PROBIT TIME-TO-EVENT REGRESSION FOR MISCLASSIFIED GROUP TESTING DATA

Lijun Fang[1], Tao Hu[1], Shuwei Li[*2], Lianming Wang[3],

Christopher S. McMahan[4], and Joshua M. Tebbs[3]

[1] *Capital Normal University*, [2] *Guangzhou University*,

[3] *University of South Carolina*, and [4] *Clemson University*

Supplementary Material

1 Conditional expectations

We derive expressions for μ_{ij}^{+}, μ_{ij}^{-} and $E(\phi_{ij})$ in Section 3 in the manuscript. Each quantity is a conditional expectation given the observed data $\mathcal{O}_i = \{(Y_i, X_{ij}, \boldsymbol{Z}_{ij}); i = 1, \ldots, n, \ j = 1, \ldots, J_i\}$ and current parameter values $\boldsymbol{\beta}^{(m)}$ and $\boldsymbol{\xi}^{(m)}$, but we do not emphasize this in the notation. Recall μ_{ij}^{+} and μ_{ij}^{-} are conditional expectations of G_{ij} under the constraints $G_{ij} > 0$ and $G_{ij} < 0$, respectively. To calculate these, we use the following known result for truncated normal random variables.

*Corresponding author.

Result: If W is normally distributed with mean μ and variance σ^2, then

$$\begin{aligned} E(W \mid W > b) &= \mu + \frac{\sigma^2 f_W(b)}{1 - F_W(b)} \\ E(W \mid W < b) &= \mu - \frac{\sigma^2 f_W(b)}{F_W(b)}, \end{aligned}$$

where $f_W(\cdot)$ and $F_W(\cdot)$ are the probability density function and the cumulative distribution function of W, respectively.

Recall $G_{ij} = \alpha_n(X_{ij}) + \boldsymbol{\beta}^\top \boldsymbol{Z}_{ij} + \varepsilon_{ij}$, where ε_{ij} is a standard normal random variable and aggregate observation times and covariates for the ith pool into $D_i = \{X_{ij}, \boldsymbol{Z}_{ij}, j = 1, \dots, J_i\}$. From the result above, we have

$$\begin{aligned} \mu_{ij}^+ &= E(G_{ij} \mid G_{ij} > 0, D_i) \\ &= E\left[\left\{\boldsymbol{\beta}^\top \boldsymbol{Z}_{ij} + \alpha_n(X_{ij}) + \varepsilon_{ij}\right\} \mid \boldsymbol{\beta}^\top \boldsymbol{Z}_{ij} + \alpha_n(X_{ij}) + \varepsilon_{ij} > 0,\, D_i\right] \\ &= \boldsymbol{\beta}^\top \boldsymbol{Z}_{ij} + \alpha_n(X_{ij}) + E\{\varepsilon_{ij} \mid \varepsilon_{ij} > -\boldsymbol{\beta}^\top \boldsymbol{Z}_{ij} - \alpha_n(X_{ij}), D_i\} \\ &= \boldsymbol{\beta}^\top \boldsymbol{Z}_{ij} + \alpha_n(X_{ij}) + \frac{\varphi\{-\boldsymbol{\beta}^\top \boldsymbol{Z}_{ij} - \alpha_n(X_{ij})\}}{1 - \Phi\{-\boldsymbol{\beta}^\top \boldsymbol{Z}_{ij} - \alpha_n(X_{ij})\}} \\ &= \boldsymbol{\beta}^\top \boldsymbol{Z}_{ij} + \alpha_n(X_{ij}) + \frac{\varphi\{\boldsymbol{\beta}^\top \boldsymbol{Z}_{ij} + \alpha_n(X_{ij})\}}{\Phi\{\boldsymbol{\beta}^\top \boldsymbol{Z}_{ij} + \alpha_n(X_{ij})\}} \\ &= \boldsymbol{\beta}^\top \boldsymbol{Z}_{ij} + \log\left\{\sum_{l=1}^{L_n} \xi_l b_l(X_{ij})\right\} + \frac{\varphi[\boldsymbol{\beta}^\top \boldsymbol{Z}_{ij} + \log\{\sum_{l=1}^{L_n} \xi_l b_l(X_{ij})\}]}{\Phi[\boldsymbol{\beta}^\top \boldsymbol{Z}_{ij} + \log\{\sum_{l=1}^{L_n} \xi_l b_l(X_{ij})\}]}, \end{aligned}$$

where $\varphi(\cdot)$ and $\Phi(\cdot)$ are the standard normal probability density function and cumulative distribution function, respectively. This expression, when evaluated at the current parameter estimates $\boldsymbol{\beta}^{(m)}$ and $\boldsymbol{\xi}^{(m)}$, is

$$\mu_{ij}^+ = \boldsymbol{Z}_{ij}^\top \boldsymbol{\beta}^{(m)} + \log\left\{\sum_{l=1}^{L_n} \xi_l^{(m)} b_l(X_{ij})\right\} + \frac{\varphi[\boldsymbol{Z}_{ij}^\top \boldsymbol{\beta}^{(m)} + \log\{\sum_{l=1}^{L_n} \xi_l^{(m)} b_l(X_{ij})\}]}{\Phi[\boldsymbol{Z}_{ij}^\top \boldsymbol{\beta}^{(m)} + \log\{\sum_{l=1}^{L_n} \xi_l^{(m)} b_l(X_{ij})\}]},$$

as given in Section 3. The quantity

$$\mu_{ij}^{-} = \boldsymbol{Z}_{ij}^{\top}\boldsymbol{\beta}^{(m)} + \log\left\{\sum_{l=1}^{L_n} \xi_l^{(m)} b_l(X_{ij})\right\} - \frac{\varphi[\boldsymbol{Z}_{ij}^{\top}\boldsymbol{\beta}^{(m)} + \log\{\sum_{l=1}^{L_n} \xi_l^{(m)} b_l(X_{ij})\}]}{1 - \Phi[\boldsymbol{Z}_{ij}^{\top}\boldsymbol{\beta}^{(m)} + \log\{\sum_{l=1}^{L_n} \xi_l^{(m)} b_l(X_{ij})\}]}$$

is derived similarly.

We now derive $E(\phi_{ij}) = E(\phi_{ij} \mid Y_i, D_i)$. When $Y_i = 1$, we have

$$\begin{aligned} E(\phi_{ij} \mid Y_i = 1, D_i) &= P(\phi_{ij} = 1 \mid Y_i = 1, D_i) \\ &= \frac{P(Y_i = 1 \mid \phi_{ij} = 1, D_i)\, P(\phi_{ij} = 1 \mid D_i)}{P(Y_i = 1 \mid D_i)}. \end{aligned}$$

Under our assumptions, the sensitivity ν and specificity ω are independent of the observation times and covariates so that $P(Y_i = 1 \mid \phi_{ij} = 1, D_i) = P(Y_i = 1 \mid \phi_{ij} = 1) = \nu$. Furthermore, under the assumed model,

$$P(\phi_{ij} = 1 \mid D_i) = P(T_{ij} \le X_{ij} \mid D_i) = \Phi\left[\boldsymbol{\beta}^{\top}\boldsymbol{Z}_{ij} + \log\left\{\sum_{l=1}^{L_n} \xi_l b_l(X_{ij})\right\}\right].$$

Finally, because $P(Y_i = 1 \mid D_i) = \nu - \gamma P(\Delta_i = 0 \mid D_i)$, where $\gamma = \nu + \omega - 1$, and

$$P(\Delta_i = 0 \mid D_i) = \prod_{j=1}^{J_i}\left(1 - \Phi\left[\boldsymbol{\beta}^{\top}\boldsymbol{Z}_{ij} + \log\left\{\sum_{l=1}^{L_n} \xi_l b_l(X_{ij})\right\}\right]\right),$$

it follows that

$$E(\phi_{ij} \mid Y_i = 1, D_i) = \frac{\nu\, \Phi[\boldsymbol{\beta}^{\top}\boldsymbol{Z}_{ij} + \log\{\sum_{l=1}^{L_n} \xi_l b_l(X_{ij})\}]}{\nu - \gamma \prod_{j=1}^{J_i}\Big(1 - \Phi[\boldsymbol{\beta}^{\top}\boldsymbol{Z}_{ij} + \log\{\sum_{l=1}^{L_n} \xi_l b_l(X_{ij})\}]\Big)}.$$

An analogous calculation shows

$$E(\phi_{ij} \mid Y_i = 0, D_i) = \frac{(1 - \nu)\Phi[\boldsymbol{\beta}^{\top}\boldsymbol{Z}_{ij} + \log\{\sum_{l=1}^{L_n} \xi_l b_l(X_{ij})\}]}{1 - \nu + \gamma \prod_{j=1}^{J_i}\Big(1 - \Phi[\boldsymbol{\beta}^{\top}\boldsymbol{Z}_{ij} + \log\{\sum_{l=1}^{L_n} \xi_l b_l(X_{ij})\}]\Big)}.$$

Combining both conditional expectations and evaluating these at the current parameter estimates $\boldsymbol{\beta}^{(m)}$ and $\boldsymbol{\xi}^{(m)}$ yields

$$E(\phi_{ij}) = \frac{\nu Y_i\, \Phi[\boldsymbol{Z}_{ij}^\top \boldsymbol{\beta}^{(m)} + \log\{\sum_{l=1}^{L_n} \xi_l^{(m)} b_l(X_{ij})\}]}{\nu - \gamma \prod_{j=1}^{J_i} \Big(1 - \Phi[\boldsymbol{Z}_{ij}^\top \boldsymbol{\beta}^{(m)} + \log\{\sum_{l=1}^{L_n} \xi_l^{(m)} b_l(X_{ij})\}]\Big)} \\ + \frac{(1-\nu)(1-Y_i)\, \Phi[\boldsymbol{Z}_{ij}^\top \boldsymbol{\beta}^{(m)} + \log\{\sum_{l=1}^{L_n} \xi_l^{(m)} b_l(X_{ij})\}]}{1 - \nu + \gamma \prod_{j=1}^{J_i} \Big(1 - \Phi[\boldsymbol{Z}_{ij}^\top \boldsymbol{\beta}^{(m)} + \log\{\sum_{l=1}^{L_n} \xi_l^{(m)} b_l(X_{ij})\}]\Big)},$$

as given in Section 3.

2 Proofs

We present proofs of Theorems 1-3 in Section 4. In what follows, for a measurable function f and a random variable W with distribution P, we define $\mathbb{P}f = \int f(w)\mathrm{d}P(w)$ and $\mathbb{P}_n f = n^{-1}\sum_{i=1}^n f(W_i)$ so that $\sqrt{n}(\mathbb{P}_n - \mathbb{P})$ is an empirical process. In the proofs below, K is a positive constant whose value may change from place to place when it is used.

The observed data log-likelihood function for a single pool of size J is

$$l(\boldsymbol{\theta}) = l(\boldsymbol{\beta}, \alpha) = Y \log\left(\nu - \gamma \prod_{j=1}^{J} \left[1 - \Phi\left\{\alpha(X_j) + \boldsymbol{\beta}^\top \boldsymbol{Z}_j\right\}\right]\right) \\ + (1-Y)\log\left(1 - \nu + \gamma \prod_{j=1}^{J} \left[1 - \Phi\left\{\alpha(X_j) + \boldsymbol{\beta}^\top \boldsymbol{Z}_j\right\}\right]\right).$$

Define $\boldsymbol{\Theta}_n = \{\boldsymbol{\theta}_n = (\boldsymbol{\beta}, \alpha_n) \in \mathcal{B} \otimes \mathcal{A}_n\}$ as in Section 4, let $\mathcal{L}_1 = \{l(\boldsymbol{\theta}_n) : \boldsymbol{\theta}_n \in \boldsymbol{\Theta}_n\}$, and suppose $\epsilon > 0$. From Pollard (1984), we define the covering number of the class of functions $\mathcal{L}_1$, denoted by $N(\epsilon, \mathcal{L}_1, L_1(\mathbb{P}_n))$, to be the smallest integer M_ϵ for which there exists $\{\boldsymbol{\theta}_n^{(1)}, \ldots, \boldsymbol{\theta}_n^{(M_\epsilon)}\}$ such that

$$\min_{\tilde{m}\in\{1,\ldots,M_\epsilon\}} \mathbb{P}_n|l(\boldsymbol{\theta}_n) - l(\boldsymbol{\theta}_n^{(\tilde{m})})| < \epsilon,$$

for each $\boldsymbol{\theta}_n \in \boldsymbol{\Theta}_n$, where $\boldsymbol{\theta}_n^{(\tilde{m})} = (\boldsymbol{\beta}^{(\tilde{m})}, \alpha_n^{(\tilde{m})}) \in \boldsymbol{\Theta}_n$, $\tilde{m} = 1\ldots, M_\epsilon$. We define $N(\epsilon, \mathcal{F}, L_1(\mathbb{P}_n)) = \infty$ if no such M_ϵ exists.

Lemma: Under conditions (A1) – (A3) stated in Section 4, the covering number of $\mathcal{L}_1$ satisfies $N(\epsilon, \mathcal{L}_1, L_1(\mathbb{P}_n)) \leq K\epsilon^{-(p+L_n)}$, where $L_n = q_n + k$ is the number of basis functions and p is the dimension of $\boldsymbol{\beta}$.

Proof. Recall $\alpha_n(t) = \log\{\sum_{l=1}^{L_n} \xi_l b_l(t)\}$ and let $\Lambda_n(t) = \sum_{l=1}^{L_n} \xi_l b_l(t)$. For any $\boldsymbol{\theta}_n^{(1)} = (\boldsymbol{\beta}^{(1)}, \alpha_n^{(1)})$ and $\boldsymbol{\theta}_n^{(2)} = (\boldsymbol{\beta}^{(2)}, \alpha_n^{(2)}) \in \boldsymbol{\Theta}_n$, it follows that $|l(\boldsymbol{\theta}_n^{(1)}) - l(\boldsymbol{\theta}_n^{(2)})| \leq K(\|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\| + \|\alpha_n^{(1)} - \alpha_n^{(2)}\|_\infty)$ by the Mean Value Theorem, where $\|\cdot\|_\infty$ denotes the infinite norm and $\|g_1(x) - g_2(x)\|_\infty = \sup_x |g_1(x) - g_2(x)|$ for functions $g_1(\cdot)$ and $g_2(\cdot)$. Let $\boldsymbol{\xi}^{(\tilde{j})} = (\xi_1^{(\tilde{j})}, \ldots, \xi_{L_n}^{(\tilde{j})})^\top$ denote the coefficients corresponding to $\Lambda_n^{(\tilde{j})}$, for $\tilde{j} = 1, 2$. It follows that

$$\begin{aligned}\|\Lambda_n^{(1)} - \Lambda_n^{(2)}\|_\infty &= \sup_{t\in(\tau_1,\tau_2]} \left| \sum_{l=1}^{L_n} \xi_l^{(1)} b_l(t) - \sum_{l=1}^{L_n} \xi_l^{(2)} b_l(t) \right| \\ &\leq K \max_{1\leq l\leq L_n} |\xi_l^{(1)} - \xi_l^{(2)}| \leq K\|\boldsymbol{\xi}^{(1)} - \boldsymbol{\xi}^{(2)}\|.\end{aligned}$$

From the Mean Value Theorem again, we have $\|\alpha_n^{(1)} - \alpha_n^{(2)}\|_\infty = \|\log\Lambda_n^{(1)} - \log\Lambda_n^{(2)}\|_\infty \leq K\|\Lambda_n^{(1)} - \Lambda_n^{(2)}\|_\infty \leq K\|\boldsymbol{\xi}^{(1)} - \boldsymbol{\xi}^{(2)}\|$, from which it follows $|l(\boldsymbol{\theta}_n^{(1)}) - l(\boldsymbol{\theta}_n^{(2)})| \leq K\|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\| + K\|\boldsymbol{\xi}^{(1)} - \boldsymbol{\xi}^{(2)}\|$ and $\mathbb{P}_n\big|l(\boldsymbol{\theta}_n) - l(\boldsymbol{\theta}_n^{(\tilde{j})})\big| \leq$

$K\|\boldsymbol{\beta}-\boldsymbol{\beta}^{(\tilde{j})}\|+K\|\boldsymbol{\xi}-\boldsymbol{\xi}^{(\tilde{j})}\|$, for any $\boldsymbol{\theta}_n=(\boldsymbol{\beta},\alpha_n)\in\boldsymbol{\Theta}_n$ and $\boldsymbol{\xi}=(\xi_1,\ldots,\xi_{L_n})^\top$, $\tilde{j}=1,2$. From Lemma 2.5 in van de Geer (2000), one can show $\{\boldsymbol{\beta}\in\mathbb{R}^p,\|\boldsymbol{\beta}\|\leq M_{\boldsymbol{\beta}}\}$ is covered by $[5M_{\boldsymbol{\beta}}/\{\epsilon/(2K)\}]^p$ balls with radius $\epsilon/(2K)$, where $M_{\boldsymbol{\beta}}$ is a large positive constant. Similarly, one can find the number of balls with radius $\epsilon/(2K)$ to cover $\{\boldsymbol{\xi}\in\mathbb{R}^{L_n},M_n^{-1}\leq\xi_l\leq M_n,\ l=1,\ldots,L_n\}$, where M_n is a large positive constant. Therefore,

$$N(\epsilon,\mathcal{L}_1,L_1(\mathbb{P}_n))\leq\left(\frac{10M_{\boldsymbol{\beta}}K}{\epsilon}\right)^p\left(\frac{10M_nK}{\epsilon}\right)^{L_n}\leq K\epsilon^{-(p+L_n)}.\ \square$$

We now prove Theorems 1-3 in Section 4. Conditions (A1)–(A6) mentioned in the proofs are stated in the manuscript.

Theorem 1: Under conditions (A1)–(A4), the sieve estimator is strongly consistent, that is, $\|\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\|\to0$ and $\sup_{t\in[\tau_1,\tau_2]}|\hat{\alpha}_n(t)-\alpha_0(t)|\to0$ almost surely as $n\to\infty$.

Proof. From the lemma above, the covering number of $\mathcal{L}_1$ satisfies $N(\epsilon,\mathcal{L}_1,L_1(\mathbb{P}_n))\leq K\epsilon^{-(p+L_n)}$. From Inequality (31) on page 31 of Pollard (1984) and the Borel-Cantelli lemma, it follows that almost surely

$$\sup_{\boldsymbol{\theta}_n\in\boldsymbol{\Theta}_n}|\mathbb{P}_n(l(\boldsymbol{\theta}_n))-\mathbb{P}(l(\boldsymbol{\theta}_n))|\to0. \tag{B.1}$$

Let $M(\boldsymbol{\theta})=-l(\boldsymbol{\theta})$, $\zeta_{1n}=\sup_{\boldsymbol{\theta}_n\in\boldsymbol{\Theta}_n}|\mathbb{P}_nM(\boldsymbol{\theta}_n)-\mathbb{P}M(\boldsymbol{\theta}_n)|$, and $\zeta_{2n}=\mathbb{P}_nM(\boldsymbol{\theta}_0)-\mathbb{P}M(\boldsymbol{\theta}_0)$. Define $K_\epsilon=\{\boldsymbol{\theta}_n:d(\boldsymbol{\theta}_n,\boldsymbol{\theta}_0)\geq\epsilon,\boldsymbol{\theta}_n\in\boldsymbol{\Theta}_n,\boldsymbol{\theta}_0\in\boldsymbol{\Theta}\}$, for any $\epsilon>0$.

We can conclude

$$\inf_{\boldsymbol{\theta}_n \in K_\epsilon} \mathbb{P}M(\boldsymbol{\theta}_n) = \inf_{\boldsymbol{\theta}_n \in K_\epsilon} \{\mathbb{P}M(\boldsymbol{\theta}_n) - \mathbb{P}_n M(\boldsymbol{\theta}_n) + \mathbb{P}_n M(\boldsymbol{\theta}_n)\}$$
$$\leq \zeta_{1n} + \inf_{\boldsymbol{\theta}_n \in K_\epsilon} \mathbb{P}_n M(\boldsymbol{\theta}_n). \quad \text{(B.2)}$$

Furthermore, if $\hat{\boldsymbol{\theta}}_n \in K_\epsilon$, we have

$$\inf_{\boldsymbol{\theta} \in K_\epsilon} \mathbb{P}_n M(\boldsymbol{\theta}) = \mathbb{P}_n M(\hat{\boldsymbol{\theta}}_n) \leq \mathbb{P}_n M(\boldsymbol{\theta}_0) = \zeta_{2n} + \mathbb{P}M(\boldsymbol{\theta}_0). \quad \text{(B.3)}$$

Define $\delta_\epsilon = \inf_{\boldsymbol{\theta}_n \in K_\epsilon} \{\mathbb{P}M(\boldsymbol{\theta}_n) - \mathbb{P}M(\boldsymbol{\theta}_0)\}$. One can show $\delta_\epsilon > 0$ when condition (A4) holds. In fact, if $\delta_\epsilon = 0$, it follows that $l(\boldsymbol{\theta}) = l(\boldsymbol{\theta}_0)$. In particular, by considering $Y = 1$ or $Y = 0$, we have

$$\prod_{j=1}^{J} \left[1 - \Phi\left\{\alpha(X_j) + \boldsymbol{\beta}^\top \boldsymbol{Z}_j\right\}\right] = \prod_{j=1}^{J} \left[1 - \Phi\left\{\alpha_0(X_j) + \boldsymbol{\beta}_0^\top \boldsymbol{Z}_j\right\}\right]. \quad \text{(B.4)}$$

For $j \neq 1$, letting $X_j \to 0$ in (B.4) leads to $\alpha(X_1) + \boldsymbol{\beta}^\top \boldsymbol{Z}_1 = \alpha_0(X_1) + \boldsymbol{\beta}_0^\top \boldsymbol{Z}_1$. By condition (A4), we have $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ and $\alpha(X_1) = \alpha_0(X_1)$, for $X_1 \in [\tau_1, \tau_2]$. Thus, model parameters are identifiable, and we can conclude $\delta_\epsilon > 0$. It follows from (B.2) and (B.3) that $\inf_{\boldsymbol{\theta} \in K_\epsilon} \mathbb{P}M(\boldsymbol{\theta}) \leq \zeta_{1n} + \zeta_{2n} + \mathbb{P}M(\boldsymbol{\theta}_0)$ so that $\{\hat{\boldsymbol{\theta}}_n \in K_\epsilon\} \subseteq \{\zeta_{1n} + \zeta_{2n} \geq \delta_\epsilon\}$. Combining (B.1) with the Strong Law of Large Numbers, we have $\zeta_{1n} = o(1)$ and $\zeta_{2n} = o(1)$ almost surely. Therefore, because $\bigcup_{d=1}^{\infty} \bigcap_{n=d}^{\infty} \{\hat{\boldsymbol{\theta}}_n \in K_\epsilon\} \subseteq \bigcup_{d=1}^{\infty} \bigcap_{n=d}^{\infty} \{\zeta_{1n} + \zeta_{2n} \geq \delta_\epsilon\}$, we conclude $d(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) = o(1)$. $\square$

Theorem 2: Under conditions (A1)–(A5),

$$d(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) = O_p\left(n^{-\min\{r\kappa,(1-\kappa)/2\}}\right).$$

Proof. We verify the conditions of Theorem 3.4.1 in van der Vaart and Wellner (1996). Define $\alpha_{n0}(t) = \log \Lambda_{n0}(t) = \log\{\sum_{l=1}^{L_n} \xi_{l0} b_l(t)\} \in \mathcal{A}_n$, where ξ_{l0} is the true value of ξ_l, for $l = 1, \ldots, L_n$. Set $\Lambda_0 = \exp(\alpha_0)$. From condition (A5) and Lemma A1 in Lu et al. (2007), there exists a Λ_{n0} with order $k \geq r+2$ and knots $\mathcal{T}_n$ such that $\|\Lambda_{n0}(t) - \Lambda_0(t)\|_\infty = O(n^{-r\kappa})$. From the Mean Value Theorem,

$$\begin{aligned} \|\alpha_{n0}(t) - \alpha_0(t)\|_\infty &= \|\log \Lambda_{n0}(t) - \log \Lambda_0(t)\|_\infty \\ &= K\|\Lambda_{n0}(t) - \Lambda_0(t)\|_\infty \leq O(n^{-r\kappa}). \end{aligned}$$

Let $\boldsymbol{\theta}_{n0} = (\boldsymbol{\beta}_0, \alpha_{n0}) \in \boldsymbol{\Theta}_n$. By the Triangle Inequality, it follows that $d(\boldsymbol{\theta}_n, \boldsymbol{\theta}_0) > d(\boldsymbol{\theta}_n, \boldsymbol{\theta}_{n0}) - d(\boldsymbol{\theta}_0, \boldsymbol{\theta}_{n0}) \geq \delta/2 - Kn^{-r\kappa} \geq K\delta$, for large n and $\delta > 0$. Using arguments similar to those in Lemma 25.85 in van der Vaart (1998), we obtain $\mathbb{P}l(\boldsymbol{\theta}_n) - \mathbb{P}l(\boldsymbol{\theta}_0) \leq -Kd^2(\boldsymbol{\theta}_n, \boldsymbol{\theta}_0) \leq -K\delta^2$. In addition, we can easily obtain $\mathbb{P}l(\boldsymbol{\theta}_0) - \mathbb{P}l(\boldsymbol{\theta}_{n0}) \leq Kd^2(\boldsymbol{\theta}_0, \boldsymbol{\theta}_{n0}) \leq Kn^{-2r\kappa}$. Thus, $\mathbb{P}l(\boldsymbol{\theta}_n) - \mathbb{P}l(\boldsymbol{\theta}_{n0}) = \mathbb{P}l(\boldsymbol{\theta}_n) - \mathbb{P}l(\boldsymbol{\theta}_0) + \mathbb{P}l(\boldsymbol{\theta}_0) - \mathbb{P}l(\boldsymbol{\theta}_{n0}) \leq -K\delta^2 + Kn^{-2r\kappa}$, which converges to $-K\delta^2$ as $n \to \infty$. Define the class of functions $\mathcal{L}_2(\delta) = \{l(\boldsymbol{\theta}_n) - l(\boldsymbol{\theta}_{n0}) : \boldsymbol{\theta}_n \in \boldsymbol{\Theta}_n, \boldsymbol{\theta}_{n0} \in \boldsymbol{\Theta}_n \text{ and } \delta/2 < d(\boldsymbol{\theta}_n, \boldsymbol{\theta}_{n0}) < \delta\}$. From the method in Shen and Wong (1994), we show $\log N_{[]}(\epsilon, \mathcal{L}_2(\delta), L_2(\mathbb{P})) \leq$

$KL_n \log(\delta/\epsilon)$, for $0 < \epsilon < \delta$. Furthermore, after algebra, we verify $\|l(\boldsymbol{\theta}_n) - l(\boldsymbol{\theta}_{n0})\|_2^2 \leq K\delta^2$ for any function in $\mathcal{L}_2(\delta)$. The bracketing integral

$$J_{[]}(\delta, \mathcal{L}_2(\delta), L_2(\mathbb{P})) = \int_0^\delta \sqrt{1 + \log N_{[]}(\epsilon, \mathcal{L}_2(\delta), L_2(\mathbb{P})}\, \mathrm{d}\epsilon \leq KL_n^{1/2}\delta.$$

Hence, by Lemma 3.4.2 of van der Vaart and Wellner (1996), we have

$$\begin{aligned} E^*\|\sqrt{n}(\mathbb{P}_n - \mathbb{P})\|_{\mathcal{L}_2(\delta)} &\leq KJ_{[]}(\delta, \mathcal{L}_2(\delta), L_2(\mathbb{P}))\left\{1 + \frac{J_{[]}(\delta, \mathcal{L}_2(\delta), L_2(\mathbb{P}))}{\delta^2\sqrt{n}}\right\} \\ &\leq KL_n^{1/2}\delta\left(1 + \frac{KL_n^{1/2}\delta}{\delta^2\sqrt{n}}\right) = O(L_n^{1/2}\delta + L_n/n^{1/2}), \end{aligned}$$

where E^* is the outer expectation. Let $\phi_n(\delta) = L_n^{1/2}\delta + L_n/n^{1/2}$. It is easy to show $\phi_n(\delta)/\delta$ is decreasing with respect to δ, $n^{2r\kappa}\phi_n(1/n^{r\kappa}) = n^{1/2}\{n^{r\kappa-(1-\kappa)/2} + n^{2r\kappa-(1-\kappa)}\}$, and $n^{1-\kappa}\phi_n(1/n^{(1-\kappa)/2}) = 2n^{1/2}$. Therefore, $r_n^2\phi_n(1/r_n) \leq Kn^{1/2}$ when $r_n = n^{\min\{r\kappa,(1-\kappa)/2\}}$. Because $\mathbb{P}l(\hat{\boldsymbol{\theta}}_n) - \mathbb{P}l(\boldsymbol{\theta}_{n0}) \geq 0$ and $d(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_{n0}) \leq d(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) + d(\boldsymbol{\theta}_0, \boldsymbol{\theta}_{n0}) \to 0$ in probability, we have $r_n d(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_{n0}) = O_p(1)$ by Theorem 3.4.1 in van der Vaart and Wellner (1996). These facts lead to $r_n d(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) \leq r_n d(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_{n0}) + r_n d(\boldsymbol{\theta}_{n0}, \boldsymbol{\theta}_0) = O_p(1)$, which completes the proof. □

Theorem 3: Under conditions (A1)–(A6), if $1/2(1+r) < \kappa < 1/2r$, then $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \to N(0, I^{-1}(\boldsymbol{\beta}_0))$ in distribution as $n \to \infty$, where the information matrix $I(\boldsymbol{\beta}_0)$ is given in the Supplementary Material.

Proof. The form of the information matrix $I(\boldsymbol{\beta}_0)$ is given in the proof below. The score function for a single pool of size J is

$$\begin{aligned} l_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \alpha) &= \frac{\partial l(\boldsymbol{\beta}, \alpha)}{\partial \boldsymbol{\beta}} \\ &= \frac{\gamma Y \left(\sum_{j=1}^{J} \varphi\{\alpha(X_j) + \boldsymbol{\beta}^\top \boldsymbol{Z}_j\} \boldsymbol{Z}_j \prod_{i \neq j}^{J} \left[1 - \Phi\{\alpha(X_i) + \boldsymbol{\beta}^\top \boldsymbol{Z}_i\}\right] \right)}{\nu - \gamma \prod_{j=1}^{J} [1 - \Phi\{\alpha(X_j) + \boldsymbol{\beta}^\top \boldsymbol{Z}_j\}]} \\ &- \frac{\gamma(1-Y) \left(\sum_{j=1}^{J} \varphi\{\alpha(X_j) + \boldsymbol{\beta}^\top \boldsymbol{Z}_j\} \boldsymbol{Z}_j \prod_{i \neq j}^{J} \left[1 - \Phi\{\alpha(X_i) + \boldsymbol{\beta}^\top \boldsymbol{Z}_i\}\right] \right)}{1 - \nu + \gamma \prod_{j=1}^{J} [1 - \Phi\{\alpha(X_j) + \boldsymbol{\beta}^\top \boldsymbol{Z}_j\}]}. \end{aligned}$$

Consider the submodel $\alpha_{\epsilon,\boldsymbol{h}}(t) = \alpha(t) + \epsilon \boldsymbol{h}(t)$, where $\boldsymbol{h} = (h_1, \ldots, h_p)^\top$ is a p-dimensional vector with all components in $L_2([\tau_1, \tau_2])$. The score function of $\alpha(\cdot)$ along this submodel is

$$\begin{aligned} l_{\alpha}(\boldsymbol{\beta}, \alpha)[\boldsymbol{h}] &= \left. \frac{\partial l(\boldsymbol{\beta}, \alpha_{\epsilon,\boldsymbol{h}})}{\partial \epsilon} \right|_{\epsilon=0} \\ &= \frac{\gamma Y \left(\sum_{j=1}^{J} \varphi\{\alpha(X_j) + \boldsymbol{\beta}^\top \boldsymbol{Z}_j\} \boldsymbol{h}(X_j) \prod_{i \neq j}^{J} \left[1 - \Phi\{\alpha(X_i) + \boldsymbol{\beta}^\top \boldsymbol{Z}_i\}\right] \right)}{\nu - \gamma \prod_{j=1}^{J} [1 - \Phi\{\alpha(X_j) + \boldsymbol{\beta}^\top \boldsymbol{Z}_j\}]} \\ &- \frac{\gamma(1-Y) \left(\sum_{j=1}^{J} \varphi\{\alpha(X_j) + \boldsymbol{\beta}^\top \boldsymbol{Z}_j\} \boldsymbol{h}(X_j) \prod_{i \neq j}^{J} \left[1 - \Phi\{\alpha(X_i) + \boldsymbol{\beta}^\top \boldsymbol{Z}_i\}\right] \right)}{1 - \nu + \gamma \prod_{j=1}^{J} (1 - \Phi\{\alpha(X_j) + \boldsymbol{\beta}^\top \boldsymbol{Z}_j\})}. \end{aligned}$$

For $\boldsymbol{h} \in L_2([\tau_1, \tau_2])$, take $\boldsymbol{h}^* = \arg\min_{\boldsymbol{h}} E \left\| l_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \alpha) - l_{\alpha}(\boldsymbol{\beta}, \alpha)[\boldsymbol{h}] \right\|^2$, the so-called least favorable direction. By the Lax-Milgram Theorem (Zeidler, 1995) and arguments similar to those in Zeng et al. (2016), it can be shown that $\boldsymbol{h}^*$ exists. From Bickel et al. (1993), the efficient score function for $\boldsymbol{\beta}$ is $l^*(\boldsymbol{\beta}, \alpha) = l_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \alpha) - l_{\alpha}(\boldsymbol{\beta}, \alpha)[\boldsymbol{h}^*]$. The information matrix of $\boldsymbol{\beta}$ is $I(\boldsymbol{\beta}) = E\{l^*(\boldsymbol{\beta}, \alpha)\}^{\otimes 2} = E\{l_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \alpha) - l_{\alpha}(\boldsymbol{\beta}, \alpha)[\boldsymbol{h}^*]\}^{\otimes 2}$, where $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^\top$ for the vector $\boldsymbol{a}$.

To establish asymptotic normality of $\hat{\boldsymbol{\beta}}_n$, it suffices to verify the following three conditions of Theorem 8.1 in Huang et al. (2008):

(C1) $\mathbb{P}_n l_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_n, \hat{\alpha}_n) = o_p(n^{-1/2})$ and $\mathbb{P}_n l_\alpha(\hat{\boldsymbol{\beta}}_n, \hat{\alpha}_n)[\boldsymbol{h}^*] = o_p(n^{-1/2})$

(C2) $(\mathbb{P}_n - \mathbb{P})\{l^*(\hat{\boldsymbol{\beta}}_n, \hat{\alpha}_n) - l^*(\boldsymbol{\beta}_0, \alpha_0)\} = o_p(n^{-1/2})$

(C3) $\mathbb{P}\{l^*(\hat{\boldsymbol{\beta}}_n, \hat{\alpha}_n) - l^*(\boldsymbol{\beta}_0, \alpha_0)\} = -I(\boldsymbol{\beta}_0)(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) + o_p(\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\|) + o_p(n^{-1/2})$.

We first establish (C1). Because $\hat{\boldsymbol{\beta}}_n$ is a sieve maximum likelihood estimate, $\mathbb{P}_n l_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_n, \hat{\alpha}_n) = 0$. Thus, we need to show $\mathbb{P}_n l_\alpha(\hat{\boldsymbol{\beta}}_n, \hat{\alpha}_n)[\boldsymbol{h}^*] = o_p(n^{-1/2})$. This can be done by showing $\mathbb{P}_n l_\alpha(\hat{\boldsymbol{\beta}}_n, \hat{\alpha}_n)[h_s^*] = o_p(n^{-1/2})$, where h_s^* is the sth component of $\boldsymbol{h}^*$, for $s = 1, \ldots, p$. Suppose condition (A5) holds. From Jackson's Theorem in De Boor (2001), there exists a spline function $h_{s,n}^* \in \mathcal{A}_n$ of order $k \geq r + 2$ and knots $\mathcal{T}_n$ such that $\|h_{s,n}^* - h_s^*\|_\infty = O(n^{-r\kappa}) \leq O(n^{-\kappa})$. Moreover, we can conclude $\mathbb{P}_n l_\alpha(\hat{\boldsymbol{\beta}}_n, \hat{\alpha}_n)[h_{s,n}^*] = 0$ and $\mathbb{P}\{l_\alpha(\boldsymbol{\beta}_0, \alpha_0)[h_s^* - h_{s,n}^*]\} = 0$. Therefore, $\mathbb{P}_n l_\alpha(\hat{\boldsymbol{\beta}}_n, \hat{\alpha}_n)[h_s^*]$ can be decomposed as $\mathbb{P}_n l_\alpha(\hat{\boldsymbol{\beta}}_n, \hat{\alpha}_n)[h_s^*] = I_{1,n} + I_{2,n}$ for each s, where

$$I_{1,n} = (\mathbb{P}_n - \mathbb{P})\{l_\alpha(\hat{\boldsymbol{\beta}}_n, \hat{\alpha}_n)[h_s^* - h_{s,n}^*]\}$$

and

$$I_{2,n} = \mathbb{P}\{l_\alpha(\hat{\boldsymbol{\beta}}_n, \hat{\alpha}_n)[h_s^* - h_{s,n}^*] - l_\alpha(\boldsymbol{\beta}_0, \alpha_0)[h_s^* - h_{s,n}^*]\}.$$

Define $\mathcal{L}_3^s = \{l_\alpha(\boldsymbol{\beta}, \alpha_n)[h_s^* - h_{s,n}^*] : (\boldsymbol{\beta}, \alpha_n) \in \boldsymbol{\Theta}_n, h_{s,n}^* \in \mathcal{A}_n, d(\boldsymbol{\theta}_n, \boldsymbol{\theta}_0) \leq \delta, \|h_s^* - h_{s,n}^*\|_\infty \leq \delta\}$, for $s = 1, \ldots, p$. From Shen and Wong (1994), we obtain $N_{[]}(\epsilon, \mathcal{L}_3^s, L_2(\mathbb{P})) \leq N_{[]}(\epsilon, \mathcal{L}_3^s, \|\cdot\|_\infty) \leq K(\delta/\epsilon)^{KL_n+p}$, and by Theorem 19.5 in van der Vaart (1998), we conclude $\mathcal{L}_3^s$ is a Donsker class. Furthermore, for any $l_\alpha(\boldsymbol{\beta}, \alpha_n)[h_s^* - h_{s,n}^*] \in \mathcal{L}_3^s$, we have $\mathbb{P}\{l_\alpha(\boldsymbol{\beta}, \alpha_n)[h_s^* - h_{s,n}^*]\}^2 \leq K\|h_s^* - h_{s,n}^*\|_\infty$, which converges to 0 as $n \to \infty$. Therefore, following the arguments in Corollary 2.13.12 in van der Vaart and Wellner (1996), we have $I_{1,n} = o_p(n^{-1/2})$. Under conditions (A1)–(A3) and by using the Cauchy-Schwartz Inequality, we obtain

$$\begin{aligned} I_{2,n} &\leq K\, d(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0)\, \|h_s^* - h_{s,n}^*\|_\infty \\ &\leq O_p\left(n^{-\min\{r\kappa,(1-\kappa)/2\}}\, n^{-\kappa}\right) = O_p\left(n^{-\min\{\kappa(r+1),(1+\kappa)/2\}}\right) = o_p\left(n^{-1/2}\right). \end{aligned}$$

Therefore, $\mathbb{P}_n l_\alpha(\hat{\boldsymbol{\theta}}_n, \hat{\alpha}_n)[h_s^*] = I_{1,n} + I_{2,n} = o_p(n^{-1/2})$, for $s = 1, \ldots, p$. This establishes (C1).

We next establish (C2). It follows similarly that $\mathcal{L}_4(\delta) = \{l^*(\boldsymbol{\theta}_n) - l^*(\boldsymbol{\theta}_0) : \boldsymbol{\theta}_n \in \boldsymbol{\Theta}_n$ and $d(\boldsymbol{\theta}_n, \boldsymbol{\theta}_0) \leq \delta\}$ is a Donsker class. Moreover, for any $l^*(\boldsymbol{\theta}_n) - l^*(\boldsymbol{\theta}_0) \in \mathcal{L}_4(\delta)$, we have $\mathbb{P}\{l^*(\boldsymbol{\theta}_n) - l^*(\boldsymbol{\theta}_0)\}^2 \leq K\, d^2(\boldsymbol{\theta}_n, \boldsymbol{\theta}_0)$, which converges to 0 as $n \to \infty$. Thus, by Corollary 2.3.12 in van der Vaart and Wellner (1996), condition (C2) holds.

Finally, we show that condition (C3) holds. First note that

$$\mathbb{P}\{l^*(\hat{\boldsymbol{\beta}}_n, \hat{\alpha}_n) - l^*(\boldsymbol{\beta}_0, \alpha_0)\}$$
$$= \mathbb{P}\{l_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_n, \hat{\alpha}_n) - l_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \alpha_0)\} - \mathbb{P}\{l_\alpha(\hat{\boldsymbol{\beta}}_n, \hat{\alpha}_n)[\boldsymbol{h}^*] - l_\alpha(\boldsymbol{\beta}_0, \alpha_0)[\boldsymbol{h}^*]\}.$$

We now write $l_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_n, \hat{\alpha}_n)$ and $l_\alpha(\hat{\boldsymbol{\beta}}_n, \hat{\alpha}_n)[\boldsymbol{h}^*]$ in their Taylor series expansions about $(\boldsymbol{\beta}_0, \alpha_0)$. Doing so gives

$$\mathbb{P}\{l_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_n, \hat{\alpha}_n) - l_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \alpha_0)\}$$
$$= \mathbb{P}\{l_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \alpha_0)(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) + l_{\boldsymbol{\beta}\alpha}(\boldsymbol{\beta}_0, \alpha_0)[\hat{\alpha}_n - \alpha_0]\}$$
$$+ O_p\left(\{n^{-\min\{r\kappa,(1-\kappa)/2\}}\}^2\right) \quad \text{(B.5)}$$

and

$$\mathbb{P}\{l_\alpha(\hat{\boldsymbol{\beta}}_n, \hat{\alpha}_n)[\boldsymbol{h}^*] - l_\alpha(\boldsymbol{\beta}_0, \alpha_0)[\boldsymbol{h}^*]\}$$
$$= \mathbb{P}\{l_{\alpha\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \alpha_0)[\boldsymbol{h}^*](\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) + l_{\alpha\alpha}(\boldsymbol{\beta}_0, \alpha_0)[\boldsymbol{h}^*, \hat{\alpha}_n - \alpha_0]\}$$
$$+ o_p(\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\|) + O_p\left(\{n^{-\min\{r\kappa,(1-\kappa)/2\}}\}^2\right), \quad \text{(B.6)}$$

where $l_{\boldsymbol{\beta}\boldsymbol{\beta}}$ is the second derivative of $l(\boldsymbol{\beta}, \alpha)$ with respect to $\boldsymbol{\beta}$, $l_{\boldsymbol{\beta}\alpha}[\boldsymbol{h}^*]$ is the derivative of $l_{\boldsymbol{\beta}}$ along the submodel $\alpha_{\epsilon,\boldsymbol{h}^*}$, $l_{\alpha\boldsymbol{\beta}}[\boldsymbol{h}^*]$ is the derivative of $l_\alpha[\boldsymbol{h}^*]$ with respect to $\boldsymbol{\beta}$, and $l_{\alpha\alpha}[\boldsymbol{h}^*, \hat{\alpha}_n - \alpha_0]$ is the derivative of $l_\alpha[\boldsymbol{h}^*]$ along the submodel $\alpha_0 + \epsilon(\hat{\alpha}_n - \alpha_0)$. Note the last terms of (B.5) and (B.6) are both $o_p(n^{-1/2})$ if κ satisfies $1/2(1 + r) < \kappa < 1/2r$. Because $\boldsymbol{h}^*$ is the least favorable direction, $\boldsymbol{h}^*$ satisfies $l_\alpha^{(*)} l_{\boldsymbol{\beta}} = l_\alpha^{(*)} l_\alpha$, where $l_\alpha^{(*)}$ is the adjoint operator of l_α. It follows that

$$\mathbb{P}\{l_{\alpha\alpha}[\boldsymbol{h}^*, \hat{\alpha}_n - \alpha_0]\} = -\mathbb{P}\{l_\alpha[\boldsymbol{h}^*] l_\alpha[\hat{\alpha}_n - \alpha_0]\} = -\mathbb{P}\{l_\alpha^{(*)} l_\alpha[\boldsymbol{h}^*][\hat{\alpha}_n - \alpha_0]\}$$
$$= -\mathbb{P}\{l_\alpha^{(*)} l_{\boldsymbol{\beta}}[\hat{\alpha}_n - \alpha_0]\} = \mathbb{P}\{l_{\boldsymbol{\beta}\alpha}[\hat{\alpha}_n - \alpha_0]\}. \qquad \text{(B.7)}$$

By the definition of $\boldsymbol{h}^*$ and Theorem 11.1 in van der Vaart (1998), we have

$$I(\boldsymbol{\beta}_0) = \mathbb{P}\{l^*(\boldsymbol{\beta}_0, \alpha_0)\}^{\otimes 2} = \mathbb{P}[\{l_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \alpha_0) - l_\alpha(\boldsymbol{\beta}_0, \alpha_0)[\boldsymbol{h}^*]\}\{l_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \alpha_0)\}^\top]$$
$$- \mathbb{P}\{l_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \alpha_0) - l_{\alpha\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \alpha_0)[\boldsymbol{h}^*]\}. \qquad \text{(B.8)}$$

Combining (B.5)–(B.8), it follows that

$$\mathbb{P}\{l^*(\hat{\boldsymbol{\beta}}_n, \hat{\alpha}_n) - l^*(\boldsymbol{\beta}_0, \alpha_0)\} = -I(\boldsymbol{\beta}_0)(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) + o_p(\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\|) + o_p(n^{-1/2}),$$

which establishes (C3).

Finally, we show $I(\boldsymbol{\beta}_0)$ is nonsingular. If $I(\boldsymbol{\beta}_0)$ is singular, then there exists a nonzero vector $\boldsymbol{u}$ such that

$$\boldsymbol{u}^\top E\left[\{l_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \alpha_0) - l_\alpha(\boldsymbol{\beta}_0, \alpha_0)[\boldsymbol{h}^*]\}\{l_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \alpha_0) - l_\alpha(\boldsymbol{\beta}_0, \alpha_0)[\boldsymbol{h}^*]\}^\top\right] \boldsymbol{u} = 0.$$

This implies $\|\boldsymbol{u}^\top\{l_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \alpha_0) - l_\alpha(\boldsymbol{\beta}_0, \alpha_0)[\boldsymbol{h}^*]\}\|_2^2 = 0$ and thus $\boldsymbol{u}^\top\{l_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \alpha_0) - l_\alpha(\boldsymbol{\beta}_0, \alpha_0)[\boldsymbol{h}^*]\} = 0$. By considering $Y = 1$ or $Y = 0$, we have

$$\sum_{j=1}^J \varphi\{\alpha(X_j) + \boldsymbol{\beta}^\top \boldsymbol{Z}_j\} \prod_{i \neq j}^J \left[1 - \Phi\{\alpha(X_i) + \boldsymbol{\beta}^\top \boldsymbol{Z}_i\}\right] (\boldsymbol{u}^\top \boldsymbol{Z}_j - \boldsymbol{u}^\top \boldsymbol{h}^*) = 0.$$

Because $\varphi\{\alpha(X_j) + \boldsymbol{\beta}^\top \boldsymbol{Z}_j\} > 0$ and $\prod_{i \neq j}^J [1 - \Phi\{\alpha(X_i) + \boldsymbol{\beta}^\top \boldsymbol{Z}_i\}] > 0$ for any j and $X_j \in [\tau_1, \tau_2]$, it must be true that $\boldsymbol{u} = \boldsymbol{0}$ by condition (A6). This is a contradiction and hence $I(\boldsymbol{\beta}_0)$ is nonsingular. $\square$

3 Second simulation study

We performed a second simulation to evaluate the finite-sample properties of our regression methods. All settings are the same as the first study except for as follows.

- We use 5 covariates $Z_{ij1}, ..., Z_{ij5}$; each follows a Bernoulli(0.5) distribution. The true $\boldsymbol{\beta} = (\beta_1, ..., \beta_5)^\top = (0.5, 0.5, -0.5, -0.5, -0.5)^\top$. These configurations provide an average right censoring rate of approximately 92%.

- We use pool sizes 1, 2, 3 or 4. These are selected according to a discrete uniform distribution with probability 0.25 for each pool size (after selection, they are regarded to be fixed).

Table S.1 (page 17) shows the results for three configurations of the assay sensitivity and specificity: $(\nu, \omega) = (1, 1)$, $(0.90, 0.95)$, and $(0.85, 0.85)$. Figure S.1 (page 18) shows averaged estimates of the baseline survival function $S(t)$ for $(\nu, \omega) = (1, 1)$ and $(\nu, \omega) = (0.85, 0.85)$. The results for this study are in agreement with those from the first study. Estimating the probit model and the large-sample covariance matrix took approximately 8 minutes on average for each group testing data set.

(continued on the next page)

Table S.1: Second simulation study. Empirical bias (Bias) and sample standard deviation (SSD) of 500 sieve maximum likelihood estimates. The averaged estimated standard error (ESE) and empirical coverage probabilities (CP) are also included. The second and third columns show results for individual testing when fixing the number of individuals and the number of tests, respectively.

(ν, ω)		Group testing				Number of individuals fixed				Number of tests fixed			
		Bias	SSD	ESE	CP	Bias	SSD	ESE	CP	Bias	SSD	ESE	CP
$(1, 1)$	$\hat{\beta}_1$	-0.001	0.043	0.045	96.0	0.000	0.035	0.032	94.6	-0.003	0.052	0.051	96.0
	$\hat{\beta}_2$	-0.002	0.047	0.045	94.4	-0.002	0.032	0.032	95.2	-0.001	0.050	0.051	94.8
	$\hat{\beta}_3$	-0.001	0.044	0.045	94.0	0.000	0.034	0.033	92.6	0.001	0.051	0.052	95.0
	$\hat{\beta}_4$	-0.001	0.047	0.045	95.4	0.000	0.034	0.033	94.6	0.000	0.051	0.052	94.2
	$\hat{\beta}_5$	-0.002	0.044	0.045	95.4	0.000	0.032	0.033	95.2	0.001	0.054	0.052	94.2
$(0.90, 0.95)$	$\hat{\beta}_1$	0.000	0.058	0.057	95.8	0.000	0.047	0.044	94.2	-0.002	0.073	0.071	94.2
	$\hat{\beta}_2$	-0.003	0.059	0.057	93.4	-0.003	0.045	0.044	94.2	-0.003	0.071	0.071	95.4
	$\hat{\beta}_3$	-0.002	0.058	0.057	95.2	0.002	0.046	0.045	94.2	-0.002	0.071	0.071	96.0
	$\hat{\beta}_4$	-0.006	0.060	0.057	94.4	0.000	0.046	0.045	94.0	-0.004	0.075	0.072	94.6
	$\hat{\beta}_5$	-0.002	0.06.	0.057	93.6	0.001	0.045	0.045	94.0	0.000	0.072	0.071	94.8
$(0.85, 0.85)$	$\hat{\beta}_1$	0.001	0.077	0.075	94.4	-0.002	0.065	0.062	94.2	-0.001	0.099	0.098	94.4
	$\hat{\beta}_2$	-0.003	0.076	0.075	94.0	-0.004	0.063	0.062	93.8	-0.002	0.104	0.098	93.2
	$\hat{\beta}_3$	-0.006	0.078	0.076	94.6	-0.003	0.063	0.063	95.4	-0.011	0.098	0.100	96.4
	$\hat{\beta}_4$	-0.007	0.078	0.076	94.0	-0.007	0.066	0.064	93.6	-0.011	0.107	0.101	93.4
	$\hat{\beta}_5$	-0.007	0.084	0.076	93.9	-0.002	0.065	0.063	95.0	-0.007	0.101	0.101	95.0

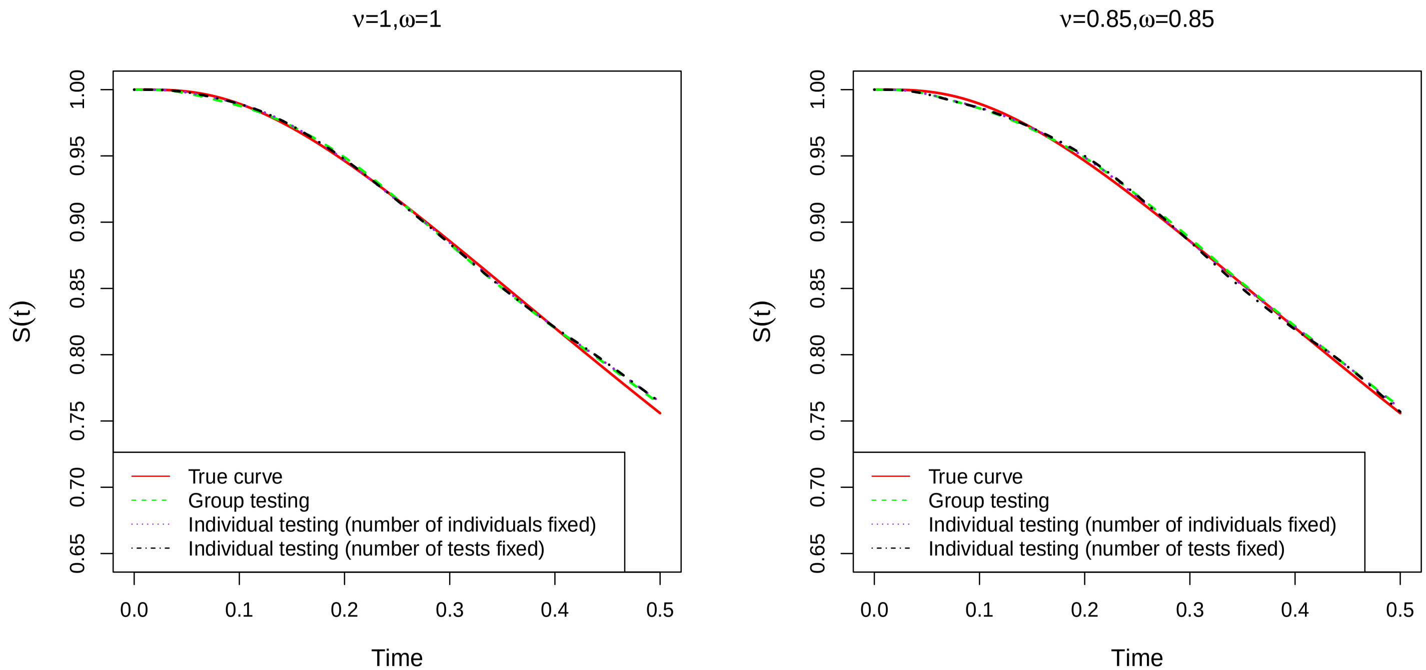

Figure S.1: Second simulation study. Estimated baseline survival curves for group testing and individual testing. Left: No misclassification. Right: Misclassification with sensitivity $\nu = 0.85$ and specificity $\omega = 0.85$.

4 PH analysis of Iowa data

For comparison purposes, we also estimated the Cox proportional hazards (PH) model

$$S(t \mid \boldsymbol{Z}_{ij}) = \exp\{-\Lambda(t)\exp(\boldsymbol{Z}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\},$$

with the Iowa data using the approach in Li et al. (2024). In this model, $\boldsymbol{Z}_{ij} = (Z_{ij1}, Z_{ij2})^{\top}$, $\boldsymbol{\beta} = (\beta_1, \beta_2)^{\top}$, and $\Lambda(t)$ is the unknown cumulative baseline hazard function. The covariates Z_{ij1} and Z_{ij2} are indicator variables for race as defined in Section 6 in the manuscript. Here are the relevant results:

- the sieve ML estimate of β_1 is 0.336 with estimated standard error 0.104 (p-value = 0.001)
- the sieve ML estimate of β_2 is 0.216 with estimated standard error 0.150 (p-value = 0.150).

As in the probit analysis, the time to chlamydial disease onset is stochastically smaller for African American subjects when compared to Caucasian subjects. When making the same comparison with subjects of other races, the difference is not statistically significant. We note that estimating the PH model and the large-sample covariance matrix of the regression parameter estimators took approximately 2 hours.

References

Bickel, P., Klaassen, C., Ritov, Y., and Wellner, J. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press: Baltimore.

De Boor, C. (2001). *A Practical Guide to Splines*. Springer-Verlag: New York.

Huang, J., Zhang, Y., and Hua, L. (2008). A least-squares approach to consistent information estimation in semiparametric models. Technical Report, Department of Biostatistics, University of Iowa.

Li, S., Hu, T., Wang, L., McMahan, C., and Tebbs, J. (2024). Regression analysis of group-tested current status data. *Biometrika* **111**, 1047–1061.

Lu, M., Zhang, Y., and Huang, J. (2007). Estimation of the mean function with panel count data using monotone polynomial splines. *Biometrika* **94**, 705–718.

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer: New York.

Shen, X. and Wong, W. (1994). Convergence rate of sieve estimates. *Annals of Statistics* **22**, 580–615.

van de Geer, S. (2000). *Applications of Empirical Process Theory*. Cambridge University Press: Cambridge.

van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge University Press: Cambridge.

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes*. Springer: New York.

Zeidler, E. (1995). *Applied Functional Analysis: Applications to Mathematical Physics*. Springer: New York.

Zeng, D., Mao, L., and Lin, D. (2016). Maximum likelihood estimation for semiparametric transformation models with interval-censored data. *Biometrika* **103**, 253–271.