Appendix: Asymptotic Distribution Theory and Additional Simulation Results

---

We now provide some detail on the asymptotic distributional results listed in §5, from which those in §3 follow as a special case.

As in §5 assume that the processes $\{(N_{ji}, Y_{ji}), j = 1, \ldots, q; Z(S_{1i}, \ldots, S_{qi})\}$ are i.i.d. for $i = 1, \ldots, n$. Also suppose that the modeled regression variables in (16) and (17) have bounded total variation, so that

$$||X_{ki}(0)|| + \int_0^{\tau_k} ||X_{ki}(dt)|| \text{ for each } k = 1, \ldots, K \text{ and } i = 1, \ldots, n,$$

and

$$||X_{kgi}(0,0)|| + \int_0^{\tau_k}\int_0^{\tau_g} ||X_{kgi}(dt_1, dt_2)|| \text{ for each } 1 \leq k < g \leq K \text{ and } i = 1, \ldots, n,$$

are bounded by a constant almost surely, where $||\cdot||$ denotes vector length. Here the regions of integration, $[0, \tau_k]$ and $[0, \tau_k] \times [0, \tau_g]$, are such that $P\{S_{ji} \geq \tau_k; Z(0, \ldots, S_{ji}, 0, \ldots, 0)\} > 0$ for some j such that $M(j) = k$ for each $k = 1, \ldots, K$, and $P\{S_{ji} \geq \tau_k, S_{hi} \geq \tau_g; Z(0, \ldots, S_{ji}, \ldots, S_{hi}, \ldots, 0)\} > 0$ for some (j, h) such that $M(j) = k$ and $M(h) = g$ for each $1 \leq k < g \leq K$, for any $i = 1, \ldots, n$. Finally, to ensure that there is definitive information for estimating β in (16) and γ in (17) we require the positive semidefinite off-diagonal blocks in the negative partial derivative matrix A in §5 to have positive definite almost sure limits, so that

$$A_1(\beta) = E\left[\sum_{j=1}^{q}\sum_{k=1}^{K} I\{M(j) = k\} \int_0^{\tau_k} \{X_k(t) - \bar{x}_k(t;\beta)\}^{\otimes 2} Y_j(t)\Gamma_k(dt)\exp\{X_k(t)\beta\}\right],$$

and

$$A_2(\gamma) = E\Bigg[\sum_{j=1}^{g}\sum_{h=j+1}^{q}\sum_{k=1}^{K}\sum_{g=k+1}^{K} I\{M(j) = k\}I\{M(h) = g\}$$
$$\int_0^{\tau_k}\int_0^{\tau_g} \{X_{kg}(t_1, t_2) - \bar{x}_{kg}(t_1, t_2;\gamma)\}^{\otimes 2} Y_j(t_1)Y_h(t_2)\Gamma_{kg}(dt_1, dt_2)\exp\{X_{kg}(t_1, t_2)\gamma\}\Bigg],$$

are required to be positive definite at β_0 and γ_0, the 'true' values of β and γ respectively. These condition naturally extend those of Spiekerman and Lin (1998). We also, temporarily, require $T_1, \ldots, T_q$ to be absolutely continuous.

Under these conditions, the arguments of Lin et al. (2000) adapt to yield the asymptotic results listed in §5. In particular, their Lemma 1 generalizes directly to include both univariate functions defined on $[0, \tau]$ and bivariate functions defined on $[0, \tau_1] \times [0, \tau_2]$. For the latter we have

Lemma 1(b). *Let f_n and g_n be two sequences of bounded bivariate real functions such that for some τ_1 and τ_2*

(i) $\sup_{\substack{0 \le t_1 \le \tau_1 \\ 0 \le t_2 \le \tau_2}} |f_n(t_1, t_2) - f(t_1, t_2)| \to 0,$ *where f is continuous on $[0, \tau_1] \times [0, \tau_2]$.*

(ii) $\{g_n\}$ *are monotone on $[0, \tau_1] \times [0, \tau_2]$ and*

(iii) $\sup_{\substack{0 \le t_1 \le \tau_1 \\ 0 \le t_2 \le \tau_2}} |g_n(t_1, t_2) - g(t_1, t_2)| \to 0,$ *for some bounded function g.*

Then $\sup_{\substack{0 \le t_1 \le \tau_1 \\ 0 \le t_2 \le \tau_2}} | \int_0^{t_1} \int_0^{t_2} f_n(s_1, s_2) g_n(ds_1, ds_2) - \int_0^{t_1} \int_0^{t_2} f(s_1, s_2) g(ds_1, ds_2)| \to 0,$

and $\sup_{\substack{0 \le t_1 \le \tau_1 \\ 0 \le t_2 \le \tau_2}} | \int_0^{t_1} \int_0^{t_2} g_n(s_1, s_2) f_n(ds_1, ds_2) - \int_0^{t_1} \int_0^{t_2} g(s_1, s_2) f(ds_1, ds_2)| \to 0,$

as $n \to \infty$.

The proof of the first assertion follows exactly as in Lin et al. (2000) for the univariate case (denoted here as Lemma 1(a)) and the second assertion follows from the first by integration by parts.

Now consider the consistency of $(\hat{\beta}, \hat{\gamma})$ solving (18) and (19) as estimator of (β_0, γ_0).

Denote

$$D_1(\beta) = n^{-1} \sum_{j=1}^{q} \sum_{k=1}^{K} I\{M(j)=k\} \Bigg[\sum_{i=1}^{n} \int_0^{\tau_k} X_{ki}(t)(\beta-\beta_0) N_{ji}(dt) - \int_0^{\tau_k} \log\left\{ \frac{Q_k^{(0)}(t;\beta)}{Q_k^{(0)}(t;\beta_0)} \right\} \sum_{i=1}^{n} N_{ji}(dt) \Bigg]$$

and

$$D_2(\gamma) = n^{-1} \sum_{j=1}^{g} \sum_{h=j+1}^{g} \sum_{k=1}^{K} \sum_{g=k+1}^{K} I\{M(j)=k\} I\{M(h)=g\} \Bigg[\sum_{i=1}^{n} \int_0^{\tau_k} \int_0^{\tau_g} X_{kgi}(t_1,t_2)(\gamma-\gamma_0) N_{ji}(dt_1) N_{hi}(dt_2) - \int_0^{\tau_k} \int_0^{\tau_g} \log\left\{ \frac{Q_{kg}^{(0)}(t_1,t_2;\gamma)}{Q_{kg}^{(0)}(t_1,t_2;\gamma_0)} \right\} \sum_{i=1}^{n} N_{ji}(dt_1) N_{hi}(dt_2) \Bigg].$$

Under the conditions listed above, the strong law of large numbers, the bounded variation of $Q_k^{(0)}(t;\beta)$, $Q_{kg}^{(0)}(t_1,t_2;\gamma)$, $n^{-1}\sum_{i=1}^n \int_0^t N_{ji}(ds)$ and $n^{-1}\sum_{i=1}^n \int_0^{t_1}\int_0^{t_2} N_{ji}(ds_1)N_{hi}(ds_2)$ for all (k,g,j,h) implies that $(D_1(\beta)', D_2(\gamma)')'$ converges almost surely to $(\mathcal{D}_1(\beta)', \mathcal{D}_2(\gamma)')'$ where

$$\mathcal{D}_1(\beta) = E\left(\sum_{j=1}^{q} \sum_{k=1}^{K} I\{M(j)=k\} \left[\int_0^{\tau_k} X_k(t)(\beta-\beta_0) N_j(dt) - \int_0^{\tau_k} \log\left\{ \frac{q_k^{(0)}(t;\beta)}{q_k^{(0)}(t;\beta_0)} \right\} N_j(dt) \right] \right)$$

and

$$\mathcal{D}_2(\gamma) = E\Bigg(\sum_{j=1}^{q} \sum_{h=j+1}^{q} \sum_{k=1}^{K} \sum_{g=k+1}^{K} I\{M(j)=k\} I\{M(h)=g\} \Bigg[\int_0^{\tau_k} \int_0^{\tau_g} X_{kg}(t_1,t_2)(\gamma-\gamma_0) N_j(dt_1) N_h(dt_2) - \int_0^{\tau_k} \int_0^{\tau_g} \log\left\{ \frac{q_{kg}^{(0)}(t_1,t_2;\gamma)}{q_{kg}^{(0)}(t_1,t_2;\gamma_0)} \right\} N_j(dt_1) N_h(dt_2) \Bigg] \Bigg)$$

for all (β,γ). In these expressions N_j denotes the counting process corresponding to a random observation on the jth failure time T_j, $q_k^{(\ell)}(t;\beta) = E\{Q_k^{(\ell)}(t;\beta)\}$, and $q_{kg}^{(\ell)}(t_1,t_2;\gamma) = E\{Q_{kg}^{(\ell)}(t_1,t_2;\gamma)\}$ for $\ell = 0,1,2$, and all $k = 1,\ldots,K$ and all (k,g) such that $1 \le k < g \le K$.

Straightforward calculations give

$$\frac{\partial^2 D_1(\beta)}{\partial\beta^2} = -n^{-1}\sum_{j=1}^{q}\sum_{k=1}^{K} I\{M(j)=k\}\sum_{i=1}^{n}\int_0^{\tau_k}\{X_{ki}(t)-\bar{X}_k(t;\beta)\}^{\otimes 2} Y_{ji}(t)\exp\{X_{ki}(t)\beta\}\frac{\sum_{\ell=1}^{n} N_{j\ell}(dt)}{Q_k^{(0)}(t;\beta)}$$

and

$$\frac{\partial^2 D_2(\gamma)}{\partial\gamma^2} = -n^{-1}\sum_{j=1}^{q}\sum_{h=j+1}^{q}\sum_{k=1}^{K}\sum_{g=k+1}^{K} I\{M(j)=k\}I\{M(h)=g\}$$
$$\sum_{i=1}^{n}\int_0^{\tau_k}\int_0^{\tau_g}\{X_{kg}(t_1,t_2)-\bar{X}_{kg}(t_1,t_2;\gamma)\}^{\otimes 2} Y_{ji}(dt_1)Y_{hi}(dt_2)\exp\{X_{kg}(t_1,t_2)\gamma\}$$
$$\frac{\sum_{\ell=1}^{n} N_{j\ell}(dt_1)N_{h\ell}(dt_2)}{Q_{kg}^{(0)}(t_1,t_2;\gamma)},$$

which are each negative semidefinite. Hence $\mathcal{D}_1(\beta)$ and $\mathcal{D}_2(\gamma)$ are concave, so that convergence of $D_1(\beta)$ to $\mathcal{D}_1(\beta)$ and of $D_2(\gamma)$ to $\mathcal{D}_2(\gamma)$ is uniform on a compact set, such as $\|\beta-\beta_0\|\le r_1$ and $\|\gamma-\gamma_0\|\le r_2$, for $r_1>0$, and $r_2>0$ and

$$\sup_{\|\beta-\beta_0\|\le r_1}\|D_1(\beta)-\mathcal{D}_1(\beta)\| \text{ and } \sup_{\|\gamma-\gamma_0\|\le r_2}\|D_2(\gamma)-\mathcal{D}_2(\gamma)\|$$

converge to zero almost surely. Also $\mathcal{D}_1(\beta)$ is concave with $\partial\mathcal{D}_1(\beta_0)/\partial\beta_0=0$ and $\partial^2\mathcal{D}_1(\beta_0)/\partial\beta_0^2=-A_1(\beta_0)$ under (16), and $\mathcal{D}_2(\gamma)$ is concave with $\partial\mathcal{D}(\gamma_0)/\partial\gamma_0=0$ and $\partial^2\mathcal{D}_2(\gamma_0)/\partial\gamma_0^2=-A_2(\gamma_0)$ under (17), so that $(\mathcal{D}_1(\beta),\mathcal{D}_2(\gamma))$ is uniquely maximized at (β_0,γ_0). One can now argue as in Lin et al. (2000, Appendix A.1) to show that $\hat\beta$ converges to β_0 almost surely under (16), and $\hat\gamma$ converges almost surely to γ_0 under (17).

Now consider the asymptotic distribution of the left sides of (18) and (19) at (β_0,γ_0).

One can define a composite 'process'

$$[\{\bar{L}_1(\cdot;\beta_0),\bar{H}_1(\cdot;\beta_0)\},\dots,\{\bar{L}_q(\cdot;\beta_0),\bar{H}_q(\cdot;\beta_0)\},\{\bar{L}_{12}(\cdot;\cdot;\gamma_0),\bar{H}_{12}(\cdot,\cdot;\gamma_0)\},\dots,$$
$$\{\bar{L}_{q-1,q}(\cdot,\cdot;\gamma_0),\bar{H}_{q-1,q}(\cdot,\cdot;\gamma_0)\}],$$

where, continuing the notation of §5,

$$\bar{L}_j(t;\beta)=n^{-1/2}\sum_{i=1}^{n}L_{ji}(t;\beta),$$

$$\bar{H}_j(t;\beta)=n^{-1/2}\sum_{i=1}^{n}\int_0^t\sum_{k=1}^{K}I\{M(j)=k\}Y_{ji}(s)X_{ki}(s)\exp\{X_{ki}(s)\beta\}\Gamma_k(ds)$$

for $j=1,\dots,q,$ and

$$\bar{L}_{jh}(t_1,t_2;\gamma)=n^{-1/2}\sum_{i=1}^{n}L_{jhi}(t_1,t_2;\gamma),\ \text{and}$$

$$\bar{H}_{jh}(t_1,t_2;\gamma)=n^{-1/2}\sum_{i=1}^{n}\int_0^{t_1}\int_0^{t_2}\sum_{k=1}^{K}\sum_{g=k+1}^{K}I\{M(j)=k\}I\{M(h)=g\}$$
$$Y_{ji}(s_1)Y_{hi}(s_2)X_{kgi}(s_1,s_2)\exp\{X_{kgi}(s_1,s_2)\gamma\}\Gamma_{kg}(ds_1,ds_2)$$

for $1\le j<h\le q$.

This composite process derives from sums of zero mean processes under (16) and (17) to which the functional central limit theorem (e.g. Pollard, 1990, p. 53) applies over the integration regions defined above. Arguing as in Lin et al. (2000, Appendix A.2), and using the total variation boundedness of the modeled covariates in (16) and (17), which allows one to restrict attention to modeled covariates that are non-negative, one has that this composite process derives from processes that converge weakly and jointly to zero mean Gaussian processes uniformly in their time arguments, over the designated follow-up periods. Moreover sample paths for the limiting Gaussian process are continuous for absolutely continuous $T_1,\dots,T_q$.

The strong embedding theorem (Shorack and Wellner, 1986, pp.47–48) allows this weak convergence to be replaced by almost sure convergence in a new probability space. The monotonicity of $Q_k^{(0)}(t;\beta_0)$ and of each component of $Q_k^{(1)}(t;\beta_0)$ for each $k=1,\dots,K$, and the monotonicity of $Q_{kg}^{(0)}(t_1,t_2;\gamma_0)$ and of each component of $Q_{kg}^{(1)}(t_1,t_2;\gamma_0)$ for $1\le k<g\le K$ using the total variation boundedness conditions on the modeled covariates, then allows Lemma 1(a) and 1(b) to be applied to functions that pertain to the left sides of (18)

and (19). Specifically, Lemma 1 and the almost sure convergence of the above composite process in the new probability space implies that the processes given by

$$\sum_{j=1}^{q} \int_0^t \bar{L}_j(ds;\beta_0)/[\sum_{k=1}^{K} I\{M(j)=k\} Q_k^{(0)}(s;\beta_0)]$$
$$\text{and } \sum_{j=1}^{q} \int_0^t \bar{L}_j(ds;\beta_0) \left[\sum_{k=1}^{K} I\{M(j)=k\} \frac{Q_k^{(1)}(s;\beta_0)}{Q_k^{(0)}(s;\beta_0)}\right]$$

converge uniformly in t almost surely to their Gaussian limits. These limits involve replacement of $Q_k^{(\ell)}$ by $q_k^{(\ell)}$ for $k=1,\ldots,K$ and $\ell=0,1$. Similarly the processes

$$\sum_{j=1}^{q} \sum_{h=j+1}^{q} \int_0^{t_1} \int_0^{t_2} \bar{L}_{jh}(ds_1, ds_2;\gamma_0)/\left[\sum_{k=1}^{K} \sum_{g=k+1}^{K} I\{M(j)=k\} I\{M(h)=g\} Q_{kg}^{(0)}(s_1,s_2;\gamma_0)\right]$$

and

$$\sum_{j=1}^{q} \sum_{h=j+1}^{q} \int_0^{t_1} \int_0^{t_2} \bar{L}_{jh}(ds_1, ds_2;\gamma_0) \left[\sum_{k=1}^{K} \sum_{g=k+1}^{K} I\{M(j)=k\} I\{M(h)=g\} \frac{Q_{kg}^{(1)}(s_1,s_2;\gamma_0)}{Q_{kg}^{(0)}(s_1,s_2:\gamma_0)}\right]$$

converge uniformly in (t_1,t_2) to their Gaussian limits that involve replacement of $Q_{kg}^{(\ell)}$ by $q_{kg}^{(\ell)}$ for $1 \leq k < g \leq K$ and $\ell=0,1$. In conjunction with the almost sure convergence of $\bar{H}_j(t;\beta_0), j=1,\ldots,K$ and of $\bar{H}_{jh}(t_1,t_2;\gamma_0), 1 \leq j < h \leq K$ in the new probability space one then has that the left sides of (18) and (19), denoted by $(U_1(\beta)', U_2(\gamma)')'$ are such that

$$\begin{pmatrix} n^{-1/2} & U_1(\beta_0) \\ n^{-1/2} & U_2(\gamma_0) \end{pmatrix}$$

converges almost surely to a mean zero Gaussian variate with covariance matrix

$$\Sigma = E\begin{pmatrix} \sum_{j=1}^{q}\sum_{k=1}^{K} I\{M(j)=k\} \int_0^{\tau_k} \{X_k(t) - \bar{x}_k(t;\beta_0)\} L_j(dt;\beta_0) \\ \sum_{j=1}^{q} \sum_{h=j+1}^{q} \sum_{k=1}^{K} \sum_{g=k+1}^{K} I\{M(j)=k\} I\{M(h)=g\} \int_0^{\tau_k} \int_0^{\tau_q} \{X_{kg}(t_1,t_2) - \bar{x}_{kg}(t_1,t_2;\gamma_0)\} L_{jk}(dt_1,dt_2;\gamma_0) \end{pmatrix}^{\otimes 2}.$$

The almost sure convergence in the new probability space implies weak convergence in the original probability space.

A Taylor expansion gives

$$\begin{pmatrix} n^{1/2} & (\hat\beta - \beta_0) \\ n^{1/2} & (\hat\gamma - \gamma_0) \end{pmatrix} = \begin{pmatrix} \hat A_1(\beta_*) & 0) \\ 0 & \hat A_2(\gamma_*) \end{pmatrix}^{-1} \begin{pmatrix} n^{-1/2} & U_1(\beta_0) \\ n^{-1/2} & U_2(\gamma_0) \end{pmatrix},$$

where the elements of β_* are each on the corresponding line segment between $\hat\beta$ and β_0, and the elements of γ_* are each on the corresponding line segment between $\hat\gamma$ and γ_0. The consistency of $\hat\beta$ and $\hat A_1(\beta_0)$ for β_0 and $A_1(\beta_0)$ respectively, and consistency of $\hat\gamma$ and $\hat A_2(\gamma_0)$ for γ_0 and $A_2(\gamma_0)$ respectively, along with the weak convergence of

$$\begin{pmatrix} n^{-1/2} & U_1(\beta_0) \\ n^{-1/2} & U_2(\gamma_0) \end{pmatrix},$$

implies that

$$\begin{pmatrix} n^{1/2} & (\hat\beta - \beta_0) \\ n^{1/2} & (\hat\gamma - \gamma_0) \end{pmatrix}$$

converges weakly to a zero mean normal variate with covariance matrix $A^{-1}\Sigma A^{-1}$. Distribution theory for the baseline hazard rates in (16) and (17) is required to show that the covariance matrix is consistently estimated by $\hat A^{-1}\hat\Sigma\hat A^{-1}$, with $\hat A$ and $\hat\Sigma$ as specified in §5.

One can write

$$\hat\Gamma_k(t;\beta) = \sum_{j=1}^{q} I\{M(j)=k\} \int_0^t \bar N_j(ds)/\{nQ_k^{(0)}(s;\beta)\},$$

and

$$\hat\Gamma_{kg}(t_1,t_2;\gamma) = \sum_{j=1}^{q}\sum_{h=j+1}^{q} I\{M(j)=k\}I\{M(h)=g\} \int_0^{t_1}\int_0^{t_2} \overline{N_jN_h}(ds_1,ds_2)/\{nQ_{kg}^{(0)}(s_1,s_2;\gamma)\},$$

where $\bar N_j(dt) = \sum_{i=1}^n N_{ji}(dt)$ and $\overline{N_jN_h}(dt_1,dt_2) = \sum_{i=1}^n N_{ji}(dt_1)N_{hi}(dt_2)$.

The uniform law of large numbers (Pollard, 1990, p. 41) implies that $n^{-1}\sum_{j=1}^q I\{M(j)=k\}\bar N_j(t)$ converges to its expectation uniformly in (t,β) for β in a neighborhood of β_0, that $n^{-1}\sum_{j=1}^q\sum_{h=j+1}^q I\{M(j)=k\}I\{M(h)=g\}\overline{N_jN_h}(t_1,t_2)$ converges to its expectation uniformly in (t_1,t_2,γ) for γ in a neighborhood of γ_0, giving the uniform convergence of $\hat\Gamma_k(t;\beta)$ to

$$\int_0^t \frac{q_k^{(0)}(s;\beta_0)}{q_k^{(0)}(s;\beta)}\Gamma_k(ds;\beta_0)$$

under (16) for $k = 1, \ldots, K$, and the uniform convergence of $\hat{\Gamma}_{kg}(t_1, t_2; \gamma)$ to

$$\int_0^{t_1} \int_0^{t_2} \frac{q_{kg}^{(0)}(s_1, s_2; \gamma_0)}{q_{kg}^{(0)}(s_1, s_2; \gamma)} \Gamma_{kg}(ds_1, ds_2; \gamma_0)$$

under (17) for $1 \le k < g \le K$. The derivatives of $\hat{\Gamma}_k(t; \beta)$ and $\hat{\Gamma}_{kg}(t; \gamma)$ with respect to β and γ respectively are uniformly bounded for n sufficiently large, for (β, γ) in a bounded region. Hence the strong consistency of $(\hat{\beta}', \hat{\gamma}')'$ for $(\hat{\beta}_0', \hat{\gamma}_0')'$ implies that $\hat{\Gamma}_k(t; \hat{\beta})$ converges almost surely to $\Gamma_k(t; \beta_0)$ uniformly for any $k = 1, \ldots K$, and that $\hat{\Gamma}_{kg}(t_1, t_2; \hat{\gamma})$ converges almost surely to $\hat{\Gamma}_{kg}(t_1, t_2; \gamma_0)$ uniformly in (t_1, t_2) for any $1 \le k < g \le K$. This along with the almost sure convergence of $\bar{X}_k(t; \beta_0)$ to $\bar{x}_k(t; \beta_0)$ for $k = 1, \ldots, K$, and $\bar{X}_{kg}(t_1, t_2; \gamma_0)$ to $\bar{x}_{kg}(t_1, t_2; \gamma_0)$ for $1 \le k < g \le K$, implies the convergence of n^{-1} times the square of the norm of

$$\begin{pmatrix} \displaystyle\sum_{i=1}^{n}\sum_{j=1}^{q}\sum_{k=1}^{K} I\{M(j)=k\}\Big[\int_0^{\tau_k}\{X_{ki}(t_j) - \bar{X}_k(t_j;\hat{\beta})\}\hat{L}_{ki}(dt_j;\hat{\beta}) \\ \displaystyle\qquad - \int_0^{\tau_k}\{X_{ki}(_j) - \bar{x}_k(t_j;\beta_0)\}L_{ki}(dt_j;\beta_0)\Big] \\ \displaystyle\sum_{i=1}^{n}\sum_{j=1}^{q}\sum_{h=j+1}^{q} I\{M(j)=k\}I\{M(h)=g\}\Big[\int_0^{\tau_k}\int_0^{\tau_g}\{X_{kgi}(t_j,t_h) - \bar{X}_{kgi}(t_j,t_h;\hat{\gamma})\}\hat{L}_{kgi}(dt_j,dt_h;\hat{\gamma}) \\ \displaystyle\qquad - \int_0^{\tau_k}\int_0^{\tau_g}\{X_{kgi}(t_j,t_h) - \bar{x}_{kg}(t_j,t_h;\gamma_0)\}L_{kgi}(dt_j,dt_h;\gamma_0)\Big] \end{pmatrix}$$

to zero almost surely. Now, applying the strong law of large numbers to

$$\begin{pmatrix} \displaystyle n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{q}\sum_{k=1}^{K} I\{M(j)=k\}\int_0^{\tau_k}\{X_{ki}(t_j) - \bar{x}_k(t_j;\beta_0)\}L_{ki}(dt_j;\beta_0) \\ \displaystyle n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{q}\sum_{h=j+1}^{q}\sum_{k=1}^{K}\sum_{g=k+1}^{K} I\{M(j)=k\}I\{M(h)=g\}\int_0^{\tau_k}\int_0^{\tau_g}\{X_{kg}(t_j,t_h) - \bar{x}(t_j,t_h;\gamma_0)\}L_{kgi}(dt_j,dt_h;\gamma_0) \end{pmatrix}^{\otimes 2}$$

shows that $\hat{\Sigma}$ is consistent for Σ. Furthermore the almost sure convergence of $\hat{\beta}$ and $\hat{A}(\beta_0)$ to β_0 and $A_1(\beta_0)$ respectively, and of $\hat{\gamma}$ and $\hat{A}_2(\gamma_0)$ to γ_0 and $A_2(\gamma_0)$ respectively, implies the almost sure convergence of $\hat{A}$ to

$$\begin{pmatrix} A_1(\beta_0) & 0 \\ 0 & A_2(\gamma_0) \end{pmatrix}$$

and implies the consistency of $\hat{A}^{-1}\hat{\Sigma}\hat{A}^{-1}$ as estimator of the covariance matrix of $\{n^{1/2}(\hat{\beta}-\beta_0)', n^{1/2}(\hat{\gamma}-\gamma_0)')\}'$.

The collection of processes $\hat{V}_k(\cdot;\beta_0) = n^{1/2}\{\hat{\Gamma}_k(\cdot;\hat{\beta}) - \Gamma_k(\cdot;\beta_0)\}$, for $k = 1,\ldots,K$ and $\hat{V}_{kg}(\cdot,\cdot;\gamma_0) = n^{1/2}\{\hat{\Gamma}_{kg}(\cdot,\cdot;\hat{\gamma}) - \Gamma_{kg}(\cdot,\cdot;\gamma_0)\}$ for $1 \le k < g \le K$ also converges to a zero mean zero Gaussian 'process' under the conditions listed above. Briefly, one can write

$$\begin{aligned}\hat{V}_k(t;\beta_0) =& n^{1/2}\left\{\sum_{j=1}^{q} I\{M(j)=k\}\int_0^t \bar{N}_j(ds)/\{nQ_k^{(0)}(s;\beta_0)\}\right\}\\ &+ n^{1/2}\sum_{j=1}^{q} I\{M(j)=k\}\left[\int_0^t \bar{N}_j(ds)/\{nQ_k^{(0)}(s;\hat{\beta})\} - \int_0^t \bar{N}_j(ds)/\{nQ_k^{(0)}(s;\beta_0)\}\right].\end{aligned}$$

The difference between the first term in this expression and

$$n^{1/2}\sum_{i=1}^{n}\sum_{j=1}^{q} I\{M(j)=k\}\int_0^t L_{ji}(ds;\beta_0)/q_k^{(0)}(s;\beta_0)$$

converges in probability to zero as $n\to\infty$, almost surely uniformly over $t\in[0,\tau_k]$. Also, following Taylor expansion about β_0, the difference between the second term and

$$\begin{aligned}-\sum_{j=1}^{q} I\{M(j)=k\}\int_0^t \bar{x}_k(s;\beta_0)\Gamma_k(ds;\beta_0)A_1(\beta_0)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{q} I\{M(j)=k\}\\ \int_0^{\tau_k}\{X_{ki}(s)-\bar{x}_k(s;\beta_0)\}L_{ji}(ds;\beta_0)\end{aligned}$$

can be seen to converge in probability to zero, almost surely uniformly over $t\in[0,\tau_k]$. It follows that the difference between $\hat{V}_k(t;\beta)$ and $n^{-1/2}\sum_{i=1}^{n}\Psi_{ki}(t;\beta_0)$, where

$$\begin{aligned}\Psi_{ki}(t;\beta_0) = \sum_{j=1}^{q} I\{M(j)=k\}\Bigg[&\int_0^t L_{ji}(ds;\beta_0)/q_k^{(0)}(s;\beta_0) - h_k(t;\beta_0)A_1(\beta_0)^{-1}\\ &\int_0^{\tau_k}\{X_{ki}(s;\beta_0)-\bar{x}_k(s;\beta_0)\}L_{ji}(ds;\beta_0)\Bigg]\end{aligned}$$

with $h_k(t;\beta_0) = \sum_{i=1}^{q} I\{M(j)=k\}\int_0^t \bar{x}_k(s;\beta_0)\Gamma_k(ds;\beta_0)$, converges in probability to zero, almost surely uniformly over $t\in[0,\tau_k]$, for each $k=1,\ldots,K$. Similarly the difference between $\hat{V}_{kg}(t_1, t_2; \gamma)$ and $n^{-1/2} \sum_{i=1}^n \Psi_{kgi}(t_1, t_2; \gamma_0)$, where

$$\Psi_{kgi}(t_1, t_2; \gamma_0) = \sum_{j=1}^q \sum_{h=j+1}^q I\{M(j) = k\} I\{M(h) = k\} \Bigg[\int_0^{t_1} \int_0^{t_2} L_{jhi}(ds_1, ds_2; \gamma_0) / q_{kg}^{(0)}(s_1, s_2; \gamma_0)$$
$$- h_{kg}(t_1, t_2; \gamma_0) A_2(\gamma_0)^{-1} \int_0^{\tau_k} \int_0^{\tau_g} \{X_{kgi}(s_1, s_2) - \bar{x}_{kg}(s_1, s_2; \gamma_0)\} L_{kgi}(ds_1, ds_2; \gamma_0) \Bigg],$$

with $h_{kg}(t_1, t_2; \gamma_0) = \sum_{j=1}^q \sum_{h=j+1}^q I\{M(j) = k\} I\{M(h) = g\} \int_0^{t_1} \int_0^{t_2} \bar{x}_{kg}(s_1, s_2; \gamma_0) \Gamma_{kg}(ds_1, ds_2; \gamma_0),$

converges in probability to zero as $n \to \infty$, uniformly almost surely over $(t_1, t_2) \in [0, \tau_k] \times [0, \tau_g]$ for each $1 \le k < g \le K$.

Application of the functional central limit theorem then shows the collection of processes $\hat{V}_k(\cdot; \beta_0), k = 1, \ldots, K$ and $\hat{V}_{kg}(\cdot, \cdot, \gamma_0)$, $1 \le k < g \le K$ to converge jointly to a zero mean Gaussian field. The covariance function between $\hat{V}_k(\cdot, \beta_0)$ and $\hat{V}_g(\cdot, \beta_0)$ can be consistently estimated at follow-up times (t, s) by the empirical covariance

$$n^{-1} \sum_{i=1}^n \hat{\Psi}_{ki}(t; \hat{\beta}) \hat{\Psi}_{gi}(s; \hat{\beta}),$$

almost surely uniformly in t and s where, for example, $\hat{\Psi}_{ki}$ equals Ψ_{ki} with β_0 replaced by $\hat{\beta}$, $\Gamma_k(\cdot, \beta_0)$ replaced by $\hat{\Gamma}_k(\cdot, \hat{\beta})$, $L_{ji}(\cdot; \beta_0)$ replaced $\hat{L}_{ji}(\cdot; \hat{\beta})$, $q_k^{(0)}(\cdot; \beta_0)$ replaced by $Q_k^{(0)}(\cdot, \hat{\beta})$, and $\bar{x}_k(\cdot; \beta_0)$ replaced by $\bar{X}_k(\cdot; \hat{\beta})$.

Similarly, the covariance function between $\hat{V}_{kg}(\cdot, \cdot; \gamma_0)$ and $\hat{V}_{\ell m}(\cdot, \cdot; \gamma_0)$ can be consistently estimate at follow-up times (t_1, s_1) and (t_2, s_2) by the empirical covariance estimator

$$n^{-1} \sum_{i=1}^n \hat{\Psi}_{kgi}(t_1, s_1; \hat{\gamma}) \hat{\Psi}_{\ell mi}(t_2, s_2; \hat{\gamma})$$

almost surely uniformly in t_1, s_1, t_2 and s_2 where, for example, $\hat{\Psi}_{kgi}(t_1, s_1; \gamma_0)$ is obtained by everywhere inserting sample estimates in $\Psi_{kgi}(t_1, s_1; \gamma_0)$.

Finally the covariance function between $\hat{V}_k(\cdot, \beta_0)$ and $\hat{V}_{\ell m}(\cdot, \cdot, \gamma_0)$ can be consistently estimated at follow-up times t_1 and (t_2, s_2) by the empirical covariance estimator

$$n^{-1} \sum_{i=1}^n \hat{\Psi}_{ki}(t_1; \hat{\beta}) \hat{\Psi}_{\ell mi}(t_2, s_2; \hat{\gamma})$$

almost surely uniformly in t_1, t_2, and s_2.

These developments allow pointwise confidence intervals to be estimated for marginal single and double baseline hazard functions. For general covariate history Z_0 one can re-center covariate values so that modeled covariate values corresponding to Z_0 are identically zero. The baseline hazard function estimators as described above then estimate marginal single and double failure hazard rates at covariate history Z_0.

These estimators can be used to induce asymptotic Gaussian distributions for estimators of parameters that arise through compact differentiable transformations of these hazard rates, with corresponding 'delta function' formula giving consistent variance estimators. For example, with fixed or external covariates the joint survivor function estimators $\hat{F}(\cdot, \cdot; Z)$ discussed in §3 have the necessary differentiability properties, but with rather complex derivative function connecting F to corresponding single and double hazard rate functions via (11).

For parameters having the requisite differentiability properties, but for which the derivative function is too complex to be useful, or for parameters that arise from transformations on parameters in (16) and (17) that are not compact differentiable, such as the supremum statistics of §3.3, a bootstrap resampling procedure can be used to estimate distributional characteristics.

Finally, straightforward generalization of the asymptotic theory sketched above allows the failure time variates $T_1, \ldots, T_q$ to be discrete with a finite number of mass points in the integration region defining the parameter estimates, or to include both continuous and discrete components. For estimators of β and $\Gamma_k, k = 1, \ldots, K$ one can apply the above arguments without change at all continuity points for each of the failure time variates, in conjunction with almost sure convergence of hazard rates to their expectation under (16) at mass points for each of the failure time variates, leading to weak Gaussian convergence for corresponding parameter estimates. Similarly, for estimators of γ and Γ_{kg}, one can divide the sample space for (T_k, T_q) into its four components comprised of the set of (t_k, t_q) continuity points, continuity point for T_k and mass point for T_g, continuity point for T_g and mass point for T_k, and mass points (t_k, t_q), for $1 \leq k < g \leq K$. The above arguments apply directly to the set of continuity points for both variates, and with minor variations to each of the other three sample space components as well. The estimators of covariances, or covariance processes, given above are applicable with continuous, discrete or mixed failure time variates. Of course, some care may be needed in specifying regression variates in (16) and (17) if failure times include discrete components, owing to the restriction that discrete hazard rates necessarily take values in $[0, 1]$.

This Appendix ends with tables showing additional simulation results that were called out in the manuscript narrative:

Appendix References

Lin, D., L. Wei, I. Yang, and Z. Ying (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 62*(4), 711–730.

Pollard, D. (1990). *Empirical Processes: Theory and Applications*. Hayward: Institute of Mathematical Statistics.

Shorack, G. and J. Wellner (1986). *Empirical Processes with Applications to Statistics*. New York: Wiley.

Spiekerman, C. F. and D. Lin (1998). Marginal regression models for multivariate failure time data. *Journal of the American Statistical Association 93*(443), 1164–1175.

Supplementary Table 1: Simulation summary statistics[a] for empirical and bootstrap confidence interval estimators for $\hat{\Lambda}_{11}$ at selected marginal survival probabilities under the Clayton–Oakes model (9) with $\beta_{10} = \beta_{01} = \gamma = \log 2$ and dependency parameter $\theta = 2$, at both $z = 0$ and $z = 1$.

Sample size (n)		500				1000				250			
T_1 and T_2 Failing %		22.6				22.6				100			
		$\hat{\Lambda}_{11}$				$\hat{\Lambda}_{11}$				$\hat{\Lambda}_{11}$			
(T_1, T_2) Marginal Survival Rates	Λ_{11}	Empirical SD	95% CI Coverage	Bootstrap SD	95% CI Coverage	Empirical SD	95% CI Coverage	Bootstrap SD	95% CI Coverage	Empirical SD	95% CI Coverage	Bootstrap SD	95% CI Coverage
$z = 0$													
(0.85, 0.85)	0.060	0.020	0.916	0.020	0.912	0.014	0.921	0.015	0.919	0.015	0.925	0.015	0.932
(0.85, 0.70)	0.114	0.039	0.910	0.039	0.913	0.028	0.934	0.028	0.935	0.024	0.925	0.024	0.932
(0.85, 0.55)	0.161	0.060	0.855	0.063	0.855	0.045	0.898	0.047	0.898	0.032	0.916	0.032	0.920
(0.70, 0.70)	0.226	0.076	0.900	0.079	0.901	0.055	0.927	0.056	0.932	0.040	0.932	0.041	0.934
(0.70, 0.55)	0.330	0.115	0.848	0.127	0.861	0.090	0.901	0.095	0.910	0.055	0.917	0.055	0.918
(0.55, 0.55)	0.500	0.162	0.750	0.196	0.783	0.139	0.843	0.159	0.858	0.078	0.935	0.079	0.983
$z = 1$													
(0.85, 0.85)	0.046	0.015	0.903	0.015	0.893	0.011	0.927	0.011	0.929	0.017	0.928	0.017	0.930
(0.85, 0.70)	0.092	0.027	0.900	0.027	0.912	0.019	0.928	0.019	0.929	0.026	0.946	0.027	0.944
(0.85, 0.55)	0.140	0.042	0.876	0.043	0.913	0.030	0.921	0.030	0.927	0.035	0.947	0.035	0.949
(0.70, 0.70)	0.189	0.049	0.900	0.049	0.919	0.034	0.926	0.035	0.930	0.041	0.937	0.041	0.934
(0.70, 0.55)	0.290	0.076	0.906	0.078	0.907	0.054	0.927	0.054	0.927	0.054	0.943	0.055	0.946
(0.55, 0.55)	0.453	0.120	0.880	0.124	0.908	0.085	0.924	0.086	0.922	0.074	0.937	0.075	0.943

[a]Based on 1000 simulations at each sample configuration. Empirical SD is the average of SD estimates based on the empirical variance estimator for $\hat{\Lambda}_{11}$. Bootstrap SD is the mean SD estimate from averaging the sample variances for $\hat{\Lambda}_{11}$ based on 200 replicates for each generated sample. 95% CI coverage is the fraction of the 1000 simulated samples where the asymptotic confidence interval using either the empirical or bootstrap SD includes the true Λ_{11} value.

Supplementary Table 2: Simulation summary statistics[a] for average cross ratio $\hat{C}$ and concordance ($\hat{\mathcal{T}}$) estimators at selected marginal survival function percentiles under the Clayton–Oakes model (9) with $\beta_{10} = \beta_{01} = \gamma = \log 2$ and cross ratio parameter $\theta = 2$, at both $z = 0$ and $z = 1$.

Sample size (n)		1000	250		1000	250
T_1 and T_2 failure %		22.6	100		22.6	100
		$\hat{C}$			$\mathcal{T}$	
(T_1, T_2) Percentiles	C	Mean(SD)	Mean (SD)	$\mathcal{T}$	Mean (SD)	Mean (SD)
				$z = 0$		
(0.85, 0.85)	3	3.075 (0.794)	3.090 (0.696)	0.5	0.492 (0.096)	0.498 (0.085)
(0.85, 0.70)	3	3.128 (0.951)	3.066 (0.550)	0.5	0.493 (0.107)	0.500 (0.067)
(0.85, 0.55)	3	3.282 (1.489)	3.058 (0.502)	0.5	0.490 (0.138)	0.501 (0.061)
(0.70, 0.70)	3	3.157 (1.071)	3.046 (0.460)	0.5	0.493 (0.117)	0.501 (0.055)
(0.70, 0.55)	3	3.200 (1.443)	3.039 (0.448)	0.5	0.486 (0.143)	0.500 (0.054)
(0.55, 0.55)	3	3.149 (1.802)	3.041 (0.448)	0.5	0.470 (0.169)	0.500 (0.054)
				$z = 1$		
(0.85, 0.85)	2	2.019 (0.532)	2.116 (0.880)	0.333	0.318 (0.122)	0.312 (0.186)
(0.85, 0.70)	2	2.020 (0.470)	2.086 (0.643)	0.333	0.324 (0.102)	0.327 (0.139)
(0.85, 0.55)	2	2.034 (0.533)	2.085 (0.570)	0.333	0.325 (0.108)	0.333 (0.119)
(0.70, 0.70)	2	2.002 (0.422)	2.069 (0.484)	0.333	0.324 (0.091)	0.335 (0.103)
(0.70, 0.55)	2	2.011 (0.476)	2.051 (0.419)	0.333	0.324 (0.097)	0.336 (0.090)
(0.55, 0.55)	2	2.029 (0.541)	2.036 (0.352)	0.333	0.326 (0.104)	0.336 (0.076)

[a]Sample mean and standard deviation (SD) based on 1000 simulations at each sample configuration..

Supplementary Table 3: Double failure hazard function estimators ($\hat{\Lambda}_{11}$) for breast cancer incidence and total mortality in the Women's Health Initiative Dietary Modification Trial (n =48,835, with 1764 women with breast cancer, 2508 deaths, and 134 women with both breast cancer and death during the 8.5 year average trial intervention period.

Follow-up Years for Breast Cancer (T_1)		Comparison Group ($z=0$) Follow-up Years for Mortality (T_2)			Intervention Group ($z=1$) Follow-up Years for Mortality (T_2)		
		3	6	9	3	6	9
3	$\hat{\Lambda}_{11}(\times 10^3)$	0.07[c]	0.67	1.74	0.05	0.43	1.11
	95% CI[a]	(0,0.16)	(0.41,0.93)	(1.25,2.22)	(0,0.10)	(0.22,0.63)	(0.68,1.54)
	95% CB[b]	(0,0.89)	(0,1.49)	(0.92,2.56)	(0,0.76)	(0,1.13)	(0.04,1.81)
6	$\hat{\Lambda}_{11}(\times 10^3)$	0.07[c]	1.18	3.15	0.05	0.75	2.01
	95% CI	(0,0.16)	(0.82,1.54)	(2.43,3.88)	(0,0.10)	(0.45,1.05)	(1.34,2.67)
	95% CB	(0,0.89)	(0.04,2.00)	(2.33,3.97)	(0,0.76)	(0.05,1.45)	(1.30,2.71)
9	$\hat{\Lambda}_{11}(\times 10^3)$	0.07[c]	1.18	3.42	0.05	0.75	2.18
	95% CI	(0,0.16)	(0.82,1.54)	(2.64,4.21)	(0,0.10)	(0.45,1.05)	(1.47,2.90)
	95% CB	(0,0.89)	(0.04,2.00)	(2.61,4.24)	(0,0.076)	(0.05,1.45)	(1.46,2.89)

[a] 95% confidence intervals for Λ_{11} given z based on 200 bootstrap replicates.
[b] 95% supremum-type confidence bands for Λ_{11} given z over the region $[0,9]\times[0,9]$ years, based on 200 bootstrap replicates.
[c] Repetition of some estimators, confidence intervals and bands occurs because double failure hazard rates are zero below the main diagonal.