SUPPLEMENT TO "SEMIPARAMETRIC LATENT-CLASS MODELS FOR MULTIVARIATE LONGITUDINAL AND SURVIVAL DATA"

BY KIN YAU WONG, DONGLIN ZENG, AND D. Y. LIN

S1. Simulation studies on the inconsistency of the NPMLE. We demonstrate the inconsistency of the NPMLE under a simple latent-class model. Suppose that there are two latent classes. For a sample of size n, let C_i denote the latent class membership and (π_{i1}, π_{i2}) denote subject-specific probabilities of class membership for the ith subject, such that $P(C_i = g) = \pi_{ig}$ for $g = 1, 2$ and $i = 1, \ldots, n$. Given $C_i = g$, the quantitative outcome variable Y_i follows the normal distribution with mean μ_g and variance σ_g^2, and the event time T_i is independent of Y_i, with the hazard function $\lambda_g(t)\exp(X_i\beta_g)$, where λ_g is a class-specific baseline hazard function, X_i is a covariate, and β_g is a regression parameter. Assume that the event time is subject to right censoring. For the ith subject, let U_i be the censoring time, $\widetilde{T}_i = T_i \wedge U_i$, and $\Delta_i = I(T_i \le U_i)$. The likelihood function is proportional to

$$\prod_{i=1}^{n}\sum_{g=1}^{2}\pi_{ig}\sigma_g^{-1}\exp\Big\{-\frac{1}{2\sigma_g^2}(Y_i-\mu_g)^2\Big\}\lambda_g(\widetilde{T}_i)^{\Delta_i}e^{\Delta_i X_i\beta_g}\exp\Big\{-\Lambda_g(\widetilde{T}_i)e^{X_i\beta_g}\Big\},$$

where $\Lambda_g(t) = \int_0^t \lambda_g(s)\,\mathrm{d}s$. The NPMLE approach sets the class-specific cumulative baseline hazard functions to be step functions that jump at the observed event times and maximizes the above likelihood function with $\lambda_g(\widetilde{T}_i)$ replaced by $\Lambda_g\{\widetilde{T}_i\}$, where $\Lambda_g\{t\}$ is the jump size of Λ_g at t. The estimators can be computed using the EM algorithm with C_i $(i = 1, \ldots, n)$ treated as missing data.

We generated $\pi_{i2} = 0.5W_i + 0.9(1 - W_i)$, where $W_i \sim \text{Bernoulli}(0.5)$, and $X_i \sim \mathrm{N}(0, 1)$ $(i = 1, \ldots, n)$. We set $\lambda_1(t) = 0.5$ and $\lambda_2(t) = \exp(0.25t)$; the other parameter values are given in Table S5. We generated the censoring variable from $\mathrm{Uniform}(0, 5)$, such that the censoring proportion is about 20%. We considered sample sizes of $n = 1000$ and 2000. To ensure that any biased results are not due to numerical issues, we set the initial values of the Euclidean parameter estimates at the true values, and the initial values of Λ_g matched the true values at the observed event times. The empirical biases of the estimators based on 1000 replicates are presented in Table S5. Estimators of the parameters of the first latent class are clearly biased. The bias does not reduce as the sample size increases from 1000 to 2000.

S2. Conditions of noninformative censoring and longitudinal measurement times. Let $\mathbb{N}_R \equiv (\mathbb{N}_{R,1}, \ldots, \mathbb{N}_{R,J})$ be a multivariate counting process, such that $\mathbb{N}_{R,j}(t)$ is the number of measurements of the jth longitudinal outcome by time t $(j = 1, \ldots, J; t \in [0, \tau])$. Let $\mathbb{Y}$ be the process of longitudinal outcomes and $\mathbb{X}$ and $\widetilde{\mathbb{X}}$ be the processes of covariates in the longitudinal outcome model, such that $(\boldsymbol{Y}, \boldsymbol{X}, \widetilde{\boldsymbol{X}})$ consist of corresponding values of $\mathbb{Y}(t)$, $\mathbb{X}(t)$, and $\widetilde{\mathbb{X}}(t)$ at $\{t : \mathrm{d}\mathbb{N}_{R,j}(t) = 1\}$ for $j = 1, \ldots, J$. Let $\mathbb{N}_T(\cdot) \equiv I(T \le \cdot)$ and $\mathbb{N}_U(\cdot) \equiv I(U \le \cdot)$ be the counting processes for the event of interest and censoring, respectively. For $t \in [0, \tau]$, let $\mathbb{W}(t)$ denote $(\boldsymbol{W}, \mathbb{X}(t), \widetilde{\mathbb{X}}(t), \boldsymbol{Z}(t))$, $\mathcal{H}_W(t) = \{\mathbb{W}(s) : 0 \le s < t\}$, $\mathcal{H}_R(t) = \{\mathbb{R}(s) : 0 \le s < t\}$, and $\mathcal{H}_{RY}(t) = \{\mathbb{Y}_j(s) : 0 \le s < t, \mathrm{d}\mathbb{N}_{R,j}(s) = 1, j = 1, \ldots, J\}$. We assume the following conditions.

(A1) The covariate process $\mathbb{W}(t)$ is fully observed for $t \in [0, \tau]$. In addition, the conditional distribution of $\mathbb{W}(t)$ given $T \ge t$, $U \ge t$, $\mathcal{H}_W(t)$, $\mathcal{H}_{RY}(t)$, $\mathcal{H}_R(t)$, $\boldsymbol{b}$, and C and the conditional distribution of $\mathbb{W}(t)$ given $T \wedge U = s$, $\mathcal{H}_W(t)$, $\mathcal{H}_{RY}(t)$, $\mathcal{H}_R(t)$, $\boldsymbol{b}$, and C do not depend on $(\boldsymbol{b}, C)$ for $t \in (0, \tau]$ and $s < t$.

(A2) The intensity function of $\mathbb{N}_T$ at t given $T \ge t$, $U \ge t$, $\mathbb{W}(t)$, $\mathcal{H}_W(t)$, $\mathcal{H}_{RY}(t)$, $\mathcal{H}_R(t)$, $\boldsymbol{b}$, and C depends only on $(\boldsymbol{Z}(t), \boldsymbol{b}, C)$ for $t \in [0, \tau]$.

(A3) The conditional distribution of $\mathbb{Y}_j(t)$ given T, U, $\mathbb{W}(t)$, $\mathcal{H}_W(t)$, $\mathcal{H}_{RY}(t)$, $\{\mathbb{Y}_k(t) : k \ne j\}$, $\mathbb{N}_R(t)$, $\mathcal{H}_R(t)$, $\boldsymbol{b}$, and C depends only on $(\mathbb{X}(t), \widetilde{\mathbb{X}}(t), \boldsymbol{b}, C)$ for $t \in [0, \tau]$ and $j = 1, \ldots, J$, where $\mathbb{Y}_j(t)$ is the jth component of $\mathbb{Y}(t)$.

(A4) The intensity function of $\mathbb{N}_U$ at t given $T > t$, $U \geq t$, $\mathbb{W}(t)$, $\mathcal{H}_W(t)$, $\mathcal{H}_{RY}(t)$, $\mathcal{H}_R(t)$, $\boldsymbol{b}$, and C does not depend on $(\boldsymbol{b}, C)$ for $t \in [0, \tau]$.

(A5) Both the intensity function of $\mathbb{N}_R$ at t given $T > t$, $U > t$, $\mathbb{W}(t)$, $\mathcal{H}_W(t)$, $\mathcal{H}_R(t)$, $\mathcal{H}_{RY}(t)$, $\boldsymbol{b}$, and C and the intensity function of $\mathbb{N}_R$ at t given $T \wedge U = s$, $\mathbb{W}(t)$, $\mathcal{H}_W(t)$, $\mathcal{H}_R(t)$, $\mathcal{H}_{RY}(t)$, $\boldsymbol{b}$, and C do not depend on $(\boldsymbol{b}, C)$ for $t \in [0, \tau]$ and $s < t$.

These conditions are analogous to conditions (A.2)–(A.6) of Zeng and Cai [5]. We allow the measurement process $\mathbb{N}_R$ to jump at $t > T$, because measurements beyond the event occurrence is possible for nonterminal events.

Under conditions (A1)–(A5), the joint density of the observed variables for a generic subject is

$$
\begin{aligned}
&\sum_{C=1}^{G} \int f(C, \boldsymbol{b} \mid \boldsymbol{W}) \prod_{t \leq T \wedge U} \Bigg(f\{\mathbb{W}(t) \mid T \geq t, U \geq t, \mathcal{H}_W(t), \mathcal{H}_R(t), \mathcal{H}_{RY}(t)\} \\
&\times \Big[1 - \mathrm{E}\{\mathrm{d}\mathbb{N}_T(t) \mid T \geq t, \boldsymbol{Z}(t), \boldsymbol{b}, C\}\Big]^{1-\mathrm{d}\mathbb{N}_T(t)} \mathrm{E}\{\mathrm{d}\mathbb{N}_T(t) \mid T \geq t, \boldsymbol{Z}(t), \boldsymbol{b}, C\}^{\mathrm{d}\mathbb{N}_T(t)} \\
&\times \Big[1 - \mathrm{E}\{\mathrm{d}\mathbb{N}_U(t) \mid T > t, U \geq t, \mathcal{H}_W(t), \mathbb{W}(t), \mathcal{H}_R(t), \mathcal{H}_{RY}(t)\}\Big]^{1-\mathrm{d}\mathbb{N}_U(t)} \\
&\times \mathrm{E}\{\mathrm{d}\mathbb{N}_U(t) \mid T > t, U \geq t, \mathcal{H}_W(t), \mathbb{W}(t), \mathcal{H}_R(t), \mathcal{H}_{RY}(t)\}^{\mathrm{d}\mathbb{N}_U(t)} \\
&\times f\{\mathrm{d}\mathbb{N}_R(t) \mid T > t, U > t, \mathcal{H}_W(t), \mathbb{W}(t), \mathcal{H}_R(t), \mathcal{H}_{RY}(t)\} \\
&\times \prod_{j=1}^{J} f\{\mathbb{Y}_j(t) \mid \mathbb{X}(t), \widetilde{\mathbb{X}}(t), \boldsymbol{b}, C\}^{\mathrm{d}\mathbb{N}_{R,j}(t)} \Bigg) \\
&\times \prod_{T \wedge U < t \leq \tau} \Bigg(f\{\mathbb{W}(t) \mid T \wedge U, \mathcal{H}_W(t), \mathcal{H}_R(t), \mathcal{H}_{RY}(t)\} \\
&\times f\{\mathrm{d}\mathbb{N}_R(t) \mid T \wedge U, \mathcal{H}_W(t), \mathbb{W}(t), \mathcal{H}_R(t), \mathcal{H}_{RY}(t)\} \\
&\times \prod_{j=1}^{J} f\{\mathbb{Y}_j(t) \mid \mathbb{X}(t), \widetilde{\mathbb{X}}(t), \boldsymbol{b}, C\}^{\mathrm{d}\mathbb{N}_{R,j}(t)} \Bigg) \, \mathrm{d}\boldsymbol{b}.
\end{aligned}
$$

Because the conditional densities of $\mathbb{W}(t)$, $\mathrm{d}\mathbb{N}_U(t)$, and $\mathrm{d}\mathbb{N}_R(t)$ do not depend on $(C, \boldsymbol{b})$, expression (4) in the main text is proportional to the likelihood function.

S3. Technical proofs.

PROOF OF LEMMA B.1. Consider $(\boldsymbol{\theta}_1, \Lambda_1, \mathcal{B}_1), (\boldsymbol{\theta}_2, \Lambda_2, \mathcal{B}_2) \in \Xi$. By the Taylor series expansion,

$$
\begin{aligned}
&\log \Psi(\boldsymbol{\theta}_1, \Lambda_1, \mathcal{B}_1) - \log \Psi(\boldsymbol{\theta}_2, \Lambda_2, \mathcal{B}_2) \\
&= \frac{\dot{\boldsymbol{\Psi}}_\theta(\widetilde{\boldsymbol{\theta}}, \widetilde{\Lambda}, \widetilde{\mathcal{B}})}{\Psi(\widetilde{\boldsymbol{\theta}}, \widetilde{\Lambda}, \widetilde{\mathcal{B}})}^{\mathrm{T}} (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) - \frac{\sum_{g=1}^{G} \pi_g \int_{\boldsymbol{b}} Q_g(\mathcal{O}, \boldsymbol{b}) \int_0^{\widetilde{T}} e^{\widetilde{\gamma}_g^{\mathrm{T}} \boldsymbol{Z}(t) + \widetilde{\boldsymbol{\eta}}_g^{\mathrm{T}} \boldsymbol{b} + \widetilde{\psi}_g(t)} \, \mathrm{d}(\Lambda_1 - \Lambda_2)(t) \, \mathrm{d}\boldsymbol{b}}{\sum_{g=1}^{G} \pi_g \int_{\boldsymbol{b}} Q_g(\mathcal{O}, \boldsymbol{b}) \, \mathrm{d}\boldsymbol{b}} \\
&\quad + \frac{\sum_{g=1}^{G} \pi_g \int_{\boldsymbol{b}} \Delta Q_g(\mathcal{O}, \boldsymbol{b}) (\psi_{g1} - \psi_{g2})(\widetilde{T}) \, \mathrm{d}\boldsymbol{b}}{\sum_{g=1}^{G} \pi_g \int_{\boldsymbol{b}} Q_g(\mathcal{O}, \boldsymbol{b}) \, \mathrm{d}\boldsymbol{b}} \\
&\quad - \frac{\sum_{g=1}^{G} \pi_g \int_{\boldsymbol{b}} Q_g(\mathcal{O}, \boldsymbol{b}) \int_0^{\widetilde{T}} e^{\widetilde{\gamma}_g^{\mathrm{T}} \boldsymbol{Z}(t) + \widetilde{\boldsymbol{\eta}}_g^{\mathrm{T}} \boldsymbol{b} + \widetilde{\psi}_g(t)} (\psi_{g1} - \psi_{g2})(t) \, \mathrm{d}\Lambda(t) \, \mathrm{d}\boldsymbol{b}}{\sum_{g=1}^{G} \pi_g \int_{\boldsymbol{b}} Q_g(\mathcal{O}, \boldsymbol{b}) \, \mathrm{d}\boldsymbol{b}}, \qquad \text{(S1)}
\end{aligned}
$$

where $(\widetilde{\boldsymbol\theta},\widetilde\Lambda,\widetilde{\mathcal B})\in\Xi$, and $Q_g(\mathcal O,\boldsymbol b)$ is evaluated at $(\widetilde{\boldsymbol\theta},\widetilde\Lambda,\widetilde{\mathcal B})$. By straightforward differentiation, we can show that

$$\left\|\frac{\dot{\boldsymbol\Psi}_\theta(\boldsymbol\theta,\Lambda,\mathcal B)}{\Psi(\boldsymbol\theta,\Lambda,\mathcal B)}\right\|\le\frac{\sum_{g=1}^G\int Q_g(\mathcal O,\boldsymbol b)(c_1+c_2\|\boldsymbol Y\|+c_3\|\boldsymbol Y\|^2+c_4\|\boldsymbol b\|+c_5e^{\boldsymbol\eta_g^{\mathrm T}\boldsymbol b}\|\boldsymbol b\|)\,\mathrm d\boldsymbol b}{\sum_{g=1}^G\int Q_g(\mathcal O,\boldsymbol b)\,\mathrm d\boldsymbol b}$$
$$\le\sum_{g=1}^G\frac{\int Q_g(\mathcal O,\boldsymbol b)(c_1+c_2\|\boldsymbol Y\|+c_3\|\boldsymbol Y\|^2+c_4\|\boldsymbol b\|+c_5e^{\boldsymbol\eta_g^{\mathrm T}\boldsymbol b}\|\boldsymbol b\|)\,\mathrm d\boldsymbol b}{\int Q_g(\mathcal O,\boldsymbol b)\,\mathrm d\boldsymbol b}$$

for some constants $c_1,\dots,c_5$. By Lemma S.1 of Wong, Zeng and Lin [3], each term in the summation of the right-hand side of the above inequality is $e^{O_p(1+\|\boldsymbol Y\|)}$. The second term on the right-hand side of (S1) is (up to a scaling factor) bounded above by

$$\left|\frac{\sum_{g=1}^G\pi_g\int_{\boldsymbol b}Q_g(\mathcal O,\boldsymbol b)e^{\widetilde{\boldsymbol\eta}_g^{\mathrm T}\boldsymbol b}\,\mathrm d\boldsymbol b\int_0^{\widetilde T}(\Lambda_1-\Lambda_2)(t)e^{\widetilde{\boldsymbol\gamma}_g^{\mathrm T}\boldsymbol Z(t)+\widetilde\psi_g(t)}\,\mathrm d|\widetilde{\boldsymbol\gamma}_g^{\mathrm T}\boldsymbol Z(t)+\widetilde\psi_g(t)|}{\sum_{g=1}^G\pi_g\int_{\boldsymbol b}Q_g(\mathcal O,\boldsymbol b)\,\mathrm d\boldsymbol b}\right|$$
$$+\left|\frac{\sum_{g=1}^G\pi_g\int_{\boldsymbol b}Q_g(\mathcal O,\boldsymbol b)e^{\widetilde{\boldsymbol\gamma}_g^{\mathrm T}\boldsymbol Z(\widetilde T)+\widetilde{\boldsymbol\eta}_g^{\mathrm T}\boldsymbol b+\widetilde\psi_g(\widetilde T)}\,\mathrm d\boldsymbol b}{\sum_{g=1}^G\pi_g\int_{\boldsymbol b}Q_g(\mathcal O,\boldsymbol b)\,\mathrm d\boldsymbol b}(\Lambda_1-\Lambda_2)(\widetilde T)\right|$$
$$\le e^{O_p(1+\|\boldsymbol Y\|)}\left\{\int_0^\tau|\Lambda_1-\Lambda_2|(t)\,\mathrm d\mu_1(t)+|\Lambda_1-\Lambda_2|(\widetilde T)\right\},$$

where $\mu_1(t)$ is some finite positive measure. By similar arguments, we conclude that

$$\begin{aligned}&\log\Psi(\boldsymbol\theta_1,\Lambda_1,\mathcal B_1)-\log\Psi(\boldsymbol\theta_2,\Lambda_2,\mathcal B_2)\\&=e^{O_p(1+\|\boldsymbol Y\|)}\Big\{\|\boldsymbol\theta_1-\boldsymbol\theta_2\|+\int_0^\tau|\Lambda_1-\Lambda_2|(t)\,\mathrm d\mu_1(t)+|\Lambda_1-\Lambda_2|(\widetilde T)\\&\quad+\sum_{g=2}^G\int_0^\tau|\psi_{g1}-\psi_{g2}|(t)\,\mathrm d\mu_2(t)+\sum_{g=2}^G|\psi_{g1}-\psi_{g2}|(\widetilde T)\Big\}\end{aligned}$$

for some finite positive measure $\mu_2(t)$, so that

$$\begin{aligned}\Big\|\log\Psi(\boldsymbol\theta_1,\Lambda_1,\mathcal B_1)-\log\Psi(\boldsymbol\theta_2,\Lambda_2,\mathcal B_2)\Big\|^2_{L_2(\mathbb P)}&\le O(1)\Big\{\|\boldsymbol\theta_1-\boldsymbol\theta_2\|^2+\int_0^\tau|\Lambda_1-\Lambda_2|(t)^2\,\mathrm d\mu_3(t)\\&\quad+\sum_{g=2}^G\int_0^\tau|\psi_{g1}-\psi_{g2}|(t)^2\,\mathrm d\mu_3(t)\Big\}\end{aligned}$$

for some finite positive measure $\mu_3(t)$. Note that $N(\epsilon,\Theta,\|\cdot\|)\lesssim(1/\epsilon)^d$, and by Theorem 2.7.5 of van der Vaart and Wellner [2], $N(\epsilon,\mathcal M_K,\|\cdot\|_{\mu_3})\lesssim e^{O(1/\epsilon)}$ and $N(\epsilon,\mathrm{BV}_K[0,\tau],\|\cdot\|_{\mu_3})\lesssim e^{O(1/\epsilon)}$, where $\|f\|^2_{\mu_3}=\int_0^\tau f^2\,\mathrm d\mu_3$. We conclude that

$$N_{[]}\{\epsilon,\mathcal G_1,L_2(\mathbb P)\}\lesssim\left(\frac1\epsilon\right)^de^{O(1/\epsilon)}.$$

By Theorem 2.5.6 of van der Vaart and Wellner [2], $\mathcal G_1$ is Donsker. Similarly, the classes $\mathcal G_2$, $\mathcal G_2$, and $\mathcal G_4$ can also be shown to be Donsker. □

PROOF OF LEMMA B.2. Suppose that there is a set of parameters $(\widetilde{\boldsymbol\theta},\widetilde\Lambda,\widetilde{\mathcal B})$ in a small neighborhood of the true values such that

$$\sum_{g=1}^G\widetilde\pi_g\int\{\widetilde\lambda_g(\widetilde T)e^{\widetilde{\boldsymbol\gamma}_g^{\mathrm T}\boldsymbol Z(\widetilde T)+\widetilde{\boldsymbol\eta}_g^{\mathrm T}\boldsymbol b}\}^\Delta\exp\Big\{-\int_0^{\widetilde T}e^{\widetilde{\boldsymbol\gamma}_g^{\mathrm T}\boldsymbol Z(s)+\widetilde{\boldsymbol\eta}_g^{\mathrm T}\boldsymbol b}\,\mathrm d\widetilde\Lambda_g(s)\Big\}\widetilde f_g(\boldsymbol Y,\boldsymbol b)\,\mathrm d\boldsymbol b$$

$$\text{(S2)}\qquad = \sum_{g=1}^{G} \pi_{0g} \int \big\{\lambda_{0g}(\widetilde{T}) e^{\boldsymbol{\gamma}_{0g}^{\mathrm{T}} \boldsymbol{Z}(\widetilde{T}) + \boldsymbol{\eta}_{0g}^{\mathrm{T}} \boldsymbol{b}}\big\}^{\Delta} \exp\Big\{ - \int_0^{\widetilde{T}} e^{\boldsymbol{\gamma}_{0g}^{\mathrm{T}} \boldsymbol{Z}(s) + \boldsymbol{\eta}_{0g}^{\mathrm{T}} \boldsymbol{b}} \,\mathrm{d}\Lambda_{0g}(s)\Big\} f_{0g}(\boldsymbol{Y}, \boldsymbol{b}) \,\mathrm{d}\boldsymbol{b},$$

where $\widetilde{\pi}_g$ and π_{0g} are the probabilities of membership of the gth latent group evaluated at $\widetilde{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$, respectively, and $\widetilde{f}_g$ and f_{0g} are the joint densities of $(\boldsymbol{Y}, \boldsymbol{b})$ of the gth latent group evaluated at $\widetilde{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$, respectively. Set $\Delta = 1$ and $\widetilde{T} = t$, then integrate t from 0 to τ on (S2). Adding the resulting equation to (S2) with $\Delta = 0$ and $\widetilde{T} = \tau$ yields

$$\sum_{g=1}^{G} \widetilde{\pi}_g \int \widetilde{f}_g(\boldsymbol{Y}, \boldsymbol{b}) \,\mathrm{d}\boldsymbol{b} = \sum_{g=1}^{G} \pi_{0g} \int f_{0g}(\boldsymbol{Y}, \boldsymbol{b}) \,\mathrm{d}\boldsymbol{b}.$$

Setting $N_1 = n_1, \ldots, N_K = n_K$ and $(\boldsymbol{X}, \widetilde{\boldsymbol{X}})$ to some value in $\mathcal{X}$, each side of the above equation is the density of a mixture normal distribution. By Proposition 2 of Yakowitz and Spragins [4] and linear independence of $\boldsymbol{W}$ and the rows of $\boldsymbol{X}$ and $\widetilde{\boldsymbol{X}}$, we conclude that $\widetilde{\boldsymbol{\alpha}}_g = \boldsymbol{\alpha}_{0g}$, $\widetilde{\boldsymbol{\beta}}_g = \boldsymbol{\beta}_{0g}$, $\widetilde{\boldsymbol{\xi}}_g = \boldsymbol{\xi}_{0g}$, and $\widetilde{\sigma}_g^2 = \sigma_{0g}^2$ for $g = 1, \ldots, G$.

Define $\boldsymbol{\Sigma}_{0g}$, $\boldsymbol{\Gamma}_{0g}$, and $\boldsymbol{\Sigma}_{0Yg}$ as in condition (C5). We have $\boldsymbol{b} \mid (\boldsymbol{Y}, C = g) \sim \mathrm{N}\{\boldsymbol{\Gamma}_{0g} \widetilde{\boldsymbol{X}}^{\mathrm{T}} \boldsymbol{\Sigma}_{0Yg}^{-1} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}_{0g}), \boldsymbol{\Gamma}_{0g}\}$ and $\boldsymbol{Y} \sim \mathrm{N}(\boldsymbol{X}\boldsymbol{\beta}_{0g}, \boldsymbol{\Sigma}_{0Yg})$. Setting $\Delta = 1$ and $\widetilde{T} = t$ and transforming the variable $\boldsymbol{b}$, (S2) becomes

$$\begin{aligned}
&\sum_{g=1}^{G} \pi_{0g} \int \widetilde{\lambda}_g(t) e^{\widetilde{\boldsymbol{\gamma}}_g^{\mathrm{T}} \boldsymbol{Z}(t) + \widetilde{\boldsymbol{\eta}}_g^{\mathrm{T}} \boldsymbol{\Gamma}_{0g}^{1/2} \boldsymbol{b} + \widetilde{\boldsymbol{\eta}}_g^{\mathrm{T}} \boldsymbol{\Gamma}_{0g} \widetilde{\boldsymbol{X}}^{\mathrm{T}} \boldsymbol{\Sigma}_{0Yg}^{-1} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}_{0g})} \\
&\qquad \times \exp\Big\{ - \int_0^t e^{\widetilde{\boldsymbol{\gamma}}_g^{\mathrm{T}} \boldsymbol{Z}(s) + \widetilde{\boldsymbol{\eta}}_g^{\mathrm{T}} \boldsymbol{\Gamma}_{0g}^{1/2} \boldsymbol{b} + \widetilde{\boldsymbol{\eta}}_g^{\mathrm{T}} \boldsymbol{\Gamma}_{0g} \widetilde{\boldsymbol{X}}^{\mathrm{T}} \boldsymbol{\Sigma}_{0Yg}^{-1} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}_{0g})} \,\mathrm{d}\widetilde{\Lambda}_g(s)\Big\} e^{-\frac{1}{2}\boldsymbol{b}^{\mathrm{T}}\boldsymbol{b}} \,\mathrm{d}\boldsymbol{b} \\
&\qquad \times \exp\Big\{ - \frac{1}{2} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}_{0g})^{\mathrm{T}} \boldsymbol{\Sigma}_{0Yg}^{-1} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}_{0g})\Big\} \\
&= \sum_{g=1}^{G} \pi_{0g} \int \lambda_{0g}(t) e^{\boldsymbol{\gamma}_{0g}^{\mathrm{T}} \boldsymbol{Z}(t) + \boldsymbol{\eta}_{0g}^{\mathrm{T}} \boldsymbol{\Gamma}_{0g}^{1/2} \boldsymbol{b} + \boldsymbol{\eta}_{0g}^{\mathrm{T}} \boldsymbol{\Gamma}_{0g} \widetilde{\boldsymbol{X}}^{\mathrm{T}} \boldsymbol{\Sigma}_{0Yg}^{-1} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}_{0g})} \\
&\qquad \times \exp\Big\{ - \int_0^t e^{\boldsymbol{\gamma}_{0g}^{\mathrm{T}} \boldsymbol{Z}(s) + \boldsymbol{\eta}_{0g}^{\mathrm{T}} \boldsymbol{\Gamma}_{0g}^{1/2} \boldsymbol{b} + \boldsymbol{\eta}_{0g}^{\mathrm{T}} \boldsymbol{\Gamma}_{0g} \widetilde{\boldsymbol{X}}^{\mathrm{T}} \boldsymbol{\Sigma}_{0Yg}^{-1} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}_{0g})} \,\mathrm{d}\Lambda_{0g}(s)\Big\} e^{-\frac{1}{2}\boldsymbol{b}^{\mathrm{T}}\boldsymbol{b}} \,\mathrm{d}\boldsymbol{b} \\
\text{(S3)}\quad &\qquad \times \exp\Big\{ - \frac{1}{2} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}_{0g})^{\mathrm{T}} \boldsymbol{\Sigma}_{0Yg}^{-1} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}_{0g})\Big\}.
\end{aligned}$$

We show that there exists some $l \in \{1, \ldots, G\}$ such that the lth terms of both sides of (S3) are equal. By the definition of $\mathcal{X}$, if $g \neq l$, then either $\boldsymbol{\Sigma}_{0Yg} \neq \boldsymbol{\Sigma}_{0Yl}$ or $\boldsymbol{\Sigma}_{0Yg}^{-1} \boldsymbol{X}\boldsymbol{\beta}_{0g} + \boldsymbol{\Sigma}_{0g}^{-1} \widetilde{\boldsymbol{X}} \boldsymbol{\Gamma}_{0g}^{\mathrm{T}} \boldsymbol{\eta}_{0g} \neq \boldsymbol{\Sigma}_{0Yl}^{-1} \boldsymbol{X}\boldsymbol{\beta}_{0l} + \boldsymbol{\Sigma}_{0l}^{-1} \widetilde{\boldsymbol{X}} \boldsymbol{\Gamma}_{0l}^{\mathrm{T}} \boldsymbol{\eta}_{0l}$. Define an ordering $\preccurlyeq$ on $\{1, \ldots, G\}$ as follows. We say that $l \preccurlyeq g$ if either $\boldsymbol{\Sigma}_{0Yl}^{-1}$ and $\boldsymbol{\Sigma}_{0Yg}^{-1}$ have the same set of eigenvalues or $\lambda_k(\boldsymbol{\Sigma}_{0Yl}^{-1}) < \lambda_k(\boldsymbol{\Sigma}_{0Yg}^{-1})$ with k being the smallest number such that $\lambda_k(\boldsymbol{\Sigma}_{0Yl}^{-1}) \neq \lambda_k(\boldsymbol{\Sigma}_{0Yg}^{-1})$, where $\lambda_j(\boldsymbol{A})$ is the jth smallest eigenvalue of a matrix $\boldsymbol{A}$. Find an l such that $l \preccurlyeq g$ for $g = 1, \ldots, G$, and let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p$ be the eigenvectors of $\boldsymbol{\Sigma}_{0Yl}^{-1}$ that correspond to the eigenvalues of $\boldsymbol{\Sigma}_{0Yl}^{-1}$ in ascending order. Let $\mathcal{L}_0 = \{g : \boldsymbol{\Sigma}_{0Yg} = \boldsymbol{\Sigma}_{0Yl}\}$ and $\mathcal{L}_0^{(-)} = \{g \in \mathcal{L}_0 : \boldsymbol{\eta}_{0g}^{\mathrm{T}} \boldsymbol{\Gamma}_{0g} \widetilde{\boldsymbol{X}}^{\mathrm{T}} \boldsymbol{\Sigma}_{0Yg}^{-1} \boldsymbol{u}_1 \leq 0\}$. If $\mathcal{L}_0^{(-)}$ is nonempty, then set $a_1 = 1$ and $\mathcal{L}_1$ as

$$\{g \in \mathcal{L}_0^{(-)} : \boldsymbol{u}_1^{\mathrm{T}} (\boldsymbol{\Sigma}_{0Yg}^{-1} \boldsymbol{X}\boldsymbol{\beta}_{0g} + \boldsymbol{\Sigma}_{0g}^{-1} \widetilde{\boldsymbol{X}} \boldsymbol{\Gamma}_{0g}^{\mathrm{T}} \boldsymbol{\eta}_{0g}) \geq \boldsymbol{u}_1^{\mathrm{T}} (\boldsymbol{\Sigma}_{0Yk}^{-1} \boldsymbol{X}\boldsymbol{\beta}_{0k} + \boldsymbol{\Sigma}_{0k}^{-1} \widetilde{\boldsymbol{X}} \boldsymbol{\Gamma}_{0k}^{\mathrm{T}} \boldsymbol{\eta}_{0k}) \text{ for } k \in \mathcal{L}_0^{(-)}\}.$$

Alternatively, if $\mathcal{L}_0^{(-)}$ is empty, then set $a_1 = -1$ and $\mathcal{L}_1$ as

$$\{g \in \mathcal{L}_0 : \boldsymbol{u}_1^{\mathrm{T}} (\boldsymbol{\Sigma}_{0Yg}^{-1} \boldsymbol{X}\boldsymbol{\beta}_{0g} + \boldsymbol{\Sigma}_{0g}^{-1} \widetilde{\boldsymbol{X}} \boldsymbol{\Gamma}_{0g}^{\mathrm{T}} \boldsymbol{\eta}_{0g}) \leq \boldsymbol{u}_1^{\mathrm{T}} (\boldsymbol{\Sigma}_{0Yk}^{-1} \boldsymbol{X}\boldsymbol{\beta}_{0k} + \boldsymbol{\Sigma}_{0k}^{-1} \widetilde{\boldsymbol{X}} \boldsymbol{\Gamma}_{0k}^{\mathrm{T}} \boldsymbol{\eta}_{0k}) \text{ for } k \in \mathcal{L}_0\}.$$

We define $\mathcal{L}_1^{(-)}, a_2, \mathcal{L}_2, \ldots, \mathcal{L}_{p-1}^{(-)}, a_p, \mathcal{L}_p$ in a similar manner. The set $\mathcal{L}_{k-1}^{(-)}$ contains all indices $g \in \mathcal{L}_{k-1}$ such that $\boldsymbol{\eta}_{0g}^{\mathrm{T}} \boldsymbol{\Gamma}_{0g} \widetilde{\boldsymbol{X}}^{\mathrm{T}} \boldsymbol{\Sigma}_{0Yg}^{-1} \boldsymbol{u}_k \leq 0$. The value a_k indicates, by the values -1 or 1, whether $\mathcal{L}_{k-1}^{(-)}$ is empty or not. If $a_k = 1$, then the set $\mathcal{L}_k$ contains all indices $g \in \mathcal{L}_{k-1}^{(-)}$ such that the value $\boldsymbol{u}_k^{\mathrm{T}}(\boldsymbol{\Sigma}_{0Yg}^{-1} \boldsymbol{X} \boldsymbol{\beta}_{0g} + \boldsymbol{\Sigma}_{0g}^{-1} \widetilde{\boldsymbol{X}} \boldsymbol{\Gamma}_{0g}^{\mathrm{T}} \boldsymbol{\eta}_{0g})$ is maximized. Alternatively, if $a_k = -1$, then the set $\mathcal{L}_k$ contains all indices $g \in \mathcal{L}_{k-1}$ such that the value $\boldsymbol{u}_k^{\mathrm{T}}(\boldsymbol{\Sigma}_{0Yg}^{-1} \boldsymbol{X} \boldsymbol{\beta}_{0g} + \boldsymbol{\Sigma}_{0g}^{-1} \widetilde{\boldsymbol{X}} \boldsymbol{\Gamma}_{0g}^{\mathrm{T}} \boldsymbol{\eta}_{0g})$ is minimized. Note that by the definition of $\mathcal{X}$, the values of $(\boldsymbol{\Sigma}_{0Yg}^{-1} \boldsymbol{X} \boldsymbol{\beta}_{0g} + \boldsymbol{\Sigma}_{0g}^{-1} \widetilde{\boldsymbol{X}} \boldsymbol{\Gamma}_{0g}^{\mathrm{T}} \boldsymbol{\eta}_{0g})$ for $g \in \mathcal{L}_0$ are distinct, so that $\mathcal{L}_p$ contains one and only one element; assume without loss of generality that $\mathcal{L}_p = \{1\}$. Let

$$\boldsymbol{y} = c_1 a_1 \boldsymbol{u}_1 + \cdots + c_p a_p \boldsymbol{u}_p,$$

where c_1 is some positive constant, and $c_j = c_{j-1}/\log(c_1)$ for $j = 2, \ldots, p$.

We divide both sides of (S3) by

$$\exp\Big\{ -\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_{01})^{\mathrm{T}} \boldsymbol{\Sigma}_{0Y1}^{-1} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_{01}) + \boldsymbol{\eta}_{01}^{\mathrm{T}} \boldsymbol{\Gamma}_{01} \widetilde{\boldsymbol{X}}^{\mathrm{T}} \boldsymbol{\Sigma}_{0Y1}^{-1} \boldsymbol{y} \Big\}.$$

By the construction of $\mathcal{L}_p$ and $\boldsymbol{y}$, all but the first term in the resulting summation of the right-hand side of (S3) are $o_p(1)$ as $c_1 \to \infty$. Note that when $\widetilde{\boldsymbol{\eta}}_g$ is in a small enough neighborhood of $\boldsymbol{\eta}_{0g}$ ($g = 2, \ldots, G$), all but the first term in the resulting summation of the left-hand side of (S3) are $o_p(1)$ as well. The resulting equation is

$$\begin{aligned}
&\int \widetilde{\lambda}_1(t) e^{\widetilde{\boldsymbol{\gamma}}_1^{\mathrm{T}} \boldsymbol{Z}(t) + \widetilde{\boldsymbol{\eta}}_1^{\mathrm{T}} \boldsymbol{\Gamma}_{01}^{1/2} \boldsymbol{b} + \widetilde{\boldsymbol{\eta}}_1^{\mathrm{T}} \boldsymbol{\Gamma}_{01} \widetilde{\boldsymbol{X}}^{\mathrm{T}} \boldsymbol{\Sigma}_{0Y1}^{-1} \boldsymbol{X}\boldsymbol{\beta}_{01} + (\widetilde{\boldsymbol{\eta}}_1 - \boldsymbol{\eta}_{01})^{\mathrm{T}} \boldsymbol{\Gamma}_{01} \widetilde{\boldsymbol{X}}^{\mathrm{T}} \boldsymbol{\Sigma}_{0Y1}^{-1} \boldsymbol{y}} \\
&\quad \times \exp\Big\{ -\int_0^t e^{\widetilde{\boldsymbol{\gamma}}_1^{\mathrm{T}} \boldsymbol{Z}(s) + \widetilde{\boldsymbol{\eta}}_1^{\mathrm{T}} \boldsymbol{\Gamma}_{01}^{1/2} \boldsymbol{b} + \widetilde{\boldsymbol{\eta}}_1^{\mathrm{T}} \boldsymbol{\Gamma}_{01} \widetilde{\boldsymbol{X}}^{\mathrm{T}} \boldsymbol{\Sigma}_{0Y1}^{-1} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_{01})} \, \mathrm{d}\widetilde{\Lambda}_1(s) \Big\} e^{-\frac{1}{2}\boldsymbol{b}^{\mathrm{T}}\boldsymbol{b}} \, \mathrm{d}\boldsymbol{b} + o_p(1) \\
= &\int \lambda_{01}(t) e^{\boldsymbol{\gamma}_{01}^{\mathrm{T}} \boldsymbol{Z}(t) + \boldsymbol{\eta}_{01}^{\mathrm{T}} \boldsymbol{\Gamma}_{01}^{1/2} \boldsymbol{b} + \boldsymbol{\eta}_{01}^{\mathrm{T}} \boldsymbol{\Gamma}_{01} \widetilde{\boldsymbol{X}}^{\mathrm{T}} \boldsymbol{\Sigma}_{0Y1}^{-1} \boldsymbol{X}\boldsymbol{\beta}_{01}} \\
&\quad \times \exp\Big\{ -\int_0^t e^{\boldsymbol{\gamma}_{01}^{\mathrm{T}} \boldsymbol{Z}(s) + \boldsymbol{\eta}_{01}^{\mathrm{T}} \boldsymbol{\Gamma}_{01}^{1/2} \boldsymbol{b} + \boldsymbol{\eta}_{01}^{\mathrm{T}} \boldsymbol{\Gamma}_{01} \widetilde{\boldsymbol{X}}^{\mathrm{T}} \boldsymbol{\Sigma}_{0Y1}^{-1} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_{01})} \, \mathrm{d}\Lambda_{01}(s) \Big\} e^{-\frac{1}{2}\boldsymbol{b}^{\mathrm{T}}\boldsymbol{b}} \, \mathrm{d}\boldsymbol{b} + o_p(1).
\end{aligned}$$

Note that by the construction of $\boldsymbol{y}$, $\boldsymbol{\eta}_{01}^{\mathrm{T}} \boldsymbol{\Gamma}_{01} \widetilde{\boldsymbol{X}}^{\mathrm{T}} \boldsymbol{\Sigma}_{0Y1}^{-1} \boldsymbol{y} \leq 0$, so that the right-hand side of the above equation is bounded away from 0. Comparing the two sides of the above equation, we conclude that $\widetilde{\boldsymbol{\eta}}_1 = \boldsymbol{\eta}_{01}$, $\boldsymbol{Z}(t)^{\mathrm{T}} \widetilde{\boldsymbol{\gamma}}_1 = \boldsymbol{Z}(t)^{\mathrm{T}} \boldsymbol{\gamma}_{01}$, and $\widetilde{\lambda}_1(t) = \lambda_{01}(t)$ for $t \in [0, \tau]$. Repeating the above arguments on the remaining terms of (S3) yields $\widetilde{\boldsymbol{\eta}}_g = \boldsymbol{\eta}_{0g}$, $\boldsymbol{Z}(t)^{\mathrm{T}} \widetilde{\boldsymbol{\gamma}}_g = \boldsymbol{Z}(t)^{\mathrm{T}} \boldsymbol{\gamma}_{0g}$, and $\widetilde{\lambda}_g(t) = \lambda_{0g}(t)$ for $t \in [0, \tau]$ and $g = 2, \ldots, G$. Finally, repeating the above arguments with different values of $(\boldsymbol{X}, \widetilde{\boldsymbol{X}}) \in \mathcal{X}$ and by the linear independence on $\boldsymbol{Z}(\cdot)$ on $\mathcal{X}$, we conclude that $\widetilde{\boldsymbol{\gamma}}_g = \boldsymbol{\gamma}_{0g}$ for $g = 1, \ldots, G$. □

PROOF OF LEMMA B.3. Consider a submodel indexed by ϵ given by

$$\boldsymbol{\alpha}_g = \boldsymbol{\alpha}_{0g} + \epsilon \boldsymbol{h}_{\alpha g}, \ \boldsymbol{\mu}_g = \boldsymbol{\mu}_{0g} + \epsilon \boldsymbol{h}_{\mu g}, \text{ and } \boldsymbol{\Omega}_g = \boldsymbol{\Omega}_{0g} + \epsilon \boldsymbol{H}_g$$

for $g = 1, \ldots, G$, where $\boldsymbol{\alpha}_{0g}$ is the true value of $\boldsymbol{\alpha}_g$. Let π_g denote $P(C = g)$ and f_g denote the density of $\boldsymbol{Y}$ given $C = g$; the parameters and covariates are suppressed. If the score statistic at $\epsilon = 0$ is zero, then

$$\begin{aligned}
&\sum_{g=1}^G \pi_g f_g(\boldsymbol{Y}) \Big[\boldsymbol{W}^{\mathrm{T}} \boldsymbol{h}_{\alpha g} \Big\{ 1 - \sum_{l=1}^G \pi_l \frac{f_l(\boldsymbol{Y})}{f_g(\boldsymbol{Y})} \Big\} + (\boldsymbol{Y} - \boldsymbol{\mu}_{0g})^{\mathrm{T}} \boldsymbol{\Omega}_{0g}^{-1} \boldsymbol{h}_{\mu g} - \frac{1}{2} \mathrm{tr}(\boldsymbol{\Omega}_{0g}^{-1} \boldsymbol{H}_g) \\
&\qquad + \frac{1}{2} (\boldsymbol{Y} - \boldsymbol{\mu}_{0g})^{\mathrm{T}} \boldsymbol{\Omega}_{0g}^{-1} \boldsymbol{H}_g \boldsymbol{\Omega}_{0g}^{-1} (\boldsymbol{Y} - \boldsymbol{\mu}_{0g}) \Big] = 0 \qquad \text{(S4)}
\end{aligned}$$

for almost surely all $(\boldsymbol{W},\boldsymbol{X},\boldsymbol{Y})$. We find a value l such that $f_g(\boldsymbol{y})/f_l(\boldsymbol{y})\to 0$ as $\boldsymbol{y}\to\infty$ (along some direction) for $g\neq l$. To show that such a choice of l exists, we define an ordering such that $l\preccurlyeq g$ if either $\boldsymbol{\Omega}_{0l}^{-1}$ and $\boldsymbol{\Omega}_{0g}^{-1}$ have the same set of eigenvalues or $\lambda_k(\boldsymbol{\Omega}_{0l}^{-1})<\lambda_k(\boldsymbol{\Omega}_{0g}^{-1})$ with k being the smallest number such that $\lambda_k(\boldsymbol{\Omega}_{0l}^{-1})\neq\lambda_k(\boldsymbol{\Omega}_{0g}^{-1})$. Find an l such that $l\preccurlyeq g$ for $g=1,\ldots,G$, and let $\boldsymbol{u}_1,\ldots,\boldsymbol{u}_p$ be the eigenvectors of $\boldsymbol{\Omega}_{0l}^{-1}$ that correspond to $\lambda_1(\boldsymbol{\Omega}_{0l}^{-1}),\ldots,\lambda_p(\boldsymbol{\Omega}_{0l}^{-1})$, where p is the dimension of $\boldsymbol{Y}$. By the uniqueness of $(\boldsymbol{\mu}_{0g},\boldsymbol{\Omega}_{0g})$, we can assume without loss of generality that if $\boldsymbol{\Omega}_{0g}=\boldsymbol{\Omega}_{0l}$, then there exists some $k\in\{1,\ldots,G\}$ such that $\boldsymbol{\mu}_{0l}^{\mathrm{T}}\boldsymbol{\Omega}_{0l}^{-1}\boldsymbol{u}_k>\boldsymbol{\mu}_{0g}^{\mathrm{T}}\boldsymbol{\Omega}_{0g}^{-1}\boldsymbol{u}_k$ and $\boldsymbol{\mu}_{0l}^{\mathrm{T}}\boldsymbol{\Omega}_{0l}^{-1}\boldsymbol{u}_j=\boldsymbol{\mu}_{0g}^{\mathrm{T}}\boldsymbol{\Omega}_{0g}^{-1}\boldsymbol{u}_j$ for $j<k$. Let $\boldsymbol{y}=c_1\boldsymbol{u}_1+\cdots+c_p\boldsymbol{u}_p$, where c_1 is some positive constant, and $c_j=c_{j-1}/\log(c_1)$ for $j=2,\ldots,p$. We can see that $f_g(\boldsymbol{y})/f_l(\boldsymbol{y})\to 0$ as $c_1\to\infty$ for $g\neq l$.

Without loss of generality, assume that $l=1$. Dividing both sides of (S4) by $f_1(\boldsymbol{Y})$ and setting $\boldsymbol{Y}=\boldsymbol{y}$, we have

$$\boldsymbol{W}^{\mathrm{T}}\boldsymbol{h}_{\alpha 1}(1-\pi_1)+(\boldsymbol{y}-\boldsymbol{\mu}_{01})^{\mathrm{T}}\boldsymbol{\Omega}_{01}^{-1}\boldsymbol{h}_{\mu 1}-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Omega}_{01}^{-1}\boldsymbol{H}_1)$$

$$+\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu}_{01})^{\mathrm{T}}\boldsymbol{\Omega}_{01}^{-1}\boldsymbol{H}_1\boldsymbol{\Omega}_{01}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_{01})-\sum_{g=2}^{G}\pi_g\boldsymbol{W}^{\mathrm{T}}\boldsymbol{h}_{\alpha g}=o(1),$$

where the right-hand side vanishes as $c_1\to\infty$. By considering the coefficients of $c_1^2,\ldots,c_p^2,c_1,\ldots,c_p$ in turn, we conclude that $\boldsymbol{H}_1=\boldsymbol{0}$ and $\boldsymbol{h}_{\mu 1}=\boldsymbol{0}$. Therefore,

$$\boldsymbol{W}^{\mathrm{T}}\boldsymbol{h}_{\alpha 1}=\frac{1}{1-\pi_1}\sum_{g=2}^{G}\pi_g\boldsymbol{W}^{\mathrm{T}}\boldsymbol{h}_{\alpha g}.$$

Plugging these results into (S4), we have

$$\sum_{g=2}^{G}\frac{\pi_g}{1-\pi_1}f_g(\boldsymbol{Y})\Big[\boldsymbol{W}^{\mathrm{T}}\boldsymbol{h}_{\alpha g}\Big\{1-\sum_{l=2}^{G}\frac{\pi_l}{1-\pi_1}\frac{f_l(\boldsymbol{Y})}{f_g(\boldsymbol{Y})}\Big\}+(\boldsymbol{Y}-\boldsymbol{\mu}_{0g})^{\mathrm{T}}\boldsymbol{\Omega}_{0g}^{-1}\boldsymbol{h}_{\mu g}-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Omega}_{0g}^{-1}\boldsymbol{H}_g)$$

$$+\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{\mu}_{0g})^{\mathrm{T}}\boldsymbol{\Omega}_{0g}^{-1}\boldsymbol{H}_g\boldsymbol{\Omega}_{0g}^{-1}(\boldsymbol{Y}-\boldsymbol{\mu}_{0g})\Big]=0.$$

Iteratively applying the above arguments yields $\boldsymbol{H}_g=\boldsymbol{0}$, $\boldsymbol{h}_{\mu g}=\boldsymbol{0}$, and

$$\boldsymbol{W}^{\mathrm{T}}\Big(\boldsymbol{h}_{\alpha g}-\sum_{l=g}^{G}\frac{\pi_l}{\sum_{k=g}^{G}\pi_k}\boldsymbol{h}_{\alpha l}\Big)=0$$

for $g=1,\ldots,G$. By linear independence of $\boldsymbol{W}$, we have $\boldsymbol{h}_{\alpha g}=\boldsymbol{0}$ for $g=1,\ldots,G-1$.

□

PROOF OF LEMMA B.4. We consider the one-dimensional submodel:

$$\boldsymbol{\alpha}_g=\boldsymbol{\alpha}_{0g}+\epsilon\boldsymbol{h}_{\alpha g},\ \boldsymbol{\beta}_g=\boldsymbol{\beta}_{0g}+\epsilon\boldsymbol{h}_{\beta g},\ \sigma_g^2=\sigma_{0g}^2+\epsilon h_{\sigma g},\ \boldsymbol{\xi}_g=\boldsymbol{\xi}_{0g}+\epsilon\boldsymbol{h}_{\xi g},$$

$$\boldsymbol{\gamma}_g=\boldsymbol{\gamma}_{0g}+\epsilon\boldsymbol{h}_{\gamma g},\ \eta_g=\eta_{0g}+\epsilon h_{\eta g},\ \text{and}\ \Lambda_g=\Lambda_{0g}+\epsilon\int h_{\Lambda g}(s)\,\mathrm{d}\Lambda_{0g}(s)$$

for $g=1,\ldots,G$; note that $\boldsymbol{\alpha}_{0G}=\boldsymbol{h}_{\alpha G}=\boldsymbol{0}$. Let

$$\widetilde{f}_g(\widetilde{T},\Delta,\boldsymbol{Y},\boldsymbol{b})=\{\lambda_{0g}(\widetilde{T})e^{\boldsymbol{Z}(\widetilde{T})^{\mathrm{T}}\boldsymbol{\gamma}_{0g}+\boldsymbol{\eta}_{0g}^{\mathrm{T}}\boldsymbol{b}}\}^{\Delta}\exp\Big\{-e^{\boldsymbol{\eta}_{0g}^{\mathrm{T}}\boldsymbol{b}}\int_0^{\widetilde{T}}e^{\boldsymbol{Z}(s)^{\mathrm{T}}\boldsymbol{\gamma}_{0g}}\,\mathrm{d}\Lambda_{0g}(s)\Big\}f_g(\boldsymbol{Y},\boldsymbol{b}).$$

Setting the score statistic along this submodel to zero, we have

$$\sum_{g=1}^{G-1}\pi_{0g}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{h}_{\alpha g}\Big\{\int \widetilde{f}_g(\widetilde{T},\Delta,\boldsymbol{Y},\boldsymbol{b})\,\mathrm{d}\boldsymbol{b}-\sum_{l=1}^{G}\pi_{0l}\int \widetilde{f}_l(\widetilde{T},\Delta,\boldsymbol{Y},\boldsymbol{b})\,\mathrm{d}\boldsymbol{b}\Big\}$$
$$+\sum_{g=1}^{G}\pi_{0g}\int \widetilde{f}_g(\widetilde{T},\Delta,\boldsymbol{Y},\boldsymbol{b})\Big[\Delta\{\boldsymbol{Z}(\widetilde{T})^{\mathrm{T}}\boldsymbol{h}_{\gamma g}+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{h}_{\eta g}+h_{\Lambda g}(\widetilde{T})\}$$
$$-\int_0^{\widetilde{T}} e^{\boldsymbol{Z}(s)^{\mathrm{T}}\boldsymbol{\gamma}_{0g}+\boldsymbol{\eta}_{0g}^{\mathrm{T}}\boldsymbol{b}}\{\boldsymbol{Z}(s)^{\mathrm{T}}\boldsymbol{h}_{\gamma g}+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{h}_{\eta g}+h_{\Lambda g}(s)\}\,\mathrm{d}\Lambda_g(s)+\frac{\boldsymbol{f}_g^{(1)}(\boldsymbol{Y},\boldsymbol{b})^{\mathrm{T}}\boldsymbol{h}_{Yg}}{f_g(\boldsymbol{Y},\boldsymbol{b})}\Big]\,\mathrm{d}\boldsymbol{b}=0 \tag{S5}$$

almost surely. We show that if (S5) holds for all possible values of the observed data, then $\boldsymbol{h}_{Yg}=\mathbf{0}$, $\boldsymbol{h}_{\alpha g}=\mathbf{0}$, and $h_{\Lambda g}(t)=0$ for $t\in[0,\tau]$. Set $\Delta=1$ and $\widetilde{T}=t$, then integrate t from 0 to τ on (S5). Adding the resulting equation to (S5) with $\Delta=0$ and $\widetilde{T}=\tau$ yields

$$\sum_{g=1}^{G-1}\pi_{0g}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{h}_{\alpha g}\Big\{\int f_g(\boldsymbol{Y},\boldsymbol{b})\,\mathrm{d}\boldsymbol{b}-\sum_{l=1}^{G}\pi_{0l}\int f_l(\boldsymbol{Y},\boldsymbol{b})\,\mathrm{d}\boldsymbol{b}\Big\}+\sum_{g=1}^{G}\pi_{0g}\int f_g(\boldsymbol{Y},\boldsymbol{b})^{\mathrm{T}}\boldsymbol{h}_{Yg}\,\mathrm{d}\boldsymbol{b}=0.$$

At $N_1=n_1,\ldots,N_K=n_K$ and any fixed $(\boldsymbol{X},\widetilde{\boldsymbol{X}})\in\mathcal{X}$, the left-hand side of the above equation is the score statistic for a mixture normal model along a one-dimensional submodel. By Lemma B.3 and the linear independence of $\boldsymbol{W}$ and the rows of $\boldsymbol{X}$ and $\widetilde{\boldsymbol{X}}$, the equation implies that $\boldsymbol{h}_{\alpha g}=\mathbf{0}$ and $\boldsymbol{h}_{Yg}=\mathbf{0}$ for $g=1,\ldots,G$.

Set $N_1=n_1,\ldots,N_K=n_K$ and $(\boldsymbol{X},\widetilde{\boldsymbol{X}})$ to some value in $\mathcal{X}$. Setting $\boldsymbol{h}_{\alpha g}$ and $\boldsymbol{h}_{Yg}$ to zero, setting $\Delta=1$ and $\widetilde{T}=t$, and transforming the variable $\boldsymbol{b}$, (S5) becomes

$$\sum_{g=1}^{G}\pi_{0g}\int \lambda_{0g}(t)e^{\boldsymbol{Z}(t)^{\mathrm{T}}\boldsymbol{\gamma}_{0g}+\boldsymbol{\eta}_{0g}^{\mathrm{T}}\boldsymbol{\Gamma}_{0g}^{1/2}\boldsymbol{b}+\boldsymbol{\eta}_{0g}^{\mathrm{T}}\boldsymbol{\Gamma}_{0g}\widetilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{\Sigma}_{0Yg}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}_{0g})}$$
$$\times\exp\Big\{-\int_0^t e^{\boldsymbol{Z}(s)^{\mathrm{T}}\boldsymbol{\gamma}_{0g}+\boldsymbol{\eta}_{0g}^{\mathrm{T}}\boldsymbol{\Gamma}_{0g}^{1/2}\boldsymbol{b}+\boldsymbol{\eta}_{0g}^{\mathrm{T}}\boldsymbol{\Gamma}_{0g}\widetilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{\Sigma}_{0Yg}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}_{0g})}\,\mathrm{d}\Lambda_{0g}(s)\Big\}$$
$$\times e^{-\frac{1}{2}\boldsymbol{b}^{\mathrm{T}}\boldsymbol{b}}\Big[\boldsymbol{Z}(t)^{\mathrm{T}}\boldsymbol{h}_{\gamma g}+\boldsymbol{h}_{\eta g}^{\mathrm{T}}\boldsymbol{\Gamma}_{0g}^{1/2}\boldsymbol{b}+\boldsymbol{h}_{\eta g}^{\mathrm{T}}\boldsymbol{\Gamma}_{0g}\widetilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{\Sigma}_{0Yg}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}_{0g})+h_{\Lambda g}(t)$$
$$-\int_0^t e^{\boldsymbol{Z}(s)^{\mathrm{T}}\boldsymbol{\gamma}_{0g}+\boldsymbol{\eta}_{0g}^{\mathrm{T}}\boldsymbol{\Gamma}_{0g}^{1/2}\boldsymbol{b}+\boldsymbol{\eta}_{0g}^{\mathrm{T}}\boldsymbol{\Gamma}_{0g}\widetilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{\Sigma}_{0Yg}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}_{0g})}$$
$$\times\Big\{\boldsymbol{Z}(s)^{\mathrm{T}}\boldsymbol{h}_{\gamma g}+\boldsymbol{h}_{\eta g}^{\mathrm{T}}\boldsymbol{\Gamma}_{0g}^{1/2}\boldsymbol{b}+\boldsymbol{h}_{\eta g}^{\mathrm{T}}\boldsymbol{\Gamma}_{0g}\widetilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{\Sigma}_{0Yg}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}_{0g})+h_{\Lambda g}(s)\Big\}\,\mathrm{d}\Lambda_{0g}(s)\Big]\,\mathrm{d}\boldsymbol{b}$$
$$\times\exp\Big\{-\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}_{0g})^{\mathrm{T}}\boldsymbol{\Sigma}_{0Yg}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}_{0g})\Big\}=0. \tag{S6}$$

Next, we divide both sides of the above equation by

$$\exp\Big\{-\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}_{0l})^{\mathrm{T}}\boldsymbol{\Sigma}_{0Yl}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}_{0l})+\boldsymbol{\eta}_{0l}^{\mathrm{T}}\boldsymbol{\Gamma}_{0l}\widetilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{\Sigma}_{0Yl}^{-1}\boldsymbol{Y}\Big\} \tag{S7}$$

for some $l=1,\ldots,G$, such that each term in the summation in (S6) except for $g=l$ vanishes as we set $\boldsymbol{Y}$ to infinity along some direction. We can use the argument in the proof of Lemma B.2 to show that such an l exists. In particular, define $\mathcal{L}_p$, $(a_1,\ldots,a_p)$, and $(\boldsymbol{u}_1,\ldots,\boldsymbol{u}_p)$ as in the proof of Lemma B.2, and assume without loss of generality that $\mathcal{L}_p=\{1\}$. Let

$$\boldsymbol{y}=c_1a_1\boldsymbol{u}_1+\cdots+c_pa_p\boldsymbol{u}_p,$$

where c_1 is some positive constant, and $c_j = c_{j-1}/\log(c_1)$ for $j = 2, \ldots, p$. Dividing both sides of (S6) by (S7) with $l = 1$, each term in the resulting summation for $g > 1$ tends to 0 as $c_1 \to \infty$. If $\boldsymbol{\eta}_{01} \neq \mathbf{0}$, then (S6) with both sides divided by (S7) becomes

$$\int e^{\boldsymbol{\eta}_{01}^{\mathrm{T}}\boldsymbol{\Gamma}_{01}^{1/2}\boldsymbol{b}-\frac{1}{2}\boldsymbol{b}^{\mathrm{T}}\boldsymbol{b}}\big\{\boldsymbol{Z}(t)^{\mathrm{T}}\boldsymbol{h}_{\gamma 1}+\boldsymbol{h}_{\eta 1}^{\mathrm{T}}\boldsymbol{\Gamma}_{01}^{1/2}\boldsymbol{b}+\boldsymbol{h}_{\eta 1}^{\mathrm{T}}\boldsymbol{\Gamma}_{01}\widetilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{\Sigma}_{0Y1}^{-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}_{01})+h_{\Lambda 1}(t)\big\}\,\mathrm{d}\boldsymbol{b}=o(1)$$

as $c_1 \to \infty$. Therefore, $\boldsymbol{h}_{\eta 1} = \mathbf{0}$, $\boldsymbol{Z}(t)^{\mathrm{T}}\boldsymbol{h}_{\gamma 1} + h_{\Lambda 1}(t) = 0$ for $t \in [0, \tau]$. Alternatively, if $\boldsymbol{\eta}_{01} = \mathbf{0}$, then (S6) with both sides divided by (S7) becomes

$$\int e^{-\frac{1}{2}\boldsymbol{b}^{\mathrm{T}}\boldsymbol{b}}\Big[\boldsymbol{Z}(t)^{\mathrm{T}}\boldsymbol{h}_{\gamma 1}+\boldsymbol{h}_{\eta 1}^{\mathrm{T}}\boldsymbol{\Gamma}_{01}^{1/2}\boldsymbol{b}+\boldsymbol{h}_{\eta 1}^{\mathrm{T}}\boldsymbol{\Gamma}_{01}\widetilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{\Sigma}_{0Y1}^{-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}_{01})+h_{\Lambda 1}(t)-\int_0^t e^{\boldsymbol{Z}(s)^{\mathrm{T}}\boldsymbol{\gamma}_{01}}$$
$$\times\{\boldsymbol{Z}(s)^{\mathrm{T}}\boldsymbol{h}_{\gamma 1}+\boldsymbol{h}_{\eta 1}^{\mathrm{T}}\boldsymbol{\Gamma}_{01}^{1/2}\boldsymbol{b}+\boldsymbol{h}_{\eta 1}^{\mathrm{T}}\boldsymbol{\Gamma}_{01}\widetilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{\Sigma}_{0Y1}^{-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}_{01})+h_{\Lambda 1}(s)\}\,\mathrm{d}\Lambda_{01}(s)\Big]\,\mathrm{d}\boldsymbol{b}=o(1).$$

Likewise, we can see that $\boldsymbol{h}_{\eta 1} = \mathbf{0}$, and the resulting equation is a homogeneous integral equation for $\boldsymbol{Z}(t)^{\mathrm{T}}\boldsymbol{h}_{\gamma 1} + h_{\Lambda 1}(t)$. We conclude that $\boldsymbol{Z}(t)^{\mathrm{T}}\boldsymbol{h}_{\gamma 1} + h_{\Lambda 1}(t) = 0$ for $t \in [0, \tau]$. Applying similar arguments on the remaining terms in (S6), we have $\boldsymbol{h}_{\eta g} = \mathbf{0}$ and $\boldsymbol{Z}(t)^{\mathrm{T}}\boldsymbol{h}_{\gamma 1} + h_{\Lambda g}(t) = 0$ for $t \in [0, \tau]$ for $g = 2, \ldots, G$. Finally, repeating the above arguments with different values of $(\boldsymbol{X}, \widetilde{\boldsymbol{X}}) \in \mathcal{X}$ and by the linear independence of $\boldsymbol{Z}(\cdot)$, we conclude that $\boldsymbol{h}_{\gamma g} = \mathbf{0}$ and $h_{\Lambda g}(t) = 0$ for $t \in [0, \tau]$ and $g = 1, \ldots, G$. □

PROOF OF THEOREM 4.3. We prove that for any $\boldsymbol{v} \in \mathbb{R}^d$ such that $\|\boldsymbol{v}\| \leq 1$, $\boldsymbol{v}^{\mathrm{T}}\widetilde{\boldsymbol{I}}\boldsymbol{v} - \boldsymbol{v}^{\mathrm{T}}\widehat{\boldsymbol{V}}_n^{-1}\boldsymbol{v} \to_p 0$. Let $\mathcal{H} = \{(h_\Lambda, h_{\psi 2}, \ldots, h_{\psi G}) : h_{\psi 2}, \ldots, h_{\psi G} \in C^2[0, \tau],\ \|h_\Lambda\|_{\mathrm{V}} + \sum_{g=2}^{G} \|h''_{\psi g}\|_2 < K\}$ for some large enough constant K. We can write $\boldsymbol{v}^{\mathrm{T}}\widetilde{\boldsymbol{I}}\boldsymbol{v}$ as

$$\min_{(h_\Lambda,h_{\psi 2},\ldots,h_{\psi G})\in\mathcal{H}}\mathbb{P}\Big\{\boldsymbol{v}^{\mathrm{T}}\dot{\boldsymbol{\ell}}_\theta(\boldsymbol{\theta}_0,\Lambda_0,\mathcal{B}_0)-\dot{\boldsymbol{\ell}}_\Lambda(\boldsymbol{\theta}_0,\Lambda_0,\mathcal{B}_0)[\int h_\Lambda\,\mathrm{d}\Lambda_0]-\sum_{g=2}^{G}\dot{\boldsymbol{\ell}}_\psi(\boldsymbol{\theta}_0,\Lambda_0,\mathcal{B}_0)[h_{\psi g}]\Big\}^2.$$

For any $(h_\Lambda, h_{\psi 2}, \ldots, h_{\psi G}) \in \mathcal{H}$,

$$\mathbb{P}\Big\{\boldsymbol{v}^{\mathrm{T}}\dot{\boldsymbol{\ell}}_\theta(\boldsymbol{\theta}_0,\Lambda_0,\mathcal{B}_0)-\dot{\ell}_\Lambda(\boldsymbol{\theta}_0,\Lambda_0,\mathcal{B}_0)[\int h_\Lambda\,\mathrm{d}\Lambda_0]-\sum_{g=2}^{G}\dot{\ell}_\psi(\boldsymbol{\theta}_0,\Lambda_0,\mathcal{B}_0)[h_{\psi g}]\Big\}^2$$
$$=-\,\mathbb{P}\ddot{\ell}_{(\theta\Lambda\psi,\theta\Lambda\psi)}(\boldsymbol{\theta}_0,\Lambda_0,\mathcal{B}_0)\big[\boldsymbol{v},-\int h_\Lambda\,\mathrm{d}\Lambda_0,-h_{\psi 2},\ldots,-h_{\psi G}\big]$$
$$=-\,\mathbb{P}\ddot{\ell}_{(\theta\Lambda\psi,\theta\Lambda\psi)}(\boldsymbol{\theta}_0,\Lambda_0,\mathcal{B}_0)\big[\boldsymbol{v},-\int h_\Lambda\,\mathrm{d}\Lambda_0,-\check{h}_{\psi 2},\ldots,-\check{h}_{\psi G}\big]+o(1)$$
$$=-\,\mathbb{P}_n\ddot{\ell}_{(\theta\Lambda\psi,\theta\Lambda\psi)}(\widehat{\boldsymbol{\theta}}_n,\widehat{\Lambda}_n,\widehat{\mathcal{B}}_n)\big[\boldsymbol{v},-\int h_\Lambda\,\mathrm{d}\widehat{\Lambda}_n,-\check{h}_{\psi 2},\ldots,-\check{h}_{\psi G}\big]+o_p(1),\tag{S8}$$

where $\check{h}_{\psi 2}, \ldots, \check{h}_{\psi G}$ are the approximations of $h_{\psi 2}, \ldots, h_{\psi G}$ based on the B-spline functions $(B_1, \ldots, B_{m_n})$ such that $\|\check{h}_{\psi g} - h_{\psi g}\|_\infty = O(m_n^{-3})$ for $g = 2, \ldots, G$, the $o_p(1)$ term denotes a variable $r_n(\boldsymbol{v}, h_\Lambda, h_{\psi 2}, \ldots, h_{\psi G})$ such that $\sup_{\|\boldsymbol{v}\|\leq 1,(h_\Lambda,h_{\psi 2},\ldots,h_{\psi G})\in\mathcal{H}} |r_n| \to_p 0$, and $\ddot{\ell}_{(\theta\Lambda\psi,\theta\Lambda\psi)}$ is the second derivative of the log-likelihood such that

$$\ddot{\ell}_{(\theta\Lambda\psi,\theta\Lambda\psi)}(\boldsymbol{\theta},\Lambda,\mathcal{B})[\boldsymbol{v},\int h_\Lambda\,\mathrm{d}\Lambda,h_{\psi 2},\ldots,h_{\psi G}]$$
$$=\frac{\partial}{\partial\epsilon}\Big\{\boldsymbol{v}^{\mathrm{T}}\dot{\boldsymbol{\ell}}_\theta\big(\boldsymbol{\theta}+\epsilon\boldsymbol{v},\Lambda+\epsilon\int h_\Lambda\,\mathrm{d}\Lambda,\mathcal{B}+\epsilon\boldsymbol{h}_\psi\big)$$
$$+\,\dot{\ell}_\Lambda\big(\boldsymbol{\theta}+\epsilon\boldsymbol{v},\Lambda+\epsilon\int h_\Lambda\,\mathrm{d}\Lambda,\mathcal{B}+\epsilon\boldsymbol{h}_\psi\big)[\int h_\Lambda\,\mathrm{d}\Lambda]$$

$$+\sum_{g=2}^{G} \dot{\ell}_{\psi}\left(\boldsymbol{\theta}+\epsilon \boldsymbol{v}, \Lambda+\epsilon \int h_{\Lambda} \,\mathrm{d}\Lambda, \mathcal{B}+\epsilon \boldsymbol{h}_{\psi}\right)[h_{\psi g}]\Big\}\Big|_{\epsilon=0}$$

with $\boldsymbol{h}_{\psi}=(h_{\psi 2}, \ldots, h_{\psi G})$. The second equality of (S8) follows from the dominated convergence theorem, and the third equality follows from the Glivenko–Cantelli property of the class of $\ddot{\ell}_{(\theta\Lambda\psi,\theta\Lambda\psi)}(\boldsymbol{\theta}, \Lambda, \mathcal{B})[\boldsymbol{v}, \int h_{\Lambda} \,\mathrm{d}\Lambda, h_{\psi 2}, \ldots, h_{\psi G}]$, which can be established using arguments in the proof of Lemma B.1. Note that the right-hand side of (S8) depends on h_{Λ} only through $\{h_{\Lambda}(\widetilde{T}_j): \Delta_j=1\}$, and denote these values by $(h_{\Lambda 1}, \ldots, h_{\Lambda k_n})$, where k_n is the number of observed event times. Let $\boldsymbol{q}=(-h_{\Lambda j} \widehat{\Lambda}\{t_j\})_{j=1,\ldots,k_n}$, where t_j is the jth observed event time, and $\boldsymbol{\alpha}$ be the vector of the negative of the coefficients of the B-spline functions in $\check{h}_{\psi 2}, \ldots \check{h}_{\psi G}$. The first term on the right-hand side of (S8) is equal to $-n^{-1}(\boldsymbol{v}^{\mathrm{T}}\ \boldsymbol{q}^{\mathrm{T}}\ \boldsymbol{\alpha}^{\mathrm{T}})\boldsymbol{I}_n(\boldsymbol{v}^{\mathrm{T}}\ \boldsymbol{q}^{\mathrm{T}}\ \boldsymbol{\alpha}^{\mathrm{T}})^{\mathrm{T}}$, so

$$\boldsymbol{v}^{\mathrm{T}} \widetilde{\boldsymbol{I}} \boldsymbol{v}=\min_{\boldsymbol{q}, \boldsymbol{\alpha}}\left(\boldsymbol{v}^{\mathrm{T}}\ \boldsymbol{q}^{\mathrm{T}}\ \boldsymbol{\alpha}^{\mathrm{T}}\right)\left(\frac{1}{n} \boldsymbol{I}_n\right)\begin{pmatrix}\boldsymbol{v} \\ \boldsymbol{q} \\ \boldsymbol{\alpha}\end{pmatrix}+o_p(1).$$

Note that $(\boldsymbol{v}^{\mathrm{T}}\ \boldsymbol{q}^{\mathrm{T}}\ \boldsymbol{\alpha}^{\mathrm{T}})(n^{-1}\boldsymbol{I}_n)(\boldsymbol{v}^{\mathrm{T}}\ \boldsymbol{q}^{\mathrm{T}}\ \boldsymbol{\alpha}^{\mathrm{T}})^{\mathrm{T}}$ is asymptotically equal to the expected value of the squared score statistic along some direction, so it is strictly larger than 0. Therefore, $\boldsymbol{I}_n$ is positive definite for large enough n. By simple matrix algebra, the first term on the right-hand side of the above equation is $\boldsymbol{v}^{\mathrm{T}} \widehat{\boldsymbol{V}}_n^{-1} \boldsymbol{v}$. Because $\widetilde{\boldsymbol{I}}$ is positive definite, the desired result follows. □

S4. Simulation studies for misspecified longitudinal outcome models. We generated simulation data sets as described in Section 5. Instead of fitting the correct model, we mimicked Liu et al. [1] and included only a single random effect. In particular, for the longitudinal outcomes, we fit the model

$$Y_{jk}\mid_{C=g}=\boldsymbol{\beta}_{g1}^{\mathrm{T}} \boldsymbol{X}_k+\boldsymbol{\beta}_{g2}^{\mathrm{T}} \boldsymbol{X}_k I(j=2)+b+\epsilon_{jk},$$

where $b \mid (C=g) \sim \mathrm{N}(0, \xi_g^2)$ and $\epsilon_{jk} \mid (C=g) \sim \mathrm{N}(0, \sigma_g^2)$ for $j=1,2$, $g=1,\ldots,G$, and $k=1,\ldots,10$. Note that the error variances of the two types of longitudinal outcomes are (correctly) assumed to be equal. We fit model (1) of the main text for the latent class membership and model (3) of the main text with a single random effect b for the event time. The fitted model is the same as the true model except that a single random effect is used to account for the association among the longitudinal outcome measurements and the event time.

We used the B-spline functions as described in Section 5 and fixed $G=3$. We set the sample size to be $n=1000$ or 2000 and considered 1000 simulation replicates. Because the parameters in the longitudinal outcome models may not be comparable with those in the true model, we only present the estimation results for the parameters in the multinomial model and the survival model in Tables S6–S7. Due to the inability to correctly identify the latent class structure, some parameter estimators have substantial bias and serious undercoverage.

REFERENCES

[1] LIU, Y., LIN, Y., ZHOU, J. and LIU, L. (2020). A semi-parametric joint latent class model with longitudinal and survival data. *Stat. Interface* **13** 411–422. MR4091806 https://dx.doi.org/10.4310/SII.2020.v13.n3.a10

[2] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer.

[3] WONG, K. Y., ZENG, D. and LIN, D. Y. (2018). Efficient estimation for semiparametric structural equation models with censored data. *J. Amer. Statist. Assoc.* **113** 893–905. MR3832235 https://doi.org/10.1080/01621459.2017.1299626

[4] YAKOWITZ, S. J. and SPRAGINS, J. D. (1968). On the identifiability of finite mixtures. *Ann. Math. Statist.* **39** 209–214. https://doi.org/10.1214/aoms/1177698520

[5] ZENG, D. and CAI, J. (2005). Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time. *Ann. Statist.* **33** 2132–2163. MR2211082 https://doi.org/10.1214/009053605000000480

TABLE S1
Simulation results under the correct model and $n = 1000$

Param	True	Bias	SE	SEE	CP	Param	True	Bias	SE	SEE	CP
α_{11}	−0.250	0.006	0.155	0.151	0.95	σ^2_{21}	1.000	−0.001	0.051	0.049	0.94
α_{12}	−0.250	0.003	0.226	0.215	0.94	σ^2_{22}	1.000	0.001	0.051	0.049	0.94
α_{13}	0.500	0.002	0.215	0.204	0.94	σ^2_{31}	1.500	−0.002	0.073	0.069	0.93
α_{21}	0.500	0.005	0.287	0.264	0.94	σ^2_{32}	1.500	−0.003	0.068	0.069	0.96
α_{22}	−0.500	0.005	0.106	0.105	0.95	ξ^2_{11}	1.000	−0.005	0.132	0.131	0.94
α_{23}	0.500	0.011	0.130	0.122	0.94	ξ^2_{12}	1.000	−0.005	0.133	0.129	0.94
β_{111}	−1.000	−0.001	0.122	0.120	0.95	ξ^2_{13}	0.500	−0.009	0.099	0.098	0.94
β_{112}	−1.000	0.000	0.039	0.040	0.96	ξ^2_{21}	1.000	−0.025	0.203	0.191	0.93
β_{113}	−1.000	−0.003	0.150	0.149	0.95	ξ^2_{22}	1.000	−0.013	0.198	0.192	0.94
β_{114}	−1.000	0.001	0.079	0.079	0.95	ξ^2_{23}	1.000	−0.009	0.198	0.183	0.93
β_{121}	1.000	0.004	0.118	0.119	0.95	ξ^2_{31}	1.000	−0.024	0.229	0.221	0.93
β_{122}	1.000	−0.001	0.040	0.040	0.95	ξ^2_{32}	1.000	−0.022	0.226	0.218	0.94
β_{123}	0.000	−0.002	0.140	0.148	0.97	ξ^2_{33}	1.500	−0.032	0.242	0.236	0.93
β_{124}	0.000	0.003	0.077	0.078	0.96	γ_{11}	−0.100	−0.014	0.158	0.152	0.95
β_{211}	0.000	0.008	0.191	0.178	0.95	γ_{12}	0.000	0.005	0.151	0.150	0.95
β_{212}	1.000	0.001	0.096	0.091	0.94	γ_{21}	0.100	0.000	0.174	0.174	0.95
β_{213}	0.000	0.002	0.220	0.210	0.94	γ_{22}	0.100	−0.001	0.083	0.081	0.94
β_{214}	1.000	−0.005	0.123	0.114	0.94	γ_{31}	0.000	0.005	0.075	0.075	0.95
β_{221}	0.000	−0.001	0.186	0.178	0.94	γ_{32}	−0.100	−0.005	0.090	0.087	0.94
β_{222}	−1.000	−0.001	0.095	0.091	0.95	η_1	0.100	0.004	0.169	0.168	0.96
β_{223}	−1.000	0.007	0.246	0.218	0.94	η_2	−0.100	−0.003	0.111	0.103	0.95
β_{224}	1.000	0.001	0.123	0.114	0.94	η_3	−0.300	−0.019	0.107	0.096	0.95
β_{311}	1.000	0.004	0.181	0.165	0.94	$\Lambda_1(0.3)$	0.150	0.000	0.033	0.032	0.93
β_{312}	0.000	0.000	0.117	0.108	0.93	$\Lambda_1(0.7)$	0.350	0.000	0.057	0.055	0.94
β_{313}	1.000	−0.006	0.275	0.261	0.94	$\Lambda_1(1.4)$	0.700	0.007	0.096	0.091	0.93
β_{314}	−1.000	0.003	0.158	0.136	0.92	$\Lambda_2(0.3)$	0.312	−0.003	0.053	0.053	0.95
β_{321}	−1.000	−0.002	0.171	0.166	0.93	$\Lambda_2(0.7)$	0.765	−0.001	0.106	0.105	0.95
β_{322}	0.000	0.004	0.109	0.108	0.95	$\Lambda_2(1.4)$	1.680	0.008	0.216	0.213	0.94
β_{323}	1.000	0.009	0.270	0.256	0.94	$\Lambda_3(0.3)$	0.300	−0.006	0.047	0.048	0.94
β_{324}	−1.000	−0.001	0.150	0.134	0.93	$\Lambda_3(0.7)$	0.700	−0.007	0.088	0.090	0.96
σ^2_{11}	0.500	−0.001	0.016	0.017	0.95	$\Lambda_3(1.4)$	1.400	0.006	0.156	0.165	0.96
σ^2_{12}	0.500	0.000	0.017	0.017	0.96						

NOTE: "Param" is the parameter to be estimated; "True" is the true parameter value; "Bias" is the empirical bias; "SE" is the empirical standard error; "SEE" is the empirical mean of the standard error estimator; and "CP" is the empirical coverage probability of the 95% confidence interval. Eight replicates have been removed due to extreme estimates.

TABLE S2
Simulation results under the correct model and $n = 2000$

Param	True	Bias	SE	SEE	CP	Param	True	Bias	SE	SEE	CP
α_{11}	−0.250	0.006	0.109	0.105	0.94	σ^2_{21}	1.000	−0.002	0.036	0.035	0.94
α_{12}	−0.250	0.006	0.156	0.149	0.94	σ^2_{22}	1.000	0.000	0.035	0.035	0.95
α_{13}	0.500	0.005	0.148	0.143	0.95	σ^2_{31}	1.500	0.000	0.049	0.049	0.94
α_{21}	0.500	0.003	0.184	0.184	0.96	σ^2_{32}	1.500	0.000	0.048	0.049	0.95
α_{22}	−0.500	0.004	0.074	0.074	0.95	ξ^2_{11}	1.000	−0.002	0.095	0.093	0.94
α_{23}	0.500	0.005	0.085	0.085	0.94	ξ^2_{12}	1.000	−0.001	0.092	0.092	0.95
β_{111}	−1.000	−0.003	0.085	0.085	0.95	ξ^2_{13}	0.500	−0.004	0.069	0.070	0.95
β_{112}	−1.000	0.000	0.028	0.028	0.95	ξ^2_{21}	1.000	−0.011	0.142	0.135	0.94
β_{113}	−1.000	0.000	0.105	0.105	0.95	ξ^2_{22}	1.000	−0.005	0.134	0.135	0.95
β_{114}	−1.000	0.000	0.057	0.056	0.95	ξ^2_{23}	1.000	−0.005	0.136	0.129	0.93
β_{121}	1.000	0.001	0.081	0.084	0.96	ξ^2_{31}	1.000	−0.007	0.162	0.157	0.94
β_{122}	1.000	0.001	0.028	0.028	0.95	ξ^2_{32}	1.000	−0.019	0.160	0.154	0.93
β_{123}	0.000	−0.002	0.101	0.105	0.96	ξ^2_{33}	1.500	−0.013	0.165	0.168	0.95
β_{124}	0.000	0.001	0.053	0.055	0.96	γ_{11}	−0.100	−0.010	0.104	0.106	0.95
β_{211}	0.000	0.001	0.122	0.125	0.95	γ_{12}	0.000	0.002	0.104	0.104	0.96
β_{212}	1.000	−0.002	0.065	0.064	0.94	γ_{21}	0.100	−0.002	0.124	0.121	0.95
β_{213}	0.000	0.002	0.149	0.148	0.95	γ_{22}	0.100	−0.002	0.057	0.056	0.95
β_{214}	1.000	0.000	0.079	0.080	0.94	γ_{31}	0.000	0.002	0.050	0.052	0.96
β_{221}	0.000	−0.004	0.128	0.125	0.94	γ_{32}	−0.100	−0.004	0.060	0.060	0.95
β_{222}	−1.000	0.001	0.065	0.064	0.95	η_1	0.100	0.007	0.111	0.114	0.96
β_{223}	−1.000	0.002	0.156	0.153	0.94	η_2	−0.100	−0.003	0.071	0.070	0.95
β_{224}	1.000	0.005	0.083	0.080	0.93	η_3	−0.300	−0.010	0.070	0.065	0.94
β_{311}	1.000	0.006	0.116	0.116	0.96	$\Lambda_1(0.3)$	0.150	0.000	0.023	0.023	0.95
β_{312}	0.000	0.001	0.079	0.076	0.94	$\Lambda_1(0.7)$	0.350	0.001	0.039	0.038	0.94
β_{313}	1.000	−0.006	0.185	0.184	0.95	$\Lambda_1(1.4)$	0.700	0.006	0.065	0.063	0.95
β_{314}	−1.000	−0.001	0.098	0.096	0.95	$\Lambda_2(0.3)$	0.312	−0.001	0.036	0.037	0.96
β_{321}	−1.000	0.001	0.114	0.116	0.96	$\Lambda_2(0.7)$	0.765	0.001	0.073	0.073	0.96
β_{322}	0.000	0.003	0.077	0.077	0.95	$\Lambda_2(1.4)$	1.680	0.007	0.148	0.148	0.96
β_{323}	1.000	0.004	0.180	0.180	0.95	$\Lambda_3(0.3)$	0.300	−0.002	0.034	0.034	0.95
β_{324}	−1.000	−0.003	0.094	0.094	0.96	$\Lambda_3(0.7)$	0.700	−0.002	0.062	0.063	0.95
σ^2_{11}	0.500	0.000	0.012	0.012	0.96	$\Lambda_3(1.4)$	1.400	0.007	0.112	0.114	0.96
σ^2_{12}	0.500	0.000	0.012	0.012	0.96						

NOTE: See NOTE to Table S1 for interpretations of the column names.

TABLE S3

Estimation results for parameters in the multinomial model for the ARIC data

Parameter	Estimate	SE	p-value	Parameter	Estimate	SE	p-value
$\alpha_{1,\mathrm{Int}}$	1.3659	0.3086	9.59E−06	$\alpha_{2,\mathrm{Smoke}}$	0.3319	0.2149	1.22E−01
$\alpha_{1,\mathrm{Center}}$	−0.2549	0.2883	3.77E−01	$\alpha_{2,\mathrm{Sex}}$	0.1183	0.2232	5.96E−01
$\alpha_{1,\mathrm{BMI}}$	−0.0169	0.0963	8.61E−01	$\alpha_{2,\mathrm{Age}}$	0.3118	0.0996	1.74E−03
$\alpha_{1,\mathrm{Glucose}}$	−0.4576	0.1065	1.75E−05	$\alpha_{3,\mathrm{Int}}$	0.7514	0.3719	4.33E−02
$\alpha_{1,\mathrm{Smoke}}$	−0.2122	0.2044	2.99E−01	$\alpha_{3,\mathrm{Center}}$	−0.0626	0.3377	8.53E−01
$\alpha_{1,\mathrm{Sex}}$	−0.3396	0.2023	9.33E−02	$\alpha_{3,\mathrm{BMI}}$	0.2211	0.0914	1.55E−02
$\alpha_{1,\mathrm{Age}}$	−0.1541	0.0916	9.24E−02	$\alpha_{3,\mathrm{Glucose}}$	−0.2109	0.0584	3.03E−04
$\alpha_{2,\mathrm{Int}}$	0.4061	0.3694	2.72E−01	$\alpha_{3,\mathrm{Smoke}}$	−0.3412	0.2302	1.38E−01
$\alpha_{2,\mathrm{Center}}$	0.0574	0.3363	8.65E−01	$\alpha_{3,\mathrm{Sex}}$	0.5665	0.2136	7.99E−03
$\alpha_{2,\mathrm{BMI}}$	0.1652	0.1037	1.11E−01	$\alpha_{3,\mathrm{Age}}$	0.1668	0.0960	8.23E−02
$\alpha_{2,\mathrm{Glucose}}$	−0.1394	0.0494	4.79E−03				

NOTE: For each parameter, the first subscript represents the latent class, and the second subscript represents the covariate that corresponds to the parameter. “Int” represents the intercept. See NOTE to Table 1 for the labels of the other covariates.

TABLE S4

Estimation results for parameters in the longitudinal outcome and latent-variable models for the ARIC data

Parameter	Estimate	SE	p-value	Parameter	Estimate	SE	p-value
$\beta_{1,\mathrm{Systolic,Int}}$	−0.0014	0.0781	9.86E−01	$\beta_{3,\mathrm{Chol,Center}}$	−0.1570	0.1117	1.60E−01
$\beta_{1,\mathrm{Systolic,Time}}$	0.0397	0.0022	1.53E−75	$\beta_{3,\mathrm{Chol,BMI}}$	−0.0958	0.0313	2.24E−03
$\beta_{1,\mathrm{Systolic,Center}}$	0.0892	0.0785	2.56E−01	$\beta_{3,\mathrm{Chol,Glucose}}$	−0.0084	0.0324	7.96E−01
$\beta_{1,\mathrm{Systolic,BMI}}$	0.1042	0.0295	4.11E−04	$\beta_{3,\mathrm{Chol,Smoke}}$	−0.0433	0.0868	6.18E−01
$\beta_{1,\mathrm{Systolic,Glucose}}$	0.0650	0.0603	2.81E−01	$\beta_{3,\mathrm{Chol,Sex}}$	−0.4833	0.0631	1.86E−14
$\beta_{1,\mathrm{Systolic,Smoke}}$	−0.0202	0.0628	7.48E−01	$\beta_{3,\mathrm{Chol,Age}}$	−0.0066	0.0296	8.23E−01
$\beta_{1,\mathrm{Systolic,Sex}}$	0.0787	0.0667	2.38E−01	$\beta_{4,\mathrm{Systolic,Int}}$	0.3396	0.1810	6.07E−02
$\beta_{1,\mathrm{Systolic,Age}}$	0.2366	0.0338	2.55E−12	$\beta_{4,\mathrm{Systolic,Time}}$	0.0198	0.0057	5.42E−04
$\beta_{1,\mathrm{Chol,Int}}$	−0.5915	0.0859	5.73E−12	$\beta_{4,\mathrm{Systolic,Center}}$	0.1012	0.1704	5.53E−01
$\beta_{1,\mathrm{Chol,Time}}$	0.0028	0.0017	1.02E−01	$\beta_{4,\mathrm{Systolic,BMI}}$	0.0653	0.0487	1.80E−01
$\beta_{1,\mathrm{Chol,Center}}$	0.0800	0.0812	3.25E−01	$\beta_{4,\mathrm{Systolic,Glucose}}$	−0.0011	0.0286	9.69E−01
$\beta_{1,\mathrm{Chol,BMI}}$	−0.0657	0.0246	7.60E−03	$\beta_{4,\mathrm{Systolic,Smoke}}$	−0.0413	0.1143	7.18E−01
$\beta_{1,\mathrm{Chol,Glucose}}$	−0.0794	0.0428	6.38E−02	$\beta_{4,\mathrm{Systolic,Sex}}$	−0.0640	0.1034	5.36E−01
$\beta_{1,\mathrm{Chol,Smoke}}$	−0.0620	0.0628	3.23E−01	$\beta_{4,\mathrm{Systolic,Age}}$	0.1355	0.0519	9.07E−03
$\beta_{1,\mathrm{Chol,Sex}}$	−0.3289	0.0625	1.45E−07	$\beta_{4,\mathrm{Chol,Int}}$	0.7522	0.2195	6.11E−04
$\beta_{1,\mathrm{Chol,Age}}$	0.0617	0.0280	2.76E−02	$\beta_{4,\mathrm{Chol,Time}}$	−0.0525	0.0063	4.60E−17
$\beta_{2,\mathrm{Systolic,Int}}$	1.2243	0.1668	2.13E−13	$\beta_{4,\mathrm{Chol,Center}}$	0.3062	0.212	1.49E−01
$\beta_{2,\mathrm{Systolic,Time}}$	0.0231	0.0064	3.03E−04	$\beta_{4,\mathrm{Chol,BMI}}$	−0.0106	0.0611	8.62E−01
$\beta_{2,\mathrm{Systolic,Center}}$	0.0157	0.1497	9.17E−01	$\beta_{4,\mathrm{Chol,Glucose}}$	0.0335	0.0314	2.85E−01
$\beta_{2,\mathrm{Systolic,BMI}}$	0.0664	0.0443	1.34E−01	$\beta_{4,\mathrm{Chol,Smoke}}$	−0.2579	0.1277	4.34E−02
$\beta_{2,\mathrm{Systolic,Glucose}}$	0.0725	0.0308	1.85E−02	$\beta_{4,\mathrm{Chol,Sex}}$	−0.1583	0.1203	1.89E−01
$\beta_{2,\mathrm{Systolic,Smoke}}$	−0.0125	0.1011	9.01E−01	$\beta_{4,\mathrm{Chol,Age}}$	0.1695	0.0632	7.27E−03
$\beta_{2,\mathrm{Systolic,Sex}}$	0.1420	0.0991	1.52E−01	$\sigma^2_{1,\mathrm{Systolic}}$	0.3418	0.0216	
$\beta_{2,\mathrm{Systolic,Age}}$	0.1246	0.0480	9.38E−03	$\sigma^2_{1,\mathrm{Chol}}$	0.1374	0.0066	
$\beta_{2,\mathrm{Chol,Int}}$	−0.0312	0.1400	8.24E−01	$\sigma^2_{2,\mathrm{Systolic}}$	1.4248	0.0691	
$\beta_{2,\mathrm{Chol,Time}}$	−0.0313	0.0048	5.00E−11	$\sigma^2_{2,\mathrm{Chol}}$	0.2599	0.0181	
$\beta_{2,\mathrm{Chol,Center}}$	0.0866	0.1224	4.79E−01	$\sigma^2_{3,\mathrm{Systolic}}$	0.2938	0.0178	
$\beta_{2,\mathrm{Chol,BMI}}$	0.0119	0.0379	7.54E−01	$\sigma^2_{3,\mathrm{Chol}}$	0.2467	0.0156	
$\beta_{2,\mathrm{Chol,Glucose}}$	0.0108	0.0277	6.97E−01	$\sigma^2_{4,\mathrm{Systolic}}$	0.6404	0.0489	
$\beta_{2,\mathrm{Chol,Smoke}}$	−0.3363	0.0871	1.12E−04	$\sigma^2_{4,\mathrm{Chol}}$	1.2111	0.0885	
$\beta_{2,\mathrm{Chol,Sex}}$	−0.1595	0.0834	5.60E−02	$\xi^2_{1,\mathrm{Systolic}}$	0.2528	0.0338	
$\beta_{2,\mathrm{Chol,Age}}$	0.0113	0.0388	7.70E−01	$\xi^2_{1,\mathrm{Chol}}$	0.3182	0.0476	
$\beta_{3,\mathrm{Systolic,Int}}$	−0.1568	0.0802	5.06E−02	$\xi^2_{1,\mathrm{Common}}$	0.0517	0.0267	
$\beta_{3,\mathrm{Systolic,Time}}$	0.0233	0.0023	1.72E−23	$\xi^2_{2,\mathrm{Systolic}}$	0.2953	0.0631	
$\beta_{3,\mathrm{Systolic,Center}}$	0.0813	0.0751	2.79E−01	$\xi^2_{2,\mathrm{Chol}}$	0.4900	0.0593	
$\beta_{3,\mathrm{Systolic,BMI}}$	0.1082	0.0254	2.09E−05	$\xi^2_{2,\mathrm{Common}}$	0.1304	0.0390	
$\beta_{3,\mathrm{Systolic,Glucose}}$	0.0058	0.0215	7.87E−01	$\xi^2_{3,\mathrm{Systolic}}$	0.2104	0.0300	
$\beta_{3,\mathrm{Systolic,Smoke}}$	−0.0784	0.0637	2.18E−01	$\xi^2_{3,\mathrm{Chol}}$	0.3427	0.0333	
$\beta_{3,\mathrm{Systolic,Sex}}$	0.0531	0.0516	3.03E−01	$\xi^2_{3,\mathrm{Common}}$	0.0130	0.0152	
$\beta_{3,\mathrm{Systolic,Age}}$	0.1007	0.0239	2.51E−05	$\xi^2_{4,\mathrm{Systolic}}$	0.2630	0.0582	
$\beta_{3,\mathrm{Chol,Int}}$	0.8089	0.1088	1.03E−13	$\xi^2_{4,\mathrm{Chol}}$	0.3197	0.0701	
$\beta_{3,\mathrm{Chol,Time}}$	−0.0657	0.0029	9.43E−116	$\xi^2_{4,\mathrm{Common}}$	0.0647	0.0389	

NOTE: For all parameters, the first subscript represents the latent class. For the parameters labeled β and σ^2, the second subscript represents the type of longitudinal outcome, and the third subscript (if available) represents the covariate that corresponds to the parameter. The parameters $\xi^2_{g,\mathrm{Systolic}}$, $\xi^2_{g,\mathrm{Chol}}$, and $\xi^2_{g,\mathrm{Common}}$ correspond respectively to the variances of b_1, b_2, and b_3 for the gth latent class. “Time” represents the examination time in years; “Systolic” represents systolic blood pressure; “Chol” represents total cholesterol level. See NOTEs to Table 1 and Table S3 for the labels of the other covariates.

TABLE S5
Simulation results for NPMLE

Param	True	Bias ($n = 1000$)	Bias ($n = 2000$)
μ_1	0.500	−0.139	−0.141
σ_1^2	0.500	−0.163	−0.161
β_1	0.500	−0.101	−0.105
$\Lambda_1(0.5)$	0.250	0.113	0.114
$\Lambda_1(1.0)$	0.500	0.209	0.210
$\Lambda_1(1.5)$	0.750	0.290	0.289
μ_2	1.000	0.113	0.113
σ_2^2	1.000	0.004	0.004
β_2	0.500	0.036	0.032
$\Lambda_2(0.5)$	0.533	−0.047	−0.048
$\Lambda_2(1.0)$	1.136	−0.108	−0.110
$\Lambda_2(1.5)$	1.820	−0.191	−0.193

NOTE: See NOTE to Table S1 for interpretations of the column names.

TABLE S6
Simulation results under a univariate random effect and $n = 1000$

Param	True	Bias	SE	SEE	CP	Param	True	Bias	SE	SEE	CP
α_{11}	−0.250	−0.212	0.368	0.143	0.62	η_1	0.100	−0.045	0.092	0.089	0.91
α_{12}	−0.250	0.227	0.392	0.152	0.44	η_2	−0.100	0.043	0.060	0.048	0.84
α_{13}	0.500	−0.215	0.317	0.195	0.68	η_3	−0.300	0.209	0.043	0.043	0.00
α_{21}	0.500	−0.511	0.435	0.217	0.40	$\Lambda_1(0.3)$	0.150	0.003	0.037	0.034	0.93
α_{22}	−0.500	0.030	0.150	0.098	0.79	$\Lambda_1(0.7)$	0.350	0.002	0.061	0.058	0.94
α_{23}	0.500	−0.426	0.286	0.101	0.15	$\Lambda_1(1.4)$	0.700	0.002	0.101	0.096	0.94
γ_{11}	−0.100	−0.049	0.172	0.168	0.95	$\Lambda_2(0.3)$	0.312	−0.038	0.045	0.043	0.83
γ_{12}	0.000	0.068	0.170	0.138	0.88	$\Lambda_2(0.7)$	0.765	−0.105	0.087	0.076	0.68
γ_{21}	0.100	−0.078	0.175	0.136	0.88	$\Lambda_2(1.4)$	1.680	−0.322	0.181	0.140	0.39
γ_{22}	0.100	0.021	0.094	0.091	0.94	$\Lambda_3(0.3)$	0.300	−0.001	0.042	0.042	0.95
γ_{31}	0.000	0.096	0.099	0.070	0.69	$\Lambda_3(0.7)$	0.700	0.002	0.077	0.075	0.95
γ_{32}	−0.100	0.091	0.088	0.068	0.69	$\Lambda_3(1.4)$	1.400	0.029	0.138	0.140	0.95

NOTE: See NOTE to Table S1 for interpretations of the column names.

TABLE S7
Simulation results under a univariate random effect and $n = 2000$

Param	True	Bias	SE	SEE	CP	Param	True	Bias	SE	SEE	CP
α_{11}	−0.250	−0.230	0.373	0.103	0.55	η_1	0.100	−0.045	0.064	0.063	0.89
α_{12}	−0.250	0.205	0.384	0.109	0.27	η_2	−0.100	0.049	0.050	0.034	0.69
α_{13}	0.500	−0.212	0.234	0.139	0.59	η_3	−0.300	0.211	0.031	0.030	0.00
α_{21}	0.500	−0.491	0.335	0.155	0.28	$\Lambda_1(0.3)$	0.150	0.003	0.026	0.024	0.93
α_{22}	−0.500	0.019	0.124	0.070	0.75	$\Lambda_1(0.7)$	0.350	0.002	0.043	0.041	0.93
α_{23}	0.500	−0.449	0.286	0.071	0.02	$\Lambda_1(1.4)$	0.700	0.002	0.072	0.068	0.94
γ_{11}	−0.100	−0.047	0.117	0.118	0.93	$\Lambda_2(0.3)$	0.312	−0.035	0.034	0.030	0.76
γ_{12}	0.000	0.052	0.151	0.097	0.79	$\Lambda_2(0.7)$	0.765	−0.105	0.065	0.054	0.51
γ_{21}	0.100	−0.073	0.125	0.095	0.86	$\Lambda_2(1.4)$	1.680	−0.327	0.156	0.098	0.18
γ_{22}	0.100	0.017	0.065	0.064	0.94	$\Lambda_3(0.3)$	0.300	0.001	0.030	0.029	0.94
γ_{31}	0.000	0.086	0.084	0.049	0.63	$\Lambda_3(0.7)$	0.700	0.006	0.055	0.053	0.93
γ_{32}	−0.100	0.101	0.067	0.048	0.42	$\Lambda_3(1.4)$	1.400	0.030	0.098	0.098	0.94

NOTE: See NOTE to Table S1 for interpretations of the column names.

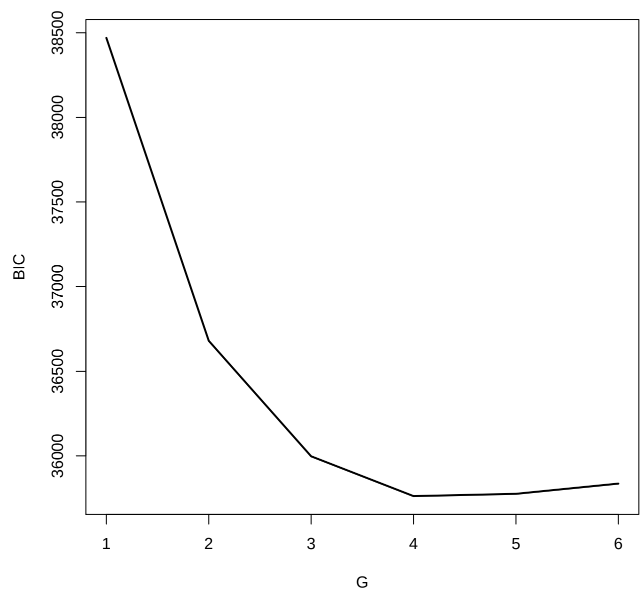

FIG S1. *BIC values for $G = 1, \ldots, 6$ for the ARIC data. Constant terms in the likelihood function were omitted in the calculation of the BIC.*

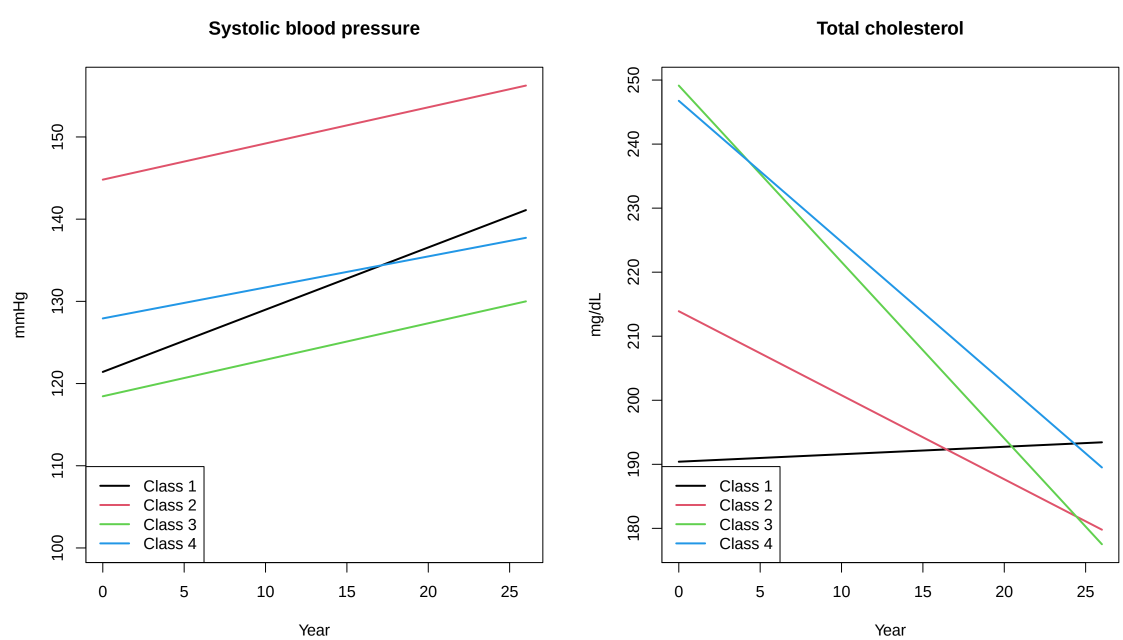

FIG S2. *Estimated class-specific trajectories of the longitudinal outcomes for a female subject from the Jackson center with all continuous covariates equal to the corresponding sample mean values and $b_3 = 0$ for the ARIC data.*

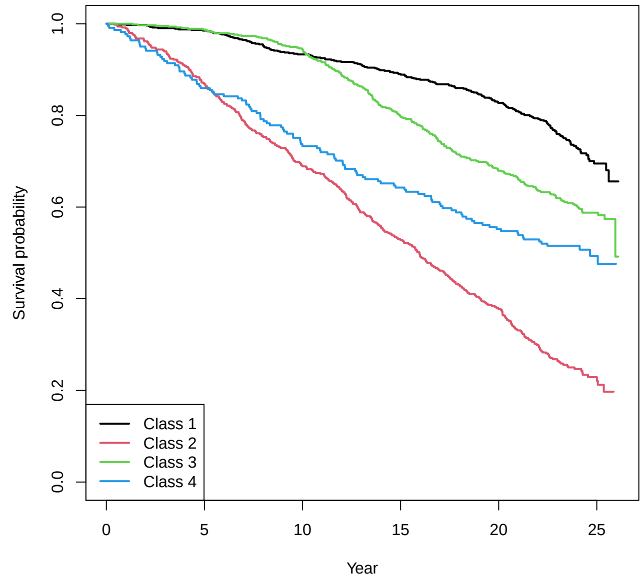

FIG S3. *Empirical class-specific survival functions for the ARIC data.*

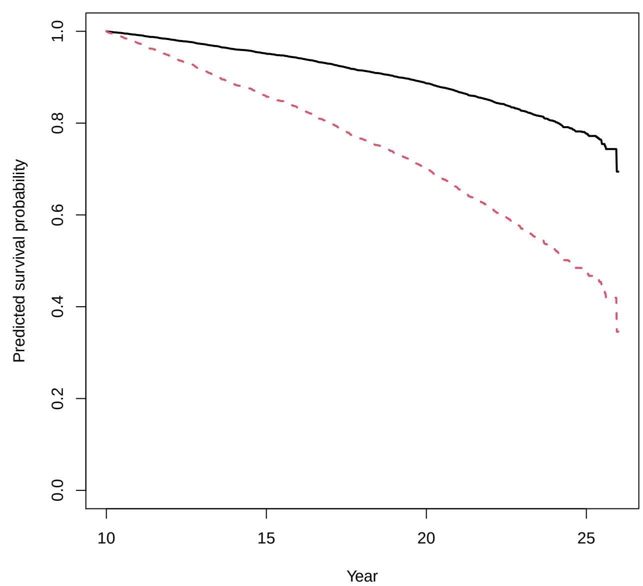

FIG S4. *Estimated conditional survival curves for the two hypothetical subjects from the ARIC study. Both subjects were from the Forsyth center, female, aged 53, and were measured for both systolic blood pressure and total cholesterol level at baseline, year 5, and year 10. The total cholesterol level measurements were 210, 205, and 200. The solid black curve pertains to a subject who had baseline BMI and glucose level at the sample median values, was a nonsmoker, and had systolic blood pressure measurements of 120, 125, and 130. The dashed red curve pertains to a subject who had baseline BMI and glucose level at the 75% quantiles, was a smoker, and had systolic blood pressure measurements of 140, 145, and 150.*