Statistica Sinica **29** (2019), 2083-2104
doi:https://doi.org/10.5705/ss.202017.0175

NONPARAMETRIC INFERENCE FOR MARKOV PROCESSES WITH MISSING ABSORBING STATE

Giorgos Bakoyannis, Ying Zhang and Constantin T. Yiannoutsos

Indiana University

Abstract: This study examines nonparametric estimations of a transition probability matrix of a nonhomogeneous Markov process with a finite state space and a partially observed absorbing state. We impose a missing-at-random assumption and propose a computationally efficient nonparametric maximum pseudolikelihood estimator (NPMPLE). The estimator depends on a parametric model that is used to estimate the probability of each absorbing state for the missing observations based, potentially, on auxiliary data. For the latter model, we propose a formal goodness-of-fit test based on a residual process. Using modern empirical process theory, we show that the estimator is uniformly consistent and converges weakly to a tight mean-zero Gaussian random field. We also provide a methodology for constructing simultaneous confidence bands. Simulation studies show that the NPMPLE works well with small sample sizes and that it is robust against some degree of misspecification of the parametric model for the missing absorbing states. The method is illustrated using HIV data from sub-Saharan Africa to estimate the transition probabilities of death and disengagement from HIV care.

Key words and phrases: Aalen-Johansen estimator, competing risks, cumulative incidence function, double-sampling, finite state space, missing cause of failure, pseudolikelihood.

1. Introduction

Continuous-time nonhomogeneous Markov processes with a finite state space and absorbing states play an important role in medicine, epidemiology, and public health. Modern medical decision-making is frequently based on estimates of the transition probability matrix of an absorbing continuous-time Markov process, with the goal of evaluating the cost-effectiveness of different medical strategies. Additionally, absorbing Markov processes are crucial in studies of natural history and disease prognoses, evaluating the health needs of populations, and monitoring and evaluating public health programs.

A common problem in studies involving absorbing Markov processes is that ascertaining the absorbing state is incomplete, owing to nonresponse or to the study design. A design with planned missing observations on absorbing states can be used to reduce the total cost of the study when the absorbing-state diagnostic procedures are expensive. Moreover, such a design can be used to provide the information necessary to deal with an absorbing-state misclassification in studies that, by default, use imperfect diagnostics, such those that use electronic health record data (Ladha and Eikermann (2015); Kolek et al. (2016)). In such cases, a gold-standard diagnostic procedure is used in a small sample of cases in an absorbing state, owing to financial or other constraints. For the remaining cases in an absorbing state, a gold-standard diagnosis is missing. The study design with planned missing observations in the absorbing state can be regarded as a special case of a double-sampling design. Double-sampling designs have been used in the past to deal with misclassification in simpler settings (Tenenbein (1970); Rahardja and Young (2011); Rahardja and Yang (2015)).

Recent studies have examined nonparametric estimation with a missing absorbing state in the competing risks model, the simplest Markov process with multiple absorbing states, under a missing-at-random (MAR) assumption. Effraimidis and Dahl (2014) proposed a fully nonparametric estimation approach that does not utilize auxiliary information. This estimator was shown to converge at a rate slower than the usual rate, $\sqrt{n}$. Lee, Dignam and Han (2014) proposed a $\sqrt{n}$-consistent estimator based on parametric multiple-imputation (Wang and Robins (1998); Lu and Tsiatis (2001)). Recently, Gouskova, Lin and Fine (2017) proposed a fully nonparametric estimator that is $\sqrt{n}$-consistent. Note that none of the aforementioned works developed a methodology for constructing simultaneous confidence bands for the transition probabilities, also known as cumulative incidence functions (CIF) in the competing-risks setting. Moreover, Effraimidis and Dahl (2014) and Gouskova, Lin and Fine (2017) established only the pointwise asymptotic normality for their CIF estimators. Finally, these fully nonparametric estimation approaches do not utilize auxiliary information, which may be needed to make the MAR assumption plausible in practice (Lu and Tsiatis (2001)).

In this work, we examine nonparametric inferences for general continuous-time nonhomogeneous Markov processes with a finite state space and a missing absorbing state, with right-censored and/or left-truncated data, under the MAR assumption. We use auxiliary variables in a parametric model for the true absorbing-state probabilities, and derive a closed-form nonparametric maximum pseudolikelihood estimator (NPMPLE) for the transition probability matrix. The basic idea is to replace the missing absorbing state-specific counting processes with the expected state-specific processes, according to the fitted parametric model. A similar approach was developed by Cook and Kosorok (2004) for analyzing the time to the first event of interest in clinical trials where event ascertainment is delayed. Our method can be regarded as an extension of the modified Kaplan–Meier estimator proposed by Cook and Kosorok (2004) in that we provide an estimator of a general Markov process that describes the complete event history of the population under study, where some absorbing states are missing or their ascertainment is delayed. Using modern empirical process theory, we study the asymptotic properties of the NPMPLE for the transition probability matrix, and evaluate its performance in finite samples via simulation studies. We show that the estimator is $\sqrt{n}$-consistent and converges weakly to a tight zero-mean Gaussian random field. We also develop a methodology for the construction of simultaneous confidence bands. The performance of our NPMPLE with small to moderate samples is satisfactory. In particular, the NPMPLE seems to be robust against some degree of misspecification of the parametric model for the true absorbing-state probabilities. We also propose a formal goodness-of-fit approach for evaluating the parametric assumption of this model. As an illustration, the NPMPLE is used to estimate the transition probabilities of disengagement from HIV care and death while in care, using data from the East Africa Regional International Epidemiologic Databases to Evaluate AIDS (IeDEA) Consortium, where death status is incompletely ascertained owing to a double-sampling design.

The rest of this article is organized as follows. Section 2 provides an overview of nonhomogeneous Markov processes, presents our nonparametric estimation approach, and describes a formal goodness-of-fit procedure for the model of the absorbing-state probabilities. Section 3 states the asymptotic theory for the NPMPLE and the goodness-of-fit procedure. Sections 4 and 5 present our simulation studies and our data analysis for the motivating HIV study, respectively. Finally, Section 6 concludes the paper. The proofs of the asymptotic theorems and additional simulation results are provided in the Supplementary Material.

2. Data and Method

Let $\{X(t) : t \geq 0\}$ be a continuous-time nonhomogeneous Markov process with a finite state space $\mathcal{I} = \{0, 1, \ldots, q\}$. The stochastic behavior of X can be described by the $(q+1) \times (q+1)$ transition probability matrix $\mathbf{P}_0(s,t) = (P_{hj}(s,t))$, with elements

$$P_{hj}(s,t) = \Pr(X(t) = j | X(s) = h, \mathcal{X}_{s^-}) = \Pr(X(t) = j | X(s) = h), \qquad h, j \in \mathcal{I},$$

where $\mathcal{X}_t = \sigma\left\langle\{N_{hj}(u) : 0 \le u \le t, \quad h, j \in \mathcal{I}\}\right\rangle$ is the σ-algebra generated by the counting processes $N_{hj}(t)$, which count the direct transitions from state $h \in \mathcal{I}$ to state $j \in \mathcal{I}$, with $h \neq j$, in $[0,t]$. The conditional independence of the transition probabilities from the past history of the process is the so-called Markov property. An absorbing state h is a state for which $P_{hj}(s,t) = 0$, for all $j \neq h$, and $t \in (s, \tau]$, whereas a transient state is not absorbing. Let $\mathcal{T} = \{h_1, \ldots, h_k\} \subset \mathcal{I}$ denote the absorbing-state subspace. The transition probability matrix for Markov processes with absorbing states can be expressed as

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_{\mathcal{T}^c} & \mathbf{P}_{\mathcal{T}} \\ \mathbf{0}_{k\times(q-k+1)} & I_k \end{pmatrix},$$

where $\mathbf{P}_{\mathcal{T}^c}$ and $\mathbf{P}_{\mathcal{T}}$ are the transition probability submatrices for the transitions from the transient states to the transient states and to the absorbing states, respectively, $\mathbf{0}_{k\times(q-k+1)}$ is a $k \times (q-k+1)$ matrix containing zeros, and I_k is the $k \times k$ identity matrix. The transition intensities are defined as

$$\alpha_{hj}(t) = \lim_{u\downarrow 0} \frac{1}{u} P_{hj}(t, t+u), \qquad h \neq j, \quad h, j \in \mathcal{I}.$$

Additionally, define the $(q+1) \times (q+1)$ integrated transition intensity matrix $\mathbf{A}(t)$ with elements

$$A_{hj}(t) = \int_0^t \alpha_{hj}(u) du, \qquad h, j \in \mathcal{I},$$

where $\alpha_{hh} \equiv -\sum_{j\neq h} \alpha_{hj}$ because $\sum_j P_{hj}(s,t) = 1$, for all $s, t \in [0, \tau]$, by the definition of a stochastic matrix. Then, the transition probability matrix can be defined as the product integral of the cumulative transition intensity matrix (Andersen et al. (1993)),

$$\mathbf{P}(s,t) = \prod_{(s,t]} \left[\mathbf{I} + d\mathbf{A}(u)\right].$$

The observations from n independent and identically distributed (i.i.d.) subjects followed over the interval $[0, \tau]$, with $\tau < \infty$, are the counting processes $N_{ihj}(t)$, which count the observed direct transitions from h to j of subject $i = 1, \ldots, n$ on $[0,t]$, and the at-risk processes $Y_{ih}(t)$, which indicate that the ith subject is at state $h \in \mathcal{I}$ just before t. Note that $N_{ihj}(t)$ can be > 1 for transient states, although we restrict our discussion to the case where the counting processes are uniformly bounded by some finite constant. Right censoring and/or left truncation can be directly incorporated in the at-risk process $Y_{ih}(t)$, which is no longer a monotonic function, owing to both the left truncation and the fact that subjects may visit a transient state more than once. The counting processes

$N_{ihj}(t)$ are governed by the transition intensities of the form $\lambda_{hj}(t) = \alpha_{hj}(t)Y_h(t)$, with $h \neq j$ and $t \in [0, \tau]$.

We can estimate the elements of the integrated transition intensity matrix using the Nelson–Aalen estimator,

$$\hat{A}_{hj}^{NA}(t) = \int_0^t \frac{dN_{\cdot hj}(u)}{Y_{\cdot h}(u)}, \qquad h \neq j, \tag{2.1}$$

where $Y_{\cdot h}(t) = \sum_{i=1}^n Y_{ih}(t)$ and $N_{\cdot hj}(t) = \sum_{i=1}^n N_{ihj}(t)$. In addition, we can estimate the transition probability matrix using the Aalen–Johansen estimator (Aalen and Johansen (1978)),

$$\hat{\mathbf{P}}_n^{AJ}(s,t) = \prod_{(s,t]} \left[\mathbf{I} + d\hat{\mathbf{A}}_n^{NA}(u)\right], \tag{2.2}$$

where $\hat{\mathbf{A}}_n^{NA}$ is a matrix with elements $\hat{A}_{hj}^{NA}(t)$, for $h, j \in \mathcal{I}$. In fact, $\hat{\mathbf{A}}_n^{NA}$ is the nonparametric maximum likelihood estimator (NPMLE) of $\mathbf{A}_0$, the true integrated transition intensity matrix, based on the likelihood for discrete-time Markov chains under the assumption of independent and noninformative right censoring and left truncation (Andersen et al. (1993)):

$$\prod_t \prod_h \left\{ \prod_{j \neq h} [Y_{\cdot h}(t) dA_{hj}(t)]^{dN_{\cdot hj}(t)} [1 - dA_{h\cdot}(t)]^{Y_{\cdot h}(t) - dN_{\cdot h \cdot}(t)} \right\}, \tag{2.3}$$

where $A_{h\cdot}(t) = \sum_{j \neq h} A_{hj}(t)$ and $N_{\cdot h \cdot}(t) = \sum_{j \neq h} N_{\cdot hj}(t)$. Because the Aalen–Johansen estimator (2.2) is a one-to-one function of the NPMLE $\hat{\mathbf{A}}_n^{NA}$, it is also an NPMLE of $\mathbf{P}_0$, the true transition probability matrix (Andersen et al. (1993)).

2.1. Inferences with missing absorbing states

In this study, we assume that the absorbing states are MAR. In cases with incomplete absorbing-state ascertainment, let R_i be the "response" indicator, with $R_i = 1$ if the absorbing state has been observed, and $R_i = 0$ otherwise. Additionally, let $\mathbf{Z}_i \in \mathcal{Z} \subset \mathbb{R}^p$ be an auxiliary covariate vector that may contain information about the true unobserved absorbing state, such as a diagnosis obtained from an imperfect absorbing-state ascertainment procedure and the last state visited prior to the arrival at the absorbing state. Such information is critical, in practice, to making the MAR assumption plausible (Lu and Tsiatis (2001)) and potentially increasing the efficiency of the estimator. Next, let δ_{ij} and $\delta_i = \sum_{j=1}^k \delta_{ij}$ be indicators that the ith subject has reached the absorbing state $j \in \mathcal{T}$ and any absorbing state, respectively. The observed data for the ith subject are

$$D_i = \begin{cases} (\mathbf{N}_i, \mathbf{Y}_i, \delta_i, R_i, \mathbf{Z}_i) \text{ if } \delta_i = 0, \\ (\mathbf{N}_i, \mathbf{Y}_i, \boldsymbol{\delta}_i, R_i, \mathbf{Z}_i) \text{ if } \delta_i = 1 \text{ and } R_i = 1, \\ (\mathbf{N}_i^\star, \mathbf{Y}_i, \delta_i, R_i, \mathbf{Z}_i) \text{ if } \delta_i = 1 \text{ and } R_i = 0, \end{cases}$$

where $\mathbf{N}_i = (N_{ihj} : h \neq j)$, $\mathbf{N}_i^\star$ is equal to $\mathbf{N}_i$, with $N_{ihj}(t)$ replaced by $N_{ih\cdot}(t) = \sum_{j \in \mathcal{T}} N_{ihj}(t)$, for all $j \in \mathcal{T}$, which is a one-jump counting process, $\mathbf{Y}_i = (Y_{i0}, \ldots, Y_{iq})^T$, and $\boldsymbol{\delta}_i = (\delta_{i1}, \ldots, \delta_{ik})^T$. The absorbing-state-specific counting processes can be expressed as $N_{ihj}(t) = \delta_{ij} N_{ih\cdot}(t)$, for $h \notin \mathcal{T}$, $j \in \mathcal{T}$. We propose replacing the missing $dN_{\cdot hj}(t)$, for $j \in \mathcal{T}$, in the logarithm of the likelihood (2.3), which is linear in the missing data $dN_{\cdot hj}(t)$, by

$$\begin{aligned} E[dN_{\cdot hj}(t)|\mathbf{D}] &\equiv d\tilde{N}_{\cdot hj}(t) \\ &= \sum_{i=1}^{n} [R_i \delta_{ij} + (1 - R_i) E(\delta_{ij}|\mathbf{D})] dN_{ih\cdot}(t), \end{aligned} \tag{2.4}$$

where $\mathbf{D}$ denotes the observed data D_i, for all $i = 1, \ldots, n$. Following Cook and Kosorok (2004), we propose replacing $E(\delta_{ij}|\mathbf{D})$ by an estimate $\pi_j(\mathbf{Z}_i, \hat{\boldsymbol{\beta}}_n)$, based on the maximum likelihood under a parametric "working" model (such as the multinomial logit model), using the subjects in any absorbing state with known δ_{ij} and utilizing the auxiliary information $\mathbf{Z}_i$. This approach is valid under the MAR assumption because

$$\Pr(\delta_{ij} = 1 | R_i = 1, \mathbf{Z}_i) = \Pr(\delta_{ij} = 1 | R_i = 0, \mathbf{Z}_i) \equiv \pi_j(\mathbf{Z}_i, \boldsymbol{\beta}_0),$$

where $\boldsymbol{\beta}_0$ is the true parameter value. Maximizing the resulting pseudolikelihood, which involves $\pi_j(\mathbf{Z}_i, \hat{\boldsymbol{\beta}}_n)$, gives the NPMPLEs

$$\hat{A}_{n,hj}(t) = \int_0^t \frac{d\tilde{N}_{\cdot hj}(u; \hat{\boldsymbol{\beta}}_n)}{Y_{\cdot h}(u)}, \quad h \notin \mathcal{T}, \quad j \in \mathcal{T}, \quad t \in [0, \tau],$$

and $\hat{A}_{n,hj}(t) = \hat{A}_{hj}^{NA}(t)$ if $j \notin \mathcal{T}$, with $h \neq j$, of the integrated transition intensities. Now, the NPMPLE of the transition probability matrix is given by the plug-in estimator

$$\hat{\mathbf{P}}_n(s,t) = \prod_{(s,t]} \left[\mathbf{I} + d\hat{\mathbf{A}}_n(u) \right], \tag{2.5}$$

where the components of $\hat{\mathbf{A}}_n$ are $\hat{A}_{n,hj}$, which are given above. In Section 3, we present our methodology for constructing $1 - \alpha$ pointwise confidence intervals and simultaneous confidence bands for the components of $\mathbf{P}_0(s,t)$.

2.2. Goodness-of-fit procedure

To simultaneously evaluate the parametric model assumption for $\pi_j(\mathbf{Z}_i, \boldsymbol{\beta}_0)$, for $j \in \mathcal{T}$, we provide a goodness-of-fit procedure. First, we define the (estimated) residual processes $L_j(t;\hat{\boldsymbol{\beta}}_n) = n^{-1}\sum_{i=1}^n L_{ij}(t;\hat{\boldsymbol{\beta}}_n)$, where

$$L_{ij}(t;\hat{\boldsymbol{\beta}}_n) = R_i[N_{i\cdot j}(t) - \pi_j(\mathbf{Z}_i,\hat{\boldsymbol{\beta}}_n)N_{i\cdot\cdot}(t)], \quad j \in \mathcal{T}^{(-1)}, \ \ t \in [0,\tau],$$

with $N_{i\cdot j}(t) = \sum_{h\notin\mathcal{T}} N_{ihj}(t)$ and $N_{i\cdot\cdot}(t) = \sum_{h\notin\mathcal{T}}\sum_{j\in\mathcal{T}} N_{ihj}(t)$ being the processes that count the transitions to the absorbing state j and to any absorbing state by time $t \in [0,\tau]$, respectively. In addition, $\mathcal{T}^{(-1)} \subset \mathcal{T}$ denotes the absorbing-state subspace that includes $k-1$ absorbing states. Note that only $k-1$ residual processes are considered, because the model for one absorbing state is determined completely by the models for the remaining $k-1$ absorbing states. To construct a formal statistical test for the goodness of fit and a diagnostic plot for the parametric absorbing-state probability model, we follow a procedure similar to that developed by Pan and Lin (2005). First, it can be shown (Supplementary Material) that under the null hypothesis $E[L_j(t;\boldsymbol{\beta}_0)] = 0$, we have

$$V_{nj}(t) \equiv \sqrt{n}L_j(t;\hat{\boldsymbol{\beta}}_n) = n^{-1/2}\sum_{i=1}^n \psi_{ij}^L(t) + o_p(1),$$

where $\psi_{ij}^L(t) = L_{ij}(t;\boldsymbol{\beta}_0) - \boldsymbol{\omega}_i^T E[\dot{\pi}_j(\mathbf{Z}_i,\boldsymbol{\beta}_0)R_iN_{i\cdot\cdot}(t)]$. Here, $\boldsymbol{\omega}_i$ is the ith individual influence function for $\hat{\boldsymbol{\beta}}_n$, given by $\boldsymbol{\omega}_i = I^{-1}(\boldsymbol{\beta}_0)U_i(\boldsymbol{\beta}_0)$, where $I(\boldsymbol{\beta}_0)$ is the Fisher information about $\boldsymbol{\beta}_0$, $U_i(\boldsymbol{\beta}_0)$ is the individual score function for the ith subject, and $\dot{\pi}_j(\mathbf{Z}_i,\boldsymbol{\beta}_0) = \partial\pi_j(\mathbf{Z}_i,\boldsymbol{\beta})/\partial\boldsymbol{\beta}|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$. The influence functions $\psi_{ij}^L(t)$ can be estimated by replacing the unknown components with the corresponding estimated components and the expectation with the sample average; that is,

$$\hat{\psi}_{ij}^L(t) = L_{ij}(t;\hat{\boldsymbol{\beta}}_n) - \hat{\boldsymbol{\omega}}_i^T n^{-1}\sum_{i=1}^n[\dot{\pi}_j(\mathbf{Z}_i,\hat{\boldsymbol{\beta}}_n)R_iN_{i\cdot\cdot}(t)],$$

where $\hat{\boldsymbol{\omega}}_i = \hat{I}^{-1}(\hat{\boldsymbol{\beta}}_n)U_i(\hat{\boldsymbol{\beta}}_n)$. Now, define $\hat{V}_{nj}(t) = n^{-1/2}\sum_{i=1}^n \hat{\psi}_{ij}^L(t)\xi_{ij}$, with ξ_{ij}, $i = 1,\ldots,n$, drawn randomly from $N(0,1)$. The goodness of fit for the parametric model can be evaluated as follows:

1. Simulate many $\{\xi_{ij}\}_{i\in\{1,\ldots,n\},j\in\mathcal{T}^{(-1)}}$ sets of values from $N(0,1)$.

2. For each simulated set $\{\xi_{ij}\}_{i\in\{1,\ldots,n\},j\in\mathcal{T}^{(-1)}}$, given $\hat{\psi}_{ij}^L(t)$, calculate the quantity $\sup_{t\in[0,\tau]}\max_{j\in\mathcal{T}^{(-1)}}|\hat{V}_{nj}(t)|$.

3. Calculate the $1-\alpha$ percentile of the distribution of $\sup_{t\in[0,\tau]}\max_{j\in\mathcal{T}^{(-1)}}|\hat{V}_{nj}(t)|$ values, denoted by $c_{1-\alpha}$.

4. Calculate the simultaneous confidence band for $E[L_j(t;\boldsymbol{\beta}_0)] = 0$ as $\pm n^{-1/2}\hat{c}_{1-\alpha}$,

and plot it along with the residual processes $L_j(t;\hat{\boldsymbol{\beta}}_n)$, for $j \in \mathcal{T}^{(-1)}$ and $t \in [0,\tau]$.

5. Calculate the p-value for the null hypothesis of the overall goodness of fit as the proportion of $\sup_{t\in[0,\tau]} \max_{j\in\mathcal{T}^{(-1)}} |\hat{V}_{nj}(t)|$ values that are greater than or equal to $\sup_{t\in[0,\tau]} \max_{j\in\mathcal{T}^{(-1)}} |v_{nj}(t)|$, where $v_{nj}(t)$ is the observed value of the $V_{nj}(t)$ statistic, based on the data.

A lack of fit for the parametric model $\pi_j(\mathbf{Z}_i, \boldsymbol{\beta}_0)N_{i\cdot\cdot}(t)$, for $j \in \mathcal{T}$, is indicated by a type-I error α if the residual process for at least one $j \in \mathcal{T}^{(-1)}$ is not contained in the confidence band for $E[L_j(t;\boldsymbol{\beta}_0)] = 0$. Equivalently, a p-value less than α provides evidence of a lack of fit for at least one absorbing-state model. The validity of this approach is ensured by Theorem 3, which is stated in Section 3.

3. Asymptotic Theory

Assume that the following regularity conditions hold:

C1. The follow-up interval is $[0,\tau]$, with $\tau < \infty$.

C2. $\Pr(N_{hj}(\tau) \leq C) = 1$, for some constant $C \in (0,\infty)$, for all $h, j \in \mathcal{I}$. In addition, $\inf_{t\in[0,\tau]} E[Y_h(t)] > 0$, for all $h \notin \mathcal{T}$, which implies that the expected number of observations at all transient states is positive for any time $t \in [0,\tau]$.

C3. $\mathbf{A}_0$ is a $(q+1)\times(q+1)$ matrix-valued function with elements that are continuous functions of bounded variation on $[0,\tau]$.

C4. The inverse of the link function for the model of the absorbing-state vector $\boldsymbol{\delta}$ has a continuous derivative on compact sets. In addition, the corresponding parameter space $\mathcal{B}$ is a bounded subset of $\mathbb{R}^p$.

C5. The estimator $\hat{\boldsymbol{\beta}}_n$ of the true model parameter $\boldsymbol{\beta}_0$ for the absorbing states $\boldsymbol{\delta}$ is strongly consistent and asymptotically linear; that is, $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = n^{-1/2}\sum_{i=1}^n \boldsymbol{\omega}_i + o_p(1)$, with $\boldsymbol{\omega}_i$ being i.i.d., $E\boldsymbol{\omega}_i = \mathbf{0}$, and $E\|\boldsymbol{\omega}_i\|^2 < \infty$. Additionally, the plug-in estimators of $\boldsymbol{\omega}_i$, $\hat{\boldsymbol{\omega}}_i$, for $i = 1,\ldots,n$, satisfy $n^{-1}\sum_{i=1}^n \|\hat{\boldsymbol{\omega}}_i - \boldsymbol{\omega}_i\|^2 = o_p(1)$.

C6. The auxiliary covariate vector $\mathbf{Z}$ is bounded in the sense that there exists a constant $K \in (0,\infty)$, such that $\Pr(\|\mathbf{Z}\| \leq K) = 1$.

Note that estimating $\boldsymbol{\beta}_0$ using the maximum likelihood under a correctly specified generalized linear model and assuming that the proportion of missing data is independent of the sample size imply C5. Before stating the asymptotic theory results, we introduce some further notation. First, the NPMPLE can be expressed as

$$\hat{\mathbf{P}}_n(s,t) = \begin{pmatrix} \hat{\mathbf{P}}_{n,\mathcal{T}^c}(s,t) & \hat{\mathbf{P}}_{n,\mathcal{T}}(s,t) \\ \mathbf{0}_{k\times(q-k+1)} & I_k \end{pmatrix}.$$

Next, define the influence functions

$$\gamma_{ihj}(s,t) = \sum_{l\notin\mathcal{T}}\sum_{m\in\mathcal{I}}\int_s^t P_{0,hl}(s,u-)P_{0,mj}(u,t)d\psi_{ilm}(u),$$

for $h \notin \mathcal{T}$, $j \in \mathcal{I}$, and $i = 1,\ldots,n$, where

$$\psi_{ilm}(t) = \begin{cases} \displaystyle\int_0^t \frac{d\tilde{N}_{ilm}(u;\boldsymbol{\beta}_0)}{EY_l(u)} - \int_0^t \frac{Y_{il}(u)}{EY_l(u)}dA_{0,lm}(u) + \boldsymbol{\omega}_i^T\mathbf{R}_{lm}(t) & \text{if } \ m \in \mathcal{T}, \\ \displaystyle\int_0^t \frac{dN_{ilm}(u)}{EY_l(u)} - \int_0^t \frac{Y_{il}(u)}{EY_l(u)}dA_{0,lm}(u) & \text{if } \ m \notin \mathcal{T}, \end{cases}$$

for $l \neq m$, where $\mathbf{R}_{lm}(t) = E\{(1-R)\dot{\pi}_m(\mathbf{Z},\boldsymbol{\beta}_0)\int_0^t[EY_l(u)]^{-1}dN_{l\cdot}(u)\}$. If $l = m$, then $\psi_{ill}(t) = -\sum_{h\neq l}\psi_{ilh}(t)$. Moreover, define

$$\hat{W}_{n,hj}(s,t) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{\gamma}_{ihj}(s,t)\xi_i, \qquad 0 \le s < t \le \tau,$$

where $\hat{\gamma}_{ihj}(s,t)$ denotes an estimated influence function, where the unknown quantities have been replaced by their consistent estimators and the expectations have been replaced by sample averages, and ξ_i are independent draws from $N(0,1)$. Given the regularity conditions C1–C6, the following theorems hold.

Theorem 1. *The NPMPLE is uniformly consistent in the sense that*

$$\sup_{t\in(s,\tau]}\left\|\hat{\mathbf{P}}_n(s,t) - \mathbf{P}_0(s,t)\right\| \overset{as*}{\to} 0,$$

for any $s \in [0,\tau)$*, where the norm* $\|\mathbf{A}\|$ *stands for* $\sup_h\sum_l|a_{hl}|$ *for the matrix* $\mathbf{A} = [a_{hl}]$.

Theorem 2. *The NPMPLE is an asymptotically linear estimator, with*

$$\sqrt{n}\left[\hat{\mathbf{P}}_n(s,t) - \mathbf{P}_0(s,t)\right] = \frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{\gamma}_i(s,t) + \boldsymbol{\epsilon},$$

where $\boldsymbol{\gamma}_i(s,t)$ *is a matrix-valued function with elements* $\gamma_{ihj}(s,t)$ *that belong to Donsker classes, and* $\boldsymbol{\epsilon}$ *is a* $(q+1)\times(q+1)$ *matrix with elements that are* $o_p(1)$.

Moreover, $\hat{W}_{n,hj}(s,\cdot)$ *converges weakly conditional on the observed data* $\mathbf{D}$ *to the same limiting process as that of* $\sqrt{n}[\hat{P}_{n,hj}(s,\cdot) - P_{0,hj}(s,\cdot)]$ *(unconditionally), for any* $s \in [0,\tau)$, $h \notin \mathcal{T}$, *and* $j \in \mathcal{I}$.

Theorem 3. *The goodness-of-fit statistic* $\sup_{t\in[0,\tau]} \max_{j\in\mathcal{T}^{(-1)}} |\hat{V}_{nj}(t)|$ *converges weakly conditional on the data* $\mathbf{D}$ *to the same limiting process as that of* $\sup_{t\in[0,\tau]} \max_{j\in\mathcal{T}^{(-1)}} |V_{nj}(t)|$ *(unconditionally).*

The proofs for the theorems are provided in the Supplementary Material.

Remark 1. The asymptotic result of Theorem 2 can be also expressed in conventional vector form as $\sqrt{n}[\hat{\mathcal{P}}_n(s,t) - \mathcal{P}_0(s,t)]$, where

$$\hat{\mathcal{P}}_n = \left(\mathrm{vec}^T\hat{\mathbf{P}}_{n,\mathcal{T}^c}, \mathrm{vec}^T\hat{\mathbf{P}}_{n,\mathcal{T}}\right)^T \text{ and } \mathcal{P}_0 = \left(\mathrm{vec}^T\mathbf{P}_{0,\mathcal{T}^c}, \mathrm{vec}^T\mathbf{P}_{0,\mathcal{T}}\right)^T.$$

Here, vec$A$ is the column vector formed by concatenating the columns of the matrix A, and $\mathrm{vec}^T A$ is the transpose of vec$A$. As a consequence of Theorem 2 and an application of the Cramer–Wold device, $\sqrt{n}[\hat{\mathcal{P}}_n(s,\cdot) - \mathcal{P}_0(s,\cdot)]$ converges weakly to a Gaussian random field, with each of its elements a tight mean-zero Gaussian process in the space $D[s,\tau]$ of cadlag functions on $[s,\tau]$. Owing to the asymptotic linearity of the NPMPLE, the corresponding asymptotic variance-covariance matrix-valued function, given the starting time point s, is equal to $\Sigma(t,w;s) = E\left[\mathrm{vec}\boldsymbol{\gamma}_i(s,t)\mathrm{vec}^T\boldsymbol{\gamma}_i(s,w)\right]$, $0 \le s < t, w \le \tau$, where $\Sigma(t,w;s)$ is a $(q-k+1)(q+1) \times (q-k+1)(q+1)$ matrix-valued process. Using this asymptotic variance-covariance matrix-valued function, and after some algebra, it can be shown that the asymptotic variance of

$$\sqrt{n}[\hat{P}_{n,hj}(s,t) - P_{0,hj}(s,t)], \quad t \in (s,\tau],$$

the hj-element of the transition probability matrix for given $s \ge 0$, can be decomposed as

$$E[\gamma^F_{ihj}(s,t)]^2 + E[\gamma^M_{ihj}(s,t)]^2 + 2E[\gamma^F_{ihj}(s,t)\gamma^M_{ihj}(s,t)], \tag{3.1}$$

where

$$\gamma^F_{ihj}(s,t) = \sum_{l\notin\mathcal{T}}\sum_{m\in\mathcal{I}}\int_s^t P_{0,hl}(s,u-)P_{0,mj}(u,t)d\psi^F_{ilm}(u),$$

with

$$\psi^F_{ilm}(t) = \int_0^t \frac{dN_{ilm}(u)}{EY_l(u)} - \int_0^t \frac{Y_{il}(u)}{EY_l(u)}dA_{0,lm}(u),$$

and where

$$\gamma_{ihj}^{M}(s,t) = \begin{cases} \displaystyle\sum_{l \notin \mathcal{T}} \int_s^t P_{0,hl}(s,u-) d\psi_{ilj}^{M}(u) & \text{if } j \in \mathcal{T}, \\ 0 \quad \text{if } j \notin \mathcal{T}, \end{cases}$$

with

$$\psi_{ilj}^{M}(t) = (1 - R_i) \int_0^t \frac{d[\pi_j(\mathbf{Z}_i, \boldsymbol{\beta}_0) N_{il\cdot}(u) - N_{ilj}(u)]}{EY_l(u)} + \boldsymbol{\omega}_i^T \mathbf{R}_{lj}(t),$$

for $j \in \mathcal{T}$. The influence function $\gamma_{ihj}^{F}(s,t)$ is the influence of the ith observation on the estimator in the ideal situation without missing absorbing states. Then, $\gamma_{ihj}^{M}(s,t)$ is the influence associated with missingness and the fact that we impute the unobserved jumps $dN_{ilj}(t)$ with $\pi_j(\mathbf{Z}_i, \hat{\boldsymbol{\beta}}_n) dN_{il\cdot}(t)$, for $j \in \mathcal{T}$. Therefore, based on decomposition (3.1), the asymptotic variance of the transition probability estimator to an absorbing state is equal to the variance of this estimator in the absence of missing data $E[\gamma_{ihj}^{F}(s,t)]^2$, plus the additional variability due to missingness $E[\gamma_{ihj}^{M}(s,t)]^2$ and two times the covariance between the influence function of the estimator without missingness and the influence function related to missingness $E[\gamma_{ihj}^{F}(s,t)\gamma_{ihj}^{M}(s,t)]$. Furthermore, the variability $E[\gamma_{ihj}^{M}(s,t)]^2$ due to missingness depends on the variability of $\hat{\boldsymbol{\beta}}_n$, through its influence function $\boldsymbol{\omega}_i$, weighted by the fixed quantity $\mathbf{R}_{lj}(t)$, which is proportional to the percent of missingness, as well as the difference between the imputed expected jump $\pi_j(\mathbf{Z}_i, \boldsymbol{\beta}_0) dN_{il\cdot}(t)$ and the actual unobserved jump $dN_{ilj}(t)$ for the missing cases. Hence, the variability of our proposed estimator for incorporating missing absorbing states is influenced by the missing rate and the total sample size.

Using regularity conditions C1–C6 and Theorems 1 and 2, it can be shown that the asymptotic variance-covariance function of the transition probability matrix estimator can be uniformly consistently (in probability) estimated by $\hat{\Sigma}_n(t,w;s) = n^{-1} \sum_{i=1}^{n} \left[\text{vec}\hat{\boldsymbol{\gamma}}_i(s,t)\text{vec}^T \hat{\boldsymbol{\gamma}}_i(s,w)\right]$, where the components of $\hat{\boldsymbol{\gamma}}_i(s,\cdot)$ are $\hat{\gamma}_{ihj}(s,\cdot)$, defined above, for $h \notin \mathcal{T}$, and $j \in \mathcal{I}$ and zero otherwise. These results, along with the functional delta method, can be used to construct $1-\alpha$ pointwise confidence intervals for $P_{0,hj}(s,t)$, under a known and differentiable transformation g (e.g. $g(x) = \log[-\log(x)]$) that ensures that the corresponding limits of the interval lie in (0, 1). For the construction of simultaneous confidence bands, consider the process $\sqrt{n} q_{hj}(s,t)\{g[\hat{P}_{n,hj}(s,t)] - g[P_{0,hj}(s,t)]\}$, where $q_{hj}(s,t)$ is a time-dependent weight that converges uniformly in probability to a nonnegative bounded function on $[t_1, t_2]$, with $0 \le s \le t_1 \le t_2 < \tau$. This weight function can be set equal to $\hat{P}_{n,hj}(s,t)/\hat{\sigma}_{hj}(s,t)$, where $\hat{\sigma}_{hj}(s,t) = [n^{-1}\sum_{i=1}^{n} \hat{\gamma}_{ihj}^2(s,t)]^{1/2}$ is the estimated standard error of $W_{n,hj}(s,t) = n^{-1/2}$

$\sum_{i=1}^n \gamma_{ihj}(s,t)$, or to $\hat{P}_{n,hj}(s,t)/[1+\hat{\sigma}^2_{hj}(s,t)]$. The first weight is equivalent to an equal precision weight (Nair (1984)) and the second to a Hall-Wellner weight (Hall and Wellner (1980)). Using Theorem 2 and the functional delta method, it can be easily shown that the process $\sqrt{n}q_{hj}(s,t)\{g[\hat{P}_{n,hj}(s,t)] - g[P_{0,hj}(s,t)]\}$ is asymptotically equivalent to the process $\hat{B}_{n,hj}(s,t) = q_{hj}(s,t)\dot{g}[\hat{P}_{n,hj}(s,t)]\hat{W}_{n,hj}(s,t)$. Next, similarly to Spiekerman and Lin (1998), define c_α as the $1-\alpha$ percentile from a large number of realizations of $\sup_{t\in[t_1,t_2]} |\hat{B}_{n,hj}(s,t)|$, generated by repeated simulations of $\{\xi_i\}_{i=1}^n$. Now, the $1-\alpha$ confidence band is

$$g^{-1}\left\{g[\hat{P}_{n,hj}(s,t)] \pm \frac{c_a}{\sqrt{n}q_{hj}(s,t)}\right\}, \qquad t\in[t_1,t_2],$$

for a given $s\in[0,t_1]$. In general, the confidence band can be unstable in the tails of the observable time domain (Yin and Cai (2004)). To resolve this issue, we can restrict the domain of the confidence band to $[u_1,u_2]$, where these limits can be set equal to the solutions of $c_l = \hat{\sigma}^2_{hj}(s,u_l)/[1+\hat{\sigma}^2_{hj}(s,u_l)]$, $l=1,2$, with $\{c_1,c_2\}=\{0.1,0.9\}$ or $\{c_1,c_2\}=\{0.05,0.95\}$ (Nair (1984); Yin and Cai (2004)).

4. Simulation Study

To evaluate the performance of the proposed estimator with finite samples and to study its robustness against a misspecification of the parametric model for the probability of the absorbing states, we conducted extensive simulation studies. We considered a nonhomogeneous Markov process with two absorbing states, denoted by 1 and 2, and one initial transient state, denoted by 0. This model is equivalent to the competing-risks model with two causes of failure. The transition probabilities for the two absorbing states were $P_{01}(0,t) = 0.4\{1-\exp[-(t/\lambda_1)^{\nu_1}]\}$ and $P_{02}(0,t) = 0.6\{1-\exp[-(t/\lambda_2)^{\nu_2}]\}$. The probability of remaining in the transient state was $P_{00}(0,t) = 1-\sum_{j=1}^2 P_{0j}(0,t)$. Four scenarios were considered: 1) $(\lambda_1,\nu_1,\lambda_2,\nu_2)^T = (1,1,0.5,1)^T$; 2) $(\lambda_1,\nu_1,\lambda_2,\nu_2)^T = (1,0.8,0.5,1)^T$; 3) $(\lambda_1,\nu_1,\lambda_2,\nu_2)^T = (1,0.4,0.5,1)^T$; and 4) $(\lambda_1,\nu_1,\lambda_2,\nu_2)^T = (1,0.2,0.5,1)^T$. Right-censoring times were simulated based on the uniform distribution, $U(0,5)$. Under these simulation settings, the average proportion of right-censored observations was 15%, and the proportion of noncensored observations in absorbing state 1 was 37%. The probability of a missing absorbing state was set to 0.8 or 0.6. To mimic a setting with planned missingness due to a double-sampling design, such as the design in our motivating HIV study (see Section 5), we considered the auxiliary covariate $\mathbf{Z} = (T, C^\star)^T$, where T is the arrival time at an absorbing state, and $C^\star$ is the absorbing state according to an imperfect diagnostic procedure. Let C denote the true, but incompletely observed absorbing state. $C^\star$ was simulated conditional on C from a Bernoulli distribution with probabilities $\pi^\star_{11} = \Pr(C^\star = 1|C = 1) = 0.9$ and $\pi^\star_{22} = \Pr(C^\star = 2|C = 2) = 0.7$. Therefore, the misclassification probabilities of the imperfect diagnostic were 0.1 and 0.3 for absorbing states 1 and 2, respectively. Note that $C^\star$ was completely observed. We considered sample sizes of $n = 200$ and $n = 400$.

In this simulation study, we evaluated the usual Aalen–Johansen estimator for the observed data by only using the misclassified absorbing state $C^\star$ and ignoring the nonmissing C values (Naïve), the Aalen-Johansen estimator under a complete-case analysis, where the observations with a missing C were discarded from the analysis (CC), and the proposed estimator. For the proposed estimator, we considered a "working" logistic model with $\text{logit}[\pi_1(\mathbf{Z}, \boldsymbol{\beta})] = \beta_0 + \beta_1 T + \beta_2 I_{\{C^\star=1\}}$. Note that the true probability of absorbing state 1 under the four simulation scenarios is given by

$$\text{logit}[\pi_1(T, C^\star; \boldsymbol{\beta})] = \beta_0 + f(T; \lambda_1, \nu_1, \lambda_2, \nu_2) + \beta_2 I_{\{C^\star=1\}},$$

where

$$f(T; \lambda_1, \nu_1, \lambda_2, \nu_2) = -\lambda_1^{-\nu_1} T^{\nu_1} - \lambda_2^{-\nu_2} T^{\nu_2} + (\nu_1 - \nu_2)\log(T).$$

Setting $\nu_1 = \nu_2 = 1$ in Scenario 1 implied a linear logit model for the probability of absorbing state 1 in T, of the form $\text{logit}[\pi_1(\mathbf{Z}, \boldsymbol{\beta}_0)] = \beta_0 + \beta_1 T + \beta_2 I_{\{C^\star=1\}}$; Scenarios 2–4 employed nonlinear logit models in T. Therefore, our "working" linear logit model was correctly specified in Scenario 1 and was misspecified in Scenarios 2–4. The nonlinear dependence of $\text{logit}[\pi_1(\mathbf{Z}, \boldsymbol{\beta}_0)]$ on T in Scenarios 2–4, which corresponds to a misspecification of our "working" model, is depicted in Figure S1 in the Supplementary Material. The Figure shows that the degree of nonlinearity on T, and thus the degree of the linear logit model misspecification, increases as ν_1 decreases. To construct the 95% simultaneous confidence bands, we performed 1,000 simulations of sets $\{\xi_i\}_{i=1}^n$ of i.i.d. random variables from $N(0, 1)$, and considered both equal-precision and Hall–Wellner-type weights.

Pointwise simulation results for absorbing state 1 under Scenario 1 are presented in Table 1. In all cases, the naïve approach yielded highly biased estimates. The CC analysis also provided biased estimates, as well as coverage probabilities that were lower than the nominal 95% level. In contrast, the proposed NPMPLE provided virtually unbiased estimates. Furthermore, the estimated standard errors (ASE) were close to the Monte Carlo standard deviations (MCSD) of the estimates, and the coverage probabilities were close to the nominal 95% level, even with 80% missing absorbing states and $n = 200$. Interestingly, the MCSD

Table 1. Pointwise simulation results for estimating the transition probability of absorbing state 1 at $t_1 = 0.4$, $t_2 = 0.8$, and $t_3 = 1.2$, under the naïve approach, the complete case analysis (CC), and the proposed method, under Scenario 1.

	Bias$\times 10^2$			MCSD$\times 10^3$			ASE$\times 10^3$			CP$\times 10^2$		
Method (missing)	t_1	t_2	t_3	t_1	t_2	t_3	t_1	t_2	t_3	t_1	t_2	t_3
$n = 200$												
Naïve	8.5	12.2	13.4	29.4	35.0	36.8	29.8	35.0	36.9	15.6	4.6	3.7
CC (80%)	−3.9	−5.5	−5.6	39.4	53.8	64.7	38.6	53.6	65.0	74.4	75.7	80.7
CC (60%)	−1.9	−2.5	−2.5	33.7	44.7	50.7	33.5	44.0	50.6	86.2	88.5	89.0
Proposed (80%)	−0.2	0.0	0.0	42.2	56.4	62.7	40.0	53.9	60.1	92.9	93.4	92.4
Proposed (60%)	−0.1	−0.1	−0.1	30.4	40.0	44.7	30.8	40.8	45.5	93.8	95.0	95.4
$n = 400$												
Naïve	8.7	12.2	13.6	21.3	24.6	26.0	21.1	24.7	26.1	0.8	0.1	0.0
CC (80%)	−3.9	−5.2	−5.4	28.4	39.0	46.8	27.5	38.2	45.8	65.1	69.2	74.3
CC (60%)	−1.8	−2.3	−2.2	23.9	31.6	37.0	23.7	31.1	35.6	84.1	85.4	88.0
Proposed (80%)	−0.1	−0.1	−0.1	28.6	38.6	42.9	27.9	37.9	42.5	92.2	92.8	93.3
Proposed (60%)	−0.1	−0.1	−0.1	21.7	28.6	31.3	21.6	28.7	32.0	94.4	94.8	94.8

MCSD, Monte Carlo standard deviation; ASE, average standard error; CP, coverage probability

Table 2. Pointwise simulation results for estimating the transition probability of absorbing state 1 at $t_1 = 0.4$, $t_2 = 0.8$, and $t_3 = 1.2$, under the naïve approach, the complete case analysis (CC), and the proposed method, under Scenario 2.

	Bias$\times 10^2$			MCSD$\times 10^3$			ASE$\times 10^3$			CP$\times 10^2$		
Method (missing)	t_1	t_2	t_3	t_1	t_2	t_3	t_1	t_2	t_3	t_1	t_2	t_3
$n = 200$												
Naïve	8.3	12.1	13.5	30.7	35.2	36.6	30.6	35.0	36.7	20.6	5.4	3.8
CC (80%)	−4.9	−6.5	−6.9	41.2	51.6	59.9	39.8	51.5	60.2	68.0	67.0	73.6
CC (60%)	−2.4	−3.0	−3.2	35.8	43.8	48.7	35.1	43.4	48.7	84.2	85.8	85.4
Proposed (80%)	−0.4	0.2	0.3	44.7	57.1	62.6	42.6	54.8	60.2	93.2	93.4	93.1
Proposed (60%)	−0.3	0.0	0.1	32.5	40.7	44.4	32.8	41.3	45.3	93.8	95.2	95.2
$n = 400$												
Naïve	8.5	12.1	13.6	21.9	24.8	26.1	21.7	24.7	25.9	1.8	0.1	0.0
CC (80%)	−4.9	−6.3	−6.7	28.7	37.7	43.5	28.5	36.6	42.6	55.9	58.3	61.4
CC (60%)	−2.3	−2.9	−2.9	24.8	31.6	35.6	24.8	30.7	34.3	81.3	80.3	83.0
Proposed (80%)	−0.4	0.0	0.2	29.8	38.9	43.0	29.7	38.5	42.5	92.7	93.0	93.3
Proposed (60%)	−0.4	0.0	0.1	22.5	28.6	31.3	23.0	29.0	31.9	94.5	95.7	95.0

MCSD, Monte Carlo standard deviation; ASE, average standard error; CP, coverage probability

of our estimator was larger than that of the CC analysis at time points t_1 and t_2 with 80% missingness and $n = 200$. This is attributed to i) the large variability of $\hat{\boldsymbol{\beta}}_n$, the estimated parameter of the model for the probability of absorbing

Table 3. Simulation results on the coverage probability of the proposed 95% simultaneous confidence bands based on equal-precision (EP) and Hall–Wellner-type (HW) weights, under Scenarios 1 and 2.

		Scenario 1		Scenario 2	
n	missing	EP	HW	EP	HW
200	80%	93.6	92.9	94.3	93.2
	60%	95.3	96.1	95.6	96.2
400	80%	93.6	93.3	94.2	93.8
	60%	94.7	95.5	96.3	95.6

state 1, because it was estimated using only 34 observations, on average, and ii) the fact that $\hat{\boldsymbol{\beta}}_n$ was used to impute the missing counting process jumps in a relatively large number of cases (i.e., 136, on average, or 80% of the non-right-censored cases). When the sample size was 400 or the missing rate was 60%, this phenomenon was almost gone. In addition, our estimator was more efficient than the CC analysis, except at time point t_1 for $n = 400$ and a missing rate of 80%. The proposed estimator may have a slightly larger standard error in some cases compared with that of the CC analysis when the sample size is not large and the missing rate is high. However, our estimator still outperforms the CC estimator in terms of the mean squared error, because the CC analysis usually yields biased estimates.

The simulation results under Scenario 2, where the proposed method was evaluated under a misspecified parametric model $\pi_1(\mathbf{Z}, \boldsymbol{\beta})$, are presented in Table 2. Again, the naïve approach and the CC analysis provided biased estimates. The proposed approach performed well, as in Scenario 1. The simulation results for the performance of the simultaneous confidence bands are presented in Table 3. The coverage probabilities for the 95% simultaneous confidence bands were close to the nominal level, even with 80% missing absorbing states, $n = 200$, and a misspecified parametric model $\pi_1(\mathbf{Z}, \boldsymbol{\beta})$. The simulation results for a more pronounced misspecification of the probability model of absorbing state 1 (Scenarios 3 and 4) are reported in Tables S1–S3 and Figure S2 in the Supplementary Material. The pointwise results in Tables S1 and S2 reveal that the more pronounced misspecification of $\pi_1(\mathbf{Z}, \boldsymbol{\beta})$ led to greater bias in the transition probability estimates. However, the degree of bias under the misspecified models was still much smaller than that in the naïve and CC analyses, and was almost negligible compared with the corresponding true values (Figure S2). Moreover, the ASEs were close to the corresponding MCSDs, and the coverage probabilities were close to the nominal level, in all cases. When considering the overall estimated transition probability functions (Figure S2), it appears that the bias levels were small in general, even under a severely misspecified model $\pi_1(\mathbf{Z}, \boldsymbol{\beta})$ (Scenario 4). Thus, it is evident that the proposed estimator is robust against some degree of misspecification of the "working" model $\pi_1(\mathbf{Z}, \boldsymbol{\beta})$. Nevertheless, the impact of a misspecification was more pronounced in the coverage of the simultaneous confidence bands, especially under Scenario 4 (Table S3 in the Supplementary Material).

The efficiency of our estimator is expected to depend on the missing rate and the accuracy of the auxiliary variable $C^\star$. To evaluate this efficiency dependence numerically, we performed further simulation experiments by varying the missing rate from 0% to 80%, while keeping $\pi^\star_{11}$ and $\pi^\star_{22}$ fixed at 0.9 and 0.7, respectively. We also varied $\pi^\star_{11}$ from 0.5 to 0.9, setting $\pi^\star_{22} = \pi^\star_{11}$ and the missing rate to 80%. The simulation results on the MCSD of the estimated transition probability at $t = 0.4$, 0.8, and 1.2, based on 1,000 simulations, are presented in Figure S3 and Table S4 in the Supplementary Material. As expected, a higher missing rate led to a larger estimation standard error, and a higher accuracy of $C^\star$ led to a smaller estimation standard error. Interestingly, the effect of the accuracy of $C^\star$ on the standard error was not pronounced.

We also compared our method with that of Gouskova, Lin and Fine (2017) (GLF) for the competing-risks model, which does not incorporate auxiliary covariates, by considering Scenarios 1–3 with $n = 400$. We did not consider Scenario 4, because the GLF estimator was highly unstable in this case. In these simulations, we generated the missingness according to the following two scenarios: i) missing completely at random (MCAR), where the probability of missingness did not depend on the auxiliary variable $C^\star$, with $\Pr(R = 0) = 0.6$; and ii) MAR, where the probability of missingness depended on the auxiliary variable $C^\star$, with $\Pr(R = 0|C^\star) = 0.5 + 0.2I_{\{C^\star=1\}}$. These simulation results, presented in Tables S5 and S6 in the Supplementary Material, revealed that the GLF estimator always had a larger mean squared error than that of our proposed method, even when our parametric model $\pi_1(\mathbf{Z}, \boldsymbol{\beta})$ was misspecified (Scenarios 2 and 3). Moreover, the GLF estimator was severely biased when the probability of missingness depended on the auxiliary variable $C^\star$.

To illustrate the computational efficiency of our estimator, we present the average computation times (in seconds) and the corresponding standard deviations, based on 100 simulations, in Table S7 in the Supplementary Material. These figures correspond to the time needed to compute the transition probability estimates and the associated standard errors, with and without simultaneous confidence bands, for sample sizes $n = 200$ to $n = 1,500$, under Scenario 1. The computation times under Scenarios 2–4 were similar. Finally, we investigated the performance of the naïve approach based on the diagnostic accuracy of $C^{\star}$ under Scenario 1. These results are presented in Table S8 in the Supplementary Material. As expected, a lower accuracy of $C^{\star}$ was associated with a larger bias in the naïve approach because of the higher misclassification rate of $C^{\star}$.

In summary, our extensive simulation studies provided sufficient evidence to numerically justify the superior statistical and computational efficiency properties of our proposed method for estimating the transition probabilities of nonhomogeneous Markov processes with partially observed absorbing states.

5. HIV Data Analysis

From an implementation science perspective, the primary outcome of interest in HIV care is how adhesive patients are to care; this was the main objective in our motivating study. As such, the proposed method was applied to estimate the transition probabilities of disengagement from care and death while in care, based on data from the East Africa IeDEA study. A major issue in this ongoing study is the significant under-reporting of deaths. Here, unreported deaths are incorrectly classified as disengagements from care, because deceased patients do not return to care. To deal with this issue, a double-sampling design was applied in the IeDEA study, where a small sample of patients who were lost to the clinic were found in the community by outreach workers, who then ascertained the correct vital status of each patient. The database consisted of 58,876 HIV-infected individuals who initiated antiretroviral therapy (ART) and had a CD4 count below 350 cells/μl. Throughout the study, 3,338 (5.7%) patients were (passively) recorded as dead, and 27,034 (45.9%) were lost to clinic. The remaining patients were alive and in care at the data closure date; their arrival times at an absorbing state were considered administratively right-censored. In this data set, 4,020 (14.9%) of the 27,034 patients who were lost were doubly sampled and outreached within a short period after they were flagged as disengagers by the clinicians. Among these doubly sampled patients, 917 (22.8%) were actually dead, indicating a significant under-reporting issue. The vital status was missing for the remaining 85.1% of the lost patients who were not doubly sampled.

At the first stage of the analysis, we considered a logistic regression model for the probability of death among those who were flagged as disengagers, with a linear effect of time from the ART initiation. We evaluated the goodness of fit

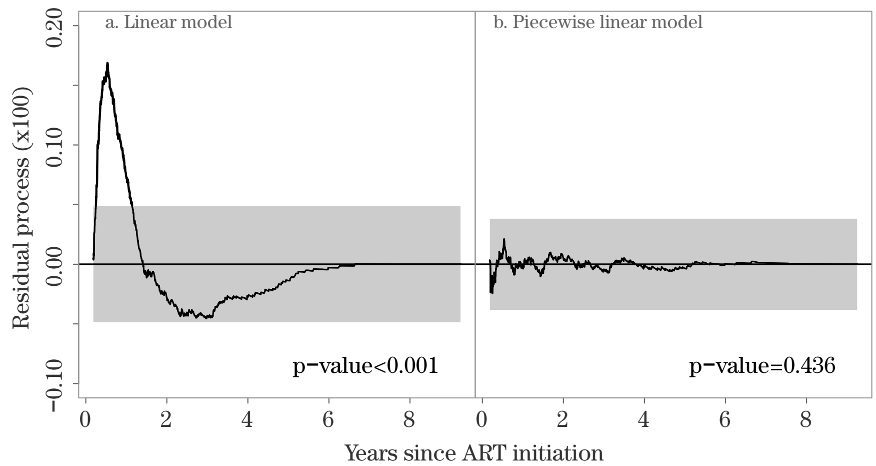

Figure 1. Residual process for the parametric model $\pi_1(\mathbf{Z}, \boldsymbol{\beta})$, based on the IeDEA HIV data, along with the 95% goodness-of-fit band (grey area) and the corresponding p-value.

of this model using the residual process presented in Section 2. The left panel of Figure 1 clearly indicates the lack of fit of this model. More specifically, the model seems to overestimate the true probability of death during the first year after ART initiation. We then considered a model with a piecewise linear effect of time, with a change in slope 12 months after ART initiation. The residual process for this model (right panel of Figure 1) was close to zero at all time points, and remained within the 95% goodness-of-fit band (p-value $= 0.436$). This was the model used in our proposed NPMPLE for this analysis.

The estimates of the transition probabilities of death while in HIV care and disengagement from care are presented in Figure 2. Compared with the proposed NPMPLE method, the naïve analysis, which ignores the information from double-sampling, significantly underestimated mortality while in HIV care (left panel of Figure 2), but overestimated disengagement from HIV care (right panel of Figure 2). The CC analysis underestimated the probabilities of both death and disengagement from care, compared with the proposed estimator. Note that the findings from the CC analysis were similar to the findings from the simulation study. However, the results from the naïve analysis were not similar to the results from the simulation study. In the HIV data example, $\Pr(C^\star = 2|C = 2) = 1$, that is, the imperfect state classification was always correct when the true state was "disengagement." In contrast, in the simulation study, we considered the more general case of $\Pr(C^\star = 2|C = 2) < 1$. The computing time for estimating the transition probabilities over the whole study period for our data set of 58,876

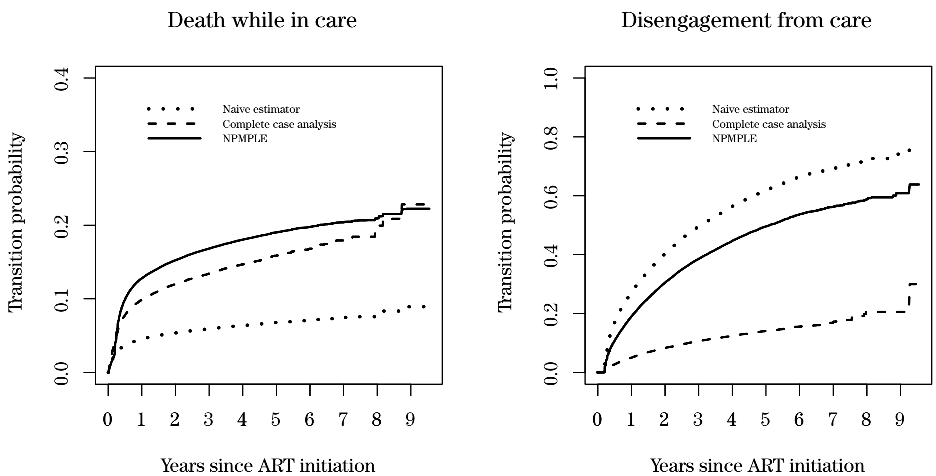

Figure 2. Transition probability estimates in the HIV study based on the naïve approach, the complete case analysis (CC), and the proposed NPMPLE method.

observations was only 15 seconds, using a computer with an i7 processor.

6. Concluding Remarks

In this paper, we proposed a computationally efficient nonparametric estimation approach for the transition probability matrix of a nonhomogeneous Markov process with a missing absorbing state, allowing for both right censoring and left truncation. Additionally, we derived a covariance function estimator based on the estimated influence functions, and proposed a methodology for constructing simultaneous confidence bands. The validity of our methodology was studied both theoretically and numerically. Even though our approach uses the parametric model $\pi_j(\mathbf{Z}, \boldsymbol{\beta})$ to estimate the probabilities of each absorbing state for the missing cases, it seems, based on our simulation studies, to be robust against some degree of misspecification of this model. Moreover, we proposed a formal goodness-of-fit approach for evaluating the "working" model for $\pi_j(\mathbf{Z}, \boldsymbol{\beta})$.

Alternative approaches for the competing-risks model, which is a special case of an absorbing Markov process with a single transient state, are the estimators proposed by Effraimidis and Dahl (2014) and Gouskova, Lin and Fine (2017). These methods nonparametrically estimate the probabilities of the absorbing states $\pi_j(t)$ as functions of time. However, unlike our approach, these methods do not incorporate auxiliary variables and, thus, impose stronger MAR assumptions. Therefore, these estimators can be biased when the probability of missingness depends on variables other than time, as was illustrated in the simulation study.

While the proposed method is computationally efficient and has superior statistical properties compared with existing methods, it is not clear if it is fully statistically efficient. Therefore, it would be interesting to study the efficiency of our pseudolikelihood estimator theoretically. Here, we can consider either the full class of nonparametric estimators of the transition probability matrix of a Markov process that utilize a parametric model for the probabilities of the absorbing states, or the subclass of the union of pseudolikelihood estimators considered in this article and potential augmented inverse probability estimators. The latter approach is useful when deriving the efficient influence function is challenging. The study of efficiency within a restricted class of estimators has been considered by Kulich and Lin (2004) and Breslow et al. (2009) for the class of augmented inverse-probability weighting estimators for the Cox proportional-hazards model under case-cohort study designs. Such efficiency considerations in the framework of the proposed method are technically challenging, but constitute an interesting topic for future research.

Supplementary Material

The Supplementary Material contains the proofs of the theorems presented in Section 3, as well as additional simulation results.

Acknowledgments

We are grateful to the co-editor, associate editor and two anonymous referees for their insightful suggestions. The research reported in this publication was supported by the National Institute Of Allergy And Infectious Diseases (NIAID), Eunice Kennedy Shriver National Institute Of Child Health & Human Development (NICHD), National Institute On Drug Abuse (NIDA), National Cancer Institute (NCI), and the National Institute of Mental Health (NIMH), in accordance with the regulatory requirements of the National Institutes of Health under Award Number U01AI069911 East Africa IeDEA Consortium. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This research has also been supported by the National Institutes of Health grant R01-AI102710 "Statistical Designs and Methods for Double-Sampling for HIV/AIDS" and by the President's Emergency Plan for AIDS Relief (PEPFAR) through USAID under the terms of Cooperative Agreement No. AID-623-A-12-0001, and was made possible by the joint support of the United States Agency for International Development (USAID). The contents of this article are the sole responsibility of AMPATH and do not necessarily reflect the views of USAID or the United States government.

References

Aalen, O. O. and Johansen, S. (1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scand. J. Stat.* **5**, 141–150.

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

Breslow, N. E., Lumley, T., Ballantyne, C. M., Chambless, L. E. and Kulich M. (2009). Improved horvitz-thompson estimation of model parameters from two-phase stratified samples: Applications in epidemiology. *Stat. Biosci.* **1**, 32–49.

Cook, T. D. and Kosorok, M. R. (2004). Analysis of time-to-event data with incomplete event adjudication. *J. Amer. Statist. Assoc.* **99**, 1140–1152.

Effraimidis, G. and Dahl, C. M. (2014). Nonparametric estimation of cumulative incidence functions for competing risks data with missing cause of failure. *Statist. Probab. Lett.* **89**, 1–7.

Gouskova, N. A., Lin, F. C. and Fine, J. P. (2017). Nonparametric analysis of competing risks data with event category missing at random. *Biometrics* **73**, 104–113.

Hall, W. J. and Wellner, J. A. (1980). Confidence bands for a survival curve from censored data. *Biometrika* **67**, 133–143.

Kolek, M. J., Graves, A. J., Xu, M., Bian, A., Teixeira, P. L., Shoemaker, M. B., Parvez, B., Xu, H., Heckbert, S. R., Ellinor, P. T. and Benjamin, E. J. (2016). Evaluation of a prediction model for the development of atrial fibrillation in a repository of electronic medical records. *JAMA Cardiol.* **1**, 1007–1013.

Kulick, M. and Lin, D. Y. (2004). Improving the efficiency of relative-risk estimation in case-cohort studies. *J. Amer. Statist. Assoc.* **99**, 832–844.

Ladha, K. S. and Eikermann, M. (2015). Codifying healthcare–big data and the issue of misclassification. *BMC Anesthesiol.* **15**, 179.

Lee, M., Dignam, J. J. and Han, J. (2014). Multiple imputation methods for nonparametric inference on cumulative incidence with missing cause of failure. *Stat. Med.* **33**, 4605–4626.

Lu, K. and Tsiatis, A. A. (2001). Multiple imputation methods for estimating regression coefficients in the competing risks model with missing cause of failure. *Biometrics* **57**, 1191–1197.

Nair, V. N. (1984). Confidence bands for survival functions with censored data: a comparative study. *Technometrics* **26**, 265–275.

Pan, Z. and Lin, D. Y. (2005). Goodness-of-fit methods for generalized linear mixed models. *Biometrics* **61**,1000–1009.

Rahardja, D. and Yang, Y. (2015). Maximum likelihood estimation of a binomial proportion using onesample misclassified binary data. *Stat. Neerl.* **69**, 272–280.

Rahardja, D. and Young, D. M. (2011). Confidence intervals for the risk ratio using double-sampling with misclassified binomial data. *J. Data Sci.* **9**, 529–548.

Spiekerman, C. F. and Lin, D. Y. (1998). Marginal regression models for multivariate failure time data. *J. Amer. Statist. Assoc.* **93**, 1164–1175.

Tenenbein, A. (1970). A double-sampling scheme for estimating from binomial data with misclassifications. *J. Amer. Statist. Assoc.* **65**, 1350–1361.

Wang, N. and Robins, J. M. (1998). Large-sample theory for parametric multiple imputation procedures. *Biometrika* **85**, 935–948.

Yin, G. and Cai, J. (2004). Additive hazards model with multivariate failure time data. *Biometrika* **91**, 801–818.

410 West 10th Street, Suite 3000, Indianapolis, Indiana 46202, USA.

E-mail: gbakogia@iu.edu

410 West 10th Street, Suite 3000, Indianapolis, Indiana 46202, USA.

E-mail: yz73@iu.edu

410 West 10th Street, Suite 3000, Indianapolis, Indiana 46202, USA.

E-mail: cyiannou@iupui.edu

(Received April 2017; accepted February 2018)