UNIVERSITY OF GHANA, LEGON.

UNIVERSITY OF GHANA

A HYBRID TWO-STAGE ESTIMATION METHOD FOR JOINTLY MODELLING LONGITUDINAL AND MULTI-STATE PROCESSES: APPLICATION TO BIOMEDICAL DATA.

BY

ERNEST YEBOAH BOATENG

10637591

A THESIS SUBMITTED TO THE DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE, UNIVERSITY OF GHANA, LEGON, IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN STATISTICS

DECEMBER, 2022

Declaration

I, Ernest Yeboah Boateng, hereby declare that I am the original author of this thesis and no part of this document has been submitted anywhere for the award of a degree.

Ernest Yeboah Boateng	.....	02/December/2023
10637591	Signature	Date
Student		

Certified by:		
Prof. Samuel Iddi	.....	4/12/2023
Lead Supervisor	Signature	Date

Certified by:		
Prof. Louis Asiedu	.....	04-12-2023
Co-Supervisor	Signature	Date

Certified by:		
Prof. K. Doku-Amponsah	.....	04-12-2023
Co-Supervisor	Signature	Date

Dedication

This thesis is dedicated to my two kids, Nana Yaa Konama Boateng and Krysmael Kofi Yeboah Boateng, and their lovely mother, Vida Boateng. Also, to my late father, Mr. Sampson Yeboah Boateng whom I did not grow up to see. Lastly, to my grandmother Obaapanyin Sophia Adwoa Konama and my mother Mrs. Janet Appiagyei for the love and support they have shown me all these years.

Abstract

Joint modelling of longitudinal outcomes and a single survival event has been the focus of current research, although, data collected in practice may be more complicated most especially when multiple event outcomes occur. Consequently, an extension to a joint modelling approach to handle multi-states is needed to portray the interplay between a longitudinal observation and the multi- state outcome. In this thesis, a hybrid two-stage estimation method for jointly modelling longitudinal and survival, and a multi-state process is proposed. The proposed two-step modelling approach uses a Bayesian estimation for the submodel of the longitudinal process and then uses all the posterior predictive estimates from the first stage of the estimation process as inputs in stage two of the process. The proposed estimation method is first applied to time to a single survival outcome and extended to a multi-state process. These models are validated using a simulation study and empirical data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and Parkinson Progressive Marker Initiative (PPMI) studies. The approach was compared to Rizopollous joint, Bayesian joint, and Bayesian two-stage models. When applied to joint model with a single event time, it was evident from the simulation results that the proposed hybrid method performs effectively, both in estimating the fixed-effects and association parameters, especially with samples with larger sizes. When extended to the multi-state model, it was evident from the empirical results that the hybrid model with the last 100 posterior estimates, estimates both the fixed effects and association parameters precisely compared to the hybrid model with all 8000 posterior estimates and the frequentist model. To speed up the computational time and produce more precise estimates, the hybrid model with fewer posterior predictive estimates at the tail end of the converged Bayesian model from the longitudinal sub-model is recommended. Also, for samples with

larger sizes, the hybrid model is a recommended approach as it yields less bias and precise estimates. An area of future work is to extend the hybrid models to multiple longitudinal and survival-type and multi-state outcomes.

Acknowledgments

A mighty praise to God for blessing me with the energy to journey through this academic stage of my life, with Him in our boat, everything is indeed possible.

Again, I convey my sincere and profound appreciation to my supervisors, Prof. Samuel Iddi (lead supervisor), Prof. Louis Asiedu and Prof. Kwabena Doku Amponsah. Mostly, my sincere gratitude goes to my affable, ever helpful and modest lead supervisor, Prof. Samuel Iddi, for his leadership, direction, mentorship and inspiration during the period of my thesis. I will forever be indebted in gratitude to my supervisors, more especially Prof. Samuel Iddi.

Also, I say a big thank you, to the entire faculty of the Department of Statistics and Actuarial Science for the immense support they provided me during my stay in the Department. To my wonderful classmates, I enjoyed every moment of the times we spent together either academically or socially, and I will always miss those memories.

Again, I extend a grateful appreciation to the Institute of Health Research (IHR) at the University of Health and Allied Sciences, Ghana, for awarding me the Faculty Development Grant to aid the conduct of this thesis.

I also thank my siblings Mavis, Elvis, Phyllis, and Eunice for their diverse support.

Lastly, I wish to thank a very good friend, Ms. Blessing Gaitu, for her wonderful support during this stage of my career and also taking time off her busy schedule to type and edit the entire thesis.

CONTENTS

ABBREVIATION

Analysis of Variance (ANOVA)

Alzheimer's Disease (AD)

Alzheimer's Disease Neuro-Imaging Initiative (ADNI)

Bayesian Joint Specification (BJS)

Bayesian Standard Two Stage (BSTS)

Bayesian Regression Models(BRMS)

Confidence Length (CL)

Hybrid Standard Two Stage (HSTS)

Generalized Linear Mixed Models (GLMMs)

Linear Mixed-Effects Models (LMMs)

Maximum Likelihood Estimation (MLE)

Markov Chain Monte Carlo (MCMC)

Multi-Level Models (MLMs)

Multi-State Model(MSM)

Parkinson's Disease (PD)

Parkinson's Progression Markers Initiative (PPMI)

Restricted Maximum Likelihood (ReML)

Rizopollous Joint Approach (RJA)

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Study Background

Although it is common for many prospective research designs to collect both longitudinal and survival-type outcomes, many researchers analyze their data mostly focusing on either the longitudinal outcomes or the survival outcomes. In real-world situations, it is common to encounter longitudinal data, which entails taking multiple observations on the same subjects over a certain period, and survival data, which pertains to time-to-event outcomes occurring simultaneously. For instance, in most research work in the life sciences, patients' information such as temperature is often collected repetitively over a period of time, and also the period to relapse of an illness is of substantial interest. Furthermore, in longitudinal clinical investigations, biomarkers are typically measured repeatedly until a medical event, for instance, onset of a new medical condition or symptom, hospitalization or admission to a medical facility or mortality, takes place.

When the analysis of data concentrates on longitudinal outcomes, the issue of informative dropouts often needs to be addressed since they often arise in longitudinal research work. Moreover, the analysis of longitudinal data poses challenges due to intricate correlation patterns, unevenly spaced observations, data that is missing, and the impact of the combination of time-dependent and fixed explanatory variable. However, it offers researchers the opportunity to explore various facets of a phenomenon, such as changes in its outcome(s) over time in association with related risk factors, the timing of its onset, and the patterns experienced by individuals and groups over time. Furthermore, in contrast to cross-sectional studies, longitudinal studies typically exhibit reduced

variability and enhanced statistical power. When focusing on survival data, it is common for longitudinal studies to include time-changing variables, for example, temperature or blood pressure, because the time to the event may be influenced by the changes in these covariates over time.

Because the time-to-event is potentially linked to longitudinal trajectories, there exists notable relationship between longitudinal data and event time data. Consequently, analyzing these two types of data separately might result in less effective or distorted outcomes. In numerous instances, the primary focus revolves around understanding the connection between the longitudinal phenomenon and the event time process, or using biomarkers to forecast or elucidate the occurrence of the time event. In situations like these, joint models that incorporate correlated repeated measurements and event time procedures are necessary.

Recently, joint models integrating longitudinal and event time data have gained significant consideration and popularity in real-life applications and method development. This is primarily due to their capability to reduce estimation bias, enhance statistical efficiency, provide comprehensive inference, and facilitate outcome predictions (Muthen et al. 2009; Ibrahim et al. 2010). Typically, joint models entail of two submodels: a longitudinal model, such as the linear mixed model (LMM), and a survival model, like the Cox model. These two models are linked by a common term. The integration of longitudinal and survival model through joint models has proven to be a crucial statistical tool (Cho et al. 2017; Wu et al. 2012) and has found utility across diverse fields, including clinical and biomedical studies.

The prevailing form of joint models assumes a connection between the longitudinal and event time procedures via shared random effects, which often leads to a joint likelihood that cannot be analytically evaluated (Luts et al. 2012; Kery, 2010; Rizopoulos et al. 2010). To address this, likelihood-based methods are commonly employed, with Expectation Maximization (EM) algorithms being a popular choice. However, the primary challenge in fitting these models

lies in the computational aspect. The incorporation of shared information inherently increases the time required for the inferential procedure coupled with the complex nature of the models further compounds the computational demands. As the number of random effects increases, the need for numerical integration becomes more pronounced (Huong et al. 2018; Verbeke et al. 2014). Consequently, computationally intensive methods like adaptive Gauss-Hermite quadrature (Jensen and Ritz, 2018; Pinheiro and Bates, 1995) become necessary for evaluating both the cumulative hazard and the overall joint likelihood.

Another frequently employed strategy involves two-stage methods, which offer computational simplicity. Two-stage approaches address the challenges associated with simultaneous inference in joint models (Tsiatis et al. 1995; Mauff et al. 2020). In this method, the longitudinal submodel is initially estimated, and subsequently, the fitted parameters are employed to incorporate the longitudinal trajectory as an internal time-dependent explanatory variable in the event time submodel (Luo et al. 2016; Locascio and Atri, 2011). This strategy is relatively straightforward to implement and enables the utilization of versatile models provided by commonly used packages for analyzing longitudinal and event time data.

In numerous real-life scenarios, a subject might undergo multiple progression events of a phenomenon before the occurrence of the specified event. For instance, a subject with a terminal disease may encounter mild symptoms, followed by a severe state, and eventually, death. Therefore, rather than considering the happening of a lone event, it becomes essential to model the progression of the phenomenon as a multi-state procedure. This approach focuses on analyzing the movements among different states of the phenomenon and examining the impact of longitudinal biomarkers on these transitions to achieve enhanced inference (Tsiatis and Davidian, 2004; Molenberghs and Verbeke, 2006).

In cases where both a longitudinal measurement and multi-state or event time outcomes are observed, there is a need for alternative two-stage estimations.

These estimations should be user-friendly and easily implementable by applied researchers. The purpose of this thesis is to suggest a hybrid two-stage estimation procedure for effectively modelling both repeated measurements and single event, and multiple event processes with a specific application to biomedical data. In this approach, the repeated measurements sub-model employs an LMM, which allows for flexibility through the utilization of constant and varying effects. On the other hand, the event time and multi-state submodels adopt a Proportional Hazard (PH) model.

1.2 Problem Statement

The literature suggests the joint likelihood method and the two-stage approach as the main estimation methods for joint models. The two-stage approach speeds up computational time by fitting two simpler submodels compared to the joint models, which integrate two different modelling frameworks, leading to increased model complexity (Murawska et al. 2012; Newsom, 2015). While joint models provide unbiased estimates, they can require significant computational resources, especially when dealing with large datasets with numerous individuals and covariates, resulting in long computational times or non-converging integrals (Guler, 2017; Viviani et al. 2014). On the other hand, the main drawback of the two-stage procedure is that by disregarding the joint behavior between both procedures, the event time model parameter estimates may be mostly biased (Yuen et al. 2016; Ye et al. 2008).

In many real-life scenarios, individuals may transition through various states of a phenomenon over time. Therefore, it becomes crucial to extend the modelling framework to simultaneously address movements between multiple states of the phenomenon (Iddi et al. 2018). However, the inclusion of movements among several states, rather than focusing on the time to a single event, can introduce additional complexity to the modelling framework.

Addressing the aforementioned challenges requires careful attention to developing

effective estimation methods that improve computational time and the precision of the estimates. With this goal in mind, this study proposes a hybrid two-stage estimation method for jointly fitting longitudinal and single event data, and extend it to multiple event data.

The novelty of this approach lies in its emphasis on fitting the sub-model for the longitudinal process using a Bayesian approach and subsequently employing all the posterior predictive estimates as inputs in the survival process, which is fitted in the frequentist domain. This approach differs much from the existing literature, which typically involves inserting just a single estimate from stage one into stage two. The main advantage of this method over transitional two-stage joint models is that the proposed model can efficiently incorporate the fluctuation of the estimation from stage one into stage two of the modelling framework.

1.3 Study Objectives

1.3.1 General Objective

The core objective of this thesis is to propose a hybrid two-step estimation method that enables the joint modelling of longitudinal and time to a single survival outcome and extend to a multi-state process. This approach is specifically applied to biomedical data and aims to enhance the precision of inference in such contexts.

1.3.2 Specific Objectives

This study specifically seeks;

- To formulate an estimation method for modelling single (univariate) longitudinal and survival processes and evaluate the effectiveness of the proposed approach using a simulation study.

- To formulate an estimation method for single (univariate) longitudinal and multi-state processes and assess the effectiveness of the proposed approach using an empirical study.

- To apply the proposed estimation strategy to address real-life problems using biomedical data, demonstrating its practical utility in real-world scenarios.

1.4 Study Design

This thesis involves the utilization of a longitudinal study design, which involves repeated examinations of the same individuals to observe any changes that may occur over a specific time period. This design allows for the tracking of movements among different states and the identification of factors that may contribute to these changes or influence transitions (Jiang et al. 2021). In this study, the aim is specifically on examining repeated measurements outcomes and multi-state or survival type outcomes. Longitudinal and multi-state study designs are commonly employed in neurodegenerative research, particularly in studies related to Alzheimer's and Parkinson's diseases. These designs enable the evaluation of disease progression over time and across various stages, providing valuable insights into the development and course of these diseases.

This study utilizes secondary data sourced from two longitudinal, observational, and multi-center natural history studies: the Alzheimer's Disease Neuro-Imaging Initiative (ADNI) and the Parkinson's Progression Markers Initiative (PPMI). Details of the ADNI and PPMI datasets have been profiled in chapter three of this thesis.

A typical longitudinal study design that leads to both biomedical longitudinal outcomes and multi-state or survival type outcome mostly consist of multiple stages (Jiang et al. 2021) including:

- Recruitment of Participants: Often participants are recruited from community clinics, and research centers among others.
- Baseline assessment: Participants undergo a comprehensive evaluation, including medical history, physical and neurological exams, cognitive testing, and brain imaging.
- Follow-up assessments: Participants are followed, for example, every 6 months for up to 10 years. Assessments may include cognitive testing, functional evaluations, and imaging studies.
- Transition to different stages: Participants who progress from, say, MCI to AD or from cognitive normal (CN) to AD may be followed in different study paths to allow for examination of disease progression and treatment outcomes.
- Data analysis: Data will be analyzed using statistical models that account for disease progression and time-dependent covariates.

To draw valid inferences from data obtained in a longitudinal study design, it is crucial to verify that the analysis properly aligns with the study design and considers specific characteristics of the data (Garcia et al. 2017; Iddi and Molenberghs, 2012; Diggle et al. 2002). Due to researchers curiosity for an in-depth understanding into data emerging from longitudinal study designs with a single statistical formulation, joint modelling has received enormous attention in the recent past.

1.5 Justification of the Study

When the inferential process for joint models becomes overly time-consuming due to their complex structure or faces challenges in achieving convergence of Markov chains caused by an increased parameter space dimension, alternative approaches to joint modelling become necessary. To overcome this computational challenge

without compromising the precision of parameter estimates, this thesis introduces the hybrid two-stage estimation model as a proposed solution.

The proposed hybrid model capitalizes on the advantages of both single models, leading to enhanced inference and more precise parameter estimates. By combining the strengths of each method, this hybrid model enables researchers to accurately identify trends or alterations in the attributes of the target population, both collectively and for individual cases. When the interest lies in comparing longitudinal trends between outcomes or investigating how the relationship between the outcomes changes over time, the proposed hybrid two-stage modelling approach proves to be a suitable choice. Its ability to capture intricate relationships between different variables makes it a favorable option for addressing such research objectives.

In clinical research, proposing an efficient hybrid joint model for disease classification is a significant step towards addressing the difficulties and refining diagnosis and assessment procedures for diseases. The traditional approach to disease diagnosis relies on medical screening and evaluation of clinical signs, such as the description and analysis of different motor symptoms(Wu et al. 2021). However, this approach may be subject to subjectivity since it relies on evaluating movements that can be subtle and difficult to classify accurately, which may result in possible misclassification (Maier et al. 2015; Muthen et al. 2009). The proposed hybrid estimation procedure, utilizing statistical techniques, offers a solution to these challenges. By employing this approach, a practitioner can recognize important characteristics that might not be conventionally employed in clinical disease diagnosis. This enables the practitioner to consider alternative measures and detect diseases in their early stages or atypical forms, thereby improving the accuracy and dependability of disease classification.

While the existing literature has predominantly focused on studying longitudinal and survival outcomes, this thesis proposes extensions towards multi-state settings. By developing a robust joint model capable of efficiently handling

longitudinal, survival, and multi-state data, researchers will have an effective alternative to conventional statistical methods. The main objective is to enhance the precision of parameter estimates, leading to improved inference in data analysis. Consequently, policymakers can make better decisions drawing insights from the findings obtained through the analysis of such data, thereby benefiting from the increased accuracy and reliability of the analysis.

1.6 Thesis Overview and Lay-out

The study begins with Chapter One as the introduction which details the background of the study, the statement of the problem, study objectives, the design of the study, justification of the study, and an overview and lay-out of the thesis.

Chapter Two presents the ADNI and the PPMI datasets which are the motivating datasets this study utilizes for both the empirical and simulation study.

Chapter Three presents the review of literature. In this chapter, fundamental concepts of LMMs, survival models and MSMs which constitutes the univariate models this thesis seeks to combine to formulate the proposed hybrid two-step joint model. These concepts are discussed within the Bayesian and frequentist perspectives since the first stage of the proposed hybrid model entails a Bayesian estimation while the second stage entails a frequentist estimation.

Chapter Three also presents the fundamental concepts on joint models, and the various existing joint models, namely, the Bayesian joint likelihood, Bayesian two-step and the Rizopollous joint likelihood against which the proposed hybrid model was compared, are introduced. Also, the various association structures that link the longitudinal model to the time-to-event model are discussed.

Chapter Four, the heartbeat of the thesis, is where the proposed hybrid two-stage estimation methods for jointly modelling longitudinal, and survival and MSMs are introduced.

Chapter Five presented the simulation results when the proposed hybrid model

was applied to time to a single event. It was evident from the simulation study that, the suggested hybrid model outperformed all the other approaches in fitting the longitudinal submodel in terms of yielding relatively less bias and precise estimates at smaller sample sizes, and outperformed the BJS and the BSTS, but had similar results to the RJA at larger sample sizes. Regarding the event time submodel, the hybrid model's performance in estimating the group (γ) and association (α) parameters, was significantly enhanced at larger sample sizes, and also performed competitively well at small sample sizes. The results of the hybrid model outperformed the BSTS approach, in terms of precision at larger sample sizes and similar to the RJA and BJS.

Chapter Six constitutes the empirical application of the proposed hybrid model and the univariate models discussed in Chapter Three.

Chapter Seven extended the proposed two-stage hybrid model from time to a single event considered in Chapter Six to time to multiple events. Two versions of the hybrid multi-state model were proposed, that is, the hybrid model with 100 posterior estimates and the hybrid model with all 8000 posterior estimates. The performance of the proposed hybrid models as against the frequentist model were examined on the precision of the model estimates considering their standard errors, confidence intervals (CI), and confidence lengths (CL). It was evident that the hybrid models and the frequentist model yielded very similar estimates as well as relatively identical standard errors, CI and CL in estimating the longitudinal submodel. For the multi-state submodel, the group parameter γ is most precisely estimated by the hybrid model with 100 posterior predictive estimates while the frequentist model yields the least precise estimates. Also, the association parameter α was well estimated by the proposed hybrid models as compared to the frequentist model. The hybrid model with 100 posterior predictive estimates in this instance again, produced the most precise estimates followed by the hybrid model with all 8000 posterior estimates and then the frequentist model.

Chapter Eight, which is the last chapter, presents the summary, conclusion, and

recommendations this study suggested.

CHAPTER 2

MOTIVATING DATA SETS

2.1 Introduction

This study utilizes secondary data obtained from the ADNI and PPMI.

2.2 Parkinson's Progression Markers Initiative (PPMI) Data

The PPMI is a pioneering longitudinal, observational, and a forward-looking, multi-center research established in 2010 with the primary objective of advancing research in PD. PPMI is a collaborative effort involving academic institutions, private industry, and governmental agencies, including the Michael J. Fox Foundation for Parkinson's Research. The study has collected data from approximately 4000 participants over an 11-year period, and it is being conducted in approximately 50 centers throughout the United States and internationally (Simuni et al. 2018).

The main goal of PPMI is to identify biomarkers that can aid in the early detection and better understanding of PD progression, with the ultimate aim of accelerating the development of new therapies and personalized treatments for individuals with PD (Marek et al. 2018). The study enrolls participants in various categories, including individuals diagnosed with PD, individuals considered to be at risk of developing PD (such as those with hyposmia or REM sleep behavior disorder), and healthy individuals serving as controls.

Subjects were classified based on their disease state, which includes healthy controls (HC), PD, scans without any supporting evidence of dopaminergic deficit,

prodromal individuals, genetic cohort patients, and genetic registry patients. The focus of this thesis was on patients belonging to the HC, PD, and prodromal groups. The timing of visits was irregular, as the assessment frequency depended on the specific group. During the first year, HCs were visited every 6 months, while PD and prodromal subjects were assessed every 3 months within the first 12 months, and subsequently every 6 months (Iddi et al. 2018).

Before commencing the study, at all study sites, institutional review boards approved the study, and written informed consent was obtained from all study participants for research purposes. Participants in PPMI undergo a comprehensive set of assessments, including clinical evaluations, cognitive testing, imaging studies (MRI and DAT-SPECT scans), genetic testing, and biological sample collection (blood, CSF, and DNA). These assessments are repeated at regular intervals to track disease progression and identify potential biomarkers associated with the development and advancement of PD. Comprehensive information on the study design and other relevant details can be accessed on the dedicated website (www.ppmi-info.org).

PPMI is a groundbreaking initiative that has significantly contributed to our understanding of PD pathophysiology and the identification of potential markers for disease progression. The data collected from PPMI are openly shared with the scientific community, promoting collaborative research efforts and enabling the development of novel diagnostic tools and treatment approaches for PD. Written permission for its use was obtained from the PPMI. The PPMI data ensures strict anonymity, meaning it contains no identifying information and has been appropriately coded to prevent researchers from accessing any personal identifiers. Consequently, there is no identifiable risk that may harm the reputation of the individuals whose datasets are being utilized (Marek et al. 2011).

In this research, two outcomes from the PPMI data, namely tremor and Total Unified Parkinson's Disease Rating Scale (UPDRS Total), are modeled. The covariates considered include age, family history, diagnosis group, baseline year,

and gender.

2.3 Alzheimer's Disease Neuro-Imaging Initiative (ADNI) Data

Alzheimer's disease (AD) is a degenerative neurological condition that worsens over time. It is a prevalent driver of dementia in individuals aged 65 and above, affecting millions of people globally. However, despite its significant impact, there are currently no known prevention methods or cures for AD (Crane et al. 2012; Weiner et al. 2017).

ADNI is a landmark observational, longitudinal, and multi-center natural history study aimed at advancing research in AD and related neurodegenerative conditions. ADNI is conducted at multiple sites, researchers from 63 sites in the United States and Canada track disease progression by monitoring biological markers in conjunction with medical assessments used to evaluate the structure and function of the brain (Weiner et al. 2017). Initiated in 2004, ADNI represents a collaborative effort between academic institutions, private industry, and the National Institutes of Health (NIH) in the United States.

The primary objective of ADNI is to identify and authenticate markers for the progression of AD, which aids in early identification and disease monitoring as well as facilitating the development of potential treatments (Jagust et al. 2015). As a key resource for researchers and clinicians, ADNI has contributed a crucial role in the advancement of novel diagnostic criteria, treatment strategies, and clinical trial design in the pursuit of effective interventions for AD (Weiner and Veitch, 2015).

The ADNI study enrolls participants from four categories: subjects without cognitive impairment, subjects with early-stage mild cognitive impairment (EMCI), subjects with late-stage mild cognitive impairment (LMCI), and eventually subjects with dementia or AD. Participation in the study is voluntary,

free of charge, and does not involve medication. Individuals between the ages of 55 and 90 years are eligible to participate (Weiner et al. 2013). Participants undergo comprehensive assessments, including clinical evaluations, cognitive testing, neuroimaging (MRI and PET scans), and biological sample collection (blood and cerebrospinal fluid). These assessments are conducted at regular intervals over an extended follow-up period to track disease progression across different stages.

By collecting and analyzing a vast array of data, ADNI has significantly contributed to the understanding of AD pathophysiology and the identification of potential biomarkers associated with disease onset and progression. The data generated by ADNI have been made publicly available to researchers worldwide, promoting collaborative research efforts and fostering advancements in the field of AD and related neurodegenerative disorders.

ADNI has undergone various phases, beginning with ADNI1, which commenced in October 2014. Subsequently, ADNI2 was initiated in 2011, and the current phase, ADNI3, was launched in 2016. These phases have been funded through both private and public sectors. While participants with dementia had an optimum follow-up period of two years, previous subjects were called back for progressive follow-up. Remarkably, some ADNI-1 participants have been followed for a little over a decade (Crane et al. 2012).

For this study, the most up-to-date data from the ADNI dataset, which is ADNI3, was utilized. The selected variables considered in this thesis are mPACCtrailsB, patient age, patient gender, years from baseline, diagnostic group at baseline, and study end with states categorized as cognitive normal (CN), mild cognitive impairment (MCI), and Alzheimer's disease (AD).

CHAPTER 3

LITERATURE REVIEW

3.1 An Overview of the Existing Literature on Joint Models

While the investigation of joint models appears to be extremely promising, it is still in its initial phases (Hickey et al. 2018). This may be attributed to the complexities involved in interpreting parameter estimates within the derived joint model or the computational challenges arising when dealing with numerous or mixed types of outcomes, such as discrete or continuous (Lazaro et al. 2021). He and Luo (2016) employed a multilevel item response theory model to analyze multiple longitudinal outcomes and a Cox PH model to assess the event time outcome. The investigation aimed to understand the effect of tocopherol on individuals experiencing initial stage of Parkinson's Disease (PD) and utilized a Bayesian Markov Chain Monte Carlo (MCMC) algorithm for more accuracy in estimating the parameters of the model. Their findings demonstrated that the suggested joint model yielded superior estimates of the parameter compared to separate analyses for each outcome.

Yuen and Mackinnon (2016) conducted a comparative study utilizing both simulations and real data to assess joint models of survival outcomes against the traditional Cox regression model. Additionally, they compared various statistical software packages for estimating joint models. Their conclusions suggested that joint models offer advantages over univariate models in terms of reliability and validity, despite potential challenges they may present. They also highlighted that the results of analyses could be influenced by the specific implementation

and methodology used in the joint modelling approach.

Ha et al. (2021), Sattar and Sinha (2019), Collett (2014), Luo (2014), Rizopoulos (2012), Wu et al. (2012), Ibrahim et al. (2010), Scharfstein et al. (1999), Wulfsohn and Tsiatis (1997), and Lee and Nelder (1996) have extensively discussed the current advancements and challenges in joint modelling of longitudinal and time-to-event data. As initially suggested by Self and Pawitan (1992) and DeGruttola and Tu (1994), the fundamental formulation of joint models involves two submodels: one for the repeated measurements outcome and another for the event time outcome. This formulation has also been highlighted by Tsiatis et al. (1995), Faucett and Thomas (1996), and Wulfsohn and Tsiatis (1997).

In particular, Self and Pawitan (1992) introduced a two-stage estimation method where they incorporated survival information to calculate the expected values of covariates. They utilized partial likelihood to attain estimates of the risk parameter and obtained related variances to consider the uncertainty in the resulting covariate values. On the other hand, DeGruttola and Tu (1994) explored a joint model where the survival outcome was fitted using a parametric estimation approach, resulting in straightforward likelihood inference.

Rizopoulos et al. (2010) examined a joint model that focused on the survival process, which was linked to a longitudinal time-changing explanatory variables with measurement error. In contrast, Philipson et al. (2020) introduced a shared random effects model, emphasizing both the time-to-event and longitudinal processes. Tsiatis et al. (1995) suggested one of the very common two-step estimation methods from a frequentist perspective. In the first stage, they fitted the repeated measurements submodel and computed the trajectory function using the random effects and estimated parameters. The second stage involved fitting the event time submodel, with the trajectory expression from stage one serving as an internal time-changing covariate.

Wulfsohn and Tsiatis (1997) utilized a full likelihood approach in their proposed

joint model, incorporating an LMM for the repeated measurements outcome and a PH model for the event time outcome. Faucett and Thomas (1996) employed the MCMC technique to obtain the joint posterior distribution of all unknown parameters in their model based on the available data. Henderson et al. (2000) recommended employing two stationary Gaussian processes, involving random effects and serial correlation, to model the repeated measurements outcomes and event times, respectively.

Tsiatis and Davidian (2001; 2004) and Song et al. (2002) expanded on Wulfsohn and Tsiatis' (1997) research by allowing for a departure from the normality assumption for random effects and considering a distribution with a smooth density instead. Ibrahim (2003) proposed a semi-parametric Bayesian joint hierarchical model, offering a versatile approach for individual-specific profiles. Rogers et al. (2009), Rizopoulos and Ghosh (2011), and Rizopoulos (2012) also introduced joint models with various association functions between repeated measurements and event time outcomes.

In another approach, Li et al. (2010) put forth a joint model utilizing the partial proportional odds model to analyze the ordinal longitudinal outcome, which was then connected to the survival process, specifically competing risks events, through latent random variables. Chi and Ibrahim (2006) suggested a joint model that handles both multivariate longitudinal data and multivariate survival data. On the other hand, Andrinopoulou et al. (2017) expanded the joint model to incorporate competing risks and two longitudinal biomarkers.

Only a limited number of studies have been dedicated to examining joint models for longitudinal and multi-state data. For instances involving a longitudinal outcome and illness-death data, Dantan et al. (2011) proposed a joint multi-state model (MSM) with a latent state. Ferrer et al. (2016) approached the shared random effect domain from a frequentist perspective to introduce a joint model for repeated measurements and multi-state processes. Krol et al. (2017) conducted studies considering three simultaneous outcomes: a longitudinal outcome, time to

terminal event, and times of recurrent events. To analyze the cause of event time outcomes in the context of competing risk, Huang et al. (2011) extended joint models. In a multifaceted approach, Farewell et al. (2017) suggested a MSM that integrated an irregularly observed longitudinal categorical outcome and a multi-state outcome.

However, as far as the available knowledge goes, the hybrid estimation of longitudinal and survival or multi-state outcome has on no occasion been suggested or applied in the literature. This study therefore proposes a hybrid joint two-step estimation method with shared random effects for the repeated measurements data and transition times among several states. The proposed approach aims to address computational and implementation challenges associated with complex joint models, while ensuring that parameter estimates are obtained with minimal bias and high precision.

3.2 Review of Univariate Models

3.2.1 Introduction

Recently, there have been notable advancements in statistical techniques for analyzing longitudinal data. Longitudinal data contains two origins of variation, namely intra-individual variability and inter-individual variability. Fitting the within-subject variability allows for monitoring the changes in longitudinal data over time during follow-ups, while fitting the between-subject variability facilitates understanding the variations among individual profiles (Diggle et al. 2002; Diggle et al. 2008; Wu et al. 2010).

The choice of longitudinal analysis depends on the type of outcomes. For continuous and approximately normally distributed outcomes, the preferred method of analysis is the LMM. GLMM, an extension of the LMM, are recommended for non-Gaussian and discrete longitudinal data. Additionally, Zeger et al. (1988) suggested the use of GEE for analyzing discrete repeated

measurements observations.

3.3 Longitudinal Data Analysis

Longitudinal data involves conducting repeated measurements on subjects over time (Hedeker and Gibbons, 2006). Such data can be viewed as clustered data, where the clusters represent multiple measurements taken from a single subject at different time points, rather than measurements from different subjects or groups within the cluster (Vonesh and Chinchilli, 1997).

A distinguishing feature of longitudinal data is the presence of multiple observations from the same subject are often related, while measurements between different subjects are generally independent. This property prompts many statistical methods to focus on examining the associations within subjects (Fitzmaurice et al. 2012; Wu et al. 2010). In contrast, cross-sectional research consider data at a single time point, where all outcomes are observed simultaneously throughout the study period. This study design allows for estimating changes between subjects in the outcome variables. This fundamental contrast between longitudinal data and cross-sectional data lies in the fact that cross-sectional data records only one outcome for each subject (Diggle et al. 2002). The popularity of longitudinal data in the literature can be attributed to the following reasons;

1. The frequency of multiple observations varies across individuals
2. Large variability between subjects, that is, between-individual variability
3. Large variability within multiple measurements of a given individual
4. Missing data at some follow-up visits for some individuals

One of the advantages of repeated measurements research is the availability of more detailed data on each individual, allowing researchers to observe individual trajectories. These trajectories illustrate how the changes in the response variable

over time for each particular subject evolves. Collecting trajectories for all participants allows researchers to assess the overall trend and its association with relevant covariates. In contrast, cross-sectional data does not enable the distinction of temporal variations within subjects (Diggle et al. 2002; Thijs et al. 2000). Having repetitive observations from the same individual offers greater independent information compared to a single observation from a single individual in cross-sectional studies (Hedeker and Gibbons, 2006). Consequently, longitudinal studies possess greater statistical power compared to cross-sectional studies due to the richer and more detailed data they provide (Hedeker and Gibbons, 2006; Scholkopf et al. 2001).

Longitudinal data is distinct from other multilevel or hierarchical data types due to its arrangement in chronological sequence, leading to numerous small clusters. The measurements within the same subject tend to be associated, violating the assumption of independent measurements in conventional statistical techniques (Diggle et al. 2002). Therefore, it is crucial to consider this correlation to avoid incorrect inferences, less precise estimates, and biases from missing data. To address this, mixed effects models, such as the LMM, extend traditional regression models by incorporating random effects to capture the correlations among multiple measurements within the same subject (Vonesh and Chinchilli, 1997). The longitudinal data can comprise diverse types of response variables, including binary, categorical, continuous, or ordinal, and it can be collected either in real-time or retrospectively. Real-time data collection is generally favored to reduce the impact of recollection bias (Skrondal and Rabe-Hesketh, 2008).

3.4 A Brief Introduction to Regression Models

Regression models are widely favored due to their ability to include multiple predictors in a single model, allowing the assessment of the effect of individual predictors while controlling for the effects of other variables. These models offer great flexibility and can be adapted to different types of predictors and

outcome variables. Additionally, they produce interpretable result, making them conceptually straightforward. From a mathematical standpoint, regression models are relatively simple and not overly complex. Furthermore, most statistical packages come equipped with ready-made functions, simplifying the implementation of various forms of regression models.

The most commonly employed regression methods include linear regression (appropriate for numerical dependent variables without outliers), logistic regression (used for binary dependent variables), ordinal regression (applied to ordered factors like Likert items), robust regression (used to handle outliers), and multinomial regression (utilized for categorical dependent variables). These models vary primarily based on the types of dependent variables they can handle. Fixed-effects regression models are suited for data with a non-hierarchical structure, where data points are not nested or grouped in higher-order categories. In contrast, mixed-effects regression models incorporate a random effect structure to account for the interconnection or hierarchical structure such data points exhibit. In summary, there are two main types of regression models: fixed-effects and mixed-effects regression models.

3.5 General Linear Models (GLMs)

3.5.1 Simple Linear Regression Model

It is the most basic type of GLMs, and its formulation and distribution can be described as:

$$y_i = \boldsymbol{\beta}' \boldsymbol{X} + \varepsilon_i \tag{3.1}$$

where

$$y_i \overset{iid}{\sim} N(\boldsymbol{\beta}' \boldsymbol{X}, \sigma^2 I_n)$$

and

$$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

for $i = 1, 2, ..., N$

3.5.2 Linear Mixed Models (LMMs)

LMMs, an extended version of linear regression models to incorporate both fixed and random effects (Huong et al. 2018; Viviani et al. 2014), have gained significant popularity in various fields of research. These models are especially valuable for analyzing data with non-independence, multilevel structure, unbalanced design, longitudinal observations, or correlations among observations. In fact, they have become a fundamental tool in experimental research across disciplines, especially in studies with designs involving multiple observations (Fitzmaurice et al. 2012; Rizopoulos, 2012; Wu, 2010).

One noteworthy feature of LMM is their capacity to estimate both the average trends within the population and the individual-specific trajectories over time (Fitzmaurice et al. 2012; Herring, 2013). This allows LMMs to provide valuable insights into the overall trends within the entire population while also forecasting individual-specific patterns or changes over a given time period.

The expression for the outcome variable $y_i(t)$ of a subject i at time t is given as:

$$y_i(t) = \mu_i(t) + e_i(t) = \boldsymbol{X}'_{L,i}(t)\boldsymbol{\beta}_i + \boldsymbol{Z}'_i(t)\boldsymbol{b}_i + e_i(t) \tag{3.2}$$

where

$$\boldsymbol{b}_i \overset{iid}{\sim} N(\boldsymbol{0}, \boldsymbol{\Sigma})$$

and

$$e_i(t) \overset{iid}{\sim} N(0, \sigma^2)$$

Where, $\mu_i(t)$, is the predicted value of the repeated measurements outcome at time, t, which is linearly characterised by $\boldsymbol{X}_{L,i}(t)$ and $\boldsymbol{Z}_i(t)$, $\boldsymbol{\beta}$, is a $p \times 1$ fixed effects vector, $\boldsymbol{b}_i$, is the random effects vector with a corresponding $k \times k$ variance-covariance matrix, $\boldsymbol{\Sigma}$, where k is the count of random effects and $e_i(t)$ is an $N \times 1$ vector which represents the error term with σ^2 as its variance. $\boldsymbol{y}_i(t)$ is a matrix of

size $N \times 1$, the predictor matrix, $\boldsymbol{X}$ is of size $N \times p$ with p number of predictors. For the J groups and q number of random effects with size $qJ \times 1$, $\boldsymbol{Z}$, represents the design matrix with size $N \times qJ$. The random effects and errors all assume normality.

The distribution of Y given the random effects is expressed as;

$$Y|b_i \overset{iid}{\sim} N(\boldsymbol{X}'\beta + \boldsymbol{Z}'b, \sigma^2 I_n) \tag{3.3}$$

Let $\sigma^2 I_n = \Sigma$, then, one can write

$$Y|b_i \overset{iid}{\sim} N(\beta'\boldsymbol{X} + b_i'\boldsymbol{Z}, \boldsymbol{\Sigma}) \tag{3.4}$$

For the outcome Y_i, the marginal distribution is given as;

$$Y \overset{iid}{\sim} N(\boldsymbol{X}'\beta,\ \boldsymbol{ZDZ}' + \boldsymbol{\Sigma})$$

It can be shown that the within subject correlation will not be equal to zero but rather a ratio of the inter-individual variability and the total variability.

$$Cor(y_{ij}, y_{ij'}) = \frac{Cov(y_{ij}, y_{ij'})}{\sqrt{Var(y_{ij})Var(y_{ij'})}} \tag{3.5}$$

$$Var(y_{ij}) = \boldsymbol{ZDZ}' + \sigma^2 I_n \tag{3.6}$$

$$Cov(y_{ij}, y_{ij'}) = Cov(\boldsymbol{\beta}'\boldsymbol{X} + \boldsymbol{b}_{oi} + \boldsymbol{b}_{1i}'\boldsymbol{Z} + \varepsilon_{ij},\ \boldsymbol{\beta}'\boldsymbol{X} + \boldsymbol{b}_{oi} + \boldsymbol{b}_{1i}'\boldsymbol{Z} + \varepsilon_{ij'}) \tag{3.7}$$

$$Cov(y_{ij}, y_{ij'}) = Cov(\boldsymbol{b}_{1i}'\boldsymbol{Z},\ \boldsymbol{b}_{1i}'\boldsymbol{Z}) = Var(\boldsymbol{b}_{1i}'\boldsymbol{Z}) = \boldsymbol{ZDZ}' \tag{3.8}$$

hence,

$$Cor(y_{ij}, y_{ij'}) = \frac{\boldsymbol{ZDZ}'}{\sqrt{(\boldsymbol{ZDZ}' + \sigma^2 I_n)^2}} \tag{3.9}$$

$$Cor(y_{ij}, y_{ij'}) = \frac{\boldsymbol{ZDZ}'}{\boldsymbol{ZDZ}' + \sigma^2 I_n} \tag{3.10}$$

3.5.3 Random-intercept model

It is the simplest case of the LMM, where each subject is expected to have an individual-specific intercept but equal gradients within groups, such as treatment groups if no significant variations are observed (Drikvandi et al. 2017; Fitzmaurice et al. 2008). This model, which includes only one random effect, has an assumption that the longitudinal measurements for all subjects progress in time with identical patterns, but the individuals differ in their baseline levels (Fitzmaurice et al. 2012; Laird (1982)).

In this model, the random intercepts represent deviations of the cluster-specific intercepts (Rizopoulos, 2012; Stefanski and Carroll, 1987). It is expressed as:

$$y_{ij} = \beta_0 + \boldsymbol{b}_i + \varepsilon_{ij} \tag{3.11}$$

$$\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$$

$$\boldsymbol{b}_i \overset{iid}{\sim} N(0, d)$$

It is evident from the derivation below that the correlation between subjects is not zero but rather a ratio between the intra-variability and the total variability. The association between any pair of multiple measurements is expressed as;

$$\text{Cor}(y_{ij}, y_{ij'}) = \frac{\text{Cov}(y_{ij}, y_{ij'})}{\sqrt{\text{Var}(y_{ij})\text{Var}(y_{ij'})}} \tag{3.12}$$

but,

$$\text{Var}(y_{ij}) = \text{Var}(\beta_0) + \text{Var}(\boldsymbol{b}_i) + \text{Var}(\varepsilon_{ij}) \tag{3.13}$$

that implies,

$$\text{Var}(y_{ij}) = d + \sigma^2 = \text{Var}(y_{ij'}) \tag{3.14}$$

also,

$$\text{Cov}(y_{ij}, y_{ij'}) = \text{Cov}(\beta_0 + \boldsymbol{b}_i + \varepsilon_{ij},\ \beta_0 + \boldsymbol{b}_i + \varepsilon_{ij'}) \tag{3.15}$$

$$Cov(y_{ij}, y_{ij'}) = Cov(\boldsymbol{b}_i\boldsymbol{b}_i) + Cov(\boldsymbol{b}_i\varepsilon_{ij'}) + Cov(\varepsilon_{ij}\boldsymbol{b}_i) + Cov(\varepsilon_{ij}\varepsilon_{ij'}) \tag{3.16}$$

$$Cov(y_{ij}, y_{ij'}) = Cov(\boldsymbol{b}_i\boldsymbol{b}_i) \tag{3.17}$$

$$Cov(y_{ij}, y_{ij'}) = Var(\boldsymbol{b}_i) = d \tag{3.18}$$

Substituting the above equations into equation (2.26) yields

$$Cor(y_{ij}, y_{ij'}) = \frac{d}{\sqrt{(d+\sigma^2)(d+\sigma^2)}} \tag{3.19}$$

hence,

$$Cor(y_{ij}, y_{ij'}) = \frac{d}{d+\sigma^2} \tag{3.20}$$

3.5.4 Random Intercept and Random Slope Model

While the random intercept model is straightforward and easy to apply, assuming equal associations among multiple observations is not realistic (Rizopoulos, 2012). In reality, it is expected that the outcome trajectory for each individual varies not only at the start point but also in the progression at successive time points (Fitzmaurice et al. 2008). To address this limitation, the random intercept and slope model extends the random intercept model by considering variations in both intercepts and slopes. It is expressed as;

$$y_{ij} = \beta_0 + b_{0i} + b_{1i}t_{ij} + \varepsilon_{ij} \tag{3.21}$$

$$y_{ij} = \beta_0 + \boldsymbol{b}_i'\boldsymbol{Z} + \varepsilon_{ij} \tag{3.22}$$

$$\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$$

$$\boldsymbol{b}_i'\boldsymbol{Z} \overset{iid}{\sim} N(\boldsymbol{0},\ \boldsymbol{Z}\boldsymbol{\Sigma}\boldsymbol{Z}')$$

$$Var(y_{ij}) = \boldsymbol{Z}\boldsymbol{\Sigma}\boldsymbol{Z}' + \sigma^2 = Var(y_{ij}') \tag{3.23}$$

The model Y is marginally distributed as;

$$Y \overset{iid}{\sim} N(\beta^{'}\boldsymbol{X},\ \boldsymbol{Z\Sigma Z}^{'} + \sigma^2) \tag{3.24}$$

The covariance and correlation of the model are expressed as;

$$Cov(y_{ij}, y_{ij}^{'}) = Cov(\beta_0 + \boldsymbol{b}_i^{'}\boldsymbol{Z} + \varepsilon_{ij},\ \beta_0 + \boldsymbol{b}_i^{'}\boldsymbol{Z} + \varepsilon_{ij}^{'}) \tag{3.25}$$

$$= Cov(\boldsymbol{b}_i^{'}\boldsymbol{Z}, \boldsymbol{b}_i^{'}\boldsymbol{Z}) = Var(\boldsymbol{b}_i^{'}\boldsymbol{Z}) = \boldsymbol{Z\Sigma Z}^{'} \tag{3.26}$$

that implies,

$$Cor(y_{ij}, y_{ij}^{'}) = \frac{\boldsymbol{Z\Sigma Z}^{'}}{\boldsymbol{Z\Sigma Z}^{'} + \sigma^2} \tag{3.27}$$

3.5.5 Deriving the Probability Density Function of Y_{ij}

Given the general LMM of the form;

$$Y_{ij} = \beta^{'}\boldsymbol{X} + \boldsymbol{b}_i^{'}\boldsymbol{Z} + \varepsilon_{ij} \tag{3.28}$$

With the random effects given Y is conditionally distributed as;

$$Y|\boldsymbol{b}_i \overset{iid}{\sim} N(\beta^{'}\boldsymbol{X} + \boldsymbol{b}_i^{'}\boldsymbol{Z},\ \sigma^2 I_n) \tag{3.29}$$

If we let $\sigma^2 I_n = V$, then, we can write

$$Y|\boldsymbol{b}_i \overset{iid}{\sim} N(\beta^{'}\boldsymbol{X} + b_i^{'}\boldsymbol{Z},\ V) \tag{3.30}$$

From the Bayes Theorem, the joint probability density function (pdf) is given as;

$$\Rightarrow f(Y, \boldsymbol{b}_i) = f(Y|\boldsymbol{b}_i) \times f(\boldsymbol{b}_i) \tag{3.31}$$

Hence to get the distribution of Y marginally, we integrate the joint pdf with respect to b_i given as below;

$$f(Y) = \int f(Y, \boldsymbol{b}_i) d(\boldsymbol{b}_i) \tag{3.32}$$

$$f(Y|b_i) = \frac{1}{\sqrt{2\pi}|V|^{\frac{1}{2}}} \exp -\frac{1}{2V}\left[(Y - \beta' \boldsymbol{X})^2 + 2b_i(Y - \beta' \boldsymbol{X}) + b_i^2\right] \tag{3.33}$$

The pdf of the random effects is given as;

$$f(b_i) = \frac{1}{\sqrt{2\pi}|D|^{\frac{1}{2}}} \exp \frac{1}{2D}(b_i)^2 \tag{3.34}$$

$$f(Y) = \int \left[\frac{1}{\sqrt{2\pi}|V|^{\frac{1}{2}}} \exp -\frac{1}{2V}[(Y - \beta' \boldsymbol{X})^2 + 2b_i(Y - \beta' \boldsymbol{X}) + b_i^2] * \frac{1}{\sqrt{2\pi}|D|^{\frac{1}{2}}} \exp \frac{1}{2D}(b_i)^2\right] db_i \tag{3.35}$$

$$f(Y) = \frac{1}{\sqrt{2\pi}|\Sigma|^{\frac{1}{2}}} \exp -\frac{1}{2}\left(Y - \beta' \boldsymbol{X}\right)' \boldsymbol{\Sigma}^{-1} \left(Y - \beta' \boldsymbol{X}\right) \tag{3.36}$$

where

$\boldsymbol{\Sigma} = \boldsymbol{Z}' \boldsymbol{D} \boldsymbol{Z} + V$

3.5.6 Maximum Likelihood Estimation of β and V

The pdf of Y is defined as;

$$f(Y, \beta' \boldsymbol{X}, V) = \frac{1}{\sqrt{2\pi}|V|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(Y - \boldsymbol{X}\beta)' V^{-1} (Y - \boldsymbol{X}\beta)\right] \tag{3.37}$$

where

$$V = \boldsymbol{Z}\boldsymbol{D}\boldsymbol{Z}' + \sigma^2$$

The log-likelihood is given as;

$$lnl = \frac{-n}{2} \log(2\pi) - \frac{n}{2} \log V - \frac{1}{2}\left[(Y - \boldsymbol{X}\beta)' V^{-1} (Y - \boldsymbol{X}\beta)\right] \tag{3.38}$$

Next, we differenciate the function with respect to β

$$\frac{dl}{d\beta} = \boldsymbol{X}'V^{-1}Y - \boldsymbol{X}'V^{-1}\boldsymbol{X}\beta \tag{3.39}$$

Equating,$\frac{dl}{d\beta}$ to zero and solving for β yields;

$$\hat{\beta} = (\boldsymbol{X}'V^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'V^{-1}Y \tag{3.40}$$

The above is the MLE,$\hat{\beta}$, for β

Again, the likelihood function is differenciated with respect to V

$$\frac{dl}{dV} = \frac{-n}{2V} + \frac{1}{2}(Y - \boldsymbol{X}\beta)'V^{-2}(Y - \boldsymbol{X}\beta) \tag{3.41}$$

Equating,$\frac{dl}{dV}$ to zero and solving for V yields;

$$V = \frac{(Y - \boldsymbol{X}\beta)'(Y - \boldsymbol{X}\beta)}{n} \tag{3.42}$$

3.5.7 Computing the Expectation and Variance of β

It is known that,

$$\hat{\beta} = (\boldsymbol{X}'V^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'V^{-1}Y \tag{3.43}$$

then,

$$E(\hat{\beta}) = (\boldsymbol{X}'V^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'V^{-1}E(Y) \tag{3.44}$$

but,

$$E(Y) = \boldsymbol{X}\beta \tag{3.45}$$

that implies,

$$E(\hat{\beta}) = (\boldsymbol{X}'V^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'V^{-1}\boldsymbol{X}\beta \tag{3.46}$$

but,

$$(\boldsymbol{X}'V^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'V^{-1}\boldsymbol{X} = I_n \tag{3.47}$$

Thus,

$$E(\hat{\beta}) = \beta \tag{3.48}$$

Also, the variance is given as;

$$Var(\hat{\beta}) = Var((\boldsymbol{X}^{'}V^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{'}V^{-1}Y) \tag{3.49}$$

$$Var(\hat{\beta}) = (\boldsymbol{X}^{'}V^{-1}\boldsymbol{X})^{-1} \tag{3.50}$$

3.5.8 Bayesian Estimation of the LMMs

Since its inception, Bayesian techniques have faced much opposition from the frequentist statisticians (Barr et al. 2013; Sorensen and Vasishth, 2015) primarily due to their insolvability when computing the posterior distribution (Kruschke and Vanpaemel, 2015; Wu, 2010) which is mostly very complicated and dissimilar from the normal distributions (Kruschke, 2018; Lindstrom and Bates, 1988). But, with the invention of MCMC principles, which is employed in Bayesian techniques to pull samples from the posterior distribution, Bayesian methods have gained momentum (Kruschke, 2018; Wu and Xu, 2016). The accomplishment of Bayesian methods is also attributable to the robustness and accuracy of its output (Eager and Roy, 2017; Vasishth et al. 2018; Matuschek et al. 2017). Bayesian techniques are progressively employed to surmount the challenges of the classical approaches in dealing with complex data structures (Gelman et al. 2012; Scott and Berger, 2010).

In Bayesian statistics, the first crucial element involves incorporating prior knowledge about the parameter before observing the data is achieved through the use of prior distributions (Gelman et al. 2012; Koster and McElreath, 2017). This knowledge is often obtained from meta-analyses, systematic reviews, or other studies with comparable data (O'Hagan et al. 2006). The prior distribution's variance or precision indicates the degree of certainty or uncertainty concerning the parameter's value of focus; a bigger prior variance reflects lower certainty

regarding the value of the parameter. Bayesian inference typically involves three types of priors, each expressing varying degrees of certainty about the parameter (Kruschke and Liddell, 2018). These classes include non-informative priors, informative priors, and weakly-informative priors. Non-informative or vague priors are employed when there is no empirical evidence available about the parameter of interest. In such cases, a prior with minimal impact on the posterior distribution is selected, reflecting a significant degree of uncertainty about the population parameter.

Weakly-informative priors are designed to include some relevant information in the model and convey a higher degree of certainty about the population parameter relative to non-informative priors. However, they typically exert minimal influence on the final parameter estimate (Scott and Berger, 2010). On the other hand, informative priors provide the most certainty about the population parameter, containing precise numerical information that significantly impacts the final estimates of the model. These priors are grounded in solid empirical evidence concerning the distribution of the parameter under consideration. The degree of informativeness for these three types of priors is achieved by adjusting the hyperparameters of the prior, such as the prior mean and prior variance, to reflect precise information and levels of certainty or uncertainty concerning the model parameters under consideration.

Next, the information conveyed by the data itself becomes the next crucial element to consider. This information is represented by the likelihood function of the observed data. Subsequently, both the prior and data are integrated through Bayes' theorem. The resulting posterior distribution portrays the updated information, striking a balance between the background information provided by the prior and the evidence from the data observed (the likelihood). When a non-informative or weakly-informative prior is used, the posterior estimate may not be significantly affected by the preference of the prior. On the other hand, when informative priors are employed, the posterior results will have reduced variance.

In cases where the prior conflicts with the data information, the posterior will represent a trade-off between the two, enabling the learner to genuinely acquire new insights about the data or the underlying theory.

In classical statistics, the assumption is that there exists a single true population parameter, such as a fixed but unknown regression coefficient. However, in the Bayesian probability perspective, all unknown parameters can accommodate uncertainty, which is represented by a probability function. As a result, Bayesian techniques do not provide a single population value but instead offer an interval distribution that expresses the probability of containing the regression coefficient. In other words, each parameter is assumed to follow a distribution that accounts for the uncertainty about its true value. While the Bayesian technique combines prior distributions and likelihood to derive the posterior function, as described by Hoekstra et al. (2014) and Nalborczyk et al. (2019), frequentist techniques solely rely on the likelihood for making inferences (Kruschke and Liddell, 2018; Morey et al. 2016).

Both Bayesian and frequentist methods approach the definition of probability from distinct viewpoints. For Bayesian researchers, probability is interpreted as the representation of uncertainty based on prior knowledge or experience, whereas for frequentists, probability is seen as the limit of relative frequency in repeated experiments (Kruschke and Liddell, 2018). Bayesian methods are particularly valuable for addressing estimation problems, providing more accurate results with smaller sample sizes, and allowing the incorporation of prior knowledge into the estimation procedure.

The advantages of Bayesian methods also come with certain challenges. Firstly, for conventional models, for instance, multilevel models, Bayesian estimation can be more time-consuming compared to frequentist models due to the need to generate samples from the posterior distribution (Gelman and Hill, 2006; Lazaro et al. 2020). While this may be a drawback, the additional time spent is considered worthwhile if the Bayesian approach yields more accurate results or

enhances the understanding of the model (Natarajan and Kass, 2000). Secondly, many Bayesian models require programming, but the availability of user-friendly programs like brms in R, has alleviated some of the challenges associated with programming (Kass and Natarajan, 2006).

Thirdly, Bayesian modelling demands a deeper appreciation of probability and its distributions, which may present a challenge for many researchers who lack such expertise. Fourthly, it is not a prerequisite for users of some Bayesian software programs to specify a prior explicitly; instead, they utilize default priors hinged on the model type (Gelman et al. 2014; Schuurman et al. 2016). However, the suitability of these default priors can vary, making it difficult to determine their appropriateness for a given analysis. Therefore, it is advisable for researchers to explicitly set the priors for their models. Fifthly, when dealing with small sample sizes, obtaining reliable estimates may require careful selection of informed priors (Ciapanna and Taboga, 2019).

3.6 Survival Models

3.6.1 Review of Survival Data

Survival data pertains to the analysis of the time it takes for a predefined event to occur (Moore, 2016). Such data can be encountered in various fields, such as sociology, where researchers might explore the time until a person commits their first crime, or in economics, where the focus could be on the time until a person receives their first promotion (Collett, 2014). Survival data exhibits distinct characteristics that set it apart from other types of data (Wu et al. 2012; Rizopoulos, 2012). Firstly, the survival time is typically right-skewed, which means it does not follow symmetric distributions. Secondly, since participants may join the study at varied times, they do not have an equal number of observation times.

Moreover, survival data contains truncated or censored observations. Where the

actual event time for a subject is not fully recorded, it is referred to as censoring (Kalbfleisch and Prentice, 1973). There can be various reasons for censoring, such as the end of the study, competing risks, dropout from the study, or individuals being lost to further follow-up visits (Peto and Peto, 1972). Censoring can be categorized as right, left, or interval censoring. Left truncation occurs when participants are included in the study only after surviving a certain period, while right truncation involves participants being included based on the occurrence of the event within a designated time frame (Collett, 2014; Rizopoulos, 2012).

3.6.2 Data Structure

There are two variable outcomes, namely;

1. In this context, the time variable, t_i, represents the duration until the last disease-free observation or until the event occurs.

2. The censoring variable, c_i, takes the value of 1 if the event has happened, and 0 if there is no event by the time, t_i.

The optimal observation plan in survival analysis (SA) is prospective, which involves following a group of individuals from a specific starting point and recording the times when the event of interest occurs over a certain period. Some events, like death, occur only once, while others, such as accidents or promotions, may happen multiple times for the same individual (repeatable events). SA can also utilize retrospective data, where individuals are asked to recall the dates of past events like marriages, child-births, or promotions.

Nevertheless, it is crucial to recognize the potential limitations of retrospective data, including significant errors in recalling event times, the possibility of forgetting events altogether, challenges in providing accurate time-dependent covariates, and the potential bias from excluding individuals who may have been at risk but were lost to follow-up due to death or relocation. While prospective data are preferred, valuable insights can still be gained from retrospective data.

3.6.3 Aims of Survival Analysis

Survival analysis (SA) seeks to estimate and compare the survival experiences of different groups, where survival experience is characterized by the cumulative survival function expressed as:

$$S(t) = 1 - P(T \leq t) = 1 - F(t) \tag{3.51}$$

$F(t)$ represents the cumulative distribution function of $f(t)$, and is more interesting than $f(t)$. Specifically, the objectives of SA include;

1. To estimate the event-time for a cohort of participants, for instance the time until divorce for a cohort of subjects.
2. To compare event-time among two or more cohorts, such as treated against control subjects in a controlled trial that is randomized.
3. To investigate the impact of covariates on event-time, for instance, whether factors like temperature impact the survival time of subjects.

3.6.4 Why use Survival Analysis

SA addresses two common aspects of survival data that conventional statistical techniques may not handle effectively, namely, censoring and time-covariates. Researchers might wonder:

1. Why not use a t-test or linear regression to compare the average event time among groups? This approach does not consider censoring.
2. Why not use risk/odds ratios or logistic regression to compare the proportion of events in groups? This approach does not consider time.

All SA methods account for censoring, and many also consider time-dependent covariates. To handle censoring, a strategy involves devising a procedure that combines information from censored and uncensored cases to yield consistent

estimates of the parameters of interest. This is achieved through methods like maximum likelihood or partial likelihood estimation. Time-varying observations can also be incorporated using these likelihood-based approaches. SA is versatile enough to accommodate data from various study designs, including randomized clinical trials, whether prospective or retrospective.

3.6.5 Basic Quantities and Models

Let X be a random variable representing the time until a specific event of interest occurs, originating from a uniform population. This event could encompass various scenarios, such as the termination of breast-feeding, the recurrence of a disease, or transitions above or below a meaningful clinical threshold, like CD4 count, among others. The distribution of X is defined by several functions, including the survival function, hazard function, probability density or mass function, and the mean residual life at time x. If any of these functions are known, the others can be derived. In practical applications, the cumulative hazard function, in combination with these functions, is commonly employed to analyze various characteristics of the distribution of X.

Survival Functions

It represents the probability that an individual will survive beyond a given time, x. Mathematically, it can be expressed as:

$$S(x) = Pr(X > x) = \int_x^{\infty} f(t)dt = 1 - F(x) \tag{3.52}$$

The survival function is the complement of the cumulative distribution function (CDF), which is a monotonically increasing function of time, denoted as t, ranging from 0 to 1. Its mathematical expression is as:

$$S(x) = 1 - F(x) \tag{3.53}$$

where

$$F(x) = Pr(X \leq x) \tag{3.54}$$

The cumulative distribution function represents the probability that the event will occur before or at time t. It is derived by integrating the probability density function, denoted as $f(x)$, and is mathematically related to the survival function as:

$$S(x) = Pr(X > x) = \int_x^{\infty} f(t)dt \tag{3.55}$$

It then follows that,

$$f(x) = -\frac{d}{dx}S(x) \tag{3.56}$$

The survival function for a discrete random variable X is a step function that decreases at specific points and is expressed as:

$$S(x) = Pr(X > x) = \sum_{x_j > x} Pr(x_j) \tag{3.57}$$

Probability density function,f(t)

It represents the probability of the event occurring exactly at time t and is expressed as:

$$f(t) = \lim_{\Delta t \longrightarrow 0} \frac{P(t < T < t + \Delta t)}{\Delta t} \tag{3.58}$$

2.6.9 Hazard Function

The hazard function represents the probability of experiencing the event immediately after time t, given that you have survived up to time t. Mathematically, it is expressed as:

$$h(t) = \lim_{\Delta t \longrightarrow 0} \frac{P(t \leq X \leq t + \Delta t | X > t)}{\Delta t} \tag{3.59}$$

It can be shown that

$$f(t) = \lim_{\Delta t \longrightarrow 0} \frac{P(t < X < t + \Delta t)}{\Delta t} \tag{3.60}$$

that implies,

$$h(t) = \frac{f(t)}{S(t)} \tag{3.61}$$

For a random variable, X, which is continuous, then,

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dx} ln[S(x)] \tag{3.62}$$

The cumulative hazard function $H(x)$, is also expressed as;

$$H(x) = \int_0^x h(u)du = -ln[S(x)] \tag{3.63}$$

Hence, for a lifetime that is continuous

$$S(x) = exp[-H(x)] = exp[-\int_0^x h(u)du] \tag{3.64}$$

The hazard function is normally the leading estimator for describing time-to-event data as it has much information about the inherent feature of failure than the survival function.

For a random variable which is discrete, X, the hazard is expressed as;

$$h(x_j) = Pr(X = x_j | X \geq x_j) = \frac{Pr(x_j)}{S(x_{j-1})} \tag{3.65}$$

where $j = 1, 2, 3, ...$ and $S(x_0) = 1$

It can be shown that;

$$h(x_j) = 1 - \frac{S(x_j)}{S(x_{j-1})} \tag{3.66}$$

where $j = 1, 2, 3, ...$

The survival function can also be defined in relation to the conditional survival

probabilities as follows:

$$S(x) = \prod_{x_j \leq x} S(x_j)|S(x_{j-1}) \tag{3.67}$$

Hence the survival and hazard function are associated as:

$$S(x) = \prod_{x_j \leq x} [1 - h(x_j)] \tag{3.68}$$

The Mean Residual Life (mrl)

It measures the anticipated remaining lifespan at time x, of a person aged x. It is expressed as;

$$mrl(x) = E(X - x|X > x) \tag{3.69}$$

also,

$$mrl(x) = \frac{\int_x^\infty (t-x)f(t)dt}{S(t)} \tag{3.70}$$

therefore,

$$\mu = E[X] = \int_0^\infty tf(t)dt = \int_0^\infty S(t)dt \tag{3.71}$$

3.6.6 Non-parametric Methods

An initial step in analyzing survival data involves providing a numeric or pictorial representation of survival experiences by estimating the hazard or survivor function. If the interest is to comparatively assess the survival times between two cohorts of participants, for instance a placebo group versus treatment group, then the frequently used non-parametric models also known as distribution-free procedures are the Kaplan-Meier estimate, the (weighted) log-rank test and Wilcoxon test (Collett, 2014).

When there are no censored observations in survival times, the survival function

$S(t)$ can be computed using the equation;

$$\hat{S} = \frac{Number\ of\ participants\ with\ survival\ time\ \geq t}{Total\ number\ of\ participants} \tag{3.72}$$

Where $t_1 < t_2 <, ..., < t_r$.

When dealing with situations that involve censored observations, the survival function $S(t)$ can be estimated using the Kaplan-Meier non-parametric estimator denoted as $\hat{S}$, as proposed by Kaplan and Meier in 1958 as:

$$\hat{S} = \prod_{i=1}^{m} (\frac{n_i - di}{n_i}) \tag{3.73}$$

for $t_{(m)} \leq t < t_{(m+1)}, m = 1, 2, ..., n$ where n_i is the count of participants who are yet to experience any of the events before time t_i and d_i the count of participants who experience the event at time t_i.

This leads to the Product-Limit estimator, expressed as; In the case $t_1 \leq t$ then,

$$\hat{S}(t) = \prod_{t_i \leq t} \left[1 - \frac{d_i}{Y_i}\right] \tag{3.74}$$

Again, in the case where $t < t_1$ then,

$$\hat{S}(t) = 1 \tag{3.75}$$

For right-censored data,$\hat{S}(t)$, has proven to be an efficient estimation method for the survival and cumulative hazard function. The estimator is;

$$\hat{H}(t) = -ln[\hat{S}(t)] \tag{3.76}$$

In estimating the cumulative hazard, under small sample size, Nelson (1972), proposed the Nelson-Aalen estimator and was re-echoed in 1978 by Aalen (Colosimo et al. 2002) as an efficient model. It is expressed as;

In the case where $t_1 \leq t$ then

$$\tilde{H}(t) = \sum_{t_i \leq t} \frac{d_i}{Y_i} \tag{3.77}$$

Also, in the situation where $t \leq t_1$ then

$$\tilde{H}(t) = 0 \tag{3.78}$$

3.6.7 Semi-parametric Model

A primary focus in analyzing survival data is to navigate how the survival times of a cohort of participants are influenced by the values of certain covariates. For instance, participants in a clinical study may vary in biomarker measurements as well as demographic characteristics such as educational status among others. It is possibles, that all these explanatory variables or some may have an influence on the survival times of participants. These variables may be fixed or vary over time. The Cox regression or PH model suggested by Cox in 1972 is the fundamental model employed to model this kind of survival data (Colosimo et al. 2002). It is a combination of a non-parametric (distribution-free) function in the hazard at baseline with a parametric functional form, often step-wise, in the explanatory variable part, and this accounts for semi-parametric attribute. It is expressed in the form;

$$log\left(\frac{h_i(t)}{h_0(t)}\right) = \boldsymbol{w}_{1i}\boldsymbol{\gamma}_1 + \boldsymbol{w}_{2i}\boldsymbol{\gamma}_2 + ... + \boldsymbol{w}_{pi}\boldsymbol{\gamma}_p \tag{3.79}$$

Where $h_0(t)$ is unspecified and non-negative , $\boldsymbol{w}_i = (\boldsymbol{w}_1, ..., \boldsymbol{w}_p)'$ represents the predictor variables vector, and $\boldsymbol{\gamma}$ denotes the regression coefficients vector. The partial likelihood approach is mostly employed for inference for the Cox estimator, Cox (1975), which is expressed as;

$$L_p(\boldsymbol{\gamma}) = \prod_{i=1}^{N} \left(\frac{exp(\boldsymbol{w}_i'\boldsymbol{\gamma})}{\Sigma_{l \in R(t_i)} exp(\boldsymbol{w}_i'\boldsymbol{\gamma})}\right)^{\delta_i} \tag{3.80}$$

Here $R(t_i)$ are the subjects that are susceptible to the phenomenon of interest at a time t_i. The resulting log-likelihood is;

$$logL_p(\boldsymbol{\gamma}) = \Sigma_{i=1}^{N}\delta_i(\boldsymbol{w}_i'\boldsymbol{\gamma} - log\Sigma_{l\in R(t_i)}exp(\boldsymbol{w}_i'\boldsymbol{\gamma})) \tag{3.81}$$

The optimum partial likelihood estimates of γ are attained by solving the equation

$$\frac{\delta L_p(\boldsymbol{\gamma})}{\delta \boldsymbol{\gamma}} = \Sigma_{i=1}^{N}\delta_i\left(\boldsymbol{w}_i - \frac{\Sigma_{l\in R(t_i)}\ \boldsymbol{w}_i \exp(\boldsymbol{w}_i'\boldsymbol{\gamma})}{\Sigma_{l\in R(t_i)}exp(\boldsymbol{w}_i'\boldsymbol{\gamma})}\right) = 0 \tag{3.82}$$

The MLE $\hat{\boldsymbol{\gamma}}$ is unbiased, efficient, and asymptotically normally distrbuted.

If $N \to \infty$ then

$$\hat{\boldsymbol{\gamma}} \overset{iid}{\sim} N(\boldsymbol{\gamma}[E(I(\boldsymbol{\gamma}))]^{-1}) \tag{3.83}$$

where $I(\boldsymbol{\gamma})$ is the Fisher information matrix.

3.6.8 Parametric Model

As previously discussed, the Cox regression model offers advantages as it does not require making assumptions about a specific probability distribution for survival times. However, in certain scenarios, assuming a parametric distribution, like exponential or Weibull, may be reasonable for the survival data (Wienke, 2010). In such cases, a parametric proportional hazards (PH) model can provide more precise inferences compared to the Cox PH model. Parametric survival models, including the PH model, describe survival times using parametric specifications (Cox, 1972; Cox, 1975).

3.6.9 The Weibull Proportional Hazards Model

Parametric survival models assuming Weibull distribution have gained much popularity due to its hazard function being able to accommodate different forms (Rizopoulos et al. 2014; Prentice, 1982), offering much flexibility in analyzing survival data; it shares the characteristics of both the PH and AFT models; since

survival observations are normally biased, summary measures, for example, the median and percentile values can be effortlessly be attained to describe the data instead of relying on the mean and standard deviation (Kalbfleisch and Prentice, 2011). It is denoted by $W(\lambda exp(\boldsymbol{\omega}_i^{'}\boldsymbol{\gamma}), \tau)$ and its hazard function expressed as;

$$h_i(t) = h_0(t) exp(\boldsymbol{\omega}_i^{'}\boldsymbol{\gamma}) \tag{3.84}$$

for $i = 1, 2, 3, ..., N$

where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, ..., \gamma_p)^{'}$, $\boldsymbol{\omega}_i = (\omega_{1i}, \omega_{2i}, ..., \omega_{pi})^{'}$ is a vector of model coefficients and covariates respectively, and $h_0(t)$ is the function for the baseline. For a survival experience with a Weibull distribution, $W(\lambda, \tau)$, then its hazard function is $h_0(t) = \lambda\tau t^{\tau-1}$. Using the Weibull PH model, the hazard function takes the form of:

$$h_i(t) = \tau t^{\tau-1}[\lambda exp(\boldsymbol{\omega}_i^{'}\boldsymbol{\gamma})] \tag{3.85}$$

$for\ i = 1, 2, 3, ...N$

and its survival function expressed as;

$$S_i(t) = exp\left\{-\lambda t^{\tau} exp(\boldsymbol{\omega}_i^{'}\boldsymbol{\gamma})\right\} \tag{3.86}$$

Where the scale parameter is $\lambda exp(\boldsymbol{\omega}_i^{'}\boldsymbol{\gamma})$ and the shape parameter is τ. The model shares a lot of similarity with the Cox regression model but the distinct variety between the two is that the distribution of baseline hazard is not defined in the Cox model and its shape is driven by the observed data (Ren and Gui, 2021; Collett, 2014).

3.6.10 Statistical Inference and Computational Techniques

Take into account the survival data $(t_i, \delta_i), i = 1, 2, ..., N$. The likelihood of the Weibull PH is expressed as;

$$L(\lambda, \tau, \gamma) = \prod_{i=1}^{N} [h_i(t_i)]^{\delta_i} S_i(t_i) \tag{3.87}$$

$$L(\lambda, \tau, \gamma) = \prod_{i=1}^{N} \left[\lambda \tau t_i^{\tau-1} exp(\boldsymbol{\omega}_i' \boldsymbol{\gamma})\right]^{\delta_i} exp\left(-\lambda t_i^{\tau} exp(\boldsymbol{\omega}_i' \boldsymbol{\gamma})\right) \tag{3.88}$$

$$L(\lambda, \tau, \gamma) = \prod_{i=1}^{N} \left[\tau t_i^{\tau-1} exp(\boldsymbol{\omega}_i' \boldsymbol{\gamma})\right]^{\delta_i} exp\left(-t_i^{\tau} exp(\boldsymbol{\omega}_i' \boldsymbol{\gamma})\right) \tag{3.89}$$

This results in the log-likelihood;

$$log\ L(\lambda, \tau, \gamma) = \Sigma_{i=1}^{N} \delta_i log(\tau) + (\tau - 1) \Sigma_{i=1}^{N} \delta_i log(t_i) + \Sigma_{i=1}^{N} \delta_i \boldsymbol{\omega}_i' \gamma - \Sigma_{i=1}^{N} t_i^{\tau} exp\left(\boldsymbol{\omega}_i' \boldsymbol{\gamma}\right) \tag{3.90}$$

The Newton-Raphson procedure is employed to estimate $\boldsymbol{\theta} = (\lambda, \tau, \gamma)'$ which results in the iterative equations;

$$\theta^{(m+1)} = \boldsymbol{\theta}^{(m)} + \left[-\frac{\delta^2 logL(\boldsymbol{\theta})}{\delta \boldsymbol{\theta} \delta \boldsymbol{\theta}'}\right]^{-1} \left[-\frac{\delta logL(\boldsymbol{\theta})}{\delta \boldsymbol{\theta}}\right]^{-1} |_{\theta^{(m)}} \tag{3.91}$$

for $m = 0, 1, 2, ...$, with some baseline starting value $\boldsymbol{\theta}^0$ of $\boldsymbol{\theta}$.

The score function becomes;

$$\begin{bmatrix} \frac{\delta logL(\boldsymbol{\gamma})}{\delta \boldsymbol{\gamma}} \\ \frac{\delta logL(\tau)}{\delta \tau} \end{bmatrix} = \begin{bmatrix} \Sigma_{i=1}^{N} \delta_i \boldsymbol{\omega}_i - t_i^{\tau} \boldsymbol{\omega}_i\ exp(\boldsymbol{\omega}_i' \boldsymbol{\gamma}) \\ \Sigma_{i=1}^{N} \frac{\delta_i}{\tau} + \Sigma_{i=1}^{N} \delta_i log(t_i) - \Sigma_{i=1}^{N} t_i^{\tau} log(t_i)\ exp(\boldsymbol{\omega}_i' \boldsymbol{\gamma}) \end{bmatrix} \tag{3.92}$$

Which results in the Fisher Information (Hession matrix) as

$$I(\boldsymbol{\theta}) = \begin{bmatrix} -\frac{\delta^2 l}{\delta\boldsymbol{\gamma}\delta\boldsymbol{\gamma}'} & -\frac{\delta^2 l}{\delta\boldsymbol{\gamma}\delta\tau'} \\ -\frac{\delta^2 l}{\delta\tau\delta\boldsymbol{\gamma}'} & -\frac{\delta^2 l}{\delta\tau\delta\boldsymbol{\gamma}'} \end{bmatrix} \tag{3.93}$$

For the Weibull distribution, it results in

$$I(\theta) = \begin{bmatrix} \Sigma_{i=1}^{N} t_i^{\tau}\boldsymbol{\omega}_i \; exp(\boldsymbol{\omega}_i'\boldsymbol{\gamma})\boldsymbol{\omega}_i & \Sigma_{i=1}^{N} t_i^{\tau} log(t_i)\boldsymbol{\omega}_i \; exp(\boldsymbol{\omega}_i'\boldsymbol{\gamma}) \\ \Sigma_{i=1}^{N} t_i^{\tau} log(t_i)\boldsymbol{\omega}_i \; exp(\boldsymbol{\omega}_i'\boldsymbol{\gamma}) & \Sigma_{i=1}^{N} \frac{\delta_i}{\tau^2} + \Sigma_{i=1}^{N} t_i^{\tau}(log(t_i))^2 exp(\boldsymbol{\omega}_i'\boldsymbol{\gamma}) \end{bmatrix} \tag{3.94}$$

3.6.11 Formulating the Likelihood for Censored and Truncated Data

In computing the functions of the likelihood for data that is censored or truncated, it is essential to be attentive and diligent in regards to information that each measurement presents. It is assumed that the event-time and censoring time are independent, and if they are dependent, specific methods are necessary. In computing the likelihoods for the different forms of censoring strategies, the following elements are incorporated;

$$Exact\ lifetime \rightarrow f(x) \tag{3.95}$$

$$Right\ censored\ measurements \rightarrow S(C_r) \tag{3.96}$$

$$left\ censored\ measurements \rightarrow 1 - S(C_l) \tag{3.97}$$

$$interval\ censored\ measurements \rightarrow S(L) - S(R) \tag{3.98}$$

$$left\ truncated\ measurements \rightarrow f(x)|S(Y_L) \tag{3.99}$$

$$right\ truncated\ measurements \rightarrow f(x)|[1 - S(Y_R) \tag{3.100}$$

$$interval\ truncated\ measurements \rightarrow f(x)|[S(Y_L) - S(Y_R) \tag{3.101}$$

The expression for the likelihood is thus computed as;

$$L \propto \prod_{i \in D} f(x_i) \prod_{i \in R} S(C_r) \prod_{i \in L} (1 - S(C_L)) \prod_{i \in I} (S(L_i) - S(R_i)) \tag{3.102}$$

Here D is the death times, R is the right-censored measurements, L is the left-censored measurements, and I is the interval censored measurements.

3.7 Multi-State Models (MSM)

3.7.1 Review of MSM

SA is concerned with the study to time to a single event but the study of survival data consists of a multiple times of interest, not just one. For instance, re-occurring events or more generally multi-state data (Prentice et al. 1978). In a survival study, it is common to observe a process with two stages, where clients may move from a "credit worthy" stage to a "default" stage. However, for some research, the "default" state may be further subdivided into intermediate stages, each relating to a specific state of interest. In such cases, MSMs are employed to analyze the transition of participants between these several states. In sum, MSMs are helpful in estimating evolution rates, examining the impact of subject-specific risk factors, predicting rates of survival, and making prognostic forecasts (Meira-Machado et al. 2009 ;Slama et al. (2002).

MSMs are mostly employed to examine life history data, such as disease progression, where each individual is assumed to belong to one of several discrete states at any given time. These models are preferred in many clinical applications because they align well with the underlying biological mechanisms or disease dynamics, which suggest categorical state occupancy over time (Hougaard, 1999). MSM offer a preferred analytical approach for understanding disease pathways as they can reveal associations across multiple states that may not be directly apparent when analyzing survival separately for each state or outcome. In MSM,

a change of state is referred to as a transition or event, and states can be either transient or absorbing, with absorbing states having no transitions to other states, for example, death.

The complex nature of an MSM is contingent significantly on the count of stages of the phenomenon and probable movements involved. Eulenburg et al. (2016) highlighted in their recent study on breast cancer endpoints that MSMs allow for the simultaneous examination of movements among different stages and accommodate adjustments for intermediate events, making them a preferable choice over traditional Cox proportional hazards models. This perspective has been further reinforced by Aralis (2016) in their recent study advocating the use of multistate illness-death models for cancer research rather than mostly employed survival models that rely on composite endpoints such as disease-free progression. Examples of multi-state processes include survival models, competing risk models, illness-death models with death from any cause, and staged disease progression models, among others.

3.7.2 General framework of MSM

A brief summary of the multistate methodology is provided, focusing on the Markovian multistate models (MSM). We will begin by explaining the concept of a general continuous-time multistate process and subsequently introduce a series of assumptions that make MSMs applicable to real-world data. In the context of MSMs, independent samples of units, such as individuals, are collected, and each unit (individual) is considered as its own continuous-time multistate process. A multistate process refers to a situation where an entity can be in any of a finite number of states at different points in time. The key assumption for Markovian MSMs is that they are continuous and homogeneous.

At a given time t, the homogeneous Markov process $Y(t)$ is defined with values taken from a state space $K = \{1, 2, ..., N\}$. The Markov assumption ensures that the future behavior of the phenomenon under study is contingent purely on its

current state, given its past, and the period of each movement should be more than the former one. That is:

$$Pr(Y(t+u) = r|Y(t) = h, Y(s), s < t) = Pr(Y(t+u) = r|Y(t) = h) \quad (3.103)$$

$$\forall (h, r) \in K, \forall u \geq 0)$$

The homogeneous assumption implies that the process evolves in a consistent manner over time.

A multi-state phenomenon is characterized as a stochastic process $(X(t), t \in T)$ with a finite state space $K = \{1, 2, ..., N\}$ within the time space $\Gamma = [0, \tau]$, where $\tau < \infty$. At any given time t, $X(t)$ represents the particular stage the process is in. As the phenomenon progresses over time, it generates past records H_{t-} (a σ-algebra), which comprises measurements of the phenomenon over the time interval $[0, t)$, including previously visited states and transition times, among other information. Where $K = 2$, the multi-state processes assumes one of its special cases, the survival process. As in SA, the MSM hinges on several functions which describe important measures, such as the transition probabilities. The complete description of the process is determined by the intensity of transition among states $h \in K$ and $r \in K$ at time t, which can be represented as:

$$\lambda^i_{hr}(t) = \lim_{dt \to 0} \frac{Pr\left(X(t+dt) = r|X(t) = h\right)}{dt} \quad (3.104)$$

with

$$\lambda^i(t) = \left\{\lambda^i_{hr}(t)\right\} \quad (3.105)$$

the matrix of size $M \times M$ subject-specific intensities of transition, denoted $\Lambda^i(t)$, consists of non-diagonal elements

$$\Lambda^i_{hr}(t) = \int_0^t \lambda^i_{hr}(u)\, du, \ \forall h \neq r \quad (3.106)$$

and diagonal elements

$$\Lambda_{hh}^{i}(t) = -\sum_{r \neq h} \Lambda_{hr}^{i}(t) \tag{3.107}$$

The transition probability is also defined as:

$$P_{hr}^{i}(s,t) = Pr(X(t) = k | X(s) = h) \tag{3.108}$$

with $s \leq t$.

Let

$$P^{i}(s,t) = \left\{ P_{hr}^{i}(s,t) \right\} \tag{3.109}$$

the transition probabilities matrix, satisfied by the Chapman-Kolmogorov expression is:

$$P^{i}(s,t) = P^{i}(s,u) P^{i}(u,t) \tag{3.110}$$

with $0 \leq s \leq u \leq t$.

$P^{i}(s,t)$ is the sole outcome of the forward differential formula as specified by Kolmogorov as:

$$P^{i}(s,s) = I \tag{3.111}$$

$$\frac{\delta}{\delta t} P^{i}(s,t) = P^{i}(s,t)\, \lambda^{i}(t) \tag{3.112}$$

This outcome is hinged on the differential equations system for all the potential states $(h, r, l) \in K^3$:

$$\frac{\delta P_{hl}^{i}(s,t)}{\delta t} = \sum_{r} P_{hr}^{i}(s,t) \lambda_{hr}^{i}(t); P_{hr}(s,s) = \delta_{hr} \tag{3.113}$$

with $\delta_{hr} = 1$ if $h = k$ and 0 otherwise.

Similar to SA, researchers have proposed both parametric and non-parametric functions for these important measures, along with regression models.

3.7.3 Non-parametric estimators of MSM

Let $N_{hr}(t)$ represent the count of observed movements from states h to k until time t, and $Y_h(t)$ indicate the count of subjects in state h prior to time t. The elements $h_r(t)$ of the cumulative intensities are estimated in a non-parametric manner using the Nelson-Aalen estimator, given by:

$$\hat{\Lambda}^*_{hr}(t) = \int_0^t \frac{dN_{hr}(u)}{Y_h(u)}, \ h \neq r \tag{3.114}$$

and

$$\Lambda^*_{hh}(t) = -\sum_{r \neq h} \Lambda^*_{hr}(t) \tag{3.115}$$

The answers satisfying the Kolmogorov equations allows the probabilities of transition $P^*(s,t)$ to be non-parametrically expressed employing the product-integral as:

$$P^*(s,t) = \prod_{(s,t)} \left(I + d\hat{\Lambda}^*(u)\right) \tag{3.116}$$

The estimator $\hat{\Lambda}^*(u)$, also known as the Aalen-Johansen estimator, provides non-parametric estimates of the cumulative transition intensities at time u. For all u, the condition $d\hat{\Lambda}_{hh}(u) \geq -1$ holds, and I represents the identity matrix with dimensions $(M \times M)$.

Let $s < T_1^* < ... < T_{m^*}^* \leq t$ denote the times in an ordered manner of transitions directly observed between s and t for all the subjects. From the above equation, it can be proven that;

$$\hat{P}^*(s,t) = \prod_{l=1}^{m^*} \left(I + \Delta\hat{\Lambda}^*(T_l^*)\right) \tag{3.117}$$

where

$$\Delta\hat{\Lambda}^*(T_l^*) = \hat{\Lambda}^*(T_l^*) - \hat{\Lambda}^*(T_{l-1}^*) \tag{3.118}$$

and

$$\Delta\hat{\Lambda}^*_{hh}(T_l^*) \geq -1 \tag{3.119}$$

for all T_l^*.

3.7.4 Semi-Parametric and Parametric Estimators of MSM

Within the framework of MSM, it is possible to examine subject-specific intensities of transition of an event say $\lambda_{hk}(t)$ by incorporating explanatory variables, employing for example proportional hazards regression models:

$$\lambda_{hk}^{i}(t) = \lambda_{0,hr}(t) exp\left\{X_{hr,i}^{T}\gamma_{hk}\right\} \tag{3.120}$$

with $(h,k) \in S^2$ and $X_{hk,i}$ the covariates vector observed at baseline connected with the intensity of transition from state h to r via the parameters vector γ_{hk}. The intensity of transition at baseline $\lambda_{0,hk}(t)$ can either be non-parametric or parametric (De Wreede et al. 2010). Other researchers have suggested estimating the baseline transitions using splines and incorporating penalized likelihood methods. This approach includes a loess parameter (Joly and Commenges, 1999) to address localized fluctuations and discontinuities in $\lambda_{hk}^{i}(t)$.

Within the parametric context, the MLE method is commonly employed to estimate the model parameters, while the semi-parametric framework extends the Cox model by stratifying on the transitions. Upon obtaining the estimated vector of parameters $\hat{\theta}$, subject-specific transition probability $\hat{P}^{i}(s,t|\hat{\theta})$ is estimated as;

$$\hat{P}_{hh}^{i}(s,t|\hat{\theta}) = exp\left\{\hat{\Lambda}_{hh}^{i}(t|\hat{\theta}) - \hat{\Lambda}_{hh}^{i}(s|\hat{\theta})\right\} \tag{3.121}$$

$$\hat{P}_{hr}^{i}(s,t|\hat{\theta}) = \int_{s}^{t} \hat{P}_{hh}^{i}\left\{(s,u|\hat{\theta})\hat{\lambda}_{hr}^{i}\left(u|\hat{\theta}\right)\hat{P}_{rr}^{i}\left(u,t|\hat{\theta}\right)du, h \neq r\right\} \tag{3.122}$$

In situations where the state space K is extensive, the integrals mentioned above become computationally challenging, and in such cases, one can estimate

$\hat{P}^i(s,t|\hat{\theta})$ using the product integral method.

$$P^*(s,t|\hat{\theta}) = \prod_{(s,t]} \left(I + d\hat{\Lambda}^*(u)\right) \tag{3.123}$$

with $d\hat{\Lambda}^i(u) \geq -1$ for all u.

Characteristically, a multistate procedure can also be defined via the transition intensities as:

$$p_{hr}(s,t) = Pr(X(t) = r|X(s) = h, H_{s-}), h, r \in K; s, t \in \Gamma; s \leq t \tag{3.124}$$

where $H_{s-} = X(t), 0 \leq t < s$ represents the observations of the phenomenon prior to time s.

And also as:

$$q_{hr}(s) = \log_{\Delta s \longrightarrow 0} \frac{p_{hr}(s, s + \Delta s)}{\Delta s} \tag{3.125}$$

The term q_{hr} represents the instantaneous risk of transitioning from states h to r, given that the process is currently in state h. These measures can be likened to the hazard function used in the Cox model. Both p_{hr} and q_{hr} are generally influenced by the past measurements H_{s-} and the current time.

The semi-Markov process represent another bloc of models where the intensity of transition q_{hr} is permitted to depend on the period in state h and it is described as homogenous where dependence is only on duration. Upon violation of the Markov assumption, it becomes challenging to get transition probabilities estimates that are not biased. In such situations, state occupation probabilities $p_h(t)$ are normally used to describe the likelihood of being state h at time t:

$$p_h(t) = \sum_{k=1}^{N} p_k(0)\, p_{kh}(0,t) \tag{3.126}$$

The Aalen-Johansen estimator of occupation probabilities is unbiased even if the Markov assumption is violated, where $p_k(0)$ represents the initial distribution at time 0.

3.8 Software Programs; R Packages

Although Bayesian techniques outweigh the frequentist methods in terms of advantages, for example, the ability of including into the model prior information about parameters, their usage was hindered for an extended period due to the inability to compute the posterior distributions of highly complex models, such as GLMMs, analytically. MCMC algorithms, which enable random sampling from the posterior, were either unavailable or too time-consuming. However, this situation has changed with the advancement in novel algorithms as well as technological and computing power in the recent past. Presently, the majority of probabilistic programming languages incorporate these techniques, including WinBugs (Lunn et al. 2000; Spiegelhalter et al. 2003), OpenBugs (Spiegelhalter et al. 2007), JAGS (Plummer 2012), MCMCglmm (Hadfield 2010), and Stan (Alvarez et al. 2014; Carpenter et al. 2015; Gelman, 2015), among other examples.

Apart from Stan, all these techniques fundamentally merge Gibbs sampling (Geman and Geman 1984; Gelfand and Smith 1990), and Metropolis-Hastings updates (Metropolis et al. 1953; Hastings 1970) and occasionally characterized with slice-sampling (Damlen et al. 1999; Neal 2003). Although it is easy to employ, it converges slowly for models with high-dimension coupled with associated parameters (Gelman et al. 2014). Again, for the Gibbs sampling to function properly, there must be a conjugate relation between the priors and the likelihood of parameters (Gelman et al. 2014; Hoffman and Gelman, 2014), and this limits the liberty of the investigator to opt for a prior that mirrors his or her opinion.

On the contrary, Stan utilizes Hamiltonian Monte Carlo (Duane et al. 1987; Neal 2011) and its extended version, the No-U-Turn Sampler (NUTS) (Hoffman and Gelman, 2014). Stan has become the preferred programming language over JAGS and WinBUGS due to its capability to fit highly intricate models. For instance,

it can handle models with 14 fixed effects predictors and two crossed random effects by participant and item, each entailing a 14×14 variance-covariance matrix (Bates et al. 2015), which is often not feasible or too time-consuming to fit in JAGS or WinBUGS. Irrespective of the priors being conjugate or not, Stan techniques relatively converge very quick particularly for models with high dimension (Hoffman and Gelman 2014). Like other packages, with WinBugs as an example, Stan presents a programming language of its own enhancing greater modelling flexibility (Gelman et al. 2015; Carpenter et al. 2015). Since every model in Stan has to be written, debugged, and optimized, its use is relatively limited. For investigators even conversant with Bayesian techniques, this usually consumes a lot of time as well as susceptible to errors.

The brms package, employed in this thesis, addresses this challenge for GLMMs by providing a formula syntax similar to lme4 (Bates et al. 2015; R Core Team, 2017). Conventionally, the LMM is frequently estimated within the frequentist framework using the nlme package (Pinheiro and Bates, 2000) or its successors, such as the lme4 package or the lmer package (Harrison et al. 2018), in R (Team, 2000). Since the brms package (Burkner, 2013) implements Bayesian LMMs in R with an lme4-like syntax, it will be familiar to researchers already well-versed in lme4. Moreover, brms can fit generalized linear and non-linear multilevel models, offering a wide array of distribution and link functions (O'Hagan, 2006).

3.9 Review of Joint Modesl

3.9.1 Introduction

In the study of most clinical and longitudinal data, it is progressively frequent to follow-up with participants repetitively over time and continuous or discrete response data, example, weight, height, biomarkers are collected, also, lifetime data, that is, time-to-event data, example, heart attack, organ transplantation, diagnose of disease, death and information relating to other explanatory variables

on each participant concurrently are also recorded (Diggle et al. 2002; Murray and Philipson, 2022). The association between multiple measurements and event time outcomes has sparked significant interest among researchers. Typically, the LMM is the preferred method for analyzing longitudinal data, while the Cox PH model is mostly employed for event time data (Tsiatis et al. 1995). This separate analysis approach is often referred to as two-stage modelling in scientific research (Murray and Philipson, 2022).

In the analysis of data with survival and longitudinal observations obtained over a period ot time, joint modelling techniques have proven to be an option worthy of consideration (Hickey et al. 2018), despite its numerical and computational challenges. It entails the concurrent fitting of two models, that is, the survival and longitudinal models. Current developments in joint models have culminated into software packages that have helped to overcome these challenges therefore opening the flood gates for the frequent use of joint models (Yuen and Mackinnon, 2016).

Longitudinal and event-time outcomes have common characteristics as;

1. Generally, longitudinal measurements are obtained sequentially and occasionally, and potentially with several discrepancies and prone to errors in measurement.

2. When the event time in focus occurs, it may not always eliminate the underlying longitudinal trajectory, which may still hold informative insights.

3. The hazard for the event time of concern could be impacted by the longitudinal trajectory that underlies the process.

These features give a clue that if the two processes are associated, the separate fitting of these outcomes can result to estimates that are biased (Wulfsohn and Tsiatis, 1997). An innovative and efficient technique to overcome these challenges is joint modelling method either the two stage estimation approach or the joint

likelihood estimation method (Ibrahim et al. 2010). It can guarantee that longitudinal trajectory is adequately included into the event time model. The joint modelling of longitudinal and survival data provides an integrated method that combines both types of data into a single model. This allows researchers to navigate the interdependency and association among the longitudinal trajectory and event time, leading to a more comprehensive assessment of intervention effectiveness (Zhang and Wu, 2019; Wu et al. 2012).

In the development of joint models, three key aspects need to be considered: the longitudinal sub-model, the event time sub-model, and the function that connects the two sub-models. The statistical literature proposes two main forms of joint modelling: shared latent classes and shared random effects joint models (Diggle et al. 2002). In the shared latent classes approach, participants are assumed to belong to heterogeneous subgroups, where each subgroup shares the same longitudinal trajectory and risk of the event (Lin et al. 2002; Proust-Lima et al. 2014).

Conversely, the shared random effects approach makes the assumption of a diverse population and establishes a connection between the longitudinal and event models through subject-specific random effects, which account for the variation in the longitudinal process among individuals (Tsiatis and Davidian, 2004; Wulfsohn and Tsiatis, 1997). Among these approaches, joint models with shared random effects are more prevalent in the current literature (Proust-Lima et al. 2014), and this thesis predominantly concentrates on this particular method.

Furthermore, a review of the literature reveals several joint modelling approaches that utilize random effects in their parameterization (Huong et al. 2018; Viviani et al. 2014). Rizopoulos (2010) suggested a joint model that primarily concentrates on the survival outcome, which is impacted by a time-changing and error-prone longitudinal outcome. As an extension, Philipson et al. (2020) suggested a random effects model, which focused on both the event time and longitudinal outcomes. Tsiatis et al. (1995) from a classical standpoint suggested

the famouns two-stage technique where stage one requires the fitting of the longitudinal process and computing its corresponding trajectory, $\mu_i(t)$. Then in the stage two, this computed function is included as a covariate when estimating the event time sub-model.

3.10 The Fundamental Joint Model

The fundamental joint model has two main components: the longitudinal and survival sub-models, which are connected by a random effects function denoted as b_i.

3.10.1 The Longitudinal Submodel

In this study, the repeated measures over time are modeled using the LMM specification. Here, the outcome variable $y_i(t)$ for participant i at time t is expressed as follows:

$$y_i(t) = \mu_i(t) + e_i(t) = \boldsymbol{X}_i^{'}(t)\beta_i + \boldsymbol{Z}_i^{'}(t)\boldsymbol{b}_i + e_i(t) \tag{3.127}$$

this implies

$$\hat{\mu}_i(t) = \boldsymbol{X}_{L,i}^{'}(t)\hat{\boldsymbol{\beta}}_i + \boldsymbol{Z}_i^{'}(t)\hat{\boldsymbol{b}}_i \tag{3.128}$$

Here, $\boldsymbol{X}_{L,i}(t)$ and $\boldsymbol{Z}_i(t)$ are covariate vectors and $\boldsymbol{\beta}$ and $\boldsymbol{b}_i$ are the vectors of their fixed and random effect respectively. $\mu_i(t)$ is the estimated longitudinal trajectory.

$$\boldsymbol{b}_i \overset{iid}{\sim} N(\boldsymbol{0}, \boldsymbol{\Sigma})$$

$$\boldsymbol{e}_i \overset{iid}{\sim} N(0, \sigma^2)$$

for $i = 1, ..., n$

3.11 The Survival Submodel

The event time sub-model is also expressed as:

$$h_i(t) = h_0(t)exp\left[\boldsymbol{X}'_{s,i}\boldsymbol{\gamma} + \alpha\left(\boldsymbol{X}'_{L,i}(t)\boldsymbol{\beta}_i + \boldsymbol{Z}'_i(t)\boldsymbol{b}_i\right)\right] \tag{3.129}$$

then

$$h_i(t|\mu_i(t)) = h_0(t)exp\left[\boldsymbol{X}'_{s,i}\boldsymbol{\gamma} + \alpha\hat{\mu}_i(t)\right] \ , t > 0 \tag{3.130}$$

Equation 3.130 consists of several components. The baseline hazard $h_0(t)$, which can be defined as a Weibull distribution, a piecewise-constant function, or a regression spline, represents the initial hazard at time t (Faucett and Thomas, 1996). The covariate vector $X_{s,i}$ is connected to the coefficients vector γ. The estimated longitudinal trajectory $\mu_i(t)$ links the longitudinal and event time sub-models. The parameter α represents the degree of connection between the underlying longitudinal process and the event risk.

Different association structures have been suggested in previous research (Schluchter, 1992), and the subsequent section presents an overview of some of these association structures.

3.12 A Review of Association Structures

1. **Shared Random Effects**

   This function that takes into consideration the random trend in time, in the survival process. It is expressed as;

$$\mu_i(t) = b_{0i} + b_{1i}t \tag{3.131}$$

   Hence, the survival submodel becomes:

$$h_i(t) = h_0(t)exp(X'_{s,i}\gamma + \alpha(b_{0i} + b_{1i})t) \tag{3.132}$$

It defines how each participant deviates from the population intercept and slope.

2. **Current Value**

It considers the real-time value of the longitudinal process at time t in the survival model, which is expressed as:

$$\mu_i(t) = \beta_0 + \beta_1 t + b_{0i} + b_{1i} t \tag{3.133}$$

Consequently, the survival model is then formulated as:

$$h_i(t) = h_0(t) exp(X_{s,i}^{'} \gamma + \alpha(\beta_0 + \beta_1 t + b_{0i} + b_{1i} t) \tag{3.134}$$

This study employs the current value structure to establish the connection between the longitudinal and survival processes.

3. **Lagged Value**

It is expressed as:

$$\mu_i(t) = \beta_0 + \beta_1(t-c) + b_{0i} + b_{1i}(t-c) \tag{3.135}$$

The survival model becomes

$$h_i(t) = h_0(t)exp(X_{s,i}^{'}\gamma + \alpha(\beta_0 + \beta_1(t-c) + b_{0i} + b_{1i}(t-c)) \tag{3.136}$$

4. **Current slope**

This structure includes the derivative of the actual value of the longitudinal process at time t in the survival model. Mathematically, it can be represented as:

$$\mu_i(t) = \beta_0 + \beta_1 t + b_{0i} + b_{1i}t \tag{3.137}$$

$$\mu_i(t)^{'} = \frac{d}{dt}(\beta_0 + \beta_1 t + b_{0i} + b_{1i}t) \tag{3.138}$$

Taking the first derivative of equation 3.138 yields;

$$\mu_i(t)^{'} = \beta_1 + b_{1i} \tag{3.139}$$

As a result, the survival sub-model can be formulated as;

$$h_i(t) = h_0(t)exp(X_{s,i}^{'}\gamma + \alpha(\beta_1 + b_{1i}) \tag{3.140}$$

5. **Current Value and Current slope**

This structure incorporates both the current value of the longitudinal process and its rate of change at time t into the survival model. It is mathematically represented as:

$$\mu_i(t) = \beta_0 + \beta_1 t + b_{0i} + b_{1i}t \tag{3.141}$$

$$\mu_i(t)^{'} = \frac{d}{dt}(\beta_0 + \beta_1 t + b_{0i} + b_{1i}t) \tag{3.142}$$

Taking the first derivative of equation 3.142 yields;

$$\mu_i(t)^{'} = \beta_1 + b_{1i} \tag{3.143}$$

Consequently, the survival model is then formulated as:;

$$h_i(t) = h_0(t)exp(X_{s,i}^{'}\gamma + \alpha_1(\beta_0 + \beta_1 t + b_{0i} + b_{1i}t) + \alpha_2(\beta_1 + b_{1i}) \tag{3.144}$$

This structure is suitable for situations where two subjects have similar actual values of the longitudinal trajectory at time t, but may have different rates of progression in their trajectories (Ye et al. 2008).

6. **Cumulative Effects**

This structure takes into account the cumulative impact of the longitudinal process, which means considering the total area under the longitudinal trajectory up to time t, in the survival model. It is expressed as:

$$\mu_i(t) = \int_0^t (\beta_0 + \beta_1 s + b_{0i} + b_{1i}s)ds \tag{3.145}$$

Therefore, the survival sub-model is formulated as follows:

$$h_i(t) = h_0(t)exp(X_{s,i}^{'}\gamma + \alpha(\int_0^t (\beta_0 + \beta_1 s + b_{0i} + b_{1i}s)ds \tag{3.146}$$

The key difference among the association functions discussed above lies in how they connect the longitudinal trajectory to the risk of the event. In the cumulative effect structure, the entire history of the longitudinal trajectory is considered in relation to the event risk. On the other hand, in the other structures, the event risk is only associated with specific characteristics of the longitudinal trajectory at certain time points (Rizopoulos, 2012).

7. **Weighted Cumulative Effects**

It is expressed as:

$$\mu_i(t) = \int_0^t (\omega(t-s)(\beta_0 + \beta_1 s) + b_{0i} + b_{1i}s)ds \tag{3.147}$$

Therefore, the survival sub-model is formulated as follows:

$$h_i(t) = h_0(t)\, exp X'_{s,i}\gamma + \alpha(\int_0^t (\omega(t-s)(\beta_0 + \beta_1 s) + (b_{0i} + b_{1i}s))ds \tag{3.148}$$

where $\omega()$ is some known weight function.

3.13 Methods for Estimating Joint Models

Various estimation techniques have been proposed to address the challenges in estimating joint model parameters, including Expectation-Maximization (EM) techniques, Newton-Raphson technique, Maximum Likelihood Estimation (MLE) approaches, and Bayesian techniques such as MCMC methods. In the literature, MLE and restricted MLE estimation techniques are commonly used, although Bayesian estimation has been gaining more attention in recent times. These estimation procedures can be implemented using two-stage or joint likelihood methods.

As a way of validating the the hybrid model this thesis seeks to propose, it is compared to already established models such as the joint approach proposed by Rizopoulos (2010; 2012), the Tsiatis et al. (1995) method situated within the Bayesian context by Leiva-Yamaguchi and Alvares (2020). The Bayesian joint likelihood model and the Bayesian standard two-step model are models that pick inspiration from the Tsiatis et al. (1995) appproach. In summary, three established joint models, namely, Rizopoulos (2010; 2012) joint likelihood model, the Bayesian joint likelihood method and the Bayesian standard two-step method suggested by Leiva-Yamaguchi and Alvares (2020) are evaluated against

the proposed hybrid two-step joint model. In this thesis, all three of these joint models are examined and analyzed in the following sections.

3.13.1 Rizopoulos Joint Likelihood Method

The joint model suggested by Rizopoulos (2010) fits the model by optimizing the joint likelihood of the longitudinal and survival data. Currently this procedure has gained much fame because it optimally utilizes all the existing information by simultaneously modelling the repeated measurements and event time data and linking them by random effects b_i via the joint likelihood.

In this context, the longitudinal outcome is represented by Y, the event time by T, and the shared random effects by b. The joint distribution of Y and T is formulated as follows:

$$f(Y,T) = \int_b f\left(Y|b\right) f\left(T|b\right) f(b)db \tag{3.149}$$

Both models are associated through the random effects, which is accountable for the connection between both models.

Conventionally, the MLE or Bayesian techniques are the estimation procedures most commonly employed in the framework of joint modelling and they function under the following set of assumptions with regards to conditional independence as follows:

$$p(T_i, \Delta_i|b_i;\theta) = p(T_i, \Delta_i|b_{i;\theta_t})p(Y_i|b_i;\boldsymbol{\theta}_y) \tag{3.150}$$

$$p(T_i|b_i;\theta) = \prod_{j=1}^{m_i} p(Y_i(t_{ij})|b_i;\boldsymbol{\theta}_y) \tag{3.151}$$

$\boldsymbol{\theta}^{'} = \theta_y^{'}, \theta_t^{'}, \theta_b^{'},$ is respectively the parameters vector of the submodels for the longitudinal and survival processes and random effects. To compute the likelihood, the general assumption is that, the repeated measurements outcome $\{y_i\}$ and event time process $\{t_i, \delta_i\}$ are independent from each other conditioned on the random effects $\{b_i\}$ and also, the longitudinal measurements are deemed

independent with respect to each other (Henderson et al. 2000; Diggle et al. 2008). If data from both the survival and longitudinal processes is considered as ,$\{y_i, t_i, \delta_i, \{i = 1, 2, ..., N\}\}$, then the joint likelihood can be expressed as:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N} f\left[t_i, \delta_i, y_i | \boldsymbol{\theta}\right] \tag{3.152}$$

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N} \int f\left[t_i, \delta_i, y_i, b_i | \boldsymbol{\theta}\right] db_i \tag{3.153}$$

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N} \int f\left[t_i | b_i, \gamma, \tau, \alpha\right] f\left[y_{ij} | b_i, \beta, \sigma_\varepsilon^2, \right] f(b_i | D) db_i \tag{3.154}$$

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N} \int f\left[t_i | b_i, \gamma, \tau, \alpha\right] \prod_{j=1}^{n_i} f\left[y_{ij} | b_i, \beta, \sigma_\varepsilon^2, \right] f(b_i | D) db_i \tag{3.155}$$

Here $\boldsymbol{\theta}$ represents the set of all model parameters which requires to be estimated. Within this framework,

- For the Weibull survival time t_i its conditional density is expressed as:

$$f(t_i | b_i; \gamma, \tau, \alpha) = h_i(t_i | b_i; \gamma, \tau, \alpha)^{\delta_i} S(t_i | b_i; \gamma, \tau, \alpha) \tag{3.156}$$

$$= \left\{\tau t_i^{t-1} exp(X_{s,i}^T \gamma + \alpha \mu_i(t))\right\}^{\delta_i} exp\left\{-\tau t_i^{\tau-1} exp(X_{s,i}^T \gamma + \alpha \mu_i(t))\right\} \tag{3.157}$$

- The assumption is that the conditional probability distribution for the longitudinal process y_{ij} is normally distributed, with a mean of $\mu_{ij} = \boldsymbol{X}_i^T \boldsymbol{\beta} + \boldsymbol{Z}_{ij}^T \boldsymbol{b}_i$ and variance σ_ε^2. Given $\boldsymbol{y}_i = (y_{i1}, y_{i2}, ..., y_{in_i})'$ the conditional density is defined as:

$$f(y_i | u_i; \beta, \sigma_\varepsilon^2) = \prod_{j=1}^{n_i} f\left[y_{ij} | b_i, \beta, \sigma_\varepsilon^2\right] \tag{3.158}$$

$$= \left(2\pi\sigma_\varepsilon^2\right)^{\frac{-n_i}{2}} exp\left\{-||y_i - x_{ij}'\beta - z_{ij}' b_i||^2 / 2\sigma_\varepsilon^2\right\} \tag{3.159}$$

where $||x|| = \{x_i^2\}^{\frac{1}{2}}$ is the Euclidian vector's norm. The random effects , $\boldsymbol{b}_i$,

which serve as the association function also assume normality, $\boldsymbol{b}_i \overset{iid}{\sim} N(0, D)$. If the association parameter, $\alpha = 0$, the joint model resets to the separate analysis of the longitudinal and survival sub-models. For the data observed, its log-likelihood is expressed as:

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{N} \log \int f\left(t_i, \delta_i, y_i, b_i | \boldsymbol{\theta}\right) db_i \tag{3.160}$$

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N} \int f\left[t_i | b_i, \gamma, \tau, \alpha\right] \prod_{j=1}^{n_i} f\left[y_{ij} | b_i, \beta, \sigma_\varepsilon^2, \right] f(b_i | D) db_i \tag{3.161}$$

It is important to highlight that the log-likelihood involves calculating integrals related to the conditional distribution $b_i|(y_i, t_i, \delta_i)$, which can be challenging to converge, making the evaluation of the integral in $l(\boldsymbol{\theta})$ very difficult(Liu, 2009; Zhang et al. 2021). As a result, the Maximum Likelihood Estimation (MLE) technique was adopted to overcome this challenge (Culpepper, 2016; Sweeting and Thompson, 2011). The MLE of $\boldsymbol{\theta}$ is achieved by optimizing the log-likelihood function concerning $\boldsymbol{\theta}$ through an iterative algorithm, such as the Newton-Raphson technique (Man et al. 2019; Cross et al. 2010).

3.13.2 Formulating Joint Model From A Bayesian Perspective

Assume that repeated observations and an associated event time of interest on n subjects are collected. Specifically, the characteristics of the longitudinal trajectory, which capture the repeated observations, are shared with the event time process (Murawska et al. 2012; Huong et al. 2018; Viviani et al. 2014).

3.13.3 Bayesian Joint Likelihood Model

Represent the repeated outcome with y and the event time by s. The vector $\boldsymbol{\theta}$ includes all parameters and hyperparameters, while b denotes the random effects.

The combined distribution of $(y, s, b, \boldsymbol{\theta})$ can be expressed as a multiple of the conditional distribution $f(y, s|b, \boldsymbol{\theta})$, the conditional distribution of the random effects $f(b|\boldsymbol{\theta})$, and the prior distribution $p(\boldsymbol{\theta})$. Mathematically,

$$f(y, s, b, \boldsymbol{\theta}) = f(y, s|b, \boldsymbol{\theta}) f(b|\boldsymbol{\theta}) p(\boldsymbol{\theta}) \tag{3.162}$$

Various methods have been suggested to define the joint conditional distribution $f(y, s|b, \boldsymbol{\theta})$, and the shared-parameter specification is widely adopted. For this approach, it is assumed that the longitudinal pattern is conditionally independent of the event time procedure, given the shared information. That is:

$$f(y, s|b, \boldsymbol{\theta}) = f(y|b, \boldsymbol{\theta}) f(s|b, \boldsymbol{\theta}) \tag{3.163}$$

The functions $f(y|b, \boldsymbol{\theta})$ and $f(s|b, \boldsymbol{\theta})$ are predominantly determined based on the specifications of the repeated measurements and event time submodels, respectively.

Within the context of joint modelling framework, the estimation of $(b, \boldsymbol{\theta})$ is carried out simultaneously hinged on the specified longitudinal and survival submodels. However, because of the increase in parameter numbers and non-converging integrals in computing the joint likelihood function, the inferential process takes a lot of time (Pinheiro and Bates, 2000).

3.13.4 Bayesian Two-Stage Model

To fasten the inferential procedure and limit the complex nature of joint models, two-stage techniques have proven to be very advantageous. From a classical perspective, Tsiatis et al. (1995) suggested one of the widely used two-phase procedures. Phase one requires the estimation of the longitudinal component and then the evolution expression $\mu_i(t)$ is computed utilizing the estimated coefficients and random effects. In phase two, this estimated expression is inputted as a time-changing covariate when estimating the event time sub-model.

As an extension, the Tsiatis et al. (1995) procedure is situated within a Bayesian framework. Precisely, in stage one, the posterior mean of the repeated observations submodel coefficients and random effects shared with the event time component is computed as:

$$\hat{\boldsymbol{\beta}} = E(\boldsymbol{\beta}|\boldsymbol{y}) \tag{3.164}$$

and

$$\hat{\boldsymbol{b}_i} = E(\boldsymbol{b_i}|\boldsymbol{y}) \tag{3.165}$$

Then in stage two, the calculated evolution function

$$\hat{\mu}_i(t) = \boldsymbol{X}'_{L,i}(t)\hat{\beta} + \boldsymbol{Z}'_i(t)\hat{\boldsymbol{b}}_i \tag{3.166}$$

is included in the event time submodel. Hence substituting into the survival sub-model

$$h_i(t|M_i(t)) = h_0(t)exp(\boldsymbol{X}'_{s,i}\boldsymbol{\gamma} + \boldsymbol{\alpha}\mu_i(t)) \tag{3.167}$$

yields

$$h_i(t|M_i(t)) = h_0(t)exp\left[\boldsymbol{X}'_{s,i}\boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\boldsymbol{X}'_{L,i}(t)\hat{\beta} + \boldsymbol{Z}'_i(t)\hat{\boldsymbol{b}}_i\right)\right] \tag{3.168}$$

where

$$\boldsymbol{b}_i \overset{iid}{\sim} N(\boldsymbol{0}, \boldsymbol{\Sigma})$$

and $i = 1, ..., n$

The Bayesian posterior distribution of $(\boldsymbol{\gamma}, \boldsymbol{\alpha})$ is subsequently computed.

3.13.5 Frequentist Standard Two-Stage Approach

For the frequentist standard two-step model adopted in this thesis, the parameters of the repeated measurements and multi-state sub-models specified in equations 4.2 and 4.4 respectively are estimated using the MLE procedure. The parameters $(\beta, \boldsymbol{b})$ of the repeated measurements submodel and $(\boldsymbol{\gamma}, \alpha)$ of the multi-state

submodel are all estimated using the MLE method. That means the estimated longitudinal trajectory estimated from equation 4.2 in the first stage of the modelling framework to be plugged into equation 4.4 in the second stage is a frequentist point estimate as opposed to a Bayesian posterior distribution in the proposed hybrid model.

3.14 Multi-state submodel

To estimate the times of transition, a Markov MSM with PH that considers the repeated measurements trajectory via the shared random effects b_i is employed. Hence, for a switch from state $d \in S$ to state $k \in S$, the intensity of transition at time t is expressed as:

$$h^i_{dk}(t|b_i) = \lim_{dt \longrightarrow 0} \frac{Pr\left(E_i\left(t+dt\right) = k | E_i(t) = d; b_i\right)}{dt} \tag{3.169}$$

$$= h_{dk,0}(t) exp\left\{ \boldsymbol{X}^S_{dk,i}(t)' \boldsymbol{\gamma}_{dk} + \alpha_{dk} \mu_{dk,i}(t)' \right\}. \tag{3.170}$$

More generally,

$$h_{dk,i}(t) = h_{dk,0}(t) exp(\boldsymbol{X}^S_{dk,i}(t)' \boldsymbol{\gamma}_{dk} + \sum_g \alpha_{dk,i,g} \mu_{dk,i,g}(t)') \tag{3.171}$$

with $h_{dk,0}(t)$ the parametric baseline intensity which could be specified different forms, for instance, for transition from d to k but for this study follows a Weibull distribution and $\boldsymbol{X}^S_{dk,i}$ the vector of prognostic factors, that is, the vector of time-changing transition-specific variables linked with the s-vector of fixed coefficients $\boldsymbol{\gamma}_{dk}$. The dependence between the two models is defined by $\mu_{dk,i}(b_i, t)$, that is, it describes the link between the g^{th} repeated measure outcome and the movement from state d to k. $\mu_{dk,i}(b_i, t)$ can assume any of the association structures outlined under section 4.4 of this chapter. The degree of connection between the longitudinal trajectory and the intensity of the transitions between the states is quantified by the s-vector of coefficients $\alpha_{dk,g}$.

3.15 Estimation of Joint Model Parameters

3.15.1 Maximum Likelihood

The prevalent method for estimating joint model parameters is MLE. Consider a set of data with n participants having longitudinal and multi-state data. The observed time for the i^{th} participant is represented as $T_i = \min(T_i, C_i)$, where T_i is the actual survival time, and C_i represents the censoring time for the i^{th} participant $(i = 1, ..., n)$. In the multi-state data, an event indicator is specified as $\delta_i = I(T_i^* \leq C_i)$. Additionally, the repeated measurements consist of data for the i^{th} participant, denoted as $y_{ij} = y_i(t_{ij}), j = 1, ..., n_i$, recorded at specific time points t_{ij}. Assuming the repeated measurements and multiple event processes are independent given the random effects, the log-likelihood function of the combined models can be expressed as:

$$l(\theta) = \sum_i \log\ p(T_i, \delta_i, y_i; \boldsymbol{\theta}) \tag{3.172}$$

$$= \sum_i \log \int_{b_i} p(T_i, \delta_i, y_i, b_i; \boldsymbol{\theta})\, db_i \tag{3.173}$$

$$= \sum_i \log \int_{b_i} p(T_i, \delta_i | b_i; \boldsymbol{\theta}_t, \beta)\, p(y_i | b_i; \boldsymbol{\theta}_y)\, p(b_i; \boldsymbol{\theta}_b)\, db_i \tag{3.174}$$

The observed data score vector for the joint model is defined as follows:

$$S(\theta) = \sum_i \frac{\delta}{\delta \theta^T} \log \int p(T_i, \delta_i | b_i; \boldsymbol{\theta}_t, \beta)\, p(y_i | b_i; \boldsymbol{\theta}_y)\, p(b_i; \boldsymbol{\theta}_b)\, db_i \tag{3.175}$$

$$S(\theta) = \sum_i \int \frac{\delta}{\delta \boldsymbol{\theta}^T} \log p(T_i, \delta_i | b_i; \boldsymbol{\theta}_t, \beta)\, p(y_i | b_i; \boldsymbol{\theta}_y)\, p(b_i; \boldsymbol{\theta}_b)\, db_i \tag{3.176}$$

It is obvious that the challenges with the estimation procedure originates from computing many integrals. These integrals normally do not converge, (Rizopoulos, 2012) particularly, as the number of random effects grows, the computation time becomes significantly high (Rizopoulos, 2011).

3.15.2 Restricted Maximum Likelihood Estimation (ReML)

The ReML procedure is a particular type of MLE that utilizes a likelihood function derived from a transformed dataset, and it does not rely on estimating parameters through a maximum likelihood fit of all the available information. ReML has become one of the most commonly used procedures for estimating variance and covariance components (Brennan, 2010; Ogden, 2014). When computing variance components, the original set of data is substituted with a set of contrasts derived from the data, and the likelihood function is computed based on the probability distribution of these contrasts, as specified by the model for the complete set of data (Davis, 2002). ReML is specifically used as a method for fitting the survival model because it results in an unbiased estimator of variance and covariance parameters (Gelman, 2013). A key features of ReML is its ability to handle data involving multiple sources of error variation and unbalanced designs. Additionally, ReML employs a powerful prediction algorithm that extends the concepts from the regression prediction algorithm to accommodate both random and fixed effects (Dempster, 1997; Hunter, 2004).

CHAPTER 4

PROPOSED HYBRID TWO-STAGE ESTIMATION METHODS FOR JOINT MODELS

4.1 Introduction

In the consideration of this thesis, the proposed hybrid two-stage joint model is first extended to time-a-single event model and again to a multi-state time event model.

4.2 Proposed Hybrid Two-Stage Method For Longitudinal and Survival Models

The hybrid two-stage model proposed is comparable to the conventional two-stage procedure, however, instead of the posterior estimates of the repeated measurements submodel coefficients and random effects being computed as an average across the posterior samples, the hybrid model extracts all these posterior estimates and estimate the second stage model using all these extracted posterior estimates. As earlier on clarified, the mathematical calculation of the likelihood for the joint model is very complex but yields relatively unbiased estimates compared to the two-stage approach which is not faced with computational complexities. Consequently, the two-stage procedure is mostly used compared to the joint likelihood model (Ibrahim, 2010; Wulfsohn and Tsiatis, 1997; Ren et al. 2019). However, the two-step model is a biased procedure because no event time information is used when fitting the longitudinal outcome and also, likely bias in

selection as a result of dropout of information not taken into consideration. This means that the procedure does not rely on information from both the repeated measurements and the event time processes concurrently during each step of model estimation (Yuen and Mackinnon, 2016; Wu et al. 2012). When the longitudinal process is only used, the LMM potentially may result in estimates that are not accurate in the first step of the estimation procedure. As a result, the estimated values of the survival sub-model parameters may be skewed and less accurate in the second phase (Ibrahim, 2010; Wu et al. 2012).

It is on this premise that the hybrid standard two-stage approach proposes a two-step estimation method with the objective of effecting a drop in the bias of the estimates of the model parameter . At phase one, from a Bayesian standpoint, an LMM is fitted to estimate the true univariate repeated measurements sub-model and then, at the step two, all the posterior predictive estimates are plugged into in the multi-state process as a covariate. The benefit of this procedure to the conventional two-step procedure is that we are able to take into consideration the variability of the estimation in step one into step two of the modelling framework. In summary, the proposed hybrid two-step estimation procedure works as outlined below:

- Step One: The longitudinal covariates from a Bayesian perspective are modelled using an LMM in order to compute the individual-specific covariate values.
- Step Two: The event time submodel is estimated employing the fitted estimates in step one as time-changing covariate values and the model parameters estimated within a frequentist domain.

Specifically, the hybrid method adapts the Tsiatis et al. (1995) approach and situate it within the frequentist and the Bayesian frameworks. Precisely, in step one, all the posterior predictive values of the parameters of the repeated measurements submodel as well as random effects shared with the event time

submodel are extracted from the Bayesian LMM. For each posterior estimate, $m = 1,\ 2,\ 3,\ 4, ...,\ M$, the population and individual effects are given as:

$$\hat{\boldsymbol{\beta}}^{(m)} = E(\boldsymbol{\beta}^{(m)}|y) \tag{4.1}$$

and

$$\hat{\boldsymbol{b}}_i^{(m)} = E(\boldsymbol{b}_i^{(m)}|y) \tag{4.2}$$

In step two, the trajectory functions defined above are incorporated into the event time submodel taking into account

$$\hat{\mu}_i^{(m)}(t) = \boldsymbol{X}'_{L,i}(t)\hat{\boldsymbol{\beta}}^{(m)} + \boldsymbol{Z}'_i(t)\hat{\boldsymbol{b}}_i^{(m)} \tag{4.3}$$

$$\hat{\boldsymbol{b}}_i^{(m)} \overset{iid}{\sim} N\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$$

That is, every posterior predictive estimate from all the iterations specified in the Bayesian LMM at the first stage of the joint modelling process is substituted in turns in the above equation when estimating the longitudinal trajectory, $\hat{\mu}_i(t)$. For this, the event time submodel is expressed as

$$\boldsymbol{h}_i^{(m)}(t|\mu_i(t)) = h_0^{(m)}(t)exp(\boldsymbol{X}'_{s,i}\boldsymbol{\gamma}^{(m)} + \boldsymbol{\alpha}\hat{\mu}_i{}^{(m)}(t)) \tag{4.4}$$

is now expressed as

$$\boldsymbol{h}_i^{(m)}(t|\mu_i(t)) = h_0^{(m)}(t)exp(\boldsymbol{X}'_{s,i}\boldsymbol{\gamma}^{(m)} + \boldsymbol{\alpha}\left[\boldsymbol{X}'_{L,i}(t)\hat{\boldsymbol{\beta}}^{(m)} + \boldsymbol{Z}'_i(t)\hat{\boldsymbol{b}}_i^{(m)}\right] \tag{4.5}$$

Moreover, assuming a Weibull PH model, the above specified risk equation may be expressed as:

$$\boldsymbol{h}_i^{(m)}(t|\mu_i(t)) = \left(\tau t^{t-1}\right)^m exp\left(\boldsymbol{X}'_{s,i}\boldsymbol{\gamma}^{(m)} + \boldsymbol{\alpha}\left[\boldsymbol{X}'_{L,i}(t)\hat{\boldsymbol{\beta}}^{(m)} + \boldsymbol{Z}'_i(t)\hat{\boldsymbol{b}}_i^{(m)}\right]\right) \tag{4.6}$$

where

$$\tau = exp\,(\gamma_0) \tag{4.7}$$

let

$$h_0(t) = \tau t^{t-1} \tag{4.8}$$

then

$$h_i^{(m)}(t|\mu_i(t)) = h_0^{(m)} exp(\boldsymbol{X}'_{s,i}\boldsymbol{\gamma}^{(m)} + \boldsymbol{\alpha}\left[\boldsymbol{X}'_{L,i}(t)\hat{\boldsymbol{\beta}}^{(m)} + \boldsymbol{Z}'_i(t)\hat{\boldsymbol{b}}_i^{(m)}\right] \tag{4.9}$$

for $i = 1, ..., n$, and $m = 1, ..., M$

The parameters (γ, α, h_0) are estimated from a frequentist perspective. $\boldsymbol{X}'_{L,i}(t)\hat{\boldsymbol{\beta}} + \boldsymbol{Z}'_i(t)\hat{\boldsymbol{b}}_i$ is the estimated longitudinal trajectory from the first stage of the joint modelling process using a Bayesian approach and it is linked to the parameter $\boldsymbol{\alpha}$, which measures the strength of relationship between the longitudinal response at time t and the hazard for an event at that time.

In the repeated measures submodel, $\boldsymbol{X}_i(t)$ and $\boldsymbol{Z}_i(t)$ represent vectors of potentially time-changing explanatory variables related to the p-dimensional vector of population effects $\boldsymbol{\beta}$ and the q-vector of subject-specific effects $\boldsymbol{b}_i$, respectively. The vector $\boldsymbol{X}'s, i$ in the survival submodel refers to baseline covariates linked to the coefficients vector $\boldsymbol{\gamma}$. The hazard function $h0(t) = \lambda^\tau t^{\tau-1}$ in the baseline function assumes the Weibull distribution.

Joint models offer the advantage of maintaining the same interpretation as the LMM for the repeated measurements data and the Weibull model for the event time submodel (Van, 2000; Hsieh, 2006). The parameter $exp(\gamma)$ in the joint modelling framework represents the proportional change in the hazard at time t when the corresponding longitudinal response $\mu_i(t)$ increases by one unit, taking into account the covariates $\boldsymbol{X}^T_{s,i}$.

4.3 Proposed Hybrid Standard Two-Stage Method For Longitudinal and Multi-State Models

Under this section, the proposed hybrid two-stage estimation procedure is extended from a single event of interest to a multiple of event of interest. Assume n subjects with longitudinal measurements and an individually associated time to a several events of interest. Particularly, the longitudinal process exerts an influence on the timing of multiple events process, and both processes are interconnected through shared random effects (Murawska et al. 2012). Still extending the idea of Tsiatis et al. (1995), the repeated measurements are fitted using the LMM, considering:

$$y_{hk,i}(t) = \boldsymbol{X}'_{L,i}(t)\boldsymbol{\beta} + \boldsymbol{Z}'_i(t)\boldsymbol{b}_i + \boldsymbol{e}_i(t) = \mu_{hk,i}(t) + \boldsymbol{e}_i(t) \tag{4.10}$$

$$\hat{y}_{hk,i}(b_i, t) = \boldsymbol{X}'_{L,i}(t)\hat{\boldsymbol{\beta}} + \boldsymbol{Z}'_i(t)\hat{\boldsymbol{b}}_i = \hat{\mu}_{hk,i}(b_i, t) \tag{4.11}$$

Here, $\hat{\mu}_{hk,i}(b_i, t)$, represents the calculated repeated measures outcome at time points t of the i^{th} individual where $\mu_i(t) = \{\mu_i(t_1), ..., \mu_i(t_j),\ t_j \leq t\}$.

For the proposed hybrid two-stage estimation method for multi-state process, $\hat{\mu}_{hk,i}(b_i, t)$, assumes the value of each of the posterior predictive estimates from the Bayesian LMM sub-model.

Hence, for each of the $m = 1,\ 2,\ 3,\ 4, ...,\ M$ posterior estimates, that is;

$$\hat{\boldsymbol{\beta}}^{(m)} = E(\boldsymbol{\beta}^m | y) \tag{4.12}$$

and

$$\hat{\boldsymbol{b}}^{(m)} = E(\boldsymbol{b}_i^m | y) \tag{4.13}$$

$\hat{\mu}_{hk,i}(b_i, t)$, is expressed as:

$$\hat{\mu}_{hk,i}(t) = \boldsymbol{X}'_{L,i}(t)\hat{\boldsymbol{\beta}}^m + \boldsymbol{Z}'_i(t)\hat{\boldsymbol{b}}_i^m \tag{4.14}$$

for $m = 1, 2, 3, ..., M$

In step two, this evolution function computed is plugged in as an endogenous time-dependent explanatory variable when estimating the multiple event submodel. Considering the estimated longitudinal measure , $\hat{\mu}_{hk,i}(t)$, as a covariate in fitting the multi-state model permits this study to extend the idea of Dafni and Tsiatis (1998), where the hazard function $h^i_{hk}(t|\mu_{hk,i}(t))$ in

$$h^i_{hk}(t|\mu_{hk,i}(t)) = h_{hk,0}(t)exp\left\{\boldsymbol{X}^S_{hk,i}(t)'\boldsymbol{\gamma}_{hk} + \boldsymbol{\alpha}_{hk}\mu_{hk,i}(t)'\right\} \tag{4.15}$$

is replaced by the function $\hat{h}^i_{hk}(t|\mu_{hk,i}(t))$. This results in:

$$\hat{h}^i_{hk}(t|\hat{\mu}_{hk,i}(t)) = h_{hk,0}(t)exp\left\{\boldsymbol{X}^S_{hk,i}(t)'\boldsymbol{\gamma}_{hk} + \boldsymbol{\alpha}_{hk}\hat{\mu}_{hk,i}(t)'\right\} \tag{4.16}$$

By substitution, the hazard becomes:

$$\hat{h}^i_{hk}(t|\hat{\mu}_{hk,i}(t)) = h^m_{hk,0}(t)exp\left\{\boldsymbol{X}^S_{hk,i}(t)'\boldsymbol{\gamma}^m_{hk} + \boldsymbol{\alpha}_{hk}\left(\boldsymbol{X}'_{L,i}(t)\hat{\boldsymbol{\beta}}^m + \boldsymbol{Z}'_i(t)\hat{\boldsymbol{b}}_i^m\right)\right\} \tag{4.17}$$

On the assumption of a Weibull distribution for the risk at baseline yields:

$$\hat{h}^i_{hk}(t|\hat{\mu}_{hk,i}(t)) = (\tau t^{t-1})^m_{hk}(t)exp\left\{\boldsymbol{X}^S_{hk,i}(t)'\boldsymbol{\gamma}^m_{hk} + \boldsymbol{\alpha}_{hk}\left(\boldsymbol{X}'_{L,i}(t)\hat{\boldsymbol{\beta}}^m + \boldsymbol{Z}'_i(t)\hat{\boldsymbol{b}}_i^m\right)\right\} \tag{4.18}$$

In estimating the parameters of MSM as well as the baseline hazard parameters the MLE method is employed.

The estimates of the multi-state parameters, $\boldsymbol{\gamma}$ and $\boldsymbol{\alpha}$, are attained by optimizing the log partial likelihood which is expressed as:

$$pl\left(\boldsymbol{\gamma}, \boldsymbol{\alpha}\right) =$$

$$\sum_{i=1}^{n} \int_{0}^{\infty} \left\{ R_i(t) \left\{ \boldsymbol{\gamma}^{'} \boldsymbol{\omega}_i + \boldsymbol{\alpha} \hat{\mu}_i(t) \right\} - log \left[\sum_{j} R_i(t) exp \left\{ \boldsymbol{\gamma}^{'} \boldsymbol{\omega}_i + \boldsymbol{\alpha} \hat{\mu}_i(t) \right\} \right] \right\} dN_i(t) \tag{4.19}$$

In this context, $N_i(t)$ represents the count of events observed for the i^{th} individual at time t, while $R_i(t)$ is an indicator function that takes the value 1 if the i^{th} individual is susceptible to the phenomenon at time t, and 0 otherwise. The vector $\boldsymbol{\omega}$ contains the explanatory variables for the multi-state submodel.

The estimates

$$(\boldsymbol{\gamma}, \boldsymbol{\alpha})$$

are then calculated from a MLE perspective.

One advantage of this method is that it is fast to execute as there are conventional LMM software employed for the step one and multi-state software also employed for step two of the model fitting process.

4.4 Estimating Measures of Interest for the Proposed Hybrid Model

In order to compare the new hybrid two-stage estimation method with existing joint models, several measures are calculated for the parameters of the suggested model. These measures include the bias, expectation, variance, and mean square error (MSE) for each parameter. This comparative assessment helps evaluate the performance and accuracy of the proposed approach.

4.4.1 Estimating Measures of Interest for the Fixed Effects Parameter,$\boldsymbol{\beta}$, for the Longitudinal Submodel

$$\hat{\boldsymbol{\beta}} = E(\boldsymbol{\beta}^m) = \frac{1}{M}\sum_{m=1}^{M}\boldsymbol{\beta}^m \tag{4.20}$$

$$Var(\hat{\boldsymbol{\beta}}) = \frac{1}{M-1}\sum_{m=1}^{M}(\hat{\boldsymbol{\beta}^m} - \boldsymbol{\beta})^2 \tag{4.21}$$

$$Bias(\boldsymbol{\beta}, \hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} - \hat{\boldsymbol{\beta}} \tag{4.22}$$

$$MSE(\hat{\boldsymbol{\beta}}) = Bias(\boldsymbol{\beta}, \hat{\boldsymbol{\beta}})^2 + Var(\hat{\boldsymbol{\beta}}) \tag{4.23}$$

4.4.2 Estimating Measures of Interest for the Fixed Effects Parameter, $\boldsymbol{\gamma}$, for the Survival Submodel

$$\hat{\boldsymbol{\gamma}} = E(\boldsymbol{\gamma}^m) = \frac{1}{M}\sum_{m=1}^{M}\boldsymbol{\gamma}^m \tag{4.24}$$

$$Var(\hat{\boldsymbol{\gamma}}) = \frac{1}{M-1}\sum_{m=1}^{M}(\hat{\boldsymbol{\gamma}}^m - \boldsymbol{\gamma})^2 \tag{4.25}$$

$$Bias(\boldsymbol{\gamma}, \hat{\boldsymbol{\gamma}}) = \boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}} \tag{4.26}$$

$$MSE(\hat{\boldsymbol{\gamma}}) = Bias(\boldsymbol{\gamma}, \hat{\boldsymbol{\gamma}})^2 + Var(\hat{\boldsymbol{\gamma}}) \tag{4.27}$$

4.4.3 Estimating Measures of Interest for the Association Parameter, $\boldsymbol{\alpha}$

$$\hat{\alpha} = E(\boldsymbol{\alpha}^{(m)}) = \frac{1}{M}\sum_{m=1}^{M}\boldsymbol{\alpha}^m \tag{4.28}$$

$$Var(\hat{\boldsymbol{\alpha}}) = \frac{1}{M-1}\sum_{m=1}^{M}(\hat{\boldsymbol{\alpha}}^m - \boldsymbol{\alpha})^2 \tag{4.29}$$

$$Bias(\boldsymbol{\alpha}, \hat{\boldsymbol{\alpha}}) = \boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}} \tag{4.30}$$

$$MSE(\hat{\boldsymbol{\alpha}}) = Bias(\boldsymbol{\alpha}, \hat{\boldsymbol{\alpha}})^2 + Var(\hat{\boldsymbol{\alpha}}) \tag{4.31}$$

4.5 Prior Distributions

The initial step of the suggested hybrid two-step estimation procedure involves computing the posterior distributions of the parameters and random effects for the repeated measurements submodel. To achieve this, prior distributions are specified for all unknown parameters in $\boldsymbol{\theta} = (\beta, b, \boldsymbol{\Sigma}, \sigma^2)$, and the full likelihood function is specified. Proper distributions with large variances are used for all parameters in $\boldsymbol{\theta}$. Precisely, the assumption is that, the prior distributions for parameter β and random error e_i are independent and distributed normally with mean 0 and variance 100. The covariance matrix $\boldsymbol{\Sigma}$ is left unspecified, using the default distribution, and declared as a covariance matrix to ensure it remains positive-definite by discarding samples that would not produce positive-definite matrices Σ.

4.6 Posterior Distributions

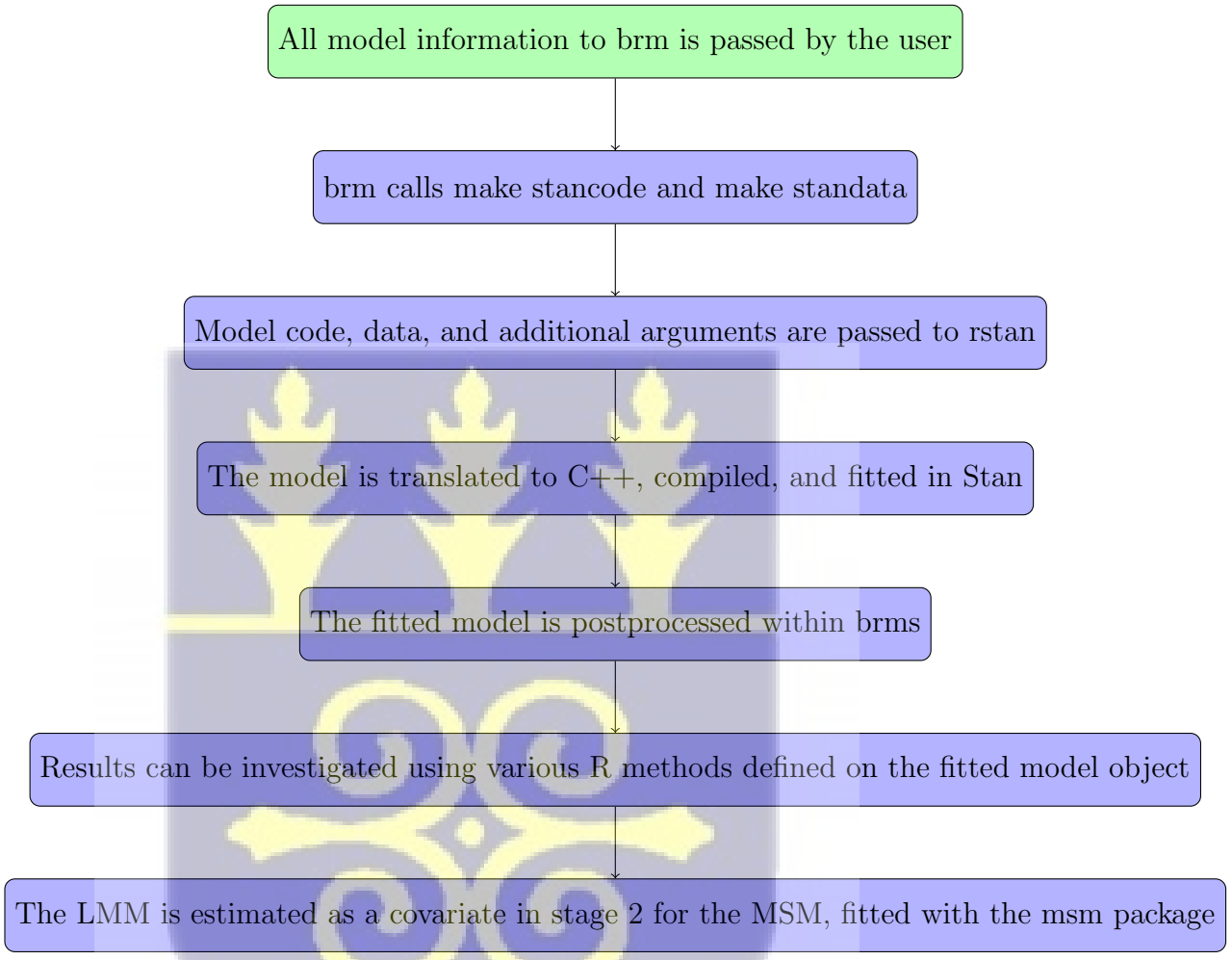
Bayesian techniques rely on MCMC posterior sampling to infer the vector $\boldsymbol{\theta}$, which contains the unknown parameters. In contrast to likelihood-based estimation that provides a single point estimate of the parameters along with asymptotic standard errors, MCMC algorithms can accurately estimate the full posterior distributions of the parameters. The MCMC algorithm is employed to sample from the full conditional distribution of each unknown parameter, enabling the generation of posterior samples. This process is facilitated through Stan, a probabilistic programming language specifically designed for statistical inference. To ensure convergence of the Markov chain, diagnostic tools such as trace plots are used. The absence of obvious patterns in these plots is considered as evidence of convergence, indicating that the MCMC algorithm has sufficiently explored the posterior distribution.

Moreover, the Gelman-Rubin test is employed to guarantee that that the scale reduction factor (Rhat) for all parameters is below 1.1, indicating good

convergence of the MCMC chains. The choice of MCMC over the Metropolis-Hastings algorithm is motivated by its ability to reduce the correlation between consecutive sampled states through a Hamiltonian evolution and by focusing on states that meet a higher acceptance criterion, the algorithm achieves faster convergence to the target distribution. Additionally, when dealing with large datasets, Stan demonstrates greater efficiency compared to the BUGS language, leading to quicker convergence, and requires relatively fewer samples to obtain accurate estimates (Carpenter et al. 2015).

4.7 Implementation of the Hybrid Two-Stage Model

The diagram below details the conceptual framework in implementing the hybrid two-stage model.



CHAPTER 5

SIMULATION STUDY AND RESULTS

5.1 Introduction

5.1.1 Simulation Study

To assess whether the proposed hybrid two-step estimation method decreases the bias, enhances estimation precision, as well as reduce the computational time, a simulation study is conducted to compare it with established models: the Rizopoulos joint approach (RJA), Bayesian joint specification (BJS), and Bayesian standard two-stage (BSTS) approach. The joint model employed in the simulation study is hinged on the longitudinal and survival submodels specified in chapter five of this thesis.

Specifically, the longitudinal outcome for i^{th} subject at time t is expressed as:

$$y_i(t) = \mu_i(t) + e_i(t) = \beta_0 + b_{0i} + \beta_1 x_{1i} + \beta_2 x_{2i} + (\beta_3 + b_{1i})t + e_i(t) \tag{5.1}$$

where;

$$\boldsymbol{b}_i = (b_{0i}, b_{1i})' \overset{iid}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma})$$

and

$$e_i(t) \overset{iid}{\sim} N(0, \sigma^2)$$

This thesis adopted the following hazard specification for individual i:

$$h_i(t) = h_0(t) exp\left\{\gamma_1 x_{3i} + \alpha \mu_i(t)\right\} \tag{5.2}$$

The fixed effect covariate x_{1i} and the group parameter, x_{3i}, are continuous explanatory variables simulated from a normal distribution with a mean of 74.19 and a standard deviation of 2.1. Also the fixed effect covariate, x_{2i}, is a categorical explanatory variable simulated from a Binomial distribution specified as Binom$(1, 0.5)$. Here the baseline hazard is specified from the exponential family of distributions. Specifically,

$$h_0(t) = exp(\gamma_0) \tag{5.3}$$

5.1.2 Data Simulation for the Joint Models

As an initial requirement in the simulation process, all the model parameters and hyperparameters, denoted by $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\Sigma}, \sigma, \boldsymbol{\gamma}, \alpha)$, along with the number of subjects (n), minimum count of longitudinal outcomes (m_{min}), and optimum follow-up time (t_{max}) must be specified (Huong et al. 2018; Lazaro et al. 2020). The explanatory variable x_i and the random effects b_i for subjects $i = 1, ..., n$ are then generated through simulation. The actual time of event for each subject, denoted by T_i, is simulated employing the inverse transform sampling method (Crowther and Lambert, 2013), where $T_i = S_i(u)$, with u being generated from a uniform distribution, and S_i representing the survival function obtained from the survival submodel specified in equation 6.2. Additionally, for each participant, the censoring time C_i is simulated from a uniform distribution over the interval $(0, t_{max})$ (Lee and Go, 1997), and then the observed time is defined as:

$$T_i = min\left\{T_i^*, C_i\right\}. \tag{5.4}$$

While the indicator event is expressed as:

$$\delta_i = I\left(T_i^* \leq C_i\right) \tag{5.5}$$

The count of repeated measurements of subject, i, denoted as n_i, is determined as m_{min} plus the greatest whole number less than T_i. The follow-up periods of the repeated observations are evenly spaced from 0 to T_i. Random errors $e_i(t_1), ..., e_i(t_{n_i})$ are generated from a normal distribution with a mean of zero and variance σ^2. Finally, the repeated measurements of subject i, $y_i(t_1), ..., y_i(t_{n_i})$, are computed based on the longitudinal submodel specified in equation (6.1). The simulation procedure to generate the longitudinal and survival data together is outlined in Table 6.1.

Table 5.1: Simulation Procedure.

0	INITIALISATION: Set $\boldsymbol{\theta}$, n, m_{min}, and t_{max}.
1	LONGITUDINAL SUBMODEL:
	$Set\ n_i = m_{min} + \lfloor T_i \rfloor \forall_i$.
	$Set\ t_1, ..., t_{n_i} = \lfloor T_i \rfloor \forall_i$ equispaced.
	$Simulate\ e_i(t) \sim N(0, \sigma^2), t_1, ..., t_{n_i} = \lfloor T_i \rfloor \forall_i$.
	$b_i \sim N(\mathbf{0}, \boldsymbol{\Sigma})\ \forall\ i$.
	Simulate $x_{1i} \sim norm(\mu, \sigma^2)$; Simulate $x_{2i} \sim Binom(1, 0.5)$.
	$Compute\ y_i(t_1), ..., y_i(t_{n_i})\ \forall_i$ based on the longitudinal submodel.
2	SURVIVAL SUBMODEL:
	Simulate $x_{3i} \sim norm(\mu, \sigma^2)$.
	Simulate T_i^* based on the survival submodel
	and sample $C_i \sim U(0, t_{max}) \forall i$.
	Set $T_i = min\{T_i^*, C_i\}$ and $\delta_i = I\{T_i^*, C_i\}\ \forall\ i$.

5.1.3 Scenarios

This section presents scenarios of the simulation schemes generated using the ADNI dataset, which is a publicly accessible data and has been vividly described in chapter three of this thesis. First, the joint models for submodels 6.1 and 6.2 were fitted to the ADNI data employing the "jointModel" function available in the R software's JM package. Then, the estimates that resulted from the combined model are utilized as the "true parameter" values during the generation

of simulated data.

The jointly estimated parameters are $\hat{b_0} = 9.3928$, $\hat{b_1} = -0.1270$, $\hat{b_2} = -0.0574$, $\hat{\sigma^2} = 1.5909$, $\hat{\Sigma_{11}} = 4.1747$, $\hat{\Sigma_{22}} = 0.04635$, and $\hat{\Sigma_{12}} = 0.1593 = \hat{\Sigma_{21}}$ for the longitudinal submodel (7.1) and $\hat{\gamma_0} = -6.9188$, $\hat{\gamma_1} = 0.0369$, and $\hat{\alpha} = -0.2818$ for the survival submodel 7.2. Finally, 100 datasets are simulated with $n = 200,\ 500,\ 1000$, $m_{min} = 3$, and $t_{max} = 15$. Although other set-ups were simulated, for the purpose of the limited space in this thesis, the results of the above scenarios are presented.

The MCMC configuration for both the Bayesian joint model and Bayesian standard two-stage model involved 2000 iterations with a burn-in of 1000 for the joint model and the longitudinal submodel in the two-step approach. Also, 1000 iterations with a burn-in of 500 were used to run the event time submodel in the two-step approach. For the proposed hybrid model, 4000 iterations with a burn-in of 2000 were specified. The Rizopollous joint model was estimated employing the "jointModel" function from the JM package in R, while all the Bayesian models were implemented using Stan. The simulations were performed on an 11th Generation HP laptop with an Intel Core i5 processor (2.40GHz) and 16 GB RAM, running on the Windows operating system.

5.1.4 Results

Tables 6.2, 6.4 and 6.6 present the comparative results for the longitudinal sub-models, among the models considered in this thesis, that is, the RJA, BJS, BSTS, and proposed hybrid standard two-step (HSTS) procedures for 100 simulated datasets from the joint models of submodels 6.1 and 6.2 for sample size, $N = 200,\ 500$ and 1000 respectively, using the parameters set above.

It is evident from Tables 6.2, 6.4 and 6.6 that, when the sample size was relatively small, precisely at, $N = 200$, all the four models satisfactorily estimated the longitudinal submodel parameters well, however, the proposed hybrid model outperformed all the other three models in terms of yielding relatively less bias

estimates and at par with the Rizopollous joint model in terms of yielding precise estimates. At higher sample sizes, $N = 1000$, the BJS and the BSTS yield relatively very bias estimates especially for the intercept while the RJA and the HSTS estimated all the model parameters relatively very well. In terms of yielding precise estimates for the longitudinal submodel, all four models performed relatively well with the exception of the BJS and the BSTS which yielded estimates with relatively high mean square error for the model intercept. In sum, in terms of the submodel for the longitudinal process, the results from the proposed hybrid model are better than the BJA and the BSTS methods and similar to the RJA, which theoretically has been suggested to be the ideal choice to handle these kind of estimation processes.

The comparative results for the event time submodel, among the models considered in this thesis, that is, the RJA, BJS, BSTS and proposed HSTS methods for 100 simulated datasets from the joint models of submodels 6.1 and 6.2 for sample size, $N = 200, 500$ and 1000, are presented in Tables 6.2, 6.4 and 6.6 respectively, using the parameters set above.

It can be seen, from Table 6.3, 6.5 and 6.7, that the group parameter (γ) is satisfactorily estimated using all four methods in terms of precision and bias. On the other hand, in all scenarios, the proposed hybrid model competed fairly with the other established approaches in terms of estimating the association parameter (α). The results of the hybrid model outperformed the BSTS approach, in terms of precision at larger sample sizes and are similar to the RJA and BJS, which theoretically are the preferred ways to address the estimation process as the information sharing among the submodel is deemed apex compared to the two-stage approach to joint modelling.

However, surprisingly, the bias of the posterior distributions using the proposed hybrid methodology is marginally lower than other methods at larger sample sizes. Furthermore, the computational time of the hybrid model, especially when the posterior predictive estimates at the tail end of the converged longitudinal

submodel are used, is far less than all the other approaches. It is crucial to emphasize that in theory, the full joint likelihood approaches are always preferred ones. The other procedures are opted for when the complex nature of the joint model renders the inferential process consuming too much of time or when as a result of high-dimensional parameter space the Markov chains encounter challenges of convergence.

Table 5.2: Simulation results for the longitudinal sub-model from 100 simulated datasets for N=200 comparing the results from each estimation approach.

Measure	Para	True Value	JM Frequentist/ Rizopoulos N = 200	JM Bayesian Joint Likelihood N=200	JM Bayesian STS N=200	Proposed JM Hybrid STS N=200
Estimate						
	β_0	9.3928	10.2849	11.7362	12.2961	9.9866
	β_1	-0.1270	-0.1378	-0.0266	-0.0048	-0.1344
	β_2	-0.0574	-0.1097	-0.1609	-0.1687	-0.0530
Std. Error						
	β_0		5.0723	5.0328	5.2508	5.1592
	β_1		0.0687	0.0264	0.0266	0.0698
	β_2		0.0300	0.0678	0.0707	0.0286
95% CI						
Lower CI	β_0		0.3432	1.8719	2.0045	-0.1254
Upper CI	β_0		20.2266	21.6004	22.5877	20.0986
Lower CI	β_1		-0.2725	-0.0783	-0.0569	-0.2712
Upper CI	β_1		-0.0031	0.0251	0.0473	0.0024
Lower CI	β_2		-0.1685	-0.2938	-0.3072	-0.1091
Upper CI	β_2		-0.0509	-0.0280	-0.0301	0.0031
Confidence Length						
	β_0		19.8834	19.7285	20.5832	20.2240
	β_1		0.2694	0.1034	0.1042	0.2736
	β_2		0.1176	0.2658	0.2771	0.1122
Bias						
	β_0		-0.8921	-2.3434	-2.9033	-0.5938
	β_1		0.0108	-0.1004	-0.1222	0.0074
	β_2		0.0523	0.1035	0.1113	-0.0044
Mean Square Error						
	β_0		26.5241	30.8210	36.0001	26.9699
	β_1		0.0048	0.0108	0.0156	0.0049
	β_2		0.0036	0.0153	0.0174	0.0008

Table 5.3: Simulation results for the survival sub-model from 100 simulated datasets for N=200 comparing the results from each estimation approach.

Measure	Para	True Value	JM Frequentist/ Rizopoulos N = 200	JM Bayesian Joint Likelihood N=200	JM Bayesian STS N=200	Proposed JM Hybrid STS N=200
Estimate						
	γ	0.0369	0.0126	0.0412	0.0234	0.0686
	α	-0.2818	-0.2460	-0.3129	-0.2948	-0.2191
Std. Error						
	γ		0.0657	0.0589	0.0598	0.0631
	α		0.0569	0.0574	0.0564	0.0359
95% Confidence Interval (CI)						
Lower CI	γ		-0.1162	-0.0742	-0.0938	-0.0551
Upper CI	γ		0.1414	0.1566	0.1406	0.1923
Lower CI	α		-0.3575	-0.4254	-0.4053	-0.2895
Upper CI	α		-0.1345	-0.2004	-0.1843	-0.1487
95% Confidence Length						
	γ		0.2576	0.2308	0.2344	0.2474
	α		0.2230	0.2250	0.2210	0.1408
Bias						
	γ		0.0243	-0.0043	0.0135	-0.0317
	α		-0.0358	0.0311	0.0130	-0.0627
Mean Square Error						
	γ		0.0049	0.0035	0.0038	0.0050
	α		0.0045	0.0043	0.0033	0.0052

Table 5.4: Simulation results for the longitudinal sub-model from 100 simulated datasets for N=500 comparing the results from each estimation approach.

Measure	Para	True Value	JM Frequentist/ Rizopoulos N = 500	JM Bayesian Joint Likelihood N=500	JM Bayesian STS N=500	Proposed JM Hybrid STS N=500
Estimate						
	β_0	9.3928	9.6613	11.8247	11.7577	11.7262
	β_1	-0.1270	-0.1303	-0.0602	-0.0339	-0.1593
	β_2	-0.0574	-0.0572	-0.1599	-0.1594	-0.0683
Std. Error						
	β_0		3.1109	3.5431	3.2676	3.5244
	β_1		0.0419	0.0179	0.0173	0.0474
	β_2		0.0166	0.0478	0.0441	0.0160
95% Confidence Interval (CI)						
Lower CI	β_0		3.5639	4.8802	5.3532	4.9723
Upper CI	β_0		15.7587	18.7692	18.1622	18.7127
Lower CI	β_1		-0.2124	-0.0953	-0.0678	-0.2530
Upper CI	β_1		-0.0482	-0.0251	0.0000	-0.0677
Lower CI	β_2		-0.0897	-0.2536	-0.2458	-0.1004
Upper CI	β_2		-0.0247	-0.0662	-0.0730	-0.0377
95% Confidence Length						
	β_0		12.1948	13.8890	12.8090	13.7403
	β_1		0.1642	0.0702	0.0678	0.1853
	β_2		0.0650	0.1874	0.1728	0.0627
Bias						
	β_0		-0.2685	-2.4319	-2.3649	-2.3334
	β_1		0.0033	-0.0668	-0.0931	0.0323
	β_2		-0.0001	0.1025	0.1020	0.0109
Mean Square Error						
	β_0		9.7495	18.4677	16.2700	17.8662
	β_1		0.0018	0.0048	0.0090	0.0033
	β_2		0.0003	0.0128	0.0123	0.0004

Table 5.5: Simulation results for the survival sub-model from 100 simulated datasets for N=500 comparing the results from each estimation approach.

Measure	Para	True Value	JM Frequentist/ Rizopoulos N = 500	JM Bayesian Joint Likelihood N=500	JM Bayesian STS N=500	Proposed JM Hybrid STS N=500
Estimate						
	γ	0.0369	0.0429	0.0225	0.0240	0.0509
	α	-0.2818	-0.2821	-0.3075	-0.2906	-0.3178
Std. Error						
	γ		0.0410	0.0392	0.0381	0.0375
	α		0.0340	0.0365	0.0354	0.0224
95% Confidence Interval						
Lower CI	γ		-0.0375	-0.0543	-0.0507	-0.0226
Upper CI	γ		0.1233	0.0993	0.0987	0.1244
Lower CI	α		-0.3487	-0.3790	-0.3600	-0.3617
Upper CI	α		-0.2155	-0.2360	-0.2212	-0.2739
95% Confidence Length						
	γ		0.1608	0.1536	0.1494	0.1470
	α		0.1332	0.1430	0.1388	0.0878
Bias						
	γ		-0.0061	0.0144	0.0129	-0.0140
	α		0.0003	0.0257	0.0088	0.0360
Mean Square Error						
	γ		0.0017	0.0017	0.0016	0.0016
	α		0.0012	0.0020	0.0013	0.0018

Table 5.6: Simulation results for the longitudinal sub-model from 100 simulated datasets for N=1000 comparing the results from each estimation approach.

Measure	Para	True Value	JM Frequentist/ Rizopolous N = 1000	JM Bayesian Joint Likelihood N= 1000	JM Bayesian STS N=1000	Proposed JM Hybrid STS N=1000
Estimate						
	β_0	9.3928	9.2250	5.8691	5.9192	8.0485
	β_1	-0.1270	-0.1247	-0.0631	-0.0437	-0.1114
	β_2	-0.0574	-0.0567	-0.0784	-0.0796	-0.0455
Std. Error						
	β_0		2.4367	2.4558	2.2455	2.5610
	β_1		0.0327	0.0123	0.0120	0.0346
	β_2		0.0136	0.0331	0.0303	0.0128
95% Confidence Interval (CI)						
Lower CI	β_0		4.4491	1.1037	1.5645	3.1686
Upper CI	β_0		14.0009	10.6955	10.3931	12.9678
Lower CI	β_1		-0.1888	-0.0872	-0.0655	-0.1779
Upper CI	β_1		-0.0606	-0.0403	-0.0200	-0.0451
Lower CI	β_2		-0.0834	-0.1433	-0.1399	-0.0728
Upper CI	β_2		-0.0300	-0.0143	-0.0211	-0.0207
95% Confidence Length						
	β_0		9.5518	9.5918	8.8286	9.7992
	β_1		0.1282	0.0469	0.0455	0.1327
	β_2		0.0534	0.1290	0.1188	0.0521
Bias						
	β_0		0.1678	3.5237	3.4736	1.3443
	β_1		-0.0023	-0.0639	-0.0833	-0.0156
	β_2		-0.0007	0.0210	0.0222	-0.0118
Mean Square Error						
	β_0		5.9657	18.4474	17.1082	8.3659
	β_1		0.0011	0.0042	0.0071	0.0014
	β_2		0.0002	0.0015	0.0014	0.0003

Table 5.7: Simulation results for the survival sub-model from 100 simulated datasets for N=1000 comparing the results from each estimation approach.

Measure	Para	True Value	JM Frequentist/ Rizopoulos N = 1000	JM Bayesian Joint Likelihood N= 1000	JM Bayesian STS N=1000	Proposed JM Hybrid STS N=1000
Estimate						
	γ	0.0369	0.0428	0.0313	0.0325	0.0342
	α	-0.2818	-0.2808	-0.2968	-0.2813	-0.2982
Std. Error						
	γ		0.0288	0.0271	0.0268	0.0289
	α		0.0232	0.0250	0.0244	0.0261
95% Confidence Interval (CI)						
Lower CI	γ		-0.0136	-0.0218	-0.0200	-0.0224
Upper CI	γ		0.0992	0.0844	0.0850	0.0908
Lower CI	α		-0.3263	-0.3458	-0.3291	-0.3494
Upper CI	α		-0.2353	-0.2478	-0.2335	-0.2470
95% Confidence Length						
	γ		0.1128	0.1062	0.1050	0.1133
	α		0.0910	0.0980	0.0956	0.1024
Bias						
	γ		-0.0059	0.0056	0.0044	0.0027
	α		-0.0010	0.0150	-0.0005	0.0164
Mean Square Error						
	γ		0.0009	0.0008	0.0007	0.0008
	α		0.0005	0.0009	0.0006	0.0010

Table 6.8 below displays the time for each estimation approach to converge for the 100 simulated datasets when was the sample size was specified to be 200. It is evident from the table that the proposed model achieved a competitive converging time in relation to the other models.

Table 5.8: Average computational time(Hours) for the 100 simulated datasets for N=200 for each estimation approach to converge.

Measure	JM Frequentist/ Rizopolous N = 200	JM Bayesian Joint Likelihood N= 200	JM Bayesian STS N=200	Proposed JM Hybrid STS N=200
Average Computational Time (Hours)	1.50	7.50	5.25	2.25

5.2 Summary of Chapter

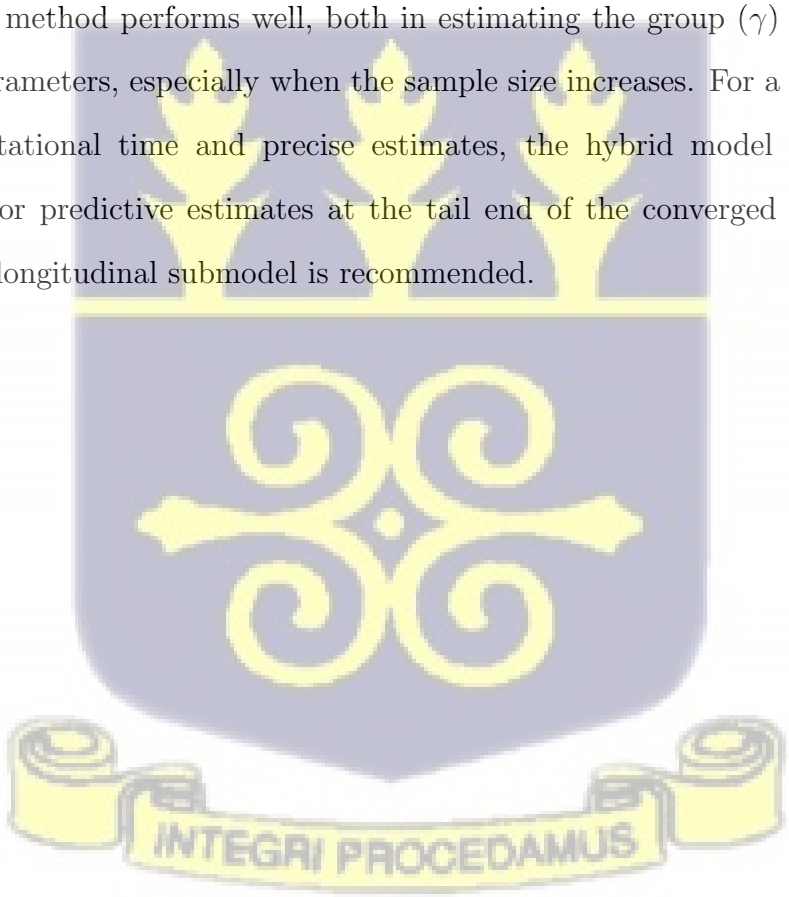
This chapter presented the simulation results when the proposed hybrid model was extended to time to a single event. The frequentist models were estimated employing the function, jointModel, from the JM package in R statistical software while all the Bayesian models were implemented using stan. It was evident from the simulation study that, the suggested hybrid model outperformed all the other approaches in fitting the longitudinal submodel in terms of yielding relatively less bias and precise estimates at smaller sample sizes. At larger sample sizes, N > 500, the hybrid model, again, outperformed the BJS and the BSTS, in estimating the longitudianl submodel parameters, but had similar results to the RJA, which in literature has been acclaimed to be the recommended way to deal with the joint estimation process (Mauff et al. 2020).

Regarding the event time submodel, the hybrid model's performance in estimating the group (γ) and association (α) parameters, was significantly enhanced at larger sample sizes. On the score of 95% CI , 95% CL, bias and the mean square

error the proposed hybrid model performed competitively well in relation to the other methods that has been touted in the literature of joint models and in some instances outperformed these existing methods. The results of the hybrid model outperformed the BSTS approach, in terms of precision at larger sample sizes and similar to the RJA and BJS.

In terms of computational time, the proposed hybrid model performed competitively well in relation to the RJA and outperformed the BJS and the BSTS models.

Surprisingly, the bias of the posterior distributions using the proposed hybrid methodology is marginally lower than other methods at larger sample sizes. It can be conveniently concluded that, the simulation study reveals that the suggested hybrid method performs well, both in estimating the group (γ) and association (α) parameters, especially when the sample size increases. For a relatively faster computational time and precise estimates, the hybrid model which uses the posterior predictive estimates at the tail end of the converged Bayesian LMM of the longitudinal submodel is recommended.



CHAPTER 6

EMPIRICAL APPLICATION

6.1 Empirical Application of the Hybrid Two-Stage Joint Longitudinal and Multi-State Models

6.1.1 Introduction

The hybrid two-stage approach is utilized on the ADNI data to assess the link between dynamics of ADNI and intensities of transition among disease states and also times of transitions. In this joint hybrid model, the relation between longitudinal trajectory and the transitions between three states of AD are modeled via a Bayesian LMM at the first stage and a frequentist continuous-time MSM at the second stage. In general, the joint model under consideration assumes that, at the second stage, subject-specific transitions among states are independent and assume a continuous homogeneous time stochastic process. Since detailed computational descriptions of both sub-models have been provided in previous chapters, a brief mathematical description of these sub-models is provided. In this proposed method, the longitudinal model is initially fitted independently, and the resulting fitted values, denoted as $\hat{\mu}_i(t)$, are then used as a covariate in the subsequent fitting of the multi-state submodel.

6.1.2 Data Description

The ADNI is a longitudinal clinical trial aimed at examining the dynamics of dementia over its progression defined on four states of the disease, that is,

cognitive normal (CN), EMCI, and LMCI to dementia or AD. For the purpose of this thesis, EMCI, and LMCI were merged as a single disease state, and labelled mild cognitive impairment (MCI) due to the relatively low frequency recorded for these two states. The criteria for including and excluding study participants, follow-up periods of evaluations as well as other relevant information can be located on http://adni.loni.usc.edu/. mPACCtrailsB is the assessment tool this study focuses on. Hence mPACCtrailsB becomes the dependent variable in the longitudinal sub-model. The selected covariates this thesis considered are patient age, patient gender, years followed from baseline, diagnostic group at baseline and study end and disease states as CN, MCI and AD. These covariates are utilized to investigate the impact of different factors on the transitions between various disease states in patients. Apart from the movement among disease states, the period (in years) taken for a subject to move from one disease state to another state as well other interesting measures are also computed.

6.1.3 Stage 1: Fitting the Longitudinal Sub-Model

The univariate LMM is fitted using the brms package. Step one of the joint model formulation has to do with fitting the longitudinal sub-model. Specifically in this study, the grouping variable is the Record Identification (RID). The RID ($J = 6688$) is indexed by the subscript, j, each participant has n_j observations. The number of observations per participant varies, with some having as few as 2 and others as many as 16, averaging around 9. The overall count of measurements for each participant in this study is the sum of the observations per participant. In this study, $N = 14119$ observations. The dependent variable of the model is a continuous variable, mPACCtrailsB, and has three fixed effects explanatory variables, age (in years), gender (0 = female, 1 = male), and years from baseline in addition to a fixed intercept and one random intercept and one random slope for each of the $J = 6688$ participants. All other effects are kept fixed at the moment. The inclusion of random effects is due to the anticipation of potential

correlations in mPACCtrailsB scores within participants. Several factors could contribute to such correlations.

For instance, observations from the same participant are more homogeneous than they are between participants. As described earlier, varied dependent structures can be employed to describe the relationship of the repeated measurements and the multi-state submodels. This study focuses on the current value predictor association structure described in Chapter four. The LMM for the n_i dimensional dependent variable y_i for participant i at time t for $m = 1, 2, 3, ..., M$ posterior predictive estimates is expressed as:

$$y_i(t)^{(m)} = \mu_i(t)^{(m)} + \boldsymbol{e}_i(t)^{(m)} = \boldsymbol{X}_{L,i}^T(t)\boldsymbol{\beta}_i^{(m)} + \boldsymbol{Z}_i^T(t)\boldsymbol{b}_i^{(m)} + \boldsymbol{e}_i(t)^{(m)} \tag{6.1}$$

$$\hat{\mu_i(t)}^{(m)} = \boldsymbol{X}_{L,i}^T(t)\hat{\boldsymbol{\beta}}_i^{(m)} + \boldsymbol{Z}_i^T(t)\hat{\boldsymbol{b}}_i^{(m)} \tag{6.2}$$

The mixed model for the n_i dimensional outcome $mPACCtrailsB_i$ for participant i at time t then becomes:

$$mPACCtrailsB_i(t)^{(m)} =$$

$$\beta_0^{(m)} + b_{0i}^{(m)} + \text{Age}\beta_1^{(m)} + \text{Gender}\beta_2^{(m)} + \text{Years.bl}\ \beta_3^{(m)} + \text{Years.bl}\ b_{1i}^{(m)} + e_i(t)^{(m)} \tag{6.3}$$

$$mPACC\hat{C}trailsB_i(t)^{(m)} =$$

$$\beta_0^{(m)} + b_{0i}^{(m)} + \text{Age}\beta_1^{(m)} + \text{Gender}\beta_2^{(m)} + \text{Years.bl}\beta_3^{(m)} + \text{Years.bl}\ b_{1i}^{(m)} \tag{6.4}$$

This implies,

$$\hat{\mu}_i(t)^{(m)} = mPACC\hat{C}trailsB_i(t)^{(m)} \tag{6.5}$$

Where;

$$\boldsymbol{b}_i^{(m)} \overset{iid}{\sim} N(\boldsymbol{0}, \boldsymbol{\Sigma})$$

and

$$\boldsymbol{e}_i(t)^{(m)} \overset{iid}{\sim} N(0, \sigma^2)$$

$mPACCtrailsB_i(t)$ is a 14119 $\times$ 1 matrix; β_0 is a 14119 $\times$ 1 matrix; $X_{L,i}^T(t)$ 14119 *by* 3 matrix; β_i is a 3×1 matrix; b_{0i} is a 6688×1 matrix; b_{01} is a 6688×1 matrix; $Z_i^T(t)$ is a 14119×6688 matrix. $y_i(t)$ represents a $N \times 1$ matrix, which serves as the explanatory variable, and X is an $N \times p$ matrix consisting of p variables. β is a $p \times 1$ matrix containing the fixed-effects regression coefficients (the β's).

The matrix Z is an $N \times qJ$ design matrix that corresponds to the q random effects for each of the J groups. Additionally, b is a $qJ \times 1$ vector comprising the random effects (the random complement to the fixed β) for the J groups. Lastly, e is a $N \times 1$ column vector representing the residuals, that is, the unexplained part of $mPACCtrailsB$, after considering the submodel contributions $X_{L,i}^T(t)\beta_i + Z_i^T(t)b_i$.

The novelty of the hybrid estimation procedure proposed in this thesis picks inspiration from extending the existing current value predictor association structure for joint models. In the existing current value structure, $\hat{\mu}_i(t)$, is estimated using either the point estimate of the model parameters at time t from the frequentist perspective or an average of the posterior distribution at time t from the Bayesian perspective. In estimating, $\hat{\mu}_i(t)$, for the proposed hybrid model, each value of the posterior predictive estimate of the fixed and random effects from the first stage of the joint estimation process was used. That is, for the number of iterations specified for the Bayesian LMM, all the posterior estimates are plugged in turns into the longitudinal sub-model to estimate $\hat{\mu}_i(t)$. Meaning all 4000 posterior estimates corresponding to all the 4000 iterations specified in fitting the Bayesian longitudinal sub-model are utilized in estimating $\hat{\mu}_i(t)$.

6.1.4 Stage 2: Fitting the Multi-State Model

Reference the proposed MSM in Chapter five where each of the posterior predictive estimates from the Bayesian LMM where used in estimating the MSM

as:

$$h^i_{hk}(t|\mu_{hk,i}) = h_{hk,0}(t)exp\left\{\gamma_{hk}X^S_{hk,i}(t)' + \alpha_{hk}\hat{\mu}_{hk,i}(t)'\right\} \tag{6.6}$$

Specifically, the study includes age and the estimated longitudinal trajectory as covariates. The multi-state transition models for the modelled states are thus:

$$h^i_{cn,mci}(t|\mu_{cn,mci,i}) = h_{cn,mci,0}(t)exp\left\{AGE_{cn,mci,i}(t)\gamma_{cn,mci} + \alpha_{cn,mci}\left(\hat{\mu}_{cn,mci,i}(t)\right)\right\} \tag{6.7}$$

$$h^i_{mci,cn}(t|\mu_{mci,cn,i}) = h_{mci,cn,0}(t)exp\left\{AGE_{mci,cn,i}(t)\gamma_{mci,cn} + \alpha_{mci,cn}\left(\hat{\mu}_{mci,cn,i}(t)\right)\right\} \tag{6.8}$$

$$h^i_{mci,ad}(t|\mu_{mci,ad,i}) = h_{mci,ad,0}(t)exp\left\{AGE_{mci,ad,i}(t)\gamma_{mci,ad} + \alpha_{mci,ad}\left(\hat{\mu}_{mci,ad,i}(t)\right)\right\} \tag{6.9}$$

$$h^i_{ad,mci}(t|\mu_{ad,mci,i}) = h_{ad,mci,0}(t)exp\left\{AGE_{ad,mci,i}(t)\gamma_{ad,mci} + \alpha_{ad,mci}\left(\hat{\mu}_{ad,mci,i}(t)\right)\right\} \tag{6.10}$$

The MLE approach is adopted in estimating the parameters of the multi-state submodel. Detailed methodology on the estimation procedure has been dealt with in Chapter five. In the two-step estimation method, the parameters for longitudinal and multi-state submodels are calculated separately. In estimating the second stage model, all the posterior estimates of the first stage estimation are included. Phase two of the analysis of the joint model formulation is implemented in R software and the msm package is utilized to fit the model. The multistate data is prepared using appropriate functions in the msm package and the parameters of the transition specific multistate submodel are then estimated employing the coxph() function.

6.1.5 The Longitudinal Submodel: Linear Mixed Model (Bayesian Estimation)

The model was fitted with a total of 8000 ($4\ chains \times (4000 - 2000)$) samples acquired from the posterior distribution, derived from 4 chains. Each chain underwent 4000 iterations, but the first 2000 iterations were discarded

as warm-up. Credible intervals(CrI) that contain zero point to a true null hypothesis, meaning, intervals that do not contain zero depicts statistical significance. It can be seen from the estimated model that, age and gender were found to have statistical significance on mPACCtrailsB while years from baseline (95% CrI = [-0.09, 0.01]) was found to have no statistical effect on mPACCtrailsB. It can be seen from the model output that when all other covariates are zero, the average mPACCtrailsB, defined by the intercept, is 9.14. An increase in year of a person's age (Est.= -0.12, 95% CrI = [-0.15, -0.09]) averagely decreases mPACCtrailsB by 0.12. Relative to a female, a male (Est.= -0.79, 95% CrI = [-1.17, -0.43]) has an average decrease of 0.79 in mPACCtrailsB.

Table 6.1: Group-Level Effects for Model with mPACCtrailsB as Dependent Variable

Group-Level Effects	Est.	Est.Error	l-95% CI	U-95% CI	Rhat	Bulk ESS	Tail ESS
sd(Intercept)	2.00	0.08	1.86	2.16	1.00	9435	16028
sd(Years.baseline)	0.21	0.03	0.16	0.27	1.00	5969	11041
cor(Intercept,Years)	0.34	0.14	0.07	0.62	1.00	8967	11634

Table 6.2: Population-Level Effects for Model with mPACCtrailsB as Dependent Variable

Population-Level Effects	Est.	Est.Error	l-95% CI	U-95% CI	Rhat	Bulk ESS	Tail ESS
(Intercept)	9.14	1.16	6.87	11.41	1.00	9349	14196
Age	-0.12	0.02	-0.15	-0.09	1.00	9221	13870
Gender.Male	-0.79	0.19	-1.17	-0.43	1.00	9191	14067
Years.baseline	-0.04	0.02	-0.09	0.01	1.00	22786	19693

Table 6.3: Family Specific Parameters for Model with mPACCtrailsB as Dependent Variable

Family Specific Parameters	Est.	Est.Error	l-95% CI	U-95% CI	Rhat	Bulk ESS	Tail ESS
sigma	1.60	0.03	1.54	1.65	1.00	15021	19215

6.1.6 Multi-State Submodel

This section presents the output of the hybrid multi-state submodel of the two-stage joint model when applied to the ADNI data. To decide on the states to include in the MSM a summary of the multi-state data in a frequency table of pairs of consecutive states is tabulated in Table 7.21 below. From the table, 388 subjects transited from MCI state to AD state while only 6 persons transited from CN to AD. Majority of persons (3366 persons) were in the MCI state followed by 2605 person in the CN state while 1480 persons were in the AD state.

Generally, it can be observed from Table 7.21 that fewer subjects, n = 6, transitioned from CN state to the AD state and also no subject back transitioned from state AD to state CN. Transitions that had few subjects as such were not included in modelling the MSM because of the sample size for these transitions being small. Therefore, the hypothesized conceptual figure for the three state transition exhibited in Figure 7.12 decreases to Figure 7.13. In the section that follows, the study assesses the variables that are linked with movements from one state of AD to a more advanced state of the disease and the reverse of each of these movements. In addition to the hazard ratios for the for transitions between the various states other fascinating measures such as transition probability matrix, mean sojourn times, probability that each state is next and the ratio of transition intensities are also explored.

The model output starts with the baseline estimates of transition intensity attained from the continuous-time MSM. These estimates are presented in

Tables 7.22 and 7.23, showing both the unmodified estimates and those adjusted for covariates, along with their corresponding 95% confidence intervals. The multi-state sub-model that includes explanatory variables recorded an enhanced fit as compared to the sub-model without covariates as evident by a lesser $-2 \times log - likelihood$ estimate of 5018.768 as against 5069.938 for the model without covariates.

Table 6.4: Frequency Distribution of Number of Transitions Among States

	State 1(CN)	State 2(MCI)	State 3(AD)
To / From			
State 1(CN)	2605	159	6
State 2(MCI)	114	3366	388
State 3(AD)	0	28	1480

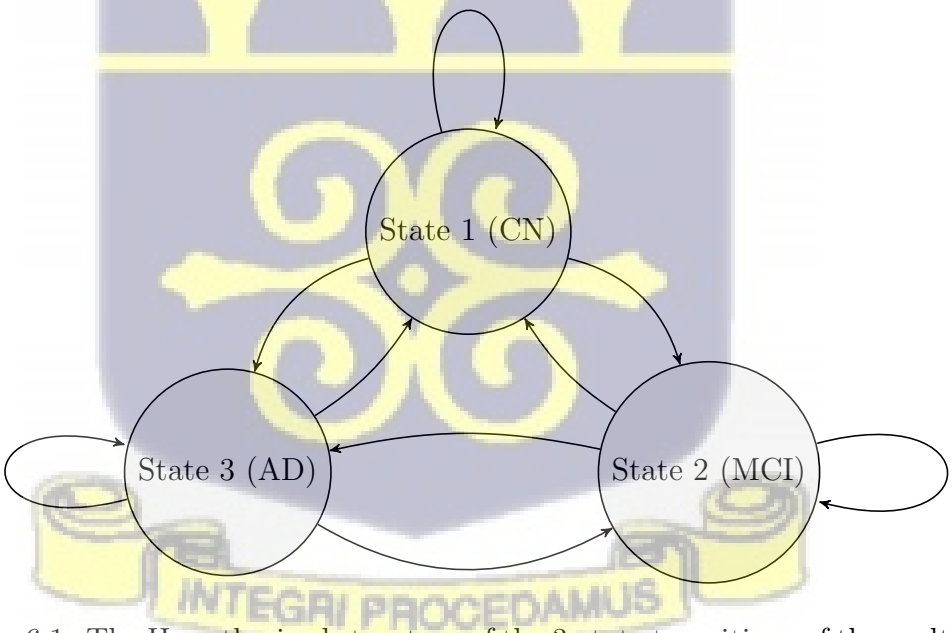


Figure 6.1: The Hypothesized structure of the 3-state transitions of the multistate data.

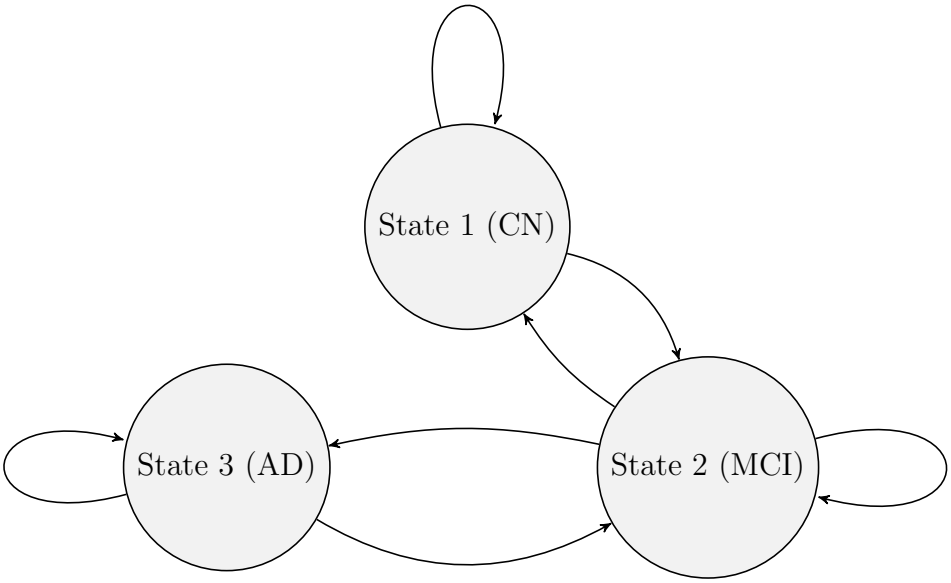


Figure 6.2: The structure of the fitted transition model

6.2 Specifying the transition intensity matrix

The likelihood is optimized using numerical methods, requiring an initial set of values to initiate the search for the maximum. The transition intensity matrix (T) is defined with the assumption of equal probabilities for progression and recovery. It also considers the structure of the fitted transition model in Figure 7.13 where transitions with relatively few counts were not included in the model estimation. The initial transition values were specified as:

$$T = \begin{bmatrix} 0.25 & 0.10 & 0 \\ 0.10 & 0.25 & 0.10 \\ 0 & 0.10 & 0.25 \end{bmatrix} \tag{6.11}$$

6.2.1 Hazard Ratios for Transition

Table 7.22 below shows the baseline transition intensities for the multi-state model with no covariate while Table 7.23 shows the baseline intensity ratio and

hazard ratios for the covariates age and the estimated longitudinal trajectory $\hat{\mu_{ij}}$ and these covariates are set to their means. It is evident from Table 7.23 that all the baseline transitions are statistically significant since none of the confidence intervals (CI's) contain one. It can be seen from Table 7.23 that age is only significant [$\gamma = 0.9411660$, CI(0.9176708 - 0.9652628)] for a subject transiting from MCI state to CN state. Also, from Table 7.23, it could be seen that the covariate $\hat{\mu_{ij}}$ is significant [$\gamma = 0.9456759$, CI(0.9107342 - 0.9819582)] for a person transiting from CN state to MCI state.

Table 6.5: Hazard ratios and corresponding 95% Confidence Interval from the Multi-state model without covariates.
State 1: Cognitive Normal, State 2: Mild Cognitive Impairment, State 3: Alzheimer Disease

Baseline			Estimate	l-95% CI	U-95% CI
State 1	-	State 1	-0.0592	-0.0689	-0.0508
State 1	-	State 2	0.0592	0.0508	0.0689
State 2	-	State 1	0.0404	0.0335	0.0486
State 2	-	State 2	-0.1749	-0.1909	-0.1603
State 2	-	State 3	0.1345	0.1219	0.1485
State 3	-	State 2	0.0241	0.0164	0.0354
State 3	-	State 3	-0.0241	-0.0354	-0.0164
-2 * log-likelihood: 5069.938					

Table 6.6: Hazard ratios and corresponding 95% Confidence Interval from the Multi-state model with covariates.
State 1: Cognitive Normal, State 2: Mild Cognitive Impairment, State 3: Alzheimer's Disease

Baseline			Estimate	l-95% CI	U-95% CI
State 1	-	State 1	-0.0540	-0.0639	-0.0456
State 1	-	State 2	0.0540	0.0456	0.0639
State 2	-	State 1	0.03534	0.0286	0.0437
State 2	-	State 2	-0.17189	-0.1881	-0.1570
State 2	-	State 3	0.1365	0.1235	0.1508
State 3	-	State 2	0.0240	0.0162	0.0356
State 3	-	State 3	0.0240	-0.0356	-0.0162
-2 * log-likelihood: 5018.768					
Age					
State 1	-	State 2	1.0209	0.9948	1.0477
State 2	-	State 1	0.9412	0.9177	0.9653
State 2	-	State 3	1.0238	1.0097	1.0380
State 3	-	State 2	0.9771	0.9271	1.0297
$\hat{\mu_{ij}}$					
State 1	-	State 2	0.9457	0.9107	0.9820
State 2	-	State 1	0.9846	0.9333	1.0386
State 2	-	State 3	1.0344	1.0021	1.0677
State 3	-	State 2	0.9565	0.8682	1.0538

Apart the from the ever popular hazard estimates for each covariate as well as baseline intensities, other interesting measures to consider when making inference from multi-state models are considered in the sections below.

6.2.2 Transition Probability Matrices

The transition probability matrix within a given time is estimated. The estimated ten-year transition probabilities for the multi-state model are presented below. It is evident from the table that within the next ten years, there is a probability of 87% for a subject to transit from MCI state to AD state while there is a 62%

chance that a person in the CN state would still be in this state.

Table 6.7: Transition Probability Matrices

	State 1	State 2	State 3
State 1	0.6235	0.2017	0.1748
State 2	0.1322	0.2606	0.6072
State 2	0.0201	0.1066	0.8733

6.2.3 Mean Sojourn Times

For the estimated model, the average length of time a subject spends in a transient state is presented below. It is evident from the table that, a subject could spend as long as 18.5 years in the cognitive normal state before transiting to any other state while a subject could also spend a whopping 41.7 years in the AD state before transiting.

Table 6.8: Mean Sojourn Times

	Estimates	SE	L	U
State 1	18.5252	1.6005	15.6395	21.9434
State 2	5.8191	0.2686	5.3157	6.3702
State 3	41.7355	8.4054	28.1239	61.9350

6.2.4 Probability That Each State Is Next

For the model under consideration, the probability that each state is next is presented in the table below which shows that there is a high probability (79.41%) for a subject to transit from the MCI stage to the AD stage while there is a relatively low probability (20.59%) for a subject to transit from CN state to MCI state.

Table 6.9: Probability That Each State is Next

	State 1	State 2	State 3
State 1	0	1.0000 (1.0000,1.0000)	0
State 2	0.2059 (0.1707,0.2488)	0	0.7941 (0.7512,0.8293)
State 3	0	1.0000 (1.0000,1.0000)	0

6.2.5 Ratio of Transition Intensities

For the estimated model, recovery is 1.9 ($e^{0.6554}$) [Est=0.6554, std. err=0.0887, 95% CI(0.05027, 0.8546)] times as likely as progression as evident in the table below.

6.3 Survival Model: Cox Proportional Hazard Model

This section provides the empirical output of the survival model applied to the ADNI data. The section starts with Table 8.16 which presents the output of the Cox PH model with gender and age as covariates and MCI as the event of interest. It is evident from the table that while age is significant [$\gamma = 0.05454, p = 0.000833$] on a subject transiting from CN state to MCI state, on the other hand, gender with male as the reference is not.

Table 6.10: Cox Proportional Hazard Model with MCI as event: Model Parameters

	Est.	exp(Est.)	Std. Error(Est.)	z	Pr(> |z|)	
Gender(Ref=Male)	0.34313	1.40935	0.18743	1.831	0.067138	.
Age	0.05454	1.05605	0.01632	3.342	0.000833	***

Table 7.17 presents a 95% C.I. for model parameters of the Cox PH model with MCI as an event. It also presents exponential estimates of the model parameters. Age has a 1.1 times effect on a subject transiting from CN state to MCI state.

Table 6.11: Cox Proportional Hazard Model with MCI as event: 95% C.I. for Model Parameters

	exp(Estimate)	exp(-Estimate)	lower 95%	upper 95%
Gender(Ref=Male)	1.409	0.7095	0.9761	2.035
Age	1.056	0.9469	1.0228	1.090

Concordance=0.625 (Std. Error=0.031)
Likelihood ratio test = 15.3 on 2 df, $p = 5e^{-4}$
Wald Test = 15.5 on 2 df, $p = 4e^{-4}$
Score (Logrank) Test = 15.63 on 2 df, $p = 4e^{-4}$

6.4 Fitting Linear Mixed Model to the PPMI Data

LMMs have emerged as the primary tool for experimental research, where repeated measures designs are the norm (Jacobi et al. 2015; Lazaro et al. 2020). The LMM can be estimated using different procedures, including the popular MLE approach and the emerging Bayesian estimation technique(; Alsefri et al. 2020; Lazaro et al. 2021).

In the sections below, the study employs both estimation methods and applies them to the PPMI data which is publicly available data. Covariates that this thesis considers include gender, age, family history with yes as a reference, years from baseline and lastly diagnostic group with cognitive normal as a reference and Prodromal and PD as the other categories. The dependent variables this study considers are important assessment tools of PD namely, tremor and UPDRS Total. These covariates are used to examine the influence on these assessment tools on the progression of PD. First, from a Bayesian perspective, the model is fitted employing the brm function in the brms package of the R software followed by a frequentist approach which was also fitted employing the lme4 package in R.

6.4.1 Analyzing the results from the brms model

All parameters were estimated employing the mean and standard deviation of the posterior distribution. Additionally, two-tailed 95% credible intervals were calculated hinged on quantiles. The estimates, Efficient Sample (Eff. Sample), and Rhat were used to assess the quality of the algorithm in estimating the posterior distribution of the parameters (Burkner, 2017b; Gelman et al. 2014). Convergence of the model is considered achieved if Rhat is less than 1.1; otherwise, more iterations or stronger priors may be needed. The convergence of the chains and posterior distributions can also be visually examined using a graphical approach (Carpenter et al. 2017).

6.4.2 Fitting LMM with Tremor as Dependent, Bayesian Estimation

The Bayesian LMMs output starts with information about the model, the data used as well as the sampling procedure. It could be noticed from the output below that, in total, $25,000$ samples were obtained from the posterior distribution using 5 chains, each consisting of $10,000$ iterations, but the first $5,000$ iterations were discarded as warm-up. For the fitted model, there are 5 chains, each creating 10000 iterations (draws). Nevertheless, only half of these iterations are retained because the other half is utilized for warm-up to ensure algorithm convergence. Consequently, the total number of posterior draws is equivalent to;

$$5\ chains * (10000\ iterations - 5000\ warm - up) = 4000$$

Often, presenting the entire posterior distributions as graphs can be impractical. Instead, it is essential to find a concise way to summarize the information. This can be achieved through a point estimate, similar to β in frequentist regressions, which provides a single value summary. Additionally, a credible interval is used

to represent the associated uncertainty, along with significance indices that offer insights into the relative importance of the effect.

The Population-Level Effects table is particularly useful for evaluating hypotheses. It presents information in a concise format, similar to The Family Specific Parameters table, but for a different type of model parameter, namely the standard deviation (sigma) of the assumed normal distributions describing the distribution of measures in each design cell. Lastly, the model output includes general information about the model fit. If the model fails to converge or other issues arise, an informative message will be displayed in the last part of the summary.

Focusing on the Population-Level Effects table, which holds the most theoretical interest as it provides answers to the research hypotheses. Each row in the table corresponds to a model parameter, that is, the model coefficients. It presents essential summary statistics based on the samples obtained from the model fit. The information in credible intervals columns is particularly relevant, as it gives the lower and upper bounds of the 95% credible interval for each parameter, estimated from the posterior samples (Heck et al. 2022).

In various scientific fields, merely describing the effects is not enough as researchers also seek to determine the practical or statistical significance of these effects, in other words, whether they are important. One common approach, similar to the frequentist framework, to assess the significance of an effect is to check if the credible interval contains zero. If zero is not within the interval, it indicates that the effect is significant. That is credible intervals that contain zero point to a true null hypothesis, meaning, intervals that do not contain zero depicts statistical significance (Nalborczyk et al. 2019).

It can be seen from the estimated model output that, diagnostic group and gender were found to have statistical significance on tremor while family history (95% CrI = [-0.09, 0.22]) and years (95% CrI = [-0.03, 0.01]) were found to have no statistical effect on tremor. Interestingly, age (95% CrI

= [0.00, 0.02]) was found not to have statistical significance on tremor. It can be seen from the model output that, when all other covariates are zero, the average tremor, defined by the intercept, is -3.44. Relative to a healthy person (HC), a person with Parkinson's Disease (PD) (Est. = 3.06, 95% CrI = [2.88, 3.24]) has an average increase of 3.06 in tremor while those in the prodromal (Est. = 0.59, 95% CrI = [0.36, 0.83]) are also associated with an average increase of 0.59 in tremor. A person's gender (Est. = 0.22, 95% CrI = [0.07, 0.37]) is positively associated with an average tremor by 0.22. This means relative to females, males experience a higher tremor by an average of 0.22.

Table 6.12: Group-Level Effects for Model with Tremor as Dependent Variable

Group-Level Effects	Est.	Est.Error	l-95% CI	U-95% CI	Rhat	Bulk ESS	Tail ESS
sd(Intercept)	1.02	0.03	0.96	1.08	1.00	7325	13819
sd(Years)	0.19	0.01	0.17	0.20	1.00	7088	12740
cor(Intercept, Years)	-0.31	0.05	-0.41	-0.22	1.00	5700	11473

Table 6.13: Population-Level Effects for Model with Tremor as Dependent Variable

Pop-Level Effects	Est.	Est.Error	l-95% CI	U-95% CI	Rhat	Bulk ESS	Tail ESS
(Intercept)	-3.22	0.25	-3.71	-2.74	1.00	6133	10675
Age	0.01	0.00	0.00	0.02	1.00	5918	105511
Diagnosis (Parkinson Disease)	3.06	0.09	2.88	3.24	1.00	6466	10881
Diagnosis (Prodromal)	0.59	0.12	0.36	0.83	1.00	7226	12282
Gender	0.22	0.08	0.07	0.37	1.00	5325	10223
Years	-0.01	0.01	-0.03	0.01	1.00	8118	13197
Family History (Yes as reference)	0.07	0.08	-0.09	0.22	1.00	5670	9815

Table 6.14: Family Specific Parameters for Model with Tremor as Dependent Variable

Family Specific Parameters	Est.	Est.Error	l-95% CI	U-95% CI	Rhat	Bulk ESS	Tail ESS
sigma	0.86	0.01	0.84	0.87	1.00	28046	18287

6.4.3 Fitting LMM with UPDRS Total as Dependent, Bayesian Estimation

It could be noticed from the output below that, a total of 25,000 (5 *chains* $*$ (10000 − 5000)) samples were obtained from the posterior distribution using 5 chains. Each chain consisted of 10,000 iterations, but the first 5,000 iterations were discarded as warm-up. Credible intervals that contain zero points to a true null hypothesis, meaning, intervals that do not contain zero depicts statistical significance. It can be seen from the estimated model that, age, diagnostic group and years were found to have statistical significance on UPDRS Total while family history (95% CrI = [-0.06, 0.11]) and gender (95% CrI = [-0.02, 0.14]) were found to have no statistical effect on UPDRS Total. It can be seen from the model output that when all other covariates are zero, the average UPDRS Total, defined by the intercept, is -1.97. An increase in the year of a person's age (Est.= 0.01, 95% CrI = [0.01, 0.02]) averagely increases UPDRS Total by 0.01. Relative to a healthy person (HC), a person with PD (Est.= 1.93, 95% CrI = [1.83, 2.03]) has an average increase of 1.93 in UPDRS Total while those in the prodromal group (Est.= 0.75, 95% CrI = [0.62, 0.87]) are also associated with an average increase of 0.75. A unit increase in the baseline year (Est.= 0.09, 95% CrI = [0.08, 0.10]) increases UPDRS Total by an average of 0.09.

Table 6.15: Group-Level Effects for Model with UPDRS Total as Dependent Variable

Group. Level Eff.	Est.	Est.Error	l-95% CI	U-95% CI	Rhat	Bulk ESS	Tail ESS
sd(Intercept)	0.57	0.02	0.54	0.60	1.00	6401	10883
sd(Years)	0.13	0.01	0.12	0.14	1.00	6283	10242
cor(Intercept, Years)	-0.29	0.05	-0.38	-0.20	1.00	5920	11630

Table 6.16: Population-Level Effects for Model with UPDRS Total as Dependent Variable

Pop-Level Eff.	Est.	Est.Error	l-95% CI	U-95% CI	Rhat	Bulk ESS	Tail ESS
(Intercept)	-1.97	0.14	-2.24	-1.70	1.00	4025	8134
Age	0.01	0.00	0.01	0.02	1.00	3745	7510
Diagnostic (Parkinson Disease)	1.93	0.05	1.83	2.03	1.00	4237	8101
Diagnostic (Prodromal)	0.75	0.06	0.62	0.87	1.00	4353	8667
Gender	0.06	0.04	-0.02	0.14	1.00	3601	7184
Years	0.09	0.01	0.08	0.10	1.00	5773	10163
Family History	0.02	0.04	-0.06	0.11	1.00	3090	6619
(Yes as reference)	0.02	0.04	-0.06	0.11	1.00	3090	6619

Table 6.17: Family Specific Parameters for Model with UPDRS Total as Dependent Variable

Family Specific Parameters	Est.	Est.Error	l-95% CI	U-95% CI	Rhat	Bulk ESS	Tail ESS
sigma	0.35	0.00	0.34	0.36	1.00	21222	16995

6.4.4 Extracting the Posterior

Once the models are estimated from the data, the parameters, that is, the model coefficients can be extracted. Table 7.7 below shows the summary of the tremor model based on the posterior distribution. It presents the model parameters in

Table 6.18: Posterior Summary for Model with Tremor as Dependent Variable

Posterior Summary	Estimate	Est.Error	Q_2.5	Q_97.5
Intercept	-3.2245	0.2485	-3.7087	-2.7357
Age	0.0093	0.0038	0.0017	0.0168
Diagnostic (Parkinson's Disease)	3.0627	0.0910	2.8833	3.2406
Diagnostic (Prodromal)	0.5928	0.1210	0.3552	0.8281
Gender(Male)	0.2189	0.0767	0.0694	0.3702
Years	-0.0084	0.0092	-0.0262	0.0096
Family History (No)	0.0679	0.0784	-0.0871	0.2219
sd RID Intercept	1.0187	0.0314	0.9589	1.0823
sd RID Years	0.1877	0.0082	0.1719	0.2038
cor RID Intercept Years	-0.3138	0.0483	-0.4055	-0.2152
sigma	0.8567	0.0076	0.8420	0.8718
r RID[3000,Intercept]	-0.1130	0.4666	-1.0305	0.7978
...	...	...	...	...
r RID[92516,Years]	4.6315e-02	0.1835	-3.1387e-01	4.0731e-01
r RID[92560,Years]	4.9165e-02	0.1813	-3.0666e-01	4.0596e-01
r RID[92834,Years]	-6.3073e-03	0.1096	-2.2402e-01	2.0693e-01
lp	-1.2403e+04	42.6026	-1.2488e+04	-1.2320e+04

the form of a comprehensive data frame with four columns: the intercept and the coefficients (parameters) of the model. These columns contain the posterior distributions of the parameters, which represent a range of plausible values for each parameter. This differs from the frequentist linear regression model results obtained using lm and lmer, where only single values were reported for each effect of the model, without any distribution. This distinction is a key aspect of Bayesian analysis, as it eliminates the need for p-values, test statistics, or degrees of freedom. All the necessary information is encompassed within this posterior distribution. Next is the visualization of the posterior distribution of the parameters of interest.

Figures 7.3 through to 7.11 below show the trace plots and density plots of the Bayesian models with tremor and UPDRS Total as dependent variables. It is evident from the trace plots that, there are no apparent anomalies and there seems to be a very slight serial correlation between successive draws, and the chains appear to have extensively explored the sample space. The figures are also indicative of the fact that the initial distributions and the distributions of the subsequent terms of the chain were not very different from the distribution targeted. From the trace plots, it is evident that the chains converged to the distribution targeted relatively very quickly. In sum, the graph points to the fact that a greater chunk of the sample was pulled from distributions that are not significantly in variance from the distribution targeted and also the effective sample size was large enough.

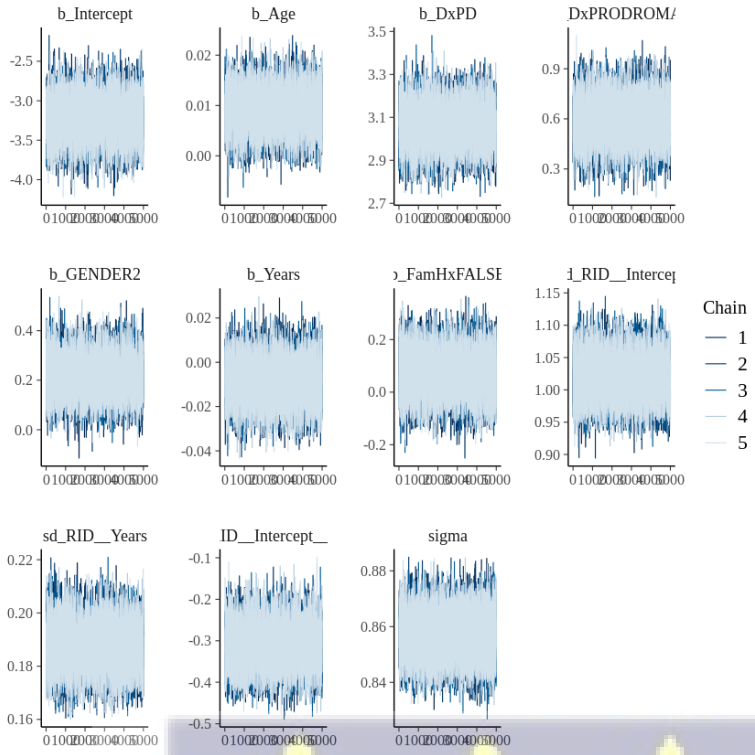


Figure 6.3: Trace Plot for Bayesian Model with Tremor as Dependent Variable

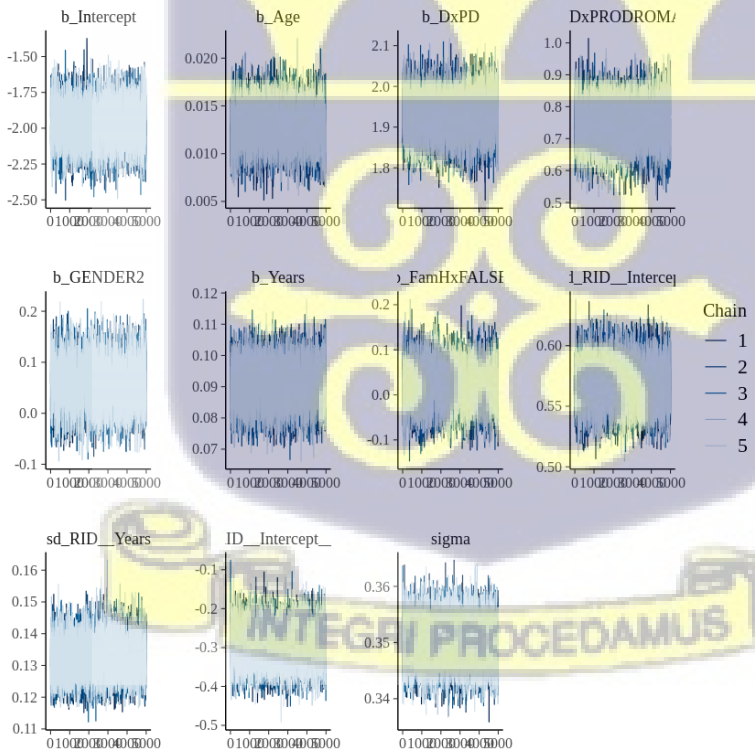


Figure 6.4: Trace Plot for Bayesian Model with UPDRS Total as Dependent Variable

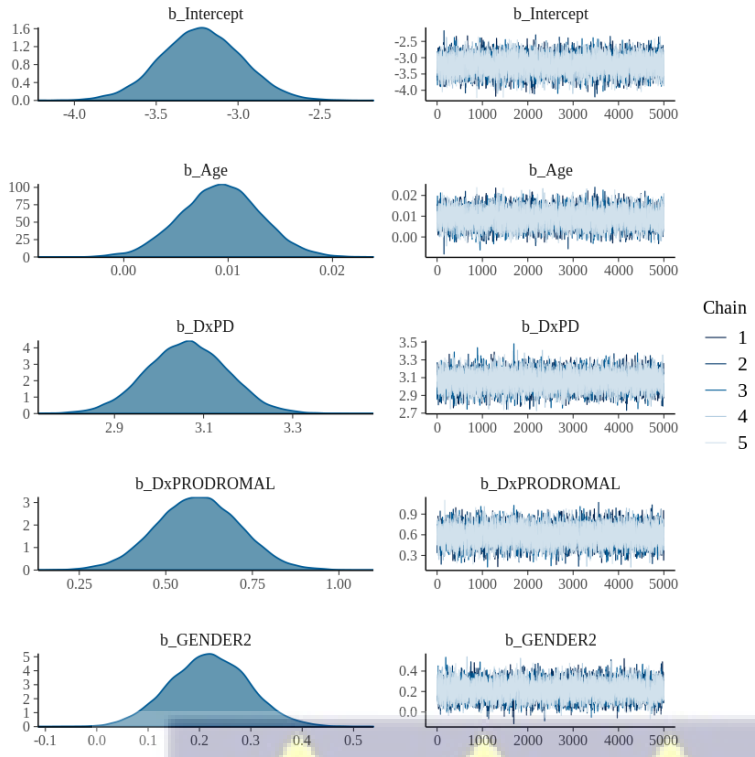


Figure 6.5: Trace and Density Plot for Bayesian Model with Tremor as Dependent Variable

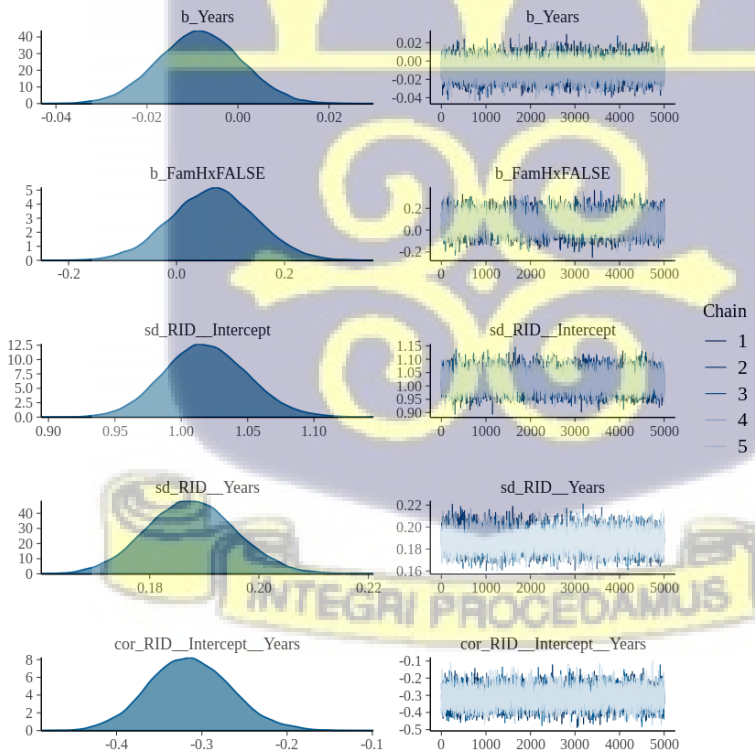


Figure 6.6: Trace and Density Plot for Bayesian Model with Tremor as Dependent Variable

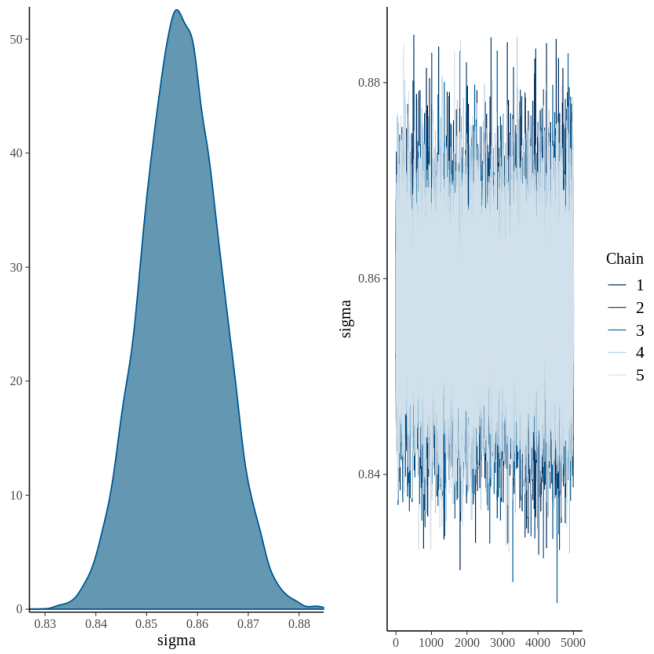


Figure 6.7: Trace and Density Plot for Bayesian Model with Tremor as Dependent Variable

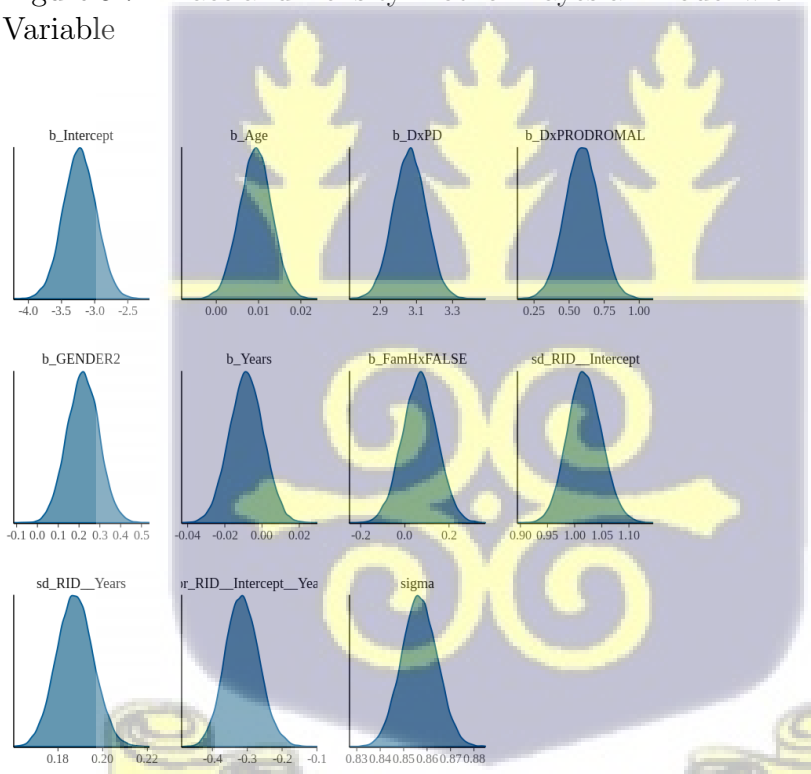


Figure 6.8: Density Plot for Bayesian Model with Tremor as Dependent Variable

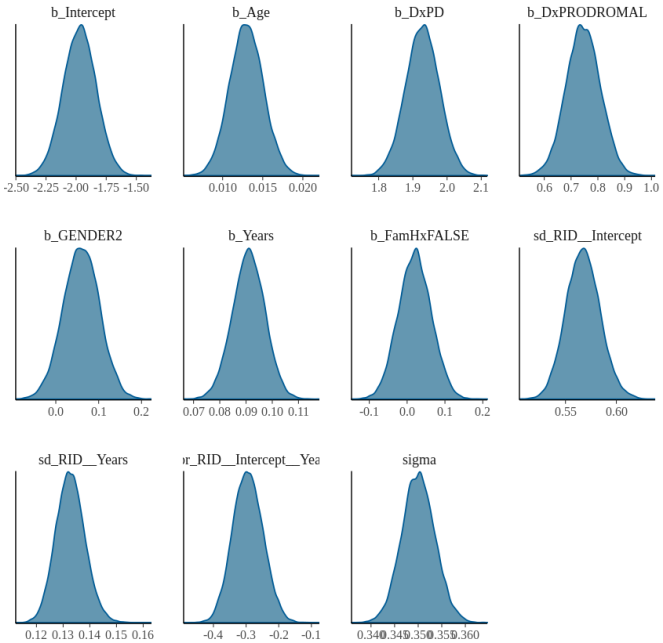


Figure 6.9: Density Plot for Bayesian Model with UPDRS Total as Dependent Variable

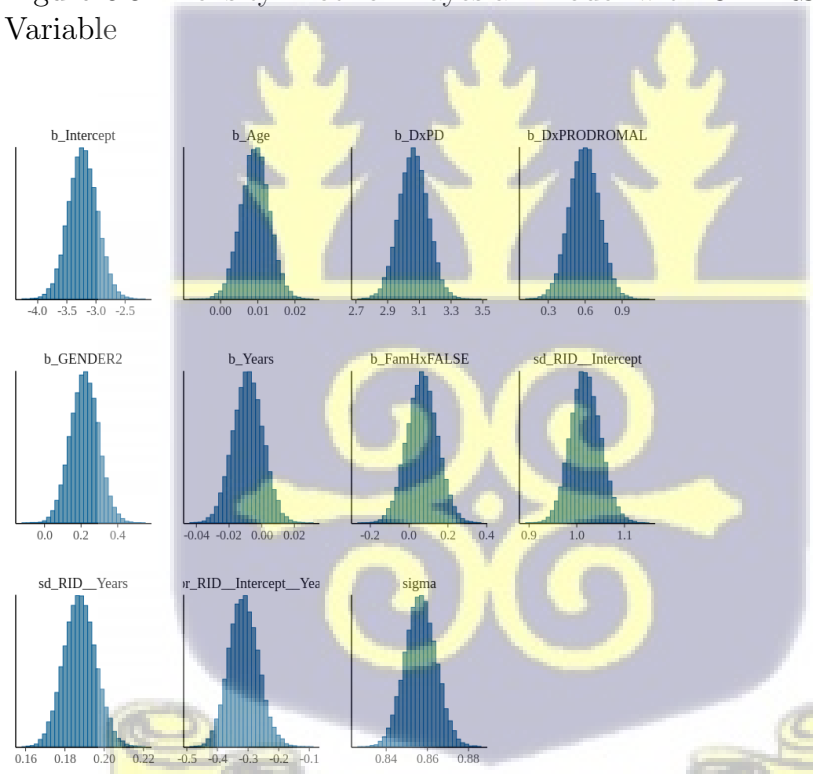


Figure 6.10: Histogram Plot for Bayesian Model with Tremor as Dependent Variable

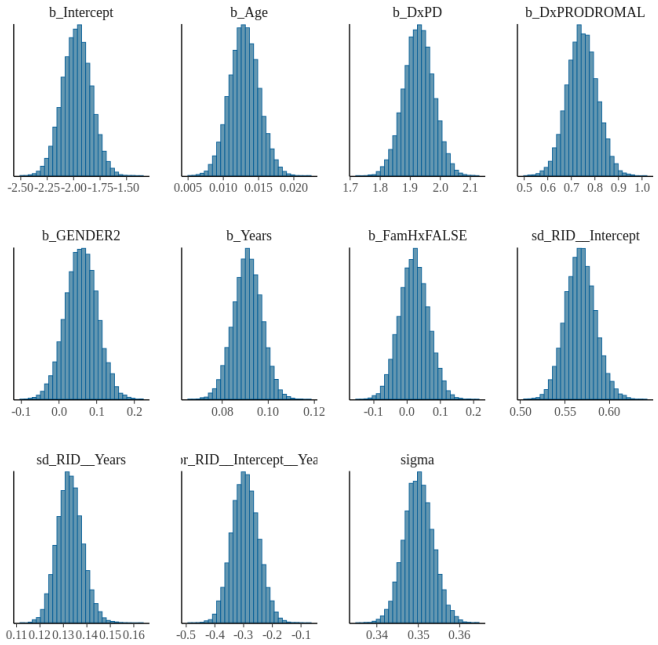


Figure 6.11: Histogram Plot for Bayesian Model with UPDRS Total as Dependent Variable

6.4.5 Autocorrelation Function (ACF) Plots

The ACFs of the tremor and UPDRS Total models are displayed in Figures 7.12 and 7.13 below. The two figures clearly show that there is significant autocorrelation at short lags, but it quickly diminishes to zero, indicating the absence of anomalies. While ACF plots are useful for identifying problems and measuring autocorrelation, they might not provide specific information about the issue. Therefore, it is recommended to use multiple diagnostics, including trace plots, density plots, and ACF plots, to thoroughly evaluate the model's quality.

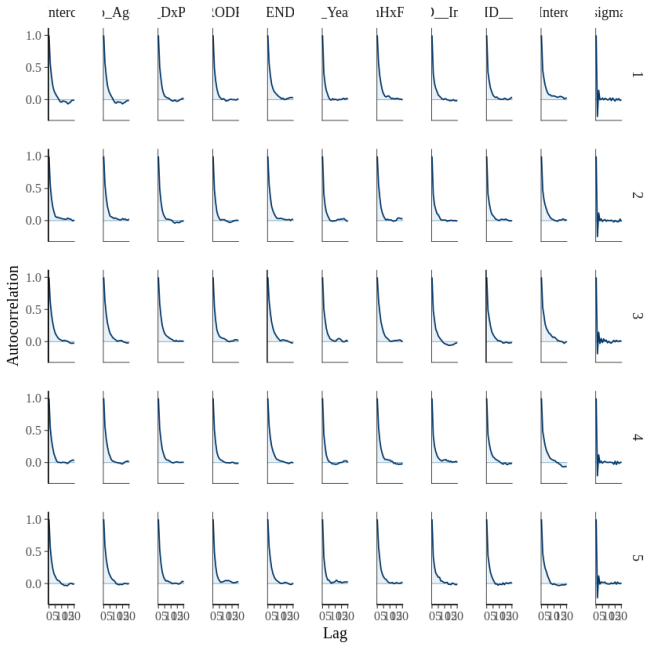


Figure 6.12: ACF Plot for Bayesian Model with Tremor as Dependent Variable

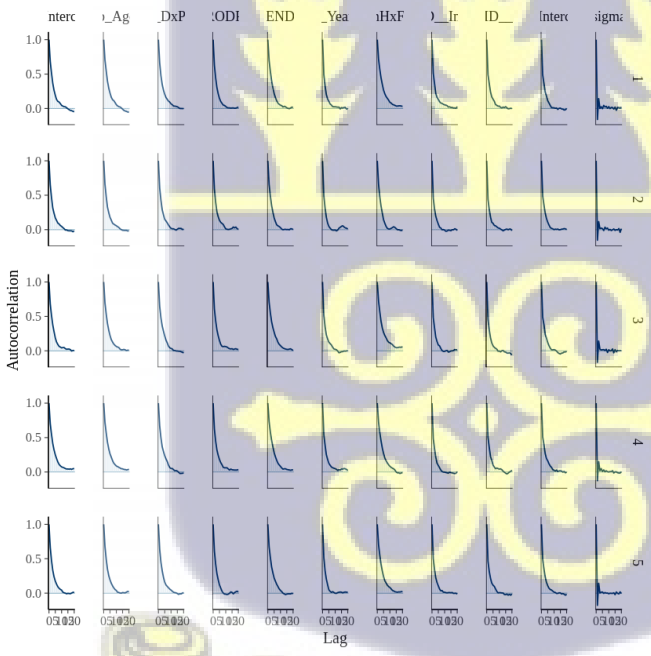


Figure 6.13: ACF Plot for Bayesian Model with UPDRS Total as Dependent Variable

6.5 Maximum Likelihood Estimation Method (MLE)

The next section presents the output when the MLE method is fitted to the PPMI data.

6.5.1 Fitting LMM with Tremor as Dependent

The output as presented below shows an LMM fitted using REML criterion. The output starts by displaying the dataset used, followed by two rows of fit statistics: the Akaike Information Criterion (Akaike, 1974; AIC = 22681.67) and the Bayesian Information Criterion (Schwarz, 1978; BIC = 22758.35), along with the log-likelihood (Log-Lik = -11329.84). In this study, AIC and BIC are utilized to compare models and assess their fit. Lower values of AIC or BIC indicate better model fit to the data. For instance, a BIC of -55.22 indicates a superior fit compared to a BIC of $+23.56$. To assess the model's fit, a common approach is to rerun the analysis with only the intercept terms, known as the null model, and compare its AIC and BIC to those of the hypothesized (full) model.

Following that, the subsequent section of the output provides estimates for the random effects in terms of correlations and standard deviations. Three values are presented: the intercept, the random effect of the trend variable 'Years', and the residual component, which defines the error. The estimates of the variance are of importance because they are summed up to get the total variability attributed to the random effects. To obtain the variability attributable to each random effect, the total variance is divided by the variability of each random effect, and this measure is synonymous to R^2 in conventional regression. Summation of the variability components results in:

$$= 1.0176^2 + 0.1873740^2 + 0.8564546^2$$

$$= 1.8041$$

To arrive at the percentage of variability accounted for, a measure which specifies whether or not an effect is significant, this total variance is then divided by the nested effect variance.

$$= (0.1873740^2)/1.804133$$

$$= 0.0195$$

It is evident that only 1.95% of the total variability of the random effects can be attributed to the nested effect. If the proportion of variability for each random effect is very small, they are considered negligible, and it is not recommended to use LMM. In such cases, the random effects are removed from the model, and generalized linear modeling is employed instead. The subsequent section of the output provides estimates of the fixed effects, starting with the model formula. These estimates are interpreted in a similar manner as one would interpret estimates from a traditional ordinary least squares linear regression.

The intercept is the average of the outcome (tremor) when all the covariates equal zero in value. It could be noticed from Table 7.10 that a unit increase in the covariate age (β = 0.0093,[p =0.0157]) represents a 0.0093 rise in tremor. It could also be seen from the output that family history (β = 0.0665,[p =0.4008]) and base line year (β = -0.0084,[p =0.3501] had no significant effect on tremor while age, diagnostic group, and gender were found to be statistically significant with the dependent variable tremor. Relative to a person with PD, a healthy person (HC) (β = -3.0611,[p =0.0000]) has a decrease of -3.0611 in tremor. Again, relative to a person with PD, a person in the prodromal group (β = -2.4681,[p =0.0000])also has a decrease of -2.4681 in tremor. Furthermore, the categorical predictor gender (β = 0.2190,[p =0.0000]) has a coefficient of 0.2190398 which indicates that the average tremor of the second group (males) is 0.2190 higher than the average tremor of the first group (females).

The output's last section, displays the associations between the fixed effects, and

it is used to examine multicollinearity. From Table 7.11, it could be noticed that, the predictors are not correlated, with the exception of age and the intercept. Therefore, multicollinearity is not a concern.

Table 6.19: Linear Mixed Effects Model fit by REML with Tremor as Dependent; Fixed Statistics

AIC	BIC	logLik
22681.67	22758.35	-11329.84

Table 6.20: Random Effects for Model with tremor as Dependent Variable

	StdDev	Corr
Intercept	1.0176	(Intr)
Years	0.1874	-0.3170
Residual	0.8645	

Table 6.21: Fixed Effects for Model with tremor as Dependent Variable

Fixed Effects	Estimate	Std. Error	DF	t value	p-value
(Intercept)	-0.3800	0.2662	7022	-1.42735	0.1535
Age	0.0093	0.0038	849	2.4208	0.0157
Diagnostic (Healthy Control)	-3.0611	0.0894	849	-34.2363	0.0000
Diagnostic (Prodromal)	-2.4681	0.1025	849	-24.0688	0.0000
Gender	0.2190	0.0764	849	2.8670	0.0042
Years	-0.0084	0.0090	7022	-0.9344	0.3501
Family History (Yes as reference)	0.0665	0.0792	849	0.8406	0.4008

Table 6.22: Correlation Matrix for Model with tremor as Dependent Variable

Correlation	(Intercept)	Age	Diagnosis (Healthy Control)	Diagnosis (Prodromal)	Gender	Years
Age	-0.857					
Diagnosis(Healthy Control)	-0.145	0.043				
Diagnosis (Prodromal)	0.138	-0.253	0.235			
Gender	-0.396	-0.082	0.001	-0.004		
Years	-0.062	0.009	-0.038	0.073	-0.013	
Family History (Yes as Reference)	-0.179	0.098	0.134	0.020	-0.033	0.017

6.5.2 Fitting LMM with UPDRS Total as Dependent, MLE Method

The fixed statistics as presented in the first part of the output show an AIC of 7983.743, BIC of 8058.261 and a log-likelihood of -3980.872. The following section of the output presents the estimates for the random effects and the variance component of interest.

Table 6.23: LMMs Fit by REML with UPDRS Total as Dependent Variable; Fixed Statistics

AIC	BIC	logLik
7983.743	8058.261	-3980.872

Table 6.24: Random Effects for Model with UPDRS Total as Dependent Variable

	StdDev	Corr
Intercept	0.5663	(Intr)
Years	0.1325	-0.2970
Residual	0.3499	

Table 6.25: Fixed Effects for Model with UPDRS Total as Dependent Variable

Fixed Effects	Estimate	Std. Error	DF	t value	p-value
(Intercept)	-0.1035	0.1453	5619	-0.7123	0.4763
Age	0.0130	0.0021	847	6.2189	0.0000
Diagnosis (Healthy Control)	-1.9274	0.0489	847	-39.4095	0.0000
Diagnosis (Prodromal)	-1.1806	0.0546	847	-21.6200	0.0000
Gender	0.0604	0.0415	847	1.4555	0.1459
Years	0.0911	0.0062	5619	14.7249	0.0000
Family History (Yes as Reference)	0.0209	0.0430	847	0.4867	0.6266

Table 6.26: Correlation Matrix for Model with UPDRS Total as Dependent Variable

Correlation	(Intercept)	Age	Diagnosis (Healthy Control)	Diagnosis (Prodromal)	Gender	Years
Age	-0.858					
Diagnostic (Healthy Control)	-0.147	0.045				
Diagnostic (Prodromal)	0.140	-0.260	0.242			
Gender	-0.394	-0.082	0.000	0.001		
Years	-0.055	0.006	-0.034	0.082	-0.010	
Family History (Yes as Reference)	-0.183	0.100	0.134	0.009	-0.029	0.014

The variability of the intercept, the random effect, years, and the residual terms are 0.3207513, 0.017547, and 0.1224566 respectively. The addition of the variance components yields a total variance of the random effects as shown below;

$$= 0.5663491^2 + 0.1324653^2 + 0.3499380^2$$

$$= 0.4607$$

This total variance is then divided by the nested effect variance which results in the percentage of variability accounted for, which shows the meaningfulness of

the effect or not.

$$= 0.0175471/0.4607457878$$

$$= 0.0381$$

It can be noticed that only 3.81% of the total variability of the random effects is associated with the nested effect. Moving on, the subsequent part of the output displays the estimates of the fixed effects. The output of the model provided in Table 7.14 shows that family history (β = 0.0209, [p =0.6266]) and gender (β = 0.0604, [p =0.1459]) were not statistically significant while age, diagnostic group, and year were found to be statistically significant on the outcome variable UPDRS. A unit rise in the age (β = 0.0130, [p =0.0000]) of a person corresponds to 0.0129772 increase in the person's UPDRS Total. Also, relative to a person with PD, a healthy person (HC) (β = -1.9274, [p =0.0000]) has a decrease of -1.9274 in his or her UPDRS Total. Again, relative to a person with PD, a person identified as prodromal (β = -1.1806, [p =0.0000]) also is expected to experience a decrease of -1.1806 in UPDRS Total. A one-unit increase in the base year (β = 0.0911, [p =0.0000]) increases UPDRS Total by 0.0911.

The correlation matrix of the fixed effects variables presented in the last section of the output, assesses multicollinearity among the independent variables under consideration. As can be seen from the output, the predictors are not related; with the exception of age and the intercept. Therefore, multicollinearity is not a concern.

CHAPTER 7

A COMPARATIVE ANALYSIS OF THE HYBRID AND FREQUENTIST TWO-STAGE ESTIMATION METHOD FOR JOINTLY MODELLING LONGITUDINAL AND MULTI-STATE MODELS: AN EMPIRICAL PERSPECTIVE

7.1 Introduction

The proposed hybrid joint two-step estimation technique adopted for a univariate longitudinal and time to a single event data in Chapter Six is extended to multi-state data in this Chapter. As time-to-single event is often limited compared to time-to-multiple events, a much more curious application would be to investigate the connection between a univariate longitudinal and multi-state data. It is on the backdrop of this that this thesis navigates to establish a joint two-stage estimation procedure that yields relatively precise estimates when applied to univariate longitudinal and multi-state data.

Since validated models for joint analysis of longitudinal and time to a single event data exist in the literature, a simulation study was conducted in Chapter Six to compare the suggested hybrid model with these existing models for improved inference. However, for longitudinal and time to multiple event data, no validated joint models were found in the literature, whether through joint likelihood or two-stage estimation procedures. This highlights the fact that the study of joint

models for longitudinal and time to multiple event data is still in its early stages, as confirmed by Hickey et al. (2016) and Sweeting and Thompson (2011). As an alternative to the simulation study, an empirical application was undertaken to compare the proposed two-step hybrid estimation method with a frequentist two-step approach to determine which one demonstrates improved inferential performance. This empirical application utilized data from the ADNI, which has been discussed in Chapters two and seven.

Both the multi-state hybrid model and multi-state frequentist model were assessed on their standard errors, confidence intervals, and confidence lengths in ascertaining which model yields precise estimates. It is a theoretically established fact that the Bayesian models yield stable estimates after the model has converged and it is on the backdrop of this that the multi-state hybrid model was extended to incorporate the last 100 posterior predictive estimates from the Bayesian model. Meaning, this Chapter presents two versions of the hybrid model, one with all the posterior predictive estimates from the Bayesian model and the other with the last 100 predictive posterior estimates after the Bayesian model has converged.

7.2 ADNI Data Structure and Modelled Transitions

As already elucidated in Chapter seven that transitions with relatively few numbers are not included in the modelling process, the same applies in this empirical application too. Hence the structure of the fitted transition model in Figure 8.1 is as a result of the frequency distribution table shown in Table 8.1. For the multi-state submodel, the three disease states are classified as CN as State 1, MCI as State 2 and finally, AD as State 3. The continuous longitudinal dependent variable is mPACCtrailsB, an assessment tool for the diagnosis of Dementia (AD). The explanatory variables considered were age, gender, and years followed from baseline.

Table 7.1: Frequency Distribution of Number of Transitions Among States

	State 1(CN)	State 2(MCI)	State 3(AD)
To / From			
State 1(CN)	2605	159	6
State 2(MCI)	114	3366	388
State 3(AD)	0	28	1480

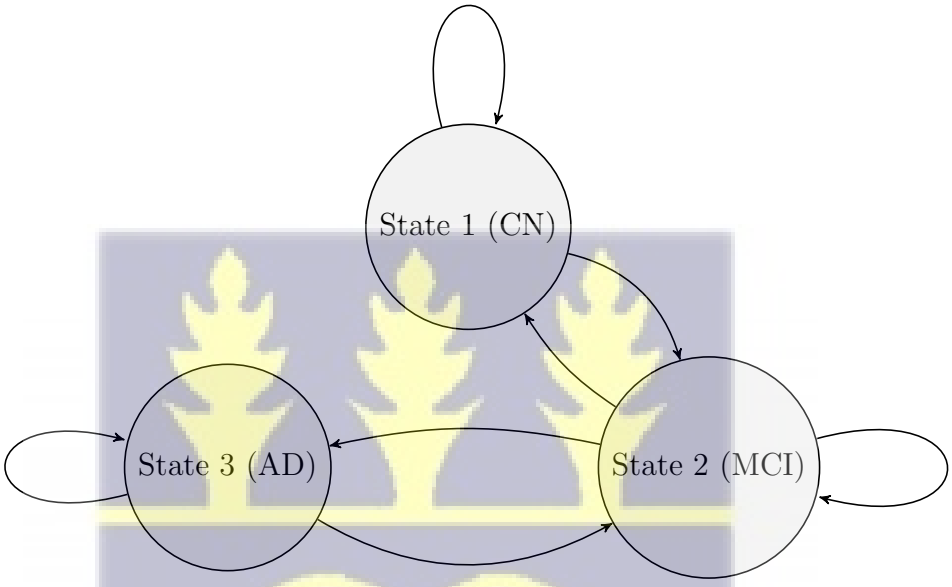


Figure 7.1: The structure of the fitted transition model

7.3 Hybrid Two-Stage Longitudinal and Multi-State Model Formulation

The novelty this thesis seeks to propose is to defy the existing estimation procedure and plug in each of the posterior predictive estimates from the Bayesian LMM from stage one of the joint modelling framework. Specifically, during the initial stage, the Bayesian technique is used to obtain estimated values of the posterior estimates for the parameters in the longitudinal submodel using the LMM procedures. This enables the evaluation of true unobserved longitudinal data continuously over time, taking into account the complete longitudinal history

when estimating the multi-state function.

The longitudinal submodel is specified as:

$$mPACĈtrailsB_i(t) = \beta_0 + b_{0i} + \beta_1 \text{Age} + \beta_2 \text{Gender} + \beta_3 \text{Years.bl} + b_{1i} \text{Years.bl} \quad (7.1)$$

This implies, for the $m = 1, 2, 3, ..., M$ posterior predictive estimates specified yields:

$$(\hat{\mu}_i(t))^m = \left(mPACĈtrailsB_i(t)\right)^m \quad (7.2)$$

In the second stage model, each $(\hat{\mu}_i(t))^m$ is plugged in the Markov MSM with PH which is utilized to fit the transition times. Specifically, the following hazards functions modelled are as follows:

$$h^i_{cn,mci}(t|\mu_{cn,mci,i}) = h_{cn,mci,0}(t) exp\left\{AGE_{cn,mci,i}(t)\gamma_{cn,mci} + \alpha_{cn,mci}\left(\hat{\mu}_{cn,mci,i}(t)\right)\right\} \quad (7.3)$$

$$h^i_{mci,cn}(t|\mu_{mci,cn,i}) = h_{mci,cn,0}(t) exp\left\{AGE_{mci,cn,i}(t)\gamma_{mci,cn} + \alpha_{mci,cn}\left(\hat{\mu}_{mci,cn,i}(t)\right)\right\} \quad (7.4)$$

$$h^i_{mci,ad}(t|\mu_{mci,ad,i}) = h_{mci,ad,0}(t) exp\left\{AGE_{mci,ad,i}(t)\gamma_{mci,ad} + \alpha_{mci,ad}\left(\hat{\mu}_{mci,ad,i}(t)\right)\right\} \quad (7.5)$$

$$h^i_{ad,mci}(t|\mu_{ad,mci,i}) = h_{ad,mci,0}(t) exp\left\{AGE_{ad,mci,i}(t)\gamma_{ad,mci} + \alpha_{ad,mci}\left(\hat{\mu}_{ad,mci,i}(t)\right)\right\} \quad (7.6)$$

In estimating the parameters of the multi-state submodel, the MLE method was employed.

7.4 The Frequentist Two-Stage Longitudinal and Multi-State Model Formulation

For the frequentist standard two-step model, the parameters of both the repeated measurements, (β, b), in equation 8.1 and multi-state sub-models, (γ, α), specified in equations 8.3 to 8.6 are estimated using the MLE procedure. That means the

estimated longitudinal trajectory estimated from equation 8.1 in the first stage of the modelling framework is plugged into the second stage equations. The frequentist method utilizes a point estimate from the first phase into the second phase as opposed to a Bayesian posterior distribution in the proposed hybrid model which uses all the posterior predictive estimates.

7.5 Results

The details of the results when the suggested hybrid model was contrasted with the frequentist model are presented in this section.

Table 8.1 below displays the model output of the longitudinal sub-model of the joint two-stage model. It presents the model output for the novel multi-state hybrid model and the frequentist multi-state model when applied to the ADNI data. Both models were compared and assessed on their estimates, standard errors, 95% confidence interval as well as their confidence length. It could be noticed from the Table 8.1 that both models produce nearly identical estimates in terms of magnitude and direction. That is, the values of the estimates from both models are identically the same. The model estimates from the hybrid model are for $[\beta_0,\ \beta_1,\ \beta_2,\ \beta_3]$ are $[2.7556,\ -0.1002,\ -0.8047, 1.3505]$ respectively while that from the frequentist approach are also $[2.7121,\ -0.0995,\ -0.8092,\ -1.3503]$ respectively. This results is not too surprising as the two models are expected to produce similar estimates. The findings here confirm earlier assertions by Albert and Shih, (2010) the maximum likelihood and Bayesian estimation methods on most occasions yield similar results.

It is an established statistical fact that Bayesian models yield relatively less bias results when compared to the maximum likelihood methods. But surprisingly for the models under consideration, they were very marginal differences in the standard errors reported. Under the hybrid specification, the standard errors recorded for $[\beta_0,\ std.\ error\ =\ 1.2673]$ and $[\beta_1,\ std.\ error\ =\ 0.0172]$ are relatively lower compared to that from the frequentist method $[\beta_0,\ std.\ error =$

1.2777] and [β_1, $std.\ error = 0.0174$]. Very unusual, the standard errors recorded for [β_2, $std.\ error = 0.2462$] and [β_3, $std.\ error = 0.0502$] under the frequentist approach are relatively lesser in contrast to that of the hybrid approach [β_2, $std.\ error = 0.2495$] and [β_3, $std.\ error = 0.0521$]. The findings here contradict the assertions of Huong et al. (2017) that Bayesian estimates yield by far unbiased estimates compared to the frequentist estimates. However, this study confirms the findings of Proust-Lima et al. (2014) that in terms of producing unbiased estimates the Bayesian and frequentist approaches have parallel performances.

The confidence interval (CI) and confidence length (CL) are both functions of the standard error and have a directly proportional relation with it. This means an increase in the standard error has a direct increase in the CI and CL. It is evident from Table 8.1 that the CL of [β_0, $CL = 4.9077$, $\beta_1, CL = 0.0671$] of the hybrid model is marginally narrower compared to that from the frequentist approach [β_0, $CL = 5.0085$, $\beta_1, CL = 0.1989$]. The narration was the other way around for the [β_2, β_3]. The CL for [β_2, $CL = 0.9652$, $\beta_3, CL = 0.1967$] were a very little bit shorter than that of the hybrid model [β_2, $CL = 0.9764$, β_3, $CL = 0.2065$].

Table 7.2: Longitudinal Sub-Model Output for the Joint Hybrid and the Frequentist Model.

Sub-Model	Measure	Parameter	JM Hybrid STS	JM Frequentist Approach
Longitudinal Sub-Model				
	Estimate			
		β_0	2.7556	2.7121
		β_1	-0.1002	-0.0995
		β_2	-0.8047	-0.8092
		β_3	-1.3505	-1.3503
	Std. Error			
		β_0	1.2673	1.2777
		β_1	0.0172	0.0174
		β_2	0.2495	0.2462
		β_3	0.0521	0.0502
	95% CI			
	Lower CI	β_0	0.2662	0.2078
	Upper CI	β_0	5.1739	5.2163
	Lower CI	β_1	-0.1332	-0.1335
	Upper CI	β_1	-0.0662	-0.0654
	Lower CI	β_2	-1.2878	-1.2918
	Upper CI	β_2	-0.3114	-0.3266
	Lower CI	β_3	-1.4550	-1.4487
	Upper CI	β_3	-1.2485	-1.2520
	Confidence Length			
		β_0	4.9077	5.0085
		β_1	0.0671	0.1989
		β_2	0.9764	0.9652
		β_3	0.2065	0.1967

Next to report on is the multi-state sub-model output displayed in Tables 8.2 and 8.3. Table 8.2 reports the multi-state model output when all 8000 posterior estimates of the Bayesian LMM are plugged into the multi-state submodel to estimate the model parameters. As it is an established fact that the Bayesian models yield relatively stable estimates after the model has converged, the last 100 posterior estimates out of the total 8000 posterior estimates of the Bayesian LMM were used in estimating the model parameters and the model output is reported in Table 8.3. As could be noticed from Tables 8.2 and 8.3, all three models yield identical estimates in terms of magnitude and direction.

Inferring from Tables 8.2 and 8.3 the group parameter γ for transitions from Cognitive Normal(CN) to (Mild Cognitive Impairment (MCI), MCI to AD, MCI to Alzheimer Disease (AD), and AD to MCI for the hybrid model with all 8000 posterior estimates are [$\gamma_{1_{CN,MCI}} = 1.0201$, $\gamma_{1_{MCI,CN}} = 0.9409$, $\gamma_{1_{MCI,AD}} = 1.0244$, *and* $\gamma_{1_{AD,MCI}} = 0.9766$] respectively, that of the hybrid model with the last 100 posterior estimates are [$\gamma_{1_{CN,MCI}} = 1.0195$, $\gamma_{1_{MCI,CN}} = 0.9407$, $\gamma_{1_{MCI,AD}} = 1.0248$, *and* $\gamma_{1_{AD,MCI}} = 0.9763$] and that of the frequentist approach are [$\gamma_{1_{CN,MCI}} = 1.0200$, $\gamma_{1_{MCI,CN}} = 0.9409$, $\gamma_{1_{MCI,AD}} = 1.0244$, *and* $\gamma_{1_{AD,MCI}} = 0.9765$]. From the above narration, it is evident that the estimate of the group parameter γ from all three models is averagely identical in magnitude and direction.

Focus is now shifted to the association parameter, that is, the estimate that quantifies the connection between the longitudinal sub-model and the multi-state sub-model. It could be noticed from Tables 8.2 and 8.3 that the association parameter α for transitions from CN to MCI, MCI to CN, MCI to AD and AD to MCI for the hybrid model with all 8000 posterior estimate are [$\alpha_{1_{CN,MCI}} = 0.9453$, $\alpha_{1_{MCI,CN}} = 0.9841$, $\alpha_{1_{MCI,AD}} = 1.0350$, *and* $\alpha_{1_{AD,MCI}} = 0.9571$] respectively, that of the hybrid model with the last 100 posterior estimates are [$\alpha_{1_{CN,MCI}} = 0.9453$, $\alpha_{1_{MCI,CN}} = 0.9836$, $\alpha_{1_{MCI,AD}} = 1.0354$, *and* $\alpha_{1_{32}} = 0.9580$] respectively and that of the frequentist approach are [$\alpha_{1_{12}} = 0.9453$, $\alpha_{1_{MCI,CN}} =$

0.9838, $\alpha_{1_{MCI,AD}} = 1.0351$, *and* $\alpha_{1_{AD,MCI}} = 0.9575$] respectively. It is obvious from the above that the estimate of the association parameter α from all three models are almost the same in magnitude and direction.

With regards to the standard errors, there were significant differences between the three models with the hybrid model with 100 posterior estimates recording the least value followed by the hybrid model with all 8000 posterior estimates, and finally the frequentist model. It could be seen from Tables 8.2 and 8.3 that the standard error of the group parameter γ for transitions from CN to MCI , MCI to CN , MCI to AD and AD to MCI for the hybrid model with all 8000 posterior estimates are [std. $\text{error}_{CN,MCI} = 0.0011$, std. $\text{error}_{MCI,CN} = 0.0011$, std. $\text{error}_{MCI,AD} = 0.0007$, *and* std. $\text{error}_{AD,MCI} = 0.0009$] respectively, that of the hybrid model with the last 100 posterior estimates are [std. $\text{error}_{CN,MCI} = 0.0009$, std. $\text{error}_{MCI,CN} = 0.0002$, std. $\text{error}_{MCI,AD} = 1.0005$, *and* std. $\text{error}_{32} = 0.0007$] respectively and that of the frequentist approach are [std. $\text{error}_{CN,MCI} = 0.0265$, std. $\text{error}_{MCI,CN} = 0.0240$, std. $\text{error}_{MCI,AD} = 0.0143$, *and* std. $\text{error}_{AD,MCI} = 0.0516$] respectively. From the foregone, it is clear that judging from the standard error perspective, the hybrid model with 100 posterior estimates yields precise group estimates followed by that of the hybrid model with all 8000 posterior estimates, and finally the frequentist model. That is, in terms of estimating the group parameter, the proposed hybrid models outwit the frequentist model in performance with the model with 100 estimates recording the best performance.

Attention is now shifted to the precision of the estimation of the association parameter and one of the assessment indicators this thesis considers is the standard error. It could be seen from Tables 8.2 and 8.3 that the standard error of the association parameter α for transitions from CN to MCI, MCI to CN, MCI to AD and AD to MCI for the hybrid model with all 8000 posterior estimate are [std. $\text{error}_{CN,MCI} = 0.0023$, std. $\text{error}_{21} = 0.0008$, std. $\text{error}_{MCI,AD} = 0.0015$, std. $\text{error}_{32} = 0.0029$] respectively,

that of the hybrid model with the last 100 posterior estimates are [std. error$_{CN.MCI}$ = 0.0021, std. error$_{MCI,CN}$ = 0.0007, std. error$_{MCI,AD}$ = 0.0015, std. error$_{AD,MCI}$ = 0.0028] respectively and that of the frequentist approach are [std. error$_{CN,MCI}$ = 0.0730, std. error$_{MCI,CN}$ = 0.1076, std. error$_{MCI,AD}$ = 0.0672, std. error$_{AD,MCI}$ = 0.1912] respectively. In respect of the standard errors, it is evident that the hybrid models are estimating the association parameter more precisely compared to the frequentist model.

The next to discuss is the CL and CI which are all functions of the standard error and as it has already been indicated, they both have a direct proportional relationship with both the standard error. Hence not too surprisingly, a comparison of the 95% CL of all three models reveals the hybrid model with 100 posterior estimates has the most narrow CI and CL followed by the model with all 8000 posterior estimates and then the frequentist model. In an effort to authenticate the precision of the estimates obtained from all three models, the CI and CL were also considered.

It could be seen from Tables 8.2 and 8.3 that the CL of the group parameter γ for transitions from CN to MCI, MCI to CN, MCI to AD, and AD to MCI for the hybrid model with all 8000 posterior estimates are [CL$_{CN,MCI}$ = 0.0043, CL$_{MCI,CN}$ = 0.0012, CL$_{MCI,AD}$ = 0.0026, *and* CL$_{AD,MCI}$ = 0.0034] respectively, that of the hybrid model with the last 100 posterior estimates are [CL$_{CN,MCI}$ = 0.0034, CL$_{MCI,CN}$ = 0.0010, CL$_{MCI,AD}$ = 0.0021, *and* CL$_{AD,MCI}$ = 0.0028] and that of the frequentist approach is [CL$_{CN,MCI}$ = 0.0530, CL$_{MCI,CN}$ = 0.0480, CL$_{MCI,AD}$ = 0.0286, *and* CL$_{AD,MCI}$ = 0.1031]. From the foregone, it is clear that judging from the standard error perspective, the hybrid model with 100 posterior estimates yields precise group estimates followed by that of the hybrid model with all 8000 posterior estimates, and finally the frequentist model. That is, in terms of estimating the group parameter, the proposed hybrid models outwit the frequentist model in performance with the model with 100 estimates recording

the best performance.

In assessing the precision of the estimation of the association parameter, the CL was also considered. It could be seen from Tables 8.2 and 8.3 that the CL of the association parameter α for transitions from CN to MCI, MCI to CN, MCI to AD, and AD to MCI for the hybrid model with all 8000 posterior estimates are $[\alpha_{CN,MCI} = 0.0092,\ \alpha_{MCI,CN} = 0.0030,\ \alpha_{MCI,AD} = 0.0057,\ and\ \alpha_{AD,MCI} = 0.0113]$ respectively, that of the hybrid model with the last 100 posterior estimates are $[\alpha_{CN,MCI} = 0.0084,\ \alpha_{MCI,CN} = 0.0029,\ \alpha_{MCI,AD} = 0.0057,\ and\ \alpha_{AD,MCI} = 0.0110]$ and that of the frequentist approach are $[\alpha_{CN,MCI} = 0.0730,\ \alpha_{MCI,CN} = 0.1076,\ \alpha_{MCI,AD} = 0.0672,\ and\ \alpha_{AD,MCI} = 0.1912]$. Inferring from the above, it is apparent that from the standpoint of the CL, the hybrid model with 100 posterior estimates yields efficient association estimates followed by that of the hybrid model with all 8000 posterior estimates, and finally the frequentist model. That is, in terms of estimating the association parameter, the proposed hybrid model outperform the frequentist model and the model with 100 posterior estimates recorded the most efficient estimates.

Table 7.3: Multi-State Sub-Model Output for the Joint Hybrid with All 8000 Posterior Estimates and the Frequentist Model.

Measure	Parameter	States	JM Hybrid STS (All 8000 Posterior Est.)	JM Frequentist Approach
Estimate	γ_1	CN - MCI	1.0201	1.0200
	γ_1	MCI - CN	0.9409	0.9409
	γ_1	MCI - AD	1.0244	1.0244
	γ_1	AD - MCI	0.9766	0.9765
	α_1	CN - MCI	0.9453	0.9453
	α_1	MCI - CN	0.9841	0.9838
	α_1	MCI - AD	1.0350	1.0351
	α_1	AD - MCI	0.9571	0.9575
Std. Error				
	γ_1	CN - MCI	0.0011	0.0265
	γ_1	MCI - CN	0.0003	0.0240
	γ_1	MCI - AD	0.0007	0.0143
	γ_1	AD - MCI	0.0009	0.0516
	α_1	CN - MCI	0.0023	0.0365
	α_1	MCI- CN	0.0008	0.0538
	α_1	MCI - AD	0.0015	0.0336
	α_1	AD- MCI	0.0029	0.0956
95% CI				
	γ_1	CN - MCI	(1.0179, 1.0222)	(0.9938, 1.0468)
	γ_1	MCI - CN	(0.9403, 0.9415)	(0.9172, 0.9651)
	γ_1	MCI - AD	(0.9749, 0.9783)	(1.0102, 1.0388)
	γ_1	AD - MCI	(1.0231, 1.0257)	(0.9263, 1.0295)
	α_1	CN - MCI	(0.9407, 0.9499)	(0.9095, 0.9826)
	α_1	MCI - CN	(0.9826, 0.9856)	(-0.9314, 1.0391)
	α_1	MCI - AD	(1.0322, 1.0379)	(-1.0020, 1.0692)
	α_1	AD - MCI	(0.9514, 0.9627)	(-0.8667, 1.0578)

Table 7.4: 95% Confidence Length for the Multi-State Sub-Model Output for the Joint Hybrid with All 8000 Posterior Estimates and the Frequentist Model.

Measure	Parameter	States	JM Hybrid STS (All 8000 Posterior Est.)	JM Frequentist Approach
95% Confidence Length				
	γ_1	CN - MCI	0.0043	0.0530
	γ_1	MCI - CN	0.0012	0.0480
	γ_1	MCI- AD	0.0026	0.0286
	γ_1	AD - MCI	0.0034	0.1031
	α_1	CN - MCI	0.0092	0.0730
	α_1	MCI - CN	0.0030	0.1076
	α_1	MCI - AD	0.0057	0.0672
	α_1	AD - MCI	0.0113	0.1912

Table 7.5: Multi-State Sub-Model Output for the Joint Hybrid with the Last 100 Posterior Estimates and the Frequentist Model.

Measure	Parameter	States	JM Hybrid STS (Last 100 Posterior Est.)	JM Frequentist Approach
Estimate	γ_1	CN - MCI	1.0195	1.0200
	γ_1	MCI - CN	0.9407	0.9409
	γ_1	MCI - AD	1.0248	1.0244
	γ_1	AD - MCI	0.9763	0.9765
	α_1	CN - MCI	0.9453	0.9453
	α_1	MCI - CN	0.9836	0.9838
	α_1	MCI- AD	1.0354	1.0351
	α_1	AD - MCI	0.9580	0.9575
Std. Error				
	γ_1	CN - MCI	0.0009	0.0265
	γ_1	MCI - CN	0.0002	0.0240
	γ_1	MCI - AD	0.0005	0.0143
	γ_1	AD - MCI	0.0007	0.0516
	α_1	CN - MCI	0.0021	0.0365
	α_1	MCI - CN	0.0007	0.0538
	α_1	MCI - AD	0.0015	0.0336
	α_1	AD - MCI	0.0028	0.0956
95% CI				
	γ_1	CN - MCI	(1.0178, 1.0212)	(0.9938, 1.0468)
	γ_1	MCI - CN	(0.9402, 0.9412)	(0.9172, 0.9651)
	γ_1	MCI - AD	(1.0238, 1.0259)	(1.0102, 1.0388)
	γ_1	AD - MCI	(0.9749, 0.9777)	(0.9263, 1.0295)
	α_1	CN - MCI	(0.9411, 0.9495)	(0.9095, 0.9826)
	α_1	MCI - CN	(0.9822, 0.9851)	(-0.9314, 1.0391)
	α_1	MCI - AD	(1.0325, 1.0382)	(-1.0020, 1.0692)
	α_1	AD - MCI	(0.9525, 0.9635)	(-0.8667, 1.0578)

Table 7.6: 95% Confidence Length for Multi-State Sub-Model Output for the Joint Hybrid with the last 100 Posterior Estimates and the Frequentist Model.

Measure	Parameter	States	JM Hybrid STS (Last 100 Posterior Est.)	JM Frequentist Approach
95% Confidence Length				
	γ_1	CN - MCI	0.0034	0.0530
	γ_1	MCI - CN	0.0010	0.0480
	γ_1	MCI - AD	0.0021	0.0286
	γ_1	AD - MCI	0.0028	0.1031
	α_1	CN - MCI	0.0084	0.0730
	α_1	MCI - CN	0.0029	0.1076
	α_1	MCI - AD	0.0057	0.0672
	α_1	AD - MCI	0.0110	0.1912

7.6 Summary of Chapter

This chapter extended the novel two-stage hybrid model from time to a single event considered in Chapter Six of this thesis to time to multiple events. Three models, that is, the multi-state hybrid model with all 8000 posterior estimates, the multi-state hybrid model with the last 100 posterior estimates, and the frequentist model were considered. The precision of the estimates from the proposed hybrid model was compared to the frequentist model. As it is a theoretically established fact that the Bayesian model yields stable estimates after convergence, this study curiously considered a form of the multi-state hybrid model with the last 100 posterior predictive estimates from the first stage of the joint modelling process. All three models were empirically applied to the ADNI data. The performance of the proposed hybrid model as against the frequentist model was examined on the precision of the association (α) and group (γ) estimates of the models. In assessing the precision of the models, their standard errors, confidence intervals

(CI), and confidence lengths (CL) were considered.

Table 8.1 presented the results of the first stage of the modelling process and it was obvious that the hybrid model and the frequentist model yielded very similar estimates as well as relatively identical standard errors, CI and CL. The difference in standard error, CI, and CL between the hybrid model and the frequentist model were very marginal and did not arouse the curiosity of this thesis. The major concentration was on the multi-state sub-models as that is the actual focus of this study. The multi-state sub-model outputs for all three models are presented in Tables 8.3 and 8.4. It can be seen from both Tables 8.2 and 8.3 that the group parameter γ is most precisely estimated by the hybrid model with 100 posterior predictive estimates while the frequentist model yields the least precise estimates. Conversely, the association parameter α was also well estimated by the proposed hybrid models as compared to the frequentist model. The hybrid model with 100 posterior predictive estimates in this instance again, produced the most precise estimates followed by the hybrid model with all 8000 posterior estimates and then the frequentist model.

For all three models, the hybrid model with 100 posterior estimates, estimates both the group and association parameters relatively precisely compared to the hybrid model with all 8000 posterior estimates which comes next to it, and finally the frequentist model. Also, the standard errors of the estimates using the proposed hybrid models are very lower than the frequentist model, and hence the hybrid models have narrow CL compared to the frequentist model.

CHAPTER 8

SUMMARY, CONCLUSION, AND RECOMMENDATION

8.1 Summary

Current research has primarily focused on the joint modelling of a longitudinal outcome and a single survival outcome, although, data collected in practice will be more complicated most especially when multiple event outcomes occur. Therefore, it becomes crucial to extend the modelling framework to simultaneously address movements between multiple states of the phenomenon. However, the inclusion of movements among several states, rather than focusing on the time to a single event, can introduce additional computational challenge and complexity to the modelling framework.

Addressing these challenges requires careful attention to developing an effective estimation method that improves computational time and the precision of the estimates. With this goal in mind, this study proposes a hybrid two-stage estimation method for jointly fitting longitudinal and single event data, and extend it to multiple event data. This proposed approach fills a gap in the current research landscape by proposing a more comprehensive and effective way to handle longitudinal and multi-state data simultaneously.

This thesis has eight chapters beginning with Chapter One as the introduction which outlines the background of the study, presents a brief review of literature, situates the thesis in the context of the problem it sought to solve, states the objectives of this research, briefly describes the design of the research, justifies the significance of the thesis and finally wraps up with the way the thesis is

organized in terms of structure. Chapter Two introduces the data sets to test the proposed joint models while Chapter Three presents the review of literature which entails the fundamental concepts on the univariate and joint models this thesis hinges on. This chapter introduces the concept of LMM, survival models, MSMs, and joint models, which is the main vehicle driving the purpose of this thesis. It also presents some various forms of association structures linking both sub-models in joint models and zooms in to the opted association structure for this thesis. The chapter also discusses other existing joint models which the proposed hybrid model is compared to in a bid to validate it. Chapter Four, the heartbeat of this thesis presents the proposed hybrid joint model. After presenting the proposed hybrid model in Chapter Four, a simulation study was triggered and the results are reported in Chapter Five. Chapter Six of this thesis presents an empirical application of the fundamental concepts to the ADNI and PPMI data. The chapter also demonstrates the practical application of the suggested two-stage hybrid model using the ADNI data to explore the association between ADNI dynamics and transition intensities between disease states, and times of transitions among other interesting measures. Chapter Seven, extends the proposed joint hybrid model from time to a single event to time to multiple events and goes beyond to do a comparative assessment of the proposed hybrid model to the frequentist model. Chapter Eight, which is the last chapter of this thesis also presents the summary, conclusion and recommendations this study suggested.

8.2 Conclusion

In this thesis, a hybrid two-stage estimation method for jointly modelling longitudinal and survival, and a multi-state process is proposed. The proposed two-step modelling approach uses a Bayesian estimation for the submodel of the longitudinal process and then uses all the posterior predictive estimates from the first stage of the estimation process as inputs in stage two of the process. This

distinguishes it from much of the existing literature that relies on a single point estimate from phase one as input in phase two. The proposed estimation method is first applied to time to a single survival outcome and extended to a multi-state process. These models are validated using a simulation study and empirical data from the ADNI and PPMI studies. The proposed model is evaluated for its performance by comparing it to the RJA, BJS and BSTS, which are already established joint models, based on various metrics like bias, mean square error, confidence length, standard error, and confidence interval. The advantage of this approach to the transitional two-stage joint models lies in its ability to consider the variability of the first step estimation in the second step modelling.

The study conducted simulations to examine the performance of the suggested hybrid model when applied to time to a single event scenarios. The simulations were run on an 11th Generation HP laptop with a 2.40GHz Intel Core i5 processor, 16 GB RAM, and Windows OS. The frequentist models were fitted utilizing the jointModel function from the R package JM, while all the Bayesian models were implemented using stan. For the longitudinal submodel, the simulation results clearly demonstrate that the proposed hybrid model outperformed all other approaches in estimating the longitudinal submodel. It exhibited relatively less bias and provided more precise estimates, especially at smaller sample sizes. At larger sample sizes, N $>$ 500, the hybrid model, again, outperformed the BJS and the BSTS, in estimating the longitudinal submodel parameters, but had similar results to the RJA, which in literature has been acclaimed to be the recommended way to deal with the joint estimation process (Mauff et al. 2020). For the survival submodel, the performance of the hybrid model, in estimating the group (γ) and association (α) parameters, was significantly enhanced at larger sample sizes. On the score of 95% CI, 95% CL, bias and the mean square error the proposed hybrid model performed competitively well in relation to the other methods that has been touted in the literature of joint models and in some instances outperformed these existing methods. The results of the hybrid model

outperformed the BSTS approach, in terms of precision at larger sample sizes and were similar to the RJA and BJS.

In terms of computational time, the proposed hybrid model performed competitively well in relation to the RJA and outperformed the BJS and the BSTS models.

Surprisingly, the bias of the posterior distributions using the proposed hybrid methodology is marginally lower than other methods at larger sample sizes. It can be conveniently concluded that the simulation results revealed that the suggested hybrid approach performs well, both in estimating the group (γ) and association (α) parameters, especially when the sample size increases.

When the proposed hybrid model was extended to the MSM, it was evident from the study that it outperformed the frequentist model in terms of yielding precise estimates. Both the multi-state hybrid model and multi-state frequentist models were assessed on their standard errors, confidence intervals, and confidence lengths in ascertaining which model yields precise estimates. It is a theoretically established fact that the Bayesian models yield stable estimates after the model has converged and it is on the backdrop of this that the multi-state hybrid model was extended to incorporate the last 100 posterior predictive estimates from the Bayesian model. Meaning, this study presented two versions of the proposed hybrid model for the MSM, one with all the posterior predictive estimates from the Bayesian model and the other with the last 100 predictive posterior estimates. These two versions of the hybrid model were compared to the frequentist model and among all three models, the hybrid model with 100 posterior predictive estimates proved superior in terms of yielding precise estimates followed by the model with all 8000 posterior estimates, and finally the frequentist model. The difference between the standard errors of the hybrid models and the frequentist model is quite stricken with the hybrid models yielding far smaller standard errors. This renders the hybrid models to have narrow CL compared to the frequentist model. For all three models that were considered when the hybrid model was

extended to the MSM, the hybrid model with 100 posterior estimates, estimates both the group and association parameters relatively precisely compared to the hybrid model with all 8000 posterior estimates which comes next to it, and finally the frequentist model.

8.3 Recommendation

To speed up the computational time with more precise estimates, the hybrid model with less posterior predictive estimates at the tail end of the converged Bayesian model from the longitudinal sub-model is recommended. The simulation results have confirmed that the proposed hybrid approach performs admirably in estimating both the group (γ) and association (α) parameters when applied to time to a single event scenario, particularly as the sample size increases. Therefore, for larger sample sizes, the hybrid model is highly recommended, as it yields less biased and more precise estimates.

It is essential to acknowledge that in theory, the joint likelihood specification approach to joint modelling is generally preferred over the two-stage approach. However, in practical situations where the joint specification model becomes highly time-consuming due to its complexity or when issues with convergence of Markov chains arise due to a high-dimensional parameter space, alternative approaches like the hybrid approach proposed in this thesis are recommended.

This thesis proposes that future research in joint modelling of longitudinal and survival data should prioritize the development and refinement of joint models that incorporate more sophisticated techniques for accommodating complex data, taking into consideration the computational speed and convergence, as this thesis has set the tone with the hybrid two-stage estimation method. Again, as this thesis focused on simulation study when the hybrid model was applied to time to a single event, future research can also extend the empirical application this thesis adopted, when the hybrid model was extended to the multi-stage submodel, to a simulation scenario.

To ascertain the scope and effectiveness of the estimation method proposed in this thesis, it would be intriguing to apply the hybrid models to more intricate longitudinal data, such as skewed or multivariate longitudinal data, and survival data, such as competing-risks submodels. Additionally, exploring the extension to multivariate survival outcomes, which involves considering multiple failure times per subject, like recurrent events, could be valuable. Nonetheless, it should be noted that incorporating recurrent events might lead to increased computational complexity due to the additional submodel required.

8.4 References

Albert, P. S., & Shih, J. H. (2010). On estimating the relationship between longitudinal measurements and time-to-event data using a simple two-stage procedure. *Biometrics*, 66(3), 983-987.

Alsefri, M., Sudell, M., Garcia-Finana, M., Kolamunnage-Dona, R. (2020). Bayesian joint modelling of longitudinal and time to event data: A methodological review. *BMC Med. Res. Methodol*, (20), 1-17.

Alvarez, I., Niemi, J., & Simpson, M. (2014). *Bayesian inference for a covariance matrix*. arXiv preprint arXiv:1408.4050.

Andrinopoulou, E. R., Rizopoulos, D., Takkenberg, J. J., & Lesaffre, E. (2017). Combined dynamic predictions using joint models of two longitudinal outcomes and competing risk data. *Statistical methods in medical research*, 26(4), 1787-1801.

Aralis, H. J. (2016). Modeling Multistate Models with Back Transitions: Statistical Challenges and Applications.

Avery, M., Wu, Y., Helen Zhang, H., & Zhang, J. (2014). RKHS-based functional nonparametric regression for sparse and irregular longitudinal data. *Canadian Journal of Statistics*, 42(2), 204-216.

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2015). *lme4: Linear mixed-effects models using Eigen and S4: R package version 1.1-9* [software package].

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3), 255-278.

Brown, R. E., & G. Ibrahim, J. (2003). A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics*, 59(2), 221-228.

Burkner, P. C. (2013). Brms: An R package for Bayesian generalized linear

mixed models using Stan. *Journal of statistical software*, (80), 1.

Bycott, P., & Taylor, J. (1998). A comparison of smoothing techniques for CD4 data measured with error in a time-dependent Cox proportional hazards model. *Statistics in medicine*, 17(18), 2061-2077.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of statistical software*, 76(1).

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancour, M., ... & Ridell, A. (2015). pkg Stan: A Probabilistic Programming Language. *Journal of Statistical Software*.

Cho, H., Wang, P., & Qu, A. (2017). Personalize treatment for longitudinal data using unspecified random-effects model. *Statistica Sinica*, 187-206.

Chi, Y. Y., & Ibrahim, J. G. (2006). Joint models for multivariate longitudinal and multivariate survival data. *Biometrics*, 62(2), 432-445.

Ciapanna, E., & Taboga, M. (2019). Bayesian analysis of coefficient instability in dynamic regressions. *Econometrics*, 7(3), 29.

Collett, D. (2014). Sample size determination in survival analysis. *Wiley StatsRef: statistics reference online*.

Colosimo, E., Ferreira, F. V., Oliveira, M., & Sousa, C. (2002). Empirical comparisons between Kaplan-Meier and Nelson-Aalen survival function estimators. *Journal of Statistical Computation and Simulation*, 72(4), 299-308.

Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62(2), 269-276.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), 187-202.

Crane, P. K., Carle, A., Gibbons, L. E., Insel, P., Mackin, R. S., Gross, A., ... & Mungas, D. (2012). Development and assessment of a composite score for memory in the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Brain imaging and behavior*, 6(4), 502-516.

Cross, P. C., Heisey, D. M., Scurlock, B. M., Edwards, W. H., Ebinger, M. R., &

Brennan, A. (2010). Mapping brucellosis increases relative to elk density using hierarchical Bayesian models. *PLoS One*, 5(4), e10322.

Crowther, M. J., & Lambert, P. C. (2013). Simulating biologically plausible complex survival data. *Statistics in medicine*, 32(23), 4118-4134.

Culpepper, S. A. (2016). Revisiting the 4-parameter item response model: Bayesian estimation and application. *Psychometrika*, 81(4), 1142-1163.

Dafni, U. G., & Tsiatis, A. A. (1998). Evaluating surrogate markers of clinical outcome when measured with error. *Biometrics*, 1445-1462.

Damlen, P., Wakefield, J., & Walker, S. (1999). Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(2), 331-344.

Dantan, E., Joly, P., Dartigues, J. F., & Jacqmin-Gadda, H. (2011). Joint model with latent state for longitudinal and multistate data. *Biostatistics*, 12(4), 723-736.

Davis, C. S. (2002). *Statistical methods for the analysis of repeated measurements* (No. 04; QA278, D38.). New York: Springer.

DeGruttola, V., & Tu, X. (1994). Modeling the relationship between disease progression and survival time. *Biometrics*, 50, 1003-1014.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), 1-22.

Diggle, P., Diggle, P. J., Heagerty, P., Liang, K. Y., & Zeger, S. (2002). *Analysis of longitudinal data*. Oxford university press.

Diggle, P. J., Sousa, I., & Chetwynd, A. G. (2008). Joint modelling of repeated measurements and time-to-event outcomes: the fourth Armitage lecture. *Statistics in Medicine*, 27(16), 2981-2998.

Drikvandi, R., Verbeke, G., & Molenberghs, G. (2017). Diagnosing misspecification of the random-effects distribution in mixed models. *Biometrics*, 73(1), 63-71.

Duane, S., Kennedy, A. D., Pendleton, B. J., & Roweth, D. (1987). Hybrid monte carlo. *Physics letters B*, 195(2), 216-222.

Eager, C., & Roy, J. (2017). Mixed effects models are sometimes terrible. *arXiv preprint arXiv:1701.04858*.

Eulenburg, C., Schroeder, J., Obi, N., Heinz, J., Seibold, P., Rudolph, A.,... & Flesch-Janys, D. (2016). A comprehensive multistate model analyzing associations of various risk factors with the course of breast cancer in a population-based cohort of breast cancer cases. *American journal of epidemiology*, 183(4), 325-334.

Farewell, V. T., Long, D. L., Tom, B. D. M., Yiu, S., & Su, L. (2017). Two-part and related regression models for longitudinal data. *Annual review of statistics and its application*, 4, 283.

Faucett, C. L., & Thomas, D. C. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in medicine*, 15(15), 1663-1685.

Ferrer, L., Rondeau, V., Dignam, J., Pickles, T., Jacqmin-Gadda, H., & Proust-Lima, C. (2016). Joint modelling of longitudinal and multi-state processes: application to clinical progressions in prostate cancer. *Statistics in medicine*, 35(22), 3933-3948.

Fitzmaurice, G., Davidian, M., Verbeke, G., & Molenberghs, G. (Eds.). (2008). *Longitudinal data analysis*. CRC press.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2012). *Applied longitudinal analysis* (Vol. 998). John Wiley & Sons.

Garcia, T. P., Ma, Y., Marder, K., & Wang, Y. (2017). Robust mixed effects model for clustered failure time data: Application to Huntington's disease event measures. *The annals of applied statistics*, 11(2), 1085.

Gelfand, A. E., & Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410), 398-409.

Gelman, A. (2013). Commentary: P values and statistical practice. *Epidemiology*, 24(1), 69-72.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014a). *Bayesian data analysis* (Vol. 2).

Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of research on educational effectiveness*, 5(2), 189-211.

Gelman, A.,& Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and computing*, 24(6), 997-1016.

Gelman, A., Lee, D., & Guo, J. (2015). Stan: A probabilistic programming language for Bayesian inference and optimization. *Journal of Educational and Behavioral Statistics*, 40(5), 530-543.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6), 721-741.

Guler, I., Faes, C., Cadarso-Suarez, C., Teixeira, L., Rodrigues, A., & Mendonca, D. (2017). Two-stage model for multivariate longitudinal and survival data with application to nephrology research. *Biometrical Journal*, 59(6), 1204-1220.

Ha, I. D., & Lee, Y. (2021). A review of h-likelihood for survival analysis. *Japanese Journal of Statistics and Data Science*, 4(2), 1157-1178.

Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of statistical software*, 33, 1-22.

Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D. N., Goodwin, C. E., ... & Inger, R. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*, 6, e4794.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications.

Heck, D. W., Boehm, U., Boing-Messing, F., Burkner, P. C., Derks, K., Dienes, Z., ... & Hoijtink, H. (2022). A review of applications of the bayes factor in psychological research. *Psychological Methods*.

Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal data analysis*. Wiley-Interscience.

Herring, A. H. (2013). *Applied Longitudinal Analysis, by Garrett M. Fitzmaurice, Nan M.*

Henderson, R., Diggle, P., & Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(4), 465-480.

Hickey, G. L., Philipson, P., Jorgensen, A., & Kolamunnage-Dona, R. (2018). A comparison of joint models for longitudinal and competing risks data, with application to an epilepsy drug randomized controlled trial. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 181(4), 1105-1123.

Hickey, G. L., Philipson, P., Jorgensen, A., & Kolamunnage-Dona, R. (2016). Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC medical research methodology*, 16(1), 1-15.

Hoekstra, R., Morey, R. D., Rouder, J. N., & Wagenmakers, E. J. (2014). Robust misinterpretation of confidence intervals. *Psychonomic bulletin & review*, 21(5), 1157-1164.

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.*, 15(1), 1593-1623.

Hougaard, P. (1999). Multi-state models: a review. *Lifetime data analysis*, 5(3), 239-264.

Hsieh, F., Tseng, Y. K., & Wang, J. L. (2006). Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics*, 62(4), 1037-1043.

Hunter, D. R., & Lange, K. (2004). A tutorial on MM algorithms. *The American*

Statistician, 58(1), 30-37.

Huong, P. T. T., Nur, D., Pham, H., & Branford, A. (2018). A modified two-stage approach for joint modelling of longitudinal and time-to-event data. *Journal of Statistical Computation and Simulation*, 88(17), 3379-3398.

Huong, P. T. T., Nur, D., & Branford, A. (2017). Penalized spline joint models for longitudinal and time-to-event data. *Communications in Statistics-Theory and Methods*, 46(20), 10294-10314.

Huang, X., Li, G., Elashoff, R. M., & Pan, J. (2011). A general joint model for longitudinal measurements and competing risks survival data with heterogeneous random effects. *Lifetime data analysis*, 17(1), 80-100.

Ibrahim, J. G., Chu, H., & Chen, L. M. (2010). Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology*, 28(16), 2796.

Iddi, S., Li, D., Aisen, P. S., Rafii, M. S., Litvan, I., Thompson, W. K., & Donohue, M. C. (2018). Estimating the evolution of disease in the Parkinson's progression markers initiative. *Neurodegenerative Diseases*, 18(4), 173-190.

Iddi, S., & Molenberghs, G. (2012). A combined overdispersed and marginalized multilevel model. *Computational Statistics & Data Analysis*, 56(6), 1944-1951.

Jacobi, H., du Montcel, S. T., Bauer, P., Giunti, P., Cook, A., Labrum, R., ... & Klockgether, T. (2015). Long-term disease progression in spinocerebellar ataxia types 1, 2, 3, and 6: a longitudinal cohort study. *The Lancet Neurology*, 14(11), 1101-1108.

Jagust, W. J., Landau, S. M., Koeppe, R. A., Reiman, E. M., Chen, K., Mathis, C. A., ... & Wang, A. Y. (2015). The Alzheimer's disease neuroimaging initiative 2 PET core: 2015. *Alzheimer's & Dementia*, 11(7), 757-771.

Jensen, S. M., & Ritz, C. (2018). A comparison of approaches for simultaneous inference of fixed effects for multiple outcomes using linear mixed models. *Statistics in medicine*, 37(16), 2474-2486.

Jiang, Y., Li, J., Schmitt, F. A., Jicha, G. A., Munro, N. B., Zhao, X., ... &

Abner, E. L. (2021). Memory-related frontal brainwaves predict transition to mild cognitive impairment in healthy older individuals five years before diagnosis. *Journal of Alzheimer's Disease*, 79(2), 531-541. Kalbfleisch, J. D., & Prentice, R. L. (2011). The statistical analysis of failure time data. John Wiley & Sons.

Kalbfleisch, J. D., & Prentice, R. L. (1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika*, 60(2), 267-278.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282), 457-481.

Kass, R. E., & Natarajan, R. (2006). A default conjugate prior for variance components in generalized linear mixed models (comment on article by Browne and Draper). *Bayesian Analysis*, 1(3), 535-542.

Kery, M. (2010). *Introduction to WinBUGS for ecologists: Bayesian approach to regression, ANOVA, mixed models and related analyses.* Academic Press.

Koster, J., & McElreath, R. (2017). Multinomial analysis of behavior: statistical methods. *Behavioral Ecology and Sociobiology*, 71(9), 1-14.

Krol, A., Mauguen, A., Mazroui, Y., Laurent, A., Michiels, S., & Rondeau, V. (2017). Tutorial in joint modeling and prediction: a statistical software for correlated longitudinal outcomes, recurrent events and a terminal event. *arXiv preprint arXiv:1701.03675*.

Kruschke, J. K. (2018). Rejecting or accepting parameter values in Bayesian estimation. *Advances in methods and practices in psychological science*, 1(2), 270-280.

Kruschke, J. K.,& Liddell, T. M. (2018). Bayesian data analysis for newcomers. *Psychonomic bulletin and review*, 25(1), 155-177.

Kruschke, J. K., & Vanpaemel, W. (2015). Bayesian estimation in hierarchical models. *The Oxford handbook of computational and mathematical psychology*, 279-299.

Laird, and James H. Ware, John Wiley & Sons, 2011: ISBN 978-0-470-38027-7, 740 pp., $125. Laird, N. M., & Ware, J. H. (1982). Random-effects models for

longitudinal data. *Biometrics*, 963-974.

Lazaro, E., Armero, C., & Alvares, D. (2021). Bayesian regularization for flexible baseline hazard functions in Cox survival models. *Biometrical Journal*, 63(1), 7-26.

Lazaro, E., Armero, C., & Gomez-Rubio, V. (2020). Approximate Bayesian inference for mixture cure models. *TEST*, 29(3), 750-767.

Lee, E. T., & Go, O. T. (1997). Survival analysis in public health research. *Annual review of public health*, 18, 105.

Lee, Y., & Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(4), 619-656.

Leiva-Yamaguchi, V., & Alvares, D. (2020). A two-stage approach for Bayesian joint models of longitudinal and survival data: Correcting bias with informative prior. *Entropy*, 23(1), 50.

Li, N., Elashoff, R. M., Li, G., & Saver, J. (2010). Joint modeling of longitudinal ordinal data and competing risks survival times and analysis of the NINDS rt-PA stroke trial. *Statistics in medicine*, 29(5), 546-557.

Lin, H., McCulloch, C. E., & Mayne, S. T. (2002). Maximum likelihood estimation in the joint analysis of time-to-event and multiple longitudinal variables. *Statistics in Medicine*, 21(16), 2369-2382.

Lindstrom, M. J., & Bates, D. M. (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association*, 83(404), 1014-1022.

Liu, L. (2009). Joint modeling longitudinal semi-continuous data and survival, with application to longitudinal medical cost data. *Statistics in medicine*, 28(6), 972-986.

Locascio, J. J., & Atri, A. (2011). An overview of longitudinal data analysis methods for neurological research. *Dementia and geriatric cognitive disorders extra*, 1(1), 330-357.

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS-a

Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and computing*, 10(4), 325-337.

Luo, S., Lawson, A. B., He, B., Elm, J. J., & Tilley, B. C. (2016). Bayesian multiple imputation for missing multivariate longitudinal data from a Parkinson's disease clinical trial. *Statistical Methods in Medical Research*, 25(2), 821-837.

Luo, S., Lawson, A. B., He, B., Elm, J. J., & Tilley, B. C. (2016). Bayesian multiple imputation for missing multivariate longitudinal data from a Parkinson's disease clinical trial. *Statistical Methods in Medical Research*, 25(2), 821-837.

Luo, S. (2014). A Bayesian approach to joint analysis of multivariate longitudinal data and parametric accelerated failure time. *Statistics in medicine*, 33(4), 580-594.

Luts, J., Molenberghs, G., Verbeke, G., Van Huffel, S., & Suykens, J. A. (2012). A mixed effects least squares support vector machine model for classification of longitudinal data. *Computational Statistics & Data Analysis*, 56(3), 611-628.

Maier, R., Moser, G., Chen, G. B., Ripke, S., Absher, D., Agartz, I., ... & Kennedy, J. L. (2015). Joint analysis of psychiatric disorders increases accuracy of risk prediction for schizophrenia, bipolar disorder, and major depressive disorder. *The American Journal of Human Genetics*, 96(2), 283-294.

Man, K., Harring, J. R., Jiao, H., & Zhan, P. (2019). Joint modeling of compensatory multidimensional item responses and response times. *Applied Psychological Measurement*, 43(8), 639-654.

Marek, K., Chowdhury, S., Siderowf, A., Lasch, S., Coffey, C. S., Caspell-Garcia, C., ... & Larsen, L. (2018). The Parkinson's progression markers initiative (PPMI)-establishing a PD biomarker cohort. *Annals of clinical and translational neurology*, 5(12), 1460-1477.

Marek, K., Jennings, D., Lasch, S., Siderowf, A., Tanner, C., Simuni, T., ... & Parkinson Progression Marker Initiative. (2011). The Parkinson progression marker initiative (PPMI). *Progress in neurobiology*, 95(4), 629-635.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017).

Balancing Type I error and power in linear mixed models. *Journal of memory and language*, 94, 305-315.

Mauff, K., Steyerberg, E., Kardys, I., Boersma, E., & Rizopoulos, D. (2020). Joint models with multiple longitudinal outcomes and a time-to-event outcome: a corrected two-stage approach. *Statistics and Computing*, 30(4), 999-1014.

Meira-Machado, L., de Una-Alvarez, J., Cadarso-Suarez, C., & Andersen, P. K. (2009). Multi-state models for the analysis of time-to-event data. *Statistical methods in medical research*, 18(2), 195-222.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6), 1087-1092.

Molenberghs, G., & Verbeke, G. (2006). Longitudinal data analysis. *Wiley StatsRef: Statistics Reference Online*, 1-28.

Moore, D. F. (2016). *Applied survival analysis using R* (Vol. 473). New York, NY: Springer.

Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E. J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic bulletin & review*, 23(1), 103-123.

Murray, J., & Philipson, P. (2022). A fast approximate EM algorithm for joint models of survival and multivariate longitudinal data. *Computational Statistics & Data Analysis*, 170, 107438.

Muthen, B., Asparouhov, T., Boye, M., Hackshaw, M., & Naegeli, A. (2009). *Applications of continuous-time survival in latent variable models for the analysis of oncology randomized clinical trial data using Mplus*. Los Angeles, CA: Muthen & Muthen.

Murawska, M., Rizopoulos, D., & Lesaffre, E. (2012). A two-stage joint model for nonlinear longitudinal response and a time-to-event with application in transplantation studies. *Journal of Probability and Statistics*, 2012.

Nalborczyk, L., Batailler, C., Loevenbruck, H., Vilain, A., & BÃ¼rkner, P. C.

(2019). An introduction to Bayesian multilevel models using brms: A case study of gender effects on vowel variability in standard Indonesian. *Journal of Speech, Language, and Hearing Research*, 62(5), 1225-1242.

Natarajan, R., & Kass, R. E. (2000). Reference Bayesian methods for generalized linear mixed models. *Journal of the American Statistical Association*, 95(449), 227-237.

Neal, R. M. (2011). MCMC using Hamiltonian dynamics. *Handbook of markov chain monte carlo*, 2(11), 2.

Neal, R. M. (2003). Slice sampling. *The annals of statistics*, 31(3), 705-767.

Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics*, 14(4), 945-966.

Newsom, J. T. (2015). *Longitudinal structural equation modeling: A comprehensive introduction*. Routledge.

Ogden, H. (2014). Robustness properties of marginal composite likelihood estimators. *arXiv preprint arXiv:1401.1383*.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., ... & Rakow, T. (2006). *Uncertain judgements: eliciting experts' probabilities*.

Peto, R., & Peto, J. (1972). Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society: Series A (General)*, 135(2), 185-198.

Philipson, P., Hickey, G. L., Crowther, M. J., & Kolamunnage-Dona, R. (2020). Faster Monte Carlo estimation of joint models for time-to-event and multivariate longitudinal data. *Computational Statistics & Data Analysis*, 151, 107010.

Pinheiro, J. C., & Bates, D. M. (2000). Linear mixed-effects models: basic concepts and examples. *Mixed-effects models in S and S-Plus*, 3-56.

Pinheiro, J. C., & Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of computational and Graphical Statistics*, 4(1), 12-35.

Plummer, M. (2012). JAGS Version 3.3. 0 user manual.

Prentice, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*, 69(2), 331-342.

Prentice, R. L., Kalbfleisch, J. D., Peterson Jr, A. V., Flournoy, N., Farewell, V. T., & Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, 541-554.

Proust-Lima, C., Sene, M., Taylor, J. M., & Jacqmin-Gadda, H. (2014). Joint latent class models for longitudinal and time-to-event data: a review. *Statistical methods in medical research*, 23(1), 74-90.

Rabe-Hesketh, S., & Skrondal, A. (2008). Multilevel and longitudinal modeling using Stata. *STATA press*.

Ren, J., & Gui, W. (2021). Statistical analysis of adaptive type-II progressively censored competing risks for Weibull models. *Applied Mathematical Modelling*, 98, 323-342.

Ren, K., Qin, J., Zheng, L., Yang, Z., Zhang, W., Qiu, L., & Yu, Y. (2019, July). Deep recurrent survival analysis. *In Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 33, No. 01, pp. 4798-4805).

Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data: With applications in R*. CRC press.

Rizopoulos, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, 67(3), 819-829.

Rizopoulos, D. (2010). JM: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software*, 35, 1-33.

Rizopoulos, D., & Ghosh, P. (2011). A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in medicine*, 30(12), 1366-1380.

Rizopoulos, D., Hatfield, L. A., Carlin, B. P., & Takkenberg, J. J. (2014). Combining dynamic predictions from joint models for longitudinal and time-to-event data using Bayesian model averaging. *Journal of the American Statistical*

Association, 109(508), 1385-1397.

Rizopoulos, D., Verbeke, G., & Molenberghs, G. (2010). Multiple-imputation-based residuals and diagnostic plots for joint models of longitudinal and survival outcomes. *Biometrics*, 66(1), 20-29.

Rogers, S. N., Brown, J. S., Woolgar, J. A., Lowe, D., Magennis, P., Shaw, R. J.,...& Vaughan, D. (2009). Survival following primary surgery for oral cancer. *Oral oncology*, 45(3), 201-211.

Tsiatis, A. A., & Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, 809-834.

Sattar, A., & Sinha, S. K. (2019). Joint modeling of longitudinal and survival data with a covariate subject to a limit of detection. *Statistical methods in medical research*, 28(2), 486-502.

Scharfstein, D. O., Rotnitzky, A., & Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, 94(448), 1096-1120.

Scholkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., & Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. *Neural computation*, 13(7), 1443-1471.

Schluchter, M. D. (1992). Methods for the analysis of informatively censored longitudinal data. *Statistics in medicine*, 11(14-15), 1861-1870.

Schuurman, N. K., Grasman, R. P. P. P., & Hamaker, E. L. (2016). A comparison of inverse-wishart prior specifications for covariance matrices in multilevel autoregressive models. *Multivariate Behavioral Research*, 51(2-3), 185-206.

Scott, J. G., & Berger, J. O. (2010). Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. *The Annals of Statistics*, 2587-2619.

Self, S., & Pawitan, Y. (1992). Modeling a marker of disease progression and onset of disease. *In AIDS epidemiology* (pp. 231-255). BirkhÃ¤user, Boston,

MA.

Simuni, T., Caspell-Garcia, C., Coffey, C. S., Weintraub, D., Mollenhauer, B., Lasch, S., ... & Marek, K. (2018). Baseline prevalence and longitudinal evolution of non-motor symptoms in early Parkinson's disease: the PPMI cohort. *Journal of Neurology, Neurosurgery & Psychiatry*, 89(1), 78-88.

Slama, R., Eustache, F., Ducot, B., Jensen, T. K., JÃ¸rgensen, N., Horte, A., ... & Jouannet, P. (2002). Time to pregnancy and semen parameters: a cross-sectional study among fertile couples from four European cities. *Human Reproduction*, 17(2), 503-515.

Song, X., Davidian, M., & Tsiatis, A. A. (2002). A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics*, 58(4), 742-753.

Sorensen, T., & Vasishth, S. (2015). Bayesian linear mixed models using Stan: A tutorial for psychologists, linguists, and cognitive scientists. *arXiv preprint arXiv:1506.06201*.

Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2007). OpenBUGS user manual. *Version*, 3(2), 2007.

Spiegelhalter, D. J., Thomas, A., Best, N., & Lunn, D. (2003). *WinBUGS version 1.4 user manual*. MRC Biostatistics Unit, Cambridge. URL http://www.mrc-bsu. cam. ac. uk/bugs.

Stefanski, L. A., & Carroll, R. J. (1987). Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika*, 74(4), 703-716.

Sweeting, M. J., & Thompson, S. G. (2011). Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. *Biometrical Journal*, 53(5), 750-763.

Team, R. C. (2000). *R language definition*. Vienna, Austria: R foundation for statistical computing.

Thijs, H., Molenberghs, G., & Verbeke, G. (2000). The milk protein trial: influence analysis of the dropout process. *Biometrical Journal: Journal of*

Mathematical Methods in Biosciences, 42(5), 617-646.

Tsiatis, A. A., & Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, 809-834.

Tsiatis, A. A., & Davidian, M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika*, 88(2), 447-458.

Tsiatis, A. A., Degruttola, V., & Wulfsohn, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association*, 90(429), 27-37.

Van Houwelingen, H. C. (2000). Validation, calibration, revision and combination of prognostic survival models. *Statistics in medicine*, 19(24), 3401-3415.

Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of phonetics*, 71, 147-161.

Verbeke, G., Fieuws, S., Molenberghs, G., & Davidian, M. (2014). The analysis of multivariate longitudinal data: a review. *Statistical methods in medical research*, 23(1), 42-59.

Vonesh, E., & Chinchilli, V. M. (1997). *Linear and Non-Linear Models for the Analysis of Repeated Measurements*. Marcel Decker. Inc, New York, NY.

Viviani, S., Alfo, M., & Rizopoulos, D. (2014). Generalized linear mixed joint model for longitudinal and survival outcomes. *Statistics and Computing*, 24(3), 417-427.

Wienke, A. (2010). Frailty models in survival analysis. Chapman and Hall/CRC.

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., ... & Alzheimer's Disease Neuroimaging Initiative. (2017). The Alzheimer's Disease Neuroimaging Initiative 3: Continued innovation for clinical trial improvement. *Alzheimer's & Dementia*, 13(5), 561-571.

Weiner, M. W., & Veitch, D. P. (2015). Introduction to special issue: overview

of Alzheimer's Disease Neuroimaging Initiative. *Alzheimer's & Dementia*, 11(7), 730-733.

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., ... & Trojanowski, J. Q. (2013). The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. *Alzheimer's & Dementia*, 9(5), e111-e194.

Wu, C. O., Tian, X., & Yu, J. (2010). Nonparametric estimation for time-varying transformation models with longitudinal data. *Journal of Nonparametric Statistics*, 22(2), 133-147.

Wu, L., Liu, W., Yi, G. Y., & Huang, Y. (2012). Analysis of longitudinal and survival data: joint modeling, inference methods, and issues. *Journal of Probability and Statistics*, 2012.

Wu, S., & Xu, X. (2016). A study of three intrinsic problems of the classic discrete element method using flat-joint model. *Rock Mechanics and Rock Engineering*, 49(5), 1813-1830.

Wu, Y. H., Gao, S. H., Mei, J., Xu, J., Fan, D. P., Zhang, R. G., & Cheng, M. M. (2021). Jcs: An explainable covid-19 diagnosis system by joint classification and segmentation. *IEEE Transactions on Image Processing*, 30, 3113-3126.

Wulfsohn, M. S., & Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, 330-339.

Xu, J., & Zeger, S. L. (2001). Joint analysis of longitudinal data comprising repeated measures and times to events. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(3), 375-387.

Ye, W., Lin, X., & Taylor, J. M. (2008). Semiparametric modeling of longitudinal measurements and time-to-event data-a two-stage regression calibration approach. *Biometrics*, 64(4), 1238-1246.

Yuen, H. P., & Mackinnon, A. (2016). Performance of joint modelling of time-to-event data with time-dependent predictors: an assessment based on transition to psychosis data. *PeerJ*, 4, e2582.

Zhang, Z., Li, J., Fukumoto, F., & Ye, Y. (2021). Abstract, Rationale, Stance: A Joint Model for Scientific Claim Verification. *arXiv preprint arXiv:2110.15116*.

Zhang, H., & Wu, L. (2019). Joint model of accelerated failure time and mechanistic nonlinear model for censored covariates, with application in HIV/AIDS. *The Annals of Applied Statistics*, 13(4), 2140-2157.

Zeger, S. L., Liang, K. Y., & Albert, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, 1049-1060.