Published in final edited form as:
Biostatistics (2016), **17**(2): 350-363
DOI: 10.1093/biostatistics/kxv042

Sieve estimation in a Markov illness-death process under dual censoring

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

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

Abstract

Semiparametric methods are well-established for the analysis of a progressive Markov illness-death process observed up to a noninformative right censoring time. However often the intermediate and terminal events are censored in different ways, leading to a dual censoring scheme. In such settings unbiased estimation of the cumulative transition intensity functions cannot be achieved without some degree of smoothing. To overcome this problem we develop a sieve maximum likelihood approach for inference on the hazard ratio. A simulation study shows that the sieve estimator offers improved finite-sample performance over common imputation-based alternatives and is robust to some forms of dependent censoring. The proposed method is illustrated using data from cancer trials.

Keywords: Cox model; Interval censoring; Method of sieves; Profile likelihood; Progression-free survival.

1 Introduction

Vital status for individuals in a clinical trial is often readily available. Detection of non-fatal events requires closer surveillance, which can prove difficult and costly to maintain over time. As a result survival times are subject to right censoring, but the occurrence of intermediate events may be right-censored earlier or interval-censored between assessments. In general we refer to this scenario as dual censoring. Various forms of dual censoring arise in trials involving tumor progression. Guidelines call for the analysis of so-called time to progression (TTP), coinciding with detection of progression, or progression-free survival (PFS), given by the earliest of TTP and death (FDA, 2007). TTP is typically right-censored at death or the preceding (negative) assessment, which induces dependent censoring. PFS is thus deemed preferable to TTP (FDA, 2007, p. 8), but this outcome is subject to systematic imputation.

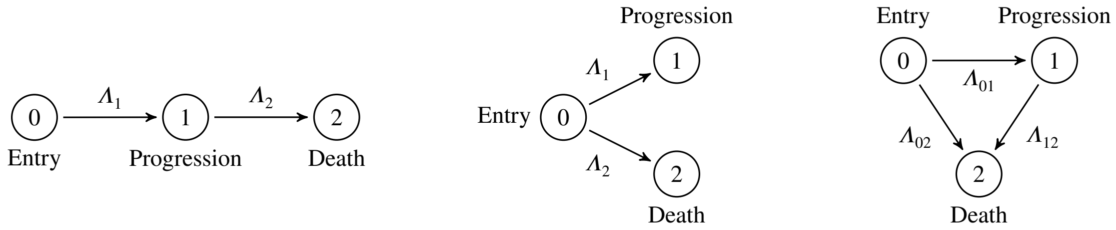

Figure 1: Multistate alternatives to TTP and PFS: chain of events (left), semicompeting risks (middle), and progressive illness-death (right) models.

Multistate models have been suggested as a more natural framework for assessing treatment effects on progression and death. A chain of events model (Fig. 1, left), for example, is useful for settings in which progression always precedes death (Frydman, 1995b). Semicompeting risks (Fig. 1, middle) have been proposed for the case where death may precede progression (Hu and Tsodikov, 2014). Xu et al. (2010) observe that semicompeting risks essentially amount to the progressive illness-death model (Fix and Neyman, 1951; Fig. 1, right), which is fully specified by the state-transition intensity functions.

Among the three state-transition structures, methods to deal with specific instances of dual censoring are most developed for the illness-death model. Frydman (1995a) considers the nonparametric maximum likelihood estimator (NPMLE) from interval-censored progression times with known progression status. This is generalized by Frydman and Szarek (2009) to account for unknown status, which often arises when the last assessment is negative and long precedes right-censoring or death. Bebchuk and Betensky (2001) combine local likelihood and multiple imputation to estimate transition intensities under progression times right-censored before death. Joly et al. (2002) propose spline-based penalized likelihood for the (Cox, 1972) proportional hazards model for an interval-censored variant of this observation scheme. Jackson (2011) considers a piecewise exponential analog by way of time-dependent covariates.

These works recognize that progression and death are observed in different ways, but the broader problem of dual censoring has not yet been considered. Methods for time-to-event endpoints that leave any dependence on time unspecified are generally preferred in practice. However non- and semi-parametric maximum likelihood estimators require the locations of support for the distribution of each transition time, and these are ambiguous whenever the progression status is unknown. To address these issues we develop a sieve estimator for a multistate extension of the Cox model and compare its numerical performance with routine analysis of imputation-based PFS under a variety of censoring scenarios.

2 Dual censoring of the progressive illness-death process

Let N_{hj} be a one-jump counting process representing the transition from state h to state j ($h \neq j$) in the progressive illness-death model and T_{hj} be the corresponding transition time. So T_{01} is the time to progression, T_{02} is the time to progression-free death, and T_{12} is the time of death following progression. Over the observation period $[0, \tau]$, $\tau < \infty$, suppose that the survival time $T_{02} \wedge T_{12}$ is observed up to a right censoring time D, $0 < D \leq \tau$, but progression status $1(T_{01} \leq t)$ is not necessarily known for all $t \in (0, V]$, $V = T_{02} \wedge T_{12} \wedge D$. For example progression may be right-censored at some random time preceding D. Alternatively progression status could assessed periodically, leading to interval censoring.

Whatever the form of this inspection process, we presume that it yields a potential censoring interval $(L, R]$ for the progression time T_{01}. We say "potential" because we may not know with certainty that $T_{01} \in (L, R]$. Put $\Delta_2 = 1(T_{02} \wedge T_{12} \leq D)$ to denote whether or not the survival time is

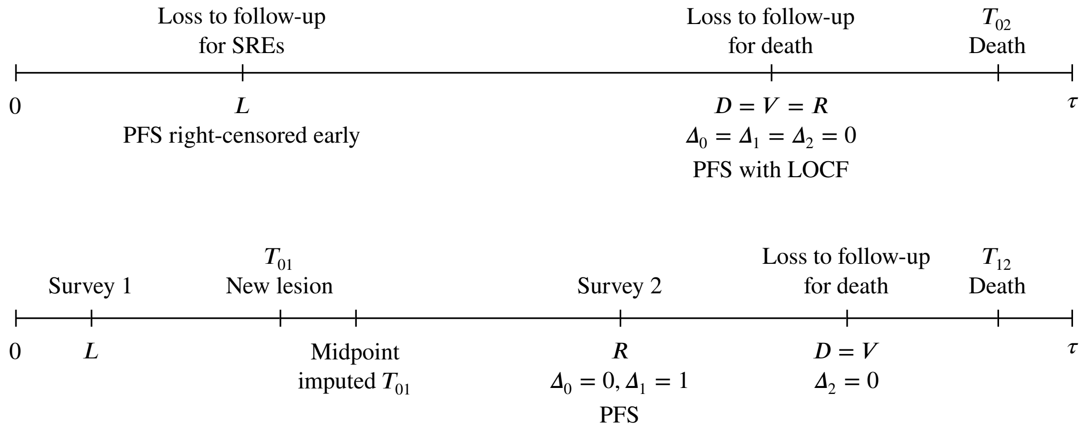

Figure 2: Top: A dual right-censored observation in which progression, coinciding here with first SRE, status is unknown at the last observation time V. Bottom: A dual-censored observation in which lesion progression status is observed to be positive, but the progression time is known only up to the interval $(L, R]$.

observed. Let $\Delta_0 = 1$ whenever progression status is known to be negative at V and $\Delta_0 = 0$ otherwise. Similarly let Δ_1 indicate that progression status is known to be positive at V. So $\Delta_1 = 1$ denotes that, based on the available data, we are certain $T_{01} \in (L, R]$ for some $L < R \leq V$. Otherwise either $\Delta_0 = 1$, indicating that $T_{01} > V$, or progression status is unknown at V. If the status is unknown, then $\Delta_0 = \Delta_1 = 0$ and we cannot rule out the possibility that either $T_{01} \in (L, R]$ or $T_{01} > R$.

2.1 Example: Bone lesions and their complications

Dual right-censored data are encountered in cancer trials evaluating the effect of bisphosphonates on bone metastases and their complications, known as skeletal-related events (SREs). The time of an SRE is often self-evident, but can otherwise be measured accurately through frequent clinic visits, so SREs are typically considered subject only to right censoring. Growth of new or existing bone lesions is assessed by radiographic surveys, which are carried out less frequently. This results in interval-censored lesion progression times. Standard practice is to evaluate SREs and lesions as separate endpoints, as SREs provide the most direct measure of clinical benefit. Time-to-event analysis of either outcome is complicated by the fact that the mortality rate is non-negligible. Use of PFS can circumvent this issue. However since the treatment is intended to manage symptoms rather than prolong survival, the measured effect on PFS will likely underestimate any symptom benefit. The illness-death model offers an alternative that isolates the effect of interest.

A dual-censored observation from this multistate process with first SRE as the intermediate event is illustrated in the top panel of Fig. 2, where we know that no SREs occurred within an initial loss to follow-up time and that the subject survived at least up to final right-censoring time D. We cannot rule out the possibility that progression may have occurred between these two times, so $\Delta_0 = \Delta_1 = 0$. The censoring interval $(L, R]$ here is indeed “potential” from the observed data because in reality, the subject did not experience any SREs. The PFS endpoint has no standard definition in this setting. Practitioners might simply discard all data collected after the initial loss to follow-up time so that PFS is right-censored early. Alternatively the negative progression status at this earlier right censoring time could be carried forward to V, giving PFS with a form of last observation carried forward (LOCF) imputation.

The bottom panel of Fig. 2 considers lesion progression rather than SRE. Here a new lesion developed some time between the first and second radiographic surveys, which gives a censoring interval $(L, R]$ that captures the progression time with certainty: $\Delta_0 = 0$ and $\Delta_1 = 1$. Loss to follow-up occurs after the second survey but before death, giving $V = D$ and $\varDelta_2 = 0$. Guidelines suggest imputing PFS to the time at which progression is first detected, carrying forward the last known progression status to death, and sensitivity analysis to examine variations on this imputation scheme (FDA, 2007, Appendix 2).

Let Y_h be the at-risk process for any transition out of state h, so that $Y_h(t) = 1$ if state h is occupied at time $t-$ and $Y_h(t) = 0$ otherwise. Define the $h \to j$ transition probability $P_{hj}(s,t) = \Pr(N_{hj}(t) = 1 \mid Y_h(s) = 1, N(u), u \leq s)$, with $s \leq t$ and $N = (N_{hj})_{h \neq j}$. Suppose that the observation scheme renders the complete data (T_{01}, T_{02}, T_{12}) coarsened at random in the sense of Heitjan and Rubin (1991). Then the likelihood of a dual-censored observation $X = (L, R, V, \varDelta_0, \varDelta_1, \varDelta_2)$ is

$$(1 - \varDelta_0)P_{00}(0, L)P_{01}(L, R)P_{11}(R, V)\lambda_{12}(V)^{\varDelta_2} + (1 - \varDelta_1)P_{00}(0, V)\lambda_{02}(V)^{\varDelta_2}, \qquad (1)$$

where $Y_h(t)\lambda_{hj}(t)$ is the transition intensity process or instantaneous transition probability at time t. Whatever model we choose for the transition intensity function $\lambda = (\lambda_{hj})_{h \neq j}$, the likelihood *a priori* maximizes to infinity; λ_{h2} $(h = 0, 1)$ can be made arbitrarily large at any time we observe T_{h2} exactly. The usual way out is to replace λ_{h2} by the jump discontinuity $\Delta\varLambda_{h2}$ in the cumulative transition intensity function $\varLambda_{h2} = \int \lambda_{h2}$. However consider an individual with unknown progression status $(\varDelta_0 = \varDelta_1 = 0)$ and known survival time $(\varDelta_2 = 1)$. Surely we need $\varLambda_{02} + \varLambda_{12}$ to increase at $V = T_{02} \wedge T_{12}$, but the observed data are insufficient to *jointly* estimate $\Delta\varLambda_{02}(V)$ and $\Delta\varLambda_{12}(V)$. Nonparametric maximum likelihood will assign mass to at least one of the two potential transition times, but the manner in which support is allocated is subject to bias. Since the so-called risk set for $1 \to 2$ transitions is empty at $t = 0$, the likelihood can be increased appreciably by allocating more mass to potential or observed support for the distribution of T_{12} early in the observation period. So the initial increments in the NPMLE for $\varLambda_{12}$ will tend to be large in finite samples. All of these difficulties can be mitigated by maximizing the likelihood with respect to a sieve—a finite-dimensional approximation to $\{\varLambda\}$ whose size increases with n. Such an approach is generally known as the method of sieves (Grenander, 1981).

3 METHOD OF SIEVES FOR DUAL-CENSORED DATA

Here a sieve is defined for a given random sample $X_i = (L_i, R_i, V_i, \varDelta_0^i, \varDelta_1^i, \varDelta_2^i)$, $i = 1, \ldots, n$, of dual-censored observations. Each element of a sieve corresponds to a piecewise parametric cumulative intensity function $\varLambda_n = (\varLambda_{hj,n})_{h \neq j}$ defined on a data-driven partition of the observation period $[0, \tau]$. Let $\mathcal{L}$ and $\mathcal{R}$ respectively denote the set of left- and right-endpoints from the collection of known censoring intervals $\mathcal{I} = \{(L_i, R_i], i = 1, \ldots, n : \varDelta_1^i = 1\}$. Define $\mathcal{U}_{01}$ as the set of right-endpoints from the maximal intersections (Wong and Yu, 1999) for $\mathcal{I}$; that is, the set of r from $(l, r]$ such that $l \in \mathcal{L}$, $r \in \mathcal{R}$ and $(l, r] \cap (L_j, R_j]$ is either $(l, r]$ or $\varnothing$ for every $(L_j, R_j] \in \mathcal{I}$. In addition let $\mathcal{U}_{h2} = \{V_i, i = 1, \ldots, n : \varDelta_h^i \varDelta_2^i = 1, \sum_{j=1}^n Y_h^j(V_i+) > 0\}$, for $h = 0, 1$, denote the set of exactly-observed terminal event times T_{h2} with known progression status and non-empty $h \to 2$ risk set at $T_{h2}+$.

From Frydman (1995a, Theorem 1), the NPMLE for $\varLambda$ based on the subsample with known progression status $(\varDelta_0^i \vee \varDelta_1^i = 1)$ can be uniquely defined as the discrete maximizer concentrating its support on $\mathcal{U}_{hj}$, $h \neq j$. This implies that $\varLambda_{hj,n}$ should, at minimum, have support on $\mathcal{U}_{hj}$. To ensure that $\varLambda_{02}$ and $\varLambda_{12}$ are jointly estimable, the sieve partition must not isolate any V_i with $\varDelta_2^i = 1$ and unknown progression status $\varDelta_0^i = \varDelta_1^i = 0$. We can achieve this by defining $\varLambda_{hj,n}$ on a partition $\mathcal{T}_{hj,n} = (t_0, \ldots, t_{K_{hj,n}+1})$, such that $K_{hj,n} = O(n^\kappa)$ for $0 < \kappa < 1$, $0 = t_0 < t_1 < \cdots < t_{K_{hj,n}+1} = \tau$, $\max_k(t_k - t_{k-1}) = O(n^{-\kappa})$, and every subinterval $(t_{k-1}, t_k]$ contains at least one point from $\mathcal{U}_{hj}$. Here κ is a tuning parameter that determines the rate at which the sieve or partition size $K_{hj,n}$ increases with n.

For consistency we need $\{\Lambda_n\} \to \{\Lambda_0\}$ as $n \to \infty$. This can be met if the true parameter Λ_0 is sufficiently smooth and increasing on (σ, τ) and the distribution of the inspection times has positive support on (σ, τ) for some small $0 < \sigma < \tau$. We express these and other assumptions throughout this paper more precisely in Appendix A of the supplementary material, but essentially this means $\mathcal{U}_{hj}$ must be dense in (σ, τ) as $n \to \infty$. Such a requirement is stronger than the ones imposed by Joly et al. (2002) and Frydman and Szarek (2009), which allow for unobservable terminal event times with negative progression status: $\mathcal{U}_{02,n} = \varnothing$. A consequence of this is that the support for the distribution of T_{02} is not apparent from the available data, so imposing at least a weakly parametric model for Λ_{02} is needed to achieve consistency.

Apart from the location of support points, estimation from dual-censored data poses two additional challenges: (1) inference under convergence rates possibly slower than the parametric rate, as encountered with various forms of interval censoring, and (2) consideration of Λ_{12} that depends arbitrarily on aspects of the event history, such as the duration in state 1. We avoid these complications by considering a sieve estimator for the Cox model with fixed covariates. This permits inference on the familiar hazard ratio via Murphy and van der Vaart's (2000) profile likelihood theory and, barring the standard Markov assumption, puts little restriction on any dependence with time. A variety of extensions or alternatives could be considered, but we adopt this model as a starting point.

4 SIEVE ESTIMATION OF THE COX MODEL

Sieve estimators have been previously proposed for interval-censored survival data. Huang and Rossini (1997) examine the proportional odds model. Zhang et al. (2010) devise a spline-based sieve for the Cox model. Our setting is complicated by multiple event types and censoring schemes, but these works provide a useful basis for extension. Herein assume that each N_{hj} has cumulative intensity function

$$\Lambda_{hj}(t) \exp(Z'_{hj}\theta), \tag{2}$$

where Z_{hj} is an $h \to j$ transition-type–specific d-vector based on the fixed covariate Z, $\theta \in \Theta \subset \mathbb{R}^d$ is a regression parameter, and $\Lambda_{hj} = \int \lambda_{hj}$ is now a nondecreasing cumulative baseline $h \to j$ transition intensity function. Note that the parameter θ is common to each transition type, but Z_{hj} can be suitably constructed from Z to give type-specific covariate effects (Andersen and Borgan, 1985, pp. 478–480). For example if we wish to estimate the effect of the scalar covariate Z on each transition type separately, we may put $\theta = (\theta_1, \theta_2, \theta_3)'$, $Z_{01} = (Z, 0, 0)'$, $Z_{02} = (0, Z, 0)'$ and $Z_{12} = (0, 0, Z)'$. The effect of Z on the risk of the $0 \to 1$, $0 \to 2$ and $1 \to 2$ transitions then corresponds to θ_1, θ_2 and θ_3, respectively.

Under (2) the cumulative $h \to j$ transition intensity process $Y_h \Lambda_{hj} \exp(Z'_{hj}\theta)$ depends on only the current state occupied and thus satisfies the Markov property $P_{hj}(s,t) = \Pr(N_{hj}(t) = 1 \mid Z, Y_h(s) = 1, N(u), u \le s) = \Pr(N_{hj}(t) = 1 \mid Z, Y_h(s) = 1)$. So $P_{hh}(s,t) = \exp\{-\sum_{j \ne h} \int_s^t \Lambda_{hj}(\mathrm{d}u) \exp(Z'_{hj}\theta)\}$ and $P_{01}(s,t) = \int_s^t P_{00}(s,u) \Lambda_{01}(\mathrm{d}u) \exp(Z'_{01}\theta) P_{11}(u,t)$, for any $s < t$ (e.g. Andersen and Borgan, 1985, Theorem II.6.7). Let $\mathrm{lik}_{\theta,\Lambda}(X)$ denote the likelihood of an observation X given by (1), evaluated under these transition probabilities. Then the sieve maximum likelihood estimator (SMLE), $(\hat{\theta}_n, \hat{\Lambda}_n)$, corresponds to the maximizer of the log-likelihood function $\ell_n(\theta, \Lambda) = \mathbb{P}_n \log \mathrm{lik}_{\theta,\Lambda}(X)$ over $(\theta, \Lambda) \in \Theta \times \{\Lambda_n\}$. The sieve $\{\Lambda_n\}$ is defined by its piecewise parametric family and partition $(\mathcal{T}_{hj,n})_{h \ne j}$. For a sufficiently large partition size, one would not anticipate $\hat{\theta}_n$ to be particularly sensitive to the parametric form on the subintervals. This general notion is demonstrated in the survival case by Huang and Rossini (1997). We defer a discussion on selection of the partition to the end of Section 4.2. The remainder of this section describes the estimation scheme, with illustrations for the piecewise exponential sieve. Numerical results for this sieve are examined in Sections 5 and 6.

4.1 PARAMETER ESTIMATION

Suppose that the cumulative baseline intensity functions from a given sieve $\{\Lambda_n\}$ are specified by the (finite-dimensional) parameter ϕ. The piecewise exponential sieve, for example, is characterized by the piecewise constant values taken by the intensity function. These range through positive values in $(K_{01,n} + K_{02,n} + K_{12,n})$-space. In general, let $\phi_{hj,k}$ denote the parameters specifying $\Lambda_{hj,n}(t)$ for $t \in (t_{k-1}, t_k]$ and $k = 1, \ldots, K_{hj,n}$. Then the SMLE satisfies the score equations $\nabla_{\theta,\phi}\ell_n(\theta, \phi) = 0$. These can be solved using the following self-consistency algorithm, which is akin to the routine outlined by Frydman (1995b) under the null model with θ fixed at zero and no dual censoring.

STEP 4.1 For $c_{hj} > 0$ and $0 < \kappa < 1$, define $\mathcal{T}_{hj,n}$ as a partition of $[0, \tau)$ in which each subinterval contains $\lceil n_{hj}/(c_{hj} n^{\kappa})\rceil$ points from $\mathcal{U}_{hj}$. Set $r = 0$, $\theta^{(0)} = 0$ and $\phi^{(0)}$ to some "neutral" value that ensures Λ_n is increasing. For example with the piecewise exponential sieve $\phi^{(0)} = 1/\tau$.

STEP 4.2 Find a candidate increment $\eta^{(j)} = (\eta_\theta^{(j)}, \eta_\phi^{(j)})$. For $\eta_\theta^{(j)}$, apply the Newton-Raphson method: $\eta_\theta^{(j)} = -\nabla_\theta^2 \ell_n(\theta^{(j)}, \phi^{(j)})^{-1} \nabla_\theta \ell_n(\theta^{(j)}, \phi^{(j)})$. Obtain $\eta_\phi^{(j)}$ via the self-consistency equations (Turnbull, 1976) that result from re-arranging the score equation $\nabla_\phi \ell_n(\theta, \phi) = 0$ to give a recursive expression for ϕ. For the piecewise exponential sieve this is loosely:

$$\eta_{\phi_{hj,k}}^{(j)} = \frac{\mathbb{P}_n \operatorname{E}\Big(\int_{t_{k-1}}^{t_k} \mathrm{d}N_{hj}(s) \,\Big|\, X\Big)}{\mathbb{P}_n \exp(Z_{hj}'\theta) \operatorname{E}\Big(\int_{t_{k-1}}^{t_k} Y_h(s)\,\mathrm{d}s \,\Big|\, X\Big)} - \phi_{hj,k}^{(j)},$$

where the conditional expectations are evaluated under $\theta = \theta^{(j)} + \eta_\theta^{(j)}$ and $\phi = \phi^{(j)}$. These are equal to one if X provides (T_{01}, T_{02}, T_{12}) exactly. A precise expression for this ratio is provided in Appendix D of the supplementary material.

STEP 4.3 Increment $\theta^{(j)}$ and $\phi^{(j)}$ by $\eta_\theta^{(j)}/2^j$ and $\eta_\phi^{(j)}/2^j$, respectively, with j the smallest nonnegative integer ensuring no decrease in the log-likelihood. This gives $(\theta^{(j+1)}, \phi^{(j+1)})$. If, for some small positive value ε, $\|\theta^{(j+1)} - \theta^{(j)}, \phi^{(j+1)} - \phi^{(j)}\|_\infty < \varepsilon$ then stop. Otherwise set j to $j + 1$ and return to Step 4.2.

For each Λ the log-likelihood is concave in θ, which implies that the Newton-Rapshon method yields a profile maximizer for θ. Similar properties are not readily available for Λ. So the score equations may neither uniquely characterize the SMLE nor identify global maxima. Multiple (local) maxima may be detected with different starting values and examination of the profile log-likelihood. Our experience thus far has uncovered rare instances where the increment-halving procedure in Step 4.3 reduces the first and only candidate increment to its starting value. In this narrow form of local maxima, the algorithm could be initialized with starting values based on imputed data.

4.2 VARIANCE ESTIMATION

In Appendices B and C of the supplementary material we show that if the rth ($r = 1$ or 2) derivative of Λ_0 is continuous, positive and bounded on (σ, τ) and some regularity conditions hold, then the SMLE $(\hat\theta_n, \hat\Lambda_n)$ converges to the truth (θ_0, Λ_0) at the rate $O_P(\min(n^{(1-\kappa)/2}, n^{r\kappa}))$ with $1/(4r) < \kappa < 1/2$. However $\hat\theta_n$ achieves the semiparametric efficiency bound. Both the limiting distribution of $\max(n^{-(1-\kappa)/2}, n^{-r\kappa})(\hat\Lambda_n - \Lambda_0)$ and interval estimation for Λ remain as open problems.

Holding $\theta^{(r)}$ fixed in the self-consistency algorithm described in Section 4.1 evaluates the profile log-likelihood needed to estimate standard error for $\hat\theta_n$ under Murphy and van der Vaart (2000, Corollary 3), which gives an approximation to the curvature in the profile log-likelihood at $\hat\theta_n$ akin to numerical differentiation. This entails successively perturbing the entries in $\hat{\theta}_n$ by a chosen value $h_n = o(1/\sqrt{n})$. The data-driven procedure outlined in Boruvka and Cook (2015, Section 6) reduces the choice to specifying typical $\mathrm{typ}\,\theta$ and extreme $\sup\theta$ (absolute) values for any given entry in θ.

Estimation thus entails setting a number of parameters—namely the sieve constants (c_{01}, c_{02}, c_{12}), sieve rate κ, typical and large values for θ, and the threshold ε. The κ achieving the fastest rate of convergence is $\kappa = 1/(1 + 2r)$, although better finite sample properties may be obtained with a larger sieve. In practice we set κ to the (presumed) asymptotically optimal value for discrete inspection processes and closer to $1/2$ under dual right censoring. We have not formally investigated performance for different values of c_{hj}, but this could be set to some positive value invariant to n that represents the presumed degree of non-linearity in Λ_{hj} relative to the other cumulative transition intensity functions. Empirical motivation for these choices is provided in Section 5, but further study is warranted. Our experience suggests that estimates are not particularly sensitive to the choice of the remaining parameters provided that $\mathrm{typ}\,\theta$ is moderately-valued, $\sup\theta$ is relatively large, and ε is sufficiently small. In the simulation studies described below we set $\mathrm{typ}\,\theta = 1$, $\sup\theta = 10$, and $\varepsilon = 10^{-7}$ to ensure convergence within a reasonable number of iterations over the censoring schemes and sample sizes considered.

5 Simulation study

Numerical properties of the piecewise exponential SMLE were investigated for right- and interval-censored variants of dual censoring. In both cases we considered the same model with cumulative $h \to j$, $h \neq j$, transition intensity $\Lambda_{hj}(t)\exp(\theta_1 Z_{01} + \theta_2 Z_{02} + \theta_3 Z_{12})$, where $\Lambda_{01}(t) = t^{4/5}$, $\Lambda_{02}(t) = 3t/4$, $\Lambda_{12}(t) = (3t/2)^{5/4}$, Z uniform on $\{0, 1\}$, Z_{hj} the product of Z and the $h \to j$ transition type indicator, $\theta_1 = \theta_2 = -\log(2)$ and $\theta_3 = 0$. Here Z influences only the exit time from initial state 0 and its effect is the same for each transition type. However neither of these properties were assumed in estimating θ. Throughout $T_{02} \wedge T_{12}$ was right-censored by the fixed time $D = \tau = 2$ representing study closure. Under these fixed parameters, roughly 56% of subjects in the sample progressed ($T_{01} < T_{02}$), 12% were event-free at τ ($T_{01} \wedge T_{02} > \tau$), and 16% survived to study closure ($T_{02} \wedge T_{12} > \tau$).

The censoring scheme acting on the progression status is described in the subsections below, where we summarize findings from 10,000 Monte Carlo replicates of the sample sizes $n = 250, 500, 1000$ under four general scenarios: (1) independent dual censoring, (2) independent dual censoring with increased censoring of progression, (3) conditionally independent dual censoring given Z, and (4) dependent dual censoring. The sieve parameters were held fixed at $c_{hj} = 1$, $\kappa = 2/5$ for dual right-censored data, and $\kappa = 1/3$ for interval-censored progression. The first scenario was revisited with alternative values for κ.

In each scenario we also considered estimates of the Cox model obtained by some combination of early right censoring, mid- or right-endpoint imputation the progression time, or carrying the last negative progression status forward to the final right censoring time or death (LOCF). Details on these alternatives and referenced displays can be found in Appendices E and F of the supplementary material.

5.1 Dual right censoring

To obtain dual right-censored data, an early censoring time C was generated by $C = D = \tau$ with probability $1 - p(Z)$, $\mathrm{logit}(p(z)) = \beta_0 + \beta_1 Z$. Otherwise C followed some distribution with $\Pr(C < D) > 0$. This gave a dual right censoring scheme in which D coincides with administrative censoring and C is a dropout time taking place earlier in the observation period. The four scenarios were respectively specified as: (1) $C = \xi \wedge D$, where ξ is exponential-distributed with mean $\tau/2 = 1$, $e^{\beta_0} = 1/2$ and $e^{\beta_1} = 1$; (2) $C = \xi \wedge D$, $\xi \sim \mathrm{Exponential}(1)$, $e^{\beta_0} = 9$ and $e^{\beta_1} = 1$; (3) $C = \xi \wedge D$, $\xi \sim \mathrm{Exponential}(1)$, $e^{\beta_0} = 1/3$ and $e^{\beta_1} = 3/2$; and (4) $C = (T_{02} \wedge T_{12} - \xi) \wedge D$, where ξ follows

$\text{Exponential}(1)$ truncated to $(0, T_{02} \wedge T_{12})$, $e^{\beta_0} = 1/2$ and $e^{\beta_1} = 1$. The rates of exact observation, singly right-censored and doubly right-censored data were roughly 70, 15 and 15%, respectively, under Scenarios 1 and 3. Under Scenario 2 the censoring rates were approximately 15 and 40%. In Scenario 4 these were 25 and 20%.

The SPMLE based on "singly" right-censored data was also considered for three alternative outcomes given respectively by the observed transition times right-censored by C, PFS right-censored at C, and PFS with LOCF imputation under exactly-observed survival times with unknown progression status. These two variants of PFS are depicted in Fig. 2.

Table E.1 summarizes performance in estimating θ. Results for the SMLE support the asymptotic properties stated in Appendices B and C with average bias generally diminishing in larger samples, average standard error estimates nearing the Monte Carlo sample standard deviations, and empirical coverage probabilities of the 95% confidence intervals at or near the nominal level. The SPMLE from right-censored data at C shows higher variability and lower bias under independent censoring (Scenarios 1 to 3). Under dependent censoring (Scenario 4), the SPMLE has larger finite-sample bias. The SPMLE for PFS right-censored at C also performed relatively well under independent censoring, however its regression coefficient is defined on the basis of the restrictive assumption that $\theta_1 = \theta_2 = \theta$. The PFS variant incorporating LOCF imputation is clearly biased under independent censoring, particularly when the rate of dual censoring is higher. LOCF imputation fared better in Scenario 4. This is not surprising since $C < T_{01} \wedge T_{02}$ often closely preceded T_{02} in this dependent censoring scheme.

Figure E.1 depicts the pointwise average and percentiles of the SMLE $\hat{\Lambda}_n$ under Scenarios 1 to 4 with $n = 1000$. Estimates appear unbiased, with the exception of overestimates for $\hat{\Lambda}_{12}$ early in the observation period. Results under $n = 250$ and 500 (not shown here or the supplement) indicate that bias and variability decreases with increasing sample size, but are otherwise similar. The SPMLE obtained by right censoring observations at C demonstrate little to no bias under independent censoring. This is however not the case under the dependent censoring scheme of Scenario 4. Estimates for Λ_{02} are clearly biased (Figure E.2), with the pointwise 97.5th percentiles consistently smaller than the truth.

From Table E.2, it is apparent that the largest sieve size ($\kappa = 2/5$) achieves the smallest finite-sample bias with little to no increase in variability compared to the sieves under $\kappa = 4/15$ and $1/3$. A larger sieve, sample size or degree of dual censoring increased computational demands for estimation, but the routine we implemented typically converged within a few seconds in all settings considered (Table E.3 of the supplementary material).

5.2 Interval-censored progression times

To generate interval-censored progression times, status was inspected on the basis of m "scheduled" visits, evenly spaced on $(0, \tau)$. "Actual" visit times followed m independent normal distributions centered at the scheduled times with common standard deviation $\sigma_m = \tau/(4(m+1)) = 1/(2(m+1))$ and truncated at zero, τ, and the midpoints between consecutive scheduled times. So the inspection times were continuously distributed on $(0, \tau)$ with greater density around the scheduled targets. This setup is similar to the one in Zeng et al. (2015), however here the spread of the inspection times better cover $(0, \tau)$ so that we can reasonably expect the SMLE to be consistent over the observation period.

Under the independent censoring schemes of Scenarios 1 to 3, every inspection after the first was missed with probability $p(W, Z)$, where $\text{logit}(p(W, Z)) = \beta_0 + \beta_1 Z$. Dependent censoring in Scenario 4 was obtained by discarding inspections taking place after $D = T_{02} \wedge T_{12} - \xi$. In all scenarios, the last observation time $V = T_{02} \wedge T_{12} \wedge D$ offered one further inspection of progression status with a fixed probability of 0.2. Parameters in each scenario were set to: (1) $m = 8$, $e^{\beta_0} = 1/9$ and $e^{\beta_1} = 1$; (2) $m = 4$, $e^{\beta_0} = 1/9$ and $e^{\beta_1} = 1$; (3) $m = 8$, $e^{\beta_0} = 1/4$ and $e^{\beta_1} = 4/9$; and (4) $m = 8$ and ξ follows $\text{Weibull}(3/4, 1)$ truncated to $(0, T_{02} \wedge T_{12})$. With $e^{\beta_0} = 1/9$ and $e^{\beta_1} = 1$ the probability of a missing inspection was $p = 0.1$, irrespective of Z. Under $e^{\beta_0} = 1/4$ and $e^{\beta_1} = 4/9$ the probability remained the same for subjects with $Z = 1$. With $Z = 0$, inspections were two times more likely to be missed. Under Scenario 4, D can be interpreted as a dropout time closely preceding death. In Scenarios 1 and 3 progression status was known by V in just over half of the sample. For Scenarios 2 and 4 this rate was 44% and 40%, respectively. In all four scenarios the rate was 20% among progression-free subjects, so status was known more often among subjects who progressed.

We also fit the SPMLE to two forms of singly right-censored data. The first arises by midpoint-imputing progression times if progression status is known to be positive, as depicted in the lower panel of Fig. 2; otherwise the negative progression status is carried forward to V. The second is the guideline-based definition for PFS, given by the earliest of progression detection, death, and right-censoring at D.

From Table F.1 numerical results for the SMLE $\hat{\theta}_n$ support the asymptotic properties in Appendices B and C with bias generally decreasing with increasing sample size, average standard error estimates reasonably approximating the Monte Carlo sample standard deviations, and empirical coverage probabilities of the 95% confidence intervals close to the nominal level. The SPMLE based on based on midpoint- and LOCF-imputed data and PFS had, on average, larger finite-sample bias. Bias in both of these estimators generally did not diminish with increasing sample size.

Pointwise average and percentiles of the SMLE $\hat{\Lambda}_n$ are depicted in Figure F.1. The SMLE overestimates increments in Λ_{12} early in the observation period. This pattern persists across scenarios and sample sizes, but the bias decreases with larger n. The imputation-based SPMLE for Λ is clearly biased (Figure F.2), with the degree of bias largest under survival-dependent interval censoring (Scenario 4). Imputation-based estimates for survivor and hazard functions typically exhibit a step pattern according to the density of the inspection times, as noted by Panageas et al. (2007). For Λ_{01} this artifact of the observation scheme is a persistent departure from the true shape of the cumulative intensity function.

On average, a smaller sieve with $\kappa = 4/15$ achieved increased bias and similar variability compared to $\kappa = 1/3$ (Table F.2). A larger sieve ($\kappa = 2/5$) gave similar variability, but did not always yield an improvement in average bias. Parameter estimation under interval censoring is more computationally demanding than under dual right-censored data, with average processing times about 100 times slower than those seen in Section 5.1 (Table F.3 of the supplementary material).

6 APPLICATIONS

We return to the examples of Section 2.1, which demonstrate two variants of dual censoring—one arising from loss to follow-up for SREs and the other from periodic assessment for lesion progression. Data were obtained from similarly-designed trials where SREs were recorded at clinic visits every three weeks and lesion progression was diagnosed on the basis of radiographic surveys every three to six months. Actual assessment times roughly followed this schedule, but with enough variation to justify use of the sieve. Since the assessment times were largely determined by a prespecified schedule, one might guess that coarsening at random assumption is plausible. However loss to follow-up for both SREs and lesion progression occurred due to treatment discontinuation and death, although discontinuation rates were similar in the treatment groups. The simulation study offers some reassurance that the SMLE performs relatively well under survival-dependent loss to follow-up. Another consideration is the plausibility of the Markov proportional hazards assumption, though one could argue that this model offers an adequate tool for detecting a difference in the risk of progression between treatment groups. Further investigation of the SMLE's requirements as they relate to the study design and features of the data is warranted, but out of scope for a simple demonstration of the proposed estimator.

Table 1: Regression coefficients for zolendronic acid versus pamidronate specific to first SRE, θ_1, death without SRE, θ_2, and death following SRE, θ_3.

				Early-censored				LOCF
	SMLE			SPMLE			PFS	PFS
	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ	θ
Estimate	0.02	0.16	-0.01	0.02	0.16	-0.01	0.03	0.06
SE	0.11	0.19	0.15	0.11	0.28	0.14	0.10	0.09
p-value	0.89	0.40	0.94	0.86	0.58	0.97	0.78	0.51
HR 95% LCL	0.82	0.80	0.74	0.82	0.67	0.75	0.85	0.89
UCL	1.26	1.72	1.32	1.26	2.04	1.32	1.24	1.27

6.1 DUAL RIGHT CENSORING: SKELETAL-RELATED EVENTS

Rosen et al. (2001) reported that two bisphosphonates, zolendronic acid and pamidronate disodium, showed equivalent efficacy and safety in preventing SREs among patients with breast cancer and multiple myeloma. This conclusion was partly drawn from the evaluation of time to the first SRE within nine months of randomization in an international trial of 1,600 patients. Here we evaluate time to first SRE and death via an illness-death model among the trial's North American breast cancer cohort. Within this subsample of 777 patients, the available trial data provide SREs up to 30 months following randomization. The majority of patients died during this period, so observation of SREs typically ceased earlier.

Under this three-state outcome (T_{01}, T_{02}, T_{12}) was observed exactly in just over one third of the sample. Incomplete transition times and known progression status was observed for 28% of the patients. Almost 15% had unknown progression status but exact survival time, leaving the remaining 23% of the sample dual-censored. Table 1 and Fig. 3 give the SMLE under $c_{hj} = 1$ $(h \neq j)$ and $\kappa = 2/5$. Estimates obtained from both smaller and larger sieves provide similar results, with changes in $\hat{\theta}_n$ less than 0.009. Also depicted is the SPMLE obtained by discarding any observations after the initial right censoring time. The same conclusion can be drawn from both approaches; under the assumed Markov illness-death process, any influence of zoledronic acid on the risk of bone interventions and death is not significantly different from that of pamidronate. The two methods diverge in estimating Λ_{02} (Fig. 3). Since patients near death would presumably be unable to attend clinic visits, early right censoring likely yields underestimates. This may explain why the SPMLE for Λ_{02} is substantially smaller.

6.2 INTERVAL-CENSORED PROGRESSION TIMES: LESION PROGRESSION

Hortobagyi et al. (1996) showed that pamidronate reduced SREs in a placebo-controlled trial of 380 breast cancer patients with bone metastases. Lesion progression was considered as a secondary outcome. This was assessed using radiographic surveys scheduled at three- to six-month intervals over the course of follow-up, rendering the time to lesion progression interval-censored. Surveys were carried out up to 30 months after randomization, but over half of the patients died during this observation period. To account for interval censoring and the occurrence of death, we analyze lesion progression and survival as an illness-death process. Both the progression status and survival time was observed in 28% of the sample. An additional 13% had known progression status but right-censored survival time. In the remaining subjects, right-censoring (11%) or survival (48%) took place long after the last (negative) radiographic survey, resulting in unknown progression status. We defined "long after" as more than six weeks, which enabled us to carry forward recent lesion status to the last observation time. Similar results were obtained by carrying forward fewer weeks. This narrow form of LOCF

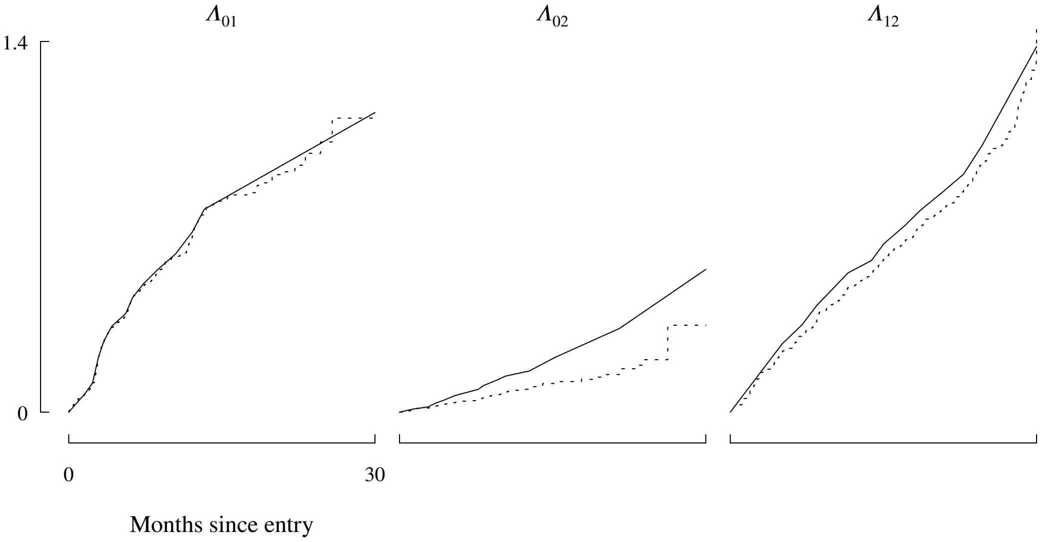

Figure 3: The SMLE (solid) and early-censored SPMLE (dotted) for the cumulative baseline transition intensity functions between study entry (state 0), first SRE (state 1) and death (state 2).

Table 2: Regression coefficients for pamidronate versus control specific to lesion progression, θ_1, death without lesion progression, θ_2, and death following lesion progression, θ_3.

				Imputation-based			
	SMLE			SPMLE			PFS
	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ
Estimate	-0.39	-0.04	-0.05	-0.23	-0.11	-0.03	-0.17
SE	0.18	0.21	0.20	0.17	0.14	0.21	0.11
p-value	0.03	0.85	0.82	0.20	0.42	0.88	0.12
HR 95% LCL	0.47	0.64	0.65	0.57	0.68	0.65	0.68
UCL	0.97	1.45	1.41	1.12	1.18	1.45	1.04

imputation is problematic, but can be avoided when it is possible to randomly assess progression at death.

Table 2 suggests that pamidronate had no influence on mortality, but there is evidence that the bisphosphonate reduces the risk of lesion progression. Based on the SMLE with $c_{hj} = 1$ and $\kappa = 1/3$, an individual treated with pamidronate had 0.68 (95% CI 0.47–0.97) times the rate of progression versus a patient who received placebo. Results obtained under different sieve sizes were similar. The SPMLE from midpoint- and LOCF-imputed data did not detect any significant treatment effect. The difference between the SMLE and SPMLE for the cumulative transition intensities is large (Fig. 4) and likely indicative of bias due to imputation, considering the simulation results in Figure F.2 of the supplementary material.

7 DISCUSSION

This paper examined dual censoring and its challenges for semiparametric maximum likelihood estimation. Methods for special cases of dual-censored data have been previously developed, but the issue of support finding and the resulting imperative for smoothing has not been granted much attention. Our proposed estimator addresses the problem in a general manner, using a model familiar

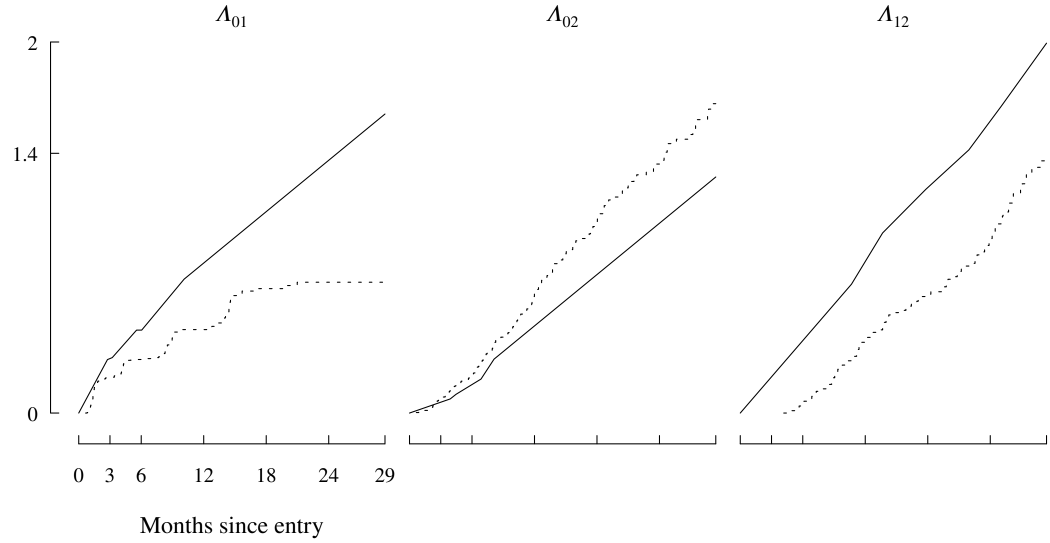

Figure 4: The SMLE (solid) and imputation-based SPMLE (dotted) for the cumulative baseline transition intensity functions between study entry (state 0), lesion progression (state 1) and death (state 2).

to practitioners. The result gives a multistate alternative to PFS that enables separate assessment of treatment effect on progression and survival without progression. A primary assumption of the maximum likelihood approach is that the observation scheme renders the underlying transition times coarsened at random. The simulation study shows that this requirement implies that the estimator is robust to survival-dependent censoring of progression provided that the censoring rate for survival is relatively low.

Supplementary Material

Details on assumptions, proofs, estimation, simulation results are available in the supplementary material at the bottom of this document.

Acknowledgements

The authors thank Novartis Pharmaceuticals for permission to use data from the bone metastases trials and Jerry Lawless for comments on an earlier version of this manuscript. *Conflict of Interest:* None declared.

Funding

This work was supported by Natural Sciences and Engineering Research Council of Canada (RGPIN 155849); Canadian Institutes for Health Research (FRN 13887); Canada Research Chair (Tier 1) - CIHR funded (950-226626).

References

Andersen, P. K. and Borgan, Ø. (1985). Counting process models for life history data: A review. *Scandinavian Journal of Statistics*, 12(2):97–158.

Bebchuk, J. D. and Betensky, R. A. (2001). Local likelihood analysis of survival data with censored intermediate events. *Journal of the American Statistical Association*, 96:449–457.

Boruvka, A. and Cook, R. J. (2015). A Cox-Aalen model for interval-censored data. *Scandinavian Journal of Statistics*, 42(2):414–426.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 34(2):187–220.

FDA (2007). Clinical trial endpoints for the approval of cancer drugs and biologics: Guidance for industry.

Fix, E. and Neyman, J. (1951). A simple stochastic model of recovery, relapse, death and loss of patients. *Human Biology*, 23(3):205–241.

Frydman, H. (1995a). Nonparametric estimation of a Markov 'illness-death' process from interval-censored observations, with application to diabetes survival data. *Biometrika*, 82(4):773–789.

Frydman, H. (1995b). Semiparametric estimation in a three-state duration-dependent Markov model from interval-censored observations with application to AIDS data. *Biometrics*, 51(2):502–511.

Frydman, H. and Szarek, M. (2009). Nonparametric estimation in a Markov 'illness-death' process from interval censored observations with missing intermediate transition status. *Biometrics*, 65(1):143–151.

Grenander, U. (1981). *Abstract Inference*. Wiley.

Heitjan, D. F. and Rubin, D. B. (1991). Ignorability and coarse data. *Annals of Statistics*, 19(4):2244–2253.

Hortobagyi, G. N., Theriault, R. L., Porter, L., Blayney, D., Lipton, A., Sinoff, C., Wheeler, H., Simeone, J. F., Seaman, J., Knight, R. D., Heffernan, M., Reitsma, D. J., Kennedy, I., Allan, S. G., Mellars, K., and the Protocol 19 Aredia Breast Cancer Study Group (1996). Efficacy of pamidronate in reducing skeletal complications in patients with breast cancer and lytic bone metastases. *New England Journal of Medicine*, 335(24):1785–1792.

Hu, C. and Tsodikov, A. (2014). Joint modeling approach for semicompeting risks data with missing nonterminal event status. *Lifetime Data Analysis*, 20(4):563–583.

Huang, J. and Rossini, A. J. (1997). Sieve estimation for the proportional-odds failure-time regression model with interval censoring. *Journal of the American Statistical Association*, 92(439):960–967.

Jackson, C. (2011). Multi-state models for panel data: The msm package for R. *Journal of Statistical Software*, 38(8):1–28.

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

Murphy, S. A. and van der Vaart, A. W. (2000). On profile likelihood. *Journal of the American Statistical Association*, 95(450):449–465.

Panageas, K. S., Ben-Porat, L., Dickler, M. N., Chapman, P. B., and Schrag, D. (2007). When you look matters: The effect of assessment schedule on progression-free survival. *Journal of the National Cancer Institute*, 99(6):428–432.

Rosen, L. S., Gordon, D., Kaminski, M., Howell, A., Belch, A., Mackey, J., Apffelstaedt, J., Hussein, M., Coleman, R. E., Reitsma, D. J., Seaman, J. J., Chen, B. L., and Ambros, Y. (2001). Zoledronic acid versus pamidronate in the treatment of skeletal metastases in patients with breast cancer or osteolytic lesions of multiple myeloma: A phase iii, double-blind, comparative trial. *Cancer Journal*, 7(5):377–387.

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

Wong, G. Y. C. and Yu, Q. (1999). Generalized MLE of a joint distribution function with multivariate interval-censored data. *Journal of Multivariate Analysis*, 69(2):155–166.

Xu, J., Kalbfleisch, J. D., and Tai, B. (2010). Statistical analysis of illness-death processes and semicompeting risks data. *Biometrics*, 66(3):716–725.

Zeng, L., Cook, R. J., Wen, L., and Boruvka, A. (2015). Bias in progression-free survival analysis due to intermittent assessment of progression. *Statistics in Medicine*, 34(24):3181–3193.

Zhang, Y., Hua, L., and Huang, J. (2010). A spline-based semiparametric maximum likelihood estimation method for the Cox model with interval-censored data. *Scandinavian Journal of Statistics*, 37(2):338–354.

Web-based Supplementary Materials for
Sieve estimation in a Markov illness-death process under dual censoring

AUDREY BORUVKA

Department of Statistics and Actuarial Science,

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

RICHARD J. COOK

Department of Statistics and Actuarial Science,

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

E-mail: rjcook@uwaterloo.ca

LIST OF APPENDICES

SIMULATION RESULTS — LIST OF FIGURES

SIMULATION RESULTS — LIST OF TABLES

A BASIC ASSUMPTIONS

Let J be a process indicating the times at which progression status is under assessment. Then, for an observation dual-right–censored by the loss to follow-up times $C < D$, we have $J(t) = 1$ for $t \in [0, C]$. Interval-censored progression times arise when J represents a discrete inspection process taking the value one only at a random number K of assessment times $Y_{K,1}, \ldots, Y_{K,K}$. For brevity here and in the sequel, let $S = T_{01} \wedge T_{02}$ be the exit time from the initial state, $T = T_{02} \wedge T_{12}$ be the survival time, $\alpha_{hj}(t \mid Z) = \lambda_{hj}(t) \exp(Z'_{hj}\theta)$, $A_{hj} = \int \alpha_{hj}$, F_ξ is the distribution function of the random variable ξ. For clarity, we now indicate the dependence of the $h \to j$ transition probability over $(s, t]$ with $P_{hj}(s, t \mid Z)$.

CONDITION A.1 The distribution of $\{J(t) : 0 \le t \le \tau\}$ and D are such that the conditional probability of the events $\{\Delta_0 = 1\}$, $\{\Delta_1 = 1\}$, $\{\Delta_0\Delta_2 = 1\}$, $\{\Delta_1\Delta_2 = 1\}$, and $\{\Delta_0 = 1, S > \tau = D\}$ given Z are almost surely positive and (σ, τ) is a subset of $\{t \in (0, D] : \Pr(J(t) = 1 \mid Z) > 0, \text{almost surely}\}$.

CONDITION A.2 For a random sample of dual-censored observations X_i ($i = 1, \ldots, n$), let $n_{01} = \sum_{i=1}^n \Delta_1^i$, $n_{02} = \sum_{i=1}^n \Delta_0^i \Delta_2^i$ and $n_{12} = \sum_{i=1}^n \Delta_1^i \Delta_2^i$. There exist $\pi_{hj} > 0$, $h \neq j$, such that $n_{hj}/n \to \pi_{hj}$ as $n \to \infty$.

CONDITION A.3 Let Θ be a compact subset of $\mathbb{R}^d$ and $H = (H_{hj})_{h \neq j}$ the set of cumulative $h \to j$ transition intensity functions $\Lambda = (\Lambda_{hj})_{h \neq j}$ with $\Lambda_{hj}(0) \equiv 0$, $\Lambda_{hj}(\infty) \equiv \infty$ and $1/M < \Lambda_{hj}(\sigma-) < \Lambda_{hj}(\tau) < M$ for some fixed $0 < M < \infty$. The true parameter (θ_0, Λ_0) belongs to $\Theta \times H$ with θ_0 an interior point of Θ and Λ_0 having continuous bounded derivative λ_0 on $[0, \tau]$.

CONDITION A.4 (Coarsening at random) The observation X is a "coarsening" of (S, T) arising from some partially observable random variable G whose conditional distribution given (S, T, Z) is specified by a parameter distinct from (θ, Λ) and is invariant with respect to all (s, t) compatible with X. So for each (s, t) such that $s \le t$, $s \in (L, t]$ if $\Delta_1 = 0$, $s \in (L, R]$ if $\Delta_1 = 1$, $t = V$ if $\Delta_2 = 1$ and $t \in (V, \infty)$ if $\Delta_2 = 0$. Moreover our ability to observe the progression time exactly $\Delta_1 1(L = R+)$ is conditionally independent of (S, T) given Z.

Condition A.1 ensures that the times at which we assess progression status become dense in the observation period. Condition A.4 implies that (S, T) is coarsened at random in the sense of **?**] and **?** , p. 274]. This can be motivated by assuming conditional independence between (S, T) and $(\{J(t) : 0 \le t \le \tau\}, D)$ given X. The coarsening element G in the dual right censoring case corresponds to (C, D). Under interval-censored progression times, $G = (K, Y_{K,1}, \ldots, Y_{K,K}, D)$.

B CONSISTENCY AND RATE OF CONVERGENCE

Consistency of the SMLE follows by way of **?** , Theorem 3.4.1], a result commonly used to derive the global rate of convergence for a sieve maximum likelihood estimator.

CONDITION B.1 The distribution of Z has support on a bounded subset of $\mathbb{R}^d$. For each $h \neq j$, $\Pr(Z'_{hj}a \neq c) > 0$ for every $a \in \mathbb{R}^d$ and $c \in \mathbb{R}$.

CONDITION B.2 For $k = 1$ or 2, the kth derivative of Λ_0 continuous, positive and bounded on $[\sigma, \tau]$.

THEOREM B.3 Let $\|\hat{\Lambda}_n - \Lambda_0\|_2 = \sum_{h \neq j} (\int_\sigma^\tau |\hat{\Lambda}_{hj,n} - \Lambda_{hj}^0|^2(u)\,\mathrm{d}u)^{1/2}$ be the L_2-distance between $\hat{\Lambda}_n$ and Λ_0 on (σ, τ). Under the above conditions $\|\hat{\theta}_n - \theta_0\| + \|\hat{\Lambda}_n - \Lambda_0\|_2 \to 0$ at the rate $O_P(\max(n^{(1-\kappa)/2}, n^{k\kappa}))$.

Let $(\Lambda^0_{hj,n})_{h\neq j} = \Lambda_{0,n} \in H_n$ be the piecewise linear interpolant of Λ_0 given by

$$\Lambda_{hj,n}(t) = \sum_{t_k \in \mathcal{T}_{hj,n}} I_k(t) \left\{ \left(1 - \frac{L_k(0,t)}{L_k(0,\tau)}\right) \Lambda^0_{hj}(t_{k-1}) + \frac{L_k(0,t)}{L_k(0,\tau)} \Lambda^0_{hj}(t_k) \right\},$$

where $I_k(t) = 1_{[t_{k-1},t_k)}(t)$ and $L_k(s,t)$ is the length of $[t_{k-1},t_k) \cap [s,t)$. Then the distance between the SMLE $(\hat{\theta}_n, \hat{\Lambda}_n)$ and the true parameter value (θ_0, Λ_0) has order

$$\|\hat{\theta}_n - \theta_0\| + \|\hat{\Lambda}_n - \Lambda_{0,n}\|_2 + \|\Lambda_{0,n} - \Lambda_0\|_2. \tag{B.1}$$

Taylor expansion shows that the third term is $O(n^{-k\kappa})$, where $k = 1$ or 2 as assumed in Condition B.2. An order for the first two terms is derived by application of **?** , Theorem 3.4.1], considering the SMLE as an M-estimator under the criterion $m_{\theta,\Lambda} = \log((\mathrm{lik}_{\theta,\Lambda} + \mathrm{lik}_{0,n})/2)$ with the subscript $0,n$ a shorthand for $\theta_0, \Lambda_{0,n}$. Since the logarithm is concave and the SMLE is a maximizer of the log-likelihood function $n\,\mathbb{P}_n \log \mathrm{lik}_{\theta,\Lambda}$ on the sieve $\Theta \times H_n$,

$$\mathbb{P}_n\Big(m_{\hat{\theta}_n,\hat{\Lambda}_n} - m_{0,n}\Big) = \mathbb{P}_n \log\left(\frac{\mathrm{lik}_{\hat{\theta}_n,\hat{\Lambda}_n} + \mathrm{lik}_{0,n}}{2\,\mathrm{lik}_{0,n}}\right) \geq \tfrac{1}{2}\,\mathbb{P}_n\Big(\log \mathrm{lik}_{\hat{\theta}_n,\hat{\Lambda}_n} - \log \mathrm{lik}_{0,n}\Big) \geq 0.$$

So the SMLE is a "near maximizer" of $\mathbb{P}_n\, m_{\theta,\Lambda}$.

Let $K_n = \sum_{h\neq j} K_{hj,n}$. Then from Condition A.3, $\Theta \times H_n$ is a compact parametric class with bracketing number

$$N_{[]}(\varepsilon, \Theta \times H_n, L_2(\mathbb{P}_n)) \lesssim (\mathrm{diam}\,\Theta/\varepsilon)^d (M/\varepsilon)^{K_n}.$$

The corresponding bracketing integral is

$$J_{[]}(\delta, \Theta \times H_n, L_2(\mathbb{P}_n)) = \int_0^\delta \sqrt{\log N_{[]}(\varepsilon, \Theta \times H_n, L_r(\mathbb{P}_n))}\,\mathrm{d}\varepsilon \lesssim \delta\sqrt{d + K_n}. \tag{B.2}$$

This is finite, so by **?** , Theorem 19.5] $\Theta \times H_n$ is P-Donsker for each n. The criterion function is pointwise Lipschitz in the transition intensities, which implies that $\{m_{\theta,\Lambda} : \theta \in \Theta, \Lambda \in H_n\}$ is also P-Donsker. Since $\log a \leq 2(\sqrt{a} - 1)$ for $a \geq 0$,

$$\begin{aligned} P(m_{\theta,\Lambda} - m_{0,n}) &\leq 2 \int \sqrt{\frac{(\mathrm{lik}_{\theta,\Lambda} + \mathrm{lik}_{0,n})\,\mathrm{lik}_{0,n}}{2}}\,\mathrm{d}\nu - 2 \\ &\leq -\int \Big(\sqrt{\mathrm{lik}_{\theta,\Lambda}} - \sqrt{\mathrm{lik}_{0,n}}\Big)^2 \mathrm{d}\nu \lesssim -\|\theta - \theta_0\|^2 - \|\Lambda - \Lambda_{0,n}\|_2^2. \end{aligned}$$

The upper bound in the last inequality corresponds to the negative Hellinger distance between $\mathrm{lik}_{\theta,\Lambda}$ and $\mathrm{lik}_{0,n}$, which is zero only if $(\theta,\Lambda) = (\theta_0, \Lambda_{0,n})$ by Lemma B.4 below. From Lemma B.5 we obtain the inequality up to a constant. Of the remaining requirements for **?** , Theorem 3.4.1], we need the "modulus of continuity" $\phi_n(\delta)$ of the centered process $\sqrt{n}(\mathbb{P}_n - P)m_{\theta,\Lambda}$ over the sieve $\theta \times H_n$. From (B.2) and **?** , Lemma 3.4.2] $\phi_n(\delta)$ has order $\delta n^{\kappa/2} + n^\kappa/\sqrt{n}$. We need $\phi_n(\delta_n) \leq \sqrt{n}\delta_n^2$, which is satisfied with equality by $\delta_n = n^{-(1-\kappa)/2}/2$. Returning to (B.1) the rate at which the SMLE converges to (θ_0, Λ_0) is thus $\min(n^{(1-\kappa)/2}, n^{k\kappa})$.

LEMMA B.4 For every $(\theta,\Lambda) \neq (\theta_0, \Lambda_{0,n})$ on (σ,τ), $\mathrm{lik}_{\theta,\Lambda} \neq \mathrm{lik}_{0,n}$, almost surely.

Proof. Assume that $\mathrm{lik}_{\theta,\Lambda} = \mathrm{lik}_{0,n}$, almost surely. Then by Conditions A.1 and A.3 and Duhamel's equation [e.g. **?** , Theorem 6],

$$0 = \big|P_{00}(0,\tau \mid Z) - P^0_{00,n}(0,\tau \mid Z)\big| = \int_0^\tau P_{00}(0,u \mid Z)\big|A_{00} - A^0_{00,n}\big|(\mathrm{d}u \mid Z)P^0_{00,n}(u,\tau \mid Z),$$

almost surely. This is satisfied only if $A_{00} = A^0_{00,n}$ almost surely on $(0,\tau)$. Then our assumption and the same conditions further imply that $A_{02} = A^0_{02,n}$ almost surely on $(0,\tau)$. Put $t^* = \inf\{\sigma \le t < \tau : \Lambda^0_{02,n}(t) > 0\}$. Thus $e^{Z'_{02}(\theta-\theta_0)} = \Lambda_{02}(t^*)/\Lambda^0_{02,n}(t^*)$ and $Z'_{02}(\theta-\theta_0)$ is degenerate. Under Condition B.1 this implies that $\theta = \theta_0$ and thus $\Lambda_{02} = \Lambda^0_{02,n}$. This in turn gives $\Lambda_{01} = \Lambda^0_{01,n}$ and $\Lambda_{12} = \Lambda^0_{12,n}$ on (σ,τ). □

LEMMA B.5 Under the previous conditions $\int(\sqrt{\mathrm{lik}_{\theta,\Lambda}} - \sqrt{\mathrm{lik}_{0,n}})^2\,\mathrm{d}\nu \gtrsim \|\theta-\theta_0\|^2 + \|\Lambda - \Lambda_{0,n}\|^2$.

Proof. Since $\mathrm{lik}_{\theta,\Lambda} + \mathrm{lik}_{0,n}$ can be uniformly bounded under Conditions A.3 and B.1, the Hellinger distance has lower bound up to a constant

$$\int\Big(\sqrt{\mathrm{lik}_{\theta,\Lambda}} - \sqrt{\mathrm{lik}_{0,n}}\Big)^2\,\mathrm{d}\nu = \int \frac{(\mathrm{lik}_{\theta,\Lambda} - \mathrm{lik}_{0,n})^2}{(\sqrt{\mathrm{lik}_{\theta,\Lambda}} + \sqrt{\mathrm{lik}_{0,n}})^2}\,\mathrm{d}\nu \gtrsim \int (\mathrm{lik}_{\theta,\Lambda} - \mathrm{lik}_{0,n})^2\,\mathrm{d}\nu.$$

Let $\mathcal{Z}$ denote the support of the distribution. By Conditions A.1 and B.2,

$$\begin{aligned}
&\int (\mathrm{lik}_{\theta,\Lambda} - \mathrm{lik}_{0,n})^2\,\mathrm{d}\nu \\
&\ge p_{02}\int_{\mathcal{Z}}\int_{0,n}^{\tau}\Big(P_{00}(0,s\mid z)A_{02}(\mathrm{d}s\mid z) - P^0_{00,n}(0,s\mid z)A^0_{02,n}(\mathrm{d}s\mid z)\Big)^2\,\mathrm{d}F_Z(z)\\
&\ge p_{02}\int_{\mathcal{Z}}\int_{0}^{\tau}(P_{00}-P^0_{00,n})^2(0,s\mid z)\,\mathrm{d}A^0_{02,n}(s\mid z)^2\,\mathrm{d}F_Z(z)\\
&\gtrsim \int_{\mathcal{Z}}\int_{\sigma}^{\tau}(P_{00}-P^0_{00,n})^2(0,s\mid z)\,\mathrm{d}s\,\mathrm{d}F_Z(z)\\
&= \int_{\mathcal{Z}}\int_{0}^{\tau}\Big(\int_0^s P_{00}(0,s\mid z)(A_{00}-A^0_{00,n})(\mathrm{d}u\mid z)P^0_{00,n}(u,s\mid z)\Big)^2\,\mathrm{d}s\,\mathrm{d}F_Z(z)\\
&\gtrsim \int_{\mathcal{Z}}\int_{\sigma}^{\tau}\Big(\int_0^s (A_{00}-A^0_{00,n})(\mathrm{d}u\mid z)\Big)^2\,\mathrm{d}s\,\mathrm{d}F_Z(z),\\
&\ge \int_{\mathcal{Z}}\int_{\sigma}^{\tau}(A_{0j}-A^0_{0j,n})^2(\mathrm{d}s\mid z)\,\mathrm{d}F_Z(z), \quad j=1,2,
\end{aligned}$$

where the inequalities up to a constant holds because p_{02}, $\alpha^0_{02,n}$, P_{hh} and $P^0_{hh,n}$ are bounded away from zero on $[\sigma,\tau]$ and the equality follows from Duhamel's equation. Similarly

$$\begin{aligned}
&\int (\mathrm{lik}_{\theta,\Lambda} - \mathrm{lik}_{0,n})^2\,\mathrm{d}\nu \\
&\ge p_{12}\int_{\mathcal{Z}}\int_0^{\tau}\int_s^{\tau}\Big(P_{00}(0,s\mid z)\alpha_{01}(s\mid z)P_{11}(s,t\mid z)A_{12}(\mathrm{d}t\mid z)\\
&\qquad\qquad - P^0_{00,n}(0,s\mid z)\alpha^0_{01,n}(s\mid z)P^0_{11,n}(s,t\mid z)A^0_{12,n}(\mathrm{d}t\mid z)\Big)^2\,\mathrm{d}F_Z(z)\\
&\ge p_{12}\int_{\mathcal{Z}}\int_0^{\tau}\int_s^{\tau}\Big(P_{11}(s,t\mid z)-P^0_{11,n}(s,t\mid z)\Big)^2\Big(P^0_{00,n}(0,s\mid z)\alpha^0_{01,n}(s\mid z)A^0_{12,n}(\mathrm{d}t\mid z)\Big)^2\,\mathrm{d}F_Z(z)\\
&\gtrsim \int_{\mathcal{Z}}\int_{\sigma}^{\tau}\int_s^{\tau}\Big(P_{11}(s,t\mid z)-P^0_{11,n}(s,t\mid z)\Big)^2\,\mathrm{d}t\,\mathrm{d}s\,\mathrm{d}F_Z(z)\\
&= \int_{\mathcal{Z}}\int_{\sigma}^{\tau}\int_s^{\tau}\Big(\int_s^t P_{11}(s,u\mid z)(A_{11}-A^0_{11,n})(\mathrm{d}u\mid z)P^0_{11,n}(u,t\mid z)\Big)^2\,\mathrm{d}t\,\mathrm{d}s\,\mathrm{d}F_Z(z)\\
&\gtrsim \int_{\mathcal{Z}}\int_{\sigma}^{\tau}\int_s^{\tau}\Big(\int_s^t (A_{11}-A^0_{11,n})(\mathrm{d}u\mid z)\Big)^2\,\mathrm{d}t\,\mathrm{d}s\,\mathrm{d}F_Z(z)
\end{aligned}$$

$$\geq \int_{\mathcal{Z}} \int_{\sigma}^{\tau} (A_{12} - A_{12,n}^{0})^2 (\mathrm{d}s \mid z) \, \mathrm{d}F_Z(z).$$

Let $\theta_t = t\theta + (1-t)\theta_0$, $\Lambda_{hj,t} = t\Lambda_{hj} + (1-t)\Lambda_{hj,n}^0$. From the mean value theorem, there is some $t \in (0,1)$ depending on (y, z, h, j) such that

$$\begin{aligned}(A_{hj} - A_{hj,n}^0)(y \mid z) &= \frac{\partial}{\partial t} A_{hj,t}(y \mid z) \\ &= t\exp(z_{hj}'\theta_t)\big\{1 + z_{hj}'(\theta-\theta_0)t\big\}(\Lambda_{hj} - \Lambda_{hj,n}^0)(y) + t\exp(z_{hj}'\theta_t) z_{hj}'(\theta-\theta_0)\Lambda_{hj,n}^0(y).\end{aligned}$$

For $(Y, Z) \sim \mu = 1_{[\sigma,\tau]} \times F_Z$, put $f_{hj,1}(Z) = 1 + Z_{hj}'(\theta-\theta_0)t$, $f_{hj,2}(Z) = (\Lambda_{hj} - \Lambda_{hj,n}^0)(Z)$ and $f_{hj,3}(Z, Z) = Z_{hj}'(\theta-\theta_0)\Lambda_{hj,n}^0(Z)$. So $(A_{hj} - A_{hj,n}^0)(Y \mid Z)$ is equal to $f_{hj,1}(Z) f_{hj,2}(Y) + f_{hj,3}(Y, Z)$ up to the factor $t\exp(Z_{hj}'\theta_t)$, which is bounded away from zero under Conditions A.3 and B.1. Also by Condition B.1,

$$\big(\mathrm{E}_\mu(f_{hj,2} f_{hj,3})\big)^2 < \mathrm{E}_\mu(f_{hj,2}^2)\,\mathrm{E}_\mu(f_{hj,3}^2).$$

Since $f_{hj,1}(z)$ is uniformly close to 1 for θ close to θ_0 and $\Lambda_{hj,n}^0$ is bounded away from zero on $[\sigma,\tau]$,

$$\int (\mathrm{lik}_{\theta,\Lambda} - \mathrm{lik}_{0,n})^2 \, \mathrm{d}\nu \gtrsim \mu(f_{hj,1} f_{hj,1} + f_{hj,3})^2 \gtrsim \mu f_{hj,3}^2 + \mu f_{hj,2}^2 \gtrsim \|\theta - \theta_0\|^2 + \|\Lambda - \Lambda_{0,n}\|^2,$$

by **?** , Lemma A.6]. □

C ASYMPTOTIC NORMALITY

The following result is a version of **?** , Corollary 3] that enables us to consistently estimate the standard error of $\hat\theta_n$ by approximating the curvature of the profile log-likelihood using tuning parameters specific to the entries of θ.

CONDITION C.1 There is some $y_0 > 0$ such that $L < R-$ implies $R - L > y_0$, almost surely.

THEOREM C.2 Let k be the order of the derivative of Λ_0 satisfying Condition B.2. If $1/(4k) < \kappa < 1/2$ then, under the above conditions, the sequence $\sqrt{n}(\hat\theta_n - \theta_0)$ is asymptotically normal with mean zero and variance equal to the inverse of the efficient information matrix $\tilde I_0$. Moreover for any symmetric matrix d-matrix h_n whose entries h_n^{ij} tend to zero in probability as $n \to \infty$ such that $(h_n^{ij}\sqrt{n})^{-1} = o_P(1)$,

$$\begin{aligned}\frac{1}{n(e_i'h_ne_i)^2}\Big(\ell_n^p(\hat\theta_n + e_i'h_ne_i) - \ell_n^p(\hat\theta_n)\Big) + \frac{1}{n(e_j'h_ne_j)^2}\Big(\ell_n^p(\hat\theta_n + e_j'h_ne_j) - \ell_n^p(\hat\theta_n)\Big)& \\ - \frac{1}{n(e_i'h_ne_j)^2}\Big(\ell_n^p(\hat\theta_n + e_i'h_ne_j) - \ell_n^p(\hat\theta_n)\Big)& \quad \text{(C.1)}\end{aligned}$$

is a consistent estimator for the (i,j)th entry of $\tilde I_0$ $(i, j = 1, \ldots, d)$.

To prove that this holds we show existence of a least favorable submodel meeting the requirements of **?** , Theorem 1] under a single covariate $d = 1$. The same result for $d > 1$ follows by repeated application of this special case. The latter assumption in Condition A.4 enables us to consider only the two extreme forms of dual censoring (namely dual right censoring and interval-censored progression times). We consider these types of dual censoring separately.

C.1 DUAL RIGHT CENSORING

Let $U = S \wedge C$, where $C \leq D$ is the right-censoring time for S. Then $\varDelta_1 = 1(U = S < V)$ and $\varDelta_0 = 1(U = V)$. The log-likelihood of a single realization x can be re-expressed as

$$\begin{aligned}\ell_{\theta,\varLambda}(x) = &- A_{01}(u \mid z) - A_{02}(u \mid z) + \delta_1(\log \alpha_{01}(u \mid z) - A_{12}(u, v \mid z) + \delta_2 \log \alpha_{12}(v \mid z)) \\ &+ (1 - \delta_0\delta_1) \log\bigl(P_{01}(u, v \mid z)\alpha_{12}(v \mid z)^{\delta_2} + P_{00}(u, v \mid z)\alpha_{02}(v \mid z)^{\delta_2}\bigr) - \delta_0\delta_2 \log \alpha_{02}(v \mid z).\end{aligned}$$

A score for θ is defined the usual way: $\partial\ell_{\theta,\varLambda}/\partial\theta \equiv \dot{\ell}_{\theta,\varLambda}$. For $\varLambda$, we differentiate with respect to a d-dimensional submodel $t \to \varLambda_t$ that indexes the "direction" in which $\varLambda_t \in H_n$ approaches $\varLambda_0$. Since (4.1) is a multiplicative intensity model, the baseline cumulative intensity functions are variation independent. Thus we can consider a one-dimensional submodel $y \to \varLambda_{hj,y}$ for each $h \neq j$. For now we will simply assume that g_{hj} is defined so that $\partial/\partial y_{|y=0}\varLambda_{hj,y} = \int g_{hj}\, \mathrm{d}\varLambda_{hj}$. Then, for $h = 0, 1$,

$$\frac{\partial}{\partial y}_{|y=0} P_{hh,y}(s, t \mid z) = -P_{hh}(s, t \mid z) \sum_{j \neq h} \int_s^t g_{hj}(y)\, \mathrm{d}A_{hj}(y \mid z).$$

For the $0 \to 1$ transition probability we have

$$\begin{aligned}&\frac{\partial}{\partial y}_{|y=0} P_{01,y}(s, t \mid z) \\ &= \frac{\partial}{\partial t}_{|y=0} \int_s^t P_{00,y}(s, y \mid z)\, \mathrm{d}A_{01,y}(y \mid z) P_{11,y}(y, t \mid z) \\ &= -\int_s^t P_{00}(s, u) \left(\sum_{j=1,2} \int_s^u g_{0j}(y)\, \mathrm{d}A_{0j}(y \mid z) + \int_u^t g_{12}(y)\, \mathrm{d}A_{12}(y \mid z)\right) g_{01}(u)\, \mathrm{d}A_{01}(u \mid z) P_{11}(u, t \mid z) \\ &= \int_s^t P_{00}(s, y \mid z) g_{01}(y)\, \mathrm{d}A_{01}(y \mid z) P_{11}(y, t \mid z) - \int_s^t P_{00}(s, y \mid z) P_{01}(y, t \mid z) \sum_{j=1,2} g_{0j}(y)\, \mathrm{d}A_{0j}(y \mid z) \\ &\quad - \int_s^t P_{01}(s, y \mid z) g_{12}(y)\, \mathrm{d}A_{12}(y \mid z) \\ &\equiv \sum_{h \neq j} \int_s^t g_{hj}(y) \varphi_{hj}(ys, t, z)\, \mathrm{d}A_{hj}(y \mid z),\end{aligned}$$

where the third equality is obtained by interchanging integrals. Put $g = (g_{hj})_{h \neq j}$. Then a score function for $\varLambda$ is

$$\begin{aligned}B_{\theta,\varLambda} g(x) \equiv &\frac{\partial}{\partial t}_{|t=0} \ell_{\theta,\varLambda_t}(x) \\ = &- \sum_{j=1,2} \int_0^u g_{0j}(y)\, \mathrm{d}A_{0j}(y \mid z) + \delta_1 \left(g_{01}(u) - \int_u^v g_{12}(y)\, \mathrm{d}A_{12}(y \mid z) + \delta_2 g_{12}(v)\right) \\ &+ (1 - \delta_0\delta_1) \frac{P_{00}(u, v \mid z)\alpha_{02}(v \mid z)^{\delta_2} \left(\delta_2 g_{02}(v) - \sum_{j=1,2} \int_u^v g_{0j}(y)\, \mathrm{d}A_{0j}(y \mid z)\right)}{\sum_{j=0,1} P_{0j}(u, v \mid z)\alpha_{j2}(v \mid z)^{\delta_2}} \\ &+ (1 - \delta_0\delta_1) \frac{P_{01}(u, v \mid z)\alpha_{12}(v \mid z)^{\delta_2} \left(\sum_{h \neq j} \int_u^v g_{hj}(y)\varphi_{hj}(ys, t, z)\, \mathrm{d}A_{hj}(y \mid z) + \delta_2 g_{12}(v)\right)}{\sum_{j=0,1} P_{0j}(u, v \mid z)\alpha_{j2}(v \mid z)^{\delta_2}} \\ &- \delta_0\delta_2 g_{02}(v \mid z).\end{aligned}$$

A score for $\theta \in \mathbb{R}$ has the same form, but with z_{hj} in place of each $g_{hj}(\cdot)$. If $B^*_{\theta,\varLambda}$ is the adjoint of the score operator $B_{\theta,\varLambda}$, then a least favorable direction $g_{\theta,\varLambda}$ satisfies $B^*_{\theta,\varLambda}\dot{\ell}_{\theta,\varLambda} = B^*_{\theta,\varLambda} B_{\theta,\varLambda} g_{\theta,\varLambda}$, $h \neq j$.

Under an information loss model (**?** , Section 25.5.2; **?** , Theorem 4.1) with "unobservable" (S, T), the adjoint is the conditional expectation operator given $\{S = s, T = t\}$. Under Condition A.4 the conditional expectation of the score for Λ given $\{S = s, T = t, Z = z\}$ is

$$
\begin{aligned}
&\mathrm{E}(B_{\theta,\Lambda} g(X) \mid S = s, T = t, Z = z) \\
=& -\sum_{j=1,2} \int_0^s g_{0j}(y)\,\mathrm{d}A_{0j}(y \mid z) \bar{F}_{C|Z}(y \mid z) \\
&+ 1(s < t)\Big(g_{01}(s) \bar{F}_{C|Z}(s \mid z) - \int_s^t g_{12}(y)\,\mathrm{d}A_{12}(y \mid z) \bar{F}_{C,D|Z}(s, y \mid z) + g_{12}(t) \bar{F}_{C,D|Z}(s, t \mid z) \Big) \\
&+ g_{02}(t) \int_0^s \frac{P_{00}(c, t \mid z) \alpha_{02}(t \mid z)^{\delta_2}}{\sum_{j=0,1} P_{0j}(c, t \mid z) \alpha_{j2}(t \mid z)^{\delta_2}} \bar{F}_{D|C,Z}(t \mid c, z)\,\mathrm{d}F_{C|Z}(c, d \mid z) \\
&- \sum_{j=1,2} \int_\sigma^\tau g_{0j}(y)\,\mathrm{d}A_{0j}(y \mid z) \int_0^s \int_c^\tau \frac{P_{00}(c, v \mid z) \alpha_{02}(v \mid z)^{\delta_2}}{\sum_{j=0,1} P_{0j}(c, v \mid z) \alpha_{j2}(v \mid z)^{\delta_2}}\,\mathrm{d}F_{C,D|Z}(c, d \mid z) \\
&+ g_{12}(t) \int_0^s \frac{P_{01}(c, t \mid z) \alpha_{12}(t \mid z)^{\delta_2}}{\sum_{j=0,1} P_{0j}(c, t \mid z) \alpha_{j2}(t \mid z)^{\delta_2}} \bar{F}_{D|C,Z}(t \mid c, z)\,\mathrm{d}F_{C|Z}(c, d \mid z) \\
&+ \sum_{h \neq j} \int_\sigma^\tau g_{hj}(y)\,\mathrm{d}A_{hj}(y \mid z) \int_0^s \int_c^\tau \frac{P_{01}(c, v \mid z) \alpha_{12}(v \mid z)^{\delta_2}}{\sum_{j=0,1} P_{0j}(c, v \mid z) \alpha_{j2}(v \mid z)^{\delta_2}} \varphi_{hj}(yc, t, z)\,\mathrm{d}F_{C,D|Z}(c, d \mid z) \\
&- 1(s = t) g_{02}(t) \bar{F}_{C,D|Z}(t, t \mid z) \\
\equiv& \; 1(s < t) g_{01}(s) C_{01}(s, t, z) + \int_0^s g_{01}(y) Q_{01}(y, t, z)\,\mathrm{d}\Lambda_{01}(y) + \sum_{h=0,1} g_{h2}(t) C_{h2}(s, t, z) \\
&+ \sum_{h \neq j} \int_\sigma^\tau g_{hj}(y) R_{hj}(y, s, t, z)\,\mathrm{d}\Lambda_{hj}(y) \\
\equiv& \sum_{h \neq j} B_{hj}(s, t, z),
\end{aligned}
$$

where $\bar{F}_\xi = 1 - F_\xi$. Similarly

$$
\begin{aligned}
\mathrm{E}(\dot{\ell}_{\theta,\Lambda}(X) \mid S = s, T = t, Z = z) &\equiv 1(s < t) z_{01} C_{01}(s, t, z) + \int_0^s z_{01} Q_{01}(y, t, z)\,\mathrm{d}\Lambda_{01}(y) \\
&\quad + \sum_{h=0,1} z_{h2} C_{h2}(s, t, z) + \sum_{h \neq j} \int_\sigma^\tau z_{hj} R_{hj}(y, s, t, z)\,\mathrm{d}\Lambda_{hj}(y) \\
&\equiv \sum_{h \neq j} L_{hj}(s, t, z).
\end{aligned}
$$

Put $b_{hj}(s, t) = \mathrm{E}_Z(B_{hj}(s, t, Z))$ and $l_{hj}(s, t) = \mathrm{E}_Z(L_{hj}(s, t, Z))$. Let $g_{hj,0}$, $h \neq j$, denote the solution to the integral equations $b_{01}(s, t) = l_{01}(s, t)$ on $s < t$, $\partial b_{01}(s, t)/\partial s = \partial l_{01}(s, t)/\partial s$ on $s = t$, $b_{02}(s, t) = l_{02}(s, t)$, and $b_{12}(s, t) = l_{12}(s, t)$ at the truth (θ_0, Λ_0). Existence and uniqueness of $g_{hj,0}$ follows from Fredholm's first theorem [e.g. **?** , p. 48]. Let $g_{hj,0n}$ be the piecewise linear interpolant having the same knots as $\hat{\Lambda}_n$. Consider the approximately least favorable submodel $\Lambda_y = (\Lambda_{hj,y})_{h \neq j}$ with

$$
\Lambda_{hj,y}(\theta, \Lambda) = \int \{1 + (\theta - y) g_{hj,0n}\}\,\mathrm{d}\Lambda_{hj}
$$

defined so that $\partial \Lambda_{hj,y}/\partial y_{|y=0} = \int g_{hj}\,\mathrm{d}\Lambda_{hj}$, $\Lambda_{hj,\theta} = \Lambda_{hj}$ and $\Lambda_{hj,y} \in H_{hj,n}$ for y sufficiently close to θ. The "no-bias" condition and remaining structural requirements of Λ_y imposed in **?** , Theorem 1] largely follow by assumption, the proof Theorem B.3 and our restrictions on the size of κ.

C.2 Interval-censored progression times

Let $Y_{K,1} < \ldots < Y_{K,K}$ denote a random number K of inspection times on (σ, τ) with $Y_{K,0} \equiv \sigma$ and $Y_{K,K+1} \equiv \tau$. For $j = 0, \ldots, K$, put

$$\varDelta_{K,j} = \varDelta_1 1(Y_{K,j} < S \leq Y_{K,j+1}) \vee (1 - \varDelta_1) 1(Y_{K,j} < V \leq Y_{K,j+1}).$$

Then the log-likelihood of a single realization x is

$$\begin{aligned}\ell_{\theta,\varLambda}(x) = \sum_{j=1}^{k} & \delta_{k,j}(-A_{01}(y_{k,j} \mid z) - A_{01}(y_{k,j} \mid z)) \\ & + \delta_1 \delta_{k,j}(\log P_{01}(y_{k,j}, y_{k,j+1} \mid z) - A_{12}(y_{k,j+1}, v \mid z) + \delta_2 \log \alpha_{12}(v \mid z)) \\ & + (1 - \delta_0 \delta_1)\delta_{k,j} + \log\big(P_{01}(y_{k,j}, v \mid z)\alpha_{12}(v \mid z)^{\delta_2} + P_{00}(y_{k,j}, v \mid z)\alpha_{02}(v \mid z)^{\delta_2}\big) \\ & + \delta_0 \delta_{k,j}(\delta_2 \log \alpha_{02}(v \mid z) - A_{01}(v \mid z) - A_{02}(v \mid z)).\end{aligned}$$

Here the score for $\varLambda$ is

$$\begin{aligned} & B_{\theta,\varLambda} g(x) \\ = & -\delta_{k,j} \sum_{j=1}^{k} \sum_{h \neq 0} \int_0^{y_{k,j}} g_{0h}(y)\, \mathrm{d}A_{0h}(y \mid z) \\ & + \sum_{j=1}^{k} \delta_1 \delta_{k,j} \Bigg(\delta_2 g_{12}(v) - \int_{y_{k,j}}^{v} g_{12}(y)\, \mathrm{d}A_{12}(y \mid z) \\ & \qquad + \sum_{h \neq l} \int_{y_{k,j}}^{y_{k,j+1}} g_{hl}(y)\, \mathrm{d}A_{hl}(y \mid z) \frac{\varphi_{hl}(y, y_{k,j}, y_{k,j+1}, z)}{P_{01}(y_{k,j}, y_{k,j+1} \mid z)} \Bigg) \\ & + \sum_{j=1}^{k} (1 - \delta_0 \delta_1)\delta_{k,j} \frac{P_{00}(y_{k,j}, v \mid z)\alpha_{02}(v \mid z)^{\delta_2}\big(\delta_2 g_{02}(v) - \sum_{h \neq 0} \int_{y_{k,j}}^{v} g_{0h}(y)\, \mathrm{d}A_{0h}(y \mid z)\big)}{\sum_{h=0,1} P_{0h}(y_{k,j}, v \mid z)\alpha_{h2}(v \mid z)^{\delta_2}} \\ & \qquad + P_{01}(y_{k,j}, v \mid z)\alpha_{12}(v \mid z)^{\delta_2} \Bigg(\delta_2 g_{12}(v) + \sum_{h \neq l} \int_{y_{k,j}}^{v} g_{hl}(y)\, \mathrm{d}A_{hl}(y \mid z)\varphi_{hl}(y, s, t, z) \Bigg) \\ & + \sum_{j=1}^{k} (1 - \delta_0 \delta_1)\delta_{k,j} \frac{P_{00}(y_{k,j}, v \mid z)\alpha_{02}(v \mid z)^{\delta_2}\big(\delta_2 g_{02}(v) - \sum_{h \neq 0} \int_{y_{k,j}}^{v} g_{0h}(y)\, \mathrm{d}A_{0h}(y \mid z)\big)}{\sum_{h=0,1} P_{0h}(y_{k,j}, v \mid z)\alpha_{h2}(v \mid z)^{\delta_2}} \\ & + \delta_0 \Bigg(\delta_2 g_{02}(v) - \sum_{j=1,2} \int_{y_{k,j}}^{v} g_{0j}(y)\, \mathrm{d}A_{0j}(y \mid z) \Bigg). \end{aligned}$$

Put $P_k(d, z) = \Pr(K = k \mid D = d, Z = z)$. Under Condition A.4 and Condition A.2, the conditional expectation of the score for $\varLambda$ given $\{S = s, T = t, Z = z\}$ can be written in the form

$$\begin{aligned} & \mathrm{E}(B_{\theta,\varLambda} g(X) \mid S = s, T = t, Z = z) \\ = & -\int_0^s g_{01}(y) \int_{d=\sigma}^{\tau} \sum_{k=1}^{\infty} P_k(d, z) \int_{u=y}^{s \wedge d} \int_{s \wedge d}^{\tau} 1(w - u > y_0)\, \mathrm{d}F(u, w \mid k, d, z)\, \mathrm{d}F_{D|Z}(d \mid z)\, \mathrm{d}A_{01}(y \mid z) \\ & + 1(s < t) \int_{\sigma}^{\tau} g_{01}(y)\, \mathrm{d}A_{01}(y \mid z) \int_{d=\sigma}^{\tau} \sum_{k=1}^{\infty} P_k(d, z) \\ & \qquad \times \int_{u=\sigma}^{s} \int_{s}^{d} \frac{1(u < y < w)\varphi_{01}(y, u, w, z)}{P_{01}(u, w \mid z)} 1(w - u > y_0)\, \mathrm{d}F(u, w \mid k, d, z)\, \mathrm{d}F_{D|Z}(d \mid z) \end{aligned}$$

$$
\begin{aligned}
&- \int_{\sigma}^{\tau} g_{01}(y)\,\mathrm{d}A_{01}(y \mid z) \int_{d=\sigma}^{\tau} \sum_{k=1}^{\infty} P_k(d,z) \\
&\qquad \times \int_{u=\sigma}^{s\wedge d} \int_{s\wedge d}^{\tau} \frac{1(u<y<v) P_{00}(u,v \mid z)\alpha_{02}(v \mid z)^{\delta_2}}{\sum_{j=0,1} P_{0j}(u,v \mid z)\alpha_{j2}(v \mid z)^{\delta_2}}\,\mathrm{d}F(u,w \mid k,d,z)\,\mathrm{d}F_{D|Z}(d \mid z) \\
&+ \int_{\sigma}^{\tau} g_{01}(y)\,\mathrm{d}A_{01}(y \mid z) \int_{d=\sigma}^{\tau} \sum_{k=1}^{\infty} P_k(d,z) \\
&\qquad \times \int_{u=\sigma}^{s\wedge d} \int_{s\wedge d}^{\tau} \frac{1(u<y<v) P_{01}(u,v \mid z)\alpha_{12}(v \mid z)^{\delta_2}}{\sum_{j=0,1} P_{0j}(u,v \mid z)\alpha_{j2}(v \mid z)^{\delta_2}} \varphi_{01}(y,u,t,z)\,\mathrm{d}F(u,w \mid k,d,z)\,\mathrm{d}F_{D|Z}(d \mid z) \\
&+ \sum_{h=0,1} g_{h2}(t) Q_{h2}(s,t,z) + \sum_{h=0,1} \int_{\sigma}^{\tau} g_{h2}(y) R_{h2}(y,s,t,z)\,\mathrm{d}\Lambda_{h2}(y) \\
\equiv &\int_0^s g_{01}(y) Q_{01}(y,s,t,z)\,\mathrm{d}\Lambda_{01}(y) + \sum_{h=0,1} g_{h2}(t) C_{h2}(s,t,z) + \sum_{h\neq j} \int_{\sigma}^{\tau} g_{hj}(y) R_{hj}(y,s,t,z)\,\mathrm{d}\Lambda_{hj}(y),
\end{aligned}
$$

where F is the conditional distribution function of $(Y_{k,j-1}, Y_{k,j})$ given (D,Z) unless otherwise indicated. Again the least favorable directions can be obtained as the solution to Fredholm integral equations. The remainder of the proof now proceeds according to the dual right censoring case.

D Self-consistency equations

Under the exponential sieve, ϕ represents the piecewise constant baseline transition intensities λ_n with $\Lambda_n = \int \lambda_n$. Define a common partition $\mathcal{S} = (s_0, \ldots, s_J)$ from $\mathcal{T} = (\mathcal{T}_{hj,n})_{h\neq j}$ so that λ_n is constant on each $(s_{j-1}, s_j]$, $j = 1, \ldots, J$. Consider $t_{k-1}, t_k \in \mathcal{T}_{hj,n}$, $s_{j-1}, s_j \in \mathcal{S}$, and any $s < t$. Let $L_{hj,k}(s,t)$ be the length of $(t_{k-1}, t_k] \cap (s,t]$, $I_{hj,k}(s,t) = 1(L_{hj,k}(s,t) > 0)$, $L_j(s,t)$ the length of $(s_{j-1}, s_j] \cap (s,t]$, and $I_j(s,t) = 1(L_j(s,t) > 0)$. For $s_{j-1}, s_j \in \mathcal{S}$, put

$$
\begin{aligned}
q_j(X;\theta,\phi) &= \lambda_{01,n}(s_j) e^{Z_{01}'\theta} + \lambda_{02,n}(s_j) e^{Z_{02}'\theta} - \lambda_{12,n}(s_j) e^{Z_{12}'\theta}, \\
r_{hk,j}(X;\theta,\phi) &= \exp\Big(-L_j(L,R)\lambda_{hk,n}(s_j) e^{Z_{hk}'\theta}\Big), \quad h \neq k, \\
p_{1,j}(X;\theta,\phi) &= (1-\Delta_0) I_j(L,R) P_{00}(0, s_{j-1}\vee L \mid Z) P_{01}(s_{j-1}\vee L, s_j \wedge R \mid Z) \\
&\qquad \times P_{11}(s_j \wedge R, V \mid Z)\Big(\lambda_{12,n}(V) e^{Z_{12}'\theta}\Big)^{\Delta_2}, \\
r_{0,j}(X;\theta,\phi) &= (1-\Delta_0) I_j(L,R) P_{00}(0, s_{j-1}\vee L \mid Z)\lambda_{01,n}(s_j) e^{Z_{01}'\theta} \frac{r_{01,j}(X;\theta,\phi) r_{02,j}(X;\theta,\phi)}{q_j(Z;\theta,\phi)} \\
&\qquad \times P_{11}(s_j \wedge R, V \mid Z)\Big(\lambda_{12,n}(V) e^{Z_{12}'\theta}\Big)^{\Delta_2}, \\
r_{1,j}(X;\theta,\phi) &= (1-\Delta_0) I_j(L,R) P_{00}(0, s_{j-1}\vee L \mid Z)\lambda_{01,n}(s_j) e^{Z_{01}'\theta} \frac{r_{12,j}(X;\theta,\phi)}{q_j(X;\theta,\phi)} \\
&\qquad \times P_{11}(s_j \wedge R, V \mid Z)\Big(\lambda_{12,n}(V) e^{Z_{12}'\theta}\Big)^{\Delta_2}.
\end{aligned}
$$

Then it is straightforward to show that

$$
\begin{aligned}
P_{01}(s_{j-1}\vee L, s_j\wedge R \mid Z) &= \int_{s_{j-1}\vee L}^{s_j\wedge R} P_{00}(s_{j-1}\vee L, u \mid Z)\lambda_{01}(u) e^{Z_{01}'\theta} P_{11}(u, s_j \wedge R \mid Z)\,\mathrm{d}u \\
&= \lambda_{01,n}(s_j) e^{Z_{01}'\theta}\left(\frac{r_{01,j}(X;\theta,\phi) r_{02,j}(X;\theta,\phi) - r_{12,j}(X;\theta,\phi)}{q_j(X;\theta,\phi)}\right)
\end{aligned}
$$

and $p_1(X;\theta,\phi) = \sum_{j=1}^{J} I_j(L,R)p_{1,j}(X;\theta,\phi)$ is the likelihood contribution under (θ,ϕ) through the states $0 \to 1 \to 2$. Similarly let $p_2(X;\theta,\phi) = (1-\Delta_1)P_{00}(0,V \mid Z)\alpha_{02}(V \mid Z)^{\Delta_2}$ be the contribution through $0 \to 2$, under (θ,ϕ).

In Step 4.2 of the self-consistency algorithm from Section 4, the ratio in the candidate increment $\eta^{(j)}_{\phi_{hj,k}}$ corresponds to a candidate value for $\lambda^{(j+1)}_{hj,n}(t)$, $t \in (t_{k-1},t_k]$ with $t_{k-1},t_k \in \mathcal{T}_{hj,n}$, obtained by re-arranging the score equation $\nabla_{\phi_{hj,k}}\ell_n(\theta,\phi) = 0$ to give an expression for $\phi_{hj,k}$ that we evaluate for $(\theta,\phi) = (\theta^{(j)},\phi^{(j)})$. Using the convention $0/0 \equiv 0$, this gives the candidate values

$$\left(\mathbb{P}_n \frac{1}{\mathrm{lik}_{\theta^{(j)},\phi^{(j)}}(X)} \sum_{j=1}^{J} I_{01,k}(s_{j-1},s_j)p_{1,j}(X;\theta^{(j)},\phi^{(j)})\right) \times \left[\mathbb{P}_n \frac{e^{Z_{01}'\theta}}{\mathrm{lik}_{\theta^{(j)},\phi^{(j)}}(X)}\left\{\sum_{j=1}^{J} p_{1,j}(X;\theta^{(j)},\phi^{(j)})\left(L_{01,k}(0,s_{j-1}) + \frac{I_{01,k}(s_{j-1},s_j)}{q_j(Z;\theta^{(j)},\phi^{(j)})}\right) - I_{01,k}(s_{j-1},s_j)r_{0,j}(X;\theta^{(j)},\phi^{(j)}) + L_{01,k}(0,V)p_2(X;\theta^{(j)},\phi^{(j)})\right\}\right]^{-1},$$

for $\lambda^{(j+1)}_{01,n}(t_k)$ with $t_k \in \mathcal{T}_{01,n}$,

$$\left(\mathbb{P}_n \frac{I_{02,k}(0,V)p_2(X;\theta^{(j)},\phi^{(j)})}{\mathrm{lik}_{\theta^{(j)},\phi^{(j)}}(X)}\right) \times \left[\mathbb{P}_n \frac{e^{Z_{02}'\theta}}{\mathrm{lik}_{\theta^{(j)},\phi^{(j)}}(X)}\left\{\sum_{j=1}^{J} p_{1,j}(X;\theta^{(j)},\phi^{(j)})\left(L_{02,k}(0,s_{j-1}) + \frac{I_{01,k}(s_{j-1},s_j)}{q_j(Z;\theta^{(j)},\phi^{(j)})}\right) - I_{02,k}(s_{j-1},s_j)r_{0,j}(X;\theta^{(j)},\phi^{(j)}) + L_{02,k}(0,V)p_2(X;\theta^{(j)},\phi^{(j)})\right\}\right]^{-1},$$

for $\lambda^{(j+1)}_{02,n}(t_k)$ with $t_k \in \mathcal{T}_{02,n}$, and

$$\left(\mathbb{P}_n \frac{I_{12,k}(0,V)p_1(X;\theta^{(j)},\phi^{(j)})}{\mathrm{lik}_{\theta^{(j)},\phi^{(j)}}(X)}\right) \times \left[\mathbb{P}_n \frac{e^{Z_{12}'\theta}}{\mathrm{lik}_{\theta^{(j)},\phi^{(j)}}(X)}\left\{\sum_{j=1}^{J} p_{1,j}(X;\theta^{(j)},\phi^{(j)})\left(L_{12,k}(s_j,V) - \frac{I_{12,k}(s_{j-1},s_j)}{q_j(Z;\theta^{(j)},\phi^{(j)})}\right) + I_{12,k}(s_{j-1},s_j)r_{1,j}(X;\theta^{(j)},\phi^{(j)})\right\}\right]^{-1}$$

for $\lambda^{(j+1)}_{12,n}(t_k)$ with $t_k \in \mathcal{T}_{12,n}$.

E SIMULATION RESULTS: DUAL RIGHT CENSORING

Reiterating the four simulation scenarios here for convenience, we considered:

(1) Independent censoring: $C = \xi \wedge D$, where ξ is exponentially-distributed with mean $\tau/2 = 1$, $\beta_0 = \log(1/2)$ and $\beta_1 = 0$.

(2) Increased independent censoring: $C = \xi \wedge D$, $\xi \sim \exp(1)$, $\beta_0 = \log(9)$ and $\beta_1 = 0$.

(3) Conditionally independent censoring: $C = \xi \wedge D$, $\xi \sim \exp(1)$, $\beta_0 = \log(1/3)$ and $\beta_1 = \log(3/2)$.

(4) Dependent censoring: $C = (T_{02} \wedge T_{12} - \xi) \wedge D$, where ξ is exponential with mean 1 and truncated to $(0, T_{02} \wedge T_{12})$, $\beta_0 = \log(1/2)$ and $\beta_1 = \log(3/2)$.

Estimators compared to the piecewise exponential SMLE considered (singly) right-censored data arising four ways:

- SPMLE: Cox-type transition intensity model under right censoring at C, so any information on survival status observed after C is discarded.
- PFS1: Cox model for PFS right-censored at C.
- PFS2: Cox model for PFS with last (negative) progression status carried forward to $V = T_{02} \wedge T_{12} \wedge D$.
- Latent SPMLE: Cox-type transition intensity model based on the underlying or "latent" data right-censored at D.

Each of these were fit using **?**] coxph function from the survival package for R [**?**]. The SMLE was fit with an implementation of the self-consistency algorithm described in Section 4.1 provided by the coxinterval package for R, which is available from CRAN at the following address.

http://cran.r-project.org/web/packages/coxinterval/index.html

			SMLE			SPMLE			PFS1	PFS2	Latent SPMLE		
Scenario	n		θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ	θ	θ_1	θ_2	θ_3
1	250	Mean	-0.700	-0.696	0.001	-0.697	-0.698	-0.002	-0.697	-0.663	-0.698	-0.698	-0.002
		SD	0.196	0.219	0.206	0.198	0.230	0.208	0.149	0.140	0.186	0.212	0.195
		ASE	0.192	0.218	0.199	0.195	0.228	0.200	0.148	0.139	0.183	0.212	0.189
		CP	0.947	0.953	**0.943**	0.946	0.954	**0.941**	0.952	**0.941**	0.947	0.953	**0.942**
	500	Mean	-0.696	-0.695	0.003	-0.695	-0.697	0.001	-0.695	-0.661	-0.694	-0.696	0.001
		SD	0.137	0.154	0.142	0.138	0.162	0.143	0.103	0.097	0.129	0.150	0.134
		ASE	0.135	0.154	0.140	0.138	0.161	0.141	0.105	0.098	0.129	0.149	0.134
		CP	0.949	0.950	0.948	0.952	0.950	0.947	**0.956**	**0.940**	0.951	0.950	0.948
	1000	Mean	-0.694	-0.694	0.001	-0.693	-0.694	-0.000	-0.693	-0.660	-0.694	-0.694	-0.001
		SD	0.097	0.109	0.099	0.098	0.115	0.100	0.074	0.069	0.093	0.106	0.095
		ASE	0.096	0.108	0.099	0.097	0.114	0.100	0.074	0.069	0.091	0.106	0.094
		CP	0.948	0.953	0.949	0.950	0.949	0.949	0.950	**0.921**	0.947	0.952	0.952
2	250	Mean	-0.703	-0.698	0.004	-0.699	-0.701	0.001	-0.698	-0.582	-0.698	-0.698	-0.002
		SD	0.218	0.235	0.229	0.224	0.273	0.235	0.171	0.140	0.186	0.212	0.195
		ASE	0.213	0.232	0.220	0.221	0.267	0.225	0.171	0.140	0.183	0.212	0.189
		CP	0.948	0.950	**0.942**	0.952	0.948	**0.937**	0.951	**0.872**	0.947	0.953	**0.942**
	500	Mean	-0.698	-0.697	0.005	-0.696	-0.700	0.003	-0.697	-0.581	-0.694	-0.696	0.001
		SD	0.152	0.165	0.157	0.157	0.191	0.160	0.119	0.098	0.129	0.150	0.134
		ASE	0.150	0.163	0.155	0.156	0.188	0.159	0.120	0.099	0.129	0.149	0.134
		CP	0.950	0.948	0.949	0.954	0.949	0.947	0.954	**0.789**	0.951	0.950	0.948
	1000	Mean	-0.696	-0.694	0.001	-0.694	-0.694	0.001	-0.694	-0.579	-0.694	-0.694	-0.001
		SD	0.108	0.115	0.109	0.111	0.134	0.112	0.085	0.070	0.093	0.106	0.095
		ASE	0.106	0.115	0.109	0.110	0.133	0.112	0.085	0.070	0.091	0.106	0.094
		CP	0.946	0.951	0.951	0.948	0.947	0.949	0.950	**0.619**	0.947	0.952	0.952
3	250	Mean	-0.700	-0.696	0.001	-0.697	-0.698	-0.002	-0.696	-0.661	-0.698	-0.698	-0.002
		SD	0.195	0.218	0.205	0.197	0.227	0.207	0.148	0.140	0.186	0.212	0.195
		ASE	0.191	0.217	0.198	0.193	0.226	0.199	0.147	0.139	0.183	0.212	0.189
		CP	0.947	0.951	**0.944**	0.948	0.952	**0.942**	0.950	**0.940**	0.947	0.953	**0.942**
	500	Mean	-0.696	-0.695	0.003	-0.695	-0.697	0.001	-0.695	-0.659	-0.694	-0.696	0.001
		SD	0.136	0.154	0.141	0.137	0.160	0.142	0.102	0.097	0.129	0.150	0.134
		ASE	0.134	0.153	0.140	0.136	0.159	0.140	0.104	0.098	0.129	0.149	0.134
		CP	0.949	0.948	0.949	0.952	0.950	0.948	0.955	**0.938**	0.951	0.950	0.948
	1000	Mean	-0.694	-0.694	0.001	-0.694	-0.694	-0.000	-0.694	-0.658	-0.694	-0.694	-0.001
		SD	0.096	0.108	0.099	0.097	0.114	0.100	0.074	0.070	0.093	0.106	0.095
		ASE	0.095	0.108	0.098	0.096	0.112	0.099	0.073	0.069	0.091	0.106	0.094
		CP	0.947	0.953	0.950	0.949	0.949	0.949	0.950	**0.918**	0.947	0.952	0.952
4	250	Mean	-0.704	-0.695	0.004	-0.733	-0.736	-0.014	-0.733	-0.693	-0.698	-0.698	-0.002
		SD	0.199	0.225	0.204	0.201	0.263	0.207	0.158	0.139	0.186	0.212	0.195
		ASE	0.194	0.223	0.197	0.197	0.261	0.200	0.158	0.139	0.183	0.212	0.189
		CP	0.946	0.953	**0.942**	**0.945**	0.949	**0.941**	**0.944**	0.951	0.947	0.953	**0.942**
	500	Mean	-0.698	-0.695	0.006	-0.728	-0.736	-0.011	-0.730	-0.690	-0.694	-0.696	0.001
		SD	0.138	0.157	0.140	0.140	0.184	0.142	0.110	0.097	0.129	0.150	0.134
		ASE	0.137	0.157	0.139	0.139	0.184	0.141	0.111	0.098	0.129	0.149	0.134
		CP	0.948	0.950	0.949	**0.944**	**0.944**	0.947	**0.942**	0.954	0.951	0.950	0.948
	1000	Mean	-0.696	-0.694	0.004	-0.726	-0.734	-0.012	-0.729	-0.689	-0.694	-0.694	-0.001
		SD	0.098	0.111	0.098	0.099	0.131	0.100	0.079	0.069	0.093	0.106	0.095
		ASE	0.097	0.110	0.098	0.098	0.130	0.100	0.078	0.069	0.091	0.106	0.094
		CP	0.948	0.950	0.950	**0.936**	**0.936**	0.950	**0.924**	0.948	0.947	0.952	0.952

Table E.1: Mean, standard deviation (SD), average standard error (ASE) and coverage probabilities of 95% confidence intervals (CP) of $\hat{\theta}_n$ with $\theta_0 = (-\log(2), -\log(2), 0) \approx (-0.693, -0.693, 0)$, $\kappa = 2/5$ and dual right censoring. CPs in boldface are significantly different from 0.95 at the 5% level.

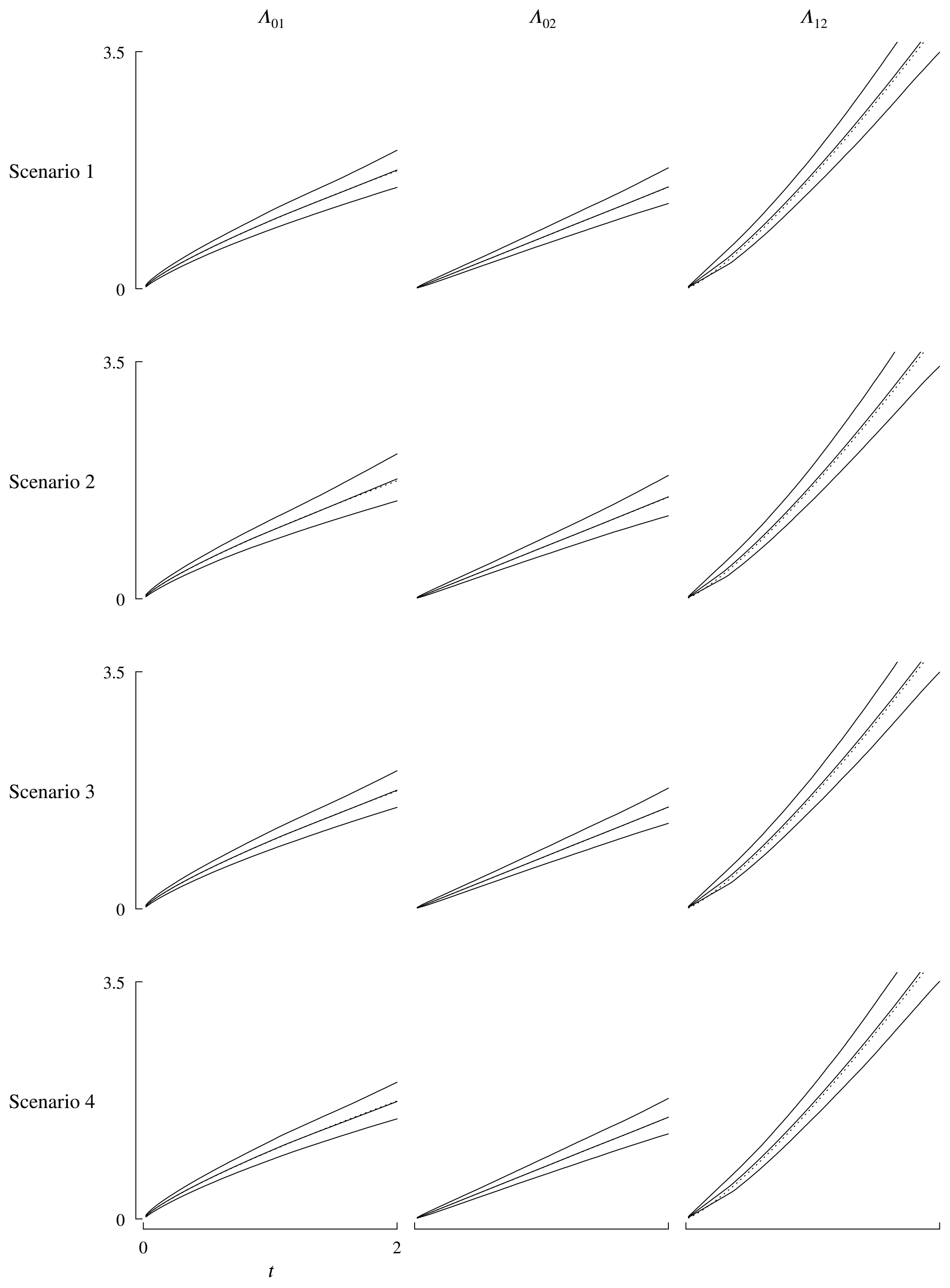

Figure E.1: True values for Λ (dotted) depicted with the pointwise average and 2.5th percentiles of the SMLE (solid) under dual right censoring, $n = 1000$ and Scenarios 1 to 4.

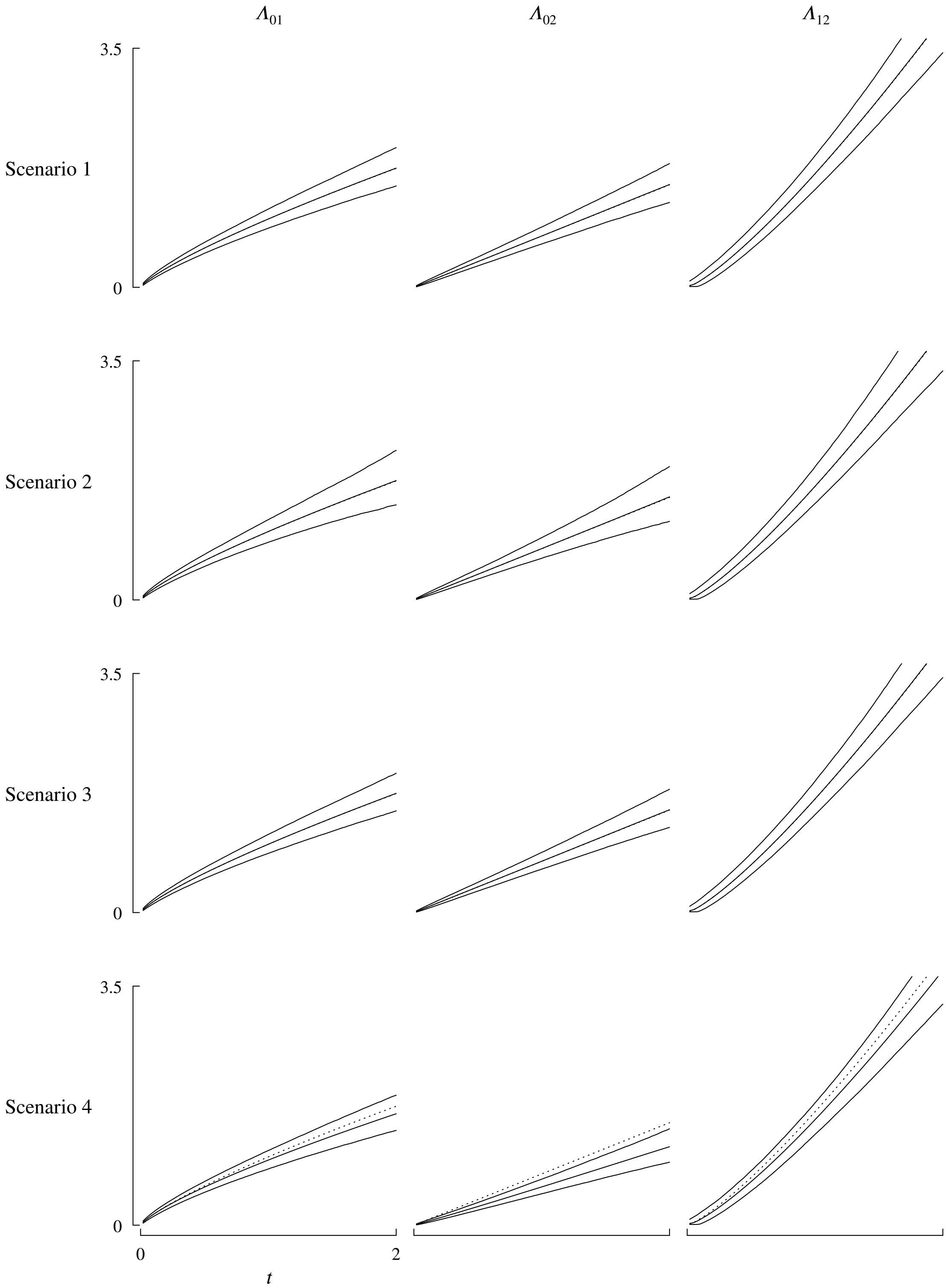

Figure E.2: True values for Λ (dotted) depicted with the pointwise average, 2.5th and 97.5th percentiles of the SPMLE right-censored at C (solid) under dual right censoring, $n = 1000$ and Scenarios 1 to 4.

		$\kappa = 4/15$			$\kappa = 1/3$		
n		θ_1	θ_2	θ_3	θ_1	θ_2	θ_3
250	Relative bias	2.65	1.29	4.70	1.47	1.08	1.92
	Relative precision	1.00	1.00	0.98	1.00	1.00	1.00
500	Relative bias	3.94	1.29	1.97	2.35	1.18	1.36
	Relative precision	1.00	1.00	0.99	1.00	1.00	1.00
1000	Relative bias	5.19	1.73	4.80	2.47	1.25	2.23
	Relative precision	1.00	1.00	0.99	1.00	1.00	1.00

Table E.2: Ratio of empirical bias (relative bias) and standard deviation (relative precision) for $\hat{\theta}_n$ between the specified κ and $\kappa = 2/5$ under dual right censoring Scenario 1.

Scenario	κ	n	CPU time	Iterations (j)	$\log_{10} \Vert\phi\nabla_{\phi}^{(j)}\ell_n(\phi^{(j)})\Vert_\infty$	Alternate $\phi^{(0)}$	Alternate $(\operatorname{typ}\theta, \sup\theta)$
1	4/15	250	0.09	31	-8.40	0	0
1	4/15	500	0.22	30	-8.35	0	0
1	4/15	1000	0.50	30	-7.25	0	0
1	1/3	250	0.12	32	-7.39	0	0
1	1/3	500	0.28	31	-5.12	0	0
1	1/3	1000	0.65	30	-6.53	0	0
1	2/5	250	0.17	32	-6.72	0	0
1	2/5	500	0.41	31	-7.49	0	0
1	2/5	1000	1.17	30	-6.34	0	0
2	2/5	250	0.46	48	-6.75	0	0
2	2/5	500	1.22	48	-5.63	0	0
2	2/5	1000	3.50	49	-7.16	0	0
3	2/5	250	0.16	31	-7.02	0	0
3	2/5	500	0.38	30	-7.78	0	0
3	2/5	1000	1.04	29	-6.55	0	0
4	2/5	250	0.16	34	-8.34	0	0
4	2/5	500	0.38	33	-6.60	0	0
4	2/5	1000	1.03	33	-6.67	0	0

Table E.3: Average processing time in seconds for parameter and variance estimation (CPU time), average number of iterations to convergence, average maximum norm of the inner product between the estimate and score function at the estimate in $\log_{10}$ scale, number of replicates associated with alternate starting value $\phi^{(0)}$, number of replicates with alternate variance tuning parameters $(\operatorname{typ}\theta, \sup\theta)$—each specific to the scenario, sieve size (κ) and sample size under dual right censoring.

F SIMULATION RESULTS: INTERVAL-CENSORED PROGRESSION TIMES

Reiterating the censoring scenarios considered:

(1) Independent censoring: $k = 8$, $\beta_0 = \log(1/4)$ and $\beta_1 = 0$;

(2) Increased independent censoring: $k = 4$, $\beta_0 = \log(1/4)$ and $\beta_1 = 0$;

(3) Conditionally independent censoring: $k = 8$, $\beta_0 = \log(1/4)$ and $\beta_1 = \log(4/9)$; and

(4) Dependent censoring: $k = 8$, ξ follows the Weibull distribution with shape $3/4$, scale 1 and truncated to $(0, T_{02} \wedge T_{12})$.

Performance of the SMLE was compared with estimators fit to (singly) right-censored data arising four different ways:

- Imputation-based SPMLE: Cox-type transition intensity model based on midpoint-imputed progression times, if status is known, and last negative status carried forward, if status is unknown.
- PFS1: Cox model for PFS imputed to the first detection time (i.e., right endpoint of the censoring interval), if status is known, and last negative status carried forward, if status is unknown.
- PFS2: Similar to PFS1, but right-censor at the last assessment that precedes two or more missed scheduled inspections times.
- Latent SPMLE: Cox-type transition intensity model based on the underlying or "latent" data right-censored at D.

The above two variants of PFS are based on guidelines [**? ?**]. As with the dual right censoring alternatives, estimates were obtained from the coxph and coxinterval packages.

From Table F.3 note that, of the 180,000 samples generated, 169 (0.09%) converged under the starting value $\theta^{(0)} = (1/2, 1/2, 1/2)$ rather than $\theta^{(0)} = 0$, and 23 (0.01%) needed smaller values for the tuning parameters in variance estimation. The majority in both of these exceptions were encountered for samples generated with $n = 250$ under the dependent censoring in Scenario 4.

			SMLE			Imputed SPMLE			PFS1	PFS2	Latent SPMLE		
Scenario	n		θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ	θ	θ_1	θ_2	θ_3
1	250	Mean	-0.706	-0.690	0.005	-0.685	-0.698	-0.042	-0.688	-0.690	-0.696	-0.695	0.001
		SD	0.210	0.248	0.211	0.206	0.192	0.201	0.141	0.142	0.186	0.214	0.194
		ASE	0.205	0.247	0.207	0.204	0.190	0.196	0.139	0.140	0.183	0.212	0.189
		CP	**0.944**	0.954	0.950	0.949	0.951	**0.937**	0.948	0.947	0.949	0.950	**0.943**
	500	Mean	-0.699	-0.690	0.004	-0.682	-0.696	-0.041	-0.687	-0.689	-0.693	-0.693	0.002
		SD	0.148	0.172	0.146	0.146	0.135	0.139	0.100	0.101	0.131	0.150	0.134
		ASE	0.145	0.171	0.145	0.144	0.134	0.138	0.098	0.099	0.129	0.149	0.133
		CP	0.946	0.952	0.949	0.947	0.949	**0.942**	0.946	0.946	0.948	0.951	**0.945**
	1000	Mean	-0.697	-0.691	0.000	-0.682	-0.697	-0.045	-0.687	-0.689	-0.693	-0.695	-0.001
		SD	0.103	0.121	0.103	0.102	0.095	0.098	0.069	0.070	0.092	0.106	0.095
		ASE	0.102	0.120	0.103	0.102	0.095	0.098	0.069	0.070	0.091	0.106	0.094
		CP	0.950	0.948	0.949	0.948	0.951	**0.928**	0.950	0.950	0.952	0.948	0.949
2	250	Mean	-0.707	-0.688	0.006	-0.661	-0.700	-0.060	-0.678	-0.680	-0.696	-0.695	0.001
		SD	0.232	0.274	0.227	0.225	0.182	0.219	0.142	0.142	0.186	0.214	0.194
		ASE	0.224	0.271	0.219	0.222	0.179	0.211	0.140	0.140	0.183	0.212	0.189
		CP	0.946	0.953	**0.943**	0.950	0.948	**0.932**	0.946	0.946	0.949	0.950	**0.943**
	500	Mean	-0.701	-0.687	0.004	-0.660	-0.697	-0.061	-0.677	-0.679	-0.693	-0.693	0.002
		SD	0.162	0.189	0.157	0.158	0.128	0.150	0.099	0.100	0.131	0.150	0.134
		ASE	0.158	0.186	0.153	0.157	0.126	0.149	0.099	0.099	0.129	0.149	0.133
		CP	0.946	0.947	0.946	**0.945**	0.946	**0.934**	**0.944**	**0.945**	0.948	0.951	**0.945**
	1000	Mean	-0.699	-0.689	0.000	-0.658	-0.699	-0.063	-0.678	-0.679	-0.693	-0.695	-0.001
		SD	0.114	0.131	0.110	0.111	0.090	0.105	0.070	0.070	0.092	0.106	0.095
		ASE	0.111	0.129	0.108	0.111	0.089	0.105	0.070	0.070	0.091	0.106	0.094
		CP	**0.944**	0.948	0.946	**0.937**	0.951	**0.910**	**0.944**	0.947	0.952	0.948	0.949
3	250	Mean	-0.706	-0.689	0.005	-0.643	-0.720	-0.022	-0.674	-0.654	-0.696	-0.695	0.001
		SD	0.212	0.251	0.212	0.208	0.191	0.202	0.141	0.143	0.186	0.214	0.194
		ASE	0.207	0.250	0.208	0.206	0.189	0.197	0.139	0.142	0.183	0.212	0.189
		CP	**0.944**	**0.955**	0.949	**0.941**	0.948	**0.942**	0.947	**0.937**	0.949	0.950	**0.943**
	500	Mean	-0.699	-0.690	0.004	-0.640	-0.718	-0.022	-0.672	-0.653	-0.693	-0.693	0.002
		SD	0.149	0.173	0.147	0.148	0.134	0.140	0.100	0.102	0.131	0.150	0.134
		ASE	0.146	0.173	0.146	0.145	0.133	0.139	0.098	0.100	0.129	0.149	0.133
		CP	0.947	0.952	0.950	**0.928**	**0.943**	0.948	**0.941**	**0.926**	0.948	0.951	**0.945**
	1000	Mean	-0.697	-0.691	0.000	-0.640	-0.718	-0.026	-0.673	-0.653	-0.693	-0.695	-0.001
		SD	0.104	0.122	0.104	0.103	0.094	0.098	0.069	0.071	0.092	0.106	0.095
		ASE	0.103	0.121	0.103	0.103	0.094	0.098	0.069	0.071	0.091	0.106	0.094
		CP	0.951	0.949	0.949	**0.917**	**0.943**	**0.942**	**0.943**	**0.911**	0.952	0.948	0.949
4	250	Mean	-0.710	-0.692	0.006	-0.648	-0.698	-0.058	-0.720	-0.720	-0.697	-0.698	0.002
		SD	0.244	0.289	0.242	0.239	0.177	0.231	0.167	0.167	0.185	0.215	0.194
		ASE	0.235	0.280	0.222	0.234	0.174	0.224	0.164	0.164	0.183	0.212	0.189
		CP	**0.940**	0.948	**0.941**	**0.940**	0.947	**0.937**	**0.944**	**0.944**	0.950	0.946	**0.944**
	500	Mean	-0.704	-0.695	0.010	-0.638	-0.696	-0.062	-0.726	-0.726	-0.695	-0.698	0.001
		SD	0.172	0.197	0.164	0.170	0.121	0.165	0.120	0.120	0.129	0.151	0.136
		ASE	0.170	0.195	0.158	0.170	0.121	0.161	0.119	0.119	0.129	0.150	0.133
		CP	**0.945**	0.950	**0.942**	**0.937**	0.950	**0.932**	**0.939**	**0.939**	0.951	0.952	**0.945**
	1000	Mean	-0.700	-0.694	0.009	-0.636	-0.695	-0.062	-0.725	-0.725	-0.694	-0.696	0.000
		SD	0.120	0.137	0.113	0.119	0.085	0.114	0.084	0.084	0.091	0.106	0.094
		ASE	0.120	0.135	0.111	0.120	0.085	0.114	0.084	0.084	0.091	0.106	0.094
		CP	0.948	0.950	0.948	**0.922**	0.952	**0.916**	**0.933**	**0.933**	0.953	0.949	0.953

Table F.1: Mean, standard deviation (SD), average standard error (ASE) and coverage probabilities of 95% confidence intervals (CP) of $\hat{\theta}_n$ with $\theta_0 = (-\log(2), -\log(2), 0) \approx (-0.693, -0.693, 0)$, $\kappa = 1/3$ and interval-censored progression times. CPs in boldface are significantly different from 0.95 at the 5% level.

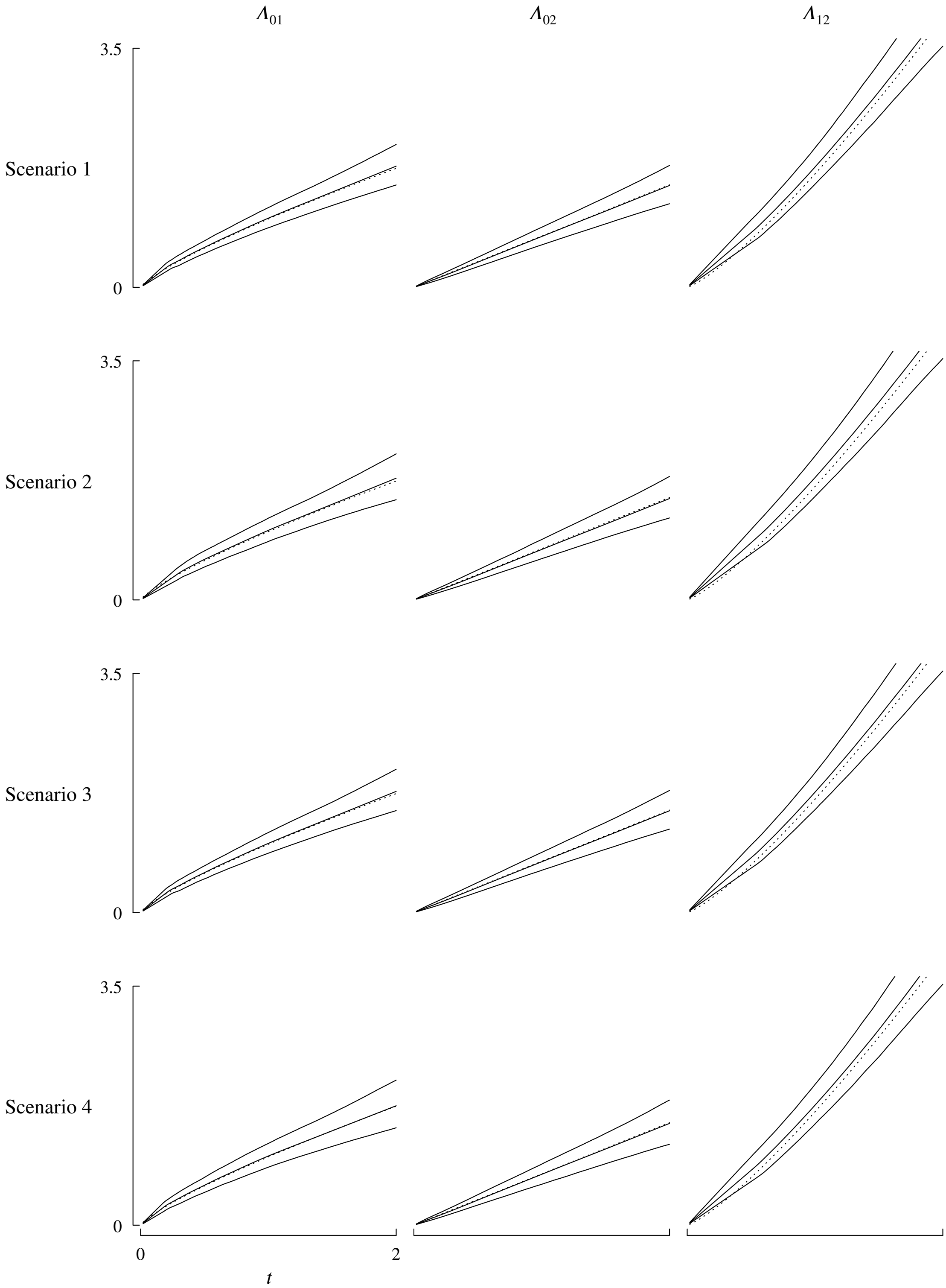

Figure F.1: True values for Λ (dotted) depicted with the pointwise average, 2.5th and 97.5th percentiles of the SMLE (solid) under interval-censored progression times, $n = 1000$ and Scenarios 1 to 4.

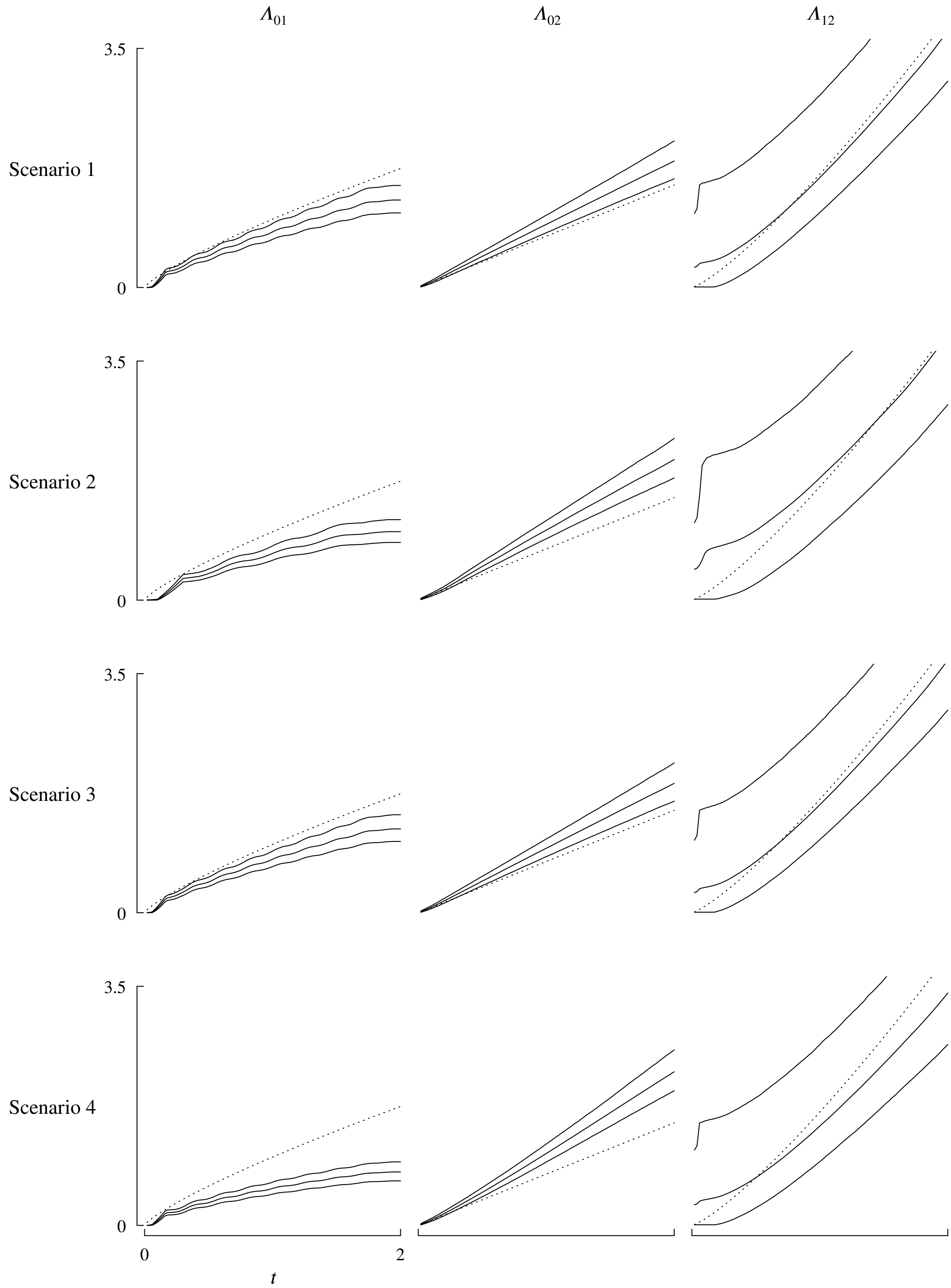

Figure F.2: True values for Λ (dotted) depicted with the pointwise average and 2.5th percentiles of the imputation-based SPMLE (solid) under interval-censored progression times, $n = 1000$ and Scenarios 1 to 4.

n		$\kappa = 4/15$ θ_1	θ_2	θ_3	$\kappa = 2/5$ θ_1	θ_2	θ_3
250	Relative bias	1.45	1.29	1.37	0.80	0.85	0.93
	Relative precision	1.00	1.00	0.99	1.00	1.00	1.01
500	Relative bias	1.58	1.23	1.37	0.85	1.13	0.83
	Relative precision	1.00	1.00	0.99	1.00	1.00	1.01
1000	Relative bias	1.51	1.25	7.95	0.86	0.97	-1.95
	Relative precision	1.00	1.00	0.99	1.00	1.00	1.00

Table F.2: Ratio of empirical bias (relative bias) and standard deviation (relative precision) for $\hat{\theta}_n$ between the specified κ and $\kappa = 1/3$ under interval-censored progression times in Scenario 1.

Scenario	κ	n	CPU time	Iterations (j)	$\log_{10} \lVert \phi \nabla_{\phi}^{(j)} \ell_n(\phi^{(j)}) \rVert_\infty$	Alternate $\phi^{(0)}$	Alternate $(\operatorname{typ} \theta, \sup \theta)$
1	4/15	250	0.18	47	-8.28	0	0
1	4/15	500	0.42	51	-7.40	0	0
1	4/15	1000	1.14	59	-4.10	0	0
1	1/3	250	0.47	116	-4.70	0	0
1	1/3	500	2.72	314	-6.03	0	0
1	1/3	1000	13.15	677	-4.74	0	1
1	2/5	250	3.20	668	-2.52	1	0
1	2/5	500	19.31	1610	-3.35	0	1
1	2/5	1000	103.17	3123	-1.14	1	0
2	1/3	250	2.51	622	-5.15	4	0
2	1/3	500	12.49	1239	-2.79	1	0
2	1/3	1000	60.05	2363	-2.72	0	2
3	1/3	250	0.56	135	-7.10	0	0
3	1/3	500	2.96	332	-4.19	0	0
3	1/3	1000	15.62	747	-4.61	0	0
4	1/3	250	1.35	270	-3.50	148	18
4	1/3	500	7.89	663	-4.54	14	1
4	1/3	1000	43.81	1443	-4.59	0	2

Table F.3: Average processing time in seconds for parameter and variance estimation (CPU time), average number of iterations to convergence, average maximum norm of the inner product between the estimate and score function at the estimate in $\log_{10}$ scale, number of replicates associated with alternate starting value $\phi^{(0)}$, number of replicates with alternate variance tuning parameters $(\operatorname{typ} \theta, \sup \theta)$—each specific to the scenario, sieve size (κ) and sample size under interval-censored progression times.