Supplementary material to "Efficient inference of parent-of-origin effect using case-control mother-child genotype data" by

Yuang Tian[1], Hong Zhang[2], Alexandre Bureau[3], Hagit Hochner[4], and Jinbo Chen[5]

[1]Shanghai Center for Mathematical Sciences, Fudan University, Shanghai, China

[2]Department of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei, Anhui, China

[3]Department of Social and Preventive Medicine, Université Laval, Québec, Canada

[4]Braun School of Public Health, The Hebrew University of Jerusalem, Jerusalem, Israel

[5]Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, U.S.A.

Appendix A Derivation of the profile log-likelihood function (10) in the main text.

We adopt the notations in the main text. The aim is to maximize the empirical log-likelihood function

$$\log L(\boldsymbol{\Theta}, \boldsymbol{\pi}) = l_1(\boldsymbol{\Theta}) + \sum_{u=1}^{n} \log \pi_u$$

under the constraints

$$\sum_{u=1}^{n} L_u(\boldsymbol{\Theta})\pi_u = f \tag{S1}$$

and

$$\sum_{u=1}^{n} \pi_u = 1. \tag{S2}$$

According to the Lagrange multiplier method, the above maximization problem is equivalent to maximizing

$$M(\boldsymbol{\Theta}, \boldsymbol{\pi}, \zeta, \lambda) = l_1(\boldsymbol{\Theta}) + \sum_{u=1}^{n} \log \pi_u - \zeta \left(\sum_{u=1}^{n} \pi_u - 1 \right) - n\lambda \left(\sum_{u=1}^{n} L_u(\boldsymbol{\Theta}) \pi_u - f \right) \tag{S3}$$

with respect to $\boldsymbol{\Theta}$, $\boldsymbol{\pi}$, ζ, and λ. The maximizer of (S3) satisfies

$$\frac{\partial M}{\partial \pi_u}(\boldsymbol{\Theta}, \boldsymbol{\pi}, \zeta, \lambda) = 0, \tag{S4}$$

or equivalently,

$$1 = \zeta \pi_u + n\lambda L_u(\boldsymbol{\Theta}) \pi_u. \tag{S5}$$

Summing the above equation over $u = 1, \ldots, n$, we have

$$\zeta = n(1 - \lambda f). \tag{S6}$$

Plugging this into (S5), we have

$$\pi_u = \frac{1}{n \left\{ 1 + \lambda \left(L_u(\boldsymbol{\Theta}) - f \right) \right\}}. \tag{S7}$$

Plugging (S7) into (S1) and combining (S2), we obtain the following constraint on the multiplier λ:

$$\sum_{u=1}^{n} \frac{L_u(\boldsymbol{\Theta}) - f}{1 + \lambda \left(L_u(\boldsymbol{\Theta}) - f \right)} = 0. \tag{S8}$$

Plugging (S7) and (S6) into (S3), we obtain the resulting profile log-likelihood (10) in the main text, with λ satisfying the constraint (S8).

Appendix B Proof of Lemma 1 in the main text.

We need to prove

$$E \left[\frac{\partial l_{\mathrm{mp}}(\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}} \right] \bigg|_{\boldsymbol{\Theta} = \boldsymbol{\Theta}_0} = 0, \tag{S9}$$

under condition (C2) in the main text.

Denote $Z = (G^m, G^c, X)$, $Z_u = (G_u^m, G_u^c, X_u)$, $W = (g^m, X)$, $W_u = (g_u^m, X_u)$, $p_Z(Z; \boldsymbol{\Theta}) = P_{\boldsymbol{\Theta}}(Y = 1|Z)$, $p_X(W; \boldsymbol{\Theta}) = P_{\boldsymbol{\Theta}}(Y = 1|W)$. Let G denote a vector containing all genotypes except g^m, so that $Z = (G, W)$. For any measurable function t, let $E\{t(Z)|Y = 1\}$

and $E\{t(Z)|Y=0\}$ be denoted by $E_1\{t(Z)\}$ and $E_0\{t(Z)\}$, respectively. Furthermore, let $E\{t(Z)\}$ denote the expectation of $t(Z)$ in the population where cases and controls arise.

We assume that the expectations appearing in the following context exist and are finite, as will be proved at the end of this section under condition (C2).

For any measurable function t, it can be immediately checked that the following equations hold true:

$$E_1\{t(Z)\} = E\left\{\frac{p_Z(Z;\mathbf{\Theta}_0)}{f}t(Z)\right\} \text{ and } E_0\{t(Z)\} = E\left\{\frac{1-p_Z(Z;\mathbf{\Theta}_0)}{1-f}t(Z)\right\}, \tag{S10}$$

$$E\left\{\sum_{u=1}^{n} t(Z_u)\right\} = nE\left[\{1+\lambda_0(p_Z(Z;\mathbf{\Theta}_0)-f)\}t(Z)\right], \tag{S11}$$

$$E\left\{\sum_{u=1}^{n}\frac{(2Y_u-1)t(Z_u)}{1-Y_u+(2Y_u-1)p_Z(Z_u;\mathbf{\Theta}_0)}\right\} = n\lambda_0 E\{t(Z)\}. \tag{S12}$$

Note that

$$\begin{aligned} l_{\mathrm{mp}}(\mathbf{\Theta}) &= \sum_{u=1}^{n}\log P_{\mathbf{\Theta}}(Y_u|Z_u) + \sum_{u=1}^{n}\log P_{\mathbf{\Theta}}(Z_u|W_u) - \sum_{u=1}^{n}\log\left[n\{1+\lambda_0(p_W(W_u;\mathbf{\Theta})-f)\}\right] \\ &:= \ell^{(1)}(\mathbf{\Theta}) + \ell^{(2)}(\mathbf{\Theta}) - \ell^{(3)}(\mathbf{\Theta}). \end{aligned} \tag{S13}$$

In view of (S12), we have that

$$E\left\{\frac{\partial}{\partial\mathbf{\Theta}}\ell^{(1)}(\mathbf{\Theta}_0)\right\} = E\left\{\sum_{u=1}^{n}\frac{(2Y_u-1)\partial p_Z(Z_u;\mathbf{\Theta}_0)/\partial\mathbf{\Theta}}{1-Y_u+(2Y_u-1)p_Z(Z_u;\mathbf{\Theta}_0)}\right\} = n\lambda_0 E\left\{\frac{\partial p_Z(Z;\mathbf{\Theta}_0)}{\partial\mathbf{\Theta}}\right\}. \tag{S14}$$

In view of (S10), we have that

$$\begin{aligned} E\left\{\frac{\partial}{\partial\mathbf{\Theta}}\ell^{(2)}(\mathbf{\Theta}_0)\right\} &= n_1 E\left\{\frac{p_Z(Z;\mathbf{\Theta}_0)}{f}\frac{\partial}{\partial\mathbf{\Theta}}\log P_{\mathbf{\Theta}_0}(Z|W)\right\} \qquad \text{(S15)} \\ &\quad + n_0 E\left\{\frac{1-p_Z(Z;\mathbf{\Theta}_0)}{1-f}\frac{\partial}{\partial\mathbf{\Theta}}\log P_{\mathbf{\Theta}_0}(Z|W)\right\} \\ &= \frac{n_0}{1-f}E\left\{\frac{\partial}{\partial\mathbf{\Theta}}\log P_{\mathbf{\Theta}_0}(Z|W)\right\} + n\lambda_0 E\left\{p_Z(Z;\mathbf{\Theta}_0)\frac{\partial}{\partial\mathbf{\Theta}}\log P_{\mathbf{\Theta}_0}(Z|W)\right\} \\ &= 0 + n\lambda_0 E\left\{P_{\mathbf{\Theta}_0}(Y=1|G,W)\frac{\partial}{\partial\mathbf{\Theta}}\log P_{\mathbf{\Theta}_0}(G|W)\right\} \\ &= n\lambda_0 E\left\{\sum_{g\in\mathcal{G}}P_{\mathbf{\Theta}_0}(Y=1|G=g,W)\frac{\partial}{\partial\mathbf{\Theta}}P_{\mathbf{\Theta}_0}(G=g|W)\right\}, \qquad \text{(S16)} \end{aligned}$$

where $\mathcal{G}$ is the set of all possible values of G. In view of (S11)-(S16), we have that

$$
\begin{aligned}
E\left\{\frac{\partial}{\partial \boldsymbol{\Theta}} \ell^{(3)}\left(\boldsymbol{\Theta}_0\right)\right\} &= \lambda_0 E\left[\sum_{u=1}^n \frac{\partial p_W\left(W_u ; \boldsymbol{\Theta}_0\right) / \partial \boldsymbol{\Theta}}{1+\lambda_0\left\{p_W\left(W_u ; \boldsymbol{\Theta}_0\right)-f\right\}}\right] \\
&= n \lambda_0 E\left[\frac{1+\lambda_0\left\{p_Z\left(Z ; \boldsymbol{\Theta}_0\right)-f\right\}}{1+\lambda_0\left\{p_W\left(W ; \boldsymbol{\Theta}_0\right)-f\right\}} \frac{\partial p_W\left(W ; \boldsymbol{\Theta}_0\right)}{\partial \boldsymbol{\Theta}}\right] \\
&= n \lambda_0 E\left[\frac{\partial p_W\left(W ; \boldsymbol{\Theta}_0\right)}{\partial \boldsymbol{\Theta}}\right] \qquad \text{(S17)} \\
&= n \lambda_0 E\left\{\frac{\partial}{\partial \boldsymbol{\Theta}} \sum_{g \in \mathcal{G}} P_{\boldsymbol{\Theta}_0}(Y=1 | G=g, W) P_{\boldsymbol{\Theta}_0}(G=g | W)\right\} \\
&= n \lambda_0 E\left\{\sum_{g \in \mathcal{G}} \frac{\partial}{\partial \boldsymbol{\Theta}} P_{\boldsymbol{\Theta}_0}(Y=1 | G=g, W) P_{\boldsymbol{\Theta}_0}(G=g | W)\right. \\
&\qquad \left.+\sum_{g \in \mathcal{G}} P_{\boldsymbol{\Theta}_0}(Y=1 | G=g, W) \frac{\partial}{\partial \boldsymbol{\Theta}} P_{\boldsymbol{\Theta}_0}(G=g | W)\right\} \\
&= E\left\{\frac{\partial}{\partial \boldsymbol{\Theta}} \ell^{(1)}\left(\boldsymbol{\Theta}_0\right)\right\}+E\left\{\frac{\partial}{\partial \boldsymbol{\Theta}} \ell^{(2)}\left(\boldsymbol{\Theta}_0\right)\right\}, \qquad \text{(S18)}
\end{aligned}
$$

which is equivalent to (S9) according to (S13).

Now we show that the above expectations exist and are finite. First,

$$
E\left\{\left|\sum_{u=1}^n t\left(Z_u\right)\right|\right\} \leq n E\left[\left\{1+\lambda_0\left(p_Z\left(Z ; \boldsymbol{\Theta}_0\right)-f\right)\right\}|t(Z)|\right],
$$

$$
E\left\{\left|\sum_{u=1}^n \frac{\left(2 Y_u-1\right) t\left(Z_u\right)}{1-Y_u+\left(2 Y_u-1\right) p_Z\left(Z_u ; \boldsymbol{\Theta}_0\right)}\right|\right\} \leq\left(\frac{n_1}{f}+\frac{n_0}{1-f}\right) E\{|t(Z)|\}.
$$

Then, following the process we went through before, we have

$$
E\left\{\left|\frac{\partial}{\partial \boldsymbol{\Theta}} \ell^{(1)}\left(\boldsymbol{\Theta}_0\right)\right|\right\} \leq\left(\frac{n_1}{f}+\frac{n_0}{1-f}\right) E\left\{\left|\frac{\partial p_Z\left(Z ; \boldsymbol{\Theta}_0\right)}{\partial \boldsymbol{\Theta}}\right|\right\},
$$

$$
E\left\{\left|\frac{\partial}{\partial \boldsymbol{\Theta}} \ell^{(2)}\left(\boldsymbol{\Theta}_0\right)\right|\right\} \leq n \lambda_0 E\left\{\left|\frac{\partial}{\partial \boldsymbol{\Theta}} \log P_{\boldsymbol{\Theta}_0}(Z | W)\right|\right\},
$$

and

$$
E\left\{\left|\frac{\partial}{\partial \boldsymbol{\Theta}} \ell^{(3)}\left(\boldsymbol{\Theta}_0\right)\right|\right\} \leq n \lambda_0\left(E\left\{\left|\frac{\partial p_Z\left(Z ; \boldsymbol{\Theta}_0\right)}{\partial \boldsymbol{\Theta}}\right|\right\}+E\left\{\left|\frac{\partial}{\partial \boldsymbol{\Theta}} \log P_{\boldsymbol{\Theta}_0}(Z | W)\right|\right\}\right).
$$

Under condition (C2), the above expectations are finite. Consequently,

$$
E\left[\left|\frac{\partial l_{\mathrm{mp}}\left(\boldsymbol{\Theta}_0\right)}{\partial \boldsymbol{\Theta}}\right|\right] \leq E\left\{\left|\frac{\partial}{\partial \boldsymbol{\Theta}} \ell^{(1)}\left(\boldsymbol{\Theta}_0\right)\right|\right\}+E\left\{\left|\frac{\partial}{\partial \boldsymbol{\Theta}} \ell^{(2)}\left(\boldsymbol{\Theta}_0\right)\right|\right\}+E\left\{\left|\frac{\partial}{\partial \boldsymbol{\Theta}} \ell^{(3)}\left(\boldsymbol{\Theta}_0\right)\right|\right\}<\infty .
$$

According to the Dominated Convergence Theorem, all relevant expectations shown above exist and are finite under condition (C2).

Finally, we illustrate that the previously defined λ_0 is indeed the limiting value of the multiplier $\lambda_{\boldsymbol{\Theta}}$, i.e.,

$$E\left[\frac{\partial l_0(\boldsymbol{\Theta}, \lambda_0)}{\partial \lambda}\right] = E\left[\sum_{u=1}^{n} \frac{p_W(W_u; \boldsymbol{\Theta}_0) - f}{1 + \lambda_0 \{p_W(W_u; \boldsymbol{\Theta}_0) - f\}}\right] = 0.$$

Applying (S11), we have

$$\begin{aligned} E\left[\sum_{u=1}^{n} \frac{p_W(W_u; \boldsymbol{\Theta}_0) - f}{1 + \lambda_0 \{p_W(W_u; \boldsymbol{\Theta}_0) - f\}}\right] &= nE\left[\frac{1 + \lambda_0 \{p_Z(Z; \boldsymbol{\Theta}_0) - f\}}{1 + \lambda_0 \{p_W(W; \boldsymbol{\Theta}_0) - f\}} \{p_W(W; \boldsymbol{\Theta}_0) - f\}\right] \\ &= E\{p_W(W; \boldsymbol{\Theta}_0) - f\} = 0, \end{aligned}$$

where the second equation follows from $E\{p_Z(Z; \boldsymbol{\Theta}_0) \mid W\} = p_W(W; \boldsymbol{\Theta}_0)$ and the last equation follows from $E\{p_W(W; \boldsymbol{\Theta}_0)\} = P_{\boldsymbol{\Theta}_0}(Y = 1) = f$.

Appendix C Proof of Theorem 1 in the main text.

First we prove (1), i.e., $l_{\mathrm{mp}}(\boldsymbol{\Theta})$ takes the value of negative infinity on the boundary of the parameter space. When $\|\boldsymbol{\beta}\| \to \infty$, with probability tend to 1, there exists a family such that each compatible haplotype combination $h^m_{iu}h^m_{ij}, h^c_{wu}, h^c_{lu}$ satisfies $P(Y_u | h^m_{iu}h^m_{ij}, h^c_{wu}, h^c_{lu}, X_u) \to 0$, which is equivalent to $l_{\mathrm{mp}}(\boldsymbol{\Theta}) \to -\infty$. Then we examine the performance of $l_{\mathrm{mp}}(\boldsymbol{\Theta})$ when $\mu_s \to 0$ for some s. Note that the true value of μ_s is not equal to 0, so with probability tend to 1 we can find a family that both mother and child have diplotype $h_s h_s$. For that family we have $H^m_u = \{h^m_{su}h^m_{su}\}$, $H^c_u = \{(h^c_{su}, h^c_{su})\}$ and $P(h^m_{su}h^m_{su}, h^c_{su}, h^c_{su}) \to 0$, which means $l_{\mathrm{mp}}(\boldsymbol{\Theta}) \to -\infty$. Under condition (C1), the maximizer of $l_{\mathrm{mp}}(\boldsymbol{\Theta})$ is an interior point of the parameter space, which ensures that there exists a solution $\hat{\boldsymbol{\Theta}}_{\mathrm{mp}}$ in the interior of the parameter space.

Then we prove (2), i.e., $\hat{\boldsymbol{\Theta}}_{\mathrm{mp}}$ is consistent for $\boldsymbol{\Theta}_0$. According to Lemma 1 and conditions (C4)-(C5), for any sufficiently small $a > 0$, we have

$$E[l_{\mathrm{mp}}(\boldsymbol{\Theta}_0)] > \sup\{E[l_{\mathrm{mp}}(\boldsymbol{\Theta})] : \boldsymbol{\Theta} \in \partial B_a\},$$

where $B_a = \{\boldsymbol{\Theta} : \|\boldsymbol{\Theta} - \boldsymbol{\Theta}_0\| \leq a\}$ and ∂B_a is the boundary of B_a. It follows that

$$P\left(\lim_{n\to\infty} \inf\{l_{\mathrm{mp}}(\boldsymbol{\Theta}_0) - l_{\mathrm{mp}}(\boldsymbol{\Theta}) : \boldsymbol{\Theta} \in \partial B_a\} > 0\right) = 1$$

according to the Strong Law of Large Number, which means there exists a solution to the score function in B_a with a large n. Using the standard theory for Z-statistic, we can find $\hat{\boldsymbol{\Theta}}_{\mathrm{mp}}$, a series of solutions to the score equation, which strongly consistent for $\boldsymbol{\Theta}_0$.

Next we prove (3), i.e., $\hat{\boldsymbol{\Theta}}_{\mathrm{mp}}$ is asymptotically normal. Under condition (C6) and the consistence of $\hat{\boldsymbol{\Theta}}_{\mathrm{mp}}$, Taylor's series expansion gives that

$$\frac{\partial l_{\mathrm{mp}}(\hat{\boldsymbol{\Theta}}_{\mathrm{mp}})}{\partial \boldsymbol{\Theta}} = \frac{\partial l_{\mathrm{mp}}(\boldsymbol{\Theta}_0)}{\partial \boldsymbol{\Theta}} + \frac{\partial^2 l_{\mathrm{mp}}(\boldsymbol{\Theta}_0)}{\partial \boldsymbol{\Theta} \partial \boldsymbol{\Theta}^\tau} \left(\hat{\boldsymbol{\Theta}}_{\mathrm{mp}} - \boldsymbol{\Theta}_0\right) + \left(\hat{\boldsymbol{\Theta}}_{\mathrm{mp}} - \boldsymbol{\Theta}_0\right) o_p(n). \tag{S19}$$

We have $\partial l_{\mathrm{mp}}(\hat{\boldsymbol{\Theta}}_{\mathrm{mp}})/\partial \boldsymbol{\Theta} = 0$ because $\hat{\boldsymbol{\Theta}}_{\mathrm{mp}}$ is an interior point solution of $\partial l_{\mathrm{mp}}(\boldsymbol{\Theta})/\partial \boldsymbol{\Theta} = 0$. Under conditions (C3) and (C5), we have $\partial^2 l_{\mathrm{mp}}(\boldsymbol{\Theta}_0)/\partial \boldsymbol{\Theta} \partial \boldsymbol{\Theta}^\tau = nA_{\mathrm{mp}}(\boldsymbol{\Theta}_0) + o_p(n)$. Then (S19) can be rewritten as

$$\sqrt{n}\left(A_{\mathrm{mp}}(\boldsymbol{\Theta}_0) + o_p(1)\right)\left(\hat{\boldsymbol{\Theta}}_{\mathrm{mp}} - \boldsymbol{\Theta}_0\right) = -\frac{1}{\sqrt{n}} \frac{\partial l_{\mathrm{mp}}(\boldsymbol{\Theta}_0)}{\partial \boldsymbol{\Theta}}. \tag{S20}$$

Under conditions (C3) and (C7), Levy's Central Limit Theorem gives

$$\frac{1}{\sqrt{n}} \frac{\partial l_{\mathrm{mp}}(\boldsymbol{\Theta}_0)}{\partial \boldsymbol{\Theta}} \xrightarrow{d} N\{0, \Sigma_{\mathrm{mp}}(\boldsymbol{\Theta}_0)\}. \tag{S21}$$

Then, combining (S20) and (S21), we have

$$\sqrt{n}(\hat{\boldsymbol{\Theta}}_{\mathrm{mp}} - \boldsymbol{\Theta}_0) \xrightarrow{d} N\left\{0, A_{\mathrm{mp}}^{-1}(\boldsymbol{\Theta}_0)\, \Sigma_{\mathrm{mp}}(\boldsymbol{\Theta}_0)\, A_{\mathrm{mp}}^{-1}(\boldsymbol{\Theta}_0)\right\}.$$

Finally we show (4), i.e., $\hat{\boldsymbol{\Theta}}_{\mathrm{mp}}$ is asymptotically efficient. It surfaces to show that the profile MLE $\hat{\boldsymbol{\Theta}}_p$ and the modified profile MLE $\hat{\boldsymbol{\Theta}}_{\mathrm{mp}}$ are asymptotic equivalent since the former is asymptotically efficient. Let the corresponding score functions be denoted by

$$U_p(\boldsymbol{\Theta}) = \frac{\partial l_{\mathrm{p}}(\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}} = \frac{\partial l_0(\boldsymbol{\Theta}, \lambda)}{\partial \boldsymbol{\Theta}}\Big|_{\lambda = \lambda(\boldsymbol{\Theta})} \quad \text{and} \quad U_{\mathrm{mp}}(\boldsymbol{\Theta}) = \frac{\partial l_0(\boldsymbol{\Theta}, \lambda_0)}{\partial \boldsymbol{\Theta}},$$

and denote

$$S_{\mathrm{mp}}(\boldsymbol{\Theta}) = \frac{\partial^2 l_0(\boldsymbol{\Theta}, \lambda_0)}{\partial \boldsymbol{\Theta} \partial \boldsymbol{\Theta}^\tau}.$$

Note that the maximizer of the profile log-likelihood $(\hat{\mathbf{\Theta}}_p, \lambda(\hat{\mathbf{\Theta}}_p))$ is consistent for $(\mathbf{\Theta}_0, \lambda_0)$, which can be proved similar to Appendix B. Consequently, we have

$$n^{1/2}\left(\hat{\mathbf{\Theta}}_p - \mathbf{\Theta}_0\right) = -\left\{S_{\mathrm{mp}}\left(\mathbf{\Theta}_0\right)/n\right\}^{-1} n^{-1/2} U_p\left(\mathbf{\Theta}_0\right)\left\{1 + o_p(1)\right\} \tag{S22}$$

and

$$n^{1/2}\left(\hat{\mathbf{\Theta}}_{\mathrm{mp}} - \mathbf{\Theta}_0\right) = -\left\{S_{\mathrm{mp}}\left(\mathbf{\Theta}_0\right)/n\right\}^{-1} n^{-1/2} U_{\mathrm{mp}}\left(\mathbf{\Theta}_0\right)\left\{1 + o_p(1)\right\}. \tag{S23}$$

Using Taylor's series expansion, we have that

$$n^{-1/2}\left\{U_p\left(\mathbf{\Theta}_0\right) - U_{\mathrm{mp}}\left(\mathbf{\Theta}_0\right)\right\} = \frac{1}{n}\frac{\partial^2 l_0\left(\mathbf{\Theta}_0, \lambda_0\right)}{\partial \mathbf{\Theta} \partial \lambda^\tau} n^{1/2}\left\{\lambda\left(\mathbf{\Theta}_0\right) - \lambda_0\right\}\left(1 + o_p(1)\right). \tag{S24}$$

Condition (C8) and the Weak Law of Large Numbers give that

$$\frac{1}{n}\frac{\partial^2 l_0\left(\mathbf{\Theta}_0, \lambda_0\right)}{\partial \mathbf{\Theta} \partial \lambda^\tau} \to 0 \text{ in probability.} \tag{S25}$$

It follows from (S24) and (S25) that

$$n^{-1/2}\left\{U_p\left(\mathbf{\Theta}_0\right) - U_{\mathrm{mp}}\left(\mathbf{\Theta}_0\right)\right\} = o_p(1). \tag{S26}$$

Now we establish the asymptotic equivalence between $\hat{\mathbf{\Theta}}_{\mathrm{mp}}$ and $\hat{\mathbf{\Theta}}_p$:

$$n^{-1/2}\{\hat{\mathbf{\Theta}}_{\mathrm{mp}} - \hat{\mathbf{\Theta}}_p\} = o_p(1) \tag{S27}$$

in view of (S22), (S23), and (S26). According to the asymptotic efficiency of the standard profile MLE $\hat{\mathbf{\Theta}}_p$ (Qin and Lawless, 1994), we have the asymptotic efficiency of $\hat{\mathbf{\Theta}}_{\mathrm{mp}}$.

Appendix D A simulation study for checking the regularity condition (C8) in the main text

We conducted a simulation study to check the rationality of regularity condition (C8) in the main text, that is,

$$E\left\{\partial^2 l_0\left(\mathbf{\Theta}_0, \lambda_0\right)/\partial \mathbf{\Theta} \partial \lambda^\tau\right\} = 0.$$

We used the sample mean $n^{-1}\partial^2 l_0\left(\mathbf{\Theta}_0, \lambda_0\right)/\partial \mathbf{\Theta} \partial \lambda^\tau$ to estimate the expectation. We generated data in the way described in Section 3.1 in the main text, with $\beta_{g^m} = \log(1.8)$,

$\beta_{g^c} = \log(1.5)$, $\beta_X = \log(1.8)$, $\beta_{im} = 0$ or $\beta_{im} = \log(1.5)$, and $\eta = 0$. The simulation results with different sample sizes are summarized in Table S1, which were based on 10,000 replicates of simulations. For various sample sizes, the estimates are very close to 0, which indicates that the assumption that $E\left\{\partial^2 l_0\left(\boldsymbol{\Theta}_0, \lambda_0\right)/\partial\boldsymbol{\Theta}\partial\lambda^\tau\right\} = 0$ approximately holds true in the simulation situation. We also tried some other parameter combinations and reached the same conclusion (results not shown).

Appendix E Proof of Theorem 2 in the main text.

The idea of proof follows from Zangwill (1969) and Wu (1983). Let g_u^m be the target SNP genotype extracted from the joint genotype G_u^m, then we have

$$P_{\boldsymbol{\Theta}}(Y_u, G_u^c, G_u^m, h_u | g_u^m, X_u) = P_{\boldsymbol{\Theta}}(h_u | Y_u, G_u^c, G_u^m, X_u) P_{\boldsymbol{\Theta}}(Y_u, G_u^c, G_u^m | g_u^m, X_u)$$

for any haplotype h_u. Consequently, we can decompose $l_1(\boldsymbol{\Theta})$ as follows:

$$\begin{aligned}\sum_{u=1}^n \log P_{\boldsymbol{\Theta}}(Y_u, G_u^c, G_u^m | g_u^m, X_u) = & \sum_{u=1}^n \log P_{\boldsymbol{\Theta}}(Y_u, G_u^c, G_u^m, h_u | g_u^m, X_u) \\ & - \sum_{u=1}^n \log P_{\boldsymbol{\Theta}}(h_u | Y_u, G_u^c, G_u^m, X_u)\end{aligned}$$

with h_u being any haplotype, which implies that

$$\begin{aligned}&\sum_{u=1}^n \log P_{\boldsymbol{\Theta}}(Y_u, G_u^c, G_u^m | g_u^m, X_u) \\ =& E\left[\sum_{u=1}^n \log P_{\boldsymbol{\Theta}}(Y_u, G_u^c, G_u^m, H_u | g_u^m, X_u)\middle|\boldsymbol{\Theta}', D\right] - E\left[\sum_{u=1}^n \log P_{\boldsymbol{\Theta}}(H_u | Y_u, G_u^c, G_u^m, X_u)\middle|\boldsymbol{\Theta}', D\right]\end{aligned}$$

for any given $\boldsymbol{\Theta}'$, where $D = \{(Y_u, G_u^c, G_u^m, X_u), u = 1, \ldots, n\}$. As a result,

$$\begin{aligned} l_{\mathrm{mp}}(\boldsymbol{\Theta}) =& \sum_{u=1}^n \log P_{\boldsymbol{\Theta}}(Y_u, G_u^c, G_u^m | g_u^m, X_u) - l_2(\boldsymbol{\Theta}, \lambda_0) \\ =& \left(E\left[\sum_{u=1}^n \log P_{\boldsymbol{\Theta}}(Y_u, G_u^c, G_u^m, H_u | g_u^m, X_u)\middle|\boldsymbol{\Theta}', D\right] - l_2(\boldsymbol{\Theta}, \lambda_0)\right) \\ & - E\left[\sum_{u=1}^n \log P_{\boldsymbol{\Theta}}(H_u | Y_u, G_u^c, G_u^m, X_u)\middle|\boldsymbol{\Theta}', D\right] \\ :=& Q(\boldsymbol{\Theta}, \boldsymbol{\Theta}') - H(\boldsymbol{\Theta}, \boldsymbol{\Theta}'). \end{aligned} \tag{S28}$$

Here $Q(\boldsymbol{\Theta}, \boldsymbol{\Theta}')$ is the objective function in the M-step of the rth iteration when $\boldsymbol{\Theta}' = \boldsymbol{\Theta}^{(r)}$. Let $M(\boldsymbol{\Theta}) = \underset{\boldsymbol{\Theta}' \in \Phi}{\operatorname{argmax}} Q(\boldsymbol{\Theta}', \boldsymbol{\Theta})$ denote a point-to-set map on the parameter space Φ, and the parameter sequence $\{\boldsymbol{\Theta}^{(r)}\}$ can be generated by $\boldsymbol{\Theta}^{(r+1)} \in M(\boldsymbol{\Theta}^{(r)})$.

Let Γ denote the set containing all stationary points of l_{mp}. In what follows, we prove that all limit points of $\{\boldsymbol{\Theta}^{(r)}\}$ belong to Γ. We have that

$$\boldsymbol{\Theta}^{(r)} \text{ belongs to a compact set } \Phi_P \subset \Phi \text{ for } r = 1, 2 \ldots \tag{S29}$$

because $l_{\text{mp}}(\boldsymbol{\Theta})$ tends to minus infinity as $\|\boldsymbol{\Theta}\| \to \infty$. We can prove $Q(\boldsymbol{\Theta}^{(r+1)}, \boldsymbol{\Theta}^{(r)}) \geq Q(\boldsymbol{\Theta}^{(r)}, \boldsymbol{\Theta}^{(r)})$ and $H(\boldsymbol{\Theta}^{(r+1)}, \boldsymbol{\Theta}^{(r)}) \leq H(\boldsymbol{\Theta}^{(r)}, \boldsymbol{\Theta}^{(r)})$ using Jensen's inequality. It immediately follows that

$$l_{\text{mp}}(\boldsymbol{\Theta}^{(r+1)}) \geq l_{\text{mp}}(\boldsymbol{\Theta}^{(r)}) \tag{S30}$$

by the definition of $l_{\text{mp}}(\boldsymbol{\Theta})$ given in (S28). Furthermore, for $\boldsymbol{\Theta}^{(r)} \notin \Gamma$, in what follows we prove

$$l_{\text{mp}}(\boldsymbol{\Theta}^{(r+1)}) > l_{\text{mp}}(\boldsymbol{\Theta}^{(r)}). \tag{S31}$$

In fact, if $l_{\text{mp}}(\boldsymbol{\Theta}^{(r+1)}) = l_{\text{mp}}(\boldsymbol{\Theta}^{(r)})$, then $Q(\boldsymbol{\Theta}^{(r+1)}, \boldsymbol{\Theta}^{(r)}) = Q(\boldsymbol{\Theta}^{(r)}, \boldsymbol{\Theta}^{(r)})$ and $H(\boldsymbol{\Theta}^{(r+1)}, \boldsymbol{\Theta}^{(r)}) = H(\boldsymbol{\Theta}^{(r)}, \boldsymbol{\Theta}^{(r)})$. As a result,

$$\frac{\partial l_{\text{mp}}(\boldsymbol{\Theta}^{(r)})}{\partial \boldsymbol{\Theta}} = \left[\frac{\partial Q(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{(r)})}{\partial \boldsymbol{\Theta}}\right]\Bigg|_{\boldsymbol{\Theta}=\boldsymbol{\Theta}^{(r)}} - \left[\frac{\partial H(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{(r)})}{\partial \boldsymbol{\Theta}}\right]\Bigg|_{\boldsymbol{\Theta}=\boldsymbol{\Theta}^{(r)}} = 0,$$

which establishes a contradiction. The inequality (S31) is thus proved. By (S29), the parameter sequence $\{\boldsymbol{\Theta}^{(r)}\}$ must have a convergent subsequence, which means that there exists a subset $\mathbb{N}^1 \subset \mathbb{N}$ such that

$$\boldsymbol{\Theta}^{(r)} \to \boldsymbol{\Theta}^{(\infty)} \quad \text{for } r \in \mathbb{N}^1, \tag{S32}$$

where $\boldsymbol{\Theta}^{(\infty)}$ denotes the limit of this subsequence. Consequently, we have

$$\lim_{r \to +\infty} l_{\text{mp}}(\boldsymbol{\Theta}^{(r)}) = l_{\text{mp}}(\boldsymbol{\Theta}^{(\infty)})$$

in view of the inequality (S30) and the continuity of l_{mp}. Now consider $\{\boldsymbol{\Theta}^{(r+1)}\}_{r\in\mathbb{N}^1}$. Again, it follows from (S29) that there exists a subset $\mathbb{N}^2 \subset \mathbb{N}^1$ such that

$$\boldsymbol{\Theta}^{(r+1)} \to \boldsymbol{\Theta}^{(\infty+1)} \quad \text{for } r \in \mathbb{N}^2, \tag{S33}$$

where $\boldsymbol{\Theta}^{(\infty+1)}$ denotes the limit of $\{\boldsymbol{\Theta}^{(r+1)}\}_{r\in\mathbb{N}^2}$. Similarly, we have

$$l_{\text{mp}}(\boldsymbol{\Theta}^{(\infty+1)}) = \lim_{r\to+\infty} l_{\text{mp}}(\boldsymbol{\Theta}^{(r+1)}) = \lim_{r\to+\infty} l_{\text{mp}}(\boldsymbol{\Theta}^{(r)}) = l_{\text{mp}}(\boldsymbol{\Theta}^{(\infty)}). \tag{S34}$$

To prove $\boldsymbol{\Theta}^{(\infty)} \in \Gamma$, we first assume $\boldsymbol{\Theta}^{(\infty)} \notin \Gamma$ then establish a contradiction in what follows. Note that $Q(\boldsymbol{\Theta}', \boldsymbol{\Theta})$ is continuous with respect to both $\boldsymbol{\Theta}'$ and $\boldsymbol{\Theta}$, which indicates that M is a closed operator. According to (S32) and (S33), we have

$$\boldsymbol{\Theta}^{(\infty+1)} \in M(\boldsymbol{\Theta}^{(\infty)}).$$

As a result, $\boldsymbol{\Theta}^{(\infty)} \notin \Gamma$ implies

$$l_{\text{mp}}(\boldsymbol{\Theta}^{(\infty+1)}) > l_{\text{mp}}(\boldsymbol{\Theta}^{(\infty)})$$

in view of (S31), which is contradicted with (S34). This proves $\boldsymbol{\Theta}^{(\infty)} \in \Gamma$.

According to the assumption of Theorem 2, l_{mp} takes different values at different points in Γ, which ensures that $\{\boldsymbol{\Theta}^{(r)}\}$ has a unique limit point due to (S30). In summary, $\{\boldsymbol{\Theta}^{(r)}\}$ converge to a stationary point of l_{mp}. Furthermore, $\{\boldsymbol{\Theta}^{(r)}\}$ would converge to the maximizer of l_{mp} if an appropriate initial value is chosen.

Appendix F Additional simulation studies

In this section, we conduct additional simulation studies on ROB-HAP under different scenarios to complement the simulation results in the main text. The considered scenarios differ slightly from the setting described in Section 3.1 so as to investigate how parameter estimates are affected by sample size, true parameters, complex covariates and the underlying model. All the results are based on 2,500 Monte Carlo simulations.

Appendix F.1 Simulations with varied sample sizes

To examine how the sample size affects the parameter estimates, we increased the sample size to $n_0 = n_1 = 400$ and $n_0 = n_1 = 800$ while keeping other settings the same as those in Section 3.3. The corresponding estimation results are summarized in Table S4. Similar to the results presented in Table 1, the estimates were virtually unbiased and the SEs were close to the SEEs. The SEs decrease to around $2^{-1/2} \approx 70\%$ when the sample size was doubled, agreeing with the $n^{1/2}$-consistent theory provided in Theorem 1.

Appendix F.2 Simulations with varied true parameters

To mimic a real world situation, we considered generating data with the underlying parameters being estimated from the Jerusalem Perinatal Study. Specifically, we set the true parameters as $\eta = 0.037$, $\beta_{g^m} = -0.078$, $\beta_{g^c} = -0.258$, $\beta_{im} = 0.340$, $\beta_X = -0.082$. Here $\beta' s$ were estimates for SNP rs8192678 in the Jerusalem Perinatal Study (refer to Section 4 of the main text), and η was estimated from the corresponding independence analysis (Zhang et al., 2020, Table S13). In this example, the confounding effect of the maternal genotype was weak. With this set of realistic parameters, the estimation results obtained by ROB-HAP still performed well, as shown in Table S5.

Appendix F.3 Simulations with complex covariates

In Section 3 of the main text, we only considered a single environmental risk factor X. We conducted additional simulations by incorporating covariates with different scales. More precisely, we considered a three-dimentional covariate vector $X = (X_1, X_2, X_3)$:

$$
\begin{aligned}
&X_1 = \log(3)\{g^m - E(g^m)\} + e_1, \\
&\tilde{X}_2 = \log(2)\{g^m - E(g^m)\} + e_2, \quad X_2 = \lfloor \tilde{X}_2 \rfloor, \\
&\tilde{X}_3 = \log(1.5)\{g^m - E(g^m)\} + e_3, \quad X_3 = I(\tilde{X}_3 \geqslant 0),
\end{aligned}
$$

where random errors e_1, e_2, and e_3 were independently generated from the standard normal distribution. Here the function $\lfloor \cdot \rfloor$ denotes the floor function, and $I(\cdot)$ denotes the indicator function. Obviously, the covariates X_1, X_2, and X_3 are binary, categorical, and continuous, respectively. As summarized in Table S6, the estimates were virtually unbiased and the confidence intervals had coverage probabilities close to the nominal level.

Appendix F.4 Simulations under varied underlying models

Finally, we examined the robustness of ROB-HAP with respect to mis-specification of the underlying model. Specifically, we generated data under two alternative penetrance models, the probit model with the link function being inverse of a normal distribution function and the complementary log-log (cloglog) model with the link function being $\log[1-\log(1-\cdot]$, while the logistic model was adopted for the penetrenace in ROB-HAP. To ensure compatibility of the estimation results, the standard deviations of the two distributions were set to be the same as that of the logit distribution. We summarize the estimation results in Table S7. The estimation biases were minor when the phenotype prevalence was high but they could be large when the phenotype prevalence was low. These results coincide with existing literature on general case-control studies (Prentice and Pyke, 1979; Cramer, 2007; Xiang and Langholz, 1999).

References

Cramer, J. S. (2007). Robustness of logit analysis: Unobserved heterogeneity and misspecified disturbances. *Oxford Bulletin of Economics and Statistics*, **69** 545–555.

Lin, D., Weinberg, C. R., Feng, R., Hochner, H. and Chen, J. (2013). A multi-locus likelihood method for assessing parent-of-origin effects using case-control mother-child pairs. *Genetic Epidemiology*, **37** 152–162.

Prentice, R. L. and Pyke, R. (1979). Logistic disease incidence models and case-control studies. *Biometrika*, **66** 403–411.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics*, **22** 300–325.

Wu, C. J. (1983). On the convergence properties of the em algorithm. *The Annals of Statistics*, **11** 95–103.

Xiang, A. H. and Langholz, B. (1999). Miscellanea. comparison of case-control to full cohort analyses under model misspecification. *Biometrika*, **86** 221–226.

Zangwill, W. I. (1969). *Nonlinear programming: a unified approach*, vol. 52. Prentice-hall Englewood Cliffs, NJ.

Zhang, H., Mukherjee, B., Arthur, V., Hu, G., Hochner, H. and Chen, J. (2020). An efficient and computationally robust statistical method for analyzing case-control mother-offspring pair genetic association studies. *Annals of Applied Statistics*, **14** 560–584.

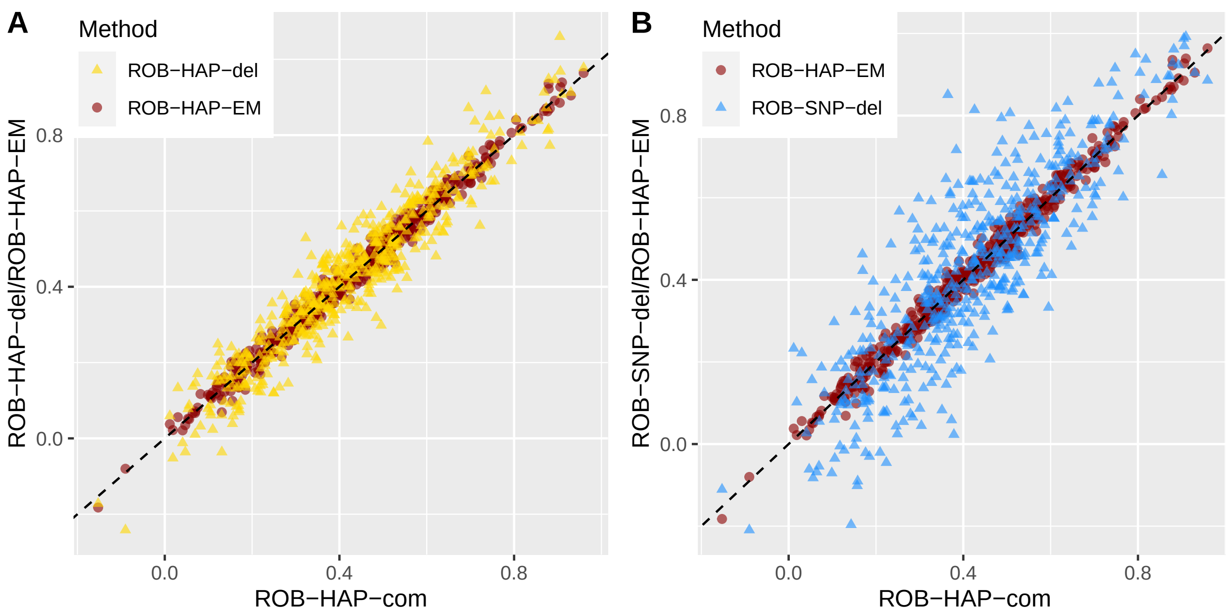

Figure S1: **Panel A**. Estimation results for POE by three versions of ROB-HAP mentioned in Section 3.7 in the main text. Data were generated under $\beta_{im} = \log 1.5$. The other underlying parameters were $f = 0.01$, $\eta = \log 3$, $\beta_{g^m} = \beta_X = \log 1.8$, and $\beta_{g^c} = \log 1.5$. ROB-HAP-com, a version of ROB-HAP utilizing complete genotypes; ROB-HAP-del, a version of ROB-HAP deleting families with incomplete genotypes; ROB-HAP-EM, a version of ROB-HAP handling missing genotypes through EM algorithm. **Panel B**. Estimation results for POE by ROB-HAP-com, ROB-SNP-del, and ROB-HAP-EM. The simulated data is the same as in Panel A. ROB-SNP-del, a version of ROB-SNP deleting families with incomplete genotypes at the target SNP.

Table S1: Simulation results for checking regularity condition (C8) in the main text[a]

		$n^{-1}\partial^2 l_0/\partial\beta_{g^m}\partial\lambda$		$n^{-1}\partial^2 l_0/\partial\beta_{g^c}\partial\lambda$		$n^{-1}\partial^2 l_0/\partial\beta_{im}\partial\lambda$		$n^{-1}\partial^2 l_0/\partial\beta_X\partial\lambda$	
β_{im}	n	EST[b]	SD[c]	EST[b]	SD[c]	EST[b]	SD[c]	EST[b]	SD[c]
0	200	4.19e-03	3.46e-04	4.56e-03	1.96e-04	2.57e-04	1.97e-04	2.50e-03	5.42e-04
	400	4.18e-03	2.45e-04	4.56e-03	1.39e-04	2.54e-04	1.39e-04	2.49e-03	3.89e-04
	800	4.19e-03	1.74e-04	4.56e-03	9.86e-05	2.55e-04	9.87e-05	2.50e-03	2.74e-04
log(1.5)	200	4.35e-03	3.34e-04	4.32e-03	2.14e-04	1.28e-03	2.11e-04	2.39e-03	5.37e-04
	400	4.35e-03	2.35e-04	4.32e-03	1.51e-04	1.28e-03	1.48e-04	2.38e-03	3.79e-04
	800	4.35e-03	1.66e-04	4.32e-03	1.06e-04	1.28e-03	1.04e-04	2.39e-03	2.73e-04

[a]Simulation setup is described in Appendix E; [b]Mean of sample mean based on 10,000 replicates of simulations; [c]Standard error of sample mean based on 10,000 replicates of simulations.

Table S2: Haplotype configurations and frequencies used in simulation studies[a]

Haplotype	SNP1[b]	SNP2[b]	SNP3[b]	SNP4[b]	SNP5[b]	Frequency[c]
H_1	0	0	0	0	1	0.298
H_2	0	0	0	0	0	0.267
H_3	0	0	1	0	1	0.152
H_4	1	1	0	1	0	0.117
H_5	0	0	1	0	0	0.099
H_6	1	0	0	1	0	0.034
H_7	1	1	1	0	0	0.032
MAF[d]	0.183	0.149	0.283	0.151	0.450	

[a]SNP1-SNP5 are located in the gene GPX1 Lin et al. (2013). [b]0 and 1 refer to the common and rare allele, respectively; [c]Haplotype frequency; [d]Minor allele frequency.

Table S3: Sensitivity analysis of ROB-HAP w.r.t. Hardy-Weinberg equilibrium[a]

Fixation index F	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Type-I error rate	0.05	0.052	0.042	0.046	0.064	0.056	0.038	0.03	0.056	0.056

[a]The target SNP was SNP1 in gene GPX1 shown in Table S2.

Table S4: Estimation results of ROB-HAP with varied sample sizes[a]

		$n_0 = n_1 = 400$				$n_0 = n_1 = 800$			
Log-OR	True[b]	Bias[c]	SE[d]	SEE[e]	CP[f]	Bias[c]	SE[d]	SEE[e]	CP[f]
β_{g^m}	0.588	-0.003	0.180	0.184	0.950	0.011	0.126	0.129	0.957
β_{g^c}	0.405	-0.003	0.121	0.122	0.958	-0.002	0.086	0.086	0.950
β_{im}	0.405	0.009	0.131	0.132	0.957	-0.008	0.092	0.093	0.948
β_X	0.588	0.009	0.087	0.088	0.954	0.002	0.060	0.062	0.951

[a]The simulation scenario is the same as in Table 1 apart from the sample size; [b]the true value of the log-OR parameter; [c]difference between the mean estimate and true parameter value; [d]empirical standard error; [e]mean estimated standard error; [f]empirical coverage probability of the 95% confidence intervals.

Table S5: Estimation results of ROB-HAP with varied true parameters[a]

Log-OR	True[b]	Bias[c]	SE[d]	SEE[e]	CP[f]
β_{g^m}	-0.078	-0.009	0.238	0.236	0.948
β_{g^c}	-0.258	-0.009	0.201	0.201	0.958
β_{im}	0.340	0.003	0.215	0.213	0.949
β_X	-0.082	-0.003	0.105	0.101	0.944

[a]The simulation scenario is the same as in Table 1 apart from the true parameters; [b]the true value of the log-OR parameter; [c]difference between the mean estimate and true parameter value; [d]empirical standard error; [e]mean estimated standard error; [f]empirical coverage probability of the 95% confidence intervals.

Table S6: Estimation results of ROB-HAP with complex covariates[a]

Log-OR	True[b]	Bias[c]	SE[d]	SEE[e]	CP[f]
β_{g^m}	0.588	0.013	0.303	0.299	0.945
β_{g^c}	0.405	0.006	0.186	0.181	0.946
β_{im}	0.405	0.002	0.196	0.195	0.953
β_{X_1}	0.588	0.015	0.147	0.142	0.948
β_{X_2}	0.588	0.009	0.138	0.136	0.951
β_{X_3}	0.588	0.010	0.292	0.285	0.951

[a]The simulation scenario is the same as in Table 1 apart from the considered covariates; [b]the true value of the log-OR parameter; [c]difference between the mean estimate and true parameter value; [d]empirical standard error; [e]mean estimated standard error; [f]empirical coverage probability of the 95% confidence intervals.

Table S7: Estimation results of ROB-HAP under varied underlying models[a]

		probit model ($f = 0.01$)				probit model ($f = 0.2$)			
Log-OR	True[b]	Bias[c]	SE[d]	SEE[e]	CP[f]	Bias[c]	SE[d]	SEE[e]	CP[f]
β_{g^m}	0.588	0.341	0.306	0.299	0.814	-0.020	0.289	0.285	0.942
β_{g^c}	0.405	0.161	0.182	0.182	0.868	-0.023	0.216	0.214	0.946
β_{im}	0.405	0.146	0.195	0.195	0.890	-0.025	0.240	0.235	0.944
β_X	0.588	0.287	0.150	0.151	0.538	0.010	0.123	0.122	0.950
		cloglog model ($f = 0.01$)				cloglog model ($f = 0.2$)			
Log-OR	True[b]	Bias[c]	SE[d]	SEE[e]	CP[f]	Bias[c]	SE[d]	SEE[e]	CP[f]
β_{g^m}	0.588	-0.159	0.247	0.248	0.898	-0.095	0.275	0.281	0.943
β_{g^c}	0.405	-0.124	0.172	0.172	0.885	-0.043	0.216	0.211	0.938
β_{im}	0.405	-0.114	0.188	0.185	0.899	-0.052	0.233	0.231	0.939
β_X	0.588	-0.165	0.114	0.113	0.668	-0.109	0.116	0.116	0.835

[a]The simulation scenario is the same as in Table 1 apart from the underlying model and the phenotype prevalence; [b]the true value of the log-OR parameter; [c]difference between the mean estimate and true parameter value; [d]empirical standard error; [e]mean estimated standard error; [f]empirical coverage probability of the 95% confidence intervals.