A new perspective on loss to follow-up in failure time and life history studies

JERALD F. LAWLESS

Department of Statistics and Actuarial Science,

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

E-mail: jlawless@uwaterloo.ca

RICHARD J. COOK

Department of Statistics and Actuarial Science,

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

Summary

A framework is proposed for the joint modeling of life history and loss to follow-up (LTF) processes in cohort studies. This framework provides a basis for discussing independence conditions for LTF and censoring and examining the implications of dependent LTF. We consider failure time and more general life history processes. The joint models are based on multistate processes with expanded state spaces encompassing both the life history and LTF processes. Tracing studies are discussed as a means of investigating the presence of dependent censoring and providing valid estimates of transition intensities and state occupancy probabilities. Simulation studies and an illustration based on a cohort of individuals with systemic lupus erythematosus demonstrate the usefulness and properties of the proposed methods.

Keywords: auxiliary data, independent censoring, life history process, loss to follow-up, multistate model, tracing studies

1 INTRODUCTION

Multistate models are widely used for the analysis of life history processes involving health, employment and many other aspects of human lifetimes (see the works of Andersen et al. (1993), Willekens (2014), Cook and Lawless (2018)). They include as special cases failure time models, competing risks (Beyersmann et al., 2012), and recurrent events (Cook and Lawless, 2007). We consider the analysis of multistate processes that are under continuous observation from some initial time until observation ceases at a right-censoring time.

Standard methods of analysis rely on independence assumptions regarding the process of interest and the censoring time. A discussion of independence conditions for censoring can be found in many books, including those by Kalbfleisch and Prentice (2002, Section 6.2), Aalen et al. (2008, Section

2.2.8) and Andersen et al. (1993, Section 3.2). It is well known that independence assumptions cannot be assessed using only the observed censored data except under conditions which themselves cannot be checked (see the works of Tsiatis (1975), Prentice et al. (1978) and Crowder (1994)). Many authors have proposed joint models that accommodate dependent failure and censoring times (e.g. Fisher and Kanarek (1974), Lagakos and Williams (1978), Slud and Rubinstein (1983), Scharfstein and Robins (2002), Siannis (2011)). Since such models cannot be checked, their primary use has been to investigate the sensitivity of inferences based on independent censoring to violations of this assumption. A few authors (e.g. Lee and Wolfe (1998), Baker et al. (1993), Frangakis and Rubin (2001), Farewell et al. (2003)) have discussed how auxiliary data can enable assessment of the independent-censoring assumption, and facilitate fitting of a failure time model in the presence of dependent censoring. Such auxiliary information is typically obtained through tracing studies in which individuals censored due to loss to follow-up (LTF) are contacted and additional data is acquired.

Our work in this paper is motivated by experiences with observational cohorts of persons with rheumatic diseases and other chronic conditions. For example, two cohort studies from the Centre for Prognosis Studies in Rheumatic Disease at the University of Toronto have motivated us to consider specific issues of LTF. The University of Toronto Lupus Clinic maintains a registry of over 1800 patients diagnosed with systemic lupus erythematosus (SLE), a complex autoimmune disorder in which several organ systems can be affected episodically over the course of the disease. Individuals in the registry are scheduled to make biannual clinic visits, during which disease activity is assessed, medical imaging may be carried out, and treatments may be changed. Vital organs (e.g. brain, heart, lungs) may be affected by the disease (Legge et al., 2016) and strong treatments such as steroids, methotrexate and biologics are routinely prescribed. Mortality rates for individuals with SLE are of interest, as are aspects of disease activity. However, individuals may cease to attend the clinic and thus become lost to follow-up. A question of considerable interest is whether persons lost to follow-up experience similar disease histories and mortality to persons remaining in the cohort and undergoing regular follow-up. Farewell et al. (2003) reported on the results of a tracing study in which effort was made to contact those lost to follow-up and to determine vital status, in order to assess mortality rates; we use data from this study for illustration later in the paper. A second cohort of patients called The University of Toronto Psoriatic Arthritis Cohort is comprised of individuals diagnosed with psoriatic arthritis, another autoimmune disorder characterized by skin and joint involvement. This cohort was established 1976 and is now comprised of over 1800 individuals (Gladman and Chandran, 2011); LTF is also common in this cohort, and there is a concern that disease histories and mortality for individuals lost to follow-up may be different than for persons remaining in the clinic.

We have two main objectives in this paper. The first objective is to provide a framework for joint modeling of life history and LTF processes, which facilitates a careful discussion of dependent LTF and censoring. Models within this framework provide a useful basis for studying the effects of dependent censoring, and for sensitivity analyses for specific data sets. The second objective is to consider tracing studies and other types of auxiliary information that allow independence assumptions to be checked and potentially relaxed. We extend previous work on dependent censoring and LTF by i) considering multistate life history processes that include as special cases failure time, competing risks and recurrent event models; ii) providing a joint modeling framework for the process of interest and LTF, which makes it easy to study underlying issues and to conduct numerical studies and sensitivity analyses; iii) characterizing important types and effects of non-independence; iv) providing new nonparametric estimates of failure time distributions and multistate occupancy probabilities under dependent-censoring conditions when tracing data are available; and v) considering different types of auxiliary data for assessment of independence assumptions. We also point out and discuss complicating issues in studies involving delayed entry. The framework used for joint modeling is

based on multistate models (see, e.g. the works of Andersen et al. (1993), Beyersmann et al. (2012), Willekens (2014), Cook and Lawless (2018)).

The remainder of this article is organized as follows. Section 2 introduces joint models for multistate life history processes and LTF and characterizes independence and the effects of non-independent LTF. Section 3 considers tracing studies, extended follow-up of selected individuals, and auxiliary data that allow assessment of independence assumptions. New nonparametric estimation procedures are also introduced. Section 4 presents numerical studies on the effects of non-independence and the power to detect it using auxiliary data. Section 5 contains an illustration involving a cohort of persons with SLE, and Section 6 has concluding remarks.

2 MULTISTATE MODELS AND CONTINUOUS OBSERVATION PROCESSES

2.1 MULTISTATE PROCESSES AND INDEPENDENT LTF

In social sciences, interest often lies in modeling educational attainment, employment status, and income level over lifetimes of individuals. In public health, it is of interest to model the onset and progression of chronic diseases such as diabetes, arthritis, and dementia. In these and many other settings, the processes involve a set of distinct states that individuals occupy and move among over time. Such processes can be represented using multistate stochastic models; state-space diagrams for several illustrative models are given in Figure 1.
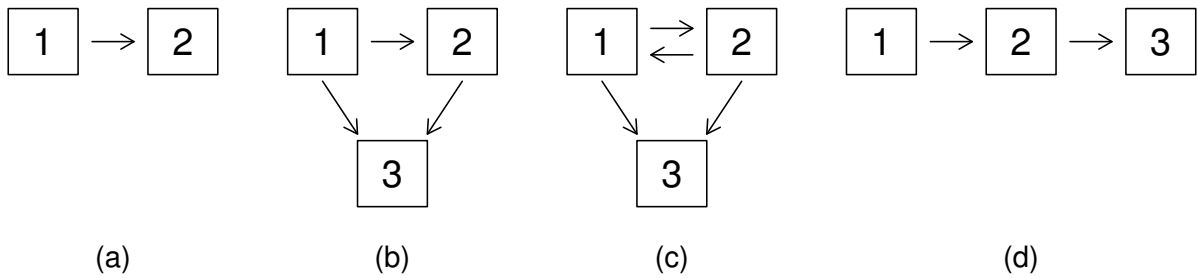

Figure 1: Some multistate processes: (a) failure process, (b) illness-death process, (c) reversible illness-death process, (d) progressive process

We label the states $1, 2, \ldots, K$ for a general K-state process and let $Z(t)$ denote the state occupied by a generic individual at time $t \geq 0$. We consider models for the process $\{Z(t), t \geq 0\}$, which may incorporate information on fixed or time-varying covariates $\{X(t), t \geq 0\}$. We focus on continuous-time multistate models, which can be specified in terms of their transition intensity functions (Andersen et al., 1993)

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

where $H(t^-) = \{Z(s), X(s), 0 \leq s < t\}$ denotes the history of state occupancy and covariate values up to time t. We may also define $\{N(t), t > 0\}$ to be a right-continuous multivariate counting process recording the times and types of all transitions of the multistate process (Andersen et al., 1993). We use bars to denote process histories; hence, let $\bar{N}(t) = \{N(s), 0 < s \leq t\}$ and $\bar{X}(t) = \{X(s), 0 \leq s \leq t\}$. We may alternatively write the process history as $H(t^-) = \{\bar{N}(t^-), \bar{X}(t^-), Z(0)\}$. In what follows, we suppress the notation for the covariates, but, if they are present, assume they are included

in the process history. We let θ denote the full set of parameters specifying the intensities for the multistate process.

We consider LTF and censoring under a continuous observation scheme. We assume for the main development that individuals in a study are all followed from the time origin $t = 0$ for the process in question. In some studies, an individual might enter at time $L > 0$; then, L is a left-truncation (LT) time, and a relevant question is whether L is independent of the life history process (Keiding, 1992; Lakhal-Chaieb et al., 2006). Even if this is the case, however, LT complicates the assessment of LTF. We discuss this further in Section 2.4 and subsequently, but for now, we assume that an individual is observed from $t = 0$ up to an administrative censoring time A, unless they are prematurely lost to follow-up. We assume A is a fixed time that is independent of $\{N(t), t > 0\}$ given the covariates.

An individual may be prematurely lost to follow-up at a random time $C < A$; hence, we define the left-continuous function $C(t) = I(C < t)$ for $t > 0$, the counting process $\{C(t), t \geq 0\}$, and let $\bar{C}(t) = \{C(s), 0 < s \leq t\}$ denote the history of the censoring (or LTF) process. To facilitate a careful discussion of independent censoring, Aalen et al. (2008, Section 2.2.8) consider a typically hypothetical situation in which the process of interest is fully observed regardless of the outcome of the LTF process. To distinguish the history in this setting, we use a script H and let $\mathcal{H}(t) = \{N(s), X(s), C(s), 0 < s \leq t, Z(0), X(0)\}$.

We use $dN(t), dC(t)$ to indicate transition or censoring events at time t, respectively, as per common notation (Aalen et al., 2008). The random LTF or censoring process is said to be (conditionally) independent of the multistate process if

$$P(dN(t) \mid \mathcal{H}(t^-)) = P(dN(t) \mid H(t^-)), \qquad (2)$$

for $0 < t \leq A$; see the works of Kalbfleisch and Prentice (2002, Section 6.2) and Aalen et al. (2008, Section 2.2.8). Equation (2) states that whether a process remains under observation or not at time t the intensities for event occurrence are the same as the intensities for the process in the absence of censoring. More formally, if

$$\lambda_{kl}(t \mid \mathcal{H}(t^-)) = \lim_{\Delta t \downarrow 0} \frac{P(Z(t + \Delta t^-) = l \mid Z(t^-) = k, \mathcal{H}(t^-))}{\Delta t}, \quad k \neq l, \qquad (3)$$

then (2) implies that, for $0 < t \leq A$ and $k \neq l$,

$$\lambda_{kl}(t \mid \mathcal{H}(t^-), C \geq t) = \lambda_{kl}(t \mid H(t^-)) \qquad (4a)$$

$$\lambda_{kl}(t \mid \mathcal{H}(t^-), C = c, C < t) = \lambda_{kl}(t \mid H(t^-)) \qquad (4b)$$

where $\mathcal{H}(t^-)$ and $H(t^-)$ include the information that $Z(t^-) = k$.

If we let $C' = \min(C, A)$, under noninformative censoring (see the work of Kalbfleisch and Prentice (2002, Section 6.2)) and assumption (4a), the partial likelihood based on continuous observation of $Z(t)$ over $(0, C']$ is proportional to

$$L(\theta) = \prod_{k=1}^{K} \prod_{l \neq k=1}^{K} \left\{ \left[\prod_{t_j \in \mathcal{D}_{kl}} \lambda_{kl}(t_j \mid H(t_j^-)) \right] \exp\left[- \int_0^{C'} Y_k(u) \lambda_{kl}(u \mid H(u^-)) \, du \right] \right\}, \qquad (5)$$

where $Y_k(u) = I(Z(u^-) = k)$, and $\mathcal{D}_{kl}$ is the set of times t_j at which the individual makes a k to l transition over $(0, C']$ (see, e.g. the works of Andersen et al. (1993), Kalbfleisch and Prentice (2002,

Section 6.1), Cook and Lawless (2018, Section 2.2)). When independent life history processes are observed for a sample of n independent individuals, partial likelihoods are a product of terms having the form of (5).

Much attention has been given to ways of addressing violations of the condition in (4a). One scenario in which (4a) may be violated arises when individuals are selectively withdrawn from a study according to values of latent or unobserved covariates, so that their distribution among individuals still being followed is different from what it would be in the absence of LTF. If there is sufficient information on auxiliary fixed or time-dependent covariates $\{W(t), t > 0\}$ such that

$$\lambda_{kl}(t \mid \mathcal{H}(t^-), \bar{W}(t^-), C \geq t) = \lambda_{kl}(t \mid H(t^-), \bar{W}(t^-)) \,, \tag{6}$$

such violations can be dealt with by generalizing the intensity model for transitions to include $\bar{W}(t^-)$, as in the right-hand side of (6). This would mean that (4a) is satisfied, but this is achieved at the expense of fitting a model that may not be of primary interest. In particular, certain features in $W(t)$ may be on the causal path of a baseline covariate effect, and conditioning on such features may therefore be undesirable. Use of inverse probability of censoring (IPC) weights (Scharfstein and Robins, 2002; Robins, 1993; Satten et al., 2001) is appealing in such contexts. This involves specifying the intensity for LTF. We note, however, that based on the observed censored data, there is no way to check the validity of the conditional independence assumption in (6). We also note that in some contexts, (4a) might hold, but (4b) might not. For example, this might occur because the environment for persons who become lost to follow-up changes in a way that affects their life history process; we comment further on this later.

When censoring is random the full likelihood is based on data on both the life history and censoring events and so involves the censoring intensity, defined as

$$\alpha(t \mid \mathcal{H}(t^-)) = \lim_{\Delta t \downarrow 0} \frac{P(\Delta C(t) = 1 \mid C(t^-) = 0, \mathcal{H}(t^-))}{\Delta t} \,. \tag{7}$$

If the intensity in (7) does not involve parameters θ specifying the intensities of the life history process, censoring is non-informative, and terms involving censoring intensities can be omitted from the likelihood with no impact on efficiency. However, the censoring process does affect the *amount* of information about θ one can anticipate at the planning stage of a study, and therefore, it must be addressed in study design. Events such as death, which terminate the process, should be represented as absorbing states in the multistate model and, thus, handled differently than premature LTF.

We now specify joint models for the life history and LTF processes. Such models allow an incisive analysis of dependent censoring and the conditions (4a) and (4b). Because of their importance and relative simplicity, we begin with failure time processes and LTF, and then consider general multistate processes.

2.2 FAILURE TIME MODELS

Censoring here corresponds to premature termination of follow-up, but the failure time process is viewed as continuing to evolve for times $t > C$. To deal with independent or dependent censoring, we consider an expanded model including life history states before and after LTF. We introduce a superscript p to distinguish life history states and intensities after (post) LTF. Figure 2 shows the expanded multistate diagram for the failure time process in Figure 1(a); when LTF occurs prior to failure, individuals remain at risk and will eventually fail. We restrict attention here to a setting in which observation ends upon failure, and in the more general multistate setting, upon entry to an

absorbing state, but in some settings, individuals might remain under observation for some reason even after failure has occurred.
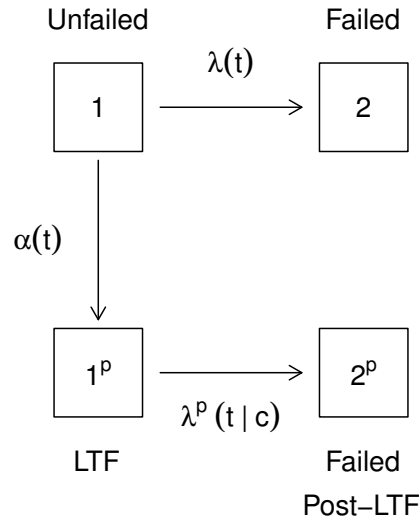

Figure 2: A four-state process $\{\mathcal{Z}(t), t > 0\}$ for failure, premature loss to followup (LTF), and failure post-LTF

We let $\mathcal{Z}(t)$ denote the state occupied in the expanded state space and $\{\mathcal{Z}(t), t > 0\}$ denote the joint failure time and censoring process. Failure time T is defined as the time of entry to state 2 or 2^p, and $N(t) = I(T \leq t)$. The failure intensity post-LTF in Figure 2 is denoted as

$$\lambda^p(t \mid c) = \lim_{\Delta t \downarrow 0} \frac{P(\mathcal{Z}(t + \Delta t^-) = 2^p \mid \mathcal{Z}(t^-) = 1^p, C = c)}{\Delta t} = \lim_{\Delta t \downarrow 0} \frac{P(\Delta N(t) = 1 \mid T \geq t > c, C = c)}{\Delta t},$$

since it may depend on the time C of LTF; we continue to suppress notationally any dependence on covariates. LTF is independent if $\lambda^p(t|c) = \lambda(t)$, though we discuss a weaker condition in Theorem 1.

As mentioned earlier, with standard censored data of the form $\{\min(T, C), \Delta = I(T < C)\}$, transitions from state 1^p to state 2^p are unobserved; thus, we cannot estimate $\lambda^p(t|c)$ nor assess whether censoring is independent. If, however, we are able to trace some individuals who were lost to follow-up and obtain extended follow-up data on them, then this may be possible; we consider such data in Section 3.

Various joint models for failure and censoring that allow for non-independence have been proposed previously. Most of them consider joint densities $f(t, c)$ for (T, C) or equivalent specifications in terms of marginal and conditional hazard functions (see, e.g. the works of Baker et al. (1993), Frangakis and Rubin (2001), Moeschberger and Klein (1995), Scharfstein and Robins (2002), Siannis (2011)). Some earlier authors have implicitly or explicitly considered special cases of our model. Lee and Wolfe (1998) considered a Markov model where $\lambda^p(t|c) = \lambda(t) \exp(\beta I(c < t))$, and Lee and Tsai (2005) considered a semi-Markov model with $\lambda^p(t|c) = \lambda^p(t - c)$; both papers consider tracing data. The Markov model with $\lambda^p(t|c) = \lambda^p(t)$ was examined by Slud and Rubinstein (1983). However, they introduced this in terms of a joint model for (T, C) and did not consider tracing of persons lost to follow-up in order to estimate $\lambda^p(t)$ but, instead, assumed that the ratio $\lambda^p(t)/\lambda(t)$ was a known function and conducted sensitivity analysis. Lee and Wolfe (1998) and Farewell et al. (2003) did not use the model in Figure 2 explicitly but suggested related tests of independent censoring by

using tracing data. The general intensity-based model we consider here has the advantage that all intensities are directly estimable with auxiliary data obtained from tracing studies; the presence and consequences of non-independent LTF are then easily assessed. As we discuss in Section 3, it allows the selection of persons to be traced to depend on data observed up to LTF, and time-varying covariates that may affect failure or LTF are readily incorporated; this is not the case with joint models $f(t, c)$. Finally, the discussion easily extends to LTF for general multistate life history models, as we describe in the next section.

To explore the effect of LTF further, we let $\mathcal{S}_T(t)$ and $\lambda_T(t) = -d \log \mathcal{S}_T(t)/dt$ denote the survivor and hazard functions for the failure time T before or after LTF in Figure 2. Slud and Rubinstein (1983) showed that when the intensity for failure after LTF is independent of the censoring time (i.e. $\lambda^p(t|c) = \lambda^p(t)$), then $\lambda(t) = \lambda^p(t)$ if and only if $\lambda_T(t) = \lambda(t)$. In the more general case, here, we note that

$$\mathcal{S}_T(t) = P(\mathcal{Z}(t) = 1) + P(\mathcal{Z}(t) = 1^p) = e^{-\Lambda(t)-A(t)} + \int_0^t \alpha(s)\, e^{-\Lambda(s)-A(s)}\, e^{-\Lambda^p(s,t)}\, ds\,, \quad (8)$$

where $A(t) = \int_0^t \alpha(s)\, ds$ and $\Lambda^p(s, t) = \int_s^t \lambda^p(u|s)\, du$. Some simple algebra shows that

$$\lambda_T(t) = \lambda(t) \left\{ \frac{\exp(-\Lambda(t) - A(t)) + \int_0^t g(s)\, \lambda^p(t \mid s)\, \lambda^{-1}(t)\, \exp(-\Lambda^p(s,t))\, ds}{\exp(-\Lambda(t) - A(t)) + \int_0^t g(s)\, \exp(-\Lambda^p(s,t))\, ds} \right\}, \quad (9)$$

where, for convenience, we define $g(s) = \alpha(s)\exp(-\Lambda(s) - A(s))$ as the sub-density for the time of LTF (see Appendix APPENDIX A). This immediately gives an extension of Slud and Rubinstein's result:

Theorem 1: $\lambda_T(t) = \lambda(t)$ if and only if $\lambda(t) = \lambda^p(t)$, where

$$\lambda^p(t) = \frac{\int_0^t g(s)\exp(-\Lambda^p(s,t))\lambda^p(t \mid s)\, ds}{\int_0^t g(s)\exp(-\Lambda^p(s,t))\, ds} = \lim_{\Delta t \downarrow 0} \frac{P(\mathcal{Z}(t + \Delta t^-) = 2^p \mid \mathcal{Z}(t^-) = 1^p)}{\Delta t}.$$

is the marginal failure rate following LTF, obtained by marginalizing $\lambda^p(t|c)$ over the LTF time.

The result that $\lambda_T(t) = \lambda(t)$ states that the distribution of the failure time T is the same in the absence of LTF (i.e. when $\alpha(t) = 0$) as it is when there is (possibly dependent) LTF. However, neither $\lambda_T(t) = \lambda(t)$ nor $\lambda^p(t) = \lambda(t)$ seem to imply that $\lambda^p(t|s) = \lambda(t)$ for all s and $t \geq s$.

Example 1: Consider the setting portrayed in Figure 3 in which an unobserved random factor U acts multiplicatively on time-homogeneous intensities for both failure and LTF; we assume LTF is conditionally independent given $U = u$. While the assumption that U affects all transitions in exactly the same way is implausible for most applications, we consider it for illustrative purposes as it allows easy calculation of intensities and related quantities for the "observable" intensities represented in Figure 2.

The conditional transition intensities in Figure 3 for failure before and after LTF are the same given $U = u$, but the marginal intensities obtained by averaging over u are different. To see this suppose U is gamma distributed with $E(U) = 1$ and $\mathrm{Var}(U) = \phi$. Straightforward calculations in
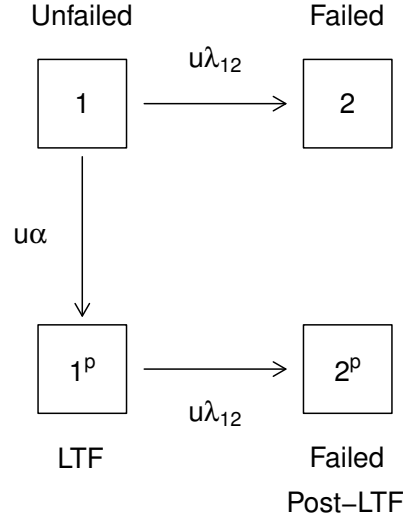
Figure 3: A multiplicative random effects model for censoring and failure

Appendix APPENDIX B show that the marginal intensities in Figure 2 are

$$\alpha(t) = \lambda_{11^p}(t \mid \mathcal{Z}(t^-) = 1) = \frac{\alpha}{1 + \phi(\lambda_{12} + \alpha)t}, \tag{10}$$

$$\lambda(t) = \lambda_{12}(t \mid \mathcal{Z}(t^-) = 1) = \frac{\lambda_{12}}{1 + \phi(\lambda_{12} + \alpha)t}, \tag{11}$$

and

$$\lambda^p(t \mid c) = \lambda^p_{12}(t \mid C = c, \mathcal{Z}(t^-) = 1^p, \mathcal{H}(t^-)) = \frac{(1 + \phi)\lambda_{12}}{1 + \phi(\alpha c + \lambda_{12}t)}. \tag{12}$$

We find that $\lambda^p(t)$ takes a complicated form (see Appendix APPENDIX C), and $\lambda^p(t) = \lambda(t)$ if and only if $\phi = 0$ and U is identically one, in which case both equal λ_{12}. We also note that, for a given $\phi > 0$, the magnitude of $\alpha > 0$ determines how different $\lambda^p(t|c)$ and $\lambda(t)$ are.

In the absence of censoring (i.e. when $\alpha = 0$), the hazard function for the failure time is

$$\lambda^\circ(t) = \frac{\lambda_{12}}{1 + \phi\lambda_{12}t},$$

from (11), which defines $\lambda_{12}(t|H(t^-))$ in (4a) and (4b). The survivor function estimand in this case is $S^\circ(t) = \exp(-\Lambda^\circ(t))$, where $\Lambda^\circ(t) = \int_0^t \lambda^\circ(s)ds$. Under non-independent censoring (i.e. when $\alpha > 0$, $\phi > 0$), standard methods based on the assumption of independent censoring estimate $\lambda(t) = \lambda^\circ(t)\{(1 + \phi\lambda_{12}t)/(1 + \phi(\lambda_{12} + \alpha)t)\}$ rather than $\lambda^\circ(t)$; this is a consequence of a violation of (4a).

We can, of course, estimate the three parameters λ_{12}, α, ϕ in the dependent censoring model considered here from the right-censored observations $\{\min(T, C), \Delta = I(\min(T, C) = T)\}$. If $\widehat{\phi}$ were close to zero, one might take this as supporting an assumption of independent censoring, but this would be risky, as would over-emphasis of sensitivity analysis based on the model, because the assumption concerning equality of $\lambda_{12}(t|\mathcal{Z}(t^-) = 1, u)$ and $\lambda^p_{12}(t|\mathcal{Z}(t^-) = 1^p, u)$ cannot be assessed.

We consider auxiliary information that allows this, based on observation of individuals after they become lost to follow-up, in Section 3. Finally, we note that $\lambda(t)$, which can be estimated from the standard right-censored data, may be of interest; it is the failure intensity for individuals who are still under follow-up in a study.

2.3   GENERAL MULTISTATE MODELS

Loss to followup can be incorporated similarly into more general multistate models. Figure 4 depicts a model that combines LTF information with the illness-death process in Figure 1(b); a similar model can be described for the joint process $\{Z(s), C(s), s > 0\}$ for any life history process. We label states $1, 2, \ldots, K$ for the process of interest before LTF and label the corresponding states $1^p, 2^p, \ldots, K^p$ post-LTF. We again use $\mathcal{Z}(t)$ to denote the state occupied in the joint model for the life history and LTF process. Let $H(t^-)$ represent the history of the process $\{Z(t), t > 0\}$ and covariates as before, and let $\mathcal{H}(t^-) = \{\mathcal{Z}(s), 0 < s < t\} = (H(t^-), \bar{C}(t^-))$ denote the expanded "full" history that includes information on LTF and the life history process beyond LTF. For the processes in Figure 4 censoring is independent if $\lambda_{rs}^p(t|\mathcal{H}(t^-)) = \lambda_{rs}(t|H(t^-))$ for $(rs) = (12), (13)$ and (23). The intensities $\alpha_1(t|\mathcal{H}(t^-))$ and $\alpha_2(t|\mathcal{H}(t^-))$ are not needed for estimation of the life history process intensities when censoring is independent. As in the preceding discussion of failure time models, under dependent censoring there is a distinction between marginal properties of the joint process such as $P(\mathcal{Z}(t) = 3$ or $3^p)$ or the distribution of the time of entry to state 3 or 3^p, and properties such as $P(Z(t) = 3)$ or the distribution of time T_3 to entry to state 3 in a setting where premature LTF cannot occur. Tracing of individuals LTF and other types of auxiliary information for the assessment of dependent censoring are considered in subsequent sections.
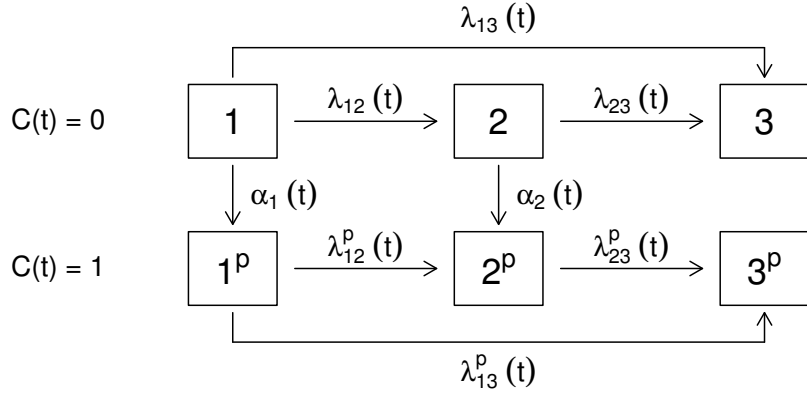

Figure 4: A combined illness-death and loss to followup model for $\{\mathcal{Z}(t), t > 0\} = \{Z(t), C(t), t > 0\}$. Notationally, $\lambda_{kl}(t)$ represents $\lambda_{kl}(t|H(t^-), C \geq t)$, $\alpha_k(t)$ represents $\alpha_k(t|H(t^-), C \geq t)$ and $\lambda_{kl}^p(t)$ represents $\lambda_{kl}(t|H(t^-), C = c < t)$

For simplicity, we assume that $Z(0) = 1$ with probability one, but the results below apply if instead we specify a distribution for $Z(0)$. A result analogous to Theorem 1 holds for general multistate models and is expressed in terms of transition rates (as opposed to transition intensities) and state occupancy probabilities. Unlike the intensities (1), transition rates are conditional only on the current state and are defined as

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

which we write as $q_{kl}(t)$ for $k, l = 1, \ldots, K$ ($l \neq k$) and $q_{kl}^p(t)$ for $k, l = 1^p, \ldots, K^p$ ($l \neq k$); we also let

$$a_k(t) = \lim_{\Delta t \downarrow 0} \frac{P(\mathcal{Z}(t + \Delta t^-) = k^p \mid \mathcal{Z}(t^-) = k)}{\Delta t} , \quad k = 1, \ldots, K .$$

Transition rates may be defined for any multistate process. For Markov models the transition intensities and transition rates coincide, but in the following development the processes are not assumed to be Markov.

Let $\mathcal{P}_k(t) = P(\mathcal{Z}(t) = k)$, $\mathcal{P}_{k^p}(t) = P(\mathcal{Z}(t) = k^p)$ and $\mathcal{P}_k^{\bullet}(t) = \mathcal{P}_k(t) + \mathcal{P}_{k^p}(t)$ for $k = 1, \ldots, K$. To $o(\Delta t)$, we have

$$\mathcal{P}_l^{\bullet}(t + \Delta t) = \sum_{k \neq l} \{\mathcal{P}_k(t) q_{kl}(t) \Delta t + \mathcal{P}_{k^p}(t) q_{kl}^p(t) \Delta t\}$$

$$+ \mathcal{P}_l(t)\{1 - q_{l\cdot}(t)\Delta t - a_l(t)\Delta t\} + \mathcal{P}_l(t) a_l(t) \Delta t + \mathcal{P}_{l^p}(t)\{1 - q_{l\cdot}^p(t)\Delta t\}$$

$$= \mathcal{P}_l^{\bullet}(t) + \sum_{k=1}^K \{\mathcal{P}_k(t) \mathcal{Q}_{kl}(t) + \mathcal{P}_{k^p}(t) \mathcal{Q}_{kl}^p(t)\} \Delta t$$

$$= \mathcal{P}_l^{\bullet}(t) + \sum_{k=1}^K \mathcal{P}_k^{\bullet}(t) \mathcal{Q}_{kl}(t) \Delta t + \sum_{k=1}^K \mathcal{P}_{k^p}(t)\{\mathcal{Q}_{kl}^p(t) - \mathcal{Q}_{kl}(t)\} \Delta t \tag{13}$$

where $\mathcal{Q}_{kl}(t) = q_{kl}(t)$ and $\mathcal{Q}_{kl}^p(t) = q_{kl}^p(t)$ for $l = 1, \ldots, K$ ($l \neq k$), and $\mathcal{Q}_{kk}(t) = -q_{k\cdot}(t) - \alpha_k(t)$ and $\mathcal{Q}_{kk}^p(t) = -q_{k\cdot}^p(t)$, $k = 1, \ldots, K$. Note that if $P(Z(t) = l)$ is the occupancy probability for state l when there is no LTF, then these probabilities satisfy

$$P_l(t + \Delta t) = P_l(t) + \sum_{k=1}^K P_k(t) Q_{kl}(t) \Delta t , \tag{14}$$

where $Q_{kl}(t) = q_{kl}(t)$ for $l = 1, \ldots, K$ ($l \neq k$) and $Q_{kk}(t) = -q_{k\cdot}(t)$, $k = 1, \ldots, K$.

Thus, by comparison of (13) and (14) and the fact that $\mathcal{P}_l^{\bullet}(0) = P_l(0) = I(l = 1)$, we see that if $\mathcal{Q}_{kl}^p(t) = \mathcal{Q}_{kl}(t)$ for all k, l and t, then $\mathcal{P}_l^{\bullet}(t) = P_l(t)$ for all l and $t \geq 0$. Conversely, if this result holds, then we must have

$$\sum_{k=1}^K \mathcal{P}_{k^p}(t)\{\mathcal{Q}_{kl}^p(t) - \mathcal{Q}_{kl}(t)\} = 0 \tag{15}$$

for all l and $t > 0$. It is not clear whether this implies that $Q_{kl}^p(t) = Q_{kl}(t)$ for all k, l and $t > 0$.

As in the failure time case, whether censoring can be considered as independent depends on the life history model of interest. In some settings we consider marginal or partially conditional models: for example, instead of intensity-based models (7) we might consider a marginal model $P_k(t|x) = P(Z(t) = k|x)$ that conditions only on fixed covariates. If there are other factors W in the process history $H(t)$ that affect both LTF and the $Z(t)$ process, then the outcome $\tilde{Y}(t) = I(Z(t) = k)$ is not independent of the event $C \geq t$, conditional only on X. In such cases, inverse-probability-of-censoring (IPC) weighted estimating functions can be used to estimate $P(t|x)$ (e.g. Datta and Satten (2002), Scheike and Zhang (2007), Cook et al. (2009)). Such methods require that the censoring process be modeled and estimated. Once again, such approaches still rely on the assumption of conditionally independent censoring given the full process history $H(t)$.

2.4   Delayed Entry and Left-Truncation

In some studies enrolment and followup of individuals may begin at delayed entry times $L_i > 0$; the L_i are also referred to as left-truncation (LT) times. Delayed entry is said to be independent if, in the notation of Section 2.2, $P(dN(t)|H(t^-)) = P(dN(t)|H(t^-), L < t)$; see for example Andersen et al. (1993, p. 196), Aalen et al. (2008, p. 32), Keiding (1992). In that case estimation proceeds as usual, but with likelihood contributions for individual i based on $P(dN_i(t), t > L_i|H(L_i))$. A practical difficulty is that, depending on the family of models, we may not know all the relevant information in $H(L_i)$. For Markov models this is not a problem, since all we need is the state $Z(L_i)$ at entry. A second remark is that independent LT should not be assumed lightly. In many settings, including the one we consider in Section 5, life history process intensities for $t > L_i$ may depend on L_i.

Delayed entry complicates the the consideration of dependent loss to followup. For simplicity we focus on the failure time setting in the following discussion. The LTF process begins at the time L of study entry, so we now denote the LTF intensity for individual i as $\alpha(t|L_i), t > L_i$. In many cases it is most natural to model this as a function of time time since entry $(t - L_i)$, but it might also depend on t and L_i in a more complicated way. In terms of the joint process for life history and LTF, as exemplified in Figures 2 and 4, left-truncation is independent if the process intensities do not depend on L_i for $t > L_i$. For the failure time setting this means that $\lambda(t|L_i)$, $\alpha(t|L_i)$ and $\lambda^p(t|C, L_i)$ for are the same for $t > L_i$ as the intensities $\lambda(t)$, $\alpha(t)$, $\lambda^p(t|C)$ that apply when there is no delayed entry $(L_i = 0)$. This applies for $\alpha(t|L_i)$ only when it is of Markov form, but the important question is whether it is plausible for the two failure time intensities.

The random effects model in Example 1 illustrates these issues in a setting where differences in pre - and post - LTF transition intensities arise from unobserved factors that affect both the life history and LTF processes. Even if L is a random variable that is completely independent of the random variable U and the life history process conditional on U, the state $Z(L)$ occupied at time L affects the distribution of U that applies to an individual. For the model in Figure 2, we find for example that when U in the population has a gamma distribution with mean 1 and variance ϕ, then

$$E(U \mid L, Z(L) = 1, \mathcal{Z}(t^-) = 1) = \{1 + \phi(\lambda t + \alpha(t - L))\}^{-1}$$

and

$$\lambda(t \mid L) = \lambda_{12}(t \mid L, Z(L) = 1, \mathcal{Z}(t^-) = 1) = \frac{\lambda}{1 + \phi(\lambda t + \alpha(t - L))}, \quad t > L\,.$$

where $A(t_i|L_i) = \int_{L_i}^{t_i} \alpha(t|L_i)\, dt$. This differs from (11), the intensity that applies when $L = 0$, so the random variable U induces non-independent left-truncation. The intensities for LTF and failure post-LTF similarly depend on L.

If we are convinced that LT is indeed independent then estimation of $\lambda(t)$, $\alpha(t)$ and $\lambda^p(t|c)$ in Figure 2, and the intensities in more general processes, can proceed as described throughout the paper. However, marginal rate functions such as $\lambda^p(t)$ involve averaging over both L and C with $L < C$, and therefore have a different interpretation than they do in the absence of delayed entry. This may have a descriptive interpretation for a given study, but that interpretation is tied to the delayed entry and LTF processes of the study and not just the life history process. The Aalen-Johansen estimates that we describe in Section 3.3 will likewise have an interpretation specific to the delayed entry and LTF processes of the study and are inconsistent for the population life history process quantities.

When a study has delayed entry, results such as those discussed above for the model of Example 1 suggest that it is wise to consider the possible dependence of all the transition intensities in Figures 2

and 4 on entry time L. The methods of estimation and the numerical investigations in the next two sections assume, except when indicated, that there is no delayed entry. In the illustration of Section 5 we have delayed entry and we discuss there how it can be handled.

3 TRACING AND EXTENDED FOLLOW-UP

Consider a study in which observation of an individual i begins at $t = 0$ but may be terminated by premature LTF at some time $C_i < \min(T_i, A_i)$, where A_i is an administrative censoring time independent of the life history process and T_i is the time of entry to an absorbing state, if they exist. The standard right-censored data consist of $\{N_i(s), 0 \le s \le t_i = \min(C_i, T_i, A_i)\}$ and $\Delta_i^c = I(t_i = C_i)$ along with covariates X_i. To assess the independent-LTF assumption, we require auxiliary data; this may also allow us to understand sources of dependent LTF and, in some cases, to estimate process features in the presence of dependent LTF. Auxiliary data could include (i) information on $\{N_i(s), C_i < s \le A_i\}$ for persons lost to follow-up, (ii) information on additional covariates $\{W_i(s), s > 0\}$ that may be related to both the life history and LTF processes, and (iii) information on the reasons a person became lost to follow-up. In the cases of (i) and (iii), such information could be obtained by tracing persons following LTF. For (ii), data could be collected on a subsample of individuals designated at the start of a study, if it were deemed too expensive to collect it on everyone.

We focus on auxiliary information from extended follow-up and briefly consider auxiliary covariate data in Section 3.4. By tracing and extended follow-up, we mean obtaining information for the time interval (C_i, A_i) on persons who are prematurely lost to follow-up at time C_i. It is usually feasible to do this only for a subset of persons lost to follow-up, and we therefore define R_i = I(person i is traced and provides extended follow-up). In practice, some individuals selected for tracing may not be found or may refuse to provide auxiliary information; we let $R_i = 1$ indicate individuals on whom auxiliary information is obtained. We assume, for simplicity, that the administrative censoring time A_i is independent of the joint life history, covariate, and LTF process history $\mathcal{H}_i(0, \infty) = (H_i(0, \infty), \bar{C}_i(0, \infty))$ and that, for persons lost to follow-up, R_i is conditionally independent of $\mathcal{H}_i(C_i^+, \infty)$, given C_i and $\mathcal{H}_i(0, C_i)$, where $\mathcal{H}_i(s, t)$ denotes the process history over $[s, t]$.

We denote the probabilities of extended followup as

$$\pi_i = P(R_i = 1 \mid t_i, \Delta_i^c, H_i(0, t_i)), \tag{16}$$

where $\pi_i = 0$ if $\Delta_i^c = 0$ and $0 \le \pi_i \le 1$ otherwise. This is a type of two-phase study (Whittemore, 1997; Zeng and Lin, 2014) where individuals with $R_i = 1$ make up the phase 2 subsample.

The selection and extended followup of individuals can be done in different ways, with the key requirement being the conditional independence condition giving (16). For example, at the beginning of a study, we might identify a random subsample of individuals who will undergo extended follow-up if needed (Baker et al., 1993). Another design involves extended follow-up on randomly selected individuals with $\Delta_i^c = 1$, possibly based on C_i and $H_i(0, C_i)$; in this case, individuals are usually identified at some time after LTF.

Under the tracing schemes described, the likelihood function for the joint life history and LTF processes, conditional on covariates X_i, is proportional to

$$L(\theta) = \prod_{i=1}^{n} P(H_i(0, t_i)) \, P(H_i^{obs}(C_i^+, A_i) \mid H_i(0, C_i))^{R_i \Delta_i^c}, \tag{17}$$

where $H_i^{obs}(C_i^+, A_i)$ is the information observed from extended follow-up and θ represents the parameters in the joint model. If complete information $H_i(C_i^+, A_i)$ on transitions can be obtained, we could, in principle, estimate all intensities or rate functions in the joint model. However, nonparametric estimation would, as usual, be over restricted ranges for specific cumulative intensities or rates, depending on the minimum time that the initial state was entered.

3.1 FAILURE TIME MODELS

We first consider failure time processes with the joint model portrayed in Figure 2; intensities may depend on covariates, but for convenience, we continue to suppress this notationally. We let T_i denote the time of failure (entry to state 2 or 2^p) and C_i the time of censoring (entry to state 1^p). Standard follow-up ends at $t_i = \min(T_i, C_i, A_i)$; when $C_i < \min(T_i, A_i)$, extended follow-up yields data over $(C_i, \min(T_i, A_i))$. We consider the case where failure times occurring in (C_i^+, A_i) can be obtained for persons who are traced.

In addition to the notation $t_i = \min(T_i, C_i, A_i)$, we define $\Delta_i = I(t_i = T_i)$ and $\Delta_i^c = I(t_i = C_i)$, and if $\Delta_i^c = 1$ and $R_i = 1$, we also define $t_i^p = \min(T_i, A_i)$ and $\Delta_i^p = I(t_i^p = T_i)$. The likelihood contribution for an individual then takes the form

$$L_i(\theta) = \lambda(t_i)^{\Delta_i} \alpha(t_i)^{\Delta_i^c} \exp(-\Lambda(t_i) - A(t_i)) \left[\lambda^p(t_i^p|c_i)^{\Delta_i^p} \exp(-\Lambda^p(c_i, t_i^p)) \right]^{\Delta_i^c R_i}, \qquad (18)$$

where $\Lambda^p(c, t) = \int_c^t \lambda^p(u|c)du$, as in Section 2.2. Note that the term in square brackets in (18) appears only for individuals who are lost to follow-up and then traced. Maximum likelihood estimates for fully parametric models are readily obtained by maximizing the log-likelihood $\ell(\theta) = \sum_{i=1}^n \log L_i(\theta)$.

Nonparametric estimation of $\lambda(t)$ and $\alpha(t)$ (or $\Lambda(t)$ and $A(t)$) when covariates are absent is also straightforward, but for $\lambda^p(t|c)$, we require some type of restriction when C is continuous. One approach is to stratify on C by assuming that $\lambda^p(t|c)$ is of the form

$$\lambda^p(t \mid c) = \sum_{g=1}^{G} \lambda_g^p(t) \, I(b_{g-1} \le c < b_g), \qquad (19)$$

where $G \ge 1$ and $0 = b_0 < b_1 < \cdots < b_G = \infty$ are specified cut-points. It then follows (see Appendix APPENDIX D) that discrete nonparametric estimates of intensities are

$$\widehat{\lambda}(t) = \sum_{i=1}^n \Delta_i \, I(t_i = t) \Big/ \sum_{i=1}^n I(t_i \ge t)$$

$$\widehat{\alpha}(t) = \sum_{i=1}^n \Delta_i^c \, I(t_i = t) \Big/ \sum_{i=1}^n I(t_i \ge t)$$

$$\widehat{\lambda}_g^p(t) = \sum_{i=1}^n R_i \, \Delta_i^p \, \Delta_i^c \, I_g(C_i) \, I(C_i \le t) \, I(t_i^p = t) \Big/ \sum_{i=1}^n R_i \, \Delta_i^c \, I_g(C_i) \, I(C_i \le t) \, I(t_i^p \ge t)$$

where, for convenience, we define $I_g(C_i) = I(b_{g-1} \le C_i < b_g)$ for individuals with $\Delta_i^c = 1$.

Independent LTF can be assessed by comparing estimates of $\lambda(t)$ and $\lambda^p(t|c)$. This is best done by comparison of $\widehat{\Lambda}(c, t)$ and $\widehat{\Lambda}^p(c, t)$ for selected values of c; note that $\widehat{\Lambda}(c, t) = \widehat{\Lambda}(t) - \widehat{\Lambda}(c)$ and that, for the model (19), for $c \in [b_{g-1}, b_g)$,

$$\widehat{\Lambda}^p(c, t) = \widehat{\Lambda}_g^p(c, t) = \int_c^t d\widehat{\Lambda}_g^p(s), \qquad (20)$$

where $d\widehat{\Lambda}_g^p(s) = \widehat{\lambda}_g^p(s)$ is non-zero only at times $s = t_i$ for which $\Delta_i^c = \Delta_i^p = 1$ and $C_i \in [b_{g-1}, b_g)$. Nonparametric estimates described above are useful for sufficiently large data sets, but a good alternative is often to consider models for which $\lambda(t)$ and $\lambda_g^p(t)$ are piecewise constant.

Other models for $\lambda^p(t|c)$ can be considered. For example, we could take $\lambda^p(t|c) = \lambda_0^p(t)e^{\beta g(c,t)}$ for $t > c$, where $g(c, t)$ is a specified function. Farewell et al. (2003) considered $g(c, t) = I(t - c \geq \delta)$, where δ is a specified time, with $\lambda_0^p(t) = \lambda(t)$; Lee and Wolfe (1998) considered $g(c|t) = I(t > c)$ with $\lambda_0^p(t) = \lambda(t)$. Models with nonparametric $\lambda_0^p(t)$ can be fitted using Cox model software. Interestingly, this is even the case when $\lambda_0^p(t) = \lambda(t)$; see the work of Farewell et al. (2003), who note that this follows from the fact that the Cox partial likelihood estimates for β and $\lambda(t)$ can also be obtained via semiparametric maximum likelihood. The models are conveniently fitted by defining the time-varying covariate $w_i(t) = I(C_i \leq t)g(C_i, t)$, and considering the contribution over the time period $(0, t_i = \min(T_i, C_i, A_i))$ for individual i and, if $t_i = C_i$ and $R_i = 1$, the subsequent contribution over the time period (C_i, t_i^p), where $t_i^p = \min(T_i, A_i)$.

When there is independent delayed entry as discussed in Section 2.4, the estimation of $\lambda(t)$ and $\lambda^p(t|c)$ can proceed in a similar way. In the failure time case, the leading terms in (18) would then be replaced by

$$\lambda(t_i)^{\Delta_i}\alpha(t_i|L_i)^{\Delta_i^c} \exp(-\Lambda(t_i) + \Lambda(L_i) - A(t_i|L_i)),$$

where $A(t_i|L_i) = \int_{L_i}^{t_i} \alpha(t|L_i) \, dt$. However, as discussed in Section 2.4, we recommend that models in which the failure intensities may depend on the L_i be considered, so that independence can be assessed.

Estimates for specific intensities can be obtained from survival and event history analysis software, with data displayed in standard counting process data frames (see the works of Cook and Lawless (2007, Appendix C1), Cook and Lawless (2018)). For example, the nonparametric estimates above can be obtained using the `etm` package in R (Allignol et al., 2011) or the R/SPLUS function `coxph` (Therneau and Grambsch, 2000). Parametric models with piecewise-constant intensities and other forms can be fitted using the `phreg` function in the R `eha` package (Broström, 2016). An illustrative data frame is given in Appendix APPENDIX E.

Another comparison of interest is of estimates of the survivor function $\mathcal{S}_T(t)$ in (8) and $\widehat{S}(t) = \exp(-\widehat{\Lambda}(t))$, which is consistent for $S_T(t)$ under independent censoring. IPC weights can also be used to compute a weighted estimate $\widehat{S}^w(t)$ to account for the possibility of selective LTF, if relevant covariates are present. Estimation of $\mathcal{S}_T(t)$ under possibly dependent LTF can be achieved under extended follow-up by inserting estimates into (8). With the model in (19), we obtain

$$\widehat{S}_T(t) = e^{-\widehat{\Lambda}(t)-\widehat{A}(t)} + \sum_{i=1}^{n}\sum_{g=1}^{G} I(C_i \leq t) \frac{\Delta_i^c I_g(C_i)}{n(C_i)} e^{-\widehat{\Lambda}(C_i)-\widehat{A}(C_i)-\widehat{\Lambda}_g^p(C_i,t)} \tag{21}$$

where estimates of $\Lambda(t)$, $A(t)$ and $\Lambda_g^p(c, t)$ are sums of estimates $\widehat{\lambda}(t)$, $\widehat{\alpha}(t)$ and $\widehat{\lambda}_g^p(t)$, respectively. For example, $\widehat{\Lambda}(t) = \sum_{u \leq t} \widehat{\lambda}(u)$. The special case with $G = 1$ is the setting with $\lambda^p(t|c) = \lambda^p(t)$ considered by earlier authors. We note that the estimate of $S_T(t)$ in (21) or others based on specific models for $\lambda^p(t|c)$ are susceptible to model misspecification. In Section 3.3, we develop an estimator that is consistent in general.

If extended follow-up can provide only failure status $I(T_i \leq A_i)$, and not exact times of failure, then nonparametric estimation following LTF is much more challenging, and weakly parametric assumptions are more convenient. We consider this situation briefly in the following section.

3.2 GENERAL MULTISTATE MODELS

We consider the framework of Section 2.3 and suppose that persons with $C_i < A_i$ may be randomly chosen for extended follow-up, possibly based on C_i, A_i and their observed history at C_i; selection probabilities π_i are as defined in (16). The likelihood (17) can be used to fit models that allow differences in the transition probabilities $\lambda_{rs}(t|H(t), C \geq t)$ and $\lambda_{rs}(t|H(t), C < t)$, denoted $\lambda_{rs}^p(t|\mathcal{H}(t^-))$ in Section 2.3. The crucial issues are the choice of models for the transition intensities and the completeness of the information on $\mathcal{H}(C^+, A)$ that is obtained. Modulated Markov and semi-Markov forms are convenient and flexible, and can be fitted using software such as phreg in R for parametric models or coxph for semi-parametric models. Data for persons with either normal or extended follow-up can be organized in standard multivariate counting process data frames (Cook and Lawless, 2018).

If complete information on $\mathcal{H}(C^+, A)$ cannot be obtained, we are generally forced to make parametric assumptions that cannot be fully checked. In the most extreme case where only $Z(A)$ can be ascertained, we usually must rely on constant or piecewise-constant transition intensities. This allows rudimentary checks on the independent-LTF assumption. Another approach that does not depend on assumed models post-LTF is to consider whether the multinomial outcomes $\mathcal{Z}_i(A_i)$, given $\mathcal{Z}_i(C_i)$, are consistent with the model fitted under the independent-LTF assumption. This may be informal, given that the multinomial categories and their probabilities vary across individuals. In the case of failure time models, this is simpler and involves considering the outcomes $\mathcal{Z}_i(A_i)$ relative to estimated probabilities $\widehat{P}(\mathcal{Z}_i(A_i) = 2^p|C_i, \mathcal{Z}_i(C_i) = 1^p)$.

Given that $H_i^{obs}(C_i^+, A_i)$ in (17) may be a highly coarsened version of $H_i(C_i^+, A_i)$ for a multistate process, there can be an increased risk that the success of tracing might depend on missing process history. If reasonably complete and accurate information on $H_i(C_i^+, A_i)$ cannot be obtained, then it is judicious to include covariates in the analysis that could be related to both the life history process and the success of tracing, but some risk of bias will remain.

3.2.1 ILLUSTRATION: AN ILLNESS-DEATH MODEL

Consider the model in Figure 4, which portrays an illness-death process jointly with LTF. We assume that Markov intensities are appropriate for transitions prior to LTF. For transitions after LTF, the censoring time C_i is both a delayed entry time and a potential covariate. If no covariates X_i are being considered, two possible approaches for intensities post-LTF are (i) to obtain nonparametric estimates, with stratification on C_i as in Section 3.1, and (ii) to consider modulated Markov models with $\lambda_{rs}^p(t|c_i) = \lambda_{0rs}^p(t)\exp(\beta_{rs}g(t, c_i))$, where $t > c_i$ and $g(t, c_i)$ is a specified function. Nonparametric estimates can be obtained using software such as the R etm package, and models (ii) can be fitted with software for multiplicative intensity modeling such as coxph or phreg. The approach can also be employed when covariates X_i are present. In the case of coxph in R/SPLUS, it is necessary to make $g(t, c_i)$ piecewise constant over $t > c_i$ for full flexibility in terms of output. Some defined time-dependent functions $g(t, c)$ can be handled by the tt() option, but estimates of baseline cumulative hazard functions are not provided. Useful models are ones with $g(t, c_i)$ either a vector of indicators for ranges of c_i (as in Section 3.1) or for ranges of $t - c_i$. Estimated transition intensities $\widehat{\lambda}_{rs}(t)$ and $\widehat{\lambda}_{rs}^p(t|c)$ can be compared, as can $\widehat{\Lambda}_{rs}(c, t)$ and $\widehat{\Lambda}_{rs}^p(c, t)$ for $t > c$. We can also compare estimates of state occupancy probabilities $P(Z(t) = r)$, $r = 1, 2, 3$ based on the ordinary censored data with LTF assumed independent and estimates of probabilities $P(\mathcal{Z}(t) = r$ or $r^p)$, $r = 1, 2, 3$ based on the data with extended follow-up. As in the failure time setting, estimates based on models for $\lambda_{rs}^p(t|c)$ are susceptible to misspecification. In the next section, we develop generally consistent nonparametric estimates for transition rates and state occupancy probabilities. Appendix APPENDIX E contains

a data frame involving three individuals with different observed outcomes, which can be used to fit models with `etm`, `phreg` or `coxph`.

3.3 NONPARAMETRIC ESTIMATION OF MARGINAL TRANSITION RATES AND SURVIVAL OR OCCUPANCY PROBABILITIES

For a general Markov multistate model, the transition intensities (1) take the form $q_{kl}(t)$ for $k \neq l$, and the $K \times K$ transition intensity matrix $Q(t)$ has entries $q_{kl}(t)$, $k \neq l$ and $q_{kk}(t) = -\sum_{l \neq k} q_{kl}(t)$, $k = 1, \ldots, K$ (Cook and Lawless, 2018). Nonparametric estimates of the integrals $Q_{kl}(t) = \int_0^t q_{kl}(u)du$ are given under independent LTF by Nelson-Aalen (NA) estimators $\widehat{Q}_{kl}(t)$, and corresponding Aalen-Johansen (AJ) estimators of transition probabilities are given by

$$\widehat{P}(0,t) = \prod_{(0,t]}\{I + d\widehat{Q}(u)\}, \tag{22}$$

where $P(0,t)$ contains transition probabilities $P_{kl}(0,t) = P(Z(t) = k|Z(0) = l)$ for $k, l = 1, \ldots, K$ (see the works of Andersen et al. (1993, Sections 4.1 and 4.4) and Cook and Lawless (2018, Section 3.2)). Furthermore, for non-Markov models, the first row of (22) gives robust estimates of transition probabilities $P_{1k}(Z(t) = k|Z(0) = 1)$, provided LTF is independent of the multistate process (see the work of Cook and Lawless (2018, Section 3.4)). If LTF is process dependent, inverse probability weighting can be used to provide consistent estimates. We now show how to extend these procedures in order to estimate integrated transition rates $Q_{kl}(t)$, $Q_{kl}^p(t)$ and $A_k(t) = \int_0^t a_k(u)du$ for the joint models described in Section 2.2, and state occupancy probabilities $P_{1k}(0,t)$ or $P_{1k^p}(0,t)$. In view of the discussion in Section 2.4, it is assumed here that all individuals are followed from $t = 0$.

Let $\{N_{kl}(u), u > 0\}$ be the counting process for transitions from life history state k to life history state l ($k \neq l$), whether this occurs before or after premature LTF. Note that, using the notation in Section 2.2, we have $E\{dN_{kl}(u)|Z(u^-) = k, C \geq u\} = dQ_{kl}(u)$, $E\{dN_{kl}(u)|Z(u^-) = k, C < u\} = dQ_{kl}^p(u)$ and $E\{dC(u)|Z(u^-) = k, C \geq u\} = dA_k(u)$. Given a sample of n independent individuals we introduce a subscript i for individual i, $i = 1, \ldots, n$. Then, letting $Y_{ik}(u) = I(Z_i(u) = k)$, we see that the following estimating functions have expectation zero, for $u > 0$:

$$G_1(u) = \sum_{i=1}^n I(C_i \geq u)Y_{ik}(u^-)\{dN_{ikl}(u) - dQ_{kl}(u)\} \tag{23}$$

$$G_2(u) = \sum_{i=1}^n I(C_i \geq u)Y_{ik}(u^-)\{dC_i(u) - dA_k(u)\} \tag{24}$$

$$G_3(u) = \sum_{i=1}^n \frac{R_i I(C_i < u)Y_{ik}(u^-)}{\pi_i}\{dN_{ikl}(u) - dQ_{kl}^p(u)\} \tag{25}$$

where π_i is given by (16). Estimating equations $G_1(u) = 0$ and $G_2(u) = 0$ give standard NA estimates $d\widehat{Q}_{kl}(u)$ and $d\widehat{A}_k(u)$, whereas $G_3(u) = 0$ gives a weighted NA estimate, $d\widehat{Q}_{kl}^p(u)$, i.e.,

$$d\widehat{Q}_{kl}(u) = \sum_{i=1}^n I(C_i \geq u)Y_{ik}(u^-)dN_{ikl}(u) \Big/ \sum_{i=1}^n I(C_i \geq u)Y_{ik}(u^-) \tag{26}$$

$$d\widehat{A}_k(u) = \sum_{i=1}^n I(C_i = u)Y_{ik}(u^-) \Big/ \sum_{i=1}^n I(C_i \geq u)Y_{ik}(u^-) \tag{27}$$

$$d\widehat{Q}_{kl}^p(u) = \sum_{i=1}^n R_i\pi_i^{-1}I(C_i < u)Y_{ik}(u^-)dN_{ikl}(u) \Big/ \sum_{i=1}^n R_i\pi_i^{-1}I(C_i < u)Y_{ik}(u^-). \tag{28}$$

Estimates of state occupancy probabilities can be obtained by using (22), with the $Q(u)$ matrix replaced with the matrix $Q^*(u)$ that applies to the joint model with states $1, 2, \ldots, K$ and $1^p, 2^p, \ldots, K^p$. The first row of (22) gives the desired probabilities $P_{1k}(0, t)$ and $P_{1k^p}(0, t)$ for k or $k^p = 1, \ldots, K$. When those selected for tracing are a simple random sample (SRS) of individuals lost to followup, π_i is constant for all i with $C_i < A_i$, and π_i can be dropped from (25). In this case, the R package etm can be used to obtain the estimates $\widehat{Q}_{kl}(\cdot)$, $\widehat{A}_k(\cdot)$, $\widehat{Q}^p_{kl}(\cdot)$ and $\widehat{P}(0, t)$; see the work of Cook and Lawless (2018, Section 3.4.3). With a weighted estimator for $Q^p_{kl}(\cdot)$, direct calculation using the expression above is straightforward. We provide illustrations below for $\widehat{Q}^p_{kl}(\cdot)$ and $\widehat{P}(0, t)$. If values of π_i are specified by design, then their values may be used directly; however, efficiency can be improved by replacing the π_i values with consistent estimates. For example, if we randomly sample from n_1 persons lost to follow-up and end up with r_1 individuals traced, we would use $\widehat{\pi}_i = r_1/n_1$ in place of π_i. For variance estimation, we use the nonparametric bootstrap based on selecting n individuals with replacement from the data. We illustrate this in Section 5.

A reviewer has noted that the methods of Sections 3.1 and 3.2 allow conditional survival probabilities $P(T > t | T > s)$ or, more generally, transition probabilities $P_{kl}(s, t) = P(Z(t) = l | Z(s) = k)$ to be estimated. They also suggested that the methods of Section 3.3 might be extended by using the landmark AJ approach of Putter and Spitoni (2018), which was developed in the context of independent LTF. This is correct: for a given $s > 0$, we simply base estimation of transition rates only on individuals who were in state k at time s, and compute $\widehat{P}(s, t)$ using the product integral over $(s, t]$ in (22). The first row of this provides robust estimates of the $P_{kl}(s, t)$, $l = 1, \ldots, K$. A limitation of this approach is that there may be few individuals in a specified state k at time s.

3.3.1 Illustration: Failure Time Model

The extended failure time model is shown in Figure 2. Replacing $\lambda^p(t|c)$ with the marginal rate function $\lambda^p(t)$, we have the following Q-matrix, with states in the order $1, 2, 1^p, 2^p$:

$$Q(u) = \begin{pmatrix} -\Lambda(u) - A(u) & \Lambda(u) & A(u) & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -\Lambda^p(u) & \Lambda^p(u) \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

NA estimates are given by (26) – (28), where, using notation of Section 3.1, we have

$$d\widehat{\Lambda}(u) = d\widehat{Q}_{12}(u) = \sum_{i=1}^{n} \Delta_i \, I(t_i = u) \Big/ \sum_{i=1}^{n} I(t_i \geq u),$$

$$d\widehat{A}(u) = \sum_{i=1}^{n} \Delta_i^c \, I(t_i = u) \Big/ \sum_{i=1}^{n} I(t_i \geq u),$$

$$d\widehat{\Lambda}^p(u) = \sum_{i=1}^{n} \Delta_i^c \, \Delta_i^p \, I(t_i^p = u) \, R_i \, \widehat{\pi}_i^{-1} \Big/ \sum_{i=1}^{n} \Delta_i^c \, I(t_i^p \geq u) \, R_i \, \widehat{\pi}_i^{-1}.$$

Estimates $\widehat{P}_{11}(0, t)$ and $\widehat{P}_{11^p}(0, t)$ are then obtained from (22); note that $d\widehat{Q}(u) = 0$ except at times where a transition of some type appears so the product in (22) has a finite number of terms. The survivor function for T is estimated by

$$\widehat{S}_T(t) = \widehat{P}_{11}(0, t) + \widehat{P}_{11^p}(0, t). \tag{29}$$

Frangakis and Rubin (2001) proposed an alternative estimator, based on estimates $\widehat{\Lambda}(u)$ and $\widehat{\Lambda}^p(u)$. However, their estimator applies only for the case where individuals traced are an SRS of those lost

to follow-up. Ours is more general and, when we have an SRS, can be calculated using the R function `etm`, which also provides variance estimates. Moreover, ours is fully efficient when the extended model is Markov, with $\lambda^p(t|c) = \lambda^p(t)$.

We remark that (29) is different from (21) with $G = 1$, which is an intensity-based estimate that uses the censoring times C_i. The intensity-based model for $\lambda^p(t|c)$ will, in general, be misspecified to some extent, and we show in simulations in Section 4 that (21) is biased under misspecification, whereas (29) is not.

3.4   OTHER AUXILIARY INFORMATION

If tracing of persons lost to follow-up is used routinely in a setting where LTF times are unambiguous, then an individual is usually selected for tracing once he or she has been lost to follow-up for a specified length of time. In many settings, however, tracing studies may be done infrequently, as resources permit, and exact times when persons become lost to follow-up may not be known exactly. In that case, a decision as to when a person is declared lost to follow-up is needed, and the success of tracing and collection of accurate life history data is generally reduced, especially for individuals who have been lost to follow-up for a long period. The illustration in Section 5 involves a setting where LTF times need to be defined. When individuals are in a study cohort with a protocol for observation, it is best to clearly define under what circumstances a person will be declared lost to follow-up.

In some studies, LTF can have multiple assignable and observable causes. For example, in randomized trials, a subject may be withdrawn because of adverse events, the necessity to switch treatment, a decision by the subject or the study's funder, and so on. This information allows expansion or enrichment of the life history process describing the primary outcome(s), intermediate or concurrent events, and LTF. Such data should be collected, even if it requires some degree of tracing for certain subjects in order to ascertain the reason for LTF. We note the increasing interest in dependent and informative LTF in randomized trials, and the recent expansion of clinical trial guidelines to reflect this (Unkel et al., 2018). Indeed, to arrive at a comprehensive understanding of disease treatment, it is desirable to collect data on life histories after LTF. This might be done for a selected subset of study subjects, perhaps as a new study. In any case, it is important that the pre-LTF data from the randomized trial be included in the subsequent analysis.

The collection of data on fixed or time-varying covariates that may be related to both life history events and LTF is also highly desirable. This has been stressed by Robins and his collaborators; in the context of this paper, see comments of the aforementioned authors (Robins et al., 2001) on the work of Frangakis and Rubin (2001). Such variables provide insight into factors that may produce non-independent LTF. Importantly, they can also serve as a basis for IPC weighting that can facilitate consistent estimation of marginal or partly conditional features of a life history process in which not all such covariates are represented. However, such variables do not necessarily remove the need for tracing studies, since unmeasured factors that produce non-independence may exist.

Finally, it is of interest in tracing studies to collect any other data that might explain differences in pre- and post-LTF transition intensities. As noted in Section 1, these may sometimes be related to differences in environment pre- and post-LTF.

4 ILLUSTRATIVE CALCULATIONS AND NUMERICAL STUDIES

We present here some studies to investigate the nature and extent of biases that can arise from dependent LTF when it is not addressed in estimation of process intensities. We also investigate the effectiveness of tracing and extended follow-up in identifying and dealing with dependent censoring. We consider models for which dependence is induced by an unobserved covariate; more general dependent LTF is discussed briefly in Section 6.

4.1 EFFECTS OF DEPENDENT LTF

In this section, we consider extensions of the failure time model in Figure 1(a) and the progressive model in Figure 1(b). Figure 5 shows dependence of transition intensities on a scalar covariate X. We assume that X has a standard normal distribution with a density function denoted by $g(x)$. Conditional on X, the transition intensities before and after LTF are the same, but we assume that X is unobserved, which induces dependent LTF.
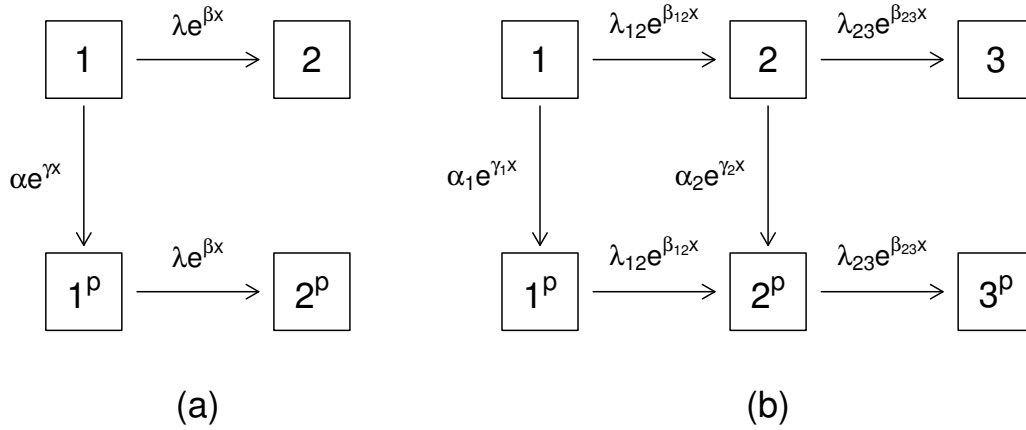

Figure 5: Failure time and progressive models with LTF and a covariate X

We first obtain the marginal intensities and occupancy probabilities for each process. For the failure time process, in the absence of LTF (i.e. when $\alpha = 0$), we have

$$S^\circ(t) = \int_{-\infty}^{\infty} \exp(-\lambda e^{\beta x} t)\, g(x)\, dx \tag{30}$$

and $\lambda^\circ(t) = -d \log S^\circ(t)/dt$; this is obtained in the same spirit as Example 1. For the four-state joint failure time-LTF process $\{\mathcal{Z}(t), t > 0\}$ in Figure 2(a), the survivor function for the failure time $T = \min\{t : \mathcal{Z}(t) = 2 \text{ or } 2^p\}$ is $S_T(t) = P(\mathcal{Z}(t) = 1) + P(\mathcal{Z}(t) = 1^p)$. Because the four-state model accommodates failure after LTF and LTF is independent given $X = x$ (i.e. the $1 \to 2$ and $1^p \to 2^p$ intensities are the same given X), we have $S_T(t) = S^\circ(t)$ and $\lambda_T(t) = -d \log S_T(t)/dt = \lambda^\circ(t)$. We can also obtain the marginal intensities for the transitions depicted in Figure 2(a) as

$$\lambda(t) = E_X\{\lambda e^{\beta X} \mid \mathcal{Z}(t^-) = 1\} = \frac{\int_{-\infty}^{\infty} \lambda e^{\beta x} P_{11}(t^- \mid x)\, g(x)\, dx}{\int_{-\infty}^{\infty} P_{11}(t^- \mid x)\, g(x)\, dx}, \tag{31}$$

and

$$\lambda^p(t \mid c) = E_X\{\lambda e^{\beta X} \mid \mathcal{Z}(t^-) = 1^p, C = c\} = \frac{\int_{-\infty}^{\infty} \alpha e^{\gamma x} P_{11}(c^- \mid x) \lambda e^{\beta x} \exp(-\lambda e^{\beta x}(t - c)) g(x) dx}{\int_{-\infty}^{\infty} \alpha e^{\gamma x} P_{11}(c^- \mid x) \exp(-\lambda e^{\beta x}(t - c)) g(x) dx},$$

where $P_{11}(s^- \mid x) = \exp(-[\lambda e^{\beta_{12} x} + \alpha e^{\gamma x}] s)$. The marginal failure rate post-LTF is

$$\lambda^p(t) = E_C\{\lambda^p(t \mid C) \mid \mathcal{Z}(t^-) = 1^p\} = \frac{\int_{-\infty}^{\infty} \left\{\int_0^{t^-} \alpha e^{\gamma x} P_{11}(c^- \mid x) \lambda e^{\beta x} \exp(-\lambda e^{\beta x}(t - c)) dc\right\} g(x) dx}{\int_{-\infty}^{\infty} \left\{\int_0^{t^-} \alpha e^{\gamma x} P_{11}(c^- \mid x) \exp(-\lambda e^{\beta x}(t - c)) dc\right\} g(x) dx}.$$

To illustrate the effects of dependent LTF, we consider the following scenarios. We let $A_i = A = 1$ be a common administrative censoring time, set $\beta = \log 1.5$, and set λ so that $S(1) = 0.2$, where $S(t)$ is given in (30); thus, in the absence of LTF, the probability of failure before $A = 1$ is 0.80. We then consider $\gamma \in \{\log 0.5, 0, \log 2\}$ and, for each value of γ, set α so that the probability of LTF over $[0, 1]$, i.e.,

$$\int_{-\infty}^{\infty} \left[\int_0^1 \alpha e^{\gamma x} \exp(-(\lambda e^{\beta x} + \alpha e^{\gamma x}) s) ds\right] g(x) dx,$$

is equal to 0.2, 0.4 or 0.6. This yields a total of 9 scenarios, with $\gamma = 0$ corresponding to independent LTF.

Under independent or dependent LTF the survivor function for T is $S^\circ(t)$ given by (30). However, when LTF is dependent, the standard nonparametric Kaplan-Meier estimate based on the censored failure time observations $(\min(T, C), I(T \le C))$ over $0 < t < A$, converges in probability to

$$S_c(t) = \exp(-\Lambda(t)) = \exp\left(-\int_0^t \lambda(s) ds\right),$$

where $\lambda(s)$ is given by (31). Figure 6 displays plots of $\Lambda^\circ(t) = \int_0^t \lambda^\circ(s) ds$ and $\Lambda(t) = \int_0^t \lambda(s) ds$ in panel (a) and $S^\circ(t)$ and $S_c(t)$ in panel (b) for the cases of $\gamma = \log 0.5$ and $\log 2$ with 40% and 60% random censoring. It is apparent and not surprising that the asymptotic bias $S_c(t) - S^\circ(t)$ is larger the higher the rate of random censoring, and that, for a given value of β, the direction of the bias depends on the nature of the dependence between the LTF time and the covariate X through γ.

We next consider the three state process in Figure 5(b) and set $\lambda_{23} = 2\lambda_{12}$ and $\beta_{12} = \beta_{23} = \beta = \log 1.5$. We then determine λ_{12} so that the probability of entering the absorbing state 3 by $A = 1$ is 0.8 in the absence of random censoring (i.e. when $\alpha_1 = \alpha_2 = 0$). For the random censoring process we let $\gamma_1 = \gamma_2 = \gamma \in \{\log 0.5, 0, \log 2\}$ and set $\alpha_2 = 2\alpha_1$ which represents a higher rate of LTF among individuals in the intermediate state. The value of α_1 is then set so that the probability of random LTF over $[0, A]$ (i.e. $P(\mathcal{Z}(1) \in \{1^p, 2^p, 3^p\})$) is 0.2, 0.4 or 0.6.

We will show the biasing effects of dependent LTF on estimation of the cumulative transition rates $\Lambda_{12}(t)$, $\Lambda_{23}(t)$ and state occupancy probabilities $P_{12}(t)$, $P_{13}(t)$. In the absence of censoring (i.e. when $\alpha_1 = \alpha_2 = 0$), the transition rate for $1 \to 2$ transitions is

$$\lambda_{12}(t) = E_X\{\lambda_{12} e^{\beta_{12} X} \mid Z(t^-) = 1\} \tag{32}$$

where the density of X given $Z(t^-) = 1$ is

$$g(x \mid Z(t^-) = 1) = \frac{\exp(-\lambda_{12} e^{\beta_{12} x} t) g(x)}{\int_{-\infty}^{\infty} \exp(-\lambda_{12} e^{\beta_{12} x} t) g(x)}.$$
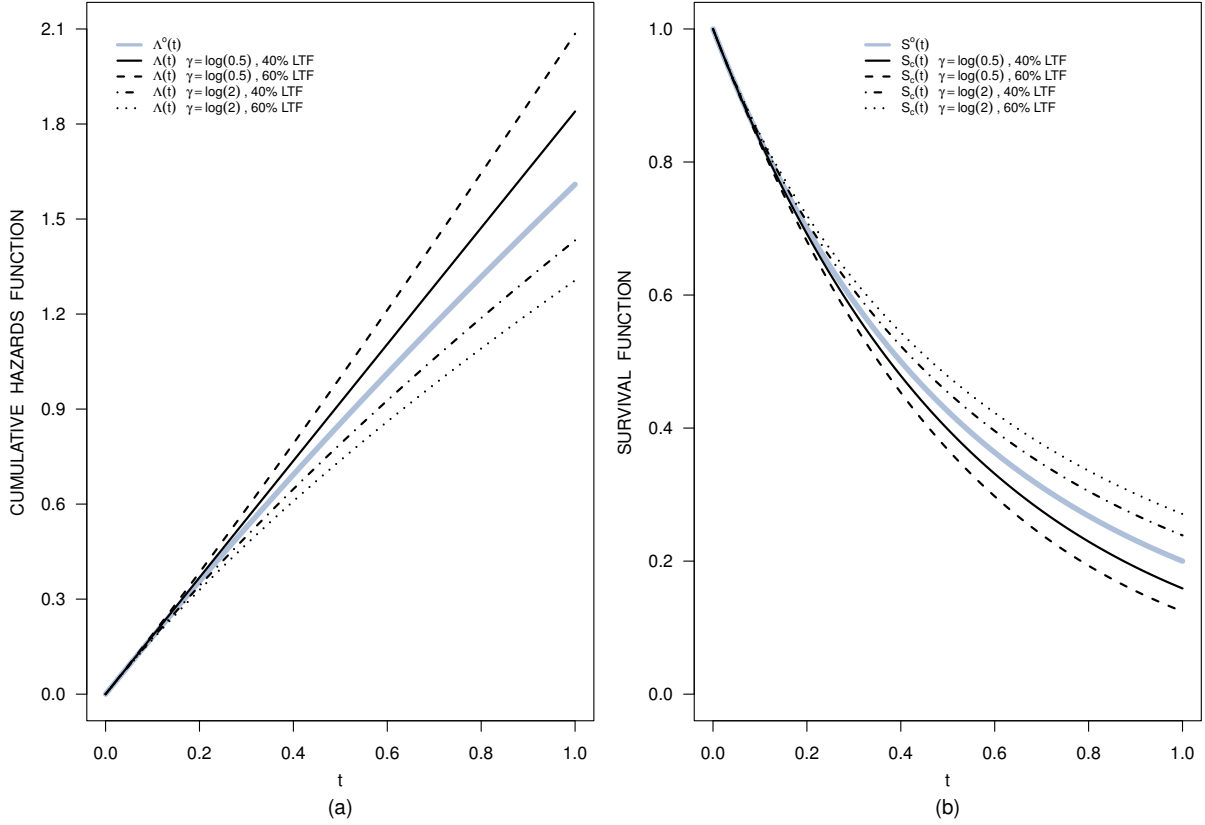
Figure 6: Plots of the cumulative hazard functions $\Lambda^\circ(t)$ and $\Lambda(t)$ (panel (a)) and survival functions $S^\circ(t)$ and $S_c(t)$ (panel (b)) for different combinations of γ and random censoring rates under dependent LTF based on Figure 5(a)

Letting T_2 denote the entry time of state 2, the intensity for a $2 \to 3$ transition is

$$\lambda_{23}(t \mid t_2) = E_X\{\lambda_{23}e^{\beta_{23}X} \mid Z(t^-) = 2, T_2 = t_2\} \tag{33}$$

where the expectation here uses

$$g(x \mid Z(t^-) = 2, t_2) = \frac{\lambda_{12}e^{\beta_{12}x} \exp(-\lambda_{12}e^{\beta_{12}x} t_2) \exp(-\lambda_{23}e^{\beta_{23}x}(t - t_2)) g(x)}{\int_{-\infty}^{\infty} \lambda_{12}e^{\beta_{12}x} \exp(-\lambda_{12}e^{\beta_{12}x} t_2) \exp(-\lambda_{23}e^{\beta_{23}x}(t - t_2)) g(x) \, dx}.$$

The marginal transition rate is then

$$\lambda_{23}(t) = E_{T_2}\{\lambda_{23}(t \mid T_2) \mid Z(t^-) = 2\} \tag{34}$$

$$= \frac{\int_{-\infty}^{\infty} \left\{ \int_0^t \lambda_{23}(t \mid t_2) \lambda_{12}e^{\beta_{12}x} \exp(-\lambda_{12}e^{\beta_{12}x} t_2) \exp(-\lambda_{23}e^{\beta_{23}x}(t - t_2)) \, dt_2 \right\} g(x) \, dx}{\int_{-\infty}^{\infty} \left\{ \int_0^t \lambda_{12}e^{\beta_{12}x} \exp(-\lambda_{12}e^{\beta_{12}x} t_2) \exp(-\lambda_{23}e^{\beta_{23}x}(t - t_2)) \, dt_2 \right\} g(x) \, dx}.$$

The occupancy probabilities for this process can be computed using the intensities $\lambda_{12}(t)$ and $\lambda_{23}(t|t_2)$ or $\lambda_{23}(t)$ in (32) - (34), respectively, or simply by noting that

$$P_{1k}(t) = P(Z(t) = k \mid Z(0) = 1) = E_X\{P(Z(t) = k \mid Z(0) = 1, X)\}. \tag{35}$$

The estimands of interest include the rates and corresponding cumulative rates $\Lambda_{12}(t)$, $\Lambda_{23}(t)$, and

the occupancy probabilities $P_{1k}(t)$, $k = 1, 2, 3$. They can be estimated nonparametrically by NA and AJ estimators, respectively, either in the absence of LTF or under independent LTF. However, these estimators are biased in the presence of dependent LTF, as we now illustrate. In the presence of censoring, the calculation of intensities and rates is analogous but requires working with the full 6-state model in Figure 5(b). Hence, analogous to (31), we write

$$\lambda_{12}^+(t) = E_X\{\lambda_{12}e^{\beta_{12}X} \mid \mathcal{Z}(t^-) = 1\} = \int_{-\infty}^{\infty} \lambda_{12}e^{\beta_{12}x}\, g(x \mid \mathcal{Z}(t^-) = 1)\, dx \qquad (36)$$

where

$$g(x \mid \mathcal{Z}(t^-) = 1) = \frac{\exp(-[\lambda_{12}e^{\beta_{12}}x + \alpha_1 e^{\gamma_1 x}]\, t)\, g(x)}{\int_{-\infty}^{\infty} \exp(-[\lambda_{12}e^{\beta_{12}}x + \alpha_1 e^{\gamma_1 x}]\, t)\, g(x)\, dx}$$

and

$$\lambda_{23}^+(t \mid t_2) = E_X\{\lambda_{23}e^{\beta_{23}X} \mid \mathcal{Z}(t^-) = 2, t_2\} = \int_{-\infty}^{\infty} \lambda_{23}e^{\beta_{23}x}\, g(x \mid \mathcal{Z}(t^-) = 2, t_2)\, dx \qquad (37)$$

where t_2 now represents the entry time to state 2 in the 6-state model and $g(x \mid \mathcal{Z}(t^-) = 2, t_2)$ is

$$\frac{\lambda_{12}e^{\beta_{12}x}\, \exp(-[\Lambda_{12}(t_2 \mid x) + A_1(t_2 \mid x)])\, \exp(-\int_{t_2}^{t}[\lambda_{23}(s \mid x) + \alpha_2(s \mid x)]\, ds)\, g(x)}{\int_{-\infty}^{\infty} \lambda_{12}e^{\beta_{12}x}\, \exp(-[\Lambda_{12}(t_2 \mid x) + A_1(t_2 \mid x)])\, \exp(-\int_{t_2}^{t}[\lambda_{23}(s \mid x) + \alpha_2(s \mid x)]\, ds)\, g(x)\, dx}.$$

Finally the marginal transition rate is computable as

$$\lambda_{23}^+(t) = E_{T_2}\{\lambda_{23}(t \mid T_2) \mid \mathcal{Z}(t^-) = 2\} \qquad (38)$$

given by

$$\frac{\int_{-\infty}^{\infty} \lambda_{23}e^{\beta_{23}x} \left\{ \int_0^t \lambda_{12}e^{\beta_{12}x}\, \exp(-[\Lambda_{12}(t_2 \mid x) + A_1(t_2 \mid x)])\, \exp(-\int_{t_2}^{t}[\lambda_{23}(s \mid x) + \alpha_2(s \mid x)]\, ds)\, dt_2 \right\} g(x)\, dx}{\int_{-\infty}^{\infty} \left\{ \int_0^t \lambda_{12}e^{\beta_{12}x}\, \exp(-[\Lambda_{12}(t_2 \mid x) + A_1(t_2 \mid x)])\, \exp(-\int_{t_2}^{t}[\lambda_{23}(s \mid x) + \alpha_2(s \mid x)]\, ds)\, dt_2 \right\} g(x)\, dx}.$$

Given the forms of (36) and (37), we can compute occupancy probabilities $P_k^+(t)$ in a (3-state) model for $Z(t)$ with these intensities. We have

$$P_{11}^+(t) = \exp(-\Lambda_{12}^+(t)), \quad P_{12}^+(t) = \int_0^t \lambda_{12}^+(s)\, \exp(-\Lambda_{12}^+(s))\, \exp\left(-\int_s^t \lambda_{23}^+(u|s)\, du\right) ds, \quad (39)$$

and $P_{13}^+(t) = 1 - P_{11}^+(t) - P_{12}^+(t)$. Figure 7 contains plots analogous to those of Figure 6 for the 3-state model in Figure 5(b). Note that when $\gamma = 0$, corresponding to independent censoring, $\Lambda_{rs}^+ = \Lambda_{rs}(t)$, $(rs) = (1, 2), (2, 3)$, and $P_{1k}^+(t) = P_{1k}(t)$, $k = 2, 3$. When $\gamma \neq 0$, we have dependent LTF, and the functions $\Lambda_{rs}^+(t)$, $P_{1k}^+(t)$ that are estimated by naive NA and AJ estimators are not desired estimands $\Lambda_{rs}(t)$ and $P_{1k}(t)$.

4.2 ASSESSMENT OF DEPENDENT LTF: EMPIRICAL STUDIES

4.2.1 EFFICIENCY GAINS FROM TRACING

As discussed in Section 3, standard censored data can be thought of as phase I data consisting of $H_i(0, C_i') = \{N_i(s), 0 < s < C_i' = \min(C_i, A_i), \Delta_i^c = I(C_i < \min(T_i, A_i))\}$. Among individuals with $\Delta_i^c = 1$, we may select a phase II subsample for tracing. We first consider the case in which $\gamma = 0$ in the failure time model of Figure 5(a), and with other parameters β, λ, α set to the values in Section 4.1. With $A_i = A = 1$, the three values of α there give the probability of LTF equal to 0.20, 0.40 and 0.60. Here, we consider tracing as a way to improve efficiency through extended follow-up. We consider three analyses.
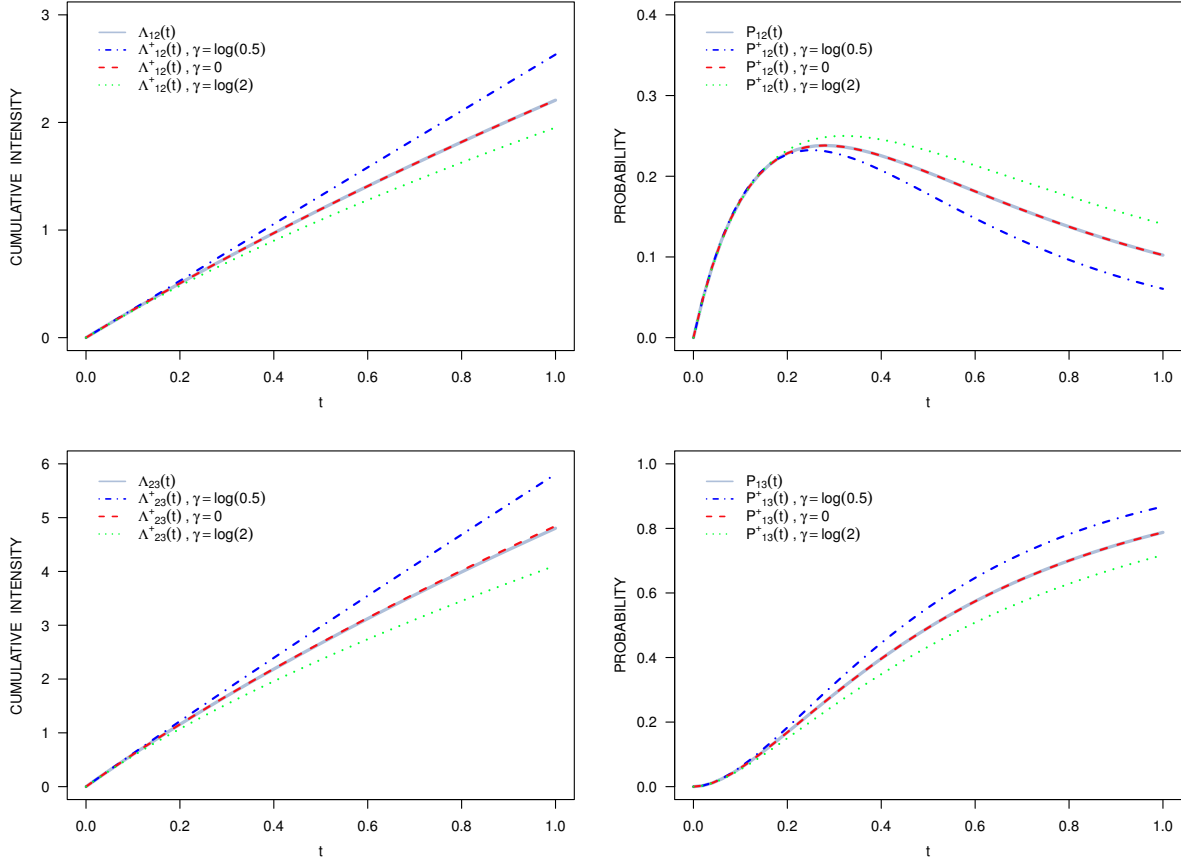
Figure 7: Plots of the cumulative transition rates $\Lambda_{12}(t)$ and $\Lambda_{12}^+(t)$ (top left panel) and $P_{12}(t)$ and $P_{12}^+(t)$ (top right panel) for different values of γ under 60% LTF; the lower panels contain the corresponding plots for the $2 \to 3$ transition intensities (bottom left panel) and $P_{13}(t)$ and $P_{13}^+(t)$ (bottom right panel)

(i) Analysis of phase I data alone based on $L(\theta) \propto \prod_{i=1}^n P(H_i(0, C_i'))$.

(ii) Analysis of phase I and II data based on $L(\theta)$ in (17), where selection for tracing is based on simple random sampling with $\pi_i = P(R_i = 1 \mid H_i(0, C_i'), \Delta_i^c = 1) = \pi = 0.25$ or 0.75.

(iii) Analysis of phase I and II data based on (17) with

$$P(R_i = 1 \mid H_i(0, C_i'), \Delta_i^c = 1) = \delta_0 + \delta_1 I(C_i < 0.5)$$

where $\delta_1 = \log 10$ and δ_0 is chosen so that the marginal probability of selection for tracing is 0.25 and 0.75, respectively. This model will select more individuals with early LTF and might be expected to yield more observed failures in the phase II sample and, hence, more precise estimates of the hazard and survival functions.

In settings (ii) and (iii), we used Bernoulli sampling with independent R_i terms; thus, the number of persons traced varies from sample to sample. We assess the empirical relative efficiency of estimation of the quantiles of $S^\circ(t)$ using the Kaplan-Meier estimator. For settings (ii) and (iii), we add the additional follow-up data to their pre-LTF data for persons who are traced. We report the empirical standard errors in Table 1 for 500 simulated samples with $n = 1000$ individuals each. We show only settings (i) and (ii) in the Table; results for (ii) and (iii) were very similar, contrary to expectations, but

Table 1: Empirical standard errors ($\times 100$) of estimators for quantiles of the failure time distribution under independent LTF ($\gamma = 0$) for different degrees of LTF and tracing

			Quantile			
% LTF	Selection Model	% Traced	0.20	0.40	0.60	0.80
20	(i)	0	0.896	1.654	2.549	2.930
	(ii)	25	0.889	1.637	2.528	2.780
		75	0.882	1.600	2.407	2.649
40	(i)	0	0.927	1.668	3.015	3.653
	(ii)	25	0.914	1.658	2.921	3.276
		75	0.900	1.589	2.532	2.743
60	(i)	0	0.964	1.925	3.769	5.060
	(ii)	25	0.942	1.765	3.248	3.858
		75	0.914	1.625	2.795	2.705

consistent with some other two-phase studies involving failure times and expensive covariates (Lawless, 2018). We see that, as is plausible, tracing mainly improves estimation of upper quantiles when censoring is fairly heavy.

4.2.2 PERFORMANCE OF AJ ESTIMATORS WITH TRACING DATA

In this section, we verify the performance of the AJ estimators of Section 3.3, and demonstrate the bias in semiparametric estimators of Section 3.3.1. We consider the same failure time scenarios as in Section 4.1, and the three types of data (i), (ii) and (iii) in Section 4.2.1. The object of estimation here is $S_T(t)$ in (8), the survivor function for failure time T in the expanded model, whether pre- or post-LTF; for the scenarios in Section 4.1, this is the same as $S^\circ(t)$. Under type (i), which is standard censored data, we use the Kaplan-Meier estimator for $S^\circ(t)$. Under type (ii) data with an SRS of persons lost to follow-up traced, we consider both the unweighted AJ estimator give by (29), and two semiparametric estimators. The first is given by (21) with $G = 1$, 2 or 4, and the second uses a Cox model with $\lambda^p(t|c) = \lambda(t) \exp(\delta' v(c))$, where $v(c)$ indicates intervals $g = 1, \ldots, G$ for C. For type (iii) data, we use the weighted AJ estimator using $d\widehat{\Lambda}^p(u)$ in Section 3.3.1 but the same semiparametric estimators as for the type (ii) data.

We show in Figure 8 the averages of different estimates from 100 samples of size $n = 1000$ for two scenarios with 40% LTF and 25% traced, where $\gamma = \log 0.5$ and $\log 2.0$. Other scenarios gave similar patterns with biases increasing as the % LTF increases. We observe that the mean AJ estimate is very close to the true $S^\circ(t)$, but the mean semiparametric estimates are biased, as is the Kaplan-Meier estimate. For simplicity, we show only the semiparametric estimates (21); the Cox model estimates had slightly more bias for each of $G = 1$ and 4. The estimate with $G = 2$, not shown here, lies between those for $G = 1$ and $G = 4$; as would be expected, the estimates with larger G were less biased. We note that in scenario (iii) involving stratified selection of persons for tracing, the unweighted AJ estimator was close to the weighted estimator. The results are consistent with the case

of 40% LTF in Figure 6.

We also considered the 3-state model in Figure 5(b) for the scenarios discussed in Section 4.1. We found that the naive NA and AJ estimates based on the right-censored data alone tracked closely the functions $\Lambda^+_{r,r+1}(t)$, $r = 1, 2$ and $P^+_{1r}(t)$, $r = 2, 3$ shown in Figure 7. On the other hand, when tracing data from persons lost to follow-up are added, the weighted AJ estimates of $P_{1r}(r)$ based on (22) and the estimators (26) – (28) for the 6-state model are close to the true $P_{1r}(t)$ in Figure 7, $r = 2, 3$. In addition, the weighted AJ estimators of $P^+_{1r}(t) + P^+_{1r^p}(t) = P(\mathcal{Z}(t) = r \text{ or } r^p)$ are consistent.
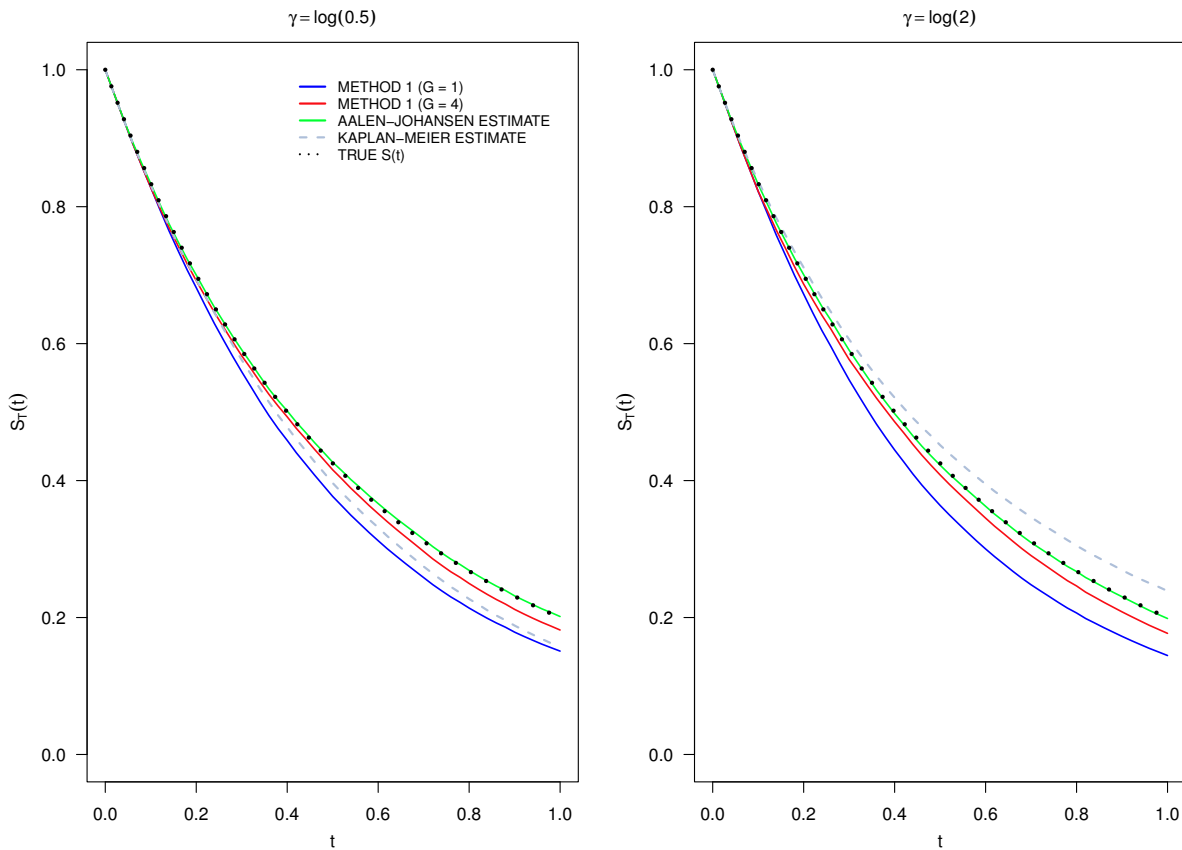

Figure 8: Averages of Kaplan-Meier and Aalen-Johansen estimates, semiparametric estimates (21) with $G = 1$ and 4, and true survivor function $S(t) = S^\circ(t)$ for model (ii) with tracing probability of 0.25

4.2.3 TESTING FOR DEPENDENT LTF

A main use of tracing is to assess whether LTF is independent. We consider the failure time scenarios where $\gamma = \log 0.5$ and $\log 2$ and generate 500 samples of $n = 1000$ individuals. For each data set, we obtain estimates of $\Lambda(t)$ based on the phase I data, as described in Section 3.3.1. We then consider the tracing models in (ii) and (iii) of Section 4.2.1 with tracing probabilities of 0.25 and 0.75. For each phase II data set obtained from tracing, we estimate $\Lambda^p(t) = \int_0^t \lambda^p(s)ds$ as described in Section 3.3.1. In Figure 9, we plot the average estimates of $\Lambda(t)$ and $\Lambda^p(t)$ along with their true values for the cases with $\gamma = \log 0.5$ and $\log 2.0$ at 40% LTF. We see good agreement between the true functions and the average estimates. We also observe the separation between $\Lambda(t)$ and $\Lambda^p(t)$, due to the dependent LTF.
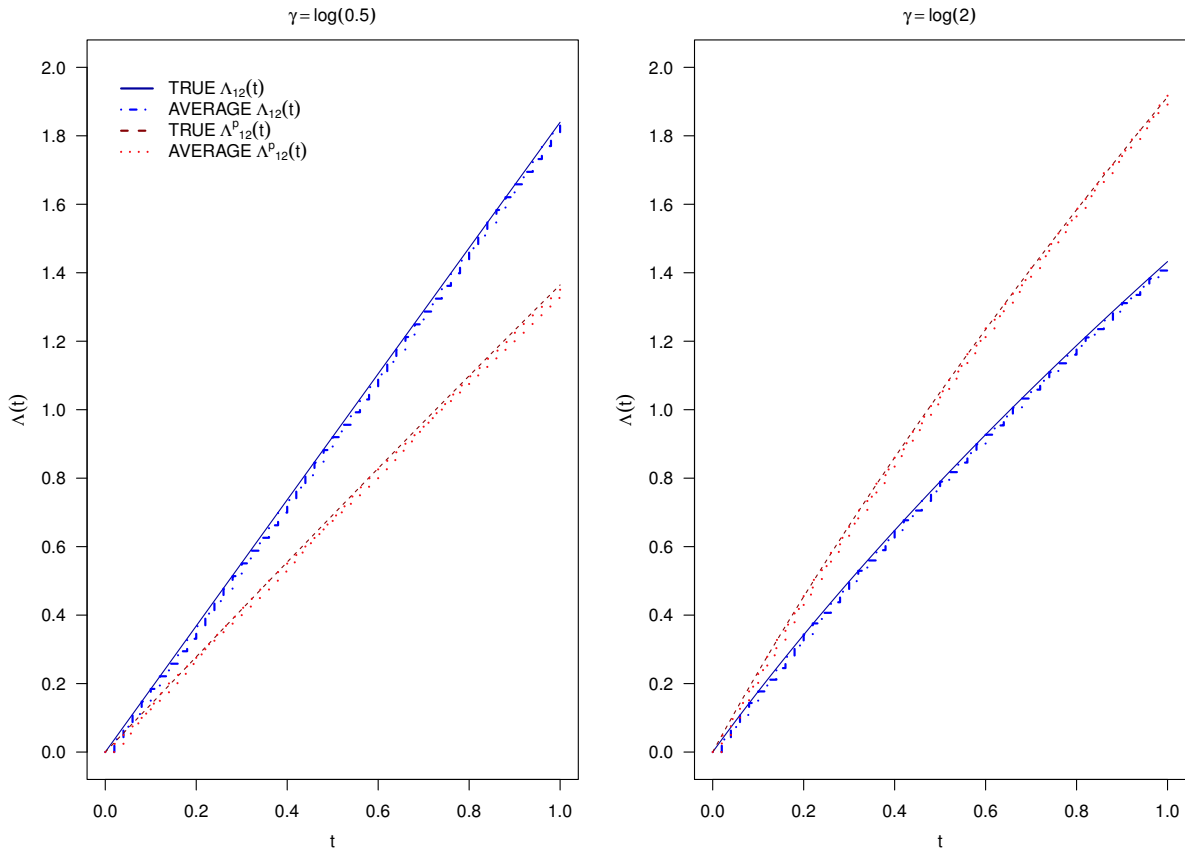
Figure 9: Plots of $\Lambda_{12}(t)$ and $\Lambda_{12}^p(t)$ with $\gamma = \log 0.5$ and $\log 2$ with $\delta_1 = \log 10$ and 40% loss to follow-up; the average estimates are calculated based on 500 samples of $n = 1000$ individuals for model (ii) with tracing probability of 0.25

Formal tests of independent LTF can be directed at either of the null hypotheses H_0:$\lambda^p(t|c) = \lambda(t)$ or H_0': $\lambda^p(t) = \lambda(t)$. The former can be considered using regression models; two of such models that include H_0 and alternatives are the following.

1. A model with

$$\lambda(t \mid c) = \lambda_{12}(t) \exp(\psi_1 V_1(t,c) + \psi_2 V_2(t,c) + \cdots + \psi_G V_g(t,c))$$

   where $V_g(t,c) = I(c \in [b_{g-1}, \min(t, b_g)))$ for cut-points b_g, $g = 1, \ldots, G$; we consider $G = 2$ with $b_0 = 0$, $b_1 = 0.25$ and $b_2 = 1$ and $G = 4$ with $b_0 = 0$, $b_1 = 0.1$, $b_2 = 0.2$, $b_3 = 0.3$ and $b_4 = 1$. The test for independent LTF is a likelihood ratio test with G d.f. based on H_0: $\psi = 0$ versus H_A: $\psi \neq 0$, where $\psi = (\psi_1, \ldots, \psi_G)'$.

2. A model with
$$\lambda(t \mid c) = \lambda_{12}(t) \exp([\psi_1 + \psi_2 c] \, I(c < t)).$$

   A test for independent LTF is a test of H_0: $\psi = 0$.

Both models can be fitted and hypotheses tested, using Cox model software; we use the R/SPLUS function `coxph`. Under the simulation scenario above with $n = 1000$, we generated 500 samples and considered the power of these tests for tracing selection models (ii) and (iii) of Section 4.2.1. For tests with nominal size 0.05, empirical powers for $\gamma = \log 0.5$ and $\log 2$ are shown in Table 2. Although models 1 and 2 are misspecified to some extent, they are correct under H_0, and we see in Table 2 that empirical rejection rates are close to 0.05 when $\gamma = 0$. We observe that the tests based on the stratified

model 1 above are substantially more powerful than those based on the simpler model 2. Moreover, for model 1, the case with $G = 2$ performs better than the one with $G = 4$. We observe also that the selection model (iii) performs similarly to the simple random sampling of model (ii); this is consistent with the similar estimation efficiencies noted earlier. Table 2 also indicates that for substantial power in the scenarios considered, it is necessary to trace a rather high proportion of individuals lost to follow-up.

Tests of H_0' can be based on a comparison of $\widehat{\Lambda}(t)$ and $\widehat{\Lambda}^p(t)$ using some type of distance measure or based on a regression model such as $\lambda^p(t) = \exp(\delta)\lambda(t)$. An investigation of such tests, and a detailed look at sample size and power, is left to a subsequent paper.

5   ILLUSTRATION: A TRACING STUDY IN A LUPUS COHORT

We consider a tracing study undertaken in connection with mortality within a cohort of persons with SLE ("lupus", for convenience) at the University of Toronto Lupus Clinic. Farewell et al. (2003) discussed a tracing study initiated in 1991: briefly, as of September 1, 1991 there were 623 patients in the cohort, but 243 had not been seen at the clinic for 18 months or more, and their death had not been reported. These individuals were deemed lost to follow-up (LTF), and an attempt was made to contact each of them personally or through physician and patient organizations. The efforts to trace individuals continued until December 30, 1992, and by then, there were 170 successful contacts. Among this group, 24 individuals were found to have died. A primary question is whether there is any difference in mortality pre- and post-LTF. We note that, here, it is unclear how precisely to define LTF times and that delays in the reporting of deaths are also a factor in specifying who is lost to follow-up. Farewell et al. (2003) considered two LTF times: the time last seen at the clinic, and 18 months after that. We adopt the latter definition as it accommodates some delays in clinic visits and is close to how we might define LTF prospectively. No information on reporting delays for death is available, and similarly to Farewell et al. (2003), we do not consider its effect.

Farewell et al Farewell et al. (2003) used time since diagnosis of lupus as their time scale. We use age since it would allow comparisons with persons without lupus. Using the notation introduced in Section 3 for the failure time model in Figure 2, we let T_i denote age at death and L_i denote the age at which individual i registered in the lupus clinic. The L_i are therefore LT times in our analyses, and the points discussed in Section 2.4 must be considered. For individuals who become lost to follow-up, C_i denotes age at LTF; the administrative censoring time A_i is a person's age on December 31, 1992. It is assumed that the 175 persons traced among the 248 deemed lost to follow-up constitute a random sample, but this cannot be checked with the data available and is a potential source of bias.

As we discussed in Section 2.4, failure intensities both before and after LTF may not be independent of the L_i, and the analysis below indicates this is the case here. Consequently, both $\lambda(t)$ and $\lambda^p(t)$ involve averages over the distribution of LT times L. This makes their interpretation and comparison difficult, since the distribution of L_i may be different for persons who do and do not become lost to follow-up. We have nevertheless elected to show the NA estimates to illustrate that naive calculation and interpretation of nonparametric estimates can lead to erroneous conclusions. Figure 10 shows NA estimates of pre- and post-LTF cumulative death rates $\Lambda(t)$ and $\Lambda^p(t)$ described in Section 3.3. The estimates are shown along with approximate 95% confidence limits using a standard error estimated from 200 nonparametric bootstrap samples. The bootstrap samples of size $n = 623$ were drawn with replacement from the 623 individuals in the cohort; thus, different bootstrap samples included varying numbers of persons lost to follow-up and persons traced. Confidence limits based on standard variance estimates used under independent LTF (see, e.g. the work of Lawless (2003, Section 3.2.4))

Table 2: Empirical power† of tests for independent LTF with different selection models; $n = 1000$, $nsim = 500$

% LTF	Selection Model	% Traced	$\gamma = \log 0.5$			$\gamma = 0$			$\gamma = \log 2$		
			Model 1		Model 2	Model 1		Model 2	Model 1		Model 2
			$G = 2$	$G = 4$		$G = 2$	$G = 4$		$G = 2$	$G = 4$	
20	(ii)	25	22.4	19.4	17.2	5.2	6.0	4.8	23.2	20.0	14.2
		75	54.8	42.8	37.8	5.6	7.2	6.0	53.6	40.4	28.0
	(iii)	25	26.0	19.4	22.6	4.6	6.4	5.0	25.2	19.6	17.4
		75	57.6	45.2	40.4	4.4	6.0	5.0	56.6	44.8	34.0
40	(ii)	25	51.2	40.0	29.8	4.4	4.2	5.6	48.6	40.4	27.2
		75	86.2	79.4	34.8	3.4	3.8	5.0	86.0	78.2	39.8
	(iii)	25	55.4	45.6	37.0	5.0	5.8	6.6	53.6	41.8	32.4
		75	90.2	83.2	44.8	5.2	4.2	4.8	88.0	81.0	47.8
60	(ii)	25	73.4	64.2	17.2	4.8	4.4	5.0	72.4	62.4	22.0
		75	95.6	93.8	8.6	3.8	3.4	4.6	96.2	92.0	14.4
	(iii)	25	77.6	66.4	23.2	7.6	6.2	6.6	76.0	65.6	29.8
		75	95.6	91.2	8.2	3.8	3.6	5.6	95.4	91.6	20.0

† Entries are the percentage of simulation samples in which H_0 is rejected, based on a Wald test using the `coxph` function in R.
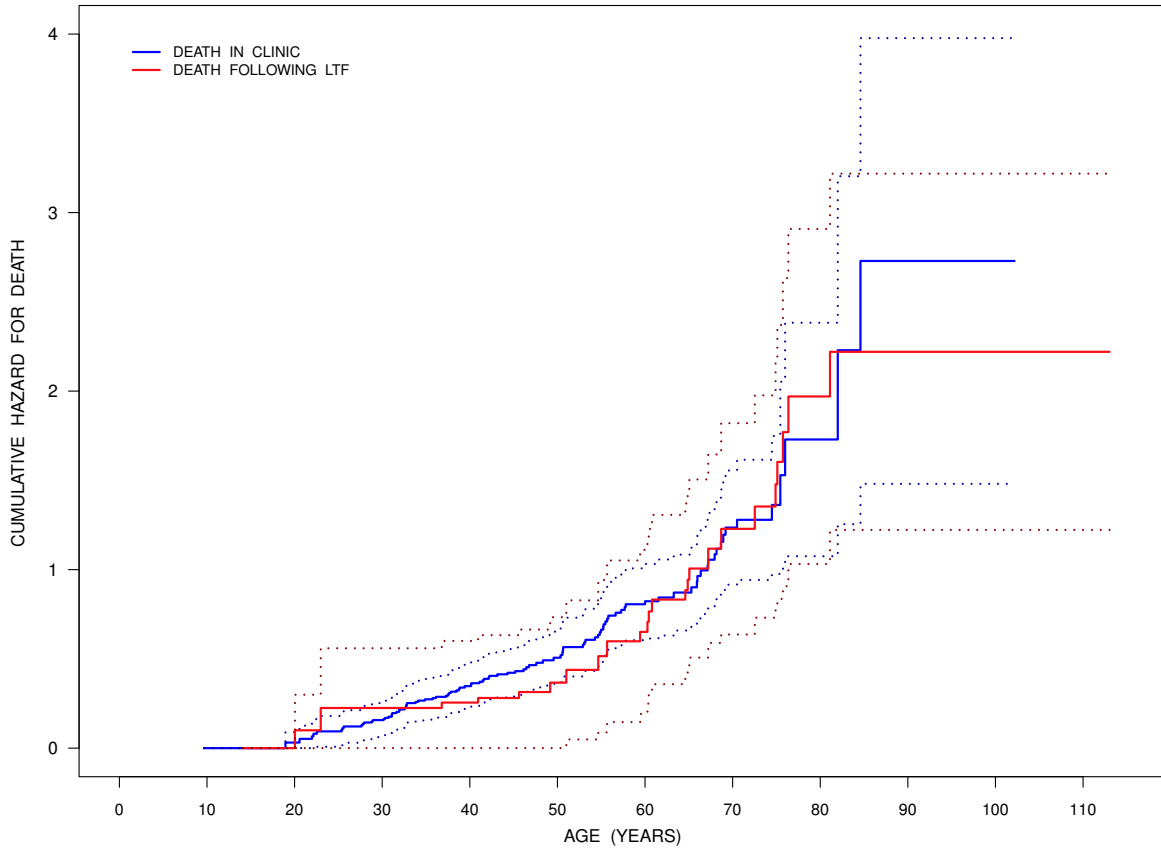
Figure 10: Plot of cumulative hazard rates for death based on Nelson-Aalen estimates with pointwise 95% confidence intervals based on 200 bootstrap samples

are close to the bootstrap limits plotted here. No difference in the average mortality rates pre- and post-LTF is suggested by these results.

We fitted regression models for for the two failure intensities depicted in Figure 2. We allow for dependent LT and consider a combined Cox model defined as follows: with LTF indicator covariate $X_i(t) = I(C_i \leq t)$ and $W_i(t) = (L_i, X_i(t), (C_i - L_i)X_i(t))'$, let

$$\lambda(t|W_i(t)) = \lambda_0(t) \exp(\beta_1 L_i + \beta_2 X_i(t) + \beta_3 (C_i - L_i)X_i(t)). \tag{40}$$

Because LTF is actually defined retrospectively in this particular study, the covariate $X_i(t)$ is not predictable as required for standard martingale-based properties of the Cox partial likelihood. However, the estimates can be interpreted as valid semi-parametric maximum likelihood estimates, as noted by Farewell et al. (2003). The model in (40) allows the death intensity to depend on the age at entry L_i to the lupus clinic, and the intensity post-LTF may depend on the time $C_i - L_i$ between clinic entry and LTF. Table 3 shows relative risk estimates $RR = \exp(\widehat{\beta}_j)$, also referred to as mortality ratios, approximate 95% confidence intervals, and p values for tests of hypotheses $\beta_j = 0$ ($j = 1, 2, 3$). Corresponding results for model 2 with $\beta_3 = 0$ are also shown. All results shown here are obtained using the `coxph` function. Checks of the proportional hazards assumptions were carried out using the `cox.zph` function; the need for expanded models incorporating non-linear covariate effects and the addition of age at lupus diagnosis as a covariate was also assessed. These checks and expanded models did not suggest the need to generalize model 1.

Table 3: Relative risks from Cox models (40) fitted to combined data on deaths pre- and post-LTF, with the latter obtained by tracing patients.

Model	Covariate	RR	95% CI	p
1	L	1.085	(1.035, 1.138)	0.0008
	$X(t)$	0.540	(0.263, 1.107)	0.0925
	$(C - L) X(t)$	1.171	(1.049, 1.308)	0.0048
2	L	1.066	(1.020, 1.114)	0.0047
	$X(t)$	1.022	(0.633, 1.651)	0.9284

$^\dagger \, RR = \exp(\widehat{\beta})$

We observe from fitted model 2 that if the post-LTF covariate $C_i - L_i$ is not included, there is no indication of a difference in death intensities pre- and post-LTF, consistent with Figure 10. However, model 1 indicates there is a difference but that it depends on the time between clinic entry and LTF. Moreover, the failure intensity pre-LTF is seen to depend on L_i. These facts indicate that analyses based on model 2 and the plot in Figure 10 are misleading. Although times between onset of lupus and clinic entry vary widely across patients, the higher death intensity for persons who are older at entry (L_i) and older at LTF (larger $C_i - L_i$) may indicate effects related to age at onset. Farewell et al. (2003), whose analyses used time since onset as the time scale for T_i, found a mildly significant effect ($p = 0.05$) of time between disease onset and LTF. Finally, we remark that when we reran analyses using C_i as age at last visit, the patterns were the same as in Table 3, but with somewhat larger p values (0.026 for β_2; 0.013 for β_3).

6   CONCLUDING REMARKS

If dependent LTF is apparent in a failure time setting, then three survival functions of potential interest are as follows: the function $S^\circ(t)$ in the absence of LTF, the estimand $S_c(t)$ of a naive analysis assuming independent LTF and no tracing, the function $S_T(t)$ in a model accommodating failure post-LTF. If LTF is dependent, the target estimand might be either $S^\circ(t)$ or $S_T(t)$. A caveat about the former is the assumption that removing LTF does not alter the failure time process. For the latter, the question is whether $S_T(t)$ is of direct interest in a given setting. We note that in regression settings involving, for example, an exposure covariate X, simple models for $\lambda(t|x)$ and $\lambda^p(t|x)$ do not lead to simple models for $S_T(t|x)$. Direct regression approaches (see, e.g. the work of Andersen and Klein (2007)) do not appear to help here, since both $P(Z(t) = 2|x)$ and $P(Z(t) = 2^p|x)$ need to be considered. In some cases, it may be best simply to consider death intensities pre- and post-LTF and occupancy probabilities for each of the four states in the joint model of Figure 2. We note that for the settings in Section 4.1, $S^\circ(t)$ and $S_T(t)$ are the same due to the LTF mechanism considered. More generally, they may differ; for example, this will be the case for the model in Figure 5(a) if we have distinct regression coefficient β_{12} and β_{12}^p or different baseline rates λ and λ^p for the $1 \rightarrow 2$ and $1^p \rightarrow 2^p$ transitions. As noted earlier, such differences might arise because of differing environments pre- and post-LTF. For example, Lagakos and Williams (1978) and later authors saw this effect in a randomized treatment trial for lung cancer, where premature withdrawals were due to metastases or an increase in primary tumor size and where subjects were given new treatments post-withdrawal.

Relatively little work has been done on the impact of dependent LTF on multistate processes, but the biases can be appreciable for features such as state occupancy probabilities, as illustrated in Section 4. As discussed in Section 3.4, in both the failure time and general case, IPC weights can be used to estimate process features in the absence of LTF, provided covariates rendering LTF conditionally independent are known (see, e.g. the works of Satten et al. (2001), Datta and Satten (2002), Cook et al. (2009)). However, this assumption cannot be checked solely on the basis of the censored data.

We assumed that the administrative censoring time A is independent of the process of interest. This is often reasonable, but may not be in some settings where individuals are recruited to cohorts or disease registries over a long period. Analyses can be stratified on age or time of recruitment in such cases, or made conditional on related covariates.

We also assumed in this paper that continuous observation of individuals in a study is possible. In some studies, the status of individuals can be ascertained only at intermittent observation times (see, e.g. the work of Cook and Lawless (2018, Chapter 5)). We defer discussion of dependent LTF in such studies to a subsequent paper. We noted the difficulties that delayed entry can create in Section 2.4 and are studying these issues in ongoing work.

Finally, we remark that it would be useful to have further empirical study of tracing studies or two-phase designs, including the collection of information on important covariates. Consideration of LTF and tracing at the onset of a study is valuable and can help mitigate issues that were seen in the lupus clinic illustration in Section 5. Tracing studies should be conducted with a view to potential costs and benefits related to the degree of bias if dependent LTF is ignored. However we make the point, that in many settings, including randomized trials, extended followup of some portion of a study's individuals can yield important scientific information, aside from providing checks on independence of LTF. Among other things, it allows an assessment of long term effects of treatment and of disease processes, after versus during a designed study. In some cases, extended followup might be through registries or administrative databases; this is an area of considerable current activity.

ACKNOWLEDGEMENTS

REFERENCES

Aalen, O., Borgan, O., and Gjessing, H. (2008). *Survival and Event History Analysis: A Process Point of View*. Springer Science & Business Media, New York, NY.

Allignol, A., Schumacher, M., and Beyersmann, J. (2011). Empirical Transition Matrix of Multi-State Models: The etm Package. *Journal of Statistical Software*, 38(4):1–15.

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

Andersen, P. and Klein, J. (2007). Regression analysis for multistate models based on a pseudo-value approach, with applications to bone marrow transplantation studies. *Scandinavian Journal of Statistics*, 34(1):3–16.

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, New York, NY.

Brostrøm, G. (2016). eha: Event History Analysis. R package version 2.4-4.

Cook, R. and Lawless, J. (2007). *The Statistical Analysis of Recurrent Events*. Springer Science & Business Media, 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., Lawless, J., Lakhal-Chaieb, L., and Lee, K.-A. (2009). Robust estimation of mean functions and treatment effects for recurrent events under event-dependent censoring and termination: application to skeletal complications in cancer metastatic to bone. *Journal of the American Statistical Association*, 104(485):60–75.

Crowder, M. (1994). Identifiability crises in competing risks. *International Statistical Review/Revue Internationale de Statistique*, 62(3):379–391.

Datta, S. and Satten, G. (2002). Estimation of integrated transition hazards and stage occupation probabilities for non-Markov systems under dependent censoring. *Biometrics*, 58(4):792–802.

Farewell, V., Lawless, J., Gladman, D., and Urowitz, M. (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.

Kalbfleisch, J. and Prentice, R. (2002). *The Statistical Analysis of Failure Time Data, 2nd Edition*. John Wiley & Sons, Hoboken, NJ.

Keiding, N. (1992). Independent delayed entry. In Klein, J. and Goel, P., editors, *Survival Analysis: State of the Art.* , pages 309–326. Kluwer Academic Publishers, Boston, MA.

Lagakos, S. and Williams, J. (1978). Models for censored survival analysis: A cone class of variable-sum models. *Biometrika*, 65(1):181–189.

Lakhal-Chaieb, L., Rivest, L.-P., and Abdous, B. (2006). Estimating survival under a dependent truncation. *Biometrika*, 93(3):655–669.

Lawless, J. (2003). *Statistical Models and Methods for Lifetime Data, Second Edition*. John Wiley & Sons, Hoboken, NJ.

Lawless, J. (2018). Two-phase outcome-dependent studies for failure times and testing for effects of expensive covariates. *Lifetime Data Analysis*, 24(1):28–44.

Lee, S.-Y. and Tsai, W.-Y. (2005). An estimator of the survival function based on the semi-Markov model under dependent censorship. *Lifetime Data Analysis*, 11(2):193–211.

Lee, S.-Y. and Wolfe, R. (1998). A simple test for independent censoring under the proportional hazards model. *Biometrics*, 54(3):1176–1182.

Legge, A., Doucette, S., and Hanly, J. (2016). Predictors of organ damage progression and effect on health-related quality of life in systemic lupus erythematosus. *Journal of Rheumatology*, 43(6):1050–1056.

Moeschberger, M. and Klein, J. (1995). Statistical methods for dependent competing risks. *Lifetime Data Analysis*, 1(2):195–204.

Prentice, R., Kalbfleisch, J., Peterson, Jr., A., Flournoy, N., Farewell, V., and Breslow, N. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, 34(4):541–554.

Putter, H. and Spitoni, C. (2018). Non-parametric estimation of transition probabilities in non-Markov multi-state models: the landmark Aalen-Johansen estimator. *Statistical Methods in Medical research*, 27(7):2081–2092.

Robins, J. (1993). Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. *Proceedings of the Biopharmaceutical Section, American Statistical Association*, 24(3):24–33.

Robins, J., Rotnitzky, A., and Bonetti, M. (2001). Discussion of the Frangakis and Rubin article. *Biometrics*, 57(2):343–347.

Satten, G., Datta, S., and Robins, J. (2001). Estimator for the survival function when data are subject to dependent censoring. *Statistics and Probability Letters*, 54:397–493.

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

Scheike, T. and Zhang, M.-J. (2007). Direct modelling of regression effects for transition probabilities in multistate models. *Scandinavian Journal of Statistics*, 34(1):17–32.

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.

Therneau, T. and Grambsch, P. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer Science & Business Media, New York, NY.

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences of the United States of America*, 72(1):20–22.

Unkel, S., Amiri, M., Benda, N., Beyersmann, J., Knoerzer, D., Kupas, K., Langer, F., Leverkus, F., Loos, A., Ose, C., and Proctor, T. (2018). On estimands and the analysis of adverse events in the presence of varying follow-up times within the benefit assessment of therapies. *arXiv*, page 2805.01834.

Whittemore, A. (1997). Multistage sampling designs and estimating equations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(3):589–602.

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

Zeng, D. and Lin, D. (2014). Efficient estimation of semiparametric transformation models for two-phase cohort studies. *Journal of the American Statistical Association*, 109(505):371–383.

APPENDIX A   THEOREM 1 AND RELATED RESULTS

From (8) or a direct argument it follows that

$$f_T(t) = -dS_T(t)/dt = \lambda(t)\exp(-\Lambda(t) - A(t)) + \int_0^t g(s)\exp(-\Lambda(s) - A(s) - \Lambda^p(s,t))\lambda^p(t \mid s)\, ds \,.$$

Thus $\lambda_T(t) = f_T(t)/S_T(t)$ equals (9), and $\lambda_T(t) = \lambda(t)$ if and only if

$$\lambda^{-1}(t)\int_0^t g(s)\lambda^p(t \mid s)\exp(-\Lambda^p(s,t))\, ds = \int_0^t g(s)\exp(-\Lambda^p(s,t))\, ds \,,$$

which gives (9).

APPENDIX B   CALCULATION OF TRANSITION INTENSITY FUNCTIONS IN EX-AMPLE 1

The observable transition intensities for failure are given by

$$\lambda(t \mid \mathcal{H}(t))\, dt = P(dN(t) = 1 \mid \mathcal{Z}(t^-) = 1)$$
$$\lambda^p(t \mid \mathcal{H}(t), C = c)\, dt = P(dN^p(t) = 1 \mid \mathcal{Z}(c^-) = 1, \mathcal{Z}(c) = 1^p, \mathcal{Z}(t^-) = 1^p) \,.$$

where $dN^p(t)$ indicates a $1^p \to 2^p$ transition occurs at time t. Thus, we have

$$\lambda(t \mid \mathcal{H}(t^-), \mathcal{Z}(t^-) = 1) = \frac{\int_0^\infty u\lambda\exp(-u(\lambda + \alpha)t)g(u)\, du}{\int_0^\infty \exp(-u(\lambda + \alpha)t)g(u)\, du} = \frac{\lambda}{1 + \phi(\lambda + \alpha)t} \tag{41}$$

where $g(u) = u^{a-1}a^a\Gamma(u)^{-1}e^{-au}$, with $\phi = a^{-1} = \text{Var}(U)$. In addition,

$$\lambda^p(t \mid \mathcal{H}(t), C = c) = \frac{\int_0^\infty u\alpha\exp(-u(\lambda + \alpha)c)(u\lambda)\exp(-u\lambda(t - c))\, g(u)\, du}{\int_0^\infty u\alpha\exp(-u(\lambda + \alpha)c)\exp(-u\lambda(t - c))g(u)\, du} = \frac{(1 + \phi)\lambda}{1 + \phi(\alpha c + \lambda t)} \,.$$

APPENDIX C   FORM OF $\lambda^p(t)$ FOR EXAMPLE 1

We have

$$\lambda^p(t)\, dt = P(dN^p(t) = 1 \mid \mathcal{Z}(t^-) = 1^p\}$$
$$= \frac{\int_0^\infty \{\int_0^{t^-} \exp(-u(\lambda_{12} + \alpha)s)\, (u\alpha)\,\exp(-u\lambda_{12}(t - s))\, ds\}\, (u\lambda_{12})\, dG(u)}{\int_0^\infty \{\int_0^{t^-} \exp(-u(\lambda_{12} + \alpha)s)\, (u\alpha)\,\exp(-u\lambda_{12}(t - s))\, ds\}\, dG(u)}$$
$$= \frac{(a + 1)\lambda_{12}\int_0^{t^-} (\alpha s + \lambda_{12}t + a)^{-a-2}\, ds}{\int_0^{t^-} (\alpha s + \lambda_{12}t + a)^{-a-1}\, ds} \,,$$

when $dG(u) = g(u)\, du$ is the gamma density with mean 1 and variance a^{-1}. This gives

$$\lambda^p(t) = \frac{a\lambda_{12}\{(\alpha t + \lambda_{12}t + a)^{-a-1} - (\lambda_{12}t + a)^{-a-1}\}}{(\alpha t + \lambda_{12}t + a)^{-a} - (\lambda_{12}t + a)^{-a}}$$

APPENDIX D   NONPARAMETRIC ESTIMATES FOR FAILURE TIME MODEL

We consider the log-likelihood function based on (18) in Section 3.1 and consider discrete times t that take values in the set of observed failure times. It then follows that

$$\frac{\partial \ell}{\partial \lambda(t)} = \sum_{i=1}^{n} \left\{ \frac{\Delta_i I(t_i = t)}{\lambda(t)} - I(t_i \geq t) \right\}$$

$$\frac{\partial \ell}{\partial \alpha(t)} = \sum_{i=1}^{n} \left\{ \frac{\Delta_i^c I(t_i = t)}{\alpha(t)} - I(t_i \geq t) \right\}$$

$$\frac{\partial \ell}{\partial \lambda_g^p(t)} = \sum_{i=1}^{n} \Delta_i^c R_i I_g(C_i) I(C_i \leq t) \left\{ \frac{\Delta_i^p I(t_i^p = t)}{\lambda_g^p(t)} - I(t_i^p \geq t) \right\}.$$

By setting each of these expressions equal to zero, we obtain the estimates $\widehat{\lambda}(t)$, $\widehat{\alpha}(t)$, and $\widehat{\lambda}_g^p(t)$.

APPENDIX E   AN ILLUSTRATIVE DATAFRAME

Consider the joint model for an illness-death process and LTF shown in Figure 4. For illustration, suppose that three individuals have the following observed data, where all three have $Z(0) = 1$.

(i) Transition to state 2 at time s_1; LTF from state 2 at time $C = s_2$; selected for extended follow-up, transition to state 3^p at time s_3.

(ii) Transition to state 2 at time s_1; still in state 2 at time $A = s_2$

(iii) Transition to state 2 at time s_1; LTF from state 2 at time $C = s_2$; not selected for extended follow-up.

The following data frame shows the data in the form of at-risk intervals and observed events. States 1^p, 2^p, 3^p have been relabelled here as 4, 5, 6. Columns for covariates are not shown, but can be included.

id	start	stop	from	to	status	R
1	0	s_1	1	2	1	NA
1	0	s_1	1	3	0	NA
1	0	s_1	1	4	0	NA
1	s_1	s_2	2	3	0	NA
1	s_1	s_2	2	5	1	1
1	s_2	s_3	5	6	1	1
2	0	s_1	1	2	1	NA
2	0	s_1	1	3	0	NA
2	0	s_1	1	4	0	NA
2	s_1	s_2	2	3	0	NA
2	s_1	s_2	2	5	0	NA
3	0	s_1	1	2	1	NA
3	0	s_1	1	3	0	NA
3	0	s_1	1	4	0	NA
3	s_1	s_2	2	3	0	NA
3	s_1	s_2	2	5	1	0