Recurrent Event Models: An Application to Offenders Found Not Criminally Responsible on Account of Mental Disorder and their Interactions with the Health Care and Criminal Justice Systems

by

Lillian Yan Lin

B.Sc., Simon Fraser University, 2016

Project Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Statistics and Actuarial Science
Faculty of Science

Approval

Name:	**Lillian Yan Lin**
Degree:	**Master of Science (Statistics)**
Title:	**Recurrent Event Models: An Application to Offenders Found Not Criminally Responsible on Account of Mental Disorder and their Interactions with the Health Care and Criminal Justice Systems**
Examining Committee:	**Chair:** Jinko Graham Professor
	Rachel Altman Senior Supervisor Associate Professor
	Martin Andresen Supervisor Professor School of Criminology
	David Alexander Campbell Internal Examiner Associate Professor
Date Defended:	**August 21, 2018**

Abstract

Prior to committing an offence for which they are ultimately found not criminally responsible (NCR), offenders may have contact with the health care and criminal justice systems. Understanding the frequency of these contacts can potentially help to prevent such offences by informing strategies for intervention. In particular, escalation in contact frequency could foreshadow the committing of an index offence. Inspired by real data, in this project, we investigate models that describe such escalation. In particular, we consider two classes of models: time-to-event models that are framed in terms of *numbers* of contacts in an interval, and time-between-events models that are framed in terms of *times* between two successive contacts. Both classes of models can incorporate predictor variables and between-subject heterogeneity (via random effects). The properties of the maximum likelihood estimators of the escalation rate and the performance of the Kolmogorov-Smirnov test of goodness-of-fit are assessed using simulations under various scenarios.

Keywords: NCR-Accused; Recurrent Events; Poisson Process; Gap Time Analysis

Dedication

To my beloved parents.

Acknowledgements

I first would like to thank my supervisor Dr. Rachel Altman for the opportunities, her endless guidance, support, patience and time throughout the past two years. I greatly appreciate the effort she has put into helping me become a more well-rounded statistician. Thank you Rachel for seeing the potential in me. And I would like to thank my committee members, Dr. Martin Andresen and Dr. Dave Campbell, for taking their time to participate in my defence. Thank you Martin for giving me the opportunity to work on this project. Thank you Dave for sparking my interest in the statistics graduate program here.

Many thanks to all the members in the Statistics and Actuarial Sciences Department at Simon Fraser University for creating an incredible learning environment. My gratitude goes out to all of the instructors whom I have taken courses with: Rachel Altman, Jinko Graham, Joan Hu, Tom Loughin, Brad McNeney and Boxin Tang. Thank you for your valuable lectures. To Charlene Bradbury, Kelly Jay and Sadika Jungic, thank you for all the support along the way.

To my fellow graduate students and friends here, specifically JinChoel Choi, Sarah Bailey, Richard Hsia, Grace Hsu, Jinwan Kim, Tian Li, Yue Ma, Ben Nie, Will Ruth, Abdollah Safari, Perry Sang, Chenlu Shi, Michelle Thiessen, Trevor Thomson, Jing Wang, Ran Wang, Adelaide Wu, Lucas Wu, Yuping Yang, Charlie Zhou. Thank you for your company, your help and laughs we shared alongside my journey here. Special mentions go to Yuping and Trevor for all the lessons you gave and time you spent to help me understand the materials and refine my codes.

To my other friends who are not always around, thank you for bringing joy to my life and for being there regardless. To Lindsay and Robbie, thank you for your encouragement and comfort. Most importantly, thank you for taking me to food when I don't eat properly.

Last but not least, thank you to my family for their unconditional love and support. A big thank you to my mom and dad who are always proud of me.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

According to Section 16 of the Criminal Code (Criminal Code, 1985), individuals are found not criminally responsible on account of mental disorder (NCRMD) for "an act committed or an omission made while suffering from a mental disorder that rendered the person incapable of appreciating the nature and quality of the act or omission or of knowing that it was wrong." Based on the summary provided by Miladinovic and Lukassen (2014), from the fiscal year 2005/2006 through to 2011/2012, across all provinces and territories in Canada, 63% of NCRMD cases involved crimes against a person, whereas only 24% of non-NCRMD cases involved crimes against a person. Major assault is the most frequent offence type in NCRMD cases (20%), while only 5% of non-NCPMD cases are classified as major assault.

Some studies on the NCRMD population have been conducted to investigate the forensic psychiatric system and potentially help with early intervention and prevention of the offence. In a study conducted by Golding *et al.* (1989), all persons found not guilty by reason of insanity (NGRI) between November, 1975 and January 1, 1984 in British Columbia were included. They found that only 16% of NGRI-acquitted individuals had no mental health contacts of any kind prior to indexed offences. The data showed that the majority of individuals found NGRI were in the forensic and/or mental health system prior to their indexed offences. Similarly, Livingston *et al.* (2003) reported that 63% of persons found NCRMD between February 1992 and February 1998 in British Columbia, Canada, were involved with the criminal justice system, and 51.8% of them had been admitted to a general psychiatric facility at least 4 times prior to the indexed offence.

The studies mentioned above imply that offences committed by NCRMD-accused people are potentially foreseeable and thus preventable. This project lays the statistical groundwork for an intervention strategy based on prior health-care admissions and criminal justice system history.

The specific goal of this project is to explore appropriate models and preliminary tools for data analysis in this setting. One challenge is the handling of non-conventional left-censoring. We develop special techniques for this purpose. Using our models, we examine

the performance of inference methods for parameters that represent the escalation in rate of contact with the two systems.

The outline of this paper is as follows. In Chapter 2, we describe the NCRMD data context. In Chapter 3, we provide details of the proposed methods. In Chapter 4, we describe simulation studies for investigating inference methods for the parameters of primary interest. We discuss limitations of the data and our approaches, as well as possibilities for extensions to our methods, in Chapter 5.

Chapter 2

NCRMD Data

2.1 Data Description

Due to a confidentiality agreement, only the *context* surrounding the real data will be described. Plots and parameter estimates derived using real data will not be shown. All analysis is based on simulated data that have been generated to mimic the real data.

In this setting, a *contact* with a system is an interaction with the system. In other words, a contact occurs each time the person has an interaction with the police or with health agencies. In statistical terms, a contact is called an *event*.

The data constitute a retrospective cohort study of individuals who committed an index offence but were subsequently deemed NCRMD. The cohort comprises both men and women, with a very low female-to-male ratio. In addition to date of birth and sex, information about each subject's contacts with the criminal justice and health care systems were also gathered for the 12-month period prior to the date of the index offence.

Criminal justice events encompass events concerning criminal violence, criminal nonviolence, and general police assistance. Health care events encompass events concerning social services, medical physical examinations, medical voluntary psychological treatments, etc. Most subjects have more health care events than criminal justice events. In fact, some subjects have only a single criminal justice event, i.e., no criminal justice event prior to the index offence. All subjects have at least two health care events.

Since all subjects were admitted to the forensic hospital for treatment following the index offence, we treat the index offence as both a criminal justice and health care event. In other words, the index offence of each subject represents the last event this person had with each system. We will therefore call the index offence the terminal event. Note that the terminal event corresponds to the only index offence committed by a given subject. However, some subjects may have committed less serious offences over the one-year period (for which they were not deemed NCRMD). These events were also recorded as criminal justice contacts.

The one-year observation window *ends* at the time of a subject's index offence and begins 365 days prior (at time $t = 0$). Therefore, the first event time recorded provides only partial information about the time between this event and the preceding event (which occurred prior to time 0). In other words, a subject's first inter-event time is censored.

Miladinovic and Lukassen (2014) reported that fewer than 1% of the adult criminal court cases are NCRMD cases, which means that NCRMD-accused people form a very small group. Given this fact, in our work, we assume that the sample size is relatively small (on the order of 100).

2.2 Question of Interest

Figure 2.1 displays simulated health event times for each individual (in elapsed days since the start of their observation period). The calendar dates of the observation period for each subject can be different, but the total observation time is the same for all subjects (365 days). Each horizontal line represents one subject's observation period, and each black dot represents an event. The event time plot shows that events tend to occur more frequently towards the end of the observation period. We consider a model to quantify this feature of the data via changes in the intensity function over time. Looking at the data from a different perspective, wait times between two successive events tend to be shorter than previous wait times. Therefore, we also consider a class of models that is specified in terms of the relationship between inter-event times.

Other plots, e.g., the cumulative sample mean function plot, can also be used to show the escalation of the event rate. Specifically, suppose there are m subjects, each with an event process. Let $N_i(t)$ denote the number of events over the time interval $[0, t]$ for subject i. Then, the cumulative sample mean function is

$$\hat{\mu}(t) = \frac{1}{m} \sum_{i=1}^{m} N_i(t),$$

where we assume that events occur in continuous time.

Figure 2.2 shows the plot of $\hat{\mu}(t)$ using simulated health care event times for a collection of individuals. The convex shape of $\hat{\mu}(t)$ suggests an increase in event rate over the observation period. This notion leads to the main objective of this project: to model the escalation in intensity of events. We will model the individuals' contacts with the two systems (the criminal justice system and the health care system) separately over time.

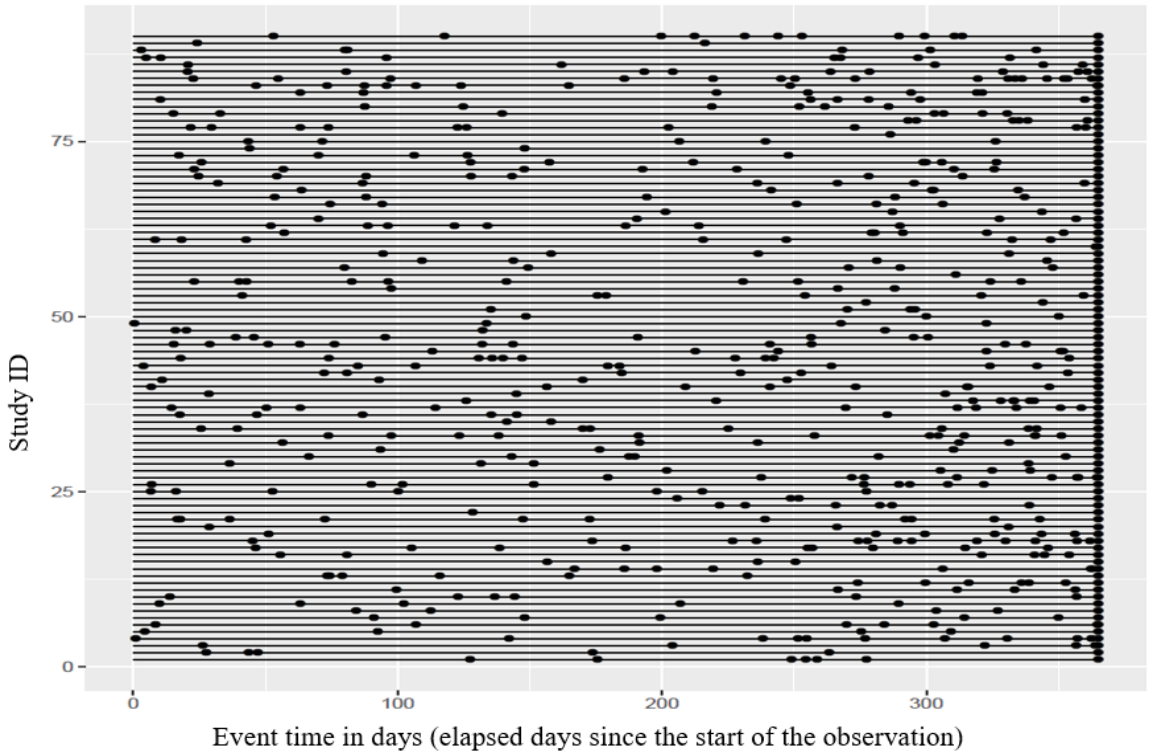


Figure 2.1: (Simulated) event time plot for health care system events (all subjects)

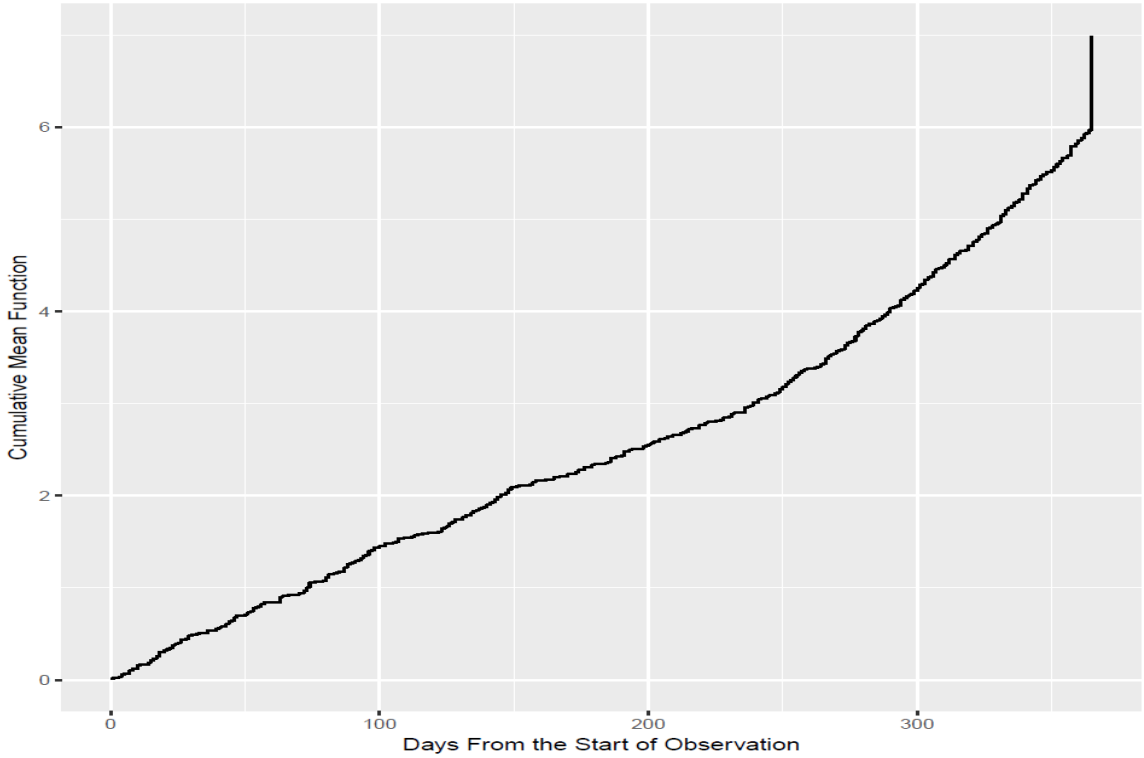


Figure 2.2: Cumulative sample mean function for (simulated) recurring contacts with the health care system

Chapter 3

Methods

We have recurrent event data, and thus consider time-to-event models and time-between-events models – two broad classes of modelling techniques for such data. Time-to-event models are specified in terms of the occurrence rate of events (the intensity function), whereas time-between-events models are specified in terms of gap times between events (the hazard function). We use the Poisson process (a counting process that treats event counts in non-overlapping time intervals as independent) as a basis for the time-to-event models we investigate. In contrast, models for gap times are specified in terms of the distribution of the current gap time conditional on the history of the process (and possibly other variables as well, such as random effects). We consider only gap time models where the gap times are independent given the most recent gap times.

In the following sections, we describe how these models can be applied to NCRMD data and, in particular, how to handle the particular type of censored data observed in this context. We also outline the goodness-of-fit methods that we consider.

3.1 Time-to-Event Models

In this section, we describe the time-to-event models considered in this project. We define notation and the notion of an intensity function in Section 3.1.1, discuss our approach to censoring in Section 3.1.2, describe the specific models we apply in the NCRMD context in Section 3.1.3, and outline our methods of estimation in Section 3.1.4.

3.1.1 The Intensity Function

For a recurrent event process starting at $t_0 = 0$, let $0 = T_{i0} < T_{i1} < T_{i2} < ... < T_{i,n_i} = 365$ denote the event times for individual i, $i = 1, \ldots, m$, where T_{i0} is the start of the observation period. Since the terminal event is also the end of the observation period, T_{i,n_i} is an event time as well as the end of the observation period. We define $N_i(t) = \sum_{j=1}^{n_i} \mathbf{1}(T_j \leq t)$ as the cumulative number of events occurring over the time interval $[0, t]$. We let $\Delta N_i(t)$ denote the number of events occurring over the interval $[t, t + \Delta t)$, and let $H_i(t) = \{N_i(s) : 0 \leq s < t\}$

denote the history of the process up to time t. The intensity function is defined as

$$\lambda_i(t|H(t)) = \lim_{\Delta t \downarrow 0} \frac{\Pr(\Delta N_i(t) = 1|H_i(t))}{\Delta t}. \tag{3.1}$$

This intensity function implies that the instantaneous probability of an event occurring between times t and $t + \Delta t$, conditional on the event history $H(t)$, is $\lambda_i(t|H(t))dt$. Given $H(t)$, Cook and Lawless (2007) defined the likelihood contribution from subject i (who is assumed to have had n_i events at times $t_1 < \ldots < t_{n_i}$ over the time interval $[0, \tau]$) as

$$\left(\prod_{j=1}^{n_i} \lambda_i(t_{ij}|H(t)) \right) \exp\left(- \int_0^\tau \lambda_i(t|H(t))dt \right),$$

where τ is the end of the observation period ($\tau = 365$ in our context).

3.1.2 Backward Right Censoring of First Event Times

Since the data collection was retrospective, for modelling time to events, we treat the process as going backwards in time, starting at the time of the terminal event ($t = 365$). Since the data reflect only event times that occurred within the year prior to the terminal event, it is possible that some individuals had events before the start time, i.e., there might be observations before $t = 0$ for some individuals. In these cases, because of the data collection protocol, we will only partially observe the first (in chronological time) gap time. By modelling the backward process, we can then treat this first gap time as conventional right-censored data.

As an aside, neither conventional left-censoring nor left-truncation occurs in our case. Conventional left-censoring is present when we know the censoring time, C, and we know the event has occurred before the censoring time (Klein and Moeschberger, 2006). For example, if we want to know the age when patients with a mental disorder first receive treatment, then we will have observations of actual event times on patients who started treatment after they entered the study. We will also know the left-censoring times of those who started treatment before they entered. I.e., we observe the maximum of the event time and left-censoring time for each patient. In contrast, left-truncation is present when we observe event times only on those individuals who have responses above a certain truncation limit. For example, if we want to know the age when adult patients with a mental disorder first receive a treatment, patients who are younger than 18 years old at the time of recruitment are excluded from the study.

Let t be the elapsed time after the start of the observation period ($t = 0$), and u be the time remaining until the time of the terminal event ($t = 365$). The forward counting process is $N(t) = \int_0^t dN(s)$, the cumulative event counts over the time interval $[0, t]$. The

backward counting process is $M(u) = \int_{\tau-u}^{\tau} dM(s)$, the number of events that occur over the time interval $[\tau - u, \tau]$. Figure 3.1 illustrates the forward and backward counting processes.

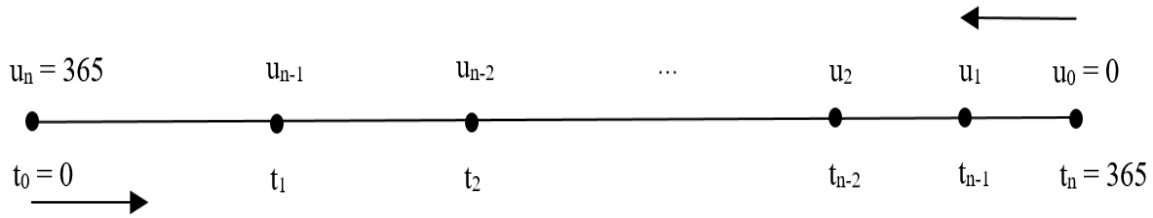


Figure 3.1: Forward and backward counting processes, $N(t)$ and $M(u)$

3.1.3 Model Specification

For this project, the only time-to-event models we consider are Poisson process-based, where, given subject-specific random effects, $v_1, \ldots, v_m$, inter-event times are assumed independent. In this case, the intensity function (3.1) becomes

$$\lambda_i(t|H(t), v_i) = \lim_{\Delta t \downarrow 0} \frac{\Pr(\Delta N_i(t) = 1 \mid v_i)}{\Delta t} = \rho_i(t \mid v_i), \tag{3.2}$$

i.e., the intensity function does not depend on the event history $H(t)$.

We propose a model that reflects the increasing rate of events over time by including a time effect parameter, α. In addition, Nagin and Land (2006) pointed out that rate of offending may vary with age and other individual-level characteristics, observable and unobservable. To characterize this heterogeneity among subjects, in addition to including a subject-specific random effect, we also extend the rate function (3.2) to incorporate the covariate age. Other predictors, such as gender, could be also be incorporated using the same framework. .

More specifically, let x_i denote the i^{th} subject's age. Let v_i follow a gamma distribution with mean 1 (for model identifiability) and variance θ. Given the random effect, the conditional rate function forwards in time is assumed to be of the form

$$\rho_i^*(t|v_i) = v_i \rho_i^*(t) = v_i e^{\gamma + \alpha t} e^{\nu x_i}.$$

In terms of the backward process, the conditional rate function is

$$\rho_i(u|v_i) = v_i \rho_i(u) = v_i e^{\gamma + \alpha(\tau - u)} e^{\nu x_i}.$$

By implication, the conditional rate function for subject i at the time origin ($u = 365$) is $e^{\gamma + \nu x_i}$, which we interpret as the conditional baseline rate for this subject.

The conditional cumulative rate function, Λ_i, for subject i over an interval $[0, u]$ is then

$$\Lambda_i(u|v_i) = \int_0^u \rho_i(s|v_i)ds = \frac{v_i}{\alpha} e^{\gamma+\alpha\tau}(1 - e^{-\alpha u})e^{\nu x_i}.$$

Let $\boldsymbol{\eta} = \{\gamma, \alpha, \nu, \theta\}$ be the vector of all parameters. As discussed by Cook and Lawless (2007), the contribution to the conditional likelihood (given the random effect) from subject i can be expressed as

$$\mathcal{L}_i(\boldsymbol{\eta}|v_i) = \exp\Big(-\int_0^\tau v_i\rho_i(s)ds\Big)\left(\prod_{j=1}^{n_i-1} v_i\rho_i(u_{ij})\right)$$

Note that this likelihood function also includes the contribution from the last (censored) event time (where "last" is defined in backwards time). The marginal likelihood for subject i is therefore

$$\mathcal{L}_i(\boldsymbol{\eta}) = \int_0^\infty \exp\Big(-\int_0^\tau v_i\rho_i(s)ds\Big)\left(\prod_{j=1}^{n_i-1} v_i\rho_i(u_{ij})\right) G(v_i)dv_i,$$

where $G(\cdot)$ is the density function of the random effect. When the random effect is gamma distributed, the backward counting process $(M(u_{i0}), ..., M(u_{i,n_i}))$ defined in Section 3.1.2 forms a negative binomial process (Cook and Lawless, 2007). The resulting marginal likelihood function is

$$\mathcal{L}_i(\boldsymbol{\eta}) = \frac{\Gamma(n_i + \theta^{-1})}{\Gamma(\theta^{-1})} \frac{(\theta\mu_i(\tau))^{n_i}}{(1 + \theta\mu_i(\tau))^{n_i+\theta^{-1}}} \left\{\prod_{j=1}^{n_i-1} \frac{\rho_0(u_{ij})}{\mu_0(\tau)}\right\}, \tag{3.3}$$

where $\rho_0(u) = e^{\gamma+\alpha(\tau-u)}$, $\mu_0(\tau) = \int_0^\tau \rho_0(s)ds$ and $\mu_i(\tau) = \int_0^\tau \rho_i(s)ds$.

The complete log-likelihood is then

$$\begin{aligned}
\ell(\boldsymbol{\eta}) &= \log\Big(\prod_{i=1}^m \mathcal{L}_i(\boldsymbol{\eta})\Big) \\
&= \sum_{i=1}^m \Big\{ n_i \log\mu_i(\tau) - (n_i + \theta^{-1})\log(1 + \theta\mu_i(\tau)) \\
&\quad + \sum_{j=1}^{n_i^*} [\log\rho_0(u_{ij}) - \log\mu_0(u_{ij})] + \sum_{j=1}^{n_i^*} \log(1 + \theta j) \Big\},
\end{aligned}$$

where $n_i^* = \max(1, n_i - 1)$. The parameters are estimated as the maximizer of $\ell(\boldsymbol{\eta})$.

The rate function can also be modified to reflect special cases of interest (e.g., no time trend, no age, no random effect, other predictors). Below we consider two such cases that are practical in the present context.

Rate Function with Time-Independent Predictors, $\mathbf{X}_i$, Only

Consider the case where the rate function does not depend on time or random effects, but may depend on other predictors. Representing these predictors by $\mathbf{x}_i$, the (constant) rate function for subject i is

$$\rho_i(u) = \rho_i = e^{\gamma} e^{\mathbf{x}_i'\boldsymbol{\nu}},$$

where $\boldsymbol{\nu}$ is a vector of regression parameters.

The complete log-likelihood is

$$\begin{aligned}
\ell(\gamma, \boldsymbol{\nu}) &= \log\left(\prod_{i=1}^{m} \mathcal{L}_i(\gamma, \boldsymbol{\nu})\right) \\
&= \log\left(\prod_{i=1}^{m} \exp\Big(-\int_0^{\tau} \rho_i ds\Big)\left(\prod_{j=1}^{n_i-1} \rho_i\right)\right) \\
&= \log\left(\prod_{i=1}^{m} \exp\Big(-\rho_i\tau\Big){\rho_i}^{n_i}\right) \\
&= \sum_{i=1}^{m}\Big[-\tau e^{\gamma} e^{\mathbf{x}_i'\boldsymbol{\nu}} + n_i(\gamma + \mathbf{x}_i'\boldsymbol{\nu})\Big] \\
&= -\tau e^{\gamma}\sum_{i=1}^{m} e^{\mathbf{x}_i'\boldsymbol{\nu}} + \gamma\sum_{i=1}^{m} n_i + \sum_{i=1}^{m} n_i \mathbf{x}_i'\boldsymbol{\nu},
\end{aligned}$$

Rate Function with Time Trend α only

Alternatively, the rate function may depend on time but not on other predictors. In this case, the (time varying) rate function for subject i is

$$\rho_i(u) = e^{\gamma + \alpha(\tau - u)}.$$

The cumulative rate function Λ over an interval $[0, u]$ is then

$$\Lambda_i(u) = \int_0^u \rho_i(s) ds = \tfrac{1}{\alpha} e^{\gamma + \alpha\tau}(1 - e^{-\alpha u}).$$

The contribution to the likelihood from subject i can be expressed as

$$\begin{aligned}
\mathcal{L}_i(\gamma, \alpha) &= \exp\left(-\Lambda_i(\tau)\right)\left(\prod_{j=1}^{n_i-1} \rho_i(u_{ij})\right) \\
&= \exp\Big(-\tfrac{1}{\alpha} e^{\gamma + \alpha\tau}(1 - e^{-\alpha\tau})\Big) \exp\left(n_i\gamma + \alpha\sum_{j=1}^{n_i-1}(\tau - u_{ij})\right).
\end{aligned}$$

The complete log-likelihood is

$$\begin{aligned}\ell(\gamma,\alpha) &= \log\left(\prod_{i=1}^{m}\mathcal{L}_i(\gamma,\alpha)\right)\\ &= -\tfrac{m}{\alpha}e^{\gamma+\alpha\tau}(1-e^{-\alpha\tau})+\sum_{i=1}^{m}n_i\gamma+\sum_{i=1}^{m}\sum_{j=1}^{n_i-1}\alpha(\tau-u_{ij}).\end{aligned}$$

3.1.4 Model Estimation

Let $U(\boldsymbol{\eta})$ be a $p\times 1$ vector of partial derivatives of $\ell(\boldsymbol{\eta})$, where $\boldsymbol{\eta}$ is the vector of p parameters, i.e.,

$$U(\boldsymbol{\eta}) = \frac{\partial\ell(\boldsymbol{\eta})}{\partial\boldsymbol{\eta}} = \begin{pmatrix}\frac{\partial\ell}{\partial\eta_1}\\ \vdots \\ \frac{\partial\ell}{\partial\eta_p}\end{pmatrix}$$

The vector $U(\boldsymbol{\eta})$ is the score function. We let $I(\boldsymbol{\eta})$ be a $p\times p$ matrix of the negative second derivatives of $\ell(\boldsymbol{\eta})$,

$$I(\boldsymbol{\eta}) = -\frac{\partial^2\ell}{\partial\eta_i\partial\eta_j},$$

where $i,j = 1,2,\ldots p$. The matrix $I(\boldsymbol{\eta})$ is the observed information matrix. Under some conditions, the maximum likelihood estimates $\hat{\boldsymbol{\eta}}$ are unique and are solutions to $U(\boldsymbol{\eta})=0$.

For the case of $\rho_i(u) = e^{\gamma+\alpha(\tau-u)}$, the elements of the score vector are

$$\begin{aligned}U_\alpha(\boldsymbol{\eta}) &= \frac{\partial\ell(\boldsymbol{\eta})}{\partial\alpha} = \tfrac{m}{\alpha^2}e^{\gamma+\alpha\tau}-\tfrac{m\tau}{\alpha}e^{\gamma+\alpha\tau}-\tfrac{m\tau}{\alpha}e^{\gamma}+\sum_{i=1}^{m}\sum_{j=1}^{n_i-1}(\tau-u_{ij})\\ U_\gamma(\boldsymbol{\eta}) &= \frac{\partial\ell(\boldsymbol{\eta})}{\partial\gamma} = -\tfrac{m}{\alpha}e^{\gamma+\alpha\tau}-\tfrac{m}{\alpha}e^{\gamma}+\sum_{i=1}^{m}n_i,\end{aligned}$$

and the elements of the information matrix are

$$\begin{aligned}I_{\alpha\alpha}(\boldsymbol{\eta}) &= -\frac{\partial^2\ell}{\partial\alpha^2} = \tfrac{2m}{\alpha^3}e^{\gamma}\left(e^{\alpha\tau}-1\right)-\tfrac{2m\tau}{\alpha^2}e^{\gamma+\alpha\tau}+\tfrac{m\tau^2}{\alpha}e^{\gamma+\alpha\tau}\\ I_{\alpha\gamma}(\boldsymbol{\eta}) &= -\frac{\partial^2\ell}{\partial\alpha\partial\gamma} = \tfrac{m}{\alpha^2}e^{\gamma}\left(1-e^{\alpha\tau}\right)+\tfrac{m\tau}{\alpha}e^{\gamma+\alpha\tau}\\ I_{\gamma\alpha}(\boldsymbol{\eta}) &= -\frac{\partial^2\ell}{\partial\gamma\partial\alpha} = I_{\alpha\gamma}(\boldsymbol{\eta})\\ I_{\gamma\gamma}(\boldsymbol{\eta}) &= -\frac{\partial^2\ell}{\partial\gamma^2} = \tfrac{m}{\alpha}e^{\gamma}\left(e^{\alpha\tau}-1\right).\end{aligned}$$

Therefore, the information matrix is

$$I(\boldsymbol{\eta}) = \begin{pmatrix} I_{\alpha\alpha}(\boldsymbol{\eta}) & I_{\alpha\gamma}(\boldsymbol{\eta}) \\ I_{\gamma\alpha}(\boldsymbol{\eta}) & I_{\gamma\gamma}(\boldsymbol{\eta}) \end{pmatrix}.$$

For the case of $\rho_i(u) = v_i e^{\gamma+\alpha(\tau-u)}$, the elements of the score vector are

$$\begin{aligned} U_\alpha(\boldsymbol{\eta}) &= \frac{\partial \ell(\boldsymbol{\eta})}{\partial \alpha} = \sum_{i=1}^{m}\sum_{j=1}^{n_i-1}(\tau - u_{ij}) - \frac{\frac{\theta}{\alpha}e^{\gamma}\left(\tau e^{\alpha\tau} - \frac{1}{\alpha}(e^{\alpha\tau}-1)\right)}{\left(1+\frac{\theta}{\alpha}e^{\gamma}(e^{\alpha\tau}-1)\right)}\sum_{i=1}^{m}(n_i+\theta^{-1}) \\ U_\gamma(\boldsymbol{\eta}) &= \frac{\partial \ell(\boldsymbol{\eta})}{\partial \gamma} = \sum_{i=1}^{m} n_i - \frac{\frac{\theta}{\alpha}e^{\gamma}\left(\tau e^{\alpha\tau}-1)\right)}{\left(1+\frac{\theta}{\alpha}e^{\gamma}(e^{\alpha\tau}-1)\right)}\sum_{i=1}^{m}(n_i+\theta^{-1}), \end{aligned}$$

and the information matrix is

$$I(\boldsymbol{\eta}) = \begin{pmatrix} -\frac{\partial U_\alpha(\boldsymbol{\eta})}{\partial \alpha} & -\frac{\partial U_\alpha(\boldsymbol{\eta})}{\partial \gamma} \\ -\frac{\partial U_\gamma(\boldsymbol{\eta})}{\partial \alpha} & -\frac{\partial U_\gamma(\boldsymbol{\eta})}{\partial \gamma} \end{pmatrix}.$$

Standard errors of the estimates can be obtained as the square root the diagonal elements of $I^{-1}(\hat{\boldsymbol{\eta}})$.

3.2 Time-Between-Events Models

In this section, we describe the second class of models we consider in this project: time-between-events models, also known as gap time models.

3.2.1 Gap Times

We define $W_{ij} = t_{ij} - t_{i,j-1}$ as the gap time, that is the wait time between events $j-1$ and j on subject i. For this class of models, each event time is considered the new time origin for next event, and gap times may be correlated.

Let $f_{W_{ij}}(w)$ be the probability density function (pdf) of W_{ij} and let the cumulative distribution function (cdf) be

$$F_{W_{ij}}(w) = \Pr(W_{ij} \leq w) = \int_0^w f_{W_{ij}}(t)dt.$$

Then, the probability of observing a gap time longer than w is

$$S_{W_{ij}}(w) = \Pr(W_{ij} \geq w) = \int_w^\infty f_{W_{ij}}(t)dt.$$

Gap time models are specified in terms of the conditional distribution of each gap time:

$$h_{ij}(w|H(w)) = \lim_{\Delta w \to 0} \frac{\Pr(w \leq W_{ij} < w + \Delta w | W_{ij} \geq w, H(w))}{\Delta w} = \frac{f_{W_{ij}|H(w)}(w)}{S_{W_{ij}|H(w)}(w)}. \quad (3.4)$$

Here, $h_{ij}(w|H(w))$ is called the hazard function of the gap time W_{ij}. Similar to the intensity function, the hazard function defines the instantaneous rate of an event at time w, conditional on the gap time being at least w and on the history of the process. In fact, the hazard function is related to the intensity function (3.1):

$$\lambda_{ij}(t \mid H(t)) = h_{ij}(t - T_{iN(t^-)}|H(w)),$$

where $t - T_{iN(t^-)}$ is the time since subject i's most recent event (measured at time t).

The survival function $S_{W_{ij}|H(w)}(w)$ can be expressed in terms of the hazard function since

$$h_{ij}(w|H(w)) = -\frac{S'_{W_{ij}|H(w)}(w)}{S_{W_{ij}|H(w)}(w))} = \frac{d}{dw} \log S_{W_{ij}|H(w)}(w).$$

Specifically,

$$S_{W_{ij}|H(w)}(w) = \exp\left(-\int_0^w h_{ij}(s|H(s))ds\right).$$

3.2.2 Censoring

As described in Section 3.1.2, some events could have occurred earlier than the start time $t = 0$. In this case, the first observed event time, t_{i1}, is likely less than the first gap time, w_{i1}. If we assume $w_{i1} = t_{i1}$, then we may introduce bias in our parameter estimates. Alternatively, if we ignore the (incomplete) first gap times, we lose substantial information. Therefore, as in our time-to-events models, in our time-between-event models, we treat the first gap time as censored. However, for time-between-events models, we work in forwards time.

3.2.3 Model Specification

We use the generic notation $f(x)$ to denote the density of a random variable X. Let x_i denote the i^{th} subject's age at time $t = 0$. We allow gap time to be dependent on previous gap times by considering hazard functions of the form

$$h_{ij}(w|H(w)) = h_{ij}(w \mid x_i, w_{i,j-1}).$$

This hazard function can capture increasing intensity of events via a negative association between consecutive gap times. Following Pénichoux *et al.* (2015), event-dependence is included by treating the previous gap time $w_{i,j-1}$ as an internal covariate.

The likelihood is a function not only of the conditional densities of the fully observed gap times, but also of the marginal density of the first (censored) gap time. Let the last event time in $(-\infty, 0]$ be t_{i0} and let $\boldsymbol{\eta}$ be the vector of all parameters. Suppose the gap time distribution for subject i is specified by the survivor function $S_{W_{ij}}(w)$ (or, equivalently, the density $f_{W_{ij}}(w)$). The contribution to the likelihood from subject i can be written as

$$\begin{aligned}
\mathcal{L}_i(\boldsymbol{\eta}) &= f(t_{i1}, \ldots, t_{in_i}) \\
&= \int_{t_{i1}}^{\infty} f(t_{i1}, \ldots, t_{in_i} \mid w_{i1}) f(w_{i1} \mid W_{i1} > t_{i1}) dw_{i1} \\
&= \left[\int_{t_{i1}}^{\infty} \mathbf{1}_{\left\{ t_{i1} = \tau - \sum_{j=2}^{n_i} w_{ij} \right\}} \frac{f_{W_{i1}}(w_{i1})}{S_{W_{i1}}(t_{i1})} f_{W_{i2}|W_{i1}}(w_{i2} \mid w_{i1}) dw_{i1} \right] \left[\prod_{j=3}^{n_i} f_{W_{ij}|W_{i,j-1}}(w_{ij} \mid w_{i,j-1}) \right]
\end{aligned}$$

The complete log-likelihood is then

$$\ell(\boldsymbol{\eta}) = \sum_{i=1}^{m} \log \mathcal{L}_i(\boldsymbol{\eta}).$$

In both the CJS and health care system data sets, some subjects have only two events. The product $\prod_{j=3}^{n_i} f_{W_{ij}|W_{i,j-1}}(w_{ij} \mid w_{i,j-1})$ will vanish in this case. The other special case is where a subject has no events other than last terminal event (a scenario that happens more often in the CJS data than in the health care system data). The contribution to the likelihood from a subject, i, with only a terminal event is $\mathcal{L}_i(\boldsymbol{\eta}) = \int_{\tau}^{\infty} f(w_{i1}) dw_{i1} = S_{W_{i1}}(\tau)$, i.e., the probability of observing a first gap time longer than τ.

We consider a few common parametric distributions for gap times here. We specify the a marginal hazard rate for the first gap time W_{i1} and conditional hazard functions for subsequent gap times (i.e., conditional on previous gap times).

Exponential Distribution

If W_{i1} and $W_{ij} \mid W_{i,j-1}$ follow an exponential distribution, then we assume the following hazard function for subject i:

$$h_{ij}(w) = \begin{cases} e^{\gamma + \nu x_i} & \text{for } j = 1 \\ \\ e^{\gamma + \beta w_{i,j-1} + \nu x_i} & \text{for } j > 1 \end{cases}$$

The density of W_{i1} is

$$f_{W_{i1}}(w) = \exp(\gamma + \nu x_i) \exp(-w e^{\gamma + \nu x_i}),$$

the survivor function for W_{i1} is

$$S_{W_{i1}}(w) = \exp(-we^{\gamma+\nu x_i}),$$

and the density of $W_{ij} \mid W_{i,j-1}$, $j = 2, \ldots, n_i$ is

$$f_{W_{ij}|W_{i,j-1}}(w) = \exp(\gamma + \beta w_{i,j-1} + \nu x_i) \exp(-we^{\gamma+\beta w_{i,j-1}+\nu x_i}).$$

Weibull Distribution

If W_{i1} and $W_{ij} \mid W_{i,j-1}$ follow a Weibull distribution with shape $a > 0$, then we assume the following hazard function for subject i:

$$h_{ij}(w) = \begin{cases} \left(\frac{a}{b}\right)\left(\frac{w}{b}\right)^{a-1} e^{\nu x_i} & \text{for } j = 1 \\ \left(\frac{a}{b}\right)\left(\frac{w}{b}\right)^{a-1} e^{\beta w_{i,j-1}+\nu x_i} & \text{for } j > 1 \end{cases}$$

The density of W_{i1} is

$$f_{W_{i1}}(w) = \left(\frac{a}{b}\right)\left(\frac{w}{b}\right)^{a-1} \exp\left(\nu x_i\right) \exp\left(-\left(\frac{w}{b}\right)^{a} e^{\nu x_i}\right),$$

the survivor function for W_{i1} is

$$S_{W_{i1}}(w) = \exp\left(-\left(\tfrac{w}{b}\right)^{a} e^{\nu x_i}\right),$$

and the density of $W_{ij} \mid W_{i,j-1}$, $j = 2, \ldots, n_i$ is

$$f_{W_{ij}|W_{i,j-1}}(w) = \left(\frac{a}{b}\right)\left(\frac{w}{b}\right)^{a-1} \exp\left(\beta w_{i,j-1} + \nu x_i\right) \exp\Big\{-\left(\frac{w}{b}\right)^{a} e^{\beta w_{i,j-1}+\nu x_i}\Big\}.$$

Note that the exponential distribution is a special case of the Weibull distribution with $a = 1$.

Log-normal Distribution

If W_{i1} and $W_{ij} \mid W_{i,j-1}$ follow a log-normal distribution, then we assume the following hazard function for subject i:

$$h_{ij}(w) = \begin{cases} h_Z(z_{i1}) \exp\left(-\mu_{i1} - \sigma z_{i1}\right)/\sigma & \text{for } j = 1 \\ h_Z(z_{ij}) \exp\left(-\mu_{ij} - \sigma z_{ij}\right)/\sigma & \text{for } j > 1 \end{cases}$$

where $h_Z(x) = \frac{\phi(x)}{\Phi(-x)}$ is the hazard function associated with the standard normal distribution. (We use $\phi(\cdot)$ to denote the cumulative distribution function of the standard normal

distribution and $\Phi(\cdot)$ to denote the probability density function of the standard normal distribution.) Here $\mu_{i1} = \nu_0 + \nu_1 x_i$, $\mu_{ij} = \nu_0 + \beta w_{i,j-1} + \nu_1 x_i$, and $z_{ij} = \frac{\log(w_{ij}) - \mu_{ij}}{\sigma}$.

The density of W_{i1} is

$$f_{W_{i1}}(w) = \frac{1}{w\sigma\sqrt{2\pi}} e^{-\frac{(\ln w - \mu_{i1})^2}{2\sigma^2}},$$

the survivor function for W_{i1} is

$$S_{W_{i1}}(w) = 1 - \Phi\left(\frac{\ln w - \mu_{i1}}{\sigma}\right),$$

and the density of $W_{ij} \mid W_{i,j-1}$, $j = 2, \ldots, n_i$ is

$$f_{W_{ij}|W_{i,j-1}}(w) = \frac{1}{w\sigma\sqrt{2\pi}} e^{-\frac{(\ln w - \mu_{ij})^2}{2\sigma^2}}.$$

Log-Logistic Distribution

If W_{i1} and $W_{ij} \mid W_{i,j-1}$ follow a Log-Logistic distribution with positive shape parameter κ, then we assume the following hazard function for subject i:

$$h_{ij}(w) = \begin{cases} \frac{\lambda\kappa(\lambda w)^{\kappa-1}}{1+(\lambda w)^\kappa} e^{\nu x_i} & \text{for } j = 1 \\ \frac{\lambda\kappa(\lambda w)^{\kappa-1}}{1+(\lambda w)^\kappa} e^{\beta w_{i,j-1} + \nu x_i} & \text{for } j > 1 \end{cases}$$

The density of W_{i1} is

$$f_{W_{i1}}(w) = \frac{\lambda\kappa(\lambda)^{\kappa-1}}{(1+(\lambda w)^\kappa)^2} \exp(\nu x_i) \exp(e^{\nu x_i}),$$

the survivor function for W_{i1} is

$$S_{W_{i1}}(w) = \frac{1}{1+(\lambda w)^\kappa} \exp(e^{\nu x_i}),$$

and the density of $W_{ij} \mid W_{i,j-1}$, $j = 2, \ldots, n_i$ is

$$f_{W_{ij}|W_{i,j-1}}(w) = \frac{\lambda\kappa(\lambda w)^{\kappa-1}}{(1+(\lambda w)^\kappa)^2} \exp(\beta w_{i,j-1} + \nu x_i) \exp(e^{\beta w_{i,j-1} + \nu x_i}).$$

3.2.4 Model Estimation

Direct maximization of the likelihood function requires the integration of a complex function with respect to w_{i1}, and the integral does not usually have a closed form solution. We use Gauss-Laguerre quadrature rules to numerically approximate the integral by weighted sums.

A Gaussian quadrature rule is

$$\int_a^b w(x)f(x)dx \approx \sum_{k=1}^{n} \alpha_k f(x_k), \qquad x_1 < x_2 < \ldots x_n,$$

where $w(x)$ is the weight function, α_k are weights, and x_k are quadrature points. When the weight function is $w(x) = \exp(-x)$, Gaussian quadrature is called Gauss-Laguerre quadrature. We can use this method to evaluate all of the likelihood functions that we consider since they all include an exponential factor (doing some variable transformations to derive the corresponding function $f(x)$, as necessary). We use the `glaguerre.quadrature` function in the `gaussquad` package to evaluate the integral. General optimization software can produce the Hessian matrix, $-I(\hat{\boldsymbol{\eta}})$, where $\boldsymbol{\eta}$ is the vector of parameters for a specific distribution. An estimate of the asymptotic covariance matrix of $\hat{\boldsymbol{\eta}}$ can be obtained as the inverse of the negative of the Hessian matrix evaluated at $\hat{\boldsymbol{\eta}}$.

3.3 Model Assessment

In this section, we present two methods of model assessment that we investigate in the NCRMD data context: the Kolmogorov-Smirnov test based on generalized residuals and the likelihood ratio test applied to nested models. The advantage of the former is its flexibility, e.g., it can be used to compare non-nested models across different classes, while the advantage of the latter is its power (at least asymptotically).

3.3.1 Adjusted Generalized Residuals and the Kolmogorov-Smirnov Test

We use adjusted generalized residuals and the Kolmogorov-Smirnov (K-S) test to assess the goodness-of-fit of the proposed model.

Time-to-Event Models

Let a general event process starting at $u_{i0} = 0$ (in backwards time) have events at $u_{i0} < u_{i1} < \ldots < u_{i,n_i} = 365$, and let $u_{i,n_i} = 365$ be the time at the end of the observation period, as displayed in Figure 3.1. For a rate function $\rho_i(u \mid H(u))$, the generalized residuals are

$$E_{ij} = \int_{U_{i,j-1}}^{U_{ij}} \rho_i(u|H(u))du \qquad i = 1, \ldots, m,\ j = 1, \ldots, n_i,$$

and the realizations of the generalized residuals e_{ij} are obtained by replacing U_{ij} with the observed u_{ij} and $\rho_i(s|H(s))$ with the maximum likelihood estimate $\hat{\rho}_i(s|H(s))$. Because the event times u_{i,n_i} are censoring times (not contact times), adjusted residuals (Lawless, 2011) are used instead. Specifically, let δ_{ij} be the indicator of censoring, $\delta_{ij} = \mathbf{1}(t_{ij}$ is observed).

Then the adjusted realizations of the residuals are

$$e_{ij}^{\text{adj}} = \delta_{ij} e_{ij} + (1 - \delta_{ij}) E(E_{ij} \mid E_{ij} \geq e_{ij}) \qquad i = 1, \ldots, m,\ j = 1, \ldots, n_i + 1,$$

where $E(E_{ij} \mid E_{ij} \geq e_{ij})$ denotes the conditional expected value of E_{ij}. If the proposed model is correct, the distribution of adjusted residuals is the standard exponential distribution. By the memoryless property, $E(E_{ij} \mid E_{ij} \geq e_{ij}) = e_{ij} + E(E_{ij})$. Since E_{ij} follows a standard exponential distribution, $E(E_{ij}) = 1$, and the adjusted residuals are

$$e_{ij}^{\text{adj}} = \delta_{ij} e_{ij} + (1 - \delta_{ij})(e_{ij} + 1) \qquad i = 1, \ldots, m,\ j = 1, \ldots, n_i + 1,$$

Time-Between-Events Models

Let a general event process starting at $t_{i0} = 0$ (in forwards time) have events at $t_{i0} < t_{i1} < \ldots < t_{i,n_i} = 365$. For a hazard $h_{ij}(t|H(t))$ and corresponding survival function $S(W_{ij})$, the generalized residuals are

$$\begin{aligned} E_{ij} &= \int_{T_{i,j-1}}^{T_{ij}} h_{ij}(s|H(s))ds \\ &= \int_0^{W_{ij}} -\frac{S'\left(s + u_{i,j-1}\right)}{S\left(s + u_{i,j-1}\right)} ds \\ &= -\log S(W_{ij}), \end{aligned}$$

where $i = 1, \ldots, m$ and $j = 1, \ldots, n_i$. Note that the event times t_{i0} are censoring times (not contact times).

Since

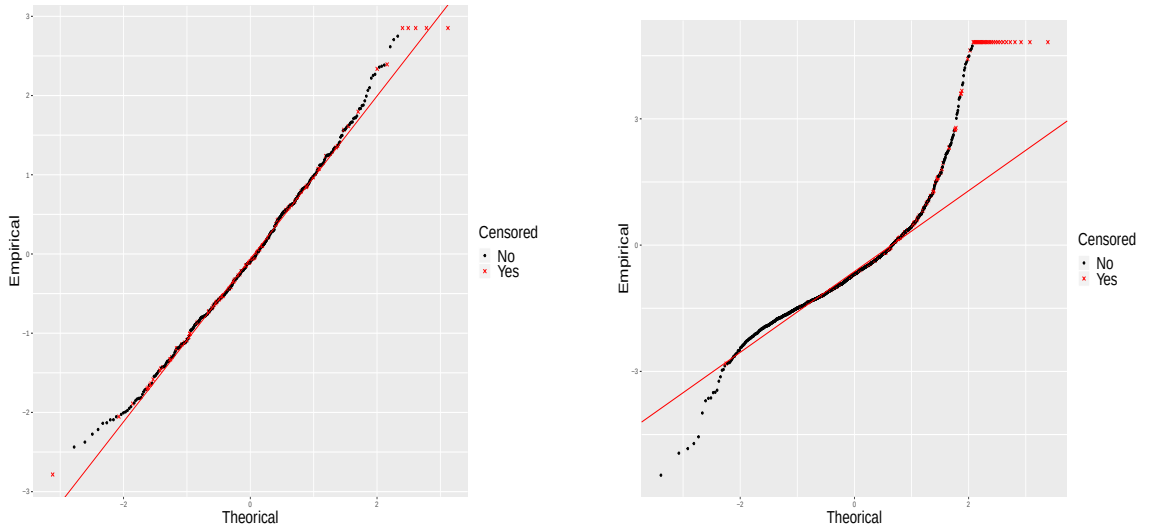
$$\begin{aligned} \Pr(-\log S(W_{ij}) \leq x) &= \Pr(S(W_{ij}) \geq e^{-x}) \\ &= \Pr\left(W_{ij} \leq S^{-1}(e^{-x})\right) \\ &= 1 - S(S^{-1}(e^{-x})) \\ &= 1 - e^{-x}, \end{aligned}$$

Again, because the event times t_{i0} are censoring times, adjusted residuals are also used for time-between-events models. As explained above, The adjusted residuals have a standard exponential distribution.

A quantile-quantile (q-q) plot is used to compare the $\hat{e}_{ij}^{\text{adj}}$ with quantiles from the standard exponential distribution, and hence to detect model misspecification.

An example of residual plots is given in Figure 3.2. The left and right panels represent residuals $\hat{e}_{ij}^{\text{adj}}$ that were computed by fitting the correct and incorrect models, respectively, to simulated data. The residuals based on the correct model lie roughly on the 45° line,

particularly those corresponding to the non-censored observations. Some subjects have only a terminal event (i.e., a single, censored event time of length 365). These residuals are the largest, and correspond to the tied residuals in the upper right corners of the plots. The residuals based on the wrong model form a non-linear shape and deviate away from the 45° line as the size of the residual increases.



(a) Generalized residual plot where the correct model is fit

(b) Generalized residual plot where the wrong model is fit

Figure 3.2: Generalized residual plots (simulated data)

We can also conduct the Kolmogorov-Smirnov (K-S) goodness-of-fit test (Chakravarti *et al.*, 1967) to assess the proximity of the distribution of the adjusted generalized residuals to the standard exponential distribution. The K-S test compares an empirical distribution with a specified parametric distribution. If the test statistic is large enough, then we can reject the null hypothesis that the adjusted generalized residuals follow a standard exponential distribution. Rejecting the null in this case implies poor fit of our chosen model. Referring again to Figure 3.2, the K-S statistics are $D = 0.045$ ($p-$value=0.2177) and $D = 0.286$ ($p-$value≈ 0) for the residuals depicted in the left and right panels, respectively.

3.3.2 Likelihood Ratio Tests

In addition to the K-S test based on the adjusted generalized residuals, a formal model check can be performed via a likelihood ratio test (LRT). Specifically, by nesting our proposed model within a class of more complex models, a LRT can be used to test whether the data provide evidence against the fit of the simpler model (for a given class of alternatives). One possible exploratory approach is to conduct a series of LRTs, first specifying the model under the null hypothesis as the simplest model under consideration, and then extending the model by one parameter at a time, each time conducting a LRT to test for evidence against

the smaller model. For instance, we could start with $H_0 : \rho_0 = e^{\gamma}$ vs. $H_A : \rho_1 = e^{\gamma+\alpha t}$ to test if there is time effect term in the rate function. Similarly, a LRT of $H_0 : \rho_0 = e^{\gamma}$ vs. $H_A : \rho_1 = e^{\gamma+\beta x_i}$ can be used to examine if the predictor variable x_i has an effect on the rate. As a final example, a LRT of $H_0 : \theta = 0$ vs. $H_A : \theta > 0$ could be used to test for the existence of random effects.

For our class of time-to-event models, a Poisson process results when the variance of the random effects, θ, is zero. Note that when we test $H_0 : \theta = 0$, the value of θ under the null is on the boundary of the parameter space. Under the null hypothesis, the limiting distribution of the test statistic is a fifty-fifty mixture distribution of a χ^2_0 distribution (a point mass at zero) and a χ^2_1 distribution (Miller, 1977).

Chapter 4

Simulation Studies

In this section, we discuss the simulation studies that we conducted to evaluate the behaviour of the MLEs of the parameters of our models of primary interest, α and β. The parameter α determines the increase in the event rate in our time-to-event models and β determines the increase in the hazard (conditional on previous gap time) in our time-between-events models. Both α and β can be considered measures of escalation rate. We also present simulation studies for studying the performance of the goodness-of-fit tests outlined in Sections 3.3.1 and 3.3.2.

4.1 Overview

Our most complex time-to-event model is the random effects model. We use $\rho_{ij}(u) = v_i e^{\gamma+\alpha(\tau-u)}$ as a basis for studying $\hat{\alpha}$. In terms of time-between-events models, we consider the time-dependent Weibull hazard model conditional on the previous gap time (as specified in Section 3.2.3) as a basis for studying $\hat{\beta}$. We evaluate the behaviour of $\hat{\alpha}$ and $\hat{\beta}$ for various values of the parameters in the models. Specifically, we study the biases of the estimators, the standard deviations of the estimators, the standard errors of estimators, and the coverage probability of the associated Wald CIs.

In terms of goodness-of-fit methods, we assess the ability of the LRTs and K-S test to detect departures from the true model. To explore the validity and sensitivity of the tests, we estimate their size and power under various conditions. The size of a test is the probability of falsely rejecting the null hypothesis when it is true, i.e., the probability of making a Type I error. True models are always used to fit the simulated data when assessing the size. The power of a test is the probability of correctly rejecting the null hypothesis when it is false. To assess the size of LRTs, we simulate data from the complex models described above with the escalation rate parameters set to 0. We then compute the proportion of times that the null hypothesis of no escalation is rejected at a significance level of 5%. To assess the power of LRTs, we simulate data using the complex models with true values of the escalation rate parameters other than 0, and then fit the same models but with the escalation rate

parameters set to 0. We record the proportion of times that the null hypothesis is rejected at a significance level of 5%. Because the K-S test can be used to test the fit of non-nested models, separate studies were conducted. We assess the performance of the K-S test for both classes of models. To assess the size of K-S test, we simulate two data sets: one from the complex models with no random effects for time-to-event models and one from the complex models for time-between-events models. For each simulated data set, we fit the true model and then compute the proportion of times that the null hypothesis of standard exponentially distributed generalized residuals is rejected at a significance level of 5%. To assess the power of the K-S test, two different wrong models are also used to fit, one within the same class and one from the other class. For each class of the models, we simulate data using the complex model within that class with true values of the escalation rate parameter other than 0, and then fit two models: the same model with the escalation rate parameter set to 0 and the complex model from other class with the the other escalation rate parameter also set to 0.

We conduct the simulation study in the spirit of a designed experiment. For the study of $\hat{\alpha}$, we consider three factors: α, m (the number of subjects), and θ (the variance of the distribution of the random effects). For each factor, we consider three levels: low, medium, and high. We also consider the special case of $\theta = 0$ (no random effect). The chosen levels were inspired by parameter estimates computed using real data. The values of the factors are listed in Table 4.1.

	α	m	θ
special case			0.0
low	0.0005	40	0.1
medium	0.0050	90	0.5
high	0.0100	150	1.0

Table 4.1: Factor levels for the study of $\hat{\alpha}$

For the study of $\hat{\beta}$, we consider three factors: β, a, and b, the parameters of the Weibull distribution. The values of the factors are listed in Table 4.2.

	β	a	b
low	-0.0050	0.3	10
medium	-0.0010	0.6	30
high	-0.0003	1.0	50

Table 4.2: Factor levels for the study of $\hat{\beta}$

In each study, simulations are conducted using all combinations of levels, i.e., we use a full factorial design. The performance of the LRT of $H_0 : \beta = 0$ is assessed using the same

design. Because γ is not a parameter of interest in this project, we fix its value at -5.12 in all time-to-event models used for generating data.

For the study of the size of the K-S test, we consider two factors for time-to-event models (α and m) and three factors for time-between-events models (β, a, and b). Each factor has the three levels listed in Tables 4.1 and 4.2. Simulations are conducted using all combinations of levels. For the study of the power of the K-S test, we simulate data sets using our time-to-event model with random effects and the time-dependent hazard model (with parameters chosen according to Tables 4.1 and 4.2).

4.2 Simulation Settings

We generate data using the steps explained in Sections 4.2.1 and 4.2.2 for the two classes of models. We repeat this process n =1,000 times to generate 1,000 datasets ("replicates") for every combination of factor levels ("run"). For each replicate, we obtain the following five quantities associated with the parameter of interest (α or β):

- The MLE
- The standard error of the estimate based on the information matrix or Hessian matrix
- The 95% Wald confidence interval
- The p-value from the LRT of $H_0 : \alpha = 0$ or $\beta = 0$

After completing this process for all n replicates, we then obtain the following summaries for the run:

- Sample bias of the estimator, computed by taking the difference between the average of the n estimates and the true parameter value
- Sample standard deviation of the n parameter estimates (which we call the empirical standard deviation)
- The proportion of LRTs where the null hypothesis was rejected at a significance level of 5%

For the K-S test, we also use 1,000 replicates for every combination of parameter levels for each chosen model ("run"). For each replicate, we use the following procedure to compute the size:

- Use a time-to-event model with $\rho(u) = e^{\gamma+\alpha(\tau-u)}$ as the true model. Fit the simulated data using the same model. Calculate the residuals using the estimates and record the p-value from the K-S test.

- Use a time-between-events model with $h_{ij}(w) = \left(\frac{a}{b}\right)\left(\frac{w_{ij}}{b}\right)^{a-1} e^{\beta w_{i,j-1}}$ as the true model. Fit the simulated data using the same model. Calculate the residuals using the estimates and record the p-value from the K-S test.

We use the proportion of replicates within each run for which the null hypothesis was rejected at level 5% as an estimate of the size of the test for the given setting. Similarly, for each replicate, we use the following settings to assess the power:

- Generate data using the time-to-event model with rate function conditional on random effects $\rho_i(u) = v_i e^{\gamma+\alpha(\tau-u)}$ (as described in 3.1.3). Fit the model with rate function with no random effects $\rho(u) = e^{\gamma+\alpha(\tau-u)}$ and an event-independent hazard function $h(w) = \left(\frac{a}{b}\right)\left(\frac{w}{b}\right)^{a-1}$. Calculate the residuals using the estimates from each fitted model and record the p-value from the K-S test.

- Generate data using the event-dependent Weibull hazard function $h_{ij}(w) = \left(\frac{a}{b}\right)\left(\frac{w_{ij}}{b}\right)^{a-1} e^{\beta w_{i,j-1}}$ (as described in 3.2.3). Fit the simulated data using the rate function without random effect $\rho(u) = e^{\gamma+\alpha(\tau-u)}$ and an event-independent hazard function $h(w) = \left(\frac{a}{b}\right)\left(\frac{w}{b}\right)^{a-1}$. Calculate the residuals using the estimates from each fitted model and record the p-value from the K-S test.

For each run, we estimate power as the proportion of K-S tests (across all replicates within the run) where the null hypothesis was rejected at a significance level of 5%

4.2.1 Generating Data Using Time-to-Events Models

We generate event times backwards in time from the terminal event time until the length of the observation window reached at least 365 days. We then truncate the last gap time. Specifically, for each subject, we took the following steps:

1. For the models with random effects, we first simulate each subject's random effect, v_i, from a gamma distribution with shape k and scale θ, where we fix the mean to be 1 (and hence the variance to be θ).

2. Gap times are simulated using the inversion method. For the j^{th} event of individual i, simulate $Y_{ij} \sim U[0,1]$. Compute $W_{ij} = 1 - S_{ij}^{-1}(Y_{ij})$, where $S_{ij}(W_{ij}) = \exp\left(-\Lambda_i(w_j)\right)$ is the survival function associated with the j^{th} gap time of individual i. The backwards event time is then $U_{ij} = U_{i,j-1} + W_{ij}$.

3. Repeat Step 2 until $U_{ij} > 365$. Set n_i equal to the current value of j. The n_i^{th} gap and event times are right-censored by setting $W_{in_i} = 365 - \sum_{j=1}^{n_i-1} W_{i,j}$ and $U_{i,n_i} = \sum_{j=1}^{n_i} W_{i,j}$.

4.2.2 Generating Data Using Time-Between-Events Models

To generate data from time-between-events models (including a censored first gap time), we first simulate complete gap times. We then truncate the first (complete) gap time to mimic the censoring observed in the real data. Specifically, for each subject, we took the following steps:

1. Simulate a series of event-dependent gap times $(w_{i1}, w_{i2}, \ldots)$ forwards in time using a parametric gap time distribution, stopping when the sum of the gap times is greater than 365 days. Denote the number of events generated by n_i.

2. Define the last event time as $t_{in_i} = 365$ days. Working backwards in time, calculate the j^{th} event time as $t_{ij} = 365 - \sum_{k=j+1}^{n_i} w_{ik}$, $j = n_i - 1, n_i - 2, \ldots, 1$.

4.3 Results of Simulation Studies

4.3.1 Time-to-Events Models

Results concerning $\hat{\alpha}$ are presented in Table 4.3. Boxplots of the differences between estimated values and true values are also presented in Figure 4.1.

		m = 40				m=90				m=150			
	α	Average Bias	Empirical SD	Average SE	Coverage rate	Average Bias	Empirical SD	Average SE	Coverage rate	Average Bias	Empirical SD	Average SE	Coverage rate
	0.0005	0.00510	0.00095	0.00090	0.00000	0.00506	0.00062	0.00060	0.00000	0.00506	0.00048	0.00046	0.00000
$\theta = 0$	0.005	0.00115	0.00064	0.00061	0.54000	0.00115	0.00043	0.00041	0.19500	0.00118	0.00032	0.00032	0.03300
	0.01	−0.00045	0.00042	0.00040	0.76100	−0.00046	0.00027	0.00026	0.59400	−0.00046	0.00021	0.00020	0.38900
	0.0005	0.00506	0.00095	0.00090	0.00000	0.00509	0.00062	0.00060	0.00000	0.00506	0.00050	0.00046	0.00000
$\theta = 0.1$	0.005	0.00116	0.00063	0.00061	0.52700	0.00113	0.00041	0.00041	0.20300	0.00113	0.00031	0.00032	0.04500
	0.01	−0.00048	0.00043	0.00040	0.74100	−0.00049	0.00029	0.00026	0.50200	−0.00050	0.00021	0.00026	0.33200
	0.0005	0.00510	0.00107	0.00091	0.00000	0.00508	0.00069	0.00060	0.00000	0.00505	0.00053	0.00046	0.00000
$\theta = 0.5$	0.005	0.00107	0.00060	0.00061	0.58500	0.00105	0.00043	0.00041	0.26000	0.00106	0.00033	0.00031	0.08200
	0.01	−0.00058	0.00044	0.00039	0.65000	−0.00060	0.00028	0.00026	0.37600	−0.00060	0.00022	0.00020	0.18400
	0.0005	0.00509	0.00116	0.00090	0.00000	0.00507	0.00074	0.00060	0.00000	0.00505	0.00060	0.00046	0.00000
$\theta = 1$	0.005	0.00100	0.00060	0.00060	0.62900	0.00097	0.00041	0.00040	0.31200	0.00098	0.00031	0.00031	0.10700
	0.01	−0.00069	0.00045	0.00039	0.57600	−0.00068	0.00030	0.00026	0.28300	−0.00068	0.00023	0.00020	0.10300

Table 4.3: Estimated properties of $\hat{\alpha}$ when the true model is the time-to-event model with conditional rate function $\rho_i(u) = v_i e^{\gamma+\alpha(\tau-u)}$

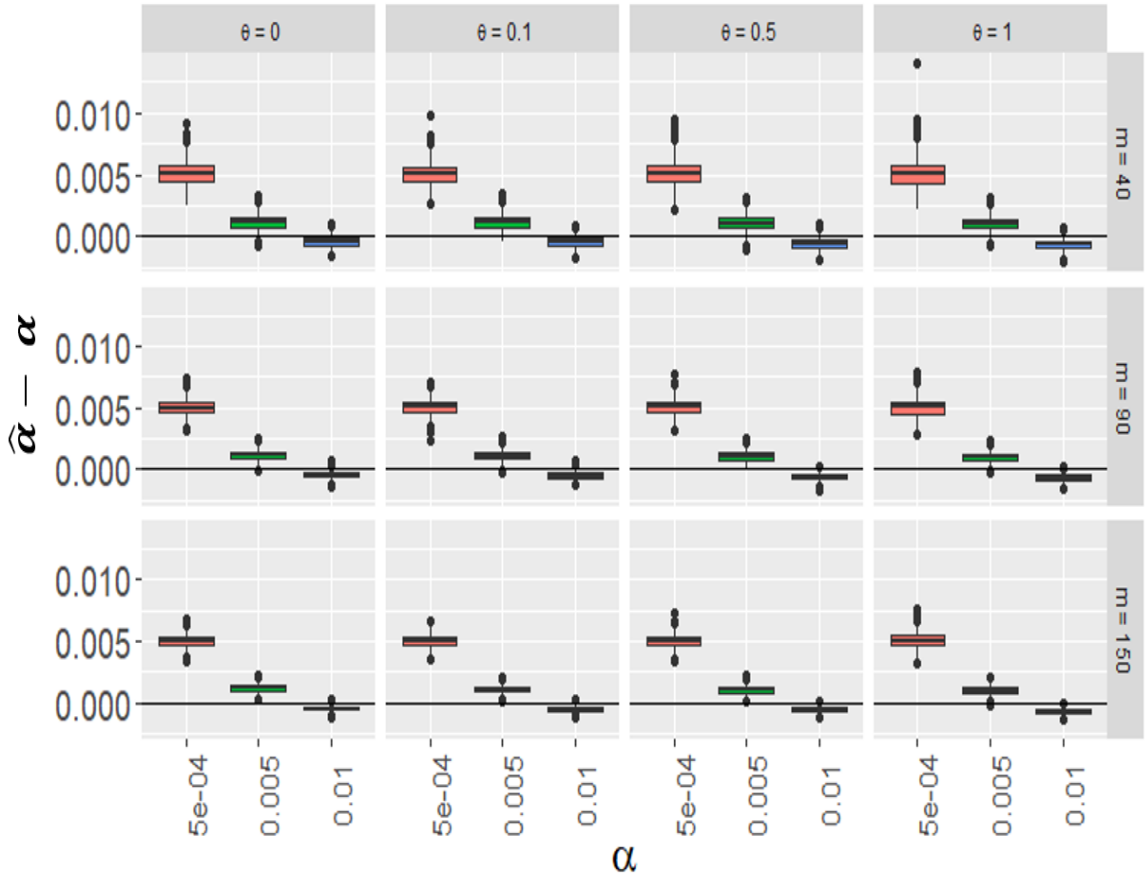


Figure 4.1: Differences between $\hat{\alpha}$ and α using 1,000 replicates per run

The higher the level of α, the less biased the estimator $\hat{\alpha}$. When $\alpha = 0.01$, the estimator is biased low. As evident in Figure 4.2, the average standard error and empirical standard deviation are close when $\theta < 1$ or $\alpha > 0.0005$, indicating that the information matrix provides reasonably accurate estimates of the standard deviation of $\hat{\alpha}$ in these cases. In other cases, particularly when m is small, the standard error appears to be biased low.

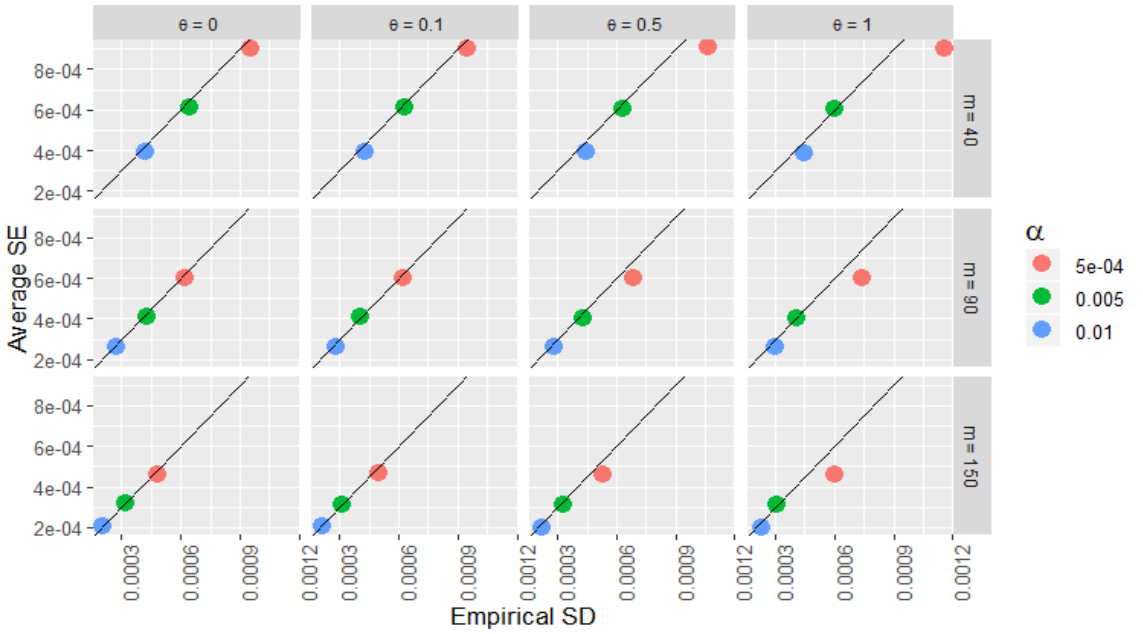


Figure 4.2: Empirical SD against average SE of $\hat{\alpha}$ using 1,000 replicates per run

Figure 4.3 shows the estimated coverage rate of the 95% Wald CI. Surprisingly, we find that the coverage rate decreases as m increases. A likely explanation is that, over the levels of m we considered, the bias remains approximately constant while the standard error decreases, resulting in a lower coverage rate.

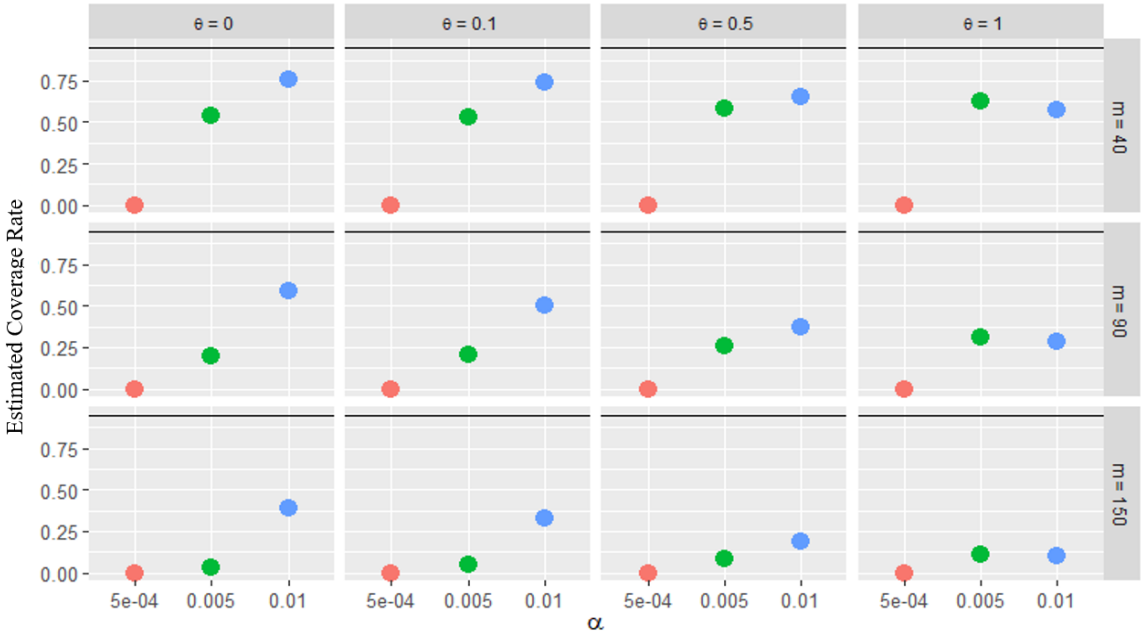


Figure 4.3: Estimated coverage rate of the 95% Wald CI for $\hat{\alpha}$. Grey horizontal lines represent the nominal 95% coverage level.

With respect to the performance of the LRT of $H_0 : \alpha = 0$, when there is no random effect (i.e., $\theta = 0$), for all values of m, the size of LRT is 0, meaning that the test always fails to reject the null hypothesis when it is true. When $\theta = 0.1$, the estimated size is 5%. But, when θ is far from 0, the estimated size is 1. In terms of power, in all cases, the LRT always correctly rejects the null hypothesis (i.e., the estimated rejection rates are all 1).

	$\alpha = 0$				$\alpha \neq 0$			
	θ=0	θ=0.1	θ=0.5	θ=1	θ=0	θ=0.1	θ=0.5	θ=1
m=40	0	0.051	1.000	1.000	1.000	1.000	1.000	1.000
m=90	0	1.000	1.000	1.000	1.000	1.000	1.000	1.000
m=150	0	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 4.4: Estimated rejection rate of the LRT of $H_0 : \alpha = 0$

4.3.2 Time-Between-Events Models

We choose the case of $(a, b) = (0.6, 50)$ as an example to present here. The results for other values of the Weibull parameters are similar. Table 4.5 displays the results for the time-dependent Weibull hazard function and Figure 4.4 illustrates the differences between $\hat{\beta}$ and β. Consistent with our results from time-to-event models, higher values of β lead to less bias. And, $\hat{\beta}$ is consistently biased low.

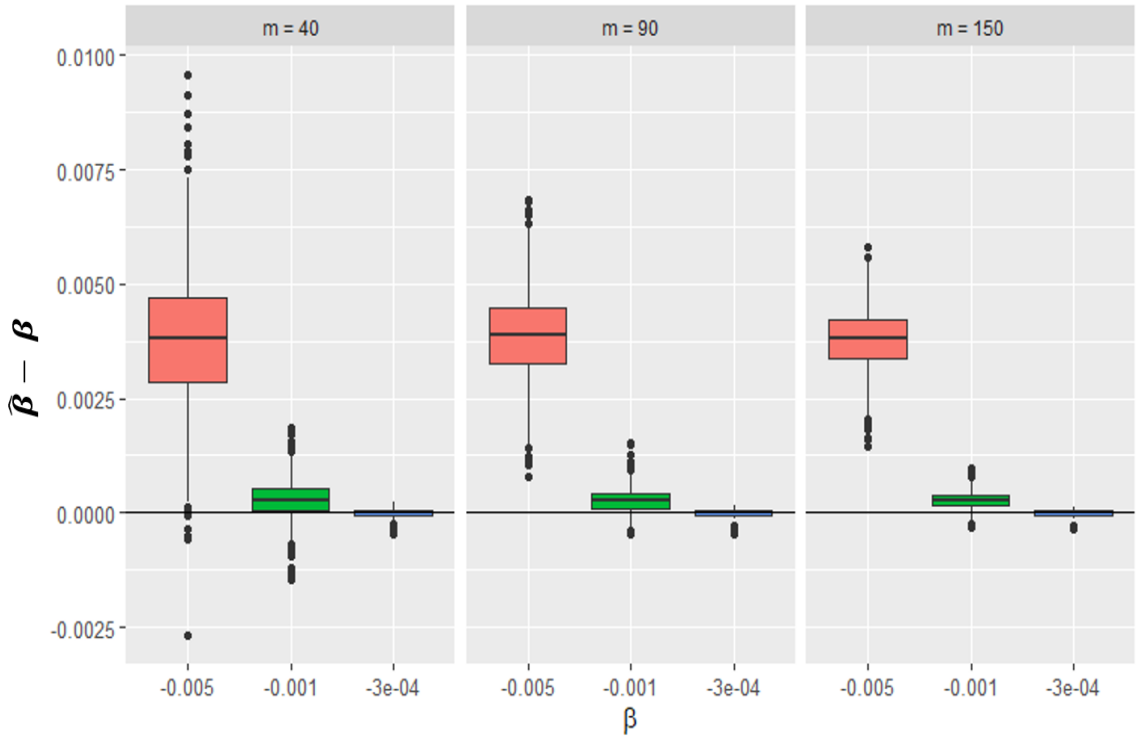


Figure 4.4: Differences between $\hat{\beta}$ and β using 1,000 replicates per run

	m = 40				m=90				m=150			
β	Bias	Sample Std Dev	Std Err	Coverage rate	Bias	Sample Std Dev	Std Err	Coverage rate	Bias	Sample Std Dev	Std Err	Coverage rate
-0.005	0.00383	0.00136	0.00176	0.37876	0.00388	0.00089	0.00115	0.06613	0.00381	0.00070	0.00089	0.01002
-0.001	0.00030	0.00042	0.00143	0.99800	0.00026	0.00025	0.00093	1.00000	0.00026	0.00021	0.00072	1.00000
-0.0003	-1.83e-05	0.00013	0.00138	1.00000	-2.38e-06	0.00008	0.00089	1.00000	2.19e-06	0.00006	0.00069	1.00000

Table 4.5: Estimated properties of $\hat{\beta}$ when the true model is the time-between-events model with time-dependent Weibull hazard function $h_{ij}(w) = \left(\frac{a}{b}\right)\left(\frac{w_{ij}}{b}\right)^{a-1} e^{\beta w_{i,j-1}}$

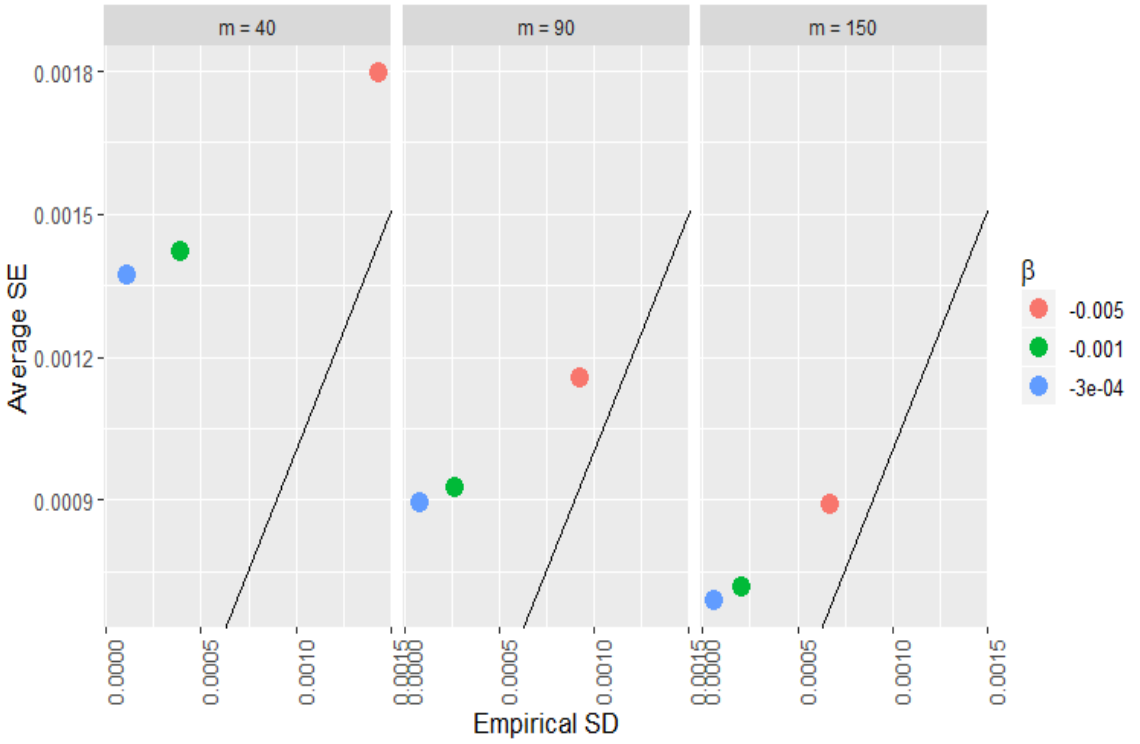


Figure 4.5: Empirical SD against average SE of $\hat{\beta}$ using 1,000 replicates per run

Both the average standard error and the empirical standard deviation are smaller for higher values of β, possibly because as β increases, the number of events tends to increase. As exhibited in Figure 4.5, the average standard error seems to be consistently greater than the empirical standard deviation. This difference indicates that the Hessian provided by the minimization routine that we used fails lead to reliably accurate standard errors. However, inference based on these standard errors is likely to be conservative, and, as m increases, the average standard error looks closer to empirical standard deviation.

Figure 4.6 shows the estimated coverage rate of the 95% Wald CI for β. When β is small, the estimated coverage rate is at most 45%, whereas when β is large, the estimated coverage rate is 1.

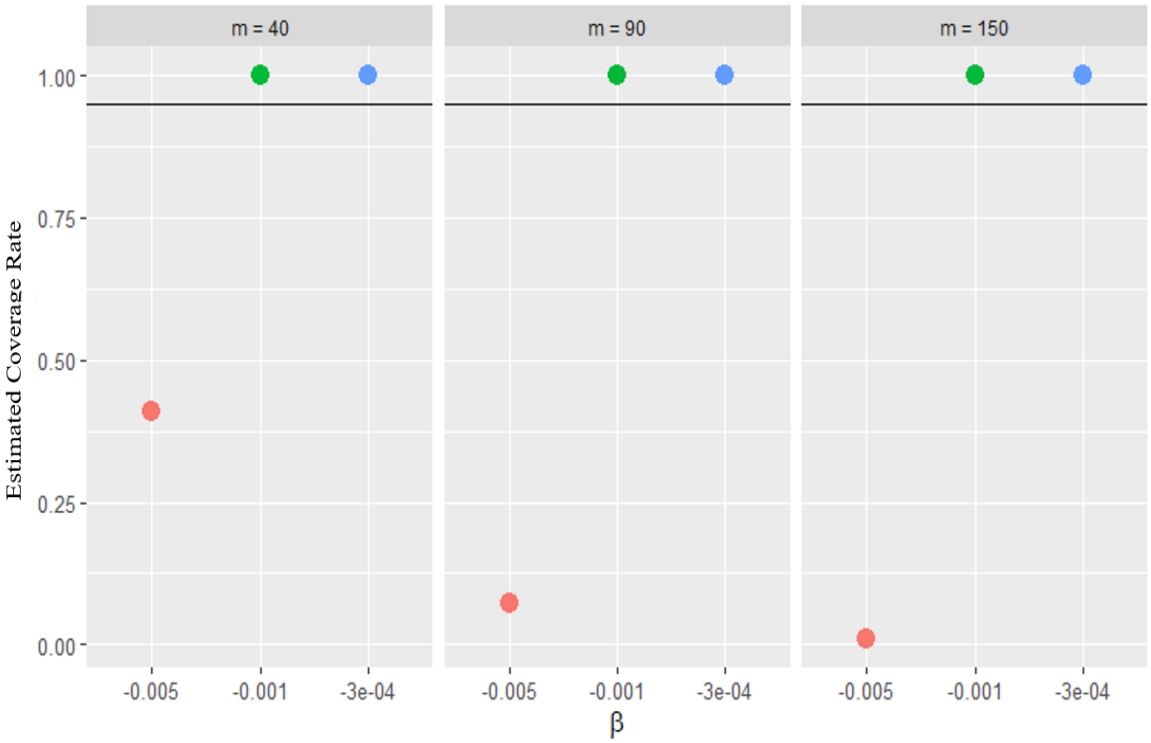


Figure 4.6: Estimated coverage rate of the 95% Wald CI for $\hat{\beta}$. The black horizontal line represents the nominal 95% coverage level.

Figure 4.7 shows the estimated size and power of the LRT of $H_0 : \beta = 0$. The estimated size of the test is less than 5%, suggested that the test is valid. In contrast, the power of the LRT is at most 26%, meaning that only 25% of time the LRT correctly rejects the null hypothesis when it is not true.

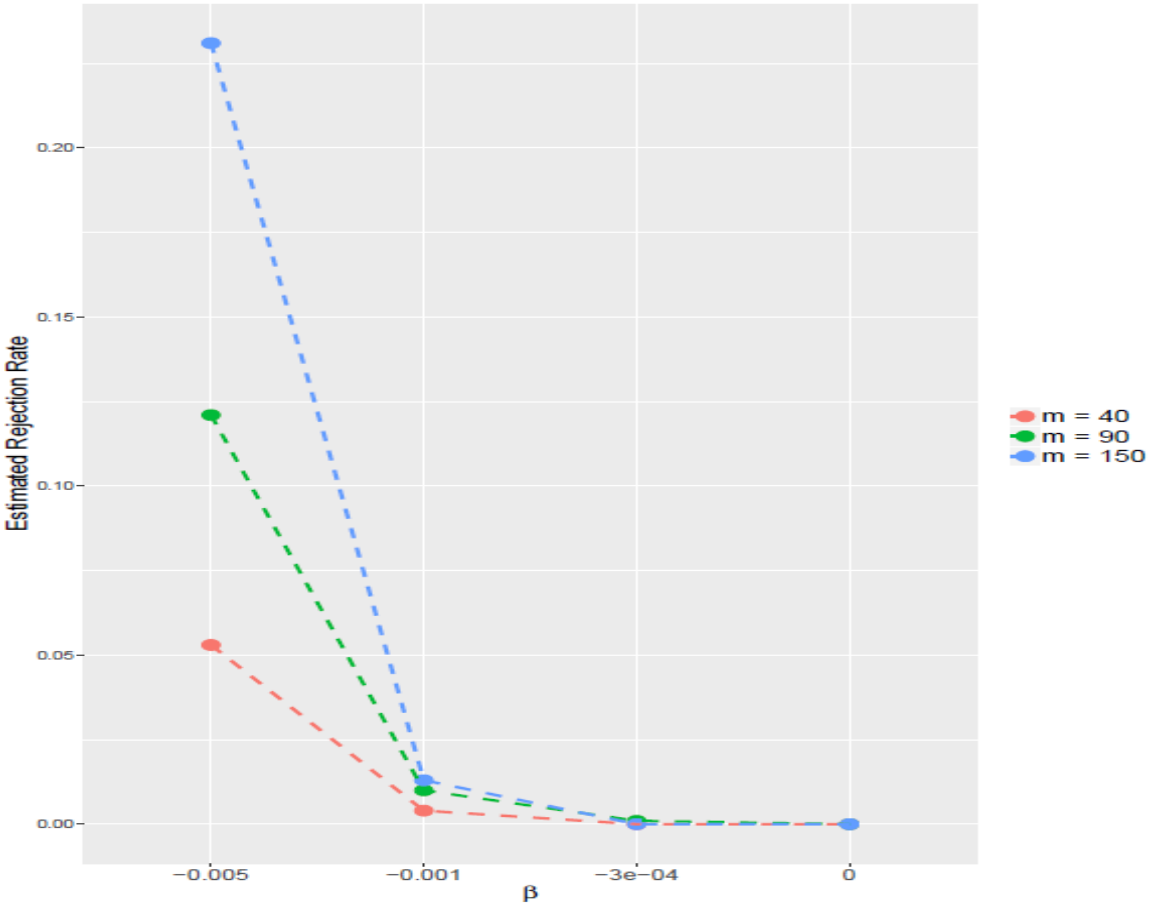


Figure 4.7: Estimated rejection rate of the LRT of $H_0 : \beta = 0$ using 1,000 replicates per run

4.3.3 Kolmogorov-Smirnov Test

We assess the performance of the K-S test of the distribution of the generalized residuals. Our goal is to evaluate its usefulness as a goodness-of-fit test, focusing on the case where the true and proposed models are not nested (i.e., the case where the conditions of the LRT don't hold). As described in Section 4.2, two settings are considered for evaluating the size and power of the K-S test.

Figure 4.8 shows the estimated size of the K-S test when a time-to-event model with $\rho(u) = e^{\gamma+\alpha(\tau-u)}$ is used to simulate the data (and the same model is used to fit the data). Figure 4.9 shows the estimated size of the K-S test when the dependent gap time model is used to simulate and fit the data. We show results only for the gap time model with Weibull parameters $(a, b) = (0.6, 50)$; results for other values of (a, b) are similar. Under both settings, the estimated size of K-S test is often higher than 5%, indicating that this test is not valid.

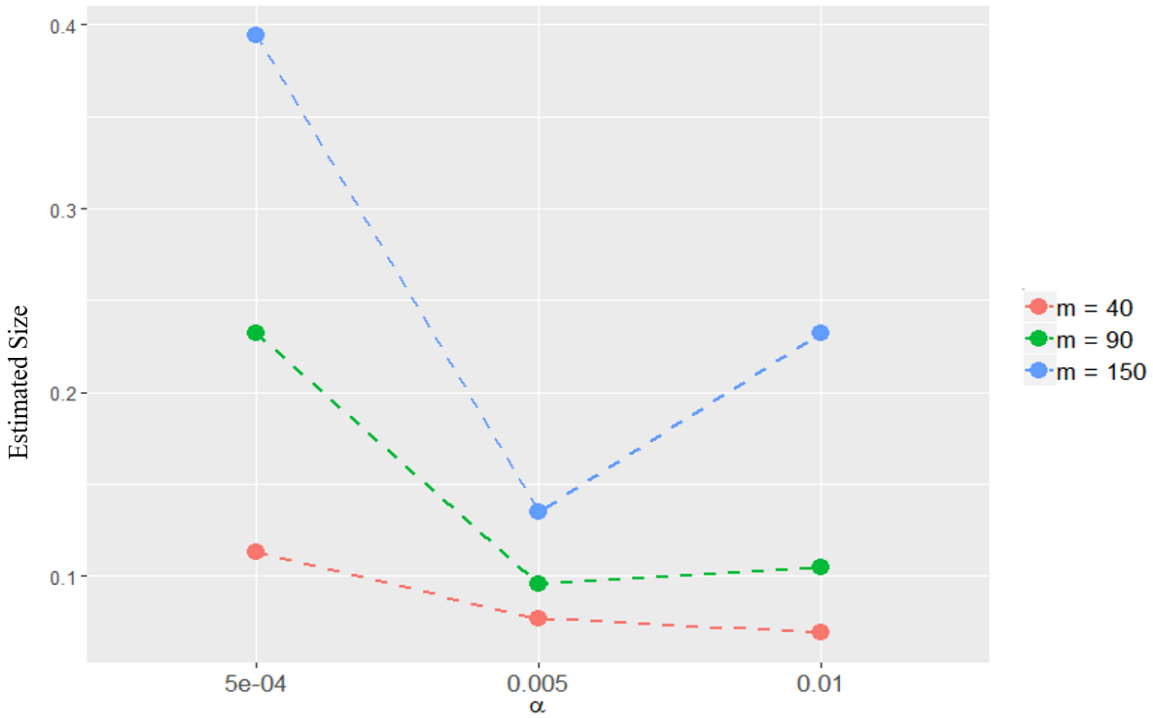


Figure 4.8: Estimated size of the K-S test when a time-to-event model with a random rate function is used to simulate data

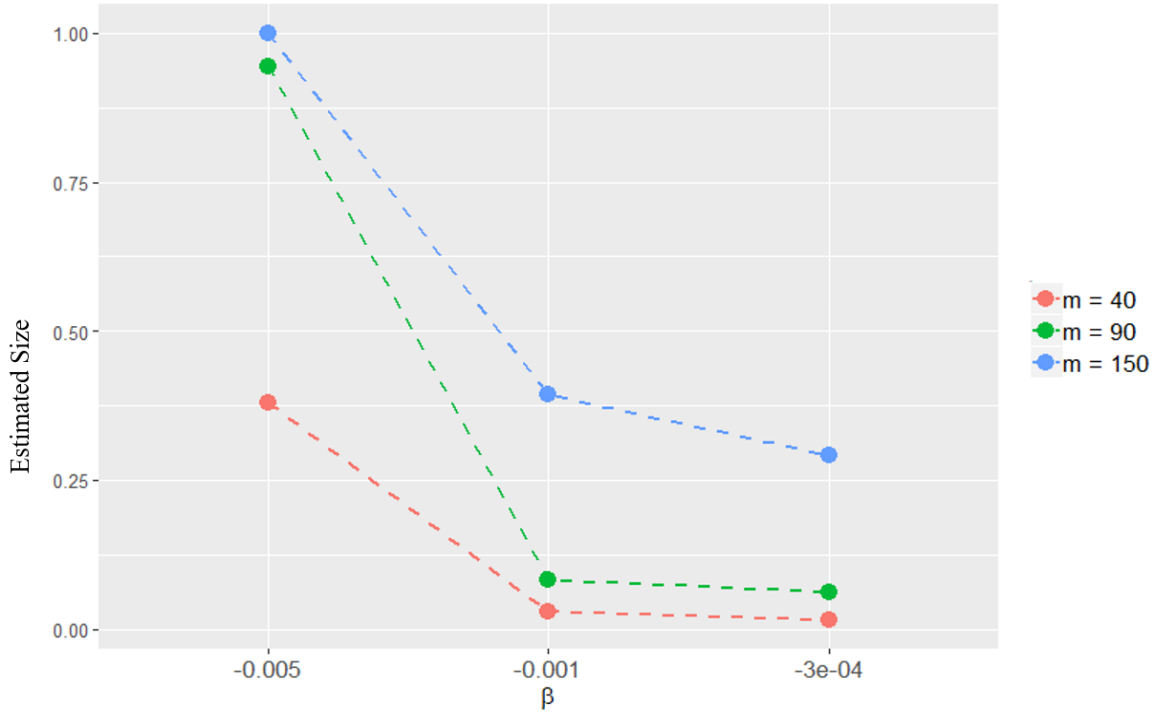


Figure 4.9: Estimated size of the K-S test when a time-between-events model with event-dependent hazard function is used to simulate data

In terms of power, Figures 4.10 and Figure 4.11 depict the estimated power when the time-to-event model with random rate function $\rho_i(u) = v_i e^{\gamma+\alpha(\tau-u)}$ and the time-between-events model with event-dependent hazard function, respectively, are used to simulate the data. In both cases, two incorrect models (a time-to-event model with rate function $\rho(u) = e^{\gamma+\alpha(\tau-u)}$ and a time-between-events model with an event-independent gap time hazard function) are used to fit the data. Across all the cases, when m is small, the power is relatively weak, but may be satisfactory when $m \geq 150$. The K-S test appears to be more powerful in detecting misspecified models from a different class than within the same class.

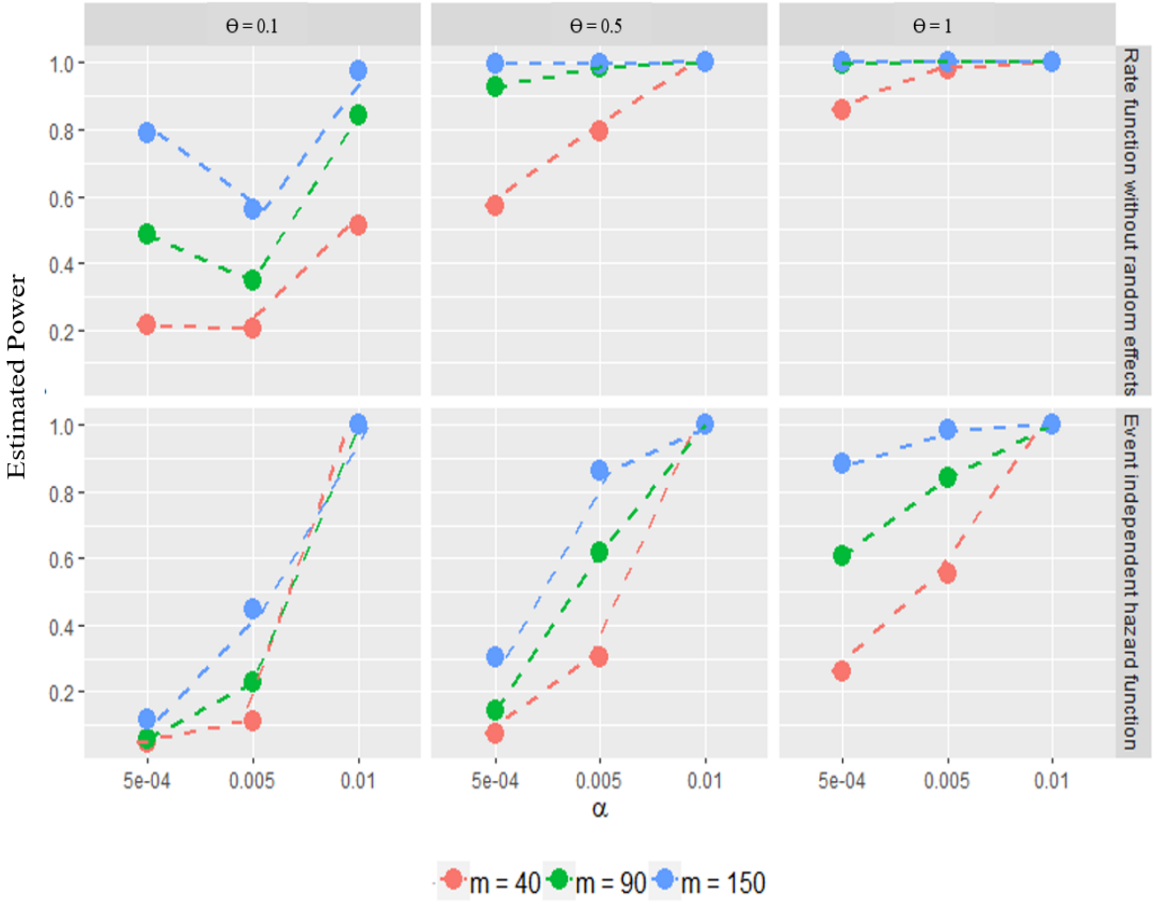


Figure 4.10: Estimated power of the K-S test when the time-to-event model with random rate function is used to simulate the data

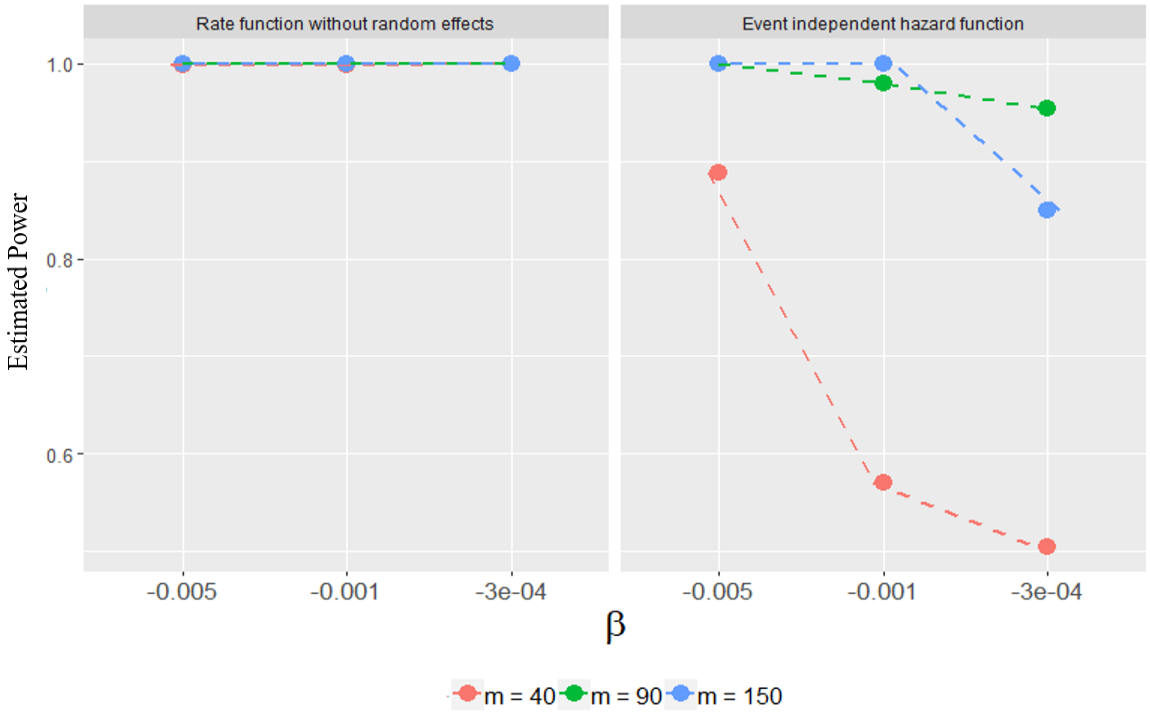


Figure 4.11: Estimated power of the K-S test when the time-between-events model with event-dependent hazard function is used to simulate the data

Chapter 5

Discussion

In this preliminary study, we use a recurrent events framework to model the escalation in event frequency a year prior to an index event for which the offender is found NCRMD. Two particular classes of models are used: time-to-event models and time-between-events models. We adapt these models to account for the unconventional left-censoring in the data.

The simulation results show that our sophisticated time-to-event and time-between-event methods are not practical in this context. The (short) one-year observation time frame negatively impacts the performance of the estimators of escalation rate; their finite sample properties can be quite different from their expected asymptotic properties. The main issue appears to be the sparse event history data from each subject. Both parameters of interest (α and β) are closely related to the event occurrence rate and gap times. The goodness-of-fit approaches that we considered seem similarly sensitive to relatively little information available per individual. Overall, our simulation studies indicate that none of the proposed approaches works reliably in this context.

Our results concerning the invalidity of the K-S test could be due to the bias in the parameter estimators, since the K-S test is very sensitive to the values of the parameter estimates. If the estimators are biased, even if correct model is used to fit the data, the generalized residuals may not lie close to the 45° line. One example is given in Figure 5.1. The data are simulated and fit using a time-to-event model with rate function without random effects and $\alpha = 0.05$. The residuals are calculated using the biased estimators $\hat{\gamma}$ and $\hat{\alpha}$. The K-S test returns $D = 0.086$ (p-value ≈ 0), which could lead to the wrong conclusion that the model is misspecified. Although we don't recommend the K-S test (in light of concerns about its size), we note that it is relatively more powerful in testing the goodness-of-fit when the true and proposed models are from different classes.

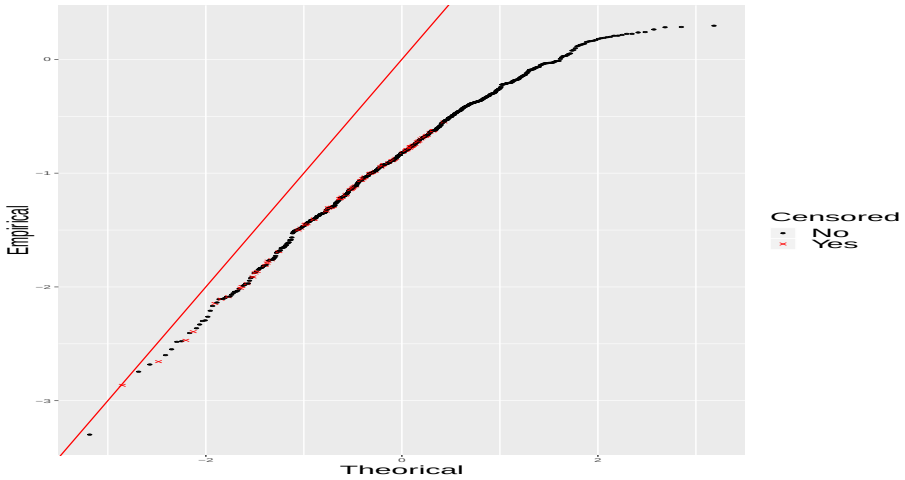


Figure 5.1: Example of an adjusted generalized residual plot when the same model is used to generate and fit the data

In the initial phase of this project, a Poisson GLM was used as a simple preliminary approach to analyzing the data. Using this approach, we were able to detect an increase in mean number of events in the second 6 months of the observation period compared to the first 6 months of the observation period, i.e., we found evidence of escalation. For future research, an expansion of this approach could be considered. Specifically, a quality control chart based on $N_i(t)$ could be explored where signals of increasing mean number of events in a time interval could be monitored.

For the time-between-events models, we assumed that the first gap time, W_{i1}, has the same baseline hazard function as the later gap times, W_{ij}, $j > 1$. This assumption may not be realistic and requires further investigation.

We modelled the data from the CJS and health care systems separately. It is possible that these two data sets are related. For example, perhaps a subject who has had more frequent events with the health care system is likely to have fewer events with the CJS. A joint model for multitype recurrent events could be considered for modelling event times from the two systems simultaneously. In addition, some events are hospitalization or/and incarceration. Unlike the events considered in this project (which we treated as having zero durations), hospitalizations and incarcerations have a positive duration. When subjects are hospitalized or incarcerated, they are unlikely to have more events within the same system. A class of multilevel discrete-time event history models could be used to handle the transitions between two-states (in jail/out of jail or in hospital/out of hospital) (Steele, 2011). We could potentially model the four different (but possibly related) processes for each individual simultaneously.

Moreover, in order for us to use these two classes of models, event times were required to be measured on a continuous scale (i.e., so that two events could not occur simultaneously). However, in practice, the day – but not the exact time – of events are usually recorded. In

other words, the real data could contain multiple events with the same event times. In our simulation studies, we simulated precise times, but, in the future, methods for dealing with tied event times could be considered.

Finally, we notice that about half of the cohort didn't have any prior events with the CJS system, suggesting the possibility of latent classes of subjects within this system. As shown in Eggleston *et al.* (2004), group-based approaches have been developed to discover crime trajectories. Extensions of our methods to include latent classes (which allow subjects from different latent groups to be treated differently) could be another avenue for future research.

Bibliography

Chakravarti, I., Laha, R., and Roy, J. (1967). *Handbook of methods of applied statistics*. Number v. 1 in Wiley series in probability and mathematical statistics. Wiley.

Cook, R. and Lawless, J. (2007). *The Statistical Analysis of Recurrent Events*. Springer: New York.

Criminal Code (1985). R.S.C. c. C-46.

Eggleston, E., Laub, J. H., and Sampson, R. J. (2004). Methodological sensitivities to latent class analysis of long-term criminal trajectories. **20**, 1–26.

Golding, S., Eaves, D., and Kowaz, A. (1989). The assessment, treatment and community outcome of insanity acquittees: Forensic history and response to treatment. *International Journal of Law and Psychiatry*, **12**(2), 149 – 179.

Klein, J. and Moeschberger, M. (2006). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer: New York.

Lawless, J. (2011). *Statistical Models and Methods for Lifetime Data; 2nd ed.* John Wiley Sons.

Livingston, J., Wilson, D., Tien, G., and Bond, L. (2003). A follow-up study of persons found not criminally responsible on account of mental disorder in british columbia. *The Canadian Journal of Psychiatry*, **48**(6), 408–415.

Miladinovic, Z. and Lukassen, J. (2014). Verdicts of not criminally responsible on account of mental disorder in adult criminal courts, 2005/2006-2011/2012.

Miller, J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. **5**(4), 746–762.

Nagin, D. and Land, K. (2006). Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric, mixed poisson model. **31**, 327 – 362.

Pénichoux, J., Moreau, T., and Latouche, A. (2015). Simulating recurrent events that mimic actual data: a review of the literature with emphasis on event-dependence. *ArXiv e-prints*.

Steele, F. (2011). Multilelvel discrete-time event history models with applications to the analysis of recurrent employment transitions. *Australian & New Zealand Journal of Statistics*, **53**(1), 1–20.