Biostatistics (2015), **0**, 0, *pp.* 1–29
doi:10.1093/biostatistics/suppr3

Sparse meta-analysis with high-dimensional data supplementary material

QIANCHUAN HE

Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA

HAO HELEN ZHANG

Department of Mathematics, The University of Arizona, Tucson, AZ 85721, USA

CHRISTY L. AVERY

Department of Epidemiology, University of North Carolina, Chapel Hill, NC 27599, USA

D. Y. LIN*

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

lin@bios.unc.edu

---

*To whom correspondence should be addressed.

SECTION A: PROOFS OF TECHNICAL RESULTS

We assume throughout that for each k, $n_k\tilde{\boldsymbol{V}}_k \to_p \boldsymbol{\Sigma}^{(k)}$ and $n_k/n \to \nu_k$ as $n \to \infty$. Let $\nu_{\min} = \min_{1\leqslant k\leqslant K}\{\nu_k\}$ and $\nu_{\max} = \max_{1\leqslant k\leqslant K}\{\nu_k\}$. Assume that all the nonzero β^0_{jk}'s are bounded by some constants r_1 and r_2.

Proof for the equivalence of equations (2.1) and (2.5)

By the Cauchy-Schwartz inequality,

$$
\begin{aligned}
&\sum_{k=1}^{K}(\tilde{\boldsymbol{\beta}}_k-\boldsymbol{\beta}_k)^{\mathrm{T}}\tilde{\boldsymbol{V}}_k^{-1}(\tilde{\boldsymbol{\beta}}_k-\boldsymbol{\beta}_k)+\lambda_1\sum_{j=1}^{p}\gamma_j+\sum_{j=1}^{p}\gamma_j^{-1}\left(\sum_{k\in\mathcal{S}_j}w_{jk}|\beta_{jk}|\right)\\
\geqslant&\sum_{k=1}^{K}(\tilde{\boldsymbol{\beta}}_k-\boldsymbol{\beta}_k)^{\mathrm{T}}\tilde{\boldsymbol{V}}_k^{-1}(\tilde{\boldsymbol{\beta}}_k-\boldsymbol{\beta}_k)+\sum_{j=1}^{p}\left(2\sqrt{\lambda_1\sum_{k\in\mathcal{S}_j}w_{jk}|\beta_{jk}|}\right),
\end{aligned}
$$

where the equality holds if and only if $\gamma_j = (1/\lambda_1)^{\frac{1}{2}}\left(\sum_{k\in\mathcal{S}_j}w_{jk}|\beta_{jk}|\right)^{\frac{1}{2}}$ for all j. Let $\lambda_1 = (\lambda/2)^2$. Then, the proof is completed.

Proof for the consistency of the proposed MIC criterion

Let $\mathcal{M}$ denote an arbitrary model, $\mathcal{M}_\lambda$ denote the model under λ, and $\mathcal{M}_T$ denote the true model. We say that $\mathcal{M}$ is an under-fitted model if $\mathcal{M} \not\supset \mathcal{M}_T$ and an over-fitted model if $\mathcal{M} \supset \mathcal{M}_T$ and $\mathcal{M} \neq \mathcal{M}_T$. Define

$$
\hat{\boldsymbol{\beta}}_{k,\mathcal{M}} = \operatorname{argmin}_{\{\boldsymbol{\beta}_k\in R^p:\beta_{k,j}=0,\forall j\notin\mathcal{M}\}}(\boldsymbol{\beta}_k-\tilde{\boldsymbol{\beta}}_k)^{\mathrm{T}}(\boldsymbol{\beta}_k-\tilde{\boldsymbol{\beta}}_k). \tag{E.1}
$$

Generally, $\hat{\boldsymbol{\beta}}_{k,\mathcal{M}} \neq \tilde{\boldsymbol{\beta}}_k$ because of the constraint that $\beta_{k,j} = 0, \forall j \notin \mathcal{M}$.

Suppose that λ_0 yields the true model, λ_L yields an under-fitted model, and λ_H yields an over-fitted model. We wish to prove that, with high probability, $MIC_{\lambda_L} > MIC_{\lambda_0}$ and $MIC_{\lambda_H} > MIC_{\lambda_0}$. We first prove the former.

Because both $\hat{\boldsymbol{\beta}}_{k,\lambda_0}$ and $\tilde{\boldsymbol{\beta}}_k$ are consistent for $\boldsymbol{\beta}^0_k$ and $q_{\lambda_0,k}\log(n_k)/n_k \to 0$, it is easy to verify that $MIC_{\lambda_0} = o_p(1)$. For λ_L,

$$\begin{aligned}
MIC_{\lambda_L} &= \sum_{k=1}^{K} (\hat{\boldsymbol{\beta}}_{k,\lambda_L} - \tilde{\boldsymbol{\beta}}_k)^{\mathrm{T}} (\hat{\boldsymbol{\beta}}_{k,\lambda_L} - \tilde{\boldsymbol{\beta}}_k) + \sum_{k=1}^{K} q_{\lambda_L,k} \times \log(n_k)/n_k \\
&\geqslant \sum_{k=1}^{K} (\hat{\boldsymbol{\beta}}_{k,\lambda_L} - \tilde{\boldsymbol{\beta}}_k)^{\mathrm{T}} (\hat{\boldsymbol{\beta}}_{k,\lambda_L} - \tilde{\boldsymbol{\beta}}_k) \\
&\geqslant \sum_{k=1}^{K} (\hat{\boldsymbol{\beta}}_{k,\mathcal{M}_{\lambda_L}} - \tilde{\boldsymbol{\beta}}_k)^{\mathrm{T}} (\hat{\boldsymbol{\beta}}_{k,\mathcal{M}_{\lambda_L}} - \tilde{\boldsymbol{\beta}}_k) \\
&\geqslant \min_{\mathcal{M} \not\supset \mathcal{M}_T} \sum_{k=1}^{K} (\hat{\boldsymbol{\beta}}_{k,\mathcal{M}} - \tilde{\boldsymbol{\beta}}_k)^{\mathrm{T}} (\hat{\boldsymbol{\beta}}_{k,\mathcal{M}} - \tilde{\boldsymbol{\beta}}_k) \\
&\to \sum_{k=1}^{K} (\boldsymbol{\beta}^0_{k,\mathcal{M}} - \boldsymbol{\beta}^0_k)^{\mathrm{T}} (\boldsymbol{\beta}^0_{k,\mathcal{M}} - \boldsymbol{\beta}^0_k) > 0.
\end{aligned}$$

The first inequality holds trivially. The second inequality follows from the definition of $\hat{\boldsymbol{\beta}}_{k,\mathcal{M}_{\lambda_L}}$; see (E.1). The third inequality holds trivially. The remaining part holds because $\hat{\boldsymbol{\beta}}_{k,\mathcal{M}} \to \boldsymbol{\beta}^0_{k,\mathcal{M}}$, $\tilde{\boldsymbol{\beta}}_k \to \boldsymbol{\beta}^0_k$, and $\mathcal{M}$ represents an underfitted model.

Next, we prove that $MIC_{\lambda_H} > MIC_{\lambda_0}$. Note that

$$\begin{aligned}
n(MIC_{\lambda_H} - MIC_{\lambda_0}) \geqslant\ & \nu_{\max}^{-1} \sum_{k=1}^{K} n_k (\hat{\boldsymbol{\beta}}_{k,\lambda_H} - \tilde{\boldsymbol{\beta}}_k)^{\mathrm{T}} (\hat{\boldsymbol{\beta}}_{k,\lambda_H} - \tilde{\boldsymbol{\beta}}_k) - \\
& \nu_{\min}^{-1} \sum_{k=1}^{K} n_k (\hat{\boldsymbol{\beta}}_{k,\lambda_0} - \tilde{\boldsymbol{\beta}}_k)^{\mathrm{T}} (\hat{\boldsymbol{\beta}}_{k,\lambda_0} - \tilde{\boldsymbol{\beta}}_k) + \\
& \nu_{\max}^{-1} \sum_{k=1}^{K} (q_{\lambda_H,k} - q_{\lambda_0,k}) \log(n_k) \\
\geqslant\ & \nu_{\max}^{-1} \sum_{k=1}^{K} n_k (\hat{\boldsymbol{\beta}}_{k,\mathcal{M}_{\lambda_H}} - \tilde{\boldsymbol{\beta}}_k)^{\mathrm{T}} (\hat{\boldsymbol{\beta}}_{k,\mathcal{M}_{\lambda_H}} - \tilde{\boldsymbol{\beta}}_k) - \\
& \nu_{\min}^{-1} \sum_{k=1}^{K} n_k (\hat{\boldsymbol{\beta}}_{k,\lambda_0} - \tilde{\boldsymbol{\beta}}_k)^{\mathrm{T}} (\hat{\boldsymbol{\beta}}_{k,\lambda_0} - \tilde{\boldsymbol{\beta}}_k) + \\
& \nu_{\max}^{-1} \sum_{k=1}^{K} (q_{\lambda_H,k} - q_{\lambda_0,k}) \log(n_k) \\
\geqslant\ & \nu_{\max}^{-1} \sum_{k=1}^{K} n_k (\hat{\boldsymbol{\beta}}_{k,\mathcal{M}_{\lambda_H}} - \tilde{\boldsymbol{\beta}}_k)^{\mathrm{T}} (\hat{\boldsymbol{\beta}}_{k,\mathcal{M}_{\lambda_H}} - \tilde{\boldsymbol{\beta}}_k) -
\end{aligned}$$

$$
\begin{aligned}
&\nu_{\min}^{-1}\sum_{k=1}^{K} n_k(\hat{\boldsymbol{\beta}}_{k,\lambda_0}-\tilde{\boldsymbol{\beta}}_k)^{\mathrm{T}}(\hat{\boldsymbol{\beta}}_{k,\lambda_0}-\tilde{\boldsymbol{\beta}}_k)+\nu_{\max}^{-1}\sum_{k=1}^{K}\log(n_k)\\
\geqslant\ &\nu_{\max}^{-1}\inf_{\mathcal{M}\supset\mathcal{M}_T}\sum_{k=1}^{K} n_k(\hat{\boldsymbol{\beta}}_{k,\mathcal{M}}-\tilde{\boldsymbol{\beta}}_k)^{\mathrm{T}}(\hat{\boldsymbol{\beta}}_{k,\mathcal{M}}-\tilde{\boldsymbol{\beta}}_k)-\\
&\nu_{\min}^{-1}\sum_{k=1}^{K} n_k(\hat{\boldsymbol{\beta}}_{k,\lambda_0}-\tilde{\boldsymbol{\beta}}_k)^{\mathrm{T}}(\hat{\boldsymbol{\beta}}_{k,\lambda_0}-\tilde{\boldsymbol{\beta}}_k)+\\
&\nu_{\max}^{-1}\sum_{k=1}^{K}\log(n_k). \qquad \text{(E.2)}
\end{aligned}
$$

The second inequality follows from the definition of $\hat{\boldsymbol{\beta}}_{k,\mathcal{M}_{\lambda_H}}$. The third inequality holds because the considered model is an overfitted model. In (E.2), the first term is $O_p(1)$ because for any $\mathcal{M}\supset\mathcal{M}_T$, $\hat{\boldsymbol{\beta}}_{k,\mathcal{M}}$ is $\sqrt{n_k}$ consistent; the second term is also $O_p(1)$; and the third term goes to infinity. Hence, $MIC_{\lambda_H} > MIC_{\lambda_0}$ with probability tending to 1.

Theoretical properties of SMA

We focus on the heterogeneous structure because it is more general than the homogeneous structure. Without loss of generality, we assume that the first p_0 covariates are active or partly active, i.e., $I=\{1,\ldots,p_0\}$ and $U=\{p_0+1,\ldots,p\}$. For $j=1,\ldots,p_0$, define $\mathcal{M}_j=\{k:\beta_{jk}^0\neq 0,k\in\mathcal{S}_j\}$ and $\mathcal{M}_j^c=\mathcal{S}_j\backslash\mathcal{M}_j$. Also, define $\mathcal{N}=\{(j,k):\beta_{jk}^0=0,j=1,\ldots,p;k\in\mathcal{S}_j\}$. For the penalty weights, define $t_{1n}=\max\{w_{jk}:1\leqslant j\leqslant p_0,k\in\mathcal{M}_j\}$, $t_{2n}=\min\{w_{jk}:1\leqslant j\leqslant p_0,k\in\mathcal{M}_j\}$, $g_{1n}=\max_{1\leqslant j\leqslant p,k\in\mathcal{S}_j}w_{jk}$, and $g_{2n}=\min\{w_{jk}:(j,k)\in\mathcal{N}\}$. Let $\boldsymbol{X}_k=(\boldsymbol{x}_{1k},\ldots,\boldsymbol{x}_{n_kk})^{\mathrm{T}}$, $n=\sum_{k=1}^{K}n_k$, and $\boldsymbol{\beta}^0$ be the vector stacked from $\boldsymbol{\beta}_1^0,\ldots,\boldsymbol{\beta}_K^0$. For $k=1,\ldots,K$, define $\mathcal{A}_k=\{j:\beta_{jk}^0\neq 0\}$ and $\mathcal{A}_k^c=\{j:\beta_{jk}^0=0\}$; let $\mathcal{D}_k=\{r=1,\ldots,p:k\in\mathcal{S}_r\}$, and assume that $\mathcal{A}_k\subseteq\mathcal{D}_k$. For $k=1,\ldots,K$, let $\boldsymbol{\Sigma}_k^{(k)}$ be the limiting covariance matrix of $\sqrt{n}_k\tilde{\boldsymbol{\beta}}_k$. Let $\boldsymbol{\beta}_{\mathcal{A}_k}^0$ be the subvector of $\boldsymbol{\beta}_k^0$ corresponding to $\mathcal{A}_k$, and let $\boldsymbol{\Sigma}_{\mathcal{A}_k}^{(k)}$ be the submatrix of $\boldsymbol{\Sigma}_k^{(k)}$ corresponding to $\mathcal{A}_k$. Let $\boldsymbol{\Sigma}_{\mathcal{A}_k\mathcal{A}_k^c}^{(k)}$ denote the submatrix of $\boldsymbol{\Sigma}^{(k)}$, with the row indices corresponding to $\mathcal{A}_k$ and the column indices corresponding to $\mathcal{A}_k^c$. Write $\boldsymbol{\Sigma}_{\mathcal{A}_k^c\mathcal{A}_k}^{(k)}=(\boldsymbol{\Sigma}_{\mathcal{A}_k\mathcal{A}_k^c}^{(k)})^{\mathrm{T}}$. Other subvectors and submatrices are to be understood in the same fashion. Let $\mathcal{I}^{\mathcal{A}_k}$ be the oracle Fisher information matrix, i.e., the one when $\mathcal{A}_k$ is known beforehand. Let $\boldsymbol{A}^{(k)}=(\mathcal{I}^{\mathcal{A}_k})^{-1}$.

We first consider p to be fixed. Let $a_n = t_{1n}t_{2n}^{-\frac{1}{2}}$ and $b_n = t_{1n}^2 t_{2n}^{-\frac{3}{2}}$. We show that if $\lambda n^{-\frac{1}{2}} a_n = O_p(1)$, $\lambda n^{-1} b_n = o_p(1)$, and $\lambda n^{-1} = o_p(1)$, then $\hat{\boldsymbol{\beta}}$ is $\sqrt{n}$ consistent. In addition, if $\lambda n^{-\frac{1}{2}} g_{2n} g_{1n}^{-\frac{1}{2}} \to \infty$, then $P(\hat{\beta}_{jk,\lambda} = 0) \to 1$ for any $(j,k) \in \mathcal{N}$, where $\hat{\beta}_{jk,\lambda}$ denotes the estimator of β_{jk}^0 for a given λ. Then, we show the asymptotic normality of the SMA estimator.

In the sequel, we consider

$$Q_n(\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_K) \equiv \sum_{k=1}^{K} (\tilde{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)^{\mathrm{T}} \boldsymbol{R}_k^{-1} (\tilde{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k) + \lambda \sum_{j=1}^{p} \left(\sum_{k \in \mathcal{S}_j} w_{jk} |\beta_{jk}| \right)^{\frac{1}{2}},$$

where $\boldsymbol{R}_k$ is an arbitrary covariance matrix.

Proof of the consistency for $\hat{\boldsymbol{\beta}}$ when p is fixed

Following Fan and Li (2001) and by the definition of Q_n,

$$\begin{aligned}
& Q_n(\boldsymbol{\beta}^0 + n^{-\frac{1}{2}} \boldsymbol{u}) - Q_n(\boldsymbol{\beta}^0) \\
\geqslant & \sum_{k=1}^{K} \boldsymbol{u}_k{}^{\mathrm{T}} (n\boldsymbol{R}_k)^{-1} \boldsymbol{u}_k + \sum_{k=1}^{K} 2\boldsymbol{u}_k{}^{\mathrm{T}} (n\boldsymbol{R}_k)^{-1} [\sqrt{n}(\boldsymbol{\beta}_k^0 - \tilde{\boldsymbol{\beta}}_k)] \\
& + \lambda \sum_{j=1}^{p_0} \left\{ \left[\sum_{k \in \mathcal{M}_j} w_{jk} |\beta_{jk}^0 + n^{-\frac{1}{2}} u_{jk}| \right]^{\frac{1}{2}} - \left[\sum_{k \in \mathcal{M}_j} w_{jk} |\beta_{jk}^0| \right]^{\frac{1}{2}} \right\} \\
\equiv & A_1 + A_2 + B,
\end{aligned}$$

where $\boldsymbol{u} = (\boldsymbol{u}_1^{\mathrm{T}}, \ldots, \boldsymbol{u}_K^{\mathrm{T}})^{\mathrm{T}}$. Assume that the eigenvalues of $(n_k \boldsymbol{R}_k)^{-1}$ are bounded by b and b'. Then, $A_1 \geqslant 0.5 \sum_{k=1}^{K} \nu_k \boldsymbol{u}_k^{\mathrm{T}} (n_k \boldsymbol{R}_k)^{-1} \boldsymbol{u}_k \geqslant 0.5b \sum_{k=1}^{K} \nu_k \|\boldsymbol{u}_k\|^2 \geqslant 0.5b\nu_{\min} C^2$ and $|A_2| \leqslant 2\sum_{k=1}^{K} \|\boldsymbol{u}_k\| \cdot \|(n\boldsymbol{R}_k)^{-1} [\sqrt{n}(\boldsymbol{\beta}_k^0 - \tilde{\boldsymbol{\beta}}_k)]\| = O_p(1)C$. Therefore, A_1 dominates A_2. The remaining proof regarding B follows from Wang et al. (2009) and Zhou and Zhu (2010) and is omitted.

Proof of the selection consistency when p is fixed

For $k = 1, \ldots, K$, let $\hat{\boldsymbol{\beta}}_{k,\lambda} = (\hat{\beta}_{jk,\lambda})^{\mathrm{T}}$ for $j \in \mathcal{D}_k$. Write $\hat{\boldsymbol{\beta}}_{\lambda} = (\hat{\boldsymbol{\beta}}_{1,\lambda}^{\mathrm{T}}, \ldots, \hat{\boldsymbol{\beta}}_{K,\lambda}^{\mathrm{T}})^{\mathrm{T}}$. For any $(j,k) \in \mathcal{N}$, if $\hat{\beta}_{jk,\lambda} \neq 0$, then it follows from the KarushKuhnTucker (KKT) conditions that

$$0 = n^{-\frac{1}{2}} \frac{\partial Q_n}{\partial \beta_{jk}}\Big|_{\hat{\boldsymbol{\beta}}_{\lambda}} = 2(n^{-1}\boldsymbol{R}_k^{-1})_{j.} \sqrt{n}(\hat{\boldsymbol{\beta}}_{k,\lambda} - \tilde{\boldsymbol{\beta}}_k) + n^{-\frac{1}{2}} \lambda \frac{1}{2} \left\{ \sum_{k' \in \mathcal{S}_j} w_{jk'} |\hat{\beta}_{jk',\lambda}| \right\}^{-\frac{1}{2}} w_{jk} \mathrm{sgn}(\hat{\beta}_{jk,\lambda}),$$

where $(\boldsymbol{R}_k^{-1})_{j\cdot}$ represents the jth row of $\boldsymbol{R}_k^{-1}$. It is straightforward to show that the first term on the right side of the above equation is $O_p(1)$. The remaining proof regarding the penalty part follows from Wang et al. (2009) and Zhou and Zhu (2010).

Proof of the asymptotic normality when p is fixed

Assume that $\lambda n^{-\frac{1}{2}} a_n = o_p(1)$, $\lambda n^{-1} b_n = o_p(1)$, $n^{-\frac{1}{2}} \lambda g_{2n} g_{1n}^{-\frac{1}{2}} \to \infty$, and $\lambda n^{-1} = o_p(1)$. Let m_k be the cardinality of $\mathcal{A}_k$. By the KKT conditions, $\frac{\partial Q_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_{\mathcal{A}}}\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = \mathbf{0}$, where $\mathcal{A} = \bigcup_{k=1}^K \mathcal{A}_k$. This implies that, for $k = 1, ..., K$,

$$\mathbf{0} = \{(n_k \boldsymbol{R}_k)^{-1}\}_{\mathcal{A}_k \mathcal{A}_k} (\hat{\boldsymbol{\beta}}_{\mathcal{A}_k} - \tilde{\boldsymbol{\beta}}_{\mathcal{A}_k}) - \{(n_k \boldsymbol{R}_k)^{-1}\}_{\mathcal{A}_k \mathcal{A}_k^c} \tilde{\boldsymbol{\beta}}_{\mathcal{A}_k^c} + n_k^{-1} \mathbf{e}_k, \tag{E.3}$$

where $\mathbf{e}_k$ is a vector of length m_k, with its sth component being $\frac{1}{2}\lambda \left(\sum_{j \in \mathcal{A}_k} w_{sk} |\hat{\beta}_{sj}| \right)^{-\frac{1}{2}} w_{sk} \mathrm{sgn}(\hat{\beta}_{sk})$. Since $\lambda n^{-\frac{1}{2}} a_n = o_p(1)$, each component of $\sqrt{n_k}(n_k^{-1} \mathbf{e}_k)$ is bounded by $o_p(1)$.

Define $\tilde{\boldsymbol{F}}^{(k)} = (n_k \boldsymbol{R}_k)^{-1}$, $\tilde{\boldsymbol{F}}^{(k)}_{\mathcal{A}_k \mathcal{A}_k} = \{(n_k \boldsymbol{R}_k)^{-1}\}_{\mathcal{A}_k \mathcal{A}_k}$, and $\tilde{\boldsymbol{F}}^{(k)}_{\mathcal{A}_k A_k^c} = \{(n_k \boldsymbol{R}_k)^{-1}\}_{\mathcal{A}_k A_k^c}$. Then, (E.3) implies that $\sqrt{n_k}(\hat{\boldsymbol{\beta}}_{\mathcal{A}_k} - \boldsymbol{\beta}^0_{\mathcal{A}_k}) = \sqrt{n}_k(\tilde{\boldsymbol{\beta}}_{\mathcal{A}_k} - \boldsymbol{\beta}^0_{\mathcal{A}_k}) + (\tilde{\boldsymbol{F}}^{(k)}_{\mathcal{A}_k \mathcal{A}_k})^{-1} \tilde{\boldsymbol{F}}^{(k)}_{\mathcal{A}_k \mathcal{A}_k^c} (\sqrt{n_k} \tilde{\boldsymbol{\beta}}_{\mathcal{A}_k^c}) + o_p(1)$.

Note that $\mathrm{Cov}(\sqrt{n_k} \tilde{\boldsymbol{\beta}}_{\mathcal{A}_k}) \to_p \boldsymbol{\Sigma}^{(k)}_{\mathcal{A}_k}$, $\mathrm{Cov}(\sqrt{n}_k \tilde{\boldsymbol{\beta}}_{\mathcal{A}_k^c}) \to_p \boldsymbol{\Sigma}^{(k)}_{\mathcal{A}_k^c \mathcal{A}_k^c}$, and $\mathrm{Cov}(\sqrt{n_k} \tilde{\boldsymbol{\beta}}_{\mathcal{A}_k}, \sqrt{n} \tilde{\boldsymbol{\beta}}_{\mathcal{A}_k^c}) \to_p \boldsymbol{\Sigma}^{(k)}_{\mathcal{A}_k \mathcal{A}_k^c}$. Therefore, we have the following results:

(1) For an arbitrary positive definite matrix $\boldsymbol{R}_k$,

$$\sqrt{n_k}(\hat{\boldsymbol{\beta}}_{\mathcal{A}_k} - \boldsymbol{\beta}^0_{\mathcal{A}_k}) \to N(0, \boldsymbol{G}_k),$$

where $\boldsymbol{G}_k = \boldsymbol{\Sigma}^{(k)}_{\mathcal{A}_k} + 2(\tilde{\boldsymbol{F}}^{(k)}_{\mathcal{A}_k \mathcal{A}_k})^{-1} \tilde{\boldsymbol{F}}^{(k)}_{\mathcal{A}_k \mathcal{A}_k^c} \boldsymbol{\Sigma}^{(k)}_{\mathcal{A}_k^c \mathcal{A}_k} + (\tilde{\boldsymbol{F}}^{(k)}_{\mathcal{A}_k \mathcal{A}_k})^{-1} \tilde{\boldsymbol{F}}^{(k)}_{\mathcal{A}_k \mathcal{A}_k^c} \boldsymbol{\Sigma}^{(k)}_{\mathcal{A}_k^c} \tilde{\boldsymbol{F}}^{(k)}_{\mathcal{A}_k^c \mathcal{A}_k} (\tilde{\boldsymbol{F}}^{(k)}_{\mathcal{A}_k \mathcal{A}_k})^{-1}$.

(2) If $\boldsymbol{R}_k = \tilde{\boldsymbol{V}}_k$, then because $n_k \tilde{\boldsymbol{V}}_k \to \boldsymbol{\Sigma}^{(k)}$, we have $(\tilde{\boldsymbol{F}}^{(k)}_{\mathcal{A}_k \mathcal{A}_k})^{-1} \tilde{\boldsymbol{F}}^{(k)}_{\mathcal{A}_k \mathcal{A}_k^c} \to_p -\boldsymbol{\Sigma}^{(k)}_{\mathcal{A}_k A_k^c} (\boldsymbol{\Sigma}^{(k)}_{\mathcal{A}_k^c})^{-1}$. Consequently, the asymptotic covariance matrix can be simplified to $\boldsymbol{\Sigma}^{(k)}_{\mathcal{A}_k} - \boldsymbol{\Sigma}^{(k)}_{\mathcal{A}_k \mathcal{A}_k^c} (\boldsymbol{\Sigma}^{(k)}_{\mathcal{A}_k^c})^{-1} \boldsymbol{\Sigma}^{(k)}_{\mathcal{A}_k^c \mathcal{A}_k}$, which can be shown to be equal to $\boldsymbol{A}^{(k)}$, the oracle covariance matrix.

Proof of the consistency for $\hat{\beta}$ when p is diverging

Now, we consider the situation in which the dimension p_n (indexed by n to emphasize its growth with n) tends to infinity (under the condition that $p_n^2/n \to 0$). By the definition of $\hat{\boldsymbol{\beta}}_k$, it can be shown that

$$\sum_{k=1}^{K} n_k \left\{ (\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k^0)^{\mathrm{T}} (n_k \boldsymbol{R}_k)^{-1} (\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k^0) - 2(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k^0)^{\mathrm{T}} (n_k \boldsymbol{R}_k)^{-1} (\tilde{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k^0) \right\} \leqslant \lambda \sum_{j=1}^{p_n} \left(\sum_{k \in \mathcal{S}_j} w_{jk} |\beta_{jk}^0| \right)^{\frac{1}{2}}.$$

Decompose $(n_k \boldsymbol{R}_k)^{-1}$ as $\boldsymbol{Q}_k^{\mathrm{T}} \boldsymbol{Q}_k$ and define $\boldsymbol{S}_k = \boldsymbol{Q}_k (\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k^0)$ and $\boldsymbol{T}_k = \boldsymbol{Q}_k (\tilde{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k^0)$. Also, define $\psi_n = \lambda \sum_{j=1}^{p_n} \left(\sum_{k \in \mathcal{S}_j} w_{jk} |\beta_{jk}^0| \right)^{\frac{1}{2}}$. Then,

$$2 \sum_{k=1}^{K} \frac{n_k}{n} \left(\boldsymbol{S}_k^{\mathrm{T}} \boldsymbol{S}_k - 2 \boldsymbol{S}_k^{\mathrm{T}} \boldsymbol{T}_k \right) + \sum_{k=1}^{K} 4 \frac{n_k}{n} \boldsymbol{T}_k^{\mathrm{T}} \boldsymbol{T}_k \leqslant \sum_{k=1}^{K} 4 \frac{n_k}{n} \boldsymbol{T}_k^{\mathrm{T}} \boldsymbol{T}_k + \frac{2}{n} \psi_n. \tag{E.4}$$

The left side of (E.4) equals $\sum_{k=1}^{K} n_k/n \left\{ ||\boldsymbol{S}_k||^2 + ||\boldsymbol{S}_k - 2\boldsymbol{T}_k||^2 \right\} \geqslant \sum_{k=1}^{K} n_k/n ||\boldsymbol{S}_k||^2$. Since $\sum_{k=1}^{K} n_k/n ||\boldsymbol{S}_k||^2 \geqslant b\nu_{\min} \sum_{k=1}^{K} ||\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k^0||^2$ and $\sum_{k=1}^{K} 4n_k/n \boldsymbol{T}_k^{\mathrm{T}} \boldsymbol{T}_k \leqslant 4b'\nu_{\max} \sum_{k=1}^{K} ||\tilde{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k^0||^2$, it follows that $b\nu_{\min} \sum_{k=1}^{K} ||\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k^0||^2 \leqslant 4b'\nu_{\max} \sum_{k=1}^{K} ||\tilde{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k^0||^2 + 2/n\psi_n$. For the M-estimators of exponential families, if $p_n^2/n \to 0$, then $||\tilde{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k^0||^2 = O_p(p_n/n)$ (Portnoy, 1988). Thus,

$$||\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0||^2 = \sum_{k=1}^{K} ||\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k^0||^2 = O_p \left(\frac{4Kp_n b'\nu_{\max}}{nb\nu_{\min}} + \frac{2\psi_n}{nb\nu_{\min}} \right),$$

which can be further shown to be

$$O_p \left(\left\{ \frac{p_n + \lambda\sqrt{t_{1n}}}{n} \right\}^{1/2} \right).$$

Consequently, if $\lambda\sqrt{t_{1n}} = O_p(p_n)$, then $\hat{\boldsymbol{\beta}}$ is root-(n/p_n) consistent.

Next, we prove that, even if p_n grows large, we can still estimate the zero coefficients exactly at zero.

Proof of the selection consistency when p is diverging

Assume that $\lambda\sqrt{t_{1n}} = O_p(p_n)$, $\lambda/\sqrt{n} \to 0$, and $\lambda(np_n)^{-1/2} g_{2n} g_{1n}^{-\frac{1}{2}} \to \infty$. We wish to prove that $P(\hat{\beta}_{jk,\lambda} = 0) \to 1$ for any $(j,k) \in \mathcal{N}$.

For any $(j,k) \in \mathcal{N}$, if $\hat{\beta}_{jk,\lambda} \neq 0$, then by the KKT conditions,

$$\begin{aligned}
0 &= \frac{\partial Q_n(\boldsymbol{\beta})}{\partial \beta_{jk}}\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_\lambda} \\
&= 2n_k \left[(n_k \boldsymbol{R}_k)^{-1}\right]_{j\cdot} (\hat{\boldsymbol{\beta}}_{k,\lambda} - \tilde{\boldsymbol{\beta}}_k) + \lambda \frac{1}{2} \left\{ \sum_{k' \in \mathcal{S}_j} w_{jk'} |\hat{\beta}_{jk',\lambda}| \right\}^{-\frac{1}{2}} w_{jk} \mathrm{sgn}(\hat{\beta}_{jk,\lambda}) \\
&\equiv E_1 + E_2,
\end{aligned}$$

where $[\boldsymbol{H}]_{j\cdot}$ represents the jth row of $\boldsymbol{H}$. Because $||\hat{\boldsymbol{\beta}}_{k,\lambda} - \boldsymbol{\beta}_k^0|| = O_p(\sqrt{p_n/n})$ and $||\tilde{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k^0|| = O_p(\sqrt{p_n/n})$, we have $||\hat{\boldsymbol{\beta}}_{k,\lambda} - \tilde{\boldsymbol{\beta}}_k|| = O_p(\sqrt{p_n/n})$. Define $\boldsymbol{T} = (n_k \boldsymbol{R}_k)^{-1}$. Because the eigenvalues of $\boldsymbol{T}$ are bounded,

$$|E_1| \leqslant 2n O_p(\sqrt{p_n/n}) \left\{ \sum_{i=1}^{p_n} \boldsymbol{T}_{ij}^2 \right\}^{1/2} = O_p(\sqrt{np_n}).$$

The remaining proof regarding E_2 follows from Wang et al. (2009) and Zhou and Zhu (2010).

Finally, we investigate the asymptotic normality of the SMA estimator, which indicates that the SMA (with the full estimated covariance matrix) has the oracle property.

Proof of asymptotic normality when p is diverging

Assume that $\lambda\sqrt{t_{1n}} = O_p(p_n)$, $\lambda/\sqrt{n} \to 0$, $\lambda(np_n)^{-1/2} g_{2n} g_{1n}^{-\frac{1}{2}} \to \infty$, and $p_n^2/n \to 0$. Assume also that $\boldsymbol{R}_k = \tilde{\boldsymbol{V}}_k$. Let $\boldsymbol{\gamma}_n$ be a vector of length p_0 with norm 1, and define $s_{n,k}^2 = \boldsymbol{\gamma}_n^{\mathrm{T}} \boldsymbol{A}^{(k)} \boldsymbol{\gamma}_n$ $(k = 1, \ldots, K)$. We wish to prove that $\sqrt{n_k} s_{n,k}^{-1} \boldsymbol{\gamma}_n^{\mathrm{T}} (\hat{\boldsymbol{\beta}}_{\mathcal{A}_k} - \boldsymbol{\beta}_{\mathcal{A}_k}^0) \to N(0,1)$.

The treatment of the penalty part in this proof is similar to Zhou and Zhu (2010). Because we have shown that, with an arbitrarily large probability, the estimator of $\{\beta_{jk}^0 : (j,k) \in \mathcal{N}\}$ must be 0, we can decompose $\hat{\boldsymbol{\beta}}$ into $\{\hat{\boldsymbol{\beta}}_{\mathcal{A}}, \mathbf{0}\}$. By the KKT conditions, $\hat{\boldsymbol{\beta}}$ should satisfy

$$\frac{\partial Q_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_{\mathcal{A}_k}}\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = \mathbf{0}.$$

We first consider linear models. By the definition of Q_n and the fact that $\tilde{\boldsymbol{V}}_k^{-1} = \hat{\sigma}_k^{-2} \boldsymbol{X}_k^{\mathrm{T}} \boldsymbol{X}_k$, where $\hat{\sigma}_k^2$ is the estimated error variance,

$$
\begin{aligned}
Q_n(\boldsymbol{\beta}) = \sum_{k=1}^{K} [(\boldsymbol{\beta}_{\mathcal{A}_k} - \tilde{\boldsymbol{\beta}}_{\mathcal{A}_k})^{\mathrm{T}} (\boldsymbol{X}_{\mathcal{A}_k}^{\mathrm{T}} \boldsymbol{X}_{\mathcal{A}_k}) (\boldsymbol{\beta}_{\mathcal{A}_k} - \tilde{\boldsymbol{\beta}}_{\mathcal{A}_k}) + 2(\boldsymbol{\beta}_{\mathcal{A}_k} - \tilde{\boldsymbol{\beta}}_{\mathcal{A}_k})^{\mathrm{T}} (\boldsymbol{X}_{\mathcal{A}_k}^{\mathrm{T}} \boldsymbol{X}_{\mathcal{A}_k^c}) (\boldsymbol{\beta}_{\mathcal{A}_k^c} - \tilde{\boldsymbol{\beta}}_{\mathcal{A}_k^c}) \\
+ (\boldsymbol{\beta}_{\mathcal{A}_k^c} - \tilde{\boldsymbol{\beta}}_{\mathcal{A}_k^c})^{\mathrm{T}} (\boldsymbol{X}_{\mathcal{A}_k^c}^{\mathrm{T}} \boldsymbol{X}_{\mathcal{A}_k^c}) (\boldsymbol{\beta}_{\mathcal{A}_k^c} - \tilde{\boldsymbol{\beta}}_{\mathcal{A}_k^c})] / \hat{\sigma}_k^2 + \lambda \sum_{j=1}^{p} \left(\sum_{k \in \mathcal{S}_j} w_{jk} |\beta_{jk}| \right)^{\frac{1}{2}}.
\end{aligned}
$$

Define $\boldsymbol{y}_k = (y_{1k}, \ldots, y_{n_k k})^{\mathrm{T}}$. Then, for each $k = 1, \ldots, K$, we have

$$
\mathbf{0} = \frac{\hat{\sigma}_k^2}{2} \frac{\partial Q_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_{\mathcal{A}_k}} \Big|_{(\hat{\boldsymbol{\beta}}_{\mathcal{A}_k}, \mathbf{0})} = (\boldsymbol{X}_{\mathcal{A}_k}^{\mathrm{T}} \boldsymbol{X}_{\mathcal{A}_k}) (\hat{\boldsymbol{\beta}}_{\mathcal{A}_k} - \tilde{\boldsymbol{\beta}}_{\mathcal{A}_k}) + (\boldsymbol{X}_{\mathcal{A}_k}^{\mathrm{T}} \boldsymbol{X}_{\mathcal{A}_k^c}) (\hat{\boldsymbol{\beta}}_{\mathcal{A}_k^c} - \tilde{\boldsymbol{\beta}}_{\mathcal{A}_k^c}) + \hat{\boldsymbol{e}}_k,
$$

where $\hat{\boldsymbol{e}}_k$ is a vector with its sth component being $\lambda \hat{\sigma}_k^2 \{\sum_{l \in \mathcal{S}_s} w_{sl} |\hat{\beta}_{sl}|\}^{-\frac{1}{2}} w_{sk} \mathrm{sgn}(\hat{\beta}_{sk})/4$. It can be shown that the last equation is equivalent to

$$
\mathbf{0} = (\boldsymbol{X}_{\mathcal{A}_k}^{\mathrm{T}} \boldsymbol{X}_{\mathcal{A}_k}) (\hat{\boldsymbol{\beta}}_{\mathcal{A}_k} - \boldsymbol{\beta}_{\mathcal{A}_k}^0) + \boldsymbol{X}_{\mathcal{A}_k}^{\mathrm{T}} \boldsymbol{X}_{\mathcal{A}_k^c} \hat{\boldsymbol{\beta}}_{\mathcal{A}_k^c} - \boldsymbol{X}_{\mathcal{A}_k}^{\mathrm{T}} \boldsymbol{\varepsilon}_k + \hat{\boldsymbol{e}}_k,
$$

where $\boldsymbol{\varepsilon}_k = (\epsilon_{1k}, \ldots, \epsilon_{n_k k})$ is the random error. Because $P(\hat{\boldsymbol{\beta}}_{\mathcal{A}_k^c} = \mathbf{0}) \to 1$ (by the selection consistency), we have

$$
(\boldsymbol{X}_{\mathcal{A}_k}^{\mathrm{T}} \boldsymbol{X}_{\mathcal{A}_k}) (\hat{\boldsymbol{\beta}}_{\mathcal{A}_k} - \boldsymbol{\beta}_{\mathcal{A}_k}^0) = \sum_{i=1}^{n_k} \boldsymbol{x}_{i\mathcal{A}_k} \epsilon_{ik} - \hat{\boldsymbol{e}}_k,
$$

where $\boldsymbol{x}_{i\mathcal{A}_k}$ represents the subvector corresponding to the $\mathcal{A}_k$ part of $\boldsymbol{x}_{ik}$. Thus,

$$
\sqrt{n}_k \boldsymbol{\gamma}_n^{\mathrm{T}} (\hat{\boldsymbol{\beta}}_{\mathcal{A}_k} - \boldsymbol{\beta}_{\mathcal{A}_k}^0) = \frac{1}{\sqrt{n}_k} \sum_{i=1}^{n_k} \epsilon_{ik} \boldsymbol{\gamma}_n^{\mathrm{T}} \left(\frac{\boldsymbol{X}_{\mathcal{A}_k}^{\mathrm{T}} \boldsymbol{X}_{\mathcal{A}_k}}{n_k} \right)^{-1} \boldsymbol{x}_{i\mathcal{A}_k} - \frac{1}{\sqrt{n}_k} \boldsymbol{\gamma}_n^{\mathrm{T}} \left(\frac{\boldsymbol{X}_{\mathcal{A}_k}^{\mathrm{T}} \boldsymbol{X}_{\mathcal{A}_k}}{n_k} \right)^{-1} \hat{\boldsymbol{e}}_k.
$$

Note that

$$
\left| \frac{1}{\sqrt{n}_k} \boldsymbol{\gamma}_n^{\mathrm{T}} \left(\frac{\boldsymbol{X}_{\mathcal{A}_k}^{\mathrm{T}} \boldsymbol{X}_{\mathcal{A}_k}}{n_k} \right)^{-1} \hat{\boldsymbol{e}}_k \right| \leqslant \frac{1}{\sqrt{n}_k} ||\boldsymbol{\gamma}_n|| \times b^{-1} \times ||\hat{\boldsymbol{e}}_k|| = O_p(n_k^{-1/2} \lambda),
$$

where b is the lower bound of the eigenvalues of $n_k^{-1} \boldsymbol{X}_k' \boldsymbol{X}_k (k = 1, \ldots, K)$. The equality holds because $||\boldsymbol{\gamma}_n|| = 1$ and $||\hat{\boldsymbol{e}}_k|| = O_p(\lambda)$. Since $n_k^{-1/2} \lambda = o_p(1)$, we have

$$
\frac{1}{\sqrt{n}_k} \boldsymbol{\gamma}_n^{\mathrm{T}} \left(\frac{\boldsymbol{X}_{\mathcal{A}_k}^{\mathrm{T}} \boldsymbol{X}_{\mathcal{A}_k}}{n_k} \right)^{-1} \hat{\boldsymbol{e}}_k = o_p(1).
$$

Therefore,

$$\sqrt{n}_k s_{n,k}^{-1} \boldsymbol{\gamma}_n^{\mathrm{T}} (\hat{\boldsymbol{\beta}}_{\mathcal{A}_k} - \boldsymbol{\beta}_{\mathcal{A}_k}^0) = \frac{s_{n,k}^{-1}}{\sqrt{n}_k} \sum_{i=1}^{n_k} \epsilon_{ik} \boldsymbol{\gamma}_n^{\mathrm{T}} \left(\frac{\boldsymbol{X}_{\mathcal{A}_k}^{\mathrm{T}} \boldsymbol{X}_{\mathcal{A}_k}}{n_k} \right)^{-1} \boldsymbol{x}_{i\mathcal{A}_k} + o_p(1).$$

This equation is equivalent to equation (14) in the proof of Theorem 2 in Huang *et al.* (2008). Following their arguments, we can show that $\sqrt{n}_k s_{n,k}^{-1} \boldsymbol{\gamma}^{\mathrm{T}} (\hat{\boldsymbol{\beta}}_{\mathcal{A}_k} - \boldsymbol{\beta}_{\mathcal{A}_k}^0) \to N(0,1)$. For generalized linear models, we can use Theorem 3.1 of Portnoy (1988) to obtain the asymptotic distribution of $\tilde{\boldsymbol{\beta}}_k$ and then show a similar asymptotic normality of $\hat{\boldsymbol{\beta}}_{\mathcal{A}_k}$.

SECTION B: RESULTS OF ADDITIONAL SIMULATION STUDIES

B.1 Performance of the SMA under the heterogeneous structure

Five studies were simulated under the heterogeneous structure, as shown in the lower panel of Figure 1 in the main text. The effect sizes of the covariates are described in Sections 3.1 and 3.2. To demonstrate the differences between our heterogeneous structure and separate variable selection in individual studies, we included an approach that conducts the adaptive-LASSO for each individual study and called it Indv-aLASSO. The results are shown in Tables S1a and S1b. We conclude that 1) the SMA performs similarly to the Raw method and better than the other methods; and 2) the Indv-aLASSO tends to select a model considerably larger than the true model, and its prediction error tends to be higher than the SMA.

B.2 Performance of the SMA under the subgroup structure

We have considered the homogeneous and heterogeneous structures. We now consider the subgroup structure. We first provide a motivating example for the sub-group structure. Suppose that we analyze 5 studies for a certain disease and that two of them belong to a subcategory of the disease, whereas the other three belong to another subcategory. This is a particularly common scenario in psychiatric research, where one major disease may include many subcategories. Thus, the 5 studies can be treated as two subgroups based on their clinical information. For this situation, it is reasonable to assume that, in addition to the active covariates shared by all subgroups, each subgroup has its own set of active covariates.

Next, we describe the simulation studies for the sub-group structure. We assume that there are H subgroups in the K studies, and we let $\mathcal{S}_h$ denote the set of studies in the hth subgroup. In our simulation studies, we let $K = 5$, $H = 2$, $\mathcal{S}_1 = \{1, 2\}$ and $\mathcal{S}_2 = \{3, 4, 5\}$. That is, we let the 5 studies contain 2 subgroups, where subgroup 1 consists of studies 1 and 2, and subgroup 2 consists of studies 3, 4, and 5. The set of covariates with nonzero coefficients in subgroup 1 was set to $\{1, 4, 6, 9, 11, 14, 16, 21, 26, 31\}$, and the set of covariates with nonzero coefficients in subgroup 2 was set to $\{1, 4, 6, 9, 11, 14, 16, 21\}$. In other words, covariates 1, 4, 6, 9, 11, 14, 16, and 21 were active in both subgroups, whereas covariates 26 and 31 were active only in subgroup 1. We considered $p = 200$. The objective function for the subgroup structure was the same as equation (1) in the main text, except that the penalty weights were chosen as follows

$$w_{jk} = \left\{ \sum_{h=1}^{H} \left[I(k \in \mathcal{S}_h) |\mathcal{S}_h|^{-1} \sum_{l \in \mathcal{S}_h} |\tilde{\beta}_{jl}| \right] \right\}^{-1}, j = 1, ..., p, \text{ and } k = 1, ..., K,$$

where $|\mathcal{S}_h|$ is the cardinality of $\mathcal{S}_h$. As shown in Table S2, our method has a high probability of identifying the important covariates and tends to outperform the aLASSO-U and aLASSO-I methods.

B.3 Comparison of the SMA with individual aLASSO on the capture rate

One of the underlying assumptions of meta-analysis is that analyzing multiple studies together may lead to a better chance of capturing weak covariates than analyzing each study separately. To investigate this issue, we revisited the model with the homogeneous structure in Section 3.1 of the main text and let all of the nonzero β^0_{jk} equal to θ^0 (for example, θ^0=0.2, 0.25, or 0.3). Then, we calculated the proportion of all nonzero β^0_{jk}s that were captured by a given variable selection method, and we call this proportion the "capture rate". We compared the SMA with the method that applies the aLASSO to each study separately. The results are shown in Table S3. The SMA tends to have a higher capture rate than the individual aLASSO for the considered θ^0.

B.4 The influence of sparsity p_0 on regression analysis

We studied whether the sparsity p_0/p has any influence on variable selection. Considering $p = 100$ under the homogeneous structure, we let $p_0 = (10, 15, 25)$, which corresponds to the sparsity of $(10\%, 15\%, 25\%)$. The indices of the nonzero covariates for the three sparsity situations are (1,2,4,5,6,7,9,10,11,12), (1,2,4,5,6,7,9,10,11,12,14,16,18,19,20), and (1,2,4,5,6,7,9,10,11,12,14,16,18, 19,20,21,22,23,26,28,30,31,33,34,35), respectively. The results in Table S4 show that the parameter estimation and prediction errors are both influenced by the ratio of p_0/p, but the SMA performs always better than the aLASSO-U and aLASSO-I methods.

B.5 Performance of the SMA with different sets of covariates from different studies

In the main text, the dimensions of the K sets of summary statistics were assumed to be the same. We now consider the situation in which the summary statistics were calculated from different (but overlapping) sets of covariates. In this section, we require that each covariate is present in at least one study. (We consider the situation in which some covariates are absent in all of the K studies in Section B.6.) For illustration, we considered the homogeneous structure with 100 covariates.

First, we let 10% of the noise covariates from each study be "missing" at different places among the K studies. Because the Raw method is for the situation where all the raw data are available, we omitted the Raw method for comparison. We let p_0 be 10, 15, and 25. Table S5a shows that the SMA handles this situation well.

Next, we allowed both the noise and the important covariates to be "missing" at a rate of 10%. The results are shown in Table S5b. Clearly, the missing of important covariates drives up both the parameter estimation error and the prediction error. However, the SMA is still a competitive method compared to the aLASSO-based methods.

B.6 The influence of p_0 and unmeasured covariates on prediction errors

It is possible that some covariates that have influence on the trait of interest are not measured in any of the K studies. For example, gene expression and DNA methylation can be important genomic features for certain traits. Likewise, other non-genetic covariates, such as exercise, food choice, and life style, may also contribute to the trait of interest. For various reasons, these covariates may not be measured for the K studies. We conducted simulation studies to examine the impact of unmeasured covariates on prediction errors. For demonstration, we considered $p = 50$ and simulated two unmeasured covariates, each with an effect size of 0.8. Then we ignored these two covariates in our analysis in order to assess their influence on the prediction accuracy. We also studied how the sparsity, p_0, influences the regression analysis. We considered $p_0 = (10, 15, 25)$.

The results in Table S6 show that when the trait involves some unmeasured covariates, using only the measured covariates can lead to inflated prediction errors (for all of the considered methods). In addition, p_0 can also influence the prediction errors. In all of the situations, the SMA has superior performance to the aLASSO-U and aLASSO-I.

B.7 The influence of screening procedures on the performance of variable selection

For $p > n$, a screening step is performed before the variable selection is conducted, as shown in Section 3.3 of the main text. To study the impact of the screening step on the performance of variable selection, we considered several screening procedures that are relevant to meta-analysis. Consider K studies with sample sizes of $n_k (k = 1, \ldots, K)$ and assume that a marginal screening is conducted for each study. Let $n^* = \max_{1 \leqslant k \leqslant K}\{n_k\}$. The first screening procedure (Screening Procedure I) involved selecting the top candidate covariates based on the marginal screening of the study that has the largest sample size, i.e., we select the top $n_k/(3\log(n_k))$ covariates from the study with the largest n_k. The second screening procedure (Screening Procedure II) involved first using the summary statistics (from the marginal screening) of the K studies to conduct a meta-analysis for each covariate and then picking the top $n^*/(3\log(n^*))$ covariates based on the meta-analysis results. The third screening procedure (Screening Procedure III) involved picking the top $n_k/(3\log(n_k))$ from each study and then taking a union of the K sets of top covariates as the selected covariates. After the screening procedure is performed, we subjected the selected candidate covariates to the considered variable selection methods to examine the influence of the screening procedures. We considered $p = 1000$ and sample sizes of $(600, 500, 600, 500, 400)$. We first studied the screening procedures under the homogeneous structure. The results in Table S7a show that for the SMA approach, Screening Procedure II yields quite similar estimation error and prediction error as Screening Procedure III, and both procedures outperform Screening

Procedure I. Next, we studied the screening procedures under the heterogeneous structure. The results in Table S7b show that for the Raw and SMA approaches, both Screening Procedure II and III are better than Screening Procedure I. Overall, these results demonstrate that using information from multiple studies tends to be more powerful in retaining important covariates than using information from only one study.

In the previous simulations, we assumed that the regression coefficients of a covariate across the K studies have the same sign with large probability. To evaluate situations in which this assumption is violated, we changed the sign for one of the K regression coefficients associated with the first covariate (i.e., β_{15}^0) and studied how the change of sign influences the screening procedures and subsequent variable selection. The results are shown in Table S7c. The parameter estimation error and prediction error with Screening Procedure II tend to be higher than those with Screening Procedure III for the Raw and SMA methods. This is mainly because Screening Procedure II tends to cancel out a covariate's signal if that covariate has regression coefficients with opposite signs (among the K studies). Overall, our simulation studies suggest that Screening Procedure III is more robust than Screening Procedure II for reducing dimensions.

In the simulation studies reported in Section 3.3 of the main text, we reduced the dimension to $n/(3\log(n))$. To keep a larger number of variables for downstream analysis, we reduced the dimension to $n/(2\log(n))$ instead. The corresponding simulation results are shown in Table S7d. We see that the Correct_0 rate and Incorrect_0 rate of SMA are still reasonable.

B.8 Sensitivity analysis

We conducted simulation studies to examine the sensitivity of our methods to data perturbation. We perturbed the trait by adding additional noise, 0.5×Binomial(0,1), to the random error $N(0,1)$. The results in Table S8 show that the SMA still performs well.

B.9 Weaker genetic effects

We examined the performance of the SMA with relatively weak genetic effects. We considered the set-up in Section 3.1 of the main text but simulated the non-zero values of the coefficients for the jth variable from the $(-1)^j \times N(0.25, 0.125^2)$ distribution, such that the variance explained by individual SNPs fluctuates around 2.7% and is below 1% for 25% of the SNPs. As shown in Table S9, the SMA still competes well with the other methods.

Section C: Additional Details about Real Data Analysis

The Atherosclerosis Risk in Communities Study (ARIC) is a population-based cohort study of ∼16,000 Caucasian and African-American subjects, who were recruited in 1987−1989 to investigate the development of cardiovascular and pulmonary disease over time. The Coronary Artery Risk Development in Young Adults (CARDIA), which was initiated in 1985−1986, is a longitudinal study of ∼5,000 African American and Caucasian subjects on the evolution of coronary heart disease risk. The Cardiovascular Health Study (CHS) is a prospective, longitudinal cohort study of risk factors for cardiovascular disease. It was initiated in 1989, with ∼6,000 Caucasian and African-American adults who were sampled from four United States communities. The Framingham Heart Study (FHS) was initiated in 1948 with randomly ascertained participants to investigate cardiovascular disease and related risk factors. The Multi-Ethnic Study of Atherosclerosis (MESA) was initiated in year 2000 to investigate subclinical cardiovascular disease and the risk factors that predict progression to clinically overt cardiovascular disease.

The subjects from the five cohorts were genotyped on ∼230,000 SNPs. We conducted marginal screening for each study and selected SNPs with p-value $\leqslant 5 \times 10^{-4}$. We took the union of the selected SNPs and identified the SNPs that have genotypes in all five cohorts. We removed highly correlated SNPs both before and after the union procedure. The final set that was subject to variable selection contained 276 SNPs.

The SNPs selected by the SMA are located in various regions of the corresponding genes. For example, rs4420638 is located downstream of the APOC1 gene, while rs1367117 is located in an exon of the APOB gene; see Supplementary Figures S1 and S2. According to the GTEx portal, rs4420638 is associated with the expression of APOE in spleen, with a p-value of 0.095, and APOE sits next to APOC1 and is known to encode an important lipoprotein. SNP rs1367117 results in a change of amino acid from Thr to Ile, which may potentially change its binding activity with lipids. Most of the selected SNPs are located in non-coding regions, and their functions remain to be elucidated. In Table S10, we show the models selected by the Raw method for the 7 studies (i.e., FHS, MESA-A, MESA-E, CARDIA-A, CARDIA-E, CHS-A, and CHS-E). In Table S11, we show the prediction errors for the compared methods under the heterogeneous structure, with the prediction error divided by the trait variance.

We next explored the homogeneous structure for the 7 studies, and the results are shown in Tables S12 and S13. It can be seen that the SMA still performs better than the aLASSO methods.

Finally, we included more SNPs for variable selection by relaxing the marginal screening p-value to $\leqslant 8 \times 10^{-4}$. After the removal of highly correlated SNPs, the final set that was subject to variable selection contained 379 SNPs. Applying the SMA to those 379 SNPs, we achieved the prediction error of 2.63. Thus, including more candidate SNPs does not significantly improve the prediction accuracy.

Additional References

Huang, J., Horowitz, J. L. and Ma, S. (2008). Asymptotic properties of Bridge estimators in sparse high-dimensional regression models. *The Annals of Statistics* **36**(2), 587-613.

Portnoy, S. (1988). Asymptotic behavior of likelihood methods for exponential families when the numbers of parameters tends to infinity. *The Annals of Statistics* **16**(1), 356–366.

Wang, S., Nan, B., Zhou, N. and Zhu, J. (2009). Hierarchically penalized Cox regression with grouped variables *Biometrika* **96**(2), 307-322.

Zhou, N. and Zhu, J. (2010). Group variable selection via a hierarchical lasso and its oracle property. *Statistics And Its Interface* **3**(4), 557-574.

Table S1a. Comparison of the SMA with other methods under the heterogeneous structure for $p = 50$

Method	$\sum_{k=1}^{5}\lvert\hat{\mathcal{M}}^{(k)}\rvert$	Correct_0 (%)	Incorrect_0 (%)	$\lvert\lvert\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\rvert\rvert^2/5$	Prediction error
Correlation structure: Auto-regressive ($\rho = 0.3$)					
Raw	39.92	99.8	3.9	0.016	1.008
SMA	39.97	99.7	3.9	0.016	1.008
SMA-Diag	40.07	99.6	4.1	0.018	1.009
SMA-I	39.79	99.8	4.0	0.017	1.008
SMA-E	40.30	99.6	3.7	0.017	1.009
aLASSO-U	164.45	40.9	0.1	0.028	1.014
aLASSO-I	30.40	100.0	26.0	0.613	1.684
HT-U	88.85	77.1	0	0.033	1.016
HT-I	28.00	100.0	31.7	0.739	1.834
Indv-aLASSO	72.41	84.4	2.9	0.028	1.014
Correlation structure: Auto-regressive ($\rho = 0.6$)					
Raw	40.14	99.6	4.0	0.016	1.008
SMA	40.14	99.6	3.9	0.016	1.008
SMA-Diag	42.56	98.5	3.8	0.025	1.012
SMA-I	40.08	99.7	3.9	0.017	1.008
SMA-E	40.38	99.5	3.9	0.018	1.008
aLASSO-U	164.90	40.7	0	0.033	1.014
aLASSO-I	30.30	100.0	26.2	0.629	1.726
HT-U	114.70	64.7	0	0.056	1.024
HT-I	28.30	100.0	31.0	0.722	1.850
Indv-aLASSO	72.79	84.2	3.0	0.033	1.014
Correlation structure: Compound symmetry ($\rho = 0.3$)					
Raw	40.07	99.7	3.9	0.016	1.008
SMA	40.13	99.7	3.8	0.016	1.008
SMA-Diag	40.76	99.4	3.9	0.018	1.009
SMA-I	40.05	99.7	4.1	0.017	1.009
SMA-E	40.65	99.4	3.9	0.018	1.009
aLASSO-U	163.25	41.5	0.1	0.030	1.014
aLASSO-I	29.8	100.0	27.4	0.645	1.711
HT-U	94.10	74.6	0	0.036	1.017
HT-I	27.9	100.0	32.0	0.749	1.856
Indv-aLASSO	71.70	84.7	3.1	0.030	1.014
Correlation structure: Compound symmetry ($\rho = 0.5$)					
Raw	39.65	99.8	4.5	0.019	1.008
SMA	39.72	99.7	4.5	0.019	1.008
SMA-Diag	41.66	98.9	4.1	0.024	1.012
SMA-I	39.96	99.6	4.4	0.020	1.009
SMA-E	40.13	99.5	4.5	0.021	1.009
aLASSO-U	165.65	40.36	0	0.035	1.014
aLASSO-I	29.45	1	28.2	0.659	1.741
HT-U	107.00	68.4	0	0.051	1.023
HT-I	27.45	100.0	33.0	0.767	1.896
Indv-aLASSO	72.61	84.2	3.2	0.035	1.014

Table S1b. Comparison of the SMA with other methods under the heterogeneous structure for $p = 200$

Method	$\sum_{k=1}^{5} \lvert\hat{\mathcal{M}}^{(k)}\rvert$	Correct_0 (%)	Incorrect_0 (%)	$\lvert\lvert\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\rvert\rvert^2/5$	Prediction error
		Correlation structure: Auto-regressive($\rho = 0.3$)			
Raw	40.74	99.9	2.9	0.011	1.006
SMA	40.69	99.9	3.0	0.011	1.006
SMA-Diag	42.53	99.7	3.0	0.014	1.008
SMA-I	41.93	99.8	2.9	0.013	1.007
SMA-E	45.29	99.4	2.5	0.016	1.009
aLASSO-U	528.75	49.1	0	0.038	1.019
aLASSO-I	31.50	100.0	23.4	0.562	1.670
HT-U	134.15	90.3	0.1	0.042	1.021
HT-I	28.20	100.0	31.2	0.722	1.893
Indv-aLASSO	164.38	87.0	2.0	0.038	1.019
		Correlation structure: Auto-regressive ($\rho = 0.6$)			
Raw	40.96	99.9	3.1	0.011	1.006
SMA	40.72	99.9	3.1	0.011	1.006
SMA-Diag	51.57	98.8	2.8	0.024	1.012
SMA-I	41.94	99.8	3.0	0.013	1.007
SMA-E	45.34	99.4	2.8	0.017	1.009
aLASSO-U	508.20	51.3	0	0.040	1.018
aLASSO-I	31.00	100.0	24.5	0.586	1.706
HT-U	216.45	81.7	0.1	0.085	1.038
HT-I	28.10	100.0	31.5	0.730	1.945
Indv-aLASSO	158.47	87.6	2.5	0.040	1.018
		Correlation structure: Compound symmetry ($\rho = 0.3$)			
Raw	40.88	99.9	2.9	0.011	1.006
SMA	40.75	99.9	2.9	0.011	1.006
SMA-Diag	43.1	99.7	2.7	0.015	1.008
SMA-I	41.96	99.8	2.8	0.014	1.007
SMA-E	45.83	99.4	2.3	0.017	1.009
aLASSO-U	532.35	48.8	0	0.040	1.019
aLASSO-I	31.40	100.0	23.6	0.549	1.627
HT-U	137.65	89.9	0.1	0.045	1.022
HT-I	28.20	100.0	31.2	0.718	1.896
Indv-aLASSO	166.66	86.8	2.1	0.040	1.019
		Correlation structure: Compound symmetry ($\rho = 0.5$)			
Raw	40.76	99.9	3.0	0.013	1.006
SMA	40.70	99.9	3.1	0.013	1.006
SMA-Diag	49.69	99.0	2.3	0.023	1.011
SMA-I	42.19	99.8	2.8	0.016	1.007
SMA-E	46.23	99.4	2.4	0.019	1.009
aLASSO-U	516.90	50.4	0	0.043	1.019
aLASSO-I	31.40	100.0	23.5	0.540	1.632
HT-U	196.40	83.8	0	0.076	1.035
HT-I	28.40	100.0	30.7	0.696	1.866
Indv-aLASSO	160.53	87.4	2.2	0.043	1.019

Table S2. Comparison of the SMA with other methods under the subgroup structure for $p = 200$

Method	$\sum_{k=1}^{5} \lvert\hat{\mathcal{M}}^{(k)}\rvert$	Correct_0 (%)	Incorrect_0 (%)	$\lvert\lvert\boldsymbol{\beta}^0 - \hat{\boldsymbol{\beta}}\rvert\rvert^2/5$	Prediction error
Correlation structure: Auto-regressive ($\rho = 0.3$)					
Raw	43.92	99.9	1.6	0.015	1.006
SMA	43.87	99.9	1.6	0.015	1.006
aLASSO-U	558.05	46.2	0	0.063	1.030
aLASSO-I	33.50	100.0	23.9	0.582	1.670
Correlation structure: Auto-regressive ($\rho = 0.6$)					
Raw	43.84	99.9	1.5	0.016	1.006
SMA	43.88	99.9	1.5	0.016	1.006
aLASSO-U	522.15	50.0	0	0.064	1.027
aLASSO-I	33.35	100.0	24.2	0.596	1.697
Correlation structure: Compound symmetry ($\rho = 0.3$)					
Raw	43.80	100.0	1.5	0.016	1.006
SMA	43.81	100.0	1.5	0.016	1.006
aLASSO-U	545.7	47.5	0	0.064	1.028
aLASSO-I	34.10	100.0	22.5	0.554	1.681
Correlation structure: Compound symmetry ($\rho = 0.5$)					
Raw	43.81	99.9	1.8	0.018	1.006
SMA	43.74	99.9	1.8	0.018	1.006
aLASSO-U	545.95	47.5	0	0.071	1.029
aLASSO-I	32.85	100.0	25.3	0.624	1.736

Note: The sample sizes range from 1100 to 1500.

Table S3. Comparison of the SMA with the individual aLASSO on the capture rate

Method	Capture Rate
$\theta_0 = 0.2$	
SMA	0.979
Individual aLASSO	0.737
$\theta_0 = 0.25$	
SMA	0.999
Individual aLASSO	0.921
$\theta_0 = 0.3$	
SMA	1
Individual aLASSO	0.984

Table S4. Influence of p_0 on SMA and the compared methods ($p = 100$)

Method	$\sum_{k=1}^{5} \lvert\hat{\mathcal{M}}^{(k)}\rvert$	Correct_0 (%)	Incorrect_0 (%)	$\lvert\lvert\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\rvert\rvert^2/5$	Prediction error
			$p_0 = 10$		
Raw	49.66	99.9	0.9	0.023	1.009
SMA	49.67	99.9	1.0	0.023	1.009
SMA-Diag	50.93	99.3	0.8	0.026	1.011
aLASSO-U	169.00	40.5	0	0.043	1.016
aLASSO-I	42.00	100.0	16.0	0.419	1.552
			$p_0 = 15$		
Raw	74.45	99.9	0.9	0.032	1.016
SMA	74.47	99.9	0.9	0.032	1.016
SMA-Diag	75.44	99.5	0.7	0.036	1.018
aLASSO-U	188.2	35.3	0	0.053	1.023
aLASSO-I	64.55	100.0	14.0	0.560	1.796
			$p_0 = 25$		
Raw	124.38	99.9	0.6	0.056	1.025
SMA	124.35	99.9	0.6	0.056	1.025
SMA-Diag	125.12	99.4	0.6	0.063	1.029
aLASSO-U	216.35	26.9	0	0.079	1.033
aLASSO-I	109.25	99.9	12.7	0.855	2.246

Table S5a. Performance of the SMA when unequal sets of covariates are reported by different studies but important covariates are not allowed to be missing ($p = 100$)

Method	$\sum_{k=1}^{5} \lvert\hat{\mathcal{M}}^{(k)}\rvert$	Correct_0 (%)	Incorrect_0 (%)	$\lvert\lvert\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\rvert\rvert^2/5$	Prediction error
			$p_0 = 10$		
SMA	49.97	99.9	0.9	0.018	1.013
SMA-Diag	50.10	99.9	1.0	0.019	1.014
aLASSO-U	272.75	50.5	0	0.041	1.024
aLASSO-I	43.2	100.0	13.6	0.372	1.536
			$p_0 = 15$		
SMA	74.66	100.0	0.7	0.026	1.016
SMA-Diag	74.95	99.9	0.6	0.029	1.017
aLASSO-U	298.65	47.4	0	0.052	1.028
aLASSO-I	65.45	100.0	12.8	0.513	1.745
			$p_0 = 25$		
SMA	124.64	99.9	0.7	0.044	1.020
SMA-Diag	124.59	99.9	0.7	0.048	1.022
aLASSO-U	334.20	44.2	0	0.070	1.032
aLASSO-I	109.05	100.0	12.8	0.818	2.087

Table S5b. Comparison of the SMA with other methods when important covariates are allowed to be missing ($p = 100$)

Method	$\sum_{k=1}^{5} \lvert\hat{\mathcal{M}}^{(k)}\rvert$	Correct_0 (%)	Incorrect_0 (%)	$\lvert\lvert\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\rvert\rvert^2/5$	Prediction error
			$p_0 = 10$		
SMA	45.46	99.8	11.0	0.374	1.166
SMA-Diag	46.50	99.6	10.9	0.376	1.167
aLASSO-U	285.2	47.7	0	0.406	1.178
aLASSO-I	24.3	100.0	51.6	1.617	2.535
			$p_0 = 15$		
SMA	68.94	99.6	10.1	0.489	1.220
SMA-Diag	70.39	99.3	9.9	0.490	1.220
aLASSO-U	313.35	43.9	0	0.526	1.233
aLASSO-I	39.55	100.0	47.3	2.189	3.331
			$p_0 = 25$		
SMA	113.66	99.3	11.2	0.897	1.390
SMA-Diag	117.68	98.3	10.9	0.902	1.391
aLASSO-U	358.45	37.7	0	0.946	1.403
aLASSO-I	61.10	100.0	51.2	3.915	5.120

Table S6. Comparison of the SMA with other methods when some covariates are unmeasured in all of the studies ($p = 50$)

Method	$\sum_{k=1}^{5} \lvert\hat{\mathcal{M}}^{(k)}\rvert$	Correct_0 (%)	Incorrect_0 (%)	$\lvert\lvert\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\rvert\rvert^2/5$	Prediction error
			$p_0 = 10$		
Raw	49.29	99.9	1.8	0.041	2.296
SMA	49.29	99.9	1.8	0.041	2.296
SMA-Diag	57.21	96.2	0.9	0.051	2.300
aLASSO-U	174.80	37.6	0	0.081	2.314
aLASSO-I	39.30	100.0	21.4	0.586	3.000
			$p_0 = 15$		
Raw	74.15	99.9	1.3	0.062	2.310
SMA	74.10	99.9	1.4	0.062	2.310
SMA-Diag	81.48	96.0	0.6	0.071	2.315
aLASSO-U	192.90	32.6	0	0.102	2.328
aLASSO-I	61.10	99.9	18.7	0.795	3.448
			$p_0 = 25$		
Raw	123.59	99.9	1.2	0.102	2.324
SMA	123.55	99.9	1.2	0.102	2.324
SMA-Diag	130.21	95.3	0.6	0.109	2.328
aLASSO-U	217.00	26.4	0	0.140	2.341
aLASSO-I	102.95	99.8	17.8	1.261	3.879

Note: Correct_0(%), Incorrect_0(%), and $\lvert\lvert\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\rvert\rvert^2/5$ are based on the measured covariates.

Table S7a. Influence of screening procedure on the performance of variable selection (under the homogeneous structure)

Method	$\sum_{k=1}^{5} \lvert\hat{\mathcal{M}}^{(k)}\rvert$	Correct_0 (%)	Incorrect_0 (%)	$\lvert\lvert\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\rvert\rvert^2/5$	Prediction error
			Screening Procedure I		
Raw	42.30	98.2	19.2	0.517	1.235
SMA	42.41	98.0	19.4	0.518	1.236
aLASSO-U	139.25	6.5	17.9	0.587	1.264
aLASSO-I	32.55	99.4	36.1	1.012	1.893
			Screening Procedure II		
Raw	49.92	99.3	1.6	0.049	1.024
SMA	50.07	99.2	1.6	0.049	1.024
aLASSO-U	134.4	19.6	0.1	0.100	1.044
aLASSO-I	39.85	99.9	20.5	0.592	1.738
			Screening Procedure III		
Raw	51.25	99.5	1.3	0.049	1.024
SMA	52.46	99.2	1.3	0.052	1.025
aLASSO-U	396.70	14.1	0.0	0.235	1.112
aLASSO-I	40.05	99.9	20.8	0.621	1.882

Note: The Correct_0 (%) and Incorrect_0 (%) were based on the dimensions after the screening procedure. The three screening procedures are described in Section B.7.

Table S7b. Influence of screening procedures on the performance of variable selection (under the heterogeneous structure)

Method	$\sum_{k=1}^{5} \lvert\hat{\mathcal{M}}^{(k)}\rvert$	Correct_0 (%)	Incorrect_0 (%)	$\lvert\lvert\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\rvert\rvert^2/5$	Prediction error
		Screening Procedure I			
Raw	34.33	97.2	24.0	0.483	1.217
SMA	34.67	96.9	24.1	0.483	1.218
aLASSO-U	137.10	8.7	19.4	0.533	1.238
aLASSO-I	22.00	99.6	47.6	1.167	1.973
		Screening Procedure II			
Raw	40.18	98.2	7.1	0.069	1.035
SMA	40.01	98.3	7.2	0.069	1.036
aLASSO-U	135.6	16.3	1.9	0.114	1.053
aLASSO-I	28.05	99.8	32.1	0.794	1.859
		Screening Procedure III			
Raw	46.73	98.2	5.7	0.065	1.033
SMA	48.78	97.8	5.4	0.069	1.035
aLASSO-U	435.30	12.9	0.4	0.249	1.119
aLASSO-I	28.55	99.8	32.4	0.814	1.927

Note: The Correct_0 (%) and Incorrect_0 (%) were based on the dimensions after the screening procedure. The three screening procedures are described in Section B.7.

Table S7c. Evaluation of screening procedures under the sign-change situation

Method	$\sum_{k=1}^{5}\lvert\hat{\mathcal{M}}^{(k)}\rvert$	Correct_0 (%)	Incorrect_0 (%)	$\lvert\lvert\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\rvert\rvert^2/5$	Prediction error
Screening Procedure I					
Raw	34.41	97.1	24.0	0.479	1.216
SMA	34.79	96.8	24.0	0.479	1.216
aLASSO-U	136.50	9.2	19.4	0.530	1.237
aLASSO-I	21.75	99.5	48.3	1.179	2.025
Screening Procedure II					
Raw	39.59	98.1	8.7	0.097	1.047
SMA	39.67	98.1	8.5	0.097	1.047
aLASSO-U	135.8	15.7	3.3	0.143	1.065
aLASSO-I	27.25	99.7	34.5	0.850	1.914
Screening Procedure III					
Raw	47.24	98.1	5.7	0.066	1.033
SMA	48.61	97.8	5.6	0.067	1.034
aLASSO-U	429.20	13.1	0.4	0.244	1.116
aLASSO-I	28.40	99.8	33.2	0.835	1.947

Note: The heterogeneous structure is considered. The Correct_0 (%) and Incorrect_0 (%) were based on the dimensions after the screening procedure. The three screening procedures are described in Section B.7.

Table S7d. Comparison of different methods when p is screened down to $n/(2\log(n))$

Method	$\sum_{k=1}^{5}\lvert\hat{\mathcal{M}}^{(k)}\rvert$	Correct_0 (%)	Incorrect_0 (%)	$\lvert\lvert\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\rvert\rvert^2/5$	Prediction error
Correlation structure: Auto-regressive ($\rho = 0.3$)					
Raw	57.26	99.1	1.0	0.059	1.034
SMA	68.20	97.7	1.0	0.079	1.044
aLASSO-U	811.45	7.4	0	0.548	1.274
aLASSO-I	43.15	99.6	20.9	0.633	2.048
Correlation structure: Auto-regressive ($\rho = 0.6$)					
Raw	57.89	98.9	1.3	0.066	1.038
SMA	67.54	97.8	0.8	0.083	1.046
aLASSO-U	778.05	8.8	0.1	0.535	1.261
aLASSO-I	40.45	99.7	23.9	0.735	2.152

Note: $p = 10,000$ and the sample sizes range from 400 to 600. The Correct_0 (%) and Incorrect_0 (%) were based on the dimensions after the SIS procedure.

Table S8. Sensitivity analysis ($p = 200$ under the homogeneous structure)

Method	$\sum_{k=1}^{5} \lvert\hat{\mathcal{M}}^{(k)}\rvert$	Correct_0 (%)	Incorrect_0 (%)	$\lvert\lvert\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\rvert\rvert^2/5$	Prediction error
Correlation structure: Compound symmetry ($\rho = 0.3$)					
Raw	49.84	100.0	0.8	0.012	1.068
SMA	49.85	100.0	0.8	0.012	1.068
SMA-Diag	51.00	99.9	0.7	0.015	1.070
aLASSO-U	526.7	49.8	0	0.044	1.083
aLASSO-I	43.70	100.0	12.6	0.338	1.515
Correlation structure: Compound-symmetry ($\rho = 0.5$)					
Raw	49.89	100.0	0.8	0.014	1.068
SMA	49.90	100.0	0.8	0.014	1.068
SMA-Diag	52.98	99.7	0.5	0.021	1.071
aLASSO-U	540.25	48.4	0	0.051	1.083
aLASSO-I	43.9	100.0	12.2	0.330	1.521

Table S9. Comparison of the SMA with other methods under smaller effects

Method	$\sum_{k=1}^{5} \lvert\hat{\mathcal{M}}^{(k)}\rvert$	Correct_0 (%)	Incorrect_0 (%)	$\lvert\lvert\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\rvert\rvert^2/5$	Prediction error
Correlation structure: Auto-regressive ($\rho = 0.6$)					
Raw	49.05	99.8	2.5	0.018	1.009
SMA	49.04	99.9	2.5	0.018	1.009
SMA-Diag	49.95	99.4	2.6	0.025	1.013
SMA-S	50.26	99.3	2.4	0.025	1.013
aLASSO-U	182.15	33.9	0	0.042	1.018
aLASSO-I	35.35	100.0	29.3	0.205	1.269
Correlation structure: Compound symmetry ($\rho = 0.5$)					
Raw	48.98	99.9	2.5	0.020	1.009
SMA	48.91	99.9	2.6	0.020	1.009
SMA-Diag	49.83	99.5	2.5	0.025	1.012
SMA-S	49.99	99.4	2.4	0.025	1.012
aLASSO-U	178.55	35.7	0	0.044	1.017
aLASSO-I	34.7	100.0	30.6	0.212	1.250

Table S10. The models selected by the Raw method for the 7 studies (i.e., FHS, MESA-A, MESA-E, CARDIA-A, CARDIA-E, CHS-A, and CHS-E)

Study	Model size	Selected SNPs
MESA-A	29	rs660240, rs6511720, rs964184, rs9534262, rs520861, rs6669785 rs4148817, rs11869143, rs262669, rs13911, rs4243233, rs7739248, rs6601431,...(remaining ones omitted)
CARDIA-A	37	rs445925, rs6511720, rs9302635, rs7254892, rs9534262, rs405509, rs520861, rs4803770, rs6902897, rs1470641, rs262669, rs312, ...(remaining ones omitted)
CHS-A	16	rs445925, rs520861, rs2360546, rs7961290, rs651028, rs2442623, rs9579604, rs2953168, rs7747551, rs759163, rs12439189, rs7335275, rs2304482, rs11207307, rs14868, rs3738702
MESA-E	33	rs660240, rs445925 , rs6511720 , rs2954021 , rs9302635 , rs520861, rs4803770, rs1367117, rs1544410 , rs4384362 , rs10983319, rs1470641 ...(remaining ones omitted)
CARDIA-E	32	rs4420638, rs660240 , rs445925 , rs6511720 , rs157582 , rs964184, rs2954021, rs9302635, rs520861, rs4803770, rs1367117, rs1513251, ...(remaining ones omitted)
CHS-E	26	rs4420638 , rs660240 , rs445925 , rs6511720 , rs964184 , rs2954021 , rs520861 , rs4803770, rs2867314 , rs1367117, rs3817588 , rs1470641 , rs7099478 , rs3846662 , rs1995775 , rs2859369 , rs12982656 , rs2442623 , rs12027341 , rs10037737 , rs17627064 ,rs8044769 , rs9321358 , rs9893040 , rs13214952 , rs701831
FHS	31	rs4420638, rs660240 , rs445925 , rs6511720 , rs157582 , rs964184 , rs2954021 , rs9302635 , rs9939224 , rs12357364, rs11869143, rs136335 , ...(remaining ones omitted)

Table S11. Prediction errors divided by the trait variance

	Raw	SMA	SMA-Diag	SMA-S	SMA-I	SMA-E	aLASSO-U	HT-U
Ratio	95.1%	95.4%	97.6%	95.8%	95.1%	95.1%	103.0%	100.9%

Note: The heterogeneous structure is adopted.

Table S12. Variable selection in seven studies (i.e., FHS, MESA-A, MESA-E, CARDIA-A, CARDIA-E, CHS-A, and CHS-E) under the homogeneous structure

	Model sizes							
Method	MESA-A	CARDIA-A	CHS-A	MESA-E	CARDIA-E	CHS-E	FHS	Pred-error
Raw	3	3	1	5	5	6	4	2.63
SMA	32	22	24	36	21	40	35	2.62
SMA-Diag	51	45	31	66	36	71	61	2.69
SMA-S	50	55	23	71	55	67	63	2.65
SMA-I	16	13	8	17	9	23	14	2.61
SMA-E	45	32	24	58	32	66	52	2.65
aLASSO-U	258	258	258	258	258	258	258	2.84
HT-U	241	241	241	241	241	241	241	2.68

Table S13. Prediction errors divided by the trait variance for the homogeneous structure

	Raw	SMA	SMA-Diag	SMA-S	SMA-I	SMA-E	aLASSO-U	HT-U
Ratio	95.4%	95.1%	97.6%	96.1%	94.7%	96.1%	103.0%	97.2%

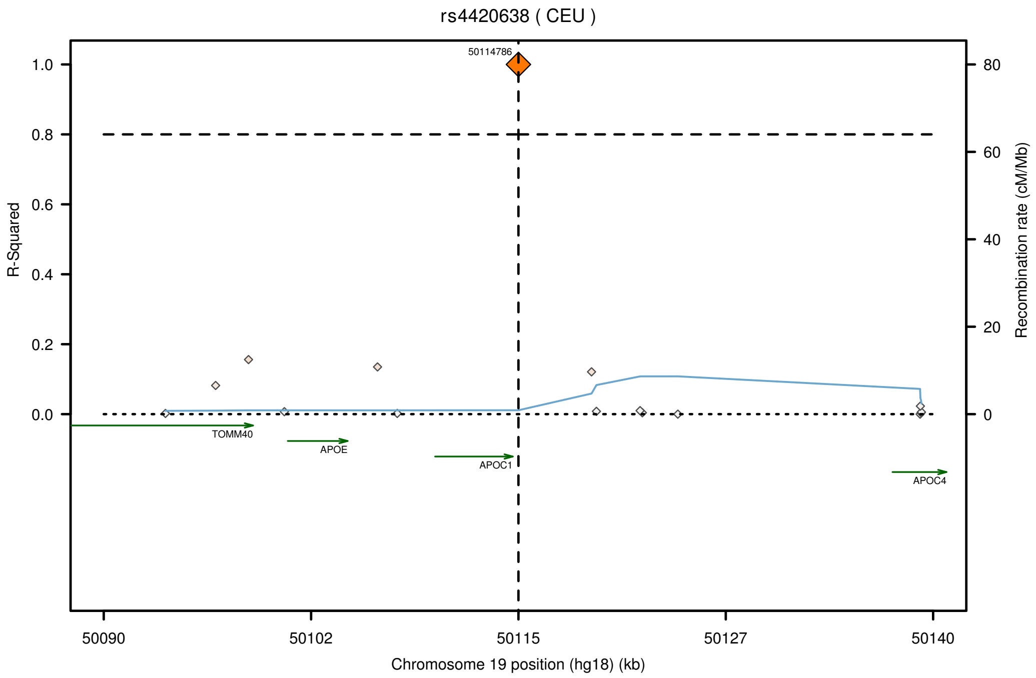

Fig. S1. Regional LD plot of SNP rs4420638 of the APOC1 gene

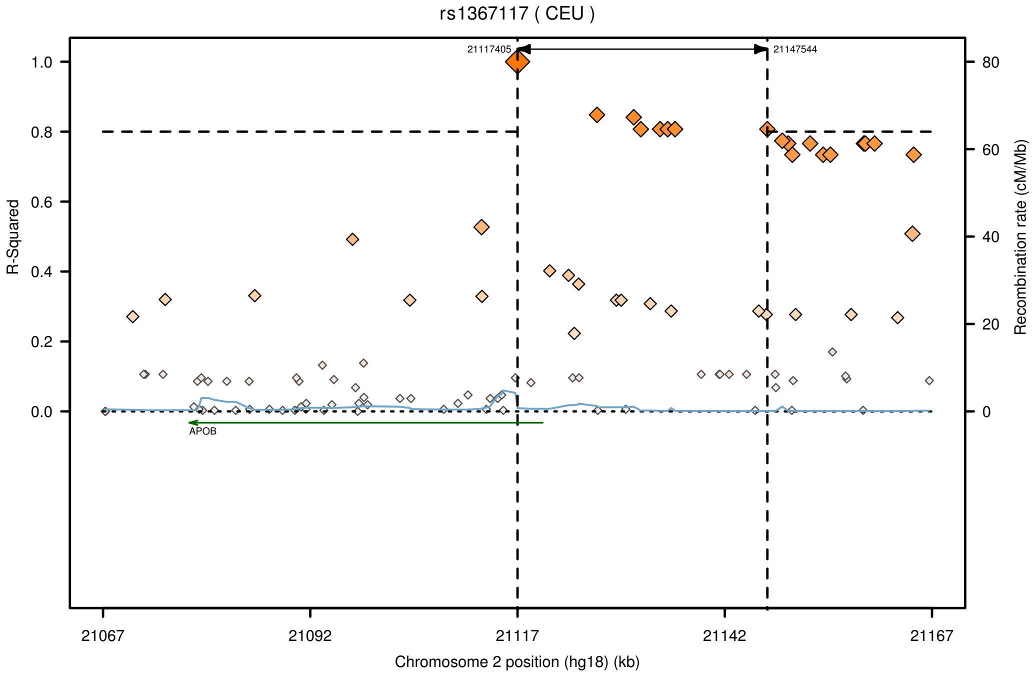

Fig. S2. Regional LD plot of SNP rs1367117 of the APOB gene