MENDELIAN RANDOMIZATION TEST OF CAUSAL EFFECT
USING HIGH-DIMENSIONAL SUMMARY DATA

1Lu Deng, 2William Wheeler and 3Kai Yu

1School of Statistics and Data Science, Nankai University, Tianjin, P. R. China
2Information Management Services, Rockville, Maryland, USA
3Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, USA

Supplementary Material

This supplementary material contains all the proofs of Theorem 1–4, and Corollary 1–2, and additional simulation results.

S1 Proof of Theorem 1

For a fixed $s \in (0, 1)$, we evaluate the mean and variance of $Q(s)$ under $H_0 : \theta = 0$. Based on the Independent Assumption, $\tilde{\beta}_{Yk} \sim N(\theta \beta_{Xk} + \alpha_k, \sigma^2_{Yk})$, we have $E_{\theta=0}(\tilde{\beta}_{Yk} - \mu)^2 = \sigma^2_{Yk} + \omega^2$. Furthermore, using the fact that the $2p$ variables $(\tilde{\beta}_{Yk})^p_{k=1}$ and $(\beta_{Xk})^p_{k=1}$ are mutually independent given the true values $\beta_{Xk}$, we have

$E_{\theta=0}\{Q(s)\}$
\[
\vartheta = 0 \sum_{k=1}^{p} \left\{ E_{\theta=0} \left( \hat{\beta}_{Yk} - \mu \right) / \left( \omega^2 + \sigma^2_{Yk} \right)^{1/2} \right\} E \left\{ \hat{\beta}_{Xk}/\sigma_{Xk} \left( \hat{\beta}_{Xk}^2/\sigma_{Xk}^2 \geq 2s \log p \right) \right\} = 0
\]

where the expectation \( E_{\theta=0} \) in the last equation is calculated over the summary data from the risk factor GWAS, thus it is independent of the causal effect.

Denote \( \phi(\cdot) \), \( \Phi(\cdot) \) as the density and survival functions of standard normal distribution, and \( \eta_k = \beta_{Xk}/\sigma_{Xk}, \hat{\eta}_k = \hat{\beta}_{Xk}/\sigma_{Xk}, \lambda_s = (2s \log p)^{1/2} \).

Next, we show the asymptotic normality of \( Q(s) \). By the framework of central limiting theorem (CLT) for independent but not identical variables, we need to verify the following Lindeberg’s condition (Durrett, 2004): for any \( \epsilon > 0 \),

\[
\lim_{p \to \infty} s_p^{-2} \sum_{k=1}^{p} E_{\theta=0} \left[ \left\{ \left( \hat{\beta}_{Yk} - \mu \right)^2 / \left( \omega^2 + \sigma^2_{Yk} \right) \right\} \hat{\eta}_k^2 I\left( \hat{\eta}_k^2 \geq \lambda_s^2 \right) \left( \hat{\beta}_{Yk} - \mu \right)^2 / \left( \omega^2 + \sigma^2_{Yk} \right) \geq \epsilon^2 s_p^2 \right] = 0,
\]

(A1)

which is equivalent to

\[
\lim_{p \to \infty} s_p^{-2} \sum_{k=1}^{p} E \left[ \left\{ 2\left| \epsilon_{\beta k} \right| \phi\left( \left| \epsilon_{\beta k} \right| \right) + 2\Phi\left( \left| \epsilon_{\beta k} \right| / \eta_k \right) \right\} \hat{\eta}_k^2 I\left( \hat{\eta}_k^2 \geq \lambda_s^2 \right) \right] = 0
\]

To show (A1), we first assume \( s_p^{-2} \max_k \hat{\eta}_k^2 \to 0 \), which implies \( s_p^{-2} \max_k \hat{\eta}_k^2 \to 0 \)
0, then we will prove \((A1)\) when \(s^{-2} \max_k \eta_k^2 \not\to 0\). Let \(g_k = 2 |\epsilon_s p \hat{\eta}_k| |\phi(|\epsilon_s p \hat{\eta}_k|) + 2\Phi(|\epsilon_s p \hat{\eta}_k|)|\), it is clear that \(g_k \to 0\) uniformly under the condition \(s^{-2} \max_k \hat{\eta}_k^2 \not\to 0\). Hence, we have

\[
E \left\{ \sum_{k=1}^{p} \{ g_k \hat{\eta}_k^2 \hat{\eta}_k^2 \geq \lambda_k^2 \} \right\} \\
= \sum_{k=1}^{p} E \left\{ g_k \hat{\eta}_k^2 \hat{\eta}_k^2 \geq \lambda_k^2 \right\} | s^{-2} \hat{\eta}_k^2 \to 0 \} \Pr(s^{-2} \hat{\eta}_k^2 \to 0) \\
+ \sum_{k=1}^{p} E \left\{ g_k \hat{\eta}_k^2 \hat{\eta}_k^2 \geq \lambda_k^2 \right\} | s^{-2} \hat{\eta}_k^2 \not\to 0 \} \Pr(s^{-2} \hat{\eta}_k^2 \not\to 0) \\
\leq E \left\{ \max_k g_k | s^{-2} \hat{\eta}_k^2 \to 0 \} s_p^2 + \max_k \Pr(s^{-2} \hat{\eta}_k^2 \not\to 0) s_p^2 = o(s_p^2),
\]

which implies the Lindeberg’s condition is satisfied. Therefore, \(s_p^{-1} Q_p(s) \to N(0, 1)\).

Next, we will verify \((A1)\) when \(s^{-2} \max_k \eta_k^2 \not\to 0\), let \(A = \{k : s^{-2} \eta_k^2 \to \rho_k^2 > 0\}\), for \(k \in A\),

\[
\eta_k^2 = \rho_k^2 s_p^2 \\
= \rho_k^2 \sum_{k=1}^{p} E \left\{ \beta_{X_k}^2 / \sigma_{X_k}^2 \hat{\eta}_k^2 \geq 2s \log p \right\} \\
\geq \rho_k^2 \int t^2 I(t^2 \geq 2s \log p) \phi(t) dt \\
\gg 2 \log p
\]

Hence, \(\Pr(\hat{\eta}_k^2 \geq 2s \log p) = 1 - o(p^{-1}) \to 1\). Then, we have

\[
Q(s) = \frac{s_p(A) \sum_{k \in A} \delta_2 \hat{\eta}_k I(\hat{\eta}_k^2 \geq 2s \log p)}{s_p(A) \sqrt{\omega^2 + \sigma_{Y_k}^2}} + \sum_{k \in A} \frac{|\hat{\eta}_k| \delta_2 \hat{\eta}_k I(\hat{\eta}_k^2 \geq 2s \log p)}{s_p(A) \sqrt{\omega^2 + \sigma_{Y_k}^2}}
\]
\[ \sum_{k \in A} \rho_k \hat{\eta}_k / |\hat{\eta}_k| \sqrt{\omega^2 + \sigma^2_{Y_k}} + o_p(1), \]

where \( s_p^2(A) = \sum_{k \notin A} E\{ \hat{\eta}_k^2 I(\hat{\eta}_k^2 \geq 2s \log p) \}. \)

By the same arguments above, we can show \( T_0 \to N(0, 1). \) Together with the fact that \( (\omega^2 + \sigma^2_{Y_k})^{-1/2}(\hat{\beta}_{Y_k} - \mu) \sim N(0, 1), \) it is clear that \( s_p^{-1} Q_p(s) \) is still normal.

Finally, we show that \( \hat{V}^2(s) \) is a consistent estimate of \( V^2(s), \) i.e.,

\[ V^{-2}(s)\{ \hat{V}^2(s) - V^2(s) \} \to 0, \]

where

\[ \hat{V}^2(s) = \sum_{k=1}^{p} \frac{\hat{\beta}_{X_k}^2}{\sigma^2_{X_k}} \frac{I(\hat{\beta}_{X_k}/\sigma_{X_k} \geq 2s \log p)}. \]

Let \( \frac{\hat{\beta}_{X_k}}{\sigma_{X_k}} = 2r_k \log p, \) for \( k = 1, \ldots, p. \) Simple calculations show that

\[ E_{\theta=0}\{ \hat{V}^2(s) \} = V^2(s) \]

\[ = \sum_{k=1}^{p} \{ L_k^{(1)}(s < r_k) + 2^{-1} L_k^{(1)}(s = r_k) + L_k^{(2)} \ p^{-s} + L_k^{(3)} \ p^{-s} \ I(s > r_k) \} \{ 1 + o(1) \}, \]

\[ \text{var}_{\theta=0}\{ \hat{V}^2(s) \} \]

\[ = \sum_{k=1}^{p} \{ L_k^{(3)}(s < r_k) + 2^{-1} L_k^{(3)}(s = r_k) + L_k^{(4)} \ p^{-s} + L_k^{(5)} \ p^{-s} \ I(s > r_k) \} \{ 1 + o(1) \}. \]

where \( L_k^{(1)} = 2r_k \log p + 1, \ L_k^{(2)} = s(\sqrt{s} - \sqrt{T_k})^{-1} \sqrt{\log p / \pi}, \ L_k^{(3)} = 8r_k \log p, \)

\( L_k^{(4)} = 4s^{3/2} \pi^{-1/2} (\log p)^{3/2} \) and \( L_k^{(5)} = 2s^2 (\sqrt{s} - \sqrt{T_k})^{-1} (\log p)^{3/2} \pi^{-1/2}. \)

Note that \( V^2(s) \geq O(p^{-1}) \) and \( \text{var}_{\theta=0}\{ \hat{V}^2(s) \} = L^{(4)} p^{1-s} \{ 1 + o(1) \} + O\{ V^2(s) \log p \} = o\{ V^4(s) \}. \) By Markov inequality, we have \( V^{-2}(s)\{ \hat{V}^2(s) - V^2(s) \} \to 0. \)
Therefore, $\hat{V}^{-1}(s)Q(s) \rightarrow N(0,1)$.

S2 Proof of Theorem 2

Denote

$$T_{p,1} = \max_{s \in [s_a, s_b]} V^{-1}(s)Q(s) = \max_{s \in [s_a, s_b]} V^{-1}(s) \sum_{k=1}^{p} q_k(s) = \max_{s \in [s_a, s_b]} F(s),$$

with $F(s) = \sum_{k=1}^{p} V^{-1}(s)\sigma_k(s)f_k(s)$, $\sigma_k^2(s) = \text{var}_{\theta=0}\{q_k(s)\}$. Using the same arguments for the proof of Theorem 1 in Zhong, Chen and Xu (2013), we can show that $F(s)$ is joint asymptotic normal at any finite points $(s_1, \ldots, s_d)^T$. To verify the stochastic convergence of the process $F(s)$, we want to show the tightness of the process $F(s)$. Based on the finite dimensional convergence of $F(s)$ and Theorem 1.5.6 in Van der Vaart and Wellner (1996), we only need to show that for any $\varepsilon > 0$ and $\xi > 0$ there exists a finite partition $\Lambda := [s_a, s_b] = \cup_{i=1}^{L} \Lambda_i$ such that

$$\limsup_{p \to \infty} P^*\left\{ \max_{1 \leq i \leq L} \sup_{s \in \Lambda_i} |F(s) - F(t)| > \varepsilon \right\} < \xi, \quad (A2)$$

where $P^*$ is the outer probability measure.

Define $\rho^2(f(s) - f(t)) = \sup_{\theta} E_{\theta=0}\{(f_k(s) - f_k(t))^2\}$. Let $\mathcal{F} = \{f(s) : s \in \Lambda\}$, and $N_0 = N(\varepsilon, \mathcal{F}, \rho)$ be the bracketing number, the smallest number of functions $f(s_1), \ldots, f(s_{N_0})$ in $\mathcal{F}$ such that for each $f \in \mathcal{F}$ there exists an
To show (A2), similar to the results of Theorem 2.2 and Corollary 2.3 in Andrews and Pollard (1994), using the fact that the $2p$ variables $(\hat{\beta}_{Yk})_{k=1}^p$ and $(\hat{\beta}_{Xk})_{k=1}^p$ are mutually independent given the true values $\beta_{Xk}$, we need only to verify the following condition: for some even integer $Q \geq 2$ and some $\gamma > 0$,
\[
\int_0^1 \varepsilon^{-\gamma/(2+\gamma)} N(\varepsilon, \mathcal{F}, \rho)^{1/Q} d\varepsilon < \infty.
\] (A3)

It can be shown that if $s < t$,
\[
\rho^2(f(s) - f(t)) = \sup_k \{2 - 2\sigma_{s}^{-1}(s)\sigma_{k}(t)\}\{1 + o(1)\}.
\]

Noting that $\sigma^2_k(s)$ is uniformly twice continuously differentiable as a function of $s$. By using the similar argument as the one in the proof of the consistency of $V^2(s)$, we have $\rho^2(f(s) - f(t)) \leq C|s - t|$ for a universal constant $C$.

Thus, for $p$ large enough and any $0 < \varepsilon^2 < 1$, $\Lambda$ can be partitioned into finitely many set $\Lambda_1, \ldots, \Lambda_L$ satisfying
\[
\max_{1 \leq i \leq L} \sup_{s,t \in \Lambda_i} \rho^2(f(s) - f(t)) < \varepsilon^2.
\]

where $L \leq C\varepsilon^{-2}$. It is clear that $N(\varepsilon, \mathcal{F}, \rho) \leq L \leq C\varepsilon^{-2}$. Thus, (A3) can be verified if $Q > 2 + \gamma$.

Hence, $F(s)$ converge to a zero mean Gaussian process $\mathcal{N}(s)$ with
\[
\text{Cov}\{\mathcal{N}(s), \mathcal{N}(t)\} = \exp[-|\log\{V(s)\} - \log\{V(t)\}|].
\]

Then it can be shown
S3. PROOF OF COROLLARY 1

that there exists an Ornstein-Uhlenbeck (O-U) process \( U(\cdot) \) with mean zero and \( E\{U(u)U(v)\} = \exp\{-|u - v|\} \) such that \( \mathcal{N}(s) = U(\log\{V(s)\}). \)

Let \( T_{p,2} = \min_{s \in [s_a, s_b]} V^{-1}(s)Q(s) \), clearly \( T_{p,2} = -\max_{s \in [s_a, s_b]} -V^{-1}(s)Q(s) \), and hence \(-T_{p,2}\) has the same asymptotic behaviours as \( T_{p,1} \). By the asymptotic distribution results for the O-U process in Leadbetter et al. [(1983), Theorem 11.1.5 and Theorem 12.2.9], for each finite \( \tau_p > 0 \), we can approximate the tail distribution of our maximal test statistics as

\[
\lim_{x \to +\infty} \frac{1}{x\phi(x)} \Pr\{T > x\} = \lim_{x \to +\infty} \frac{1}{x\phi(x)} \Pr\{ \max_{u \in (0, \tau_p)} |U(u)| > x\} \to 2\tau_p,
\]

where \( \tau_p = 2^{-1}\log\{V^2(s_a)/V^2(s_b)\} \). Noting that \( \tau = 2^{-1}\log\{\hat{V}^2(s_a)/\hat{V}^2(s_b)\} \) is a consistent estimate of \( \tau_p \), we therefore have proved the result.

S3  Proof of Corollary 1

Denote \( \hat{Q}(s) = \sum_{k=1}^{p} \{(\hat{\beta}_Y - \hat{\mu})/\sqrt{\sigma_X^2\sigma_Y^2}\} \hat{\beta}_X/k/\sigma_X I(\hat{\beta}_X^2/\sigma_X^2 \geq 2s\log p) \), then

\[
\hat{Q}(s) = \sum_{k=1}^{p} \left\{ \frac{1}{(\omega^2 + \sigma_Y^2)^{1/2}} + \frac{\omega^2 - \hat{\omega}^2}{(\omega^2 + \sigma_Y^2)^{1/2}(\hat{\omega}^2 + \sigma_Y^2)^{1/2}} \right\} \hat{\beta}_X/k/\sigma_X I(\hat{\beta}_X^2/\sigma_X^2 \geq 2s\log p) \]

(\( \hat{\beta}_Y - \mu + \mu - \hat{\mu} \))
\[= \sum_{k=1}^{p} \left\{ \frac{\hat{\beta}_{Yk} - \mu}{(\omega^2 + \sigma_{Yk}^2)^{1/2}} \right\} \hat{\beta}_{Xk}/\sigma_{Xk}I(\hat{\beta}_{Xk}^2/\sigma_{Xk}^2 \geq 2s \log p) + (\omega^2 - \omega_0^2) \sum_{k=1}^{p} \frac{\hat{\beta}_{Yk} - \mu}{2(\omega^2 + \sigma_{Yk}^2)^{3/2}} \hat{\beta}_{Xk}/\sigma_{Xk}I(\hat{\beta}_{Xk}^2/\sigma_{Xk}^2 \geq 2s \log p) \{1 + o_p(1)\} + (\mu - \hat{\mu}) (\omega^2 - \omega_0^2) \sum_{k=1}^{p} \frac{1}{2(\omega^2 + \sigma_{Yk}^2)^{3/2}} \hat{\beta}_{Xk}/\sigma_{Xk}I(\hat{\beta}_{Xk}^2/\sigma_{Xk}^2 \geq 2s \log p) \{1 + o_p(1)\} \]
\[\equiv Q(s) + R_1 \{1 + o_p(1)\} + R_2 + R_3 \{1 + o_p(1)\}.\]

We evaluate \(R_1\) first. By using the similar argument as the one in the proof of Theorem 1, we have
\[E_{\theta=0} \left[ \sum_{k=1}^{p} \frac{\hat{\beta}_{Yk} - \mu}{2(\omega^2 + \sigma_{Yk}^2)^{3/2}} \hat{\beta}_{Xk}/\sigma_{Xk}I(\hat{\beta}_{Xk}^2/\sigma_{Xk}^2 \geq 2s \log p) \right] = 0,\]
\[\text{var}_{\theta=0} \left[ \sum_{k=1}^{p} \frac{\hat{\beta}_{Yk} - \mu}{2(\omega^2 + \sigma_{Yk}^2)^{3/2}} \hat{\beta}_{Xk}/\sigma_{Xk}I(\hat{\beta}_{Xk}^2/\sigma_{Xk}^2 \geq 2s \log p) \right] \leq 4\omega^{-4}V^2(s).\]

Hence, by the condition \((\omega^2 - \omega_0^2) = o_p(\omega^2), R_1 \leq O_p((\omega^2 - \omega_0^2)\cdot 2\omega^{-2}V(s) = o_p(V(s)).\]

On the other hand, noting that \(E\{\hat{\beta}_{Xk}/\sigma_{Xk}I(\hat{\beta}_{Xk}^2/\sigma_{Xk}^2 \geq 2s \log p)\} = 0\)
if \(\beta_{Xk} = 0\), by Cauchy inequality we have
\[E^2 \left[ \sum_{k=1}^{p} \frac{1}{(\omega^2 + \sigma_{Yk}^2)^{1/2}} \hat{\beta}_{Xk}/\sigma_{Xk}I(\hat{\beta}_{Xk}^2/\sigma_{Xk}^2 \geq 2s \log p) \right] \]
\[= E^2 \sum_{k \in M} \frac{1}{(\omega^2 + \sigma_{Yk}^2)^{1/2}} \hat{\beta}_{Xk}/\sigma_{Xk}I(\hat{\beta}_{Xk}^2/\sigma_{Xk}^2 \geq 2s \log p) \]
\[\leq m\omega^{-2} E \left[ \sum_{k \in M} \hat{\beta}_{Xk}^2/\sigma_{Xk}^2I(\hat{\beta}_{Xk}^2/\sigma_{Xk}^2 \geq 2s \log p) \right] \]
S4. PROOF OF THEOREM 3

For simplicity of calculations, we assume $\mu$, $\omega$, and $V(\cdot)$ are known in the following discussions.

(i) Let $\alpha(p)$ be a sequence of the probabilities of type I error, $x_{\alpha(p)}$ be the corresponding critical values which satisfy $2x_{\alpha(p)}\phi(x_{\alpha(p)})\tau = \alpha(p)$. The power function of our test with nominal size $\alpha(p)$ is approximately

$$\text{Power}(\alpha(p)) = \text{pr}(\sup_{s \in S} V(s)^{-1}|Q(s)| \geq x_{\alpha(p)} \mid H_1)$$

$$= \text{pr}\left(\bigcup_{s \in S} |V_\theta(s)^{-1}\{Q(s) - E_\theta(s)\} + V_\theta(s)^{-1}E_\theta(s)| \geq x_{\alpha(p)}V_\theta(s)^{-1}V(s) \mid H_1\right)$$

Hence, by the condition $(\hat{\mu} - \mu) = o_p(m^{-1/2} \omega)$, $R_2 \leq O_p(|\hat{\mu} - \mu|m^{1/2} \omega^{-1}V(s) = o_p(V(s))$.

Similarly, we can show that $R_3 = o_p(V(s))$. Since we have shown $\hat{V}^2(s)$ is a consistent estimate of $V^2(s)$ uniformly in the proof of Theorem 1, we know $\hat{Q}(s)/\hat{V}(s) - Q(s)/V(s) = o_p(1)$. Therefore, it is clear that $T_{GW}$ has the same asymptotic distribution as $T$ when plugging in those consistent estimates.

S4 Proof of Theorem 3

For simplicity of calculations, we assume $\mu$, $\omega$, and $V(\cdot)$ are known in the following discussions.

(i) Let $\alpha(p)$ be a sequence of the probabilities of type I error, $x_{\alpha(p)}$ be the corresponding critical values which satisfy $2x_{\alpha(p)}\phi(x_{\alpha(p)})\tau = \alpha(p)$. The power function of our test with nominal size $\alpha(p)$ is approximately

$$\text{Power}(\alpha(p)) = \text{pr}(\sup_{s \in S} V(s)^{-1}|Q(s)| \geq x_{\alpha(p)} \mid H_1)$$

$$= \text{pr}\left(\bigcup_{s \in S} |V_\theta(s)^{-1}\{Q(s) - E_\theta(s)\} + V_\theta(s)^{-1}E_\theta(s)| \geq x_{\alpha(p)}V_\theta(s)^{-1}V(s) \mid H_1\right)$$

$$\leq m\omega^{-2}V^2(s), \quad \text{var} \left[ \sum_{k=1}^{p} \frac{1}{(\omega^2 + \sigma^2_{Y_k})^{1/2}} \frac{\hat{\beta}_{X_k}}{\sigma_{X_k}}I(\hat{\beta}_{X_k}^2/\sigma_{X_k}^2 \geq 2s \log p) \right] \leq \omega^{-2}V^2(s).$$
Because $x_{\alpha(p)}$ is slowly varying, $0 < V_\theta(s)^{-1}V(s) \leq 1$, and $V_\theta(s)^{-1}\{Q(s) - E_\theta(s)\}$ is asymptotic normal, a necessary and sufficient condition to ensure $\text{Power}(\alpha(p)) \to 1$ is that there exists a $s$ such that

$$\triangle(s; \theta, r, \kappa) = V_\theta(s)^{-2}E_\theta^2(s) \to +\infty.$$  \hspace{1cm} (A4)

By the fact that $\bar{\Phi}(\sqrt{2t \log p})$ and $\phi(\sqrt{2t \log p})$ are both at the order of $p^{-t}$, we have, up to a factor $1 + o(1)$,

$$\triangle(s; \theta, r, \kappa) = \begin{cases} 
C_1\theta^2 p^{1+s-2\kappa}, & \text{if } s \leq r \text{ and } s \leq \kappa; \\
C_2\theta^2 p^{1-\kappa}, & \text{if } s \leq r \text{ and } s > \kappa; \\
C_3\theta^2 p^{1-2\kappa+s-2(\sqrt{s} - \sqrt{r})^2}, & \text{if } s > r \text{ and } s \leq (\sqrt{s} - \sqrt{r})^2 + \kappa; \\
C_4\theta^2 p^{1-\kappa-(\sqrt{s} - \sqrt{r})^2}, & \text{if } s > r \text{ and } s > (\sqrt{s} - \sqrt{r})^2 + \kappa;
\end{cases}$$

where $C_1, C_2, C_3, C_4$ are between the order of $p^{-\delta}$ and $p^{\delta}$, for any $\delta > 0$ that does not influence the order of $\triangle(s; \theta, r, \kappa)$.

Note that for $\kappa \in [0, 1/2]$, (A4) is easily satisfied. Then, here we only consider the case $\kappa \in (1/2, 1)$, and define

$$\rho^*(\kappa) = \begin{cases} 
\kappa - 1/2, & 1/2 < \kappa \leq 3/4; \\
(1 - \sqrt{1 - \kappa})^2, & 3/4 < \kappa < 1.
\end{cases}$$

Next, we show that there exists a $s_0$ that make sure $\triangle(s_0; \theta, r, \kappa) \to +\infty$ for $r > \rho^*(\kappa)$.

Case 1: $s \leq r$ and $s \leq \kappa$. To make $\triangle(s; \theta, r, \kappa) = C_1\theta^2 p^{1+s-2\kappa} \to +\infty$, 

it need $s > 2\kappa - 1$. we can select $s = \min\{r, \kappa\}$, and arrive at the best divergence rate for $\Delta(s; \theta, r, \kappa)$ of order $C_1 p^{1+\min\{r, \kappa\} - 2\kappa}$;

Case 2: $s \leq r$ and $s > \kappa$. Similarly, the best divergence rate of $\Delta(s; \theta, r, \kappa)$ is of order $C_2 p^{1-\kappa}$ for any $\kappa < s \leq r$;

Case 3: $s > r$ and $s \leq (\sqrt{s} - \sqrt{r})^2 + \kappa$. To make $\Delta(s; \theta, r, \kappa) = C_3 \theta^2 p^{1-2s+s-2(\sqrt{s} - \sqrt{r})^2} \to +\infty$, it needs

$$2\sqrt{r} - \sqrt{1 - 2\kappa + 2r} < \sqrt{s} < 2\sqrt{r} + \sqrt{1 - 2\kappa + 2r}.$$ 

Combine the restrictions above, it must satisfy

$$1 - 2\kappa + 2r > 0, \sqrt{r} < (r + \kappa)/(2\sqrt{r}), 2\sqrt{r} - \sqrt{1 - 2\kappa + 2r} \leq (r + \kappa)/(2\sqrt{r})$$

This translates to

$$\kappa - 1/2 < r < \kappa, \ r \leq \kappa/3$$

or

$$\kappa - 1/2 < r < \kappa, \ r > \kappa/3, \ r \geq (1 - \sqrt{1 - \kappa})^2;$$

Case 4: $s > r$ and $s > (\sqrt{s} - \sqrt{r})^2 + \kappa$. Similarly, it must satisfy $r > (1 - \sqrt{1 - \kappa})^2$.

In summary of Case 1–4, the union of the restrictions is $r > \rho^*(\kappa)$.

Since the nominal size $\alpha(p) = O\{(\log \log p)^{1/2}(\log p)^{-1}\} \to 0$ and the corresponding critical value $x_{\alpha(p)}$ satisfies $2x_{\alpha(p)}\phi(x_{\alpha(p)})\tau = \alpha(p)$, we have
\[
x_{\alpha(p)} = O\{(\log \log p)^{1/2}\}, \text{ which is at a lower order of } \sqrt{\Delta(s_0; \theta, r, \kappa)}.
\]

Thus,

\[
\Pr(V(s_0)^{-1}|Q(s_0)| \geq x_{\alpha} | H_1) \to 1.
\]

Combing the fact that \( T_{GW} \geq V(s)^{-1}|Q(s)| \) for any \( s \in (0, 1) \), we complete the proof of part (i).

(ii) Rewriting \( T_{GW} \) as

\[
T_{GW} = \max_{s \in S} \left| \frac{Q(s) - E_\theta(s)V_\theta(s)}{V_\theta(s)} + \frac{E_\theta(s)}{V(s)} \right|.
\]

If \( r < \rho^*(\kappa) \), then \( r < \kappa \) and \( r < (r+\kappa)^2/(4r) \). Hence, \( V(s)^{-1}V_\theta(s) = 1+o(1) \) if \( s \leq r; 1+o(1) \) if \( r < s < (r+\kappa)^2/(4r) \). Also note that \( (r+\kappa)^2/(4r) > 1 \) for \( r < \rho^*(\kappa) \). Therefore, for all \( s \in \mathcal{S}, V(s)^{-1}V_\theta(s) = 1 + o(1) \).

Using the similar arguments in the proof of part (i), we can show that \( \max_{s \in S} E_\theta(s)V(s)^{-1} \to 0 \) if \( r < \rho^*(\kappa) \). Then, together with the result \( V(s)^{-1}V_\theta(s) = 1+o(1) \) shown above, we have \( T_{GW} = \max_{s \in S} V_\theta(s)^{-1}|Q(s) - E_\theta(s)|\{1 + o(1)\} \).

By employing the same argument of the proof of Theorem 2, the result of Theorem 2 still holds for \( \max_{s \in S} V_\theta(s)^{-1}|Q(s) - E_\theta(s)| \). Hence, under \( H_1 \), when the nominal size \( \alpha(p) \to 0 \), we have \( \Pr(T_{GW} > x_{\alpha(p)}) \to 0 \), which completes the proof of part (ii).
S5  Proof of Corollary 2

Since \( G_{l} \subseteq \mathcal{M} \) or \( G_{l} \subseteq \overline{\mathcal{M}} \), for \( 1 \leq l \leq L \), we define \( \mu_{l} = \mu_{1}, \omega_{l}^{2} = \omega_{1}^{2} \), if \( G_{l} \subseteq \mathcal{M} \), and \( \mu_{l} = \mu_{2}, \omega_{l}^{2} = \omega_{2}^{2} \), if \( G_{l} \subseteq \overline{\mathcal{M}} \). Then, let

\[
Q^{\ast}_{L}(s) = \sum_{l=1}^{L} \sum_{k \in G_{l}} \frac{(\hat{\beta}_{Yk} - \mu_{l}) \hat{\beta}_{Xk}}{(\omega_{l}^{2} + \sigma_{Yk}^{2})^{1/2}} \frac{1}{\sigma_{Xk}} I\left(\frac{\hat{\beta}_{Xk}^{2}}{\sigma_{Xk}^{2}} \geq 2s \log p\right),
\]

it is clear that \( Q^{\ast}_{L}(s) = Q^{\ast}(s) \). By using the same argument as the one in the proof of Theorem 1 and Theorem 2, we can show that the results of Theorem 1 and Theorem 2 still hold for \( Q^{\ast}_{L}(s) \).

Next, we will show that \( \hat{Q}^{\ast}(s) \) and \( Q^{\ast}(s) \) are asymptotically equivalent. By using the similar argument as the one in the proof of Corollary 1, we have

\[
\hat{Q}^{\ast}(s) - Q^{\ast}(s)
\]

\[
= \sum_{l=1}^{L} (\omega_{l}^{2} - \hat{\omega}_{l}^{2}) \sum_{k \in G_{l}} \frac{\hat{\beta}_{Yk} - \mu_{l}}{2(\omega_{l}^{2} + \hat{\omega}_{l}^{2})^{3/2}} \frac{1}{\sigma_{Xk}} I\left(\frac{\hat{\beta}_{Xk}^{2}}{\sigma_{Xk}^{2}} \geq 2s \log p\right)\{1 + o_{p}(1)\}
\]

\[
+ \sum_{l=1}^{L} (\mu_{l} - \hat{\mu}_{l}) \sum_{k \in G_{l}} \frac{1}{(\omega_{l}^{2} + \hat{\omega}_{l}^{2})^{1/2}} \frac{1}{\sigma_{Xk}} I\left(\frac{\hat{\beta}_{Xk}^{2}}{\sigma_{Xk}^{2}} \geq 2s \log p\right)
\]

\[
+ \sum_{l=1}^{L} (\mu_{l} - \hat{\mu}_{l})(\omega_{l}^{2} - \hat{\omega}_{l}^{2}) \sum_{k \in G_{l}} \frac{1}{2(\omega_{l}^{2} + \hat{\omega}_{l}^{2})^{3/2}} \frac{1}{\sigma_{Xk}} I\left(\frac{\hat{\beta}_{Xk}^{2}}{\sigma_{Xk}^{2}} \geq 2s \log p\right)\{1 + o_{p}(1)\}
\]

\[
\equiv R_{1}\{1 + o_{p}(1)\} + R_{2} + R_{3}\{1 + o_{p}(1)\}.
\]

Denote

\[
\sigma_{l}^{2}(s) = \text{var}_{\theta=0} \left\{ \sum_{k \in G_{l}} \frac{(\hat{\beta}_{Yk} - \mu_{l}) \hat{\beta}_{Xk}}{(\omega_{l}^{2} + \hat{\omega}_{l}^{2})^{1/2}} \frac{1}{\sigma_{Xk}} I\left(\frac{\hat{\beta}_{Xk}^{2}}{\sigma_{Xk}^{2}} \geq 2s \log p\right) \right\},
\]

\[ \sigma_0^2(s) = \sum_{l=1}^{L} \sigma_l^2(s). \]

We evaluate \( R_2 \) first. By the assumption that \( \beta_{Xk}/\sigma_{Xk} \) follow a symmetrical distribution around 0, and the way how \( G_l \) constructed, we have \( \beta_{Xk}/\sigma_{Xk} \) are still symmetrical around 0, for \( k \in G_l \). Noting that

\[
E \left[ \hat{\beta}_{Xk}/\sigma_{Xk} I(\hat{\beta}_{Xk}^2/\sigma_{Xk}^2 \geq 2s \log p) \right]
\]

is an odd function of \( \beta_{Xk}/\sigma_{Xk} \), then

\[
E \hat{\theta}^2 = 0 \sum_{k \in G_l} \frac{1}{(\omega_l^2 + \sigma_{Yk}^2)^{1/2}} \hat{\beta}_{Xk}/\sigma_{Xk} I(\hat{\beta}_{Xk}^2/\sigma_{Xk}^2 \geq 2s \log p)
\]

\[
\var \left[ \sum_{k \in G_l} \frac{1}{(\omega_l^2 + \sigma_{Yk}^2)^{1/2}} \hat{\beta}_{Xk}/\sigma_{Xk} I(\hat{\beta}_{Xk}^2/\sigma_{Xk}^2 \geq 2s \log p) \right] \leq \omega_l^{-2} \sigma_l^2(s).
\]

Hence, by the condition \( |G_l| \to \infty \) uniformly for \( l \), we have \( (\mu_l - \hat{\mu}_l) = o_p(\omega_l) \) uniformly. Then, \( R_2 \leq O_p \left\{ (\sum_{l=1}^{L} |\mu_l - \hat{\mu}_l|^2 \omega_l^{-2} \sigma_l^2(s))^{1/2} \right\} = o_p(\sigma_0(s)). \)

Similarly, by the fact that \( (\omega_l^2 - \bar{\omega}_l^2) = o_p(\omega_l^2) \) uniformly, we can show that \( R_1 = R_3 = o_p(\sigma_0(s)) \). Then, it is straightforward to show that \( T_{\text{cond}} \) has the same asymptotic distribution when plugging in those consistent estimates of \( \mu_l \) and \( \omega_l \).

**S6 Proof of Theorem 4**

Rewriting \( Q(s) = \sum_{k=1}^{p} q_k(s) \). We will first prove that the \( \rho \)-mixing of \( \{Z_k\}_{k=1}^{p} \) leads to the \( \rho \)-mixing of \( \{q_k(s)\}_{k=1}^{p} \) for any \( s \in S \).

Define \( w_Y = \{w_{Yk} = (\hat{\beta}_{Yk} - \mu)/\sqrt{\omega^2 + \sigma_{Yk}^2}\}_{k=1}^{p} \) and \( w_X(s) = \{w_{Xk}(s) = \hat{\beta}_{Xk}/\sigma_{Xk} I(\hat{\beta}_{Xk}^2/\sigma_{Xk}^2 \geq 2s \log p)\}_{k=1}^{p} \). By the fact that \( \hat{\beta}_{Yk}, \hat{\beta}_{Xk}, k = 1, \ldots, p, \)
are mutually independent given $\beta X_k$, simple calculations show that
\[
|\text{Corr}(w Y_k w X_k(s), w Y_l w X_l(s))| = |\text{Corr}(w Y_k, w Y_l)\text{Corr}(w X_k(s), w X_l(s))| 
\leq |\text{Corr}(w Y_k, w Y_l)|.
\]
Therefore,
\[
\rho_{q(s)}(t) \leq \rho_{wY}(t).
\]
Since the fact that $\hat{\beta}_{Y_k}$ is the least-squared estimate of regression for $Z_k$ on $Y$, we have
\[
\text{Corr}(\hat{\beta}_{Y_k}, \hat{\beta}_{Y_l}) = C_0 \text{Corr}(Z_k, Z_l)/\sqrt{\text{var}(Z_k)\text{var}(Z_l)},
\]
for some constant value $C_0$. Then,
\[
\rho_{\hat{\beta}_Y}(t) \leq C_1 \rho_Z(t),
\]
for some constant value $C_1$. Also note that $\rho_{wY}(t) = \rho_{\hat{\beta}_Y}(t)$. Therefore,
\[
\rho_{q(s)}(t) \leq \rho_{\hat{\beta}_Y}(t) \leq C_1 \rho_Z(t).
\]
That is, the assumption of $\{q_k(s)\}_{k=1}^p$ being $\rho$-mixing for each $s$ is weaker than the requirement of the SNP genotype data $\{Z_k\}_{k=1}^p$ to be $\rho$-mixing.

Next, we establish the asymptotic normality of $V^{-1}(s)Q(s)$ under the assumption that $\{q_k(s)\}_{k=1}^p$ is $\rho$-mixing for each $s$. 
By the framework of central limiting theorem for dependent variables, we use the Bernstein’s blocking method (Leadbetter et., al , 1983), and divide the sequence \( \{q_k(s)\}_{k=1}^p \) evenly into \( R \) groups, so that each group consists of \( b \) elements. For each group, we denote the first \( a \) elements as the large block and the remaining \( a' \) elements as the small block. Therefore, for \( j = 1, \ldots, R \), the large blocks are

\[
Q_j(s) = \sum_{i=1}^{a} q_{(j-1)b+i}(s),
\]

the small blocks are

\[
Q'_j(s) = \sum_{i=1}^{a'} q_{(j-1)b+a'+i}(s)
\]

and the residual block is

\[
\delta_p(s) = \sum_{i=Rb+1}^{p} q_i(s).
\]

Then,

\[
V^{-1}(s)Q(s) = V^{-1}(s) \sum_{j=1}^{R} Q_j(s) + V^{-1}(s) \sum_{j=1}^{R} Q'_j(s) + V^{-1}(s)\delta_p(s),
\]

where clearly the expectation of each block is 0.

Denote \( \sigma_k^2(s) = \text{var}_{\theta=0}\{q_k(s)\} \) and \( \Sigma_j(s) = \text{var}_{\theta=0}\{Q'_j(s)\} \). Then,

\[
\text{var}_{\theta=0}\{\sum_{j=1}^{R} Q'_j(s)\} = \sum_{j=1}^{R} \text{var}_{\theta=0}\{Q'_j(s)\} + \sum_{j\neq t} E_{\theta=0}\{Q'_j(s)Q'_t(s)\}
\]

\[
\leq \sum_{j=1}^{R} \Sigma_j(s) + \sum_{j=1}^{R-1} \sum_{i=j+1}^{R} C \alpha^{(t-1-j)b+a} \{\Sigma_j(s) + \Sigma_i(s)\}
\]
\[ \leq \sum_{j=1}^{R} \sum_{j=1}^{R} \alpha^a + \cdots + \alpha^{(R-2)b+a} \sum_{j=1}^{R} \sum_{j=1}^{R} (s) + 2C \{ \alpha a + \cdots + \alpha (R-2)b + a \} \sum_{j=1}^{R} \sum_{j=1}^{R} (s) \]
\[ \leq O(\sum_{j=1}^{R} \sum_{j=1}^{R} (s)). \]

Clearly noting that if we let \( a' \to +\infty, a'/a \to 0, \) and \( a/p \to 0 \) as \( p \to +\infty, \) that is, the number of elements in the large blocks is much larger than the one in the small blocks. Then

\[ \text{var}_{a=0} \{ \sum_{j=1}^{R} Q_j(s) \} \leq O(\sum_{j=1}^{R} \sum_{j=1}^{R} (s)) = o(V^2(s)). \]

Similarly, we have \( \text{var} \{ V^{-1}(s) \delta_p(s) \} \to 0. \) Hence,

\[ V^{-1}(s)Q(s) = V^{-1}(s) \sum_{j=1}^{R} Q_j(s) + o_p(1). \]

By Bradley’s lemma (Bradley, 1992), let \( R = p^c \) for \( c \in (0,1), b = p^{1-c} \) and \( a' = p^d \) for \( d \in (0,1-c), \) which satisfy the above conditions, then there exist independent random variables \( W_j(s) \) such that \( W_j(s) \) and \( Q_j(s) \) are identically distributed and for any \( \epsilon > 0, \)

\[ \text{pr}(\left| V^{-1}(s) \sum_{j=1}^{R} Q_j(s) - V^{-1}(s) \sum_{j=1}^{R} W_j(s) \right| > \epsilon) \to 0 \]

Therefore, we only need to show the asymptotic normality of \( V^{-1}(s) \sum_{j=1}^{R} W_j(s). \)

To this end, we only need to verify the Lindeberg’s condition. This can be verified using similar arguments as the ones in the proof of Theorem 1.

Finally, we show that the result of Theorem 2 still holds for \( \rho \)-mixing correlated SNPs under \( H_0. \)
Similar to the proof of Theorem 2, we also only need to show (A2). Since the fact that the SNP genotype data \( \{Z_k\}_{k=1}^{p} \) is \( \rho \)-mixing, then the \( 2p \) variables \( (\hat{\beta}_{Yk})_{k=1}^{p} \) and \( (\hat{\beta}_{Xk})_{k=1}^{p} \) are not mutually independent given the true values \( \beta_{Xk} \). Therefore, based on the results of Theorem 2.2 and Corollary 2.3 in Andrews and Pollard (1994), besides (A3), we also need to verify the following condition: for some \( \gamma > 0 \) and an even integer \( Q > 2+\gamma \),

\[
\sum_{d=1}^{\infty} d^{Q-2} \rho_{Z}(d)^{\gamma/(Q+\gamma)} < \infty. \tag{A5}
\]

Noting that \( \rho_{Z_k}(d) \leq C\alpha^d \) for some \( \alpha \in (0, 1) \), we have

\[
\sum_{d=1}^{\infty} d^{Q-2} \rho_{Z}(d)^{\gamma/(Q+\gamma)} \leq C \sum_{d=1}^{\infty} d^{Q-2} \alpha^{d\gamma/(Q+\gamma)}.
\]

Then, by using the d’Alembert’s ratio test,

\[
\lim_{d \to \infty} \frac{(d + 1)^{Q-2} \alpha^{(d+1)\gamma/(Q+\gamma)}}{d^{Q-2} \alpha^{d\gamma/(Q+\gamma)}} = \alpha^{\gamma/(Q+\gamma)} < 1,
\]

we shown that (A5) is satisfied.

References


**Supplementary Table 1**: Simulation results on type I errors under the genome-wide InSIDE assumption with unbalanced pleiotropy. Results are summarized based on performances over 2000 datasets generated from two GWAS of equal sample size ($N$). Each simulated dataset consists of summary statistics on 200,000 independent SNPs.

<table>
<thead>
<tr>
<th>MR method</th>
<th>Sample size $N$ (average number of significant SNPs)$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>300k (26)</td>
</tr>
<tr>
<td>MaxK-1</td>
<td>0.046</td>
</tr>
<tr>
<td>MaxK-2</td>
<td>0.043</td>
</tr>
<tr>
<td>IVW</td>
<td>0.041</td>
</tr>
<tr>
<td>W-Median</td>
<td>0.026</td>
</tr>
<tr>
<td>W-Mode</td>
<td>0.001</td>
</tr>
<tr>
<td>IVW-Robust</td>
<td>0.041</td>
</tr>
<tr>
<td>MR-Egger</td>
<td>0.043</td>
</tr>
<tr>
<td>Con-mix</td>
<td>0.116</td>
</tr>
<tr>
<td>MRMix</td>
<td>0.025</td>
</tr>
</tbody>
</table>

---

a. Each summary data is generated from two GWAS of equal sample size $N$. The average number of significant SNPs is the number of SNPs with their risk factor association p-values less than $5 \times 10^{-8}$, averaged over 2000 simulated datasets.
b. Both MaxK-1 and MaxK-2 use summary statistics on 200,000 independent SNPs. All other tests use summary statistics on SNPs that are genome-wide significantly associated with the risk factor.
Supplementary Table 2: Simulation results on type I errors under the conditional InSIDE assumption with unbalanced pleiotropy. Results are summarized based on performances over 2000 datasets generated from two GWAS of equal sample size (N). Each simulated dataset consists of summary statistics on 200,000 independent SNPs.

<table>
<thead>
<tr>
<th>MR method</th>
<th>Sample size N (average number of significant SNPs)$^a$</th>
<th>300k (26)</th>
<th>350k (41)</th>
<th>400k (59)</th>
<th>450k (80)</th>
<th>500k (104)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MaxK-1</td>
<td></td>
<td>0.610</td>
<td>0.653</td>
<td>0.679</td>
<td>0.706</td>
<td>0.718</td>
</tr>
<tr>
<td>MaxK-2</td>
<td></td>
<td>0.048</td>
<td>0.052</td>
<td>0.052</td>
<td>0.053</td>
<td>0.052</td>
</tr>
<tr>
<td>IVW</td>
<td></td>
<td>0.065</td>
<td>0.052</td>
<td>0.056</td>
<td>0.054</td>
<td>0.048</td>
</tr>
<tr>
<td>W-Median</td>
<td></td>
<td>0.117</td>
<td>0.099</td>
<td>0.101</td>
<td>0.106</td>
<td>0.101</td>
</tr>
<tr>
<td>W-Mode</td>
<td></td>
<td>0.023</td>
<td>0.012</td>
<td>0.010</td>
<td>0.006</td>
<td>0.005</td>
</tr>
<tr>
<td>IVW-Robust</td>
<td></td>
<td>0.059</td>
<td>0.048</td>
<td>0.050</td>
<td>0.047</td>
<td>0.045</td>
</tr>
<tr>
<td>MR-Egger</td>
<td></td>
<td>0.057</td>
<td>0.059</td>
<td>0.060</td>
<td>0.052</td>
<td>0.052</td>
</tr>
<tr>
<td>Con-mix</td>
<td></td>
<td>0.191</td>
<td>0.174</td>
<td>0.166</td>
<td>0.166</td>
<td>0.173</td>
</tr>
<tr>
<td>MRMix</td>
<td></td>
<td>0.092</td>
<td>0.055</td>
<td>0.051</td>
<td>0.041</td>
<td>0.041</td>
</tr>
</tbody>
</table>

a. Each summary data is generated from two GWAS of equal sample size N. The average number of significant SNPs is the number of SNPs with their risk factor association p-values less than $5 \times 10^{-8}$, averaged over 2000 simulated datasets.

b. Both MaxK-1 and MaxK-2 use summary statistics on 200,000 independent SNPs. All other tests use summary statistics on SNPs that are genome-wide significantly associated with the risk factor.
Supplementary Table 3: Simulation results on type I errors when the InSiDE assumption is violated. Results are summarized based on performances over 2000 datasets generated from two GWAS of equal sample size (N). Each simulated dataset consists of summary statistics on 200,000 independent SNPs.

<table>
<thead>
<tr>
<th>% IVs</th>
<th>Sample size N (average number of significant SNPs)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>300k (26)</td>
</tr>
<tr>
<td>5%</td>
<td>0.048</td>
</tr>
<tr>
<td>10%</td>
<td>0.048</td>
</tr>
<tr>
<td>20%</td>
<td>0.092</td>
</tr>
<tr>
<td>30%</td>
<td>0.127</td>
</tr>
</tbody>
</table>

a. Each summary data is generated from two GWAS of equal sample size N. The average number of significant SNPs is the number of SNPs with their risk factor association p-values less than $5 \times 10^{-8}$, averaged over 2000 simulated datasets.
b. Percentage of risk factor associated SNPs that have correlated pleiotropic effect on the outcome.
Supplementary Figure 1. Simulation results on power comparisons under the genome-wide InSIDE assumption with unbalanced pleiotropy. Results are summarized based on performances over 2000 simulated datasets under a given causal effect (theta) and sample size (N). Each simulated dataset consists of summary statistics on 200,000 independent SNPs generated from two GWAS of equal sample size N. (a) Both GWAS have sample size \( N = 300k \), with an average number of 26 SNPs that are genome-wide significantly associated with the risk factor; (b) Sample size \( N = 400k \), and an average number of 59 significant SNPs; (c) Sample size \( N = 450k \), and an average number of 80 significant SNPs; and (d) Sample size \( N = 500k \), and an average number of 104 significant SNPs.
Supplementary Figure 2. Simulation results on power comparisons under the conditional InSIDE Assumption with unbalanced pleiotropy. Results are summarized based on performances over 2000 simulated datasets under a given causal effect (theta) and sample size (N). Each simulated dataset consists of summary statistics on 200,000 independent SNPs generated from two GWAS of equal sample size N. (a) Both GWAS have sample size $N = 300k$, with an average number of 26 SNPs that are genome-wide significantly associated with the risk factor; (b), Sample size $N = 400k$, and an average number of 59 significant SNPs; (c), Sample size $N = 450k$, and an average number of 80 significant SNPs; and (d), Sample size $N = 500k$, and an average number of 104 significant SNPs.
Supplementary Figure 3. Simulation results on power comparisons with K statistics of different thresholds under the conditional InSIDE Assumption with balanced pleiotropy. Results are summarized based on performances over 2000 simulated datasets under a given causal effect (theta) and sample size (N). Each simulated dataset consists of summary statistics on 200,000 independent SNPs generated from two GWAS of equal sample size N. (a) Both GWAS have sample size $N = 300k$, with an average number of 26 SNPs that are genome-wide significantly associated with the risk factor; (b). Sample size $N = 400k$, and an average number of 59 significant SNPs; (c). Sample size $N = 450k$, and an average number of 80 significant SNPs; and (d). Sample size $N = 500k$, and an average number of 104 significant SNPs.
(a) N=300k (26)

(b) N=400k (59)

(c) N=450k (80)

(d) N=500k (104)
Supplementary Figure 4. Simulation results on power comparisons with K statistics of different thresholds under the conditional InSIDE Assumption with unbalanced pleiotropy. Results are summarized based on performances over 2000 simulated datasets under a given causal effect (theta) and sample size ($N$). Each simulated dataset consists of summary statistics on 200,000 independent SNPs generated from two GWAS of equal sample size $N$. (a) Both GWAS have sample size $N = 300k$, with an average number of 26 SNPs that are genome-wide significantly associated with the risk factor; (b) Sample size $N = 400k$, and an average number of 59 significant SNPs; (c) Sample size $N = 450k$, and an average number of 80 significant SNPs; and (d) Sample size $N = 500k$, and an average number of 104 significant SNPs.