Supplementary material to A general framework for studying genetic effects and gene-environment interactions with missing data

Y. J. HU, D. Y. LIN*, D. ZENG

Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599-7420, USA

*email: lin@bios.unc.edu

1. ASYMPTOTIC PROPERTIES

We state in Theorems S.1-S.4 the asymptotic properties of the NPMLEs described in Sections 2.2-2.4 of the main paper and provide the proofs of the theorems. For each theorem, it is necessary to verify that the parameters are identifiable and the information matrices along all non-trivial parametric submodels are non-singular. We state those intermediate results in Lemmas S.1-S.8.

1.1 *Cross-sectional studies*

We impose the following conditions.

CONDITION S.1 If $P_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\xi}}(\mathbf{Y}|\mathbf{X},H) = P_{\widetilde{\boldsymbol{\alpha}},\widetilde{\boldsymbol{\beta}},\widetilde{\boldsymbol{\xi}}}(\mathbf{Y}|\mathbf{X},H)$ for any $H=(h,h)$ and $H=(h,h^\dagger)$, then $\boldsymbol{\alpha}=\widetilde{\boldsymbol{\alpha}}$, $\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}$, and $\boldsymbol{\xi}=\widetilde{\boldsymbol{\xi}}$.

CONDITION S.2 If there exists a constant vector $\boldsymbol{\nu}$ such that $\boldsymbol{\nu}^{\mathrm{T}}\nabla_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\xi}}\log P_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\xi}}(\mathbf{Y}|\mathbf{X},H) = 0$ for any $H=(h,h)$ and $H=(h,h^\dagger)$, then $\boldsymbol{\nu}=\mathbf{0}$.

CONDITION S.3 If there exists a function $a(H)$ and a constant vector $\mathbf{b}$ such that $a(H) + \mathbf{b}^{\mathrm{T}}\mathcal{D}(\mathbf{X},H) = 0$ with probability one, then $a=0$ and $\mathbf{b}=\mathbf{0}$.

REMARK S.1 Condition S.1 ensures that $(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\xi})$ are identifiable from the genotype data while Condition S.2 ensures nonsingularity of the information matrix. All commonly used regression models, particularly generalized linear (mixed) models with design vectors in the form of (2.1), satisfy these two conditions. Condition S.3 pertains to the identifiability of $\boldsymbol{\zeta}$.

This condition holds under all common modes of inheritance for the $\zeta_{s,k,l}$ provided that $\mathbf{X}$ is linearly independent given H.

LEMMA S.1 If two sets of parameters $(\boldsymbol{\theta}, F)$ and $(\widetilde{\boldsymbol{\theta}}, \widetilde{F})$ yield the same joint distribution of the data, then $\boldsymbol{\theta} = \widetilde{\boldsymbol{\theta}}$ and $F = \widetilde{F}$.

Proof: Suppose that

$$\sum_{H \in \mathcal{S}(G)} P_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\xi}}(\mathbf{Y}|\mathbf{X}, H) P_{\boldsymbol{\zeta},F}(\mathbf{X}|H) P_{\boldsymbol{\gamma}}(H) = \sum_{H \in \mathcal{S}(G)} P_{\widetilde{\boldsymbol{\alpha}},\widetilde{\boldsymbol{\beta}},\widetilde{\boldsymbol{\xi}}}(\mathbf{Y}|\mathbf{X}, H) P_{\widetilde{\boldsymbol{\zeta}},\widetilde{F}}(\mathbf{X}|H) P_{\widetilde{\boldsymbol{\gamma}}}(H).$$

Letting $G = 2h$ or $G = h + h^{\dagger}$ and integrating over $\mathbf{Y}$ on both sides, we obtain

$$P_{\boldsymbol{\zeta},F}(\mathbf{X}|H) P_{\boldsymbol{\gamma}}(H) = P_{\widetilde{\boldsymbol{\zeta}},\widetilde{F}}(\mathbf{X}|H) P_{\widetilde{\boldsymbol{\gamma}}}(H).$$

Integrating over $\mathbf{X}$ on both sides then yields that $P_{\boldsymbol{\gamma}}(H) = P_{\widetilde{\boldsymbol{\gamma}}}(H)$. By Lemma 1 of Lin and Zeng (2006), $\boldsymbol{\gamma} = \widetilde{\boldsymbol{\gamma}}$. Thus, $P_{\boldsymbol{\zeta},F}(\mathbf{X}|H) = P_{\widetilde{\boldsymbol{\zeta}},\widetilde{F}}(\mathbf{X}|H)$. It follows from the definition of $P_{\boldsymbol{\zeta},F}(\mathbf{X}|H)$ that

$$\exp\{(\boldsymbol{\zeta} - \widetilde{\boldsymbol{\zeta}})^{\mathrm{T}} \mathcal{D}(\mathbf{X}, H)\} \frac{f(\mathbf{X})}{\widetilde{f}(\mathbf{X})} = \frac{\int_{\mathbf{x}} \exp\{\boldsymbol{\zeta}^{\mathrm{T}} \mathcal{D}(\mathbf{x}, H)\} dF(\mathbf{x})}{\int_{\mathbf{x}} \exp\{\widetilde{\boldsymbol{\zeta}}^{\mathrm{T}} \mathcal{D}(\mathbf{x}, H)\} d\widetilde{F}(\mathbf{x})}.$$

By setting $H = (h_0, h_0')$, we obtain $\mathcal{D}(\mathbf{X}, H) = \mathbf{0}$, so the above equation reduces to $f(\mathbf{x}) = \widetilde{f}(\mathbf{x})$ for any $\mathbf{x}$. It then follows from Condition S.3 that $\boldsymbol{\zeta} = \widetilde{\boldsymbol{\zeta}}$. Therefore, $P_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\xi}}(Y|\mathbf{X}, H) = P_{\widetilde{\boldsymbol{\alpha}},\widetilde{\boldsymbol{\beta}},\widetilde{\boldsymbol{\xi}}}(Y|\mathbf{X}, H)$ for any $H = (h, h)$ or $H = (h, h^{\dagger})$. By Condition S.1, $\boldsymbol{\alpha} = \widetilde{\boldsymbol{\alpha}}$, $\boldsymbol{\beta} = \widetilde{\boldsymbol{\beta}}$ and $\boldsymbol{\xi} = \widetilde{\boldsymbol{\xi}}$.

LEMMA S.2 If there exist a vector $\boldsymbol{\mu}_{\boldsymbol{\theta}} \equiv (\boldsymbol{\mu}_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\xi}}^{\mathrm{T}}, \boldsymbol{\mu}_{\boldsymbol{\gamma}}^{\mathrm{T}}, \boldsymbol{\mu}_{\boldsymbol{\zeta}}^{\mathrm{T}})^{\mathrm{T}}$ and a function $\psi(\mathbf{x})$ with $E[\psi(\mathbf{X})] = 0$ such that $\boldsymbol{\mu}_{\boldsymbol{\theta}}^{\mathrm{T}} l_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, F_0) + l_{F_0}(\boldsymbol{\theta}_0, F_0)[\int \psi \, dF_0] = 0$, where $l_{\boldsymbol{\theta}}$ is the score function for $\boldsymbol{\theta}$, and $l_{F_0}[\int \psi \, dF_0]$ is the score function for F along the submodel $F_0 + \epsilon \int \psi \, dF_0$ with scalar ϵ, then $\boldsymbol{\mu}_{\boldsymbol{\theta}} = \mathbf{0}$ and $\psi = 0$.

Proof: We wish to verify that if there exist a vector $\boldsymbol{\mu}_{\boldsymbol{\theta}} \equiv (\boldsymbol{\mu}_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\xi}}^{\mathrm{T}}, \boldsymbol{\mu}_{\boldsymbol{\gamma}}^{\mathrm{T}}, \boldsymbol{\mu}_{\boldsymbol{\zeta}}^{\mathrm{T}})^{\mathrm{T}}$ and a function $\psi(\mathbf{x})$ with $E[\psi(\mathbf{X})] = 0$ such that

$$\boldsymbol{\mu}_{\boldsymbol{\theta}}^{\mathrm{T}} l_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, F_0) + l_{F_0}(\boldsymbol{\theta}_0, F_0)[\int \psi \, dF_0] = 0, \tag{S.1}$$

where $l_{\boldsymbol{\theta}}$ is the score function for $\boldsymbol{\theta}$, and $l_{F_0}[\int \psi \, dF_0]$ is the score function for F along the submodel $F_0 + \epsilon \int \psi \, dF_0$ with scalar ϵ, then $\boldsymbol{\mu}_{\boldsymbol{\theta}} = \mathbf{0}$ and $\psi = 0$. To this end, we set $G = 2h$ or $G = h + h^{\dagger}$. Then (S.1) becomes

$$\begin{aligned} &\boldsymbol{\mu}_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\xi}}^{\mathrm{T}} \nabla_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\xi}} \log P_{\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0,\boldsymbol{\xi}_0}(\mathbf{Y}|\mathbf{X}, H) + \boldsymbol{\mu}_{\boldsymbol{\gamma}}^{\mathrm{T}} \nabla_{\boldsymbol{\gamma}} \log P_{\boldsymbol{\gamma}_0}(H) \\ &+\boldsymbol{\mu}_{\boldsymbol{\zeta}}^{\mathrm{T}} \mathcal{D}(\mathbf{X}, H) - \frac{\boldsymbol{\mu}_{\boldsymbol{\zeta}}^{\mathrm{T}} \int_{\mathbf{x}} \exp\{\boldsymbol{\zeta}_0^{\mathrm{T}} \mathcal{D}(\mathbf{x}, H)\} \mathcal{D}(\mathbf{x}, H) dF_0(\mathbf{x})}{\int_{\mathbf{x}} \exp\{\boldsymbol{\zeta}_0^{\mathrm{T}} \mathcal{D}(\mathbf{x}, H)\} dF_0(\mathbf{x})} \\ &+\psi(\mathbf{X}) - \frac{\int_{\mathbf{x}} \exp\{\boldsymbol{\zeta}_0^{\mathrm{T}} \mathcal{D}(\mathbf{x}, H)\} \psi(\mathbf{x}) dF_0(\mathbf{x})}{\int_{\mathbf{x}} \exp\{\boldsymbol{\zeta}_0^{\mathrm{T}} \mathcal{D}(\mathbf{x}, H)\} dF_0(\mathbf{x})} = 0. \end{aligned} \tag{S.2}$$

Taking the expectation with respect to $P_{\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0,\boldsymbol{\xi}_0}(\mathbf{Y}|\mathbf{X}, H)$ yields

$$\begin{aligned} &\boldsymbol{\mu}_{\boldsymbol{\gamma}}^{\mathrm{T}} \nabla_{\boldsymbol{\gamma}} \log P_{\boldsymbol{\gamma}_0}(H) + \boldsymbol{\mu}_{\boldsymbol{\zeta}}^{\mathrm{T}} \mathcal{D}(\mathbf{X}, H) - \frac{\boldsymbol{\mu}_{\boldsymbol{\zeta}}^{\mathrm{T}} \int_{\mathbf{x}} \exp\{\boldsymbol{\zeta}_0^{\mathrm{T}} \mathcal{D}(\mathbf{x}, H)\} \mathcal{D}(\mathbf{x}, H) dF_0(\mathbf{x})}{\int_{\mathbf{x}} \exp\{\boldsymbol{\zeta}_0^{\mathrm{T}} \mathcal{D}(\mathbf{x}, H)\} dF_0(\mathbf{x})} \\ &+\psi(\mathbf{X}) - \frac{\int_{\mathbf{x}} \exp\{\boldsymbol{\zeta}_0^{\mathrm{T}} \mathcal{D}(\mathbf{x}, H)\} \psi(\mathbf{x}) dF_0(\mathbf{x})}{\int_{\mathbf{x}} \exp\{\boldsymbol{\zeta}_0^{\mathrm{T}} \mathcal{D}(\mathbf{x}, H)\} dF_0(\mathbf{x})} = 0. \end{aligned} \tag{S.3}$$

Since $\mathcal{D}(\mathbf{x}, H) = \mathbf{0}$ for any $\mathbf{x}$ under $H = (h_0, h_0')$, we have

$$\boldsymbol{\mu}_{\boldsymbol{\gamma}}^{\mathrm{T}} \nabla_{\boldsymbol{\gamma}} \log P_{\boldsymbol{\gamma}_0}(h_0, h_0') + \psi(\mathbf{X}) - \int_{\mathbf{x}} \psi(\mathbf{x}) dF_0(\mathbf{x}) = 0.$$

This implies that $\psi(\mathbf{x})$ is constant over $\mathbf{x}$, so $\psi = 0$. Thus, (S.3) reduces to

$$\boldsymbol{\mu}_{\boldsymbol{\gamma}}^{\mathrm{T}} \nabla_{\boldsymbol{\gamma}} \log P_{\boldsymbol{\gamma}_0}(H) + \boldsymbol{\mu}_{\boldsymbol{\zeta}}^{\mathrm{T}} \mathcal{D}(\mathbf{X}, H) - \frac{\boldsymbol{\mu}_{\boldsymbol{\zeta}}^{\mathrm{T}} \int_{\mathbf{x}} \exp\{\boldsymbol{\zeta}_0^{\mathrm{T}} \mathcal{D}(\mathbf{x}, H)\} \mathcal{D}(\mathbf{x}, H) dF_0(\mathbf{x})}{\int_{\mathbf{x}} \exp\{\boldsymbol{\zeta}_0^{\mathrm{T}} \mathcal{D}(\mathbf{x}, H)\} dF_0(\mathbf{x})} = 0.$$

By Condition S.3, $\boldsymbol{\mu}_{\boldsymbol{\zeta}} = \mathbf{0}$. It then follows from Lemma 1 of Lin and Zeng (2006) that $\boldsymbol{\mu}_{\boldsymbol{\gamma}} = \mathbf{0}$. Hence, (S.2) reduces to $\boldsymbol{\mu}_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\xi}}^{\mathrm{T}} \nabla_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\xi}} \log P_{\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0,\boldsymbol{\xi}_0}(\mathbf{Y}|\mathbf{X}, H) = 0$. By Condition S.2, $\boldsymbol{\mu}_{\boldsymbol{\alpha}} = \mathbf{0}$, $\boldsymbol{\mu}_{\boldsymbol{\beta}} = \mathbf{0}$, and $\boldsymbol{\mu}_{\boldsymbol{\xi}} = \mathbf{0}$.

THEOREM S.1 Under Conditions S.1-S.3, $|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0| + \sup_{\mathbf{x}} |\widehat{F}(\mathbf{x}) - F_0(\mathbf{x})| \to 0$ almost surely. In addition, $n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ converges in distribution to a zero-mean normal random vector whose covariance matrix attains the semiparametric efficiency bound.

Proof: We first prove the consistency of $\widehat{\boldsymbol{\theta}}$ and $\widehat{F}$. Because $\widehat{\boldsymbol{\theta}}$ is bounded and $\widehat{F}$ is a distribution function, it follows from Helly's selection theorem that, for any subsequence of $\widehat{\boldsymbol{\theta}}$ and $\widehat{F}$, there exists a further subsequence, still denoted as $\widehat{\boldsymbol{\theta}}$ and $\widehat{F}$, such that $\widehat{\boldsymbol{\theta}} \to \boldsymbol{\theta}^*$ and $\widehat{F} \to F^*$ in distribution. It suffices to show $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$ and $F^* = F_0$. Since $\widehat{F}$ maximizes the likelihood function and its jump sizes are positive, there exists a Lagrange multiplier $\widehat{\lambda}$ such that

$$\frac{1}{\widehat{F}\{\mathbf{X}_k\}} - \sum_{i=1}^{n} \frac{\sum_{H \in \mathcal{S}(G_i)} P_{\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\xi}}}(\mathbf{Y}_i|\mathbf{X}_i, H) P_{\widehat{\boldsymbol{\gamma}}}(H) \frac{\exp\{\widehat{\boldsymbol{\zeta}}^{\mathrm{T}}\mathcal{D}(\mathbf{X}_i,H)\}\exp\{\widehat{\boldsymbol{\zeta}}^{\mathrm{T}}\mathcal{D}(\mathbf{X}_k,H)\}}{[\int_{\mathbf{x}} \exp\{\widehat{\boldsymbol{\zeta}}^{\mathrm{T}}\mathcal{D}(\mathbf{x},H)\}d\widehat{F}(\mathbf{x})]^2}}{\sum_{H \in \mathcal{S}(G_i)} P_{\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\xi}}}(\mathbf{Y}_i|\mathbf{X}_i, H) P_{\widehat{\boldsymbol{\gamma}}}(H) \frac{\exp\{\widehat{\boldsymbol{\zeta}}^{\mathrm{T}}\mathcal{D}(\mathbf{X}_i,H)\}}{\int_{\mathbf{x}} \exp\{\widehat{\boldsymbol{\zeta}}^{\mathrm{T}}\mathcal{D}(\mathbf{x},H)\}d\widehat{F}(\mathbf{x})}} - \widehat{\lambda} = 0,$$

where $\widehat{F}\{\mathbf{X}_k\}$ is the jump size of $\widehat{F}$ at $\mathbf{X}_k$. Due to the constraint that $\sum_k \widehat{F}\{\mathbf{X}_k\} = 1$, the above equation implies that $\widehat{\lambda} = 0$. Define $\widetilde{F}$ as a distribution function with jumps at the $\mathbf{X}_k$'s such that the jump size is proportional to

$$\left[\sum_{i=1}^{n} \frac{\sum_{H \in \mathcal{S}(G_i)} P_{\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0,\boldsymbol{\xi}_0}(\mathbf{Y}_i|\mathbf{X}_i, H) P_{\boldsymbol{\gamma}_0}(H) \frac{\exp\{\boldsymbol{\zeta}_0^{\mathrm{T}}\mathcal{D}(\mathbf{X}_i,H)\}\exp\{\boldsymbol{\zeta}_0^{\mathrm{T}}\mathcal{D}(\mathbf{X}_k,H)\}}{[\int_{\mathbf{x}} \exp\{\boldsymbol{\zeta}_0^{\mathrm{T}}\mathcal{D}(\mathbf{x},H)\}dF_0(\mathbf{x})]^2}}{P_{\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0,\boldsymbol{\xi}_0}(\mathbf{Y}_i|\mathbf{X}_i, H) P_{\boldsymbol{\gamma}_0}(H) \frac{\exp\{\boldsymbol{\zeta}_0^{\mathrm{T}}\mathcal{D}(\mathbf{X}_i,H)\}}{\int_{\mathbf{x}} \exp\{\boldsymbol{\zeta}_0^{\mathrm{T}}\mathcal{D}(\mathbf{x},H)\}dF_0(\mathbf{x})}} \right]^{-1}.$$

By the Glivenko-Cantelli theorem, $\widetilde{F}$ uniformly converges to F_0. In addition, $\widehat{F}$ is absolutely continuous with respect to $\widetilde{F}$, and $d\widehat{F}/d\widetilde{F}$ converges uniformly to some positive function g. Finally, since $n^{-1}\log\{L_n(\widehat{\boldsymbol{\theta}}, \widehat{F})/L_n(\boldsymbol{\theta}_0, \widetilde{F})\} \geq 0$, we can take the limit as $n \to \infty$. Thus, the Kullback-Leibler information for $(\boldsymbol{\theta}^*, F^*)$ is non-positive, so the density under $(\boldsymbol{\theta}^*, F^*)$ is the same as the true density. It then follows from Lemma S.1 that $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$ and $F^* = F_0$. This establishes the consistency of $(\widehat{\boldsymbol{\theta}}, \widehat{F})$. The weak convergence of $\widehat{F}$ to F_0 can be strengthened to the uniform convergence since F_0 is a continuous distribution function.

To derive the asymptotic distribution, we consider the score equation along the submodel $(\widehat{\boldsymbol{\theta}} + \epsilon\mathbf{v}, d\widehat{F} + \epsilon(\psi - \int \psi d\widehat{F}))$, where $\mathbf{v}$ is a vector with norm bounded by 1, and ψ is any function with $\int \psi dF_0 = 0$ and with total variation bounded by 1. The score equation takes the form

$$\sqrt{n}\, \boldsymbol{\Omega}_1(\mathbf{v}, \psi)^{\mathrm{T}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \sqrt{n} \int \Omega_2(\mathbf{v}, \psi) d(\widehat{F} - F_0) = \mathbf{G}_n \left\{ l_{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{v} + l_F[\psi] \right\} + o_p(1),$$

where $\mathbf{G}_n$ denotes the empirical measure, $l_{\boldsymbol{\theta}}$ is the score function for $\boldsymbol{\theta}_0$, l_F is the score operator for F_0, $(\boldsymbol{\Omega}_1, \Omega_2)$ is a linear operator of the first-order Fredholm-type which maps $(\mathbf{v}, \psi)$ to the same space as $(\mathbf{v}, \psi)$, and $o_p(1)$ means a random variable converging in probability to zero uniformly in $\mathbf{v}$ and ψ. By some algebra, $(\boldsymbol{\Omega}_1, \Omega_2)[\mathbf{v}, \psi] = 0$ implies that the Fisher information along the submodel is zero, so $\mathbf{v} = \mathbf{0}$ and $\psi = 0$ by Lemma S.2. Thus, $(\boldsymbol{\Omega}_1, \Omega_2)$ is invertible. We then verify all the conditions in Theorem 3.3.1 of van der Vaart and Wellner (1996). Hence, $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0, \widehat{F} - F_0)$ weakly converges to a mean-zero Gaussian process.

In light of the above derivation, the influence function for $\widehat{\boldsymbol{\theta}}$ is a linear combination of some $l_{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{v} + l_F[\psi]$. Thus, the influence function lies on the tangent space spanned by the score functions and thus must be the efficient influence function. This means that $\widehat{\boldsymbol{\theta}}$ is asymptotically efficient in that its limiting covariance matrix attains the semiparametric efficiency bound.

1.2 *Case-control studies with rare disease*

We impose the following identifiability condition.

CONDITION S.4 If $\alpha + \boldsymbol{\beta}^{\mathrm{T}}\mathcal{Z}(\mathbf{X}, H) = \widetilde{\alpha} + \widetilde{\boldsymbol{\beta}}^{\mathrm{T}}\mathcal{Z}(\mathbf{X}, H)$ for any $H = (h, h)$ and $H = (h, h^{\dagger})$, then $\alpha = \widetilde{\alpha}$ and $\boldsymbol{\beta} = \widetilde{\boldsymbol{\beta}}$.

LEMMA S.3 If two sets of parameters $(\boldsymbol{\theta}, F)$ and $(\widetilde{\boldsymbol{\theta}}, \widetilde{F})$ yield the same joint distribution, then $\boldsymbol{\theta} = \widetilde{\boldsymbol{\theta}}$ and $F = \widetilde{F}$.

Proof: Suppose that

$$\left\{\frac{\sum_{H\in\mathcal{S}(G)}\exp\{\boldsymbol{\beta}^{\mathrm{T}}\mathcal{Z}(\mathbf{X},H)\}P_{\boldsymbol{\zeta},F}(\mathbf{X}|H)P_{\boldsymbol{\gamma}}(H)}{\int_{\mathbf{x}}\sum_H\exp\{\boldsymbol{\beta}^{\mathrm{T}}\mathcal{Z}(\mathbf{x},H)\}P_{\boldsymbol{\zeta},F}(\mathbf{x}|H)P_{\boldsymbol{\gamma}}(H)d\mathbf{x}}\right\}^Y\left\{\sum_{H\in\mathcal{S}(G)}P_{\boldsymbol{\zeta},F}(\mathbf{X}|H)P_{\boldsymbol{\gamma}}(H)\right\}^{1-Y}$$

$$=\left\{\frac{\sum_{H\in\mathcal{S}(G)}\exp\{\widetilde{\boldsymbol{\beta}}^{\mathrm{T}}\mathcal{Z}(\mathbf{X},H)\}P_{\widetilde{\boldsymbol{\zeta}},\widetilde{F}}(\mathbf{X}|H)P_{\widetilde{\boldsymbol{\gamma}}}(H)}{\int_{\mathbf{x}}\sum_H\exp\{\widetilde{\boldsymbol{\beta}}^{\mathrm{T}}\mathcal{Z}(\mathbf{x},H)\}P_{\widetilde{\boldsymbol{\zeta}},\widetilde{F}}(\mathbf{x}|H)P_{\widetilde{\boldsymbol{\gamma}}}(H)d\mathbf{x}}\right\}^Y\left\{\sum_{H\in\mathcal{S}(G)}P_{\widetilde{\boldsymbol{\zeta}},\widetilde{F}}(\mathbf{X}|H)P_{\widetilde{\boldsymbol{\gamma}}}(H)\right\}^{1-Y}. \quad \text{(S.4)}$$

Setting $Y = 0$ and $G = 2h$ or $G = h + h^{\dagger}$ in (S.4), we obtain

$$P_{\boldsymbol{\zeta},F}(\mathbf{X}|H)P_{\boldsymbol{\gamma}}(H) = P_{\widetilde{\boldsymbol{\zeta}},\widetilde{F}}(\mathbf{X}|H)P_{\widetilde{\boldsymbol{\gamma}}}(H).$$

Integrating over $\mathbf{X}$ on both sides yields $P_{\boldsymbol{\gamma}}(H) = P_{\widetilde{\boldsymbol{\gamma}}}(H)$, so $\boldsymbol{\gamma} = \widetilde{\boldsymbol{\gamma}}$. Thus, $P_{\boldsymbol{\zeta},F}(\mathbf{X}|H) = P_{\widetilde{\boldsymbol{\zeta}},\widetilde{F}}(\mathbf{X}|H)$. By the arguments in the proof of Lemma S.1, $f = \widetilde{f}$ and $\boldsymbol{\zeta} = \widetilde{\boldsymbol{\zeta}}$. Letting $Y = 1$

and $G = 2h$ or $G = h + h^\dagger$ in (S.4), we see that $\exp\{(\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}})^{\mathrm{T}} \mathcal{Z}(\mathbf{X}, H)\}$ must be a constant. It then follows from Condition S.4 that $\boldsymbol{\beta} = \widetilde{\boldsymbol{\beta}}$.

LEMMA S.4 If there exist a vector $\boldsymbol{\mu_\theta} \equiv (\boldsymbol{\mu}_{\boldsymbol{\beta}}^{\mathrm{T}}, \boldsymbol{\mu}_{\boldsymbol{\gamma}}^{\mathrm{T}}, \boldsymbol{\mu}_{\boldsymbol{\zeta}}^{\mathrm{T}})^{\mathrm{T}}$ and functions $\psi(\mathbf{x})$ with $E[\psi(\mathbf{X})] = 0$ such that

$$\boldsymbol{\mu}_{\boldsymbol{\theta}}^{\mathrm{T}} l_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, F_0) + l_F(\boldsymbol{\theta}_0, F_0)[\int \psi \, dF_0] = 0,$$

where $l_{\boldsymbol{\theta}}$ is the score function for $\boldsymbol{\theta}$, and $l_F[\int \psi \, dF_0]$ is the score function for F along the submodel $F_0 + \epsilon \int \psi \, dF_0$, then $\boldsymbol{\mu_\theta} = \mathbf{0}$ and $\psi = 0$.

Proof: We wish to show that if there exist a vector $\boldsymbol{\mu_\theta} \equiv (\boldsymbol{\mu}_{\boldsymbol{\beta}}^{\mathrm{T}}, \boldsymbol{\mu}_{\boldsymbol{\gamma}}^{\mathrm{T}}, \boldsymbol{\mu}_{\boldsymbol{\zeta}}^{\mathrm{T}})^{\mathrm{T}}$ and functions $\psi(\mathbf{x})$ with $E[\psi(\mathbf{X})] = 0$ such that

$$\boldsymbol{\mu}_{\boldsymbol{\theta}}^{\mathrm{T}} l_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, F_0) + l_F(\boldsymbol{\theta}_0, F_0)[\int \psi \, dF_0] = 0, \tag{S.5}$$

where $l_{\boldsymbol{\theta}}$ is the score function for $\boldsymbol{\theta}$, and $l_F[\int \psi \, dF_0]$ is the score function for F along the submodel $F_0 + \epsilon \int \psi \, dF_0$, then $\boldsymbol{\mu_\theta} = \mathbf{0}$ and $\psi = 0$. To this end, we choose $Y = 0$ and $G = 2h$ or $G = h + h^\dagger$. Then (S.5) becomes

$$\boldsymbol{\mu}_{\boldsymbol{\gamma}}^{\mathrm{T}} \nabla_{\boldsymbol{\gamma}} \log P_{\boldsymbol{\gamma}_0}(H) + \boldsymbol{\mu}_{\boldsymbol{\zeta}}^{\mathrm{T}} \mathcal{D}(\mathbf{X}, H) - \frac{\boldsymbol{\mu}_{\boldsymbol{\zeta}}^{\mathrm{T}} \int_{\mathbf{x}} \exp\{\boldsymbol{\zeta}_0^{\mathrm{T}} \mathcal{D}(\mathbf{x}, H)\} \mathcal{D}(\mathbf{x}, H) dF_0(\mathbf{x})}{\int_{\mathbf{x}} \exp\{\boldsymbol{\zeta}_0^{\mathrm{T}} \mathcal{D}(\mathbf{x}, H)\} dF_0(\mathbf{x})}$$
$$+\psi(\mathbf{X}) - \frac{\int_{\mathbf{x}} \exp\{\boldsymbol{\zeta}_0^{\mathrm{T}} \mathcal{D}(\mathbf{x}, H)\} \psi(\mathbf{x}) dF_0(\mathbf{x})}{\int_{\mathbf{x}} \exp\{\boldsymbol{\zeta}_0^{\mathrm{T}} \mathcal{D}(\mathbf{x}, H)\} dF_0(\mathbf{x})} = 0. \tag{S.6}$$

With $H = (h_0, h_0')$, (S.6) reduces to $\boldsymbol{\mu}_{\boldsymbol{\gamma}}^{\mathrm{T}} \nabla_{\boldsymbol{\gamma}} \log P_{\boldsymbol{\gamma}_0}(h_0, h_0') + \psi(\mathbf{X}) - \int_{\mathbf{x}} \psi(\mathbf{x}) dF_0(\mathbf{x}) = 0$. This implies that $\psi(\mathbf{x})$ is constant, so it must be zero. Thus, (S.6) reduces to

$$\boldsymbol{\mu}_{\boldsymbol{\gamma}}^{\mathrm{T}} \nabla_{\boldsymbol{\gamma}} \log P_{\boldsymbol{\gamma}_0}(H) + \boldsymbol{\mu}_{\boldsymbol{\zeta}}^{\mathrm{T}} \mathcal{D}(\mathbf{X}, H) - \frac{\boldsymbol{\mu}_{\boldsymbol{\zeta}}^{\mathrm{T}} \int_{\mathbf{x}} \exp\{\boldsymbol{\zeta}_0^{\mathrm{T}} \mathcal{D}(\mathbf{x}, H)\} \mathcal{D}(\mathbf{x}, H) dF_0(\mathbf{x})}{\int_{\mathbf{x}} \exp\{\boldsymbol{\zeta}_0^{\mathrm{T}} \mathcal{D}(\mathbf{x}, H)\} dF_0(\mathbf{x})} = 0.$$

By Condition S.3, $\boldsymbol{\mu}_{\boldsymbol{\zeta}} = \mathbf{0}$, so (S.6) further reduces to $\boldsymbol{\mu}_{\boldsymbol{\gamma}}^{\mathrm{T}} \nabla_{\boldsymbol{\gamma}} \log P_{\boldsymbol{\gamma}_0}(H) = 0$. By Lemma 1 of Lin and Zeng (2006), $\boldsymbol{\mu}_{\boldsymbol{\gamma}} = \mathbf{0}$. Setting $Y = 1$ and $G = 2h$ or $G = h + h^\dagger$, we see that $\boldsymbol{\mu}_{\boldsymbol{\beta}}^{\mathrm{T}} \mathcal{Z}(\mathbf{X}, H)$ must be a constant. By Condition S.4, $\boldsymbol{\mu}_{\boldsymbol{\beta}} = \mathbf{0}$.

We provide a mathematical definition of rare disease in Condition S.5 and state the asymptotic results in Theorem S.2.

CONDITION S.5 $\Pr(Y_i = 1|\mathbf{X}_i, H_i) = a_n \exp\{\boldsymbol{\beta}_0^{\mathrm{T}} \mathcal{Z}(\mathbf{X}_i, H_i)\}/[1 + a_n \exp\{\boldsymbol{\beta}_0^{\mathrm{T}} \mathcal{Z}(\mathbf{X}_i, H_i)\}]$, $i = 1, \ldots, n$, where $a_n = o(n^{-1/2})$.

THEOREM S.2 Assume that Conditions S.3-S.5 hold and $n_1/n \to q \in (0, 1)$. Then $|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0| + \sup_{\mathbf{x}} |\widehat{F}(\mathbf{x}) - F_0(\mathbf{x})| \to 0$ almost surely, and $n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ converges in distribution to a zero-mean normal random vector whose covariance matrix attains the semiparametric efficiency bound.

Proof: Let $\widetilde{P}_n$ be the probability measure generated by the likelihood function given in (2.8) and let P_{n0} be the true likelihood function. Since $a_n = o(n^{-1/2})$, we have $\log \widetilde{P}_n/P_{n0} \to_{\widetilde{P}_n \text{ or } P_{n0}} 1$. By LeCam's lemma, $\widetilde{P}_n$ and P_{n0} are equivalent. Thus, the asymptotic properties under the true likelihood is equivalent to those under the the approximate likelihood given in (2.8). In other words, we can assume that data are generated from (2.8). Hence, the conclusion of the theorem follows from the arguments in the proof of Theorem S.1.

1.3 *Case-control studies with known disease rate*

LEMMA S.5 If two sets of parameters $(\boldsymbol{\theta}, F)$ and $(\widetilde{\boldsymbol{\theta}}, \widetilde{F})$ yield the same joint distribution, then $\boldsymbol{\theta} = \widetilde{\boldsymbol{\theta}}$ and $F = \widetilde{F}$.

Proof: Suppose that

$$\sum_{H \in \mathcal{S}(G)} P_{\alpha,\boldsymbol{\beta}}(Y|\mathbf{X}, H) P_{\boldsymbol{\zeta},F}(\mathbf{X}|H) P_{\boldsymbol{\gamma}}(H) = \sum_{H \in \mathcal{S}(G)} P_{\widetilde{\alpha},\widetilde{\boldsymbol{\beta}}}(Y|\mathbf{X}, H) P_{\widetilde{\boldsymbol{\zeta}},\widetilde{F}}(\mathbf{X}|H) P_{\widetilde{\boldsymbol{\gamma}}}(H).$$

Letting $G = 2h$ or $G = h + h^\dagger$, we have

$$P_{\alpha,\boldsymbol{\beta}}(Y|\mathbf{X}, H) P_{\boldsymbol{\zeta},F}(\mathbf{X}|H) P_{\boldsymbol{\gamma}}(H) = P_{\widetilde{\alpha},\widetilde{\boldsymbol{\beta}}}(Y|\mathbf{X}, H) P_{\widetilde{\boldsymbol{\zeta}},\widetilde{F}}(\mathbf{X}|H) P_{\widetilde{\boldsymbol{\gamma}}}(H). \tag{S.7}$$

Set $Y = 0$ or 1 in (S.7). The summation of the two resulting equations yields

$$P_{\boldsymbol{\zeta},F}(\mathbf{X}|H) P_{\boldsymbol{\gamma}}(H) = P_{\widetilde{\boldsymbol{\zeta}},\widetilde{F}}(\mathbf{X}|H) P_{\widetilde{\boldsymbol{\gamma}}}(H).$$

By the arguments in the proof of Lemma S.3, $\boldsymbol{\gamma} = \widetilde{\boldsymbol{\gamma}}$, $f = \widetilde{f}$, and $\boldsymbol{\zeta} = \widetilde{\boldsymbol{\zeta}}$. Then (S.7) reduces to $\exp\big\{(\alpha - \widetilde{\alpha}) + (\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}})^{\mathrm{T}} \mathcal{Z}(\mathbf{X}, H)\big\} = 1$. By Condition S.4, $\alpha = \widetilde{\alpha}$ and $\boldsymbol{\beta} = \widetilde{\boldsymbol{\beta}}$.

LEMMA S.6 If there exist a vector $\boldsymbol{\mu}_{\boldsymbol{\theta}} \equiv (\mu_\alpha, \boldsymbol{\mu}_{\boldsymbol{\beta}}^{\mathrm{T}}, \boldsymbol{\mu}_{\boldsymbol{\gamma}}^{\mathrm{T}}, \boldsymbol{\mu}_{\boldsymbol{\zeta}}^{\mathrm{T}})^{\mathrm{T}}$ and a function ψ with $E[\psi(\mathbf{X})] = 0$ such that

$$\boldsymbol{\mu}_{\boldsymbol{\theta}}^{\mathrm{T}} l_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, F_0) + l_F(\boldsymbol{\theta}_0, F_0)[\int \psi \; dF_0] = 0,$$

where $l_{\boldsymbol{\theta}}$ is the score function for $\boldsymbol{\theta}$, and $l_F[\int \psi \; dF_0]$ is the score function for F along the submodel $F_0 + \epsilon \int \psi dF_0$ that satisfies the constraint $\Pr(Y=1) = p_1$, then $\boldsymbol{\mu}_{\boldsymbol{\theta}} = \mathbf{0}$ and $\psi = 0$.

Proof: We wish to show that if there exist a vector $\boldsymbol{\mu}_{\boldsymbol{\theta}} \equiv (\mu_\alpha, \boldsymbol{\mu}_{\boldsymbol{\beta}}^{\mathrm{T}}, \boldsymbol{\mu}_{\boldsymbol{\gamma}}^{\mathrm{T}}, \boldsymbol{\mu}_{\boldsymbol{\zeta}}^{\mathrm{T}})^{\mathrm{T}}$ and functions ψ with $E[\psi(\mathbf{X})] = 0$ such that

$$\boldsymbol{\mu}_{\boldsymbol{\theta}}^{\mathrm{T}} l_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, F_0) + l_F(\boldsymbol{\theta}_0, F_0)[\int \psi \; dF_0] = 0, \tag{S.8}$$

where $l_{\boldsymbol{\theta}}$ is the score function for $\boldsymbol{\theta}$, and $l_F[\int \psi \; dF_0]$ is the score function for F along the submodel $F_0 + \epsilon \int \psi dF_0$ that satisfies the constraint $\Pr(Y=1) = p$, then $\boldsymbol{\mu}_{\boldsymbol{\theta}} = \mathbf{0}$ and $\psi = 0$. With $G = 2h$ or $G = h + h^{\dagger}$, (S.8) becomes

$$(\mu_\alpha + \boldsymbol{\mu}_{\boldsymbol{\beta}}^{\mathrm{T}} \mathcal{Z}(\mathbf{X}, H)) \left[Y - \frac{\exp\{\alpha_0 + \boldsymbol{\beta}_0^{\mathrm{T}} \mathcal{Z}(\mathbf{X}, H)\}}{1 + \exp\{\alpha_0 + \boldsymbol{\beta}_0^{\mathrm{T}} \mathcal{Z}(\mathbf{X}, H)\}} \right] + \boldsymbol{\mu}_{\boldsymbol{\gamma}}^{\mathrm{T}} \nabla_{\boldsymbol{\gamma}} \log P_{\boldsymbol{\gamma}_0}(H) + \boldsymbol{\mu}_{\boldsymbol{\zeta}}^{\mathrm{T}} \mathcal{D}(\mathbf{X}, H)$$
$$- \frac{\boldsymbol{\mu}_{\boldsymbol{\zeta}}^{\mathrm{T}} \int_{\mathbf{x}} \exp\{\boldsymbol{\zeta}_0^{\mathrm{T}} \mathcal{D}(\mathbf{x}, H)\} \mathcal{D}(\mathbf{x}, H) dF_0(\mathbf{x})}{\int_{\mathbf{x}} \exp\{\boldsymbol{\zeta}_0^{\mathrm{T}} \mathcal{D}(\mathbf{x}, H)\} dF_0(\mathbf{x})} + \psi(\mathbf{X}) - \frac{\int_{\mathbf{x}} \exp\{\boldsymbol{\zeta}_0^{\mathrm{T}} \mathcal{D}(\mathbf{x}, H)\} \psi(\mathbf{x}) dF_0(\mathbf{x})}{\int_{\mathbf{x}} \exp\{\boldsymbol{\zeta}_0^{\mathrm{T}} \mathcal{D}(\mathbf{x}, H)\} dF_0(\mathbf{x})} = 0.$$

The difference of the two equations under $Y = 1$ and $Y = 0$ yields $\mu_\alpha + \boldsymbol{\mu}_{\boldsymbol{\beta}}^{\mathrm{T}} \mathcal{Z}(\mathbf{X}, H) = 0$. By Condition S.4, $\mu_\alpha = 0$ and $\boldsymbol{\mu}_{\boldsymbol{\beta}} = \mathbf{0}$. It then follows from the arguments in the proof of Lemma S.4 that $\boldsymbol{\mu}_{\boldsymbol{\zeta}} = \mathbf{0}$, $\boldsymbol{\mu}_{\boldsymbol{\gamma}} = \mathbf{0}$, and $\psi = 0$.

THEOREM S.3 Under Conditions S.3-S.4, the results of Theorem S.2 hold.

Proof: First, we prove the consistency. Since $\widehat{\boldsymbol{\theta}}$ is bounded and $\widehat{F}$ is a distribution function, for any subsequence of $(\widehat{\boldsymbol{\theta}}, \widehat{F})$, there exists a further subsequence, still denoted as $(\widehat{\boldsymbol{\theta}}, \widehat{F})$, such that $\widehat{\boldsymbol{\theta}} \to \boldsymbol{\theta}^*$, and $\widehat{F}$ weakly converge to F^*. The consistency will hold if we can show that $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$ and $F^* = F_0$. We abbreviate $\eta(\mathbf{x}, \mathbf{x}_0, (h, h'), (h_0, h_0'))$ and $P_{\alpha,\boldsymbol{\beta}}(Y|\mathbf{x}, H) P_{\boldsymbol{\gamma}}(H)$ as $\eta(\mathbf{x}, H)$ and $q(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{x}, H, Y)$, respectively. After differentiating the log-likelihood function with respect to the jump sizes of F, we see that there exist some Lagrange multipliers $\widehat{\lambda}_1$ and $\widehat{\lambda}_2$ such that, for $k = 1, \ldots, n$,

$$\frac{1}{\widehat{F}\{\mathbf{X}_k\}} - \sum_{i=1}^n \frac{\sum_{H \in \mathcal{S}(G_i)} q(\widehat{\alpha}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}, \mathbf{X}_i, H, Y_i)\eta(\mathbf{X}_i, H)\eta(\mathbf{X}_k, H)/\{\int_{\mathbf{x}} \eta(\mathbf{x}, H) d\widehat{F}(\mathbf{x})\}^2}{\sum_{H \in \mathcal{S}(G_i)} q(\widehat{\alpha}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}, \mathbf{X}_i, H, Y_i)\eta(\mathbf{X}_i, H)/\int_{\mathbf{x}} \eta(\mathbf{x}, H) d\widehat{F}(\mathbf{x})}$$
$$-\widehat{\lambda}_2 \sum_H \left[\frac{q(\widehat{\alpha}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}, \mathbf{X}_k, H, 1)\eta(\mathbf{X}_k, H)}{\int_{\mathbf{x}} \eta(\mathbf{x}, H) d\widehat{F}(\mathbf{x})} - \frac{\eta(\mathbf{X}_k, H) \int_{\mathbf{x}} q(\widehat{\alpha}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}, \mathbf{x}, H, 1)\eta(\mathbf{x}, H) d\widehat{F}(\mathbf{x})}{\{\int_{\mathbf{x}} \eta(\mathbf{x}, H) d\widehat{F}(\mathbf{x}))\}^2} \right] - \widehat{\lambda}_1 = 0.$$

In addition, $\widehat{\lambda}_1$ and $\widehat{\lambda}_2$ satisfy the constraint equations

$$\sum_{k=1}^n \widehat{F}\{\mathbf{X}_k\} = 1,$$

$$\sum_{k=1}^n \sum_H q(\widehat{\alpha}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}, \mathbf{X}_k, H, 1) \frac{\eta(\mathbf{X}_k, H)}{\int_{\mathbf{x}} \eta(\mathbf{x}, H) d\widehat{F}(\mathbf{x})} \widehat{F}\{\mathbf{X}_k\} = p_1.$$

It follows that $\widehat{\lambda}_1 = 0$. Thus,

$$\left\{ \sum_{i=1}^n \frac{\sum_{H \in \mathcal{S}(G_i)} q(\widehat{\alpha}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}, \mathbf{X}_i, H, Y_i)\eta(\mathbf{X}_i, H)\eta(\mathbf{X}_k, H)/\{\int_{\mathbf{x}} \eta(\mathbf{x}, H) d\widehat{F}(\mathbf{x})\}^2}{\sum_{H \in \mathcal{S}(G_i)} q(\widehat{\alpha}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}, \mathbf{X}_i, H, Y_i)\eta(\mathbf{X}_i, H)/\int_{\mathbf{x}} \eta(\mathbf{x}, H) d\widehat{F}(\mathbf{x})} \right.$$
$$\left. +\widehat{\lambda}_2 \sum_H \left[\frac{q(\widehat{\alpha}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}, \mathbf{X}_k, H, 1)\eta(\mathbf{X}_k, H)}{\int_{\mathbf{x}} \eta(\mathbf{x}, H) d\widehat{F}(\mathbf{x})} - \frac{\eta(\mathbf{X}_k, H) \int_{\mathbf{x}} q(\widehat{\alpha}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}, \mathbf{x}, H, 1)\eta(\mathbf{x}, H) d\widehat{F}(\mathbf{x})}{\{\int_{\mathbf{x}} \eta(\mathbf{x}, H) d\widehat{F}(\mathbf{x}))\}^2} \right] \right\}^{-1} = 1,$$

and each denominator on the left-hand side should be positive. This equation for $\widehat{\lambda}_2$ has a unique solution satisfying the above constraints. In addition, we can show that $\widehat{\lambda}_2/n$ is bounded with probability one. Thus, we can choose a further subsequence such that $\widehat{\lambda}_2/n \to \lambda_2^*$.

We construct a discrete distribution function $\widetilde{F}$ such that $\widetilde{F} \to F_0$ uniformly. The sequence can be constructed along the lines of Lin and Zeng (2006, §A.4.6). Although $\widetilde{F}$ is a distribution function, it may not satisfy the constraint that

$$\int_{\mathbf{x}} \sum_H P_{\alpha_0, \boldsymbol{\beta}_0}(Y = 1|\mathbf{x}, H) P_{\boldsymbol{\gamma}}(H) P_{\boldsymbol{\zeta}_0, F}(\mathbf{x}|H) f(\mathbf{x}) d\mathbf{x} = p_1.$$

Thus, we modify the jump size of $\widetilde{F}$ at $\mathbf{X}_k$ as $[\widetilde{F}\{\mathbf{X}_k\} + \xi/n]/(1 + \xi)$ for some constant ξ such that ξ satisfies the above constraint. It can be shown that the solution exists and $\xi \to 0$. The modified distribution function $\widetilde{F}$ then satisfies all the constraints. By the Glivenko-Cantelli theorem, $\widehat{F}$ is absolutely continuous with respect to $\widetilde{F}$, and $d\widehat{F}/d\widetilde{F}(\mathbf{x}) \to q(\mathbf{x})$ uniformly in $\mathbf{x}$ for some positive function $q(\cdot)$. Since $n^{-1}\log\{L_n(\widehat{\boldsymbol{\theta}}, \widehat{F})/L_n(\boldsymbol{\theta}_0, \widetilde{F})\} \geq 0$, we take limits. We conclude that the Kullback-Leibler information for $(\boldsymbol{\theta}^*, F^*)$ is non-positive. Hence, Lemma S.5 entails that $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$ and $F^* = F_0$.

We now derive the asymptotic distribution. We obtain score functions by differentiating $\log L_n(\boldsymbol{\theta}, F)$ with respect to $\widehat{\boldsymbol{\theta}}$ along the direction $\mathbf{v}$ and with respect to $\widehat{F}$ along submodels with tangent direction ψ satisfying all the constraints and with the total variation bounded by 1. The linearization of the score functions around the true parameter value, together with the Donsker theorem, yields

$$n^{1/2}\left[\boldsymbol{\Omega}_1(\mathbf{v}, \psi)^{\mathrm{T}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \int \Omega_2(\mathbf{v}, \psi) d(\widehat{F} - F_0)\right] = n^{-1/2}\sum_{i=1}^{n}\left(\mathbf{v}^{\mathrm{T}} l_{\boldsymbol{\theta}} + l_F[\psi]\right) + o_p(1),$$

where $\boldsymbol{\Omega} \equiv (\boldsymbol{\Omega}_1, \Omega_2)$ corresponds to the information operator and has the form of the first-order Fredholm type, and $l_{\boldsymbol{\theta}}$ and l_F are the score operators for $\boldsymbol{\theta}$ and F, respectively. According to Lemma S.6, $\boldsymbol{\Omega}$ is invertible. Thus, the weak convergence follows from Theorem 3.3.1 of van der Vaart and Wellner (1996). In addition, $\widehat{\boldsymbol{\theta}}$ is an asymptotically linear estimator for $\boldsymbol{\theta}_0$ with the influence function in the score space, so it follows from Proposition 3.3.1 of Bickel et al. (1993) that the limiting covariance matrix of $n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ attains the semiparametric efficiency bound.

1.4 *Cohort studies*

We impose the following conditions:

CONDITION S.6 There exists a positive constant δ_0 such that $\Pr(C \geq \tau|\mathbf{X}, G) = \Pr(C = \tau|\mathbf{X}, G) \geq \delta_0$ almost surely, where τ corresponds to the end of the study.

CONDITION S.7 The true value $\Lambda_0(t)$ of $\Lambda(t)$ is a strictly increasing function in $[0, \tau]$ and is continuously differentiable. In addition, $\Lambda_0(0) = 0$ and $\Lambda_0'(0) > 0$.

LEMMA S.7 If two sets of parameters $(\boldsymbol{\theta}, F, \Lambda)$ and $(\widetilde{\boldsymbol{\theta}}, \widetilde{F}, \widetilde{\Lambda})$ yield the same joint distribution, then $\boldsymbol{\theta} = \widetilde{\boldsymbol{\theta}}$, $F = \widetilde{F}$ and $\Lambda = \widetilde{\Lambda}$.

Proof: Suppose that

$$\sum_{H\in\mathcal{S}(G)}\Big[\Lambda'(\widetilde{Y})e^{\boldsymbol{\beta}^{\mathrm{T}}\mathcal{Z}(\mathbf{X},H)}Q'(\Lambda(\widetilde{Y})e^{\boldsymbol{\beta}^{\mathrm{T}}\mathcal{Z}(\mathbf{X},H)})\Big]^{\Delta}\exp\Big\{-Q(\Lambda(\widetilde{Y})e^{\boldsymbol{\beta}^{\mathrm{T}}\mathcal{Z}(\mathbf{X},H)})\Big\}P_{\boldsymbol{\zeta},F}(\mathbf{X}|H)P_{\boldsymbol{\gamma}}(H)$$
$$=\sum_{H\in\mathcal{S}(G)}\Big[\widetilde{\Lambda}'(\widetilde{Y})e^{\widetilde{\boldsymbol{\beta}}^{\mathrm{T}}\mathcal{Z}(\mathbf{X},H)}Q'(\widetilde{\Lambda}(\widetilde{Y})e^{\widetilde{\boldsymbol{\beta}}^{\mathrm{T}}\mathcal{Z}(\mathbf{X},H)})\Big]^{\Delta}\exp\Big\{-Q(\widetilde{\Lambda}(\widetilde{Y})e^{\widetilde{\boldsymbol{\beta}}^{\mathrm{T}}\mathcal{Z}(\mathbf{X},H)})\Big\}P_{\widetilde{\boldsymbol{\zeta}},\widetilde{F}}(\mathbf{X}|H)P_{\widetilde{\boldsymbol{\gamma}}}(H).$$

We choose $\Delta = 1$ and integrate $\widetilde{Y}$ from 0 to y on both sides to obtain the equation

$$\sum_{H\in\mathcal{S}(G)}\Big[1-\exp\{-Q(\Lambda(y)e^{\boldsymbol{\beta}^{\mathrm{T}}\mathcal{Z}(\mathbf{X},H)})\}\Big]P_{\boldsymbol{\zeta},F}(\mathbf{X}|H)P_{\boldsymbol{\gamma}}(H)$$
$$=\sum_{H\in\mathcal{S}(G)}\Big[1-\exp\{-Q(\widetilde{\Lambda}(y)e^{\widetilde{\boldsymbol{\beta}}^{\mathrm{T}}\mathcal{Z}(\mathbf{X},H)})\}\Big]P_{\widetilde{\boldsymbol{\zeta}},\widetilde{F}}(\mathbf{X}|H)P_{\widetilde{\boldsymbol{\gamma}}}(H). \tag{S.9}$$

We obtain a second equation by setting $\Delta = 0$ and $\widetilde{Y} = y$. The summation of the two equations yields

$$\sum_{H\in\mathcal{S}(G)}P_{\boldsymbol{\zeta},F}(\mathbf{X}|H)P_{\boldsymbol{\gamma}}(H)=\sum_{H\in\mathcal{S}(G)}P_{\widetilde{\boldsymbol{\zeta}},\widetilde{F}}(\mathbf{X}|H)P_{\widetilde{\boldsymbol{\gamma}}}(H).$$

By the arguments in the proof of Lemma S.1, $\boldsymbol{\gamma} = \widetilde{\boldsymbol{\gamma}}$, $f = \widetilde{f}$ and $\boldsymbol{\zeta} = \widetilde{\boldsymbol{\zeta}}$. By letting $G = 2h$ or $G = h + \widetilde{h}$ in (S.9), we have $\Lambda(y)e^{\boldsymbol{\beta}^{\mathrm{T}}\mathcal{Z}(\mathbf{X},H)} = \widetilde{\Lambda}(y)e^{\widetilde{\boldsymbol{\beta}}^{\mathrm{T}}\mathcal{Z}(\mathbf{X},H)}$, which entails $\Lambda = \widetilde{\Lambda}$ and $\boldsymbol{\beta} = \widetilde{\boldsymbol{\beta}}$ under Condition S.4.

LEMMA S.8 If there exist a vector $\boldsymbol{\mu}_{\boldsymbol{\theta}} \equiv (\boldsymbol{\mu}_{\boldsymbol{\beta}}^{\mathrm{T}}, \boldsymbol{\mu}_{\boldsymbol{\gamma}}^{\mathrm{T}}, \boldsymbol{\mu}_{\boldsymbol{\zeta}}^{\mathrm{T}})^{\mathrm{T}}$ and functions $\psi(\mathbf{x})$ and $\phi(t)$ with $E[\psi(\mathbf{X})] = E[\phi(Y)] = 0$ such that

$$\boldsymbol{\mu}_{\boldsymbol{\theta}}^{\mathrm{T}}l_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0,F_0,\Lambda_0)+l_F(\boldsymbol{\theta}_0,F_0,\Lambda_0)[\int\psi\,dF_0]+l_{\Lambda}(\boldsymbol{\theta}_0,F_0,\Lambda_0)[\int\phi\,d\Lambda_0]=0,$$

where $l_{\boldsymbol{\theta}}$ is the score function for $\boldsymbol{\theta}$, $l_F[\int\psi\,dF_0]$ is the score function for F along the sub-model $F_0+\epsilon\int\psi\,dF_0$, and $l_\Lambda[\int\phi\,d\Lambda_0]$ is the score function for Λ along the sub-model $\Lambda_0+\epsilon\int\phi\,d\Lambda_0$, then $\boldsymbol{\mu}_{\boldsymbol{\theta}} = \mathbf{0}$, $\psi = 0$ and $\phi = 0$.

Proof: We wish to show that if there exist a vector $\boldsymbol{\mu}_{\boldsymbol{\theta}} \equiv (\boldsymbol{\mu}_{\boldsymbol{\beta}}^{\mathrm{T}}, \boldsymbol{\mu}_{\boldsymbol{\gamma}}^{\mathrm{T}}, \boldsymbol{\mu}_{\boldsymbol{\zeta}}^{\mathrm{T}})^{\mathrm{T}}$ and functions $\psi(\mathbf{x})$ and $\phi(t)$ with $E[\psi(\mathbf{X})] = E[\phi(Y)] = 0$ such that

$$\boldsymbol{\mu}_{\boldsymbol{\theta}}^{\mathrm{T}}l_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0,F_0,\Lambda_0)+l_F(\boldsymbol{\theta}_0,F_0,\Lambda_0)[\int\psi\,dF_0]+l_{\Lambda}(\boldsymbol{\theta}_0,F_0,\Lambda_0)[\int\phi\,d\Lambda_0]=0, \tag{S.10}$$

where $l_{\boldsymbol{\theta}}$ is the score function for $\boldsymbol{\theta}$, $l_F[\int \psi\, dF_0]$ is the score function for F along the sub-model $F_0 + \epsilon \int \psi\, dF_0$, and $l_\Lambda[\int \phi\, d\Lambda_0]$ is the score function for Λ along the sub-model $\Lambda_0 + \epsilon \int \phi\, d\Lambda_0$, then $\boldsymbol{\mu}_{\boldsymbol{\theta}} = \mathbf{0}$, $\psi = 0$ and $\phi = 0$. With $\Delta = 1$, (S.10) becomes

$$\begin{aligned}
&\sum_{H\in\mathcal{S}(G)} \Lambda_0'(\widetilde{Y})e^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathcal{Z}}Q'(\Lambda_0(\widetilde{Y})e^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathcal{Z}})\exp\Big\{-Q(\Lambda_0(\widetilde{Y})e^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathcal{Z}})\Big\}P_{\boldsymbol{\gamma}_0}(H)\frac{\exp\{\boldsymbol{\zeta}_0^{\mathrm{T}}\mathcal{D}(\mathbf{X},H)\}f_0(\mathbf{X})}{\int_{\mathbf{x}}\exp\{\boldsymbol{\zeta}_0^{\mathrm{T}}\mathcal{D}(\mathbf{x},H)\}dF_0(\mathbf{x})} \\
&\quad\times\Bigg\{\boldsymbol{\mu}_{\boldsymbol{\beta}}^{\mathrm{T}}\mathcal{Z} + \frac{\Big[Q''(\Lambda_0(\widetilde{Y})e^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathcal{Z}}) - \big(Q'(\Lambda_0(\widetilde{Y})e^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathcal{Z}})\big)^2\Big]\Lambda_0(\widetilde{Y})e^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathcal{Z}}\boldsymbol{\mu}_{\boldsymbol{\beta}}^{\mathrm{T}}\mathcal{Z}}{Q'(\Lambda_0(\widetilde{Y})e^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathcal{Z}})} \\
&+\phi(\widetilde{Y}) + \frac{\Big[Q''(\Lambda_0(\widetilde{Y})e^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathcal{Z}}) - \big(Q'(\Lambda_0(\widetilde{Y})e^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathcal{Z}})\big)^2\Big]\int_0^{\widetilde{Y}}\phi(t)d\Lambda_0(t)e^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathcal{Z}}}{Q'(\Lambda_0(\widetilde{Y})e^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathcal{Z}})} + \boldsymbol{\mu}_{\boldsymbol{\gamma}}^{\mathrm{T}}\nabla_{\boldsymbol{\gamma}}\log P_{\boldsymbol{\gamma}_0}(H) \\
&\quad+\boldsymbol{\mu}_{\boldsymbol{\zeta}}^{\mathrm{T}}\mathcal{D}(\mathbf{X},H) - \frac{\boldsymbol{\mu}_{\boldsymbol{\zeta}}^{\mathrm{T}}\int_{\mathbf{x}}\exp\{\boldsymbol{\zeta}_0^{\mathrm{T}}\mathcal{D}(\mathbf{x},H)\}\mathcal{D}(\mathbf{x},H)dF_0(\mathbf{x})}{\int_{\mathbf{x}}\exp\{\boldsymbol{\zeta}_0^{\mathrm{T}}\mathcal{D}(\mathbf{x},H)\}dF_0(\mathbf{x})} \\
&\quad+\psi(\mathbf{X}) - \frac{\int_{\mathbf{x}}\exp\{\boldsymbol{\zeta}_0^{\mathrm{T}}\mathcal{D}(\mathbf{x},H)\}\psi(\mathbf{x})dF_0(\mathbf{x})}{\int_{\mathbf{x}}\exp\{\boldsymbol{\zeta}_0^{\mathrm{T}}\mathcal{D}(\mathbf{x},H)\}dF_0(\mathbf{x})}\Bigg\} = 0. \qquad \text{(S.11)}
\end{aligned}$$

In the above equation, we integrate $\widetilde{Y}$ from 0 to τ. We also let $\Delta = 0$ and $\widetilde{Y} = \tau$ in (S.10). The summation of these two equations with $G = 2h$ or $G = h + h^\dagger$ yields

$$\begin{aligned}
\boldsymbol{\mu}_{\boldsymbol{\gamma}}^{\mathrm{T}}\nabla_{\boldsymbol{\gamma}}\log P_{\boldsymbol{\gamma}_0}(H) + \boldsymbol{\mu}_{\boldsymbol{\zeta}}^{\mathrm{T}}\mathcal{D}(\mathbf{X},H) - \frac{\boldsymbol{\mu}_{\boldsymbol{\zeta}}^{\mathrm{T}}\int_{\mathbf{x}}\exp\{\boldsymbol{\zeta}_0^{\mathrm{T}}\mathcal{D}(\mathbf{x},H)\}\mathcal{D}(\mathbf{x},H)dF_0(\mathbf{x})}{\int_{\mathbf{x}}\exp\{\boldsymbol{\zeta}_0^{\mathrm{T}}\mathcal{D}(\mathbf{x},H)\}dF_0(\mathbf{x})}& \\
+ \psi(\mathbf{X}) - \frac{\int_{\mathbf{x}}\exp\{\boldsymbol{\zeta}_0^{\mathrm{T}}\mathcal{D}(\mathbf{x},H)\}\psi(\mathbf{x})dF_0(\mathbf{x})}{\int_{\mathbf{x}}\exp\{\boldsymbol{\zeta}_0^{\mathrm{T}}\mathcal{D}(\mathbf{x},H)\}dF_0(\mathbf{x})} &= 0.
\end{aligned}$$

It follows from the arguments in the proof of Lemma S.2 that $\boldsymbol{\mu}_{\boldsymbol{\gamma}} = \mathbf{0}$, $\boldsymbol{\mu}_{\boldsymbol{\zeta}} = \mathbf{0}$, and $\psi = 0$. By letting $G = 2h$ or $G = h + h^\dagger$ and $Y = 0$ in (S.11), we obtain $\boldsymbol{\mu}_{\boldsymbol{\beta}}^{\mathrm{T}}\mathcal{Z}(\mathbf{X},H) + \phi(0) = 0$. It then follows from Condition S.4 that $\boldsymbol{\mu}_{\boldsymbol{\beta}} = \mathbf{0}$ and $\phi(0) = 0$. Thus, (S.11) reduces to

$$\phi(\widetilde{Y}) + \frac{\Big[Q''(\Lambda_0(\widetilde{Y})e^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathcal{Z}}) - \big(Q'(\Lambda_0(\widetilde{Y})e^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathcal{Z}})\big)^2\Big]\int_0^{\widetilde{Y}}\phi(t)d\Lambda_0(t)e^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathcal{Z}}}{Q'(\Lambda_0(\widetilde{Y})e^{\boldsymbol{\beta}_0^{\mathrm{T}}\mathcal{Z}})} = 0$$

for $H = (h, h)$. Since Q is strictly increasing, we conclude that $\phi(y) = 0$ for any y.

THEOREM S.4 Under the conditions of Theorem S.3 and Conditions S.6-S.7, $|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0| + \sup_{\mathbf{x}}|\widehat{F}(\mathbf{x}) - F_0(\mathbf{x})| + \sup_{t\in[0,\tau]}|\widehat{\Lambda}(t) - \Lambda_0(t)| \to 0$ almost surely. In addition, $n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0, \widehat{\Lambda} - \Lambda_0)$

converges weakly to a zero-mean Gaussian process in $R^d \times l^\infty([0,\tau])$, where d is the dimension of $\boldsymbol{\theta}_0$, and $l^\infty([0,\tau])$ is the space of all bounded functions on $[0,\tau]$ equipped with the supremum norm. Furthermore, the limiting covariance matrix of $\widehat{\boldsymbol{\theta}}$ attains the semiparametric efficiency bound.

Proof: First, we show that $\widehat{\Lambda}$ is uniformly bounded in $[0,\tau]$ as $n \to \infty$. Note that $\widehat{\Lambda}$ maximizes $\widetilde{L}_n(\Lambda) \equiv L_n(\widehat{\boldsymbol{\theta}}, \Lambda, \widehat{F}) / \prod_{i=1}^n \widehat{F}\{\mathbf{X}_i\}$. Clearly,

$$\widetilde{L}_n(\Lambda) \le c_0 \prod_{i=1}^n \sum_{H \in \mathcal{S}(G_i)} \left\{ \Lambda'(\widetilde{Y}_i) e^{\boldsymbol{\beta}^{\mathrm{T}} \mathcal{Z}(\mathbf{X}_i, H)} Q'\big(-\Lambda(\widetilde{Y}_i) e^{\boldsymbol{\beta}^{\mathrm{T}} \mathcal{Z}(\mathbf{X}_i, H_i)}\big) \right\}^{\Delta_i} \exp\left\{ -Q\big(\Lambda(\widetilde{Y}_i) e^{\boldsymbol{\beta}^{\mathrm{T}} \mathcal{Z}(\mathbf{X}_i, H_i)}\big) \right\}$$

for some constant c_0. According to the conditions of this theorem and Appendix B of Zeng and Lin (2007), $\widetilde{L}_n(\Lambda) \le c_1 \prod_{i=1}^n \left[\Lambda'(\widetilde{Y}_i)^{\Delta_i} (1 + \Lambda(\widetilde{Y}_i))^{-(\Delta_i + \delta_0)} \right]$ for some positive constants c_1 and δ_0. By the partitioning arguments in the proof of Theorem 1 of Zeng and Lin (2007), we can show that if $\widehat{\Lambda}(\tau)$ is unbounded, then the difference between $L_n(\widehat{\boldsymbol{\theta}}, \widehat{\Lambda}, \widehat{F})$ and $L_n(\boldsymbol{\theta}_0, \widetilde{\Lambda}, \widehat{F})$, where $\widetilde{\Lambda}$ is a step function converging to Λ_0, diverges to $-\infty$. Thus, $\widehat{\Lambda}(\tau)$ must be bounded with probability one.

Using the above result and the arguments in the proof of Theorem S.3, we choose a uniformly convergent subsequence from any subsequence of $(\widehat{\boldsymbol{\theta}}, \widehat{\Lambda}, \widehat{F})$. By the Glivenko-Cantelli theorem and the property of the Kullback-Leibler information, the limit of the convergent subsequence must be the true parameters $(\boldsymbol{\theta}_0, \Lambda_0, F_0)$. The asymptotic distribution of $\widehat{\boldsymbol{\theta}}, \widehat{\Lambda}$ and $\widehat{F}$ follows from the arguments used in the proof of Theorem S.3.

2. NUMERICAL ALGORITHMS

In this section, we present the EM algorithms (treating H as missing data) for all the designs considered in this paper.

2.1 *Cross-sectional studies*

Suppose that there are J distinct values of $\mathbf{X}$, denoted by $\mathbf{x}_1, \ldots, \mathbf{x}_J$. Let $F\{\mathbf{x}_j\}$ be the jump size of F at $\mathbf{x}_j$. To incorporate the restriction that $\sum_j F\{\mathbf{x}_j\} = 1$, we estimate $\log(F\{\mathbf{x}_j\}/F\{\mathbf{x}_J\})$ $(j = 1, \ldots, J-1)$ instead. Define $\mathcal{D}_{jkl} = \mathcal{D}(\mathbf{x}_j, h_k, h_l)$, $\mathcal{Z}_{jkl} = \mathcal{Z}(\mathbf{x}_j, h_k, h_l)$,

$$\mathcal{W}_{kl} = \begin{pmatrix} I(h_k = h_1) + I(h_l = h_1) \\ \vdots \\ I(h_k = h_{K-1}) + I(h_l = h_{K-1}) \end{pmatrix}, \mathcal{M}_{jkl} = \begin{pmatrix} \mathcal{D}_{jkl} \\ I(j = 1) \\ \vdots \\ I(j = J-1) \end{pmatrix}, \boldsymbol{\delta} = \begin{pmatrix} \boldsymbol{\zeta} \\ \log(F\{\mathbf{x}_1\}/F\{\mathbf{x}_J\}) \\ \vdots \\ \log(F\{\mathbf{x}_{(J-1)}\}/F\{\mathbf{x}_J\}) \end{pmatrix}.$$

To incorporate the constraint that $\sum_k \pi_k = 1$, we define $\nu_k = \log(\pi_k/\pi_K)$ and $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_{K-1})^{\mathrm{T}}$, so $P_{\boldsymbol{\gamma}}(H = (h_k, h_l)) = \exp(\boldsymbol{\nu}^{\mathrm{T}}\mathcal{W}_{kl})/\sum_{k,l}\exp(\boldsymbol{\nu}^{\mathrm{T}}\mathcal{W}_{kl})$. Under $\mathbf{X} = \mathbf{x}_j$ and $H = (h_k, h_l)$,

$$\frac{\exp\{\boldsymbol{\zeta}^{\mathrm{T}}\mathcal{D}(\mathbf{X}, H)\}f(\mathbf{X})}{\int_{\mathbf{x}}\exp\{\boldsymbol{\zeta}^{\mathrm{T}}\mathcal{D}(\mathbf{x}, H)\}dF(\mathbf{x})} = \frac{\exp(\boldsymbol{\delta}^{\mathrm{T}}\mathcal{M}_{jkl})}{\sum_{j'}\exp(\boldsymbol{\delta}^{\mathrm{T}}\mathcal{M}_{j'kl})}.$$

The complete-data log-likelihood is

$$l_n^c = \sum_{i,j,k,l} I\{\mathbf{X}_i = \mathbf{x}_j, H_i = (h_k, h_l)\}\Bigg\{\log P_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\xi}}(\mathbf{Y}_i|\mathbf{x}_j, (h_k, h_l)) + \boldsymbol{\nu}^{\mathrm{T}}\mathcal{W}_{kl} + \boldsymbol{\delta}^{\mathrm{T}}\mathcal{M}_{jkl}$$

$$- \log\sum_{j'}\exp(\boldsymbol{\delta}^{\mathrm{T}}\mathcal{M}_{j'kl})\Bigg\} - n\log\sum_{k,l}\exp(\boldsymbol{\nu}^{\mathrm{T}}\mathcal{W}_{kl}).$$

In the E-step, we evaluate $E\{I(\mathbf{X}_i = \mathbf{x}_j, H_i = (h_k, h_l))\big|\mathbf{X}_i, \mathbf{Y}_i, G_i\}$, which can be shown to be

$$\omega_{ijkl} \equiv \frac{I\{\mathbf{X}_i = \mathbf{x}_j, (h_k, h_l) \in \mathcal{S}(G_i)\}P_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\xi}}(\mathbf{Y}_i|\mathbf{x}_j, (h_k, h_l))e^{\boldsymbol{\nu}^{\mathrm{T}}\mathcal{W}_{kl} + \boldsymbol{\delta}^{\mathrm{T}}\mathcal{M}_{jkl}}/\sum_{j'}e^{\boldsymbol{\delta}^{\mathrm{T}}\mathcal{M}_{j'kl}}}{\sum_{(h_{k'}, h_{l'}) \in \mathcal{S}(G_i)} P_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\xi}}(\mathbf{Y}_i|\mathbf{x}_j, (h_{k'}, h_{l'}))e^{\boldsymbol{\nu}^{\mathrm{T}}\mathcal{W}_{k'l'} + \boldsymbol{\delta}^{\mathrm{T}}\mathcal{M}_{jk'l'}}/\sum_{j'}e^{\boldsymbol{\delta}^{\mathrm{T}}\mathcal{M}_{j'k'l'}}}.$$

In the M-step, we maximize l_n^c with $I\{\mathbf{X}_i = \mathbf{x}_j, H_i = (h_k, h_l)\}$ replaced by ω_{ijkl}. The maximization is carried out by the quasi-Newton algorithm. Starting with $\boldsymbol{\alpha} = \mathbf{0}$, $\boldsymbol{\beta} = \mathbf{0}$, $\boldsymbol{\delta} = \mathbf{0}$ and $\nu_k = \log(\widetilde{\pi}_k/\widetilde{\pi}_K)$ $(k = 1, \ldots, K-1)$, where the $\widetilde{\pi}_k$'s are the MLEs of the π_k's based on G_i

$(i = 1, \ldots, n)$, we iterate between the E-step and M-step until the change in the observed-data log-likelihood is negligible.

We can estimate the limiting covariance matrix of $\widehat{\boldsymbol{\theta}}$ and $\widehat{F}$ by inverting the (observed-data) information matrix for all the parameters including the jump sizes of $\widehat{F}$. The information matrix is obtained via the Louis (1982) formula. We can also estimate the limiting covariance matrix of $\widehat{\boldsymbol{\theta}}$ by using the profile likelihood function $pl_n(\boldsymbol{\theta}) \equiv \max_F \log L_n(\boldsymbol{\theta}, F)$. Particularly, the (s,t)th element of the inverse covariance matrix of $\widehat{\boldsymbol{\theta}}$ can be estimated by $-\epsilon_n^{-2}\Big\{pl_n(\widehat{\boldsymbol{\theta}} + \epsilon_n \mathbf{e}_s + \epsilon_n \mathbf{e}_t) - pl_n(\widehat{\boldsymbol{\theta}} + \epsilon_n \mathbf{e}_s) - pl_n(\widehat{\boldsymbol{\theta}} + \epsilon_n \mathbf{e}_t) + pl_n(\widehat{\boldsymbol{\theta}})\Big\}$, where ϵ_n is a constant of order $n^{-1/2}$, and $\mathbf{e}_s$, and $\mathbf{e}_t$ are the sth and tth canonical vectors. We calculate $pl_n(\boldsymbol{\theta})$ via the EM algorithm by holding $\boldsymbol{\theta}$ constant in both the E-step and M-step.

2.2 *Case-control studies with rare disease*

We adopt the notation of Section 2.1. The E-step of the EM algorithm is the same as in Section 2.1. In the M-step, the objective function to be maximized is

$$\begin{aligned}\widetilde{l}_n(\boldsymbol{\beta}, \boldsymbol{\nu}, \boldsymbol{\delta}) = \sum_{i,j,k,l} \omega_{ijkl} \Big\{ Y_i \boldsymbol{\beta}^{\mathrm{T}} \mathcal{Z}_{jkl} + \boldsymbol{\nu}^{\mathrm{T}} \mathcal{W}_{kl} + \boldsymbol{\delta}^{\mathrm{T}} \mathcal{M}_{jkl} - \log\Big(\sum_{j'} e^{\boldsymbol{\delta}^{\mathrm{T}} \mathcal{M}_{j'kl}} \Big) \Big\} \\ - n_1 \log\Bigg\{ \sum_{j,k,l} e^{\boldsymbol{\beta}^{\mathrm{T}} \mathcal{Z}_{jkl} + \boldsymbol{\nu}^{\mathrm{T}} \mathcal{W}_{kl}} \frac{e^{\boldsymbol{\delta}^{\mathrm{T}} \mathcal{M}_{jkl}}}{\sum_{j'} e^{\boldsymbol{\delta}^{\mathrm{T}} \mathcal{M}_{j'kl}}} \Bigg\} - n_0 \log\Bigg\{ \sum_{k,l} e^{\boldsymbol{\nu}^{\mathrm{T}} \mathcal{W}_{kl}} \Bigg\},\end{aligned}$$

where ω_{ijkl} is defined in Section 2.1. We use the Louis formula to calculate the observed-data information matrix, whose inverse is used to estimate the covariance matrix of the NPMLEs; the profile likelihood method can also be used to estimate the covariance matrix of $\widehat{\boldsymbol{\theta}}$.

2.3 *Case-control studies with known disease rate*

The E-step is similar to that of Section 2.1. In the M-step, we use the Lagrange multiplier λ for the constraint

$$\sum_{j,k,l} P_{\alpha,\boldsymbol{\beta}}(Y = 1 | \mathbf{x}_j, h_k, h_l) \frac{\exp(\boldsymbol{\nu}^{\mathrm{T}} \mathcal{W}_{kl} + \boldsymbol{\delta}^{\mathrm{T}} \mathcal{M}_{jkl})}{\sum_{j'} \exp(\boldsymbol{\delta}^{\mathrm{T}} \mathcal{M}_{j'kl})} = p_1 \sum_{k,l} \exp(\boldsymbol{\nu}^{\mathrm{T}} \mathcal{W}_{kl}). \tag{S.12}$$

The objective function to be maximized in the M-step is

$$\widetilde{l}_n(\alpha,\boldsymbol{\beta},\boldsymbol{\nu},\boldsymbol{\delta},\lambda)=\sum_{i,j,k,l}\omega_{ijkl}\Big\{\log P_{\alpha,\boldsymbol{\beta}}(Y_i|\mathbf{x}_j,h_k,h_l)+\boldsymbol{\nu}^{\mathrm{T}}\mathcal{W}_{kl}+\boldsymbol{\delta}^{\mathrm{T}}\mathcal{M}_{jkl}-\log\Big(\sum_{j'}e^{\boldsymbol{\delta}^{\mathrm{T}}\mathcal{M}_{j'kl}}\Big)\Big\}$$
$$-\lambda\Big\{\sum_{j,k,l}P_{\alpha,\boldsymbol{\beta}}(Y=1|\mathbf{x}_j,h_k,h_l)e^{\boldsymbol{\nu}^{\mathrm{T}}\mathcal{W}_{kl}+\boldsymbol{\delta}^{\mathrm{T}}\mathcal{M}_{jkl}}\Big/\sum_{j'}e^{\boldsymbol{\delta}^{\mathrm{T}}\mathcal{M}_{j'kl}}-p_1\sum_{k,l}e^{\boldsymbol{\nu}^{\mathrm{T}}\mathcal{W}_{kl}}\Big\}$$
$$-n\log\Big\{\sum_{k,l}e^{\boldsymbol{\nu}^{\mathrm{T}}\mathcal{W}_{kl}}\Big\}. \tag{S.13}$$

We can treat λ as a free parameter in (S.13), so that (S.12) is automatically met by setting the derivative with respect to λ to zero. The maximization can be carried out by the quasi-Newton method. The variances and covariances can be estimated by the inverse information matrix or by the profile-likelihood method.

2.4 *Cohort studies*

We present the EM algorithm for the proportional hazards model. Suppose that there are L distinct failure times $t_1,\ldots,t_L$. Let $\Lambda\{t_l\}$ denote the jump size of Λ at t_l, and d_l the number of failures at t_l. In the E-step, we evaluate the conditional expectations

$$\omega_{ijkl}\equiv E\{I(\mathbf{X}_i=\mathbf{x}_j,H_i=(h_k,h_l))\big|\widetilde{Y}_i,\Delta_i,\mathbf{X}_i,G_i\}$$
$$=\frac{I(\mathbf{X}_i=\mathbf{x}_j,(h_k,h_l)\in\mathcal{S}(G_i))R_{ijkl}(\boldsymbol{\beta},\boldsymbol{\nu},\boldsymbol{\delta})/\sum_{j'}\exp(\boldsymbol{\delta}^{\mathrm{T}}\mathcal{M}_{j'kl})}{\sum_{(h_{k'},h_{l'})\in\mathcal{S}(G_i)}R_{ijk'l'}(\boldsymbol{\beta},\boldsymbol{\nu},\boldsymbol{\delta})/\sum_{j'}\exp(\boldsymbol{\delta}^{\mathrm{T}}\mathcal{M}_{j'k'l'})},$$

where $R_{ijkl}(\boldsymbol{\beta},\boldsymbol{\nu},\boldsymbol{\delta})=\exp(\Delta_i\boldsymbol{\beta}^{\mathrm{T}}\mathcal{Z}_{jkl}+\boldsymbol{\nu}^{\mathrm{T}}\mathcal{W}_{kl}+\boldsymbol{\delta}^{\mathrm{T}}\mathcal{M}_{jkl}-e^{\boldsymbol{\beta}^{\mathrm{T}}\mathcal{Z}_{jkl}}\sum_{m:t_m\le\widetilde{Y}_i}\Lambda\{t_m\})$. In the M-step, we maximize

$$\widetilde{l}_n(\boldsymbol{\beta},\boldsymbol{\nu},\boldsymbol{\delta},\Lambda)=\sum_{i,j,k,l}\omega_{ijkl}\Delta_i\log\Lambda\{\widetilde{Y}_i\}+\sum_{i,j,k,l}\omega_{ijkl}\Big(\Delta_i\boldsymbol{\beta}^{\mathrm{T}}\mathcal{Z}_{jkl}+\boldsymbol{\nu}^{\mathrm{T}}\mathcal{W}_{kl}+\boldsymbol{\delta}^{\mathrm{T}}\mathcal{M}_{jkl}$$
$$-\log\Big\{\sum_{j'}\exp(\boldsymbol{\delta}^{\mathrm{T}}\mathcal{M}_{j'kl})\Big\}-e^{\boldsymbol{\beta}^{\mathrm{T}}\mathcal{Z}_{jkl}}\sum_{m:t_m\le\widetilde{Y}_i}\Lambda\{t_m\}\Big)-n\log\Big\{\sum_{k,l}\exp(\boldsymbol{\nu}^{\mathrm{T}}\mathcal{W}_{kl})\Big\}.$$

The estimate for $\Lambda\{t_m\}$ is given explicitly by $d_m/\sum_{i:\widetilde{Y}_i\ge t_m}\sum_{j,k,l}\omega_{ijkl}e^{\boldsymbol{\beta}^{\mathrm{T}}\mathcal{Z}_{jkl}}$, and the estimate for $\boldsymbol{\beta}$ solves the equation

$$\sum_{i,j,k,l}\omega_{ijkl}\Delta_i\mathcal{Z}_{jkl}-\sum_{m=1}^{L}d_m\frac{\sum_{i:\widetilde{Y}_i\ge t_m}\sum_{j,k,l}\omega_{ijkl}\mathcal{Z}_{jkl}e^{\boldsymbol{\beta}^{\mathrm{T}}\mathcal{Z}_{jkl}}}{\sum_{i:\widetilde{Y}_i\ge t_m}\sum_{j,k,l}\omega_{ijkl}e^{\boldsymbol{\beta}^{\mathrm{T}}\mathcal{Z}_{jkl}}}=0.$$

The remaining parameters can be estimated by maximizing

$$\sum_{i,j,k,l} \omega_{ijkl} \Big[\boldsymbol{\nu}^{\mathrm{T}} \mathcal{W}_{kl} + \boldsymbol{\delta}^{\mathrm{T}} \mathcal{M}_{jkl} - \log\Big\{ \sum_{j'} \exp(\boldsymbol{\delta}^{\mathrm{T}} \mathcal{M}_{j'kl}) \Big\} \Big] - n \log\Big\{ \sum_{k,l} \exp(\boldsymbol{\nu}^{\mathrm{T}} \mathcal{W}_{kl}) \Big\}.$$

We can estimate the asymptotic variances and covariances by the inverse information matrix or the profile-likelihood method. For other transformation models, we may use the Laplace transformation to convert the estimation problem into that of the proportional hazards model with a random effect; see Zeng and Lin (2007).

3. NUMERICAL RESULTS

We conducted simulation studies in the set-up of Chen et al. (2009). Specifically, we generated haplotypes under HWE from the distribution given in Table 1 of Chen et al. (2009) and generated a binary environmental covariate X with $\Pr(X = 1) = 0.3$, $\zeta_{1,3} = 0$ or $-.4$ and $\zeta_{1,j} = 0$ $(j \neq 3)$. Given H and X, the disease status was generated from model (13) of Chen et al. (2009).

For each simulated data set, we calculated the proposed estimator of $\boldsymbol{\beta}$ allowing for gene-environment dependence and the Lin-Zeng estimator assuming gene-environment independence, denoted as $\widehat{\boldsymbol{\beta}}_{\text{dep}}$ and $\widehat{\boldsymbol{\beta}}_{\text{ind}}$, respectively. Given these two estimators, we constructed two empirical Bayes estimators using formula (7) of Chen et al. (2009). Specifically, the multivariate shrinkage estimator of $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}}_{\text{EB1}} = \widehat{\boldsymbol{\beta}}_{\text{dep}} + \mathbf{K}(\widehat{\boldsymbol{\beta}}_{\text{ind}} - \widehat{\boldsymbol{\beta}}_{\text{dep}}),$$

where $\mathbf{K} = \mathbf{V}\big[\mathbf{V} + (\widehat{\boldsymbol{\beta}}_{\text{ind}} - \widehat{\boldsymbol{\beta}}_{\text{dep}})(\widehat{\boldsymbol{\beta}}_{\text{ind}} - \widehat{\boldsymbol{\beta}}_{\text{dep}})^{\text{T}}\big]^{-\mathbf{1}}$, and $\mathbf{V}$ is the estimated covariance matrix of $(\widehat{\boldsymbol{\beta}}_{\text{ind}} - \widehat{\boldsymbol{\beta}}_{\text{dep}})$; the component-wise shrinkage estimator of the jth component of $\boldsymbol{\beta}$ is

$$\widehat{\beta}_{\text{EB2},j} = \widehat{\beta}_{\text{dep},j} + k_j(\widehat{\beta}_{\text{ind},j} - \widehat{\beta}_{\text{dep},j}),$$

where $\widehat{\beta}_{\text{ind},j}$ and $\widehat{\beta}_{\text{dep},j}$ are the jth components of $\widehat{\boldsymbol{\beta}}_{\text{ind}}$ and $\widehat{\boldsymbol{\beta}}_{\text{dep}}$, $k_j = v_j/\big[v_j + (\widehat{\beta}_{\text{ind},j} - \widehat{\beta}_{\text{dep},j})^2\big]$, and v_j is the jth diagonal element of $\mathbf{V}$.

Write $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\text{T}}, \boldsymbol{\chi}^{\text{T}})^{\text{T}}$, where $\boldsymbol{\chi}$ denotes all nuisance parameters (including finite-dimensional nuisance parameters and jump sizes of nuisance functions). Also, let $\boldsymbol{\theta}^*_{\text{ind}}$ and $\boldsymbol{\theta}^*_{\text{dep}}$ be the probability limits of $\widehat{\boldsymbol{\theta}}_{\text{ind}}$ and $\widehat{\boldsymbol{\theta}}_{\text{dep}}$. We note the following representations

$$\widehat{\boldsymbol{\beta}}_{\text{ind}} - \boldsymbol{\beta}^*_{\text{ind}} = \begin{pmatrix}\mathbf{I}_p & \mathbf{0}\end{pmatrix} \mathcal{I}^{-1}_{\text{ind}}(\boldsymbol{\theta}^*_{\text{ind}}) \sum_{i=1}^{n} U_{\text{ind},i}(\boldsymbol{\theta}^*_{\text{ind}}) + o_p(n^{-1/2}),$$

and

$$\widehat{\boldsymbol{\beta}}_{\text{dep}} - \boldsymbol{\beta}^*_{\text{dep}} = \begin{pmatrix}\mathbf{I}_p & \mathbf{0}\end{pmatrix} \mathcal{I}^{-1}_{\text{dep}}(\boldsymbol{\theta}^*_{\text{dep}}) \sum_{i=1}^{n} U_{\text{dep},i}(\boldsymbol{\theta}^*_{\text{dep}}) + o_p(n^{-1/2}),$$

where $U_{\text{ind},i}(\boldsymbol{\theta})$ and $U_{\text{dep},i}(\boldsymbol{\theta})$ are the ith subject's contributions to the score functions of $\boldsymbol{\theta}$ under the Lin-Zeng and proposed methods, respectively, $\mathcal{I}_{\text{ind}}(\boldsymbol{\theta})$ and $\mathcal{I}_{\text{dep}}(\boldsymbol{\theta})$ are the corresponding information matrices, $\mathbf{I}_p$ is the $p \times p$ identity matrix, and $\mathbf{0}$ is the $p \times q$ zero matrix, with p and q being the dimensions of $\boldsymbol{\beta}$ and $\boldsymbol{\chi}$, respectively. Thus, we estimate the covariance matrices for $\widehat{\boldsymbol{\beta}}_{\text{ind}}$ and $\widehat{\boldsymbol{\beta}}_{\text{dep}}$ as follows:

$$\widehat{\text{var}}(\widehat{\boldsymbol{\beta}}_{\text{ind}}) \equiv \begin{pmatrix} \mathbf{I}_p & \mathbf{0} \end{pmatrix} \mathcal{I}_{\text{ind}}^{-1}(\widehat{\boldsymbol{\theta}}_{\text{ind}}) \Big\{ \sum_{i=1}^{n} U_{\text{ind},i}(\widehat{\boldsymbol{\theta}}_{\text{ind}}) U_{\text{ind},i}^{\text{T}}(\widehat{\boldsymbol{\theta}}_{\text{ind}}) \Big\} \mathcal{I}_{\text{ind}}^{-1}(\widehat{\boldsymbol{\theta}}_{\text{ind}}) \begin{pmatrix} \mathbf{I}_p & \mathbf{0} \end{pmatrix}^{\text{T}},$$

$$\widehat{\text{var}}(\widehat{\boldsymbol{\beta}}_{\text{dep}}) \equiv \begin{pmatrix} \mathbf{I}_p & \mathbf{0} \end{pmatrix} \mathcal{I}_{\text{dep}}^{-1}(\widehat{\boldsymbol{\theta}}_{\text{dep}}) \Big\{ \sum_{i=1}^{n} U_{\text{dep},i}(\widehat{\boldsymbol{\theta}}_{\text{dep}}) U_{\text{dep},i}^{\text{T}}(\widehat{\boldsymbol{\theta}}_{\text{dep}}) \Big\} \mathcal{I}_{\text{dep}}^{-1}(\widehat{\boldsymbol{\theta}}_{\text{dep}}) \begin{pmatrix} \mathbf{I}_p & \mathbf{0} \end{pmatrix}^{\text{T}},$$

$$\widehat{\text{cov}}(\widehat{\boldsymbol{\beta}}_{\text{ind}}, \widehat{\boldsymbol{\beta}}_{\text{dep}}) \equiv \begin{pmatrix} \mathbf{I}_p & \mathbf{0} \end{pmatrix} \mathcal{I}_{\text{ind}}^{-1}(\widehat{\boldsymbol{\theta}}_{\text{ind}}) \Big\{ \sum_{i=1}^{n} U_{\text{ind},i}(\widehat{\boldsymbol{\theta}}_{\text{ind}}) U_{\text{dep},i}^{\text{T}}(\widehat{\boldsymbol{\theta}}_{\text{dep}}) \Big\} \mathcal{I}_{\text{dep}}^{-1}(\widehat{\boldsymbol{\theta}}_{\text{dep}}) \begin{pmatrix} \mathbf{I}_p & \mathbf{0} \end{pmatrix}^{\text{T}}.$$

The simulation results for the dominant and recessive models are presented in Table S.1, in the same format as Tables 2 and 3 of Chen et al. (2009). Our results for the Lin-Zeng estimator (i.e., $\widehat{\boldsymbol{\beta}}_{\text{ind}}$) are similar to those of Chen et al.'s (2009) model-based estimator, especially under the recessive model. Under the dominant model, the proposed estimator (i.e., $\widehat{\boldsymbol{\beta}}_{\text{dep}}$) tends to be more efficient than Chen et al.'s (2009) model-free estimator, particularly in estimating gene-environment interactions. The efficiency gain is much more substantial under the recessive model, for both main effects and interactions. Consequently, our empirical Bayes estimators are more efficient than Chen et al.'s, especially under the recessive model.

Table S.1. *Simulation results for the empirical Bayes estimators under dominant and recessive models*

		$n_1 = n_0 = 150$		$n_1 = n_0 = 300$		$n_1 = n_0 = 600$	
		MSE(Bias)		MSE(Bias)		MSE(Bias)	
Dominant Model		H	$H \times X$	H	$H \times X$	H	$H \times X$
$\zeta_{1,3} = 0$	$\widehat{\beta}_{\text{dep}}$	.109(-.016)	.292(.024)	.054(-.008)	.141(.003)	.025(-.006)	.069(-.007)
	$\widehat{\beta}_{\text{ind}}$	.097(-.001)	.203(-.002)	.049(.004)	.095(-.022)	.023(.004)	.047(-.031)
	$\widehat{\beta}_{\text{EB1}}$	.106(-.018)	.274(.024)	.054(-.007)	.137(.001)	.025(-.005)	.067(-.008)
	$\widehat{\beta}_{\text{EB2}}$	.101(-.013)	.234(.016)	.052(-.003)	.118(-.007)	.024(-.002)	.059(-.016)
$\zeta_{1,3} = -.4$	$\widehat{\beta}_{\text{dep}}$	.111(-.008)	.310(.044)	.051(-.005)	.156(.005)	.026(-.005)	.074(-.013)
	$\widehat{\beta}_{\text{ind}}$	.120(.133)	.375(-.398)	.062(.128)	.275(-.416)	.038(.122)	.225(-.418)
	$\widehat{\beta}_{\text{EB1}}$	.107(-.004)	.293(.026)	.051(-.001)	.155(-.008)	.026(-.002)	.074(-.022)
	$\widehat{\beta}_{\text{EB2}}$	.107(.028)	.290(-.072)	.051(.021)	.163(-.079)	.026(.012)	.081(-.065)
Recessive Model							
$\zeta_{1,3} = 0$	$\widehat{\beta}_{\text{dep}}$	.099(-.048)	.261(-.049)	.049(-.027)	.127(-.031)	.029(-.023)	.073(-.023)
	$\widehat{\beta}_{\text{ind}}$	.095(-.057)	.197(-.043)	.047(-.030)	.092(-.034)	.026(-.023)	.052(-.031)
	$\widehat{\beta}_{\text{EB1}}$	.097(-.050)	.239(-.046)	.049(-.028)	.115(-.030)	.028(-.023)	.066(-.024)
	$\widehat{\beta}_{\text{EB2}}$	.096(-.054)	.225(-.047)	.048(-.030)	.108(-.031)	.027(-.024)	.062(-.026)
$\zeta_{1,3} = -.4$	$\widehat{\beta}_{\text{dep}}$	.087(-.050)	.339(-.065)	.044(-.026)	.173(-.031)	.026(-.022)	.088(-.032)
	$\widehat{\beta}_{\text{ind}}$	.095(.117)	.778(-.720)	.065(.147)	.621(-.699)	.047(.149)	.536(-.678)
	$\widehat{\beta}_{\text{EB1}}$	.087(-.039)	.352(-.107)	.044(-.020)	.177(-.053)	.026(-.018)	.090(-.044)
	$\widehat{\beta}_{\text{EB2}}$	.087(-.029)	.370(-.170)	.044(-.015)	.185(-.080)	.026(-.015)	.094(-.069)

NOTE: $\widehat{\beta}_{\text{dep}}$ and $\widehat{\beta}_{\text{ind}}$ pertain to the proposed estimator allowing for gene-environment dependence and the Lin-Zeng estimator assuming gene-environment independence, respectively. H and $H \times X$ stand for main haplotype effect and haplotype-environment interaction. MSE and Bias are the mean square error and bias of the parameter estimator. Each entry is based on 1,000 replicates.