0.4pt=0pt

NONPARAMETRIC INFERENCE FOR MARKOV PROCESSES WITH MISSING ABSORBING STATE

Indiana University

Supplementary Material

We provide outlines of the proofs of the theorems stated in Section 3 using modern empirical process theory. Let $\mathcal{D}$ denote the sample space and D denote an arbitrary sample point with $D \in \mathcal{D}$. Also, let $\mathbb{P}_n f = n^{-1} \sum_{i=1}^{n} f(D_i)$, for a measurable function $f : \mathcal{D} \mapsto \mathbb{R}$, and $Pf = \int_{\mathcal{D}} f dP$ denote the expectation of f under P, the probability measure of the data on the measurable space $(\mathcal{D}, \mathcal{A})$, with $\mathcal{A}$ being a σ-algebra on $\mathcal{D}$. In what follows, C denotes a universal constant that may vary from place to place. We start with two technical lemmas that will be used later in the proofs.

Lemma 1. *Let $h(t)$ be a fixed uniformly bounded function on $[0, \tau]$ and $\phi(t)$ a non-decreasing random function on $[0, \tau]$ that belongs to the P-Donsker class Φ. Then the class of functions*

$$\mathcal{F}_1 = \left\{ \int_0^t h(u) d\phi(u) : t \in [0, \tau] \right\}$$

is P-Donsker.

Proof. For any probability measure Q and any $t_1, t_2 \in [0, \tau]$ we can easily demonstrate that

$$\left\| \int_0^{t_1} h(u) d\phi(u) - \int_0^{t_2} h(u) d\phi(u) \right\|_{Q,2} \leq C \|\phi(t_2) - \phi(t_1)\|_{Q,2},$$

where $\|h\|_{Q,2} = (\int h^2 dQ)^{1/2}$. Now, for any $t \in [0, \tau]$ there exists a $t_i \in [0, \tau]$, $i = 1, \ldots, N(\epsilon, \Phi, L_2(Q))$, such that $\|\phi(t)-\phi(t_i)\|_{Q,2} < \epsilon$. Consequently, for any member of $\mathcal{F}_1$ there exists a $\int_0^{t_i} h(u)d\phi(u)$, for $i = 1, \ldots, N(\epsilon, \Phi, L_2(Q))$, such that

$$\left\| \int_0^t h(u)d\phi(u) - \int_0^{t_i} h(u)d\phi(u) \right\|_{Q,2} \leq C\epsilon,$$

and thus we can cover the whole $\mathcal{F}_1$ with $N(\epsilon, \Phi, L_2(Q))$ $L_2(Q)$ ϵ'-balls centered at $\int_0^{t_i} h(u)d\phi(u)$. By the uniform entropy bound (2.5.1) in van der Vaart and Wellner (1996) and the P-Donsker theorem, Theorem 2.5.2 of van der Vaart and Wellner (1996), it follows that $\mathcal{F}_1$ is P-Donsker. □

Lemma 2. *Let $g(t)$ be a data-dependent and uniformly bounded function and $f(t)$ a continuous fixed function of bounded variation on $[0, \tau]$. Then, the class*

$$\mathcal{F}_2 = \left\{ \int_0^t g(s)df(s) : t \in [0, \tau] \right\},$$

is P-Donsker.

Proof. The assumption on $f(t)$ implies that $f(t) = f_1(t) - f_2(t)$, where f_1 and f_2 are non-decreasing continuous functions. The classes of fixed functions $\mathcal{F}_{3,l} = \{f_l(t) : t \in [0, \tau]\}$, $l = 1, 2$, are P-Donsker because they are totally bounded by the $|\cdot|$ metric. The total boundedness of $\mathcal{F}_{3,1}$ and $\mathcal{F}_{3,2}$ is a consequence of the fact that their members are continuous functions defined on the compact set $[0, \tau]$, and therefore $\mathcal{F}_{3,1}$ and $\mathcal{F}_{3,2}$ are compact. Now, for any $t_1, t_2 \in [0, \tau]$ and any finitely discrete probability measure Q it follows that

$$\left\| \int_0^{t_1} g(s)df_l(s) - \int_0^{t_2} g(s)df_l(s) \right\|_{Q,2} \leq C|f_l(t_1) - f_l(t_2)|, \qquad l = 1, 2.$$

Similar arguments to those used in the proof of Lemma 1 can be used to show the Donsker property of $\mathcal{F}_{2,l} = \left\{ \int_0^t g(s)df_l(s) : t \in [0, \tau] \right\}$, $l = 1, 2$. Finally, the Donsker property of $\mathcal{F}_2$ is a consequence of Corollary 9.31 of Kosorok (2008), since $\mathcal{F}_2$ is formed by differences of functions that belong to the Donsker classes $\mathcal{F}_{2,1}$ and $\mathcal{F}_{2,2}$. □

S1 Proof of Theorem 1

Consider the underlying stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t : t \geq 0\}, P_\Omega)$, with

$$\mathcal{F}_t = \sigma\Big\langle\{\mathbf{N}(u), \mathbf{N}^\star(u), \mathbf{Y}(u), R, \mathbf{Z} : 0 \leq u \leq t\}\Big\rangle,$$

and define $\mathcal{G} = \sigma\Big\langle\{\mathbf{N}^\star(u), \mathbf{Y}(u), R, \mathbf{Z} : 0 \leq u \leq \tau\} \cup \{\mathbf{N}(u) : 0 \leq u \leq \tau, R = 1\}\Big\rangle$, the σ-algebra generated by the observable random variables on $(0, \tau]$. In the remaining of this proof, we present the arguments for the estimators concerning the absorbing states which may be missing. The corresponding arguments for the estimators regarding the transient states that are completely observed, follow the same arguments as those given in Andersen et al. (1993). The expected cause-specific counting process for an absorbing state $j \in \mathcal{T}$ is

$$E[N_{ihj}(t)] = E[\delta_{ij} N_{ih\cdot}(t)] = E\{E[\delta_{ij} N_{ih\cdot}(t)|\mathcal{G}]\} = E[N_{ih\cdot}(t) E(\delta_{ij}|\mathcal{G})],$$

which immediately results in

$$E[\tilde{N}_{ihj}(t)] \equiv E\left\{[R_i\delta_{ij} + (1 - R_i)E(\delta_{ij}|\mathcal{G})]N_{ih\cdot}(t)\right\} = E[N_{ihj}(t)], \quad \text{(S1.1)}$$

for all $h \notin \mathcal{T}$, $j \in \mathcal{T}$ and $t \in [0, \tau]$. It is also noted that after observing D_i we have that $E(\delta_{ij}|\mathcal{G}) = E(\delta_{ij}|D_i) = \pi_j(\mathbf{Z}_i, \boldsymbol{\beta}_0)$ with $\boldsymbol{\beta}_0$ denoting the true regression parameter for the probability of the absorbing state $j \in \mathcal{T}$, π_j. Hence the corresponding absorbing state-specific counting process can be expressed as $\tilde{N}_{ihj}(t; \boldsymbol{\beta}_0) = [R_i\delta_{ij} + (1 - R_i)\pi_j(\mathbf{Z}_i, \boldsymbol{\beta}_0)]N_{ih\cdot}(t)$.

Define three classes of functions, $\mathcal{L}_{1,h,j} = \{N_{hj}(t) : t \in [0,\tau]\}$ for $h \notin \mathcal{T}$ and $j \in \mathcal{I}$, $\mathcal{L}_{2,h} = \{Y_h(t) : t \in [0,\tau]\}$, $h \notin \mathcal{T}$, and $\mathcal{L}_{3,j} = \{f_j(\boldsymbol{\beta}) = R\delta_j + (1-R)\pi_j(\mathbf{Z}_i, \boldsymbol{\beta}) : \boldsymbol{\beta} \in \mathcal{B}\}$, $j \in \mathcal{T}$. $\mathcal{L}_{1,h,j}$ is a class of monotone cadlag functions satisfying C2 and hence it is P-Donsker for any $h \notin \mathcal{T}$ and $j \in \mathcal{I}$, with $h \neq j$, by Lemma 4.1 of Kosorok (2008), because $P[N_{hj}(\tau)]^2 \leq C^2 < \infty$.

Note that $Y_h(t) = N_{\cdot h}(t-) - N_{h\cdot}(t-)$, with $N_{\cdot h}(t) = \sum_{j \neq h} N_{jh}(t)$ and $N_{h\cdot}(t) = \sum_{j \neq h} N_{hj}(t)$. The classes of functions $N_{\cdot h}(t-)$ and $N_{h\cdot}(t-)$ are P-Donsker because they can be expressed as the finite sum of P-Donsker classes (Kosorok, 2008). Thus, using the same argument as before, $\mathcal{L}_{2,h}$ is also P-Donsker.

Next, note that the fixed class $\mathcal{B}$ is trivially Donsker by condition C4. Conditions C4 and C6 imply that $\pi_j(\mathbf{Z}_i, \boldsymbol{\beta})$ is a Lipschitz continuous function of $\boldsymbol{\beta}$ on compacts and, therefore, the class $\mathcal{L}_{3,j}$ is P-Donsker for any $j \in \mathcal{T}$, by Corollary 9.31 in Kosorok (2008). Now, the class of functions $\mathcal{L}_{3,j}\mathcal{L}_{1,h} = \{f_j(\boldsymbol{\beta}) N_{h\cdot}(t) = \tilde{N}_{hj}(t; \boldsymbol{\beta}) : t \in [0,\tau], \boldsymbol{\beta} \in \mathcal{B}\}$, for all $h \notin \mathcal{T}$ and $j \in \mathcal{T}$, is P-Donsker because it is formed by products of bounded functions that belong to Donsker classes.

Since a P-Donsker class is also Glivenko-Cantelli, we have that

$$\sup_{t \in [0,\tau]} |\mathbb{P}_n Y_h(t) - P Y_h(t)| \overset{as*}{\to} 0 \quad h \notin \mathcal{T},$$

and

$$\sup_{t\in[0,\tau]} |\mathbb{P}_n \tilde{N}_{hj}(t;\boldsymbol{\beta}_0) - PN_{hj}(t)| \stackrel{as*}{\to} 0, \quad h \notin \mathcal{T}, \quad j \in \mathcal{T}, \tag{S1.2}$$

by (S1.1). Then it is straightforward to show

$$\begin{aligned}\sup_{t\in[0,\tau]} |\mathbb{P}_n \tilde{N}_{hj}(t;\hat{\boldsymbol{\beta}}_n) - PN_{hj}(t)| &\equiv \|\mathbb{P}_n \tilde{N}_{hj}(t;\hat{\boldsymbol{\beta}}_n) - PN_{hj}(t)\|_\infty \\ &\leq C\left(\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\| + \|\mathbb{P}_n \tilde{N}_{hj}(t;\boldsymbol{\beta}_0) - PN_{hj}(t)\|_\infty\right) \stackrel{as*}{\to} 0.\end{aligned}$$

for any $h \notin \mathcal{T}, j \in \mathcal{T}$ by (S1.2) and C5. Next, performing the same calculation as that in Stute (1995), we have

$$\hat{A}_{n,hj}(t) - A_{0,hj}(t) = I_1(t) + I_2(t) + I_3(t),$$

where,

$$I_1(t) = \int_0^t \frac{PY_h(u) - \mathbb{P}_n Y_h(u)}{[PY_h(u)]^2} d\mathbb{P}_n \tilde{N}_{hj}(u;\hat{\boldsymbol{\beta}}_n), \;\; I_2(t) = \int_0^t \frac{d[\mathbb{P}_n \tilde{N}_{hj}(u;\hat{\boldsymbol{\beta}}_n) - PN_{hj}(u)]}{PY_h(u)}$$

and

$$I_3(t) = \int_0^t \frac{[PY_h(u) - \mathbb{P}_n Y_h(u)]^2}{[PY_h(u)]^2 \mathbb{P}_n Y_h(u)} d\mathbb{P}_n \tilde{N}_{hj}(u;\hat{\boldsymbol{\beta}}_n).$$

We shall show that each of the above terms converges almost surely to zero under the $\|\cdot\|_\infty$ metric. First,

$$\|I_1(t)\|_\infty \leq \frac{\|PY_h(t) - \mathbb{P}_n Y_h(t)\|_\infty}{\inf_{t\in[0,\tau]}[PY_h(t)]^2} \mathbb{P}_n \tilde{N}_{hj}(\tau;\hat{\beta}_n) \stackrel{a.s.}{\to} 0,$$

due to the fact that $\mathcal{L}_{2,h}$ is Glivenko-Cantelli.

Next, note that $\mathbb{P}_n\tilde{N}_{hj}(u;\hat{\boldsymbol{\beta}}_n) - PN_{hj}(u)$ is a right-continuous process of bounded variation on any interval $A \subset [0,\tau]$. Now, we have

$$\int_0^t \frac{d[\mathbb{P}_n\tilde{N}_{hj}(u;\hat{\boldsymbol{\beta}}_n) - PN_{hj}(u)]}{PY_h(u)} = \int_0^t \frac{d\mathbb{P}_n\tilde{N}_{hj}(u;\hat{\boldsymbol{\beta}}_n)}{PY_h(u)} - \int_0^t \frac{dPN_{hj}(u)}{PY_h(u)}$$
$$= \mathbb{P}_n\left[f_j(\hat{\boldsymbol{\beta}}_n) - f_j(\boldsymbol{\beta}_0)\right]\int_0^t \frac{dN_{h\cdot}(u)}{PY_h(u)} + (\mathbb{P}_n - P)\left[\int_0^t \frac{d\tilde{N}_{hj}(u;\boldsymbol{\beta}_0)}{PY_h(u)}\right]$$

and hence

$$\|I_2(t)\|_\infty \leq C\left(\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\| + \left\|(\mathbb{P}_n - P)\left[\int_0^t \frac{d\tilde{N}_{hj}(u;\boldsymbol{\beta}_0)}{PY_h(u)}\right]\right\|_\infty\right) \overset{as}{\to} 0$$

by C2, C4–C6 and Lemma 1.

Finally, since $\mathcal{L}_{2,h}$ is Donsker and hence Glivenko-Cantelli, for any small enough $\epsilon > 0$, $\mathbb{P}_nY_h(t) \geq PY_h(t) - \epsilon \geq \frac{1}{2}PY_h(t)$, it follows that for any $t \in [0,\tau]$

$$\frac{[PY_h(t) - \mathbb{P}_nY_h(t)]^2}{[PY_h(t)]^2\mathbb{P}_nY_h(t)} \leq \frac{2|PY_h(t) - \mathbb{P}_nY_h(t)|^2}{[PY_h(t)]^3} \leq \frac{2\|PY_h(t) - \mathbb{P}_nY_h(t)\|_\infty^2}{\inf_{t\in[0,\tau]}[PY_h(t)]^3}, \quad t \in [0,\tau].$$

Thus for any $h \notin T$ and $j \in \mathcal{T}$

$$\|I_3(t)\|_\infty \leq \frac{2\|P_DY(t) - \mathbb{P}_nY(t)\|_\infty^2}{\inf_{t\in[0,\tau]}[P_DY_h(t)]^3}\mathbb{P}_n\tilde{N}_j(\tau;\hat{\beta}_n) \overset{a.s.}{\to} 0,$$

which leads to $\|\hat{A}_{n,hj}(t) - A_{0,hj}(t)\|_\infty \overset{as*}{\to} 0$, for all $h \notin \mathcal{T}$ and $j \in \mathcal{T}$. Thus, we conclude that $\hat{\mathbf{A}}_n(t) \overset{as*}{\to} \mathbf{A}_0(t)$, uniformly on $[0,\tau]$, and therefore it follows that

$$\prod_{(s,t]}\left[\mathbf{I} + d\hat{\mathbf{A}}_n(u)\right] \overset{as*}{\to} \prod_{(s,t]}\left[\mathbf{I} + d\mathbf{A}_0(u)\right],$$

uniformly on $s, t \in [0, \tau]$, with $s < t$, as a consequence of a continuity result from the Duhamel equation (Andersen et al., 1993). Thus the proof of Theorem 1 is complete.

S2 Proof of Theorem 2

We start the proof of Theorem 2 by studying the asymptotic distribution of $\tilde{N}_{hj}(t; \hat{\boldsymbol{\beta}}_n)$, for $h \notin \mathcal{T}$ and $j \in \mathcal{T}$. First, we have the following decomposition

$$\begin{aligned}\sqrt{n}[\mathbb{P}_n \tilde{N}_{hj}(t; \hat{\boldsymbol{\beta}}_n) - P\tilde{N}_{hj}(t; \boldsymbol{\beta}_0)] &= \mathbb{G}_n[\tilde{N}_{hj}(t; \hat{\boldsymbol{\beta}}_n) - \tilde{N}_{hj}(t; \boldsymbol{\beta}_0)] + \mathbb{G}_n \tilde{N}_{hj}(t; \boldsymbol{\beta}_0) \\ &\quad + \sqrt{n}[P\tilde{N}_{hj}(t; \hat{\boldsymbol{\beta}}_n) - P\tilde{N}_{hj}(t; \boldsymbol{\beta}_0)], \qquad \text{(S2.1)}\end{aligned}$$

where $\mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n - P)f$. For the first part, we have

$$\begin{aligned}\sup_{t \in [0,\tau]} P[\tilde{N}_{hj}(t; \hat{\boldsymbol{\beta}}_n) - \tilde{N}_{hj}(t; \boldsymbol{\beta}_0)]^2 &\leq \sup_{t \in [0,\tau]} P[\pi_j(D; \hat{\boldsymbol{\beta}}_n) - \pi_j(D; \boldsymbol{\beta}_0)]^2 \\ &\leq C\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\|^2.\end{aligned}$$

Hence, by C5 it follows that $\sup_{t \in [0,\tau]} P[\tilde{N}_{hj}(t; \hat{\boldsymbol{\beta}}_n) - \tilde{N}_{hj}(t; \boldsymbol{\beta}_0)]^2 \overset{as*}{\to} 0$ for all $h \notin \mathcal{T}$ and $j \in \mathcal{T}$. Additionally, we have that $\Pr(\hat{\boldsymbol{\beta}}_n \in \mathcal{B}) \to 1$ and that $\mathcal{L}_{3,j}\mathcal{L}_{1,h,j} = \{\tilde{N}_{hj}(t; \boldsymbol{\beta}) : t \in [0, \tau], \boldsymbol{\beta} \in \mathcal{B}\}$ is P-Donsker for all $h \notin \mathcal{T}$ and $j \in \mathcal{T}$ as shown in the proof of Theorem 1. Consequently, it follows by Theorem 2.1 in van der Vaart and Wellner (2007) that

$$\left\| \mathbb{G}_n[\tilde{N}_{hj}(t; \hat{\boldsymbol{\beta}}_n) - \tilde{N}_{hj}(t; \boldsymbol{\beta}_0)] \right\|_\infty = o_p(1).$$

The second part in (S2.1) is asymptotically equivalent to a tight zero mean Gaussian process $\mathbb{G}_{hj}$ as the class of functions $\{\tilde{N}_{hj}(t;\boldsymbol{\beta}_0) : t \in [0,\tau]\}$ is a subclass of the P-Donsker class $\{\tilde{N}_{hj}(t;\boldsymbol{\beta}) : t \in [0,\tau], \boldsymbol{\beta} \in \mathcal{B}\}$ and it is therefore a P-Donsker for all $h \notin \mathcal{T}$ and $j \in \mathcal{T}$ as well. Next, applying a Taylor expansion at $\boldsymbol{\beta}_0$ to the third part of (S2.1) along with conditions C4–C6 leads to

$$\sqrt{n}P[\tilde{N}_{hj}(t;\hat{\boldsymbol{\beta}}_n) - \tilde{N}_{hj}(t;\boldsymbol{\beta}_0)] = \sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)^T\tilde{\mathbf{R}}_{hj}(t) + o_p(1),$$

where $\tilde{\mathbf{R}}_{hj}(t) = P[(1-R)N_{h\cdot}(t)\dot{\pi}_j(\mathbf{Z},\boldsymbol{\beta}_0)]$. Using condition C5 we have that

$$\begin{aligned}\sqrt{n}[\mathbb{P}_n\tilde{N}_{hj}(t;\hat{\boldsymbol{\beta}}_n) - P\tilde{N}_{hj}(t;\boldsymbol{\beta}_0)] &= \sqrt{n}\mathbb{P}_n[\tilde{N}_{hj}(t;\boldsymbol{\beta}_0) - P\tilde{N}_{hj}(t;\boldsymbol{\beta}_0) + \boldsymbol{\omega}^T\tilde{\mathbf{R}}_{hj}(t)] \\ &\quad + o_p(1).\end{aligned}$$

Since $\{\tilde{N}_{hj}(t;\boldsymbol{\beta}_0) : t \in [0,\tau]\}$ is P-Donsker, $\boldsymbol{\omega}$ have bounded second moment and zero expectation, and $P\tilde{N}_{hj}(t;\boldsymbol{\beta}_0)$ is non-random, we conclude that

$$\sqrt{n}[\mathbb{P}_n\tilde{N}_{hj}(\cdot;\hat{\boldsymbol{\beta}}_n) - P\tilde{N}_{hj}(\cdot;\boldsymbol{\beta}_0)] \rightsquigarrow \mathbb{G}^*_{hj}, \quad h \notin \mathcal{T}, \;\; j \in \mathcal{T},$$

with $\mathbb{G}^*_{hj}$ being a tight zero mean Gaussian process. Note that the map $(PN_{hj}, PY_h) \mapsto \int_0^t \frac{1}{PY_h} dPN_{hj}$ is Hadamaard differentiable with derivative at (a,b) over the whole domain

$$\mathcal{H} = \left\{(PN_{hj}, B) : \int_{[0,\tau]} |dPN_{hj}(t)| \leq C, C \in (0,\infty), \inf_{t\in[0,\tau]} |PY_h(t)| \geq \epsilon, \epsilon > 0\right\}$$

given by $\int_0^t \frac{da}{PY_h} - \int_0^t \frac{bdPN_{hj}}{(PY_{hj})^2}$, (Kosorok, 2008). By C1 and C2, we have that $(PN_{hj}, PY_h) \in \mathcal{H}$ and $\Pr((\mathbb{P}_n\tilde{N}_{hj}, \mathbb{P}_nY_h) \in \mathcal{H}) \to 1$ as $n \to \infty$, for $h \notin \mathcal{T}$

and $j \in \mathcal{T}$. Hence simple algebra followed by the Functional Delta Method (van der Vaart, 2000) leads to

$$\sqrt{n}[\hat{A}_{n,hj}(\cdot)-A_{0,hj}(\cdot)] = \sqrt{n}\mathbb{P}_n[\psi_{hj,1}(\cdot)-\psi_{hj,2}(\cdot)+\boldsymbol{\omega}^T\mathbf{R}_{hj}(\cdot)]+o_p(1), \quad h \notin \mathcal{T}, j \in \mathcal{T},$$

where

$$\psi_{ihj,1}(\cdot) = \int_0^\cdot \frac{d\tilde{N}_{ihj}(u;\boldsymbol{\beta}_0)}{PY_h(u)}, \quad \psi_{ihj,2}(\cdot) = \int_0^\cdot \frac{Y_{ih}(u)}{PY_h(u)} dA_{0,hj}(u),$$

$$\text{and} \quad \mathbf{R}_{hj}(\cdot) = P\left[(1-R)\dot{\pi}_j(\mathbf{Z},\boldsymbol{\beta}_0)\int_0^\cdot \frac{dN_{h\cdot}(u)}{PY_h(u)}\right].$$

Under the regularity conditions and Lemma 1, it can be shown that $\psi_{hj,1}(t)$, $t \in [0,\tau]$, forms a Donsker class. The same is true for $\psi_{hj,2}(t)$, $t \in [0,\tau]$, as a consequence of the regularity conditions and Lemma 2. Therefore, $\sqrt{n}[\hat{A}_{n,hj}(\cdot) - A_{0,hj}(\cdot)]$ converges weakly to a mean-zero Gaussian process for all $h \notin \mathcal{T}$ and $j \in \mathcal{T}$. Using similar calculations, it is straightforward to show for the transient states, which are completely observed, that

$$\sqrt{n}[\hat{A}_{n,hj}(\cdot) - A_{0,hj}(\cdot)] = \sqrt{n}\mathbb{P}_n[\psi_{hj1}(\cdot) - \psi_{hj2}(\cdot)] + o_p(1), \qquad h, j \notin \mathcal{T},$$

for $h \neq j$. Thus it follows that

$$\sqrt{n}[\hat{\mathbf{A}}_n(\cdot) - \mathbf{A}_0(\cdot)] = \sqrt{n}\mathbb{P}_n[\boldsymbol{\psi}_1(\cdot) - \boldsymbol{\psi}_2(\cdot) + \boldsymbol{\omega}^T I_{q+1}\mathbf{R}(\cdot)] + \boldsymbol{\epsilon},$$

where

$$\boldsymbol{\psi}_l = \begin{pmatrix} -\sum_{j\neq 0}\psi_{0jl} & \psi_{01l} & \cdots & \psi_{0ql} \\ \psi_{10l} & -\sum_{j\neq 1}\psi_{1jl} & \cdots & \psi_{1ql} \\ \vdots & \vdots & \vdots & \vdots \\ \psi_{(q-k)0l} & \psi_{(q-k)1l} & \cdots & \psi_{(q-k)ql} \\ \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times 1} & \cdots & \mathbf{0}_{k\times 1} \end{pmatrix}, \quad l = 1, 2,$$

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

and $\boldsymbol{\epsilon}$ is a $(q+1)\times(q+1)$ matrix that includes $o_p(1)$ terms. In the matrix $\mathbf{R}$, $\mathbf{R}_{\mathcal{T}^c}$ is a $(q-k+1)\times(q-k+1)$ diagonal matrix with diagonal elements $-\sum_{j\in\mathcal{T}}\mathbf{R}_{hj}$, $h\notin\mathcal{T}$, and $\mathbf{R}_{\mathcal{T}}$ a $(q-k+1)\times k$ matrix with elements $\mathbf{R}_{hj}$, $h\notin\mathcal{T}$ and $j\in\mathcal{T}$.

Next, by the regularity conditions and the Hadamard differentiability of the product integral (Andersen et al., 1993)

$$\mathbf{A}_0 \mapsto \prod\left[\mathbf{I} + d\mathbf{A}_0\right],$$

on the whole domain space of the matrix-value functions with elements that are continuous functions of bounded variation on $[0,\tau]$, and the functional delta method (van der Vaart, 2000) it follows that

$$\begin{aligned}\sqrt{n}[\hat{\mathbf{P}}_n(s,\cdot)-\mathbf{P}_0(s,\cdot)] &= \sqrt{n}\mathbb{P}_n\int_s^{\cdot}\prod_{[s,u)}[\mathbf{I}+d\mathbf{A}_0(v)]\,\boldsymbol{\psi}(du)\prod_{(u,\cdot]}[\mathbf{I}+d\mathbf{A}_0(v)]+\boldsymbol{\epsilon}\\ &\equiv \sqrt{n}\mathbb{P}_n\boldsymbol{\gamma}(s,\cdot)+\boldsymbol{\epsilon},\qquad s\in[0,\tau)\end{aligned}$$

where $\boldsymbol{\psi}(t)=\boldsymbol{\psi}_1(t)-\boldsymbol{\psi}_2(t)+\boldsymbol{\omega}^T I_{q+1}\mathbf{R}(t)$. The components of the matrix-valued influence functions $\boldsymbol{\gamma}_i(s,t)$ are

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

where

$$\psi_{ilm}(t)=\begin{cases}\psi_{ilm,1}(t)-\psi_{ilm,2}(t)+\boldsymbol{\omega}_i^T\mathbf{R}_{lm}(t) & \text{if } m\in\mathcal{T}\\ \\ \psi_{ilm,1}(t)-\psi_{ilm,2}(t) & \text{if } m\notin\mathcal{T}\end{cases}$$

for $l\neq m$, and $\psi_{ill}(t)=-\sum_{h\neq l}\psi_{ilh}(t)$, $l\notin\mathcal{T}$, otherwise.

Without loss of generality, set $s=0$. To show the Donsker property of the class $\Gamma_{hj}=\{\gamma_{hj}(0,t):t\in[0,\tau]\}$, for all $h\notin\mathcal{T}$ and $j\in\mathcal{I}$, with $h\neq j$, consider the classes

$$\Gamma_{hj,1}=\left\{\int_0^t P_{0,hl}(s,u-)P_{0,mj}(u,t)d\psi_{lm,1}(u):t\in[0,\tau]\right\},$$

$$\Gamma_{hj,2}=\left\{\int_0^t P_{0,hl}(s,u-)P_{0,mj}(u,t)d\psi_{lm,2}(u):t\in[0,\tau]\right\},$$

and

$$\Gamma_3=\left\{\boldsymbol{\omega}^T\int_0^t P_{0,hl}(s,u-)P_{0,mj}(u,t)d\mathbf{R}_{lm}(u):t\in[0,\tau]\right\}.$$

The class $\Gamma_{hj,1}$ is P-Donsker as a result of C2, C4, C6 and Lemma 1. The Donsker property of $\Gamma_{hj,2}$ follows from C2, C3 and Lemma 2. The class $\Gamma_{hj,3}$ is P-Donsker because it is formed by products of the random quantities $\boldsymbol{\omega}$ which have bounded second moments by C5, with a fixed and uniformly bounded function. Now, the classes Γ_{hj} for $h \notin \mathcal{T}$ and $j \in \mathcal{I}$, with $h \neq j$ are P-Donsker because they are formed by finite sums of functions that belong to P-Donsker classes.

Next, consider the processes $W_{n,hj}(0,t) = \sqrt{n}\mathbb{P}_n \gamma_{hj}(0,t)$, $\tilde{W}_{n,hj}(0,t) = \sqrt{n}\mathbb{P}_n \gamma_{hj}(0,t)\xi$ and $\hat{W}_{n,hj}(0,t) = \sqrt{n}\mathbb{P}_n \hat{\gamma}_{hj}(0,t)\xi$, where ξ_i are random draws from $N(0,1)$,

$$\hat{\gamma}_{ihj}(0,t) = \sum_{l \notin \mathcal{T}} \sum_{m \in \mathcal{I}} \int_0^t \hat{P}_{n,hl}(0,u-)\hat{P}_{n,mj}(u,t)d\hat{\psi}_{ilm}(u),$$

with

$$\hat{\psi}_{ilm}(t) = \begin{cases} \hat{\psi}_{ilm,1}(t) - \hat{\psi}_{ilm,2}(t) + \hat{\boldsymbol{\omega}}_i^T \hat{\mathbf{R}}_{lm}(t) & \text{if } \ m \in \mathcal{T} \\ \\ \hat{\psi}_{ilm,1}(t) - \hat{\psi}_{ilm,2}(t) & \text{if } \ m \notin \mathcal{T} \end{cases}$$

and the remaining components of the estimated influence functions are

$$\hat{\psi}_{ilm,1}(t) = \int_0^t \frac{d\tilde{N}_{ilm}(u;\hat{\boldsymbol{\beta}}_n)}{\mathbb{P}_n Y_l(u)}, \quad \hat{\psi}_{ilm,2}(t) = \int_0^t \frac{Y_{il}(u)}{\mathbb{P}_n Y_l(u)} d\hat{A}_{n,lm}(u),$$

$$\text{and} \ \ \hat{\mathbf{R}}_{lm}(t) = \mathbb{P}_n \left[(1-R)\dot{\pi}_m(\mathbf{Z},\hat{\boldsymbol{\beta}}_n) \int_0^t \frac{dN_{l\cdot}(u)}{\mathbb{P}_n Y_l(u)} \right].$$

Given the Donsker property of the classes $\{\gamma_{hj}(0,t) : t \in [0,\tau]\}$ for all

$h \notin \mathcal{T}$ and $j \in \mathcal{I}$ and the conditional multiplier central limit theorem (van der Vaart and Wellner, 1996), $\tilde{W}_{n,hj}(0,t)$ converges weakly, conditionally on the observed data $\mathbf{D}$, to the same limiting process as that of $W_{n,hj}(0,t)$ (unconditionally). To complete the proof of the second part of theorem 2 it remains to show that $\|\hat{W}_{n,hj}(0,t) - \tilde{W}_{n,hj}(0,t)\|_\infty = o_p(1)$, for all $h \notin \mathcal{T}$ and $j \in \mathcal{I}$, unconditionally. After some algebra it can be shown that

$$\|\hat{W}_{n,hj}(0,t) - \tilde{W}_{n,hj}(0,t)\|_\infty \leq \sum_{l \notin \mathcal{T}} \sum_{m \in \mathcal{I}} (I_{n,lm1} + I_{n,lm2} + I_{n,lm3}),$$

for all $h \notin \mathcal{T}$ and $j \in \mathcal{T}$, where

$$\begin{aligned} I_{n,lm1} &= \left\| \sqrt{n}\mathbb{P}_n \xi \int_0^t [\hat{P}_{n,hl}(0,u-)\hat{P}_{n,mj}(u,t) - P_{0,hl}(0,u-)P_{0,mj}(u,t)] \right. \\ &\quad \left. \times d[\hat{\psi}_{lm}(u) - \psi_{lm}(u)] \right\|_\infty, \end{aligned}$$

$$I_{n,lm2} = \left\| \sqrt{n}\mathbb{P}_n \xi \int_0^t d[\hat{\psi}_{lm}(u) - \psi_{lm}(u)] \right\|_\infty,$$

and

$$\begin{aligned} I_{n,lm3} &= \left\| \int_0^t [\hat{P}_{n,hl}(0,u-)\hat{P}_{n,mj}(u,t) - P_{0,hl}(0,u-)P_{0,mj}(u,t)] \right. \\ &\quad \left. \times d[\sqrt{n}\mathbb{P}_n \psi_{lm}(u)] \right\|_\infty O_p(1). \end{aligned}$$

Due to the weak convergence result for the NPMPLE that was shown above and after some algebra, we have

$$\sqrt{n}[\hat{P}_{n,hl}(0,u-)\hat{P}_{n,mj}(u,t) - P_{0,hl}(0,u-)P_{0,mj}(u,t)] = O_p(1).$$

Thus, after some algebra and C2 and C3, we have

$$\begin{aligned}
I_{n,lm1} &\le \left\|\mathbb{P}_n \int_0^t d[\hat{\psi}_{ilm}(u) - \psi_{ilm}(u)]\right\|_\infty O_p(1) \\
&\le \left\{\left\|\mathbb{P}_n \int_0^t \left[\frac{f_m(\hat{\boldsymbol{\beta}}_n)}{\mathbb{P}_n Y_l(u)} - \frac{f_m(\boldsymbol{\beta}_0)}{PY_l(u)}\right] dN_{l\cdot}(u)\right\|_\infty \right. \\
&\quad + \|\hat{A}_{n,lm}(t) - A_{0,lm}(t)\|_\infty + \left\|\frac{1}{\mathbb{P}_n Y_l(t)} - \frac{1}{PY_l(t)}\right\|_\infty \\
&\quad + (\mathbb{P}_n\|\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}\| + \|\mathbb{P}_n \boldsymbol{\omega}\|) \sup_{t\in[0,\tau]} \left\|\int_0^t d[\hat{\mathbf{R}}_{lm}(u) - \mathbf{R}_{lm}(u)]\right\| \\
&\quad \left. + (\mathbb{P}_n\|\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}\|) \sup_{t\in[0,\tau]} \left\|\int_0^t d\mathbf{R}_{lm}(u)\right\| \right\} O_p(1). \qquad \text{(S2.2)}
\end{aligned}$$

After some algebra and by C2, C4 and C5, it can be shown that first term in the right side of (S2.2) is $o_p(1)$. The second and third terms are also $o_p(1)$ as a result of the uniform consistency of $\hat{A}_{n,lm}(t)$ and the Glivenko-Cantelli property of the class $\{Y_l(t) : t \in [0,\tau]\}$ which were shown in the proof of Theorem 1. For the last two terms, we have that $\mathbb{P}_n\|\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}\| = o_p(1)$ and $\mathbb{P}_n \boldsymbol{\omega} = o_{as}(1)$ by C5 and the strong law of large numbers. Next, C2, C4 and C5, the continuous mapping theorem and the strong law of large numbers, imply that

$$\sup_{t\in[0,\tau]} \left\|\int_0^t d[\hat{\mathbf{R}}_{lm}(u) - \mathbf{R}_{lm}(u)]\right\| = o_p(1), \qquad \text{(S2.3)}$$

and therefore the forth term in (S2.2) is $o_p(1)$. Finally, C2, C4 and C6 ensure that $\sup_{t\in[0,\tau]} \left\|\int_0^t d\mathbf{R}_{lm}(u)\right\| = O_p(1)$ and thus the fifth term is also $o_p(1)$ and, therefore, $I_{n,lm1} = o_p(1)$.

Using C2 and C3, (S2.3) and the fact that $\left\|\frac{1}{\mathbb{P}_n Y_l(t)} - \frac{1}{PY_l(t)}\right\|_\infty = o_{as}(1)$ and $\sup_{t\in[0,\tau]}\left\|\int_0^t d\mathbf{R}_{lm}(u)\right\| = O_p(1)$, it can be shown that

$$\begin{aligned}
I_{n,lm2} &\leq \left\|\sqrt{n}\mathbb{P}_n\xi\int_0^t d[\hat{\psi}_{ilm}(u) - \psi_{ilm}(u)]\right\|_\infty \\
&\leq \left\|\sqrt{n}\mathbb{P}_n\xi\int_0^t \left[\frac{f_m(\hat{\boldsymbol{\beta}}_n)}{\mathbb{P}_n Y_l(u)} - \frac{f_m(\boldsymbol{\beta}_0)}{PY_l(u)}\right] dN_{l\cdot}(u)\right\|_\infty \\
&\quad +\|\sqrt{n}[\hat{A}_{n,lm}(t) - A_{0,lm}(t)]\|_\infty|\mathbb{P}_n\xi|O_p(1) \\
&\quad +\|A_{0,lm}(t)\|_\infty|\sqrt{n}\mathbb{P}_n\xi|o_{as}(1) \\
&\quad +\|\sqrt{n}\mathbb{P}_n(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega})\xi\|o_p(1) + \|\sqrt{n}\mathbb{P}_n\boldsymbol{\omega}\xi\|o_p(1) \\
&\quad +\|\sqrt{n}\mathbb{P}_n(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega})\xi\|O_p(1) + o_p(1). \qquad \text{(S2.4)}
\end{aligned}$$

The first term in (S2.4) can be shown to be $o_p(1)$ by conditions C2, C4–C6, and the central limit theorem. The second term is also $o_p(1)$ because $\sqrt{n}[\hat{A}_{n,lm}(t) - A_{0,lm}(t)] = O_p(1)$ and the fact that $\mathbb{P}_n\xi = o_{as}(1)$ by the strong law of large numbers. Conditions C3 and C5, and the central limit theorem ensure that the third term converges in probability to 0. Finally, the fourth, fifth and sixth terms are all $o_p(1)$ by the fact that $\|\sqrt{n}\mathbb{P}_n(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega})\xi\| = o_p(1)$, which follows from C5 and Lemma A.3 of Spiekerman and Lin (1998), and also the fact that $\sqrt{n}\mathbb{P}_n\boldsymbol{\omega}\xi$ converges in distribution by C5 and the central limit theorem. Therefore, $I_{n,lm2} = o_p(1)$.

Finally, $I_{n,lm3} = o_p(1)$ by Lemma 4.2 in Kosorok (2008), because

$$\|\hat{P}_{n,hl}(0,u-)\hat{P}_{n,mj}(u,t) - P_{0,hl}(0,u-)P_{0,mj}(u,t)\|_\infty = o_{as}(1),$$

as a result of Theorem 1, and the fact that $\sqrt{n}\mathbb{P}_n\psi_{lm}(t)$ converges weakly to a tight mean-zero Gaussian process by the Donsker property of the class $\{\psi_{lm}(t) : t \in [0,\tau]\}$. Taking all the pieces together it follows that

$$\|\hat{W}_{n,hj}(0,t) - \tilde{W}_{n,hj}(0,t)\|_\infty = o_p(1), \tag{S2.5}$$

for all $h \notin \mathcal{T}$ and $j \in \mathcal{T}$. The fact that (S2.5) also holds for all $h, j \notin \mathcal{T}$ can be shown by similar but simpler arguments because the influence functions do not involve the terms $\boldsymbol{\omega}^T \mathbf{R}_{lm}(t)$. Therefore, the proof of Theorem 2 is complete.

S3 Proof of Theorem 3

Using similar arguments to those used in the proof of Theorem 2, it can be shown that under the null hypothesis $PL_j(t;\boldsymbol{\beta}_0) = 0$,

$$\begin{aligned} V_{nj}(t) &= \sqrt{n}\mathbb{P}_n L_j(t;\boldsymbol{\beta}_0) - \sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)^T P[\dot{\pi}_j(\mathbf{Z},\boldsymbol{\beta}_0)RN_{\cdot\cdot}(t)] + o_p(1) \\ &\equiv \sqrt{n}\mathbb{P}_n\psi_j^L(t) + o_p(1), \quad j \in \mathcal{T}^{(-1)}. \end{aligned} \tag{S3.1}$$

where $\psi_j^L(t) = L_j(t;\boldsymbol{\beta}_0) - \boldsymbol{\omega}^T P[\dot{\pi}_j(\mathbf{Z},\boldsymbol{\beta}_0)RN_{\cdot\cdot}(t)]$. Note that the class $\mathcal{L}_j^L = \{\psi_j^L(t) : t \in [0,\tau]\}$ is P-Donsker, due to the Donsker property of the class $\{N_{\cdot j}(t) : t \in [0,\tau]\}$ and the fact that the class $\mathcal{L}_j^L$ is formed by (finite) sums of functions that belong to Donkser classes, which are multiplied by bounded random variables (conditions C4 and C6) and added to bounded random variables (condition C5). Now, we can make use of the conditional multiplier central limit theorem (van der Vaart and Wellner, 1996) and argue that the asymptotic distribution of $V_{nj}(t)$ is the same as the conditional on the data limiting distribution of $\tilde{V}_{nj}(t) = \sqrt{n}\mathbb{P}_n\psi_j^L(t)\xi$, where ξ_i is a random draw from $N(0,1)$. Next, we need to show that $\hat{V}_{nj}(t) = \sqrt{n}\mathbb{P}_n\hat{\psi}_j^L(t)\xi_j$ and $\tilde{V}_{nj}(t)$ converge weakly (unconditionally) to the same distribution. As in the proof of Theorem 2, this requires showing that $\|\hat{V}_{nj}(t) - \tilde{V}_{nj}(t)\|_\infty = o_p(1)$. It can be shown that

$$\begin{aligned}
\|\hat{V}_{nj}(t) - \tilde{V}_{nj}(t)\|_\infty \leq\ & \|\sqrt{n}\mathbb{P}_n\hat{\boldsymbol{\omega}}^T[\mathbb{P}_n\dot{\pi}_j(\mathbf{Z},\hat{\boldsymbol{\beta}}_n)RN_{\cdot\cdot}(t) - P\dot{\pi}_j(\mathbf{Z},\boldsymbol{\beta}_0)RN_{\cdot\cdot}(t)]\xi_j\|_\infty \\
& + \sum_{h\notin\mathcal{T}}\sum_{l\in\mathcal{T}}\|\sqrt{n}\mathbb{P}_n[\pi_j(D;\hat{\boldsymbol{\beta}}_n) - \pi_j(D;\boldsymbol{\beta}_0)]RN_{hl}(t)\xi_j\|_\infty \\
& + \|\sqrt{n}\mathbb{P}_n(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega})^T[P\dot{\pi}_j(\mathbf{Z},\boldsymbol{\beta}_0)RN_{\cdot\cdot}(t)]\xi_j\|_\infty. \qquad \text{(S3.2)}
\end{aligned}$$

The right side of (S3.2) is $o_p(1)$ and this can be shown by using similar arguments to those used in the proof of Theorem 2. Therefore, the proof of Theorem 3 is complete.

S4 Additional Simulation Results

Additional simulation results that are referred in Section 4 of the main text are presented below.

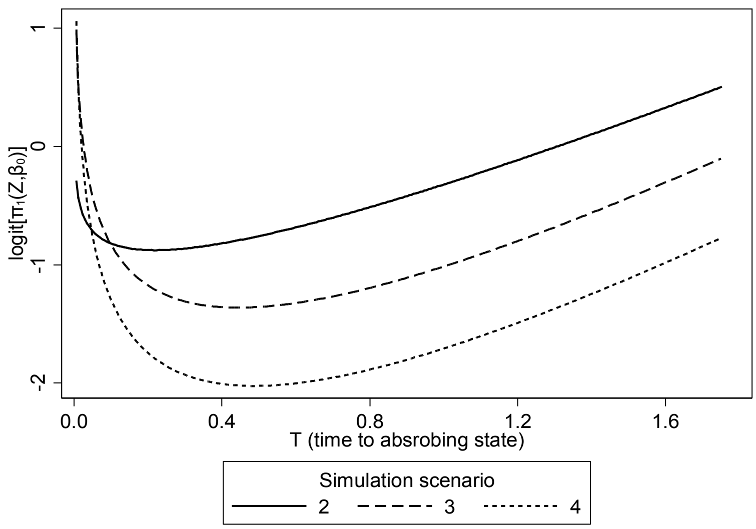

Figure S1: Dependence of the parametric model $\text{logit}[\pi_1(\mathbf{Z}, \boldsymbol{\beta}_0)]$ on time to absorbing state T, according to the misspecification simulation scenarios 2–4 with $\nu_1 = 0.8$, $\nu_1 = 0.4$ and $\nu_1 = 0.2$, respectively.

References

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

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference.* Springer, New York.

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

Stute, W (1995). The statistical analysis of Kaplan-Meier integrals. In *Analysis of censored data* (pp. 231-254). Institute of Mathematical Statistics, Hayward.

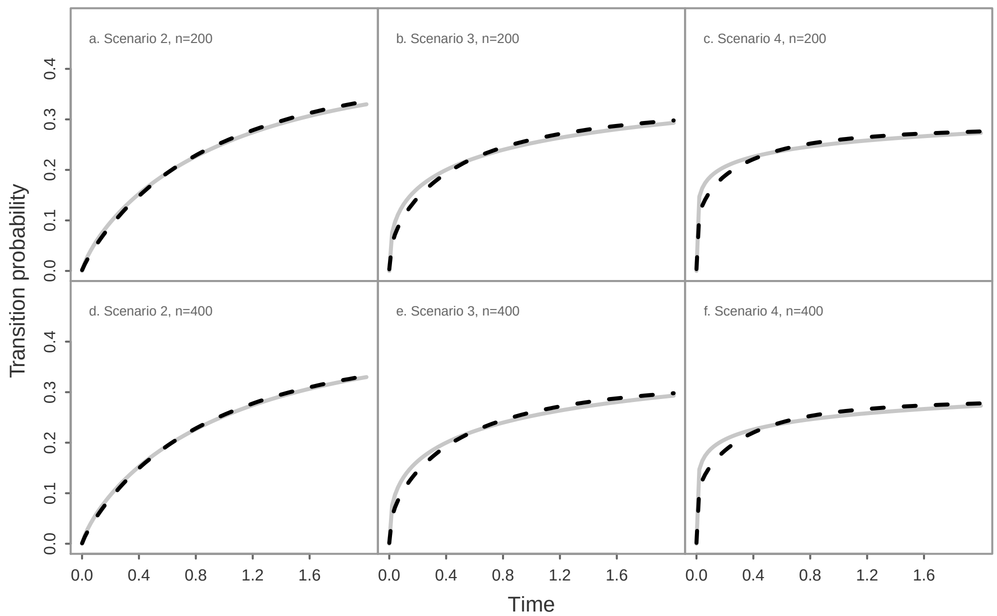

Figure S2: Transition probability estimates (dashed black lines) and corresponding true transition probabilities (solid grey lines) according to the sample size and the misspecification simulation scenarios 2–4 with $\nu_1 = 0.8$, $\nu_1 = 0.4$ and $\nu_1 = 0.2$, respectively. The missingness probability was 0.8.

van der Vaart, A. W. (2000). *Asymptotic statistics.* Cambridge University Press, Cambridge.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics.* Springer-Verlag, New York.

van der Vaart, A. W. and Wellner, J. A. (2007). Empirical processes indexed by estimated functions. In *Asymptotics: Particles, Processes and Inverse Problems. Institute of Mathematical Statistics Lecture Notes–Monograph Series* **55**:234–252. IMS, Beachwood, OH.

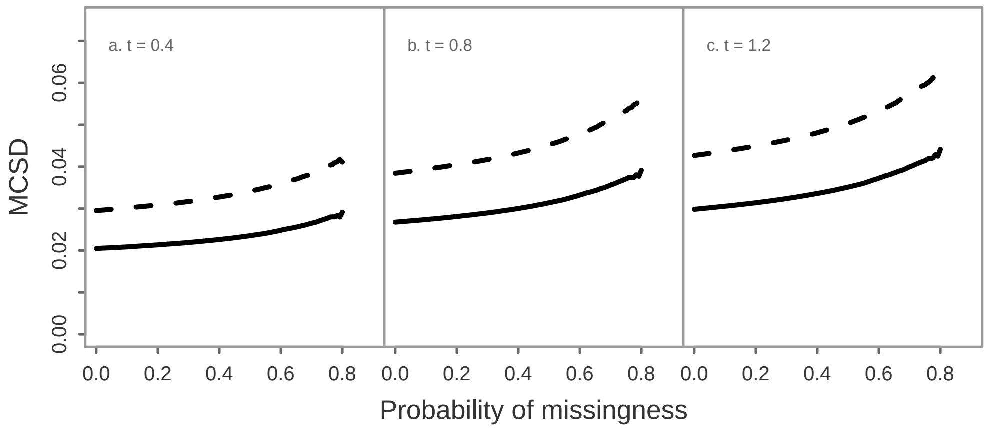

Figure S3: Monte Carlo standard deviation (MCSD) of the estimates for $t = 0.4$, $t = 0.8$ and $t = 1.2$, according to the probability of missingness and the sample size (dashed lines: $n = 200$; solid lines: $n = 400$).

Table S1: Pointwise simulation results for absorbing state 1 with $n = 200$ at $t_1 = 0.4$, $t_2 = 0.8$ and $t_3 = 1.2$, under model misspecification scenarios 3 and 4 with $\nu_1 = 0.4$ and $\nu_1 = 0.2$, respectively.

	Bias$\times 10^2$			MCSD$\times 10^3$			ASE$\times 10^3$			CP$\times 10^2$		
Scenario (ν_1)	t_1	t_2	t_3	t_1	t_2	t_3	t_1	t_2	t_3	t_1	t_2	t_3
Missing: 80%												
3 (0.4)	-0.8	0.4	0.7	49.1	57.7	62.1	47.2	55.2	58.9	92.9	93.4	92.2
4 (0.2)	-0.5	0.7	0.8	51.3	56.6	59.0	47.8	53.2	55.5	92.4	92.7	91.4
Missing: 60%												
3 (0.4)	-0.6	0.4	0.6	36.5	41.7	43.9	36.5	41.8	44.1	93.8	94.9	95.1
4 (0.2)	-0.5	0.4	0.6	39.0	41.6	43.2	38.2	41.3	42.9	94.2	94.5	94.6

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

Table S2: Pointwise simulation results for absorbing state 1 with $n = 400$ at $t_1 = 0.4$, $t_2 = 0.8$ and $t_3 = 1.2$, under model misspecification scenarios 3 and 4 with $\nu_1 = 0.4$ and $\nu_1 = 0.2$, respectively.

	Bias$\times 10^2$			MCSD$\times 10^3$			ASE$\times 10^3$			CP$\times 10^2$		
Scenario (ν_1)	t_1	t_2	t_3	t_1	t_2	t_3	t_1	t_2	t_3	t_1	t_2	t_3
Missing: 80%												
3 (0.4)	-1.1	0.1	0.5	33.0	38.7	40.7	32.9	38.8	41.5	93.0	94.2	94.5
4 (0.2)	-1.0	0.3	0.5	35.1	38.0	39.4	34.6	38.3	40.2	93.0	94.8	94.9
Missing: 60%												
3 (0.4)	-0.8	0.2	0.5	25.4	29.4	30.9	25.7	29.4	31.2	92.9	94.7	94.3
4 (0.2)	-0.7	0.4	0.7	27.1	29.2	29.9	27.2	29.3	30.4	92.6	94.7	95.0

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

Table S3: 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 model misspecification scenarios 3 and 4 with $\nu_1 = 0.4$ and $\nu_1 = 0.2$, respectively.

		Scenario 3		Scenario 4	
n	missing	EP	HW	EP	HW
200	80%	89.1	90.8	81.8	89.4
	60%	94.1	93.7	88.3	92.0
400	80%	82.9	92.6	62.3	87.3
	60%	89.2	93.6	73.0	89.5

Table S4: Simulation results on Monte Carlo standard deviation ($\times 10^3$) of the estimator according to the probability $\pi^\star_{11} = \Pr(C^\star = 1|C = 1)$, while setting $\pi^\star_{22} = \Pr(C^\star = 2|C = 2) = \pi^\star_{11}$ in all cases. A larger value of $\pi^\star_{11}$ indicates a higher accuracy of the auxiliary imperfect diagnostic test $C^\star$.

	$n = 200$ Missing: 80%			$n = 200$ Missing: 60%			$n = 400$ Missing: 80%			$n = 400$ Missing: 60%		
$\pi^\star_{11}$	t_1	t_2	t_3	t_1	t_2	t_3	t_1	t_2	t_3	t_1	t_2	t_3
0.5	46.1	63.0	72.1	33.9	46.5	52.5	32.1	44.1	50.2	23.4	32.0	36.6
0.6	45.8	63.5	72.1	33.3	45.9	51.9	31.0	42.9	48.6	23.2	31.4	35.7
0.7	45.3	62.4	70.6	33.1	45.2	50.8	31.3	43.1	48.9	23.2	31.4	35.6
0.8	44.3	60.8	68.7	32.8	44.7	50.1	30.6	41.9	47.3	22.8	30.6	34.8
0.9	42.8	58.2	65.8	32.0	43.3	48.5	29.6	40.4	45.5	22.4	30.0	33.9

Table S5: Pointwise simulation results for absorbing state 1 with $n = 400$ at $t_1 = 0.4$, $t_2 = 0.8$ and $t_3 = 1.2$, based on the Gouskova, Lin and Fine nonparametric approach (GLF) which does not incorporate auxiliary covariates, and the proposed method that incorporates the auxiliary covariate, under a correctly specified (Scenario 1), a mildly misspecified (Scenario 2), and a moderately misspecified (Scenario 3) model for $\pi_1(\mathbf{Z}, \boldsymbol{\beta})$. The probability of missingness was $\Pr(R = 0) = 0.6$, and did not depend on the auxiliary variable.

	Bias$\times 10^2$			MCSD$\times 10^3$			MSE$\times 10^4$		
Method	t_1	t_2	t_3	t_1	t_2	t_3	t_1	t_2	t_3
Scenario 1									
GLF	-0.1	0.0	-0.2	25.5	33.4	38.8	6.5	11.1	15.1
Proposed	0.1	0.1	0.2	22.8	29.9	33.6	5.2	8.9	11.3
Scenario 2									
GLS	0.1	0.1	0.1	27.0	34.3	55.1	7.3	11.7	30.4
Proposed	-0.2	0.2	0.3	24.3	30.6	33.7	6.0	9.4	11.5
Scenario 3									
GLS	0.4	0.4	0.5	29.5	55.6	96.0	8.9	31.1	92.5
Proposed	-0.6	0.4	0.7	27.1	31.1	33.1	7.7	9.9	11.5

MCSD, Monte Carlo standard deviation; MSE, mean squared error

Table S6: Pointwise simulation results with $n = 400$, for absorbing state 1 at $t_1 = 0.4$, $t_2 = 0.8$ and $t_3 = 1.2$, based on the Gouskova, Lin and Fine nonparametric approach (GLF) which does not incorporate auxiliary covariates, and the proposed method that incorporates the auxiliary covariate, under a correctly specified (Scenario 1), a mildly misspecified (Scenario 2), and a moderately misspecified (Scenario 3) model for $\pi_1(\mathbf{Z}, \boldsymbol{\beta})$. The probability of missingness was $\Pr(R = 0) = 0.5 + 0.2I_{\{C^{\star}=1\}}$, depending on the auxiliary variable.

	Bias×10^2			MCSD×10^3			MSE×10^4		
Method	t_1	t_2	t_3	t_1	t_2	t_3	t_1	t_2	t_3
Scenario 1									
GLS	-2.7	-4.4	-5.4	23.3	31.2	36.6	13.0	28.7	42.9
Proposed	0.1	0.2	0.2	24.2	32.0	36.3	5.9	10.2	13.2
Scenario 2									
GLS	-3.0	-4.4	-5.2	25.0	31.9	51.7	15.1	29.5	53.7
Proposed	-0.2	0.2	0.4	25.7	32.6	36.2	6.7	10.7	13.3
Scenario 3									
GLS	-2.7	-3.6	-4.4	28.7	38.3	175.2	15.3	27.4	326.2
Proposed	-0.6	0.5	0.8	29.0	33.5	35.8	8.7	11.5	13.6

MCSD, Monte Carlo standard deviation; MSE, mean squared error

Table S7: Simulation results on computing time in seconds for point estimates and standard errors under the proposed estimator for Scenario 1, according to the sample size.

	Missing: 80%				Missing: 60%			
	No SCB		With SCB		No SCB		With SCB	
n	Mean	SD	Mean	SD	Mean	SD	Mean	SD
200	0.6	0.1	0.9	0.1	0.7	0.1	0.9	0.1
300	1.7	0.2	2.6	0.3	1.9	0.1	2.7	0.1
400	3.2	0.1	4.9	0.2	3.4	0.1	4.9	0.1
500	5.1	0.2	7.8	0.2	5.4	0.2	7.7	0.2
600	8.1	0.3	11.7	0.4	8.0	0.3	11.3	0.3
700	11.8	0.3	17.0	0.5	12.0	0.5	16.5	0.5
800	17.7	0.7	24.6	0.9	17.0	0.8	22.9	0.8
900	24.1	0.7	32.5	0.9	23.7	0.9	30.8	0.9
1000	30.2	0.9	40.7	1.1	30.1	0.8	38.7	1.0
1100	40.0	1.2	51.5	1.5	38.7	1.0	49.3	1.4
1200	50.9	1.9	65.7	2.3	48.9	2.0	61.8	1.7
1300	62.9	1.9	79.9	1.8	60.0	1.7	75.2	1.8
1400	77.7	2.3	97.3	2.2	73.8	2.0	91.4	2.0
1500	94.2	2.4	116.3	3.1	89.4	2.3	109.6	2.4

SCB, simultaneous confidence band based on 1,000 simulations; SD, standard deviation

Table S8: Pointwise simulation results for the naïve approach under Scenario 1, according to the probability $\pi^{\star}_{11} = \Pr(C^{\star} = 1|C = 1)$, while setting $\pi^{\star}_{22} = \Pr(C^{\star} = 2|C = 2) = \pi^{\star}_{11}$ in all cases. A larger value of $\pi^{\star}_{11}$ indicates a higher accuracy and a lower misclassification rate of the imperfect diagnostic test $C^{\star}$.

	Bias×10^2			MCSD×10^3			ASE×10^3			CP×10^2		
$\pi^{\star}_{11}$	t_1	t_2	t_3	t_1	t_2	t_3	t_1	t_2	t_3	t_1	t_2	t_3
$n = 200$												
0.9	1.9	2.5	2.5	24.8	31.3	34.1	25.4	31.4	34.2	89.7	88.4	90.2
0.8	3.9	5.1	5.1	26.1	32.2	34.4	26.7	32.4	35.0	68.1	64.5	70.4
0.7	5.9	7.8	7.9	27.9	34.1	36.1	28.0	33.4	35.7	36.9	32.0	39.6
0.6	7.8	10.3	10.5	28.9	34.3	36.5	29.1	34.2	36.2	16.9	10.5	14.0
0.5	9.9	13.0	13.2	29.4	34.1	36.3	30.1	34.9	36.7	4.8	2.6	3.3
$n = 400$												
0.9	2.0	2.6	2.6	18.8	22.8	24.2	18.1	22.3	24.3	79.0	76.8	81.3
0.8	4.0	5.2	5.3	19.3	23.2	24.9	19.1	23.0	24.8	41.6	36.8	42.2
0.7	6.1	7.8	8.0	20.2	23.6	25.4	20.0	23.7	25.3	10.2	6.0	10.0
0.6	8.1	10.4	10.8	20.8	23.8	26.2	20.8	24.3	25.6	0.9	0.5	1.1
0.5	10.0	13.0	13.4	21.3	24.0	26.3	21.4	24.7	25.9	0.0	0.0	0.1

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