MODELING AND ANALYSIS OF COMPETING RISKS DATA

Thesis Submitted to the
Cochin University of Science and Technology
for the Award of Degree of

DOCTOR OF PHILOSOPHY

under the Faculty of Science

by

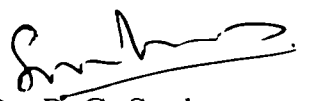
SREEDEVI E. P.

DEPARTMENT OF STATISTICS
COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY
COCHIN-682 022

APRIL 2010

Certificate

Certified that the thesis entitled '**Modeling and Analysis of Competing Risks Data**' is a bonafide record of works done by Smt. Sreedevi. E. P under my guidance in the Department of Statistics, Cochin University of Science and Technology, Cochin-22, Kerala, India and that no part of it has been included anywhere previously for the award of any degree or title.



Cochin-22

09 April 2010

Dr. P. G. Sankaran

(Supervising Guide)

Declaration

The thesis entitled '**Modeling and Analysis of Competing Risks Data**' contains no material which has been accepted for the award of any Degree or Diploma in any University and to the best of my knowledge and belief, it contains no material previously published by any other person, except where due references are made in the text of the thesis.

Sreedevi

Cochin-22 Sreedevi E. P.

09 April 2010

Acknowledgements

I am obliged to my supervising guide Dr. P. G. Sankaran, Reader, Department of Statistics, Cochin University of Science and Technology, for his inspiring guidance, substantial support and encouragement, thanks to which, the thesis reached completion.

I owe a lot to Dr. N. Balakrishna, Professor and Head, Department of Statistics, Cochin University of Science and Technology, for his valuable suggestions and timely advice.

I am also grateful to all the teaching staff of the Department of Statistics, all of whom have whole-heartedly provided encouragement and moral support during my study period. With endless respect and love, I thank all my teachers who gave me light in life through education.

My most heartleaf thanks to Prof. N. Balakrishnan, Professor, McMaster University, Canada who helped me with his beneficial suggestions and guidance during my research period.

I am indebted to Prof. N. Unnikrishnan Nair, Department of Statistics, CUSAT and to Prof. Isha Dewan, ISI, New Delhi for their technical support to complete my research work. I am also thankful to Dr. M. Sivaram, Scientist, Kerala Forest Research Institute, Peechi, Kerala for giving me an exposure to new fields of research work.

It is a pleasure to express my gratitude to Dr. Alan Fredeen, Professor, NSAC, Truro, Canada and Dr. Alex Georgallas, Assistant Professor, NSAC, Truro, Canada for giving me an opportunity to work in their extramural grant thus helping improving my technical skills and providing financial support. I also thank Prof. Greg Bishop, NSAC, Truro, Canada for his technical support.

I place on record my profound gratitude to the non-teaching staff of the Department of Statistics for their sincere co-operation. With immense pleasure, I convey my gratitude to my friends, especially to the research scholars of the Department of Statistics for their help, encouragement and motivation.

I am obliged to the care, support and sacrifice of my family members; especially to the love and devotion of my mother and my husband without whom I may not have been able to fulfill this work. I also thankfully mention the support of my father, brother, in laws and my loving son with fond remembrance.

The financial support extended by Kerala State Council for Science Technology and Environment to start my research work is duly acknowledged.

Above all, for the blessings showered upon me during times of difficulties, I offer my salutations to the Almighty for giving me strength to bring this thesis to fruition.

SREEDEVI E. P

Contents

Chapter One

Preliminaries

1.1 Introduction

Lifetime data analysis encompasses a wide variety of methods for analyzing time to occurrence of some event of interest. The event may be death, appearance of some disease, relapse from remission, equipment breakdown etc. The response variable is then the time until the event of interest occurs, which is variously referred to as lifetime, failure time or survival time. Applications of lifetime distribution methodology range from investigations of the durability of manufactured items to study of human diseases and their treatments. The modern era in lifetime data analysis started some decades back with applications to engineering. Starting from 1970, the field expanded rapidly with respect to theory, methodology and fields of applications. The development proceeded to two inter mingling streams; viz. reliability theory and survival analysis. The reliability theory deals with models and methods for lifetime of components and systems in engineering and industrial fields where as the survival analysis concerns with data arising from medical studies.

The definition of lifetime includes a time scale and a time origin as well as the specification of the event that determines the lifetime. The lifetime is not always the real or chronological time. The prototypical event is death, which accounts for the name given to these methods. The following examples illustrate various types of lifetime data that arise in practical situations.

Example 1.1: A standard experiment in the investigation of carcinogenic substance is one in which laboratory animals are subjected to doses of the substance and then

observed to see if they develop tumors. A main variable of interest is the time to appearance of a tumor, measured from when the dose is administrated.

Example 1.2: Nelson (1972) considered a life test experiment in which specimens of a type of electrical insulating fluid were subjected to constant voltage stress. Then the time until each specimen failed or broke down is termed as the event time.

Example 1.3: Demographers and social scientists are interested in the duration of certain life 'states' for humans. For example, the duration of marriages formed during a particular year is the lifetime variable where a marriage may end due to annulment, divorce or death.

Example 1.4: Gehan (1965) discussed the results of a clinical trial, in which the drug 6-mercaptopurine (6-MP) was compared to a placebo with respect to the ability to maintain remission in acute leukemia patients. Remission time for each patient is the lifetime variable in this study, which is used to compare the effect of two drugs on patients.

1.2 Basic Concepts

Let T be a nonnegative random variable representing the lifetime of an individual with an absolutely continuous distribution function $F(t)$. The basic concepts that are usually employed to specify the probability distribution of T in survival analysis are the survivor function and the hazard rate function.

Let $f(t)$ be the probability density function (p.d.f) of T The probability of an individual surviving up to time t is defined as the survivor function which is given by

$$S(t) = P(T \geq t) = \int_{t}^{\infty} f(u)\,du . \tag{1.1}$$

$S(t)$ is also referred to as the reliability function in the framework involving lifetimes of manufactured items. $S(t)$ is a monotone, non-increasing, right continuous function with $S(0) = 1$ and $S(\infty) = \lim_{t \to \infty} S(t) = 0$.

The hazard rate function $\lambda(t)$ is defined as

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \,|\, T \geq t)}{\Delta t}$$

If the density $f(t)$ exists, $\lambda(t)$ can be written as

$$\lambda(t) = \frac{f(t)}{S(t)} \tag{1.2}$$

The hazard rate function specifies the instantaneous rate of death or failure at time t, given that the individual has survived up to time t $\lambda(t)$ is referred to as the failure rate function and force of mortality in different contexts.

Any of the three functions discussed above fully specifies the distribution of the lifetime variable T From equation (1.1), we have $S(t) = 1 - F(t)$. Now, from (1.2), we obtain

$$\lambda(t) = -\frac{d}{dt} \log S(t)$$

which provides

$$S(t)=\exp\left(-\int_0^t \lambda(u)\,du\right). \tag{1.3}$$

By defining the cumulative hazard rate function $\Lambda(t)$ as

$$\Lambda(t)=\int_0^t \lambda(u)\,du \tag{1.4}$$

we have

$$S(t)=\exp\left(-\Lambda(t)\right). \tag{1.5}$$

Now from (1.2) and (1.5), it follows that the probability density function can be written in terms of the hazard rate function as

$$f(t)=\lambda(t)\exp\left(-\Lambda(t)\right) \tag{1.6}$$

Another concept that is used for modeling survival data is the mean residual life function, which is also known as the expected residual life. The mean residual life of time t is given by

$$m(t)=E\left(T-t\,|\,T\geq t\right). \tag{1.7}$$

$m(t)$ represents the average remaining lifetime for an individual at time t, given that the individual has survived up to time t If $m(t)$ exists, then by the definition (1.7), it follows that

$$m(t)=\frac{\int_t^\infty S(u)\,du}{S(t)} \tag{1.8}$$

which yields $m(0) = E(T) = \int_0^\infty S(u)\,du$. From (1.8), the survivor function can be written in terms of the mean residual life function as

$$S(t) = \frac{m(0)}{m(t)} \exp\left(-\int_0^t \left(m(u)\right)^{-1} du \right). \tag{1.9}$$

1.3 Censoring

Lifetime data often includes different patterns of missing or incomplete observations. The situation in which the exact lifetime values are unavailable for certain individuals is known as censoring. Practical constraints such as time and cost in continuing the study may result in censoring.

Various types of censoring schemes deal with different moulds of incomplete observations.

1.3.1 Right Censoring

In right censoring, only the lower bounds of lifetimes are available for some individuals. This may be happened when a decision is made to terminate a life test before all items have failed, or when a person in a prospective study is 'lost to follow-up' because they move away from the region where the study take place (Lawless, 2003). Right censoring is the most common censoring scheme observed in survival and reliability studies.

Example 1.5: Prentice (1973) discussed the data on 40 lung cancer patients taken from a study designed to compare the effect of two chemotherapy treatments in prolonging survival time. All patients in the study were received prior therapy and then randomly assigned to one of the two treatments termed 'standard' and 'test' Survival times are measured in days from the start of the treatment for each patient.

Observations correspond to patients who were still alive at the time of data collection are considered to be right censored.

Type I and Type II censoring schemes are two different forms of right censoring.

1.3.1.1 Type I Censoring

In type I censoring scheme, each individual has a fixed potential censoring time $R>0$ such that T is observed if $T \le R$; otherwise we know only that $T > R$ Type I censoring often happens, when a study is conducted over a specified time period. Suppose that there are n individuals under study. Then under type I censoring, we have $t_i = \min(T_i, R_i)$ and the indicator variable $\delta_i = I(T_i \le R_i)$ for $i = 1, 2, ..., n$. Due to the simplicity in the censoring pattern, type I is the most commonly employed censoring scheme in practical studies.

1.3.1.2 Type II Censoring

Consider a situation where only the r smallest lifetimes $t_{(1)} \le .... \le t_{(r)}$ in a random sample of n individuals are observed, where r is a number between 1 and n This situation of incomplete data is referred to as type II censoring. In type II censoring scheme, the value of r is chosen before the data are collected and the data consist of the r smallest lifetimes in a random sample of $T_1, T_2, ..., T_n$. A significant advantage of type II censoring is that, we know the number of observed lifetimes in advance, which helps enormously when planning adequate tests. Type II censoring is often discussed in theoretical works than in practical studies.

1.3.1.3 Progressive Type II Censoring

Progressive type II censoring is a generalisation of type II censoring. In this set up, the first r_1 failures in a life test of n items are observed; then n_1 of the remaining

$n-r_1$ unfailed items are removed from the experiment, leaving $n-r_1-n_1$ items still present in the study. When further r_2 items have failed, n_2 of the still unfailed items are removed, and so on. The experiment terminates after some prearranged series of repetitions of this procedure.

1.3.1.4 Independent Random Censoring

A very simple random censoring process that is often realistic is the one in which each individual is assumed to have a lifetime T and a censoring time Z, where T and Z are independent, continuous random variables with survivor functions $S(t)$ and $G(t)$ respectively. This implies that, at any time t, the survival experience in the future is not statistically altered by survival and censoring experience in the past. If $G(t)$ does not depend on any of the parameters of $S(t)$, then we call it as non informative censoring process. Then, the observed variable will be (Y,δ), where $Y=\min(T,Z)$ and the censoring indicator δ is defined such that $\delta=1$ if $T\le Z$ and 0 if $T>Z$. The data on n individuals consist of the pairs (Y_i,δ_i), $i=1,2,...,n$.

1.3.2 Left Censoring

A lifetime T associated with an individual is considered to be left censored if it is less than a censoring time L. This means that the event of interest has already occurred before the individual entered into the study at time L and that exact lifetime is unknown.

Example 1.6: Baboons in the Amboseli Reserve, Kenya, sleep in the trees and descend for foraging at some time of the day. Observers often arrive later in the day than this descent and for such days they can only ascertain that descent took place

before a particular time, so that the descent times are left censored (see Andersen et al., 1993).

1.3.3 Double Censoring

When individuals are experiencing both left censoring and right censoring, the situation is considered to be doubly censored. Doubly censored data may also occur, when T represents an outcome variable that can only be accurately measured within a certain range $[U,V]$. The observed data for n individuals can be represented by $\{\min\{V_i, \max(T_i,U_i)\}, I(T_i < U_i), I(T_i > V_i)\}, i = 1,2,...,n$.

Example 1.7: Consider the AIDS clinical trial discussed by Cai and Cheng (2004). The study was designed to compare the virological responses to three treatments for HIV infected children. One major end point of the study was the plasma HIV-1 RNA level obtained by the NucliSens assay. As a result of the limit of the quantification of assay, the observed RNS copies per millilitre of plasma are highly unreliable if it is below 400 or above 750000. The lifetimes for the individuals with plasma level within this window are only exactly observed and if the plasma level is below 400, the lifetimes are left censored and if it is above 750000, the lifetimes are right censored.

1.4 Truncation

A possibility for incomplete data in survival or reliability studies arises, due to the limited time span of the study or dropouts of the subjects for various reasons. Such situations are described by the phenomenon, truncation. In many life testing situations, individuals cannot be randomly selected and followed prospectively from the time origin $t = 0$, but some value $u > 0$. If the selection of i th individual at time u_i requires that $T_i \geq u_i$ and the observed data for individual i consist of (u_i, t_i, δ_i)

where $t_i \geq u_i$ is a lifetime or censoring time, we say that the lifetime T_i is left truncated at u_i. In many occasions, at least some of the data arise chronologically before the time by which individuals are selected for the study. Then, the condition for i th individual to be included in the data set is $T_i \leq v_i$. This is referred to as right truncation of the lifetime T_i. Truncated samples of this type arise in reliability and epidemiology studies (Kalbfleisch and Lawless, 1988). For various kinds of truncation, one could refer to Lawless (2003).

Example 1.8: Kalbfleisch and Lawless (1989) analyzed data on patients infected with HIV via blood transfusion, who were subsequently diagnosed with AIDS. The data were used to estimate the distribution of the time T between HIV infection and AIDS diagnosis. The study group was assembled in 1987 and consisted of individuals who had a diagnosis of AIDS prior to July 1, 1986. For each patient the date of HIV infection could also be ascertained, because the individuals selected were deemed to have contracted the HIV through a blood transfusion on a particular date. The condition for i th individual to be included in the data set was therefore that, $T_i \leq v_i$, where v_i is the time between the individual's HIV infection and July 1, 1986. This is an example of right truncated data.

1.5 Inference Procedures

In survival studies, various techniques are employed for modeling and analysis of lifetime data. Three different approaches to model survival data are parametric methods, semiparametric methods and nonparametric methods. In conventional parametric methods, we assume that the random variable T follows some probability distribution $f(t;\theta)$, where the functional form of $f(t;\theta)$ is known, but the parameter θ is unknown. Continuous probability models such as exponential, Weibull,

lognormal, log logistic, Pareto and inverse Gaussian are commonly used in the analysis of lifetime data. The estimation of parameters of the model is done using different procedures such as maximum likelihood estimation, method of moments, Bayesian techniques etc. For more details on parametric lifetime models and on their inference procedures, one may refer to Martz and Waller (1982), Sinha (1986) and Lawless (2003).

In many practical situations, the functional form of $f(t)$ is unknown. Semiparametric and nonparametric methods, which allow a distribution free approach, are useful in such situations. Semiparametric methods do not make assumptions on the distribution of lifetimes, but make assumption on how the covariates influence the lifetimes. However, nonparametric methods allows a completely assumption free approach for the modeling of survival data. If there are no censored observations in a sample of size n, the nonparametric estimate of $S(t)$ is the empirical survivor function, which is given by

$$\hat{S}_{ESF}(t) = \frac{Number\ of\ observations \geq t}{n}$$

When censored observations are present, the number of lifetimes greater than or equal to t will not be known exactly. Thus, $\hat{S}_{ESF}(t)$ do not give a correct estimate of the underlying survival distribution. Accordingly, Kaplan and Meier (1958) defined a product-limit estimator for $\hat{S}(t)$.

1.5.1 Kaplan-Meier Estimator (Product-Limit Estimator)

Let (t_i^*, δ_i) be a random sample of n lifetimes which may include censored observations, where t_i^* is the observed lifetime or censoring time and δ_i is the

indicator function which equals to one if t_i^* is a lifetime and 0 otherwise. Suppose that there are k distinct lifetimes $t_1 < t_2 < .... < t_k$ where the possibility of more than one death at t_j is allowed. Let $d_j = \sum I\left(t_i^* = t_j, \delta_i = 1\right)$ represents the number of deaths at t_j where $I(.)$ is the indicator function. Then the product-limit estimator of $S(t)$ is defined as

$$\hat{S}(t) = \prod_{j:t_j \le t} \frac{n_j - d_j}{n_j} \tag{1.10}$$

where $n_j = \sum I\left(t_i^* \ge t_j\right)$ is the number of individuals at risk at time t_j. The product-limit estimator does not change at censoring time points.

Alternatively, the survivor function $S(t)$ can be estimated nonparametrically using the estimator of the cumulative hazard rate function.

1.5.2 Nelson-Aalen Estimator

The estimator of the cumulative hazard rate function (1.4) is given by

$$\hat{\Lambda}(t) = \sum_{j:t_j \le t} \frac{d_j}{n_j} \tag{1.11}$$

which is referred as the Nelson-Aalen estimator where d_j and n_j are defined as in Section 1.5.1.

Now using (1.5), $S(t)$ can be estimated by

$$\tilde{S}(t) = \exp\left(-\hat{\Lambda}(t)\right). \tag{1.12}$$

Both Kaplan-Meier and Nelson-Aalen estimators possess desirable large sample properties like strong consistency and asymptotic normality. It is important to note that both $\hat{S}(t)$ and $\hat{\Lambda}(t)$ are nonparametric maximum likelihood estimators of $S(t)$ and $\Lambda(t)$ respectively. For more properties of theses estimators, one may refer to Lawless (2003).

1.6 Regression Models

The use of explanatory variables or covariates is a significant method of describing the heterogeneity in a population under consideration. In most survival studies, the focus of interest is to establish the relationship between lifetime variable and covariates. In a lung cancer study, factors such as age, general conditions of the patient and the type of tumor can be considered as covariates. In reliability context, the voltage level during the break down time of equipment is an example of covariate. Regression models are employed to understand and exploit the relationship between the lifetime variable and the covariates. In some practical situations, the effect of covariates on lifetime variable may change over time and such covariates are referred to as time-dependent or time-varying covariates.

Parametric and semiparametric regression models are often employed to analyse lifetime data with covariates. Parametric regression models have been studied by Feigl and Zelen (1965), Zippin and Armitage (1966), Glasser (1967) and Prentice (1973) among many others. In semiparametric regression models, even though we do not make any assumption about the underlying form of the lifetime distribution, some postulations about how the covariates affect the lifetime variable are necessary. The proportional hazards model introduced by Cox (1972) is the commonly employed semiparametric regression model in survival analysis. For recent advances on semiparametric models, one could refer to Chen and Cheng (2005) and Gimenez et al. (2006).

1.6.1 Cox Proportional Hazards (PH) Model

The proportional hazards model assumes that the covariates have a multiplicative effect on the hazard rate function of individuals. In proportional hazards model, the hazard rate function of time T given the $p\times 1$ vector of covariates X takes the form

$$\lambda(t|X)=\lambda_0(t)r(X) \tag{1.13}$$

where $\lambda_0(t)$ is the baseline hazard rate function which is common to all individuals and $r(X)$ is a positive valued real function. $\lambda_0(t)$ is the hazard rate function for an individual with covariate vector X such that, $r(X)=1$. The proportional hazards model possesses the property that, any two individuals have hazard rate functions that are constant multiples of each other. A specification of the proportional hazards model proposed by Cox (1972) with $r(X)=\exp(\beta' X)$ is widely used in literature and known as Cox's proportional hazards (PH) model, where β denote the $p\times 1$ vector of regression parameters.

Under Cox proportional hazards model, the hazard rate function of time t, in presence of the covariate vector X can be specified by

$$\lambda(t|X)=\lambda_0(t)\exp(\beta' X) \tag{1.14}$$

where $\lambda_0(t)$ is an arbitrary unspecified baseline hazard rate function which is common to all individuals. Cox's proportional hazards model is also referred to as relative risk model when the covariates are time-dependent. The model given by (1.14) assumes that covariates have multiplicative effect on the lifetime variable. If

the multiplicative assumption holds well, Cox's model does not depend on the true form of the baseline hazard rate function, so the model can be considered as distribution free.

When the model (1.14) is true, the survivor function for T given X will be

$$S(t|X) = S_0(t)^{\exp(\beta X)} \tag{1.15}$$

Now, the primary objective is to estimate the regression parameters and the cumulative baseline hazard rate function. To estimate the vector β, Cox (1972) proposed a partial likelihood method.

Suppose that the lifetime variable T is randomly right censored by the censoring variable Z Now, we observe (t, δ, X) where $t = \min(T, Z)$, $\delta = I(T = t)$ and X is the corresponding covariate vector and $I(.)$ is the usual indicator function. Let (t_i, δ_i, X_i), $i = 1, 2, ..., n$ be independent and identically distributed copies of (t, δ, X). Define $Y_i(t) = I(t_i \geq t)$ for $i = 1, 2, ..., n$. The partial likelihood proposed by Cox (1972) for estimating β is given by

$$L(\beta) = \prod_{i=1}^{n} \left(\frac{\exp(\beta' X_i)}{\sum_{l=1}^{n} Y_l(t_i) \exp(\beta X_l)} \right)^{\delta_i} \tag{1.16}$$

Maximum likelihood estimates of β can be obtained by maximising the partial likelihood $L(\beta)$ given in (1.16). The nonparametric estimate of the cumulative baseline hazard rate function is then given by

$$\hat{\Lambda}_0(t) = \sum_{i:t_i \le t} \left(\frac{\delta_i}{\sum_{l=1}^{n} Y_l(t_i) \exp\left(\hat{\beta}' X_l\right)} \right). \tag{1.17}$$

Since $S_0(t) = \exp\left(-\Lambda_0(t)\right), S_0(t)$ can be estimated as

$$\hat{S}_0(t) = \exp\left(-\hat{\Lambda}_0(t)\right) \tag{1.18}$$

which leads to the estimator of $S(t|X)$ given by

$$\hat{S}(t|X) = \exp\left(-\hat{\Lambda}(t|X)\right) = \hat{S}_0(t)^{\exp\left(\hat{\beta}' X\right)} \tag{1.19}$$

1.7 Competing Risks Models

In the analysis of medical data or industrial data, the failure of individuals or items may be attributed to more than one cause or factor. These causes (factors) in some sense compete each other for the failure of the experimental unit. The term competing risks refers to such situations in which an individual, either a living organism or an inanimate object is exposed to two or more causes of death (failure), but its eventual death (failure) can be attributed to exactly one of these causes of failure. Accordingly, models which are used to analyse such data are referred to as competing risks models.

The competing risks models arise in public health, demography, actuarial science, industrial reliability applications and experiments in medical therapeutics. The theory of competing risks dates back to 1760, when Daniel Bernoulli studied the effect of small pox eradication on the mortality structure of the overall population.

The following examples demonstrate some situations in which competing risks data arise.

Example 1.9: Hoel (1972) discussed the data of survival times for two groups of laboratory mice, all of which were exposed to a fixed dose of radiation at an age of 5 to 6 weeks. The first group of mice was lived in a conventional lab environment and the second group was kept in a germ-free environment. After autopsy, the cause of death for each mouse was assigned to be one of the three causes of failure: thymic lymphoma, reticulum cell sarcoma or other causes. Another example of competing risks data in survival analysis is from a study of breast cancer patients where the cause of death was recorded as 'cancer' or 'other' (Boag, 1949).

Example 1.10: There are numerous examples in industrial experiments, where items may fail due to one of the several causes. Hinds (1996) observed the data on failure of engines fitted to heavy vehicles which is an example of competing risks problem in industry. Five causes of failures were identified viz. 1- the cooling system, 2-dirt contamination, 3- mechanical failure, 4-ignition fault and 5- fuel fault. For each unit, the miles travelled before failure and the cause of failure were reported.

Example 1.11: In economics, Flinn and Heckman (1983) applied a competing risks model for modeling the unemployment time, where T is the waiting time till the end of unemployment and C indexes the reason for leaving unemployment.

Consider a competing risks situation with k causes (modes) of failure, so that the cause of failure C takes values on the set $\{1,2,...,k\}$ Now a pair (T,C) is defined for each individual where $T>0$ is the continuous lifetime variable and C is the cause (mode) of failure for that individual. In the traditional analysis of competing risks data, the failures due to all other causes except the cause of interest are combined and treated as censored data under the assumption that the causes of failure are

independent each other. Another approach which is useful in the study of competing risks data is the latent failure time approach. In a latent failure time model, we assume that with j th cause of failure, there exists a nonnegative random variable T_j which represents the observed lifetime if all causes except the j th one are inoperative. T_j is referred to as the latent or conceptual lifetime of an individual. In the simultaneous presence of all k causes, only the smallest of such nonnegative random variables $T = \min(T_j)$ is observable, together with the actual cause of failure. Now the data consist of the time and type of first failure. Once the interest is focused upon a particular failure type, failure times due to other causes can be viewed as random right censored observations. Crowder (2001) and Kalbfleisch and Prentice (2002) gave exclusive reviews on this topic.

A finite mixture model can also be used for the analysis of competing risks data (Crowder, 2001). The observed failure time data may be partitioned into separate sets for each failure mode and a lifetime distribution can be fitted to each mode separately. Then,

$$f(t) = p_1 f_1(t) + p_2 f_2(t) + ... + p_k f_k(t)$$

where $f(t)$ and $f_j(t)$ are the probability density functions corresponding to T and T_j respectively with p_j be the probability of an individual to fail by cause j ($j = 1, 2, .., k$). For the analysis of lifetime data using finite mixture models, one could refer to McLachlan and Peel (2000).

Analysis of competing risks data using parametric models has been carried out by Sampford (1952), David (1957), Cox (1959), Berkson and Elveback (1960), Boardman and Kendell (1970), Herman and Patel (1971), Moeschberger and David

(1971) and Hoel (1972). However, such parametric assumptions may be unrealistic in some situations, especially in the context of medical studies. Even when a certain parametric form can be assumed, there is no guarantee that the joint survivor function of $(T_1, T_2, ..., T_k)$ is identifiable. Accordingly, although the concept of latent failure times provides a theoretical basis for discussion, it is not to be recommended except in special types of applications where the unobserved potential failure times can be given a clear meaning.

Recently, various models have been developed to assess the lifetimes of a specific risk in presence of the other competing risk factors. The causes of failure may be assumed to be dependent or independent. Even though the assumption of dependence among the causes may be more realistic, there is some concern about the identifiability of the underlying model. For the analysis of survival data with dependent competing risks, one could refer to Aras and Deshpande (1992), Moeschberger (1974), Moeschberger and Klein (1995) and Matsuyama and Yamaguchi (2008). Markov process and counting process models can be employed for the analysis of competing risks data. Aalen (1976), Fleming and Harrington (1991) and Andersen et al. (1993) provided comprehensive reviews on this area. For an exhaustive treatment of different competing risks models, we can refer to David and Moeschberger (1978) and Crowder (2001).

Two important concepts that are used to specify the distribution of the observable random pair (T, C) in competing risks set up are cause specific hazard rate functions $(\lambda_j(t))$ and cause specific subdistribution functions (cumulative incidence functions).

1.7.1 Cause Specific Hazard Rate Function

Suppose that an individual is subjected to k causes of failure and for each individual we observe the time to failure and cause of failure. Then cause specific hazard rate function $\lambda_j(t)$ is defined as

$$\lambda_j(t) = \lim_{\Delta t \to 0} \frac{P(T < t + \Delta t, C = j | T \geq t)}{\Delta t} \qquad j = 1, 2, ..., k \qquad (1.20)$$

It can be noted that, $\lambda_j(t)$ is the instantaneous rate of failure by cause j in the presence of all other failure types, given that the individual has survived up to time t Cause specific hazard rate functions often have intuitive and scientific appeal. $\lambda_j(t)$ was termed 'decremental forces' by the English actuary Makeham (1874) and 'cause specific hazard rate function' by Prentice et al. (1978). It is identical to the 'force of transition' function in Aalen's Markov formulation. Cause specific hazard rate function is also identical with the 'forces of mortality' commonly employed in demography.

We assume that the k failure types are mutually exclusive and exhaustive so that an individual can have at most one realized lifetime. Assuming the existence of the quantities $\lambda_j(t)$, the overall hazard rate function $\lambda(t)$ is given by

$$\lambda(t) = \sum_{j=1}^{k} \lambda_j(t). \qquad (1.21)$$

Prentice et al. (1978) established that only the probabilities that can be expressed as functions of $\{\lambda_j(t)\}$ can be estimated from the observed data (T, C).

Using (1.3) and (1.21), we can represent the survivor function in terms of cause specific hazard rate functions as

$$S(t)=\exp\left(-\int_0^t\sum_{j=1}^{k}\lambda_j(u)\,du\right). \tag{1.22}$$

1.7.2 Cause Specific Subdistribution Function (Cumulative Incidence Function)

In a competing risks set up with k causes of failure, cause specific subdistribution function, $F_j(t)$ of the observable random pair (T,C) is defined as

$$F_j(t)=P(T\le t,C=j)\qquad j=1,2,\ldots,k \tag{1.23}$$

Each $F_j(t)$ is a subdistribution function in the sense that $F_j(\infty)\le 1$ The significance of the cause specific subdistribution functions in modeling competing risks data is well recognised in demography, epidemiology and survival analysis. Comparison of cause specific subdistribution functions is useful in selecting appropriate treatment for a patient (Gray, 1988). Cause specific subdistribution functions are directly estimable from the observed data (T,C) without making any untestable assumptions and avoid the identifiability problem in competing risks set up (Prentice et al., 1978). Pepe and Mori (1993) pointed that, even though, the cause specific hazard rate functions are estimable from the data (T,C), they do not directly indicate the magnitude of the proportion of patients suffering each of the cause specific end points. For more examples involving the use of cause specific subdistribution functions in survival analysis, we may refer to Lin (1997), Cheng et al. (1998), Gooley et al. (1999), Cronin and Feuer (2000) and Farley et al. (2001).

Either set of functions described above fully specifies the joint distribution of T and C, but they lead to different types of regression models when covariates are

present. Hougaard (2000), Crowder (2001), Kalbfleisch and Prentice (2002) and Lawless (2003) provided reviews on this area.

The problem of identifiability in modeling the competing risks data in terms of latent failure times is well known. In competing risks frame work, the identifiability problem arise because for a particular individual, we could only observe the random vector (T,C), the occurrence of j th type of failure effectively censoring the remaining latent failure times due to other causes. Tsiatis (1975) has proved that, given any joint survivor function with arbitrary dependence between component variates, there exists a different joint survivor function in which the variates are independent. It is also shown that the two survivor functions reproduces the same cause specific subdistribution function $F_j(t)$. Thus from the observations of (T,C) alone, one cannot make sure which of the two models is correct since the both will fit the data equally well. Heckman and Honore (1989) and Bedford and Lindqvist (2004) studied the identifiability problem in detail.

Censoring can occur both in reliability life testing and medical follow up studies under competing risks set up as well. Many researchers have studied the problem of nonparametric estimation of survivor function, cause specific hazard rate functions and cause specific subdistribution functions of competing risks models. To overcome the identifiability problem discussed above, we suppose that each individual may subject to censoring and along with the lifetime variable T, the cause of failure (death) for each individual is also observed. Now the observed data of n individuals consist of $(t_i,\delta_i,\delta_i C_i), i=1,2,...,n$, where t_i is the observed lifetime, δ_i is the censoring indicator and C_i is the cause of failure for i th individual. In such situations, the nonparametric estimator of survivor function proposed by Kaplan and Meier (1958) can be directly generalized to include competing risks. However,

Prentice et al. (1978) emphasized that, only those quantities which can be expressed in terms of cause specific hazard rate functions can be estimated from the competing risks data even if the risks are dependent. So, it is more common to use the Nelson-Aalen estimator of the cumulative hazard rate function in modeling competing risks data (see Lawless, 2003). Let $\delta_{ij} = I(C_i = j)$, and n_i denote the number of individuals at risk just prior to time $t_i, i = 1,2,...,n$, $j = 1,2,...,k$ Then the estimator of the cumulative hazard rate function for j th cause of failure can be obtained as

$$\hat{\Lambda}_j(t) = \sum_{i:t_i \le t} \frac{\delta_{ij}}{n_i} \qquad j = 1,2,...,k$$

which directly gives the estimator of survivor function using (1.22).

1.7.3 Regression Models in Competing Risks

Consider a competing risks situation with k possible causes of failure. Let $T > 0$ be the lifetime variable and $C \in \{1,2,...,k\}$ be the cause of failure. Assume that a $p \times 1$ vector of covariates X is observed for each individual. The Cox proportional hazards model given by (1.14) can be extended into the competing risks set up as

$$\lambda_j(t|X) = \lambda_{0j}(t)\exp(\beta_j X) \qquad j = 1,2,...,k \qquad (1.24)$$

where β_j is the $p \times 1$ vector of regression parameters for cause j and $\lambda_{0j}(t)$ is the baseline cause specific hazard rate function common to all individuals and $\lambda_j(t|X)$ is the cause specific hazard rate function at time t in presence of the covariate vector X

Regression parameters and cumulative baseline cause specific hazard rate functions can be estimated using the partial likelihood approach as in non competing

risks situation described as in Section 1.6.1. Let $(t_i, \delta_i, \delta_i C_i, X_i)$, $i = 1, 2, ..., n$ be the observed data where δ_i is the indicator variable defined as $\delta_i = I(T_i = t_i)$. The partial likelihood for estimating β_j, $j = 1, 2, ..., k$ is given by

$$L(\beta_1, ..., \beta_k) = \prod_{i=1}^{n} \prod_{j=1}^{k} \left(\frac{\exp(\beta_j' X_i)}{\sum_{l=1}^{n} Y_l(t_i) \exp(\beta_j' X_l)} \right)^{\delta_{ij}} \tag{1.25}$$

where $Y_i(t) = I(t_i \geq t)$ and $\delta_{ij} = I(C_i = j)$ for $i = 1, 2, ..., n$; $j = 1, 2, ..., k$ Then (1.25) is a product of k similar terms, of which j th one is given by

$$L(\beta_j) = \prod_{i=1}^{n} \left(\frac{\exp(\beta_j' X_i)}{\sum_{l=1}^{n} Y_l(t_i) \exp(\beta_j' X_l)} \right)^{\delta_{ij}} \qquad j = 1, 2, ..., k \tag{1.26}$$

Regression parameters can be estimated by maximizing the likelihood function given in (1.25). The generalised Nelson-Aalen estimator of cumulative baseline cause specific hazard rate functions is given by

$$\hat{\Lambda}_{0j}(t) = \sum_{i: t_i \leq t} \left(\frac{\delta_{ij}}{\sum_{l=1}^{n} Y_l(t_i) \exp(\hat{\beta}_j' X_l)} \right) \qquad j = 1, 2, ..., k \tag{1.27}$$

Consequently, the survivor function for T given X can be estimated as

$$\hat{S}(t|X) = \exp\left(-\sum_{j=1}^{k} \hat{\Lambda}_{0j}(t) \exp(\hat{\beta}_j' X) \right). \tag{1.28}$$

Thus, (1.28) leads to the estimator of cause specific subdistribution functions as

$$\hat{F}_j(t|X) = \sum_{i:t_i \le t} \delta_{ij} \hat{S}(t|X) \frac{\exp(\hat{\beta}_j' X)}{\sum_{l=1}^{n} Y_l(t_i) \exp(\hat{\beta}_j' X_l)} \quad j = 1, 2, \ldots, k \qquad (1.29)$$

For various types of regression models employed for the analysis of competing risks data, one could refer to Kalbfleisch and Prentice (2002) and Lawless (2003).

1.8 Testing

In life testing and reliability experiments, situations to make decisions about some assumptions under consideration are quite often. Statistical hypothesis tests are employed to validate a postulate under consideration and make appropriate decisions.

One of the major testing problems in classical survival analysis is to compare the underlying survivor distributions of two or more populations. Log rank test, a generalisation of the rank test proposed by Savage (1956) to include censoring times is widely used to compare two survivor distributions. Log rank test is also referred to as Mantel-Cox test in literature.

Two sample tests for comparing survivor distributions were also considered by Mantel (1966), Cox (1972), Peto and Peto (1972). Two sample tests with prominent stratification feature was studied by Peto et al. (1976, 1977). Collet (2003) discussed about the sample size requirement for two sample log rank tests. Linear rank tests for the equality of m - survivor distributions with uncensored data were considered by Hajek and Sidak (1967) and Hettmansperger (1984). Johnson and Mehrotra (1972), Struthers (1984) and Cuzick (1985) considered m-sample tests with censored data among many others. Further, the counting process formulation of the rank tests was studied by Aalen (1978) and Gill (1980).

In competing risks situations, various problems of statistical testing arise with data sets. Two major research problems are to test equality of cause specific hazard rate functions or cause specific subdistribution functions and to test independence of time to failure and cause of failure. When the risks are independent, different authors have proposed nonparametric tests for testing the equality of two or more cause specific hazard rate functions against ordered alternatives. Froda (1987) proposed locally most powerful rank tests for testing the equality of two risks against scaled alternatives. Bagai et al. (1989a, 1989b) developed distribution free rank tests for testing the equality of two cause specific hazard rate functions against stochastic ordering alternatives. Neuhaus (1991) has proposed asymptotically optimal rank tests for comparing several independent competing risks differing in their location or scale parameters. Yip and Lam (1992, 1993) suggested a class of weighted log rank type statistics for testing the equality of cause specific hazard rate functions. By generalising the approach of Harrington and Fleming (1982), Gray (1988) proposed a class of k-sample tests for comparing the cumulative incidence functions. Recently, a procedure based on U-statistic for testing the equality of cause specific hazard rate functions was proposed by Molinari (2005). The situation with two dependent risks has been considered by Deshpande (1990), Aras and Deshpande (1992), Aly et al. (1994) and Sun and Tiwari (1995). One can refer to Sun (2001), Kulathinal and Gasbarra (2002) and Tiwari et al. (2006) for different approaches to test the equality of cause specific hazard rate functions. In the case of discrete (grouped) data, Dykstra et al. (1995) studied the likelihood ratio test for testing the equality of cause specific hazard rate functions against ordered alternatives.

Testing the independence of time to failure and cause of failure is another significant problem in the analysis of competing risks data. Kochar and Proschan (1991) considered the problem of testing the independence of time to failure and cause of failure in the multiple dependent competing risks model. Recently, Dykstra et

al. (1998), Dewan et al. (2004), Gasbarra et al. (2006) and Dewan and Kulathinal (2007) addressed the same testing problem in different situations.

Crowder (1997) proposed a test for independence of competing risks when failure times are discrete. Other major objectives of hypothesis testing in competing risks analysis are to test proportionality of the cause specific hazard rate functions and to test stochastic dominance of a particular type of failure over other causes of failure. Sen (1979) proposed nonparametric tests with maximum asymptotic relative efficiency for testing the interchangeability of two competing risks. Deshpande and Senguptha (1995) considered the problem of testing proportionality assumption of the cause specific hazard rate functions in competing risks set up.

For more approaches to the testing problems in competing risks set up, one could refer to Stanish et al. (1978), Bagai and Prakasha Rao (1992), Kochar (1995), Wada and Sen (1995), Sun and Tiwari (1997), Hu and Tsai (1999), Dewan (2005), Friedlin and Korn (2005), Williamson et al. (2006), Chen et al. (2008) and Dignam and Kocherginsky (2008).

1.9 Neural Networks

Neural networks are persuasive data modeling tools that are able to capture and represent input/output relationships. The network consists of an interconnected group of artificial neurons. Information given as input is processed through the neurons to produce some desired output. The advantages of neural networks lie in their ability to learn and represent both linear and nonlinear variable relationships directly from the data being modeled. Further, neural network models need not make any assumptions about the distribution of the data or about the relationship of covariates with the lifetime. Neural network models use an algorithmic approach which helps to solve complexities beyond the reach of empirical statistical methods.

Neural networks have been successfully applied to a broad spectrum of data intensive applications. Medical diagnosis, signal processing, pattern recognition, credit assignment, stock market prediction and speech recognition are some of the real life problems where neural network models can be applied successfully.

Multilayer Perceptron (MLP) is the most commonly used form of neural network model for prediction and classification problems. It is proved that with appropriate number of hidden units, a multilayer perceptron can approximate any functional relationship (Ripley, 1996). A multilayer perceptron network consists of one input layer of units, one output layer of units together with one or more layers of hidden units. The input signal propagates through the network in a forward direction, on a layer by layer basis. All computed values from the previous units are combined into a single value using a combination function before feeding into the next unit. Linear combination functions are commonly employed for this purpose, where each of the hidden layer units and output units take a weighted sum of its inputs and adds a constant. The next step is to calculate a fixed function of the result. The fixed function is referred to as the activation function. A commonly used form of activation function is logistic activation function given by

$$a(x) = \frac{1}{1+\exp(-x)}.$$

Threshold and identity activation functions are also common in a multilayer perceptron network. The results are then passed to the hidden units in the next layer or to the output units.

The weights of the network are the parameters of the model, which are determined by minimizing some objective function. Usually the objective function will be either the sum of squared distances between the target values and the network

calculated outputs or the (negative of the) log likelihood of the data. To calculate the partial derivatives of the objective function with respect to the weights, a standard back propagation algorithm is used. The optimization techniques such as quasi-Newton or Levenberg-Marquardt algorithms are employed to find the local minima. In practice, we divide the data into training and validation sets. Data from the training set are used for preliminary model fitting whereas data from the validation set are used to assess the adequacy of the model. To avoid over fitting, either the early stopping rule or the regularization techniques such as weight decay or weight elimination method can be used. The typical architecture of a multilayer perceptron neural network is given in Figure 1.1.

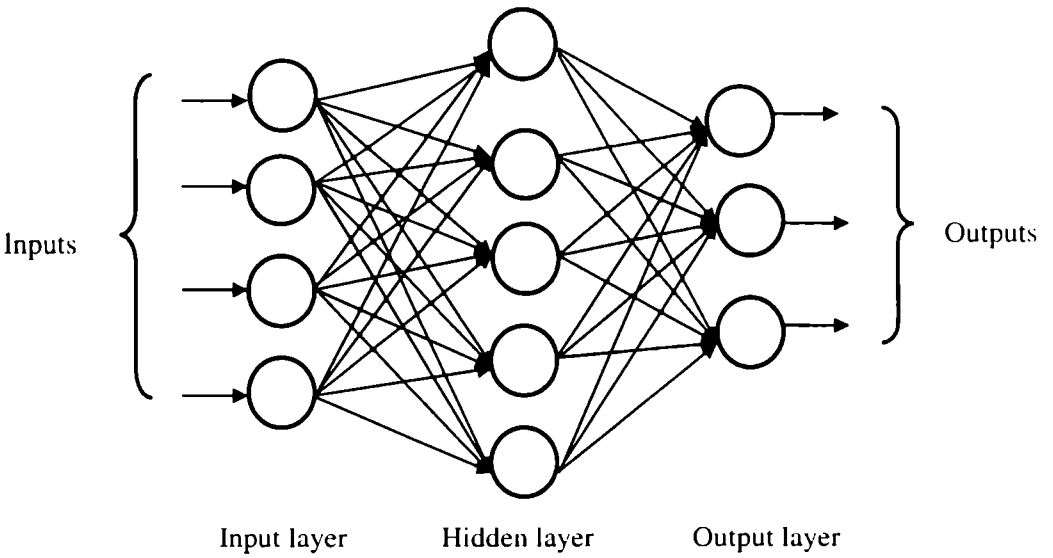


Figure 1.1 The typical architecture of a multilayer perceptron network.

The basic element of a neural network model is a single unit perceptron. A single perceptron, can be treated as a regression model, represented by $y = w_0 + \sum_{j=1}^{p} w_j x_j$ where y is the expected response, $x = (x_1 . x_2 .... x_p)$ is the input vector,

w_j s are the corresponding weights and w_0 is the bias added. For more details on this topic, one could refer to Cheng and Titterington (1994).

Neural networks have been recently used for prediction and classification problems in survival analysis. Kappen and Neijt (1993) trained the 'single time point models' to predict the t year survivor probability, where t is a pre fixed constant. De Laurentiis and Ravdin (1994) mentioned that artificial neural networks are better than Cox regression model under the situations when the proportionality assumption of the hazard rate functions cannot be applied to the data, the relationship of explanatory variables to the outcome is complex and unknown and there are interactions among explanatory variables. Biganzoli et al. (1998) pointed out that the feed forward neural networks provide flexible nonlinear modeling of censored survival data. Street (1998) used a single network to predict survivor probability at each time point. Smith and Anand (2000) transformed the output from Cox regression into survival function estimation using neural networks. Ripley et al. (2004) described the models based on multilayer perceptrons for survival analysis. A neural Bayesian approach to survival analysis was introduced by Bakker and Heskes (1999) and Bakker et al. (2004). For more studies about the use of neural networks in various situations in survival analysis, one could refer to Ravdin and Clark (1992), Macahdo (1997) and Eleuteri et al. (2003a, 2003b) among many others. A survey on the use of neural networks in survival analysis was given by Ripley and Ripley (2001).

Recently, Biganzoli et al. (2006) proposed a partial logistic artificial neural network model for the analysis of competing risks data in the discrete set up. Later, Lisboa et al. (2009) employed the automatic relevance determination technique to regularize this model, which is commonly employed in Bayesian modeling as a regularization technique.

1.10 Present Study

The discussions in previous sections reveal that there has been much research on analyzing various forms of competing risks data. Nevertheless, there are several occasions in survival studies, where the existing models and methodologies are inadequate for the analysis competing risks data. Identifiabilty problem and various types of and censoring induce more complications in the analysis of competing risks data than in classical survival analysis. Parametric models are not adequate for the analysis of competing risks data since the assumptions about the underlying lifetime distributions may not hold well. Motivated by this, in the present study, we develop some new inference procedures, which are completely distribution free for the analysis of competing risks data.

The thesis is organized into seven chapters, of which first chapter is the introductory chapter, where we have pointed out the relevance and scope the study along with a review of literature. In Chapter 2, we introduce a semiparametric Bayesian approach for the analysis of competing risks data. In the context of Bayesian analysis of competing risks data, Gasbarra and Karia (2000) proposed to model competing risks data using overall hazard rate functions and conditional probabilities. Gelfand et al. (2000) considered the modeling of overall baseline hazard rate function using latent failure time models. However, as pointed out in Prentice et al. (1978), modeling and analysis of competing risks data using cause specific hazard rate functions is more realistic. Motivated by this, we develop a semiparametric Bayesian procedure using cause specific hazard rate functions, for the analysis of competing risks data. We assume that each cumulative baseline cause specific hazard rate function has a gamma prior distribution and the estimation of regression parameters is carried out by considering the cumulative baseline cause specific hazard rate functions as nuisance parameters. We derive posterior distribution of the cumulative baseline

cause specific hazard rate functions and then propose Bayes estimators for the cumulative baseline cause specific hazard rate functions under squared error loss function. The proposed method generalises the work of Kalbfleisch (1978) into the competing risks set up. Simulation studies are carried out to assess the finite sample performance of the proposed estimators. We illustrate the practical utility of the method using a real lifetime data given in Klein and Moeschberger (2003).

In many survival studies, patients are only watched within a window of observational time; otherwise we only know that the event time is below or above the window. The data obtained in such situations are doubly censored. The complex nature of the doubly censored data, do not allow us to use the well-known partial likelihood approach given in Fine and Gray (1999) for the analysis of data in the competing risks set up. It may be noted that, the semiparametric regression model given in Cai and Cheng (2004) cannot directly be employed to model cause specific subdistribution functions. Motivated by this, in Chapter 3, we propose a semiparametric transformation model for cause specific subdistribution functions, using the hazard of Gray (1988). We derive estimators for the regression parameters and the cumulative baseline cause specific hazard rate functions. Asymptotic properties of the estimators are discussed. A simulation study is conducted to assess the finite sample behaviour of proposed estimators. We apply the proposed model to a real life data.

Modeling and analysis of lifetime data using neural network models is a topic of recent interest. In Chapter 4, we use different neural network models for various prediction and classification problems in competing risks set up. Time to failure and cause of failure are embedded as input variables. The novelty of our models is that cause of failure is treated as an input variable for the analysis. Applications of the proposed models to real life data sets have been well established. The estimates of

cumulative cause specific hazard rate functions, cause specific subdistribution functions and survivor function are compared to the smoothed estimates of those, using the kernel density method, given in Wells (1994). It is shown that neural network models perform equally well, comparing to the standard techniques but in a more smoothed manner.

The independence of time to failure T and cause of failure C is an important research question in the context of competing risks. In Chapter 5, we develop tests for testing independence of T and C against the alternative that they are not independent. We consider both continuous lifetime data and categorical lifetime data. First, we discuss the tests for independence of time to failure and cause of failure for continuous lifetime data. We introduce a class of test statistics using martingale approach and discuss the limiting distribution of the proposed test statistics. A test statistic using likelihood ratio procedure is developed and its asymptotic distribution is also derived. Simulation studies are conducted to assess the finite sample performance of the proposed test statistics. The methods are also illustrated using two real life data sets. We, then consider the same testing problem with categorical lifetime data. Occasionally in competing risks data, the cause of failure for an individual has not been exactly observed but has only been narrowed down to a subset of all potential risks. Situations with incomplete information about the cause of failure are referred to as masking. We consider four different situations with categorical data. First, we derive likelihood ratio test for the independence of T and C when the data are unmasked. The situations with uncensored and censored lifetimes are discussed subsequently. We, then derive likelihood ratio test statistics for masked data. We carry out simulation studies to assess the power of the proposed test statistics. The proposed procedures are well demonstrated with real data sets.

Quantile function, as an alternative to the distribution function can be employed for modeling and analysis of statistical data. The role of quantile function and other concepts derived from it is well established in exploratory data analysis and in different areas of applied statistics (see Parzen, 1979 and Gilchrist, 2000). In survival studies with heavy tailed lifetime models, a single long term survivor can have a marked effect on reliability measures based on a distribution function. It is therefore more convenient to work with quantile functions that are less influenced by extreme observations. Motivated by this, in Chapter 6, we develop a test statistic based on quantile functions for testing equality of cause specific subdistribution functions. We derive the asymptotic distribution of the test statistic and it is shown to be chi-square. A simulation study is carried out to assess the finite sample behaviour of the test statistic. The practical utility of the procedure is illustrated using two real life data sets.

Finally, Chapter 7 summarizes major conclusions of the study and discusses future works to be carried out in this area.

Chapter Two

A Semiparametric Bayesian approach for the Analysis of Competing Risks Data

2.1 Introduction

In many practical situations, it is possible that we may have some additional information on the lifetime of individuals in terms of past data. For example, in survival studies, disease history of the patient may be often available. Conventional statistical techniques do not provide a method to incorporate this prior information in the data analysis. This motivated researchers to develop Bayesian techniques for the analysis of survival data.

In classical survival analysis, Hjort (1990) used discrete time independent increment beta process to model discrete failure times. Later, Walker and Mallick (1997) employed independent gamma process for modeling competing risks data using piece wise constant hazard rate functions. For a detailed review on different semiparametric Bayesian approaches to model survival data, one may refer to Sinha and Dey (1997), Ibrahim et al. (2001a), Nieto-Barajas and Walker (2002), Cai (2003), Mallick and Walker (2003), Ibrahim et al. (2004), Nieto-Barajas and Walker (2005), Laud et al. (2006) and Henschel et al. (2009).

In competing risks situations without covariates, Gasbarra and Karia (2000) proposed to model the survival data using overall hazard rate functions and

---

The results in this chapter have been accepted for publication as entitled 'A Semiparametric Bayesian Approach for the Analysis of Competing Risks data', in Communications in Statistics- Theory and Methods (see Sreedevi and Sankaran, 2010).

conditional probabilities. Gelfand et al. (2000) considered the modeling of competing risks data using overall baseline hazard rate function through latent failure time approach. Recently, a semiparametric mixture model for the analysis of clustered competing risks data was proposed by Naskar et al. (2005). However, as pointed out in Prentice et al. (1978), modeling and analysis of competing risks data using cause specific hazard rate functions is more realistic. Motivated by this, we introduce a semiparametric Bayesian method, based on cause specific hazard rate functions for the analysis of competing risks data. The proposed method generalises the work of Kalbfleisch (1978) into the competing risks set up.

The chapter is organized as follws. In Section 2.2, we consider the semiparametric Cox proportional hazards model in the competing risks set up. The likelihood for the estimation of regression parameters is derived by treating each cumulative baseline cause specific hazard rate function as a nuisance parameter with a certain prior distribution. We, in Section 2.3, derive the posterior distributions of the cumulative baseline cause specific hazard rate functions and then propose Bayes estimators for the cumulative baseline hazard rate functions under squared error loss function. Simulation studies are carried out in Section 2.4, to assess the finite sample behaviour of the proposed estimator. We, in Section 2.5, demonstrate the practical utility of the method using a real life example given in Klein and Moeschberger (2003). Finally, Section 2.6 summarizes major conclusions of the study.

2.2 Prior Distributions

In this section, we propose a marginal likelihood for the estimation of regression parameters. Consider a competing risks situation with k causes of failure. Let T be a nonnegative random variable representing the lifetime of an individual, $C \in \{1, 2, ..., k\}$ be the cause of failure and X denote the $p \times 1$ vector of covariates. Now, we consider the Cox proportional hazards model in competing risks set up given by

$$\lambda_j(t|X) = \lambda_{0j}(t)\exp(\beta_j X) \qquad j = 1,2,...,k$$

where β_j is the $p\times 1$ vector of regression parameters for cause j and $\lambda_{0j}(t)$ is the baseline cause specific hazard rate function common to all individuals and $\lambda_j(t|X)$ is the cause specific hazard rate function at time t in presence of covariate vector X

Suppose that the lifetime variable T is randomly right censored by the variable Z Now, we observe $(t,\delta,\delta C,X)$ where $t=\min(T,Z)$ and $\delta = I(T=t)$. Let $(t_i,\delta_i,\delta_i C_i,X_i)$, $i=1,2,...,n$ be independent and identically distributed copies of $(t,\delta,\delta C,X)$. Let $S(t|X)$ be the survivor function and $f(t|X)$ be the probability density function in presence of covariate vector X Now, the likelihood function for the observed data can be formulated as

$$L = \prod_{i=1}^{n}\prod_{j=1}^{k}\left[f_j(t_i|X_i)\right]^{\delta_i}\left[S(t_i|X_i)\right]^{1-\delta_i} \tag{2.1}$$

where $f_j(t|X)$ is the cause specific subdensity function in presence of the covariate vector X .Using (1.22), we obtain

$$S(t|X) = \exp\left(-\sum_{j=1}^{k}\Lambda_j(t|X)\right) \tag{2.2}$$

where $\Lambda_j(t|X) = \Lambda_{0j}(t)\exp(\beta_j' X)$ is the cumulative cause specific hazard rate function given X Now, we can write

$$S(t|X) = \prod_{j=1}^{k} G_j(t|X) \tag{2.3}$$

where $G_j(t|X) = \exp(-\Lambda_j(t|X))$ for $j=1,2,...,k$

Note that, even though the functions $G_j(t|X)$ for $j=1,2,...,k$ are not the survivor functions of any observable random variables, they do satisfy the mathematical properties of a survivor function and can be referred as survivor-like

functions (Porta et al., 2007). When the distinct causes of failures are assumed to be independent, $1-G_j(t|X)$ is fully interpretable as the probability of failure due to cause j, if the other causes of failures were removed (Gooley et al., 1999). Porta et al. (2008) provided a detailed review of the survivor-like functions that can be defined in competing risks set up. By defining the indicator variable $\delta_{ij} = I(C_i = j)$, we can rewrite the likelihood function (2.1) in terms of survivor-like functions as

$$L = \prod_{i=1}^{n}\prod_{j=1}^{k}\left[g_j(t_i|X_i)\right]^{\delta_{ij}}\left[G_j(t_i|X_i)\right]^{1-\delta_{ij}} \tag{2.4}$$

where $g_j(y) = -G_j'(y)$ with prime represents the derivative of $G_j(y)$ with respect to y

Now, we modify the likelihood (2.4) to incorporate the prior information. Conditional on the cumulative baseline cause specific hazard rate functions, $G_j(t|X)$ can be written as

$$G_j\left(t|X,\Lambda_{0j}(t)\right) = \exp\left(-\Lambda_{0j}(t)\exp\left(\beta_j' X\right)\right) \quad j = 1,2,...,k \tag{2.5}$$

Consider a partition of $[0,\infty)$ into finite number of n disjoint intervals, say $[a_i, a_{i+1})$ for $i = 0,1,...,n$ with $a_0 = 0$ and $a_{n+1} = \infty$ Define the hazard contribution of the ith interval due to jth cause as

$$q_{ij} = P\left[T \in [a_{i-1}, a_i), C_i = j \middle| T \geq a_{i-1}, \Lambda_{0j}(t_i)\right] \tag{2.6}$$

where $q_{ij} = 1$ if $P\left[T \geq a_{i-1}, C_i = j \middle| \Lambda_{0j}(t_i)\right] = 0$ for $i = 0,1,...,n \quad j = 1,2,...,k$

Note that, $\Lambda_{0j}(a_0) = 0$ for each j and hence

$$\Lambda_{0j}(a_i) = \sum_{l=1}^{i} -\log(1-q_{lj}) = \sum_{l=1}^{i} r_{lj} \quad i = 1,2...n \quad j = 1,2...,k \tag{2.7}$$

As mentioned in Kalbfleisch (1978), the probability distributions of the prior parameters can be specified on the space $\{\Lambda_{0j}(t)\}$ by specifying the finite

dimensional distributions of $q_{1j},....,q_{nj}$ for each partition $[a_{i-1},a_i)$ for $i=0,1,...,n$. It is clear from (2.7) that each $\Lambda_{0j}(t)$ is a non decreasing process with independent prior by its construction. Now the problem reduces to the specification of a non decreasing independent increment process prior for $\Lambda_{0j}(t)$ and for this, we need to specify independent prior for each r_{ij}. Accordingly, we follow the prior elicitation method given in Ibrahim et al. (2001b).

Let $D_0=(n_0,T_0,\upsilon_0,C_0\upsilon_0,X_0)$ denote the data from the previous study, where n_0 denotes the sample size of the previous study, T_0 denotes a right censored vector of survival times with censoring indicators υ_0, X_0 denotes the $p\times 1$ vector of covariates and $C_0\in\{1,2,....k\}$ be the possible causes of failure. Since $\Lambda_{0j}(t)$ is treated as a nuisance parameter, we assume that the cumulative baseline cause specific hazard rate functions and the vector of regression parameters are independent.
Let $\pi_0(\beta_j,\Lambda_{0j}(t))$ denote the prior for $(\beta_j,\Lambda_{0j}(t))$. Now the joint prior density for $(\beta_j,\Lambda_{0j}(t))$ is given by

$$\pi(\beta_j,\Lambda_{0j}(t)|D_0)=\pi(\beta_j|D_0)\pi(\Lambda_{0j}(t)|D_0)\quad j=1,2,...,k \tag{2.8}$$

We take, $\pi(\Lambda_{0j}(t)|D_0)$ as a product of M independent gamma densities where the mean and variance of the gamma density can be elicited from D_0 (Ibrahim et al., 2001b).

The prior given by (2.8) has several advantages. First, it has a closed form and it is easy to interpret. Second, the prior elicitation is straight forward. Third, (2.8) is computationally feasible and easy to interpret. Fourth, the prior given by (2.8) assumes a prior independence between $(\beta_j,\Lambda_{0j}(t))$, which further simplifies the interpretations as well as the elicitation scheme.

Let $\Lambda_{0j}^{*}(t)$ be an initial guess at $\Lambda_{0j}(t)$ and γ_j be the specification of weight attached to that guess. Also take, $\alpha_{ij} = \gamma_j \Lambda_{0j}^{*}(a_i)$ for each j

Assume that , $r_{ij} = -\log(1-q_{ij})$ has independent gamma distributions

$$r_{ij} \sim G(\alpha_{ij} - \alpha_{i-1j}, \gamma_j) \; i = 1,2...,n \; ; \; j = 1,2...,k \tag{2.9}$$

where $\gamma_j > 0$ is the scale parameter and $\alpha_{ij} - \alpha_{i-1j}$ is the shape parameter of the gamma prior distribution.

Now, $\exp(-\Lambda_{0j}^{*}(t))$ is a completely specified survivor-like function and $\Lambda_{0j}(t)$ is a gamma process by its construction. By considering the partition $(0,t),[t,\infty)$ and (2.9) it shows that $\Lambda_{0j}(t) \sim G(\gamma_j \Lambda_{0j}(t), \gamma_j)$ and hence $E\{\Lambda_{0j}(t)\} = \Lambda_{0j}^{*}(t)$ and $\text{var}\{\Lambda_{0j}(t)\} = \Lambda_{0j}^{*}(t)/\gamma_j$, $j = 1,2,...,k$

Now, we derive the marginal likelihood based on parameters of the prior distribution for estimating $\beta_j, j = 1,2,...,k$ First, we suppose that, there is no censoring and all individuals have observed lifetimes. Now conditional on $\Lambda_{0j}(t)$, we have

$$\prod_{i=1}^{n} G_j\left(t_i \middle| \beta_j, X_i, \Lambda_{0j}(t_i)\right) = \exp\left(-\sum_{i=1}^{n} \Lambda_{0j}(t_i) \exp(\beta_j' X_i) \right)$$

$$j = 1,2,...,k \, . \tag{2.10}$$

Without loss of generality we can assume, $t_1 \le t_2 \le .... \le t_n$ and $r_{ij} = \Lambda_{0j}(t_i) - \Lambda_{0j}(t_{i-1})$ for $i = 1,2,...,n+1$ and $j = 1,2,...,k$, where $t_0 = 0$ and $t_{n+1} = \infty$. Here t_i's define the intervals of partition of $[0,\infty)$, with $t_0 = 0$ and $t_{n+1} = \infty$.

Now $r_{ij} \sim G(\gamma_j(\Lambda_{0j}^{*}(t_i) - \Lambda_{0j}^{*}(t_{i-1})), \gamma_j)$ for $i = 1,2,...,n+1$; $j = 1,2,...,k$,and they are independently distributed. Since, $\Lambda_{0j}(t_i) = \sum_{l=1}^{i} r_{lj}$ for $i = 1,2,...,n+1$; $j = 1,2,...,k$,it follows from (2.10) that

$$\prod_{i=1}^{n} G_j\left(t_i \middle| \beta_j, X_i, r_{1j}, r_{2j}, ..., r_{n+1j}\right) = \exp\left(-\sum_{i=1}^{n} r_{ij} A_{ij}\right) \tag{2.11}$$

where

$$A_{ij} = \sum_{l \in R(t_i)} \exp\left(\beta_j' X_l\right) \quad l = 1, 2, ..., n\,; \quad j = 1, 2, ..., k\,, \tag{2.12}$$

with $R(t_i)$ is the set all individuals at risk just prior to time t_i.

Integrating (2.11) with respect to the distributions of $r_{1j}, r_{2j}, ..., r_{nj}$, we get,

$$\prod_{i=1}^{n} G_j\left(t_i \middle| \beta_j, X_i\right) = \exp\left(-\sum_{i=1}^{n} \gamma_j B_{ij} \Lambda_{0j}^{*}(t_i)\right) \tag{2.13}$$

where

$$B_{ij} = -\log\left(1 - \exp\left(\beta_j' X_i\right) / \left(\gamma_j + A_{ij}\right)\right) \quad i = 1, 2, ..., n \quad j = 1, 2, ..., k\,. \tag{2.14}$$

The expression (2.13) is true for any cumulative baseline cause specific hazard rate function $\Lambda_{0i}^{*}(t)$. We assume that, $\Lambda_{0i}^{*}(t)$ is absolutely continuous. Following Kalbfleisch (1978), under the assumption of no ties in the data, the likelihood function (2.4) is obtained as

$$L(\beta_1, ..., \beta_k) = \prod_{j=1}^{k}\left(\gamma_j{}^{n} \exp\left\{-\sum_{i=1}^{n} \gamma_j B_{ij} \Lambda_{0j}^{*}(t_i)\right\} \prod_{i=1}^{n}\left\{\lambda_{0j}^{*}(t_i) B_{ij}\right\}\right) \tag{2.15}$$

where $\lambda_{0j}^{*}(t) = (d/dt)\Lambda_{0j}^{*}(t)$, $j = 1, 2, ..., k$.

The likelihood given by (2.15) is a product of k identical terms, of which j th factor is given by

$$L_j(\beta_j) = \gamma_j{}^{n} \exp\left\{-\sum_{i=1}^{n} \gamma_j B_{ij} \Lambda_{0j}^{*}(t_i)\right\} \prod_{i=1}^{n}\left\{\lambda_{0j}^{*}(t_i) B_{ij}\right\} \quad j = 1, 2, ..., k \tag{2.16}$$

In presence of censored lifetimes (2.15) becomes

$$L(\beta_1, ..., \beta_k) = \prod_{j=1}^{k}\left(\exp\left\{-\sum_{i=1}^{n} \gamma_j B_{ij} \Lambda_{0j}^{*}(t_i)\right\} \prod_{i=1}^{n}\left\{\gamma_j \lambda_{0j}^{*}(t_i) B_{ij}\right\}^{\delta_{ij}}\right). \tag{2.17}$$

Under the assumption of piecewise hazard rates, we have $\Lambda_{0j}(t) = \lambda_{0j} t$, where λ_{0j} is left unspecified. Now from (2.17), it follows that the likelihood of $(\beta_1,...,\beta_k)$ and $(\lambda_{01},...,\lambda_{0k})$ for each specified γ_j, $j = 1,2,...,k$ is given by

$$L(\beta_1,...,\beta_k) = \prod_{j=1}^{k}\left(\exp\left(-\gamma_j \lambda_{0j} \sum_{i=1}^{n} t_i B_{ij} \right) \prod_{i=1}^{n} \left(\gamma_i \lambda_{0j} B_{ij} \right)^{\delta_{ij}} \right). \qquad (2.18)$$

Estimates of $\beta = (\beta_1, \beta_2,..., \beta_k)$ are obtained by maximizing the likelihood (2.18), which can be done using Newton-Raphson iterative algorithm.

Consider a situation, where $\Lambda_{0j}(t)$ is fully specified at the starting point. Then the j th factor of $L(\beta_1,...,\beta_k)$ will be given by

$$\lim_{\gamma_j \to \infty} L_j(\beta_j) = \exp\left\{ -\sum_{i=1}^{n} \Lambda^{*}_{0j}(t_i) \exp(\beta_j X_i) \right\} \prod_{i=1}^{n} \lambda^{*}_{0j}(t_i) \exp(\beta_j X_i)$$

$$j = 1,2,...,k \qquad (2.19)$$

On the other hand, suppose a situation with little faith in the prior estimates of $\Lambda_{0j}(t)$. Then to a first order approximation, the j th factor of $L(\beta_1,...,\beta_k)$ will be given by

$$L_j(\beta_j) \equiv K_j \prod_{i=1}^{n} \exp(\beta_j X_i) / A_{ij} \qquad j = 1,2,...,k \qquad (2.20)$$

where K_j is the normalising constant.

This shows that $L(\beta_1,...,\beta_k)$ gives a spectrum of likelihoods ranging from truly nonparametric situations (γ_j near to 0, $j = 1,2,...,k$) to situations where $\Lambda_{0j}(t)$, $j = 1,2,...,k$ is assumed to be completely known. By allowing $\Lambda^{*}_{0j}(t)$ to depend on one or more unknown parameters, for example if $\Lambda^{*}_{0j}(t) = \lambda_{0j} t$, $j = 1,2,...,k$, the likelihood (2.18) corresponds to a generalization of the usual parametric analysis $(\gamma_j \to \infty, j = 1,2,...,k)$ with exponential survivals. The dependence of prior

parameters on the estimate of β can be evaluated by choosing different sets of values for prior parameters in (2.18).

In the case of ties

$$T_{11} = ... = T_{1b_1} = t_1 < T_{21} = ... = T_{2b_2} = t_2 < .... < T_{m1} = ... = T_{mb_m} = t_m,$$

the j th factor of the likelihood for estimating $\beta_j; j = 1,2,...,k$ is obtained as

$$L_j^*(\beta_j) = \prod_{i=1}^{m} \left\{ \frac{\gamma_j + A_{i+1}}{\gamma_j + A_i} \right\}^{\gamma_j \Lambda_{0j}^*(t)} \sum_{l=0}^{b_i} (-1)^l \sum_{p_l \in P_l} \exp\left\{ -\left(\gamma_j + A_{i+1} + \sum_{s \in p_l} \exp\left(\beta_j' X_{sl}\right) \right) u_j \right\}$$

$$j = 1,2,...,k \quad (2.21)$$

where P_l is the class of all subsets of l items chosen from $1,2,...b_i$, and A_{ij} is defined as in (2.12) and $A_{m+1\,j} = 0$.

2.3 Posterior Distributions

In the following, we first derive the posterior distribution of the cumulative baseline cause specific hazard rate function $\Lambda_{oj}(t)$ and then we propose Bayes estimator for $\Lambda_{oj}(t)$ under squared error loss function.

Let $(t_i, \delta_i, \delta_i C_i, X_i)$, $i = 1,2,....,n$ be a set of independent and identically distributed observations as defined in Section 2.2. Let δ_i and δ_{ij} denote the indicator variables as defined in Section 2.2.

The conditional survivor-like function of T for cause j is

$$G_j\left(t \middle| X, \Lambda_{0j}(t)\right) = \exp\left(-\Lambda_{0j}(t) \exp\left(\beta_j' X\right)\right). \quad (2.22)$$

As mentioned in Section 2.2, we assume that the prior distribution of $\Lambda_{0j}(t)$ is the gamma process with parameters γ_j and $\Lambda_{0j}^*(t)$

Consider a partition of $[0,\infty)$ into m disjoint intervals $[a_o = 0, a_1), [a_1, a_2), ..., [a_{m-1}, a_m = \infty)$ and suppose that $n = 1$ The extension to $n \geq 2$

values follows directly from the results for $n=1$ Assume that $a_{i-1} \le t_1 < a_i$ and $r_{i1j} = -\log\left(1-q_{i1j}\right)$ and $r_{i2j} = -\log\left(1-q_{i2j}\right)$ where q_{i1j}, q_{i2j} are the hazard contributions of the intervals $[a_{i-1}, t_1)$ and $[t_1, a_i)$ respectively. Now r_{ij} has independent gamma distributions for $i=1,2,....,m$ and $j=1,2,...,k$ Then the conditional survivor function for t_1 is given by

$$S\left(t_1 \middle| r_j, r_{i1j}, X_1\right) = \prod_{j=1}^{k} G_j\left(t_1 \middle| r_j, r_{i1j}, X_1\right) = \prod_{j=1}^{k} \exp\left\{-\left(r_{1j} + ... + r_{i-1j} + r_{i1j}\right)\exp\left(\beta_j' X_1\right)\right\} \tag{2.23}$$

where $r_j = \left(r_{1j}, r_{2j}, ..., r_{i-1j}\right)$.

By taking the product of (2.23) with independent gamma probability densities for the priors of $r_{1j}, ..., r_{i-1j}, r_{i1j}, r_{i2j}, r_{i+1j}, ..., r_{mj}$ and integrating over $r_{i0j} = r_{i1j} + r_{i2j}$, we get

$$S\left(t_1, r_{lj} \in \left(r_{l0j}, r_{l0j} + dr_{l0j}\right), l = 1,2...m \middle| X_1\right) = \prod_{j=1}^{k} h_j\left(t_1, r_{10j}, ..., r_{m0j} \middle| X_1\right) \prod_{l=1}^{m} dr_{l0j} \tag{2.24}$$

where $h_j(.)$ is the joint density of t_1 and r_{i0j} given the covariate $X_1, i=1,2,...,m$

The posterior density of r_j given $T_1 = t_1$ is obtained as

$$f_j\left(r_j \middle| T_1 = t_1, r_{1j}\right) = \frac{\left(\partial/\partial t_1\right) h_j\left(t_1, r_{1j}, ...r_{mj}, C_1 = j \middle| X_1\right)}{\gamma_j \lambda_j^*(t_1) \log\left(\gamma_j/\gamma_{1j}\right)\left(\gamma_j/\gamma_{1j}\right)^{\gamma_j \lambda_{0j}^*(t_1)}} \tag{2.25}$$

where $\gamma_{1j} = \gamma_j + \exp\left(\beta_j' X_1\right)$ and the denominator is the marginal density of T_1 from the expression for likelihood (2.15). A closed form expression for equation (2.25) does not exist, but the probability laws can be characterized using moment generating functions. Now $h_j\left(t_1, r_{1j}, ..., r_{mj}, C_1 = j \middle| X_1\right)$ is a product of m factors, of which lth one involves r_{lj} only, while t_1 is the only lifetime such that, $a_{i-1} \le t_1 < a_i$. It follows that

$r_{1j},...,r_{mj}$ are independent with respect to the posterior distribution and that the density of $r_{lj}\ (l \neq 1)$ is gamma. Now the moment generating functions of r_{ij} can be written as

$$M_{r_j}\left(\theta_j | T_1 = t_1\right) = E\left\{\exp\left(\theta_{1j} r_{1j} + ... + \theta_{mj} r_{mj}\right) | T_1 = t_1\right\} = \prod_{l=1}^{m} M_{r_{lj}}\left(\theta_{lj} | T_1 = t_1\right)$$

with $\theta_j = \left(\theta_{1j}, \theta_{2j}, ..., \theta_{mj}\right)$,

where

$$M_{r_{lj}}\left(\theta_{lj} | T_1 = t_1\right) = \left\{\gamma_{1j} / \left(\gamma_{1j} - \theta_{lj}\right)\right\}^{\alpha_{lj} - \alpha_{l-1,j}} \qquad l = 1,2,...,i-1$$

and

$$M_{r_{lj}}\left(\theta_{lj} | T_1 = t_1\right) = \left\{\gamma_j / \left(\gamma_j - \theta_{lj}\right)\right\}^{\alpha_{lj} - \alpha_{l-1,j}} \qquad l = i+1,...,m.$$

For the i th interval we evaluate the differential,

$$\frac{\partial}{\partial t_1} \int\int \exp\left\{\theta_{ij}\left(r_{i1j} + r_{i2j}\right)\right\} \exp\left\{-r_{i1j} \exp\left(\beta_j' X_1\right)\right\} g_{1j}\left(r_{i1j}\right) g_{2j}\left(r_{i2j}\right) dr_{i1j} dr_{i2j}$$

where g_{1j} and g_{2j} are independent gamma prior densities for r_{i1j} and r_{i2j}. The ratio of these gamma prior densities to the marginal density of t_1 is proportional to the moment generating function of r_{ij},

$$M = \left(\frac{\gamma_{1j}}{\gamma_{1j} - \theta_{ij}}\right)^{\alpha_j(t_1) - \alpha_{i-1,j}} \left(\frac{\gamma_j}{\gamma_j - \theta_{ij}}\right)^{\alpha_{ij} - \alpha(t_1)} \left\{\log\left(\frac{\gamma_{1j} - \theta_{ij}}{\gamma_j - \theta_{ij}}\right) \Big/ \log\left(\frac{\gamma_{1j}}{\gamma_j}\right)\right\} \qquad (2.26)$$

where $\alpha_{lj} = \gamma_j \Lambda_{0j}^{*}\left(a_l\right)$ for $l = 1,2...m$ and $\alpha_j\left(t_1\right) = \gamma_j \Lambda_{0j}^{*}\left(t_1\right)$ It follows that, r_{ij} is distributed as the sum of three independent random variables W_j, Y_j and U_j where W_j and Y_j are gamma variables and $U_j \sim L\left(\gamma_{1j}, \gamma_j\right)$ with density,

$$u_j^{-1} \exp\left\{\left(-\gamma_j u_j\right) - \exp\left(-\gamma_j u_j\right)\right\} \Big/ \log\left(\gamma_{1j} / \gamma_j\right).$$

Moment generating function of the density of U_j is the last factor of (2.26). Now, given $t_1 < t_2 < ..... < t_n$, $\Lambda_{0j}\left(t\right)$ is an independent increment process. Then

$$U_{ij} \sim L\left(\gamma_j + A_{ij}, \gamma_j + A_{i+1\,j}\right) \qquad i = 1,2,...,m$$

where $A_{ij} = \sum_{l \in R(t_i)} \exp\left(\beta_j' X_l\right)$ ($i = 1,2,...,m$, $j = 1,2,...,k$) as defined in (2.12).

Under squared error loss function, $E\left\{\Lambda_{0j}(t) \middle| t, X, \beta_j\right\}$ is the estimator of $\Lambda_{0j}(t)$ If $t_1 < t_2 < ..... < t_n$ and $t_{i-1} \le t < t_i$ then the posterior distribution of $\Lambda_{0j}(t)$ is the sum of independent variables $W_{1j} + U_{1j} + ... + W_{i-1\,j} + U_{i-1\,j} + \rho_{ij}$,

where

$W_{lj} \sim G\left(\gamma_j \Lambda_{0j}^*(t_l) - \gamma_j \Lambda_{0j}^*(t_{l-1}), \gamma_j + A_{lj}\right), U_{lj} \sim L\left(\gamma_j + A_{lj}, \gamma_j + A_{l+1\,j}\right)\ l = 1,2,...,i-1$

and $\rho_{ij} \sim G\left(\gamma_j \Lambda_{0j}^*(t_i) - \gamma_j \Lambda_{0j}^*(t_{i-1}), \gamma_j + A_{ij}\right)$ for $j = 1,2,...,k$

Now it follows that,

$$E\left(U_{lj}\right) = \frac{\exp\left(\beta_j' X_l\right)}{\left(\gamma_j + A_{lj}\right)\left(\gamma_j + A_{l+1\,j}\right)} \Bigg/ \log\left(\frac{\gamma_j + A_{lj}}{\gamma_j + A_{l+1\,j}}\right)$$

and the Bayes estimator of $\Lambda_{oj}(t)$ is given by

$$\tilde{\Lambda}_{oj}(t_i) = E\left\{\Lambda_{0j}(t_i) \middle| t_i, X, \beta_j\right\} = \sum_{l=1}^{i-1}\left\{E\left(X_{lj} + U_{lj}\right)\right\} + E\left(\rho_{ij}\right)$$

$$i = 1,2,...,m\,;\ j = 1,2,...,k \qquad (2.27)$$

Using the estimators of β_j and Λ_{0j}, $j = 1,2,...,k$, we obtain the estimator of $S\left(t \middle| X\right)$ from (1.28). The estimator of $F_j\left(t \middle| X\right)$, $j = 1,2,...,k$ is obtained from (1.29). However, note that the estimators of $S\left(t \middle| X\right)$ and $F_j\left(t \middle| X\right)$ are not true Bayesian estimators.

When parameter values of prior distribution are near to zero, $E\left(U_{lj}\right) \cong A_{lj}^{-1}$

$l = 1,2,...,n-1$ and $E\left(X_{lj}\right) \cong E\left(\rho_{ij}\right) \cong 0$ and it follows that

$E\left\{\Lambda_{0j}(t) \middle| t, X, \beta_j\right\} \cong \sum_{l:t_l \le t} A_{lj}^{-1}$ $i = 1,2,...,m\,;\ j = 1,2,...,k$.

Note that for a single censored observation, the moment generating function of r_{ij} is (2.26) except for the last factor. Thus a censored observation does not insert a random jump into the posterior distribution of $\Lambda_{0j}(t)$ and doesn't upshot any further complication in the analysis.

Remark 2.1: When $j=1$, the procedures developed in Section 2.3 and 2.4 reduce to the work by Kalbfleisch (1978).

2.4 Simulation Study

We carry out an extensive simulation study to assess the finite sample performance of the proposed estimators. We consider two causes of failure. For simplicity, we consider a single covariate X, which is generated using a Bernoulli random number. The data are generated from the following model. The subdistribution function due to failures from cause 1 is given by

$$F_1(t|X) = 1-\left[1-m\left(1-\exp(-t)\right)\right]^{\exp(\beta_1' X)}$$

which is a unit exponential mixture with mass $(1-m)$ at ∞ when $X=0$. The subdistribution function due to failures from cause 2 is given by

$$F_2(t|X) = 1-\left[1-(1-m)\left(1-\exp(-t)\right)\right]^{\exp(\beta_2 X)}$$

Censoring variable is generated from uniform distribution over the interval $(0,a)$ where a is chosen in such a way that 20% or 40% of the lifetimes are censored. We generate random samples of size $n=$ 100 and 250.

To calculate the absolute value of the bias and mean square error (MSE) of the estimates, 1000 samples are generated. As the values of bias and MSE do not vary with m, we present the simulation results for $m=0.5$. To study the effect of censoring, we consider three situations viz. no censoring, mild censoring (20% of the observations are censored) and heavy censoring (40% of the observations are censored). The pattern of simulation results seen to be similar for all the combinations

of prior parameters and hence we show the results for only two different combinations of prior parameter values viz. $\gamma_1 = \gamma_2 = 1$ and $\gamma_1 = 10,\ \gamma_2 = 50$. Absolute bias and MSE of the estimates of β_j's; $j = 1,2$ in the three censoring situations for both the combinations are given in Tables 2.1 and 2.2 respectively. We note that, both absolute bias and MSE decrease with increase in sample size. Also it is clear that, if the censoring percentage increases absolute bias and MSE also increases. Absolute value of bias and MSE of $\hat{F}_j(t|X)$; $j = 1,2$ are estimated at five arbitrarily selected time points. Absolute bias and MSE of the cause specific subdistribution functions in different censoring situations for the two prior parameter combinations $\gamma_1 = \gamma_2 = 1$ and $\gamma_1 = 10,\ \gamma_2 = 50$ are given in Tables 2.3(a)-(c) and 2.4(a)-(c) respectively. It can be noted from Tables 2.3(a)-(c)- 2.4 (a)-(c) that, both absolute bias and MSE decrease with increase in sample size and increase if censoring percentage increases.

Table 2.1 Absolute bias and MSE of $\hat{\beta}_j$; $j = 1,2$ for $\gamma_1 = 1,\ \gamma_2 = 1$

β_1	β_2	Sample size	Absolute bias $\hat{\beta}_1$	MSE $\hat{\beta}_1$	Absolute bias $\hat{\beta}_2$	MSE $\hat{\beta}_2$
No Censoring						
1	1	100	0.0102	0.0004	0.0166	0.0022
		250	0.0073	0.0001	0.0071	0.0002
Mild(20%) Censoring						
1	1	100	0.0163	0.0005	0.0184	0.0121
		250	0.0093	0.0002	0.0163	0.0003
Heavy (40%) Censoring						
1	1	100	0.0198	0.0008	0.0214	0.0118
		250	0.0100	0.0005	0.0184	0.0003

Table 2.2 Absolute bias and MSE of $\hat{\beta}_j$; $j = 1,2$ for $\gamma_1 = 10$, $\gamma_2 = 50$

β_1	β_2	Sample size	Absolute bias $\hat{\beta}_1$	MSE $\hat{\beta}_1$	Absolute bias $\hat{\beta}_2$	MSE $\hat{\beta}_2$
No Censoring						
-1	2	100	0.0024	0.0032	0.0132	0.0009
		250	0.0018	0.0022	0.0098	0.0008
Mild (20%) Censoring						
-1	2	100	0.0029	0.0099	0.0165	0.0045
		250	0.0023	0.0055	0.0104	0.0012
Heavy (40%) Censoring						
-1	2	100	0.0034	0.0024	0.0176	0.0158
		250	0.0027	0.0222	0.0110	0.0092

Table 2.3a Absolute bias and MSE of $\hat{F}_j(t|x)$; $j = 1,2$ with no censoring for $\gamma_1 = 1, \gamma_2 = 1$ and β_1=1, β_2=1

Sample size	Time points	Absolute bias $\hat{F}_1(t\|x)$	MSE $\hat{F}_1(t\|x)$	Absolute bias $\hat{F}_2(t\|x)$	MSE $\hat{F}_2(t\|x)$
100	0.10	0.0111	0.0185	0.0020	0.0003
	0.25	0.0249	0.1322	0.0071	0.0127
	0.50	0.0425	0.1532	0.0157	0.0532
	1.00	0.0797	0.1121	0.0353	0.1121
	2.00	0.0999	0.1902	0.0490	0.1732
250	0.10	0.0081	0.0066	0.0001	0.0001
	0.25	0.0200	0.0398	0.0027	0.0007
	0.50	0.0321	0.1030	0.0113	0.0128
	1.00	0.0786	0.0021	0.0307	0.1032
	2.00	0.0807	0.0921	0.0380	0.1421

Table 2.3b Absolute bias and MSE of $\hat{F}_j(t|x)$, $j=1,2$ with mild (20%) censoring for $\gamma_1=1, \gamma_2=1$ and $\beta_1=1$, $\beta_2=1$

Sample size	Time points	Absolute bias $\hat{F}_1(t\|x)$	MSE $\hat{F}_1(t\|x)$	Absolute bias $\hat{F}_2(t\|x)$	MSE $\hat{F}_2(t\|x)$
100	0.10	0.0129	0.0107	0.0104	0.0012
	0.25	0.0574	0.0329	0.0430	0.0172
	0.50	0.1119	0.0908	0.1024	0.0156
	1.00	0.1818	0.1289	0.1818	0.0832
	2.00	0.2089	0.1876	0.2089	0.0932
250	0.10	0.0104	0.0098	0.0085	0.0001
	0.25	0.0235	0.0231	0.0235	0.0111
	0.50	0.0501	0.0543	0.0501	0.0098
	1.00	0.0851	0.0983	0.0851	0.0721
	2.00	0.1045	0.1093	0.1045	0.0762

Table 2.3c Absolute bias and MSE of $\hat{F}_j(t|x)$; $j=1,2$ with heavy (40%) censoring for $\gamma_1=1, \gamma_2=1$ and $\beta_1=1$, $\beta_2=1$

Sample size	Time points	Absolute bias $\hat{F}_1(t\|x)$	MSE $\hat{F}_1(t\|x)$	Absolute bias $\hat{F}_2(t\|x)$	MSE $\hat{F}_2(t\|x)$
100	0.10	0.0197	0.0187	0.0201	0.0098
	0.25	0.0784	0.0456	0.0560	0.0221
	0.50	0.1459	0.1098	0.1198	0.0224
	1.00	0.1918	0.1342	0.2098	0.1234
	2.00	0.2234	0.1911	0.2564	0.1256
250	0.10	0.0154	0.0167	0.0187	0.0008
	0.25	0.0236	0.0434	0.0467	0.0195
	0.50	0.0703	0.0709	0.0809	0.0098
	1.00	0.1291	0.1004	0.1567	0.0876
	2.00	0.1651	0.1198	0.1897	0.1098

Table 2.4a Absolute bias and MSE of $\hat{F}_j(t|x)$; $j=1,2$ with no censoring for $\gamma_1 = 10$, $\gamma_2 = 50$ and β_1=-1, β_2=2

Sample size	Time points	Absolute bias $\hat{F}_1(t\|x)$	MSE $\hat{F}_1(t\|x)$	Absolute bias $\hat{F}_2(t\|x)$	MSE $\hat{F}_2(t\|x)$
100	0.10	0.0048	0.0024	0.0001	0.0001
	0.50	0.0135	0.0351	0.0023	0.0007
	1.00	0.0312	0.0321	0.0065	0.0009
	2.00	0.0581	0.0666	0.0098	0.0033
	5.00	0.0982	0.1084	0.0132	0.0155
250	0.10	0.0023	0.0012	0.0001	0.0001
	0.50	0.0098	0.0209	0.0014	0.0004
	1.00	0.0231	0.0245	0.0041	0.0006
	2.00	0.0322	0.0321	0.0067	0.0021
	5.00	0.0876	0.0897	0.0092	0.0098

Table 2.4b Absolute bias and MSE of $\hat{F}_j(t|x)$; $j=1,2$ with mild (20%) censoring for $\gamma_1 = 10$, $\gamma_2 = 50$ and β_1=-1, β_2=2

Sample size	Time points	Absolute bias $\hat{F}_1(t\|x)$	MSE $\hat{F}_1(t\|x)$	Absolute bias $\hat{F}_2(t\|x)$	MSE $\hat{F}_2(t\|x)$
100	0.10	0.0062	0.0083	0.0004	0.0004
	0.50	0.0203	0.0382	0.0080	0.0009
	1.00	0.0543	0.0555	0.0082	0.0013
	2.00	0.0672	0.0843	0.0132	0.0100
	5.00	0.1093	0.1012	0.0145	0.0421
250	0.10	0.0034	0.0045	0.0003	0.0004
	0.50	0.0122	0.0212	0.0058	0.0006
	1.00	0.0342	0.0387	0.0079	0.0009
	2.00	0.0367	0.0755	0.0134	0.0085
	5.00	0.0933	0.0908	0.0142	0.0143

Table 2.4c Absolute bias and MSE of $\hat{F}_j(t|x)$; $j=1,2$ with heavy (40%) censoring for $\gamma_1 = 10$, $\gamma_2 = 50$ and β_1=-1, β_2=2

Sample size	Time points	Absolute bias $\hat{F}_1(t\|x)$	MSE $\hat{F}_1(t\|x)$	Absolute bias $\hat{F}_2(t\|x)$	MSE $\hat{F}_2(t\|x)$
100	0.10	0.0101	0.0087	0.0008	0.0007
	0.50	0.0232	0.0421	0.0088	0.0013
	1.00	0.0772	0.0592	0.0121	0.0045
	2.00	0.0921	0.0932	0.0196	0.0120
	5.00	0.1041	0.1211	0.0202	0.0401
250	0.10	0.0077	0.0062	0.0008	0.0007
	0.50	0.0189	0.0219	0.0081	0.0011
	1.00	0.0432	0.0382	0.0100	0.0021
	2.00	0.0532	0.0721	0.0163	0.0090
	5.00	0.0991	0.1013	0.0192	0.0140

2.5 Data Analysis

We apply the proposed method for the analysis of the bone marrow transplantation data of 43 patients given in Klein and Moeschberger (2003). Time to relapse for a patient in days was recorded as survival time. Each individual is suffered by one of the two diseases Non Hodgkin Lymphoma (NHL) or Hodgkins Disease (HD), which eventually lead to death. The causes of death are classified into NHL and HD in this study. Three explanatory variables associated with each patient, which are patient's Karnofsky score, the waiting time to transplant recorded in months and the graft type of transplantation. In the present study we considered the graft type of transplant as the covariate. Estimates of β_j; $j=1,2$ are obtained by maximising (2.18) using Newton-Raphson algorithm. One tie is present in the above data, which is broken in random order. Estimated values of β_j; $j=1,2$, using the standard partial likelihood method and the proposed Bayesian method with different sets of prior

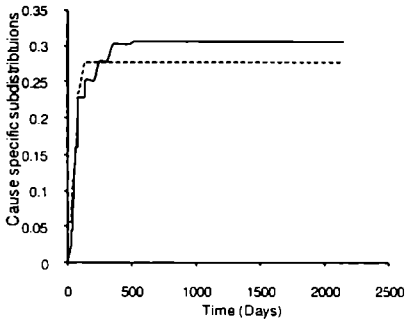
parameter values are given in Table 2.5. Standard Error (S.E) of the estimates are also given in brackets.

Table 2.5 Estimates of $\beta_j \quad j=1,2$ using partial likelihood method and Bayesian methods with different sets of prior parameter values.

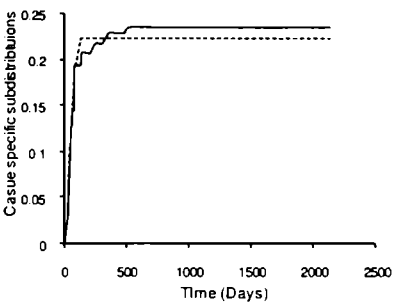
Method	$\hat{\beta}_1$ (S.E of $\hat{\beta}_1$)	$\hat{\beta}_2$ (S.E of $\hat{\beta}_2$)
Partial likelihood	0.01531(0.5626)	-0.3332(0.5863)
$\gamma_1 = 161.5,\ \gamma_2 = 0.61$	0.01551(3.12e-05)	-0.33142(0.00016)
$\gamma_1 = 0.75,\ \gamma_2 = 0.61$	-1.7894(0.00043)	-0.33046(0.00017)
$\gamma_1 = \gamma_2 = 10$	-0.8966(0.00018)	0.2437(3.3e-05)
$\gamma_1 = 50,\ \gamma_2 = 100$	-0.3371(5.30e-05)	0.2994(1.91e-05)
$\gamma_1 = 161.5,\ \gamma_2 = 200$	0.01556(3.11e-05)	0.3453(1.70e-05)
$\gamma_1 = \gamma_2 = 500$	0.2899(2.02e-05)	0.3964(1.52e-05)
$\gamma_1 = \gamma_2 = 1000$	0.4186(1.64e-05)	0.4227(1.44e-05)

From Table 2.5, it can be noted that, for the prior parameter combination $\gamma_1 = 161.5$ and $\gamma_2 = 0.61$, the estimates of β_j for $j=1,2$ given by the proposed likelihood are close to the estimates obtained by the standard partial likelihood given in Lawless (2003). Estimates of β_j are stable over the range $0 < \gamma_1, \gamma_2 < \infty$ for $j=1,2$

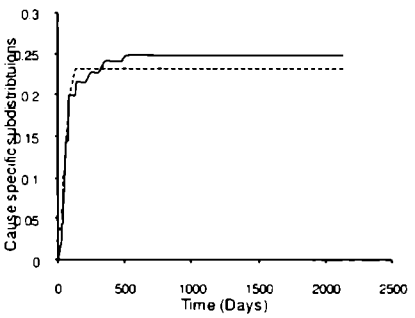
We estimate the cumulative baseline cause specific hazard rate functions using (2.27). Then the cause specific subdistribution functions are also estimated for each set of prior values. Plots of the estimates of cause specific subdistribution functions using the standard partial likelihood method and Bayesian approach with different prior distributions are given in Figure 2.1.



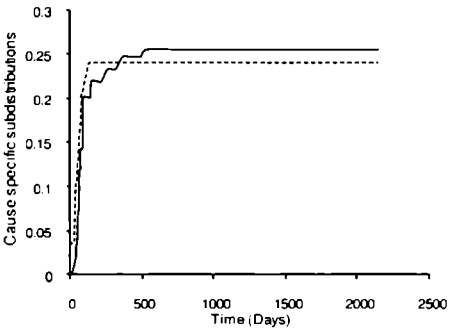
(a) Partial likelihood method



(b) $\gamma_1 = 50,\ \gamma_2 = 100$



(c) $\gamma_1 = 161.5,\ \gamma_2 = 200$



(d) $\gamma_1 = \gamma_2 = 1000$

Figure 2.1 Estimates of cause specific subdistribution functions for patients with NHL and HD.

In Figure 2.1, solid line represents the estimates of subdistribution function due to failures from NHL and dotted line represents the estimates of subdistribution function due to failures from HD. Plots of cause specific subdistribution functions given in Figure 2.1 show that, values of subdistribution function due to failures from HD is smaller than the subdistribution function due to failures from NHL. In both causes, the cause specific subdistribution functions estimated using Bayesian method with different prior parameter values, yield smaller values compared to the same estimated using standard partial likelihood method. When prior parameter values are

near to zero, estimates of the cause specific subdistribution functions yield relatively small values.

2.5 Conclusion

In this chapter, we proposed a semiparametric Bayesian method for the analysis of competing risks data, by treating each cumulative baseline cause specific hazard rate function as a nuisance parameter with a gamma prior. The proposed method generalised the work of Kalbfleisch (1978) to the competing risks set up. The Bayesian approach proposed here can be used for the estimation of $F_j(t)$ which overcomes the limitation of the study of Gelfand et al. (2000). The simulation studies show that the performance of the method is efficient in terms of absolute bias and MSE. The method is applied to a real life data given in Klein and Moeschberger (2003). The proposed method can be extended by modeling the dependence among cause specific hazard rate functions in each interval using Markov Processes.

Chapter Three

A Semiparametric Regression Model for Doubly Censored Competing Risks Data

3.1 Introduction

In many survival studies, patients are only watched within a window of observational time; otherwise we only know that the event time is below or above the window. This pattern of incomplete observations is referred to as double censoring and accordingly the data obtained in such situations are termed as doubly censored data. Consider an example from a sociological study given below. In early childhood learning centers, interest is to determine when a child learns to accomplish certain specified tasks. The age at which a child learns the tasks would be considered as the lifetime. Often some children can already perform the tasks, when they enter into the study. Such lifetimes are considered to be left censored. It is also possible that, some children may learn the tasks during the study period, which are the exactly observed lifetimes while some others may not learn the tasks during the entire study period in which case, the lifetimes would be right censored. An example for a doubly censored survival data is discussed in Example 1.7. For more examples of doubly censored data, one could refer to Gehan (1965), Mantel (1967), Peto (1973) and Turnbull (1974) among others. As mentioned in Cai and Cheng (2004), doubly censored data differs from doubly interval censored data in which the occurrences of both the originating event and the terminating event are either right or interval censored.

The results in this chapter have been communicated as entitled 'A Semiparametric Regression Model for Doubly Censored Competing Risks Data' (See Sankaran and Sreedevi, 2010a).

Analysis of doubly censored survival data is considered by many researchers in literature. Turnbull (1974) proposed the self consistent estimator (SCE) of survivor function for doubly censored data where the concept of SCE was first introduced by Efron (1967). The properties of SCE were studied by many authors including Tsai and Crowley (1985), Chang and Yang (1987), Gill (1989), Chang (1990), Gu and Zhang (1993) and Mykland and Ren (1996). Recently, Cai and Cheng (2004) introduced a semiparametric transformation model to study the effect of covariates on hazard rate functions. For various approaches to the analysis of doubly censored survival data, one could refer to Bravo and Esteban (1993), Wellner and Zhan (1997), Chen and Zhou (2003) and Zhou (2004). The analysis of doubly censored data in the competing risks set up without covariates was recently discussed in Adamic (2008). He developed a self-consistent expectation-maximization algorithm to find the nonparametric estimator of the cumulative incidence functions.

As mentioned in Chapter 1, a standard procedure for the analysis of competing risks data with covariates is to model cause specific hazard rate functions under a proportional hazards assumption (see Prentice et.al, 1978 and Larson, 1984). However, cause specific hazard rate functions do not have a straight forward interpretation in terms of survival probabilities for a particular failure cause. Further, many researchers pointed out that, the effect of a covariate on cause specific hazard rate function of a particular failure type may be very different from the effect of covariate on the corresponding cause specific subdistribution function. Accordingly, Fine and Gray (1999) presented a proportional hazards model for cause specific subdistribution functions based on the $\log(-\log)$ transformation model under right censoring. To estimate the parameters of the transformation model, Fine and Gray (1999) explored the hazard of the subdistribution function, originally given in Gray (1988).

In this chapter, we propose a semiparametric transformation model for the regression analysis of doubly censored competing risks data. The complex nature of the data, as seen in Section 3.2, does not allow us to use the well-known partial

likelihood approach given in Fine and Gray (1999). It may be noted that, the semiparametric regression model given in Cai and Cheng (2004) cannot directly be employed to model cause specific subdistribution functions. Motivated by this, we modify the approach of Cai and Cheng (2004), using the hazard of Gray (1988) to model cause specific subdistribution functions.

In Section 3.2, we describe the model and inference procedures. We estimate the regression parameters and the cumulative baseline cause specific hazard rate functions using estimating equation approach. We, then discuss asymptotic properties of the estimators in Section 3.3. The estimators are shown to be asymptotically unique and consistent. Limiting distributions of the estimators are derived. A simulation study is carried out in Section 3.4 to assess the finite sample performance of the proposed estimators. We apply the proposed method to a real life data in Section 3.5. Finally, Section 3.6 provides brief conclusions of the study.

3. 2 Model and Inference Procedures

Let T be a nonnegative random variable representing the lifetime of an individual with distribution function $F(.)$ and X be the corresponding $p\times 1$ vector of covariates. Let U and V represent the left and right censoring times respectively. We assume that both U and V are always observed and also that U and V are independent of each other. It is also assumed that both the left censoring and right censoring times are independent of the lifetime T Let $C\in\{1,2,...,k\}$ be the cause of failure associated with each observed lifetime. When T is subjected to double censoring, one can only observe the vector $\{\tilde{T},\delta_1,\delta_2,(1-\delta_1\delta_2)C,X\}$,where $\tilde{T}=\max\{U,\min(T,V)\}$, $\delta_1=I(T<U)$, $\delta_2=I(T>V)$ with $I(.)$ as the indicator function. Our objective is to model cause specific subdistribution functions

$F_j(t|X) = P(T \le t, C = j|X)$ based on the data, to study the effect of covariates on lifetime variable.

Gray (1988) introduced the concept of a new hazard rate function in the competing risks set up. Following Gray (1988), the hazard rate function for j th cause is defined as

$$\tilde{\lambda}_j(t|X) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P\{t \le T \le t + \Delta t, C = j | T \ge t \cup (T \le t \cap C \ne j), X\}$$
$$j = 1, 2, .., k \qquad (3.1)$$

In (3.1), the risk set associated with $\tilde{\lambda}_j(t|X)$ at time t, includes all the individuals who has not failed by cause j just prior to time t This follows that, $\tilde{\lambda}_j(t|X)$ can be viewed as the hazard rate function for the improper random variable, $T^* = I(C = j) \times T + \{1 - I(C = j)\} \times \infty$ with distribution $F_j(t|X)$, where $F_j(t|X)$ satisfies $\tilde{\lambda}_j(t|X) = -\frac{d \log(1 - F_j(t|X))}{dt}$, $j = 1, 2, .., k$

Now, we consider the Cox proportional hazards model in competing risks situation using (3.1) as

$$\tilde{\lambda}_j(t|X) = \tilde{\lambda}_{0j}(t) \exp(\beta_j' X) \qquad j = 1, 2, .., k \qquad (3.2)$$

where $\tilde{\lambda}_j(t|X)$ is the cause specific hazard rate function in presence of the covariate vector X $\tilde{\lambda}_{0j}(t)$ is the baseline cause specific hazard rate function which is common to all individuals and β_j is the $p \times 1$ vector of regression parameters for cause j

Now we consider a semiparametric transformation model for the cause specific subdistribution function

$$F_j(t|X) = g(h_{0j}(t) + \beta_j' X) \qquad (3.3)$$

where $g(.)$ is a continuous, strictly increasing link function and $h_{0j}(.)$ is a completely unspecified, invertible and monotone increasing function and β_j is the $p\times 1$ vector of regression parameters for $j=1,2,..,k$

Under a proportional hazards assumption given in (3.2), $g(u)=1-\exp(-\exp(u))$, which provides $h_{oj}(t)=\log(\tilde{\Lambda}_{0j}(t))$, with $\tilde{\Lambda}_{0j}(t)=\int_0^t \tilde{\lambda}_{0j}(u)du$, the cumulative baseline cause specific hazard rate function.

Thus, (3.3) can be written as

$$F_j(t|X)=1-\exp\left[-\left(\tilde{\Lambda}_{0j}(t)\exp(\beta_j' X)\right)\right] \qquad j=1,2,...,k \tag{3.4}$$

Let $\{\tilde{T}_i,\delta_{1i},\delta_{2i},(1-\delta_{1i}\delta_{2i})C_i,X_i\}$ for $i=1,2,...,n$ be n independent and identically distributed copies of $\{\tilde{T},\delta_1,\delta_2,(1-\delta_1\delta_2)C,X\}$. Now, our objective is to estimate β_j and $\tilde{\Lambda}_{oj}(t)$ for $j=1,2,...,k$, using the observed data. We note that

$$P(\tilde{T}_i\le t<V_i;C_i=j|X_i)=P(U_i\le t<V|X_i)P(\tilde{T}_i\le t;C_i=j|X_i)$$
$$j=1,2,...,k\ ;\ i=1,2,...,n. \tag{3.5}$$

Under the model (3.4), we have

$$P(\tilde{T}_i\le t<V_i;C_i=j|X_i)=P(U_i\le t<V|X_i)\left\{1-\exp\left[-\left(\tilde{\Lambda}^*_{oj}(t)\exp(\beta_j^{*'}X_i)\right)\right]\right\}$$
$$j=1,2,...,k\ ;i=1,2,...,n \tag{3.6}$$

where $\tilde{\Lambda}^*_{0j}(t)$ and β_j^* be the true values of $\tilde{\Lambda}_{0j}(t)$ and β_j respectively. This induces the formulation of estimating equation for $\tilde{\Lambda}_{oj}(.)$ at any given β_j as

$$\sum_{i=1}^{n}\left[I(\tilde{T}_i\le t<V_i;C_i=j)-I(U_i\le t<V)\left\{1-\exp\left[-\left(\tilde{\Lambda}_{oj}(t)\exp(\beta_j'X_i)\right)\right]\right\}\right]=0$$
$$(\theta_a\le t\le\theta_b);\ j=1,2,...,k \tag{3.7}$$

where θ_a and θ_b are pre specified constants such that, both $P(\tilde{T}_1<\theta_a)$ and $P(\tilde{T}_1>\theta_b)$ are positive.

Let $\hat{\Lambda}_{0j}\left(t|\beta_j\right)$ be a solution to the equation (3.7). Note that $\hat{\Lambda}_{0j}\left(t|\beta_j\right)$ is a step function in t that arises at the distinct jump points of

$$\left\{I\left(\tilde{T}_i \le t; C_i = j\right); i = 1,2,...,n\right\},\ \left\{I\left(U_i \le t < V_i\right); i = 1,2,...,n,\ j = 1,2,...,k\right\}$$

Now, following the generalized estimating equation method, we obtain the following class of estimating equations for β_j^* as

$$\sum_{i=1}^{n}\int_{\theta_a}^{\theta_b} X_i\left[I\left(\tilde{T}_i \le t < V_i; C_i = j\right) - I\left(U_i \le t < V\right)\Psi\right]d\hat{\eta}(t) = 0 \qquad j = 1,2,...,k \qquad (3.8)$$

with $\Psi = \left(1-\exp\left[-\left(\hat{\Lambda}_{oj}\left(t|\beta_j\right)\exp\left(\beta_j' X_i\right)\right)\right]\right)$ and where $\hat{\eta}(t)$ is an increasing, data dependent weight function, which converges to a deterministic function $\eta(t)$ uniformly in $t \in [\theta_a, \theta_b]$. We choose weight function $\hat{\eta}$ to be counting process of $\left\{\tilde{T}_i\right\}$ in our analysis. We employ Newton –Raphson procedure to solve the equations (3.7) and (3.8) which provides the estimates of the regression parameters and the cumulative baseline cause specific hazard rate functions. Let $\hat{\beta}_j$ denote the root of (3.8) and let $\hat{\Lambda}_{oj}(t) = \hat{\Lambda}_{oj}\left(t|\hat{\beta}_j\right)$. In Section 3.3, we show that, under mild conditions, $\hat{\beta}_j$ and $\hat{\Lambda}_{0j}(t)$ are unique and consistent for large n.

The estimation of $F_j\left(t|X\right)$ is straightforward, following the estimation of β_j and $\tilde{\Lambda}_{0j}$; $j = 1,2,..,k$. The estimator of $F_j\left(t|X\right)$ can be obtained as

$$\hat{F}_j\left(t|X\right) = 1-\exp\left[-\left(\hat{\Lambda}_{oj}(t)\exp\left(\hat{\beta}_j' X\right)\right)\right] \qquad j = 1,2,..,k \qquad (3.9)$$

The consistency of $\hat{\beta}_j$ and $\hat{\Lambda}_{0j}$ ensures the consistency of $\hat{F}_j\left(t|X\right)$. To obtain the distribution of $\hat{F}_j\left(t|X\right)$ at a given covariate level X_0 we prove that, the process

$$v_j\left(t|X_0\right) = n^{1/2}\left(g^{-1}\left\{\hat{F}_j\left(t|X_0\right)\right\} - g^{-1}\left\{F_j\left(t|X_0\right)\right\}\right) \qquad j = 1,2,..,k$$

is asymptotically equivalent to $\tilde{v}_j(t|X_0) = n^{-1/2}\sum_{i=1}^{n}\tilde{v}_j(t|X_{0i})$, where $\tilde{v}_j(t|X_{0i}) = \left\{X_0 - X\left(t|\beta_j^*\right)\right\} A_j^{-1}Q_{ij} + a(t)^{-1}\left[I\left(\tilde{T}_i \le t < V_i; C_i = j\right) - I\left(U_i \le t < V_i\right)\Psi\right]$ with $\Psi = \left(1-\exp\left[-\left(\tilde{\Lambda}_{oj}^*(t)\exp\left(\beta_j^{*'}X_i\right)\right)\right]\right)$ and $a_j(t)$ is the limit of $n^{-1}\sum_{i=1}^{n} I\left(U_i \le t < V_i\right)\dot{g}\left(\tilde{\Lambda}_{0j}^*(t), \beta_j^{*'}X_i\right)$ with $\dot{g}(.)$ as defined in (3.14) and $Q_{ij} = \int_{\theta_a}^{\theta_b}\left(X_i - X\left(t|\beta_j^*\right)\right)*\left[I\left(T_i \le t < V_i; C_i = j\right) - I\left(U_i \le t < V_i\right)\Psi\right]d\eta(t)$

with $\Psi = \left(1-\exp\left[-\left(\tilde{\Lambda}_{oj}^*(t)\exp\left(\beta_j^{*'}X_i\right)\right)\right]\right)$ and A_j as the expectation defined by (3.16). In Section 3.3 we also show that $\tilde{v}_j(t|X_0)$, converges weakly to a zero mean Gaussian process and hence assures the consistency of $\hat{F}_j(t|X)$ for $j = 1, 2, ..., k$

3.3 Asymptotic Properties

In the following, we discuss the uniqueness, consistency and limiting distributions of the proposed estimators.

Theorem 3.1

The estimators $\hat{\beta}_j$ and $\hat{\Lambda}_{oj}(t)$ are unique and are strongly consistent for β_j and $\tilde{\Lambda}_{oj}(t)$ respectively for $j = 1, 2, ..., k$

Proof

To establish this, we follow the approach given in Cai and Cheng (2004). We assume that, the cumulative baseline cause specific hazard rate function $\tilde{\Lambda}_{0j}(t)$ is continuous. By strong law of large numbers for large $n, \Delta \ge 0$, $D_\Delta = \left\{\beta_j : \left\|\beta_j - \beta_j^*\right\| \le \Delta\right\}, t \in [\theta_a, \theta_b]$,

$$\frac{1}{n}\sum_{i=1}^{n}\left[I\left(\tilde{T}_i \le t < V_i; C_i = j\right) - I\left(U_i \le t < V_i\right)\left\{1-\exp\left[-\left(\left(\tilde{\Lambda}^{*}_{oj}(t) - \in\right)\exp\left(\beta_j' X_i\right)\right)\right]\right\}\right] < 0,$$

$$j = 1, 2, .., k \quad (3.10)$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\left[I\left(\tilde{T}_i \le t < V_i; C_i = j\right) - I\left(U_i \le t < V_i\right)\left\{1-\exp\left[-\left(\left(\tilde{\Lambda}^{*}_{oj}(t) + \in\right)\exp\left(\beta_j' X_i\right)\right)\right]\right\}\right] > 0,$$

$$j = 1, 2, .., k \quad (3.11)$$

when $\in$ is sufficiently large, and hence there exists a unique $\hat{\Lambda}_{0j}\left(t|\beta_j\right)$ such that,

$$\sum_{i=1}^{n}\left[I\left(\tilde{T}_i \le t < V_i; C_i = j\right) - I\left(U_i \le t < V_i\right)\left\{1-\exp\left[-\left(\hat{\Lambda}_{oj}\left(t|\beta_j\right)\exp\left(\beta_j' X_i\right)\right)\right]\right\}\right] = 0.$$

$$j = 1, 2, .., k \quad (3.12)$$

By differentiating both sides of (3.12) with respect to β_j, we can have the identity,

$$-\frac{\partial}{\partial \beta_j}\hat{\Lambda}_{0j}\left(t|\beta_j\right) = \bar{X}\left(t|\beta_j\right) = \frac{\sum_{i=1}^{n} I\left(U_i \le t < V_i\right)\dot{g}\left(\hat{\Lambda}_{0j}\left(t|\beta_j\right), \beta_j' X_i\right) X_i}{\sum_{i=1}^{n} I\left(U_i \le t < V_i\right)\dot{g}\left(\hat{\Lambda}_{0j}\left(t|\beta_j\right), \beta_j' X_i\right)}$$

$$j = 1, 2, ..., k \quad (3.13)$$

where

$$\dot{g}\left(\tilde{\Lambda}_{oj}(t), \beta_j' X\right) = \left(\exp\left[-\left(\tilde{\Lambda}_{oj}(t)\exp\left(\beta_j' X\right)\right)\right]\tilde{\Lambda}_{oj}(t)\, X \exp\left(\beta_j' X\right)\right)$$

$$j = 1, 2, ..., k \quad (3.14)$$

To show the existence and uniqueness of $\hat{\beta}_j$, let $Q_j\left(\beta_j\right)$ be the left- hand side of (3.8). It follows from (3.13) that, $n^{-1}\dfrac{\partial Q_j\left(\beta_j\right)}{\partial \beta_j} = -\hat{A}_j\left(\beta_j\right)$; $j = 1, 2, ..., k$, where

$$\hat{A}_j\left(\beta_j\right) = \frac{1}{n}\sum_{i=1}^{n}\int_{\theta_a}^{\theta_b}\left\{X_i - \bar{X}\left(t|\beta_j\right)\right\}^{\otimes 2} I\left(U_i \le t < V_i\right)\dot{g}\left(\hat{\Lambda}_{oj}\left(t|\beta_j\right), \beta_j' X_i\right) d\hat{\eta}(t)$$

$$j = 1, 2, ..., k \quad (3.15)$$

which is non positive definite and for any vector $b, b^{\otimes 0}=1, b^{\otimes 1}=b$ and $b^{\otimes 2}=bb$ Furthermore, since (3.10) and (3.11) hold for any $\epsilon>0$ when and only when $\beta_j=\beta_j^*$, we can have that $\hat{\Lambda}_{0j}\left(t\middle|\beta_j^*\right)\to\Lambda_{0j}(t)$ uniformly in $t\in[\theta_a,\theta_b]$ and $n^{-1}\dfrac{\partial Q_j\left(\beta_j^*\right)}{\partial\beta_j}=-A_j$, where

$$A_j=E\left[\int_{\theta_a}^{\theta_b}\left\{X_i-\bar{X}\left(t\middle|\beta_j^*\right)\right\}^{\otimes 2}\dot{g}\left(\tilde{\Lambda}_{oj}(t),\beta_j^{*'}X_1\right)G\left(t\middle|X_1\right)d\eta(t)\right]$$

$$j=1,2,..,k \quad (3.16)$$

with $G(t|X)=\Pr(U\le t<V|X)$, and $X\left(t\middle|\beta_j\right)$ as the limit of $\bar{X}\left(t\middle|\beta_j\right)$ for $j=1,2,..,k$ When X_i is non-degenerate and A_j is negative definite for all j, since $n^{-1}Q_j\left(\beta_j^*\right)\to 0$, by the standard inverse theorem, there exists a unique solution $\hat{\beta}_j$ to the equation $Q_j\left(\beta_j\right)=0$ in a neighborhood of β_j^* for $j=1,2,..,k$. Thus, coupled with the non positivity of $\hat{A}_j\left(\beta_j\right)$ for large n, ensures the uniqueness of the root of $Q_j\left(\beta_j\right)=0$ in the entire domain of β_j, asymptotically for $j=1,2,..,k$ It also follows that, $\hat{\beta}_j$ is strongly consistent with $\hat{\Lambda}_{0j}\left(t\middle|\hat{\beta}_j\right)\to\tilde{\Lambda}_{0j}^*(t)$ for $j=1,2,..,k$, almost surely uniformly in $t\in[\theta_a,\theta_b]$.

Theorem 3.2

The process $v_j\left(t\middle|X_0\right)$ converges weakly to a zero mean Gaussian process with dispersion matrix $\Sigma_j=n^{-1}A_j^{-1}\left(\sum_{i=1}^{n}Q_{ij}Q_{ij}\right)A_j^{-1}$ for $j=1,2,..,k$ as $n\to\infty$.

Proof

The consistency of $\hat{\beta}_j$ and Taylor series expansion of $Q_j\left(\hat{\beta}_j\right)$ around β_j^* give $n^{1/2}\left(\hat{\beta}_j-\beta_j^*\right)=A_j^{-1}n^{-1/2}Q_j\left(\beta_j^*\right)$ for $j=1,2,..,k$ Taking a Taylor series expansion of $\hat{\Lambda}_{0j}\left(t\middle|\beta_j^*\right)$ around $\tilde{\Lambda}_{0j}^*(t)$ we have,

$$n^{-1/2}Q_j\left(\beta_j^*\right)\simeq n^{-1/2}\sum_{i=1}^{n}\int_{\theta_a}^{\theta_b}\left[X_i-\bar{X}\left(t\middle|\beta_j^*\right)\right]\rho_{ij}\left(t\middle|\tilde{\Lambda}_{0j}^*,\beta_j^*\right)d\eta(t)\quad j=1,2,..,k$$

where

$$\rho_{ij}\left(t\middle|\tilde{\Lambda}_{oj},\beta_j\right)=I\left(\tilde{T}_i\le t<V_i;C_i=j\right)-I\left(U_i\le t<V_i\right)\left\{1-\exp\left[-\left(\tilde{\Lambda}_{oj}(t)\exp\left(\beta_j'X_i\right)\right)\right]\right\}$$

Furthermore, it follows from uniform law of large numbers (Pollard,1990) that $\sup_{t\in[\theta_a,\theta_b]}\left|\bar{X}\left(t\middle|\beta_j^*\right)-X\left(t\middle|\beta_j^*\right)\right|\to 0$, almost surely as $n\to\infty$ for $j=1,2,..,k$ The functional central limit theorem (Pollard, 1990) ensures the weak convergence of $n^{-1/2}\sum_{i=1}^{n}\rho_{ij}\left(t\middle|\tilde{\Lambda}_{0j}^*,\beta_j^*\right)$ for $j=1,2,..,k$ This coupled with the strong representation theorem and the uniform convergence of $\bar{X}\left(t\middle|\beta_j^*\right)$ to $X\left(t\middle|\beta_j^*\right)$ entails that $n^{-1/2}Q_j\left(\beta_j^*\right)$ is asymptotically equivalent to $n^{-1/2}\sum_{i=1}^{n}Q_{ij}$ where

$$Q_{ij}=\int_{\theta_a}^{\theta_b}\left(X_i-X\left(t\middle|\beta_j^*\right)\right)\left[I\left(T_i\le t<V_i;C_i=j\right)-I\left(U_i\le t<V_i\right)\Psi\right]d\eta(t)$$

with $\Psi=\left(1-\exp\left[-\left(\tilde{\Lambda}_{oj}^*(t)\exp\left(\beta_j^{*'}X_i\right)\right)\right]\right)$ for $j=1,2,..,k$ as defined in Section 3.2.

Consider $n^{1/2}\left(\hat{\beta}_j-\beta_j^*\right)\simeq n^{-1/2}A_j^{-1}\sum_{i=1}^{n}Q_{ij}$, where A_j is defined by (3.16), and $X\left(t\middle|\beta_j\right)$ is the limit of $\bar{X}\left(t\middle|\beta_j\right)$ which is also defined by (3.13). From standard

central limit theorem, it follows that, the distribution of $n^{1/2}\left(\hat{\beta}_j - \beta_j^*\right)$ can be approximated by a zero mean normal random vector, with dispersion matrix,

$$\Sigma_j = n^{-1} A_j^{-1} \left(\sum_{i=1}^{n} Q_{ij} Q_{ij} \right) A_j^{-1} \qquad j = 1, 2, .., k$$

An estimator of Σ_j is given by $\hat{\Sigma}_j = n^{-1} \hat{A}_j^{-1}\left(\hat{\beta}_j\right)\left(\sum_{i=1}^{n} \hat{Q}_{ij} \hat{Q}_{ij} \right) \hat{A}_j^{-1}\left(\hat{\beta}_j\right)$ where $\hat{Q}_{ij}$ is obtained by replacing all the theoretical quantities in Q_{ij} with their empirical counterparts and $\hat{A}_j\left(\beta_j\right)$ is defined by (3.16).

To show the asymptotic distribution of,

$v_j\left(t|X_0\right) = n^{-1/2}\left\{\hat{\Lambda}_{0j}\left(t|\hat{\beta}_j\right) - \tilde{\Lambda}_{0j}^*(t) + \left(\hat{\beta}_j - \beta_j^*\right) X_0\right\}$ we take Taylor series expansion of $\hat{\Lambda}_{0j}\left(t|\hat{\beta}_j\right)$ around β_j^* and $\hat{\Lambda}_{0j}\left(t|\beta_j^*\right)$ around $\tilde{\Lambda}_{0j}^*(t)$ and obtain

$$n^{1/2}\left\{\hat{\Lambda}_{0j}\left(t|\hat{\beta}_j\right) - \tilde{\Lambda}_{0j}^*(t)\right\} = n^{-1/2} \sum_{i=1}^{n} \left\{ -X\left(t|\beta_j^*\right) A_j^{-1} Q_{ij} + a_j(t)^{-1} \rho_{ij}\left(t|\tilde{\Lambda}_{0j}^*, \beta_j^*\right)\right\}$$

$$j = 1, 2, .., k$$

It follows that $v_j\left(t|X_0\right)$ is asymptotically equivalent to $\tilde{v}_j\left(t|X_0\right) = n^{-1/2} \sum_{i=1}^{n} \tilde{v}_j\left(t|X_{0i}\right)$ for $j = 1, 2, .., k$. To show the weak convergence of $\tilde{v}_j\left(t|X_0\right)$ to a zero mean Gaussian process, it is enough to show the finite dimensional convergence and tightness of $\tilde{v}_j\left(t|X_0\right)$. It is straight forward to see that, for any finite number of time points, $\{t_1, ..., t_m\}$, the joint distribution of $\left\{\tilde{v}\left(t_1|X_0\right), ..., \tilde{v}\left(t_m|X_0\right)\right\}$ for $j = 1, 2, .., k$ is asymptotically normal with mean zero. Since $a_j(t)$, A_j and $X_0 - X\left(t|\beta_j^*\right)$ for $j = 1, 2, .., k$ are nonrandom, it remains only to show that

$n^{-1/2}\sum_{i=1}^{n} Q_{ij}$ and $n^{-1/2}\sum_{i=1}^{n}\rho_{ij}\left(t\left|\tilde{\Lambda}_{0j},\beta_j^*\right.\right)$ for $j=1,2,..,k$ are tight. Since Q_{ij} does not involve t, $n^{-1/2}\sum_{i=1}^{n} Q_{ij}$ is tight. The tightness of $n^{-1/2}\sum_{i=1}^{n}\rho_{ij}\left(t\left|\tilde{\Lambda}_{0j}^*,\beta_j^*\right.\right)$ for $j=1,2,..,k$ follows from the basic properties of empirical processes (Shorack and Wellner, 1986).

3.4 Simulation Study

We carry out a simulation study to assess the finite sample behavior of the proposed estimators. We consider two causes of failure. For simplicity, we consider a single covariate X which is generated using a Bernoulli random number. The data are generated from the following model. The subdistribution due to failures from cause 1 is given by

$$\Pr\left(T_i \le t; C_i = 1 \middle| X_i\right) = 1-\left[1-m\left(1-\exp(-t)\right)\right]^{\exp\left(\beta_1' X_i\right)}$$

which is a unit exponential mixture with mass $(1-m)$ at ∞ when $X_i = 0$. The subdistribution due to failures from cause 2 is given by

$$\Pr\left(T_i \le t; C_i = 2 \middle| X_i\right) = 1-\left[1-(1-m)\left(1-\exp(-t)\right)\right]^{\exp\left(\beta_2 X_i\right)}$$

Left censoring variable U is generated from uniform distribution over the interval $(0,a)$ and right censoring variable V is generated from $U+Uniform(0,b)$ where a and b are chosen in such a way that 15% of the observations are left censored and 20% of the observations are right censored. We generate random samples of size $n=$ 100 and 250.

Three different combinations of parameter realizations are considered for simulations. 1000 samples are generated for each combination, to calculate the absolute value of the bias and mean square error (MSE) of the estimates. We use values in the neighborhood of zero as initial values for estimating regression parameters. Empirical coverage probability (ECP) for $\hat{\beta}_j$ and $\hat{F}_j\left(t\middle|X\right)$ $j=1,2$ are

also estimated. The absolute bias, MSE and ECP of the estimates of $\beta_j \quad j=1,2$ for different combinations of parameter realizations are given in Table 3.1. As the values of bias and MSE do not vary with m, we present the simulation results for m=0.5. It can be noted that, absolute bias and MSE are negligible for both $\hat{\beta}_1$ and $\hat{\beta}_2$ and decrease as sample size increases. From Table 3.1, it is also clear that, ECP for $\hat{\beta}_1$ and $\hat{\beta}_2$ increase as sample size increases. The absolute value of the bias, MSE and ECP of the estimates of cause specific subdistribution functions are also computed, at five arbitrarily selected time points, which are given in Tables 3.2-3.4. Tables 3.2-3.4 show that, absolute bias and MSE are insignificant for $\hat{F}_j(t|X); j=1,2$ and also that the values decrease with an increase in sample size. The ECP for $\hat{F}_1(t|X)$ and $\hat{F}_2(t|X)$ increase as sample size increases.

Table 3.1 Absolute bias, MSE and ECP of $\beta_j \quad j=1,2$

β_1	β_2	Sample size	Absolute bias $\hat{\beta}_1$	MSE $\hat{\beta}_1$	ECP for $\hat{\beta}_1$	Absolute bias $\hat{\beta}_2$	MSE $\hat{\beta}_2$	ECP for $\hat{\beta}_2$
-1	0.5	100	0.0136	0.1449	94.5	0.0136	0.1649	90.1
		250	0.0118	0.1394	99.1	0.0127	0.1333	95.2
0.5	-1	100	0.0304	0.1312	94.8	0.0130	0.1396	91.7
		250	0.0188	0.1196	100	0.0114	0.1294	96.2
-0.6	-0.75	100	0.0590	0.1327	94.4	0.0178	0.1163	90.8
		250	0.0143	0.1083	99.8	0.0149	0.0849	95.2

Table 3.2 Absolute bias, MSE and ECP of $\hat{F}_j(t|X)$ $j=1,2$ for β_1=-1, β_2=0.5

Sample size	Time points	Absolute bias $\hat{F}_1(t\|X)$	MSE $\hat{F}_1(t\|X)$	ECP for $\hat{F}_1(t\|X)$	Absolute bias $\hat{F}_2(t\|X)$	MSE $\hat{F}_2(t\|X)$	ECP for $\hat{F}_2(t\|X)$
100	0.1	0.0023	0.0027	94.8	0.0013	0.0008	92.6
	0.25	0.0110	0.0645	93.1	0.0051	0.0178	91.5
	0.5	0.0199	0.1423	93.5	0.0134	0.0981	92.1
	1.0	0.0295	0.1691	92.8	0.0238	0.1562	91.0
	2.0	0.0354	0.1809	91.0	0.0350	0.1983	90.3
250	0.1	0.0012	0.0007	99.3	0.0009	0.0004	98.1
	0.25	0.0093	0.0343	98.9	0.0020	0.0021	98.5
	0.5	0.0162	0.1013	99.4	0.0033	0.0054	97.3
	1.0	0.0156	0.1281	98.7	0.0051	0.0131	96.9
	2.0	0.0172	0.1645	97.1	0.0138	0.0956	95.0

Table 3.3 Absolute bias, MSE and ECP of $\hat{F}_j(t|X)$ $j=1,2$ for β_1=0.5, β_2=-1

Sample size	Time points	Absolute bias $\hat{F}_1(t\|X)$	MSE $\hat{F}_1(t\|X)$	ECP for $\hat{F}_1(t\|X)$	Absolute bias $\hat{F}_2(t\|X)$	MSE $\hat{F}_2(t\|X)$	ECP for $\hat{F}_2(t\|X)$
100	0.1	0.0022	0.0005	95.1	0.0043	0.0019	91.6
	0.25	0.0085	0.0072	94.8	0.0243	0.0158	92.8
	0.5	0.0174	0.0304	94.7	0.0661	0.0456	92.2
	1.0	0.0596	0.1551	93.0	0.0906	0.0876	92.5
	2.0	0.0882	0.1775	92.8	0.0223	0.0089	91.3
250	0.1	0.0009	0.0004	99.9	0.0015	0.0011	99.2
	0.25	0.0013	0.0008	99.1	0.0029	0.0042	98.7
	0.5	0.0011	0.0006	99.4	0.0047	0.1110	98.3
	1.0	0.0012	0.0004	98.6	0.0057	0.0164	97.0
	2.0	0.0019	0.0009	96.7	0.0037	0.0070	96.1

Table 3.4 Absolute bias, MSE and ECP of of $\hat{F}_j(t|X)$ $j=1,2$ for β_1=-0.6, β_2=-0.75

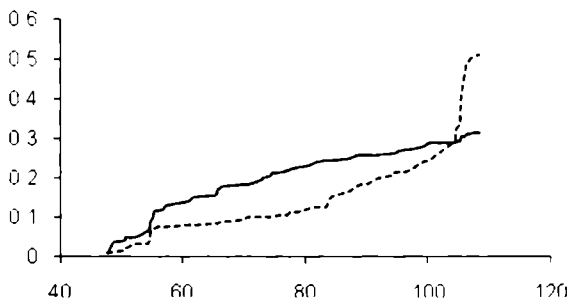
Sample size	Time points	Absolute bias $\hat{F}_1(t\|X)$	MSE $\hat{F}_1(t\|X)$	ECP for $\hat{F}_1(t\|X)$	Absolute bias $\hat{F}_2(t\|X)$	MSE $\hat{F}_2(t\|X)$	ECP for $\hat{F}_2(t\|X)$
100	0.2	0.0017	0.0014	94.1	0.0031	0.0049	91.5
	0.5	0.0107	0.0575	93.8	0.0188	0.1759	90.9
	1.0	0.0295	0.1278	93.9	0.0484	0.1168	92.3
	2.0	0.0548	0.1498	91.9	0.0900	0.1850	91.7
	5.0	0.0815	0.1467	91.0	0.1204	0.1909	90.4
250	0.2	0.0003	0.0001	99.3	0.0005	0.0001	98.8
	0.5	0.0063	0.0198	99.1	0.0095	0.0453	98.6
	1.0	0.0143	0.1020	98.4	0.0211	0.1019	97.9
	2.0	0.0234	0.1248	98.3	0.0388	0.1719	95.9
	5.0	0.0401	0.1343	97.9	0.0535	0.1423	95.1

3.5 Data Analysis

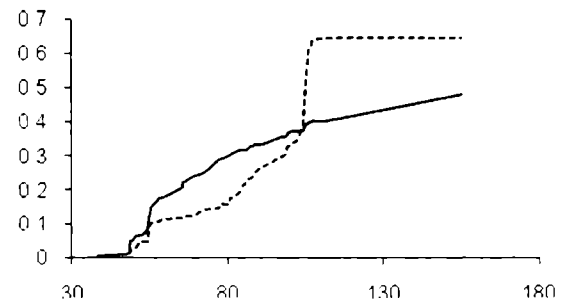
We apply the model to a real data of Nowinski et al. (1979) given in Kalbflescih and Prentice (2002, pp: 390-395). 204 mice were observed to study the genetic and viral factors that may influence the development of spontaneous tumors leukemia in AKR mice. The mice were followed over a 2-year period starting from their birth date, for mortality due to any of the causes and the surviving mice were then sacrificed. We choose a base date and the failures before the base date are considered to be left censored. The possible causes of death were identified as 1: thymic leukemia, 2: non thymic leukemia, 3: non leukemia and no other tumors, 4: unknown causes, 5: other tumors and 6: accidental death. Since thymic leukemia seems to the major cause of mortality in AKR mice in the above data set, we consider the data as a two risks problem with the two causes of death viz. thymic leukemia and other causes. We choose the variable sex as covariate in our analysis.

We find the initial estimates of cumulative baseline cause specific hazard rate functions by ignoring left censored observations as suggested in Klein and

Moeschberger (2003) for the estimation of the survivor function. These estimates are substituted in (3.8) to calculate initial estimates of $\beta_j; j=1,2$ Then the estimates of β_j values are substituted in (3.7) to solve for $\Lambda_{0j}; j=1,2$. The process is repeated until the estimates converge. We get the estimates as $\hat{\beta}_1$=0.31345 and $\hat{\beta}_2$=0.36119 for the above data set. The estimates of $\hat{F}_j(t|X); j=1,2$ are computed using (3.9). The plots of cause specific subdistribution functions for male mice and female mice are given in Figure 3.1(a)-(b). In Figures 3.1 (a)-(b), solid line represents the subdistribution function due to failures from thymic leukemia and dotted line represents the subdistribution function due to failures from other causes.



(a) Male mice

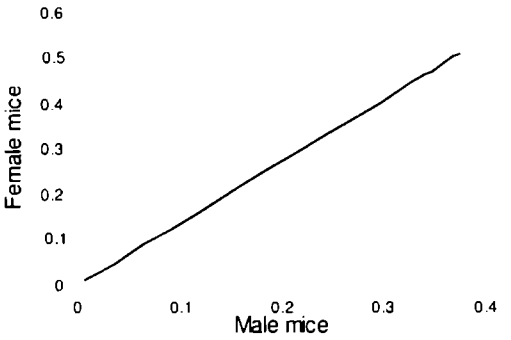


(b) Female mice

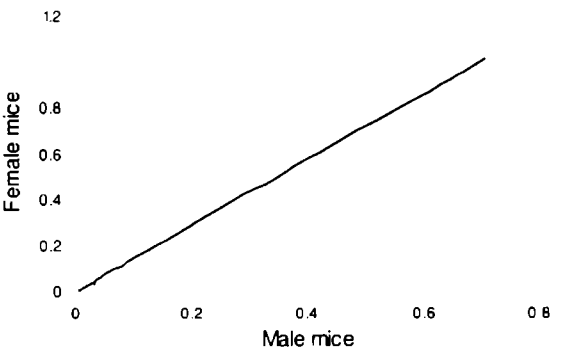
Figure 3.1(a)-(b) Plots of cause specific subdistribution functions for male mice and female mice

Figure 3.1(a)-(b) shows that female mice have a high probability of remission comparing to male mice.

To check the proportionality assumption of hazard rates for each cause of the model (3.2) for the above data set, we use a graphical method, known as Andersen (1982) plots. The covariate sex is discrete and takes only two values. Let $\hat{\Lambda}_{rj0}(t)$ be the estimate of the cumulative baseline cause specific hazard rate for r th stratum, $r = 1,2\,;\, j = 1,2$. Here, we plot $\hat{\Lambda}_{2j0}(t)$ versus $\hat{\Lambda}_{1j0}(t)$ for $j = 1,2$ If the proportional hazards model holds, the curves should be straight lines through the origin. Figures 3.2 (a)-(b) give the Andersen plots for failures due to thymic leukemia and other causes respectively.



(a) Thymic leukemia



(b) Other causes

Figure 3.2 (a)-(b) Andersen plots for the two causes

From Figure 3.2(a)-(b), it follows that the proportionality assumption holds well for both causes.

3.6 Conclusion

In this chapter, we proposed a semiparametric transformation model for the analysis of doubly censored competing risks data. The present study extended the work of Fine and Gray (1999) into the doubly censored set up. The model proposed by Fine and Gray (1999) assumed that the covariate has common effect on cause specific hazard rates, but in the proposed model we assumed that covariate has different effect on different causes. The present study also extended the work of Cai and Cheng (2004) into competing risks set up. The proposed method can be applied to other semiparametric transformation models by considering different choices of $g(.)$ and $h(.)$. The method can also be directly extended to the time dependent covariate set up.

Chapter Four

Neural Network Models for Competing Risks Data

4.1 Introduction

In survival studies, nonparametric methods are very popular due to the fact that the lifetime data may not always meet parametric model assumptions. Gasbarra and Karia (2000), Crowder (2001), Kvam and Singh (2001) and Lawless (2003) provide comprehensive reviews on various nonparametric methods employed in survival analysis. Recently, researchers paid attention on neural network models for the analysis of survival data. Multilayer Perceptrons (MLP) are the most commonly employed neural network models for prediction and classification problems in survival analysis. The main advantage of neural network models is that, the inherent nonlinear structure of the network provides a platform to deal with complex input-output relationships. Further, neural network models need not make any assumptions about the distribution of data or about the relationship of covariates with the survival time. Neural network models use an algorithmic approach to solve complexities beyond the reach of empirical statistical methods.

Another significant advantage in using neural network models is that, the network provides smoothed estimates of cumulative hazard rate functions or survivor function without using any additional smoothing techniques. There are many nonparametric smoothing techniques viz. spline smoothing, kernel density smoothing, and additive smoothing are available in literature for the analysis of the survival data. However, smoothing techniques like kernel density smoothing may result computational issues. For example, in a kernel smoothing method, the selection of

---

The results in this chapter have been communicated as entitled 'Neural Network Models for Competing Risks Data' (see Sankaran and Sreedevi, 2010b).

kernel function and band width is an important issue, since the wrong choice of kernel function and band width may result in wrong conclusions (Wells, 1994). As the neural network models inherently smooth the estimates, it resolves such issues of wrong selection of methods and models.

Multilayer perceptron neural network models were employed in survival studies to improve the estimates of the survivor function (see Bakker and Heskes, 1999 and Bakker et al., 2004). Ambrogi et al. (2007) discussed the role of neural network models with genetic algorithms in the analysis of survival data. One can refer to MacKay (1992), Faraggi and Simon (1995), Bishop (1996), Machado (1997), Neal (1996), Ripley (1998), Neal (2001), Biganzoli et al. (2002) and Ahmed (2005) among many others for various applications of neural network models in survival analysis.

Recently, Biganzoli et al. (2006) employed partial logistic artificial neural networks to model competing risks data in the discrete set up. Later, Lisboa et al. (2009) employed automatic relevance determination technique to regularize this model, which is a commonly employed regularization technique in Bayesian modeling. However, neural network models are less explored for modeling and analysis of competing risks data when the lifetime variable is continuous. Motivated by this, in this chapter, we present neural network models for prediction and classification problems in the analysis of competing risks data. The novelty of our models is that, causes of failure are treated as an input variable, which allows a straightforward accommodation of censoring times. The estimates of the cumulative cause specific hazard rate functions, cause specific subdistribution functions and survivor function are compared with the estimates obtained using nonparametric kernel density smoothing method of Wells (1994).

The chapter is organized as follows. We, in Section 4.2, discuss neural network models for the estimation of cumulative cause specific hazard rate functions, cause specific subdistribution functions and survivor function. We describe the specifications of the proposed multilayer perceptron neural network models in Section

4.2.1. The models are illustrated with two real data examples in Section 4.2.2. Two different classification problems in competing risks set up are addressed in Section 4.3. A binary network to classify individuals into two groups of survived and relapsed patients at a pre specified time point is presented in Section 4.3.1. In Section 4.3.2, a softmax neural network is presented to classify individuals according to their cause of failure. Neural network models developed for classification problems are well demonstrated with a real life example. Finally, Section 4.4 summarizes major conclusions of the study.

4.2 Estimation Problems

In this section, we present multilayer perceptron neural network models for the estimation of basic quantities of the competing risks models. As a first step, we only use the simplest possible network architecture. The extension of our models to handle complex situations is straightforward. From the existing neural network models, our models distinguished by the inclusion of causes of failure as input variables. Further, this input variable is also used for estimation. One advantage of including the cause of failure as an input variable is that, by denoting the censored observations with a cause labeled '0', we can easily include the censored observations as input varaibles.

4.2.1 Neural Network Models

We first introduce multilayer perceptron neural network models for the estimation of cumulative cause specific hazard rate functions and cause specific subdistribution functions of competing risks models in absence of covariates. We present multiple time point models, which predict the desired output at each specified time point. The data is divided into training and validation sets. The training set consists of 80% of the total data points which are used for preliminary model fitting and the validation set contains the remaining 20% of the data which are used to assess the adequacy of the model. We choose negative of the log likelihood function as the objective function to train the networks. Further, a standard back propagation learning

algorithm is employed with quasi-Newton optimisation technique. We use normal error functions in our models. Weight decay procedure is used as the complexity regularisation technique, with weight decay constant 1.

Time to failure and cause of failure are given as input variables for the estimation of cumulative cause specific hazard rate functions and cause specific subdistribution functions. The network is fully connected with one neuron in the hidden layer. Activation functions are selected according to the nature of target variables. For the estimation of cumulative cause specific hazard rate functions, we use exponential activation functions. For estimating cause specific subdistribution functions, logistic activation functions are used. Estimates of cumulative cause specific hazard rate functions and cause specific subdistribution functions given by the nonparametric methods (Lawless, 2003) are given as target variables.

The output of the models, excluding bias terms is given by

$$a_o\left(\sum_h w_{ho} a_h\left(\sum_j w_{jh} x_j\right)\right) \tag{4.1}$$

where j ranges over the inputs, h ranges over the hidden units o denote the output units. a_o denote the activation function used in output layer and a_h denote the activation function used in the hidden layer.

We modify the above specified network to estimate the survivor function directly in presence of covariates. The explanatory variables along with the survival or censoring time and cause of failure are given as input variables. Hidden layer has two nodes. The first hidden node is connected to the input from explanatory variables and survival or censoring time. The second hidden node is connected to the input from the cause of failure. Estimates of the survivor function obtained using Cox's partial likelihood method are given as the target variables. We use logistic activation functions. The focal advantage is that, our model is assumption free and allows a direct estimation of the survivor function. To estimate cause specific subdistribution functions, along with time to failure and cause of failure, explanatory variables are

also given as input variables. Estimates of cause specific subdistribution functions (Lawless, 2003) are given as target variables. The output given by (4.1) can be modified into the situation with covariates.

The efficiency of the neural network models is measured in terms of the mean square error of the estimates. We compare the estimates given by neural network models with the corresponding smoothed estimates using the kernel density technique given in Wells (1994).

Now, consider a competing risks situation with k causes of failure. The Cox proportional hazards model for competing risks data, in which, the hazard rate function for cause j at time t, in presence of the covariate vector X can be specified by

$$\lambda_j(t|X) = \lambda_{0_j}(t)\exp(\beta_j X) \qquad j = 1,2,...,k \tag{4.2}$$

where $\lambda_{0_j}(t)$ is an arbitrary unspecified baseline cause specific hazard rate function and β_j is the vector of regression parameters for cause j

The observed data consist of $(t_i, \delta_i, \delta_i C_i, X_i)$, $i = 1,2,...,n$ where t_i is the observed lifetime or failure time, δ_i is the censoring indicator and X_i is the corresponding covariate and $C_i \in \{1,2,...,k\}$ is the cause of failure.

Let $\hat{\beta}_j$ be the maximum partial likelihood estimate of β_j. An estimator of the cumulative baseline cause specific hazard rate function (Breslow type) $\hat{\Lambda}_{0nj}(t, \beta_j)$ is given by

$$\hat{\Lambda}_{0nj}(t, \hat{\beta}_j) = \int_0^t \left(\sum_{i=1}^{n} Y_i(u)\exp(\hat{\beta}_j X) \right)^{-1} d\bar{N}_j(u) \tag{4.3}$$

where $\bar{N}_j(t) = \sum_{i=1}^{n} I(Y_i \le t, \delta_{ij} = 1)$ and $\delta_{ij} = I(\delta_i = 1, C_i = j)$.

Let K be a density function on $[-1,1]$, symmetric about zero with unit integral and let the parameters $\{b_n\}$ be a sequence of positive numbers tending to zero. Now, a

smoothed estimate of baseline cause specific hazard rate function is $\lambda_{0nj}(t,\beta_j)$ is given by

$$\hat{\lambda}_{0nj}(t,\hat{\beta}_j)=\int_0^1 \frac{1}{b_n} K\left(\frac{t-u}{b_n}\right) d\hat{\Lambda}_{0nj}(u,\hat{\beta}_j) \qquad j=1,2,...,k \tag{4.4}$$

where $\hat{\Lambda}_{0nj}(t,\hat{\beta}_j)$ is given by (4.3). In absence of covariates we can write (4.4) as

$$\hat{\lambda}_{0nj}(t)=\int_0^1 \frac{1}{b_n} K\left(\frac{t-u}{b_n}\right) d\hat{\Lambda}_{0nj}(u) \qquad j=1,2,...,k \tag{4.5}$$

with $\hat{\Lambda}_{0nj}(t)$ as the Nelson-Aalen estimate of cumulative cause specific hazard rate function.

Following the estimation of (4.4), the smoothed estimator of survivor function for T given X is given by

$$\hat{S}(t|X)=\exp\left(-\sum_{j=1}^{k}\hat{\Lambda}_{0nj}(t)\exp\left(\hat{\beta}_j' X\right)\right). \tag{4.6}$$

Thus, (4.6) leads to the smoothed estimator of cause specific subdistribution functions as

$$\hat{F}_j(t|X)=\sum_{i:t_i\le t}\delta_{ij}\hat{S}(t|X)\frac{\exp\left(\hat{\beta}_j' X\right)}{\sum_{l=1}^{n} Y_l(t_i)\exp\left(\hat{\beta}_j' X_l\right)} \qquad j=1,2,...,k \tag{4.7}$$

Remark 4.1: When $j=1$, the estimator given by (4.4) reduces to the smoothed estimator of baseline hazard rate function proposed by Wells (1994).

Remark 4.2: Asymptotic properties of the estimators follow by extending the results of Wells (1994) under certain assumptions.

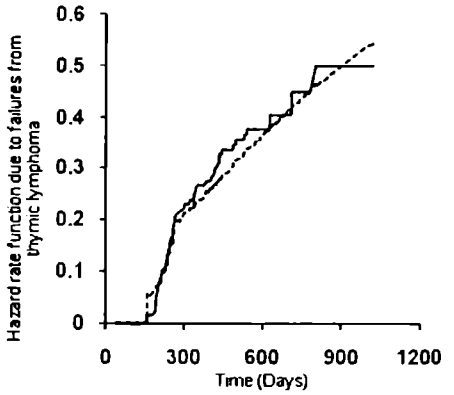
4.2.2 Data Analysis

In this section, we illustrate the models given in Section 4.2.1 with two real life data sets.

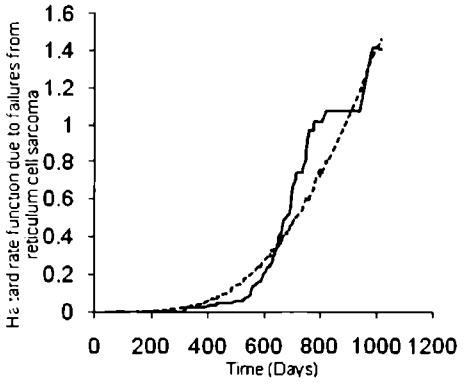
We, first consider the mortality data on RFM strain male mice given by Hoel (1972) to illustrate the estimation of cumulative cause specific hazard rate functions and cause specific subdistribution functions. Mice were divided into two groups and one group of mice lived in a conventional lab environment and the second group was kept in a germ-free environment. There are three causes of death viz. thymic lymphoma, reticulum cell sarcoma and other causes. The mice all died by the end of experiment, so there are no censoring. The data set contains 181 time points. Mean square error (MSE) of the neural network models for estimating cumulative cause specific hazard rate functions and cause specific subdistribution functions are given in Table 4.1. Figures 4.1(a)-(c) compare the smoothed estimates of cumulative cause specific hazard rate functions obtained using nonparametric methods (Lawless, 2003) and neural network models. To employ the kernel smoothing method, we consider $K(x) = 1 - |x|$ and b_n is chosen in such a way that the mean square error of the estimates is minimum. The estimates of cause specific subdistribution functions using the two approaches are given in Figures 4.2(a)-(c). In Figures 4.1(a)-(c) and 4.2(a)-(c) dark line represents the smoothed estimates of Wells (1994) and dotted line represents the corresponding neural network estimates.

Table 4.1 MSE of the neural network models for estimating $\hat{\Lambda}_j(t)$ and $\hat{F}_j(t)$ for mice mortality data, $j = 1, 2, 3$.

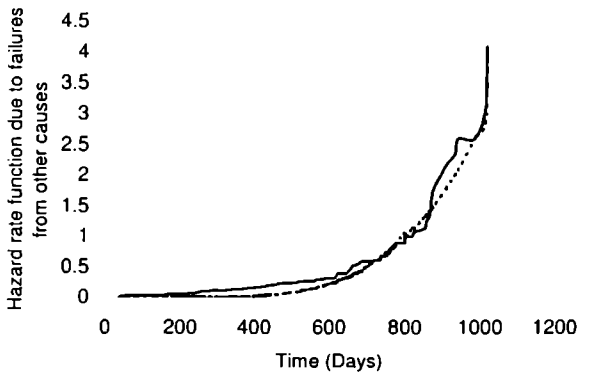
Estimate	MSE
$\hat{\Lambda}_1(t)$	0.00493
$\hat{\Lambda}_2(t)$	0.00266
$\hat{\Lambda}_3(t)$	0.00112
$\hat{F}_1(t)$	0.00347
$\hat{F}_2(t)$	0.00329
$\hat{F}_3(t)$	0.00209



(a)



(b)



(c)

Figure 4.1 (a)-(c) Plots of the estimates of cumulative cause specific hazard rate functions for mice died of thymic lymphoma, reticulum cell sarcoma and other causes

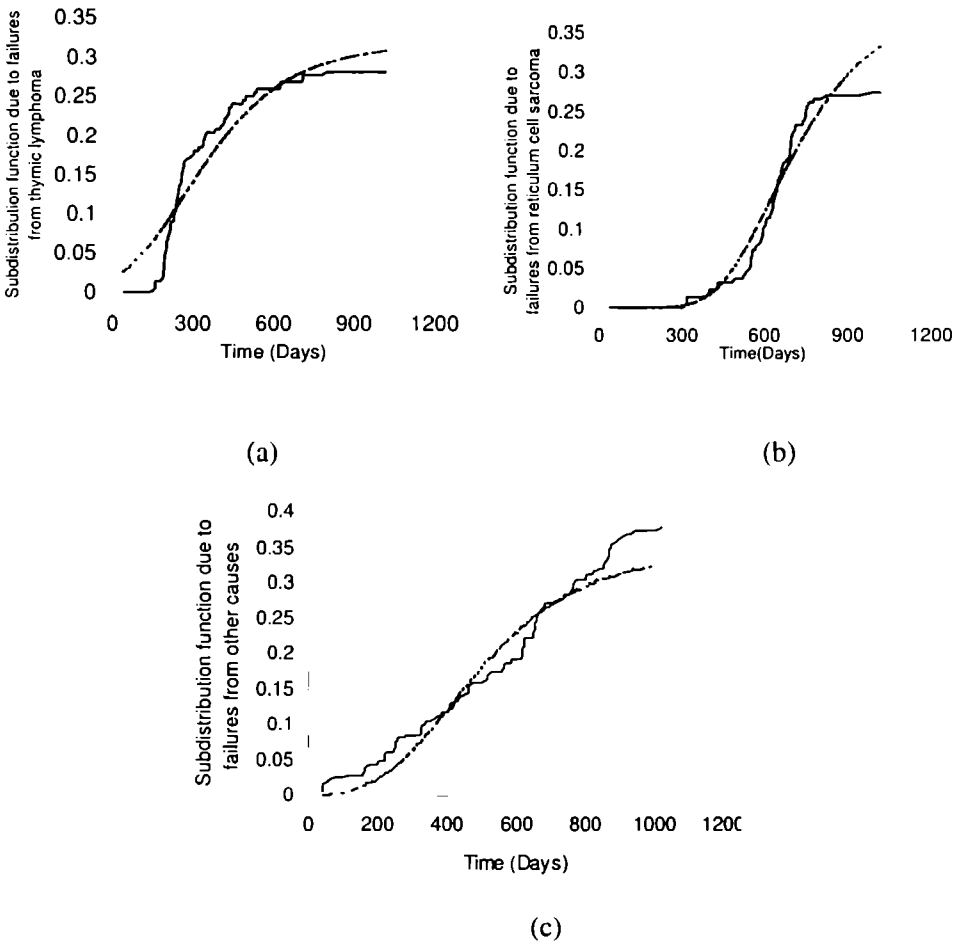


Figure 4.2 (a)-(c) Plots of the estimates of cause specific subdistribution functions for mice died of thymic lymphoma, reticulum cell sarcoma and other causes

From Figures 4.1(a)-(c) and 4.2(a)-(c), it is clear that, neural network models directly give smoothed estimates of both cumulative cause specific hazard rate functions and cause specific subdistribution functions without using any further smoothing techniques.

Now, we consider a competing risks data with covariates given in Andersen et al. (1993, p:709) for the illustration. The data consist of survival times of 202 melanoma patients with cause of death and covariates viz. age, sex, an indicator variable of patient's condition, year of operation, tumor thickness and ulceration. We consider the three covariates namely, age, sex and tumor size of the patient for illustration. The covariate age is in years, sex (1-man, 0- woman) and survival time in days. $C=1$ (death from malignant melanoma); $C=2$ (death from any other causes); $C=3$ (alive on 1, Jan 1978).

Mean square error (MSE) given by neural network model for estimating survivor function and cause specific subdistribution functions for the melanoma data set is given in Table 4.2. Figure 4.3 compares the estimates obtained using two approaches for survivor function. To employ the kernel smoothing method, we consider $K(x)=1-|x|$ and b_n is chosen in such a way that the mean square error of the estimates is minimum. We use (4.6) and (4.7) respectively to obtain the smoothed estimates of survivor function and cause specific hazard rate functions. Figures 4.4(a)-(b) plot the smoothed nonparametric estimates of cause specific subdistribution functions given by (4.7) and the corresponding neural network estimates. In Figures 4.3 and 4.4(a)-(b) dark line represents smoothed estimates of Wells (1994) and dotted line represents the corresponding neural network estimates.

Table 4.2 MSE of the neural network models for estimating $\hat{S}(t)$ and $\hat{F}_j(t)$ for the melanoma data set, $j=1,2$.

Estimate	MSE
$\hat{S}(t)$	0.00234
$\hat{F}_1(t)$	0.00057
$\hat{F}_2(t)$	0.00158

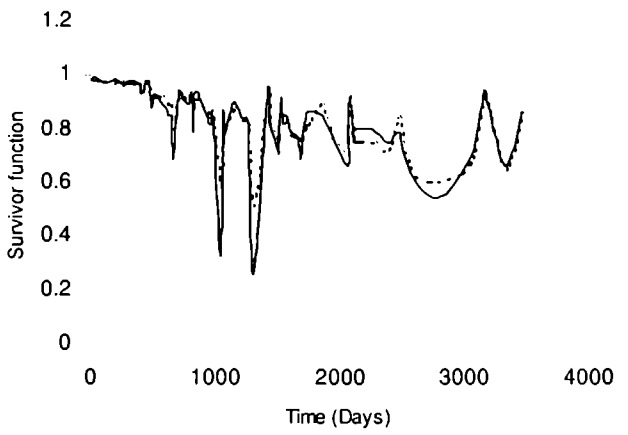
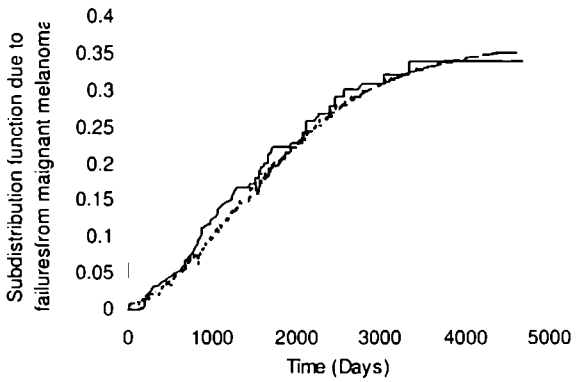
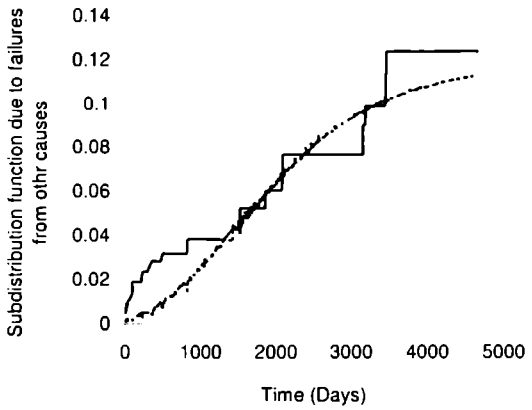


Figure 4.3 Plot of the estimates of survivor function for melanoma data

It can be noted that Figure 4.3 is not monotone, since the difference in covariate levels affect the estimates of the survivor function. Figure 4.3 shows that the absolute difference between two estimates of the survivor function exceeds 0.1 only for three data points. The maximum difference observed is 0.2578. This corresponds to the value 17.42, for the covariate tumor thickness, where as the average value of tumor thickness for patients is 2.906. Considering the other two data points, for which the absolute difference between estimates exceeds 0.1, the value of covariate variable tumor thickness exceeds 10. But, for those data points which deviate extremely from the average value of the input variable age, the estimated values do not show much deviation. It follows that, tumor thickness has significant effect on the estimates of survivor function.



(a)



(b)

Figure 4.4 (a)-(b) Plots of the estimates of cause specific subdistribution functions for melanoma data

4.3 Classification Problems

In this section, we present neural network models for two different classification problems in competing risks set up. We consider the situations in presence of covariates.

4.3.1 Binary Model Network

We present a neural network model to estimate the survivor probability within a pre specified time point and hence to classify the individuals into two groups at any specified time point according to the relapse or survival of the patient at that particular time point. We modify the data by splitting the time into two periods, before the pre specified time point and after the pre specified time point. This model is a standard classification neural network. This model is an extension of the binary model introduced by Ripley et al. (2004) into the competing risks set up.

Let $\tilde{T}$ denote the pre specified time point. Assume that p_{ij} be the probability of relapse for i th patient due to cause j before $\tilde{T}$ and δ_{ij} be the indicator variable which equals 1 if i th patient is relapsed due to cause j before $\tilde{T}$ and 0 otherwise.

Now the likelihood function of the observed data is given by

$$L = \prod_{i=1}^{n} \prod_{j=1}^{k} p_{ij}^{\delta_{ij}} (1 - p_{i.})^{1-\delta_i}$$

where $p_{i.} = \sum_{j=1}^{k} p_{ij}$ and $\delta_{i.} = \sum_{j=1}^{k} \delta_{ij}$ with $0 \le p_{ij} \le 1$ and $0 \le \sum_{j=1}^{k} p_{ij} \le 1$

To incorporate censorship, we follow Ripley et al. (2004). Each censored patient is included twice in the data, with indicator 1 and 0 with appropriate weights. We used multilayer perceptron network with one hidden layer and logistic activation functions.

4.3.2 Softmax Neural Network

In the following, we present a softmax neural network to classify the individuals into different groups according to their causes of death. The individual's survival time or censoring time and the explanatory variables are given as inputs. Censored individuals are considered as a separate class. Target variable is the indicator function denoting the class membership. The model is fitted with cross

entropy error function and softmax activation function. With a softmax activation function, the probability of the membership for class j is given by

$$\frac{\exp(h_j)}{\sum_{j=1}^{k}\exp(h_j)}$$

where h_j is the output from the previous unit.

For illustration of classification problems, we use the same mice mortality data due to Hoel (1972). Living environment of the mice is selected as covariate. The accuracy of the classification is assessed in terms of sensitivity and specificity. Sensitivity is the proportion of event responses that are predicted to be events and specificity is the proportion of non-event responses that are predicted to be non-events.

The time point is selected as 550 days and we run the binary model to predict the survival before 550 days for three observed causes. The sensitivity and specificity of the model is calculated using a 0-1 loss function. The results of the classification in terms of sensitivity and specificity are given in Table 4.3. Both models seem to yield best results regarding classification.

Table 4.3 Sensitivity and Specificity based on a 0-1 loss function for the binary model

Cause	Sensitivity	Specificity
Thymic lymphoma	94.1	45.9
Reticulum cell sarcoma	93.2	44.6
Other causes	95.4	49.0

To assess the accuracy of classification using softmax neural network models, we calculate the specificity and sensitivity based on a 0-1 loss function. The results are given in Table 4.4.

Table 4.4 Sensitivity and Specificity based on a 0-1 loss function for softmax neural networks

Cause	Sensitivity	Specificity
Thymic lymphoma	87.7	75.8
Reticulum cell sarcoma	82.3	73.4
Other causes	91.2	78.7

4.4 Conclusion

In this chapter, we developed neural network models for various estimation and classification problems in the analysis of competing risks data. We developed multiple time point neural network models to estimate cumulative cause specific hazard rate functions, cause specific subdistribution functions and survivor functions. When covariates are present, we introduced a multilayer perceptron neural network model for the direct estimation of survivor probability. We extended the smoothing procedure proposed by Wells (1994) into the competing risks set up and compare the smoothed estimates with the same given by neural network models. It has been shown that neural network models give the smoothed estimates inherently without using any other smoothing techniques. Another important advantage of the neural network models is that, they are free from any assumption about the distribution of data. The flexibility that the neural network models offer for modeling the data is also an additional advantage.

We, further proposed a binary model to classify individuals into two groups of survived and relapsed patients at a pre specified time point. Classification of failure times according to the cause of failure is important in testing the independence of time of failure and cause of failure. A softmax neural network model was also employed to classify the individuals according to their cause of failure. The heuristic and algorithmic approaches help us to solve complexities in the real life situation that are beyond the reach of empirical statistical methods.

Chapter Five

Tests for Independence of Time to Failure and Cause of Failure

5.1 Introduction

In the analysis of competing risks data, decision making problems arise quite often. In competing risks set up, the important problems of hypothesis testing are to test whether any one of the risks acts more seriously than the other risks and to test whether the time to failure and cause of failure are independent or not.

In Chapter 1, we note that latent failure time approach for modeling the competing risks data is not recommended except in special types of applications where the unobserved potential failure times can be given a clear meaning. Deshpande (1990), Aras and Deshpande (1992) and many others employed an approach based on the observable random pair (T, C) for modeling competing risks data as an alternative to the latent failure time approach. The nature of dependence of T and C is vital and important in such modeling. This motivated researchers to develop tests for independence of T and C The problem of testing the independence of time to failure and cause of failure can be represented by the null hypothesis

$$H_0 : \lambda_j(t) = \pi_j \lambda_j(t) \text{ for } j = 1, 2, ..., k \text{ and for all } t$$

Some results of this chapter have been entitled and communicated as A Class of General Tests for Testing Independence of Failure Time and Cause of Failure in a Competing Risks Model' (see Sankaran, Dewan and Sreedevi, 2010) and some other results are entitled and communicated as 'On Testing Independence of Failure Time and Cause of Failure in a Competing Risks Model for Grouped Data (see Dewan, Sankaran and Sreedevi, 2010).

against the alternative

$$H_1 : \lambda_j(t) \neq \pi_j \lambda_j(t) \text{ at least for one } j \text{ and for some } t$$

where $\pi_j = P(C = j)$ for $j = 1, 2, ..., k$, represents the probability of a failure due to cause j

Kochar and Proschan (1991) considered the testing problem for a multiple dependent competing risks model when the lifetime data are continuous. A class of restricted tests for testing independence of T and C was derived by Dykstra et al. (1998). Dewan et al. (2004) developed a test procedure using conditional probabilities for testing independence against specific alternatives. When the risks are dependent, Gasbarra et al. (2006) proposed a class of tests using the crude hazard rate functions and conditional probabilities. Testing independence of T and C when causes of failure are missing was discussed by Dewan and Kulathinal (2007). Most of these tests were based on U -statistics.

In survival studies, categorical (grouped) failure grouped time data may arise from a life test. For example, Mendenhall and Hader (1958) discussed the grouped data on failure times of radio transmitter receivers with two risks of failure. When the lifetime data are categorical, the exact time to failure of an individual may not be known, but we know only the interval in which it lies. Censoring within the intervals may also be present (see Byar and Green, 1980). Chiang (1968) studied competing risks problem when the lifetimes are categorical and derived the relationship between net, crude and partial crude probabilities. David and Moeschberger (1978) have considered the parametric analysis of categorical competing risks data. Dykstra et al. (1995) considered likelihood based inference for cause specific hazard rate functions under order restrictions for categorical lifetime data. Test for independence of time to failure and cause of failure is vital in the analysis of categorical lifetime data also, which has not been developed yet.

Motivated by this, in this chapter, we develop tests for independence of time to failure and cause of failure for both continuous and categorical competing risks data.

Occasionally in competing risks situations, the cause of failure for an individual has not been exactly observed, but has only been narrowed down to a subset of all potential risks. Such situations with incomplete information on the causes of failure are referred to as masking. Recently, a vast amount of research has been carried out in the statistical analysis of competing risks data with masking. For example, one can refer to Kodell and Chen (1987), Lapidus et al. (1994), Dewanji and Sengupta (2003), Craiu and Duchesne (2004), Craiu and Lee (2005), Craiu and Reiser (2006) and Antony and Sankaran (2008) among many others. In the present study, we also consider the testing problem with masked data when the lifetime is categorical in nature.

In Section 5.2, we discuss the tests for independence of time to failure and cause of failure for continuous lifetime data. In Section 5.2.1, we introduce a general class of tests using martingale approach and discuss its properties. A test statistic using likelihood ratio procedure is derived in Section 5.2.2. Asymptotic distribution of the test statistic is also derived. In Section 5.2.3, we carry out simulation studies to assess the finite sample performance of the proposed test statistics. The methods are illustrated using two real life data sets in Section 5.2.4. We, in Section 5.3, consider the testing problem with categorical lifetime data. We consider four different situations with categorical lifetime data. In Section 5.3.1, we derive likelihood ratio test statistics for independence of T and C when all causes of failures are known (unmasked data). We discuss the testing problem when the data are censored in unmasked set up. In Section 5.3.2, we derive likelihood ratio test statistic for the same testing problem, when the data are masked. In the masked data set up, we discuss situations with uncensored and censored lifetimes separately. In Section 5.3.3, we carry out a series of simulation studies to calculate the power of the proposed test statistics. We illustrate our procedures with real data sets in Section 5.3.4. Finally, Section 5.4 summarizes major conclusions of the study.

5.2 Tests for Continuous Lifetime Data

We use two approaches viz. martingale approach and likelihood ratio approach to develop tests for testing independence of time to failure and cause of failure for a continuous lifetime data. The asymptotic distributions of the proposed test statistics are derived.

5.2.1 Martingale Approach

This approach is based on cause specific hazard rate functions which are the fundamental quantities in competing risks data as they can be estimated on the basis of data on failure time and cause of failure.

Suppose that there are n individuals under study. Let $N_j(t)$ denote the number of transitions from alive to death due to cause j during the time interval $(0,t]$, for $j=1,2,...,k$ We now consider the situation where the lifetime variable T is right censored by the censoring variable Z In practice, one could observe, $Y_i=\min(T_i,Z_i)$ and $\delta_{ij}=I(Y_i\le Z_i, C_i=j)$ on n individuals. Note that, C_i is observed only if $Y_i=T_i$ and $I(.)$ is the usual indicator function for $i=1,2,...,n;\ j=1,2,...,k$. In this situation $N_j(t)=\sum_{i=1}^{n} I(Y_i\le t,\delta_{ij}=1)$ and $Y(t)=\sum_{i=1}^{n} I(Y_i>t)$, is the number of individuals who have survived beyond the time t.

The history of the entire process up to time t is represented by $\mathbb{F}_{t-}$, where $\mathbb{F}_{t-}$ represent the σ-field generated by the counting process $\{N_j(t), j=1,2,...,k\}$ The cause specific hazard process is given by $Y(t)\lambda_j(t)$ for $j=1,2,...,k$ and $t\ge 0$ As discussed in Chapter 1, the cumulative cause specific hazard rate functions $\Lambda_j(t)$ are defined as

$$\Lambda_j(t) = \int_0^t \lambda_j(s)\,ds \qquad j = 1,2,...,k$$

and under the assumption that the k failure causes are mutually exclusive and exhaustive, the overall cumulative hazard rate function $\Lambda(t)$ is given by

$$\Lambda(t) = \sum_{j=1}^{k} \Lambda_j(t)$$

If we assume that T and C are independent, $\lambda_j(t)$ can be written as $\lambda_j(t) = \pi_j \lambda(t)$ where $\pi_j = P(C = j)$ is the probability of a failure due to cause j for $j = 1,2....,k$ and $\lambda(t) = \sum_{j=1}^{k} \lambda_j(t)$.

Now, we test the null hypothesis

$$H_0: \ \lambda_j(t) = \pi_j \lambda(t) \text{ for } j = 1,2,...,k \text{ and for all } t \tag{5.1}$$

against the alternative

$$H_1 \quad \lambda_j(t) \neq \pi_j \lambda(t) \quad \text{for at least one } j \text{ and for some } t \tag{5.2}$$

The construction of the test statistic is based on the simple idea of comparing the estimates of both sides of (5.1). The Nelson-Aalen estimator (Andersen et al., 1993) of $\Lambda_j(t)$ and $\Lambda(t)$ are given by

$$\hat{\Lambda}_j(t) = \int_0^t \frac{c(u)\,dN_j(u)}{Y(u)} \qquad j = 1,2,...,k \tag{5.3}$$

and

$$\hat{\Lambda}(t) = \int_0^t \frac{c(u)\,dN(u)}{Y(u)} \tag{5.4}$$

where $c(u) = I\{Y(u) > 0\}$ and $dN(u) = \sum_{j=1}^{k} dN_j(u)$.

The probabilities $\pi_j \quad j = 1,2,...,k$ can be written as

$$\pi_j = F_j(\tau) = \int_0^{\tau} S(u)\, d\Lambda_j(u) \tag{5.5}$$

where $S(t) = P(T > t)$ is the survivor function of T and $\tau = \sup\{t : F(t) \leq 1\}$ Thus the estimator of π_j can be obtained as

$$\hat{\pi}_j = \hat{F}_j(\tau) = \int_0^{\tau} \hat{S}(u)\, d\hat{\Lambda}_j(u) \qquad j = 1,2,...,k \tag{5.6}$$

where $\hat{S}(t)$ is the well known Kaplan-Meier estimator of $S(t)$. Note that $\hat{\pi}_j$ is independent of t, and in the uncensored set up $\hat{\pi}_j$ is nothing but the probability of a failure due to the cause j In case of right censoring, $\sum_{j=1}^{k} \hat{\pi}_j$ need not be equal to one. In such cases, the estimator of π_j is $\tilde{\pi}_j$ which is the normalized version of the estimate. Consider a measure of departure from the null hypothesis as follows

$$Z_j(t) = \int_0^t w_j(u)\left\{d\hat{\Lambda}_j(u) - \hat{\pi}_j d\hat{\Lambda}(u)\right\}$$

$$= \int_0^t c(u) w_j(u)\left\{\frac{dN_j(u)}{Y(u)} - \hat{\pi}_j \frac{dN(u)}{Y(u)}\right\} \qquad j = 1,2,...,k \tag{5.7}$$

where $w_j(t)$ is a locally bounded predictable weight process. By the Doob-Meyer decomposition theorem, we have

$$M_j(t) = N_j(t) - \pi_j \int_0^t \lambda_j(u) Y(u)\, du \qquad j = 1,2,...,k \tag{5.8}$$

are zero mean martingales with respect to the increasing family of σ-fields $\{\mathbb{F}_{t-}, t \geq 0\}$ Taking summation on (5.8), we get

$$M(t) = N(t) - \int_0^t \lambda(u) Y(u)\, du$$

where $M(t)=\sum_{j=1}^{k}M_j(t)$.

Now (5.7) can be written as

$$Z_j(t)=\int_0^t c(u)w_j(u)\left\{\frac{dN_j(u)}{Y(u)}-\pi_j\frac{dN(u)}{Y(u)}+\pi_j\frac{dN(u)}{Y(u)}-\hat{\pi}_j\frac{dN(u)}{Y(u)}\right\}$$

$$=\int_0^t c(u)w_j(u)\left\{\frac{dN_j(u)}{Y(u)}-\pi_j\frac{dN(u)}{Y(u)}\right\}+(\pi_j-\hat{\pi}_j)\int_0^t c(u)w_j(u)\frac{dN(u)}{Y(u)}$$

$$j=1,2,...,k \qquad (5.9)$$

Since $\tilde{\pi}_j$ is a consistent estimator of π_j (Lawless,2003), $(\pi_j-\tilde{\pi}_j)\xrightarrow{p}0$ and hence the second factor of (5.9) vanishes.

Define

$$c(t)w_j(t)=k(t)Y(t) \qquad j=1,2,...,k \qquad (5.10)$$

Then for large n, (5.9) is asymptotically equals to

$$Z_j(t)\simeq\int_0^t k(u)\left(\delta_j-\pi_j dN(u)\right) \qquad j=1,2,...,k \qquad (5.11)$$

where $k(t)$ is a locally bounded predictable process and $\delta_j=I(C=j)$. We note that $\sum_{j=1}^{k}Z_j(t)=0$. Further, Z_j's are locally square integrable martingales. Under H_0, it can be shown from Andersen et al. (1993) that $Z(t)=\left[Z_1(t),...,Z_k(t)\right]^T$ is asymptotically distributed as a k variate normal with mean vector 0 and covariance matrix $\Sigma(t)$, which can be consistently estimated by $W(t)=\left\{W_{jm}(t)\right\}$, where

$$W_{jm}(t)=\int_0^t k^2(u)\left(\Delta_{jm}-\hat{\pi}_m\right)\hat{\pi}_j dN(u) \qquad j,m=1,2,...,k \qquad (5.12)$$

with Δ_{jm} as the Kronecker delta. Thus a test statistic is given by

$$\chi^2 = Z^T(\tau)\bar{W}(\tau)Z(\tau) \tag{5.13}$$

where $\bar{W}(\tau)$ is the generalized inverse of $W(\tau)$. If we delete the last row and last column of $W(t)$, to give say $W_0(t)$, and let $Z_0(t) = \left[Z_1(t),...,Z_{k-1}(t)\right]^T$ (5.13) can be alternatively given as,

$$\chi^2 = Z_0^T(\tau)W_0(\tau)^{-1}Z_0(\tau) \tag{5.14}$$

where $W_0(\tau)^{-1}$ is the ordinary inverse of $W(\tau)$. Under H_0, χ^2 is asymptotically distributed as chi-square with $k-1$ degrees of freedom (see Andersen et al., 1993). Possible choices for weight function $k(t)$ in (5.10) are

a) $k(t) = S(t)$, b) $k(t) = S^2(t)$, c) $k(t) = Y(t)S(t)$ and d) $k(t) = Y(t)$

When $k(t) = Y(t)$, we get

$$Z_j(\infty) = \int_0^\infty Y(u)dN_j(u) - \int_0^\infty Y(u)dN(u), \quad j = 1,2,...,k.$$

Thus, (5.11) is the generalization of the Wilcoxon and Kruskal-Wallis tests to right censored data due to Gehan (1965) and Breslow (1970). The efficiency of these tests obviously depends on the choice of weight function. A possible approach to find the optimal value of $k(t)$ is that to choose $k(t)$ which minimizes the mean square error of the test statistics. However, the exact mean square error of the test statistic is not available. In practice, we choose an optimal weight function using bootstrap procedure, which will be discussed in Section 5.2.4.

5.2.2 Likelihood Ratio Test

Let $N^*(t) = \left\{\left(N_1(t),...,N_k(t)\right): t \in \mathbb{F}_{t-}\right\}$ be a k-variate counting process with intensity process $\lambda^* = \left(\lambda_1^*,...,\lambda_k^*\right)$ Then the likelihood L based on a sample is given by (Andersen et al., 1993)

$$L=\prod_{t\in\mathbb{F}}\left(\prod_{j=1}^{k}\lambda_j^*(t)^{dN_j(t)}\right)\exp\left(-\int_0^{\tau}\sum_{j=1}^{k}\lambda_j^*(t)dt\right) \tag{5.15}$$

where $\prod_{t\in\mathbb{F}}$ represents the product integral.

Since $\lambda_j^*(t)=\lambda_j(t)Y(t)$, under H_0, the likelihood (5.15) becomes

$$L^*=\prod_{t\in\mathbb{F}}\left(\prod_{j=1}^{k}\left(\lambda_j(t)Y(t)\right)^{dN_j(t)}\right)\exp\left(-\int_0^{\tau}\sum_{j=1}^{k}\lambda_j(t)Y(t)dt\right). \tag{5.16}$$

As (5.16) is a semiparametric model, we follow the approach given in El Barmi et al. (2006) to obtain the likelihood ratio test statistic. Differentiating $\log L^*$ with respect to $\lambda_j(t)$ and equating to zero, we get the maximum likelihood estimate of $\lambda_j(t)$ as

$$\hat{\lambda}_j(t)=\frac{c(t)dN_j(t)}{Y(t)} \qquad j=1,2,...,k.$$

Since, under H_0 , $\lambda_j(t)=\pi_j\lambda(t)$, the likelihood function (5.16) becomes

$$L^0=\prod_{t\in\mathbb{F}}\left(\prod_{j=1}^{k}\left(\lambda(t)\pi_jY(t)\right)^{dN_j(t)}\right)\exp\left(-\int_0^{\tau}\sum_{j=1}^{k}\lambda(t)\pi_jY(t)dt\right).$$

Differentiating $\log L^0$ with respect to $\lambda(t)$ and equating to zero, we get the maximum likelihood estimate of $\lambda(t)$ as

$$\hat{\lambda}(t)=\frac{c(t)dN(t)}{Y(t)}$$

Similarly, differentiating $\log L^0$ with respect to π_j and equating to zero, we get the maximum likelihood estimate of π_j $j=1,2,...,k$ as

$$\hat{\pi}_j=\frac{N_j(\tau)}{N(\tau)}$$

where $\tau=\sup\{t:F(t)\le 1\}$

Thus we can obtain the likelihood ratio test statistic for testing (5.1) as

$$Q = \frac{\prod_t \left\{ \prod_{j=1}^{k} \hat{\lambda}(t)^{dN_j(t)} \hat{\pi}_j^{dN_j(t)} \right\}}{\prod_t \left\{ \prod_{j=1}^{k} \hat{\lambda}_j(t)^{dN_j(t)} \right\}} \tag{5.17}$$

where $\hat{\lambda}_j(t), \hat{\pi}_j$ and $\hat{\lambda}(t)$ are the maximum likelihood estimates of $\lambda_j(t), \pi_j$ and $\lambda(t)$ respectively.

In the following, we can establish that the asymptotic distribution of $-2\log Q$ is a chi-square distribution with $k-1$ degrees of freedom.

From (5.17), we can have that

$$-2\log Q = 2\int_0^\tau \sum_{j=1}^{k} \left(\log \hat{\lambda}_j(t) - \log\left(\hat{\pi}_j \hat{\lambda}(t)\right) \right) dN_j(t)$$

$$= 2\int_0^\tau \sum_{j=1}^{k} \left(\log \hat{\lambda}_j(t) - \log\left(\hat{\pi}_j \hat{\lambda}(t)\right) \right) \hat{\lambda}_j(t) Y(t)\, dt \tag{5.18}$$

By Taylor's series expansion, (5.18) can be written as

$$-2\log Q = 2\int_0^\tau \sum_{j=1}^{k} \left(\hat{\lambda}_j(t) - \hat{\pi}_j \hat{\lambda}(t) \right) Y(t)\, dt + \int_0^\tau \sum_{j=1}^{k} \frac{\left(\hat{\lambda}_j(t) - \hat{\pi}_j \hat{\lambda}(t) \right)^2}{\hat{\lambda}_j(t)} Y(t)\, dt + O\left(\frac{1}{n}\right). \tag{5.19}$$

Under H_0, $\sum_{j=1}^{k} \left(\hat{\lambda}_j(t) - \hat{\pi}_j \hat{\lambda}(t) \right) Y(t) = 0$ and thus (5.19) can be asymptotically written as

$$-2\log Q = \int_0^\tau \sum_{j=1}^{k} \frac{\left(\hat{\lambda}_j(t) - \hat{\pi}_j \hat{\lambda}(t) \right)^2}{\hat{\pi}_j \hat{\lambda}(t)} Y(t)\, dt \tag{5.20}$$

From Andersen et al. (1993, p: 403) it is seen that, $-2\log Q$ is a partial score statistic and it follows that $-2\log Q$ has chi-square distribution with $k-1$ degrees of freedom. Then, we reject the null hypothesis when $-2\log Q$ is large.

5.2.3 Simulation Study

We carry out an extensive simulation study to assess the performance of the test statistics. In martingale approach, the four different weight functions (a)-(d) described in Section 5.2.1 are considered. We consider two causes of failure. Lifetimes are generated from exponential distributions with parameters λ_1 and λ_2 which correspond to two different causes. To study the effect of censoring on the proposed test statistics, we consider three different situations viz. no censoring, mild censoring (20% of the observations are censored) and heavy censoring (40 % of the observations are censored). In the censored situations, observations are generated from a uniform random variable over $(0,a)$, where a is chosen in such a way that 20% or 40% of the observations are censored. Random samples of size $n=$ 50,100 and 250 are generated 1000 times. To calculate the empirical type I error, we generate lifetimes from the exponential models with parameter values $\lambda_1=2,\lambda_2=8,\pi_1=0.2$ and $\pi_2=0.8$. Empirical power of the test is calculated by generating lifetimes from the exponential models with parameter values $\lambda_1=8,\lambda_2=2,\pi_1=0.2$ and $\pi_2=0.8$.

To assess the performance of the likelihood ratio test discussed in Section 5.2.2, 1000 random samples of size $n=$ 50,100 and 250 are generated. The exponential distributions with same parameters and proportions described above are used for computing empirical type I error and power of the test statistic.

Tables 5.1 and 5.2 provide the empirical type I error and power, in percent, for different test statistics developed using martingale approach. Table 5.3 presents the empirical type I error and power of the likelihood ratio test.

Table 5.1 Empirical type I error (in percentage) of different test statistics

n	Alpha	Test Statistic	No Censoring	20% Censoring	40% Censoring
50	5	Q_1	4.4	4.4	4.6
50	5	Q_2	5.2	5.1	5.8
50	5	Q_3	4.9	4.9	5.5
50	5	Q_4	5.1	5.6	5.7
50	1	Q_1	1.2	1.2	1.3
50	1	Q_2	1.4	1.5	1.6
50	1	Q_3	1.4	1.4	1.6
50	1	Q_4	1.6	1.8	1.9
100	5	Q_1	4.3	4.1	4.6
100	5	Q_2	5.2	5.1	5.4
100	5	Q_3	4.8	4.8	5.0
100	5	Q_4	5.0	5.0	5.2
100	1	Q_1	1.0	1.0	1.2
100	1	Q_2	1.3	1.3	1.5
100	1	Q_3	0.9	1.2	1.3
100	1	Q_4	1.4	1.6	1.8
250	5	Q_1	3.8	4.1	4.3
250	5	Q_2	5.1	4.9	5.4
250	5	Q_3	4.6	4.5	5.5
250	5	Q_4	4.8	4.9	5.2
250	1	Q_1	0.9	1.0	1.2
250	1	Q_2	0.8	1.1	1.3
250	1	Q_3	0.9	0.9	1.2
250	1	Q_4	1.4	1.4	1.5

From Table 5.1, it is clear that all the four test statistics have type I error, close to the chosen significance level for all values of n. When sample size is large, the effect of censoring on type I error is marginal.

Table 5.2 Empirical power (in percentage) of different test statistics

n	Alpha	Test Statistic	No Censoring	20% Censoring	40% Censoring
50	5	Q_1	50.0	48.9	44.3
50	5	Q_2	51.0	51.9	48.3
50	5	Q_3	52.8	51.6	46.7
50	5	Q_4	34.3	23.9	23.3
50	1	Q_1	12.4	10.2	10.3
50	1	Q_2	48.8	43.1	42.3
50	1	Q_3	49.1	44.7	42.1
50	1	Q_4	14.1	14.3	12.9
100	5	Q_1	70.5	72.1	71.4
100	5	Q_2	76.4	75.4	75.0
100	5	Q_3	79.2	78.2	76.5
100	5	Q_4	44.1	43.0	43.2
100	1	Q_1	21.1	20.2	19.3
100	1	Q_2	66.8	63.2	62.0
100	1	Q_3	68.1	64.3	60.1
100	1	Q_4	30.9	28.2	23.2
250	5	Q_1	98.6	95.6	93.2
250	5	Q_2	98.1	97.4	94.5
250	5	Q_3	98.2	94.3	94.6
250	5	Q_4	98.5	94.1	92.1
250	1	Q_1	87 1	77.2	74.2
250	1	Q_2	88.6	87.8	84.3
250	1	Q_3	89.1	90.1	88.6
250	1	Q_4	87.2	90.6	88.4

Table 5.2 shows that all the proposed test statistics have good power in general. As sample size increases, the power of the four test statistics increases, especially for Q_4 However, Q_2 and Q_3 behaves in similar fashion in most of the simulations, essentially because of the choice of weight functions. The presence of censoring does not provide much difference to the power of the tests.

Table 5.3 Empirical type I error and power (in percentage) of the likelihood ratio test statistic

n	Alpha	No Censoring		20% Censoring		40% Censoring	
		Type I error	Power	Type I error	Power	Type I error	Power
50	5	5.5	95.9	5.4	84.8	5.4	82.3
100	5	4.3	97.1	5.0	96.1	5.1	94.3
250	5	4.0	99.3	3.9	98.8	4.8	96.2
50	1	1.0	77.8	1.0	71.2	1.2	68.5
100	1	0.8	88.5	0.7	80.6	0.9	78.6
250	1	0.6	98.1	0.6	87.1	0.9	85.4

The results of the simulation study using likelihood ratio test statistic show that the test has good power.

5.2.4 Data Analysis

We consider two real life data sets to establish the utility of the proposed test procedures in practical situations. First, we consider a competing risks data given in Hoel (1972). The data obtained from a laboratory experiment on RFM strain male mice, which had received a radiation dose of 300 rads at ages of 5 to 6 weeks. Three causes of death were observed viz. thymic lymphoma, reticulum cell sarcoma and other causes. The data were studied by different authors including Aly et al. (1994), Kochar et al. (2002) and Dewan et al. (2004). The estimates of the cumulative hazard rate functions due to the three different causes are given in Figure 5.1.

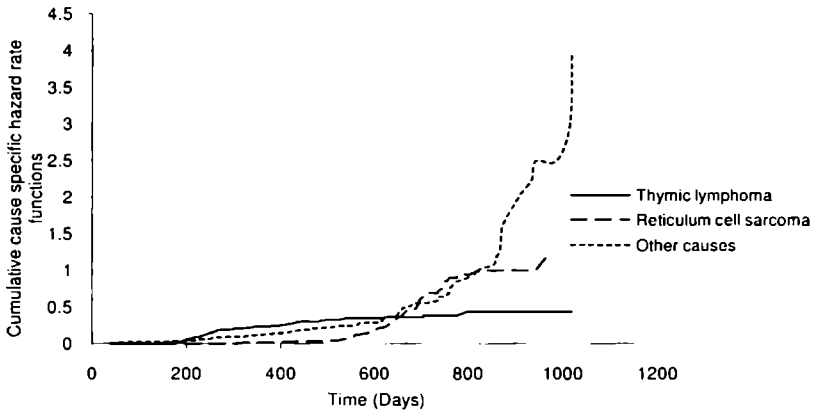


Figure 5.1 The estimates of the cumulative hazard rate functions of death from thymic lymphoma, reticulum cell sarcoma and other causes.

Figure 5.1 shows that, the estimates of cumulative hazard rate functions due to failures from 'other causes' yields large values comparing to the cumulative hazard rate functions due to failures from thymic lymphoma and reticulum cell sarcoma.

Further, we consider the 'other causes' as censored variables and analyze the same data set as a two risks problem. The estimates of the cumulative hazard rate functions due to the two types of cancer thymic lymphoma and reticulum cell sarcoma, when considering 'other causes' as censored variables are given in Figure 5. 2.

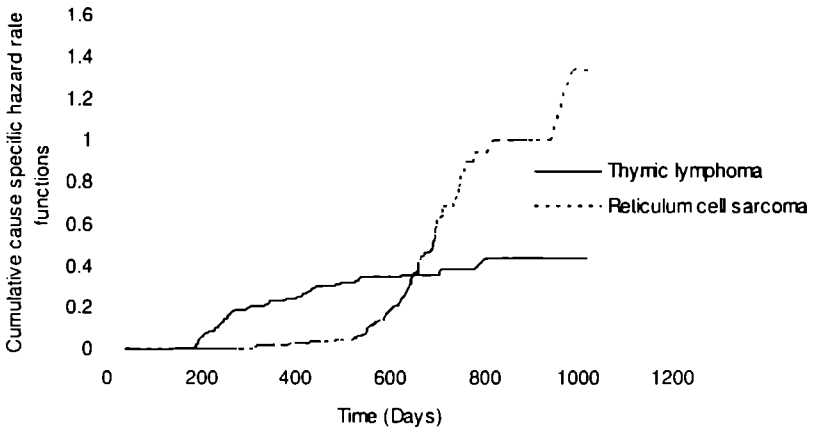


Figure 5.2 The estimates of the cumulative hazard rate functions due to causes thymic lymphoma and reticulum cell sarcoma in the censored case

Figure 5.2 shows that, the estimates of cumulative hazard rate function due to failures from reticulum cell sarcoma possess large values comparing to the cumulative hazard rate functions due to failures from thymic lymphoma.

To apply the test procedure (5.13) in the above two situations, we consider different weight functions (a)-(d) given in Section 5.2.1. Test statistics on these cases were calculated along with their P-values. Let Q_1, Q_2, Q_3, Q_4 and Q^* represent the test statistics obtained using weight functions (a),(b),(c),(d) and the likelihood ratio test respectively. Table 5.4 provides the values of the test statistics along with their P-values.

From Table 5.4, it follows that, the test statistic values are highly significant for all the four weight functions (a)-(d). Likelihood ratio test statistic is highly significant in both censored and uncensored set up. Thus failure time and cause of failure can be assumed to be not independent for Hoel's data in both uncensored and censored situations.

Table 5.4 Test statistic values using different weight functions and likelihood ratio test statistic for mice mortality data

Test statistic	Test statistic value for 3 causes (uncensored data)	P value	Test statistic value for 2 causes (censored data)	P value
Q_1	13.348	<0.01	18.799	< 0.01
Q_2	26.206	<0.01	32.811	< 0.01
Q_3	26.08	<0.01	38.597	< 0.01
Q_4	13.223	<0.01	27.411	<0.01
Q^*	162.597	< 0.01	64.339	<0.01

To find the optimal $k(t)$, we use the bootstrap procedure. We choose optimal $k(t)$, as the value of $k(t)$, which minimizes the bootstrap mean square error estimate for the proposed test statistics. The bootstrap technique for determining optimal $k(t)$ is applied to the real data given above. The bootstrap estimates of the absolute value of the biases and the mean square errors (MSE) of the test statistics Q_1, Q_2, Q_3 and Q_4 are computed from the real data based on 1000 bootstrap samples of size 181, which are given in Table 5.5.

From Table 5.5, it is clear that the estimates of the mean square error based on the bootstrap samples yields the lowest value for the statistic Q_4 in both uncensored and censored set up. Thus we can conclude that the weight function $k(t) = Y(t)$, is the optimal choice of weight function and Q_4 is the best test statistic, if mean square error is the optimality criterion.

Table 5.5 Bootstrap estimates of absolute value of the biases and mean square error of test statistics Q_1, Q_2, Q_3 and Q_4 based on 1000 bootstrap samples for mice mortality data

Test statistic	Bias and MSE without censoring		Bias and MSE with censoring	
Q_1	0.50059	0.97267	1.1213	2.3693
Q_2	0.52775	0.908	0.37406	1.2235
Q_3	0.44099	0.84408	1.07	2.4634
Q_4	0.15795	0.70372	0.27757	0.83607

Now, we consider a censored data from a laboratory test on pneumatic tires given in Davis and Lawrence (1989). The test involves rotating the tires against a steel drum until some type of failure occurred. Failures were classified into 6 modes or categories. 1-open joint on the inner linear; 2-rubber chunking on the shoulder; 3-loose casing low on the side wall; 4-cracking of the tread rubber; 5-cracking on the side wall; 6-any other causes. C=0 denotes that the tire did not fail under test, so that failure time is censored. Figure 5.3 plots, the estimates of cumulative cause specific hazard rate functions for pneumatic tire data set.

From Figure 5.3, it can be noted that the cumulative hazard rate function due to failures from cause 4 is significantly different from the cumulative hazard rate functions due to failures from all other causes.

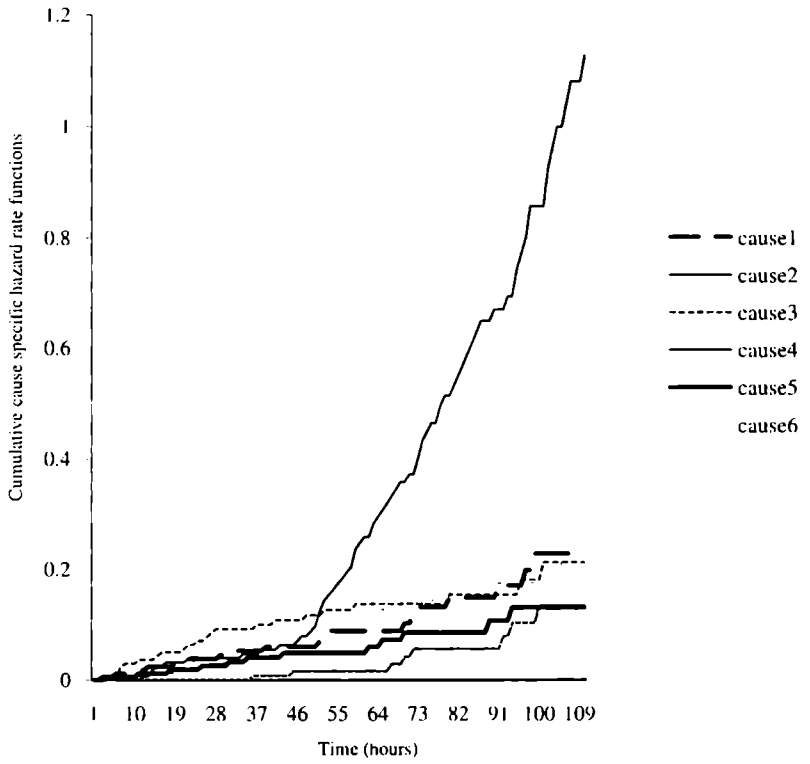


Figure 5.3 Plots of cumulative cause specific hazard rate functions for data on pneumatic tires

Table 5.6 provides the values of the test statistics along with their P-values for various weight functions for the above discussed data set. The bootstrap estimates of the absolute value of the biases and the mean square errors of the test statistics Q_1, Q_2, Q_3 and Q_4 are computed from the real data based on 1000 bootstrap samples of size 172, those are given in Table 5.7.

Table 5.6 Test statistic values using different weight functions and likelihood ratio test statistic for the data on pneumatic tires

Test statistic	Test statistic value (censored data)	P value
Q_1	8.1618	<0.01
Q_2	16.4635	<0.01
Q_3	16.6499	<0.01
Q_4	8.5328	<0.01
Q^*	12.939	<0.01

Table 5.7 Bootstrap estimates of absolute value of the biases and mean square error of test statistics Q_1, Q_2, Q_3 and Q_4 based on 1000 bootstrap samples for the data on pneumatic tires

Test statistic	Bias	MSE
Q_1	0.4005	0.5726
Q_2	0.2897	0.5392
Q_3	0.3998	0.5983
Q_4	0.1121	0.3835

The bootstrap estimates of the absolute value of the biases and the mean square errors of test statistics Q_1, Q_2, Q_3 and Q_4 computed from the real data. As earlier, Q_4 provides small bias (0.1121) and MSE (0.3835). Thus we can conclude that time of failure and causes of failure are not independent for the failure time data on pneumatic tires.

5.3 Tests for Categorical Lifetime Data

In this section, we consider the tests for testing independence of time to failure and cause of failure for a grouped lifetime data. Four different situations in categorical competing risks data; viz. (1)- unmasked uncensored data (2)- unmasked censored data (3)- masked uncensored data and (4) -masked censored data are considered here.

5.3.1 Unmasked Data

In this section, we derive the likelihood ratio tests for independence of time to failure and cause of failure when all causes of failure are known. First we consider the testing problem when the data are uncensored.

5.3.1.1 Uncensored Case

Suppose that an individual is exposed to k risks of death and also that the life span of the individual is split into m intervals, $I_i = [t_{i-1}, t_i)$, $i = 1, 2, ..., m$ with $t_0 = 0$. Then, the probability that the individual fails in the i th interval due to j th cause is denoted by p_{ij} ,where

$$p_{ij} = P[T \in I_i, C = j] \tag{5.21}$$

with $\sum_{i=1}^{m}\sum_{j=1}^{k} p_{ij} = 1$. Now, $p_{i.} = \sum_{j=1}^{k} p_{ij}$ and $p_{.j} = \sum_{i=1}^{m} p_{ij}$ for $i = 1, 2, ..., m \quad j = 1, 2, ..., k$,

where $p_{i.}$ is the probability of failure in the i th interval and $p_{.j}$ is the probability of failure due to cause j

Assume that n individuals are being observed. Let d_{ij} be the number of failures in the i th interval due to j th cause. Let $d_{i.}$ denote the number of failures in the i th interval and $d_{.j}$, the number of failures due to j th cause. Then

$$d_{i.} = \sum_{j=1}^{k} d_{ij}, d_{.j} = \sum_{i=1}^{m} d_{ij} \text{ and } n = \sum_{i=1}^{m}\sum_{j=1}^{k} d_{ij}$$

We wish to test the hypothesis that the time to failure is independent of the cause of failure. That is,

$$H_0 : p_{ij} = p_{i.}p_{.j} \text{ for all } i = 1,2,...,m\,;\, j = 1,2,...,k \tag{5.22}$$

against the alternative that the equality does not hold at least for one i or j

Now the likelihood function of the observed data is given by

$$L = \prod_{i=1}^{n}\prod_{j=1}^{k} p_{ij}^{\,d_{ij}}$$

Then, the unrestricted maximum likelihood estimate of p_{ij} is given by

$$\hat{p}_{ij} = \frac{d_{ij}}{n} \qquad i = 1,2,...,m\,;\, j = 1,2,...,k \tag{5.23}$$

Under H_0, the likelihood function is given by

$$L^0 = \prod_{i=1}^{n} p_{i.}^{\,d_{i.}} \prod_{j=1}^{k} p_{.j}^{\,d_{.j}}$$

Then, under H_0, maximum likelihood estimates of $p_{i.}$ and $p_{.j}$ are given by

$$\hat{p}_{i.} = \frac{d_{i.}}{n} \text{ and } \hat{p}_{.j} = \frac{d_{.j}}{n} \qquad i = 1,2,...,m\,;\, j = 1,2,...,k \tag{5.24}$$

Using (5.23) and (5.24), the likelihood ratio is given by

$$\Lambda_1 = \frac{\prod_{i=1}^{m}(\hat{p}_{i.})^{d_{i.}} \prod_{j=1}^{k}(\hat{p}_{.j})^{d_{.j}}}{\prod_{i=1}^{m}\prod_{j=1}^{k}(\hat{p}_{ij})^{d_{ij}}}$$

which gives

$$-2\log\Lambda_1 = 2\left[\sum_{i=1}^{m}\sum_{j=1}^{k} d_{ij}\log\frac{d_{ij}}{n} - \sum_{i=1}^{m} d_{i.}\log\frac{d_{i.}}{n} - \sum_{j=1}^{k} d_{.j}\log\frac{d_{.j}}{n}\right] \tag{5.25}$$

Using Taylor series expansion of (5.25), we get

$$-2\log\Lambda_1 = 2n\left[\sum_{i=1}^{m}\sum_{j=1}^{k}\frac{d_{ij}}{n}\left(\log\frac{d_{ij}}{n} - \log\frac{d_{i.}}{n}\frac{d_{.j}}{n}\right)\right]$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{k}\frac{\left(d_{ij} - \dfrac{d_{i.}d_{.j}}{n}\right)^2}{\dfrac{d_{i.}d_{.j}}{n}} + O\left(\frac{1}{n}\right).$$

Notice that $\sum_{i=1}^{m}\sum_{j=1}^{k}\left(\hat{p}_{ij} - \hat{p}_{i.}\hat{p}_{.j}\right) = 0$. Then, under H_0,

$T_1 = -2\log\Lambda_1 = \sum_{i=1}^{m}\sum_{j=1}^{k}\dfrac{\left(d_{ij} - \dfrac{d_{i.}d_{.j}}{n}\right)^2}{\dfrac{d_{i.}d_{.j}}{n}}$ has asymptotically a $\chi^2_{(k-1)(m-1)}$. Large values of the statistic are significant.

5.3.1.2 Censored Case

Let d_{ic} be the number of individuals who are censored in the i th interval. The total number of failures observed is given by

$$d = \sum_{i=1}^{m}\sum_{j=1}^{k} d_{ij} \tag{5.26}$$

Then,

$$n = \sum_{i=1}^{m}\sum_{j=1}^{k} d_{ij} + \sum_{i=1}^{m} d_{ic}$$

Let $p_{ic} > 0$ be the probability that an individual is censored in the i th interval. Then

$$\sum_{i=1}^{m}\sum_{j=1}^{k} p_{ij} + \sum_{i=1}^{m} p_{ic} = 1.$$

The likelihood is given by

$$L = \prod_{i=1}^{m}\prod_{j=1}^{k} p_{ij}^{d_{ij}} p_{ic}^{d_{ic}} \tag{5.27}$$

The expression for the likelihood in (5.27) is not useful for defining the test statistic. Accordingly, we consider the conditional probability

$$q_{ij} = P \text{ (an individual fails in the } i \text{ th interval} \mid \text{it is alive at the beginning of the } i \text{ th interval)} \qquad (5.28)$$

Then, $q_{i.} = \sum_{j=1}^{k} q_{ij}$ is the conditional probability of individual failing in the i th given he was alive at the beginning of the i th interval. Let n_i denote the number of individuals at risk in the beginning of the i th interval,

$$n_i = n - \sum_{l=1}^{i-1} d_{l.} - \sum_{l=1}^{i-1} d_{lc}$$

One can also write the likelihood given in (5.27) as follows

$$L = \prod_{i=1}^{m} \prod_{j=1}^{k} q_{ij}^{d_{ij}} (1 - q_{i.})^{n_i - d_{i.}} \qquad (5.29)$$

Note that there are mk parameters $q_{ij}, 0 \le q_{ij} \le 1, 0 \le \sum_{j=1}^{k} q_{ij} \le 1$, for each i

The unrestricted maximum likelihood estimates of the parameters in the censored case are given by

$$\hat{q}_{ij} = \frac{d_{ij}}{n_i}, \quad i = 1, 2, ..., m \quad j = 1, 2, ..., k \qquad (5.30)$$

Under the hypothesis of the independence of failure time and cause of failure, we have

$$q_{ij} = p_{.j} q_{i.} \quad i = 1, 2, ..., m \quad j = 1, 2, ..., k$$

The likelihood function under H_0 is given by

$$L^0 = \prod_{i=1}^{m} \prod_{j=1}^{k} p_{.j}^{d_{ij}} q_{i.}^{d_{ij}} (1 - q_{i.})^{n_i - d_{i.}}$$

Then under H_0 , the maximum likelihood estimates are given by

$$\hat{p}_{.j} = \frac{d_{.j}}{d} \text{ and } \hat{q}_{i.} = \frac{d_{i.}}{n_i}, \quad i = 1,2,...,m\,;\, j = 1,2,...,k\,. \tag{5.31}$$

where d is as defined in (5.26).

Thus, using (5.30) and (5.31), the likelihood ratio is

$$\Lambda_2 = \frac{\prod_{j=1}^{k}\left(\frac{d_{.j}}{d}\right)^{d_{.j}} \prod_{i=1}^{m} d_{i.}^{\,d_{i.}}}{\prod_{i=1}^{m}\prod_{j=1}^{k} d_{ij}^{\,d_{ij}}}$$

which leads to

$$-2\log\Lambda_2 = 2\left[\sum_{i=1}^{m}\sum_{j=1}^{k} d_{ij}\log d_{ij} - \sum_{j=1}^{k} d_{.j}\log d_{.j} - \sum_{i=1}^{m} d_{i.}\log d_{i.} + d\ \log d\ \right]$$

$$= 2\sum_{i=1}^{m}\sum_{j=1}^{k} d_{ij}\left[\log d_{ij} - \log\frac{d_{i.}d_{.j}}{d}\right] \tag{5.32}$$

Using Taylor series expansion of (5.32) and the fact that

$$\sum_{i=1}^{m}\sum_{j=1}^{k}\left[d_{ij} - \frac{d_{i.}d_{.j}}{d}\right] = 0$$

we have,

$$-2\log\Lambda_2 = \sum_{i=1}^{m}\sum_{j=1}^{k}\frac{\left(d_{ij} - \frac{d_{i.}d_{.j}}{d_{..}}\right)^2}{\frac{d_{i.}d_{.j}}{d}} + O\left(\frac{1}{n}\right).$$

Thus under $H_0, T_2 = -2\log\Lambda_2 = \sum_{i=1}^{m}\sum_{j=1}^{k}\frac{\left(d_{ij} - \frac{d_{i.}d_{.j}}{d_{..}}\right)^2}{\frac{d_{i.}d_{.j}}{d}}$ has asymptotically a χ^2 distribution with $(k-1)(m-1)$ degrees of freedom. Again, large values of the test statistic reject the null hypothesis.

5.3.2 Masked Data

In a competing risks scenario, sometimes the experimenter may not able to obtain complete information on the cause of failure for all individuals. Identifying the cause for all individuals might be too expensive or just may not be feasible. Several authors have considered likelihood based inference for the analysis of such competing risks data (see Dinse (1982), Dewanji (1992), Goetghebeur and Ryan (1995), Dewanji and Sengupta (2003) and Lu and Tsiatis (2005)).

In the following, we derive the likelihood ratio test for testing independence of T and C when only partial information about the cause of failure is available.

5.3.2.1 Uncensored Case

Consider a situation where n individuals are being observed. Let p_{ij} and d_{ij} are defined as in the previous section. Let d_{iu} denote the number of failures in the i th interval whose causes are unknown. Let p_{iu} be the probability that an individual fails in the i th interval and the cause of failure is unknown.

Then

$$n = \sum_{i=1}^{m}\sum_{j=1}^{k} d_{ij} + \sum_{i=1}^{m} d_{iu}$$

and

$$1 = \sum_{i=1}^{m}\sum_{j=1}^{k} p_{ij} + \sum_{i=1}^{m} p_{iu}$$

Now the likelihood based on observed data is given by

$$L = \prod_{i=1}^{m}\prod_{j=1}^{k} p_{ij}^{\,d_{ij}} \left\{ \prod_{i=1}^{m} p_{iu}^{\,d_{iu}} \right\}.$$

The unrestricted maximum likelihood estimates of p_{ij} and p_{iu} given as

$$\hat{p}_{ij} = \frac{d_{ij}}{n} \text{ and } \hat{p}_{iu} = \frac{d_{iu}}{n} \qquad i = 1,2,...,m\,;\ j = 1,2,...,k \tag{5.33}$$

Under H_0, we have

$$p_{ij} = (p_{i.} + p_{iu})p_{.j} \text{ and } p_{iu} = (p_{i.} + p_{iu})\left(1 - \sum_{j=1}^{k} p_{.j}\right), i = 1,2,...,m\,;\ j = 1,2,...,k$$

The second relationship follows from the fact that the sum of entries corresponding to i th interval is constant.

Thus, under H_0, the likelihood is given by

$$L^{c} = \prod_{j=1}^{k} p_{.j}^{\,d_{.j}} \left(1 - \sum_{j=1}^{k} p_{.j}\right) \prod_{i=1}^{m} (p_{i.} + p_{iu})^{d_{i.} + d_{iu}}$$

Then, under H_0, the maximum likelihood estimates are given as

$$\hat{p}_{.j} = \frac{d_{.j}}{n},\ \hat{p}_{i.} + \hat{p}_{iu} = \frac{d_{i.} + d_{iu}}{n};\quad i = 1,2,...,m\,;\ j = 1,2,...,k \tag{5.34}$$

Now using (5.33) and (5.34) the likelihood ratio is given by

$$\Lambda_3 = \frac{\prod_{j=1}^{k} d_{.j}^{\,d_{.j}} \prod_{i=1}^{m} (d_{i.} + d_{iu})^{d_{i.} + d_{iu}} \left(\sum_{i=1}^{m} d_{iu}\right)^{\sum_{i=1}^{m} d_{iu}}}{n^{n} \prod_{i=1}^{m}\prod_{j=1}^{k} d_{ij}^{\,d_{ij}} d_{iu}^{\,d_{iu}}}$$

Therefore

$$-2\log\Lambda_3 = 2\sum_{i=1}^{m}\sum_{j=1}^{k} d_{ij}\log d_{ij} + 2\sum_{i=1}^{m} d_{iu}\log d_{iu} + 2n\log n$$
$$-2\sum_{j=1}^{k} d_{.j}\log d_{.j} - 2\sum_{i=1}^{m}(d_{i.}+d_{iu})\log(d_{i.}+d_{iu}) - 2\left(\sum_{i=1}^{m} d_{iu}\right)\log\left(\sum_{i=1}^{m} d_{iu}\right)$$
$$= 2\sum_{i=1}^{m}\sum_{j=1}^{k} d_{ij}\left[\log d_{ij} - \log\frac{d_{.j}(d_{i.}+d_{iu})}{n}\right]$$
$$+ 2\sum_{i=1}^{m} d_{iu}\left[\log\frac{d_{iu}}{\sum_{i=1}^{m} d_{iu}} - \log\frac{(d_{i.}+d_{iu})}{n}\right]$$
$$\simeq \sum_{i=1}^{m}\sum_{j=1}^{k}\frac{\left[d_{ij} - \frac{d_{.j}(d_{i.}+d_{iu})}{n}\right]^2}{\frac{d_{.j}(d_{i.}+d_{iu})}{n}} + \sum_{i=1}^{m}\frac{\left[d_{iu} - \frac{(d_{i.}+d_{iu})\sum_{i=1}^{m} d_{iu}}{n}\right]}{\frac{(d_{i.}+d_{iu})\sum_{i=1}^{m} d_{iu}}{n}}$$

The last step follows from the fact that

$$\sum_{i=1}^{m}\left[d_{ij} - \frac{d_{.j}(d_{i.}+d_{iu})}{n}\right] = \sum_{i=1}^{m}\left[d_{iu} - \frac{(d_{i.}+d_{iu})\sum_{i=1}^{m} d_{iu}}{n}\right] = 0.$$

Thus, under H_0,

$$T_3 = -2\log\Lambda_3 = \sum_{i=1}^{m}\sum_{j=1}^{k}\frac{\left[d_{ij} - \frac{d_{.j}(d_{i.}+d_{iu})}{n}\right]^2}{\frac{d_{.j}(d_{i.}+d_{iu})}{n}} + \sum_{i=1}^{m}\frac{\left[d_{iu} - \frac{(d_{i.}+d_{iu})\sum_{i=1}^{m} d_{iu}}{n}\right]}{\frac{(d_{i.}+d_{iu})\sum_{i=1}^{m} d_{iu}}{n}}$$

has a χ^2 distribution $k(m-1)$ d.f. Large values of T_3 are significant.

5.3.2.2 Censored Case

We now consider the situation with censored observations in a masked failure time data. The total number of individuals is given by d and let d_{ic} be the number of individuals who are censored in i th interval. Suppose that, d_{iu} is the number of individuals who die in i th interval and whose cause of failure is unknown. Let the probability of this event as p_{iu}. Then,

$$n = \sum_{i=1}^{m}\sum_{j=1}^{k} d_{ij} + \sum_{i=1}^{m} d_{ic} + \sum_{i=1}^{m} d_{iu}$$

and $$1 = \sum_{i=1}^{m}\sum_{j=1}^{k} p_{ij} + \sum_{i=1}^{m} p_{ic} + \sum_{i=1}^{m} p_{iu}$$

The likelihood is given by

$$L = \prod_{i=1}^{m}\prod_{j=1}^{k} p_{ij}^{d_{ij}} \left(p_{iu}\right)^{d_{iu}} \left(p_{ic}\right)^{d_{ic}}$$

Let q_{iu} be the probability that an individual fails in the i th interval with unknown cause of failure given that he is alive at the beginning of the i th interval.

The likelihood function in terms of the conditional probabilities is given by

$$L = \prod_{i=1}^{m}\prod_{j=1}^{k} q_{ij}^{d_{ij}} \left(q_{iu}\right)^{d_{iu}} \left(1 - q_i - q_{iu}\right)^{n - d_{i.} - d_{iu}}$$

where the number of individuals at risk at the beginning of i th interval is

$$n_i = n - \sum_{l=1}^{i-1} d_l - \sum_{l=1}^{i-1} d_{lc} - \sum_{l=1}^{i-1} d_{lu} \quad i = 1, 2, ..., m.$$

In this case, there are $mk + m$ parameters $q_{ij}, 0 \le q_{ij} \le 1, 0 \le \sum_{j=1}^{k} q_{ij} \le 1$, for each i and $q_{iu}, i = 1, 2, ..., m$.

The unrestricted maximum likelihood estimates of the parameters are given by

$$\hat{q}_{ij} = \frac{d_{ij}}{n_i} \quad \text{and} \quad \hat{q}_{iu} = \frac{d_{iu}}{n_i} \quad \text{for } i = 1, 2, ..., m,\ j = 1, 2, ...k \tag{5.35}$$

Since the causes are missing completely at random, then under the null hypothesis of the independence of failure time and cause of failure, we have

$$q_{ij} = p_{.j}q_{i.} \text{ and } q_{iu} = (q_{i.} + q_{iu})\left(1 - \sum_{j=1}^{k} p_{.j}\right) \; i = 1,2,...,m \; , \; j = 1,2,...,k$$

The second condition follows from the fact that the sum of individuals failing in the i th interval is constant.

Thus, the likelihood function under H_0 is given by

$$L^0 = \prod_{j=1}^{k} p_{.j}^{\,d_{.j}} \left(1 - \sum_{j=1}^{k} p_{.j}\right)^{\sum_{i=1}^{m} d_{iu}} \prod_{i=1}^{m} (q_{i.} + q_{iu})^{(d_{i.} + d_{iu})} (1 - q_{i.} - q_{iu})^{n - d_{i.} - d_{iu}}$$

Then, under H_0, maximum likelihood estimates are given by

$$\hat{q}_{i.} + \hat{q}_{iu} = \frac{d_{i.} + d_{iu}}{n_i} \; , i = 1,2,...,m \; , \; j = 1,2,...,k \tag{5.36}$$

Hence, using (5.35) and (5.36) the likelihood ratio is given by

$$\Lambda_4 = \frac{\prod_{j=1}^{k} d_{.j}^{\,d_{.j}} \prod_{i=1}^{m} (d_{i.} + d_{iu})^{(d_{i.} + d_{iu})} \left(\sum_{i=1}^{m} d_{iu}\right)^{\sum_{i=1}^{m} d_{iu}}}{\left(d + \sum_{i=1}^{m} d_{iu}\right)^{d + \sum_{i=1}^{m} d_{iu}} \prod_{i=1}^{m} \prod_{j=1}^{k} d_{ij}^{\,d_{ij}} d_{iu}^{\,d_{iu}}}$$

which gives

$$-2\log\Lambda_4 = 2\sum_{i=1}^{m}\sum_{j=1}^{k} d_{ij}\log d_{ij} + 2\sum_{i=1}^{m} d_{iu}\log d_{iu} + 2\left(d + \sum_{i=1}^{m} d_{iu}\right)\log\left(d + \sum_{i=1}^{m} d_{iu}\right)$$
$$- 2\sum_{j=1}^{k} d_{.j}\log d_{.j} - 2\sum_{i=1}^{m}(d_{i.} + d_{iu})\log(d_{i.} + d_{iu}) - 2\left(\sum_{i=1}^{m} d_{iu}\right)\log\left(\sum_{i=1}^{m} d_{iu}\right)$$

$$= 2\sum_{i=1}^{m}\sum_{j=1}^{k} d_{ij}\left[\log d_{ij} - \log \frac{d_{.j}(d_{i.}+d_{iu})}{\left(d + \sum_{i=1}^{m} d_{iu}\right)}\right]$$

$$+ 2\sum_{i=1}^{m} d_{iu}\left[\log \frac{d_{iu}}{\sum_{i=1}^{m} d_{iu}} - \log \frac{(d_{i.}+d_{iu})}{\left(d + \sum_{i=1}^{m} d_{iu}\right)}\right]$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{k} \frac{\left[d_{ij} - \dfrac{d_{.j}(d_{i.}+d_{iu})}{\left(d + \sum_{i=1}^{m} d_{iu}\right)}\right]^2}{\dfrac{d_{.j}(d_{i.}+d_{iu})}{\left(d + \sum_{i=1}^{m} d_{iu}\right)}} + \sum_{i=1}^{m} \frac{\left[d_{iu} - \dfrac{(d_{i.}+d_{iu})\sum_{i=1}^{m} d_{iu}}{\left(d + \sum_{i=1}^{m} d_{iu}\right)}\right]^2}{\dfrac{(d_{i.}+d_{iu})\sum_{i=1}^{m} d_{iu}}{\left(d + \sum_{i=1}^{m} d_{iu}\right)}}.$$

The last step follows from the identity

$$\sum_{i=1}^{m}\sum_{j=1}^{k}\left[d_{ij} - \frac{d_{.j}(d_{i.}+d_{iu})}{\left(d + \sum_{i=1}^{m} d_{iu}\right)}\right] = \sum_{i=1}^{m}\left[d_{iu} - \frac{(d_{i.}+d_{iu})\sum_{i=1}^{m} d_{iu}}{\left(d + \sum_{i=1}^{m} d_{iu}\right)}\right] = 0.$$

Let,

$$T_4 = -2\log\Lambda_4 = \sum_{i=1}^{m}\sum_{j=1}^{k} \frac{\left[d_{ij} - \dfrac{d_{.j}(d_{i.}+d_{iu})}{\left(d + \sum_{i=1}^{m} d_{iu}\right)}\right]^2}{\dfrac{d_{.j}(d_{i.}+d_{iu})}{\left(d + \sum_{i=1}^{m} d_{iu}\right)}} + \sum_{i=1}^{m} \frac{\left[d_{iu} - \dfrac{(d_{i.}+d_{iu})\sum_{i=1}^{m} d_{iu}}{\left(d + \sum_{i=1}^{m} d_{iu}\right)}\right]^2}{\dfrac{(d_{i.}+d_{iu})\sum_{i=1}^{m} d_{iu}}{\left(d + \sum_{i=1}^{m} d_{iu}\right)}}.$$

Under H_0, T_4 has asymptotically a χ^2 distribution with d. f $k(m-1)$. Large values of the statistic are significant.

5.3.3 Simulation Study

We carry out a series of simulation studies to assess the performance of the proposed test statistics. We consider two causes of failure. We generate 1000 random samples of size n = 50,100 and 250 from exponential distributions. To calculate the empirical type I error, lifetimes are generated from the exponential models with parameter values $\lambda_1 = 2, \lambda_2 = 8, \pi_1 = 0.2$ and $\pi_2 = 0.8$. We consider both 1% and 5% significance level for the tests. Censored observations are generated by a uniform random variable over $(0, a)$, where a is chosen in such a way that, the desired proportion of the observations are censored. Masked observations are chosen at random.

Empirical power of the test is calculated by generating lifetimes from the exponential models with parameter values $\lambda_1 = 8, \lambda_2 = 2, \pi_1 = 0.2$ and $\pi_2 = 0.8$. The generated data is grouped into intervals with non zero frequencies in each cell. Tables 5.8-5.10 provide empirical type I errors and empirical powers of $T_1 - T_3$.

Table 5.8 Empirical type I error and power (in percentage) of the test statistic T_1

Significance level (%)	Sample size	Type I error	Power
α =1	50	1.2	71
	100	1.2	99
	250	0.8	100
α =5	50	5.3	92
	100	5.4	100
	250	5.2	100

Table 5.9 Empirical type I error and power (in percentage) of the test statistic T_2 with 20% censored observations

Significance level (%)	Sample size	Type I error	Power
$\alpha=1$	50	1.3	58
	100	1.2	89
	250	1.1	100
$\alpha=5$	50	5.2	51
	100	5.1	99
	250	4.95	100

Table 5.10 Empirical type I error and power (in percentage) of the test statistic T_3 with 20% masked data

Significance level (%)	Sample size	Type I error	Power
$\alpha=1$	50	0.6	50
	100	0.8	83
	250	0.9	100
$\alpha=5$	50	4.6	58
	100	4.9	95
	250	5.1	100

To investigate the effect of censoring and masking on the performance of the test statistics, we carry out a series of simulation studies using exponential distribution with the same parameters given above. Empirical power and empirical type I error of the test statistic T_4 are computed for various amounts of censoring and masking. The results are given in Tables 5.11-5.16.

Table 5.11 Empirical type I error and power (in percentage) of the test statistic T_4 with 20% censored and 20% masked data

Significance level (%)	Sample size	Type I error	Power
$\alpha=1$	50	0.6	52
	100	0.7	82
	250	0.8	99
$\alpha=5$	50	5.3	56
	100	5.2	94
	250	4.9	100

Table 5.12 Empirical type I error and power (in percentage) of the test statistic T_4 with 30% censoring and 20% masking

Significance level (%)	Sample size	Type I error	Power
$\alpha=1$	50	0.6	49
	100	0.7	81
	250	0.8	98
$\alpha=5$	50	4.3	55
	100	4.4	89
	250	4.5	99

Table 5.13 Empirical type I error and power (in percentage) of the test statistic T_4 with 40% censoring and 20% masking

Significance level (%)	Sample size	Type I error	Power
$\alpha=1$	50	0.5	48
	100	0.6	77
	250	0.7	95
$\alpha=5$	50	3.9	51
	100	4.0	80
	250	4.3	97

Table 5.14 Empirical type I error and power (in percentage) of the test statistic T_4 with 20% censoring and 30% masking

Significance level (%)	Sample size	Type I error	Power
$\alpha=1$	50	0.5	48
	100	0.6	86
	250	0.6	98
$\alpha=5$	50	4.1	52
	100	4.3	90
	250	4.4	99

Table 5.15 Empirical type I error and power (in percentage) of the test statistic T_4 with 20% censoring and 40% masking

Significance level (%)	Sample size	Type I error	Power
α=1	50	0.6	46
	100	0.7	80
	250	0.7	96
α=5	50	4.2	50
	100	4.3	86
	250	4.4	97

Table 5.16 Empirical type I error and power (in percentage) of the test statistic T_4 with 30% censoring and 30% masking

Significance level (%)	Sample size	Type I error	Power
α=1	50	0.5	44
	100	0.6	75
	250	0.7	95
α=5	50	4.0	50
	100	4.1	82
	250	4.2	96

From the above tables we observe that the tests based on T_1 and T_2 attain their levels but T_3 and T_4 do not attain their levels. It is seen that masking slightly affects the levels attained for sample sizes 50 and 100. However, there is no affect of masking when sample size is large (250). When proportion of masked data is fixed and censoring proportion changes we observe that the levels are not attained. The empirical type I error decreases with increase in censoring, but increases with increase in sample size. Similar conclusions can be drawn when proportion of censored data is

fixed and proportion of masked data changes. It can also be noted that, in all cases, power of the test statistics increases as sample size increases.

5.3.4 Data Analysis

We now consider four different types of grouped data to test whether the time of failure and the cause of failure are independent or not. First, we discuss the analysis of data on failure times of radio transmitter receivers, given in Lawless (2003, p: 460). Failures are classified to two types; those confirmed on arrival at maintenance center (cause I) and those unconfirmed (cause II). We only consider observed failure times by deleting last 44 observations. There are 325 observed failure times; those are given in Table 5.17.

Table 5.17 Data on failure times of radio transmitter receivers

Time interval	Failures due to cause 1	Failures due to cause 1
[0,50)	26	15
[50,100)	29	15
[100,150)	28	22
[150,200)	35	13
[200,250)	17	11
[250,300)	21	8
[300,350)	11	7
[350,400)	11	5
[400.450)	12	3
[450,500)	7	4
[500,550)	6	1
[550,600)	9	2
[600,630)	6	1

The test statistic T_1 has the value 9.4707, with $P-$value >0.01. Thus, we do not reject the null hypothesis that time to failure and cause of failure are independent.

For the illustration of procedure in the censored case, we consider data from a laboratory test on pneumatic tires (Davis and Lawrance,1989). The data is discussed

in Section 5.2.4. There are 172 failure times. We have grouped this data into 3 intervals, 0-149, 15-249 and >250. We have combined modes 2,3 and 5,6 to get non-zero frequencies in each cell. Finally, we have four modes of failure with 21 censored observations and the data is given in Table 5.18. In Table 5.18, cause 1 failures denote the failures due to failure mode 1 in the original data, cause 2 failures denote the failures due to failure modes 2 and 3, cause 3 failures denote the failures by failure mode 4 and failures due to cause 4 denote the failures by failure modes 5 and 6.

Table 5.18 Data on failure times of pneumatic tires

Time interval	Cause 1	Cause 2	Cause 3	Cause4	Censored observations
[0,150)	6	13	6	8	2
[150,250)	10	12	42	20	2
>=250	3	5	21	4	17
Total	19	30	69	32	21

The value of the test statistic is T_2=18.89. The corresponding $P-$ value is less than 0.005 which means that the failure times and causes of failure are not independent.

For example of masked data, consider the hard drive manufacturing data of 172 computers given in Flehinger et al. (2002). Originally there are three causes of failure with two masked sets of failures viz. (1,2,3) and (1,3). There are 66 masked observations. For these individuals, the exact cause of failure is not known, but we only know that the cause of failure is anyone within the group (1,2,3) and (1,3). We grouped the data into 4 intervals. The grouped data is presented in Table 5.19.

Table 5.19 Data on failure times of hard drives

Time interval	Cause 1	Cause 2	Cause 3	Masked observations
[0,1.0)	14	8	2	9
[1.0,2.0)	8	2	12	13
[2.0,3.0)	5	4	17	20
>=3.0	8	5	21	24
Total	35	19	52	66

We get the value of T_3 as 27.343, with $P-$ value less than 0.005. Thus we reject the null hypothesis that time to failure and cause of failure are independent.

For the illustration of the procedure in censored and masked set up, we consider the same data given in Table 5.18 with one observation in every cell (except censored cells) is masked. The modified data set is given in Table 5.20.

Table 5.20 Data on failure times of pneumatic tires with masking

Time interval	Cause 1	Cause 2	Cause 3	Cause4	Censored obser-vations	Masked obser-vations
[0,150)	5	12	5	7	2	4
[150,250)	9	11	41	19	2	4
>=250	2	4	20	3	17	4
Total	16	27	66	29	21	12

The value of the test statistic T_4 is 22.79. The corresponding $P-$value is less than 0.005 which means that the time to failure and causes of failure are not independent.

5.4 Conclusion

In this chapter, we developed nonparametric test procedures for testing independence of time to failure and cause of failure for both continuous and categorical competing risks data. For continuous lifetime data, a class of tests using martingale approach and a test statistic using likelihood ratio method are derived. Asymptotic properties of the test statistics were also studied. For categorical lifetime data, we considered the situations when the lifetimes are censored within intervals and some of the causes of failures are unknown. Simulation studies were conducted to assess the power of the proposed test statistics. We illustrated our procedures with real life data sets.

Chapter Six

A Quantile Based Test for Comparing Cumulative Incidence Functions

6.1 Introduction

Quantile function, as an alternative to the distribution function can be employed for modeling and analysis of statistical data. The role of quantile function and other concepts derived from it is well established in exploratory data analysis and in different areas of applied statistics (see Parzen, 1979 and Gilchrist, 2000). In survival studies, with heavy tailed lifetime models, a single long term survivor can have a marked effect on reliability measures based on a distribution function. It is therefore more convenient to work with quantile functions that are less influenced by extreme observations. One can refer to Reid (1981), Slud et al. (1984), Su and Wei (1993), Nair et al. (2008), Nair and Sankaran (2009) and Sankaran and Nair (2009) for modeling and analysis of lifetime data using the quantile based reliability measures.

Recently, Peng and Fine (2007) and Jeong and Fine (2009) have studied nonparametric quantile inference for competing risks models. In this chapter, we consider the problem of testing equality of cumulative incidence functions using quantile functions.

The chapter is organized as follows. In Section 6.2, we present the basic concepts of competing risks models in terms of quantile functions. A test statistic based on quantile functions is proposed in Section 6.3. The asymptotic distribution of

---

The results of this chapter have been accepted for publication as entitled 'A Quantile Based Test for Comparing Cumulative Incidence Functions' in Statistics and Probability Letters (see Sankaran, Nair and Sreedevi, 2010).

the test statistic is shown to be chi-square in Section 6.4. A simulation study is carried out in Section 6.5, to assess the performance of the test statistic. Section 6.6 illustrates the practical utility of the proposed procedure using two real life data sets. Finally, in Section 6.7, we provide a brief conclusion of the study.

6.2 Quantile Functions

Let T be a random variable representing the lifetime of a subject (or an individual) having an absolutely continuous distribution function $F(t)$ and $C \in \{1,2,...,k\}$ be the possible causes of failure. Let $S(t)$ be the survivor function of T The cumulative incidence function $F_j(t)$ is defined as

$$F_j(t) = P(T \le t, C = j) \qquad j = 1,2,...,k \tag{6.1}$$

Note that, the probability of failure $F(t)$ is

$$F(t) = \sum_{j=1}^{k} F_j(t). \tag{6.2}$$

Recently, attention has been paid to the problem of studying possible differences in mortality from different causes. This problem can be studied by comparing cumulative incidence functions. Accordingly, we consider the problem of testing the null hypothesis

$$H_0 : F_1(t) = F_2(t) = ... = F_k(t) \text{ for all } t > 0. \tag{6.3}$$

For various approaches for testing (6.3), one can refer to Aras and Deshpande (1992), Aly et al. (1994), Carriere and Kochar (2000), Kochar et al. (2002) and El Barmi et al. (2008) and references there in. In the present work, we study the problem of testing (6.3) using quantile functions.

Define the quantile functions

$$Q_j(u) = \inf_t \{t \mid F_j(t) \ge u\} \qquad j = 1,2,...,k \tag{6.4}$$

Then (6.4) is the smallest time for which the probability of failure due to cause j having occurred exceeds u, in the presence of other risks, which prevent the occurrence of cause j

Now, the hypothesis (6.3) can be represented as

$$H_0 : Q_1(u) = Q_2(u) = ... = Q_k(u) \qquad \text{for all } 0 < u < 1 \tag{6.5}$$

From (6.2) and (6.3), we obtain

$$H_0 : F_j(t) = \frac{1}{k} F(t) \qquad j = 1, 2, .., k,\ t > 0. \tag{6.6}$$

Now using (6.6), the hypothesis (6.5) can be written as

$$H_0 : Q_j\left(\frac{u}{k}\right) = Q\ (u) \qquad j = 1, 2, .., k,\ 0 < u < 1 \tag{6.7}$$

where $Q\ (u) = \inf_t \left\{ t \middle| F\ (t) \geq u \right\}$

6. 3 A Test Statistic

Let T be the lifetime random variable and $C \in \{1, 2, ..., k\}$ be the possible causes of failure as defined in Section 6.2. Suppose that the lifetime variable T is randomly right censored by the censoring variable Z Assume that T and Z are independent. Let G be the distribution function of Z Under right censoring, we observe n independent and identically distributed samples $(X_i, C_i), i = 1, 2, .., n$ of (X, C) where $X = \min(T, Z)$, and C is observed only when $X = T$ .With usual counting process notation, let $N_{ij}(t) = I(X_i \leq t, C_i = j)$. Denote $N_j(t) = \sum_{i=1}^{n} N_{ij}(t)$, $Y_i(t) = I(X_i \geq t)$ and $Y(t) = \sum_{i=1}^{n} Y_i(t)$. A nonparametric estimator of $F_j(t)$ is obtained as

$$\hat{F}_j(t) = \int_0^t \frac{\hat{S}(x)}{Y(x)} I\left(Y(x) > 0\right) dN_j(x) \qquad j = 1, 2, .., k \tag{6.8}$$

where $\hat{S}(t)$ is the Kaplan-Meier estimate of $S(t)$.

Let $\gamma = \min_j \left\{ P(C = j) \wedge F_j(\tau_j) \right\}$, where $\tau_j = \sup_t \left\{ t \middle| F_j(t) < 1 \right\}$ and $a \wedge b$ is the minimum of a and b For fixed $u, 0 < u < \gamma$ a nonparametric estimate of $Q_j(u)$ is given by

$$\hat{Q}_j(u) = \inf_t \left\{ t \middle| \hat{F}_j(t) \geq u \right\} \qquad j = 1, 2, .., k . \tag{6.9}$$

We can obtain the estimate of $Q(u)$ as

$$\hat{Q}(u) = \inf_t \left\{ t \middle| \hat{F}\ (t) \geq u \right\} \tag{6.10}$$

where $\hat{F}\ (t) = 1 - \hat{S}(t)$. Now we consider the quantity

$$Z_j(u) = \sqrt{n} \left(\hat{Q}_j\left(\frac{u}{k}\right) - \hat{Q}\ (u) \right) \qquad j = 1, 2, .., k \tag{6.11}$$

Then a test statistic for testing (6.5) is given by

$$\chi^2(u) = Z'(u) \hat{\Sigma}(u)^{-} Z(u) \tag{6.12}$$

where $Z(u) = \left(Z_1(u), ..., Z_k(u) \right)$ and $\hat{\Sigma}(u)^{-}$ is the generalized inverse of $\hat{\Sigma}(u)$ with $\hat{\Sigma}(u)$ as a consistent estimator of variance-covariance matrix $\Sigma(u)$ of $Z(u)$. We show, in Section 6.4, that under H_0, for fixed u, $\chi^2(u)$ follows χ^2 distribution with $k-1$ degrees of freedom. In practice, we reject H_0 if $\chi^2 > \chi^2_{\alpha, k-1}$, where $\chi^2 = \sup_{0<u<\gamma} \chi^2(u)$ and $\chi^2_{\alpha, k-1}$ is the ordinate value of chi-square distribution with $k-1$ degrees of freedom at α level.

Note that $\hat{\Sigma}(u)$ has maximum rank $k-1$. If we delete, for instance, the last row and last column of $\hat{\Sigma}(u)$, to give $\hat{\Sigma}_0(u)$, say and let $Z_0(u)=(Z_1(u),...,Z_{k-1}(u))$ then (6.12) may be obtained as

$$\chi^2(u)=Z_0'(u)\hat{\Sigma}_0^{-1}(u)Z_0(u)$$

where $\hat{\Sigma}_0^{-1}(u)$ is the ordinary inverse of $\hat{\Sigma}_0(u)$.

As shown in Section 6.4, the expression for Σ is complex and it involves the cause specific hazard rate functions. The estimation of Σ, thus generally requires smoothing. To obtain the approximate value of $\hat{\Sigma}(u)$, we can employ the method given in Peng and Fine (2007). Let $y(t)=P(X>t)$. Let $\eta^{(l)}$ be the solution of $F_l(\eta^{(l)})=\min(\max\{0,-n^{-1}\sum_{i=1}^{n}\tilde{\iota}_{il}(\tilde{Q}_l(u))\xi_i\},1)$, where $\tilde{\iota}_{il}$ is obtained as ι_{il} with

$$\iota_{il}(t)=\int_0^t\frac{S(u)}{y(u)}dN_{il}(u)-\int_0^t\frac{Y_i(u)S(u)}{y(u)}\lambda_l(u)du+\int_0^t\{-S(u)\}\{\int_0^u\frac{dN_{il}(v)}{y(v)}-\int_0^t\frac{Y_i(v)\lambda_l(v)}{y(v)}dv\}du$$

and $S(t), y(t)$ and $\lambda_l(t)dt$ are replaced by corresponding empirical quantities and ξ_i's are independent standard normal variables; $i=1,2...n$ and $l=1,2...k$ The estimate of the variance of $\tilde{Q}_l(u), l=1,2...k$;the diagonal elements of $\hat{\Sigma}(u)$, are obtained with the empirical variance of $\eta^{(l)}$ calculated by repeatedly generating $(\xi_1,\xi_2,...\xi_n)$ while fixing the data at their observed points, after omitting infinite values $l=1,2...k$. The estimate of the variance of the off diagonal elements of $\hat{\Sigma}(u)$ is obtained with the empirical covariance between $\eta^{(l)}$ and $\eta^{(m)}$ calculated by generating $(\xi_1,\xi_2,...\xi_n)$ again while fixing the data at their observed values, $l,m=1,2...k$ and $l\neq m$.

6.4 Asymptotic Distribution

We, now find the asymptotic distribution of the proposed test statistic.

Theorem 6.1

Assume that, $F_j(t)$ is continuous, twice differentiable and the density $f_j(t)$ is uniformly bounded below by a positive constant for $t \in [p,q], 0 \le p < q < \gamma^*$ $j = 1,2,...,k$, where $\gamma^* = \min_j \tau_j$, where $\tau_j = \sup_t \{t | F_j(t) < 1\}$ Then for fixed $u, 0 < u < \gamma, \chi^2(u)$ follows chi-square distribution with $k-1$ degrees of freedom.

Proof

Consider

$$Z_j(u) = \sqrt{n}\left(\hat{Q}_j\left(\frac{u}{k}\right) - \hat{Q}\ (u)\right) \qquad j = 1,2,...,k$$

which can be written as

$$Z_j(u) = \sqrt{n}\left(\hat{Q}_j\left(\frac{u}{k}\right) - Q_j\left(\frac{u}{k}\right)\right) + \sqrt{n}\left(Q_j\left(\frac{u}{k}\right) - Q\ (u)\right) + \sqrt{n}\left(Q\ (u) - \hat{Q}\ (u)\right)$$

$$j = 1,2,...,k \qquad (6.13)$$

Under H_0, since $Q_j\left(\frac{u}{k}\right) = Q\ (u)$ for all $u\ (0 < u < \gamma)$, (6.13) becomes

$$Z_j(u) = \sqrt{n}\left(\hat{Q}_j\left(\frac{u}{k}\right) - Q_j\left(\frac{u}{k}\right)\right) + \sqrt{n}\left(Q\ (u) - \hat{Q}\ (u)\right)\ j = 1,2,...,k \qquad (6.14)$$

From Andersen et al. (1993) and Peng and Fine (2007), we can obtain that for fixed u, $(0 < u < \gamma)$, $\sqrt{n}\left(\hat{Q}_j\left(\frac{u}{k}\right) - Q_j\left(\frac{u}{k}\right)\right)$ is asymptotically normal with mean 0 and variance $\sigma_j^2(u)$, where

$$\sigma_j^2(u)=\frac{(1-(u/k))^2}{f_j^2\left(Q_j(u/k)\right)}\int_0^{Q_j(u/k)}\frac{dF_j(x)}{(1-F_j(x))}\quad j=1,2,...,k \tag{6.15}$$

Similarly, we can prove that for $0<u<\gamma,\sqrt{n}\left(\hat{Q}(u)-Q(u)\right)$ is asymptotically normal with mean 0 and variance $\sigma^2(u)$,where

$$\sigma^2(u)=\frac{(1-u)^2}{f^{*2}(Q(u))}\int_0^{Q(u)}\frac{dF(x)}{(1-F(x))} \tag{6.16}$$

where $f^*(t)$ is the density corresponding to $F(t)$. Thus, for fixed u, $Z_j(u)$ asymptotically normal with mean 0 and variance $\sigma^{2(j)}(u)$, where

$$\sigma^{2(j)}(u)=\sigma_j^2(u)+\sigma^2(u)-2E\left[\hat{Q}_j(u)-Q_j(u)\right]\left[\hat{Q}^*(u)-Q^*(u)\right]$$

$$j=1,2,...,k \tag{6.17}$$

This implies that, for fixed $u,0<u<\gamma,Z(u)=\left(Z_1(u),...,Z_k(u)\right)$ is asymptotically a $k-$variate normal with mean zero vector and variance covariance matrix ,$\Sigma(u)$ Thus the quadratic form $\chi^2(u)$ follows χ^2 distribution with $k-1$ degrees of freedom.

6.5 Simulation Study

We carry out a simulation study to assess the performance of the test statistic. We consider two causes of failure. The variables (X,Y) are generated based on Block and Basu's (1974) absolutely continuous bivariate exponential distribution with density

$$f(x_1,x_2)=\begin{cases}\dfrac{\lambda\lambda_1(\lambda_2+\lambda_0)}{\lambda_1+\lambda_2}\exp\left(-\lambda_1x_1-(\lambda_2+\lambda_0)x_2\right) \text{ if } x_1<x_2,\\ \dfrac{\lambda\lambda_2(\lambda_1+\lambda_0)}{\lambda_1+\lambda_2}\exp\left(-\lambda_2x_2-(\lambda_1+\lambda_0)x_1\right) \text{ if } x_1>x_2,\end{cases}$$

where $(\lambda_0, \lambda_1, \lambda_2)$ are the parameters and $\lambda = \lambda_0 + \lambda_1 + \lambda_2$. Then the failure time is $T = \min(X_1, X_2)$ and cause specific hazard rate functions $h_j(t) = \frac{\lambda_j \lambda}{\lambda_1 + \lambda_2}; j = 1,2$ are proportional. $\lambda_1 = \lambda_2$ is equivalent to $Q_1(u) = Q_2(u)$ for all u When $\lambda_1 \neq \lambda_2$, then $Q_1(u) \neq Q_2(u)$. In particular, $\lambda_1 < \lambda_2$ is equivalent to $Q_1(u) > Q_2(u)$ for all u .We fix $\lambda_1 = 1$. We set $\lambda_0 = 0$ and 1 and considered different values for λ_2. Now, $\lambda_0 = 0$ controls the degree of dependence between X_1 and X_2. We considered uncensored, mildly censored (20% censoring) and heavily censored (40% censoring) situations. In the censored situations, observations are generated from a uniform random variable over $(0, a)$, where a is chosen in such a way that 20% or 40% of the observations are censored. We generate random sample of size $n =$ 50,100 and 250. We use asymptotic critical values at 5% level. Empirical type I errors and empirical powers of the test are calculated by generating 1000 repeated samples. For each sample, 1000 bootstrap samples are selected to compute the estimate of the variance-covariance matrix $\hat{\Sigma}(u)$.

Table 6.1 gives the empirical type I errors and empirical powers (both in percent) of the quantile based χ^2-test statistic. Table 6.1 show that, the proposed statistic has empirical type I error values close to 0.05 when the null hypothesis is true. Note that, empirical type 1 error doesn't show much effect on censoring percentage, when the sample size is large. By varying the values of parameter λ_2, we estimate the empirical power of the test statistic. From Table 6.1, it is clear that, increase in censoring percentage doesn't have much effect on the power of the quantile based test statistic when sample size is large. The power of the test increases, as sample size increases. When the degree of departure from null hypothesis increases, empirical power of the test increases. The proposed method can be compared with the simulation studies conducted by Aly et al. (1994, p: 997) for $\lambda_1 \leq \lambda_2$. When sample size is small the performance of our test is much better than

that of the test by Aly et al. (1994). For large samples, our test is equally powerful with the test by Aly et al. (1994).

Table 6.1 Empirical type I errors and powers (in percentage) of the proposed test statistic at an asymptotic level of 5%.

λ_2	n=50		n=100		n=250	
	$\lambda_0 = 0$	$\lambda_0 = 1$	$\lambda_0 = 0$	$\lambda_0 = 1$	$\lambda_0 = 0$	$\lambda_0 = 1$
No Censoring						
0.5	90.8	91.2	94.3	94.4	99.7	99.7
1.0	5.2	4.8	4.9	5.1	4.9	4.9
1.5	80.9	76.5	84.3	83.2	99.7	98.4
2.0	94.1	94.7	96.3	96.1	99.8	99.8
2.5	96.5	96.4	99.3	99.5	100	100
Mild (20%) Censoring						
0.5	86.5	87.1	90.4	90.8	98.6	98.8
1.0	5.4	4.7	4.7	4.8	4.8	4.8
1.5	78.7	75.3	82.8	82.2	98.9	98.2
2.0	86.8	84.1	88.0	87.8	99.9	99.9
2.5	94.3	93.2	95.4	93.2	100	100
Heavy (40%) Censoring						
0.5	80.2	81.1	84.5	86.5	98.7	98.7
1.0	3.8	3.9	4.0	4.1	4.5	4.6
1.5	76.5	73.3	81.2	81.0	97.6	95.4
2.0	83.2	79.1	84.4	85.9	99.3	99.6
2.5	92.1	93.1	93.6	94.2	100	100

6.6 Data Analysis

We illustrate the utility of the method with two real life data sets. First, we consider the data from a laboratory experiment on 181 mice discussed by Hoel (1972). There are three causes of death viz. thymic lymphoma, reticulum cell sarcoma and

other causes. All the mice are died at the end of study, so that there is no censoring. The data were studied by different authors including Aly et al. (1994), Lam (1998), Kochar et al. (2002), Dewan et al. (2004) and Sankaran et al. (2009). We consider two situations with the above data set. First, the data is trated as a three risks problem. We employ the bootstrap technique to estimate $\Sigma(u)$. The test statistic value is obtained as $Q =$ 296.95 with P-value <0.00001, which indicates that the cumulative incidence functions due to failures from the three causes are significantly different. It can be noted that, Lam (1998) also has arrived the same conclusion using a test procedure based on cause specific hazard rate functions. Figure 6.1 gives the plot of $\hat{Q}_j(u)$ against u for three different causes. In Figure 6.1, the solid line represents the sub quantile function due to failures from cause 1, the dotted line denotes the sub quantile function due to failures from cause 2 and the dashed line represents the same due to failures from cause 3.

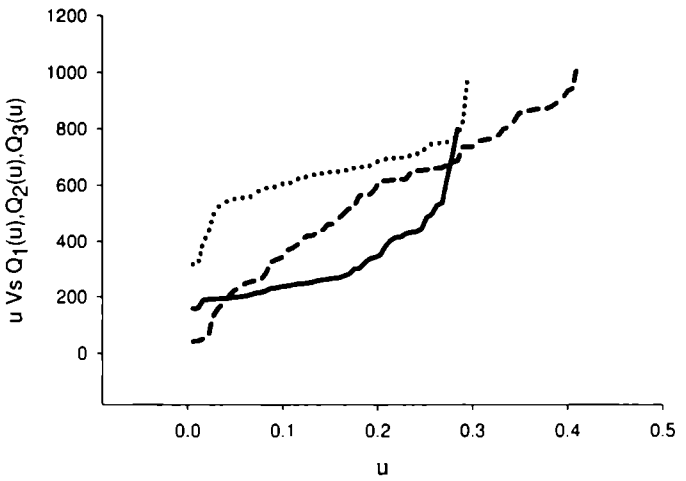


Figure 6.1 Plot of $\hat{Q}_j(u)$; $j = 1,2,3$ against u for mice mortality data set

Figure 6.1 shows that, the sub quantile functions due to different causes of failure are different each other. $\hat{Q}_1(u)$ yields a smaller value, comparing to $\hat{Q}_2(u)$ and $\hat{Q}_3(u)$,

starting from $u=0.04$ to $u=0.28$ approximately. However, $\hat{Q}_2(u)$ is significantly greater than $\hat{Q}_1(u)$ and $\hat{Q}_3(u)$ for almost all points even though the two sub quantiles coincide at $u = 0.3$.

Now we consider the same data as a two risks problem to compare with the Aly et al. (1994) procedure. All the lifetimes with cause of failure 'other causes' are considered to be censored. The test statistic value is obtained as $Q = 269.52$ with P value <0.00001, which indicates that the cumulative incidence functions due to the two causes, thymic lymphoma and reticulum cell sarcoma are significantly different. Using the test procedure by Aly et al. (1994), we obtained $D^{*} = 8.97$, which also infer the same conclusion. Figure 6.2 gives the plot of $\hat{Q}_j(u)$ against u for two different causes. In Figure 6.2, the solid line represents the sub quantile function due to failures from cause 1, and the dotted line denotes the sub quantile function due to failures from cause 2.

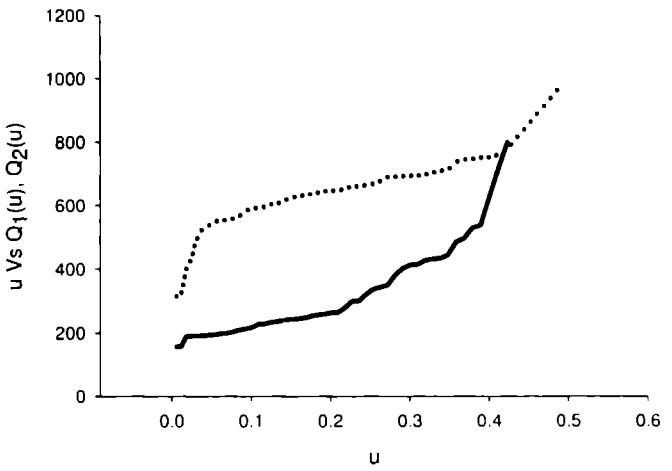


Figure 6.2 Plot of $\hat{Q}_j(u)$; $j=1,2$ against u for mice mortality data set

Figure 6.2 shows that, the sub quantile functions due to different causes of failure are different each other. $\hat{Q}_1(u)$ yields a smaller value, comparing to $\hat{Q}_2(u)$ for most

points. Even if at u =0.4 they coincide, $\hat{Q}_1(u)$ and $\hat{Q}_2(u)$ are significantly different for all other points.

We also consider a competing risks data from a laboratory experiment on pneumatic tires given in Davis and Lawrence (1989) to illustrate the utility of the method. The failures were classified into six modes. 1- open joint on the inner liner, 2-rubber chunking on the shoulder, 3-loose casing low on the side wall, 4-cracking on the tread rubber, 5-cracking on the side wall, 6-all other causes. In the present study, we first consider the data as a three risks problem. We merge the modes of failure 1,2 and 3 into a single failure mode and consider as cause 1 failures, while keeping the mode of failure 4 as such and consider as cause 2 failures. Further, modes of failure 5 and 6 are merged into a single failure mode and denote as cause 3 failures. There are 172 failure times including 22 censoring times. The test statistic value Q = 37.8244 with P-value <0.001, indicates that the cumulative incidence functions are significantly different for all the causes. The plots of $\hat{Q}_j(u), j=1,2$ are given in Figure 6.3. In Figure 6.3, the solid line represents the sub quantile function due to failures from cause 1, the dotted line denotes the sub quantile function due to failures from cause 2 and the dashed line represents the same due to failures from cause 3.

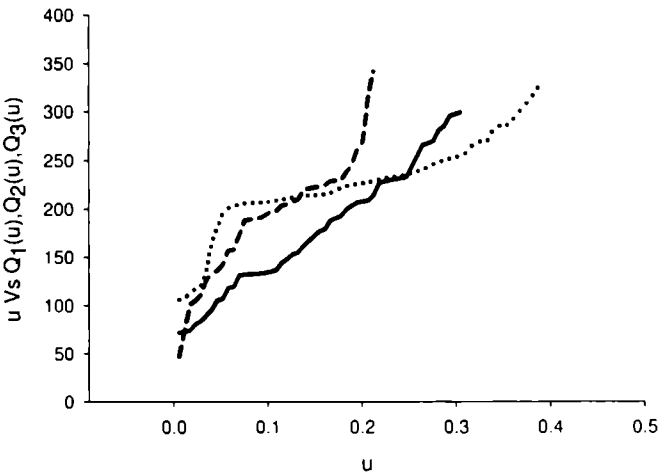


Figure 6.3 Plot of $\hat{Q}_j(u)$; $j=1,2,3$ against u for pneumatic tire data set.

Figure 6.3 shows that, the sub quantile functions due to the different causes of failure are different each other. $\hat{Q}_1(u)$ is smaller than $\hat{Q}_2(u)$ for the values of u value less than 0.22 and grater for u value above 0.22. Initially $\hat{Q}_3(u)$ has a very low value, but starting dominates $\hat{Q}_1(u)$ immediately and $\hat{Q}_2(u)$ near to the u value 0.15.

Now, we consider the above data set as a two risks problem to compare with the test statistic by Aly et al. (1994). The mode of failure 4, cracking on the tread rubber seems to be the major cause of failure. So we merge all other modes of failure into a single mode of failure and denote as cause 1 failures, while keeping the failures due to mode 4 as such and consider as cause 2 failures. For the two risks problem, we compare our statistic with the one proposed by Aly et al. (1994). The test statistic Q = 212.62 with P-value <0.00001, which indicates that the cumulative incidence functions due to the two causes are significantly different. Using the test procedure by Aly et al. (1994), we obtain D^* = 8.432, where $D = \sqrt{n}D_{3n}$ and D_{3n} is the statistic proposed by Aly et al. (1994), which also infer the same conclusion. Figure 6.4 gives the plot of $\hat{Q}_j(u)$ against u for two different causes. In Figure 6.4, the solid line represents the sub quantile function due to failure from cause 1, and the dotted line denotes the sub quantile function due to failures from cause 2.

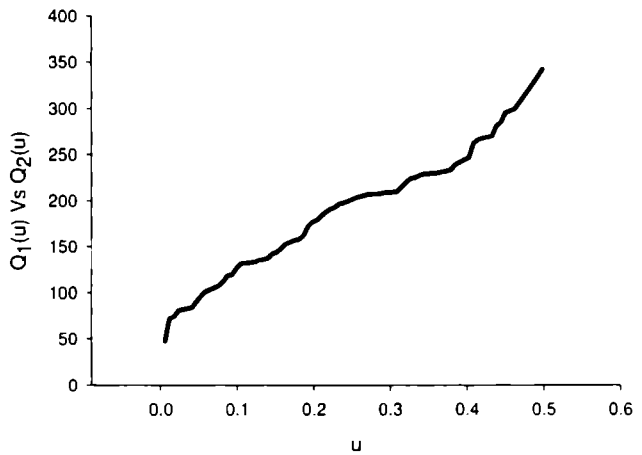


Figure 6.4 Plot of $\hat{Q}_j(u)$; $j = 1, 2$ against u for pneumatic tire data set.

Figure 6.4 shows that $\hat{Q}_1(u)$ and $\hat{Q}_2(u)$ are significantly different for all points. $\hat{Q}_2(u)$ obviously yields a higher value comparing to $\hat{Q}_1(u)$ for all u values. Probability of failure due to cause 2 is clearly less than probability of failure due to cause 1 for all time points.

6.7 Conclusion

In this chapter, we developed a quantile based test procedure for testing the equality of cumulative incidence functions. The performance of the procedure was studied using simulated examples. We demonstrated the practical utility of the method using real life data sets. There are several advantages for the proposed procedure over the existing techniques using distribution functions. Firstly, there exist simple quantile functions as lifetime models that are highly flexible than distribution functions. Further, quantile functions are very good approximations to lifetime distributions, so that it can work well in many practical situations. Secondly, the proposed procedure would be more suitable to work with data, for which distribution functions do not have simple closed forms. Moreover, simulation studies show that the quantile based test statistic seems to be robust in censored situations, than the existing ones.

In the present work, the Kaplan-Meier (1958) estimator of the survivor function is employed to find out quantile estimators. When Q is a continuous function, it may be more suitable to use a smoothed estimator rather than the step function $\hat{F}^{-1}$, since smoothing reduces the random variation in the data and allows a better display of interesting features of the lifetime distribution. Smoothed estimator of the quantile function using kernel density function given in Padgett (1986) can be employed in such contexts.

Chapter Seven

Conclusion

7.1 Introduction

Survival data is a term used to refer the data measuring time to occurrence of certain event of interest. Survival data frequently come from medical studies, but may come from other applied fields like demography, engineering, economics and social sciences. In many practical situations, individuals under study are exposed to the failure due to more than one cause or factor and the eventual failure of an individual can be attributed to exactly one of the causes or factors. Data arise from such situations are referred to as competing risks data and accordingly the models employed for the analysis of such data are referred to as competing risks models.

The literature on competing risks models reveal that there are several occasions in survival studies where the existing models and methodologies are inadequate for the analysis of the lifetime data. Identifiabilty problem and various types of censoring induce more complications in the analysis of competing risks data. This lead us to a consider some of important research questions on inference procedures in the analysis of competing risks data.

Classical survival analysis does not provide a tool to incorporate the prior information on the failure mechanism that generates the data during the modeling and analysis of data. In Chapter 2, we proposed a semiparametric Bayesian approach for modeling competing risks data by considering the cumulative baseline cause specific hazard rate functions as nuisance parameters. We carried out simulation studies to assess the asymptotic behaviour of the proposed method. The proposed method was illustrated using a real life example.

In many practical situations, lifetimes are observed to be doubly censored. The complex nature of the doubly censored competing risks data, does not allow us to use the well-known partial likelihood approach given in Fine and Gray (1999) for modeling data. In Chapter 3, we proposed a semiparametric transformation model for the cause specific subdistribution functions, conditional on the covariates for the modeling and analysis of doubly censored competing risks data. The parameters of the model were estimated using estimating equation approach. Asymptotic properties of the estimators were discussed. A simulation study was conducted to assess the finite sample behaviour of the estimators. The utility of the proposed method was well demonstrated using real a data set.

Modeling and analysis of lifetime data using neural network models is a topic of recent interest. Neural network models are less explored for the analysis of competing risks data when the lifetime variable is continuous. In Chapter 4, we proposed neural network models for various prediction and classification problems in the competing risks set up. The proposed neural network models were applied to real life data sets.

Testing independence of time to failure and cause of failure is an important research problem in the analysis of competing risks data. Many authors proposed tests based on U -statistics to address this problem. In Chapter 5, we introduced a class of tests using martingale approach. We also developed a test statistic using likelihood ratio test procedure. When lifetime data are categorical, we provided likelihood ratio test statistic that can be employed in different practical situations. Asymptotic distributions of the test statistics were proved to be chi-square. A series of simulation studies was carried out with continuous as well as categorical lifetime data to assess the power of the proposed test statistics. The practical utility of the proposed test statistics was demonstrated using different real life data sets.

In survival studies, with heavy tailed lifetime models, a single long term survivor can have a noticeable effect on reliability measures based on a distribution function. It is therefore more convenient to work with quantile functions that are less influenced by extreme observations. Accordingly, in Chapter 6, we developed a quantile based procedure to test the equality of cumulative incidence functions. The asymptotic distribution of the test statistic was shown to be chi-square. We carried out a simulation study to understand the finite sample behaviour of the proposed test statistic. Two real life data sets were used to illustrate the practical implementation of the proposed procedure.

7.2 Future Works

In Chapter 2, we proposed a semiparametric Bayesian approach for the modeling and analysis of competing risks data by assuming a prior distribution for each cumulative baseline cause specific hazard rate functions. The analysis was carried out by assuming independent increment process prior for cumulative baseline cause specific hazard rate functions in each interval. However, this assumption is unrealistic in many situations. Modeling the dependence among cause specific hazard rate functions in each interval using Markov Processes is an interesting research problem which is to be addressed. We can also attempt a true Bayesian approach to model competing risks data by assuming suitable prior distribution for cause specific subdistribution functions.

In survival studies, the individuals may observe to be grouped into a cluster. The heterogeneity among variables may be due to certain unobserved common risk factors present in the data. To model such unobserved factors, frailty models are usually employed in survival analysis. Frailty models are basically random effect models for survival data, where one of the random effects is specified by means of the hazard rate function. The semiparametric Bayesian analysis of frailty models for

competing risks data is an area yet to be explored. The extension of the semiparametric transformation model to the frailty set up for doubly censored competing risks data is another topic of research interest.

There are situations in survival studies where the exact lifetime of an event is not known, but it known to lie in some interval. Various semiparametric models have been developed for the analysis of such interval censored data. The analysis of interval censored competing risks data is complicated and the research work in this direction will be worth exploring. Apart from censoring, truncation is very common in life testing experiments. As the literature on the analysis of truncated data in competing risks set up is limited, the extension of our models to the truncated set up is a topic of future work.

Occasionally in competing risks data, the cause of failure for an individual has not been exactly observed, but has only been narrowed down to a subset of all potential risks. Situation with incomplete information about the cause of failure is referred to as masking. The semiparametric regression model in Chapters 3 can be studied to analyze masked competing risks data with proper modification. Neural network models proposed in Chapter 4 can also be modified to deal with masked data. With masked causes of failure, the testing of independence of time to failure and cause of failure for continuous lifetime data can be revisited. The problem of testing equality of cumulative incidence functions or cause specific hazard rate functions, when the data are masked is also worth exploring.

In biostatistical applications, there are situations where one can only observe the lifetime T belongs to certain interval $(0,C]$ or (C,∞) where C is known as the status time. Such data are called as current status data. The analysis of current status data with multiple causes of failure in presence of covariates is less explored in

literature. The models developed in Chapters 2, 3 and 4 can be extended to this set up, which is not straight forward.

Multivariate survival data often arise when each study subject experiences several events or when we study repeated occurrence of the same event. Lifetimes of left and right front brake pads in a car and times until a particular condition appears in the left and right eyes of a person are examples of multivariate survival data. The literature on the analysis of multivariate survival data with multiple causes of failure is limited. The semiparametric Bayesian analysis of multivariate competing risks data is not yet discussed in literature. Regression model described in Chapter 3 is worth studying in multivariate set up. The neural network models proposed in Chapter 4 can be attempted to handle multivariate competing risks data with proper modifications. Testing the independence of time to failure and cause of failure for multivariate competing risks data is a topic of research that remains to be explored. Testing the equality of cause specific hazard rate functions or cause specific subdistribution functions for multivariate competing risks data is another research problem that is yet to be extensively studied.

Bibliography

1. Aalen, O. (1976). Nonparametric inference in connection with multiple decrement models. *Scandinavian Journal of Statistics*, **3**, 15-27.
2. Aalen, O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics*, **6**, 701-726.
3. Adamic, P. F. (2008). Modeling multiple risks in the presence of double censoring. *Scandinavian Actuarial Journal*, 1-14, i First article.
4. Ahmed, F. E. (2005). Artificial neural networks for diagnosis and survival prediction in colon cancer. *Molecular Cancer*, **4**, 29-41.
5. Aly, E. A. A., Kochar, S. C. and McKeague, I. W. (1994). Some tests for comparing cumulative incidence functions and cause specific hazard rates. *Journal of the American Statistical Association*, **89**, 994-999.
6. Ambrogi, F., Lama, N., Boracchi, P. and Biganzoli, E. (2007). Selection of artificial neural network models for survival analysis with genetic algorithms. *Computational Statistics and Data Analysis*, **52**, 30-42.
7. Andersen, P. K. (1982). Testing goodness of fit of Cox's regression and life model. *Biometrics*, **38**, 67-77. Correction (1984) **40**, 12-17.
8. Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlang, New York.
9. Antony, A. A. and Sankaran, P. G. (2008). Nonparametric estimation of bivariate survivor function under masked causes of failure. *Journal of Nonparametric Statistics*, **20**, 77-89.
10. Aras, G. and Deshpande, J. V. (1992). Statistical analysis of dependent competing risks. *Statistics and Decisions*, **10**, 323-336.
11. Bagai, I., Deshpande, J. V. and Kochar, S. C. (1989a). A distribution free test for the equality of failure rates due to two competing risks. *Communications in Statistics -Theory and Methods*, **18**, 107-120.

12. Bagai, I., Deshpande, J. V. and Kochar, S. C. (1989b). Distribution free tests for stochastic ordering in the competing risks model. *Biometrika*, **76**, 775-781.

13. Bagai, I. and Prakasa Rao, B. L. S. (1992). Analysis of survival data with two dependent competing risks. *Biometrical Journal*, **34**,801-814.

14. Bakker, B. and Heskes, T. (1999). A neural Bayesian approach to survival analysis. *Proceedings of international Conference in Artificial Neural Networks*, 832-837.

15. Bakker, B., Heskes, T., Neijt, J. and Kappen, B. (2004). Improving Cox survival analysis with a neural Bayesian approach. *Statistics in Medicine*, **23,** 2989-3012.

16. Bedford, T. and Lindqvist, B. H. (2004). The identifiability problem for repairable systems subject to competing risks. *Advances in Applied Probability*, **36**, 774-790.

17. Berkson, J. and Elveback, L. (1960). Competing exponential risks with particular reference to the study of smoking and lung cancer. *Journal of the American Statistical Association*, **55**, 415-428.

18. Biganzoli, E., Boracchi, P., Mariani, L. and Marubini, E. (1998). Feed forward neural networks for the analysis of censored survival data: a partial logistic regression approach. *Statistics in Medicine*, **17,** 1169-1186.

19. Biganzoli. E., Boracchi, P. and Marubini, E. (2002). A general framework for neural network models on censored survival data. *Neural Networks*, **15**, 209–218.

20. Biganzoli, E., Boracchi, P., Ambrogi, F. and Marubini, E. (2006). Artificial neural network for the joint modelling of discrete cause specific hazards. *Artificial Intelligence in Medicine*, **37,** 119-130.

21. Bishop, C. M. (1996). *Neural Networks for Pattern Recognition*. Oxford University Press, New York.

22. Block, H. W. and Basu, A. P. (1974). A continuous bivariate exponential distribution. *Journal of the American Statistical Association*, **69**, 1031-1037.

23. Boag, J. W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society-Series B*, **11**, 15-53.

24. Boardman, T. J. and Kendell, P. J. (1970). Estimation in compound exponential failure models. *Technometrics*, **12**, 891-900.

25. Bravo, J. I. and Esteban, M. D. (1993). Nonparametric estimation of survival functions for doubly censored observations when the lifetime hazard rate is proportionally related to the censoring hazard rates. *Communications in Statistics- Theory and Methods*, **22**, 3237-3253.

26. Breslow, N. (1970). A generalized Kruskal-Wallis test for comparing k samples subject to unequal patterns of censorship. *Biometrika*, **57**, 579-594.

27. Byar, D. P. and Green, S. B. (1980). The choice of treatment for cancer patients based on covariate information: Applications to prostate cancer. *Bulletin du Cancer*, **67**,477-490.

28. Cai, B. (2003). *Adaptive Sampling Schemes and Bayesian Semiparametric Survival Analysis*. PhD Thesis, The University of Auckland, New Zealand. http://hdl.handle.net/2292/1037

29. Cai, T. and Cheng, S. (2004). Semiparametric regression analysis for doubly censored data. *Biometrika*, **91**, 277-290.

30. Carriere, K. C. and Kochar, S. C. (2000). Comparing sub-survival functions in a competing risks model. *Lifetime Data Analysis*, **6**, 85-97.

31. Chang, M. N. and Yang, G. L. (1987). Strong consistency of a nonparametric estimator of the survival function with doubly censored data. *The Annals of Statistics*, **15**, 1536-1547.

32. Chang, M. N. (1990). Weak convergence of a self-consistent estimator of the survival function with doubly censored data. *The Annals of Statistics*, **18**, 391-404.

33. Chen, B. E., Kramer, J. L., Greene, M. H. and Rosenberg, P. S. (2008). Competing risks analysis of correlated failure time data. *Biometrics*, **64**, 172-179.

34. Chen, K. and Zhou, M. (2003). Nonparametric hypothesis testing and confidence intervals with doubly censored data. *Lifetime Data Analysis*, **9**, 71-91.

35. Chen, Y. Q. and Cheng, S. (2005). Semiparametric regression analysis of mean residual life with censored survival data. *Biometrika*, **92**, 19-29.

36. Cheng, B. and Titterington, D. M. (1994). Neural networks: A review from a statistical perspective. *Statistical Science*, **9**, 2-30.

37. Cheng, S.C., Fine, J. P. and Wei, L. J. (1998). Predictions of cumulative incidence function under the proportional hazards model. *Biometrics*, **54**, 219-228.

38. Chiang, C. L. (1968). *Introduction to Stochastic Processes in Biostatistics*. Wiley, New York.

39. Collett, D. (2003). *Modelling Survival Data in Medical Research*. Chapman and Hall/CRC, Boca Raton.

40. Cox, D. R. (1959). The analysis of exponentially distributed life times with two types of failure. *Journal of the Royal Statistical Society- Series B*, **21**, 411-421.

41. Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society-Series-B*, **34**,187-220.

42. Craiu, R. V. and Duchesne, T. (2004). Inference based on the EM algorithm for the competing risks model with masked causes of failure. *Biometrika*, **91**, 543-558.

43. Craiu, R. V and Lee, T. C. M. (2005). Model selection for the competing risks model with and without masking. *Techometrics*, **47**, 457-467.

44. Craiu, R. V. and Reiser, B. (2006). Inference for the dependent competing risks model with masked causes of failure. *Lifetime Data Analysis*, **12**, 21-33.

45. Cronin, K. A. and Feuer, E. J. (2000). Cumulative cause specific mortality for cancer patients in the presence of other causes: a crude analogue of relative survival. *Statistics in Medicine*, **19**, 1729-1740.

46. Crowder, M. (1997). A test for independence of competing risks with discrete failure times. *Lifetime Data Analysis*, **3**,215-223.

47. Crowder, M. (2001). *Classical Competing Risks*. Chapman and Hall, London.

48. Cuzick, J. (1985). Asymptotic properties of censored linear rank tests. *The Annals of Statistics*, **13**, 133-141.

49. David, H. A. (1957). Estimation of means of normal populations from observed minima. *Biometrika*, **44**, 282-286.

50. David, H. A. and Moeschberger, M. L. (1978). *The Theory of Competing Risks*. Macmillan, New York.

51. Davis, T. P and Lawrence, A. J. (1989). The likelihood for competing risks survival analysis. *Scandinavian Journal of Statistics*, **16**, 23-28.

52. De Laurentiis, M. and Ravdin, P. M. (1994). A technique for using neural network analysis to perform survival analysis of censored data. *Cancer letters*, **77**, 127-138.

53. Deshpande, J. V (1990). A test for bivariate symmetry of dependent competing risks. *Biometrical Journal*, **32**, 737-746.

54. Deshpande, J. V and Sengupta, D. (1995). Testing the hypothesis of proportional hazards in two populations. *Biometrika*, **82**, 251-261.

55. Dewan, I. (2005). A review of some statistical tests under independent competing risks model. Technical Report, ISI, New Delhi.

56. Dewan, I., Deshpande, J. V and Kulathinal, S. B. (2004). On testing dependence between time to failure and cause of failure via conditional probabilities. *Scandinavian Journal of Statistics*, **31**, 79-91.

57. Dewan, I. and Kulathinal, S. B. (2007). On testing dependence between time to failure and cause of failure when causes of failure are missing. PLos ONE, **2**, 12.

58. Dewan, I., Sankaran, P. G. and Sreedevi, E. P. (2010). On testing independence of failure time and cause of failure in a competing risks model for grouped data. Communicated.

59. Dewanji, A. (1992). A note on a test for competing risks with missing failure type. *Biometrika*, **79**, 855-857.

60. Dewanji, A. and Sengupta, D. (2003). Estimation of competing risks with general missing pattern in failure types. Biometrics, **59**, 1063-1070.

61. Dignam, J. J. and Kocherginsky, M. N. (2008). Choice and interpretation of statistical tests used when competing risks are present. *Journal of Clinical Oncology*, **26**, 4027-4034.

62. Dinse, G. E. (1982). Nonparametric estimation of partially-complete time and type of failure data. *Biometrics*, **38**, 417-431.

63. Dykstra, R., Kochar, S. C. and Robertson, T. (1995). Likelihood based inference for cause specific hazard rates under order restrictions. *Journal of Multivariate Analysis*, **54**, 163-174.

64. Dykstra, R., Kochar, S. C. and Robertson, T. (1998). Restricted tests for testing independence of time to failure and cause of failure in a competing risks model. *The Canadian Journal of Statistics*, **26**, 57-68.

65. Efron, B. (1967). The two-sample problem with censored data. *Proceedings of Fifth Berkeley Symposium on Mathematical Statistics and Probability*, **4**, 831-853, University of California Press, Berkeley.

66. El Barmi, H., Kochar, S. C. and Tsimikas, J. (2006). Likelihood ratio tests for and against ordering of the cumulative incidence functions in multiple competing risks and discrete mark variable models. *Journal of Statistical Planning and Inference*, **136**, 1588-1607.

67. El Barmi, H., Kochar, S. C. and Mukerjee, H. (2008). Order restricted inference for comparing the cumulative incidence of a competing risk over several populations. *IMS Collections- Beyond parametrics in interdisciplinary research: Festschrift in Honor of Professor Pranab. K. Sen*, **1**, 50-61.

68. Eleuteri, A., Tagliaferri, R., Milano, L., De Placido, S. and De Laurentiis, M. (2003a). A novel neural network based survival analysis model. *Neural Networks*, **16,** 855-864.

69. Eleuteri, A., Tagliaferri, R., Milano, L., Sansone, G., D'agostino,D., De Placido, S. and De Laurentiis, M. (2003b). Survival analysis and neural networks. *Proceedings of the International Joint Conference on Neural Networks*, **4,** 2631-2636.

70. Faraggi, D. and Simon, R. (1995). A neural network model for survival data. *Statistics in Medicine*, **14**, 73–82.

71. Farley, T. M. M., Ali, M. M. and Slaymaker, E. (2001). Competing approaches to analysis of failure times with competing risks. *Statistics in Medicine*, **20**, 3601-3610.

72. Feigl, P. and Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information. *Biometrics*, **21**, 826-838.

73. Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, **94**, 496-509.

74. Flehinger, B. J., Reiser, B. and Yashchin, E. (2002). Parametric modeling for survival with competing risks and masked failure causes. *Lifetime Data Analysis*, **8**. 177-203

75. Fleming, T. R. and Harrington, D. P (1991). *Counting Processes and Survival Analysis*. Wiley, New York.

76. Flinn, C. J. and Heckman, J. J. (1983). Are unemployment and out of the labor force behaviourally distinct labor force states?. *Journal of Labor Economics*, **1**, 28-42.

77. Friedlin, B. and Korn, E. L. (2005). Testing treatment effects in the presence of competing risks. *Statistics in Medicine*, **24**, 1703-1712.

78. Froda, S. (1987). A signed rank test for censored paired lifetimes. *Communications in Statistics- Theory and Methods*, **16**, 3497-3517.

79. Gasbarra, D. and Karia, S. R. (2000). Analysis of competing risks by using Bayesian smoothing. *Scandinavian Journal of Statistics*, **27,** 605-617.

80. Gasbarra, D., Kulathinal, S. B. and Nissinen, A. (2006). Testing dependence between the failure time and failure modes: An application of enlarged filtration. *Journal of Statistical Planning and Inference*, **136**, 1669-1686.

81. Gehan, E. A. (1965). A generalized two-sample Wilcoxon test for doubly censored data. *Biometrika*, **52**, 650-653.

82. Gelfand, A. E., Ghosh, S. K., Christiansen, C., Soumerai, S. B. and McLaughlin, T. J. (2000). Proportional hazards models: a latent competing risk approach. *Journal of the Royal Statistical Society- Series C*, **49**, 385-397.

83. Gilchrist, W (2000). *Statistical Modelling with Quantile Functions.* Chapman and Hall, CRC, USA.

84. Gill, R. D. (1980). *Censoring and Stochastic Integrals.* Mathematical Centre Tracts 124, Mathematisch Centrum, Amsterdam.

85. Gill, R. D. (1989). Non and semiparametric maximum likelihood estimators and the Von Mises method (Part-1). *Scandinavian Journal of Statistics*, **16**, 97-128.

86. Gimenez, O., Crainiceanu, C., Barbraud, C., Jenouvrier, S. and Morgan, B. J. (2006). Semiparametric regression in capture-recapture modeling. *Biometrics*, **62**, 691-698.

87. Glasser, M. (1967). Exponential survival with covariance. *Journal of the American Statistical Association*, **62**, 561-568.

88. Goetghebeur, E. and Ryan, L. (1995). Analysis of competing risks survival data when some failure types are missing. *Biometrika*, **82**, 821-833.

89. Gooley, T. A., Leisenring, W., Crowley, J. and Storer, B. E. (1999). Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Statistics in Medicine*, **18**,695-706.

90. Gray, R. J. (1988). A class of k sample tests for comparing the cumulative incidence of a competing risk. *The Annals of Statistics*, **16**, 1141-1154.

91. Gu, M. G. and Zhang, C. H. (1993). Asymptotic properties of self- consistent estimators based on doubly censored data. *The Annals of Statistics*, **21**, 611-624.

92. Hajek, J. and Sidak, Z. (1967). *Theory of Rank Tests*. Academic Press, New York.

93. Harrington, D. P. and Fleming, T. R. (1982). A class of rank test procedures for censored survival data. *Biometrika*, **69**, 553-566.

94. Heckman, J. J. and Honore, B. E. (1989). The identifiability of the competing risks model. *Biometrika*, **76**, 325-330.

95. Henschel, V Engel, J., Holzel, D. and Mansmann, U. (2009). A semiparametric Bayesian proportional hazards model for interval censored data with frailty effects. *BMC Medical Research Methodology*, **9**, doi:10.1186/1471-2288-9-9.

96. Herman, R. J. and Patel, R. K. N. (1971). Maximum likelihood estimation for multi-risk model. *Technometrics*, **13**, 385-396.

97. Hettmansperger, T. P. (1984). *Statistical Inference Based on Ranks*. John Wiley and Sons, New York.

98. Hinds, P. A. (1996). *Reliability Analysis and Techniques for Equipment Management*. M Sc Dissertation, City University, London.

99. Hjort, N. L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *The Annals of Statistics*, **18**, 1259-1294.

100. Hoel, D. G. (1972). A representation of mortality data by competing risks. *Biometrics*, **28**, 475-488.

101. Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. Springer-Verlang, New York.

102. Hu, X. S. and Tsai, W. Y. (1999). Linear rank tests for competing risks model. *Statistica Sinica*, **9**, 971-983.

103. Ibrahim, J. G., Chen, M. H. and Sinha, D. (2001a). Bayesian semiparamteric models for survival data with a cure fraction. *Biometrics*, **57**, 383-388.

104. Ibrahim, J. G., Chen, M. H. and Sinha, D. (2001b). *Bayesian Survival Analysis*. Springer-Verlang, New York.

105. Ibrahim, J. G., Chen, M. H. and Sinha, D. (2004). Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials. *Statistica Sinica*, **14**, 863-883.

106. Jeong, J. H. and Fine, J. P. (2009). *Parametric regression on cumulative incidence function. Biostatistics*, **8**, 184-196.

107. Johnson, R. A. and Mehrotra, K. G. (1972). Locally most powerful rank tests for the two sample problem with censored data. *The Annals of Mathematical Statistics*, **43**,823-831.

108. Kalbfleisch, J. D. (1978). Nonparametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society-Series B*, **40**, 214-221.

109. Kalbfleisch, J. D. and Lawless, J. F. (1988). Estimation of reliability from field performance studies (with discussion). *Technometrics*, **30**,365-388.

110. Kalbfleisch, J. D. and Lawless, J. F. (1989). Inference based on retrospective ascertainment: An analysis of the data on transfusion related AIDS. *Journal of the American Statistical Association*, **84**,360-372.

111. Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. John Wiley and Sons, New York.

112. Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**, 457-481.

113. Kappen, H. J. and Neijt, J. P. (1993). Neural network analysis to predict treatment outcome. *Annuals of Oncology*, **4**, Supplement, S31-S34.

114. Klein, J. P. and Moeschberger, M. L. (2003). *Survival Analysis- Techniques for Censored and Truncated Data*. Springer-Verlang, New York.

115. Kochar, S. C. (1995). A review of some distribution free tests for the equality of cause specific hazard rates. *Analysis of Censored Data-IMS Lecture Notes*, Monograph Series, **27**,147-162.

116. Kochar, S. C., Lam, K. F. and Yip, P. S. F. (2002). Generalized supremum tests for the equality of cause specific hazard rates. *Lifetime Data Analysis*, **8**,277-288.

117. Kochar, S. C. and Proschan, F. (1991). Independence of time and cause of failure in the multiple dependent competing risks model. *Statistica Sinica*, **1**, 295-299.

118. Kodell, R. L. and Chen, J. J. (1987). Handling cause of death in equivocal cases using the em algorithm. *Communications in Statistics -Theory and Methods*, **16**, 2565-2585.

119. Kulathinal, S. B. and Gasbarra, D. (2002). Testing equality of cause specific hazard rates corresponding to m competing risks among k groups. *Lifetime Data Analysis*, **8**, 147-161.

120. Kvam, P. H. and Singh, H. (2001). On nonparametric estimation of the survival function with competing risks. *Scandinavian Journal of Statistics*, **28**, 715-724.

121. Lam, K. F. (1998). A class of tests for the equality of k cause specific hazard rates in a competing risks model. *Biometrika*, **85**,179-188.

122. Lapidus, G., Braddock, M., Schwartz, R., Banco, L. and Jacobs, L. (1994). Accuracy of fatal motorcycle injury reporting on death certificates. *Accident; Analysis and Prevention*, **26**, 535-542.

123. Larson, M. G. (1984). Covariate analysis of competing risks data with log-linear models. *Biometrics*, **40**, 459-469.

124. Laud, P. W., Damien, P. and Walker, S. G. (2006). Computations via auxiliary random functions for survival models. *Scandinavian Journal of Statistics*, **33**, 219-226.

125. Lawless, J .F. (2003). *Statistical Models and Methods for Lifetime Data*. John Wiley and Sons, New York.

126. Lin, D. Y. (1997). Nonparametric inference for cumulative incidence functions in competing risks studies. *Statistics in Medicine*, **16**, 901-910.

127. Lisboa, P. J. G., Etchells, T. A., Jarman, I. H., Arsene, C. T. C., Hane Aung, M.S., Eleuteri, A., Taktak, A.F. G., Ambrogi, F., Boracchi, P. and Biganzoli, E. (2009). Partial logistic artificial neural network for competing risks regularized with automatic relevance determination. *IEEE Transactions on Neural Networks*, **20,** 1403-1416.

128. Lu, K. and Tsiatis, A.A. (2005). Comparison between two partial likelihood approaches for the competing risks model with missing cause of failure. *Lifetime Data Analysis*, **11**, 29-40.

129. Machado, L. (1997). A comparison of Cox proportional hazards and artificial neural network models for medical prognosis. *Computers in Biology and Medicine*, **27,** 55-65.

130. MacKay, D. J. C. (1992). A practical Bayesian framework for back propagation networks. *Neural Computation*, **4**, 448–472.

131. Makeham, W M. (1874). On an application of the theory of the composition of decremental forces. *Journal of the Institute of Actuaries*, **18**, 317-322.

132. Mallick, B. K. and Walker, S. G. (2003). A Bayesian semiparametric transformation model incorporating frailties. *Journal of Statistical Planning and Inference*, **112**,159-174.

133. Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports*, **50**, 163-170.

134. Mantel, N. (1967). Ranking procedures for arbitrarily restricted observations. *Biometrics*, **23**, 65-78.

135. Martz, H. F. and Waller, R. A. (1982). *Bayesian Reliability Analysis*. Wiley, New York.

136. Matsuyama, Y. and Yamaguchi, T. (2008). Estimation of the marginal survival time in the presence of dependent competing risks using inverse probability of censoring weighted (IPCW) methods. *Pharmaceutical Statistics*, **7**, 202-214.

137. McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models*. John Wiley and Sons, New York.

138. Mendenhall, W. and Hader, R. J. (1958). Estimation of parameters of mixed exponential distributed failure time distributions from censored life test data. *Biometrika*, **45**, 504-520.

139. Moeschberger, M. L. (1974). Life tests under dependent competing causes of failure. *Technometrics*, **16**, 39-47.

140. Moeschberger, M. L. and David, H. A. (1971). Life tests under competing causes of failure and the theory of competing risks. *Biometrics*, **27**, 909-933.

141. Moeschberger, M. L. and Klein, J. P. (1995). Statistical methods for dependent competing risks. *Lifetime Data Analysis*, **1**, 195-204.

142. Molinari, N. (2005). A U-statistic test in competing risk models. *Comptes Rendus Mathematique*, **341**, 317-322.

143. Mykland, P. A. and Ren, J. J. (1996). Algorithms for computing self-consistent and maximum likelihood estimators with doubly censored data. *The Annals of Statistics*, **24**, 1740-1764.

144. Nair, N. U., Sankaran, P.G. and Kumar, V. B. (2008). Total time on test transforms of order n and their implications in reliability analysis. *Journal of Applied Probability*, **45**, 1126-1139.

145. Nair, N. U. and Sankaran, P. G. (2009). Quantile based reliability analysis. *Communications in Statistics- Theory and Methods*, **38**, 222-232.

146. Naskar, M., Das, K. and Ibrahim, J. G. (2005). A semiparametric mixture model for analysing clustered competing risks data. *Biometrics*, **61**, 729-737.

147. Neal, R. M. (1996). *Bayesian Learning for Neural Networks*. Springer-Verlang, New York.

148. Neal, R. M. (2001). Survival analysis using a Bayesian neural network. *Joint Statistical Meetings Report*, Atlanta.

149. Neito-Barajas, L. E. and Walker, S. G. (2002). Markov beta and gamma processes for modelling hazard rates. *Scandinavian Journal of Statistics*, **29**,413-424.

150. Nieto-Barajas, L. E. and Walker, S. G. (2005). A semiparametric Bayesian analysis of survival data based on levy- driven processes. *Lifetime Data Analysis*, **11**,529-543.

151. Nelson, W B. (1972). Graphical analysis of accelerated life test data with the inverse power law model. *IEEE Transactions on Reliability*, **R-21**, 2-11.

152. Neuhaus, G. (1991). Some linear and non linear rank tests for competing risks models. *Communications in Statistics-Theory and Methods*, **20**, 667-701.

153. Nowinski, R. C., Brown, M., Doyle, T. and Prentice, R. L. (1979). Genetic and viral factors influencing the development of spontaneous leukemia in AKR mice. *Virology*, **96**, 186-204.

154. Padgett, W. J. (1986). A kernel type estimation of a quantile function from right censored data. *Journal of the American Statistical Association*, **81**, 215-222.

155. Parzen, E. (1979). Nonparametric statistical data modeling (with discussion). *Journal of the American Statistical Association*, **74**, 105-131.

156. Peng, L. and Fine, J. P (2007). Nonparametric quantile inference with competing risks data. *Biometrika*, **94**, 735-744.

157. Pepe, M. S. and Mori, M. (1993). Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data?. *Statistics in Medicine*, **12**, 737-751.

158. Peto, R. (1973). Experimental survival curves for interval censored data. *Applied Statistics*, **22**, 86-91.

159. Peto, R. and Peto, J. (1972). Asymptotically efficient rank invariant test procedures (with discussion). *Journal of the Royal Statistical Society- Series A*, **135**, 185-206.

160. Peto, R., Pike, M C., Armitage, P., Breslow, N. E., Cox, D. R., Howard, S. V., Mantel, M rson. Peto, J. and Smith, P. G. (1976). Design and analysis rand zed c cal trials requiri pro ged observation of each patient. I. introdu on and design. *British Jou al of ncer*, **34**,585-612.

161. Peto, R., P ke, M Armitage, P., Breslow, N ., Cox, D. R., Howard, S V Mantel, N McP erson, K., Peto, J. and Smith, P. G. (1977). Design and analysis of ndomized clinical trials requiring prolonged observation of each patient. II. analysis and examples. *British Journal of Cancer*, **35**, 1-39.

162. Pollard, D. (1990). *Empirical Processes: Theory and Applications*. Institute of Mathematica stics, Hayward, California.

163. Porta, N., C , G., Calle, M. L. and Malats, N. (2007). Competing risks iods. htt pcommons.upc.edu/e-ints/bitstre m/2117/2201/1/TR_CR.pdf.

164. Porta, N., Gomez, G. and Calle, M. L. (2008). The role of survival functions in competing risks. http://upcommons.upc.edu/e-prints/handle/2117/2202.

165. Prentice, R. L. (1973). Exponential survival with censoring and explanatory variables. *Biometrika*, **60**, 279-288.

166. Prentice, R. L., Kalbfleisch, J. D., Peterson, A.V Jr., Flournoy, N., Farewell, V.T. and Beslow E. (978). The analysis of failure times in the presence of competing risks. *Biometrics*, **34**, 541-554.

167. Ravdin, P. M. and Clark, G. M. (1992). A practical application of neural network analysis for predicting outcome of individual breast cancer patients. *Breast Cancer Research and Treatment*, **22**, 285-293.

168. Reid, N. (1981). Estimating the median survival time. *Biometrika*, **68**, 601-608.

169. Ripley, B. D. (1996). *Pattern Recognition and Neural Networks.* Cambridge University Press, Cambridge.

170. Ripley, B. D. and Ripley, R. M. (2001). Neural networks as statistical methods in survival analysis. *Clinical Applications of Artificial Neural Networks, Eds: R. Dybowsky and V. Gant,* Lands Biosciences Publishers.

171. Ripley, R. M., Harris, A. L. and Tarassenko, L. (2004). Nonlinear survival analysis using neural networks. *Statistics in Medicine*, **23,** 825-842.

172. Ripley. R. M. (1998). *Neural network models for breast cancer prognosis.* PhD Thesis, Department of Engineering Science, University of Oxford.

173. Sampford, M. R. (1952). The estimation of response-time distributions .II. Multi-stimulus distributions. *Biometrics*, **8**, 307-369.

174. Sankaran, P. G. and Nair, N. U. (2009). Nonparametric estimation of hazard quantile function. *Journal of Nonparametric Statistics*, **21**, 757- 767.

175. Sankaran, P. G. and Sreedevi, E. P. (2010a). A Semiparamertic regression model for doubly censored competing risks data. Communicated.

176. Sankaran, P. G. and Sreedevi, E. P. (2010b). Neural network models for competing risks data. Communicated.

177. Sankaran, P. G., Dewan, I. and Sreedevi, E. P.(2009). A class of general tests for testing independence of failure time and cause of failure in a competing risks model. Technical Report, Indian Statistical Institute, New Delhi

178. Sankaran, P.G., Nair, N. U. and Sreedevi, E. P. (2010). A quantile based test for comparing cumulative incidence functions. *Statistics and Probability Letters* (to appear).

179. Savage, I. R. (1956). Contributions to the theory of rank order statistics-the two sample case. *The Annals of Mathematical Statistics*, **27**, 590-615.

180. Sen, P. K. (1979). *N*onparametric tests for interchangeability under competing risks. *In Contributions to Statistics, Jaroslav Hajek Memorial Volume (J. Jureckova, ed.)*, 211-228, Reidel, Dordrecht.

181. Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

182. Sinha, D. and Dey, D. K. (1997). Semiparametric Bayesian analysis of survival data. *Journal of the American Statistical Association*, **92**, 1195-1212.

183. Sinha, S. K. (1986). *Reliability and Life Testing*. Wiley Eastern Limited, New Delhi.

184. Slud, E. V Byar, D. P and Green, S.B. (1984). A comparison of reflected versus test-based confidence intervals for the median survival time, based on censored data. *Biometrics*, **40**, 587-600.

185. Smith, A. E. and Anand, S. S. (2000). Patient survival estimation with multiple variables: Adaption of Cox's regression to give an individual's point prediction. *Proceedings of the IDAMAP*, Berlin, 47-55.

186. Sreedevi, E. P. and Sankaran, P G. (2010). A semiparametric Bayesian approach for the analysis of competing risks data. *Communications in Statistics- Theory and Methods* (to appear).

187. Stanish, W M., Chi, G. Y. H., Johnson, W D., Koch, G. G., Landis, J. R. and Liu-Chi, S. (1978). A computer program for the generalised chi-square analysis.of competing risks grouped survival data. *Computer programs in Biomedicine*, **8**,208-223.

188. Street, W N. (1998). A neural network model for prognostic prediction. *Proceedings of the Fifteenth International Conference on Machine Learning*, San Francisco, 540-546.

189. Struthers, C. A. (1984). *Asymptotic properties of linear rank tests with censored data*. Unpublished PhD Thesis, University of Waterloo, Waterloo, Ontario, Canada.

190. Su, J. Q. and Wei, L. J. (1993). Nonparametric estimation for the difference or ratio of median failure times. *Biometrics*, **49**, 603-608.

191. Sun, Y (2001). Generalized nonparametric test procedures for comparing multiple cause specific hazard rates. *Journal of Nonparametric Statistics*, **13**,171-207.

192. Sun, Y. and Tiwari, R. C (1995). Comparing cause specific hazard rates of a competing risks model with censored data. *Analysis of Censored Data, IMS Lecture Notes-Monograph Series*, **27**, 255-270.

193. Sun, Y. and Tiwari, R. C. (1997). Comparing cumulative incidence functions of a competing risks model. *IEEE Transactions on Reliability*, **46**, 247-253.

194. Tiwari, R. C., Kulasekera, K. B. and Park, C. (2006). Nonparametric tests for cause specific hazard rates with censored data for competing risks among several groups. *Journal of Statistical Planning and Inference*, **136**, 1718-1745.

195. Tsai, W. Y. and Crowley, J. (1985). A large sample study of generalized maximum likelihood estimators from incomplete data via self-consistency. *The Annals of Statistics*, **13**, 1317-1334.

196. Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of National Academy of Sciences*, USA, **72**, 20-22.

197. Turnbull, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *Journal of the American Statistical Association*, **69**, 169-173.

198. Wada, C. Y. and Sen, P. K. (1995). Restricted alternative test in a parametric model with competing risks data. *Journal of Statistical Planning and Inference*, **44**,193-203.

199. Walker, S. G. and Mallick, B. K. (1997). Hierarchical generalized linear models and frailty models with Bayesian nonparametric mixing. *Journal of the Royal Statistical Society-Series B*, **59**, 845-860.

200. Wellner, J. A. and Zhan, Y (1997). A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data. *Journal of the American Statistical Association*, **92**, 945-959.

201. Wells, M. T. (1994). Nonparametric kernel estimation in counting processes with explanatory variables. *Biometrika*, **81,** 795-801.

202. Williamson, P. R., Kolamunnage-Dona, R. and Tudur Smith, C. (2006). The influence of competing risks settings on the choice of hypothesis test for treatment effect. *Biostatistics*, **8**, 689-694.

203. Yip, P. S. F. and Lam, K. F. (1992). A class of nonparametric tests for the equality of failure rates in a competing risks model. *Communications in Statistics- Theory and Methods*, **21**, 2541-2556.

204. Yip, P. S. F. and Lam, K. F. (1993). A multivariate nonparametric test for the equality of failure rates in a competing risks model. *Communications in Statistics- Theory and Methods*, **22**, 3199-3222.

205. Zhou, M. (2004). Nonparametric Bayes estimator of survival functions for doubly/interval censored data. *Statistica Sinica,* **14**, 533-546.

206. Zippin, C. and Armitage, P. (1966). Use of concomitant variables and incomplete survival information in the estimation of an exponential survival parameter. *Biometrics*, **22**, 665-672.