Mitigating bias from intermittent measurement of time-dependent covariates in failure time analysis

SHU JIANG

Division of Public Health Sciences,

Department of Surgery,

Washington University School of Medicine in St. Louis, St. Louis, MO, 63110, U.S.A.

RICHARD J. COOK

Department of Statistics and Actuarial Science,

University of Waterloo, Waterloo, ON, N2L 3G1, Canada
E-mail: rjcook@uwaterloo.ca

LEILEI ZENG

Department of Statistics and Actuarial Science,

University of Waterloo, Waterloo, ON, N2L 3G1, Canada

Summary

Cox regression models are routinely fitted to examine the association between time-dependent markers and a failure time when analyzing data from clinical registries. Typically, the marker values are measured periodically at clinic visits with the recorded value carried forward until the next assessment. We examine the asymptotic behavior of estimators from Cox regression models under this observation and data handling scheme when the true relationship is based on a Cox model using the current value of the marker. Specifically, we explore the impact of the marker process dynamics, the clinic visit intensity, and the marginal failure rate on the limiting value of the estimator of the marker effect from the Cox model. We also illustrate how a joint multistate model that accommodates intermittent observation of the time-varyingmarker can be formulated. Simulation studies demonstrate that the finite sample performance of the naive estimator aligns with the asymptotic results and shows good performance of the estimators from the joint model. We apply both methods to data from a study of bone markers and their effect on the development of skeletal complications in metastatic cancer.

Keywords: Cox model, intermittent observation, model misspecification, time-dependent covariates

1 INTRODUCTION

Interest often lies in examining the relationship between a time-varying biomarker and the occurrence of a clinically important event to enhance scientific understanding about a disease process. The Cox regression model (Cox, 1972) is well-suited to address this and is routinely used, but like other regression methods accommodating time-varying covariates, it requires marker values to be known at times other than when they were measured; for the Cox model, we need the values at each of the failure times. While markers change value in continuous time, they are only measured intermittently at, say clinic visits, times that blood or urine samples are drawn or through other encounters in the health care system. As a result, the "current" values are typically unknown and an *ad hoc* approach is to carry forward the most recently recorded value.

There is a large literature on methods for dealing with covariate measurement error in cross-sectional, longitudinal, and life history settings (Carroll et al., 2006; Fuller, 2009; Buonaccorsi, 2010; Yi, 2016). Many of these methods require either a validation substudy or a replication study to acquire auxiliary information on the measurement error process.

In this failure time setting with intermittent observation of marker processes, the deviation between the true effect and the estimand arising from a naive analysis using the carry-forward approach is driven by the joint marker, failure and censoring processes, rather than an external measurement process. Specific governing factors include the temporal variation (i.e. volatility) of the marker process, the intensity of the visit process, the nature and strength of the association between the marker and failure processes, and the failure and censoring processes. To address this Bruijne et al. (2001) proposed a weighted analysis in which time-varying weights were defined as a function of time since the most recent clinic visit, say, at which time the marker was measured; this is implemented in such a way that more out-of-date marker values have less impact in the estimating equations than recent marker values obtained from more recent assessments. Smoothing the observed covariate values to impute the missing current marker values has also been explored in several articles (Raboud et al., 1993; Tsiatis et al., 1995; Andersen and Liestol, 2003). Joint modeling is another approach where it involves modeling the covariate process, but is usually accompanied by a stronger assumption about the distribution of the covariate process (Degruttola and Tu, 1994; Xu and Zeger, 2001).

In this article, we focus on a discrete biomarker process and we shall discuss both the asymptotic bias by using the carried-forward marker value and a joint model that accommodates the unobserved up-to-date marker value. We assume a two-state alternating process for the marker and adopt a semi-parametric Cox model for the failure time analysis. The visit process is assumed to be completely independent of the marker-failure time process. We use this structure to provide a detailed large sample and empirical study of the limiting value of the estimator from a Cox model using the standard approach of carrying-forward the most recently recorded marker. A joint analysis is then proposed, and the finite sample behaviour of the corresponding estimator is also examined.

The remainder of this article is organized as follows. In Section 2, we define notation and formulate a joint model for the marker and failure time process based on a three-state process. We define the Cox model and review the large sample theory associated with misspecified Cox regression models. In Section 3, we give a detailed exploration of the factors that influence the asymptotic bias of estimators from misspecified Cox models when the most recently measured marker value is carried forward. In Section 4, we give the likelihood for fitting a joint model for marker and failure process and examine the performance of the misspecified Cox model and the proposed model in simulation studies. An application involving data from a study of the relationship between bone markers and the occurrence of skeletal complications in metastatic cancer is given in Section 5 and concluding remarks are provided in Section 6.

2 NOTATION AND LIKELIHOOD

2.1 NOTATION FOR THE MARKER, FAILURE TIME AND VISIT PROCESSES

Let T_i denote a time to failure for an event of interest and $N_i(s) = I(T_i \leq s)$ indicate that failure has occurred by time s for individual i in a sample of m independent individuals $i = 1, \ldots, m$. We let $\Delta N_i(t) = N_i(t + \Delta t^-) - N_i(t^-)$ so that $dN_i(t) = \lim_{\Delta t \downarrow 0} \Delta N_i(t) = 1$ if failure occurs at time s for individual i and is zero otherwise. We let C_i^A represent an administrative censoring time and C_i^R a random censoring time so that individual i is observed over $(0, C_i]$ where $C_i = \min(C_i^A, C_i^R)$; we let $V_i = \min(T_i, C_i)$ and $\delta_i = I(V_i = T_i)$. The function $Y_i(s) = I(s \leq C_i)$ indicates that individual i is uncensored at time s, $i = 1, \ldots, m$. If $Y_i^\dagger(s) = I(s \leq T_i)$ indicates that individual i has not failed prior to time s, then $\bar{Y}_i(s) = Y_i(s)Y_i^\dagger(s)$ indicates that the individual is at risk and under observation at time s and we let $d\bar{N}_i(s) = \bar{Y}_i(s)dN_i(s)$.

Let $\{X_i(s), 0 < s\}$ denote a marker process for individual i with history denoted by $\mathcal{X}_i(t) = \{X_i(s), 0 < s < t\}$. Note that while the marker values change in continuous time, the marker process is typically under intermittent observation because it is only measured upon visits to a clinic at random times. We suppose that there is a baseline measurement of the marker at $s = 0$ corresponding to clinic entry, say, and let $\{A_i(s), 0 < s\}$ denote the counting process for follow-up clinic visits. We let $d\bar{A}_i(s) = \bar{Y}_i(s)dA_i(s)$ indicate that a visit has been made at time s and $\bar{A}_i(t) = \int_0^t d\bar{A}_i(s)$ denote the cumulative number of visits over $(0, t]$; the associated process is denoted by $\{\bar{A}_i(s), 0 < s\}$. At any time t, the cumulative number of follow-up visits to have taken place for individual i is denoted by $\bar{A}_i(t^-)$ and the realized times of these visits are denoted by a_{ir}, $r = 1, \ldots, \bar{A}_i(t^-)$.

Conventionally, researchers will use the most recently recorded value of the marker in failure time analyses and we denote this value at time t as $X_i^\circ(t) = X_i(a_{ir})$, where $r = \bar{A}_i(t^-)$, $t > 0$; the difference $X_i^\circ(t) - X_i(t)$ represents the error in the covariate value arising from the intermittent measurement process and the carry-forward approach to deal with the incomplete information. We let $\mathcal{X}_i^\circ(t) = \{(a_{ir}, X_i^\circ(a_{ir})), r = 0, \ldots, \bar{A}_i(t^-)\}$ denote the history of the observed marker measurements and write the resulting observed data for individual i as $\mathcal{D}_i = \{(V_i, \delta_i), \mathcal{X}_i^\circ(V_i)\}$. In many applications, $X_i(t)$ is a discrete marker which indicates the level of a continuous marker, the severity of a symptom, or the stage of a disease process. Binary markers are particularly common, which may indicate the presence of an acute condition such as the elevation of inflammatory markers in autoimmune diseases (Pearle et al., 2007), elevated HbA1c in diabetics indicating poor blood glucose control (Tarim et al., 1999), or elevated counts of circulating tumour cells in cancer (Tibbe et al., 2007).
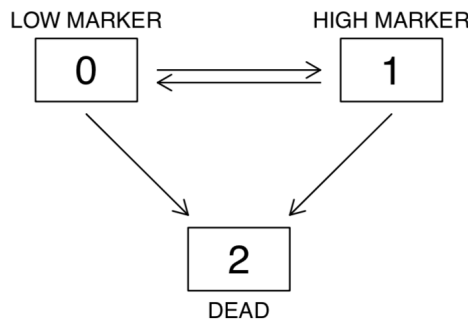

Figure 1: A three-state model for joint consideration of marker process and death

Since this is a common setting with a relatively simple structure we focus on it in order to gain insight into the consequence of using the observed value in analyses and to discuss a modeling approach

to mitigate the biases that arise from naive analyses. We therefore consider the multistate process in Figure 1 and let $Z_i(s)$ represent the state occupied at time s for individual i in this modeling framework where we consider both the marker process and failure process together. In this framework, we may define transition intensities

$$\lim_{\Delta t \downarrow 0} \frac{P(Z(t + \Delta t^-) = l \mid Z(t^-) = k, \mathcal{H}(t))}{\Delta t} = \lambda_{kl}(t \mid \mathcal{H}(t)) , \tag{1}$$

where $\mathcal{H}(t) = \{Z(s), 0 < s < t\}$ is the history (Cook and Lawless, 2018) and $k, l \in \{0, 1, 2\}$. For Markov models, we note $\lambda_{kl}(t|\mathcal{H}(t)) = \lambda_{kl}(t)$ and for time-homogeneous Markov models $\lambda_{kl}(t) = \lambda_{kl}$.

2.2 KEY ESTIMATING FUNCTION THEORY FOR COX REGRESSION

We consider a simple Cox regression model for modeling the effect of the time-dependent binary marker $X(t)$ on the intensity for death which has the form

$$d\Lambda(t|X_i(t)) = d\Lambda_0(t) \exp(X_i(t)\beta) , \tag{2}$$

where $d\Lambda_0(t)/dt = \lambda_{02}(t)$ is the baseline ($0 \to 2$) intensity for death at time t and $\lambda_{12}(t)/\lambda_{02}(t) = \exp(\beta)$ reflects the multiplicative effect of an elevated marker on the instantaneous risk of death. The Cox partial likelihood score equations for β and $d\Lambda_0(s)$ are

$$U_1 = \sum_{i=1}^{m} \int_0^{\infty} \bar{Y}_i(s)\{dN_i(s) - \bar{Y}_i(s)d\Lambda(s|X_i(s))\}X_i(s) \tag{3a}$$

$$U_2(s) = \sum_{i=1}^{m} \bar{Y}_i(s)\{dN_i(s) - d\Lambda(s|X_i(s))\}, \quad 0 < s , \tag{3b}$$

respectively (Kalbfleisch and Prentice, 2011). Setting $U_2(s) = 0$ and solving for $d\Lambda_0(s)$ gives the Breslow "profile" estimate

$$d\tilde{\Lambda}_0(s; \beta) = \frac{\sum_{i=1}^{m} \bar{Y}_i(s)dN_i(s)}{\sum_{i=1}^{m} \bar{Y}_i(s) \exp(X_i(s)\beta)} ,$$

which upon substitution into (3a) gives the partial score function for β as

$$U(\beta) = \sum_{i=1}^{m} \int_0^{\infty} \left\{ X_i(s) - \frac{R^{(1)}(s; \beta)}{R^{(0)}(s; \beta)} \right\} d\bar{N}_i(s) \tag{4}$$

with $R^{(k)}(s; \beta) = \sum_{i=1}^{m} \bar{Y}_i(s)X_i^{(k)}(s) \exp(X_i(s)\beta)$ where $a^{(k)} = 1$ for $k = 0$ and a for $k = 1$ respectively. Under a correctly specified model $\widehat{\beta}$, the maximum partial likelihood estimator of β defined as the solution to $U(\beta) = 0$, has the property that $\sqrt{m}(\widehat{\beta} - \beta) \sim N(0, I^{-1}(\beta))$ where

$$I(\beta) = \sum_{i=1}^{m} \int_0^{\infty} \left\{ \frac{R^{(2)}(s; \beta)}{R^{(0)}(s; \beta)} - \frac{R^{(1)}(s; \beta)}{R^{(0)}(s; \beta)} \left[\frac{R^{(1)}(s; \beta)}{R^{(0)}(s; \beta)} \right]' \right\} d\bar{N}_i(s) .$$

More generally, Andersen and Gill (1982) show that the solution to the partial score function (4) is consistent for the solution to

$$\int_0^{\infty} \left\{ r^{(1)}(u) - \frac{r^{(1)}(u; \beta)}{r^{(0)}(u; \beta)} r^{(0)}(u) \right\} du = 0 \tag{5}$$

where

$$r^{(k)}(s; \beta) = E\{R^{(k)}(s; \beta)\} = mE\{\bar{Y}_i(s)X_i^{(k)}(s)\exp(X_i(s)\beta)\} \qquad (6)$$

and

$$r^{(k)}(s) = E\{\bar{Y}_i(s)X_i^{(k)}(s)dN_i(s)\} , \qquad (7)$$

for $k = 0, 1$ (see also Struthers and Kalbfleisch (1986) and Lin and Wei (1989)). In the next section, we consider the calculation of these terms to explore the impact on the asymptotic bias of the regression coefficient estimator from different marker measurement schemes under the carry-forward approach. We assume the initial state occupied is 0 or 1 with equal probability so take $P(Z_i(0) = 0) = P(Z_i(0) = 1) = 0.5$.

3 ASYMPTOTIC BIAS FROM CARRYING FORWARD MARKER MEASUREMENTS

3.1 DERIVATIONS OF FUNCTIONS FOR EXPECTATIONS

We write the adopted Cox model using the marker value carried forward from the last assessment as

$$dH(s|X_i^\circ(s)) = dH_o(s)\exp(X_i^\circ(s)\gamma) , \qquad (8)$$

where we use $dH_o(s)$ and γ to indicate the parameters of the working model. To evaluate the limiting value of the estimator $\widehat{\gamma}$ and investigate the asymptotic bias resulting from using the carried-forward marker value, we require (6) and (7) for $k = 0, 1$ with the expectations taken with respect to the underlying process depicted in Figure 1, which we take under the constraint (2). Here, we consider a marker process with Markov transition intensities whereby $\lambda_{kl}(t|\mathcal{H}(t)) = \lambda_{kl}(t)$; to further simplify the calculations we make these time-homogeneous, but the following can be easily carried out for the time nonhomogeneous case. We also consider the setting where the visits arise according to a Poisson process with intensity $\rho(t) = \rho$; calculations under a time-nonhomogeneous visit process are possible in principle using product integration to derive or compute the transition probability matrix for the calculations that follow but we do not explore this here.

For $r^{(1)}(s; \beta)$ note that

$$E\{\bar{Y}_i(s)X_i^\circ(s)\exp(X_i^\circ(s)\beta)\} = \exp(\beta)P(\bar{Y}_i(s)X_i^\circ(s) = 1) , \qquad (9)$$

where

$$P(\bar{Y}_i(s)X_i^\circ(s) = 1) = \sum_{r=0}^{\infty} \int_0^s P(\bar{Y}_i(s)X_i^\circ(s) = 1, A_i(s) = r, a_{ir})da_{ir}$$

$$= P(C_i \geq s)\sum_{r=0}^{\infty}\left[\int_0^s P(Z_i(s^-) < 2|Z_i(a_{ir}) = 1, a_{ir}, A_i(s) = r)\right.$$

$$\times P(Z_i(a_{ir}) = 1|a_{ir}, A_i(s) = r)f(a_{ir}|A_i(s) = r)da_{ir}\Big] P(A_i(s) = r) .$$

The density $f(a_{ir}|A_i(s) = r)$ is that of the maximum order statistic for a set of $A_i(s) = r$ independent uniform random variables over $(0, s]$ since we consider the case where $A_i(s)$ is a Poisson random variable; this has the form $f(a_{ir}|A_i(s) = r) = ra_{ir}^{r-1}/s^r$. For $r^{(0)}(s; \beta)$, we require $r^{(0)}(s; \beta) = m \cdot E\{\bar{Y}_i(s)\exp(X_i^\circ(s)\beta)\}$. Note that

$$E\{\bar{Y}_i(s)\exp(X_i^\circ(s)\beta)\} = \left[e^\beta P(\bar{Y}_i(s) = 1, X_i^\circ(s) = 1) + P(\bar{Y}_i(s) = 1, X_i^\circ(s) = 0)\right] ,$$

and

$$P(\bar{Y}_i(s) = 1, X_i^\circ(s) = 0) = \sum_{r=0}^{\infty} \int_0^s P(\bar{Y}_i(s) = 1, X_i^\circ(s) = 0, A_i(s) = r, a_{ir}) da_{ir}$$

$$= P(C_i > s) \sum_{r=0}^{\infty} \left[\int_0^s P(Z_i(s^-) < 2 | Z_i(a_{ir}) = 0, a_{ir}, A_i(s) = r) \right.$$

$$\left. \times P(Z_i(a_{ir}) = 0 | a_{ir}, A_i(s) = r) f(a_{ir} | A_i(s) = r) da_{ij} \right] P(A_i(s) = r) .$$

For the remaining functions $r^{(0)}(s)$ and $r^{(1)}(s)$ note that

$$E\{\bar{Y}_i(s) dN_i(s)\} = P(C_i \geq s) \left[\lambda_{02}(t) P(Z_i(s^-) = 0) + \lambda_{12}(t) P(Z_i(s^-) = 1) \right] , \qquad (10)$$

and

$$E\{\bar{Y}_i(s) dN_i(s) X_i^\circ(s)\} = P(C_i \geq s) P(dN_i(s) = 1 | Y_i^\dagger(s) = 1, X_i^\circ(s) = 1) P(Y_i^\dagger(s) = 1, X_i^\circ(s) = 1)$$

$$= P(C_i \geq s) \sum_{x=0}^{1} P(dN_i(s) = 1 | Z_i(s) = x, X_i^\circ(s) = 1) P(Z_i(s) = x, X_i^\circ(s) = 1)$$

$$= P(C_i \geq s) \sum_{x=0}^{1} \lambda_{x2}(s) P(Z_i(s) = x, X_i^\circ(s) = 1)$$

respectively, where

$$P(Z_i(s) = x, X_i^\circ(s) = 1) = \sum_{r=0}^{\infty} \int_0^s P(Z_i(s) = x | X_i^\circ(s) = 1, a_{ir}, A_i(s) = r) P(X_i^\circ(s) = 1, a_{ir}, A_i(s) = r) da_{ir}$$

$$= \sum_{r=0}^{\infty} \int_0^s P(Z_i(s) = x | Z_i(a_{ir}) = 1, a_{ir}, A_i(s)) P(Z_i(a_{ir}) = 1 | a_{ir}, A_i(s) = r)$$

$$\times f(a_{ir} | A_i(s) = r) da_{ir} P(A_i(s) = r) .$$

Note that $P(Z_i(a_{ir}) = 1 | a_{ir}, A_i(s) = r) = \sum_{l=0}^{1} P(Z_i(a_{ir}) = 1 | Z_i(0) = l) P(Z_i(0) = l)$ since the visit process is independent of the multistate process.
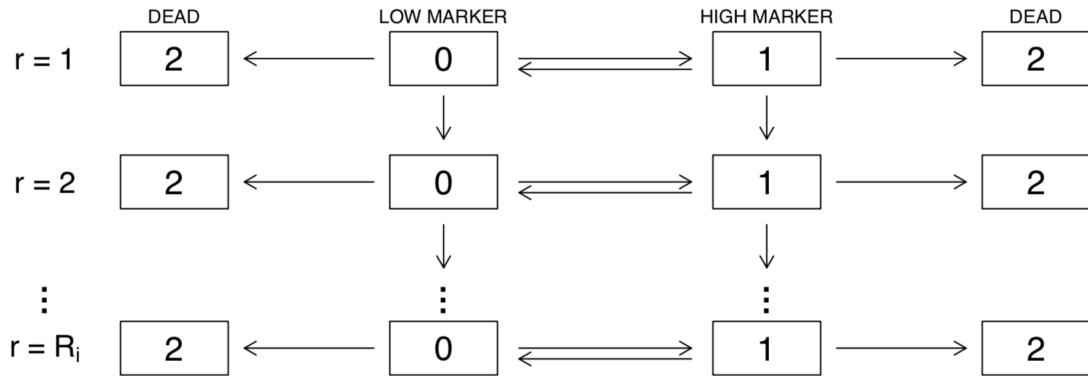

Figure 2: A joint model for a three-state reversible illness-death process and visit process

Figure 2 represents a more general joint model which incorporates the recurrent visit process and we use this in the calculations that follow. Each row represents the three-state process with transitions

between states governed by the process in Figure 1. Transitions from row r to row $r+1$ arise when a visit is made, as in Cook and Lawless (2019). We let

$$\lim_{\Delta t \downarrow 0} \frac{P(\Delta \bar{A}(t) = 1 \mid \mathcal{H}^+(t))}{\Delta t} = \rho(t|\mathcal{H}^+(t)) \tag{11}$$

where in general with $\mathcal{H}^+(t) = \{Y(s), d\bar{N}(s), X(s), d\bar{A}(s); 0 < s < t\}$ is the history of the censoring, joint multistate and visit processes (see Cook and Lawless (2018) for remarks on such joint processes). For a conditionally independent visit process, $\rho(t|\mathcal{H}^+(t)) = \rho(t|\mathcal{H}^\circ(t))$, where $\mathcal{H}^\circ(t) = \{Y(s), d\bar{N}(s), \mathcal{X}^\circ(s), d\bar{A}(s), 0 < s < t\}$ is the history of the observed process (Cook and Lawless, 2018). Here, however, we assume a completely independent visit process in which $\rho(t|\mathcal{H}^+(t)) = \rho(t)$ is a Poisson rate function.

We now have the functions for evaluation of (5), which we can then solve to assess the asymptotic bias in the regression coefficient that results from carrying forward the most recently recorded marker value.

3.2 ILLUSTRATIVE CALCULATIONS OF THE ASYMPTOTIC BIAS IN THE REGRESSION COEFFICIENT ESTIMATOR

We consider the case in which the intention is to observe an individual over an interval $(0, C^A]$ where, without loss of generality, we set $C^A = 1$ as the administrative censoring time. Time-homogenous intensities are adopted with λ_{kl} the intensity for a transition from state k to l; the 3×3 transition probability matrix $P(s, t)$ with (k, l) entry $p_{kl}(s, t) = P(Z_i(t) = l|Z_i(s) = k)$ is then easily computed using a matrix exponential of the intensity matrix. We consider the probability of failure by C^A as $P(Z_i(C^A) = 2|Z_i(0) = 0) = \pi_F$ and set it equal to 0.40 and 0.80. We set $\lambda_{12}/\lambda_{02} = \exp(\beta)$. The mean sojourn time in state 0 is denoted by $\mu_0 = (\lambda_{01} + \lambda_{02})^{-1}$ and we let $\pi_{01} = \lambda_{01}/(\lambda_{01} + \lambda_{02})$ be the probability of a $0 \rightarrow 1$ transition being realized whenever state 0 is occupied; we assume $P(Z_i(0) = 0) = 1$. We adopt a Poisson process for the visits with rate $\rho = 4$ and 8.

Figure 3 displays the asymptotic percent relative bias of the regression coefficient given by $100(\gamma^* - \beta)/\beta$, when the true value is determined so that $\exp(\beta) = 1.25$. In Figure 3 (a), we consider the setting with failure probability $\pi_F = 0.4$ and visit intensity $\rho = 4$ giving an average of four post-baseline visits per individual. The horizontal axis is π_{01} and different lines are plotted corresponding to $R = \lambda_{10}/\lambda_{01} = 4$ and 8 to illustrate how the bias depends on these two features within the given setting. Note that the bias increases as the probability of a $0 \rightarrow 1$ transition increases and is higher when the time spent in state 1 is lower (i.e. with higher values of R). Figure 3 (b) corresponds to the setting with $\pi_F = 0.4$ and $\rho = 8$, where it can be seen, as expected, that the asymptotic bias is lower with more frequent assessments. Figure 3 (c) and (d) show similar patterns for the settings with an 80% failure rate by the administrative censoring time (i.e. $\pi_F = 0.8$) with $\rho = 4$ and $\rho = 8$, respectively. Similar patterns in the asymptotic bias are seen for the regression coefficient at different values of β with the biases tending to be larger with values of β farther from zero.
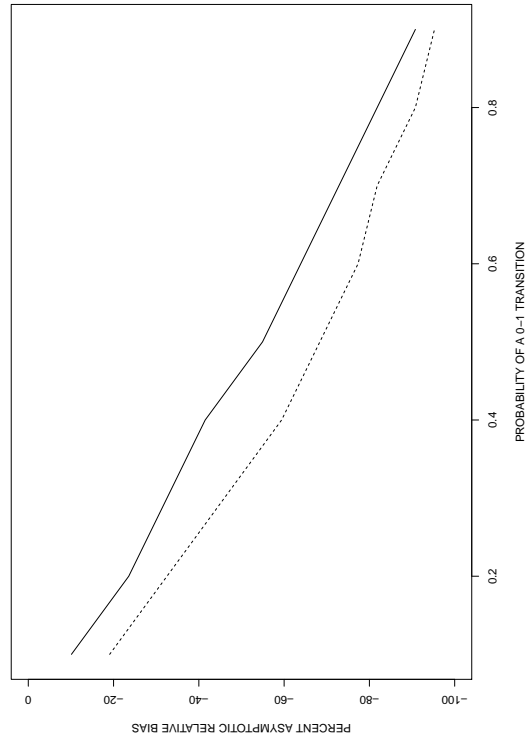
4 IMPLEMENTATION OF A JOINT MODEL FOR THE MARKER AND FAILURE TIME

4.1 LIKELIHOOD CONSTRUCTION

The bias mentioned in the previous section can be mitigated if a joint model is formed based on the three-state process of Figure 1. This approach requires modeling the covariate process and we do so here in the parametric setting of the simulation studies. Under the assumption that we have a
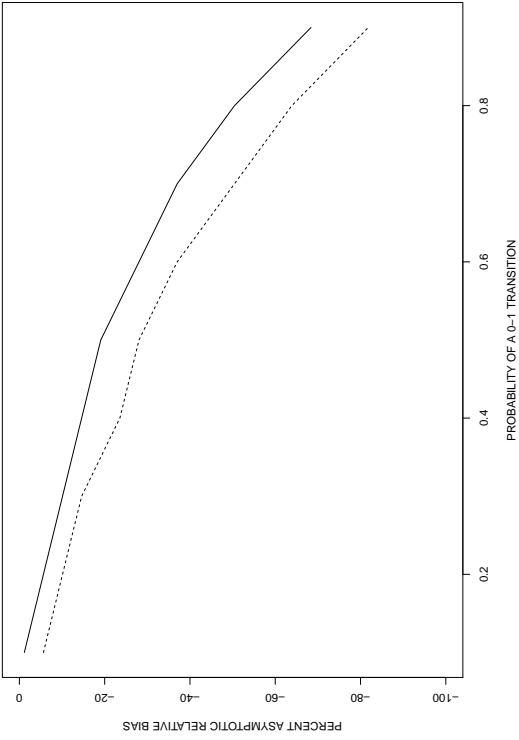
(a) 40% failure rate ($\pi_F = 0.40$), $\rho = 4$

(b) 40% failure rate ($\pi_F = 0.40$), $\rho = 8$

(c) 80% failure rate ($\pi_F = 0.80$), $\rho = 4$
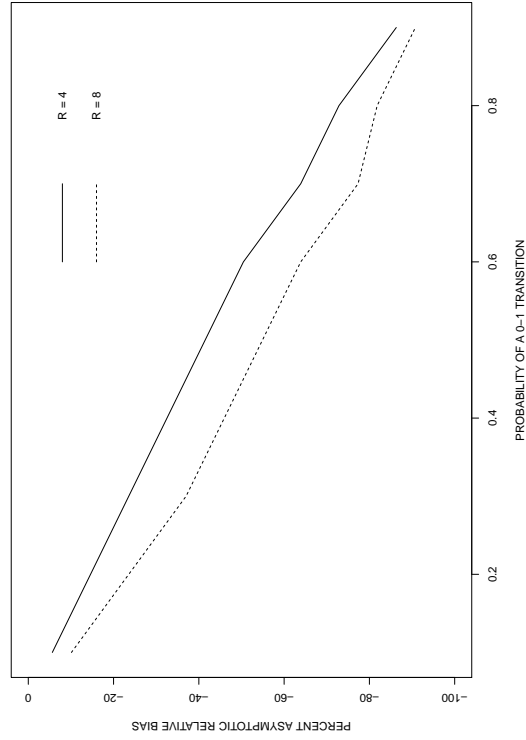
(d) 80% failure rate ($\pi_F = 0.80$), $\rho = 8$

Figure 3: Percent asymptotic relative biases of regression coefficient from Cox model ($100(\gamma^* - \beta)/\beta$) as a function of π_{01} and $R = \lambda_{10}/\lambda_{01}$ when a time-dependent marker value is carried forward from the most recent assessment for different failure rates and visit frequencies; $\beta = \log 1.25$

conditionally independent visit process, then the partial likelihood (Cook and Lawless, 2019) becomes

$$\prod_{i=1}^{m} \left\{ \prod_{r=1}^{\bar{A}_i(V_i)} P(Z_i(a_{ir})|Z_i(a_{i,r-1})) \right\} \left\{ \sum_{k=0}^{1} P(Z_i(V_i^-) = k|Z_i(a_{i\bar{A}_i(V_i)})) \, \lambda_{k2}(V_i|Z(V_i^-) = k) \right\}^{\delta_i}.$$

(12)

For many processes, the probabilities above may be difficult to express in terms of the intensity functions used to specify the models. Researchers often model such data using Markov assumptions (Cook and Lawless, 2014) for which the transition intensities are of the form $\lambda_{kl}(t|\mathcal{H}(t)) = \lambda_{kl}(t|Z(t^-) = k)$, that is, the instantaneous risk of the event, given the entire history, is governed by simply the current marker state. The R `msm` library by Jackson (2011) contains functions that facilitate the sort of multistate analyses we employ here by dealing with intermittent observation regarding the occupancy of some states and right-censored state entry times for others as above.

4.2 SIMULATION STUDIES OF FINITE SAMPLE PROPERTIES

Here, we report on simulation studies designed to examine the empirical bias of naive analyses based on the Cox model in which the values of the time-dependent marker are carried forward until their values are measured again. The goal is to assess whether the empirical bias tracks the asymptotic results and also to examine the performance of the estimators under the joint model using the likelihood in (12).

Individuals are first assigned an initial marker state with equal probability (i.e. $P(Z_i(0) = 0)$ and $P(Z_i(0) = 1)$ with probability 0.5). We consider a common administrative censoring time of $C_i^A = C^A = 1$ with no random censoring. The data on the transition times of the three-state process of Figure 1 are generated using the `msm` function with intensities given by the same parameter settings as in Section 3. Specifically, we consider the case with failure rates of 0.40 and 0.80 over $(0, 1]$, $\lambda_{10}/\lambda_{01} = R = 4$ and 8 and we consider $\pi_{01} = 0.25, 0.50$ and 0.75. The visit times are generated according to a Poisson process with visit intensities leading to an average of 4 or 8 visits over $(0, 1]$. For each simulated dataset, a Cox regression model is fitted by carrying forward the marker value at the most recent assessment; a joint three-state model is also fitted based on the likelihood in Section 4.1. Two thousand datasets of $n = 200$ or $n = 2000$ individuals each were simulated with the results displayed in Tables 1 and 2, respectively. We report the empirical bias (EBIAS) along with the asymptotic bias (ABIAS) calculated based on Section 3.2, the empirical standard error (ESE), the average robust standard error (ASE), and the empirical coverage probability percentage (ECP%) of nominally 95% confidence intervals (CIs).

For the misspecified Cox model, we find good agreement between the asymptotic and empirical bias. The empirical coverage based on the Cox model is reasonable in some settings when the failure rate is low, but as the failure rate increases the impact of model misspecification increases as the biases become larger and the empirical coverage probabilities become poor (too low). For the analysis based on the correct likelihood, the empirical biases are negligible as anticipated, as the maximum likelihood estimators are consistent, and the empirical coverage probabilities are generally compatible with the nominal 95% level. We note that when the sample size is $n = 200$ and visits are infrequent ($\rho = 4$), there are some settings where the empirical coverage probability is slightly above the nominal level, but we attribute this to the finite sample performance of maximum likelihood estimators; the liberal nature of the confidence intervals in modest samples can be kept in mind when interpreting the results of the analyses that follow.

Table 1: Empirical performance of estimators based on 2000 simulations with $n = 200$ individuals per sample.

π_F	R	π_{01}	Method	Average of 4 Visits ($\rho = 4$)					Average of 8 Visits ($\rho = 8$)				
				EBIAS	ABIAS	ESE	ASE	ECP%	EBIAS	ABIAS	ESE	ASE	ECP%
0.4	4	0.25	Cox	-0.032	-0.031	0.604	0.609	95.6	-0.038	-0.024	0.549	0.568	96.5
			Joint	-0.045		0.710	0.703	95.9	-0.040		0.585	0.595	96.0
		0.50	Cox	-0.074	-0.064	0.447	0.428	95.3	-0.050	-0.044	0.383	0.381	96.2
			Joint	-0.069		0.586	0.595	95.8	-0.058		0.490	0.472	95.7
		0.75	Cox	-0.127	-0.134	0.389	0.362	94.8	-0.129	-0.094	0.228	0.328	95.0
			Joint	-0.090		0.708	0.702	95.0	-0.086		0.604	0.598	95.7
	8	0.25	Cox	-0.023	-0.044	0.606	0.648	95.8	-0.007	-0.025	0.576	0.583	95.5
			Joint	-0.031		0.691	0.704	96.3	-0.030		0.669	0.701	95.7
		0.50	Cox	-0.127	-0.094	0.536	0.511	96.9	-0.106	-0.064	0.528	0.519	96.5
			Joint	-0.064		0.727	0.780	96.1	-0.060		0.654	0.690	95.7
		0.75	Cox	-0.175	-0.154	0.497	0.461	97.0	-0.135	-0.124	0.423	0.410	97.1
			Joint	-0.039		0.851	0.890	95.3	-0.053		0.751	0.780	95.3
0.8	4	0.25	Cox	-0.061	-0.064	0.333	0.334	96.5	-0.044	-0.044	0.292	0.294	95.8
			Joint	-0.035		0.457	0.446	95.6	-0.022		0.362	0.348	95.8
		0.50	Cox	-0.131	-0.124	0.279	0.274	93.8	-0.095	-0.094	0.252	0.245	93.3
			Joint	-0.055		0.534	0.520	95.6	-0.037		0.379	0.375	94.7
		0.75	Cox	-0.193	-0.174	0.258	0.255	91.1	-0.148	-0.154	0.230	0.229	91.1
			Joint	-0.043		0.810	0.844	96.9	-0.081		0.587	0.568	96.2
	8	0.25	Cox	-0.090	-0.094	0.397	0.393	97.0	-0.065	-0.066	0.364	0.342	95.2
			Joint	-0.065		0.640	0.671	95.3	-0.061		0.510	0.499	95.8
		0.50	Cox	-0.156	-0.151	0.357	0.342	94.1	-0.135	-0.124	0.313	0.304	94.8
			Joint	-0.022		0.837	0.868	95.6	-0.072		0.620	0.642	96.0
		0.75	Cox	-0.218	-0.194	0.339	0.328	93.6	-0.218	-0.174	0.300	0.294	92.3
			Joint	-0.045		0.810	0.823	95.5	-0.027		0.787	0.789	95.2

Abbreviations: ABIAS, asymptotic bias; ASE, average robust standard error; EBIAS, empirical bias; ECP, empirical coverage probability; ESE, empirical standard error.

Table 2: Empirical performance of estimators based on 2000 simulations with $n = 2000$ individuals per sample.

π_F	R	π_{01}	Method	Average of 4 Visits ($\rho = 4$)					Average of 8 Visits ($\rho = 8$)				
				EBIAS	ABIAS	ESE	ASE	ECP%	EBIAS	ABIAS	ESE	ASE	ECP%
0.4	4	0.25	Cox	-0.033	-0.031	0.181	0.177	95.5	-0.025	-0.024	0.160	0.158	94.9
			Joint	-0.008		0.194	0.192	95.4	-0.007		0.170	0.167	94.8
		0.50	Cox	-0.064	-0.064	0.127	0.128	93.7	-0.047	-0.044	0.117	0.116	94.2
			Joint	-0.009		0.164	0.163	95.5	<0.001		0.134	0.136	95.5
		0.75	Cox	-0.125	-0.134	0.110	0.110	82.5	-0.094	-0.094	0.099	0.100	84.2
			Joint	-0.009		0.186	0.191	95.8	-0.001		0.149	0.150	96.0
	8	0.25	Cox	-0.043	-0.044	0.195	0.193	94.8	-0.031	-0.025	0.168	0.171	95.3
			Joint	-0.019		0.227	0.226	96.0	-0.005		0.185	0.189	95.1
		0.50	Cox	-0.095	-0.094	0.151	0.149	92.2	-0.064	-0.064	0.137	0.135	93.4
			Joint	-0.014		0.223	0.223	95.5	-0.009		0.174	0.178	95.9
		0.75	Cox	-0.162	-0.154	0.137	0.136	80.7	-0.129	-0.124	0.123	0.124	84.5
			Joint	-0.020		0.313	0.308	95.6	-0.013		0.224	0.227	95.4
0.8	4	0.25	Cox	-0.062	-0.064	0.103	0.102	91.6	-0.042	-0.044	0.093	0.091	92.3
			Joint	-0.004		0.123	0.126	96.0	<0.001		0.105	0.105	94.9
		0.50	Cox	-0.123	-0.124	0.086	0.085	71.3	-0.095	-0.094	0.077	0.076	77.7
			Joint	-0.003		0.139	0.142	95.6	-0.004		0.112	0.112	94.8
		0.75	Cox	-0.175	-0.174	0.079	0.079	40.4	-0.154	-0.154	0.071	0.071	42.7
			Joint	-0.008		0.209	0.207	96.0	-0.006		0.157	0.155	94.6
	8	0.25	Cox	-0.093	-0.094	0.122	0.118	88.5	-0.068	-0.066	0.104	0.104	91.2
			Joint	-0.007		0.170	0.171	95.5	-0.003		0.137	0.135	94.8
		0.50	Cox	-0.152	-0.151	0.106	0.104	68.2	-0.128	-0.124	0.092	0.093	73.4
			Joint	-0.009		0.225	0.223	95.1	-0.007		0.168	0.167	95.2
		0.75	Cox	-0.193	-0.194	0.102	0.100	49.0	-0.177	-0.174	0.089	0.090	50.1
			Joint	-0.025		0.362	0.356	95.5	-0.013		0.257	0.255	95.4

Abbreviations: ABIAS, asymptotic bias; ASE, average robust standard error; EBIAS, empirical bias; ECP, empirical coverage probability; ESE, empirical standard error.

5 MARKERS AND SKELETAL COMPLICATIONS IN METASTATIC BREAST CANCER

Individuals with cancer metastatic to bone are at risk of skeletal complications such as bone pain, hypercalcemia and need for surgery, all of which impact quality of life and disability. We consider the relationship between levels of the marker bone alkaline phosphatase (BALP), a marker of bone formation, and the risk of the composite event of skeletal complication or death in prostate cancer patients receiving standard therapy from the trial reported in Saad et al. (2004). Levels of BALP change continuously over time, but they are only measured when blood samples are taken at the periodic follow-up assessments. We carry out a standard Cox regression analysis using the most recently recorded value of the bone marker, as well as an analysis based on jointly modeling the marker state and the composite endpoint in a five-state model; R code for the joint analysis is available on request. Figure 4 contains a multistate diagram with marker states defined based on quartiles of BALP at the time of recruitment where 1 to 4 corresponding to ranges $0 \leq \text{BALP} < 150.25$, $150.25 \leq \text{BALP} < 267.50$, $267.50 \leq \text{BALP} < 529.75$, $529.75 \leq \text{BALP}$, and state 5 representing the clinical event of interest (skeletal event or death). We let $Z(t)$ represent the state occupied at time t so that $Z(t) = k$ if an individual is in state k at time t, for example, $Z(t) = 3$ if at time t an individual is alive and has a BALP value in the interval $[267.50, 529.75)$. The transitions between states are again governed by intensity functions, which in this context reflect the instantaneous risk of movement from one state to another given the history of the process. We adopt functionally independent piecewise constant transition intensities for all transitions between the marker states with two cut points at 90 days and 180 days giving three pieces. For the transitions into the absorbing state defined by the clinical event of interest, we set

$$\lambda_{k5}(t|\mathcal{H}(t)) = \lambda_{15}(t|\mathcal{H}(t)) \exp(I(Z(t^-) = 2)\beta_2 + I(Z(t^-) = 3)\beta_3 + I(Z(t^-) = 4)\beta_4)$$

so that the regression coefficients characterize the ratio of the instantaneous risk of the event across the different marker states.
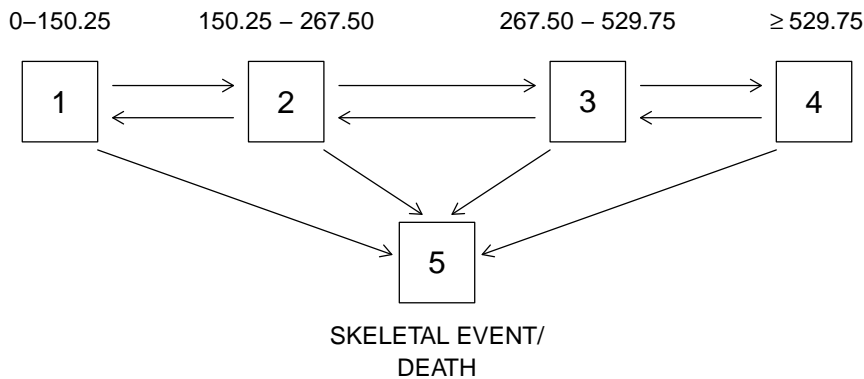

Figure 4: A five-state model for bone alkaline phosphatase and the composite event of skeletal complication and death

In the first set of columns of Table 3, we report the hazard ratio estimates from fitting a Cox regression model using the customary strategy of carrying forward the most recently recorded marker value; the covariate levels are defined as they are in Figure 4. In addition we report the relative risks (RRs) for transitions into the absorbing state based on the fit of the multistate model with the higher marker states compared to the reference marker state of $0 \leq \text{BALP} < 150.25$. When comparing the estimates from the Cox regression to the estimates from the multistate analysis, it is apparent that the point

Table 3: Estimates of relative risk obtained from fitting a Cox regression model and a joint model with piecewise-constant intensities to the skeletal event-free survival in $N = 202$ placebo patients from Saad et al. (2004)

BALP Level (IU/ℓ)	Cox Regression Model					Joint Model				
	Est.	S.E.	RR	95% CI	p$-$value	Est.	S.E.	RR	95% CI	p$-$value
$0 - 150.25$	-	-	-	-	-	-	-	-	-	-
$150.25 - 267.50$	0.418	0.331	1.518	(0.793, 2.907)	0.2077	0.585	0.483	1.794	(0.696, 4.628)	0.2264
$267.50 - 529.75$	1.007	0.305	2.738	(1.505, 4.981)	0.0001	1.053	0.431	2.867	(1.232, 6.675)	0.0145
≥ 529.75	1.356	0.282	3.880	(2.233, 6.743)	< 0.0001	1.580	0.356	4.855	(2.417, 9.751)	< 0.0001

Abbreviations: BALP, bone alkaline phosphatase; CI, confidence interval; RR, relative risk.

estimates of the RRs from the Cox model are attenuated compared to those of the multistate model, a relation in alignment with the simulation results. Generally, we see an increased risk with higher values of BALP; for example the RR for the highest category compared to the reference category is $RR = 4.85$ (95% CI: 2.42, 9.75) from the joint model and $RR = 3.88$ (95% CI: 2.23, 6.74) from the Cox model. Figure 5 contains plots of estimates of the cumulative probability of the composite event based on a marginal Kaplan-Meier estimate and based on the five-state joint model of the BALP and failure process; the latter is obtained by computing an estimate of the absorption probability in the five-state model of Figure 4. Since the former is robust, the agreement between these two estimates does not suggest evidence that the Markov assumptions of the five-state model are implausible.
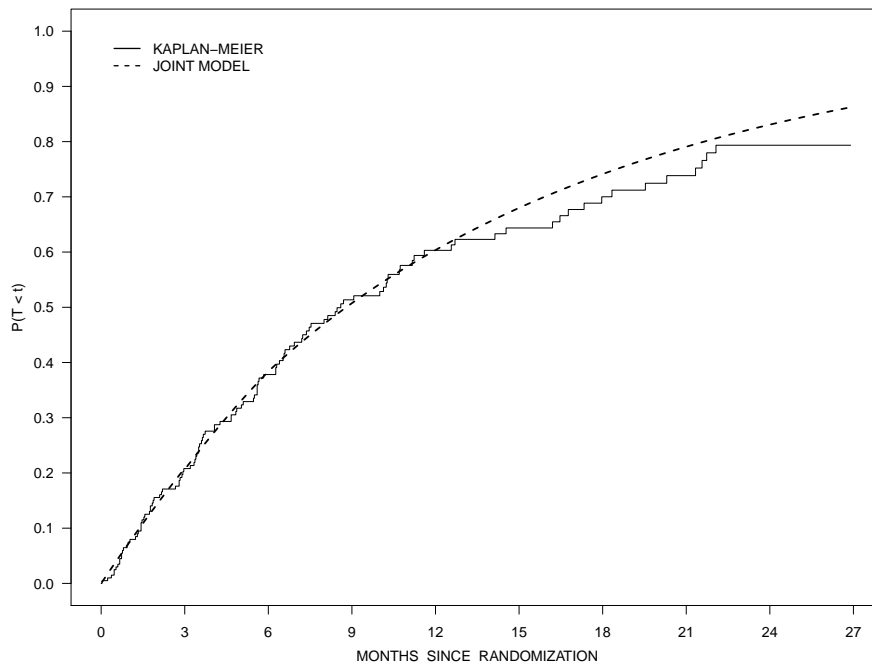

Figure 5: The cumulative probability of the composite endpoint of skeletal complication or death based on a marginal Kaplan-Meier estimate and the joint five-state model

6 DISCUSSION

We consider the common setting in which markers that vary over time are measured at periodic assessment times, which we refer to as clinic visits. It is well known that if the current covariate acts multiplicatively on the hazard function for failure inconsistent estimates are obtained when the most recently recorded value is used in the analysis (Bruijne et al., 2001; Andersen and Liestol, 2003), though much of this work has been directed at the study of continuous markers. We have investigated the asymptotic bias of estimators arising from model fitting with the common convention of carrying forward the marker value that was most recently recorded. We found that there is generally an attenuation of the estimated effect, with a greater attenuation arising when the visit intensity is lower and assessments are less frequent. For a given visit intensity, the bias is greater with a more rapidly changing marker value, as reflected here by the ratio $\lambda_{10}/\lambda_{01}$. Moreover, as noted for other settings involving misspecified Cox regression models, the asymptotic bias is lower when the censoring rate is higher since there is less information conveyed about the effect (Wu and Cook, 2012; Rufibach, 2019). Other failure time models could be considered of course, including additive models (Aalen, 1989), the Cox-Aalen model (Scheike and Zhang, 2002) and accelerated failure time models (Robins and Tsiatis, 1992), but we anticipate the same general dependence on these factors would remain.

The approach to dealing with the intermittent observation of the marker process is the formation of a joint model for the marker and failure times. There is a large literature on joint modeling of longitudinal marker data and failure times, but this too has primarily been directed at continuous markers (Tsiatis and Davidian, 2004; Rizopoulos, 2012). Handling of discrete marker processes is naturally done by casting the joint model into the multistate framework, an approach which is particularly amenable under Markov assumptions. We considered parametric or weakly parametric joint multistate models and assumed that the censoring time and visit process were completely independent of the marker process. Tests of the assumption that the visit process is independent of the joint process are possible by conceptualizing a more general joint model of the failure, marker, and visit processes. Cook and Lawless (2019) consider joint multistate-visit processes and discuss model fitting and identifiability issues. This expanded state space would require larger datasets if model fitting were desired to provide protection against various types of dependent visit processes, but they could also be used as a basis for score tests if simple diagnostic checks were desired.

Robustness of the model could be further improved by adopting a piecewise constant transition intensities between the marker states. Random effects model could also be considered by adding in the individual level variation accounting for more frequent or slower cycling through the states. Finally, we note that the joint modeling approach can also be used when there is more than one clinical event of interest; joint marker and illness-death models can be considered in the semi-competing risks setting.

The approach in this article could be extended to deal with higher dimensional marker processes by expanding the state space to accommodate different combination of marker states; of course, larger datasets and more frequent visit times would be desirable to ensure estimability challenges do not create problems in model fitting. As an example, N-telopeptide of type I collagen is a marker of bone resorption that has been found to be associated with the presence and volume of skeletal metastases and it could be studied in the prostate cancer setting of the illustrative example in combination with the bone alkaline phosphatase marker examined here. The more markers or other time-dependent covariates considered, the more states are required in the state space of the joint model and so one may need to be careful in selecting a modest number of states for each marker to avoid estimability problems.

ACKNOWLEDGEMENTS

DATA ACCESSIBILITY

The data used in this analysis are from a third party and are not available for distribution.

CONFLICT OF INTEREST

The authors declare no potential conflicts of interests.

REFERENCES

Aalen, O. (1989). A linear regression model for the analysis of life times. *Statistics in Medicine*, 8(8):907–925.

Andersen, P. and Gill, R. (1982). Cox's regression model for counting processes: a large sample study. *Annals of Statistics*, 10(4):1100–1120.

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

Bruijne, M., Cessie, S., Kluin-Nelemans, H., and Houwelingen, H. (2001). On the use of cox regression in the presence of an irregularly observed time-dependent covariate. *Statistics in Medicine*, 20:3817 – 3829.

Buonaccorsi, J. (2010). *Measurement Error: Models, Methods, and Applications*. Chapman and Hall/CRC.

Carroll, R., Ruppert, D., Stefanski, L., and Crainiceanu, C. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*. Chapman and Hall/CRC.

Cook, R. and Lawless, J. (2014). Statistical issues in modeling chronic disease in cohort studies. *Statistics in Biosciences*, 6(1):127–161.

Cook, R. and Lawless, J. (2018). *Multistate Models for the Analysis of Life History Data*. Chapman and Hall/CRC.

Cook, R. and Lawless, J. (2019). Independence conditions and the analysis of life history processes under intermittent observation. *Biostatistics (on-line ahead of print)*, pages 1–25.

Cox, D. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202.

Degruttola, V. and Tu, X. (1994). Modeling progression of CD-4 lymphocyte count and its relation to survival time. *Biometrics*, 50:1003 – 1014.

Fuller, W. (2009). *Measurement Error Models*, volume 305. John Wiley & Sons.

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

Kalbfleisch, J. and Prentice, R. (2011). *The Statistical Analysis of Failure Time Data*, volume 360. John Wiley & Sons.

Lin, D. and Wei, L. (1989). The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association*, 84(408):1074–1078.

Pearle, A., Scanzello, C., George, S., Mandl, L., DiCarlo, E., Peterson, M., Sculco, T., and Crow, M. (2007). Elevated high-sensitivity c-reactive protein levels are associated with local inflammatory findings in patients with osteoarthritis. *Osteoarthritis and Cartilage*, 15(5):516–523.

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*, 156:393 – 406.

Rizopoulos, D. (2012). *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. Chapman and Hall/CRC.

Robins, J. and Tsiatis, A. (1992). Semiparametric estimation of an accelerated failure time model with time-dependent covariates. *Biometrika*, 79(2):311–319.

Rufibach, K. (2019). Treatment effect quantification for time-to-event endpoints–estimands, analysis strategies, and beyond. *Pharmaceutical Statistics*, 18(2):145–165.

Saad, F., Gleason, D., Murray, R., Tchekmedyian, S., Venner, P., Lacombe, L., Chin, J., Vinholes, J., Goas, J., Zheng, M., and for the Zoledronic Acid Prostate Cancer Study Group (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.

Scheike, T. and Zhang, M. (2002). An additive–multiplicative cox–aalen regression model. *Scandinavian Journal of Statistics*, 29(1):75–88.

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

Tarim, Ö., Küçükerdogan, A., Günay, Ü., Eralp, Ö., and Ercan, İ. (1999). Effects of iron deficiency anemia on hemoglobin A1c in type 1 diabetes mellitus. *Pediatrics International*, 41(4):357–362.

Tibbe, A., Miller, M., and Terstappen, L. (2007). Statistical considerations for enumeration of circulating tumor cells. *Cytometry Part A: The Journal of the International Society for Analytical Cytology*, 71(3):154–162.

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

Tsiatis, A., Degruttola, V., and Wulfsogn, M. (1995). Modeling the relationship of survival to longitudinal data measured with error. application to survival data and CD4 counts in patients with AIDS. *Journal of the American Statistical Association*, 90:27 – 37.

Wu, L. and Cook, R. (2012). Misspecification of cox regression models with composite endpoints. *Statistics in Medicine*, 31(28):3545–3562.

Xu, J. and Zeger, S. (2001). Joint analysis of longitudinal data comprising repeated measures and times to events. *Applied Statistics*, 50:375 – 387.

Yi, G. (2016). *Statistical Analysis with Measurement Error or Misclassification*. Springer.