UNIVERSITY OF CALIFORNIA

Los Angeles

New Approaches to Joint Modeling of Longitudinal and Time-to-Event Outcomes: with Applications to Dynamic Prediction of Health Outcomes Using Massive Biobank Data

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Biostatistics

by

Shanpeng Li

2023

ABSTRACT OF THE DISSERTATION

New Approaches to Joint Modeling of Longitudinal and Time-to-Event Outcomes: with Applications to Dynamic Prediction of Health Outcomes Using Massive Biobank Data

by

Shanpeng Li
Doctor of Philosophy in Biostatistics
University of California, Los Angeles, 2023
Professor Gang Li, Chair

It is often of interest to study the temporal patterns of longitudinal biomarker(s) that are potentially correlated and predictive of time-to-event outcomes in biomedical studies. In this dissertation, I develop new approaches for joint modeling of longitudinal and time-to-event data. My dissertation consists of three projects. In Chapter 2, I develop customized linear scan algorithms to speed up the computation of semi-parametric joint models by linearizing the computational burden of the estimation procedure from $O(n^2)$ or $O(n^3)$ to $O(n)$. Compared to the existing software and packages on semi-parametric joint models, our implementations can provide more than thousands of speed-ups when the sample size goes large. In Chapter 3, motivated by the Multi-Ethnic Study of Atherosclerosis (MESA), I propose a novel joint model to account for the heterogeneity of within-subject variability of a longitudinal outcome and demonstrate that it improves the dynamic prediction accuracy of predicting the future event probabilities of both heart failure and death across MESA individuals. In Chapter 4, I extend the joint model described in Chapter 3 to handle interval-censored covariates as missing data due to the unknown initial event time. Using age at diagnosis of

diabetes as an interval-censored covariate, we revisit the UK-Biobank data to illustrate that our proposed joint model can yield clinically meaningful parameter estimates, compared to the existing methods such as midpoint imputation, which can lead to problematic conclusion on the effect of covariates on the outcomes.

The dissertation of Shanpeng Li is approved.

Zuo-Feng Zhang

Donatello Telesca

Hua Zhou

Gang Li, Committee Chair

University of California, Los Angeles

2023

To my parents and professors

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express my sincerest gratitude to Professor Gang Li, my advisor for his dedicated encouragement, patience, and insightful guidance. He has always been generous with this time whenever I need assistance on statistical methodology and manuscript writing. He also helped me to improve my writing style to polish my dissertation as more readable and well-organized. He provided financial support on my research assistance-ship through Biostatistics Department.

I am also very grateful to Professor Jin Zhou for providing fruitful resources on application studies and useful suggestions on data analyses in my dissertation. She provided comprehensive guidelines to help me analyze the application studies with in-depth data exploration and clinically meaningful data interpretation.

I also would like to give thanks to Dr. Ning Li for providing technical support on implementing the estimation procedure of joint modeling of longitudinal and time-to-event data. My interaction with her helped me to have a solid understanding of the principles and computational aspects of joint modeling.

Special thanks go to the other members of my dissertation committee: Professor Hua Zhou, Donatello Telesca, and Zuo-Feng Zhang.

Finally, I want to thank my friends Mr. Jonathan Tsang and Wilson Lee for providing insightful suggestions on my presentation and offering mental support on my dissertation preparation.

VITA

2012 - 2017 Bachelor of Science, Statistics & Economics, South China Normal University, China

2017 - 2019 Master of Science, Biostatistics, University of California, Los Angeles, USA

PUBLICATIONS

Kimberly L. Yan, **Shanpeng Li**, Chi-Hong Tseng, Jiyoon Kim, Dalena T. Nguyen, Nardeen B. Dawood, Masha J. Livhits, Michael W. Yeh, and Angela M. Leung."Rising incidence and incidence-based mortality of thyroid cancer in California, 2000-2017." The Journal of Clinical Endocrinology & Metabolism 105, no. 6 (2020): 1770-1777.

Ellingson, Benjamin M., Patrick Y. Wen, Susan M. Chang, Martin van den Bent, Michael A. Vogelbaum, Gang Li, **Shanpeng Li** et al. "Objective response rate (ORR) targets for recurrent glioblastoma clinical trials based on the historic association between ORR and median overall survival." Neuro-oncology (2023).

Ning Li, Yi Liu, **Shanpeng Li**, Robert M. Elashoff, and Gang Li. "A flexible joint model for multiple longitudinal biomarkers and a time-to-event outcome: With applications to dynamic prediction using highly correlated biomarkers." Biometrical Journal 63, no. 8 (2021): 1575-1586.

Shanpeng Li, Ning Li, Hong Wang, Jin Zhou, Hua Zhou, and Gang Li. "Efficient Algorithms and Implementation of a Semiparametric Joint Model for Longitudinal and Compet-

ing Risk Data: With Applications to Massive Biobank Data." Computational and Mathematical Methods in Medicine 2022 (2022).

Na Eun Kim, Rajam S. Raghunathan, Elena G. Hughes, Xochitl R. Longstaff, Chi-Hong Tseng, **Shanpeng Li**, Dianne S. Cheung et al."Bethesda III and IV Thyroid Nodules Managed Nonoperatively after Molecular Testing with Afirma GSC or Thyroseq v3." The Journal of Clinical Endocrinology & Metabolism (2023): dgad181.

Shanpeng Li, Robyn McClelland, Peter D. Reaven, Jin Zhou, Hua Zhou, and Gang Li. "A joint model of the individual mean and within-subject variability of a longitudinal outcome with a competing risks time-to-event outcome." arXiv preprint arXiv:2301.06584 (2023).

Eileen F. Shiuan, Lu Sun, Jeremy Reynoso, **Shanpeng Li**, Gang Li, Jiyoon Kim, Won Kim, and Robert Prins. "Abstract B032: Immune landscape of resected brain metastases in patients treated with and without immune checkpoint blockade immunotherapy." Cancer Research 83, no. 2 Supplement (2023): B032-B032.

Arnold I. Chin, Alexandra Drakaki, Sara Rodriguez, Ankush Sachadeva, Margaret Bradley, Nazy Zomorodian, Hanwei Zhang, **Shanpeng Li** et al. "Interim results of a phase Ib single-center study of pembrolizumab in combination with chemotherapy in patients with locally advanced or metastatic small cell/neuroendocrine cancers of the prostate and urothelium." (2022): 510-510.

Hong Wang, Ning Li, **Shanpeng Li**, and Gang Li. "JMcmprsk: An R Package for Joint Modelling of Longitudinal and Survival Data with Competing Risks." R J. 13, no. 1 (2021): 53.

CHAPTER 1

Introduction

My dissertation consists of three projects: 1) efficient algorithms and implementation of a semi-parametric joint model of longitudinal and competing risks time-to-event data; 2) a joint model of the individual mean and within-subject variability of a longitudinal outcome with a competing risks time-to-event outcome; 3) an efficient and unbiased estimator of joint modeling in the presence of interval-censored covariates due to the unknown initial event time. Below gives the introduction of the summary of the three projects.

1.1 Efficient algorithms and implementation of a semi-parametric joint model of longitudinal and competing risks time-to-event data

In project 1, we will study efficient algorithms and implementation of a semi-parametric joint model of longitudinal and competing risks time-to-event data. In clinical research and other longitudinal studies, it is common to collect both longitudinal and time-to-event data on each participant, and these two endpoints are often correlated. Joint models of longitudinal and survival data have been widely used to mitigate incorrect estimation and statistical inferences associated with separate analysis of each endpoint [Elashoff et al., 2017, Rizopoulos, 2010]. For instance, in longitudinal data analysis, joint models are often used to handle nonignorable missing data due to a terminal event, which cannot be properly accounted for by a standard mixed effects model or generalized estimating equation (GEE) method that relies on the

ignorable missing-at-random or missing-completely-at-random assumption [Henderson et al., 2000, Elashoff et al., 2008, Sattar and Sinha, 2019]. Joint models are also popularly employed in survival analysis to study the effects of a time-dependent covariate that is measured intermittently or subject to measurement error [Tsiatis and Davidian, 2004, Wang, 2006, Song et al., 2002, Crowther et al., 2013], and for dynamic prediction of an event outcome from the past history of a biomarker [Yu et al., 2008, Barrett and Su, 2017, Proust-Lima et al., 2015, Rizopoulos, 2011, Garre et al., 2008]. Comprehensive reviews of joint models for longitudinal and time-to-event data and their applications can be found in [Elashoff et al., 2017, Rizopoulos, 2012b, Hickey et al., 2018], [Sudell et al., 2016] and the references therein.

Despite the explosive growth of literature on joint models for longitudinal and time-to-event data during the past three decades, efficient implementation of joint models has lagged behind, which limits the application of joint models to only small to moderate studies. Recently, massive sample size data collected from electronic health records (EHR) and insurance claim databases warrants great opportunities to conduct clinical studies in a real-world setting. For example, the UK Biobank is a prospective cohort study with approximately 500,000 individuals, aged 37-73 years, from the general population between 2006 and 2010 in the United Kingdom [Collins, 2012, Sudlow et al., 2015]. Aggregated and quality controlled EHR data purchasable from Optum [`https://www.optum.com/EHR`] includes 80+ millions patients with longitudinal lab measures. The Million Veteran Project [Gaziano et al., 2016], and IBM MarketScan Database [Butler et al., 2021] are some of many other big biobank databases that contain rich yet complex longitudinal and time-to-event data on $10^5 \sim 10^8$ patients. However, current implementations of many joint models are inefficient and not scalable to the size of biobank data as demonstrated later in Sections 2.2 and 2.3. There is a pressing need to develop efficient implementations of joint models to enable the analysis of these rich data sources.

The purpose of this paper is to develop and implement efficient algorithms for a semiparametric joint model of longitudinal and competing risks time-to-event data. As specified

in Section 2.1.1, the joint model consists of a linear mixed effects sub-model for a longitudinal outcome and a semiparametric proportional cause-specific hazards sub-model for a competing risks survival outcome. These two sub-models are linked together by shared random effects or features of an individual's longitudinal biomarker trajectory. In Section 2.1, we identify key computational bottlenecks in the semiparametric maximum likelihood inference procedure for the joint model. Specifically, we point out that in a standard implementation, the computational complexities for numerical integration, risk set calculation, and standard error estimation are of the order $O(n^2)$, $O(n^2)$, and $O(n^3)$, respectively, where n is the number of subjects. Consequently, current implementation grinds to a halt as n becomes large (for example, $n > 10^4$). We further develop tailored linear scan algorithms to reduce the computational complexity to $O(n)$ in each of the aforementioned components. We illustrate by simulation and real-world data that the linearization algorithms can result in a drastic speed-up by a factor of many thousands when $n > 10^4$, reducing the runtime from days to minutes for big data. Finally, we have developed a user-friendly R package `FastJM` to fit the shared parameter semiparametric joint model using the proposed efficient algorithms and made it publicly available on the Comprehensive R Archive Network (CRAN) at `https://CRAN.R-project.org/package=FastJM`.

The rest of the paper is organized as follows. Section 2.1.1 outlines the semiparametric shared random effect joint model framework and reviews a customized expectation-maximization (EM) algorithm for semiparametric maximum likelihood estimation as well as a standard error estimation method. Section 2.1.2 develops various linear scan algorithms to address the key computational bottlenecks in the EM algorithm and standard error estimation for large data. Section 2.2 presents simulation studies to illustrate the computational efficiency of the proposed linear scan algorithms. Section 2.3 demonstrates the improved computational performance of our `FastJM` R package over some established joint model R packages on two moderate to large real-world data sets. Concluding remarks are provided in Section 2.4.

1.2 A joint model of the individual mean and within-subject variability of a longitudinal outcome with a competing risks time-to-event outcome

In biomedical studies, it is often of interest to study the temporal patterns of longitudinal biomarkers and their association with health outcomes. This paper studies a new joint model of a longitudinal biomarker and a time-to-event outcome that takes into account of possible heterogeneous within-subject (WS) variability of the longitudinal biomarker. Our research is motivated by recent evidence that the WS visit-to-visit variations of some longitudinal biomarkers for diabetes patients are strong risk factors for various health outcomes [Rothwell et al., 2010, Zhou et al., 2018, Ceriello et al., 2019]. For example, using several clinical trials, including UK Prospective Diabetes Study (UKPDS) [Group et al., 1998], Action to Control Cardiovascular Risk in Diabetes (ACCORD) [Group, 2008, Zhou et al., 2018, Ismail-Beigi et al., 2010], and Veterans Affairs Diabetes Trial (VADT) [Duckworth et al., 2009, Reaven et al., 2019], several investigators have demonstrated that individual glycemic variability and blood pressure variability were associated with the risk of cardiovascular diseases (CVD), e.g., myocardial infarction or stroke [Zhou et al., 2018, Nuyujukian et al., 2021], heart failure [Nuyujukian et al., 2020], nephropathy [Zhou et al., 2020, 2021], and retinopathy [Zhou et al., 2021], independent of traditional glycemic and blood pressure control. These studies typically used an ad hoc two-stage approach that first estimates the individual variability of a longitudinal biomarker in the stage 1 analysis and then correlates the individual variability from stage 1 with a time-to-event outcome in the stage 2 analysis. However, the ad hoc two-stage approach has some obvious practical and theoretical drawbacks and limitations.

To illustrate, in a two-stage analysis, participants may be excluded if they have insufficient repeated measures of a biomarker to generate reliable estimates of visit-to-visit variability during the first stage. Second, the first stage analysis is unable to account for the variability measures' uncertainty caused by the uneven number of repeated measures across

individuals. Additionally, the first stage analysis cannot adjust for time-varying effects, such as medication, during variability estimation, leading to significant estimation bias and inaccurate inference during the second stage's risk assessment. Lastly, the second stage analysis cannot perform dynamic prediction of event of interest relying on the underlying trajectory of biomarker(s) as time-dependent covariates. As a result, there is a pressing need to develop a comprehensive statistical framework that can simultaneously model the individual mean and within-subject variability trends of a longitudinal biomarker, allowing for valid statistical inference on their connections with a time-to-event outcome. Furthermore, this framework should enable dynamic prediction of the event outcome based on the observed biomarker history.

During the past three decades, joint modeling of longitudinal and time-to-event data has been extensively studied in the statistical literature and has emerged as a versatile tool to address many challenging issues in longitudinal and survival data analysis. For instance, joint models have been used to handle non-ignorable missing data due to a terminal event in longitudinal data analysis, model intermittently observed time-varying covariates and/or measurement error in survival analysis, and make dynamic prediction of a time-to-event outcome based on the observed trajectories of longitudinal biomarkers [Tsiatis and Davidian, 2004, Wu et al., 2012, Elashoff et al., 2016, Hickey et al., 2018, Papageorgiou et al., 2019, Alsefri et al., 2020, Huang et al., 2011]. However, current statistical literature on joint models of longitudinal and time-to-event data has mostly focused on modeling the mean attribute of a longitudinal biomarker trajectory and its association with a time-to-event outcome and typically assume homogeneous WS variability for the longitudinal biomarker. More recently, there have been efforts to model a longitudinal biomarker while accounting for heterogeneous WS variability [Hedeker et al., 2008, Dzubur et al., 2020, German et al., 2021, Parker et al., 2021]. [Hedeker et al., 2008] introduced a *mixed-effects location-scale model* for the longitudinal data, which allows both WS and between-subject (BS) variability to be modeled by time-dependent covariates. [Dzubur et al., 2020] further extended the *mixed-effects location-*

scale model to *mixed-effects multiple location-scale models* that allow for multiple random effects in the mean component. [German et al., 2021] developed a robust and scalable estimation method to study the effects of both time-varying and time-invariant predictors on WS variability. [Parker et al., 2021] proposed a joint modeling approach to associate both mean trajectory and WS variability of systolic blood pressure with left ventricular mass in early adulthood. However, these longitudinal models have some major limitations. They rely on the stringent missing at random (MAR) assumption, which is often violated by non-ignorable missing data due to a terminal event. Moreover, they are not formally linked to an event outcome and thus cannot be readily used to study the association between a longitudinal biomarker with an event outcome of interest. As mentioned above, joint modeling of longitudinal and time-to-event outcomes can help to address these limitations. However, to the best of our knowledge, joint models of longitudinal and survival outcomes in the presence of heterogeneous WS variability have not yet been reported.

This paper aims to fill the above-mentioned gap in the statistical literature on joint models. The newly developed joint models can capture distinct attributes in a longitudinal biomarker's mean and WS variability and link them to a time-to-event outcome. We illustrate the utility of our method using data from the Multi-Ethnic Study of Atherosclerosis (MESA). The new joint model consists of a linear mixed-effects multiple location scale submodel for the mean and WS variability of a longitudinal biomarker and a semiparametric cause-specific Cox proportional hazards submodel for a competing risks survival outcome. The submodels are linked by shared random effects. We derive an expectation-maximization (EM) algorithm for semiparametric maximum likelihood parameter estimation and adopt a profile-likelihood approach for standard error estimation. We also develop an efficient implementation of our procedure that scales to large-scale biobank data. To this end, we point out that the EM algorithm and standard error estimation for our semiparametric joint model will require $O(n^2)$ and $O(n^3)$ operations, respectively, when implemented naively due to repetitive risk set calculations (see Supplementary Materials B.3). By developing

customized linear scan algorithms for both the EM algorithm and standard error estimation, we reduce the computational complexity to $O(n)$ and thus make the implementation scalable to massive sample size data with tens of thousands to millions of subjects. Furthermore, our empirical results show that in the presence of heterogeneous WS variability, a classical joint model with homogeneous WS variability can lead to biased estimation and invalid inference, which can be remedied by our proposed joint model (see Table 3.1). Finally, we have developed a user-friendly R package **JMH** for the proposed methodology and made it publicly available at `https://github.com/shanpengli/JMH`.

1.3 A joint model of longitudinal and interval-censored post-diagnosis time to event data in the presence of interval-censored covariates due to the unknown diagnosis time: with application to UK-Biobank data

Interval-censored data arise frequently in biomedical studies when the event is known only to occur within a certain time interval. For example, diagnosis of a disease such as diabetes is usually performed at clinical visits and the onset of the disease is only known to fall between two consecutive visits.

The fact that the event time is not directly observed imposes theoretical and computational challenges on semiparametric regression in survival analysis. In the recent decades, there is an extensive literature on modeling and analyzing interval-censored event time data, see [Huang and Wellner, 1997, Sun, 2006, Zhang and Zhao, 2013, Zeng et al., 2016, Wang et al., 2016, Gómez et al., 2009, Sun et al., 2013], among others. For example, Goetghebeur and Ryan [2000] proposed a semiparametric approach to the proportional hazards regression analysis based on the Turnbull algorithm [Turnbull, 1976], which has been used to determine times at which the discrete hazard can take positive mass. Zeng et al. [2016] proposed

a broad class of semiparametric transformation models to account for the effects of possibly time-dependent covariates and showed that the estimated coefficients are consistent and asymptotically efficient.

Instead of interval censored event time, interval censored covariate can also occur in many applications. Langohr and Melis [2014] created a discrete support with corresponding probability masses for the interval-censored covariate. In this paper, a likelihood approach [Gómez et al., 2003], referred to as the Gómez, Espinal, and Lagakos (GEL) approach, was introduced to jointly estimate the regression coefficients for modeling a continuous response variable as well as the marginal distribution of the interval-censored covariate. Morrison et al. [2021] proposed a joint modeling approach to model the distribution of HIV seroconversion dates (interval-censored) without covariates directly and then derive the distribution of infection durations (longitudinal outcome). In this paper, they conducted numerous simulation studies to compare with the existing approaches such as midpoint imputation, uniform imputation, and the GEL approach [Gómez et al., 2003], and demonstrated that their method can produce less bias. Ahn et al. [2018] proposed a available-data, doubly robust, and maximum likelihood estimator to model the primary event of interest, in which an interval-censored secondary event is included as a covariate characterized by Cox proportional hazards model for interval-censored data.

During the past three decades, joint modeling of longitudinal and time-to-event data has been extensively studied in the statistical literature and has emerged as a versatile tool to address many challenging issues in longitudinal and survival data analysis. For instance, joint models have been used to handle non-ignorable missing data due to a terminal event in longitudinal data analysis, model intermittently observed time-varying covariates and/or measurement error in survival analysis, and make dynamic prediction of a time-to-event outcome based on the observed trajectories of longitudinal biomarkers [Tsiatis and Davidian, 2004, Wu et al., 2012, Elashoff et al., 2016, Hickey et al., 2018, Papageorgiou et al., 2019, Alsefri et al., 2020, Huang et al., 2011]. Separate analyses or two-stage modeling of both

outcomes can yield biased results and invalid inference. Despite the aforementioned efforts on modeling longitudinal or survival outcomes together with interval-censored covariates [Gómez et al., 2003, Morrison et al., 2021, Gómez et al., 2003, Ahn et al., 2018], none of these works exist under the joint modeling framework. Methodological development of joint modeling of longitudinal and time-to-event data in the presence of interval-censored covariates due to the unknown diagnosis time has not yet been studied.

The purpose of this paper is to fill the aforementioned gap in the statistical literature on joint models. The newly developed joint model can handle interval-censored covariates that appear in both longitudinal and survival sub-models, and we will demonstrate our estimation procedure can yield efficient and unbiased estimates in our simulation studies. We illustrate the utility of our proposed model using data from UK-Biobank. The new joint model consists of a linear mixed-effects multiple location scale submodel for the mean and WS variability of a longitudinal biomarker and a semiparametric Cox proportional hazards submodel together with interval-censored covariates. The submodels are linked by shared random effects. We derive an expectation-maximization (EM) algorithm for semiparametric maximum likelihood parameter estimation and adopt a profile-likelihood approach for standard error estimation. We also develop an efficient implementation of our procedure that scales to large-scale biobank data.

The rest of the paper is organized as follows. Section 4.1 describes the semiparametric shared random effects joint model framework and derives an efficient, customized EM algorithm for semiparametric maximum likelihood estimation and a standard error estimation method. Section 4.2 evaluates the empirical performance of the proposed joint model in contrast to existing imputation methods via simulation studies. In Section 4.3, we apply our developed method to the UK-Biobank data. Concluding remarks and further discussion are given in Section 4.4.

CHAPTER 2

Efficient algorithms and implementation of a semi-parametric joint model of longitudinal and competing risks time-to-event data with applications to massive biobank data

In this chapter, we identify and address the key computational issues on semi-parametric joint modeling of longitudinal and competing risks time-to-event data. We develop efficient algorithms and implementation of a semi-parametric joint model to reduce the computational complexity from $O(n^2)$ or $O(n^3)$ to $O(n)$. We perform simulations to compare estimation accuracy and precision between different implementations and compare the model performance between our model and other R packages on two real data applications. Below we will discuss the details.

2.1 Efficient algorithms for a semi-parametric joint model of longitudinal and competing risks data

2.1.1 Notations and preliminaries

Let $Y_i(t)$ be the longitudinal outcome at time t for subject i, $i = 1, 2, ..., n$, and n is the total number of subjects. Suppose the longitudinal outcome $Y_i(t)$ is observed at time points t_{ij}, $j = 1, 2, ..., n_i$, and denote $Y_i = (Y_{i1}, ..., Y_{in_i})$. Let $C_i = (T_i, D_i)$ be the competing risks data

on subject i, where T_i is the observed time to either failure or censoring, and D_i takes value in $\{0, 1, 2, \ldots, K\}$, with $D_i = 0$ indicating a censored event and $D_i = k$ implying the kth type of failure is observed on subject i, $k = 1, 2, ..., K$. The censoring mechanism is assumed to be independent of the failure time.

2.1.1.1 Model

Consider the following joint model in which the longitudinal outcome $Y_i(t)$ is characterized by a linear mixed effects model:

$$\begin{aligned} Y_i(t) &= m_i(t) + \epsilon_i(t), \\ &= X_i^{(1)}(t)^T \beta + \tilde{X}_i^{(1)}(t)^T b_i + \epsilon_i(t), \end{aligned} \tag{2.1}$$

and the competing risks survival outcome is modeled by a proportional cause-specific hazards model:

$$\begin{aligned} \lambda_k(t \mid X_i^{(2)}, b_i, \gamma_k, \nu_k) = & \lim_{h \to 0} \frac{P(t \leq T_i < t + h, D_i = k \mid T_i \geq t, X_i^{(2)}, b_i)}{h} \\ = & \quad \lambda_{0k}(t) \exp\{X_i^{(2)T} \gamma_k + \nu_k^T b_i\}, \quad k = 1, \ldots, K. \end{aligned} \tag{2.2}$$

In the longitudinal sub-model (2.1), $m_i(t)$ is the mean of the longitudinal outcome at time t, $X_i^{(1)}(t)$ and $\tilde{X}_i^{(1)}(t)$ are column vectors of possibly time-varying covariates associated with the longitudinal outcome $Y_i(t)$, β represents a $p \times 1$ vector of fixed effects of $X_i^{(1)}(t)$, $b_i \sim N_q(0, \Sigma)$ denotes a $q \times 1$ vector of random effects for $\tilde{X}_i^{(1)}(t)$, $\epsilon_i(t) \sim N(0, \sigma^2)$ is the measurement error independent of b_i, and $\epsilon_i(t_1)$ is independent of $\epsilon_i(t_2)$ for any $t_1 \neq t_2$. In the competing risks survival sub-model (2.2), $\lambda_k(t \mid X_i^{(2)}, b_i, \gamma_k, \nu_k)$ is the conditional cause-specific hazard rate for type k failure at time t, given time-invariant covariates $X_i^{(2)}$ and the shared random effects b_i, and $\lambda_{0k}(t)$ is a completely unspecified baseline cause-specific hazard function for type k failure. The two sub-models are linked together by the shared random effects b_i and the strength of association is quantified by the association parameters $\nu_1, ..., \nu_K$.

2.1.1.2 Semiparametric maximum likelihood estimation via an EM algorithm

Let $\Psi = \{\beta, \sigma^2, \gamma, \nu, \Sigma, \lambda_{01}(\cdot), ..., \lambda_{0K}(\cdot)\}$ denote all unknown parameters for the joint model (2.1) and (2.2), where $\gamma = (\gamma_1^T, ..., \gamma_K^T)^T$ and $\nu = (\nu_1^T, ..., \nu_K^T)^T$. Let $I(D_i = k)$ be the competing event indicator for type k failure, which takes the value 1 if the condition $D_i = k$ is satisfied, and 0 otherwise. The observed-data likelihood for Ψ is then given by

$$
\begin{aligned}
L(\Psi; Y, C) &\propto \prod_{i=1}^{n} f(Y_i, C_i \mid \Psi) \\
&= \prod_{i=1}^{n} \int_{b_i} f(Y_i \mid C_i, b_i, \Psi) f(C_i \mid b_i, \Psi) f(b_i \mid \Psi) db_i \\
&\propto \prod_{i=1}^{n} \int_{b_i} \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2} \left\{ Y_{ij} - X_i^{(1)}(t_{ij})^T \beta - \tilde{X}_i^{(1)}(t_{ij})^T b_i \right\}^2 \right] \\
&\quad \times \left\{ \prod_{k=1}^{K} \lambda_k(t \mid X_i^{(2)}, b_i, \gamma_k, \nu_k)^{I(D_i=k)} \right\} \exp\left[-\int_0^{T_i} \left\{ \sum_{k=1}^{K} \lambda_k(t \mid X_i^{(2)}, b_i, \gamma_k, \nu_k) \right\} dt \right] \\
&\quad \times \frac{1}{\sqrt{(2\pi)^q |\Sigma|}} \exp(-\frac{1}{2} b_i^T \Sigma^{-1} b_i) db_i,
\end{aligned}
$$

which follows from the assumption that Y_i and C_i are independent conditional on the covariates and the random effects.

Because Ψ involves K unknown hazard functions, finding its maximum likelihood estimate by maximizing the above observed-data likelihood is nontrivial. However, a customized EM algorithm can be derived to compute the maximum likelihood estimate of Ψ by regarding the latent random effects b_i as missing data [Elashoff et al., 2008]. The complete-data likelihood based on (Y, C, b) is

$$
\begin{aligned}
L(\Psi; Y, C, b) &\propto \prod_{i=1}^{n} \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2} \left\{ Y_{ij} - X_i^{(1)}(t_{ij})^T \beta - \tilde{X}_i^{(1)}(t_{ij})^T b_i \right\}^2 \right] \\
&\quad \times \prod_{k=1}^{K} \left\{ \Delta\Lambda_{0k}(T_i) \exp(X_i^{(2)T} \gamma_k + \nu_k^T b_i) \right\}^{I(D_i=k)} \\
&\quad \times \exp\left\{ -\sum_{k=1}^{K} \Lambda_{0k}(T_i) \exp(X_i^{(2)T} \gamma_k + \nu_k^T b_i) \right\} \times \frac{1}{\sqrt{(2\pi)^q |\Sigma|}} \exp(-\frac{1}{2} b_i^T \Sigma^{-1} b_i),
\end{aligned}
$$

where $\Lambda_{0k}(.)$ is the cumulative baseline hazards function for type k failure and $\Delta\Lambda_{0k}(T_i) = \Lambda_{0k}(T_i) - \Lambda_{0k}(T_i-)$.

The EM algorithm iterates between an expectation step (E-step):

$$Q(\Psi; \Psi^{(m)}) \equiv E^{(m)}_{b|Y,C,\Psi^{(m)}} \left\{ \log L(\Psi; Y, C, b) \right\}, \tag{2.3}$$

and a maximization step (M-step):

$$\Psi^{(m+1)} = \arg\max_{\Psi} Q(\Psi; \Psi^{(m)}), \tag{2.4}$$

until the algorithm converges, where $\Psi^{(m)}$ is the estimate of Ψ from the m-th iteration. The E-step (2.3) involves calculating the following expected values across all subjects: $E^{(m)}(b_i)$, $E^{(m)}(b_i b_i^T)$, $E^{(m)}\{\exp(\nu_k b_i)\}$, $E^{(m)}\{b_i \exp(\nu_k b_i)\}$, and $E^{(m)}\{b_i b_i^T \exp(\nu_k b_i)\}$, where

$$E^{(m)}\{h(b_i)\} = \int h(b_i) f(b_i|Y_i, C_i, \Psi^{(m)})\, db_i \tag{2.5}$$

for any function $h(\cdot)$. Furthermore, it can be shown that the M-step (2.4) has closed-form solutions for the parameters β, σ^2, Σ, and $\Lambda_{0k}(t)$ and that the other parameters γ and ν can be updated using the one-step Newton-Raphson method. Details are provided in equations (A.1)-(A.8) of the supplementary materials.

2.1.1.3 Standard error estimation

As discussed in Elashoff et al. [2016] (Section 4.1, p.72), standard errors of the parametric components of the semiparametric maximum likelihood estimate can be estimated by profiled likelihood, observed information matrix, or bootstrap. All three methods can be computationally intensive when n is large. Here we focus on the profiled likelihood-based method and show that its computation can be linearized with respect to n.

Let $\Omega = (\beta, \Sigma, \sigma^2, \gamma_1, ..., \gamma_K, \nu_1, ..., \nu_K)$ denote the parametric component of Ψ and $\widehat{\Omega}$ its maximum likelihood estimate. The variance-covariance matrix of $\widehat{\Omega}$ can be estimated by inverting the following approximate empirical Fisher information [Lin et al., 2004, Zeng

et al., 2005, Zeng and Cai, 2005]:

$$\sum_{i=1}^{n} \nabla_{\Omega} l^{(i)}(\hat{\Omega}; Y, C) \nabla_{\Omega} l^{(i)}(\hat{\Omega}; Y, C)^T, \tag{2.6}$$

where $\nabla_{\Omega} l^{(i)}(\Omega; Y, C)$ is the observed score vector from the profiled likelihood $l^{(i)}(\Omega; Y, C)$ of Ω on the ith subject by profiling out the baseline hazards. Details of the observed score vector for each parametric component are provided in equations (A.9)-(A.13) of the supplementary materials.

2.1.2 Efficient algorithms and implementation of the EM algorithm and standard error estimation

With naive implementation, multiple quantities in the above E-step, M-step, and standard error estimation will involve $O(n^2)$ or $O(n^3)$ operations, which become computationally prohibitive at large sample size n. Below we identify these bottlenecks and discuss appropriate linear scan algorithms to reduce their computational complexities to $O(n)$.

2.1.2.1 Efficient implementation of the E-step

At each EM iteration, the E-step (2.3) requires calculating integrals (2.5) across all subjects. Below we discuss how to accelerate two commonly used numerical integration methods for evaluating these integrals.

Standard Gauss-Hermite quadrature rule for numerical integration A commonly used method for numerical evaluation of integral (2.5) is based on the standard Gauss-Hermite quadrature rule [Press et al., 2007]:

$$\begin{aligned} E^{(m)}\{h(b_i)\} &= \frac{\int h(b_i) f(Y_i, C_i, b_i \mid \Psi^{(m)}) db_i}{f(Y_i, C_i \mid \Psi^{(m)})} \\ &= \frac{\int h(b_i) f(Y_i, C_i \mid b_i, \Psi^{(m)}) f(b_i \mid \Psi^{(m)}) db_i}{\int f(Y_i, C_i \mid b_i, \Psi^{(m)}) f(b_i \mid \Psi^{(m)}) db_i} \end{aligned}$$

$$\approx \frac{\sum_{t_1,t_2,...,t_q} \pi_t h(\tilde{b}_t^{(m)}) f(Y_i, C_i \mid \tilde{b}_t^{(m)}, \Psi^{(m)}) f(\tilde{b}_t^{(m)} \mid \Psi^{(m)}) \exp(||c_t||^2)}{\sum_{t_1,t_2,...,t_q} \pi_t f(Y_i, C_i \mid \tilde{b}_t^{(m)}, \Psi^{(m)}) f(\tilde{b}_t^{(m)} \mid \Psi^{(m)}) \exp(||c_t||^2)}, \quad (2.7)$$

where

$$\begin{aligned} f(Y_i, C_i \mid \tilde{b}_t^{(m)}, \Psi^{(m)}) &= f(Y_i \mid \tilde{b}_t^{(m)}, \Psi_y^{(m)}) \times f(C_i \mid \tilde{b}_t^{(m)}, \Psi_c^{(m)}) \\ &= \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\sigma^{2(m)}}} \exp\left[-\frac{1}{2\sigma^{2(m)}} \left\{ Y_{ij} - X_i^{(1)}(t_{ij})^T \beta^{(m)} - \tilde{X_i}^{(1)}(t_{ij})^T \tilde{b}_t^{(m)} \right\}^2 \right] \\ &\quad \times \prod_{k=1}^{K} \left\{ \Delta\Lambda_{0k}^{(m)}(T_i) \exp(X_i^{(2)T} \gamma_k^{(m)} + \nu_k^{T(m)} \tilde{b}_t^{(m)}) \right\}^{I(D_i=k)} \\ &\quad \times \exp\left\{ -\sum_{k=1}^{K} \Lambda_{0k}^{(m)}(T_i) \exp(X_i^{(2)T} \gamma_k^{(m)} + \nu_k^{T(m)} \tilde{b}_t^{(m)}) \right\}, \end{aligned} \quad (2.8)$$

with $\Lambda_{0k}^{(m)}(.)$ the right-continuous and non-decreasing cumulative baseline hazard function for type k failure at the m-th EM iteration as defined in Section A.1 (equation (A.4)) of the supplementary material and $\Delta\Lambda_{0k}^{(m)}(T_i)$ the jump size of $\Lambda_{0k}^{(m)}(.)$ at T_i, q the dimension of the random effects vector, $\sum_{t_1,t_2,...,t_q}$ the shorthand for $\sum_{t_1=1}^{n_q} ... \sum_{t_q=1}^{n_q}$, n_q the number of quadrature points, $c_t = (c_{t_1}, c_{t_2}, ..., c_{t_q})^T$ the abscissas with corresponding weights π_t, $\tilde{b}_t^{(m)} = \sqrt{2}\hat{\Sigma}^{(m)1/2} c_t$ the re-scaled alternative abscissas, and $\hat{\Sigma}^{(m)1/2}$ the square root of $\hat{\Sigma}^{(m)}$ [Elashoff et al., 2008]. However, this method is computationally intensive due to multiple factors. First, it usually requires many quadrature points to approximate an integral with sufficient accuracy because the mode of the integrand is often located in a region different from zero. Second, the computational cost increases exponentially with q because the Cartesian product of the abscissas is used to evaluate the integrand with respect to each random effect. Lastly, the alternative abscissas $\tilde{b_t}^{(m)}$ need to be re-calculated at every EM iteration.

Pseudo-adaptive Gauss-Hermite quadrature rule for numerical integration When the number of longitudinal measurements per subject is relatively large, [Rizopoulos, 2012a] introduced a pseudo-adaptive Gauss-Hermite quadrature rule for numerical approximation of integral (2.5), which achieves good approximation accuracy with only a small number (n_q) of quadrature points and is thus computationally more efficient. The pseudo-adaptive

Gauss-Hermite quadrature rule proceeds as follows. First, fit the linear mixed effects model (2.1) to extract the empirical Bayes estimates of the random effects and its covariance matrix $\tilde{H}_i^{-1}$:

$$\begin{aligned}
\tilde{b}_i &= \hat{\Sigma}\tilde{\boldsymbol{X}}_i^{(1)}\hat{V}_i^{-1}\left\{Y_i - \boldsymbol{X}_i^{(1)T}\hat{\beta}\right\}, \qquad (2.9)\\
\tilde{H}_i^{-1} &= \hat{\Sigma} - \hat{\Sigma}\tilde{\boldsymbol{X}}_i^{(1)}\left[\hat{V}_i^{-1} - \hat{V}_i^{-1}\boldsymbol{X}_i^{(1)T}\left\{\sum_{i=1}^{n}\boldsymbol{X}_i^{(1)}\hat{V}_i^{-1}\boldsymbol{X}_i^{(1)T}\right\}^{-1}\boldsymbol{X}_i^{(1)}\hat{V}_i^{-1}\right]\\
&\quad \times\tilde{\boldsymbol{X}}_i^{(1)T}\hat{\Sigma}, \qquad (2.10)
\end{aligned}$$

where $\boldsymbol{X}_i^{(1)} = (X_i^{(1)}(t_{i1}), ..., X_i^{(1)}(t_{in_i}))$, $\tilde{\boldsymbol{X}}_i^{(1)} = (\tilde{X}_i^{(1)}(t_{i1}), ..., \tilde{X}_i^{(1)}(t_{in_i}))$, and $V_i = \tilde{\boldsymbol{X}}_i^{(1)T}\Sigma\tilde{\boldsymbol{X}}_i^{(1)}$ $+\sigma^2 I$. Then, define the alternative abscissas $\tilde{r}_t = \tilde{b}_i + \sqrt{2}\tilde{H}_i^{-1/2}c_t$ and approximate $E^{(m)}\{h(b_i)\}$ by

$$E^{(m)}\{h(b_i)\} \approx \frac{\sum_{t_1,t_2,...,t_q}\pi_t h(\tilde{r}_t)f(Y_i, C_i \mid \tilde{r}_t, \Psi^{(m)})f(\tilde{r}_t \mid \Psi^{(m)})\exp\{||c_t||^2\}}{\sum_{t_1,t_2,...,t_q}\pi_t f(Y_i, C_i \mid \tilde{r}_t, \Psi^{(m)})f(\tilde{r}_t \mid \Psi^{(m)})\exp\{||c_t||^2\}}, \qquad (2.11)$$

where

$$\begin{aligned}
f(Y_i, C_i \mid \tilde{r}_t, \Psi^{(m)}) &= f(Y_i \mid \tilde{r}_t, \Psi_y^{(m)}) \times f(C_i \mid \tilde{r}_t, \Psi_c^{(m)})\\
&= \prod_{j=1}^{n_i}\frac{1}{\sqrt{2\pi\sigma^{2(m)}}}\exp\left[-\frac{1}{2\sigma^{2(m)}}\left\{Y_{ij} - X_i^{(1)}(t_{ij})^T\beta^{(m)} - \tilde{X}_i^{(1)}(t_{ij})^T\tilde{r}_t\right\}^2\right]\\
&\quad \times\prod_{k=1}^{K}\left\{\Delta\Lambda_{0k}^{(m)}(T_i)\exp(X_i^{(2)T}\gamma_k^{(m)} + \nu_k^{(m)T}\tilde{r}_t)\right\}^{I(D_i=k)}\\
&\quad \times\exp\left\{-\sum_{k=1}^{K}\Lambda_{0k}^{(m)}(T_i)\exp\{X_i^{(2)T}\gamma_k^{(m)} + \nu_k^{(m)T}\tilde{r}_t\}\right\}, \qquad (2.12)
\end{aligned}$$

with the notations defined similarly to equation (2.8).

The pseudo-adaptive Gauss-Hermite quadrature rule is computationally appealing because the alternate re-scaled quadrature points $\tilde{r}_t$ are computed only once before the EM algorithm and do not need to be updated in the EM algorithm. Additionally, the pseudo-adaptive Gauss-Hermite quadrature rule requires fewer quadrature points than the standard Gauss-Hermite quadrature rule to achieve the same numerical approximation accuracy [Rizopoulos, 2012a]. For example, our simulation results in the supplementary materials (Section A.3, Table A.1) illustrate that the pseudo-adaptive Gauss-Hermite quadrature rule with

$n_q = 6$ quadrature points produces almost identical results to the standard Gauss-Hermite quadrature rule with $n_q = 20$ quadrature points.

Remark 1. *(Linear calculation of $\tilde{H}_i^{-1}$'s across all subjects) At the first sight, calculating $\tilde{H}_i^{-1}$ (2.10) across all subjects would involve $O(n^2)$ operations since each $\tilde{H}_i^{-1}$ involves a summation over n subjects. However, because the same quantity $\sum_{i=1}^{n} \boldsymbol{X}_i^{(1)} \hat{V}_i^{-1} \boldsymbol{X}_i^{(1)T}$ appears in every $\tilde{H}_i^{-1}$, one can pre-compute this quantity and then use the cached value to calculate $\tilde{H}_i^{-1}$ across all subjects. This way, one can compute the $\tilde{H}_i^{-1}$'s across all subjects in $O(n)$ operations. Our simulation study in the supplementary material (Section A.4, Figure A.1) shows that applying this simple linearization algorithm can yield a speed-up by a factor of 10 to 10,000 when n grows from 10 to 10^5. Our implementation is also significantly faster than a popular R package* ***lme4*** *[Bates et al., 2007] with over a speed-up by a factor of 10 to 500 as n grows from 10 to 10^5.*

Linear scan algorithm for calculating $f(Y_i, C_i \mid \tilde{r}_t, \Psi^{(m)})$ across all subjects Both the standard Gauss-Hermite quadrature rule (2.7) and the pseudo-adaptive Gauss-Hermite quadrature rule (2.11) require evaluating $f(Y_i, C_i \mid b_i, \Psi^{(m)})$ at their pre-specified abscissas across all subjects (See equations (2.8) and (2.12)). Hence, calculating $f(Y_i, C_i \mid \tilde{r}_t, \Psi^{(m)})$ requires evaluation of $\Lambda_{0k}^{(m)}(T_i)$ across all subjects for each k. We observe from equation (A.4) that for each k, $\{\Lambda_{0k}^{(m)}(t_{k1}), \ldots, \Lambda_{0k}^{(m)}(t_{kq_k})\}$ have already been calculated from the m-th EM iteration, where $t_{k1} > \cdots > t_{kq_k}$ are q_k distinct observed type k event times. For each i, calculating $\Lambda_{0k}^{(m)}(T_i)$ would involve $O(n)$ operations if a global search is performed to find the interval of two adjacent type k event times containing T_i. Consequently, calculating $\{\Lambda_{0k}^{(m)}(T_1), \ldots, \Lambda_{0k}^{(m)}(T_n)\}$ would require $O(n^2)$ operations. However, by taking advantage of the fact that $\Lambda_{0k}^{(m)}(t)$ is a right-continuous and non-decreasing step function, one can obtain $\{\Lambda_{0k}^{(m)}(T_1), \ldots, \Lambda_{0k}^{(m)}(T_n)\}$ from $\{\Lambda_{0k}^{(m)}(t_{k1}), \ldots, \Lambda_{0k}^{(m)}(t_{kq_k})\}$ in $O(n)$ operations using the following linear scan algorithm. First, sort the observation times T_i, $i = 1, \ldots, n$, in

descending order. Denote by (i) the ranked index of a subject. Then, define a mapping

$$\{\Lambda_{0k}^{(m)}(t_{k1}), \Lambda_{0k}^{(m)}(t_{k2}), \ldots \Lambda_{0k}^{(m)}(t_{kq_k})\} \mapsto \{\Lambda_{0k}^{(m)}(T_{(1)}), \Lambda_{0k}^{(m)}(T_{(2)}), \ldots, \Lambda_{0k}^{(m)}(T_{(n)})\}, \quad (2.13)$$

where $t_{k1}, \ldots, t_{kq_k}$ are scanned forward from the largest to the smallest, and for each t_{kj}, only a subset of the ranked observation times $T_{(i)}$ are scanned forward to calculate $\Lambda_{0k}^{(m)}(T_{(i)})$ as follows

$$\Lambda_{0k}^{(m)}(T_{(i)}) = \begin{cases} \Lambda_{0k}^{(m)}(t_{k1}), & \text{if } T_{(i)} \geq t_{k1}, \\ \Lambda_{0k}^{(m)}(t_{k(j+1)}), & \text{if } T_{(i)} \in [t_{k(j+1)}, t_{kj}), \text{ for some } j \in \{1, \ldots, q_k - 1\}, \\ 0, & T_{(i)} < t_{kq_k}. \end{cases}$$

Specifically, start with t_{k1} and scan through the first set of observation times $T_{(1)} \geq ... \geq T_{(i_{k1})}$ where $T_{(i_{k1})} \geq t_{k1} > T_{(i_{k1}+1)}$, and the corresponding $\Lambda_{0k}(T_{(i)})$'s take the value $\Lambda_{0k}(t_{k1})$. Next, move forward to t_{k2} and scan through the second set $T_{(i_{k1}+1)} \geq ... \geq T_{(i_{k2})}$, where $T_{(i_{k2})} \geq t_{k2} > T_{(i_{k2})+1}$, and the corresponding $\Lambda_{0k}(T_{(i)})$'s take the value evaluated at $\Lambda_{0k}(t_{k2})$. Repeat the same process until $T_{(n)}$ is scanned. Because the scanned $T_{(i)}$'s for different t_{kj}'s do not overlap, the entire algorithm costs only $O(n)$ operations.

2.1.2.2 Linear risk set scan for the M-step

In the M-step, multiple quantities in equations, (A.4)-(A.8) such as the cumulative baseline hazard functions, and the Hessian matrix and score vector for γ_k and ν_k $(k = 1, 2, \ldots, K)$, involve aggregating information over the risk set $R(t_{kj}) = \{r : T_r \geq t_{kj}\}$ at each uncensored event time t_{kj}. These quantities are further aggregated across all t_{kj}'s. If all subjects are scanned to determine the risk set $R(t_{kj})$ at each t_{kj}, then aggregating information over the risk set for all uncensored event times would obviously require $O(n^2)$ operations. Below we explain how to linearize the number of operations for risk set calculations across all uncensored event times by taking advantage of the fact that the risk set is decreasing over time for right censored data.

First, to calculate $\Lambda_{0k}^{(m+1)}(t_{kj})$, $j = 1, \ldots, q_k$, one needs to compute $\sum_{r \in R(t_{kj})} \exp(X_r^{(2)T} \gamma_k^{(m)})$ $E\left\{\exp(\nu_k^{(m)} b_r)\right\}$, $j = 1, \ldots, q_k$. Because the distinct uncensored event times $t_{k1} > \cdots > t_{kq_k}$ are arranged in decreasing order, the risk set $R(t_{k(j+1)})$ can be decomposed into two disjoint sets: $R(t_{k(j+1)}) = R(t_{kj}) \cup \{r : T_{(r)} \in [t_{k(j+1)}, t_{kj})\}$, and consequently,

$$\sum_{r \in R(t_{k(j+1)})} a_r = \sum_{r \in R(t_{kj})} a_r + \sum_{\{r : T_{(r)} \in [t_{k(j+1)}, t_{kj})\}} a_r, \tag{2.14}$$

for any sequence of real numbers $a_1, \ldots, a_n$. It follows from the recursive formula (2.14) and the fact that the subjects in $R(t_{kj})$ do not need to be scanned to calculate the second term of (2.14), one can calculate $\sum_{r \in R(t_{kj})} a_r$, $j = 1, \ldots q_k$, in $O(n)$ operations when $T_{(r)}$'s are scanned backward in time.

Next, to calculate the Hessian matrix $I_{\gamma_k}^{(m)}$ for γ_k in (A.5), we first rewrite it as

$$I_{\gamma_k}^{(m)} = \sum_{j=1}^{q_k} \left[\Delta\Lambda_{0k}(t_{kj})^{(m+1)} \sum_{r \in R(t_{kj})} \exp(X_r^{(2)T} \gamma_k^{(m)}) E\left\{\exp(\nu_k^{(m)T} b_r)\right\} X_r^{(2)} X_r^{(2)T} \right],$$

which allows one to linearize its calculation based on (2.14) with $a_r = \Delta\Lambda_{0k}(t_{kj})^{(m+1)} \exp(X_r^{(2)T} \gamma_k^{(m)})$ $E\left\{\exp(\nu_k^{(m)T} b_r)\right\} X_r^{(2)} X_r^{(2)T}$ similar to the linear scan algorithm for $\Lambda_{0k}^{(m+1)}(t_{kj})$'s.

Finally, the above linear risk set scan algorithm can be adapted to calculate the Hessian matrix and score vector for γ_k and ν_k in equations (A.6)-(A.8) in $O(n)$ operations in a similar fashion.

2.1.2.3 Linear scan algorithm for standard error estimation

The standard error estimation formula in (2.6) relies on the observed score vector from the profiled likelihood where the baseline hazard is profiled out. However, for each subject i, two components of the score vector, $\nabla_{\gamma_k} l^{(i)}(\hat{\Omega}; Y, C)$ and $\nabla_{\nu_k} l^{(i)}(\hat{\Omega}; Y, C)$ as given in equations (A.12) and (A.13), involve aggregating information either over $\{r \in R(T_i)\}$ or over both $\{r \in R(t_{kj})\}$ and $\{j : t_{kj} \leq T_i\}$. If implemented naively, the aggregation can take either $O(n)$ or $O(n^2)$ operations, respectively. As a result, the observed information matrix can

take $O(n^3)$ operation as it requires summing up the information across all subjects. Below we describe a sequential linear scan algorithm to reduce the computational complexity from $O(n^3)$ to $O(n)$.

Our algorithm can be easily explained by considering the calculation of the following expression in the second term of $\nabla_{\gamma_k} l^{(i)}(\hat{\Omega}; Y, C)$ in (A.12):

$$\sum_{j:t_{kj}\le T_i} \frac{d_{kj}\sum_{r\in R(t_{kj})}\exp(\gamma_k^T X_r^{(2)})E\left\{\exp(\nu_k^T b_r)\right\}X_r^{(2)}}{\left[\sum_{r\in R(t_{kj})}\exp(\gamma_k^T X_r^{(2)})E\left\{\exp(\nu_k^T b_r)\right\}\right]^2}, \quad \text{for } i=1,...,n.$$

In other words, we need to compute $B(T_i)$ for $i=1,\ldots,n$, where $B(t)\equiv\sum_{j:t_{kj}\le t} b_{kj}$ and

$$b_{kj}=\frac{d_{kj}\sum_{r\in R(t_{kj})}\exp(\gamma_k^T X_r^{(2)})E\left\{\exp(\nu_k^T b_r)\right\}X_r^{(2)}}{\left[\sum_{r\in R(t_{kj})}\exp(\gamma_k^T X_r^{(2)})E\left\{\exp(\nu_k^T b_r)\right\}\right]^2}.$$

Before going further, we recall that the distinct uncensored event times $t_{k1}>\cdots>t_{kq_k}$ are in descending order and that the subjects are sorted so that the observation times T_i's are in descending order.

First of all, because the risk set is decreasing over time for right censored data, it follows from equation (2.14) that $B(t_{k1}),\ldots,B(t_{kq_k})$ can be computed in $O(n)$ operations as one scans through $t_{k1},\ldots,t_{kq_k}$ backward in time. Second, analogous to (2.13), the following linear scan algorithm can be used to calculate $\{B(T_{(1)}),B(T_{(2)}),\cdots,B(T_{(n)})\}$ from $\{B(t_{k1}),\ldots,B(t_{kq_k})\}$:

$$\{B(t_{k1}),\ldots,B(t_{kq_k})\}\mapsto\{B(T_{(1)}),B(T_{(2)}),\cdots,B(T_{(n)})\},$$

where $t_{k1},\ldots,t_{kq_k}$ are scanned forward from the largest to the smallest, and for each t_{kj}, only a subset of the ranked observation times $T_{(i)}$'s are scanned forward to calculate $B(T_{(i)})$'s as follows

$$B(T_{(i)})=\begin{cases} B(t_{k1}), & \text{if } T_{(i)}\ge t_{k1},\\ B(t_{k(j+1)}), & \text{if } T_{(i)}\in[t_{k(j+1)},t_{kj}),\ \text{for some } j\in\{1,\ldots,q_k-1\},\\ 0, & \text{otherwise.}\end{cases}$$

The details are essentially the same as those discussed following equation (2.13), and thus omitted here.

2.2 Simulation studies

We present a simulation study to illustrate the computational speed-up rendered by the proposed linear algorithms as the sample size n grows from 100 to 1,000,000. All simulations were run on a MacBook Pro with 6-Core Intel Core i7 processor (2.6GHz) and 16 GB RAM running MacOS.

We generated longitudinal measurements Y_{ij} from

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 X_{2i} + b_{0i} + b_{1i} t_{ij} + \epsilon_{ij} \tag{2.15}$$

which corresponds to model (2.1) with $X_i^{(1)}(t_{ij})^T = (1, t_{ij}, X_{2i})$ and $\tilde{X}_i^{(1)}(t_{ij})^T = (1, t_{ij})$, and competing risks event times from a proportional cause-specific hazards model

$$\lambda_1(t; X_{1i}, X_{2i}, b_i, \gamma_1, \nu_1) = \lambda_{01}(t) \exp\{\gamma_{11} X_{1i} + \gamma_{12} X_{2i} + \nu_1^T b_i\} \tag{2.16}$$

$$\lambda_2(t; X_{1i}, X_{2i}, b_i, \gamma_2, \nu_2) = \lambda_{02}(t) \exp\{\gamma_{21} X_{1i} + \gamma_{22} X_{2i} + \nu_2^T b_i\}, \tag{2.17}$$

where the two sub-models (2.15), (2.16), and (2.17) are linked together through the shared random effects $b_i = (b_{0i}, b_{1i})^T$. In the above joint model, $t_{ij} = 0, 1, ...$ represent scheduled visit times, X_{1i} follows $N(2, 1.0)$, $X_{2i} \sim$ Bernoulli(0.5) is a binary covariate, the random effects $b_i = (b_{0i}, b_{1i})^T$ follows a $N_2(0, \Sigma)$ distribution with $\Sigma_{11} = 0.5, \Sigma_{22} = 0.25,$ and $\Sigma_{12} = 0$, the measurement errors ϵ_{ij} are iid $N(0, 0.5)$ and independent of b_i, and the baseline hazards $\lambda_{01}(t)$ and $\lambda_{01}(t)$ are constants 0.05 and 0.1, respectively. We simulated noninformative censoring time V_i following exp(20) and let $T_i = \min\{T_{i1}^*, T_{i2}^*, V_i\}$ be the observed time (possibly censored) for subject i. The longitudinal measurements for subject i at t_{ij} are assumed missing after T_i.

We first compared the runtime between three different implementations of the EM algorithm for fitting the joint model (2.1) and (2.2) as described in Section 2.1.2:

1. Method 1: Standard implementation of the EM algorithm using the standard Gauss-Hermite quadrature rule in the E-step (equation (2.7) with $n_q = 20$) without any linear computation;

2. Method 2: Standard implementation of the EM algorithm using the pseudo-adaptive quadrature rule in the E-step (equation (2.11) with $n_q = 6$) with the linear calculation of $\tilde{H}_i^{-1}$'s described in Remark 1 and without any other linear computation;

3. Method 3: Method 2 + linear scan for calculating $f(Y_i, C_i \mid \tilde{r}_t, \Psi^{(m)})$'s + linear risk set scan for M-step as described in Section 2.1.2;

where the number of quadrature points n_q for methods 1 and 2 were determined by first trying different values, {10, 20, 30} for method 1 and {6, 9, 12, 15} for method 2, and then choosing the smallest value for which the estimation results are stabilized and similar between the two implementation methods. For comparison purposes, we have also included the runtime of an established joint model R package `joineR`, which uses a similar EM algorithm for parameter estimation to fit a semiparametric joint model with a slightly different latent association structure in the competing risks sub-model [Philipson et al., 2018]. The results are depicted in Figure 2.1.

It is seen from Figure 2.1a that the runtime of method 3 increases linearly with the sample size, while the runtime of the other three methods grows exponentially. For moderate sample size, method 2 is computationally more efficient than method 1 because it requires fewer quadrature points for numerical integration. However, its computational advantage diminishes as the sample size increases due to the exponentially increasing computational cost of $f(Y_i, C_i \mid \tilde{r}_t, \Psi^{(m)})$'s and risk set calculation in the M-step. By further linearizing the computation of these key components, method 3 has yielded more than 100-fold speed-up over method 2 when $n = 10^5$, and the speed-up is expected to increase exponentially as n increases (Figure 2.1b). Furthermore, method 3 has demonstrated more than 30-fold speed-up over `joineR` when $n = 10^4$. We also note that `joineR` failed to run when $n = 10^5$ due to

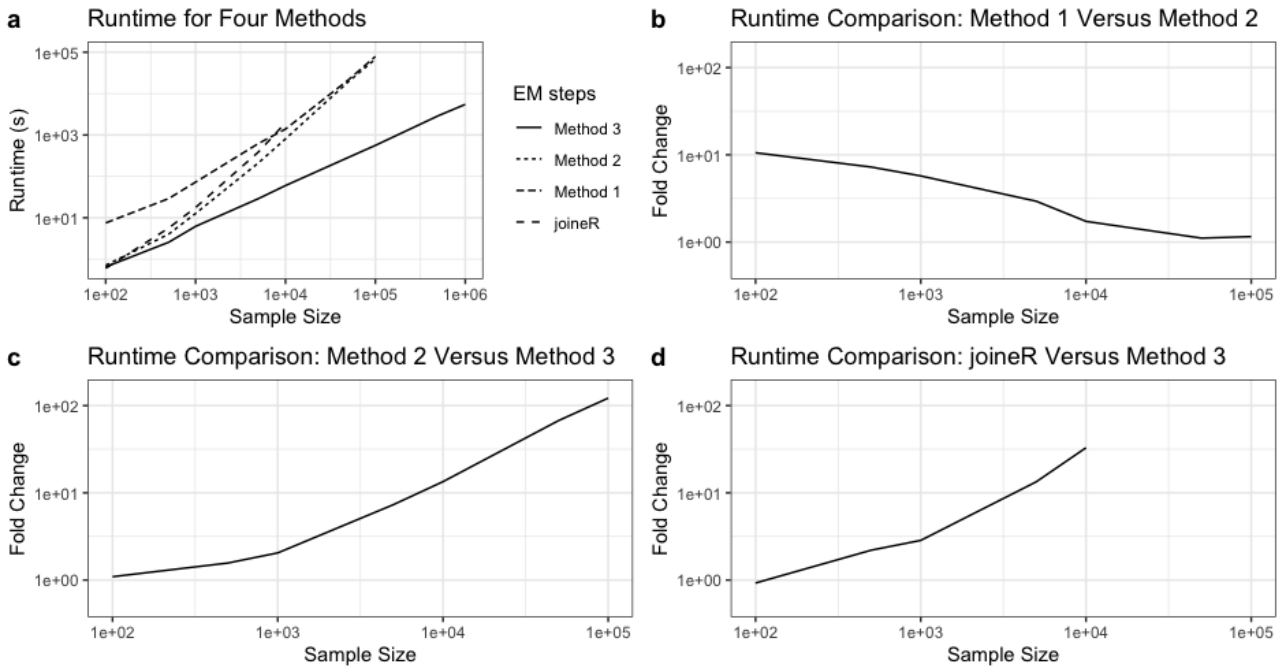

Figure 2.1: Runtime (seconds) comparison between three different implementations of the EM algorithm for fitting the joint model (2.1) and (2.2) and the `joineR` package. The details of Methods 1-3 are given in Section 2.2. `joineR` is an established R package which fits a similar semiparametric joint model with a slightly different latent association structure in the competing risks sub-model [Philipson et al., 2018]. Fold change is calculated as the ratio of runtime between two methods.

the overload of memory.

We also compared the runtime of two implementations of the standard error estimation: with and without linear scan as described in Section 2.1.2.3, and the bootstrap method employed by the `joineR` package [Philipson et al., 2018]. The results are shown in Figure 2.2.

It is seen from Figure 2.2a that the implementation with linear scan easily scales to a million subjects, taking only minutes to finish, while the naive implementation without linear scan grinds to a halt when the sample size is 10,000 or larger. Figure 2.2b shows that

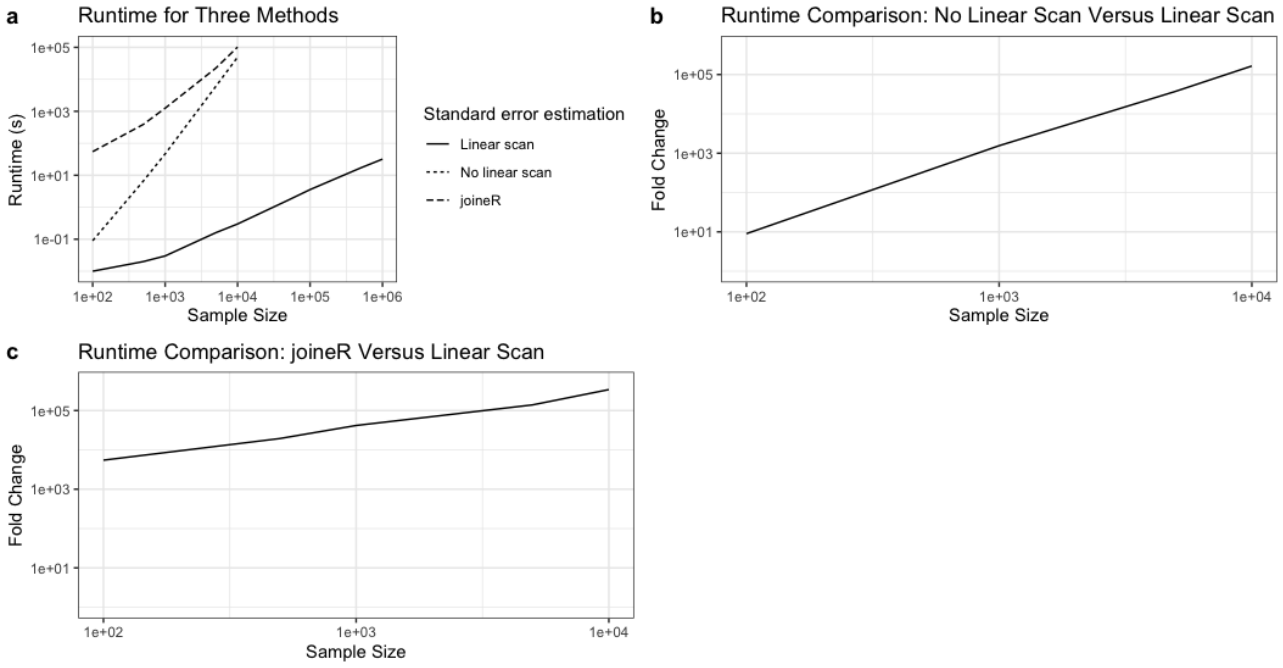

Figure 2.2: Runtime (seconds) comparison between two implementations of standard error estimation for fitting the joint model (2.1) and (2.2): linear scan and no linear scan as described in Section 2.1.2.3, and the bootstrap method employed by the `joineR` package [Philipson et al., 2018]. Fold change is calculated as the ratio of runtime between two methods.

linear scan can generate a speed-up by a factor of greater than 100,000 when $n \geq 10,000$. Similarly, in comparison with `joineR` that used 100 bootstrap samples for standard error estimation, our standard error estimation method with linear scan generated a speed-up by a factor of greater than 100,000 when $n \geq 5,000$.

Finally, Figures 2.1 and 2.2 in Section 2.2 have focused on contrasting the computational efficiency of different implementations for parameter estimation and standard error estimation in terms of the runtime. We have also compared their parameter estimates and standard error in Section A.5 of the supplementary materials. As one would expect, our three different implementations (methods 1-3) yielded almost identical estimation results, whereas `joineR` produced similar estimation results for the longitudinal model, but slightly different results for the competing risks model due to its different latent association structure.

2.3 Real Data Examples

We have developed an R package `FastJM` to implement the efficient algorithms described in Section 2.1. Below we illustrate the improved computational performance of `FastJM` in comparison to existing joint model R packages on a lung health study (LHS) data with $n = 5,887$ subjects and a UK-Biobank data with $n = 193,287$ participants.

2.3.1 Lung health study

The lung health study (LHS) data were collected from a ten-center randomized clinical trial on 5,887 middle-aged smokers with mild to moderate chronic obstructive pulmonary disease (COPD) [Tashkin et al., 2012]. Patients were randomized into three arms: usual care, smoking intervention and placebo inhaler (SIP), and smoking intervention and active bronchodilator inhaler (SIA). An important objective of the study was to determine if the intervention program with the combination of intensive smoking cessation counseling and an inhaled anticholinergic bronchodilator can slow down the decline in forced expired volume in

1s (FEV_1) during a 5-year follow-up period. Patients' FEV_1 values were collected annually upon recruitment into the study. FEV_1 was chosen as the primary outcome since its trajectory is an indicator of a patient's natural loss of lung function during the progression of COPD. Since not all patients completed the whole study period, about 9.47% of longitudinal measurements were missing. One of the possible reasons for dropout is that treatment was not effective, and hence missing longitudinal measurements after dropout are non-ignorable.

Joint modeling of FEV_1 together with the possible informative dropout time provides an attractive approach to deal with nonignorable missing longitudinal data due to dropout. Based on previous findings, we considered the following covariates when characterizing the trajectory of $Y = \text{FEV}_1$: time (year), sex, age, body mass index (BMI), baseline number of cigarettes smoked per day, the logarithm of two-point methacholine concentration-FEV1 O'Connor slope (logslope) [O'Connor et al., 1995]. We also included two interaction terms between treatment indicators SIP and SIA and time, so that the difference in the slope of FEV_1 between SIP (or SIA) and usual care can be evaluated by testing if the interactions are zero or not. Specifically, we considered the following linear mixed effect model:

$$\begin{aligned} Y_{ij} &= \beta_0 + \beta_1 t_{ij} + \beta_2 X_{age_i} + \beta_3 X_{F10CIGS_i} + \beta_4 D_{sex_i} + \beta_5 X_{logslope_i} + \beta_6 X_{BMI_i} \\ &\quad + \beta_7 D_{SIP_i} + \beta_8 D_{SIA_i} + \beta_9 D_{SIP_i} \times t_{ij} + \beta_{10} D_{SIA_i} \times t_{ij} + b_{0i} + b_{1i} t_{ij} + \epsilon_{ij}, \end{aligned} \quad (2.18)$$

which corresponds to model (2.1) with $X_i^{(1)}(t_{ij})^T = (1, t_{ij}, X_{age_i}, X_{F10CIGS_i}, D_{sex_i}, X_{logslope_i}, X_{BMI_i}, D_{SIP_i}, D_{SIA_i}, D_{SIP_i} \times t_{ij}, D_{SIA_i} \times t_{ij})$ and $\tilde{X}_i^{(1)}(t_{ij})^T = (1, t_{ij})$. The random error term $\epsilon_{ij} \overset{\text{iid}}{\sim} N(0, \sigma^2)$ and the random effects $b_i = (b_{0i}, b_{1i})^T$ are assumed normally distributed with zero mean and a covariance matrix Σ. For the dropout time T_i (possibly censored at the end of the study), we assume the Cox proportional hazard sub-model:

$$\lambda_i(t) = \lambda_0(t) \exp\{\gamma_1 X_{BMI_i} + \gamma_2 D_{SIP_i} + \gamma_3 D_{SIA_i} + \gamma_4 X_{logslope_i} + \gamma_5 D_{sex_i} + \gamma_5 X_{age_i} + W_i(t)\}, \quad (2.19)$$

where $\lambda_0(t)$ denotes the baseline hazard function and $W_i(t)$ is a latent association structure that links the two sub-models.

Table 2.1: Runtime comparison between different R packages for joint modeling of a longitudinal and a single event time on the lung health study data

	Semiparametric Joint Models					Parametric Joint Models			
Package	`FastJM`	`joineR`	$\texttt{JSM}_a$	$\texttt{JSM}_b$	$\texttt{JM}_b$	$\texttt{JM}_{a_1}$	$\texttt{JM}_{a_2}$	$\texttt{JMbayes}_a$	$\texttt{JMbayes}_b^*$
Baseline hazard	Unspecified	Unspecified	Unspecified	Unspecified	Unspecified	Weibull	B-spline	B-spline	B-spline
$W_i(t)$	$\nu^T b_i$	$\nu\tilde{X}_i^{(1)}(t)^T b_i$	$\nu m_i(t)$	$\nu\tilde{X}_i^{(1)}(t)^T b_i$	$\nu m_i(t)$	$\nu m_i(t)$	$\nu m_i(t)$	$\nu m_i(t)$	$\nu^T b_i$
Runtime	0.3min	20.4min	1h36min	1h51min	*	0.9min	1min	19.8min	43min

* Failed to produce any result due to convergence issue.

Table 2.1 compares the runtime of `FastJM` and some existing joint model packages including `joineR` [Philipson et al., 2018], different versions of `JM` [Rizopoulos, 2010], `JMBayes` [Rizopoulos, 2014], and `JSM` [Xu et al., 2020] with various specifications of $\lambda_0(t)$ and $W_i(t)$.

Among all the semi-parametric models (`FastJM`, `joineR`, $\texttt{JSM}_a$, $\texttt{JSM}_b$), `FastJM` finished in 0.3 minutes while other methods took 20.4 minutes to 111 minutes. As a matter of fact, the runtime of `FastJM` was even shorter than those of some parametric joint models ($\texttt{JSM}_a$ and $\texttt{JSM}_b$). We also observed that `JMbayes` based on a Bayesian MCMC framework is considerably slower than its frequentist counterpart `JM`. Finally, the parameter estimates and inference results for the longitudinal outcome were almost identical between all packages, but slightly different for the survival sub-model because of their slightly different latent structure $W_i(t)$. Detailed analysis results are summarized in Section A.6 of the supplementary materials.

2.3.2 UK Biobank primary care (UKB-PC) study

The UK Biobank (UKB) is a prospective cohort study with deep genetic and phenotypic data collected on approximately 500,000 individuals, aged 37-73 years, from the general population between 2006 and 2010 in the United Kingdom [Collins, 2012, Sudlow et al., 2015]. Participants attended assessment at their closest clinic center where they completed questionnaires, took physical measurements, and provided biological samples (blood,

urine, and saliva) as a baseline assessment visit. Hospital admission records were available until February, 2018, for the full UKB cohort, whereas linkage to primary care records was available for 45% of the UKB cohort (approximately 230,000 participants) until May, 2017, for Scotland, September, 2017, for Wales, and August, 2017, for England. The detailed linkage procedures relating to primary care records are available online [`https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/primary_care_data.pdf`].

In this example we consider a joint model of longitudinal systolic blood pressure (SBP) measurements and a competing risks event time defined as age-at-onset of Type 2 diabetes (T2D) as the first risk and age-at-onset of stroke, myocardial infarction (MI), or all-cause death as the second risk, whichever occurred first. Age-at-onset of outcomes were based on participants' primary care or hospital records, whichever occurred first. Followup was censored at the primary care data end date for the relevant country or the date of outcomes, if this occurred earlier. SBP measures were extracted from either baseline assessment visit or primary care data. Covariates include sex, ethnicity, and BMI measured during baseline visit. However, considering the imbalanced racial distribution in this case study, we only considered white vs non-whites ethnicity groups. Specifically, the joint model consists of a linear mixed effects model for the longitudinal outcome (SBP),

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 X_{BMI_i} + \beta_3 D_{male_i} + \beta_4 D_{non-white_i} + b_{0i} + b_{1i} t_{ij} + \epsilon_{ij}, \qquad \text{(Model-L)}$$

which corresponds to model (2.1) with $X_i^{(1)}(t_{ij})^T = (1, t_{ij}, X_{BMI_i}, D_{male_i}, D_{non-white_i})$ and $\tilde{X}_i^{(1)}(t_{ij})^T = (1, t_{ij})$, and a proportional cause-specific hazards model for the competing risks event outcome

$$\lambda_{ik}(t) = \lambda_{k0}(t) \exp\{\gamma_{k1} X_{BMI_i} + \gamma_{k2} D_{male_i} + \gamma_{k3} D_{non-white_i} + W_{ik}(t)\}, \qquad \text{(Model-PCH)}$$

for $k = 1, 2$. In Model-L, the random error term $\epsilon_{ij} \overset{\text{iid}}{\sim} N(0, \sigma^2)$ and the random effects $b_i = (b_{0i}, b_{1i})^T$ are assumed normally distributed with zero mean and covariance matrix Σ. In Model-PCH, $k = 1$ denotes type-2 diabetes and $k = 2$ stroke, $\lambda_{k0}(t)$ denotes the baseline

Table 2.2: Runtime comparison between different R packages for semi-parametric joint modeling of longitudinal SBP trajectory and competing risks event time on the UK-Biobank primary care (UKB-PC) data

Package	`FastJM`	`joineR`
UKB-PC subset ($n = 5,000$)	1min	3.3h
UKB-PC subset ($n = 20,000$)	4.4min	33h
UKB-PC full data ($n = 193,287$)	1h	*

* Failed to produce any result due to computational failure.

cause-specific hazard function for cause k, and $W_{ik}(t)$ the latent association structure of SBP with cause k risk.

To our knowledge, besides our `FastJM` package, `joineR`, `JM` are two other current joint model R packages that are capable of handling competing risks event outcomes. However, because `JM` encountered convergence issues, we will focus on `FastJM` and `joineR` in this case study. Table 2.2 compares the runtime of `FastJM` and `joineR` on a subset of 5,000 and 20,000 participants randomly selected from the UKB-PC data and the full UKB-PC data with 193,287 participants.

Table 2.2 shows that for the UKB-PC subset of 5,000 participants, `FastJM` finished within 1 minute, while `joineR` took 3.3 hours to finish. For the UKB-PC subset of 20,000 participants, `FastJM` finished within 5 minutes, while `joineR` took 33 hours to run. For the UKB-PC full data with 193,287 participants, `FastJM` finished within 1 hour, whereas `joineR` encountered a computational failure.

Finally, the analysis results produced by `FastJM` and `joineR` are similar for the longitudinal sub-model for the UKB-PC subset of 5,000 and 20,000 participants and for UKB-PC full data. For the survival sub-model, the analysis results are also similar for most parameters except for the association parameters due to the different latent structure $W_i(t)$ between

two packages. Detailed analysis results are provided in Section A.7 of the supplementary materials.

2.4 Discussion

We have developed customized linear scan algorithms to reduce the computational complexity from $O(n^2)$ or $O(n^3)$ to $O(n)$ within each iteration of the EM algorithm and in the standard error estimation step for a semiparametric joint model of a longitudinal biomarker and a competing risks event time outcome. Through simulation and case studies, we have demonstrated that the efficient implementation can generate a speed-up by a factor of up to hundreds of thousands and oftentimes reduce the runtime from days to minutes when the sample size is large ($n > 10^4$), making it feasible to fit a joint model on a large data in real time.

The ideas and techniques of this paper can potentially be adapted to improve computational efficiency for other joint models. For instance, the linear computational algorithm in Remark 1 for computing the variance-covariance matrices of empirical Bayes estimates of the random effects is not specific to the joint model considered in this paper and can be used in any procedure that uses the pseudo-adaptive quadrature rule. Also, although we have focused on joint modeling of a single biomarker with a time-to-event outcome, our methodology can be easily extended to handle multiple biomarkers in a similar fashion. It is also important to note that the linear risk set scan algorithm is limited to the share random effects joint model in which the Cox sub-model (2.2) only involves time-independent covariates. If the Cox sub-model contains time-dependent covariates such as the present value of the longitudinal marker, then one may have to impose more restrictive assumptions such as assuming a parametric baseline hazard in order to linearize the computation costs with respect to the sample size.

This paper has focused on linearizing the computation with respect to the sample size

within the framework of classical EM algorithm that is coupled with the pseudo-adaptive quadrature rule for numerical integration in the E-step. It would be interesting to investigate if coupling our algorithms with other numerical integration methods such as quasi-Monte Carlo method [Owen, 2013] in the E-step, or with other variations of EM algorithms such as the stochastic EM algoirthm (stEM) [Nielsen, 2000] or Turbo-EM [Bobb and Varadhan, 2021] could further enhance the computational efficiency, especially when there are 3 or more random effects in the model. Finally, current joint model implementations are generally not scalable as the number of longitudinal measurement grows large, rendering it infeasible to fit dense longitudinal data such as those generated from modern wearable devices for dynamic prediction of a health outcome. Future research is warranted to develop novel joint modeling procedures that are scalable to large number of subjects, random effects, and longitudinal measurements.

2.5 Software

A user-friendly R package `FastJM` [Li et al., 2022b] has been developed to fit the shared parameter joint model using the efficient algorithms developed in this paper and is publicly available on the Comprehensive R Archive Network (CRAN) at `https://CRAN.R-project.org/package=FastJM`.

2.6 Data Availability

The lung health study (LHS) data used to support the findings in Section 2.3.1 have not been made available because the authors did not have permission for data sharing from the data provider. The data that support the findings in Section 2.3.2 are available from UK Biobank repositories. The UK Biobank data are retrieved under Project ID: 48152. Data are available at `https://www.ukbiobank.ac.uk` with the permission of UK Biobank.

CHAPTER 3

A joint model of the individual mean and within-subject variability of a longitudinal outcome with a competing risks time-to-event outcome

In this chapter, we discuss a joint modeling approach to take into account of the impact of both individual mean and WS variability of a longitudinal outcome on a time-to-event outcome. We develop an efficient estimation procedure of the joint model via a customized EM algorithm. We perform simulations to compare estimation accuracy and precision between our proposed model and the classical joint model that assumes homogeneous WS variability. Lastly, we apply our proposed model to the Multi-Ethnic Study of Atherosclerosis (MESA) study, showing that our method can yield higher dynamic prediction accuracy than the classical joint model.

3.1 Methods

3.1.1 Model and data specifications

Assume that there are n subjects in the study. For subject i, one observes a longitudinal outcome $Y_i(t)$ at multiple time points t_{ij}, $j = 1, \ldots, n_i$, $i = 1, \ldots, n$. In addition, each subject may experience one of K distinct failure types or be right censored during follow-up. Let $\tilde{T}_i$ denote the failure time of interest, $\tilde{D}_i$ the failure type taking values in $\{1, \ldots, K\}$, and C_i be a censoring time for subject i. Then the observed right-censored competing

risks time-to-event data for subject i has the form $(T_i, D_i) \equiv \left\{\min(\tilde{T}_i, C_i), \tilde{D}_i I(\tilde{T}_i \leq C_i)\right\}$, $i = 1, \ldots, n$.

Assume that the longitudinal outcome $Y_i(t)$ is characterized by the following *mixed-effects multiple location-scale submodel*:

$$\begin{aligned} Y_i(t) &= m_i(t) + \sigma_i(t)\epsilon_i(t), \\ m_i(t) &= X_i^{(1)T}(t)\beta + Z_i^T(t)b_i, \qquad (3.1) \\ \sigma_i^2(t) &= \exp\left\{W_i^T(t)\tau + V_i^T(t)\omega_i\right\}, \qquad (3.2) \end{aligned}$$

where $X_i^{(1)}(t)$, $Z_i(t)$, $W_i(t)$, and $V_i(t)$ are vectors of possibly time-varying covariates, β and b_i represent the fixed effects and random effects, respectively, associated with the location component $m_i(t)$ for the mean trajectory, and τ and ω_i represent the fixed effects and random effects, respectively, associated with the scale component $\sigma_i(t)$ for the WS variability. Assume that the measurement error $\epsilon_i(t) \sim N(0,1)$ is independent of b_i and ω_i, and mutually independent across all time points and subjects, and the random effects $\theta_i \equiv (b_i^T, \omega_i^T)^T$ follow a multivariate normal distribution $MVN(0, \Sigma_\theta)$ with mean 0 and variance-covariance matrix

$$\Sigma_\theta = \begin{pmatrix} \Sigma_{bb} & \Sigma_{b\omega} \\ \Sigma_{b\omega}^T & \Sigma_{\omega\omega} \end{pmatrix},$$

where $\Sigma_{b\omega} = cov(b_i, \omega_i)$, $\Sigma_{bb} = cov(b_i, b_i)$, and $\Sigma_{\omega\omega} = cov(\omega_i, \omega_i)$.

Assume further that the competing risks time-to-event outcome follows the following cause-specific Cox proportional hazards submodel:

$$\begin{aligned} \lambda_k(t \mid X_i^{(2)}(t), M_i(\theta_i, t)) &= \lim_{h \to 0} \frac{P(t \leq \tilde{T}_i < t + h, \tilde{D}_i = k \mid T_i \geq t, X_i^{(2)}(t), M_i(\theta_i, t))}{h} \\ &= \lambda_{0k}(t) \exp\{X_i^{(2)T}(t)\gamma_k + M_i^T(\theta_i, t)\alpha_k\}, \quad k = 1, \ldots, K, \qquad (3.3) \end{aligned}$$

where $\lambda_{0k}(t)$ is a completely unspecified baseline hazard function, $X_i^{(2)}(t)$ is a vector of possibly time-varying covariates for the competing risks time-to-event outcome, γ_k is a vector of fixed effects of $X_i^{(2)}(t)$, $M_i(\theta_i, t)$ is a vector of pre-specified functions of θ_i and t, and α_k is a vector of association parameters between the longitudinal and time-to-event outcomes.

Note that the three submodels (3.1)-(3.3) are linked together via the latent association structure $M_i^T(\theta_i, t)\alpha_k$. Some useful examples of $M_i^T(\theta_i, t)\alpha_k$ include (1) "shared random effects" parameterization: $M_i^T(\theta_i, t)\alpha_k = \alpha_{bk}^T b_i + \alpha_{\omega k}^T \omega_i$, (2) "present value" parameterization: $M_i^T(\theta_i, t)\alpha_k = \alpha_{bk} m_i(t) + \alpha_{\omega k} \log\{\sigma_i^2(t)\}$, and (3) "present value of latent process" parametrization: $M_i^T(\theta_i, t)\alpha_k = \alpha_{bk} Z_i^T(t) b_i + \alpha_{\omega k} V_i^T(t)\omega_i$, where α_{bk} and $\alpha_{\omega k}$ are the association parameters for the individual mean trajectory and WS variability, respectively. It is also worth noting that our joint model (3.1)-(3.3) reduces to the joint model of [Li et al., 2022a] if the submodel (3.2) is replaced by a homogeneous WS variance $\sigma_i^2(t) \equiv \sigma^2$.

Throughout the paper we assume that C_i is independent of $\tilde{T}_i$ conditional on the observed covariates and the unobserved random effects θ_i, $i = 1, \ldots, n$. We further assume that the longitudinal measurements are independent of the competing risks time-to-event data conditional on the observed covariates and the unobserved random effects.

3.1.2 Likelihood and EM estimation

Denote by $\Psi = (\beta,\ \tau,\ \gamma,\ \alpha,\ \Sigma_\theta,\ \lambda_{01}(\cdot),\ ...,\ \lambda_{0K}(\cdot))$ the collection of all unknown parameters and functions from the submodels (3.1)-(3.3), where $\gamma = (\gamma_1^T, \ldots, \gamma_K^T)^T$ and $\alpha = (\alpha_{b1}^T, \ldots, \alpha_{bK}^T, \alpha_{\omega 1}^T, \ldots, \alpha_{\omega K}^T)^T$. Denote by $Y_i = (Y_{i1}, ..., Y_{in_i})^T$, where $Y_{ij} = Y_i(t_{ij})$. Omitting the covariates for the sake of brevity, the observed-data likelihood is given by

$$
\begin{aligned}
L(\Psi; Y, T, D) \ &\propto \ \prod_{i=1}^{n} f(Y_i, T_i, D_i \mid \Psi) \\
&= \ \prod_{i=1}^{n} \int f(Y_i \mid \theta_i, \Psi) f(T_i, D_i \mid \theta_i, \Psi) f(\theta_i \mid \Psi) d\theta_i \\
&= \ \prod_{i=1}^{n} \int \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\sigma_i^2(t_{ij})}} \exp\left[-\frac{\{Y_i(t_{ij}) - m_i(t_{ij})\}^2}{2\sigma_i^2(t_{ij})} \right] \\
&\quad \times \prod_{k=1}^{K} \lambda_k \left\{ T_i \mid X_i^{(2)}(T_i), M_i(\theta_i, T_i) \right\}^{I(D_i = k)}
\end{aligned}
$$

$$\times \exp\left[-\sum_{k=1}^{K}\int_0^{T_i}\lambda_k\left\{t \mid X_i^{(2)}(t), M_i(\theta_i,t)\right\}dt\right]$$
$$\times \frac{1}{\sqrt{(2\pi)^q \mid \Sigma_\theta \mid}}\exp\left(-\frac{1}{2}\theta_i^T\Sigma_\theta^{-1}\theta_i\right)d\theta_i,$$

where the first equality follows from the assumption that Y_i and (T_i, D_i) are independent conditional on the covariates and the random effects.

Because Ψ involves K unknown hazard functions, directly maximizing the above observed-data likelihood is difficult. To tackle this issue, we derive an EM algorithm to compute the semiparametric maximum likelihood estimate (SMLE) of Ψ by regarding the latent random effects θ_i as missing data [Dempster et al., 1977, Elashoff et al., 2008].The complete-data likelihood based on (Y, T, D, θ) is given by

$$\begin{aligned}
L(\Psi; Y, T, D, \theta) \;\propto\; & \prod_{i=1}^{n}\prod_{j=1}^{n_i}\frac{1}{\sqrt{2\pi\exp\left\{W_i^T(t_{ij})\tau + V_i(t_{ij})^T\omega_i\right\}}} \\
& \times \exp\left[-\frac{\left\{Y_i(t_{ij}) - X_i^{(1)T}(t_{ij})\beta - Z_i^T(t_{ij})b_i\right\}^2}{2\exp\left\{W_i^T(t_{ij})\tau + V_i(t_{ij})^T\omega_i\right\}}\right] \\
& \times \prod_{k=1}^{K}\left[\Delta\Lambda_{0k}(T_i)\exp\left\{X_i^{(2)T}(T_i)\gamma_k + M_i^T(\theta_i, T_i)\alpha_k\right\}\right]^{I(D_i=k)} \\
& \times \exp\left[-\sum_{k=1}^{K}\int_0^{T_i}\exp\left\{X_i^{(2)T}(t)\gamma_k + M_i^T(\theta_i,t)\alpha_k\right\}d\Lambda_{0k}(t)\right] \\
& \times \frac{1}{\sqrt{(2\pi)^q \mid \Sigma_\theta \mid}}\exp\left(-\frac{1}{2}\theta_i^T\Sigma_\theta^{-1}\theta_i\right),
\end{aligned}$$

where $\Lambda_{0k}(.)$ is the cumulative baseline hazard function for type k failure and $\Delta\Lambda_{0k}(T_i) = \Lambda_{0k}(T_i) - \Lambda_{0k}(T_i-)$. The EM algorithm iterates between an expectation step (E-step):

$$Q(\Psi; \Psi^{(m)}) \equiv E^{(m)}_{\theta|Y,T,D,\Psi^{(m)}}\left\{\log L(\Psi; Y, T, D, \theta)\right\}, \tag{3.4}$$

and a maximization step (M-step):

$$\Psi^{(m+1)} = \arg\max_{\Psi} Q(\Psi; \Psi^{(m)}), \tag{3.5}$$

until the algorithm converges, where $\Psi^{(m)}$ is the estimate of Ψ from the m-th iteration. Each E-step involves calculating integrals of the form

$$E^{(m)}\{h(\theta_i)\} \quad = \quad \int h(\theta_i) f(\theta_i \mid Y_i, T_i, D_i, \Psi^{(m)}) d\theta_i \tag{3.6}$$

for every subject i, $i = 1, \ldots, n$, which are evaluated using the standard Gauss-Hermite quadrature rule [Press et al., 2007]. As shown in Supplementary Materials B.1.1, the M-step (3.5) has closed-form solutions for a number of parameters including the nonparametric baseline cumulative hazard functions $\Lambda_{0k}(t)$, $k = 1, \ldots, K$, which is a key advantage of the EM-algorithm. Other parameters without closed-form solutions in the M-step are updated using the one-step Newton-Raphson method. Details of the EM algorithm are provided in equations (B.4)-(B.9) of the Supplementary Materials.

3.1.3 Standard error estimation

As discussed in Elashoff et al. [2016] (Section 4.1, p.72), several approaches including profile-likelihood, observed information matrix, and bootstrap method have been proposed in the literature for estimating the standard errors of the parametric components of the SMLE. Here we adopt the profile-likelihood approach because it can be readily computed from the EM algorithm and performed well in our simulation studies. Let $\Omega = (\beta, \tau, \gamma, \alpha, \Sigma_\theta)$ denote the parametric component of Ψ and $\hat{\Omega}$ its SMLE. We propose to estimate the variance-covariance matrix of $\hat{\Omega}$ by inverting the empirical Fisher information obtained from the profile likelihood of Ω [Lin et al., 2004, Zeng et al., 2005, Zeng and Cai, 2005] as follows:

$$\sum_{i=1}^{n} [\nabla_\Omega l^{(i)}(\hat{\Omega}; Y, T, D)][\nabla_\Omega l^{(i)}(\hat{\Omega}; Y, T, D)]^T, \tag{3.7}$$

where $\nabla_\Omega l^{(i)}(\hat{\Omega}; Y, T, D)$ is the observed score vector from the profile-likelihood $l^{(i)}(\Omega; Y, T, D)$ of Ω on the ith subject by profiling out the baseline hazards. Details of the observed score vector for each parametric component are provided in equations (B.10)-(B.14) of the Supplementary Materials.

3.1.4 Computational aspects

We point out in Supplementary Materials B.3 that a naive implementation of the proposed EM algorithm and the standard error estimate for our joint model typically involves $O(n^2)$ and $O(n^3)$ operations, respectively, which will become computationally prohibitive for large scale studies with a massive number of subjects. However, by applying the linear scan algorithms of [Li et al., 2022a], one can reduce the computational complexity of our EM algorithm and standard error estimation to $O(n)$ for the shared random effects joint model (3.1)-(3.3), where $X_i^{(2)}(t)$ is assumed to be time-independent and $M_i^T(\theta_i, t)\alpha_k = \alpha_{bk}^T b_i + \alpha_{\omega k}^T \omega_i$. We have implemented these efficient algorithms and developed an R package, JMH `https://github.com/shanpengli/JMH`, which is capable of fitting the aforementioned shared random effects joint model in real time with tens of thousands patients.

3.1.5 Dynamic prediction for competing risks time-to-event data

The proposed joint model (3.1)-(3.3) not only offers a general framework to model the individual mean and WS variability of a longitudinal outcome and study their association with a competing risks time-to-event outcome, but also facilitates subject-level dynamic prediction of cumulative incidence probabilities of a competing risks event for a new subject i^* based on the longitudinal outcome history. Specifically, given the longitudinal outcome history $Y_{i^*}^{(s)} = \{Y_{i^*}(t_{i^*j}), t_{i^*j} \leq s\}$ prior to a landmark time $s > 0$ and that an event has yet to happen by time s, the cumulative incidence probability for type k failure at a horizon time $u > s$ is given by

$$
\begin{aligned}
P_{i^*k}(u, s|\Psi) &= \Pr(T_{i^*} \leq u, D_{i^*} = k | T_{i^*} > s, Y_{i^*}^{(s)}, \Psi) \\
&= \int \Pr(T_{i^*} \leq u, D_{i^*} = k | T_{i^*} > s, Y_{i^*}^{(s)}, \theta_{i^*}, \Psi) f(\theta_{i^*} | T_{i^*} > s, Y_{i^*}^{(s)}, \Psi) d\theta_{i^*} \\
&= \int \Pr(T_{i^*} \leq u, D_{i^*} = k | T_{i^*} > s, \theta_{i^*}, \Psi) f(\theta_{i^*} | T_{i^*} > s, Y_{i^*}^{(s)}, \Psi) d\theta_{i^*} \\
&= \int \frac{\Pr(T_{i^*} \leq u, D_{i^*} = k, T_{i^*} > s | \theta_{i^*}, \Psi)}{\Pr(T_{i^*} > s | \theta_{i^*}, \Psi)} f(\theta_{i^*} | T_{i^*} > s, Y_{i^*}^{(s)}, \Psi) d\theta_{i^*}
\end{aligned}
$$

$$
\begin{aligned}
&= \int \frac{CIF_{i^*k}(u,s|\theta_{i^*},\Psi)}{S_{i^*}(s|\theta_{i^*},\Psi)} f(\theta_{i^*}|T_{i^*}>s, Y_{i^*}^{(s)},\Psi)d\theta_{i^*} \\
&= \frac{\int \frac{CIF_{i^*k}(u,s|\theta_{i^*},\Psi)}{S_{i^*}(s|\theta_{i^*},\Psi)} f(Y_{i^*}^{(s)}|\theta_{i^*},\Psi) f(T_{i^*}>s|\theta_{i^*},\Psi) f(\theta_{i^*}|\Psi)d\theta_{i^*}}{\int f(Y_{i^*}^{(s)}|\theta_{i^*},\Psi) f(T_{i^*}>s|\theta_{i^*},\Psi) f(\theta_{i^*}|\Psi)d\theta_{i^*}},
\end{aligned} \tag{3.8}
$$

where $S_{i^*}(.) = \exp\left\{-\sum_{k=1}^{K}\int_0^s d\Lambda_k(t|\theta_{i^*},\Psi)\right\}$ is the overall survival function and $CIF_{i^*k}(.) = \int_s^u S_{i^*}(t|\theta_{i^*},\Psi)d\Lambda_k(t|\theta_{i^*},\Psi)$ the cumulative incidence function (CIF) for type k failure and the last step of integral (3.8) can be evaluated using a standard Gauss-Hermite quadrature rule. An estimate of $P_{i^*k}(u,s|\Psi)$ can be obtained by replacing $S_{i^*}(.)$, $CIF_{i^*k}(.)$, and Ψ by their sample estimates $\hat{S}_{i^*}(.)$, $\widehat{CIF}_{i^*k}(.)$, and $\hat{\Psi}$, respectively.

3.2 Simulation studies

3.2.1 Simulation 1

We present extensive simulations to examine the finite sample performance of the proposed joint model with heterogeneous WS variability for the longitudinal outcome. We also demonstrate that ignoring heterogeneous WS variability can lead to biased parameter estimates, biased standard error estimates, invalid inferences, and inferior prediction accuracy.

We first evaluate the performance of parameter estimation, standard error estimation, and confidence intervals in comparison to a classical joint model that ignores heterogeneous WS variability. In this simulation, the longitudinal measurements $Y_i(t_{ij})$ were generated from the mixed-effects multiple location scale model (3.1)-(3.2) with

$$
m_i(t_{ij}) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 t_{ij} + b_i, \tag{3.9}
$$

$$
\sigma_i^2(t_{ij}) = \exp(\tau_0 + \tau_1 X_{1i} + \tau_2 X_{2i} + \tau_3 X_{3i} + \tau_4 t_{ij} + \omega_i), \tag{3.10}
$$

and the competing risks event data were generated from the following proportional cause-

specific hazards models:

$$\lambda_1(t) = \lambda_{01}(t)\exp\{\gamma_{11}X_{1i} + \gamma_{12}X_{2i} + \gamma_{13}X_{3i} + \alpha_{b1}b_i + \alpha_{\omega 1}\omega_i\}, \quad (3.11)$$

$$\lambda_2(t) = \lambda_{02}(t)\exp\{\gamma_{21}X_{1i} + \gamma_{22}X_{2i} + \gamma_{23}X_{3i} + \alpha_{b2}b_i + \alpha_{\omega 2}\omega_i\}, \quad (3.12)$$

where the four sub-models (3.9)-(3.12) are linked together through the shared random effects $\theta_i = (b_i, \omega_i)^T$. Here t_{ij}'s represent the scheduled visiting times for subject i with an increment of 0.25, $X_{1i} \sim Bernoulli(0.5)$, $X_{2i} \sim Uni(-1, 1)$, and $X_{3i} \sim N(1, 4)$. $\theta_i \sim N_2(0, \Sigma_\theta)$, with $\sigma_b^2 = 0.5, \sigma_\omega^2 = 0.5,$ and $\sigma_{b\omega} = \sigma_b\sigma_\omega\rho_{bw}$. The WS variance $\sigma_i^2(t_{ij})$ modeled by equation (3.10) includes a fixed effect induced by a time-varying covariate t_{ij} (visit-to-visit variation), three covariates X_{i1}, X_{i2}, and X_{i3}, and a subject level random effect ω_i. The baseline hazards $\lambda_{01}(t), \lambda_{02}(t)$ are set to constants 0.05 and 0.1, respectively. We simulated non-informative censoring time $C_i \sim Uni(5, 10)$ and let $T_i = min\{T_{i1}^*, T_{i2}^*, C_i\}$ be the observed survival time (possibly censored) for subject i, where T_{i1}^* and T_{i2}^* are independent event times from models (3.11) and (3.12), respectively, $i = 1, \ldots, n$. The longitudinal measurements for subject i are assumed missing when $t_{ij} > T_i$. The censoring rate is about 21%, the rate of event 1 is about 47%, and the rate of event 2 is about 32%. The average number of longitudinal measurements per subject is around 10.

We fitted both our proposed joint model (3.1)-(3.3) with heterogeneous WS variability (Model 1) using our developed methods and R package "JMH" as described in Section 3.1 and a classical joint model (Model 2) with homogeneous WS variance ($\sigma_i^2(t_{ij}) \equiv \sigma^2$) using the R-package FastJM [Li et al., 2022a]. Table 3.1 summarizes simulated results on the bias, sample standard deviations of the parameter estimates (SE), average estimated standard errors of the parameter estimates (Est. SE), and coverage probabilities of 95% confidence intervals (CP) for a scenario with medium correlation ($\rho_{b\omega} = 0.50$) between the random effects, and sample size $n = 800$. Each entry in the table is based on 500 Monte Carlo samples.

In Table 3.1, our proposed joint model method (Model 1) demonstrates a smaller bias

Table 3.1: Comparison of the bias, standard error (SE), estimated standard error (Est. SE), and coverage probability (CP) between the proposed joint model with heterogeneous WS variability (Model 1) and a classical joint model with homogeneous WS variability (Model 2) for the longitudinal outcome ($n = 800$)

		Model 1 (heterogeneous WS variability)				Model 2 (homogeneous WS variability)			
Parameter	True	Bias	SE	Est. SE	CP (%)	Bias	SE	Est. SE	CP (%)
Longitudinal									
Fixed effects									
Mean trajectory									
β_0	5	0.002	0.050	0.049	95.0	0.004	0.052	0.065	98.6
β_1	1.5	<0.001	0.066	0.070	95.6	-0.004	0.071	0.075	95.4
β_2	2	0.001	0.061	0.060	95.8	-0.003	0.063	0.066	96.0
β_3	1	<0.001	0.019	0.018	94.2	-0.001	0.019	0.020	95.8
β_4	2	<0.001	0.002	0.002	97.2	<0.001	0.003	0.003	93.2
WS variability									
τ_0	0.5	0.001	0.048	0.050	95.8	-	-	-	-
τ_1	0.5	-0.001	0.063	0.065	95.2	-	-	-	-
τ_2	-0.2	0.005	0.054	0.057	95.2	-	-	-	-
τ_3	0.2	<0.001	0.017	0.017	94.6	-	-	-	-
τ_4	0.05	<0.001	0.002	0.002	96.0	-	-	-	-
Competing risks									
Fixed effects									
γ_{11}	1	0.008	0.114	0.119	96.2	-0.045	0.111	0.114	94.4
γ_{12}	0.5	0.005	0.096	0.099	94.8	-0.005	0.094	0.095	95.0
γ_{13}	0.5	0.004	0.033	0.034	95.8	-0.012	0.032	0.033	93.8
γ_{21}	-0.5	-0.010	0.150	0.139	95.0	0.028	0.149	0.136	92.6
γ_{22}	0.5	-0.002	0.118	0.116	94.8	-0.002	0.116	0.113	95.0
γ_{23}	0.25	<0.001	0.039	0.038	94.2	0.004	0.038	0.037	93.8
Association									
α_{b1}	0.1	-0.004	0.143	0.150	95.6	**0.211**	0.116	0.124	**62.2**
α_{b2}	-0.1	-0.007	0.144	0.153	96.2	**-0.258**	0.145	0.161	**65.8**
$\alpha_{\omega 1}$	0.5	<0.001	0.139	0.141	95.8	-	-	-	-
$\alpha_{\omega 2}$	-0.5	-0.007	0.142	0.158	96.8	-	-	-	-
Covariance matrix									
of random effects									
σ_b^2	0.5	-0.005	0.045	0.044	93.8	**-0.030**	0.053	0.045	**82.2**
$\sigma_{b\omega}$	0.25	-0.003	0.038	0.040	96.0	-	-	-	-
σ_ω^2	0.5	-0.003	0.031	0.032	95.0	-	-	-	-

Note: Large error in confidence interval coverage probability (CP) compared to the 95% nominal level are highlighted in boldface. Each entry is based on 500 Monte Carlo samples.

for all parameters and standard error estimates and that CPs are close to the 95% nominal level. On the other hand, ignoring the heterogeneous WS variability (Model 2) induced non-negligible bias in some parameter and standard error estimates, leading to significant under-coverage of the associated confidence intervals. For example, the parameter estimate and standard error estimate of α_{b2} are both substantially biased, and its associated confidence interval coverage probability (65.8%) is unreasonably low compared to the 95% nominal level.

We have performed additional simulations with different sample sizes and random effects correlations. The results are consistent with those in Table 3.1 and thus are not included here.

3.2.2 Simulation 2

We present another simulation to demonstrate the improved prediction performance of our proposed joint model (Model 1) over the classical joint model (Model 2) that ignores heterogeneous WS variability for dynamic prediction of a future event probability from a subject's history prior to a landmark time.

Using the same setting as Simulation 1, we generated a random dataset from (3.9) - (3.12) with the sample size of $n = 3,000$. For the simulated dataset, we did 10 times of random splits as 10 different collections of training sets and validation sets based on 4-fold cross-validated mean absolute prediction error (MAPE4) by contrasting the predicted and empirical cumulative incidence rates for 2 different failure types. MAPE4 is defined as below.

Step 1. Randomly partition all the MESA in study subjects into 4 equal-sized disjoint subsets, $\mathcal{D}_n^{(1)}, \ldots, \mathcal{D}_n^{(4)}$.

Step 2. For each $l = 1, \ldots, 4$, designate $\mathcal{D}_n^{(l)}$ as the validation set and the remaining 3 subsets, denoted by $\mathcal{D}_n^{(-l)}$, as the training set. The training set $\mathcal{D}_n^{(-l)}$ is used to fit a joint model, and then for each subject i^* in the validation set $\mathcal{D}_n^{(l)}$, the fitted model is used to perform a dynamic prediction of its risk-k cumulative incidence rate $\hat{P}_{i^*k}(u, s)$ at a

horizon time u from some landmark time s using formula (3.8).

Step 3. Rank the subjects in the validation $\mathcal{D}_n^{(l)}$ according to the predicted risk k cumulative incidence rate $\hat{P}_{i^*k}^{(l)}(u,s)$, and denote by $Q_{kq}^{(l)}$ the q^{th} quartile group, $q = 1, \ldots, 4$. Define the risk k mean absolute prediction error (MAPE) for the q^{th} quartile group $Q_{kq}^{(l)}$ by

$$\text{MAPE}_k^{(l)}(u,s) = \frac{1}{4}\sum_{q=1}^{4}\left|\hat{F}_k(u|s,Q_{kq}^{(l)}) - \frac{\sum I(i^* \in Q_{kq}^{(l)})\hat{P}_{i^*k}(u,s)}{\sum I(i^* \in Q_{kq}^{(l)})}\right|, \tag{3.13}$$

where $\hat{F}_k(u|s,Q_{kq}^{(l)}) = \int_s^u \hat{S}(t|s,Q_{kq}^{(l)})d\hat{H}_k(t|s,Q_{kq}^{(l)})$ is the empirical cumulative risk k incidence in the q^{th} quartile group $Q_{kq}^{(l)}$, with $\hat{S}(t|s,Q_{kq}^{(l)})$ being the Kaplan-Meier estimator of the all risk survival function and $\hat{H}_k(t|s,Q_{kq}^{(l)})$ the Nelson-Aalen estimator of the risk k cause-specific cumulative hazard function within the q^{th} quartile group.

Step 4. Lastly, the 4-fold cross-validated mean absolute prediction error (MAPE4) for risk k is defined as

$$\text{MAPE4}_k(u,s) \;=\; \frac{1}{4}\sum_{l=1}^{4}\text{MAPE}_k^{(l)}(u,s), \quad k = 1,2.$$

In this simulation, we adopt 4-fold cross validation, by setting the landmark time $s = 3$ so that all the longitudinal information of the subjects at risk in a validation set are available just prior to time s, and predicting the future event probabilities of each failure at the predetermined horizon times $u = (4,5,6,7)$. To reduce the variance of prediction error, we repeat this procedure by proceeding with 10 random splits of training and validation sets, and take the average of MAPE4 sores across 10 splits on all the horizon times as specified above. Figure 3.1 shows the trajectories of the average MAPE4 scores of both Model 1 and Model 2 based on 10 random splits of cross validation. We can see that Model 2 (proposed joint model) always outperforms Model 1 (classical joint model) at all horizon times when heterogeneous WS variability exist in a dataset.

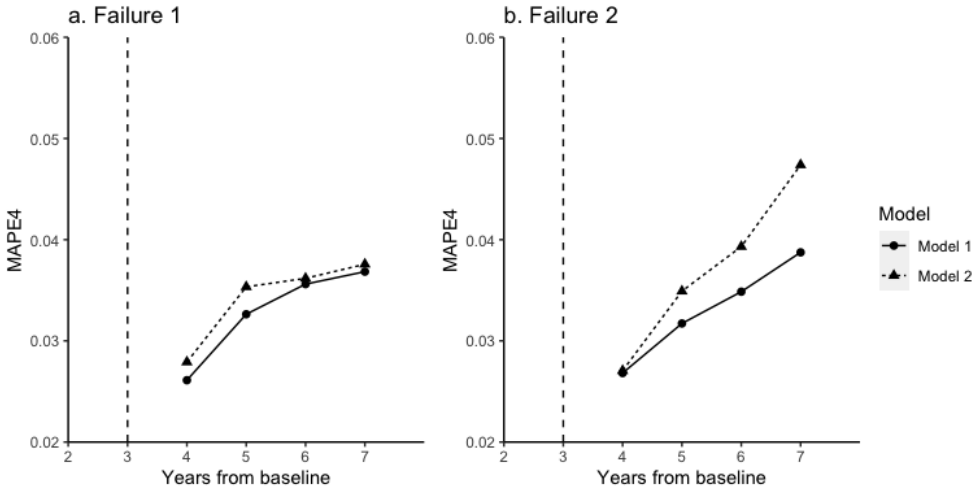

Figure 3.1: Average MAPE4 scores of 10 random splits of cross-validation on the horizon times $u = (4, 5, 6, 7)$ at the landmark time $s = 3$ based on one simulated dataset following the equations (3.9) - (3.12).

3.3 An application: Multi-Ethnic Study of Atherosclerosis (MESA)

The Multi-Ethnic Study of Atherosclerosis (MESA) started in July 2000 and included a community-based sample of 6,814 men and women, who were aged 45-84 years, free of clinical CVD events at baseline, and constituted four racial/ethnic groups from six US field centers [Bild et al., 2002]. The first examination was assigned at the enrollment (i.e., baseline), and the following exams were conducted with up to five exams in total. During the study, blood pressures, i.e., systolic and diastolic blood pressure (SBP and DBP, mmHg), were measured at each exam, and the event surveillance was completely separate from the exams, from phone follow-ups at intervals of 9-12 months. Eligible events such as CVD were collected and abstracted for central adjudication. As an illustration, we consider joint modeling of longitudinal SBP with two competing risk time-to-event outcomes: heart failure (risk 1) and death (risk 2). We adjust for age at baseline (in years), sex (1=female, 0=male), and race (1=non-white, 0=white) in both SBP and competing risk time-to-event models. After excluding 29 individuals missing the event time, a cohort of 6,785 participants with an

average of 4.3 SBP measures per subject and a total of 29,283 SBP measures are kept for analysis. Among them, 380 (5.6%) participants had heart failure, 1,327 (19.6%) died, and 5,078 (74.8%) were right censored without experiencing any of the two events.

In our joint model (Model 1), the longitudinal SBP ($Y_i(t_{ij})$) is assumed to follow the mixed-effects multiple location scale model (3.1)-(3.2) as

$$\begin{aligned} m_i(t_{ij}) &= \beta_0 + \beta_1 t_{ij} + \beta_2 \text{Age}_i + \beta_3 \text{Race}_i + \beta_4 \text{Sex}_i + b_i, && (3.14) \\ \sigma_i^2(t_{ij}) &= \exp(\tau_0 + \tau_1 t_{ij} + \tau_2 \text{Age}_i + \tau_3 \text{Race}_i + \tau_4 \text{Sex}_i + \omega_i), && (3.15) \end{aligned}$$

and the competing risk outcomes, heart failure, and death are modeled by the following cause-specific hazards model:

$$\begin{aligned} \lambda_1(t) &= \lambda_{01}(t) \exp(\gamma_{11} \text{Age}_i + \gamma_{12} \text{Race}_i + \gamma_{13} \text{Sex}_i + \alpha_{b1} b_i + \alpha_{\omega 1} \omega_i), && (3.16) \\ \lambda_2(t) &= \lambda_{02}(t) \exp(\gamma_{21} \text{Age}_i + \gamma_{22} \text{Race}_i + \gamma_{23} \text{Sex}_i + \alpha_{b2} b_i + \alpha_{\omega 2} \omega_i), && (3.17) \end{aligned}$$

where the shared random effects $\theta_i = (b_i, \omega_i)^T$ follow bivariate normal distribution with mean zero and variance-covariance matrix

$$\Sigma_\theta = \begin{pmatrix} \sigma_b^2 & \sigma_{b\omega} \\ \sigma_{b\omega} & \sigma_\omega^2 \end{pmatrix}.$$

Several more complex models were also explored, including additional covariate adjustment, their interactions, and random slopes. The results are similar; thus, we only report results for the shared random intercept model in Table 3.2. For comparison purposes, Table 3.2 also includes an analysis based on the following classical joint model (Model 2) of SBP, heart failure, and death, assuming homogeneous SBP WS variability:

$$\begin{aligned} Y_i(t_{ij}) &= m_i(t_{ij}) + \sigma \epsilon(t_{ij}), && (3.18) \\ m_i(t_{ij}) &= \beta_0 + \beta_1 t_{ij} + \beta_2 \text{Age}_i + \beta_3 \text{Race}_i + \beta_4 \text{Sex}_i + b_i, && (3.19) \\ \lambda_1(t) &= \lambda_{01}(t) \exp(\gamma_{11} \text{Age}_i + \gamma_{12} \text{Race}_i + \gamma_{13} \text{Sex}_i + \alpha_{b1} b_i), && (3.20) \\ \lambda_2(t) &= \lambda_{02}(t) \exp(\gamma_{21} \text{Age}_i + \gamma_{22} \text{Race}_i + \gamma_{23} \text{Sex}_i + \alpha_{b2} b_i). && (3.21) \end{aligned}$$

Table 3.2: Joint analysis of systolic blood pressure (SBP, mmHg), time to heart failure, and time to death using the MESA data (Model 1 is the joint model (3.14)-(3.17) assuming heterogeneous SBP individual mean trajectory and heterogeneous WS variability and Model 2 is the joint model (3.18)-(3.21) assuming heterogeneous SBP individual mean trajectory and homogeneous SBP WS variability. Abbreviations: SE=standard error; HR=hazard ratio; CI=confidence interval.)

	Model 1		Model 2
Longitudinal outcome	Mean trajectory	WS variability	Mean trajectory
(Systolic blood pressure (SBP, mmHg))	Estimate (SE)	Estimate (SE)	Estimate (SE)
Intercept	79.59 (0.74)***	2.34 (0.08)***	79.55 (1.31)***
Time (Years from baseline)	-0.02 (0.03)	0.08 (<0.01)***	-0.14 (0.03)***
Age at baseline	0.66 (0.01)***	0.03 (<0.01)***	0.67 (0.02)***
Race (Non-White/White)	6.58 (0.24)***	0.23 (0.02)***	5.95 (0.41)***
Sex (Female/Male)	-0.22 (0.26)	0.24 (0.03)***	0.89 (0.42)*
Sex × Time	0.26 (0.04)***	-0.02 (0.01)**	0.21 (0.04)***
Random effects			
(variance-covariance matrix)	Estimate (SE)		Estimate (SE)
σ_b^2	204.15 (4.45)***		213.49 (5.42)***
$\sigma_{b\omega}$	7.34 (0.23)***		N/A
σ_ω^2	0.46 (0.02)***		N/A
Cause-specific hazard			
(Heart failure)	HR (95% CI)		HR (95% CI)
Age at baseline	1.08 (1.07-1.10)***		1.08 (1.07-1.09)***
Race (Non-White/White)	1.00 (0.81-1.23)		0.96 (0.78-1.18)
Sex (Female/Male)	0.57 (0.46-0.70)***		0.58 (0.47-0.71)***
Random effects			
Mean trajectory (α_{b1})	0.99 (0.97-1.01)		1.03 (1.02-1.03)***
WS variability ($\alpha_{\omega 1}$)	2.04^+ (1.42-2.94)***		N/A
Cause-specific hazard			
(Death)	HR (95% CI)		HR (95% CI)
Age at baseline	1.11 (1.10-1.12)***		1.10 (1.10-1.11)***
Race (Non-White/White)	1.11 (0.99-1.24)		1.08 (0.97-1.21)
Sex (Female/Male)	0.64 (0.57-0.71)***		0.64 (0.58-0.72)***
Random effects			
Mean trajectory (α_{b2})	0.98 (0.97-0.99)***		1.01 (1.00-1.01)***
WS variability ($\alpha_{\omega 2}$)	1.91^+ (1.55-2.35)***		N/A

* p-value<0.05; ** p-value<0.01; *** p-value<0.001.

$^+$ Standardized HR of association parameter for WS variability is reported to show the effect on the cause-specific hazard by 1 SD change of WS variability of SBP.

There are several interesting observations from the results (Table 3.2). Our proposed model (Model 1) reveals that the heterogeneity of SBP WS variability is substantial ($\sigma^2_\omega = 0.46, \text{SE} = 0.02$, p-value$< 0.001$) across MESA individuals. It also suggests that standardized SBP WS variability (ω_i/σ_ω) is highly predictive of heart failure (HR = 2.04, 95% CI=(1.42, 2.94)) and death (HR = 1.91, 95% CI=(1.55, 2.35)). As SBP WS variability (ω_i) and individual mean (b_i) are two highly correlated covariates, caution needs to be taken when interpreting their effects. In contrast, Model 2 only relates subject-specific SBP to the event outcomes besides the baseline covariates. Results of Model 2 show that every 10 mmHg elevation of subject-specific SBP contributes to 34% and 10% increased risk of heart failure and death, respectively.

Next, we assess and compare the dynamic prediction accuracy of Model 1 and Model 2 using a 4-fold cross-validated mean absolute prediction error (MAPE4) by contrasting the predicted and empirical cumulative incidence rates for heart failure and death, respectively.

Figure 3.2 depicts the 4-fold cross-validated MAPE4 score with landmark time $s = 7$ and horizon times $u = 9, 11$, and 13 years from baseline for heart failure (risk 1) and death (risk 2). Figure 3.2b demonstrated that Model 1 outperforms both Model 2 for dynamic prediction of the risk of death at all three horizon times u =9, 11, and 13, with a noticeable smaller MAPE4 prediction error. Given the incidence rate of heart failure is low ($< 6\%$), MAPE4 may not be an informative prediction accuracy measure.

Figure 3.3 provides an illustration of the significant impact of SBP WS variability on the event outcomes. In Figure 3.3, we ranked all subjects by the empirical Bayes estimate of the random effect ω_i based on Model 1 and then plotted the cumulative incidence of heart failure (Figure 3.3a) and death (Figure 3.3b) for the top 20% high SBP WS variability group versus the bottom 20% low SBP WS variability group. Our results demonstrate participants in the top 20% high SBP WS variability group had a much higher cumulative incidence than the bottom 20% low SBP WS variability group for both clinical events. The same phenomenon is also observed in the spaghetti profile plot of the longitudinally measured

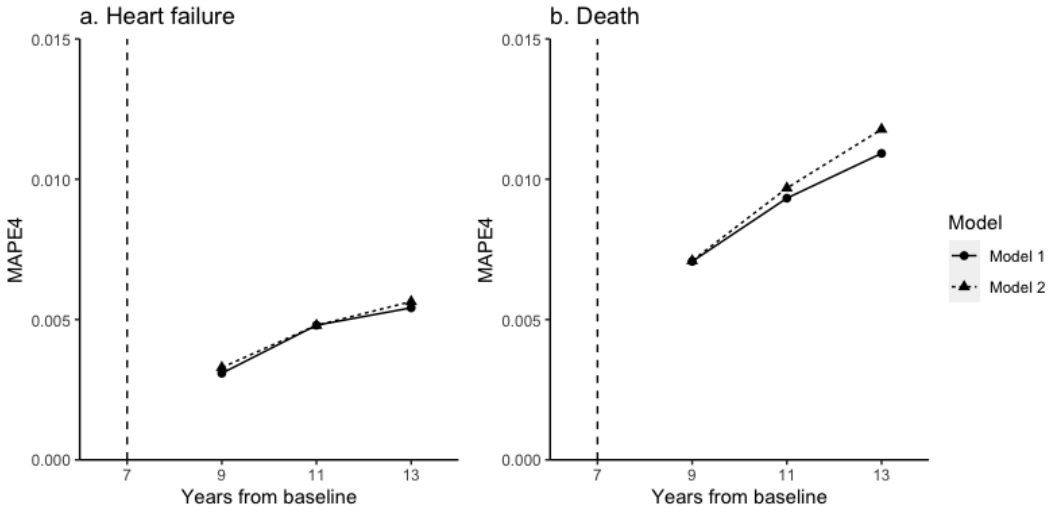

Figure 3.2: The average 4-fold cross-validated MAPE4 score at $s = 7$ and $u = (9, 11, 13)$ years from baseline based on 10 random splits. Model 1 is the joint model (3.14)-(3.17) assuming heterogeneous SBP individual mean trajectory and heterogeneous WS variability, Model 2 is the joint model (3.18)-(3.21) assuming heterogeneous SBP individual mean trajectory and homogeneous SBP WS variability.

SBP for 25 randomly selected participants from each group over the study period, as shown in Figure 3.4, where both events are more likely to happen among the high WS variability cohort than the low WS variability cohort.

3.4 Discussion

Current literature on joint models of longitudinal and time-to-event outcomes has focused primarily on modeling the mean trajectory of a longitudinal biomarker and associating it with a time-to-event outcome. To our knowledge, the joint model introduced in this paper is the first to model both the individual mean and WS variability of a longitudinal biomarker and to utilize both features for dynamic prediction of the time-to-event outcomes. We have developed an EM algorithm for SMLE, a profile-likelihood method for standard error esti-

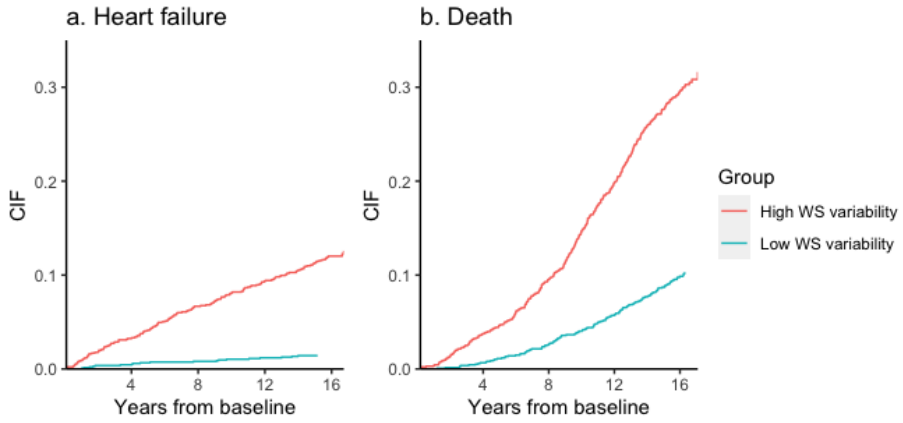

Figure 3.3: Empirical cause-specific cumulative incidence function (CIF) of the high (top 20%, red) and low (bottom 20%, blue) WS variability of SBP (mmHg) groups for heart failure (left) and death (right) in MESA data.

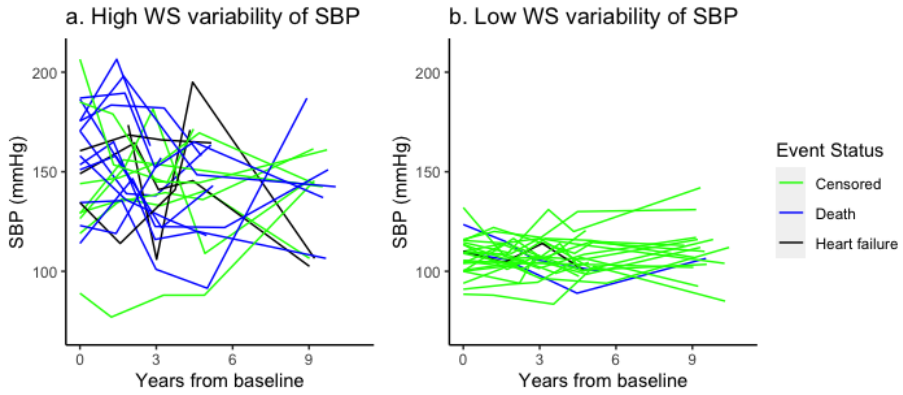

Figure 3.4: Spaghetti profile plot of 25 randomly selected participants from the high (top 20%, left) and low (bottom 20%, right) SBP WS variability groups in the MESA cohort.

mation, and scalable linear scan algorithms for large data. The advantages of our proposed model over classical joint modeling approach are demonstrated using simulations as well as through an application to the MESA study. Our analysis of the MESA study not only reveals that the SBP WS variability is another important predictor for heart disease besides the SBP mean level but also leads to a better dynamic prediction model for heart failure and death by accounting for the mean and WS variability of SBP in the joint model.

The utility of joint modeling of mean trajectory and WS variability of a longitudinal biomarker with a time-to-event outcome is not limited to the simple setting considered in this paper. It could be informative to extend the current work to many other applications, including joint models with multiple longitudinal biomarkers, multivariate time-to-event outcomes, recurrent events, and other types of time-to-event data such as left-truncated or interval-censored data. Extensions to more flexible models for the time trend of the longitudinal outcome are also warranted.

It should be noted that the estimation method based on the EM algorithm may become computationally infeasible as the number of random effects, the number of longitudinal measurements, or the number of subjects increases. For example, the EM algorithm involves numerically evaluating multiple intractable integrals for all subjects in each EM iteration, and the computational cost of the standard Gauss Hermite quadrature rule used in this paper for numerical integration grows exponentially with the number of random effects. In future research, further development of scalable estimation methods is warranted.

Lastly, our model and inference procedure are based on normality assumptions for both the random effects and WS error term in the longitudinal submodel. It would be of interest to develop more robust joint models and inference procedure in future research.

3.5 Software

A user-friendly R package **JMH** to fit the shared parameter joint model developed in this paper is publicly available at `https://github.com/shanpengli/JMH`.

3.6 Acknowledgement

The authors thank the investigators, staff, and participants of MESA for their valuable contributions. A full list of participating MESA investigators and institutions can be found at `http://www.mesa-nhlbi.org`.

3.7 Funding

This research was partially supported by National Institutes of Health (P30 CA-16042, UL1TR000124-02, and P01AT003960, GL), the National Institute of General Medical Sciences (R35GM141798, HZ), the National Human Genome Research Institute (R01HG006139, HZ and JJZ), the National Science Foundation (DMS-2054253, HZ and JJZ; IIS-2205441, GL, HZ, and JJZ), the National Heart, Lung, and Blood Institute (R21HL150374, JJZ, 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, and N01-HC-95169, GL), and the National Center for Advancing Translational Sciences (UL1-TR-000040, UL1-TR-001079, and UL1-TR-001420, GL). This paper has been reviewed and approved by the MESA Publications and Presentations Committee.

CHAPTER 4

A joint model of longitudinal and interval-censored post-diagnosis time to event data in the presence of interval censored covariates due to the unknown diagnosis time: with applications to large-scale biobank data

In this chapter, we develop a novel joint modeling approach to take into account of interval-censored time-to-event and interval-censored covariates due to the unknown initial event time. We develop an estimation procedure of the joint model via a customized EM algorithm. We perform simulations to compare estimation accuracy and precision between our proposed model and the classical joint model with midpoint imputation. We also apply computationally efficient algorithms to make the proposed method scalable to large sample size data. Lastly, we apply our proposed model to the UK-Biobank (UKB) data and demonstrate that our method can yield more clinically plausible parameter estimates and statistical inference.

4.1 Joint model

4.1.1 Notations and preliminaries

Let n be the number of subjects and let $i \in \{1, 2, .., n\}$ index the subjects. For each subject, let $\tilde{B}_i$ denote the calendar time of origin (e.g. date of birth), $\tilde{S}_i$ the calendar time of occurrence of an initial event (e.g. diagnosis of diabetes). We formulate the interval censoring interval by assuming a random, independent set of inspection times, denoted by $0 < \tilde{U}_{i1} < \ldots \tilde{U}_{ik_i} < +\infty$, with k_i the total number of inspection times for subject i [Zeng et al., 2016]. We do not model $(\tilde{U}_{i1}, \ldots, \tilde{U}_{ik_i})$. For subject i, instead of observing $\tilde{S}_i$, the interval such that $\tilde{S}_i \in (\tilde{U}_{ij}, \tilde{U}_{i(j+1)}]$ is observed and thus denote the calendar time to the last negative diagnosis $\tilde{L}_i = \tilde{U}_{ij}$ and the calendar time to the first positive diagnosis $\tilde{R}_i = \tilde{U}_{i(j+1)}$. Let $\left\{\tilde{O}_{ij}, Y_i(t_{ij})\right\}$, $j \in \{1, .., n_i\}$ be the post-initial-event observations, where $\tilde{O}_{ij}$'s are calendar observation times, $t_{ij} = \tilde{O}_{ij} - \tilde{S}_i$ the observation time since the initial event time $\tilde{S}_i$, and $Y_i(t_{ij})$ the value of a longitudinal biomarker collected at the corresponding observation time $\tilde{O}_{ij}$. Let $\tilde{\tilde{E}}_i$ be the true calendar time for the target event, which is subject to right censoring by a censoring time $\tilde{C}_i$. Let $\tilde{E}_i = \min(\tilde{\tilde{E}}_i, \tilde{C}_i)$ be the observed calendar time for the target event and $\delta_i = I(\tilde{\tilde{E}}_i \leq \tilde{C}_i)$ the censoring indicator, with $\delta_i = 1$ if the condition $\tilde{E}_i \leq \tilde{C}_i$ is satisfied, and 0 otherwise. As a result, the observed data is $(\tilde{B}_i, \tilde{L}_i, \tilde{R}_i, \{\tilde{O}_{ij}\}, \{Y_i(t_{ij})\}, \tilde{\tilde{E}}_i, \delta_i), i = 1, \ldots, n$. Figure 4.1 shows an illustrative example of the notations for subject i.

Denote $S_i = \tilde{S}_i - \tilde{B}_i$ as the time to occurrence of the initial event since time of origin, $T_i = \tilde{E}_i - \tilde{S}_i$ as the time since the initial event, $L_i = \tilde{L}_i - \tilde{B}_i$, and $R_i = \tilde{R}_i - \tilde{B}_i$. Let $\tilde{\boldsymbol{B}} = (\tilde{B}_1, \ldots, \tilde{B}_n)$, $\boldsymbol{L} = (L_1, \ldots, L_n)$, $\boldsymbol{R} = (R_1, \ldots, R_n), \tilde{\boldsymbol{O}} = (\tilde{O}_1, \ldots, \tilde{O}_n)$ with $\tilde{O}_i = (\tilde{O}_{i1}, \ldots, \tilde{O}_{in_i})$, $\boldsymbol{S} = (S_1, \ldots, S_n)$, $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)$ with $\boldsymbol{Y}_i = \{Y_i(t_{i1}), \ldots, Y_i(t_{in_i})\}$, and $\boldsymbol{T} = (T_1, \ldots, T_n)$. Our interest is to (1) jointly model $\boldsymbol{S}, \boldsymbol{Y}, \boldsymbol{T}$ and (2) use the proposed joint model to make dynamic prediction of $\boldsymbol{T}$ based on the post-initial-event observations $(\tilde{\boldsymbol{B}}, \boldsymbol{L}, \boldsymbol{R}, \tilde{\boldsymbol{O}}, \boldsymbol{Y})$.

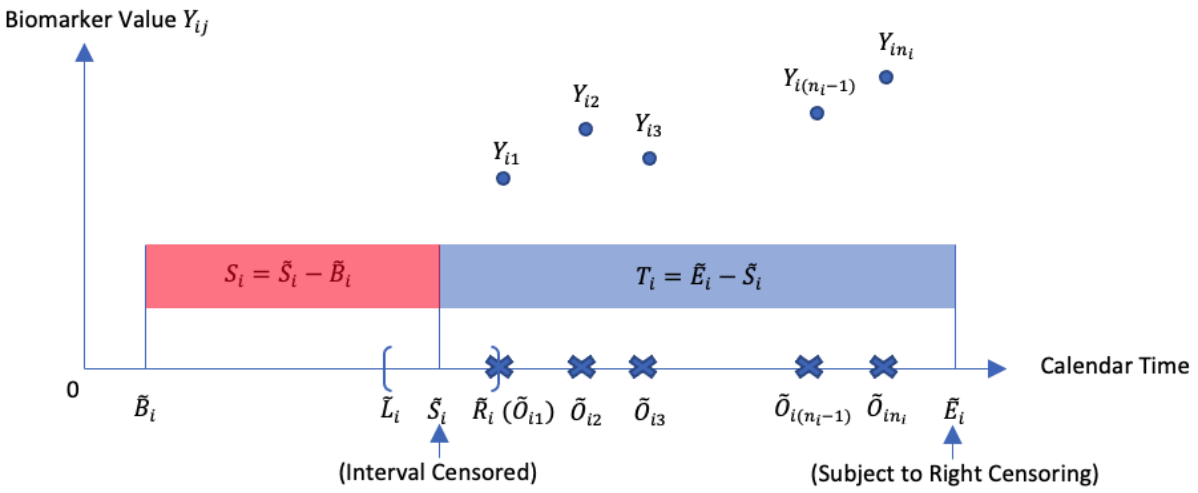

Figure 4.1: An example of longitudinal and event process after the unknown initial event occurs at S_i for subject i.

4.1.1.1 Model formulation

We assume that the initial event time S_i is continuous but only observed at the censoring interval $(L_i, R_i]$.

Longitudinal sub-model for the biomarker The longitudinal outcome $Y_i(t_{ij})$ is characterized by a *mixed-effects multiple location scale model*:

$$\begin{aligned} Y_i(t_{ij}) &= m_i(t_{ij}) + \sigma_i(t_{ij})\epsilon_i(t_{ij}), \\ &= X_i^{(1)T}(t_{ij}, S_i)\beta + Z_i^T(t_{ij})b_i + \sigma_i(t_{ij})\epsilon_i(t_{ij}), \qquad (4.1) \\ \sigma_i^2(t_{ij}) &= \exp\left\{W_i^T(t_{ij}, S_i)\tau + V_i^T(t_{ij})\omega_i\right\}, \qquad (4.2) \end{aligned}$$

where $X_i^{(1)}(t_{ij}, S_i)$, $Z_i(t_{ij})$, $W_i(t_{ij}, S_i)$, and $V_i(t_{ij})$ are vectors of possibly time-varying covariates measured at $t_{ij} = \tilde{O}_{ij} - S_i - \tilde{B}_i$. The simplest case is to consider $X_i^{(1)}(t_{ij}, S_i) = W_i(t_{ij}, S_i) = (1, S_i, t_{ij})^T$. β and b_i represent the fixed effects and random effects, respectively, associated with the location component $m_i(t)$ for the mean trajectory, and τ and ω_i represent the fixed effects and random effects, respectively, associated with the scale component $\sigma_i(t)$

for the WS variability. Assume that the measurement error $\epsilon_i(t) \sim N(0,1)$ is independent of b_i and ω_i, and mutually independent across all time points and subjects, and the random effects $\theta_i \equiv (b_i^T, \omega_i^T)^T$ follow a multivariate normal distribution $MVN(0, \Sigma_\theta)$ with mean 0 and variance-covariance matrix

$$\Sigma_\theta = \begin{pmatrix} \Sigma_{bb} & \Sigma_{b\omega} \\ \Sigma_{b\omega}^T & \Sigma_{\omega\omega} \end{pmatrix},$$

where $\Sigma_{b\omega} = cov(b_i, \omega_i)$, $\Sigma_{bb} = cov(b_i, b_i)$, and $\Sigma_{\omega\omega} = cov(\omega_i, \omega_i)$.

Cox sub-model for the target event The target event outcome is characterized by a Cox proportional hazards model:

$$\begin{aligned} \lambda_i(t) &= \lim_{h\to 0} \frac{P(t \le T_i < t+h | T_i \ge t, X_i^{(2)}, \theta_i)}{h} \\ &= \lambda_0(t) \exp\{X_i^{(2)T}(S_i)\gamma + M_i^T(\theta_i, t)\alpha\}, \end{aligned} \tag{4.3}$$

with $T_i = \tilde{E}_i - S_i - \tilde{B}_i$ (e.g. time to cardiovascular disease since the diagnosis of diabetes). It is worth noting that since S_i is interval censored and $\tilde{E}_i$ is possibly right censored, T_i is possibly interval censored. In the sub-model (4.3), $\lambda_i(t)$ is the conditional hazard rate for at time t since the initial event, given the covariates $X_i^{(2)}(.)$ measured at time S_i and the latent association structure $M_i(\theta_i, t)$. The simplest case is to consider $X_i^{(2)}(S_i) = S_i$.

Note that the three submodels (4.1)-(4.3) are linked together via the latent association structure $M_i^T(\theta_i, t)\alpha$. Some useful examples of $M_i^T(\theta_i, t)\alpha$ include (1) "shared random effects" parameterization: $M_i^T(\theta_i, t)\alpha = \alpha_b^T b_i + \alpha_\omega^T \omega_i$, (2) "present value" parameterization: $M_i^T(\theta_i, t)\alpha = \alpha_b m_i(t) + \alpha_\omega \log\{\sigma_i^2(t)\}$, and (3) "present value of latent process" parametrization: $M_i^T(\theta_i, t)\alpha_k = \alpha_b Z_i^T(t) b_i + \alpha_\omega V_i^T(t)\omega_i$, where α_b and α_ω are the association parameters for the individual mean trajectory and WS variability, respectively.

Throughout the paper we assume that 1) the calendar time to the target event $\tilde{E}_i$ is always observed to be after the calendar time to the first positive diagnosis $\tilde{R}_i$; 1) $\tilde{C}_i - \tilde{S}_i$ is independent of T_i; 2) the longitudinal process is independent of the time-to-event process conditional on the observed covariates, θ_i, and S_i; 3) θ_i and S_i are mutually independent.

4.1.2 Semi-parametric maximum likelihood estimate (SMLE)

4.1.2.1 Assumptions

Non-parametric estimation for the initial event distribution Similar to Zeng et al. [2016], to derive the estimation procedure, we restrict the survival function for S_i to have a finite support $\mathbb{S} = \{l : 0 = s_0 < s_1 < \ldots < s_k < +\infty\}$ consisting of the unique values of L_i and R_i $(i = 1, \ldots, n)$. Under $\mathbb{S}$, the survival function of the initial event $S(.)$ will be evaluated as a non-increasing step function. Let ϕ_{s_l} be the hazard rate at time $s_l \in \mathbb{S}$, then the probability that S_i equals s_l is given by

$$\begin{aligned} p(S_i = s_l|\Psi_\phi) &= S(s_l^-) - S(s_l) \\ &= \exp\left(-\sum_{j:s_j<s_l} \phi_{s_j}\right) - \exp\left(-\sum_{j:s_j\leq s_l} \phi_{s_j}\right), \quad l = 1, 2, \ldots, k. \end{aligned} \tag{4.4}$$

To enable our joint modeling analysis, we make three key modeling assumptions about the relationships among these variables. First, the longitudinal outcome $\boldsymbol{Y}_i$ and the target event outcome T_i are independent conditional on the covariates, random effects θ_i, and time to the-initial-event S_i. Second, θ_i and S_i are independent. Third, the interval censoring mechanism (L_i, R_i) is a non-informative and independent process, that is, $p(L_i, R_i|S_i) = \mathbf{1}\{S_i \in (L_i, R_i]\}\, p(L_i, R_i)$, where $p(L_i, R_i)$ is the probability that subject i's pre-diabetic follow-up schedule includes tests at L_i followed by R_i.

4.1.2.2 Likelihood

For subject i the observed data is $\boldsymbol{D}_i = \{L_i, R_i, T_i, \delta_i, \boldsymbol{Y}_i, \tilde{\boldsymbol{O}}_i\}$. Let $\Psi = (\Psi_Y^T, \Psi_T^T, \Psi_\theta^T, \Psi_\phi^T)^T$ be a vector of all unknown parameters from the submodels (4.1) - (4.3), where $\Psi_Y^T = (\beta, \tau)$, $\Psi_T^T = (\gamma, \alpha, \nu, \Lambda_0(.))$, $\Psi_\theta^T = \text{vech}(\Sigma_\theta)$, and $\Psi_\phi^T = \{\phi_{s_l}\}$. The joint distribution of $(\boldsymbol{Y}_i, T_i, \delta_i, L_i, R_i)$ is fully determined by $p_i(\boldsymbol{Y}_i|S_i, \theta_i, \Psi_Y), p_i(T_i, \delta_i|S_i, \theta_i, \Psi_T), p(\theta_i|\Psi_\theta), p_i(S_i|\Psi_\phi)$, and $p(L_i, R_i|S_i)$. For the sake of brevity, here we omit the covariates $X_i^{(1)}(.,.), Z_i(.), w_i(.,.)$, and $X_i^{(2)}(.)$ from the subject-specific density functions $p_i(.)$. The observed-data log likelihood

is given by

$$\begin{aligned} l(\Psi) &= \sum_{i=1}^{n} \log p_i(\boldsymbol{Y}_i, T_i, \delta_i, L_i, R_i|\Psi) \\ &= \sum_{i=1}^{n} \log \int\int p_i(\boldsymbol{Y}_i|S_i, \theta_i, \Psi_Y) p_i(T_i, \delta_i|S_i, \theta_i, \Psi_T) p(\theta_i|\Psi_\theta) \\ &\quad \times p(L_i, R_i|S_i) p(S_i|\Psi_\phi) dS_i d\theta_i. \end{aligned} \tag{4.5}$$

Since S_i is assumed to be restricted on the finite support $\mathbb{S}$, the log likelihood (4.5) can be rewritten as

$$\begin{aligned} l(\Psi) &= \sum_{i=1}^{n} \log \int \sum_{s_l \in \mathbb{S}} p_i(\boldsymbol{Y}_i|S_i = s_l, \theta_i, \Psi_Y) p_i(T_i, \delta_i|S_i = s_l, \theta_i, \Psi_T) p(\theta_i|\Psi_\theta) \\ &\quad p(L_i, R_i|S_i = s_l) p(S_i = s_l|\Psi_\phi) d\theta_i \\ &= \sum_{i=1}^{n} \log \int \sum_{s_l \in \mathbb{S}} \left(\prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi \exp(W_i^T(t_{ij,l}, s_l)\tau + \omega_i)}} \right. \\ &\quad \times \exp\left\{ -\frac{(Y_i(t_{ij,l}) - X_i^{(1)T}(t_{ij,l}, s_l)\beta - Z_i^T(t_{ij,l})b_i)^2}{2\exp(W_i^T(t_{ij,l}, s_l)\tau + \omega_i)} \right\} \\ &\quad \times \left[\Delta\Lambda_0(t_{i,s_l}) \exp\left\{ X_i^{(2)T}(s_l)\gamma + W_i^T(\theta_i|t_{i,s_l})\alpha \right\} \right]^{\delta_i} \\ &\quad \exp\left[-\int_0^{T_i} \exp\left\{ X_i^{(2)T}(s_l)\gamma + W_i^T(\theta_i|s)\alpha \right\} d\Lambda_0(s) \right] \\ &\quad \times \frac{1}{\sqrt{(2\pi)^{q+1}|\Sigma_\theta|}} \exp\left(-\frac{1}{2}\theta_i^T \Sigma_\theta^{-1} \theta_i \right) \\ &\quad \left. \times \mathbf{1}\{s_l \in (L_i, R_i]\} p(L_i, R_i) \left\{ \exp(-\sum_{s_u < s_l} \phi_{s_u}) - \exp(-\sum_{s_u \le s_l} \phi_{s_u}) \right\} \right) d\theta_i, \end{aligned}$$

where $\Lambda_0(.)$ is the unknown cumulative baseline hazard function, $\Delta\Lambda_0(t_{i,s_l}) = \Lambda_0(t_{i,s_l}) - \Lambda_0(t_{i,s_l}-)$, and $t_{ij,l} = \tilde{O}_{ij} - s_l - \tilde{B}_i$ and $t_{i,s_l} = \tilde{E}_i - s_l - \tilde{B}_i$ when $S_i = s_l$.

4.1.2.3 EM algorithm

One can estimate Ψ by maximizing the observed-data log likelihood $l(\Psi)$. However, the log likelihood of each subject consists of the logarithm of a sum, which becomes computationally challenging to maximize directly, and following the EM-steps will often increase $l(\Psi)$ toward a local maximum or saddlepoint [Morrison et al., 2021]. To address this issue, we can

obtain the maximum likelihood estimate $\hat{\Psi}$ by maximizing its complete-data log likelihood $l^*(\Psi;\theta,S)$ via the EM algorithm:

$$\begin{aligned} l^*(\Psi;\theta,S) \quad \propto \quad & \sum_{i=1}^{n} \log p_i(\boldsymbol{Y}_i|S_i,\theta_i,\Psi_Y) + \log p_i(T_i|S_i,\theta_i,\Psi_T) + \log p(\theta_i|\Psi_\theta) + \log p(S_i|\Psi_\phi) \\ & + \log p(L_i,R_i|S_i). \end{aligned}$$

In the E step, we calculate the expectation of the complete-data log-likelihood, conditional on the observed data, given the current estimates $\Psi^{(m)}$. Since each of these terms involves a disjoint set of parameters, to maximize the expected complete-data log likelihood, we can maximize each term's expectation separately. $p(L_i,R_i|S_i)$ is non-informative so we can ignore this term and maximize

$$\begin{aligned} Q(\Psi;\Psi^{(m)}) &= Q_Y(\Psi;\Psi^{(m)}) + Q_T(\Psi;\Psi^{(m)}) + Q_\theta(\Psi;\Psi^{(m)}) + Q_\phi(\Psi;\Psi^{(m)}) \\ &= \sum_{i=1}^{n} E^{(m)}_{\theta,S} \left\{ \log p_i(\boldsymbol{Y}_i|S_i,\theta_i,\Psi_Y) + \log p_i(T_i|S_i,\theta_i,\Psi_T) + \log p(\theta_i|\Psi_\theta) + \log p(S_i|\Psi_\phi) \right\} \\ &= \sum_{i=1}^{n} \sum_{s_l \in \mathbb{S} \cap (L_i,R_i]} E^{(m)}_\theta \left(-\sum_{j=1}^{n_i} \left[\frac{1}{2} \log 2\pi \exp\left\{ W_i^T(t_{ij},S_i)\tau + \omega_i \right\} \right.\right. \\ &\quad \left.\left. + \frac{\left\{ Y_i(t_{ij}) - X_i^{(1)T}(t_{ij},S_i)\beta - Z_i^T(t_{ij})b_i \right\}^2}{2\exp\left\{ W_i^T(t_{ij},S_i)\tau + \omega_i \right\}} \right] \right. \\ &\quad + \delta_i \left\{ \log \Delta\Lambda_0(T_i) + X_i^{(2)T}(S_i)\gamma + W_i^T(\theta_i|T_i)\alpha \right\} \\ &\quad - \int_0^{T_i} \exp\left\{ X_i^{(2)T}(S_i)\gamma + W_i^T(\theta_i|s)\alpha \right\} d\Lambda_0(s) \\ &\quad \left. - \left\{ \frac{q+1}{2} \log 2\pi + \frac{1}{2} \log|\Sigma_\theta| + \frac{\theta_i^T \Sigma_\theta^{-1} \theta_i}{2} \right\} \right) p_i^{(m)}(S_i = s_l|\boldsymbol{Y}_i,T_i,L_i,R_i,\Psi^{(m)}) \\ &\quad + \sum_{i=1}^{n} \sum_{s_l \in \mathbb{S} \cap (L_i,R_i]} \log \left\{ \exp(-\sum_{s_u < s_l} \phi_{s_u}) - \exp(-\sum_{s_u \le s_l} \phi_{s_u}) \right\} \\ &\quad \times p_i^{(m)}(S_i = s_l|\boldsymbol{Y}_i,T_i,L_i,R_i,\Psi^{(m)}). \end{aligned} \tag{4.6}$$

Use $\boldsymbol{p}_i^{(m)}(s_l)$ to denote $p_i^{(m)}(S_i = s_l|\boldsymbol{Y}_i,T_i,L_i,R_i,\Psi^{(m)})$ for brevity. $\boldsymbol{p}_i^{(m)}(s_l)$ is calculated by

$$\begin{aligned} \boldsymbol{p}_i^{(m)}(s_l) &= \int_{\theta_i} p_i^{(m)}(\theta_i, S_i = s_l|\boldsymbol{Y}_i,T_i,L_i,R_i,\Psi^{(m)}) d\theta_i \\ &= \frac{\int_{\theta_i} p_i(\boldsymbol{Y}_i|S_i,\theta_i,\Psi_Y^{(m)}) p_i(T_i|S_i,\theta_i,\Psi_T^{(m)}) p(\theta_i|\Psi_\theta^{(m)}) p_i^{(m)}(S_i=s_l|L_i,R_i,\Psi_\phi^{(m)}) d\theta_i}{\int_{\theta_i} \sum_{s_l \in \mathbb{S}} p_i(\tilde{\boldsymbol{Y}}_i|S_i,\theta_i,\Psi_Y^{(m)}) p_i(T_i|S_i,\theta_i,\Psi_T^{(m)}) p(\theta_i|\Psi_\theta^{(m)}) p_i^{(m)}(S_i=s_l|L_i,R_i,\Psi_\phi^{(m)}) d\theta_i}, \end{aligned} \tag{4.7}$$

where

$$p_i^{(m)}(S_i = s_l | L_i, R_i, \Psi_\phi^{(m)}) \;=\; \mathbf{1}\{s_l \in (L_i, R_i]\} \frac{\exp\left(-\sum_{L_i < s_u < s_l} \phi_{s_u}\right)\{1 - \exp(-\phi_{s_l})\}}{1 - \exp\left(-\sum_{L_i < s_u \le R_i} \phi_{s_u}\right)} \tag{4.8}$$

The detailed derivation of equation (4.8) is provided in Section C.1 of the supplementary materials.

In the M-step, we find the update of Ψ by maximizing equation (4.6):

$$\Psi^{(m+1)} = \arg\max_{\Psi} Q(\Psi; \Psi^{(m)}), \tag{4.9}$$

until the algorithm converges, where $\Psi^{(m)}$ is the estimate of Ψ from the m-th iteration. It is easy to see that the M-step (4.9) requires the following expected values across all subjects that are calculated in the E-step (4.6): $E_{\theta,S}^{(m)}\{\exp(-\omega_i)\}$, $E\{b_i \exp(-\omega_i)\}$, $E_{\theta,S}^{(m)}\{b_i b_i^T \exp(-\omega_i)\}$, $E_{\theta,S}^{(m)}\left[\exp\left\{W_i(\theta_i|t)^T\alpha\right\}\right]$, $E_{\theta,S}^{(m)}\left[W_i(\theta_i|t)\exp\left\{W_i(\theta_i|t)^T\alpha\right\}\right]$, $E_{\theta,S}^{(m)}\left[W_i(\theta_i|t)W_i(\theta_i|t)^T\exp\left\{W_i(\theta_i|t)^T\alpha\right\}\right]$, $E_{\theta,S}^{(m)}(\theta_i\theta_i^T)$, where

$$\begin{aligned}
E_{\theta,S}^{(m)}\{h(\theta_i|t)\} &= \sum_{s_l \in \mathbb{S}} E_\theta^{(m)} h(\theta_i|t) \boldsymbol{p}_i^{(m)}(s_l) \\
&= \sum_{s_l \in \mathbb{S}} \int_{\theta_i} h(\theta_i|t) p_i^{(m)}(\theta_i, S_i = s_l | \boldsymbol{Y}_i, T_i, L_i, R_i, \Psi^{(m)}) d\theta_i \\
&= \int_{\theta_i} h(\theta_i|t) \sum_{s_l \in \mathbb{S}} p_i^{(m)}(\theta_i, S_i = s_l | \boldsymbol{Y}_i, T_i, L_i, R_i, \Psi^{(m)}) d\theta_i \\
&= \frac{\int_{\theta_i} h(\theta_i|t) \sum_{s_l \in \mathbb{S}} p_i(\boldsymbol{Y}_i|S_i,\theta_i,\Psi_Y^{(m)}) p_i(T_i|S_i,\theta_i,\Psi_T^{(m)}) p(\theta_i|\Psi_\theta^{(m)}) p_i^{(m)}(S_i = s_l|L_i,R_i,\Psi_\phi^{(m)}) d\theta_i}{\int_{\theta_i} \sum_{s_l \in \mathbb{S}} p_i(\tilde{\boldsymbol{Y}}_i|S_i,\theta_i,\Psi_Y^{(m)}) p_i(T_i|S_i,\theta_i,\Psi_T^{(m)}) p(\theta_i|\Psi_\theta^{(m)}) p_i^{(m)}(S_i = s_l|L_i,R_i,\Psi_\phi^{(m)}) d\theta_i},
\end{aligned}$$

for any function $h(.)$. Furthermore, it can be shown that the M-step (4.9) has closed form solutions for the parameters β, Σ_θ, $\{\phi_{s_l}\}$ and $\Lambda_0(.)$. The other parameters τ, γ, α can be updated using the one-step Newton-Raphson method. Details are provided in equations (C.2) - (C.11) of Section C.2 of the supplementary materials.

Remark 2. *It is worth noting that in the EM algorithm, the instantaneous baseline hazard function $\lambda_0(.)$ has positive mass only at the jump points $t_{i,s_l}, i = 1, \ldots, n, s_l \in \mathbb{S}$, i.e.,*

$$\lambda_0^{(m)}(t) = \begin{cases} \Delta\Lambda_0^{(m)}(t), & t \in \{t_{i,s_l}\} \\ 0, & elsewhere. \end{cases}$$

We observe that the non-continuity of $\lambda_0(.)$ *will lead* $\psi^{(m)}$ *to a saddlepoint because* $\Delta\Lambda_0^{(m)}(.)$ *will disturb the update of* $\boldsymbol{p}_i^{(m)}(s_l)$ *in equation (4.7), which will result in a severe bias issue. To mitigate this bias issue in the EM steps, we replace* $\Delta\Lambda_0^{(m)}(.)$ *with the following quantity to enhance the smoothness of the hazard function* $\lambda_0(.)$*:*

$$\Delta\Lambda_0^{(m)*}(t) = \int \frac{1}{b} K(\frac{t-s}{b}) d\Delta\Lambda_0^{(m)}(s), \tag{4.10}$$

is a kernel estimate of the instantaneous hazard function of $\lambda_0^{(m)}(.)$*. In equation (4.10),* b *is the smoothing bandwidth,* K *is a nonnegative kernel function satisfying certain properties. For example, one may use an Epanechnikov kernel defined as* $K(x) = 0.75(1-x^2), |x| \leq 1$ *and the global bandwidth* $b = c * max\{t_{i,s_l}\}/(8n_v^{1/5})$ *with* c *the pre-determined constant and* n_v *the total number of jump points in* $\Lambda_0(.)$*. Altstein and Li [2013] discussed the choice of* c *in the sensitivity analysis and found that when* $c = 0.95$*, the estimation procedure would yield good estimation accuracy in the most of the cases. We adopted this specification in Section 4.2 and found that the performance of our proposed method is also satisfying in all our simulations, which is consistent with the conclusions found in [Altstein and Li, 2013].*

4.1.3 Standard error estimation

As discussed in Elashoff et al. [2016] (Section 4.1, p.72), standard errors of the parametric components of the SMLE can be estimated by profiled likelihood, observed information matrix, or bootstrap. All three methods can be computationally intensive when n is large. Here we focus on the profiled likelihood-based method, which can be linearized with n [Li et al., 2022a].

Let $\Omega = (\beta, \tau, \text{vech}(\Sigma_\theta), \gamma, \alpha, \{\phi_{s_l}\})$ denote the parametric component of Ψ and $\widehat{\Omega}$ its maximum likelihood estimate and $\boldsymbol{D} = \{\boldsymbol{D}_1, \ldots, \boldsymbol{D}_n\}$ the observed data for all the subjects. The variance-covariance matrix of $\widehat{\Omega}$ can be estimated by inverting the following approximate

empirical Fisher information [Lin et al., 2004, Zeng et al., 2005, Zeng and Cai, 2005]:

$$\sum_{i=1}^{n} \nabla_{\Omega} l^{(i)}(\hat{\Omega}; \boldsymbol{D}) \nabla_{\Omega} l^{(i)}(\hat{\Omega}; \boldsymbol{D})^T, \tag{4.11}$$

where $\nabla_{\Omega} l^{(i)}(\Omega; \boldsymbol{D})$ is the observed score vector from the profiled likelihood $l^{(i)}(\Omega; \boldsymbol{D})$ of Ω on the ith subject by profiling out the baseline hazards. Details of the observed score vector for each parametric component are provided in equations (C.12)-(C.17) in Section C.3 of the supplementary materials.

4.2 Simulation studies

We present extensive simulations to examine the finite sample performance of the proposed joint model and the classical joint model with midpoint imputation. We also demonstrate that midpoint imputation can lead to biased parameter estimates, biased standard error estimates and invalid inferences.

We began with the cohort size $n = 600$ and $5,000$ to examine the large sample properties. First of all, the initial event time $S_i \sim \text{uni}(0.1, 5)$ was observed to be interval-censored through a series of non-informative finite inspection times [Lawless and Babineau, 2006]

$$U = \{U_{ij} : i = 1, \ldots, n, j = 1, \ldots, K_i \text{ with } U_{i0} = 0, U_{ij} < U_{i(j+1)}\},$$

with $U_{i(j+1)} = U_{ij} + \text{uni}(0.1, 1)$. For each subject i, there exists a smallest censoring interval $(U_{ij}, U_{i(j+1)}]$ that contains S_i, and thus $L_i = U_{ij}$ and $R_i = U_{i(j+1)}$ was observed.

Next, longitudinal measurements $Y_i(t_{ij})$ were generated from a *mixed-effects location scale model* (4.1) and (4.2) with

$$m_i(t_{ij}) = \beta_0 + \beta_1 S_i + \beta_2 t_{ij} + \beta_3 X_{1i} + \beta_4 X_{2i} + b_i, \tag{4.12}$$

$$\sigma_i^2(t_{ij}) = \exp(\tau_0 + \tau_1 S_i + \tau_2 t_{ij} + \tau_3 X_{1i} + \tau_4 X_{2i} + \omega_i), \tag{4.13}$$

where $t_{ij} = \tilde{O}_{ij} - S_i$ is a scheduled follow-up visit time after the initial event, the first scheduled follow-up visit time $\tilde{O}_{i1}$ starts at R_i, $X_{1i} \sim \text{uni}(-1, 1)$, and $X_{2i} \sim N(1, 4)$. Therefore,

we observed $\boldsymbol{Y}_i(t_{ij})$ measured at $\tilde{O}_{ij}$. The random effects $\theta_i = (b_i, \omega_i)^T$ follows $N_2(0, \Sigma_\theta)$, where $\sigma_b^2 = 1, \sigma_\omega^2 = 0.5$, and $\sigma_{bw} = 0.1$. The mean trajectory $m_i(t_{ij})$ and WS variance $\sigma_i^2(t_{ij})$, modeled by equations (4.12) - (4.13), include fixed effects induced by the interval-censored covariates S_i, t_{ij}, and two baseline covariates X_{i1}, X_{i2}, and a subject level random effect b_i, ω_i, respectively.

Lastly, the target event process was generated from a Cox proportional hazards model:

$$\lambda_i(t) \quad = \quad \lambda_0(t) \exp\{\gamma_1 S_i + \gamma_2 X_{1i} + \gamma_3 X_{2i} + \alpha_b b_i + \alpha_\omega \omega_i\}, \tag{4.14}$$

where the baseline hazards $\lambda_0(t)$ is held constant at 0.3 for $t \geq 1$ and 0 otherwise to assure that the target event can be always observed after the first positive diagnosis of the initial event. We simulated the non-informative censoring time $C_i \sim \text{uni}(5, 10)$ and let $T_i = min\{\tilde{T}_i, C_i\}$ be the survival time after the initial event occurs (possibly censored) for subject i, where $\tilde{T}_i$ are independent event times from equation (4.14), $i = 1, \ldots, n$. The longitudinal measurements for subject i at t_{ij} are assumed missing after T_i. Finally, we observed the survival time $E_i = T_i + S_i$ in a total time scale. Overall, we obtained an event rate at about 75% and the average number of longitudinal measurements per subject is about 10.

We fitted both our proposed joint model (4.1) - (4.3) (Model 1) that accounts for interval-censored covariates as missing data with a well developed R package "iCenJMH" and a classical joint model (Model 2) using midpoint imputation [Li et al., 2023]. Table 4.1 - 4.4 summarize simulated results on the bias, sample standard deviations of the parameter estimates (SE), average estimated standard errors of the parameter estimates (Est. SE), and coverage probabilities of 95% confidence intervals (CP) between the random effects under different scenarios. Each entry in the table is based on 300 Monte Carlo samples.

Table 4.1 shows the performance of both models under the sample size of $n = 600$ and the 100% interval censoring rate of the initial event. Compared to midpoint imputation (Model 2), our proposed joint model method (Model 1) demonstrates a smaller bias for all parameters, a larger standard error estimates, and CPs reach the 95% nominal level. On

the other hand, Model 2 ignores the uncertainty of interval-censored covariates as missing data induced severe bias in some parameter and lower standard error estimates, leading to significant under-coverage of the associated confidence intervals. For example, the parameter estimate and standard error estimate of β_0, α_b are both substantially biased, and its associated confidence interval coverage probability (18% and 0%) is unreasonably low compared to the 95% nominal level. When the interval censoring rate of the initial event reduces to 30% (See Table 4.2), the performance of Model 1 behaves similar as shown in Table 4.1 but Model 2 still encounters undercoverage issue due to non-ignorable estimation bias and underestimated standard errors.

We further investigate the large sample theory of Model 1 and Model 2 by increasing the sample size to be 5,000 and results are summarised in Table 4.3 - 4.4. We find that the performance of Model 1 satisfies the large sample theory such that the CPs are similar to 95% nominal level. In contrast, the performance of Model 2 worsens under the simulation settings with both high and low interval-censoring rates of the initial event, which confirms the conclusion we draw from Table 4.1 - 4.2.

4.3 An application: UK-Biobank data

The UK-Biobank (UKB) is a prospective cohort study with both genetic and phenotypic data collected on approximately 500,000 individuals, aged 37-73 years, from the general population between 2006 and 2010 in the United Kingdom [Collins, 2012, Sudlow et al., 2015]. Record linkage to Health Episode Statistics (England), Patient Episode Database for Wales, and the Scottish Morbidity Records (Scotland) was used to identify the date and cause of hospital admissions. Hospital admission records were available until February 2018 for the full UKB cohort (noted as "UKB data"), whereas linkage to primary care records was available for 45% of the UKB cohort until the end of 2017 (noted as "UKB Primary Care data"). The detailed linkage procedures relating to primary care records are available online

Table 4.1: Comparison of the bias, standard error (SE), estimated standard error (Est. SE), and coverage probability (CP) between the proposed joint model (Model 1) and a classical joint model using midpoint imputation (Model 2) for the interval-censored initial event time. Abbreviation: IC = interval-censored; RC = right-censored. ($n = 600$, IC rate for S_i=100%, RC rate for T_i=75%)

		Model 1 (proposed model)				Model 2 (midpoint imputation)			
Covariates	True	Bias	SE	Est. SE	CP (%)	Bias	SE	Est. SE	CP (%)
Longitudinal									
Mean trajectory									
Intercept	5	-0.028	0.131	0.159	99.3	0.379	0.173	0.110	**18.0**
S_i (Interval-censored)	2	0.010	0.047	0.052	97.0	-0.057	0.063	0.034	**56.3**
t_{ij} (Interval-censored)	-3	-0.001	0.013	0.015	97.7	<0.001	0.013	0.014	95.3
X_{1i}	-3	-0.003	0.104	0.116	97.0	-0.007	0.140	0.084	**74.7**
X_{2i}	2	-0.002	0.033	0.033	96.0	-0.010	0.042	0.022	**68.0**
WS variability									
Intercept	1	-0.005	0.075	0.085	97.0	-0.016	0.073	0.079	96.0
S_i (Interval-censored)	0.05	0.001	0.025	0.027	96.7	0.003	0.025	0.025	96.3
t_{ij} (Interval-censored)	0.2	<0.001	0.009	0.010	98.7	<0.001	0.009	0.010	98.0
X_{1i}	0.1	-0.002	0.063	0.065	97.7	-0.003	0.064	0.063	97.0
X_{2i}	-0.2	-0.001	0.018	0.019	95.3	<0.001	0.018	0.018	95.3
Survival									
S_i (Interval-censored)	-0.05	<0.001	0.039	0.050	99.7	0.009	0.036	0.036	93.0
X_{1i}	0.2	0.008	0.102	0.102	95.0	-0.013	0.094	0.088	92.3
X_{2i}	-0.1	-0.003	0.028	0.030	96.7	0.005	0.026	0.026	93.3
Association									
b_i	0.5	0.029	0.109	0.116	97.0	-0.311	0.041	0.041	**0.0**
ω_i	-0.5	-0.012	0.115	0.124	96.7	0.116	0.091	0.093	**78.0**
Covariance matrix of random effects									
σ_b^2	1	-0.032	0.118	0.130	95.0	1.105	0.146	0.149	**0.0**
σ_ω^2	0.5	-0.004	0.045	0.046	95.0	-0.005	0.044	0.044	94.0
$\sigma_{b\omega}$	0.1	0.002	0.061	0.060	95.3	-0.031	0.059	0.058	93.0

Note: Large error in confidence interval coverage probability (CP) compared to the 95% nominal level are highlighted in boldface. Each entry is based on 300 Monte Carlo samples.

Table 4.2: Comparison of the bias, standard error (SE), estimated standard error (Est. SE), and coverage probability (CP) between the proposed joint model (Model 1) and a classical joint model using midpoint imputation (Model 2) for the interval-censored initial event time. Abbreviation: IC = interval-censored; RC = right-censored. ($n = 600$, IC rate for S_i=30%, RC rate for T_i=75%)

		Model 1 (proposed model)				Model 2 (midpoint imputation)			
Covariates	True	Bias	SE	Est. SE	CP (%)	Bias	SE	Est. SE	CP (%)
Longitudinal									
Mean trajectory									
Intercept	5	-0.006	0.099	0.114	98.3	0.108	0.116	0.106	**79.3**
S_i (Interval-censored)	2	<0.001	0.036	0.037	94.7	-0.019	0.043	0.034	**83.7**
t_{ij} (Interval-censored)	-3	<0.001	0.012	0.014	97.7	0.001	0.012	0.013	97.7
X_{1i}	-3	0.003	0.085	0.090	96.3	0.002	0.096	0.083	**89.3**
X_{2i}	2	0.001	0.028	0.025	94.7	-0.002	0.035	0.022	**82.0**
WS variability									
Intercept	1	-0.002	0.073	0.082	96.7	-0.003	0.072	0.078	96.3
S_i (Interval-censored)	0.05	<0.001	0.023	0.026	97.3	0.001	0.023	0.025	96.0
t_{ij} (Interval-censored)	0.2	<0.001	0.010	0.010	96.3	<0.001	0.010	0.009	94.3
X_{1i}	0.1	0.001	0.064	0.064	96.7	0.001	0.063	0.062	96.0
X_{2i}	-0.2	-0.001	0.018	0.019	95.0	-0.001	0.018	0.018	94.7
Survival									
S_i (Interval-censored)	-0.05	<0.001	0.037	0.041	96.3	0.004	0.036	0.037	94.7
X_{1i}	0.2	0.005	0.102	0.099	94.3	-0.004	0.100	0.091	91.0
X_{2i}	-0.1	-0.002	0.027	0.029	96.3	0.001	0.027	0.027	94.7
Association									
b_i	0.5	0.010	0.080	0.079	95.7	-0.153	0.059	0.053	**24.0**
ω_i	-0.5	-0.005	0.096	0.104	97.7	0.057	0.088	0.094	**88.3**
Covariance matrix of random effects									
σ_b^2	1	-0.007	0.094	0.092	94.0	0.287	0.112	0.096	**17.0**
σ_ω^2	0.5	-0.003	0.042	0.045	97.0	-0.004	0.042	0.043	96.7
$\sigma_{b\omega}$	0.1	0.003	0.044	0.048	95.7	-0.005	0.047	0.048	93.3

Note: Large error in confidence interval coverage probability (CP) compared to the 95% nominal level are highlighted in boldface. Each entry is based on 300 Monte Carlo samples.

Table 4.3: Comparison of the bias, standard error (SE), estimated standard error (Est. SE), and coverage probability (CP) between the proposed joint model (Model 1) and a classical joint model using midpoint imputation (Model 2) for the interval-censored initial event time. Abbreviation: IC = interval-censored; RC = right-censored. ($n = 5,000$, IC rate for S_i=100%, RC rate for T_i=75%)

		Model 1 (proposed model)				Model 2 (midpoint imputation)			
Covariates	True	Bias	SE	Est. SE	CP (%)	Bias	SE	Est. SE	CP (%)
Longitudinal									
Mean trajectory									
Intercept	5	-0.004	0.045	0.051	95.3	0.387	0.061	0.036	**0.0**
S_i (Interval-censored)	2	0.001	0.014	0.017	98.0	-0.062	0.022	0.011	**3.3**
t_{ij} (Interval-censored)	-3	<0.001	0.005	0.005	94.0	0.001	0.005	0.005	94.0
X_{1i}	-3	-0.004	0.037	0.039	94.7	0.001	0.053	0.027	**70.7**
X_{2i}	2	<0.001	0.011	0.011	92.0	-0.004	0.019	0.007	**51.7**
WS variability									
Intercept	1	0.002	0.027	0.027	97.0	-0.007	0.026	0.027	94.7
S_i (Interval-censored)	0.05	-0.001	0.008	0.009	97.3	0.001	0.008	0.009	96.0
t_{ij} (Interval-censored)	0.2	<0.001	0.004	0.003	93.0	<0.001	0.004	0.003	94.0
X_{1i}	0.1	-0.001	0.021	0.021	96.0	-0.002	0.021	0.021	95.3
X_{2i}	-0.2	-0.001	0.006	0.006	95.3	<0.001	0.006	0.006	95.7
Survival									
S_i (Interval-censored)	-0.05	<0.001	0.014	0.014	95.0	0.007	0.013	0.012	**88.3**
X_{1i}	0.2	0.001	0.034	0.033	93.7	-0.016	0.031	0.030	91.0
X_{2i}	-0.1	-0.001	0.010	0.010	92.0	0.007	0.010	0.009	**84.7**
Association									
b_i	0.5	0.002	0.036	0.035	93.3	-0.312	0.014	0.014	**0**
ω_i	-0.5	-0.002	0.037	0.038	95.7	0.111	0.030	0.031	**4.3**
Covariance matrix of random effects									
σ_b^2	1	-0.011	0.044	0.043	92.7	1.006	0.050	0.051	**0.0**
σ_ω^2	0.5	-0.002	0.014	0.015	96.0	-0.006	0.014	0.015	94.7
$\sigma_{b\omega}$	0.1	<0.001	0.020	0.020	96.7	-0.031	0.021	0.020	**61.7**

Note: Large error in confidence interval coverage probability (CP) compared to the 95% nominal level are highlighted in boldface. Each entry is based on 300 Monte Carlo samples.

Table 4.4: Comparison of the bias, standard error (SE), estimated standard error (Est. SE), and coverage probability (CP) between the proposed joint model (Model 1) and a classical joint model using midpoint imputation (Model 2) for the interval-censored initial event time. Abbreviation: IC = interval-censored; RC = right-censored. ($n = 5,000$, IC rate for S_i=30%, RC rate for T_i=75%)

		Model 1 (proposed model)				Model 2 (midpoint imputation)			
Covariates	True	Bias	SE	Est. SE	CP (%)	Bias	SE	Est. SE	CP (%)
Longitudinal									
Mean trajectory									
Intercept	5	-0.001	0.042	0.037	92.7	0.117	0.045	0.035	**12.3**
S_i (Interval-censored)	2	<0.001	0.014	0.012	92.0	-0.021	0.015	0.011	**50.3**
t_{ij} (Interval-censored)	-3	0.001	0.004	0.004	94.7	0.001	0.004	0.004	94.7
X_{1i}	-3	-0.003	0.032	0.029	92.0	-0.003	0.036	0.027	**86.3**
X_{2i}	2	<0.001	0.010	0.008	95.0	<0.001	0.013	0.007	**86.0**
WS variability									
Intercept	1	0.002	0.025	0.027	96.3	<0.001	0.025	0.026	96.3
S_i (Interval-censored)	0.05	-0.001	0.008	0.009	96.0	-0.001	0.008	0.009	96.3
t_{ij} (Interval-censored)	0.2	<0.001	0.003	0.003	95.3	<0.001	0.003	0.003	94.7
X_{1i}	0.1	-0.002	0.020	0.021	95.7	-0.002	0.020	0.021	95.7
X_{2i}	-0.2	<0.001	0.006	0.006	94.0	<0.001	0.006	0.006	94.7
Survival									
S_i (Interval-censored)	-0.05	-0.001	0.014	0.013	94.0	0.003	0.013	0.013	94.3
X_{1i}	0.2	0.001	0.034	0.032	92.0	-0.008	0.033	0.031	90.3
X_{2i}	-0.1	-0.001	0.010	0.009	92.0	0.003	0.010	0.009	90.7
Association									
b_i	0.5	0.001	0.025	0.025	94.3	-0.161	0.020	0.018	**0**
ω_i	-0.5	-0.005	0.034	0.033	95.0	0.053	0.032	0.031	**56.7**
Covariance matrix of random effects									
σ_b^2	1	-0.001	0.028	0.030	96.0	0.296	0.034	0.033	**0**
σ_ω^2	0.5	-0.002	0.014	0.014	95.3	-0.003	0.014	0.014	95.3
$\sigma_{b\omega}$	0.1	<0.001	0.015	0.016	94.7	-0.006	0.016	0.016	93.0

Note: Large error in confidence interval coverage probability (CP) compared to the 95% nominal level are highlighted in boldface. Each entry is based on 300 Monte Carlo samples.

[https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/primary_care_data.pdf].

In this illustrating example, we identified 42,785 participants, who were free from cardiovascular diseases (CVD) at baseline and diagnosed with both type 1 and type 2 diabetes but known to be observed at two consecutive clinical visits. The primary event outcome is a CVD composite outcome, reflecting myocardial infarction (MI), unstable angina (UA), ischemic stroke (IS), and percutaneous coronary intervention (PCI) [Kim et al., 2023]. During the study period, CVD-related biomarkers were measured during the follow-up visits, such as systolic and diastolic blood pressure (SBP and DBP, mmHg), High-density and low-density lipoprotein cholesterol (HDL and LDL, mmol/L), and triglycerides (mmol/L). As an illustration, we considered a joint model of SBP and CVD event outcome in the presence of the interval-censored age-onset diagnosis of type 2 diabetes and analyzed the association between the two outcomes after type 2 diabetes diagnosis. We adjusted for the selected baseline variables including self-reported blood pressure lowering medication use at the assessment center (BP drug, 1=yes, 0=no), body mass index (BMI, kg/m^2), smoking status (1=smoker, 0=non-smoker), International Standard Classification of Education (ISCED), race/ethnicity (1=non-British, 0=British), and sex (1=female, 0=male) in both SBP and CVD time-to-event sub-models.

After excluding the patients without both longitudinal and event information, a cohort of 11,821 participants with an average of 8.2 repeated measures and a total of 97,317 SBP measures are kept for this analysis. Among them, 1,694 participants (14.3%) experienced CVD events and 10,127 (85.7%) were right-censored during the study period.

In our proposed joint model (Model 1), the longitudinal SBP ($\boldsymbol{Y}_i(t_{ij})$) is assumed to follow the mixed-effects multiple location scale model (4.1)-(4.2) as

$$\begin{aligned} m_i(t_{ij}) &= \beta_0 + \beta_1 S_i + \beta_2 t_{ij} + \beta_3 \text{BP durg}_i + \beta_4 \text{BP durg}_i \times t_{ij} + \\ &\quad \beta_5 \text{BMI}_i + \beta_6 \text{Smoking}_i + \beta_7 \text{ISCED}_i + \beta_8 \text{Race}_i + \beta_9 \text{Sex}_i + b_i, \qquad (4.15) \\ \sigma_i^2(t_{ij}) &= \exp(\tau_0 + \tau_1 S_i + \tau_2 t_{ij} + \tau_3 \text{BP durg}_i + \tau_4 \text{BP durg}_i \times t_{ij} + \end{aligned}$$

$$\tau_5 \text{BMI}_i + \tau_6 \text{Smoking}_i + \tau_7 \text{ISCED}_i + \tau_8 \text{Race}_i + \tau_9 \text{Sex}_i + \omega_i), \qquad (4.16)$$

and the CVD event outcome is modeled by the following Cox proportional hazards model:

$$\begin{aligned} \lambda_i(t) &= \lambda_0(t) \exp(\gamma_1 S_i + \gamma_2 \text{BP durg}_i + \gamma_3 \text{BMI}_i + \gamma_4 \text{Smoking}_i + \gamma_5 \text{ISCED}_i + \\ &\quad \gamma_6 \text{Race}_i + \gamma_7 \text{Sex}_i + \alpha_b b_i + \alpha_\omega \omega_i), \end{aligned} \qquad (4.17)$$

where the shared random effects $\theta_i = (b_i, \omega_i)^T$ follow bivariate normal distribution with mean zero and variance-covariance matrix

$$\Sigma_\theta = \begin{pmatrix} \sigma_b^2 & \sigma_{b\omega} \\ \sigma_{b\omega} & \sigma_\omega^2 \end{pmatrix}.$$

The results of joint analysis of SBP and time-to-CVD are summarised in Table 4.5. Both Model 1 and Model 2 show that the heterogeneity of SBP WS variability is substantial ($\sigma_\omega^2 = 0.27, \text{SE} = 0.01$, p-value$< 0.001$) across UKB participants and suggest that standardized SBP WS variability (ω_i/σ_ω) is highly predictive of CVD event. We notice that the individual mean of SBP is protective of CVD and one possible reason is the non-ignorable correlation between SBP WS variability (ω_i) and individual mean (b_i). Both models also suggest that those who took the blood pressure lowering medication at the assessment center had around 8mmHg higher SBP at the baseline but the mean trajectory significantly declined over time, compared to those never self-reported their medication use. One interesting finding is that Model 1 shows that age at diagnosis of diabetes is positively associated with the mean SBP, however, Model 2 does not detect any significant signal of age at diagnosis of diabetes. In general, the results from Model 1 would be more clinically meaningful since one's physical condition makes it harder to regulate the blood pressure as aging. In the survival sub-model, Model 1 suggests that a smoker had an 12% increased risk of CVD event than a non-smoker with a p-value of 0.05, while Model 2 only shows a marginal relationship. In summary, the results demonstrate that midpoint imputation can yield biased results and invalid inference and thus lead to problematic conclusion about the relationship between outcomes and covariates.

Table 4.5: Joint analysis of systolic blood pressure (SBP, mmHg) and time to CVD using the UK-Biobank data (Model 1 is the proposed joint model (4.15)-(4.17) and Model 2 is the classical joint model using midpoint imputation. Abbreviations: SE=standard error; HR=hazard ratio; CI=confidence interval.)

	Model 1		Model 2	
Longitudinal outcome	Mean trajectory	WS variability	Mean trajectory	WS variability
(Systolic blood pressure (SBP, mmHg))	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)
Intercept	125.79 (0.62)***	4.76 (0.08)***	129.76 (0.57)***	4.91 (0.08)***
Interval-censored covariates				
Age at diagnosis of diabetes	**0.06 (0.01)*** **	-0.01 (<0.01)***	**<0.01 (0.01)**	-0.01 (<0.01)***
Time (Years from diagnosis of diabetes)	0.09 (0.01)***	<0.01 (<0.01)	0.08 (0.01)***	<0.01 (<0.01)
BP drug × Time	-0.50 (0.01)***	-0.02 (<0.01)***	-0.46 (0.01)***	-0.01 (<0.01)***
Baseline covariates				
BP drug	8.33 (0.15)***	0.26 (0.02)***	8.31 (0.15)***	0.26 (0.02)***
BMI	0.13 (0.01)***	<0.01 (<0.01)	0.11 (0.01)***	<0.01 (<0.01)
smoking status	0.12 (0.13)	0.05 (0.02)***	0.15 (0.12)	0.05 (0.02)***
ISCED	-0.26 (0.04)***	-0.02 (<0.01)***	-0.30 (0.04)***	-0.02 (<0.01)***
Race (Non-British/British)	-0.44 (0.17)***	0.02 (0.02)	-0.82 (0.16)***	0.01 (0.02)
Sex (Female/Male)	-1.08 (0.13)***	0.03 (0.02)	-1.19 (0.13)***	0.03 (0.02)
Random effects				
(variance-covariance matrix)	Estimate (SE)		Estimate (SE)	
σ_b^2	72.52 (1.28)***		74.06 (1.30)***	
$\sigma_{b\omega}$	1.77 (0.07)***		1.81 (0.07)***	
σ_ω^2	0.27 (0.01)***		0.27 (0.01)***	
Target event outcome (CVD)	HR (95% CI)		HR (95% CI)	
Interval-censored covariates				
Age at diagnosis of diabetes	1.15 (1.14-1.16)***		1.14 (1.13-1.14)***	
Baseline covariates				
BP drug	1.09 (0.97-1.23)		1.07 (0.95-1.20)	
BMI	1.04 (1.03-1.05)***		1.03 (1.02-1.04)***	
smoking status	**1.12 (1.00-1.24)***		**1.10 (1.00-1.23)**	
ISCED	0.96 (0.93-0.99)**		0.96 (0.93-0.99)***	
Race (Non-British/British)	1.28 (1.12-1.46)***		1.26 (1.10-1.43)***	
Sex (Female/Male)	0.60 (0.53-0.68)***		0.61 (0.54-0.69)***	
Random effects				
Mean trajectory (α_b)	0.97 (0.97-0.98)***		0.97 (0.96-0.98)***	
WS variability (α_ω)	1.27^+ (1.16-1.39)***		1.26^+ (1.15-1.38)***	

* p-value<0.05; ** p-value<0.01; *** p-value<0.001.

$^+$ Standardized HR of association parameter for WS variability is reported to show the effect on the risk of CVD by 1 SD change of WS variability of SBP.

4.4 Discussion

We have developed a novel joint model of longitudinal and time-to-event outcomes that handles the interval-censored covariates as missing data due to the unknown initial event time. The non-parametric estimator of unknown initial event times is originally proposed by Morrison et al. [2021], in which a discrete, finite support is well defined, and integrating the complete-data log likelihood becomes a summation over the designated support and thus relaxes the distributional assumption on the hazard of the initial event time. We develop a customized EM algorithm under the joint modeling framework and demonstrate its estimation advantage via extensive simulations, comparing with existing approaches such as midpoint imputation, which ignores the uncertainty of interval-censored covariates and thus leads to severe bias and undercoverage. Our analysis of UKB data reveals that imputation strategy can result in problematic conclusion on parameter estimation.

The utility of joint modeling of longitudinal and time-to-event outcomes is not limited to the simple setting considered in this paper. It could be possibly extended to many other applications, such as joint multi-state model, joint modeling of recurrent events, or when the target event is also interval-censored. Extensions to more sophisticated joint models for multiple longitudinal biomarkers are also warranted in our future research.

APPENDIX A

Supplementary Materials for Project 1

A.1 M-step solutions for equation (2.4) in Section 2.1.1.2

It can be shown that the parameters β, σ^2, and Σ, as well as $\Lambda_{0k}(t)$ have closed-form solutions in the M-step (See equation (2.4)). Using $E^{(m)}$ to denote $E^{(m)}_{b_i|Y_i,C_i,\Psi^{(m)}}$, we have

$$\begin{aligned}
\beta^{(m+1)} &= \left\{\sum_{i=1}^{n}\sum_{j=1}^{n_i} X_i^{(1)}(t_{ij})X_i^{(1)}(t_{ij})^T\right\}^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{Y_{ij} - E^{(m)}(\tilde{X}_i{}^{(1)}(t_{ij})^T b_i)\right\} \\
&\quad \times X_i^{(1)}(t_{ij}), \qquad \text{(A.1)} \\
\sigma^{2(m+1)} &= \frac{1}{\sum_{i=1}^{n} n_i}\sum_{i=1}^{n}\sum_{j=1}^{n_i} E^{(m)}\left\{Y_{ij} - X_i^{(1)}(t_{ij})^T\beta^{(m+1)} - \tilde{X}_i{}^{(1)}(t_{ij})^T b_i\right\}^2, \qquad \text{(A.2)} \\
\Sigma^{(m+1)} &= \frac{1}{n}\sum_{i=1}^{n} E^{(m)}(b_i b_i^T), \qquad \text{(A.3)} \\
\Lambda_{0k}^{(m+1)}(t) &= \sum_{l:t_{kl}\le t}\frac{d_{kl}}{\sum_{r\in R(t_{kl})}\exp(X_r^{(2)T}\gamma_k^{(m)})E^{(m)}\left\{\exp(\nu_k^{(m)T}b_r)\right\}}, \qquad \text{(A.4)}
\end{aligned}$$

where $t_{k1} > ... > t_{kq_k}$ are the distinct uncensored failure times for risk k, $R(t_{kl})$ is the risk set at time t_{kl}, $l = 1, ..., q_k$, and d_{kl} is the number of type k failures, for $k = 1, \ldots, K$. It is clear from equation (A.4) that $\Lambda_{0k}^{(m+1)}(t)$ is a right-continuous and non-decreasing function.

The parameters γ and ν do not have closed-form solutions and we update them using the one-step Newton-Raphson method

$$\begin{aligned}
\gamma_k^{(m+1)} &= \gamma_k^{(m)} + I_{\gamma_k}^{(m)-1}S_{\gamma_k}^{(m)}, \quad k = 1, \ldots, K, \\
\nu_k^{(m+1)} &= \nu_k^{(m)} + I_{\nu_k}^{(m)-1}S_{\nu_k}^{(m)}, \quad k = 1, \ldots, K,
\end{aligned}$$

where

$$I^{(m)}_{\gamma_k} = \sum_{i=1}^{n}\sum_{t_{kj}\le T_i}\Delta\Lambda_{0k}(t_{kj})^{(m+1)}\exp(X_i^{(2)T}\gamma_k^{(m)})E^{(m)}\left\{\exp(\nu_k^{(m)T}b_i)\right\}X_i^{(2)}X_i^{(2)T}, \quad (A.5)$$

$$S^{(m)}_{\gamma_k} = \sum_{i=1}^{n}\Big[I(D_i=k)X_i^{(2)}(T_i) - \sum_{t_{kj}\le T_i}\Delta\Lambda_{0k}(t_{kj})^{(m+1)}\exp(X_i^{(2)T}\gamma_k^{(m)})E^{(m)}\left\{\exp(\nu_k^{(m)T}b_i)\right\}X_i^{(2)}\Big], \quad (A.6)$$

$$I^{(m)}_{\nu_k} = \sum_{i=1}^{n}\sum_{t_{kj}\le T_i}\Delta\Lambda_{0k}^{(m+1)}\exp(X_i^{(2)T}\gamma_k^{(m)})E^{(m)}\left\{b_ib_i^T\exp(\nu_k^{(m)T}b_i)\right\}, \quad (A.7)$$

$$S^{(m)}_{\nu_k} = \sum_{i=1}^{n}\Big[I(D_i=k)E^{(m)}(b_i) - \sum_{t_{kj}\le T_i}\Delta\Lambda_{0k}(t_{kj})^{(m+1)}\exp(X_i^{(2)T}\gamma_k^{(m)})E^{(m)}\left\{b_i\exp(\nu_k^{(m)T}b_i)\right\}\Big]. \quad (A.8)$$

A.2 The observed score vector in equation (2.6) in Section 2.1.1.3

The components of the observed score vector in equation (2.6) are defined by

$$\nabla_\beta l^{(i)}(\hat\Omega;Y,C) = \frac{1}{\sigma^2}\sum_{i=1}^{n_i}E\left\{Y_{ij}-X_i^{(1)}(t_{ij})^T\beta-\tilde X_i{}^{(1)}(t_{ij})^Tb_i\right\}X_i^{(1)}(t_{ij})\Big|_{\beta=\hat\beta,\sigma^2=\hat\sigma^2}, \quad (A.9)$$

$$\nabla_\Sigma l^{(i)}(\hat\Omega;Y,C) = \frac{1}{2}\left[2\Sigma^{-1}E(b_ib_i^T)\Sigma^{-1}-\left\{\Sigma^{-1}E(b_ib_i^T)\Sigma^{-1}\circ I\right\}-2\Sigma^{-1}+\Sigma^{-1}\circ I\right]\Big|_{\Sigma=\hat\Sigma}, \quad (A.10)$$

$$\nabla_{\sigma^2} l^{(i)}(\hat\Omega;Y,C) = \left[\frac{1}{2\sigma^4}\sum_{i=1}^{n_i}E\left\{Y_{ij}-X_i^{(1)}(t_{ij})^T\beta-\tilde X_i{}^{(1)}(t_{ij})^Tb_i\right\}^2-\frac{n_i}{2\sigma^2}\right]\Big|_{\beta=\hat\beta,\sigma^2=\hat\sigma^2} \quad (A.11)$$

$$\nabla_{\gamma_k} l^{(i)}(\hat\Omega;Y,C) = I(D_i=k)\left[X_i^{(2)}-\frac{\sum_{r\in R(T_i)}\exp(\gamma_k^TX_r^{(2)})E\left\{\exp(\nu_k^Tb_r)\right\}X_r^{(2)}}{\sum_{r\in R(T_i)}\exp(\gamma_k^TX_r^{(2)})E\left\{\exp(\nu_k^Tb_r)\right\}}\right] + \left(\sum_{j:t_{kj}\le T_i}\frac{d_{kj}\sum_{r\in R(t_{kj})}\exp(\gamma_k^TX_r^{(2)})E\left\{\exp(\nu_k^Tb_r)\right\}X_r^{(2)}}{\left[\sum_{r\in R(t_{kj})}\exp(\gamma_k^TX_r^{(2)})E\left\{\exp(\nu_k^Tb_r)\right\}\right]^2} - \sum_{j:t_{kj}\le T_i}\frac{d_{kj}}{\sum_{r\in R(t_{kj})}\exp(\gamma_k^TX_r^{(2)})E\left\{\exp(\nu_k^Tb_r)\right\}}X_i^{(2)}\right)$$

$$\times \exp(\gamma_k^T X_i^{(2)}) E\left\{\exp(\nu_k^T b_i)\right\}\Big|_{\gamma_k=\hat{\gamma}_k,\nu_k=\hat{\nu}_k}, \tag{A.12}$$

$$\begin{aligned}
\nabla_{\nu_k} l^{(i)}(\hat{\Omega}; Y, C) &= I(D_i = k)\left[E(b_i) - \frac{\sum_{r\in R(T_i)} \exp(\gamma_k^T X_r^{(2)}) E\left\{b_r \exp(\nu_k^T b_r)\right\}}{\sum_{r\in R(T_i)} \exp(\gamma_k^T X_r^{(2)}) E\left\{\exp(\nu_k^T b_r)\right\}}\right] \\
&+ \left(\sum_{j:t_{kj}\le T_i} \frac{d_{kj} \sum_{r\in R(t_{kj})} \exp(\gamma_k^T X_r^{(2)}) E\left\{b_r \exp(\nu_k^T b_r)\right\}}{\left[\sum_{r\in R(t_{kj})} \exp(\gamma_k^T X_r^{(2)}) E\left\{\exp(\nu_k^T b_r)\right\}\right]^2} E\left\{\exp(\nu_k^T b_i)\right\}\right. \\
&\left. - \sum_{j:t_{kj}\le T_i} \frac{d_{kj}}{\sum_{r\in R(t_{kj})} \exp(\gamma_k^T X_r^{(2)}) E\left\{\exp(\nu_k^T b_r)\right\}} E\left\{b_i \exp(\nu_k^T b_i)\right\}\right) \\
&\times \exp(\gamma_k^T X_i^{(2)})\Big|_{\gamma_k=\hat{\gamma}_k,\nu_k=\hat{\nu}_k}.
\end{aligned} \tag{A.13}$$

A.3 Comparison of estimation results using the standard Gauss-Hermite quadrature rule and the pseudo-adaptive Gauss-Hermite quadrature rule

We ran a small simulation study to compare the estimation results based on the standard Gauss-Hermite rule ($n_q = 20$) with those based on the pseudo-adaptive quadrature rule ($n_q = 6$). We generated 500 Monte Carlo samples of size $n = 1,000$ from the joint model (2.15) - (2.17) described in Section 2.2, where the maximum scheduled follow-up time was set as 5 and the average number of longitudinal measurements was about 3 per subject. The overall censoring rate was about 34% (35% for risk 1 and 30% for risk 2). The bias, the standard error (SE), and the estimated standard error (Est. SE) of the parameter estimates are summarized in Table A.1.

It is seen that the performance of the two methods are similar. Both methods have small bias and their Est. SEs are close to the SEs in most cases.

Table A.1: Comparison of bias, standard error (SE), and estimated standard error (Est. SE) between the standard Gauss-Hermite quadrature rule ($n_q = 20$) and the pseudo-adaptive Gauss-Hermite quadrature rule ($n_q = 6$)($n = 1000$)

		Standard			Pseudo-adaptive		
Parameter	True	Bias	SE	Est. SE	Bias	SE	Est. SE
Longitudinal							
Fixed effects							
β_0	10	-0.008	0.043	0.040	0.001	0.043	0.040
β_1	1	-0.003	0.048	0.027	0.002	0.033	0.031
β_2	-1.5	-0.005	0.059	0.058	-0.002	0.058	0.058
σ^2	0.5	0.002	0.018	0.019	-0.001	0.018	0.018
Competing risks							
Fixed effects							
γ_{11}	0.8	-0.025	0.178	0.182	-0.006	0.179	0.182
γ_{12}	-1	0.004	0.126	0.127	-0.005	0.125	0.127
γ_{21}	0.5	0.005	0.193	0.191	-0.005	0.189	0.191
γ_{22}	-1.5	-0.010	0.135	0.138	-0.017	0.134	0.138
Association							
ν_{11}	1	-0.015	0.132	0.129	0.006	0.133	0.128
ν_{12}	0.5	-0.058	0.228	0.202	0.012	0.204	0.209
ν_{21}	0.7	-0.011	0.131	0.131	0.008	0.130	0.131
ν_{22}	0.25	0.022	0.227	0.210	-0.004	0.216	0.215
Random effects							
Σ_{11}	0.5	-0.006	0.041	0.041	-0.001	0.040	0.041
Σ_{22}	0.25	0.001	0.023	0.022	0.001	0.022	0.022
Σ_{12}	0	0.002	0.026	0.024	0.001	0.024	0.024

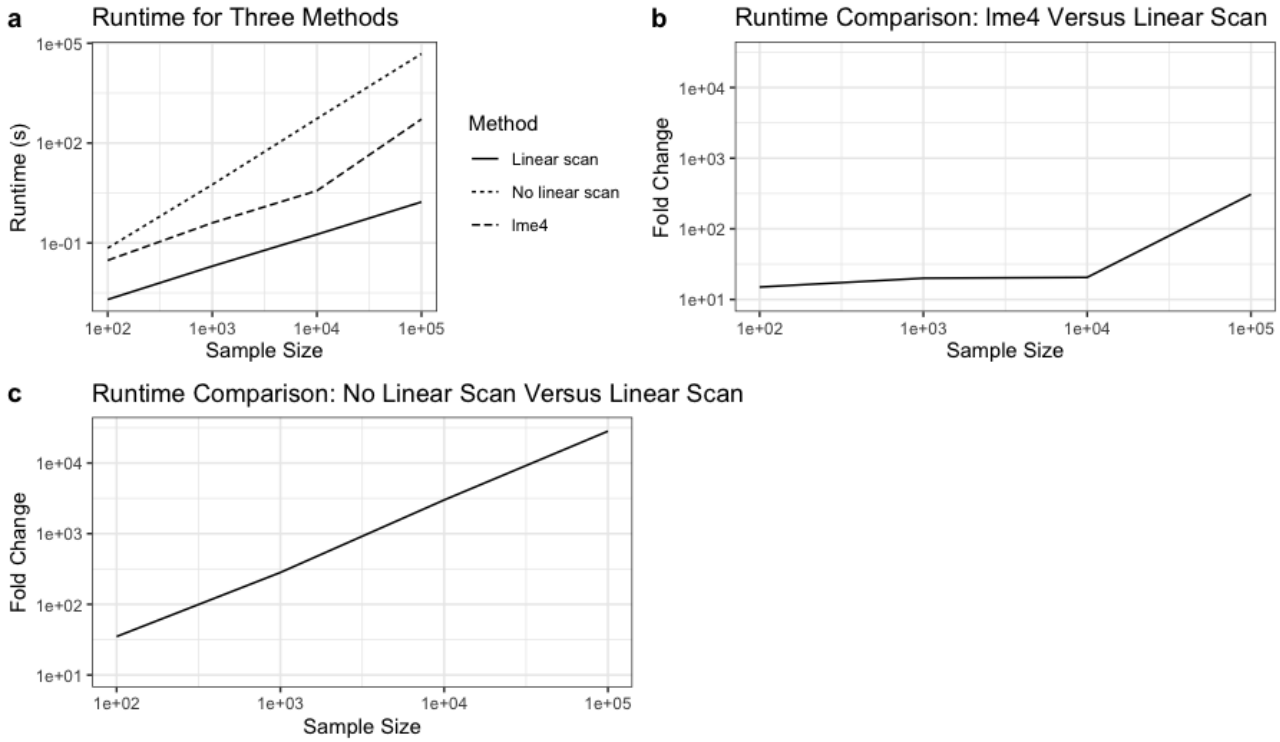

Figure A.1: Runtime (seconds) for three implementations of Empirical Bayes estimates: `lme4`, linear scan, and no linear scan. Fold change is calculated as the ratio of runtime between two methods.

A.4 Contrast of runtime between three implementations of the variance-covariance matrices of the Empirical Bayes estimates for the linear mixed effects model

We conducted a small simulation study to compare the runtime between the linear calculation algorithm of the variance-covariance matrices of the Empirical Bayes estimates described in Remark 1 of Section 2.1.2 and the direct implementation (2.10) over various samples sizes. We also report the runtime of a popular R package **lme4** as a reference. The results are depicted in Figure A.1 below.

It is seen from Figure A.1 that applying the simple linear calculation algorithm can yield

a speed-up by a factor of 10 to 10,000 when n grows from 10 to 10^5. Our implementation was also significantly faster than the popular R package **lme4** by a factor of 10 to 500 as n grows from 10 to 10^5.

A.5 Parameter and standard error estimation results of different implementations of semi-parametric joint models with competing risks data

Figures 2.1 and 2.2 in Section 2.2 have focused on contrasting the computational efficiency of different implementations for parameter estimation and standard error estimation in terms of the runtime. As an illustration, we ran a small simulation to compare their parameter estimates and estimated standard error based on 100 simulated data sets of size $n = 1{,}000$ from the model (2.15) - (2.17). The maximum scheduled follow-up time was set as 5 and the average number of longitudinal measurements was about 3 per subject. The overall censoring rate was about 34% (35% for risk 1 and 30% for risk 2). The estimation results are summarized in Table A.2, where each entry is the average of the parameter and standard error estimates over the 100 simulated data sets. As one would expect, the three different implementations (methods 1-3) yielded almost identical results, whereas `joineR` produced similar estimation results for the longitudinal model, but slightly different results for the competing risks model due to its different latent association structure.

A.6 Analysis results of the lung health study data

Table A.3 summarizes the parameter estimates with 95% confidence intervals for the joint model (2.18) and (2.19) using different joint model packages for the lung health study data.

It is seen that for the longitudinal sub-model, the parameter estimates obtained from these different packages are almost identical, and that they do not seem to be much influ-

Table A.2: Parameter estimates and standard error (SE) for three implementations of the joint model (2.1) and (2.2) and the `joineR` package based on 100 data sets of size $n = 1{,}000$ generated from the model (2.15) - (2.17). The details of Methods 1-3 are given in Section 2.2. Each entry is the average of the parameter and standard error estimates over the 100 simulated data sets.

		Method 1	Method 2	Method 3	`joineR`*
Parameter	True	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)
Longitudinal					
Fixed effects					
β_0	10	10.00 (0.04)	10.00 (0.04)	10.00 (0.04)	9.94 (0.05)
β_1	1	1.00 (0.03)	1.00 (0.03)	1.00 (0.03)	0.95 (0.02)
β_2	-1.5	-1.51 (0.06)	-1.51 (0.06)	-1.51 (0.06)	-1.52 (0.06)
σ^2	0.5	0.50 (0.02)	0.50 (0.02)	0.50 (0.02)	0.49 (0.02)
Competing risks					
Fixed effects					
γ_{11}	0.8	0.78 (0.18)	0.78 (0.18)	0.78 (0.18)	0.72 (0.20)
γ_{12}	-1	-1.02 (0.13)	-1.02 (0.13)	-1.02 (0.13)	-0.90 (0.11)
γ_{21}	0.5	0.50 (0.19)	0.50 (0.19)	0.50 (0.19)	0.45 (0.19)
γ_{22}	-1.5	-1.52 (0.14)	-1.52 (0.14)	-1.52 (0.14)	-1.44 (0.14)
Association					
ν_{11}	1	1.00 (0.13)	1.00 (0.13)	1.00 (0.13)	N/A
ν_{12}	0.5	0.49 (0.21)	0.49 (0.21)	0.49 (0.21)	N/A
ν_{21}	0.7	0.70 (0.13)	0.70 (0.13)	0.70 (0.13)	N/A
ν_{22}	0.25	0.29 (0.21)	0.29 (0.21)	0.29 (0.21)	N/A
Random effects					
Σ_{11}	0.5	0.50 (0.04)	0.50 (0.04)	0.50 (0.04)	0.52 (0.04)
Σ_{22}	0.25	0.25 (0.02)	0.25 (0.02)	0.25 (0.02)	0.25 (0.02)
Σ_{12}	0	0.00 (0.04)	0.00 (0.02)	0.00 (0.02)	-0.01 (0.03)

*No estimates were available for the association parameters ν because `joineR` uses a different latent association structure.

Table A.3: Parameter estimates using different joint model R packages for the lung health study data

Packages	FastJM	joineR	$\texttt{JSM}_a$	$\texttt{JSM}_b$	$\texttt{JM}_{a_1}$	$\texttt{JM}_{a_2}$	$\texttt{JM}_b^*$	$\texttt{JMbayes}_a$	$\texttt{JMbayes}_b^*$
Parameters (95% CI)									
Longitudinal outcome (FVC%)									
Intercept	5.07	5.07	5.07	5.07	5.07	5.07	-	5.06	5.06
	(4.97, 5.18)	(4.95, 5.17)	(4.97, 5.18)	(4.97, 5.18)	(4.97, 5.18)	(4.97, 5.18)	-	(4.96, 5.17)	(4.96, 5.16)
Time	-0.05	-0.05	-0.05	-0.05	-0.05	-0.05	-	-0.05	-0.04
	(-0.06, -0.05)	(-0.06, -0.05)	(-0.06, -0.05)	(-0.06, -0.05)	(-0.06, -0.05)	(-0.06, -0.05)	-	(-0.07, -0.04)	(-0.06, -0.03)
Age	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-	-0.03	-0.03
	(-0.03, -0.03)	(-0.03, -0.03)	(-0.03, -0.03)	(-0.03, -0.03)	(-0.03, -0.03)	(-0.03, -0.03)	-	(-0.03, -0.03)	(-0.03, -0.03)
F10CIGS	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003	-	-0.003	-0.003
	(-0.004, -0.002)	(-0.004, -0.002)	(-0.004, -0.002)	(-0.004, -0.002)	(-0.004, -0.002)	(-0.004, -0.002)	-	(-0.004, -0.002)	(-0.004, -0.002)
Sex									
Female	-	-	-	-	-	-	-	-	-
Male	-0.86	-0.86	-0.86	-0.86	-0.86	-0.86	-	-0.86	-0.86
	(-0.88, -0.83)	(-0.88, -0.84)	(-0.88, -0.83)	(-0.88, -0.83)	(-0.88, -0.83)	(-0.88, -0.83)	-	(-0.88, -0.84)	(-0.88, -0.84)
Logslope	-0.14	-0.14	-0.14	-0.14	-0.14	-0.14	-	-0.13	-0.13
	(-0.15, -0.12)	(-0.15, -0.12)	(-0.15, -0.12)	(-0.15, -0.12)	(-0.15, -0.12)	(-0.15, -0.12)	-	(-0.15, -0.12)	(-0.15, -0.12)
BMI	0.0007	0.0007	0.0007	0.0007	0.0007	0.001	-	0.0004	0.0004
	(-0.002, 0.003)	(-0.002, 0.003)	(-0.002, 0.003)	(-0.002, 0.003)	(-0.002, 0.003)	(-0.002, 0.003)	-	(-0.002, 0.003)	(-0.002, 0.003)
SIP	0.02	0.02	0.02	0.02	0.02	0.02	-	0.02	0.02
	(-0.01, 0.04)	(-0.01, 0.04)	(-0.01, 0.04)	(-0.01, 0.04)	(-0.01, 0.04)	(-0.01, 0.04)	-	(-0.01, 0.04)	(-0.01, 0.04)
SIA	0.03	0.03	0.03	0.03	0.03	0.03	-	0.03	0.03
	(0.01, 0.06)	(0.00, 0.05)	(0.01, 0.06)	(0.01, 0.06)	(0.01, 0.06)	(0.01, 0.06)	-	(0.01, 0.06)	(0.01, 0.06)
SIP:time	0.01	0.01	0.01	0.01	0.01	0.01	-	0.00	0.03
	(0.01, 0.01)	(0.01, 0.01)	(0.01, 0.01)	(0.01, 0.01)	(0.01, 0.01)	(0.01, 0.01)	-	(-0.02, 0.02)	(0.01, 0.05)
SIA:time	0.01	0.01	0.01	0.01	0.01	0.01	-	0.01	0.02
	(0.01, 0.02)	(0.01, 0.02)	(0.01, 0.02)	(0.01, 0.02)	(0.01, 0.02)	(0.01, 0.02)	-	(0.00, 0.02)	(0.01, 0.03)
Error									
σ^2	0.02	0.02	0.02	0.02	0.02	0.02	-	0.02	0.02
	(0.02, 0.02)	(0.02, 0.02)	(0.02, 0.02)	(0.02, 0.02)	(0.01, 0.02)	(0.02, 0.02)	-	(0.02, 0.02)	(0.02, 0.02)
Covariance matrix of b_i									
(Intercept)	0.13	0.13	0.13	0.13	0.13	0.13	-	0.14	0.14
	(0.13, 0.14)	(0.13, 0.14)	(0.13, 0.14)	(0.13, 0.14)	(0.12, 0.15)	(0.11, 0.15)	-	(0.13, 0.14)	(0.13, 0.14)
(Intercept):(time)	0.0012	0.0012	0.0012	0.0012	0.0012	0.001	-	-0.003	-0.0044
	(0.0007, 0.0016)	(0.0011, 0.0012)	(0.0012, 0.0012)	(0.0012, 0.0012)	(0.0011, 0.0012)	(-0.03, 0.032)	-	(-0.0039, -0.0023)	(-0.0090, 0.0002)
(Time)	0.0015	0.0015	0.0015	0.0015	0.0015	0.002	-	0.233	0.234
	(0.0015, 0.0016)	(0.0014, 0.0017)	(0.0015, 0.0016)	(0.0015, 0.0016)	(-0.001, 0.004)	(0, 0.003)	-	(0.225, 0.241)	(0.226, 0.243)
survival outcome (drop out)									
BMI	0.01	0.01	0.02	0.02	0.04	0.02	-	0.01	-0.01
	(-0.01, 0.04)	(-0.01, 0.03)	(0.00, 0.04)	(0.00, 0.04)	(0.02, 0.06)	(0.00, 0.04)	-	(-0.01, 0.03)	(-0.03, 0.02)
SIP	-0.10	-0.10	-0.09	-0.09	-0.06	-0.08	-	-0.07	0.28
	(-0.30, 0.10)	(-0.34, 0.10)	(-0.30, 0.11)	(-0.30, 0.11)	(-0.27, 0.14)	(-0.28, 0.12)	-	(-0.28, 0.10)	(-0.03, 0.59)
SIA	-0.18	-0.18	-0.16	-0.17	-0.14	-0.15	-	-0.15	-0.06
	(-0.38, 0.03)	(-0.39, 0.02)	(-0.37, 0.05)	(-0.38, 0.04)	(-0.35, 0.07)	(-0.36, 0.05)	-	(-0.35, 0.11)	(-0.28, 0.15)
Logslope	0.10	0.09	0.05	0.09	0.16	0.08	-	0.03	0.20
	(-0.02, 0.21)	(-0.07, 0.16)	(-0.07, 0.17)	(-0.02, 0.21)	(0.04, 0.27)	(-0.04, 0.20)	-	(-0.08, 0.15)	(0.09, 0.31)
Sex									
Female	-	-	-	-	-	-	-	-	-
Male	-0.11	-0.09	-0.37	-0.08	0.04	-0.28	-	-0.45	-0.11
	(-0.30, 0.09)	(-0.32, 0.11)	(-0.62, -0.11)	(-0.27, 0.10)	(-0.23, 0.30)	(-0.54, -0.02)	-	(-0.71, -0.15)	(-0.37, 0.15)
Age	0.03	0.03	0.02	0.03	0.05	0.03	-	0.01	0.04
	(0.02, 0.04)	(0.01, 0.04)	(0.01, 0.04)	(0.02, 0.04)	(0.04, 0.07)	(0.01, 0.04)	-	(0.00, 0.03)	(0.02, 0.06)
Association*, ν	Shared Random effect	Current value of Latent process	Current value	Current value of Latent process	Current value	Current value	-	Current value	Shared Random effect
	-0.43 (Intercept)	-0.37	-0.33	-0.37	0.11	-0.23	-	-0.45	-0.60 (Intercept)
	(-0.69, -0.17)	(-0.64, -0.17)	(-0.55, -0.11)	(-0.59, -0.15)	(-0.12, 0.34)	(-0.45, -0.01)	-	(-0.69, -0.15)	(-1.11, -0.28)
	0.93 (Time)								14.86 (Time)
	(-3.56, 5.42)								(13.19, 16.42)

*Association parameters that capture the latent relationship between two sub-models have different interpretations among different packages and cannot be compared directly.

*$\texttt{JM}_b$ failed to converge and thus no parameter estimates were provided.

*Some parameter estimates obtained from $\texttt{JMbayes}_b$ might not converge. Diagnostic plots of the posterior parameter estimates are provided in supplementary materials (See Figure A.2).

enced by the different latent association parameterization $W_i(t)$ in the survival sub-model. For the survival sub-model, the results are also similar between different packages. The only exception is the association parameter because of the different latent association parameterization $W_i(t)$ in different packages.

It is worth pointing out that both `FastJM` and $\texttt{JMbayes}_b$ fit the same shared random effects model while $\texttt{JMbayes}_b$ assumes a parametric B-spline baseline hazard, and they have produced similar results for all parameters except for the association parameter of random slope of time. A diagnostic analysis for the posterior estimates of the association parameter (Figure A.2) reveals that $\texttt{JMbayes}_b$ suffered a convergence issue for this parameter, which explains the discrepancy between the `FastJM` and $\texttt{JMbayes}_b$ results.

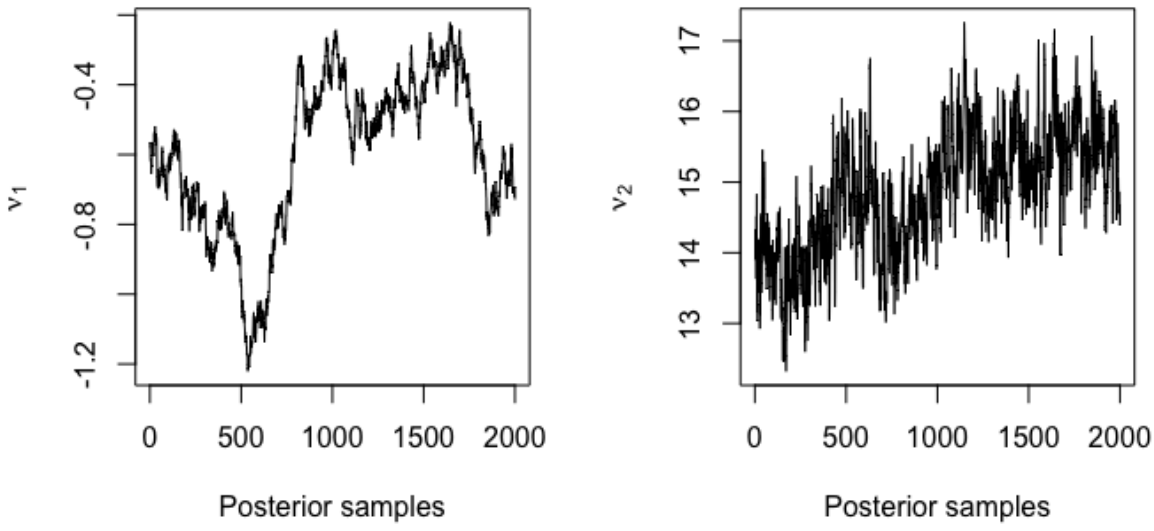

Figure A.2: Diagnostic plot of posterior estimates of the association parameters ν_1 (intercept) and ν_2 (time) based on a Bayesian MCMC sample of 2,000 points obtained from $\texttt{JMbayes}_b$ for lung health study data. (MCMC setup: number of iterations = 40,000; number of burn-in = 15,000; number of thinning = 20.)

A.7 Analysis results of the UK-Biobank primary care (UKB-PC) data

Table A.4 shows the analysis results of the `FastJM` and `joineR` packages for the joint model (Model-L) and (Model-PCH) based on two UKB-PC random subsets of sizes n=5,000 and 20,000, as well as the full UKB-PC data (n=193,287).

It is seen that the analysis results produced by `FastJM` and `joineR` are similar for the longitudinal sub-model for the UKB-PC subset of 5,000 and 20,000 participants. For the survival sub-model, the results are also similar for most parameters except for the association parameters due to the different latent structure $W_i(t)$ between the two packages. It is worth noting that for the full UKB-PC data, `FastJM` was able to finish the analysis in real time (within 1 hour), whereas `joineR` failed to produce any result due to computational failure.

Table A.4: Comparisons of parameter estimates for the longitudinal and competing risks survival outcomes for the UK-biobank primary care data between different R packages (n = 5,000, 20,000, 193,287; Type 1 failure = t2d (type 2 diabetes), Type 2 failure = stroke, MI, or all-cause death)

Package	FastJM						joineR					
Sample size	5,000		20,000		193,287		5,000		20,000		193,287	
Parameter	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
Longitudinal outcome (SBP)												
Intercept	88.13	<0.0001	88.18	<0.0001	88.48	<0.0001	88.55	<0.0001	88.50	<0.0001	-	-
Age at visit	0.39	<0.0001	0.40	<0.0001	0.40	<0.0001	0.38	<0.0001	0.39	<0.0001	-	-
BMI	0.75	<0.0001	0.74	<0.0001	0.73	<0.0001	0.75	<0.0001	0.75	<0.0001	-	-
Sex												
Female	-	-	-	-	-	-	-	-	-	-	-	-
Male	4.18	<0.0001	3.88	<0.0001	4.01	<0.0001	4.10	<0.0001	3.78	<0.0001	-	-
Ethnicity												
White	-	-	-	-	-	-	-	-	-	-	-	-
Non-white	-2.91	0.0006	-1.49	0.0004	-1.32	<0.0001	-2.97	0.0002	-1.46	0.0001	-	-
σ^2	160.47	<0.0001	160.91	<0.0001	162.65	<0.0001	160.22	<0.0001	160.69	<0.0001	-	-
Covariance matrix of b_i												
(Intercept)	1162.76	<0.0001	1143.11	<0.0001	1088.65	<0.0001	1227.14	<0.0001	1200.78	<0.0001	-	-
(Intercept):(age at visit)	-19.21	<0.0001	-18.92	<0.0001	-18.03	<0.0001	-20.34	<0.0001	-19.94	<0.0001	-	-
(age at visit)	0.34	<0.0001	0.34	<0.0001	0.33	<0.0001	0.36	<0.0001	0.36	<0.0001	-	-
Type 1 Failure												
BMI	0.14	<0.0001	0.13	<0.0001	0.14	<0.0001	0.14	<0.0001	0.13	<0.0001	-	-
Sex												
Female	-	-	-	-	-	-	-	-	-	-	-	-
Male	0.64	<0.0001	0.47	<0.0001	0.48	<0.0001	0.63	<0.0001	0.46	<0.0001	-	-
Ethnicity												
White	-	-	-	-	-	-	-	-	-	-	-	-
Non-white	1.42	<0.0001	1.54	<0.0001	1.28	<0.0001	1.40	<0.0001	1.55	<0.0001	-	-
Association*, ν	0.02	0.0077	0.01	0.0039	0.02	<0.0001	0.01	0.2314	0.0042	0.2439	-	-
	1.17	0.0096	0.52	0.0151	0.97	<0.0001	-	-	-	-	-	-
Type 2 Failure												
BMI	0.05	<0.0001	0.03	<0.0001	0.03	<0.0001	0.04	0.0003	0.03	<0.0001	-	-
Sex												
Female	-	-	-	-	-	-	-	-	-	-	-	-
Male	0.68	<0.0001	0.63	<0.0001	0.62	<0.0001	0.66	<0.0001	0.60	<0.0001	-	-
Ethnicity												
White	-	-	-	-	-	-	-	-	-	-	-	-
Non-white	0.27	0.3671	0.30	0.0258	0.23	<0.0001	0.28	0.3335	0.29	0.0180	-	-
Association*, ν	0.02	0.0217	0.02	<0.0001	0.02	<0.0001	0.02	0.0061	0.02	<0.0001	-	-
	1.42	0.0013	1.51	<0.0001	1.35	<0.0001	-	-	-	-	-	-

*Association parameters that capture the latent relationship between the longitudinal and survival sub-models are different between `FastJM` and `joineR` because of their different latent association structures: `FastJM` uses shared random effects, whereas `joineR` uses the current value of the latent process.

- Fail to produce any result due to computational failure.

APPENDIX B

Supplementary Materials for Project 2

B.1 The EM algorithm

B.1.1 The M-step (equation (3.5))

It can be shown that the parameters β, Σ_θ, as well as $\Lambda_{0k}(t_{kj})$ at the q_k distinct observed type k event times $t_{k1} > \cdots > t_{kq_k}$, have closed form solutions in the M-step. Using $E^{(m)}$ to denote $E^{(m)}_{b_i,\omega_i|Y_i,T_i,D_i,\Psi^{(m)}}$ and $a^{\otimes 2} = aa^T$, $\forall a \in \mathbb{R}^d$, we have

$$\begin{aligned}
\beta^{(m+1)} &= \left\{ \sum_{i=1}^{n} \sum_{j=1}^{n_i} \frac{X_i^{(1)\otimes 2}(t_{ij})}{\exp\{W_i^T(t_{ij})\tau^{(m)}\}} E^{(m)}\left[\exp\left\{-V_i(t_{ij})^T\omega_i\right\}\right] \right\}^{-1} \\
&\quad \times \left(\sum_{i=1}^{n} \sum_{j=1}^{n_i} \frac{E^{(m)}\left[\exp\left\{-V_i(t_{ij})^T\omega_i\right\}\right] X_i^{(1)}(t_{ij}) Y_i(t_{ij})}{\exp\{W_i^T(t_{ij})\tau^{(m)}\}} \right. \\
&\quad \left. - \frac{Z_i^T(t_{ij}) E^{(m)}\left[b_i \exp\left\{-V_i(t_{ij})^T\omega_i\right\}\right] X_i^{(1)}(t_{ij})}{\exp\{W_i^T(t_{ij})\tau^{(m)}\}} \right) \qquad \text{(B.1)}
\end{aligned}$$

$$\Sigma_\theta^{(m+1)} = \frac{1}{n}\sum_{i=1}^{n} E^{(m)}(\theta_i^{\otimes 2}) = \frac{1}{n}\sum_{i=1}^{n} \begin{pmatrix} E^{(m)}(b_i^{\otimes 2}) & E^{(m)}(b_i\omega_i^T) \\ E^{(m)}(\omega_i b_i^T) & E^{(m)}(\omega_i^{\otimes 2}) \end{pmatrix}, \qquad \text{(B.2)}$$

$$\Lambda_{0k}^{(m+1)}(t_{kj}) = \sum_{l=j}^{q_k} \frac{d_{kl}}{\sum_{r\in R(t_{kj})} \exp\left\{X_r^{(2)T}(t_{kj})\gamma_k^{(m)}\right\} E^{(m)}\left[\exp\{M_i^T(\theta_i, t_{kj})\alpha_k\}\right]}, \qquad \text{(B.3)}$$

where $R(t_{kj})$ is the risk set at time t_{kj}, and d_{kl} is the number of type k failures at time t_{kl}, for $k = 1, \ldots, K$.

The other parameters τ, γ, α, and ν do not have closed form solutions and we update

them using the one-step Newton-Raphson method

$$\begin{aligned}
\tau^{(m+1)} &= \tau^{(m)} + I_{\tau}^{(m)-1} U_{\tau}^{(m)}, \\
\gamma_k^{(m+1)} &= \gamma_k^{(m)} + I_{\gamma_k}^{(m)-1} U_{\gamma_k}^{(m)}, \quad k = 1, \ldots, K, \\
\alpha_k^{(m+1)} &= \alpha_k^{(m)} + I_{\alpha_k}^{(m)-1} U_{\alpha_k}^{(m)}, \quad k = 1, \ldots, K,
\end{aligned}$$

where

$$\begin{aligned}
I_{\tau}^{(m)} &= \sum_{i=1}^{n} \sum_{j=1}^{n_i} \frac{1}{2} \exp\left\{-W_i^T(t_{ij})\tau^{(m)}\right\} \\
&\quad \times \Big(r_i^{(m+1)2}(t_{ij}) E^{(m)} \left[\exp\left\{-V_i(t_{ij})^T \omega_i\right\}\right] \\
&\quad - 2 r_i^{(m+1)}(t_{ij}) Z_i^T(t_{ij}) E^{(m)} \left[b_i \exp\left\{-V_i(t_{ij})^T \omega_i\right\}\right] \\
&\quad + \operatorname{tr}\left[Z_i^{\otimes 2}(t_{ij}) E^{(m)} \left\{ b_i^{\otimes 2} \exp\left\{-V_i(t_{ij})^T \omega_i\right\}\right\}\right] \Big) W_i^{\otimes 2}(t_{ij}), \qquad \text{(B.4)} \\
U_{\tau}^{(m)} &= \sum_{i=1}^{n} \sum_{j=1}^{n_i} \frac{1}{2} \Big[\exp\left\{-W_i^T(t_{ij})\tau^{(m)}\right\} \\
&\quad \times \Big\{ r_i^{(m+1)2}(t_{ij}) E^{(m)} \left[\exp\left\{-V_i(t_{ij})^T \omega_i\right\}\right] \\
&\quad - 2 r_i^{(m+1)}(t_{ij}) Z_i^T(t_{ij}) E^{(m)} \left[b_i \exp\left\{-V_i(t_{ij})^T \omega_i\right\}\right] \\
&\quad + \operatorname{tr}\left(Z_i^{\otimes 2}(t_{ij}) E^{(m)} \left[b_i^{\otimes 2} \exp\left\{-V_i(t_{ij})^T \omega_i\right\}\right]\right) \Big\} - 1 \Big] W_i(t_{ij}), \qquad \text{(B.5)}
\end{aligned}$$

with $r_i^{(m+1)}(t_{ij}) = Y_i(t_{ij}) - X_i^{(1)T}(t_{ij})\beta^{(m+1)}$,

$$\begin{aligned}
I_{\gamma_k}^{(m)} &= \sum_{i=1}^{n} \sum_{t_{kj} \le T_i} \Delta\Lambda_{0k}(t_{kj})^{(m+1)} \exp\left\{ X_i^{(2)T}(t_{kj}) \gamma_k^{(m)} \right\} \\
&\quad \times E^{(m)} \left[\exp\left\{ M_i^T(\theta_i, t_{kj}) \alpha_k \right\} \right] X_i^{(2)\otimes 2}(t_{kj}), \qquad \text{(B.6)} \\
U_{\gamma_k}^{(m)} &= \sum_{i=1}^{n} I(D_i = k) X_i^{(2)}(T_i) \\
&\quad - \sum_{i=1}^{n} \sum_{t_{kj} \le T_i} \Delta\Lambda_{0k}(t_{kj})^{(m+1)} \exp\left\{ X_i^{(2)T}(t_{kj}) \gamma_k^{(m)} \right\} \\
&\quad \times E^{(m)} \left[\exp\left\{ M_i^T(\theta_i, t_{kj}) \alpha_k \right\} \right] X_i^{(2)}(t_{kj}), \qquad \text{(B.7)} \\
I_{\alpha_k}^{(m)} &= \sum_{i=1}^{n} \sum_{t_{kj} \le T_i} \Delta\Lambda_{0k}(t_{kj})^{(m+1)} \exp\left\{ X_i^{(2)T}(t_{kj}) \gamma_k^{(m)} \right\}
\end{aligned}$$

$$\times E^{(m)}\left[M_i^{\otimes 2}(\theta_i, t_{kj}) \exp\left\{M_i^T(\theta_i, t_{kj})\alpha_k\right\}\right], \tag{B.8}$$

$$\begin{aligned} U_{\alpha_k}^{(m)} &= \sum_{i=1}^{n} I(D_i = k) E(\theta_i) \\ &\quad - \sum_{i=1}^{n} \sum_{t_{kj} \le T_i} \Delta\Lambda_{0k}(t_{kj})^{(m+1)} \exp\left\{X_i^{(2)T}(t_{kj})\gamma_k^{(m)}\right\} \\ &\quad \times E^{(m)}\left[M_i(\theta_i, t_{kj}) \exp\left\{M_i^T(\theta_i, t_{kj})\alpha_k\right\}\right]. \end{aligned} \tag{B.9}$$

B.1.2 Numerical integration for the E-step

We approximate the integral in (3.6) using the standard Gauss-Hermite quadrature rule [Press et al., 2007]. Specifically, the integral in (3.6) can be evaluated numerically as follows:

$$\begin{aligned} E^{(m)}\{h(\theta_i)\} &= \int h(\theta_i) f(\theta_i | Y_i, T_i, D_i, \Psi^{(m)}) d\theta_i \\ &= \frac{\int h(\theta_i) f(Y_i, T_i, D_i, \theta_i | \Psi^{(m)}) d\theta_i}{f(Y_i, T_i, D_i | \Psi^{(m)})} \\ &= \frac{\int h(\theta_i) f(Y_i | \theta_i, \Psi^{(m)}) f(T_i, D_i | \theta_i, \Psi^{(m)}) f(\theta_i | \Psi^{(m)}) d\theta_i}{\int f(Y_i | \theta_i, \Psi^{(m)}) f(T_i, D_i | \theta_i, \Psi^{(m)}) f(\theta_i | \Psi^{(m)}) d\theta_i} \\ &\approx \frac{\sum_{t_1,t_2,\ldots,t_q} \pi_t h(\tilde{\theta}_t^{(m)}) f(Y_i \mid \tilde{\theta}_t^{(m)}, \Psi^{(m)}) f(T_i, D_i \mid \tilde{\theta}_t^{(m)}, \Psi^{(m)}) f(\tilde{\theta}_t^{(m)} \mid \Psi^{(m)}) \exp(||\theta_t||^2)}{\sum_{t_1,t_2,\ldots,t_q} \pi_t f(Y_i \mid \tilde{\theta}_t^{(m)}, \Psi^{(m)}) f(T_i, D_i \mid \tilde{\theta}_t^{(m)}, \Psi^{(m)}) f(\tilde{\theta}_t^{(m)} \mid \Psi^{(m)}) \exp(||\theta_t||^2)}, \end{aligned}$$

where $\sum_{t_1,t_2,\ldots,t_q}$ is the shorthand for $\sum_{t_1=1}^{n_q} \ldots \sum_{t_q=1}^{n_q}$, n_q the number of quadrature points, $\theta_t = (\theta_{t_1}, \theta_{t_2}, ..., \theta_{t_q})^T$ the abscissas with corresponding weights π_t, $\tilde{\theta}_t^{(m)} = \sqrt{2}\Sigma_\theta^{(m)1/2}\theta_t$ the re-scaled alternative abscissas, and $\Sigma_\theta^{(m)1/2}$ the square root of $\Sigma_\theta^{(m)}$ [Elashoff et al., 2008].

B.2 Formulas for standard error estimation

The observed score vector $\nabla_\Omega l^{(i)}(\hat{\Omega}; Y, T, D)$ is needed in equation (3.7) for calculating the empirical Fisher information matrix. Using $\nabla_\Omega l^{(i)}$ to denote $\nabla_\Omega l^{(i)}(\hat{\Omega}; Y, T, D)$ and E to denote $E_{b_i,\omega_i|Y_i,T_i,D_i,\hat{\Psi}}$, the parametric components of the observed score vector in equation (3.7) is given by

$$\nabla_\beta l^{(i)} = \sum_{i=1}^{n_i} \exp\left\{-W_i^T(t_{ij})\tau\right\}$$

$$\times \left(r_i(t_{ij})E\left[\exp\left\{-V_i(t_{ij})^T\omega_i\right\}\right] - Z_i^T(t_{ij})E\left[b_i\exp\left\{-V_i(t_{ij})^T\omega_i\right\}\right]\right) X_i^{(1)}(t_{ij}), \quad \text{(B.10)}$$

$$\begin{aligned}\nabla_\tau l^{(i)} &= \sum_{j=1}^{n_i}\frac{1}{2}\left[\exp\left\{-W_i^T(t_{ij})\tau\right\}\right.\\ &\times\left\{r_i^2(t_{ij})E\left[\exp\left\{-V_i(t_{ij})^T\omega_i\right\}\right]\right.\\ &-2r_i(t_{ij})Z_i^T(t_{ij})E\left[b_i\exp\left\{-V_i(t_{ij})^T\omega_i\right\}\right]\\ &\left.\left.+\mathrm{tr}\left(Z_{ij}^{\otimes 2}E\left[b_i^{\otimes 2}\exp\left\{-V_i(t_{ij})^T\omega_i\right\}\right]\right)\right\}-1\right]W_i(t_{ij}), \quad \text{(B.11)}\end{aligned}$$

$$\nabla_\Sigma l^{(i)} = \frac{1}{2}\left[2\Sigma^{-1}E(\theta_i^{\otimes 2})\Sigma^{-1} - \left\{\Sigma^{-1}E(\theta_i^{\otimes 2})\Sigma^{-1}\circ I\right\} - 2\Sigma^{-1} + \Sigma^{-1}\circ I\right], \quad \text{(B.12)}$$

$$\begin{aligned}\nabla_{\gamma_k} l^{(i)} &= I(D_i=k)\left(X_i^{(2)}(T_i) - \frac{\sum_{r\in R(T_i)}\exp\left\{X_r^{(2)T}(T_i)\gamma_k\right\}E\left[\exp\left\{M_r^T(\theta_r,T_i)\alpha_k\right\}\right]X_r^{(2)}(T_i)}{\sum_{r\in R(T_i)}\exp\left\{X_r^{(2)T}(T_i)\gamma_k\right\}E\left[\exp\left\{M_r^T(\theta_r,T_i)\alpha_k\right\}\right]}\right)\\ &+\sum_{j=1}^{t_{kj}\le T_i}\left\{\frac{d_{kj}\sum_{r\in R(t_{kj})}\exp\left\{X_r^{(2)T}(t_{kj})\gamma_k\right\}E\left[\exp\left\{M_r^T(\theta_r,t_{kj})\alpha_k\right\}\right]X_r^{(2)}(t_{kj})}{\left(\sum_{r\in R(t_{kj})}\exp\left\{X_r^{(2)T}(t_{kj})\gamma_k\right\}E\left[\exp\left\{M_r^T(\theta_r,t_{kj})\alpha_k\right\}\right]\right)^2}\right.\\ &\times\exp\left\{X_i^{(2)T}(t_{kj})\gamma_k\right\}E\left[\exp\left\{M_i^T(\theta_i,t_{kj})\alpha_k\right\}\right]\\ &\left.-\frac{d_{kj}\exp\left\{X_i^{(2)T}(t_{kj})\gamma_k\right\}E\left[\exp\left\{M_i^T(\theta_i,t_{kj})\alpha_k\right\}\right]X_i^{(2)}(t_{kj})}{\sum_{r\in R(t_{kj})}\exp\left\{X_r^{(2)T}(t_{kj})\gamma_k\right\}E\left[\exp\left\{M_r^T(\theta_r,t_{kj})\alpha_k\right\}\right]}\right\}, \quad \text{(B.13)}\end{aligned}$$

$$\begin{aligned}\nabla_{\alpha_k} l^{(i)} &= I(D_i=k)\left[E(\theta_i) - \frac{\sum_{r\in R(T_i)}\exp\left\{X_r^{(2)T}(T_i)\gamma_k\right\}E\left[M_r(\theta_r,T_i)\exp\left\{M_r^T(\theta_r,T_i)\alpha_k\right\}\right]}{\sum_{r\in R(T_i)}\exp\left\{X_r^{(2)T}(T_i)\gamma_k\right\}E\left[\exp\left\{M_r^T(\theta_r,T_i)\alpha_k\right\}\right]}\right]\\ &+\sum_{j=1}^{t_{kj}\le T_i}\left\{\frac{d_{kj}\sum_{r\in R(t_{kj})}\exp\left\{X_r^{(2)T}(t_{kj})\gamma_k\right\}E\left[M_r(\theta_r,t_{kj})\exp\left\{M_r^T(\theta_r,t_{kj})\alpha_k\right\}\right]}{\left(\sum_{r\in R(t_{kj})}\exp\left\{X_r^{(2)T}(t_{kj})\gamma_k\right\}E\left[\exp\left\{M_r^T(\theta_r,t_{kj})\alpha_k\right\}\right]\right)^2}\right.\\ &\times E\left[\exp\left\{M_i^T(\theta_i,t_{kj})\alpha_k\right\}\right]\exp\left\{X_i^{(2)T}(t_{kj})\gamma_k\right\}\\ &\left.-\frac{d_{kj}\exp\left\{X_i^{(2)T}(t_{kj})\gamma_k\right\}E\left[M_i(\theta_i,t_{kj})\exp\left\{M_i^T(\theta_i,t_{kj})\alpha_k\right\}\right]}{\sum_{r\in R(t_{kj})}\exp\left\{X_r^{(2)T}(t_{kj})\gamma_k\right\}E\left[\exp\left\{M_r^T(\theta_r,T_i)\alpha_k\right\}\right]}\right\}. \quad \text{(B.14)}\end{aligned}$$

B.3 Computational aspects

When the latent association structure of the joint model (3.1) - (3.3) is assumed to be shared random effects and the survival covariates are assumed to be time-independent, i.e., $X_i^{(2)}(t) =$

$X_i^{(2)}$, we derive linear scan algorithms to reduce the computational burden using similar ideas to [Li et al., 2022a]. Below we discuss some linear scan algorithms for implementation of the EM steps and standard error estimation.

B.3.1 Linear scan for the E-step

As discussed in Supplementary Materials B.1.1, the E-step involves evaluating expected values of multiple $h(\theta_i)$'s at each EM iteration, which requires calculating $f(T_i, D_i \mid \theta_i, \Psi^{(m)})$ across all subjects (See Supplementary Materials B.1.2). Note that $f(T_i, D_i \mid \theta_i, \Psi^{(m)})$ can be rewritten as

$$\begin{aligned} f(T_i, D_i \mid \theta_i, \Psi) &= \prod_{k=1}^{K} \left[\Delta\Lambda_{0k}^{(m)}(T_i) \exp^{(m)} \left\{X_i^{(2)T}(T_i)\gamma_k + M_i^T(\theta_i, T_i)\alpha_k\right\}\right]^{I(D_i=k)} \\ &\quad \times \exp\left[-\sum_{k=1}^{K} \int_0^{T_i} \exp^{(m)} \left\{X_i^{(2)T}(t)\gamma_k + M_i^T(\theta_i, t)\alpha_k\right\} d\Lambda_{0k}^{(m)}(t)\right] \\ &= \prod_{k=1}^{K} \left\{\Delta\Lambda_{0k}^{(m)}(T_i) \exp^{(m)}(X_i^{(2)T}\gamma_k + \alpha_{bk}^T b_i + \alpha_{\omega k}^T \omega_i)\right\}^{I(D_i=k)} \\ &\quad \times \exp\left\{-\sum_{k=1}^{K} \Lambda_{0k}^{(m)}(T_i) \exp^{(m)}(X_i^{(2)T}\gamma_k + \alpha_{bk}^T b_i + \alpha_{\omega k}^T \omega_i)\right\}. \end{aligned}$$

For each subject i, calculating $\Lambda_{0k}(T_i)$ would involve $O(n)$ operations if a global search is performed to find an interval of two adjacent uncensored event times that contains T_i. Consequently, calculating all $\Lambda_{0k}(T_i)$'s will require $O(n^2)$ operations. Similar to [Li et al., 2022a], by taking advantage of the fact that $\Lambda_{0k}(t)$ is a right-continuous and non-decreasing step function, we define the following a linear scan map

$$\{\Lambda_{0k}^{(m)}(t_{k1}), \Lambda_{0k}^{(m)}(t_{k2}), \ldots \Lambda_{0k}^{(m)}(t_{kq_k})\} \mapsto \{\Lambda_{0k}^{(m)}(T_{(1)}), \Lambda_{0k}^{(m)}(T_{(2)}), \ldots, \Lambda_{0k}^{(m)}(T_{(n)})\}, \tag{B.15}$$

where $t_{k1} > \cdots > t_{kq_k}$ are scanned forward from the largest to the smallest, and for each t_{kj}, only a subset of the ranked observation times $T_{(i)}$ are scanned forward to calculate $\Lambda_{0k}^{(m)}(T_{(i)})$

as follows

$$\Lambda_{0k}^{(m)}(T_{(i)}) = \begin{cases} \Lambda_{0k}^{(m)}(t_{k1}), & \text{if } T_{(i)} \geq t_{k1}, \\ \Lambda_{0k}^{(m)}(t_{k(j+1)}), & \text{if } T_{(i)} \in [t_{k(j+1)}, t_{kj}), \text{ for some } j \in \{1, \ldots, q_k - 1\}, \\ 0, & T_{(i)} < t_{kq_k}. \end{cases}$$

Consequently, the entire algorithm for calculating all $\Lambda_{0k}(T_i)$'s costs only $O(n)$ operations since the scanned $T_{(i)}$'s for different t_{kj}'s do not overlap.

B.3.2 Linear risk set scan for the M-step

Multiple quantities in (B.6)-(B.9) including the cumulative baseline hazard functions involve aggregating information over the risk set $R(t_{kj}) = \{r : T_r \geq t_{kj}\}$ at each uncensored event time t_{kj}, which are further aggregated across all t_{kj}'s. All subjects are scanned to determine the risk set $R(t_{kj})$ for all uncensored event times will require $O(n^2)$ operations. Specifically, to update $\Lambda_{0k}^{(m+1)}(t_{kj})$, $\gamma_k^{(m+1)}$ and $\alpha_k^{(m+1)}$, one needs to compute $\sum_{r \in R(t_{kj})} a_r(t_{kj})$, where $a_r(.)$ is any time-dependent quantity defined in equations (B.6)-(B.9). Note that when $a_r(.)$ is assumed to be time-independent, i.e., $a_r(.) = a_r$, the risk set $R(t_{k(j+1)})$ can be decomposed into two disjoint sets:

$$\sum_{r \in R(t_{k(j+1)})} a_r = \sum_{r \in R(t_{kj})} a_r + \sum_{\{r: T_{(r)} \in [t_{k(j+1)}, t_{kj})\}} a_r, \tag{B.16}$$

where the distinct uncensored event times $t_{k1} > \cdots > t_{kq_k}$ are arranged in a decreasing order. it is easy to see that calculating $\sum_{r \in R(t_{kj})} a_r$, $j = 1, \ldots q_k$, takes $O(n)$ operations when $T_{(r)}$'s are scanned backward in time, by following the recursive formula (B.16) where the subjects in $R(t_{kj})$ do not need to be scanned to calculate the second term.

B.3.3 Linear risk set scan for standard error estimation

Standard error estimation formula in (3.7) relies on the observed score vectors from the profile likelihood where the baseline hazards are profiled out. It is seen from equations (B.13)-(B.14)

that obtaining the observed score vectors $\nabla_{\gamma_k} l^{(i)}(\hat{\Omega}; Y, T, D)$ and $\nabla_{\alpha_k} l^{(i)}(\hat{\Omega}; Y, T, D)$ involve aggregating information either over $\{r \in R(T_i)\}$ or over both $\{r \in R(t_{kj})\}$ and $\{j : t_{kj} \leq T_i\}$, which can takes either $O(n)$ or $O(n^2)$ operations, respectively, if not optimized. As a result, the empirical Fisher information matrix can take $O(n^3)$ operations as it requires summing up the information across all subjects. Specifically, to calculate the gradient $\nabla_{\gamma_k} l^{(i)}(\hat{\Omega}; Y, T, D)$ and $\nabla_{\alpha_k} l^{(i)}(\hat{\Omega}; Y, T, D)$, one needs to compute

$$B(T_i) = \sum_{j: t_{kj} \leq T_i} b_{kj}(t_{kj}), \quad \text{for } i = 1, ..., n,$$

where $B(.)$ is a right-continuous non-decreasing step function and $b_{kj}(t_{kj}) = \sum_{r \in R(t_{kj})} a_r(t_{kj})$ is any time-dependent quantity defined in equations (B.13) - (B.14). When $a_r(.)$ is assumed to be time-independent, $b_{kj}(.) = b_{kj}$. Note that $B(t_{k1}), \ldots, B(t_{kq_k})$ can be computed in $O(n)$ operations as one scans through $t_{k1}, \ldots, t_{kq_k}$ backward in time, following the recursive formula (B.16). Furthermore, analogous to (B.15), the following linear scan algorithm can be used to calculate $\{B(T_{(1)}), B(T_{(2)}), \cdots, B(T_{(n)})\}$ from $\{B(t_{k1}), \ldots, B(t_{kq_k})\}$:

$$\{B(t_{k1}), \ldots, B(t_{kq_k})\} \mapsto \{B(T_{(1)}), B(T_{(2)}), \cdots, B(T_{(n)})\},$$

where for each t_{kj}, only a subset of the ranked observation times $T_{(i)}$'s are scanned forward to calculate $B(T_{(i)})$'s as follows

$$B(T_{(i)}) = \begin{cases} B(t_{k1}), & \text{if } T_{(i)} \geq t_{k1}, \\ B(t_{k(j+1)}), & \text{if } T_{(i)} \in [t_{k(j+1)}, t_{kj}), \text{ for some } j \in \{1, \ldots, q_k - 1\}, \\ 0, & \text{otherwise.} \end{cases}$$

Consequently, calculating all $B(T_{(i)})$'s takes $O(n)$ operations.

APPENDIX C

Supplementary Materials for Project 3

C.1 Derivation for equation (4.8) in the E-step in Section 4.1.2.3

$$
\begin{aligned}
p_i^{(m)}(S_i = s_l | L_i, R_i, \Psi_\phi^{(m)}) &= \frac{p_i^{(m)}(S_i = s_l, S_i \in (L_i, R_i] | \Psi_\phi^{(m)})}{p_i^{(m)}(S_i \in (L_i, R_i] | \Psi_\phi^{(m)})} \\
&= \frac{\mathbf{1}\{s_l \in (L_i, R_i]\}\, p_i^{(m)}(S_i = s_l | \Psi_\phi^{(m)})}{p_i^{(m)}(S_i \in (L_i, R_i] | \Psi_\phi^{(m)})} \\
&= \mathbf{1}\{s_l \in (L_i, R_i]\}\, \frac{p_i^{(m)}(S_i = s_l, S_i > L_i | \Psi_\phi^{(m)})}{p_i^{(m)}(S_i > L_i, S_i \le R_i | \Psi_\phi^{(m)})} \\
&= \mathbf{1}\{s_l \in (L_i, R_i]\}\, \frac{p_i^{(m)}(S_i = s_l | S_i > L_i, \Psi_\phi^{(m)}) p_i^{(m)}(S_i > L_i | \Psi_\phi^{(m)})}{p_i^{(m)}(S_i \le R_i | S_i > L_i, \Psi_\phi^{(m)}) p_i^{(m)}(S_i > L_i | \Psi_\phi^{(m)})} \\
&= \mathbf{1}\{s_l \in (L_i, R_i]\}\, \frac{p_i^{(m)}(S_i = s_l | S_i > L_i, \Psi_\phi^{(m)})}{p_i^{(m)}(S_i \le R_i | S_i > L_i, \Psi_\phi^{(m)})} \\
&= \mathbf{1}\{s_l \in (L_i, R_i]\}\, \frac{p_i^{(m)}(S_i = s_l | S_i > L_i, \Psi_\phi^{(m)})}{1 - p_i^{(m)}(S_i > R_i | S_i > L_i, \Psi_\phi^{(m)})} \\
&= \mathbf{1}\{s_l \in (L_i, R_i]\}\, \frac{\exp\left(-\sum_{L_i < s_u < s_l} \phi_{s_u}\right)\{1 - \exp(-\phi_{s_l})\}}{1 - \exp\left(-\sum_{L_i < s_u \le R_i} \phi_{s_u}\right)}.
\end{aligned}
$$

C.2 M-step solutions for equation (4.9) in Section 4.1.2.3

C.2.1 Update for Ψ_ϕ

From equation (4.6), $Q_\phi(\Psi;\Psi^{(m)})$ can be rewritten as

$$\begin{aligned} Q_\phi(\Psi;\Psi^{(m)}) &= \sum_{i=1}^{n}\sum_{s_l}\log\left\{\exp(-\sum_{s_u<s_l}\phi_{s_u})-\exp(-\sum_{s_u\le s_l}\phi_{s_u})\right\}\boldsymbol{p}_i^{(m)}(s_l) \\ &= \sum_{i=1}^{n}\sum_{s_l}\log\left\{\exp(-\sum_{s_u<s_l}\phi_{s_u})(1-\exp(-\phi_{s_l}))\right\}\boldsymbol{p}_i^{(m)}(s_l) \\ &= \sum_{i=1}^{n}\sum_{s_l}\left\{-\sum_{s_u<s_l}\phi_{s_u}+\log(1-\exp(-\phi_{s_l}))\right\}\boldsymbol{p}_i^{(m)}(s_l). \end{aligned} \tag{C.1}$$

We aim to maximize equation (C.1) in the M-step. ϕ_{s_u} has an analytical solution by letting

$$\begin{aligned} &\frac{\partial Q_\phi(\Psi;\Psi^{(m)})}{\partial\phi_{s_u}}=0 \\ &\Rightarrow \phi_{s_u}^{(m+1)} = -\log\left(1-\frac{\sum_{i=1}^{n}\boldsymbol{p}_i^{(m)}(s_u)}{\sum_{i=1}^{n}\sum_{s_l\ge s_u}\boldsymbol{p}_i^{(m)}(s_l)}\right). \end{aligned} \tag{C.2}$$

C.2.2 Update for $\Psi_Y, \Psi_T, \Psi_\theta$

The parameters $\Psi_Y=(\beta,\tau)^T, \Psi_T=(\gamma,\alpha)^T, \Psi_\theta=\text{vech}(\Sigma_\theta)$ can be updated in the M-step (4.9). It can be shown that β and Σ_θ have closed form solutions:

$$\begin{aligned} \beta^{(m+1)} &= \left[\sum_{i=1}^{n}\sum_{s_l\in\mathbb{S}}\left\{\sum_{j=1}^{n_i}\frac{X_i^{(1)}(t_{ij,l},s_l)X_i^{(1)T}(t_{ij,l},s_l)}{\exp(W_i^T(t_{ij,l},s_l)\tau^{(m)})}\right\}\boldsymbol{p}_i^{(m)}(s_l)E^{(m)}(\exp(-\omega_i))\right]^{-1} \\ &\quad\times\left[\sum_{i=1}^{n}\sum_{s_l\in\mathbb{S}}\left\{\sum_{j=1}^{n_i}\frac{Y_i(t_{ij,l})E^{(m)}(\exp(-\omega_i))X_i^{(1)}(t_{ij,l},s_l)-Z_i^T(t_{ij,l})E^{(m)}(b_i\exp(-\omega_i))X_i^{(1)}(t_{ij,l},s_l)}{\exp(W_i^T(t_{ij,l},s_l)\tau^{(m)})}\right\}\right. \\ &\quad\left.\times\,\boldsymbol{p}_i^{(m)}(s_l)\right], \end{aligned} \tag{C.3}$$

$$\Sigma_\theta^{(m+1)} = \frac{1}{n}\sum_{i=1}^{n}E^{(m)}(\theta_i\theta_i^T). \tag{C.4}$$

Denote $t_1>...>t_k$ the distinct uncensored failure time consisting of all possible values of t_{i,s_l} when $\delta_i=1$. The cumulative baseline hazard function has a closed form solution, given

by

$$\Lambda_0^{(m+1)}(t) = \sum_{t_v \le t} \frac{\boldsymbol{p}^{(m)}(t_v)}{\sum_{r \in R(t_v)} \exp(X_r^{(2)T}(s_l)\gamma^{(m)}) E^{(m)}\left[\exp\left\{W_r^T(\theta_r|t_v)\alpha^{(m)}\right\}\right] \boldsymbol{p}_r^{(m)}(s_l)}, \quad \text{(C.5)}$$

where $\boldsymbol{p}^{(m)}(t_v) = \sum_i \delta_i \mathbf{1}(t_v = t_{i,s_l}) \boldsymbol{p}_i^{(m)}(s_l)$ is the jump size at t_v, $R(t_v)$ is the risk set at time $t_v, v = 1, \ldots, k$. Note that $\Lambda_0^{(m+1)}(.)$ only has jumps at the uncensored event times t_v's. It is clear that $\Lambda_0^{(m+1)}(t)$ is a right-continuous and non-decreasing function.

The rest of the parameters τ, γ, α, and ν do not have closed-form solutions and we update them using the one-step Newton-Raphson method

$$\begin{aligned}
\tau^{(m+1)} &= \tau^{(m)} + I_\tau^{(m)-1} U_\tau^{(m)}, \\
\gamma^{(m+1)} &= \gamma^{(m)} + I_\gamma^{(m)-1} U_\gamma^{(m)}, \\
\alpha^{(m+1)} &= \alpha^{(m)} + I_\alpha^{(m)-1} U_\alpha^{(m)},
\end{aligned}$$

where

$$\begin{aligned}
I_\tau^{(m)} &= \sum_{i=1}^{n} \sum_{s_l \in \mathbb{S}} \sum_{j=1}^{n_i} \frac{1}{2} w_i(t_{ij,l}, s_l) W_i^T(t_{ij,l}, s_l) \exp\left\{-W_i^T(t_{ij,l}, s_l)\tau^{(m)}\right\} \\
&\quad \times \Bigg(\left\{ Y_i(t_{ij,l}) - X_i^{(1)T}(t_{ij,l}, s_l)\beta^{(m+1)} \right\}^2 E^{(m)}\left\{\exp(-\omega_i)\right\} \\
&\quad - 2 \left\{ Y_i(t_{ij,l}) - X_i^{(1)T}(t_{ij,l}, s_l)\beta^{(m+1)} \right\} Z_i^T(t_{ij,l}) E^{(m)}\left\{b_i \exp(-\omega_i)\right\} \\
&\quad + \operatorname{tr}\left[Z_i(t_{ij,l}) Z_i^T(t_{ij,l}) E^{(m)}\left\{ b_i b_i^T \exp(-\omega_i) \right\} \right] \Bigg) \boldsymbol{p}_i^{(m)}(s_l), \qquad \text{(C.6)} \\
U_\tau^{(m)} &= \sum_{i=1}^{n} \sum_{s_l \in \mathbb{S}} \sum_{j=1}^{n_i} \frac{1}{2} w_i(t_{ij,l}) \Big\{ \exp(-W_i^T(t_{ij,l}, s_l)\tau^{(m)}) \\
&\quad \times \Bigg(\left\{ Y_i(t_{ij,l}) - X_i^{(1)T}(t_{ij,l}, s_l)\beta^{(m+1)} \right\}^2 E^{(m)}\left\{\exp(-\omega_i)\right\} \\
&\quad - 2 \left\{ Y_i(t_{ij,l}) - X_i^{(1)T}(t_{ij,l}, s_l)\beta^{(m+1)} \right\} Z_i^T(t_{ij,l}) E^{(m)}\left\{b_i \exp(-\omega_i)\right\} \\
&\quad + \operatorname{tr}\left[Z_i(t_{ij,l}) Z_i^T(t_{ij,l}) E^{(m)}\left\{ b_i b_i^T \exp(-\omega_i) \right\} \right] \Bigg) - 1 \Big\} \boldsymbol{p}_i^{(m)}(s_l), \qquad \text{(C.7)} \\
I_\gamma^{(m)} &= \sum_{i=1}^{n} \sum_{s_l \in \mathbb{S}} \Bigg[\sum_{t_v \le t_{i,s_l}} \Delta\Lambda_0^{(m+1)}(t_v) \exp(X_i^{(2)T}(s_l)\gamma^{(m)}) E^{(m)}\left[\exp\left\{W_i^T(\theta_i|t_v)\alpha^{(m)}\right\}\right] \\
&\quad \times X_i^{(2)}(s_l) X_i^{(2)T}(s_l) \Bigg] \boldsymbol{p}_i^{(m)}(s_l), \qquad \text{(C.8)} \\
U_\gamma^{(m)} &= \sum_{i=1}^{n} \sum_{s_l \in \mathbb{S}} \Big[\delta_i X_i^{(2)}(s_l)
\end{aligned}$$

$$
\begin{aligned}
&- \sum_{t_v \le t_{i,s_l}} \Delta\Lambda_0^{(m+1)}(t_v) \exp(X_i^{(2)T}(s_l)\gamma^{(m)}) E^{(m)} \left[\exp\left\{W_i^T(\theta_i|t_v)\alpha^{(m)}\right\}\right] X_i^{(2)}(s_l)\Bigg] \\
&\times \boldsymbol{p}_i^{(m)}(s_l), \qquad \text{(C.9)}
\end{aligned}
$$

$$
\begin{aligned}
I_\alpha^{(m)} &= \sum_{i=1}^{n}\sum_{s_l\in\mathbb{S}} \left(\sum_{t_v \le t_{i,s_l}} \Delta\Lambda_0^{(m+1)}(t_v) \exp(X_i^{(2)T}(s_l)\gamma^{(m)}) E^{(m)} \left[W_i(\theta_i|t_v)^{\otimes 2} \exp\left\{W_i^T(\theta_i|t_v)\alpha^{(m)}\right\}\right] \right) \\
&\times \boldsymbol{p}_i^{(m)}(s_l), \qquad \text{(C.10)}
\end{aligned}
$$

$$
\begin{aligned}
U_\alpha^{(m)} &= \sum_{i=1}^{n}\sum_{s_l\in\mathbb{S}} \delta_i E^{(m)} \left\{W_i^T(\theta_i|t_{i,s_l})\right\} \boldsymbol{p}_i^{(m)}(s_l) \\
&- \sum_{i=1}^{n}\sum_{s_l\in\mathbb{S}} \left[\sum_{t_v \le t_{i,s_l}} \Delta\Lambda_0^{(m+1)}(t_v) \exp(X_i^{(2)T}(s_l)\gamma^{(m)}) E^{(m)} \left\{W_i(\theta_i|t_v) \exp(W_i^T(\theta_i|t_v)\alpha^{(m)})\right\} \right] \\
&\times \boldsymbol{p}_i^{(m)}(s_l). \qquad \text{(C.11)}
\end{aligned}
$$

C.3 Formulas for standard error estimation to equation (4.11) in Section 4.1.3

The observed score vector $\nabla_\Omega l^{(i)}(\hat{\Omega}; \boldsymbol{D})$ is needed for calculating the empirical Fisher information matrix. The components of the observed score vector in equation (4.11) is given by

$$
\begin{aligned}
\nabla_\beta l^{(i)}(\hat{\Omega}; \boldsymbol{D}) &= \sum_{s_l\in\mathbb{S}} \boldsymbol{p}_i(s_l) \sum_{j=1}^{n_i} \exp\left\{-W_i^T(t_{ij,l}, s_l)\tau\right\} \left[\left\{ Y_i(t_{ij,l}) - X_i^{(1)T}(t_{ij,l}, s_l)\beta \right\} E\left\{\exp(-\omega_i)\right\} \right. \\
&\left. - Z_{ij}^T E\left\{b_i \exp(-\omega_i)\right\} \right] X_{ij}^{(1)}|_{\beta=\hat{\beta},\tau=\hat{\tau}}, \qquad \text{(C.12)} \\
\nabla_\tau l^{(i)}(\hat{\Omega}; \boldsymbol{D}) &= \sum_{s_l\in\mathbb{S}} \boldsymbol{p}_i(s_l) \sum_{j=1}^{n_i} \frac{1}{2} W_i^T(t_{ij,l}, s_l) \left\{\exp\left\{-W_i^T(t_{ij,l}, s_l)\tau\right\}\right. \\
&\times \left(\left\{ Y_i(t_{ij,l}) - X_i^{(1)T}(t_{ij,l}, s_l)\beta \right\}^2 E\left\{\exp(-\omega_i)\right\} \right. \\
&\left. - 2 \left\{ Y_i(t_{ij,l}) - X_i^{(1)T}(t_{ij,l}, s_l)\beta \right\} Z_{ij}^T E\left\{b_i \exp(-\omega_i)\right\} \right.
\end{aligned}
$$

$$+ \text{ tr}\left[Z_{ij}Z_{ij}^T E\left\{b_i b_i^T \exp(-\omega_i)\right\}\right]) - 1\Big\} \mid_{\beta=\hat{\beta},\tau=\hat{\tau}}, \tag{C.13}$$

$$\nabla_\Sigma l^{(i)}(\hat{\Omega};\boldsymbol{D}) = \frac{1}{2}\left[2\Sigma^{-1}E(\theta_i\theta_i^T)\Sigma^{-1} - \left\{\Sigma^{-1}E(\theta_i\theta_i^T)\Sigma^{-1}\circ I\right\} - 2\Sigma^{-1} + \Sigma^{-1}\circ I\right] \mid_{\Sigma=\hat{\Sigma}}, \tag{C.14}$$

$$\begin{aligned}
\nabla_\gamma l^{(i)}(\hat{\Omega};\boldsymbol{D}) = & \sum_{s_l\in\mathbb{S}} \delta_i \boldsymbol{p}_i(s_l) X_i^{(2)}(s_l) \\
& - \sum_{s_l\in\mathbb{S}} \delta_i \boldsymbol{p}_i(s_l) \frac{\sum_{r\in R(t_{i,s_l})} \boldsymbol{p}_r(s_l)\exp\left\{X_r^{(2)T}(s_l)\gamma\right\} E\left\{\exp\left\{W_r^T(\theta_r|t_{i,s_l})\alpha\right\}\right\} X_r^{(2)}(s_l)}{\sum_{r\in R(t_{i,s_l})} \boldsymbol{p}_r(s_l)\exp\left\{X_r^{(2)T}(s_l)\gamma\right\} E\left\{\exp\left\{W_r^T(\theta_r|t_{i,s_l})\alpha\right\}\right\}} \\
& + \sum_{s_l\in\mathbb{S}} \boldsymbol{p}_i(s_l) \left(\sum_{v:t_v\le t_{i,s_l}} \frac{\boldsymbol{p}(t_v)\sum_{r\in R(t_v)} \boldsymbol{p}_r(s_l)\exp\left\{X_r^{(2)T}(s_l)\gamma\right\} E\left\{\exp\left\{W_r^T(\theta_r|t_v)\alpha\right\}\right\} X_r^{(2)}(s_l)}{\left[\sum_{r\in R(t_v)} \boldsymbol{p}_r(s_l)\exp\left\{X_r^{(2)T}(s_l)\gamma\right\} E\{\exp\{W_r^T(\theta_r|t_v)\alpha\}\}\right]^2} \right. \\
& \times \exp\left\{X_i^{(2)T}(s_l)\gamma\right\} E\left\{\exp(W_i^T(\theta_i|t_v)\alpha)\right\} \\
& \left. - \sum_{v:t_v\le t_{i,s_l}} \frac{\boldsymbol{p}(t_v)\exp\left\{X_i^{(2)T}(s_l)\gamma\right\} E\left\{\exp(W_i^T(\theta_i|t_v)\alpha)\right\} X_i^{(2)}(s_l)}{\sum_{r\in R(t_v)} \boldsymbol{p}_r(s_l)\exp\left\{X_r^{(2)T}(s_l)\gamma\right\} E\{\exp\{W_r^T(\theta_r|t_v)\alpha\}\}} \right) \mid_{\gamma=\hat{\gamma},\alpha=\hat{\alpha}},
\end{aligned} \tag{C.15}$$

$$\begin{aligned}
\nabla_\alpha l^{(i)}(\hat{\Omega};\boldsymbol{D}) = & \sum_{s_l\in\mathbb{S}} \delta_i \boldsymbol{p}_i(s_l) E\left\{W_i^T(\theta_i|t_{i,s_l})\right\} \\
& - \sum_{s_l\in\mathbb{S}} \delta_i \boldsymbol{p}_i(s_l) \frac{\sum_{r\in R(t_{i,s_l})} \boldsymbol{p}_r(s_l)\exp\left\{X_r^{(2)T}(s_l)\gamma\right\} E\left\{W_r(\theta_r|t_{i,s_l})\exp\left\{W_r^T(\theta_r|t_{i,s_l})\alpha\right\}\right\}}{\sum_{r\in R(t_{i,s_l})} \boldsymbol{p}_r(s_l)\exp\left\{X_r^{(2)T}(s_l)\gamma\right\} E\left\{\exp\left\{W_r^T(\theta_r|t_{i,s_l})\alpha\right\}\right\}} \\
& + \sum_{s_l\in\mathbb{S}} \boldsymbol{p}_i(s_l) \left(\sum_{v:t_v\le t_{i,s_l}} \frac{\boldsymbol{p}(t_v)\sum_{r\in R(t_v)} \boldsymbol{p}_r(s_l)\exp\left\{X_r^{(2)T}(s_l)\gamma\right\} E\left\{W_r(\theta_r|t_v)\exp\left\{W_r^T(\theta_r|t_v)\alpha\right\}\right\}}{\left[\sum_{r\in R(t_v)} \boldsymbol{p}_r(s_l)\exp\left\{X_r^{(2)T}(s_l)\gamma\right\} E\{\exp\{W_r^T(\theta_r|t_v)\alpha\}\}\right]^2} \right. \\
& \times \exp\left\{X_i^{(2)T}(s_l)\gamma\right\} E\left\{\exp(W_i^T(\theta_i|t_v)\alpha)\right\} \\
& \left. - \sum_{v:t_v\le t_{i,s_l}} \frac{\boldsymbol{p}(t_v)\exp\left\{X_i^{(2)T}(s_l)\gamma\right\} E\left\{W_i(\theta_i|t_v)\exp(W_i^T(\theta_i|t_v)\alpha)\right\}}{\sum_{r\in R(t_v)} \boldsymbol{p}_r(s_l)\exp\left\{X_r^{(2)T}(s_l)\gamma\right\} E\{\exp\{W_r^T(\theta_r|t_v)\alpha\}\}} \right) \mid_{\gamma=\hat{\gamma},\alpha=\hat{\alpha}},
\end{aligned} \tag{C.16}$$

$$\frac{\partial l^{(i)}(\hat{\Omega};\boldsymbol{D})}{\partial \phi_{s_u}} = \sum_{s_l\in(L_i,R_i]} \left\{-\mathbf{1}(s_u < s_l) + \frac{\exp(-\phi_{s_l})}{1-\exp(-\phi_{s_l})}\mathbf{1}(s_u = s_l)\right\} \boldsymbol{p}_i(s_l)\big|_{\phi_{s_l}=\hat{\phi}_{s_l}}. \tag{C.17}$$

Bibliography

Soohyun Ahn, Johan Lim, Myunghee Cho Paik, Ralph L Sacco, and Mitchell S Elkind. Cox model with interval-censored covariate in cohort studies. *Biometrical Journal*, 60(4): 797–814, 2018.

Maha Alsefri, Maria Sudell, Marta García-Fiñana, and Ruwanthi Kolamunnage-Dona. Bayesian joint modelling of longitudinal and time to event data: a methodological review. *BMC Medical Research Methodology*, 20:1–17, 2020.

L Altstein and Gang Li. Latent subgroup analysis of a randomized clinical trial through a semiparametric accelerated failure time mixture model. *Biometrics*, 69(1):52–61, 2013.

Jessica Barrett and Li Su. Dynamic predictions using flexible joint models of longitudinal and time-to-event data. *Statistics in medicine*, 36(9):1447–1460, 2017.

Douglas Bates, Deepayan Sarkar, Maintainer Douglas Bates, and L Matrix. The lme4 package. *R package version*, 2(1):74, 2007.

Diane E Bild, David A Bluemke, Gregory L Burke, Robert Detrano, Ana V Diez Roux, Aaron R Folsom, Philip Greenland, David R JacobsJr, Richard Kronmal, Kiang Liu, et al. Multi-ethnic study of atherosclerosis: objectives and design. *American journal of epidemiology*, 156(9):871–881, 2002.

Jennifer F. Bobb and Ravi Varadhan. *turboEM: A Suite of Convergence Acceleration Schemes for EM, MM and Other Fixed-Point Algorithms*, 2021. URL `https://CRAN.R-project.org/package=turboEM`. R package version 2021.1.

Anne M Butler, Katelin B Nickel, Robert A Overman, and M Alan Brookhart. Ibm marketscan research databases. In *Databases for Pharmacoepidemiological Research*, pages 243–251. Springer, 2021.

Antonio Ceriello, Louis Monnier, and David Owens. Glycaemic variability in diabetes: clinical and therapeutic implications. *The lancet Diabetes & endocrinology*, 7(3):221–230, 2019.

Rory Collins. What makes uk biobank special? *The Lancet*, 9822(379):1173–1174, 2012.

Michael J Crowther, Keith R Abrams, and Paul C Lambert. Joint modeling of longitudinal and survival data. *The Stata Journal*, 13(1):165–184, 2013.

Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.

William Duckworth, Carlos Abraira, Thomas Moritz, Domenic Reda, Nicholas Emanuele, Peter D Reaven, Franklin J Zieve, Jennifer Marks, Stephen N Davis, Rodney Hayward, et al. Glucose control and vascular complications in veterans with type 2 diabetes. *New England Journal of Medicine*, 360(2):129–139, 2009.

Eldin Dzubur, Aditya Ponnada, Rachel Nordgren, Chih-Hsiang Yang, Stephen Intille, Genevieve Dunton, and Donald Hedeker. Mixwild: A program for examining the effects of variance and slope of time-varying variables in intensive longitudinal data. *Behavior research methods*, 52(4):1403–1427, 2020.

Robert Elashoff, Ning Li, et al. *Joint modeling of longitudinal and time-to-event data*. CRC Press, 2016.

Robert M Elashoff, Gang Li, and Ning Li. A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics*, 64(3):762–771, 2008.

Robert M. Elashoff, Gang Li, and Ning Li. *Joint Modeling of Longitudinal and Time-to-Event Data*, volume 151 of *Monographs on Statistics and Applied Probability*. CRC Press, Boca Raton, FL, 2017. ISBN 978-1-4398-0782-8.

Francisca Galindo Garre, Aeilko H Zwinderman, Ronald B Geskus, and Yvo WJ Sijpkens. A joint latent class changepoint model to improve the prediction of time to graft failure. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(1):299–308, 2008.

John Michael Gaziano, John Concato, Mary Brophy, Louis Fiore, Saiju Pyarajan, James Breeling, Stacey Whitbourne, Jennifer Deen, Colleen Shannon, Donald Humphries, Peter Guarino, Mihaela Aslan, Daniel Anderson, Rene LaFleur, Timothy Hammond, Kendra Schaa, Jennifer Moser, Grant Huang, Sumitra Muralidhar, Ronald Przygodzki, and Timothy J. O'Leary. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *Journal of Clinical Epidemiology*, 70:214–223, 2016.

Christopher A German, Janet S Sinsheimer, Jin Zhou, and Hua Zhou. Wiser: Robust and scalable estimation and inference of within-subject variances from intensive longitudinal data. *Biometrics*, 2021.

Els Goetghebeur and Louise Ryan. Semiparametric regression analysis of interval-censored data. *Biometrics*, 56(4):1139–1144, 2000.

Guadalupe Gómez, Anna Espinal, and Stephen W. Lagakos. Inference for a linear regression model with an interval-censored covariate. *Statistics in medicine*, 22(3):409–425, 2003.

Guadalupe Gómez, M Luz Calle, Ramon Oller, and Klaus Langohr. Tutorial on methods for interval-censored data and their implementation in r. *Statistical Modelling*, 9(4):259–297, 2009.

ACCORD Study Group. Effects of intensive glucose lowering in type 2 diabetes. *New England journal of medicine*, 358(24):2545–2559, 2008.

UK Prospective Diabetes Study Group et al. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (ukpds 33). *The Lancet*, 352(9131):837–853, 1998.

Donald Hedeker, Robin J Mermelstein, and Hakan Demirtas. An application of a mixed-effects location scale model for analysis of ecological momentary assessment (ema) data. *Biometrics*, 64(2):627–634, 2008.

Robin Henderson, Peter Diggle, and Angela Dobson. Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(4):465–480, 2000.

Graeme Hickey, Pete Philipson, Andrea Jorgensen, and Ruwanthi Kolamunnage-Dona. 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, 2018.

Jian Huang and Jon A Wellner. Interval censored survival data: a review of recent progress. In *Proceedings of the First Seattle Symposium in Biostatistics*, pages 123–169. Springer, 1997.

Xin Huang, Gang Li, Robert M Elashoff, and Jianxin Pan. A general joint model for longitudinal measurements and competing risks survival data with heterogeneous random effects. *Lifetime data analysis*, 17(1):80–100, 2011.

Faramarz Ismail-Beigi, Timothy Craven, Mary Ann Banerji, Jan Basile, Jorge Calles, Robert M Cohen, Robert Cuddihy, William C Cushman, Saul Genuth, Richard H Grimm Jr, et al. Effect of intensive treatment of hyperglycaemia on microvascular outcomes in type 2 diabetes: an analysis of the accord randomised trial. *The Lancet*, 376 (9739):419–430, 2010.

Do Hyun Kim, Aubrey Jensen, Kelly Jones, Sridharan Raghavan, Lawrence S Phillips, Adriana Hung, Yan V Sun, Gang Li, Peter Reaven, Hua Zhou, et al. A platform for phenotyping disease progression and associated longitudinal risk factors in large-scale ehrs, with application to incident diabetes complications in the uk biobank. *JAMIA open*, 6(1):ooad006, 2023.

Klaus Langohr and Guadalupe Gómez Melis. Estimation and residual analysis with r for a linear regression model with an interval-censored covariate. *Biometrical Journal*, 56(5): 867–885, 2014.

Jerald F Lawless and Denise Babineau. Models for interval censoring and simulation-based inference for lifetime distributions. *Biometrika*, 93(3):671–686, 2006.

Shanpeng Li, Ning Li, Hong Wang, Jin Zhou, Hua Zhou, and Gang Li. Efficient algorithms and implementation of a semiparametric joint model for longitudinal and competing risk data: With applications to massive biobank data. *Computational and mathematical methods in medicine*, 2022, 2022a.

Shanpeng Li, Ning Li, Hong Wang, Jin Zhou, Hua Zhou, and Gang Li. *FastJM: Semi-Parametric Joint Modeling of Longitudinal and Survival Data*, 2022b. URL `https://CRAN.R-project.org/package=FastJM`. R package version 1.0.1.

Shanpeng Li, Robyn McClelland, Peter D Reaven, Jin Zhou, Hua Zhou, and Gang Li. A joint model of the individual mean and within-subject variability of a longitudinal outcome with a competing risks time-to-event outcome. *arXiv preprint arXiv:2301.06584*, 2023.

Haiqun Lin, Charles E McCulloch, and Robert A Rosenheck. Latent pattern mixture models for informative intermittent missing data in longitudinal studies. *Biometrics*, 60(2):295–305, 2004.

Doug Morrison, Oliver Laeyendecker, and Ron Brookmeyer. Regression with interval-censored covariates: Application to cross-sectional incidence estimation. *Biometrics*, 2021.

Søren Feodor Nielsen. The stochastic em algorithm: estimation and asymptotic results. *Bernoulli*, 6(3):457–489, 2000.

Daniel S. Nuyujukian, Juraj Koska, Gideon Bahn, Peter D. Reaven, and Jin J. Zhou. Blood

pressure variability and risk of heart failure in ACCORD and the VADT. *Diabetes Care*, 43(7):1471–1478, 2020.

Daniel S Nuyujukian, Jin J Zhou, Juraj Koska, and Peter D Reaven. Refining determinants of associations of visit-to-visit blood pressure variability with cardiovascular risk: results from the action to control cardiovascular risk in diabetes trial. *Journal of Hypertension*, 39(11):2173–2182, 2021.

George T O'Connor, David Sparrow, and Scott T Weiss. A prospective longitudinal study of methacholine airway responsiveness as a predictor of pulmonary-function decline: the normative aging study. *American journal of respiratory and critical care medicine*, 152 (1):87–92, 1995.

Art B Owen. Monte carlo theory, methods and examples. 2013.

Grigorios Papageorgiou, Katya Mauff, Anirudh Tomer, and Dimitris Rizopoulos. An overview of joint modeling of time-to-event and longitudinal outcomes. *Annual review of statistics and its application*, 6:223–240, 2019.

Richard MA Parker, George Leckie, Harvey Goldstein, Laura D Howe, Jon Heron, Alun D Hughes, David M Phillippo, and Kate Tilling. Joint modeling of individual trajectories, within-individual variability, and a later outcome: systolic blood pressure through childhood and left ventricular mass in early adulthood. *American journal of epidemiology*, 190 (4):652–662, 2021.

Pete Philipson, Ines Sousa, Peter J. Diggle, Paula Williamson, Ruwanthi Kolamunnage-Dona, Robin Henderson, and Graeme L. Hickey. *joineR: Joint Modelling of Repeated Measurements and Time-to-Event Data*, 2018. URL `https://github.com/graemeleehickey/joineR/`. R package version 1.2.5.

William H Press, Saul A Teukolsky, William T Vetterling, and Brian P Flannery. *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge university press, 2007.

Cécile Proust-Lima, Viviane Philipps, and Benoit Liquet. Estimation of extended mixed models using latent classes and latent processes: the r package lcmm. *arXiv preprint arXiv:1503.00890*, 2015.

Peter D Reaven, Nicholas V Emanuele, Wyndy L Wiitala, Gideon D Bahn, Domenic J Reda, Madeline McCarren, William C Duckworth, and Rodney A Hayward. Intensive glucose control in patients with type 2 diabetes?15-year follow-up. *New England Journal of Medicine*, 380(23):2215–2224, 2019.

Dimitris Rizopoulos. JM: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software*, 35(9):1–33, 2010.

Dimitris Rizopoulos. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, 67(3):819–829, 2011.

Dimitris Rizopoulos. Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive gaussian quadrature rule. *Computational Statistics & Data Analysis*, 56(3):491–501, 2012a.

Dimitris Rizopoulos. *Joint Models For Longitudinal and Time-to-Event Data: With Applications in R*. CRC press, 2012b.

Dimitris Rizopoulos. The r package jmbayes for fitting joint models for longitudinal and time-to-event data using mcmc. *arXiv preprint arXiv:1404.7625*, 2014.

Peter M Rothwell, Sally C Howard, Eamon Dolan, Eoin O'Brien, Joanna E Dobson, Bjorn Dahlöf, Peter S Sever, and Neil R Poulter. Prognostic significance of visit-to-visit variability, maximum systolic blood pressure, and episodic hypertension. *The Lancet*, 375(9718): 895–905, 2010.

Abdus Sattar and Sanjoy K Sinha. 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, 2019.

Xiao Song, Marie Davidian, and Anastasios A. Tsiatis. A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics*, 58(4):742–753, 2002.

Maria Sudell, Ruwanthi Kolamunnage-Dona, and Catrin Tudur-Smith. Joint models for longitudinal and time-to-event data: a review of reporting quality with a view to meta-analysis. *BMC medical research methodology*, 16(1):1–11, 2016.

Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, Paul Elliott, Jane Green, Martin Landray, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine*, 12(3):e1001779, 2015.

Jianguo Sun. *The statistical analysis of interval-censored failure time data*, volume 3. Springer, 2006.

Xing Sun, Xiaoyun Li, Cong Chen, and Yang Song. A review of statistical issues with progression-free survival as an interval-censored time-to-event endpoint. *Journal of Biopharmaceutical Statistics*, 23(5):986–1003, 2013.

Donald P Tashkin, He-Jing Wang, David Halpin, Eric C Kleerup, John Connett, Ning Li, and Robert Elashoff. Comparison of the variability of the annual rates of change in FEV1 determined from serial measurements of the pre-versus post-bronchodilator FEV1 over 5 years in mild to moderate COPD: Results of the lung health study. *Respiratory research*, 13(1):1–10, 2012.

A.A. Tsiatis and M. Davidian. Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, 14(3):809–834, 2004.

Bruce W Turnbull. The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 38(3):290–295, 1976.

C. Y. Wang. Corrected score estimator for joint modeling of longitudinal and failure time data. *Statistica Sinica*, 16(1):235–253, 2006.

Lianming Wang, Christopher S McMahan, Michael G Hudgens, and Zaina P Qureshi. A flexible, computationally efficient method for fitting the proportional hazards model to interval-censored data. *Biometrics*, 72(1):222–231, 2016.

Lang Wu, Wei Liu, Grace Y Yi, and Yangxin Huang. Analysis of longitudinal and survival data: joint modeling, inference methods, and issues. *Journal of Probability and Statistics*, 2012, 2012.

Cong Xu, Pantelis Z Hadjipantelis, and Jane-Ling Wang. Semi-parametric joint modeling of survival and longitudinal data: The r package jsm. *Journal of Statistical Software*, 93(2), 2020.

Menggang Yu, Jeremy M. G Taylor, and Howard M Sandler. Individual prediction in prostate cancer studies using a joint longitudinal survival–cure model. *Journal of the American Statistical Association*, 103(481):178–187, 2008.

Donglin Zeng and Jianwen Cai. Simultaneous modelling of survival and longitudinal data with an application to repeated quality of life measures. *Lifetime Data Analysis*, 11: 151–174, 2005.

Donglin Zeng, Jianwen Cai, et al. Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time. *The Annals of Statistics*, 33 (5):2132–2163, 2005.

Donglin Zeng, Lu Mao, and DY Lin. Maximum likelihood estimation for semiparametric transformation models with interval-censored data. *Biometrika*, 103(2):253–271, 2016.

Zhigang Zhang and Yichuan Zhao. Empirical likelihood for linear transformation models with interval-censored failure time data. *Journal of Multivariate Analysis*, 116:398–409, 2013.

Jin J Zhou, Dawn C Schwenke, Gideon Bahn, and Peter Reaven. Glycemic variation and cardiovascular risk in the veterans affairs diabetes trial. *Diabetes Care*, 41(10):2187–2194, 2018.

Jin J Zhou, Ruth Coleman, Rury R Holman, and Peter Reaven. Long-term glucose variability and risk of nephropathy complication in UKPDS, ACCORD and VADT trials. *Diabetologia*, 63(11):2482–2485, 2020.

Jin J Zhou, Juraj Koska, Gideon Bahn, Peter Reaven, VADT Investigators, et al. Fasting glucose variation predicts microvascular risk in ACCORD and VADT. *The Journal of Clinical Endocrinology & Metabolism*, 106:1150–1162, 2021.