Supplementary material

In the supplementary material, we provide a table of notations and a table of biomarkers selected by all the methods. We also provide one more simulation setting under missing completely at random, and proofs of Theorems 1–2 and Proposition 1. The table of notations is in the following.

Notation	Definition
$\boldsymbol{X}$	Design matrix consisting of all samples
$\boldsymbol{y}$	The vector of samples for the response variable
y	The response variable
$\boldsymbol{\varepsilon}$	The vector of samples for the error term
ε	The error term
σ_ε	The standard deviation of the error term
N	Number of total samples
p	Number of all covariates
q	Number of relevant covariates in the true model
$\boldsymbol{x}_i$	The i-th row (sample) of $\boldsymbol{X}$
y_i	The i-th sample of the response variable
X_{ij}	The i-th sample of the j-th random covariate
$\boldsymbol{\mathcal{X}}$	A random vector consisting of all random covariates
X_j	The j-th random covariate
$\boldsymbol{C}$	The true covariance matrix of all covariates
c_{ij}	The covariance between the i-th covariate and the j-th covariate
S	Number of data sources
R	Number of different missing patterns or groups
$\mathcal{H}(r)$	Index set of samples in the r-th group

Notation	Definition
n_r	Number of samples in the r-th group
$a(r)$	Index sets of the observed covariates in the r-th group
$m(r)$	Index sets of the missing covariates in the r-th group
$\boldsymbol{\mathcal{X}}_{a(r)}$	The vector of random variables which are observed in the r-th group
$\boldsymbol{\mathcal{X}}_{m(r)}$	The vector of random variables which are missing in the r-th group
$\mathcal{G}(r)$	Index set of the groups where missing variables $\boldsymbol{\mathcal{X}}_{m(r)}$ and variables in at least one of the other sources are observed
M_r	Number of elements in $\mathcal{G}(r)$
$\boldsymbol{\beta}^0$	The vector of true coefficients
β_j^0	The j-th element in $\boldsymbol{\beta}^0$
A_1	Index set of relevant covariates in the vector of all covariates
A_2	Index set of irrelevant covariates in the vector of all covariates
$J(r,k)$	Index set of covariates which are observed in Groups r and k
B_k	Index set of covariates from Source k
$\hat{\boldsymbol{X}}_{m(r)}^{(k)}$	Imputation for all missing values in Group r through estimating conditional expectation $E(X_j \vert \boldsymbol{\mathcal{X}}_{J(r,k)})$ for each $j \in m(r)$
$\hat{E}(X_l \vert \boldsymbol{X}_{iJ(r,k)})$	An estimator of $E(X_l \vert \boldsymbol{\mathcal{X}}_{J(r,k)} = \boldsymbol{X}_{iJ(r,k)})$
$\boldsymbol{\beta}_{a(r)}^0$	The sub-vector of $\boldsymbol{\beta}^0$ consisting of β_j^0 for $j \in a(r)$
$\boldsymbol{\beta}_{m(r)}^0$	The sub-vector of $\boldsymbol{\beta}^0$ consisting of β_j^0 for $j \in m(r)$
$\boldsymbol{\beta}_{A_1}$	The sub-vector of coefficients $\boldsymbol{\beta}$ consisting of β_j for $j \in A_1$
$\boldsymbol{\beta}_{A_2}$	The sub-vector of coefficients $\boldsymbol{\beta}$ consisting of β_j for $j \in A_2$
$X_{ij}^{(k)}$	The $X_{ij}^{(k)}$ equals X_{ij} if the j-th covariate is observed in the sample $\boldsymbol{x}_j$, otherwise $X_{ij}^{(k)}$ is the imputed value of X_{ij} in $\hat{\boldsymbol{X}}_{m(r)}^{(k)}$

Notation	Definition
$\boldsymbol{x}_i^{(k)}$	The vector of $X_{ij}^{(k)}$ for $1 \leq j \leq p$
$\boldsymbol{z}_i^{(k)}$	The sub-vector of $\boldsymbol{x}_i^{(k)}$ consisting of $X_{ij}^{(k)}$ for $j \in a(k)$
$p_\lambda(\cdot)$	The penalty function with tuning parameter λ
$\mu_i^{(k)}(\boldsymbol{\beta})$	The function defined as $\boldsymbol{x}_i^{(k)}\boldsymbol{\beta}$
$\boldsymbol{g}_i^{(r,k)}(\boldsymbol{\beta})$	The vector defined as $\{\boldsymbol{z}_i^{(k)}\}^T\{y_i - \mu_i^{(k)}(\boldsymbol{\beta})\}$
$\boldsymbol{g}_i^{(r)}(\boldsymbol{\beta})$	The vector consisting of $\boldsymbol{g}_i^{(r,k)}(\boldsymbol{\beta})$ for $k \in \mathcal{G}(r)$
$\boldsymbol{g}^{(r)}(\boldsymbol{\beta})$	The vector consisting of estimating functions $\sum_{i\in\mathcal{H}(r)} \boldsymbol{g}_i^{(r)}(\boldsymbol{\beta})/n_r$
$\boldsymbol{g}(\boldsymbol{\beta})$	The vector consisting of $\boldsymbol{g}^{(r)}(\boldsymbol{\beta})$ for $1 \leq r \leq R$
$\boldsymbol{W}^{(r)}(\boldsymbol{\beta})$	The matrix defined as $\sum_{i\in\mathcal{H}(r)} \boldsymbol{g}_i^{(r)}(\boldsymbol{\beta})\{\boldsymbol{g}_i^{(r)}(\boldsymbol{\beta})\}^T/n_r$
$\boldsymbol{W}(\boldsymbol{\beta})$	The block diagonal matrix with diagonal blocks $\boldsymbol{W}^{(r)}(\boldsymbol{\beta})$ for $1 \leq r \leq R$
$f(\boldsymbol{\beta})$	The function defined as $\{\boldsymbol{g}(\boldsymbol{\beta})\}^T\boldsymbol{W}(\boldsymbol{\beta})^{-1}\boldsymbol{g}(\boldsymbol{\beta}) + \sum_{j=1}^p p_\lambda(\lvert\beta_j\rvert)$
$\boldsymbol{g}_{(1)}^{(r)}$	The vector consisting of the estimating functions in $\boldsymbol{g}^{(r)}(\boldsymbol{\beta})$ whose imputation is based on complete observations
$\boldsymbol{g}_{(2)}^{(r)}$	The vector consisting of the estimating functions in $\boldsymbol{g}^{(r)}(\boldsymbol{\beta})$ but not in $\boldsymbol{g}_{(1)}^{(r)}$
$\boldsymbol{W}_{11}^{(r)}$	The sub-matrix of $\boldsymbol{W}$ corresponding to $\boldsymbol{g}_{(1)}^{(r)}$
$\boldsymbol{W}_{22}^{(r)}$	The sub-matrix of $\boldsymbol{W}$ corresponding to $\boldsymbol{g}_{(2)}^{(r)}$
$\boldsymbol{U}^{(r)}(\boldsymbol{\beta})$	The transformation matrix for $\boldsymbol{g}^{(r)}$ to obtain essential information from $\boldsymbol{g}^{(r)}$
$\boldsymbol{U}(\boldsymbol{\beta})$	The block diagonal matrix with diagonal blocks $\boldsymbol{U}^{(r)}$ for $1 \leq r \leq R$
$f^*(\boldsymbol{\beta})$	The object function defined as $(\boldsymbol{U}\boldsymbol{g})^T(\boldsymbol{U}\boldsymbol{W}\boldsymbol{U}^T)^{-1}\boldsymbol{U}\boldsymbol{g} + \sum_{j=1}^p p_\lambda(\lvert\beta_j\rvert)$
$\hat{\boldsymbol{\beta}}$	The proposed estimator which is also the minimizer of $f^*(\boldsymbol{\beta})$
$\hat{\boldsymbol{\beta}}_\lambda$	The proposed estimator with a given λ
df_λ	The number of non-zero estimated coefficients in $\hat{\boldsymbol{\beta}}_\lambda$
$\boldsymbol{s}_k$	The conjugate direction at the k-th iteration in Algorithm 1

Notation	Definition
α_k	The step size at the k-th iteration in Algorithm 1
$\boldsymbol{\beta}^{(k)}$	The updated coefficient estimator at the k-th iteration in Algorithm 1
$RSS_r(\boldsymbol{\beta})$	Average of residual sum of squares for samples in Group r over all imputations
$RSS(\boldsymbol{\beta})$	The sum of RSS_r for $1 \le r \le R$
ξ_i	The random group label for the i-th subject
$\boldsymbol{g}_i(\boldsymbol{\beta})$	The i-th sample of all available estimating functions, that is, $(\boldsymbol{g}_i^{(1)}(\boldsymbol{\beta})^T, \ldots, \boldsymbol{g}_i^{(R)}(\boldsymbol{\beta})^T)^T$
$\boldsymbol{G}(\boldsymbol{\beta})$	The sample matrix for all estimating equations, that is, $(\boldsymbol{g}_1(\boldsymbol{\beta}), \ldots, \boldsymbol{g}_N(\boldsymbol{\beta}))^T$
$\boldsymbol{G}^{(r)}(\boldsymbol{\beta})$	The sub-matrix of $\boldsymbol{G}(\boldsymbol{\beta})$ with columns representing estimating functions of the r-th group
$\boldsymbol{G}_0^*(\boldsymbol{\beta})$	The sample matrix for transformed estimating functions which are linearly independent at $\boldsymbol{\beta} = \boldsymbol{\beta}^0$, that is, $\boldsymbol{G}_0^*(\boldsymbol{\beta}) = \boldsymbol{G(\beta)\{U(\beta^0)\}^T}$
$\boldsymbol{V}$	The asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}_{A_1}$
$\hat{\boldsymbol{V}}$	The empirical covariance matrix of $\hat{\boldsymbol{\beta}}_{A_1}$
$\boldsymbol{U}_{(1)}$	The transformation matrix selecting estimating functions corresponding to the single imputation
$\boldsymbol{G}_{(1)}(\boldsymbol{\beta})$	The sample matrix defined as $\boldsymbol{G}(\boldsymbol{\beta})\{\boldsymbol{U}_{(1)}\}^T$
$\hat{\boldsymbol{V}}^{(1)}$	The empirical covariance matrix of the estimator induced by $\boldsymbol{G}_{(1)}(\boldsymbol{\beta})$
$\mathcal{B}_0$	A neighborhood of $\boldsymbol{\beta}_0$
$\beta_{\min}$	The minimum signal, that is, the minimum value of non-zero elements in $\boldsymbol{\beta}^0$
$\lambda_{\min}(\cdot)$	The smallest eigenvalue of a matrix
n_v	Number of random samples in a validation set
β_{sk}	The shared signal strength in Source k
ρ	The correlation among covariates in simulations

Notation	Definition
$\boldsymbol{H}(\boldsymbol{\beta})$	The matrix defined as $\left(\{\mathbf{1}^T\partial_1\boldsymbol{G}(\boldsymbol{\beta})/N\}^T, \ldots, \{\mathbf{1}^T\partial_p\boldsymbol{G}(\boldsymbol{\beta})/N\}^T\right)^T$
$\boldsymbol{H}_{A_1}(\boldsymbol{\beta})$	The sub-matrix of $\boldsymbol{H}(\boldsymbol{\beta})$ consisting of rows corresponding to covariates indexed by A_1
$\boldsymbol{H}_{A_2}(\boldsymbol{\beta})$	The sub-matrix of $\boldsymbol{H}(\boldsymbol{\beta})$ consisting of rows corresponding to covariates indexed by A_2
$\boldsymbol{W}_0^*(\boldsymbol{\beta})$	The matrix defined as $\boldsymbol{G}_0^*(\boldsymbol{\beta})^T\boldsymbol{G}_0^*(\boldsymbol{\beta})/N$
$\widehat{\boldsymbol{W}}(\boldsymbol{\beta})$	An estimator of the weighting matrix for all estimating functions that is defined as $\{\boldsymbol{U}(\boldsymbol{\beta}^0)\}^T\{\boldsymbol{W}_0^*(\boldsymbol{\beta})\}^{-1}$
$\widetilde{\boldsymbol{W}}(\boldsymbol{\beta})$	The matrix defined as $\boldsymbol{H}_{A_1}(\boldsymbol{\beta})\widehat{\boldsymbol{W}}(\boldsymbol{\beta})\{\boldsymbol{H}_{A_1}(\boldsymbol{\beta})\}^T$
$\partial_j\boldsymbol{G}(\boldsymbol{\beta})$	The first derivative of $\boldsymbol{G}(\boldsymbol{\beta})$ with respect to β_j

Table 8: Notations.

Method	Biomarkers selected
Proposed method	ST39SA, ST49TA, ST26TA, **ST29SV**, X11730174_at, ST88SV, ST51TA, ST48TA, ST107TA, 11720564_a_at, 11744781_at, ST108TA, ST32CV, ST91TA, ST83CV, **11723246_s_at**, 11726615_a_at, ST102TA, ST24TA, ST32TA, **ST40TA**, 11724946_at, 11745173_a_at, ST113SA, **ST119TS**, **ST60TS**, ST26CV, ST39CV, ST82TA, **RIGHT_LATERAL_VENTRICLE**, 11754084_x_at, ST107TS, ST13CV, ST37SV, **RIGHT_CHOROID_PLEXUS**, 11743801_at, 11759895_at, ST113CV, ST13SA, **ST43TA**, ST83TS
CC-SCAD	11723246_s_at, 11736422_s_at, ST60TS, 11752811_a_at, ST45CV, 11716721_a_at, ST18SV, ST40TA, ST113TA, ST114SA, ST29SV, ST40CV, ST47SA

Method	Biomarkers selected
SI-SCAD	ST30SV, ST107TS, ST39SA, 11753548_x_at, ST103SA, 11737596_x_at, ST46TA, 11743801_at, ST58TA, ST57SA, ST45TA, 11746766_a_at, ST105TA, CTX_LH_TEMPORALPOLE, ST31TA, ST90TA, 11746527_a_at, ST57TS, ST60TA, 11736223_a_at, ST3SV, 11718714_x_at, 11737767_a_at, 11746953_a_at, ST83CV, 11749470_a_at, ST113TA, ST35TA, ST72TA, CTX_RH_PARAHIPPOCAMPAL, 11723246_s_at, ST94CV, ST6SV, 11750206_a_at, ST74CV, 11724946_at, 11727402_at, ST85CV, 11737155_at, ST32CV, ST107CV, ST46TS, ST98TA, SUMMARYSUVR_COMPOSITE_REFNORM, 11758477_s_at, ST24TS, ST52TA, ST62CV, 11761965_at, 11747813_a_at, ST119TA
DISCOM	ST89SV, 11737596_x_at, 11743252_a_at, ST30SV, 11724191_a_at, ST80SV, 11724946_at, 11737155_at, 11743980_a_at, 11730174_at, 11718745_s_at, 11746766_a_at, ST32CV, 11723960_at, 11734187_a_at, ST130TS, 11726615_a_at, 11727371_a_at, 11737767_a_at, 11749470_a_at, 11759181_at, ST69SV, SUMMARYSUVR_COMPOSITE_REFNORM, 11762182_a_at, ST21SV, ST24TA, 11724133_s_at, 11742849_a_at, 11744198_s_at, 11736223_a_at, 11744781_at, 11747813_a_at, 11754084_x_at, ST114TA, CTX_LH_FRONTALPOLE, 11745173_a_at, 11750337_a_at, 11726492_a_at, 11748676_a_at, CTX_LH_INFERIORTEMPORAL, 11719316_s_at, 11725683_at, 11743801_at, 11747512_s_at, ST24CV, ST29SV, ST83TS, 11720564_a_at, 11728396_a_at, 11728584_s_at, ST73TS, SUMMARYSUVR_WHOLECEREBNORM, 11722305_at, 11760100_x_at, 11716212_a_at, 11721575_at, 11727402_at, 11738311_at, 11746953_a_at, ST83CV, CTX_RH_INFERIORTEMPORAL, 11737936_at, 11762124_x_at, ST37SV, ST96SV, 11715588_x_at, 11717488_s_at, 11751416_a_at, 11759812_at, 11734307_at, 11745344_a_at, 11746127_a_at, 11751481_a_at, ST14TS, 11724699_a_at, 11730853_s_at, 11736641_x_at, 11753548_x_at, 11762618_a_at, ST40CV, ST85CV, 11716721_a_at, ST127SV, CTX_RH_CAUDALMIDDLEFRONTAL, 11715936_x_at, 11717693_at, 11735003_a_at, 11746527_a_at, 11747491_a_at, RIGHT_CHOROID_PLEXUS, CTX_RH_ROSTRALMIDDLEFRONTAL, 11718714_x_at, 11737372_at, 11755402_x_at, 11755826_a_at, 11763162_a_at, CTX_LH_ROSTRALMIDDLEFRONTAL, 11727226_at, 11737668_at, 11740134_a_at, 11758148_s_at, ST105TA, ST71SV, ST88SV, 11719499_at, 11721576_s_at, 11729697_at, 11735407_at, 11739708_x_at, 11745583_a_at, 11759396_a_at, 11715683_a_at,

Method	Biomarkers selected
	11725272_a_at, 11737392_at, 11743543_x_at, ST58TA, CSF, CTX_LH_CAUDALMIDDLEFRONTAL, CTX_LH_MIDDLETEMPORAL, 11716298_x_at, 11717134_a_at, 11719838_a_at, 11722407_a_at, 11727994_at, ST32SA, ST90CV, RIGHT_HIPPOCAMPUS, WM_HYPOINTENSITIES, CTX_LH_SUPERIORFRONTAL, 11716639_a_at, 11732139_at, 11732891_a_at, ST104TS, ST12SV, ST31CV, ST60TS, RIGHT_INF_LAT_VENT, 11723246_s_at, 11744011_s_at, 11752945_a_at, 11759078_at, ST129TA, ST46TA, ST77SV, ST91CV, LEFT_INF_LAT_VENT, FRONTAL, CTX_RH_PARSORBITALIS, CTX_RH_MIDDLETEMPORAL, 11726884_a_at, 11730171_a_at, 11736422_s_at, 11737442_x_at, 11741631_a_at, 11742242_at, 11744036_a_at, 11750206_a_at, ST106TA, ST24TS, ST46TS, ST84TA, RIGHT_LATERAL_VENTRICLE, CTX_LH_PARSOPERCULARIS, CTX_LH_PARSORBITALIS, 11716460_x_at, 11745763_a_at, 11758010_s_at, ST115TA, ST34CV, ST56TS, ST58CV, ST60SA, ST94CV, CTX_LH_TEMPORALPOLE, LEFT_HIPPOCAMPUS, SUMMARYSUVR_WHOLECEREBNORM_1.11CUTOFF, CTX_LH_LATERALORBITOFRONTAL, CTX_LH_PARSTRIANGULARIS, CTX_RH_LATERALORBITOFRONTAL, 11716973_a_at, 11739586_a_at, 11742765_at, 11746321_s_at, 11754191_x_at, 11755647_a_at, ST103SA, ST103TA, ST105CV, ST111CV, ST115TS, ST121TS, ST18SV, ST26TA, ST49TA, ST52CV, ST55TA, ST57SA, ST70SV, ST83TA, ST90TA, CC_MID_POSTERIOR, CTX_RH_BANKSSTS, TEMPORAL, CTX_RH_SUPERIORFRONTAL, CTX_LH_ROSTRALANTERIORCINGULATE, 11742273_a_at, 11752491_x_at, 11752811_a_at, 11759455_s_at, 11760709_at, 11762385_at, ST102CV, ST107TS, ST108TA, ST116CV, ST120SV, ST13TS, ST14SA, ST31TA, ST32TA, ST49TS, ST53SV, ST7SV, ST99CV, CTX_LH_FUSIFORM, LEFT_LATERAL_VENTRICLE, 11717221_s_at, 11726149_a_at, 11735696_a_at, 11746891_x_at, 11748045_x_at, 11763698_x_at, ST118CV, ST31TS, ST38SA, ST39SA, ST72CV, VENTRICLE_3RD, CC_CENTRAL, CTX_RH_FUSIFORM, NON_WM_HYPOINTENSITIES, CTX_LH_SUPRAMARGINAL, 11717764_x_at, 11718885_a_at, 11719114_a_at, 11720429_at, 11722008_a_at, 11735132_s_at, 11743393_x_at, 11753357_a_at, 11757675_s_at, 11758006_s_at, 11758379_s_at, ST107SA, ST112SV, ST39TA, ST45TS, ST46CV, ST61SV, ST74CV, ST82TA, ST85TA, CC_POSTERIOR, RIGHT_VENTRALDC, CTX_RH_FRONTALPOLE, CTX_RH_ROSTRALANTERIORCINGULATE, CTX_LH_INFERIORPARIETAL, 11718135_at, 11721817_a_at, 11725804_a_at, 11732597_at, 11743189_a_at, 11745007_s_at, 11750975_x_at,

Method	Biomarkers selected
	11763513_at, ST105SA, ST111TA, ST119CV, ST57TA, ST58SA, ST68SV, ST76SV, ST148SV, ST91TA, ST95SA, CC_MID_ANTERIOR, CTX_RH_TEMPORALPOLE, LEFT_VENTRALDC, CTX_LH_SUPERIORTEMPORAL, 11716507_at, 11727547_s_at, 11732526_s_at, 11734017_a_at, 11741744_x_at, 11745780_x_at, 11745906_at, 11747247_x_at, 11749588_a_at, 11750137_a_at, 11755073_a_at, 11756884_a_at, 11761223_at, ST107CV, ST108CV, ST124SV, ST129TS, ST147SV, ST23CV, ST26SA, ST40TA, ST54CV, ST60TA, ST74TS, ST82SA, ST90TS, ST94TA, ST95TA, ST99TA, VENTRICLE_4TH, CTX_LH_PRECENTRAL, CTX_RH_PARAHIPPOCAMPAL, LEFT_CHOROID_PLEXUS, CTX_RH_PARSTRIANGULARIS, CTX_RH_CAUDALANTERIORCINGULATE, 11715986_s_at, 11716947_at, 11721093_a_at, 11722152_s_at, 11725075_a_at, 11727169_x_at, 11731936_at, 11744125_x_at, 11745802_a_at, 11752835_a_at, 11763423_x_at, ST104TA, ST106CV, ST107TA, ST108TS, ST113TA, ST114SA, ST116SA, ST13CV, ST25CV, ST34SA, ST38TA, ST3SV, ST42SV, ST48TA, ST50CV, ST59SA, ST62TA, ST84CV, ST94TS, ST97TS, ST98CV, ST98SA, CC_ANTERIOR, CTX_RH_PERICALCARINE, LEFT_PUTAMEN, RIGHT_AMYGDALA, CTX_LH_MEDIALORBITOFRONTAL, 11722280_at, 11723294_x_at, 11732175_s_at, 11736364_s_at, 11744443_x_at, 11745219_a_at, 11754088_a_at, 11763530_a_at, ST103CV, ST106SA, ST110TS, ST13TA, ST23TS, ST25TS, ST26CV, ST2SV, ST35TA, ST38CV, ST52TA, ST56SA, ST56TA, ST59CV, ST59TA, ST60CV, ST6SV, ST73CV, ST82CV, ST85SA, ST93TS, ST97TA
DISCOM-Huber	11743252_a_at, 11737767_a_at, 11730174_at, 11744781_at, ST21SV, 11747512_s_at, ST29SV, ST88SV, 11716460_x_at, 11725683_at, 11754084_x_at, ST80SV, 11724946_at, 11726615_a_at, SUMMARYSUVR_COMPOSITE_REFNORM, 11715588_x_at, 11743980_a_at, 11719838_a_at, 11745763_a_at, 11750337_a_at, ST24CV, 11720564_a_at, 11750975_x_at, 11759181_at, 11762124_x_at, ST83TS, ST24TA, ST83CV, 11736641_x_at, 11737596_x_at, 11746127_a_at, 11724699_a_at, 11734307_at, ST129TS, 11721817_a_at, 11739708_x_at, 11717134_a_at, 11726492_a_at, 11739826_s_at, 11743801_at, 11744036_a_at, 11728396_a_at, 11723960_at, ST12SV, 11719016_at, 11716721_a_at, 11727371_a_at, 11737155_at, 11745173_a_at, 11760945_at, ST71SV, ST73TS, 11719499_at, 11737936_at, 11747813_a_at, 11748676_a_at, ST130TS, SUMMARYSUVR_COMPOSITE_REFNORM_0.79CUTOFF, 11717488_s_at, 11721575_at, 11724191_a_at, 11743543_x_at,

Method	Biomarkers selected
	11724133_s_at, 11737372_at, 11742242_at, 11718714_x_at, ST7SV, 11715683_a_at, 11716973_a_at, ST104TS, SUMMARYSUVR_WHOLECEREBNORM, SUMMARYSUVR_WHOLECEREBNORM_1.11CUTOFF, 11722152_s_at, 11726665_a_at, 11746953_a_at, 11749470_a_at, 11763162_a_at, CTX_LH_FRONTALPOLE, 11718745_s_at, 11727169_x_at, 11735003_a_at, 11742765_at, 11750199_a_at, 11762618_a_at, ST46TS, RIGHT_CHOROID_PLEXUS, FRONTAL, CTX_LH_CAUDALMIDDLEFRONTAL, CTX_LH_ROSTRALMIDDLEFRONTAL, CTX_RH_CAUDALMIDDLEFRONTAL, CTX_RH_ROSTRALMIDDLEFRONTAL, 11716298_x_at, 11744011_s_at, ST106TA, ST69SV, ST85CV, CSF, 11730767_x_at, ST24TS, ST32CV, CTX_LH_INFERIORTEMPORAL, 11736223_a_at, 11738311_at, 11746527_a_at, 11750206_a_at, ST91CV, 11727226_at, 11745583_a_at, 11755402_x_at, 11759078_at, 11762182_a_at, ST127SV, ST77SV, ST89SV, CTX_RH_INFERIORTEMPORAL, RIGHT_HIPPOCAMPUS, CTX_LH_SUPERIORFRONTAL, 11722421_at, 11722989_a_at, 11752491_x_at, 11762385_at, CTX_RH_SUPERIORFRONTAL, 11715936_x_at, 11725272_a_at, 11725998_a_at, 11739586_a_at, 11746766_a_at, 11747828_x_at, ST103TA, ST129TA, ST60SA, LEFT_LATERAL_VENTRICLE, RIGHT_INF_LAT_VENT, CTX_RH_PARSORBITALIS, 11725861_a_at, 11737392_at, CC_MID_POSTERIOR, CC_POSTERIOR, RIGHT_LATERAL_VENTRICLE, CTX_RH_FRONTALPOLE, 11726865_a_at, 11734187_a_at, 11736571_x_at, 11737442_x_at, 11742849_a_at, 11759895_at, ST18SV, ST58TA, CTX_LH_MEDIALORBITOFRONTAL, CTX_LH_PARSOPERCULARIS, 11732891_a_at, 11753548_x_at, 11754810_a_at, LEFT_HIPPOCAMPUS, NON_WM_HYPOINTENSITIES, WM_HYPOINTENSITIES, CTX_RH_MEDIALORBITOFRONTAL, 11721433_a_at, 11721576_s_at, 11727547_s_at, 11747491_a_at, 11751412_x_at, 11760100_x_at, ST103CV, ST115TS, ST31CV, ST32SA, ST57SA, ST58CV, ST83TA, LEFT_INF_LAT_VENT, TEMPORAL, CTX_LH_ROSTRALANTERIORCINGULATE, CTX_RH_MIDDLETEMPORAL, 11718853_a_at, 11727402_at, 11730171_a_at, 11746321_s_at, 11755826_a_at, ST111CV, ST120SV, ST121TS, ST34CV, ST76SV, CTX_RH_PARSOPERCULARIS, 11759455_s_at, 11759812_at, 11759909_x_at, 11763530_a_at, ST49TA, ST94CV, ST95TA, ST96SV, VENTRICLE_4TH, CTX_RH_POSTCENTRAL, LEFT_VENTRALDC, CTX_RH_LATERALORBITOFRONTAL, CTX_RH_PARSTRIANGULARIS, CTX_LH_SUPERIORTEMPORAL, CTX_RH_SUPERIORTEMPORAL, 11759148_at, ST26CV, ST70SV, ST148SV, ST93CV, CTX_LH_TEMPORALPOLE,

Method	Biomarkers selected
	RIGHT_CEREBELLUM_WHITE_MATTER, RIGHT_VENTRALDC, CTX_LH_PARSORBITALIS, CTX_LH_PARSTRIANGULARIS, CTX_RH_ROSTRALANTERIORCINGULATE, 11717693_at, 11722407_a_at, 11722922_s_at, 11737668_at, 11744198_s_at, 11748045_x_at, 11751416_a_at, 11755647_a_at, 11757786_s_at, 11758379_s_at, ST108TA, ST40TA, ST52CV, ST54TS, ST60TS, ST73CV, ST90CV, VENTRICLE_3RD, CC_CENTRAL, WHOLECEREBELLUM, CTX_LH_SUPRAMARGINAL, CTX_LH_MIDDLETEMPORAL, 11723608_a_at, 11724787_s_at, 11745219_a_at, 11745780_x_at, 11749258_a_at, 11752811_a_at, 11752945_a_at, 11758010_s_at, ST104TA, ST105TA, ST107SA, ST116CV, ST147SV, ST49TS, ST56TS, ST61SV, ST90TA, CTX_LH_POSTCENTRAL, CTX_LH_PRECENTRAL, CTX_RH_BANKSSTS, CTX_RH_PRECENTRAL, LEFT_CHOROID_PLEXUS, BRAINSTEM, CINGULATE, CTX_RH_CAUDALANTERIORCINGULATE, CTX_RH_SUPRAMARGINAL, 11718135_at, 11719316_s_at, 11722008_a_at, 11723246_s_at, 11731161_at, 11732175_s_at, 11733074_s_at, 11735132_s_at, 11737753_s_at, 11742273_a_at, 11743180_at, 11744017_at, 11746891_x_at, 11754088_a_at, 11756401_x_at, 11763423_x_at, ST114SA, ST23TS, ST26TA, ST30SV, ST34SA, ST37SV, ST39TA, ST45TA, ST50SA, ST54TA, ST56SA, ST62TA, ST72CV, ST85TA, CTX_LH_PARACENTRAL
iSFSF	ST39SA, ST60SA, ST114TA, ST13SA, 11724699_a_at, ST111TA, ST14TS, ST62TS, 11745007_s_at, 11749470_a_at, 11739586_a_at, 11761965_at, 11760945_at, 11762182_a_at, ST84SA, ST98SA, ST112SV, 11751481_a_at, ST113SA, ST39CV, ST108TA, ST38CV, ST130SA, 11736571_x_at, 11727994_at, ST84TA, 11743324_a_at, ST118SA, 11748045_x_at, ST50CV, ST109CV, ST23CV, ST119SA, ST34SA, 11726865_a_at, 11746527_a_at, ST54SA, 11744583_s_at, ST49TA, ST98CV, 11730767_x_at, ST69SV, 11722407_a_at

Table 9: Biomarkers selected by all the methods.

	$N \cdot \log\left\{RSS(\hat{\boldsymbol{\beta}}_\lambda)/N\right\}$	$df_\lambda \cdot \log(N)$	MBI-BIC$_\lambda$
1	8809.12	262.04	9071.16
2	8809.08	262.04	9071.12
3	8809.03	262.04	9071.07
4	8808.99	262.04	9071.03
5	8808.96	262.04	9071.00
6	8808.56	262.04	9070.60
7	8757.83	262.04	9019.87
8	8767.55	242.39	9009.94
9	8781.33	209.63	8990.96
10	8777.70	163.78	8941.47
11	8777.70	163.78	8941.48
12	8818.59	104.82	8923.41
13	8818.57	104.82	8923.39
14	8871.46	65.51	8936.97
15	8871.45	65.51	8936.96
16	8982.54	26.20	9008.74
17	9037.99	19.65	9057.64
18	9037.99	19.65	9057.64
19	9044.92	19.65	9064.57
20	9044.92	19.65	9064.57

Table 10: The two terms in MBI-BIC for Setting 1 with $\rho = 0.7$.

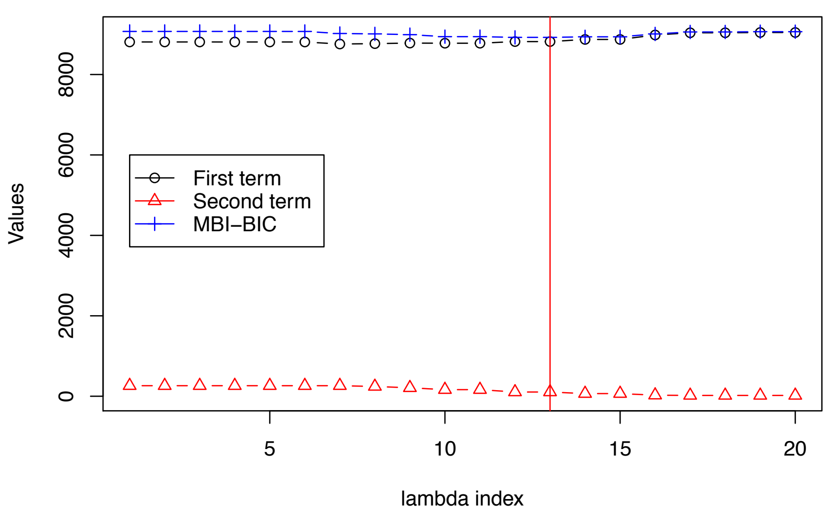

Figure 3: MBI-BIC plot under Setting 1 with $\rho = 0.7$. In the plot, circles represent values of the first term $N \cdot \log\left\{RSS(\hat{\boldsymbol{\beta}}_\lambda)/N\right\}$ as λ increases. Triangles and pluses represent values of the second term $df_\lambda \cdot \log(N)$ and MBI-BIC$_\lambda$, respectively.

Setting 7 in the following is simulated with an unstructured correlation matrix. We first generate a $p \times p$ matrix $\boldsymbol{A}_0$ with each element simulated from Unif$(0,1)$, then let $\boldsymbol{A}_1 = \boldsymbol{A}_0^T \boldsymbol{A}_0$, and finally transform the $\boldsymbol{A}_1$ to a correlation matrix $\boldsymbol{A}_2$. In Setting 7, each row of $\boldsymbol{X}$ is independent and identically distributed from $N(\mathbf{0}, \boldsymbol{A}_2)$. Table 11 shows that the proposed method outperforms other methods in terms of overall FNR+FPR, which is consistent with results of Settings 1–5.

Setting 7: Let $N = 800$, $p = 30$, $q = 6$, $R = 4$, $S = 3$, $p_1 = p_2 = 8$, $p_3 = 14$, $(\beta_{s1}, \beta_{s2}, \beta_{s3}) = (6, 7, 8)$. Each source contains two relevant covariates. The four groups have the same missing structure as Groups 1–4 illustrated in Figure 1. The probability of a sample to be assigned to Group r is b_r for $r = 1, 2, 3, 4$, where $b_1 = 1/31$ and $b_2 = b_3 = b_4 = 10/31$. The validation set consists of half of random samples from Group 1.

Method	FNR	FPR	FNR+FPR	Time
Proposed method	0.027	0.303	0.329	3.588
CC-SCAD	0.587	0.089	0.676	0.020
SI-SCAD	0.140	0.525	0.665	0.061
DISCOM	0.000	0.619	0.619	0.027
DISCOM-Huber	0.000	0.611	0.611	0.037
iSFS	0.340	0.338	0.678	0.289

Table 11: FNR, FPR, and FNR+FPR under Setting 7.

For the proofs, we define the following notations. For any vector $\boldsymbol{v} = (v_1, \ldots, v_k)$ and set J, $\boldsymbol{v}_J$ refers to a sub-vector of $\boldsymbol{v}$, which consists of v_j for all $j \in J$. If J is a singleton with an element j, then $\boldsymbol{v}_J$ becomes v_j. For any matrix $\boldsymbol{M}$, any set J_1 and J_2, $\boldsymbol{M}_{J_1 J_2}$ refers to a sub-matrix of $\boldsymbol{M}$, whose rows are indexed by J_1 and columns are indexed by J_2; $\boldsymbol{M}_{J_1,\cdot}$ refers to a matrix containing rows of $\boldsymbol{M}$ which are indexed by J_1; $\boldsymbol{M}_{\cdot,J_2}$ refers to a matrix containing columns of $\boldsymbol{M}$ which are indexed by J_2. For example, $\boldsymbol{X}_{iJ}$ is a row vector containing X_{ij} for all $j \in J$.

Proof of Theorem 1. (i) and (ii): Since we have multiple imputations in each group, columns in $\boldsymbol{G}(\boldsymbol{\beta})$ could be linearly dependent. Suppose that the columns in $\boldsymbol{G}(\boldsymbol{\beta})$ span a $d_{\boldsymbol{\beta}}$-dimensional space $\mathcal{D}_{\boldsymbol{\beta}}$. Then, $d_{\boldsymbol{\beta}} \geq p$ for sufficiently large N. Since $\log(n_r d)/(n_r d)$ in (5) goes to zero as $Nge \to \infty$ for each $1 \leq r \leq R$, $\boldsymbol{U}$ contains $d_{\boldsymbol{\beta}}$ eigenvectors of $\boldsymbol{W}$. Then, the first term in (4) is a projection of $\mathbf{1}$ onto the $d_{\boldsymbol{\beta}}$-dimensional space $\mathcal{D}_{\boldsymbol{\beta}}$, and thus does not depend on the choice of basis. For convenience and without loss of generality, we let $\boldsymbol{U}$ have $d_{\boldsymbol{\beta}}$ rows with $\boldsymbol{U}_{kj_k} = 1$ for $1 \leq k \leq d_{\boldsymbol{\beta}}$ and other elements 0, and select a basis of the space spanned by the columns through $\boldsymbol{G}^*(\boldsymbol{\beta}) = \boldsymbol{G}(\boldsymbol{\beta})[\boldsymbol{U}(\boldsymbol{\beta})]^{\boldsymbol{T}}$. Here, $\{j_1, \ldots, j_{d_{\boldsymbol{\beta}}}\}$ are indexes of the first $d_{\boldsymbol{\beta}}$ linearly independent columns in $\boldsymbol{G}$ which include estimating functions in the form of $X_j(y - \boldsymbol{\mathcal{X}}\boldsymbol{\beta})$ for each $1 \leq j \leq p$, where elements in $\boldsymbol{\mathcal{X}}$ are imputed if not observed. Then, we have $\|\boldsymbol{U}(\boldsymbol{\beta})\|_2 = 1$ and

$$f^*(\boldsymbol{\beta}) = \mathbf{1}^T \boldsymbol{G}^*(\boldsymbol{\beta}) \left(\boldsymbol{G}^*(\boldsymbol{\beta})^T \boldsymbol{G}^*(\boldsymbol{\beta})\right)^{-1} \boldsymbol{G}^*(\boldsymbol{\beta})^T \mathbf{1}/N + \sum_{i=1}^{p} p_{\lambda_N}(|\beta_i|),$$

where $\mathbf{1}$ is a $N \times 1$ vector of ones.

Let $\boldsymbol{G}_0^*(\boldsymbol{\beta}) = \boldsymbol{G}(\boldsymbol{\beta})[\boldsymbol{U}(\boldsymbol{\beta}^{\mathbf{0}})]^{\boldsymbol{T}}$ and $a_N = N^{-1/2}\zeta_N$. For sufficiently large N, any given constant K_0, and any $\boldsymbol{\beta} \in \mathcal{B}_1 = \{\boldsymbol{\beta} : \|\boldsymbol{\beta} - \boldsymbol{\beta}^0\|_2 \leq K_0 a_N\}$, since $\boldsymbol{G}_0^*(\boldsymbol{\beta})$ has linearly independent columns, we can define

$$f_0^*(\boldsymbol{\beta}) = \mathbf{1}^T \boldsymbol{G}_0^*(\boldsymbol{\beta}) \left(\boldsymbol{G}_0^*(\boldsymbol{\beta})^T \boldsymbol{G}_0^*(\boldsymbol{\beta})\right)^{-1} \boldsymbol{G}_0^*(\boldsymbol{\beta})^T \mathbf{1}/N + \sum_{i=1}^{p} p_{\lambda_N}(|\beta_i|).$$

Let $\mathcal{B}_2 = \{\boldsymbol{\beta} : \beta_j = 0 \text{ if } j \in A_2\}$. Then, $f_0^*(\boldsymbol{\beta}) \leq f^*(\boldsymbol{\beta})$ for any $\boldsymbol{\beta} \in \mathcal{B}_1$, and $f^*(\boldsymbol{\beta}) = f_0^*(\boldsymbol{\beta})$ for any $\boldsymbol{\beta} \in \mathcal{B}_1 \cap \mathcal{B}_2$ by Condition 2.

We first show that $f^*(\boldsymbol{\beta}^0 + a_N \boldsymbol{h}) > f^*(\boldsymbol{\beta}^0)$, where $\boldsymbol{h}$ is a $p \times 1$ vector such that $\|\boldsymbol{h}\|_2 = K_0$ and $\boldsymbol{h}_{A_2} = \mathbf{0}$. Since $(\boldsymbol{\beta}^0 + a_N \boldsymbol{h}) \in \mathcal{B}_1 \cap \mathcal{B}_2$, it suffices to show that $f_0^*(\boldsymbol{\beta}^0 + a_N \boldsymbol{h}) > f_0^*(\boldsymbol{\beta}^0)$.

Let $\boldsymbol{W}_0^*(\boldsymbol{\beta}) = \boldsymbol{G}_0^*(\boldsymbol{\beta})^T \boldsymbol{G}_0^*(\boldsymbol{\beta})/N$ and $L_N(\boldsymbol{\beta}) = \mathbf{1}^T \boldsymbol{G}_0^*(\boldsymbol{\beta}) \boldsymbol{W}_0^*(\boldsymbol{\beta})^{-1} \boldsymbol{G}_0^*(\boldsymbol{\beta})^T \mathbf{1}/N^2$. By Taylor expansion, there exists $\boldsymbol{\beta}^*$ lying on the segment joining $\boldsymbol{\beta}^0 + a_N \boldsymbol{h}$ and $\boldsymbol{\beta}^0$ such that

$$\begin{aligned} f_0^*(\boldsymbol{\beta}^0 + a_N \boldsymbol{h}) - f_0^*(\boldsymbol{\beta}^0) &= a_N \boldsymbol{h}^T \nabla L_N(\boldsymbol{\beta}^0) + \frac{1}{2} a_N^2 \boldsymbol{h}^T \nabla^2 L_N(\boldsymbol{\beta}^*) \boldsymbol{h} \\ &\quad + \sum_{i=1}^{p} \left[p_{\lambda_N}(|\beta_i + a_N h_i|) - p_{\lambda_N}(|\beta_i|)\right]. \end{aligned} \tag{13}$$

For any $1 \le j \le p$,

$$\frac{\partial L_N}{\partial \beta_j}(\boldsymbol{\beta}^0) = 2 \cdot \mathbf{1}^T \frac{\partial \boldsymbol{G}_0^*}{\partial \beta_j}(\boldsymbol{\beta}^0) \boldsymbol{W}_0^*(\boldsymbol{\beta}^0)^{-1} \boldsymbol{G}_0^*(\boldsymbol{\beta}^0)^T \mathbf{1}/N^2 \tag{14}$$

$$- 2 \cdot \mathbf{1}^T \boldsymbol{G}_0^*(\boldsymbol{\beta}^0) \boldsymbol{W}_0^*(\boldsymbol{\beta}^0)^{-1} \frac{\partial \boldsymbol{G}_0^*}{\partial \beta_j}(\boldsymbol{\beta}^0)^T \boldsymbol{G}_0^*(\boldsymbol{\beta}^0) \boldsymbol{W}_0^*(\boldsymbol{\beta}^0)^{-1} \boldsymbol{G}_0^*(\boldsymbol{\beta}^0)^T \mathbf{1}/N^3. \tag{15}$$

By the Central Limit Theorem and missing at random assumption, for any $1 \le r \le R$ and $j \in a(r)$, we have

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N I(\xi_i = r) X_{ij} \varepsilon_i \xrightarrow{d} N(0, c_{jj}\sigma_\varepsilon^2) \text{ as } n_r \to \infty, \tag{16}$$

which implies $\sum_{i \in \mathcal{H}(r)} X_{ij}\varepsilon_i / N = O_p(1/\sqrt{N})$. Similarly, for any $k \in \mathcal{G}(r)$ and $j \in J(r,k)$,

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N I(\xi_i = r) X_{ij} \left[\varepsilon_i + (\boldsymbol{X}_{i,m(r)} - E(\boldsymbol{X}_{i,m(r)} | \boldsymbol{X}_{iJ(r,k)}))\boldsymbol{\beta}_{m(r)}^0\right] \xrightarrow{d} N(0, \sigma_j^2), \tag{17}$$

where $\sigma_j^2 = E[X_j(\varepsilon + (\boldsymbol{\mathcal{X}}_{m(r)} - E(\boldsymbol{\mathcal{X}}_{m(r)} | \boldsymbol{\mathcal{X}}_{J(r,k)}))\boldsymbol{\beta}_{m(r)}^0)]^2$. Then, by Condition 1,

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N I(\xi_i = r) X_{ij} \left[\varepsilon_i + (\boldsymbol{X}_{i,m(r)} - \hat{E}(\boldsymbol{X}_{m(r)} | \boldsymbol{X}_{iJ(r,k)}))\boldsymbol{\beta}_{m(r)}^0\right] = O_p\left(\frac{\zeta_N}{\sqrt{N}}\right).$$

Similarly, for any $j \in m(r)$, we have

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N I(\xi_i = r) \hat{E}(X_j | \boldsymbol{X}_{iJ(r,k)}) \left[\varepsilon_i + (\boldsymbol{X}_{i,m(r)} - \hat{E}(\boldsymbol{X}_{m(r)} | \boldsymbol{X}_{iJ(r,k)}))\boldsymbol{\beta}_{m(r)}^0\right] = O_p\left(\frac{\zeta_N}{\sqrt{N}}\right),$$

by the Central Limit Theorem and Condition 1. Thus, $\|\boldsymbol{G}_0^*(\boldsymbol{\beta}^0)^T \mathbf{1}/N\|_2 = \|\boldsymbol{G}(\boldsymbol{\beta}^0)^T \mathbf{1}/N\|_2 = O_p(\zeta_N/\sqrt{N})$. Since each element in $\boldsymbol{W}_0^*(\boldsymbol{\beta}^0)$ converges in probability to a constant by the Law of Large Numbers and Condition 1, $\|\boldsymbol{W}_0^*(\boldsymbol{\beta}^0)^{-1}\|_2 = O_p(1)$ and $\|\boldsymbol{W}_0^*(\boldsymbol{\beta}^0)\|_2 = O_p(1)$. By the definition of $\boldsymbol{G}_0^*$, we have

$$\left\|\frac{1}{\sqrt{N}}\frac{\partial \boldsymbol{G}_0^*}{\partial \beta_j}(\boldsymbol{\beta}^0)\right\|_2 \leq \left\|\frac{1}{\sqrt{N}}\frac{\partial \boldsymbol{G}}{\partial \beta_j}(\boldsymbol{\beta}^0)\right\|_2 = \sqrt{\lambda_{\max}\left(\frac{1}{N}\frac{\partial \boldsymbol{G}}{\partial \beta_j}(\boldsymbol{\beta}^0)^T\frac{\partial \boldsymbol{G}}{\partial \beta_j}(\boldsymbol{\beta}^0)\right)} = O_p(1), \qquad (18)$$

where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of a matrix. The last equality in (18) follows by the fact that each diagonal element of $\frac{\partial \boldsymbol{G}}{\partial \beta_j}(\boldsymbol{\beta}^0)^T\frac{\partial \boldsymbol{G}}{\partial \beta_j}(\boldsymbol{\beta}^0)/N$ converges in probability to a constant due to the Law of Large Numbers and Condition 1. Thus, the term in (15) is $O_p(\zeta_N/\sqrt{N})$. Let

$$\boldsymbol{H}_{N,1}^*(\boldsymbol{\beta}) = \left(\left[\frac{1}{N}\boldsymbol{1}^T\frac{\partial \boldsymbol{G}_0^*}{\partial \beta_1}(\boldsymbol{\beta})\right]^T, \ldots, \left[\frac{1}{N}\boldsymbol{1}^T\frac{\partial \boldsymbol{G}_0^*}{\partial \beta_p}(\boldsymbol{\beta})\right]^T\right)^T.$$

By the definition of $\boldsymbol{G}_0^*$, we have

$$\|\boldsymbol{H}_{N,1}^*(\boldsymbol{\beta}^0)\|_2 \;\leq\; \left\|\left(\left[\frac{1}{N}\boldsymbol{1}^T\frac{\partial \boldsymbol{G}}{\partial \beta_1}(\boldsymbol{\beta}^0)\right]^T, \ldots, \left[\frac{1}{N}\boldsymbol{1}^T\frac{\partial \boldsymbol{G}}{\partial \beta_p}(\boldsymbol{\beta}^0)\right]^T\right)^T\right\|_2. \qquad (19)$$

The norm on the right-hand side of (19) is $O_p(1)$ since each entry of the matrix converges in probability to a covariance in $\boldsymbol{C}$ by the Law of Large Numbers and Condition 1. It follows that

$$\|\boldsymbol{h}^T\nabla L_N(\boldsymbol{\beta}^0)\|_2 \leq \|\boldsymbol{h}\|_2\|\nabla L_N(\boldsymbol{\beta}^0)\|_2 = O_p\left(\frac{\zeta_N}{\sqrt{N}}\right). \qquad (20)$$

For any $1 \leq j \leq q$, there exists $1 \leq r \leq R$ such that $j \in a(r) \cap [\cup_{k\in\mathcal{G}(r)}a(k)]$ by Condition 3, which implies that rows of $[\boldsymbol{H}_{N,1}^*(\boldsymbol{\beta}^0)]_{A_1,\cdot}$ are linearly independent. Then for any vector $\boldsymbol{v} \neq \boldsymbol{0}$, $\{[\boldsymbol{H}_{N,1}^*(\boldsymbol{\beta}^0)]_{A_1,\cdot}\}^T\boldsymbol{v} \neq \boldsymbol{0}$. Since $\|\boldsymbol{W}_0^*(\boldsymbol{\beta}^0)\|_2 = O_p(1)$ and $\boldsymbol{h}_{A_2} = \boldsymbol{0}$, there exists constant κ_0 such that

$$\begin{aligned}\boldsymbol{h}^T\boldsymbol{H}_{N,1}^*(\boldsymbol{\beta}^0)\boldsymbol{W}_0^*(\boldsymbol{\beta}^0)^{-1}\boldsymbol{H}_{N,1}^*(\boldsymbol{\beta}^0)^T\boldsymbol{h} &= \boldsymbol{h}_{A_1}^T[\boldsymbol{H}_{N,1}^*(\boldsymbol{\beta}^0)]_{A_1,\cdot}\boldsymbol{W}_0^*(\boldsymbol{\beta}^0)^{-1}\left\{[\boldsymbol{H}_{N,1}^*(\boldsymbol{\beta}^0)]_{A_1,\cdot}\right\}^T\boldsymbol{h}_{A_1}\\ &\geq \kappa_0\|\boldsymbol{h}\|_2^2\end{aligned}$$

w.p.a.1. Other terms in $\nabla^2 L_N(\boldsymbol{\beta}^0)$ are dominated by $\boldsymbol{H}_{n,1}^*(\boldsymbol{\beta}^0)\boldsymbol{W}_0^*(\boldsymbol{\beta}^0)^{-1}\boldsymbol{H}_{n,1}^*(\boldsymbol{\beta}^0)^T$, since each of them contains $\boldsymbol{G}_0^*(\boldsymbol{\beta}^0)^T\mathbf{1}/N$ whose norm goes to 0 as $N \to \infty$. Thus, there exists constant κ_1 such that $\boldsymbol{h}^T\nabla^2 L_N(\boldsymbol{\beta}^0)\boldsymbol{h} \geq \kappa_1\|\boldsymbol{h}\|_2^2$ w.p.a.1. Since $\boldsymbol{\beta}^*$ is close to $\boldsymbol{\beta}^0$, $\boldsymbol{h}^T\nabla^2 L_N(\boldsymbol{\beta}^*)\boldsymbol{h} \geq \kappa_1/2\|\boldsymbol{h}\|_2^2 = K_0^2\kappa_1/2$ w.p.a.1. Since $\lambda_N \to 0$, $\max_{j\in A_1}\{p'_{\lambda_N}(|\beta_j^0|)\} = O_p(\zeta_N/\sqrt{n})$ and $\max_{j\in A_1}\{p''_{\lambda_N}(|\beta_j^0|)\} \to 0$ by the definition of SCAD penalty. Thus, by (20), the right-hand side of (13) is dominated by the second term for sufficiently large N. It follows that $f_0^*(\boldsymbol{\beta}^0 + a_N\boldsymbol{h}) > f_0^*(\boldsymbol{\beta}^0)$. Since $f_0^*(\boldsymbol{\beta})$ is continuous, $\min\limits_{\boldsymbol{\beta}\in\mathcal{B}_1\cap\mathcal{B}_2} f_0^*(\boldsymbol{\beta})$ is achieved. The minimizer $\hat{\boldsymbol{\beta}}$ is in the interior of $\mathcal{B}_1 \cap \mathcal{B}_2$ because $f_0^*(\boldsymbol{\beta}^0 + a_N\boldsymbol{h}) > f_0^*(\boldsymbol{\beta}^0)$.

Let $\hat{\boldsymbol{h}} = \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0$. For any $1 \leq j \leq p$, by Taylor expansion of $\partial L_N/\partial\beta_j(\hat{\boldsymbol{\beta}})$ at $\boldsymbol{\beta}^0$,

$$\frac{\partial L_N}{\partial \beta_j}(\hat{\boldsymbol{\beta}}) = \frac{\partial L_N}{\partial \beta_j}(\boldsymbol{\beta}^0) + \hat{\boldsymbol{h}}^T\nabla^2 L_N(\boldsymbol{\beta}^0)\boldsymbol{e}_j + O(\|\hat{\boldsymbol{h}}\|^2),$$

where $\boldsymbol{e}_j$ is a $p\times 1$ vector with 1 in the j-th entry and zeros everywhere else. Since $\hat{\boldsymbol{\beta}}$ is a minimizer of $f_0^*(\boldsymbol{\beta})$ in $\mathcal{B}_1\cap\mathcal{B}_2$, $\partial L_N/\partial\beta_j(\hat{\boldsymbol{\beta}}) = -p'_{\lambda_N}(|\hat{\beta}_j|)\operatorname{sign}(\hat{\beta}_j) = -p'_{\lambda_N}(|\hat{\beta}_j|)\operatorname{sign}(\beta_j^0)$ for any $j \in A_1$. The last equality follows from $\|\hat{\boldsymbol{h}}\|_2 = O(a_N)$. Then, we have

$$\hat{\boldsymbol{h}}_{A_1} = -\left\{[\nabla^2 L_N(\boldsymbol{\beta}^0)]_{A_1A_1}\right\}^{-1}\left\{[\nabla L_N(\boldsymbol{\beta}^0)]_{A_1} + \dot{\mathcal{P}}_{\lambda_N}(\hat{\boldsymbol{\beta}}_{A_1})\operatorname{sign}(\boldsymbol{\beta}^0_{A_1}) + \boldsymbol{r}_N\right\},$$

where $\boldsymbol{r}_N$ is a vector such that $\|\boldsymbol{r}_N\| = O(\|\hat{\boldsymbol{h}}\|^2)$, and $\dot{\mathcal{P}}_{\lambda_N}(\boldsymbol{\beta})$ is a square diagonal matrix with $p'_{\lambda_N}(|\boldsymbol{\beta}|)$ on the main diagonal. Let $\boldsymbol{T}_N = [\nabla^2 L_N(\boldsymbol{\beta}^0)]_{A_2A_1}\left\{[\nabla^2 L_N(\boldsymbol{\beta}^0)]_{A_1A_1}\right\}^{-1}$. Then we have

$$[\nabla L_N(\hat{\boldsymbol{\beta}})]_{A_2} \quad = \quad [\nabla L_N(\boldsymbol{\beta}^0)]_{A_2} + [\nabla^2 L_N(\boldsymbol{\beta}^0)]_{A_2A_1}\hat{\boldsymbol{h}}_{A_1} + \boldsymbol{r}_N. \tag{21}$$

Note that $\left|[\nabla L_N(\hat{\boldsymbol{\beta}})]_{A_2}\right| < \lambda_N\mathbf{1}$, since $\lambda_N\sqrt{n}/\zeta_N \to \infty$, $\|\hat{\boldsymbol{h}}_{A_1}\|_2 = O_p(\zeta_N/\sqrt{N})$, and $\|\nabla L_N(\boldsymbol{\beta}^0)\|_2 = O_p(\zeta_N/\sqrt{N})$, where $\mathbf{1}$ is a $(p-q)\times 1$ vector with all entries 1. Thus, $\hat{\boldsymbol{\beta}}$ is a local minimizer of $f_0^*(\boldsymbol{\beta})$ in $\mathcal{B}_1$. Moreover, $\hat{\boldsymbol{\beta}}$ is a local minimizer of $f^*(\boldsymbol{\beta})$ since $f^*(\hat{\boldsymbol{\beta}}) = f_0^*(\hat{\boldsymbol{\beta}})$.

(iii): By the Taylor expansion on $[\nabla f_0^*(\hat{\boldsymbol{\beta}})]_{A_1}$ at $\boldsymbol{\beta}^0$, we have

$$[\nabla f_0^*(\hat{\boldsymbol{\beta}})]_{A_1} = [\nabla L_N(\boldsymbol{\beta}^0)]_{A_1} + [\nabla^2 L_N(\boldsymbol{\beta}^0)]_{A_1A_1}\hat{\boldsymbol{h}}_{A_1} + \dot{\mathcal{P}}_{\lambda_N}(\boldsymbol{\beta}^0_{A_1})\operatorname{sign}(\boldsymbol{\beta}^0_{A_1}) + \ddot{\mathcal{P}}_{\lambda_N}(\boldsymbol{\beta}^0_{A_1})\hat{\boldsymbol{h}}_{A_1} + \boldsymbol{r}_N,$$

where $\ddot{\mathcal{P}}_{\lambda_N}(\boldsymbol{\beta})$ is a square diagonal matrix with $p''_{\lambda_N}(|\boldsymbol{\beta}|)$ on the main diagonal. Since $\hat{\boldsymbol{\beta}}$ is a local minimizer of $f_0^*(\boldsymbol{\beta})$,

$$\sqrt{N}\hat{\boldsymbol{h}}_{A_1} = -\left\{[\nabla^2 L_N(\boldsymbol{\beta}^0)]_{A_1A_1} + \ddot{\mathcal{P}}_{\lambda_N}(\boldsymbol{\beta}^0_{A_1})\right\}^{-1}\sqrt{N}\left\{[\nabla L_N(\boldsymbol{\beta}^0)]_{A_1} + \dot{\mathcal{P}}_{\lambda_N}(\boldsymbol{\beta}^0_{A_1})\operatorname{sign}(\boldsymbol{\beta}^0_{A_1}) + \boldsymbol{r}_N\right\}.$$

We have $\sqrt{N}\dot{\mathcal{P}}_{\lambda_N}(\boldsymbol{\beta}^0_{A_1})\operatorname{sign}(\boldsymbol{\beta}^0_{A_1}) \xrightarrow{p} \mathbf{0}$, $\ddot{\mathcal{P}}_{\lambda_N}(\boldsymbol{\beta}^0_{A_1}) \xrightarrow{p} \mathbf{0}_{p\times p}$, and $\sqrt{N}\boldsymbol{r}_N \xrightarrow{p} \mathbf{0}$ as $n \to \infty$. Let $\boldsymbol{H}^*_{N,2} = \left(\boldsymbol{v}^T_{N,1}, \ldots, \boldsymbol{v}^T_{N,p}\right)^T$, where $\boldsymbol{v}_{N,j} = \mathbf{1}^T\boldsymbol{G}^*_0(\boldsymbol{\beta}^0)\boldsymbol{W}^*_0(\boldsymbol{\beta}^0)^{-1}\frac{\partial \boldsymbol{G}^*_0}{\partial \beta_j}(\boldsymbol{\beta}^0)^T\boldsymbol{G}^*_0(\boldsymbol{\beta}^0)/N^2$. Then

$$\nabla L_N(\boldsymbol{\beta}^0) = 2\cdot(\boldsymbol{H}^*_{N,1} + \boldsymbol{H}^*_{N,2})\boldsymbol{W}^*_0(\boldsymbol{\beta}^0)^{-1}\boldsymbol{G}^*_0(\boldsymbol{\beta}^0)^T\mathbf{1}/N.$$

By Condition 4,

$$\frac{1}{\sqrt{N}}\boldsymbol{G}^*_0(\boldsymbol{\beta}^0)^T\mathbf{1} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{V}_1).$$

By the Law of Large Numbers, $\boldsymbol{W}^*_0(\boldsymbol{\beta}^0)^{-1} \xrightarrow{p} \boldsymbol{V}_1^{-1}$, $[\boldsymbol{H}^*_{N,1}]_{A_1,\boldsymbol{\cdot}} \xrightarrow{p} \boldsymbol{V}_2$, and $\boldsymbol{H}^*_{N,2} \xrightarrow{p} \mathbf{0}$. Thus,

$$\sqrt{N}[\nabla L_N(\boldsymbol{\beta}^0)]_{A_1} \xrightarrow{d} N(\mathbf{0}, 4\cdot\boldsymbol{V}_2\boldsymbol{V}_1^{-1}\boldsymbol{V}_2^T),$$

Similarly, we have

$$[\nabla^2 L_N(\boldsymbol{\beta}^0)]_{A_1A_1} \xrightarrow{p} 2\cdot\boldsymbol{V}_2\boldsymbol{V}_1^{-1}\boldsymbol{V}_2^T.$$

Therefore, $\sqrt{N}(\hat{\boldsymbol{\beta}}_{A_1} - \boldsymbol{\beta}^0_{A_1}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{V})$.

□

Proof of Proposition 1. We need to show that $\hat{\boldsymbol{V}}_2\hat{\boldsymbol{V}}_1^{-1}\hat{\boldsymbol{V}}_2^T \geq \hat{\boldsymbol{V}}_2^{(1)}(\hat{\boldsymbol{V}}_1^{(1)})^{-1}(\hat{\boldsymbol{V}}_2^{(1)})^T$ in the sense of Loewner ordering for matrices. Since $\hat{\boldsymbol{V}}_1$ and $\hat{\boldsymbol{V}}_1^{(1)}$ are block diagonal matrices,

$$\hat{\boldsymbol{V}}_2\hat{\boldsymbol{V}}_1^{-1}\hat{\boldsymbol{V}}_2^T = \sum_{r=1}^{R}(\hat{\boldsymbol{V}}_2)_{\boldsymbol{\cdot},\mathcal{E}(r)}(\hat{\boldsymbol{V}}_1)_{\mathcal{E}(r)\mathcal{E}(r)}(\hat{\boldsymbol{V}}_2)^T_{\boldsymbol{\cdot},\mathcal{E}(r)},$$

$$\hat{\boldsymbol{V}}_2^{(1)}(\hat{\boldsymbol{V}}_1^{(1)})^{-1}(\hat{\boldsymbol{V}}_2^{(1)})^T = \sum_{r=1}^{R}(\hat{\boldsymbol{V}}_2^{(1)})_{\boldsymbol{\cdot},\mathcal{E}^{(1)}(r)}(\hat{\boldsymbol{V}}_1^{(1)})_{\mathcal{E}^{(1)}(r)\mathcal{E}^{(1)}(r)}(\hat{\boldsymbol{V}}_2^{(1)})^T_{\boldsymbol{\cdot},\mathcal{E}^{(1)}(r)},$$

where $\mathcal{E}(r)$ is the index set for estimating functions of the r-th group in $\boldsymbol{G}_0^*(\boldsymbol{\beta})$, and $\mathcal{E}^{(1)}(r)$ is the index set for estimating functions of the r-th group in $\boldsymbol{g}_i^T\boldsymbol{U}_{(1)}$. Then it suffices to show that for any $1 \leq r \leq R$,

$$(\hat{\boldsymbol{V}}_2)_{\boldsymbol{\cdot},\mathcal{E}(r)}(\hat{\boldsymbol{V}}_1)^{-1}_{\mathcal{E}(r)\mathcal{E}(r)}(\hat{\boldsymbol{V}}_2)^T_{\boldsymbol{\cdot},\mathcal{E}(r)} \geq (\hat{\boldsymbol{V}}_2^{(1)})_{\boldsymbol{\cdot},\mathcal{E}^{(1)}(r)}(\hat{\boldsymbol{V}}_1^{(1)})^{-1}_{\mathcal{E}^{(1)}(r)\mathcal{E}^{(1)}(r)}(\hat{\boldsymbol{V}}_2^{(1)})^T_{\boldsymbol{\cdot},\mathcal{E}^{(1)}(r)}. \tag{22}$$

Without loss of generality, assume that $\boldsymbol{U}_{(1)}$ is a rectangular diagonal matrix whose each main diagonal element is 0 or 1. Since columns of $\boldsymbol{G}_{(1)}(\boldsymbol{\beta})$ are linearly independent with a sufficiently large n, without loss of generality, we assume that each row of $\boldsymbol{U}_{(1)}$ is also a row of $\boldsymbol{U}(\boldsymbol{\beta}^0)$. Let $\boldsymbol{G}_{(2)}(\boldsymbol{\beta})$ be the matrix of estimating equations which are in $\boldsymbol{G}_0^*(\boldsymbol{\beta})$ but not selected by $\boldsymbol{U}_{(1)}$. Let $\mathcal{E}^{(2)}(r)$ be the index set of columns representing estimating functions of the r-th group in $\boldsymbol{G}_{(2)}(\boldsymbol{\beta})$. Then for any $1 \leq r \leq R$, $[\boldsymbol{G}_0^*(\boldsymbol{\beta}^0)]_{\mathcal{H}(r)\mathcal{E}(r)} = ([\boldsymbol{G}_{(1)}(\boldsymbol{\beta}^0)]_{\mathcal{H}(r)\mathcal{E}^{(1)}(r)}, [\boldsymbol{G}_{(2)}(\boldsymbol{\beta}^0)]_{\mathcal{H}(r)\mathcal{E}^{(2)}(r)})$. We orthogonalize $[\boldsymbol{G}_{(2)}(\boldsymbol{\beta}^0)]_{\mathcal{H}(r)\mathcal{E}^{(2)}(r)}$ from $[\boldsymbol{G}_{(1)}(\boldsymbol{\beta}^0)]_{\mathcal{H}(r)\mathcal{E}^{(1)}(r)}$ as

$$\boldsymbol{G}_2(\boldsymbol{\beta}^0) = [\boldsymbol{G}_{(2)}(\boldsymbol{\beta}^0)]_{\mathcal{H}(r)\mathcal{E}^{(2)}(r)} - [\boldsymbol{G}_{(1)}(\boldsymbol{\beta}^0)]_{\mathcal{H}(r)\mathcal{E}^{(1)}(r)}(\hat{\boldsymbol{V}}_1^{(1)})^{-1}_{\mathcal{E}^{(1)}(r)\mathcal{E}^{(1)}(r)}\boldsymbol{C}_{21},$$

where $\boldsymbol{C}_{21} = \frac{1}{n_r}[\boldsymbol{G}_{(1)}(\boldsymbol{\beta}^0)]^T_{\mathcal{H}(r)\mathcal{E}^{(1)}(r)}[\boldsymbol{G}_{(2)}(\boldsymbol{\beta}^0)]_{\mathcal{H}(r)\mathcal{E}^{(2)}(r)}$. Then, $\boldsymbol{G}_2(\boldsymbol{\beta}^0)^T[\boldsymbol{G}_{(1)}(\boldsymbol{\beta}^0)]_{\mathcal{H}(r)\mathcal{E}^{(1)}(r)} = \boldsymbol{0}$. Let $\boldsymbol{Q}$ be a matrix such as $[\boldsymbol{G}_0^*(\boldsymbol{\beta}^0)]_{\mathcal{H}(r)\mathcal{E}(r)}\boldsymbol{Q} = \big([\boldsymbol{G}_{(1)}(\boldsymbol{\beta}^0)]_{\mathcal{H}(r)\mathcal{E}^{(1)}(r)}, \boldsymbol{G}_2(\boldsymbol{\beta}^0)\big)$. We have

$$\begin{aligned}
& (\hat{\boldsymbol{V}}_2)_{\boldsymbol{\cdot},\mathcal{E}(r)}(\hat{\boldsymbol{V}}_1)^{-1}_{\mathcal{E}(r)\mathcal{E}(r)}(\hat{\boldsymbol{V}}_2)^T_{\boldsymbol{\cdot},\mathcal{E}(r)} \\
= \ & (\hat{\boldsymbol{V}}_2)_{\boldsymbol{\cdot},\mathcal{E}(r)}\boldsymbol{Q}\left[\boldsymbol{Q}(\hat{\boldsymbol{V}}_1)_{\mathcal{E}(r)\mathcal{E}(r)}\boldsymbol{Q}\right]^{-1}\boldsymbol{Q}^T(\hat{\boldsymbol{V}}_2)^T_{\boldsymbol{\cdot},\mathcal{E}(r)} \\
= \ & (\hat{\boldsymbol{V}}_2^{(1)})_{\boldsymbol{\cdot},\mathcal{E}^{(1)}(r)}(\hat{\boldsymbol{V}}_1^{(1)})^{-1}_{\mathcal{E}^{(1)}(r)\mathcal{E}^{(1)}(r)}(\hat{\boldsymbol{V}}_2^{(1)})^T_{\boldsymbol{\cdot},\mathcal{E}^{(1)}(r)} + \hat{\boldsymbol{V}}_2^{(2)}(\hat{\boldsymbol{V}}_1^{(2)})^{-1}(\hat{\boldsymbol{V}}_2^{(2)})^T,
\end{aligned} \tag{23}$$

where $\hat{\boldsymbol{V}}_1^{(2)} = \boldsymbol{G}_2(\boldsymbol{\beta}^0)^T\boldsymbol{G}_2(\boldsymbol{\beta}^0)/N$ and $\hat{\boldsymbol{V}}_2^{(2)} = [\nabla(\boldsymbol{1}^T\boldsymbol{G}_2(\boldsymbol{\beta}^0)/N)]_{A_1,\boldsymbol{\cdot}}$. Since the second term in (23) is a non-negative positive matrix, (22) holds in the sense of Loewner ordering for matrices, which implies $\hat{\boldsymbol{V}} = (\hat{\boldsymbol{V}}_2\hat{\boldsymbol{V}}_1^{-1}\hat{\boldsymbol{V}}_2^T)^{-1} \leq [\hat{\boldsymbol{V}}_2^{(1)}(\hat{\boldsymbol{V}}_1^{(1)})^{-1}(\hat{\boldsymbol{V}}_2^{(1)})^T]^{-1} = \hat{\boldsymbol{V}}^{(1)}$. This completes the proof.

□

Proof of Theorem 2. We use the same notations as in the proof of Theorem 1. For each $1 \leq r \leq R$ and $k \in \mathcal{G}(k)$, let $\boldsymbol{\varepsilon}^{(r,k)}$ be the error vector in the model (1) for Group r. Note that $\boldsymbol{\varepsilon}^{(r,1)} = \cdots = \boldsymbol{\varepsilon}^{(r,M_r)}$. For each $1 \leq r \leq R$ and $k \in \mathcal{G}(k)$, let $\hat{\boldsymbol{X}}^{(r,k)}$ be the completed data in Group r using imputed values based on Group k, and $\bar{\boldsymbol{X}}^{(r,k)}$ be the samples in Group r but with missing values in the i-th sample replaced by $E(\boldsymbol{\mathcal{X}}_{m(r)}|\boldsymbol{X}_{iJ(r,k)})$. Let $\boldsymbol{X}_0^{(r,k)}$ be the same as $\hat{\boldsymbol{X}}^{(r,k)}$ except that the missing values in Group r are replaced by corresponding true values. Let

$$\hat{\boldsymbol{X}} = \left(\left[\hat{\boldsymbol{X}}^{(1,1)}\right]^T, \left[\hat{\boldsymbol{X}}^{(1,k_1)}\right]^T, \ldots, \left[\hat{\boldsymbol{X}}^{(R,k_R)}\right]^T\right)^T,$$

$$\bar{\boldsymbol{X}} = \left(\left[\bar{\boldsymbol{X}}^{(1,1)}\right]^T, \left[\bar{\boldsymbol{X}}^{(1,k_1)}\right]^T, \ldots, \left[\bar{\boldsymbol{X}}^{(R,k_R)}\right]^T\right)^T,$$

$$\boldsymbol{X}_0 = \left(\left[\boldsymbol{X}_0^{(1,1)}\right]^T, \left[\boldsymbol{X}_0^{(1,k_1)}\right]^T, \ldots, \left[\boldsymbol{X}_0^{(R,k_R)}\right]^T\right)^T,$$

$$\boldsymbol{\varepsilon}_0 = \left(\left[\boldsymbol{\varepsilon}^{(1,1)}\right]^T, \left[\boldsymbol{\varepsilon}^{(1,k_1)}\right]^T, \ldots, \left[\boldsymbol{\varepsilon}^{(R,k_R)}\right]^T\right)^T,$$

$$\bar{\boldsymbol{X}}_D = diag\left\{\bar{\boldsymbol{X}}_{a(1)}^{(1,1)}, \bar{\boldsymbol{X}}_{a(1)}^{(1,k_1)}, \ldots, \bar{\boldsymbol{X}}_{a(k_R)}^{(R,k_R)}\right\},$$

and

$$\hat{\boldsymbol{X}}_D = diag\left\{\hat{\boldsymbol{X}}_{a(1)}^{(1,1)}, \hat{\boldsymbol{X}}_{a(1)}^{(2,1)}, \ldots, \hat{\boldsymbol{X}}_{a(k_R)}^{(R,k_R)}\right\},$$

where k_r is the last element in $\mathcal{G}(r)$ for $1 \leq r \leq R$, and $\hat{\boldsymbol{X}}_{a(k)}^{(r,k)}$ represents covariates (columns) indexed by $a(k)$ in $\hat{\boldsymbol{X}}^{(r,k)}$.

We first show that $f^*(\boldsymbol{\beta}^0 + a_N\boldsymbol{h}) > f^*(\boldsymbol{\beta}^0)$, where $a_N = N^{-\kappa_0}\zeta_N$, $\boldsymbol{h}$ is a $p_N \times 1$ vector such that $\|\boldsymbol{h}\|_\infty = K_0$ and $\boldsymbol{h}_{A_2} = \boldsymbol{0}$. Let $\mathcal{B}_1 = \{\boldsymbol{\beta} : \|\boldsymbol{\beta} - \boldsymbol{\beta}^0\|_\infty \leq K_0 a_N\}$ and $\mathcal{B}_2 = \{\boldsymbol{\beta} : \beta_j = 0 \text{ if } j \in A_2\}$. Similarly as in the proof of Theorem 1, since $(\boldsymbol{\beta}^0 + a_N\boldsymbol{h}) \in \mathcal{B}_1 \cap \mathcal{B}_2$, it suffices to show that $f_0^*(\boldsymbol{\beta}^0 + a_N\boldsymbol{h}) > f_0^*(\boldsymbol{\beta}^0)$.

We have

$$\nabla L_N(\boldsymbol{\beta}^0) \;=\; \frac{2}{N}\boldsymbol{H}(\boldsymbol{\beta}^0)[\boldsymbol{U}(\boldsymbol{\beta}^0)]^T\boldsymbol{W}_0^*(\boldsymbol{\beta}^0)^{-1}\boldsymbol{U}(\boldsymbol{\beta}^0)\boldsymbol{G}(\boldsymbol{\beta}^0)^T\boldsymbol{1} - 2\boldsymbol{R}_N(\boldsymbol{\beta}^0), \tag{24}$$

where $\boldsymbol{R}_N(\boldsymbol{\beta})$ is a $p_N \times 1$ vector whose j-th element is $\mathbf{1}^T \boldsymbol{G}_0^*(\boldsymbol{\beta}) \boldsymbol{W}_0^*(\boldsymbol{\beta})^{-1} \frac{\partial \boldsymbol{G}_0^*}{\partial \beta_j}(\boldsymbol{\beta})^T \boldsymbol{G}_0^*(\boldsymbol{\beta}) \boldsymbol{W}_0^*(\boldsymbol{\beta})^{-1}$ $\boldsymbol{G}_0^*(\boldsymbol{\beta})^T \mathbf{1}/N^3$.

By Condition 7, elements in $\bar{\boldsymbol{X}}_D$ and $\left(\boldsymbol{X}_0 - \bar{\boldsymbol{X}}\right)$ are all sub-Gaussian distributed. Due to the missing at random assumption, $E\left(\boldsymbol{X}_0 - \bar{\boldsymbol{X}}\right) = \mathbf{0}$. Let $\mathcal{E}_m$ be the set of indexes of non-zero columns in $\left(\boldsymbol{X}_0 - \bar{\boldsymbol{X}}\right)$. Let $n_m = |\mathcal{E}_m|$. Then $n_m \leq \sum_{r=1}^R |m(r)| k_r$. Let $[\bar{\boldsymbol{X}}_D]_{\cdot,j}$ be the j-th column in $\bar{\boldsymbol{X}}_D$. For each $i \in \mathcal{E}_m$ and j, there exist constants c_0, c_2, c_3, and c_4 such that

$$
\begin{aligned}
& P\left(\left|\{[\bar{\boldsymbol{X}}_D]_{\cdot,j}\}^T [\boldsymbol{X}_0 - \bar{\boldsymbol{X}}]_{\cdot,i} \beta_i^0\right| > c_0 \frac{N^{7/6-\kappa_1}\zeta_N}{n_m+1}\right) \\
\leq\ & P\left(\frac{\left|\{[\bar{\boldsymbol{X}}_D]_{\cdot,j}\}^T [\boldsymbol{X}_0 - \bar{\boldsymbol{X}}]_{\cdot,i}\right|}{\left\|[\bar{\boldsymbol{X}}_D]_{\cdot,j}\right\|_2} > c_0 \frac{N^{7/6-\kappa_1}\zeta_N}{(n_m+1)|\beta_i^0| \left\|[\bar{\boldsymbol{X}}_D]_{\cdot,j}\right\|_2}, \left\|[\bar{\boldsymbol{X}}_D]_{\cdot,j}\right\|_2 \leq \sqrt{N}\right) \\
& + P\left(\left\|[\bar{\boldsymbol{X}}_D]_{\cdot,j}\right\|_2 > \sqrt{N}\right) \\
\leq\ & P\left(\left|\{[\bar{\boldsymbol{X}}_D]_{\cdot,j}\}^T [\boldsymbol{X}_0 - \bar{\boldsymbol{X}}]_{\cdot,i} \beta_i^0\right| > c_0 \left\|[\bar{\boldsymbol{X}}_D]_{\cdot,j}\right\|_2 \frac{N^{4/6-\kappa_1}\zeta_N}{(n_m+1)|\beta_i^0|}\right) + P\left(\left\|[\bar{\boldsymbol{X}}_D]_{\cdot,j}\right\|_2 > \sqrt{N}\right) \\
\leq\ & c_2 \exp\left(-c_3 \frac{N^{4/3-2\kappa_1}\zeta_N^2}{(n_m+1)^2 |\beta_i^0|^2}\right) + c_2 \exp\left(-c_4 N\right)
\end{aligned} \tag{25}
$$

The last inequality in (25) follows from properties of sub-Gaussian random vectors [30, Theorem 1]. Similarly, we can show that

$$
P\left(\left|\{[\bar{\boldsymbol{X}}_D]_{\cdot,j}\}^T \boldsymbol{\varepsilon}_0\right| > c_0 \frac{N^{7/6-\kappa_1}\zeta_N}{n_m+1}\right) \leq c_2 \exp\left(-c_3 \frac{N^{4/3-2\kappa_1}\zeta_N^2}{(n_m+1)^2}\right) + c_2 \exp\left(-c_4 N\right). \tag{26}
$$

Since $\max\limits_{1\leq r\leq R} |m(r)| = o(N^{1/6})$ and $\max\limits_{1\leq r\leq R} \|\boldsymbol{\beta}_{m(r)}\|_\infty = O(\zeta_N)$, we have $n_m = o(N^{1/6})$ and $|\beta_i^0| = O(\zeta_N)$ for each $i \in \mathcal{E}_m$. Then,

$$
\begin{aligned}
& P\left(\left\|\bar{\boldsymbol{X}}_D^T\left[\boldsymbol{\varepsilon}_0 + \left(\boldsymbol{X}_0 - \bar{\boldsymbol{X}}\right)\boldsymbol{\beta}^0\right]/N\right\|_\infty > c_0 N^{1/6-\kappa_1}\zeta_N\right) \\
\leq\ & c_1 p_N \cdot P\left(\{[\bar{\boldsymbol{X}}_D]_{\cdot,j}\}^T\left[\boldsymbol{\varepsilon}_0 + \left(\boldsymbol{X}_0 - \bar{\boldsymbol{X}}\right)\boldsymbol{\beta}^0\right] > c_0 N^{7/6-\kappa_1}\zeta_N\right) \\
\leq\ & c_1 p_N \cdot P\left(\left|\{[\bar{\boldsymbol{X}}_D]_{\cdot,j}\}^T \boldsymbol{\varepsilon}_0\right| > c_0 \frac{N^{7/6-\kappa_1}\zeta_N}{n_m+1}\right) \\
& + c_1 p_N \cdot \sum_{i\in\mathcal{E}_m} P\left(\left|\{[\bar{\boldsymbol{X}}_D]_{\cdot,j}\}^T [\boldsymbol{X}_0 - \bar{\boldsymbol{X}}]_{\cdot,i} \beta_i^0\right| > c_0 \frac{N^{7/6-\kappa_1}\zeta_N}{n_m+1}\right) \\
=\ & o(1).
\end{aligned} \tag{27}
$$

The last equality in (27) follows from $\log p_N = O(N^{1-2\kappa_1})$, $n_m = o(N^{1/6})$ and $|\beta_i^0| = O(\zeta_N)$. Thus, we have $\left\|\bar{\boldsymbol{X}}_D^T\left[\boldsymbol{\varepsilon}_0 + \left(\boldsymbol{X}_0 - \bar{\boldsymbol{X}}\right)\boldsymbol{\beta}^0\right]/N\right\|_\infty = O_p(N^{1/6-\kappa_1}\zeta_N)$. Since

$$
\begin{aligned}
\boldsymbol{G}(\boldsymbol{\beta}^0)^T\mathbf{1}/N &= \left(\hat{\boldsymbol{X}}_D^T - \bar{\boldsymbol{X}}_D^T\right)\left[\boldsymbol{\varepsilon}_0 + \left(\boldsymbol{X}_0 - \bar{\boldsymbol{X}}\right)\boldsymbol{\beta}^0\right]/N + \bar{\boldsymbol{X}}_D^T\left[\boldsymbol{\varepsilon}_0 + \left(\boldsymbol{X}_0 - \bar{\boldsymbol{X}}\right)\boldsymbol{\beta}^0\right]/N \\
&\quad + \hat{\boldsymbol{X}}_D^T\left[\bar{\boldsymbol{X}} - \hat{\boldsymbol{X}}\right]\boldsymbol{\beta}^0/N,
\end{aligned}
$$

$\|\boldsymbol{G}(\boldsymbol{\beta}^0)^T\mathbf{1}/N\|_\infty = O_p(N^{1/6-\kappa_1}\zeta_N)$ by Condition 7. By Condition 5, the first term in (24) dominates $\boldsymbol{R}_N(\boldsymbol{\beta}^0)$. Thus, $\|\nabla L_N(\boldsymbol{\beta}^0)\|_\infty = O_p(N^{1/6-\kappa_1+2\kappa_3}\zeta_N)$. Also, by Condition 5, other terms in $\boldsymbol{h}^T\nabla^2 L_N(\boldsymbol{\beta}^*)\boldsymbol{h}$ are dominated by $\boldsymbol{h}^T\boldsymbol{H}(\boldsymbol{\beta}^*)[\boldsymbol{U}(\boldsymbol{\beta}^*)]^T\boldsymbol{W}_0^*(\boldsymbol{\beta}^*)^{-1}\boldsymbol{U}(\boldsymbol{\beta}^*)[\boldsymbol{H}(\boldsymbol{\beta}^*)]^T\boldsymbol{h} \geq \kappa_4\|\boldsymbol{h}\|_2^2 \geq \kappa_4 K_0^2$ w.p.a.1. The last term in (13) is

$$
\begin{aligned}
\left|\sum_{i=1}^{p_N}\left[p_{\lambda_N}(|\beta_i + a_N h_i|) - p_{\lambda_N}(|\beta_i|)\right]\right| &\leq \sum_{i\in A_1}|p_{\lambda_N}(|\beta_i + a_N h_i|) - p_{\lambda_N}(|\beta_i|)| \\
&\leq a_N\|\boldsymbol{h}\|_\infty\sum_{i\in A_1}p'_{\lambda_N}(\beta_i^*) \\
&\leq K_0 a_N q_N p'_{\lambda_N}\left(\frac{\beta_{\min}}{2}\right),
\end{aligned}
$$

where β^* sits in between $\beta_i + a_N h_i$ and β_i. By Condition 6, $q_N p'_{\lambda_N}(\beta_{\min}/2) = O_p(a_N)$. Thus, the right-hand side of (13) is dominated by the second term for sufficiently large N. It follows that $f_0^*(\boldsymbol{\beta}^0 + a_N\boldsymbol{h}) > f_0^*(\boldsymbol{\beta}^0)$. Since $f_0^*(\boldsymbol{\beta})$ is continuous, $\min_{\boldsymbol{\beta}\in\mathcal{B}_1\cap\mathcal{B}_2} f_0^*(\boldsymbol{\beta})$ is achieved. The minimizer $\hat{\boldsymbol{\beta}}$ is in the interior of $\mathcal{B}_1\cap\mathcal{B}_2$ because $f_0^*(\boldsymbol{\beta}^0 + a_N\boldsymbol{h}) > f_0^*(\boldsymbol{\beta}^0)$. Since $[\nabla^2 L_N(\boldsymbol{\beta})]_{A_1A_1}$ is positive definite for $\boldsymbol{\beta}\in\mathcal{B}_1\cap\mathcal{B}_2$ by Condition 5, $\hat{\boldsymbol{\beta}}$ is a strict local minimizer.

Next, we will show that $\|[\nabla L_N(\hat{\boldsymbol{\beta}})]_{A_2}\|_\infty < \lambda_N$. Let $\hat{\boldsymbol{h}} = \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0$. For any $1 \leq j \leq p_N$, by Taylor expansion of $\partial L_N/\partial\beta_j(\hat{\boldsymbol{\beta}})$ at $\boldsymbol{\beta}^0$,

$$
\frac{\partial L_N}{\partial\beta_j}(\hat{\boldsymbol{\beta}}) = \frac{\partial L_N}{\partial\beta_j}(\boldsymbol{\beta}^0) + \hat{\boldsymbol{h}}^T\nabla^2 L_N(\boldsymbol{\beta}^*)\boldsymbol{e}_j,
$$

where $\boldsymbol{e}_j$ is a $p_N\times 1$ vector with 1 in the j-th entry and zeros everywhere else, and $\boldsymbol{\beta}^*$ is on the segment joining $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}^0$. Since $\hat{\boldsymbol{\beta}}$ is a minimizer of $f_0^*(\boldsymbol{\beta})$ in $\mathcal{B}_1\cap\mathcal{B}_2$, $\partial L_N/\partial\beta_j(\hat{\boldsymbol{\beta}}) =$

$-p'_{\lambda_N}(|\hat{\beta}_j|)\,\mathrm{sign}(\hat{\beta}_j) = -p'_{\lambda_N}(|\hat{\beta}_j|)\,\mathrm{sign}(\beta_j^0)$ for any $j \in A_1$ w.p.a.1. The last equality follows from $\|\hat{\boldsymbol{h}}\|_\infty = O_p(a_N)$ and $\beta_{\min} > N^{-\kappa_0} \log N$.

$$\hat{\boldsymbol{h}}_{A_1} = -\left\{[\nabla^2 L_N(\boldsymbol{\beta}^*)]_{A_1A_1}\right\}^{-1} \left\{[\nabla L_N(\boldsymbol{\beta}^0)]_{A_1} + \dot{\mathcal{P}}_{\lambda_N}(\hat{\boldsymbol{\beta}}_{A_1})\,\mathrm{sign}(\boldsymbol{\beta}^0_{A_1})\right\},$$

where $\dot{\mathcal{P}}_{\lambda_N}(\boldsymbol{\beta})$ is a square diagonal matrix with $p'_{\lambda_N}(|\boldsymbol{\beta}|)$ on the main diagonal. Let $\boldsymbol{T}_N = [\nabla^2 L_N(\boldsymbol{\beta}^*)]_{A_2A_1} \left\{[\nabla^2 L_N(\boldsymbol{\beta}^*)]_{A_1A_1}\right\}^{-1}$, Then we have

$$[\nabla L_N(\hat{\boldsymbol{\beta}})]_{A_2} = [\nabla L_N(\boldsymbol{\beta}^0)]_{A_2} - \boldsymbol{T}_N[\nabla L_N(\boldsymbol{\beta}^0)]_{A_1} + \boldsymbol{T}_N\dot{\mathcal{P}}_{\lambda_N}(\hat{\boldsymbol{\beta}}_{A_1})\,\mathrm{sign}(\boldsymbol{\beta}^0_{A_1}).$$

Since $\|\nabla L_N(\boldsymbol{\beta}^0)\|_\infty = O_p(N^{1/6-\kappa_1+2\kappa_3}\zeta_N)$, $\left|[\nabla L_N(\hat{\boldsymbol{\beta}})]_{A_2}\right| < \lambda_N \mathbf{1}$ by Condition 6, where $\mathbf{1}$ is a $(p_N - q_N) \times 1$ vector with all entries 1. Thus, $\hat{\boldsymbol{\beta}}$ is a strict local minimizer of $f_0^*(\boldsymbol{\beta})$ in $\mathcal{B}_1$ for sufficiently large N. Moreover, $\hat{\boldsymbol{\beta}}$ is a strict local minimizer of $f^*(\boldsymbol{\beta})$ since $f^*(\hat{\boldsymbol{\beta}}) = f_0^*(\hat{\boldsymbol{\beta}})$.

□