds : i = l,2,\cdots,n \right\}, \tag{3.65}$$

where for any vector a, $diag\{a\}$ is the diagonal matrix with diagonal elements as the elements of a. Because $[\frac{\partial}{\partial\theta}] \log \lambda_i(\cdot;\theta,\beta) = \psi_i(\cdot;\beta), (i = 1,2,\cdots,n)$

$$U_\theta(\cdot;0,\beta_0) = \sum_{i=1}^{n} \int_0^{\cdot} \psi_i(s;\beta_0) dM_i(s). \tag{3.66}$$

But the score process $U_\theta(\cdot;0,\beta_0)$ above is not observable, because β_0 is unknown and the $\psi_i's$ and $M_i's$ depend on it. To derive a score test statistic, β_0 above is replaced by an estimator $\hat{\beta}$ based on F. Typically, the estimator used is the full maximum likelihood estimator (MLE) of β subject to the restriction $\theta = 0$. The Cox model is used to estimate β using the partial MLE of β, which is the β that maximizes the Cox partial likelihood function given by Andersen (1982) and Andersen et al. (2012)

$$L(t;\beta) = \prod_{i=1}^{n} \prod_{s=0}^{t} \left[\frac{Y_i \exp\{\beta^t X_i(s)\}}{\sum_{j=1}^{n} Y_j \exp\{\beta^t X_j(s)\}} \right] \tag{3.67}$$

Notice that this partial likelihood process does not depend on the baseline hazard rate function; hence it is automatically (functionally) independent of θ. There is a tradeoff in using the partial MLE of β instead of the full MLE of β restricted to $\theta = 0$. It is that this partial

MLE will be less efficient than the full MLE; on the other hand, it is the estimator under the Cox model.

The score process arising from the partial likelihood is given by Aalen (1978) and Andersen et al. (2012)

$$U_\beta(\cdot;\beta) = \sum_{i=1}^{n} \int_0^{\cdot} [X_i(s) - \mathrm{E}(s;\beta)]\, dM_i(s;0,\beta), \tag{3.68}$$

where, for $m = 0$, 1, or 2, (Pena, 1998b,a; Andersen and Gill, 1982)

$$S_{(m)}(t;\beta) = \frac{1}{n} \sum_{i=1}^{n} X_i(t)^{\otimes} Y_i(t) \exp\{\beta^t X_i(t)\}, \tag{3.69}$$

and

$$\mathrm{E}(s;\beta) = \frac{S_{(1)}(t;\beta)}{S_{(0)}(t;\beta)} \tag{3.70}$$

The other expression for variance will yield the following process

$$V(t;\beta) = \frac{S_{(2)}(t;\beta)}{S_{(0)}(t;\beta)} - \mathrm{E}(t;\beta)^{\otimes 2} \tag{3.71}$$

where for vector $\mathbf{a}$, $\mathbf{a}^{\otimes 0} = 1$, $\mathbf{a}^{\otimes 1} = \mathbf{a}$ and $\mathbf{a}^{\otimes 2} = \mathbf{a}^t\mathbf{a}$. The partial MLE of β, denoted by $\hat{\beta}$, satisfies the equation $U_\beta(T;\hat{\beta}) = 0$. The resulting observable score process used for testing $H_0^* : \theta = 0$ is given by (Pena, 1998b)

$$U_\theta(\cdot;0,\hat{\beta}) = \sum_{i=1}^{n} \int_0^{\cdot} \psi_i(s;\hat{\beta}) dM_i(s;0,\hat{\beta}) \tag{3.72}$$

and, in particular, the score test statistic based on F_τ, is obtained by evaluating $U_\theta(\cdot;0,\hat{\beta})$ at $t = \tau$. The processes $\{M_i(\cdot;0,\hat{\beta}) : i = 1,2,\cdots,n\}$ are special cases of the "martingale" residual processes. An important thing to note is that the test process in above is a score process, and consequently, test procedures arising from this process will have certain asymptotic optimality properties similar to score tests.

Because there is freedom to choose ψ_i's by varying these processes, we generate a sequence

of GOF tests.

The covariance matrix function of the limiting process is (Pena, 1998a)

$$\Xi(t_i,t_2;\beta_0) = \Sigma_{11}(t_1,t_2;\beta_0) - \Sigma_{12}(t_l;\beta_0)\Sigma_{22}(\tau;\beta_0)^{-}\Sigma_{12}(t_2;\beta_0)^t + \Delta(t_l;\beta_0)\Sigma_{22}(\tau;\beta_0)^{-1}\Delta(t_2;\beta_0)^t, \tag{3.73}$$

where the matrix function is given as

$$\Sigma(t;\beta) = \begin{pmatrix} \Sigma_{11}(t;\beta) & \Sigma_{12}(t;\beta) \\ \Sigma_{12}(t;\beta) & \Sigma_{22}(t;\beta) \end{pmatrix}. \tag{3.74}$$

The definition of submatrices in matrix, $\Sigma(t;\beta)$ is given by

$\Sigma_{11}(t;\beta) = \int_0^t \psi^{(0)}(s;\beta)^{\otimes 2} y(s)\lambda_0(s;\beta)ds$

$\Sigma_{12}(t;\beta) = \int_0^t \psi^{(0)}(s;\beta)\rho(s;\beta)^t y(s)\lambda_0(s;\beta)ds$

$\Sigma_{22}(t;\beta) = \int_0^t \rho(s;\beta)^{\otimes 2} y(s)\lambda_0(s;\beta)ds$

$\Sigma_{21}(t;\beta) = \Sigma_{12}(t;\beta)^t$ See Pena (1998a); Agustin and Peña (2001); Koziol (1987); Kraus (2007b), for elaborate descriptions of these functions.

The covariance matrix function $\Xi(t_i,t_2;\beta_0)$ can be consistently estimated by the matrix $\hat{\Xi}(t_i,t_2;\hat{\beta}_0)$. The test statistic

$$S(\tau;\hat{\beta}) = \frac{1}{n} U_\theta(\tau;0,\hat{\beta})^t \hat{\Xi}(t_i,t_2;\hat{\beta}_0)^{-1} U_\theta(\tau;0,\hat{\beta}). \tag{3.75}$$

$S(\tau;\hat{\beta})$ has a limiting chi-squared distribution with degrees of freedom k^*= rank$[\Xi(\tau,\tau;\beta_0)]$. The asymptotic smooth goodness-of-fit test for $H_0 : \lambda(\cdot) = \lambda_0(\cdot)$ versus $H_1 : \lambda(\cdot) \neq \lambda_0(\cdot)$ is reject H_0 if

$$S(\tau;\hat{\beta}) = \frac{1}{n} U_\theta(\tau;0,\hat{\beta})^t \hat{\Xi}(t_i,t_2;\hat{\beta}_0)^{-1} U_\theta(\tau;0,\hat{\beta}) \geq \chi^2_{k^*;\alpha}, \tag{3.76}$$

where $\chi^2_{k^*;\alpha}$ a is the $(1-\alpha)100^{th}$ percentile of the chi-squared distribution with k^* degree of freedom. Effects of replacing the unknown nuisance parameter β_0 by the partial MLE $\hat{\beta}$ in $\frac{U_\theta(\tau;\beta_0)}{\sqrt{n}}$ have covered extensively by Pena (1998b,a); Agustin and Peña (2001). Since $\hat{\beta}$ is a consistent estimator (Pena, 1998b) of β_0, the test statistic can be expressed

$$S^\circ(\tau;\hat{\beta}) = \frac{1}{n} U_\theta(\tau;\hat{\beta})^t \hat{\Sigma}_{11}(\tau;\hat{\beta})^{-1} U_\theta(\tau;\hat{\beta}). \tag{3.77}$$

The limiting distribution of $S^\circ(\tau;\hat{\beta})$ is not even a chi-squared distribution. In such a case, misleading conclusions thus may be reached. Also note that if β_0 were estimated by the MLE based on the full likelihood, then the term involving $\Delta(\cdot;\beta_0)$ in $\Xi(t_1,t_2;\beta_0)$ disappears, so the effect of estimating β_0 is a decrease in asymptotic variance. Consequently, the term involving $\Delta(\cdot;\beta_0)$ in $\Xi(\cdot,\cdot;\cdot)$ can be interpreted as the increase in asymptotic variance attributed to the use of the less efficient partial likelihood MLE.

The ψ_i Processes

Each choice or specification leads to a score test process, a limiting variance process, a quadratic test statistic, and ultimately an asymptotic test of H_0.
The associated score test process is denoted by (Pena, 1998a)

$$Q(t;\beta:\xi) = \frac{1}{\sqrt{n}} U_\theta(t;0,\beta), \tag{3.78}$$

where ξ is the label to identify the choice of ψ. The limiting variance process is denoted by (Pena, 1998a)

$$\begin{aligned}\Xi(t;\beta:\xi) =& \Sigma_{l1}(t;\beta:\xi) - \Sigma_{l2}(t;\beta:\xi)\Sigma_{22}(\tau;\beta)^{-1}\Sigma_{l2}(t;\beta:\xi)^t + \\ & \Delta(t;\beta:\xi)\Sigma_{22}(\tau;\beta)^{-1}\Delta(t;\beta:\xi)^t.\end{aligned} \tag{3.79}$$

This problem is related to that of validating whether this model fits observed data. The goodness-of-fit procedures is derived as score tests and is obtained by embedding $\lambda_0(\cdot)$ in a larger class of hazard rate functions; with this class developed through smooth transformation of $\lambda_0(\cdot)$. The score tests depend on the unknown vector of regression coefficients β, so to obtain usable tests, this β is replaced by its partial MLE. The effect of this substitution is ascertained in the asymptotic properties of the score process. The resulting goodness-of-fit procedures are also related to model validation procedures that utilize generalized residuals and, consequently, through the asymptotic results, the appropriate adjustments needed to properly use procedures based on generalized residuals are obtained. Several classes of goodness-of-fit tests, both omnibus and directional can be generated. Explicit expressions of these tests are presented for the randomly censored model. Furthermore, some goodness-of-fit tests that have been proposed in the literature can actually be derived and justified more formally through counting processes, and because these tests can be viewed as score tests, they are automatically endowed with certain optimality properties. This test can be applicable in more complex dynamic models in survival analysis, reliability, and the econometric settings, where the specification of the model is through hazard rates or failure intensities. The resulting goodness-of-fit tests, both omnibus and directional, are also appealing.

3.2.3 Smooth test for Composite Baseline Hazard

We extend the Neyman goodness-of-fit approach discussed in section 1.2 here in order to cover composite situations. This extension allows situations where censoring variable exists; a situation that is typically useful in time-to-failure scenario.

Let $N = \{N(t) : t \in T \equiv [0, \tau]\}$ be a counting process defined on probability space (Ω, F, P) with filtration $F = \{f_t : t \in \mathtt{T}\}$ (Pena, 1998a). Let observable predictor process be $Y = \{Y(t) : t \in F\}$. Assuming that the filtration considered here is the natural filtration $F^N =$

$\{f_t^N : t \in \mathrm{T}\}$ where f_t^N is the $\sigma-$field generated by $\{(N(s), Y(s)) : s \leq t\}$ and f_0. It is assumed f_0 contains all information at time 0.

The compensator process of N is therefore A=$\{A(t) : t \in \mathrm{T}\}$ where

$$A(t) = \int_0^t Y(\omega)\lambda(\omega)d(\omega) \tag{3.80}$$

and $\lambda(\cdot)$ is unknown hazard rate function (Aalen, 1978).

The smooth test of goodness-of-fit is implemented by testing the null hypothesis $H_0 : \lambda(\cdot)$ belongs to a parametric class $\xi = \{\lambda_0(\cdot;\eta) : \eta \in F \subseteq \Re\}$ of hazard function verses the alternative hypothesis $\lambda(\cdot) \neq \xi$ (McKeague and Utikal, 1991).

We assume the following basic conditions on the $\psi(\cdot;\eta)$ process and the class ζ. If H_0 is true, we let η_0 denote the true value of η. Furthermore, $\psi(\cdot;\eta)$ is a locally bounded and predictable process, and $\frac{\partial}{\partial\eta}\lambda_0(\cdot;\eta)$ exists with $\lambda_0(\cdot;\eta) > 0$ for each $(t,\eta) \in F \times \Gamma$. Under our model, the partial log-likelihood process for (θ,η) [it will be the full log-likelihood under additional assumptions such as noninformative censoring] is given as (Andersen et al., 2012)

$$\log L(t;\theta,\eta) = \int_0^t \log[Y(s)\lambda_k(s;\theta,\eta)]dN(s) - \int_0^t Y(s)\lambda_k(s;\theta,\eta)ds. \tag{3.81}$$

Let

$$M(t;\theta,\eta) = N(t) - \int_0^t Y(s)\lambda_k(s;\theta,\eta)ds, t \in F. \tag{3.82}$$

Then, the score process is defined as

$$U(t;\theta,\eta) = \begin{pmatrix} U_1(t;\theta,\eta) \\ U_2(t;\theta,\eta) \end{pmatrix} = \int_0^t \begin{pmatrix} \psi(s;\eta) \\ \frac{\partial}{\partial\eta}\log\lambda_k(s;\theta,\eta) \end{pmatrix} dM(;\theta,\eta), \tag{3.83}$$

which, when evaluated at $(\theta, \eta) = (0, \eta)$, becomes

$$U(t;\theta,\eta) = \begin{pmatrix} U_1(t;\theta,\eta) \\ U_2(t;\theta,\eta) \end{pmatrix} = \int_0^t H(s;\eta_0)dM(s;\eta_0), \tag{3.84}$$

where $H(s;\eta) = [\psi(s;\eta), \frac{\partial}{\partial\eta}\log\lambda_k(s;\theta,\eta)]^t$ and $M(s;\eta) = M(s;0,\eta)$.
Under H_0, $\{M(t;\eta_0) : t \in \mathcal{F}\}$ is a square-integrable local martingale, and its predictable variation process is

$$\langle M(\cdot;\eta_0)\rangle(t) = A(t;\eta_0) = \int_0^t Y(s)\lambda_0(s;\eta_0)ds. \tag{3.85}$$

Since $\{H(t,\eta_0) : t \in \mathcal{F}\}$ is a locally bounded predictable process, then under H_0, $\{U(t,\eta_0) : t \in \mathcal{F}\}$ is a square-integrable local martingale with predictable quadratic variation process

$$\langle U(\cdot;\eta_0)\rangle(t) = \int_0^t H(s;\eta_0)^{\otimes 2}Y(s)\lambda_0(s;\eta_0)ds. \tag{3.86}$$

If, under H_0, the true value η_0 is known, a test of Ho can be based on the statistic

$$U_1(\tau;\eta_0) = \int_0^t \psi(s;\eta_0)dM(s;\eta_0). \tag{3.87}$$

However, since η_0 is unknown, an asymptotically optimal test can be obtained from the efficient score vector

$$\hat{U_1}(\tau;\eta_0) = U_1(\tau;\eta_0) - \Sigma_{12}(\tau;\eta_0)\Sigma_{22}(\tau;\eta_0)^{-1}U_2(\tau;\eta_0), \tag{3.88}$$

upon replacing η_0 by a suitable estimator, and where

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \tag{3.89}$$

is a (possibly limiting) covariance matrix of U. Replace η_0 by its restricted maximum likelihood estimator (Pena, 1998a) $\hat{\eta}$ obtained under the restriction $\theta = 0$.

Asymptotic properties of this estimator have been discussed by Andersen and Gill (1982), Andersen et al. (2012) and McKeague and Utikal (1991).

Substituting $\hat{\eta}$ for η_0, the estimated efficient score vector becomes

$$U_1(\tau;\hat{\eta}) = \hat{U_1}(\tau;\hat{\eta}) = \int_0^\tau \psi(s;\hat{\eta}))dM(s;\hat{\eta}). \tag{3.90}$$

The process $M(t;\hat{\eta}) : t \in F$ is the martingale residual process (Fleming and Harrington, 2011).

To obtain the exact form of the test, we need the sampling distribution of $U_1(\tau;\hat{\eta})$ and its covariance matrix $\Xi(\tau;\eta_0)$ under H_0. The smooth goodness-of-fit test for H_0 of order k associated with $\psi(\cdot;\eta)$ is to reject H_0 if

$$S_k(\tau;\hat{\eta}) = U_l(\tau;\hat{\eta})^t \hat{\Xi}(\tau;\hat{\eta})^{-1} U_1(\tau;\hat{\eta}) \geq c_\alpha, \tag{3.91}$$

where c_α is such that $Pr[S_k(\tau;\hat{\eta}) \geq c_\alpha | H_0)]$ equals α.

The asymptotic properties of the test have been covered extensively by Pena (1998a). He showed that the test statistic has an asymptotic chi-squared distribution with k^* degrees-of-freedom under H_0. The asymptotic α-level smooth test then rejects H_0 whenever $S_k(\tau;\hat{\eta}) \geq \chi^2_{\hat{k^*},\alpha}$, where $\hat{k^*}$ is the rank of the covariance matrix $\Sigma_{11.2}(\tau;\hat{\eta})$. The definition and components of $\Sigma_{11.2}(\tau;\hat{\eta})$ have been covered in section 3.2.2.

3.2.4 Testing baseline hazard Function for Weibull Distribution

One of the most common family is the Weibull Distribution, and its survival function can be defined as

$$S(t) = exp\{-(\lambda t)^p\} \tag{3.92}$$

and hazard function

$$\lambda(t) = p\lambda(\lambda t)^{p-1} \tag{3.93}$$

for parameters $\lambda > 0$ and$p > 0$. If $p = 1$, the model reduces to the exponential and has constant risk over time. If $p > 1$, then the risk increases over time. If $p < 1$, then the risk decreases over time. If we pick the Weibull Distribution as a baseline risk and then multiply the hazard by a constant in a proportional hazards framework, the resulting distribution turns out to be still a Weibull Distribution, so the family is closed under proportionality of hazards. If we pick the Weibull Distribution as a baseline survival and then speed up the passage of time in an accelerated life framework, dividing time by a constant , the resulting distribution is still a Weibull Distribution, so the family is closed under acceleration of time. Suppose we want to test $\lambda_0(\cdot)$ belongs to the two-parameter Weibull Distribution class of hazard rate functions

$$\{\lambda_0(t;\alpha,\eta) = \alpha\eta(\eta t)^{\alpha-1} : \alpha > 0, \eta > 0\}, \tag{3.94}$$

where α and η is the scale and shape parameter respectively. Immediate substitution yield

$\rho^{(0)}(t;\alpha,\eta) = \begin{pmatrix} \frac{\alpha}{\eta} \\ \frac{1}{\eta}(1+\log t) \end{pmatrix}$,

$\gamma_1 = \frac{r^{(0)}\tau^{(0)}}{\alpha} \begin{pmatrix} 0 \\ C(\tau^{(0)}) \end{pmatrix}$

and $\Psi = r^{(0)}\tau^{(0)} \begin{pmatrix} \frac{\alpha^2}{\eta^2} & \frac{1}{\eta}E^{(1)}(\tau^{(0)}) \\ \frac{1}{\eta}E^{(1)}(\tau^{(0)}) & \frac{1}{\alpha^2}E^{(2)}(\tau^{(0)}) \end{pmatrix}$

where,

$D^{(j)}(\tau^{(0)}) = \int_0^{\tau^{(0)}} (\log w)^j d\psi_1^{(0)}(w), (j = 0,1,2),$

$E^{(1)}(\tau^{(0)}) = D^{(0)}(\tau^{(0)}) + D^{(1)}(\tau^{(0)}),$

$E^{(2)}(\tau^{(0)}) = D^{(0)}(\tau^{(0)}) + 2D^{(1)}(\tau^{(0)}) + D^{(2)}(\tau^{(0)})$

and $C(\tau^{(0)}) = \int_0^{\tau^{(0)}} (1+\log w)\psi_1^{(0)}(w) d\psi_1^{(0)}(w).$

Also, we define

$$V(\tau^{(0)}) = E^{(2)}(\tau^{(0)}) - [E^{(1)}(\tau^{(0)})]^2. \tag{3.95}$$

The empirical versions of these quantities are $\hat{D}^{(j)}(\hat{\tau}^{(0)})'s$, $\hat{E}^{(j)}(\hat{\tau}^{(0)})'s$, $\hat{C}(\hat{\tau}^{(0)})$ and $\hat{V}(\hat{\tau}^{(0)})$, which are found by replacing $\psi_1^{(0)}(\cdot)$ with $\psi_1^{(n)}(\cdot) = \frac{R_0^R(\cdot)}{R_0^R(\hat{\tau^{(0)}})} - 1/2$ in their definitions.
The test statistic is then obtained by setting

$$\hat{\Delta}(\hat{\tau}^{(0)}) = \frac{\hat{C}(\hat{\tau}^{(0)})^2}{\hat{V}(\hat{\tau}^{(0)})}. \tag{3.96}$$

3.2.5 Smooth Tests of Goodness-of-fit for 2-sample hazard functions

This is achieved by nesting the Cox PH model defined in equation (1) to yield .

$$\lambda_i(t) = Y_i(t)\lambda_0(t)\exp\{\beta^t X_i(t) + \theta^t \Psi_i(\cdot,\beta)X_{iq}(t)\}, \tag{3.97}$$

where $\Psi_i(\cdot,\beta) = \{(\psi_1(\cdot,\beta), \psi_2(\cdot,\beta), \cdots, \psi_k(\cdot,\beta))\}$ is the smooth alternative chosen as some basis functions in standardised time i.e.

$$\psi_i(\cdot,\beta) = \varpi_i(1 - \exp\{-\hat{\gamma}_0(\cdot,\beta)\}) \tag{3.98}$$

where $\hat{\gamma}_0(\cdot,\beta)$ is the Breslow estimator of $\gamma_0(\cdot,\beta) = \int_0^t \lambda_0(s)ds$ and $\varpi_i(\cdot)$ are some bounded functions on $[0,1]$ (Kraus, 2007a). Details on derivation of the test statistic, asymptotic properties and how to determine the choice k including data-driven version of the test have been elucidated by Kraus (2007a), Kraus (2007b) and Kraus (2009).
The score test statistic for testing $H_0 : \theta = 0$ against $H_A : \theta \neq 0$ is defined by

$$S_k = U(\cdot;\hat{\beta})^t \Sigma(\cdot;\hat{\beta}) U(\cdot;\hat{\beta}) \tag{3.99}$$

where $U(\cdot;\hat{\beta})$ is a score process for θ evaluated at $(\beta^T,\theta^T)=(\beta^T,(0,0,\cdots,0))$ and has been shown by Kraus (2007a) to be

$$\sum_{i=1}^{n}\int_0^t \psi(s,\beta)X_{\varpi}(s)dN_i(s)-\int_0^t \frac{\sum_{i=1}^n Y_i(t)\Psi(s,\beta)X_{\varpi}(s)\exp\{\beta^T X_i(t)\}}{\sum_{i=1}^n Y_i(t)\exp\{\beta^T X_i(t)\}}d\bar{N}(s). \quad (3.100)$$

The estimated variance (Σ) of the score is given as

$$\hat{\Sigma}(\cdot;\hat{\beta})=\hat{\Sigma}_{22}(\cdot;\hat{\beta})-\hat{\Sigma}_{21}(\cdot;\hat{\beta})\hat{\Sigma}_{11}(\cdot;\hat{\beta})^{-1}\hat{\Sigma}_{12}(\cdot;\hat{\beta}) \quad (3.101)$$

where $\hat{\Sigma}_{11}(\cdot;\hat{\beta})=[U_1(\cdot;\hat{\beta})](t)$, $\hat{\Sigma}_{22}(\cdot;\hat{\beta})=[U_2(\cdot;\hat{\beta})](t)$ and
$\hat{\Sigma}_{21}(\cdot;\hat{\beta})=[U_2(\cdot;\hat{\beta}),U_1(\cdot;\hat{\beta})](t)$.
$S_k\to\chi^2_k$ for $n\to\infty$ under H_0 and H_o is rejected for large values of S_k.

3.2.6 Other Conventional GOF Tests

In the classical non-censored one-sample goodness-of-fit problem, one observes a random sample $X_1,\cdots,X_n$ from a population with distribution function $F(x)=Pr(X\le x)$; the corresponding survival function is $\bar{F}(x)=Pr(X>x)=1-F(x)$. The null hypothesis asserts that $F(x)=G(x)$, where G is completely specified. The need to generalize this problem to encompass censored data arises because in some situations, such as clinical trials, or life testing, the X's may represent times to the occurrence of an end-point event and the data are usually analyzed before all patients, or items on test, have experienced the event. In the clinical trials context the end-point event could, for example , be relapse , pregnancy , or death. In the life-testing framework, the end-point event could be failure of the inner ring of ball bearings which are on test (Gatsonis et al., 1985).

Kolmogorov-Smirnov D_n two-sample Goodness-of-fit for Hazard functions

Kolmogorov-Smirnov test for two sample proportional hazard with right censoring has been covered by Lin et al. (1993) and Therneau et al. (1990). More on asymptotic properties have been covered by Arjas (1988); Wei (1984) and Persson (2002). The test is based on the simplified partial likelihood score process and it tests the hypothesis that transformation hazards are proportional in two samples of right censored data. The test uses Kolmogorov-Smirnov supremum statistic based on the simplified partial likelihood score process and martingale simulations are used to compute the p-value.

Let ρ be a variable that indicates group 1 or 2 and

$$U(\hat{\beta},t) = \{U_1(\hat{\beta},t), U_2(\hat{\beta},t), \cdots U_p(\hat{\beta},t)\} \tag{3.102}$$

be the empirical score process. The standardized score process $F^{1/2}U(\hat{\beta},t)$ where $\hat{\beta}$ is the NPMLE estimator β. For the supremum test, if the dimension, p $\geq$1, each of the proportional hazards test statistics (Lin et al., 1993),

$$\sup_t \{F^{-1}(\hat{\beta})_{jj}\}^{\frac{1}{2}} |U_j(\hat{\beta},t)| \ (j = 1,2,\cdots,p) \tag{3.103}$$

has the asymptotic distribution of $\sup_{0\leq U\leq} |B^0(u)|$ if $\{V(t)\}_{jk} = 0$ for $(j \neq k)$ for all t, where $V(\cdot)$ is the limiting covariance matrix for $n^{-\frac{1}{2}}U(\beta_0,\cdot)$.

However if the independence of covariates used to determine $V(\cdot)$ fails, assessing the overall proportionality, can be approached by the following test statistic

$$\sup_t ||U(\hat{\beta},t)|| \ or \ \sup_t \sum_{j=1}^{p} \{F^{-1}(\hat{\beta})_{jj}\}^{\frac{1}{2}} |U_j(\hat{\beta},t)|. \tag{3.104}$$

More on the consistency of the test against the nonproportional hazards alternative have been covered by Lin et al. (1993) and Kraus (2009).

Global test for two-sample Proportional Goodness-of-fit test

The global test is widely used to test the proportional hazards assumption for a CPH model. The test was first proposed by Quantin et al. (1996). The test is based on a semi-parametric generalization of the proportional hazards regression model. The hazard function corresponding to a covariate vector X and has the time function defined as

$$\Phi(t) = 1 + \log[\Lambda_0(t)] \tag{3.105}$$

where $\Lambda_0(t)$ is the cumulative baseline hazard function which is essentially Breslows maximum likelihood estimator under H_0 (Lin et al., 1993; Lin, 1991; Lim and Zhang, 2011). The hypothesis of proportional hazards, $H_0 : \theta = 0$, is tested by using a score statistic derived from the partial likelihood. The Breslow estimator for $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ is

$$\hat{\Lambda}_0(t) = \sum_{i=1}^{n} \int_0^t \frac{dN_i(u)}{\sum_{j=1}^{n} Y_j(u) \exp(\hat{\beta}^t X_j(u))} \tag{3.106}$$

where $\hat{\beta}$ maximizes the partial log-likelihood of β. As in the special case of the proportional hazards model, β and $\hat{\Lambda}_0(t)$ are NPMLEs. For the more exhaustive coverage of $\Lambda_0(t)$ see Lin (1991); Lin et al. (1993); Chen et al. (2004) and Persson (2002).

3.3 Smooth Tests for Baseline Hazard in Recurrent Events

3.3.1 General Overview of Recurrent Events

Recurrent events are processes that generate events repeatedly over time. The events occurs in many areas such as public health, medicine, engineering, reliability studies etc. Examples of recurrent events in public health settings are the re-occurrence of polio, low CD4

count in an HIV patient during ART, relapse of a drug and alcohol after rehabilitation, recurring migraines, TB recurrences etc. From engineering studies we have breakdown of a machine, the failure of an operating system, or malfunctioning parts on an assembly line. The deteriorating episodes of visual acuity, and the turnover rate for a company are also examples of recurrent events in other settings. Studies on recurrent events that involves a variety of statistical fields have been developed. See Cook and Lawless (2007), Lim and Zhang (2011), Aalen and Husebye (1991), Adekpedjou et al. (2012) etc for more on examples of recurrent events. Recurrent events are structured to be of naturally ordered failure time and the different events within an individual may be correlated.

Methods for recurrent events analysis include nonhomogeneous Poisson process (NHPP), Andersen-Gill (AD), Wei-Lin-Weissfeld (WLW), Prentice, Williams and Peterson total time (PWP-CP), Prentice, Williams and Peterson gap time (PWP-GT) and Lee, Wei and Amato (LWA). Generally, nonhomogeneous Poisson process (NHPP) models and renewal process models have been studied with the requirements of strictly minimal repair and perfect repair, respectively (Cox, 1972), LWA model allows a subject to be at risk several times for the same event, WLW model overestimates treatment effect where as PWP-GT and TT-R are useful models for analysing recurrent event data (Kelly and Lim, 2000). Models that deal with recurrent data and incorporate a fraity component include those of Block et al. (1985) and Stocker IV and Peña (2012).

Recurrent events can either be time-to-events model or gap times model. Time-to-events models focuses on occurrence rate of recurrent events over time; time is measured from time-origin to time of event of interest. Time-origin could be a fixed calendar time, onset of treatment, or a biological event. On the other hand, in gap time models, outcome variables of interest are gap times between events and is more relevant when cycling pattern of recurrent events is strong.

3.3.2 The Gap Time Model

Modeling hazard functions using calendar time domain has been the main conventional counting processes in studying past recurrences and are widely used for asymptotic theory derivation. When the focus switches to hazard function of gap time, it is more convenient to consider processes that restart the time clock every time a recurrence occurs. For example, a gap time counting process counts only the immediate event after the last recurrence. This leads us to a generalised minimal repair model Block et al. (1985) which offers a general framework gap-time events model.

Let a component start functioning at time 0. If the component fails at time t, either a perfect repair is done with probability $\rho(t)$ or a minimal repair is undertaken with probability $q(t) = 1 - \rho(t)$. A perfect repair restores the component to the original state whereas, a minimal repair restores the effective age to that just before the failure. The process of perfectly or minimally repairing the component takes place at each subsequent failure with the probability associated with the type of repair dependent on the effective age of the system. This model is also applicable to other areas since it admits as special cases some of the models commonly encountered in practice (i.e. $\rho(t) = 1$), under which we, recover the independent and identically distributed (i.i.d.) model. A common model used for recurrent events in the biomedical setting is the nonhomogeneous Poisson process which is a special case of the BBS model.

Let $\omega_0 = 0 < \omega_0 < \omega_1 < \cdots$ be the successive failure times of a component, and let $U_1, U_2, \cdots$ be a sequence of i.i.d. Uniform[0,1] random variables which are independent of the failure times. The sequence $(\omega_1, \omega_2, \cdots, \omega_v)$, where $v = \inf\{k \in (1,2,\cdots) : U_k < p(\omega_k)\}$, is an episode of the BBS model. Consider observing n independent BBS episodes $\{\omega_{jk} : 1 \leq j \leq n, 1 \leq k \leq v_j\}$ associated with n units where the j^{th} unit has a possibly time-dependent covariate process $X_j(\cdot)$.

Let the counting process $\mathbf{N} = \{N_1(t), N_2(t), \cdots, N_n(t)\}$ and $\mathbf{Y} = \{Y_1(t), Y_2(t), \cdots, Y_n(t)\}$,

where

$$N_j(t) = \sum_{k=1}^{\infty} I\{\omega_{jk} \leq t \wedge \omega_{jv}\} \tag{3.107}$$

and

$$Y_j(t) = I\{\omega_{jv_j} \geq t\}. \tag{3.108}$$

With respect to the filtration $F = \{\Im_t : t \in T\}$, where $F_t = F_0 \vee v_j = \ln \sigma\{(Nj(s), Yj(s)) : s \leq t\}$, and with $\Im_t$ containing all information available at time 0, the compensator of N is $A = \{(A_1(t;\beta), (A_2(t;\beta), \cdots, (A_n(t;\beta) : t \in T)\}$ with

$$A_j(t;\beta) = \int_0^t Y_j(s)\lambda(s)\exp\{\beta^t X_j\}ds \tag{3.109}$$

where $\lambda(s)$ is a baseline hazard function, β is a $q \times 1$ vector of regression coefficients, and $X_1(s), \cdots, X_n(s)$ are $q \times 1$ vectors of locally bounded predictable covariate processes.
The following condition is holds:

$$\int_0^{\infty} \rho(t)\lambda(t)\exp\{\beta^t X_j(t)\}dt = \infty \tag{3.110}$$

The condition above allows the waiting time to the first perfect repair to be almost surely finite with hazard rate function (Block et al., 1985; Agustin and Peña, 2005).
Let $W_j = T_j - T_{j-1}$ be the gap time between the $(j-1)^{st}$ and the j^{th} event. Papers dealing with BBS models as renewal processes include Agustin and Pena (1999); Aalen and Husebye (1991); Hollander et al. (1992); Adekpedjou et al. (2012); Dorado (1995); Dorado et al. (1997); Kijima (1989); Agustin and Peña (2001).

Renewal process are mostly applicable for gap-time and are defined as a process where (Cook and Lawless, 2007)

$$\lambda(t|H(t)) = h(t - T_{N(t^-)}) \tag{3.111}$$

Where $h(\cdot)$ is the hazard function for the gap time between events which are independent and identically distributed. Assuming covariates are exogenous, let $x^{(t)} = \{x(s) : 0 \leq s \leq t\}$ be the history of covariates over $[0,t]$ and x^{∞} is the complete covariate path. Assuming that $\lambda(t|H(t))$ depends only on $x^{(t)}$, we partition the interval as $[a,b]$ i.e. $\mu_0 < \mu_1 < \cdots < \mu_R$ and define $\triangle\mu_r = \mu_{r+1} - \mu_r$.

The product integral of $g(\mu)$ over $[a,b]$ is defined as

$$\prod_{[a,b]}\{1 + g(\mu)d\mu\} = \lim_{R\to\infty}\prod_{r=0}^{R}(1 + g(\mu_r)\triangle\mu_r) \tag{3.112}$$

We start by defining methods that are based on renewal processes then extend the methods to generalised situations where censored data exist.

Let an individual i be observed over the time interval $[o, \tau_i]$ and $t = 0$ refers to the start of the event process. The event intensity function is defined by Cook and Lawless (2007) as

$$\lambda(t|H(t)) = h(B(t)), t > 0 \tag{3.113}$$

Where $B(t) = t - T_{N(t^-)}$ is the time since the most recent event before t and $h(\omega)$ is the hazard function for the variable W_i. That is, if $W_i$have a common density function $f\omega$ and the survival function $S(\omega) = Pr\{W \geq \omega\}$ then

$$h(\omega) = \frac{f(\omega)}{S(\omega)} = \lim_{\triangle\omega\to 0}\frac{Pr\{W < \omega + \triangle\omega | W > \omega\}}{\triangle\omega} \tag{3.114}$$

We can allow the baseline hazard W_1 to have a different distribution from that of $W_2, W_3, \cdots$. Allowing for fixed covariates x_i means that the gap times W_{ij} between events have hazard functions $h(\omega|x_i)$. If n_i events are observed at time $0 < t_{i1} < t_{i2}, \cdots \leq \tau_i$. Let $\omega_{ij} = t_{ij} - t_{ij-1}, (j = 1, \cdots, n_i)$ and $\omega_{i,n_{i+1}} = \tau_i - t_{in_i}$ where $t_{i0} = 0$. These are the observed gap times for individual i with the final time being possibly censored. The likelihood function from

m independent individuals will take the form of

$$L = \prod_{i=1}^{m} \{ \prod_{j=1}^{n_i} h(\omega_{ij}|x_i) exp(-H(\omega_{ij}|x_i)) \} exp(-H(\omega_{ij}|x_i)) \tag{3.115}$$

where $H(\omega|x) = \int_0^{\omega} h(\mu|x) d\mu$ is the cumulative hazard function for W_{ij} given x_i. If τ_i is the censoring variable then the expression in equation 3.115 can be expressed as

$$Pr\{W_{i,1} = \omega_{i,1}, \cdots, W_{i,n_i} = \omega_{i,n_i}, W_{i,n_i+1} > \omega_{i,n_i+1}\} = P(\cdot). \tag{3.116}$$

The expression above (equation 3.116) denotes the probability density.
Let $f(\omega|x) = \{h(\omega|x) exp(-H(\omega|x))\} exp[-H(\omega|x)]$ be density function and $S(\omega|x) = exp[-H(\omega|x)]$ be survival function. Then

$$L = \prod_{i=1}^{m} \{ \prod_{j=1}^{n_i} f(\omega_{ij}|x_i) \cdot S(\omega_{i,n_{i+1}}|x_i) \tag{3.117}$$

which is the likelihood function for a random sample involving failure times $\omega_{ij}, (j = 1, 2, \cdots, n_i)$ and the right censored times $\omega_{i,n_i+1}, (i = 1, 2, \cdots, m)$.
If observation terminates after the n_i^{th} event i.e. $\omega_{i,n_i+1} = 0$, then the term $S(\omega_{i,n_i+1})$ disappears.

3.3.3 Smooth GOF tests for Baseline Hazard Functions in Recurrent Events

Gap times between recurrent events are common in many clinical and observational studies. Here, we extend Neyman's smooth test to assesses the hypothesis for recurrence (Liu et al., 2016) for the baseline hazards function. An extension of chapter 3 is used to obtain the score test statistic of the baseline hazard function. The asymptotic properties of the test have been examined in literature.

Suppose that there are n subjects and that each subject can experience K recurrent events. Assuming non-informative censoring, let T_{ik} be the time when the k^{th} failure occurs for the i^{th} subject and C_{ik} be the corresponding censoring time. T_{ik} is measured from an individual's entry into the study and the censoring C_{ik} occurs after the individual has been entered into a study to the right of the last known failure time; it is therefore, right censoring. When T_{ik} is subject to right censoring, the k^{th} failure time X_{ik} is a minimum of (T_{ik}, C_{ik}), i.e., X_{ik} is equal to T_{ik} if the event was observed and is equal to C_{ik} if it is censored. Let $\delta_{ik} = I(T_{ik} \leq C_{ik})$, where $I(\cdot)$ is an indicator function (Lim and Zhang, 2011) and takes the value 1 when $T_{ik} \leq C_{ik}$ and is 0 otherwise. Let Z_{ik} be a covariate vector of p-dimensions for the i^{th} subject at the k^{th} failure time. For each of the K failures, the hazard function for the i^{th} subject with respect to the k^{th} event, $l_{ik}(t)$, is assumed to take additive or multiplicative forms (Lim and Zhang, 2011).

$$\lambda_{ik}(t) = \lambda_{0k}(t - t_{k-1}) \exp\{\beta^t Z_{ik}(t)\}, \tag{3.118}$$

where t is the time since a patients study enrollment and t_{k-1} is the time of the $(k-1)^{th}$ failure. Note that $\lambda_k(t)$ are unspecified baseline hazard functions varying with $k = 1, \cdots, K$. Embedding the baseline hazard, $\lambda_{0k}(t - t_{k-1})$ in equation 3.120 to form a larger family of order k through smooth transformation.

$$\Phi_k = \{\lambda_k(\cdot;\theta) = \lambda_{0k}(\cdot) \exp[\theta^{'} \Psi(\cdot)] : \theta \in \Re^k\} \tag{3.119}$$

where k is a fixed positive integer and $\Psi(\cdot)$ is a $k \times 1$ vector of locally bounded predictable processes. The null hypothesis $H_0 : \lambda(\cdot) = \lambda_0(\cdot)$ is recovered when setting $\theta = 0$. The score process associated with θ have been derived by Pena (1998b), Pena (1998a), Agustin and Peña (2001).

3.3.4 Modelling Smooth Test in BBS model

Let $\{\omega_0 = 0, \omega_1, \omega_2, \cdots\}$ be a sequence of failure age generated under a minimal repair model with ω_i being continuously distributed with probability density function f and hazard function $\lambda = \frac{f}{\bar{S}}$, where $\bar{S}$ is the survival function. Considering $\omega_0 = 0 < \omega_1 < \omega_2 < \cdots$ are successive failure times of a component, and $U_1, U_2, \cdots$ is a sequence of i.i.d Uniform $[0,1]$ random variables which are independent of the failure times. The sequence $(\omega_0 = 0, \omega_1, \omega_2, \cdots, \omega_\nu)$, where $\nu = \inf\{k \in \{1,2,\cdots\} : U_k < (p\omega_k)\}$ is an epoch of the BBS model. Articles dealing with applications of BBS models include Aalen and Husebye (1991); Agustin and Peña (2005); Baxter et al. (1996); Dorado (1995); Dorado et al. (1997); Hollander et al. (1992); Kijima (1989); Nelson (2003); Agustin and Pena (1999); Block et al. (1985); Brown and Proschan (1983). Since a perfect repair restores a component to as good-as-new state, it suffices to observe a component only until the time of its first repair; a situation that is naturally similar to time to first loss to follow-up (LTFU) in a typically HIV clinical setting. BBS Model (Block et al., 1985) allows the probability of a perfect repair to depend on the age of the failed item. In the BBS model $\rho(\cdot)$ is a measurable function $\rho : [0,\infty] \to [0,1]$. The waiting time between perfect repairs is almost-surely finite with distribution H given by

$$H(t) = 1 - \exp\{-\int_o^t \frac{\rho(s)}{F(s)} dF(s)\} \tag{3.120}$$

Considering counting process $N(t) = \{(N_1(t), N_2(t), \cdots, N_n(t)) : t \in \digamma\}$ with $N_i(t) = N_i^*(t \wedge W_{iv_i}), i = 1,2,\cdots,n$. The corresponding observable filtration $\mathbf{F} = \{\digamma_t : t \in T\}$ is given by $\digamma_t = V_{j=1}^* \digamma^*_{j(t \wedge W_{iv_i})}$. The compensator of $N(t)$ is given as $\mathbf{A} = \{A_1(t), A_2(2), \cdots, A_n(t)\}$, with

$$A_i(t) = \int_t^0 Y_i(s)\lambda(s)ds, i = 1,2,\cdots,n, \tag{3.121}$$

where $Y_i(s) = I\{W_{iv_i}\}$ and $\lambda(\cdot)$ is unknown baseline hazard function.

Generally, the BBS model has two parameters Agustin and Pena (1999); a lifetime distribution function F, which in our case we assume here to be continuous, and a function $p[0,\infty) \to [0,1]$. Thus we write the model as BBS (F,p). Because there is a one-to-one correspondence between F and its hazard function Λ, given by

$$\Lambda(t) = \int_0^t \frac{dF(\omega)}{1-F(\omega)} = -\log(l-F(t)) \tag{3.122}$$

and

$$F(t) = 1-\exp\{-\Lambda(t)\}. \tag{3.123}$$

We can therefore specify a BBS model and rewrite it as BBS(Λ,p) (Agustin and Pena, 1999). Under a model of minimal repair, described at the introduction section, the sequence $\{\omega_j\}_{j=1}^{\infty}$ is a Markov process, and the conditional survivor function of ω_j given $\omega_0,\omega_1,\cdots,\omega_{j-1}$ is

$$\bar{S}(t|\omega_{j-l}) = \frac{\bar{S}(t)}{\bar{S}(\omega_{j-1}}, t \geq \omega_{j-1}, j \geq 1,$$

where $S = 1-F$ is the survivor function. Let $U_1,U_2,\cdots$ be a sequence of identically distributed and independent standard uniform variables, which are independent of the $\omega_j's$. Let $\nu = \min\{k \in \{1,2,\cdots\} : U_k \leq p(\omega_k)\}$. An epoch in the BBS$(\Lambda,p)$ model is the sequence $\omega_1,\omega_1,\cdots,\omega_\nu$. Because the system's effective age is restored to 0 after a perfect repair is performed, it suffices to observe the system until the ν^{th} failure, which occurs at time ω_ν. Hence we focus on the feature of an epoch of a BBS(Λ,p) model. The probability mass function ν for a BBS(Λ,p) is given by (see Agustin and Pena (1999))

$$f_\nu(k) = \frac{1}{(k-1)!}\int_0^\infty \exp\{-\Lambda(\omega)\} \times [\Lambda^*(\omega)]^{k-1} p(\omega)\lambda(\omega)d\omega, k = 1,2,\cdots \tag{3.124}$$

where $\Lambda^*(\omega) = \int_0^\omega \lambda^*(s)ds$ and $\lambda^*(s) = [1-p(s)]\lambda(s)$. Considering stochastic formulation developed by Hollander et al. (1992).

Let $\mathbf{N}^* = \{(\mathbf{N_1^*}(t), \mathbf{N_2^*}(t), \cdots, \mathbf{N_n^*}(t))\}$ be a multivariate counting process defined by $\mathbf{N_j^*}(t) = \sum_{k=1}^{\infty} I(\omega_{jk} \leq t), j = 1,2,\cdots$ and filtration $\mathbf{F}^* = \{\mathfrak{F}_t^* : t \in \mathfrak{F}\}$ defined by $\mathfrak{F}_t^* = \mathfrak{F}_o \bigvee \bigwedge_{j=1}^{n} \mathfrak{F}_{jt}^*$, where

$$\mathfrak{F}_{jt}^* = \sigma\{\{N_j^*(s) : s \leq t\} U \{U_{jk} : k \geq 1\}\}, \tag{3.125}$$

with $\mathfrak{F}_o$ containing all null sets of $\mathfrak{F}$ (Agustin and Peña, 2001). Consider observation of n independent BBS epochs (Agustin and Peña, 2005)$\{\omega_{jk} : 1 \leq j \leq n, 1 \leq k \leq \nu_j\}$ associated with n units where the j^{th} unit has time-dependent covariate process $\mathbf{X}_j(\cdot)$. Define the stochastic processes $\mathbf{N} = \{N_1(t), N_2(t), \cdots, N_n(t) : i \in \mathfrak{F}\}$ with $N_j(t) = N_j^*(t \wedge \omega_{j\nu_j}), j = 1,2,\cdots$ and the corresponding filtration $\mathbf{F} = \{\mathfrak{F}_t : t \in \mathfrak{F}\}$ is given by $\mathfrak{F}_t = \vee_{j=1}^{n} \mathfrak{F}_{j(t \wedge \omega_{j\vee_j})}^*$. The compensator $\mathbf{F}$ of $\mathbf{N}$ is given as $\mathbf{A} = \{A_1(t), A_2(t), \cdots, A_n(t) : t \in \mathfrak{F}\}$ with

$$A_j(t) = \sum_{0}^{t} Y_j(s)\lambda(s)ds, j = 1,2,\cdots,n, \tag{3.126}$$

where $Y_j(s) = I\{\omega_{j\nu_j} \geq s\}$, and $\lambda(\cdot)$ is the baseline hazard function.

The interest is to test the null hypothesis, $H_0 : \lambda(\cdot) = \lambda_0(\cdot)$, where $\lambda_0(\cdot)$ is completely specified. Nesting the null hypothesis to get the larger family of order k by smoothly transforming $\lambda_0(\cdot)$ yields

$$\beth_k = \{\lambda_k(\cdot;\theta) = \lambda_0(\cdot)\exp[\theta'\Psi(\cdot)] : \theta \in \Re^k\} \tag{3.127}$$

where k is a fixed positive integer and $\Psi(\cdot)$ is a $k \times 1$ vector of locally bounded predictable processes. The null hypothesis $H_0 : \lambda(\cdot) = \lambda_0(\cdot)$ is recovered when setting $\theta = 0$.

The score process associated with θ have been derived by Pena (1998b), Pena (1998a), Agustin and Peña (2001) to yield

$$U_\theta(t;\theta) = \sum_{j=1}^{n} \int_0^t \Psi(s)dM_j(s;\theta) \tag{3.128}$$

where $M_j(s;\theta) = N_j(s) - A_j(s;\theta),\ j = 1,2,\cdots,n$. The distribution of $U_\theta(t;\theta)$ under the null hypothesis can be achieved under four regularity conditions outlined in appendix B.

Under BBS model (Agustin and Peña, 2001) if the four conditions hold and $H_0 : \theta = 0$ is true, then as $n \to \infty$ converges to a normal distribution with mean zero and covariance matrix $\Sigma(\tau)$, based on theorem 2.1 in Agustin and Peña (2001). Through Pena (1998b), and assuming $X = 0$ and risk process $Y_j = I\{W_{iv_j} \geq s\}$, the asymptotic $\alpha-$ level smooth test of GOF of $H_0 : \lambda(\cdot) = \lambda_0(\cdot)$. We reject H_0 whenever

$$S(\tau) = \frac{1}{n} U_\theta(\tau;0)' \Sigma^{-1} U_\theta(\tau;0) \geq \chi^2_{k^*;\alpha}, \tag{3.129}$$

where $\Sigma^{-1}(\cdot)$ is a generalized inverse of $\Sigma(\cdot)$ and $\chi^2_{k^*;\alpha}$ is the $(1-\alpha)100\%$ of the chi-square distribution with degrees of freedom k =rank$[\Sigma(\tau)]$. Since $p(\cdot)$ is not known, $\Sigma(\tau)$ is estimated by

$$\hat{\Sigma(\tau)} = \frac{1}{n} \sum_{j=1}^{n} \int_0^\tau \psi(s)\psi(s)^t Y_j(s)\lambda_0(s)ds. \tag{3.130}$$

For comprehensive coverage on the processes of $S(\tau)$, polynomial specification (k), smoothing process of $S(\tau)$, achieved power, asymptotic properties, small sample properties, efficiency and other limiting distribution, see Agustin and Peña (2001), Pena (1998a), Agustin and Peña (2001), Baxter et al. (1996), Aalen and Husebye (1991) and Agustin and Pena (1999).

Chapter 4

Research Findings

4.1 Application of Smooth Test of GOF to Probability Distributions

4.1.1 Simulations

Here we begin by estimating the power of empirical GOF tests and smooth test when assessing the two parameter Weibull distribution. Specifically, we compare efficiencies of the smooth tests of order 3 and order 4 for the Weibull distribution against three common empirical distribution function (**EDF**) tests (**KS, CVM and AD**).

We analysed simulation results in order to compare the critical values and powers of empirical GOF tests (Anderson-Darling ($\mathbf{A}^2$), Kolmogorov-Smirnov test ($\mathbf{D}_n$) and Cramér-von Mises (ω^2)) and smooth test. All computations were performed using ***R package EW-GoF***. We generated samples from two-parameter Weibull Distribution with scale and shape parameters set at 30 and 6 respectively. The number of Monte Carlo runs in each situation was 1000. Samples of size $n \in \{5, 20, 50, 100, 500, 1000\}$ were generated and estimates of rejection probabilities computed. We eventually compared the performance of the GOF

tests against significance level of 1%, 5% and 10%. The power of the test was determined by the percentage rejection of the null hypothesis.

Similar tests have been examined by Bargal and Thomas (1983); Agustin and Peña (2001); Hollander et al. (1992); Agustin and Peña (2005) in an extensive simulation study. The performance of all the tests are strongly linked to the shape of the simulated distribution. Empirical tests ($\mathbf{A}^2$, ω^2 and $\mathbf{D}_n$) are biased in situations where $\hat{\beta}$ is fairly minimal (close to 1) whereas when $\hat{\eta}$ is sufficiently large i.e. $\hat{\eta} > 30$, they fail to detect the right distribution (Weibull Distribution). For smaller samples, smooth test tend to be unbiased compared to the empirical GOF tests. The components of smooth test tend to be unbiased when $\beta = 5$, $\eta = 35$ for large samples i.e. $n \geq 500$. The calculation of the asymptotic distributions of the EDF statistics follows $\phi_n = (F_n(z) - z)/\sqrt{n}$, where $F_n(z)$ is the EDF of the set of z_i and tends to standard normal distribution $\phi(z)$ as $n \to \infty$ and the statistics are functions of the process.

Table 4.1: The power of goodness-of-fit tests for the simple hypothesis H_0 (the Weibull distribution with parameters: scale=30 and shape=6) versus the hypothesis H_1 : a class of the Weibull distribution (for the smooth test) and Not Weibull (for EDF tests)

The power of Smooth Test (order 3)							
α	n=5	n=20	n=50	n=100	n=200	n=500	n=1,000
0.01	0.015	0.006	0.007	0.006	0.006	0.007	0.008
0.05	0.053	0.041	0.046	0.035	0.039	0.040	0.047
0.1	0.103	0.103	0.090	0.089	0.085	0.091	0.088
The power of Smooth Test (order 4)							
α	n=5	n=20	n=50	n=100	n=200	n=500	n=1,000
0.01	0.016	0.007	0.008	0.008	0.010	0.008	0.009
0.05	0.055	0.042	0.049	0.043	0.042	0.042	0.051
0.1	0.105	0.104	0.096	0.087	0.088	0.101	0.090
The power of Kolmogorov-Smirnov Test							
α	n=5	n=20	n=50	n=100	n=200	n=500	n=1,000
0.01	0.014	0.010	0.012	0.012	0.012	0.012	0.012
0.05	0.049	0.055	0.049	0.056	0.060	0.065	0.069
0.1	0.107	0.106	0.097	0.098	0.098	0.113	0.108
The power of Anderson-Darling Test							
α	n=5	n=20	n=50	n=100	n=200	n=500	n=1,000
0.01	0.013	0.010	0.010	0.011	0.011	0.013	0.013
0.05	0.047	0.052	0.048	0.055	0.055	0.060	0.066
0.1	0.113	0.122	0.095	0.104	0.115	0.118	0.120
The power of Cramer-Von Mises Test							
α	n=5	n=20	n=50	n=100	n=200	n=500	n=1,000
0.01	0.014	0.013	0.011	0.011	0.012	0.013	0.013
0.05	0.050	0.055	0.050	0.053	0.047	0.060	0.067
0.1	0.112	0.117	0.096	0.099	0.107	0.116	0.123

4.1.2 Modelling HIV Retention data-uncensored situation

Data Description

We conducted a retrospective data analysis for all patients who were initiated ART at two government hospitals in Nairobi, Kenya (Makadara Health Center and Lungalunga Health Center) between 1st of October 2011 to 31st December 2014. Considering that ART services were initiated in Kenya in 2003 across all government hospital, we specifically extracted data from 2011 because by then all the public systems, processes and structures for defaulter tracing were expected to have picked up effectively. Our event of interest was time to first LTFU. The clinical setting considered here is routine regular Comprehensive Care Center (CCC) in typical government hospitals. Data is collected routinely whenever patients come for clinical check-up or drug refill. Since time to first LTFU was the event of interest, other exits (i.e. transfer outs and deaths) were not considered in the analysis. Patients who were actively receiving ART services and did not experience the event were also removed. Only patients who were observed from the time of ART initiation between 1st November 2011 to 31st December, 2014 were included in the analysis. The time between ART initiation to first LTFU was given in months. Time to first LTFU was defined as missing routine clinical appointment within 48 hours from the scheduled appointment date and not identified as “Active on AR”, “dead”, or “transferred-out”. The time to first LTFU was calculated as the time interval between the dates of ART initiation and first drop out, as recorded by the ART database IQCare. The cohort was stratified into gender (male and female), WHO Staging (WHO Stage 1, WHO Stage 2, WHO Stage 3 and WHO Stage 4) considered at the time of ART initiation and age groups ($<$10 years, 10-14 years, 15-24 years and 25$+$ years). Data was retrieved from an Health Information System (HIS) called *IQCare* without patients identifiers. Only variables of interest were pulled out to excel spreadsheet. Data was stored in excel and thereafter analysed in R. Approval was obtained from Pathfinder International.

Modelling LTFU

The focus is on time to first LTFU, and data in this perspective is primary and has not been published or utilised in any publication. This is a typical Kenyan case, however, different types of LTFU are expected to reflect the general evolution of HIV programming. LTFU is expected to be a stable event that does not evolve much with time, at least in adults. However, in young children, the risk of the event is not likely. Heterogeneity is to be expected as it is well-known that there are various degrees of LTFU. The fact that LTFU is considered as a stable event in any HIV programming suggests that at least in adults there is no event-dependence and no time-dependence. The start time is the time of enrolment on ART. Patients are expected to come for drug refill and routine check-up. During the observation period, a patient can remain active (i.e. does not miss regular appointments), die, transferred-out or LTFU.

Cohort description

A total of 4,981 patients were initiated ART between November 1, 2011 and December 31, 2014 in two public hospitals. Out of those initiated on ART, 854 patients experienced LFTU and were therefore included in the analysis. The table below shows the patients' status.

Table 4.2: Patients' status

Status	Frequency	Percentage
Active on ART	2,392	48
Transfer Out	1,405	28
Dead	330	7
First Lost-to-follow up	854	17
Total	4,981	100

The median age of those lost to follow-up (n=854) was 34.2 years (IQR $30.4 - 38.4$), and 59% (n=509) of them were female. Forty five percent of patients had advanced/severe immunodeficiency at the start of treatment, and 20% had WHO clinical stage 3 or 4 disease. The mean CD4 count was 449 (SD 9.3) at baseline. Characteristics at baseline during ART initiation is given below.

Table 4.3: Patients Baseline Characteristics at ART initiation

Patients Characteristics	$<$ 15 years (Children) (%)	15-24 years (Adolescents) (%)	25+ years (Adults) (%)
Gender (N=854)	87(4%)	458(52%)	309(34%)
Male (N=345)	45(5%)	185(28%)	115(68%)
Female (N=509)	42(8%)	273(54%)	194(38%)
WHO Staging (N=854)	60(7%)	258(30%)	536(63%)
WHO Stage I(N=376)	18(5%)	112 (30%)	216(65%)
WHO Stage II(N=308)	29(9%)	87(28%)	192(62%)
WHO Stage III(N=136)	8 (6%)	41(30%)	87 (87%)
WHO Stage IV(N=34)	5(15%)	18 (53%)	81 (81%)

Graphical Assessment

We obtained probability plots to assess the validity of statistical distributions (Kimber, 1985)to time to first LTFU data (Computing, 1991). Graphically, the Weibull distribution seems to be close to the P-P plot line compared to the Gamma and log-normal distribution. See Figure 4.1, 4.2, 4.3 and 4.4.

Figure 4.1: Test for Theoretical Distributions. The Weibull distribution is closer to the distribution

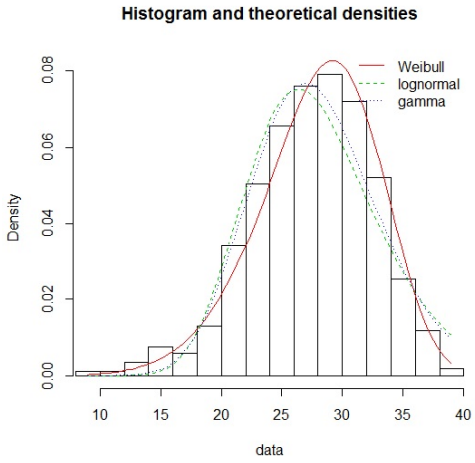


Figure 4.2: Test for Theoretical Distributions. Here, the data appears to be more coherent with Weibull distribution

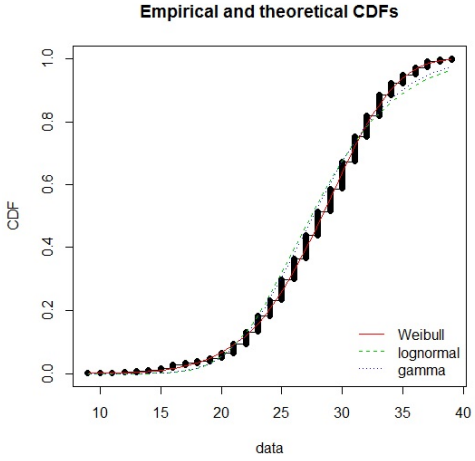


Figure 4.3: Test for Theoretical Distributions. The PP-plot indicates that the Weibull distribution is the ideal distribution.

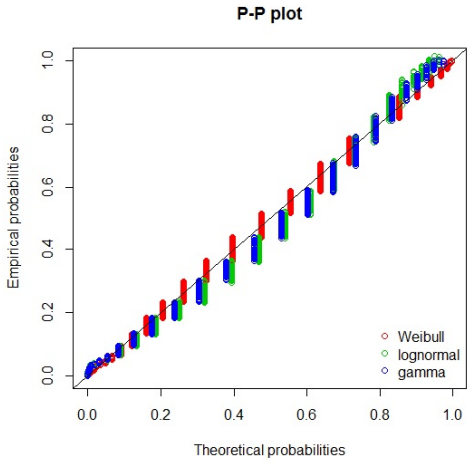
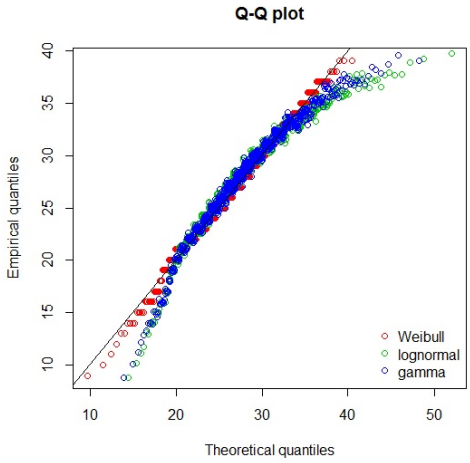


Figure 4.4: Test for Theoretical Distributions. The QQ-plot shows the Weibull distribution is much coherent with the data.



Model fitting

In order to assess the model fit for the Weibull distribution, we obtained the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). We also fitted the log-normal and the Gamma distributions to the same data. The AIC and BIC for the Weibull distribution (AIC = 5,276.6, BIC = 5,286.1) are lower than that of the Gamma distribution (AIC = 5,777.8, BIC = 5,787.3) and the log-normal (AIC = 6,009.4, BIC = 6,018.9) distributions. Therefore the model that fits the data best is the Weibull distribution.

Table 4.4: Comparison of the AIC and BIC for the Weibull, Gamma and Lognormal Distributions. All the parameter estimates were obtained by the MLE method

	Weibull	Gamma	Log-normal
Parameter	shape=6.786; scale=30.145	shape=14.958; rate=0.532	meanlog*=3.303; sdlog**=0.299
AIC	5,276.633	5,777.821	6,009.374
BIC	5,286.133	5,787.321	6,018.874
* is the mean of the natural logarithm of LTFU, **is the standard deviation of the natural logarithm of LTFU			

Tests Performance

The smooth test generated here is constructed using orthonormal functions as opposed to quadratic forms (Bargal and Thomas, 1983). For the smooth test, the score statistics of order 3 and order 4 are given. The hypothesis regarding the distributional form is rejected by the three empirical distribution tests if their respective test statistic, D_n , ω^2 and A_n are greater than the critical value obtained from their tabulated values. Also their p-values are considerably lower than the significance level of 0.01. These tests (D_n , ω^2 and A_n) are more powerful whenever the sample size is not large. In our situation, however, with a sample size of size of 854, the tests are misleading. The smooth test, on the other hand, rejects the hypothesis when considering score statistics up to order four and the p-value is quite large compared with the D_n , ω^2 and A_n tests.

Table 4.5: Tests comparison (N = 864) for the One-sample Kolmogorov-Smirnov test, Cramer-von Mises test, Anderson-Darling test and smooth test of order 3 and order 4. H_0: Weibull distribution (6,30) vs H_1: Not Weibull distribution for AD, KS and CVM and Generalised Weibull distribution for smooth test.

Test type	Test Statistics	p-value
One-sample Kolmogorov-Smirnov test	D_n = 0.055232	0.01092
Cramer-von Mises test	$\omega^2 = 1.0238$	7.947e-12
Anderson-Darling test	A_n = 1.9603	0.09659
S_3	S = 1.2529	0.308
S_4	S = 0.66308	0.409

4.1.3 Application Results

To demonstrate the importance of the smooth test of goodness-of-fit in a real life application, we examined an HIV retention data and fit a two parameter Weibull distribution to LFTU data. We assessed the fit using smooth tests of order 3 and 4 and then compared the results with the three empirical GOF tests. Other exits from the program (i.e. death, transfer-outs and active-on-ART) were removed. Essentially, we tested the null hypothesis that the Weibull distribution is the underlying distribution of time to first LTFU. The maximum likelihood estimates of the scale and shape parameters under the Weibull model were $\eta = 30.145$ and $\beta = 6.786$, respectively, and the resulting values of the test statistics were $S_3 = 1.2529, p = 0.308)$ and $S_4 = 1.66308, p = 0.409)$. Hence the Weibull null hypothesis could not be rejected when using smooth tests, suggesting that the Weibull model is the best model for the duration between the start of ART and first LTFU. In comparison, the three empirical GOF tests rejected the null hypothesis. The one-sample Kolmogorov-Smirnov test ($D_n = 0.055232, p = 0.01092$), Cramer-von Mises test ($\omega^2 = 1.0238, p = 7.947e-12$) and the Anderson-Darling test ($A_n = 1.9603, p = 0.09659$ indicated significance deviations

from the null distribution. This suggests that the smooth test is the most reliable test compared with the rest whenever the sample size is sufficiently large.

4.2 Application of Smooth Tests to Hazard Functions: two-Sample Problem

4.2.1 Simulations

Similarly, in this section we conducted a simulation study to ascertain proportionality under right censoring in the CPH model. Independent samples of size 10, 50, 100, 200, 500 and 1,000 were simulated and adjusted to give a chosen percentage of censored observations before the end of follow-up (i.e. 25% to 35% censoring, 45% to 55% censoring and 65% to 75% censoring). Each simulated dataset had a treatment covariate stratified by group (i.e. 1 or 2) and one other covariate arranged to contain equal numbers of observations. The power of the test was calculated as the percentage of rejection at the 5% level of significance. All simulations and comparative analyses were performed using the ***R*** packages ***survival, eha, prodlim*** and ***surv2sample***. For each sample size (i.e $n \in (10, 50, 100, 200, 500, 1000)$), 1,000 samples were generated and percentage rejection was computed as the number of cases rejected (with $p < 0.05$). Simulation studies show that as the censoring percentage increases, the percentage rejection of the global and Kolmogorov-Smirnov tests increases. Also the Kolmogorov-Smirnov test is strongly affected by sample size, such that as the sample size increases, percentage rejection increases as well. It fails to detect proportionality. The smooth tests are not affected by the percentage of censoring and sample size (see Fig. 4.5 to 4.10).

Figure 4.5: Graphs of percentage rejections for 1000 simulations with varying censoring of sample size 10

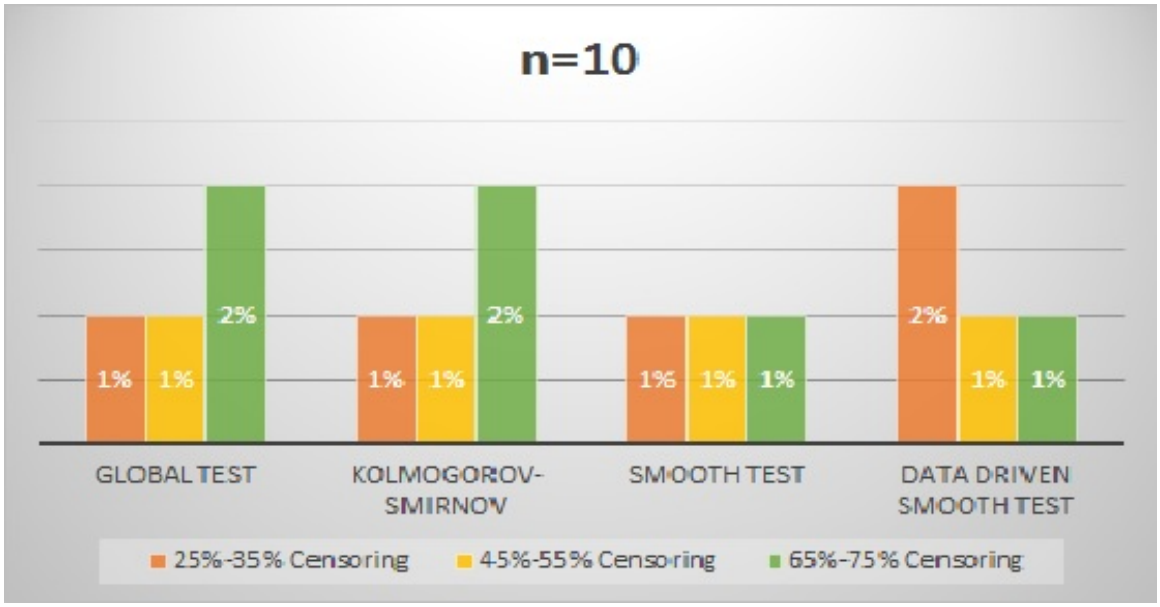


Figure 4.6: Graphs of percentage rejections for 1000 simulations with varying censoring of sample size 50

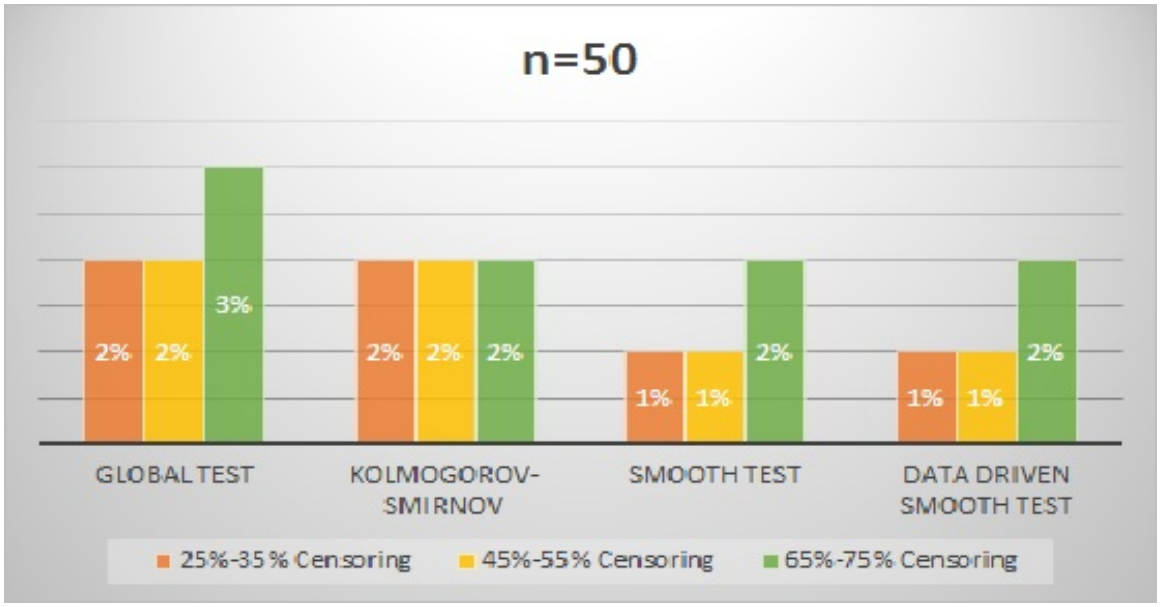


Figure 4.7: Graphs of percentage rejections for 1000 simulations with varying censoring of sample size 100

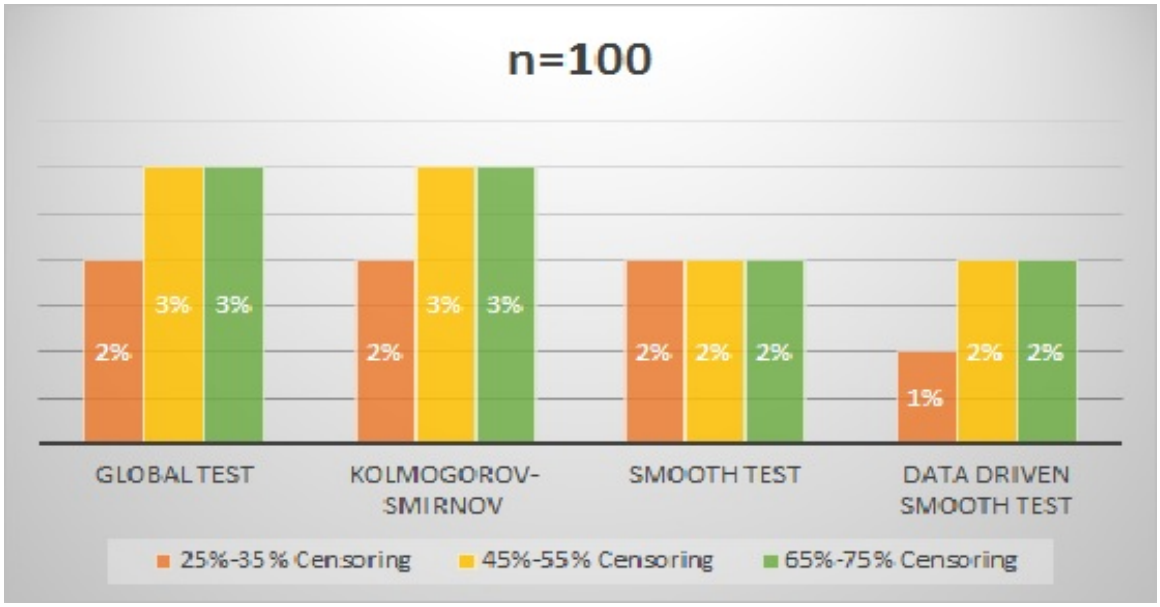


Figure 4.8: Graphs of percentage rejections for 1000 simulations with varying censoring of sample size 200

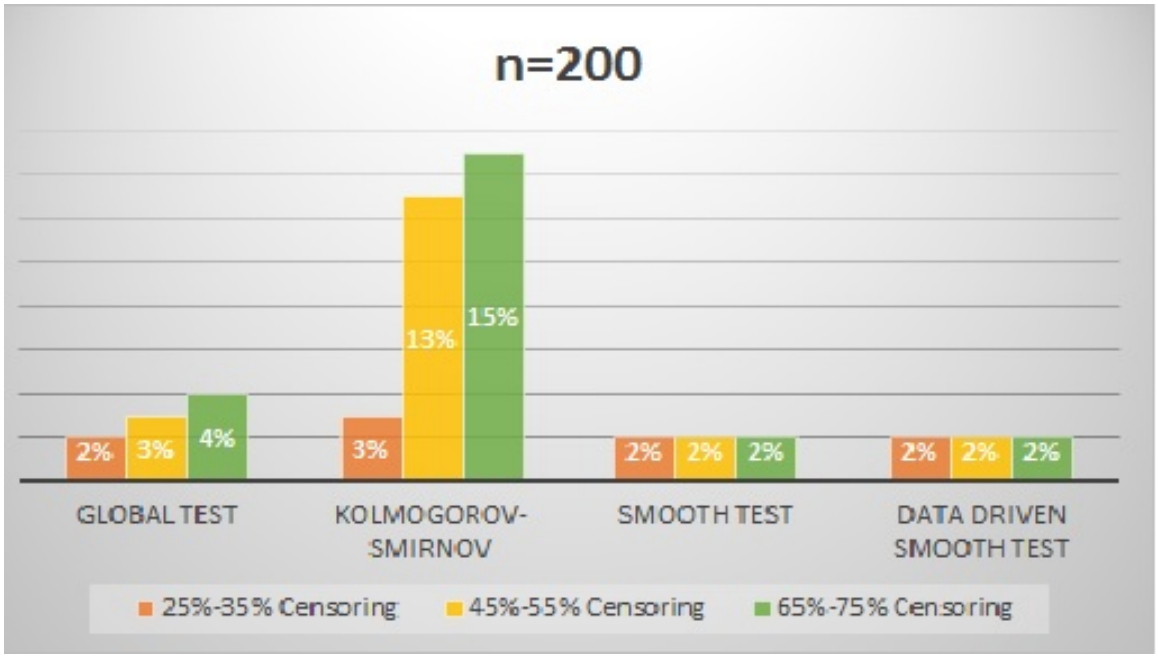


Figure 4.9: Graphs of percentage rejections for 1000 simulations with varying censoring of sample size 500

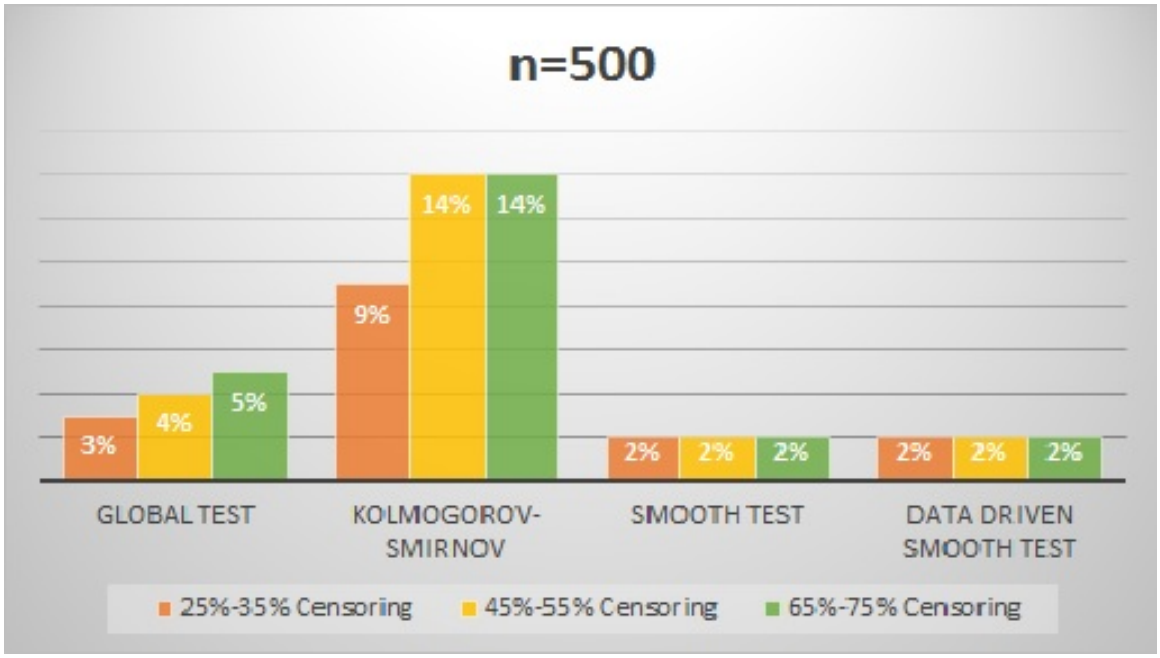
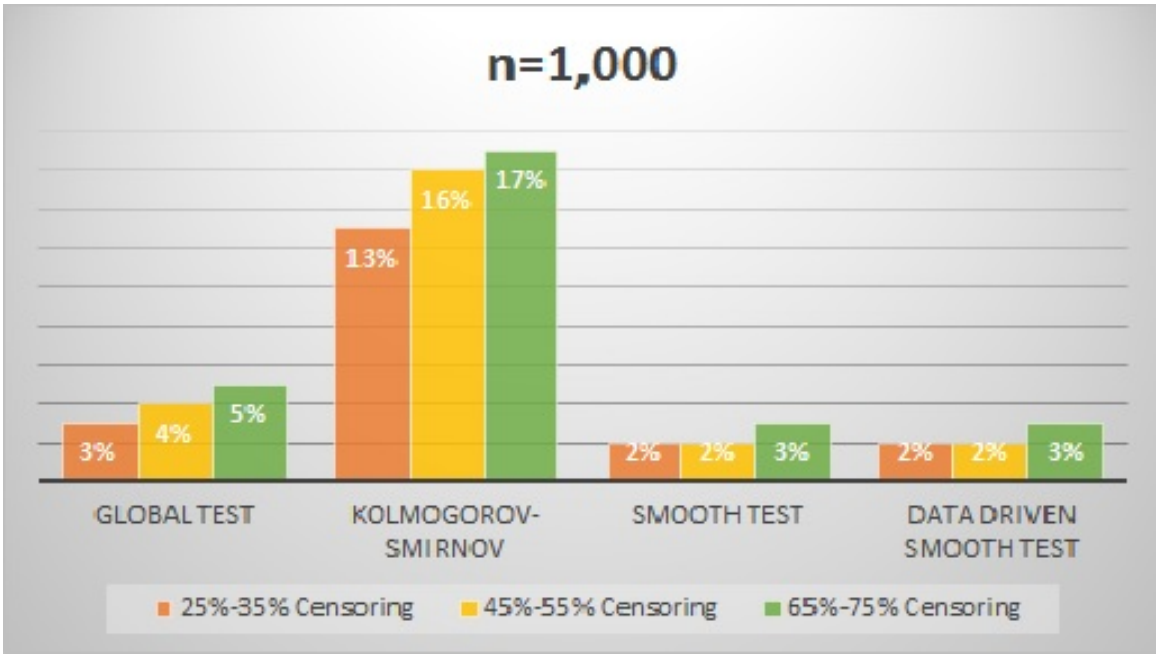


Figure 4.10: Graphs of percentage rejections for 1000 simulations with varying censoring of sample size 1,000



4.2.2 Data Setting and Analysis of Cancer Studies

Dataset 1: Survival with Malignant Melanoma

This dataset consists of measurements made on patients with malignant melanoma. Each patient had their tumour removed by surgery at the Department of Plastic Surgery, University Hospital of Odense, Denmark during the period 1962 to 1977. The surgery consisted of complete removal of the tumour together with about 2.5cm of the surrounding skin. Measurements taken included the thickness of the tumour and whether it was ulcerated or not. Patients were followed until the end of 1977. Time was defined as survival time in days since the operation, possibly censored. The patients' status at the end of the study were death from melanoma, alive and death from causes unrelated to their melanoma. Other variables include survival time in days since the operation; the patients status at the end of the study, 1 indicates that they had died from melanoma, 2 indicates that they were still

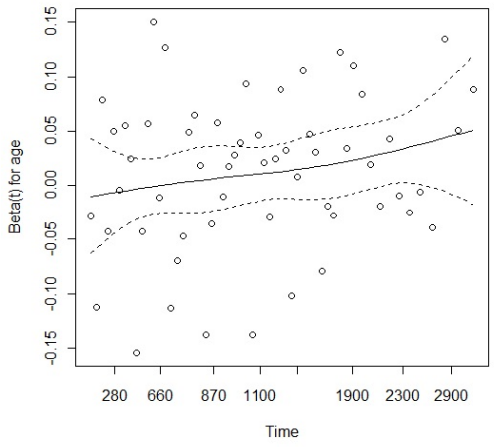
alive and 3 indicates that they had died from causes unrelated to their melanoma (status);the patients sex 1=male, 0=female(sex); age in years at the time of the operation (age); year of operation (year). tumour thickness in mm (thickness); indicator of ulceration; 1=present, 0=absent (ulcer). Data is described in Andersen et al. (2012).
We begin by fitting the Cox PH model with sex, thick, ulcer and age as covariates.

Table 4.6: Fitting Cox model to Malignant Melanoma data

Covariate	β	chisq	p-value
sex	0.151	1.35	0.2456
tumour thickness	-0.249	3.02	0.0823
ulceration	0.163	1.52	0.2182
age	0.207	3.08	0.0791

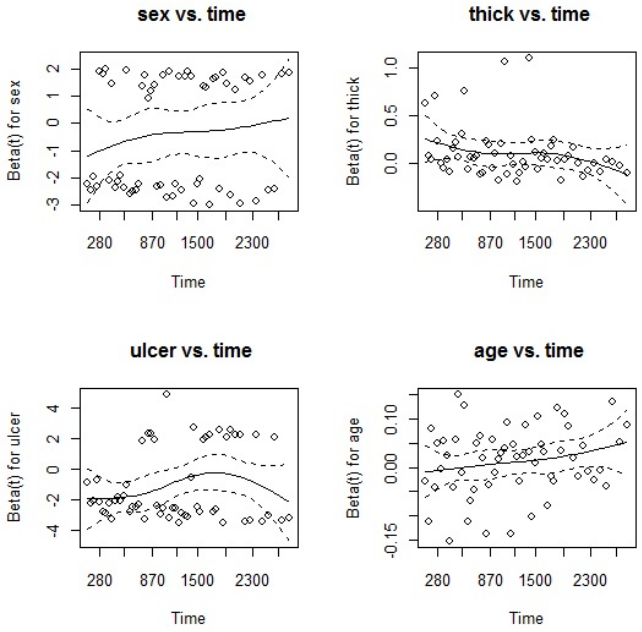
From above (Table 4.6), all the covariates are insignificant at $\alpha = 0.05$. Our focus, however, was on testing the proportionality assumption and so we created the plots of the Schoenfeld residuals versus log(time) for the overall fit. Testing the time-dependent covariates is equivalent to testing for a non-zero slope. A non-zero slope indicates a violation of the proportional hazards assumption. We started by looking at the graphs of the Cox regression models before performing the tests of non-zero slopes.

Figure 4.11: Schoenfeld residuals versus Time for the overall fit: Malignant Melanoma Data



The overall fit of the CPH model shows residuals scattered all over with a general zero slope (Figure 4.11). Hence proportionality exists despite the fact that the covariates are insignificant. The next step was to create Schoenfeld residual plots for each of the four covariates, including a lowess smoothing curve. The graphs for the residuals were still scattered for the four covariates (Figure 4.12).

Figure 4.12: Schoenfeld residuals versus Time for each covariates: Malignant Melanoma Data



Like in the plots, we expect all tests to fail to reject the null hypothesis, indicating that the proportionality assumption holds. We then compared the power of rejection between Kolmogorov-Smirnov test for proportional hazard, the smooth tests (Legendre polynomials with $d = 3$ with 3 degrees of freedom), data-driven smooth test (Legendre polynomials as the basis functions, nested with 5 dimensions) and the global test for all the interactions tested at once. Note that a p-value less than 0.05 indicates a violation of the proportionality assumption.

Table 4.7: Tests of Hazard Proportionality in CPH: Melanoma cancer data

Test	Statisic	p-value
Global test	10.03	0.04
Two-sample Kolmogorov-Smirnov test	3.89	0.09
Smooth test of order 3	5.84	0.12
Data-driven Smooth test	2.28	0.15

Both the smooth test of order 3 and the data-driven version fail to reject the null hypothesis, with p-values of 0.12 and 0.15, respectively, whereas the global test rejects the null hypothesis at $\alpha < 0.05$. On the other hand, the Kolmogorov-Smirnov test also fails to reject the null at $\alpha < 0.05$ but does not do well at $\alpha < 0.1$ (null hypothesis is rejected).

Dataset 2: Cohort Study On Breast Cancer Patients From Netherlands

This dataset contains follow-up data on 2,982 women with breast cancer who went through breast surgery. The women were followed from the time of surgery until death, relapse or censoring. Only female patients diagnosed with primary epithelial breast cancer between 1 January 1990 and 31 December 2010 were selected from the Netherlands Cancer Registry (NCR). The register is a population-based independent cancer registry containing clinical administrative data of every newly diagnosed cancer patient in the Netherlands. Topography and morphology is coded according to the International Classification of Diseases for Oncology and staging according to the TNM-classification. Patients were included from hospitals in the Northern Netherlands and the Rotterdam region. Patients from hospitals from other regions that never participated before 2009 were included in the control group. Patients that were diagnosed with neuroendocrine tumors, synchronous tumors, diagnosed at autopsy and that had any type of previous malignancy were excluded. Hospitals from the intervention group were categorized by the implementation proportion (IP)

of recommendations that were given in the final reports of each peer review. Rating the implementation was performed by studying final reports from subsequent reviews, follow-up correspondence, hospital documents and interviews with shareholders when necessary. Implementation of a recommendation was ranked on a scale from 0 to 4. The IP per hospital was expressed as a percentage of the total possible score. When implementation of a recommendation could not be determined (lost to follow-up), this recommendation was subtracted from the total possible score. The average IP of all peer reviews per hospital was used because it is not known what the time period is in which changes based on organizational change can occur and quality improvement is a continuous process. Ranking the implementation of recommendations was performed by the principal investigator and is described in. Other variables are defined as follows: patient ID number(pid);year of breast surgery (i.e. year of enrollment into the study), between the years 1978-1993 (year); relapse free interval measured in months(rf); relapse indicator(rfi); metastasis free (m); metastasis status(mfi); overall survival(os); overall survival indicator(osi);age at surgery measured in years (age); menopausal status with levels "pre" and "post" (meno); tumor size in three classes: $<$=20mm, $>$20-50mmm and $>$50mm (size); differentiation grade with levels 2 or 3(grade); progesterone receptors, fmol/l (pr) oestrogen receptors, fmol/l (er); the number of positive lymph nodes (nodes); hormonal therapy with levels "no" and "yes" (hormon); categorical variable indicating whether the patient received chemotherapy or not, with levels "no" and "yes" (chemo); a numeric indicator of whether the tumor was discovered recently with levels "1978-87" and "1988-93" (recent); a numerical indicator of whether the patient did not received chemotherapy (no.chemo). Date is described in Moons et al. (2009).

We begin by fitting the Cox PH model with sex, thick, ulcer and age as covariates.

Table 4.8: Fitting Cox model: Breast Cancer data

Covariate	β	chisq	p-value
age	0.0519	4.28	3.85e-02
menopausal status with levels ”pre” and ”post” (meno)	0.0293	1.10	2.94e-01
progesterone receptors, fmol/l (pr) oestrogen receptors, fmol/l (er)	0.1375	26.10	3.24e-07
differentiation grade with levels 2 or 3(grade)	-0.0304	1.20	2.74e-01
the number of positive lymph nodes (nodes)	-0.0569	2.54	1.11e-01
progesterone receptors, fmol/l (pr)	0.1077	18.42	1.77e-05

From the table (Table 4.8), all the covariates are significant at $\alpha = 0.5$. Figure 3 below shows the Schoenfeld residuals versus log(time) plot for the overall fit. The solid line is a smoothing spline fit to the plot, with the broken lines representing a $\pm$2-standard-error band around the fit.

Figure 4.13: Schoenfeld residuals versus Time for overall fit: Breast Cancer Patient Data

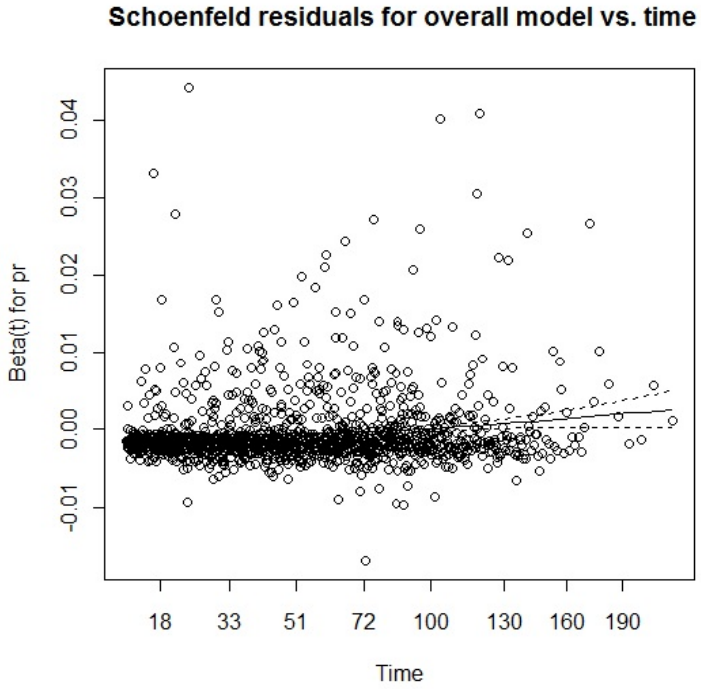
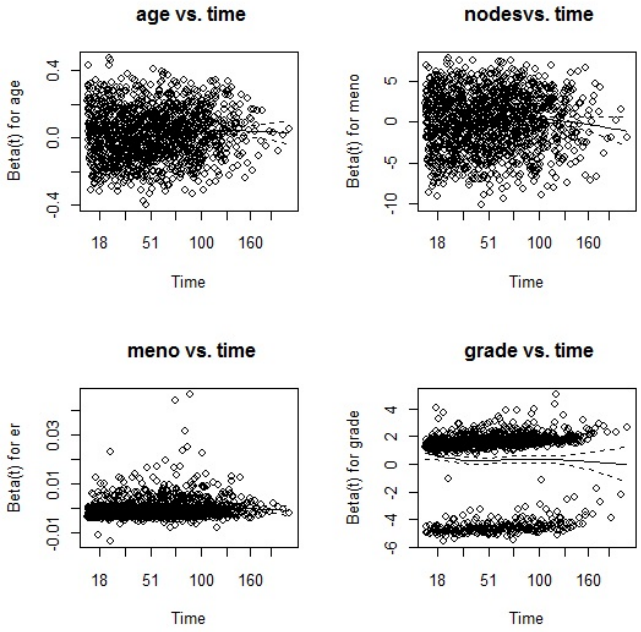


Figure 4.14: Schoenfeld residuals versus Time for the covariates: Breast Cancer Patient Data



The Schoenfeld residual plots show scatter plots with general non-zero slopes, indicating time-dependence (Figure 4.13 and Figure 4.14). The proportionality assumption does not hold in this dataset. Table 4.9 shows the proportionality tests for this dataset. We also compared the power of rejection for the Kolmogorov-Smirnov test for proportional hazard, the smooth tests (Legendre $d = 3$ with 3 degrees of freedom), the data-driven smooth test (Legendre functions as basis, nested with 5 dimensions) and the global test for all the interactions tested at once.

Table 4.9: Tests of Proportionality: Breast Cancer Patient data

Test	Statisic	p-value
Global test	83.48	6.66e-16
Two-sample Kolmogorov-Smirnov test	29.63	0.01
Smooth test of order 3	17.73	0.00
Data-driven Smooth test	8.48	0.01

All the four tests are consistent in the rejection of the null hypothesis, which is supported by the Schoenfeld residual plots as well.

Dataset 3: Ovarian Cancer Survival Data

Between mid-1974 to mid-1977, 82 patients with advanced ovarian carcinoma and 29 patients with minimal residual disease were followed. Patients included in the minimal disease group had surgical excision of all tumor > 2 cm in diameter at the time of total abdominal hysterectomy, bilateral salpingo-oophorectomy and omentectomy within one month before enrolment. Following surgery they were classified according to the distribution of residual diseases in arbitral defined stages II to IIIA. All patients in each of the groups had histologically proved epithelial type ovarian carcinoma and all had adequate renal hepatic and marrow functions. The dataset is described in Edmonson et al. (1979).

Survival in a randomised trial comparing two treatments for ovarian cancer. The variables includes survival or censoring time (futime); censoring status (fustat); age in years; residual disease present (1=no,2=yes)(resid.ds); treatment group (rx); ECOG performance status (1 is better) (ecog.ps). We fit a Cox PH model for futime and fustat with covariates age, ecog.ps, rx and resid.ds.

Table 4.10: Schoenfeld residuals versus Time for the overall fit

Covariates	β	chisq	p-value
age in years	-0.0399	0.0262	0.871
ECOG performance status (1 is better) (ecog.ps)	0.4845	1.8819	0.170
treatment group (rx)	0.1325	0.2001	0.655
residual disease present (1=no,2=yes)(resid.ds)	-0.1417	0.2463	0.620

From Table 4.10, all covariates are insignificant at $\alpha = 0.05$.

For testing the proportionality assumption, we plotted the Schoenfeld residuals versus log(time) for the overall fit and each of the four covariates (Figure 4.15 and Figure 4.16).

Figure 4.15: Schoenfeld residuals versus Time for overall fit: Ovarian Cancer Survival Data

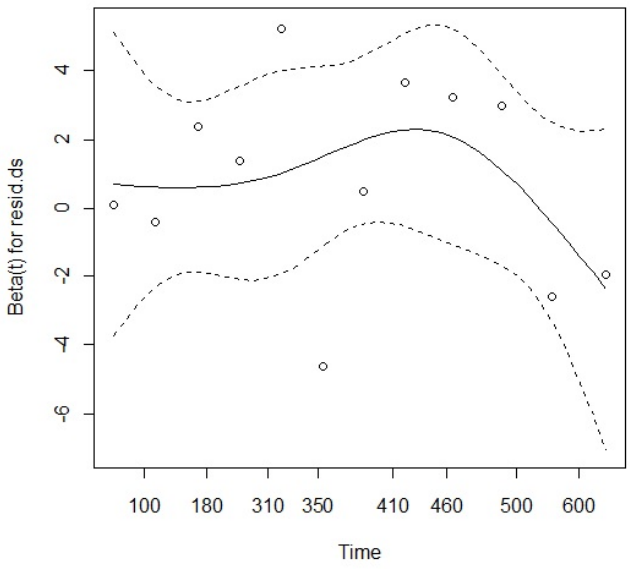
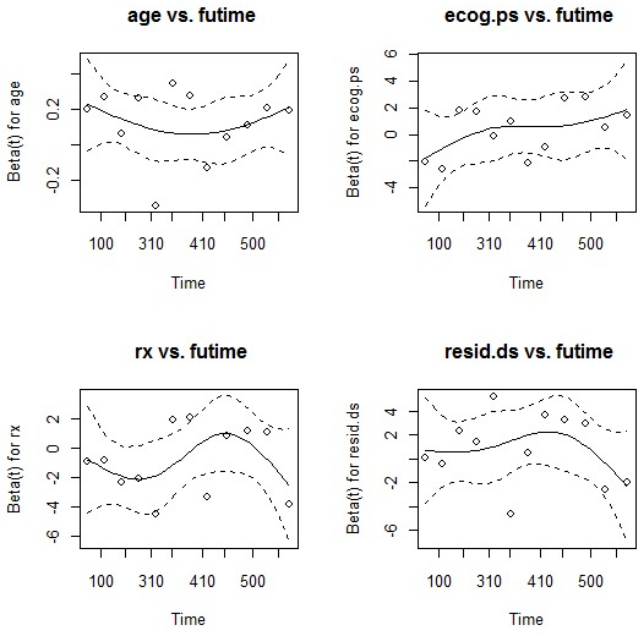


Figure 4.16: Schoenfeld residuals versus Time for each covariate



The Schoenfeld residual plots show non-zero slopes, suggesting time-dependence. The proportionality assumption holds in this dataset. Table 4.11 shows the proportionality tests for this dataset. We compared the power of rejection between Kolmogorov-Smirnov test for proportional hazard, the smooth tests (Legendre $d = 3$ with 3 degrees of freedom), data-driven smooth test (Legendre functions as basis, nested with 5 dimensions) and the global test for all the interactions tested at once.

Table 4.11: Tests of Proportionality: Ovarian Cancer Data

Test	Statisic	p-value
Global test	3.36	0.50
Two-sample Kolmogorov-Smirnov test	1.89	0.07
Smooth test of order 3	4.84	0.18
Data-driven Smooth test	2.23	0.20

The global test, smooth tests fixed dimension and data-driven smooth test fail to reject the null hypothesis. This is in agreement with the Schoenfeld residual plots for the general zero slope. However, the two-sample Kolmogorov-Smirnov test rejects the null hypothesis at $\alpha < 10\%$. This is misleading and inconsistent with the Schoenfeld residual plots.

Dataset 4: Remission Times for Acute Myelogenous Leukaemia

Acute myeloid leukemia (AML) represents a group of clonal hematopoietic stem cell disorders in which both a block in differentiation and unchecked proliferation result in the accumulation of myeloblasts at the expense of normal hematopoietic precursors. The patients in the study of maintenance therapy included 22 adults with AML, two with promyelocytic leukemia and two who had subacute myelogenous leukemia before conversion to classical AML. Patients had received no previous therapy for AML and there had been complete remission with standardized induction regimens supervised by the Stanford University Hematology Division. The median age of patients entered on the study was 45 years, with a range of 18 to 72 years. The induction program was modified from the programs of Clarkson, Gee and colleagues by the addition of daunarubicin. With minor modifications, therapy was administered as follows: daunarubicin, 60 mg per sq meter by rapid intravenous infusion, was given on the first day. This was followed in 12 hours by cytarabine, 3 mg per kg of body weight by rapid intravenous infusion, and 6-thioguanine, 2.5 mg per kg of body weight given orally. Administration of the last two agents was continued every 12 hours until biopsy-proven marrow hypoplasia was achieved. A second dose of daunarubicin between days 7 and 10 was nearly always given, the dose varying, depending on the cellularity of a marrow biopsy specimen. Changes in therapy from the original program were undertaken so as to shorten the treatment program and decrease the time at risk from severe neutropenia and thrombopenia. That this was achieved is reflected in the shorter treatment period required to reach hypoplasia with the current drug program compared with earlier regimen employing only a single daily dose of cytarabine and 6-thioguanine. The question

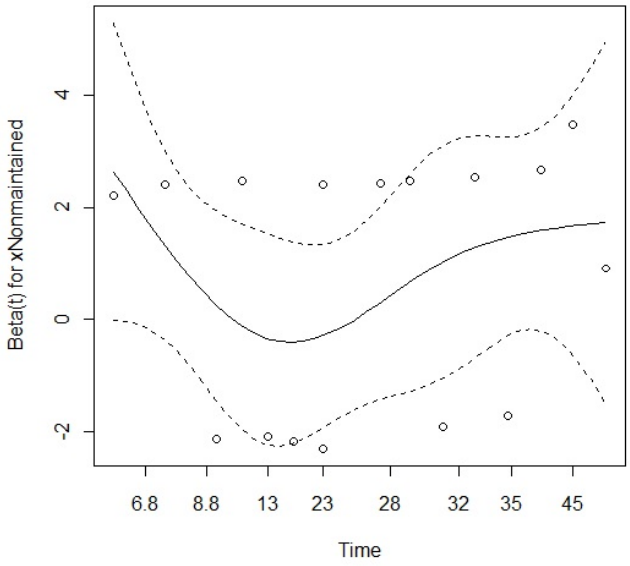
at the time was whether the standard course of chemotherapy should be extended ('maintenance') for additional cycles. The dataset is described in Miller Jr (2011). Variables and covariates are defined as The length of the complete remission-in weeks (time); an indicator of right censoring. 1 indicates that the patient had a relapse and so time is the length of the remission. 0 indicates that the patient had left the study or was still in remission in October 1974, that is the length of remission is right-censored (cens); The group into which the patient was randomized-Group 1 received maintenance chemotherapy, group 2 did not (x). We fitted a Cox PH model for remission time and status with covariate X, representing 'maintenance' or 'non-maintenance' of patients in chemotherapy.

Table 4.12: Fitting the CPH model: Acute Myelogenous Leukaemia Data

Covariates	β	chisq	p-value
x('non-maintenance)'	0.0198	0.00691	0.934

The covariate in this case ('non-maintenance') is not significant.

Figure 4.17: Schoenfeld residuals versus Time for overall fit: Acute Myelogenous Leukaemia Data



The Schoenfeld residual plots show a general zero-slope indicating proportionality. The proportionality tests are indicated in Table 8 below. We compared the power of rejection between Kolmogorov-Smirnov test for proportional hazard, the smooth tests (Legendre $d = 3$ with 3 degrees of freedom), data-driven smooth test (Legendre functions as basis, nested with 5 dimensions). The global tests did not yield any result.

Table 4.13: Tests of Proportionality: Acute Myelogenous Leukaemia Data

Test	Statisic	p-value
Global test	NA	NA
Two-sample Kolmogorov-Smirnov test	1.131	0.63
Smooth test of order 3	3.07	0.38
Data-driven Smooth test	0.15	0.75

The global test did not give any result but the other 3 tests (i.e. two-sample Kolmogorov-Smirnov test, smooth test and data-driven smooth test) failed rejects the null hypothesis. That is, they detected proportionality.

Dataset 5: North Central Cancer Treatment Group Lung Cancer Data

This data shows survival of patients with advanced lung cancer from the North Central Cancer Treatment Group (NCCTG). The study looked at how performance scores can rate how well a patient performs usual daily activities. An initial detailed questionnaire was administered to approximately 150 patients with advanced cancer. This questionnaire was subsequently revised and given to a total of 1,115 patients with advanced colorectal or lung cancer. Thirty six variables showed significant prognostic information for survival in univariate analyses, even though many of these variables were associated with only a minimal increase in risk. A multivariate analysis demonstrated that there was a high correlation between many variables. Three major groups of variables became apparent as providing strong prognostic information (i.e. physician's assessment, patient's assessment and nutritional factor such as appetite). Data contained 228 patients with advanced lung cancer and includes measurements of the survival time in days, as well as other demographic and biological information for each patient. Variables such as weight loss was categorized by quartiles, and ECOG scores were grouped into categories with subjects rated as either 0/1 or 2/3, with 0/1 representing the best and 2/3 representing a poor score. The data set was 28% censored, with a median observed failure time of 256 days. The baseline group (n = 16) were males with ECOG scores equal to 1 and a weight loss measure in the first quartile (Loprinzi et al., 1994). Variables and covariates were defined as institution code (inst); survival time in days (time); censoring status 1=censored, 2=dead (status); age in years (age); Male=1 Female=2 (sex); ECOG performance score (0=good, 5=dead) (ph.ecog); Karnofsky performance score (bad=0-good=100) rated by physician (ph.karno); Karnofsky performance score as rated by patient (pat.karno); Calories consumed at meals (meal.cal);

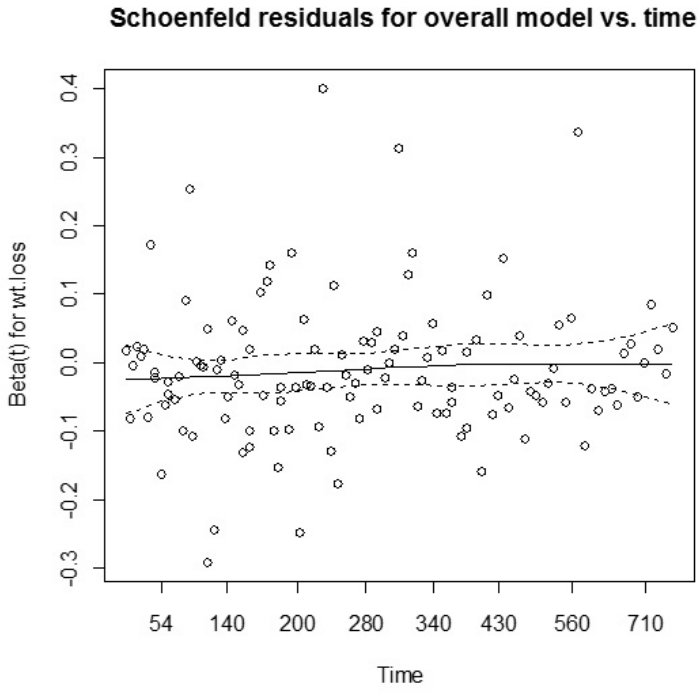
Weight loss in last six months (wt.loss). We fit a Cox PH model for time and status with covariates age, ecog.ps, rx and resid.ds.

Table 4.14: Fitting Cox PH model to NCCTG Lung Cancer Data

Covariates	β	chisq	p-value
age in years	0.0710	0.6553	0.4182
sex	0.1773	3.7609	0.0525
ECOG performance score (0=good, 5=dead) (ph.ecog)	-0.0189	0.0491	0.8247
Karnofsky performance score (bad=0-good=100) rated by physician (ph.karno)	0.1718	2.5791	0.1083
Karnofsky performance score as rated by patient (pat.karno)	0.0298	0.1403	0.7080
Calories consumed at meals (meal.cal)	0.1793	4.1493	0.0417
Weight loss in last six months (wt.loss)	0.0764	1.0117	0.3145

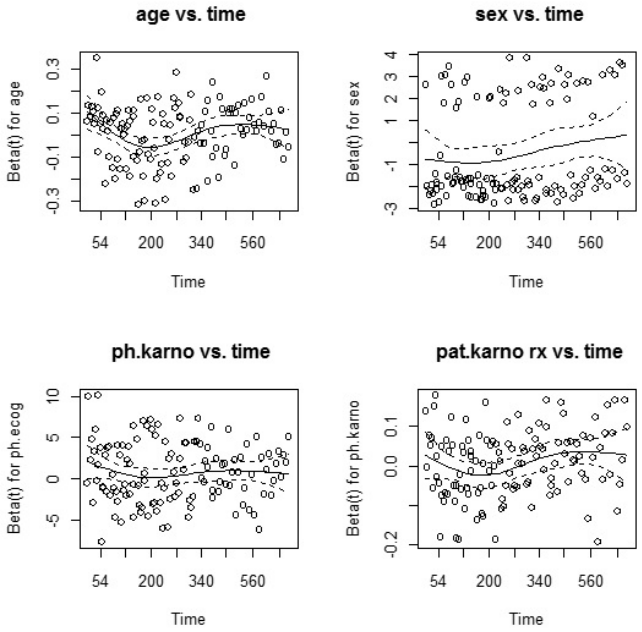
From Table 4.14 above, meal.cal and sex are the only significant covariates. The other covariates are insignificant at $\alpha = 0.05$ significance level.

Figure 4.18: Schoenfeld residuals versus Time for the overall fit: NCCTG Lung Cancer Data



The Schoenfeld residual plots for four covariates in the model including a lowess smoothing curve yields

Figure 4.19: Schoenfeld residuals versus Time for the covariates :NCCTG Lung Cancer Data



The overall Schoenfeld residual plot together with the plots for the four covariates show a zero slope, indicating proportionality. The hazard proportionality tests for the two-sample results are provided in Table 4.15 below. We compare the power of rejection between Kolmogorov-Smirnov test for proportional hazard, the smooth tests (Legendre $d = 3$ with 3 degrees of freedom), data-driven smooth test (Legendre functions as basis, nested with 5 dimensions) and the global test.

Table 4.15: Tests of Proportionality in CPH: Lung cancer data

Test	Statisic	p-value
Global test	13.8	0.06
Two-sample Kolmogorov-Smirnov test	1.97	0.99
Smooth test of order 3	0.81	0.85
Data-driven Smooth test	0.01	0.96

Results show that the global test rejects the null hypothesis at $\alpha < 0.1$ but does well for $\alpha < 0.05$. The other three tests fail to reject the null hypothesis. This is coherent with the Schoenfeld residual plots.

Dataset 6: Stage C Prostate Cancer

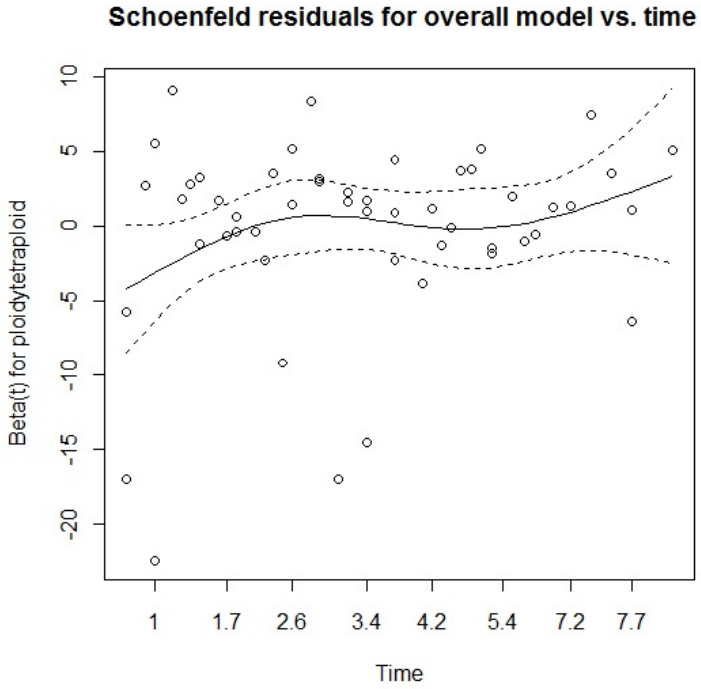
Data contained 146 patients with stage C prostate cancer, from a study exploring the prognostic value of flow cytometry. Patients were followed and variables for time to progression or last follow-up (years) recorded. other measurements were status (1= progression observed, 0 = censored), age in years, status for endocrine captured (i.e. early endocrine therapy, 1 = no, 2 = yes), percent of cells in G2 phase, as found by flow cytometry, grade of the tumor, grade of the tumor, the ploidy status of the tumor, from flow cytometry, values for diploid, tetraploid, and aneuploid. A tumor was determined to be diploid (normal complement of dividing cells) if the fraction of cells in G2 phase was determined to be 13% or less. Aneuploid cells were given a measurable fraction with a chromosome count that is neither 24 nor 48, for these the G2 percent is difficult or impossible to measure (Hankey et al., 1999). Variables and covariates include time to progression or last follow-up in years (pgtime); 1 = progression observed, 0 = censored (pgstat); age in years (age); early endocrine therapy, 1 = no, 2 = yes (eet); percent of cells in G2 phase, as found by flow cytometry (g2); grade of the tumor, Farrow system (grade); grade of the tumor, Gleason system (gleason); the ploidy status of the tumor, from flow cytometry. Values are diploid, tetraploid, and aneuploid (ploidy). We fit Cox PH model for time and status with covariates age, sex, rx, obstruct, adhere, differ, extent, surg, node4 and etype.

Table 4.16: Fitting Cox PH model to Stage C Prostate Cancer data

Covariates	β	chisq	p-value
age in years	0.0529	0.2361	0.6270
early endocrine therapy, 1 = no, 2 = yes (eet)	-0.0406	0.1106	0.7395
percent of cells in G2 phase, as found by flow cytometry (g2)	0.0134	0.0109	0.9170
grade of the tumor, Farrow system (grade)	-0.0716	0.2733	0.6012
grade of the tumor, Gleason system (gleason)	-0.1328	0.9858	0.3208
ploidydiploid	0.1311	1.2196	0.2694
ploidytetraploid	0.1945	2.9799	0.0843

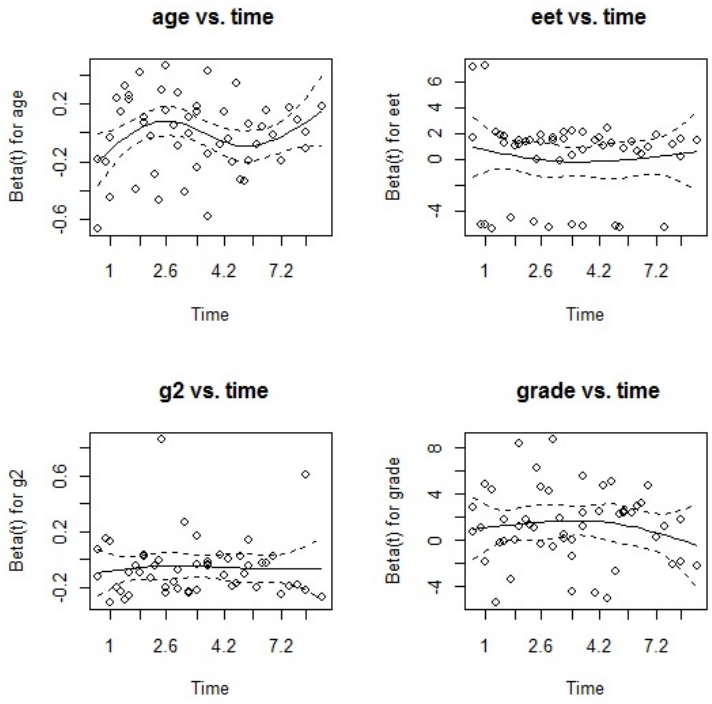
From Table 4.16 above all covariates are also insignificant at $\alpha = 0.05$. The Schoenfeld residuals versus log(time) for the overall fit is shown in Figure 4.20 below.

Figure 4.20: Schoenfeld residuals versus Time for the overall fit: Stage C Prostate Cancer Data



We also created the Schoenfeld residual plots for four covariates, including a lowess smoothing curve.

Figure 4.21: Schoenfeld residuals versus Time for each covariate: Stage C Prostate Cancer Data



The overall Schoenfeld residual plot together with the four covariates plots shows a general zero slope indicating proportionality. The proportionality tests for the two-sample results are provided in Table 4.17. We compare the power of rejection between Kolmogorov-Smirnov test for proportional hazard, the smooth tests (Legendre $d = 3$ with 3 degrees of freedom), data-driven smooth test (Legendre functions as basis, nested with 5 dimensions) and the global tests results to

Table 4.17: Tests of Proportionality in CPH: Stage C Prostate Cancer data

Test	Statisic	p-value
Global test	7.30	0.40
Two-sample Kolmogorov-Smirnov test	2.35	0.50
Smooth test of order 3	1.31	0.73
Data-driven Smooth test	1.12	0.31

Results show that all the tests are consistent and fail to rejects the null hypothesis at $\alpha < 0.05$.

Dataset 7: Chemotherapy for Stage B/C colon cancer data

This was a national intergroup trial that was sponsored by the National Cancer Institute and involved the Eastern Cooperative Oncology Group, the NCCTG, the Southwest Oncology Group, and the Mayo Clinic. Enrollment of patients started in March 1984, when a preliminary analysis of the NCCTG study indicated the likelihood of a treatment advantage for levamisole plus fluorouracil and for levamisole alone, with regard to time to recurrence. Enrollment was completed in October 1987. All patients were required to have undergone a potentially curative adenocarcinoma of the colon without gross or microscopic evidence of residual disease. Patients with rectal carcinoma were excluded for the study. The resected specimen in eligible patients showed one of two indicators of poor prognosis - invasion extending at least to the serosa or pericolonic fat (Stage B2) or metastasis to regional lymph nodes (Stage C). It was further required that the patient be able to swallow oral medication and have a leukocyte count of at least 4000 per microliter and a platelet count of at least 130,000 per microliter. Eligibility was determined by careful review of study forms, operative reports, and pathology reports. Entry into the study was allowed no earlier than one week and no later than five weeks after surgery. These are data from one of the first

successful trials of adjuvant chemotherapy for colon cancer. Levamisole is a low-toxicity compound previously used to treat worm infestations in animals. There are two records per person, one for recurrence and one for death (Moertel et al., 1990).

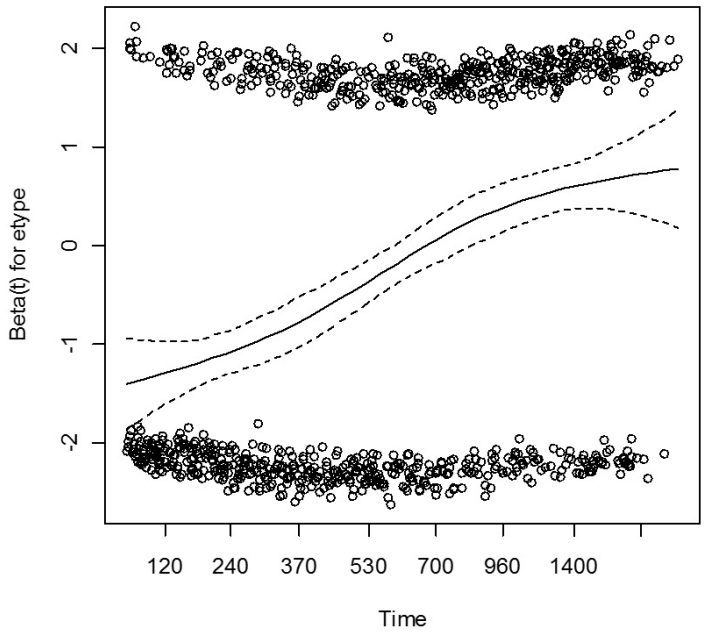
Variables include patients id; treatment -Obs(ervation), Lev(amisole), Lev(amisole)+5-FU (rx); sex; age in years; obstruction of colon by tumour(obstruct); perforation of colon (perfor); adherence to nearby organs (adhere); number of lymph nodes with detectable cancer (nodes); days until event or censoring (time); censoring status (status); differentiation of tumour (1=well, 2=moderate, 3=poor)(differ); Extent of local spread (1=submucosa, 2=muscle, 3=serosa, 4=contiguous structures)(extent); time from surgery to registration (0=short, 1=long)(surg); more than 4 positive lymph nodes (node4); event type: 1=recurrence,2=death (etype). We fit Cox PH model for time and status with covariates age, sex, rx, obstruct, adhere, differ, extent, surg, node4 and etype.

Table 4.18: Fitting Cox PH model to Chemotherapy for Stage B/C colon cancer data

Covariates	β	chisq	p-value
age in years	-0.0107	0.111	7.39e-01
sex	0.0563	2.869	9.03e-02
Treatment Lev(amisole)(rxLev)	-0.0475	2.068	1.50e-01
Treatment Lev(amisole)+5-FU (rxLev+5FU)	-0.0198	0.355	5.51e-01
obstruction of colon by tumour(obstruct)	-0.1093	11.113	8.57e-04
adherence to nearby organs (adhere)	0.0470	2.032	1.54e-01
differentiation of tumour (1=well, 2=moderate, 3=poor)(differ)	-0.1480	22.515	2.08e-06
Extent of local spread (1=submucosa, 2=muscle, 3=serosa, 4=contiguous structures)(extent)	-0.0364	1.229	2.68e-01
time from surgery to registration (0=short, 1=long)(surg)	0.0135	0.166	6.84e-01
more than 4 positive lymph nodes (node4	-0.1121	10.788	1.02e-03
event type: 1=recurrence,2=death (etype)	0.3561	112.926	0.00e+00

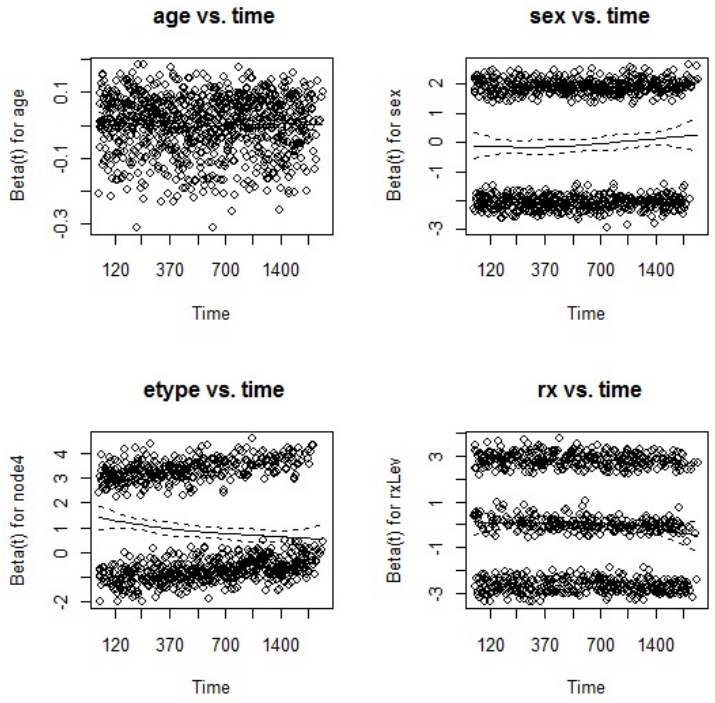
Results from Table 4.18 show that all the selected covariates are also significant at $\alpha = 0.05$. For testing proportionality the Schoenfeld residuals versus log(time) plot for the overall fit is shown in Figure 4.22.

Figure 4.22: Schoenfeld residuals versus Time for the overall fit: Stage B/C Colon Cancer Data



The Schoenfeld residual plots for four selected covariates in the model including a lowess smoothing curve yields The overall Schoenfeld residual plot together with the plots for the four covariates show a general non-zero slope indicating non-proportionality. The proportionality tests for the two-sample results in are indicated in Table 4.19. We compare the power of rejection between Kolmogorov-Smirnov test for proportional hazard, the smooth tests (Legendre $d = 3$ with 3 degrees of freedom), data-driven smooth test (Legendre functions as basis, nested with 5 dimensions) and the global tests results to Results show that all the tests are consistent in rejecting the null hypothesis. Despite the fact that all covariates incorporated in this model are significant, proportionality does not hold. The covariates are therefore time dependent $\alpha < 0.05$.

Figure 4.23: Schoenfeld residuals versus Time for four selected covariates: Stage B/C Colon Cancer Data



Dataset 8: Veteran Administration Lung Cancer study

The study population consisted of 109 patients with newly diagnosed Small Cell Lung Cancer (SCLC) investigated at the Pulmonary Division of Mainz University Hospital between 1989 and 1999. Clinical data were collected from chart review. The staging procedure for the majority of patients was standardized including a fiberoptic bronchoscopy, routine laboratory parameters, chest CT, abdomen CT and bone scan. In 89% of the patients chemotherapy was performed as first-line treatment. Three different standard combinations were applied with a median of four cycles. Response was first evaluated after two cycles of chemotherapy and every second cycles thereafter or if new clinical symptoms occurred. Response to chemotherapy was classified according to the WHO criteria in complete response, partial response, stable disease or progressive disease. Complete response was

Table 4.19: Tests of Hazard Proportionality in CPH: Colon cancer data

Test	Statisic	p-value
Global test	161.05	0.00
Two-sample Kolmogorov-Smirnov test	73.17	$<$ 2.22e-16
Smooth test of order 3	97.76	$<$ 2.22e-16
Data-driven Smooth test	97.21	$<$ 2.22e-16

achieved in 24% of patients, partial response in 29%, and stable disease in 5% of patients. 42% of patients progressed during therapy. In 35% of all patients chemotherapy was followed by radiotherapy of the primary tumor. From all subjects four patients with complete response underwent surgical resection of the primary tumor side. The majority of patients were followed-up regularly in a time frame of 2 to 3 months. The survival time was calculated from the date of histological diagnosis Micke et al. (2002).

Variables include treatment 1=standard 2=test (trt); celltype 1=squamous, 2=smallcell, 3=adeno, 4=large (celltype); survival time (time); censoring status (status); Karnofsky performance score (100=good) (karno); months from diagnosis to randomisation (diagtime);age in years; prior therapy 0=no, 1=yes (prior).

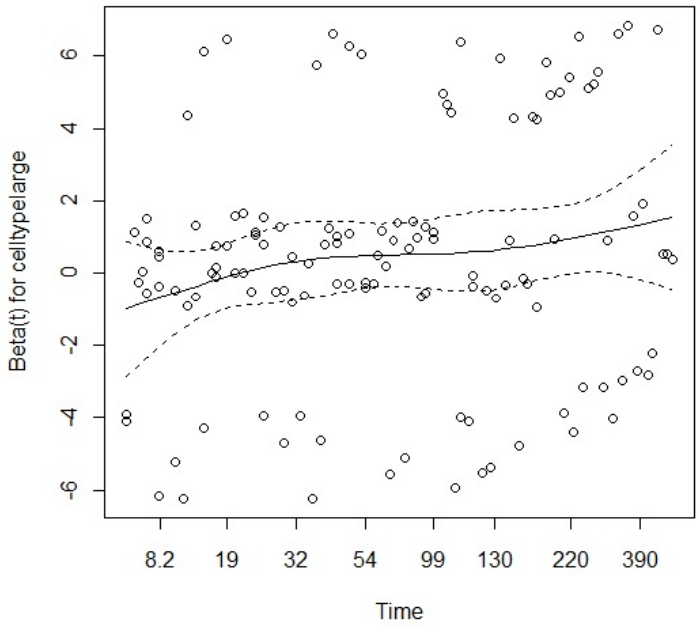
We fit Cox PH Model for time and status with covariates age, karno, trt, diagtime, prior and celltype

Table 4.20: Fitting Cox PH model to Veteran Administration Lung Cancer study

Covariates	β	chisq	p-value
age in years	0.1890	5.3476	0.020750
Karnofsky performance score (100=good) (karno)	0.3073	13.0449	0.000304
treatment 1=standard 2=test (trt)	-0.0273	0.1227	0.726104
months from diagnosis to randomisation (diagtime)	0.1491	2.9436	0.086217
age in years; prior therapy 0=no, 1=yes (prior)	-0.1767	4.4714	0.034467
celltype smallcell	0.0128	0.0261	0.871621
celltype adeno	0.1424	2.9794	0.084329
celltype large	0.1712	4.1093	0.042649

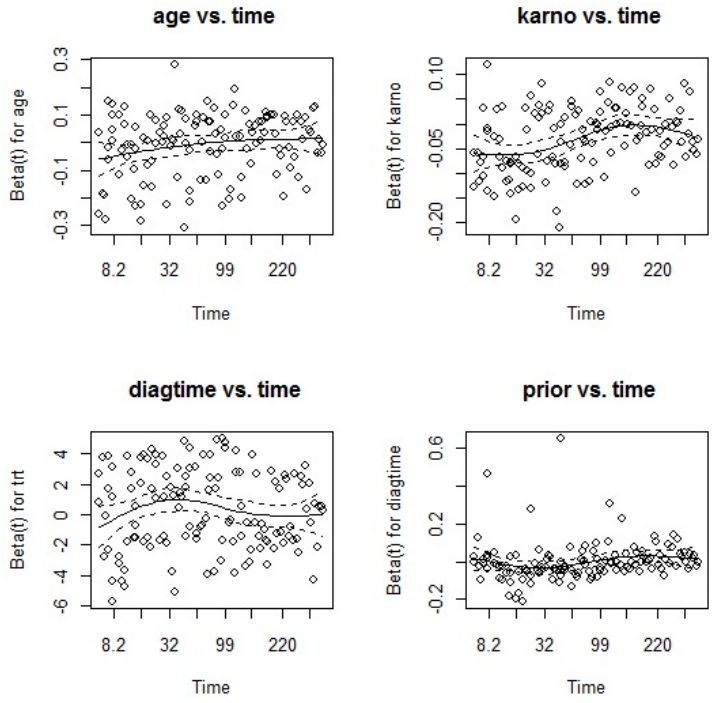
From Table 4.20 above some covariates (i.e. celltypelarge, prior, karno and age) were significance at $\alpha < 0.05$. The other covariates were not significant at $\alpha = 0.05$. The Schoenfeld residuals versus log(time) plot for the overall fit is shown in Figure 4.24 below.

Figure 4.24: Schoenfeld residuals versus Time for the overall fit: Veteran Administration Lung Cancer study Data



The Schoenfeld residual plots for four selected covariates including a lowess smoothing curve is shown in Figure 4.25 below.

Figure 4.25: Schoenfeld residuals versus Time for four selected covariates: Veteran Administration Lung Cancer study Data



The overall Schoenfeld residual plot together with the plots for the four covariates, except prior, shows a non-zero slope, indicating proportionality. The proportionality tests for the two-sample results are indicated in Table 4.21 below. We compared the power of rejection between Kolmogorov-Smirnov test for proportional hazard, the smooth tests (Legendre $d = 3$ with 3 degrees of freedom), data-driven smooth test (Legendre functions as basis, nested with 5 dimensions) and the global tests.

Table 4.21: Tests of Proportionality in CPH model: Lung Cancer Data

Test	Statisic	p-value
Global test	28.00	0.00
Two-sample Kolmogorov-Smirnov test	9.09	0.06
Smooth test of order 3	6.84	0.08
Data-driven Smooth test	2.03	0.18

In this case the global test strongly rejects proportionality, whereas the two-sample Kolmogorov-Smirnov test and smooth test of order 3 rejects the null (proportionality) at $\alpha < 0.10$. The data-driven version of the smooth test however remains stable and fails to reject the null, an indication of proportionality.

The results are summarized in Table 4.22

Table 4.22: Summary of 8 Real Datasets when Testing of Rejection* of Hazard Proportionality at ($\alpha < 0.05$) in CPH. Table also include articles that used the dataset without verifying proportionality in CPHM and Description of Schoenfeld plots

Dataset	Articles	Schoenfeld residual plots	Global test	2-sample K-S test	Smooth test (order 3)	Data-driven smooth test
Dataset 1	Andersen et al (2012)	zero line slope	rejects H_0	fails to reject H_0	fails to reject H_0	fails to reject H_0
Dataset 2	Royston et al (2011)	non-zero slope	rejects H_0	rejects H_0	rejects H_0	rejects H_0
Dataset 3	Edmonson et al (1979)	zero slope	fails to reject H_0	fails to reject H_0	fails to reject H_0	fails to reject H_0
Dataset 4	Embury et al (1977)	zero slope	NA	fails to reject H_0	fails to reject H_0	fails to reject H_0
Dataset 5	Loprinzi et al (1994)	zero slope	marginal	fails to reject H_0	fails to reject H_0	fails to reject H_0
Dataset 6	Zagars et al (1993)	non-zero slope	fails to reject H_0	fails to reject H_0	fails to reject H_0	fails to reject H_0
Dataset 7	Moertel et al (1990)	non-zero slope	rejects H_0	rejects H_0	rejects H_0	rejects H_0
Dataset 8	Micke et al (2002)	zero slope	rejects H_0	marginal	fails to reject H_0	fails to reject H_0

*H_0: Hazard proportionality hold

4.3 Application of Smooth Tests to Recurrent Event

4.3.1 Simulations

We conducted Monte Carlo simulations to investigate the performance of the smooth tests. Similar tests have been examined in extensive simulation studies by Agustin and Peña (2001, 2005); Pena (1998a). The goal of simulations was to compare the empirical significance levels of the tests with the specified nominal asymptotic levels as the sample size and the degree of censoring are varied. These comparisons indicate which tests qualify as good omnibus tests and which tests have good control of Type I error among a wide range of alternatives. Simulations were also helpful in determining appropriate values of the smoothing parameter k (i.e. $k = \{1,2,3,4\}$). Since we are fitting the BBS model with the probability of perfect repair $\rho(\cdot) = 1$, we considered simulations for $n = \{20,50,100,200,500,1000\}$.

In the simulations, the initial failure-time variables were generated according to the exponential distribution with mean $\theta = 8$, the Weibull distribution with shape parameter γ and scale parameter β and the Gamma ($\varsigma = 3, \alpha = 4$) distribution. The failure-time variables ($T_1, T_2, \cdots T_n$) were generated using the chosen alternative, and the censoring variables ($C_1, C_2, \cdots, C_m$) were distributed according to the exponential distribution with mean 1. By utilizing the resulting randomly censored data $\{(Z,i), i = 1, \cdots, n\}$, the null hypothesis was tested according to the different 5% asymptotic level tests. The percentage out of the replicates that a test rejected H_0 was then calculated. The bootstrapping procedure was applied 1,000 times to each generated dataset to obtain the significance level of the test. Within the context of model selection, size estimates were based on the proportion of replications that indicate acceptable fit, with a larger number of replications resulting in smaller CIs (higher power, more accuracy) around the estimates. The data-generating process was performed using the ***SimSurv*** function of the ***prodlim*** package from ***R***.

Table 4.23: Empirical control of the Type I error rate under $H_0 : \lambda(\cdot)$ is distributed as exponential ($\theta = 8$) at $\alpha = 0.05$. The failure times under the null hypothesis were generated according to a BBS model.

	Empirical Type I error rate (70% Censoring)					
Sample size, n	n=20	n=50	n=100	n=200	n=500	n=1,000
ψ_1	0.0452	0.0406	0.0422	0.0394	0.0323	0.0401
ψ_2	0.0418	0.0455	0.0497	0.0453	0.0498	0.0428
ψ_3	0.0560	0.0485	0.0443	0.0490	0.0545	0.0563
ψ_4	0.0568	0.0491	0.0553	0.0598	0.0654	0.0570
Censoring %	Empirical Type I error rate (50% Censoring)					
Sample size, n	n=20	n=50	n=100	n=200	n=500	n=1,000
ψ_1	0.0366	0.041	0.0333	0.0394	0.0303	0.0371
ψ_2	0.0422	0.0459	0.0373	0.0453	0.0363	0.034
ψ_3	0.0457	0.049	0.0398	0.05	0.0401	0.0482
ψ_4	0.0564	0.0496	0.0403	0.0528	0.0509	0.0491
Censoring %	Empirical Type I error rate (20% Censoring)					
Sample size, n	n=20	n=50	n=100	n=200	n=500	n=1,000
ψ_1	0.0311	0.0322	0.0334	0.0304	0.0423	0.0309
ψ_2	0.0384	0.0385	0.0398	0.0376	0.0486	0.0394
ψ_3	0.043	0.0424	0.0438	0.0422	0.0525	0.0447
ψ_4	0.044	0.0432	0.0446	0.0431	0.0533	0.0558

Table 4.24: Empirical control of the Type I error rate under $H_0 : \lambda(\cdot)$ is distributed as Weibull ($\beta = 6$, $\gamma = 10$) at $\alpha = 0.05$. The failure times under the null hypothesis were generated according to a BBS model.

	Empirical Type I error rate (70% Censoring)					
Sample size, n	n=20	n=50	n=100	n=200	n=500	n=1,000
ψ_1	0.0352	0.0406	0.0422	0.0494	0.0423	0.0371
ψ_2	0.0431	0.0411	0.0401	0.0411	0.0411	0.0431
ψ_3	0.0444	0.0423	0.0677	0.069	0.0663	0.0471
ψ_4	0.0798	0.0919	0.081	0.0788	0.0703	0.0998
	Empirical Type I error rate (50% Censoring)					
Sample size, n	n=20	n=50	n=100	n=200	n=500	n=1,000
ψ_1	0.0466	0.041	0.0333	0.0494	0.0403	0.0301
ψ_2	0.0471	0.0466	0.0411	0.0494	0.0455	0.0411
ψ_3	0.0516	0.0511	0.0594	0.0503	0.0571	0.0479
ψ_4	0.0533	0.0666	0.0811	0.0777	0.0603	0.0578
	Empirical Type I error rate (20% Censoring)					
Sample size, n	n=20	n=50	n=100	n=200	n=500	n=1,000
ψ_1	0.0311	0.0322	0.0334	0.0304	0.0323	0.0309
ψ_2	0.0471	0.0466	0.0371	0.0394	0.0403	0.0371
ψ_3	0.0499	0.0511	0.0454	0.0422	0.0471	0.0455
ψ_4	0.0521	0.0596	0.0511	0.0484	0.0509	0.0588

Table 4.25: Empirical control of the Type I error rate under $H_0 : \lambda(\cdot)$ is distributed as Gamma($\varsigma = 3, \alpha = 4$) at $\alpha = 0.05$. The failure times under the null hypothesis were generated according to a BBS model.

	Empirical Type I error rate (70% Censoring)					
Sample size, n	n=20	n=50	n=100	n=200	n=500	n=1,000
ψ_1	0.0444	0.0354	0.0423	0.0394	0.0323	0.0371
ψ_2	0.0497	0.0396	0.0474	0.0453	0.0398	0.0428
ψ_3	0.0531	0.0423	0.0506	0.049	0.0445	0.0563
ψ_4	0.0537	0.0529	0.0512	0.0598	0.0854	0.077
	Empirical Type I error rate (50% Censoring)					
Sample size, n	n=20	n=50	n=100	n=200	n=500	n=1,000
ψ_1	0.0366	0.031	0.0333	0.0394	0.0403	0.0371
ψ_2	0.0422	0.0459	0.0373	0.0453	0.0463	0.044
ψ_3	0.0557	0.049	0.0498	0.059	0.0601	0.0482
ψ_4	0.0964	0.0696	0.0503	0.0598	0.0909	0.0991
	Empirical Type I error rate (20% Censoring)					
Sample size, n	n=20	n=50	n=100	n=200	n=500	n=1,000
ψ_1	0.0331	0.0322	0.0355	0.0304	0.0323	0.0309
ψ_2	0.044	0.0385	0.0422	0.0476	0.041	0.0394
ψ_3	0.0599	0.0624	0.0663	0.0722	0.0564	0.0447
ψ_4	0.0691	0.1122	0.1072	0.0931	0.0875	0.1058

Table 4.26: Empirical control of the Type I error rate under $H_0 : \lambda(\cdot)$ is distributed as Weibull ($\beta = 5, \gamma = 15$) at $\alpha = 0.05$. The failure times under the null hypothesis were generated according to a BBS model.

	Empirical Type I error rate (70% Censoring)					
Sample size, n	n=20	n=50	n=100	n=200	n=500	n=1,000
ψ_1	0.0415	0.0331	0.0395	0.0361	0.0382	0.044
ψ_2	0.0464	0.037	0.0442	0.0417	0.0452	0.0493
ψ_3	0.0496	0.0395	0.0472	0.0551	0.0695	0.0526
ψ_4	0.0502	0.04	0.0478	0.0559	0.0704	0.0533
	Empirical Type I error rate (50% Censoring)					
Sample size, n	n=20	n=50	n=100	n=200	n=500	n=1,000
ψ_1	0.0435	0.0383	0.0311	0.0361	0.037	0.0333
ψ_2	0.0487	0.0429	0.0348	0.0417	0.0426	0.0397
ψ_3	0.052	0.0458	0.0472	0.0551	0.0561	0.0437
ψ_4	0.0527	0.0464	0.0677	0.0759	0.1069	0.0946
	Empirical Type I error rate (20% Censoring)					
Sample size, n	n=20	n=50	n=100	n=200	n=500	n=1,000
ψ_1	0.0333	0.0388	0.0318	0.0364	0.0375	0.0362
ψ_2	0.0397	0.0446	0.0381	0.0432	0.0456	0.0442
ψ_3	0.0437	0.0583	0.0619	0.0574	0.0607	0.0791
ψ_4	0.0846	0.089	0.1028	0.0683	0.0817	0.0802

To investigate the performance of the different tests, we evaluated the exponential, Weibull and Gamma null hypothesis against their generalised alternatives for the initial distribution of failure ages. Hence, given the values of different parameters, values for each alternative were generated. Simulations were done for the 5% asymptotic level tests. We also performed simulations for values of $\beta = 6$, $\gamma = 10$ and $\beta = 5, \gamma = 15$ for the

Weibull initial distribution and found the results to be consistent with those presented by Agustin and Peña (2001, 2005); Pena (1998a); Odhiambo et al. (2017b). Examining the performance of the directional tests, we again noticed that we are able to achieve required significance levels. Table 1, 2, 3 and 4 summarize the percentage rejection of the tests for the exponential alternatives, Gamma alternatives and Weibull alternatives. For the exponential alternative, the ψ_1, ψ_2 and ψ_3, have the highest power under H_0. The directional tests based on $\psi_i, i = 1,2,3,4$, are sensitive for the Gamma-type distributions. The fact that the ψ_1 test are powerful for this alternative is expected because this test was derived against such alternatives, whereas the observation that the ψ_4 test is not powerful for this alternative is also expected because the normalized total-time-on-test statistic is invariant to changes in scale Agustin and Peña (2001). Against the Weibull-type, Gamma-type and exponential alternatives, the ψ_1 test performed best, followed by the ψ_2 tests.

4.3.2 Fitting HIV retention data to BBS model

Motivation for Analysis of HIV Retention Data

HIV/AIDS have consistently been a major challenge in Kenya. The national prevalence is currently estimated to be 6% and there are at least 1.6 million Kenyans living with HIV (PLHIV) with at least 800,000 of PLHIV on ART (Council, 2014). In practice, the quality of the ART services is measured against the rate of retention of PLHIV on ART. With the advent of the United Nation AIDS (UNAIDS) programme on HIV/AIDS targets in 2013, the focus have turned to interventions that quicken elimination of HIV/AIDS at the global, regional, country, province, district and city levels (UNAIDS, 2014). The strategy popularly known as $90-90-90$ targets that by 2020, 90% of people living with HIV know their HIV status, 90% of people who know their status are receiving ART treatment and 90% of people on HIV treatment have a suppressed viral load so that their immune system

remains stronger and the likelihood of their infection being passed to others is greatly reduced. This strategy is currently being implemented in Kenya and this section focuses on statistical innovation that hinges on one of the pillars: the third 90% i.e. viral suppression. Viral suppression is achievable by retaining patients on ART for long. There are potential benefits whenever a PLHIV's viral load is reduced to an undetectable level (i.e. people with undetectable viral loads are generally healthier than those people with higher levels of virus in their blood and are also less likely to transmit HIV to their sexual partners (Rachlis et al., 2014)). High retention rate in ART treatment plays a crucial role in maintaining viral load suppression (Megerso et al., 2016; Rachlis et al., 2014; Rasschaert et al., 2012; Gwynn et al., 2015; Ramadhani et al., 2007; Haddow et al., 2003). See Odhiambo et al. (2017a) for excellent summary on the relevance of LTFU and the rationale for fitting LTFU data to a parametric distribution.

Loss to follow-up (LTFU) in a typical ART treatment facility translates to increased morbidity and mortality through sub-optimal viral suppression, increased risk of drug resistance, and increased risk of HIV transmission. Consequently, there is increasing concern about inadequate retention and adherence due to LTFU (Rachlis et al., 2014). This paper revisits HIV retention and fits a primary data from typical HIV setting to a time to first occurrence recurrent events model. The main applications, is to assess the hazard rate of the first LTFU time using smooth test of GOF. The work is motivated by the analysis of HIV retention data from two typical government Hospitals in Kenya. Models such as CPH have been employed to assess the effects of baseline covariates, such as onset age, gender, and WHO staging, on the risk of LTFU. The validity of statistical inference, however, depends critically on the adequacy of the assumptions. Smooth test of GOF have been developed and studied in the literature when assessing models in gap time setting by Agustin and Peña (2005, 2001); Pena (1998a). The focus of this paper is pegged on application of smooth tests to hazard function in recurrent event situation.

Patients receiving ART can experience LTFU which may result in discontinuation of

treatment, drug toxicity, treatment failure due to poor adherence and drug resistance. This can result in an increased risk of death of up to 40% of ART patients in sub-Saharan Africa (Berheto et al., 2014). Studies have shown that LTFU has negative impact on immunological benefit of ART and increases AIDS-related morbidity, mortality, and hospitalizations (Berheto et al., 2014). Individuals who miss visits in the first year of treatment have a higher mortality rate (Rachlis et al., 2014). Asiimwe et al. (2016) showed that retention of patients who are on ART treatment remains stable after 12 months of ART initiation, with loss to follow-up(LTFU) being the main cause of attrition (Rasschaert et al., 2012). Previous studies have also illustrated associations between frequent LTFU and more severe opportunistic illnesses (Haddow et al., 2003). Analysis of LTFU have also been used in HIV care to monitor and improve programme effectiveness, using patient retention as a measure of quality of care (Sengayi et al., 2013). The main objective in analysis of LTFU data is to check retention of patients in care. Retaining patients for long allows provision of long term Highly Active Antiretroviral Therapy (HAART), tracking WHO staging, tracking immunosuppression profiles and evaluation of emergence of medication toxicities. In resource-limited settings, it is common to find patients dropping out of ART treatment. Due to significant drop-outs, patients may not realize the benefits of ART if they are LTFU. More innovation is therefore required for further ART scale-up and improve retention in care.

4.3.3 Data Description

Data comprised all patients who were initiated ART at two government hospitals in Kenya. Patients under observation were enrolled between 1st of October 2011 and 31st December 2014. The event of interest was time to first LTFU. Data was collected routinely whenever patients came for clinical check-up or drug refill. The time between ART initiation to first LTFU was given in months. Time to first LTFU was defined as missing routine clinical

appointment within 48 hours from the scheduled appointment date. Out of those initiated on ART 854 patients experienced LFTU while the rest were right-censored. See Odhiambo et al. (2017a) for more information on data description. We have also described the data in section 4.1.2.

4.3.4 Application to HIV retention

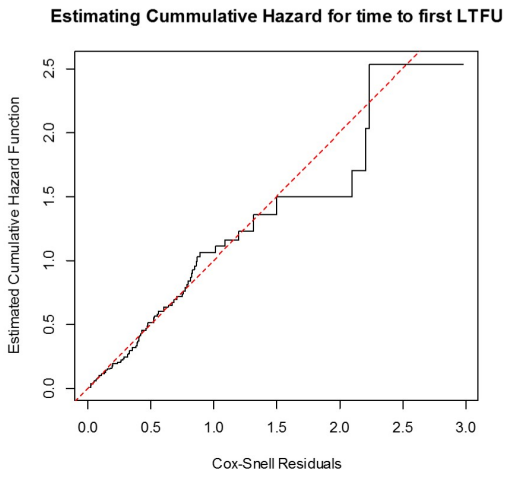
Modelling LTFU

Let $\omega_1, \omega_2, \cdots, \omega_n$ be the gap-time between started on ART to LTFU for PLHIV. The date of start of ART for each patient is independent and the start time of observation is set at 0. The gap time is the period between start of ART until time t, when the patients experiences LTFU (failure) during observation period. Patients who experience other exits (i.e. transfer-out and death) are censored. Patients who remain active on ART at the end of the observation period are also censored. A patient who experiences LTFU (failure) can still be recovered back (repaired) to ART treatment through a mechanism called defaulter tracing. Such patients are still exposed to the risk of LTFU even after recovery and so the event can recur several times. The setting is naturally public health, hence the risk for the 2nd, 3rd and proceeding episodes of LTFU is the same as the fist. Let $S_1, S_2, \cdots$ be the successive event times (LTFU) for the A process, and let $T_1, T_2, \cdots$ be the successive event times of the B process. To obtain a realization of a $\mathrm{BBS}(\Lambda, p)$ model, events in the A process correspond to the imperfectly repaired failures, whereas events in the B process are associated with the perfectly repaired failures. The equivalence is then seen by noting that the process $A+B$ is a nonhomogeneous Poisson process with intensity function $\int_0^{\cdot} \lambda(s)ds$, and, given that at time t an event has occurred, the probability that it is from the B process is $p(t)$.

Model fitting

Here we begin by checking difference in residual behaviour and result to be detected in the hypothesis testing.

Figure 4.26: Comparing baseline hazard for time to first occurrence of LTFU



The residuals graphically showed moderate signs of a different behaviour for violated models when the sample size is small and censoring is present. This indicates that for small samples with a higher degree of censoring the residuals could be sensitive for model violations. The advantage of using smooth test in the next subsection is that, the test is not

Table 4.27: Assessing the initial distribution. Only the Weibull distribution fails to reject the H_0 :Initial distribution is Weibull, at $\alpha < 0.05$.

Distribution	Scale	Chi-square value	p-value
Weibull	2.11	4.13	0.213
Exponential	1	8.29	0.016
Logistic	3.73	21.13	2.6e-05
Rayleigh	0.5	9.37	0.0093
Extreme Value	3.97	16.25	3e-04
Gaussian	7.45	29.83	3.3e-07
Student-t	4.89	21.58	2.1e-05
Log-normal	0.287	30.39	6.3e-05

affected by sample size, number of covariates and the level of censoring. We now check the fit for the distribution of the initial distribution The results of smooth test of goodness-of-fit when fitting LTFU data to BBS model with initial distribution Weibull($\beta = 6;\gamma = 30$) yields

Table 4.28: Result of Smooth test up to order 4 for BBS model with initial distribution Weibull ($\beta = 6, \gamma = 30$) against an initial distribution of generalized Weibull family.

Ψ	Test statistic	p-value
ψ_1	2.211	1.0
ψ_2	3.334	0.811
ψ_3	4.121	0.565
ψ_4	5.011	0.323

Chapter 5

Comparative Discussion of GOF Methods

We have revisited the problem of testing whether a complete sample or censored sample with right-censored survival data come from a given distribution or hazard functions. The aim is to apply smooth tests of GOF to a wide areas in classical probabilities, survival analysis and recurrent events models. The tests are based on Neyman's embedding idea. The null hypothesis is tested against a model where the probability or hazard function of the survival distributions is expressed by several smooth functions. We have also incorporated data-driven versions to the selection of basis functions. We have also explored sample size performance and consistencies through simulations in order to compare the power of smooth tests with other conventional tests.

We have also explored Neyman's smooth test idea and its data-driven versions and demonstrated how it can be extended with applications in survival analysis and recurrent events. The smooth tests and their data-driven versions is a viable competitor to many existing procedures, especially when one seeks a test without a clear advance idea of the alternative. Certainly, methods of this type can be developed for many other branches of statistics, though now they do not seem to be used often.

In Chapter 1, we have looked at motivation of the study by elaborating on the significance of retaining patients on ART and the risk of experience LTFU which may result in discontinuation of treatment, drug toxicity, treatment failure due to poor adherence and drug resistance. We also described the main objective in analysis of LTFU data in order to check retention of patients in care. We discussed the fact that retaining patients for long allows provision of long term ART, tracking WHO staging, tracking immunosuppression profiles and evaluation of emergence of medication toxicities. Due to significant drop-outs, patients may not realize the benefits of ART. More innovation is therefore required for further ART scale-up and improve retention in care. Another motivation for this study is the need to validate smooth test under different practical setting in variety of cancer studies. This have been captivated by the 2-sample problem in cancer studies.

In Chapter 2 we have revisited the general mathematical setting and framework when analysing smooth test of GOF introduced by Neyman (1937). We have considered extension of the smooth tests to cover non-censored data, censored data and recurrent events data.

5.1 HIV Retention

Human Immunodeficiency Virus/Acquired Immuno Deficiency Syndrome (HIV/AIDS) have consistently been a major challenge in Kenya. The national prevalence is currently estimated to be 6% (Council, 2014)and there are at least 1.6 million Kenyans living with HIV (PLHIV) with 59% of PLHIV are on antiretroviral therapy (ART) (Council, 2014). In practice, the quality of the ART services is measured against the rate of retention of PLHIV on ART. Program data and national survey shows that, the percentage of PLHIV initiated on ART reduces with time progression and that retention rate in ART is higher in the first 12 months (about 92%) and reduced to about 70% at month 60 (Council, 2014). This data depicts a critical need to establish robust measures to reduce loss to follow-up (LTFU) among

PLHIV on ART. With the advent of the United Nation AIDS (UNAIDS) programme on HIV/AIDS targets in 2013 (UNAIDS, 2014), the focus have turned to interventions that quicken elimination of HIV/AIDS at the global, regional, country, province, district and city levels. The strategy popularly known as 909090 targets: that by 2020, 90% of people living with HIV know their HIV status, 90% of people who know their status are receiving ART treatment and 90% of people on HIV treatment have a suppressed viral load so that their immune system remains stronger and the likelihood of their infection being passed to others is greatly reduced. This strategy is currently being implemented in Kenya. This thesis focuses on statistical innovation that hinges to the third 90 i.e. viral suppression.

Viral suppression is achievable by retaining patients on ART for long. There are potential benefits whenever a PLHIV's viral load is reduced to an undetectable level. PLHIV with undetectable viral loads are generally healthier than those people with higher levels of virus in their blood and are also less likely to transmit HIV to their sexual partners (Stricker et al., 2014). High retention rate in ART treatment plays a crucial role in maintaining viral load suppression (Ochieng-Ooko et al., 2010; Makunde et al., 2012; Gwynn et al., 2015).

Loss to follow-up (LTFU) in a typical ART treatment facility translates to increased morbidity and mortality through sub-optimal viral suppression, increased risk of drug resistance, and increased risk of HIV transmission. Consequently, there is increasing concern about inadequate retention and adherence due to LTFU (Gwynn et al., 2015).

The main application in this work is to assesses the hazard rate of the first LTFU time using smooth test of GOF. The work is motivated by the analysis of HIV retention data from two typical government Hospitals in Kenya. From the literature, Cox proportional hazards models (CPHM) have been employed to assesses the effects of baseline covariates, such as onset age, gender, and WHO staging, on the risk of LTFU. The validity of statistical inference, however, depends critically on the adequacy of the assumed model. The smooth test of GOF as a model checking procedures have been developed studied in the literature for assessing the CPH model but has not been extended to HIV retention setting.

5.2 Smooth Tests for Probability Distribution

For Weibull Distribution application with non-censored data, we provided the orthonormal structure so that the smooth tests of order k=3 and k=4 are computed. We then evaluate the goodness-of-fit for the two-parameter Weibull Distribution fitted on HIV retention data. The empirical tests GOF considered here (i.e. KS, CVM and AD) are powerful whenever the sample size is small. This is consistent with Kallenberg and Ledwina (1995); Rayner et al. (2009); Rayner and Best (1990, 1986). Our contribution in this article is unique in the sense that we are considering LTFU data generated from a typical clinical setting.

Lemeshko et al. (2009) investigated the gamma distribution with its parameters chosen so that it is closest to the Weibull Distribution. The power test was used to assess both simple and composite hypotheses against the simple alternative. Although he found the KolmogorovSmirnov, Cramérvon Mises and AndersonDarling type nonparametric tests to be most powerful compared with the case when the estimates are found by minimizing the corresponding statistics. The comparison did not take into consideration the smooth tests. In this study, we have considered smooth test and assessed its performance with other conventional tests.

Sürücü (2008) also did power comparison, and a simulation study of GOF tests but smooth test was not included in his assessment. Few authors have incorporated smooth tests when testing power of goodness-of-fit test. In particular, (Kallenberg and Ledwina, 1995) showed through simulation results that the data-driven version of Neyman's smooth tests performs very well over a wide range of alternatives and is competitive with other data-driven procedures. They also showed that the data driven smooth tests are consistent against essentially all alternatives (Kallenberg and Ledwina, 1995). Rayner and Best (1986) also showed that Neyman smooth test for location-scale families are flexible and can be chosen to improve detection of particular alternatives. The tests was shown to perform well against its competitors. This assessment is also consistent with our simulations output

though, our major contribution is in the setting of typical HIV programming data.

The smooth test approach fails to reject the hypothesis when considering up to order four. The p-value is quite large compared with KS test, CVM test and AD test and therefore more appropriate compared to other alternatives whenever we have large sample size. The smooth GOF test also produced the best estimate of the distribution of the data (Weibull Distribution), which will ultimately result in a better estimate of the hazard function for time to LTFU for predicting hazard rates. The Weibull Distribution is the best choice from the density plots and graphs and is validated by smooth GOF test. It is important to carefully choose the best GOF tests in order to make the correct inference about underlying distribution. We have demonstrated that the smooth test is superior and can be used to analyze time to LTFU data in order to determine the underlying distribution. These results agree with those of Kang (1979), who showed by simulation that the test for normality based on smooth test has much greater power than the generalized χ^2 test and the Kolmogorov-Smirnov test. Kang (1979)also demonstrated that the test performs generally as well as the Shapiro-Wilk, skewness, and kurtosis tests for a wide range of alternatives.

Several studies have shown that LTFU poses challenges to the successful implementation of ART programs (Berheto et al., 2014). Studies have shown that patients who discontinued ART developed a rapid increase in viral load and depletion of CD4 cells, putting them at risk of opportunistic infections and early death. Therefore, understanding the underlying pattern and distribution of LTFU is necessary to making sound interventions that maintain adherence to ART treatment. In this study, the two parameter Weibull Distribution fits the time to first LTFU well. Several authors (Megerso et al., 2016; Wang et al., 2011; Rasschaert et al., 2012; Ramadhani et al., 2007; Haddow et al., 2003; Asiimwe et al., 2016; Berheto et al., 2014; Rachlis et al., 2014)have shown that the main reason in rising cases of LTFU to be poor patient's defaulter tracing in resource-limited settings. This is likely to compromise positive outcomes of ART in a large scale HIV care center. Pattern of LFTU are therefore crucial in developing practical programmatic interventions.

5.3 Smooth Tests for 2-sample problem in Cancer Studies

We have considered CPH model; commonly used to determine risk factors. The assumption of proportional hazards is therefore important whenever the model is applied. Numerous methods for assessing the assumption of hazard proportions have been proposed. These methods (e.g. global test, G-test, Kolmogorov-Smirnov test, smooth test etc.) together with their asymptotic properties have been studied theoretically by several authors. However, validation of these tests in light of real settings have generally utilized either none or at most two real datasets. Furthermore, the combined use of graphical and non-graphical analysis which is one of the contribution of this manuscript have been studied comparatively by few authors. Also, in practice there exist variations in real data settings, particularly in cancer studies and validations of these tests in multiple settings have not been done. More so, most researchers, usually fit CPH models using several explanatory variables in order to identify risk factors. However, in the fitted CPH model, the covariates included in the model should satisfy the assumption that the relative risk is proportional over the time for different levels. This study's main objective is to validate the performance of smooth test in different cancer settings and compare with that of global goodness-of-fit test and Kolmogorov-Smirnov proportional hazard test. Particularly, we assesses the performance of these tests under different cancer study setting when testing for the PH assumption. With many variations in cancer studies, this paper however, does not aim to provide an exhaustive performance of smooth tests for proportionality for all types of cancer, but instead it aims to statistically validate its performance in selected eight different practical cancer settings. Ultimately, we hope that the issues and features we comment on will result in higher overall standards and quality of oncological research and limit the risk of using invalid models. For the eight data sets with the respective settings we have displayed the projected hazard plots together with their $\log(\beta(t))$ projection. We chose graphs that are based on the Schoenfeld residuals because they are more robust compared to Kaplan-Meier(K-M) survival graphs

because K-M graph does not work well for continuous predictor or categorical predictors that have many levels. Furthermore, the K-M curves are sparse when there are fewer time points and is usually not straight-forward when detecting how close to parallel is close enough. In these cases the resulting power of rejection is compared with the graphical presentation. Whereas, there are certain types on non-proportionality that cannot be detected by the tests of non-zero slopes alone, it become obvious when looking at the graphs of the residuals such as nonlinear relationship between the residuals and the function of time. In this regard, the behavior of smooth tests is similar as the other tests if we have ”sizeable” sample size. The two versions of smooth tests provide a procedure with power that is more stable than power of other methods. The smooth test is analyzed with a fixed dimension of order 3 with 3 degrees of freedom. For the data-driven version of the smooth test, we nested subsets in order to avoid the use of many components. The nested subsets selection procedure is not sensitive with respect to the choice of the maximum dimension (d) if d is large enough to cover realistic departures from the hypothesis (Kraus, 2009). Since, we have analyze these datasets in order to validate the smooth test in different settings, we have utilized Schoenfeld residuals plots to see indication if either proportionality does or does not hold for Time (Time defined as log (time)). We have consequently applied the four tests to determine consistency between the plots and the level of rejection of null hypothesis (proportionality). We have done this simultaneously for comparison and verification of results obtained with both the graphical and data analysis. The results of applying the tests to the eight cancer datasets are reported in each subsection under methods section. The eight dataset were pulled from already published articles and are readily accessible in *R* program. For the graphical presentation testing the time dependent covariates is equivalent to testing for a non-zero gradient. Therefore if the proportional hazards assumption is true, beta(t) will be a horizontal line. All analysis in this section were performed using ***coxph, cox.zph, survival*** and ***Surv2sample*** packages in ***R*** program. For each situation we describe the general setting of the study and the power of each test is computed. Assessment of

CPH model fitting in identifying proportional hazards on each categorical covariate taken individually are clearly observed in Schoenfeld residuals plots. The variables that showed approximate proportional hazards. A CPH model was then fitted to the data, using forward selection procedure that ended up including as many covariates as possible into the model. It is important, to note that, our interest is not how good covariates fit in CPH model, but how accurate the hazard proportionality assumption is determined. Then, the Schoenfeld residuals for overall and consequently four covariates were studied. A closer look at results for the various settings summarised below;

Data 1: The setting under this sub-section is Malignant Melanoma with 205 patients. The overall Schoenfeld residuals plot shows non-zero slope. Further, all the four covariates for Schoenfeld residuals plots except sex shows non-zero slope. Analytic results on the other hand show smooth tests both with fixed order 3 and data driven version fail to reject the null hypothesis with their p-values 0.12 and 0.15 respectively whereas the global test rejects the null hypothesis at $\alpha < 0.05$. On the other hand Kolmogorov-Smirnov test also fails to rejects the null at $\alpha < 0.05$ but doesn't do well at $\alpha < 0.1$ (null hypothesis is rejected). The smooth test in this setting does better than the other two tests in determining hazard proportionality. The smooth test is generally, coherent with the Schoenfeld residuals plots.

Data 2: For the cohort study on breast cancer patients analyzed in this sub-section, the overall Schoenfeld residuals plot and three of the selected four covariates depicts zero slope. This is an indication of time dependent covariates. In this setting, the sample size is also significantly large (n=2,982). Results shows all the four test are consistent in rejected the null hypothesis; a situation that is consistent with Schoenfeld residual plots.

Data 3: The setting here is ovarian cancer with 82 patients being observed. Results show the global test, smooth tests fixed dimension and data-driven smooth test fails to rejects the null hypothesis. This is in agreement with the Schoenfeld residual plots for none zero slope. However, the two-sample Kolmogorov-Smirnov test rejects the null hypothesis at $\alpha < 0.10$. Kolmogorov-Smirnov test will still be consistent with the other two test at $\alpha < 0.05$ but

may be misleading at $\alpha < 0.1$.

Data 4: The setting here involves patients with clonal hematopoietic stem cell disorder (acute myeloid leukaemia). With a sample size of 23 and one covariate,, the global test did not give any result but the other 3 tests (i.e. two-sample Kolmogorov-Smirnov test, smooth test and data-driven smooth test) failed rejects the null hypothesis. In other words, they detected proportionality. Despite the fact that the covariate was insignificant (chiq= 0.00691 and p-value=0.934), our interest was to check proportionality and not the best fit and it is after ascertaining proportionality assumption that we can objectively say; the covariate is insignificant. This is an indication that researchers can utilize smooth test and two-sample Kolmogorov-Smirnov test whenever other variables (e.g. sample size, number of covariate etc.) hinders accurate global test

Data 5: Survival data in this setting involves 228 patients with advanced lung cancer. The overall Schoenfeld residual plot depicts a none-zero slope. Three of the four selected covariates (age, ph.karno and pat.karno rx) also shows none-zero slope. Results shows, the global test rejects the null hypothesis at $\alpha < 0.1$ by does well for $\alpha < 0.05$. The other three test fails to reject the null hypothesis; this is coherent with the Schoenfeld residual plots. This is an indication that the global test may not be accurate.

Data 6: Data analyzed in this setting involved 146 patients with stage C prostate cancer. All the seven covariates were statistically insignificant at $\alpha < 0.05$. The overall Schoenfeld residual plot depicts a none-zero slope. However, only one (eet) of the four selected covariates two showed a zero slope. Analytic result showed that all the test are consistent and fail to rejects the null hypothesis at $\alpha < 0.05$.

Data 7: This is a setting of national intergroup trial and involved 1,858 patients with stage B and stage C colon cancer. Results show that all the tests are consistent in rejecting the null hypothesis; a situation that is coherent with both the overall Schoenfeld residual plot and the four selected covariates Schoenfeld residual plots. Despite the fact that all covariates incorporated in this model are significant, proportionality does not hold. The covariates are

therefore time dependent $\alpha < 0.05$.

Data 8: The study setting here involves a population consisting of 109 patients with small cell lung cancer. The overall Schoenfeld residual plot depicts a none-zero slope. only one coveriate (prior) of the four selected covariates shows a zero slope. In this case the global test strongly reject proportionality where as two-sample Kolmogorov-Smirnov test and Smooth test of order 3 rejects the null (proportionality) at $\alpha < 0.10$. The data-driven version of the smooth test however remains stable and fails to reject the null; an indication of proportionality. This is a situation where data-driven smooth test performs better than the other test.

Cancer data analysis showed that the smooth test and its data-driven version is a stable compared to the global and the Kolmogorov-Smirnov tests when assessing the proportional hazards assumption in variety of practical settings. Furthermore, although the smooth test does not universally dominate the other two tests in different cancer study settings, it remains relatively stable irrespective of the sample size and the number of covariates. The application of the smooth test and its data-driven version to assesses proportionality, illustrates how global test and Kolmogorov-Smirnov test inadequacies can result in invalid models. We have offered remedial measures whenever Schoenfeld residual plots shows proportionality and the global test and Kolmogorov-Smirnov test are in conflict with the inference about the plot. We therefore, implore researchers to use smooth tests of goodness-of-fit.

5.4 Smooth Tests for Baseline Hazard Function in Recurrent Events

We have extended application of BBS model to cover HIV retention data by setting loss to follow-up data to represent recurrent event scenario. This was motivated by LTFU data

comprising of typical records of patients with repeated LTFU and repeated time-to-failure (LTFU) measurements of multiple patients. First we revisited BBS model and generally discussed its features and applicability. The BBS model is used to estimate baseline hazard function. Furthermore, applying the BBS model to fit LTFU is more straightforward, particularly when the analysis involves several recurrent observation. Method of smooth test for assessing model fit for BBS model have been presented. The application of the test to assess overall fit of BBS model have also been revisited. The BBS model is often used in reliability studies. In this paper, we demonstrate that the model can be extended to cover other scenarios particularly in public health. BBS models are flexible and a typical HIV retention data can be fitted to the model. This finding stresses the point that BBS model and time-to-event analyses can be used to model LTFU. An analysis of the time to first event also shows the flexibility of the model. The application to the LTFU data is special in that we have attempted to show varied application of BBS Model. The procedures for estimating the parameters of a general and flexible class of the models for recurrent events have been revisited and its properties examined through simulation studies. Some data sets in the biomedical and reliability/engineering settings can be reanalyzed using BBS models. The importance of HIV retention and adherence is also reflected in the 2011 General Assembly Political Declaration on HIV/AIDS, which emphasizes the need to address factors that limit treatment uptake and contribute to poor adherence and calls for the mobilization and capacity building of communities to support treatment scale-up and patient retention as well as programmes that support improved treatment adherence (Assembly, 2011). A focus on the client side was also underscored by UNAIDS in their 20112015 strategy, which stated that the demand side of treatment the factors that make people enrol for treatment and adhere to ithas not received enough attention (UNAIDS, 2010; Stricker et al., 2014). One of the main challenges to the response to HIV treatment is insufficient adherence to treatment. Suboptimal viral suppression as a result of LTFU may yield a higher risk of developing drug resistances, as well as the transmission of drug resistance. We consider the

problem of testing that the baseline hazard function $\lambda(\cdot)$ of the time to first LTFU equals some specified baseline hazard function $\lambda_0(\cdot)$. The goodness-of-fit procedures were derived as score tests obtained by nesting the null hypothesis in a larger family of hazard rate functions and has been studied by Agustin and Peña (2001), Agustin and Peña (2005), Pena (1998b) and Pena (1998a). The resulting smooth test of goodness-of-fit procedures are also related to model validation procedures that utilize generalized residuals and, consequently, through the asymptotic results, the appropriate adjustments needed to properly use procedures based on generalized residuals can be obtained (Pena, 1998a). Several classes of goodness-of-fit tests, both omnibus and directional, can also be generated. Because the smooth tests are viewed as score tests, they posses optimality properties. Through a simulation study, the acceptability of the asymptotic approximations were ascertained for the BBS model, and the powers of the different tests were obtained for a wide range of alternatives.

One of the aims of this study is to provide an assessment of smooth test when modelling LTFU. The most illustrative way of describing recurrent event data is showing the different response patterns observed. Several statistical analyses are performed on the recurrent event data to determine baseline hazard function. Though the CPH model is used in recurrent events, our interest however, lies in the validation of the baseline function. This is directly important since invalid model assumption can lead to wrong statistical inference. Although patients can experience several LTFU events in a typical HIV program is a quite common situation, it should be taken into account that it is different from a recurrent event situation, in which the events are short lasting. In the study presented in this work, the events are not short lasting, but they can be more or less considered as states. When the duration of the events under consideration is short relative to the total follow up time, the definition of the time at risk for a Cox regression for recurrent events is somewhat easier. Another issue that should be taken into account is that the measurements in this on predefined time points. Although this is the situation in most experimental studies, it is also

possible to measure on a continuous time scale. This basically means that measurements are taken exactly at the moment the event of interest occurs. Therefore, the number of measurements per subject or patient and the spacing of these measurements are highly dependent on the number and spacing of the recurrent events. In these situations the definition of the time at risk can be slightly different from the ones described in this work.

We have attempted to illustrate methodological issues of smooth test through analyses of recurrence events in LTFU data. Even though main conclusions did not change in our analyses. We fitted BBS models for LTFU application in order to illustrate the use of smooth tests. We singled out that our dataset focuses on time to first LTFU since the risk is considered the same in across several epoch.

In summary, the choice of the approach for analysis of LTFU data will be determined by many factors, including: number of the drop-outs; relationship between subsequent drop-outs; effects varying or not across recurrences; HIV programming; and existing referral structure. Usually the stratified models, as PWP (total or gap times) or multi-state models, are used when there are few recurrent events per subject and the risk of LTFU recurrence varies between recurrences. On the other hand, models that account for correlation between recurrent LTFU using robust covariance matrix, time-varying covariates or frailties are indicated for frequent events with constant hazard between recurrences. Many statistical challenges arise when performing analyses of repeated LTFU data and the researcher should be careful to address them adequately.

Furthermore, simulation studies in Agustin and Peña (2001) demonstrated that the directional components of the test based on the polynomial specification tend to be flexible when the sample size is small, in contrast to the behavior of the directional components of the test based on the orthogonal specification. Also, each component of the test based on the orthogonal specification has the potential of detecting specific departures from the hypothesized hazard rate.

We have extended application of hazard-based smooth GOF tests to model LTFU data.

The appeal of the class of smooth test lies in the fact that a rich family of tests can be generated by varying the smoothing process $\Psi(\cdot)$. Simulation studies showed that the orthonormal choices leads to powerful omnibus tests and there individual components good directional tests. Simulation results showed doers $k = 2$ and $k = 3$ is appropriate. This is consistent with Agustin and Peña (2001).

Chapter 6

Conclusions and Recommendation

6.1 Conclusion

As mentioned in the introduction, applications of smooth test of GOF to models fitted to LTFU has not been considered before. Authors who have looked at smooth tests of GOF have concentrated on almost two real data applications that didn't include LTFU data. The smooth test applied in this study can be extended to cover data-driven applications and can also be adjusted to detect particular alternatives. They are characterized by the following properties: The test statistics are asymptotically distribution free; The asymptotic distribution of the test statistics can be determined under both the null and alternative hypotheses; In the important i.i.d. case consistency can be proven; The tests can be tailored to detect specific alternatives, in particular Cox's model, by an appropriate choice of $\Psi(\cdot)$; No estimate of the baseline $\lambda_0(\cdot)$ is needed in the test against CPH model and BBS model.

The smooth test of GOF approach performs better than empirical GOF test when fitting a parametric distribution to time-to-event complete data. We describe how to fit a two parameter Weibull Distribution to retention data and assesses the fit using goodness-of-fit procedures. Smooth test is comparatively better when fitting a parametric distribution to time to LTFU data in a typical clinical setting. Our results highlight the need to better

understand LTFU of patients initiated on ART. Nuisance parameter estimation can be performed without changing the test statistic and since the tests rely on maximum likelihood techniques, they asymptotically meet the conditions of the Neyman-Pearson lemma against any simple alternative hypothesis. Further studies that address fitting hazard functions in the presence of censored data and determinants of risks to LTFU are required for clarity.

Neyman (1937) paper have had significant contribution in the statistical scene. Starting from the first principles of testing, Neyman derived an optimal test statistic and discussed its applications along with its possible drawbacks. In terms of its significance in the history of hypothesis testing, Neyman's smooth test comparably performs well when testing principles that satisfied certain optimality criteria. There is a lot of potential when borrowing suitable statistical techniques and adapting them for real data applications. Given the fundamental nature of Neyman's contribution, the smooth test has not been formally used in many different practical situations across different sectors despite the fact that the test is known to perform well with both censored and uncensored data. This thesis is our modest attempt to bring Neyman's smooth test application to HIV research. Further numerical investigation into the performance of the test in various practical applications is required.

6.2 Strength and Limitation of The Study

6.2.1 Strength of The Study

Primary data used in the study was obtained from Kenya's typical HIV setting and is therefore more broadly generalizable. Data was pulled from a database that have been verified. The other strength is that, although the effect of LTFU on clinical outcomes is not a new issue, there is no study that have systematically endeavoured to assesses the fit of a parametric distribution or hazard function to loss to follow-up data.

6.2.2 Limitation of the study

The study also has several limitations. Because data for missed clinic visits only began to be collected in after 2007, they are available on only a small sample of the population and therefore must be interpreted with caution. In addition, it is possible that some individuals lost to follow-up had died, moved or transferred to another health-care provider without informing the clinic, and that this resulted in some outcomes being misclassified. However, individuals were considered lost to follow-up only after the outreach programme had attempted to locate them. The duration between the point of first LTFU and retracing patients back to care may be significantly long. Its a random time and requires further investigation. Several factors associated with being lost to follow-up are potentially modifiable and provide opportunities for improving HIV care programmes in sub-Saharan Africa.

6.3 Future Direction

This work have only covered continuous case. We have identified potential counting data sets that can be fit to probability and hazard models and use smooth tests of GOF to assesses underling distribution and hazard functions. Future potential areas involving count data include

1. Family planing and couple year of protection (CYP) data. This can be fitted to Poisson distribution and Negative binomial distribution in order to test over dispersion

2. Number pregnant mothers in the first and fourth Antenatal visits. Test for dispersion by fitting Poisson distribution and Negative binomial distribution

3. HIV Sero-conversion among discordant couple data. We can model a Geometric distribution

Bibliography

Aalen, O. (1978). Non-Parametric Inference for a Family of Counting Processes. *The Annals of Statistics*, pages 701–726.

Aalen, O. O. and Husebye, E. (1991). Statistical Analysis of Repeated Events forming Renewal Processes. *Statistics in medicine*, 10(8):1227–1240.

Adekpedjou, A., Zamba, K., et al. (2012). A Chi-squared Goodness-of-fit Test for Recurrent-Event Data. *Journal of Statistical Theory and Applications*, 11(2):97–119.

Agustin, M. Z. N. and Pena, E. A. (1999). Order Statistic Properties, Random Generation, and Goodness-of-fit Testing for a Minimal Repair Model. *Journal of the American statistical association*, 94(445):266–272.

Agustin, M. Z. N. and Peña, E. A. (2001). Goodness-of-fit of the Distribution of Time-to-first-occurrence in Recurrent Event Models. *Lifetime Data Analysis*, 7(3):289–306.

Agustin, M. Z. N. and Peña, E. A. (2005). A Basis Approach to Goodness-of-fit Testing in Recurrent Event Models. *Journal of statistical planning and inference*, 133(2):285–303.

Andersen, P. K. (1982). Testing Goodness of fit of Cox's Regression and Life Model. *Biometrics*, pages 67–77.

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (2012). *Statistical Models Based on Counting Processes*. Springer Science & Business Media.

Andersen, P. K. and Gill, R. D. (1982). Cox's Regression Model for Counting Processes: A Large Sample Study. *The annals of statistics*, pages 1100–1120.

Arjas, E. (1988). A Graphical Method for Assessing Goodness of fit in Cox's Proportional Hazards Model. *Journal of the American Statistical Association*, 83(401):204–212.

Asiimwe, S. B., Kanyesigye, M., Bwana, B., Okello, S., and Muyindike, W. (2016). Predictors of Dropout from Care among HIV-infected Patients Initiating Antiretroviral Therapy at a Public Sector HIV Treatment Clinic in sub-Saharan Africa. *BMC infectious diseases*, 16(1):1.

Assembly, U. G. (2011). Political Declaration on HIV and AIDS: Intensifying our Efforts to Eliminate HIV and AIDS. *65th Session. New York: UN General Assembly (UNGASS)*.

Baltazar-Aban, I. and Pena, E. A. (1995). Properties of Hazard-based Residuals and Implications in Model Diagnostics. *Journal of the American Statistical Association*, 90(429):185–197.

Bargal, A. I. and Thomas, D. R. (1983). Smooth Goodness of fit Tests for the Weibull Distribution with Singly Censored Data. *Communications in Statistics-Theory and Methods*, 12(12):1431–1447.

Barlow, W. E. and Prentice, R. L. (1988). Residuals for Relative Risk Regression. *Biometrika*, 75(1):65–74.

Barton, D. E. (1956). Neyman's Test of Goodness of fit when the Null Hypothesis is Composite. *Scandinavian Actuarial Journal*, 1956(2):216–245.

Baxter, L. A., Kijima, M., and Tortorella, M. (1996). A Point Process Model for the Reliability of a Maintained System Subject to General Repair. *Stochastic models*, 12(1):12–1.

Berheto, T. M., Haile, D. B., Mohammed, S., et al. (2014). Predictors of Loss to Follow-up in Patients Living with HIV/AIDS after Initiation of Antiretroviral Therapy. *North American journal of medical sciences*, 6(9):453.

Best, D. and Rayner, J. (1985). Uniformity Testing when Alternatives have Low Order. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 25–35.

Block, H. W., Borges, W. S., and Savits, T. H. (1985). Age-Dependent Minimal Repair. *Journal of applied probability*, pages 370–385.

Brown, M. and Proschan, F. (1983). Imperfect Repair. *Journal of Applied probability*, pages 851–859.

Chambers, J. M., Cleveland, W. S., Kleiner, B., and Tukey, P. A. (1983). *Graphical Methods for Data Analysis*.

Chen, H.-S., Lai, K., and Ying, Z. (2004). Goodness-of-fit Tests and Minimum Power Divergence Estimators for Survival Data. *Statistica Sinica*, pages 231–248.

Computing, S. (1991). R Foundation for Statistical Computing, Vienna, Austria.

Cook, R. J. and Lawless, J. (2007). *The Statistical Analysis of Recurrent Events*. Springer Science & Business Media.

Council, K. A. C. (2014). Response Progress Report 2014: Progress Towards Zero. *Nairobi, Kenya: Government Printer*.

Cousineau, D. (2009). Fitting the Three-parameter Weibull Distribution: Review and Evaluation of Existing and New Methods. *IEEE Transactions on Dielectrics and Electrical Insulation*, 16(1):281–288.

Cox, D. R. (1972). Regression Models and Life Tables (with discussion). *Journal of the Royal Statistical Society*, 34:187–220.

Cox, D. R. (1975). Partial Likelihood. *Biometrika*, 62(2):269–276.

Crowley, J. and Hu, M. (1977). Covariance Analysis of Heart Transplant Survival Data. *Journal of the American Statistical Association*, 72(357):27–36.

David, F. (1939). On Neyman's "Smooth" Test for Goodness of fit: I. Distribution of the Criterion ψ^2 when the Hypothesis Tested is True. *Biometrika*, 31(1/2):191–199.

De Boeck, B., Thas, O., Rayner, J., and Best, D. J. (2011). Smooth Tests for the Gamma Distribution. *Journal of Statistical Computation and Simulation*, 81(7):843–855.

Dorado, C., Hollander, M., and Sethuraman, J. (1997). Non-Parametric Estimation for a General Repair Model. *The Annals of Statistics*, pages 1140–1160.

Dorado, C. A. (1995). *On a General Repair Model for Repairable Systems*. PhD thesis, Florida State University.

Edmonson, J. H., Fleming, T. R., Decker, D., Malkasian, G., Jorgensen, E., Jefferies, J., Webb, M., and Kvols, L. (1979). Different Chemotherapeutic Sensitivities and Host factors Affecting Prognosis in Advanced Ovarian Carcinoma versus Minimal Residual Disease. *Cancer treatment reports*, 63(2):241–247.

Fleming, T. R. and Harrington, D. P. (2011). *Counting Processes and Survival Analysis*, volume 169. John Wiley & Sons.

Gatsonis, C., Hsieh, H., and Korwar, R. (1985). Simple Non-Parametric Tests for a Known Standard Survival Based on Censored Data. *Communications in Statistics-Theory and Methods*, 14(9):2137–2162.

Gwynn, R. C., Fawzy, A., Viho, I., Wu, Y., Abrams, E. J., and Nash, D. (2015). Risk Factors for Loss to Follow-up Prior to ART Initiation among Patients Enrolling in HIV

Care with CD4+ cell Count $\geq$ 200 cells/μl in the Multi-country MTCT-Plus Initiative. *BMC health services research*, 15(1):1.

Haddow, L., Edwards, S., Sinka, K., and Mercey, D. (2003). Patients Lost to Follow up: Experience of an HIV Clinic. *Sexually transmitted infections*, 79(4):349–350.

Hahn, G. J. and Shapiro, S. S. (1968). Statistical Models in Engineering. In *Statistical models in engineering*. John Wiley & Sons.

Hankey, B. F., Feuer, E. J., Clegg, L. X., Hayes, R. B., Legler, J. M., Prorok, P. C., Ries, L. A., Merrill, R. M., and Kaplan, R. S. (1999). Cancer Surveillance Series: Interpreting Trends in Prostate Cancerpart I: Evidence of the Effects of Screening in recent Prostate Cancer Incidence, Mortality, and Survival Rates. *Journal of the National Cancer Institute*, 91(12):1017–1024.

Hirose, Y. (2011). Asymptotic Linear Expansion of Profile Likelihood in the Cox Model. *Mathematical Methods of Statistics*, 20(3):224–231.

Hollander, M., Presnell, B., and Sethuraman, J. (1992). Non-Parametric Methods for Imperfect Repair Models. *The Annals of Statistics*, pages 879–896.

Huber-Carol, C., Balakrishnan, N., Nikulin, M., and Mesbah, M. (2012). *Goodness-of-fit Tests and Model Validity*. Springer Science & Business Media.

Inglot, T., Kallenberg, W. C., and Ledwina, T. (1994). Power Approximations to and Power Comparison of Smooth Goodness-of-fit Tests. *Scandinavian Journal of Statistics*, pages 131–145.

Inglot, T., Kallenberg, W. C., and Ledwina, T. (1997). Data driven smooth tests for composite hypotheses. *The Annals of Statistics*, pages 1222–1250.

Janic-Wróblewska, A. (2004). Data-Driven Smooth Tests for the Extreme Value Distribution. *Statistics*, 38(5):413–426.

Kalbfleisch, J. D. and Prentice, R. L. (2011). *The Statistical Analysis of Failure-time Data*, volume 360. John Wiley & Sons.

Kallenberg, W. C. and Ledwina, T. (1995). Consistency and Monte Carlo Simulation of a Data-Driven Version of Smooth Goodness-of-fit Tests. *The Annals of Statistics*, pages 1594–1608.

Kang, S.-I. (1979). *Performance of Generalized Neyman Smooth Goodness of fit Tests*. PhD thesis, Department of Statistics, Oregon State University.

Kay, R. (1977). Proportional Hazard Regression Models and the Analysis of Censored Survival Data. *Applied Statistics*, pages 227–237.

Kelly, P. J. and Lim, L. L.-Y. (2000). Survival Analysis for Recurrent-event Data: An Application to Childhood Infectious Diseases. *Statistics in medicine*, 19(1):13–33.

Kijima, M. (1989). Some Results for Repairable Systems with General Repair. *Journal of Applied probability*, pages 89–102.

Kimber, A. (1985). Tests for the Exponential, Weibull and Gumbel Distributions Based on the Stabilized Probability Plot. *Biometrika*, 72(3):661–663.

Koziol, J. (1987). An Alternative Formulation of Neyman's Smooth Goodness-of-fit Tests under Composite Alternatives. *Metrika*, 34(1):17–24.

Kraus, D. (2007a). Data-Driven Smooth Tests of the Proportional Hazards Assumption. *Lifetime data analysis*, 13(1):1–16.

Kraus, D. (2007b). *Neyman's Smooth Tests in Survival Analysis*. PhD thesis, PhD thesis. Charles University in Prague, Dept. of Statistics.

Kraus, D. (2009). Adaptive Neyman's Smooth Tests of Homogeneity of Two-samples of Survival Data. *Journal of Statistical Planning and Inference*, 139(10):3559–3569.

Krit, M. (2014). *Goodness-of-fit Tests in Reliability: Weibull Distribution and Imperfect Maintenance Models*. PhD thesis, Université de Grenoble.

Langovoy, M. (2007). Data-Driven Goodness-of-fit Tests. *arXiv preprint arXiv:0708.0169*.

LaRiccia, V. (1991). Smooth Goodness-of-fit Tests: A Quantile Function Approach. *Journal of the American Statistical Association*, 86(414):427–431.

Lawless, J. F. (2011). *Statistical Models and Methods for Lifetime Data*, volume 362. John Wiley & Sons.

Ledwina, T. (1994). Data-Driven Version of Neyman's Smooth Test of fit. *Journal of the American Statistical Association*, 89(427):1000–1005.

Lemeshko, B. Y., Lemeshko, S., and Postovalov, S. (2009). Comparative Analysis of the Power of Goodness-of-fit Tests for near Competing Hypotheses I. the Verification of Simple Hypotheses. *Journal of Applied and Industrial Mathematics*, 3(4):462–475.

Lim, H. J. and Zhang, X. (2011). Additive and Multiplicative Hazards Modeling for Recurrent-event Data Analysis. *BMC medical research methodology*, 11(1):1.

Lin, D. (1991). Goodness-of-fit Analysis for the Cox Regression Model Based on a Class of Parameter Estimators. *Journal of the American Statistical Association*, 86(415):725–728.

Lin, D. Y., Wei, L.-J., and Ying, Z. (1993). Checking the Cox Model with Cumulative Sums of Martingale-based Residuals. *Biometrika*, 80(3):557–572.

Liu, J.-c., Liu, H.-b., and Zhang, R.-q. (2016). The Proportional Hazards Model for Multiple type Recurrent Gap-times. *Acta Mathematicae Applicatae Sinica, English Series*, 32(1):221–230.

Liu, X. (2012). *Survival Analysis: Models and Applications*. John Wiley & Sons.

Loprinzi, C. L., Laurie, J. A., Wieand, H. S., Krook, J. E., Novotny, P. J., Kugler, J. W., Bartel, J., Law, M., Bateman, M., and Klatt, N. E. (1994). Prospective Evaluation of Prognostic Variables from Patient-completed Questionnaires. North Central Cancer Treatment Group. *Journal of Clinical Oncology*, 12(3):601–607.

Makunde, W. H., Francis, F., Mmbando, B. P., Kamugisha, M. L., Rutta, A. M., Mandara, C. I., and Msangeni, H. A. (2012). Lost to Follow up and Clinical Outcomes of HIV Adult Patients on Antiretroviral Therapy in Care and Treatment Centres in Tanga City, North-Eastern Tanzania. *Tanzania journal of health research*, 14(4).

McHugh, M. L. (2013). The Chi-square Test of Independence. *Biochemia Medica*, 23(2):143–149.

McKeague, I. W. and Utikal, K. J. (1991). Goodness-of-fit Tests for Additive Hazards and Proportional Hazards Models. *Scandinavian Journal of Statistics*, pages 177–195.

Megerso, A., Garoma, S., Tolosa Eticha, T. W., Daba, S., Tarekegn, M., and Habtamu, Z. (2016). Predictors of Loss to Follow-up in Antiretroviral Treatment for Adult Patients in the Oromia region, Ethiopia. *HIV/AIDS (Auckland, NZ)*, 8:83.

Micke, P., Faldum, A., Metz, T., Beeh, K.-M., Bittinger, F., Hengstler, J.-G., and Buhl, R. (2002). Staging Small Cell Lung Cancer: Veterans Administration Lung Study Group versus International Association for the Study of Lung Cancerwhat Limits Limited Disease? *Lung cancer*, 37(3):271–276.

Miller Jr, F. L. and Quesenberry, C. (1979). Power Studies of Tests for Uniformity II. *Communications in Statistics-Simulation and Computation*, 8(3):271–290.

Miller Jr, R. G. (2011). *Survival Analysis*, volume 66. John Wiley & Sons.

Moertel, C. G., Fleming, T. R., Macdonald, J. S., Haller, D. G., Laurie, J. A., Goodman, P. J., Ungerleider, J. S., Emerson, W. A., Tormey, D. C., Glick, J. H., et al. (1990). Levamisole and Fluorouracil for Adjuvant Therapy of Resected Colon Carcinoma. *New England Journal of Medicine*, 322(6):352–358.

Moons, K. G., Royston, P., Vergouwe, Y., Grobbee, D. E., and Altman, D. G. (2009). Prognosis and Prognostic Research: What, Why, and How? *Bmj*, 338:b375.

Munk, A., Stockis, J.-P., Valeinis, J., and Giese, G. (2011). Neyman Smooth Goodness-of-fit Tests for the Marginal Distribution of Dependent Data. *Annals of the Institute of Statistical Mathematics*, 63(5):939–959.

Nagelkerke, N., Oosting, J., and Hart, A. (1984). A Simple Test for Goodness of fit of Cox's Proportional Hazards Model. *Biometrics*, pages 483–486.

Nelson, W. B. (2003). *Recurrent Events Data Analysis for Product Repairs, Disease Recurrences, and Other Applications*. SIAM.

Neyman, J. (1937). Smooth Test for Goodness of fit. *Scandinavian Actuarial Journal*, 1937(3-4):149–199.

Ochieng-Ooko, V., Ochieng, D., Sidle, J. E., Holdsworth, M., Wools-Kaloustian, K., Siika, A. M., Yiannoutsos, C. T., Owiti, M., Kimaiyo, S., and Braitstein, P. (2010). Influence of Gender on Loss to Follow-up in a Large HIV Treatment Programme in Western Kenya. *Bulletin of the World Health Organization*, 88(9):681–688.

Odhiambo, Collins, J., Odhiambo, and Omolo, B. (2017a). A Smooth Test of Goodness-of-fit for the Weibull Distribution: An Application to an HIV Retention Data. *International Journal of Statistics in Medical Research*, 6(1):68–78.

Odhiambo, Collins, J., Odhiambo, and Omolo, B. (2017b). Validation of the Smooth Test of Goodness-of-fit for Proportional Hazards in Cancer Survival Studies. *International Journal of Statistics in Medical Research*, 6(2):49–67.

O'Quigley, J. and Pessione, F. (1989). Score Tests for Homogeneity of Regression Effect in the Proportional Hazards Model. *Biometrics*, pages 135–144.

Pearson, K. (1992). On the Criterion that a Given System of Deviations from the Probable in the Case of a Correlated System of Variables is Such that it Can be Reasonably Supposed to have Arisen from Random Sampling. In *Breakthroughs in Statistics*, pages 11–28. Springer.

Pena, E. A. (1998a). Smooth Goodness-of-fit Tests for Composite Hypothesis in Hazard based Models. *The Annals of Statistics*, 26(5):1935–1971.

Pena, E. A. (1998b). Smooth Goodness-of-fit Tests for the Baseline Hazard in Cox's Proportional Hazards Model. *Journal of the American Statistical Association*, 93(442):673–692.

Persson, I. (2002). *Essays on the Assumption of Proportional Hazards in Cox Regression*. Acta Universitatis Upsaliensis,.

Preedalikit, K., Liu, I., Hirose, Y., Sibanda, N., and Fernández, D. (2016). Joint Modeling of Survival and Longitudinal Ordered Data Using a Semiparametric Approach. *Australian & New Zealand Journal of Statistics*, 58(2):153–172.

Purohit, S. G. (1994). Goodness of fit Tests Based on Recurring Failure Data. *Sankhyā: The Indian Journal of Statistics, Series B*, pages 76–85.

Quantin, C., Moreau, T., Asselain, B., Maccario, J., and Lellouch, J. (1996). A Regression Survival Model for Testing the Proportional Hazards Hypothesis. *Biometrics*, pages 874–885.

Rachlis, B., Cole, D. C., van Lettow, M., Escobar, M., Muula, A. S., Ahmad, F., Orbinski, J., and Chan, A. K. (2014). Follow-Up Visit Patterns in an Antiretroviral Therapy (ART) Programme in Zomba, Malawi. *PloS one*, 9(7):e101875.

Ramadhani, H. O., Thielman, N. M., Landman, K. Z., Ndosi, E. M., Gao, F., Kirchherr, J. L., Shah, R., Shao, H. J., Morpeth, S. C., McNeill, J. D., et al. (2007). Predictors of Incomplete Adherence, Virologic failure, and Antiviral Drug Resistance among HIV-infected Adults Receiving Antiretroviral Therapy in Tanzania. *Clinical Infectious Diseases*, 45(11):1492–1498.

Rasschaert, F., Koole, O., Zachariah, R., Lynen, L., Manzi, M., and Van Damme, W. (2012). Short and Long term Retention in Antiretroviral Care in Health Facilities in Rural Malawi and Zimbabwe. *BMC health services research*, 12(1):1.

Rayner, G. D. and Rayner, J. C. (2001). Power of the Neyman Smooth Tests for the Uniform Distribution. *Advances in Decision Sciences*, 5(3):181–191.

Rayner, J. and Best, D. (1986). Neyman-type Smooth Tests for Location-scale Families. *Biometrika*, 73(2):437–446.

Rayner, J. and Best, D. (1990). Smooth Tests of Goodness of fit: An Overview. *International Statistical Review/Revue Internationale de Statistique*, pages 9–17.

Rayner, J., Thas, O., and De Boeck, B. (2008). A Generalized Emerson Recurrence Relation. *Australian & New Zealand Journal of Statistics*, 50(3):235–240.

Rayner, J. C., Thas, O., and Best, D. J. (2009). *Smooth Tests of Goodness of fit: Using R*. John Wiley & Sons.

Rayner, J. C., Thas, O., Best, D. J., et al. (2015). Statistical Distribution Models: Goodness of fit Tests— NOVA. the University of Newcastle's Digital Repository.

Schoenfeld, D. (1980). Chi-squared Goodness-of-fit Tests for the Proportional Hazards Regression Model. *Biometrika*, 67(1):145–153.

Schoenfeld, D. (1982). Partial Residuals for the Proportional Hazards Regression Model. *Biometrika*, 69(1):239–241.

Sengayi, M., Dwane, N., Marinda, E., Sipambo, N., Fairlie, L., and Moultrie, H. (2013). Predictors of Loss to Follow-up among Children in the First and Second years of Antiretroviral Treatment in Johannesburg, South Africa. *Glob Health Action*, 6:19248.

Siegel, R. L., Miller, K. D., and Jemal, A. (2015). Cancer Statistics, 2015. *CA: A cancer journal for clinicians*, 65(1):5–29.

Stigler, S. M. et al. (2008). Karl Pearson's Theoretical Errors and the Advances they Inspired. *Statistical Science*, 23(2):261–271.

Stocker IV, R. S. and Peña, E. A. (2012). A General Class of Parametric Models for Recurrent Event Data. *Technometrics*.

Stricker, S. M., Fox, K. A., Baggaley, R., Negussie, E., de Pee, S., Grede, N., and Bloem, M. W. (2014). Retention in Care and Adherence to ART are Critical Elements of HIV Care Interventions. *AIDS and Behavior*, 18(5):465–475.

Stuart, A., Kendall, M. G., et al. (1968). *The Advanced Theory of Statistics*. Charles Griffin.

Sürücü, B. (2008). A Power Comparison and Simulation Study of Goodness-of-fit Tests. *Computers & Mathematics with Applications*, 56(6):1617–1625.

Therneau, T. M., Grambsch, P. M., and Fleming, T. R. (1990). Martingale-based Residuals for Survival Models. *Biometrika*, 77(1):147–160.

Thomas, D. R. and Pierce, D. A. (1979). Neyman's Smooth Goodness-of-fit Test when the Hypothesis is Composite. *Journal of the American Statistical Association*, 74(366a):441–445.

Tiku, M. L. (1986). *Robust Inference*. Marcel Dekker, Inc.

UNAIDS (2010). Getting to Zero: 2011–2015 Strategy.

UNAIDS (2014). 90-90-90: An Ambitious Treatment Target to Help end the AIDS Epidemic.

Wang, B., Losina, E., Stark, R., Munro, A., Walensky, R. P., Wilke, M., Martin, D., Lu, Z., Freedberg, K. A., and Wood, R. (2011). Loss to Follow-up in a Community Clinic in South Africa: Roles of Gender, Pregnancy and CD4+ Count. *SAMJ: South African Medical Journal*, 101(4):253–257.

Wei, L. (1984). Testing Goodness of fit for Proportional Hazards Model with Censored Observations. *Journal of the American Statistical Association*, 79(387):649–652.

Xue, X., Xie, X., Gunter, M., Rohan, T. E., Wassertheil-Smoller, S., Ho, G. Y., Cirillo, D., Yu, H., and Strickler, H. D. (2013). Testing the Proportional Hazards Assumption in Case-cohort Analysis. *BMC medical research methodology*, 13(1):1.

Appendix A

Formulae

A.1 Smooth Tests of GOF for Probability Distribution

A.1.1 Orthonormal Polynomials of Probability Distribution Using Emerson Recurrence Relation

The Emerson relation provides computational relation to enable construction of orthonormal polynomials (Rayner et al., 2008).

Let $\{h_i(x)\}$ be the set of `orthonormal` polynomials of a given probability distribution. Let $E[\cdot]$ be expectation of the probability distribution i.e. $E[X] = \mu$ and $\mu_i = E[(X-\mu)^i]$ for $i = 2,3,\cdots$. Let $\{h_i^*(x)\}$ be a set of `orthogonal` polynomials with $\{h_0^*(x)\} = 1$ for all x. We write $c_i = E[\{h_i^*(x)\}^2]$ for $i = 0,1,2,\cdots$. Therefore

$$\{h_i(x)\} = \frac{\{h_i^*(x)\}}{\sqrt{c_i}}, \tag{A.1}$$

for all x and $i = 0,1,2,\cdots$, then $\{h_i(x)\}$ is the corresponding set of `orthonormal` polynomials. The initial polynomials are $h_0(x) = h_0^*(x) = 1$ for all x, $h_1^*(x) = (x-\mu)$ and $h_1(x) = \frac{(x-\mu)}{\sqrt{\mu_2}}$. Note that $c_0 = 1$ and $c_1 = \mu_2$.

In expectation form, the Emerson recurrence relation is (Rayner et al., 2008; Rayner and Best, 1990)

$$h_i^*(x) = (x - E[Xh_{i-1}^2(x)])h_{i-1}(x) - E[Xh_{i-2}(x)Xh_{i-1}(x)]h_{i-1}(x). \tag{A.2}$$

Example, for $r = 2$ equation (A.2) generates $h_2^*(x) = \frac{[(x-\mu)^2 - \mu_3/\mu_2(x-\mu) - \mu_2]}{\sqrt{\mu_2}}$.
The following two theorems are can be generated from equation (A.2)

- If $h_0(x) = 1$ for all x and $h_1(x) = (x-\mu)/\sqrt{\mu_2}$, then for $r = 2, 3, \cdots, h_r^*$, defined by equation (A.2), is orthogonal to $h_0(x), h_1(x), \cdots, h_{r-1}(x)$.

- For $r = 2, 3, \cdots$ and $i = 0, 1, \cdots, r$,

$$a_{i,r} = \frac{a_{i-1,r-1}}{\sqrt{c_{r-1}}} + \frac{b_{r-1,r}a_{i,r-1}}{\sqrt{c_{r-1}}} + \frac{b_{r-2,r}a_{i,r-2}}{\sqrt{c_{r-2}}},$$

  with boundary conditions $a_{0,0} = 1$, $c_0 = 1$, $a_{1,1} = 1/\sqrt{\mu_2}$, $a_{0,1} = -\mu/\sqrt{\mu_2}$, $c_1 = 1$. This follows by equating coefficient in

$$h_r^*(x) = \sum_{i=0}^{r} a_{i,r}x^i = xh_{r-1}(x) + b_{r-1,r}h_{r-1}(x) + b_{r-2,r}h_{r-2}(x).$$

- Assuming the polynomials $h_0^*(x), h_1^*(x), \cdots, h_{r-1}^*(x)$ and the constants $c_0, c_1, \cdots, c_{r-1}$ are known and the necessary moments required exist and are known. We write $\mu_r^{'} = E[X^r]$ for $r = 1, 2, \cdots$. For $r = 2, 3, \cdots$, the quantities in equation (A.2), and c_r can be obtained from

$$E[Xh_{r-1}^2(X) = -b_{r-1,r} = \frac{\sum_{j=0}^{r-1}\sum_{k=0}^{r-1} a_{j,r-1}a_{k,r-1}\mu_{j+k+1}^{'}}{c_{r-1}},$$

$$E[Xh_{r-1}(X)h_{r-2}(X)] = -b_{r-2,r} = \frac{\sum_{j=0}^{r-1}\sum_{k=0}^{r-2} a_{j,r-1}a_{k,r-2}\mu_{j+k+1}^{'}}{\sqrt{c_{r-1}c_{r-2}}}$$

and

$$c_r = \frac{\sum_{j=0}^{r-1}\sum_{k=0}^{r-1} a_{j,r-1}a_{k,r-1}\mu'_{j+k+1}}{c_{r-1}} - b^2_{r-1,r} - b^2_{r-2,r}.$$

A.1.2 Prepositions for Data Driven version of Smooth Tests

The idea of data-driven version of smooth tests was introduced by Ledwina (1994). He applied the Bayesian information criterion (BIC, Schwarzs selection rule) to the task of testing uniformity (or other single distribution).

Let $X_1, X_2, \cdots, X_n$ be iid random variables with probability density $f(x;\theta)$ belonging to μ_s for some s. Let the likelihood function be defined as

$$L_s(\theta) = \log \prod_{i=1}^{n} f(x_i;\theta), \tag{A.3}$$

and

$$L_s = \sup_{\theta\in\Omega_s} L_s(\theta), \tag{A.4}$$

i.e.

$$\beth_s = L_s - \frac{1}{2} s \log n. \tag{A.5}$$

We use Schwarz's technique to select the model with index S defined by

$$S = \min\{j, 1 \le j \le K : \beth_j = \max_{1\le s\le k} \beth_s\}. \tag{A.6}$$

The index of μ_s is selected with respect to the sample size n so that the density function yields high likelihood on the data.

Let N_s be the data-driven version smooth test. The main properties of Schwarz's technique and N_s are stated below.

With $f(\cdot,\theta) \in \mu_s$ for $s = 1, 2, \cdots, K$.

1. *Preposition 1*: Let the law of the observed random variables belongs to the model μ_s

for $s = 1, 2, \cdots, K$. Then, the probability $\lim_{n \longrightarrow \infty} f_\theta^n = 1$, the Schwarz's rule leads to the choice of the lowest-dimensional model containing the true value of θ.

2. *Proposition 2*: If $\theta \in \Omega_s \Omega_{s-1}$ for $s = 1, 2, \cdots, K$ then

$$P_\theta^n = 0(e^{n\beta})$$

for some $\beta > 0$ and for $1 < i < k - s$ it holds that

$$P_\theta^n(S \geq s + i) = 0(n^{i/2}(\log n)^{(s+i-2)/2}).$$

3. *Preposition 3*: Suppose that H_0 is true. Then, as $n \to \infty$, the distribution of N_s is approximated by the central chi-squared distribution with 1 degree of freedom.

4. *Preposition 4*: If $\theta \in \Omega_s \Omega_{s-1}$ for some $s = 1, 2, \cdots, K$ and $\theta \neq 0$, then the power of N_s tends to 1 as $n \to \infty$.

A.2 Smooth Tests of GOF for Hazard Functions

Regularity conditions

The regularity conditions required to generate asymptotic results for the Ψ process.

For a vector $\mathbf{v} = (v_1, v_2, \cdots, V_m)^t, |\mathbf{V}| = \sqrt{\mathbf{V}^t \mathbf{V}}$, while for a matrix $\mathbf{V} = (v_{ij})$, let $||\mathbf{V}|| = \max_{i,j} |v_{ij}|$

1. There exists a neighborhood $\Gamma_0 \subseteq \Gamma$ of η_0 such that on $F \times \Gamma_0, \lambda_O(t; \eta) > 0$, and the partial derivatives $(\frac{\partial}{\partial \eta_i})\lambda_O(t; \eta), (\frac{\partial^2}{\partial \eta_i \eta_j})\lambda_O(t; \eta)$ and $(\frac{\partial^3}{\partial \eta_i \eta_j \eta_k})\lambda_O(t; \eta)$ exist and are continuous at $\eta_0 = \eta$.

2. On $F \times \Gamma_0$, the log-likelihood process

$$l(t;\theta,\eta) = \int_0^t \log[Y(s)\lambda(s;\theta,\eta)]dN(s) - \int_0^t Y(s)\lambda(s;\theta,\eta)ds$$

can be differentiated three times with respect to η and with the order of the differentiation and integration operations being interchangeable.

3. On $F \times \Gamma_0$, the partial derivatives $(\frac{\partial}{\partial\eta_i})\lambda_O(t;\eta)$, $(\frac{\partial^2}{\partial\eta_i\eta_j})\lambda_O(t;\eta)$ and $(\frac{\partial^3}{\partial\eta_i\eta_j\eta_k})\lambda_O(t;\eta)$ exist and are continuous at $\eta_0 = \eta$ and with the processes $\{\psi(t;\eta_0) : t \in F\}$ and $\left\{(\frac{\partial}{\partial\eta_i})\psi(t;\eta_0) : t \in F\right\}$ being locally bounded and predictable.

4. There exist deterministic functions $y(\cdot)$ defined on F, and $\psi^{(0)}(\cdot;\cdot)$ defined on $F \times \Gamma_0$, such that for $(t,\eta) \in F \times \Gamma_0$ and as $n \to \infty$

$$\left|\left|a_n^{-2}\int_0^t \psi(s;\eta)^{\otimes 2}Y(s)\lambda_0(s;\eta)ds - \int_0^t \psi^{(0)}(s;\eta)^{\otimes 2}y(s)\lambda_0(s;\eta)ds\right|\right| \xrightarrow{pr} 0;$$

$$\left|\left|a_n^{-2}\int_0^t \psi(s;\eta)^{\otimes 2}\rho(t;\eta)^t Y(s)\lambda_0(s;\eta)ds - \int_0^t \psi^{(0)}(s;\eta)^{\otimes 2}\rho(t;\eta)^t y(s)\lambda_0(s;\eta)ds\right|\right| \xrightarrow{pr} 0;$$

$$\left|\left|a_n^{-2}\int_0^t \rho(t;\eta)^{\otimes 2}Y(s)\lambda_0(s;\eta)ds - \int_0^t \rho(t;\eta)^{\otimes 2}y(s)\lambda_0(s;\eta)ds\right|\right| \xrightarrow{pr} 0.$$

5. Defining the matrices of functions on $F \times \Gamma_0$ given by

$$\Sigma_{11}(t;\eta) = \int_0^t \psi^{(0)}(s;\eta)^{\otimes 2}y(s)\lambda_0(s;\eta)ds;$$

$$\Sigma_{21}(t;\eta) = \Sigma_{12}(t;\eta)^t = \int_0^t \psi^{(0)}(s;\eta)^{\otimes 2}\rho(t;\eta)^t y(s)\lambda_0(s;\eta)ds;$$

$$\Sigma_{22}(t;\eta) = \int_0^t \rho(t;\eta)^{\otimes 2}y(s)\lambda_0(s;\eta)ds$$

the $(k+q)\times(q+k)$ matrix

$$\begin{pmatrix} \Sigma_{11}(t;\eta) & \Sigma_{12}(t;\eta) \\ \Sigma_{21}(t;\eta) & \Sigma_{22}(t;\eta) \end{pmatrix}$$

have finite elements for each $t \in F$, and $\Sigma(\tau;\eta)$ is positive definite.

6. For each $\varepsilon > 0$ and $t \in F$, as $n \to \infty$

$$a_n^{-2} \int_0^t |\psi(s;\eta_0)|^2 I\{|\psi(s;\eta_0)| > \varepsilon a_n\} Y(s)\lambda_0(s;\eta_0)ds \xrightarrow{pr} 0;$$

$$a_n^{-2} \int_0^t |\rho(s;\eta_0)|^2 I\{|\rho(s;\eta_0)| > \varepsilon a_n\} Y(s)\lambda_0(s;\eta_0)ds \xrightarrow{pr} 0.$$

7. There exist functions G_1 and H_1 defined on F such that for each $t \in F$

$$\sup_{\eta\in\Gamma_0} \left| \frac{\partial^3}{\partial\eta_i\eta_j\eta_k} \lambda_O(t;\eta) \right| \le G_1(t)$$

$$\sup_{\eta\in\Gamma_0} \left| \frac{\partial^3}{\partial\eta_i\eta_j\eta_k} \log\lambda_O(t;\eta) \right| \le H_1(t)$$

and such that, as $n \to \infty$,

$$a_n^{-2} \int_0^\tau G_1(s)Y(s)ds \xrightarrow{pr} \int_0^\tau G_1(s)y(s)ds < \infty,$$

$$a_n^{-2} \int_0^\tau H_1(s)Y(s)\lambda_O(s;\eta_0)ds \xrightarrow{pr} \int_0^\tau H_1(s)y(s)\lambda_O(s;\eta_0)ds < \infty,$$

$$a_n^{-2} \int_0^\tau H_1^2(s)Y(s)\lambda_O(s;\eta_0)ds \xrightarrow{pr} \int_0^\tau H_1^2(s)y(s)\lambda_O(s;\eta_0)ds < \infty.$$

8. There exists a predictable process $G_2(\cdot)$ with $\sup_{\eta\in\Gamma} || \frac{\partial^2}{\partial\eta_i\partial\eta_j} \psi_i(s;\eta) || \le G_2(s)$ for

each $i = 1, 2, \cdots, k$ and such that, as $n \to \infty$,

$$a_n^{-2} \int_0^\tau \left|\left| \frac{\partial}{\partial \eta} \psi(s; \eta_0) \right|\right|^2 Y(s) \lambda_0(s; \eta) ds =)_p(1)$$

and

$$a_n^{-2} \int_0^\tau G_2(s)^2 Y(s) \lambda_O(s; \eta_0) ds \xrightarrow{pr} < \infty.$$

A.3 Smooth Tests of GOF for Baseline Hazard in Recurrent Events

In order to derive appropriate test procedure we obtain the distribution of $\boldsymbol{U}_\theta(t; \theta)$ under the null hypothesis. Required regularity conditions needed are (Agustin and Peña, 2001);

$$U_\theta(t; \theta) = \sum_n^{j=1} \int_0^t \Psi(s) dM_j(s; \theta)$$

1. $\int_0^\tau \lambda(\cdot)(s) ds < \infty$.

2. There exist a $k \times k$ matrix function D such that as $n \to \infty$,

$$\sup_{t \in F} \left|\left| \frac{1}{n} \Psi(t) \Psi(t)' Y_j(t) - D(t) \right|\right| \xrightarrow{pr} 0.$$

3. The matrix $\Sigma(\tau) = \int_0^t D(t) \lambda_0(t) dt$ is a positive definite.

4. For each $\varepsilon > 0, l = 1, 2, \cdots, k$ and for every $t \in \mathscr{F}$

$$\frac{1}{n} \sum_{j=1}^n \int_0^t \Psi_l(s)^2 I\{| \geq n\varepsilon|\} Y_j(s) ds \xrightarrow{pr} 0.$$

Appendix B

R Syntax

B.1 Analysis of Loss to follow-up uncensored Data

```
R scripts for testing Kolmogorov Sminov, Cram$\acute{\text{e}}$r Von M
data_uncensoredweib<-read.csv("Uncensored_weibdata.csv")
data_uncensoredweib
c.weib<-as.numeric(data_censoredweib$Time)
unc.weib<-as.numeric(data_uncensoredweib$Time)

ks.test(c.weib,"pweibull", scale = 39.076276, shape = 6.770389)
cvm.test(c.weib,"pweibull", scale = 39.076276, shape = 6.770389)
ad.test(c.weib,"pweibull", scale = 39.076276, shape = 6.770389)

ks.test(unc.weib,"pweibull", scale=30.145206, shape= 6.785959)
cvm.test(unc.weib,"pweibull", scale=30.145206, shape= 6.785959)
ad.test(unc.weib,"pweibull", scale=30.145206, shape= 6.785959)
```

B.1.1 Simulations R Codes

Sampled R scripts

```
##########Simulated 2-parameter Weibull random samples#########
sample_n5_1 = rweibull(5, 6, 30)
sample_n5_2 = rweibull(5, 6, 30)
sample_n5_3 = rweibull(5, 6, 30)
sample_n5_4 = rweibull(5, 6, 30)
sample_n5_5 = rweibull(5, 6, 30)
sample_n5_6 = rweibull(5, 6, 30)
sample_n5_7 = rweibull(5, 6, 30)
sample_n5_8 = rweibull(5, 6, 30)
sample_n5_9 = rweibull(5, 6, 30)
sample_n5_10 = rweibull(5, 6, 30)
###############Smooth test for order 3#######################
WPP.test(sample_n5_1,type="ST1")
WPP.test(sample_n5_2,type="ST1")
WPP.test(sample_n5_3,type="ST1")
WPP.test(sample_n5_4,type="ST1")
WPP.test(sample_n5_5,type="ST1")
WPP.test(sample_n5_6,type="ST1")
WPP.test(sample_n5_7,type="ST1")
WPP.test(sample_n5_8,type="ST1")
WPP.test(sample_n5_9,type="ST1")
WPP.test(sample_n5_10,type="ST1")
#############Smooth test for order 4#######################
WPP.test(sample_n5_1,type="ST2")
WPP.test(sample_n5_2,type="ST2")
WPP.test(sample_n5_3,type="ST2")
```

```
WPP.test(sample_n5_4,type="ST2")
WPP.test(sample_n5_5,type="ST2")
WPP.test(sample_n5_6,type="ST2")
WPP.test(sample_n5_7,type="ST2")
WPP.test(sample_n5_8,type="ST2")
WPP.test(sample_n5_9,type="ST2")
WPP.test(sample_n5_10,type="ST2")
##############Kolmogorov-Smirnov test#####################
WEDF.test(sample_n5_1,type="KS",funEstimate="MLE")
WEDF.test(sample_n5_2,type="KS",funEstimate="MLE")
WEDF.test(sample_n5_3,type="KS",funEstimate="MLE")
WEDF.test(sample_n5_4,type="KS",funEstimate="MLE")
WEDF.test(sample_n5_5,type="KS",funEstimate="MLE")
WEDF.test(sample_n5_6,type="KS",funEstimate="MLE")
WEDF.test(sample_n5_7,type="KS",funEstimate="MLE")
WEDF.test(sample_n5_8,type="KS",funEstimate="MLE")
WEDF.test(sample_n5_9,type="KS",funEstimate="MLE")
WEDF.test(sample_n5_10,type="KS",funEstimate="MLE")
##################Anderson-Darling Test#####################
WEDF.test(sample_n5_1,type="AD",funEstimate="MLE")
WEDF.test(sample_n5_2,type="AD",funEstimate="MLE")
WEDF.test(sample_n5_3,type="AD",funEstimate="MLE")
WEDF.test(sample_n5_4,type="AD",funEstimate="MLE")
WEDF.test(sample_n5_5,type="AD",funEstimate="MLE")
WEDF.test(sample_n5_6,type="AD",funEstimate="MLE")
WEDF.test(sample_n5_7,type="AD",funEstimate="MLE")
WEDF.test(sample_n5_8,type="AD",funEstimate="MLE")
```

```
WEDF.test(sample_n5_9,type="AD",funEstimate="MLE")
WEDF.test(sample_n5_10,type="AD",funEstimate="MLE")
#####################Cramer Von Mises Tests#####################
WEDF.test(sample_n5_1,type="CM",funEstimate="MLE")
WEDF.test(sample_n5_2,type="CM",funEstimate="MLE")
WEDF.test(sample_n5_3,type="CM",funEstimate="MLE")
WEDF.test(sample_n5_4,type="CM",funEstimate="MLE")
WEDF.test(sample_n5_5,type="CM",funEstimate="MLE")
WEDF.test(sample_n5_6,type="CM",funEstimate="MLE")
WEDF.test(sample_n5_7,type="CM",funEstimate="MLE")
WEDF.test(sample_n5_8,type="CM",funEstimate="MLE")
WEDF.test(sample_n5_9,type="CM",funEstimate="MLE")
WEDF.test(sample_n5_10,type="CM",funEstimate="MLE")
```

B.2 Analysis for Real Cancer Data

```
#################Lung Cancer###################
data(lung)
fit.cancer1 <- coxph(Surv(time, status) ~ age + sex+ ph.ecog
   +ph.karno+pat.karno+meal.cal+wt.loss, data=cancer1)
temp.canl<- cox.zph(fit.cancer1)
temp.canl
plot(temp.canl, main=" Schoenfeld residuals for overall
   model vs. time")
par(mfrow=c(2,2))
plot(temp.canl[1], main=" age vs. time")
plot(temp.canl[2], main=" sex vs. time")
```

```
plot(temp.canl[3], main=" ph.karno vs. time")
plot(temp.canl[4], main=" pat.karno rx vs. time")
col1= proprate2.ks(Surv(cancer1$time, cancer1$status),
   cancer1$ph.ecog, model = 1)
col1
## Neyman's test of proportional odds
## test with fixed dimension
proprate2.neyman(Surv(cancer1$time, cancer1$status),
   cancer1$ph.ecog, model = 1, data.driven = FALSE)
## data-driven test
print(can1.sm <- proprate2.neyman(Surv(cancer1$time,
   cancer1$status),cancer1$ph.ecog, model = 1, data.driven =
    TRUE))
#####################Colon Cancer###################
data(colon)
colon
fit.col <- coxph(Surv(time, status) ~ age + sex+ rx +
   obstruct + adhere+differ+  extent+surg+node4+etype, data=
   colon)
temp.col<- cox.zph(fit.col)
temp.col
plot(temp.vet[3], main=" Schoenfeld residuals for rx vs.
   time")
par(mfrow=c(2,2))
plot(temp.col[1], main=" Schoenfeld residuals for age vs.
   time")
```

```
plot(temp.col[2], main=" Schoenfeld residuals for sex vs.
   time")
plot(temp.col[10], main=" Schoenfeld residuals for etime vs.
    time")
plot(temp.col[3], main=" Schoenfeld residuals for rx vs.
   time")
col1= proprate2.ks(Surv(colon$time, colon$status),
   colon$etype, model = 1)
col1
## Neyman's test of proportional odds
## test with fixed dimension
proprate2.neyman(Surv(colon$time, colon$status),colon$etype,
    model = 1, data.driven = FALSE)
## data-driven test
print(vet1.sm <- proprate2.neyman(Surv(colon$time,
   colon$status),colon$etype, model = 1, data.driven = TRUE)
   )
# print results
print(vet1.gof<- surv2.ks(Surv(colon$time, colon$status),
   colon$etype))
##################Leukemia###################
data(leukemia)
leukemia
fit.leukemia1 <- coxph(Surv(time, status) ~ x, data=
   leukemia1)
temp.leukemia1<- cox.zph(fit.leukemia1)
temp.leukemia1
```

```
fit.leukemia1 <- coxph(Surv(time, status) ~ x, data=
   leukemia1)
temp.leukemia1<- cox.zph(fit.leukemia1)
temp.leukemia1
plot(temp.leukemia1)
plot(temp.leukemia1)
test.leukemia1= proprate2.ks(Surv(leukemia1$time,
   leukemia1$status),leukemia1$x1, model = 1)
test.leukemia1
## Neyman's test of proportional odds
## test with fixed dimension
proprate2.neyman(Surv(leukemia1$time, leukemia1$status),
   leukemia1$x1, model = 1, data.driven = FALSE)
## data-driven test
print(test.leukemia1<- proprate2.neyman(Surv(leukemia1$time,
    leukemia1$status),leukemia1$x1, model = 1, data.driven =
    TRUE))
# print results
print(test.leukemia1.gof<- surv2.ks(Surv(leukemia1$time,
   leukemia1$status),leukemia1$x1))
##############Melanoma###################
data(melanoma)
mel1=melanoma
fit.mel1 <- coxph(Surv(time, state) ~ sex+thick +ulcer +age
   , data=mel1)
temp.mel1<- cox.zph(fit.mel1)
temp.mel1
```

```
plot(temp.mel1)
par(mfrow=c(2,2))
plot(temp.mel1[1], main=" Schoenfeld residuals for sex vs.
   time")
plot(temp.mel1[2], main=" Schoenfeld residuals for thick vs.
    time")
plot(temp.mel1[3], main=" Schoenfeld residuals for ulcer vs.
    time")
plot(temp.mel1[4], main=" Schoenfeld residuals for age vs.
   time")
test.mel1= proprate2.ks(Surv(mel1$time, mel1$state),
   mel1$ulcer, model = 1)
test.mel1
## Neyman's test of proportional odds
## test with fixed dimension
proprate2.neyman(Surv(mel1$time, mel1$state),mel1$ulcer,
   model = 1, data.driven = FALSE)
## data-driven test
print(pbc1.sm <- proprate2.neyman(Surv(mel1$time, mel1$state
   ),mel1$ulcer, model = 1, data.driven = TRUE))
# print results
print(pbc1.gof<- surv2.ks(Surv(mel1$time, mel1$state),
   mel1$ulcer))
###############Ovarian Cancer###################
data(ovarian)
ovarian
```

```
fit <- coxph(Surv(futime, fustat) ~ age + ecog.ps+ rx +
   resid.ds, data=ovarian)
temp <- cox.zph(fit)
plot(temp)
par(mfrow=c(2,2))
plot(temp[1], main=" Schoenfeld residuals for age vs. futime
   ")
plot(temp[2], main=" Schoenfeld residuals for ecog.ps vs.
   futime")
plot(temp[3], main=" Schoenfeld residuals for rx vs. futime
   ")
plot(temp[4], main=" Schoenfeld residuals for resid.ds vs.
   futime")
temp
proprate2.ks(Surv(ovarian$futime, ovarian$fustat),ovarian$rx
   , model = 1)
ova1
## test with fixed dimension
proprate2.neyman(Surv(ovarian$futime, ovarian$fustat),
   ovarian$rx, model = 1, data.driven = FALSE)
## data-driven test
print(ova1.sm <- proprate2.neyman(Surv(ovarian$futime,
   ovarian$fustat),ovarian$rx, model = 1, data.driven = TRUE
   ))
# print results
print(ova1.gof<- surv2.ks(Surv(ovarian$futime,
   ovarian$fustat),ovarian$rx))
```

```
##################rott2###################
rott2<- read.csv("rott2.csv")
rott2
fit.rott2 <- coxph(Surv(time, status) ~ age + meno+er+ grade
   +nodes+nodes+pr, data=rott2)
temp.rott2<- cox.zph(fit.rott2)
temp.rott2
plot(temp.rott2, main=" Schoenfeld residuals for overall
   model vs. time")
par(mfrow=c(2,2))
plot(temp.rott2[1], main=" age vs. time")
plot(temp.rott2[2], main=" nodesvs. time")
plot(temp.rott2[3], main=" meno vs. time")
plot(temp.rott2[4], main=" grade vs. time")
col1= proprate2.ks(Surv(rott2$time, rott2$status),rott2$meno
   , model = 1)
col1
## Neyman's test of proportional odds
## test with fixed dimension
proprate2.neyman(Surv(rott2$time, rott2$status),rott2$meno,
   model = 1, data.driven = FALSE)
## data-driven test
print(can1.sm <- proprate2.neyman(Surv(rott2$time,
   rott2$status),rott2$meno, model = 1, data.driven = TRUE))
#######################Stage B/C Prostate Cancer
   ############################
stagec<- read.csv("stagec.csv")
```

```
stagec
fit.stagec <- coxph(Surv(time, status) ~ age + eet+ g2+grade
   +gleason+ploidy, data=stagec)
temp.stagec<- cox.zph(fit.stagec)
temp.stagec
plot(temp.stagec, main=" Schoenfeld residuals for overall
   model vs. time")
par(mfrow=c(2,2))
plot(temp.stagec[1], main=" age vs. time")
plot(temp.stagec[2], main=" eet vs. time")
plot(temp.stagec[3], main=" g2 vs. time")
plot(temp.stagec[4], main=" grade vs. time")
stagec1= proprate2.ks(Surv(stagec$time, stagec$status),
   stagec$eet, model = 1)
stagec1
## Neyman's test of proportional odds
## test with fixed dimension
proprate2.neyman(Surv(stagec$time, stagec$status),stagec$eet
   , model = 1, data.driven = FALSE)
## data-driven test
print(stagec1.sm <- proprate2.neyman(Surv(stagec$time,
   stagec$status),stagec$eet, model = 1, data.driven = TRUE)

###############Veteran###################
data(veteran)
vet1=veteran
plot(vet1)
```

```
fit.vet <- coxph(Surv(time, status) ~ age + karno+trt +
   diagtime + prior+celltype, data=veteran)
temp.vet <- cox.zph(fit.vet)
temp.vet
par(mfrow=c(2,2))
plot(temp.vet[1], main=" Schoenfeld residuals for age vs.
   time")
plot(temp.vet[2], main=" Schoenfeld residuals for karno vs.
   time")
plot(temp.vet[3], main=" Schoenfeld residuals for diagtime
   vs. time")
plot(temp.vet[4], main=" Schoenfeld residuals for prior vs.
   time")
vet1= proprate2.ks(Surv(veteran$time, veteran$status),
   veteran$trt, model = 1)
vet1
## Neyman's test of proportional odds
## test with fixed dimension
proprate2.neyman(Surv(veteran$time, veteran$status),
   veteran$trt, model = 1, data.driven = FALSE)
## data-driven test
print(vet1.sm <- proprate2.neyman(Surv(veteran$time,
   veteran$status),veteran$trt, model = 1, data.driven =
   TRUE))
# print results
 print(vet1.gof<- surv2.ks(Surv(veteran$time, veteran$status
    ),veteran$trt))
```

B.3 Analysis of Recurrent Events

B.3.1 Simulations

```
sim.data561<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
sim.data562<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
sim.data563<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
```

```
sim.data564<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
sim.data565<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
sim.data566<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
```

```
sim.data567<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
sim.data568<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
sim.data569<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
```

```
sim.data570<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
sim.data571<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
sim.data572<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
```

```
sim.data573<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
sim.data574<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
sim.data575<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
```

```
sim.data576<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
sim.data577<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
sim.data578<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
```

```
sim.data579<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
sim.data580<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
sim.data581<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
```

```
sim.data582<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
sim.data583<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
sim.data584<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
```

```
sim.data585<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
sim.data586<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
sim.data587<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
```

```
sim.data588<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
sim.data589<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
sim.data590<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
```

```
sim.data591<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
sim.data592<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
sim.data593<- crisk.sim(n=5,foltime=100, dist.ev=c("lnorm","
    lnorm"), anc.ev=c(1.479687, 0.5268302),beta0.ev=c
    (3.80342, 2.535374),dist.cens="lnorm", anc.cens=1.242733,
    beta0.cens=5.421748,z=list(c("unif", 0.8,1.2), c("unif",
    0.9, 1.5)), beta=list(c(0.1698695,0.0007010932),c
    (0.3735146,0.5591244)), x=list(c("bern", 0.381), c("bern
    ", 0.564)), nsit=2)
> simple.surv.sim(n, time, dist.ev, anc.ev, beta0.ev, dist.
    cens = "weibull", anc.cens, beta0.cens, z = NA, beta = NA
    , x = NA)
simDat <- function(n, a, b, tau) {
z <- rgamma(n, 2, 0.2)
X1 <- rbinom(n, 1, 0.5)
```

```
X2 <- runif(n, 0 ,1)
lambda <- z * (1/10) * exp(cbind(X1, X2) %*% a)
mt <- rpois(n, tau * lambda)
eventT <- lapply(sapply(mt, function(x) runif(x) * tau),
   sort)
D <- exp(- cbind(X1, X2) %*% b) * sqrt(800 * (-log(runif(n))
    / z))
C <- (X1 == 1) * rexp(n, 0.1) + (X1 == 0) * rexp(n, ifelse(z
    == 0, 10^-10, z)^2/300)
Y <- pmin(C, tau, D)
Delta <- 1 * (D <= Y)
for (i in 1:n) {
eventT[[i]] <- c(eventT[[i]][eventT[[i]] < Y[i]], Y[i])}
out <- data.frame(id = rep(1:n, unlist(lapply(eventT, length
   ))),
T = unlist(eventT),
X1 = rep(X1, unlist(lapply(eventT, length))),
X2 = rep(X2, unlist(lapply(eventT, length))),
Delta = rep(Delta, unlist(lapply(eventT, length))))
out}
set.seed(123)
mydat <- simDat(30, a = c(-1, 1), b = c(-1, 1), tau = 10)
(fit.sim <- reReg(reSurv(T, id, Delta) ~ X1 + X2, data =
   mydat, B = 0))
plot(fit.sim, se = TRUE, B = 10)
```