Published in final edited form as:
Scandinavian Journal of Statistics (2020), **47**(4): 1035–1063
DOI: 10.1111/sjos.12471

Failure time studies with intermittent observation and losses to follow-up

RICHARD J. COOK*

Department of Statistics and Actuarial Science,
University of Waterloo, Waterloo, ON, N2L 3G1, Canada
E-mail: rjcook@uwaterloo.ca

JERALD F. LAWLESS*

Department of Statistics and Actuarial Science,
University of Waterloo, Waterloo, ON, N2L 3G1, Canada

*Richard J. Cook and Jerald F. Lawless contributed equally to this study.

Summary

In health research interest often lies in modeling a failure time process but in many cohort studies failure status is only determined at scheduled assessment times. While the assessment times may be fixed upon study entry, individuals may become lost to follow-up and miss visits subsequent to the time of loss to follow-up. We consider a three-state model to characterize a joint failure and loss to follow-up process, and use it to investigate the impact of dependent loss to follow-up on standard parametric, nonparametric, and semiparametric analysis. The effect of dependent loss to follow-up is mitigated by fitting the joint model. The performance of standard methods is studied using the asymptotic theory of misspecified models, and the finite sample performance is examined for the standard and joint analyses through simulation studies. An application to data from a youth smoking prevention study is presented for illustration.

Keywords: dependent censoring, failure time analysis, joint modeling, loss to follow-up, model misspecification

This is the peer reviewed version of the following invited article: Richard J. Cook and Jerald F. Lawless, Failure time studies with intermittent observation and losses to follow-up, *Scandinavian Journal of Statistics* (2020), 47(4): 1035–1063 which has been published in final form at https://doi.org/10.1111/sjos.12471.

1 INTRODUCTION

Longitudinal studies of event occurrences or failure time outcomes typically involve intermittent observation of study individuals at scheduled times; for convenience we refer to the occasions at which individuals are seen as visits (e.g. to a clinic or assessment facility) and the observation times as visit times. The status (failed versus unfailed) of an individual is determined and covariates may be measured at each visit. When an individual is determined to have failed for the first time (i.e. they have failed since the previous visit) their failure time is interval-censored. Sometimes the exact failure time can be ascertained retrospectively; we consider both cases here. Individuals may also become lost to follow-up (LTF). For example, a person is often deemed to be LTF when they do not appear for a scheduled visit.

Problems arise in dealing with LTF under intermittent observation schemes. First, an exact LTF time is rarely observed; all we know is that a person has not had a visit, scheduled or not, for some period of time. This has lead to discussion about what to use as a putative censoring time for persons LTF (e.g. Shepherd et al., 2011; Lesko et al., 2018). In practice, analysts have used a variety of choices, including the time of the last observed visit, the time of the first missed visit, and a defined time in between these visit times. It has been noted (e.g. Lawless, 2013) that if LTF intensities are different for failed and unfailed individuals, then the independent censoring condition needed by standard methods of analysis does not hold, and standard estimates of failure time distributions are biased under the first choice and, in general, under the other choices as well. Another issue arises if there is a possibility that failure time intensities are different for persons before versus after loss to follow-up. This type of dependent or non-ignorable LTF has been studied extensively in the case where individuals can be continuously monitored; see for example, Fisher and Kanarek (1974); Slud and Rubinstein (1983); Scharfstein and Robins (2002); Siannis (2011). Lawless and Cook (2019) review this area and discuss the use of tracing studies (e.g. Baker et al., 1993; Frangakis and Rubin, 2001; Farewell et al., 2003) that collect failure information on persons after LTF. However, aside from aforementioned studies on how to define LTF times, there has been very little study of settings with intermittent observation. Exceptions include a brief investigation of bias in standard estimates (Lawless, 2013) and studies of the special situation where LTF corresponds to death, for which exact times are ascertainable (e.g. Joly et al., 2002; Binder and Schumacher, 2014; Binder et al., 2017; Binder et al., 2019). We focus here on situations where exact LTF times are not generally ascertainable. Examples are ubiquitous and include, for example, registry studies (Gladman and Chandran, 2011), large national cohort studies in which follow-up assessments are scheduled every 3 years (Raina et al., 2009), and high-school based smoking prevention studies (Cameron et al., 1999).

Our objectives in this paper are to provide a framework through which dependent, non-ignorable LTF can be addressed in studies involving intermittent observation. The framework uses joint failure-loss to follow-up models based on multistate processes in the spirit of Commenges and Gégout-Petit (2007), Lange et al. (2015), and Cook and Lawless (2018). This allows us to discuss the effects of dependent LTF on estimation based on standard follow-up until failure or LTF is confirmed, as well as the utility of auxiliary information collected on individuals after LTF or failure or, in some cases, between visits. Thus, we are able to examine bias in standard failure time analyses that ignore dependent LTF, and to consider how more appropriate analyses may be conducted. We focus first on settings where failure is not a terminal event such as death; we consider death, either on its own or as a competing risk, later in the article. We assume throughout that visit times are conditionally independent in the sense of Grüger et al. (1991) or Cook and Lawless (2007, 2018, 2019); we discuss violations of this condition in Sections 6.1 and 7.

Section 2 introduces joint multistate models for failure time and loss to follow-up and considers three potential observation schemes: standard follow-up until either failure or loss to follow-up is confirmed, and two schemes with extended follow-up for at least some individuals. Maximum likelihood estimation is developed for each observation scheme, along with ways to assess independent LTF conditions. Section 3 examines the bias arising from standard parametric failure time analysis which ignores the LTF mechanism, as well as estimation with joint models. Section 4 considers analogous biases from nonparametric or semiparametric methods under independent LTF assumptions. Section 5 reports the results of an application involving data from a school-based smoking prevention study, where the aim is to illustrate the impact of naive and joint modeling. Section 6 considers the challenges arising when observation times are random as well as when processes are more complex and there is non-negligible risk of death or some other terminal event; time-dependent covariates are also discussed briefly. Concluding remarks are given in Section 7.

2 Intermittent Observation Schemes and Loss to Follow-up

2.1 A Joint Multistate Model

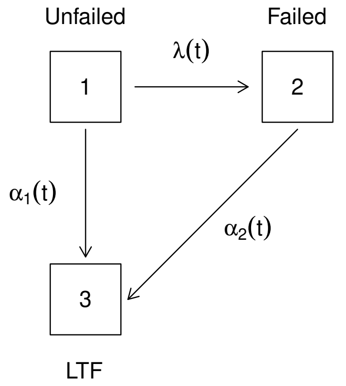
Methods for formulating and fitting multistate models can be found in books such as Cook and Lawless (2018), Andersen et al. (1993), Beyersmann et al. (2012) and Willekens (2014). Here we consider multistate models that include states representing the condition of being lost to follow-up. Lawless (2013) used such models to discuss association for disease processes and loss to follow-up, and to consider joint modeling. In one model shown in Figure 1(a) we display intensity functions (Cook and Lawless, 2018, Section 1.3) for each of three types of transitions; covariates are not included for simplicity but are considered later. We consider failure events that do not preclude further follow-up of an individual, so transitions to the LTF state are possible either before failure (i.e. from state 1) or after failure (i.e. from state 2). In settings where an individual's failure status is only known at the intermittent visit times, consider an individual observed to be in state 1 at visit time a_{j-1} and declared lost to follow-up (and hence in state 3) at their next scheduled visit time a_j. In this case we would not know their underlying disease state at this time; that is, we would not know whether they had passed through state 2 between times a_{j-1} and a_j.

An expanded model is shown in Figure 1(b) where we distinguish the failure status of individuals lost to follow-up; the label 1^p represents the state of being unfailed and LTF, and 2^p the state of being failed and LTF; note that state 2^p can be entered from either state 2 or state 1^p. Such a formulation is appropriate when failures can occur after LTF, which is typically the case. If we are able to obtain data on some individuals after LTF, say through a tracing study, such models can be fit. Figure 1(a) applies to settings where the observation process terminates upon LTF and acquiring data following LTF is not possible. In Figure 1(a), (b), the intensity for LTF from states 1 and 2 is allowed to be different. In Figure 1(b), we also allow the failure intensity to differ for persons before and after LTF; Lawless and Cook (2019) consider this for the case where individuals are continuously rather than intermittently observed.

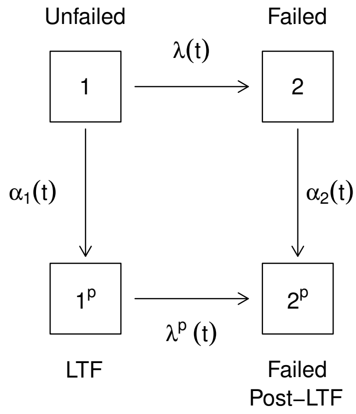
We denote time of failure (entry to state 2 in Figure 1(a) or to states 2 or 2^p in Figure 1(b)) as T. We assume that for a generic individual there is an administrative end of follow-up time A, but that an individual may be prematurely lost to follow-up at time $C < A$; this corresponds to the time of entry to state 3 in Figure 1(a) and 1^p or 2^p in Figure 1(b). As noted earlier, we assume that failure does not preclude further visits, and that the rate of terminal events such as death is negligible and can be ignored; cases where this is not possible are considered in Section 6. We now describe intermittent observation schemes, some types of data, and likelihood functions based on them.

2.2 Observation Schemes and Types of Outcomes

We consider observation schemes where an individual attends a clinic upon study entry at $a_0 = 0$ and is then scheduled for future visits at times $a_1 < a_2 < \cdots$; we consider the case in which data are only collected at these clinic visit times. When the visit times are stochastic, we let A_j denote the



(a)



(b)

Figure 1: Joint models for a failure and loss to follow-up process: (a) a three-state model where LTF terminates the observation and (b) an expanded model accommodating failure post-LTF and LTF post-failure.

random time of the jth visit, which may depend on data observed up to the previous visit time a_{j-1} but, conditional on that, is independent of unobserved process history before and after a_{j-1} (Cook and Lawless, 2018, Section 5.1). At time a_j, data D_j are collected on the failure status and possibly covariates $X(t)$ for the time interval $(a_{j-1}, a_j]$; in some settings it may be possible to ascertain the exact time of failure in an interval, or information on changes in covariate values between a_{j-1} and a_j, but often all that can be obtained are current values $I(T \leq a_j)$, $X(a_j)$.

Here we make a number of simplifying assumptions. First, we assume that follow-up of an individual begins at time $t = 0$ and define $a_0 = 0$. Second, we assume for convenience of discussion that the visit times a_j are prespecified by design, but the methods also apply to cases where a_j is specified according to information gathered at visits $a_0, a_1, \ldots, a_{j-1}$. In general, individuals may have different administrative censoring times and so the numbers of potential visits may be different. Third, we assume that visit times are the same across individuals; this is common in many longitudinal studies, but most of the methods here extend easily to cases where visit times may vary across individuals. We discuss random visit times in Section 6.1 and Section 7. Loss to follow-up is assumed confirmed when an individual fails to appear for a specified visit.

In planned studies or surveys, a baseline visit will take place at $a_0 = 0$ and R subsequent visit times at $a_1, a_2 \ldots, a_R$ may be set for an individual; individuals, however, may become LTF before a_R. In such cases an exact LTF time is usually unobserved and the notion of such a time may be ambiguous since all that is known is that the person dropped out of the study between two (potential) visit times. In some settings, for example, an individual may decide not to attend the visit at time a_j just an instant before a_j. We describe a framework for considering a discrete LTF process in Appendix C and note that the data generated and types of likelihood functions presented there correspond to those we construct with our continuous-time LTF processes.

For this part of the discussion we consider the model in Figure 1(b) for generality. We reiterate that in this framework if a person is in state 1 at time a_{j-1} but LTF at time a_j then we do not know whether they made transition to LTF $1 \to 1^p$ or $1 \to 2 \to 2^p$ between a_{j-1} and a_j. In this setting, LTF is dependent and non-ignorable if $\lambda^p(t) \neq \lambda(t)$ or if $\alpha_1(t) \neq \alpha_2(t)$; in the latter case LTF may depend on an unobserved failure since the preceding visit. In the language of missing data, observations at a_j are not missing at random in this setting as missingness depends not just on observed data, but on missing data after time a_{j-1}.

To accommodate LTF we introduce a random variable M, which denotes the subscript labeling the visit time they were last observed to be in state 1; lower case m denotes its realized value. Let $\Delta_j = I(\text{individual is seen at } a_j)$ and let D_j denote the information obtained at a_j when $\Delta_j = 1$. For now we ignore covariates and suppose D_j just involves failure information. We now consider three observation schemes that might be used for a given individual.

Scheme I: Observation ends as soon as either failure or LTF is recorded.

Scheme II: Observation continues until LTF.

Scheme III: Observation continues until time a_R or, for some individuals, to the first visit at which entry to state 2^p is recorded.

Scheme I is standard practice in failure time studies, and Schemes II and III involve collecting additional information after failure or loss to follow-up. Scheme II is common when time to event analyses are based on data from a prospective study where interest may lie in multiple events. Scheme III would occur when some individuals are traced following LTF to obtain extended follow-up data. We assume here for simplicity that this begins immediately after LTF is recorded, but the discussion is readily extended to cases where, for example, persons LTF are contacted only at time a_R, and their vital status is recorded then. We focus on Schemes I and II in the main part of this article; Scheme III is discussed briefly in Section 7 and Appendix B.

2.3 Likelihood Construction for Joint Models

We now consider likelihood construction under observation Schemes I and II, and then discuss estimability of parameters in the joint model of Figure 1(a).

2.3.1 Likelihood Construction for Observation Scheme I

We let $Z(t)$ denote the state occupied by an individual in Figure 1 at time $t \geq 0$, and assume $Z(0) = 1$. It is easy to write down the types of observable histories under each of Schemes I and II, along with their probabilities. We initially discuss the data and corresponding probabilities for a generic individual with potential visits at specified times $a_1 < \cdots < a_R$ and let a_m denote the time they are last observed in state 1; if $m < R$ then at time a_{m+1} the individual is in either state 2 or 1^p or 2^p in Figure 1(b). We consider two scenarios for the type of data collected:

Case A: We observe only whether or not failure occurred over an interval $(a_{j-1}, a_j]$, and

Case B: If failure occurs in $(a_{j-1}, a_j]$, we can ascertain the exact failure time $T = t$.

We assume that exact LTF times may be conceptualized but are not observable. In what follows we write $P_{kl}(a_{j-1}, a_j)$ for $P(Z(a_j) = l | Z(a_{j-1}) = k;\ Z(a_{j-2}), \ldots, Z(a_0))$ for convenience where $a_0 = 0$.

There are three distinct observation types for Case A in Scheme I which, since they are defined by a sequence of observations, we refer to as paths; they are portrayed in Figure 2(a). For a generic individual with $a_R < \min(T, C)$, $m = R$ and path 1 defined by $Z(a_1) = \cdots = Z(a_R) = 1$ occurs with probability $P_{11}(0, a_R)$. For path 2, $a_m < T < a_{m+1} < C$ for some $m < R$ so they are known to have failed, and $Z(a_1) = \cdots = Z(a_m) = 1, Z(a_{m+1}) = 2$; this occurs with probability $P_{11}(0, a_m)\, P_{12}(a_m, a_{m+1})$. Finally, path 3 occurs when $a_m < T$ and $a_m < C < a_{m+1}$ for some $m < R$ so that $Z(a_1) = \cdots = Z(a_m) = 1$, $Z(a_{m+1}) = 1^p$ or 2^p; this occurs with probability $P_{11}(0, a_m)\{P_{11^p}(a_m, a_{m+1}) + P_{12^p}(a_m, a_{m+1})\}$. In what follows we assume that $\lambda^p(t) = \lambda(t)$ in Figure 1(b), in which case it is sufficient to consider the reduced model in Figure 1(a) and for simplicity we use it in describing types of outcomes and estimation procedures.

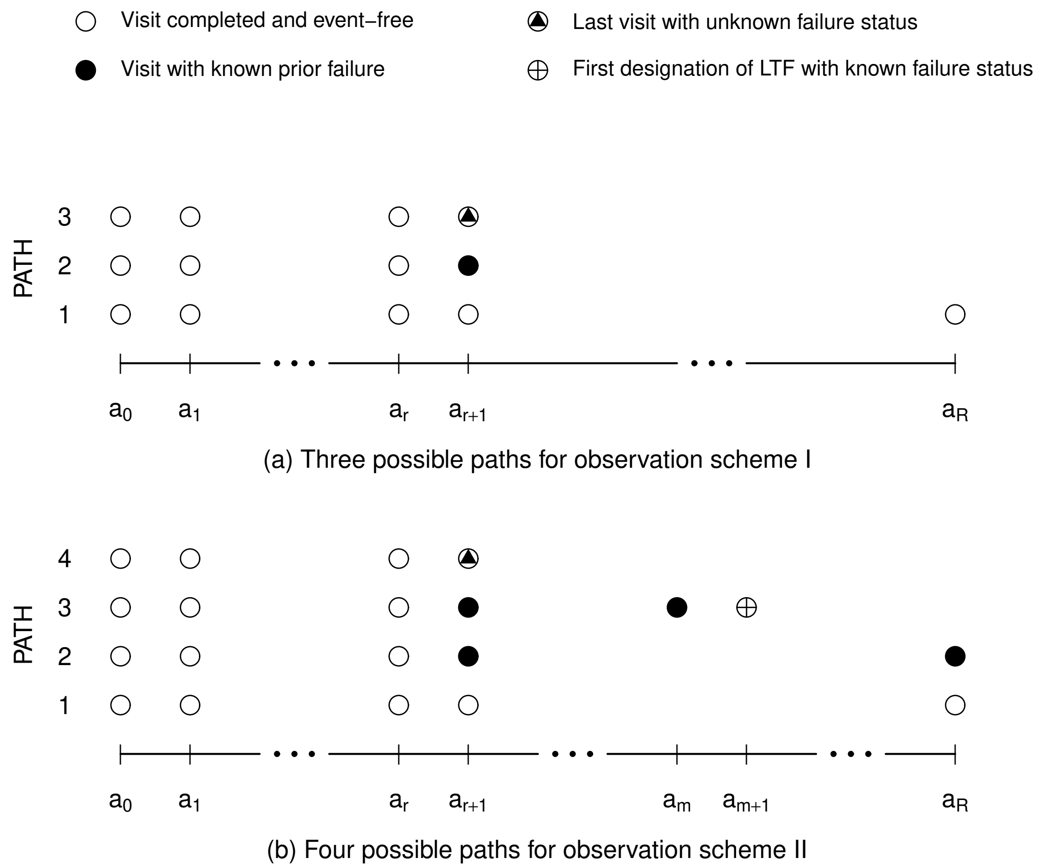


Figure 2: Schematic of the distinct possible paths under (a) Scheme I and (b) Scheme II.

We next add a subscript i to identify terms from individual i and consider likelihood construction based on a sample of n independent individuals, $i = 1, \ldots, n$. If individuals are serially accrued to a study they may have different administrative censoring times A_i and potential numbers of follow-up visits R_i, $i = 1, \ldots, n$. For simplicity, however, we assume that the potential visits are scheduled in study time, so that $0 = a_0 < a_1 < \cdots$ are the same across individuals. As in the preceding discussion, a_{m_i} ($a_{m_i} \leq a_{R_i} \leq A_i$) denotes the last scheduled assessment time individual i is observed in state 1. We let δ_{ij} indicate whether the observed path for individual i is of type j, $j = 1, 2, 3$. When considering the three possible paths the likelihood under Scheme I, Case A observation is

$$L_A^I(\psi) = \prod_{i=1}^{n} P_{11}(0, a_{R_i})^{\delta_{i1}} \; [P_{11}(0, a_{m_i})\, P_{12}(a_{m_i}, a_{m_i+1})]^{\delta_{i2}} \; [P_{11}(0, a_{m_i})\, P_{13}(a_{m_i}, a_{m_i+1})]^{\delta_{i3}} \; , \quad (1)$$

where ψ parameterizes the joint process. Covariates are left out of the notation for now, but if the transition intensities involve covariates, the same likelihood functions apply, with transition probabilities depending on the covariate values for each individual.

For settings in which the failure time can be retrospectively determined (i.e. Case B) the only modification is the contribution for path 2, in which case $Z_i(a_{m_i+1}) = 2$ is replaced by $Z_i(a_{m_i+1}) = 2$, $T_i = t_i$ and the probability $P_{12}(a_{m_i}, a_{m_i+1})$ in (1) is replaced by

$$P_{11}(a_{m_i}, t_i^-)\, \lambda(t_i)\, P_{22}(t_i, a_{m_i+1})\, .$$

We thus obtain the likelihood for Scheme I, Case B as

$$L_B^I(\psi) = \prod_{i=1}^n P_{11}(0, a_{R_i})^{\delta_{i1}} \left[P_{11}(0, t_i^-)\, \lambda(t_i)\, P_{22}(t_i, a_{m_i+1})\right]^{\delta_{i2}} \left[P_{11}(0, a_{m_i})\, P_{13}(a_{m_i}, a_{m_i+1})\right]^{\delta_{i3}} .$$

Note that we can write

$$L_B^I(\psi) = L_A^I(\psi) L_C^I(\psi) \tag{2}$$

where

$$L_C^I(\psi) = \prod_{i=1}^n \left[\frac{P_{11}(0, t_i^-)\, \lambda(t_i)\, P_{22}(t_i; a_{m_i+1})}{P_{11}(0, a_{m_i})\, P_{12}(a_{m_i}, a_{m_i+1})}\right]^{\delta_{i2}} .$$

If we let $J_A^I(\psi) = -\partial^2 \log L_A^I(\psi)/\partial\psi\partial\psi'$ and $J_C^I(\psi) = -\partial^2 \log L_C^I(\psi)/\partial\psi\partial\psi'$ then

$$J_B^I(\psi) = -\partial^2 \log L_A^I(\psi)/\partial\psi\partial\psi' = J_A^I(\psi) + J_C^I(\psi) \tag{3}$$

where $J_C^I(\psi)$ is the information added by the retrospective collection of the data on the event time for individuals known to have failed under Case B.

2.3.2 Likelihood Construction for Observation Scheme II

For Scheme II observation individuals are followed until LTF and so there are four possible paths for Case A observation as shown in Figure 2(b). As before, for a generic individual path 1 has $a_R < \min(T, C)$ giving $m = R$ and $Z(a_1) = \cdots = Z(a_R) = 1$, with probability $P_{11}(0, a_R)$. For path 2 failure is known to have occurred (i.e. $a_m < T < a_{m+1}$) for some $m < R$ and there is no subsequent LTF (i.e. $a_R < C$), so $Z(a_1) = \cdots = Z(a_m) = 1$ and $Z(a_{m+1}) = \cdots = Z(a_R) = 2$ for $m < R$. This occurs with probability $P_{11}(0, a_m)\, P_{12}(a_m, a_{m+1})\, P_{22}(a_{m+1}, a_R)$. If failure is known to have occurred at some point prior to LTF then $a_m < T < a_{m+1}$ and $a_r < C < a_{r+1}$ for some $m < r < M$ represented by path 3 with $Z(a_1) = \cdots = Z(a_m) = 1$, $Z(a_{m+1}) = \cdots = Z(a_r) = 2$ and $Z(a_{r+1}) = 2^p$ which occurs with probability

$$P_{11}(0, a_m)\, P_{12}(a_m, a_{m+1})\, P_{22}(a_{m+1}, a_r)\, P_{23}(a_r, a_{r+1}) .$$

Finally if $a_m < T$ and $a_m < C < a_{m+1}$ for some $m < R$ then path 4 is realized, with $Z(a_1) = \cdots = Z(a_m) = 1$ and $Z(a_{m+1}) = 1^p$ or 2^p since the failure status is unknown at a_{m+1}; this occurs with probability $P_{11}(0, a_m)\{P_{11^p}(a_m, a_{m+1}) + P_{12^p}(a_m, a_{m+1})\}$.

Reintroducing the subscript i to distinguish individuals the likelihood for a sample of n independent individuals under Scheme II, Case A is then

$$\begin{aligned} L_A^{II}(\psi) = \prod_{i=1}^n & P_{11}(0, a_{R_i})^{\delta_{i1}} \left[P_{11}(0, a_{m_i})\, P_{12}(a_{m_i}, a_{m_i+1})\, P_{22}(a_{m_i+1}, a_{R_i})\right]^{\delta_{i2}} \\ & \times \left[P_{11}(0, a_{m_i})\, P_{12}(a_{m_i}, a_{m_i+1})\, P_{22}(a_{m_i+1}, a_{r_i})\, P_{23}(a_{r_i}, a_{r_i+1})\right]^{\delta_{i3}} \\ & \times \left[P_{11}(0, a_{m_i})\, P_{13}(a_{m_i}, a_{m_i+1}))\right]^{\delta_{i4}} , \end{aligned} \tag{4}$$

where δ_{ij} indicates a path of type j, $j = 1, 2, 3, 4$. In Case B the likelihood contributions change for paths 2 and 3 with the additional recording of $T_i = t_i$ for both. In particular the contributions $P_{12}(a_{m_i}, a_{m_i+1})$ to (4) are replaced by $P_{11}(a_{m_i}, t_i^-)\lambda(t_i)P_{22}(t_i, a_{m_i+1})$ giving

$$\begin{aligned} L_B^{II}(\psi) = \prod_{i=1}^n & P_{11}(0, a_{R_i})^{\delta_{i1}} \left[P_{11}(0, t_i^-)\, \lambda(t_i)\, P_{22}(t_i, a_{R_i})\right]^{\delta_{i2}} \\ & \times \left[P_{11}(0, t_i^-)\, \lambda(t_i)\, P_{22}(t_i, a_{r_i})\, P_{23}(a_{r_i}, a_{r_i+1})\right]^{\delta_{i3}} \\ & \times \left[P_{11}(0, a_{m_i})\, P_{13}(a_{m_i}, a_{m_i+1})\right]^{\delta_{i4}} . \end{aligned} \tag{5}$$

As in Scheme I, we can write $L_B^{II}(\psi) = L_A^{II}(\psi) L_C^{II}(\psi)$ here, where

$$L_C^{II}(\psi) = \prod_{i=1}^{n} \left[\frac{P_{11}(0,t_i^-)\,\lambda(t_i)\,P_{22}(t_i;a_{m_i})}{P_{11}(0,a_{m_i})\,P_{12}(a_{m_i},a_{m_i+1})\,P_{22}(a_{m_i+1},a_{M_i})} \right]^{\delta_{i2}} \times \left[\frac{P_{11}(0,t_i^-)\,\lambda(t_i)\,P_{22}(t_i,a_{r_i})\,P_{23}(a_{r_i},a_{r_i+1})}{P_{11}(0,a_{m_i})\,P_{12}(a_{m_i},a_{m_i+1})\,P_{22}(a_{m_i+1},a_{r_i})\,P_{23}(a_{r_i},a_{r_i+1})} \right]^{\delta_{i3}} .$$

2.3.3 Remarks on Identifiability and Estimability

Estimability of the intensities in Figure 1 is an important concern. For observation Scheme I in which individuals are followed to the first visit where failure or LTF is observed, estimation of intensities in Figure 1(b) essentially requires the constraints $\alpha_1(t) = \alpha_2(t)$ and $\lambda^p(t) = \lambda(t)$. Lawless (2013) considers the case where $\lambda^p(t) = \lambda(t)$ but $\alpha_1(t) \neq \alpha_2(t)$ and describes the bias in estimation of $\lambda(t)$ that results from treating $Z(a_{j-1})$ as the final observation in a standard analysis when LTF is recorded at a_j. The proper approach in such cases is based on conditional probabilities $P(D_j|Z(a_{j-1}) = 1, Z(a_j) \in \{1,2\})$ or likelihoods (1) and (2) that can be considered with the joint models of Figures 1(a), (b). However estimation of all the parameters in Figure 1(a) is problematic. We show below that when visits are equi-spaced ($\Delta a_j = a_j - a_{j-1} = \Delta$), parameters are non-identifiable for Scheme I, Case A. For Case B, and for both cases when Δa_j values vary, parameters appear identifiable but very large samples are needed to estimate α_2 precisely. We discuss this and provide some numerical results in Section 3.3.

To demonstrate the non-identifiability issue for Scheme I, Case A we consider the time-homogeneous model for Figure 1(a) with $\lambda(t;\eta) = \eta$, $\alpha_1(t;\theta_1) = \theta_1$ and $\alpha_2(t;\theta_2) = \theta_2$; we let $\psi = (\eta', \theta')'$. We focus on the setting where $\Delta a_j = \Delta$ for $j = 1, 2, \ldots$. In this time time-homogeneous setting we let $P_{kl}(t, t+\Delta) = P_{kl}(\Delta)$ for $\Delta > 0$, and we see that the likelihood $L_A^I(\psi)$ in (1) reduces to

$$\begin{aligned} L_A^I(\psi) &= \prod_{i=1}^{n} \left[P_{11}(\Delta)^{R_i}\right]^{\delta_{i1}} \left[P_{11}(\Delta)^{m_i}\,P_{12}(\Delta)\right]^{\delta_{i2}} \left[P_{11}(\Delta)^{m_i}\,P_{13}(\Delta)\right]^{\delta_{i3}} \\ &= P_{11}(\Delta;\psi)^{n_1}\,P_{12}(\Delta;\psi)^{n_2}\,P_{13}(\Delta;\psi)^{n_3} , \end{aligned} \tag{6}$$

where $n_1 = \sum_{i=1}^n \delta_{i1} R_i + \delta_{i2} m_i + \delta_{i3} m_i$, $n_2 = \sum_{i=1}^n \delta_{i2}$ and $n_3 = \sum_{i=1}^n \delta_{i3}$. Since $P_{11}(\Delta) + P_{12}(\Delta) + P_{13}(\Delta) = 1$, we see that $L_A^I(\psi)$ involves only two parametric functions of ψ, say $P_{11}(\Delta;\psi)$ and $P_{12}(\Delta;\psi)$. As there are three functionally independent parameters in $\psi = (\eta, \theta_1, \theta_2)$ the parameter vector is non-identifiable based on (6); that is, for any data set there will be an infinite set of ψ-vectors maximizing $L_A^I(\psi)$. Consequently we need data from Scheme II, or external information about $\alpha_2(t)$ versus $\alpha_1(t)$ for estimation. The same results hold for other parametric models.

We remark that if $\alpha_1(t) = \alpha_2(t)$ in Figure 1(a), corresponding to independent LTF, then the likelihood in (1) factors into a term involving θ $(= \theta_1 = \theta_2)$ only and a term involving η only. A unique maximum likelihood estimate $\widehat{\psi}$ exists in this case. For $\alpha_1(t;\theta_1) = \alpha_2(t;\theta_2) = \theta$ in the time-homogeneous case above it is easily seen that this occurs at

$$\tilde{\eta} = \Delta^{-1}\,\log(1 + \tilde{P}_{12}/\tilde{P}_{11})\,, \quad \tilde{\theta} = -\Delta^{-1}\,\log(\tilde{P}_{11} + \tilde{P}_{12})\,,$$

where $\tilde{P}_{11} = n_1/n$ and $\tilde{P}_{12} = n_2/n$. More details are given in Sections 3.1 and 4.1.

For Case B and equi-spaced visits and for both Cases A and B for unequally spaced visits, all parameters η, θ_1, θ_2 in the time-homogeneous joint model appear estimable. We have not proven consistency, but extensive simulations show that with large enough samples estimators have small bias, and coverages for confidence intervals based on normal approximations for $\log\widehat{\eta}$, $\log\widehat{\theta}_1$, $\log\widehat{\theta}_2$ are close to nominal levels. For smaller sample sizes, and depending on the proportions of individuals failed and prematurely lost to follow-up, θ_2 is less well estimated; $\widehat{\theta}_2$ can equal zero and its distribution can have a long left tail. In regression models with $\lambda(t|X) = \lambda e^{X\beta}$ and X binary, β was well estimated. The main limitation of observation Scheme I, however, is that information about parameters in more complex models is limited unless the Δa_j are small. This makes adequate model checking and specification challenging, and we therefore caution against an over-reliance on the models in Figure 1(a) unless auxiliary information about loss to follow-up or data obtained under observation Scheme II are available.

For Scheme II where individuals are followed after failure, under $\lambda^p(t) = \lambda(t)$ we can fit parametric models $\lambda(t;\eta)$, $\alpha_1(t;\theta_1)$ and $\alpha_2(t;\theta_2)$. There may be limited information about certain parameters, however, if visits are far apart, if particular paths occur infrequently in a given sample, or if the sample size is small.

In the next section we examine bias in standard failure time analysis when $\alpha_1(t) \neq \alpha_2(t)$. Section 3.3 gives numerical results on estimability for the model in Figure 1(a) under observation Scheme II.

3 Standard Analysis and Biases from Dependent LTF

3.1 Partial Likelihoods under Independent LTF

Under the assumption that the LTF time is independent of T (i.e. $\alpha_1(t) = \alpha_2(t) = \alpha(t)$ in Figure 1(a)) then the likelihood contributions in (1) or (2), which are expressed in terms of the joint model, can be simplified. Let $\mathcal{F}(t;\eta) = P(T \geq t;\eta)$ and $\mathcal{G}(t;\theta) = P(C \geq t;\theta)$ denote the marginal survivor functions for the failure and LTF processes. Then under the assumption of independent LTF, $P_{11}(0,t;\psi) = \mathcal{F}(t;\eta)\,\mathcal{G}(t;\theta)$, $P_{12}(s,t;\psi) = [1-\mathcal{F}(t;\eta)/\mathcal{F}(s;\eta)]\,\mathcal{G}(t;\theta)/\mathcal{G}(s;\theta)$, and $P_{13}(s,t;\psi) = 1-\mathcal{G}(t;\theta)/\mathcal{G}(s;\theta)$. In this case, (1) can be written as $L_A^I(\psi) = L_{A1}^I(\eta)\,L_{A2}^I(\theta)$ where

$$L_{A1}^I(\eta) = \prod_{i=1}^n \mathcal{F}(a_{m_i^\dagger};\eta)^{1-\delta_{i2}}\,[\mathcal{F}(a_{m_i}) - \mathcal{F}(a_{m_i+1})]^{\delta_{i2}} \tag{7}$$

with $m_i^\dagger = \delta_{i1}\,R_i + (1-\delta_{i1})\,m_i$, $i = 1,\dots,n$. Under the assumption of non-informative LTF we would typically ignore

$$L_{A2}^I(\theta) \propto \prod_{i=1}^n \mathcal{G}(a_{m_i^\ddagger};\theta)^{1-\delta_{i3}}\,[\mathcal{G}(a_{m_i}) - \mathcal{G}(a_{m_i+1})]^{\delta_{i3}}\ , \tag{8}$$

where $m_i^\ddagger = \delta_{i1}\,R_i + \delta_{i2}\,(m_i+1)$, and focus on the partial likelihood $L_{A1}^I(\eta)$. A similar simplification for Case B gives the partial likelihood for η as

$$L_{B1}^I(\eta) = \prod_{i=1}^n P_{11}(0,a_{m_i^\star})^{1-\delta_{i2}}\,\left[P_{11}(0,t_i^-)\lambda(t_i)\right]^{\delta_{i2}} = \prod_{i=1}^n \lambda(v_i)^{\delta_i}\,\mathcal{F}(v_i)\,, \tag{9}$$

where $m_i^\star = \delta_{i1}R_i + \delta_{i3}m_i$, $v_i = \min(a_{m_i^\star}, t_i)$, and $\delta_i = \delta_{i2}$, $i = 1,\dots,n$.

The resulting expressions in (7) and (9) are of the conventional form for interval-censored and right-censored failure time data, respectively. Note that if failure is not recorded to have occurred in both (7) and (9) the censoring time is the last visit they were known to be failure-free (in state 1). This confirms the appropriateness of adopting this as the censoring time rather than some other value in $[a_{m_i}, a_{m_i+1}]$ as has sometimes been suggested. We now use the joint model to derive the asymptotic bias of estimators based on (7) or (9) when dependent LTF occurs.

3.2 Bias in Parametric Models Assuming Independent LTF

We now consider models involving fixed covariates and write $\lambda(t|x_i;\eta)$ for the failure intensity given x_i, $i=1,\ldots,n$. The score functions based on (7) and (9) are denoted by $U^I_{A1}(\eta)=\partial \log L^I_{A1}(\eta)/\partial\eta$ and $U^I_{B1}(\eta)=\partial \log L^I_{B1}(\eta)/\partial\eta$, respectively. If the joint model in Figure 1(a) represents the true process, then under mild conditions the estimators $\widehat{\eta}_A$ and $\widehat{\eta}_B$ obtained by solving $U^I_{A1}(\eta)=0$ or $U^I_{B1}(\eta)=0$ converge in probability as $n\to\infty$ to vectors $\eta^\star_A$ and $\eta^\star_B$ obtained as solutions to $E\{U^I_{A1}(\eta^\star_A)\}=0$ and $E\{U^I_{B1}(\eta^\star_B)\}=0$, respectively (White, 1982). We assume that $\{R_i, X_i, Z_i(t), t>0\}$ are i.i.d. across individuals and use the corresponding probability model for the required expectations. It is feasible in some cases to solve these equations; in other situations simulation can be used to examine the extent to which $\eta^\star_A$ and $\eta^\star_B$ differ from the true value η_0.

We consider Case A for illustration and note that

$$U^I_{A1}(\eta)=\sum_{i=1}^{n}\left\{-(1-\delta_i)\,\Lambda^\eta(a_{m_i^\dagger};\eta)+\delta_i\,B^\eta(a_{m_i};\eta)\right\}, \tag{10}$$

where $\Lambda(t;\eta)=-\log\mathcal{F}(t;\eta)$ is the cumulative hazard for failure and we let $\delta_i=\delta_{i2}$, $\Lambda^\eta(t;\eta)=\partial\Lambda(t;\eta)/\partial\eta$, and

$$B^\eta(a_m;\eta)=\partial\log(\exp(-\Lambda(a_m;\eta))-\exp(-\Lambda(a_{m+1};\eta)))/\partial\eta\,.$$

The expectation is to be taken with respect to the joint model for the failure and LTF process under the intermittent observation scheme and for this it is helpful to re-express the contribution to (10) in terms of the joint multistate process. The score vector in (10) can be written as $U^I_{A1}(\eta)=\sum_{i=1}^n U^I_{iA1}(\eta)$ where

$$\begin{aligned}U^I_{iA1}(\eta)=&-I(Z_i(a_{R_i})=1)\,\Lambda^\eta(a_{R_i};\eta)-\sum_{m=0}^{R_i-1}I(Z_i(a_m)=1,Z_i(a_{m+1})=3)\,\Lambda^\eta(a_m;\eta)\\&+\sum_{m=0}^{R_i-1}I(Z_i(a_m)=1,Z_i(a_{m+1})=2)\,B^\eta(a_m;\eta)\,.\end{aligned}\tag{11}$$

We consider the case where covariates X may be present and write $\Lambda^\eta(a|X;\eta)$ and $B^\eta(a|X;\eta)$ to indicate this. Dropping the subscript i we note that the expectation of an individual contribution to (11) under the joint model in Figure 1(a) is then

$$\begin{aligned}E\{U^I_{A1}(\eta^\star)\}=&-E_{RX}\{P(Z(a_R)=1\mid X)\,\Lambda^\eta(a_R|X;\eta)\\&-\sum_{m=0}^{R-1}P(Z(a_m)=1,Z(a_{m+1})=3\mid X)\,\Lambda^\eta(a_m|X;\eta)\}\\&+E_{RX}\left\{\sum_{m=0}^{R-1}P(Z(a_m)=1,Z(a_{m+1})=2\mid X)\,B^\eta(a_m|X;\eta)\right\},\end{aligned}\tag{12}$$

under the assumption that the processes $\{R_i, X_i, Z_i(s), 0<s\}$ are independent and identically distributed.

We consider a particular setting where R is fixed to illustrate the impact of a dependent LTF process. We consider a proportional hazards failure time process with $\lambda(t|X)=\lambda_0(t)e^{X\beta}$, where $\lambda_0(t)$ is a baseline hazard and X is a Bernoulli random variable with $P(X=1)=0.5$. The baseline hazard function is taken to be piecewise-constant with break-points $0=b_0<b_1<\cdots<b_K$ and $\lambda_0(t)=\rho_k$ if $t\in[b_{k-1},b_k)$, $k=1,\ldots,K$. We consider the process over the period $[0,A]$ where $A=1$ is a fixed and common administrative censoring time, and set the break-points as $b_k=k/K$,

$k = 0, 1, \ldots, K$. We let $\rho_k = \rho_{k-1}\, e^{\zeta}$, $k = 2, \ldots, K$ where $e^{\zeta} = 1.1$ so there is a 10% relative increase in the risk of failure for each successive interval. We let $\rho = (\rho_1, \ldots, \rho_K)'$ so that $\eta = (\rho', \beta)'$ parameterizes the failure time process. The probability of failure by time $A = 1$ is denoted by π_F and given by

$$\pi_F = F(1) = E_X\{F(1|X)\} = 1 - \sum_{x=0}^{1} \exp\left(-\Lambda_0(1)\, e^{\beta x}\right)\, 0.5^{(1-x)}\, 0.5^{x}\ , \tag{13}$$

where $\Lambda_0(t) = \int_0^t \lambda_0(s)\, ds$. Under Scheme I we follow individuals to the first visit with evidence of failure or the last visit before LTF. Under the joint model in Figure 1(a), let π_F° denote the probability an individual is observed to fail by $A = a_R = 1$; this is given by computing

$$\begin{aligned} \pi_F^{\circ}(X) &= P(Z(a_1) = 2 \mid Z(a_0) = 1; X) \\ &\quad + \sum_{j=2}^{R} \left[\prod_{s=1}^{j-1} P(Z(a_s) = 1 \mid Z(a_{s-1}) = 1; X)\right] P(Z(a_j) = 2 \mid Z(a_{j-1}) = 1; X)\,, \end{aligned} \tag{14}$$

and averaging over X.

We next discuss how we set the particular parameter values. For a specified K, $e^{\zeta} = 1.1$ and $e^{\beta} = 0.75$, we solve for ρ_1 such that π_F equals 0.8 (i.e. the probability of failure by $A = 1$ is 0.80). For the visit schedule we set $a_j = j/R$, $j = 0, 1, \ldots, R$. For simplicity we set $K = R$ and let $a_j = b_j$, $j = 1, \ldots, K = R$ for $K = R = 5$ or 10. For the LTF process we let $\alpha_k(t) = \alpha_k$, set $\alpha_2/\alpha_1 = 1$ (corresponding to independent LTF), 2 or 5, and for each of these values we solve for α_1 in order to achieve a desired value of π_F° using (14). Then $\theta = (\alpha_1, \alpha_2)'$ and $\psi = (\eta', \theta')'$. In the top two panels of Figure 3 we plot the asymptotic percent relative bias of the naive estimator for β as a function of π_F°/π_F, the probability that failure is observed given that it occurred before $A = 1$. We consider the setting with $K = R = 5$ in panel (a) and $K = R = 10$ in panel (b). Separate lines are given for the cases where $\alpha_2/\alpha_1 = 1$ where no asymptotic bias is anticipated since LTF is independent, $\alpha_2/\alpha_1 = 2$ for moderately dependent LTF, and $\alpha_2/\alpha_1 = 5$ for more strongly dependent LTF. We find as expected that the asymptotic percent relative bias for the regression coefficient is zero under independent LTF and small when the rate of censoring is low. The bias can be more substantial under heavier censoring reflected by lower values of π_F°/π_F , but it does not exceed 4.5% for the scenarios examined even when $\alpha_2/\alpha_1 = 5$ in panel (a). With more frequent assessments and more flexible baseline hazards (i.e. when $K = R = 10$ as in panel (b)) the asymptotic bias is much lower.

In the bottom two panels of Figure 3 we plot the asymptotic percent relative bias for estimators corresponding to different levels of random censoring (i.e. with $\pi_F^{\circ}/\pi_F = 0.2$, 0.4, 0.6 and 0.8) for the baseline survival distribution $\mathcal{F}(t|X = 0)$ over $t \in [0, 1]$ under the strongly dependent LTF (i.e. $\alpha_2/\alpha_1 = 5$). Here the percent relative bias can be large with $K = R = 5$, particularly for lower values of π_F°/π_F and as t approaches the administrative censoring time. With $K = R = 10$ the bias remains large but is much smaller (panel (d)).

In summary there is a small relative bias in the relative risk parameters (covariate effects) but much larger bias in the absolute survival probabilities; this is a phenomenon often seen in scenarios involving dependent censoring (Cook et al., 2003). We caution, however, that these results hold for a proportional hazards failure time process and other types of processes would need to be investigated on a case-by-case basis.

3.3 Empirical Studies and Numerical Results

In Table 1 we report on the results of simulation studies to assess biases of naive (standard) methods of analysis that assume independent LTF under observation Scheme I, Case A, as well as the performance of estimators based on the joint model for Scheme II under Cases A and B. Additional data

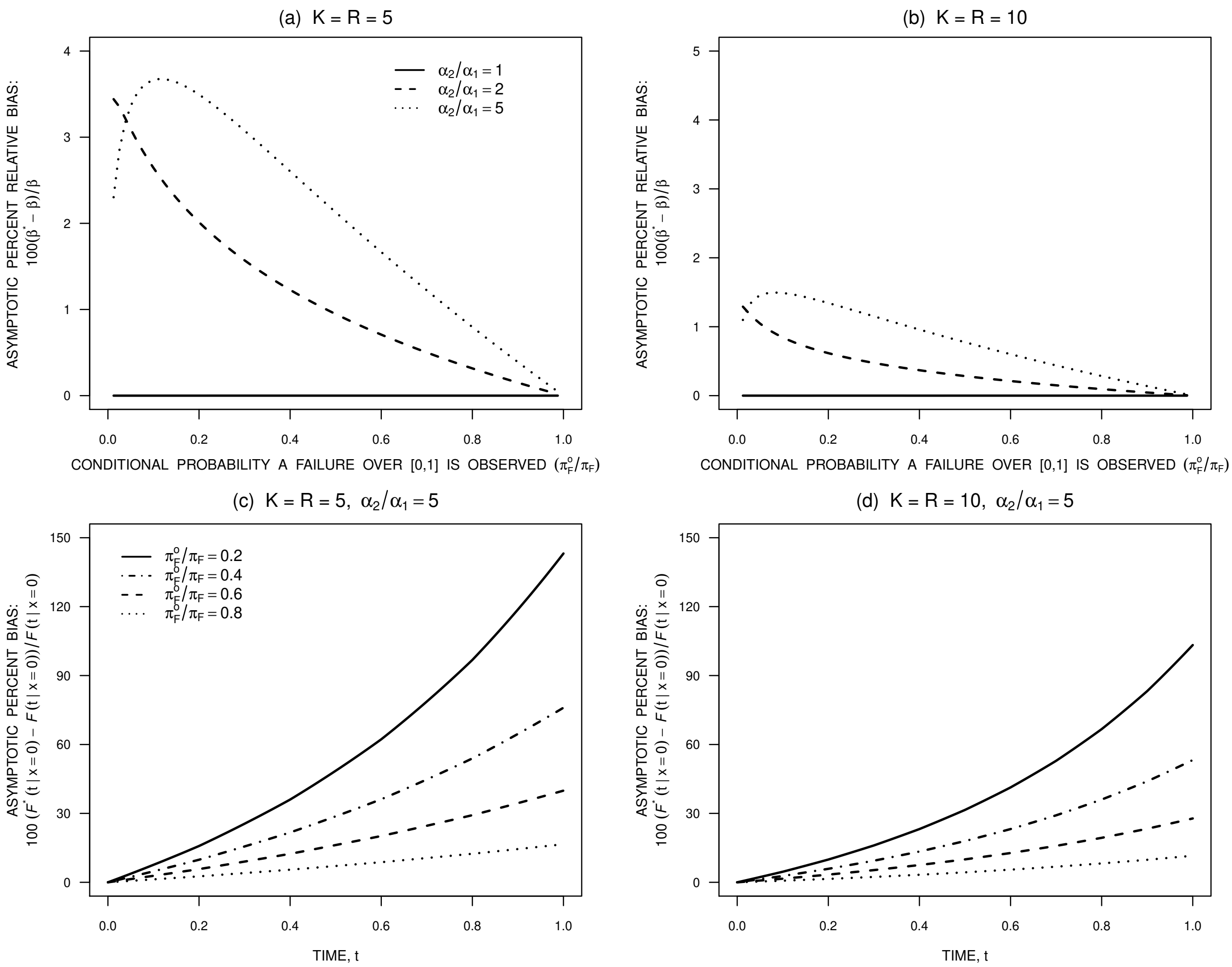


Figure 3: Asymptotic percent relative bias $100 \times (\beta^* - \beta)/\beta$ versus π_F°/π_F for $K = 5$ (panel (a)) and $K = 10$ (panel (b)) and asymptotic percent relative bias $100 \times (\mathcal{F}^*(t|x = 0) - \mathcal{F}(t|x = 0))/\mathcal{F}(t|x = 0)$ of baseline survivor function with $K = 5$, $\alpha_2/\alpha_1 = 5$ (panel (c)) and $K = 10$, $\alpha_2/\alpha_1 = 5$ (panel (d)) under standard parametric analysis assuming independent LTF; for all figures $\pi_F = 0.80$, $e^\zeta = 1.1$, $e^\beta = 0.75$ and the assumed model is $\lambda(t|x) = \lambda_0(t)\, e^{X\beta}$ with $\lambda_0(t)$ piecewise constant baseline hazard with break-points at $b_k = a_k = k/K$, $k = 1, \ldots, K$ with $K = 5$ or 10.

under the more informative Scheme II enables fitting of the joint model to mitigate the bias from dependent LTF.

We consider the same parametric setting as in Section 3.2, but set $K = 1$ so the fitted model is a proportional hazards model with a constant baseline hazard. Here $K \neq R$ as visits are scheduled at $a_j = j/R$, $j = 0, 1, \ldots, R$ with $R = 5$, and we consider $\alpha_2/\alpha_1 = 2$ and 5 to represent moderate and more severe dependent LTF. The failure time model is an exponential model with $\lambda(t|x) = \lambda \exp(x\beta)$ with $\beta = \log 0.75$. Table 1 shows the empirical bias (BIAS) and empirical standard error (ESE), the average standard error based on the observed information (ASE) and the empirical coverage probability of nominally 95% confidence intervals for the case where $\pi_F = 0.8$. The sample size is set to $n = 1000$ and the results are based on $nsim = 2000$ simulated samples.

The first set of columns in Table 1 is from maximizing the naive likelihood $L_{A1}^I(\eta)$ in (7) based on the assumption of independent LTF. We see large bias for the estimator of λ but small bias for the estimator of β; even when censoring is heavy (as reflected by small π_F°/π_F) the empirical coverage probability for β is close to the nominal level. The magnitude of the bias is close to that expected from the results in Figure 3. The second and third set of columns give results from the joint analysis

Table 1: Empirical performance of estimators from naive and joint analyses of a proportional hazards model with an exponential (i.e. time-homogeneous) baseline hazard; visit times are equi-spaced with $a_j = j/R$, $j = 0, 1, \ldots, R$, $\pi_F = 0.8$, $\beta = \log 0.75$, $n = 1000$, $nsim = 2000$.

		CASE A								CASE B			
		NAIVE, $L^I_{A1}(\eta)$ in (7)				JOINT, $L^{II}_A(\psi)$ in (4)				JOINT, $L^{II}_B(\psi)$ in (5)			
Parameter	Value	BIAS	ESE	ASE	ECP	BIAS	ESE	ASE	ECP	BIAS	ESE	ASE	ECP
LIGHT, MILDLY DEPENDENT CENSORING: $\pi^\circ_F = 0.6$, $\alpha_2/\alpha_1 = 2$													
λ	1.870	-0.079	0.101	0.101	86.0	0.004	0.106	0.106	95.0	0.004	0.106	0.105	95.0
β	-0.288	-0.001	0.082	0.082	94.5	-0.000	0.081	0.082	94.6	-0.000	0.081	0.081	94.8
α_1	0.512	-	-	-	-	0.000	0.043	0.042	95.5	0.000	0.043	0.042	95.5
α_2	1.025	-	-	-	-	0.002	0.064	0.065	95.0	0.002	0.064	0.065	95.3
LIGHT, STRONGLY DEPENDENT CENSORING: $\pi^\circ_F = 0.6$, $\alpha_2/\alpha_1 = 5$													
λ	1.870	-0.201	0.093	0.094	41.8	0.003	0.104	0.105	96.0	0.003	0.104	0.105	95.8
β	-0.288	-0.003	0.082	0.082	95.4	0.000	0.080	0.080	95.0	0.000	0.079	0.079	94.8
α_1	0.328	-	-	-	-	-0.001	0.040	0.040	94.8	-0.001	0.040	0.040	94.7
α_2	1.642	-	-	-	-	0.002	0.089	0.091	95.2	0.002	0.088	0.090	95.2
MODERATE, MILDLY DEPENDENT CENSORING: $\pi^\circ_F = 0.4$, $\alpha_2/\alpha_1 = 2$													
λ	1.870	-0.198	0.117	0.115	56.4	0.004	0.134	0.131	94.7	0.004	0.133	0.130	94.5
β	-0.288	-0.002	0.101	0.101	95.2	0.001	0.100	0.099	95.5	0.001	0.099	0.099	95.5
α_1	1.294	-	-	-	-	0.000	0.079	0.078	94.4	0.000	0.079	0.078	94.4
α_2	2.588	-	-	-	-	0.005	0.153	0.152	95.5	0.005	0.150	0.150	95.5
MODERATE, STRONGLY DEPENDENT CENSORING: $\pi^\circ_F = 0.4$, $\alpha_2/\alpha_1 = 5$													
λ	1.870	-0.457	0.100	0.097	1.0	0.005	0.134	0.130	94.3	0.005	0.133	0.129	94.5
β	-0.288	-0.007	0.104	0.101	94.5	-0.001	0.096	0.093	94.9	-0.001	0.096	0.093	94.9
α_1	0.818	-	-	-	-	-0.000	0.077	0.074	94.5	-0.000	0.076	0.074	94.6
α_2	4.091	-	-	-	-	0.007	0.234	0.231	94.8	0.006	0.224	0.223	95.1
HEAVY, MILDLY DEPENDENT CENSORING: $\pi^\circ_F = 0.2$, $\alpha_2/\alpha_1 = 2$													
λ	1.870	-0.396	0.140	0.142	22.7	0.020	0.196	0.195	95.3	0.019	0.193	0.193	95.5
β	-0.288	-0.008	0.142	0.143	95.3	-0.001	0.137	0.137	95.3	-0.001	0.136	0.137	95.5
α_1	2.811	-	-	-	-	-0.010	0.155	0.152	94.5	-0.009	0.151	0.148	95.0
α_2	5.622	-	-	-	-	0.054	0.445	0.435	95.0	0.047	0.415	0.409	95.0
HEAVY, STRONGLY DEPENDENT CENSORING: $\pi^\circ_F = 0.2$, $\alpha_2/\alpha_1 = 5$													
λ	1.870	-0.813	0.100	0.102	0.0	0.021	0.200	0.198	95.0	0.016	0.192	0.192	94.8
β	-0.288	-0.013	0.141	0.143	95.3	-0.001	0.123	0.122	94.8	-0.001	0.122	0.122	95.0
α_1	1.743	-	-	-	-	-0.013	0.154	0.150	94.6	-0.007	0.142	0.140	94.3
α_2	8.716	-	-	-	-	0.084	0.750	0.721	94.8	0.046	0.632	0.623	94.8

possible under the Scheme II observation process for Cases A and B, obtained by maximizing the likelihoods in (4) and (5), respectively; for these analyses the LTF intensities are estimated. Moreover the added value of retrospective ascertainment of the failure time in Case B is apparent by comparing the standard errors of the estimates from (5) versus (4). We see negligible empirical bias for all estimators, good agreement between the empirical standard errors and average model-based standard errors and the empirical coverage probabilities are all close to the nominal level. The added value of observing the failure time in Case B observation is very small for λ and β suggesting the term $J_c^I(\psi)$ in (3) conveys relatively little information on these parameters; gains are more substantial for the intensities α_1, α_2 for LTF at the heavier censoring rates. The very small gain in efficiency for exact versus interval-censored failure times is a phenomenon that is often observed for the analysis of constant failure rates.

4 Bias in Nonparametric and Semiparametric Analyses

4.1 Bias in Nonparametric Analyses

Nonparametric estimation of $\lambda(t)$ can be considered by discretizing time, letting $t = 1, 2, \ldots$ denote possible visit, failure and LTF times. The survivor and probability functions for T are in this case

$$\mathcal{F}(t) = P(T > t) = \prod_{s=1}^{t} (1 - \lambda(s)) \,, \quad f(t) = \mathcal{F}(t-1)\, \lambda(t) \,, \quad t = 1, 2, \ldots$$

where $\mathcal{F}(0) = 0$ and $\lambda(t) = P(T = t | T \geq t)$.

For data under Scheme I, Case A we once again consider nonparametric estimation when individuals have the same potential visit times. To do this we let $\eta_j = P(a_{j-1} < T \leq a_j | T > a_{j-1})$ for $j = 1, 2, \ldots$ In the absence of censoring $\eta_j = 1 - e^{-\Lambda(a_{j-1}, a_j)}$ where $\Lambda(a_{j-1}, a_j) = \Lambda(a_j) - \Lambda(a_{j-1})$ and with $\delta_i = \delta_{i2}$ we can rewrite (7) as

$$\begin{aligned} L_{A1}^I(\eta) &\propto \prod_{i=1}^{n} \left\{ \prod_{j=1}^{m_i^\dagger} (1 - \eta_j) \right\}^{1-\delta_i} \left\{ \eta_{m_i+1} \prod_{j=1}^{m_i} (1 - \eta_j) \right\}^{\delta_i} \\ &= \prod_{j=1}^{R} \prod_{i=1}^{n} \left\{ (1 - \eta_j)^{I(Z_i(a_j)=1)} \, \eta_j^{I(Z_i(a_j)=2)} \right\}^{I(Z_i(a_{j-1})=1, Z_i(a_j)<3)} . \end{aligned}$$

The score function for η_j is $\partial \log L_{A1}^I(\eta) / \partial \eta_j$, given by

$$\sum_{i=1}^{n} I(Z_i(a_{j-1}) = 1)\, I(Z_i(a_j) < 3) \left\{ \frac{I(Z_i(a_j) = 2)}{\eta_j} - \frac{I(Z_i(a_j) = 1)}{1 - \eta_j} \right\} \tag{15}$$

and equating this to zero gives the estimates

$$\widehat{\eta}_j = \frac{\sum_{i=1}^{n} I(Z_i(a_{j-1}) = 1, Z_i(a_j) = 2)}{\sum_{i=1}^{n} I(Z_i(a_{j-1}) = 1, Z_i(a_j) < 3)} \,. \tag{16}$$

Taking the expectation of $\partial \log L_{A1}^I(\eta) / \partial \eta_j$ with respect to the joint model or directly from (16), it can be seen that $\widehat{\eta}_j$ converges in probability to

$$\eta_j^\star = \frac{P_{12}(a_{j-1}, a_j)}{P_{11}(a_{j-1}, a_j) + P_{12}(a_{j-1}, a_j)} \,, \tag{17}$$

as $n \to \infty$, which equals $P(Z(a_j) = 2|Z(a_{j-1}) = 1, Z(a_j) = 1 \text{ or } 2)$. When $\alpha_1(t) = \alpha_2(t) = \alpha(t)$ in Figure 1(a) then

$$P_{11}(s,t) = e^{-\Lambda(s,t)+A(s,t)}\,, \quad P_{22}(s,t) = e^{-A(s,t)}\,, \quad P_{12}(s,t) = (1 - e^{-\Lambda(s,t)})\, e^{-A(s,t)} \tag{18}$$

for $s \le t$, where $\Lambda(s,t) = \int_s^t \lambda(u)\,du$ and $A(s,t) = \int_s^t \alpha(u)\,du$. In this case we find that $\eta_j^\star = \eta_j = 1 - \exp\{-\Lambda(a_{j-1}, a_j)\}$, so a consistent estimate is obtained.

A similar result holds for Case B. The partial likelihood function for case B corresponding to (9) is

$$L_{B1}^I(\eta) = \prod_{i=1}^n \{\lambda(t_i)\,\mathcal{F}(t_i)\}^{\delta_i}\,\mathcal{F}(a_{m_i^\star})^{1-\delta_i}\,,$$

where $\eta = (\lambda(1), \lambda(2), \ldots)'$. For a given t we find that $\partial \log L_{B1}^I(\eta)/\partial\lambda(t) = 0$ gives

$$\widehat{\lambda}(t) = \frac{\sum_{i=1}^n \delta_i\, I(t_i = t)}{\sum_{i=1}^n (1-\delta_i)\, I(a_{m_i^\star} \ge t) + \delta_i\, I(t_i \ge t)}\,. \tag{19}$$

By similar arguments to those above $\widehat{\lambda}(t)$ for $t \in (a_{j-1}, a_j]$ converges in probability to

$$\lambda^\star(t) = \left\{\frac{P_{22}(t, a_j)}{P_{11}(t, a_j) + P_{12}(t, a_j)}\right\}\lambda(t)\,, \tag{20}$$

as $n \to \infty$; see Lawless (2013). Under independent LTF (i.e. when $\alpha_1(t) = \alpha_2(t)$), (18) holds and the multiplier of $\lambda(t)$ in (20) equals one, so that $\lambda^\star(t) = \lambda(t)$. When $\alpha_1(t) \ne \alpha_2(t)$, the expressions

$$P_{11}(s,t) = e^{-\Lambda(s,t)+A_1(s,t)}\,, \;\; P_{22}(s,t) = e^{-A_2(s,t)}\,, \;\; P_{12}(s,t) = \int_s^t P_{11}(s, u^-)\,\lambda(u)\,P_{22}(u,t)\,du \tag{21}$$

can be used to compute $\lambda^\star(t)$. Cook and Lawless (2018, Section 7.2.5) consider this for the time-homogeneous case, where $\lambda(t) = \lambda$, $\alpha_1(t) = \alpha_1$, $\alpha_2(t) = \alpha_2$.

4.2 The Semiparametric Cox model

Here we investigate the implications of the Scheme I: Case B observation process when fitting a Cox regression model of the form $dH(s|X_i) = dH_0(s)\exp(X_i'\gamma)$ to the available data where X_i is a fixed $p \times 1$ covariate. To this end we let $Y_{ij}(s) = I(a_{j-1} \le s < a_j < C_i)$ indicate $s \in \mathcal{A}_j = [a_{j-1}, a_j)$ and the LTF time for individual i exceeds a_j, let $Y_i^\dagger(s) = I(s \le T_i)$ indicate individual i is at risk of failure at s, and let $\bar{Y}_{ij}(s) = Y_{ij}(s)Y_i^\dagger(s)$. The semiparametric maximum likelihood score equations for the working Cox model based on Scheme 1, Case B are

$$U_1(s; dH_0(\cdot), \gamma) = \sum_{i=1}^n \sum_{j=1}^R \bar{Y}_{ij}(s)\,\{dN_i(s) - dH(s \mid X_i)\}\,, \quad s > 0 \tag{22a}$$

$$U_2(dH_0(\cdot), \gamma) = \sum_{i=1}^n \sum_{j=1}^R \int_{a_{j-1}}^{a_j} \bar{Y}_{ij}(s)\,\{dN_i(s) - dH(s \mid X_i)\}\,X_i \tag{22b}$$

given a sample of size n. Under mild conditions the estimates $H_0(t)$ and $\widehat{\beta}$ obtained by setting (22a) and (22b) equal to zero and solving, converge in probability to a function $H_0^\star(t)$ and vector $\gamma^\star$ for which the expectations of (22a) and (22b) are equal to zero. We therefore derive $E\{U_1(s)\}$ where the expectation is taken with respect to the three-state process with $\lambda_{12}(t|X_i) = \lambda_0(t)\,\exp(X_i\beta)$ for the $1 \to 2$ transition intensity in Figure 1(a), and $\alpha_1(t)$ and $\alpha_2(t)$ the censoring intensities which are taken to be independent of the covariate. Setting this expectation equal to zero gives the solution

$$d\tilde{H}_0^*(s;\gamma) = \frac{E_X\{P(Z(s^-)=1, Z(s)=2, Z(a_j)=2 \mid X)\}}{E_X\{P(Z(s^-)=1, Z(a_j)<3 \mid X)\,\exp(X'\gamma)\}} = \frac{r_j^{(0)}(s)}{r_j^{(0)}(s;\gamma)}\,, \tag{23}$$

for $s \in \mathcal{A}_j = [a_{j-1}, a_j)$ where the numerator is $E_X\{P(Z(s^-)=1 \mid X)\,\lambda_{12}(s|X)\,P(Z(a_j)=2 \mid Z(s)=2, X)\}$. Substituting (23) for $dH_0(s)$ in (22b), taking the expectation of the resulting function, and setting it equal to zero, gives the equation

$$\sum_{j=1}^{R} \int_{a_{j-1}}^{a_j} \left\{ r_j^{(1)}(u) - \frac{r_j^{(1)}(u;\gamma)}{r_j^{(0)}(u;\gamma)}\, r_j^{(0)}(u) \right\} du = 0\,, \tag{24}$$

where for $k = 0, 1, 2$ and $u \in [a_{j-1}, a_j)$,

$$r_j^{(k)}(u) = E_X\{P(Z(u^-)=1 \mid X)\,\lambda_{12}(u \mid X)\,P(Z(a_j)=2 \mid Z(u)=2, X)\,X^{\otimes k}\}\,, \tag{25a}$$

$$r_j^{(k)}(u;\gamma) = E_X\{P(Z(u^-)=1, Z(a_j)<3 \mid X)\,X^{\otimes k}\,\exp(X'\gamma)\}\,. \tag{25b}$$

The estimator $\widehat{\gamma}$ obtained by setting (22a) and (22b) equal to zero and solving is consistent for γ^* defined as the solution to (24) and the limiting value of $\widehat{H}_0(t) = \int_0^t d\widehat{H}_0(s)$ is given by replacing γ with γ^* in (23) and integrating (Struthers and Kalbfleisch, 1986; Lin and Wei, 1989).

Here we consider the same parametric setting as in Section 3.2 and plot the limiting value γ^* in Figure 4 for $K = R = 5$ (panel (a)) and $K = R = 10$ (panel (b)) with separate lines for $\alpha_2/\alpha_1 = 1, 2$ and 5. Note that the data are generated under a proportional hazards model with a piecewise-constant baseline hazard described in Section 3.2, but here the analysis is based on a Cox regression model so the value of K simply reflects the number of pieces in the true baseline hazard here. Interestingly, we see a very similar trend in the bias for the regression coefficient in panel (a) to what was seen earlier in the results for the parametric analyses, and a smaller bias is also seen in panel (b) reflecting the decreasing effect of dependent LTF when the assessments are more frequent; this arises because the $2 \to 3$ intensity α_2 is at play for a shorter period of time in such settings. Figure 4 (bottom panels) shows the asymptotic bias of $H_0^*(t;\gamma^*) - \Lambda_0(t)$ with $\alpha_2/\alpha_1 = 5$ for $\pi_F^\circ/\pi_F = 0.2$, 0.4, 0.6 and 0.8. We find in panel (c) that the bias can be appreciable under heavy censoring but decreases as one might expect with more frequent assessments (panel (d)). The trend of the asymptotic bias within intervals $[b_{j-1}, b_j)$ reflects the fact that failures early during the corresponding interval are less likely to be reported due to LTF later during the interval, but the LTF process has a weaker effect on the estimator of the baseline hazard later during the interval since the time at risk of LTF is shorter.

5 The Waterloo Smoking Prevention Study

Here we report on an application involving data from a large smoking prevention study in which 100 schools were randomized to receive a health curriculum enhanced with more information on the consequences of smoking delivered by a nurse or specially trained teacher, or the regular curriculum. Students in participating schools entered the study in grade 6 and completed follow-up assessments at school visits in grades 7 and 8 along with follow-up assessments during high school grades 9 to 12. Among the 100 schools recruited there were 4,456 students taking part. We restrict attention in the following analysis to 527 children who had a high social models risk score, which was obtained if both of their parents smoked. We consider the failure time to be the first incident of smoking and consider a missed visit as representing loss to follow-up.

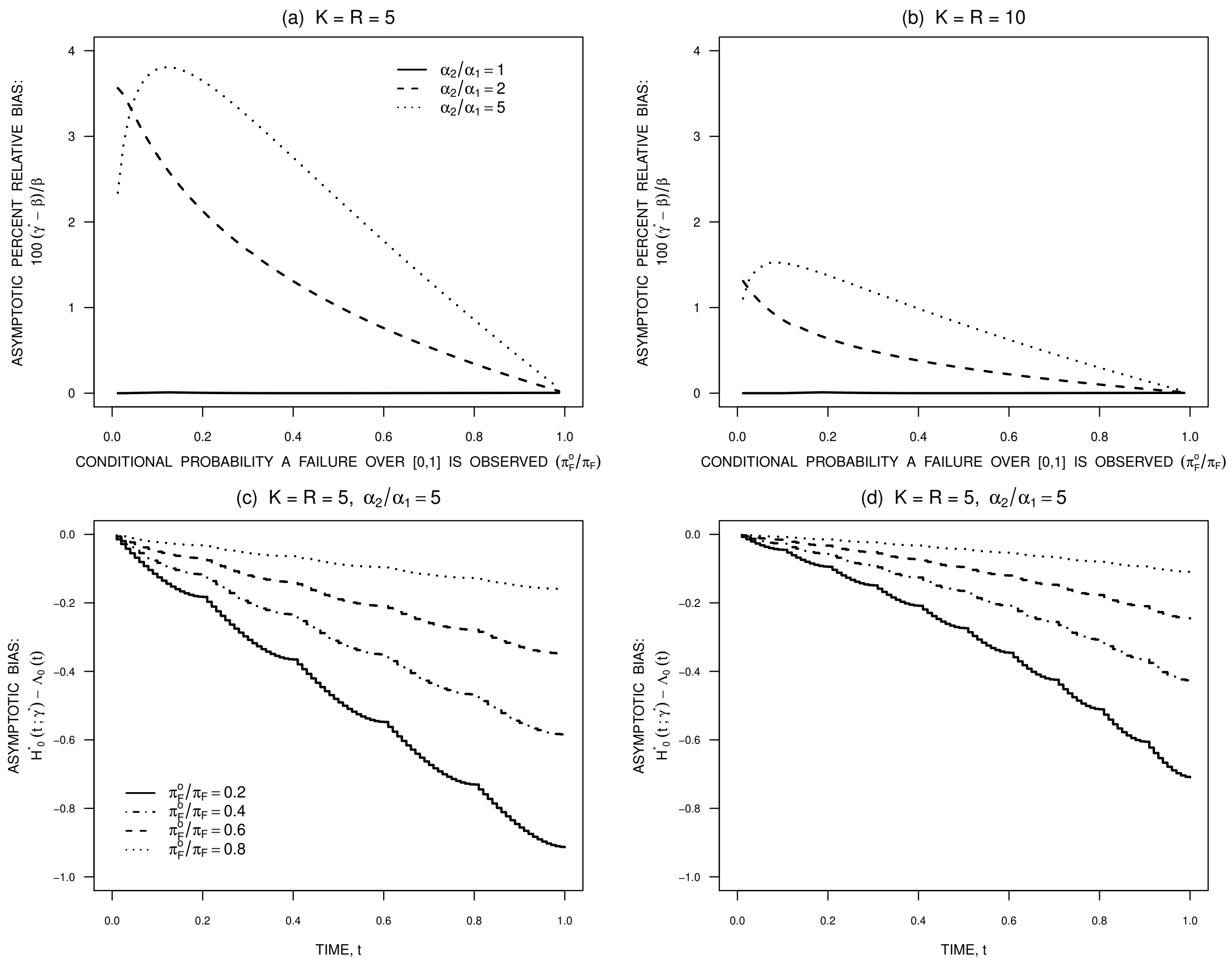


Figure 4: Plot of asymptotic percent relative bias $100\times(\gamma^*-\beta)/\beta$ versus π_F°/π_F (top panels) and plot of asymptotic bias $H_0^*(t;\gamma^*)-\Lambda_0(t)$ versus t with $\alpha_2/\alpha_1=5$ (bottom panels) from semiparametric analyses under different settings including dependent LTF; $\pi_F=0.80$

The intervention effect was assessed based on a proportional hazards model with $\lambda(t|X) = \lambda_0(t)\exp(X\beta)$, where $\lambda_0(t)$ was again specified to be piecewise constant with break-points each year, and $X=1$ if the child received the intensive anti-smoking program and $X=0$ otherwise. An analysis was carried out based on a simple two-state failure process under the assumption of independent LTF. For the joint analysis based on the three-state model, the LTF intensity had a piecewise-constant form motivated by the fact that the LTF process is known to be quite different following the completion of elementary school. In particular we assumed a constant LTF rate between each consecutive pair of assessments but allowed this rate to change from year to year. An intervention effect on the LTF process and a dependence between the failure and LTF times were accommodated by specifying $\alpha_1(t|X)=\alpha_{10}(t)\,\exp(X\xi)$ and $\alpha_2(t|X)=\alpha_{20}(t)\,\exp(X\xi)$, where $\alpha_{10}(t)$ reflects the intensity of a $1\to 3$ transition for a child receiving the standard curriculum. We further set $\alpha_{20}(t)=\alpha_{10}(t)\exp(\delta)$ so that δ reflects the dependence between the failure and LTF times, and take $\alpha_{10}(t)$ to be piecewise-constant with break-points each year.

The results of a standard interval-censored failure time analysis based on the proportional hazards model, along with the results for an analysis involving the joint model are given in Table 2. Note that fewer individuals contribute to the two-state analysis because individuals LTF following their initial assessment do not contribute to the corresponding likelihood. Estimates were obtained for the failure time and three-state analysis under a working independence assumption in which the outcomes for the

527 students are treated as independent, but within-school dependencies were accommodated through use of a robust variance estimate (Cook et al., 2002; Cook and Lawless, 2018).

We find no evidence of an intervention effect on the intensity for smoking from either the standard or joint models. This lack of effect is aligned with the results of the primary analysis reported by Cameron et al. (1999), which was directed at estimating the intervention effect on the yearly smoking prevalence rate. The hazard ratios are estimated to be 1.080 (95%CI: 0.789, 1.477) based on the standard analysis and 1.090 (95% CI: 0.801, 1.470) from the joint analysis. There is some variation in the intensity for onset of smoking over the grades as expected from both models. There is highly significant evidence of dependent LTF with $\exp(\widehat{\delta}) = 2.351$ (95% CI: 1.581, 3.501; p <0.001), and $\exp(\widehat{\xi}) = 1.406$ (95% CI: 0.947, 2.090; p = 0.091), which is suggestive of a higher withdrawal rate for individuals in the intervention arm. Despite this strong evidence of dependent LTF the impact on the estimate of the intervention effect is negligible; this finding is consistent with the results of our numerical studies which showed mild sensitivity of covariate effects to dependent LTF. More appreciable (but still quite modest) differences are seen in the estimates of the baseline intensity for the onset of smoking from the standard and joint models. Here we see an expected pattern of higher smoking rates in higher grades.

Table 2: Estimates obtained by fitting a two-state failure time model and three-state joint model to data from the Waterloo Smoking Prevention Project; restrict attention to students with high social model risk score in grade 6

			Two-State Model (N = 296 subjects)			Three-State Model (N = 527 subjects)		
	Parameter	Interval	Est.[a]	95% CI[b]	p^b	Est.[a]	95% CI[b]	p^b
					Failure Process			
$1 \to 2$ Intensity	$\log \lambda(t)$	$[6,7)$	-0.981	(-1.301, -0.661)		-0.933	(-1.248, -0.619)	
		$[7,8)$	-1.007	(-1.344, -0.669)		-0.973	(-1.309, -0.636)	
		$[8,9)$	-1.049	(-1.457, -0.642)		-0.940	(-1.349, -0.531)	
		$[9,10)$	-0.888	(-1.370, -0.405)		-0.886	(-1.368, -0.405)	
		$[10,11)$	-1.871	(-2.716, -1.026)		-1.819	(-2.651, -0.987)	
		$[11,12)$	-1.859	(-2.781, -0.938)		-1.846	(-2.743, -0.950)	
Intervention Effect	β		0.077	(-0.236, 0.390)	0.631	0.082	(-0.222, 0.385)	0.598
					Loss to Follow-up Process			
$1 \to 3$ Intensity	$\log \alpha_1(t)$	$[6,7)$	-	-	-	-2.603	(-3.058, -2.147)	
		$[7,8)$	-	-	-	-3.013	(-3.519, -2.507)	
		$[8,9)$	-	-	-	-2.317	(-2.909, -1.725)	
		$[9,10)$	-	-	-	-3.356	(-3.912, -2.800)	
		$[10,11)$	-	-	-	-2.794	(-3.355, -2.234)	
		$[11,12)$	-	-	-	-3.102	(-3.633, -2.570)	
$\log(\alpha_{20}(t)/\alpha_{10}(t))$	δ		-	-	-	0.855	(0.458, 1.253)	< 0.001
Intervention Effect	ξ		-	-	-	0.341	(-0.054, 0.737)	0.091

[a] Estimates are obtained based on proportional intensity models

[b] 95% CIs and p-values are computed based on sandwich-type S.E. (clustered by school an individual attended in Grade 6)

6 Some Further Remarks

6.1 Random Visit Times

We have focussed on settings where visits are scheduled at predetermined times, as in the protocols of many large cohort studies. In this case LTF is confirmed, or declared, at the first missed visit for an individual. The concept of a time of LTF is usually idealized in the sense that even with complete data a precise time at which an individual decided to drop out of a study might not be ascertainable. The methods considered in this article treat LTF time as interval-censored, so this is not a problem.

In many observational cohorts, visit times are random and may vary considerably within and between individuals. Cook and Lawless (2019) consider this situation and formalize the concepts of independent and dependent visit processes. This involves defining a visit counting process $\{N^A(t), t > 0\}$, where $N^A(t)$ is the number of visits for an individual up to time t. Using clinic-based cohorts of persons with rheumatic disease for illustration, they consider models that allow the intensity functions for visit processes to depend on the disease history $\{Z(t), t > 0\}$. They consider situations where LTF times are observable, however, so that a combined visits-LTF process as shown in Figure 5 applies. Cook and Lawless (2019) mainly consider independent LTF, but the methods in the paper can readily be extended to deal with settings where LTF may depend on disease history.

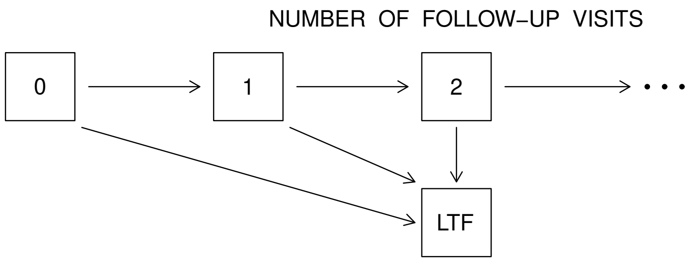


Figure 5: A joint visit and LTF process.

Our focus in this article is on settings where exact LTF times are not observed. This is difficult to deal with when visit times are random, and especially when they are highly variable. The key problem is that the observed data simply record in this case that for an individual with administrative censoring time A, there was a final visit at some time a_m and then no further visits over (a_m, A). It is therefore uncertain whether LTF occurred or not, and observation schemes I and II in this article do not apply here because of this. Researchers are then faced with the issue of how to specify end-of-follow-up times for individuals. The use of a_m is typically inadvisable if visit processes are related to disease histories, since it uses the information that no visit occurred over (a_m, A). There have been empirical studies of the effect of different choices for end-of-follow-up, for example, in connection with HIV-AIDS cohorts (e.g. Shepherd et al., 2011; Lesko et al., 2018) but these have been in the context of independent visit processes and have not involved joint modeling. Our current recommendation is to use the joint visits-disease process models in Cook and Lawless (2019), perhaps combined with a protocol for declaring LTF. For example, in a setting where inter-visit times vary widely but are typically 2 years or less, we might declare LTF once no visit has occurred for 2 years. In this case we treat $a_m + 2$ years as the end-of-follow-up time; note that there is no new information on the disease process observed at that time. More informed approaches can be considered if some level of tracing is undertaken for persons who are thus declared LTF, as discussed for Scheme III in this article; see Appendix B. In particular, for persons successfully traced we can obtain information concerning their disease history over the time period (a_m, A). Research on this topic is ongoing and we hope to report further in a future communication.

6.2 Illness-Death and Other Multistate Processes

We can consider a similar approach to joint modeling when a general multistate process (i.e. not simply a failure time process) is of interest. Joint models analogous to that in Figure 1(b) can be used; Figure A.1 in Appendix A shows such a joint model for an illness-death process and LTF. In the general case, if there are no absorbing states in the disease process model, then observation Schemes I and II are the same; both involve observation of an individual until they become LTF. When there is an absorbing state, then Scheme I involves observation up to the first time when a person is either in an absorbing state or LTF, whereas Scheme II follows individuals until they are LTF. The standard likelihood function used with an assumption of independent LTF does not use the joint model and includes data on the disease process up to the last visit where an individual is not LTF. We give this likelihood in Appendix A and note how asymptotic bias in standard estimates can be assessed when there is dependent LTF, as represented by the joint model. The likelihood functions under observation Scheme II and the joint model are also easy to write down. They include terms for cases where an individual was LTF at some visit time a_{m+1} but were not LTF and in some observed state, absorbing or otherwise, at time a_m.

A death state is an important special case of an absorbing state. In this case we can typically ascertain the exact time of entry to the death state. Slightly different settings can arise. In the simplest case, a person who does not appear for a visit at time a_{m+1} can be identified as either being dead or LTF; if they are dead the time of death can be obtained. In studies with planned visit times this is often the case. A more complicated case is where the vital status of an individual who does not appear for a visit at time a_{m+1} cannot immediately be ascertained. In that case we can treat them provisionally as LTF, but in some cases deaths are eventually reported, but subject to random reporting delays. Many cohort studies link periodically to death registries, for example. We stress that such registries usually involve random reporting delays and that information about the distribution of delays is needed to avoid potential bias. A number of authors, including Binder and Schumacher (2014) and Binder et al. (2017, 2019) have considered bias in the estimation of failure time models caused by death being treated as independent LTF. There have not been thorough investigations of the effects of reporting delays, however, and we are currently undertaking this.

6.3 Intermittent Observation of Time-Varying Covariates

Information is routinely collected on time-varying covariates at visits which may be used for modeling the life history process. In this case we may consider the failure intensity corresponding to a $1 \to 2$ transition in Figure 1(b) as

$$\lim_{\Delta t \downarrow 0} \frac{P(Z(t+\Delta t^-) = 2 \mid Z(t^-) = 1, X(t))}{\Delta t} = \lambda(t \mid X(t))$$

and the LTF intensities as

$$\lim_{\Delta t \downarrow 0} \frac{P(Z(t+\Delta t^-) = 3 \mid Z(t^-) = k, X(t))}{\Delta t} = \alpha_k(t \mid X(t)), \quad k = 1, 2\,.$$

If $T \perp C | \{X(s), 0 < s\}$ then it is sufficient to model the failure time process. In practice, however, models can only be fit of the form

$$\lim_{\Delta t \downarrow 0} \frac{P(Z(t+\Delta t^-) = 2 \mid Z(t^-) = 1, X(a_{A(t^-)}))}{\Delta t} = \lambda(t \mid X(a_{A(t^-)}))\,,$$

where $a_{A(t^-)}$ denotes the time of the most recent assessment, which may be viewed as a reasonable strategy when visits are regularly scheduled, but inadequate when they are random and highly variable as this can lead to long lags between t and the most recent measurement of the covariates. That is, if the temporal variation of the covariate values is high over the typical intervals between the scheduled visits, the most recently recorded covariate value may be a poor approximation of the current covariate value and models are misspecified. This leads to a type of covariate measurement error problem which has been studied by de Bruijne et al. (2001) and Andersen and Liestøl (2003).

The conditioning on time-dependent covariates weakens the nature of the conditionally independent LTF assumption, however if the covariate values have high temporal variation, conditioning on the most recently recorded value may not render LTF as conditionally independent and mitigate biases from dependent LTF. If measurements on time-dependent covariates are available that can render LTF conditionally independent, but interest lies in the marginal hazard and failure time distribution, inverse probability of censoring weighting can be carried out (Satten et al., 2001).

7 Concluding Remarks

We have focussed primarily on the setting of many cohort studies where recruited individuals are scheduled to be examined according to protocol at fixed visit times. When individuals are found to have experienced the event between the previous and current assessment the failure times may be available retrospectively in some cases through self-report, a review of medical records, or other means. In the present context we found this retrospective data did not add an appreciable amount of information. More importantly we found that in both the setting where the data are retrospectively available and when it is not, use of a joint model can mitigate the bias that would otherwise arise from dependent LTF. We have considered joint models which are Markov. Sensitivity analyses can be conducted to assess the robustness of findings to misspecification of the LTF process in joint models. More flexible models which retain Markov LTF intensities are straightforward to implement, but ones with semi-Markov LTF intensities can be challenging to fit.

The main focus of this article has been on failure time processes; more general multistate processes were discussed briefly in Section 6.2. In other settings interest may lie in modeling time-varying marker processes such as prostate-specific antigen (Proust-Lima et al., 2008). In such cases a fundamentally different framework for joint modeling of the marker and failure processes seems necessary, with potentially stronger assumptions which are more difficult to check; see Tsiatis and Davidian (2004) for a discussion of the conceptual issues and an overview of the modeling approaches. Meaningful estimands are often difficult to specify in such cases (Commenges, 2019) and the impact of dependent LTF on related inference warrants further study for this setting.

As discussed in Section 6.1 matters are more complicated when visit times are random. In circumstances where patients are being treated by study clinicians, this will often be the case since they will naturally seek care when symptoms are manifest. Data from such visits can be excluded from a study database but interventions administered during such visits can change the course of the failure or multistate process of interest so that is not a suitable approach. A further complication is that such spontaneous visits can alter the probability that future scheduled visits are made, thereby more substantially impacting the visit process. Further work is warranted on the development and use of hybrid visit process models, which involve an intensity for random visits along with a model for an indicator of whether scheduled visits are made. Embedding these into a joint model with potential LTF would enable comprehensive modeling of the failure and observation processes but these models could require collection of additional information such as the nature of any visit process.

The discrete visit process model discussed in Appendix C is particularly appealing as a framework for considering the challenge of intermittently missed visits. In this case there may be a sequence of visits scheduled in advance but some individuals may attend scheduled visits sporadically. This makes modeling of the LTF process more difficult and a definition of LTF may need to incorporate this complication.

Finally in related work we have discussed the idea of tracing studies as a means of strengthening inferences for joint models (Lawless and Cook, 2019). In the LTF setting this would involve use of extraordinary resources to track down individuals who are identified as LTF or perhaps use of external administrative databases to obtain partial information on the status of the individuals LTF. Dynamic tracing studies could also be carried out during the course of a study to track down individuals who have not shown up for clinic visits in order to obtain partial information; even the reason for the missed visit could facilitate fitting of more complete joint models. Further study of the utility of such methods is warranted given the many large cohort studies now being undertaken around the world.

ACKNOWLEDGEMENTS

This work was funded by the Natural Sciences and Engineering Research Council of Canada (RGPIN 155849 and RGPIN 04207 for R.J.C. and RGPIN 8597 for J.F.L.) and the Canadian Institutes of Health Research (FRN 13887 for R.J.C.). R.J.C. is a Mathematics Faculty Research Chair at the University of Waterloo. The authors thank Ker-Ai Lee for help in statistical computing in this research and Drs. Roy Cameron and Steve Brown for permission to use data from the Waterloo Smoking Prevention Project in the illustration. Helpful reviews from the editors and referees are also gratefully acknowledged.

REFERENCES

Andersen, P., Borgan, O., Gill, R., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York, NY.

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

Baker, S., Wax, Y., and Patterson, B. (1993). Regression analysis of grouped survival data: informative censoring and double sampling. *Biometrics*, 49(2):379–389.

Beyersmann, J., Allignol, A., and Schumacher, M. (2012). *Competing Risks and Multistate Models with R*. Springer Science + Business Media, LLC, New York, NY.

Binder, N., Blümle, A., Balmford, J., Motschall, E., Oeller, P., and Schumacher, M. (2019). Cohort studies were found to be frequently biased by missing disease information due to death. *Journal of Clinical Epidemiology*, 105:68–79.

Binder, N., Hernböck, A., and Schumacher, M. (2017). Estimating hazard ratios in cohort data with missing disease information due to death. *Biometrical Journal*, 59(2):251–269.

Binder, N. and Schumacher, M. (2014). Missing information caused by death leads to bias in relative risk estimates. *Journal of Clinical Epidemiology*, 67(10):1111–1120.

Cameron, R., Brown, K., Best, J., Pelkman, C., Madill, C., Manske, S., and Payne, M. (1999). Effectiveness of a social influences smoking prevention program as a function of provider type, training method, and school risk. *American Journal of Public Health*, 89(12):1827–1831.

Commenges, D. (2019). Dealing with death when studying disease or physiological marker: the stochastic system approach to causality. *Lifetime Data Analysis*, 25(3):381–405.

Commenges, D. and Gégout-Petit, A. (2007). Likelihood for generally coarsened observations from multistate or counting process models. *Scandinavian Journal of Statistics*, 34(2):432–450.

Cook, R., Kalbfleisch, J., and Yi, G. (2002). A generalized mover-stayer model for panel data. *Biostatistics*, 3(3):407–420.

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

Cook, R. and Lawless, J. (2018). *Multistate Models for the Analysis of Life History Data*. CRC Press, Boca Raton, FL.

Cook, R. and Lawless, J. (2019). Independence conditions and the analysis of life history processes under intermittent observation. *Biostatistics (submitted).*

Cook, R., Lawless, J., and Lee, K. (2003). Cumulative processes related to event histories. *SORT*, 27:13–30.

de Bruijne, M., le Cessie, S., Kluin-Nelemans, H., and van Houwelingen, H. (2001). On the use of Cox regression in the presence of an irregularly observed time-dependent covariate. *Statistics in Medicine*, 20(24):3817–3829.

Farewell, V., Lawless, J., and Gladman, D. (2003). Tracing studies and analysis of the effect of loss to follow-up on mortality estimation from patient registry data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52(4):445–456.

Fisher, L. and Kanarek, P. (1974). Presenting censored survival data when censoring and survival times may not be independent. In Proschan, F. and Serfling, R., editors, *Reliability and Biometry.* , pages 303–326. SIAM, Philadelphia, PA.

Frangakis, C. and Rubin, D. (2001). Addressing an idiosyncrasy in estimating survival curves using double sampling in the presence of self-selected right censoring. *Biometrics*, 57(2):333–342.

Gladman, D. and Chandran, V. (2011). Observational cohort studies: lessons learnt from the University of Toronto Psoriatic Arthritis Program. *Rheumatology*, 50(1):25–31.

Grüger, J., Kay, R., and Schumacher, M. (1991). The validity of inferences based on incomplete observations in disease state models. *Biometrics*, 47(2):595–605.

Joly, P., Commenges, D., Helmer, C., and Letenneur, L. (2002). A penalized likelihood approach for an illness-death model with interval-censored data: application to age-specific incidence of dementia. *Biostatistics*, 3(3):433–443.

Lange, J., Hubbard, R., Inoue, L., and Minin, V. (2015). A joint model for multistate disease processes and random informative observation times, with applications to electronic medical records data. *Biometrics*, 71(1):90–101.

Lawless, J. (2013). Armitage Lecture 2011: the design and analysis of life history studies. *Statistics in Medicine*, 32(13):2155–2172.

Lawless, J. and Cook, R. (2019). A new perspective on loss to followup in failure time and life history studies. *Statistics in Medicine*.

Lesko, C., Edwards, J., Cole, S., Moore, R., and Lau, B. (2018). When to Censor? *American Journal of Epidemiology*, 187(3):623–632.

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.

Proust-Lima, C., Taylor, J., Williams, S., Ankerst, D., Liu, N., Kestin, L., Bae, K., and Sandler, H. (2008). Determinants of change in prostate-specific antigen over time and its association with recurrence after external beam radiation therapy for prostate cancer in five large cohorts. *International Journal of Radiation Oncology Biology Physics*, 72(3):782–791.

Raina, P., Wolfson, C., Kirkland, S., Griffith, L., Oremus, M., Patterson, C., Tuokko, H., Penning, M., Balion, C., Hogan, D., Wister, A., Payette, H., Shannon, H., and Brazil, K. (2009). The Canadian longitudinal study on aging (CLSA). *Canadian Journal on Aging/La Revue Canadienne du Vieillissement*, 28(3):221–229.

Satten, G., Datta, S., and Robins, J. (2001). Estimating the marginal survival function in the presence of time dependent covariates. *Statistics and Probability Letters*, 54(4):397–403.

Scharfstein, D. and Robins, J. (2002). Estimation of the failure time distribution in the presence of informative censoring. *Biometrika*, 89(3):617–634.

Shepherd, B., Blevins, M., Vaz, L., Moon, T., Kipp, A., José, E., Ferreira, F., and Vermund, S. (2011). Impact of definitions of loss to follow-up on estimates of retention, disease progression, and mortality: application to an HIV program in Mozambique. *American Journal of Epidemiology*, 178(5):356–367.

Siannis, F. (2011). Sensitivity analysis for multiple right censoring processes: investigating mortality in psoriatic arthritis. *Statistics in Medicine*, 30(4):356–367.

Slud, E. and Rubinstein, L. (1983). Dependent competing risks and summary survival curves. *Biometrika*, 70:643–649.

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

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

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25.

Willekens, F. (2014). *Multistate Analysis of Life Histories with R*. Springer International Publishing, Switzerland.

APPENDIX A DEPENDENT LTF WITH MULTISTATE PROCESSES

Here the consequences of dependent LTF is considered when more general multistate processes are under intermittent observation according to Scheme I. Let $\mathcal{S} = \{1, \ldots, K\}$ denote the state space and let $\mathcal{A}$ denote the subset of absorbing states. Let $\{\mathcal{Z}(t), t > 0\}$ denote the multistate process governed by intensities

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

for $l \leq k$ and $k \notin \mathcal{A}$ where $\mathcal{H}(t) = \{\mathcal{Z}(s), 0 < s < t\}$ is the history. We omit covariates for convenience but fixed covariates are easily handled. We focus on Markov processes for which $\lambda_{kl}(t|\mathcal{H}(t))$ is equal to $\lambda_{kl}(t)$. In this class of models, it is possible to compute the $K \times K$ matrix of transition probabilities $\mathcal{P}(s,t)$ with (k,l) entry $P(\mathcal{Z}(t) = l|\mathcal{Z}(s) = k)$.

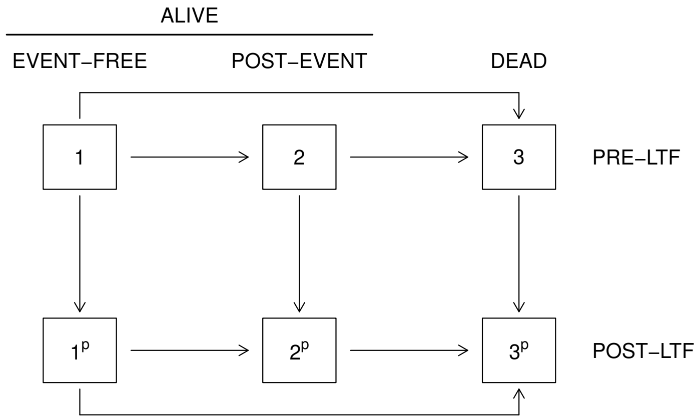


Figure A.1: A joint model for an underlying illness-death process and LTF process.

We let $\mathcal{S}^p = \{1^p, \ldots, K^p\}$ denote the states that can be occupied after loss to follow-up. An expanded state space is formed by considering the union of all possible states in $\mathcal{S}$ and $\mathcal{S}^p$ denoted $\mathcal{S}^J = \mathcal{S} \cup \mathcal{S}^p = \{1, \ldots, K, 1^p, \ldots, K^p\}$, where k^p denotes disease state k after loss to follow-up. We let $Z(t)$ denote the state of this joint process and $H(t) = \{Z(t), 0 < s < t\}$ let be the respective history at time t. Then

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

and

$$\lim_{\Delta t \downarrow 0} \frac{P(Z(t+\Delta t^-) = k^p \mid Z(t^-) = k, H(t))}{\Delta t} = \alpha_k(t;\theta)\,,$$

so that the intensities for transitions between disease states in the expanded model are the same as those in the model solely for the disease process. Figure A.1 gives an illustrative state space for a joint model involving an underlying illness-death process.

The likelihood under Scheme I can be written as follows. Let $Y_{ikr} = I(\mathcal{Z}_i(a_{r-1}) = k)$,

$$p_{ikl}(s,t) = P(\mathcal{Z}_i(t) = l \mid \mathcal{Z}_i(s) = k)\,,$$

and $p_{iklr} = p_{ikl}(a_{r-1}, a_r)$, which is indexed by η. Then if $N_{iklr} = I(\mathcal{Z}_i(a_{r-1}) = k, \mathcal{Z}_i(a_r) = l)$,

$$L^I_{A1}(\eta) \propto \prod_{i=1}^{n} \prod_{r=1}^{R} \left\{ \prod_k \left[\prod_l p_{iklr}(\eta)^{N_{iklr}} \right]^{Y_{ikr}\, I(\mathcal{Z}_i(a_r) \notin \mathcal{S}^p)} \right\}$$

where if $l = k$ we note $p_{iklr}(\eta) = 1 - \sum_{l \neq k} p_{iklr}(\eta)$. The resulting score vector can be written as

$$U^I_{A1}(\eta) = \sum_{i=1}^{n} \sum_{r=1}^{R} I(\mathcal{Z}_i(a_r) \notin \mathcal{S}^p) \sum_{k=1}^{K} Y_{ikr} \sum_{l} N_{iklr}\, \partial \log p_{iklr}(\eta)/\partial \eta \,.$$

The limiting value of $\widehat{\eta}$ is obtained by solving $E\{U^I_{A1}(\eta)\} = 0$ where the expectation is taken with respect to the joint model.

Under Scheme II the likelihood contribution from individual i based on the joint model is

$$\prod_{r=1}^{R} \Big\{ \prod_{k \in \mathcal{S}} \Big[\Big\{ \prod_{l \in \mathcal{S}} p^J_{iklr}(\psi)^{N_{iklr}} \Big\} \Big\{ \sum_{l \in \mathcal{S}^p} p^J_{iklr}(\psi) \Big\}^{I(Z_i(a_r) \in \mathcal{S}^p)} \Big]^{Y_{ikr}} \Big\}$$

where $p^J_{iklr}(\psi) = P(Z_i(a_r) = l | Z_i(a_{r-1}) = k, X_i)$ is the transition probability based on the joint Markov model and $N_{iklr} = I(Z_i(a_{r-1}) = k, Z_i(a_r) = l)$ indicates a $k \to l$ transition, where $k \in \mathcal{S}$; if $l \in \mathcal{S}$ then $Z_i(a_r)$ is observed, but otherwise it is only known that $l \in \mathcal{S}^p$ and that LTF has occurred.

APPENDIX B LIKELIHOOD FOR OBSERVATION SCHEME III

Here we consider the likelihood construction under observation Scheme III mentioned in Section 2.2 in which observation continues until a_M or for some individuals who are LTF until entry to state 2^p. The paths can be of five unique forms. We may have $a_R < \min(T, C)$ in which case the data are of the form $Z(a_1) = \cdots = Z(a_R) = 1$ and the likelihood contribution is $P_{11}(0, a_R)$. For path 2 $a_m < T < a_{m+1}$ and $a_R < C$ so failure is detected at a_{m+1} but there is no LTF; the resulting data will be of the form $Z(a_1) = \cdots = Z(a_m) = 1$ and $Z(a_{m+1}) = \cdots = Z(a_R) = 2$ for some $m < R$ and the likelihood contribution is $P_{11}(0, a_m)\, P_{12}(a_m, a_{m+1})\, P_{22}(a_{m+1}, a_R)$. For path 3 we have $a_m < T < a_{m+1}$ for some m and $a_r < C < a_{r+1}$ so LTF is determined post-failure with $m + 1 \leq r$. Here the likelihood contribution is

$$P_{11}(0, a_m)\, P_{12}(a_m, a_{m+1})\, [P_{22}(a_{m+1}, a_r)]^{I(m+1<r)}\, P_{22^p}(a_r, a_r + 1)\,.$$

For path 4 $a_r < C < a_{r+1}$ and $a_R < T$ so there is LTF detected at a_{r+1} but no failure by a_R; the likelihood contribution is $P_{11}(0, a_r)\, P_{11^p}(a_r, a_{r+1})\, P_{1^p1^p}(a_{r+1}, a_R)$. For path 5, $a_r < C < a_{r+1}$ and $a_m < T < a_{m+1}$ so LTF is detected at a_{r+1} and failure occurs over $[a_m, a_{m+1}]$ with $r + 1 \leq m$; for this the likelihood contribution is

$$P_{11}(0, a_r)\, P_{11^p}(a_r, a_{r+1})\, P_{1^p1^p}(a_{r+1}, a_m)\, P_{1^p2^p}(a_m, a_{m+1})\,.$$

For Case B the contributions for paths 2 and 3 change as they did under Scheme II. In addition, the contribution for path 5 changes with the addition $T = t$ so the term $P_{1^p2^p}(a_m, a_{m+1})$ is replaced by

$$P_{1^p1^p}(a_m, t)\, \lambda(t)\, P_{2^p2^p}(t, a_{m+1})\,,$$

where the last term is equal to 1 as 2^p is an absorbing state.

For Scheme III we can estimate all four transition intensities in Figure 1(b). Thus it is only under Scheme III, which involves tracing and extended follow-up of individuals LTF, that independent censoring can be assessed, by comparing estimates of $\lambda(t)$ and $\lambda^p(t)$. For most of the article we focus on Schemes I and II and assume $\lambda^p(t) = \lambda(t)$, since our primary objective is to investigate the dependent LTF induced by intermittent observation.

APPENDIX C CONTINUOUS-TIME AND DISCRETE-TIME LTF PROCESSES

Here we give some remarks on the relation between models for the LTF process under the current continuous-time formulation and a discrete-time one, which is often used in longitudinal data settings when interest lies in marginal analyses. We let scheduled visits be at times $a_0 < a_1 < \cdots < a_R$ and let $\Delta_j = I(\text{not LTF at } a_j) = I(a_j < C)$. Let

$$P(\Delta_j = 1 \mid H(a_{j-1}, a_j), H(a_{j-1})) = \begin{cases} p_{1j} = p_1(a_{j-1}, a_j) & \text{if} \quad Z(a_j) = 1 \\ p_2(a_{j-1}, a_j, t) & \text{if} \quad Z(a_j) = 2\,, T = t\,. \end{cases}$$

If we consider Scheme 2, then the likelihood contributions for data over $(a_{j-1}, a_j]$ are of the form

1. $P(Z(a_{j-1}) = 1 \mid Z(a_{j-1}) = 1, \Delta_j = 1) = \exp\left(-\Lambda(a_{j-1}, a_j)\right) p_{1j}$ when an individual is in state 1 for the successive visit times a_{j-1} and a_j,

2. (a) If the failure time is not retrospectively observed as in Case A observation the corresponding contribution is

   $$P(Z(a_j) = 2 \mid Z(a_{j-1}) = 1, \Delta_j = 1) = \int_{a_{j-1}}^{a_j} \exp\left(-\Lambda(a_{j-1}, t)\right) \lambda(t)\, p_2(a_{j-1}, a_j, t)\, dt$$

   when state 1 is occupied at a_{j-1} and state 2 is occupied at a_j;

   (b) For Case B observation in which the failure time is retrospectively determined the contribution is

   $$P(Z(a_j) = 2, T \mid Z(a_j) = 1, \Delta_j = 1) = \exp\left(-\Lambda(a_{j-1}, t)\right) \lambda(t)\, p_2(a_{j-1}, a_j, t)\ ;$$

3. If state 1 is occupied at a_{j-1} and state 3 is occupied at a_j (i.e. LTF occurs over (a_{j-1}, a_j)) then the contribution is

   $$P(Z(a_j) = 3 \mid Z(a_{j-1}) = 1) = 1 - \sum_{l=1}^{2} P(Z(a_j) = l \mid Z(a_{j-1}) = 1, \Delta_j = 1)\ .$$

For continuous-time models, the likelihood contributions corresponding to those above are

$$\exp\left(-\Lambda(a_{j-1}, a_j)\right) \exp\left(-A_1(a_{j-1}, a_j)\right) = \exp\left(-\Lambda(a_{j-1}, a_j)\right) p_{1j}$$

for **1**,

$$\exp\left(-\Lambda(a_{j-1}, t)\right) \lambda(t) \exp\left(-A_1(a_{j-1}, t)\right) \exp\left(-A_2(t, a_j)\right) = \exp\left(-\Lambda(a_{j-1}, t)\right) \lambda(t)\, p_2(a_{j-1}, a_j, t)$$

for **2 a.**,

$$\int_{a_{j-1}}^{a_j} \exp\left(-\Lambda(a_{j-1}, t)\right) \lambda(t)\, p_2(a_{j-1}, a_j, t)\, dt$$

for **2 b.** and

$$P(Z(a_j) = 3 \mid Z(a_{j-1}) = 1) = 1 - \exp\left(-\Lambda(a_{j-1}, a_j)\right) p_{1j} - \int_{a_{j-1}}^{a_j} \exp\left(-\Lambda(a_{j-1}, t)\right) \lambda(t)\, p_2(a_{j-1}, a_j, t)\, dt$$

for **3**.