Failure Time Analysis with Discrete Marker Processes under Intermittent Observation

by

Bing Feng Xie

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Biostatistics

Waterloo, Ontario, Canada, 2021

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: Julie Horrocks
Professor, Dept. of Mathematics and Stats ,
University of Guelph

Supervisor(s): Richard J. Cook
Professor, Dept. of Stats and ActSc,
University of Waterloo

Internal Member: Cecilia Cotton
Associate Professor, Dept. of Stats and ActSc,
University of Waterloo

Kun Liang
Associate Professor, Dept. of Stats and ActSc,
University of Waterloo

Internal-External Member: Carrie McAiney
Associate Professor, School of Public Health Sciences
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

Regression analysis for failure time data is often directed at studying the relationship between a time-dependent biomarker and failure. The Cox regression model and the associated partial likelihood on which inference is based is well-suited for this kind of investigation since the values of time-dependent biomarkers are only required at the observed failure times in the sample. It is common, however, for markers values to be obtained only at periodic clinic visits when biospecimens are acquired for testing. The convention is then to take these values as the working value of the biomarker until the next visit, failure, or censoring. In such settings the assumed biomarker value is typically out-of-date and therefore misrepresents the true value. Joint modeling can be shown to address this misspecification, where the marker process can mitigate the bias from a naive analysis using the last observation carried forward approach.

In Chapter 2 of this thesis an expectation-maximization algorithm is developed for fitting a joint (i.e. multistate) model for an intermittently-observed binary time-dependent biomarker and failure time. This is implemented and assessed empirically through simulation studies and applied to a dataset from a cancer clinical trial studying the relation between a biomarker and the occurrence of a composite endpoint defined as the time of a skeletal complication or death.

Chapter 3 involves a careful study of the asymptotic bias of regression coefficients from a Cox regression model using the conventional approach of carrying biomarker values forward in time from the time of clinic visits until the next measurement occasion, failure or censoring. Using counting process notation and large sample theory related to misspecified models we gain insights into the determinants of the limiting bias. We consider a true underlying Cox model in which the current marker value and a baseline covariate act multiplicatively on a baseline hazard so the bias in the effect of the biomarker and the baseline covariate can be examined. The determinants of the limiting bias include the proportion of time spent in the two marker states, the relation between the baseline covariates and the intensities governing transitions between the marker states, and the frequency of the

measurements. We also define a marker-dependent visit process as one in which the visit intensity depends on the latent marker value. The strength of this association is found to affect the magnitude of the asymptotic bias as well.

An expanded joint model is described in Chapter 4 which incorporates the marker process, failure process, visit process and right-censoring process. This general framework accommodates marker-dependent censoring and marker-dependent visit intensities and so is quite general. It offers a basis for joint modeling of all four processes in order to mitigate the biases from either the conventional last observation carried forward approach, or the simpler joint model of Chapter 2. Note that visit and failure times are observed exactly but are subject to right censoring, so the baseline intensities of these events can be well-estimated. The transitions between marker states are unobserved however so these intensities must be modelled parsimoniously. The focus of the investigation is primarily to study the ability to obtain good estimation of the failure process intensity under a marker-dependent visit process and so this is the setting of the simulation studies. We fit the model to data from a study of the relation between an inflammatory blood marker, the erythrocyte sedimation rate, a baseline genetic marker and the time to joint damage involving patients from the University of Toronto Psoriatic Arthritis Clinic.

A summary is given in Chapter 5 along with some discussion of topics for future research.

Acknowledgements

I would like to express my deep thank and gratitude to my supervisor Dr. Richard J. Cook. I am very grateful for his inspiration, guidance and support. I appreciate him for sharing biostatistics in so many brilliant and fascinating insights. It has been an amazing Ph.D journey working with him. I would also like to sincerely thank my thesis committee members Dr. Cecilia Cotton, Dr. Kun Liang, Dr. Carrie McAiney and Dr. Julie Horrocks for their insightful comments, suggestions and inspiring questions regarding my thesis. My special thank goes to Ker-Ai Lee for her great suggestion and immense knowledge on statistical programming. Last but not least, I would like to thank my friends and family for their love, support and encouragement.

Dedication

This is dedicated to my parents T. Xie and S. Dong.

Table of Contents

List of Figures

List of Tables

Chapter 1

Joint Modeling of Dependent Visit and Life History Data

1.1 Introduction

Multistate models provide a convenient and powerful framework for modeling life history processes (Cook and Lawless, 2018). For example, Andersen (1988) used mutltistate models to analyze the mortality and incidence of nephropathy for patients with diabetes. Gentleman et al. (1994) proposed a multistate approach to model the onset of AIDs for HIV infected patients that have incomplete information about their life history. Klein et al. (2000) and Keiding et al. (2001) discussed a set of models to analyze bone marrow transplantation. Some typical multistate models include progressive disease model for survival analysis (Gil et al., 2007), illness death models (Meira-Machado et al., 2006), and competing risk models (Andersen et al., 2002).

Jewell and Kalbfleisch (1996) termed markers as internal time-dependent covariates which could reflect the state of an indvidual. In longitudinal studies, marker values are assessed at repeated measurements over time. A large number of methods have been proposed to analyze longitudinal data. The three primary classes of methods are mixed effects models, marginal models, and transition models (Zeger and Liang, 1992). For mixed effects models, the individual trajectories are modeled based on the assumption that the

repeated measurements of marker values share a common random effect. Harville (1977) proposed to use the maximum likelihood of linear mixed effect model to iteratively estimate the variance components. Laird and Ware (1982) extended the linear mixed effects model with a two-stage expectation-maximization (EM) algorithm. Models are also developed to get a robust inference under the misspecification of random effects (Verbeke and Lesaffre, 1996, 1997). Marginal models for estimating the average response of the population based on generalized estimating equation (GEE) methods are also appealing (Zeger and Liang, 1986; Zeger et al., 1988). In GEE, estimation of parameters remain consistent if the working correlation matrix is not correctly specified. In transition models, the mean response is modeled given covariates and past responses (Heagerty, 2002). These models have a similar general formulation as many auto-regressive models used in time series as well as models commonly used for discrete state stochastic processes.

Often interest lies in markers only as covariates in models for life history processes. For example, interest may lie in the impact of blood pressure control on risk of cardiovascular events (McEvoy et al., 2016), or poor blood glucose control on the risk of complications in diabetes (Al-Kateb et al., 2008). In this case joint modelling of multistate and longitudinal data are appealing as they accommodate the fact that the marker processes are only under intermittent observation. For continuous markers, Rizopoulos et al. (2008) developed a shared random effect models with a random effect in a linear mixed model which is also a factor in the hazard model for a survival time. This classical joint modelling framework has been extended to deal with different life history processes. Elashoff et al. (2008), for example, considered joint modelling with competing risk data, Dantan et al. (2011) studied a joint model with latent states for the longitudinal process and an illness-death model, and Ferrer et al. (2016) linked a linear mixed effect model for longitudinal data to proportional hazards functions for multistate processes via a shared random effect.

This thesis is concerned with regression analysis of failure time data. A major focus is on the analysis of data when markers are measured intermittently at clinic visits, and used as covariates in Cox regression models. The convention in such settings is to carry forward the most recently recorded marker value and use it until a new updated value is

available but this can lead to biased estimators of covariate effects. de Bruijne et al. (2001) proposed to incorporate weights that were functions of time since the recording of the covariate value to reduce this bias by down-weighting contributions when there is a longer time from the marker measurement to the time of interest. Smoothing techniques and imputation techniques have also been explored (Raboud et al., 1993; Tsiatis et al., 1995; Andersen and Liestøl, 2003). Joint modeling is also a feasible way of mitigating these biases, subject to the modeling assumptions being correct (Rizopoulos, 2012). Cook and Lawless (2018, 2019) classify intermittent observation schemes according to whether the visit process leading to the measurement of the marker process is conditionally independent and hence ignorable in a likelihood analysis, or not. The likelihood in this context is the full likelihood which here would involve modeling the marker jointly with the failure process.

One aim of this thesis is to compute the asymptotic bias of naive estimators obtained from carrying forward the marker value from the time they were last measured until the next measurement occasion or the minimum of the failure and censoring times. This builds on the work of Jiang et al. (2020) who considered a simple setting of a single binary marker under a completely independent visit process. We explore the limiting behavior of additional fixed covariate effects, and also explore the setting in which the (latent) marker values influence the visit intensity-based models.

Interest in joint modeling typically lies in weakly parametric or semiparametric proportional hazards models. We therefore develop an expectation-maximization algorithm (Dempster et al., 1979) to facilitate fitting models with piecewise constant baseline hazards for the failure process and outline an extension to the semiparametric setting. Interest lies in extending these methods to accommodate marker-dependent visit processes, which will require conceptualization of joint model for the marker, visit and failure process.

1.2 Model Formulation for the Disease Process

1.2.1 An Overview of Multistate Models

We consider a multistate process with K states. Let $\mathcal{S} = \{1, \ldots, K\}$ denote the state space. We let $Z(t) = k$ if state k is occupied at time t, $k = 1, \ldots, K$ and assume $Z(0) = 1$ with probability 1. We then let $\{Z(t), t > 0\}$ be the multistate process, and $\mathcal{Z}(t) = \{Z(s), 0 \leq s < t\}$ be the history of the process. Let $\lambda_{kl}(t \,|\, \mathcal{Z}(t))$ be the complete transition intensity of the form

$$\lim_{\Delta t \to 0} \frac{P(Z(t + \Delta t^-) = l \,|\, Z(t^-) = k, \mathcal{Z}(t))}{\Delta t} = \lambda_{kl}(t \,|\, \mathcal{Z}(t)) \tag{1.1}$$

for all $k \neq l$ (Cook and Lawless, 2018). If we let $Y_k(t) = I(Z(t^-) = k)$ indicate the process is in state k at t^- and hence at risk of a transition out of state k, this can also be written as

$$\lim_{\Delta t \to 0} \frac{P(Z(t + \Delta t^-) = l \,|\, \mathcal{Z}(t))}{\Delta t} = Y_k(t)\, \lambda_{kl}(t \,|\, \mathcal{Z}(t))\,.$$

Likelihood Construction with Administrative Right Censoring Times

Let C_A be a fixed administrative censoring time where we assume without loss of generality that $C_A = 1$ so that individuals are to be observed over the interval $[0, 1]$. We partition the interval into R pieces according to the cut-points $0 = u_0 < u_1 < \ldots < u_R = 1$. To introduce counting process notation we first define $N_{kl}(t)$ to be the number of transitions from state k to l over $(0, t]$. We then let $\Delta N_{kl}(u_r) = \Delta N_{kl}(u_r^-) - \Delta N_{kl}(u_{r-1}^-)$ to be the number of $k \to l$ transitions over $[u_{r-1}, u_r)$, $r = 1, \ldots, R$. We further construct a vector $\Delta N_k(u_r) = (\Delta N_{kl}(u_r), l \neq k, l = 1, \ldots, K)'$ to denote whether a transition out of state k happens over the interval, where any non-zero elements indicate which state was entered. Let $\Delta N(u_r) = (\Delta N_k'(u_r), k = 1, \ldots, K)'$ be a full vector of all counting process increments (i.e. for all pairs of states). Let $N_{k\cdot}(t) = \sum_{l=1, l\neq k}^{K} N_{kl}(t)$ count the number of transitions departing state k over $[0, t]$ and we let $\Delta N_{k\cdot}(t) = N_{k\cdot}(t + \Delta t^-) - N_{k\cdot}(t^-)$. Note that $\Delta N_{k\cdot}(t) = 0$ when no transition is made from state k over $[t, t + \Delta t)$. We also let $\mathcal{Z}(u_r) = \{Z(u_s), s = 0, 1, \ldots, r\}$ to be the history of a multistate process at time u_r in the

partition. The likelihood of the multistate process is then given by

$$\lim_{R\to\infty}\prod_{r=1}^{R} P(\Delta N(u_r)\,|\,\mathcal{Z}(u_{r-1})) = \lim_{R\to\infty}\prod_{r=1}^{R}\left\{\prod_{k=1}^{K} P(\Delta N_k(u_r)\,|\,\mathcal{Z}(u_{r-1}))^{Y_k(u_{r-1})}\right\} \tag{1.2}$$

where $Y_k(s) = I(Z(s^-) = k), k = 1,\ldots,K$. Given a sample path $\{\Delta N(u_1),\ldots,\Delta N(u_R)\}$, the limit can be taken as $R\to\infty$, so that each sub-interval only contains at most one transition. Next note that by (1.1) $P(\Delta N_k(u_r)|\mathcal{Z}(u_{r-1}))^{Y_k(u_{r-1})}$ is

$$\left[\prod_{l=1,l\neq k}^{K}[\lambda_{kl}(u_r|\mathcal{Z}(u_{r-1}))\Delta u_r + o(\Delta u_r))^{\Delta N_{kl}(u_r)}]\Big(1-\lambda_{k\cdot}(u_r|\mathcal{Z}(u_{r-1}))\Delta u_r + o(\Delta u_r)\Big)^{1-\Delta N_{k\cdot}(u_r)}\right]^{Y_k(u_{r-1})}$$

where $\lambda_{k\cdot}(u_r|\mathcal{Z}(u_{r-1})) = \sum_{l=1,l\neq k}^{K}\lambda_{kl}(u_r|\mathcal{Z}(u_{r-1}))$. As a result, (1.2) can be rewritten as

$$\begin{aligned}\lim_{R\to\infty}\prod_{k=1}^{K}\Bigg\{\Bigg[\prod_{l=1,l\neq k}^{K}\prod_{r=1}^{R}&(\lambda_{kl}(u_r\,|\,\mathcal{Z}(u_{r-1}))\Delta u_r + o(\Delta u_r))^{\Delta N_{kl}(u_r)}\\ &\times\Big(1-\lambda_{k\cdot}(u_r\,|\,\mathcal{Z}(u_{r-1}))\Delta u_r + o(\Delta u_r)\Big)^{1-\Delta N_{k\cdot}(u_r)}\Bigg]^{Y_k(u_{r-1})}\Bigg\}\end{aligned} \tag{1.3}$$

Dividing equation (1.3) by $\prod_{k=1}^{K}\prod_{l=1,l\neq k}^{K}\prod_{r=1}^{R}(\Delta u_r)^{Y_k(u_{r-1})\Delta N_{kl}(u_r)}$ and taking the limit as $R\to\infty$ gives the likelihood

$$\mathcal{L} = \prod_{k=1}^{K}[\prod_{l=1,l\neq k}^{K}\Bigg[\prod_{t_j\in D_{kl}}\lambda_{kl}(t_j\,|\,\mathcal{Z}(t_j^-))\exp(-\int_0^1 Y_k(u^-)\lambda_{kl}(u\,|\,\mathcal{Z}(u^-)))\Bigg]$$

where D_{kl} is the set of transition times from k to l over $[0,1]$, and t_j represents an element of D_{kl}.

1.2.2 Markov Models

A K-state stochastic process is Markov if given the full history up to time t^-, the probabilistic structure of the future only depends on the state at t^-, but not the history prior to that point (Resnick, 2013). The transition intensity (1.1) then reduces to

$$\lambda_{kl}(t\,|\,\mathcal{Z}(t)) = Y_k(t)\,\lambda_{kl}(t)\,. \tag{1.4}$$

Let $P(s,t)$ denote a $K \times K$ transition probability matrix with entries

$$P_{kl}(s,t) = P(Z(t) = l \,|\, Z(s) = k)$$

satisfying $P_{kl}(s,t) \geq 0$, and $\sum_{l=1, l\neq k}^{K} P_{kl}(s,t) = 1$ for $l = 1, \ldots, K$.

Let $dQ(t)$ be a $K \times K$ matrix with off-diagonal (k,l) entry given by $\lambda_{kl}(t)dt$ and $-\sum_{l\neq k} \lambda_{kl}(t)dt$ in the diagonal entries, $k = 1, \ldots, K$. If $\mathcal{I}$ be a $K \times K$ identity matrix, then the $K \times K$ transition probability matrix is obtained from the $K \times K$ transition intensity matrix $dQ(t)$ by product integration as

$$P(s,t) = \prod_{(s,t]} \{I + dQ(u)\}\,,$$

For time-homogeneous Markov models $\lambda_{kl}(t) = \lambda_{kl}$ and the transition probability matrix can be calculated as

$$P(s,t) = \exp(dQ^* \cdot (t-s))\,,$$

via matrix exponentiation (Cox and Miller, 1965) where $dQ^* = dQ/dt$.

1.2.3 Semi-Markov Models

The transition intensity to state l from state k for a semi-Markov model is of the form

$$\lambda_{kl}(t \,|\, \mathcal{Z}(t)) = Y_k(t) h_{kl}(B(t)) \tag{1.5}$$

where $B(t) = t - t_r$ is the time an individual has spent in state k since the most recent entry to it at t_r. Such models are appealing and suitable when the trend in the risk of transition out of a particular state is best characterized by the time since state entry. For the disease process of interest which involve progressive damage or deterioration the Markov time scale is most suitable and we focus on these models in this thesis.

We next consider some natural models for visit processes.

1.3 A Recurrent Event Models for a Visit Process

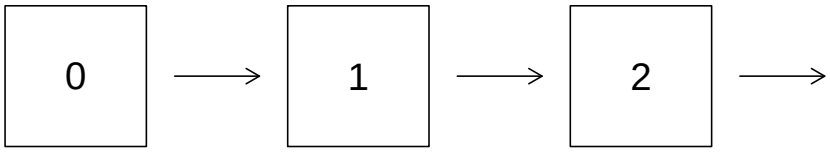


Figure 1.1: A multistate representation of a counting process for visits over time

Visits may be scheduled at particular points in time in some controlled studies, but the schedules are often imperfectly adhered to due to missed visits. Moreover in many observational cohorts, registries, or studies involving electronic medical records, the times of the visits may be considered as governed by a point process. We consider a multistate presentation of a random visit process; here the states represent the cumulative number of assessments after the baseline assessment; see Figure 1.1. We discuss visit process models initially in the absence of random right-censoring and presume visits can occur over the interval $[0, 1]$.

Let $0 = a_{i0} < a_{i1} < \cdots < a_{ir}$ denote the subject-specific visit times over $[0, t]$ for an individual i in a sample of size n, $i = 1, \ldots, n$. Let $A(t)$ denote the number of visits over $(0, t]$, and let $A(1)$ denote the total number of visits over $(0, 1]$, the period of interest for observation. We let $\Delta A(t) = A(t + \Delta t^-) - A(t^-)$ be the number of visits over $[t, t + \Delta t)$ and $dA(t) = \lim_{\Delta t \to 0} \Delta A(t) = 1$ if there is a visit at time t with $dA(t) = 0$ otherwise. Let $\mathcal{A}(t) = \{A(s), 0 \leq s < t\}$ denote the history of the visit process and consider the visit intensity

$$\lim_{\Delta t \to 0} \frac{P(A(t + \Delta t^-) = k + 1 \,|\, A(t^-) = k, \mathcal{A}(t))}{\Delta t} = \rho(t \,|\, \mathcal{A}(t)) \,, t > 0 \,. \tag{1.6}$$

This visit process intensity is based on the implicit assumption that the only history relevant for the visit process intensity is the history of the visits, and hence that the visit process is completely independent of the disease process. This assumption is relaxed in the investigation of Chapter 3 and 4.

A simple and convenient model for an independent visit process is a Poisson process in which $\rho(t \,|\, \mathcal{A}(t)) = \rho(t)$. In such processes, the inter-arrival time between two consecutive visits is also conveniently modeled. A stochastic counting process $\{A(s), 0 < s\}$ is a Poisson process with rate $\rho(t)$ if

1. $A(0) = 0$,
2. It features independent increments whereby $A(t_2) - A(t_1)$ and $A(s_2) - A(s_1)$ are independent if $0 \leq s_1 < s_2 \leq t_1 < t_2$,
3. $A(t+s) - A(s) \sim \text{Poisson}(\mu(s, s+t))$ if $\mu(s, s+t) = \int_s^{s+t} \rho(u)du$

where $\mu(s, s+t)$ is the integrated rate function over interval $[s, s+t]$. Note that the Poisson process is time homogeneous when $\rho(t) = \rho$ is a constant, and time non-homogeneous otherwise. Let $W_k = A_k - A_{k-1}$ for $k = 1, \ldots, K$ denote the waiting time of the visit process. For a time homogeneous Poisson process with rate ρ, it can be shown that the waiting time $W_1, \ldots, W_K$ are independent and identically distributed exponential random variables with the same hazard ρ (Cook and Lawless, 2007).

To allow for the extra variation of a standard Poisson process, one could consider a mixed Poisson model. In this model, we introduce a random effect U, with $G(u; \phi)$ denoting the cumulative distribution function and $\mathcal{G}(u; \phi) = 1 - G(u; \phi)$ denoting the corresponding survivor function. We assume $E(U) = 1$ and $Var(U) = \phi$. Conditional on the random effect u, the cumulative number of visits over $(0, 1]$ is Poisson with mean $u \cdot \mu(t)$, where $\mu(t) = \int_0^t \rho(s)ds$. The number of visits for each individual then arises from a mixed Poisson process with conditional rate (given $U = u$) $u\rho(t)$:

$$A(t) \,|\, u \sim \text{PP}(\text{rate} = u\rho(t)) \tag{1.7}$$

Note that if U is gamma distributed, the result is a negative binomial process (Lawless, 1987), with intensity

$$\rho(t \,|\, \mathcal{A}(t)) = \frac{1 + \phi A(t^-)}{1 + \phi\mu(t)}\rho(t).$$

Here, $P(A(1) = 0; \rho, \phi)$ corresponds to the probability that an individual does not have any visit until the administrative censoring time, where

$$\begin{aligned} P(A(1) = 0; \rho, \phi) = E_u\{e^{-u\rho}\} &= \int_0^\infty e^{-u\rho} \frac{u^{\phi^{-1}-1}\, e^{-u/\phi}}{\Gamma(\phi^{-1})\phi^{\phi^{-1}}}\, du \\ &= \frac{1}{\Gamma(\phi^{-1})\phi^{\phi^{-1}}} \int_0^\infty u^{\phi^{-1}-1} e^{-u/\frac{\phi}{1+\rho\phi}} du = \left(\frac{1}{1+\rho\phi}\right)^{\phi^{-1}}. \end{aligned} \tag{1.8}$$

More generally

$$P(A(1) = r; \rho, \phi) = \frac{\Gamma(\phi^{-1} + r)}{\Gamma(\phi^{-1})r!} \frac{(\phi\mu(1))^r}{[1 + \phi\mu(1)]^{\phi^{-1}+r}}, r = 0, 1, \ldots$$

Many joint models are constructed based on shared or correlated random effects so it would be natural to consider linking the multistate process to the visit process in this manner. We do not adopt this approach however as these models have some undesirable properties: they will not accommodate Poisson visit processes and they require introduction of random effects into the multistate process.

If we consider $\mathcal{H}(t) = \{A(s), Z(s), 0 \leq s < t\}$, the joint history of the visit and multistate process, we can write the intensity of a joint multistate-visit process model as

$$\lim_{\Delta t \to 0} \frac{P(Z(t + \Delta t^-) = l \,|\, \mathcal{H}(t))}{\Delta t} = Y_k(t)\lambda_{kl}(t \,|\, \mathcal{H}(t)), \tag{1.9}$$

$l \neq k$ and $l, k = 1, \ldots, K$, with

$$\lim_{\Delta t \to 0} \frac{P(\Delta A(t) = 1 \,|\, \mathcal{H}(t))}{\Delta t} = Y(t)\rho(t \,|\, \mathcal{H}(t)) \tag{1.10}$$

where $Y(t)$ indicates they are still at risk of a visit (here we assume $Y(t) = 1$ over $[0, 1]$). These intensities govern the joint multistate and visit processes, but the visit process

intensity in (1.10) is not always convenient to work with in isolation since $\mathcal{Z}(t)$ is not observed under the intermittent visit process. A joint model for the multistate and visit processes is possible to specify, however and we explore their use in this thesis.

1.4 Motivating Studies

1.4.1 A Bone Marker Study

In the bone marker study, our interest lies in evaluating the effect of treatment with bisphosphonate on reducing the incidence of skeletal complications for prostate cancer patients (Saad et al., 2004). In cancer patients, the disruption of bone resorption could weaken the skeleton and therefore these patients may have a higher risk of skeletal complications such as pathological and vertebral fractures. The researchers recorded the level of bone resorption marker Ntx and the level of bone formation marker BALP in prostate cancer patients as these two markers are found to be related with the occurrence of the skeletal complications in the disease process (Demers et al., 1995).

This study consists of three randomized, double-blinded, phase III clinical trials (Saad et al., 2004), where patients were randomly assigned into either the placebo treatment group or the bisphosphonate (zoledronate in the study) treatment group. Patients are followed for up to 24 months, and throughout the period, they get treatment every three to four weeks. The measurements of bone marker (Ntx and BALP) are taken 1) at randomization time; 2) after one month of treatment; 3) once every three months in the time of study.

A sample pathway for the NTX and BALP profile of two placebo treated patients (left panel) and two bisphosphonate treated patients (right panel) can be seen in Figure 1.2. In Chapter 2, we consider different methods which deals with evaluating the effect of marker value on failure time model while accommodating the intermittent observation.

1.4.2 Development of Joint Damage in Psoriatic Arthritis

The University of Toronto Psoriatic Arthritis (PsA) Cohort has registered close to 2000 patients with PsA over the past 40 years (Gladman and Chandran, 2011). Psoriatic arthri-

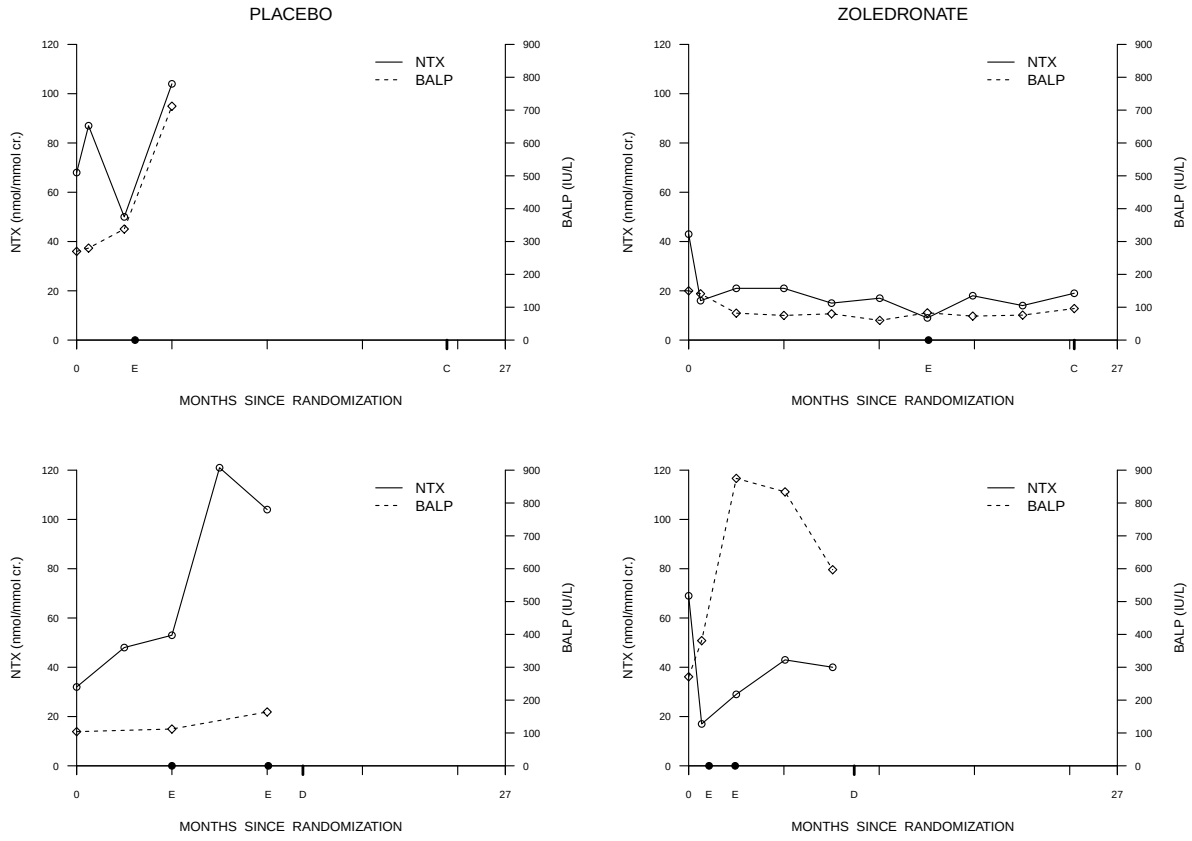


Figure 1.2: A diagram depicting the profile of two prostate cancer patients with placebo treatment (left panel) and two prostate cancer patients with zoledronate treatment (right panel). The X-axis indicates the time of skeletal event (E), death (D), and end of follow-up (C). The Y-axis indicates an osteoclast (Ntx) and osteoblast (BALP) marker for bone disease; Figure from Cook (2014)

tis is an immunological disease which could lead to destruction of joints. It mainly affects people who are diagnosed with psoriasis - featured by red patches of skin. In the study, we consider disease progression in terms of the number and severity of damaged joints. The time-dependent covariate of interest is an indicator of elevated marker of inflammation called the erythrocyte sedimentation rate and we consider a baseline covariate indicating the presence of a human leukocyte antigen (HLA) marker HLA-B27 which has been reported as a risk factor for PsA. Our interest lies in characterizing the effect of these covariates on time to joint damage.

Patients are required to have at least one radiological assessment to enter the study, and then they are followed intermittently for roughly every two years to get an X-ray for the severity of joint damage. These patients initially provided serum samples when they entered the study, which is used for genetic testing to determine the HLA-B27 status, and then they are followed up yearly for assessing biomarker levels.

1.5 Outline of the Thesis

This thesis is broadly concerned with studying the bias arising from misspecified Cox regression models when the model contains a time-dependent covariate that is only observed at intermittent inspection times and the model is misspecified because the marker value of the time-varying covariate from the last observation is carried forward. The general theory is provided but in the investigations of the asymptotic and finite sample bias we focus on the special case of a binary time-dependent covariate and a single fixed covariate. We then discuss methods for mitigating this bias through joint modeling.

In Chapter 2, a joint marker-failure time model is considered with a binary time-dependent marker and a fixed covariate vector. Piecewise constant transition intensities are considered as governing movement between states of a reversible illness-death model. Following the presentation of the complete data likelihood which is appropriate when the marker is under continuous observation with a process subject to right-censoring, we developed an expectation-maximization algorithm to accommodate the setting where the

marker process is under a conditionally independent intermittent observation scheme. Remarks on the semiparametric implementation of this algorithm are given in the appendix but this analysis can be approximated by the piecewise constant formulation by increasing the number of pieces in the failure time transition intensities. Simulation studies are then conducted to illustrate the method can give estimates with good properties, which is expected, as they are maximum likelihood estimators. We also consider an application to a study of patients with skeletal metastases at risk of skeletal complications because of these metastases, and relate the treatment and bone marker status to the risk of the composite event skeletal-event free survival.

In Chapter 3, we describe how to determine the asymptotic bias of naive Cox model estimators when the marker is under intermittent observation but the conventional approach to analysis is adopted and the marker values are carried forward from the most recent assessment time, from the time of assessment to the next assessment, censoring, or failure. We do so by considering a joint model for the marker, failure and censoring times, along with a visit process. Integrating the visit process into the joint model framework enables the evaluation of a marker-independent visit process as well as a marker-dependent visit process wherein the propensity for a visit is dependent on the current marker value. Using this framework we evaluate the bias of both the effect of the marker and the effect of the fixed covariate under the settings of independent and marker-dependent right-censoring. We consider the case where the marker transition intensities are possibly associated with the fixed baseline covariate for generality. Despite the fact that time-dependent markers are ubiquitous in health research and that markers often change to reflect exacerbations of symptoms or the severity of disease activity which may prompt individuals to seek medical care, to our knowledge this is the first work to investigate the nature of this bias.

In Chapter 4 the challenge of fitting the joint model for the marker, failure time, and dependent visit process is addressed. We do this by constructing the full likelihood and estimating the parameter governing transitions between the marker states, into the failure state, and the visit process under the assumption of independent censoring. A small simulation study is conducted and reported on to demonstrate that the joint model can be

fitted and provide estimates which have good finite sample properties which demonstrate that the joint model mitigates the bias from the conventional carry-forward approach to the marker values and the assumption that the visit intensity is independent of the disease process. We illustrate the joint model by fitting it to a dataset of patients with psoriatic arthritis where interest lies in assessing the effect of an inflammatory marker on progression of joint damage.

Concluding remarks and topics of future research are given in Chapter 5.

Chapter 2

A Joint Marker-Failure Time Analysis with Intermittent Assessments

2.1 Introduction

Measurements on markers are typically taken when individuals are seen at intermittent clinic visits. Examples can be taken from many health research settings. For example individuals participating in the Canadian Longitudinal Study of Aging (Raina et al., 2009) are seen every three years and at these follow-up visits they provide blood and urine samples which can be used to measure blood glucose, cholesterol levels, and metabolites which may in turn be used to predict onset or progression of disease, or death. In studies of patients with diabetes, blood samples are taken to measure HbA1C, a marker of blood sugar control, which is used to model the risk of diabetic complications including diabetic nephropathy or retinopathy (Mitchell et al., 2020). In studies of rheumatoid arthritis, markers of inflammation such as C-reactive protein (CRP) are measured in the blood samples taken at clinic visits and these too are used to model risk of joint damage (Haroon et al., 2020). In each of these examples the markers vary in continuous time but measurements are only taken at periodic clinic visits and the values of measurements are used for the marker covariate until it is measured next under a last-observation carried forward strategy. This is then used to assess the association between the marker and events of interest. Here we focus on

failure time models for the event of interest based on the proportional hazards assumption.

The above strategy is naive in the sense that the most recently recorded value of the marker does not represent the actual value of the marker at the time of any future failures; it therefore is akin to a covariate measurement error problem where the magnitude of the error in the marker (the difference between the most recently recorded value of the marker and the true current value) will tend to increase over time since last measurement, if there are systematic trends in the marker as a function of disease duration; the volatility of the marker process is also a key factor. Given the widespread practice of carrying the marker value forward and using the most recently recorded value as a covariate in a failure time model, we study the properties of estimators when intermittently observed markers are handled according to this strategy in the next chapter.

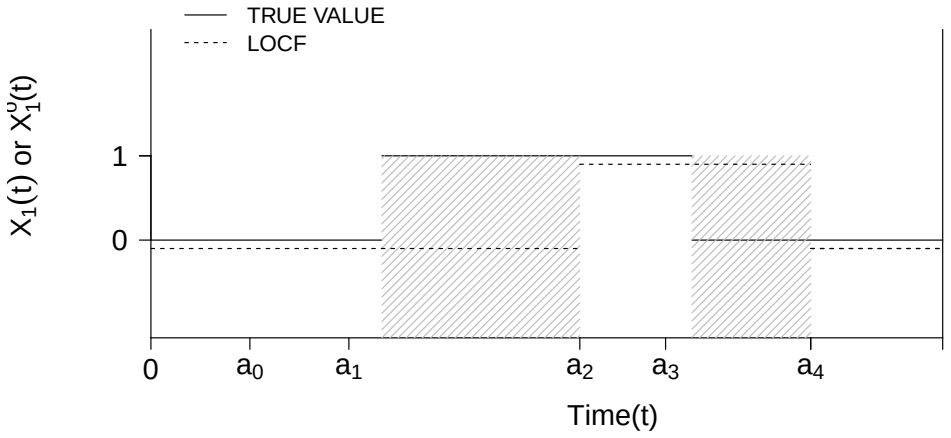


Figure 2.1: A schematic indicating the true value of a binary time-dependent marker, visit times, and the assumed value under last observation carried forward; shaded regions represent periods of time where the marker is misclassified. a_i represents the visit time on x-axis.

Figure 2.1 contains a timeline for a hypothetical individual with visits at a_0, a_1, a_2, a_3 and a_4. The true marker value $X_1(t)$ is denoted by a solid line whereas the marker value computed based on last-observation carried forward (LOCF) is denoted by $X_1^{\circ}(t)$ and a dashed line. Note that LOCF changes at visit times, while true value changes at non-visit

times. The shaded regions represent the periods of time where $X_1(t) \neq X_1^{\circ}(t)$ and hence the marker value is misclassified by LOCF.

Here we develop an expectation-maximization algorithm for suitably dealing with the intermittent observation process under specified model assumptions. This involves a parametric component for modeling the dynamics of the marker process under intermittent observation, and a weakly parametric proportional hazards regression model for the failure time given the marker process. We present it first with a piecewise-constant baseline hazard function and discuss extensions accommodating a truly semiparametric model in Appendix A. The semiparametric implementation of the expectation-maximization algorithm can be arbitrarily well approximated in the piecewise-constant formulation by increasing the number of pieces for the failure time baseline hazard so we focus on piecewise-constant primarily. We investigate the performance of the EM algorithm by simulation and illustrate it via application to data from a study of patients with skeletal metastases in which the goal is to evaluate the effect of bone markers on the risk of the composite endpoint of skeletal complication or death. We end with a discussion of topics of future research along these lines including problems involving marker processes with more than two categories, higher dimensional discrete marker processes, and interval-censored failure times.

In Section 2.2 a joint model is described for the marker and failure process along with the likelihood constructions one would use if individual processes were only subject to right censoring. In Section 2.3 a random visit process is introduced to generate the clinic visit times at which marker states are recorded. Section 2.4 then develops an EM algorithm for the case of a piecewise-constant hazard, and simulation results of the EM algorithm are given in Section 2.5. An application on bone marker study is discussed in Section 2.6 and concluding remarks are presented in Section 2.7. Appendix A covers details on EM on a semi-parametric Cox regression model are outlined and Appendix B shows steps on computation of the transition probability matrix.

2.2 A Joint Marker-Failure Time Model

2.2.1 The Model for the Marker Process

We consider a binary marker process which arises frequently in medical applications when physicians classify a marker level as normal or abnormal. When viewed jointly with the failure process we can consider a three state model described in Figure 2.2.

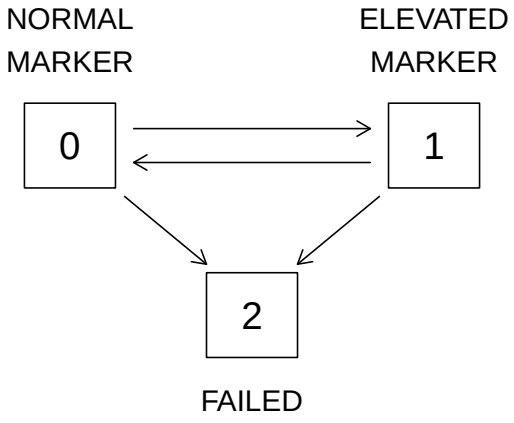


Figure 2.2: A three state process for joint modeling of a binary marker and a failure time

Let $\{Z(s), 0 < s\}$ denote the multistate process for individual i with the failure time T_i, the time individual i enters the absorbing state 2; states 0 and 1 are transient states. We let C_i denote an independent random censoring time.

For the marker process, we let $dN_{ij}(s) = 1$ if a $j \to 1-j$ transition occurs at time s and $dN_{ij}(s) = 0$ otherwise, $j = 0, 1$. Let $Y_{ij}(s) = I(Z_i(s^-) = j)$ and $\bar{Y}_{ij}(s) = Y_i(s)\, Y_{ij}(s)$, $j = 0, 1$, where $Y_i(s) = I(s \leq C_i)$ indicates individual i is under observation at time s. Suppose X_{i2} is a p$\times$1 vector of fixed covariates, we then let

$$\lim_{\Delta t \downarrow 0} \frac{P(\Delta N_{ij}(t) = 1 \mid Z_i(t^-) = j, X_{i2})}{\Delta t} = \lambda_j(t \mid X_{i2}) = \lambda_j(t)\, \exp(X'_{i2}\gamma_j)\,, \qquad (2.1)$$

be a Markov transition rate where γ_j is a $p \times 1$ vector of regression coefficients. These transition rates thus accommodate an association between the marker process and the $p \times 1$

vector of fixed covariates X_{i2} directly through $\gamma = (\gamma_0', \gamma_1')'$. We also let $\Lambda_j(t) = \int_0^t \lambda_j(s)\, ds$ be the cumulative baseline intensity and $d\Lambda_j(u) = \lambda_j(u)\, du$, $j = 0, 1$.

If the marker process was under complete observation other than being subject to termination (by T) and right censoring (by C), the partial log likelihood for the parameters governing the $j \to 1-j$ marker transitions based on a sample of n independent individuals would be

$$\ell_j = \sum_{i=1}^{n} \int_0^{\infty} \bar{Y}_{ij}(s) \left\{ dN_{ij}(s)\, \log d\Lambda_j(s \mid X_{i2}) - d\Lambda_j(s \mid X_{i2}) \right\}, \; j = 0, 1\,, \tag{2.2}$$

(Cook and Lawless, 2018). We consider a piecewise constant transition rate from state j to state $1-j$ with $K_j - 1$ finite cut-points at $0 = b_{j0} < b_{j1} < \cdots < b_{j,K_j-1} < b_{jK_j} = \infty$ defining intervals $\mathcal{B}_{jk} = [b_{j,k-1}, b_{jk})$, $k = 1, \ldots, K_j$, $j = 0, 1$. We let the baseline rate be given by $d\Lambda_j(s) = e^{\alpha_{jk}}$ if $s \in \mathcal{B}_{jk}$ and write a vector of K_j estimating functions $U_{jk} = \sum_{i=1}^{n} U_{ijk}$ with kth element for α_{jk} given by

$$\begin{aligned} U_{jk} = \frac{\partial \ell_j}{\partial \alpha_{jk}} &= \sum_{i=1}^{n} \int_0^{\infty} \bar{Y}_{ij}(s) \left[dN_{ij}(s) - e^{\alpha_{jk} + X_{i2}'\gamma_j} \right] I(s \in \mathcal{B}_{jk})\,, \; k = 1, \ldots, K_j \\ &= \sum_{i=1}^{n} \left[N_{ijk} - e^{\alpha_{jk} + X_{i2}'\gamma_j}\, S_{ijk} \right] \end{aligned} \tag{2.3}$$

where

$$N_{ijk} = \int_0^{\infty} \bar{Y}_{ij}(s)\, I(s \in \mathcal{B}_{jk})\, dN_{ij}(s)$$

is the total number of $j \to 1-j$ transitions for individual i in interval $\mathcal{B}_{jk}$ and

$$S_{ijk} = \int_0^{\infty} \bar{Y}_{ij}(s)\, I(s \in \mathcal{B}_{jk})\, ds$$

is the total time at risk for a $j \to 1-j$ transition in interval $\mathcal{B}_{jk}$ for individual i.

In order to estimate γ_j in (2.1), we also have the $p \times 1$ vector of estimating functions

$$U_{j,K_j+1} = \frac{\partial \ell_j}{\partial \gamma_j} = \sum_{i=1}^{n} \sum_{k=1}^{K_j} \int_0^{\infty} \bar{Y}_{ij}(u)\, I(u \in \mathcal{B}_{jk}) \left[dN_{ij}(u) - e^{\alpha_{jk} + X_{i2}'\gamma_j} \right] X_{i2}\,. \tag{2.4}$$

Note that from (2.3) solving $U_{jk} = 0$ gives the profile estimator

$$\exp(\tilde{\alpha}_{jk}(\gamma_j)) = \frac{\sum_{i=1}^{n} N_{ijk}}{\sum_{i=1}^{n} S_{ijk}\, e^{X_{i2}'\gamma_j}}\,,\ j = 0, 1,$$

and substituting the above for $\exp(\alpha_{jk})$ in (2.4) gives the $p \times 1$ profile score function

$$\begin{aligned} U_j(\gamma_j) &= \sum_{i=1}^{n}\sum_{k=1}^{K_j} \int_0^{\infty} \bar{Y}_{ij}(u)\, I(u \in \mathcal{B}_{jk}) \left\{ dN_{ij}(u) - \frac{\sum_{i=1}^{n} N_{ijk}}{\sum_{i=1}^{n} S_{ijk}\, e^{X_{i2}'\gamma_j}}\, e^{X_{i2}'\gamma_j} \right\} X_{i2} \\ &= \sum_{i=1}^{n}\sum_{k=1}^{K_j} \int_0^{\infty} \bar{Y}_{ij}(u)\, I(u \in \mathcal{B}_{jk}) \left\{ X_{i2} - \frac{\sum_{i=1}^{n} X_{i2}\, S_{ijk}\, e^{X_{i2}'\gamma_j}}{\sum_{i=1}^{n} S_{ijk}\, e^{X_{i2}'\gamma_j}} \right\} dN_{ij}(u) \\ &= \sum_{i=1}^{n}\sum_{k=1}^{K_j} N_{ijk} \left\{ X_{i2} - \frac{R_{jk}^{(1)}(\gamma_j)}{R_{jk}^{(0)}(\gamma_j)} \right\}, \end{aligned}$$

with

$$R_{jk}^{(r)}(\gamma_j) = \sum_{i=1}^{n} X_{i2}^{(r)}\, S_{ijk}\, e^{X_{i2}'\gamma_j}\,,\ r = 0, 1, 2\,, j = 0, 1,$$

where $X^{(0)} = 1$, $X^{(1)} = X$ and $X^{(2)} = X\, X'$. This can be solved for γ_j to obtain $\widehat{\gamma}_j$ with which we estimate α_{jk} as

$$\widehat{\alpha}_{jk} = \log(\sum_{i=1}^{n} N_{ijk} / \sum_{i=1}^{n} S_{ijk} e^{X_{i2}'\widehat{\gamma}_j})$$

2.2.2 The Failure Time Model

For the failure time process we wish to study the effect of the marker and the fixed covariates and we let $X_{i1}(t) = I(Z_i(t^-) = 1)$ and define $X_i(t) = (X_{i1}(t), X_{i2}')'$ to be the time dependent covariate vector comprised of the marker $X_{i1}(t)$ and the fixed covariate X_{i2}. We continue to consider the case in which the marker is under continuous observation, and here we construct the likelihood for the failure process parameters. The intensity for the failure time T_i is

$$\lim_{\Delta t \downarrow 0} \frac{P(t \le T_i < t + \Delta t \mid t \le T_i, X_i(t))}{\Delta t} = \lambda_2(t \mid X_i(t)) \tag{2.5}$$

which is taken to have the multiplicative form

$$\lambda_2(t \mid X_i(t)) = \lambda_2(t)\, \exp(X_i'(t)\beta) = \lambda_2(t)\, \exp(X_{i1}(t)\beta_1 + X_{i2}'\beta_2)\,,$$

where $\beta = (\beta_1, \beta_2')'$. Note that here we use subscript 2 to denote that the intensity is characterizing the risk of transition into state 2.

Let C_i denote an independent right-censoring time for individual i and $V_i = \min(T_i, C_i)$ with $\delta_i = I(V_i = T_i)$. Here we consider a baseline hazard of a piecewise-constant form and let $0 = b_{20} < b_{21} < \cdots < b_{2,K_2-1} < b_{2K_2} = \infty$ be the finite $K_2 - 1$ cut-points for the failure time hazard where the subscript "2" again reflects the fact that this is an intensity for entry into state 2. We let $\mathcal{B}_{2k} = [b_{2,k-1}, b_{2k})$ denote the kth interval and set

$$\lambda_2(t) = \lambda_{2k} = e^{\alpha_{2k}} \quad \text{if } t \in \mathcal{B}_{2k}\,,\ k = 1, \ldots, K_2\,.$$

Note that we can write

$$\lambda_2(t) = \prod_{k=1}^{K_2} \lambda_{2k}^{I(t \in \mathcal{B}_{2k})} = \prod_{k=1}^{K_2} \exp\left(\alpha_{2k}\, I(t \in \mathcal{B}_{2k})\right) = \exp\left(\sum_{k=1}^{K_2} \alpha_{2k}\, I(t \in \mathcal{B}_{2k})\right)\,, \tag{2.6}$$

or

$$\lambda_2(t) = \sum_{k=1}^{K_2} e^{\alpha_{2k}}\, I(t \in \mathcal{B}_{2k})\,. \tag{2.7}$$

Let $Y_i^\dagger(t) = I(t \le T_i)$ and $\bar{Y}_i(t) = Y_i(t)Y_i^\dagger(t)$. The partial log likelihood for the failure process can be written in the notation of counting process as

$$\begin{aligned}
\ell_2 &= \sum_{i=1}^{n} \int_0^\infty \left\{\bar{Y}_i(u)\, dN_{i2}(u)\, \log d\Lambda_2(u \mid X_i(u)) - \bar{Y}_i(u)\, d\Lambda_2(u \mid X_i(u))\right\} \\
&= \sum_{i=1}^{n} \int_0^\infty \bar{Y}_i(u)\, \left\{dN_{i2}(u)\, \left[\log d\Lambda_2(u) + X_i'(u)\,\beta\right] - d\Lambda_2(u)\, e^{X_i'(u)\,\beta}\right\}\,,
\end{aligned} \tag{2.8}$$

where $dN_{i2}(s) = I(T_i = s)$ indicates failure at time s, $s > 0$.

Since $d\Lambda_2(u)/du = \prod_{k=1}^{K_2} e^{\alpha_{2k}\, I(u\in\mathcal{B}_{2k})}$ by (2.6) and $d\Lambda_2(u)/du = \sum_{k=1}^{K_2} e^{\alpha_{2k}}\, I(u \in \mathcal{B}_{2k})$ by (2.7), we have

$$\begin{aligned}\ell_2 &= \sum_{i=1}^{n} \int_0^{\infty} \bar{Y}_i(u) \left\{ dN_{i2}(u) \left[\sum_{k=1}^{K_2} \alpha_{2k}\, I(u \in \mathcal{B}_{2k}) + X_i'(u)\,\beta \right] - \sum_{k=1}^{K_2} e^{\alpha_{2k}}\, I(u \in \mathcal{B}_{2k})\, e^{X_i'(u)\,\beta}\, du \right\} \\ &= \sum_{i=1}^{n} \sum_{k=1}^{K_2} \int_0^{\infty} \bar{Y}_i(u) \left\{ dN_{i2}(u) \left[\alpha_{2k} + X_i'(u)\,\beta \right] I(u \in \mathcal{B}_{2k}) - e^{\alpha_{2k}+X_i'(u)\,\beta}\, I(u \in \mathcal{B}_{2k})\, du \right\}. \end{aligned} \tag{2.9}$$

The estimating function for α_{2k} is of the form

$$\begin{aligned} U_{2k} = \frac{\partial \ell_2}{\partial \alpha_{2k}} &= \sum_{i=1}^{n} \int_0^{\infty} \bar{Y}_i(u) \left[dN_{i2}(u) - e^{\alpha_{2k}+X_i'(u)\,\beta}\, du \right] I(u \in \mathcal{B}_{2k}) \\ &= \sum_{i=1}^{n} \int_0^{\infty} d\bar{N}_{i2}(u)\, I(u \in \mathcal{B}_{2k}) - e^{\alpha_{2k}+X_{i2}\beta_2} \left[\mathcal{S}_{i0k} + \mathcal{S}_{i1k}\, e^{\beta_1} \right] \\ &= \sum_{i=1}^{n} \left\{ \delta_i\, I(u \in \mathcal{B}_{2k}) - e^{\alpha_{2k}+X_{i2}\beta_2} \left[\mathcal{S}_{i0k} + \mathcal{S}_{i1k}\, e^{\beta_1} \right] \right\}. \end{aligned} \tag{2.10}$$

where $\mathcal{S}_{ijk} = \int_0^{\infty} \bar{Y}_i(s)\, I(s \in \mathcal{B}_{2k})\; I(Z_i(s) = j)\, ds$.

Likewise, differentiating (2.9) with respect to β gives the $(p+1) \times 1$ score equation

$$\begin{aligned} U_{2,K_2+1} &= \sum_{i=1}^{n} \sum_{k=1}^{K_2} \int_0^{\infty} \bar{Y}_i(u) \left[dN_{i2}(u)\, I(u \in \mathcal{B}_{2k})\, X_i(u) - e^{\alpha_{2k}+X_i'(u)\,\beta}\, X_i(u)\, I(u \in \mathcal{B}_{2k}) \right] \\ &= \sum_{i=1}^{n} \sum_{k=1}^{K_2} \int_0^{\infty} \bar{Y}_i(u)\, I(u \in \mathcal{B}_{2k}) \left[dN_{i2}(u) - e^{\alpha_{2k}+X_i'(u)\,\beta} \right] X_i(u). \end{aligned} \tag{2.11}$$

Setting U_{2k} in (2.10) equal to zero gives

$$\exp(\tilde{\alpha}_{2k}(\beta)) = \frac{\sum_{i=1}^{n} \int_0^{\infty} \bar{Y}_i(u)\, I(u \in \mathcal{B}_{2k})\, dN_{i2}(u)}{\sum_{i=1}^{n} \int_0^{\infty} \bar{Y}_i(u)\, I(u \in \mathcal{B}_{2k})\, e^{X_i'(u)\,\beta}\, du} \tag{2.12}$$

Substituting (2.12) into (2.11) gives the profile score function for β:

$$U_2(\beta) = \sum_{i=1}^{n} \sum_{k=1}^{K_2} \int \bar{Y}_i(u)\, I(u \in \mathcal{B}_{2k}) \left[X_i(u) - \frac{R_{2k}^{(1)}(\beta)}{R_{2k}^{(0)}(\beta)} \right] dN_{i2}(u) \tag{2.13}$$

where

$$R_{2k}^{(r)}(\beta) = \sum_{i=1}^{n} \int_0^{\infty} \bar{Y}_i(u)\, I(u \in \mathcal{B}_{2k})\, X_k^{(r)}(u)\, \exp(X_i'(u)\beta)\, du\,,\ r = 0, 1.$$

If we consider the entry of (2.13) related to β_1 we have

$$U_{21}(\beta) = \sum_{i=1}^{n} \sum_{k=1}^{K_2} \int \bar{Y}_i(u)\, I(u \in \mathcal{B}_{2k}) \left[X_{i1}(u) - \frac{R_{2k1}^{(1)}(\beta)}{R_{2k}^{(0)}(\beta)} \right] dN_{i2}(u) \tag{2.14}$$

where

$$R_{2k1}^{(1)}(\beta) = \sum_{i=1}^{n} \int_0^{\infty} \bar{Y}_i(u)\, I(u \in \mathcal{B}_{2k})\, X_{i1}(u)\, e^{X_i'(u)\beta}\, du = \sum_{i=1}^{n} \mathcal{S}_{i1k}\, e^{\beta_1} e^{X_{i2}'\beta_2}$$

and

$$R_{2k}^{(0)}(\beta) = \sum_{i=1}^{n} \int_0^{\infty} \bar{Y}_i(u)\, I(u \in \mathcal{B}_{2k})\, e^{X_i'(u)\beta}\, du = \sum_{i=1}^{n} [\mathcal{S}_{i0k} + \mathcal{S}_{i1k} e^{\beta_1}]\, e^{X_{i2}'\beta_2}\,.$$

If we consider the entry related to β_2 we have the p×1 equation

$$U_{22}(\beta) = \sum_{i=1}^{n} \sum_{k=1}^{K_2} \int \bar{Y}_i(u)\, I(u \in \mathcal{B}_{2k}) \left[X_{i2} - \frac{R_{2k2}^{(0)}(\beta)}{R_{2k}^{(0)}(\beta)} \right] dN_{i2}(u) \tag{2.15}$$

where $R_{2k2}^{(1)}(\beta) = \sum_{i=1}^{n} X_{i2}\, [\mathcal{S}_{i0k} + \mathcal{S}_{i1k}\, e^{\beta_1}]\, e^{X_{i2}'\beta_2}$. Solving (2.14) and (2.15) for $\hat{\beta}_1$ and $\hat{\beta}_2$ respectively enables estimation of α_{2k} as

$$\widehat{\alpha}_{2k} = \log\left(\frac{\sum_{i=1}^{n} \int_0^{\infty} \bar{Y}_i(u) I(u \in \mathcal{B}_{2k})\, dN_{i2}(u)}{\sum_{i=1}^{n} \int_0^{\infty} \bar{Y}_i(u) I(u \in \mathcal{B}_{2k})\, \exp(X_i'(u)\hat{\beta}) du} \right)$$

These score equations were constructed for the setting where the marker process is under complete observation over $[0, \min(T, C)]$, but it will typically only be under intermittent observation. We turn to this problem in the next section, beginning with the description of a random visit process.

2.3 Data Under A Random Marker Visit Process

We now consider the case where the marker process $\{X_{i1}(s), 0 < s\}$ is under intermittent observation, where the observation times may correspond to, for example, clinic visits. We let A_{ir} denote the random time of the rth visit for individual i with a_{ir} being its realized value. The number of visits realized over $[0, V_i)$ is random and we let R_i denote the number made for individual i with r_i being the realization with the times $0 = a_{i0} < a_{i1} < \ldots < a_{iR_i} < V_i = \min(T_i, C_i)$. More generally, let $A_i(s) = \sum_{j=1}^{\infty} \bar{Y}_i(s)\, I(a_{ij} \leq s)$ denote the number of visits over $(0, s]$ for individual i, $\Delta A_i(s) = A_i(s + \Delta s^-) - A_i(s^-)$, and $dA_i(s) = \lim_{\Delta s \to 0} \Delta A_i(s)$ indicate if there is a visit at time s with $dA_i(s) = 0$ otherwise. We let $\{A_i(s), 0 < s\}$ be the counting process for visits and let

$$\bar{\mathcal{H}}_i(s) = \{\bar{Y}_i(u), dA_i(u), dN_i(u), 0 < u < s, Z_i(0)\}$$

denote the history of the joint process including the censoring, visit and multistate process.

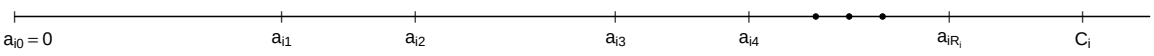


Figure 2.3: Timeline of visits and right-censoring time for a hypothetical individual

We define

$$\lim_{\Delta t \downarrow 0} \frac{P(\Delta A_i(t) = 1 \,|\, \bar{Y}_i(t) = 1, \bar{\mathcal{H}}_i(t))}{\Delta t} = \lambda_i^a(t \,|\, \bar{\mathcal{H}}_i(t)) \tag{2.16}$$

as the visit process intensity (Cook and Lawless, 2019).

Observations for the marker process are only made at visits, and we let

$$X_{i1}^{\circ}(s) = X_{i1}(a_{iA_i(s^-)})$$

denote the most recently observed value that is conventionally used at time s in standard failure time analysis, and $X_i^{\circ}(a_{ij}) = (X_{i1}^{\circ}(a_{ij}), X_{i2}')'$ be the corresponding complete

vector. The observed marker process history for individual i is denoted by $\mathcal{X}_{i1}^{\circ}(s) = \{(a_{ij}, X_{i1}(a_{ij})), j = 0, 1, \ldots, A_i(s^-)\}$, where a_{ij} is the time of jth assessment. We let

$$\bar{\mathcal{H}}_i^{\circ}(t) = \{\bar{Y}_i(u),\, dA_i(u),\, d\bar{N}_{i2}(u),\, 0 < u < t,\, X_i^{\circ}(a_{ij}),\, j = 0, 1, \ldots, A_i(t^-)\}$$

denote the history of the observed data process which can equivalently be written as $\bar{\mathcal{H}}_i^{\circ}(t) = \{\bar{Y}_i(u),\, dA_i(u),\, dN_{i2}(u),\, X_i^{\circ}(u),\, 0 < u < t\}$.

A Conditionally Independent Marker-Driven Visit Process

A *conditionally independent marker driven visit process* is one for which

$$\lim_{\Delta t \downarrow 0} \frac{P(\Delta A_i(t) = 1 \,|\, \bar{Y}_i(t) = 1, \bar{\mathcal{H}}_i(t))}{\Delta t} = \lambda^a(t \,|\, \bar{\mathcal{H}}_i^{\circ}(t)),$$

where the dependence on the marker process only lies in the observable $\mathcal{X}_i^{\circ}(t)$. This visit process is said to be conditionally independent because when conditioned on the history of the observed data $\bar{\mathcal{H}}_i^{\circ}(a_{j-1})$, the states occupied and the values of covariate since assessment time a_{j-1} do not affect the time of the next visit (Cook and Lawless, 2018).

A *conditionally dependent visit process* is accommodated in (2.16) since the intensity may depend on the marker over $(a_{iA_i(t^-)}, t)$. We assume in this chapter the visit process is conditionally independent but explore the consequences of conditionally dependent visit processes in Chapter 3 and ways of addressing them in Chapter 4.

The log likelihoods and score functions of Section 2.2 were given in the setting where the marker process was under continuous observation. This continuous observation scheme corresponds to a complete data setting where we can conceptualize when markers are only available at clinic visits. Recall that in Section 2.2.1 we specified cut-points $\{b_{jk}, k = 0, 1, \ldots, K_j, j = 0, 1, 2\}$ defining regions $\mathcal{B}_{jk}$, $k = 1, \ldots, K_j$, $j = 0, 1, 2$, within which intensities are constant. We may also conceive of intervals $\mathcal{B}_{k_0 k_1 k_2} = \mathcal{B}_{0k_0} \cap \mathcal{B}_{1k_1} \cap \mathcal{B}_{2k_2}$ within which the entire process is time homogeneous. We revisit this shortly.

For the visit process we let $a_{i0} = 0 < a_{i1} < \cdots < a_{iR_i} < V_i$ denote the R_i assessment

times for individual i and $\mathcal{A}_{ir} = [a_{i,r-1}, a_{ir})$ denote the intervals between assessments, $r = 1, \ldots, R_i$, with $\mathcal{A}_{i,R_i+1} = [a_{iR_i}, V_i)$. These assessment times lead to the observations $\{(X_i(a_{ir}), a_{ir}), r = 1, \ldots, R_i\}$ for the marker process, and the observed data from individual i can be written as

$$D_i = \{(X_i(a_{ir}), a_{ir}), r = 1, \ldots, R_i, V_i, \delta_i\}\,.$$

We also let $r_i(s) = \sum_{r=1}^{r_i+1} r\, I(s \in \mathcal{A}_{ir})$ record the label of the inter-assessment interval containing time $s > 0$ for individual i, and let $\mathcal{A}_i(s) = [a_{i,r_i(s)-1}, a_{i,r_i(s)})$ denote the corresponding interval, where for $r_i(s) = r_i + 1$, we let $a_{i,r_i+1} = V_i$. We next let

$$D_i(u) = \{(a_{i,r_i(u)-1}, Z_i(a_{i,r_i(u)-1})), (a_{ir_i(u)}, Z_i(a_{ir_i(u)}), X_{i2})\}$$

if $r_i(u) \leq r_i$ and

$$D_i(u) = \{(a_{ir_i}, Z_i(a_{ir_i})), V_i, \delta_i, X_{i2}\}$$

if $r_i(u) = r_i + 1$. This $D_i(u)$ term is useful as it identifies data that are pertinent for the conditional expectation needed for the EM algorithm developed in Section 2.4.

2.4 The EM Algorithm for Cox Regression Model

2.4.1 Complete Data Score Equations

Based on the developments of Section 2.2, the full vector of estimating functions is given by $U = (U_0', U_1', U_2')$ where $U_j = (U_{j1}, \ldots, U_{jK_j}, U_{j,K_j+1}')'$ is given by concatenating (2.3) and (2.4) for $j = 0, 1$, and (2.10) and (2.11) for $j = 2$. Note that U is a $q \times 1$ vector of estimating function with $q = (K_0 + p) + (K_1 + p) + K_2 + (p + 1)$. Note if we write the parameters corresponding to the estimating functions as $\theta_j = (\alpha_j', \gamma_j')'$ for $j = 0, 1$ and $\theta_2 = (\alpha_2', \beta')'$ for $j = 2$, we have

$$U(\theta) = (U_0'(\theta_0), U_1'(\theta_1), U_2'(\theta_2))'\,.$$

Under the complete data observation scheme we set $U(\theta) = 0$, solve for $\widehat{\theta}$ to obtain the maximum likelihood estimate. Again under continuous observation we know the required quantities $(N_{ijk},\, S_{ijk})$, $k = 1, \ldots, K_j$, $j = 0, 1$ and $(N_{i2k},\, S_{ijk})$, $j = 0, 1$, $k = 1, \ldots, K_2$, but we next consider the case where the marker is observed intermittently.

2.4.2 Observed Data Scores for the Marker Process

For the expectation of the complete data score for α_{jk} in (2.3) consider the expectation for individual i,

$$\bar{U}_{ijk} = E(U_{ijk} \mid D_i) = E(N_{ijk} \mid D_i) - e^{\alpha_{jk} + X'_{i2}\gamma_j} E(S_{ijk} \mid D_i)\,,\ j = 0, 1 \tag{2.17}$$

Adding contributions of (2.17) from each individual and setting it to 0 solves for the profile estimator of the marker process,

$$\exp(\tilde{\alpha}_{jk}(\gamma_j)) = \frac{\sum_{i=1}^{n} E(N_{ijk} \mid D_i)}{\sum_{i=1}^{n} E(S_{ijk} \mid D_i) e^{X_{i2}\gamma_j}} \tag{2.18}$$

Likewise, taking the expectation of the contribution to (2.4) from individual i, the observed data score for γ_j is

$$\begin{aligned}
\bar{U}_{ij,K_j+1} =& E(U_{ij,K_j+1} \mid D_i) \\
=& E\left\{ \sum_{k=1}^{K_j} \int_0^{\infty} \bar{Y}_{ij}(u) I(u \in B_{jk}) \{dN_{ij}(u) - e^{\alpha_{jk} + X'_{i2}\gamma_j}\} X_{i2} \mid D_i \right\} \\
=& \sum_{k=1}^{K_j} [E\{N_{ijk} \mid D_i\} X_{i2} - e^{\alpha_{jk} + X'_{i2}\gamma_j} E\{S_{ijk} \mid D_i\} X_{i2}] \\
=& \sum_{k=1}^{K_j} [E\{N_{ijk} \mid D_i\} - e^{\alpha_{jk} + X'_{i2}\gamma_j} E\{S_{ijk} \mid D_i\}] X_{i2}\ ,
\end{aligned} \tag{2.19}$$

for $j = 0, 1$.

Inserting (2.18) into (2.19) and adding up all individuals gives

$$\bar{U}_{j,K_j+1} = \sum_{i=1}^{n} \{ \sum_{k=1}^{K_j} [E\{N_{ijk} \mid D_i\} - \frac{\sum_{i=1}^{n} E(N_{ijk} \mid D_i)}{\sum_{i\dagger=1}^{n} E(S_{ijk} \mid D_i)\, e^{X_{i2}\gamma_j}} e^{X_{i2}\gamma_j} E\{S_{ijk} \mid D_i\}] X_{i2} \} \tag{2.20}$$

Remarks on the computation of $E\{N_{ijk} \mid D_i\}$ and $E\{S_{ijk} \mid D_i\}$ are given shortly and in Appendix A, but briefly we note that $E\{N_{ijk} \mid D_i\}$ is given by

$$\int_{\mathcal{B}_{jk}} \bar{Y}_i(u) \frac{P(Z_i(u) = j \mid Z_i(a_{i,r_i(u)-1}), X_{i2}) e^{\alpha_{jk} + X'_{i2}\gamma_j}\, P(Z_i(a_{ir_i(u)}) \mid Z_i(u^-) = j, X_{i2})}{P(Z_i(a_{ir_i(u)}) \mid Z_i(a_{i,r_i(u)-1}), X_{i2})}\, du \tag{2.21}$$

Exploiting Software for Poisson Regression: The Marker Process

We consider piecewise constant marker process intensities which have connections with Poisson processes. In this case, we could write the intensity as

$$d\Lambda_j(s \mid X_{i2}) = \Big[\prod_{k=1}^{K_j} e^{\alpha_{jk} I(s \in \mathcal{B}_{jk})} \Big] e^{X'_{i2}\gamma_j} = e^{v'_{ij}(s)\alpha_j + X'_{i2}\gamma_j} \tag{2.22}$$

where $v'_{ij}(s) = (v_{ij1}(s), \ldots, v_{ijK_j}(s))'$, and $v_{ijk}(s) = I(s \in \mathcal{B}_{jk})$ is the indicator if time s falls in the kth interval. If the marker process is under complete observation, then the log-likelihood based on a sample of n individuals is given by

$$\begin{aligned} \ell_j &= \sum_{i=1}^{n} \int_0^{\infty} \bar{Y}_{ij}(s) \{\, dN_{ij}(s) [v'_{ij}(s)\alpha_j + X'_{i2}\gamma_j] - e^{v'_{ij}(s)\alpha_j + X'_{i2}\gamma_j} \,\} \\ &= \sum_{i=1}^{n} \sum_{k=1}^{K_j} \{\, N_{ijk} \,[\alpha_{jk} + X'_{i2}\gamma_j] - S_{ijk} \,[\, e^{\alpha_{jk} + X'_{i2}\gamma_j} \,] \,\} \end{aligned} \tag{2.23}$$

If under intermittent observation, we take the expectation of the log likelihood and get the expectation of the complete data log likelihood as

$$E\{\ell_j \mid D_i\} = \sum_{i=1}^{n} \sum_{k=1}^{K_j} \{\, E\{N_{ijk} \mid D_i\} \,[\alpha_{jk} + X'_{i2}\gamma_j] - E\{S_{ijk} \mid D_i\} \,[\, e^{\alpha_{jk} + X'_{i2}\gamma_j} \,] \,\} \tag{2.24}$$

Aside. Notice that if we have a time homogeneous Poisson process with count N_i observed over an interval $[0, \tau_i)$ with covariate W_i acting on the rate function

$$\rho_i = \rho_0 e^{W_i'\Omega}$$

then the log likelihood for a sample of m independent individuals are

$$\begin{aligned}\ell &= \sum_{i=1}^{m} \{\, N_i(\alpha_0 + W_i'\Omega) - \exp(\alpha_0 + W_i'\Omega)\tau_i \,\} \\ &= \sum_{i=1}^{m} \{\, N_i(\alpha_0 + W_i'\Omega) - \exp(\alpha_0 + W_i'\Omega + \log\tau_i) \,\}\end{aligned}$$

Such a likelihood can be maximized by a call to any function for Poisson regression by specification of a log link (the canonical link), covariate vector W_i, and an offset of $\log\tau_i$.

Based on the remarks in the aside, in our present context, for a given individual, we construct data frame of the form

```
> data[data$id==1,]
  id        ni0k    logsi0k        ni1k    logsi1k  k x2
   1 0.006019806 -1.133993 0.007899272 -7.649370  1  1
   1 0.835741069 -1.695045 0.003743915 -1.975938  2  1
   1 0.006014236 -7.728171 0.013149845 -1.099934  3  1
```

where we take marker process with 2 cut-points as an example here (i.e. $K_j = 3$, $j = 0, 1$). For a given individual and each piece $k, k = 1, \ldots, K_j$, we calculate $E\{N_{ijk} \mid D_i; \theta^r\}$ and $E\{S_{ijk} \mid D_i; \theta^r\}$, $j = 0, 1$. For the pseudo data, at the (r+1)st maximization step we record the value of $E\{N_{i0k} \mid D_i; \theta^r\}$ as ni0k, and the value of $E\{N_{i1k} \mid D_i; \theta^r\}$ as ni1k. We calculate the offset term by taking the natural log of $E\{S_{i0k} \mid D_i; \theta^r\}$ and $E\{S_{i1k} \mid D_i; \theta^r\}$, and record them in columns logsi0k and logsi1k, respectively. Following the construction of this data-frame we maximize the expected log likelihood for the parameters of the $0 \to 1$ transition intensity by a call to the glm() function in R,

```
glm(ni0k ~ offset(logsi0k) + factor(k) + x2,
    family=poisson(link=log), data=data)
```

A similar call will yield updated estimates for the parameters of the $1 \to 0$ transition intensity.

2.4.3 Observed Data Scores for the Failure Process

For the failure process the incomplete data is the "covariate" $X_{i1}(s)$ so the observed data score equations obtained from the complete data score equations (2.10) and (2.11) have a different form from those of the previous section. First based on (2.10), we obtain the expected score

$$\begin{aligned}\bar{U}_{2k} &= E\{U_{2k} \mid D_i\} \\ &= \sum_{i=1}^{n} \{\delta_i I(V_i \in \mathcal{B}_{2k}) - e^{\alpha_{2k}+X_{i2}\beta_2}\,[\,E\{S_{i0k} \mid D_i\} + E\{S_{i1k} \mid D_i\}e^{\beta_1}\,]\,\}\,,\end{aligned} \tag{2.25}$$

Setting (2.25) to zero gives the profile estimator

$$\exp(\tilde{\alpha}_{2k}(\beta)) = \frac{\sum_{i=1}^{n} \bar{Y}_i(u)\delta_i I(V_i \in \mathcal{B}_{2k})}{\sum_{i=1}^{n} \bar{Y}_i(u)[\,E\{S_{i0k} \mid D_i\} + E\{S_{i1k} \mid D_i\}e^{\beta_1}\,]e^{X_{i2}\beta_2}}\,. \tag{2.26}$$

Plugging (2.26) into the observed data score equation based on (2.11), we have

$$\begin{aligned}&\bar{U}_{2,K_2+1}(\beta) = E\{U_{2,K_2+1} \mid D_i\} \\ =&E\left\{\sum_{i=1}^{n}\sum_{k=1}^{K_2}\int_0^{\infty} \bar{Y}_i(u)\, I(u \in \mathcal{B}_{2k})\, \left[dN_{i2}(u) - e^{\alpha_{2k}+X_i'(u)\,\beta}\right]\, X_i(u)\,|\,D_i\right\} \\ =&\sum_{i=1}^{n}\sum_{k=1}^{K_2}\int_0^{\infty} \bar{Y}_i(u)\, I(u \in \mathcal{B}_{2k})\, dN_{i2}(u) W(u \mid D_i)\end{aligned} \tag{2.27}$$

where $W(u \mid D_i) = \left[E\{X_i(u) \mid D_i\} - \frac{\sum_{i=1}^{n} \bar{Y}_i(u)E\{\,X_i(u)\,e^{X_{i1}(u)\beta_1}|D_i\,\}e^{X_{i2}\beta_2}}{\sum_{i=1}^{n} \bar{Y}_i(u)[\,\psi_{i0k}+\psi_{i1k}e^{\beta_1}\,]e^{X_{i2}\beta_2}}\right]$ and $\psi_{ijk} = E\{S_{ijk} \mid D_i\}$.

As $X_{i1}(u)$ is observed intermittently and X_{i2} is a fixed covariate, for a vector of covariate $X_i(u) = (X_{i1}(u), X_{i2})'$, if $\mathcal{A}_i = (a_{i,r_i(u)-1}, a_{i,r_i(u)})$ is the inter-visit interval spanning u, the

conditional expectation is given by

$$E\{X_i(u) \mid D_i\} = \begin{pmatrix} P(Z_{i1}(u) = 1 \mid D_i) \\ X_{i2} \end{pmatrix}$$

Similarly, $E\{\, X_i(u)\, e^{X_{i1}(u)\beta_1 + X_{i2}\beta_2} \mid D_i \,\}$ has the vector form

$$\begin{pmatrix} E\{\, X_{i1}(u)\, e^{X_{i1}(u)\beta_1 + X_{i2}\beta_2} \mid D_i \,\} \\ X_{i2}\, E\{\, e^{X_{i1}(u)\beta_1 + X_{i2}\beta_2} \mid D_i \,\} \end{pmatrix}$$

which can be written as

$$\begin{pmatrix} e^{\beta_1 + X_{i2}\beta_2}\, P(Z_{i1}(u) = 1 \mid D_i) \\ X_{i2}\ e^{X_{i2}\beta_2}\, [P(Z_{i1}(u) = 0 \mid D_i) + e^{\beta_1} P(Z_{i1}(u) = 1 \mid D_i)] \end{pmatrix}$$

where in both cases,

$$P(Z_{i1}(u) = 1 \mid D_i) = \frac{P(Z_i(u) = 1 \mid Z(a_{i,r_i(u)-1}), X_{i2})\, P(Z(a_{i,r_i(u)}) \mid Z_i(u) = 1), X_{i2})}{P(Z(a_{i,r_i(u)}) \mid Z(a_{i,r_i(u)-1}), X_{i2})} \tag{2.28}$$

Note that we need $E\{\mathcal{S}_{ijk} \mid D_i\}$ which has the form

$$\begin{aligned} &\int_0^\infty I(u \in \mathcal{B}_{jk})\, \bar{Y}_i(u)\, P(Z_i(u^-) = j \mid D_i)\, du \qquad (2.29) \\ &= \int_{\mathcal{B}_{jk}} \bar{Y}_i(u)\, \frac{P(Z_i(u) = j \mid Z_i(a_{i,r_i(u)-1}), X_{i2})\, P(Z_i(a_{ir_i(u)}) \mid Z_i(u^-) = j, X_{i2})}{P(Z_i(a_{ir_i(u)}) \mid Z_i(a_{i,r_i(u)-1}), X_{i2})}\, du\,. \end{aligned}$$

for $j = 0, 1$. If $b_{jk} < a_{i\ell}$ for some $\ell \le r_i$ then the marker state is known at some point after interval $\mathcal{B}_{jk}$. In this case

$$E\{\mathcal{S}_{ijk} \mid D_i\} = \int_0^\infty \sum_{r_i(u)=1}^{r_i+1} I(u \in \mathcal{B}_{jk})\, I(u \in \mathcal{A}_i(u))\, \bar{Y}_i(u)\, P(Z_i(u) = j \mid D_i(u), r_i(u))\, du\,.$$

If $r_i(u) \leq r_i$, then $P(Z_i(u) = j \mid D_i(u), r_i(u))$ is given by

$$\frac{P(Z_i(u) = j \mid Z_i(a_{i,r_i(u)-1}), X_{i2})\, P(Z_i(a_{ir_i(u)}) \mid Z_i(u^-) = j, X_{i2})}{P(Z_i(a_{ir_i(u)}) \mid Z_i(a_{i,r_i(u)-1}), X_{i2})}$$

and if $r_i(u) = r_i + 1$ we have

$$\begin{aligned}
&P(Z_i(u) = j \mid D_i(u), r_i(u) = r_i + 1) \\
&= \frac{\sum_{\ell=0}^{1} P(Z_i(u) = j \mid Z_i(a_{ir_i}), X_{i2})\, P(Z_i(V_i^-) = \ell \mid Z_i(u) = j, X_{i2})\, [\lambda_0(V_i)\, e^{\ell\beta_1 + X_{i2}'\beta_2}]^{\delta_i}}{\sum_{\ell=0}^{1} P(Z_i(V_i^-) = \ell \mid Z_i(a_{ir_i}), X_{i2})\, [\lambda_0(V_i)\, e^{\ell\beta_1 + X_{i2}'\beta_2}]^{\delta_i}}\,.
\end{aligned}$$

Second we need

$$P(Z_i(V_i^-) = 1 \mid D_i) = \frac{P(Z_i(V_i^-) = 1 \mid Z_i(a_{ir_i}), X_{i2})\, \lambda_2(V_i)\, e^{\beta_1 + X_{i2}'\beta_2}}{\sum_{j=0}^{1} P(Z_i(V_i^-) = j \mid Z_i(a_{ir_i}), X_{i2})\, \lambda_2(V_i)\, e^{j\beta_1 + X_{i2}'\beta_2}}\,. \tag{2.30}$$

Exploiting Software for Poisson Regression: The Failure Process

For the failure process, we could write the transition intensity in the context of Poisson process as

$$d\Lambda_2(s \mid X_{i2}) = e^{v_{i2}'(s)\alpha_2 + X_{i1}(s)\beta_1 + X_{i2}\beta_2} \tag{2.31}$$

where $v_{i2}'(s) = (v_{i21}(s), \ldots, v_{i2K_2}(s))'$, and $v_{i2k}(s) = I(s \in \mathcal{B}_{2k})$ is the indicator if failure time s falls in the kth interval, $k = 1, \ldots, K_2$. If the marker process is under complete observation, then the log-likelihood based on a sample of n individuals is given by

$$\begin{aligned}
\ell_2 =& \sum_{i=1}^{n} \int_0^{\infty} \bar{Y}_i(u) \{\, dN_{i2}(u)[v_{i2}'(s)\alpha_2 + X_{i1}(s)\beta_1 + X_{i2}\beta_2 - e^{v_{i2}'(s)\alpha_2 + X_{i1}(s)\beta_1 + X_{i2}\beta_2}] \,\} \qquad (2.32) \\
=& \sum_{i=1}^{n} \sum_{k=1}^{K_2} \{\, N_{i2k}\alpha_{2k} + N_{i2k}^{(1)}\beta_1 + N_{i2k}X_{i2}\beta_2 - S_{i2k}^{(0)}\, e^{\alpha_{2k} + X_{i2}\beta_2} - S_{i2k}^{(1)}\, e^{\alpha_{2k} + \beta_1 X_{i2}\beta_2} \,\} \\
=& \sum_{i=1}^{n} \sum_{k=1}^{K_2} \sum_{X_1=0}^{1} \{\, N_{i2k}^{(X_1)}[\alpha_{2k} + X_1\beta_1 + X_{i2}\beta_2] - S_{i2k}^{(X_1)}[\, e^{\alpha_{2k} + X_1\beta_1 + X_{i2}\beta_2}] \,\}
\end{aligned}$$

where $N_{i2k}^{(X_1)}$ is the number of failure times (0 or 1) from individual i in interval k from state X_1 (0 or 1) and $S_{i2k}^{(X_1)}$ is the total time at risk of failure in state X_1 (0 or 1) in interval

k of the partition for the failure hazard.

We generate a pseudo-dataset for a given individual as

> datai

i	k	x1	$N_{i2k}^{(X_1)}$	$\log S_{i2k}^{(X_1)}$	X_{i1}	X_2
1	1	0			0	X_{12}
1	2	0			0	X_{12}
				.....		
1	K_2	0			0	X_{12}
1	1	1			1	X_{12}
1	2	1			1	X_{12}
				.....		
1	K_2	1			1	X_{12}

where similarly to the marker process, if there are K_2 cut-points for the failure process, for a given individual and each piece $k, k = 1, \ldots, K_2$, for values of X_1, $X_1 = 0, 1$, we calculate $N_{i2k}^{(X_1)}(\theta^r)$ and $S_{i2k}^{(X_1)}(\theta^r)$ respectively. Then at the (r+1)st maximization step, we record the value $N_{i2k}^{(X_1)}(\theta^r)$ as $N_{i2k}^{(X_1)}$ and we calculate the natural log of $S_{i2k}^{(X_1)}(\theta^r)$ and record in column $\log S_{i2k}^{(X_1)}$ with the corresponding X_1 value. Similar to the marker process, we could calculate the parameters of the $0 \to 2$ transition intensity by a call to the glm() function in R, and fit the model with

```
glm( ni2k ~ offset( logsi2k ) + factor( k ) + xi1 + x2 ,
     family=poisson( link=log ) , data=data )
```

For the case where $X_1(s)$ is unknown as markers are only measured periodically at clinic assessments, note we have $E\{S_{i2k}^{(1)} \mid D_i\}$ in the form

$$E\left\{\int_0^\infty I(u \in \mathcal{B}_{2k})\, I(Z_i(u) = 1)) \mid D_i\right\} = \int_0^\infty I(u \in \mathcal{B}_{2k}) P(Z_i(u) = 1 \mid D_i) du$$

Likewise, we have

$$E\{S_{i2k}^{(0)} \mid D_i\} = \int_0^\infty I(u \in \mathcal{B}_{2k}) P(Z_i(u) = 0 \mid D_i) du$$

And if failure does not occur in interval k, then

$$E\{N_{i2k}^{(X_1)} \mid D_i\} = 0$$

Otherwise,

$$E\{N_{i2k}^{(X_1)} \mid D_i\} = P(\, Z_i(t_{i2}^-) = X_1 \mid D_i, d\bar{N}_{i2}(t_{i2}) = 1\,) \tag{2.33}$$

where t_{i2} is the observed failure time for individual i.

Under intermittent observation, N_{i2k} is known, but $N_{i2k}^{(X_1)}$ and $S_{i2k}^{(X_1)}$ are unknown. The call of the function is the same as if the dataset is under complete observation, but it is done in the EM algorithm based on a pseudo-dataset where the conditional expectations are estimated based on the parameter estimate at the previous iteration.

Defining the Finest Partition

Note these expectations all require computation of the transition probability matrix for the process $\{Z_i(s), 0 < s < C_i, X_{i2}, C_i\}$. This is achieved by product integration via

$$P(s,t \mid X_{i2}) = \prod_{(s,t]} \{\mathcal{I} + dQ(u \mid X_{i2})\}$$

where $\mathcal{I}$ is a 3×3 identity matrix and $dQ(s \mid X_{i2})$ is a transition intensity matrix.

To facilitate the calculations necessary for computation of the conditional expectations it is helpful to more fully exploit the simplifications that come from the piecewise constant intensities. Specifically we let $\mathcal{B}_k$ represent the partition of the positive real line defined by the union of all cut-points forming the partitions for $0 \to 1$, $1 \to 0$ and $0 \to 2$ transitions. Thus if $\mathcal{S}_j = \{b_{jr},\ r = 1, \ldots, K_j - 1\}$ is the set of non-zero and finite cut-points for $0 \to 1$

transition ($j = 0$), $1 \to 0$ transition ($j = 1$), and $0 \to 2$ transition ($j = 2$), let

$$\mathcal{S} = \mathcal{S}_0 \cup \mathcal{S}_1 \cup \mathcal{S}_2$$

with elements labeled $0 < b_1 < \cdots < b_{K-1} < \infty$. In order to retain the information about the piece, a new sub-interval $\mathcal{B}_k$ falls in for a particular intensity, we let ℓ_{jk} denote the sub-interval for the partition for $d\Lambda_j(\cdot)$ within which $\mathcal{B}_k$ falls. That is

$$\mathcal{B}_{j\ell_{jk}} \cap \mathcal{B}_k = \mathcal{B}_k$$

but $\mathcal{B}_{j\ell} \cap \mathcal{B}_k$ for all other values of ℓ. The full marker-failure time process is then time homogeneous within each interval $\mathcal{B}_k$, $k = 1, \ldots, K$ where $\mathcal{B}_K = [b_{K-1}, \infty)$. Let $k(u)$ denote the interval $\mathcal{B}_k$ containing u, and let $d\Lambda_{0\ell_{0k}}(\cdot)$, $d\Lambda_{1\ell_{1k}}(\cdot)$ and $d\Lambda_{2\ell_{2k}}(\cdot)$ denote the baseline intensities in $\mathcal{B}_k$. We define the matrix $dQ_k(X_2)$ as

$$\begin{bmatrix} -(d\Lambda_{0\ell_{0k}}(s)e^{\gamma_0 X_2} + d\Lambda_{2\ell_{2k}}(s)e^{\beta_2 X_2}) & d\Lambda_{0\ell_{0k}}(s)e^{\gamma_0 X_2} & d\Lambda_{2\ell_{2k}}(s)e^{\beta_2 X_2} \\ d\Lambda_{1\ell_{1k}}(s)e^{\gamma_1 X_2} & -(d\Lambda_{1\ell_{1k}}(s)e^{\gamma_1 X_2} + d\Lambda_{2\ell_{2k}}(s)e^{\beta_1+\beta_2 X_2}) & d\Lambda_{2\ell_{2k}}(s)e^{\beta_1+\beta_2 X_2} \\ 0 & 0 & 0 \end{bmatrix}.$$

and $dQ^*_k(X_2) = dQ_k(X_2)/ds$ and note that if $k(s) = k(t)$ (i.e. we are carrying out a computation at times $s < t$ within the same interval where the process is time homogeneous), we can compute

$$P(s, t \mid X_2; \theta) = \exp(dQ^*_{k(s)}(X_2)(t - s)).$$

More generally,

$$P(s, t \mid X_2; \theta) = \prod_{h=k(s)+1}^{k(t)} \exp(dQ^*_h(X_2)(\min(t, b_{k(t)+1}) - b_{k-1})).$$

2.4.4 Variance Estimation

We estimate the observed information matrix using the approach of Louis (1982). If U_i is the contribution to the complete data score from individual i then $J_i(\theta) = -\partial U_i(\theta)/\partial\theta'$ is the complete data information matrix. The contribution to the observed information

matrix from individual i is then

$$\bar{J}_i(\theta) = E\{J_i(\theta)|D_i\} - E\{U_i(\theta)U_i'(\theta)|D_i\} + \bar{U}_i(\theta)\bar{U}_i'(\theta)\ . \tag{2.34}$$

where $\bar{U}_i(\theta) = E\{U_i(\theta)\,|\,D_i\}$. It may be preferable to consider variance estimation based on estimating function theory as done by Herring and Ibrahim (2001).

If $\mathcal{A}(\theta) = -E\{\partial\bar{U}_i(\theta)/\partial\theta'\}$ and $\mathcal{B}(\theta) = E\{\bar{U}_i(\theta)\bar{U}_i'(\theta)\}$ then

$$\sqrt{n}(\widehat{\theta} - \theta) \sim N(0, \mathcal{A}^{-1}(\theta)\mathcal{B}(\theta)\mathcal{A}^{-1}(\theta))$$

as $n \to \infty$.

In practice we replace the expected matrices by their empirical counterparts evaluated at the solution $\widehat{\theta}$, where $\widehat{\theta}$ is also the solution to the estimating equation $E\{U_i(\theta)\} = 0$, and we have

$$A(\widehat{\theta}) = -n^{-1}\sum_{i=1}^{n}\partial\bar{U}_i(\theta)/\partial\theta'\,|_{\theta=\widehat{\theta}}$$
$$B(\widehat{\theta}) = -n^{-1}\sum_{i=1}^{n}\bar{U}_i(\theta)\bar{U}_i'(\theta)\,|_{\theta=\widehat{\theta}}$$

Nonparametric bootstrapping (Efron and Tibshirani, 1994) can also be carried for variance estimation. In this case the life history paths observed for each individual are numbered according to the index i labeling individuals, $i = 1, \ldots, n$. We then sample from the integers 1 to n with replacement n times to create a bootstrap sample which is analysed to obtain an estimate $\widehat{\theta}^b$. This process is repeated B times to create B bootstrap estimates $\hat{\theta}^{(1)}, ...\hat{\theta}^{(B)}$. We then compute the bootstrap standard error as the square root of the empirical variance of the estimates.

2.5 Simulation Studies

2.5.1 Design of Simulation Studies

Let the administrative censoring time be $C^A = 1$. We set $P(Z(0) = 0) = 1$ where every individual starts with a normal marker value with probability 1. We assume that the elevated marker value increases the hazard for failure by 25% and we then solve for λ_2 so that the probability of death at the censoring time to be $P(Z(1) = 2 \,|\, Z(0)) = 0.4$ and 0.6 representing moderate and high mortality rates. We constrain the transition intensities for the multistate process to be time homogeneous. In figure 2.2, we let λ_0 to be the transition intensity from state 0 to state 1 and let λ_1 to be the transition intensity starting from state 1 back to state 0. We let λ_2 to be the intensity of entering the failure state from a normal marker state. In our simulation studies, we set $\lambda_1/\lambda_0 = 2$ and $\lambda_0/\lambda_2 = 0.75$. For the auxiliary covariate X_2, we consider a binary covariate where we let $P(X_2 = 1) = 0.5$. Let the log hazard ratio of $0 \to 1$ and $1 \to 0$ transitions be $\gamma_0 = \log 1.2$ and $\gamma_1 = \log 0.8$ respectively. We set the marker coefficient for the failure time model to be $\beta_2 = \log 1.5$.

For the visit process, we let $\{A(s), 0 < s\}$ to be a time-homogeneous Poisson process so let $\lambda^a(t \,|\, \bar{\mathcal{H}}(t)) = \rho(t) = \rho$. Let the average number of visits over $[0, C^A = 1)$ be $E(A(1)) = 8$ or 12 respectively. We generate gap time between each visit using an exponential distribution with rate ρ. We then keep all visit times that are less than C. Note that we consider both the case with no random censoring and the case with random censoring.

Let $F(t) = P(T > t) = P(Z(t) < 2 \,|\, Z(0) = 0)$ be the marginal survivor function. When we have independent random censoring we specify the random censoring rate as $P(C^R < T \,|\, T < C^A)$ which is the proportion of failures over $[0, C^A)$ that are not observed due to random censoring; this definition allows us to consider the two mortality rates in a similar fashion.

Specifically we find the hazard λ_c for an exponentially distributed censoring time such

that

$$P(C^R < T \,|\, T < C^A) = \frac{P(C^R < T < C^A)}{1 - \mathcal{F}(C^A)} = \frac{\int_0^{C^A} \lambda_c e^{-\lambda_c s} \mathcal{F}(s) ds}{1 - \mathcal{F}(C^A)}$$

is equal to 0.2 or 0.4 to represent mild or moderate right censoring.

We consider sample size of $n = 500$ and carry out nsim $= 500$ simulations for each parameter configuration. Analyses are carried out based on a misspecified Cox model using $X^\circ(s)$ in lieu of $X(s)$ as well as based on the proposed EM algorithm with piecewise-constant hazards having $p = 3$ or 10 pieces.

2.5.2 Simulation Results

Even though the primary interest lies in the regression coefficients of the failure process, we need to estimate the baseline hazard of the failure process and the transition rates and covariate effects for the marker process. Table 2.1 contains simulation results with 3 cut-points at failure process based on 500 individuals per simulation under 500 simulations. Table 2.2 shows results of 10 cut-points in $[0, 1]$ timeline for the failure process with a mean of 8 or 12 visits for the Poisson process. According to the simulation results, we see good performances of the EM algorithm for both the failure process and the marker process. We find that the empirical biases of the estimators arising from the joint analysis are small and that the standard errors are smaller for the regression coefficient β_1 when assessments are more frequent (i.e. compare the standard errors for the case when $\rho = 12$ compared to $\rho = 8$ in Tables 2.1 and 2.2). The transition rates and fixed covariate effect for the marker process are also well-estimated.

		$\rho = 8$		$\rho = 12$	
Parameter	TRUE	EBIAS	ESE	EBIAS	ESE
Failure Process					
β_1	0.2231	-0.0186	0.6484	0.0075	0.2120
β_2	0.4055	-0.0166	0.2683	0.0082	0.1334
λ_{21}	0.8768	-0.0116	0.2212	0.0179	0.0691
λ_{22}	0.8768	0.0089	0.2656	0.0120	0.0723
λ_{23}	0.8768	0.0323	0.3153	-0.0015	0.0759
Marker Process					
λ_{01}	0.6576	0.0507	0.1671	0.0069	0.0673
λ_{02}	0.6576	0.0328	0.2419	0.0054	0.0841
λ_{03}	0.6576	0.0569	0.2818	0.0081	0.0938
γ_0	0.1823	-0.0229	0.3247	-0.0091	0.2074
λ_{11}	1.3152	0.0741	0.7073	0.0082	0.5160
λ_{12}	1.3152	0.0662	0.7816	0.0095	0.3320
λ_{13}	1.3152	0.0382	0.7579	0.0050	0.2960
γ_1	-0.2231	0.0047	0.5626	-0.0091	0.2074

Table 2.1: Parameter estimates with EM algorithm based on glm function. Three cut-points for marker and failure process. Number of individuals per simulation =500, Number of simulations = 500. $P(Z(A) = 2 \,|\, Z(0) = 0) = 0.4$; $\rho = 8$ or 12 are the means of the Poisson visit process.

		$\rho = 8$		$\rho = 12$	
Parameter	TRUE	EBIAS	ESE	EBIAS	ESE
Failure Process					
β_1	0.2231	-0.0213	0.2164	0.0097	0.1796
β_2	0.4055	0.0066	0.1142	0.0152	0.1149
λ_{21}	0.8768	0.0110	0.1463	-0.0064	0.1332
λ_{22}	0.8768	-0.0122	0.1408	-0.0049	0.1531
λ_{23}	0.8768	-0.0099	0.1529	0.0018	0.1698
λ_{24}	0.8768	-0.0029	0.1460	-0.0055	0.1673
λ_{25}	0.8768	-0.0053	0.1852	-0.0051	0.1645
λ_{26}	0.8768	-0.0043	0.1660	-0.0037	0.1780
λ_{27}	0.8768	0.0409	0.1928	0.0453	0.2011
λ_{28}	0.8768	-0.0068	0.2198	-0.0227	0.1807
λ_{29}	0.8768	-0.0061	0.2019	0.0156	0.2148
λ_{210}	0.8768	0.0222	0.1819	-0.0251	0.1998
Marker Process					
λ_{01}	0.6576	-0.0082	0.0978	0.0178	0.1084
λ_{02}	0.6576	-0.0116	0.1415	0.0030	0.1160
λ_{03}	0.6576	-0.0108	0.1833	0.0099	0.1606
γ_0	0.1823	0.0156	0.1681	-0.0096	0.1783
λ_{11}	1.3152	0.0045	0.7401	0.0710	0.5946
λ_{12}	1.3152	0.0658	0.4595	0.0411	0.3936
λ_{13}	1.3152	0.0264	0.5283	0.0081	0.4077
γ_1	-0.2231	-0.0510	0.3444	0.0178	0.3731

Table 2.2: Parameter estimates with EM algorithm based on glm function. Three cut-points for marker and ten for failure. Number of individuals per simulation =500, Number of simulations = 500. $P(Z(A) = 2 \,|\, Z(0) = 0) = 0.6$; $\rho = 8$ or 12 are the means of the Poisson visit process.

2.6 Application

2.6.1 Application to a Bone Marker Study

The application involves the joint analysis of bone markers and risk of skeletal related events. To avoid the problem of the competing risk of death, a composite endpoint of skeletal related event-free survival was used for the failure time. In this study, we used bone markers N-telopeptide of type I collagen (Ntx), which is a marker of bone destruction (resorption) as the time-varying covariate. Ntx is measured periodically at clinic visits and conventional analyses carry forward the most recently recorded values until the next assessment or the occurrence of the failure or censoring times. We classify individuals who have higher than 64 nmol/mmol cr of Ntx as in the elevated marker state. The semiparametric EM algorithm was applied. In this application, we examined the effect of the time-varying Ntx marker and treatment effect on skeletal-related-events(SRE) or death. The maximum number of months being followed was 29 months. We set cut-points for marker process at 9 and 18 months and cut-points for failure process at 4, 8, 12, 16 and 20 months. The standard error is calculated based on the bootstrap method where we sample n individuals with replacement from the data to create the bootstrap sample of size n, and we then calculate the bootstrap standard error based on the square root of the empirical variance.

We estimated the effects of the marker and the treatment effect on the occurrence of skeletal-related-event (SRE) or death. Table 2.3 shows parameter estimates based on the EM algorithm and Table 2.4 shows results from the naive Cox regression model when time-varying covariate value is obtained by last-observations carried forward. Both tables include estimates, standard errors, hazard ratio and the p-values associated with chi-squared statistics for all parameters. Based on the results from Table 2.3, it can be concluded that Ntx marker has a significant effect on the failure time process (HR = 2.50, 95% CI: 1.37, 4.54; p=0.0026) and treatment also has a significant effect on the 1 to 0 marker transition (HR = 9.62, 95% CI: 1.92, 48.34; p=0.0060). Note that when analyses are directed at the effects of both time-varying markers Ntx and BALP, an expanded state-space will be

required to facilitate joint analysis of the markers Ntx with the failure time.

Parameter	EST.	S.E.	H.R.	95%C.I. for H.R.	p-value
Failure Process					
β_1	0.9166	0.3044	2.5008	(1.3771,4.5413)	0.0026
β_2	0.1733	0.1520	1.1892	(0.8828,1.6019)	0.2543
λ_{21}	0.0374	0.0082			
λ_{22}	0.0391	0.0072			
λ_{23}	0.0416	0.0103			
λ_{24}	0.0282	0.0069			
λ_{25}	0.0340	0.0099			
λ_{26}	0.0223	0.0072			
Marker Process					
λ_{01}	0.1781	0.0528			
λ_{02}	0.0753	0.0370			
λ_{03}	0.0946	0.1311			
γ_0	-0.9275	1.0472	0.3955	(0.0508, 3.0803)	0.3758
λ_{11}	0.0851	0.0235			
λ_{12}	0.0108	0.0063			
λ_{13}	0.0365	0.0855			
γ_1	2.2641	0.8236	9.6224	(1.9152, 48.3446)	0.0060

Table 2.3: Parameter estimates for the bone marker analysis. The cut-points for the marker process are at 9 and 18 months, and the cut-points for the failure process are at 4, 8, 12, 16, 20 months.

Parameter	EST.	S.E.	H.R.	95% C.I. for H.R.	p-value
Failure Process					
β_1	0.8970	0.1258	2.4522	(1.9163, 3.1379)	<0.001
β_2	0.1561	0.1292	1.1689	(0.9074, 1.5058)	0.2270

Table 2.4: Parameter estimates using LOCF based on a Cox regression model.

2.7 Concluding Remarks and Future Work

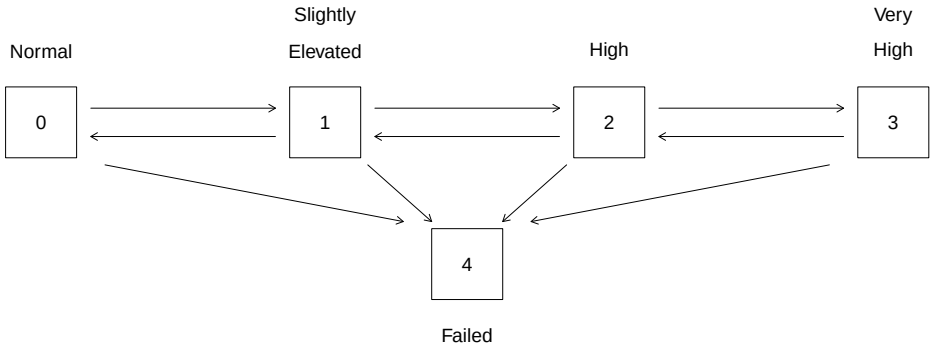
2.7.1 Concluding Remarks

In this chapter, we described an EM algorithm for fitting a joint model of the marker process and failure process when the marker process is under intermittent observation. The EM algorithm allows us to accommodate different numbers of cut-points for the failure time transition intensities, and as the number of cut-points increases, it better approximates the semiparametric Cox model for the failure process. We focus on the case when there is a binary time-dependent covariate and a binary fixed covariate in the simulations, but this algorithm can be easily extended to deal with any fixed covariate vector. Simulation studies are conducted and an application on the bone marker study in which we relate the time-varying bone marker Ntx and the fixed treatment effect to the occurrence of skeletal-event or death is discussed.

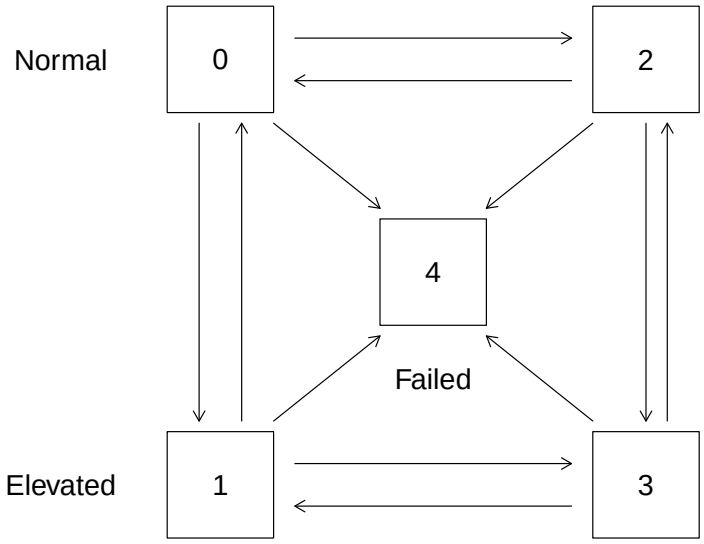
2.7.2 Future Work

The data from the University of Toronto Psoriatic Arthritis Registry yields intermittent observation of the erythrocyte sedimention rate (ESR), a marker if inflammation in autoimmune disease. This will be used as the marker of interest. The failure time in this setting, however, is typically based on the cumulative number of damaged joints and physician scientists are interest in how the ESR values are associated with the onset of additional joint damage. Joints are only radiographed periodically and so this endpoint is also subject to interval-censoring. Exension of the algorithm will be developed to accommodate an interval-censored failure time where the assessment times for failure are based on times of x-rays (and hence are different than the times the blood samples are taken to measure ESR).

If a marker involves more than two categories that can easily be accommodated. With four levels, for example, the multistate model may be expanded as in Figure 2.4a The algorithm for estimation can proceed in a very similar spirit to the way we have described it



(a) A joint marker and failure time process with four marker states



(b) A state space for joint analysis of two binary time-dependent biomarkers and failure

Figure 2.4: Multistate models for more general joint marker-failure time process

here, but more transition intensities need to be estimated between marker states. Likewise if there are two binary markers we may define the joint state-space as in Figure 2.4b.

Further extentions could be considered to deal with intermittent observation of the failure process. In some settings the failure process will be intermittently observed at different visit times than the times the blood samples, for example, may be taken to assess marker values. In this case a second visit process may be necessary to model. While this is challenging we are aided in this by the fact that visit times are observed exactly and the nature of the information needed at the visits will be known as well.

Chapter 3

Large Sample Properties Under a Marker-Dependent Visit Process

This chapter is in two parts. The first part is devoted to studying the asymptotic bias of estimators for regression coefficients in the Cox model of the time-varying marker and the fixed covariates when the "last observation carried forward" approach is used to deal with the intermittently observed binary time-dependent marker. We extend the preliminary investigation of Jiang et al. (2020) in two directions. First, as in Chapter 2, we introduce an auxiliary fixed baseline covariate which may, or may not be associated with the dynamic marker; the incorrect handling of the time-dependent marker will not only introduce a bias in the coefficient of that marker, but the coefficient of the fixed covariate will also be inconsistently estimated. We will explore how the asymptotic bias of the coefficients of both the time-dependent and fixed variables vary according to several factors including the strength of association between the two covariates and the frequency of the assessments. Second, we define a marker-dependent visit process and examine the extent to which the limiting bias of the regression coefficients depends on the strength of the association between the marker value itself and the visit intensity in an intensity-based model for the visit process. This builds on the work of Cook and Lawless (2019) who define a conditionally independent visit process for a multistate model. The distinction here is that the process of interest in their work is a failure process where the times are subject only to possibly dependent right

censoring, and the analysis approach is one in which the marker process is not modeled.

3.1 A Joint Marker-Visit-Failure Time Process

Markers are internal time-dependent covariates which reflect the health of an individual (Jewell and Kalbfleisch, 1996). Common markers include blood pressure and cholesterol level in cardiovascular disease, blood glucose in diabetes, and markers of bone destruction in osteoporosis. In many health research settings the goal is to study the relation between markers and risk of an event such as cardiac death, diabetic complications, or vertebral fracture in the previous illustrative examples respectively. We first, in addition to the binary marker, report on an extension of the investigation by Jiang et al. (2020) by considering a fixed baseline covariate arising from carrying forward the most recently recorded value. We then explore the impact of carrying forward the marker value on the coefficient of the markers and the fixed covariates. The binary time-dependent marker and the fixed covariate may be associated. We use the theory of misspecified estimation functions to determine the limiting bias of naive estimators (Struthers and Kalbfleisch, 1986).

In the three-state model of Figure 2.2 we let $X_{i1}(s) = I(Z_i(s^-) = 1)$ as before. We let $0 = a_{i0} < a_{i1} < \cdots$ denote the assessment times for individual i and $A_i(s) = \sum_{j=1}^{\infty} I(a_{ij} \leq s)$ count the number of post-baseline assessments taking place over $(0, s]$. We let $\{A_i(s), 0 < s\}$ denote the corresponding counting process. To distinguish the actual marker value from the most recently recorded value we let $X_{i1}^{\circ}(s) = X_{i1}(a_{iA_i(s^-)})$ denote the marker value recorded at the most recent visit prior to time s for individual i. In practice, interest may lie in the regression coefficient β in the Cox failure time model

$$\lambda_2(t \,|\, X_i(t)) = \lambda_2(t) \exp(X_i'(t)\beta) = \lambda_2(t) \exp(X_{i1}(t)\beta_1 + X_{i2}'\beta_2)\,, \tag{3.1}$$

where β_1 is the log hazard ratio characterising the effect of the true marker value $X_{i1}(s)$ given X_{i2}. Often, however, one fits the Cox model with $X_i^{\circ}(s) = (X_{i1}^{\circ}(s), X_{i2}')'$, following the convention of "last observation carried forward" for the time-dependent markers.

More explicitly we may specify

$$d\Gamma(s \mid X_i^\circ(s)) = d\Gamma_0(s)\exp(\psi' X_i^\circ(s)) \tag{3.2}$$

to be the working model where we use $\Gamma_0(s)$ to denote the baseline cumulative hazard and ψ for regression coefficients with $\psi = (\psi_1, \psi_2')'$. If (3.1) is the correct model then use of $X_i^\circ(s)$ in place of $X_i(s)$ will yield an estimate $\hat{\psi}$ which will be consistent for ψ^* where $\psi^* \neq \beta$. In what follows we explore the determinants of the limiting value to assess situations where the EM algorithm of Chapter 2 will be most important to use.

For a sample of size m, the working partial likelihood based on the working Cox model (3.2) is

$$\mathcal{L} \propto \prod_{i=1}^{m} \left\{ d\Gamma(v_i \mid X_i^\circ(v_i))^{\delta_i} \exp(-\int_0^\infty \bar{Y}_i(s) d\Gamma(s \mid X_i^\circ(s))) \right\}. \tag{3.3}$$

We obtain the partial likelihood estimating equations for $\alpha(s) = d\Gamma_0(s)$ and ψ as

$$U_1(s) = \frac{\partial \ell}{\partial \alpha(s)} = \sum_{i=1}^{m} \bar{Y}_i(s)\left\{ dN_i(s) - d\,\Gamma_0(s)\exp(\psi' X_i^\circ(s)) \right\}, 0 < s, \tag{3.4a}$$

$$U_2 = \frac{\partial \ell}{\partial \psi} = \sum_{i=1}^{m} \int_0^\infty \bar{Y}_i(s)\left\{ dN_i(s) - \bar{Y}_i(s) d\Gamma_0(s)\exp(\psi' X_i^\circ(s)) \right\} X_i^\circ(s), \tag{3.4b}$$

respectively. Let $U = (U_1(\cdot), U_2)'$ where here $U_1(\cdot)$ represents the estimating function $U_1(s)$ at all times $s > 0$. For a given dataset this reduces to a non-degenerate estimating function for $d\,\Gamma_0(\cdot)$ at each unique failure time in a sample. Here however, we aim to explore the limiting value of estimators from misspecified models, so we retain the general dependence on time $s > 0$ for the function $U_1(s)$ in the derivations that follows.

3.2 Asymptotic Distribution of Cox Model Estimators

If the model leading to the construction of (3.3) is incorrect, then the estimator of $\Omega = (\Gamma_0(\cdot), \psi')'$ is consistent for Ω^*, the solution to the equation formed by setting the expected working Cox model score equations equal to zero (Struthers and Kalbfleisch, 1986; White,

1982). Wei et al. (1989) showed that as $m \to \infty$,

$$\sqrt{m}(\hat{\Omega} - \Omega^*) \sim \text{MVN}(\, 0,\, \mathcal{A}^{-1}(\Omega^*)\mathcal{B}(\Omega^*)[\mathcal{A}^{-1}(\Omega^*)]')$$

where $\mathcal{A}(\Omega^*) = \mathbb{E}\{-\partial U/\partial\Omega\}$, and $\mathcal{B}(\Omega^*) = \mathbb{E}\{UU'\}$, and we use the notation $\mathbb{E}$ to convey an expectation taken with respect to the true model.

We next take the expectation of the estimating equations (3.3) and (3.4) with respect to the true distribution to solve for $\Omega^* = (\Gamma_0^*(\cdot), \psi^{*\prime})'$. At fixed ψ and for a given s, solving $\mathbb{E}\{U_1(s)\} = 0$ gives the limiting "profile" quantity

$$d\widetilde{\Gamma}_0^*(s;\psi) = \frac{\sum_{i=1}^m \mathbb{E}\{\bar{Y}_i(s)dN_i(s)\}}{\sum_{i=1}^m \mathbb{E}\{\bar{Y}_i(s)\exp(\psi' X_i^\circ(s))\}} \quad . \tag{3.5}$$

Note that each individuals data is assumed to arise from the same data generation process so we can equivalently write

$$d\widetilde{\Gamma}_0^*(s;\psi) = \frac{\mathbb{E}\{\bar{Y}_i(s)dN_i(s)\}}{\mathbb{E}\{\bar{Y}_i(s)\exp(\psi' X_i^\circ(s))\}} \quad .$$

If we substitute (3.5) into $\mathbb{E}\{U_2\}$ where U_2 is given by (3.4b), we obtain $\mathcal{U}_2(\psi) = \sum_{i=1}^m \mathcal{U}_{i2}(\psi) = m\mathcal{U}_{i2}(\psi)$ with

$$\begin{aligned}\mathcal{U}_{i2}(\psi) &= \int_0^\infty \left\{\mathbb{E}\{\bar{Y}_i(s)dN_i(s)X_i^\circ(s)\} - d\widetilde{\Gamma}_0^*(s;\psi)\mathbb{E}\{\bar{Y}_i(s)X_i^\circ(s)\exp(\psi' X_i^\circ(s))\}\right\} \\ &= \int_0^\infty \left\{r^{(1)}(s) - \frac{r^{(1)}(s;\psi)}{r^{(0)}(s;\psi)}\ r^{(0)}(s)\right\}ds \end{aligned} \tag{3.6}$$

where

$$r^{(k)}(s) = \mathbb{E}\{\, \bar{Y}_i(s)[X_i^\circ(s)]^{\otimes k}dN_i(s)\,\}, k = 0, 1, \tag{3.7a}$$

$$r^{(k)}(s;\psi) = \mathbb{E}\{\, \bar{Y}_i(s)[X_i^\circ(s)]^{\otimes k}\exp(\psi' X_i^\circ(s))\,\}, k = 0, 1. \tag{3.7b}$$

Thus the limiting value of the regression coefficient from a Cox model using $X_i^\circ(s)$ as the covariate is ψ^*, the solution to the equation $\mathcal{U}_2(\psi) = 0$.

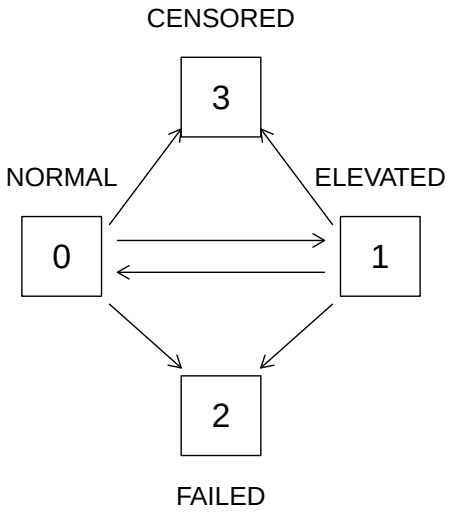


Figure 3.1: A joint model for the multistate and right-censoring processes

In order to evaluate this limiting value, a specific model for the data generating process must be provided. We consider the four state model depicted in Figure 3.1 to represent the joint marker-failure-censoring time process. The addition of the censoring state enables us to consider the impact of dependent censoring which we do briefly in what follows. We also need to give a possibly covariate and marker-dependent visit process through specification of the visit intensity. To do this we introduce a more elaborate multistate model shown in Figure 3.2. Here for convenience of representation we have two failure (death) states depending on whether failure occurred from the normal or high marker state but this is simply for presentation purposes - they represent the same absorbing states for any given row. We also introduce multiple censoring states on a given row that may be entered from the two marker states, but again this is only for display purposes. The states are organized so that a pair of rows correspond to a particular cumulative number of visits. The states in the far left and right columns are absorbing states representing failure or censoring while the inner two states reflect the marker values. Thus state $(j, 1)$, for example, is occupied by an individual alive at time s if they have had j visits and have an elevated marker and are uncensored (i.e. $A_i(s) = j, Z_i(s) = 1, Y_i(s) = 1$). A downward transition occur when a visit occurs, so the row of the figure represents the cumulative number of post-baseline

visits. Individuals are in the row corresponding to $A_i(s) = 0$ until the first follow-up visit. Thus the states are labeled with two integers with the first being the cumulative number of visits; the second reflects the state of the 3-state process or censoring. The longer an individual has spent in states $(j, 0)$ or $(j, 1)$, the more out of date their marker value $X_i^\circ(s)$ is.

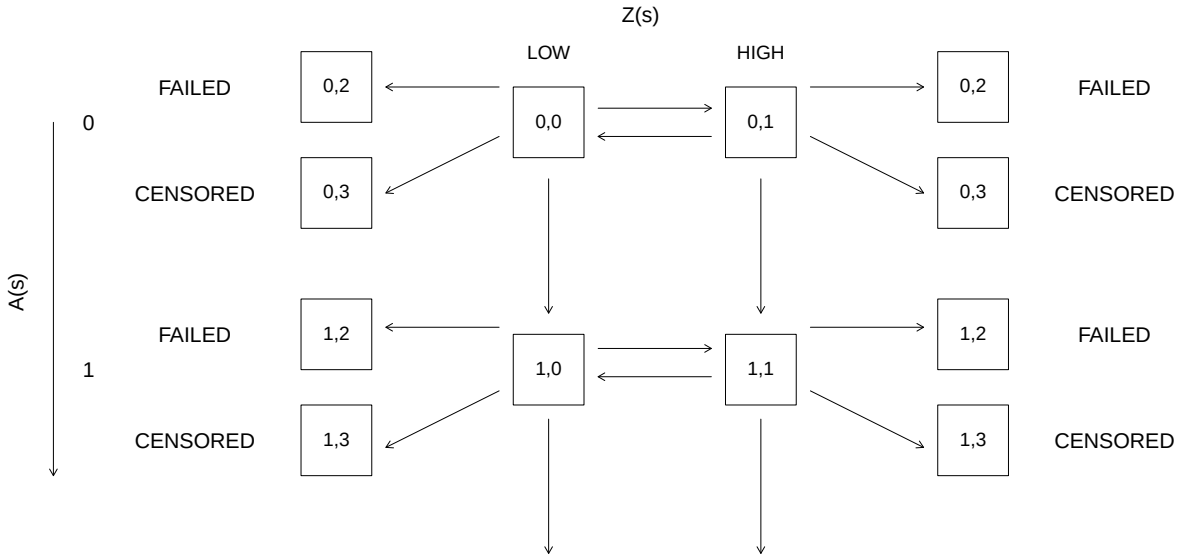


Figure 3.2: A joint model for the marker, failure, visit and right-censoring processes

We let

$$\mathcal{H}_i(t) = \{\bar{Y}_i(s), d\bar{A}_i(s), Z_i(s), 0 < s < t, X_{i2}\}$$

denote the *complete history* of the joint process, and let

$$\bar{\mathcal{H}}_i(t) = \{\bar{Y}_i(s), d\bar{A}_i(s), d\bar{N}_{i0}(s), d\bar{N}_{i1}(s), d\bar{N}_{i2}(s), 0 < s < t, X_{i2}\}$$

denote the process history accommodating right censoring, where $d\bar{A}_i(s) = \bar{Y}_i(s)dA_i(s)$, $d\bar{N}_{i0}(s) = \bar{Y}_i(s)dN_{i0}(s)$, $d\bar{N}_{i1}(s) = \bar{Y}_i(s)dN_{i1}(s)$, and $d\bar{N}_{i2}(s) = \bar{Y}_i(s)dN_{i2}(s)$. We then let

$$\bar{\mathcal{X}}_i^\circ(t) = \{(a_{ij}, X_i(a_{ij}), j = 0, 1, \ldots, \bar{A}_i(s)\} = \{d\bar{A}_i(s), X_i^\circ(s), 0 < s < t, X_i(0)\}$$

denote the observed history of the covariate process under the last observation carried forward approach, and finally we let

$$\bar{\mathcal{H}}_i^{\circ}(t) = \{\bar{Y}_i(s), d\bar{A}_i(s), d\bar{N}_{i2}(s), X_i^{\circ}(s), 0 < s < t\}$$

denote the history of the observed data process when failure is subject to right censoring and the markers are observed intermittently. The convention in this notation is that an overbar denotes a process history under right-censoring and the superscript "o" denotes an observed process history where the process of interest is subject to intermittent observation.

The intensities for the model depicted in Figure 3.2 are defined as follows. For the failure intensity we have a general formulation

$$\lim_{\Delta t \downarrow 0} \frac{P(\Delta \bar{N}_i(t) = 1 \,|\, \bar{\mathcal{H}}_i(t))}{\Delta t} = \bar{Y}_i(t)\lambda_i(t \,|\, \bar{\mathcal{H}}_i(t)), \tag{3.8}$$

where as before $\bar{Y}_i(t) = Y_i(t)Y_i^{\dagger}(t)$ with $Y_i(t) = I(t \leq C_i)$ and $Y_i^{\dagger}(t) = I(t \leq T_i)$. Note that this intensity becomes zero following either failure or censoring. Also note that this intensity is expressed in terms of the full censored data history whereby failure can depend on the current (even if unknown) marker value.

As before let $dN_{ij}(s)$ indicates a $j \to 1-j$ transition for the marker process, and $Y_{ij}(s)$ indicate individual i is at risk for such a transition, $j = 0, 1$. The intensities for the dynamic marker process are

$$\lim_{\Delta t \downarrow 0} \frac{P(\Delta \bar{N}_{ij}(t) = 1 \,|\, \bar{\mathcal{H}}_i(t))}{\Delta t} = \bar{Y}_{ij}(t)\lambda_{ij}(t \,|\, \bar{\mathcal{H}}_i(t)), \tag{3.9}$$

where $\bar{Y}_{ij}(s) = \bar{Y}_i(s)I(Z_i(s^-) = j), j = 0, 1$. The visit process intensity is

$$\lim_{\Delta t \downarrow 0} \frac{P(\Delta \bar{A}_i(t) = 1 \,|\, \bar{\mathcal{H}}_i(t))}{\Delta t} = \bar{Y}_i(t)\lambda_i^a(t \,|\, \bar{\mathcal{H}}_i(t)) \tag{3.10}$$

where $\bar{A}_i(t) = \int_0^t \bar{Y}_i(s)dA_i(s)$ and $\Delta\bar{A}_i(t) = \bar{A}_i(t + \Delta t^-) - \bar{A}_i(t^-)$. Note here that, in general, the visit process intensity at time t can depend on $X_{i1}(t)$, the latent marker value

at that time. This would be the case for a non-ignorable marker-dependent visit process which we consider how to deal with in Chapter 4.

The censoring intensity is

$$\lim_{\Delta t \downarrow 0} \frac{P(\Delta \bar{C}_i(t) = 1 \,|\, \bar{\mathcal{H}}_i(t))}{\Delta t} = \bar{Y}_i(t) \lambda_i^c(t \,|\, \bar{\mathcal{H}}_i(t)) \,, \tag{3.11}$$

where $\bar{C}_i(t) = I(C_i \leq t)$ and $\Delta \bar{C}_i(t) = \bar{C}_i(t + \Delta t^-) - \bar{C}_i(t^-)$. Again if (3.11) depends on $X_i(t)$ rather than $X_i^\circ(t)$ we will have a dependent censoring mechanism in a standard Cox regression analysis since we cannot render censoring independent by conditioning on $X_i(t)$ as $X_i(t)$ is not observed.

We have thus formulated a general stochastic model for the multistate process in Figure 3.2 in terms of full set of intensities enabling us to calculate the probability for an arbitrary path as

$$\mathcal{P}_i = \mathcal{P}_{i0} \times \left\{ \prod_{j=0}^{1} \mathcal{P}_{ij} \right\} \times \mathcal{P}_{ia} \times \mathcal{P}_{ic} \tag{3.12}$$

where if the realized process is represented as $\bar{\mathcal{H}}_i(\infty) = \{\bar{Y}_i(s), d\bar{A}_i(s), d\bar{N}_{i0}(s), d\bar{N}_{i1}(s), d\bar{N}_{i2}(s), 0 < s < V_i, (V_i, \delta_i), X_{i2}\}$,

$$\begin{aligned}
\mathcal{P}_{i0} &= \lambda_i(v_i \,|\, \bar{\mathcal{H}}_i(v_i))^{\delta_i} \exp\left(- \int_0^\infty \bar{Y}_i(u)\, \lambda_i(u \,|\, \bar{\mathcal{H}}_i(u))\, du \right), \\
\mathcal{P}_{ij} &= \prod_{l=1}^{\bar{N}_{ij}(v_i)} \lambda_{ij}(s_{il} \,|\, \bar{\mathcal{H}}_i(s_{il}), d\bar{N}_{ij}(s_i)) \exp\left(- \int_0^\infty \bar{Y}_{ij}(u) \lambda_{ij}(u \,|\, \bar{\mathcal{H}}_i(u)\, du) \right), \qquad j = 0, 1, \\
\mathcal{P}_{ia} &= \prod_{j=1}^{r_i} \lambda_i^a(a_{ij} \,|\, \bar{\mathcal{H}}_i(a_{ij})) \exp\left(- \int_0^\infty \bar{Y}_i(u)\, \lambda_i^a(u \,|\, \bar{\mathcal{H}}_i(u)\, du) \right),
\end{aligned}$$

and

$$\mathcal{P}_{ic} = \lambda_i^c(v_i \,|\, \bar{\mathcal{H}}_i(v_i))^{1-\delta_i} \exp\left(- \int_0^\infty \bar{Y}_i(u) \lambda_i^c(u \,|\, \bar{\mathcal{H}}_i(u) du) \right),$$

where $\bar{A}_i(v_i) = r_i$ is the total number of post-baseline visits and $\bar{N}_{ij}(v_i) = \int_0^\infty \bar{Y}_i(s) dN_{ij}(s)$, $j = 0, 1$ for individual i.

We aim to use this general framework to investigate the impact of naive use of $\{X_i^\circ(s),$

$0 < s\}$ as the time-dependent marker in a Cox regression model given in (3.2) in the framework of an independent or dependent visit and censoring process. To do so we consider some specific models and let

$$\lambda_i^c(t \mid \bar{\mathcal{H}}_i(t)) = \bar{Y}_i(t)\rho_c \exp(X_{i1}(t^-)\eta_c)$$

where η_c represents the strength of the marker dependence for the censoring time ($\eta_c = 0$ implies complete independence), and

$$\lambda_i^a(t \mid \bar{\mathcal{H}}_i(t)) = \bar{Y}_i(t)\rho_a \exp(X_{i1}(t^-)\eta_a)$$

where η_a represents the strength of the dependence of the visit process and the marker process. We assume the visit and censoring processes are conditionally independent of X_{i2} given $\{X_{i1}(s), 0 < s\}$ for simplicity, but it is straightforward to relax that association.

We use $\{\mathcal{Z}_i(s), 0 < s\}$ to depict the multistate process in Figure 3.2 to distinguish it from the 3-state multistate process $\{Z_i(s), 0 < s\}$. With a fully specified joint model, for a given individual in the kth row and jth column, we denote the state by (j, k). The state space for $\{\mathcal{Z}_i(s), 0 < s\}$ is therefore $\mathcal{S} = \{(j, k), j = 0, 1, \ldots, k = 0, 1, 2, 3\}$. The transition intensities (3.7) - (3.10) govern the dynamics of the complete process.

3.3 Limiting Values of Naive Estimators under LOCF

3.3.1 A General Framework

When the visit process is conditionally independent of the marker value given the observed data history, some explicit expressions can be established for the expected estimating functions. However, we consider the theory here in general and also describe how to compute the required expectations via Monte Carlo in Appendix C. In Appendix D we present an expanded joint model which facilitates direct computation of the required expectations based on a much larger but more comprehensive state space. This expanded state-space is well-suited to the generalization of the EM algorithm of Chapter 2 to accommodate a

semiparametric failure time analysis under a marker-dependent visit process but that is beyond the scope of this thesis.

Suppose X_{i2} is a $p \times 1$ vector which will make $X_i(t) = (X_{i1}(t), X'_{i2})'$ and $X_i^\circ(t) = (X_{i1}^\circ(t), X'_{i2})'$ a $(p+1) \times 1$ vector. Then for (3.7a) and (3.7b) we require the following expectations

(i) The scalar $r^{(0)}(s) = \mathbb{E}\{\bar{Y}_i(s)dN_i(s)\}$,

(ii) The $(p+1) \times 1$ vector $r^{(1)}(s) = \mathbb{E}\{\bar{Y}_i(s)X_i^\circ(s)dN_i(s)\}$,

(iii) The scalar $r^{(0)}(s;\psi) = \mathbb{E}\{\bar{Y}_i(s)\exp(\psi' X_i^\circ(s))\}$, and

(iv) The $(p+1) \times 1$ vector $r^{(1)}(s;\psi) = \mathbb{E}\{\bar{Y}_i(s)X_i^\circ(s)\exp(\psi' X_i^\circ(s))\}$

These expressions require working with the full joint model and for an example we provide the following derivation.

To compute $r^{(0)}(s)$, note

$$\begin{aligned}
\mathbb{E}\{\bar{Y}_i(s)dN_i(s)\} &= \mathbb{E}\Big\{\bar{Y}_i(s)E\{dN_i(s) \mid X_i(s), \bar{Y}_i(s) = 1\}\Big\} \\
&= \mathbb{E}\Big\{\bar{Y}_i(s)[d\Lambda_2(s)e^{\beta_1 + X'_{i2}\beta_2}P(\mathcal{Z}_i(s) \in \mathcal{C}_{01} \mid \mathcal{Z}_i(s) \in \mathcal{C}_{00} \cup \mathcal{C}_{01}, X_{i2}) \\
&\quad + d\Lambda_2(s)e^{X'_{i2}\beta_2}P(\mathcal{Z}_i(s) \in \mathcal{C}_{00} \mid \mathcal{Z}_i(s) \in \mathcal{C}_{00} \cup \mathcal{C}_{01}, X_{i2})]\Big\} \\
&= d\Lambda_2(s)\mathbb{E}\Big\{e^{X'_{i2}\beta_2}(e^{\beta_1}P(\mathcal{Z}_i(s) \in \mathcal{C}_{01} \mid X_{i2}) + P(\mathcal{Z}_i(s) \in \mathcal{C}_{00} \mid X_{i2}))\Big\}
\end{aligned}$$

where $\mathcal{C}_{0k} = \{(j,k), j = 0, 1, \ldots\}$ is the set of states with the second subscript equal to k, $k = 0, 1$. The remaining expectation is with respect to the covariate distribution for the fixed covariate X_{i2} and so the expectation will depend on the precise form of the covariate distribution.

The derivations for i) - iv) will be described in detail in Appendix C based on the full joint model of Figure 3.2, or an expanded version of it described in Appendix C. For the

failure process we let

$$\lambda_i(t \,|\, \bar{\mathcal{H}}_i(t)) = \bar{Y}_i(t)\, \lambda_2 \, \exp(X_{i1}(t)\beta_1 + X'_{i2}\beta_2\,)$$

and for the marker transitions we let

$$\lambda_{ij}(t \,|\, \bar{\mathcal{H}}_i(t)) = \bar{Y}_{ij}(t)\lambda_j \, \exp(X'_{i2}\,\gamma_j)$$

for simplicity in the following calculations. We next consider computations for a particular setting to investigate the primary sources of the limiting biases.

3.3.2 Asymptotic Bias Under a Marker-Independent Visit Process

We first explore the asymptotic bias from carrying forward the marker value from the most recent assessment under a marker-independent visit process. We again consider data arising over the interval $[0, 1]$, and here let X_{i2} be a binary covariate with $P(X_{i2} = 1) = P(X_{i2} = 0) = 0.5$. We plot the graph of the empirical bias of both ψ_1^* and ψ_2^* with respect to different values of $\log(\lambda_1/\lambda_0)$ which reflects the proportion of time in the two marker states while event-free for an individual with $X_{i2} = 0$. We set the rate ρ of a Poisson visit process so that $E\{A(1)\} = 4, 8$ or 12. We consider settings where $\lambda_0 = 4$ or 8. We let the true marker effect to be $\beta_1 = \log 1.25$ and $\beta_2 = \log 1.5$.

Given the value of λ_0, for the 3-state process, we set the $1 \to 2$ intensity to $1.25\lambda_2$ and at different values of $\log(\lambda_1/\lambda_0)$, we solve for parameters of the multistate process so that the probability of an event at administrative censoring time equals to 60%. As it is a marker-independent visit process, $\eta_a = 0$, and we solve for the visit intensity ρ_a so that

$$E\{A(1)\} = \sum_{j=0}^{\infty} j \sum_{k=0}^{1} P(Z(1) \in \mathcal{R}_j \,|\, Z(0) = (0, k))P(Z(0) = (0, k))\,.$$

Figure 3.3 - Figure 3.6 show the asymptotic bias of both $\hat{\psi}_1$ and $\hat{\psi}_2$ with respect to both the log ratio of λ_1 over λ_0 and the Poisson mean number of visits under different values of

λ_0. We see that for all plots, when the expected number of visits increases, the limiting absolute bias decreases, because we often obtain more information with increasing number of visits. We also see that the limiting bias increases when $\log(\lambda_1/\lambda_0)$ increases. This might be due to the increased number of visits to state 1 rather than to state 2, resulting a higher chance of getting mismeasured covariate value when individuals move more often in between state 0 and state 1.

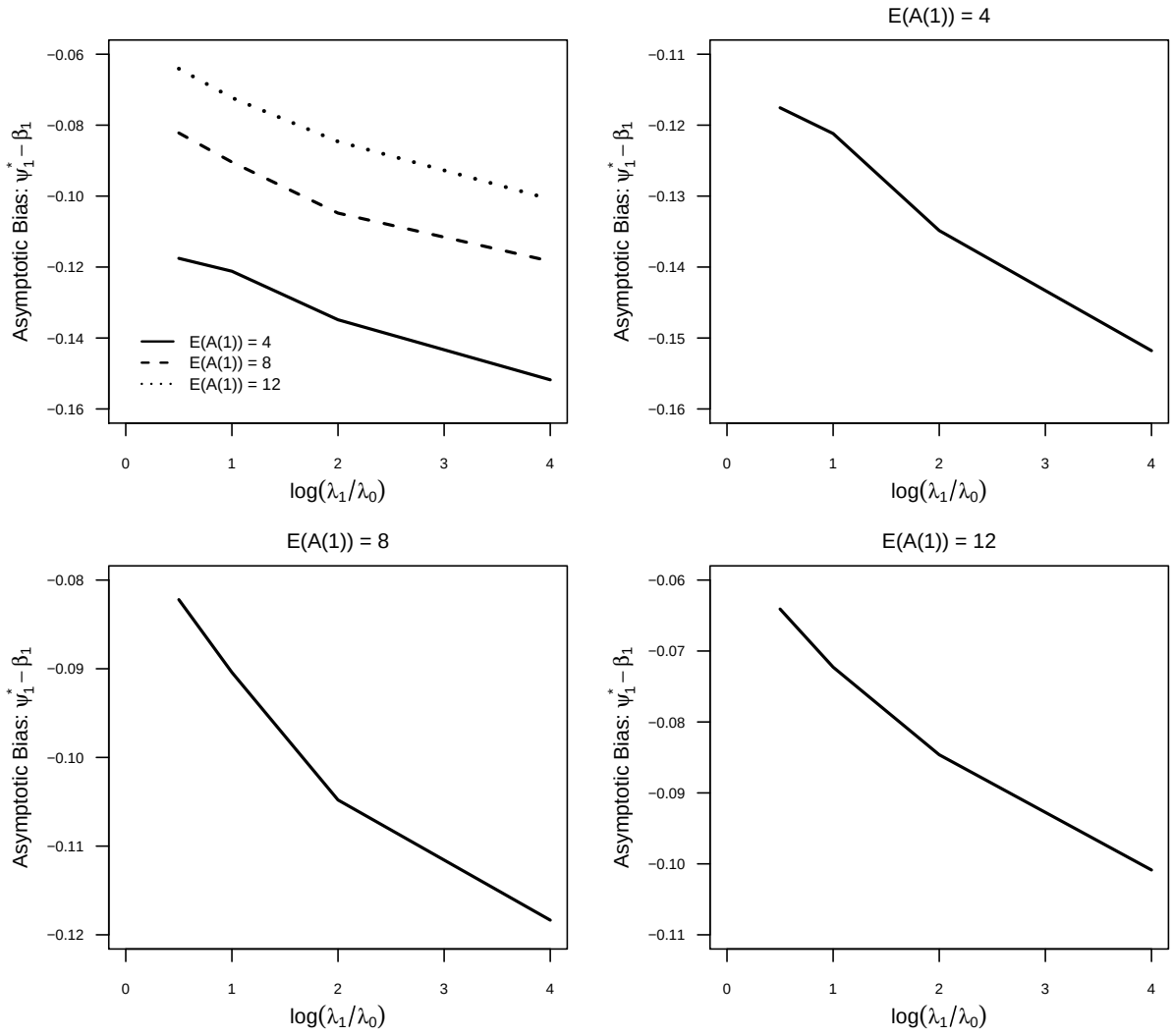


Figure 3.3: Asymptotic bias of $\hat{\psi}_1$ for β_1, when the marker effect is estimated based on a naive Cox regression model and marker value is carried forward as a function of the log of the ratio of λ_1 over λ_0 -independent censoring and independent visit are considered here; First panel shows biases for all expected number of assessments while remaining panels are for a given expected number of assessments to better display the trend as a function of the ratio of λ_1 over λ_0; $\lambda_0 = 4$ here.

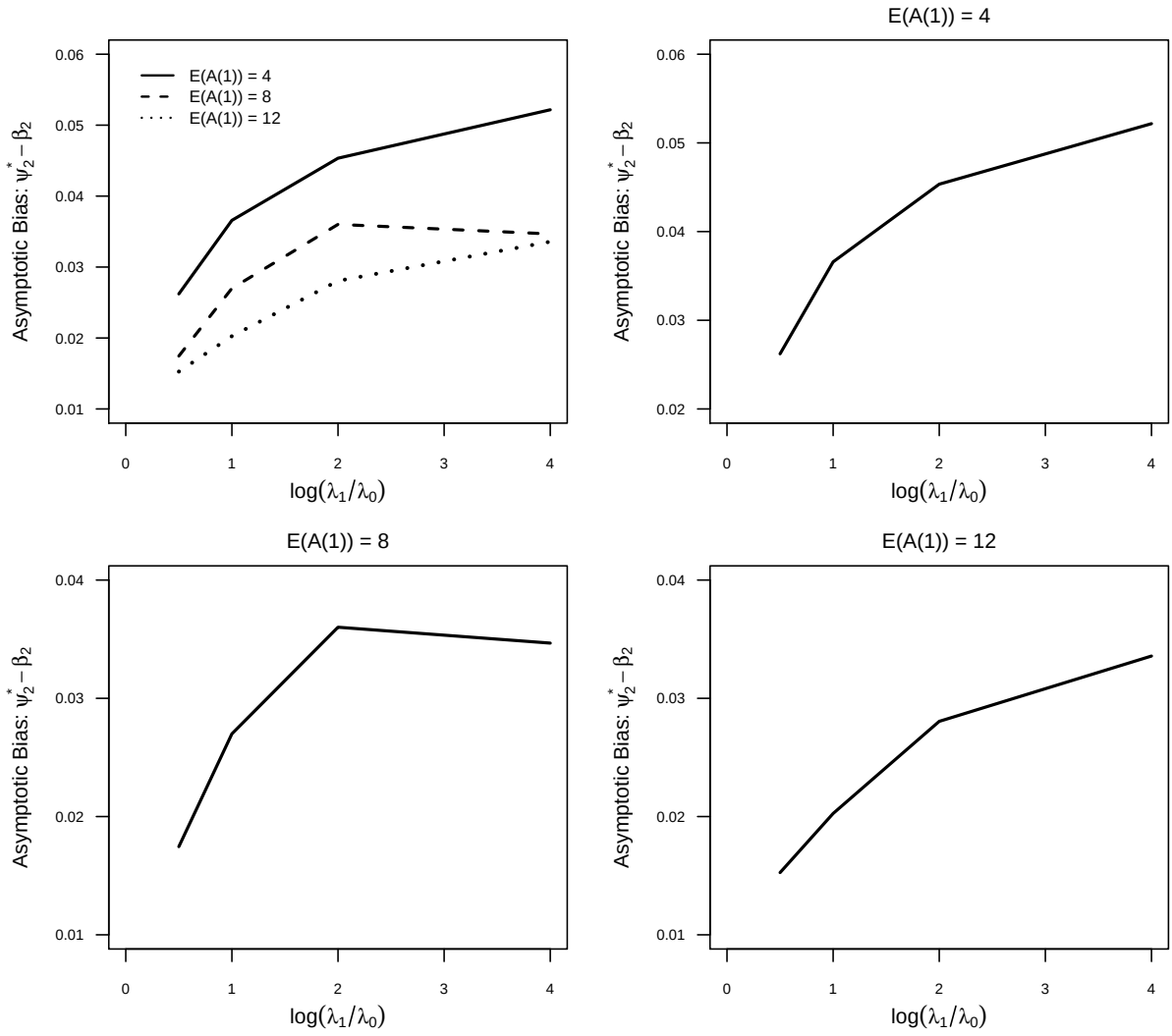


Figure 3.4: Asymptotic bias of $\hat{\psi}_2$ for marker effect β_2 when estimated based on a naive Cox regression model and marker value is carried forward as a function of the log of the ratio of λ_1 over λ_0 -independent censoring and independent visit are considered here; First panel shows biases for all expected number of assessments while remaining panels are for a given expected number of assessments to better display the trend as a function of the ratio of λ_1 over λ_0; $\lambda_0 = 4$ here.

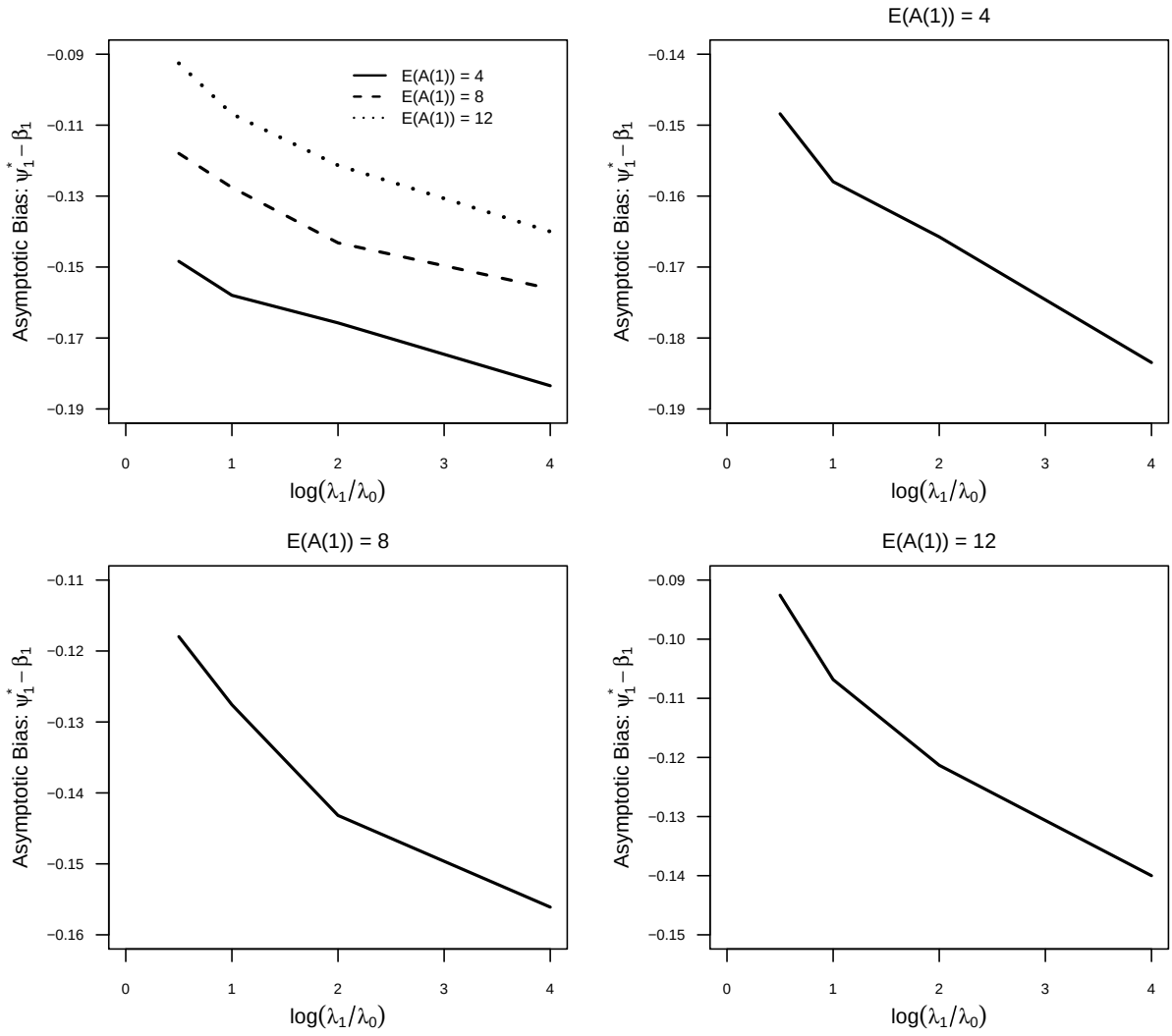


Figure 3.5: Asymptotic bias of $\hat{\psi}_1$ for marker effect β_1 when estimated based on a naive Cox regression model and marker value is carried forward as a function of the log of the ratio of λ_1 over λ_0 -independent censoring and independent visit are considered here; First panel shows biases for all expected number of assessments while remaining panels are for a given expected number of assessments to better display the trend as a function of the ratio of λ_1 over λ_0; $\lambda_0 = 8$ here.

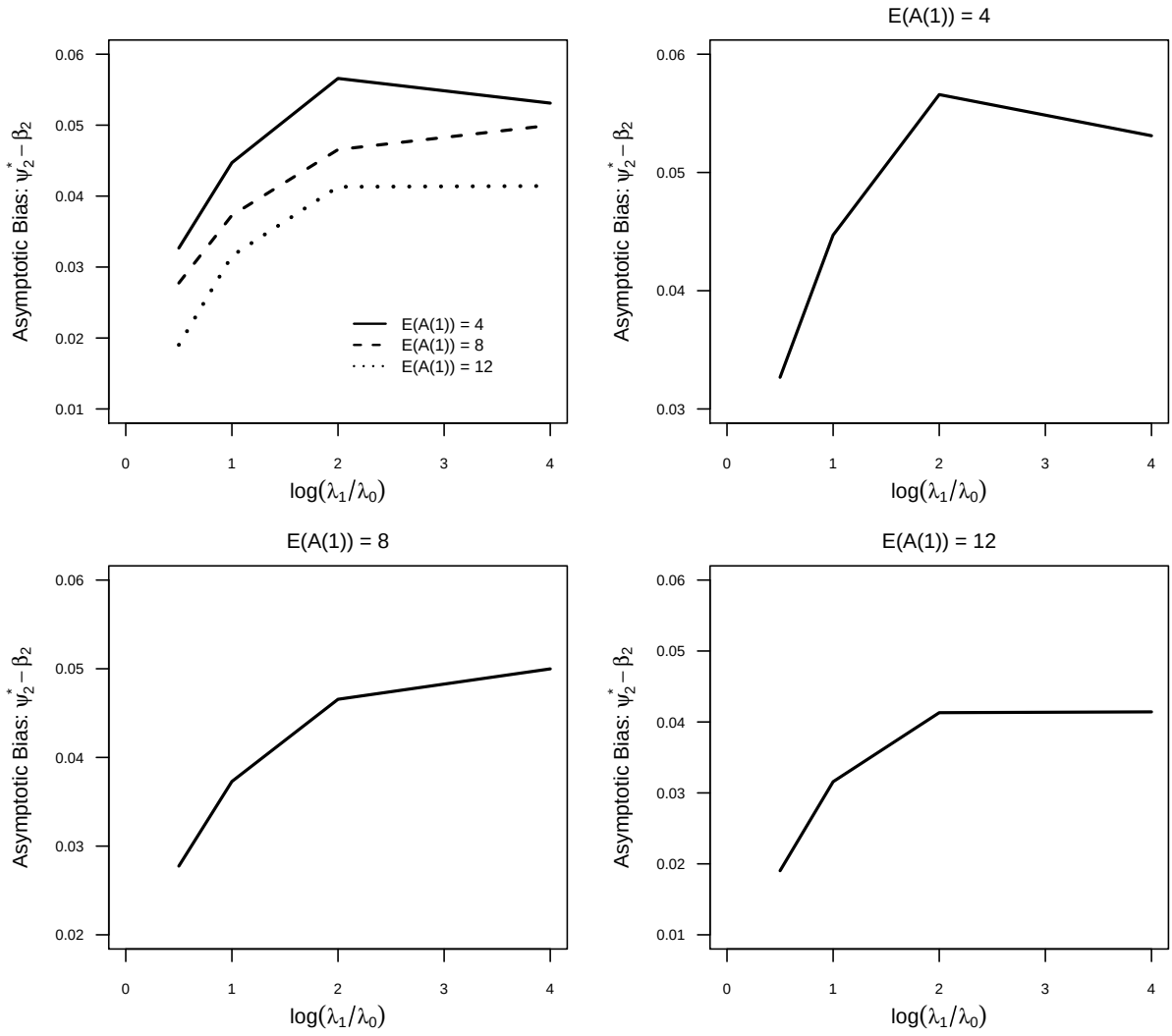


Figure 3.6: Asymptotic bias of $\hat{\psi}_2$ for marker effect β_2 when estimated based on a naive Cox regression model and marker value is carried forward as a function of the log of the ratio of λ_1 over λ_0 -independent censoring and independent visit are considered here; First panel shows biases for all expected number of assessments while remaining panels are for a given expected number of assessments to better display the trend as a function of the ratio of λ_1 over λ_0; $\lambda_0 = 8$ here.

3.3.3 Asymptotic Bias Under a Marker-Dependent Visit Process

Here we investigate the asymptotic bias arising from carrying forward the most recently recorded value of a time-dependent marker as the limiting value of a regression coefficient in a Cox model under a dependent visit process. We set an administrative censoring time $A = 1$. For the three-state process, we set the initial distribution via $P(Z(0) = 0) = 0.5$ and set the true log hazard ratio to $\exp(\gamma) = 2.5$. We set the probability of death at the administrative censoring to be $P(Z(A) = 2 \,|\, Z(0) = 0) = 0.4$ and 0.8. We then solve for transition intensities under the constraints $\lambda_1/\lambda_0 = 2$, $\lambda_0/(\lambda_0 + \lambda_2) = 0.8$. For the setting with an independent visit process we adopt a Poisson visit process with rate ρ and set ρ such that $E\{A(1)\} = \rho = 4, 8$ or 12.

To specify the censoring process, we consider $\exp(\eta_c) = 1$ for an independent censoring process and $\exp(\eta_c) = 1.2$ for a mildly dependent censoring process, and set ρ_c so that the net probability of censoring over $[0, 1]$ is 40%. That is, we solve for ρ_c so that

$$P_C = \sum_{j=0}^{\infty} \sum_{k=0}^{1} P(Z(1) = (j, 3) \,|\, Z(0) = (0, k)) P(Z(0) = (0, k)) = 0.4\,. \tag{3.13}$$

To generate a dependent visit process, we let $\mathcal{R}_j = \Big\{(j, 0), (j, 1), (j, 2), (j, 3)\Big\}$ to be the set of states at row j. We consider the values $\exp(\eta_a) = 0.6, 0.8, 1, 1.2, 2$ where $\exp(\eta_a) = 1$ stands for a conditionally independent visit process. With the administrative censoring at time 1, we then solve for ρ_a so that the expected number of visits,

$$E\{A(1)\} = \sum_{j=0}^{\infty} j \sum_{k=0}^{1} P(Z(1) \in \mathcal{R}_j \,|\, Z(0) = (0, k)) P(Z(0) = (0, k))\,, \tag{3.14}$$

is equal to 4, 8 or 12. We set the upper limit of the sum over visits j to be 50 in solving (3.13) and (3.14); that is, we assume the expanded joint model consists of no more than 50 rows. This censors the visit process but at such a high number of visits that the upper limit has negligible impact on the accuracy of the calculations.

A Dependent Visit Process and Independent Right Censoring

We consider both the independent right censoring (i.e. $\rho_c = \rho_c \exp(\eta_c) = 0$ so there is no random censoring) as well as the dependent censoring case in the following section. For both cases, we solve for ρ_a so that the number of total visit for an individual has mean $E\{A(1)\} = \sum_{j=1}^{\infty} j \sum_{k=0}^{1} P(Z(1) \in \mathcal{R}_j | Z(0) = (0,k)) P(Z(0) = (0,k)) = 4, 8, 12$ at time 1 for each value $\exp(\eta_a) = 0.6, 0.8, 1, 1.2, 2$. We plot the asymptotic bias for β with $E\{A(1)\} = \rho = 4, 8, 12$ under both independent censoring (see Figure 3.7) and dependent censoring (see Figure 3.8).

For the independent censoring case, the different values of η_a correspond to a range of settings including cases when an elevated marker decreases the visit intensity to cases where it increases the intensity of a visit. Note that the effects of a given η_a will be different according to the baseline intensity of visits since a strongly dependent visit process will have a weaker effect if visits occur frequently.

In the top left panel we see that the mean number of visits has a large impact on the magnitude of the bias as it determines how long, on average, values are carried forward. The trend in the asymptotic bias as a function of the marker dependence is weaker, so in the remaining three panels of Figure 3.7 we display the plots for a given mean number of visits so the scales better enable illustration of the trends. We generate a dataset of $n = 2000$ individuals. We then solve for the limiting value ψ_1^* and plot $\psi_1^* - \beta_1$ vs. $\exp(\eta_a)$.

Dependent Censoring

When censoring is dependent we set $\exp(\eta_c) = 1.2$ and solve for ρ_c so that $P_C = 0.40$ so 40% of the individuals should be censored by the administrative censoring time 1. Figure 3.8 contains the corresponding plots where a similar pattern is seen.

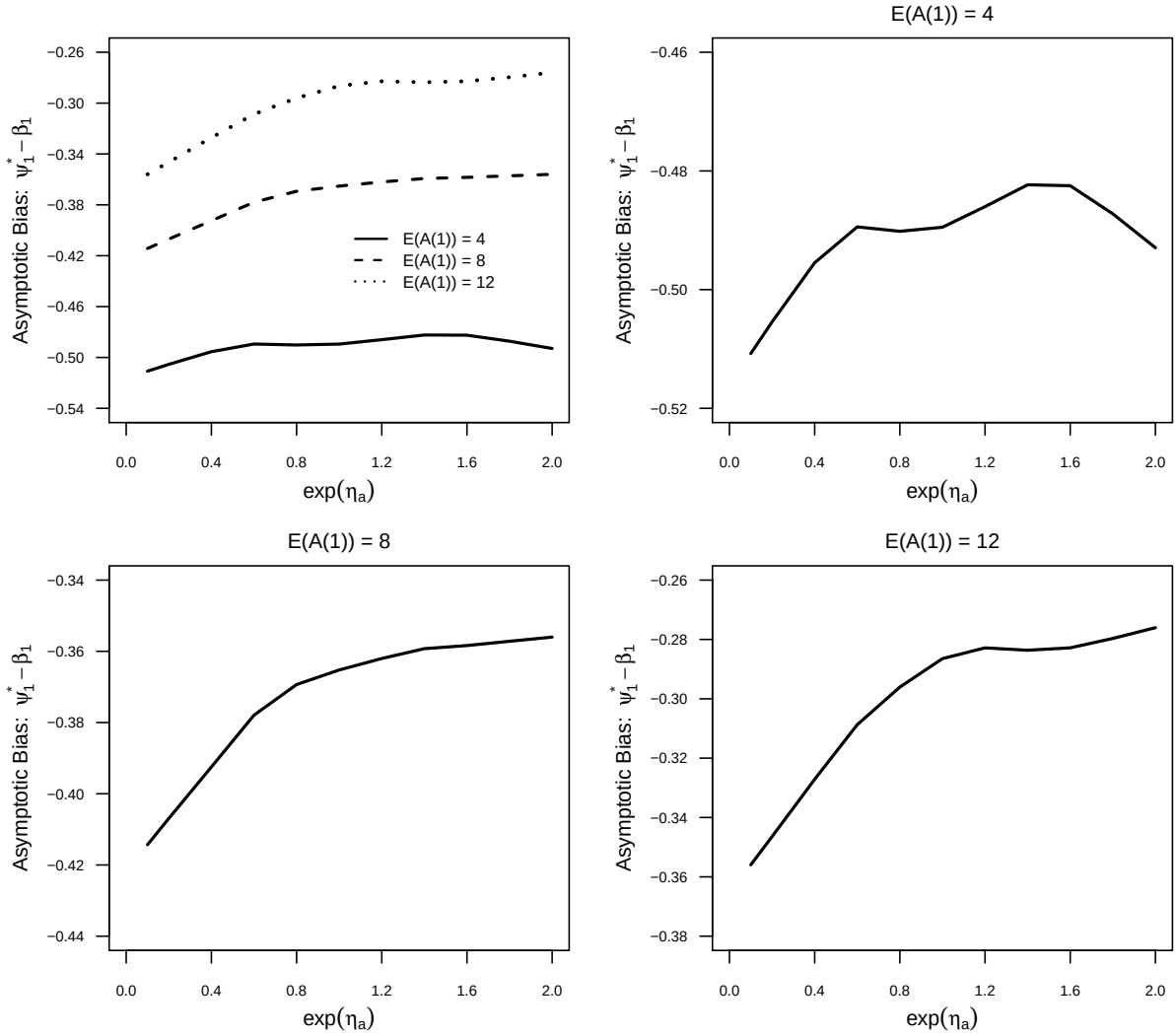


Figure 3.7: Asymptotic bias of regression coefficient for marker effect estimated based on naive Cox regression with marker value carried forward as a function of the expected number of assessments over $(0, 1]$ and the strength of the marker effect on the visit intensity - independent censoring is considered here; First panel shows biases for all expected number of assessments while remaining panels are for a given expected number of assessments to better show trend as a function of marker dependence.

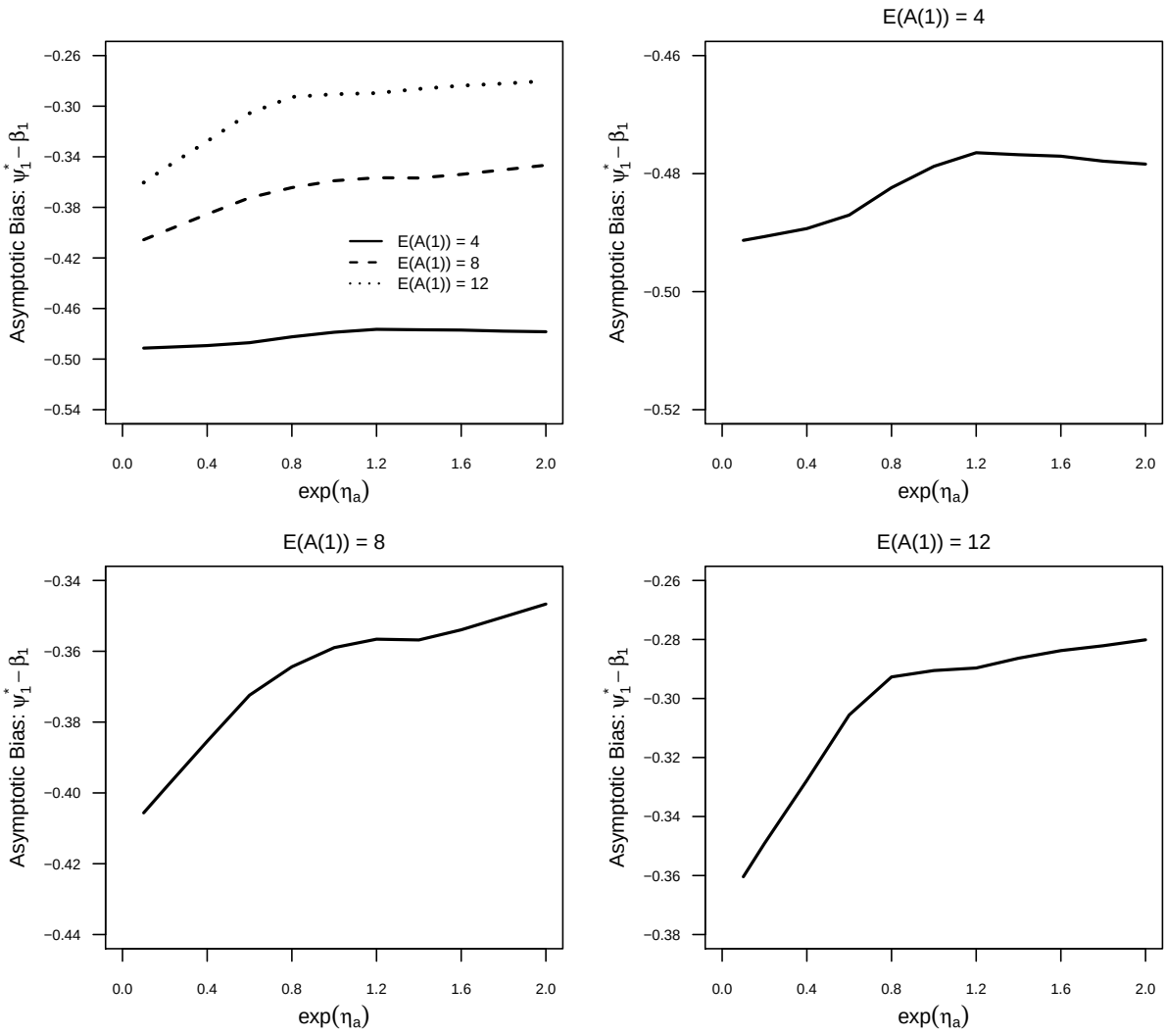


Figure 3.8: Asymptotic bias of regression coefficient for marker effect estimated based on naive Cox regression with marker value carried forward as a function of the expected number of assessments over $(0, 1]$ and the strength of the marker effect on the visit intensity - dependent censoring is considered here; First panel shows biases for all expected number of assessments while remaining panels are for a given expected number of assessments to better show trend as a function of marker dependence; The probability of dependent censoring $P_C = 40\%$.

In both plots, the x-axis reflects the extent of the dependence between the marker value and the visit process. We could see from the top left panel that the limiting bias $\psi_1^* - \beta_1$ decreases as the expected number of visits increases. This is due to the increased information obtained when we see patients more often. In general, the bias decreases when $\exp(\eta_a)$ increases. Note that the bias at $\exp(\eta_a) = 1$ may only be due to the out-of-date marker value.

3.3.4 Simulation Studies to Investigate Finite Sample Bias

We then carry out a simulation study to investigate empirical properties of the estimator with misspecified Cox model using $X^\circ(s)$ (denoted by the mean on the table) and the limiting value (β^*) under independent censoring to see how these two values agree with each other in each parameter settings. For the misspecified Cox model, we consider 500 simulations and in each simulation, $n = 1000$. For solving the limiting value based on the estimating equations, we consider $n = 2000$ individuals.

As expected, from Table 3.1 from the asymptotic calculations leading to Figure 3.7 and Figure 3.8, Cox regression estimators are biased when marker values are carried forward. Size of bias depends on the visit intensity dependence on marker value and we see good agreement between the limiting value $\psi^\star$ and the finite sample empirical average estimate. Also, the empirical standard errors (ESE) agree well with the average robust standard errors (ASE).

As in Table 3.1, we compared the values of the limiting value with the finite sample empirical average estimate when a fixed covariate X_2 is added. In Table 3.2 we compare the limiting values ψ_1^* and ψ_2^* with estimates by fitting a Cox model for 500 simulations of 2000 individuals each. It shows that there is a good agreement between the two values, and the average robust standard errors are close to the empirical standard errors.

		40% Failure Rate				80% Failure Rate			
$E\{A(1)\}$	$\exp(\eta_a)$	ψ_1^*	MEAN	ESE	ASE	ψ_1^*	MEAN	ESE	ASE
4	0.6	0.5734	0.5880	0.1686	0.1663	0.4322	0.4261	0.1157	0.1250
	0.8	0.5845	0.5801	0.1656	0.1664	0.4271	0.4443	0.1268	0.1234
	1	0.5976	0.6128	0.1620	0.1670	0.4347	0.4419	0.1184	0.1229
	1.2	0.6074	0.6060	0.1715	0.1665	0.4343	0.4337	0.1312	0.1235
	2	0.5915	0.6111	0.1520	0.1681	0.4486	0.4316	0.1348	0.1232
8	0.6	0.6724	0.6741	0.1846	0.1659	0.5403	0.5417	0.1155	0.1246
	0.8	0.6762	0.6877	0.1627	0.1670	0.5627	0.5618	0.1182	0.1240
	1	0.6886	0.6923	0.1650	0.1659	0.5607	0.5545	0.1140	0.1236
	1.2	0.6971	0.6995	0.1845	0.1655	0.5591	0.5649	0.1186	0.1233
	2	0.6922	0.6946	0.1832	0.1672	0.5637	0.5632	0.1165	0.1237
12	0.6	0.7330	0.7407	0.1563	0.1663	0.6127	0.6106	0.1274	0.1236
	0.8	0.7473	0.7463	0.1392	0.1671	0.6131	0.6282	0.1228	0.1232
	1	0.7446	0.7579	0.1440	0.1660	0.6348	0.6302	0.1115	0.1233
	1.2	0.7466	0.7719	0.1546	0.1674	0.6286	0.6207	0.1070	0.1227
	2	0.7488	0.7468	0.1806	0.1686	0.6289	0.6323	0.1226	0.1234

Table 3.1: Parameter estimates with coxph function and estimating equations for the illness-death model. Number of individuals per simulation for the coxph =1000, Number of simulations = 500. Number of individuals for estimating equations = 2000. $P(Z(A) = 2 \mid Z(0) = 0) = 0.4$.

		β_1				β_2			
$E\{A(1)\}$	$\exp(\eta_a)$	ψ_1^*	MEAN	ESE	ASE	ψ_2^*	MEAN	ESE	ASE
4	0.6	0.0816	0.0832	0.1084	0.1112	0.4524	0.4569	0.1115	0.1119
	0.8	0.0876	0.0843	0.1156	0.1110	0.4484	0.4529	0.1081	0.1121
	1	0.0866	0.0905	0.1196	0.1112	0.4502	0.4521	0.1125	0.1120
	1.2	0.0868	0.0869	0.1153	0.1116	0.4523	0.4531	0.1108	0.1122
	2	0.0955	0.0940	0.1120	0.1133	0.4463	0.4419	0.1181	0.1119
8	0.6	0.1127	0.1188	0.1153	0.1119	0.4438	0.4423	0.1095	0.1128
	0.8	0.1181	0.1187	0.1123	0.1119	0.4367	0.4434	0.1121	0.1129
	1	0.1216	0.1278	0.1112	0.1123	0.4380	0.4432	0.1126	0.1131
	1.2	0.1210	0.1225	0.1131	0.1125	0.4392	0.4361	0.1133	0.1129
	2	0.1235	0.1249	0.1137	0.1141	0.4388	0.4329	0.1149	0.1128

Table 3.2: Parameter estimates with coxph function and estimating equation. ψ_1^* and ψ_2^* are calculated by estimating function. Mean, ESE, ASE are based on coxph() function. Number of individuals per simulation =2000. $P(Z(A) = 2|Z(0) = 0) = 0.6$, $\exp(\gamma_0) = 2$ and $\exp(\gamma_1) = 0.5$. Number of individuals for coxph function =500, number of simulations =500.

3.4 Concluding Remarks

In this paper we developed a joint model for the marker, censoring and failure process that deals with both marker independent and marker dependent visit when the marker value is carried forward from the most recent assessment time. We have demonstrated that the Cox regression estimates are biased in the covariate of time-varying marker from conventional analysis when the last observation is carried forward for both marker-independent and marker dependent visit process or right censoring process. We also extended the work by adding a fixed covariate to the setting to generalize the finding.

Chapter 4

A Joint Analysis With a Marker-Dependent Visit Process

4.1 Introduction

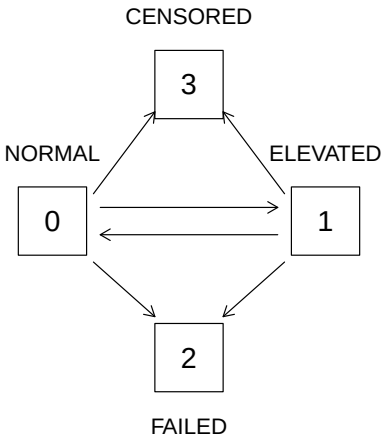
In Chapter 3 we described two elaborate multistate processes which jointly described a marker-failure process with a possibly marker-dependent visit process and marker-dependent right-censoring process. The one depicted in Figure 3.2 was used to evaluate the asymptotic bias of regression coefficient estimators in the Cox model (3.2) as estimators of the coefficients in the Cox model of (3.1). Such biases arise from carrying forward the recorded marker values between assessments when they are observed according to a marker-dependent visit process and a possibly marker-dependent censoring process. Through full specification of the joint model, the probability of a particular sample path of observed data can be evaluated as described in (3.12). Here we show how to compute the probability of all sample paths under this observation scheme and use this to construct a full likelihood for an analysis which accommodates marker dependence for the visit and censoring processes. We do this by evaluating the transition probabilities when the transition intensities are piecewise-constant. An important special feature of our goal is to accommodate a larger number of cut-points for the transition intensity to the failure state which is possible because the failures are only subject to right-censoring. The greater the number

of cut-points for this baseline hazard, the better this will approximate the results of a Cox model depicted in (3.1) which cannot be fit using standard method due to intermittent observation of the marker process.

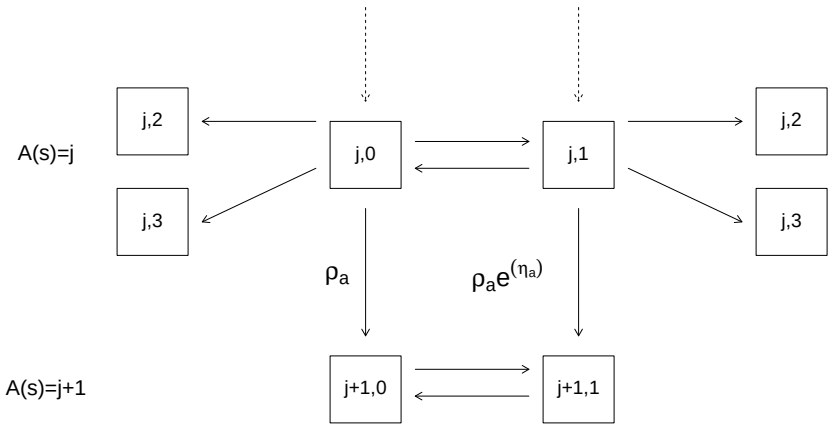
Note that the msm() function in R by Jackson (2011) can be used to fit Markov process under intermittent observation, and can handle the case that some states may be entered at exact observation times if they are terminal states. In our setting with the full joint model, exact transition times are observed when failure is observed, when right-censoring occurs, or when downward transitions are made upon visit occurrence. The states which are in the inner two columns in Figure 3.2 are not absorbing states for the full process. Thus the msm() function cannot be used for the full model fit; this is what motivates us to consider a partition of the process.

We partition the full multistate process depicted in Figure 3.2 into sub-processes corresponding to contributions between visit times, or between the last visit and the censoring or failure time. In panel (a) of Figure 4.1 we show the multistate process of interest which includes the marker states (0 and 1), a failure state (2), and a censoring state (3). Recall that this process is only partially observed because the marker process is under intermittent observation – that is transitions between states 0 and 1 are not observed. In panel (b) of Figure 4.1 we show the component of the full process in Figure 3.2 corresponding to the transitions for individual i, say, from the time of the assessment at a_{ij} when the jth increment in the visit counting process occurs, until the next observable event which could be failure, censoring or the occurrence of the next visit at time $a_{i,j+1}$. Note as in Figure 3.2 the states are labeled with two indices with the first identifying the cumulative number of visits and the second the state of the latent four state process in Figure 4.1a.

Let $A_0 < A_1 < \ldots$ represent random visit times and $a_0 < a_1 < \ldots$ be their realized values. Here, entry to state $(j, 2)$ corresponds to failure after the jth visit at a_{ij}, and entry to state $(j, 3)$ corresponds to censoring after a_{ij}. Transitions between $(j, 0)$ and $(j, 1)$ can occur but are not observed. An individual makes a downward transition if a visit occurs



(a) The multistate process $\{Z(s), 0 < s\}$ incorporating censoring



(b) The multistate representation of a component of the full joint process depicted in Figure 3.2 for information immediately after the jth visit.

Figure 4.1: A multistate model for (A, Z) depicting a joint model for the multistate, visit and failure process which accommodates a marker-dependent visit process and status of binary covariate

and at that time it is determined whether the marker is normal or elevated. We let

$$\bar{\mathcal{H}}_i(t) = \{\bar{Y}_i(s), dA_i(s), Z_i(s), 0 < s < t, X_{i2}\} \tag{4.1}$$

denote the complete history of the joint process, and let

$$\bar{\mathcal{H}}_i^\circ(t) = \{\bar{Y}_i(s), dA_i(s), d\bar{N}_i(s), 0 < s < t, X_{i1}^\circ(a_{ij}), j = 0, 1, \ldots, A_i(t^-), X_{i2}\} \tag{4.2}$$

denote the *observed history* of the multistate processes under right censoring where here $d\bar{N}_i(s) = \bar{Y}_i(s)dN_i(s)$ and $\{N_i(s), 0 < s\}$ is the counting process for failure for individual i. The visit process intensity is

$$\lim_{\Delta t \downarrow 0} \frac{P(\Delta \bar{A}_i(t) = 1 \,|\, \bar{\mathcal{H}}_i(t))}{\Delta t} = \bar{Y}_i(t)\lambda_i^a(t \,|\, \bar{\mathcal{H}}_i(t)) \tag{4.3}$$

in general where $\bar{A}_i(t) = \int_0^t \bar{Y}_i(s)dA_i(s)$ and $\Delta\bar{A}_i(t) = \bar{A}_i(t + \Delta t^-) - \bar{A}_i(t^-)$. We use the state-space diagram in Figure 4.1b to investigate the effect of marker dependence on the visit process intensity. To do this, we consider

$$\lambda_i^a(t \,|\, \bar{\mathcal{H}}_i(t)) = \bar{Y}_i(t)\, \rho_a \exp(X_{i1}(t^-)\eta_a)$$

where η_a represents the strength of the dependence of the visit process and the marker value. Note that the visit intensity could also depend on X_{i2} but that would be ignorable when we condition on X_{i2} in the working Cox model so we do not consider it here.

4.2 Computation of the Transition Probability Matrix

4.2.1 Model Specification and Preliminary Derivation

Here we aim to fit the full model accommodating a marker-dependent visit and censoring process. To do so we consider the likelihood based on the observed sample path as described in (3.12).

For the marker process, we let $dN_{ij}(s) = 1$ if a $j \to 1-j$ transition occurs at time s and $dN_{ij}(s) = 0$ otherwise, $j = 0, 1$. Let $Y_{ij}(s) = I(Z_i(s^-) = j)$ and $\bar{Y}_{ij}(s) = Y_i(s)\, Y_{ij}(s)$, $j = 0, 1$, where $Y_i(s) = I(s \leq C_i)$ indicates individual i is under observation at time s. Suppose X_{i2} is a p$\times$1 vector of fixed covariates, we then let

$$\lim_{\Delta t \downarrow 0} \frac{P(\Delta N_{ij}(t) = 1 \mid Z_i(t^-) = j, X_{i2})}{\Delta t} = \lambda_j(t \mid X_{i2}) = \lambda_j(t)\, \exp(X_{i2}'\gamma_j)\,, \tag{4.4}$$

be a Markov transition rate where γ_j is a $p\times 1$ vector of regression coefficients. The function $\lambda_j(t)$ is the baseline intensity for $j \to 1-j$ transitions, $j = 0, 1$.

For the failure time process we wish to study the effect of the marker and the fixed covariates and we let $X_{i1}(t) = I(Z_i(s^-) = 1)$ and define $X_i(t) = (X_{i1}(t), X_{i2}')'$ to be the time dependent covariate vector comprised of the marker $X_{i1}(t)$ and the fixed covariate X_{i2}. We continue to consider the case in which the marker is under continuous observation, and here we construct the likelihood for the failure process parameters. The intensity for the failure time T_i is

$$\lim_{\Delta t \downarrow 0} \frac{P(t \leq T_i < t + \Delta t \mid t \leq T_i, X_i(t))}{\Delta t} = \lambda_2(t \mid X_i(t)) \tag{4.5}$$

which is taken to have the multiplicative form

$$\lambda_2(t \mid X_i(t)) = \lambda_2(t)\, \exp(X_i'(t)\beta) = \lambda_2(t)\, \exp(X_{i1}(t)\beta_1 + X_{i2}'\beta_2)\,,$$

where $\beta = (\beta_1, \beta_2')'$.

We assume that the baseline intensities are piecewise-constant. As in Chapter 2, we let $0 = b_{j0} < b_{j1} < \ldots < b_{jk_j} = \infty$ denote the break-points for the baseline intensity for transitions from state j, $j = 0, 1$, or into state 2 ($j = 2$). We then let $\mathcal{B}_{jk} = [b_{j,k-1}, b_{jk})$, $k = 1, \ldots, K_j$, $j = 0, 1, 2$ define the set of all sub-intervals and let

$$\mathcal{B}_{k_0 k_1 k_2} = \mathcal{B}_{0k_0} \mathcal{B}_{1k_1} \mathcal{B}_{2k_2} \tag{4.6}$$

denote the sub-interval within $\mathcal{B}_{0k_0}$, $\mathcal{B}_{1k_1}$ and $\mathcal{B}_{2k_2}$, in which all baseline intensities are constant. We can relabel the sub-intervals of the positive real line sequentially and denote this by

$$\mathcal{B}_k = [b_{k-1}, b_k),\ k = 1, \ldots, K$$

where each value of (k_0, k_1, k_2) corresponds to a particular unique value of k in this row index. Thus the multistate process depicted in Figure 4.1b is time-homogeneous within each interval $\mathcal{B}_k, k = 1, \ldots, K$.

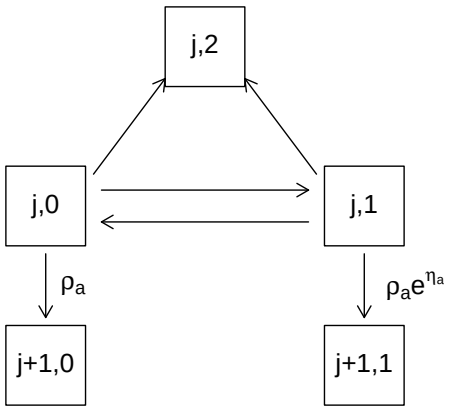


Figure 4.2: A simplified multistate model from Figure 4.1b under independent censoring.

Within each $\mathcal{B}_k$ we need to compute the transition probability matrix for the process $\{Z_i(s), 0 < s\} \,|\, X_{i2}$ depicted in Figure 4.1b. To simplify things and focus on the key features of interest - a marker-dependent visit process, we now consider here a simplified multistate model under the assumption that censoring is independent. It is displayed in Figure 4.2. Here we order the states $\{(j,0), (j,1), (j+1,0), (j+1,1), (j,2)\}$ and consider

the transition intensity matrix $dQ(s \,|\, X_{i2})$ for $s \in \mathcal{B}_k$ as

$$\begin{bmatrix} -dQ_{1\cdot}^{(k)}(X_{i2}) & dQ_{12}^{(k)}(X_{i2}) & dQ_{13}^{(k)}(X_{i3}) & 0 & dQ_{15}^{(k)}(X_{i2}) \\ dQ_{21}^{(k)}(X_{i2}) & -dQ_{2\cdot}^{(k)}(X_{i2}) & 0 & dQ_{24}^{(k)}(X_{i2}) & dQ_{25}^{(k)}(X_{i2}) \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

We suppose sub-interval k is obtained by the intersection of $\mathcal{B}_{0k_0}$, $\mathcal{B}_{1k_1}$ and $\mathcal{B}_{2k_2}$. In this case for the first row of $dQ(s \,|\, X_{i2})$, $s \in \mathcal{B}_k$, we have

$$\begin{aligned} dQ_{12}^{(k)}(X_{i2}) &= \lambda_{0k_0} \exp(X_{i2}'\gamma_{01})\, ds \\ dQ_{13}^{(k)}(X_{i2}) &= \rho_a\, ds \\ dQ_{15}^{(k)}(X_{i2}) &= \lambda_{2k_2} \exp(X_{i2}'\beta_2) ds \end{aligned}$$

where $dQ_{1\cdot}^{(k)}(X_{i2}) = \sum_{\ell \neq 1} dQ_{1\ell}^{(k)}(X_{i2})$. For the second row, we have

$$\begin{aligned} dQ_{21}^{(k)}(X_{i2}) &= \lambda_{1k_1} \exp(X_{i2}'\gamma_{10})\, ds \\ dQ_{24}^{(k)}(X_{i2}) &= \rho_a \exp(\eta_a)\, ds \\ dQ_{25}^{(k)}(X_{i2}) &= \lambda_{2k_2} \exp(\beta_1 + X_{i2}'\beta_2) ds \end{aligned}$$

where again $dQ_{2\cdot}^{(k)}(X_{i2}) = \sum_{\ell \neq 2} dQ_{2\ell}^{(k)}(X_{i2})$. Note that the visit process transition intensity could be time non-homogeneous but we consider a homogeneous (but marker-dependent) intensity here.

We can then write the time-dependent transition intensity matrix as

$$dQ(u \,|\, X_{i2}) = \sum_{k=1}^{K+1} I(u \in \mathcal{B}_k)\, dQ^{(k)}(u \mid X_{i2})\,.$$

With this defined we can proceed to compute the transition probability matrix. If $\mathcal{I}$ is a 5×5 identity matrix then the transition probability matrix for $s \in [b_{k-1}, b_k)$ is obtained

by product integration via

$$P(s,t \mid X_{i2}) = \prod_{(s,t]} \{\mathcal{I} + dQ(u \mid X_{i2})\}$$

(Cook and Lawless, 2018).

To elaborate, if we consider the fine partition of time based on $\mathcal{B}_k, k = 1, \ldots, K$ with break-points labeled $0 = b_0 < b_1 < \cdots < b_{K-1} < b_K = \infty$ we can let $dQ(u \mid X_{i2}) = dQ^{(k)}(X_{i2})$ for $u \in \mathcal{B}_k$ and $P^{(k)}(s,t)$ the corresponding transition probability matrix for $s,t \in \mathcal{B}_k$. Then if s is in the kth sub-interval and t is in the lth interval $(k < l)$, more generally we can write

$$P(s,t \mid X_{i2}) = P^{(k)}(s, b_k \mid X_{i2}) \left[\prod_{j=k+1}^{l-1} P^{(j)}(b_{j-1}, b_j \mid X_{i2}) \right] P^{(l)}(b_{l-1}, t \mid X_{i2}) \tag{4.7}$$

where $P^{(k)}(u, v \,|\, X_{i2})$ is based on the transition probability matrix for the kth interval $b_{k-1} < u < v < b_k$.

Having obtained an expression for the transition probability for the sub-process depicted in Figure 4.1b equivalently Figure 4.2 within any interval $\mathcal{B}_k$ and based on (4.7) we can compute the probability for any sample path of observed data based on (3.12) which we reproduce below for the special case of independent censoring. In terms of full set of intensities we calculate the probability for an arbitrary path here as

$$\mathcal{P}_i = \prod_{(0,\infty)} \mathcal{P}_{i0} \times \left\{ \prod_{j=0}^{1} \mathcal{P}_{ij} \right\} \times \mathcal{P}_{ia} \tag{4.8}$$

where if the realized process is represented as

$$\bar{\mathcal{H}}_i(\infty) = \{\bar{Y}_i(s), d\bar{A}_i(s), d\bar{N}_{i0}(s), d\bar{N}_{i1}(s), d\bar{N}_{i2}(s), 0 < s < V_i, (V_i, \delta_i), X_{i2}\}$$

then we have

$$\mathcal{P}_{i0} = \lambda_i(v_i \,|\, \bar{\mathcal{H}}_i(v_i))^{\delta_i} \exp\left(-\int_0^\infty \bar{Y}_i(u)\lambda_i(u \,|\, \bar{\mathcal{H}}_i(u)du)\right),$$

$$\mathcal{P}_{ij} = \prod_{l=1}^{\bar{N}_{ij}(v_i)} \lambda_{ij}(s_{il} \,|\, \bar{\mathcal{H}}_i(s_{il}), d\bar{N}_{ij}(s_i)) \exp\left(-\int_0^\infty \bar{Y}_{ij}(u)\lambda_{ij}(u \,|\, \bar{\mathcal{H}}_i(u)du)\right), \qquad j = 0, 1,$$

$$\mathcal{P}_{ia} = \prod_{j=1}^{r_i} \lambda_i^a(a_{ij} \,|\, \bar{\mathcal{H}}_i(a_{ij})) \exp\left(-\int_0^\infty \bar{Y}_i(u)\, \lambda_i^a(u \,|\, \bar{\mathcal{H}}_i(u)du)\right),$$

where r_i is the total number of post-baseline visits for individual i and $\bar{N}_{ij}(v_i) = \int_0^\infty \bar{Y}_i(s)dN_{ij}(s)$, $j = 0, 1$. Under conditionally independent censoring $\lambda_i^c(t \,|\, \bar{\mathcal{H}}_i(t)) = \lambda_i^c(t \,|\, \bar{\mathcal{H}}_i^\circ(t))$ and under the further assumption that censoring is non-informative we can omit the term

$$\mathcal{P}_{ic} = \lambda_i^c(v_i \,|\, \bar{\mathcal{H}}_i(v_i))^{1-\delta_i} \exp\left(-\int_0^\infty \lambda_i^c(u \,|\, \bar{\mathcal{H}}_i(u)du)\right),$$

of (3.12) to obtain (4.8). Of course the entire path is not observed since the transitions between the marker states are unknown (both how many there are between assessments, and their times). We therefore consider a full analysis that accommodates the incompleteness of the information which requires computation of the transition probability matrix.

4.2.2 The Transition Probability Matrix

We next discuss the calculation of the transition probability matrix. For convenience we relabel the states $(j, 0), (j, 1), (j + 1, 0), (j + 1, 1)$ and $(j, 2)$ in Figure 4.2 as states 1,2,3,4 and 5 respectively. We then let the transition intensity matrix within interval $\mathcal{B}_k$ given by $dQ^{(k)}(u \mid X_{i2})/du$ be written as

$$\begin{pmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} & 0 & \lambda_{15} \\ \lambda_{21} & \lambda_{22} & 0 & \lambda_{24} & \lambda_{25} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

for the kth sub-interval, where we drop the subscript k denoting the interval for ease of notation the same procedure should be followed within each interval. Here then λ_{13} denotes the intensity of visit for normal marker value and λ_{24} to denote the intensity of visit if marker value is elevated. We let $\lambda_{11} = -(\lambda_{12} + \lambda_{13} + \lambda_{15})$ and $\lambda_{22} = -(\lambda_{21} + \lambda_{24} + \lambda_{25})$. Given the transition matrix for $b_{k-1} < s < t \leq b_k$, we could write $P^{(k)}(s, t \,|\, X_{i2}) = \exp(dQ^{(k)}(X_{i2})/ds\,(t-s))$ as

$$\begin{pmatrix} P_{11}(s,t \,|\, X_{i2}) & P_{12}(s,t \,|\, X_{i2}) & P_{13}(s,t \,|\, X_{i2}) & P_{14}(s,t \,|\, X_{i2}) & P_{15}(s,t \,|\, X_{i2}) \\ P_{21}(s,t \,|\, X_{i2}) & P_{22}(s,t \,|\, X_{i2}) & P_{23}(s,t \,|\, X_{i2}) & P_{24}(s,t \,|\, X_{i2}) & P_{25}(s,t \,|\, X_{i2}) \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{4.9}$$

Here, the transition probabilities in (4.9) share a common denominator which we denote by $P^D(s, t \,|\, X_{i2})$. If we let

$$\lambda^* = \sqrt{\lambda_{12}^2 + (2\,\lambda_{13} + 2\,\lambda_{15} + 2\,\lambda_{21} - 2\,\lambda_{23} - 2\,\lambda_{25})\,\lambda_{12} + (\lambda_{13} + \lambda_{15} + \lambda_{22})^2}\,,$$

then $P^D(s, t \,|\, X_{i2})$ is given by

$$P^D(s,t \,|\, X_{i2}) = \lambda^*\,(-\lambda_{11} - \lambda_{22} + \lambda^*)^3\,(-\lambda_{11} - \lambda_{22} - \lambda^*)^3$$

The (1,1) element of (4.9) is given by $P_{11}(s,t\,|\,X_{i2}) = P_{11}^N(s,t\,|\,X_{i2})/P^D(s,t\,|\,X_{i2})$, where

$$P_{11}^N(s,t\,|\,X_{i2}) = -32\,\left(\left(-\lambda_{11}+\lambda_{22}-\lambda^*\right)\mathrm{e}^{-1/2\,(-\lambda_{11}-\lambda_{22}-\lambda^*)(t-s)}\right.$$
$$\left.-\,\mathrm{e}^{-1/2\,(-\lambda_{11}-\lambda_{22}+\lambda^*)(s-t)}\left(-\lambda_{11}+\lambda_{22}+\lambda^*\right)\right)\left(\left(-\lambda_{22}(\lambda_{13}+\lambda_{15})+\left(\lambda_{24}+\lambda_{25}\right)\lambda_{12}\right)^3\right.\,.$$

For $P_{12}(s,t\,|\,X_{i2}) = P_{12}^N(s,t\,|\,X_{i2})/P^D(s,t\,|\,X_{i2})$, the numerator $P_{12}^N(s,t\,|\,X_{i2})$ is given by

$$64\,\lambda_{12}\left(\left(\lambda_{24}+\lambda_{25}\right)\lambda_{12}-\left(\lambda_{22}\right)\left(\lambda_{14}+\lambda_{15}\right)\right)^3\left(\mathrm{e}^{-1/2\,(-\lambda_{11}-\lambda_{22}-\lambda^*)t}-\mathrm{e}^{-1/2\,(-\lambda_{11}-\lambda_{22}+\lambda^*)(t-s)}\right)\,.$$

Then for $P_{13}(s,t\,|\,X_{i2}) = P_{13}^N(s,t\,|\,X_{i2})/P^D(s,t\,|\,X_{i2})$ we obtain $P_{13}^N(s,t\,|\,X_{i2})$ as

$$-\,32\,\left(\left(\lambda_{14}+\lambda_{15}\right)\lambda_{21}-\left(\lambda_{24}+\lambda_{25}\right)\lambda_{11}\right)^2$$
$$\left(-\lambda_{22}\left(\mathrm{e}^{-1/2\,(-\lambda_{11}-\lambda_{22}-\lambda^*)(t-s)}+\mathrm{e}^{-1/2\,(-\lambda_{11}-\lambda_{22}+\lambda^*)(t-s)}-2\right)\lambda^*+\right.$$
$$\left({\lambda_{21}}^2+\left(\lambda_{12}-\lambda_{13}-\lambda_{15}+2\,\lambda_{24}+2\,\lambda_{25}\right)\lambda_{21}-\left(\lambda_{24}+\lambda_{25}\right)\left(-\lambda_{11}-\lambda_{24}-\lambda_{25}\right)\right)$$
$$\left(\mathrm{e}^{-1/2\,(-\lambda_{11}-\lambda_{22}-\lambda^*)(t-s)}-\mathrm{e}^{-1/2\,(-\lambda_{11}-\lambda_{22}+\lambda^*)(t-s)}\right)\lambda_{13}\,.$$

Next $P_{14}(s,t\,|\,X_{i2}) = P_{14}^N(s,t\,|\,X_{i2})/P^D(s,t\,|\,X_{i2})$ where

$$P_{14}(s,t\,|\,X_{i2}) = -32\,\lambda_{12}\left(\left(-\lambda_{11}\right)\lambda_{24}+\left(\lambda_{21}+\lambda_{25}\right)\lambda_{13}+\left(\lambda_{21}+\lambda_{25}\right)\lambda_{15}\right.$$
$$\left.+\,\lambda_{12}\lambda_{25}\right)^2\left(\left(-\lambda_{11}-\lambda_{22}+\lambda^*\right)\mathrm{e}^{-1/2\,(-\lambda_{11}-\lambda_{22}-\lambda^*)(t-s)}\right.$$
$$\left.+\left(\lambda_{11}+\lambda_{22}+\lambda^*\right)\mathrm{e}^{-1/2\,(-\lambda_{11}-\lambda_{22}+\lambda^*)(t-s)}-2\,\lambda^*\right)\lambda_{24}\,.$$

And finally $P_{15}(s,t\,|\,X_{i2}) = P_{15}^N(s,t\,|\,X_{i2})/P^D(s,t\,|\,X_{i2})$ where

$$P_{15}^N(s,t\,|\,X_{i2}) = -32\,\left(-\lambda_{22}(\lambda_{13}+\lambda_{15})+\left(\lambda_{24}+\lambda_{25}\right)\lambda_{12}\right)^2\left(\left(-\lambda_{22}\lambda_{15}+\lambda_{25}\lambda_{12}\right)\right.$$
$$\left(\mathrm{e}^{-1/2\,(-\lambda_{11}-\lambda_{22}-\lambda^*)(t-s)}+\mathrm{e}^{-1/2\,(-\lambda_{11}-\lambda_{22}+\lambda^*)(t-s)}-2\right)\lambda^*$$
$$+\left(\lambda_{22}\lambda_{15}^2+\left(\left(\lambda_{21}-\lambda_{24}\right)\lambda_{12}+\lambda_{22}\left(\lambda_{13}+\lambda_{22}\right)\right)\lambda_{15}+\lambda_{12}\lambda_{25}\left(\lambda_{12}+\lambda_{13}-\lambda_{22}\right)\right)$$
$$\left.\left(\mathrm{e}^{-1/2\,(-\lambda_{11}-\lambda_{22}-\lambda^*)(t-s)}-\mathrm{e}^{-1/2\,(-\lambda_{11}-\lambda_{22}+\lambda^*)(t-s)}\right)\right)\,.$$

Similarly, for the second row of the transition probability matrix (4.9), we have $P_{21}(s,t\,|\,X_{i2})$ $= P_{21}^N(s,t\,|\,X_{i2})/P^D(s,t\,|\,X_{i2})$ where

$$P_{21}(s,t\,|\,X_{i2}) = 64\,\left(\left(\lambda_{13}+\lambda_{15}\right)\lambda_{21}\right.$$
$$\left.-\,\left(\lambda_{24}+\lambda_{25}\right)\lambda_{11}\right)^3\left(\mathrm{e}^{-1/2\,(-\lambda_{11}-\lambda_{22}-\lambda^*)(t-s)}-\mathrm{e}^{-1/2\,(-\lambda_{11}-\lambda_{22}+\lambda^*)(t-s)}\right)\lambda_{21}$$

Next $P_{22}(s,t\,|\,X_{i2}) = P_{22}^N(s,t\,|\,X_{i2})/P^D(s,t\,|\,X_{i2})$ where

$$P_{22}^N(s,t\,|\,X_{i2}) = 32\,\Big((\lambda_{22}-\lambda_{11}+\lambda^*)\,\mathrm{e}^{-1/2\,(-\lambda_{11}-\lambda_{22}-\lambda^*)(t-s)} - \mathrm{e}^{-1/2\,(-\lambda_{11}-\lambda_{22}+\lambda^*)(t-s)}\,(\lambda_{22}-\lambda_{11}-\lambda^*)\Big)\,(-\lambda_{22}(\lambda_{13}+\lambda_{15})+(\lambda_{24}+\lambda_{25})\,\lambda_{12})^3\,,$$

and $P_{23}(s,t\,|\,X_{i2}) = P_{23}^N(s,t\,|\,X_{i2})/P^D(s,t\,|\,X_{i2})$ where $P_{23}^N(s,t\,|\,X_{i2})$ is given by

$$-32\,(-\lambda_{22}(\lambda_{13}+\lambda_{15})+(\lambda_{24}+\lambda_{25})\,\lambda_{12})^2\,\Big((-\lambda_{11}-\lambda_{22}+\lambda^*)\,\mathrm{e}^{-1/2\,(-\lambda_{11}-\lambda_{22}-\lambda^*)(t-s)} + (\lambda_{11}+\lambda_{22}+\lambda^*)\,\mathrm{e}^{-1/2\,(-\lambda_{11}-\lambda_{22}+\lambda^*)(t-s)} - 2\,\lambda^*\Big)\,\lambda_{13}\lambda_{21}\,.$$

Then $P_{24}(s,t\,|\,X_{i2}) = P_{24}^N(s,t\,|\,X_{i2})/P^D(s,t\,|\,X_{i2})$ with

$$P_{24}^N(s,t\,|\,X_{i2}) = -32\,((\lambda_{24}+\lambda_{25})\,\lambda_{12} - \lambda_{22}(\lambda_{13}+\lambda_{15}))^2\,\lambda_{24}\,\Big(-\lambda_{11}\,\Big(\mathrm{e}^{-1/2\,(-\lambda_{11}-\lambda_{22}-\lambda^*)(t-s)} + \mathrm{e}^{-1/2\,(-\lambda_{11}-\lambda_{22}+\lambda^*)(t-s)} - 2\Big)\,\lambda^* + \Big(\lambda_{12}^2 + (2\,\lambda_{13}+2\,\lambda_{15}+\lambda_{21}-\lambda_{24}-\lambda_{25})\,\lambda_{12} + (\lambda_{13}+\lambda_{15})\,(\lambda_{13}+\lambda_{15}+\lambda_{22})\Big)\,\Big(\mathrm{e}^{-1/2\,(-\lambda_{11}-\lambda_{22}-\lambda^*)(t-s)} - \mathrm{e}^{-1/2\,(-\lambda_{11}-\lambda_{22}+\lambda^*)(t-s)}\Big)\Big)$$

And finally $P_{25}(s,t\,|\,X_{i2}) = P_{25}^N(s,t\,|\,X_{i2})/P^D(s,t\,|\,X_{i2})$ where

$$P_{25}^N(s,t\,|\,X_{i2}) = 32\,(-\lambda_{11}(\lambda_{24}+\lambda_{25})+(\lambda_{13}+\lambda_{15})\,\lambda_{21})^2\,\Big(-\,(-\lambda_{11}\lambda_{25}+\lambda_{21}\lambda_{15})\,\Big(\mathrm{e}^{-1/2\,(-\lambda_{11}-\lambda_{22}-\lambda^*)(t-s)} + \mathrm{e}^{-1/2\,(-\lambda_{11}-\lambda_{22}+\lambda^*)(t-s)} - 2\Big)\,\lambda^* + \Big(\lambda_{11}{\lambda_{25}}^2 + ((\lambda_{12}-\lambda_{13})\,\lambda_{21} - \lambda_{11}\,(-\lambda_{11}-\lambda_{24}))\,\lambda_{25} + \lambda_{15}\lambda_{21}\,(-\lambda_{11}+\lambda_{21}+\lambda_{24})\Big)\,\Big(\mathrm{e}^{-1/2\,(-\lambda_{11}-\lambda_{22}+\lambda^*)(t-s)} - \mathrm{e}^{-1/2\,(-\lambda_{11}-\lambda_{22}-\lambda^*)(t-s)}\Big)\Big)$$

4.2.3 The Observed Data Likelihood for the Joint Model

Having obtained the form of the transition probability matrix over an arbitrary pair of times $s < t$, we can now give the likelihood for the data when the marker is under intermittent observation. If an individual i is seen at times $0 = a_{i0} < a_{i1} < \ldots < a_{ir_i}$ and we observe $X_{i1}(a_{ij})$ at each of these visits, if we then let $V_i = \min(T_i, C_i)$ and $\delta_i = I(V_i = T_i)$ denote the censored outcome data following the last assessment, the observed data likelihood is given by

$$\mathcal{L}_i \propto \Big\{ \prod_{j=1}^{r_i} P(\mathcal{Z}_i(a_{ij}) = X_{i1}(a_{ij}), A_i(a_{ij}) = j-1\,|\,\bar{\mathcal{H}}_i^\circ(a_{ij}^-))\,\lambda_i^a(a_{ij}\,|\,X_{i1}(a_{ij}^-),\,\bar{\mathcal{H}}_i^\circ(a_{ij}^-))\Big\}$$
$$\sum_{l=0}^{1} P(Z_i(V_i) = (r_i,l)\,|\,A_i(V_i^-) = r_i, \bar{\mathcal{H}}_i^\circ(V_i^-))\,(\lambda_2(V_i)\exp(\beta_1 + X_{i2}'\beta_2))^{\delta_i}$$
$$(\lambda_c(V_i)\exp(\eta_c))^{1-\delta_i}$$

where the probabilities are computed based on the multistate model in Figure 4.1b and the piecewise constant assumptions via (4.7). If censoring is marker-independent then the last term can be omitted and $\{\mathcal{Z}(s), 0 < s\}$ process can be viewed as the simplified multistate model depicted in Figure 4.2.

4.3 Simulation Studies

4.3.1 Design of Simulation Studies

In the simulation study, individuals are under observation over the interval from time 0 to the minimum of event and administrative right-censoring time. We first solve the parameters for the multistate process. We let $\lambda_0 = 4$ and let $\lambda_1 = 2\lambda_0$=8. We then solve for the values of λ_2 under the further constraint that the elevated marker value increases the hazard for failure by 25% and the probability of an event at the administrative censoring time $P(Z_i(1) = 2 \,|\, Z_i(0) = 1) = 60\%$. We set the covariate effect of the failure process to be $\beta_1 = \log 1.25$ and $\beta_2 = \log 1.5$. We let the dependency of the marker process when the marker is elevated to be $\eta_a = \log 1.2$. We considered two sets of values for the covariate effect on marker process $\gamma_0 = \log 1.2$, $\gamma_1 = \log 0.8$ and $\gamma_0 = \log 1.5$, $\gamma_1 = \log 0.5$.

We put 2 cut-points at 1/3 and 2/3 for the marker process and 4 cut-points at 0.2, 0.4, 0.6, 0.8 for the failure process. For the joint analysis, we perform 200 iterations of dataset of 500 individuals each, and we compare the estimates we get from this joint analysis with the estimates from the naive Cox model and naive piecewise constant model when the conventional analysis is used.

4.3.2 Simulation Results

The results of the simulation studies are displayed in Table 4.1 for the setting where there is a modest covariate effect on the marker transition intensities and Table 4.2 for the setting of a stronger effect of X_2 on the marker transitions. We find with the relatively infrequent vistis ($\rho_a = 4$) the bias from the Cox model and the piecewise constant model is appreciable - we know from the theoretical results that these biases are due to both

the carrying forward of the marker values as well as the marker-dependent visit process whereby visits occur with a higher intensity when the marker state is elevated. The joint analysis in contrast tends to yield estimators with negligible empirical bias. There is a higher empirical standard error for the marker effect on failure in the joint analysis but this price is paid to mitigate the bias from the misspecified model. The parameters of the marker process intensities are also generally well-estimated, as is the intensity for the visit process, including the parameter η_a reflecting the extent of the marker-dependence. Note these are well estimated because the visit times are observed precisely and there are many visit times informing these parameter estimates.

		Cox		PWC		Joint Analysis	
Parameter	TRUE	EBIAS	ESE	EBIAS	ESE	EBIAS	ESE
Failure Process							
β_1	0.2231	-0.1330	0.1882	-0.1401	0.1681	0.0009	0.2262
β_2	0.4055	0.0012	0.2043	0.0395	0.1481	-0.0150	0.1019
$\log \lambda_{21}$	-0.1698					-0.0056	0.1361
$\log \lambda_{22}$	-0.1698					0.0080	0.1288
$\log \lambda_{23}$	-0.1698					0.0063	0.1513
$\log \lambda_{24}$	-0.1698					0.0089	0.1535
$\log \lambda_{25}$	-0.1698					0.0020	0.1854
Marker Process							
$\log \lambda_{01}$	1.3863					0.0346	0.1731
$\log \lambda_{02}$	1.3863					0.0713	0.2182
$\log \lambda_{03}$	1.3863					0.0590	0.2185
γ_0	0.1823					-0.0195	0.2010
$\log \lambda_{11}$	2.0794					0.0285	0.1437
$\log \lambda_{12}$	2.0794					0.0658	0.2180
$\log \lambda_{13}$	2.0794					0.0836	0.2087
γ_1	-0.2231					-0.0023	0.1880
Visit Process							
$\log \rho_a$	1.3863					0.0001	0.0485
η_a	0.1823					0.0009	0.1054

Table 4.1: Parameter estimates for naive Cox, naive piecewise constant model and joint analysis. There are 5 cut-points for the failure model and 3 cut-points for marker process. Number of individuals per simulation =500 and number of simulation = 200. $\rho_a = 4$, $\gamma_0 = \log 1.2$ and $\gamma_1 = \log 0.8$.

		Cox		PWC		Joint Analysis	
Parameter	TRUE	EBIAS	ESE	EBIAS	ESE	EBIAS	ESE
Failure Process							
β_1	0.2231	-0.1227	0.1762	-0.1221	0.1611	-0.0446	0.1991
β_2	0.4055	0.0420	0.1766	0.0573	0.1643	0.0118	0.1181
$\log \lambda_{21}$	-0.1698					-0.0001	0.1233
$\log \lambda_{22}$	-0.1698					0.0415	0.1154
$\log \lambda_{23}$	-0.1698					0.0168	0.1450
$\log \lambda_{24}$	-0.1698					-0.0076	0.1473
$\log \lambda_{25}$	-0.1698					0.0086	0.1692
Marker Process							
$\log \lambda_{01}$	1.3863					-0.0034	0.1808
$\log \lambda_{02}$	1.3863					0.0275	0.2047
$\log \lambda_{03}$	1.3863					0.0480	0.2228
γ_0	0.4055					0.0167	0.1657
$\log \lambda_{11}$	2.0794					0.0285	0.1437
$\log \lambda_{12}$	2.0794					0.0469	0.2064
$\log \lambda_{13}$	2.0794					0.0812	0.2509
γ_1	-0.6931					0.0025	0.2295
Visit Process							
$\log \rho_a$	1.3863					-0.0105	0.0463
η_a	0.1823					0.0268	0.0802

Table 4.2: Parameter estimates for naive Cox, naive piecewise constant model and joint analysis. There are 5 cut-points for the failure model and 3 cut-points for marker process. Number of individuals per simulation =500 and number of simulation = 200. $\rho_a = 4$, $\gamma_0 = \log 1.5$ and $\gamma_1 = \log 0.5$.

4.4 Application

4.4.1 Application to the Psoriatic Arthritis Data

The likelihood-based approach of joint modeling was applied on patients with psoriatic arthritis (PSA). In this study, interest lies in evaluating the effect of inflammatory markers on severity and progression of joint damage. Patients are assessed both by X-ray for the level of joint damage and through clinic visits for values of markers. During radiological examinations, a damage score from 0 to 5 is evaluated for each joint, where score 0 means a normal joint and score 5 means the joint needs surgery. In the application, we consider joint with grade 1 or higher as damaged joints and we define patients with greater than or equal to 3 damaged joints as failure. For clinic visits, blood samples are taken to get values of biomarkers. We take the genetic marker HLA-B27 as fixed covariate and consider the level of ESR as a time-varying marker, where an ESR value of less than or equal to 20 mm/hr for female, and less than or equal to 13 mm/hr is considered normal. We define patients with normal ESR value in state 0 and abnormal ESR value in state 1. Note that the clinic visits and x-ray visits do not necessarily happen at the same time, and we use the age at the first clinic visit as time origin so each individual has a visit at time 0 for our data. There are 2 cut-points at 2 and 6 years for the marker process and 4 cut-points at 2, 4, 6 and 8 years for the failure process. As in Chapter 2, we sample from the integers 1 to n with replacement n times to create a bootstrap sample and then compute the bootstrap standard error as the square root of the empirical variance of the estimates.

4.4.2 Application Result

The results of the analysis for both independent visit process ($\eta_a = 0$) and dependent visit process are provided in Table 4.3, where we see that elevation of the erythrocyte sedimation marker does have a significant impact on the failure process but the presence of HLA-B27 does not, (HR = 1.13, 95% CI: 0.87, 1.42; p=0.3879) for independent visit and (HR = 1.08, 95% CI: 0.85, 1.38; p=0.5216) for dependent visit process. We also see that those individuals who are HLA-B27 positive have a statistically significant lower intensity for

a transition to the elevated ESR state (HR = 0.51, 95% CI: 0.30, 0.88; p=0.0149) and a lower, but not statistically significant transition intensity to the normal ESR state (HR = 0.75, 95% CI: 0.45, 1.25; p=0.2720) for dependent visit process. Finally we see that there does appear to be a significant effect of the marker process on the visit intensity since $\exp(\widehat{\eta}_a) = 0.63$ with 95% CI given by (0.48, 0.82) and p=0.0004.

4.5 Future Research

This chapter reports on an innovative joint modeling approach which deals with an intermittently observed binary time-dependent marker for the setting where the value of the marker affects the intensity of the visit process. Here the marker and failure intensities were piecewise constant and the visit intensity was taken to be time homogeneous. The visit intensity could be relaxed to take a piecewise constant form as well and like the failure process, it could accommodate a large number of pieces because the visit times are known precisely. An interesting area of future research would be to extend the visit intensity model to accommodate an individual specific random effect to accommodate unexplained heterogeneity in the propensity for visits in the spirit of the mixed Poisson model described in Section 1.3.

The approach taken to estimation in this chapter was maximization of the observed data likelihood, but the expanded state space of Appendix D lays the foundation for an EM algorithm in a multistate setting which accommodates a marker-dependent visit process by incorporating information on the state occupied at the time of each visit through specification of a separate marker-visit occurrence state. It is in this framework that a random effect model for the visit intensity would most naturally be fitted; a gamma distributed random effect may be most natural. Joint modeling requires stronger model assumptions than standard analyses and this investment in model assumptions is worthwhile to mitigate the bias from the last observation carried forward approach. However the model assumptions may be incorrect in which case different types of biases are induced. Model checking in the setting of an intermittently observed marker and a right-censored failure time process are

	Independent Visit			Dependent Visit		
Parameter	Est.	ESE	p-value	Est.	ESE	p-value
Failure Process						
β_1	0.5784	0.1147	< 0.0001	0.3698	0.1583	0.0195
β_2	0.1076	0.1246	0.3879	0.0803	0.1253	0.5216
$\log \lambda_{21}$	-1.8626	0.1036	< 0.0001	-1.7720	0.1068	< 0.0001
$\log \lambda_{22}$	-1.9322	0.0946	< 0.0001	-1.8703	0.1072	< 0.0001
$\log \lambda_{23}$	-2.4029	0.1360	< 0.0001	-2.3395	0.1279	< 0.0001
$\log \lambda_{24}$	-2.4575	0.1629	< 0.0001	-2.3971	0.1812	< 0.0001
$\log \lambda_{25}$	-2.4866	0.0959	< 0.0001	-2.4323	0.1074	< 0.0001
Marker Process						
$\log \lambda_{01}$	-0.9183	0.1218	< 0.0001	-0.7656	0.1380	< 0.0001
$\log \lambda_{02}$	-0.9885	0.1281	< 0.0001	-0.8681	0.1060	< 0.0001
$\log \lambda_{03}$	-1.1702	0.2113	< 0.0001	-1.0270	0.2206	< 0.0001
γ_0	-0.8166	0.3028	0.0070	-0.6596	0.2708	0.0149
$\log \lambda_{11}$	-0.2721	0.1223	0.0261	-0.4656	0.1359	0.0006
$\log \lambda_{12}$	-0.4486	0.1398	0.0013	-0.6575	0.1513	< 0.0001
$\log \lambda_{13}$	-0.8564	0.2600	0.0010	-1.1844	0.2808	< 0.0001
γ_1	-0.2419	0.2845	0.3952	-0.2836	0.2582	0.2720
Visit Process						
$\log \rho_a$	-0.1879	0.0467	0.0001	-0.0024	0.0562	0.9659
η_a				-0.4608	0.1312	0.0004

Table 4.3: Table of parameter estimates for joint analysis for applying to the PsA data for both independent visit ($\eta_a = 0$) and dependent visit process.

warranted, as are methods for conducting sensitivity analyses to investigate the possible impact of model misspecification. The types of misspecification of possible interest include misspecification of the marker intensities due to incorrect dependence modeling, incorrect time-scale, or unexplained heterogeneity and misspecification of the visit intensity for the same reasons. Model checking through fitting expanded models (or carrying out score tests for the need for expanded models) is an important area of future research.

As discussed in Section 2.7.2 it would be natural to extend this analysis approach to deal with interval-censored failure time data and settings with more complex (e.g. multicategory or multi-type) marker processes; for the former setting, the visits at which the failure status is determined may also be different from those at which the marker is measured. More elaborate responses such as clustered failure time processes, recurrent event processes, or progressive damage processes are also of interest.

Chapter 5

Summary and Future Research

5.1 A Review

This thesis has been concerned with the analysis of failure times and the examination of the relationship between a dynamic discrete biomarker and fixed covariates and the failure time. There is a long history on the development of statistical methods for the analysis of failure time data with time dependent covariates, but the vast majority of these methods presume the dynamic covariates are under continuous observation. This is not true in practice, as the time-dependent covariates representing markers measuring the severity of disease activity or more broadly the state of the underlying disease condition, can only be measured at periodic assessment times. The values at these assessment times are typically carried forward and presumed to be accurate measurements of the marker until the next assessment is made. A major contribution of this thesis is the development of a framework to investigate the biases that arise from this conventional approach to analysis. We find there is typically an attenuation in the regression coefficient for the time dependent marker and an associated smaller bias for the effect of any fixed covariates with the conventional approach.

In Chapter 2 we describe an EM algorithm that facilitates fitting a joint model for a discrete marker process and failure process which mitigates the biases arising from conventional analyses involving the carry-forward approach to the intermittently observed marker

process. The EM algorithm yields maximum likelihood estimators which are therefore consistent when modelling assumptions are valid. We restrict attention to a single binary time-dependent covariate but in principle the algorithm is naturally adapted to deal with more than two categories for the marker, or even multiple time-dependent markers. The EM algorithm accommodates a different number of cut points for the failure process which is under right censored observation and the marker process which is in completely observed measurements. As the number of pieces in the piecewise constant intensity for the failure time increases, this method better approximates the results of the semi-parametric regression model.

There is a price to pay when joint modelling the marker process and the failure process in the sense that the standard error of the regression coefficients can be increasing due to the need to estimate the parameters governing the marker process and there would be more variability of the model when the number of parameters estimated increases. We therefore conduct a careful examination of the determinants of the bias from the conventional carry forward approach in order to give some guidance on when joint modelling may be warranted. The framework we develop in Chapter 3 makes use of the theory of the misspecified models and specifically misspecified Cox regression models. We find, quite naturally, that the less frequent the visits leading to the measurements of the markers, the greater the bias from the carry forward approach. The framework was also exploited to investigate the consequences of a marker-dependent visit process. Often markers reflecting disease severity will be outdated in the time of examination if the symptoms lead individuals to seek medical care. We therefore identify a second source of bias in this situation if simple joint models are used.

In Chapter 4 we develop a likelihood-based approach to joint modelling of the marker process, failure process, visit process and the censoring process. This joint model enables one to mitigate the bias in the regression coefficients from a marker-dependent visit process and a marker-dependent loss to follow-up. The joint model involves the specifications of marker states, failure states, censoring states, and states based on the cumulative number of visits at which markers are assessed. The visit process intensities and censoring intensi-

ties can depend on the latent marker status. This investigation is carried out in the context of a Markov model. Piecewise constant intensities are specified for all transitions within this large joint model, and likelihood based inferences proposed. Existing software requires partitioning of the state space into smaller state spaces corresponding to information on the joint process between assessment times or between the final assessment time and failure. Because visit times, failure times and censoring times are observed precisely, these intensities may involve a large number of break-points, or even be estimated semi-parametrically. We focus primarily on a small number of pieces for the failure process, conditionally independent and non-informative right censoring, and a marker-modulated time homogeneous visit process intensity. Simulation studies confirm that the model has good performance and application to the motivating data set illustrates how different estimators can arise between the naive model and the joint model.

5.2 Future Work: Interval-Censored Failure Times

In many circumstances the failure process may also be under intermittent observation. This arises in osteoporosis studies when asymptomatic fractures could not be detectable other than through x-ray or bone scans. In these situations it is at the periodic assessment times that the failure status of individuals is determined which leads to type K interval-censored data (Sun, 2007). Sometimes the assessment times for the failure process will be the same as the assessment times for the marker process, but often they will also different. Blood samples need not be obtained at the same time as an x-ray or bone scan. In these settings it is of interest to jointly model the visit process for the time-dependent covariates and the visit process for the assessment of the failure status. These more elaborate joint models can be characterized by state spaces with many more states - while this may look unwieldy there will be many constraints that can reduce the number of parameters to be estimated and make the model possible to fit.

Cook and Lawless (2019) consider the situation where there may be periodic visits that are scheduled and others that are disease driven. In this context there may be an informal

schedule for the assessment of markers but elevated markers arising due to exacerbation of symptoms could lead to premature visits that the patient initiated. In such situations, point process model could be considered with one type of event corresponding to a routine visit and another type of event corresponding to a symptom driven visit. To date no such models have been considered to our knowledge and this represents an area of future research.

5.3 Heterogeneity and Truncated Data

Throughout this thesis we have assumed that the model intensities can be characterized based on either observable covariates at the baseline, or a time dependent marker process. On many occasions such models will be inadequate because there is unexplained person-to-person variation in the visit process, the dynamics of the marker process, or the failure process itself. The latter is difficult to deal with but it is possible to introduce random effects for characterizing variation from individual to individual beyond that which would be expected based on the available information. Such a random effect model could accommodate more rapid or slower than average transitions between the marker states, or more variation in the visit process than would be expected based on a modulated Markov model. The development of likelihood methods and estimation approaches for these expanded models is worthwhile both to accommodate settings where there is an additional component of variation, and to form a basis for score tests of the need for model expansion.

Different truncation schemes could be used to screen or exclude individuals from a population upon sampling. Left truncation arises when individuals are required to be event-free at a time of sampling; that is if an individual is contacted at time A_0 and they are only selected if $T > A_0$ where T is the failure time of interest, A_0 is the left-truncation time. Right truncation, on the other hand, occurs when we only sample individuals if, at a contact time R, they are recruited if $T < R$: in this case, R is the right truncation time. For example, in studies of the HIV incubation distribution, we may only select individuals retrospectively based on whether they have developed AIDS; the incubation

period is then right-truncated (Kalbfleisch and Lawless, 1989; Lagakos et al., 1988). In each of these settings, bias could arise if analysis do not account for the fact that individuals are selected based on conditions related to their failure time. There has been a long literature on time-to-event data under truncation. Turnbull (1976) extended Kaplain-Meier estimator to truncation data in the presence of censoring. The asymptotic properties of the product limit estimators were explored in several papers (Wang et al., 1986; Woodroofe et al., 1985; Keiding and Gill, 1990). Inverse probability weighted estimators are consistent under truncation, and the Kaplan-Meier estimator can be viewed as an IPW estimator in this case (Horvitz and Thompson, 1952; Shen, 2003). Regression modeling can be adopted to handle time-to-event data under truncation (Bhattacharya et al., 1983; Gross and Lai, 1996). Expansion of the joint models to accommodate truncation is an area for future work.

References

Al-Kateb, H., Boright, A. P., Mirea, L., Xie, X., Sutradhar, R., Mowjoodi, A., Bharaj, B., Liu, M., Bucksa, J. M., Arends, V. L., et al. (2008). Multiple superoxide dismutase 1/splicing factor serine alanine 15 variants are associated with the development and progression of diabetic nephropathy: the diabetes control and complications trial/epidemiology of diabetes interventions and complications genetics study. *Diabetes*, 57(1):218–228.

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., Abildstrom, S. Z., and Rosthøj, S. (2002). Competing risks as a multistate model. *Statistical methods in medical research*, 11(2):203–215.

Andersen, P. K. and Liestøl, K. (2003). Attenuation caused by infrequently updated covariates in survival analysis. *Biostatistics*, 4(4):633–649.

Bhattacharya, P., Chernoff, H., Yang, S., et al. (1983). Nonparametric estimation of the slope of a truncated regression. *The Annals of Statistics*, 11(2):505–514.

Cook, R. J. (2014). Statistical models for disease processes: Markers and skeletal complications in cancer metastatic to bone. *Statistics in Action: A Canadian Outlook*, page 177.

Cook, R. J. and Lawless, J. (2007). *The Statistical Analysis of Recurrent Events*. Springer Science & Business Media.

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. (2019). Independence conditions and the analysis of life history studies with intermittent observation. *Biostatistics*.

Cox, D. and Miller, H. (1965). *The Theory of Stochastic Processes*. Number v. 10 in The Theory of Stochastic Processes. Wiley.

Dantan, E., Joly, P., Dartigues, J.-F., and Jacqmin-Gadda, H. (2011). Joint model with latent state for longitudinal and multistate data. *Biostatistics*, 12(4):723–736.

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.

Demers, L., Costa, L., Chinchilli, V., Gaydos, L., Curley, E., and Lipton, A. (1995). Biochemical markers of bone turnover in patients with metastatic bone disease. *Clinical chemistry*, 41(10):1489–1494.

Dempster, A., Laird, N., and Rubin, D. (1979). J. royal stat. soc.

Efron, B. and Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.

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

Ferrer, L., Rondeau, V., Dignam, J., Pickles, T., Jacqmin-Gadda, H., and Proust-Lima, C. (2016). Joint modelling of longitudinal and multi-state processes: application to clinical progressions in prostate cancer. *Statistics in medicine*, 35(22):3933–3948.

Gentleman, R., Lawless, J., Lindsey, J., and Yan, P. (1994). Multi-state markov models for analysing incomplete disease history data with illustrations for hiv disease. *Statistics in medicine*, 13(8):805–821.

Gil, J., Preux, P.-M., Alioum, A., Ketzoian, C., Desport, J.-C., Druet-Cabanac, M., and Couratier, P. (2007). Disease progression and survival in als: First multi-state model approach. *Amyotrophic lateral sclerosis*, 8(4):224–229.

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.

Gross, S. T. and Lai, T. L. (1996). Nonparametric estimation and regression analysis with left-truncated and right-censored data. *Journal of the American Statistical Association*, 91(435):1166–1180.

Haroon, M., Gallaghar, P., Ahmad, M., and FitzGerald, O. (2020). Elevated crp even at the first visit to a rheumatologist is associated with long-term poor outcomes in patients with psoriatic arthritis. *Clinical rheumatology*.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72(358):320–338.

Heagerty, P. J. (2002). Marginalized transition models and likelihood inference for longitudinal categorical data. *Biometrics*, 58(2):342–351.

Herring, A. H. and Ibrahim, J. G. (2001). Likelihood-based methods for missing covariates in the cox proportional hazards model. *Journal of the American Statistical Association*, 96(453):292–302.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, 47(260):663–685.

Jackson, C. H. (2011). Multi-state models for panel data: The msm package for R. *Journal of Statistical Software*, 38(8):1–29.

Jewell, N. P. and Kalbfleisch, J. (1996). Marker processes in survival analysis. *Lifetime Data Analysis*, 2(1):15–29.

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.

Kalbfleisch, J. and Lawless, J. F. (1989). Inference based on retrospective ascertainment: an analysis of the data on transfusion-related aids. *Journal of the American Statistical Association*, 84(406):360–372.

Keiding, N. and Gill, R. D. (1990). Random truncation models and markov processes. *The Annals of Statistics*, pages 582–602.

Keiding, N., Klein, J. P., and Horowitz, M. M. (2001). Multi-state models and outcome prediction in bone marrow transplantation. *Statistics in medicine*, 20(12):1871–1885.

Klein, J. P., Szydlo, R. M., Craddock, C., and Goldman, J. M. (2000). Estimation of current leukaemia-free survival following donor lymphocyte infusion therapy for patients with leukaemia who relapse after allografting: application of a multistate model. *Statistics in medicine*, 19(21):3005–3016.

Lagakos, S. W., Barraj, L. M., and Gruttola, V. d. (1988). Nonparametric analysis of truncated survival data, with application to aids. *Biometrika*, 75(3):515–523.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, pages 963–974.

Lawless, J. F. (1987). Negative binomial and mixed poisson regression. *Canadian Journal of Statistics*, 15(3):209–225.

Louis, T. A. (1982). Finding the observed information matrix when using the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(2):226–233.

McEvoy, J. W., Chen, Y., Rawlings, A., Hoogeveen, R. C., Ballantyne, C. M., Blumenthal, R. S., Coresh, J., and Selvin, E. (2016). Diastolic blood pressure, subclinical myocardial damage, and cardiac events: implications for blood pressure control. *Journal of the American College of Cardiology*, 68(16):1713–1722.

Meira-Machado, L., De Una-Alvarez, J., and Cadarso-Suarez, C. (2006). Nonparametric estimation of transition probabilities in a non-markov illness–death model. *Lifetime Data Analysis*, 12(3):325–344.

Mitchell, D. M., Caksa, S., Joseph, T., Bouxsein, M. L., and Misra, M. (2020). Elevated hba1c is associated with altered cortical and trabecular microarchitecture in girls with type 1 diabetes. *The Journal of Clinical Endocrinology & Metabolism*, 105(4):e1648–e1656.

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.

Raina, P. S., Wolfson, C., Kirkland, S. A., Griffith, L. E., Oremus, M., Patterson, C., Tuokko, H., Penning, M., Balion, C. M., Hogan, D., et al. (2009). The canadian longitudinal study on aging (clsa). *Canadian Journal on Aging/La Revue canadienne du vieillissement*, 28(3):221–229.

Resnick, S. I. (2013). *Adventures in stochastic processes*. Springer Science & Business Media.

Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data: With applications in R*. CRC press.

Rizopoulos, D., Verbeke, G., and Molenberghs, G. (2008). Shared parameter models under random effects misspecification. *Biometrika*, 95(1):63–74.

Saad, F., Gleason, D. M., Murray, R., Tchekmedyian, S., Venner, P., Lacombe, L., Chin, J. L., Vinholes, J. J., Goas, J. A., and Zheng, M. (2004). Long-term efficacy of zoledronic acid for the prevention of skeletal complications in patients with metastatic hormone-refractory prostate cancer. *Journal of the National Cancer Institute*, 96(11):879–882.

Shen, P.-s. (2003). The product-limit estimate as an inverse-probability-weighted average. *Communications in Statistics-Theory and Methods*, 32(6):1119–1133.

Struthers, C. A. and Kalbfleisch, J. D. (1986). Misspecified proportional hazard models. *Biometrika*, 73(2):363–369.

Sun, J. (2007). *The statistical analysis of interval-censored failure time data*. springer.

Tsiatis, A. A., Degruttola, 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.

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

Verbeke, G. and Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association*, 91(433):217–221.

Verbeke, G. and Lesaffre, E. (1997). The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computational Statistics & Data Analysis*, 23(4):541–556.

Wang, M.-C., Jewell, N. P., and Tsai, W.-Y. (1986). Asymptotic properties of the product limit estimate under random truncation. *the Annals of Statistics*, pages 1597–1605.

Wei, L.-J., Lin, D. Y., 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.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, pages 1–25.

Woodroofe, M. et al. (1985). Estimating a distribution function with truncated data. *The Annals of Statistics*, 13(1):163–177.

Zeger, S. L. and Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, pages 121–130.

Zeger, S. L. and Liang, K.-Y. (1992). An overview of methods for the analysis of longitudinal data. *Statistics in medicine*, 11(14-15):1825–1839.

Zeger, S. L., Liang, K.-Y., and Albert, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, pages 1049–1060.

Appendices

A An EM Algorithm for Cox Regression

A.1 A Slight Model Re-formulation

We consider here the setting of this chapter with a right-censored failure time and an intermittently observed binary time-dependent covariate. The description here is aiming to outline the modifications to the algorithm in the body of the thesis to support a fully semiparametric Cox regression analysis for the failure time process. We let T_i denote the time of event of interest for individual i which is subject to right censoring at a random censoring time C_i. We then let $V_i = \min(T_i, C_i)$ and $\delta_i = I(V_i = T_i)$. We also define $Y_i(s) = I(s \leq C_i)$, $Y_i^{\dagger}(s) = I(s \leq T_i)$ and let $\bar{Y}_i(s) = Y_i(s)Y_i^{\dagger}(s) = I(s \leq V_i)$ indicate whether individual i is under observation and at risk of failure at time s, as before. Let $N_{i2}(s) = I(T_i \leq s)$ indicate whether individual i has failed by time s, where $dN_{i2}(s) = \lim_{\Delta s \to 0} \Delta N_{i2}(s)$ where $\Delta N_{i2}(s) = \lim_{\Delta s \to 0} N_{i2}(s + \Delta s^-) - N_{i2}(s^-)$, which takes the value 1 if failure occurs at time s. We let $d\bar{N}_{i2}(s) = \bar{Y}_i(s)dN_{i2}(s)$, $\bar{N}_{i2}(s) = \int_0^s d\bar{N}_{i2}(u)$ and define $\Delta\bar{N}_{i2}(s) = \bar{N}_{i2}(s + \Delta s^-) - \bar{N}_{i2}(s^-)$ as terms related to the censored data counting process accordingly.

We consider a discrete internal time-dependent covariate which we refer to as a marker since it is a reflection of the state of a disease in an individual at risk of failure. For a binary marker, let $X_{i1}(s) = 1$ if the marker indicates active disease and zero otherwise. As examples, this may represent a presence of high blood pressure in study of a cardiovascular disease, an acute state of inflammation in an arthritis study, or elevated marker indicating

abnormal kidney function in a study of diabetes nephropathy. When viewed as a stochastic process, we write this as $\{X_{i1}(s), 0 < s\}$. We consider a $p \times 1$ fixed covariate X_{i2} and write $X_i(s) = (X_{i1}(s), X_{i2}')'$. We then consider a regression model for failure based on a multiplicative Cox model with

$$\lim_{\Delta t \downarrow 0} \frac{P(\Delta N_{i2}(t) = 1 \,|\, T_i \geq t, X_i(t))}{\Delta t} = \lambda_2(t \,|\, X_i(t)) = \lambda_2(t) \exp(X_i'(t)\beta) \tag{5.1}$$

where we write $\Lambda_2(t) = \int_0^t \lambda_2(s)ds$ and $d\Lambda_2(t) = \lambda_2(t)\, dt$. Here, unlike the body of Chapter 2, the baseline intensity function is of an unspecified form.

For a sample of size n, the partial likelihood based on the Cox model (5.1) is

$$\mathcal{L} \propto \prod_{i=1}^{n} \left\{ d\Lambda_2(v_i \,|\, X_i(v_i))^{\delta_i} \exp(-\int_0^{\infty} \bar{Y}_i(s) d\Lambda_2(s \,|\, X_i(s))) \right\},$$

and the corresponding log-likelihood is

$$\ell = \sum_{i=1}^{n} \left\{ \int_0^{\infty} \bar{Y}_i(s) dN_{i2}(s) \log d\Lambda_2(s \,|\, X_i(s)) - \int_0^{\infty} \bar{Y}_i(s) d\Lambda_2(s \,|\, X_i(s)) \right\}.$$

The partial likelihood estimating equations for $d\Lambda_2(s)$ is

$$U_2(s) = \sum_{i=1}^{n} U_{i2}(s) = \sum_{i=1}^{n} \bar{Y}_i(s)\{dN_{i2}(s) - d\Lambda_2(s \,|\, X_i(s))\}, \quad 0 < s. \tag{5.2}$$

From (5.2) it is apparent that $d\Lambda_2(s)$ is estimated to be zero except at unique recorded failure times across the sample, denoted $s_1, \ldots, s_L$. We may effectively rewrite (5.2) as

$$U_{2l} = \sum_{i=1}^{n} U_{i2l} = \sum_{i=1}^{n} \bar{Y}_i(s_l)\{dN_{i2}(s_l) - d\Lambda_2(s_l \,|\, X_i(s_l))\}, \quad l = 1, \ldots, L, \tag{5.3}$$

and let $(U_{21}, \ldots, U_{2L})'$ be a vector of estimating functions for $d\Lambda_2(s_l)$, $l = 1, \ldots, L$. The

$(p+1) \times 1$ estimating function for β is analogous to (2.11) but written here as

$$U_{2,L+1} = \sum_{i=1}^{n} U_{i1} = \sum_{i=1}^{n} \int_0^{\infty} \bar{Y}_i(s)\{dN_{i2}(s) - d\Lambda_2(s \mid X_i(s))\}X_i(s) \tag{5.4}$$

and combining these estimating functions for $d\Lambda_2(\cdot)$ and β we write
$U_2 = (U_{21}, \ldots, U_{2L}, U'_{2,L+1})'$.

The Cox model partial likelihood is inherently conditional on the marker process. Models for such internal time-dependent covariates are naturally viewed jointly with the failure process as discussed earlier. To do this, we consider a 3-state model depicted in Figure A.1 with state space $\mathcal{S} = \{0, 1, 2\}$. We let $Z_i(s) = k$ if state k is occupied by individual i at time s, $k = 0, 1, 2$, and let $\{Z_i(s), 0 < s\}$ represent the 3-state stochastic process. Let $\mathcal{H}_i(t) = \{Z_i(s), 0 < s < t, X_{i2}\}$ denote the history of the joint process for individual i, including the fixed covariate vector X_{i2}, $i = 1, \ldots, m$. The function $Y_{ij}(s) = I(Z_i(s^-) = j)$ indicates that individual i is at risk of a transition out of state j at time s, $j = 0, 1$. Note that in the 3-state model, the failure time T_i is the time of entry to the absorbing state 2, and we may write $X_{i1}(s) = I(Z_i(s) = 1)$ to indicate that the time-varying binary marker is elevated at time s, where $X_{i1}(s) = 0$ if the marker value is low and $X_{i1}(s) = 1$ if the marker value is high. The $p \times 1$ covariate vector X_{i2} contains fixed (time independent) covariates.

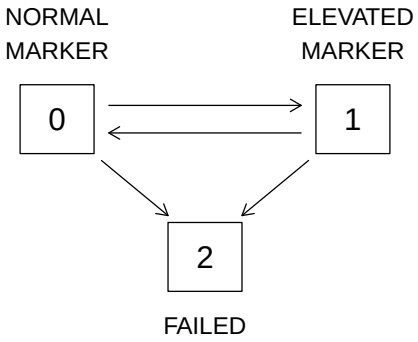


Figure A.1: A multistate diagram depicting a joint model for a dynamic binary covariate and a failure time process

When we consider the joint model of Figure A.1 we need to introduce additional count-

ing processes and intensities to reflect the marker dynamics. Let $N_{ij}(t)$ count the number of $j \to 1-j$ transitions over $(0,t]$ and $dN_{ij}(s) = 1$ if a $j \to 1-j$ transition occurs at time s for the marker with $dN_{ij}(s) = 0$ otherwise; we then write $N_i(t) = (N_{i0}(t), N_{i1}(t), N_{i2}(t))'$ to denote the multivariate counting process for Figure A.1. The history may then be equivalently defined as in terms of the multivariate counting process as $\mathcal{H}_i(t) = \{dN_i(s),\, 0 < s < t, Z_i(0), X_{i2}\}$. The intensities for the marker process dynamics can then be defined as in Section 2.2.1. Note while a semiparametric model can be adopted for the failure time we cannot easily fit a semiparametric model for the marker dynamics since transitions between states 0 and 1 are not observed. We therefore retain the piecewise constant form of them.

Recall we had $U_j(\theta_j) = (\alpha_j, \gamma_j) = (U_{j1}, \ldots, U_{jK_j}, U'_{j,K_j+1})'$ where $U_{jk}, k = 1, \ldots, K_j$ is given in (2.3) and the $p \times 1$ vector U_{j,K_j+1} is given in (2.4), $j = 0, 1$. Combining these with the Cox model estimating function, we then consider $U = (U'_0, U'_1, U'_2)'$ where $U_2 = (U_{21}, \ldots, U_{2L}, U'_{2,L+1})'$. The full parameter vector is then $\theta = (\theta'_0, \theta'_1, d\Lambda_2(\cdot), \beta')'$. We next define the conditional expectation of the complete data score equations, emphasizing the failure time terms since the marker model is unchanged from that of Section 2.2.1 and Section 2.4.2.

A.2 Observed Data Score Equations

Here we consider the intermittent observation process and the resulting observed data $D_i = \{\bar{Y}_i(s), dA_i(s), d\bar{N}_{i2}(s), X_i^\circ(s), 0 < s \le V_i\}$ where $X_i^\circ(s) = (X_{i1}^\circ(s), X'_{i2})'$ and $X_{i1}^\circ(s) = X_{i1}(a_{iA(s^-)})$ is the most recently recorded value. Let θ^r denote the value of the estimate at rth iteration. For an EM algorithm, at the E-step, we take the conditional expectation of the estimating function $U_i(\theta)$ with respect to the missing data (here $\{X_i(s), 0 < s < V_i\}$) given the observed data; here we make the iterative aspect of the EM algorithm explicit by evaluating this conditional expectation at θ^r, the value at the rth iteration. We then

write

$$\bar{U}_i(\theta;\theta^r) = E\{U_i(\theta) \mid D_i;\theta^r\} = \begin{pmatrix} \bar{U}_{i0}(\theta;\theta^r) \\ \bar{U}_{i1}(\theta;\theta^r) \\ \bar{U}_{i2}(\theta;\theta^r) \end{pmatrix}, \tag{5.5}$$

where we use an overbar to denote the conditional expectations of the corresponding estimating equations given the observed data. The expectation for U_{i0} and U_{i1} were given in Section 2.4.2 and only modest adaptations are required here. We therefore focus on the expectations of the Cox estimating functions for $d\Lambda_2(\cdot)$ and β. Recall $U_{i21} = (U_{i21},\ldots,U_{i2L})'$ is the $L \times 1$ contribution to the score function from individual i for $d\Lambda_2(\cdot)$, and U_{i22} is the $(p+1) \times 1$ score function for β.

The conditional expectation of $U_{21}(\theta;\theta^r)$

For the observed data score related to (5.3) we obtain

$$\begin{aligned} \bar{U}_{2l}(\theta;\theta^r) &= \sum_{i=1}^{n} E\{\bar{Y}_i(s_l) \mid dN_{i2}(s_l) - d\Lambda_2(s_l \mid X_i(s_l)) \mid D_i;\theta^r\} \\ &= \sum_{i=1}^{n} \bar{Y}_i(s_l) dN_{i2}(s_l) - d\Lambda_2(s_l)\exp(X_{i2}'\beta_2) E\{\exp(X_{i1}'(s_l)\beta_1) \mid D_i;\theta^r\} \end{aligned} \tag{5.6}$$

The conditional expectation of $U_{22}(\theta;\theta^r)$

From the complete data partial score in (5.4), we obtain the following expected score at θ^r

$$\bar{U}_{22}(\theta;\theta^r) = \sum_{i=1}^{n} \int_0^\infty \bar{Y}_i(s)\{dN_{i2}(s) - d\Lambda_2(s)\exp(\beta_1 + X_{i2}'\beta_2)\}\widetilde{X}_i(s;\,\theta^r) \tag{5.7}$$

for the contribution from individual i, where

$$\widetilde{X}_i(s;\,\theta^r) = \begin{pmatrix} P(X_{i1}(s) = 1 \mid D_i;\theta^r) \\ X_{i2} \end{pmatrix}$$

Adding contributions from each individual in (5.6) and solving for $d\Lambda_2(s)$ gives the profile-type semiparametric interim estimator

$$\begin{aligned} d\tilde{\Lambda}_2^r(s;\beta,\theta^r) &= \frac{\sum_{i=1}^n \bar{Y}_i(s) dN_{i2}(s)}{\sum_{i=1}^n \bar{Y}_i(s) E\{\exp X_i'(s)\beta \,|\, D_i;\theta^r\}} \\ &= \frac{\sum_{i=1}^n \bar{Y}_i(s) dN_{i2}(s)}{\sum_{i=1}^n \bar{Y}_i(s)\exp(X_{i2}'\beta_2) E\{\exp X_{i1}(s)\beta_1 \,|\, D_i;\theta^r\}} \quad , \end{aligned} \tag{5.8}$$

which can be inserted into (5.7) in an analogous way that we did for the piecewise constant model. Doing so gives a $(p+1)\times 1$ pseudo-profile observed data score

$$\bar{\mathcal{U}}_2(\beta;\theta^r) = \sum_{i=1}^n \int_0^\infty \bar{Y}_i(s)\Big\{\tilde{X}_i(s;\,\theta^r) - \frac{\sum_{i=1}^n \bar{Y}_i(s)\tilde{X}_i(s;\,\theta^r)e^{\beta_1 + X_{i2}'\beta_2}}{\sum_{i=1}^n \bar{Y}_i(s) E\{\exp(X_i'(s)\beta) \,|\, D_i;\theta^r\}}\Big\} dN_{i2}(s),$$

where we use $\bar{\mathcal{U}}_2(\beta;\theta^r)$ to distinguish it from the analogous piecewise constant function in the body of the chapter. Because $\bar{N}_{i2}(s)$ only increments at the unique failure times $s_1 < s_2 < \ldots < s_L$, this can be re-expressed as a sum

$$\sum_{i=1}^n \sum_{l=1}^L \bar{Y}_i(s_l)\Big\{\tilde{X}_i(s_l;\,\theta^r) - \frac{\sum_{i=1}^n \bar{Y}_i(s_l)\tilde{X}_i(s_l;\,\theta^r)\exp(\beta_1 + X_{i2}'\beta_2)}{\sum_{i=1}^n \bar{Y}_i(s_l) E\{\exp(X_i'(s_l)\beta) \,|\, D_i;\,\theta^r\}}\Big\} dN_{i2}(s_l) \tag{5.9}$$

over the distinct times $s_1 < s_2 < \cdots < s_L$.

In order to calculate the conditional expectation $E\{X_{i1}(s_l) \,|\, D_i;\theta^r\} = P(X_{i1}(s_l) = 1 \,|\, D_i;\theta^r)$, we consider first the case where s lies between two assessment times. Let $\mathcal{A}_i(s) = (a_{i,r_i(s)-1}, a_{i,r_i(s)})$ be the inter-visit interval for individual i spanning s where $a_{i,r_i(s)-1} < s < a_{i,r_i(s)}$. Then under a Markov model, suppressing the dependence of those expressions on θ^r, we write $E\{X_{i1}(s) \,|\, D_i\} = P(X_{i1}(s) = 1 \,|\, D_i)$ as

$$\begin{aligned} E\{X_{i1}(s) \,|\, D_i\} &= P(X_{i1}(s) = 1 \,|\, Z(a_{i,r_i(s)-1}), Z(a_{i,r_i(s)}),\, X_{i2}) \\ &= \frac{P(Z(a_{i,r_i(s)-1}), Z_i(s) = 1, Z(a_{i,r_i(s)}) \,|\, X_{i2})}{P(Z(a_{i,r_i(s)-1}), Z(a_{i,r_i(s)}) \,|\, X_{i2})} \\ &= \frac{P(Z(a_{i,r_i(s)}) \,|\, Z_i(s) = 1, X_{i2})\, P(Z_i(s) = 1 \,|\, Z(a_{i,r_i(s)-1}), X_{i2})}{P(Z(a_{i,r_i(s)}) \,|\, Z(a_{i,r_i(s)-1}), X_{i2})}. \end{aligned}$$

which is computable from the transition probability matrix as estimated at the rth iteration.

Next we consider the case in which s occurs after the last assessment time. If $A_i(V_i) = r_i$ is the number of post-baseline assessments and $a_{ir_i} < s < V_i$ then when $\delta_i = 0$

$$\begin{aligned} E\{X_{i1}(s) \,|\, D_i\} =& \frac{P(X_{i1}(s) = 1 \,|\, Z(a_{ir_i}), X_{i2})\, P(Z(C_i^-) < 2 \,|\, X_{i1}(s) = 1, X_{i2})}{P(Z(C_i^-) < 2 \,|\, Z(a_{ir_i}), X_{i2})} \\ =& \frac{P(Z_i(s) = 1 \,|\, Z(a_{ir_i}), X_{i2}) P(Z(C_i^-) < 2 \,|\, Z_i(s) = 1, X_{i2})}{P(Z(C_i^-) < 2 \,|\, Z(a_{ir_i}), X_{i2})} \,. \end{aligned}$$

Likewise, if $a_{ir_i} < s < V_i$ and $\delta_i = 1$, then $E\{X_{i1}(s) \,|\, D_i\} = P(X_{i1}(s) = 1 \,|\, Z(a_{ir_i}), T_i)$ is

$$\frac{\sum_{j=0}^{1} P(Z(T_i^-) = j \,|\, Z_i(s) = 1, X_{i2})[d\Lambda_2(T_i \mid X_{i2})]^{1-j}[d\Lambda_2(T_i \mid X_{i2})e^{\beta}]^j \, P(Z_i(s) = 1 \,|\, Z(a_{ir_i}), X_{i2})}{\sum_{j=0}^{1} P(Z(T_i^-) = j \,|\, Z(a_{ir_i}), X_{i2})[d\Lambda_2(T_i \mid X_{i2})]^{1-j}[d\Lambda_2(T_i \mid X_{i2})e^{\beta}]^j} \,.$$

Again this can be estimated at the $(r+1)$st step of the EM algorithm from the transition probability matrix estimate at rth step. Given the above conditional probabilities, we find

$$\begin{aligned} E\{\exp(X_{i1}(s)\beta_1) \,|\, D_i\} &= \exp(\beta_1) P(Z_i(s) = 1 \,|\, D_i) + P(Z_i(s) = 0 \,|\, D_i) \\ &= \exp(\beta) P(Z_i(s) = 1 \,|\, D_i) + [1 - P(Z_i(s) = 1 \,|\, D_i)] \,. \end{aligned}$$

As in equations (2.17) and (2.19) for the marker processes, we note that these expected score equations involve computing $E\{S_{ijk} \,|\, D_i; \theta^r\}$ and $E\{N_{ijk} \,|\, D_i; \theta^r\}$. To evaluate these expressions we consider the following setting. Let $a_{i0} = 0 < a_{i1} < \ldots < a_{ir_i}$ denote the r_i assessment times for individual i, $i = 1, \ldots, m$. Let $\mathcal{A}_{il} = \{a_{i,l-1}, a_{il}\}$ denote interval l between visits, $l = 1, \ldots, r_i$, and $\mathcal{A}_{i,r_i+1} = [a_{ir_i}, V_i]$ denote the interval from the last visit to failure or censoring where $V_i = \min(T_i, C_i)$.

Let $\mathcal{R}_{ijk} = \{l : \mathcal{A}_{il} \cap \mathcal{B}_{jk} \neq \emptyset\}$ be the labels for inter-visit intervals $\mathcal{A}_{il}$, $l = 1, \ldots, r_i + 1$, which span $\mathcal{B}_{jk}$ for individual i and let $\mathcal{C}_{ijkl} = \mathcal{A}_{il} \cap \mathcal{B}_{jk}$ denote the lth sub-interval, $l \in \mathcal{R}_{ijk}$. In Figure A.2, we consider interval $\mathcal{B}_{jk}$ giving labels $\mathcal{R}_{ijk} = \{r+1, r+$

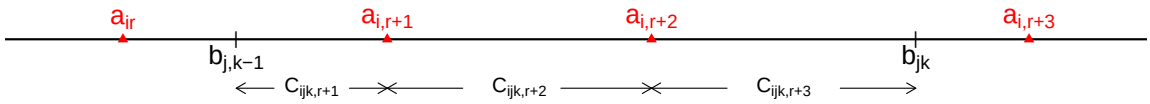


Figure A.2: Timeline depicting the visit process at a given time interval $[b_{j,k-1}, b_{jk}]$

$2, r+3\}$ labeling the inter-visit intervals $\mathcal{A}_{i,r+1}$, $\mathcal{A}_{i,r+2}$, $\mathcal{A}_{i,r+3}$ spanning $\mathcal{B}_{jk}$. We let

$$D_{ijkl} = \{(Z_i(a_{i,l-1}), a_{i,l-1}), (Z_i(a_{il}), a_{il}), X_{i2}\}$$

denote the relevant data for sub-interval $\mathcal{A}_{il}$, $l = r+1, r+2, r+3$.

To compute (2.19) in the setting with the semiparametric Cox model for failure times, we consider each interval $\mathcal{A}_{i,r+1}$, $\mathcal{A}_{i,r+2}$, $\mathcal{A}_{i,r+3}$ separately and then add the results together. Specifically, we let N_{ijkl} denote the $j \to 1-j$ transition count for interval $\mathcal{A}_{il} \cap \mathcal{B}_{jk}$ for individual i and note

$$E\{N_{ijk} \,|\, D_i; \theta^r\} = \sum_{l \in R_{ijk}} E\{N_{ijkl} \,|\, D_{ijkl}; \theta^r\}$$

where the expectation for each sub-interval $E\{N_{ijkl} \,|\, D_{ijkl}\}$ is calculated as

$$\int_0^\infty \bar{Y}_i(s) I(s \in \mathcal{C}_{ijkl}) \frac{P(Z_i(s^-) = j \,|\, Z_i(a_{i,l-1}), X_{i2})\, P(Z_i(a_{il}) \,|\, Z_i(s^-) = 1-j, X_{i2})\, e^{\alpha_{jk} + X_{i2}'\gamma_j}}{P(Z_i(a_{il}) \,|\, Z_i(a_{i,l-1}), X_{i2})}\, ds.$$

Likewise for (2.21) we have

$$E\{S_{ijk} \,|\, D_i; \theta^r\} = \sum_{l \in R_{ijk}} E\{S_{ijkl} \,|\, D_{ijkl}; \theta^r\}$$

where S_{ijkl} is the time spent in state j for intersection of interval $\mathcal{B}_{jk}$ and $\mathcal{A}_{il}$ for individual i and momentarily suppressing the dependence on θ^r we note $E\{S_{ijkl} \,|\, D_{ijkl}\}$ can be

computed as

$$\begin{aligned}
&\int_0^\infty \bar{Y}_i(s) I(s \in \mathcal{C}_{ijkl}) E\{I(Z_i(s^-) = j) \,|\, D_{ijkl}\}\, ds \\
&= \int_0^\infty \bar{Y}_i(s) I(s \in \mathcal{C}_{ijkl}) \frac{P(Z_i(s^-) = j \,|\, Z_i(a_{i,l-1}), X_{i2})\, P(Z_i(a_{il}) \,|\, Z_i(s^-) = j, X_{i2})}{P(Z_i(a_{il}) \,|\, Z_i(a_{i,l-1}), X_{i2})}\, ds.
\end{aligned}$$

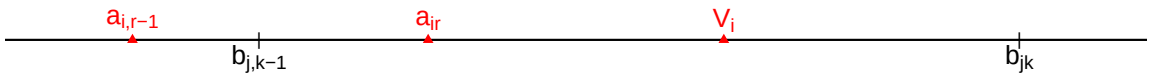


Figure A.3: An illustrative example of the visit process where an event is observed in a given time interval

We next consider the setting where an individual's follow-up ends at time V_i within the interval $\mathcal{B}_{jk}$ as depicted in Figure A.3. Here, we have $R_{ijk} = \{r_i, r_i+1\}$, where we define the inter-visit intervals to be $\mathcal{A}_{ir_i} = [a_{i,r_i-1}, a_{ir_i}]$ and $\mathcal{A}_{i,r_i+1} = [a_{ir_i}, V_i]$, where $V_i = \min(T_i, C_i)$ and we let $\delta_i = I(V_i = T_i)$. Note that the expectation for the first sub-interval is calculated as before.

For the second sub-interval, we first consider the case of an individual censored in interval $\mathcal{B}_{jk}$. In this case, we have $\delta_i = 0$ (i.e. $V_i = C_i$) and the second sub-interval is $\mathcal{A}_{i,r_i+1} = [a_{ir_i}, V_i]$. Here, the data are $D_{ijk,r_i+1} = \{Z_i(a_{ir_i}), a_{ir_i}, V_i, \delta_i = 0, X_{i2}\}$. We then calculate

$$\begin{aligned}
&E\{N_{ijk,r_i+1} \,|\, D_{ijk,r_i+1}\} \\
&= \int_0^\infty \bar{Y}_i(s) I(s \in \mathcal{C}_{ijk,r_i+1}) E\{I(Z_i(s^-) = j) dN_{ij}(s) \,|\, D_{ijk,r_i+1}\}\, ds
\end{aligned}$$

where $E\{I(Z_i(s^-) = j) dN_{ij}(s) \,|\, D_{ijk,r_i+1}\}$ is given by

$$\frac{\sum_{v=0}^{1} P(Z_i(s^-) = j \,|\, Z_i(a_{ir_i}), X_{i2})\, P(Z_i(C_i^-) = v \,|\, Z_i(s^-) = 1-j, X_{i2})\, e^{\alpha_{jk} + X_{i2}'\gamma_j}}{\sum_{v=0}^{1} P(Z_i(C_i^-) = v \,|\, Z_i(a_{ir_i}), X_{i2})}$$

. Likewise, the expectation for the time at risk $E\{S_{ijk,r_i+1} \mid D_{ijk,r_i+1}\}$ is computed as

$$\int_0^\infty \bar{Y}_i(s) I(s \in \mathcal{C}_{ijk,r_i+1}) \frac{\sum_{v=0}^{1} P(Z_i(s^-) = j \mid Z_i(a_{ir_i}), X_{i2}) P(Z_i(C_i^-) = v \mid Z_i(s^-) = j, X_{i2})}{\sum_{v=0}^{1} P(Z_i(C_i^-) = v \mid Z_i(a_{i,r_i}), X_{i2})} \, ds$$

Then we consider the case of an individual who has the event of interest at the end of the second sub-interval. Here, $\delta_i = 1$ (i.e. $V_i = T_i$) and the observed data are $D_{ijk,r_i+1} = \{Z_i(a_{ir_i}), a_{ir_i}, V_i, \delta_i = 1, X_{i2}\}$. We can then calculate (2.19) as

$$E\{N_{ijk,r_i+1} \mid D_{ijk,r_i+1}\} = \int_0^\infty \bar{Y}_i(s) I(s \in \mathcal{C}_{ijk,r_i+1}) E\{I(Z_i(s^-) = j) dN_{ij}(s) \mid D_{ijk,r_i+1}\} \, ds$$

where $E\{I(Z_i(s^-) = j) dN_{ij}(s) \mid D_{ijk,r_i+1}\}$ is

$$\frac{\sum_{v=0}^{1} P(Z_i(s^-) = j \mid Z_i(a_{ir_i}), X_{i2}) \, e^{\alpha_{jk} + X_{i2}'\gamma_j} \, P(Z_i(T_i^-) = v \mid Z_i(s^-) = 1 - j, X_{i2}) dH_0(T_i) e^{v\beta_1 + X_{i2}'\beta_2}}{\sum_{v=0}^{1} dH_0(T_i) e^{v\beta_1 + X_{i2}'\beta_2} P(Z_i(T_i^-) = v \mid Z_i(a_{i,r_i}), X_{i2})}$$

The expectation for the time at risk $E\{S_{ijk,r_i+1} \mid D_{ijk,r_i+1}\}$ is likewise computed as

$$E\{S_{ijk,r_i+1} \mid D_{ijk,r_i+1}\} = \int_0^\infty \bar{Y}_i(s) I(s \in \mathcal{C}_{ijk,r_i+1}) E\{I(Z_i(s^-) = j) dS_{ij}(s) \mid D_{ijk,r_i+1}\} \, ds$$

where $E\{I(Z_i(s^-) = j) dS_{ij}(s) \mid D_{ijk,r_i+1}\}$ is

$$\frac{\sum_{v=0}^{1} P(Z_i(s^-) = j \mid Z_i(a_{ir_i}), X_{i2}) P(Z_i(T_i^-) = v \mid Z_i(s^-) = j, X_{i2}) dH_0(T_i) e^{v\beta_1 + X_{i2}'\beta_2} \, ds}{\sum_{v=0}^{1} dH_0(T_i) e^{v\beta_1 + X_{i2}'\beta_2} P(Z_i(T_i^-) = v \mid Z_i(a_{i,r_i}), X_{i2})} .$$

A.3 The Maximization Step

We are next required to solve

$$\bar{U}(\theta; \theta^r) = 0 = \sum_{i=1}^{n} \bar{U}_i(\theta; \theta^r)$$

where $\bar{U}_i(\theta; \theta^r)$ is given by (5.5) to obtain an updated estimate θ^{r+1}.

The solutions to the expected estimating functions for the marker process transitions are obtained as in the body of the chapter. For the failure process we use the approach

adopted for semiparametric maximum likelihood estimation by noting that setting (5.9) to zero and solving gives $\widetilde{\beta}^{r+1}$. This can be substituted into (5.8) to obtain $d\widetilde{\Lambda}_2^{r+1}(s;\widetilde{\beta}^{r+1};\theta^r)$. Combining the updated estimates θ_0^{r+1} and θ_1^{r+1} with β_0^{r+1} and $d\widetilde{\Lambda}^{r+1}(\cdot)$ we use θ^{r+1} to recompute the expectations and repeat this procedure iteratively until the difference in successive estimates becomes negligible or the L_2 norm of $\bar{U}(\theta;\theta^r)$ becomes lower than a specified tolerance.

B Estimating the Transition Probability Matrix

In order to calculate the transition probability matrix we must deal with the fact that the intensities between the markers states are estimated based on piecewise constant models while the failure intensities are governed by a semi-parametric model. To illustrate how the required expectations can be computed in this context consider a setting where the two marker-to-marker transition intensities have the same location for the break-points so $K_0 = K_1 = K$ and $\lambda_0(s)$ and $\lambda_1(s)$ are constant over common intervals $[b_{k-1}, b_k), k = 1,\ldots,K$. We let R_k denote the number of failures observed to fall in $\mathcal{B}_k$ and let $\delta_k = \{s_{k_r}, r = 1,\ldots,R_k\}$ denote the set of failure times over $\mathcal{B}_k, k = 1,\ldots,K$. Since $d\widehat{\Lambda}_2(s) = 0$ unless $s = t_{k_r}$ for some r, within $\mathcal{B}_k$ the transition intensity matrix for the 3-state process has zero-valued intensities for failure between failure times. Let $\mathcal{S}_{k0} = (b_{k-1}, s_{k_1})$, $\mathcal{S}_{kr} = (s_{k_{r-1}}, s_{k_r}), r = 1,\ldots,R_k$ and $\mathcal{S}_{kR_k+1} = (s_{kR_k}, b_k)$ partition $\mathcal{B}_k$. Within intervals $\int_{kr}, r = 0,\ldots,R_k+1$, no failure occurs so the intensity matrix has estimate of the form

$$d\widehat{Q}_1(t) = \begin{pmatrix} -\widehat{\lambda}_{0k}\,dt & \widehat{\lambda}_{0k}\,dt & 0 \\ \widehat{\lambda}_{1k}\,dt & -\widehat{\lambda}_{1k}\,dt & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{for } t \in \mathcal{S}_{kr}\,,$$

$r = 0, 1, \ldots, R_k + 1$. As the failure time is only observed at an instant in time, the intensity matrix $d\widehat{Q}_2(s_{k_r} \mid X_{i2})$ at each failure time within $\mathcal{B}_k$ has the form

$$\begin{pmatrix} -[\widehat{\lambda}_{0k}\, dt + d\widehat{\Lambda}_2(s_{k_r}) e^{X_2'\widehat{\beta}_2}] & \widehat{\lambda}_{0k}\, dt & d\widehat{\Lambda}_2(s_{k_r}) e^{X_2'\widehat{\beta}_2} \\ \widehat{\lambda}_{1k}\, dt & -[\widehat{\lambda}_{1k}\, dt + d\widehat{\Lambda}_2(s_{k_r}) e^{\widehat{\beta}_1 + X_2'\widehat{\beta}_2}] & d\widehat{\Lambda}_2(s_{k_r}) e^{\widehat{\beta}_1 + X_2'\widehat{\beta}_2} \\ 0 & 0 & 0 \end{pmatrix}$$

for $s_{k_r} \in \mathcal{S}_k$ that occurred in $\mathcal{B}_k$ and dt is very small (near zero) value, such as 10^{-6}. If we let $s_{k_0} = b_{k-1}$, then the probability matrix over each piece is estimated by

$$\begin{aligned} \widehat{P}(b_{k-1}, b_k) &= \prod_{(b_{k-1}, b_k]} [\, I + d\widehat{Q}(u \mid X_{i2})\,] \\ &= \prod_{r=1}^{R_k} \Big\{ \prod_{(s_{k_{r-1}}, s_{k_r}]} [\, I + d\widehat{Q}_1(u)]\, [I + d\widehat{Q}_2(s_{k_r} \mid X_{i2})] \Big\} \prod_{(s_{R_k}, b_k]} [I + d\widehat{Q}_1(u)]\} \end{aligned}$$

over the interval $\mathcal{B}_k$ where I is a 3×3 identity matrix. For computations for $b_{k-1} \le s < t < b_k$ the limit of the product integration can be modified accordingly.

C Computation of Limiting Value Based on a Monte Carlo Algorithm

Under a marker-independent visit process the calculation of conditional expectations in (3.7a) and (3.7b) requires integration of distribution of the visit as well as the distribution of its maximum order statistics $f(a_k = t \,|\, A(s^-) = k)$. The calculations are more challenging with a marker-dependent visit process. We therefore introduce the use of Monte Carlo methods for the required integration. As in Section 3.1 of Chapter 3, we let $X_{i1}(s) = I(Z_i(s^-) = 1)$ and let $X_{i1}^\circ(s) = X_{i1}(a_{iA_i(s^-)})$ denote the most recently recorded binary marker value. Let X_{i2} denote a $p \times 1$ fixed covariate vector and we write $X_i^\circ(s) = (X_{i1}^\circ(s), X_{i2}')'$. We let nsim denote the total number of individuals for whom the data are to be simulated. For a very large dataset (i.e. if nsim $\approx$ 5000), the conditional expectations can be approximated as follows.

We let $\hat{r}^{(0)}(s;\psi)$ be

$$\begin{aligned}
&\widehat{\mathbb{E}}\{\bar{Y}(s)\exp(\psi' X^{\circ}(s))\} \\
=&\frac{1}{\text{nsim}}\sum_{i=1}^{\text{nsim}}\sum_{k=0}^{\infty}\Big\{I(s<C_i)\,I(s<T_i)\,I(A_i(s)=k)\sum_{x=0}^{1}I(Z_i(a_{ik})=x)\exp(x_{i1}\psi_1+X_{i2}'\psi_2)\Big\}
\end{aligned}$$

where $\psi=(\psi_1,\psi_2')'$, and $\sum_{x=0}^{1}I(Z_i(a_{ik})=x)\exp(x_{i1}\psi_1+X_{i2}'\psi_2)$ is given by

$$I(Z_i(a_{ik})=0)\exp(X_{i2}'\psi_2)+I(Z_i(a_{ik})=1)\exp(\psi_1+X_{i2}'\psi_2)$$

. Here, $\bar{Y}(s)\exp(\psi' X^{\circ}(s))$ only takes a non-zero value when $\bar{Y}(s)=1$. In other words, in a simulation study, we only need to select individuals occupying either of the middle two columns of states in Figure 3.2 for a given time s. Likewise for the other conditional expectations, we have the Monte Carlo approximation $\hat{r}^{(1)}(s;\psi)$ given by

$$\begin{aligned}
&\widehat{\mathbb{E}}\{\bar{Y}(s)X^{\circ}(s)\exp(\psi' X^{\circ}(s))\} \\
=&\frac{1}{\text{nsim}}\sum_{i=1}^{\text{nsim}}\Big\{\sum_{k=0}^{\infty}I(\mathcal{Z}_i(s)\in\{(0,k),(1,k)\})\sum_{x_{i1}=0}^{1}I(Z_i(a_{ik})=x_{i1})x_i^{\circ}(s)\exp(x_{i1}\psi_1+X_{i2}'\psi_2)\Big\} \\
=&\frac{1}{\text{nsim}}\sum_{i=1}^{\text{nsim}}\sum_{k=0}^{\infty}\Big\{I(s<C_i)\,I(s<T_i)\,I(A_i(s)=k) \\
&\qquad(I(Z_i(a_{ik})=0)x_i^{\circ}(s)\exp(X_{i2}'\psi_2)+I(Z_i(a_{ik})=1)x_i^{\circ}(s)\exp(\psi_1+X_{i2}'\psi_2)\Big\}.
\end{aligned}$$

For $r^{(0)}(s)$ we use the Monte Carlo approximation $\hat{r}^{(0)}(s)$ given by

$$\begin{aligned}
&\widehat{\mathbb{E}}\{\bar{Y}(s)dN(s)\} \\
=&\frac{1}{\text{nsim}}\sum_{i=1}^{\text{nsim}}\Big\{\sum_{k=0}^{\infty}I(\mathcal{Z}_i(s)\in\{(0,k),(1,k)\})\sum_{x_{i1}=0}^{1}I(Z_i(a_{ik})=x_{i1})d\Gamma_0(s)\exp(x_{i1}\psi_1+X_{i2}'\psi_2)\Big\} \\
=&\frac{1}{\text{nsim}}\sum_{i=1}^{\text{nsim}}\sum_{k=0}^{\infty}d\Gamma_0(s)\Big\{I(s<C_i)I(s<T_i)\,I(A_i(s)=k) \\
&\qquad I(Z_i(a_{ik})=0)\exp(X_{i2}'\psi_2)+I(Z_i(a_{ik})=1)\exp(x_{i1}\psi_1+X_{i2}'\psi_2)\Big]\Big\},
\end{aligned}$$

and finally $r^{(1)}(s)$ is

$$\begin{aligned}
&\widehat{\mathbb{E}}\{\bar{Y}(s)dN(s)X^{\circ}(s))\}\\
=&\frac{1}{\text{nsim}}\sum_{i=1}^{\text{nsim}}\Big\{\sum_{k=0}^{\infty}I(\mathcal{Z}_i(s)\in\{(0,k),(1,k)\})\sum_{x_{i1}=0}^{1}I(Z_i(a_{ik})=x)\,x\,d\Gamma_0(s)\exp(x\psi)\Big\}\\
=&\frac{1}{\text{nsim}}\sum_{i=1}^{\text{nsim}}\sum_{k=0}^{\infty}d\Gamma_0(s)\Big\{\sum_{k=0}^{\infty}I(s<C_i)I(s<T_i)I(A_i(s)=k)\,I(Z_i(a_{ik})=1)\exp(\psi)\Big\}.
\end{aligned}$$

D An Expanded Joint Model For a Marker, Misspecified Marker, Failure, Visit and Censoring Process

D.1 Notation and Intensities

Here we consider a joint model for a four state process involving a binary marker, failure and censoring process, and the process by which visits are made for the measurement of markers, and the misspecified marker value resulting from carrying forward a recorded marker to the next time a visit is made, or failure or right-censoring occurs. This joint model differs from the multistate model of Figure 3.2 in that it incorporates through the path of the marker state that was occupied at the time each visit occurs; in Figure 3.2 a visit-marker state could be entered upon a visit or following a marker-transition between visits so the marker state occupied at the time of a visit is not captured. This new state space enables joint modeling of $X_{i1}^{\circ}(t)$ and $X_{i1}(t)$ since $X_{i1}^{\circ}(t)$ is the marker value at the time of the most recent visit. We let $Z(t)$ denote the state of the four state process which is occupied with state 0 representing being event-free and uncensored with a normal marker value, 1 representing being event-free and uncensored with an elevated marker value, 2 representing the state of having failed, and 3 representing the state of being lost to follow-up and right-censored.

We let $A(t)$ be the counting process recording the cumulative number of post-baseline visits, and we let $X_1^{\circ}(t)$ represent the misspecified marker value resulting from carrying

forward the most recently recorded marker value. We then consider an expanded state space and let $\mathcal{Z}(t) = (A^*(t), Z(t), X_1^\circ(t))$ for the a joint model where $A^*(t)$ is used to denote the counting process for the visits, with the extension of additional states following the rth visit labeled r^p to denote the setting where some elements of $\mathcal{Z}(t)$ have changed between visits - this permits defining a state space which facilitates distinguishing between the true marker value as determined by $Z(t)$ and the mismeasured value reflected by $X_1^\circ(t)$. We let

$$\bar{\mathcal{H}}(t) = \{(A^*(s), Z(s), X_1^\circ(s)), 0 < s < t\} = \{\mathcal{Z}(s), 0 < s < t\}$$

denote the history of the expanded joint process. In what follows for simplicity we assume individuals begin in state (0,0,0) with probability 1.

Figure D.1 gives the state space with arrows reflecting the possible transitions. There are still only four types of intensities in the figure, some of which are labelled by m for marker-related transitions in the $Z(t)$ sub-process, d for death, c for censoring, and a for assessments (visits). These are defined as follows.

Marker Transition Intensities

For the marker related intensities we set

$$\lambda^m(t \,|\, \bar{\mathcal{H}}_i(t)) = \lambda_j(t \,|\, X_{i2}) \,, \qquad j = 0, 1 \,. \tag{5.10}$$

Death Transition Intensities

Here we let

$$\lambda^d(t \,|\, \bar{\mathcal{H}}_i(t)) = \lambda_{i2}(t) = \lambda_2(t \,|\, X_{i1}(t), X_{i2}) \tag{5.11}$$

where the intensity depends on the true marker value as well as fixed covariates.

Censoring Intensities

Here we let

$$\lambda^c(t \,|\, \bar{\mathcal{H}}_i(t)) \tag{5.12}$$

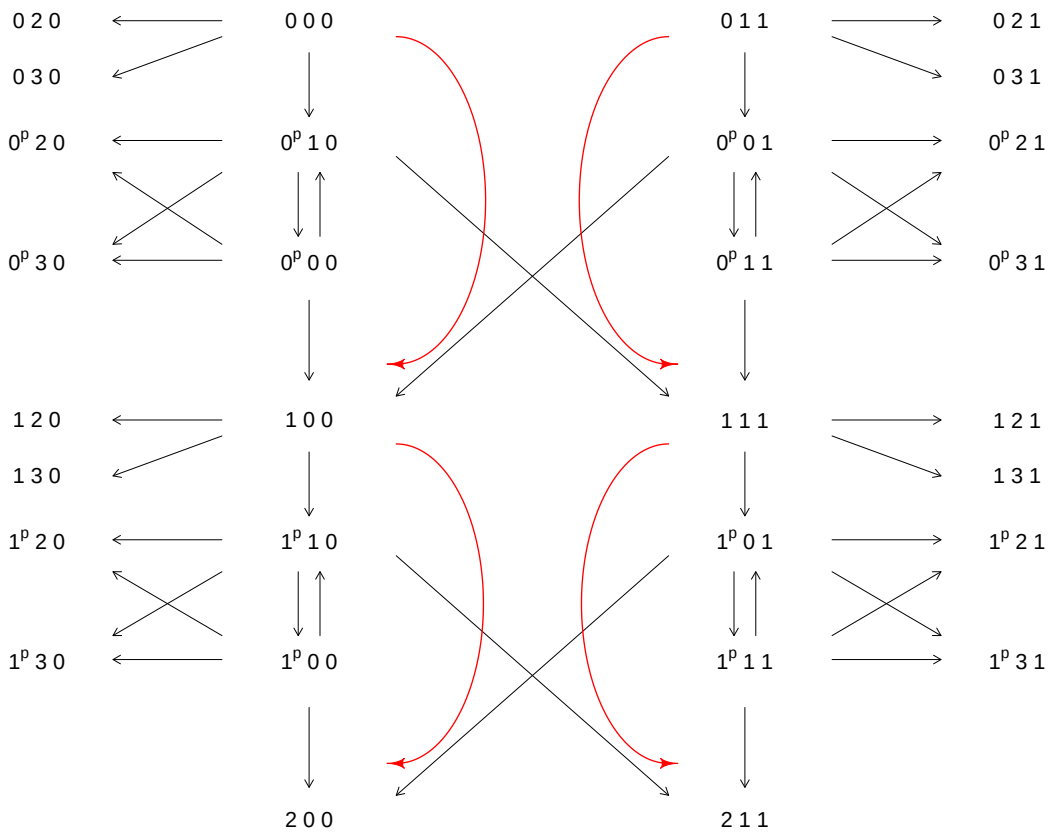


Figure D.1: A multistate model for $(A(t), Z(t), X_1^\circ(t))$ depicting a more general joint model for the multistate, visit and censoring process which accommodates a marker dependent visit process and records the misclassification status of the binary time-dependent covariate.

where the nature of the dependence on this history determines whether censoring is dependent or independent.

Visit Intensities

Here we let

$$\lambda^a(t \mid \bar{\mathcal{H}}_i(t)) \tag{5.13}$$

where the nature of the dependence on this history determines whether visit process is dependent on the marker or not.

We consider Markov models and set

$$\lambda_i^a(t|\bar{\mathcal{H}}(t)) = \rho_a \exp(X_{i1}(t^-)\eta_a)$$

to accommodate a marker-driven visit process, and

$$\lambda_i^c(t|\bar{\mathcal{H}}(t)) = \rho_c \exp(X_{i1}(t^-)\eta_c)$$

to accommodate marker-dependent censoring.

D.2 Derivations of $r^{(k)}(s)$ and $r^{(k)}(s;\psi)$ for Model in Figure D.1

Here we consider the evaluation of the key function in (3.6) to describe an alternative approach to estimating the limiting values of the estimators under misspecified Cox regression model based on the expanded multistate process of Figure D.1. We consider each function in (3.7a) and (3.7b) in turn.

I. Evaluation of $r^{(0)}(s) = E(\bar{Y}_i(s)dN_i(s))$

From (3.7a) with $k = 0$ we obtain

$$
\begin{aligned}
r^{(0)}(s) &= E(\bar{Y}_i(s)dN_i(s)) \\
&= E_{\bar{Y}(s)}\left\{ E_{X(s)}\left[E(d\bar{N}(s) \,|\, \bar{Y}(s), X(s))\right]\right\} \\
&= E_{\bar{Y}(s)}\left\{ E_{X(s)}\left[\bar{Y}(s)d\Lambda_2(s)e^{X'(s)\beta}\right]\right\} \\
&= d\Lambda_2(s)E_{X_2}\left\{ e^{X_2'\beta_2} E_{\bar{Y}(s)\,|\,X_2}\bar{Y}(s)\left[E_{X_1(s)|\bar{Y}(s),X_2}(e^{X_1(s)\beta_1})\right]\right\} \\
&= d\Lambda_2(s)E_{X_2}\left\{ e^{X_2'\beta_2} E_{\bar{Y}(s),X_1(s)\,|\,X_2}\left(\bar{Y}(s)e^{X_1(s)\beta_1}\right)\right\} \\
&= d\Lambda_2(s)E_{X_2}\left\{ e^{X_2'\beta_2} \sum_{a=0}^{\infty}\sum_{z=0}^{1}\sum_{x^\circ=0}^{1} e^{z\beta_1} P(\mathcal{Z}(s) = (a, z, x^\circ) \,|\, \mathcal{Z}(0) = (0, 0, 0), X_2)\right\}.
\end{aligned}
$$

II. Evaluation of $r^{(1)}(s) = E\{\bar{Y}_i(s)X_i^\circ(s)dN_i(s)\}$

Likewise if $k = 1$ we obtain from (3.7a) the following

$$
\begin{aligned}
r^{(1)}(s) &= E\{\bar{Y}_i(s)X_i^\circ(s)dN_i(s)\} \qquad (5.14)\\
&= E\left(E_{X^\circ(s),\bar{Y}(s)}\left[E\left\{\bar{Y}(s)X^\circ(s)d\Lambda_2(s)e^{X_1(s)\beta_1 + X_2'\beta_2} \mid X_1^\circ(s) = 1, X_2, \bar{Y}(s) = 1\right\} \mid X_2\right]\right)
\end{aligned}
$$

Consider the first element of $r^{(1)}(s)$, denoted $r_1^{(1)}(s)$. This is of the form

$$
\begin{aligned}
&E_{X_2}\left(E_{X^\circ(s),\bar{Y}(s)}\left[E\{\bar{Y}(s)X_1^\circ(s)d\Lambda_2(s \mid X(s))|X_1^\circ(s) = 1, X_2, \bar{Y}(s) = 1\} \mid X_2\right]\right) \\
&= E_{X_2}\Bigg(E_{X^\circ(s),\bar{Y}(s)}\Bigg[\bar{Y}(s)X_1^\circ(s)\sum_{k=0}^{1} d\Lambda_2(s)e^{k\beta_1 + X_2'\beta_2} \\
&\quad \times P(\mathcal{Z}(s) \in \{(a, k, 1), a = 0, 1, \ldots\}|\mathcal{Z}(s) \in \{(a, k, 1), a = 0, 1, \ldots, k = 0, 1\}, X_2)\Bigg] \mid X_2\Bigg) \\
&= E_{X_2}\left(d\Lambda_2(s)e^{X_2'\beta_2}\sum_{k=0}^{1} e^{k\beta_1} P(\mathcal{Z}(s) \in \{(a, k, 1), a = 0, 1, \ldots\}|X_2)\right)
\end{aligned}
$$

Note $\bar{Y}(s)X^\circ(s) = 1$ iff a subject is in a state in the set $\mathcal{S} = \{(*, k, 1),\ k = 0, 1\}$.

Let $\mathcal{Z}$ denote the full multistate process. Then the first element of (5.14) is equal to

$$d\Lambda_2(s)\, E_{X_2}\left(e^{X_2'\beta_2}\sum_{k=0}^{1} e^{k\beta_1} P(\mathcal{Z}(s)\in\{(a,k,1),\ a=0,1,\ldots\}\mid X_2)\right)$$

For the remaining p elements of $r^{(1)}(s)$, we have $r_2^{(1)}(s)$ in the form

$$E_{X_2}\left(d\Lambda_2(s)e^{X_2'\beta_2}X_2\sum_{x^\circ=0}^{1}\sum_{k=0}^{1} e^{k\beta_1}P(\mathcal{Z}(s)\in\{(a,k,x^\circ),a=0,1,\ldots,x^\circ=0,1\}|X_2)\right)$$

III. Evaluation of $r^{(0)}(s;\psi) = E\{\bar{Y}_i(s)\exp(X_1^\circ(s)\psi_1 + X_2'\psi_2)\}$

Now for (3.7b) with $k=0$ we obtain

$$\begin{aligned} r^{(0)}(s;\psi) &= E\left\{\bar{Y}_i(s)e^{X_1^\circ(s)\psi_1+X_2'\psi_2}\right\} \\ &= E_{X_2}\left(E_{\bar{Y}(s),X_1^\circ(s)|X_2}\left\{\bar{Y}(s)e^{X_1^\circ(s)\psi_1+X_2'\psi_2}\right\}\right) \\ &= E_{X_2}\left(e^{X_2'\psi_2}\sum_{a=0}^{\infty}\sum_{z=0}^{1}\sum_{x_1^\circ=0}^{1} e^{x_1^\circ\psi_1}P(\mathcal{Z}(s)=(a,z,x_1^\circ)|\mathcal{Z}(0)=(0,0,0),X_2)\right)\end{aligned}$$

IV. Evaluation of $r^{(1)}(s;\psi) = E\{\bar{Y}(s)X^\circ(s)\exp(X_1^\circ\psi_1 + X_2'\psi_2)\}$

Finally if $k=1$, (3.7b) gives

$$r^{(1)}(s;\psi) = E\left\{\bar{Y}(s)X^\circ(s)e^{X_1^\circ\psi_1+X_2'\psi_2}\right\},$$

where the first element is

$$
\begin{aligned}
r_1^{(1)}(s;\psi) &= E\left\{\bar{Y}(s)X_1^{\circ}(s)e^{X_1^{\circ}(s)\psi_1+X_2'\psi_2}\right\} \\
&= E_{X_2}\left\{e^{X_2'\psi_2}E_{\bar{Y}(s),X_1^{\circ}(s)|X_2}\left[\bar{Y}(s)X_1^{\circ}(s)e^{X_1^{\circ}(s)\psi_1}\right]\right\} \\
&= E_{X_2}\left\{e^{X_2'\psi_2}\sum_{a=0}^{\infty}\sum_{z=0}^{1}e^{\psi_1}P(\mathcal{Z}(s)=(a,z,1)|\mathcal{Z}(0)=(0,0,0),X_2)\right\}
\end{aligned}
$$

and the remaining elements are

$$
\begin{aligned}
r_2^{(1)} &= E\left\{\bar{Y}(s)X_2e^{X_1^{\circ}(s)\psi_1+X_2'\psi_2}\right\} \\
&= E_{X_2}\left\{X_2e^{X_2'\psi 2}E\left(\bar{Y}(s)e^{X_1^{\circ}(s)\psi_1}\mid X_2\right)\right\} \\
&= E_{X_2}\left\{X_2e^{X_2'\psi_2}\sum_{a=0}^{\infty}\sum_{z=0}^{1}\sum_{x_1^{\circ}=0}^{1}e^{x_1^{\circ}\psi_1}P(\mathcal{Z}(s)=(a,z,x_1^{\circ})\mid\mathcal{Z}(0)=(0,0,0),X_2)\right\}.
\end{aligned}
$$