QATAR UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

COMPETING RISKS MODEL BASED ON FINE AND GRAY IN PRESENCE OF INTERVAL CENSORED DATA

BY

HAMDA HAMDAN AL ABDULLA

A Thesis Submitted to

the College of Arts and Sciences

in Partial Fulfillment of the Requirements for the Degree of

Masters of Science in Applied Statistics

January 2022

COMMITTEE PAGE

The members of the Committee approve the Thesis of
Hamda Hamdan Al Abdulla defended on 05/12/2021.

_____
Dr. Faiz Ahmed Elfaki
Thesis/Dissertation Supervisor

_____
Dr. Reza Pakyari
Committee Member

_____
Dr. Mohammad Salehi
Committee Member

Approved:

_____
Ahmed Elzatahry, Dean, College of Arts and Sciences

ABSTRACT

ALABDULLA, HAMDA, HAMDAN., Masters : January : [2022:], Applied Statistics

Title: Competing Risks Model Based on Fine and Gray in Presence of Interval Censored Data

Supervisor of Thesis: Dr. Faiz Ahmed Elfaki.

Generally, survival analysis is a significant aspect of statistics that helps in anticipating possible outcomes in the various phenomena of study. A competing risk model is widely used in survival analysis since it not only studies the event of interest but also studies the other possible outcomes and this is the main topic of this research. Various models have been developed by statisticians and are widely used in examining competing risks in real-life phenomena where each model seems to have its strength and weaknesses. The Fine and Gray model is a largely employed method in competing risks analysis for its various advantages, such as the accuracy and the ability to consider multiple competing events. The main goal of this thesis is to analyze the effect of covariate on the cumulative incidence function, the Cox proportional hazards model for the subdistribution is used on both right-censored (RC) data and the model for interval-censored (IC) data. We simulate competing risks data, then we use midpoint imputation to handle the simulated interval-censored and right-censored data. In comparison to the Fine & Gray model with interval -censored data, the simulation results show that our model in this study is applicable and performs well. In additional to that both methods were applied to the MERS data set and the results of the two models show that the covariates have no effect on the cumulative incidence function.

DEDICATION

This thesis is dedicated to my husband, who has been a constant source of encouragement and support throughout my graduate studies and life. I'm grateful for your presence in my life. This work is also dedicated to my family, who have always supported me in my endeavors.

ACKNOWLEDGMENTS

First and foremost, praises and thanks to Allah. Also, I would like to acknowledge and thank to my supervisor Dr. Faiz Elfaki who made this work possible. His guidance and advice carried me through all stages of writing my thesis. Besides, I would like to express my gratitude to my committee members Dr. Reza Pakyari and Dr. Mohammad Salehi for making my defense a pleasurable experience, as well as for your insightful comments and suggestions. Special thanks to my husband and my family as a whole for their support, encouragement, help and understanding when undertaking my research and writing my thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE INTRODUCTION

CHAPTER OVERVIEW

In this chapter we give a general introduction to survival analysis, main types of censoring and introduction to Competing Risks model, followed by a brief introduction to Fine & Gray (FG) model based on Cumulative Incident Function (CIF). Also, we present the background, the statement of the problem and the objective of the research.

1.1 INTRODUCTION

1.1.1 SURVIVAL ANALYSIS

Statistics as a discipline and practice deals with the collection, analysis, interpretation and presentation of data from different sources such as economics, business, medicine, social science and others. In analyzing data, survival analysis as a branch of statistics is largely employed for its effectiveness. Survival analysis or time to failure analysis is a collection of statistical data analysis techniques where the outcome variable of interest is the time to the occurrence of an event. Death, disease, equipment failure, or a complex system failure are examples of events.

In many clinical biomedical and epidemiologic studies, survival analysis is commonly used. As an example, Cornelia Liedtke et al. (2008) studied the effect of neoadjuvant chemotherapy on triple-negative breast cancer (TNBC) patients. They compared the survival time, time from surgery until death, of TNBC patients to non-TNBC patients and they discovered that TNBC patients have a lower survival rate than non-TNBC patients.

Guillermo Salinas-Escudero et al. (2020) used survival analysis to look into the risk factors associated with COVID-19 deaths in the Mexican population. As a result, they found that men, those in older age groups, people with chronic kidney disease, and patients hospitalized in public health facilities all had a higher risk of dying at any point throughout the study.

In longitudinal studies, individuals who don't experience the event of interest during the follow-up period or withdraw for some reasons. One appealing feature of survival analysis is the ability to include data from censored observations until they are removed from the risk set. In general, right, left and interval-censored are the three main types of censoring. A subject is right censored if failure occurs sometime after the recorded follow-up period and left censoring occurs when it is known that the failure occurred prior to the recorded follow-up period. Nevertheless, if the event occurs between two times but the precise time of failure is unknown the subject is interval censored (Mark Stevenson (2007); Hudgens et al (2014)).

1.1.2 COMPETING RISKS

In survival analysis, more than one event may be considered in the same analysis, in this case the statistical problem can be characterized as a competing risks problem. Competing risks (CR) are said to be present when an object is at risk of more than one mutually exclusive event, such as death from different causes, and the occurrence of one of these will prevent any other event from ever happening. Because of its proven effectiveness and differentiation from other models, the competing risks is a commonly used method in survival analysis. When conducting research to determine the frequency and relevant data about a specific event among many other

competing events in the same phenomenon, the competing risks model is used (Fine & Gray 1999). This method beats other previously used methods but could not provide only a single event of interest among the various competing events. The method is also known for the accurate results associated with its application instead of other methods that sometimes produce inaccurate results. In most cases, there is more than one possible occurrence in each phenomenon where only one event can occur at a time. The fact that only one event can occur of the many possible events implies that the many or the two events are competing.

For instance, assuming that the main aim of the research is to find out the cause of deaths in hospitals, there are various possible outcomes. The first cause of death could be severe accidents, and diseases that doctors cannot successfully manage are common causes of death. Secondly, deaths can be caused by wrong medication or error in diagnosis, or errors in surgical procedures in the theatre. Another cause of death in the hospital could be the various infections contracted by the patients while in the wards. The death of a patient while in the hospital can result from the three competing causes or events. These three competing events can therefore be projected using the competing risks model, which helps in analyzing individual events among other competing events and factors. The fact explains the essence of the competing risks model in analyzing, interpreting, and projecting the frequency of events in survival analysis as a major branch of statistics.

Competing risks analysis is a subset of survival analysis that seeks to accurately estimate the marginal probability of an event in the presence of competing events. Kaplan Meier product-limit method that is used in standard survival analysis produce inaccurate estimates when analyzing the marginal probability for cause-specific events in the present of competing risk. As a workaround, the CIF was proposed to tackle this problem by calculating the marginal probability of a certain event as a function of its cause-specific probability and overall survival probability. This method combines the product-limit approach and competing causal pathways to offer a more interpretable estimate for a group of subjects' surviving experience of several competing events.

In the next two sections will introduce the (CIF) and Fine and Gray Model.

1.1.3. CUMULATIVE INCIDENCE FUNCTION (CIF)

It's worth noting that when we talk about the crude rate, we're talking about how quickly people are dying from cause 1 among those who could die from any cause at time t. This is a real-world event rate where an individual could fail for any number of reasons. In competing risks theory, this crude rate is called hazard rate and it is given by John et al., (2014) as;

$$h_1(t) = \lim_{\Delta t \to 0} \left(\frac{P_r[t < T \leq t + \Delta t, \delta = 1 | T \geq t}{\Delta t} \right)^n$$

The CIF is alternative to the crude hazard rate. This function is defined as

$$C_1(t) = \int_0^t h_1(z)\, S(z)\, dz$$

Where $h_1(z)$ is the hazard rate of cause 1 and $S(t)$ is the survival function of time t.

1.1.4 FINE AND GRAY MODEL

Gray (1988) proposed a K-sample test for CIF, which allows for direct inference about the functions. The effects of this work were later quantified in the Fine-Gray model in a regression setting (Fine and Gray, 1999).

According to the Fine-Gray model, the cumulative incidence for cause j and subject i is defined by

$$F_j(t; Z_i) = 1 - exp\ \{-\Lambda_0(t).exp(\beta_{0j}.Z_i)\}, i = 1, \dots, n,$$

where β_{0j} is a 1 x p vector of regression coefficients, related to j'th cause, Z_i is a vector of *p x 1* covariates for individual i, and $\Lambda_0(t)$ is an unspecified, non-decreasing baseline with $\Lambda_0(0) = 0$. This model resembles Cox's regression in some ways and was developed as a Cox model type based on the subdistribution hazard rate.

More generally, any link-function can be used to assess covariate effects directly on the CIF:

$$F_j(t; Z_i) = \text{h}\ \left(-\Lambda_0(t), \beta_{0j}, Z_i\right)\ i = 1, \dots, n, \qquad (1.1)$$

Additionally, calculate a non-decreasing baseline $\Lambda_0(t)$ and regression coefficient β_{0j}. The link then provides the Fine-Gray model as given below:

$$h_{fg}(a, b, z) = 1 - \exp\ (a\ \exp(bz)), \qquad (1.2)$$

where b is a regression coefficient, z is a covariate, and a is a non-decreasing baseline.

Fine and Gray (1999) took the approach of taking into account a time of subdistribution until the competing risks occurrence for type 1 as;

$$\tilde{T} = \inf\{t > 0 | Z_t = 1\}$$

If and only if $Z_t = 1$, this equals the real-life event time, T. Otherwise, the time required for subdistribution is infinite. Then, for $t\ \in [0, \infty)$, the subdistribution time will be;

$$P(T \leq t, Z_T = 1).$$

Fine and Gray suggested subdistribution hazard $\lambda(t)$ for fitting a Cox's model as;

$$\lambda(t) = -\frac{d}{dt} log\big(1 - P(T \leq t, Z_T = 1)\big) = \frac{P(T>t)}{1-P(T\leq t, Z_T=1)}\alpha_{01}(t) \quad (1.3)$$

The CIF is measures a direct effect on type 1 events as;

$$P(T \leq t, Z_T = 1) = 1 - \exp(-\int_0^t \lambda(u)du). \quad (1.4)$$

1.2 PROBLEM STATEMENT

Survival analysis is a collection of statistical methods for the analysis of data for which the variable outcome of interest is time until an event occurs. This type of data is commonly referred to as failure time data or lifetime data. There are various techniques used to analyze failure time data under left or right censoring. However, when competing risks (CR) are addressed in the presence of interval-censored data, there are few techniques that have been used to look into the effect of the explanatory variable on the response. Competing risks are said to be present when an object is at risk of two or more mutually exclusive events, for example failure due to different causes, and the occurrence of one cause of these is independent from the others.

In this thesis, we are interested in analyzing the effect of risk factors (covariates) on cumulative incident function (CIF) in the presence of interval-censoring. In addition to that, the performance of Fine and Gray (FG) methods will be evaluated by simulated interval censored data and secondary data set.

1.3 OBJECTIVE

The objective of this study is to evaluate and derive the MLEs of the parameters for FG method based on CR when the data are interval-censored. When the parameters are estimated, the performances of theses estimators will be evaluated. The effective of the covariates will be checked. Additionally, to apply and analyze these inferential procedures on secondary data and simulated data. Hence, the major objectives of this research are:

1. Estimate the parameters of FG models with interval-censored data using MLE and imputation methods.
2. Compare the performance of midpoint imputation methods for FG model with interval censored data.
3. Compare the performance of FG with interval censored data and FG with right censored data via simulation study.
4. Applying the proposed techniques to real medical data.

CHAPTER TWO LITERATURE REVIEW

CHAPTER OVERVIEW

In this chapter, we will provide literature review on survival analysis, censored data, competing risks model, Fine and Gray model and cumulative incidence function.

According to Hansen et al. (2017), survival analysis comprises investigating the time of event statistics. It is a branch of statistics that analyzes the expected duration for an event such as death. It attempts to answer critical questions regarding a specific event such as the proportion of a population which will survive within a particular time. The analysis can also identify the rate at which the surviving population will die (Hansen et al., 2017). Besides, the survival analysis will be used to examine how particular circumstances will raise or lower the probability of survival. The general concept in survival analysis implies that the survivor function is the integral focus of the clinical studies, which comprises the probability of non-occurrence of the event up to a particular time.

According to Austin & Fine (2017), survival analysis focuses on the expected time frame until an event occurs, either death or relapse. However, an event may not be observed for an individual in the study to constitute the censored observations. Censoring may therefore arise from the individual not experiencing the event within the study period. Also, an individual could be lost to follow-up during the research duration. Additionally, an individual may experience a different event, which makes follow-up a challenge to handle. Based on Deng (2016), censoring is common in survival analysis. It entails the missing data in which time to event is not recorded for

varied reasons such as termination of study or the subject leaving the study (Hansen et al., 2017). According to Deng (2016), right censoring is the most common type of censoring and the easiest to compute in the analysis process. It occurs when an individual has been followed from a particular time up to a given time, but they have entirely participated in the study. The right censoring may occur when an individual decides not to participate in the research study before completing the event of interest. There are two types of right censoring; fixed type one and type two censoring.

The fixed type one occurs when the research is tailored to be completed after a specific time. However, the individuals who may fail to be observed during the fixed time are censored at the fixed time. In type II censoring, the research ends when there is a predetermined. Notably, regardless of the type of censoring, it is caused by other things than failure. Left censoring is the other type that is central to the right censoring but involves missing data elements. The left censoring is considered when an individual undertakes an event before a specified time but occurs before the period of censoring (Zhang et al., 2018). The interval censoring is also possible when an individual partakes in an event between two durations, but the exact time is not recorded.

A competing risk is an occurrence that hinders the observation of the event of interest or changes the chance of the event happening. In typical survival data, subjects are expected to experience an event over follow-up. For instance, in breast cancer, the event of interest will be death; any other event would be considered as competing event or risk. In real situations, subjects may experience more than one particular event type contrary to the survival data (Heckman & Honoré (1989); Pintilie (2006); Austin & Fine (2017) and Austin et al. (2021)). Therefore, when multiple events occur when only

one event is expected, the events are termed as competing risks. In such situations, the competing events will contest each other to cause the desired event. However, the occurrence of one event will inhibit the occurrence of the selected event.

Fine and Gray (1999) emphasize the need to model cause-specific functions using similar hazard assumptions as the criterion for applying the explanatory characteristics in the competing risks data. However, while studying a particular failure type, there is no direct link between cause-specific functions and the specific survival probability. It is worth notable that there is a significant rise in the application of the marginal failure possibilities for specific events and the cumulative incident function in working the competing risk models in survival analysis (Fine & Gray 1999). The application of the marginal failure probability and the cumulative incidents, not function, is effective owing to the ease of application and the fact that nearly everyone can easily understand them. Further, the cumulative incident function is preferred since it is cost-effective as the survival probabilities can be applied to determine the most effective way to promote the chances of survival. Fine and Gray (1999) cite a significant challenge often not addressed in the various models that seek to work out the survival analysis. They cite that the various analysis does not allow the researchers to directly analyze the effects of the characteristics on the marginal probability function. The model thus proposes weighting techniques and the partial probability likelihood in assessing the semi-parametric proportional hazard functions. The model also gives a similar and consistent method of assessing and predicting the cumulative incident for a specific event of interest. This method thus introduces not only new concepts and techniques in analysis but also uses such techniques to correct the weaknesses of other methods previously used in competing for risk analysis.

Scheike and Zhang (2008) proposed what they termed as a more straightforward method of assessing and analyzing the effects of the characteristics in the cumulative incident curve in working out competing risks. This approach introduces a simplified estimator that can be fed in standard programmed software to assess the effects of the covariates automatically. This model considers both the effects that vary with time and constant effects throughout the research period (Scheike & Zhang (2008) and Chenxi (2016)). Compared to the sub-distribution approach, the method proves that it can provide finite characteristics as proven in the simulation trials of the system, making it appear better than other methods. This method further analyses the survival chances by introducing a complete remission against the competing risks, though with more concentration on the event of interest. The method involves modeling all the cause-specific hazards and estimating the cumulative incident curve based on cause-specific hazards as modeled. It also applies the FG model and the direct links between the covariates and the cumulative incident curve. This model appears more proficient as it considers the none proportional hazards, which other competing risk models often ignore. This test is also accurate compared to other methods as it indicates the exact position of none proportionality in the cumulative incident curve. This method, as presented by Scheike and Zhang (2008), improves the FG model as it introduces the application of flexible regression models in analyzing competing risks while considering the none proportional covariates.

Jeong and Fine (2007) and Shayan et al. (2011) formulate a parametric regression analysis around CIF's about the competing risk models in survival analysis. The distribution of the events of interest or the events under study uses the Gompers distribution or the improper baseline sub-distribution. This method is concerned with

analyzing the maximum probability of each competing risk to determine the chances of occurrence of the event of interest (Jeong et al. (2006) and Jeong & Fine (2007)). This method considers the cause-specific characteristics and the likelihood of the occurrences of events as linked to the chances of the desired event. While estimating the long-term probability of cases with cause-specific events, the approach takes a straightforward presentation in the parametric setting. For instance, when analyzing the deaths of patients with hypertension in the hospital, the parametric regression method produces empirical results. This method analyses the probability of the various possible events around the fate of hypertension patients in the hospital, including the desired event. The model considers the maximum likelihood of recovery of such patients based on the cause-specific scenarios around recovery as a possible event. The projective regression model, as proposed by Jeong and Fine (2007), also considers the maximum probability of the death of the patients. It then applies the cumulative incident function of these outcomes and uses such data to project the maximum likelihood of the desired event.

Generally, regression models in competing risks effects are founded on balanced hazards models and simple hazard proportions. Such measures hardly conform with notions extracted from diagrams indicating total incidence functions in every level of the risk factors. Klein and Anderson (2005) thus illustrate a method that models the CIF's directly as drawn from the empirical analysis of previous cases (Klein & Andersen 2005). The technique is founded on the derived values as drawn from a jackknife statistic designed from the cumulative incidence curve. The pseudo values used in this technique are utilized in a general approximation equation that is burger used in obtaining estimations of model parameters in competing risk models. Klein and

Anderson (2005) thus examine the characteristics of the estimator and correlate the method in analyzing the effects of alternative events that may arise besides the likelihood of the occurrence of the event of interest.

The distribution of the events of interest or the events under study uses the Gompers distribution or the improper baseline sub-distribution. The technique is founded on the derived values as drawn from a jackknife statistic designed from the cumulative incidence curve. This method considers the cause-specific characteristics and the likelihood of the occurrences of events as linked to the chances of the occurrence of the desired event (Fine, 2001). For instance, in a random examination of the drug when treating an infection, patients can experience worse conditions or die from competing events. The analysis aims at describing the effect of the drug on the likelihood of recurrence or death from other causes, all of which are competing events. A semiparametric transformation model for the crude failure likelihoods of a competing risk, conditional on covariates, is thus more applicable in this case. The criterion is formulated to expand the standard method to survival data with dominant right censoring in survival analysis. This method achieves the approximation of the regression coefficients using a rank-based least-squares criterion. The conducted simulations indicate that the method functions satisfactorily with practical sample sizes and not extensive imaginary data.

Masked data Bayesian analysis in competing risk brackets is researched to examine the effect of covariates upon functions (hazard) where the time is precisely empirical for part of the subjects but not exactly identified to occur for the other objects (Yousif et al., 2020). These data are called partially interval-censored data and are

commonly a product of periodic assessment and inspection. Gamma and Dirichlet methods are deduced as initials for masking baseline hazards and probabilities. The distribution of the events of interest or the events under study applies the Gompers distribution or the improper baseline sub-distribution. The model also gives a similar and consistent method of assessing and predicting the cumulative incident for a specific event of interest. The application of the marginal failure probability and the cumulative incidents, not function, is effective owing to the ease of application and the fact that nearly everyone can easily understand them.

The random variate is the period to events like death, a recurring illness, or an aloof metastasis. Examples of interval-censored data occur in clinical studies that comprise recurrent investigation. In this case, a person set for pre-planned observations for a medically observable variation in illness or well-being status may miss some changes and come back in a different state (Anderson, 2017). Another instance arises in the AIDS cases where AIDS is determined based on blood testing done periodically and not constantly. Interval censoring data usually represents a random variable lying within an interval instead of being observed precisely.

Competing risks models estimate the marginal likelihood of an event in the presence of competing occurrences. Regression models for interval-censored data are associated with fundamental parametric regression models (Suhaini et al 2020). The standard maximum probability methods and the approximate coverage at the rate of nth root make inferences. First, the Fine- Gray subdistribution risk model is commonly used to approximate result occurrence over time in the event of competing hazards (Anderson, 2020). The method is preferred since it instantly relates covariates to the

accumulative incidence function (CIF). Parametric models include statistical programs but, each distributional imposes rigorous assumptions on the hazard function shape. An example is the Poisson log-linear model, where the hazard is presumed constant in some interval sets of the follow-up time.

Midpoint imputation applies when the periods between successive visits are short. If the width of the interval widens, there are issues. A midpoint restores each finite value, and analysis is done, if the mid-points are precise observations. (Suhaini et al 2020). However, the midpoint imputation leads to biased approximations especially, when the perceived intervals are too long.

Other imputation methods are multiple and combined imputation methods; multiple methods transform interval to right-censored data for typical methods. They clarify complicated scenarios and include; the uniform and weighted weight methods. In line with Suhaini et al (2020), the uniform weight method presumes that the actual failure time of a subject is equally distributed (Sj, Li<Ri, for j=1,…, m). The study worked out a pseudo-hazard and collapsing risk relying on equal weights. MI methods obtain test statistic and their variance grid like the attribution of an actual collapse under a similar presumption (Suhaini et al 2020). The study explained the weighted weight based on NPMLE from the initial data made by Turnbull's technique and utilized the NPMLE as ascription weights.

The combined model begins with imputing a specific time interval and uses the subdistribution hazards model to approximate the parameter and the measured survival function. The model conducts a repetition protocol between parameter approximation,

imputation, and the measured survival function (Suhaini et al 2020). Multiple imputed variates account for ambiguity within the period intervals. In addition, different survival models are befitted in sequence to impute the period for corresponding changes based on similar initial status. The multiple imputed methods are not actual MI methods since they don't use a Bayesian structure and no prior dispersal specification for the parameter (Suhaini et al 2020). Multiple imputed variates were used to account for ambiguity within period intervals but may not hold the same features as the MI method. Therefore, multiple imputation methods fill missing data to make a complete matrix.

CHAPTER THREE METHODOLOGY

CHAPTER OVERVIEW

This chapter will present the computation of FG competing risk model and MLE will be obtained under right-censored data. Furthermore, the proposed similar model under interval-censored data via imputation methods will be introduced. Derivation of the point estimate for the survival function for both models will be discussed.

3.1 COX PROPORTIONAL HAZARD (CPH) MODEL

The distribution of survival times is the focus of survival analysis. Although there are well-known methods for estimating unconditional survival distributions, the most interesting survival modeling investigates the relationship between survival and one or more predictors, referred to as covariates in the literature on survival analysis. The Cox proportional-hazards regression model (introduced by Cox, 1972), a widely applicable and widely used method of survival analysis, is the subject of this appendix (Fox,2008).

One of the commonly used survival/mathematical model for assessing the risk factors effect (exploratory variable) on the failure time through the hazard function is CPH and it was used extensively since it was introduced by (Cox, 1972). Let T denote the time until the unit experiences failure, and let Z denote the observed vector of covariates. Then, under the CPH model the function of hazard is given as;

$$\lambda(t|X) = \lambda_0(t)\, exp(\beta' Z) \tag{3.1}$$

The model's survival function is given by

$$S(t|Z) = exp\big(-exp(\beta' Z)\, \Lambda_0(t)\big) = S_0(t)^{exp(\beta' Z)}$$

and the corresponding distribution function has the form

$$F(t|Z) = 1 - exp(-\Lambda_0(t)exp(Z'\beta)).$$

Here $\lambda_0(t)$ is the baseline hazard, β is the vector of regression parameters, Z is the vector of covariates of an individual, $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$ is cumulative baseline hazard and $S_0(t) = exp(-\Lambda_0(t))$ is the baseline survival function.

3.1.1 CPH MODEL FOR THE SUBDISTRIBUTION

CPH for the subdistribution was presented by Find-Gray (1999). This model is built on the log (-log (1-u)) transformation model that is generally used with univariate survival data. Further, this model was estimated using the subdistribution of the hazard that is originally introduced by Gray (1988) and given as;

$$\lambda_j^*(t,Z) = \lim_{dt\to 0} \frac{1}{dt} P(t \le T \le t + dt, C = j | T \ge t \cup (T \le t \cap C \ne j), Z)$$

$$= \frac{dF_j(t,Z)/\, dt}{1 - F_j(t,Z)}$$

$$= \frac{-d\, log(1 - F_j(t,Z))}{dt},$$

Where; j is the interest cause of failure and λ_j^* is the hazard function for the improper random variable $T^* = I(C = j) \times T + \big(1 - I(C = j)\big) \times \infty$. The implied failure time T^* has a distribution function equal to $F_j(t,Z)$. Obviously, the risk set connected to the hazard λ_j^* is unusual, that is, the units which have failed due to cause other than the cause of interest remain in the risk set indefinitely as long as they have not experienced the event of interest. Under a PHs specification, the hazard of subdistribution is given as;

$$\lambda_j^*(t,Z) = \lambda_{0j}^*(t)e^{\beta_j' Z}, \qquad (3.2)$$

where: λ_{0j}^* is an unspecified and nonnegative function, and the corresponding CIF is

$$F_j(t,Z) = P(T \le t, C = j|X) = 1 - e^{-\Lambda_{0j}^*(t)e^{\beta_j' Z}}, \qquad (3.3)$$

where: $\Lambda_{0j}^* = \int_0^t \lambda_{0j}^*(s)ds$.

The full likelihood associated with observed censored data is given by Kalbfleisch & Prentice (1980) as;

$$L = \prod_{i=1}^{n_1} \sum\nolimits_{j \in S_i} P(S_i|T_i, C_i = j, Z_i) f_j(T_i|Z_i) \prod_{i=n_1+1}^{n_2} S(T_i|Z_i), \qquad (3.4)$$

where respectively n_1 and n_2 represent the number of failed and censored as the type right units. As the relationship between the subdistribution hazard $\lambda_j^*(t)$ and the subdensity $f_j(t)$ and the CIF has the form

$$\lambda_j^*(t) = \frac{f_j(t)}{1 - F_j(t)}, \qquad (3.5)$$

and the relationship between the subsurvival function $S_j(t)$ and the CIF $F_j(t)$ has the form

$$F_j(t) + S_j(t) = P(C = j), \ \sum_{j=1}^{K} P(C = j) = 1. \qquad (3.6)$$

Then (3.4) can be rewritten as

$$L = \prod_{i=1}^{n_1} \sum\nolimits_{j \in S_i} P(S_i|T_i, C_i = j, Z_i)\, \lambda_j^*(T_i|Z_i)(1 - F_j(T_i|Z_i)) \times \prod_{i=n_1+1}^{n_2} [1 - \sum_{j=1}^{K} F_j(T_i|Z_i)]. \qquad (3.7)$$

Further, substituting (3.2) and (3.3) in (3.7) then the likelihood for the right-censored will be as;

$$L = \prod_{i=1}^{n_1} \sum_{j \in S_i} P(S_i|T_i, C_i = j, Z_i)\ \lambda_{0j}^*(t) e^{\beta_j' Z} e^{-\Lambda_{0j}^*(t) e^{\beta_j' Z}}$$

$$\times \prod_{i=n_1+1}^{n_2} [1 - \sum_{j=1}^{K} (1 - e^{-\Lambda_{0j}^*(t) e^{\beta_j' Z}})]. \qquad (3.8)$$

Now to derive a likelihood for Interval-Censored (IC) data we need to extend the RC data likelihood to accommodate the interval-censored observations. Using the relationships mentioned previously the likelihood can be defined as

$$L = \prod_{i=1}^{n_1} \sum_{j \in S_i} P(S_i|T_i, C_i = j, Z_i)\ \lambda_j^*(T_i|Z_i)(1 - F_j(T_i|Z_i)$$

$$\times \prod_{i=n_1+1}^{n_2} [1 - \sum_{j=1}^{K} F_j(T_i|Z_i)]$$

$$\times \prod_{i=n_2+1}^{n_3} \sum_{j \in S_i} P(S_i|T_i, C_i = j, Z_i)\ [F_j(R_i|Z_i) - F_j(L_i|Z_i)] \quad (3.9)$$

The likelihood for interval-censored will takes the form of equation (4.0) after substituting (3.1) and (3.2) into (3.9) as;

$$\mathrm{L} = \prod_{\mathrm{i}=1}^{\mathrm{n}_1} \sum_{\mathrm{j} \in \mathrm{S_i}} \mathrm{P}(\mathrm{S_i}|\mathrm{T_i}, \mathrm{C_i} = \mathrm{j}, \mathrm{Z_i})\ \lambda_{0\mathrm{j}}^*(\mathrm{T_i}) \mathrm{e}^{\beta_{\mathrm{j}}' \mathrm{Z_i}} \mathrm{e}^{-\Lambda_{0\mathrm{j}}^*(\mathrm{T_i}) \mathrm{e}^{\beta_{\mathrm{j}}' \mathrm{Z_i}}}$$

$$\times \prod_{\mathrm{i}=\mathrm{n}_1+1}^{\mathrm{n}_2} \left[1 - \sum_{j=1}^{K} \left(1 - \mathrm{e}^{-\Lambda_{0\mathrm{j}}^*(T_i) \mathrm{e}^{\beta_{\mathrm{j}}' \mathrm{Z_i}}} \right) \right]$$

$$\times \prod_{i=n_2+1}^{n_3} \sum_{j \in S_i} P(S_i | T_i, C_i = j, Z_i) \left[e^{-\Lambda_{0j}^*(L_i) e^{\beta_j' Z_i}} - e^{-\Lambda_{0j}^*(R_i) e^{\beta_j' Z_i}} \right]. \quad (4.0)$$

Further, we will take the log of the likelihood function of equations (3.8) and (4.0) that depends on the unknown parameters β, the values of Z being known. In large sample, the distribution of β can be approximated by a normal distribution with the score vector, estimated by maximizing the likelihood from the first derivative, and a variance-covariance matrix, estimated from the second derivative of the likelihood function.

The regression coefficients β are estimated by the values $\hat{\beta}$, which maximize the logarithm of the full likelihood. The values $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_n)$ are obtained by equating to zero the n first derivatives of log likelihood function of equations (3.8) and (4.0) with respect to β. An iterative process such as the EM algorithm or Newton-Raphson are adopted to solve this system of equations for β.

CHAPTER FOUR SIMULATION STUDY AND REAL DATA ANALYSIS

CHAPTER OVERVIEW

This chapter focuses on the analysis of the failure time data in competing risks model via a cumulative incidence framework. The modified Fine & Gray model with interval-censored (IC) data will be compared with right-censored model when there is more than one cause of failure that at least one is known. Furthermore, the results obtained from simulated data and secondary data are compared to these models that is based on the right-censored (RC) data with fully observed causes of failure, as proposed by Fine and Gray (1999) to our FG model with IC. All calculations were computed using R software.

4.1 SIMULATION STUDY

Under competing risks framework via the CIF, simulation data was generated based on the CIF through Cox model with subdistribution hazard function. Following Fine and Gray (1999) and Yosra (2017) we generate the failure times. If we have two events, 1 & 2, as suggested by FG the CIF to follow the model

$$P(T \leq t, C = 1|Z) = 1 - \left(1 - p(1 - e^{-t})\right)^{e^{\theta_1' z}}, \qquad (4.1)$$

where p ∈ (0,1) and $P(C = 1|Z) = 1 - (1 - p)^{e^{\theta_1' z}}$ is the probability of experiencing the event of interest 1, given the vector of covariates Z. The distribution function (4.1) results from a proportional subdistribution hazards model (4.1) with baseline hazard

$$\lambda_{01}^{*}(t) = \frac{pe^{-t}}{1 - p(1 - e^{-t})} \qquad (4.2)$$

The competing CIF was assumed to be;

$$P(T \leq t, C = 2|Z) = P(C = 2|X)P(T \leq t|C = 2, Z)$$

$$= (1-p)^{e^{\theta_1' z}} \left(1 - e^{-te^{\theta_1' z}}\right), \qquad (4.3)$$

where: $P(T \leq t, C = 2|Z)$ is an exponential distribution with hazard function $e^{\theta_1' z}$. This simulation is designed so that the unit's event type can be determined first with $P(C = 1|\, Z)$, then the corresponding failure time T is generated conditional on C with distribution

$$P(T \leq t|\, C = j, Z) = P(T \leq t, C = j|Z)/P(C = j|Z), \quad j = 1,2. \qquad (4.4)$$

Following Yosra (2017), in our simulation study, we assumed the number 0.5 or -0.5 as values of one covariate Z, whereas for sample size of 26, 50 and 100 respectively the true parameters of (p, θ_1, θ_2) represented as (0.7, -2.387, 3.183), (0.7, -1.85, 2.23) and (0.7, -1.00, 2.34). From uniform distribution U[1,7] we generated the censored times. Variety of interval-censoring rates are generated based on different samples with different sizes. The assumed models result in units that fail due to cause1 and cause 2 with different interval width as 0.2, 0.3, and 2. After generating the interval censored data, the inspection times based on imputation methods such as midpoint will be used to turn the data into right-censored data with ranging between 51% and 55%, for cause 1 and ranging between 45% and 49% for cause 2.

We will estimate the unknown parameters using the initial value of zero for the regression coefficients with their standard error and p-value of estimations based on our model (IC) and right censored (RC) model that are applied to simulated data sets.

The estimations of regression coefficients, bias and mean square error (MSE) of estimation for right and interval-censored data for three data sets with different sample sizes across three interval-censoring levels are summarize in Table 4.1. It is clear from this Table that the interval-censoring rate has a significant impact on model performance. Further, it can be seen from Table 4.1 that the parameters estimate exhibit some evidence of error (i.e., the MSE values), however, this is not significant as the MSE is more than 5% of the sample standard error. This suggests that the estimates of all regression parameters from both IC and RC models are accurate. Moreover, Figures 4.1 to 4.8 showed that the comparisons of the estimated CIF's between the IC model and RC model through three different data sets for both causes of failure. The varieties in the data sets levels basically aim to investigate how sensitive are the proposed models to interval-censored as mentioned previously in Section 3. Then, from the figures it can be inferred that interval-censored and right-censored has a slight impact on CIF curves that correspond to IC and RC models as the distance between them when the different level of interval censored changes. Nevertheless, the CIF curves show a substantial consistency with Fine & Gray CIF curves except for IC model which show an indication of little change in its pattern in the early time period. This change is attributed to the difference of the points of time. In contrast, Figures 4.4 and 4.8 compare the estimated CIF of IC and RC models with three interval-censoring rates. Obviously, the performance of the modified model is not affected by the width of the interval censoring (0.2, 0.3, and 2) as the CIF curves for the three different interval-censoring levels are nearly equal. However, the estimated CIF based on IC are compared with RC as showed in the Figures 4.1 to 4.8 and Table 4.1 mentioned above. We find that the fit is reasonable in both cause of failures, although it is not perfect for some case for cause 2. In additional to that it was observe that when the sample size

increase the MSE will decrease suggestion that the models fitting better with large sample.

4.2 DATA SET AND APPLICATION

This section discusses the application of the proposed models to the data set based on a retrospective data on the Middle East respiratory syndrome corona virus (MERS or MERS-CoV) outbreak in the Kingdom of Saudi Arabia (KSA) between 2012 and 2016. The data set was the case-by-case data list compiled and regularly maintained by Rambaut (2013) from different sources that including World Health Organization (WHO) bulletins, Ministry of Health of the KSA and media reports. We consider age, gender, patient type (patient is a Healthcare worker (HCW) or non HCW), patient comorbidity status and patient of exposure to known risk factors (animal contact and camel contact indirectly or directly) as the variables in this study.

The information on the event time, i.e., a patients time of infection until occurrence of MRES infection, end of MRES or death, whatever occurs first, and on the event type.

The dataset contains 1361 patients with MERS with age between 25 to 74 years. Of those, 901 are males, 460 are females, the number of patients with comorbidity about 1090, 629 of the patients exposure to camel, 1086 of the patients infection through the HCW and about 750 cases whose contact infection are unknown, for more details about the data reader refer to; Rambaut (2013); Oyelola et.al (2017); Oyelola and Elfaki (2018); Oyelola et al (2019. The patients are effected by MRES due to contact with HCW, comorbidity and exposure to camel in this situation we have competing risks in survival analysis. Therefore, we consider the event of interest that a person confirmed

with MRES disease due to one of these causes; through HCW, comorbidity or exposure to camel or other causes.

The estimations of regression coefficients from both model with their standard error and P-value of estimation for MERS data set with two different covariates that gender and age are summarize in Table 4.2. It is clear from the summaries that the interval-censoring width has a significant impact on model performance. Further, it can be seen from table 4.2 that the parameters estimate with their SE and P-value obtained from gender are more significant compare with one from age suggesting that older people are easily affected with MERS diseases through HCW and comorbidity. In additional to that, these results suggests that, the estimates of all regression parameters from both IC and RC models are almost similar. Since the test gives p-values below 0.001 for both covariates, respectively, suggesting that the both models fit well for the CIF for the three causes of failure.

The CIF curve is a proper summary curve, showing the CIF failure rates over time due to a particular cause. Figures 4.9 to 4.20 compare the cumulative incidence functions of the three causes of failures that infection based on the HCW, comorbidity and exposure to MERS from two models based on two covariates. The figures confirm the results in Tables 4.2 as the estimated CIFs indicate no evidence of effect by infection of MERS in cause 2, which is a little greater than the other causes across the three causes of failures. The infection the disease due to HCW, comorbidities and exposure to MERS are significant with respect to P-value for both causes. Obviously, the HCW and comorbidities for gender are more prone to infection since in the hospital the HCW acquired infection among patients and from patients to HCW. These results indicate that the performance of our model is reasonable for interval-censored compare with RC model.

4.3 CONCLUDING REMARKS

Fine & Gray models that is; the RC model and the modified with IC are used under the hazard of subdistribution framework to evaluate the impact of covariates on CIF in this chapter. The first model deals with right-censored data, whereas the second deals with interval-censored data. Both models are easy to implement and their results are comparable.

Table 4.1. Result from Simulated Data for Regression Coefficients of RC and IC Models.

No of Sample Size	Model	Cause 1			Cause 2		
		θ	Bias (θ)	MSE (θ)	θ	Bias (θ)	MSE (θ)
50	RC	-1.853	-7.50e-10	0.154	2.429	5.32e-08	0.123
	IC	-1.743	-4.54e-09	0.277	2.277	1.16e-10	0.216
100	RC	-1.005	-4.21e-09	0.124	2.347	1.89e-11	0.095
	IC	-0.9692	-3.93e-08	0.275	2.055	1.13e-08	0.049
26	RC	-2.387	-2.79e-08	0.281	3.183	3.95e-07	0.501
	IC	-2.523	-3.012e-08	0.364	2.777	8.52e-11	0.664

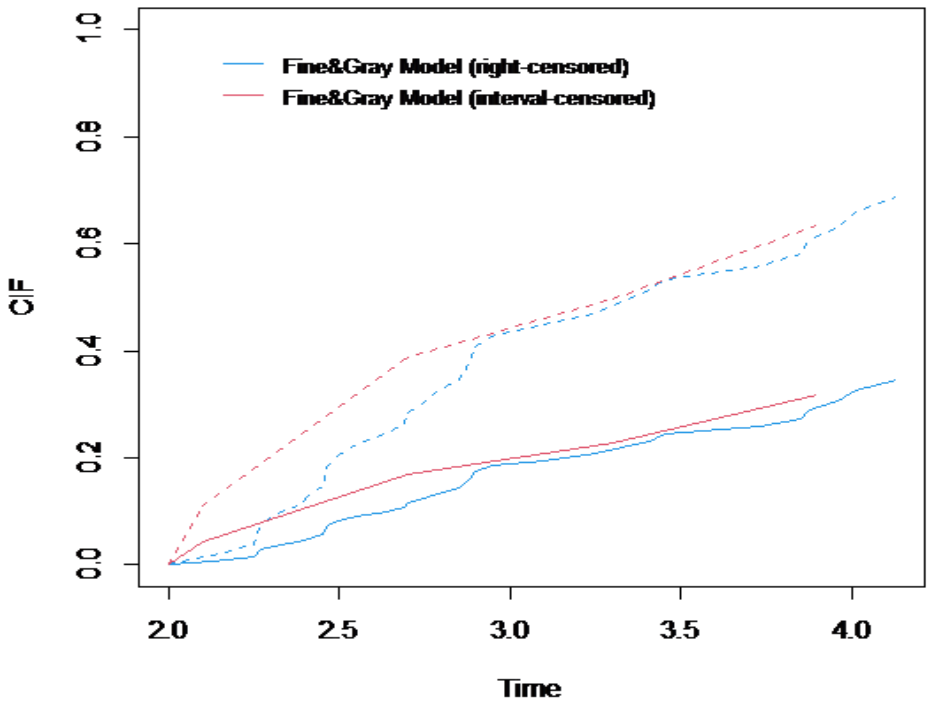

Figure 4.1: Estimation of the CIFs obtained by IC and RC models for cause 1 with sample size 100.

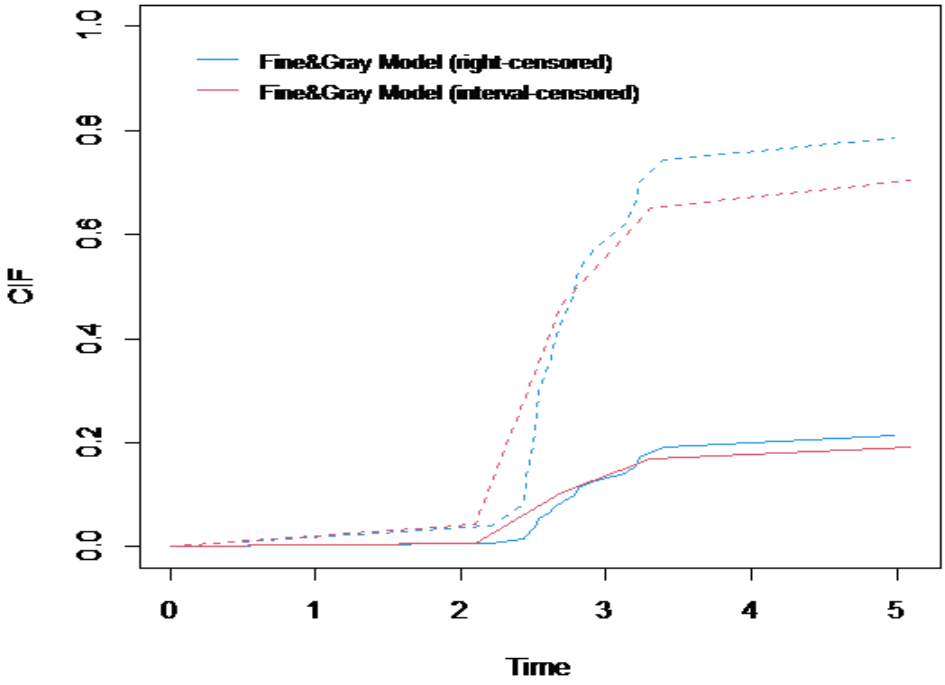

Figure 4.2: Estimation of the CIFs obtained by IC and RC models for cause 1 with sample size 50.

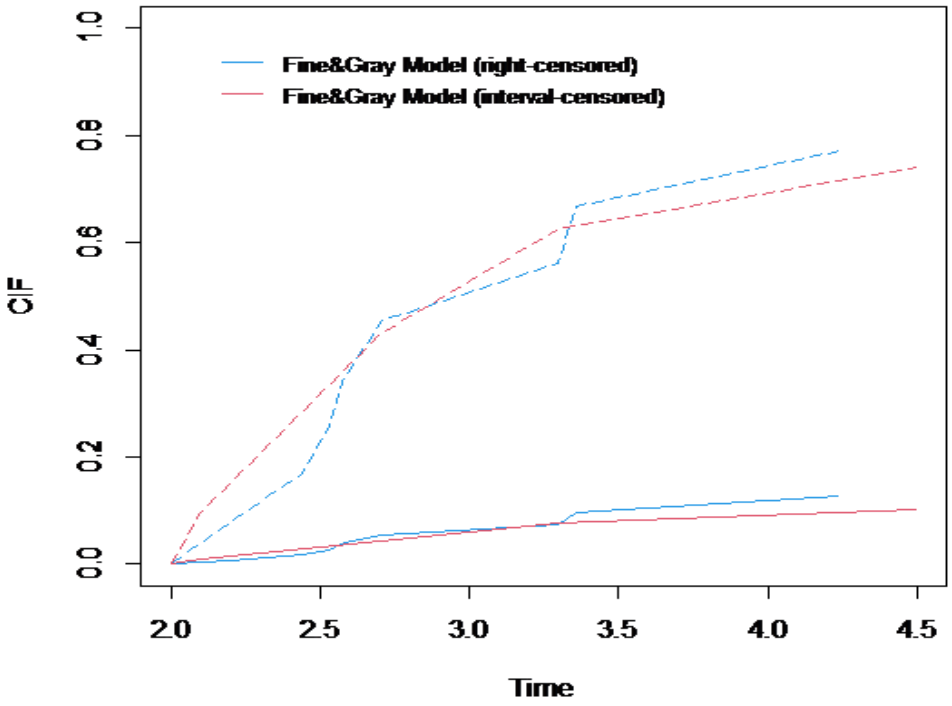

Figure 4.3: Estimation of the CIFs obtained by IC and RC models for cause 1 with sample size 26.

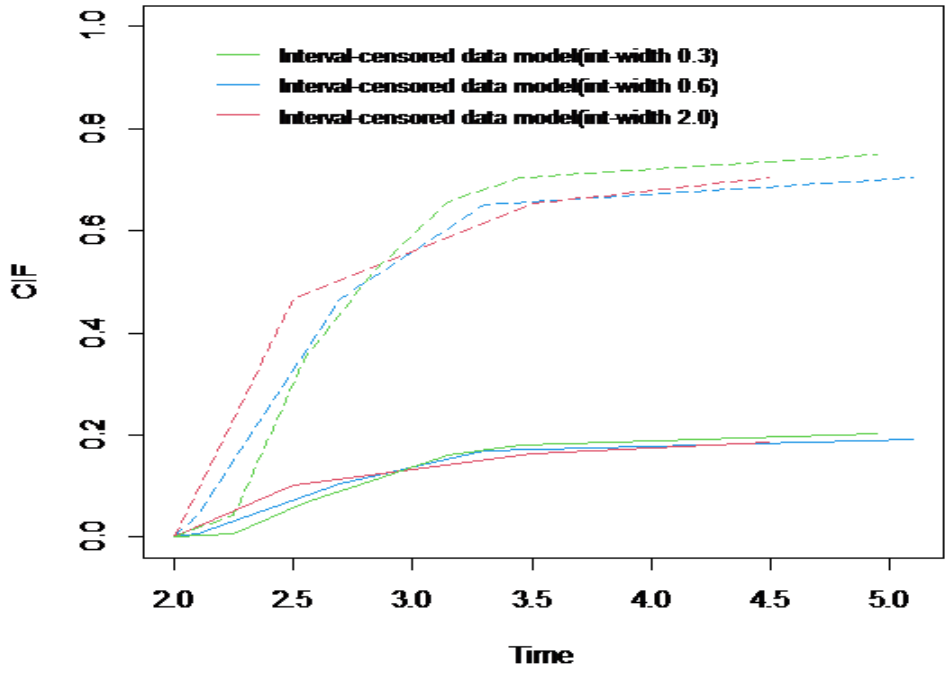

Figure 4.4: Estimation of the CIFs obtained by IC and RC models based on different width of interval-censored for Cause 1 with sample size 50.

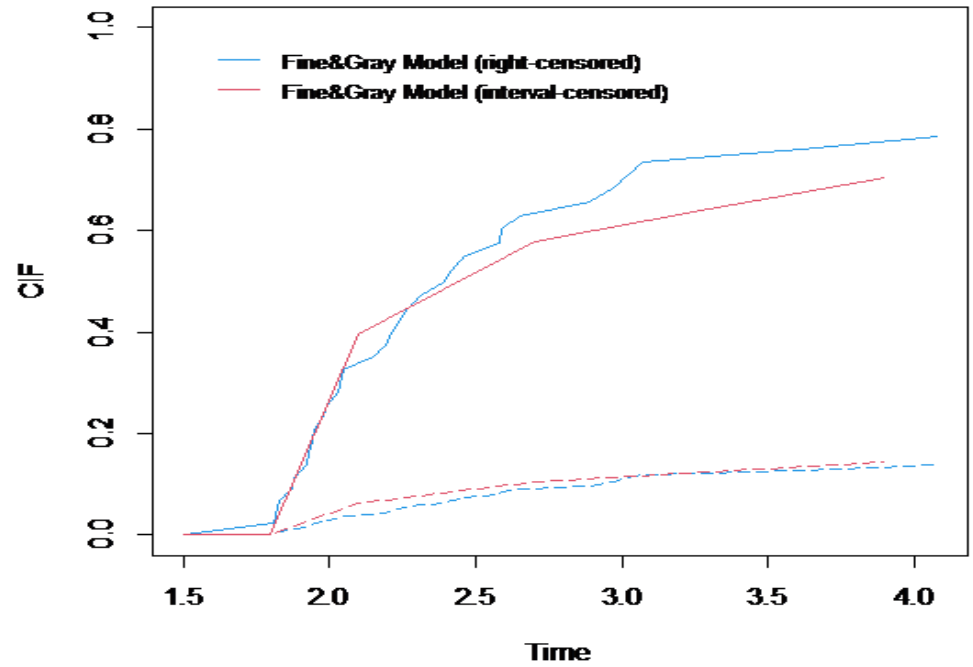

Figure 4.5: Estimation of the CIFs obtained by IC and RC models for cause 2 with sample size 100.

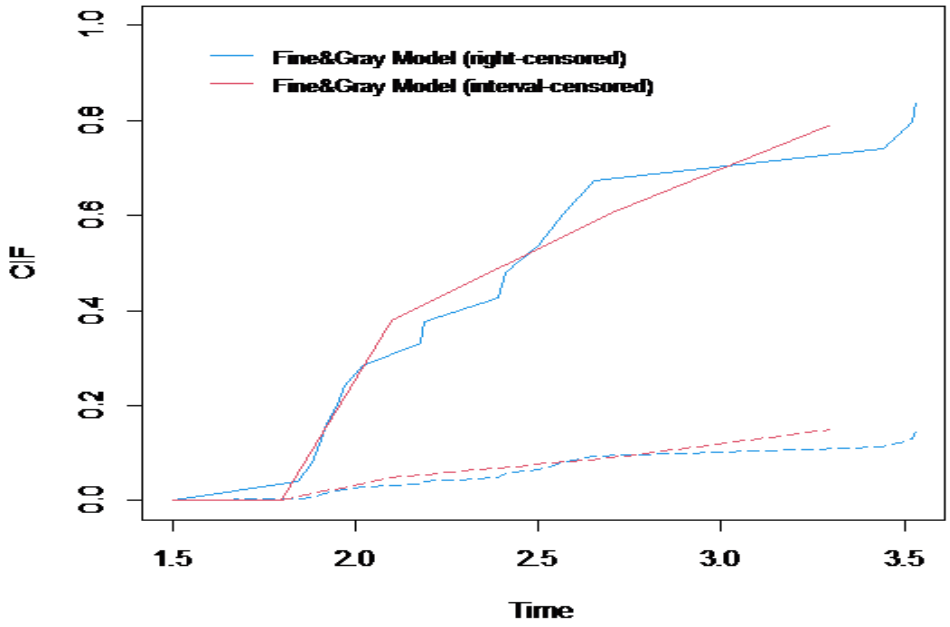

Figure 4.6: Estimation of the CIFs obtained by IC and RC models for cause 2 with sample size 50.

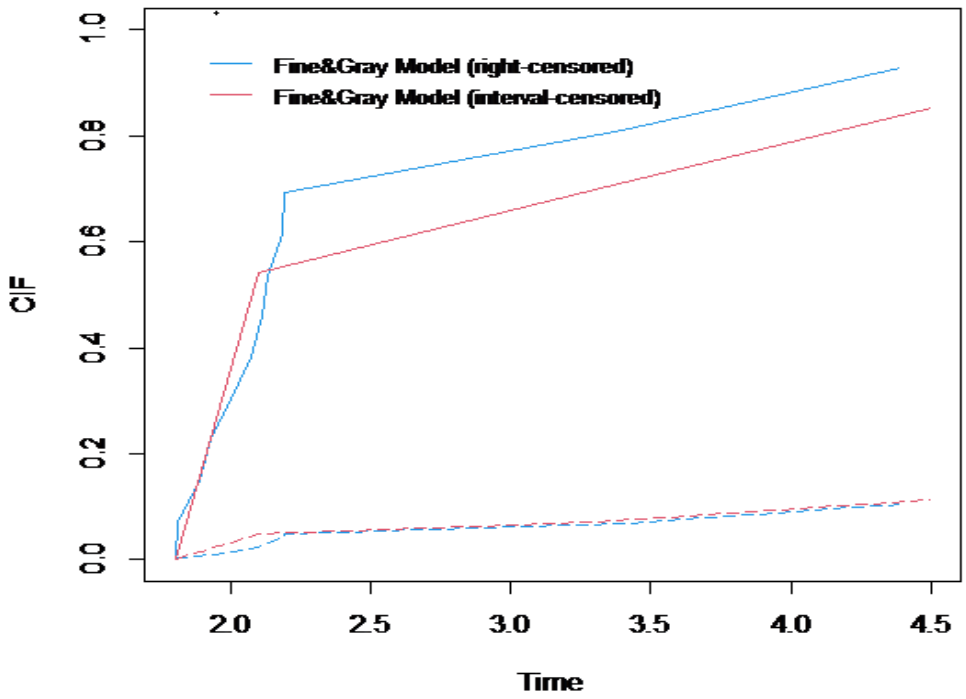

Figure 4.7: Estimation of the CIFs obtained by IC and RC models for cause 2 with sample size 26.

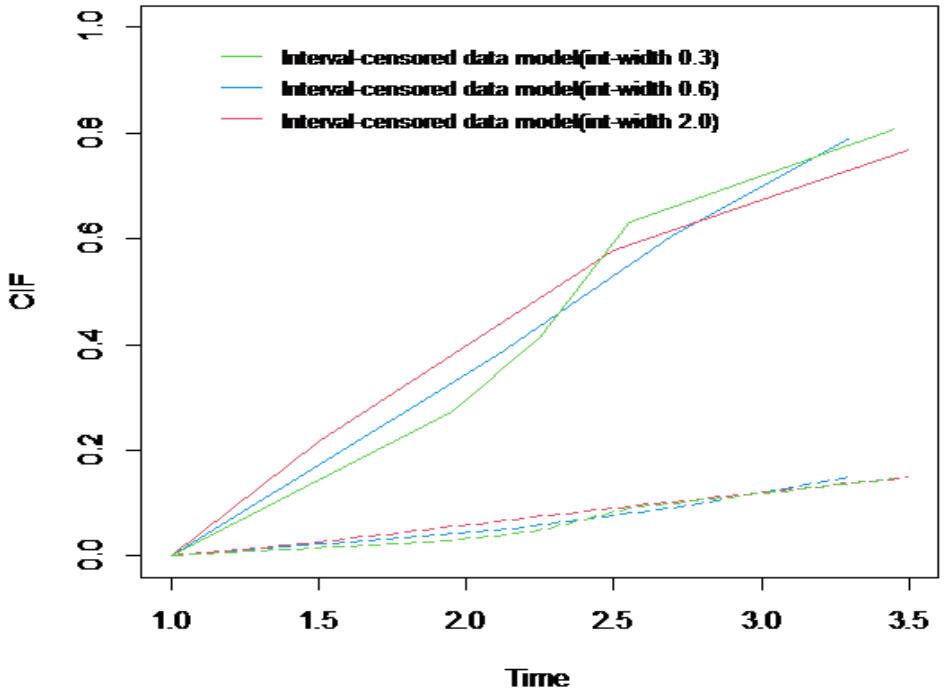

Figure 4.8: Estimation of the CIF obtained by IC and RC models based on different width of interval-censored for Cause 2 with sample size 50.

Table 4.2. Result from MERS Data for Regression Coefficients of IC and RC Models.

Causes of Failure	Method	Covariate	Parameter Estimate	Standard Error	P-value
			Cause 1		
HCW	RC	Gender	-0.6214	0.09233	1.7e-11
	IC		-0.5820	0.09209	2.6e-10
	RC	Age	-0.8508	0.09400	0.001
	IC		-0.8229	0.09388	0.001
Comorbidities	RC	Gender	0.2506	0.06070	3.6e-05
	IC		0.2694	0.05879	4.6e-06
	RC	Age	1.151	0.06081	0.001
	IC		1.127	0.05901	0.001
MERS Exposure	RC	Gender	-0.5966	0.06022	0.001
	IC		-0.5358	0.05685	0.001
	RC	Age	-0.4515	0.05971	4e-14
	IC		-0.4229	0.05689	1.1e-13
			Cause 2		
Other than HCW	RC	Gender	0.4734	0.07015	1.5e-11
	IC		0.4654	0.06801	7.7e-12
	RC	Age	0.6256	0.06390	0.001
	IC		0.6031	0.06184	0.001
Other than	RC	Gender	-0.2573	0.09460	0.00650
Comorbidities	IC		-0.2499	0.09328	0.0074
	RC	Age	-1.5940	0.11000	0.001
	IC		-1.5750	0.1087	0.001

Causes of Failure	Method	Covariate	Parameter Estimate	Standard Error	P-value
Other than	RC	Gender	0.9513	0.1032	0.001
MERS Exposure	IC		0.9440	0.1025	0.001
	RC	Age	0.6108	0.08227	1.1e-13
	IC		0.6032	0.08142	1.3e-13

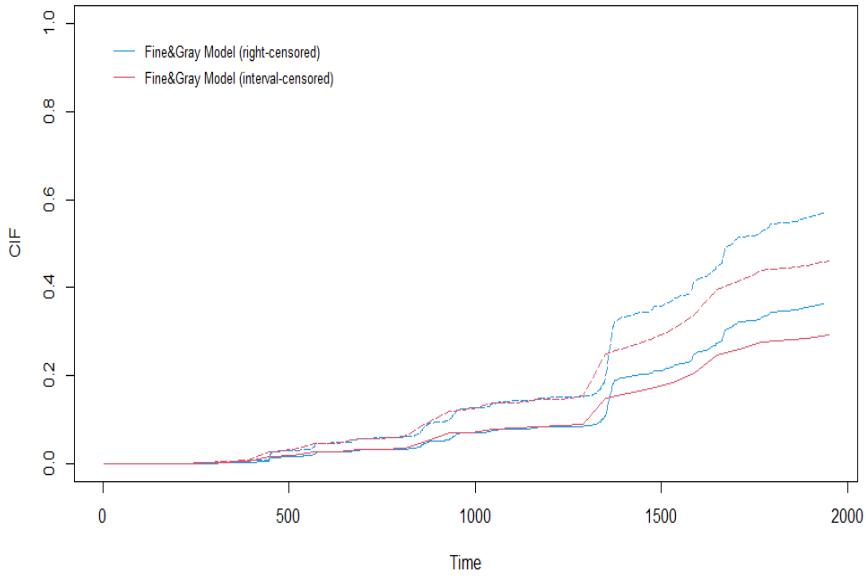

Figure 4.9: Estimation of the CIFs obtained by IC and RC models based on HCW-Gender for cause 1.

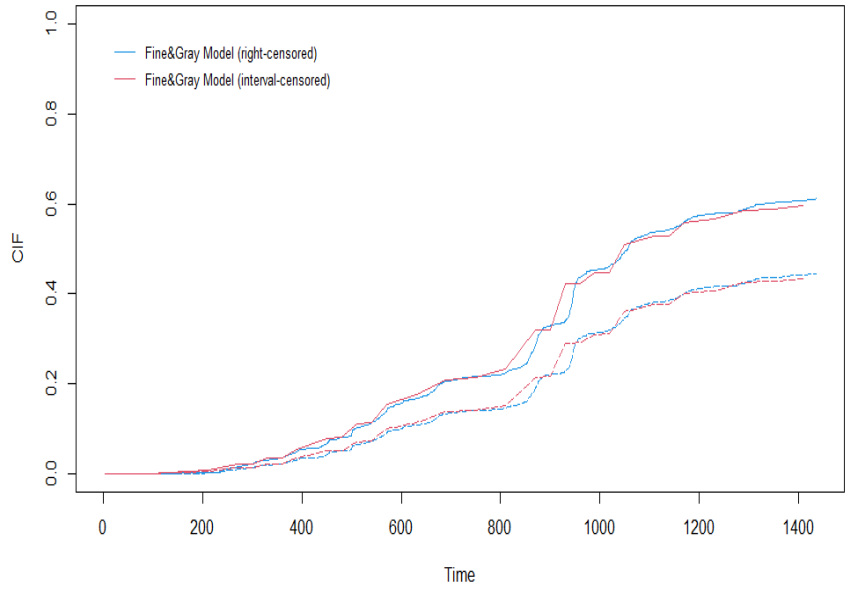

Figure 4.10: Estimation of the CIF obtained by IC and RC models based on HCW-Gender for Cause 2.

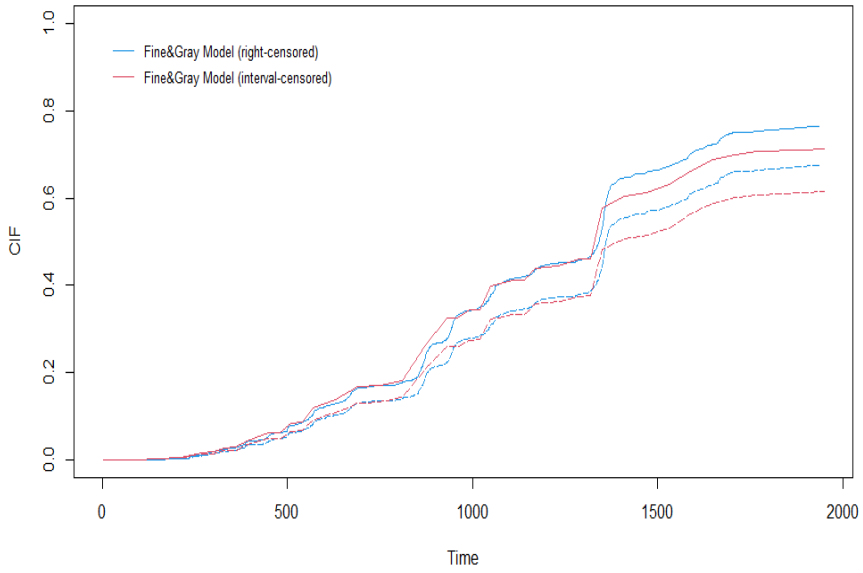

Figure 4.11: Estimation of the CIF's obtained by IC and RC models based on Comorbidities -Gender for Cause 1.

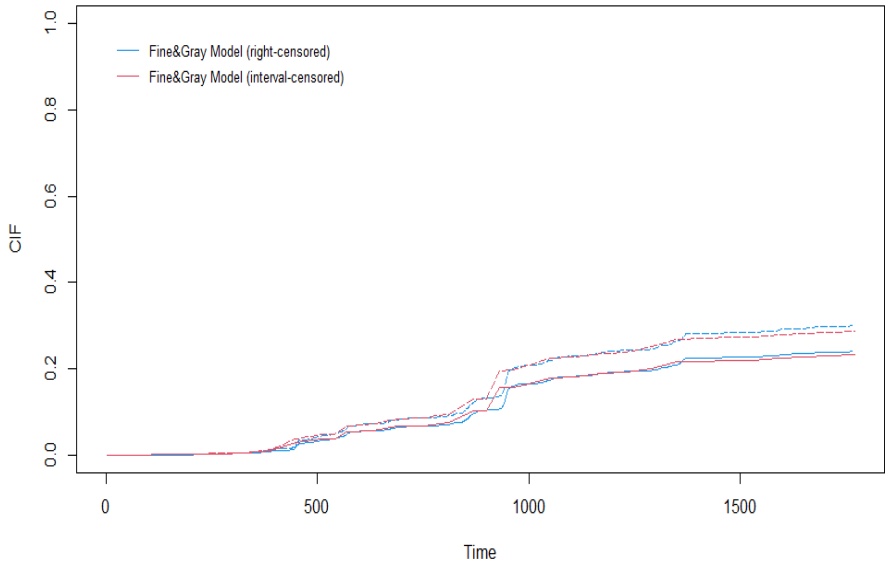

Figure 4.12: Estimation of the CIF's obtained by IC and RC models based on Comorbidities -Gender for Cause 2.

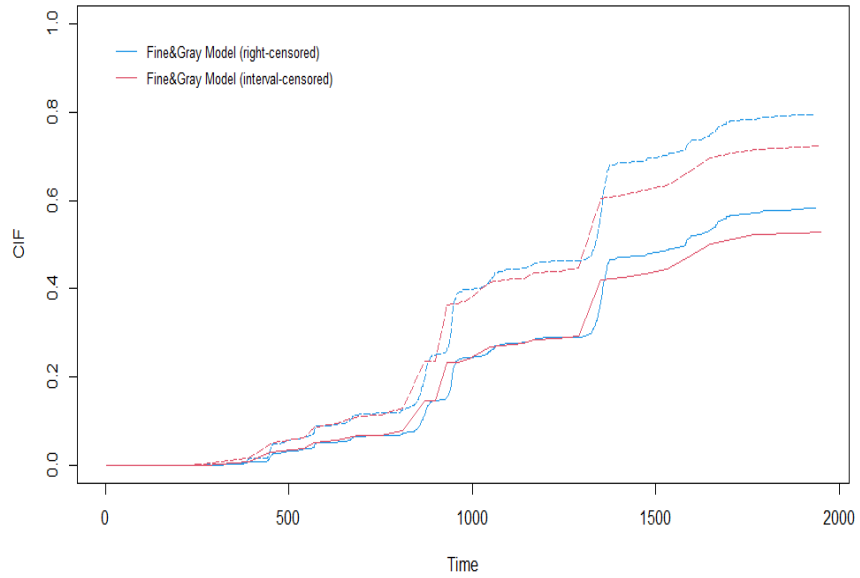

Figure 4.13: Estimation of the CIF's obtained by IC and RC models based on MERS Exposure -Gender for Cause 1.

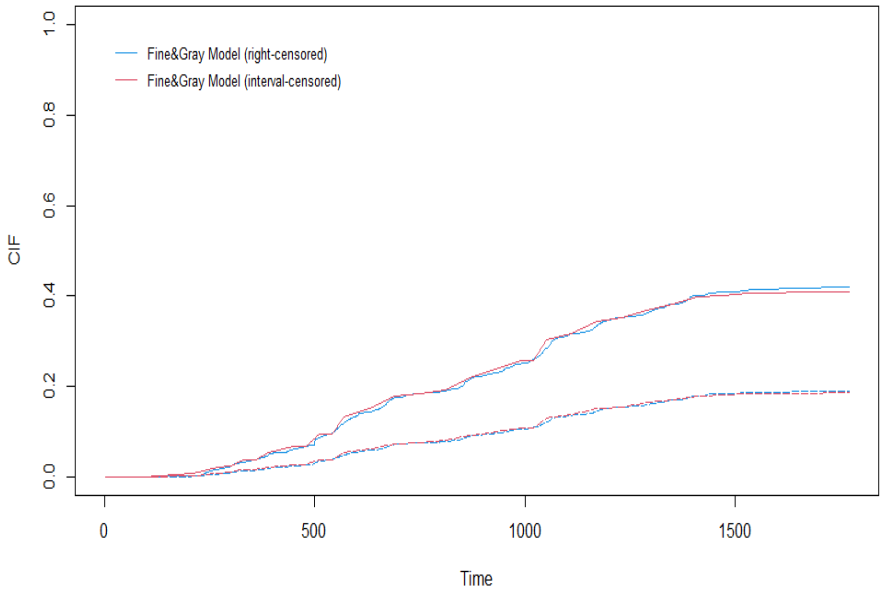

Figure 4.14: Estimation of the CIF's obtained by IC and RC models based on MERS Exposure -Gender for Cause 2.

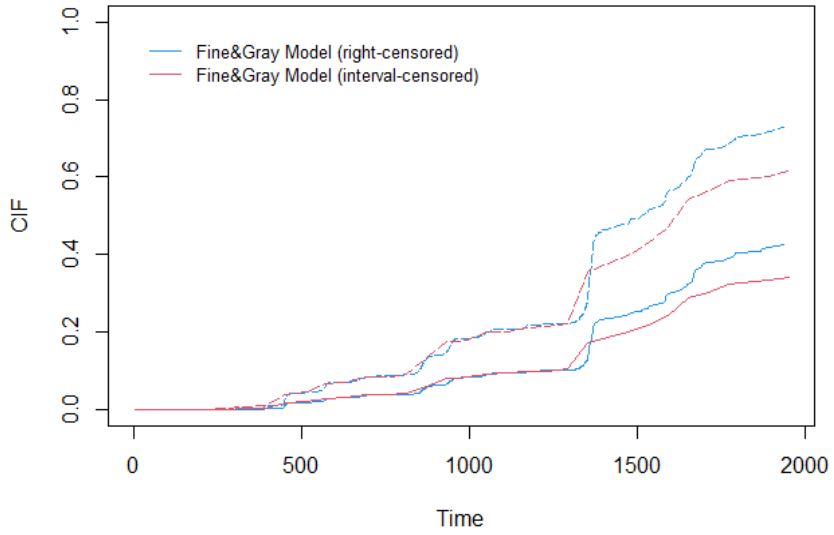

Figure 4.15: Estimation of the CIF's obtained by IC and RC models based on HCW-Age for Cause 1.

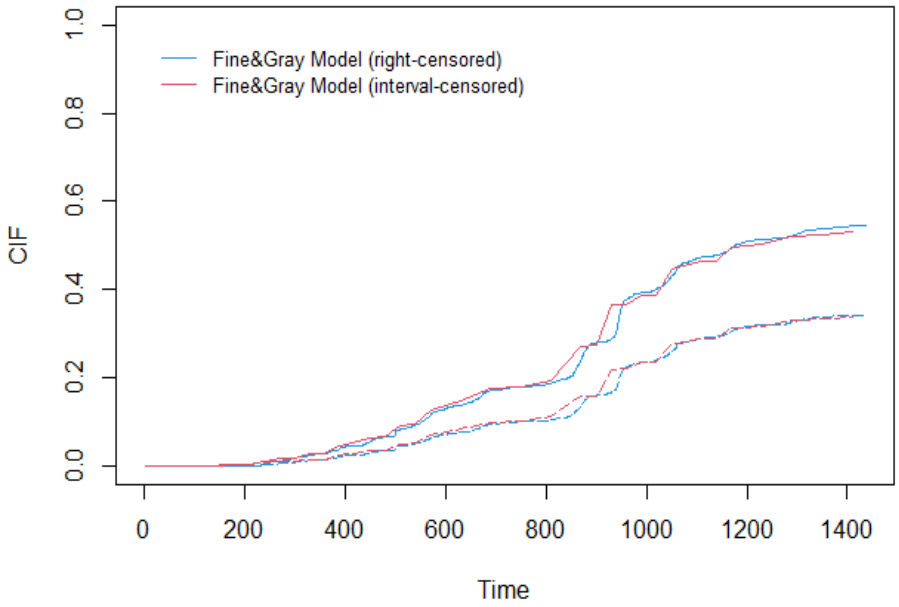

Figure 4.16: Estimation of the CIF's obtained by IC and RC models based on HCW-Age for Cause 2.

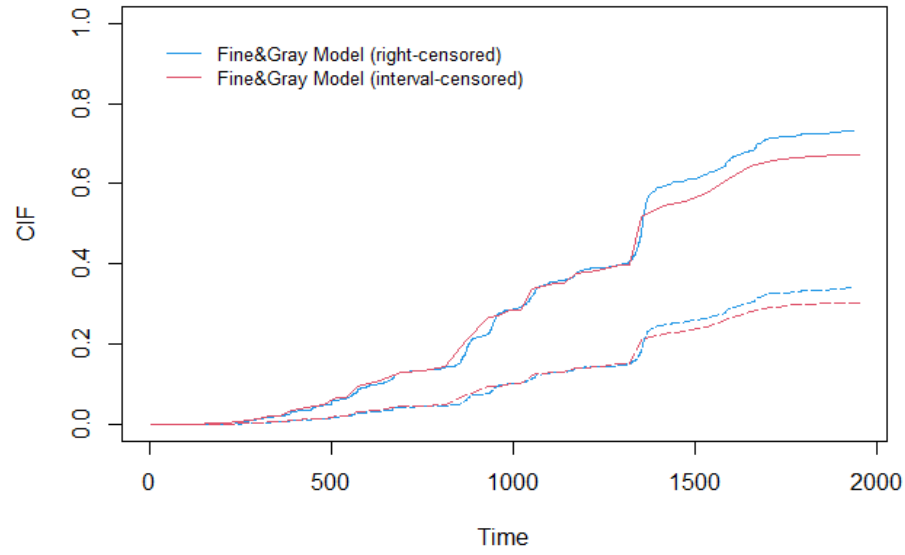

Figure 4.17: Estimation of the CIF's obtained by IC and RC models based on Comorbidities- Age for Cause 1.

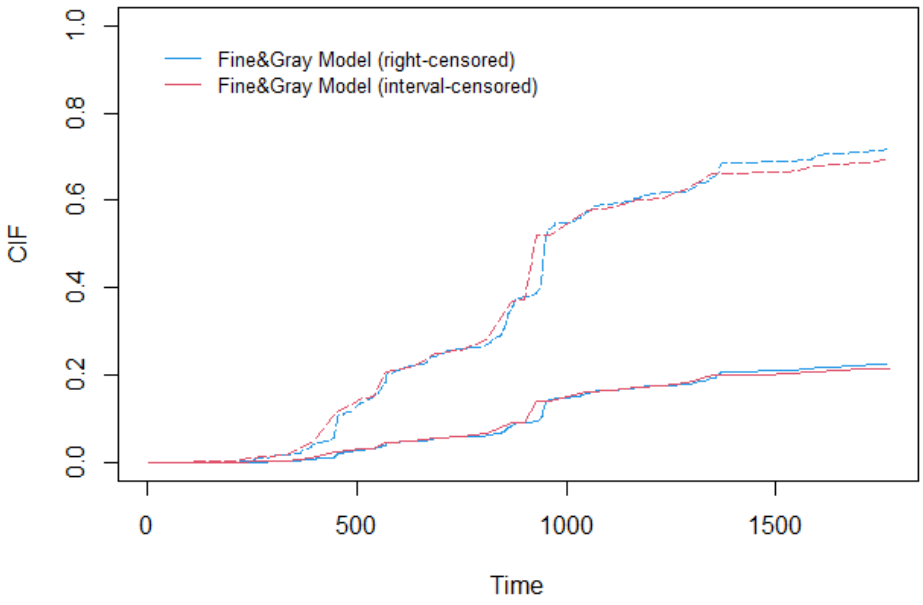

Figure 4.18: Estimation of the CIF's obtained by IC and RC models based on Comorbidities- Age for Cause 2.

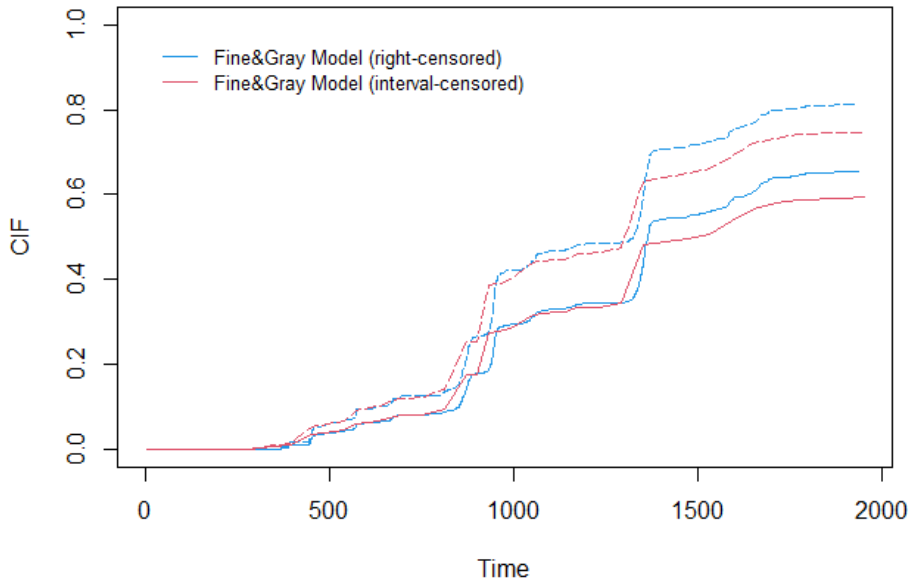

Figure 4.19: Estimation of the CIF's obtained by IC and RC models based on MERS Exposure - Age for Cause 1.

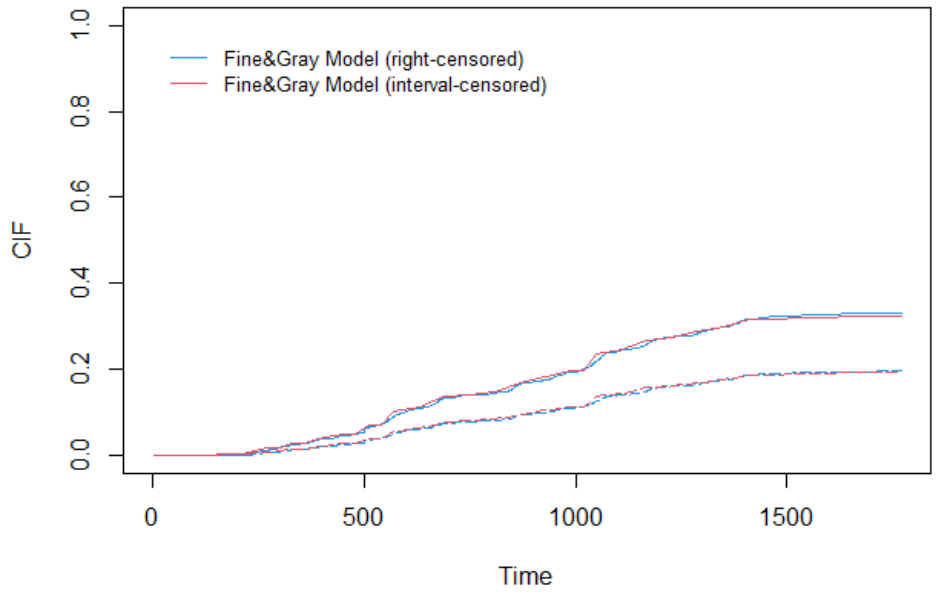

Figure 4.20: Estimation of the CIF's obtained by IC and RC models based on MERS Exposure - Age for Cause 2.

CHAPTER FIVE CONCLUSION AND SUGGESTIONS FOR FUTURE RESEARCH

CHAPTER OVERVIEW

Conclusion that discusses the results of the previous chapters and some suggestions for further research will be parented in this chapter.

5.1 CONCLUSION

Competing risks data with two causes of failure is studied in this thesis. Further, the failure times are observed exactly between an interval of time under study, instead, they are only known to be included in an interval of time. This type of data is termed interval-censored data. The main goal of this study is to provide models that can be used to assess the effect of covariates on the cumulative incidence function. The Cox's proportional hazard model and the subdistribution hazard model is used. The statistical literature has several methods developed to study the competing risks data with different censoring scenarios. However, less work for the interval-censored data has been done so far to our knowledge especially for the use of MERS data to competing risks model.

In order to achieve our objectives, we start with modified the likelihood function for interval-censored data. The later function that include the observed data, needs to be constructed that to deal with interval-censored data based on competing risks framework.

Since the major aims are to assess the effect of covariates on the CIF, therefore the Cox's proportional hazards model for the subdistribution is adopted to evaluate the

effect of covariate on the CIF. Two cases are studied here, Fine & Gray model based on interval-censored (IC) data and right-censored (RC) data. Later the method of midpoint imputation will be used to impute the data to right censoring.

To evaluate the proposed methodologies of this study, the simulation studies are conducted that is similar to the one proposed by Fine and Gray (1999). First, we simulate the competing risks data into interval censoring and then the method of midpoint is used to turn the simulated data into right censored data. The simulation results show that the our model in this study is applicable and performs well compared to the model with Fine & Gray model with right-censored data.

In conclusion, to show that this work is a useful tool that can be employed to solve real life issues, the proposed methodologies are applied to real MERS data set as this study has interest in studying the risks factors that can affect medical experiments which relate to the lifetime data analysis. However, this does not mean that the proposed models of this study cannot be applied to data sets from other study fields. The data set reported in Rambaut A. (2013) and others is modified to become interval-censored data with three causes of failure then analyzed using methods proposed in this study. The obtained results from the two models, namely, the modified IC model and RC model based on CIF formulation, indicate that the CIF are not influenced by the covariates.

5.2 SUGGESTIONS FOR FUTURE RESEARCH

In this research, we used midpoint imputation method to deal with the interval censored data for the simulation study as well as for MERS data. It will be better to use

different imputation methods such; as EM algorithm or multiple imputation among others and made a comparison between them for better performance and inference. Also, one can develop a full parametric model which might be more suitable in some situations.

REFERENCES

Anderson-Bergman, C. (2017). Icen Reg: regression models for interval censored data in R. Journal of Statistical Software, 81(1), 1-23. Retrieved (https://www.jstatsoft.org/article/view/v081i12)

Austin, P. C., & Fine, J. P. (2017). Practical recommendations for reporting Fine-Gray model analyses for competing risk data. Statistics in Medicine, 36(27), 4391–4400. Retrieved (https://doi.org/10.1002/sim.7501).

Austin, P. C., Steyerberg, E. W., & Putter, H. (2021). Fine-Gray subdistribution hazard models to simultaneously estimate the absolute risk of different event types: Cumulative total failure probability may exceed 1. Statistics in Medicine, 40(19), 4200–4212. Retrieved (https://doi.org/10.1002/sim.9023).

Chenxi, L. (2016) "The Fine–Gray model under interval censored competing risks data." Journal of Multivariate Analysis, 143:327-344.

Cox, D. R. 1972. "Regression Models and Life-Tables." Journal of the Royal Statistical Society. Series B (Methodological) 34(2):187–220. Retrieved (http://links.jstor.org/sici?sici=0035-9246(1972)34:2<187:RMAL>2.0.CO;2-6\nhttp://www.jstor.org).

Deng, D. (2016). Estimating the cumulative mean function for history process with time-dependent covariates and censoring mechanism. Statistics in Medicine, 35(25), 4624–4636. Retrieved (https://doi.org/10.1002/sim.6998).

Fine, J. P. 2001. "Regression Modeling of Competing Crude Failure Probabilities." Biostatistics (Oxford, England) 2(1):85–97. Retrieved (http://www.ncbi.nlm.nih.gov/pubmed/12933558).

Fine, Jason P. and Robert J. Gray. 1999. "A Proportional Hazards Model for the Subdistribution of a Competing Risk." American Statistcal Association 94(446):496–509. Retrieved (http://www.jstor.org/stable/2670170).

Fox, J. (2008). "Cox Proportional-Hazards Regression for Survival Data" https://socialsciences.mcmaster.ca/jfox/Books/Companion-1E/appendix-cox-regression.pdf.

Hansen, S. N., Overgaard, M., Andersen, P. K., & Parner, E. T. (2017). Estimating a population cumulative incidence under calendar time trends. BMC Medical Research Methodology, 17(1). Retrieved (https://doi.org/10.1186/s12874-016-0280-6).

Heckman, J. J., & Honoré, B. E. (1989). The Identifiability of the Competing Risks Model. Biometrika, 76(2), 325-330. Retrieved (https://academic.oup.com/biomet/article-abstract/76/2/325/265352).

Hudgens, Michael G., Chenxi Li, and Jason P. Fine. 2014. "Parametric Likelihood Inference for Interval Censored Competing Risks Data." Biometrics 70(1):1–9.

Retrieved (http://www.ncbi.nlm.nih.gov/pubmed/24400873).

Jeong, Jong Hyeon and Jason Fine. 2006. “Direct Parametric Inference for the Cumulative.” Appl. Statist. 55(2):187–200.

Jeong, Jong-Hyeon and Jason P. Fine. 2007. “Parametric Regression on Cumulative Incidence Function.” Biostatistics (Oxford, England) 8(2):184–96. Retrieved (http://www.ncbi.nlm.nih.gov/pubmed/16636138).

John P. Klein, Hans C. van Houwelingen, Joseph G. Ibrahim,Thomas H. Scheike.2014. Handbook of Survival Analysis.

Kalbfleisch, John D. and Ross L. Prentice. 1980. The Statistical Analysis of Failure Time Data. New York: John wiley & sons,inc.

Klein, John P. and Per Kragh Andersen. 2005. “Regression Modeling of Competing Risks Data Based on Pseudovalues of the Cumulative Incidence Function.” Biometrics 61(1):223–29. Retrieved (http://www.ncbi.nlm.nih.gov/pubmed/15737097).

Liedtke C, Mazouni C, Hess KR, André F, Tordai A, Mejia JA, Symmans WF, Gonzalez-Angulo AM, Hennessy B, Green M, Cristofanilli M, Hortobagyi GN, Pusztai L. Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. J Clin Oncol. 2008 Mar 10;26(8):1275-81. doi: 10.1200/JCO.2007.14.4147. Epub 2008 Feb 4. PMID: 18250347.

Oyelola A., Gayawan E, Hanna F. (2017). “Spatial modeling of contribution of individual level risk factors for mortality from Middle East Respiratory Syndrome coronavirus in the Arabian Peninsula.” PloS One, 12(7):e0181215.

Oyelola A., Saffary T., Adegboye M., and Faiz Elfaki (2019). “Individual and network characteristic associated with hospital-acquired Middle East Respiratory Syndrome coronavirus.” Journal of Infection and Public Health. Retrieved (https://doi.org/10.1016/j.jiph.2018.12.002.).

Oyelola A. Adegboye and Faiz Elfaki (2018). “Network Analysis of MERS Coronavirus within Households, Communities and Hospitals to Identify Most Centralized and Super-Spreading in the Arabian Peninsula, 2012 to 2016.” Canadian Journal of Infectious diseases and medical microbiology. Retrieved (https://doi.org/10.1155/2018/6725284.).

Pintilie, Melania. 2006. Competing Risks: A Practical Perspective. John Wiley& Sons.

Rambaut A. (2013). “MERS-CoV spatial and epidemiological information. Retrieved (http://epidemic.bio.ed.ac.uk/coronavirus_background).

Salinas-Escudero, G., Carrillo-Vega, M.F., Granados-García, V. et al. A survival analysis of COVID-19 in the Mexican population. BMC Public Health 20, 1616 (2020). Retrieved (https://doi.org/10.1186/s12889-020-09721-2)

Scheike, Thomas H. and Mei-Jie Zhang. 2008. “Flexible Competing Risks Regression Modeling and Goodness-of-Fit.” Lifetime Data Analysis 14(4):464–83. Retrieved (http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2715961&tool=pmcentrez&rendertype=abstract).

Shayan, Zahra, Seyyed Mohammad Taghi Ayatollahi, and Najaf Zare. 2011. “A Parametric Method for Cumulative Incidence Modeling with a New Four-Parameter Log-Logistic Distribution.” Theoretical Biology & Medical Modelling 8:43. Retrieved (http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3713116&tool=pmcentrez&rendertype=abstract).

Stevenson, M. (2007). “An Introduction to Survival Analysis.” http://www.biecek.pl/statystykaMedyczna/Stevenson_survival_analysis_195.721.pdf

Suhaini, S. D, N. M. and Arasan, J. (2020). Non-Parametric Tests for Comparing More than Two Survivor Functions with Interval Censored Data via Midpoint Imputation. Menemui Matematik (Discovering Mathematics), 42(2), 86-95. Retrieved (https://myjms.mohe.gov.my/index.php/dismath/article/view/13074)

Yosra Yousif (2017). “Bayesian Analysis of Partly Interval-Censored Data from Engineering Applications under Competing Risks Framework.” PhD Thesis, IIUM. https://lib.iium.edu.my/mom/services/mom/document/getFile/4TPXzgbMn1ugCSENlfTabFbJuYfeXmmB20180227115732486

Yosra Yousif, Faiz A. M. Elfaki, Meftah Hrairi and Oyelola .A. Adegboye (2020). “A Bayesian Approach to Competing Risks Model with Masked Causes of Failure and Incomplete Failure Times." Mathematical Problems in Engineering, Vol. 2020, Article ID 8248640, 7 pages, 2020. Retrieved (https://doi.org/10.1155/2020/8248640).

Zhang, Z., Cortese, G., Combescure, C., Marshall, R., Lee, M., Lim, H., & Haller, B. (2018). Overview of model validation for survival regression model with competing risks using melanoma study data. Annals of Translational Medicine, 6(16), 325–325. Retrieved (https://doi.org/10.21037/atm.2018.07.38).