Biometrika (201X), **00**, 0, *pp.* 1–10
© 201X Biometrika Trust
Printed in Great Britain
Advance Access publication on XX XXXX 201X

Supplementary material for Partition-based ultrahigh-dimensional variable screening

BY JIAN KANG
Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, U.S.A
jiankang@umich.edu

HYOKYOUNG G. HONG
Department of Statistics and Probability, Michigan State University, East Lansing, Michigan 48823, U.S.A
hhong@msu.edu

AND YI LI
Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, U.S.A
yili@umich.edu

1. REGULARITY CONDITIONS FOR SURE SCREENING PROPERTIES

We discuss the conditions for the uniform convergence results at the sample level.

Condition A. Suppose the covariates can be partitioned into G disjoint groups with group index $g \in \{1, ..., G\}$. Denote by g_j the group membership for variable X_j.

A·1. Suppose that $\{\bar{\beta}^*_{g,0}, \bar{\beta}^*_g\}$ is an interior point of a sufficiently large, compact and convex set $\mathcal{D}_g = \{(\beta^*_{g,0}, \beta^*_g) : |\beta^*_{g,0} - \bar{\beta}^*_{g,0}| + \|\beta^*_g - \bar{\beta}^*_g\|_1 < D_g\}$, where $D_g > 0$.

A·2. The Fisher information is

$$I_g(\beta^*_{g,0}, \beta^*_g) = E\{b''(\beta^*_{g,0} + X_g^{*\mathrm{T}} \beta^*_g)(1, X_g^{*\mathrm{T}})^{\mathrm{T}}(1, X_g^{*\mathrm{T}})\},$$

and

$$\sup_{(\beta^*_{g,0}, \beta^*_g) \in \mathcal{D}_g, \|X^*_g\|=1} \|I_g(\beta^*_{g,0}, \beta^*_g)^{1/2}(1, X_g^{*\mathrm{T}})^{\mathrm{T}}\| < \infty.$$

A·3. There exist positive constants r_0, r_1, s_0, s_1 and α such that

$$\mathrm{pr}(|X_j| > t) \leq r_1 \exp(-r_0 t^\alpha)$$

for a sufficiently large t and that

$$E[\exp\{b(\beta_0 + X^{\mathrm{T}}\beta + s_0) - b(\beta_0 + X^{\mathrm{T}}\beta)\}] + E[\exp\{b(\beta_0 + X^{\mathrm{T}}\beta - s_0) - b(\beta_0 + X^{\mathrm{T}}\beta)\}] \leq s_1.$$

A·4. Suppose that $b''(\theta)$ is continuous and positive, as a function of θ.

A·5. For $g = 1, \ldots, G$, there exists a sequence $R_n > 0$ and we assume that

(a) there exists an $\epsilon_1 > 0$ such that

$$\sup_{\{\beta^*_{g,0}, \beta^*_g\} \in \mathcal{D}_g, \|(\beta^*_{g,0}, \beta^*_g) - (\bar{\beta}^*_{g,0}, \bar{\beta}^*_g)\| \leq \epsilon_1} |E\{b(\beta^*_{g,0} + X_g^{*\mathrm{T}}\beta^*_g) I[|X_j| > R_n]\}| \leq o\left(S_g n^{-1}\right),$$

for all j such that $g_j = g$;

(b) for a given $\{\beta^*_{g,0}, \beta^*_g\} \in \mathcal{D}_g$, the function $l(\beta^*_{g,0} + \xi_g^{\mathrm{T}} \beta^*_g, \eta)$ is Lipschitz with a positive constant $r_{g,n}$. That is, for all $(\xi_g^{\mathrm{T}}, \eta)^{\mathrm{T}}$ in $\Psi_{n,g} = \{(\xi_g, \eta) \, \|\xi_g\|_\infty \le R_n, |\eta| \le r_0 R_n^\alpha / s_0\}$, we have that

$$|l(\beta^*_{g,0} + \xi_g^{\mathrm{T}} \beta^*_g, \eta) - l(\beta^{*'}_{g,0} + \xi_g^{\mathrm{T}} \beta^{*'}_g, \eta)| \le r_{g,n} |\beta^*_{g,0} + \xi_g^{\mathrm{T}} \beta^*_g - (\beta^{*'}_{g,0} + \xi_g^{\mathrm{T}} \beta^{*'}_g)|,$$

for any $\{\beta^*_{g,0}, \beta^*_g\}, \{\beta^{*'}_{g,0}, \beta^{*'}_g\} \in \mathcal{D}_g$.

A·6. For $g = 1, \ldots, G$ and $\{\beta^*_{g,0}, \beta^*_g\} \in \mathcal{D}_g$,

$$E\{l(\beta^*_{g,0} + X_g^{*\mathrm{T}} \beta^*_g, Y) - l(\bar\beta^*_{g,0} + X_g^{*\mathrm{T}} \bar\beta^*_g, Y)\} \ge K_0\{\|\beta^*_g - \bar\beta^*_g\|^2 + (\beta^*_{g,0} - \bar\beta^*_{g,0})^2\},$$

for a positive K_0.

Of note, many generalised linear models, such as linear regression, logistic regression and Poisson regression, satisfy Conditions A·1–A·5 for any group partition. In particular, by taking $r_{g,n} = b'(R_n D_g S_g) + r_0 R_n^\alpha / s_0$, Condition A·5 holds for all models in the exponential family for any group partition. For logistic regression, $r_{g,n}$ is a finite constant. Additionally, Condition A·6 ensures model identifiability with group partitions. Similar conditions are also used for the theoretical development of sure and conditional sure independence screening (Fan & Song, 2010; Barut et al., 2016).

2. REGULARITY CONDITIONS FOR THE UPPER BOUND OF THE FALSE POSITIVE RATE

Condition B. Suppose the covariates can be partitioned into G disjoint groups with a group index $g \in \{1, ..., G\}$:

B·1. define $e_g = Y - b'(X_g^{*\mathrm{T}} \beta^*_g)$, for $g = 1, \ldots, G$. Assume that $\mathrm{var}(e_g) > c_6$ for $c_6 > 0$ and $\sup_{1 \le g \le G} E(|e_g|^{2+l}) < \infty$ for some $l > 0$;

B·2. for $j \in \mathcal{M}^c$, we have that $E_{b'}(Y \mid X^*_{-j}) = E_{b'}(Y \mid X^*_{-j}, X_j)$;

B·3. $n^{1/2} |\mathcal{M}| = o(p)$.

Condition B·1 includes mild assumptions to ensure the asymptotic normality of the proposed screening statistics. Similar conditions have been used by others (Barut et al., 2016; Heyde, 2008). By Theorem 1, Condition B·2 is equivalent to $\bar\beta_j = 0$ for all $j \in \mathcal{M}^c$. Condition B·3 does not allow the number of true signals to grow too fast as $n \to \infty$.

3. PROOF OF THEOREM 1

Proof. When X_j is eliminated in group g, (3) in § 3 becomes

$$(\tilde\beta^*_{g,0}, \tilde\beta^*_{-j}) = \underset{(\beta_{g,0}, \beta^*_{-j})}{\mathrm{argmax}}\, E\left\{ l\left(\beta_{g,0} + X_{-j}^{*\mathrm{T}} \beta^*_{-j}, Y\right)\right\},$$

which satisfies

$$E\{b'(\tilde\beta^*_{g,0} + X_{-j}^{*\mathrm{T}} \tilde\beta^*_{-j})(1, X_{-j}^{*\mathrm{T}})^{\mathrm{T}}\} = E\{Y(1, X_{-j}^{*\mathrm{T}})^{\mathrm{T}}\}. \tag{S1}$$

On the other hand, (4) in § 3 is equivalent to

$$E\{b'(\bar\beta^*_{g,0} + X_{-j}^{*\mathrm{T}} \bar\beta^*_{-j} + X_j \bar\beta_j)(1, X_{-j}^{*\mathrm{T}})^{\mathrm{T}}\} = E\{Y(1, X_{-j}^{*\mathrm{T}})^{\mathrm{T}}\},$$
$$E\{b'(\bar\beta^*_{g,0} + X_{-j}^{*\mathrm{T}} \bar\beta^*_{-j} + X_j \bar\beta_j) X_j\} = E\{Y X_j\}.$$

When $\bar\beta_j = 0$, $E\{b'(\bar\beta^*_{g,0} + X^{*\mathrm{T}}_{-j}\bar\beta^*_{-j} + X_j\bar\beta_j)\} = E\{b'(\bar\beta^*_{g,0} + X^{*\mathrm{T}}_{-j}\bar\beta^*_{-j})(1, X^{*\mathrm{T}}_{-j})^{\mathrm{T}}\} = E\{Y(1, X^{*\mathrm{T}}_{-j})^{\mathrm{T}}\}$, implying that $(\bar\beta^*_{g,0}, \bar\beta^*_{-j})$ is a solution to (S1).

Under the assumption that the solution to (S1) is unique, we have

$$\bar\beta^*_{g,0} = \tilde\beta^*_{g,0}, \quad \bar\beta^*_g = \tilde\beta^*_g.$$

Using Definition 1 of $E_{b'}(\cdot \mid \cdot)$ completes the proof for the necessary part.

Now, we prove the sufficient condition. First,

$$E_{b'}(Y \mid X^*_{-j}) = E_{b'}(Y \mid X^*_{-j}, X_j)$$

implies that

$$b'(\tilde\beta^*_{g,0} + X^{*\mathrm{T}}_{-j}\tilde\beta^*_{-j} + X_j \times 0) = b'(\bar\beta^*_{g,0} + X^{*\mathrm{T}}_{-j}\bar\beta^*_{-j} + X_j\bar\beta_j).$$

By Definition 1, this further implies that $(\tilde\beta^*_{g,0}, \tilde\beta^{\mathrm{T}*}_{-j}, 0)^{\mathrm{T}}$ is a solution of (4). By the uniqueness of solutions to (4), and given that $(\bar\beta^*_{g,0}, \bar\beta^{\mathrm{T}*}_{-j}, \bar\beta_j)^{\mathrm{T}}$ is also a solution to (4), we have $\bar\beta_j = 0$.

4. Proof of Theorem 2

Proof. For each j, let $g = g_j$ and $\Omega_j = E\{\delta_j(1, X^{*\mathrm{T}}_{-j}, X_j)^{\mathrm{T}}(1, X^{*\mathrm{T}}_{-j}, X_j)\}$, where

$$\delta_j = \frac{b'(\bar\beta^*_{g,0} + X^{*\mathrm{T}}_{-j}\bar\beta^*_{-j} + X_j\bar\beta_j) - b'(\tilde\beta^*_{g,0} + X^{*\mathrm{T}}_{-j}\tilde\beta^*_{-j})}{(\bar\beta^*_{g,0} + X^{*\mathrm{T}}_{-j}\bar\beta^*_{-j} + X_j\bar\beta_j) - (\tilde\beta^*_{g,0} + X^{*\mathrm{T}}_{-j}\tilde\beta^*_{-j})}.$$

Because $b''(\cdot)$ is positive, $\delta_j > 0$.

From (4) in § 3 and (S1),

$$E\{b'(\bar\beta^*_{g,0} + X^{*\mathrm{T}}_{-j}\bar\beta^*_{-j} + X_j\bar\beta_j)(1, X^{*\mathrm{T}}_{-j})^{\mathrm{T}}\} = E\{b'(\tilde\beta^*_{g,0} + X^{*\mathrm{T}}_{-j}\tilde\beta^*_{-j})(1, X^{*\mathrm{T}}_{-j})^{\mathrm{T}}\} = E\{Y(1, X^{*\mathrm{T}}_{-j})^{\mathrm{T}}\}.$$

Let $\check\beta^*_{-j} = \bar\beta^*_{-j} - \tilde\beta^*_{-j}$ and $\check\beta^*_{g,0} = \bar\beta^*_{g,0} - \tilde\beta^*_{g,0}$. Recall the definition of δ_j. We have

$$E\{\delta_j(\check\beta^*_{g,0} + X^{*\mathrm{T}}_{-j}\check\beta^*_{-j} + X_j\bar\beta_j)(1, X^{*\mathrm{T}}_{-j})^{\mathrm{T}}\} = 0. \tag{S2}$$

We can partition Ω_j as

$$\Omega_j = \begin{bmatrix} E\{\delta_j(1, X^{*\mathrm{T}}_{-j})^{\mathrm{T}}(1, X^{*\mathrm{T}}_{-j})\} & E\{\delta_j(1, X^{*\mathrm{T}}_{-j})^{\mathrm{T}}X_j\} \\ E\{\delta_j X_j(1, X^{*\mathrm{T}}_{-j})\} & E(\delta_j X_j^2) \end{bmatrix} = \begin{pmatrix} \Omega^{**}_j & \Omega^*_j \\ \Omega^{*\mathrm{T}}_j & \Omega_{j,j} \end{pmatrix}.$$

Solving (S2),

$$(\check\beta^*_{g,0}, \check\beta^{*\mathrm{T}}_{-j})^{\mathrm{T}} = -(\Omega^{**}_j)^{-1}\Omega^*_j\bar\beta_j.$$

By Definition 1 and the definition of δ_j, we have

$$\begin{aligned} E[X_j\{E_{b'}(Y \mid X^*_{-j}, X_j) - E_{b'}(Y \mid X^*_{-j})\}] &= E[X_j\{\delta_j(\check\beta^*_{g,0} + X^{*\mathrm{T}}_{-j}\check\beta^*_{-j} + X_j\bar\beta_j)\}] \\ &= \{\Omega_{j,j} - \Omega^{*\mathrm{T}}_j(\Omega^{**}_j)^{-1}\Omega^*_j\}\bar\beta_j. \end{aligned}$$

To bound $\Omega_{j,j} - \Omega^{*\mathrm{T}}_j(\Omega^{**}_j)^{-1}\Omega^*_j$, we first apply a blockwise Cholesky decomposition and obtain

$$\begin{Bmatrix} I_{S_g} & 0 \\ -\Omega^{*\mathrm{T}}_j(\Omega^{**}_j)^{-1} & 1 \end{Bmatrix} \Omega_j \begin{Bmatrix} I_{S_g} & -(\Omega^{**}_j)^{-1}\Omega^*_j \\ 0 & 1 \end{Bmatrix} = \begin{Bmatrix} \Omega^{**}_j & 0 \\ 0 & \Omega_{j,j} - \Omega^{*\mathrm{T}}_j(\Omega^{**}_j)^{-1}\Omega^*_j \end{Bmatrix}, \tag{S3}$$

where I_{S_g} is an identity matrix with dimension S_g. Since Ω_j is positive definite by assumption, we have $\Omega_{j,j} - \Omega_j^{*\mathrm{T}}(\Omega_j^{**})^{-1}\Omega_j^* > 0$.

Furthermore, by Condition 2 and the monotonicity of function $b'(\cdot)$, it follows that $0 < \delta_j \leq L$, where L is the Lipschitz constant. Hence, $0 < \Omega_{j,j} \leq LE(X_j^2) \leq c_1$ for a positive constant c_1. Furthermore, because Ω_j^{**} is semi-positive definite, we have that $\Omega_j^{*\mathrm{T}}(\Omega_j^{**})^{-1}\Omega_j^* \geq 0$. Therefore, $0 < \Omega_{j,j} - \Omega_j^{*\mathrm{T}}(\Omega_j^{**})^{-1}\Omega_j^* \leq \Omega_{j,j} \leq c_1$.

By Condition 1, we have

$$|\bar{\beta}_j| \geq c_1^{-1}|E[X_j\{E_{b'}(Y \mid X_{-j}^*, X_j) - E_{b'}(Y \mid X_{-j}^*)\}]| \geq c_2 n^{-\kappa},$$

where $c_2 = c_0/c_1$ for all $j \in \mathcal{M}$, which completes the proof.

5. PROOF OF THEOREM 3

Proof. By Lemma 1 of Fan & Song (2010) and Condition A·3, for any $t > 0$,

$$\mathrm{pr}(|Y| > t) \leq s_1 \exp(-s_0 t).$$

Furthermore, for each $g = 1, \ldots, G$, we have

$$\begin{aligned}\mathrm{pr}(\Psi_{g,n}^c) &\leq \mathrm{pr}(\|X_g^*\|_\infty > R_n) + \mathrm{pr}(|Y| > r_0 R_n^\alpha / s_0)\\ &\leq \sum_{j:g_j=g} \mathrm{pr}(|X_j| > R_n) + s_1 \exp\left(-r_0 R_n^\alpha\right)\\ &\leq \left(S_g r_1 + s_1\right) \exp\left(-r_0 R_n^\alpha\right),\end{aligned}$$

where $R_n \to \infty$ as $n \to \infty$.

By taking $V_n = K_0$, $k_n = r_{g,n} = b'(R_n D_g S_g) + r_0 R_n^\alpha / s_0$ and $t = c_2 V_n n^{1/2-\kappa}/(8k_n) - 1$ in Theorem 1 of Fan & Song (2010), for each j, we have that

$$\begin{aligned}\mathrm{pr}\left(|\hat{\beta}_j - \bar{\beta}_j| \geq c_2 n^{-k}/2\right) &\leq \mathrm{pr}\left(\|\hat{\beta}_{g_j}^* - \bar{\beta}_{g_j}^*\| \geq c_2 n^{-\kappa}/2\right)\\ &\leq \exp(-c_3 Q_{g_j,n}) + n\mathrm{pr}(\Psi_{g_j,n}^c),\end{aligned} \tag{S4}$$

for a positive constant c_3 and $Q_{g,n} = n^{1-2\kappa}(r_{g,n} R_n)^{-2}$. Then we have that

$$\begin{aligned}\mathrm{pr}\left(\max_{1\leq j\leq p} |\hat{\beta}_j - \bar{\beta}_j| \geq c_2 n^{-\kappa}/2\right) &\leq \sum_{j=1}^p \exp(-c_3 Q_{g_j,n}) + \sum_{j=1}^p n\mathrm{pr}(\Psi_{g_j,n}^c)\\ &= \sum_{g=1}^G S_g \exp(-c_3 Q_{g,n}) + \sum_{g=1}^G S_g(S_g r_1 + s_1) n \exp(-r_0 R_n^\alpha).\end{aligned}$$

This completes the proof for part (a). For part (b), we consider the event

$$\mathcal{E}_n = \left\{\max_{j\in\mathcal{M}} |\hat{\beta}_j - \bar{\beta}_j| \leq c_2 n^{-\kappa}/2\right\}.$$

By Theorem 2,

$$\min_{j\in\mathcal{M}} |\bar{\beta}_j| \geq c_2 n^{-\kappa}.$$

Thus, if event $\mathcal{E}_n$ happens, it holds that for all $j \in \mathcal{M}$, $|\hat{\beta}_j| \geq c_2 n^{-\kappa}/2$.

By taking $c_4 = c_2/2$ and $\gamma = c_4 n^{-\kappa}$, it follows immediately that $\mathcal{E}_n \subset \{\mathcal{M} \subset \hat{\mathcal{M}}_\gamma\}$. This implies that

$$\mathrm{pr}(\mathcal{M} \subset \hat{\mathcal{M}}_\gamma) \geq \mathrm{pr}(\mathcal{E}_n) = 1 - \mathrm{pr}\left(\max_{j \in \mathcal{M}} |\hat{\beta}_j - \bar{\beta}_j| > c_2 n^{-\kappa}/2\right).$$

Applying (S4) and Bonferroni's inequality over all $j \in \mathcal{M}$ completes the proof.

6. Proof of Theorem 4

Proof. By Condition 5, we can bound the determinant of the Schur complement (von Neumann & Goldstine, 1947) of block Ω_j^* in Ω_j as follows:

$$|\Omega_{j,j} - \Omega_j^{*\mathrm{T}}(\Omega_j^{**})^{-1}\Omega_j^*| \geq \lambda_{\min}(\Omega_j) > K_1,$$

for $j = 1, \ldots, p$.

By (2) and the Lipschitz continuity of $b'(\theta)$,

$$\begin{aligned}
|\bar{\beta}_j| &< M_0 |E[X_j\{E_{b'}(Y \mid X^{\mathrm{T}}) - E_{b'}(Y \mid X_{-j}^{*\mathrm{T}})\}]| \\
&\leq M_0 E|X_j\{b'(\beta_0 + X^{\mathrm{T}}\beta) - b'(\tilde{\beta}_{g_j,0} + X_{-j}^{*\mathrm{T}}\tilde{\beta}_{-j}^*)\}| \\
&\leq M_1 E|X_j\{\beta_0 + X^{\mathrm{T}}\beta - \tilde{\beta}_{g_j,0} - X_{-j}^{*\mathrm{T}}\tilde{\beta}_{-j}^*\}| \\
&= M_1 E\left|X_j\left\{\beta_0 - \tilde{\beta}_{g_j,0} + X_{-j}^{*\mathrm{T}}(\beta_{-j}^* - \tilde{\beta}_{-j}^*) + \sum_{g' \neq g_j} X_{g'}^{*\mathrm{T}}\beta_{g'}^* + X_j\beta_j\right\}\right|,
\end{aligned}$$

where $M_0 = K_1^{-1}$, $M_1 = M_0 L$ and $(\beta_0, \beta^{\mathrm{T}})^{\mathrm{T}}$ are the true parameters in model (1) in § 2. Simplifying the above inequality, we have

$$|\bar{\beta}_j| < M_1 E\left|X_j(X_j\beta_j + \breve{X}_{-j}^{\mathrm{T}}\breve{\beta}_{-j}^{\delta} + \breve{X}_{-g_j}^{*\mathrm{T}}\breve{\beta}_{-g_j})\right|, \tag{S5}$$

where

$$\breve{X}_{-j} = (1, X_{-j}^{*\mathrm{T}})^{\mathrm{T}}, \quad \breve{\beta}_{-j}^{\delta} = (\beta_0 - \beta_{g_j,0}, \beta_{-j}^{*\mathrm{T}} - \tilde{\beta}_{-j}^{*\mathrm{T}})^{\mathrm{T}},$$

$$\breve{X}_{-g}^* = (X_1^{*\mathrm{T}}, \ldots, X_{g-1}^{*\mathrm{T}}, X_{g+1}^{*\mathrm{T}}, \ldots, X_G^{*\mathrm{T}})^{\mathrm{T}},$$

$$\breve{\beta}_{-g}^* = (\beta_1^{*\mathrm{T}}, \ldots, \beta_{g-1}^{*\mathrm{T}}, \beta_{g+1}^{*\mathrm{T}}, \ldots, \beta_G^{*\mathrm{T}})^{\mathrm{T}}.$$

By the properties of the generalised linear conditional expectation,

$$\begin{aligned}
&E\{E_1(X_j \mid X_{-j}^*, X_{g'}^*, g' \neq g_j)X_{-j}^{*\mathrm{T}}\} = E(X_j X_{-j}^{*\mathrm{T}}), \\
&E\{E_1(X_j \mid X_{-j}^*, X_{g'}^*, g' \neq g_j)X_g^{*\mathrm{T}}\} = E(X_j X_g^{*\mathrm{T}}),\ g \neq g_j.
\end{aligned}$$

Furthermore,

$$E\{E_1(X_j \mid X_{-j}^*, X_{g'}^*, g' \neq g_j)(\breve{X}_{-j}^{\mathrm{T}}\breve{\beta}_{-j}^{\delta} + \breve{X}_{-g_j}^{*\mathrm{T}}\breve{\beta}_{-g_j})\} = E\{X_j(\breve{X}_{-j}^{\mathrm{T}}\breve{\beta}_{-j}^{\delta} + \breve{X}_{-g_j}^{*\mathrm{T}}\breve{\beta}_{-g_j})\},$$

and

$$E\{E_1(X_j \mid X_{-j}^*, X_{g'}^*, g' \neq g_j)E_1(X_j \mid X_{-j}^*, X_{g'}^*, g' \neq g_j)\} = E\{X_j E_1(X_j \mid X_{-j}^*, X_{g'}^*, g' \neq g_j)\}.$$

Let $V_j = E\left[\{X_j - E_1(X_j \mid X^*_{-j}, X^*_{g'}, g' \neq g_j)\}^2\right]$. Then

$$\begin{aligned} V_j &= E\left[\{X_j - E_1(X_j \mid X^*_{-j}, X^*_{g'}, g' \neq g_j)\}\{X_j - E_1\{X_j \mid X^*_{-j}, X^*_{g'}, g' \neq g_j)\}\right] \\ &= E(X_j^2) - E\{E_1(X_j \mid X^*_{-j}, X^*_{g'}, g' \neq g_j)X_j\}. \end{aligned}$$

This further implies that

$$\begin{aligned} V_j\beta_j + U_j =& E(X_j^2\beta_j) - E\left\{E_1\left(X_j \mid X^*_{-j}, X^*_{g'}, g' \neq g_j\right)X_j\right\}\beta_j \\ &+ E\left\{E_1\left(X_j \mid X^*_{-j}, X^*_{g'}, g' \neq g_j\right)\left(\beta_0 - \tilde{\beta}_{g_j,0} + X^{\mathrm{T}}\beta - X^{*\mathrm{T}}_{-j}\tilde{\beta}^*_{-j}\right)\right\} \\ =& E(X_j^2\beta_j) + E\left[E_1\left(X_j \mid X^*_{-j}, X^*_{g'}, g' \neq g_j\right)\left\{\beta_0 - \tilde{\beta}_{g_j,0} + X^{*\mathrm{T}}_{-j}(\beta^*_{-j} - \tilde{\beta}^*_{-j}) + \sum_{g' \neq g_j} X^{*\mathrm{T}}_{g'}\beta^*_{g'}\right\}\right] \\ =& E(X_j^2\beta_j) + E\left\{E_1\left(X_j \mid X^*_{-j}, X^*_{g'}, g' \neq g_j\right)\left(\check{X}^{\mathrm{T}}_{-j}\check{\beta}^{\delta}_{-j} + \check{X}^{*\mathrm{T}}_{-g_j}\check{\beta}_{-g_j}\right)\right\} \\ =& E\left\{X_j\left(X_j\beta_j + \check{X}^{\mathrm{T}}_{-j}\check{\beta}^{\delta}_{-j} + \check{X}^{*\mathrm{T}}_{-g_j}\check{\beta}_{-g_j}\right)\right\}. \end{aligned}$$

Thus, we can write (S5) in a vector form:

$$\|\bar{\beta}\|^2 \leq M_1^2\,\|\Sigma\beta + U\|^2\,, \tag{S6}$$

where $\bar{\beta} = (\bar{\beta}_1, \ldots, \bar{\beta}_p)^{\mathrm{T}}$ and $\Sigma = \mathrm{diag}(V_1, \ldots, V_p)$.

Furthermore, $V_j = E(X_j^2) - \{E(X_j)\}^2 + \{E(X_j)\}^2 - E\{E_1(X_j \mid X^*_{-j}, X^*_{g'}, g' \neq g_j)^2\} \leq \mathrm{var}(X_j)$ and

$$\|\Sigma\beta\|_2^2 \leq V\beta^{\mathrm{T}}\Sigma\beta \leq V\sum_{j=1}^{p}\mathrm{var}(X_j\beta_j) = O(V).$$

The right-hand side of (S6) can be further bounded as follows:

$$\begin{aligned} \|\Sigma\beta + U\|_2^2 &= \beta^{\mathrm{T}}\Sigma^2\beta + 2U^{\mathrm{T}}\Sigma\beta + U^{\mathrm{T}}\mathrm{U} \\ &\leq V\beta^{\mathrm{T}}\Sigma\beta + 2U^{\mathrm{T}}\Sigma\beta + U^{\mathrm{T}}U \\ &\leq V\sum_{j=1}^{p}\mathrm{var}(X_j\beta_j) + 2U^{\mathrm{T}}\Sigma\beta + U^{\mathrm{T}}U, \end{aligned}$$

where the last two terms are $o(V)$ based on Condition 6. Hence,

$$\|\bar{\beta}\|^2 = O(V).$$

This further implies that the number of j's such that $|\bar{\beta}_j| > \gamma/2 = c_4n^{-\kappa}/2$ cannot exceed $O(n^{2\kappa}V)$. That is, letting $\bar{\mathcal{M}}_{\gamma/2} = \{j : |\bar{\beta}_j| > c_4n^{-\kappa}/2\}$, we have that $\bar{\mathcal{M}}_{\gamma/2} = O(n^{2\kappa}V)$. Consider

$$\mathcal{F}_n = \left\{\max_{1 \leq j \leq p}|\hat{\beta}_j - \bar{\beta}_j| \leq c_4n^{-\kappa}/2\right\}.$$

Then on the event $\mathcal{F}_n$, $\hat{\mathcal{M}}_\gamma = \left\{j : |\hat{\beta}_j| > c_4n^{-\kappa}\right\}$ is a subset of $\bar{\mathcal{M}}_{\gamma/2}$. More precisely, we have

$$\left\{|\hat{\mathcal{M}}_\gamma| \leq |\bar{\mathcal{M}}_{\gamma/2}|\right\} \supseteq \left\{\hat{\mathcal{M}}_\gamma \subset \bar{\mathcal{M}}_{\gamma/2}\right\} \supseteq \left\{\mathcal{F}_n \cap \{\hat{\mathcal{M}}_\gamma \subset \bar{\mathcal{M}}_{\gamma/2}\}\right\} = \mathcal{F}_n.$$

Thus

$$\mathrm{pr}\left\{|\hat{\mathcal{M}}_\gamma| \leq O(n^{2\kappa}V)\right\} \geq \mathrm{pr}(\mathcal{F}_n).$$

Applying Theorem 3 part (a) and Bonferroni's inequality completes the proof.

7. PROOF OF THEOREM 6

Proof. For partition $\mathcal{G}^{(k)}(k=1,\ldots,K)$, let $c_3^{(k)}$, $c_4^{(k)}$, $r_0^{(k)}$, $r_2^{(k)}$, $r_3^{(k)}$, $\gamma^{(k)}$, $Q_{g_j,n}^{(k)}$, $S_g^{(k)}$, $R_n^{(k)\alpha}$ and $\hat{\mathcal{M}}_\gamma^{(k)}$ be the corresponding terms for $c_3, c_4, r_0, r_2, r_3, \gamma, Q_{g_j,n}, S_g, R_n^\alpha$ and $\hat{\mathcal{M}}_\gamma$ respectively in Theorems 3 and 4. Since $\mathcal{G}^{(k)}$ satisfies Conditions 1–3,

$$\mathrm{pr}\left(\mathcal{M} \subset \hat{\mathcal{M}}_\gamma^{(k)}\right) \geq 1 - \sum_{j\in\mathcal{M}} \exp\left(-c_3^{(k)} Q_{g_j,n}^{(k)}\right) - nr_3^{(k)} \exp\left(-r_0^{(k)} R_n^{(k)\alpha}\right).$$

Take $c_5 = \min_{1\leq l\leq K} c_4^{(l)} \leq c_4^{(k)}$. Then $\hat{\mathcal{M}}_\gamma^{(k)} \subset \tilde{\mathcal{M}}_\gamma$ and

$$\mathrm{pr}\left(\mathcal{M} \subset \tilde{\mathcal{M}}_\gamma\right) \geq \mathrm{pr}\left(\mathcal{M} \subset \hat{\mathcal{M}}_\gamma^{(k)}\right) \geq 1 - \sum_{j\in\mathcal{M}} \exp\left(-c_3^{(k)} Q_{g_j,n}^{(k)}\right) - nr_3^{(k)} \exp\left(-r_0^{(k)} R_n^{(k)\alpha}\right).$$

Taking $n \to \infty$ on both sides completes the proof for part (a).

For part (b), since all $\mathcal{G}^{(k)}$ satisfy Conditions A·1–A·5 and Conditions 4–6, by Theorem 4 we have that

$$\mathrm{pr}\left\{|\hat{\mathcal{M}}_\gamma^{(k)}| \leq O\left(n^{2\kappa}V^{(k)}\right)\right\} \geq 1 - \sum_{g=1}^{G^{(k)}} S_g^{(k)} \exp\left(-c_3^{(k)} Q_{g,n}^{(k)}\right) - nr_2^{(k)} \exp\left(-r_0^{(k)} R_n^{(k)\alpha}\right).$$

Since $|\tilde{\mathcal{M}}_\gamma| \leq \sum_{k=1}^K |\hat{\mathcal{M}}_\gamma^{(k)}|$,

$$\begin{aligned}\left\{\left|\tilde{\mathcal{M}}_\gamma\right| > O\left(n^{2\kappa}\sum_{k=1}^K V^{(k)}\right)\right\} &\subset \left\{\sum_{k=1}^K |\hat{\mathcal{M}}_\gamma^{(k)}| > O\left(\sum_{k=1}^K n^{2\kappa}V^{(k)}\right)\right\} \\ &\subset \bigcup_{k=1}^K \left\{|\hat{\mathcal{M}}_\gamma^{(k)}| > O\left(n^{2\kappa}V^{(k)}\right)\right\}.\end{aligned}$$

This further implies that

$$1 - \mathrm{pr}\left\{\left|\tilde{\mathcal{M}}_\gamma\right| > O\left(n^{2\kappa}\sum_{k=1}^K V^{(k)}\right)\right\} \geq 1 - \sum_{k=1}^K \mathrm{pr}\left\{\left|\hat{\mathcal{M}}_\gamma^{(k)}\right| > O\left(n^{2\kappa}V^{(k)}\right)\right\}.$$

Taking $n \to \infty$ on both sides completes the proof for part (b).

8. PROOF OF THEOREM 7

Proof. First consider

$$E\left(\left|\hat{M}_\tau^{\#} \cap \mathcal{M}^c\right|\right) = \sum_{j\in\mathcal{M}^c} \mathrm{pr}\left\{I_j(\hat{\beta}_j)^{1/2}|\hat{\beta}_j| \geq \tau\right\}.$$

By Condition B·2 and Theorem 1, we have $\bar{\beta}_j = 0$. By Condition B·1 and the theory of quasi-likelihood (Heyde, 2008), we have

$$I_j(\hat{\beta}_j)^{1/2}|\hat{\beta}_j| \sim N(0,1)$$

and according to the Berry–Esseen inequality (Korolev & Shevtsova, 2010), there exists a constant $\tilde{c}_7 > 0$ such that

$$\sup_z \left| \mathrm{pr}\left\{ I_j(\hat{\beta}_j)^{1/2}|\hat{\beta}_j| > z \right\} - \Phi(z) \right| \le \tilde{c}_7 n^{-1/2}.$$

Let $\tau = \Phi^{-1}\{1 - q/(2p)\}$, then we have

$$E\left(\left| \hat{M}_\tau^{\#} \cap \mathcal{M}^c \right| \right) \le \sum_{j \in \mathcal{M}^c} \left[2\{1 - \Phi(\tau)\} + \tilde{c}_7 n^{-1/2} \right] = (p - |\mathcal{M}|) \left(q/p + \tilde{c}_7 n^{-1/2} \right).$$

Because $\left| \hat{M}_\gamma \cap \mathcal{M}^c \right| \le \left| \hat{\mathcal{M}}_\gamma \right| = \left| \hat{M}_\tau^{\#} \right| = \left| \hat{M}_\tau^{\#} \cap \mathcal{M}^c \right| + \left| \hat{M}_\tau^{\#} \cap \mathcal{M} \right| \le \left| \hat{M}_\tau^{\#} \cap \mathcal{M}^c \right| + |\mathcal{M}|$, we have

$$\mathrm{EFPR}_\gamma = E\left(\frac{|\hat{M}_\gamma \cap \mathcal{M}^c|}{|\mathcal{M}^c|} \right) \le q/p + \tilde{c}_7 n^{-1/2} + \frac{|\mathcal{M}|}{p - |\mathcal{M}|}.$$

By Condition B·3, there exist $c_7 > 0$ and $N_7 > 0$ such that for any $n > N_7$,

$$\mathrm{EFPR}_\gamma \le q/p + c_7 n^{-1/2}.$$

9. COVARIATE PROJECTION VIA PRINCIPAL COMPONENT ANALYSIS

To make Condition 6 less restrictive, this section presents a method to generate surrogate covariates with a reduced-correlation structure for the screening procedures. Specifically, we decompose X, a vector of p continuous covariates, based on principal component analysis and project it to a set of variables with the largest loadings on the leading eigenvectors. We write

$$X = \Pi Z + \breve{X},$$

where Π, a $p \times Q$ matrix, is formed by selecting the first Q leading eigenvectors of the sample covariance matrix of X, Z is a vector of Q projection loadings that can be obtained by ordinary least squares estimates given Π and X, and $\breve{X}$ is a vector of p residuals. The number of components Q can be chosen according to the proportion of variation explained by principal component analysis. When Q is large, the correlation among $\breve{X}$ can be much less than the correlation among X. However, we also need to keep Q small to ensure that the variation in $\breve{X}$ is relatively large and the variance of the screening statistics is small. In practice, we suggest choosing Q such that 50% of the variance of the covariate X is explained by $\breve{X}$. Then with adjusting for Z we compute the partition-based screening statistics $\breve{\beta}_g^* = (\breve{\beta}_j, g_j = g)$ based on $\breve{X}$ as follows:

$$(\breve{\beta}_{g,0}^*, \breve{\beta}_{g,Z}^*, \breve{\beta}_g^*) = \underset{(\beta_{g,0}, \beta_{g,Z}, \beta_g^*)}{\mathrm{argmax}} E_n l \left(\beta_{g,0} + Z^{\mathrm{T}} \beta_{g,Z} + \breve{X}_g^{*\mathrm{T}} \beta_g^*, Y \right),$$

where the definition of $\breve{X}_g^*$ is the same as X_g^* in § 2 except that we replace X with $\breve{X}$. For a given threshold γ, the selected index set is $\breve{\mathcal{M}} = \{j : |\breve{\beta}_j| > \gamma\}$. We refer to this approach as partition-based screening with covariate projection.

We also perform simulation studies to compare this approach with the original partition-based screening approach. We generate data based on settings 1 and 4 in § 5 of the main text by slightly

Table S1. *Model selection accuracy of partition-based screening with and without covariate projections for both linear regression and logistic regression under Settings 1* and 4* with high correlation among covariates*

	Linear Regression					
Setting 1^*	Original covariates			Projected covariates		
	PIT (%)	MMS	TPR (%)	PIT (%)	MMS	TPR (%)
PartS ($\mathcal{G}^{\mathrm{red}}$)	70	137	100	85	72	100
PartS ($\mathcal{G}^{\mathrm{mis}_1}$)	52	192	100	48	238	83
PartS ($\mathcal{G}^{\mathrm{mis}_2}$)	47	224	83	37	263	83
CombPartS (5 $\mathcal{G}^{\mathrm{mis}_2}$)	63	124	100	59	146	100
CombPartS (10 $\mathcal{G}^{\mathrm{mis}_2}$)	61	153	100	62	1250	100
CorrPartS ($\mathcal{G}^{\mathrm{cor}}$)	29	362	83	34	348	83
	Logistic Regression					
	Original covariates			Projected covariates		
	PIT (%)	MMS	TPR (%)	PIT (%)	MMS	TPR (%)
PartS ($\mathcal{G}^{\mathrm{red}}$)	100	6	100	100	6	100
PartS ($\mathcal{G}^{\mathrm{mis}_1}$)	47	219	83	29	419	83
PartS ($\mathcal{G}^{\mathrm{mis}_2}$)	34	328	83	21	621	83
CombPartS (5 $\mathcal{G}^{\mathrm{mis}_2}$)	55	171	100	47	234	83
CombPartS (10 $\mathcal{G}^{\mathrm{mis}_2}$)	55	159	100	49	214	83
CorrPartS ($\mathcal{G}^{\mathrm{cor}}$)	28	332	83	17	576	83
	Linear Regression					
Setting 4^*	Original covariates			Projected covariates		
	PIT (%)	MMS	TPR (%)	PIT (%)	MMS	TPR (%)
SpatPartS ($\mathcal{G}^{\mathrm{opt}}$)	100	100	100	100	100	100
SpatPartS ($\mathcal{G}^{\mathrm{red}}$)	100	100	100	100	100	100
SpatPartS ($\mathcal{G}^{\mathrm{mis}_1}$)	0	8762	59	0	8797	57
SpatPartS ($\mathcal{G}^{\mathrm{mis}_2}$)	0	9385	52	0	9460	49
CombPartS ($\mathcal{G}^{\mathrm{red}}, \mathcal{G}^{\mathrm{mis}_1}, \mathcal{G}^{\mathrm{mis}_2}$)	100	100	100	100	100	100
	Logistic Regression					
	Original covariates			Projected covariates		
	PIT (%)	MMS	TPR (%)	PIT (%)	MMS	TPR (%)
SpatPartS ($\mathcal{G}^{\mathrm{opt}}$)	0	10000	0	0	10000	0
SpatPartS ($\mathcal{G}^{\mathrm{red}}$)	0	9999	0	0	9999	0
SpatPartS ($\mathcal{G}^{\mathrm{mis}_1}$)	0	9780	18	0	9738	14
SpatPartS ($\mathcal{G}^{\mathrm{mis}_2}$)	0	9998	9	0	9998	7
CombPartS ($\mathcal{G}^{\mathrm{red}}, \mathcal{G}^{\mathrm{mis}_1}, \mathcal{G}^{\mathrm{mis}_2}$)	0	9780	18	0	9738	14

CombPartS, combined partition-based screening; CorrPartS, correlation-guided partition screening; SpatPartS, spatial-oriented partition screening; MMS, the median minimum size of the selected models that are required to have a sure screening; TPR, the average true positive rate; PIT, the estimated probability of including all true predictors in the top n selected predictors.

modifying the correlation structure of covariates. In Setting 1, we change the exchangeable correlation to 0·9, and in Setting 4, we modify the exponential square correlation structure such that $\mathrm{cor}(X_j, X_{j'})$ =0·95 when j and j' are neighbors. We refer to them as Settings 1^* and 4^*, respectively. Table S1 summarizes the comparisons of the model selection accuracy between the two methods. When the group partition is a size-reduced partition, the covariate projection can improve the model selection accuracy for linear regression in Setting 1^*. However, there are no clear improvements for linear regression in Setting 4^* and logistic regression in either setting.

10. Simulation Studies for Goodness-of-Fit Adjustment

This section presents additional simulations to evaluate the adjusted partition-based screening in § 4·1. The data were generated using the same Settings 1–4 as in § 5. The group partitions were also kept the same as those for the partition-based screening method in each scenario in § 5. Table

S2 summarizes the results. For comparisons, Table S2 also includes the model selection accuracy of the partition-based screening listed in Table 1 in § 5. Based on a size-reduced partition, $\mathcal{G}^{\rm red}$, the adjustment produces a clear improvement in the model selection accuracy for both linear regression and logistic regression in most cases except for Setting 4, where it has a comparable performance for linear regression but has a much worse accuracy for logistic regression. For the correlation-guided partitions, $\mathcal{G}^{\rm cor}$, the adjustment improves the accuracy in Setting 2 for both linear regression and logistic regression, while it decreases the selection accuracy substantially in Setting 3, where the predictors are highly correlated with an exchangeable correlation of $0{\cdot}9$. For the misspecified partitions, $\mathcal{G}^{\rm mis_1}$ and $\mathcal{G}^{\rm mis_2}$, the adjustment only improves the accuracy in Setting 1 with $\mathcal{G}^{\rm mis_1}$ for linear regression, while in all other cases it performs worse, especially in Setting 4, where the noise predictors are highly correlated with the true predictors. For combined partition-based screening, the adjustment is not helpful in all cases either. In summary, the adjusted partition-based screening may improve the model selection accuracy when the predictors are not highly correlated, but it may produce worse results in many other cases.

11. ADDITIONAL DATA ANALYSIS RESULTS

This section presents additional data analysis results. Figure S1 presents the boxplots of the mean prediction errors estimated through ten-fold cross-validation, indicating that combined partition-based screening has the best prediction accuracy. This conclusion is also confirmed by the prediction receiver operating characteristic curves in Fig. S2. Figure S3 shows seven axial slices that cut through eight important brain regions, which are selected based on the combined partition-based screening method.

REFERENCES

BARUT, E., FAN, J. & VERHASSELT, A. (2016). Conditional sure independence screening. *J. Am. Statist. Assoc.* **111**, 1266–77.

FAN, J. & SONG, R. (2010). Sure independence screening in generalized linear models with NP-dimensionality. *Ann. Statist.* **38**, 3567–604.

HEYDE, C. C. (2008). *Quasi-Likelihood and Its Application: a General Approach to Optimal Parameter Estimation*. New York: Springer.

KOROLEV, V. Y. & SHEVTSOVA, I. G. (2010). On the upper bound for the absolute constant in the Berry–Esseen inequality. *Theory Probab. Appl.* **54**, 638–58.

VON NEUMANN, J. & GOLDSTINE, H. H. (1947). Numerical inverting of matrices of high order. *Bull. Am. Math. Soc.* **53**, 1021–99.

[*Received August* 2016. *Revised XXXX* 201X]

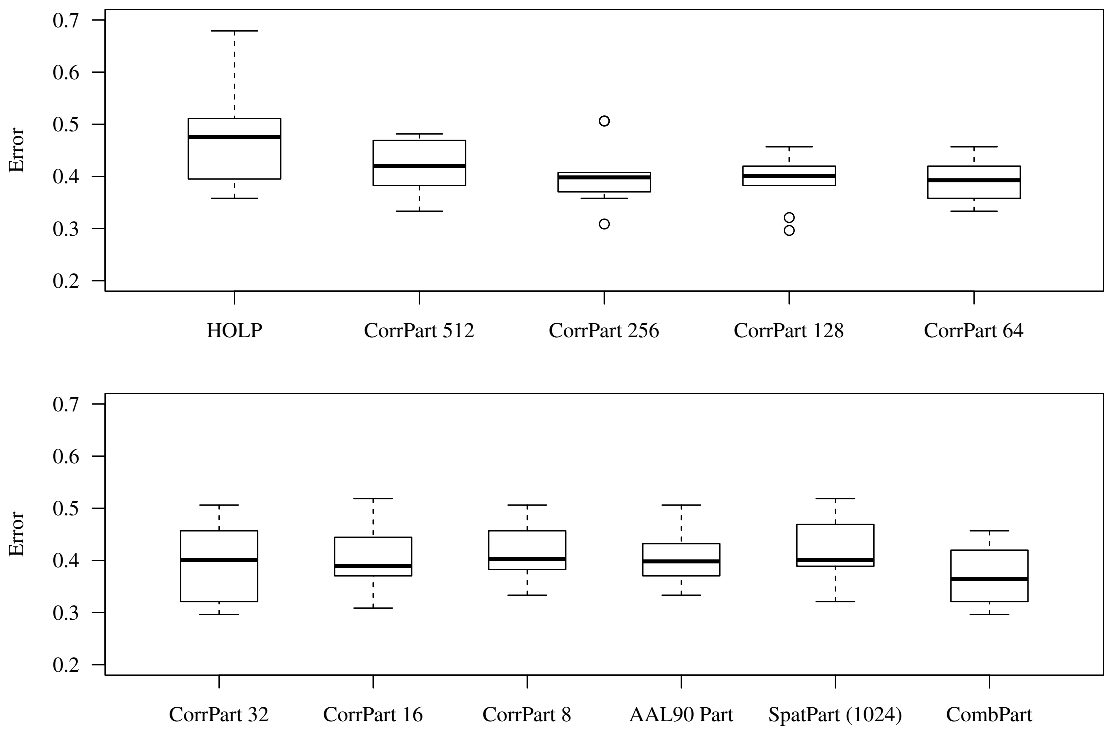

Fig. S1. Boxplots for the autism spectrum disorder risk prediction errors of ten-fold cross-validation by different variable screening methods. The mean cross-validation prediction errors: high-dimensional ordinary least squares projection (HOLP, 48%), correlation-guided partition screening (CorrPartS $G = 8, \ldots, 512$, 39%–42%), brain region partition-based screening (AAL90, 41%), spatial-oriented partition screening (SpatPartS, 42%) and combined partition-based screening (CombPartS, 37%).

Table S2. *Model selection accuracy of partition-based screening with and without the goodness-of-fit adjustment for both linear regression and logistic regression under Settings 1–4*

	Linear Regression					
Setting 1	No adjustment			GOF adjustment		
	PIT (%)	MMS	TPR (%)	PIT (%)	MMS	TPR (%)
PartS ($\mathcal{G}^{\rm red}$)	100	7	100	100	6	100
PartS ($\mathcal{G}^{\rm mis_1}$)	84	36	97	91	30	100
PartS ($\mathcal{G}^{\rm mis_2}$)	87	34	98	82	47	100
CombPartS (5 $\mathcal{G}^{\rm mis_2}$)	93	25	99	92	32	100
CombPartS (10 $\mathcal{G}^{\rm mis_2}$)	96	22	99	89	46	100
CorrPartS ($\mathcal{G}^{\rm cor}$)	78	64	96	81	56	100
	Logistic Regression					
	No adjustment			GOF adjustment		
	PIT (%)	MMS	TPR (%)	PIT (%)	MMS	TPR (%)
PartS ($\mathcal{G}^{\rm red}$)	100	6	100	97	6	100
PartS ($\mathcal{G}^{\rm mis_1}$)	46	226	88	53	175	100
PartS ($\mathcal{G}^{\rm mis_2}$)	60	145	91	33	361	83
CombPartS (5 $\mathcal{G}^{\rm mis_2}$)	64	115	93	56	159	100
CombPartS (10 $\mathcal{G}^{\rm mis_2}$)	74	105	96	52	192	100
CorrPartS ($\mathcal{G}^{\rm cor}$)	47	219	88	30	369	83
	Linear Regression					
Setting 2	No adjustment			GOF adjustment		
	PIT (%)	MMS	TPR (%)	PIT (%)	MMS	TPR (%)
PartS ($\mathcal{G}^{\rm red}$)	100	10	100	100	10	100
CorrPartS ($\mathcal{G}^{\rm cor}$)	91	10	98	96	10	100
	Logistic Regression					
	No adjustment			GOF adjustment		
	PIT (%)	MMS	TPR (%)	PIT (%)	MMS	TPR (%)
PartS ($\mathcal{G}^{\rm red}$)	100	10	100	100	10	100
CorrPartS ($\mathcal{G}^{\rm cor}$)	88	10	98	90	10	100
	Linear Regression					
Setting 3	No adjustment			GOF adjustment		
	PIT (%)	MMS	TPR (%)	PIT (%)	MMS	TPR (%)
PartS ($\mathcal{G}^{\rm red}$)	100	11	100	100	10	100
CorrPartS ($\mathcal{G}^{\rm cor}$)	62	101	100	36	574	90
	Logistic Regression					
	No adjustment			GOF adjustment		
	PIT (%)	MMS	TPR (%)	PIT (%)	MMS	TPR (%)
PartS ($\mathcal{G}^{\rm red}$)	100	10	100	100	10	100
CorrPartS ($\mathcal{G}^{\rm cor}$)	18	426	80	10	1579	0
	Linear Regression					
Setting 4	No adjustment			GOF adjustment		
	PIT (%)	MMS	TPR (%)	PIT (%)	MMS	TPR (%)
SpatPartS ($\mathcal{G}^{\rm opt}$)	100	100	100	100	101	100
SpatPartS ($\mathcal{G}^{\rm red}$)	100	100	100	100	101	100
SpatPartS ($\mathcal{G}^{\rm mis_1}$)	0	8232	65	0	7975	64
SpatPartS ($\mathcal{G}^{\rm mis_2}$)	0	9190	55	0	8672	71
CombPartS ($\mathcal{G}^{\rm red}$, $\mathcal{G}^{\rm mis_1}$, $\mathcal{G}^{\rm mis_2}$)	100	100	100	100	102	100
	Logistic Regression					
	No adjustment			GOF adjustment		
	PIT (%)	MMS	TPR (%)	PIT (%)	MMS	TPR (%)
SpatPartS ($\mathcal{G}^{\rm opt}$)	100	100	100	36	733	99
SpatPartS ($\mathcal{G}^{\rm red}$)	73	174	100	3	5617	96
SpatPartS ($\mathcal{G}^{\rm mis_1}$)	0	9033	65	0	9615	25
SpatPartS ($\mathcal{G}^{\rm mis_2}$)	0	9269	74	0	9100	59
CombPartS ($\mathcal{G}^{\rm red}$, $\mathcal{G}^{\rm mis_1}$, $\mathcal{G}^{\rm mis_2}$)	80	174	100	12	1561	98

PartS, partition-based screening; CombPartS, combined partition-based screening; CorrPartS, correlation-guided partition screening; SpatPartS, spatial-oriented partition screening; MMS, the median minimum size of the selected models that are required to have a sure screening; TPR, the average true positive rate; PIT, the estimated probability of including all true predictors in the top n selected predictors; GOF, goodness-of-fit.

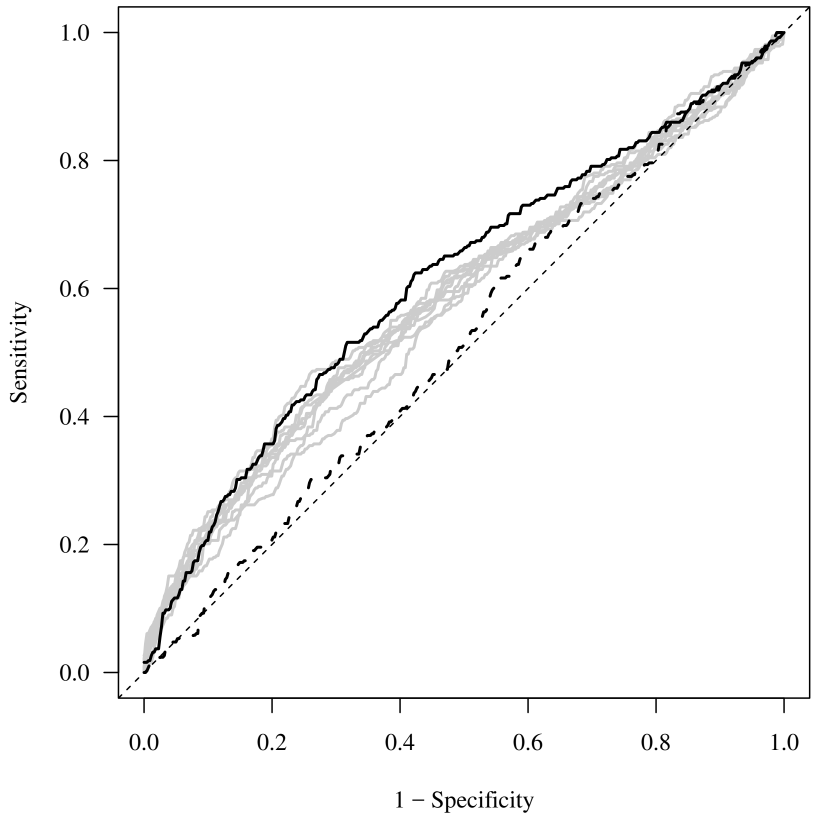

Fig. S2. Prediction receiver operating characteristic curves by combined partition-based screening (solid line), other partition-based screening (grey curves) and high-dimensional ordinary least squares projection (dashes). The reported area under the curves indicates that the proposed method improves on the prediction accuracy compared with high-dimensional ordinary least squares projection.

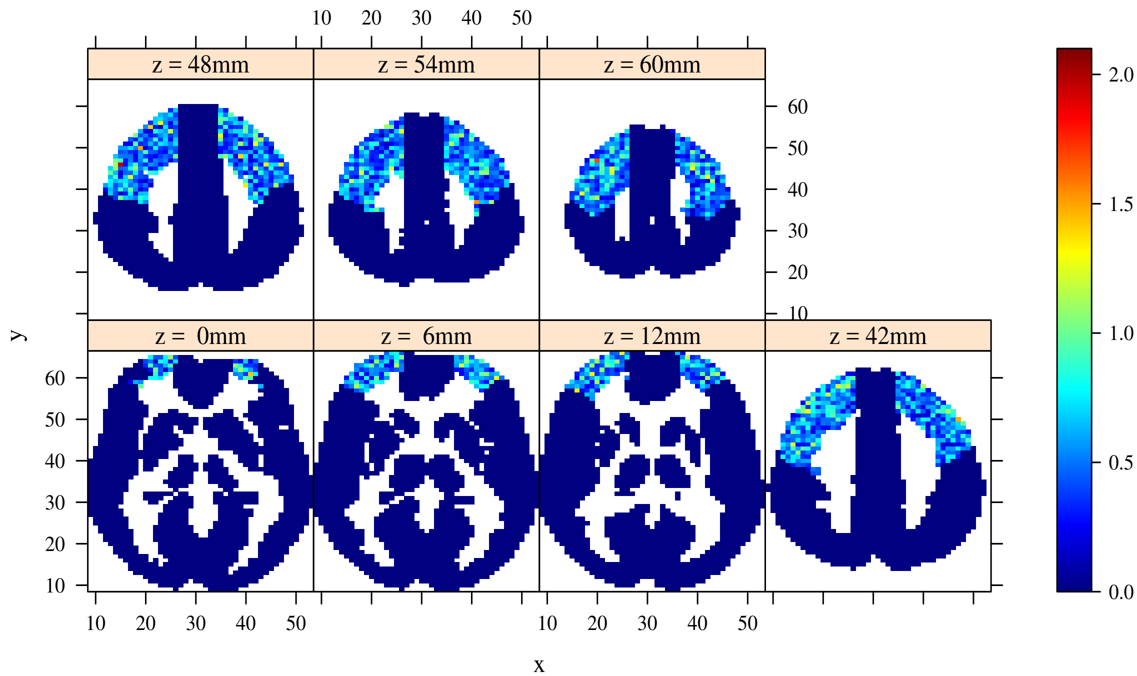

Fig. S3. Combined partition-based screening statistics are shown on seven axial slices that cut through eight important brain regions, which have more than 60 selected voxels.