Supplemental Material for "Efficient Semiparametric Inference Under Two-Phase Sampling, With Applications to Genetic Association Studies"

Tao et al.

S.1. Asymptotic Properties

Let $\boldsymbol{\Theta}$ denote the parameter space of $\boldsymbol{\theta}$, which is a bounded open set in the interior of the domain of $\boldsymbol{\theta}$, and let $\mathcal{F}$ denote the space of the joint distributions of $(\boldsymbol{X}, \boldsymbol{Z})$. Let $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ and $F_0 \in \mathcal{F}$ denote the true values of $\boldsymbol{\theta}$ and F, respectively. We impose the following regularity conditions:

(C.1) The set of covariates $(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{W})$ has bounded support.

(C.2) If there exist two sets of parameters $(\boldsymbol{\theta}_1, F_1)$ and $(\boldsymbol{\theta}_2, F_2)$ such that

$$P_{\boldsymbol{\theta}_1}(Y|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{W})F_1(\boldsymbol{X}, \boldsymbol{Z}) = P_{\boldsymbol{\theta}_2}(Y|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{W})F_2(\boldsymbol{X}, \boldsymbol{Z}),$$

where $(Y, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{W}) \in \mathcal{C} \equiv \{(y, \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{w})\colon P(R = 1|y, \boldsymbol{z}, \boldsymbol{w}) > 0\}$, then $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ and $F_1 = F_2$. In addition, if there exists a constant vector $\boldsymbol{v}$ such that

$$[\partial \log\{P_{\boldsymbol{\theta}_0}(y_1|\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{w}_1)/P_{\boldsymbol{\theta}_0}(y_2|\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{w}_2)\}/\partial\boldsymbol{\theta}]^{\mathrm{T}}\boldsymbol{v} = 0$$

for any $(y_i, \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{w}_i) \in \mathcal{C}$, $i = 1, 2$, then $\boldsymbol{v} = \boldsymbol{0}$.

(C.3) The density function of F_0 is positive in its support and q-times continuously differentiable with respect to a suitable measure.

(C.4) The function $\mathrm{E}(R|\boldsymbol{X}, \boldsymbol{Z})$ is q-times continuously differentiable with respect to $\boldsymbol{X}$ and $\boldsymbol{Z}$.

(C.5) As $n \to \infty$, $s_n \to \infty$, and $n^{1/2}s_n^{-q/d_{\boldsymbol{z}}} \to 0$.

Remark S.1 The first part of Condition (C.2) pertains to model identifiability with complete data. For commonly used regression models, the set $\mathcal{C}$, where $P(R = 1|y, \boldsymbol{z}, \boldsymbol{w}) > 0$, does not necessarily need to cover the entire support of $(Y, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{W})$. For example, in linear regression, $\mathcal{C}$ can consist of data points with extremely large or small values of Y only. The second part of Condition (C.2) ensures that the score functions for $\boldsymbol{\theta}$ are of full rank on $\mathcal{C}$. For linear regression, this condition follows from the linear independence of the covariates $(1, \boldsymbol{X}^{\mathrm{T}}, \boldsymbol{Z}^{\mathrm{T}}, \boldsymbol{W}^{\mathrm{T}})^{\mathrm{T}}$. Condition (C.3) pertains to the smoothness of the joint distribution function of $(\boldsymbol{X}, \boldsymbol{Z})$. Condition (C.4) holds for all commonly used two-phase designs, including the extreme-tail design adopted by the NHLBI ESP. Under Condition (C.5), the order of the B-spline basis q is greater than $d_{\boldsymbol{z}}/2$. Thus, when $d_{\boldsymbol{z}} = 1$, we can choose $q = 1$ and use the histogram basis $\{B_j^1(z)\}_{j=1}^{b_n+1}$ to estimate $P(\boldsymbol{X}|Z)$.

We state the asymptotic results in two theorems and provide the proofs.

Theorem S.1. Under Conditions (C.1)–(C.5), $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| + \sup_{\boldsymbol{x},\boldsymbol{z}} |\widehat{F}(\boldsymbol{x}, \boldsymbol{z}) - F_0(\boldsymbol{x}, \boldsymbol{z})| \to 0$ almost surely.

Proof. Because $\widehat{\boldsymbol{\theta}}$ is bounded and $\widehat{F}(\boldsymbol{x}, \boldsymbol{z})$ is a distribution function with bounded support, it follows from Helly's selection theorem that, for any subsequence of $\widehat{\boldsymbol{\theta}}$ and $\widehat{F}(\boldsymbol{x}, \boldsymbol{z})$, there exists a further subsequence, still denoted as $\widehat{\boldsymbol{\theta}}$ and $\widehat{F}(\boldsymbol{x}, \boldsymbol{z})$, such that $\widehat{\boldsymbol{\theta}}$ converges almost surely to some vector $\boldsymbol{\theta}^*$ and $\widehat{F}(\boldsymbol{x}, \boldsymbol{z})$ converges weakly to some function $F^*(\boldsymbol{x}, \boldsymbol{z})$. Because F_0 is a continuous function under Condition (C.3), Theorem S.1 will hold if we can show that $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$ and $F^* = F_0$.

Because $\widehat{p}_{kj}$ maximizes expression (2), differentiating expression (2) with respect to p_{kj} yields

$$\sum_{i=1}^n R_i \frac{I(\boldsymbol{X}_i = \boldsymbol{x}_k) B_j^q(\boldsymbol{Z}_i)}{p_{kj}} + \sum_{i=1}^n (1 - R_i) \frac{P_{\widehat{\boldsymbol{\theta}}}(Y_i|\boldsymbol{x}_k, \boldsymbol{Z}_i, \boldsymbol{W}_i) B_j^q(\boldsymbol{Z}_i)}{\sum_{j'=1}^{s_n} \sum_{k'=1}^m P_{\widehat{\boldsymbol{\theta}}}(Y_i|\boldsymbol{x}_{k'}, \boldsymbol{Z}_i, \boldsymbol{W}_i) B_{j'}^q(\boldsymbol{Z}_i) p_{k'j'}} = \widehat{\mu}_j, \tag{S.1}$$

where $\widehat{\mu}_j$ is the Lagrange multiplier for the constraint that $\sum_{k=1}^m \widehat{p}_{kj} = 1$. By multiplying both sides of equation (S.1) with p_{kj} and then summing over k, we have

$$\widehat{\mu}_j = \sum_{i=1}^n R_i B_j^q(\boldsymbol{Z}_i) + \sum_{i=1}^n (1 - R_i) \frac{\sum_{k'=1}^m P_{\widehat{\boldsymbol{\theta}}}(Y_i|\boldsymbol{x}_{k'}, \boldsymbol{Z}_i, \boldsymbol{W}_i) B_j^q(\boldsymbol{Z}_i) p_{k'j}}{\sum_{j'=1}^{s_n} \sum_{k'=1}^m P_{\widehat{\boldsymbol{\theta}}}(Y_i|\boldsymbol{x}_{k'}, \boldsymbol{Z}_i, \boldsymbol{W}_i) B_{j'}^q(\boldsymbol{Z}_i) p_{k'j'}}. \tag{S.2}$$

It then follows from equation (S.1) that

$$\widehat{p}_{kj} = \frac{\sum_{i=1}^n R_i I(\boldsymbol{X}_i = \boldsymbol{x}_k) B_j^q(\boldsymbol{Z}_i)}{\widehat{\mu}_j - \sum_{i=1}^n (1 - R_i) \frac{P_{\widehat{\boldsymbol{\theta}}}(Y_i|\boldsymbol{x}_k, \boldsymbol{Z}_i, \boldsymbol{W}_i) B_j^q(\boldsymbol{Z}_i)}{\sum_{j'=1}^{s_n} \sum_{k'=1}^m P_{\widehat{\boldsymbol{\theta}}}(Y_i|\boldsymbol{x}_{k'}, \boldsymbol{Z}_i, \boldsymbol{W}_i) B_{j'}^q(\boldsymbol{Z}_i) \widehat{p}_{k'j'}}}. \tag{S.3}$$

By replacing $\widehat{\mu}_j$ in equation (S.3) with the right-hand-side of equation (S.2), we obtain

$$\widehat{P}(\boldsymbol{X} = \boldsymbol{x}_k|\boldsymbol{z}) = \sum_{j=1}^{s_n} B_j^q(\boldsymbol{z}) \widehat{p}_{kj}$$

$$
= \sum_{j=1}^{s_n} B_j^q(\boldsymbol{z}) \frac{\sum_{i=1}^n R_i I(\boldsymbol{X}_i = \boldsymbol{x}_k) B_j^q(\boldsymbol{Z}_i)}{\sum_{i=1}^n \left\{ R_i + (1-R_i) \frac{\sum_{k'=1}^m P_{\widehat{\boldsymbol{\theta}}}(Y_i|\boldsymbol{x}_{k'}, \boldsymbol{Z}_i, \boldsymbol{W}_i)\widehat{p}_{k'j} - P_{\widehat{\boldsymbol{\theta}}}(Y_i|\boldsymbol{x}_k, \boldsymbol{Z}_i, \boldsymbol{W}_i)}{\sum_{j'=1}^{s_n}\sum_{k'=1}^m P_{\widehat{\boldsymbol{\theta}}}(Y_i|\boldsymbol{x}_{k'}, \boldsymbol{Z}_i, \boldsymbol{W}_i) B_{j'}^q(\boldsymbol{Z}_i)\widehat{p}_{k'j'}} \right\} B_j^q(\boldsymbol{Z}_i)}. \tag{S.4}
$$

Because the B-spline basis functions have local support, we have

$$
|B_j^q(\widetilde{\boldsymbol{z}}) - B_j^q(\boldsymbol{z}) I(\|\widetilde{\boldsymbol{z}} - \boldsymbol{z}\| \le \xi_n)| \lesssim \xi_n, \quad j = 1, \ldots, s_n, \tag{S.5}
$$

where $\xi_n = (b_n+1)^{-1}$, and "$\lesssim$" means less than or equal to up to a constant. It follows from equation (S.4) and inequality (S.5) that $\widehat{P}(\boldsymbol{X} = \boldsymbol{x}_k|\boldsymbol{z})$ is asymptotically equivalent to

$$
\begin{aligned}
&\frac{\sum_{j=1}^{s_n}\sum_{i=1}^n R_i I(\boldsymbol{X}_i = \boldsymbol{x}_k, \|\boldsymbol{Z}_i - \boldsymbol{z}\| \le \xi_n) B_j^q(\boldsymbol{z})}{\sum_{j=1}^{s_n}\sum_{i=1}^n \left\{ R_i + (1-R_i) \frac{\sum_{k'=1}^m P_{\widehat{\boldsymbol{\theta}}}(Y_i|\boldsymbol{x}_{k'}, \boldsymbol{Z}_i, \boldsymbol{W}_i)\widehat{p}_{k'j} - P_{\widehat{\boldsymbol{\theta}}}(Y_i|\boldsymbol{x}_k, \boldsymbol{Z}_i, \boldsymbol{W}_i)}{\sum_{j'=1}^{s_n}\sum_{k'=1}^m P_{\widehat{\boldsymbol{\theta}}}(Y_i|\boldsymbol{x}_{k'}, \boldsymbol{Z}_i, \boldsymbol{W}_i) B_{j'}^q(\boldsymbol{Z}_i)\widehat{p}_{k'j'}} \right\} I(\|\boldsymbol{Z}_i - \boldsymbol{z}\| \le \xi_n) B_j^q(\boldsymbol{z})} \\
&= \frac{\sum_{j=1}^{s_n}\sum_{i=1}^n R_i I(\boldsymbol{X}_i = \boldsymbol{x}_k, \|\boldsymbol{Z}_i - \boldsymbol{z}\| \le \xi_n) B_j^q(\boldsymbol{z})}{\sum_{j=1}^{s_n}\sum_{i=1}^n \left\{ 1 - (1-R_i) \frac{P_{\widehat{\boldsymbol{\theta}}}(Y_i|\boldsymbol{x}_k, \boldsymbol{Z}_i, \boldsymbol{W}_i)}{\sum_{j'=1}^{s_n}\sum_{k'=1}^m P_{\widehat{\boldsymbol{\theta}}}(Y_i|\boldsymbol{x}_{k'}, \boldsymbol{Z}_i, \boldsymbol{W}_i) B_{j'}^q(\boldsymbol{Z}_i)\widehat{p}_{k'j'}} \right\} I(\|\boldsymbol{Z}_i - \boldsymbol{z}\| \le \xi_n) B_j^q(\boldsymbol{z})}.
\end{aligned} \tag{S.6}
$$

By combining equations (A.2) and (S.6), we conclude that the distribution function $\widehat{F}(\boldsymbol{x}, \boldsymbol{z})$ is asymptotically equivalent to

$$
n^{-1}\sum_{k=1}^m\sum_{i=1}^n I(\boldsymbol{x}_k \le \boldsymbol{x}, \boldsymbol{Z}_i \le \boldsymbol{z}) \frac{\sum_{j=1}^{s_n}\sum_{i'=1}^n R_{i'} I(\boldsymbol{X}_{i'} = \boldsymbol{x}_k, \|\boldsymbol{Z}_{i'} - \boldsymbol{Z}_i\| \le \xi_n) B_j^q(\boldsymbol{Z}_i)}{g_{1n}(\boldsymbol{x}_k, \boldsymbol{Z}_i; \widehat{\boldsymbol{\theta}}, \widehat{F})},
$$

where

$$
\begin{aligned}
g_{1n}(\boldsymbol{x}, \boldsymbol{z}; \widehat{\boldsymbol{\theta}}, \widehat{F}) = \sum_{j=1}^{s_n}\sum_{i=1}^n &\left\{ 1 - (1-R_i) \frac{P_{\widehat{\boldsymbol{\theta}}}(Y_i|\boldsymbol{x}, \boldsymbol{Z}_i, \boldsymbol{W}_i)}{\sum_{j'=1}^{s_n}\sum_{k'=1}^m P_{\widehat{\boldsymbol{\theta}}}(Y_i|\boldsymbol{x}_{k'}, \boldsymbol{Z}_i, \boldsymbol{W}_i) B_{j'}^q(\boldsymbol{Z}_i)\widehat{p}_{k'j'}} \right\} \\
&\times I(\|\boldsymbol{Z}_i - \boldsymbol{z}\| \le \xi_n) B_j^q(\boldsymbol{z}).
\end{aligned} \tag{S.7}
$$

We wish to show that $(ns_n)^{-1} g_{1n}(\boldsymbol{x}, \boldsymbol{z}; \widehat{\boldsymbol{\theta}}, \widehat{F})$ is bounded away from zero for sufficiently large n. By the approximation theory of B-splines (Schumaker 1981) and Glivenko-Cantelli theorem,

$$
\begin{aligned}
&n^{-1}\sum_{j'=1}^{s_n}\sum_{k=1}^m P_{\widehat{\boldsymbol{\theta}}}(y|\boldsymbol{x}_k, \boldsymbol{z}, \boldsymbol{w}) B_{j'}^q(\boldsymbol{z}) p_{kj'} \\
&= \int_{\widetilde{\boldsymbol{x}}} P_{\widehat{\boldsymbol{\theta}}}(y|\widetilde{\boldsymbol{x}}, \boldsymbol{z}, \boldsymbol{w}) \widehat{F}(d\widetilde{\boldsymbol{x}}, \boldsymbol{z}) \to \int_{\widetilde{\boldsymbol{x}}} P_{\boldsymbol{\theta}^*}(y|\widetilde{\boldsymbol{x}}, \boldsymbol{z}, \boldsymbol{w}) F^*(d\widetilde{\boldsymbol{x}}, \boldsymbol{z})
\end{aligned} \tag{S.8}
$$

uniformly in $(y, \boldsymbol{z}, \boldsymbol{w})$. It follows from equations (S.7) and (S.8) that $(ns_n)^{-1} g_{1n}(\boldsymbol{x}, \boldsymbol{z}; \widehat{\boldsymbol{\theta}}, \widehat{F})$ converges to $g_1(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\theta}^*, F^*)$ for $(\boldsymbol{x}, \boldsymbol{z})$ in the support of $(\boldsymbol{X}, \boldsymbol{Z})$, where

$$g_1(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\theta}^*, F^*) = \mathrm{E}\left[\left\{1 - (1-R)\frac{P_{\boldsymbol{\theta}^*}(Y|\boldsymbol{x}, \boldsymbol{Z}, \boldsymbol{W}) \int_{\widetilde{\boldsymbol{x}}} F^*(d\widetilde{\boldsymbol{x}}, \boldsymbol{Z})}{\int_{\widetilde{\boldsymbol{x}}} P_{\boldsymbol{\theta}^*}(Y|\widetilde{\boldsymbol{x}}, \boldsymbol{Z}, \boldsymbol{W}) F^*(d\widetilde{\boldsymbol{x}}, \boldsymbol{Z})}\right\} f_{\boldsymbol{z}}(\boldsymbol{Z}) \middle| \boldsymbol{Z} = \boldsymbol{z}\right] \geq 0, \tag{S.9}$$

and $f_{\boldsymbol{z}}(\cdot)$ is the density function of $\boldsymbol{Z}$. Thus, it follows from equations (S.6), (S.7), and (S.9), and the approximation theory of B-splines (Schumaker 1981) that

$$\begin{aligned} 1 = \sum_{k=1}^{m} \widehat{P}(\boldsymbol{X} = \boldsymbol{x}_k | \boldsymbol{z}) &= \sum_{k=1}^{m} \frac{\sum_{j=1}^{s_n} \sum_{i'=1}^{n} R_{i'} I(\boldsymbol{X}_{i'} = \boldsymbol{x}_k, \|\boldsymbol{Z}_{i'} - \boldsymbol{z}\| \leq \xi_n) B_j^q(\boldsymbol{z})}{g_{1n}(\boldsymbol{x}_k, \boldsymbol{z}; \widehat{\boldsymbol{\theta}}, \widehat{F})} \\ \rightarrow \int \frac{\mathrm{E}\{R f_{\boldsymbol{z}}(\boldsymbol{Z}) | \boldsymbol{Z} = \boldsymbol{z}\}}{g_1(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\theta}^*, F^*)} d\boldsymbol{x}. \end{aligned}$$

If $g_1(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\theta}^*, F^*)$ is not bounded away from zero, then there exists $\boldsymbol{x}_0 \in \mathbb{D}_{\boldsymbol{x}}$, where $\mathbb{D}_{\boldsymbol{x}}$ is the support of $\boldsymbol{X}$, such that $g_1(\boldsymbol{x}_0, \boldsymbol{z}; \boldsymbol{\theta}^*, F^*) = 0$. Because $g_1(\boldsymbol{x}_0, \boldsymbol{z}; \boldsymbol{\theta}^*, F^*)$ is a smooth function of the continuous components of $\boldsymbol{x}$, there exists a positive constant δ such that for any $\epsilon > 0$,

$$\begin{aligned} 1 &\geq \int \frac{\mathrm{E}\{R f_{\boldsymbol{z}}(\boldsymbol{Z}) | \boldsymbol{Z} = \boldsymbol{z}\}}{g_1(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\theta}^*, F^*) + \epsilon} d\boldsymbol{x} \geq \int_{\|\boldsymbol{x} - \boldsymbol{x}_0\| \leq \delta} \frac{\mathrm{E}\{R f_{\boldsymbol{z}}(\boldsymbol{Z}) | \boldsymbol{Z} = \boldsymbol{z}\}}{|g_1(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\theta}^*, F^*)| + \epsilon} d\boldsymbol{x} \\ &\gtrsim \int_{\|\boldsymbol{x} - \boldsymbol{x}_0\| \leq \delta} \frac{\mathrm{E}\{R f_{\boldsymbol{z}}(\boldsymbol{Z}) | \boldsymbol{Z} = \boldsymbol{z}\}}{\|\boldsymbol{x} - \boldsymbol{x}_0\| + \epsilon} d\boldsymbol{x}, \end{aligned} \tag{S.10}$$

where "$\gtrsim$" means greater than or equal to up to a constant. Because $\int_{\|\boldsymbol{x} - \boldsymbol{x}_0\| \leq \delta} (1/\|\boldsymbol{x} - \boldsymbol{x}_0\|) d\boldsymbol{x}$ is infinite, the last integration in expression (S.10) also goes to ∞ when $\epsilon \to 0$, which is a contradiction. Thus, $g_1(\boldsymbol{x}_0, \boldsymbol{z}; \boldsymbol{\theta}^*, F^*)$ is bounded away from zero for $(\boldsymbol{x}, \boldsymbol{z})$ in the support of $(\boldsymbol{X}, \boldsymbol{Z})$. The same conclusion holds for $(ns_n)^{-1} g_{1n}(\boldsymbol{x}, \boldsymbol{z}; \widehat{\boldsymbol{\theta}}, \widehat{F})$ when n is sufficiently large.

The final step is to prove that $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$ and $F^* = F_0$ through the Kullback-Leibler inequality. Let

$$\widetilde{p}_{kj} = \frac{\sum_{i=1}^{n} R_i I(\boldsymbol{X}_i = \boldsymbol{x}_k) B_j^q(\boldsymbol{Z}_i) / P(R_i = 1 | Y_i, \boldsymbol{Z}_i, \boldsymbol{W}_i)}{\sum_{i=1}^{n} R_i B_j^q(\boldsymbol{Z}_i) / P(R_i = 1 | Y_i, \boldsymbol{Z}_i, \boldsymbol{W}_i)},$$

and let $\widetilde{F}(\boldsymbol{x}, \boldsymbol{z}) = n^{-1} \sum_{k=1}^{m} \sum_{i=1}^{n} I(\boldsymbol{x}_k \leq \boldsymbol{x}, \boldsymbol{Z}_i \leq \boldsymbol{z}) \sum_{j=1}^{s_n} B_j^q(\boldsymbol{Z}_i) \widetilde{p}_{kj}$. By the approximation theory of B-splines (Schumaker 1981), $\widetilde{F}(\boldsymbol{x}, \boldsymbol{z}) \to F_0(\boldsymbol{x}, \boldsymbol{z})$ uniformly. Furthermore, it follows from the definitions of $\widehat{F}$ and $\widetilde{F}$ that $\widehat{F}$ is absolutely continuous with respect to $\widetilde{F}$. Thus, $d\widehat{F}/d\widetilde{F}$ converges uniformly to dF^*/dF_0. By Condition (C.3), F^* is continuously differentiable with respect to $\boldsymbol{x}$ and $\boldsymbol{z}$.

By the definitions of $\widehat{\boldsymbol{\theta}}$ and $\{\widehat{p}_{kj}\}$, we have $n^{-1} l_n(\widehat{\boldsymbol{\theta}}, \{\widehat{p}_{kj}\}) \geq n^{-1} l_n(\boldsymbol{\theta}_0, \{\widetilde{p}_{kj}\})$, i.e.,

$$
\begin{aligned}
&- n^{-1} \sum_{i=1}^{n} R_i \log \frac{P_{\widehat{\boldsymbol{\theta}}}(Y_i|\boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{W}_i)}{P_{\boldsymbol{\theta}_0}(Y_i|\boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{W}_i)} - n^{-1} \sum_{i=1}^{n} R_i \sum_{k=1}^{m} I(\boldsymbol{X}_i = \boldsymbol{x}_k) \sum_{j=1}^{s_n} B_j^q(\boldsymbol{Z}_i) \log \frac{\widehat{p}_{kj}}{\widetilde{p}_{kj}} \\
&- n^{-1} \sum_{i=1}^{n} (1 - R_i) \log \frac{\int P_{\widehat{\boldsymbol{\theta}}}(Y_i|\boldsymbol{x}, \boldsymbol{Z}_i, \boldsymbol{W}_i) \widehat{F}(d\boldsymbol{x}, \boldsymbol{Z}_i)}{\int P_{\boldsymbol{\theta}_0}(Y_i|\boldsymbol{x}, \boldsymbol{Z}_i, \boldsymbol{W}_i) \widetilde{F}(d\boldsymbol{x}, \boldsymbol{Z}_i)} \leq 0. \qquad \text{(S.11)}
\end{aligned}
$$

The first term in expression (S.11) converges to

$$
-\mathrm{E}\left\{ R \log \frac{P_{\boldsymbol{\theta}^*}(Y|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{W})}{P_{\boldsymbol{\theta}_0}(Y|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{W})} \right\}. \qquad \text{(S.12)}
$$

By the approximation theory of B-splines (Schumaker 1981), $\sum_{j=1}^{s_n} B_j^q(\boldsymbol{z}) \log(\widehat{p}_{kj}/\widetilde{p}_{kj})$ is asymptotically equivalent to

$$
\log \frac{\sum_{j=1}^{s_n} B_j^q(\boldsymbol{z}) \widehat{p}_{kj}}{\sum_{j=1}^{s_n} B_j^q(\boldsymbol{z}) \widetilde{p}_{kj}} = \log \left. \frac{d\widehat{F}(\boldsymbol{x}, \boldsymbol{z})}{d\widetilde{F}(\boldsymbol{x}, \boldsymbol{z})} \right|_{\boldsymbol{x}=\boldsymbol{x}_k}.
$$

Thus $\sum_{j=1}^{s_n} B_j^q(\boldsymbol{z}) \log(\widehat{p}_{kj}/\widetilde{p}_{kj})$ converges uniformly to $\log\{dF^*(\boldsymbol{x}, \boldsymbol{z})/dF_0(\boldsymbol{x}, \boldsymbol{z})\}|_{\boldsymbol{x}=\boldsymbol{x}_k}$. As a result, the second term in expression (S.11) converges to

$$
-\mathrm{E}\left\{ R \log \frac{dF^*(\boldsymbol{X}, \boldsymbol{Z})}{dF_0(\boldsymbol{X}, \boldsymbol{Z})} \right\}. \qquad \text{(S.13)}
$$

The third term in expression (S.11) converges to

$$
-\mathrm{E}\left\{ (1 - R) \log \frac{\int P_{\boldsymbol{\theta}^*}(Y|\boldsymbol{x}, \boldsymbol{Z}, \boldsymbol{W}) F^*(d\boldsymbol{x}, \boldsymbol{Z})}{\int P_{\boldsymbol{\theta}_0}(Y|\boldsymbol{x}, \boldsymbol{Z}, \boldsymbol{W}) F_0(d\boldsymbol{x}, \boldsymbol{Z})} \right\}. \qquad \text{(S.14)}
$$

By combining expressions (S.12), (S.13), and (S.14), we conclude that the Kullback-Leibler information of the density indexed by $\boldsymbol{\theta}^*$ and F^* with respect to the true density is nonpositive and thus must be zero. Therefore, the two densities are identical almost surely. For $R = 1$, this implies that $P_{\boldsymbol{\theta}^*}(Y|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{W}) F^*(\boldsymbol{X}, \boldsymbol{Z}) = P_{\boldsymbol{\theta}_0}(Y|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{W}) F_0(\boldsymbol{X}, \boldsymbol{Z})$. It follows from Condition (C.2) that $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$ and $F^* = F_0$. Thus, Theorem S.1 holds. ■

Theorem S.2. Under Conditions (C.1)–(C.5), $n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ converges in distribution to a zero-mean normal random vector whose covariance matrix attains the semiparametric efficiency bound.

Proof. Let $l_{\boldsymbol{\theta}}$ denote the score function for $\boldsymbol{\theta}_0$ and $l_F(h)$ denote the score function along the submodel $\{1 + \epsilon h(\boldsymbol{x}, \boldsymbol{z})\} dF_0(\boldsymbol{x}, \boldsymbol{z})$ based on one complete observation $(Y, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{W})$, where $h \in L_2(\mathcal{P})$, $\mathcal{P}$ is the probability measure indexed by $(\boldsymbol{\theta}_0, F_0)$, and $\mathrm{E}\{h(\boldsymbol{X}, \boldsymbol{Z})\} = 0$. We have $l_{\boldsymbol{\theta}} = \partial \log P_{\boldsymbol{\theta}_0}(Y|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{W})/\partial\boldsymbol{\theta}$ and $l_F(h) = h$. For two-phase studies, the score operators are $l^o_{\boldsymbol{\theta}} = Rl_{\boldsymbol{\theta}} + (1-R)\mathrm{E}(l_{\boldsymbol{\theta}}|Y, \boldsymbol{Z}, \boldsymbol{W})$ and $l^o_F = Rl_F + (1-R)\mathrm{E}(l_F|Y, \boldsymbol{Z}, \boldsymbol{W})$. The information operator is

$$\begin{bmatrix} l^{o*}_{\boldsymbol{\theta}} l^o_{\boldsymbol{\theta}} & l^{o*}_{\boldsymbol{\theta}} l^o_F \\ l^{o*}_F l^o_{\boldsymbol{\theta}} & l^{o*}_F l^o_F \end{bmatrix},$$

where $l^{o*}_{\boldsymbol{\theta}}$ and l^{o*}_F are the adjoint operators of $l^o_{\boldsymbol{\theta}}$ and l^o_F, respectively. We calculate the information operator as

$$l^{o*}_{\boldsymbol{\theta}} l^o_{\boldsymbol{\theta}} = \mathrm{E}\left\{Rl_{\boldsymbol{\theta}}^{\otimes 2} + (1-R)\mathrm{E}(l_{\boldsymbol{\theta}}|Y, \boldsymbol{Z}, \boldsymbol{W})^{\otimes 2}\right\},$$

$$l^{o*}_{\boldsymbol{\theta}} l^o_F(h) = l^{o*}_F l^o_{\boldsymbol{\theta}}(h)^{\mathrm{T}} = \mathrm{E}\left[\mathrm{E}\left\{Rl_{\boldsymbol{\theta}} + (1-R)\mathrm{E}(l_{\boldsymbol{\theta}}|Y, \boldsymbol{Z}, \boldsymbol{W})|\boldsymbol{X}, \boldsymbol{Z}\right\} h(\boldsymbol{X}, \boldsymbol{Z})\right], \text{ and}$$

$$l^{o*}_F l^o_F(h) = \mathrm{E}(R|\boldsymbol{X}, \boldsymbol{Z})h(\boldsymbol{X}, \boldsymbol{Z}) + \mathrm{E}\left\{(1-R)\mathrm{E}(h(\boldsymbol{X}, \boldsymbol{Z})|Y, \boldsymbol{Z}, \boldsymbol{W})|\boldsymbol{X}, \boldsymbol{Z}\right\}.$$

This information operator is the sum of an invertible operator and a compact operator from the space $\mathbb{M} \equiv \mathbb{R}^d \times BV(\mathbb{D}_{\boldsymbol{x},\boldsymbol{z}})$ to itself, where d is the dimension of $\boldsymbol{\theta}$, and $BV(\mathbb{D}_{\boldsymbol{x},\boldsymbol{z}})$ is the space of functions with bounded total variation in the support of $(\boldsymbol{X}, \boldsymbol{Z})$. By Theorem 4.7 of Rudin (1973), the information operator is invertible if it is one to one, or equivalently, the Fisher information along any nontrivial submodel is nonzero.

Suppose that the Fisher information is zero along some submodel $[\boldsymbol{\theta}_0 + \epsilon\boldsymbol{v}, dF_0(\boldsymbol{x}, \boldsymbol{z})\{1 + \epsilon h(\boldsymbol{x}, \boldsymbol{z})\}]$. Then, the score function along this submodel, i.e., $l^{o\mathrm{T}}_{\boldsymbol{\theta}} \boldsymbol{v} + l^o_F(h)$, is zero. We set $R = 1$ to obtain $l^{\mathrm{T}}_{\boldsymbol{\theta}} \boldsymbol{v} + l_F(h) = 0$ for any $(Y, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{W}) \in \mathcal{C}$. Specifically, for any

$(y_i, \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{w}_i) \in \mathcal{C}$, $i = 1, 2$, we have

$$\left\{ \frac{\partial}{\partial \boldsymbol{\theta}} \log P_{\boldsymbol{\theta}_0}(y_1|\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{w}_1) \right\}^{\mathrm{T}} \boldsymbol{v} + h(\boldsymbol{x}, \boldsymbol{z}) = \left\{ \frac{\partial}{\partial \boldsymbol{\theta}} \log P_{\boldsymbol{\theta}_0}(y_2|\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{w}_2) \right\}^{\mathrm{T}} \boldsymbol{v} + h(\boldsymbol{x}, \boldsymbol{z}),$$

which can be rewritten as a linear equation on $\boldsymbol{v}$, i.e.,

$$\left\{ \frac{\partial}{\partial \boldsymbol{\theta}} \log P_{\boldsymbol{\theta}_0}(y_1|\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{w}_1) - \frac{\partial}{\partial \boldsymbol{\theta}} \log P_{\boldsymbol{\theta}_0}(y_2|\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{w}_2) \right\}^{\mathrm{T}} \boldsymbol{v} = 0.$$

By Condition (C.2), $\boldsymbol{v} = 0$ and $h = 0$ with probability one. Thus, the information operator is invertible. Consequently, there exists a function h such that $l_F^{o\,*} l_F^o(h) = l_F^{o\,*} l_{\boldsymbol{\theta}}^o$, i.e.,

$$\begin{aligned} &\mathrm{E}(R|\boldsymbol{X}, \boldsymbol{Z})h + \mathrm{E}\left\{(1-R)\mathrm{E}(h|Y, \boldsymbol{Z}, \boldsymbol{W})|\boldsymbol{X}, \boldsymbol{Z}\right\} \\ =&\mathrm{E}\left\{Rl_{\boldsymbol{\theta}} + (1-R)\mathrm{E}(l_{\boldsymbol{\theta}}|Y, \boldsymbol{Z}, \boldsymbol{W})|\boldsymbol{X}, \boldsymbol{Z}\right\}. \end{aligned} \tag{S.15}$$

This means that the least favorable direction for $\boldsymbol{\theta}_0$ exists. In addition, by using the arguments in the proof of Theorem 3.4 of Zeng (2005) and Conditions (C.3) and (C.4), we can show that h is q-times continuously differentiable.

Because $(\widehat{\boldsymbol{\theta}}, \widehat{F})$ maximizes expression (2), the derivatives of the log-likelihood function with respect to ϵ along the submodel $(\widehat{\boldsymbol{\theta}} + \epsilon\boldsymbol{v}, d\widehat{F})$ for any $\boldsymbol{v}$ and the submodel $\{\widehat{\boldsymbol{\theta}}, d\widehat{F}(1 + \epsilon h_n)\}$ must be zero, where h_n is the projection of h onto the tangent space of the sieve space. By the approximation theory of B-splines (Schumaker 1981), we have $\|h_n - h\|_{L_2} \lesssim s_n^{-q/d_{\boldsymbol{z}}}$. Therefore, $(\widehat{\boldsymbol{\theta}}, \widehat{F})$ is the solution to the functional $\Psi_n(\boldsymbol{\theta}, F) = 0$, where $\Psi_n(\boldsymbol{\theta}, F) = \Psi_{1n}(\boldsymbol{\theta}, F) - \Psi_{2n}(\boldsymbol{\theta}, F)$,

$$\begin{aligned} \Psi_{1n}(\boldsymbol{\theta}, F) = {} & \mathcal{P}_n \left\{ R \frac{\partial}{\partial \boldsymbol{\theta}} \log P_{\boldsymbol{\theta}}(Y|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{W}) \right\} \\ & + \mathcal{P}_n \left\{ (1-R) \int \frac{\partial}{\partial \boldsymbol{\theta}} \log P_{\boldsymbol{\theta}}(Y|\boldsymbol{x}, \boldsymbol{Z}, \boldsymbol{W}) g_2(Y, \boldsymbol{Z}, \boldsymbol{W}, \boldsymbol{x}; \boldsymbol{\theta}, F) F(d\boldsymbol{x}, \boldsymbol{Z}) \right\}, \\ \Psi_{2n}(\boldsymbol{\theta}, F) = {} & \mathcal{P}_n \left\{ R h_n(\boldsymbol{X}, \boldsymbol{Z}) \right\} \\ & + \mathcal{P}_n \left\{ (1-R) \int g_2(Y, \boldsymbol{Z}, \boldsymbol{W}, \boldsymbol{x}; \boldsymbol{\theta}, F) h_n(\boldsymbol{x}, \boldsymbol{Z}) F(d\boldsymbol{x}, \boldsymbol{Z}) \right\}, \end{aligned}$$

$\mathcal{P}_n$ is the empirical measure of the sample, and

$$g_2(Y, \boldsymbol{Z}, \boldsymbol{W}, \boldsymbol{x}; \boldsymbol{\theta}, F) = \frac{P_{\boldsymbol{\theta}}(Y|\boldsymbol{x}, \boldsymbol{Z}, \boldsymbol{W})}{\int P_{\boldsymbol{\theta}}(Y|\widetilde{\boldsymbol{x}}, \boldsymbol{Z}, \boldsymbol{W})F(d\widetilde{\boldsymbol{x}}, \boldsymbol{Z})}.$$

Let $\Psi(\boldsymbol{\theta}, F)$ be the same as $\Psi_n(\boldsymbol{\theta}, F)$ except that $\mathcal{P}_n$ is replaced by $\mathcal{P}$. Clearly, $\widehat{\boldsymbol{\theta}}$ satisfies the following equation:

$$n^{1/2}\left\{\Psi_n(\widehat{\boldsymbol{\theta}}, \widehat{F}) - \Psi(\widehat{\boldsymbol{\theta}}, \widehat{F})\right\} = -n^{1/2}\Psi(\widehat{\boldsymbol{\theta}}, \widehat{F}). \tag{S.16}$$

We wish to use Theorem 2.11.22 of van der Vaart and Wellner (1996) to show that

$$n^{1/2}\left\{\Psi_n(\widehat{\boldsymbol{\theta}}, \widehat{F}) - \Psi(\widehat{\boldsymbol{\theta}}, \widehat{F})\right\} = n^{1/2}(\mathcal{P}_n - \mathcal{P})\left\{l^o_{\boldsymbol{\theta}} - l^o_F(h_n)\right\} + o_p(1). \tag{S.17}$$

Note that the left-hand side of equation (S.17) is an empirical process of the following two classes of functions indexed by $(\widehat{\boldsymbol{\theta}}, \widehat{F})$:

$$\begin{aligned}
\mathcal{F}_{1n} =& \bigg\{ R\frac{\partial}{\partial\boldsymbol{\theta}}\log P_{\boldsymbol{\theta}}(Y|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{W}) + (1-R)\int \frac{\partial}{\partial\boldsymbol{\theta}}\log P_{\boldsymbol{\theta}}(Y|\boldsymbol{x}, \boldsymbol{Z}, \boldsymbol{W}) \\
&\times g_2(Y, \boldsymbol{Z}, \boldsymbol{W}, \boldsymbol{x}; \boldsymbol{\theta}, F)F(d\boldsymbol{x}, \boldsymbol{Z}) : |\boldsymbol{\theta} - \boldsymbol{\theta}_0| + \|F - F_0\| \le \epsilon_0 \bigg\}; \\
\mathcal{F}_{2n} =& \bigg\{ Rh_n(\boldsymbol{X}, \boldsymbol{Z}) + (1-R)\int g_2(Y, \boldsymbol{Z}, \boldsymbol{W}, \boldsymbol{x}; \boldsymbol{\theta}, F)h_n(\boldsymbol{x}, \boldsymbol{Z})F(d\boldsymbol{x}, \boldsymbol{Z}) : \\
&|\boldsymbol{\theta} - \boldsymbol{\theta}_0| + \|F - F_0\| \le \epsilon_0 \bigg\},
\end{aligned}$$

where $\|F - F_0\|$ is the supreme norm in $\mathbb{D}_{\boldsymbol{x},\boldsymbol{z}}$. By Theorem S.1 and the approximation theory of B-splines (Schumaker, 1981), it is straightforward to verify that

$$\begin{aligned}
& R\frac{\partial}{\partial\boldsymbol{\theta}}\log P_{\widehat{\boldsymbol{\theta}}}(Y|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{W}) + (1-R)\int \frac{\partial}{\partial\boldsymbol{\theta}}\log P_{\widehat{\boldsymbol{\theta}}}(Y|\boldsymbol{x}, \boldsymbol{Z}, \boldsymbol{W}) \\
&\quad \times g_2(Y, \boldsymbol{Z}, \boldsymbol{W}, \boldsymbol{x}; \widehat{\boldsymbol{\theta}}, \widehat{F})\widehat{F}(d\boldsymbol{x}, \boldsymbol{Z}) \\
\to\ & R\frac{\partial}{\partial\boldsymbol{\theta}}\log P_{\boldsymbol{\theta}_0}(Y|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{W}) + (1-R)\int \frac{\partial}{\partial\boldsymbol{\theta}}\log P_{\boldsymbol{\theta}_0}(Y|\boldsymbol{x}, \boldsymbol{Z}, \boldsymbol{W}) \\
&\quad \times \frac{P_{\boldsymbol{\theta}_0}(Y|\boldsymbol{x}, \boldsymbol{Z}, \boldsymbol{W})F_0(d\boldsymbol{x}, \boldsymbol{Z})}{\int P_{\boldsymbol{\theta}_0}(Y|\boldsymbol{x}, \boldsymbol{Z}, \boldsymbol{W})F_0(d\boldsymbol{x}, \boldsymbol{Z})} \\
=\ & Rl_{\boldsymbol{\theta}} + (1-R)\mathrm{E}\{l_{\boldsymbol{\theta}}|Y, \boldsymbol{Z}, \boldsymbol{W}\} = l^o_{\boldsymbol{\theta}},
\end{aligned}$$

and

$$
\begin{aligned}
&Rh_n(\boldsymbol{X},\boldsymbol{Z}) + (1-R)\int g_2(Y,\boldsymbol{Z},\boldsymbol{W},\boldsymbol{x};\widehat{\boldsymbol{\theta}},\widehat{F})h_n(\boldsymbol{x},\boldsymbol{Z})\widehat{F}(d\boldsymbol{x},\boldsymbol{Z})\\
\to\ & Rh(\boldsymbol{X},\boldsymbol{Z}) + (1-R)\frac{\int h(\boldsymbol{x},\boldsymbol{Z})P_{\boldsymbol{\theta}_0}(Y|\boldsymbol{x},\boldsymbol{Z},\boldsymbol{W})F_0(d\boldsymbol{x},\boldsymbol{Z})}{\int P_{\boldsymbol{\theta}_0}(Y|\boldsymbol{x},\boldsymbol{Z},\boldsymbol{W})F_0(d\boldsymbol{x},\boldsymbol{Z})}\\
=\ & Rh(\boldsymbol{X},\boldsymbol{Z}) + (1-R)\mathrm{E}\left\{h(\boldsymbol{X},\boldsymbol{Z})|Y,\boldsymbol{Z},\boldsymbol{W}\right\} = l_F^o(h)
\end{aligned}
$$

uniformly in $(Y,\boldsymbol{X},\boldsymbol{Z},\boldsymbol{W})$.

Clearly, all functions in the classes $\mathcal{F}_{1n}$ and $\mathcal{F}_{2n}$ are uniformly bounded. We wish to verify the conditions in Theorem 2.11.22 of van der Vaart and Wellner (1996). We first show that the classes of functions $\mathcal{F}_{1n}$ and $\mathcal{F}_{2n}$ satisfy the uniform entropy condition. Pick any two functions from $\mathcal{F}_{1n}$, say f_1 and f_2, which are indexed by $(\boldsymbol{\theta}_1, F_1)$ and $(\boldsymbol{\theta}_2, F_2)$, respectively. The difference between the two functions is bounded from above by

$$
\begin{aligned}
&\left|\frac{\partial}{\partial\boldsymbol{\theta}}\log P_{\boldsymbol{\theta}_1}(Y|\boldsymbol{X},\boldsymbol{Z},\boldsymbol{W}) - \frac{\partial}{\partial\boldsymbol{\theta}}\log P_{\boldsymbol{\theta}_2}(Y|\boldsymbol{X},\boldsymbol{Z},\boldsymbol{W})\right|\\
&+\left|\int\frac{\partial}{\partial\boldsymbol{\theta}}\log P_{\boldsymbol{\theta}_1}(Y|\boldsymbol{x},\boldsymbol{Z},\boldsymbol{W})g_2(Y,\boldsymbol{Z},\boldsymbol{W},\boldsymbol{x};\boldsymbol{\theta}_1,F_1)(F_1-F_2)(d\boldsymbol{x},\boldsymbol{Z})\right|\\
&+\left|\int\left\{\frac{\partial}{\partial\boldsymbol{\theta}}\log P_{\boldsymbol{\theta}_1}(Y|\boldsymbol{x},\boldsymbol{Z},\boldsymbol{W}) - \frac{\partial}{\partial\boldsymbol{\theta}}\log P_{\boldsymbol{\theta}_2}(Y|\boldsymbol{x},\boldsymbol{Z},\boldsymbol{W})\right\}\right.\\
&\qquad\left.\times\, g_2(Y,\boldsymbol{Z},\boldsymbol{W},\boldsymbol{x};\boldsymbol{\theta}_1,F_1)F_2(d\boldsymbol{x},\boldsymbol{Z})\right|\\
&+\left|\int\frac{\partial}{\partial\boldsymbol{\theta}}\log P_{\boldsymbol{\theta}_2}(Y|\boldsymbol{x},\boldsymbol{Z},\boldsymbol{W})\Big\{g_2(Y,\boldsymbol{Z},\boldsymbol{W},\boldsymbol{x};\boldsymbol{\theta}_1,F_1)\right.\\
&\qquad\left.-\, g_2(Y,\boldsymbol{Z},\boldsymbol{W},\boldsymbol{x};\boldsymbol{\theta}_2,F_2)\Big\}F_2(d\boldsymbol{x},\boldsymbol{Z})\right|\\
&\equiv (i) + (ii) + (iii) + (iv).
\end{aligned}
$$

By the mean-value theorem, $(i) \lesssim \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|$. Because the denominator in the expression of $g_2(Y,\boldsymbol{Z},\boldsymbol{W},\boldsymbol{x};\boldsymbol{\theta},F)$ is bounded away from zero, we obtain that

$$
(ii) \lesssim \int |F_1(\boldsymbol{x},\boldsymbol{Z}) - F_2(\boldsymbol{x},\boldsymbol{Z})|\,d\boldsymbol{x} \lesssim \int |F_1(\boldsymbol{x},\boldsymbol{z}) - F_2(\boldsymbol{x},\boldsymbol{z})|\,d\boldsymbol{x}d\boldsymbol{z}.
$$

By the mean-value theorem,

$$(iii) \lesssim \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| \int g_2(Y, \boldsymbol{Z}, \boldsymbol{W}, \boldsymbol{x}; \boldsymbol{\theta}_1, F_1) F_2(d\boldsymbol{x}, \boldsymbol{Z}) \lesssim \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|.$$

Likewise,

$$(iv) \lesssim \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| + \int |F_1(\boldsymbol{x}, \boldsymbol{z}) - F_2(\boldsymbol{x}, \boldsymbol{z})| d\boldsymbol{x} d\boldsymbol{z}.$$

Combining the above inequalities for (i), (ii), (iii), and (iv), we have

$$|f_1 - f_2| \lesssim \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| + \int |F_1(\boldsymbol{x}, \boldsymbol{z}) - F_2(\boldsymbol{x}, \boldsymbol{z})| d\boldsymbol{x} d\boldsymbol{z}.$$

Thus, the Cauchy-Schwartz inequality implies that, for any finite measure $\mathcal{Q}$,

$$\begin{aligned} \|f_1 - f_2\|_{L_2(\mathcal{Q})} &\lesssim \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| + \left\{ \int |F_1(\boldsymbol{x}, \boldsymbol{z}) - F_2(\boldsymbol{x}, \boldsymbol{z})|^2 d\boldsymbol{x} d\boldsymbol{z} \right\}^{1/2} \\ &= \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| + \|F_1(\boldsymbol{X}, \boldsymbol{Z}) - F_2(\boldsymbol{X}, \boldsymbol{Z})\|_{L_2(\widetilde{\mathcal{Q}})}, \end{aligned} \tag{S.18}$$

where $\widetilde{\mathcal{Q}}$ is the uniform measure on $\mathbb{D}_{\boldsymbol{x},\boldsymbol{z}}$. We conclude that

$$\begin{aligned} N\{\epsilon, \mathcal{F}_{1n}, L_2(\mathcal{Q})) \lesssim & N(\epsilon/2, (\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < \epsilon_0), |\cdot|\} \\ & \times N(\epsilon/2, (F : \|F - F_0\|_\infty < \epsilon_0), L_2(\widetilde{\mathcal{Q}})\}, \end{aligned} \tag{S.19}$$

where $N(\cdot, \cdot, \cdot)$ denotes the covering number. On the right-hand side of (S.19), the first covering number is $O(1/\epsilon^d)$. The second covering number is $O[\exp\{\epsilon^{-2V/(V+2)}\}]$, where V is some positive index. To see the latter result, we observe that $(F : \|F - F_0\|_\infty < \epsilon)$ is in the symmetric convex hull of a Vapnik-Chervonenkis class $[I\{\boldsymbol{a} < (\boldsymbol{X}^{\mathrm{T}}, \boldsymbol{Z}^{\mathrm{T}})^{\mathrm{T}} \leq \boldsymbol{b}\}\colon \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^{d_{\boldsymbol{x}}+d_{\boldsymbol{z}}}]$, where $d_{\boldsymbol{x}}$ denotes the dimension of $\boldsymbol{X}$. The result follows from Theorem 2.6.9 of van der Vaart and Wellner (1996). Therefore, expression (S.19) implies that $\mathcal{F}_{1n}$ satisfies the uniform entropy condition in Theorem 2.11.22 of van der Vaart and Wellner (1996). By similar arguments and the fact that $\|h_n\|_{L_2} \lesssim \|h\|_{L_2}$, we can show that $\mathcal{F}_{2n}$ also satisfies the uniform entropy condition.

If we replace measure $\mathcal{Q}$ by $\mathcal{P}$, then expression (S.18) implies that the functions in

$\mathcal{F}_{1n}$ and $\mathcal{F}_{2n}$ are Lipschitz continuous with respect to $(\boldsymbol{\theta}, F)$ in the metric defined as

$$\rho\{(\boldsymbol{\theta}_1, F_1), (\boldsymbol{\theta}_2, F_2)\} = \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| + \|F_1 - F_2\|_{L_2(\mathcal{P})}.$$

As a result, condition (2.11.21) in Theorem 2.11.22 of van der Vaart and Wellner (1996) holds. In addition, the total boundedness of the index set (θ, F) holds due to the precompactness of $(\boldsymbol{\theta}, F)$ under the uniform metric. We have now verified all of the conditions in Theorem 2.11.22 of van der Vaart and Wellner (1996). Thus, equation (S.17) follows from that theorem.

By combining equations (S.16) and (S.17), we have

$$-n^{1/2}\left\{\Psi_1(\widehat{\boldsymbol{\theta}}, \widehat{F}) - \Psi_2(\widehat{\boldsymbol{\theta}}, \widehat{F})\right\} = n^{1/2}(\mathcal{P}_n - \mathcal{P})\{l^o_{\boldsymbol{\theta}} - l^o_F(h_n)\} + o_p(1), \tag{S.20}$$

where $\Psi_1(\boldsymbol{\theta}, F)$ and $\Psi_2(\boldsymbol{\theta}, F)$ are the same as $\Psi_{1n}(\boldsymbol{\theta}, F)$ and $\Psi_{2n}(\boldsymbol{\theta}, F)$, respectively, except that $\mathcal{P}_n$ is replaced by $\mathcal{P}$. The left-hand side of equation (S.20) can be linearized around $(\boldsymbol{\theta}_0, F_0)$. Specifically,

$$\begin{aligned}
\Psi_1(\widehat{\boldsymbol{\theta}}, \widehat{F}) = \Psi_1(\boldsymbol{\theta}_0, F_0) &+ \mathcal{P}\{R\frac{\partial^2}{\partial \boldsymbol{\theta}^{\mathrm{T}} \partial \boldsymbol{\theta}} \log P_{\boldsymbol{\theta}^*}(Y|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{W})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\} \\
&+ \mathcal{P}\Bigg[(1-R)\int \frac{\partial}{\partial \boldsymbol{\theta}}\left\{\frac{\partial}{\partial \boldsymbol{\theta}} \log P_{\boldsymbol{\theta}^*}(Y|\boldsymbol{x}, \boldsymbol{Z}, \boldsymbol{W}) g_2(Y, \boldsymbol{Z}, \boldsymbol{W}, \boldsymbol{x}; \boldsymbol{\theta}^*, F^*)\right\} \\
&\quad \times (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\widehat{F}(d\boldsymbol{x}, \boldsymbol{Z})\Bigg] \\
&+ \mathcal{P}\Bigg[(1-R)\int \frac{\partial}{\partial \boldsymbol{\theta}} \log P_{\boldsymbol{\theta}^*}(Y|\boldsymbol{x}, \boldsymbol{Z}, \boldsymbol{W})\Bigg\{\frac{\partial}{\partial F} g_2(Y, \boldsymbol{Z}, \boldsymbol{W}, \boldsymbol{x}; \boldsymbol{\theta}^*, F^*) \\
&\quad \times F^*(d\boldsymbol{x}, \boldsymbol{Z})\Bigg\}(\widehat{F} - F_0)\Bigg],
\end{aligned}$$

where $\partial/\partial F$ denotes the pathwise derivative, and $(\boldsymbol{\theta}^*, F^*)$ lies between $(\widehat{\boldsymbol{\theta}}, \widehat{F})$ and $(\boldsymbol{\theta}_0, F_0)$. Similar expansions can be obtained for $\Psi_2(\widehat{\boldsymbol{\theta}}, \widehat{F})$. By the approximation theory of B-splines (Schumaker 1981), we can show that the left-hand side of (S.20) equals

$$\begin{aligned}
&- n^{1/2}\{1 + o_p(1)\} \mathrm{E}\left\{l^o_{\boldsymbol{\theta}\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + l^o_{\boldsymbol{\theta}F}(\widehat{F} - F_0) - l^o_{F\boldsymbol{\theta}}(h_n)(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) - l^o_{FF}(h_n, \widehat{F} - F_0)\right\} \\
&- n^{1/2}\{\Psi_1(\boldsymbol{\theta}_0, F_0) - \Psi_2(\boldsymbol{\theta}_0, F_0)\},
\end{aligned} \tag{S.21}$$

where $l^o_{\boldsymbol{\theta\theta}}$ is the derivative of $l^o_{\boldsymbol{\theta}}$ with respect to $\boldsymbol{\theta}$, $l^o_{\boldsymbol{\theta}F}(h)$ is the derivative of $l^o_{\boldsymbol{\theta}}$ with respect to F along the direction h, $l^o_{F\boldsymbol{\theta}}(h)$ is the derivative of $l^o_F(h)$ with respect to $\boldsymbol{\theta}$, and $l^o_{FF}(h_1, h_2)$ is the derivative of $l^o_F(h_1)$ with respect to F along the direction h_2.

Because we have chosen h to be the least favorable direction for $\boldsymbol{\theta}_0$ and $\|h_n - h\|_{L_2} \lesssim s_n^{-q/d_{\boldsymbol{z}}}$, we have $\mathrm{E}\{l^o_{FF}(h_n, \widehat{F}-F_0)\} = \mathrm{E}\{l^o_{\boldsymbol{\theta}F}(\widehat{F}-F_0)\}+O(s_n^{-q/d_{\boldsymbol{z}}})$ and $\mathrm{E}\{l^o_{F\boldsymbol{\theta}}(h_n)(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\} = \mathrm{E}\{l^o_{F\boldsymbol{\theta}}(h)(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\}+O(s_n^{-q/d_{\boldsymbol{z}}})$. Thus, by Condition (C.5), the first term in expression (S.21) is $n^{1/2}\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)+O(n^{1/2}s_n^{-q/d_{\boldsymbol{z}}}) = n^{1/2}\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)+o(1)$, where $\boldsymbol{\Sigma} = -\mathrm{E}\{l^o_{\boldsymbol{\theta\theta}}-l^o_{F\boldsymbol{\theta}}(h)\}$, which is an invertible matrix due to the invertibility of the information operator for $(\boldsymbol{\theta}_0, F_0)$. Because $\mathcal{P}\left\{R\partial \log P_{\boldsymbol{\theta}_0}(Y|\boldsymbol{X},\boldsymbol{Z},\boldsymbol{W})/\partial\boldsymbol{\theta}\right\} = 0$ and

$$\mathcal{P}\left\{(1-R)\int \frac{\partial}{\partial\boldsymbol{\theta}}\log P_{\boldsymbol{\theta}_0}(Y|\boldsymbol{x},\boldsymbol{Z},\boldsymbol{W})\frac{P_{\boldsymbol{\theta}_0}(Y|\boldsymbol{x},\boldsymbol{Z},\boldsymbol{W})F_0(d\boldsymbol{x},\boldsymbol{Z})}{\int P_{\boldsymbol{\theta}_0}(Y|\boldsymbol{x},\boldsymbol{Z},\boldsymbol{W})F_0(d\boldsymbol{x},\boldsymbol{Z})}\right\} = 0,$$

the last term in (S.21) equals zero. It follows from equation (S.20) that

$$n^{1/2}\{1+o_p(1)\}\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)+o_p(1) = n^{1/2}(\mathcal{P}_n-\mathcal{P})\left\{l^o_{\boldsymbol{\theta}} - l^o_F(h)\right\}.$$

Thus, we have established the asymptotic normality in Theorem S.2. Because $\boldsymbol{\Sigma}^{-1}\{l^o_{\boldsymbol{\theta}} - l^o_F(h)\}$ is the efficient influence function for $\boldsymbol{\theta}_0$, its limiting covariance matrix attains the semiparametric efficiency bound. ■

For a given $\boldsymbol{\theta}$, we define $\widehat{F}_{\boldsymbol{\theta}}$ as the joint distribution function of $(\boldsymbol{X},\boldsymbol{Z})$ that maximizes $l_n(\boldsymbol{\theta}, \{p_{kj}\})$. By the arguments in the proof of Theorem S.1, we can show that for any $\widehat{\boldsymbol{\theta}} \to \boldsymbol{\theta}_0$ in probability, the estimator $\widehat{F}_{\widehat{\boldsymbol{\theta}}} \to F_0$ uniformly. Furthermore, given the existence of the least favorable directions, we can construct the least favorable model. These two facts imply that the profile likelihood theory in Murphy and van der Vaart (2000) holds for our approach. Thus, the inverse of the negative Hessian matrix of the profile likelihood function is a consistent estimator for the limiting covariance matrix of $n^{1/2}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$.

REFERENCES

Murphy, S. A. and van der Vaart, A. W. (2000), "On Profile Likelihood," *Journal of the American Statistical Association*, 95, 449–465.

Rudin, W. (1973), *Functional Analysis*, New York: McGraw-Hill.

Schumaker, L. (1981), *Spline Functions: Basic Theory*, New York: Wiley-Interscience.

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

Zeng, D. (2005), "Likelihood Approach for Marginal Proportional Hazards Regression in the Presence of Dependent Censoring," *The Annals of Statistics*, 33, 501–521.

S.2. Supplementary Figures and Tables

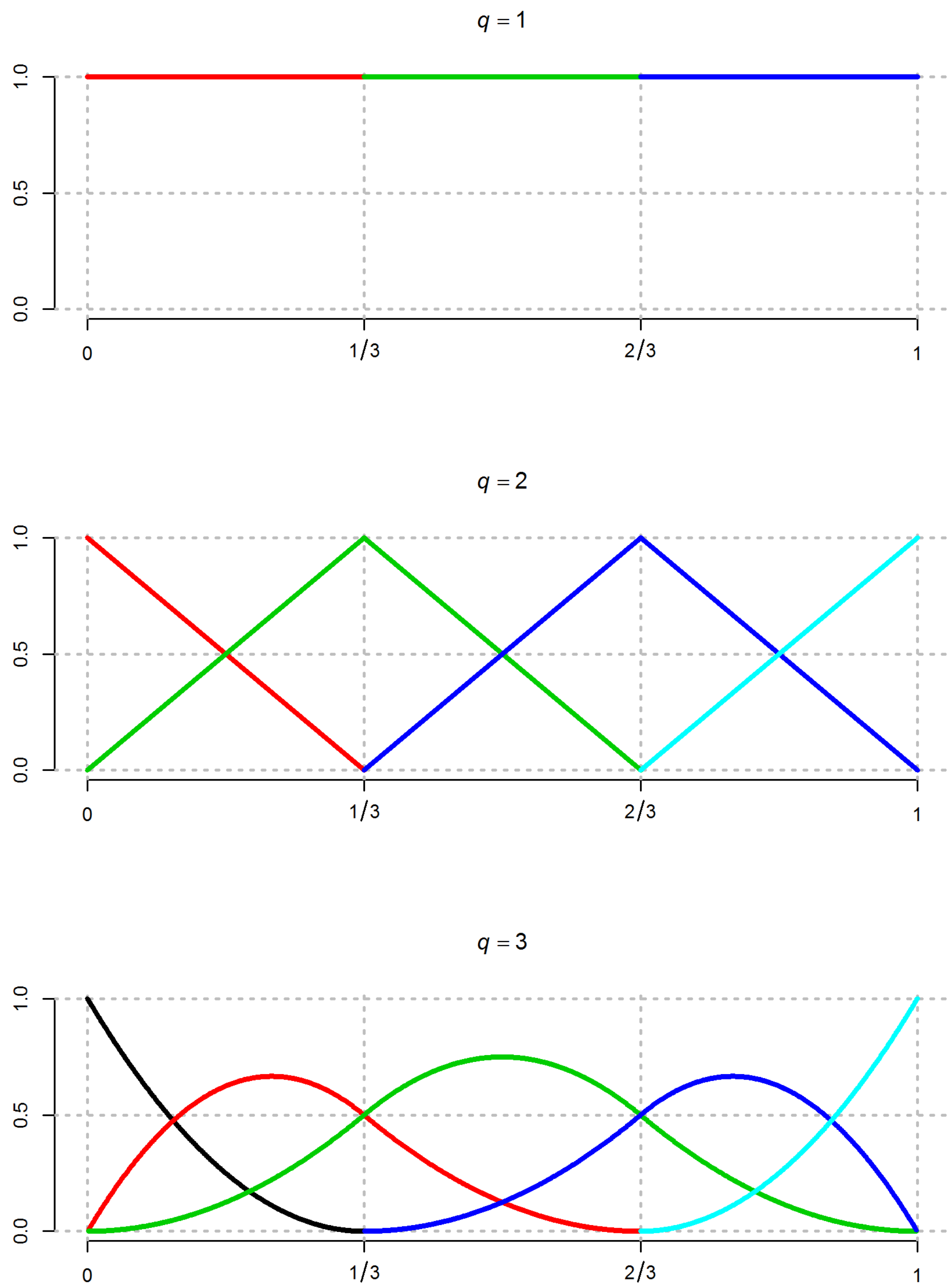

Figure S1. Plots of $\{N_l^q(z)\}_{l=-q+1}^{2}$ for $q = 1$, 2, and 3. The functions in each B-spline basis are distinguished by different colors.

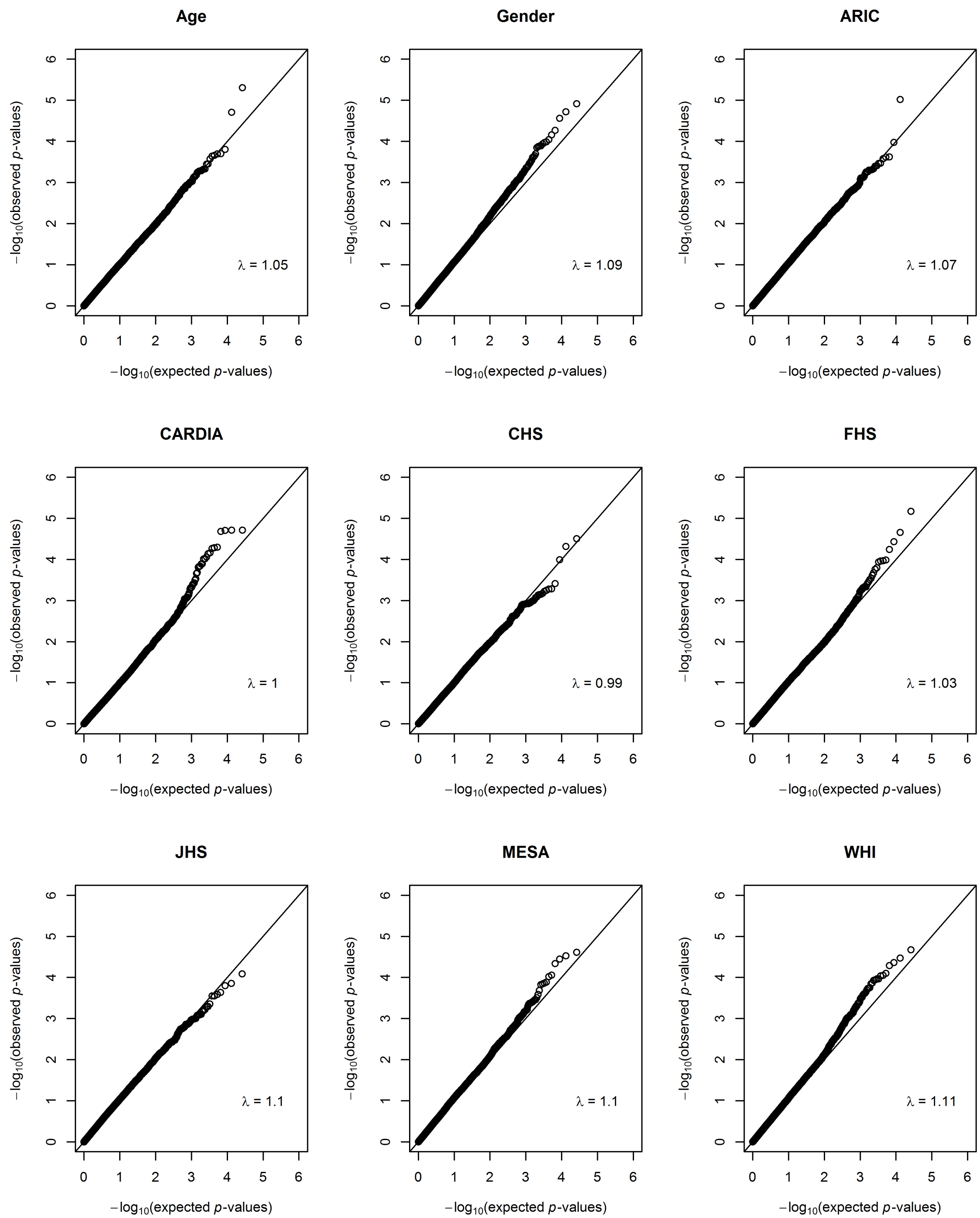

Figure S2. Quantile-quantile plots for the analysis of age, gender, and cohort indicators in the DPR group.

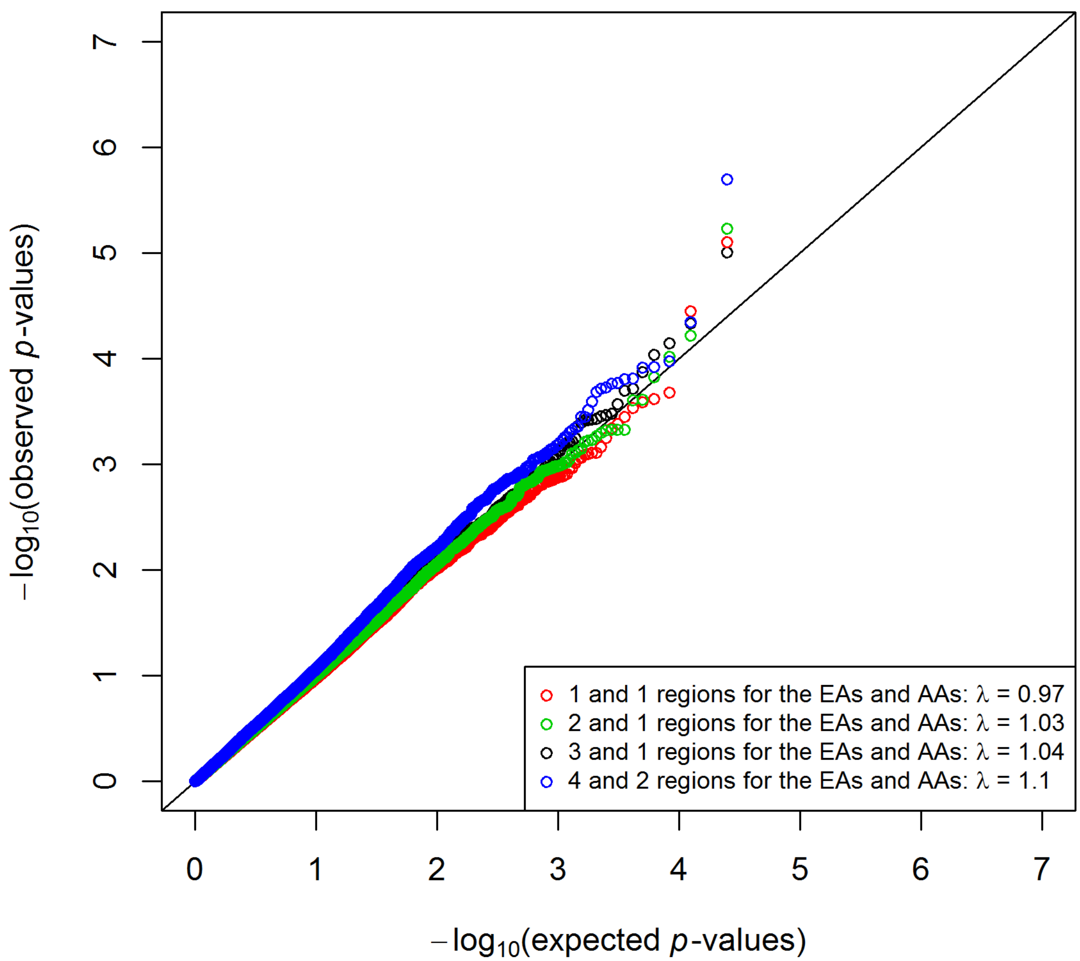

Figure S3. Quantile-quantile plots for the analysis of the BP study in the NHLBI ESP using the SMLE method with different numbers of sieve regions.

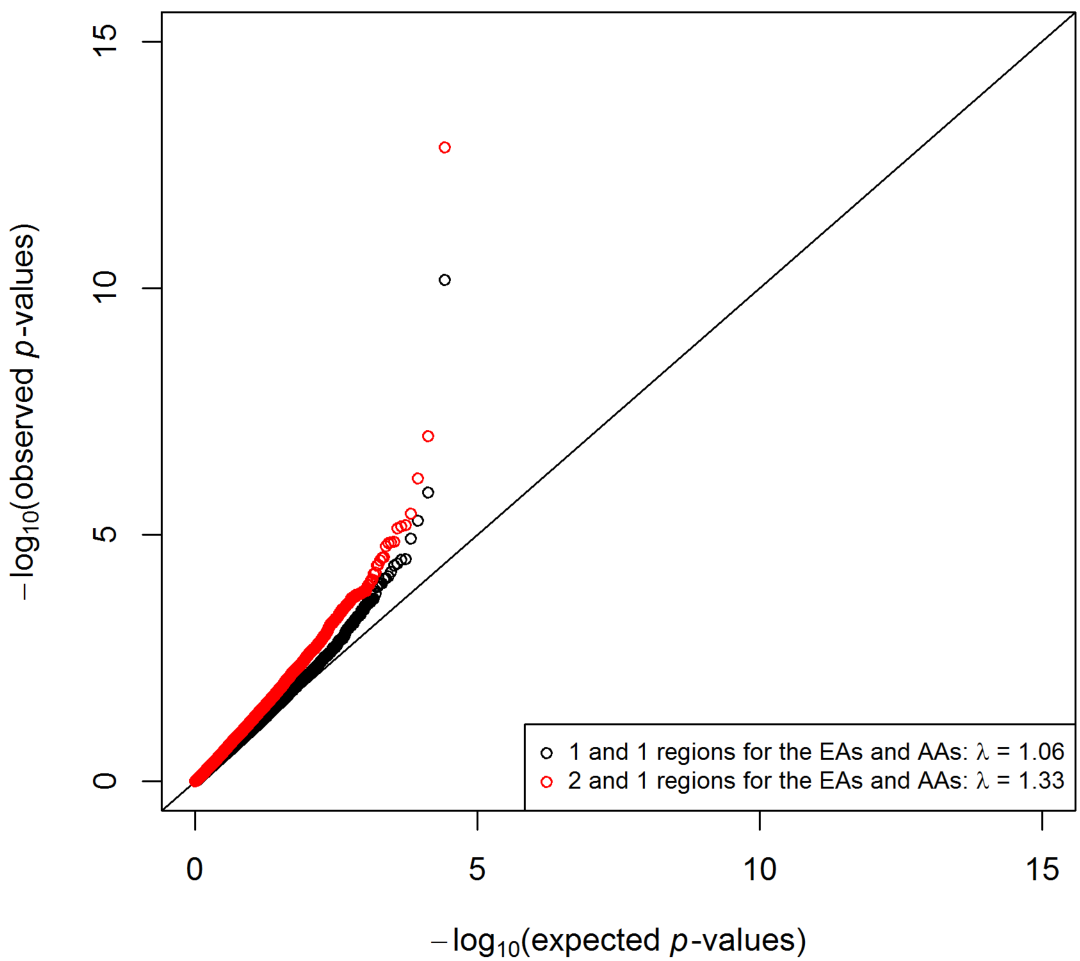

Figure S4. Quantile-quantile plots for the analysis of the LDL study in the NHLBI ESP using the SMLE method with different numbers of sieve regions.

Table S1. Additional Simulation Results Under the Model $Y = 0.5X + 0.5Z + 0.5W + \epsilon$ With the Second-Phase Sample Selection Depending Only on Y

			SMLE					MLE_0	
n	r	Covariate	Bias	SE	SEE	CP	RE	Bias	SE
4000	0.0	X	0.006	0.099	0.094	0.942	1.048	0.008	0.102
		Z	0.002	0.062	0.062	0.951	2.728	0.006	0.103
		W	-0.001	0.055	0.055	0.949	3.380	0.006	0.102
	0.1	X	0.008	0.100	0.095	0.940	1.040	0.008	0.102
		Z	0.006	0.062	0.062	0.949	2.820	0.006	0.104
		W	-0.001	0.055	0.055	0.950	3.422	0.006	0.103
	0.2	X	0.006	0.100	0.094	0.941	1.099	0.008	0.104
		Z	0.007	0.061	0.061	0.949	2.973	0.006	0.105
		W	-0.001	0.055	0.055	0.950	3.480	0.007	0.103
	0.3	X	0.003	0.099	0.095	0.941	1.155	0.008	0.107
		Z	0.006	0.061	0.061	0.950	2.971	0.006	0.105
		W	-0.001	0.056	0.055	0.949	3.509	0.006	0.104
6000	0.0	X	0.005	0.096	0.093	0.937	1.061	0.006	0.099
		Z	0.002	0.054	0.054	0.954	3.387	0.008	0.099
		W	0.000	0.045	0.045	0.952	4.740	0.007	0.098
	0.1	X	0.008	0.097	0.092	0.936	1.047	0.006	0.100
		Z	0.008	0.053	0.053	0.951	3.540	0.007	0.099
		W	0.000	0.045	0.045	0.951	4.790	0.007	0.099
	0.2	X	0.003	0.095	0.090	0.936	1.131	0.006	0.102
		Z	0.007	0.052	0.052	0.949	3.782	0.007	0.101
		W	0.000	0.045	0.045	0.951	4.926	0.007	0.100
	0.3	X	-0.001	0.095	0.089	0.936	1.217	0.006	0.104
		Z	0.007	0.052	0.052	0.948	3.785	0.007	0.101
		W	0.000	0.045	0.045	0.952	5.004	0.007	0.101

NOTE: Bias and SE are, respectively, the empirical bias and standard error of the parameter estimator; SEE is the empirical mean of the standard error estimator; CP is the coverage probability of the 95% confidence interval; RE is the empirical variance of MLE_0 over that of SMLE. Each entry is based on 10,000 replicates.

Table S2. Simulation Results When Z is Misclassified as Being Independent of X

r	Covariate	Bias	SE	SEE	CP
0.00	X	0.002	0.108	0.108	0.948
	Z	0.000	0.078	0.078	0.949
	W	-0.001	0.078	0.078	0.952
0.02	X	0.009	0.108	0.108	0.948
	Z	0.009	0.078	0.078	0.948
	W	-0.001	0.078	0.078	0.953
0.04	X	0.016	0.108	0.108	0.947
	Z	0.017	0.077	0.078	0.944
	W	-0.001	0.078	0.078	0.953
0.06	X	0.023	0.108	0.108	0.947
	Z	0.025	0.077	0.078	0.940
	W	-0.001	0.078	0.078	0.953
0.08	X	0.030	0.108	0.108	0.943
	Z	0.034	0.077	0.078	0.931
	W	-0.001	0.078	0.078	0.953
0.10	X	0.037	0.109	0.108	0.939
	Z	0.042	0.077	0.078	0.919
	W	-0.001	0.078	0.078	0.953
0.15	X	0.053	0.108	0.108	0.923
	Z	0.061	0.076	0.077	0.879
	W	-0.001	0.078	0.078	0.952
0.20	X	0.069	0.109	0.108	0.907
	Z	0.080	0.076	0.077	0.826
	W	-0.001	0.078	0.078	0.953

NOTE: See the Note to Table S1.

Table S3. Simulation Results Under the Model $Y = 0.5X + 0.5Z_1 + 0.5Z_2 + \epsilon$

Second-phase sampling	r	Covariate	Bias	SE	SEE	CP
Depend only on Y	0.0	X	0.006	0.094	0.091	0.945
		Z_1	-0.001	0.080	0.079	0.949
		Z_2	0.001	0.079	0.079	0.950
	0.1	X	0.015	0.094	0.092	0.940
		Z_1	0.002	0.079	0.079	0.948
		Z_2	0.004	0.078	0.079	0.948
	0.2	X	0.020	0.096	0.094	0.940
		Z_1	0.002	0.078	0.079	0.949
		Z_2	0.004	0.078	0.079	0.948
	0.3	X	0.022	0.100	0.098	0.934
		Z_1	0.001	0.078	0.079	0.954
		Z_2	0.002	0.078	0.079	0.951
Depend on (Y, Z_1, Z_2)	0.0	X	0.005	0.091	0.089	0.943
		Z_1	-0.001	0.079	0.079	0.950
		Z_2	0.000	0.079	0.079	0.951
	0.1	X	0.006	0.091	0.089	0.944
		Z_1	0.003	0.079	0.079	0.950
		Z_2	0.004	0.079	0.079	0.950
	0.2	X	0.005	0.093	0.092	0.946
		Z_1	0.005	0.079	0.079	0.948
		Z_2	0.006	0.078	0.079	0.952
	0.3	X	0.000	0.096	0.095	0.943
		Z_1	0.006	0.079	0.079	0.951
		Z_2	0.008	0.079	0.079	0.951

NOTE: See the Note to Table S1.