observation rate. An easy extension to account for other types of correlation would be to do a power transformation on one of the frailty terms. For example, model the intensity function of event process following Sun et al. (2007),

$$\lambda(t|\boldsymbol{x}, z) = z^{\rho}\lambda_0(t)\exp(\boldsymbol{x}'\boldsymbol{\beta}), \quad t \in [0, \tau],$$

where ρ is an unknown parameter which partly determines the correlation between the event and observation processes. Specifically, $\rho = 0$ means that the two processes are independent given $\boldsymbol{x}$. For $\rho > 0$ and subjects with the same $\boldsymbol{x}$, subjects with a higher event rate would have more frequent observations. On the other hand, $\rho < 0$ means that the two processes are negatively correlated. The estimating procedures under the two scenarios with known and unknown start time would remain the same except for estimating equations (5) and (8), where $\boldsymbol{x}_i^*$ changes to $(\log(\hat{z}_i), 1, \boldsymbol{x}_i')'$ and $\boldsymbol{\beta}^*$ changes to $(\rho, \log\Lambda_0(\tau), \boldsymbol{\beta}')'$ with $\hat{z}_i$ given by equations (7) and (9), respectively.

In conclusion, our methods were motivated by a real application of labor progression and have wider applicability to any panel count data scenario where the interest not just lies on the number of recurrences but also on the duration of time that they are recurrence free.

APPENDIX: PROOF OF LARGE SAMPLE PROPERTIES

Proof of Theorem 1:

Before we start the proof, we'd like to make a remark about the conditions. For (C7), we can set a very small ρ to truncate the gap times between observations, say, $\rho = 1e - 10$. In general, with finite sample size, $\{t_j - t_{j-1}, j = 1, \ldots, K\}$ would be greater than or equal to ρ. Condition (C8) is satisfied by a large class of distributions namely, Gamma, Weibull, Normal etc.

We focus on the proof of more difficult Scenario II, while a sketch is provided for Scenario I. In the following, we use the notation $a \lesssim b$ to denote that there exists a non-random constant $C > 0$ which is independent of n such that $a \leq Cb$; $a \gtrsim b$ is similarly defined. We will also denote

$$Pf(X) = E_P f(X) \text{ and } P_n f(X) = E_{P_n} f(X) = \frac{1}{n}\sum_{i=1}^{n} f(X_i).$$

Denote the observed data by X and the parameter $\boldsymbol{\theta} = (\boldsymbol{\eta}, F)$ where $\boldsymbol{\eta}$ is the $p-$dimensional (possibly covariate dependent) parameter for f_A and $F(t) = \frac{\Lambda_0(t)}{\Lambda_0(\tau)}, 0 < t < \tau$. Define

$$\Theta_n = \{\boldsymbol{\theta}_n = (\boldsymbol{\eta}, F_n)\} = \mathcal{B} \otimes \mathcal{M}_n,$$

where $\mathcal{B} = \{\boldsymbol{\eta} \in R^p, \|\boldsymbol{\eta}\| \leq M\}$ and $\mathcal{M}_n = \{F_n : F_n(t) = \sum_{j=1}^{l+k_n} \xi_j I_j(t), \xi_j \geq 0, \sum_{j=1}^{l+k_n} \xi_j \leq M_n\}$. The likelihood and the log-likelihood are denoted by $L(\boldsymbol{\theta}, X)$ and $l(\boldsymbol{\theta}, X)$ respectively and let $\alpha_j = F(\tilde{t}_j + A), j = 1, \ldots, K-1$ and $\alpha_{-1} = 0$. Recall also that $t_j = \tilde{t}_{j-1} + A$, $M_0 = \sum m_j \leq N(\tau)$ a.s. and

$$L(\boldsymbol{\theta}, X) = \int \prod_{j=1}^{K} \left[\frac{\alpha_{j-1} - \alpha_{j-2}}{\alpha_{K-1}}\right]^{m_j} f(A)\, dA.$$

Observe that $\left[\frac{\alpha_{j-1}-\alpha_{j-2}}{\alpha_{K-1}}\right] \leq 1$; consequently the likelihood $L(\boldsymbol{\theta}, X) \leq 1$. Moreover, by the mean value theorem and assumptions (C5), (C7) and (C3), we have

$$\alpha_{j-1} - \alpha_{j-2} = F'(t^*)(\tilde{t}_{j-1} - \tilde{t}_{j-2}) \gtrsim 1, \quad \alpha_{K-1} \gtrsim 1, \quad a.s., \tag{A.1}$$

where t^* is a point between $\tilde{t}_{j-1}$ and $\tilde{t}_{j-2}$, which in turn implies $\left[\frac{\alpha_{j-1}-\alpha_{j-2}}{\alpha_{K-1}}\right] \geq C$ for an adequate constant C. Thus,

$$|l(\boldsymbol{\theta}, X)| \lesssim N(\tau), \ a.s. \tag{A.2}$$

In other words, $N(\tau)$ is an envelope function for the class

$$\mathcal{L}_n := \{l(\boldsymbol{\theta}, X) : \boldsymbol{\theta} \in \Theta_n\}.$$

We will need the following partial derivatives:

$$\left.\begin{array}{lll} \frac{\partial l}{\partial \alpha_j} & = & \frac{1}{L(\boldsymbol{\theta},X)} \int \left(\frac{m_{j+1}}{\alpha_j - \alpha_{j-1}} - \frac{m_{j+2}}{\alpha_{j+1}-\alpha_j}\right) \prod_{j=1}^K \left[\frac{(\alpha_{j-1}-\alpha_{j-2})}{\alpha_{K-1}}\right]^{m_j} f(A)\, dA, j < K-1, \\ \frac{\partial l}{\partial \alpha_{K-1}} & = & \frac{1}{L(\boldsymbol{\theta},X)} \int \left(\frac{m_K}{\alpha_{K-1}-\alpha_{K-2}} - \frac{M_0}{\alpha_{K-1}}\right) \prod_{i=1}^K \left[\frac{(\alpha_{j-1}-\alpha_{j-2})}{\alpha_{K-1}}\right]^{m_i} f(A)\, dA \text{ and} \\ \frac{\partial l}{\partial \delta} & = & \frac{1}{L(\boldsymbol{\theta},X)} \int \prod_{j=1}^K \left[\frac{(\alpha_{j-1}-\alpha_{j-2})}{\alpha_{K-1}}\right]^{m_j} \frac{\partial f}{\partial \delta}(A)\, dA \end{array}\right\}$$

Using the analytic expression for the derivatives above, the mean value theorem, assumption (C8) and (A.1), we have

$$|l(\boldsymbol{\theta}_1, X) - l(\boldsymbol{\theta}_2, X)| \lesssim N(\tau)\{\|F_1 - F_2\|_\infty + \|\boldsymbol{\eta}_1 - \boldsymbol{\eta}_2\|\},$$

for any $\boldsymbol{\theta}_1 = (\boldsymbol{\eta}_1, F_1)$ and $\boldsymbol{\theta}_2 = (\boldsymbol{\eta}_2, F_2) \in \Theta_n$. Let $\boldsymbol{\xi}_k = (\xi_1^k, \ldots, \xi_{l+k_n}^k)$ denote the I-spline coefficients of F_k for $k = 1, 2$. It follows that

$$\|F_1 - F_2\|_\infty \leq \left(\sum_{j=1}^{l+k_n} |\xi_j^1 - \xi_j^2|\right) \max_{1 \leq j < \infty} \|I_j\|_\infty \lesssim \|\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2\|.$$

Let $\Omega_J = \{N(\tau) \leq J\}$ for a fixed $J \in \mathbb{N}$. In view of the estimate above and by the result in problem 6 on page 94 of van der Vaart and Wellner (1996) and the inequality above, it follows that on the set Ω_J, the covering number $N(\epsilon, \mathcal{L}_n, L_1(P_n))$ satisfies the estimate

$$N(\epsilon, \mathcal{L}_n, L_1(P_n)) \lesssim J^p M_n^{l+k_n} \epsilon^{-p_n}, p_n = p + l + k_n. \tag{A.3}$$

Proceeding in a similar manner, one can establish an analogous estimate for the covering number $N_2(\epsilon, \mathcal{L}_n, L_2(P_n))$ based on the L_2-norm $\|f\|_{L_2(P_n)} = (P_n f^2)^{1/2}$.

Lemma 1. *Assume that conditions (C1)–(C8) hold, we have*

$$\sup_{\boldsymbol{\theta} \in \Theta_n} |P_n l(\boldsymbol{\theta}, X) - Pl(\boldsymbol{\theta}, X)| \to 0, \quad a.s.$$

Proof. Let $\nu/2 < \phi_1 < 1/2$ and $\epsilon_n = \epsilon\, n^{-1/2+\phi_1}(\log n)^{1/2}$. For any $l(\boldsymbol{\theta}, X) \in \mathcal{L}_n$ and sufficiently large n, we have

$$Var(P_n l(\boldsymbol{\theta}, X))/16\epsilon_n^2 \le \frac{1/nPl^2(\boldsymbol{\theta}, X)}{16\epsilon_n^2} \lesssim \frac{1}{16n\epsilon_n^2} < \frac{1}{2},$$

where, we used (A.2) and the fact that $Pl^2(\boldsymbol{\theta}, X) \lesssim EN^2(\tau) < \infty$. By applying inequality (31) of Pollard (1984), we have

$$\begin{aligned}
& P\left(\{\sup_{\mathcal{L}_n} |P_n l(\boldsymbol{\theta}, X) - Pl(\boldsymbol{\theta}, X)| > 8\epsilon_n\} \cap \Omega_J\right) \\
\le \;& 8E_P N\left(\epsilon, \mathcal{L}_n, L_1(P_n)\right) \exp(-n\epsilon_n^2/128) \mathbf{1}(\sup_{\mathcal{L}_n} P_n l^2(\boldsymbol{\theta}, X) \le 64) + \\
& \qquad\qquad\qquad E_P \mathbf{1}(\sup_{\mathcal{L}_n} P_n l^2(\boldsymbol{\theta}, X) > 64/J^2) \\
\le \;& 8E_P N\left(\epsilon, \mathcal{L}_n, L_1(P_n)\right) \exp(-n\epsilon_n^2/128) + P(\sup_{\mathcal{L}_n} P_n l^2(\boldsymbol{\theta}, X) > 64/J^2) \\
= \;& I_n + II_n.
\end{aligned}$$

Using (A.3) and the calculation on page 4, supplementary material in Ma et al. (2015), we obtain

$$\sum_n I_n = 8 \sum_n E_P N\left(\epsilon, \mathcal{L}_n, L_1(P_n)\right) \exp(-n\epsilon_n^2/128) \lesssim J^p \sum_n \exp(-Cn^{2\phi_1} \log n) < \infty.$$

The last inequality holds because $\exp(-Cn^{2\phi_1} \log n) = n^{-Cn^{2\phi_1}}$ is decreasing when n increases. And $n^{-Cn^{2\phi_1}} < n^{-2}$ when $n \ge \lceil (2/C)^{1/(2\phi_1)} \rceil = C_2$. Thus

$$\sum_n \exp(-Cn^{2\phi_1} \log n) < \sum_{n=1}^{C_2} \exp(-Cn^{2\phi_1} \log n) + \sum_{n=C_2+1}^{\infty} n^{-2} < \infty$$

Now applying Lemma 33 of Pollard (1984) to the term II_n above, we have

$$II_n = P(\sup_{\mathcal{L}_n} P_n l^2(\boldsymbol{\theta}, X) > 64/J^2) \le E_P N_2\left(\epsilon, \mathcal{L}_n, L_2(P_n)\right) \exp(-n/J^2).$$

A similar calculation now yields

$$\sum_n II_n \leq \sum_n \exp(-Cn^{2\phi_1}/J^2) < \infty.$$

The last inequality holds because there exist a finite integer C_3 such that $Cn^{2\phi_1}/J^2 > 2\log(n)$ for any $n > C_3$. The calculation follows that of $\sum_n I_n$. Thus, we have shown that

$$\sum_n P\left(\{\sup_{\mathcal{L}_n} |P_n l(\boldsymbol{\theta}, X) - Pl(\boldsymbol{\theta}, X)| > 8\epsilon_n\} \cap \Omega_J\right) < \infty.$$

By the Borel-Cantelli lemma, for each J, there exists a null set $\mathcal{N}_J$ such that

$$\sup_{\boldsymbol{\theta}\in\Theta_n} |P_n l(\boldsymbol{\theta}, X) - Pl(\boldsymbol{\theta}, X)| \to 0 \ a.s.\text{on } \Omega_J \cap \mathcal{N}_J^c.$$

Since $N(\tau) < \infty$ $a.s.$, $\Omega = \cup_J \Omega_J$. By discarding the null set $\mathcal{N} = \cup_J \mathcal{N}_J$, we conclude that $\sup_{\boldsymbol{\theta}\in\Theta_n} |P_n l(\boldsymbol{\theta}, X) - Pl(\boldsymbol{\theta}, X)| \to 0$ $a.s.$

□

Thus we have established

$$N\left(\epsilon, \mathcal{L}_n, L_1(P_n)\right) \lesssim J^p M_n^{l+k_n} \epsilon^{-p_n},$$

and

$$\sup_{\boldsymbol{\theta}\in\Theta_n} |P_n l(\boldsymbol{\theta}, X) - Pl(\boldsymbol{\theta}, X)| \to 0, \quad a.s. \tag{A.4}$$

We now complete the proof of Theorem 1 following Chapter 5 of van der Vaart (2000).

Proof. Let $G(\boldsymbol{\theta}, X) = -l(\boldsymbol{\theta}, X)$ and

$$\zeta_{1n} = \sup_{\boldsymbol{\theta}\in\Theta_n} |P_n G(\boldsymbol{\theta}, X) - PG(\boldsymbol{\theta}, X)|, \quad \zeta_{2n} = P_n G(\boldsymbol{\theta}_0, X) - PG(\boldsymbol{\theta}_0, X),$$

where $P = P_{\boldsymbol{\theta}_0}$. Denote

$$K_\epsilon = \{\boldsymbol{\theta} = (\boldsymbol{\eta}, F) : \|\boldsymbol{\eta} - \boldsymbol{\eta}_0\| \geq \epsilon, \|F - F_0\|_\infty \geq \epsilon, \boldsymbol{\theta} \in \Theta_n\}.$$

It is easy to see that

$$\begin{aligned}\inf_{K_\epsilon} PM(\boldsymbol{\theta}, X) &= \inf_{K_\epsilon}\{PG(\boldsymbol{\theta}, X) - P_nG(\boldsymbol{\theta}, X) + P_nG(\boldsymbol{\theta}, X)\} \\ &\leq \zeta_{1n} + \inf_{K_\epsilon} P_nG(\boldsymbol{\theta}, X).\end{aligned}$$

If $\hat{\boldsymbol{\theta}}_n := \inf_{\boldsymbol{\theta}\in\Theta_n} P_nG(\boldsymbol{\theta}, X) \in K_\epsilon$, we have

$$\inf_{K_\epsilon} P_nG(\boldsymbol{\theta}, X) = P_nG(\hat{\boldsymbol{\theta}}_n, X) \leq P_nG(\boldsymbol{\theta}_0, X) + o(1) = \zeta_{2n} + o(1) + PG(\boldsymbol{\theta}_0, X).$$

Since $PG(\boldsymbol{\theta}, X)$ attains its unique minimum at the true value $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ (van der Vaart, 2000, p. 62) and K_ϵ is a compact set, we have that

$$\left\{\inf_{\boldsymbol{\theta}\in K_\epsilon} PG(\boldsymbol{\theta}, X)\right\} - PG(\boldsymbol{\theta}_0, X) = \delta_\epsilon > 0.$$

Thus,

$$\inf_{K_\epsilon} PG(\boldsymbol{\theta}, X) \leq \zeta_{1n} + \zeta_{2n} + o(1) + PG(\boldsymbol{\theta}_0, X) = \zeta_n + o(1) + PG(\boldsymbol{\theta}_0, X),$$

with $\zeta_n = \zeta_{1n} + \zeta_{2n}$. Hence we have $\zeta_n \geq \delta_\epsilon$ and $\{\hat{\boldsymbol{\theta}}_n \in K_\epsilon\} \subseteq \{\zeta_n \geq \delta_\epsilon\}$. By (A.4) and strong law of large numbers, we conclude that $\zeta_{1n} = o(1)$ and $\zeta_{2n} = o(1)$, almost surely. By $\cup_{k=1}^{\infty} \cap_{n=k}^{\infty} \{\hat{\boldsymbol{\theta}}_n \in K_\epsilon\} \subseteq \cup_{k=1}^{\infty} \cap_{n=k}^{\infty} \{\zeta_n \geq \delta_\epsilon\}$, we proved that $\|\hat{F}_n - F\|_\infty \to 0$ *a.s.*, which immediately implies $\|\hat{F}_n - F\|_2 \to 0$ *a.s.*

$\square$

Consistency of $\hat{\Lambda}_n$ and $\hat{\boldsymbol{\beta}}$: We will prove that $\hat{\Lambda}_n(t)$ is a consistent estimator of $\Lambda_0(t)$ for $t \in [0, \tau]$ by showing that $\hat{\alpha}$ converges to $\log\{\Lambda_0(\tau)\}$. Recall that $\hat{\alpha}$ is obtained by solving the estimating function (8). Define the function $U(\boldsymbol{\beta}^*) = n^{-1}\sum_{i=1}^{n} w_{3i}\boldsymbol{x}_i^{*'}[M_{0i}E_{A_i}\{\hat{F}_n^{-1}(t_{i\tilde{K}_i} + A_i)\} - \exp(\boldsymbol{x}_i^{*'}\boldsymbol{\beta}^*)]$. It can be shown that the function U converges to 0 almost surely when evaluated at $\boldsymbol{\beta}^* = [\log\{\Lambda_0(\tau)\}, \boldsymbol{\beta}']'$. Furthermore, it is easy to see that the derivative of U evaluated at $[\log\{\Lambda_0(\tau)\}, \boldsymbol{\beta}']'$ is negative definite. Applying Taylor expansion on $U(\boldsymbol{\beta}^*)$, one can show that the solution of estimating equation (8), i.e. $\hat{\boldsymbol{\beta}}^* = (\hat{\alpha}, \hat{\boldsymbol{\beta}}')'$, converges to $\boldsymbol{\beta}^*$ almost surely. Thus, we proved that $\hat{\alpha}$ converges to $\log\{\Lambda_0(\tau)\}$ almost surely. Along with the fact that $\|\hat{F}_n - F_0\|_2 \to 0$ almost surely as $n \to \infty$, it can be shown that $\|\hat{\Lambda}_n - \Lambda_0\|_2 = \|\hat{F}_n \exp(\hat{\alpha}) - \Lambda_0\|_2 \to 0$ almost surely as $n \to \infty$.

Consistency of the survival functions: This proof will follow given the strong consistency of $\hat{\Lambda}_n$, $\hat{\boldsymbol{\beta}}$ and $\hat{f}_z$. The consistency of $\hat{\Lambda}_n$, $\hat{\boldsymbol{\beta}}$ has already been established.

Note that if $z_i = \tilde{K}_i / \exp(\boldsymbol{x}_i^{*'} \boldsymbol{\gamma})$, then z_i are i.i.d. and, by Huang and Wang (2004), $\frac{1}{n}\sum_{i=1}^n z_i \to E(Z)$ as $n \to \infty$. Consequently, by strong law of large numbers, the kernel density estimator $\hat{f}_z(u) = \frac{1}{n}\sum_i K_h(u - z_i)$ converges uniformly to $f_z(u)$ as $n \to \infty$ and $h \to 0$, provided $|f_z''(u)| \leq C$. Since $\hat{z}_i = \tilde{K}_i / \exp(\boldsymbol{x}_i^{*'} \hat{\boldsymbol{\gamma}})$, and one can easily prove that $\hat{\gamma} \to \gamma$ $a.s.$ following the proof of strong consistency of $\hat{\boldsymbol{\beta}}$, the asymptotic consistency of $\hat{f}_z(u) = \frac{1}{n}\sum_i K_h(u - \hat{z}_i)$ follows using (C6).

The strong consistency of survival function in case of Scenario I with known start time follows analogously to Theorem 1 for the case of $F(t)$ being approximated by I-splines without the condition (C8). Alternatively, the proof of strong consistency in case of estimating Λ_0 through self-consistency algorithm in Scenario I follows by suitably modifying Huang et al. (2006) strong consistency results for $\hat{\Lambda}_0, \hat{\beta}$ and the consistency of $\hat{\gamma}$ delineated above, and the fact that the survival function of the gap time distribution is a smooth function of these quantities as shown in equation (2).

Proof of Theorem 2:

We now establish the rate of convergence of the proposed estimators. The proof is based on Theorem 3.4.1 of van der Vaart and Wellner (1996). First note that there exists I-splines F_{n0} such that $\|F_{n0} - F_0\|_\infty = O_P(n^{-r\nu})$, where r is as in (C5) and ν as in (C4) using Lu et al. (2009). Further, for any $\delta > 0$ define $\mathcal{F}_\delta = \{l(\theta_n, X) - l(\theta, X) : \theta \in \Theta_n, d(\theta, \theta_n) < \delta\}$, where $\theta_n = (\eta, F_n)$. Using (Shen and Wong, 1994, p. 597), the bracketing number can be shown to be bounded by

$$\log N_{[]}(\epsilon, \mathcal{F}_\delta, \|\cdot\|_2) < CN \log(\delta/\epsilon).$$

with $N = 2(m + k_n)$. Recall that

$$|l(\theta_n, X) - l(\theta, X)| \lesssim N(\tau)\{\|F_n - F\|_\infty + \|\eta_n - \eta\|\}.$$

Consequently,

$$\|l(\theta_n, X) - l(\theta, X)\|_2^2 \leq C\delta^2$$

for $l(\theta_n, X) - l(\theta, X) \in \mathcal{F}_\delta$. Finally, using Lemma 3.4.2 of van der Vaart and Wellner (1996) we get

$$E_P\|n^{1/2}(P_n - P)\|_{\mathcal{F}_\delta} \leq CJ_\delta(\epsilon, \mathcal{F}_\delta, \|\cdot\|_2)\left\{1 + \frac{J_\delta(\epsilon, \mathcal{F}_\delta, \|\cdot\|_2)}{\delta^2 n^{1/2}}\right\},$$

where $J_\delta(\epsilon, \mathcal{F}_\delta, \|\cdot\|_2) = \int_0^\tau (1 + \log N_{[]}(\epsilon, \mathcal{F}_\delta, \|\cdot\|_2))^{1/2} d\epsilon$. The right hand side of the above equation can be shown to be $\leq \phi_n(\delta) = C\{N^{1/2}\delta + \frac{N}{n^{1/2}}\}$. Observe that $\phi_n(\delta)/\delta$ is decreasing in δ and for $r_n = N^{-1/2}n^{1/2} = n^{(1-\nu)/2},\ 0 < \nu < 0.5$, we have $r_n^2\phi_n(r_n) < 2n^{1/2}$. Thus, invoking Theorem 3.4.1 of van der Vaart and Wellner (1996) yields $n^{(1-\nu)/2}\|\hat{\theta} - \theta_{n0})\|_2 = O_P(1)$. Combining with $\|\theta_{n0} - \theta_0\|_2 = O_P(n^{-r\nu})$, we get $\|\hat{\theta} - \theta_0\|_2 = O_P(n^{-(1-\nu)/2} + n^{-r\nu})$. We can also establish the $n^{-1/2}$ rate of convergence for $\hat{\beta}$ using standard techniques as they are solution to estimating equation U.

Further, note that for every k, $\hat{S}_{M_{k,0}}$ is a smooth and bounded function of $\hat{\Lambda}_0$ and that $\hat{\beta}$ is consistent estimator of β, using Taylor series expansion and the consistency properties proven in Theorem 1, we have

$$\|\hat{S}_{M_{k,0}} - S_{M_{k,0}}\|_2 = O_P(n^{-(1-\nu)/2} + n^{-r\nu}).$$