

Statistical Methods for Joint Modeling of Disease Processes under Intermittent Observation

by

Jianchu Chen

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics

Waterloo, Ontario, Canada, 2024

© Jianchu Chen 2024

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Lang Wu
 Professor,
 Department of Statistics,
 University of British Columbia

Supervisor(s): Richard J. Cook
 University Professor, Mathematics Faculty Research Chair,
 Department of Statistics and Actuarial Science,
 University of Waterloo

Internal Members: Cecilia Cotton
 Associate Professor,
 Department of Statistics and Actuarial Science,
 University of Waterloo

 Joel Dubin
 Professor,
 Department of Statistics and Actuarial Science,
 University of Waterloo

Internal-External Member: JM Gamble
 Clinical Associate Professor,
 School of Pharmacy,
 University of Waterloo

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

In studies of life history data, individuals often experience multiple events of interest that may be associated with one another. In such settings, joint models of event processes are essential for valid inferences. Data used for statistical inference are typically obtained through various sources, including observational data from registries or clinics and administrative records. These observation processes frequently result in incomplete histories of the event processes of interest. In settings where interest lies in the development of conditions or complications that are not self-evident, data become available only at periodic clinic visits. This thesis focuses on developing statistical methods for the joint analysis of disease processes involving incomplete data due to intermittent observation.

Many disease processes involve recurrent adverse events and an event which terminates the process. Death, for example, terminates the event process of interest and precludes the occurrence of further events. In Chapter 2, we present a joint model for such processes which has appealing properties due to its construction using copula functions. Covariates have a multiplicative effect on the recurrent event intensity function given a random effect, which is in turn associated with the failure time through a copula function. This permits dependence modeling while retaining a marginal Cox model for the terminal event process. When these processes are subject to right-censoring, simultaneous and two-stage estimation strategies are developed based on the observed data likelihood, which can be implemented by direct maximization or via an expectation-maximization algorithm - the latter facilitates semi-parametric modeling for the terminal event process. Variance estimates are derived based on the missing information principle. Simulation studies demonstrate good finite sample performance of proposed methods and high efficiency of the two-stage procedure. An application to a study of effect of pamidronate on reducing skeletal complications in patient with skeletal metastases illustrates the use of this model.

Interval-censored recurrent event data can occur when the events of interest are only evident through intermittent clinical examination. Chapter 3 addresses such scenarios and extends the copula-based joint model for recurrent and terminal events proposed in Chapter 2 to accommodate interval-censored recurrent event data resulting from intermittent observation. Conditional on a random effect, the intensity for the recurrent event process has a multiplicative form with a weak parametric piecewise constant baseline rate, and a Cox model is formulated for the terminal event process. The two processes are then linked via a copula function, which defines a joint model for the random effect and the terminal event. The observed data likelihood can be maximized directly or via an EM algorithm; the latter facilitates a semi-parametric terminal event process. A computationally convenient two-stage estimation procedure is also investigated. Variance estimates are derived

and validated by simulation studies. We apply this method to investigate the association between a biomarker (HLA-B27) and joint damage in patients with psoriatic arthritis.

Databases of electronic medical records offer an unprecedented opportunity to study chronic disease processes. In survival analysis, interest may lie in studying the effects of time-dependent biomarkers on a failure time through Cox regression models. Often however, it is too labour intensive to collect and clean data on all covariates at all times, and in such settings it is common to select a single clinic visit at which variables are measured. In Chapter 4, we consider several cost-effective ad hoc strategies for inference, consisting of: 1) selecting either the last or the first visit for a measurement of the marker value, and 2) using the measured value with or without left-truncation. The asymptotic bias of estimators based on these strategies arising from misspecified Cox models is investigated via a multistate model constructed for the joint modeling of the marker and failure processes. An alternative selection method for efficient selection of individuals is discussed under budgetary constraint, and the corresponding observed data likelihood is derived. The asymptotic relative efficiency of regression coefficients obtained from Fisher information is explored and an optimal design is provided under this selection scheme.

The contributions of this thesis are reviewed in Chapter 5 where we also discuss topics of future research.

Acknowledgements

I would like to begin by expressing my deepest gratitude to my supervisor, Dr. Richard Cook. His unwavering support, guidance, and wisdom have been invaluable to me throughout my academic journey. I have learned so much from him, not only about my research but also about the qualities of a thoughtful and dedicated scholar. I feel truly fortunate to have him as my supervisor and this thesis would not have been possible without his care and support.

I would also like to extend my sincere thanks to Dr. Cecilia Cotton, Dr. Joel Dubin, Dr. JM Gamble, and Dr. Lang Wu for serving as my committee members and for offering guidance.

I am especially grateful to the faculty and staff in the Department of Statistics and Actuarial Science at the University of Waterloo. Over the past six years, I have had the privilege of meeting many inspiring people here. Their efforts in creating a positive and friendly environment have greatly supported my development as a graduate student. I've also made so many friends during this journey, and their companionship has made this experience truly memorable.

Finally, I would like to express my deepest appreciation to my parents, whose love, encouragement, and sacrifices have made this journey possible. Their unwavering belief in me has been my constant source of strength, and I am forever grateful for their support.

Dedication

To my family

Table of Contents

Examining Committee	ii
Author's Declaration	iii
Abstract	iv
Acknowledgements	vi
Dedication	vii
List of Figures	ix
List of Tables	x
1 Introduction	1
1.1 Overview	1
1.2 Introduction to Statistical Methods	2
1.2.1 Multivariate Distributions via Copula Models	2
1.2.2 Analysis of Recurrent Events	3
1.2.3 Multistate Models	4
1.3 Motivating Studies	5
1.3.1 Skeletal Complications in Patients with Bone Metastases from Breast Cancer	5

1.3.2	The University of Toronto Psoriatic Arthritis Cohort Study	6
1.4	Outline of the Thesis	7
2	A Copula-Based Joint Model for Recurrent and Terminal Events	9
2.1	Introduction	9
2.2	A Copula-Based Joint Model	11
2.2.1	Model Formulation	11
2.2.2	Implied Rates and Intensities	13
2.2.3	Parameter Estimation and Inference	19
2.3	A Two-Stage Procedure	21
2.3.1	Estimation and Inference via Direct Optimization	21
2.3.2	Two-Stage Estimation with the EM Algorithm	22
2.4	Simulation Studies	25
2.4.1	Parametric Analysis with Cox Piecewise Constant Baseline Hazard Models	26
2.4.2	Semiparametric Analysis with a Cox Model for the Terminal Event	29
2.5	Efficacy of Pamidronate in Reducing Skeletal Complications	31
2.6	Discussion	33
3	A Joint Model for Recurrent and Terminal Events with Intermittent Observation	37
3.1	Introduction	37
3.2	A Copula-Based Joint Model	39
3.2.1	Model Framework	39
3.2.2	A Piecewise Constant Conditional Baseline Rate	41
3.2.3	EM Algorithm and Complete Data Likelihood	43
3.3	A Two-Stage Procedure	46
3.3.1	Two-Stage Estimation via Direct Optimization	46
3.3.2	Two-Stage Estimation via an EM Algorithm	48

3.4	Simulation Studies	48
3.4.1	Simulation Set-up	49
3.4.2	Parametric Analyses with Piecewise Constant Models	49
3.4.3	Analyses with a Cox Model for the Terminal Event	53
3.5	Joint Damage in Psoriatic Arthritis	53
3.6	Discussion	57
4	Cost-Effective Design of Survival Studies Involving Intermittent Observation of Time-Dependent Covariates	60
4.1	Introduction	60
4.2	Joint Modeling Framework	62
4.2.1	A Joint Marker-Failure Process	62
4.2.2	The Observation Processes	64
4.2.3	Observable Event Processes	66
4.3	Asymptotic Bias from Naive Cost-Effective Strategies	67
4.3.1	Cost-Effective Ad Hoc Strategies Using Cox Models	67
4.3.2	Calculations of Expectations	72
4.3.3	Investigating Determinants of the Limiting Value ψ^*	75
4.4	Cost-Effective Design of Selection Schemes	77
4.5	Illustrative Application to Psoriatic Arthritis Cohort Data	82
4.6	Discussion	86
5	Discussion and Future Work	87
5.1	Discussion	87
5.2	Future Work	89
5.2.1	Future Research on Chapters 2 and 3	89
5.2.2	Future Research on Chapter 4	90
	References	92

APPENDICES	99
A Appendix to Chapter 2	100
A.1 Joint Recurrent-Terminal Event Model via a Gaussian Copula	100
A.2 Calculations of Some Expectation Terms under a Gaussian Copula	102
A.3 Data Generation Procedure for Simulation Studies	104
A.4 Parametric Multiplicative Models with Piecewise Constant Baseline Functions	105
A.4.1 Complete Data Score	107
A.4.2 Complete Data Information	110
A.5 A Semiparametric Cox Model for the Terminal Event	115
A.5.1 Complete Data Score	116
A.5.2 Complete Data Information	117
A.5.3 Two-Stage Score and Information	118
B Appendix to Chapter 3	120
B.1 Calculations for the EM Algorithm	120
B.2 Data Generation for Simulation Studies	122
B.3 Proportional Hazards Models with Piecewise Constant Baseline Hazards .	123
B.3.1 Complete Data Score	124
B.3.2 Complete Data Information	125
B.4 A Cox Model for The Terminal Event	126
B.4.1 The Complete Data Score	126
B.4.2 Complete Data Information	126
B.4.3 Two-stage Score and Information	127
C Appendix to Chapter 4	128
C.1 Computation of Some Conditional Expectations	128
C.1.1 Measuring X_1° at the Last Visit	128
C.1.2 Measuring X_1° at the Last Visit with Left-Truncation	130
C.1.3 Measuring X_1° at the First Visit	132
C.1.4 Measuring X_1° at the First Visit with Left-Truncation	134

List of Figures

2.1	A multistate diagram illustrating the recurrent and terminal event processes	16
2.2	Four illustrative sample paths to investigate intensities for recurrent and terminal events implied by the model specified in Section 2.2.1.2	17
2.3	Plots of the complete intensity for the recurrent event process ($E\{dN_{i1}(s) \mathcal{H}_i(s)\}$) and the complete intensity for the terminal event process ($E\{dN_{i2}(s) \mathcal{H}_i(s)\}$) versus s for $X_i = 0$, for $\phi_1 = 0.2$ (left panels) and $\phi_1 = 0.4$ (right panels); the four rows correspond to the four illustrative paths in Figure 2.2	18
2.4	Sample profile features for various skeletal complications and survival times in the Pamidronate and placebo groups; green circles represent skeletal complications and red triangles represent observed deaths	32
2.5	Plots of estimated conditional (on being alive and $u_{i1} = 1$) and marginal cumulative rates for the SRE process (left panel), and plots of estimated cumulative hazard and survival function for the survival process (right panel)	34
3.1	Selected sample profiles of individuals recruited to the PsA cohort	56
3.2	Plots of estimated conditional (on being alive and $u_{i1} = 1$) and marginal cumulative rates for the joint damage process (left panel), and plots of estimated cumulative hazard and survival function for the survival process (right panel)	58
4.1	A joint marker-failure process model for a binary marker and a single absorbing state entered upon failure	63
4.2	A four-state process obtained by incorporating a random censoring state to the joint marker-failure process	65
4.3	A joint model for a four-state marker-failure-censoring process and a visit process	68

4.4	An illustration of the discordance between the observed value $X_1^\circ(s)$ and the true value $X_1(s)$ when marker values are carried forward under intermittent observation schemes	68
4.5	An illustration of the discordance between X_1° and the true value $X_1(s)$ based on four <i>ad hoc</i> strategies: a) choosing the last visit to measure the marker value and carry it both forward and backward; b) choosing the last visit to measure the marker value and only carry it forward via left-truncation; c) choosing the first visit to measure the marker value and carry it both forward and backward; d) choosing the last visit to measure the marker value and only carry it forward via left-truncation;	70
4.6	A multistate model of $\mathcal{Z}(s) = (\bar{A}(s), Z(s), X_1^\circ(s))$ depicting a more general joint model for the marker-failure process and the observations processes	73
4.7	Relative asymptotic biases of $\hat{\psi}_1$ (left panel) and $\hat{\psi}_2$ as a function of the baseline visit intensity ρ_0 with an independent visit process under four <i>ad hoc</i> strategies; the solids dots represent estimators averaged over 500 simulations with a sample size of 1000.	78
4.8	Relative asymptotic biases of $\hat{\psi}_1$ (left panel) and $\hat{\psi}_2$ as a function of the marker effect on the visit intensity (i.e., η_1) for $\mu = 4, 8, 12$ with a marker-dependent visit process under four <i>ad hoc</i> strategies	79
4.9	Relative asymptotic efficiency of $\hat{\psi}_1$ (left panel) and $\hat{\psi}_2$ as a function of $E\{m_L\}/\mu_b$ for $\pi_2 = 0.1, 0.25, 0.5, 0.8$, $\mu_b/n = 0.5, 1$ and $\bar{\mu} = 4, 8, 12$; the asymptotic variances are obtained from the expected information matrix $\mathcal{I}(\theta)$ with the expectation carried out by Monte-Carlo integration with 500k replications	83

List of Tables

2.1	Parameter specifications for each parameter setting	26
2.2	Tabulation of frequency properties of estimators obtained from a fully parametric specification with piecewise constant baseline hazards for both processes under no random censoring (just administrative censoring at 1; estimates are obtained via direct optimization of the observed data likelihood with simultaneous estimation ; $m = 1000$ and $n_{\text{sim}} = 500$)	27
2.3	Tabulation of frequency properties of estimators obtained from a fully parametric specification with piecewise constant baseline hazards for both processes under no random censoring (just administrative censoring at 1; estimates are obtained via direct optimization of the observed data likelihood with two-stage estimation ; $m = 1000$ and $n_{\text{sim}} = 500$)	28
2.4	Tabulation of frequency properties of estimators from either a parametric (with an 8-piece piecewise constant baseline) or a semiparametric recurrent event process, and a semiparametric terminal event process with administrative censoring at 1 (no random censoring); estimates are obtained using a two-stage EM algorithm; $m = 1000$ and $n_{\text{sim}} = 500$	30
2.5	Joint analysis of the SRE and the survival process with a Gaussian copula, a Gamma distributed random effect and a single binary covariate indicating receipt of pamidronate rather than placebo; both parametric (with 8-piece PWC baseline rates) and semiparametric models are considered (all estimates are obtained via a two-stage procedure while the semiparametric estimation exploits the EM algorithm; BSE is the bootstrap standard error of the estimates obtained from 100 bootstrap runs)	35
3.1	True parameter values for each scenario	50

3.2	Tabulation of frequency properties of estimators obtained from a fully parametric specification with an 8-piece piecewise constant baseline hazards for both processes under no random censoring for $E(A_i(1) X_i) = 4$ (just administrative censoring at 1; estimates are obtained via direct optimization of the observed data likelihood with simultaneous estimation ; $m = 1000$ and $n_{\text{sim}} = 500$) . . .	51
3.3	Tabulation of frequency properties of estimators obtained from a fully parametric specification with an 8-piece piecewise constant baseline hazards for both processes under no random censoring for $E(A_i(1) X_i) = 4$ (just administrative censoring at 1; estimates are obtained via direct optimization of the observed data likelihood with two-stage estimation ; $m = 1000$ and $n_{\text{sim}} = 500$)	52
3.4	Tabulation of frequency properties of estimators obtained from a semiparametric specification with a Cox model for the terminal event under no random censoring for $E(A_i(1) X_i) = 4$ (just administrative censoring at 1; point estimates are obtained via direct optimization of the observed data likelihood via two-stage estimation ; variance estimates are obtained through Louis formula with an EM algorithm; $m = 1000$ and $n_{\text{sim}} = 500$)	54
3.5	Joint analysis of the joint damage and the survival process with a Gaussian copula, a Gamma distributed random effect and a single covariate of HLA-B27; both parametric (with an 8-piece PWC baseline rate for death; estimates are obtained through direct optimization) and semiparametric models (estimates are obtained via a two-stage procedure with variance estimates derived through an EM algorithm) are considered	57
4.1	Joint analyses of the ESR_CRP and failure processes with a fixed covariate of HLA-B27 using subsamples selected under different specifications of μ_b and $E\{m_L\}/\mu_b$; time-homogeneous baseline transition intensities are fitted; variance estimates are obtained through observed information matrix averaged over 500 repetitions	85

Chapter 1

Introduction

1.1 Overview

Life history data pertain to the events and conditions that individuals experience over their lifetime. In some settings, events of interest are transient and can occur multiple times, making it meaningful to model the event occurrence. In other settings, we may characterize complex life history processes, such as chronic disease progression, into different states where events are more naturally viewed as representing transitions in the status of an individual. Recurrent event analysis (Cook and Lawless, 2007) and multistate models (Klein et al., 2014; Cook and Lawless, 2018) are particularly useful for such settings where interests lie in investigating the association between fixed or time-varying risk factors and event occurrence or state transitions. Often the time individuals experience multiple events of interest, and these event processes may be associated with each other. Death, for example, will terminate the disease process of interest and preclude the occurrence of further events. In such settings, joint models of the disease process incorporating death are necessary for valid inferences. A common example is the illness-death model (Andersen, 1988). Another example of joint modeling in life history data is the joint analysis of survival data and time-varying covariates. A standard modeling method is to specify a linear mixed-effects model for the time-dependent covariates and a Cox proportional hazards model for the survival data, and link the two models via shared random effects (Wu et al., 2012).

Datasets used for statistical inference are typically obtained through various sources, including longitudinal surveys and cohort studies, observational data from registries or clinics, and administrative records. Such observation processes often result in incomplete histories of the disease process of interest. Common observation schemes include censoring

or termination of the observation period due to the end of study or loss to follow-up, and intermittent measurement times at which data are collected. In settings where interest lies in the development of conditions or complications that are not self-evident, data become available at periodic clinic visits, giving rise to so-called panel data (Kalbfleisch and Lawless, 1985) where transition times are subject to interval censoring. In multistate settings with cyclical disease processes it may even be unknown how many transitions occurred between successive assessment times. This limits the scope of models that can be fitted but some (e.g. Markov processes) can, in principle, be fitted in such settings. Standard methods of life history analysis depend on independence and ignorability assumptions of the observation process (Farewell et al., 2017; Kalbfleisch and Prentice, 2011; Andersen et al., 2012; Aalen et al., 2008). It is crucial to understand the dependence between the disease process and the observation process for valid inferences from such data through statistical modeling. Cook and Lawless (2021) provide a thorough discussion of the independence conditions of the observation process in the analysis of life history data.

This thesis is concerned with the development of statistical methods for the joint analysis of disease processes involving incomplete data due to such intermittent observation. In Chapter 2 and 3, we focus on joint modeling of a recurrent event and a terminal event, accommodating the framework for incomplete recurrent event data due to intermittent observation. In Chapter 4, we focus on joint modeling of a time-varying covariate and a failure process, where the time-varying covariate can only be measured at intermittent clinical visits. The topics of the projects in this thesis are discussed in more detail in Section 1.4. A brief overview of the statistical methods covered in this thesis is presented in Section 1.2, and the motivating studies are described in Section 1.3.

1.2 Introduction to Statistical Methods

1.2.1 Multivariate Distributions via Copula Models

Let X_k be a continuous random variable with a continuous cumulative distribution function $F_k(x_k)$, $k = 1, \dots, K$, the joint distribution function of $X = (X_1, X_2, \dots, X_K)'$, denoted by

$$F(x) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_K \leq x_K),$$

can be represented by a K-dimensional copula function $\mathcal{C} : [0, 1]^K \rightarrow [0, 1]$ (Sklar, 1959; Joe, 1997) indexed by a vector of dependence parameters ρ , such that

$$F(x) = \mathcal{C}(F_1(x_1), \dots, F_K(x_K); \rho).$$

If we let $U_k = F_k(X_k)$, $k = 1, \dots, K$, and $U = (U_1, \dots, U_K)$, U_k is a uniform random variable over $[0, 1]$, $k = 1, \dots, K$, and we have

$$\mathcal{C}(u; \rho) = F(F_1^{-1}(u_1), \dots, F_K^{-1}(u_K); \rho)$$

with quantile functions F_k^{-1} for $k = 1, \dots, K$. There are a variety of ways to measure the dependence, one is known as Kendall's τ (Kendall, 1938). For each pair of random variables, Kendall's τ is defined as the probability of concordance minus the probability of discordance (Nelsen, 2006), with $\tau = 0$ signifying an independence and $\tau = 1$ signifying a perfect association. Kendall's τ is a common association measure in copula models due to its functional independence from marginal parameters.

Well-studied copula functions can be generally classified into elliptical copulas, Archimedean copulas and vine copulas. Archimedean copulas are popular because they allow modeling dependence in arbitrarily high dimensions with only one parameter, while governing the strength of dependence. Clayton copulas (Clayton, 1978) and Frank copulas (Frank, 1979) are examples of the Archimedean copula family. Other well-known copulas, such as Gaussian copulas and t-copulas, belong to the elliptical copula family. Vine copulas are constructed from a sequence of bivariate copulas organized in a hierarchical structure, allowing for flexible modeling of complex dependencies in high dimensions. One appealing feature of copula models is that the dependence structure induced by them does not depend on the marginal models but is solely characterized by the copula function itself. Consequently the marginal models can be constructed in a flexible manner.

1.2.2 Analysis of Recurrent Events

In many medical conditions, an event of interest may occur multiple times for the same subject during the period of follow-up. Such processes are referred to as recurrent event processes, and the data consisting of information on the events and covariates over time are called recurrent event data. Various statistical methods have been developed for analyzing recurrent event data, including intensity-based models, random effect models and marginal methods; see Cook and Lawless (2007).

Intensity functions and counting processes are very useful for modeling recurrent events. We detail intensity-based methods here, as they form the foundation for the projects discussed in this thesis. Suppose the recurrent event process starts at $t = 0$, and $T_1 < T_2 < \dots$, where T_j is the time of the j th event. Let $N(t) = \sum_{i=1}^{\infty} \mathbb{1}(T_i \leq t)$ denote the number of events occurred over $(0, t]$, $N(s, t) = N(t) - N(s)$ denote the number of event occurrences over $(s, t]$, and $\{N(t), t \geq 0\}$ denote the corresponding counting process. The event rate

function is the instantaneous probability of an event occurring at time t , conditional on the process history, which is given by

$$\lambda(t | H(t)) = \lim_{\Delta t \rightarrow 0} \frac{P(\Delta N(t) = 1 | H(t))}{\Delta t}$$

where $H(t) = \{N(s), 0 < s \leq t\}$ is the event history up to time t , and $\Delta N(t) = N(t + \Delta t) - N(t)$. If the individual is observed over $(0, \tau]$, the joint probability of observing $m = N(\tau)$ events at times $t_1, \dots, t_m$ can be written as

$$\prod_{j=1}^m \lambda(t_j | H(t_j)) \exp \left\{ - \int_0^\tau \lambda(s | H(s)) ds \right\}.$$

Specifications of the event rate function $\lambda(t | H(t))$ is quite flexible but Markov or semi-Markov assumptions are perhaps most often adopted. Poisson models are also commonly employed for the event rate function so that $\lambda(t | H(t)) = \lambda(t)$. Intensity-based regression models with multiplicative covariate effects are often used in life history analysis due to the appealing nature of the interpretation of covariate effects and the fact that such models can be fitted by exploiting software for Cox regression analyses.

1.2.3 Multistate Models

Multistate models are a natural way to characterize chronic disease processes with multiple stages. Let $\{Z(t), t > 0\}$ denote a continuous stochastic process with state space $\mathcal{S} = \{0, 1, \dots, K\}$. Let $H(t) = \{Z(s), 0 < s \leq t\}$ be the full history of the multistate process, we can define the transition intensity from state k to state l as

$$\lambda_{kl}(t | H(t^-)) = \lim_{\Delta t \rightarrow 0} \frac{P(Z(t + \Delta t) = l | Z(t^-) = k, H(t^-))}{\Delta t}$$

for $k \neq l$ and $k, l \in \mathcal{S}$. Under a Markov model assumption, the transition intensity can be simplified to $\lambda_{kl}(t | H(t^-)) = \lambda_{kl}(t)$. The transition probabilities $p_{kl}(s, t) = P(Z(t) = l | Z(s) = k)$, $s < t$, are then obtainable from the intensities via the differential equation (Cox, 2017)

$$\frac{\partial}{\partial t} \mathcal{P}(s, t) = \mathcal{P}(s, t) \mathcal{Q}(t)$$

where $\mathcal{P}(s, t)$ is the transition probability matrix with entries $p_{kl}(s, t)$ and $\mathcal{Q}(t)$ is the transition intensity matrix with off-diagonal entries $\lambda_{kl}(t)$ for $k \neq l$ and diagonal entries

$-\sum_{l \neq k} \lambda_{kl}(t)$. Furthermore, if a time-homogenous model is adopted for the transition intensities, we have $\mathcal{Q}(t) = \mathcal{Q}_0$ and the transition probability matrix can be obtained by matrix exponential as

$$\mathcal{P}(s, t) = \exp\{(t - s)\mathcal{Q}_0\} = \sum_{n=0}^{\infty} \mathcal{Q}_0^n (t - s)^n / n!.$$

A more flexible Markov model can be adopted for the transition intensities by specifying piecewise constant intensity functions. Let $0 = b_0 < b_1 < \dots < b_{K-1} < b_K = \infty$ be a sequence of pre-defined cut-points of the real line, then $\mathcal{Q}(t) = \mathcal{Q}_k$ if $t \in \mathcal{B}_k = [b_{k-1}, b_k)$, $k = 1, \dots, K$. The transition probability matrix $\mathcal{P}(s, t)$ can then be obtained by

$$\begin{aligned} \mathcal{P}(s, t) &= \prod_{k=k_s}^{k_t} \mathcal{P}(\max(s, b_{k-1}), \min(t, b_k)) \\ &= \prod_{k=k_s}^{k_t} \exp\{(\min(t, b_k) - \max(s, b_{k-1})) \mathcal{Q}_k\}. \end{aligned}$$

Under a Markov assumption, multiplicative intensity-based models can be used to characterize the effect of covariates on the dynamics of the disease process by specifying

$$\lambda_{kl}(t | X) = \lambda_{kl0}(t) \exp(X' \beta_{kl})$$

where $\lambda_{kl0}(t)$ is the baseline transition intensity from state k to state l , $X = (X_1, \dots, X_p)'$ is a $p \times 1$ covariate vector, and $\beta_{kl} = (\beta_{kl1}, \dots, \beta_{klp})'$ is a vector of regression coefficients specific to the $k \rightarrow l$ transition.

1.3 Motivating Studies

1.3.1 Skeletal Complications in Patients with Bone Metastases from Breast Cancer

The presence of bone metastases is common among women in advanced stages of breast cancer. The destruction of bone from these lesions is a consequence of osteoclast-induced bone resorption, possibly stimulated by osteoclast-activating factors released by tumor cells (Johnston, 1970; Coleman and Rubens, 1985). Bisphosphonates are powerful inhibitors of osteoclast-induced bone resorption and have efficacy in treating cancer-related

hypercalcemia (Fleisch, 1989; Coleman and Purohit, 1993). Pamidronate disodium, a second-generation bisphosphonate, can inhibit bone resorption without impacting bone mineralization when administered at appropriate doses (Kelliher and Mangino, 1992).

Hortobagyi et al. (1996) conducted a randomized, double-blind study to investigate the efficacy of pamidronate in prevention of skeletal complications in patients with bone lesions due to metastatic breast cancer. Women with stage IV breast cancer who were receiving cytotoxic chemotherapy and had at least one lytic bone lesion were enrolled into the study from January 1991 through March 1994 to receive either placebo or pamidronate infusion monthly for 12 cycles. After recruitment to the study, patients were evaluated for the occurrence of skeletal complications (including pathologic fractures, the need for radiation to bone or bone surgery, spinal cord compression, and hypercalcemia) at each monthly visit, and for survival until the patient’s death, the last date of contact or loss to follow-up, or February 1, 1995, whichever occurred first. Bone pain, use of analgesic drugs, performance status, and quality of life were assessed throughout the trial.

Marginal analyses of the efficacy of pamidronate on the skeletal-related event process and the survival process were conducted by Hortobagyi et al. (1996). However, these analyses did not account for a possible association between the two processes, thereby failing to give insights into the relationship between the two aspects of the disease processes. The challenge of a joint analysis of the occurrence of skeletal complications with death terminating the event process motivates the project discussed in Chapter 2.

1.3.2 The University of Toronto Psoriatic Arthritis Cohort Study

Psoriatic arthritis (PsA) a complex rheumatic disease defined as a chronic inflammatory arthritis associated with psoriasis. Patients with PsA often experience symptoms such as stiffness, swelling, pain and inflammation in the affected joints. The majority of the patients will develop nail lesions, approximately half will develop dactylitis and axial disease, and at least one third will develop enthesitis. Additionally, at least 20% of the patients suffer from a severe, rapidly destructive form of arthritis (Gladman et al., 2005).

The University of Toronto PsA Clinic at the Centre for Prognosis Studies in The Rheumatic Diseases at Toronto Western Hospital maintains a comprehensive database of patients diagnosed with PsA (Gladman and Chandran, 2011). Established in 1978, the PsA Clinic focuses on studying the course and prognosis of PsA patients. Individuals with a prior diagnosis of PsA are recruited to participate in the cohort and are followed prospectively over time to monitor the severity of disease progression. Data collected at the clinic contains retrospective information on disease history, such as age at diagnosis of psoriasis

and PsA, as well as information of disease progression collected during each follow-up visit through clinical and laboratory assessments.

The development of joint damage in PsA patients is of primary interest to clinicians since this damage impairs quality of life and functional ability. Therefore, understanding the impact of various factors on the risks of rapid and cumulative joint damage in patients with PsA is crucial. Important factors of interest include genetic markers such as HLA-B27, as well as time-varying biomarkers like CRP and ESR. Since patients are only assessed at periodic clinical visits, the data collected on joint damage recurrences and time-varying biomarkers are often incomplete. This motivates the two projects discussed in Chapters 3 and 4 of this thesis. In Chapter 3, we consider joint modeling of death and recurrent joint damage in PsA patients with intermittent observation. In Chapter 4, we consider joint modeling of a time-varying biomarker and a failure (of more than five total joint damages) with budgetary constraints.

1.4 Outline of the Thesis

The remainder of this thesis is organized as follows.

Joint modeling of recurrent and terminal event processes is considered in Chapter 2. We propose a new framework of the joint analysis by retaining the marginal hazard function for the terminal event process, adopting a rate function model for the recurrent event process conditional on a multiplicative random effect, and linking the terminal event and the random effect via a copula function. Under a Gaussian copula and a Gamma distributed random effect specification, plots of conditional recurrent and terminal event intensities given the full history are investigated to gain insights about the nature of the joint processes under our model formulation. When these processes are subject to right-censoring, the observed data likelihood can be maximized directly or via an expectation-maximization algorithm, with the variance estimates derived based on Louis method (Louis, 1982). We also present a computationally efficient two-stage procedure which enables semiparametric specifications. Simulation studies confirm that the proposed methods provide valid inference, and an application to the motivating study signifies the effect of pamidronate on reducing skeletal complications in patients with skeletal metastases due to breast cancer.

In many clinical settings, the occurrence of events can only be determined during periodic visits through radiographic examinations, blood tests, or detailed clinical examination. Consequently, instead of observing the exact times of event occurrences, we only observe the total number of occurrences during each interval between successive visits. Chapter

3 addresses such scenarios and extends the copula-based joint model for recurrent and terminal events proposed in Chapter 2 to accommodate interval-censored recurrent event data resulting from intermittent observation. A K -piece piecewise constant baseline rate function is adopted for the recurrent events due to its flexibility in modeling and ease of implementation. Each visit interval can then be divided into K subintervals based on pre-specified cut points. The complete data likelihood containing the latent random effects and the unobserved number of event occurrences within each subinterval is derived with an EM algorithm for estimation and inference. Additionally, a computationally efficient two-stage procedure allowing for semiparametric specifications is proposed. Simulation results demonstrate satisfactory finite sample performance of the proposed estimators. We apply this method to investigate the association between a biomarker (HLA-B27) and joint damage in patients with psoriatic arthritis.

In studies of chronic disease, interest often lies in studying the association between a time-dependent biomarker and a failure time of interest. Typically marker values are not available in continuous time, but are measured at periodic clinic visits when blood or urine samples are collected for testing. In Chapter 4, we present a multistate model for the joint analysis of a time-dependent biomarker process and a failure process by discretizing a continuous biomarker into two levels. This multistate model is further expanded to accommodate a marker-dependent visit process. We adopt Markov transition intensities with multiplicative proportional hazard models to govern the disease dynamics. In practice, collecting and cleaning data on a time-varying biomarker at all clinical visits can be labour-intensive. Consequently, a subsample of marker values is often accessed for statistical inference on the effect of the marker on the failure time of interest. We investigate the asymptotic bias of several naive strategies of selecting marker measurements with Cox regression models for inference under budgetary constraints. An alternative cost-effective selection method is discussed for efficient selection of individuals and the corresponding observed data likelihood is derived. The Fisher information is presented for an optimal design of the proposed selection method, and the asymptotic relative efficiency of regression coefficients under different designs is explored.

Finally, further comments regarding proposed methods and topics warranting future research are discussed in Chapter 5.

Chapter 2

A Copula-Based Joint Model for Recurrent and Terminal Events

2.1 Introduction

In many medical conditions subjects can experience recurrent clinical events. Examples include tumour recurrences in patients with superficial bladder cancer (Byar, 1980), repeated transient ischemic attacks and strokes in patients with asymptomatic carotid stenosis (Hobson et al., 1993), and recurrent exacerbations of respiratory symptoms in patients with cystic fibrosis (Fuchs et al., 1994). Various statistical methods for analyzing such recurrent events have been proposed and investigated (Gail et al., 1980; Prentice et al., 1981; Andersen and Gill, 1982; Wei et al., 1989; Lawless and Nadeau, 1995).

In addition to recurrent events, individuals are often also at risk of terminal events which preclude the occurrence of further recurrent events. In patients with advanced breast cancer and bone metastases, for example, death precludes the occurrence of future skeletal complications (Hortobagyi et al., 1996). In trials of such patients, one may wish to examine the effects of covariates on the occurrence of both the recurrent and terminal events; one may also be interested in modelling and understanding the dependence between the recurrent and terminal events. Joint modelling of the recurrent and terminal events is typically necessary to obtain estimates aligned with a data generating process and ensure valid inference.

Two general methodologies have been developed for joint analysis of recurrent and terminal events with one based on marginal methods, and the other based on shared

or correlated random effect models. Marginal methods do not require specification of a complete probability model for both recurrent and terminal event processes. [Cook and Lawless \(1997\)](#) specify a marginal distribution for the terminal event, and a conditional rate function for the recurrent events, defined as the instantaneous risk of a recurrent event conditional on the individual remaining at risk for recurrent events (i.e., conditional on them not having experienced the terminal event). [Ghosh and Lin \(2002\)](#) extended this approach to accommodate regression and proposed a two-sample test for a binary treatment effect. While marginal methods can be robust, they do not model the dependence structure between the recurrent and terminal events, inference is not likelihood-based, and the approach does not provide a good basis for prediction in terms of disease progression.

Random effect models incorporate the dependence between the recurrent and terminal events by introducing a shared random effect (frailty) and allowing it to affect the recurrent event rate and hazard function for failure where the specified recurrent event rate can be either complete ([Wang et al., 2001](#); [Huang and Wang, 2004](#); [Liu et al., 2004](#)) or marginal ([Ye et al., 2007](#); [Kalbfleisch et al., 2013](#)). Under the framework of shared random effects heterogeneity in recurrent events, along with the dependence between the recurrent and terminal events, is reflected by a single factor and it is therefore difficult to separately characterize heterogeneity and dependence. One may also consider separate random effects for the recurrent and terminal event processes and specifying conditional event intensities for the respective processes conditional on the corresponding random effect ([Cook and Lawless, 2007](#)). A joint distribution for the bivariate frailty with different covariance structures may be specified to account for various correlation between the recurrent and terminal events. A limitation of this approach is that the terminal event is modelled with a frailty model which is not appealing in the absence of the recurrent events. Moreover, the regression coefficient from the frailty Cox model is not aligned with the coefficient from a standard Cox model.

We propose a new framework for joint analysis of the recurrent and terminal event processes by retaining a focus on the marginal hazard function for the terminal event, adopting a rate function model for the recurrent event process conditional on a multiplicative random effect, and linking the terminal event and the random effect via a copula function. We also present a computationally efficient two-stage procedure for estimation and inference that allows for semiparametric analysis.

The remainder of this chapter is organized as follows. In [Section 2.2](#) we define notation, formulate the model, and give expressions for some conditional and marginal event rates of interest. The likelihood is presented, along with details on estimation and inference when this is marginalized via an expectation-maximization algorithm. An alternative two-stage estimation procedure is described in [Section 2.3](#) which facilitates semiparametric modeling.

The finite sample performance of the proposed estimators are assessed by simulation studies in Section 2.4 and we illustrate the methodology by application to a study on the effect of pamidronate on the occurrence of skeletal complications in patients with skeletal metastases from breast cancer in Section 2.5. Concluding remarks are made in Section 2.6.

2.2 A Copula-Based Joint Model

2.2.1 Model Formulation

2.2.1.1 Notation and Independence Assumptions

For the i th individual in a sample, we let $N_{i1}(s)$ denote the number of recurrent (type 1) events over $(0, s]$ experienced by this individual and $\{N_{i1}(s), 0 < s\}$ denote the corresponding counting process. Let $\Delta N_{i1}(s) = N_{i1}(s + \Delta s^-) - N_{i1}(s^-)$ denote the number of recurrent events over a small interval $[s, s + \Delta s)$ and $dN_{i1}(s) = \lim_{\Delta s \downarrow 0} \Delta N_{i1}(s)$. In addition to this recurrent event process, individuals are at risk of a terminal (type 2) event; once the terminal event occurs at time T_i the recurrent event process ends. Note $N_{i1}(s) = \int_0^s Y_i^\dagger(u) dN_{i1}(u)$ where $Y_i^\dagger(u) = \mathbb{1}(u \leq T_i)$ indicates the individual is terminal-event-free by time u . We let $N_{i2}(s) = I(T_i \leq s)$ and $\{N_{i2}(s), 0 < s\}$ be the terminal event counting process for this individual. If $\Delta N_{i2}(s) = N_{i2}(s + \Delta s^-) - N_{i2}(s^-)$ and $dN_{i2}(s) = \lim_{\Delta s \downarrow 0} \Delta N_{i2}(s)$, then we let $dN_i(s) = (dN_{i1}(s), dN_{i2}(s))'$, $N_i(s) = (N_{i1}(s), N_{i2}(s))'$ and $\{N_i(s), 0 < s\}$ be the bivariate counting process.

Let X_i denote a set of covariates for the individual and let $h(t | X_i; \theta_2)$ denote the marginal hazard function for T_i . The marginal survival function is

$$\mathcal{F}(t | X_i; \theta_2) = 1 - F(t | X_i; \theta_2) = \exp(-H(t | X_i; \theta_2)),$$

where $F(t | X_i; \theta_2)$ denotes the corresponding c.d.f. and $H(t | X_i; \theta_2) = \int h(s | X_i; \theta_2) ds$; we let $f(t | X_i; \theta_2)$ denote the corresponding p.d.f. Any type of marginal regression model can be specified for the terminal event but what follows is based on a proportional hazards model, with

$$h(t | X_i; \theta_2) = h_0(t; \alpha_2) \exp(X_i' \beta_2) \quad (2.1)$$

which can be fitted with parametric or semiparametric specifications; $h_0(t)$ is of an unspecified form for the latter.

To accommodate between-individual heterogeneity in the propensity for the type 1 (recurrent) events (which can not be explained by available covariates) we introduce a latent random effect $U_{i1} > 0$ with $U_{i1} \perp X_i$ for each individual. We assume $U_{i1}, i = 1, \dots, m$, are independently distributed with $E(U_{i1}) = 1$, $\text{var}(U_{i1}) = \phi_1$ and c.d.f. $P(U_{i1} \leq u_1; \phi_1) = G_1(u_1; \phi_1)$; the corresponding p.d.f. is denoted by $g_1(u_1; \phi_1)$. Conditional on the random effect $U_{i1} = u_{i1}$ and the full history $\mathcal{H}_i(s) = \{N_i(u), 0 < u < s, X_i\}$, the intensity function for the recurrent events is

$$\lim_{\Delta s \downarrow 0} \frac{P(\Delta N_{i1}(s) = 1 \mid \mathcal{H}_i(s), u_{i1}; \theta_1)}{\Delta s}. \quad (2.2)$$

Under the assumption that $\{N_{i1}(s), 0 < s < T_i\} \perp T_i \mid U_{i1} = u_{i1}, X_i$, we adopt a multiplicative model for the recurrent events so that (2.2) can be written as

$$\lim_{\Delta s \downarrow 0} \frac{P(\Delta N_{i1}(s) = 1 \mid X_i; \theta_1)}{\Delta s} = Y_i^\dagger(s) u_{i1} \rho(s \mid X_i; \theta_1). \quad (2.3)$$

A multiplicative model can be specified for the covariate effects so that $\rho(s \mid X_i; \theta_1)$ is of the form

$$\rho(s \mid X_i; \theta_1) = \rho_0(s; \alpha_1) \exp(X_i' \beta_1),$$

where $\rho_0(s; \alpha_1)$ is a baseline rate function of a parametric or unspecified form.

2.2.1.2 Specification of the Joint Model

In settings involving serious medical conditions, individuals with more severe or advanced forms of the disease are often at higher risk of adverse outcomes such that individuals experiencing frequent recurrent events are often at increased risk of the terminal event. In settings where the recurrent events signify declining health (e.g., when they are myocardial infarctions, stroke or hospitalizations), there may be an increased risk of death with each event occurrence. In cancer trials, individuals with more extensive tumour involvement they may be both at greater risk of recurrent metastases and death. We model a dependence between the two processes via a copula model.

Copula functions define joint distribution for a vector of uniform $[0, 1]$ random variables. Let $V_j, j = 1, 2, \dots, K$, be uniform random variables over $[0, 1]$, $V = (V_1, V_2, \dots, V_K)'$ be a vector random variable, and $v = (v_1, v_2, \dots, v_K)'$ denote the corresponding realization. A K-dimensional copula function indexed by a vector of dependence parameters ψ is

$$\mathcal{C}(v; \psi) = P(V_1 \leq v_1, V_2 \leq v_2, \dots, V_K \leq v_K; \psi).$$

The degree of association between each pair of V_j and V_l , $j \neq l$, $j, l = 1, 2, \dots, K$, can be expressed in terms of Kendall's τ (Kendall, 1938), denoted as τ_{jl} . $\tau_{jl} = 0$ signifies an independence between V_j and V_l and $\tau_{jl} = 1$ signifies a perfect association. One appealing feature of copula models is that the dependence structure induced by them is solely characterized by the copula function itself. Consequently the marginal models can be constructed in a flexible manner. Within the copula framework, regression parameters in marginal models are meaningful and maintain the same interpretation regardless of the choice of the copula function used. Several widely-used copula functions include the Gaussian copula, the Clayton copula and the Frank copula; the first one belongs to the elliptical copula family and the last two belong to the Archimedean family (Nelsen, 2006).

The proposed model uses copula functions in a unique way. Typically they are used to specify joint models for multivariate failure time data (Joe, 1997) or perhaps multivariate random effect distributions (Cook et al., 2010). Here we use the model to characterize the dependence between the multiplicative non-negative random effect on the recurrent event process and the terminal event time. This enables one to retain the simple marginal prospective of any desired failure time model, while allowing the event history to affect the intensity for death, and while allowing the event history (number of recurrent events and survival status) to effect the intensity for recurrent events. Specifically, given X_i , a joint model for U_{i1} and T_i is constructed by setting $V_1 = G_1(U_{i1}; \phi_1)$, $V_2 = \mathcal{F}(T_i | X_i; \theta_2)$ and linking these via a bivariate copula model indexed by a single parameter ψ

$$P(U_{i1} \leq u_1, T_i > t | X_i; \theta_2, \phi_1, \psi) = \mathcal{C}(G_1(u_1; \phi_1), \mathcal{F}(t | X_i; \theta_2); \psi).$$

Note $\theta_1 = (\alpha'_1, \beta'_1)'$ indexes the recurrent event process in (2.3), $\theta_2 = (\alpha'_2, \beta'_2)'$ indexes the terminal event process in (2.1), ϕ_1 indexes the random effect distribution, and ψ measures the association between the random effect (and hence the recurrent event process) and the terminal event process. We let $\theta = (\theta'_1, \theta'_2, \phi_1, \psi)'$ denote the set of full parameters of interest. Under this joint model formulation, we assume the recurrent events and the terminal event are independent given covariates and the random effect as long as we know if the bivariate process has not been terminated.

2.2.2 Implied Rates and Intensities

2.2.2.1 Computation of Intensities

Based on the model framework along with the aforementioned assumptions, some conditional and marginal intensities can be calculated for the recurrent event and terminal event

processes. We do this in what follows in order to gain insights into the propensities of this model.

The *complete intensity function for the recurrent event process* is

$$\begin{aligned} E\{dN_{i1}(s) \mid \mathcal{H}_i(s)\} &= E_{U_{i1}}\{E\{dN_{i1}(s) \mid U_{i1}, Y_i^\dagger(s), \mathcal{H}_i(s)\} \mid Y_i^\dagger(s), \mathcal{H}_i(s)\} \\ &= Y_i^\dagger(s) \rho(s \mid X_i; \theta_1) E\{U_{i1} \mid dN_{i1}(u), 0 < u < s, T_i \geq s, X_i\} \end{aligned} \quad (2.4)$$

where $E\{U_{i1} \mid dN_{i1}(u), 0 < u < s, T_i \geq s, X_i\}$ can be calculated as

$$\int_0^\infty u_1 g_1(u_1 \mid dN_{i1}(u), 0 < u < s, T_i \geq s, X_i; \theta) du_1$$

with $g_1(u_1 \mid dN_{i1}(u), 0 < u < s, T_i \geq s, X_i; \theta)$ equal to

$$\frac{\frac{\partial}{\partial u_1} \mathcal{C}(G_1(u_1; \phi_1), \mathcal{F}(s \mid X_i; \theta_2; \psi) P(dN_{i1}(u), 0 < u < s \mid u_1, X_i, Y_i^\dagger(s) = 1; \theta_1))}{\int_0^\infty \frac{\partial}{\partial u_1} \mathcal{C}(G_1(u_1; \phi_1), \mathcal{F}(s \mid X_i; \theta_2; \psi) P(dN_{i1}(u), 0 < u < s \mid u_1, X_i, Y_i^\dagger(s) = 1; \theta_1)) du_1}.$$

Note under the specification of (2.3), $P(dN_{i1}(u), 0 < u < s \mid u_1, X_i, Y_i^\dagger(s) = 1; \theta_1)$ is calculated as

$$\frac{\left\{ \prod_{j=0}^{N_{i1}(s)} u_1 \rho(s_{ij} \mid X_i; \theta_1) \right\} \exp(-u_1 R(s \mid X_i; \theta_1))}{N_{i1}(s)!} \quad (2.5)$$

where $s_{i0} = 0$, s_{ij} , $j = 1, \dots, N_{i1}(s)$ are recurrent event times, and

$$R(s \mid X_i; \theta_1) = \int_0^s \rho(u \mid X_i; \theta_1) du.$$

The rate of the recurrent event process at time s conditional only on being alive at time s is

$$\begin{aligned} E\{dN_{i1}(s) \mid X_i, Y_i^\dagger(s) = 1\} &= E_{U_{i1}}\{E\{dN_{i1}(s) \mid U_{i1}, X_i, Y_i^\dagger(s) = 1\} \mid X_i, Y_i^\dagger(s) = 1\} \\ &= \rho(s \mid X_i; \theta_1) E\{U_{i1} \mid X_i, Y_i^\dagger(s) = 1\} \end{aligned}$$

where $E\{U_{i1} \mid X_i, Y_i^\dagger(s) = 1\} = \int_0^\infty u_1 g_1(u_1 \mid X_i, Y_i^\dagger(s) = 1; \theta_2, \phi_1, \psi) du_1$ with

$$g_1(u_1 \mid X_i, Y_i^\dagger(s) = 1; \theta_2, \phi_1, \psi) = \frac{P(U_{i1} = u_1, T_i \geq t \mid X_i; \theta_2, \phi_1, \psi)}{\mathcal{F}(t \mid X_i; \theta_2)}. \quad (2.6)$$

This quantity can be estimated nonparametrically by restricting attention to individuals who are alive and estimating the proportion having an event at time s ; see [Cook and Lawless \(1997\)](#). The *marginal recurrent event rate* averaging over the survival status is

$$\begin{aligned} E\{dN_{i1}(s) \mid X_i\} &= E_{U_{i1}, Y_i^\dagger(s)}\{E\{dN_{i1}(s) \mid U_{i1}, Y_i^\dagger(s), X_i\}\} \\ &= \rho(s \mid X_i; \theta_1) E\{U_{i1} \mid Y_i^\dagger(s) = 1, X_i\} \mathcal{F}(s \mid X_i; \theta_2) \end{aligned} \quad (2.7)$$

where $E\{U_{i1} \mid Y_i^\dagger(s) = 1, X_i\} = E\{U_{i1} \mid X_i, T_i \geq s\}$. This is a quantity that can also be estimated nonparametrically as described by [Cook and Lawless \(1997\)](#) and considered by [Ghosh and Lin \(2002\)](#).

The *complete intensity for the terminal event process* is

$$E\{dN_{i2}(s) \mid \mathcal{H}_i(s)\} = Y_i^\dagger(s) P(dN_{i2}(s) = 1 \mid \mathcal{H}_i(s); \theta) \quad (2.8)$$

where $P(dN_{i2}(s) = 1 \mid \mathcal{H}_i(s); \theta)$ is given by

$$\begin{aligned} &\frac{P(T_i = s, dN_{i1}(u), 0 < u < s \mid X_i; \theta)}{P(T_i \geq s, dN_{i1}(u), 0 < u < s \mid X_i; \theta)} \\ &= \frac{\int_0^\infty P(U_{i1} = u_1, T_i = s \mid X_i; \theta_2, \phi_1, \psi) P(dN_{i1}(u), 0 < u < s \mid u_1, X_i, Y_i^\dagger(s) = 1; \theta_1) du_1}{\int_0^\infty P(U_{i1} = u_1, T_i \geq s \mid X_i; \theta_2, \phi_1, \psi) P(dN_{i1}(u), 0 < u < s \mid u_1, X_i, Y_i^\dagger(s) = 1; \theta_1) du_1} \end{aligned}$$

with $P(dN_{i1}(u), 0 < u < s \mid u_1, X_i, Y_i^\dagger(s) = 1; \theta_1)$ obtained from (2.5). Note that the complete intensity for the terminal event of death makes it clear that the recurrent event process impacts the risk of death (although through integral expressions). The appealing feature of our model is that if one only conditions on the treatment indicator a simple Cox model is appropriate and leads to a standard interpretation of the treatment effect on this outcome. Thus insights are gained regarding the relationship between the recurrent and terminal event processes through this model, but the model is compatible with a familiar Cox model used for time to event outcomes.

If $N_{i1}(s) = k$, (2.4) is the recurrent event intensity that characterizes the risk of a $k \rightarrow k + 1$ transition in the multistate diagram in Figure 2.1, and (2.8) is the terminal event intensity characterizing the risk of a $k \rightarrow D$ transition. While these are of interest to study the joint process from the perspective of intensities, marginal quantities are useful for assessing treatment effects in clinical trials.

2.2.2.2 Graphical Explorations of Complete Event Intensities

Here we aim to gain insight into the forms of the complete recurrent event intensity $E\{dN_{i1}(s) \mid \mathcal{H}_i(s)\}$ in (2.4) and the complete terminal event intensity $E\{dN_{i2}(s) \mid \mathcal{H}_i(s)\}$

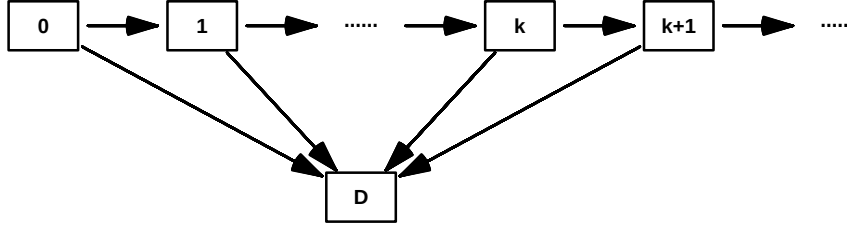


Figure 2.1: A multistate diagram illustrating the recurrent and terminal event processes

in (2.8) by specifying model parameters and plotting the intensities over time for some illustrative sample paths. For simplicity, we consider a simple setting with a binary covariate where $X_i = 0$ or 1. We consider multiplicative models for both the recurrent and terminal event processes with constant baseline functions (i.e., $\rho_0(s; \alpha_1) = \rho_0$ and $h_0(t; \alpha_2) = h_0$). We let the random effect U_{i1} be gamma distributed with mean = 1 and variance = ϕ_1 , and specify a Gaussian copula with Kendall's τ indexing the association; the copula parameter ψ is a deterministic function of τ (see Appendix A.1). For the coefficients, we set $\beta_1 = \log 0.8$ and $\beta_2 = \log 0.9$. We let $\phi_1 = 0.2, 0.4$ to reflect different levels of heterogeneity in the recurrent event process and vary τ over 0, 0.2, and 0.4 to investigate scenarios with no, weak and medium associations between the two processes. We consider a common administrative right-censoring time $C_i = 1$ for all individuals. Let $E(N_{i1}(1)|X_i = 0) = \int_0^1 E\{dN_{i1}(s)|X_i = 0\} ds$ be the expected number of recurrent events by time 1 in the control group ($X_i = 0$) with $E\{dN_{i1}(s) | X_i = 0\}$ obtained from (2.7), and let $P(T_i \leq 1|X_i = 0)$ be the probability of the terminal event by time 1 in that group. We then set $\rho_0 = 3.0543$ and $h_0 = 0.9163$ so that $E\{N_{i1}(1)|X_i = 0\} = 2$ and $P(T_i \leq 1|X_i = 0) = 0.6$ when $\tau = 0$. The set values of ρ_0 and h_0 are used for all combinations of ϕ_1 and τ .

We consider four sample paths illustrated in Figure 2.2 and evaluate the conditional intensities under different specifications of ϕ_1 and τ . Specifically, for sample paths A – D, we compute $E\{dN_{i1}(s) | \mathcal{H}_i(s)\}$ and $E\{dN_{i2}(s) | \mathcal{H}_i(s)\}$ for the group with $X_i = 0$, and plot them in Figure 2.3.

For the recurrent event process, we observe that in general, the conditional recurrent event intensity given the full history gradually decreases over time provided there are no recurrent or terminal events, while at time points where the recurrent event occurs, the conditional intensity instantly jumps. This is similar to how the intensity behaves in mixed

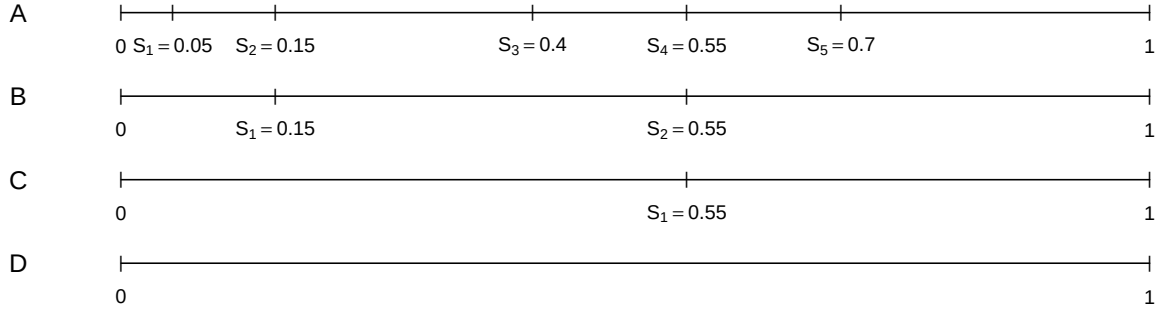


Figure 2.2: Four illustrative sample paths to investigate intensities for recurrent and terminal events implied by the model specified in Section 2.2.1.2

Poisson models (Cook and Lawless, 2007) but we note that survival up to a time $t > 0$ also has an impact on the recurrent event intensity here. We also find that as the association τ between the two processes increases, the magnitude of the slope of the conditional intensity decreases when there is no occurrence of the recurrent event. This makes sense because by our model framework, a survival up to some point suggests that this individual has a smaller than average random effect u_1 and hence a smaller recurrent event intensity. Hence a stronger association between the two processes makes the impact of the survival on the recurrent event intensity stronger. In addition, comparing left and right panels, we see that the instantaneous change of the recurrent event intensity (including the size of the jump at each recurrent event time) is greater when $\phi_1 = 0.4$.

For the terminal event process when $\tau = 0$, the event intensity is not affected by the recurrent event process, giving horizontal lines when the baseline hazard is constant. When $\tau > 0$, the terminal event intensity is similar to what we observe in the recurrent event intensity: the conditional terminal event intensity jumps to a higher value upon the occurrence of each recurrent event time and it decreases over time when there are no additional recurrent events occurring. The magnitude of the instantaneous change of the conditional terminal event intensity is influenced by both τ and ϕ_1 . A higher value of τ results in a larger change because it corresponds to a stronger association between the two processes, while a higher ϕ_1 reflects greater heterogeneity in risks for recurrent events and an event occurrence suggests a more extreme value of the frailty.

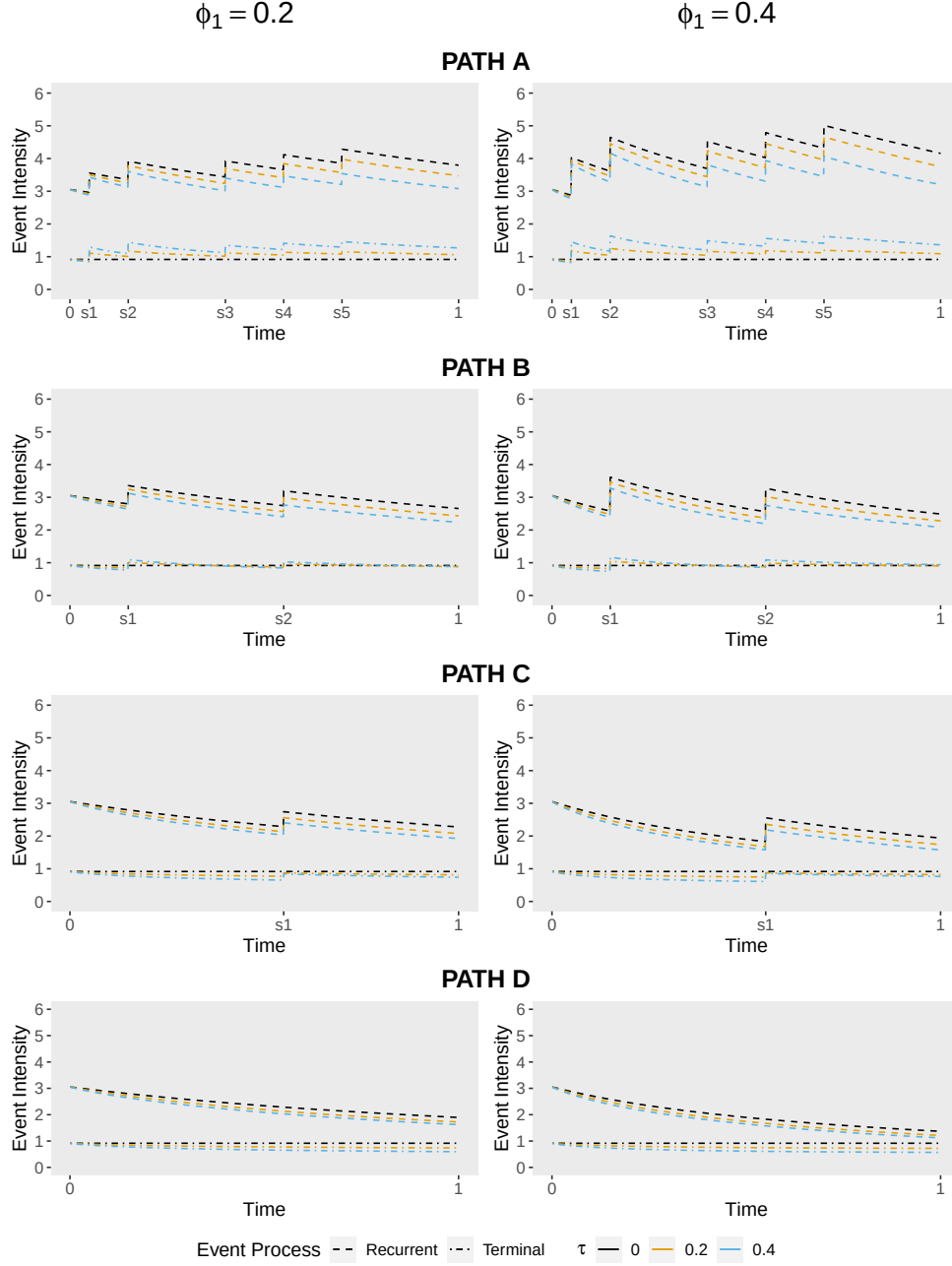


Figure 2.3: Plots of the complete intensity for the recurrent event process ($E\{dN_{i1}(s) \mid \mathcal{H}_i(s)\}$) and the complete intensity for the terminal event process ($E\{dN_{i2}(s) \mid \mathcal{H}_i(s)\}$) versus s for $X_i = 0$, for $\phi_1 = 0.2$ (left panels) and $\phi_1 = 0.4$ (right panels); the four rows correspond to the four illustrative paths in Figure 2.2

2.2.3 Parameter Estimation and Inference

To accommodate possible right-censoring, we let C_i denote an independent right-censoring time, $W_i = \min(T_i, C_i)$, $\delta_i = \mathbb{1}(T_i \leq C_i)$, $Y_i(s) = \mathbb{1}(s \leq C_i)$ and $\bar{Y}_i(s) = Y_i(s)Y_i^\dagger(s)$. We define the observable counting process notation for the recurrent and terminal events as $d\bar{N}_{ik}(s) = \bar{Y}_i(s)dN_{ik}(s)$, $\bar{N}_{ik}(s) = \int_0^s d\bar{N}_{ik}(u)$, $k = 1, 2$, and let $d\bar{N}_i(s) = (d\bar{N}_{i1}(s), d\bar{N}_{i2}(s))'$ and $\bar{N}_i(s) = (\bar{N}_{i1}(s), \bar{N}_{i2}(s))'$. The observable history of the bivariate process under right-censoring is then $\mathcal{H}_i(s) = \{\bar{Y}_i(u), d\bar{N}_i(u), 0 < u < s, X_i\}$. Conditioning on $U_{i1} = u_{i1}$ and $\mathcal{H}_i(s)$, the observed recurrent event rate is then

$$\lim_{\Delta s \downarrow 0} \frac{P(\Delta \bar{N}_{i1}(s) = 1 \mid \mathcal{H}_i(s), u_{i1}; \theta_1)}{\Delta s} = \bar{Y}_i(s)u_{i1}\rho(s \mid X_i; \theta_1).$$

Under the assumption of independent and non-informative censoring, the observed data partial likelihood contributed from individual i , denoted by $L_i(\theta)$, is

$$\begin{aligned} & h(w_i \mid X_i; \theta_2)^{\delta_i} \exp(-H(w_i \mid X_i; \theta_2)) \\ & \times \left\{ \int_0^\infty P(dN_{i1}(s), 0 < s < w_i \mid X_i, u_1, Y_i^\dagger(w_i) = 1; \theta_1) g_1(u_1 \mid X_i, T_i = w_i; \theta_2, \phi_1, \psi) du_1 \right\}^{\delta_i} \\ & \times \left\{ \int_0^\infty P(dN_{i1}(s), 0 < s < w_i \mid X_i, u_1, Y_i^\dagger(w_i) = 1; \theta_1) g_1(u_1 \mid X_i, T_i > w_i; \theta_2, \phi_1, \psi) du_1 \right\}^{1-\delta_i} \end{aligned}$$

where $P(dN_{i1}(s), 0 < s < w_i \mid X_i, u_1, Y_i^\dagger(w_i) = 1; \theta_1)$, $g_1(u_1 \mid X_i, T_i > w_i; \theta_2, \phi_1, \psi)$ can be obtained from (2.5) and (2.6) respectively, and

$$g_1(u_1 \mid X_i, T_i = t; \theta_2, \phi_1, \psi) = \frac{P(U_{i1} = u_1, T_i = t \mid X_i; \theta_2, \phi_1, \psi)}{f(t \mid X_i; \theta_2)}.$$

The corresponding observed data log-likelihood is then $\log L_i(\theta)$, and the observed data score and information contribution from this individual are $S_i(\theta) = \partial \log L_i(\theta) / \partial \theta$ and $I_i(\theta) = -\partial^2 \log L_i(\theta) / \partial \theta \partial \theta'$ respectively. With a sample of data from m independent individual processes, for parametric analysis the maximum likelihood estimator (MLE) of θ (denoted by $\hat{\theta}$) can be obtained by solving $\sum_{i=1}^m S_i(\theta) = 0$. Variance estimate of the MLE (denoted by $\text{var}(\hat{\theta})$) can be obtained by calculating $\{\sum_{i=1}^m I_i(\hat{\theta})\}^{-1}$, where $S_i(\theta)$ and $I_i(\theta)$ are the contributions from individual i , $i = 1, \dots, m$.

An EM algorithm (Dempster et al., 1977) can be appealing for estimation when latent variables are involved and particularly when existing algorithms can be used for optimization at the maximization step. We consider a complete data likelihood contributed from the

individual where U_{i1} is part of the complete data. This gives a complete data log-likelihood which can be partitioned as

$$\log \mathcal{L}_i(\theta) = \ell_{i1}(\theta_1) + \ell_{i2}(\theta_2) + \ell_{i3}(\theta_2, \phi_1, \psi) \quad (2.9)$$

where we denote

$$\ell_{i1}(\theta_1) = \sum_{j=0}^{n_i} \log \rho(s_{ij} \mid X_i; \theta_1) - u_{i1} R(w_i \mid X_i; \theta_1),$$

$$\ell_{i2}(\theta_2) = \delta_i \log h(w_i \mid X_i; \theta_2) - H(w_i \mid X_i; \theta_2),$$

and

$$\begin{aligned} \ell_{i3}(\theta_2, \phi_1, \psi) &= \delta_i \log g_1(u_{i1} \mid X_i, T_i = w_i; \theta_2, \phi_1, \psi) \\ &\quad + (1 - \delta_i) \log g_1(u_{i1} \mid X_i, T_i > w_i; \theta_2, \phi_1, \psi), \end{aligned}$$

with $n_i = N_{i1}(w_i)$. We let $\mathcal{S}_i(\theta) = \partial \log \mathcal{L}_i(\theta) / \partial \theta$ and $\mathcal{I}_i(\theta) = -\partial^2 \log \mathcal{L}_i(\theta) / \partial \theta \partial \theta'$ denote the corresponding complete data score vector and information matrix respectively. The EM algorithm is described as follows.

The E-Step: Given a current estimate $\theta^{(r-1)}$, we calculate the expectation of the complete data log-likelihood in (2.9) with respect to U_{i1} conditioning on the observed data $\bar{\mathcal{H}}_i(w_i^+)$ which is defined in Section 2.2.1.1. This gives us

$$\begin{aligned} E(\ell_{i1}(\theta_1) \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)}) &= \sum_{j=0}^{n_i} \log \rho(s_{ij} \mid X_i; \theta_1) - E(U_{i1} \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)}) R(w_i \mid X_i; \theta_1), \\ E(\ell_{i2}(\theta_2) \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)}) &= \ell_{i2}(\theta_2) \end{aligned}$$

since the random effect does not appear in $\ell_{i2}(\theta_2)$, and

$$\begin{aligned} E(\ell_{i3}(\theta_2, \phi_1, \psi) \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)}) &= \delta_i E(\log g_1(U_{i1} \mid X_i, T_i = w_i; \theta_2, \phi_1, \psi) \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)}) \\ &\quad + (1 - \delta_i) E(\log g_1(U_{i1} \mid X_i, T_i > w_i; \theta_2, \phi_1, \psi) \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)}), \end{aligned}$$

with

$$E(U_{i1} \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)}) = \frac{\int_0^\infty u_1 \mathcal{L}_i(\theta^{(r-1)}) du_1}{\int_0^\infty \mathcal{L}_i(\theta^{(r-1)}) du_1}, \quad (2.10a)$$

$E(\log g_1(U_{i1} \mid X_i, T_i = w_i; \theta_2, \phi_1, \psi) \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)})$ given by

$$\frac{\int_0^\infty \log g(u_1 \mid X_i, T_i = w_i; \theta_2, \phi_1, \psi) \mathcal{L}_i(\theta^{(r-1)}) du_1}{\int_0^\infty \mathcal{L}_i(\theta^{(r-1)}) du_1}, \quad (2.10b)$$

and $E(\log g_1(U_{i1} \mid X_i, T_i > w_i; \theta_2, \phi_1, \psi) \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)})$ given by

$$\frac{\int_0^\infty \log g_1(u_1 \mid X_i, T_i > w_i; \theta_2, \phi_1, \psi) \mathcal{L}_i(\theta^{(r-1)}) du_1}{\int_0^\infty \mathcal{L}_i(\theta^{(r-1)}) du_1}. \quad (2.10c)$$

Detailed calculations of the above conditional expectations under a joint Gaussian copula can be found in Appendix A.2.

The M-Step: We then maximize $\sum_{i=1}^m E(\log \mathcal{L}_i(\theta) \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)})$ w.r.t. to θ and obtain a new estimate $\theta^{(r)}$: for the recurrent event process, we maximize $\sum_{i=1}^m E(\ell_{i1}(\theta_1) \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)})$ w.r.t. θ_1 and obtain $\theta_1^{(r)}$; for the rest of the parameters, we maximize $\sum_{i=1}^m \{\ell_{i2}(\theta_2) + E(\ell_{i3}(\theta_2, \phi_1, \psi) \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)})\}$ w.r.t. (θ_2, ϕ_1, ψ) and obtain $(\theta_2^{(r)}, \phi_1^{(r)}, \psi^{(r)})$. We have $\theta^{(r)} = ((\theta_1^{(r)})', (\theta_2^{(r)})', \phi_1^{(r)}, \psi^{(r)})'$. The *E-Step* and *M-Step* are repeated until the absolute value of each element of the vector $\theta^{(r)} - \theta^{(r-1)}$ are smaller than 10^{-4} and $\hat{\theta}$ is obtained.

The observed data information can be obtained alternatively through the complete data score and information as

$$\begin{aligned} I_i(\theta) = & E(\mathcal{I}_i(\theta) \mid \bar{\mathcal{H}}_i(w_i^+); \theta) - E(\mathcal{S}_i(\theta) \mathcal{S}_i'(\theta) \mid \bar{\mathcal{H}}_i(w_i^+); \theta) \\ & + E(\mathcal{S}_i(\theta) \mid \bar{\mathcal{H}}_i(w_i^+); \theta) E(\mathcal{S}_i'(\theta) \mid \bar{\mathcal{H}}_i(w_i^+); \theta), \end{aligned} \quad (2.11)$$

as shown by [Louis \(1982\)](#). We use $\hat{\text{var}}(\hat{\theta}) = \{\sum_{i=1}^m I_i(\hat{\theta})\}^{-1}$ as an estimated variance-covariance matrix of $\hat{\theta}$ and evaluate the validity of it in simulation studies.

2.3 A Two-Stage Procedure

2.3.1 Estimation and Inference via Direct Optimization

An appealing feature of the model formulation in Section 2.2 is that the analyses of covariate effects on the terminal event time can be based on a standard Cox regression model ([Cox, 1972](#)). This is in contrast to most joint models for recurrent and terminal event where random effects are introduced in the failure time model to incorporate a dependence structure - in this case the estimate of the treatment effect on the terminal event has a conditional, rather than marginal, interpretation and one would not adopt such a specification if one were not fitting a joint model. In light of this appealing feature of our joint model, we now consider a computationally more convenient two-stage estimation procedure

where in stage one we focus on the marginal terminal event process and estimate related parameters θ_2 , and in stage two we estimate all remaining parameters, with θ_2 set to their estimated value obtained at stage one. We let $\bar{\theta}$ denote all the other parameters except θ_2 (i.e., $\bar{\theta} = (\theta'_1, \phi_1, \psi)'$) and reorder the parameters so $\theta = (\theta'_2, \bar{\theta})'$. The two-stage estimation procedure works as follows.

Stage 1: Estimate θ_2 by maximizing $\sum_{i=1}^m \ell_{i2}(\theta_2)$ w.r.t. θ_2 to obtain $\tilde{\theta}_2$.

Stage 2: Obtain $\tilde{\bar{\theta}} = (\tilde{\theta}'_1, \tilde{\phi}_1, \tilde{\psi})'$ by maximizing $\sum_{i=1}^m \log L_i(\tilde{\theta}_2, \bar{\theta})$, the overall observed data log-likelihood after $\tilde{\theta}_2$ is inserted. Let $\tilde{\theta} = (\tilde{\theta}'_2, \tilde{\bar{\theta}})'$ denote the estimate from this two-stage procedure.

We denote the scores of the partial log-likelihoods used in the two stages contributed from the i th individual as $\mathcal{U}_{i1}(\theta_2) = \partial \ell_{i2}(\theta_2) / \partial \theta_2$ and $\mathcal{U}_{i2}(\theta) = \partial \log L_i(\theta) / \partial \bar{\theta}$ respectively, and let $\mathcal{U}_i(\theta) = (\mathcal{U}_{i1}(\theta_2), \mathcal{U}_{i2}(\theta))'$. The two-stage procedure solves the unbiased estimating equation $\sum_{i=1}^m \mathcal{U}_i(\theta) = 0$. By the standard estimating function theory (Boos and Stefanski, 2013), we have

$$\sqrt{m}(\tilde{\theta} - \theta) \rightarrow \text{MVN}(0, \mathcal{A}^{-1}(\theta) \mathcal{B}(\theta) (\mathcal{A}^{-1}(\theta))') \quad (2.12)$$

where $\mathcal{A}(\theta) = E(-\partial \mathcal{U}_i(\theta) / \partial \theta')$ and $\mathcal{B}(\theta) = E(\mathcal{U}_i(\theta) \mathcal{U}_i'(\theta))$. Moreover, $\mathcal{A}(\theta)$ and $\mathcal{B}(\theta)$ can be estimated empirically by

$$A(\tilde{\theta}) = -\frac{1}{m} \sum_{i=1}^m \frac{\partial \mathcal{U}_i(\theta)}{\partial \theta'} \Big|_{\theta=\tilde{\theta}} \quad \text{and} \quad B(\tilde{\theta}) = \frac{1}{m} \sum_{i=1}^m \mathcal{U}_i(\theta) \mathcal{U}_i'(\theta) \Big|_{\theta=\tilde{\theta}}. \quad (2.13)$$

Hence, the asymptotic covariance matrix can be estimated by $\hat{\text{var}}(\tilde{\theta}) = A^{-1}(\tilde{\theta}) B(\tilde{\theta}) (A^{-1}(\tilde{\theta}))' / m$. Note $\mathcal{U}_{i2}(\theta)$ is a sub-matrix of the corresponding observed data score $S_i(\theta)$ and $-\partial \mathcal{U}_{i2}(\theta) / \partial \theta'$ is a sub-matrix of the corresponding observed data information $I_i(\theta)$.

2.3.2 Two-Stage Estimation with the EM Algorithm

2.3.2.1 Parametric Analysis

Stage 2 of the two-stage estimation procedure can be complemented via an EM algorithm as follows. As described in (2.9), the complete data log-likelihood $\log \mathcal{L}_i(\theta)$ can be partitioned into three parts: $\ell_{i1}(\theta_1)$, $\ell_{i2}(\theta_2)$ and $\ell_{i3}(\theta_2, \phi_1, \psi)$, so the two-stage estimation procedure can be described as follows.

Stage 1: As in Section 2.3.1 we obtain $\tilde{\theta}_2$ by maximizing $\sum_{i=1}^m \ell_{i2}(\theta_2)$ w.r.t. θ_2 .

Stage 2: We use an EM algorithm to estimate $\bar{\theta}$ as follows. Let $\theta^{(r-1)} = (\tilde{\theta}'_2, (\bar{\theta}^{(r-1)})')'$.

Step 1: Maximize $\sum_{i=1}^m E(\ell_{i1}(\theta_1) \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)})$ w.r.t θ_1 and obtain $\theta_1^{(r)}$.

Step 2: Maximize $\sum_{i=1}^m \{\ell_{i2}(\tilde{\theta}_2) + E(\ell_{i3}(\tilde{\theta}_2, \phi_1, \psi) \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)})\}$ w.r.t (ϕ_1, ψ) and obtain $(\phi_1^{(r)}, \psi^{(r)})$.

Note $\bar{\theta}^{(r)} = ((\theta_1^{(r)})', \phi_1^{(r)}, \psi^{(r)})'$. We repeat Steps 1 and 2 until $\bar{\theta}^{(r)}$ converges to obtain $\tilde{\theta}$.

The estimate $\tilde{\theta} = (\tilde{\theta}'_2, \tilde{\theta}')'$ obtained here is theoretically the same as the estimate obtained in Section 2.3.1 because we are solving the same estimating equation $\sum_{i=1}^m \mathcal{U}_i(\theta) = 0$. The standard estimating function theory therefore still applies and we calculate $\mathcal{U}_i(\theta)$, $-\partial \mathcal{U}_i(\theta)/\partial \theta'$ to obtain $A^{-1}(\tilde{\theta})$, $B(\tilde{\theta})$ using (2.13), and thus a variance estimate $\text{var}(\tilde{\theta})$.

The score of the partial log-likelihood used in the first stage contributed from the i th individual is still $\mathcal{U}_{i1}(\theta_2)$, and we denote the score of the partial complete data log-likelihood used in the second stage as $\mathcal{U}_{i2}^c(\theta) = ((\mathcal{U}_{i21}^c(\theta_1))', (\mathcal{U}_{i22}^c(\theta))')'$ where $\mathcal{U}_{i21}^c(\theta_1) = \partial \ell_{i1}(\theta_1)/\partial \theta_1$ and $\mathcal{U}_{i22}^c(\theta) = \partial \ell_{i3}(\theta_2, \phi_1, \psi)/\partial (\phi_1, \psi)'$. Note $\mathcal{U}_{i2}^c(\theta)$ is a sub-matrix of the corresponding complete data score $\mathcal{S}_i(\theta)$ and we can obtain $\mathcal{U}_{i2}(\theta)$ using

$$\mathcal{U}_{i2}(\theta) = E(\mathcal{U}_{i2}^c(\theta) \mid \bar{\mathcal{H}}_i(w_i^+); \theta).$$

For $-\partial \mathcal{U}_{i2}(\theta)/\partial \theta'$, we use (2.11) to calculate the corresponding observed information $I_i(\theta)$ and then extract $-\partial \mathcal{U}_{i2}(\theta)/\partial \theta$ from $I_i(\theta)$.

2.3.2.2 Semiparametric Analysis

The two-stage estimation procedure is particularly appealing when considering semiparametric analyses since through its use one avoids the need to directly maximize the likelihood with respect to two infinite dimensional parameters - the baseline rate and hazard functions. Here we adopt a semiparametric Cox model for both the recurrent and terminal events with multiplicative covariate effects of the forms:

$$\begin{aligned} dR(s \mid X_i; \theta_1) &= dR_0(s) \exp(X_i' \beta_1), \\ dH(t \mid X_i; \theta_2) &= dH_0(t) \exp(X_i' \beta_2). \end{aligned}$$

Estimation proceeds via partial likelihood (Breslow and Day, 1980) in a two-stage procedure using the EM algorithm. The idea is similar to what was described in Section 2.3.2.1.

Stage 1: We obtain $\tilde{\theta}_2$ by maximizing $\sum_{i=1}^m \ell_{i2}(\theta_2)$ w.r.t. $\theta_2 = (dH_0(\cdot), \beta_2')'$. Note $\sum_{i=1}^m \ell_{i2}(\theta_2)$ can be re-written using counting process notation as

$$\sum_{i=1}^m \left\{ \int_0^\infty \log dH(s | X_i) \bar{Y}_i(s) dN_{i2}(s) - \int_0^\infty \bar{Y}_i(s) dH(s | X_i) \right\}. \quad (2.14)$$

Taking its derivative w.r.t $dH_0(t)$ and setting the derivative to 0 gives an estimating equation for $dH_0(t)$:

$$\sum_{i=1}^m \bar{Y}_i(t) [dN_{i2}(t) - dH_0(t) \exp(X_i' \beta_2)] = 0, \quad t \geq 0.$$

By solving this estimating equation for $dH_0(t)$ at a fixed β_2 , a profile likelihood estimate (Murphy and Van Der Vaart, 2000) is

$$d\tilde{H}_0(t; \beta_2) = \frac{\sum_{i=1}^m \bar{Y}_i(t) dN_{i2}(t)}{\sum_{i=1}^m \bar{Y}_i(t) \exp(X_i' \beta_2)}. \quad (2.15)$$

Moreover, taking the derivative of (2.14) w.r.t. to β_2 gives

$$\sum_{i=1}^m \left\{ \int_0^\infty \bar{Y}_i(s) X_i [dN_{i2}(s) - dH_0(s) \exp(X_i' \beta_2)] \right\}, \quad (2.16)$$

and plugging $d\tilde{H}_0(s; \beta_2)$ into (2.16) gives a profile score function for β_2

$$U(\beta_2) = \sum_{i=1}^m \left\{ \int_0^\infty \bar{Y}_i(s) \left[X_i - \frac{\sum_{l=1}^m \bar{Y}_l(s) X_l \exp(X_l' \beta_2)}{\sum_{l=1}^m \bar{Y}_l(s) \exp(X_l' \beta_2)} \right] dN_{i2}(s) \right\}.$$

Setting $U(\beta_2) = 0$ and solving for β_2 gives $\tilde{\beta}_2$. Which, upon substituting in (2.15), gives $d\tilde{H}_0(t)$.

Stage 2: We use an EM algorithm to estimate $\bar{\theta}$. If the specified model for the recurrent events is parametric, we follow the steps described in Stage 2 of Section 2.3.2.1 and calculate scores directly. If the specified model is Cox semiparametric, we do the following steps.

Step 1: We maximize $\sum_{i=1}^m E(\ell_{i1}(\theta_1) | \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)})$ w.r.t $\theta_1 = (dR_0(\cdot), \beta_1')'$ and obtain $\theta_1^{(r)}$. Note $\sum_{i=1}^m E(\ell_{i1}(\theta_1) | \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)})$ can also be re-written using counting process notation as

$$\sum_{i=1}^m \left\{ \int_0^\infty \log dR(u | X_i) \bar{Y}_i(u) dN_{i1}(u) - \int_0^\infty E(U_{i1} | \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)}) dR(u | X_i) \right\}. \quad (2.17)$$

An estimating equation for $dR_0(s)$ can be obtained similarly as described in Stage 1:

$$\sum_{i=1}^m \bar{Y}_i(s) [dN_{i1}(s) - E(U_{i1} \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)}) dR_0(s) \exp(X_i' \beta_1)] = 0, \quad s \geq 0.$$

Solving the estimating equation for $dR_0(s; \beta_1)$ at a fixed β_1 value gives us

$$dR_0^{(r)}(s; \beta_1) = \frac{\sum_{i=1}^m \bar{Y}_i(s) dN_{i1}(s)}{\sum_{i=1}^m \bar{Y}_i(s) E(U_{i1} \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)}) \exp(X_i' \beta_1)}. \quad (2.18)$$

We then take the derivative of (2.17) w.r.t. to β_1 and substitute $dR_0^{(r)}(s; \beta_1)$ to obtain a profile score function for β_1

$$U(\beta_1; \theta^{(r-1)}) = \sum_{i=1}^m \left\{ \int_0^\infty \bar{Y}_i(s) \left[X_i - \frac{\sum_{l=1}^m \bar{Y}_l(s) X_l E(U_{l1} \mid \bar{\mathcal{H}}_l(w_l); \theta^{(r-1)}) \exp(X_l' \beta_1)}{\sum_{l=1}^m \bar{Y}_l(s) E(U_{l1} \mid \bar{\mathcal{H}}_l(w_l); \theta^{(r-1)}) \exp(X_l' \beta_1)} \right] dN_{i1}(s) \right\}.$$

The estimate $\beta_1^{(r)}$ is obtained by solving $U(\beta_1; \theta^{(r-1)}) = 0$ for β_1 . We can then compute $dR_0^{(r)}(s)$ by substituting β_1 with $\beta_1^{(r)}$ in (2.18).

Step 2: We then maximize $\sum_{i=1}^m \{ \ell_{i2}(\tilde{\theta}_2) + E(\ell_{i3}(\tilde{\theta}_2, \phi_1, \psi) \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)}) \}$ w.r.t. (ϕ_1, ψ) and obtain $(\phi_1^{(r)}, \psi^{(r)})$.

We repeat Steps 1 and 2 until $\tilde{\theta}^{(r)}$ converges and obtain $\tilde{\theta}$.

Again, we use (2.12) from standard estimating function theory and (2.13) to obtain a variance estimate for $\tilde{\theta}$ with $\tilde{\theta}_2 = (d\tilde{H}_0(\cdot), \tilde{\beta}_2')'$, and $\tilde{\theta}_1 = (\tilde{\alpha}_1', \tilde{\beta}_1')'$ if the recurrent event process is parametric or $\tilde{\theta}_1 = (d\tilde{R}_0(\cdot), \tilde{\beta}_1')'$ if it is semiparametric. We need to calculate the first stage score and information contribution from the i th individual $\mathcal{U}_{i1}(\theta_2)$ and $-\partial \mathcal{U}_{i1}(\theta_2)/\partial \theta'$. We also need to calculate the corresponding complete data score and information, $\mathcal{S}_i(\theta)$ and $\mathcal{I}_i(\theta)$, to get the observed data score and information $S_i(\theta)$ and $I_i(\theta)$, and thus the second stage score and information $\mathcal{U}_{i2}(\theta)$ and $-\partial \mathcal{U}_{i2}(\theta)/\partial \theta'$.

2.4 Simulation Studies

In this section we conduct simulation studies to assess and compare finite sample performance of different estimation methods described in Section 2.2 and 2.3, particularly the simultaneous estimation (for parametric analysis) and the two-stage estimation. We consider the true model specifications described in Section 2.2.2.2 and set $X_i = 1$ with

probability 0.5 and $X_i = 0$ otherwise. We still set $\beta_1 = \log 0.8$ and $\beta_2 = \log 0.9$, and vary $\phi_1 = 0.2, 0.4$ and $\tau = 0, 0.2, 0.4$. For each combination of ϕ_1 and τ , we then solve for ρ_0 and h_0 to ensure that $E\{N_{i1}(1)|X_i = 0\} = 2$ and $P(T_i \leq 1|X_i = 0) = 0.6$. The implied parameter values for each scenario are summarized in Table 2.1. We aim to assess estimators of parameters of interest. Some are specified directly (i.e., for $\beta_1, \beta_2, \phi_1, \psi$) and some are computable by plugging in estimates of all parameters into their expressions (i.e., $E(N_{i1}(1)|X_i = 0), \mathcal{F}(1|X_i = 0)$).

Table 2.1: Parameter specifications for each parameter setting

Scenario	ϕ_1	Kendall's τ	ψ	θ_1		θ_2	
				ρ_0	β_1	h_0	β_2
1	0.2	0.0	0.0	3.0543	-0.2231	0.9163	-0.1054
2	0.2	0.2	0.3090	3.2769	-0.2231	0.9163	-0.1054
3	0.2	0.4	0.5878	3.5135	-0.2231	0.9163	-0.1054
4	0.4	0.0	0.0	3.0543	-0.2231	0.9163	-0.1054
5	0.4	0.2	0.3090	3.3732	-0.2231	0.9163	-0.1054
6	0.4	0.4	0.5878	3.7368	-0.2231	0.9163	-0.1054

2.4.1 Parametric Analysis with Cox Piecewise Constant Baseline Hazard Models

We simulate a data set of $m = 1000$ individuals following the data generation procedure described in Appendix A.3. We conduct fully parametric analyses by using a Gaussian copula and fitting a proportional hazard model with a piecewise constant baseline hazard function for both the recurrent and terminal event processes. The number of pieces of the baseline hazard function for the recurrent and terminal events are denoted by K_1 and K_2 respectively, and we consider $K_1 = K_2 = 1, 4, 8$. Detailed calculations of the observed and complete data likelihood, complete data scores, and complete data information under this fully parametric joint model are presented in Appendix A.4. We investigate the finite sample performance of proposed estimators via simultaneous estimation and the two-stage estimation by evaluating and comparing the empirical bias (EBIAS), empirical standard error (ESE), average asymptotic standard error (ASE) and empirical coverage probability (ECP) for parameter estimates of interest. Simulation results based on $n_{\text{sim}} = 500$ replications are summarized in Tables 2.2 and 2.3.

Table 2.2: Tabulation of frequency properties of estimators obtained from a **fully parametric** specification with piecewise constant baseline hazards for both processes under no random censoring (just administrative censoring at 1; estimates are obtained via **direct optimization** of the observed data likelihood **with simultaneous estimation**; $m = 1000$ and $n_{\text{sim}} = 500$)

τ	(K_1, K_2)	Parameter	$\phi_1 = 0.2$					$\phi_1 = 0.4$				
			TRUE	EBIAS	ESE	ASE	ECP%	TRUE	EBIAS	ESE	ASE	ECP%
0	(1, 1)	β_1	-0.2231	-0.0007	0.0601	0.0567	94.2	-0.2231	0.0025	0.0663	0.0644	94.6
		β_2	-0.1054	-0.0038	0.0823	0.0831	94.4	-0.1054	-0.0009	0.0810	0.0831	95.0
		$\log \phi_1$	-1.6094	-0.0059	0.1725	0.1742	95.8	-0.9163	-0.0120	0.1086	0.1180	97.4
		$\log\{(1+\psi)/(1-\psi)\}$	0.0000	0.0080	0.1975	0.1936	95.4	0.0000	0.0026	0.1424	0.1521	96.4
		$E(N_{i1}(1) X_i = 0)$	2.0000	-0.0036	0.0897			2.0000	-0.0066	0.0968		
		$\mathcal{F}(1 X_i = 0)$	0.4000	-0.0005	0.0210			0.4000	-0.0001	0.0210		
	(4, 4)	β_1	-0.2231	-0.0008	0.0601	0.0567	94.6	-0.2231	0.0026	0.0663	0.0645	94.8
		β_2	-0.1054	-0.0038	0.0823	0.0831	95.0	-0.1054	-0.0010	0.0810	0.0831	95.0
		$\log \phi_1$	-1.6094	-0.0049	0.1726	0.1743	96.0	-0.9163	-0.0115	0.1085	0.1182	97.4
		$\log\{(1+\psi)/(1-\psi)\}$	0.0000	0.0098	0.2136	0.2071	95.2	0.0000	0.0016	0.1512	0.1610	96.2
		$E(N_{i1}(1) X_i = 0)$	2.0000	-0.0030	0.0914			2.0000	-0.0063	0.0985		
		$\mathcal{F}(1 X_i = 0)$	0.4000	-0.0007	0.0213			0.4000	-0.0002	0.0216		
	(8, 8)	β_1	-0.2231	-0.0008	0.0601	0.0567	94.6	-0.2231	0.0026	0.0663	0.0645	94.4
		β_2	-0.1054	-0.0038	0.0823	0.0831	94.8	-0.1054	-0.0010	0.0810	0.0831	95.0
		$\log \phi_1$	-1.6094	-0.0050	0.1723	0.1743	96.0	-0.9163	-0.0116	0.1083	0.1182	97.6
		$\log\{(1+\psi)/(1-\psi)\}$	0.0000	0.0100	0.2134	0.2084	95.2	0.0000	0.0022	0.1522	0.1620	96.6
		$E(N_{i1}(1) X_i = 0)$	2.0000	-0.0031	0.0915			2.0000	-0.0064	0.0984		
		$\mathcal{F}(1 X_i = 0)$	0.4000	-0.0006	0.0213			0.4000	-0.0002	0.0217		
0.2	(1, 1)	β_1	-0.2231	-0.0014	0.0597	0.0566	92.6	-0.2231	-0.0018	0.0638	0.0645	95.4
		β_2	-0.1054	0.0006	0.0802	0.0831	96.0	-0.1054	-0.0007	0.0837	0.0830	95.4
		$\log \phi_1$	-1.6094	-0.0153	0.1638	0.1685	97.0	-0.9163	-0.0142	0.1174	0.1117	94.2
		$\log\{(1+\psi)/(1-\psi)\}$	0.6389	0.0114	0.2051	0.2021	94.4	0.6389	0.0018	0.1594	0.1546	94.8
		$E(N_{i1}(1) X_i = 0)$	2.0000	0.0010	0.0947			2.0000	-0.0078	0.0980		
		$\mathcal{F}(1 X_i = 0)$	0.4000	-0.0003	0.0208			0.4000	-0.0010	0.0219		
	(4, 4)	β_1	-0.2231	-0.0016	0.0598	0.0566	93.2	-0.2231	-0.0018	0.0637	0.0645	95.4
		β_2	-0.1054	0.0006	0.0802	0.0831	95.8	-0.1054	-0.0007	0.0838	0.0830	95.2
		$\log \phi_1$	-1.6094	-0.0135	0.1646	0.1689	96.8	-0.9163	-0.0143	0.1174	0.1121	94.4
		$\log\{(1+\psi)/(1-\psi)\}$	0.6389	0.0166	0.2247	0.2174	93.6	0.6389	-0.0009	0.1674	0.1651	96.0
		$E(N_{i1}(1) X_i = 0)$	2.0000	0.0007	0.0958			2.0000	-0.0075	0.0999		
		$\mathcal{F}(1 X_i = 0)$	0.4000	-0.0002	0.0210			0.4000	-0.0010	0.0221		
	(8, 8)	β_1	-0.2231	-0.0017	0.0598	0.0566	92.8	-0.2231	-0.0018	0.0637	0.0645	95.4
		β_2	-0.1054	0.0006	0.0802	0.0831	95.8	-0.1054	-0.0007	0.0838	0.0830	95.2
		$\log \phi_1$	-1.6094	-0.0131	0.1648	0.1689	96.8	-0.9163	-0.0145	0.1175	0.1121	94.2
		$\log\{(1+\psi)/(1-\psi)\}$	0.6389	0.0194	0.2257	0.2189	93.6	0.6389	-0.0019	0.1671	0.1661	95.2
		$E(N_{i1}(1) X_i = 0)$	2.0000	0.0006	0.0962			2.0000	-0.0075	0.1000		
		$\mathcal{F}(1 X_i = 0)$	0.4000	-0.0002	0.0210			0.4000	-0.0010	0.0222		
0.4	(1, 1)	β_1	-0.2231	-0.0006	0.0592	0.0565	93.2	-0.2231	-0.0004	0.0622	0.0642	95.6
		β_2	-0.1054	-0.0030	0.0804	0.0831	95.6	-0.1054	-0.0011	0.0823	0.0830	96.2
		$\log \phi_1$	-1.6094	-0.0122	0.1564	0.1646	97.8	-0.9163	-0.0126	0.1064	0.1078	95.6
		$\log\{(1+\psi)/(1-\psi)\}$	1.3486	-0.0095	0.2065	0.2254	95.8	1.3486	-0.0129	0.1675	0.1655	94.8
		$E(N_{i1}(1) X_i = 0)$	2.0000	-0.0027	0.0892			2.0000	-0.0053	0.0881		
		$\mathcal{F}(1 X_i = 0)$	0.4000	-0.0006	0.0214			0.4000	-0.0002	0.0218		
	(4, 4)	β_1	-0.2231	-0.0007	0.0592	0.0565	93.0	-0.2231	-0.0005	0.0621	0.0642	95.6
		β_2	-0.1054	-0.0030	0.0804	0.0831	95.6	-0.1054	-0.0013	0.0823	0.0830	96.4
		$\log \phi_1$	-1.6094	-0.0125	0.1605	0.1679	97.6	-0.9163	-0.0134	0.1079	0.1099	95.4
		$\log\{(1+\psi)/(1-\psi)\}$	1.3486	-0.0136	0.2247	0.2387	94.4	1.3486	-0.0148	0.1762	0.1766	95.2
		$E(N_{i1}(1) X_i = 0)$	2.0000	-0.0030	0.0910			2.0000	-0.0047	0.0912		
		$\mathcal{F}(1 X_i = 0)$	0.4000	-0.0006	0.0215			0.4000	-0.0003	0.0222		
	(8, 8)	β_1	-0.2231	-0.0007	0.0592	0.0565	93.0	-0.2231	-0.0005	0.0621	0.0642	95.2
		β_2	-0.1054	-0.0029	0.0804	0.0831	95.6	-0.1054	-0.0012	0.0823	0.0830	96.4
		$\log \phi_1$	-1.6094	-0.0116	0.1612	0.1681	96.8	-0.9163	-0.0136	0.1084	0.1101	95.4
		$\log\{(1+\psi)/(1-\psi)\}$	1.3486	-0.0109	0.2250	0.2400	94.6	1.3486	-0.0151	0.1767	0.1777	95.4
		$E(N_{i1}(1) X_i = 0)$	2.0000	-0.0032	0.0911			2.0000	-0.0050	0.0914		
		$\mathcal{F}(1 X_i = 0)$	0.4000	-0.0005	0.0216			0.4000	-0.0002	0.0221		

Table 2.3: Tabulation of frequency properties of estimators obtained from a **fully parametric** specification with piecewise constant baseline hazards for both processes under no random censoring (just administrative censoring at 1; estimates are obtained via **direct optimization** of the observed data likelihood **with two-stage estimation**; $m = 1000$ and $n_{\text{sim}} = 500$)

τ	(K_1, K_2)	Parameter	$\phi_1 = 0.2$					$\phi_1 = 0.4$				
			TRUE	EBIAS	ESE	ASE	ECP%	TRUE	EBIAS	ESE	ASE	ECP%
0	(1, 1)	β_1	-0.2231	-0.0007	0.0601	0.0567	94.4	-0.2231	0.0025	0.0663	0.0645	94.6
		β_2	-0.1054	-0.0038	0.0823	0.0831	94.6	-0.1054	-0.0008	0.0810	0.0831	94.8
		$\log \phi_1$	-1.6094	-0.0059	0.1725	0.1734	95.6	-0.9163	-0.0120	0.1086	0.1175	97.0
		$\log\{(1 + \psi)/(1 - \psi)\}$	0.0000	0.0080	0.1974	0.1939	95.4	0.0000	0.0026	0.1424	0.1522	96.4
		$E(N_{i1}(1) X_i = 0)$	2.0000	-0.0036	0.0897			2.0000	-0.0066	0.0968		
		$\mathcal{F}(1 X_i = 0)$	0.4000	-0.0005	0.0210			0.4000	-0.0001	0.0210		
	(4, 4)	β_1	-0.2231	-0.0008	0.0601	0.0567	94.4	-0.2231	0.0026	0.0663	0.0646	94.6
		β_2	-0.1054	-0.0038	0.0823	0.0832	94.8	-0.1054	-0.0009	0.0811	0.0832	94.8
		$\log \phi_1$	-1.6094	-0.0049	0.1726	0.1735	95.6	-0.9163	-0.0115	0.1085	0.1177	96.8
		$\log\{(1 + \psi)/(1 - \psi)\}$	0.0000	0.0098	0.2135	0.2074	94.6	0.0000	0.0016	0.1511	0.1611	96.2
		$E(N_{i1}(1) X_i = 0)$	2.0000	-0.0030	0.0914			2.0000	-0.0063	0.0985		
		$\mathcal{F}(1 X_i = 0)$	0.4000	-0.0007	0.0213			0.4000	-0.0002	0.0216		
	(8, 8)	β_1	-0.2231	-0.0008	0.0601	0.0567	94.4	-0.2231	0.0026	0.0663	0.0646	94.4
		β_2	-0.1054	-0.0038	0.0823	0.0832	94.6	-0.1054	-0.0009	0.0811	0.0832	95.0
		$\log \phi_1$	-1.6094	-0.0050	0.1723	0.1735	95.8	-0.9163	-0.0116	0.1083	0.1177	96.8
		$\log\{(1 + \psi)/(1 - \psi)\}$	0.0000	0.0100	0.2133	0.2086	95.4	0.0000	0.0022	0.1521	0.1621	96.2
		$E(N_{i1}(1) X_i = 0)$	2.0000	-0.0031	0.0915			2.0000	-0.0064	0.0984		
		$\mathcal{F}(1 X_i = 0)$	0.4000	-0.0006	0.0213			0.4000	-0.0001	0.0217		
	(1, 1)	β_1	-0.2231	-0.0014	0.0597	0.0565	92.6	-0.2231	-0.0018	0.0638	0.0646	95.4
		β_2	-0.1054	0.0006	0.0803	0.0831	95.8	-0.1054	-0.0009	0.0839	0.0831	95.0
		$\log \phi_1$	-1.6094	-0.0153	0.1638	0.1674	96.4	-0.9163	-0.0142	0.1174	0.1111	94.0
		$\log\{(1 + \psi)/(1 - \psi)\}$	0.6389	0.0113	0.2051	0.2014	94.2	0.6389	0.0017	0.1594	0.1547	95.0
		$E(N_{i1}(1) X_i = 0)$	2.0000	0.0010	0.0948			2.0000	-0.0078	0.0979		
		$\mathcal{F}(1 X_i = 0)$	0.4000	-0.0003	0.0208			0.4000	-0.0010	0.0219		
	(4, 4)	β_1	-0.2231	-0.0016	0.0598	0.0566	93.0	-0.2231	-0.0018	0.0636	0.0646	95.4
		β_2	-0.1054	0.0006	0.0804	0.0831	95.8	-0.1054	-0.0009	0.0839	0.0831	95.2
		$\log \phi_1$	-1.6094	-0.0135	0.1646	0.1678	96.0	-0.9163	-0.0143	0.1174	0.1115	94.4
		$\log\{(1 + \psi)/(1 - \psi)\}$	0.6389	0.0164	0.2247	0.2166	93.8	0.6389	-0.0012	0.1673	0.1652	96.0
		$E(N_{i1}(1) X_i = 0)$	2.0000	0.0007	0.0959			2.0000	-0.0075	0.0999		
		$\mathcal{F}(1 X_i = 0)$	0.4000	-0.0002	0.0211			0.4000	-0.0010	0.0221		
	(8, 8)	β_1	-0.2231	-0.0017	0.0598	0.0566	92.8	-0.2231	-0.0018	0.0637	0.0646	95.2
		β_2	-0.1054	0.0006	0.0804	0.0831	95.8	-0.1054	-0.0009	0.0839	0.0831	95.2
		$\log \phi_1$	-1.6094	-0.0132	0.1648	0.1678	95.6	-0.9163	-0.0145	0.1175	0.1115	94.4
		$\log\{(1 + \psi)/(1 - \psi)\}$	0.6389	0.0191	0.2256	0.2181	93.6	0.6389	-0.0022	0.1670	0.1662	95.2
		$E(N_{i1}(1) X_i = 0)$	2.0000	0.0006	0.0962			2.0000	-0.0075	0.1000		
		$\mathcal{F}(1 X_i = 0)$	0.4000	-0.0002	0.0211			0.4000	-0.0010	0.0221		
0.4	(1, 1)	β_1	-0.2231	-0.0007	0.0592	0.0564	93.4	-0.2231	-0.0005	0.0621	0.0644	95.6
		β_2	-0.1054	-0.0031	0.0804	0.0831	95.6	-0.1054	-0.0016	0.0825	0.0832	96.8
		$\log \phi_1$	-1.6094	-0.0123	0.1563	0.1626	97.0	-0.9163	-0.0127	0.1063	0.1067	95.4
		$\log\{(1 + \psi)/(1 - \psi)\}$	1.3486	-0.0097	0.2063	0.2104	92.6	1.3486	-0.0132	0.1674	0.1615	93.4
		$E(N_{i1}(1) X_i = 0)$	2.0000	-0.0030	0.0892			2.0000	-0.0055	0.0883		
		$\mathcal{F}(1 X_i = 0)$	0.4000	-0.0009	0.0214			0.4000	-0.0003	0.0217		
	(4, 4)	β_1	-0.2231	-0.0007	0.0592	0.0564	93.0	-0.2231	-0.0006	0.0620	0.0644	95.0
		β_2	-0.1054	-0.0031	0.0805	0.0831	95.6	-0.1054	-0.0017	0.0825	0.0832	97.0
		$\log \phi_1$	-1.6094	-0.0126	0.1605	0.1661	97.0	-0.9163	-0.0134	0.1078	0.1088	95.2
		$\log\{(1 + \psi)/(1 - \psi)\}$	1.3486	-0.0140	0.2247	0.2216	91.0	1.3486	-0.0152	0.1763	0.1720	93.4
		$E(N_{i1}(1) X_i = 0)$	2.0000	-0.0034	0.0911			2.0000	-0.0052	0.0915		
		$\mathcal{F}(1 X_i = 0)$	0.4000	-0.0008	0.0215			0.4000	-0.0004	0.0220		
	(8, 8)	β_1	-0.2231	-0.0007	0.0592	0.0564	93.0	-0.2231	-0.0006	0.0620	0.0644	95.2
		β_2	-0.1054	-0.0030	0.0804	0.0831	95.6	-0.1054	-0.0017	0.0825	0.0832	97.0
		$\log \phi_1$	-1.6094	-0.0117	0.1612	0.1663	96.6	-0.9163	-0.0136	0.1084	0.1090	95.2
		$\log\{(1 + \psi)/(1 - \psi)\}$	1.3486	-0.0114	0.2249	0.2226	91.0	1.3486	-0.0155	0.1767	0.1730	94.4
		$E(N_{i1}(1) X_i = 0)$	2.0000	-0.0037	0.0912			2.0000	-0.0055	0.0917		
		$\mathcal{F}(1 X_i = 0)$	0.4000	-0.0007	0.0216			0.4000	-0.0003	0.0220		

For both the simultaneous and two-stage estimation methods, the empirical biases are small for all parameters of interest, demonstrating the consistency of proposed estimators through either estimation procedure can be achieved within a finite sample size. The empirical standard errors are in good agreement with the average estimated standard errors and the empirical coverage probabilities are close to the nominal level of 95% for either estimation procedure. Meanwhile, the finite sample performance of parameters from two-stage estimation method is comparable to those from the simultaneous estimation in terms of empirical biases, empirical standard errors and average estimated standard errors regardless of the strength of the association, showing that the two-stage estimation procedure is remarkably efficient.

2.4.2 Semiparametric Analysis with a Cox Model for the Terminal Event

Now we conduct semiparametric analyses by fitting a Cox proportional hazard model with a non-parametric baseline hazard function for the terminal event process. For the recurrent event process, we consider a proportional hazard model with either a 8-piece piecewise constant baseline hazard, or a non-parametric baseline hazard. We use a two-stage estimation procedure with an EM algorithm to obtain point estimates and variance estimates. The complete data score and information as well as the two-stage score and information under this semiparametric specification are calculated in details in [Appendix A.5](#).

Simulation results averaged over 500 replications are summarized in [Table 2.4](#). We see that for all parameters of interest from the semiparametric analyses, the empirical biases are small, the empirical and average estimated standard errors are in good agreement, and the empirical coverage probabilities are close to the nominal 95% level.

[Tables 2.2](#), [2.3](#) and [2.4](#) all demonstrate the validity of our model framework even when there is no dependence between the recurrent and terminal event processes. Moreover, for all scenarios of parameter settings, all point estimates and variance estimates are close between parametric and semiparametric model specifications.

Table 2.4: Tabulation of frequency properties of estimators from either a parametric (with an 8-piece piecewise constant baseline) or a semiparametric recurrent event process, and a **semiparametric terminal event** process with administrative censoring at 1 (no random censoring); estimates are obtained using a **two-stage EM** algorithm; $m = 1000$ and $n_{\text{sim}} = 500$

τ	Parameter	$\phi_1 = 0.2$										$\phi_1 = 0.4$									
		8-Piece PWC (EM)					semiparametric (EM)					8-Piece PWC (EM)					semiparametric (EM)				
		TRUE	EBIAS	ESE	ASE	ECP%	EBIAS	ESE	ASE	ECP%	TRUE	EBIAS	ESE	ASE	ECP%	EBIAS	ESE	ASE	ECP%		
0	β_1	-0.2231	-0.0007	0.0601	0.0567	94.4	-0.0008	0.0601	0.0567	94.6	-0.2231	0.0026	0.0662	0.0646	94.4	0.0026	0.0662	0.0646	94.4		
	β_2	-0.1054	-0.0037	0.0822	0.0831	94.6	-0.0037	0.0822	0.0831	94.6	-0.1054	-0.0009	0.0811	0.0831	95.0	-0.0009	0.0811	0.0831	95.0		
	$\log \phi_1$	-1.6094	-0.0035	0.1720	0.1739	96.0	-0.0041	0.1720	0.1741	96.0	-0.9163	-0.0065	0.1113	0.1181	96.2	-0.0070	0.1112	0.1182	96.2		
	$\log\{(1+\psi)/(1-\psi)\}$	0.0000	0.0092	0.2132	0.2086	95.4	0.0102	0.2129	0.2091	95.8	0.0000	0.0006	0.1524	0.1624	96.4	0.0017	0.1524	0.1628	96.4		
	$E(N_{it}(1) X_i = 0)$	2.0000	-0.0026	0.0915			-0.0026	0.0915			2.0000	-0.0060	0.0984			-0.0060	0.0984				
	$\mathcal{F}(1 X_i = 0)$	0.4000	-0.0003	0.0213			-0.0008	0.0213			0.4000	0.0002	0.0216			0.0002	0.0216				
0.2	β_1	-0.2231	-0.0017	0.0598	0.0566	92.8	-0.0017	0.0598	0.0566	92.8	-0.2231	-0.0018	0.0637	0.0646	95.2	-0.0018	0.0637	0.0646	95.4		
	β_2	-0.1054	0.0007	0.0803	0.0830	95.8	0.0007	0.0803	0.0830	95.8	-0.1054	-0.0008	0.0839	0.0830	95.2	-0.0008	0.0839	0.0830	95.2		
	$\log \phi_1$	-1.6094	-0.0108	0.1645	0.1658	95.0	-0.0111	0.1643	0.1660	95.2	-0.9163	-0.0112	0.1175	0.1110	94.6	-0.0114	0.1175	0.1111	94.6		
	$\log\{(1+\psi)/(1-\psi)\}$	0.6389	0.0185	0.2253	0.2176	93.4	0.0203	0.2262	0.2183	93.2	0.6389	-0.0029	0.1672	0.1663	95.0	-0.0019	0.1676	0.1668	95.0		
	$E(N_{it}(1) X_i = 0)$	2.0000	0.0012	0.0961			0.0012	0.0961			2.0000	-0.0069	0.0999			-0.0068	0.0999				
	$\mathcal{F}(1 X_i = 0)$	0.4000	0.0001	0.0210			0.0001	0.0210			0.4000	-0.0007	0.0222			-0.0007	0.0222				
0.4	β_1	-0.2231	-0.0008	0.0592	0.0564	93.0	-0.0008	0.0592	0.0565	93.2	-0.2231	-0.0007	0.0620	0.0644	95.2	-0.0006	0.0621	0.0645	95.2		
	β_2	-0.1054	-0.0030	0.0804	0.0831	95.6	-0.0030	0.0804	0.0831	95.6	-0.1054	-0.0016	0.0824	0.0831	97.0	-0.0016	0.0824	0.0831	97.0		
	$\log \phi_1$	-1.6094	0.0069	0.1584	0.1510	94.0	0.0068	0.1585	0.1512	93.8	-0.9163	0.0025	0.1093	0.1041	93.4	0.0023	0.1099	0.1043	93.6		
	$\log\{(1+\psi)/(1-\psi)\}$	1.3486	-0.0132	0.2237	0.2259	93.0	-0.0122	0.2241	0.2265	93.4	1.3486	-0.0142	0.1781	0.1758	94.2	-0.0138	0.1776	0.1763	95.4		
	$E(N_{it}(1) X_i = 0)$	2.0000	-0.0031	0.0911			-0.0031	0.0911			2.0000	-0.0054	0.0919			-0.0054	0.0919				
	$\mathcal{F}(1 X_i = 0)$	0.4000	-0.0004	0.0216			-0.0004	0.0216			0.4000	0.0000	0.0220			0.0000	0.0220				

2.5 Efficacy of Pamidronate in Reducing Skeletal Complications

Bone metastases occur in most women with advanced breast cancer. Cytotoxic chemotherapy or hormone therapy is the preferred treatment for symptomatic bone disease, but progressive skeletal destruction ultimately leads to increased pain, immobility, and deterioration in the quality of life. Bisphosphonates such as pamidronate disodium are powerful inhibitors of osteoclast-induced bone resorption associated with cancer that has metastasized to bone.

A randomized, double-blind trial was conducted to study the efficacy of pamidronate in preventing skeletal complications in patients with bone lesions due to metastatic breast cancer ([Hortobagyi et al., 1996](#)). In this trial, 382 women with stage IV breast cancer who were receiving cytotoxic chemotherapy and had at least one lytic bone lesion were enrolled across 97 study sites. The participants were randomly assigned to receive either a placebo or pamidronate infusion each month for 12 cycles, with 185 of them receiving pamidronate and 197 receiving placebo. During the 12-month trial, patients were administered either a placebo or pamidronate (90mg) as a two-hour intravenous infusion monthly for 12 cycles. The clinical features of patients when they entered the study were comparable between the placebo group and the pamidronate group. Figure 2.4 displays the timing and number of skeletal related events (SRE) for ten patients from each treatment group. This figure illustrates the fact that each patient may experience more than one skeletal complication prior to death, and that some deaths are right-censored. The mean duration of participation in the study was 9.6 months in the pamidronate group and 8.9 months in the placebo group. The median time to the first skeletal complication was 7.0 months for the placebo group and 13.1 months in the pamidronate group. 6 patients were lost to follow up and 326 were observed to die.

To account for the fact that death precludes the occurrence of further skeletal complications, we apply our proposed framework to conduct a joint analysis of skeletal complications and the survival process. The association between the two processes is accommodated by a Gaussian copula with Kendall's τ , and the random effect is specified to be a Gamma distribution. We consider the treatment of receiving either pamidronate or placebo as the sole covariate and fit a fully parametric model with a 8-piece piecewise constant baseline rate, as well as a semiparametric model with a non-parametric baseline rate for both processes. For the fully parametric model, we employ the two-stage estimation method with direct optimization of the likelihoods; for the semiparametric model, we use the two-stage estimation procedure with an EM algorithm.

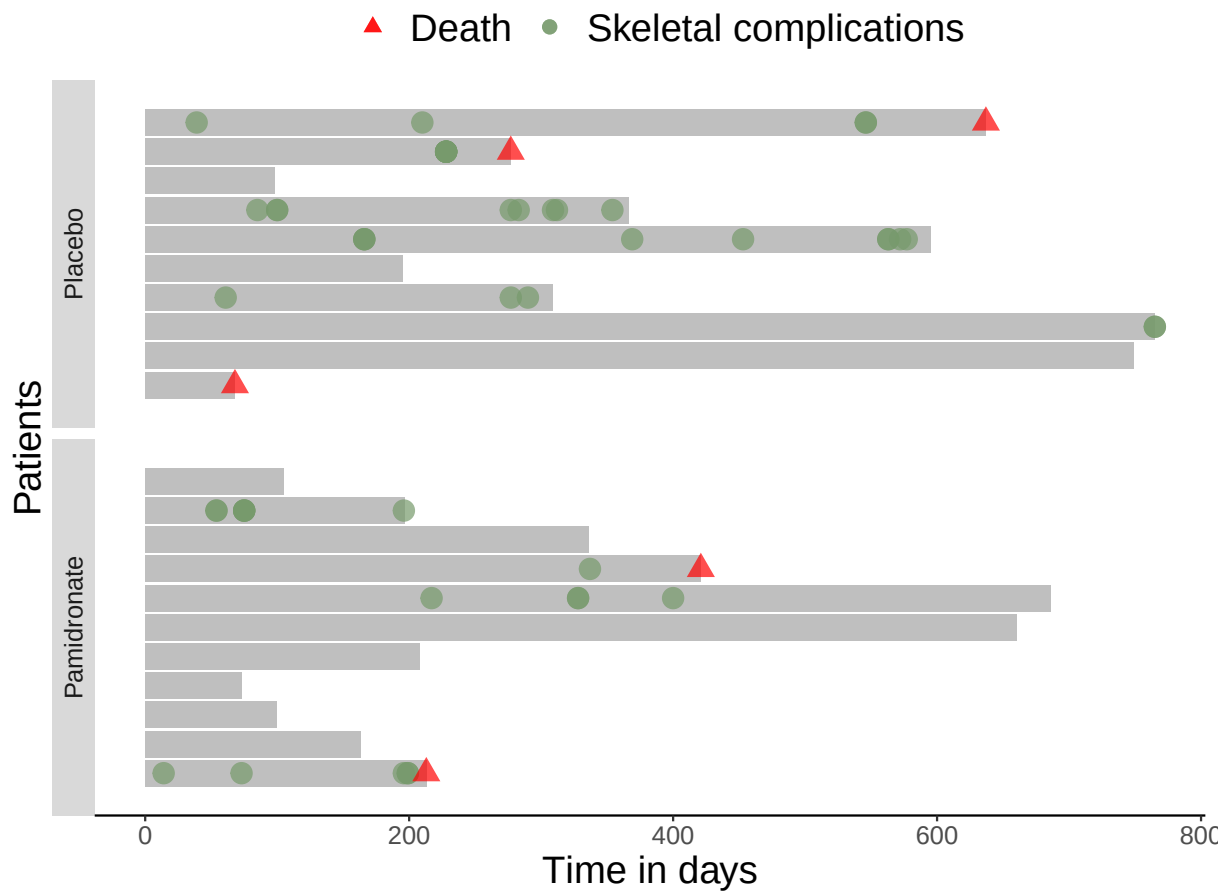


Figure 2.4: Sample profile features for various skeletal complications and survival times in the Pamidronate and placebo groups; green circles represent skeletal complications and red triangles represent observed deaths

The results obtained from the joint analysis of the SRE and the survival processes are summarized in Table 2.5. We observe that there is broad agreement between the fully parametric and semiparametric models in terms of point estimate (EST), standard error (SE) and bootstrap standard error (BSE), which demonstrates that parametric 8-piece piecewise constant baseline rate models are satisfactorily flexible. The treatment of receiving monthly intravenous infusion of pamidronate significantly reduces the occurrence of skeletal complications in patients with advanced breast cancer ($\exp(\hat{\beta}_1) = \exp(-0.4564) = 0.634$, 95% CI: (0.426, 0.943), $p = 0.024$ with the parametric model; $\exp(\hat{\beta}_1) = \exp(-0.4580) = 0.633$, 95% CI: (0.425, 0.942), $p = 0.024$ with the semiparametric model), while it doesn't have significant effects on the survival process among these patients. There is evidence of heterogeneity in the risk of skeletal complications ($\phi_1 = 2.457$, 95% CI: (1.838, 3.283) with the parametric model; $\phi_1 = 2.253$, 95% CI: (1.757, 2.890) with the semiparametric model) and there is a positive association between skeletal complications and death within patients who have bone lesions due to metastatic breast cancer ($\tau = 0.337$, 95% CI: (0.232, 0.433), $p < 0.001$ with the parametric model; $\tau = 0.310$, 95% CI: (0.196, 0.415), $p < 0.001$ with the semiparametric model). Figure 2.5 shows the estimated cumulative conditional and cumulative marginal rates for the SRE process (left panel) and the estimated cumulative death hazard and survival function (right panel), obtained from both parametric and semiparametric model specifications. The conditional cumulative rate for the SRE process is calculated given that the individual is alive and the random effect is specified to be 1, i.e., $u_{i1} = 1$. The marginal cumulative rate is an integrated version that averages over the random effect and the survival status. The marginal cumulative rate is smaller compared to the conditional one because death terminates the SRE process. The figure also indicates a decent agreement between parametric and semiparametric model specifications.

2.6 Discussion

We proposed a new joint model for recurrent and terminal events involving a copula function used to link a terminal event with a random effect that modulates a mixed Poisson process. Estimation and inference methods including a computationally more efficient two-stage procedure which enables semiparametric specifications have been developed with and without the use of an EM algorithm. Under a Gaussian copula and a Gamma distributed random effect specification, plots of conditional recurrent and terminal event intensities given the full history have been investigated to gain insight about the nature of the joint processes under our model formulation; numerical studies demonstrate satisfactory frequency properties of estimators obtained from both simultaneous and two-stage proce-

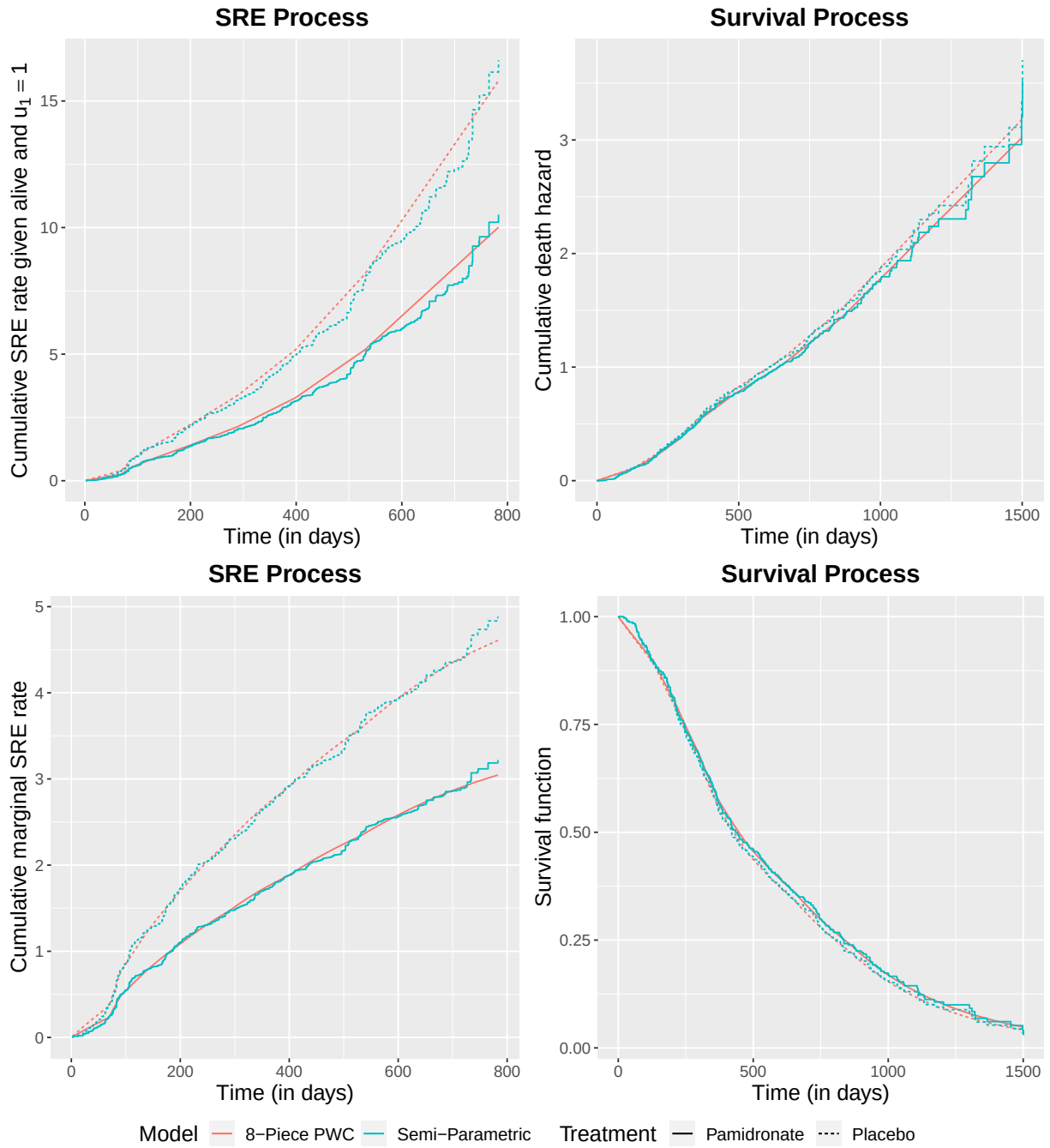


Figure 2.5: Plots of estimated conditional (on being alive and $u_{i1} = 1$) and marginal cumulative rates for the SRE process (left panel), and plots of estimated cumulative hazard and survival function for the survival process (right panel)

Table 2.5: Joint analysis of the SRE and the survival process with a Gaussian copula, a Gamma distributed random effect and a single binary covariate indicating receipt of pamidronate rather than placebo; both parametric (with 8-piece PWC baseline rates) and semiparametric models are considered (all estimates are obtained via a two-stage procedure while the semiparametric estimation exploits the EM algorithm; BSE is the bootstrap standard error of the estimates obtained from 100 bootstrap runs)

Type	Parameter	8-Piece PWC (Direct)			semiparametric (EM)		
		EST	SE	BSE	EST	SE	BSE
Pamidronate (on SRE)	β_1	-0.4564	0.1947	0.2027	-0.4580	0.1751	0.2034
Pamidronate (on death)	β_2	-0.0515	0.1095	0.1184	-0.0502	0.1101	0.1201
Variance of the Random Effect	$\log \phi_1$	0.8988	0.1587	0.1480	0.8124	0.1233	0.1270
Copula Parameter	$\log \frac{1+\psi}{1-\psi}$	1.1130	0.1710	0.1873	1.0159	0.1501	0.1993

dures in finite samples and the high relative efficiency of the two-stage procedure makes the semiparametric analysis practical. We note the copula function can be any copula function from well-studied copula families and the marginal distribution of the random effect can be other distributions, e.g., a log-normal distribution.

Compared to the joint modelling of the recurrent and terminal events based on shared or correlated random effects, our model preserves the neat marginal proportional hazard property for the terminal event process. Furthermore, our approach distinguishes itself by addressing unobservable heterogeneity in recurrent events and capturing the interdependence between the two processes through distinct structures. Notably, it remains valid even in situations where there is no association between the two event processes. An alternative framework of joint analysis via marginal approaches though is robust, it doesn't exploit the history of the joint processes that leads to more accurate and precise predictions of features of the recurrent event process.

As our model accommodates the dependence structure using a pre-specified copula function, a natural next step is to investigate the sensitivity of our model to misspecification of the copula function, particularly in terms of coefficient estimation. Copula models indexed by more than one parameter could be used, but it may require more data for accurate parameter estimation. Furthermore, this general model can be readily extended to scenarios involving multiple types of recurrent events and a terminal event process by increasing the dimension of the copula function. Gaussian copulas, vine copulas or

hierarchical copulas might be appealing in this context. Another potential extension is to adapt our work to adjust for recurrent events with intermittent observations; see [Chen et al. \(2005\)](#). This is developed in Chapter [3](#). Additionally, one may also be interested in sample size estimation when designing clinical trials with these endpoints in mind.

Chapter 3

A Joint Model for Recurrent and Terminal Events with Intermittent Observation

3.1 Introduction

Recurrent events are of interest in many fields of study. In health studies, there are many chronic disease processes for which affected individuals experience recurrent adverse events. For instance, individuals with chronic bronchitis frequently experience recurrent exacerbations (Grossman et al., 1998). Various statistical methods have been developed for analyzing recurrent events, including intensity-based models (Gail et al., 1980; Prentice et al., 1981; Andersen and Gill, 1982), random effect models (Lawless, 1987), and marginal methods (Lawless and Nadeau, 1995; Lin et al., 2000).

In many settings, interest lies in characterizing the incidence of recurrent events in conjunction with an event which terminates the recurrent event process. For example, in organ transplant studies, transient graft rejection episodes are terminated by total graft rejection or patient death. When the terminal event is related to the recurrent events, joint modelling of both events is necessary for valid statistical inference. Existing methodologies for joint analysis of recurrent and terminal events can be broadly divided into two categories: marginal methods and shared or correlated random effects models. Marginal methods involve specification of a marginal distribution of the termination time and a rate function for the recurrent events conditional on the termination time or survival to the time of interest (Cook and Lawless, 1997; Ghosh and Lin, 2002). Shared or correlated

random effects models account for the dependence between recurrent and terminal events by incorporating a shared random effect or correlated random effects into the respective event rates (Wang et al., 2001; Huang and Wang, 2004; Liu et al., 2004; Ye et al., 2007). See Chapter 6 of Cook and Lawless (2007) for future discussion of these methods.

In some settings the occurrence of events is not evident, but rather the event count can only be determined upon radiographic examination, some other form of imaging, or detailed clinical examination. Examples include the development of new tumors in bladder cancer patients (Byar, 1980), the occurrences of joint damage in arthritic patients, and the development of new skeletal metastases in cancer patients with bone metastasis (Hortobagyi et al., 1996). In such scenarios, the exact times of event occurrence are not observed; instead, only the number of events occurring during successive assessment periods is available. This type of data is often referred to as panel count data or interval-grouped recurrent events data (Lawless and Zhan, 1998).

Thall and Lachin (1988) proposed a nonparametric approach to compare the recurrence rates of two treatment groups using panel count data. Sun and Kalbfleisch (1993, 1995) and Wellner and Zhang (2000) explored nonparametric estimation of rate and mean functions in Poisson processes from interval-count data. Sun and Matthews (1997) and Sun and Wei (2000) conducted semiparametric estimation of regression parameters without estimating baseline rate or mean functions. Other methods for semiparametric regression analysis of interval-censored recurrent events have also been developed (Zhang, 2002; Zhang and Jamshidian, 2003; Wellner et al., 2004). Lawless and Zhan (1998) considered multiplicative recurrent event models with piecewise constant baseline rate functions, fitted using semiparametric methods via estimating functions as well as fully specified random effect models fitted using maximum likelihood. Such piecewise constant models share the advantages of parametric models while providing some robustness against misspecification of the parametric form of rate functions. Regression analysis on interval-censored recurrent events with informative observation times has been investigated by Huang et al. (2006); Sun et al. (2007); Zhao and Tong (2011).

While there is extensive literature on joint modeling of recurrent and terminal events and on analysis of recurrent events with interval-censored data, there has been limited discussion on the joint analysis of recurrent and terminal events subject to intermittent observations. Chen et al. (2005) developed a framework for handling multivariate interval-censored recurrent events; however, the events considered in their study do not terminate each other.

In Chapter 2, we developed a copula-based framework for the joint analysis of recurrent and terminal event processes. In this chapter, we briefly review the model and describe

a framework for analysis to situations where recurrent events are subject to intermittent observations. We achieve this by specifying piecewise constant baseline rate functions for the recurrent events, following the approach of [Lawless and Zhan \(1998\)](#). A semiparametric analysis for the terminal event is feasible via a two-stage estimation procedure and the simple marginal interpretation of the regression coefficients of the Cox model for the terminal events is retained.

The remainder of this chapter is organized as follows. In [Section 3.2](#), we briefly describe the copula-based model and specify a piecewise constant baseline rate for recurrent events under intermittent observation. Besides direct optimization of the observed data likelihood, a complete data likelihood is presented along with details on the implementation of an EM algorithm ([Dempster et al., 1977](#)) for estimation and inference. In [Section 3.3](#), we describe an alternative two-stage estimation procedure that facilitates semiparametric modeling. [Section 3.4](#) investigates the finite sample performance of the proposed estimators through simulation studies. In [Section 3.5](#), we apply our model to the University of Toronto Psoriatic Arthritis Cohort Study and investigate the effect of HLA-B27 on the occurrence of joint damages in patients with psoriatic arthritis. Concluding remarks are made in [Section 3.6](#).

3.2 A Copula-Based Joint Model

3.2.1 Model Framework

For individual i in the sample, we let $N_{i1}(s)$ denote the number of recurrent events over $(0, s]$, and $\{N_{i1}(s), 0 < s\}$ denote the corresponding counting process. Let $\Delta N_{i1}(s) = N_{i1}(s + \Delta s^-) - N_{i1}(s^-)$ denote the number of recurrent events over a small interval $[s, s + \Delta s)$ and $dN_{i1}(s) = \lim_{\Delta s \downarrow 0} \Delta N_{i1}(s)$. Let T_i denote a terminal event time at which point the recurrent event process is terminated, and let $N_{i2}(s) = I(T_i \leq s)$ and $\{N_{i2}(s), 0 < s\}$ be the terminal event counting process. If we denote $\Delta N_{i2}(s) = N_{i2}(s + \Delta s^-) - N_{i2}(s^-)$ and $dN_{i2}(s) = \lim_{\Delta s \downarrow 0} \Delta N_{i2}(s)$, then let $dN(s) = (dN_{i1}(s), dN_{i2}(s))'$, $N_i(s) = (N_{i1}(s), N_{i2}(s))'$ be the bivariate counting process for both events.

Let X_i denote a $p \times 1$ vector of covariates and consider a marginal hazard function for T_i denoted by $h(t | X_i; \theta_2)$. The corresponding marginal survival function is then

$$\mathcal{F}(t | X_i; \theta_2) = 1 - F(t | X_i; \theta_2) = \exp(-H(t | X_i; \theta_2)),$$

where $F(t \mid X_i; \theta_2)$ denotes the corresponding c.d.f. and $h(t \mid X_i; \theta_2) = dH(t \mid X_i; \theta_2)/dt$; $f(t \mid X_i; \theta_2)$ is the corresponding p.d.f. We may specify a proportional hazards model, where

$$h(t \mid X_i; \theta_2) = h_0(t; \alpha_2) \exp(X_i' \beta_2)$$

and consider parametric or nonparametric specifications for the baseline hazard $h_0(t)$.

To model the recurrent event process, we introduce i.i.d. latent random effects $U_{i1} > 0$ for each individual with $E(U_{i1}) = 1$, $\text{var}(U_{i1}) = \phi_1$ and c.d.f. $G_1(U_{i1}; \phi_1)$ to accommodate variation in the recurrent event rates between individuals which cannot be explained by available covariates; we assume $U_{i1} \perp X_i$. The rate function for the recurrent events conditional on $U_{i1} = u_{i1}$ and the full history $\mathcal{H}_i(s) = \{N_i(u), 0 < u < s, X_i\}$ is

$$\lim_{\Delta s \downarrow 0} \frac{P(\Delta N_{i1}(s) = 1 \mid \mathcal{H}_i(s), u_{i1}; \theta_1)}{\Delta s}.$$

Under the assumption that $\{N_{i1}(s), 0 < s < T_i\} \perp T_i \mid U_{i1} = u_{i1}, X_i$, we consider a multiplicative rate function for the recurrent event process of the form

$$\lim_{\Delta s \downarrow 0} \frac{P(\Delta N_{i1}(s) = 1 \mid \mathcal{H}_i(s), u_{i1}; \theta_1)}{\Delta s} = Y_i^\dagger(s) u_{i1} \rho(s \mid X_i; \theta_1)$$

with $Y_i^\dagger(s) = \mathbb{1}(s \leq T_i)$. A proportional hazard model may also be specified for $\rho(s \mid X_i; \theta_1)$ with a baseline hazard $\rho_0(s; \alpha_1)$,

$$\rho(s \mid X_i; \theta_1) = \rho_0(s; \alpha_1) \exp(X_i' \beta_1).$$

The last step to complete our model framework is to construct a dependence structure between the two processes by specifying a joint model for the random effect and the terminal event via a copula function (Nelsen, 2006). Let $\mathcal{C}(v; \psi)$ denote a copula function where $V = (V_1, V_2)'$ with V_j a uniform random variable, $v = (v_1, v_2)'$ with v_j the corresponding realizations, and ψ indexes the dependence. That is,

$$\mathcal{C}(v; \psi) = P(V_1 \leq v_1, V_2 \leq v_2; \psi).$$

Setting $V_1 = G_1(U_{i1}; \phi_1)$ and $V_2 = \mathcal{F}(T_i \mid X_i; \theta_2)$, a joint model for U_{i1} and $T_i \mid X_i$ is determined by linking these via

$$P(U_{i1} \leq u_1, T_i > t \mid X_i; \theta_2, \phi_1, \psi) = \mathcal{C}(G_1(u_1; \phi_1), \mathcal{F}(t \mid X_i; \theta_2); \psi).$$

Note that $\theta_1 = (\alpha_1', \beta_1')'$ indexes the recurrent event process, $\theta_2 = (\alpha_2', \beta_2')'$ indexes the terminal event process, ϕ_1 indexes the random effect, and ψ measures the association

between the random effect (and hence the recurrent event process) and the terminal event process. We let $\theta = (\theta'_1, \theta'_2, \phi_1, \psi)'$ denote the set of full parameters of interest. Under this model framework, we assume that the recurrent events and the terminal event are independent conditional on covariates and the random effect, provided that the whole process has not been terminated.

3.2.2 A Piecewise Constant Conditional Baseline Rate

Often in clinical studies, instead of the full history of the bivariate process $\mathcal{H}_i(s)$, we observe incomplete data denoted by $\bar{\mathcal{H}}_i(s)$ for each individual, where we only have intermittent observations of the recurrent event process due to periodic clinical examination. For the terminal event process, the events are of more serious in nature (i.e., death) and so the event times are known subject to right-censoring. In other settings alternative sources of data (e.g., electronic health records) may make this information available.

Let C_i denote an independent and non-informative right-censoring time for individual i , and suppose we observe $W_i = \min(T_i, C_i)$ for the terminal event and let $\delta_i = \mathbb{1}(T_i \leq C_i)$, $Y_i(s) = \mathbb{1}(s \leq C_i)$ and $\bar{Y}_i(s) = Y_i(s)Y_i^\dagger(s)$. We define the “complete data” observable counting process notation for the recurrent and terminal events as $d\bar{N}_{ik}(s) = \bar{Y}_i(s)dN_{ik}(s)$, $\bar{N}_{ik}(s) = \int_0^s d\bar{N}_{ik}(u)$, $k = 1, 2$, and let $d\bar{N}_i(s) = (d\bar{N}_{i1}(s), d\bar{N}_{i2}(s))'$, $\bar{N}_i(s) = (\bar{N}_{i1}(s), \bar{N}_{i2}(s))'$ denote the bivariate counting process notation. The recurrent events are subject to intermittent observation and so this “complete data” is not observed for the recurrent event process. Let $0 = a_{i0} < a_{i1} < \dots < a_{ir_i}$ denote the r_i assessment times for individual i over $(0, W_i]$. We define the counting process notation for the assessments accommodating right-censoring as $\{\bar{A}_i(s), 0 < s\}$ with $\bar{A}_i(s) = \bar{Y}_i(s)A_i(s)$ and $\bar{A}_i(s) = \sum_{j=1}^\infty \mathbb{1}(a_{ij} \leq s)$. We observe a total number of events $n_{ij} = \bar{N}_{i1}(a_{ij}) - \bar{N}_{i1}(a_{i,j-1})$ in the interval $\mathcal{A}_{ij} = (a_{i,j-1}, a_{ij}]$ for $j = 1, \dots, r_i$, where r_i is the number of intervals for which individual i is assessed. Note a_{ij} times vary across individuals and may be stochastic in this case in what follows we assume a conditionally independent visit process (Cook and Lawless, 2021). Note in particular the recurrent events over $(a_{r_i}, W_i]$ is unknown. The observed history for individual i can be explicitly written as $\bar{\mathcal{H}}_i(s) = \{\bar{Y}_i(u), d\bar{N}_{i2}(u), 0 < u < s; (a_{ij}, n_{ij}), j = 1, \dots, \bar{A}_i(s), X_i\}$. Conditioning on $U_{i1} = u_{i1}$ and $\bar{\mathcal{H}}_i(s)$, the recurrent event rate is

$$\lim_{\Delta s \downarrow 0} \frac{P(\Delta \bar{N}_{i1}(s) = 1 \mid \bar{\mathcal{H}}_i(s), u_{i1}; \theta_1)}{\Delta s} = \bar{Y}_i(s)u_{i1}\rho(s \mid X_i; \theta_1).$$

In general, the baseline rate $\rho_0(s; \alpha_1)$ for the recurrent events can be modeled parametrically or nonparametrically. For this project, we specify it to be a piecewise constant

rate function because such models are flexible and robust, and they are much easier to implement than purely nonparametric or semiparametric models, particularly when the observed data are incomplete. We let K_1 denote the number of pieces of the baseline hazard for the recurrent (type 1) events and let $0 = b_{10} < b_{11} < \dots < b_{1K_1} = \infty$ denote a pre-specified sequence of cut points of the positive real line. Let $\delta_{1k}(s) = \mathbb{1}(s \in \mathcal{B}_{1k})$ indicate if time s is in the interval $\mathcal{B}_{1k} = (b_{1,k-1}, b_{1k}]$, and $d_{1k}(s) = \int_0^s \delta_{1k}(u) du$ be the length of the intersection of the interval $(0, s]$ and $\mathcal{B}_{1k}$, for $k = 1, \dots, K_1$. The instantaneous and cumulative baseline recurrent event rates can be written as

$$\rho_0(s; \alpha_1) = \prod_{k=1}^{K_1} \rho_k^{\delta_{1k}(s)} \quad \text{and} \quad R_0(s; \alpha_1) = \int_0^s \rho_0(u; \alpha_1) du = \sum_{k=1}^{K_1} \rho_k d_{1k}(s),$$

respectively with $\alpha_1 = (\rho_1, \dots, \rho_{K_1})'$. Under a model where covariates act multiplicatively on a baseline rate function, we have

$$R(s \mid X_i; \theta_1) = \int_0^s \rho(u \mid X_i; \theta_1) du = R_0(s; \alpha_1) \exp(X_i' \beta_1).$$

The observed data likelihood contributed from individual i can now be written as

$$L_i(\theta) \propto \int_0^\infty P(\bar{\mathcal{H}}_i(w_i^+), U_{i1} = u_{i1}; \theta) du_{i1} \quad (3.1)$$

where $P(\bar{\mathcal{H}}_i(w_i^+), U_{i1} = u_{i1}; \theta)$ is equal to

$$\begin{aligned} & h(w_i \mid X_i; \theta_2)^{\delta_i} \exp(-H(w_i \mid X_i; \theta_2)) \\ & \times P(n_{i1}, \dots, n_{ir_i} \mid X_i, u_{i1}, T_i \geq w_i; \theta_1) \\ & \times \{g_1(u_{i1} \mid X_i, T_i = w_i; \theta_2, \phi_1, \psi)\}^{\delta_i} \{g_1(u_{i1} \mid X_i, T_i > w_i; \theta_2, \phi_1, \psi)\}^{1-\delta_i} \end{aligned} \quad (3.2)$$

with

$$g_1(u_1 \mid X_i, T_i > t; \theta_2, \phi_1, \psi) = \frac{P(U_{i1} = u_1, T_i > t \mid X_i; \theta_2, \phi_1, \psi)}{\mathcal{F}(t \mid X_i; \theta_2)},$$

and

$$g_1(u_1 \mid X_i, T_i = t; \theta_2, \phi_1, \psi) = \frac{P(U_{i1} = u_1, T_i = t \mid X_i; \theta_2, \phi_1, \psi)}{f(t \mid X_i; \theta_2)}.$$

Let

$$d_{ijk} = \int_0^\infty \mathbb{1}(u \in \mathcal{A}_{ij} \cap \mathcal{B}_{1k}) du = \max(0, \min(b_{1k}, a_{ij}) - \max(b_{1,k-1}, a_{i,j-1}))$$

denote the length of the interval $\mathcal{A}_{ij} \cap \mathcal{B}_{1k}$, $\mu_{0,ij}(\alpha_1) = \sum_{k=1}^{K_1} \rho_k d_{ijk}$ denote the baseline mean over the interval $\mathcal{A}_{ij}$, and $\mu_{ij}(\theta_1 | X_i) = \mu_{0,ij}(\alpha_1) \exp(X_i' \beta_1)$. We then have $P(n_{i1}, \dots, n_{ir_i} | X_i, u_{i1}, T_i \geq w_i; \theta_1)$ equal to

$$\frac{u_{i1}^{n_{i\cdot}} \exp(-u_{i1} R(a_{ir_i} | X_i; \theta_1)) \prod_{j=1}^{r_i} \{\mu_{ij}(\theta_1 | X_i)\}^{n_{ij}}}{\prod_{j=1}^{r_i} n_{ij}!}$$

where $n_{i\cdot} = \sum_{j=1}^{r_i} n_{ij}$ and $R(a_{ir_i} | X_i; \theta_1) = \sum_{j=1}^{r_i} \mu_{ij}(\theta_1)$. The corresponding observed data log-likelihood $\log L_i(\theta)$ is then

$$\begin{aligned} & \delta_i \log h(w_i | X_i; \theta_2) - H(w_i | X_i; \theta_2) \\ & + \delta_i \log \left\{ \int_0^\infty P(n_{i1}, \dots, n_{ir_i} | X_i, u_1, T_i \geq w_i; \theta_1) g_1(u_1 | X_i, T_i = w_i; \theta_2, \phi_1, \psi) du_1 \right\} \\ & + (1 - \delta_i) \log \left\{ \int_0^\infty P(n_{i1}, \dots, n_{ir_i} | X_i, u_1, T_i \geq w_i; \theta_1) g_1(u_1 | X_i, T_i > w_i; \theta_2, \phi_1, \psi) du_1 \right\}. \end{aligned}$$

We denote the corresponding observed data score and information as $S_i(\theta) = \partial \log L_i(\theta) / \partial \theta$ and $I_i(\theta) = -\partial^2 \log L_i(\theta) / \partial \theta \partial \theta'$ respectively. The overall observed data log-likelihood from a sample of m independent individuals is $\log L(\theta) = \sum_{i=1}^m \log L_i(\theta)$. For parametric analysis, the maximum likelihood estimator (MLE) of θ , $\hat{\theta}$, can be obtained by solving $\sum_{i=1}^m S_i(\theta) = 0$. The MLE $\hat{\theta}$ is consistent and asymptotically normally distributed with the variance-covariance matrix estimated by $\text{var}(\hat{\theta}) = \{\sum_{i=1}^m I_i(\hat{\theta})\}^{-1}$. Wald tests can be conducted to make inferences about θ . We next consider implementation of an expectation-maximization algorithm for estimation, motivated by the desire to exploit existing software for the maximization step.

3.2.3 EM Algorithm and Complete Data Likelihood

An expectation-maximization algorithm ([Dempster et al., 1977](#)) can be appealing for estimation when there are latent variables involved and existing functions can be used for the maximization step, which we show is the case here. In constructing the complete data likelihood, we now consider a pseudo-assessment time for the recurrent events at $w_i = a_{i,r_i+1}$ and let n_{i,r_i+1} denote the number of events in the interval $\mathcal{A}_{i,r_i+1} = (a_{ir_i}, a_{i,r_i+1}]$. By doing this, now we have $r_i + 1$ intervals of assessment periods of the recurrent events $\mathcal{A}_{ij}$, with the number of event occurrences in each interval n_{ij} , $j = 1, \dots, r_i + 1$. Note n_{ij} , $j = 1, \dots, r_i$, is observed but n_{i,r_i+1} is latent. Moreover, we let n_{ijk} be the unobserved number of recurrent events in the interval $\mathcal{A}_{ij} \cap \mathcal{B}_{1k}$ and $\mu_{ijk}(\theta_1) = \rho_k d_{ijk} \exp(X_i' \beta_1)$ be the corresponding mean

event rate for $j = 1, \dots, r_i + 1$ and $k = 1, \dots, K_1$. Note we have $n_{ij} = \sum_{k=1}^{K_1} n_{ijk}$ and $\mu_{ij}(\theta_1) = \sum_{k=1}^{K_1} \mu_{ijk}(\theta_1)$. A complete data likelihood contribution from the i th individual containing the information of u_{i1} and n_{ijk} , $j = 1, \dots, r_i + 1$ and $k = 1, \dots, K_1$, can be written as

$$\begin{aligned} \mathcal{L}_i(\theta) &\propto \prod_{j=1}^{r_i+1} \prod_{k=1}^{K_1} \frac{(u_{i1} \mu_{ijk}(\theta_1))^{n_{ijk}} \exp(-u_{i1} \mu_{ijk}(\theta_1))}{n_{ijk}!} \\ &\times h(w_i \mid X_i; \theta_2)^{\delta_i} \exp(-H(w_i \mid X_i; \theta_2)) \\ &\times \{g_1(u_{i1} \mid X_i, T_i = w_i; \theta_2, \phi_1, \psi)\}^{\delta_i} \{g_1(u_{i1} \mid X_i, T_i > w_i; \theta_2, \phi_1, \psi)\}^{1-\delta_i}. \end{aligned}$$

The corresponding complete data log-likelihood can be partitioned as

$$\log \mathcal{L}_i(\theta) = \ell_{i1}(\theta_1) + \ell_{i2}(\theta_2) + \ell_{i3}(\theta_2, \phi_1, \psi) \quad (3.3)$$

where we have

$$\begin{aligned} \ell_{i1}(\theta_1) &= \sum_{j=1}^{r_i+1} \sum_{k=1}^{K_1} n_{ijk} \log \mu_{ijk}(\theta_1) - u_{i1} \sum_{j=1}^{r_i+1} \mu_{ij}(\theta_1), \\ \ell_{i2}(\theta_2) &= \delta_i \log h(w_i \mid X_i; \theta_2) - H(w_i \mid X_i; \theta_2), \end{aligned}$$

and $\ell_{i3}(\theta_2, \phi_1, \psi)$ is

$$\delta_i \log g_1(u_{i1} \mid X_i, T_i = w_i; \theta_2, \phi_1, \psi) + (1 - \delta_i) \log g_1(u_{i1} \mid X_i, T_i > w_i; \theta_2, \phi_1, \psi).$$

We let $\mathcal{S}_i(\theta) = \partial \log \mathcal{L}_i(\theta) / \partial \theta$ and $\mathcal{I}_i(\theta) = -\partial^2 \log \mathcal{L}_i(\theta) / \partial \theta \partial \theta'$ denote the corresponding complete data score and information respectively. The EM algorithm for parameter estimation is described as follows.

The E-Step: Given a current estimate $\theta^{(r-1)}$, we calculate the expectation of the complete data log-likelihood in (3.3) with respect to U_{i1} and all N_{ijk} conditioning the observed history $\bar{\mathcal{H}}_i(w_i^+)$. This leads to $Q_{i1}(\theta_1; \theta^{(r-1)}) = E(\ell_{i1}(\theta_1) \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)})$ given by

$$\begin{aligned} Q_{i1}(\theta_1; \theta^{(r-1)}) &= \sum_{j=1}^{r_i+1} \sum_{k=1}^{K_1} E(N_{ijk} \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)}) \log \mu_{ijk}(\theta_1) \\ &\quad - E(U_{i1} \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)}) \sum_{j=1}^{r_i+1} \mu_{ij}(\theta_1), \end{aligned}$$

and $Q_{i3}(\theta_2, \phi_1, \psi; \theta^{(r-1)}) = E(\ell_{i3}(\theta_2, \phi_1, \psi) \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)})$ given by

$$\begin{aligned} & \delta_i E(\log g_1(U_{i1} \mid X_i, T_i = w_i; \theta_2, \phi_1, \psi) \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)}) \\ & + (1 - \delta_i) E(\log g_1(U_{i1} \mid X_i, T_i > w_i; \theta_2, \phi_1, \psi) \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)}). \end{aligned}$$

Also note that

$$E(\ell_{i2}(\theta_2) \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)}) = \ell_{i2}(\theta_2)$$

since there is no missing data in $\ell_{i2}(\theta_2)$. To compute these expectations for $Q_{i1}(\theta_1; \theta^{(r-1)})$, we require the expectation for the latent counts given by

$$E(N_{ijk} \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)}) = \begin{cases} n_{ij} \mu_{ijk}(\theta_1^{(r-1)}) / \mu_{ij}(\theta_1^{(r-1)}), & j \neq r_i + 1, \\ \mu_{ijk}(\theta_1^{(r-1)}) E(U_{i1} \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)}), & \text{otherwise,} \end{cases}$$

with

$$E(U_{i1} \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)}) = \frac{\int_0^\infty u_{i1} P(\bar{\mathcal{H}}_i(w_i^+), U_{i1} = u_{i1}; \theta^{(r-1)}) du_{i1}}{\int_0^\infty P(\bar{\mathcal{H}}_i(w_i^+), U_{i1} = u_{i1}; \theta^{(r-1)}) du_{i1}}. \quad (3.4)$$

For $Q_{i3}(\theta_2, \phi_1, \psi; \theta^{(r-1)})$, we require $E(\log g_1(U_{i1} \mid X_i, T_i = w_i; \theta_2, \phi_1, \psi) \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)})$ given by

$$\frac{\int_0^\infty \log g_1(u_{i1} \mid X_i, T_i = w_i; \theta_2, \phi_1, \psi) P(\bar{\mathcal{H}}_i(w_i^+), U_{i1} = u_{i1}; \theta^{(r-1)}) du_{i1}}{\int_0^\infty P(\bar{\mathcal{H}}_i(w_i^+), U_{i1} = u_{i1}; \theta^{(r-1)}) du_{i1}}, \quad (3.5a)$$

and $E(\log g_1(U_{i1} \mid X_i, T_i > w_i; \theta_2, \phi_1, \psi) \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)})$ given by

$$\frac{\int_0^\infty \log g_1(u_{i1} \mid X_i, T_i > w_i; \theta_2, \phi_1, \psi) P(\bar{\mathcal{H}}_i(w_i^+), U_{i1} = u_{i1}; \theta^{(r-1)}) du_{i1}}{\int_0^\infty P(\bar{\mathcal{H}}_i(w_i^+), U_{i1} = u_{i1}; \theta^{(r-1)}) du_{i1}}, \quad (3.5b)$$

with $P(\bar{\mathcal{H}}_i(w_i^+), U_{i1} = u_{i1}; \theta^{(r-1)})$ obtained from (3.2). Detailed calculations of the expectation terms in (3.4), (3.5a) and (3.5b) are presented in Appendix B.1.

The M-Step: We maximize $\sum_{i=1}^m E(\log \mathcal{L}_i(\theta) \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)})$ w.r.t. to θ and obtain a new estimate $\theta^{(r)}$. For the recurrent event process, we maximize $Q_{i1}(\theta_1; \theta^{(r-1)})$ w.r.t. θ_1 and obtain $\theta_1^{(r)}$. For the rest of the parameters, we maximize $\sum_{i=1}^m \{\ell_{i2}(\theta_2) + Q_{i3}(\theta_2, \phi_1, \psi; \theta^{(r-1)})\}$ w.r.t. (θ_2, ϕ_1, ψ) and obtain $(\theta_2^{(r)}, \phi_1^{(r)}, \psi^{(r)})$. This gives $\theta^{(r)} = ((\theta_1^{(r)})', (\theta_2^{(r)})', \phi_1^{(r)}, \psi^{(r)})'$.

The *E-Step* and *M-Step* are repeated until the absolute value of each element of the vector $\theta^{(r)} - \theta^{(r-1)}$ are smaller than 10^{-4} and $\hat{\theta}$ is obtained. The observed data information matrix may be obtained based on the complete data score and information

$$\begin{aligned} I_i(\theta) &= E(\mathcal{I}_i(\theta) \mid \bar{\mathcal{H}}_i(w_i^+); \theta) - E(\mathcal{S}_i(\theta) \mathcal{S}_i'(\theta) \mid \bar{\mathcal{H}}_i(w_i^+); \theta) \\ &\quad + E(\mathcal{S}_i(\theta) \mid \bar{\mathcal{H}}_i(w_i^+); \theta) E(\mathcal{S}_i'(\theta) \mid \bar{\mathcal{H}}_i(w_i^+); \theta). \end{aligned} \quad (3.6)$$

as shown by [Louis \(1982\)](#). We use $\hat{\text{var}}(\hat{\theta}) = \{\sum_{i=1}^m I_i(\hat{\theta})\}^{-1}$ as an estimate of the variance-covariance matrix of $\hat{\theta}$. Note the Louis formula requires calculations of $E(\mathcal{S}_i(\theta)\mathcal{S}_i'(\theta) \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)})$, which involves for example, calculations of $E(N_{ijk}N_{ihl} \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)})$ and $E(N_{ijk}U_{i1} \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)})$; see [Appendix B.1](#) for the details on how to calculate these terms.

3.3 A Two-Stage Procedure

As discussed in [Chapter 2](#), an appealing feature of the copula model framework is that the analyses of covariate effects on the terminal event time are based on a marginal specification. This enables a computationally more convenient two-stage estimation procedure. In stage one of the procedure, we focus on the terminal event model and estimate related parameters θ_2 . We let $\bar{\theta}$ denote all parameters in θ excluding θ_2 (i.e., $\bar{\theta} = (\theta'_1, \phi_1, \psi)'$). In stage two, we estimate $\bar{\theta}$ with θ_2 set to the estimated value obtained in stage one. This two-stage procedure thus allows for a semiparametric Cox model for the terminal events. In what follows we reorder the parameters and write $\theta = (\theta'_2, \bar{\theta}')'$.

3.3.1 Two-Stage Estimation via Direct Optimization

We describe the two-stage estimation procedure as follows.

Stage 1: We obtain an estimate of θ_2 , denoted by $\tilde{\theta}_2$, by maximizing $\sum_{i=1}^m \ell_{i2}(\theta_2)$ w.r.t. θ_2 . If we adopt a semiparametric Cox model for the terminal event, we have

$$dH(t \mid X_i; \theta_2) = dH_0(t) \exp(X_i' \beta_2)$$

with $\theta_2 = (dH_0(\cdot), \beta_2')'$, and we can rewrite $\sum_{i=1}^m \ell_{i2}(\theta_2)$ using counting process notation as

$$\sum_{i=1}^m \left\{ \int_0^\infty \log dH(s \mid X_i; \theta_2) \bar{Y}_i(s) dN_{i2}(s) - \int_0^\infty \bar{Y}_i(s) dH(s \mid X_i; \theta_2) \right\}.$$

Taking its derivative w.r.t $dH_0(t)$ and β_2 respectively gives partial score functions

$$U_1(t) = \sum_{i=1}^m \bar{Y}_i(t) [dN_{i2}(t) - dH_0(t) \exp(X_i' \beta_2)] = 0, \quad t \geq 0.$$

and

$$U_2 = \sum_{i=1}^m \left\{ \int_0^\infty \bar{Y}_i(s) X_i [dN_{i2}(s) - dH_0(s) \exp(X_i' \beta_2)] \right\}. \quad (3.7)$$

Solving $U_1(t) = 0$ for $dH_0(t)$ at a fixed β_2 value gives the Breslow profile estimate

$$d\tilde{H}_0(t; \beta_2) = \frac{\sum_{i=1}^m \bar{Y}_i(t) dN_{i2}(t)}{\sum_{i=1}^m \bar{Y}_i(t) \exp(X_i' \beta_2)}, \quad (3.8)$$

which upon substitution into (3.7) gives a partial profile score function for β_2 as

$$U(\beta_2) = \sum_{i=1}^m \left\{ \int_0^\infty \bar{Y}_i(s) \left[X_i - \frac{\sum_{l=1}^m \bar{Y}_l(s) X_l \exp(X_l' \beta_2)}{\sum_{l=1}^m \bar{Y}_l(s) \exp(X_l' \beta_2)} \right] dN_{i2}(s) \right\}.$$

Setting $U(\beta_2) = 0$ and solving for β_2 gives $\tilde{\beta}_2$. Having computed $\tilde{\beta}_2$, then $d\tilde{H}_0(t)$ can be computed by substituting $\tilde{\beta}_2$ in (3.8).

Stage 2: Having an estimate $\tilde{\theta}_2$, we obtain $\tilde{\theta} = (\tilde{\theta}_1', \tilde{\phi}_1, \tilde{\psi})'$ by maximizing $\sum_{i=1}^m \log L_i(\tilde{\theta}_2, \tilde{\theta})$, the overall observed data log-likelihood, with respect to $\tilde{\theta}$, after $\tilde{\theta}_2$ is inserted.

Let $\tilde{\theta} = (\tilde{\theta}_2', \tilde{\theta}')'$ denote the estimate from this two-stage estimation procedure. We also denote the individual contribution of scores of the partial log-likelihoods used in the two stages as $\mathcal{U}_{i1}(\theta_2) = \partial \ell_{i2}(\theta_2) / \partial \theta_2$, $\mathcal{U}_{i2}(\theta) = \partial \log L_i(\theta) / \partial \tilde{\theta}$ respectively, and the overall score $\mathcal{U}_i(\theta) = (\mathcal{U}_{i1}'(\theta_2), \mathcal{U}_{i2}'(\theta))'$. The two-stage estimation procedure described above is indeed solving the unbiased estimating equation $\sum_{i=1}^m \mathcal{U}_i(\theta) = 0$. By standard estimating function theory (Boos and Stefanski, 2013), we have

$$\sqrt{m}(\tilde{\theta} - \theta) \rightarrow \text{MVN}(0, \mathcal{A}^{-1}(\theta) \mathcal{B}(\theta) (\mathcal{A}^{-1}(\theta))') \quad (3.9)$$

where $\mathcal{A}(\theta) = E(-\partial \mathcal{U}_i(\theta) / \partial \theta')$ and $\mathcal{B}(\theta) = E(\mathcal{U}_i(\theta) \mathcal{U}_i'(\theta))$. Moreover, $\mathcal{A}(\theta)$ and $\mathcal{B}(\theta)$ can be estimated empirically by

$$A(\tilde{\theta}) = -\frac{1}{m} \sum_{i=1}^m \frac{\partial \mathcal{U}_i(\theta)}{\partial \theta'} \Big|_{\theta=\tilde{\theta}} \quad \text{and} \quad B(\tilde{\theta}) = \frac{1}{m} \sum_{i=1}^m \mathcal{U}_i(\theta) \mathcal{U}_i'(\theta) \Big|_{\theta=\tilde{\theta}}. \quad (3.10)$$

The asymptotic covariance matrix can then be estimated by $\hat{\text{var}}(\tilde{\theta}) = A^{-1}(\tilde{\theta}) B(\tilde{\theta}) (A^{-1}(\tilde{\theta}))' / m$. Note $\mathcal{U}_{i2}(\theta)$ is a sub-matrix of the corresponding observed data score $S_i(\theta)$, and $-\partial \mathcal{U}_{i2}(\theta) / \partial \theta'$ is a sub-matrix of the corresponding observed data information $I_i(\theta)$.

3.3.2 Two-Stage Estimation via an EM Algorithm

Note that *Stage 2* of the two-stage estimation procedure can also be completed via an EM algorithm as follows. As shown in (3.3), the complete data log-likelihood $\log \mathcal{L}_i(\theta)$ can be partitioned into three parts: $\ell_{i1}(\theta_1)$, $\ell_{i2}(\theta_2)$ and $\ell_{i3}(\theta_2, \phi_1, \psi)$, so the two-stage estimation procedure with an EM algorithm works as follows.

Stage 1: Same as in Section 3.3.1, we obtain $\tilde{\theta}_2$ by maximizing $\sum_{i=1}^m \ell_{i2}(\theta_2)$ w.r.t. θ_2 .

Stage 2: We use an EM algorithm to estimate $\bar{\theta}$ as follows. Let $\theta^{(r-1)} = (\tilde{\theta}'_2, (\bar{\theta}^{(r-1)})')'$.

Step 1: We maximize $\sum_{i=1}^m E(\ell_{i1}(\theta_1) \mid \bar{\mathcal{H}}_i(w_i); \theta^{(r-1)})$ w.r.t θ_1 and obtain $\theta_1^{(r)}$.

Step 2: We then maximize $\sum_{i=1}^m \{\ell_{i2}(\tilde{\theta}_2) + E(\ell_{i3}(\tilde{\theta}_2, \phi_1, \psi) \mid \bar{\mathcal{H}}_i(w_i); \theta^{(r-1)})\}$ w.r.t (ϕ_1, ψ) and obtain $(\phi_1^{(r)}, \psi^{(r)})$.

Note $\bar{\theta}^{(r)} = ((\theta_1^{(r)})', \phi_1^{(r)}, \psi^{(r)})'$. We repeat *Step 1* and *Step 2* until $\bar{\theta}^{(r)}$ converges and obtain $\tilde{\theta}$. The estimate $\tilde{\theta} = (\tilde{\theta}'_2, \tilde{\theta}')$ obtained here is the same as the estimate obtained via direct optimization in Section 3.3.1 because we are solving the same estimating equation $\sum_{i=1}^m \mathcal{U}_i(\theta) = 0$. The standard estimating function theory still applies, and we need to compute $\mathcal{U}_i(\theta)$, $-\partial \mathcal{U}_i(\theta)/\partial \theta'$ to obtain $A^{-1}(\tilde{\theta})$, $B(\tilde{\theta})$ using (3.10), and thus a variance estimate $\text{var}(\tilde{\theta})$.

The first stage score contributed from individual i is still $\mathcal{U}_{i1}(\theta_2)$. We denote the second stage complete data score contributed from individual i as $\mathcal{U}_{i2}^c(\theta) = ((\mathcal{U}_{i21}^c(\theta_1))', (\mathcal{U}_{i22}^c(\theta))')'$ where $\mathcal{U}_{i21}^c(\theta_1) = \partial \ell_{i1}(\theta_1)/\partial \theta_1$ and $\mathcal{U}_{i22}^c(\theta) = \partial \ell_{i3}(\theta_2, \phi_1, \psi)/\partial (\phi_1, \psi)'$. Note $\mathcal{U}_{i2}^c(\theta)$ is a sub-matrix of the corresponding complete data score $\mathcal{S}_i(\theta)$ and we can obtain $\mathcal{U}_{i2}(\theta)$ using

$$\mathcal{U}_{i2}(\theta) = E(\mathcal{U}_{i2}^c(\theta) \mid \bar{\mathcal{H}}_i(w_i); \theta).$$

For $-\partial \mathcal{U}_{i2}(\theta)/\partial \theta'$, we use the Louis formula (Louis, 1982) in (3.6) to calculate the observed information $I_i(\theta)$ and then extract $-\partial \mathcal{U}_{i2}(\theta)/\partial \theta$ from $I_i(\theta)$.

3.4 Simulation Studies

In this section, we perform simulation studies to evaluate and compare the finite sample performance of the proposed estimation methods outlined in Sections 3.2 and 3.3. Specifically, we focus on comparing simultaneous estimation (for parametric analysis) and the two-stage estimation.

3.4.1 Simulation Set-up

We consider a simple setting with a binary covariate where $X_i = 0$ or 1 . We consider multiplicative models for both the recurrent and terminal event processes with constant baseline functions (i.e., $\rho_0(s; \alpha_1) = \rho_0$ and $h_0(t; \alpha_2) = h_0$). We let the random effect U_{i1} be gamma distributed with mean $= 1$ and variance $= \phi_1$, and specify a Gaussian copula with Kendall's τ ([Kendall, 1938](#)) indexing the association; ψ is a deterministic function of τ with the particular function governed by the specification of the Gaussian copula. For the coefficients, we set $\beta_1 = \log 0.8$ and $\beta_2 = \log 0.9$. We set $\phi_1 = 0.2, 0.4$ to reflect different levels of heterogeneity in the recurrent event rate and vary $\tau = 0, 0.2, 0.4$ to investigate scenarios with no, weak and medium associations between the two event processes. We consider an administrative right-censoring time for the terminal event $C_i = 1$. For the assessment process $\{A_i(s), 0 < s\}$, we consider it to be a time homogeneous Poisson process with an instantaneous rate $E\{dA_i(s) \mid X_i\} = \lambda_A$. We set $E(A_i(1) \mid X_i) = 4$. Let $E(N_{i1}(1) \mid X_i = 0) = \int_0^1 E\{dN_{i1}(s) \mid X_i = 0\} ds$ be the expected number of recurrent events by time 1 in the control group ($X_i = 0$) where

$$\begin{aligned} E\{dN_{i1}(s) \mid X_i\} &= E_{U_{i1}, Y_i^\dagger(s)}\{E\{dN_{i1}(s) \mid U_{i1}, Y_i^\dagger(s), X_i\}\} \\ &= \rho(s \mid X_i; \theta_1) E\{U_{i1} \mid Y_i^\dagger(s) = 1, X_i\} \mathcal{F}(s \mid X_i; \theta_2) \end{aligned}$$

with

$$E\{U_{i1} \mid Y_i^\dagger(s) = 1, X_i\} = \int_0^\infty u_1 g_1(u_1 \mid X_i, T_i \geq s; \theta_2, \phi_1, \psi) du_1.$$

We then solve for ρ_0 and h_0 so that the expected number of events over $(0, 1]$, $E\{N_{i1}(1) \mid X_i = 0\}$, is 2 or 4, and $P(T_i \leq 1 \mid X_i = 0) = 0.6$ corresponding to a 60% failure rate if $X_i = 0$. The true parameter values corresponding to each scenario are summarized in [Table 3.1](#) for 12 scenarios. We aim to assess estimators of parameters of interest when the recurrent events are subject to intermittent observation. Some of the parameters of interest are specified directly by us (i.e., $\beta_1, \beta_2, \phi_1, \psi$) and some are computable by plugging in estimates of θ into their expressions (i.e., $E(N_{i1}(1) \mid X_i = 0), \mathcal{F}(1 \mid X_i = 0)$).

3.4.2 Parametric Analyses with Piecewise Constant Models

Here we simulate datasets of $m = 1000$ individuals according to the parameter setting given in [Section 3.4.1](#), following the data generation procedure described in [Appendix B.2](#). We conduct fully parametric analyses by using a Gaussian copula and fitting a proportional hazard model with a piecewise constant baseline hazard function for the terminal event

Table 3.1: True parameter values for each scenario

Scenario	$E(N_{i1}(1) X_i = 0)$	ϕ_1	Kendall's τ	ψ	θ_1		θ_2	
					ρ_0	β_1	h_0	β_2
1	2	0.2	0.0	0.0	3.0543	-0.2231	0.2877	-0.1054
2			0.2	0.3090	3.2769	-0.2231	0.2877	-0.1054
3			0.4	0.5878	3.5135	-0.2231	0.2877	-0.1054
4		0.4	0.0	0.0	3.0543	-0.2231	0.2877	-0.1054
5			0.2	0.3090	3.3732	-0.2231	0.2877	-0.1054
6			0.4	0.5878	3.7368	-0.2231	0.2877	-0.1054
7	4	0.2	0.0	0.0	6.1086	-0.2231	0.2877	-0.1054
8			0.2	0.3090	6.5537	-0.2231	0.2877	-0.1054
9			0.4	0.5878	7.0270	-0.2231	0.2877	-0.1054
10		0.4	0.0	0.0	6.1086	-0.2231	0.2877	-0.1054
11			0.2	0.3090	6.7464	-0.2231	0.2877	-0.1054
12			0.4	0.5878	7.4736	-0.2231	0.2877	-0.1054

process. The number of pieces of the baseline hazard function for the terminal events are denoted by K_2 and we set $K_1 = K_2 = 8$. Computation of the observed and complete data likelihood, complete data scores and information under this fully parametric joint model are presented in Appendix B.3. We investigate the finite sample performance of proposed estimators via simultaneous estimation and the two-stage estimation by evaluating and comparing the empirical bias (EBIAS), empirical standard error (ESE), average asymptotic standard error (ASE) and empirical coverage probability (ECP) for parameter estimates of interest. Simulation results obtained via simultaneous estimation and a two-stage estimation are summarized in Table 3.2 and Table 3.3 respectively. All the estimates are averaged over $n_{\text{sim}} = 500$ replications.

For both the simultaneous and two-stage estimation methods, the empirical biases are small for all parameters of interest, as expected for both simultaneous and two-stage procedures. The empirical standard errors are in good agreement with the average model-based standard errors, and the empirical coverage probabilities are close to the nominal level of 95% for both estimation procedures. In addition, the finite sample performance of parameters from the two-stage estimation method is comparable to those from the simultaneous estimation in terms of empirical standard errors, average estimated standard errors and coverage probabilities, regardless of the strength of the association between the two event processes. This demonstrates that the two-stage estimation procedure is remarkably efficient. Comparing the simulation results for scenarios 1-6 in Table 3.3 to

Table 3.2: Tabulation of frequency properties of estimators obtained from a **fully parametric** specification with an 8-piece piecewise constant baseline hazards for both processes under no random censoring for $E(A_i(1) | X_i) = 4$ (just administrative censoring at 1; estimates are obtained via **direct optimization** of the observed data likelihood with **simultaneous estimation**; $m = 1000$ and $n_{\text{sim}} = 500$)

$E(N_{i1}(1) X_i = 0)$	τ	Parameter	$\phi_1 = 0.2$					$\phi_1 = 0.4$				
			TRUE	EBIAS	ESE	ASE	ECP%	TRUE	EBIAS	ESE	ASE	ECP%
2	0	β_1	-0.2231	0.0003	0.0662	0.0665	94.6	-0.2231	0.0005	0.0827	0.0749	92.0
		β_2	-0.1054	-0.0014	0.0795	0.0831	95.2	-0.1054	-0.0030	0.0853	0.0831	95.2
		$\log \phi_1$	-1.6094	-0.0199	0.2159	0.2218	94.8	-0.9163	-0.0096	0.1500	0.1464	95.4
		$\log\{(1 + \psi)/(1 - \psi)\}$	0.0000	0.0080	0.2752	0.2861	96.8	0.0000	0.0087	0.2201	0.2196	94.6
		$E(N_{i1}(1) X_i = 0)$	2.0000	-0.0030	0.1100			2.0000	-0.0030	0.1232		
		$\mathcal{F}(1 X_i = 0)$	0.4000	-0.0013	0.0212			0.4000	-0.0006	0.0204		
	0.2	β_1	-0.2231	-0.0011	0.0625	0.0671	96.2	-0.2231	-0.0005	0.0716	0.0754	96.0
		β_2	-0.1054	0.0074	0.0828	0.0831	95.4	-0.1054	-0.0059	0.0823	0.0831	95.0
		$\log \phi_1$	-1.6094	-0.0037	0.2155	0.2120	96.4	-0.9163	-0.0128	0.1425	0.1378	93.8
		$\log\{(1 + \psi)/(1 - \psi)\}$	0.6389	0.0318	0.2989	0.2990	94.4	0.6389	0.0163	0.2349	0.2253	94.0
		$E(N_{i1}(1) X_i = 0)$	2.0000	-0.0028	0.0994			2.0000	0.0016	0.1129		
		$\mathcal{F}(1 X_i = 0)$	0.4000	0.0011	0.0215			0.4000	-0.0009	0.0213		
	0.4	β_1	-0.2231	-0.0004	0.0634	0.0671	95.6	-0.2231	-0.0008	0.0710	0.0755	96.0
		β_2	-0.1054	0.0064	0.0840	0.0831	95.2	-0.1054	-0.0054	0.0849	0.0830	95.2
		$\log \phi_1$	-1.6094	-0.0067	0.2184	0.2153	96.4	-0.9163	-0.0143	0.1403	0.1383	95.4
		$\log\{(1 + \psi)/(1 - \psi)\}$	1.3486	0.0561	0.3428	0.3580	97.2	1.3486	0.0193	0.2588	0.2500	94.2
		$E(N_{i1}(1) X_i = 0)$	2.0000	-0.0076	0.0974			2.0000	-0.0010	0.1050		
		$\mathcal{F}(1 X_i = 0)$	0.4000	0.0008	0.0210			0.4000	-0.0008	0.0220		
4	0	β_1	-0.2231	-0.0001	0.0510	0.0529	96.6	-0.2231	0.0008	0.0694	0.0627	91.4
		β_2	-0.1054	-0.0014	0.0795	0.0831	95.2	-0.1054	-0.0031	0.0853	0.0831	95.0
		$\log \phi_1$	-1.6094	-0.0118	0.1343	0.1385	97.0	-0.9163	-0.0058	0.0992	0.1010	95.4
		$\log\{(1 + \psi)/(1 - \psi)\}$	0.0000	-0.0011	0.2079	0.2162	96.2	0.0000	0.0117	0.1703	0.1740	95.8
		$E(N_{i1}(1) X_i = 0)$	4.0000	-0.0051	0.1774			4.0000	-0.0113	0.2012		
		$\mathcal{F}(1 X_i = 0)$	0.4000	-0.0013	0.0212			0.4000	-0.0007	0.0204		
	0.2	β_1	-0.2231	0.0016	0.0507	0.0533	95.6	-0.2231	-0.0014	0.0614	0.0632	95.8
		β_2	-0.1054	0.0074	0.0827	0.0831	95.2	-0.1054	-0.0058	0.0822	0.0831	94.8
		$\log \phi_1$	-1.6094	-0.0006	0.1302	0.1328	96.2	-0.9163	-0.0059	0.0970	0.0951	94.8
		$\log\{(1 + \psi)/(1 - \psi)\}$	0.6389	0.0173	0.2201	0.2217	94.8	0.6389	0.0053	0.1752	0.1760	0.1760
		$E(N_{i1}(1) X_i = 0)$	4.0000	-0.0066	0.1647			4.0000	0.0111	0.1905		
		$\mathcal{F}(1 X_i = 0)$	0.4000	0.0011	0.0216			0.4000	-0.0009	0.0213		
	0.4	β_1	-0.2231	0.0022	0.0515	0.0532	96.0	-0.2231	-0.0022	0.0606	0.0632	96.2
		β_2	-0.1054	0.0064	0.0841	0.0830	95.2	-0.1054	-0.0051	0.0847	0.0830	95.0
		$\log \phi_1$	-1.6094	0.0027	0.1317	0.1369	96.8	-0.9163	-0.0103	0.0973	0.0968	94.8
		$\log\{(1 + \psi)/(1 - \psi)\}$	1.3486	0.0450	0.2527	0.2484	94.8	1.3486	0.0036	0.1855	0.1877	94.4
		$E(N_{i1}(1) X_i = 0)$	4.0000	-0.0039	0.1608			4.0000	0.0064	0.1784		
		$\mathcal{F}(1 X_i = 0)$	0.4000	0.0008	0.0210			0.4000	-0.0007	0.0220		

Table 3.3: Tabulation of frequency properties of estimators obtained from a **fully parametric** specification with an 8-piece piecewise constant baseline hazards for both processes under no random censoring for $E(A_i(1) | X_i) = 4$ (just administrative censoring at 1; estimates are obtained via **direct optimization** of the observed data likelihood with **two-stage estimation**; $m = 1000$ and $n_{\text{sim}} = 500$)

$E(N_{i1}(1) X_i = 0)$	τ	Parameter	$\phi_1 = 0.2$					$\phi_1 = 0.4$				
			TRUE	EBIAS	ESE	ASE	ECP%	TRUE	EBIAS	ESE	ASE	ECP%
2	0	β_1	-0.2231	0.0003	0.0662	0.0665	94.4	-0.2231	0.0005	0.0827	0.0748	92.0
		β_2	-0.1054	-0.0014	0.0795	0.0831	95.2	-0.1054	-0.0031	0.0853	0.0831	95.4
		$\log \phi_1$	-1.6094	-0.0201	0.2159	0.2199	95.0	-0.9163	-0.0096	0.1500	0.1452	95.6
		$\log\{(1+\psi)/(1-\psi)\}$	0.0000	-0.0199	0.2751	0.2857	96.6	0.0000	0.0087	0.2199	0.2195	94.4
		$E(N_{i1}(1) X_i = 0)$	2.0000	-0.0031	0.1101			2.0000	-0.0030	0.1232		
		$\mathcal{F}(1 X_i = 0)$	0.4000	-0.0013	0.0212			0.4000	-0.0006	0.0204		
	0.2	β_1	-0.2231	-0.0011	0.0625	0.0670	96.2	-0.2231	-0.0005	0.0716	0.0753	96.0
		β_2	-0.1054	0.0074	0.0828	0.0831	95.2	-0.1054	-0.0058	0.0822	0.0832	95.0
		$\log \phi_1$	-1.6094	-0.0037	0.2156	0.2107	95.8	-0.9163	-0.0128	0.1425	0.1371	93.6
		$\log\{(1+\psi)/(1-\psi)\}$	0.6389	0.0315	0.2988	0.2990	94.6	0.6389	0.0160	0.2348	0.2250	93.8
		$E(N_{i1}(1) X_i = 0)$	2.0000	-0.0028	0.0994			2.0000	0.0016	0.1129		
		$\mathcal{F}(1 X_i = 0)$	0.4000	0.0011	0.0215			0.4000	-0.0009	0.0213		
	0.4	β_1	-0.2231	-0.0005	0.0634	0.0671	95.6	-0.2231	-0.0007	0.0710	0.0756	95.8
		β_2	-0.1054	0.0063	0.0841	0.0831	95.0	-0.1054	-0.0050	0.0846	0.0832	95.6
		$\log \phi_1$	-1.6094	-0.0069	0.2184	0.2145	96.2	-0.9163	-0.0144	0.1403	0.1378	95.2
		$\log\{(1+\psi)/(1-\psi)\}$	1.3486	0.0555	0.3426	0.3558	96.6	1.3486	0.0187	0.2588	0.2501	94.4
		$E(N_{i1}(1) X_i = 0)$	2.0000	-0.0077	0.0973			2.0000	-0.0010	0.1050		
		$\mathcal{F}(1 X_i = 0)$	0.4000	0.0007	0.0210			0.4000	-0.0007	0.0220		
4	0	β_1	-0.2231	-0.0001	0.0510	0.0528	96.4	-0.2231	0.0008	0.0694	0.0626	91.6
		β_2	-0.1054	-0.0014	0.0795	0.0831	95.2	-0.1054	-0.0031	0.0853	0.0831	95.4
		$\log \phi_1$	-1.6094	-0.0118	0.1343	0.1381	97.0	-0.9163	-0.0058	0.0992	0.1007	95.2
		$\log\{(1+\psi)/(1-\psi)\}$	0.0000	-0.0011	0.2078	0.2160	96.2	0.0000	0.0117	0.1702	0.1739	95.8
		$E(N_{i1}(1) X_i = 0)$	4.0000	-0.0051	0.1774			4.0000	-0.0112	0.2012		
		$\mathcal{F}(1 X_i = 0)$	0.4000	-0.0013	0.0212			0.4000	-0.0006	0.0204		
	0.2	β_1	-0.2231	0.0016	0.0507	0.0533	95.4	-0.2231	-0.0014	0.0614	0.0632	95.6
		β_2	-0.1054	0.0074	0.0828	0.0831	95.2	-0.1054	-0.0058	0.0822	0.0832	95.0
		$\log \phi_1$	-1.6094	-0.0006	0.1302	0.1319	95.8	-0.9163	-0.0059	0.0970	0.0949	94.6
		$\log\{(1+\psi)/(1-\psi)\}$	0.6389	0.0171	0.2201	0.2218	94.8	0.6389	0.0050	0.1751	0.1760	95.2
		$E(N_{i1}(1) X_i = 0)$	4.0000	-0.0067	0.1646			4.0000	0.0111	0.1903		
		$\mathcal{F}(1 X_i = 0)$	0.4000	0.0011	0.0215			0.4000	-0.0009	0.0213		
	0.4	β_1	-0.2231	0.0022	0.0515	0.0532	96.2	-0.2231	-0.0022	0.0607	0.0631	95.8
		β_2	-0.1054	0.0063	0.0841	0.0831	95.0	-0.1054	-0.0050	0.0846	0.0832	95.6
		$\log \phi_1$	-1.6094	0.0025	0.1317	0.1365	96.8	-0.9163	-0.0104	0.0973	0.0965	95.0
		$\log\{(1+\psi)/(1-\psi)\}$	1.3486	0.0443	0.2526	0.2479	94.0	1.3486	0.0029	0.1855	0.1878	94.4
		$E(N_{i1}(1) X_i = 0)$	4.0000	-0.0041	0.1609			4.0000	0.0064	0.1782		
		$\mathcal{F}(1 X_i = 0)$	0.4000	0.0007	0.0210			0.4000	-0.0007	0.0220		

those in Table 2.3 (where the recurrent events are subject to right-censoring), we find that the empirical standard errors as well as the asymptotic standard errors of estimators related to $\hat{\beta}_1$, $\hat{\phi}_1$, and $\hat{\psi}$ are larger when the recurrent events data are interval-censored. This is reasonable because we lose information on the exact recurrent event times when recurrent events are intermittently observed.

3.4.3 Analyses with a Cox Model for the Terminal Event

Now we conduct semiparametric analyses by fitting a Cox proportional hazard model with a non-parametric baseline hazard function for the terminal event process. We use a two-stage estimation procedure with direct optimization to obtain point estimates, with the variance estimates obtained using the Louis formula (Louis, 1982) in (3.6). The complete data score and information needed to compute the observed information are provided in Appendix B.4 where the two-stage score and information are also given.

Simulation results averaged over 500 replications are summarized in Table 3.4. We see that for all parameters of interest from the semiparametric analyses, the empirical biases are small, the empirical and average estimated standard errors are in good agreement, and the empirical coverage probabilities are close to the nominal 95% level. In addition, simulation results obtained from a fully parametric specification with a 8-piece piecewise constant baseline model are comparable with those obtained from semi-parametric specifications (Table 3.3 vs Table 3.4) in terms of point estimates and variance estimates. Tables 3.2, 3.3 and 3.4 all demonstrate the validity of our model framework when the recurrent events are subject to intermittent observation.

3.5 Joint Damage in Psoriatic Arthritis

HLA-B27 is a protein within the human leukocyte antigen (HLA) system, which is found on the surface of white blood cells and plays a crucial role in the body's immune response. Its presence is found to be strongly associated with several autoimmune diseases such as particularly ankylosing spondylitis, and other related inflammatory conditions like reactive arthritis, psoriatic arthritis, and inflammatory bowel disease (Gladman and Farewell, 1995). To examine the effect of the genetic marker HLA-B27 on the development of joint damage among patients with psoriatic arthritis, we apply our proposed model to the data from the University of Toronto Psoriatic Arthritis Clinic registry. The joint damage is measured by radiographic examination among patients with psoriatic arthritis and it is assessed

Table 3.4: Tabulation of frequency properties of estimators obtained from a **semiparametric** specification with a Cox model for the terminal event under no random censoring for $E(A_i(1) | X_i) = 4$ (just administrative censoring at 1; point estimates are obtained via **direct optimization** of the observed data likelihood via **two-stage estimation**; variance estimates are obtained through Louis formula with an EM algorithm; $m = 1000$ and $n_{\text{sim}} = 500$)

$E(N_{i1}(1) X_i = 0)$	τ	Parameter	$\phi_1 = 0.2$					$\phi_1 = 0.4$				
			TRUE	EBIAS	ESE	ASE	ECP%	TRUE	EBIAS	ESE	ASE	ECP%
2	0	β_1	-0.2231	0.0003	0.0662	0.0667	94.4	-0.2231	0.0005	0.0827	0.0751	92.0
		β_2	-0.1054	-0.0013	0.0794	0.0830	95.2	-0.1054	-0.0030	0.0853	0.0831	95.2
		$\log \phi_1$	-1.6094	-0.0198	0.2159	0.2199	95.0	-0.9163	-0.0096	0.1500	0.1452	95.8
		$\log\{(1 + \psi)/(1 - \psi)\}$	0.0000	0.0080	0.2753	0.2866	97.0	0.0000	0.0088	0.2203	0.2205	94.6
		$E(N_{i1}(1) X_i = 0)$	2.0000	-0.0024	0.1101			2.0000	-0.0023	0.1233		
		$\mathcal{F}(1 X_i = 0)$	0.4000	-0.0010	0.0212			0.4000	-0.0004	0.0204		
	0.2	β_1	-0.2231	-0.0011	0.0625	0.0684	96.2	-0.2231	-0.0005	0.0716	0.0789	96.8
		β_2	-0.1054	0.0075	0.0827	0.0830	95.2	-0.1054	-0.0057	0.0821	0.0831	95.2
		$\log \phi_1$	-1.6094	-0.0036	0.2155	0.2111	96.0	-0.9163	-0.0127	0.1424	0.1375	93.6
		$\log\{(1 + \psi)/(1 - \psi)\}$	0.6389	0.0323	0.2988	0.3071	95.4	0.6389	0.0168	0.2352	0.2374	95.4
		$E(N_{i1}(1) X_i = 0)$	2.0000	-0.0022	0.0994			2.0000	0.0022	0.1129		
		$\mathcal{F}(1 X_i = 0)$	0.4000	0.0014	0.0215			0.4000	-0.0006	0.0213		
	0.4	β_1	-0.2231	-0.0005	0.0634	0.5602	97.4	-0.2231	-0.0007	0.0710	0.0756	95.8
		β_2	-0.1054	0.0064	0.0840	0.0830	95.0	-0.1054	-0.0049	0.0846	0.0831	95.6
		$\log \phi_1$	-1.6094	-0.0064	0.2185	0.2145	97.0	-0.9163	-0.0140	0.1402	0.1453	96.2
		$\log\{(1 + \psi)/(1 - \psi)\}$	1.3486	0.0568	0.3431	0.3558	96.6	1.3486	0.0197	0.2592	0.2949	97.2
		$E(N_{i1}(1) X_i = 0)$	2.0000	-0.0069	0.0973			2.0000	0.0000	0.1050		
		$\mathcal{F}(1 X_i = 0)$	0.4000	0.0010	0.0210			0.4000	-0.0004	0.0220		
4	0	β_1	-0.2231	-0.0001	0.0510	0.0531	96.4	-0.2231	0.0008	0.0694	0.0633	92.2
		β_2	-0.1054	-0.0013	0.0794	0.0830	95.2	-0.1054	-0.0030	0.0853	0.0831	95.2
		$\log \phi_1$	-1.6094	-0.0118	0.1343	0.1381	96.8	-0.9163	-0.0058	0.0992	0.1007	95.2
		$\log\{(1 + \psi)/(1 - \psi)\}$	0.0000	-0.0011	0.2080	0.2172	96.2	0.0000	0.0118	0.1705	0.1754	97.2
		$E(N_{i1}(1) X_i = 0)$	4.0000	-0.0039	0.1773			4.0000	-0.0099	0.2014		
		$\mathcal{F}(1 X_i = 0)$	0.4000	-0.0010	0.0212			0.4000	-0.0004	0.0204		
	0.2	β_1	-0.2231	0.0016	0.0507	0.0567	96.6	-0.2231	-0.0014	0.0614	0.0732	97.6
		β_2	-0.1054	0.0075	0.0827	0.0830	95.2	-0.1054	-0.0057	0.0821	0.0831	95.2
		$\log \phi_1$	-1.6094	-0.0005	0.1302	0.1328	96.0	-0.9163	-0.0059	0.0970	0.0957	94.6
		$\log\{(1 + \psi)/(1 - \psi)\}$	0.6389	0.0177	0.2202	0.2381	96.4	0.6389	0.0057	0.1753	0.1760	95.2
		$E(N_{i1}(1) X_i = 0)$	4.0000	-0.0056	0.1647			4.0000	0.0123	0.1902		
		$\mathcal{F}(1 X_i = 0)$	0.4000	0.0014	0.0215			0.4000	-0.0006	0.0213		
	0.4	β_1	-0.2231	0.0022	0.0515	0.0532	96.2	-0.2231	-0.0022	0.0607	0.0631	95.8
		β_2	-0.1054	0.0064	0.0840	0.0830	95.0	-0.1054	-0.0049	0.0846	0.0831	95.6
		$\log \phi_1$	-1.6094	0.0029	0.1317	0.1365	96.8	-0.9163	-0.0101	0.0972	0.0965	95.0
		$\log\{(1 + \psi)/(1 - \psi)\}$	1.3486	0.0455	0.2527	0.2479	94.0	1.3486	0.0039	0.1856	0.1878	94.4
		$E(N_{i1}(1) X_i = 0)$	4.0000	-0.0025	0.1610			4.0000	0.0085	0.1786		
		$\mathcal{F}(1 X_i = 0)$	0.4000	0.0010	0.0210			0.4000	-0.0004	0.0220		

in each of 42 joints, including 30 hand joints (wrists, metacarpophalangeals, proximal interphalangeals and distal interphalangeals) and 12 foot joints (metatarsophalangeals and interphalangeal first toes).

The cohort we focused on consists of 1131 patients diagnosed with PsA, all of whom had at least one follow-up visit after recruitment into the study. Figure 3.1 displays ten profiles of individuals that were recruited to the PsA cohort; 2 of them are HLA-B27 positive and are coloured in blue in the figure. The dashed lines represent periods from PsA onset to the recruitment, and the solid lines represent the follow-up periods. This figure shows that the timing and number of visits vary across individuals. For each individual, the reported age of recruitment is taken as the time origin, and the dates of radiological assessments and numbers of cumulative damaged joints were recorded at subsequent assessment visits. The average follow-up period is 21.17 years (S.D. 10.61, range 3.22 to 50.22), with an average time from recruitment to the first follow-up visit of 1.27 years (S.D. 2.46, range 0.04 to 32.18). On average, patients had 18.50 visits (S.D. 15.45, range 1 to 80) and 4.95 total joint damages (S.D. 8.67, range 0 to 59) during the follow-up period. A total of 184 (16.3%) patients are HLA-B27 positive and 82 (7.3%) patients were observed to have died.

We consider a joint analysis of the recurrent joint damage process and the survival process with the association between the two processes accommodated by a Gaussian copula with Kendall's τ and the random effect specified by a Gamma distribution. We consider the HLA-B27 as the only covariate and conduct both parametric and semiparametric analyses. For the fully parametric analysis, we specify an 8-piece piecewise constant baseline intensity for death and use direct optimization of the observed data likelihood to obtain estimators. For the semiparametric analysis, we specify a Cox model for death, and estimators are obtained via a two-stage procedure, with variance estimates derived through an EM algorithm. Results obtained from the joint analyses are summarized in Table 3.5. We observe that the point estimates obtained through both parametric and semiparametric models are comparable to each other while the estimated standard errors are notably larger when we use a two-stage procedure for estimation. The presence of a positive HLA-B27 significantly increases the occurrence of joint damage in patients with PsA ($\exp(\hat{\beta}_1) = \exp(0.4410) = 1.554$, 95% CI: (1.339, 1.805), $p < 0.001$ with the parametric model; $\exp(\hat{\beta}_1) = \exp(0.4414) = 1.555$, 95% CI: (1.191, 2.030), $p = 0.001$ with the semiparametric model), while it doesn't have a significant effect on the survival process ($p > 0.05$ for both specifications). The recurrences of joint damage show heterogeneity ($\phi_1 = 1.894$, 95% CI: (1.722, 2.083) with the parametric model; $\phi_1 = 1.893$, 95% CI: (1.719, 2.085) with the semiparametric model) between individuals and there is no evidence of a significant dependence between the two processes ($p > 0.05$). Figure 3.2 shows the estimated cumulative conditional and cumulative marginal rates for the joint damage process

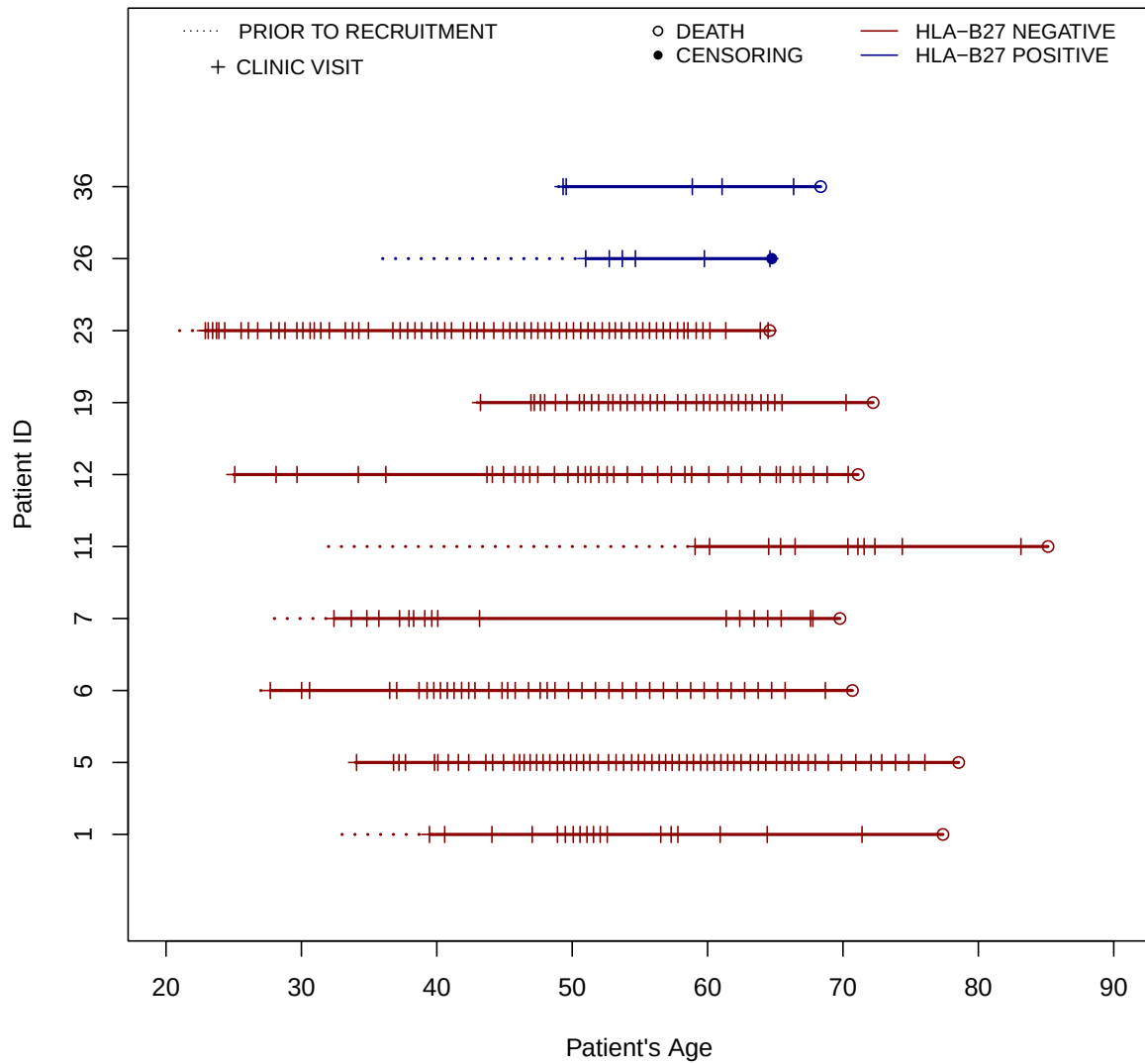


Figure 3.1: Selected sample profiles of individuals recruited to the PsA cohort

(left panel) and the estimated cumulative death hazard and survival function (right panel), obtained from both parametric and semiparametric model specifications. The conditional cumulative rate for the joint damage process is calculated given that the individual is alive and the random effect is equal to 1. The marginal cumulative rate is an integrated version that averages over the random effect and the survival status. The two cumulative rates of the joint damage process give comparable results, which may be due to the fact that death is rarely observed in the data and not significantly associated with the joint damage process.

Table 3.5: Joint analysis of the joint damage and the survival process with a Gaussian copula, a Gamma distributed random effect and a single covariate of HLA-B27; both parametric (with an 8-piece PWC baseline rate for death; estimates are obtained through direct optimization) and semiparametric models (estimates are obtained via a two-stage procedure with variance estimates derived through an EM algorithm) are considered

Type	Parameter	Parametric (Simultaneous)		Semiparametric (Two-Stage)	
		EST	SE	EST	SE
HLA-B27 (on joint damage)	β_1	0.4410	0.0762	0.4414	0.1360
HLA-B27 (on death)	β_2	0.2718	0.2671	0.2436	0.2733
Variance of the Random Effect	$\log \phi_1$	0.6385	0.0486	0.6382	0.0493
Copula Parameter	$\log \frac{1+\psi}{1-\psi}$	0.1749	0.1245	0.1723	0.1402

3.6 Discussion

In this chapter, we extended the copula-based model for joint modeling of recurrent and terminal events, initially formulated in Chapter 2, to address scenarios where the recurrent events are intermittently observed. This extension was achieved by specifying a piecewise constant baseline rate function for the recurrent events, as such models are flexible, robust, and relatively easy to implement. We developed estimation and inference methods, including a computationally efficient two-stage procedure that enables semiparametric specifications, with and without the use of an EM algorithm. For numerical studies, we

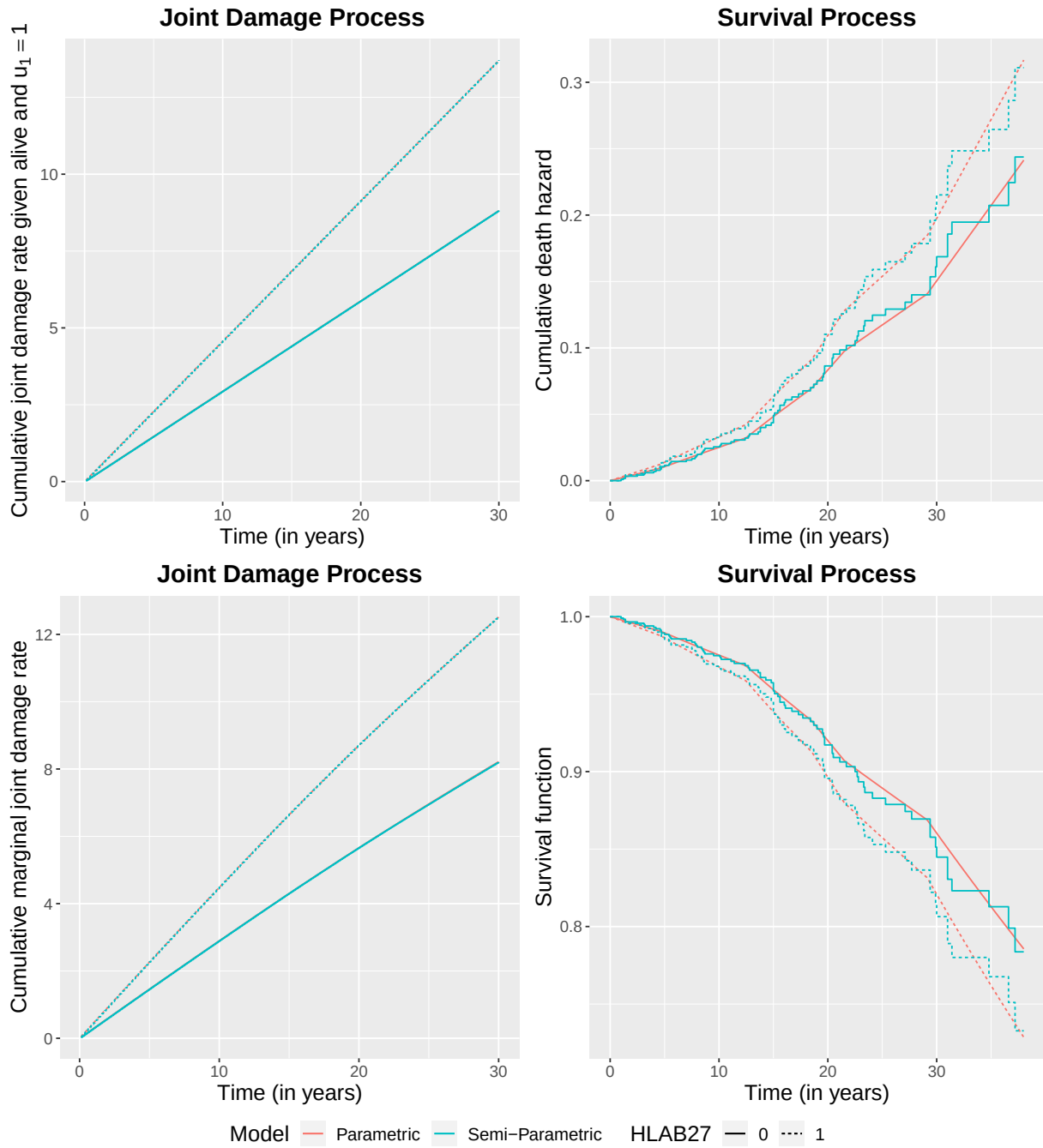


Figure 3.2: Plots of estimated conditional (on being alive and $u_{i1} = 1$) and marginal cumulative rates for the joint damage process (left panel), and plots of estimated cumulative hazard and survival function for the survival process (right panel)

specified a Gamma distributed random effect and a Gaussian copula. The results demonstrate satisfactory frequency properties of estimators obtained from both simultaneous and two-stage procedures in finite samples. The high relative efficiency of the two-stage procedure supports the practicality of semiparametric analysis. Our model preserves the neat marginal proportional hazard property for the terminal events and remains valid in situations where there is no association between the two event processes.

We note that the copula function can be chosen from any well-studied copula families, and the marginal distribution of the random effect is not limited to the Gamma distribution; other distributions, such as a log-normal distribution, can also be utilized. As our model accommodates the dependence structure using a pre-specified copula function, an important next step is to explore the sensitivity of the model to potential misspecifications of the copula function. Copula models indexed by more than one parameter could be used, but it may require larger datasets for estimability. Additionally, this model can be readily extended to scenarios involving multiple types of recurrent events and a terminal event process by increasing the dimension of the copula function. In this context, options such as Gaussian copulas, vine copulas, or hierarchical copulas may be particularly appealing.

When individuals are observed or assessed intermittently, we generally assume that the observation process is 1) non-informative about the event process and 2) independent of the event process. These assumptions are necessary to draw valid inferences from standard methods of life history analysis. However, the second condition might be violated in some settings. For example, in clinical studies, assessment times may be related to the number of event occurrences, such as when individuals visit more frequently if they experience more severe disease conditions. Dependent censoring can also occur when individuals decide to drop out of the cohort when they feel better. Such event-dependent observation processes have received attention; see [Cook and Lawless \(2007, 2021\)](#) for a general review and discussion. In this project, we assume the censoring is ignorable and independent, and the assessment times are non-informative and each is conditionally independent of the corresponding current count of the recurrent events given the past observed history. A potential extension is to accommodate event-dependent censoring; see [Cook et al. \(2010\)](#).

Chapter 4

Cost-Effective Design of Survival Studies Involving Intermittent Observation of Time-Dependent Covariates

4.1 Introduction

In studies of chronic disease, interest often lies in studying the association between time-dependent biomarker processes and a failure time of interest which may represent death, disease onset, or the progression time of a disease. Marker values are not available in continuous time, but are available at periodic clinic visits when measurements can be made on patients or blood or urine samples are drawn for testing. The Cox regression model (Cox, 1972) is routinely used to model the association between marker processes and failure but it requires the time-varying marker values to be known at all event times which may not be the times they are measured. As a result, an ad hoc approach to statistical analysis is to carry forward marker values from the times they are measured until they are updated at the next clinic visit.

The bias between the true covariate effect and the estimand arising from a naive analysis using the carry-forward approach has been investigated when the marker processes are under intermittent observation (Jiang et al., 2020; Cook et al., 2022). To address this issue, joint modeling of continuous marker processes and failure times is needed, with most models for the marker trajectory having an additive form with subject-specific random effects

accommodating heterogeneity in the marker paths and a serial dependence. The association between the marker process and the failure process is accommodated by incorporating shared or correlated random effects to the failure time model (Wulfsohn and Tsiatis, 1997; Papageorgiou et al., 2019). Conditioning on the random effects, the failure time process and the marker processes are typically assumed to be independent, but a joint model can be obtained by integrating over the random effects. Such joint models often require a strong assumption about the distributions of the marker processes (Binguet et al., 2001; Xu and Zeger, 2001) and are less appealing when the failure event is death, as the marker process does not terminate upon death (Tsiatis and Davidian, 2004). Other strategies for dealing with intermittent observation of marker processes do not involve joint modeling. de Bruijne et al. (2001) proposed a weighted Cox regression analysis where weights are introduced and defined to decay with the time since the marker was most recently measured; this reduces the bias by down-weighting contributions to the partial score function when it is a long time since the marker measurement. Other methods involving using smoothing techniques to impute the missing “current” marker values at failure times have been explored (Raboud et al., 1993; Tsiatis et al., 1995; Andersen and Liestøl, 2003).

In medical research, it is common to discretize continuous biomarkers by specifying cut-points agreed upon by researchers; in some settings these may be thresholds delineating meaningfully different disease states. Such discretization can be appealing since it enables a joint multistate model of the marker and failure processes by specification of a discrete state spacing incorporating different levels of marker states and an absorbing state for the terminal effect of death (Cook and Lawless, 2018). Moreover, in settings where marker values are recorded at clinic visits or when an individual seeks medical care, the visit times are often associated with the values of the marker process. It is then crucial to account for such association between the clinical visit process and the disease process for valid inference of the marker effect as the observation process conveys information about the disease process (Cook and Lawless, 2018, 2021).

For this project, we consider a discrete biomarker with two levels representing, say, an indicator of whether a marker is elevated (abnormal) or not. A multistate model is therefore used for the joint marker and failure processes. A more general multistate model is also formulated by incorporating a marker-dependent visit process. In practice, databases of electronic medical records offer an unprecedented opportunity to study chronic disease processes and it is too labour intensive to collect and clean data on all covariates at all times. In such settings it is common to select a single clinic visit at which covariates are measured and used for statistical inference. *Ad hoc* strategies include

- a) using the marker value measured at the last visit time before failure or censoring; the (misguided) rationale is that these values are likely the ones closest to the true

values at the times of the respective failures for these observed to fail.

- b) using the marker value measured at the last visit time under left-truncation; the rationale is to avoid carrying “backward” measurements made later in the course of follow-up.
- c) using the marker value measured at the first visit time, approximating the role of a baseline covariate and having a consistent strategy for those failing and those who are censored.
- d) using the marker value measured at the first visit time under left-truncation.

We adopt the joint multistate model framework to investigate the asymptotic bias of estimators based on these *ad hoc* strategies. We then propose and investigate a simple alternative cost-effective selection strategy for efficient choice of individuals and assessment times to measure available markers to enable consistent parameter estimation in joint models.

The remainder of this chapter is organized as follows. In Section 4.2 we define notation and formulate a three-state multistate model for joint modeling of the marker and the failure processes. The joint marker-failure multistate model is then extended to incorporate random censoring and marker-dependent visit processes. In Section 4.3 we present the Cox model and review the large sample theory associated with misspecified Cox models. We also describe four naive cost-effective *ad hoc* strategies and calculate limiting values of the estimators from misspecified Cox models based on these strategies. Determinants of the asymptotic biases of estimators based on these strategies are investigated through simulation studies. In Section 4.4, we describe an alternative simple cost-effective selection strategy where the final dataset for inference consists of a subsample of individuals with marker values measured at each visit, and a subsample of individuals with marker values measured only once at the first visit. A joint model is fitted to these selected individuals and the effect of the size of the longitudinal subsample (to the cost budget) on the efficiency of estimating the covariate effects is explored via numerical studies. We illustrate the proposed selection strategy by an application to the PsA data in Section 4.5 and concluding remarks are provided in Section 4.6.

4.2 Joint Modeling Framework

4.2.1 A Joint Marker-Failure Process

Suppose we have a time-dependent binary marker process and our primary goal is to study the effect of the marker state on the intensity of failure (e.g., death or some other clinical

endpoint). For continuous marker processes, we may discretize marker values into two levels “normal/low” or “elevated/high”. We then consider a joint model for the binary marker and the failure time depicted in Figure 4.1 where states 0 and 1 correspond to ordered marker states and state 2 is an absorbing state entered upon failure at time T . We let $Z(s)$ denote the state occupied at time s , $s \geq 0$, and $\{Z(s), 0 \leq s\}$ denote this joint 3-state marker-failure process. Note the transitions are possible in both directions between adjacent states for the marker process.

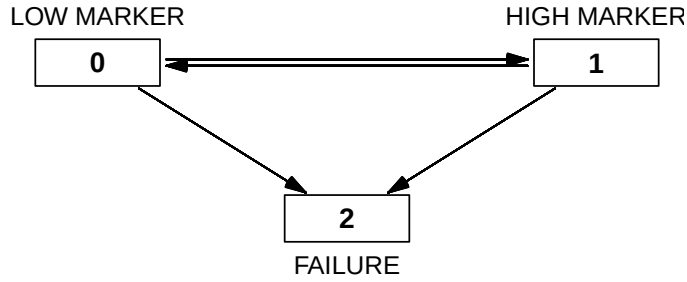


Figure 4.1: A joint marker-failure process model for a binary marker and a single absorbing state entered upon failure

Let X_2 be a $p \times 1$ vector of fixed covariates, $X_1(s) = \mathbb{1}(Z(s^-) = 1)$ indicate the time-dependent binary marker, and $X(s) = (X_1(s), X_2')'$ the overall covariates. Let $H(t) = \{Z(s), 0 \leq s < t, X_2\}$ denote the history of the joint marker-failure process up to any time t^- along with the fixed covariates. The intensity for the failure is defined as

$$\lim_{\Delta s \downarrow 0} \frac{P(Z(s + \Delta s^-) = 2 \mid Z(s^-) = k, H(s))}{\Delta s} = \lambda_{k2}(s \mid H(s)), \quad k = 0, 1, \quad (4.1a)$$

and the intensity for the transition between the marker states can be defined as

$$\lim_{\Delta s \downarrow 0} \frac{P(Z(s + \Delta s^-) = k \mid Z(s^-) = 1 - k, H(s))}{\Delta s} = \lambda_{1-k,k}(s \mid H(s)), \quad k = 0, 1. \quad (4.1b)$$

Under a conditional Markov assumption (Cook and Lawless, 2018) given the fixed covariates X_2 , we can simplify (4.1a) as

$$\lim_{\Delta s \downarrow 0} \frac{P(Z(s + \Delta s^-) = 2 \mid Z(s^-) = k, X_2)}{\Delta s} = \lambda_{k2}(s \mid X(s)), \quad k = 0, 1, \quad (4.2a)$$

and (4.1b) as

$$\lim_{\Delta s \downarrow 0} \frac{P(Z(s + \Delta s^-) = k \mid Z(s^-) = 1 - k, X_2)}{\Delta s} = \lambda_{1-k,k}(s \mid X_2), \quad k = 0, 1. \quad (4.2b)$$

Under a proportional intensity model for the transition intensities within the joint marker-failure process, (4.2a) and (4.2b) can be expressed as

$$\lambda_{k2}(s \mid X(s); \theta_2) = \lambda_2(s; \alpha_2) \exp(X'(s)\beta), \quad k = 0, 1, \quad (4.3a)$$

and

$$\lambda_{1-k,k}(s \mid X_2; \theta_k) = \lambda_k(s; \alpha_k) \exp(X'_2 \gamma_k), \quad k = 0, 1, \quad (4.3b)$$

respectively, where $\lambda_2(s; \alpha_2)$ denotes the baseline intensity to failure indexed by α_2 , and $\lambda_k(s; \alpha_k)$ denotes the baseline transition intensity from state $1 - k$ to state k between the marker states indexed by α_k . Note $\theta_2 = (\alpha'_2, \beta')'$ with $\beta = (\beta_1, \beta'_2)'$ a $(1 + p) \times 1$ vector of regression coefficients acting on the intensity to failure, and $\theta_k = (\alpha'_k, \gamma'_k)'$ with γ_k a $p \times 1$ vector of regression coefficients acting on the state $1 - k$ to state k transition between the marker states. We adopt this specification in what follows.

4.2.2 The Observation Processes

We consider the scenario where the failure time T is observed under a right-censoring scheme. Let $C = \min(\tau, W)$ be a right-censoring time where τ denotes an administrative censoring time and W a random withdrawal/censoring time. Let $Y(s) = \mathbb{1}(s \leq C)$ indicate an individual is under follow-up at time s and $Y^\dagger(s) = \mathbb{1}(Z(s^-) < 2) = \mathbb{1}(s \leq T)$ indicate an individual is at risk of failure at time s , so $\bar{Y}(s) = Y(s)Y^\dagger(s)$ indicates an individual is under follow-up and failure-free at time s . We let $V = \min(T, C)$ and assume an individual is followed up until whichever censoring or failure occurs first at time V . To incorporate random-censoring, we expand the marker-failure process $\{Z(s), 0 \leq s\}$ in Figure 4.1 by adding another absorbing state and depict the expanded joint model in Figure 4.2. By the administrative censoring time τ , an individual either has been lost to follow-up due to a random withdrawal ($Z(\tau) = 3$), has failed ($Z(\tau) = 2$), or is still under follow-up and failure-free ($Z(\tau) = 0$ or $Z(\tau) = 1$).

To model the random censoring we define $dC(s) = \mathbb{1}(W = s)$, $C(s) = \int_0^s dC(u)$, $\Delta C(s) = C(s + \Delta s^-) - C(s^-)$ and let $\{C(s), 0 \leq s\}$ denote the corresponding counting process notation. The intensity for random censoring is defined as

$$\lim_{\Delta s \downarrow 0} \frac{P(\Delta C(s) = 1 \mid \mathcal{H}(s); \theta_3)}{\Delta s} = \lambda_3(s \mid \mathcal{H}(s); \theta_3), \quad (4.4)$$

with $\mathcal{H}(s)$ explicitly defined in (4.6) and θ_3 indexing the random censoring process. A marker-dependent random censoring process can be accommodated with (4.4) but for simplicity we consider a completely independent random censoring time so that

$$\lambda_3(s \mid \mathcal{H}(s); \theta_3) = \lambda_3(s; \alpha_3), \quad (4.5)$$

where $\theta_3 = \alpha_3$ with α_3 indexing the baseline random censoring intensity.

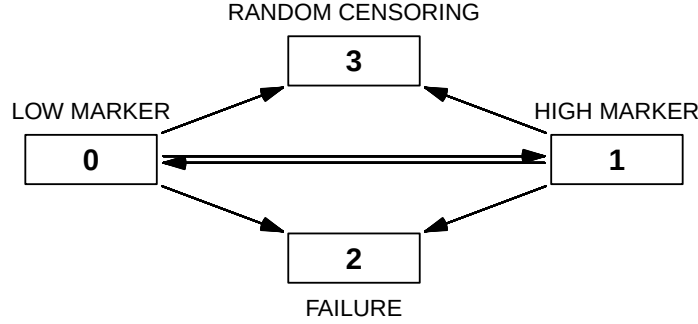


Figure 4.2: A four-state process obtained by incorporating a random censoring state to the joint marker-failure process

The marker process is not typically under continuous observation even during the follow-up period $(0, V]$, but values may be potentially available at clinical visits intermittently held during follow-up. As a result, both the number of transitions and the exact transition times within the marker process are not observable. We let $A(s)$ denote the number of clinic visits over the period $(0, s]$, $\Delta A(s) = A(s + \Delta s^-) - A(s^-)$, $dA(s) = \lim_{\Delta s \downarrow 0} \Delta A(s)$, and $\{A(s), 0 \leq s\}$ be the counting process for visits. Let

$$\mathcal{H}(t) = \{\bar{Y}(s), dA(s), Z(s), 0 \leq s < t, X_2\} \quad (4.6)$$

denote the full history of the joint marker-failure process, the censoring and the visit processes, and the covariate information up to time t^- . The intensity for the visit process is defined as

$$\lim_{\Delta s \downarrow 0} \frac{P(\Delta A(s) = 1 \mid \mathcal{H}(s))}{\Delta s} = \rho(s \mid \mathcal{H}(s)). \quad (4.7)$$

We then consider a modulated Poisson visit process (Cook and Lawless, 2007) given the covariates $X(s)$ with a proportional intensity model so that (4.7) simplifies to

$$\lim_{\Delta s \downarrow 0} \frac{P(\Delta A(s) = 1 \mid X; \theta_4)}{\Delta s} = \rho(s \mid X(s); \theta_4) = \rho_0(s; \alpha_4) \exp(X'(s)\eta), \quad (4.8)$$

where $\rho_0(s; \alpha_4)$ denotes the baseline visit intensity indexed by α_4 and $\eta = (\eta_1, \eta_2)'$ is a $(1 + p) \times 1$ vector of regression coefficients acting on the intensity to visit. Under the specification of (4.8), we have

$$E\{dA(s) \mid X(s)\} = \rho_0(s; \alpha_4) ds \exp(X'(s)\eta),$$

and we note (4.8) allows a dependent visit process governed by marker values and fixed covariates.

4.2.3 Observable Event Processes

Note that the full history $\mathcal{H}(t)$, $t > 0$ includes information from the past that is unobservable. The marker-failure process becomes unobservable if censoring occurs; similarly, the random censoring process becomes unobservable if a failure occurs. Additionally, the visit process becomes unobservable if either censoring or failure occurs. We now define the counting process notation and intensities for the observable marker, failure, and observation processes.

We let $dN_2(s) = \mathbb{1}(T = s)$ and $\{N_2(s), 0 \leq s\}$ denote the counting process for the failure with $N_2(s) = \mathbb{1}(T \leq s)$. Let $d\bar{N}_2(s) = \bar{Y}(s)dN(s)$ indicate a failure is occurred and observed at time s , $\bar{N}_2(s) = \int_0^s d\bar{N}_2(u)$, $\Delta\bar{N}_2(s) = \bar{N}_2(s + \Delta s^-) - \bar{N}_2(s^-)$ and $\{\bar{N}_2(s), 0 \leq s\}$ denote the counting process for an observable failure. The intensity of an observable failure is then

$$\lim_{\Delta s \downarrow 0} \frac{P(\Delta\bar{N}_2(s) = 1 \mid Z(s^-) = k, \mathcal{H}(s))}{\Delta s} = \bar{Y}(s)\lambda_{k2}(s \mid \mathcal{H}(s)), \quad k = 0, 1.$$

If the failure process is conditionally independent of the observation processes (Cook and Lawless, 2021), $\lambda_{k2}(s \mid \mathcal{H}(s))$ simplifies to $\lambda_{k2}(s \mid H(s))$ which is defined in (4.1a). For the observable marker process, let $dN_k(s) = \mathbb{1}(Z(s^-) = 1 - k, Z(s) = k)$, $d\bar{N}_k(s) = \bar{Y}(s)dN_k(s)$, $\bar{N}_k(s) = \int_0^s d\bar{N}_k(u)$ and $\Delta\bar{N}_k(s) = \bar{N}_k(s + \Delta s^-) - \bar{N}_k(s^-)$ for $k = 0, 1$, the observable transition intensity between the marker states is then

$$\lim_{\Delta s \downarrow 0} \frac{P(\Delta\bar{N}_k(s) = 1 \mid Z(s^-) = 1 - k, \mathcal{H}(s))}{\Delta s} = \bar{Y}(s)\lambda_{1-k,k}(s \mid \mathcal{H}(s)), \quad k = 0, 1.$$

Similarly we have $\lambda_{1-k,k}(s \mid \mathcal{H}(s)) = \lambda_{1-k,k}(s \mid H(s))$ defined in (4.1b) if the visit process does not affect the transition intensities between the marker values. For the random

censoring process, we let $d\bar{C}(s) = \bar{Y}(s)dC(s)$, $\bar{C}(s) = \int_0^s d\bar{C}(u)$ and $\Delta\bar{C}(s) = \bar{C}(s + \Delta s^-) - \bar{C}(s^-)$, the intensity for an observable random withdrawal is then

$$\lim_{\Delta s \downarrow 0} \frac{P(\Delta\bar{C}(s) = 1 \mid \mathcal{H}(s))}{\Delta s} = \bar{Y}(s)\lambda_3(s \mid \mathcal{H}(s))$$

with $\lambda_3(s \mid \mathcal{H}(s))$ defined in (4.4). For the visit process, we let $d\bar{A}(s) = \bar{Y}(s)dA(s)$, $\bar{A}(s) = \int_0^s d\bar{A}(u)$ and $\Delta\bar{A}(s) = \bar{A}(s + \Delta s^-) - \bar{A}(s^-)$. The intensity for an observable visit is then

$$\lim_{\Delta s \downarrow 0} \frac{P(\Delta\bar{A}(s) = 1 \mid \mathcal{H}(s))}{\Delta s} = \bar{Y}(s)\rho(s \mid \mathcal{H}(s))$$

where $\rho(s \mid \mathcal{H}(s))$ is defined in (4.7).

Figure 4.3 represents a more general joint model that incorporates a visit process on top of the four-state process described in Figure 4.2. Transitions from row $j - 1$ to row j arise when the j th visit is made. We let a_j denote the time of the j th visit, $j = 1, 2, \dots$, and $a_0 = 0$ denote the time of origin. Under the observation scheme described in Section 4.2.2, the observable history of the marker-failure process and the observation processes up to time t^- can be written out explicitly as

$$\bar{\mathcal{H}}(t) = \{\bar{Y}(s), d\bar{A}(s), d\bar{N}_2(s), 0 \leq s < t, \{(a_j, Z(a_j)), j = 1, \dots, \bar{A}(t^-)\}, X_2\}. \quad (4.9)$$

A conventional analysis of data under such an intermittent observation of the marker process is to carry forward recorded marker values from one assessment to the next available one and update it, or from the final assessment until censoring or failure. This means using $X_1^\circ(s) := X_1(a_{\bar{A}(s)})$ in lieu of $X_1(s)$ and fitting a standard Cox model for the failure time with covariates $X_1^\circ(s)$ and X_2 . Figure 4.4 shows the deviation of the carried-forward marker values compared to the true marker values using an illustrative example. Cook et al. (2022) investigate the asymptotic bias of covariate effects estimated from this conventional method.

4.3 Asymptotic Bias from Naive Cost-Effective Strategies

4.3.1 Cost-Effective Ad Hoc Strategies Using Cox Models

In a large database there may be biospecimens collected periodically over follow-up and stored in a biobank, but it may be prohibitively expensive to assess all of them. In settings

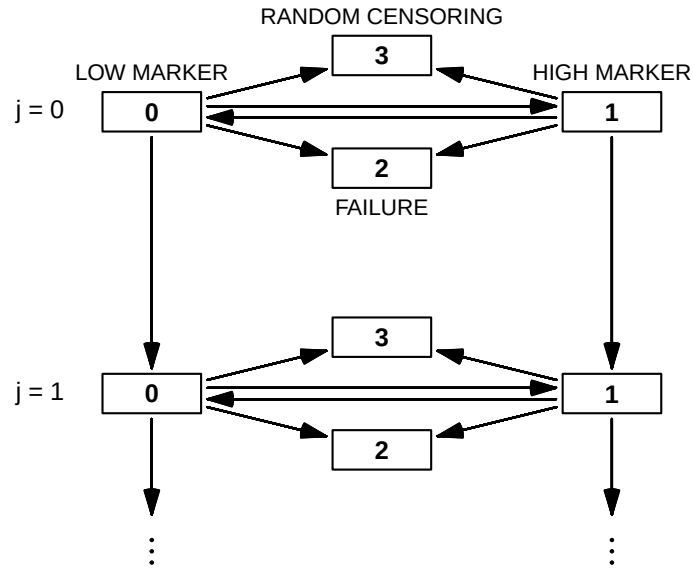


Figure 4.3: A joint model for a four-state marker-failure-censoring process and a visit process

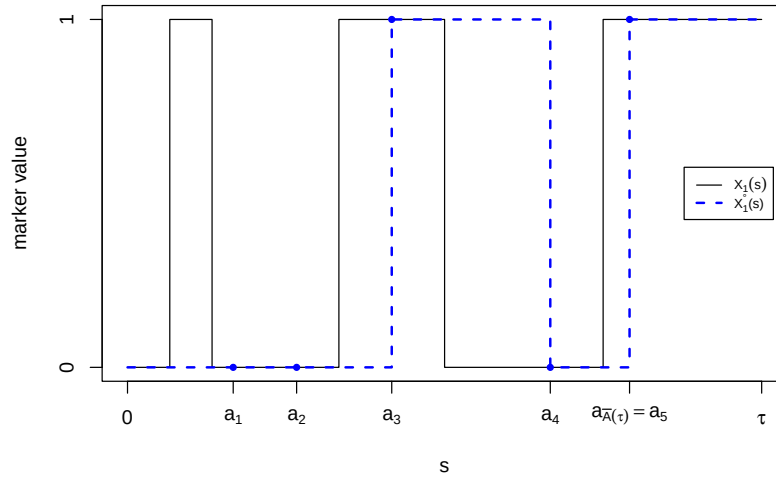


Figure 4.4: An illustration of the discordance between the observed value $X_1^o(s)$ and the true value $X_1(s)$ when marker values are carried forward under intermittent observation schemes

with electronic medical records it may be necessary to delve into clinical cohorts to retrieve more detailed information than is routinely recorded in medical records. This can be laborious and time-consuming. In such cases some researchers advocate selecting only one visit per individual for the measurement of expensive biomarkers due to budget and time constraints. Common approaches include using the first visit to measure the marker value, or the last visit before failure or censoring at time V . We let X_1° denote the marker value measured at the chosen visit, so that $X_1^\circ = X_1^\circ(V) = X_1(a_{\bar{A}(V)})$ when we select the last visit before censoring or failure, and $X_1^\circ = X_1^\circ(a_1) = X_1(a_1)$ when we select the first visit to assess the marker value. The observed history of the marker-failure and the observation processes up to time t^- is now

$$\mathcal{H}^\circ(t) = \{\bar{Y}(s), d\bar{A}(s), d\bar{N}_2(s), 0 \leq s < t, X_1^\circ, X_2\}.$$

One conventional analysis involves using X_1° in place of $X_1(s)$, treating the value as fixed either from time 0 to V which involves carrying backward and forward, or from the time of measurement to V under left-truncation (which only involves carrying forward). Fitting a Cox model for the failure time is then carried out with

$$\Gamma(s \mid X^\circ) = \Gamma_0(s) \exp(\psi' X^\circ) \quad (4.10)$$

where $X^\circ = (X_1^\circ, X_2')'$, $\psi = (\psi_1, \psi_2')'$, and $\Gamma_0(s)$ is the baseline intensity to failure under this specification. We consider four *ad hoc* strategies described in Section 4.1. The deviation of X_1° from $X_1(s)$ under these strategies is illustrated in Figure 4.5. Such a misspecification of the true model in (4.1a) with (4.10) generally results in asymptotic bias in the estimation of parameters of interest. In particular, we explore the asymptotic bias of this misspecification on estimating parameters β_1 and β_2 when the true model for the failure process is (4.3a).

Suppose we already decided which visit (e.g., the last visit) to measure the marker value and thus have a sample of observations from n independent processes yielding observed data of $\{\mathcal{H}_i^\circ(\tau), i = 1, 2, \dots, n\}$. Treating the marker value fixed with X_1° from time 0 to the time until failure or censoring at V , the Cox partial likelihood score functions based on (4.10) can be formulated as

$$U_1(s) = \sum_{i=1}^n \mathbb{1}(\bar{A}_i(\tau) \geq 1) \bar{Y}_i(s) \{dN_{i2}(s) - d\Gamma_0(s) \exp(\psi' X_i^\circ)\}, \quad s > 0, \quad (4.11a)$$

and

$$U_2 = \sum_{i=1}^n \mathbb{1}(\bar{A}_i(\tau) \geq 1) \int_0^\tau \bar{Y}_i(s) \{dN_{i2}(s) - d\Gamma_0(s) \exp(\psi' X_i^\circ)\} X_i^\circ \quad (4.11b)$$

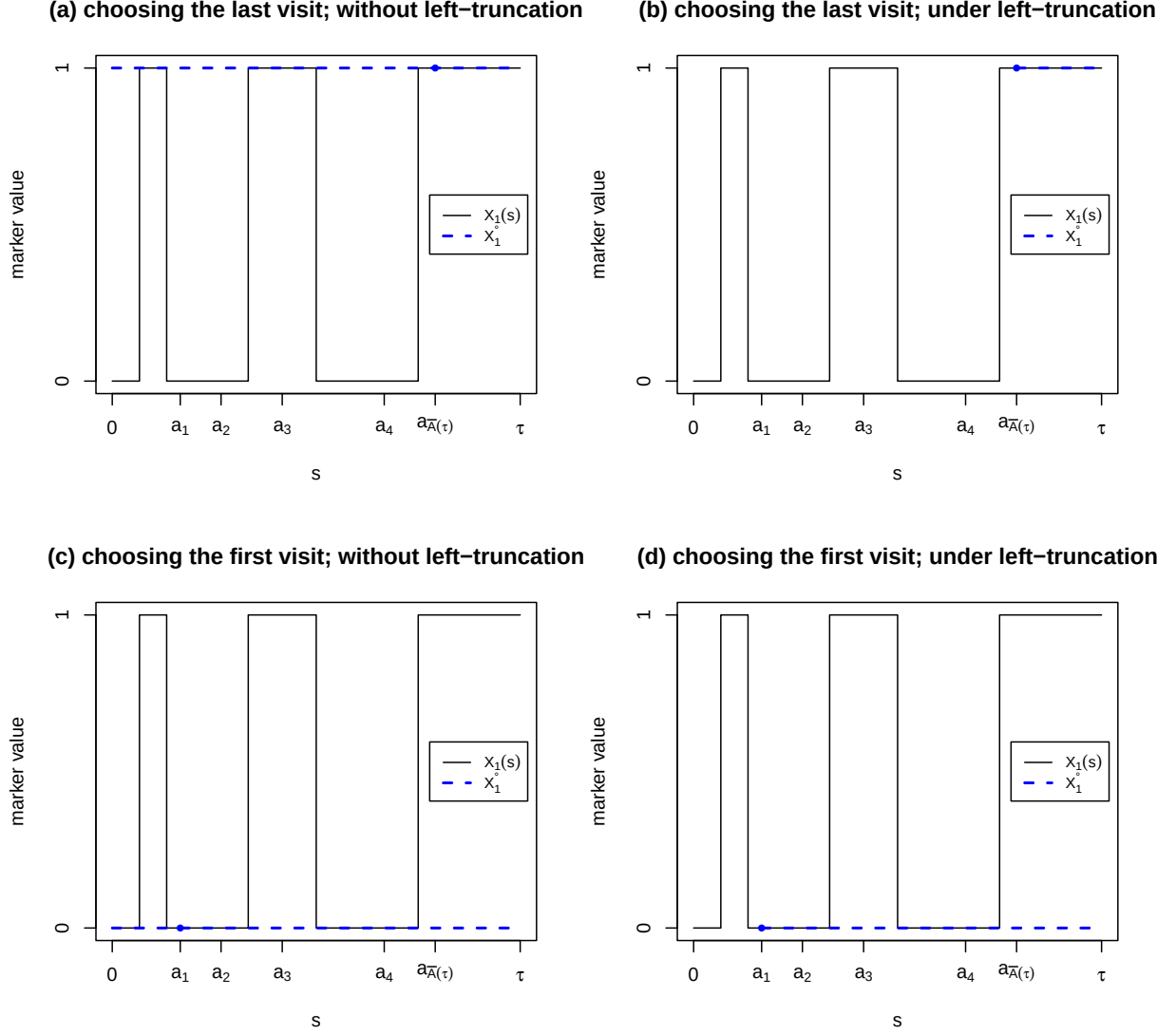


Figure 4.5: An illustration of the discordance between X_1^o and the true value $X_1(s)$ based on four *ad hoc* strategies: a) choosing the last visit to measure the marker value and carry it both forward and backward; b) choosing the last visit to measure the marker value and only carry it forward via left-truncation; c) choosing the first visit to measure the marker value and carry it both forward and backward; d) choosing the first visit to measure the marker value and only carry it forward via left-truncation;

with $X_i^\circ = (X_{i1}^\circ, X_{i2}^\circ)'$, since each individual has to make at least one visit to have a measured marker value X_{i1}° . Solving $U_1(s) = 0$ for $d\Gamma_0(s)$ at a fixed level of ψ gives the Breslow profile estimate (Lawless, 2011) modified to reflect the requirement of at least one visit

$$d\tilde{\Gamma}_0(s; \psi) = \frac{\sum_{i=1}^n \mathbb{1}(\bar{A}_i(\tau) \geq 1) \bar{Y}_i(s) dN_{i2}(s)}{\sum_{i=1}^n \mathbb{1}(\bar{A}_i(\tau) \geq 1) \bar{Y}_i(s) \exp(\psi' X_i^\circ)}, \quad s > 0. \quad (4.12)$$

Upon substitution into (4.11b) we obtain a partial profile score function for ψ as

$$U(\psi) = \sum_{i=1}^n \int_0^\tau \mathbb{1}(\bar{A}_i(\tau) \geq 1) \bar{Y}_i(s) \left\{ X_i^\circ - \frac{\sum_{i=1}^n R_i^{(1)}(s; \psi)}{\sum_{i=1}^n R_i^{(0)}(s; \psi)} \right\} dN_{i2}(s), \quad (4.13)$$

where

$$R_i^{(k)}(s; \psi) = \mathbb{1}(\bar{A}_i(\tau) \geq 1) \bar{Y}_i(s) (X_i^\circ)^{(k)} \exp(\psi' X_i^\circ) \quad (4.14)$$

for $k = 0, 1$, with $a^{(k)} = 1$ for $k = 0$ and $a^{(k)} = a$ for $k = 1$. Setting $U(\psi) = 0$ and solving for ψ gives $\tilde{\psi}$. Upon substituting $\tilde{\psi}$ into (4.12), we obtain $d\tilde{\Gamma}_0(s; \tilde{\psi})$ and thus $\Gamma_0(s; \tilde{\psi}) = \int_0^s d\tilde{\Gamma}_0(u; \tilde{\psi})$.

The limiting value of $\tilde{\psi}$ is denoted by ψ^* , and is the solution to

$$E\{U(\psi)\} = n \int_0^\tau \left\{ r^{(1)}(s) - \frac{r^{(1)}(s; \psi)}{r^{(0)}(s; \psi)} r^{(0)}(s) \right\} = 0, \quad (4.15)$$

where

$$r^{(k)}(s) = E\{R_i^{(k)}(s)\} \quad (4.16a)$$

and

$$r^{(k)}(s; \psi) = E\{R_i^{(k)}(s; \psi)\} \quad (4.16b)$$

for $k = 0, 1$, with

$$R_i^{(k)}(s) = \mathbb{1}(\bar{A}_i(\tau) \geq 1) \bar{Y}_i(s) (X_i^\circ)^{(k)} dN_{i2}(s) \quad (4.17)$$

and $R_i^{(k)}(s; \psi)$ defined in (4.14). Note the expectations are taken with respect to the true data generating model. More generally, $\tilde{\psi}$ is consistent for ψ^* , and as $n \rightarrow \infty$,

$$\sqrt{n}(\tilde{\psi} - \psi^*) \sim \text{MVN}(0, \mathcal{A}^{-1}(\psi^*) \mathcal{B}(\psi^*) [\mathcal{A}^{-1}(\psi^*)]') \quad (4.18)$$

where $\mathcal{A}(\psi) = E\{-\partial U(\psi)/\partial \psi\}$ and $\mathcal{B}(\psi) = E\{U(\psi)U'(\psi)\}$ (Andersen and Gill, 1982; Struthers and Kalbfleisch, 1986; Lin and Wei, 1989).

Alternatively, if we consider the marker value left-truncated at the selected visit time we denote generally by L_i (i.e., the time of the measurement) and only carry forward X_{i1}°

from L_i to V_i , then the Cox partial likelihood score functions, the estimators $\tilde{\psi}$ and $\Gamma_0(s; \tilde{\psi})$, as well as the limiting value ψ^* and its asymptotic property, follow the results given in (4.11a) to (4.18), but with $\bar{Y}_i(s) = \bar{Y}_i(s)Y_i^\dagger(s)Y_i^L(s)$ where $Y_i^L(s) = \mathbb{1}(s \geq L_i)$. If the last visit is chosen to measure the marker value, $L_i = a_{i\bar{A}_i(V_i)}$; if the first visit is chosen to measure the marker value, then $L_i = a_{i1}$.

For what follows we focus on the scenario where the random censoring is independent as specified in (4.5) and the visit process does not affect the nature of the transition intensities between the joint marker-failure process, so that

$$\lambda_{k2}(s \mid \mathcal{H}(s)) = \lambda_{k2}(s \mid H(s)) \quad \text{and} \quad \lambda_{1-k,k}(s \mid \mathcal{H}(s)) = \lambda_{1-k,k}(s \mid H(s))$$

for $k = 0, 1$. In addition, we consider the particular setting of a conditional Markov process with a proportional intensity assumption for the joint marker-failure process, so that the true models for the marker-failure process are $\lambda_{k2}(s \mid H(s)) = \lambda_{k2}(s \mid X(s))$ and $\lambda_{1-k,k}(s \mid H(s)) = \lambda_{1-k,k}(s \mid X_2)$ for $k = 0, 1$, with explicit forms specified in (4.3a) and (4.3b). We discuss the implications of specifying (4.10) instead of (4.3a) under the observation scheme described in Section 4.2.2. Specifically, we consider strategies of selecting the last or the first visit to measure the marker value with the Cox model conducted without or with left-truncation. Our goal is to investigate the relative asymptotic bias of $\hat{\psi}$ on the true parameter β , defined here as $(\psi^* - \beta)/\beta$.

4.3.2 Calculations of Expectations

In order to calculate the expectations required to evaluate the limiting value ψ^* through (4.15), we describe a multistate model which is a further expansion to the joint model of Figure 4.3 in Figure 4.6, where states are labeled as $\mathcal{Z}(s) := (\bar{A}(s), Z(s), X_1^\circ(s))$; the first index is the number of observable visits up to time s , the second denotes the four-state process of Figure 4.2 at time s , and the third is the marker value recorded at the most recent assessment before time s . The states are arranged so that those in the left column of Figure 4.6 are occupied by individuals with $X_1^\circ(s) = 0$ and those on the right are occupied by individuals with $X_1^\circ(s) = 1$. Transitions from the upper row to the lower row of states indicate a visit is occurred and observed, and once entered one row the first and third state indices remain the same within the row. This means transitions from one column to another column of states can only occur when a new visit is observed and the marker value recorded at this current visit is different than the marker value recorded at the last visit. This figure thus contains information on the latent marker process, the observable failure and observation processes, and the marker values recorded at each observable visit. Such

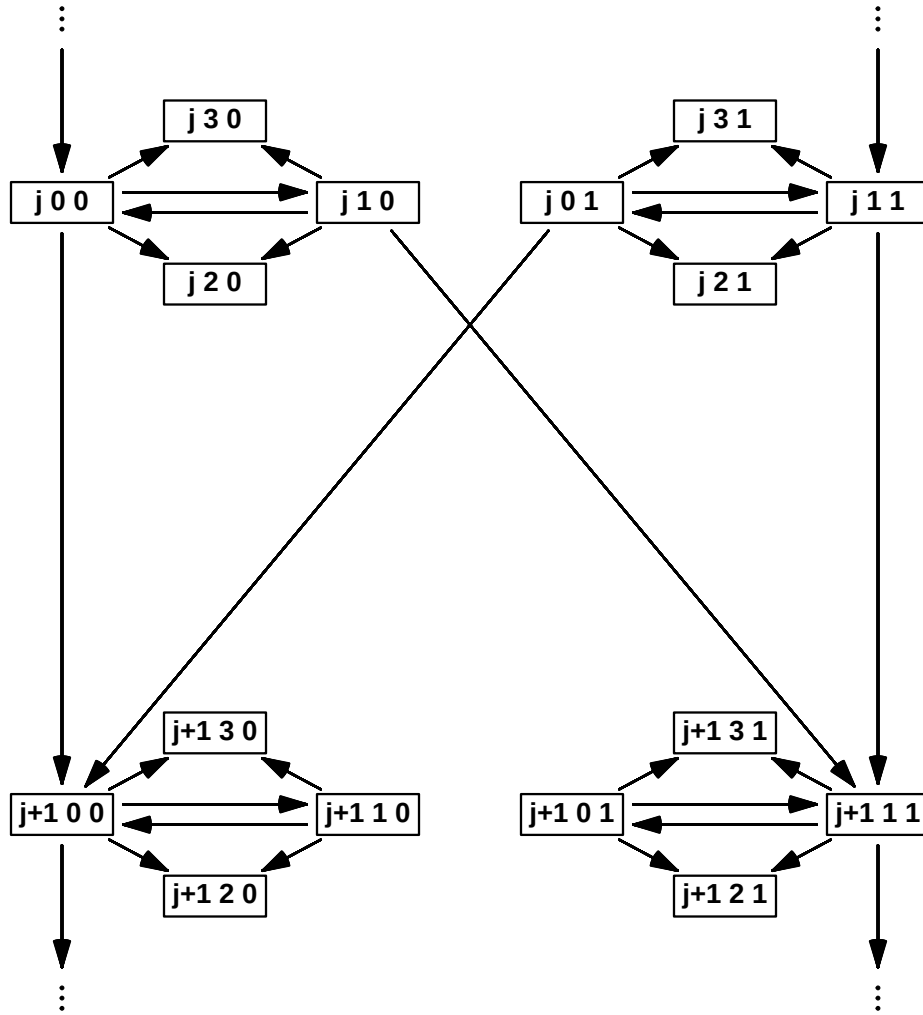


Figure 4.6: A multistate model of $\mathcal{Z}(s) = (\bar{A}(s), Z(s), X_1^o(s))$ depicting a more general joint model for the marker-failure process and the observations processes

a model was considered by [Cook et al. \(2022\)](#). We next define classes of states which are convenient for future calculations. We let $\mathcal{C}_j^a = \{(a, z, x_1^\circ) : a = j\}$ denote a set of states occupied by individuals with the j th follow-up visit, $\mathcal{C}_k^z = \{(a, z, x_1^\circ) : z = k\}$ denote a set of states occupied by individuals who are in state k for the four-state process depicted in Figure 4.2, and $\mathcal{C}_l^{x_1^\circ} = \{(a, z, x_1^\circ) : x_1^\circ = l\}$ denote a set of states occupied by individuals whose most recently recorded marker value is l .

Based on this general multistate model, we calculate $r^{(k)}(s)$ and $r^{(k)}(s; \psi)$, $k = 0, 1$, under each of the four *ad hoc* strategies described in Section 4.3.1. Note $r^{(0)}(s)$ can be written as

$$r^{(0)}(s) = E_{Z_i(0), X_{i2}} \left\{ E \left\{ R_i^{(0)}(s) \mid Z_i(0), X_{i2} \right\} \right\},$$

and $r^{(1)}(s)$ can be partitioned into $r^{(1)}(s) = (r_1^{(1)}(s), [r_2^{(1)}(s)]')'$ where

$$r_1^{(1)}(s) = E_{Z_i(0), X_{i2}} \left\{ E \left\{ R_{i1}^{(1)}(s) \mid Z_i(0), X_{i2} \right\} \right\}$$

and

$$\begin{aligned} r_2^{(1)}(s) &= E_{Z_i(0), X_{i2}} \left\{ E \left\{ R_{i2}^{(1)}(s) \mid Z_i(0), X_{i2} \right\} \right\} \\ &= E_{Z_i(0), X_{i2}} \left\{ X_{i2} E \left\{ R_i^{(0)}(s) \mid Z_i(0), X_{i2} \right\} \right\}, \end{aligned}$$

with $R_i^{(0)}(s)$ defined in (4.17),

$$R_{i1}^{(1)}(s) = \mathbb{1}(\bar{A}_i(\tau) \geq 1) \bar{Y}_i(s) X_{i1}^\circ dN_{i2}(s),$$

and

$$R_{i2}^{(1)}(s) = \mathbb{1}(\bar{A}_i(\tau) \geq 1) \bar{Y}_i(s) X_{i2} dN_{i2}(s).$$

Similarly we can write $r^{(0)}(s; \psi)$ as

$$r^{(0)}(s; \psi) = E_{Z_i(0), X_{i2}} \left\{ E \left\{ R_i^{(0)}(s; \psi) \mid Z_i(0), X_{i2} \right\} \right\}$$

and partition $r^{(1)}(s; \psi)$ into $r^{(1)}(s; \psi) = (r_1^{(1)}(s; \psi), [r_2^{(1)}(s; \psi)]')'$ with

$$r_1^{(1)}(s; \psi) = E_{Z_i(0), X_{i2}} \left\{ E \left\{ R_{i1}^{(1)}(s; \psi) \mid Z_i(0), X_{i2} \right\} \right\}$$

and

$$\begin{aligned} r_2^{(1)}(s; \psi) &= E_{Z_i(0), X_{i2}} \left\{ E \left\{ R_{i2}^{(1)}(s; \psi) \mid Z_i(0), X_{i2} \right\} \right\} \\ &= E_{Z_i(0), X_{i2}} \left\{ X_{i2} E \left\{ R_i^{(0)}(s; \psi) \mid Z_i(0), X_{i2} \right\} \right\}, \end{aligned}$$

where $R_i^{(0)}(s; \psi)$ is defined in (4.14),

$$R_{i1}^{(1)}(s; \psi) = \mathbb{1}(\bar{A}_i(\tau) \geq 1) \bar{Y}_i(s) X_{i1}^\circ \exp(\psi' X_i^\circ),$$

and

$$R_{i2}^{(1)}(s; \psi) = \mathbb{1}(\bar{A}_i(\tau) \geq 1) \bar{Y}_i(s) X_{i2} \exp(\psi' X_i^\circ).$$

Note the terms $E\{R_i^{(0)}(s) \mid Z_i(0), X_{i2}\}$, $E\{R_{i1}^{(1)}(s) \mid Z_i(0), X_{i2}\}$, $E\{R_i^{(0)}(s; \psi) \mid Z_i(0), X_{i2}\}$ and $E\{R_{i1}^{(1)}(s; \psi) \mid Z_i(0), X_{i2}\}$ may take different forms when different *ad hoc* strategies are used. Detailed calculations of these terms using the multistate model depicted in Figure 4.6 can be found in Appendix C.1.

4.3.3 Investigating Determinants of the Limiting Value ψ^*

Next we investigate the limiting value ψ^* as a function of the parameters of the transition intensities in Figure 4.6 under above *ad hoc* strategies. We set the initial state occupancy of the low marker value to be $P(Z(0) = 0) = \pi_0 = 0.5$ and set the administrative censoring time to be $\tau = 1$ without loss of generality. We consider a single binary fixed covariate X_2 with $P(X_2 = 1) = 0.5$. For the conditional Markov marker-failure process with proportional hazard transition intensities specified in (4.3a) and (4.3b), we set the baseline transition intensities to be time homogeneous so that $\lambda_2(s; \alpha_2) = \lambda_2$ and $\lambda_k(s; \alpha_k) = \lambda_k$ for $k = 0, 1$. For the failure intensities we set $\beta_1 = \log 1.5$ so that an elevated marker confers a 50% increase in the intensity to failure, and we set $\beta_2 = \log 1.25$ so that X_2 is also a risk factor. For the marker process we set $\gamma_0 = \log 0.9$ and $\gamma_1 = \log 1.1$ so $X_2 = 1$ elevates the $0 \rightarrow 1$ transition intensity but reduces the $1 \rightarrow 0$ intensity. We let $\pi_1 = \lambda_1/(\lambda_1 + \lambda_2)$ denotes the odds of a $0 \rightarrow 1$ transition when the process enters state 0 with $X_2 = 0$, and set $\pi_1 = 0.75$. We then set $r_{01} = (\lambda_0 + \lambda_2 \exp(\beta_1))/(\lambda_1 + \lambda_2) = 2$ so that the mean sojourn time in state 0 is twice the mean sojourn time in state 1 when $X_2 = 0$. With all other parameters specified as described above, we control the marginal failure rate at the administrative censoring time for the marker-failure process by letting $\pi_2 = P(Z(\tau) = 2) = E_{Z(0), X_2} \{P(Z(\tau) = 2 \mid Z(0), X_2)\}$ and solving for λ_2 to satisfy $\pi_2 = 0.8$.

Having specified the parameters for the three-state marker-failure process in Figure 4.1, we now consider the expanded four-state model in Figure 4.2. We consider an independent random censoring with an intensity specified in (4.5) with a time-homogeneous model so that $\lambda_3(s; \alpha_3) = \lambda_3$. Let $\pi_3 = P(Z(\tau) = 3) = E_{Z(0), X_2} \{P(Z(\tau) = 3 \mid Z(0), X_2)\}$ denote the marginal random censoring rate by the administrative censoring time, and we solve

$\pi_3 = 0.2$ for λ_3 to ensure a 20% random censoring rate. Lastly we consider a conditional Poisson visit process with a proportional intensity for visits specified in (4.8). We let the visit process to be time-homogenous so that $\rho_0(s; \alpha_4) = \rho_0$. We consider both a completely independent visit process ($\eta_1 = \eta_2 = 0$ in (4.8)) and a marker-dependent visit process ($\eta_1 \neq 0, \eta_2 = 0$). For the former case we investigate the limiting value ψ^* as a function of ρ_0 . For the latter case, we investigate ψ^* as a function of the marker effect on the visits η_1 ; we let

$$\mu = E\{\bar{A}(\tau)\} = E_{Z(0), X_2} \left\{ \sum_{j=0}^{\infty} j \cdot P(\mathcal{Z}(\tau) \in \mathcal{C}_j^a \mid Z(0), X_2) \right\}$$

and for each specified value of η_1 , we solve for ρ_0 such that $\mu = 4, 8, 12$. The relative asymptotic bias of $\hat{\psi}$, i.e., $(\psi_1^* - \beta_1)/\beta_1$ and $(\psi_2^* - \beta_2)/\beta_2$, obtained from the four *ad hoc* strategies with either an independent or marker-dependent visit process are shown in Figures 4.7 and 4.8 respectively. The solids dots in Figure 4.7 represent the average estimate from a naive Cox model fitted under the corresponding assumptions; 500 samples were generated of 1000 individuals each. It is clear to see that these empirical averages coincide well with the corresponding limiting values.

Figure 4.7 contains plots of the asymptotic relative bias of estimators of the effects of X_1° (left column) and X_2 (right column) on the failure time process defined as $(\psi_1^* - \beta_1)/\beta_1$ and $(\psi_2^* - \beta_2)/\beta_2$ respectively. The four naive sampling strategies are represented in the four rows of the figure. The results demonstrate that even with an independent visit process, selecting only one visit to measure the marker value and carrying it forward or backward as well will render the estimator $\hat{\psi}$ inconsistent for β . All four-strategies severely underestimate the marker effect. For the effect of fixed covariate, all strategies overestimate the effect except when X_1° is measured at the first visit and treated as left-truncated. Specifically, when X_1° is measured at the last visit and carried both backward and forward, there is a trend of decreasing bias of both $\hat{\psi}_1$ and $\hat{\psi}_2$ as the visit intensity ρ_0 increases. When X_1° is measured at the last visit but carried only forward, we still observe a decreasing bias of $\hat{\psi}_1$ as ρ_0 increases, but the bias of $\hat{\psi}_2$ goes to the other direction and significantly increases when ρ_0 increases. For the strategy where X_1° is measured at the first visit and carried both backward and forward, the asymptotic bias of $\hat{\psi}_1$ tends to increase while the bias of $\hat{\psi}_2$ tends to decrease with an increasing ρ_0 . There is no trend of change of the biases of $\hat{\psi}_1$ and $\hat{\psi}_2$ when X_1° is carried only forward from the first visit.

Figure 4.8 shows analogous results that in general all strategies provide a conservative estimate of the marker effect and a progressive estimate for the fixed covariate effect, except when X_1° is carried only forward from the last visit. A negative value of η_1 indicates a negative association between the marker and the visit intensity. As η_1 increases from

−2 to 2, the association between the marker and the visit process goes from negative to positive, and the strength of the association first decreases (to 0) and then increases. When X_1° is carried only forward from the first visit, no trend is observed as η_1 increases. In contrast, a trend of increasing ψ_1^* and ψ_2^* is apparent for all other strategies, with the rate of increase being higher with greater visit intensities.

4.4 Cost-Effective Design of Selection Schemes

Having established that any of the *ad hoc* approaches that we considered yield inconsistent parameter estimators with appreciable asymptotic bias, we next turn to the design questions of how to develop sampling strategies to facilitate joint modeling and the possibility of consistent parameter estimation. We do this by sampling some individuals for the complete evaluation of their samples over assessments and some individuals for a single assessment of the baseline (first visit) value.

The rationale for this simple scheme is that those selected for the complete assessment of their samples will enable estimation of the transition intensities for the marker process which enables suitable use of data from those where a single sample is chosen. There are, however, individuals who experience events that are not sub-sampled in this scheme so there is the potential for some loss of information. Suppose we have n individual processes, each with $m_i > 0$ clinical visits by the administrative censoring time for individual i , $i = 1, \dots, n$. For all individuals their biospecimens are sampled and stored at each clinical visit. The total number of measurements of the marker value is denoted by $m_\cdot = \sum_{i=1}^n m_i$. Suppose instead of accessing all $m_\cdot$ measurements, budgetary constraints mean only an average number of μ_b ($\mu_b < m_\cdot$) measurements can be made across all n individuals. In Section 4.3, $\mu_b = n$ and we investigated the allocation of the μ_b measurements such that for each individual we access one single measurement of the marker value at a chosen clinical visit. We demonstrated that the Cox estimators of regression coefficients are asymptotically biased regardless of which visit these marker values are measured and whether they are treated as left-truncated. The nature of the bias, however, varied according to the sampling strategy.

The bias induced by using the naive *ad hoc* strategies can be mitigated by fitting a joint model based on Figure 4.3 with intermittently observed data $\mathcal{H}(V_i^+)$ defined in (4.9), and

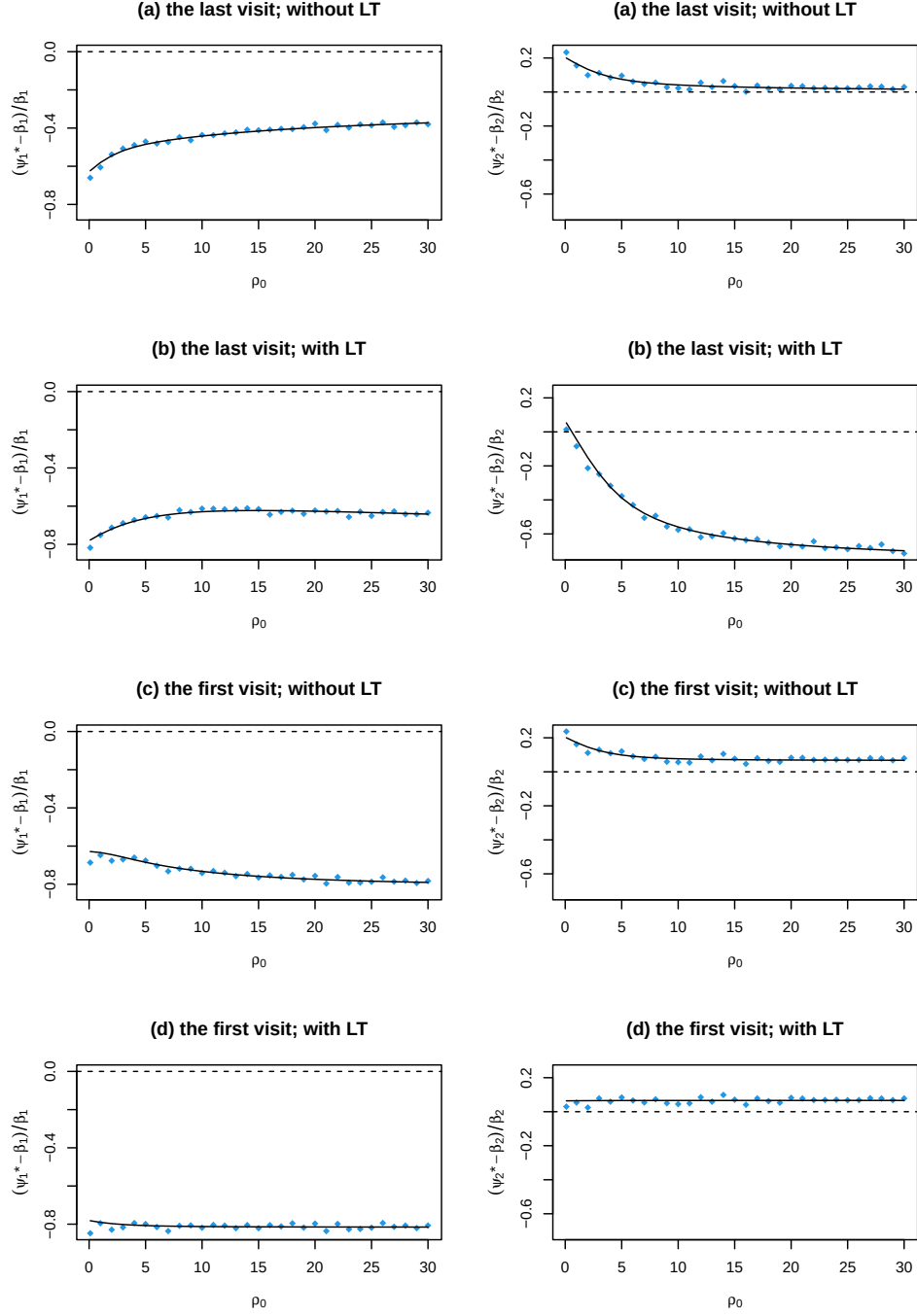


Figure 4.7: Relative asymptotic biases of $\hat{\psi}_1$ (left panel) and $\hat{\psi}_2$ as a function of the baseline visit intensity ρ_0 with an independent visit process under four *ad hoc* strategies; the solids dots represent estimators averaged over 500 simulations with a sample size of 1000.

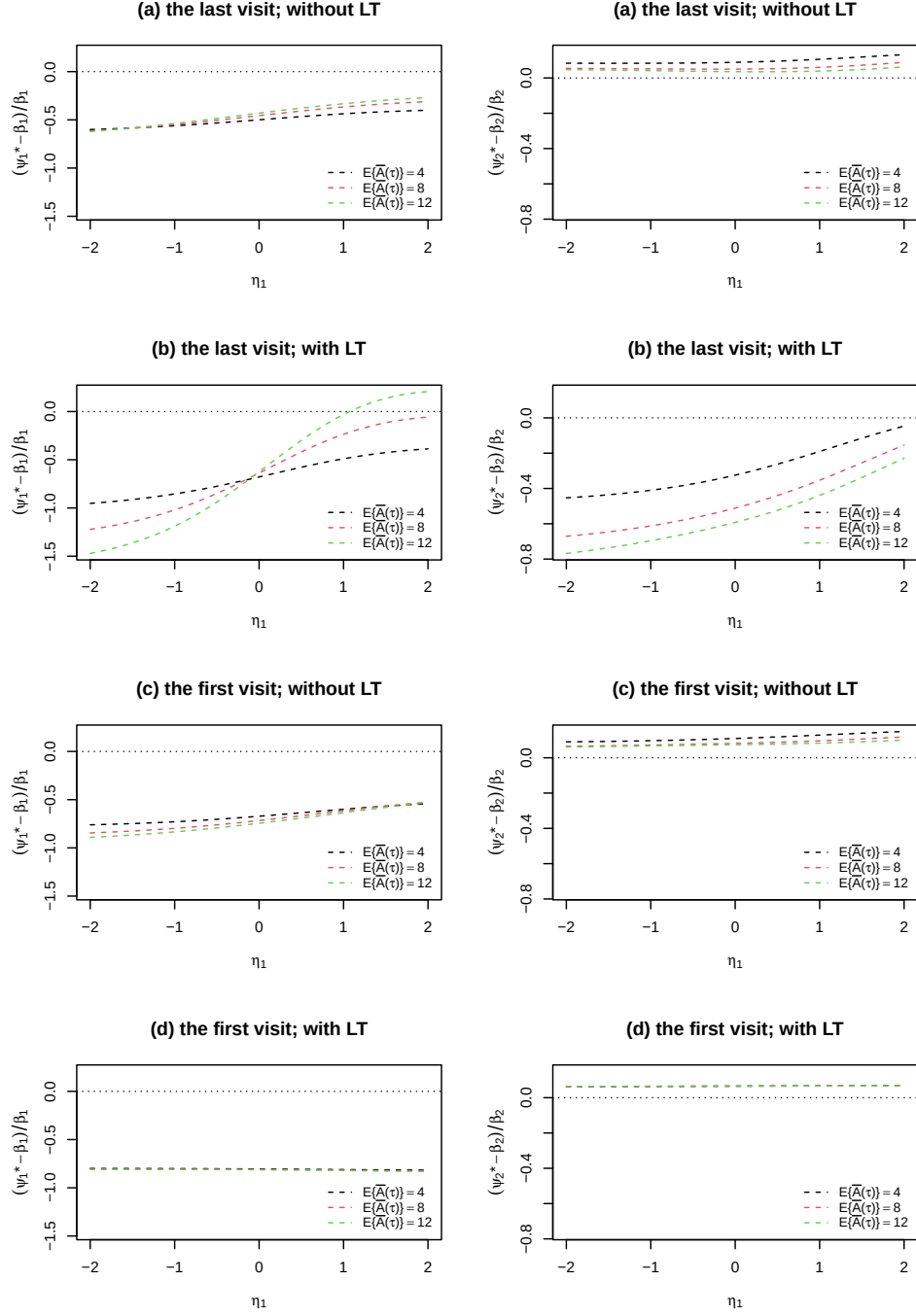


Figure 4.8: Relative asymptotic biases of $\hat{\psi}_1$ (left panel) and $\hat{\psi}_2$ as a function of the marker effect on the visit intensity (i.e., η_1) for $\mu = 4, 8, 12$ with a marker-dependent visit process under four *ad hoc* strategies

the partial likelihood is given by

$$\prod_{i=1}^n \left\{ \prod_{j=1}^{\bar{A}_i(V_i)} P(Z_i(a_{ij}) \mid Z_i(a_{i,j-1}), X_{i2}) \right. \\ \left. \times \sum_{k=0}^1 P(Z_i(V_i^-) = k \mid Z_i(a_{i\bar{A}_i(V_i)}), X_{i2}) \lambda_{k2}(V_i \mid X_{i1}(V_i^-) = k, X_{i2})^{\delta_i} \right\}.$$

Thus it is natural to consider an alternative cost-effective selection strategy that enables joint modeling: we select n_L individuals and access all of their marker values (i.e., at each clinical visit); we then select n_S individuals from the rest to access their marker values at their first visit. Let R_{iL} indicate individual i is selected to be followed longitudinally for complete assessment of marker values over time, R_{iS} indicate the individual is selected for a single measurement of the marker value at the first visit, $i = 1, \dots, n$, then $n_L = \sum_{i=1}^n R_{iL}$ and $n_S = \sum_{i=1}^n R_{iS}$. Note under this selection strategy, an individual is either selected to be in the longitudinal subsample ($R_{iL} = 1, R_{iS} = 0$), or selected to be in the single-measurement subsample ($R_{iL} = 0, R_{iS} = 1$), or not selected at all ($R_{iL} = 0, R_{iS} = 0$). Let

$$m_L = \sum_{i=1}^n R_{iL} m_i$$

denote the total number of visits in the longitudinal subsample, we denote the total number of marker measurements selected as $m_b = m_L + n_S$ with $E\{m_b\} = \mu_b$ to satisfy the cost constraint. A question then naturally arises: how should the budget be allocated between the longitudinal and single-measurement subsamples? In other words, what fraction of m_b should be allocated to m_L ?

We consider an independent random sampling so that (R_{iL}, R_{iS}) follows a trinomial distribution with $\pi_{iL} = P(R_{iL} = 1)$ and $\pi_{iS} = P(R_{iS} = 1)$. For simplicity, here we focus on the setting where the observation process is independent and non-informative. Under the conditional Markov assumption, the partial observed data likelihood contributed from one individual can be written as

$$L_i \propto \{\pi_L L_{iL}\}^{R_{iL}} \{\pi_S L_{iS}\}^{R_{iS}}$$

where L_{iL} , L_{iS} denote the contributions from the longitudinal and single-measurement

subsamples respectively, with explicit forms

$$L_{iL} \propto \prod_{j=2}^{m_i} P(Z_i(a_{ij}) \mid Z_i(a_{i,j-1}), X_{i2}) \\ \times \sum_{k=0}^1 P(Z_i(V_i^-) = k \mid Z_i(a_{im_i}), X_{i2}) \lambda_{k2}(V_i \mid X_{i1}(V_i^-) = k, X_{i2})^{\delta_i},$$

and

$$L_{iS} \propto \sum_{k=0}^1 P(Z_i(V_i^-) = k \mid Z_i(a_{i1}), X_{i2}) \lambda_{k2}(V_i \mid X_{i1}(V_i^-) = k, X_{i2})^{\delta_i}.$$

where $V_i = \min(T_i, C_i)$ and $\delta_i = \mathbb{1}(V_i = T_i)$. With a sample of n independent individual processes, the overall partial observed data likelihood is then $L(\theta) = \sum_{i=1}^n L_i(\theta)$ where $\theta = (\theta'_0, \theta'_1, \theta'_2)'$ denotes parameters associated with the marker-failure process defined in (4.3a) and (4.3b). The Fisher information $\mathcal{I}(\theta) = E\{\partial L_i(\theta)/\partial\theta\partial\theta'\}$ can be written as a weighted sum of the Fisher information contributed from the longitudinal and single-measurement subsamples

$$\mathcal{I}(\theta) = \pi_L \mathcal{I}_L(\theta) + \pi_S \mathcal{I}_S(\theta),$$

where

$$\mathcal{I}_L(\theta) = E\{\partial L_{iL}(\theta)/\partial\theta\partial\theta'\} \quad \text{and} \quad \mathcal{I}_S(\theta) = E\{\partial L_{iS}(\theta)/\partial\theta\partial\theta'\}.$$

Let $\bar{\mu} = E\{\bar{A}_i(\tau) \mid \bar{A}_i(\tau) \geq 1\}$ denote the expected number of visits per individual by the administrative censoring time conditioning on making at least one visit. Taking the expectation of $m_b = m_L + n_S$ on both sides gives the relation

$$\mu_b/n = \pi_L \bar{\mu} + \pi_S. \quad (4.21)$$

With a pre-specified value of μ_b/n and $\bar{\mu}$, we can solve π_S for a specific value of π_L . Hence the asymptotic variance of regression coefficients is a function of π_L and such we can find an optimal value of π_L that minimizes the asymptotic variance.

Next we investigate the asymptotic variance of $\hat{\beta}$ when we vary the fraction

$$E\{m_L\}/\mu_b = n\pi_L \bar{\mu}/\mu_b, \quad (4.22)$$

which represents the relative proportion of the measurements in the longitudinal subsample to the total budget. We follow the model specifications and parameter settings of Section

4.3.3 and let the failure rate be $\pi_2 = 0.1, 0.25, 0.5$, or 0.8 . We set $\mu_b/n = 0.5, 1$ and $\bar{\mu} = 4, 8$, and plot the asymptotic relative efficiency of $\hat{\beta}$, defined here as

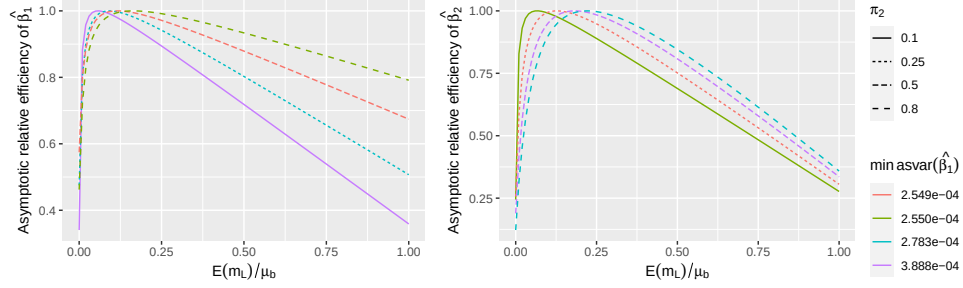
$$\text{ARE}(\hat{\beta}) = \frac{\text{asvar}(\hat{\beta})}{\min_{E\{m_L\}/\mu_b} \text{asvar}(\hat{\beta})},$$

as a function of $E\{m_L\}/\mu_b$ in Figure 4.9. The optimal values of $E\{m_L\}/\mu_b$ vary between 5% to 25% and the corresponding asymptotic variances of $\hat{\beta}$ are provided in the legends of the plots. Across all parameter settings, there is a common pattern: the asymptotic relative efficiency of β increases rapidly and then gradually decreases when the relative proportion of m_L increases from 0 to 1. This is aligned with the fact that when $E\{m_L\}/\mu_b$ is smaller, we have more individuals with fewer repeated measurements of the marker value so that we lose information about the marker transitions; while $E\{m_L\}/\mu_b$ is larger, on the other hand, we have fewer individuals with more repeated marker measurements, so we lose information on the failure time. Comparing various levels of failure rate π_2 , the optimal fraction of $E\{m_L\}/\mu_b$ is larger with higher π_2 ; additionally, the rate of change in the relative efficiency is also larger. The asymptotic variances are overall smaller when we have more budget ($\mu_b/n = 0.5$ vs $\mu_b/n = 1$).

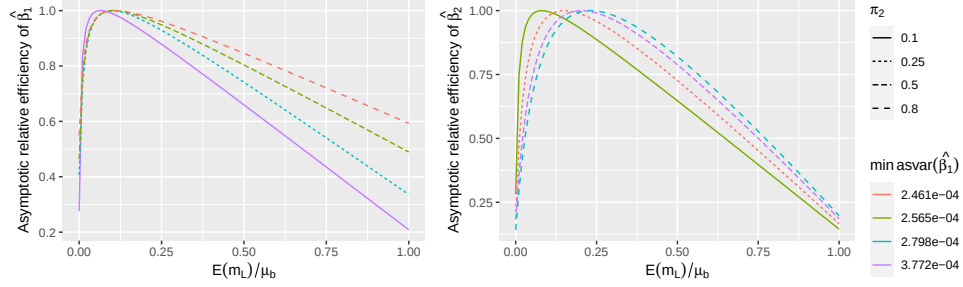
4.5 Illustrative Application to Psoriatic Arthritis Cohort Data

An illustrative example of the proposed selection strategy is given here by applying the method to the PsA cohort data described in Section 1.3.2. C-reactive protein (CRP) is a substance produced by the liver in response to inflammation, and a CRP blood test measures its level to detect inflammation from acute conditions or to monitor the severity of disease in chronic conditions. Erythrocyte sedimentation rate (ESR) is another blood test used to detect inflammation in the body. While similar to CRP in its purpose, ESR measures different aspects and can sometimes be used together with CRP for a more comprehensive view of inflammation. Both CRP and ESR measurements can only be taken during periodic clinical visits through blood tests. The primary interest here lies in studying the effect of the time-dependent covariate denoted ESR_CRP (which indicates if either ESR or CRP is elevated) on the time to the development of five or more damaged joints in patients with PsA. This represents a sufficient degree of damage that it satisfies the condition for a severe form of arthritis called arthritis mutilans (Gladman et al., 1987).

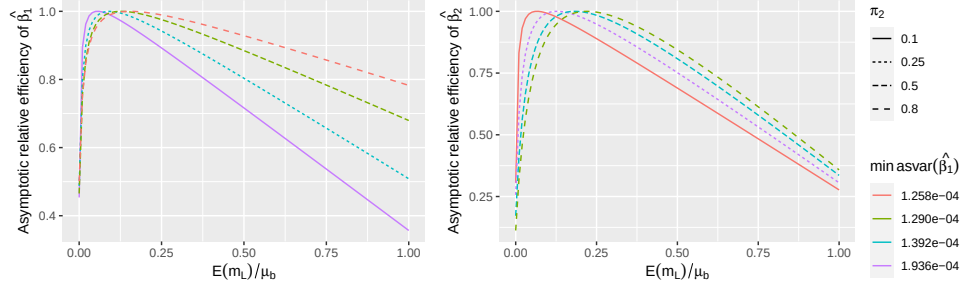
$\mu_b/n = 0.5$ and $\bar{\mu} = 4$:



$\mu_b/n = 0.5$ and $\bar{\mu} = 8$:



$\mu_b/n = 1$ and $\bar{\mu} = 4$:



$\mu_b/n = 1$ and $\bar{\mu} = 8$:

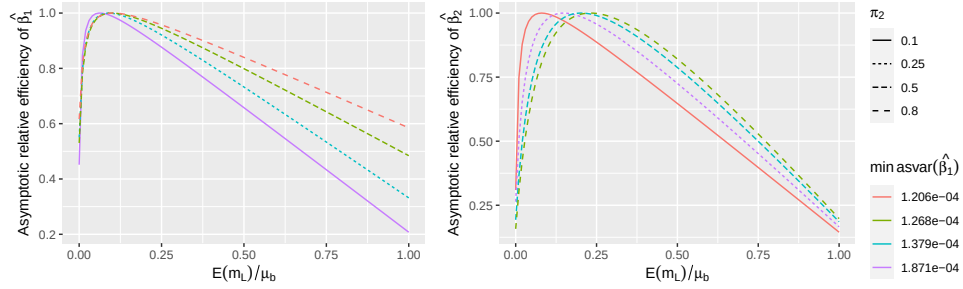


Figure 4.9: Relative asymptotic efficiency of $\hat{\psi}_1$ (left panel) and $\hat{\psi}_2$ as a function of $E\{m_L\}/\mu_b$ for $\pi_2 = 0.1, 0.25, 0.5, 0.8$, $\mu_b/n = 0.5, 1$ and $\bar{\mu} = 4, 8, 12$; the asymptotic variances are obtained from the expected information matrix $\mathcal{I}(\theta)$ with the expectation carried out by Monte-Carlo integration with 500k replications

We focus on a total of $n = 1123$ patients with a prior diagnosis of PsA who had made at least one follow-up visit after recruitment into the study. For each individual, the reported age at PsA diagnosis is taken as the time origin. At each subsequent clinical visit, data including the ESR and CRP results from the blood tests, the date of the blood test, and the number of cumulative damaged joints were recorded. The age of onset of PsA varies from 4 to 49 years (with an average of 39.07 and S.D. of 13.57) across patients in this cohort, and the number of visits made per individual during their follow-up period ranges from 1 to 62 (with an average of $\bar{\mu} = 9.98$ and S.D. of 12.42). Additionally, 274 (24.4 %) patients were observed to develop five or more damaged joints.

Instead of using all $m. = 11204$ measurements of the ESR_CRP marker value, we imagine that the measurements were not available and consider a budgetary constraint. We subsample the times at which samples are available by our proposed cost-effective sampling strategy for a joint analysis of the ESR_CRP marker and the development of arthritis mutilans. We consider a fixed covariate of HLA-B27 and consider budgets μ_b to be $0.5n$, n or $2n$. With a pre-specified value of $E\{m_L\}/\mu_b$ (i.e., the relative proportion of the marker measurements in the longitudinal subsample to the total budget), π_L and π_S can be determined via (4.21) and (4.22). We then select a subsample of marker measurements accordingly for the joint analysis. Our goal is to empirically investigate the variance of the coefficients of both ESR_CRP and HLA-B27 covariates under different specifications of $E\{m_L\}/\mu_b$. When $\mu_b = 0.5n$ or n , we choose $E\{m_L\}/\mu_b = 0.2, 0.4, 0.6$, or 0.8 ; when $\mu_b = 2n$, we choose $E\{m_L\}/\mu_b = 0.6, 0.7, 0.8$, or 0.9 , since we have to assign at least half of the budget to the longitudinal subsample (i.e., $E\{m_L\}/\mu_b > 0.5$) to ensure $\pi_S < 1$. Time-homogeneous transition intensities are fitted for the joint multistate model, and the results of the estimates of the regression coefficients under various specifications of μ_b and $E\{m_L\}/\mu_b$ are summarized in Table 4.1. For $\mu_b = 0.5n$ or n , a minimum variance of the effect of the time-dependent marker ESR_CRP on failure is achieved when we assign 40% of the marker budget to the longitudinal subsample, while a minimum variance of the effect of the fixed covariate HLA-B27 is achieved at 60%. For $\mu_b = 2n$, the minimum variance of both the marker effect and the fixed covariate effect is achieved when $E\{m_L\}/\mu_b = 60\%$. Additionally, the estimated variance of both effects decreases as we have more budget (data) for the time-dependent ESR_CRP value. The point estimates of the ESR_CRP effect are comparable when $E\{m_L\}/\mu_b < 0.6$ but not when $E\{m_L\}/\mu_b > 0.6$. This may be due to the small censoring rate of the original data, meaning that recruiting fewer individuals (and having more longitudinal marker measurements) results in less data on observed failures.

Table 4.1: Joint analyses of the ESR_CRP and failure processes with a fixed covariate of HLA-B27 using subsamples selected under different specifications of μ_b and $E\{m_L\}/\mu_b$; time-homogeneous baseline transition intensities are fitted; variance estimates are obtained through observed information matrix averaged over 500 repetitions

μ_b/n	$E\{m_L\}/\mu_b$	Covariate			
		ESR_CRP (β_1)		HLA-B27 (β_2)	
		EST	VAR	EST	VAR
0.5	0.2	1.639	2.848	0.175	0.590
	0.4	1.464	2.449	0.293	0.588
	0.6	1.361	2.608	0.225	0.532
	0.8	0.941	2.634	0.291	0.574
1	0.2	1.660	1.343	0.231	0.417
	0.4	1.522	1.288	0.270	0.339
	0.6	1.355	1.305	0.285	0.312
	0.8	0.917	1.324	0.258	0.351
2	0.6	1.229	0.705	0.283	0.200
	0.7	1.134	0.732	0.270	0.211
	0.8	0.974	0.974	0.268	0.231
	0.9	0.798	0.791	0.263	0.273

4.6 Discussion

We considered a common setting where biomarkers that vary over time are measured at periodic clinical visits. By discretizing a continuous biomarker into two levels, we formulated a joint multistate model for the marker-failure process; moreover, the multistate model was further expanded to accommodate dependent observation processes, following [Cook et al. \(2022\)](#). Markov transition intensities with proportional hazard models were adopted to govern the disease dynamics. With the marker value measured at each clinical visit, a joint analysis can be conducted for the marker-failure process. However, in practice, instead of having the marker value measured at each visit, it's common to use the marker value assessed at one single visit for statistical inference due to cost and time constraints. Four common cost-effective strategies where the marker value is treated as fixed from either time 0 or the time of measurement were described and the asymptotic bias of estimators arising from fitting misspecified Cox models were investigated. We found that there is generally an attenuation of the estimated regression coefficients, particularly for the marker effect, even when the visit process is completely independent. We then proposed an alternative cost-effective selection scheme of individuals that enables efficient and consistent estimation of parameters and formulated the observed data likelihood under the assumption of independent observation processes. The Fisher information was presented for an optimal design and the asymptotic relative efficiency of estimators obtained from the proposed strategy was explored.

As we only have focused on simple random sampling for individual selection, a natural next step is to consider other sampling mechanisms such as outcome-dependent sampling - we may select more individuals with observed failures. Additionally, we could choose specific visits for marker measurements of individuals, either dependent or independent of the failure status - the visits chosen could vary across individuals. Robustness of the joint model could be improved by adopting piecewise constant baseline transition intensities. Random effects could also be incorporated to account for unexplained heterogeneity in the individual level. More complex multistate disease process could be considered when there is more than one clinical event of interest. Moreover, higher dimensional marker process could also be considered via the use of more cut-points or through the consideration of more time-dependent markers.

Chapter 5

Discussion and Future Work

5.1 Discussion

This thesis has presented new statistical methods addressing challenging problems in the joint analysis of disease processes involving incomplete data due to intermittent observation. In Chapters 2 and 3, we concentrated on joint modeling of a recurrent event process and a terminal event process subject to right-censoring as well as intermittent observation. In Chapter 4, we focused on the joint analysis of a time-dependent marker process and a failure process, where the marker is only assessed during intermittent clinical visits.

The project discussed in Chapter 2 was motivated by a study of the efficacy of pamidronate in preventing skeletal complications in patients with bone metastases from advanced breast cancer. To accommodate the fact that death precludes further skeletal complications, we formulated a new model for a joint analysis of the recurrent skeletal complications and death. We specified a Cox model for the terminal event, and a multiplicative rate function for the recurrent events conditional on a latent random effect. These two processes are linked via a copula function, which defines a joint model for the random effect and the terminal event. Plots of conditional recurrent and terminal event rates given the full history have been investigated to gain insights about the nature of the bivariate process under our proposed model formulation with a Gaussian copula and a Gamma distributed random effect. When these two processes are subject to right-censoring, we have developed estimation and inference methods including a computationally more efficient two-stage procedure that enables semiparametric specifications, both with and without the use of an EM algorithm. Simulation studies demonstrate that the proposed methods provide valid inference in finite samples, and the two-stage procedure is highly efficient. Important in-

sights were gained about the efficacy of pamidronate in reducing skeletal complications in breast cancer patients with bone metastases. Compared to joint modelling of the recurrent and terminal events based on shared or correlated random effects, our model preserves the marginal proportional hazard property for the terminal event process. Additionally, our approach distinguishes itself by addressing unobservable heterogeneity in recurrent events and capturing the interdependence between the two processes through distinct structures. Notably, it remains valid in situations where there is no association between the two event processes. While alternative frameworks for joint analysis via marginal approaches are robust, they do not exploit the history of the joint processes which can lead to more accurate and precise predictions of features of the recurrent event process.

In Chapter 2, we focused on the settings where recurrent events are subject to right-censoring so that the exact times of event occurrence during the follow-up period can be observed. However in many clinical trials, the occurrences of events of interest are not immediately evident and can only be determined during periodic clinical visits. As described in the motivating study of patients with PsA, the number of joint damages for participants is assessed through radiographic examinations during their periodic clinical visits. This motivated the project discussed in Chapter 3 where we considered the joint analysis of the terminal event and interval-censored recurrent events. The extension of the coupled-based joint model was achieved by specifying a piecewise constant baseline rate function for the recurrent events. These piecewise constant models offer the advantages of parametric models while providing some robustness against misspecification of the parametric form of rate functions. The marginal interpretability of the regression coefficients for the terminal event is retained, and semiparametric analyses are feasible via a two-stage estimation procedure. The finite sample performance of estimators obtained through proposed methods demonstrated the validity of our model framework. Applying this method to the PsA cohort data revealed a significant positive association between the HLA-B27 marker and joint damages in patients with PsA.

In Chapter 4, we studied the association between a time-dependent biomarker and a failure time of interest, focusing on the setting where biomarkers are measured at intermittent clinical visits. We introduced a multistate model for the joint marker-failure process by discretizing the continuous biomarker into two levels. One appealing feature of this model formulation is that this multistate model can be further extended to account for dependent censoring and visit processes (Cook et al., 2022). We also considered budgetary constraints on collecting marker values at clinical visits and investigated asymptotic biases of several common but naive cost-effective strategies for statistical inference arising from misspecified Cox models. Our findings indicated that there is generally an attenuation of the estimated regression coefficients, particularly for the marker effect, even when the

visit process is completely independent of the disease process. To address this, an alternative cost-effective strategy has been proposed to mitigate bias under budgetary constraints which involves efficient selections of individuals. The asymptotic relative efficiency of estimators obtained from this proposed strategy under different designs has been explored and an optimal one can thus be determined.

5.2 Future Work

5.2.1 Future Research on Chapters 2 and 3

In Chapters 2 and 3, we accommodated the dependence structure between the recurrent and terminal events by linking the random effect and the terminal event using a pre-specified copula function. An important next step is to investigate the sensitivity of our model to potential misspecification of the copula function. We note the choice of copula functions can be quite flexible - copula models indexed by more than one parameter of association could be used and copula functions from any well-studied family could be used. More data may be required for estimability of parameters when the copula function in use is of more than one parameter. The random effect is not restricted to the Gamma distribution, a log-normal distribution may be used for example.

This joint model can be readily extended to scenarios when we are interested in multiple types of recurrent events and a terminal event process by increasing the dimension of the copula function. Specifically, if we have K types of recurrent events of interest, we can introduce a latent random effect for each type of recurrent event and specify an event rate function modulated by this random effect. The dependence structure can then be constructed by a $(K + 1)$ -dimensional copula function linking the random effects and the terminal event. Various dependence structures can be considered, such as specifying pairwise associations. Gaussian copulas, vine copulas, or other hierarchical copulas might be particularly appealing in this context due to their flexibility and ability to model complex dependencies. Additionally, one may also be interested in sample size estimation when designing clinical trials with these endpoints in mind. Developing a joint test of the hypothesis concerning the effects of covariates on both terminal and recurrent events might be of interest as well. This could provide a more comprehensive assessment of the covariate effects and improve the power of statistical tests in the context of complex event processes.

We focused on the situation where the right-censoring is non-informative about the parameters of interest and independent of the disease process. However, this assumption

might be violated in settings where individuals decide to drop out of the cohort simply because they feel better. Such event-dependent censoring or loss to follow-up in studies of failure time processes has received considerable attention (Fisher and Kanarek, 1974; Slud and Rubinstein, 1983; Scharfstein and Robins, 2002; Siannis, 2011). Models in which the censoring shares a common random effect with the event process have been suggested for settings involving event-dependent censoring or inspection (Cook and Lawless, 2007). We can follow the idea and introduce a shared random effect between the terminal event and the censoring process, or we may introduce another latent random effect that modulates the censoring process and link this random effect with other events of interest through a copula function. In addition, interval-censored data for the terminal event might be considered as well. When disease onset times are prior to selection to a cohort, for example, in the PsA cohort individuals are required to have a previous diagnosis of PsA and be alive to be recruited to the study, biased selection arises. Such outcome dependent selections might be incorporated into the likelihoods to mitigate the bias in parameter estimation.

5.2.2 Future Research on Chapter 4

In Chapter 4, budgetary constraints are considered when collecting marker values measured at clinical visit for statistical inference on the effect of the time-dependent marker on the failure of interest. To address these constraints, we proposed a cost-effective selection strategy where the final dataset for inference consists of a subsample of individuals with marker values measured at each visit, and another subsample of individuals with marker values measured only once at the first visit. We focused on simple random sampling for individual selection into the subsamples and derived an optimal design for specifying the selection probabilities. A natural alternation of the selection strategy is to consider outcome-dependent sampling for individual selection to the two subsamples. Specifically, we might intentionally select all individuals with an observed failure and collect their marker values measured at all clinical visits. Besides selecting individuals with either a single measurement or all measurements of the marker value, we could also consider selecting specific visits or time points for each individual, and the visits chosen could vary across individuals. In general, if we let $R_{ij} = 1$, $j = 1, \dots, m_i$, indicate the marker value measured at the j th visit for individual i is collected for analysis, the budgetary constraint becomes $m_b = \sum_{i=1}^n \sum_{j=1}^{m_i} R_{ij}$. Various models can be specified for the joint distribution of the selection indicators R_{ij} , allowing for the formulation of a joint likelihood of the selection indicators and the marker-failure process. Fisher information can then be derived for determining an optimal design of the parameters related to the selection distributions.

The robustness of the multistate model for the joint marker-failure process can be

improved by adopting piecewise constant baseline transition intensities. Random effects could be incorporated into the transition intensities to account for unexplained heterogeneity at the individual level. More complex multistate disease process can be considered when there is more than one clinical event of interest. Additionally, the multistate model can be extended to accommodate a higher-dimensional marker process via the use of more cut-points for the marker value or through the consideration of multiple time-dependent markers. Mixture multistate models might also be of interest when there is a portion of the population not susceptible to the event of interest. These models can separate the susceptible and non-susceptible subpopulations, providing more accurate estimates of the effects of covariates on the failure risk. This approach could be particularly useful in settings where heterogeneity in the population's response to covariates or failure susceptibility is expected.

References

- Aalen, O., Borgan, O., and Gjessing, H. (2008). *Survival and Event History Analysis: A Process Point of View*. Springer Science & Business Media.
- Andersen, P. and Gill, R. (1982). Cox’s regression model for counting processes: A large sample study. *The Annals of Statistics*, 10(4):1100–1120.
- Andersen, P. K. (1988). Multistate models in survival analysis: a study of nephropathy and mortality in diabetes. *Statistics in Medicine*, 7(6):661–670.
- Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (2012). *Statistical Models Based on Counting Processes*. Springer Science & Business Media.
- Andersen, P. K. and Liestøl, K. (2003). Attenuation caused by infrequently updated covariates in survival analysis. *Biostatistics*, 4(4):633–649.
- Binquet, C., Chene, G., Jacqmin-Gadda, H., Journot, V., Saves, M., Lacoste, D., and Dabis, F. (2001). Modeling changes in cd4-positive t-lymphocyte counts after the start of highly active antiretroviral therapy and the relation with risk of opportunistic infections the aquitaine cohort, 1996–1997. *American Journal of Epidemiology*, 153(4):386–393.
- Boos, D. and Stefanski, L. (2013). *Essential Statistical Inference: Theory and Methods*. Springer.
- Breslow, N. and Day, N. (1980). *Statistical Methods in Cancer Research - Volume 1: The Analysis of Case-Control Studies*, volume 32. IARC scientific publications.
- Byar, D. (1980). The veterans administration study of chemoprophylaxis for recurrent stage i bladder tumours: comparisons of placebo, pyridoxine and topical thiotepa. In *Bladder Tumors and Other Topics in Urological Oncology*, pages 363–370. Springer.

- Chen, B., Cook, R., Lawless, J., and Zhan, M. (2005). Statistical methods for multivariate interval-censored recurrent events. *Statistics in Medicine*, 24(5):671–691.
- Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65(1):141–151.
- Coleman, R. and Purohit, O. (1993). Osteoclast inhibition for the treatment of bone metastases. *Cancer Treatment Reviews*, 19(1):79–103.
- Coleman, R. and Rubens, R. (1985). Bone metastases and breast cancer. *Cancer Treatment Reviews*, 12(4):251–270.
- Cook, R. and Lawless, J. (1997). Marginal analysis of recurrent events and a terminating event. *Statistics in Medicine*, 16(8):911–924.
- Cook, R. and Lawless, J. (2007). *The Statistical Analysis of Recurrent Events*. Springer.
- Cook, R. J. and Lawless, J. F. (2018). *Multistate Models for the Analysis of Life History Data*. Chapman and Hall/CRC.
- Cook, R. J. and Lawless, J. F. (2021). Independence conditions and the analysis of life history studies with intermittent observation. *Biostatistics*, 22(3):455–481.
- Cook, R. J., Lawless, J. F., and Lee, K.-A. (2010). A copula-based mixed poisson model for bivariate recurrent events under event-dependent censoring. *Statistics in Medicine*, 29(6):694–707.
- Cook, R. J., Lawless, J. F., and Xie, B. (2022). Marker-dependent observation and carry-forward of internal covariates in cox regression. *Lifetime Data Analysis*, 28(4):560–584.
- Cox, D. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202.
- Cox, D. R. (2017). *The Theory of Stochastic Processes*. Routledge.
- de Bruijne, M. H., le Cessie, S., Kluin-Nelemans, H. C., and van Houwelingen, H. C. (2001). On the use of cox regression in the presence of an irregularly observed time-dependent covariate. *Statistics in Medicine*, 20(24):3817–3829.
- Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.

- Farewell, D., Huang, C., and Didelez, V. (2017). Ignorability for general longitudinal data. *Biometrika*, 104(2):317–326.
- Fisher, L. and Kanarek, P. (1974). Presenting censored survival data when censoring and survival times may not be independent. *Reliability and Biometry*, pages 303–326.
- Fleisch, H. (1989). Bisphosphonates: a new class of drugs in diseases of bone and calcium metabolism. *Bisphosphonates and Tumor Osteolysis*, pages 1–28.
- Frank, M. J. (1979). On the simultaneous associativity of $f(x, y)$ and $x + y - f(x, y)$. *Aequationes Mathematicae*, 19:194–226.
- Fuchs, C., Giovannucci, E., Colditz, G., Hunter, D., Speizer, F., and Willett, W. (1994). A prospective study of family history and the risk of colorectal cancer. *New England Journal of Medicine*, 331(25):1669–1674.
- Gail, M., Santner, T., and Brown, C. (1980). An analysis of comparative carcinogenesis experiments based on multiple times to tumor. *Biometrics*, 36(2):255–266.
- Ghosh, D. and Lin, D. (2000). Nonparametric analysis of recurrent events and death. *Biometrics*, 56(2):554–562.
- Ghosh, D. and Lin, D. (2002). Marginal regression models for recurrent and terminal events. *Statistica Sinica*, 12(3):663–688.
- Gladman, D., Antoni, C., Mease, P., Clegg, D., and Nash, P. (2005). Psoriatic arthritis: epidemiology, clinical features, course, and outcome. *Annals of the Rheumatic Diseases*, 64(suppl 2):ii14–ii17.
- Gladman, D. D. and Chandran, V. (2011). Observational cohort studies: lessons learnt from the university of toronto psoriatic arthritis program. *Rheumatology*, 50(1):25–31.
- Gladman, D. D. and Farewell, V. T. (1995). The role of hla antigens as indicators of disease progression in psoriatic arthritis. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology*, 38(6):845–850.
- Gladman, D. D., Shuckett, R., Russell, M., Thorne, J., and Schachter, R. (1987). Psoriatic arthritis (psa)-an analysis of 220 patients. *QJM: An International Journal of Medicine*, 62(2):127–141.

- Grossman, R., Mukherjee, J., Vaughan, D., Eastwood, C., Cook, R., LaForge, J., and Lampron, N. (1998). A 1-year community-based health economic study of ciprofloxacin vs usual antibiotic treatment in acute exacerbations of chronic bronchitis: the canadian ciprofloxacin health economic study group. *Chest*, 113(1):131–141.
- Hobson, R., Weiss, D., Fields, W., Goldstone, J., Moore, W., Towne, J., Wright, C., and Group, V. A. C. S. (1993). Efficacy of carotid endarterectomy for asymptomatic carotid stenosis. *New England Journal of Medicine*, 328(4):221–227.
- Hortobagyi, G., Theriault, R., Porter, L., Blayney, D., Lipton, A., Sinoff, C., Wheeler, H., Simeone, J., Seaman, J., Knight, R., et al. (1996). Efficacy of pamidronate in reducing skeletal complications in patients with breast cancer and lytic bone metastases. *New England Journal of Medicine*, 335(24):1785–1792.
- Huang, C. and Wang, M. (2004). Joint modeling and estimation for recurrent event processes and failure time data. *Journal of the American Statistical Association*, 99(468):1153–1165.
- Huang, C.-Y., Wang, M.-C., and Zhang, Y. (2006). Analysing panel count data with informative observation times. *Biometrika*, 93(4):763–775.
- Jiang, S., Cook, R. J., and Zeng, L. (2020). Mitigating bias from intermittent measurement of time-dependent covariates in failure time analysis. *Statistics in Medicine*, 39(13):1833–1845.
- Joe, H. (1997). *Multivariate Models and Multivariate Dependence Concepts*. CRC Press.
- Johnston, A. (1970). Pathology of metastatic tumors in bone. *Clinical Orthopaedics and Related Research (1976-2007)*, 73:8–32.
- Kalbfleisch, J. and Lawless, J. F. (1985). The analysis of panel data under a markov assumption. *Journal of the American Statistical Association*, 80(392):863–871.
- Kalbfleisch, J., Schaubel, D., Ye, Y., and Gong, Q. (2013). An estimating function approach to the analysis of recurrent and terminal events. *Biometrics*, 69(2):366–374.
- Kalbfleisch, J. D. and Prentice, R. L. (2011). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons.
- Kellihan, M. and Mangino, P. (1992). Pamidronate. *Annals of Pharmacotherapy*, 26(10):1262–1269.

- Kendall, M. (1938). A new measure of rank correlation. *Biometrika*, 30(1/2):81–93.
- Klein, J. P., Van Houwelingen, H. C., Ibrahim, J. G., and Scheike, T. H. (2014). *Handbook of Survival Analysis*. CRC Press Boca Raton, FL:.
- Lawless, J. and Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics*, 37(2):158–168.
- Lawless, J. and Zhan, M. (1998). Analysis of interval-grouped recurrent-event data using piecewise constant rate functions. *Canadian Journal of Statistics*, 26(4):549–565.
- Lawless, J. F. (1987). Negative binomial and mixed poisson regression. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, pages 209–225.
- Lawless, J. F. (2011). *Statistical models and methods for lifetime data*. John Wiley & Sons.
- Lin, D. Y. and Wei, L.-J. (1989). The robust inference for the cox proportional hazards model. *Journal of the American statistical Association*, 84(408):1074–1078.
- Lin, D. Y., Wei, L.-J., Yang, I., and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4):711–730.
- Liu, L., Wolfe, R., and Huang, X. (2004). Shared frailty models for recurrent events and a terminal event. *Biometrics*, 60(3):747–756.
- Louis, T. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(2):226–233.
- Murphy, S. and Van Der Vaart, A. (2000). On profile likelihood. *Journal of the American Statistical Association*, 95(450):449–465.
- Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer.
- Papageorgiou, G., Mauff, K., Tomer, A., and Rizopoulos, D. (2019). An overview of joint modeling of time-to-event and longitudinal outcomes. *Annual review of statistics and its application*, 6(1):223–240.
- Prentice, R., Williams, B., and Peterson, A. (1981). On the regression analysis of multivariate failure time data. *Biometrika*, 68(2):373–379.

- Raboud, J., Reid, N., Coates, R., and Farewell, V. (1993). Estimating risks of progressing to aids when covariates are measured with error. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 156(3):393–406.
- Scharfstein, D. O. and Robins, J. M. (2002). Estimation of the failure time distribution in the presence of informative censoring. *Biometrika*, 89(3):617–634.
- Siannis, F. (2011). Sensitivity analysis for multiple right censoring processes: investigating mortality in psoriatic arthritis. *Statistics in Medicine*, 30(4):356–367.
- Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. In *Annales de l'ISUP*, volume 8, pages 229–231.
- Slud, E. V. and Rubinstein, L. V. (1983). Dependent competing risks and summary survival curves. *Biometrika*, 70(3):643–649.
- Struthers, C. A. and Kalbfleisch, J. D. (1986). Misspecified proportional hazard models. *Biometrika*, 73(2):363–369.
- Sun, J. and Kalbfleisch, J. (1995). Estimation of the mean function of point processes based on panel count data. *Statistica Sinica*, pages 279–289.
- Sun, J. and Kalbfleisch, J. D. (1993). The analysis of current status data on point processes. *Journal of the American Statistical Association*, 88(424):1449–1454.
- Sun, J. and Matthews, D. E. (1997). A random-effect regression model for medical follow-up studies. *Canadian Journal of Statistics*, 25(1):101–111.
- Sun, J., Tong, X., and He, X. (2007). Regression analysis of panel count data with dependent observation times. *Biometrics*, 63(4):1053–1059.
- Sun, J. and Wei, L. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 62(2):293–302.
- Thall, P. F. and Lachin, J. M. (1988). Analysis of recurrent events: Nonparametric methods for random-interval count data. *Journal of the American Statistical Association*, 83(402):339–347.
- Tsiatis, A. A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, pages 809–834.

- Tsiatis, A. A., Degruittola, V., and Wulfsohn, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error. applications to survival and cd4 counts in patients with aids. *Journal of the American Statistical Association*, 90(429):27–37.
- Wang, M., Qin, J., and Chiang, C. (2001). Analyzing recurrent event data with informative censoring. *Journal of the American Statistical Association*, 96(455):1057–1065.
- Wei, L., Lin, D., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association*, 84(408):1065–1073.
- Wellner, J. A. and Zhang, Y. (2000). Two estimators of the mean of a counting process with panel count data. *The Annals of Statistics*, 28(3):779–814.
- Wellner, J. A., Zhang, Y., and Liu, R. (2004). A semiparametric regression model for panel count data: when do pseudo-likelihood estimators become badly inefficient? In *Proceedings of the Second Seattle Symposium in Biostatistics: Analysis of Correlated Data*, pages 143–174. Springer.
- Wu, L., Liu, W., Yi, G. Y., and Huang, Y. (2012). Analysis of longitudinal and survival data: joint modeling, inference methods, and issues. *Journal of Probability and Statistics*, 2012(1):640153.
- Wulfsohn, M. S. and Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, pages 330–339.
- Xu, J. and Zeger, S. L. (2001). Joint analysis of longitudinal data comprising repeated measures and times to events. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(3):375–387.
- Ye, Y., Kalbfleisch, J., and Schaubel, D. (2007). Semiparametric analysis of correlated recurrent and terminal events. *Biometrics*, 63(1):78–87.
- Zhang, Y. (2002). A semiparametric pseudolikelihood estimation method for panel count data. *Biometrika*, 89(1):39–48.
- Zhang, Y. and Jamshidian, M. (2003). The gamma-frailty poisson model for the nonparametric estimation of panel count data. *Biometrics*, 59(4):1099–1106.
- Zhao, X. and Tong, X. (2011). Semiparametric regression analysis of panel count data with informative observation times. *Computational Statistics & Data Analysis*, 55(1):291–300.

APPENDICES

Appendix A

Appendix to Chapter 2

A.1 Joint Recurrent-Terminal Event Model via a Gaussian Copula

A K -dimensional function of $v = (v_1, \dots, v_K)'$ from the Gaussian copula family has the form

$$\mathcal{C}(v; \Sigma) = \Phi_K \left(\Phi^{-1}(v_1), \dots, \Phi^{-1}(v_K); \Sigma \right)$$

where $\Phi(\cdot)$ denotes the c.d.f of a standard normal distribution and $\Phi_K(\cdot; \Sigma)$ denotes the multivariate c.d.f. of a K -variate normal distribution with mean vector zero and a $K \times K$ covariance matrix Σ . Note the Gaussian copula is determined entirely by its covariance matrix.

Using a bivariate Gaussian copula, a joint model for the random effect U_{i1} and the terminal event time T_i conditioning on covariates X_i can be written as

$$\begin{aligned} P(U_{i1} \leq u_1, T_i > t \mid X_i; \theta_2, \phi_1, \psi) &= \mathcal{C}(G_1(u_1; \phi_1), \mathcal{F}(t \mid X_i; \theta_2); \psi) \\ &= \Phi_2 \left(\Phi^{-1}(G_1(u_1; \phi_1)), \Phi^{-1}(\mathcal{F}(t \mid X_i; \theta_2)); \psi \right) \end{aligned} \quad (\text{A.1})$$

where $\Phi_2(\cdot, \cdot; \psi)$ denotes the c.d.f. of a bivariate normal distribution with a covariance matrix

$$\Sigma_{2 \times 2} = \begin{pmatrix} 1 & \psi \\ \psi & 1 \end{pmatrix}$$

and the correlation ψ . Note that under such a specification, we have

$$\text{Kendall's } \tau = \frac{2}{\pi} \arcsin(\psi),$$

and $\psi = \sin\left(\frac{\pi}{2} \times \text{Kendall's } \tau\right)$.

The first derivative of (A.1) w.r.t. u_1 gives

$$\begin{aligned}
P(U_{i1} = u_1, T_i > t \mid X_i; \theta_2, \phi_1, \psi) &= \frac{\partial}{\partial u_1} \mathcal{C}(G_1(u_1; \phi_1), \mathcal{F}(t \mid X_i; \theta_2); \psi) \\
&= \frac{\partial}{\partial G_1(u_1; \phi_1)} \mathcal{C}(G_1(u_1; \phi_1), \mathcal{F}(t \mid X_i; \theta_2); \psi) \frac{\partial G_1(u_1; \phi_1)}{\partial u_1} \\
&= \Phi\left(\frac{q_1(t; \theta_2) - \psi q_2(\phi_1)}{\sqrt{1 - \psi^2}}\right) g_1(u_1; \phi_1)
\end{aligned} \tag{A.2}$$

where

$$q_1(t; \theta_2) = \Phi^{-1}(\mathcal{F}(t \mid X_i; \theta_2)) \quad \text{and} \quad q_2(\phi_1) = \Phi^{-1}(G_1(u_1; \phi_1)). \tag{A.3}$$

Taking the negative of the derivative of (A.2) w.r.t to t , the joint density of (U_{i1}, T_i) is then given by

$$\begin{aligned}
&P(U_{i1} = u_1, T_i = t \mid X_i; \theta_2, \phi_1, \psi) \\
&= - \frac{\partial}{\partial u_1} \frac{\partial}{\partial t} \mathcal{C}(G_1(u_1; \phi_1), \mathcal{F}(t \mid X_i; \theta_2); \psi) \\
&= - \frac{\partial^2 \mathcal{C}(G_1(u_1; \phi_1), \mathcal{F}(t \mid X_i; \theta_2); \psi)}{\partial G_1(u_1; \phi_1) \partial \mathcal{F}(t \mid X_i; \theta_2)} \frac{dG_1(u_1; \phi_1)}{du_1} \frac{\partial \mathcal{F}(t \mid X_i; \theta_2)}{\partial F(t \mid X_i; \theta_2)} \frac{dF(t \mid X_i; \theta_2)}{dt} \\
&= \frac{1}{\sqrt{1 - \psi^2}} \exp\left(-\frac{\psi^2 [q_1(t; \theta_2)^2 + q_2(\phi_1)^2] - 2\psi q_1(t; \theta_2)q_2(\phi_1)}{2(1 - \psi^2)}\right) g_1(u_1; \phi_1) f(t \mid X_i; \theta_2).
\end{aligned} \tag{A.4}$$

A.2 Calculations of Some Expectation Terms under a Gaussian Copula

The EM algorithm requires calculation of some expectation terms in (2.10) and we give details on their calculation here. We take $E(U_{i1} \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)})$ as an example. Specifically,

$$\begin{aligned} E(U_{i1} \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)}) &= \frac{\int_0^\infty u_1 \mathcal{L}_i(\theta^{(r-1)}) du_1}{\int_0^\infty \mathcal{L}_i(\theta^{(r-1)}) du_1} \\ &= \left\{ \frac{\int_0^\infty u_1 M_i(u_1; \theta_1^{(r-1)}) g_1(u_1 \mid T_i = w_i, X_i; \theta_2^{(r-1)}, \phi_1^{(r-1)}, \psi^{(r-1)}) du_1}{Q_{i1}(\theta^{(r-1)})} \right\}^{\delta_i} \\ &\quad \times \left\{ \frac{\int_0^\infty u_1 M_i(u_1; \theta_1^{(r-1)}) g_1(u_1 \mid T_i > w_i, X_i; \theta_2^{(r-1)}, \phi_1^{(r-1)}, \psi^{(r-1)}) du_1}{Q_{i2}(\theta^{(r-1)})} \right\}^{1-\delta_i}, \end{aligned} \quad (\text{A.5})$$

where

$$M_i(u_1; \theta_1) = u_1^{n_i} \exp(-u_1 R(w_i \mid X_i; \theta_1)),$$

$$Q_{i1}(\theta) = \int_0^\infty M_i(u_1; \theta_1) g_1(u_1 \mid T_i = w_i, X_i; \theta_2, \phi_1, \psi) du_1,$$

and

$$Q_{i2}(\theta) = \int_0^\infty M_i(u_1; \theta_1) g_1(u_1 \mid T_i > w_i, X_i; \theta_2, \phi_1, \psi) du_1.$$

If we specify a Gaussian copula for the joint model as in Appendix A.1, and using the notation $q_1(t; \theta_2)$, $q_2(\phi_1)$ from (A.3) with

$$q_1(\theta_2) := q_1(w_i; \theta_2), \quad (\text{A.6})$$

we have

$$\begin{aligned} Q_{i1}(\theta) &= \int_0^\infty M_i(u_1; \theta_1) \frac{g_1(u_1; \phi_1)}{\sqrt{1-\psi^2}} \exp\left(-\frac{\psi^2(q_1(\theta_2)^2 + q_2(\phi_1)^2) - 2\psi q_1(\theta_2)q_2(\phi_1)}{2(1-\psi^2)}\right) du_1 \\ &= \int_{-\infty}^\infty M_i(u_1; \theta_1) \frac{\phi(q_2(\phi_1))}{\sqrt{1-\psi^2}} \exp\left(-\frac{\psi^2(q_1(\theta_2)^2 + q_2(\phi_1)^2) - 2\psi q_1(\theta_2)q_2(\phi_1)}{2(1-\psi^2)}\right) dq_2(\phi_1) \\ &= \int_{-\infty}^\infty M_i(G_1^{-1}(\Phi(q_2(\phi_1))); \phi_1; \theta_1) \frac{1}{\sqrt{2\pi(1-\psi^2)}} \exp\left(-\frac{(q_2(\phi_1) - \psi q_1(\theta_2))^2}{2(1-\psi^2)}\right) dq_2(\phi_1) \\ &= E_{V_1} \{M_i(G_1^{-1}(\Phi(V_1)); \phi_1; \theta_1)\} \end{aligned} \quad (\text{A.7})$$

where $V_1 \sim N(\psi q_1(\theta_2), 1 - \psi^2)$, and

$$\begin{aligned}
Q_{i2}(\theta) &= \int_0^\infty M_i(u_1; \theta_1) \Phi \left(\frac{q_1(\theta_2) - \psi q_2(\phi_1)}{\sqrt{1 - \psi^2}} \right) \frac{g_1(u_1; \phi_1)}{\mathcal{F}(w_i | X_i; \theta_2)} du_1 \\
&= \int_{-\infty}^\infty M_i(u_1; \theta_1) \Phi \left(\frac{q_1(\theta_2) - \psi q_2(\phi_1)}{\sqrt{1 - \psi^2}} \right) \frac{\phi(q_2(\phi_1))}{\mathcal{F}(w_i | X_i; \theta_2)} dq_2(\phi_1) \\
&= \frac{1}{\mathcal{F}(w_i | X_i; \theta_2)} E_{V_2} \left\{ M_i(G_1^{-1}(\Phi(V_2); \phi_1); \theta_1) \Phi \left(\frac{q_1(\theta_2) - \psi V_2}{\sqrt{1 - \psi^2}} \right) \right\} \quad (\text{A.8})
\end{aligned}$$

where $V_2 \sim N(0, 1)$. Similarly, we have

$$\begin{aligned}
&\int_0^\infty u_1 M_i(u_1; \theta_1^{(r-1)}) g_1(u_1 | T_i = w_i, X_i; \theta_2^{(r-1)}, \phi_1^{(r-1)}, \psi^{(r-1)}) du_1 \\
&= E_{V_1^{(r-1)}} \left\{ G_1^{-1}(\Phi(V_1^{(r-1)}); \phi_1^{(r-1)}) M_i(G_1^{-1}(\Phi(V_1^{(r-1)}); \phi_1^{(r-1)}); \theta_1^{(r-1)}) \right\} \quad (\text{A.9})
\end{aligned}$$

where $V_1^{(r-1)} \sim N(\psi^{(r-1)} q_1(\theta_2^{(r-1)}), 1 - (\psi^{(r-1)})^2)$, and

$$\begin{aligned}
&\int_0^\infty u_1 M_i(u_1; \theta_1^{(r-1)}) g_1(u_1 | T_i > w_i, X_i; \theta_2^{(r-1)}, \phi_1^{(r-1)}, \psi^{(r-1)}) du_1 \\
&= \frac{E_{V_2} \left\{ G_1^{-1}(\Phi(V_2); \phi_1^{(r-1)}) M_i(G_1^{-1}(\Phi(V_2); \phi_1^{(r-1)}); \theta_1^{(r-1)}) \Phi \left(\frac{q_1(\theta_2^{(r-1)}) - \psi^{(r-1)} V_2}{\sqrt{1 - (\psi^{(r-1)})^2}} \right) \right\}}{\mathcal{F}(w_i | X_i; \theta_2^{(r-1)})}. \quad (\text{A.10})
\end{aligned}$$

(A.5) can be calculated using (A.7), (A.8), (A.9) and (A.10).

In general, for some function of u_1 , denoted by $m_i(u_1; \theta)$, we have

$$\begin{aligned}
&\int_0^\infty m_i(u_1; \theta) M_i(u_1; \theta_1^{(r-1)}) g_1(u_1 | T_i = w_i, X_i; \theta_2^{(r-1)}, \phi_1^{(r-1)}, \psi^{(r-1)}) du_1 \\
&= E_{V_1^{(r-1)}} \left\{ m_i(G_1^{-1}(\Phi(V_1^{(r-1)}); \phi_1^{(r-1)}); \theta) M_i(G_1^{-1}(\Phi(V_1^{(r-1)}); \phi_1^{(r-1)}); \theta_1^{(r-1)}) \right\}
\end{aligned}$$

and

$$\begin{aligned}
&\int_0^\infty m_i(u_1; \theta) M_i(u_1; \theta_1^{(r-1)}) g_1(u_1 | T_i > w_i, X_i; \theta_2^{(r-1)}, \phi_1^{(r-1)}, \psi^{(r-1)}) du_1 \\
&= \frac{E_{V_2} \left\{ m_i(G_1^{-1}(\Phi(V_2); \phi_1^{(r-1)}); \theta) M_i(G_1^{-1}(\Phi(V_2); \phi_1^{(r-1)}); \theta_1^{(r-1)}) \Phi \left(\frac{q_1(\theta_2^{(r-1)}) - \psi^{(r-1)} V_2}{\sqrt{1 - (\psi^{(r-1)})^2}} \right) \right\}}{\mathcal{F}(w_i | X_i; \theta_2^{(r-1)})}.
\end{aligned}$$

Hence, we have

$$E(m_i(U_{i1}; \theta) \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)}) = \frac{\int_0^\infty m_i(u_1; \theta) \mathcal{L}_i(\theta^{(r-1)}) du_1}{\int_0^\infty \mathcal{L}_i(\theta^{(r-1)}) du_1} \quad (\text{A.11})$$

which can be further expressed as

$$\left\{ \frac{E_{V_1^{(r-1)}} \left\{ m_i(G_1^{-1}(\Phi(V_1^{(r-1)}); \phi_1^{(r-1)}); \theta) M_i(G_1^{-1}(\Phi(V_1^{(r-1)}); \phi_1^{(r-1)}); \theta_1^{(r-1)}) \right\}}{Q_{i1}(\theta^{(r-1)})} \right\}^{\delta_i} \\ \times \left\{ \frac{E_{V_2} \left\{ m_i(G_1^{-1}(\Phi(V_2); \phi_1^{(r-1)}); \theta) M_i(G_1^{-1}(\Phi(V_2); \phi_1^{(r-1)}); \theta_1^{(r-1)}) \Phi \left(\frac{q_1(\theta_2^{(r-1)}) - \psi^{(r-1)} V_2}{\sqrt{1 - (\psi^{(r-1)})^2}} \right) \right\}}{\mathcal{F}(w_i \mid X_i; \theta_2^{(r-1)}) Q_{i2}(\theta^{(r-1)})} \right\}^{1-\delta_i}.$$

Other expectation terms in (2.10) can be calculated using (A.11) by specifying $m_i(u_1; \theta) = \log g_1(u_1 \mid X_i, T_i = w_i; \theta_2, \phi_1, \psi)$ and then $m_i(u_1; \theta) = \log g_1(u_1 \mid X_i, T_i > w_i; \theta_2, \phi_1, \psi)$.

A.3 Data Generation Procedure for Simulation Studies

For each simulation, the data generation procedure for the i th individual is described as follows.

1. Generate X_i as Bernoulli with $P(X_i = 1) = 0.5$.
2. Generate $U_{i1} \sim \text{Gamma}(\text{mean} = 1, \text{variance} = \phi_1)$.
3.
 - If Kendall's $\tau = 0$, generate the failure time t_i from an Exponential distribution with

$$T_i \sim \text{Exponential}(\text{rate} = h_0 e^{x_i \beta_2}).$$

- If Kendall's $\tau \neq 0$, generate p_i from $\text{Uniform}(0, 1)$ first, and then solve for t_i such that

$$\begin{aligned} p_i &= P(T_i > t_i \mid U_{i1} = u_{i1}, X_i = x_i; \theta_2, \phi_1, \psi) \\ &= \frac{P(U_{i1} = u_{i1}, T_i > t_i \mid X_i = x_i; \theta_2, \phi_1, \psi)}{g_1(u_{i1}; \phi_1)} \\ &= \Phi \left(\frac{\Phi^{-1}(\mathcal{F}(t_i \mid X_i = x_i; \theta_2)) - \psi \Phi^{-1}(G_1(u_{i1}; \phi_1))}{\sqrt{1 - \psi^2}} \right) \end{aligned}$$

where $\mathcal{F}(t \mid X_i = x_i; \theta_2) = \exp(-h_0 t e^{x_i \beta_2})$. This gives

$$t_i = \frac{-\log \left\{ \Phi \left[\Phi^{-1}(p_i) \sqrt{1 - \psi^2} + \psi \Phi^{-1}(G_1(u_{i1}; \phi_1)) \right] \right\}}{h_0 e^{x_i \beta_2}}.$$

4. Note $C_i = 1$, let $w_i = \min(t_i, 1)$ and $\delta_i = \mathbb{1}(t_i < 1)$.
5. Let the recurrent event times be s_{ij} , $j = 0, 1, \dots$, we generate s_{ij} via

$$s_{ij} = s_{i,j-1} + e_{ij}, \quad j = 0, 1, \dots,$$

where $e_{ij} \sim \text{Exponential}(\text{rate} = u_{i1} \rho_0 e^{x_i \beta_1})$ with $s_{i0} = 0$. We keep all $s_{ij} < w_i$ and $n_i = \sum_{j=1}^{\infty} \mathbb{I}(s_{ij} < w_i)$.

We repeat Steps 1 to 5 for $i = 1, 2, \dots, m$, and a data set is finalized.

A.4 Parametric Multiplicative Models with Piecewise Constant Baseline Functions

Here the details are provided on the notation, complete data likelihoods and steps of the EM algorithm for the setting of piecewise constant baseline functions. For both recurrent and terminal events, we specify multiplicative models with piecewise constant baseline functions. Note the number of pieces of the baseline functions for the recurrent and terminal events are denoted by K_1 and K_2 respectively. Let $b_{10}, b_{11}, \dots, b_{1K_1}$ and $b_{20}, b_{21}, \dots, b_{2K_2}$ be the pre-specified cut-points of the baseline hazard function for the recurrent and terminal events respectively. We let $b_{10} = b_{20} = 0$, $b_{1K_1} = b_{2K_2} = \infty$, $b_{10} < b_{11} < \dots < b_{1K_1}$ and $b_{20} < b_{21} < \dots < b_{2K_2}$.

For the recurrent event process, we let $\delta_{1k}(s) = \mathbb{1}(s \in \mathcal{B}_{1k})$ indicate if time s is in the interval $\mathcal{B}_{1k} = (b_{1,k-1}, b_{1k}]$ and $d_{1k}(s) = \int_0^s \delta_{1k}(t) dt$ be length of the intersection of the interval $(0, s]$ and $\mathcal{B}_{1k}$, $k = 1, \dots, K_1$. So now the recurrent event rate and cumulative rate functions can be written as $\rho_0(s; \alpha_1) = \prod_{k=1}^{K_1} \rho_k^{\delta_{1k}(s)}$, $\rho(s \mid X_i; \theta_1) = \prod_{k=1}^{K_1} \rho_k^{\delta_{1k}(s)} e^{X_i' \beta_1}$, $R_0(s; \alpha_1) = \sum_{k=1}^{K_1} \rho_k d_{1k}(s)$ and $R(s \mid X_i; \theta_1) = \sum_{k=1}^{K_1} \rho_k d_{1k}(s) e^{X_i' \beta_1}$. Note $\theta_1 = (\alpha'_1, \beta'_1)'$ where $\alpha_1 = (\rho_1, \dots, \rho_{K_1})'$.

For the terminal event process, we let $\delta_{2k}(s) = \mathbb{1}(s \in \mathcal{B}_{2k})$ with $\mathcal{B}_{2k} = (b_{2,k-1}, b_{2k}]$ and $d_{2k}(t) = \int_0^t \delta_{2k}(s) ds$, $k = 1, \dots, K_2$, so we have the terminal event rate and cumulative rate

functions written as $h_0(t; \alpha_2) = \prod_{k=1}^{K_2} h_k^{\delta_{2k}(t)}$, $h(t | X_i; \theta_2) = \prod_{k=1}^{K_2} h_k^{\delta_{2k}(t)} e^{X_i' \beta_2}$, $H(t; \alpha_2) = \sum_{k=1}^{K_2} h_k d_{2k}(t)$ and $H(t | X_i; \theta_2) = \sum_{k=1}^{K_2} h_k d_{2k}(t) e^{X_i' \beta_2}$. We also have

$$f(t | X_i; \theta_2) = \prod_{k=1}^{K_2} h_k^{\delta_{2k}(t)} e^{X_i' \beta_2} \exp\left\{-\sum_{k=1}^{K_2} h_k d_{2k}(t) e^{X_i' \beta_2}\right\}$$

and

$$\mathcal{F}(t | X_i; \theta_2) = \exp\left\{-\sum_{k=1}^{K_2} h_k d_{2k}(t) e^{X_i' \beta_2}\right\}.$$

Note $\theta_2 = (\alpha_2', \beta_2')'$ with $\alpha_2 = (h_1, \dots, h_{K_2})'$.

The observed data log-likelihood $\log L_i(\theta)$ for a single individual under such settings now has a specific form:

$$\begin{aligned} & \delta_i \left(\sum_{k=1}^{K_2} \delta_{2k}(w_i) \log h_k + X_i' \beta_2 \right) - \sum_{k=1}^{K_2} h_k d_{2k}(w_i) e^{X_i' \beta_2} + \sum_{j=1}^{n_i} \left(\sum_{k=1}^{K_1} \delta_{1k}(s_{ij}) \log \rho_k + X_i' \beta_1 \right) \\ & + \delta_i \log \left\{ \int_0^\infty u_1^{n_i} \exp \left(-u_1 \sum_{k=1}^{K_1} \rho_k d_{1k}(w_i) e^{X_i' \beta_1} \right) g_1(u_1 | T_i = w_i, X_i; \theta_2, \phi_1, \psi) du_1 \right\} \\ & + (1 - \delta_i) \log \left\{ \int_0^\infty u_1^{n_i} \exp \left(-u_1 \sum_{k=1}^{K_1} \rho_k d_{1k}(w_i) e^{X_i' \beta_1} \right) g_1(u_1 | T_i > w_i, X_i; \theta_2, \phi_1, \psi) du_1 \right\}. \end{aligned}$$

The complete data log-likelihood contribution for one individual $\log \mathcal{L}_i(\theta) = \ell_{i1}(\theta_1) + \ell_{i2}(\theta_2) + \ell_{i3}(\theta_2, \phi_1, \psi)$ is then

$$\begin{aligned} \ell_{i1}(\theta_1) &= \sum_{j=1}^{n_i} \left(\log u_{i1} + \sum_{k=1}^{K_1} \delta_{1k}(s_{ij}) \log \rho_k + X_i' \beta_1 \right) - u_{i1} \sum_{k=1}^{K_1} \rho_k d_{1k}(w_i) e^{X_i' \beta_1}, \\ \ell_{i2}(\theta_2) &= \delta_i \left(\sum_{k=1}^{K_2} \delta_{2k}(w_i) \log h_k + X_i' \beta_2 \right) - \sum_{k=1}^{K_2} h_k d_{2k}(w_i) e^{X_i' \beta_2}, \end{aligned}$$

and

$$\begin{aligned} \ell_{i3}(\theta_2, \phi_1, \psi) &= \delta_i \log g_1(u_{i1} | T_i = w_i, X_i; \theta_2, \phi_1, \psi) \\ &+ (1 - \delta_i) \log g_1(u_{i1} | T_i > w_i, X_i; \theta_2, \phi_1, \psi). \end{aligned}$$

A.4.1 Complete Data Score

The complete data score for a single individual is computed as follows.

The derivative of $\log \mathcal{L}_i(\theta)$ w.r.t. θ_1 :

$$\frac{\partial \log \mathcal{L}_i(\theta)}{\partial \theta_1} = \frac{\partial \ell_{i1}(\theta_1)}{\partial \theta_1} = \sum_{j=1}^{n_i} \frac{\partial}{\partial \theta_1} \log \rho(s_{ij} \mid X_i; \theta_1) - u_{i1} \frac{\partial}{\partial \theta_1} R(w_i \mid X_i; \theta_1)$$

where

$$\begin{aligned} \frac{\partial}{\partial \theta_1} \log \rho(s_{ij} \mid X_i; \theta_1) &= \begin{pmatrix} \frac{\partial}{\partial \alpha_1} \log \rho_0(s_{ij}; \alpha_1) \\ X_i \end{pmatrix}_{(K_1+p) \times 1}, \\ \frac{\partial}{\partial \theta_1} R(w_i \mid X_i; \theta_1) &= \begin{pmatrix} \frac{\partial}{\partial \alpha_1} R_0(w_i; \alpha_1) e^{X_i' \beta_1} \\ R_0(w_i; \alpha_1) X_i e^{X_i' \beta_1} \end{pmatrix}_{(K_1+p) \times 1}, \end{aligned}$$

with

$$\frac{\partial}{\partial \alpha_1} \log \rho_0(s_{ij}; \alpha_1) = \begin{pmatrix} \vdots \\ \delta_{1k}(s_{ij})/\rho_k \\ \vdots \end{pmatrix}_{K_1 \times 1} \quad \text{and} \quad \frac{\partial}{\partial \alpha_1} R_0(w_i; \alpha_1) = \begin{pmatrix} \vdots \\ d_{1k}(w_i) \\ \vdots \end{pmatrix}_{K_1 \times 1}.$$

The derivative of $\log \mathcal{L}_i(\theta)$ w.r.t. θ_2 :

$$\frac{\partial \log \mathcal{L}_i(\theta)}{\partial \theta_2} = \frac{\partial}{\partial \theta_2} \ell_{i2}(\theta_2) + \frac{\partial}{\partial \theta_2} \ell_{i3}(\theta_2, \phi_1, \psi)$$

where

$$\begin{aligned} \frac{\partial}{\partial \theta_2} \ell_{i2}(\theta_2) &= \delta_i \frac{\partial}{\partial \theta_2} \log h(w_i \mid X_i; \theta_2) - \frac{\partial}{\partial \theta_2} H(w_i \mid X_i; \theta_2), \\ \frac{\partial}{\partial \theta_2} \ell_{i3}(\theta_2, \phi_1, \psi) &= \delta_i \frac{\partial}{\partial \theta_2} \log g_1(u_{i1} \mid T_i = w_i, X_i; \theta_2, \phi_1, \psi) \\ &\quad + (1 - \delta_i) \frac{\partial}{\partial \theta_2} \log g_1(u_{i1} \mid T_i > w_i, X_i; \theta_2, \phi_1, \psi). \end{aligned}$$

Note

$$\begin{aligned} \frac{\partial}{\partial \theta_2} \log h(w_i \mid X_i; \theta_2) &= \begin{pmatrix} \frac{\partial}{\partial \alpha_2} \log h_0(w_i; \alpha_2) \\ X_i \end{pmatrix}_{(K_2+p) \times 1}, \\ \frac{\partial}{\partial \theta_2} H(w_i \mid X_i; \theta_2) &= \begin{pmatrix} \frac{\partial}{\partial \alpha_2} H_0(w_i; \alpha_2) e^{X_i' \beta_2} \\ H_0(w_i; \alpha_2) X_i e^{X_i' \beta_2} \end{pmatrix}_{(K_2+p) \times 1}, \end{aligned}$$

with

$$\frac{\partial}{\partial \alpha_2} \log h_0(w_i; \alpha_2) = \begin{pmatrix} \vdots \\ \delta_{2k}(w_i)/h_k \\ \vdots \end{pmatrix}_{K_2 \times 1} \quad \text{and} \quad \frac{\partial}{\partial \alpha_2} H_0(w_i; \alpha_2) = \begin{pmatrix} \vdots \\ d_{2k}(w_i) \\ \vdots \end{pmatrix}_{K_2 \times 1}.$$

Moreover, under the specification of a Gaussian copula for the joint density, (A.2) and (A.4) give

$$\begin{aligned} \frac{\partial}{\partial \theta_2} \log g_1(u_{i1} \mid T_i = w_i, X_i; \theta_2, \phi_1, \psi) &= -\frac{\psi^2 q_1(\theta_2) - \psi q_2(\phi_1)}{1 - \psi^2} \cdot \frac{\partial}{\partial \theta_2} q_1(\theta_2), \\ \frac{\partial}{\partial \theta_2} \log g_1(u_{i1} \mid T_i > w_i, X_i; \theta_2, \phi_1, \psi) &= \frac{\phi(q_3(\theta_2, \phi_1, \psi))}{\Phi(q_3(\theta_2, \phi_1, \psi))} \cdot \frac{1}{\sqrt{1 - \psi^2}} \cdot \frac{\partial}{\partial \theta_2} q_1(\theta_2) \\ &\quad - \frac{1}{\mathcal{F}(w_i \mid X_i; \theta_2)} \cdot \frac{\partial}{\partial \theta_2} \mathcal{F}(w_i \mid X_i; \theta_2), \end{aligned} \quad (\text{A.12})$$

where $\phi(\cdot)$ denotes the p.d.f of a standard Normal distribution,

$$q_3(\theta_2, \phi_1, \psi) = \frac{q_1(\theta_2) - \psi q_2(\phi_1)}{\sqrt{1 - \psi^2}},$$

with $q_1(\theta_2)$, $q_2(\phi_1)$ defined in (A.3) and (A.6), and

$$\begin{aligned} \frac{\partial}{\partial \theta_2} q_1(\theta_2) &= \frac{1}{\phi(q_1(\theta_2))} \cdot \frac{\partial}{\partial \theta_2} \mathcal{F}(w_i \mid X_i; \theta_2), \\ \frac{\partial}{\partial \theta_2} \mathcal{F}(w_i \mid X_i; \theta_2) &= -\mathcal{F}(w_i \mid X_i; \theta_2) \frac{\partial}{\partial \theta_2} H(w_i \mid X_i; \theta_2). \end{aligned}$$

The derivative of $\log \mathcal{L}_i(\theta)$ w.r.t. ϕ_1 :

$$\begin{aligned} \frac{\partial \log \mathcal{L}_i(\theta)}{\partial \phi_1} &= \frac{\partial}{\partial \phi_1} \ell_{i3}(\theta_2, \phi_1, \psi) \\ &= \delta_i \frac{\partial}{\partial \phi_1} \log g_1(u_{i1} \mid T_i = w_i, X_i; \theta_2, \phi_1, \psi) \\ &\quad + (1 - \delta_i) \frac{\partial}{\partial \phi_1} \log g_1(u_{i1} \mid T_i > w_i, X_i; \theta_2, \phi_1, \psi) \end{aligned}$$

where

$$\begin{aligned}\frac{\partial}{\partial \phi_1} \log g_1(u_{i1} \mid T_i = w_i, X_i; \theta_2, \phi_1, \psi) &= \frac{1}{g_1(u_{i1}; \phi_1)} \cdot \frac{\partial}{\partial \phi_1} g_1(u_{i1}; \phi_1) \\ &\quad - \frac{\psi^2 q_2(\phi_1) - \psi q_1(\theta_2)}{1 - \psi^2} \cdot \frac{\partial}{\partial \phi_1} q_2(\phi_1), \\ \frac{\partial}{\partial \phi_1} \log g_1(u_{i1} \mid T_i > w_i, X_i; \theta_2, \phi_1, \psi) &= \frac{\phi(q_3(\theta_2, \phi_1, \psi))}{\Phi(q_3(\theta_2, \phi_1, \psi))} \cdot \frac{-\psi}{\sqrt{1 - \psi^2}} \cdot \frac{\partial}{\partial \phi_1} q_2(\phi_1) \\ &\quad + \frac{1}{g_1(u_{i1}; \phi_1)} \cdot \frac{\partial}{\partial \phi_1} g_1(u_{i1}; \phi_1),\end{aligned}$$

with

$$\frac{\partial}{\partial \phi_1} q_2(\phi_1) = \frac{1}{\phi(q_2(\phi_1))} \cdot \frac{\partial}{\partial \phi_1} G_1(u_{i1}; \phi_1).$$

Under the specification of a Gamma distributed U_{i1} with $E(U_{i1}) = 0$ and $\text{var}(U_{i1}) = \phi_1$, we have

$$\frac{\partial}{\partial \phi_1} g_1(u_{i1}; \phi_1) = g_1(u_{i1}; \phi_1) \frac{1}{\phi_1^2} \left\{ u_{i1} - \log u_{i1} + \psi^{(0)} \left(\frac{1}{\phi_1} \right) + \log \phi_1 - 1 \right\}$$

and we may approximate $\partial G_1(u_{i1}; \phi_1) / \partial \phi_1$ through numerical approaches.

The derivative of $\log \mathcal{L}_i(\theta)$ w.r.t. ψ :

$$\begin{aligned}\frac{\partial \log \mathcal{L}_i(\theta)}{\partial \psi} &= \frac{\partial}{\partial \psi} \ell_{i3}(\theta_2, \phi_1, \psi) \\ &= \delta_i \frac{\partial}{\partial \psi} \log g_1(u_{i1} \mid T_i = w_i, X_i; \theta_2, \phi_1, \psi) \\ &\quad + (1 - \delta_i) \frac{\partial}{\partial \psi} \log g_1(u_{i1} \mid T_i > w_i, X_i; \theta_2, \phi_1, \psi)\end{aligned}$$

where

$$\begin{aligned}\frac{\partial}{\partial \psi} \log g_1(u_{i1} \mid T_i = w_i, X_i; \theta_2, \phi_1, \psi) &= \frac{\psi}{1 - \psi^2} - \frac{\psi[q_1(\theta_2)^2 + q_2(\phi_1)^2] - (\psi^2 + 1)q_1(\theta_2)q_2(\phi_1)}{(1 - \psi^2)^2}, \\ \frac{\partial}{\partial \psi} \log g_1(u_{i1} \mid T_i > w_i, X_i; \theta_2, \phi_1, \psi) &= \frac{\phi(q_3(\theta_2, \phi_1, \psi))}{\Phi(q_3(\theta_2, \phi_1, \psi))} \cdot \frac{q_1(\theta_2)\psi - q_2(\phi_1)}{(1 - \psi^2)^{\frac{3}{2}}}.\end{aligned}$$

The complete data score for one individual is

$$\mathcal{S}_i(\theta) = \left(\frac{\partial \log \mathcal{L}_i(\theta)}{\partial \theta'_1}, \frac{\partial \log \mathcal{L}_i(\theta)}{\partial \theta'_2}, \frac{\partial \log \mathcal{L}_i(\theta)}{\partial \phi_1}, \frac{\partial \log \mathcal{L}_i(\theta)}{\partial \psi} \right)'.$$

A.4.2 Complete Data Information

Note by [Louis \(1982\)](#) (2.11) requires calculations of $\mathcal{I}_i(\theta)$ to obtain $I_i(\theta)$. So we present the expression of $\mathcal{I}_i(\theta)$ here.

The derivative of $\partial \log \mathcal{L}_i(\theta) / \partial \theta_1$ w.r.t. θ_1 :

$$\frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial \theta_1 \partial \theta'_1} = \frac{\partial^2 \ell_{i1}(\theta_1)}{\partial \theta_1 \partial \theta'_1} = \sum_{j=1}^{n_i} \frac{\partial^2}{\partial \theta_1 \partial \theta'_1} \log \rho(s_{ij} | X_i; \theta_1) - u_{i1} \frac{\partial^2}{\partial \theta_1 \partial \theta'_1} R(w_i | X_i; \theta_1)$$

where

$$\begin{aligned} \frac{\partial^2}{\partial \theta_1 \partial \theta'_1} \log \rho(s_{ij} | X_i; \theta_1) &= \begin{pmatrix} \frac{\partial^2}{\partial \alpha_1 \partial \alpha'_1} \log \rho_0(s_{ij}; \alpha_1) & 0 \\ 0 & 0 \end{pmatrix}_{(K_1+p) \times (K_1+p)}, \\ \frac{\partial^2}{\partial \theta_1 \partial \theta'_1} R(w_i | X_i; \theta_1) &= \begin{pmatrix} 0 & \frac{\partial}{\partial \alpha_1} R_0(w_i; \alpha_1) X'_i e^{X'_i \beta_1} \\ X_i e^{X'_i \beta_1} \frac{\partial}{\partial \alpha'_1} R_0(w_i; \alpha_1) & R_0(w_i; \alpha_1) X_i X'_i e^{X'_i \beta_1} \end{pmatrix}_{(K_1+p) \times (K_1+p)}, \end{aligned}$$

with

$$\frac{\partial^2}{\partial \alpha_1 \partial \alpha'_1} \log \rho_0(s_{ij}; \alpha_1) = \text{diag}(\dots, -\delta_{1k}(s_{ij})/\rho_k^2, \dots)_{K_1 \times K_1}.$$

The derivative of $\partial \log \mathcal{L}_i(\theta) / \partial \theta_1$ w.r.t. other parameters:

$$\begin{aligned} \frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial \theta_1 \partial \theta'_2} &= \frac{\partial^2 \ell_{i1}(\theta_1)}{\partial \theta_1 \partial \theta'_2} = 0_{(K_1+p) \times (K_2+p)} \\ \frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial \theta_1 \partial \phi_1} &= \frac{\partial^2 \ell_{i1}(\theta_1)}{\partial \theta_1 \partial \phi_1} = 0_{(K_1+p) \times 1} \\ \frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial \theta_1 \partial \psi} &= \frac{\partial^2 \ell_{i1}(\theta_1)}{\partial \theta_1 \partial \psi} = 0_{(K_1+p) \times 1} \end{aligned}$$

The derivative of $\partial \log \mathcal{L}_i(\theta) / \partial \theta_2$ w.r.t. θ_2 :

$$\frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial \theta_2 \partial \theta'_2} = \frac{\partial^2}{\partial \theta_2 \partial \theta'_2} \ell_{i2}(\theta_2) + \frac{\partial^2}{\partial \theta_2 \partial \theta'_2} \ell_{i3}(\theta_2, \phi_1, \psi)$$

where

$$\begin{aligned}\frac{\partial^2}{\partial\theta_2\partial\theta'_2}\ell_{i2}(\theta_2) &= \delta_i \frac{\partial^2}{\partial\theta_2\partial\theta'_2} \log h(w_i | X_i; \theta_2) - \frac{\partial^2}{\partial\theta_2\partial\theta'_2} H(w_i | X_i; \theta_2), \\ \frac{\partial^2}{\partial\theta_2\partial\theta'_2}\ell_{i3}(\theta_2, \phi_1, \psi) &= \delta_i \frac{\partial^2}{\partial\theta_2\partial\theta'_2} \log g_1(u_{i1} | T_i = w_i, X_i; \theta_2, \phi_1, \psi) \\ &\quad + (1 - \delta_i) \frac{\partial^2}{\partial\theta_2\partial\theta'_2} \log g_1(u_{i1} | T_i > w_i, X_i; \theta_2, \phi_1, \psi).\end{aligned}$$

Note

$$\begin{aligned}\frac{\partial^2}{\partial\theta_2\partial\theta'_2} \log h(w_i | X_i; \theta_2) &= \begin{pmatrix} \frac{\partial^2}{\partial\alpha_2\partial\alpha'_2} \log h_0(w_i; \alpha_2) & 0 \\ 0 & 0 \end{pmatrix}_{(K_2+p) \times (K_2+p)}, \\ \frac{\partial^2}{\partial\theta_2\partial\theta'_2} H(w_i | X_i; \theta_2) &= \begin{pmatrix} 0 & \frac{\partial}{\partial\alpha_2} H_0(w_i; \alpha_2) X'_i e^{X'_i \beta_2} \\ X_i e^{X_i \beta_2} \frac{\partial}{\partial\alpha'_2} H_0(w_i; \alpha_2) & H_0(w_i; \alpha_2) X_i X'_i e^{X'_i \beta_2} \end{pmatrix}_{(K_2+p) \times (K_2+p)},\end{aligned}$$

with

$$\frac{\partial^2}{\partial\alpha_2\partial\alpha'_2} \log h_0(w_i; \alpha_2) = \text{diag}(\cdots, -\delta_{2k}(w_i)/h_k^2, \cdots)_{K_2 \times K_2}.$$

Moreover, we have

$$\begin{aligned}&\frac{\partial^2}{\partial\theta_2\partial\theta'_2} \log g_1(u_{i1} | T_i = w_i, X_i; \theta_2, \phi_1, \psi) \\ &= -\frac{\psi^2 q_1(\theta_2) - \psi q_2(\phi_1)}{1 - \psi^2} \cdot \frac{\partial}{\partial\theta_2\partial\theta'_2} q_1(\theta_2) - \frac{2\psi q_1(\theta_2)}{1 - \psi^2} \cdot \frac{\partial}{\partial\theta_2} q_1(\theta_2) \frac{\partial}{\partial\theta'_2} q_1(\theta_2),\end{aligned}$$

and

$$\begin{aligned}&\frac{\partial^2}{\partial\theta_2\partial\theta'_2} \log g_1(u_{i1} | T_i > w_i, X_i; \theta_2, \phi_1, \psi) \\ &= \frac{1}{\sqrt{1 - \psi^2}} \cdot \frac{\partial}{\partial\theta_2} q_1(\theta_2) \cdot \frac{\partial}{\partial\theta'_2} \frac{\phi(q_3(\theta_2, \phi_1, \psi))}{\Phi(q_3(\theta_2, \phi_1, \psi))} + \frac{\phi(q_3(\theta_2, \phi_1, \psi))}{\Phi(q_3(\theta_2, \phi_1, \psi))} \cdot \frac{1}{\sqrt{1 - \psi^2}} \cdot \frac{\partial^2}{\partial\theta_2\partial\theta'_2} q_1(\theta_2) \\ &\quad + \frac{1}{\mathcal{F}(w_i | X_i; \theta_2)^2} \cdot \frac{\partial}{\partial\theta_2} \mathcal{F}(w_i | X_i; \theta_2) \frac{\partial}{\partial\theta'_2} \mathcal{F}(w_i | X_i; \theta_2) - \frac{1}{\mathcal{F}(w_i | X_i; \theta_2)} \cdot \frac{\partial^2}{\partial\theta_2\partial\theta'_2} \mathcal{F}(w_i | X_i; \theta_2),\end{aligned}$$

where

$$\frac{\partial^2}{\partial\theta_2\partial\theta'_2} q_1(\theta_2) = \frac{q_1(\theta_2)}{\phi(q_1(\theta_2))} \cdot \frac{\partial}{\partial\theta_2} \mathcal{F}(w_i | X_i; \theta_2) \frac{\partial}{\partial\theta'_2} q_1(\theta_2) + \frac{1}{\phi(q_1(\theta_2))} \cdot \frac{\partial}{\partial\theta_2\partial\theta'_2} \mathcal{F}(w_i | X_i; \theta_2),$$

$$\begin{aligned}\frac{\partial^2}{\partial\theta_2\partial\theta'_2}\mathcal{F}(w_i \mid X_i; \theta_2) &= -\frac{\partial}{\partial\theta_2}H(w_i \mid X_i; \theta_2)\frac{\partial}{\partial\theta'_2}\mathcal{F}(w_i \mid X_i; \theta_2) \\ &\quad - \mathcal{F}(w_i \mid X_i; \theta_2) \cdot \frac{\partial^2}{\partial\theta_2\partial\theta'_2}H(w_i \mid X_i; \theta_2),\end{aligned}$$

and

$$\frac{\partial}{\partial\theta_2}\frac{\phi(q_3(\theta_2, \phi_1, \psi))}{\Phi(q_3(\theta_2, \phi_1, \psi))} = \frac{\partial}{\partial q_3(\theta_2, \phi_1, \psi)}\frac{\phi(q_3(\theta_2, \phi_1, \psi))}{\Phi(q_3(\theta_2, \phi_1, \psi))} \cdot \frac{1}{\sqrt{1-\psi^2}} \cdot \frac{\partial}{\partial\theta_2}q_1(\theta_2),$$

with

$$\frac{\partial}{\partial q_3(\theta_2, \phi_1, \psi)}\frac{\phi(q_3(\theta_2, \phi_1, \psi))}{\Phi(q_3(\theta_2, \phi_1, \psi))} = -\frac{q_3(\theta_2, \phi_1, \psi)\phi(q_3(\theta_2, \phi_1, \psi))\Phi(q_3(\theta_2, \phi_1, \psi)) + \phi(q_3(\theta_2, \phi_1, \psi))^2}{\Phi(q_3(\theta_2, \phi_1, \psi))^2}.$$

The derivative of $\partial \log \mathcal{L}_i(\theta)/\partial\theta_2$ w.r.t. ϕ_1 :

$$\begin{aligned}\frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial\theta_2\partial\phi_1} &= \frac{\partial^2}{\partial\theta_2\partial\phi_1}\ell_{i3}(\theta_2, \phi_1, \psi) \\ &= \delta_i \frac{\partial^2}{\partial\theta_2\partial\phi_1} \log g_1(u_{i1} \mid T_i = w_i, X_i; \theta_2, \phi_1, \psi) \\ &\quad + (1 - \delta_i) \frac{\partial^2}{\partial\theta_2\partial\phi_1} \log g_1(u_{i1} \mid T_i > w_i, X_i; \theta_2, \phi_1, \psi)\end{aligned}\tag{A.13}$$

where

$$\begin{aligned}\frac{\partial^2}{\partial\theta_2\partial\phi_1} \log g_1(u_{i1} \mid T_i = w_i, X_i; \theta_2, \phi_1, \psi) &= \frac{\psi}{1-\psi^2} \cdot \frac{\partial}{\partial\theta_2}q_1(\theta_2) \frac{\partial}{\partial\phi_1}q_2(\phi_1), \\ \frac{\partial^2}{\partial\theta_2\partial\phi_1} \log g_1(u_{i1} \mid T_i > w_i, X_i; \theta_2, \phi_1, \psi) &= \frac{1}{\sqrt{1-\psi^2}} \cdot \frac{\partial}{\partial\theta_2}q_1(\theta_2) \cdot \frac{\partial}{\partial\phi_1}\frac{\phi(q_3(\theta_2, \phi_1, \psi))}{\Phi(q_3(\theta_2, \phi_1, \psi))},\end{aligned}\tag{A.14}$$

with

$$\begin{aligned}\frac{\partial}{\partial\phi_1}\frac{\phi(q_3(\theta_2, \phi_1, \psi))}{\Phi(q_3(\theta_2, \phi_1, \psi))} &= \frac{\partial}{\partial q_3(\theta_2, \phi_1, \psi)}\frac{\phi(q_3(\theta_2, \phi_1, \psi))}{\Phi(q_3(\theta_2, \phi_1, \psi))} \cdot \frac{\partial}{\partial\phi_1}q_3(\theta_2, \phi_1, \psi), \\ \frac{\partial}{\partial\phi_1}q_3(\theta_2, \phi_1, \psi) &= \frac{-\psi}{\sqrt{1-\psi^2}} \cdot \frac{\partial}{\partial\phi_1}q_2(\phi_1).\end{aligned}\tag{A.15}$$

The derivative of $\partial \log \mathcal{L}_i(\theta)/\partial \theta_2$ w.r.t. ψ :

$$\begin{aligned} \frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial \theta_2 \partial \psi} &= \frac{\partial^2}{\partial \theta_2 \partial \psi} \ell_{i3}(\theta_2, \phi_1, \psi) \\ &= \delta_i \frac{\partial^2}{\partial \theta_2 \partial \psi} \log g_1(u_{i1} \mid T_i = w_i, X_i; \theta_2, \phi_1, \psi) \\ &\quad + (1 - \delta_i) \frac{\partial^2}{\partial \theta_2 \partial \psi} \log g_1(u_{i1} \mid T_i > w_i, X_i; \theta_2, \phi_1, \psi) \end{aligned} \quad (\text{A.16})$$

where

$$\begin{aligned} \frac{\partial^2}{\partial \theta_2 \partial \psi} \log g_1(u_{i1} \mid T_i = w_i, X_i; \theta_2, \phi_1, \psi) &= - \frac{2\psi q_1(\theta_2) - (\psi^2 + 1)q_2(\phi_1)}{(1 - \psi^2)^2} \cdot \frac{\partial}{\partial \theta_2} q_1(\theta_2), \\ \frac{\partial^2}{\partial \theta_2 \partial \psi} \log g_1(u_{i1} \mid T_i > w_i, X_i; \theta_2, \phi_1, \psi) &= \frac{\partial}{\partial \theta_2} q_1(\theta_2) \left\{ \frac{1}{\sqrt{1 - \psi^2}} \cdot \frac{\partial}{\partial \psi} \frac{\phi(q_3(\theta_2, \phi_1, \psi))}{\Phi(q_3(\theta_2, \phi_1, \psi))} \right. \\ &\quad \left. + \frac{\psi}{(1 - \psi^2)^{3/2}} \cdot \frac{\phi(q_3(\theta_2, \phi_1, \psi))}{\Phi(q_3(\theta_2, \phi_1, \psi))} \right\}, \end{aligned} \quad (\text{A.17})$$

with

$$\begin{aligned} \frac{\partial}{\partial \psi} \frac{\phi(q_3(\theta_2, \phi_1, \psi))}{\Phi(q_3(\theta_2, \phi_1, \psi))} &= \frac{\partial}{\partial q_3(\theta_2, \phi_1, \psi)} \frac{\phi(q_3(\theta_2, \phi_1, \psi))}{\Phi(q_3(\theta_2, \phi_1, \psi))} \cdot \frac{\partial}{\partial \psi} q_3(\theta_2, \phi_1, \psi), \\ \frac{\partial}{\partial \psi} q_3(\theta_2, \phi_1, \psi) &= \frac{\psi q_1(\theta_2) - q_2(\phi_1)}{(1 - \psi^2)^{3/2}}. \end{aligned} \quad (\text{A.18})$$

The derivative of $\partial \log \mathcal{L}_i(\theta)/\partial \phi_1$ w.r.t. ϕ_1 :

$$\begin{aligned} \frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial \phi_1^2} &= \frac{\partial^2}{\partial \phi_1^2} \ell_{i3}(\theta_2, \phi_1, \psi) \\ &= \delta_i \frac{\partial^2}{\partial \phi_1^2} \log g_1(u_{i1} \mid T_i = w_i, X_i; \theta_2, \phi_1, \psi) \\ &\quad + (1 - \delta_i) \frac{\partial^2}{\partial \phi_1^2} \log g_1(u_{i1} \mid T_i > w_i, X_i; \theta_2, \phi_1, \psi) \end{aligned}$$

where

$$\begin{aligned} \frac{\partial^2}{\partial \phi_1^2} \log g_1(u_{i1} \mid T_i = w_i, X_i; \theta_2, \phi_1, \psi) &= \frac{\partial}{\partial \phi_1} \left\{ \frac{1}{g_1(u_{i1}; \phi_1)} \cdot \frac{\partial}{\partial \phi_1} g_1(u_{i1}; \phi_1) \right\} \\ &\quad - \frac{\partial}{\partial \phi_1} \left\{ \frac{\psi^2 q_2(\phi_1) - \psi q_1(\theta_2)}{1 - \psi^2} \cdot \frac{\partial}{\partial \phi_1} q_2(\phi_1) \right\}, \\ \frac{\partial}{\partial \phi_1} \log g_1(u_{i1} \mid T_i > w_i, X_i; \theta_2, \phi_1, \psi) &= \frac{-\psi}{\sqrt{1 - \psi^2}} \left\{ \frac{\partial}{\partial \phi_1} \frac{\phi(q_3(\theta_2, \phi_1, \psi))}{\Phi(q_3(\theta_2, \phi_1, \psi))} \cdot \frac{\partial}{\partial \phi_1} q_2(\phi_1) \right. \\ &\quad \left. + \frac{\phi(q_3(\theta_2, \phi_1, \psi))}{\Phi(q_3(\theta_2, \phi_1, \psi))} \cdot \frac{\partial^2}{\partial \phi_1^2} q_2(\phi_1) \right\} \\ &\quad + \frac{\partial}{\partial \phi_1} \left\{ \frac{1}{g_1(u_{i1}; \phi_1)} \cdot \frac{\partial}{\partial \phi_1} g_1(u_{i1}; \phi_1) \right\}, \end{aligned}$$

with

$$\begin{aligned} \frac{\partial}{\partial \phi_1} \left\{ \frac{1}{g_1(u_{i1}; \phi_1)} \cdot \frac{\partial}{\partial \phi_1} g_1(u_{i1}; \phi_1) \right\} &= \frac{\frac{\partial^2}{\partial \phi_1^2} g_1(u_{i1}; \phi_1)}{g_1(u_{i1}; \phi_1)} - \frac{\left(\frac{\partial}{\partial \phi_1} g_1(u_{i1}; \phi_1) \right)^2}{g_1(u_{i1}; \phi_1)^2}, \\ \frac{\partial}{\partial \phi_1} \left\{ \frac{\psi^2 q_2(\phi_1) - \psi q_1(\theta_2)}{1 - \psi^2} \cdot \frac{\partial}{\partial \phi_1} q_2(\phi_1) \right\} &= \frac{\psi^2}{1 - \psi^2} \left(\frac{\partial}{\partial \phi_1} q_2(\phi_1) \right)^2 + \frac{\psi^2 q_2(\phi_1) - \psi q_1(\theta_2)}{1 - \psi^2} \frac{\partial^2}{\partial \phi_1^2} q_2(\phi_1), \end{aligned}$$

and

$$\frac{\partial^2}{\partial \phi_1^2} q_2(\phi_1) = \frac{q_2(\phi_1)}{\phi(q_2(\phi_1))} \cdot \frac{\partial}{\partial \phi_1} q_2(\phi_1) \frac{\partial}{\partial \phi_1} G_1(u_{i1}; \phi_1) + \frac{1}{\phi(q_2(\phi_1))} \cdot \frac{\partial^2}{\partial \phi_1^2} G_1(u_{i1}; \phi_1).$$

Note $\frac{\partial^2}{\partial \phi_1^2} g_1(u_{i1}; \phi_1)$ and $\frac{\partial^2}{\partial \phi_1^2} G_1(u_{i1}; \phi_1)$ may be calculated using numerical approaches.

The derivative of $\partial \log \mathcal{L}_i(\theta) / \partial \phi_1$ w.r.t. ψ :

$$\begin{aligned} \frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial \phi_1 \partial \psi} &= \frac{\partial^2}{\partial \phi_1 \partial \psi} \ell_{i3}(\theta_2, \phi_1, \psi) \\ &= \delta_i \frac{\partial^2}{\partial \phi_1 \partial \psi} \log g_1(u_{i1} \mid T_i = w_i, X_i; \theta_2, \phi_1, \psi) \\ &\quad + (1 - \delta_i) \frac{\partial^2}{\partial \phi_1 \partial \psi} \log g_1(u_{i1} \mid T_i > w_i, X_i; \theta_2, \phi_1, \psi) \end{aligned}$$

where

$$\frac{\partial^2}{\partial \phi_1 \partial \psi} \log g_1(u_{i1} \mid T_i = w_i, X_i; \theta_2, \phi_1, \psi) = -\frac{2\psi q_2(\phi_1) - (\psi^2 + 1)q_1(\theta_2)}{(1 - \psi^2)^2} \cdot \frac{\partial}{\partial \phi_1} q_2(\phi_1),$$

$$\begin{aligned} & \frac{\partial}{\partial \phi_1 \partial \psi} \log g_1(u_{i1} \mid T_i > w_i, X_i; \theta_2, \phi_1, \psi) \\ &= -\frac{\partial}{\partial \phi_1} q_2(\phi_1) \left\{ \frac{\psi}{\sqrt{1-\psi^2}} \cdot \frac{\partial}{\partial \psi} \frac{\phi(q_3(\theta_2, \phi_1, \psi))}{\Phi(q_3(\theta_2, \phi_1, \psi))} + \frac{1}{(1-\psi^2)^{3/2}} \cdot \frac{\phi(q_3(\theta_2, \phi_1, \psi))}{\Phi(q_3(\theta_2, \phi_1, \psi))} \right\}. \end{aligned}$$

The derivative of $\partial \log \mathcal{L}_i(\theta) / \partial \psi$ w.r.t. ψ :

$$\begin{aligned} \frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial \psi^2} &= \frac{\partial^2}{\partial \psi^2} \ell_{i3}(\theta_2, \phi_1, \psi) \\ &= \delta_i \frac{\partial^2}{\partial \psi^2} \log g_1(u_{i1} \mid T_i = w_i, X_i; \theta_2, \phi_1, \psi) \\ &\quad + (1 - \delta_i) \frac{\partial^2}{\partial \psi^2} \log g_1(u_{i1} \mid T_i > w_i, X_i; \theta_2, \phi_1, \psi) \end{aligned}$$

where

$$\begin{aligned} & \frac{\partial^2}{\partial \psi^2} \log g_1(u_{i1} \mid T_i = w_i, X_i; \theta_2, \phi_1, \psi) \\ &= \frac{\psi^2 + 1}{(1 - \psi^2)^2} + \frac{2\psi(\psi^2 + 3)q_1(\theta_2)q_2(\phi_1) - (3\psi^2 + 1)[q_1(\theta_2)^2 + q_2(\phi_1)^2]}{(1 - \psi^2)^3}, \end{aligned}$$

and

$$\begin{aligned} & \frac{\partial^2}{\partial \psi^2} \log g_1(u_{i1} \mid T_i > w_i, X_i; \theta_2, \phi_1, \psi) \\ &= \frac{\partial}{\partial \psi} \frac{\phi(q_3(\theta_2, \phi_1, \psi))}{\Phi(q_3(\theta_2, \phi_1, \psi))} \cdot \frac{q_1(\theta_2)\psi - q_2(\phi_1)}{(1 - \psi^2)^{\frac{3}{2}}} + \frac{\phi(q_3(\theta_2, \phi_1, \psi))}{\Phi(q_3(\theta_2, \phi_1, \psi))} \cdot \frac{q_1(\theta_2)(2\psi + 1)^2 - 3q_2(\phi_1)\psi}{(1 - \psi^2)^{\frac{5}{2}}}. \end{aligned}$$

The complete data information for one individual $\mathcal{I}_i(\theta)$ is

$$-\begin{pmatrix} \partial^2 \log \mathcal{L}_i(\theta) / \partial \theta_1 \partial \theta_1' & 0 & 0 & 0 \\ 0 & \partial^2 \log \mathcal{L}_i(\theta) / \partial \theta_2 \partial \theta_2' & \partial^2 \log \mathcal{L}_i(\theta) / \partial \theta_2 \partial \phi_1 & \partial^2 \log \mathcal{L}_i(\theta) / \partial \theta_2 \partial \psi \\ 0 & (\partial^2 \log \mathcal{L}_i(\theta) / \partial \theta_2 \partial \phi_1)' & \partial^2 \log \mathcal{L}_i(\theta) / \partial \phi_1^2 & \partial^2 \log \mathcal{L}_i(\theta) / \partial \phi_1 \partial \psi \\ 0 & (\partial^2 \log \mathcal{L}_i(\theta) / \partial \theta_2 \partial \psi)' & \partial^2 \log \mathcal{L}_i(\theta) / \partial \phi_1 \partial \psi & \partial^2 \log \mathcal{L}_i(\theta) / \partial \psi^2 \end{pmatrix}.$$

A.5 A Semiparametric Cox Model for the Terminal Event

Here we present complete data score and information when the terminal event process is specified to be the semiparametric Cox model.

A.5.1 Complete Data Score

If the recurrent event process is fully parametric, $\partial \log \mathcal{L}_i(\theta) / \partial \theta_1$ has been discussed in Appendix A.4.1. Here, we focus on the scenario where the recurrent event process is also semiparametrically specified. The complete data score for a single individual is computed in the following.

The derivative of $\log \mathcal{L}_i(\theta)$ w.r.t. $dR_0(s)$:

$$\frac{\partial \log \mathcal{L}_i(\theta)}{\partial dR_0(s)} = \frac{\partial \ell_{i1}(\theta_1)}{\partial dR_0(s)} = \bar{Y}_i(s) \left[\frac{dN_{i1}(s)}{dR_0(s)} - u_{i1} e^{X_i' \beta_1} \right]$$

The derivative of $\log \mathcal{L}_i(\theta)$ w.r.t. β_1 :

$$\frac{\partial \log \mathcal{L}_i(\theta)}{\partial \beta_1} = \frac{\partial \ell_{i1}(\theta_1)}{\partial \beta_1} = n_i X_i - u_{i1} X_i e^{X_i' \beta_1} R_0(w_i)$$

The derivative of $\log \mathcal{L}_i(\theta)$ w.r.t. $dH_0(t)$:

$$\frac{\partial \log \mathcal{L}_i(\theta)}{\partial dH_0(t)} = \frac{\partial \ell_{i2}(\theta_2)}{\partial dH_0(t)} + \frac{\partial \ell_{i3}(\theta_2, \phi_1, \psi)}{\partial dH_0(t)}$$

where

$$\frac{\partial \ell_{i2}(\theta_2)}{\partial dH_0(t)} = \bar{Y}_i(t) \left[\frac{dN_{i2}(t)}{dH_0(t)} - e^{X_i' \beta_2} \right],$$

and $\partial \ell_{i3}(\theta_2, \phi_1, \psi) / \partial dH_0(t)$ has the same form as (A.12) but now with

$$\frac{\partial}{\partial dH_0(t)} \mathcal{F}(w_i | X_i; \theta_2) = \mathcal{F}(w_i | X_i; \theta_2) (-\bar{Y}_i(t) e^{X_i' \beta_2}).$$

The derivative of $\log \mathcal{L}_i(\theta)$ w.r.t. β_2 :

$$\frac{\partial \log \mathcal{L}_i(\theta)}{\partial \beta_2} = \frac{\partial \ell_{i2}(\theta_2)}{\partial \beta_2} + \frac{\partial \ell_{i3}(\theta_2, \phi_1, \psi)}{\partial \beta_2}$$

where

$$\frac{\partial \ell_{i2}(\theta_2)}{\partial \beta_2} = \delta_i X_i - X_i e^{X_i' \beta_2} H_0(w_i)$$

and $\partial \ell_{i3}(\theta_2, \phi_1, \psi) / \partial \beta_2$ has the same form as (A.12) but now with

$$\frac{\partial}{\partial \beta_2} \mathcal{F}(w_i | X_i; \theta_2) = \mathcal{F}(w_i | X_i; \theta_2) (-X_i e^{X_i' \beta_2} H_0(w_i)).$$

The derivative of $\log \mathcal{L}_i(\theta)$ w.r.t. ϕ_1 and ψ :

$\partial \log \mathcal{L}_i(\theta)/\partial \phi_1$ and $\partial \log \mathcal{L}_i(\theta)/\partial \psi$ have the same forms as for the fully parametric case discussed in Appendix A.4.1.

The complete data score for one individual is then

$$\mathcal{S}_i(\theta) = \left(\left(\frac{\partial \log \mathcal{L}_i(\theta)}{\partial dR_0(s)} \right)', \left(\frac{\partial \log \mathcal{L}_i(\theta)}{\partial \beta_1} \right)', \left(\frac{\partial \log \mathcal{L}_i(\theta)}{\partial dH_0(t)} \right)', \left(\frac{\partial \log \mathcal{L}_i(\theta)}{\partial \beta_2} \right)', \frac{\partial \log \mathcal{L}_i(\theta)}{\partial \phi_1}, \frac{\partial \log \mathcal{L}_i(\theta)}{\partial \psi} \right)'.$$

A.5.2 Complete Data Information

Same here, we focus on the scenario where both processes are semiparametrically specified. Not all entries of the complete data information are needed because what we need for the variance estimation is actually a sub-matrix of the observed data information.

The derivative of $\partial \log \mathcal{L}_i(\theta)/\partial dR_0(s)$ w.r.t. $dR_0(u)$:

$$\frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial dR_0(s) \partial dR_0(u)'} = \begin{cases} -\bar{Y}_i(s) \frac{dN_{i1}(s)}{dR_0(s)^2} & \text{if } s = u \\ 0 & \text{o.w.} \end{cases}$$

The derivative of $\partial \log \mathcal{L}_i(\theta)/\partial dR_0(s)$ w.r.t. β_1 :

$$\frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial dR_0(s) \partial \beta_1'} = -\bar{Y}_i(s) u_{i1} X_i e^{X_i' \beta_1}$$

The derivative of $\partial \log \mathcal{L}_i(\theta)/\partial \beta_1$ w.r.t. β_1 :

$$\frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial \beta_1 \partial \beta_1'} = -u_{i1} e^{X_i' \beta_1} X_i X_i' R_0(w_i)$$

The derivative of $\partial \log \mathcal{L}_i(\theta)/\partial \theta_1$ w.r.t. other parameters:

$$\begin{aligned} \frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial \theta_1 \partial \theta_2'} &= \frac{\partial^2 \ell_{i1}(\theta_1)}{\partial \theta_1 \partial \theta_2'} = 0 \\ \frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial \theta_1 \partial \phi_1} &= \frac{\partial^2 \ell_{i1}(\theta_1)}{\partial \theta_1 \partial \phi_1} = 0 \\ \frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial \theta_1 \partial \psi} &= \frac{\partial^2 \ell_{i1}(\theta_1)}{\partial \theta_1 \partial \psi} = 0 \end{aligned}$$

The derivatives of $\partial \log \mathcal{L}_i(\theta)/\partial dH_0(t)$ and the derivative of $\partial \log \mathcal{L}_i(\theta)/\partial \beta_2$ w.r.t. ϕ_1 : $\partial^2 \log \mathcal{L}_i(\theta)/\partial dH_0(t)\partial \phi_1$ and $\partial^2 \log \mathcal{L}_i(\theta)/\partial \beta_2 \partial \phi_1$ follow (A.13), (A.14) and (A.15).

The derivatives of $\partial \log \mathcal{L}_i(\theta)/\partial dH_0(t)$ and the derivative of $\partial \log \mathcal{L}_i(\theta)/\partial \beta_2$ w.r.t. ψ : $\partial^2 \log \mathcal{L}_i(\theta)/\partial dH_0(t)\partial \psi$ and $\partial^2 \log \mathcal{L}_i(\theta)/\partial \beta_2 \partial \psi$ follow (A.16), (A.17) and (A.18).

The derivatives of $\partial \log \mathcal{L}_i(\theta)/\partial \phi_1$ w.r.t. (ϕ_1, ψ) and the derivative of $\partial \log \mathcal{L}_i(\theta)/\partial \psi$ w.r.t. ψ :

$\partial^2 \log \mathcal{L}_i(\theta)/\partial \phi_1^2$, $\partial^2 \log \mathcal{L}_i(\theta)/\partial \phi_1 \partial \psi$ and $\partial^2 \log \mathcal{L}_i(\theta)/\partial \psi^2$ all have the same forms as for the fully parametric case discussed in Appendix A.4.2.

The complete data information for one individual is

$$\mathcal{I}_i(\theta) = - \begin{pmatrix} \frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial dR_0(s)\partial dR_0(s)'} & \frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial dR_0(s)\partial \beta_1'} & 0 & 0 & 0 & 0 \\ \left(\frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial dR_0(s)\partial \beta_1'}\right)' & \frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial \beta_1 \partial \beta_1'} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial dH_0(t)\partial dH_0(t)'} & \frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial dH_0(t)\partial \beta_2'} & \frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial dH_0(t)\partial \phi_1} & \frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial dH_0(t)\partial \psi} \\ 0 & 0 & \left(\frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial dH_0(t)\partial \beta_2'}\right)' & \frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial \beta_2 \partial \beta_2'} & \frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial \beta_2 \partial \phi_1} & \frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial \beta_2 \partial \psi} \\ 0 & 0 & \left(\frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial dH_0(t)\partial \phi_1}\right)' & \left(\frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial \beta_2 \partial \phi_1}\right)' & \frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial \phi_1^2} & \frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial \phi_1 \partial \psi} \\ 0 & 0 & \left(\frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial dH_0(t)\partial \psi}\right)' & \left(\frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial \beta_2 \partial \psi}\right)' & \frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial \phi_1 \partial \psi} & \frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial \psi^2} \end{pmatrix}.$$

A.5.3 Two-Stage Score and Information

Two-stage score $\mathcal{U}_i(\theta) = (\mathcal{U}'_{i1}(\theta_2), \mathcal{U}'_{i2}(\theta))'$ and two-stage information $-\partial \mathcal{U}_i(\theta)/\partial \theta'$ are needed for variance estimation using (2.12) and (2.13). Note the first-stage score

$$\mathcal{U}_{i1}(\theta_2) = \frac{\partial \ell_{i2}(\theta_2)}{\partial \theta_2}$$

has been derived in Appendix A.5.1, and the first-stage information is given in the below

$$\begin{aligned} -\frac{\partial^2 \ell_{i2}(\theta_2)}{\partial dH_0(t)\partial dH_0(u)'} &= \begin{cases} \bar{Y}_i(t) \frac{dN_{i2}(t)}{dH_0(t)^2} & \text{if } t = u \\ 0 & \text{o.w.} \end{cases}, \\ -\frac{\partial^2 \ell_{i2}(\theta_2)}{\partial dH_0(t)\partial \beta_2} &= \bar{Y}_i(t) X_i e^{X_i' \beta_2}, \\ -\frac{\partial^2 \ell_{i2}(\theta_2)}{\partial \beta_2 \partial \beta_2'} &= e^{X_i' \beta_2} X_i X_i' H_0(w_i). \end{aligned}$$

Moreover, note the second-stage score $\mathcal{U}_{i2}(\theta) = \partial \ell_i(\theta) / \partial \bar{\theta}$ is a sub-matrix of the observed data score $S_i(\theta)$, and the second-stage information

$$-\frac{\partial \mathcal{U}_{i2}(\theta)}{\partial \theta'} = - \begin{pmatrix} \frac{\partial^2 \ell_i(\theta)}{\partial \bar{\theta} \partial \theta'_2} & \frac{\partial^2 \ell_i(\theta)}{\partial \bar{\theta} \partial \bar{\theta}'} \end{pmatrix}$$

is a sub-matrix of the observed data information $I_i(\theta)$, thus we can obtain them using Louis method in (2.11).

Appendix B

Appendix to Chapter 3

B.1 Calculations for the EM Algorithm

The EM algorithm of Sections 3.2.3 and 3.3.2 requires calculation of some expectation terms (e.g. see (3.4), (3.5a) and (3.5b)). The following computational procedures follow what was present in Appendix A.2. In general, for some function of u_1 and $\bar{\mathcal{H}}_i(w_i^+)$, denoted by $m_i(u_1; \theta)$, we have

$$E(m_i(U_{i1}; \theta) \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)}) = \frac{\int_0^\infty m_i(u_1; \theta) P(\bar{\mathcal{H}}_i(w_i^+), U_1 = u_1; \theta^{(r-1)}) du_1}{\int_0^\infty P(\bar{\mathcal{H}}_i(w_i^+), U_1 = u_1; \theta^{(r-1)}) du_1}.$$

Under a specification of a Gaussian copula for the dependence model, if we let $V_2 \sim N(0, 1)$, $V_1^{(r-1)} \sim N(\psi^{(r-1)} q_1(\theta_2^{(r-1)}), 1 - (\psi^{(r-1)})^2)$ with

$$q_1(\theta_2) = \Phi^{-1}(\mathcal{F}(w_i \mid X_i; \theta_2)) \quad \text{and} \quad q_2(\phi_1) = \Phi^{-1}(G_1(u_1; \phi_1)),$$

this expectation can be further expressed as

$$\begin{aligned} & \left\{ \frac{E_{V_1^{(r-1)}} \left\{ m_i(G_1^{-1}(\Phi(V_1^{(r-1)}); \phi_1^{(r-1)}); \theta) M_i(G_1^{-1}(\Phi(V_1^{(r-1)}); \phi_1^{(r-1)}); \theta_1^{(r-1)}) \right\}}{W_{i1}(\theta^{(r-1)})} \right\}^{\delta_i} \\ & \times \left\{ \frac{E_{V_2} \left\{ m_i(G_1^{-1}(\Phi(V_1^{(r-1)}); \phi_1^{(r-1)}); \theta) M_i(G_1^{-1}(\Phi(V_2); \phi_1^{(r-1)}); \theta_1^{(r-1)}) \Phi \left(\frac{q_1(\theta_2^{(r-1)}) - \psi^{(r-1)} V_2}{\sqrt{1 - (\psi^{(r-1)})^2}} \right) \right\}}{\mathcal{F}(w_i \mid X_i; \theta_2^{(r-1)}) W_{i2}(\theta^{(r-1)})} \right\}^{1-\delta_i} \end{aligned}$$

where

$$\begin{aligned} M_i(u_1; \theta_1) &= u_1^{n_i} \exp(-u_1 R(a_{ir_i} \mid X_i; \theta_1)), \\ W_{i1}(\theta) &= \int_0^\infty M_i(u_1; \theta_1) g_1(u_1 \mid X_i, T_i = w_i; \theta_2, \phi_1, \psi) du_1 \\ &= E_{V_1} \{M_i(G_1^{-1}(\Phi(V_1)); \phi_1); \theta_1\}, \end{aligned} \quad (\text{B.1})$$

and

$$\begin{aligned} W_{i2}(\theta) &= \int_0^\infty M_i(u_1; \theta_1) g_1(u_1 \mid X_i, T_i > w_i; \theta_2, \phi_1, \psi) du_1 \\ &= \frac{1}{\mathcal{F}(w_i \mid X_i; \theta_2)} E_{V_2} \left\{ M_i(G^{-1}(\Phi(V_2)); \phi_1); \theta_1 \right\} \Phi \left(\frac{q_1(\theta_2) - \psi V_2}{\sqrt{1 - \psi^2}} \right). \end{aligned} \quad (\text{B.2})$$

Expectation terms in (3.4), (3.5a) and (3.5b) can be calculated if we specify $m_i(u_1; \theta) = u_1$, $m_i(u_1; \theta) = \log g_1(u_1 \mid X_i, T_i = w_i; \theta_2, \phi_1, \psi)$ and $m_i(u_1; \theta) = \log g_1(u_1 \mid X_i, T_i > w_i; \theta_2, \phi_1, \psi)$.

The Louis formula in (3.6) requires calculations of the conditional expectation of the square of the complete data score (i.e., $E(\mathcal{S}_i(\theta) \mathcal{S}_i'(\theta) \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)})$), which involves for example, calculations of $E(N_{ijk} N_{ihl} \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)})$ and $E(N_{ijk} U_{i1} \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)})$ (Louis, 1982). Here, we show the calculation details.

Calculations of $E(N_{ijk} N_{ihl} \mid \bar{\mathcal{H}}_i(w_i^+); \theta)$:

If $j \neq h$, $E(N_{ijk} N_{ihl} \mid \bar{\mathcal{H}}_i(w_i^+); \theta) = E(N_{ijk} \mid \bar{\mathcal{H}}_i(w_i^+); \theta) E(N_{ihl} \mid \bar{\mathcal{H}}_i(w_i^+); \theta)$.

If $j = h$,

- when $j \neq r_i + 1$, we have

$$E(N_{ijk} N_{ihl} \mid \bar{\mathcal{H}}_i(w_i^+); \theta) = \begin{cases} n_{ij} p_{i1}(\theta_1)(1 - p_{i1}(\theta_1)) + n_{ij}^2 p_{i1}^2(\theta_1), & k = l, \\ (n_{ij}^2 - n_{ij}) p_{i1}(\theta_1) p_{i2}(\theta_1), & k \neq l, \end{cases}$$

where $p_{i1}(\theta_1) = \mu_{ijk}(\theta_1)/\mu_{ij}(\theta_1)$ and $p_{i2}(\theta_1) = \mu_{ijl}(\theta_1)/\mu_{ij}(\theta_1)$;

- when $j = r_i + 1$, we have

$$E(N_{ijk} N_{ihl} \mid \bar{\mathcal{H}}_i(w_i^+); \theta) = \begin{cases} E(N_{ij} \mid \bar{\mathcal{H}}_i(w_i^+); \theta) p_{i1}(\theta_1)(1 - p_{i1}(\theta_1)) + E(N_{ij}^2 \mid \bar{\mathcal{H}}_i(w_i^+); \theta) p_{i1}^2(\theta_1), & k = l, \\ (E(N_{ij}^2 \mid \bar{\mathcal{H}}_i(w_i^+); \theta) - E(N_{ij} \mid \bar{\mathcal{H}}_i(w_i^+); \theta)) p_{i1}(\theta_1) p_{i2}(\theta_1), & k \neq l, \end{cases}$$

with

$$E(N_{ij} \mid \bar{\mathcal{H}}_i(w_i^+); \theta) = E(U_{i1} \mid \bar{\mathcal{H}}_i(w_i^+); \theta) \mu_{ij}(\theta_1)$$

and

$$E(N_{ij}^2 \mid \bar{\mathcal{H}}_i(w_i^+); \theta) = E(U_{i1} \mid \bar{\mathcal{H}}_i(w_i^+); \theta) \mu_{ij}(\theta_1) + E(U_{i1}^2 \mid \bar{\mathcal{H}}_i(w_i^+); \theta) \mu_{ij}^2(\theta_1).$$

Calculations of $E(N_{ijk} m_i(U_{i1}; \theta) \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)})$:

$$E(N_{ijk} m_i(U_{i1}; \theta) \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)}) = \begin{cases} n_{ij} \frac{\mu_{ijk}(\theta_1^{(r-1)})}{\mu_{ij}(\theta_1^{(r-1)})} E(m_i(U_{i1}; \theta) \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)}), & j \neq r_i + 1, \\ \mu_{ijk}(\theta_1^{(r-1)}) E(U_{i1} m_i(U_{i1}; \theta) \mid \bar{\mathcal{H}}_i(w_i^+); \theta^{(r-1)}), & j = r_i + 1. \end{cases}$$

B.2 Data Generation for Simulation Studies

The data generation procedure is similar to what we describe in B.2 when the recurrent events are subject to right-censoring. Besides the steps described in B.2, we also generate assessment times for each individual when recurrent events are assessed intermittently. For each simulation, the data generation procedure for the i th individual is described as follows.

1. Generate X_i as Bernoulli with $P(X_i = 1) = 0.5$.
2. Generate $U_{i1} \sim \text{Gamma}(\text{mean} = 1, \text{variance} = \phi_1)$.
3.
 - If Kendall's $\tau = 0$, generate the failure time t_i from an Exponential distribution with

$$T_i \sim \text{Exponential}(\text{rate} = h_0 e^{x_i \beta_2}).$$

- If Kendall's $\tau \neq 0$, generate p_i from Uniform(0, 1) first, and then solve for t_i such that

$$\begin{aligned} p_i &= P(T_i > t_i \mid U_{i1} = u_{i1}, X_i = x_i; \theta_2, \phi_1, \psi) \\ &= \frac{P(U_{i1} = u_{i1}, T_i > t_i \mid X_i = x_i; \theta_2, \phi_1, \psi)}{P(U_{i1} = u_{i1}; \phi_1)} \\ &= \Phi \left(\frac{\Phi^{-1}(\mathcal{F}(t_i \mid X_i = x_i; \theta_2)) - \psi \Phi^{-1}(G_1(u_{i1}; \phi_1))}{\sqrt{1 - \psi^2}} \right) \end{aligned}$$

where $\mathcal{F}(t \mid X_i = x_i; \theta_2) = \exp(-h_0 t e^{x_i \beta_2})$. This gives

$$t_i = \frac{-\log \left\{ \Phi \left[\Phi^{-1}(p_i) \sqrt{1 - \psi^2} + \psi \Phi^{-1}(G_1(u_{i1}; \phi_1)) \right] \right\}}{h_0 e^{x_i \beta_2}}.$$

4. Note $C_i = 1$, let $w_i = \min(t_i, 1)$ and $\delta_i = \mathbb{1}(t_i < 1)$.
5. Let the assessment times be a_{ij} , $j = 1, \dots$, we generate a_{ij} via

$$a_{ij} = a_{i,j-1} + \epsilon_{ij}, \quad j = 1, 2, \dots,$$

where $\epsilon_{ij} \sim \text{Exponential}(\text{rate} = \lambda_A)$ with $a_{i0} = 0$. We keep all $a_{ij} \leq w_i$ so we have r_i intervals of observation period $\mathcal{A}_{ij}$, $j = 1, \dots, r_i$.

6. Let the recurrent event times be s_{il} , $l = 0, 1, \dots$, we generate s_{il} via

$$s_{il} = s_{i,l-1} + e_{il}, \quad l = 1, 2, \dots,$$

where $e_{il} \sim \text{Exponential}(\text{rate} = u_{i1} \rho_0 e^{x_i \beta_1})$ with $s_{i0} = 0$. We then calculate $n_{ij} = \sum_l \mathbb{1}(s_{il} \in \mathcal{A}_{ij})$ for $j = 1, \dots, r_i$.

We repeat Steps 1 to 6 for $i = 1, 2, \dots, m$, and a data set is finalized.

B.3 Proportional Hazards Models with Piecewise Constant Baseline Hazards

For parametric analyses, we also specify a Cox proportional hazards model with a piecewise constant baseline hazard for the terminal event. We let K_2 denote the number of pieces of the baseline hazard function for the terminal event and let $0 = b_{20} < b_{21} < \dots < b_{2K_2} = \infty$ denote another pre-specified sequence of cut points of the positive real line. Let $\delta_{2k}(s) = \mathbb{1}(s \in \mathcal{B}_{2k})$ with $\mathcal{B}_{2k} = (b_{2,k-1}, b_{2k}]$ and $d_{2k}(s) = \int_0^s \delta_{2k}(u) du$, $k = 1, \dots, K_2$, so we have $h_0(t; \alpha_2) = \prod_{k=1}^{K_2} h_k^{\delta_{2k}(t)}$, $h(t \mid X_i; \theta_2) = \prod_{k=1}^{K_2} h_k^{\delta_{2k}(t)} e^{X_i' \beta_2}$, $H_0(t; \alpha_2) = \sum_{k=1}^{K_2} h_k d_{2k}(t)$ and $H(t \mid X_i; \theta_2) = \sum_{k=1}^{K_2} h_k d_{2k}(t) e^{X_i' \beta_2}$. We also have

$$f(t \mid X_i; \theta_2) = \prod_{k=1}^{K_2} h_k^{\delta_{2k}(t)} e^{X_i' \beta_2} \exp\left\{-\sum_{k=1}^{K_2} h_k d_{2k}(t) e^{X_i' \beta_2}\right\} \quad \text{and}$$

$$\mathcal{F}(t \mid X_i; \theta_2) = \exp\left\{-\sum_{k=1}^{K_2} h_k d_{2k}(t) e^{X_i' \beta_2}\right\}.$$

Note $\theta_2 = (\alpha'_2, \beta'_2)'$ with $\alpha_2 = (h_1, \dots, h_{K_2})'$. The observed data likelihood in (3.1) can be re-written using some notation $W_{i1}(\theta)$ and $W_{i2}(\theta)$ (defined in (B.1) and (B.2)) as

$$L_i(\theta) \propto h(w_i | X_i; \theta_2)^{\delta_i} \exp(-H(w_i | X_i; \theta_2)) \times \prod_{j=1}^{r_i} \mu_{ij}^{n_{ij}}(\theta_1 | X_i) \times \{W_{i1}(\theta)\}^{\delta_i} \{W_{i2}(\theta)\}^{1-\delta_i},$$

and the corresponding observed data log likelihood is

$$\begin{aligned} \log L_i(\theta) &= \delta_i \log h(w_i | X_i; \theta_2) - H(w_i | X_i; \theta_2) + \sum_{j=1}^{r_i} n_{ij} \log \mu_{ij}(\theta_1 | X_i) \\ &\quad + \delta_i \log W_{i1}(\theta) + (1 - \delta_i) \log W_{i2}(\theta). \end{aligned}$$

B.3.1 Complete Data Score

Recall the complete data loglikelihood can be partitioned into three parts as in (3.3) and note $\ell_{i2}(\theta_2)$ and $\ell_{i3}(\theta_2, \phi_1, \psi)$ have the same expressions as in (2.9) where the recurrent events are subject to right-censoring. Hence we only present the calculations of $\partial \log \mathcal{L}_i(\theta) / \partial \theta_1$ here. The rest of the complete data scores follow Appendix A.4.1.

The derivative of $\log \mathcal{L}_i(\theta)$ w.r.t. θ_1 :

$$\frac{\partial \log \mathcal{L}_i(\theta)}{\partial \theta_1} = \frac{\partial \ell_{i1}(\theta_1)}{\partial \theta_1} = \sum_{j=1}^{r_i+1} \sum_{k=1}^{K_1} n_{ijk} \frac{\partial}{\partial \theta_1} \log \mu_{ijk}(\theta_1) - u_{i1} \sum_{j=1}^{r_i+1} \sum_{k=1}^{K_1} \frac{\partial}{\partial \theta_1} \mu_{ijk}(\theta_1)$$

where

$$\begin{aligned} \frac{\partial}{\partial \theta_1} \log \mu_{ijk}(\theta_1) &= \left(\frac{\partial}{\partial \alpha_1} \log \mu_{ijk}(\theta_1) \right)_{X_i}^{(K_1+p) \times 1}, \\ \frac{\partial}{\partial \theta_1} \mu_{ijk}(\theta_1) &= \mu_{ijk}(\theta_1) \frac{\partial}{\partial \theta_1} \log \mu_{ijk}(\theta_1), \end{aligned}$$

with

$$\frac{\partial}{\partial \alpha_1} \log \mu_{ijk}(\theta_1) = (0 \quad \cdots \quad 0 \quad 1/\rho_k \quad 0 \quad \cdots \quad 0)'_{K_1 \times 1}.$$

The derivative of $\log \mathcal{L}_i(\theta)$ w.r.t. θ_2 , ϕ_1 and ψ :

$\partial \log \mathcal{L}_i(\theta)/\partial \theta_2$, $\partial \log \mathcal{L}_i(\theta)/\partial \phi_1$ and $\partial \log \mathcal{L}_i(\theta)/\partial \psi$ have the same forms discussed in Appendix A.4.1 where the recurrent events are right-censored.

The complete data score contributed from the i th individual is then

$$\mathcal{S}_i(\theta) = \left(\frac{\partial \log \mathcal{L}_i(\theta)}{\partial \theta'_1}, \frac{\partial \log \mathcal{L}_i(\theta)}{\partial \theta'_2}, \frac{\partial \log \mathcal{L}_i(\theta)}{\partial \phi_1}, \frac{\partial \log \mathcal{L}_i(\theta)}{\partial \psi} \right)'.$$

B.3.2 Complete Data Information

Note the method of [Louis \(1982\)](#) in (3.6) requires calculation of $\mathcal{I}_i(\theta)$ to obtain $I_i(\theta)$, so we present the expression of $\mathcal{I}_i(\theta)$ here. Similarly, we only compute $\partial^2 \log \mathcal{L}_i(\theta)/\partial \theta_1 \partial \theta'$ as the rest follow the expressions presented in Appendix A.4.2.

The derivative of $\partial \log \mathcal{L}_i(\theta)/\partial \theta_1$ w.r.t. θ_1 :

$$\frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial \theta_1 \partial \theta'_1} = \frac{\partial^2 \ell_{i1}(\theta_1)}{\partial \theta_1 \partial \theta'_1} = \sum_{j=1}^{r_i+1} \sum_{k=1}^{K_1} n_{ijk} \frac{\partial^2}{\partial \theta_1 \partial \theta'_1} \log \mu_{ijk}(\theta_1) - u_{i1} \sum_{j=1}^{r_i+1} \sum_{k=1}^{K_1} \frac{\partial^2}{\partial \theta_1 \partial \theta'_1} \mu_{ijk}(\theta_1)$$

where

$$\frac{\partial^2}{\partial \theta_1 \partial \theta'_1} \log \mu_{ijk}(\theta_1) = \begin{pmatrix} \frac{\partial^2}{\partial \alpha_1 \partial \alpha'_1} \log \mu_{ijk}(\theta_1) & 0 \\ 0 & 0 \end{pmatrix}_{(K_1+p) \times (K_1+p)},$$

$$\frac{\partial^2}{\partial \theta_1 \partial \theta'_1} \mu_{ijk}(\theta_1) = \frac{\partial}{\partial \theta_1} \mu_{ijk}(\theta_1) \frac{\partial}{\partial \theta'_1} \log \mu_{ijk}(\theta_1) + \mu_{ijk}(\theta_1) \frac{\partial^2}{\partial \theta_1 \partial \theta'_1} \log \mu_{ijk}(\theta_1),$$

with

$$\frac{\partial^2}{\partial \alpha_1 \partial \alpha'_1} \log \mu_{ijk}(\theta_1) = \text{diag}(0, \dots, 0, -1/\rho_k^2, 0, \dots, 0)_{K_1 \times K_1}.$$

The derivative of $\partial \log \mathcal{L}_i(\theta)/\partial \theta_1$ w.r.t. other parameters:

$$\begin{aligned} \frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial \theta_1 \partial \theta'_2} &= \frac{\partial^2 \ell_{i1}(\theta_1)}{\partial \theta_1 \partial \theta'_2} = 0 \\ \frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial \theta_1 \partial \phi_1} &= \frac{\partial^2 \ell_{i1}(\theta_1)}{\partial \theta_1 \partial \phi_1} = 0 \\ \frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial \theta_1 \partial \psi} &= \frac{\partial^2 \ell_{i1}(\theta_1)}{\partial \theta_1 \partial \psi} = 0 \end{aligned}$$

The complete data information contributed from the i th individual is then

$$\mathcal{I}_i(\theta) = - \begin{pmatrix} \partial^2 \log \mathcal{L}_i(\theta) / \partial \theta_1 \partial \theta_1' & 0 & 0 & 0 \\ 0 & \partial^2 \log \mathcal{L}_i(\theta) / \partial \theta_2 \partial \theta_2' & \partial^2 \log \mathcal{L}_i(\theta) / \partial \theta_2 \partial \phi_1 & \partial^2 \log \mathcal{L}_i(\theta) / \partial \theta_2 \partial \psi \\ 0 & (\partial^2 \log \mathcal{L}_i(\theta) / \partial \theta_2 \partial \phi_1)' & \partial^2 \log \mathcal{L}_i(\theta) / \partial \phi_1^2 & \partial^2 \log \mathcal{L}_i(\theta) / \partial \phi_1 \partial \psi \\ 0 & (\partial^2 \log \mathcal{L}_i(\theta) / \partial \theta_2 \partial \psi)' & \partial^2 \log \mathcal{L}_i(\theta) / \partial \phi_1 \partial \psi & \partial^2 \log \mathcal{L}_i(\theta) / \partial \psi^2 \end{pmatrix}.$$

B.4 A Cox Model for The Terminal Event

Here we present the complete data score and information when the terminal event process is specified to be a Cox model with a non-parametric baseline hazard function.

B.4.1 The Complete Data Score

The complete data score for a single individual is computed in the following.

The derivative of $\log \mathcal{L}_i(\theta)$ w.r.t. θ_1 :

$\partial \log \mathcal{L}_i(\theta) / \partial \theta_1$ has the same form as the fully parametric case of Appendix B.3.1.

The derivative of $\log \mathcal{L}_i(\theta)$ w.r.t. $dH_0(t)$, β_2 , ϕ_1 and ψ :

$\partial \log \mathcal{L}_i(\theta) / \partial dH_0(\cdot)$, $\partial \log \mathcal{L}_i(\theta) / \partial \beta_2$, $\partial \log \mathcal{L}_i(\theta) / \partial \phi_1$ and $\partial \log \mathcal{L}_i(\theta) / \partial \psi$ have the same forms discussed in Appendix A.5.1 where the recurrent events are right-censored.

The complete data score for one individual is then

$$\mathcal{S}_i(\theta) = \left(\frac{\partial \log \mathcal{L}_i(\theta)}{\partial \theta_1'}, \left(\frac{\partial \log \mathcal{L}_i(\theta)}{\partial dH_0(\cdot)} \right)', \frac{\partial \log \mathcal{L}_i(\theta)}{\partial \beta_2'}, \frac{\partial \log \mathcal{L}_i(\theta)}{\partial \phi_1}, \frac{\partial \log \mathcal{L}_i(\theta)}{\partial \psi} \right)'.$$

B.4.2 Complete Data Information

Note the semi-parametric analysis is conducted via a two-stage procedure and the second-stage information is a submatrix of the observed data information. Hence, not all entries of the complete data information are needed when using the Louis formula to calculate the observed data information. The terms we need are $\partial \log \mathcal{L}_i(\theta) / \partial \bar{\theta} \partial \theta_2'$ and $\partial \log \mathcal{L}_i(\theta) / \partial \bar{\theta} \partial \bar{\theta}'$. We compute $\partial^2 \log \mathcal{L}_i(\theta) / (\partial \theta_1 \partial \theta')$ here and the rest follow the forms presented in Appendix A.5.2 when the recurrent events are subject to right-censoring.

The derivative of $\partial \log \mathcal{L}_i(\theta) / \partial \theta_1$ w.r.t. θ_1 :

$\partial^2 \log \mathcal{L}_i(\theta) / \partial \theta_1 \partial \theta'_1$ has the same form as for the fully parametric case discussed in Appendix B.3.2.

The derivative of $\partial \log \mathcal{L}_i(\theta) / \partial \theta_1$ w.r.t. other parameters:

$$\begin{aligned} \frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial \theta_1 \partial \theta'_2} &= \frac{\partial^2 \ell_{i1}(\theta_1)}{\partial \theta_1 \partial \theta'_2} = 0 \\ \frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial \theta_1 \partial \phi_1} &= \frac{\partial^2 \ell_{i1}(\theta_1)}{\partial \theta_1 \partial \phi_1} = 0 \\ \frac{\partial^2 \log \mathcal{L}_i(\theta)}{\partial \theta_1 \partial \psi} &= \frac{\partial^2 \ell_{i1}(\theta_1)}{\partial \theta_1 \partial \psi} = 0 \end{aligned}$$

B.4.3 Two-stage Score and Information

The two-stage score $\mathcal{U}_i(\theta) = (\mathcal{U}'_{i1}(\theta_2), \mathcal{U}'_{i2}(\theta))'$ and the two-stage information $-\partial \mathcal{U}_i(\theta) / \partial \theta'$ are needed for variance estimation using (3.9) and (3.10). Note the first-stage score

$$\mathcal{U}_{i1}(\theta_2) = \frac{\partial \ell_{i2}(\theta_2)}{\partial \theta_2}$$

has been derived in Appendix A.5.1, and the first-stage information is calculated in the below

$$\begin{aligned} -\frac{\partial^2 \ell_{i2}(\theta_2)}{\partial dH_0(t) \partial dH_0(u)'} &= \begin{cases} \bar{Y}_i(t) \frac{dN_{i2}(t)}{dH_0(t)^2} & \text{if } t = u \\ 0 & \text{o.w.} \end{cases}, \\ -\frac{\partial^2 \ell_{i2}(\theta_2)}{\partial dH_0(t) \partial \beta_2} &= \bar{Y}_i(t) X_i e^{X_i' \beta_2}, \\ -\frac{\partial^2 \ell_{i2}(\theta_2)}{\partial \beta_2 \partial \beta_2'} &= e^{X_i' \beta_2} X_i X_i' H_0(w_i). \end{aligned}$$

Moreover, note the second-stage score $\mathcal{U}_{i2}(\theta) = \partial \ell_i(\theta) / \partial \bar{\theta}$ is a sub-matrix of the observed data score $S_i(\theta)$, and the second-stage information

$$-\frac{\partial \mathcal{U}_{i2}(\theta)}{\partial \theta'} = -\begin{pmatrix} \frac{\partial^2 \ell_i(\theta)}{\partial \bar{\theta} \partial \theta'_2} & \frac{\partial^2 \ell_i(\theta)}{\partial \bar{\theta} \partial \theta'} \end{pmatrix}$$

is a sub-matrix of the observed data information $I_i(\theta)$, thus we can obtain them using the Louis (1982) method.

Appendix C

Appendix to Chapter 4

C.1 Computation of Some Conditional Expectations

Here we present detailed calculations of $E\{R_i^{(0)}(s) \mid Z_i(0), X_{i2}\}$, $E\{R_{i1}^{(1)}(s) \mid Z_i(0), X_{i2}\}$, $E\{R_i^{(0)}(s; \psi) \mid Z_i(0), X_{i2}\}$ and $E\{R_{i1}^{(1)}(s; \psi) \mid Z_i(0), X_{i2}\}$ for the four naive cost-effective strategies described in Section [4.3.2](#).

C.1.1 Measuring X_1° at the Last Visit

When the marker value is accessed at the last visit and treated without left-truncated, we have $X_{i1}^\circ = X_{i1}^\circ(V_i) = X_{i1}(a_i \bar{A}_i(V_i))$ and $\bar{Y}_i(s) = Y_i(s)Y_i^\dagger(s)$. $E\{R_i^{(0)}(s) \mid Z_i(0), X_{i2}\}$ can be

calculated as

$$\begin{aligned}
& E \left\{ \mathbb{1}(\bar{A}_i(\tau) \geq 1) \bar{Y}_i(s) dN_{i2}(s) \mid Z_i(0), X_{i2} \right\} \tag{C.1} \\
&= \sum_{k=0}^1 P \left(\bar{A}_i(\tau) \geq 1, Z_i(s^-) = k, dN_{i2}(s) = 1 \mid Z_i(0), X_{i2} \right) \\
&= \sum_{k=0}^1 P \left(\bar{A}_i(s^-) \geq 1, Z_i(s^-) = k, dN_{i2}(s) = 1 \mid Z_i(0), X_{i2} \right) \\
&= \sum_{k=0}^1 P \left(\bar{A}_i(s^-) \geq 1, Z_i(s^-) = k \mid Z_i(0), X_{i2} \right) P \left(dN_{i2}(s) = 1 \mid Z_i(s^-) = k, X_{i2} \right) \\
&= \lambda_2(s) ds \sum_{k=0}^1 \sum_{j=1}^{\infty} P \left(\mathcal{Z}_i(s^-) \in \mathcal{C}_{jk}^{az} \mid Z_i(0), X_{i2} \right) \exp(k\beta_1 + X'_{i2}\beta_2)
\end{aligned}$$

where $\mathcal{C}_{jk}^{az} = \mathcal{C}_j^a \cap \mathcal{C}_k^z$. Recall $\mathcal{C}_j^a = \{(a, z, x_1^\circ) : a = j\}$ denote a set of states occupied by individuals with the j th follow-up visit, $\mathcal{C}_k^z = \{(a, z, x_1^\circ) : z = k\}$ denote a set of states occupied by individuals who are in state k for the four-state process depicted in Figure 4.2, and $\mathcal{C}_l^{x_1^\circ} = \{(a, z, x_1^\circ) : x_1^\circ = l\}$ denote a set of states occupied by individuals whose most recently recorded marker value is l . $E\{R_{i1}^{(1)}(s) \mid Z_i(0), X_{i2}\}$ can be calculated as

$$\begin{aligned}
& E \left\{ \mathbb{1}(\bar{A}_i(\tau) \geq 1) \bar{Y}_i(s) X_{i1}^\circ dN_{i2}(s) \mid Z_i(0), X_{i2} \right\} \tag{C.2} \\
&= \sum_{k=0}^1 P \left(\bar{A}_i(\tau) \geq 1, Z_i(s^-) = k, dN_{i2}(s) = 1, X_{i1}^\circ(V_i) = 1 \mid Z_i(0), X_{i2} \right) \\
&= \sum_{k=0}^1 P \left(\bar{A}_i(s^-) \geq 1, Z_i(s^-) = k, dN_{i2}(s) = 1, X_{i1}^\circ(s^-) = 1 \mid Z_i(0), X_{i2} \right) \\
&= \sum_{k=0}^1 P \left(\bar{A}_i(s^-) \geq 1, Z_i(s^-) = k, X_{i1}^\circ(s^-) = 1 \mid Z_i(0), X_{i2} \right) P \left(dN_{i2}(s) = 1 \mid Z_i(s^-) = k, X_{i2} \right) \\
&= \lambda_2(s) ds \sum_{k=0}^1 \sum_{j=1}^{\infty} P \left(\mathcal{Z}_i(s^-) \in \mathcal{C}_{jk1}^{azx_1^\circ} \mid Z_i(0), X_{i2} \right) \exp(k\beta_1 + X'_{i2}\beta_2)
\end{aligned}$$

with $\mathcal{C}_{jkl}^{ax_1^\circ} = \mathcal{C}_{jk}^{az} \cap \mathcal{C}_l^{x_1^\circ}$. Similarly, we can calculate $E\{R_i^{(0)}(s; \psi) \mid Z_i(0), X_{i2}\}$ as

$$\begin{aligned}
& E \left\{ \mathbb{1}(\bar{A}_i(\tau) \geq 1) \bar{Y}_i(s) \exp(\psi' X_i^\circ) \mid Z_i(0), X_{i2} \right\} \\
&= \sum_{l=0}^1 \sum_{k=0}^1 P \left(\bar{A}_i(\tau) \geq 1, Z_i(s^-) = k, X_{i1}^\circ(\tau) = l \mid Z_i(0), X_{i2} \right) \exp(\psi_1 l + \psi_2' X_{i2}) \\
&= \sum_{l=0}^1 \sum_{k=0}^1 P \left(Z_i(s^-) = k \mid Z_i(0), X_{i2} \right) P \left(\bar{A}_i(\tau) \geq 1, X_{i1}^\circ(\tau) = l \mid Z_i(s^-) = k, X_{i2} \right) \exp(\psi_1 l + \psi_2' X_{i2}) \\
&= \sum_{l=0}^1 \sum_{k=0}^1 P \left(\mathcal{Z}_i(s^-) \in \mathcal{C}_k^z \mid Z_i(0), X_{i2} \right) \sum_{j=1}^{\infty} P \left(\mathcal{Z}_i(\tau) \in \mathcal{C}_{jl}^{ax_1^\circ} \mid \mathcal{Z}_i(s^-) \in \mathcal{C}_k^z, X_{i2} \right) \exp(\psi_1 l + \psi_2' X_{i2}),
\end{aligned}$$

and $E\{R_{i1}^{(1)}(s; \psi) \mid Z_i(0), X_{i2}\}$ as

$$\begin{aligned}
& E \left\{ \mathbb{1}(\bar{A}_i(\tau) \geq 1) \bar{Y}_i(s) X_{i1}^\circ \exp(\psi' X_i^\circ) \mid Z_i(0), X_{i2} \right\} \\
&= \sum_{k=0}^1 P \left(\bar{A}_i(\tau) \geq 1, Z_i(s^-) = k, X_{i1}^\circ(\tau) = 1 \mid Z_i(0), X_{i2} \right) \exp(\psi_1 + \psi_2' X_{i2}) \\
&= \sum_{k=0}^1 P \left(\mathcal{Z}_i(s^-) \in \mathcal{C}_k^z \mid Z_i(0), X_{i2} \right) \sum_{j=1}^{\infty} P \left(\mathcal{Z}_i(\tau) \in \mathcal{C}_{j1}^{ax_1^\circ} \mid \mathcal{Z}_i(s^-) \in \mathcal{C}_k^z, X_{i2} \right) \exp(\psi_1 + \psi_2' X_{i2}).
\end{aligned}$$

C.1.2 Measuring X_1° at the Last Visit with Left-Truncation

When the marker value is accessed at the last visit and considered as left-truncated, we have $X_{i1}^\circ = X_{i1}^\circ(V_i) = X_{i1}(a_{i\bar{A}_i}(V_i))$ and $\bar{Y}_i(s) = Y_i(s)Y_i^\dagger(s)Y_i^L(s)$ where $L_i = a_{i\bar{A}_i}(V_i)$.

$E\{R_i^{(0)}(s) \mid Z_i(0), X_{i2}\}$ is now

$$\begin{aligned}
& \sum_{k=0}^1 P(\bar{A}_i(\tau) \geq 1, Z_i(s^-) = k, dN_{i2}(s) = 1, s \geq a_{i\bar{A}_i(V_i)} \mid Z_i(0), X_{i2}) \\
&= \sum_{k=0}^1 P(\bar{A}_i(\tau) \geq 1, Z_i(s^-) = k, dN_{i2}(s) = 1, \bar{A}_i(s) = \bar{A}_i(\tau) \mid Z_i(0), X_{i2}) \\
&= \sum_{k=0}^1 \sum_{j=1}^{\infty} P(\bar{A}_i(\tau) = j, Z_i(s^-) = k, dN_{i2}(s) = 1, \bar{A}_i(s) = \bar{A}_i(\tau) \mid Z_i(0), X_{i2}) \\
&= \sum_{k=0}^1 \sum_{j=1}^{\infty} P(\bar{A}_i(\tau) = j, Z_i(s^-) = k, dN_{i2}(s) = 1, \bar{A}_i(s) = j \mid Z_i(0), X_{i2}) \\
&= \sum_{k=0}^1 \sum_{j=1}^{\infty} P(\bar{A}_i(s^-) = j, Z_i(s^-) = k, dN_{i2}(s) = 1 \mid Z_i(0), X_{i2}) \\
&= \sum_{k=0}^1 P(\bar{A}_i(s^-) \geq 1, Z_i(s^-) = k, dN_{i2}(s) = 1 \mid Z_i(0), X_{i2})
\end{aligned}$$

and $E\{R_{i1}^{(1)}(s) \mid Z_i(0), X_{i2}\}$ is equal to

$$\begin{aligned}
& \sum_{k=0}^1 P(\bar{A}_i(\tau) \geq 1, Z_i(s^-) = k, dN_{i2}(s) = 1, X_{i1}^{\circ}(s^-) = 1, \bar{A}_i(s) = \bar{A}_i(\tau) \mid Z_i(0), X_{i2}) \\
&= \sum_{k=0}^1 \sum_{j=1}^{\infty} P(\bar{A}_i(\tau) = j, Z_i(s^-) = k, dN_{i2}(s) = 1, X_{i1}^{\circ}(s^-) = 1, \bar{A}_i(s) = j \mid Z_i(0), X_{i2}) \\
&= \sum_{k=0}^1 P(\bar{A}_i(s^-) \geq 1, Z_i(s^-) = k, dN_{i2}(s) = 1, X_{i1}^{\circ}(s^-) = 1 \mid Z_i(0), X_{i2}).
\end{aligned}$$

Note these two conditional expectations have the same forms as in (C.1) and (C.2) where the analysis is conducted without left-truncation. This makes sense since $dN_{i2}(s) = 1$ implies no more visits after failure, which means the process has made its last visit by time s , i.e., $s \geq a_{i\bar{A}_i(V_i)}$. The other two conditional expectations $E\{R_i^{(0)}(s; \psi) \mid Z_i(0), X_{i2}\}$ and

$E \left\{ R_{i1}^{(1)}(s; \psi) \mid Z_i(0), X_{i2} \right\}$ can be calculated as

$$\begin{aligned}
& \sum_{l=0}^1 \sum_{k=0}^1 \sum_{j=1}^{\infty} P \left(\bar{A}_i(\tau) = j, Z_i(s^-) = k, X_{i1}^{\circ}(\tau) = l, \bar{A}_i(s^-) = j \mid Z_i(0), X_{i2} \right) \exp(\psi_1 l + \psi'_2 X_{i2}) \\
&= \sum_{l=0}^1 \sum_{k=0}^1 \sum_{j=1}^{\infty} \left\{ P \left(\bar{A}_i(s^-) = j, Z_i(s^-) = k \mid Z_i(0), X_{i2} \right) \exp(\psi_1 l + \psi'_2 X_{i2}) \right. \\
&\quad \left. P \left(\bar{A}_i(\tau) = j, X_{i1}^{\circ}(\tau) = l \mid \bar{A}_i(s^-) = j, Z_i(s^-) = k, X_{i2} \right) \right\} \\
&= \sum_{l=0}^1 \sum_{k=0}^1 \sum_{j=1}^{\infty} \left\{ P \left(\mathcal{Z}_i(s^-) \in \mathcal{C}_{jk}^{az} \mid Z_i(0), X_{i2} \right) \exp(\psi_1 l + \psi'_2 X_{i2}) P \left(\mathcal{Z}_i(\tau) \in \mathcal{C}_{jl}^{ax_1^{\circ}} \mid \mathcal{Z}_i(s^-) \in \mathcal{C}_{jk}^{az}, X_{i2} \right) \right\}
\end{aligned}$$

and

$$\begin{aligned}
& \sum_{k=0}^1 \sum_{j=1}^{\infty} P \left(\bar{A}_i(\tau) = j, Z_i(s^-) = k, X_{i1}^{\circ}(\tau) = 1, \bar{A}_i(s^-) = j \mid Z_i(0), X_{i2} \right) \exp(\psi_1 + \psi'_2 X_{i2}) \\
&= \sum_{k=0}^1 \sum_{j=1}^{\infty} P \left(\mathcal{Z}_i(s^-) \in \mathcal{C}_{jk}^{az} \mid Z_i(0), X_{i2} \right) \exp(\psi_1 + \psi'_2 X_{i2}) P \left(\mathcal{Z}_i(\tau) \in \mathcal{C}_{j1}^{ax_1^{\circ}} \mid \mathcal{Z}_i(s^-) \in \mathcal{C}_{jk}^{az}, X_{i2} \right)
\end{aligned}$$

respectively.

C.1.3 Measuring X_1° at the First Visit

Now we use the marker value measured at the first visit time, i.e., $X_{i1}^{\circ} = X_{i1}^{\circ}(a_{i1}) = X_{i1}(a_{i1})$ and the Cox regression analysis is conducted without left-truncation, i.e., $\bar{Y}_i(s) =$

$Y_i(s)Y_i^\dagger(s)$. $E\{R_i^{(0)}(s) \mid Z_i(0), X_{i2}\}$ can be written as

$$\begin{aligned}
& \sum_{k=0}^1 P(\bar{A}_i(\tau) \geq 1, Z_i(s^-) = k, dN_{i2}(s) = 1, X_{i1}^\circ(a_{i1}) = 1 \mid Z_i(0), X_{i2}) \\
&= \sum_{k=0}^1 P(\bar{A}_i(s^-) \geq 1, Z_i(s^-) = k, dN_{i2}(s) = 1, X_{i1}^\circ(a_{i1}) = 1 \mid Z_i(0), X_{i2}) \\
&= \sum_{k=0}^1 \int_0^\tau P(\bar{A}_i(s^-) \geq 1, Z_i(s^-) = k, dN_{i2}(s) = 1, \bar{A}_i(a_1^-) = 0, Z_i(a_1^-) = 1, dA_i(a_1) = 1 \mid Z_i(0), X_{i2}) da_1 \\
&= \sum_{k=0}^1 \int_0^s P(Z_i(s^-) = k, dN_{i2}(s) = 1, \bar{A}_i(a_1^-) = 0, Z_i(a_1^-) = 1, dA_i(a_1) = 1 \mid Z_i(0), X_{i2}) da_1 \\
&= \sum_{k=0}^1 \int_0^s \left\{ P(Z_i(s^-) = k, \bar{A}_i(a_1^-) = 0, Z_i(a_1^-) = 1, dA_i(a_1) = 1 \mid Z_i(0), X_{i2}) \right. \\
&\quad \left. P(dN_{i2}(s) = 1 \mid Z_i(s^-) = k) \right\} da_1 \\
&= \lambda_2(s) ds \sum_{k=0}^1 \exp(k\beta_1 + X'_{i2}\beta_2) \int_0^s P(Z_i(s^-) = k, \bar{A}_i(a_1^-) = 0, Z_i(a_1^-) = 1, dA_i(a_1) = 1 \mid Z_i(0), X_{i2}) da_1
\end{aligned} \tag{C.3}$$

where $P(Z_i(s^-) = k, \bar{A}_i(a_1^-) = 0, Z_i(a_1^-) = l, dA_i(a_1) = 1 \mid Z_i(0), X_{i2})$ is equal to

$$P(\mathcal{Z}_i(a_1^-) \in \mathcal{C}_{0l}^{az} \mid Z_i(0), X_{i2}) P(\mathcal{Z}_i(s^-) \in \mathcal{C}_k^z \mid \mathcal{Z}_i(a_1) \in \mathcal{C}_{1ll}^{azx_1^1}, X_{i2}) \rho_0(a_1) \exp(l\alpha_1 + X'_{i2}\alpha_2) \tag{C.4a}$$

for $a_1 \leq s$, and

$$P(\mathcal{Z}_i(s^-) \in \mathcal{C}_{0k}^{az} \mid Z_i(0), X_{i2}) P(\mathcal{Z}_i(a_1^-) \in \mathcal{C}_{0l}^{az} \mid \mathcal{Z}_i(s^-) \in \mathcal{C}_{0k}^{az}, X_{i2}) \rho_0(a_1) \exp(l\alpha_1 + X'_{i2}\alpha_2) \tag{C.4b}$$

for $a_1 > s$. Moreover, we can write $E\{R_i^{(0)}(s; \psi) \mid Z_i(0), X_{i2}\}$ as

$$\begin{aligned}
& \sum_{l=0}^1 \sum_{k=0}^1 P(\bar{A}_i(\tau) \geq 1, Z_i(s^-) = k, X_{i1}^\circ(a_{i1}) = l \mid Z_i(0), X_{i2}) \exp(\psi_1 l + \psi'_2 X_{i2}) \\
&= \sum_{l=0}^1 \sum_{k=0}^1 \exp(\psi_1 l + \psi'_2 X_{i2}) \int_0^\tau P(\bar{A}_i(\tau) \geq 1, Z_i(s^-) = k, \bar{A}_i(a_1^-) = 0, Z_i(a_1^-) = l, dA_i(a_1) = 1 \mid Z_i(0), X_{i2}) da_1 \\
&= \sum_{l=0}^1 \sum_{k=0}^1 \exp(\psi_1 l + \psi'_2 X_{i2}) \int_0^\tau P(Z_i(s^-) = k, \bar{A}_i(a_1^-) = 0, Z_i(a_1^-) = l, dA_i(a_1) = 1 \mid Z_i(0), X_{i2}) da_1
\end{aligned}$$

and $E\{R_{i1}^{(1)}(s; \psi) \mid Z_i(0), X_{i2}\}$ as

$$\sum_{k=0}^1 \exp(\psi_1 + \psi'_2 X_{i2}) \int_0^\tau P(Z_i(s^-) = k, \bar{A}_i(a_1^-) = 0, Z_i(a_1^-) = 1, dA_i(a_1) = 1 \mid Z_i(0), X_{i2}) da_1,$$

where $P(Z_i(s^-) = k, \bar{A}_i(a_1^-) = 0, Z_i(a_1^-) = l, dA_i(a_1) = 1 \mid Z_i(0), X_{i2})$ can be obtained from (C.4a) and (C.4b).

C.1.4 Measuring X_1° at the First Visit with Left-Truncation

We still have $X_{i1}^\circ = X_{i1}^\circ(a_{i1}) = X_{i1}(a_{i1})$. The left truncation time L_i is now the first visit time, i.e., $L_i = a_{i1}$, and we have $\bar{Y}_i(s) = Y_i(s)Y_i^\dagger(s)Y_i^L(s)$ under this strategy. Note $E\{R_i^{(0)}(s) \mid Z_i(0), X_{i2}\}$ equals to

$$\begin{aligned} & \sum_{k=0}^1 P(\bar{A}_i(\tau) \geq 1, Z_i(s^-) = k, dN_{i2}(s) = 1, s \geq a_{i1} \mid Z_i(0), X_{i2}) \\ &= \sum_{k=0}^1 P(\bar{A}_i(\tau) \geq 1, Z_i(s^-) = k, dN_{i2}(s) = 1, \bar{A}_i(s) \geq 1 \mid Z_i(0), X_{i2}) \\ &= \sum_{k=0}^1 P(\bar{A}_i(s^-) \geq 1, Z_i(s^-) = k, dN_{i2}(s) = 1 \mid Z_i(0), X_{i2}) \end{aligned}$$

which is calculated in (C.1), and $E\{R_i^{(1)}(s) \mid Z_i(0), X_{i2}\}$ can be written as

$$\begin{aligned} & \sum_{k=0}^1 P(\bar{A}_i(\tau) \geq 1, Z_i(s^-) = k, dN_{i2}(s) = 1, X_{i1}^\circ(a_{i1}) = 1, \bar{A}_i(s) \geq 1 \mid Z_i(0), X_{i2}) \\ &= \sum_{k=0}^1 P(\bar{A}_i(s^-) \geq 1, Z_i(s^-) = k, dN_{i2}(s) = 1, X_{i1}^\circ(a_{i1}) = 1 \mid Z_i(0), X_{i2}), \end{aligned}$$

which can be obtained by (C.3). $E\{R_i^{(0)}(s; \psi) \mid Z_i(0), X_{i2}\}$ can be written as

$$\begin{aligned} & \sum_{l=0}^1 \sum_{k=0}^1 P(\bar{A}_i(\tau) \geq 1, Z_i(s^-) = k, X_{i1}^\circ(a_{i1}) = l, \bar{A}_i(s) \geq 1 \mid Z_i(0), X_{i2}) \exp(\psi_1 l + \psi_2' X_{i2}) \\ &= \sum_{l=0}^1 \sum_{k=0}^1 P(\bar{A}_i(s^-) \geq 1, Z_i(s^-) = k, X_{i1}^\circ(a_{i1}) = l \mid Z_i(0), X_{i2}) \exp(\psi_1 l + \psi_2' X_{i2}) \\ &= \sum_{l=0}^1 \sum_{k=0}^1 \exp(\psi_1 l + \psi_2' X_{i2}) \int_0^s P(\bar{A}_i(s^-) \geq 1, Z_i(s^-) = k, \bar{A}_i(a_1^-) = 0, Z_i(a_1^-) = l, dA_i(a_1) = 1 \mid Z_i(0), X_{i2}) da_1 \\ &= \sum_{l=0}^1 \sum_{k=0}^1 \exp(\psi_1 l + \psi_2' X_{i2}) \int_0^s P(Z_i(s^-) = k, \bar{A}_i(a_1^-) = 0, Z_i(a_1^-) = l, dA_i(a_1) = 1 \mid Z_i(0), X_{i2}) da_1 \end{aligned}$$

and $E\{R_{i1}^{(1)}(s; \psi) \mid Z_i(0), X_{i2}\}$ is calculated as

$$\begin{aligned} & \sum_{k=0}^1 P\left(\bar{A}_i(\tau) \geq 1, Z_i(s^-) = k, X_{i1}^\circ(a_{i1}) = 1, \bar{A}_i(s) \geq 1 \mid Z_i(0), X_{i2}\right) \exp(\psi_1 + \psi'_2 X_{i2}) \\ &= \sum_{k=0}^1 \exp(\psi_1 + \psi'_2 X_{i2}) \int_0^s P\left(Z_i(s^-) = k, \bar{A}_i(a_1^-) = 0, Z_i(a_1^-) = 1, dA_i(a_1) = 1 \mid Z_i(0), X_{i2}\right) da_1. \end{aligned}$$

Similarly we can calculate $P\left(Z_i(s^-) = k, \bar{A}_i(a_1^-) = 0, Z_i(a_1^-) = l, dA_i(a_1) = 1 \mid Z_i(0), X_{i2}\right)$ via (C.4a) and (C.4b).