Biometrics 0000, 1–14 DOI: 000
00 0000

Supporting Information for "Hierarchical Cancer Heterogeneity Analysis Based On Histopathological Imaging Features" by Mingyang Ren, Qingzhao Zhang, Sanguo Zhang, Tingyan Zhong, Jian Huang, and Shuangge Ma

This paper has been submitted for consideration for publication in *Biometrics*

A. Proof of Theorem 1

Recall that the penalized objective function is:

$$Q(\boldsymbol{\beta},\boldsymbol{\gamma}) = \frac{1}{2}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta}_i - \boldsymbol{z}_i^\top\boldsymbol{\gamma}_i)^2 + \sum_{1\leqslant j<m\leqslant n} p\left(\sqrt{\|\boldsymbol{\beta}_j-\boldsymbol{\beta}_m\|_2^2 + \|\boldsymbol{\gamma}_j-\boldsymbol{\gamma}_m\|_2^2}, \lambda_1\right) + \sum_{1\leqslant j<m\leqslant n} p\left(\|\boldsymbol{\beta}_j-\boldsymbol{\beta}_m\|_2, \lambda_2\right), \tag{A.1}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top,\cdots,\boldsymbol{\beta}_n^\top)^\top$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^\top,\cdots,\boldsymbol{\gamma}_n^\top)^\top$, and $p(\cdot,\lambda)$ is a concave penalty function with tuning parameter $\lambda > 0$.

A.1 *Notations and preparation*

It is noted that some of the following notations, definitions, and conditions have been provided in the main text. We also provide them here for completeness.

Denote the true values of parameters as $\boldsymbol{\beta}^* = (\boldsymbol{\beta}_1^{*\top},\cdots,\boldsymbol{\beta}_n^{*\top})^\top$ and $\boldsymbol{\gamma}^* = (\boldsymbol{\gamma}_1^{*\top},\cdots,\boldsymbol{\gamma}_n^{*\top})^\top$. Let $\{\boldsymbol{\alpha}_1^*,,\cdots,\boldsymbol{\alpha}_{K_1}^*\}$ and $\{\boldsymbol{\delta}_1^*,,\cdots,\boldsymbol{\delta}_{K_2}^*\}$ be the distinct values of $\boldsymbol{\beta}^*$ and $\boldsymbol{\gamma}^*$, respectively, and $\mathcal{T}_{k_2}^* = \{i : \boldsymbol{\gamma}_i^* = \boldsymbol{\delta}_{k_2}^*\}, 1 \leqslant k_2 \leqslant K_2$. With the hierarchical structure, there exists a mutually exclusive partition of $\{1,\cdots,K_2\}$: $\{\mathcal{H}_1,\cdots,\mathcal{H}_{K_1}\}$ satisfying $\boldsymbol{\beta}_i^* = \boldsymbol{\alpha}_{k_1}^*$, if $i \in \mathcal{G}_{k_1}^*$, where $\mathcal{G}_{k_1}^* = \bigcup_{k_2\in\mathcal{H}_{k_1}} \mathcal{T}_{k_2}^*$, $1 \leqslant k_1 \leqslant K_1$. Then, each of $\{\mathcal{T}_1^*,\cdots,\mathcal{T}_{K_2}^*\}$ and $\{\mathcal{G}_1^*,\cdots,\mathcal{G}_{K_1}^*\}$ constitutes a mutually exclusive partition of $\{1,\cdots,n\}$. Denote $\boldsymbol{\alpha}^* = (\boldsymbol{\alpha}_1^{*\top},\cdots,\boldsymbol{\alpha}_{K_1}^{*\top})^\top$ and $\boldsymbol{\delta}^* = (\boldsymbol{\delta}_1^{*\top},\cdots,\boldsymbol{\delta}_{K_2}^{*\top})^\top$. Denote the minimal differences of the common values between two (sub-)subgroups as $b_1 = \min\limits_{1\leqslant k\neq k'\leqslant K_2}\|\boldsymbol{\delta}_k^* - \boldsymbol{\delta}_{k'}^*\|_2 = \min_{j\in\mathcal{T}_k^*, m\in\mathcal{T}_{k'}^*, 1\leqslant k\neq k'\leqslant K_2}\|\boldsymbol{\gamma}_j^* - \boldsymbol{\gamma}_m^*\|_2$ and $b_2 = \min\limits_{1\leqslant k\neq k'\leqslant K_1}\|\boldsymbol{\alpha}_k^* - \boldsymbol{\alpha}_{k'}^*\|_2 = \min_{j\in\mathcal{G}_k^*, m\in\mathcal{G}_{k'}^*, 1\leqslant k\neq k'\leqslant K_1}\|\boldsymbol{\beta}_j^* - \boldsymbol{\beta}_m^*\|_2$. Write $|\mathcal{T}_{\min}| = \min\limits_{1\leqslant k_2\leqslant K_2}|\mathcal{T}_{k_2}^*|$ and $|\mathcal{G}_{\min}| = \min\limits_{1\leqslant k_1\leqslant K_1}|\mathcal{G}_{k_1}^*|$, where $|\cdot|$ is the cardinality of a set. Define:

$$\mathcal{M}_{\mathcal{G}^*} = \{\boldsymbol{\beta}\in\mathbb{R}^{nq} : \boldsymbol{\beta}_i = \boldsymbol{\beta}_j, \text{ for any } i,j\in\mathcal{G}_{k_1}^*, 1\leqslant k_1\leqslant K_1\},$$

$$\mathcal{M}_{\mathcal{T}^*} = \{\boldsymbol{\gamma}\in\mathbb{R}^{np} : \boldsymbol{\gamma}_i = \boldsymbol{\gamma}_j, \text{ for any } i,j\in\mathcal{T}_{k_2}^*, 1\leqslant k_2\leqslant K_2\}.$$

Let $\widetilde{\boldsymbol{W}}_1 = \{w_{ik_1}^{(1)}\}$ be an $n\times K_1$ matrix with $w_{ik_1}^{(1)} = 1$ for $i\in\mathcal{G}_{k_1}^*$ and $w_{ik_1}^{(1)} = 0$ otherwise, $\widetilde{\boldsymbol{W}}_2 = \{w_{ik_2}^{(2)}\}$ be an $n\times K_2$ matrix with $w_{ik_2}^{(2)} = 1$ for $i\in\mathcal{T}_{k_2}^*$ and $w_{ik_2}^{(2)} = 0$ otherwise, $\boldsymbol{W}_1 =$

$\widetilde{\boldsymbol{W}}_1 \otimes \boldsymbol{I}_q$, and $\boldsymbol{W}_2 = \widetilde{\boldsymbol{W}}_2 \otimes \boldsymbol{I}_p$. So each $\boldsymbol{\beta} \in \mathcal{M}_{\mathcal{G}^*}$ can be rewritten as $\boldsymbol{\beta} = \boldsymbol{W}_1\boldsymbol{\alpha}$, where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^\top, \cdots, \boldsymbol{\alpha}_{K_1}^\top)^\top$, and $\boldsymbol{\alpha}_{k_1}$ is the $q \times 1$ vector of the k_1th subgroup-specific parameter for $k_1 = 1, \cdots, K_1$. Similarly, each $\boldsymbol{\gamma} \in \mathcal{M}_{\mathcal{T}^*}$ can be rewritten as $\boldsymbol{\gamma} = \boldsymbol{W}_2\boldsymbol{\delta}$. Denote $\boldsymbol{y} = (y_1, \cdots, y_n)^\top$, $\boldsymbol{X} = \mathrm{diag}(\boldsymbol{x}_1^\top, \cdots, \boldsymbol{x}_n^\top)$, and $\boldsymbol{Z} = \mathrm{diag}(\boldsymbol{z}_1^\top, \cdots, \boldsymbol{z}_n^\top)$. Define $\boldsymbol{U} = (\boldsymbol{X}\boldsymbol{W}_1, \boldsymbol{Z}\boldsymbol{W}_2)$.

Consider the oracle estimator for $(\boldsymbol{\beta}, \boldsymbol{\gamma})$, under which the true hierarchical subgrouping memberships $\{\mathcal{G}_1^*, \cdots, \mathcal{G}_{K_1}^*\}$ and $\{\mathcal{T}_1^*, \cdots, \mathcal{T}_{K_2}^*\}$ are known. Specifically,

$$(\widehat{\boldsymbol{\beta}}^o, \widehat{\boldsymbol{\gamma}}^o) = \underset{\boldsymbol{\beta} \in \mathcal{M}_{\mathcal{G}^*}, \boldsymbol{\gamma} \in \mathcal{M}_{\mathcal{T}^*}}{\operatorname{argmin}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{\gamma}\|_2^2, \tag{A.2}$$

and correspondingly, the oracle estimators for the common coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\delta}$ are:

$$(\widehat{\boldsymbol{\alpha}}^o, \widehat{\boldsymbol{\delta}}^o) = \underset{\boldsymbol{\alpha} \in \mathbb{R}^{K_1 q}, \boldsymbol{\delta} \in \mathbb{R}^{K_2 p}}{\operatorname{argmin}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{W}_1\boldsymbol{\alpha} - \boldsymbol{Z}\boldsymbol{W}_2\boldsymbol{\delta}\|_2^2. \tag{A.3}$$

Define the 2-norm of a q-dimensional vector $\boldsymbol{d} = (d_1, \cdots, d_q)^\top$ as $\|\boldsymbol{d}\|_2 = \sqrt{\sum_{j=1}^q d_j^2}$. For a matrix $\boldsymbol{A}$, denote its smallest/largest eigenvalue as $\psi_{\min}(\boldsymbol{A})/\psi_{\max}(\boldsymbol{A})$ and 2-norm as $\|\boldsymbol{A}\|_2 = \psi_{\max}^{1/2}(\boldsymbol{A}^\top\boldsymbol{A})$. The following conditions are assumed.

Condition A.1: $\rho(t) = \lambda^{-1}p(t, \lambda)$ is concave in $t \in [0, \infty)$ with a continuous derivative $\rho'(t)$ satisfying $\rho(0+) = 1$, and $\rho'(0+)$ is independent of λ. There exists a constant $0 < a < \infty$ such that $\rho(t)$ is constant for all $t \geqslant a\lambda$.

Condition A.2: $\boldsymbol{\epsilon} = (\epsilon_1, \cdots, \epsilon_n)^\top$ is sub-Gaussian. That is, $\mathrm{pr}(|\boldsymbol{d}^\top\boldsymbol{\epsilon}| > \|\boldsymbol{d}\|_2 v) \leqslant 2\exp(-c_1 v^2)$ for any vector $\boldsymbol{d} \in \mathbb{R}^n$ and $v > 0$, where $0 < c_1 < \infty$.

Condition A.3: $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$, $i = 1, \cdots, n$ are bounded almost surely, and $\psi_{\min}(\boldsymbol{U}^\top\boldsymbol{U}) \geqslant C_0|\mathcal{T}_{\min}|$ for some constant $C_0 > 0$.

To prove Theorem 1, it is equivalent to establishing the following Results 1 and 2.

Result 1: *Suppose that Conditions A.1 - A.3 hold. Assume* $|\mathcal{T}_{\min}| \gg \sqrt{(K_1 q + K_2 p) n \log n}$. *Then, with probability tending to 1,*

$$\sup_i \|\widehat{\boldsymbol{\beta}}_i^o - \boldsymbol{\beta}_i^*\|_2 + \sup_i \|\widehat{\boldsymbol{\gamma}}_i^o - \boldsymbol{\gamma}_i^*\|_2 = O_p(\tau_n),$$

where $\tau_n = |\mathcal{T}_{\min}|^{-1}\sqrt{(K_1 q + K_2 p) n \log n}$.

Result 2: *Assume that* $b_1 > (a+0.5)\lambda_1 \gg \tau_n$, $b_2 > (a+0.5)\lambda_2 \gg \tau_n$, *where* a *is the regularization parameter of the concave penalty defined in Condition A.1, and that the conditions in Result 1 hold. Then there exists* $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}})$, *a local minimizer of* $Q(\boldsymbol{\beta}, \boldsymbol{\gamma})$ *defined in (A.1), that satisfies:*

$$\mathrm{pr}\left((\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}) = (\widehat{\boldsymbol{\beta}}^o, \widehat{\boldsymbol{\gamma}}^o)\right) \to 1.$$

A.2 *Proof of Result 1 in Theorem 1*

Recall that:

$$\begin{aligned}
(\widehat{\boldsymbol{\alpha}}^{o\top}, \widehat{\boldsymbol{\delta}}^{o\top})^\top &= \underset{\boldsymbol{\alpha}\in\mathbb{R}^{K_1 q}, \boldsymbol{\delta}\in\mathbb{R}^{K_2 p}}{\operatorname{argmin}} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{W}_1\boldsymbol{\alpha} - \boldsymbol{Z}\boldsymbol{W}_2\boldsymbol{\delta}\|_2^2 \\
&= \underset{\boldsymbol{\alpha}\in\mathbb{R}^{K_1 q}, \boldsymbol{\delta}\in\mathbb{R}^{K_2 p}}{\operatorname{argmin}} \frac{1}{2}\left\|\boldsymbol{y} - \boldsymbol{U}(\boldsymbol{\alpha}^\top, \boldsymbol{\delta}^\top)^\top\right\|_2^2 \\
&= (\boldsymbol{U}^\top\boldsymbol{U})^{-1}\boldsymbol{U}^\top\boldsymbol{y},
\end{aligned}$$

where $\boldsymbol{U} = (\boldsymbol{X}\boldsymbol{W}_1, \boldsymbol{Z}\boldsymbol{W}_2)$. Then,

$$\left((\widehat{\boldsymbol{\alpha}}^o - \boldsymbol{\alpha}^*)^\top, (\widehat{\boldsymbol{\delta}}^o - \boldsymbol{\delta}^*)^\top\right)^\top = (\boldsymbol{U}^\top\boldsymbol{U})^{-1}\boldsymbol{U}^\top\boldsymbol{\epsilon},$$

and

$$\sqrt{\|\widehat{\boldsymbol{\alpha}}^o - \boldsymbol{\alpha}^*\|_2^2 + \|\widehat{\boldsymbol{\delta}}^o - \boldsymbol{\delta}^*\|_2^2} \leqslant \|(\boldsymbol{U}^\top\boldsymbol{U})^{-1}\|_2\|\boldsymbol{U}^\top\boldsymbol{\epsilon}\|_2. \tag{A.4}$$

By Condition A.3,

$$\|(\boldsymbol{U}^\top\boldsymbol{U})^{-1}\|_2 = \psi_{\min}^{-1}(\boldsymbol{U}^\top\boldsymbol{U}) \leqslant C_0^{-1}|\mathcal{T}_{\min}|^{-1}. \tag{A.5}$$

In addition,

$$\mathrm{pr}(\|\boldsymbol{U}^\top\boldsymbol{\epsilon}\|_\infty > C\sqrt{n\log n}) \leqslant \mathrm{pr}(\|(\boldsymbol{X}\boldsymbol{W}_1)^\top\boldsymbol{\epsilon}\|_\infty > C\sqrt{n\log n}) + \mathrm{pr}(\|(\boldsymbol{Z}\boldsymbol{W}_2)^\top\boldsymbol{\epsilon}\|_\infty > C\sqrt{n\log n}),$$

for some constant $0 < C < \infty$. By Condition A.3, $\sum_{i=1}^n x_{ij}^2 I(i \in \mathcal{G}_{k_1}^*) = O(|\mathcal{G}_{k_1}^*|)$. That is, for $j = 1, \cdots, q$, there exists $0 < c_j < \infty$ satisfying $\sum_{i=1}^n x_{ij}^2 I(i \in \mathcal{G}_{k_1}^*) = c_j'|\mathcal{G}_{k_1}^*|$. Moreover,

$$\|(\boldsymbol{X}\boldsymbol{W}_1)^\top\boldsymbol{\epsilon}\|_\infty = \sup_{1\leqslant j\leqslant q, 1\leqslant k_1\leqslant K_1} |\sum_{i=1}^n x_{ij}\epsilon_i I(i \in \mathcal{G}_{k_1}^*)|. \tag{A.6}$$

By (A.6), Condition A.2, and Condition A.3, we have

$$
\begin{aligned}
&\mathrm{pr}\left(\|(\boldsymbol{X}\boldsymbol{W}_1)^\top\boldsymbol{\epsilon}\|_\infty > C\sqrt{n\log n}\right)\\
&\leqslant \sum_{1\leqslant j\leqslant q, 1\leqslant k_1\leqslant K_1} \mathrm{pr}\left(|\sum_{i=1}^n x_{ij}\epsilon_i I(i\in\mathcal{G}_{k_1}^*)| > C\sqrt{n\log n}\right)\\
&\leqslant \sum_{1\leqslant j\leqslant q, 1\leqslant k_1\leqslant K_1} \mathrm{pr}\left(|\sum_{i=1}^n x_{ij}\epsilon_i I(i\in\mathcal{G}_{k_1}^*)| > \sqrt{c_j'|\mathcal{G}_{k_1}^*|}C\sqrt{c_j^{-1}}\sqrt{\log n}\right)\\
&\leqslant 2K_1 q n^{-C_1},
\end{aligned}
$$

where $C_1 = c_1C^2/c'$, $c' = \max_{1\leqslant j\leqslant q} c_j'$. Similarly, $\mathrm{pr}\left(\|(\boldsymbol{Z}\boldsymbol{W}_2)^\top\boldsymbol{\epsilon}\|_\infty > C\sqrt{n\log n}\right) \leqslant 2K_2pn^{-C_2}$ for some constant $0 < C_2 < \infty$. So,

$$
\mathrm{pr}(\|\boldsymbol{U}^\top\boldsymbol{\epsilon}\|_\infty > C\sqrt{n\log n}) \leqslant 2(K_1q + K_2p)n^{-\min\{C_1,C_2\}}.
$$

Note that $\|\boldsymbol{U}^\top\boldsymbol{\epsilon}\|_2 \leqslant \sqrt{K_1q+K_2p}\|\boldsymbol{U}^\top\boldsymbol{\epsilon}\|_\infty$. Then

$$
\begin{aligned}
\mathrm{pr}(\|\boldsymbol{U}^\top\boldsymbol{\epsilon}\|_2 > C\sqrt{(K_1q+K_2p)n\log n}) &\leqslant \mathrm{pr}(\|\boldsymbol{U}^\top\boldsymbol{\epsilon}\|_\infty > C\sqrt{n\log n})\\
&\leqslant 2(K_1q+K_2p)n^{-\min\{C_1,C_2\}}.
\end{aligned}
\tag{A.7}
$$

By (A.4), (A.5), and (A.7), with probability at least $1 - 2(K_1q+K_2p)n^{-\min\{C_1,C_2\}}$, we have

$$
\sqrt{\|\widehat{\boldsymbol{\alpha}}^o - \boldsymbol{\alpha}^*\|_2^2 + \|\widehat{\boldsymbol{\delta}}^o - \boldsymbol{\delta}^*\|_2^2} \leqslant CC_0^{-1}|\mathcal{T}_{\min}|^{-1}\sqrt{(K_1q+K_2p)n\log n}.
$$

Then,

$$
\sup_i \|\widehat{\boldsymbol{\beta}}_i^o - \boldsymbol{\beta}_i^*\|_2 \leqslant \sup_k \|\widehat{\boldsymbol{\alpha}}_k^o - \boldsymbol{\alpha}_k^*\|_2 \leqslant \|\widehat{\boldsymbol{\alpha}}^o - \boldsymbol{\alpha}^*\|_2 \leqslant \tau_n',
$$

where $\tau_n' = CC_0^{-1}|\mathcal{T}_{\min}|^{-1}\sqrt{(K_1q+K_2p)n\log n}$. Similarly, $\sup_i \|\widehat{\boldsymbol{\gamma}}_i^o - \boldsymbol{\gamma}_i^*\|_2 \leqslant \tau_n'$. This concludes the proof of Result 1.

A.3 *Proof of Result 2 in Theorem 1*

Define

$$
Q(\boldsymbol{\beta},\boldsymbol{\gamma}) = \ell(\boldsymbol{\beta},\boldsymbol{\gamma}) + P(\boldsymbol{\beta},\boldsymbol{\gamma}),\ \ Q^o(\boldsymbol{\beta},\boldsymbol{\gamma}) = \ell^o(\boldsymbol{\beta},\boldsymbol{\gamma}) + P^o(\boldsymbol{\beta},\boldsymbol{\gamma}),
$$

where

$$\ell(\boldsymbol{\beta},\boldsymbol{\gamma})=\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Z}\boldsymbol{\gamma}\|_2^2,\ P(\boldsymbol{\beta},\boldsymbol{\gamma})=P_1(\boldsymbol{\beta},\boldsymbol{\gamma})+P_2(\boldsymbol{\beta},\boldsymbol{\gamma}),$$

$$P_1(\boldsymbol{\beta},\boldsymbol{\gamma})=\sum_{1\leqslant j<m\leqslant n}p\left(\sqrt{\|\boldsymbol{\beta}_j-\boldsymbol{\beta}_m\|_2^2+\|\boldsymbol{\gamma}_j-\boldsymbol{\gamma}_m\|_2^2},\lambda_1\right),$$

$$P_2(\boldsymbol{\beta},\boldsymbol{\gamma})=\sum_{1\leqslant j<m\leqslant n}p\left(\|\boldsymbol{\beta}_j-\boldsymbol{\beta}_m\|_2,\lambda_2\right),$$

$$\ell^o(\boldsymbol{\beta},\boldsymbol{\gamma})=\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Z}\boldsymbol{\gamma}\|_2^2,\ \text{s.t. } \boldsymbol{\beta}\in\mathcal{M}_{\mathcal{G}^*},\boldsymbol{\gamma}\in\mathcal{M}_{\mathcal{T}^*},$$

$$P^o(\boldsymbol{\beta},\boldsymbol{\gamma})=P_1^o(\boldsymbol{\beta},\boldsymbol{\gamma})+P_2^o(\boldsymbol{\beta}),$$

$$P_1^o(\boldsymbol{\beta},\boldsymbol{\gamma})=\sum_{1\leqslant j<m\leqslant n}p\left(\sqrt{\|\boldsymbol{\beta}_j-\boldsymbol{\beta}_m\|_2^2+\|\boldsymbol{\gamma}_j-\boldsymbol{\gamma}_m\|_2^2},\lambda_1\right),\ \text{s.t. } \boldsymbol{\beta}\in\mathcal{M}_{\mathcal{G}^*},\boldsymbol{\gamma}\in\mathcal{M}_{\mathcal{T}^*}$$

$$P_2^o(\boldsymbol{\beta})=\sum_{1\leqslant j<m\leqslant n}p\left(\|\boldsymbol{\beta}_j-\boldsymbol{\beta}_m\|_2,\lambda_2\right),\ \text{s.t. } \boldsymbol{\beta}\in\mathcal{M}_{\mathcal{G}^*}.$$

Let $G_1:\mathcal{M}_{\mathcal{G}^*}\to\mathbb{R}^{K_1q}$ be the mapping such that $G_1(\boldsymbol{\beta})$ is the $K_1q\times 1$ vector consisting of K_1 vectors with dimension q, and its k_1th vector component equals the common value of $\boldsymbol{\beta}_i$ for $i\in\mathcal{G}_{k_1}^*$. Let $\breve{G}_1:\mathbb{R}^{nq}\to\mathbb{R}^{K_1q}$ be the mapping such that $\breve{G}_1(\boldsymbol{\beta})=\{|\mathcal{G}_{k_1}^*|^{-1}\sum_{i\in\mathcal{G}_{k_1}^*}\boldsymbol{\beta}_i^\top,k_1=1,\cdots,K_1\}^\top$. Clearly, when $\beta\in\mathcal{M}_{\mathcal{G}^*}$, $G_1(\boldsymbol{\beta})=\breve{G}_1(\boldsymbol{\beta})$. Similarly, define $G_2:\mathcal{M}_{\mathcal{T}^*}\to\mathbb{R}^{K_2p}$ and $\breve{G}_2(\boldsymbol{\gamma})=\{|\mathcal{T}_{k_2}^*|^{-1}\sum_{i\in\mathcal{T}_{k_2}^*}\boldsymbol{\gamma}_i^\top,k_2=1,\cdots,K_2\}^\top$. For any $\boldsymbol{\beta}\in\mathbb{R}^{nq}$, $\boldsymbol{\gamma}\in\mathbb{R}^{np}$, denote $\breve{\boldsymbol{\beta}}=G_1^{-1}\left(\breve{G}_1(\boldsymbol{\beta})\right)$, $\breve{\boldsymbol{\gamma}}=G_2^{-1}\left(\breve{G}_2(\boldsymbol{\gamma})\right)$. Clearly, $\breve{\boldsymbol{\beta}}\in\mathcal{M}_{\mathcal{G}^*}$, $\breve{\boldsymbol{\gamma}}\in\mathcal{M}_{\mathcal{T}^*}$, and

$$Q(\breve{\boldsymbol{\beta}},\breve{\boldsymbol{\gamma}})=Q^o(\breve{\boldsymbol{\beta}},\breve{\boldsymbol{\gamma}}).\tag{A.8}$$

Consider the neighborhood of $(\boldsymbol{\beta}^*,\boldsymbol{\gamma}^*)$:

$$\mathcal{C}=\{\boldsymbol{\beta}\in\mathbb{R}^{qn},\boldsymbol{\gamma}\in\mathbb{R}^{pn}:\sup_i\|\boldsymbol{\beta}_i-\boldsymbol{\beta}_i^*\|_2\leqslant\tau_n',\sup_i\|\boldsymbol{\gamma}_i-\boldsymbol{\gamma}_i^*\|_2\leqslant\tau_n'\}.$$

By Result 1, there exists an event E_1 in which:

$$\sup_i\|\widehat{\boldsymbol{\beta}}_i^o-\boldsymbol{\beta}_i^*\|_2\leqslant\tau_n',\sup_i\|\widehat{\boldsymbol{\gamma}}_i^o-\boldsymbol{\gamma}_i^*\|_2\leqslant\tau_n',i=1,\cdots,n,$$

and $\mathrm{pr}(E_1^C)\leqslant 2(K_1q+K_2p)n^{-\min\{C_1,C_2\}}$. Thus $(\widehat{\boldsymbol{\beta}}^{o\top},\widehat{\boldsymbol{\gamma}}^{o\top})^\top\in\mathcal{C}$. With the following two steps, we show that $(\widehat{\boldsymbol{\beta}}^{o\top},\widehat{\boldsymbol{\gamma}}^{o\top})^\top$ is a strict local minimizer of objective function (A.1) with probability approaching 1.

(i) On event E_1, for any $(\breve{\boldsymbol{\beta}}^\top,\breve{\boldsymbol{\gamma}}^\top)\in\mathcal{C}$ and $(\breve{\boldsymbol{\beta}}^\top,\breve{\boldsymbol{\gamma}}^\top)\neq(\widehat{\boldsymbol{\beta}}^{o\top},\widehat{\boldsymbol{\gamma}}^{o\top})^\top$, $Q(\breve{\boldsymbol{\beta}},\breve{\boldsymbol{\gamma}})>Q(\widehat{\boldsymbol{\beta}}^o,\widehat{\boldsymbol{\gamma}}^o)$.

(ii) There is an event E_2 such that $\mathrm{pr}(E_2^C) \leqslant 2n^{-1}$. On event $E_1 \cap E_2$, there is a neighborhood of $(\widehat{\boldsymbol{\beta}}^{o\top}, \widehat{\boldsymbol{\gamma}}^{o\top})^\top$, denoted by $\mathcal{C}_n$, such that $Q(\boldsymbol{\beta}, \boldsymbol{\gamma}) \geqslant Q(\breve{\boldsymbol{\beta}}, \breve{\boldsymbol{\gamma}})$, for any $(\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top \in \mathcal{C} \cap \mathcal{C}_n$ and a sufficiently large n.

With the results in (i) and (ii), for any $(\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top \in \mathcal{C} \cap \mathcal{C}_n$ and $(\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top \neq (\widehat{\boldsymbol{\beta}}^{o\top}, \widehat{\boldsymbol{\gamma}}^{o\top})^\top$ on $\mathcal{C} \cap \mathcal{C}_n$, we have $Q(\boldsymbol{\beta}, \boldsymbol{\gamma}) > Q(\widehat{\boldsymbol{\beta}}^{o}, \widehat{\boldsymbol{\gamma}}^{o})$. So $(\widehat{\boldsymbol{\beta}}^{o\top}, \widehat{\boldsymbol{\gamma}}^{o\top})^\top$ is a strict local minimizer of (A.1) on event $\mathcal{C} \cap \mathcal{C}_n$ with $\mathrm{pr}(E_1 \cap E_2) \geqslant 1 - 2(K_1 q + K_2 p + 1)n^{-\min\{C_1, C_2, 1\}}$ for a sufficiently large n.

First, we prove the result in (i). For any $\boldsymbol{\beta} \in \mathcal{C}$, let $\breve{G}_1(\boldsymbol{\beta}) = (\boldsymbol{\alpha}_1^\top, \cdots, \boldsymbol{\alpha}_{K_1}^\top)^\top$. Then $\|\boldsymbol{\alpha}_{k_1} - \boldsymbol{\alpha}_{k_1'}\|_2 \geqslant \|\boldsymbol{\alpha}_{k_1}^* - \boldsymbol{\alpha}_{k_1'}^*\|_2 - 2\sup_{k_1} \|\boldsymbol{\alpha}_{k_1} - \boldsymbol{\alpha}_{k_1}^*\|_2$, and

$$\begin{aligned}
\sup_{k_1} \left\|\boldsymbol{\alpha}_{k_1} - \boldsymbol{\alpha}_{k_1}^*\right\|_2^2 &= \sup_{k_1} \||\mathcal{G}_{k_1}^*|^{-1} \sum_{i \in \mathcal{T}_{k_1}^*} \boldsymbol{\beta}_i - \boldsymbol{\alpha}_{k_1}^*\|_2^2 = \sup_{k_1} \||\mathcal{G}_{k_1}^*|^{-1} \sum_{i \in \mathcal{G}_{k_1}^*} (\boldsymbol{\beta}_i - \boldsymbol{\beta}_i^*)\|_2^2 \\
&= \sup_{k_1} \left|\mathcal{G}_{k_1}^*\right|^{-2} \|\sum_{i \in \mathcal{G}_{k_1}^*} (\boldsymbol{\beta}_i - \boldsymbol{\beta}_i^*)\|_2^2 \leqslant \sup_{k_1} \left|\mathcal{G}_{k_1}^*\right|^{-1} \sum_{i \in \mathcal{G}_{k_1}^*} \|\boldsymbol{\beta}_i - \boldsymbol{\beta}_i^*\|_2^2 \qquad \text{(A.9)} \\
&\leqslant \sup_i \|\boldsymbol{\beta}_i - \boldsymbol{\beta}_i^*\|_2^2 \leqslant \tau_n'^2.
\end{aligned}$$

Note that $b_2 = \min_{1 \leqslant k_1 \neq k_1' \leqslant K_1} \|\boldsymbol{\alpha}_{k_1}^* - \boldsymbol{\alpha}_{k_1'}^*\|_2$ is sufficiently large such that $b_2 > (a + 0.5)\lambda_2$ and $\lambda_2 \gg \tau_n'$. Then for any k_1 and k_1', $\|\boldsymbol{\alpha}_{k_1} - \boldsymbol{\alpha}_{k_1'}\|_2 \geqslant b_2 - 2\tau_n' > a\lambda_2$. So $p(\|\boldsymbol{\alpha}_{k_1} - \boldsymbol{\alpha}_{k_1'}\|_2, \lambda_2)$ is a constant. Also note that for any $\boldsymbol{\beta} \in \mathcal{M}_{\mathcal{G}^*}$,

$$P_2^o(\boldsymbol{\beta}) = \sum_{1 \leqslant j < m \leqslant n} p\left(\|\boldsymbol{\beta}_j - \boldsymbol{\beta}_m\|_2, \lambda_2\right) = \sum_{1 \leqslant k_1 < k_1' \leqslant K_1} |\mathcal{G}_{k_1}^*||\mathcal{G}_{k_1'}^*| p\left(\|\boldsymbol{\alpha}_{k_1'} - \boldsymbol{\alpha}_{k_1'}\|_2, \lambda_2\right).$$

So $P_2^o(\boldsymbol{\beta})$ is a constant that does not depend on $\boldsymbol{\beta}$, for any $\boldsymbol{\beta} \in \mathcal{C} \cap \mathcal{M}_{\mathcal{G}^*}$.

Moreover, for any $\boldsymbol{\beta} \in \mathcal{C} \cap \mathcal{M}_{\mathcal{G}^*}$ and $\boldsymbol{\gamma} \in \mathcal{C} \cap \mathcal{M}_{\mathcal{T}^*}$, let $\breve{G}_2(\boldsymbol{\gamma}) = (\boldsymbol{\delta}_1^\top, \cdots, \boldsymbol{\delta}_{K_2}^\top)^\top$. We have

$$\begin{aligned}
P_1^o(\boldsymbol{\beta}, \boldsymbol{\gamma}) &= \sum_{1 \leqslant j < m \leqslant n} p\left(\sqrt{\|\boldsymbol{\beta}_j - \boldsymbol{\beta}_m\|_2^2 + \|\boldsymbol{\gamma}_j - \boldsymbol{\gamma}_m\|_2^2}, \lambda_1\right) \\
&= \sum_{1 \leqslant k_2 < k_2' \leqslant K_2} |\mathcal{T}_{k_2}^*||\mathcal{T}_{k_2'}^*| p\left(\sqrt{\|\boldsymbol{\alpha}_{h(k_2)} - \boldsymbol{\alpha}_{h(k_2')}\|_2^2 + \|\boldsymbol{\delta}_{k_2} - \boldsymbol{\delta}_{k_2'}\|_2^2}, \lambda_1\right),
\end{aligned}$$

where $h(k_2) = \sum_{k_1=1}^{K_1} k_1 I(k_2 \in \mathcal{H}_{k_1})$. Similarly, $P_1^o(\boldsymbol{\beta}, \boldsymbol{\gamma})$ is a constant that does not depend on $\boldsymbol{\beta}$ or $\boldsymbol{\gamma}$. Therefore, $P^o(\boldsymbol{\beta}, \boldsymbol{\gamma}) = P_1^o(\boldsymbol{\beta}, \boldsymbol{\gamma}) + P_2^o(\boldsymbol{\beta})$ is a constant for any $\boldsymbol{\beta} \in \mathcal{C} \cap \mathcal{M}_{\mathcal{G}^*}$ and $\boldsymbol{\gamma} \in \mathcal{C} \cap \mathcal{M}_{\mathcal{T}^*}$.

Since $(\widehat{\boldsymbol{\beta}}^{o\top}, \widehat{\boldsymbol{\gamma}}^{o\top})^\top$ is the unique minimizer of $\ell^o(\boldsymbol{\beta}, \boldsymbol{\gamma})$, $\ell^o(\breve{\boldsymbol{\beta}}, \breve{\boldsymbol{\gamma}}) > \ell^o(\widehat{\boldsymbol{\beta}}^o, \widehat{\boldsymbol{\gamma}}^o)$ for any $(\breve{\boldsymbol{\beta}}^\top, \breve{\boldsymbol{\gamma}}^\top) \in \mathcal{C}$ and $(\breve{\boldsymbol{\beta}}^\top, \breve{\boldsymbol{\gamma}}^\top)^\top \neq (\widehat{\boldsymbol{\beta}}^{o\top}, \widehat{\boldsymbol{\gamma}}^{o\top})^\top$. Note that $\breve{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}}^o \in \mathcal{M}_{\mathcal{G}^*}$, and $\breve{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\gamma}}^o \in \mathcal{M}_{\mathcal{T}^*}$. So $P^o(\breve{\boldsymbol{\beta}}, \breve{\boldsymbol{\gamma}}) = P^o(\widehat{\boldsymbol{\beta}}^o, \widehat{\boldsymbol{\gamma}}^o)$, and we have $Q^o(\breve{\boldsymbol{\beta}}, \breve{\boldsymbol{\gamma}}) > Q^o(\widehat{\boldsymbol{\beta}}^o, \widehat{\boldsymbol{\gamma}}^o)$. By (A.8), $Q(\breve{\boldsymbol{\beta}}, \breve{\boldsymbol{\gamma}}) > Q(\widehat{\boldsymbol{\beta}}^o, \widehat{\boldsymbol{\gamma}}^o)$. Therefore, the result in (i) is proved.

Next, we prove the result in (ii). Given a positive sequence ϕ_n, consider $\mathcal{C}_n = \{\boldsymbol{\beta} \in \mathbb{R}^{nq}, \boldsymbol{\gamma} \in \mathbb{R}^{np} : \sup_i \|\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i^o\|_2 \leqslant \phi_n, \sup_i \|\boldsymbol{\gamma}_i - \widehat{\boldsymbol{\gamma}}_i^o\|_2 \leqslant \phi_n\}$. For $(\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top \in \mathcal{C} \cap \mathcal{C}_n$ and $(\breve{\boldsymbol{\beta}}^\top, \breve{\boldsymbol{\gamma}}^\top)^\top \in \mathcal{C}$, by Taylor's expansion,

$$Q(\boldsymbol{\beta}, \boldsymbol{\gamma}) - Q(\breve{\boldsymbol{\beta}}, \breve{\boldsymbol{\gamma}}) = q_1 + q_2 + q_3 + q_4,$$

where

$$\begin{aligned}
q_1 &= -(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \boldsymbol{Z}\widetilde{\boldsymbol{\gamma}})^\top \boldsymbol{X}(\boldsymbol{\beta} - \breve{\boldsymbol{\beta}}), \\
q_2 &= -(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \boldsymbol{Z}\widetilde{\boldsymbol{\gamma}})^\top \boldsymbol{Z}(\boldsymbol{\gamma} - \breve{\boldsymbol{\gamma}}), \\
q_3 &= \sum_{i=1}^n \frac{\partial P(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\gamma}})}{\partial \boldsymbol{\beta}_i^\top} (\boldsymbol{\beta}_i - \breve{\boldsymbol{\beta}}_i), \\
q_4 &= \sum_{i=1}^n \frac{\partial P(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\gamma}})}{\partial \boldsymbol{\gamma}_i^\top} (\boldsymbol{\gamma}_i - \breve{\boldsymbol{\gamma}}_i),
\end{aligned}$$

$\widetilde{\boldsymbol{\beta}} = \varsigma\boldsymbol{\beta} + (1-\varsigma)\breve{\boldsymbol{\beta}}$, and $\widetilde{\boldsymbol{\gamma}} = \varsigma\boldsymbol{\gamma} + (1-\varsigma)\breve{\boldsymbol{\gamma}}$ for some constant $\varsigma \in (0,1)$.

First, we consider q_1 and q_2. Let $\boldsymbol{D}_{1j} = (y_j - \boldsymbol{x}_j^\top \widetilde{\boldsymbol{\beta}}_j - \boldsymbol{z}_j^\top \widetilde{\boldsymbol{\gamma}}_j)\boldsymbol{x}_j$, $j = 1, \cdots, n$. q_1 can be rewritten as:

$$\begin{aligned}
q_1 &= -\sum_{j=1}^n \boldsymbol{D}_{1j}^\top (\boldsymbol{\beta}_j - \breve{\boldsymbol{\beta}}_j) = -\sum_{k_1=1}^{K_1} \sum_{j \in \mathcal{G}_{k_1}^*} \sum_{m \in \mathcal{G}_{k_1}^*} \frac{\boldsymbol{D}_{1j}^\top (\boldsymbol{\beta}_j - \boldsymbol{\beta}_m + \boldsymbol{\beta}_m - \breve{\boldsymbol{\beta}}_j)}{|\mathcal{G}_{k_1}^*|} \\
&= -\left[\sum_{k_1=1}^{K_1} \sum_{j,m \in \mathcal{G}_{k_1}^*} \frac{\boldsymbol{D}_{1j}^\top (\boldsymbol{\beta}_j - \boldsymbol{\beta}_m)}{|\mathcal{G}_{k_1}^*|} + \sum_{k_1=1}^{K_1} \sum_{j \in \mathcal{G}_{k_1}^*} \boldsymbol{D}_{1j}^\top \left(\sum_{m \in \mathcal{G}_{k_1}^*} \frac{\boldsymbol{\beta}_m}{|\mathcal{G}_{k_1}^*|} - \breve{\boldsymbol{\beta}}_j\right)\right].
\end{aligned}$$

Note that for $j, m \in \mathcal{G}^*_{k_1}$, $\breve{\boldsymbol{\beta}}_j = \breve{\boldsymbol{\beta}}_m = \sum_{m\in\mathcal{G}^*_{k_1}} \frac{\boldsymbol{\beta}_m}{|\mathcal{G}^*_{k_1}|}$. So

$$\begin{aligned}
q_1 &= -\sum_{k_1=1}^{K_1}\sum_{j,m\in\mathcal{G}^*_{k_1}} \frac{\boldsymbol{D}_{1j}^\top(\boldsymbol{\beta}_j-\boldsymbol{\beta}_m)}{|\mathcal{G}^*_{k_1}|} \\
&= -\left[\sum_{k_1=1}^{K_1}\sum_{j,m\in\mathcal{G}^*_{k_1}} \frac{\boldsymbol{D}_{1j}^\top(\boldsymbol{\beta}_j-\boldsymbol{\beta}_m)}{2|\mathcal{G}^*_{k_1}|} + \sum_{k_1=1}^{K_1}\sum_{j,m\in\mathcal{G}^*_{k_1}} \frac{\boldsymbol{D}_{1m}^\top(\boldsymbol{\beta}_m-\boldsymbol{\beta}_j)}{2|\mathcal{G}^*_{k_1}|}\right] \\
&= -\sum_{k_1=1}^{K_1}\sum_{j,m\in\mathcal{G}^*_{k_1},j<m} \frac{(\boldsymbol{D}_{1j}-\boldsymbol{D}_{1m})^\top(\boldsymbol{\beta}_j-\boldsymbol{\beta}_m)}{|\mathcal{G}^*_{k_1}|} \\
&\geqslant -\sum_{k_1=1}^{K_1}\sum_{j,m\in\mathcal{G}^*_{k_1},j<m} 2|\mathcal{G}_{\min}|^{-1}\sup_j\|\boldsymbol{D}_{1j}\|_2\|\boldsymbol{\beta}_j-\boldsymbol{\beta}_m\|_2.
\end{aligned} \tag{A.10}$$

Moreover,

$$\boldsymbol{D}_{1j} = (y_j - \boldsymbol{x}_j^\top\widetilde{\boldsymbol{\beta}}_j - \boldsymbol{z}_j^\top\widetilde{\boldsymbol{\gamma}}_j)\boldsymbol{x}_j = [\epsilon_j - \boldsymbol{x}_j^\top(\boldsymbol{\beta}^*_j-\widetilde{\boldsymbol{\beta}}_j) - \boldsymbol{z}_j^\top(\boldsymbol{\gamma}^*_j-\widetilde{\boldsymbol{\gamma}}_j)]\boldsymbol{x}_j.$$

Then, $\sup_j\|\boldsymbol{D}_{1j}\|_2 \leqslant \sup_j\left\{\|\boldsymbol{x}_j\|_2\left(|\epsilon_j| + \|\boldsymbol{x}_j\|_2\|\boldsymbol{\beta}^*_j-\widetilde{\boldsymbol{\beta}}_j\|_2 + \|\boldsymbol{z}_j\|_2\|\boldsymbol{\gamma}^*_j-\widetilde{\boldsymbol{\gamma}}_j\|_2\right)\right\}$.

Note that $\sup_j\|\breve{\boldsymbol{\beta}}_j-\boldsymbol{\beta}^*_j\|_2 = \sup_{k_1}\|\boldsymbol{\alpha}_{k_1}-\boldsymbol{\alpha}^*_{k_1}\|_2 \leqslant \tau_n'$, following from (A.9). And then

$$\sup_j\|\widetilde{\boldsymbol{\beta}}_j-\boldsymbol{\beta}^*_j\|_2 \leqslant \varsigma\sup_j\|\boldsymbol{\beta}_j-\boldsymbol{\beta}^*_j\|_2 + (1-\varsigma)\sup_j\|\breve{\boldsymbol{\beta}}_j-\boldsymbol{\beta}^*_j\|_2 \leqslant \varsigma\tau_n' + (1-\varsigma)\tau_n' = \tau_n'. \tag{A.11}$$

Similarly, $\sup_j\|\widetilde{\boldsymbol{\gamma}}_j-\boldsymbol{\gamma}^*_j\|_2 \leqslant \tau_n'$. By Condition A.2, $\mathrm{pr}(\|\boldsymbol{\epsilon}\|_\infty > \sqrt{2(\log n)/c_1}) \leqslant \sum_{i=1}^n \mathrm{pr}(|\epsilon_i| > \sqrt{2(\log n)/c_1}) \leqslant 2/n$. By Condition A.3, $\sqrt{2(\log n)/c_1} \gg \tau_n'\max\{\sup_i\|\boldsymbol{x}_i\|_2, \sup_i\|\boldsymbol{z}_i\|_2\}$. Therefore, there is an event E_2 such that $\mathrm{pr}(E_2^C) \leqslant 2/n$, and on E_2,

$$\sup_j\|\boldsymbol{D}_{1j}\|_2 \leqslant C_{D_1}\sqrt{\log n}, \tag{A.12}$$

for some constant $0 < C_{D_1} < \infty$. Combining (A.10) and (A.12), we have:

$$q_1 \geqslant -\sum_{k_1=1}^{K_1}\sum_{j,m\in\mathcal{G}^*_{k_1},j<m} 2C_{D_1}|\mathcal{G}_{\min}|^{-1}\sqrt{\log n}\|\boldsymbol{\beta}_j-\boldsymbol{\beta}_m\|_2. \tag{A.13}$$

Similarly, for some constant $0 < C_{D_2} < \infty$,

$$\begin{aligned}
q_2 &\geqslant -\sum_{k_2=1}^{K_2}\sum_{j,m\in\mathcal{T}^*_{k_2},j<m} 2C_{D_2}|\mathcal{T}_{\min}|^{-1}\sqrt{\log n}\|\boldsymbol{\gamma}_j-\boldsymbol{\gamma}_m\|_2 \\
&\geqslant -\sum_{k_2=1}^{K_2}\sum_{j,m\in\mathcal{T}^*_{k_2},j<m} 2C_{D_2}|\mathcal{T}_{\min}|^{-1}\sqrt{\log n}\sqrt{\|\boldsymbol{\beta}_j-\boldsymbol{\beta}_m\|_2^2 + \|\boldsymbol{\gamma}_j-\boldsymbol{\gamma}_m\|_2^2}.
\end{aligned} \tag{A.14}$$

Next, we consider q_3 and q_4:

$$
\begin{aligned}
q_3 &= q_{31} + q_{32},\\
q_{31} &= \lambda_2 \sum_{j=1}^{n} \sum_{j\neq m} \rho'(\|\widetilde{\boldsymbol{\beta}}_j - \widetilde{\boldsymbol{\beta}}_m\|_2)\|\widetilde{\boldsymbol{\beta}}_j - \widetilde{\boldsymbol{\beta}}_m\|_2^{-1}(\widetilde{\boldsymbol{\beta}}_j - \widetilde{\boldsymbol{\beta}}_m)^\top(\boldsymbol{\beta}_j - \breve{\boldsymbol{\beta}}_j),\\
q_{32} &= \lambda_1 \sum_{j=1}^{n} \sum_{j\neq m} \rho'(\sqrt{\|\widetilde{\boldsymbol{\beta}}_j - \widetilde{\boldsymbol{\beta}}_m\|_2^2 + \|\widetilde{\boldsymbol{\gamma}}_j - \widetilde{\boldsymbol{\gamma}}_m\|_2^2})(\|\widetilde{\boldsymbol{\beta}}_j - \widetilde{\boldsymbol{\beta}}_m\|_2^2 + \|\widetilde{\boldsymbol{\gamma}}_j - \widetilde{\boldsymbol{\gamma}}_m\|_2^2)^{-1/2}\\
&\qquad \cdot (\widetilde{\boldsymbol{\beta}}_j - \widetilde{\boldsymbol{\beta}}_m)^\top(\boldsymbol{\beta}_j - \breve{\boldsymbol{\beta}}_j),\\
q_4 &= \lambda_1 \sum_{j=1}^{n} \sum_{j\neq m} \rho'(\sqrt{\|\widetilde{\boldsymbol{\beta}}_j - \widetilde{\boldsymbol{\beta}}_m\|_2^2 + \|\widetilde{\boldsymbol{\gamma}}_j - \widetilde{\boldsymbol{\gamma}}_m\|_2^2})(\|\widetilde{\boldsymbol{\beta}}_j - \widetilde{\boldsymbol{\beta}}_m\|_2^2 + \|\widetilde{\boldsymbol{\gamma}}_j - \widetilde{\boldsymbol{\gamma}}_m\|_2^2)^{-1/2}\\
&\qquad \cdot (\widetilde{\boldsymbol{\gamma}}_j - \widetilde{\boldsymbol{\gamma}}_m)^\top(\boldsymbol{\gamma}_j - \breve{\boldsymbol{\gamma}}_j).
\end{aligned}
$$

Denote $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$. Then,

$$
\begin{aligned}
\widetilde{q}_4 &= q_4 + q_{32},\\
&= \lambda_1 \sum_{j=1}^{n} \sum_{j\neq m} \rho'(\|\widetilde{\boldsymbol{\theta}}_j - \widetilde{\boldsymbol{\theta}}_m\|_2)\|\widetilde{\boldsymbol{\theta}}_j - \widetilde{\boldsymbol{\theta}}_m\|_2^{-1}(\widetilde{\boldsymbol{\theta}}_j - \widetilde{\boldsymbol{\theta}}_m)^\top(\boldsymbol{\theta}_j - \breve{\boldsymbol{\theta}}_j)\\
&= \lambda_1 \sum_{j<m} \rho'(\|\widetilde{\boldsymbol{\theta}}_j - \widetilde{\boldsymbol{\theta}}_m\|_2)\|\widetilde{\boldsymbol{\theta}}_j - \widetilde{\boldsymbol{\theta}}_m\|_2^{-1}(\widetilde{\boldsymbol{\theta}}_j - \widetilde{\boldsymbol{\theta}}_m)^\top(\boldsymbol{\theta}_j - \breve{\boldsymbol{\theta}}_j)\\
&\quad + \lambda_1 \sum_{j>m} \rho'(\|\widetilde{\boldsymbol{\theta}}_m - \widetilde{\boldsymbol{\theta}}_j\|_2)\|\widetilde{\boldsymbol{\theta}}_m - \widetilde{\boldsymbol{\theta}}_j\|_2^{-1}(\widetilde{\boldsymbol{\theta}}_m - \widetilde{\boldsymbol{\theta}}_j)^\top(\boldsymbol{\theta}_m - \breve{\boldsymbol{\theta}}_m)\\
&= \lambda_1 \sum_{j<m} \rho'(\|\widetilde{\boldsymbol{\theta}}_j - \widetilde{\boldsymbol{\theta}}_m\|_2)\|\widetilde{\boldsymbol{\theta}}_j - \widetilde{\boldsymbol{\theta}}_m\|_2^{-1}(\widetilde{\boldsymbol{\theta}}_j - \widetilde{\boldsymbol{\theta}}_m)^\top[(\boldsymbol{\theta}_j - \breve{\boldsymbol{\theta}}_j) - (\boldsymbol{\theta}_m - \breve{\boldsymbol{\theta}}_m)]\\
&= \lambda_1 \sum_{k_2=1}^{K_2} \sum_{j,m\in\mathcal{T}_{k_2}^*, j<m} \rho'(\|\widetilde{\boldsymbol{\theta}}_j - \widetilde{\boldsymbol{\theta}}_m\|_2)\|\widetilde{\boldsymbol{\theta}}_j - \widetilde{\boldsymbol{\theta}}_m\|_2^{-1}(\widetilde{\boldsymbol{\theta}}_j - \widetilde{\boldsymbol{\theta}}_m)^\top(\boldsymbol{\theta}_j - \boldsymbol{\theta}_m)\\
&\quad + \lambda_1 \sum_{k_2<k_2'} \sum_{j\in\mathcal{T}_{k_2}^*, m\in\mathcal{T}_{k_2'}^*} \rho'(\|\widetilde{\boldsymbol{\theta}}_j - \widetilde{\boldsymbol{\theta}}_m\|_2)\|\widetilde{\boldsymbol{\theta}}_j - \widetilde{\boldsymbol{\theta}}_m\|_2^{-1}(\widetilde{\boldsymbol{\theta}}_j - \widetilde{\boldsymbol{\theta}}_m)^\top[(\boldsymbol{\theta}_j - \breve{\boldsymbol{\theta}}_j) - (\boldsymbol{\theta}_m - \breve{\boldsymbol{\theta}}_m)].
\end{aligned}
$$

The last equality is from $\breve{\boldsymbol{\theta}}_j = \breve{\boldsymbol{\theta}}_m$ for $j, m \in \mathcal{T}_{k_2}^*$, $k_2 = 1, \cdots, K_2$.

For $j \in \mathcal{T}_{k_2}^*, m \in \mathcal{T}_{k_2'}^*, k_2 \neq k_2'$, $\|\widetilde{\boldsymbol{\theta}}_j - \widetilde{\boldsymbol{\theta}}_m\|_2 \geqslant \|\widetilde{\boldsymbol{\gamma}}_j - \widetilde{\boldsymbol{\gamma}}_m\|_2 \geqslant \min_{j\in\mathcal{T}_{k_2}^*, m\in\mathcal{T}_{k_2'}^*} \|\boldsymbol{\gamma}_j^* - \boldsymbol{\gamma}_m^*\|_2 - 2\sup_j \|\widetilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_j^*\|_2 \geqslant b_1 - 2\tau_n' > a\lambda_1$, and so $\rho'(\|\widetilde{\boldsymbol{\theta}}_j - \widetilde{\boldsymbol{\theta}}_m\|_2) = 0$. Noting that for $j, m \in \mathcal{T}_{k_2}^*$,

$\widetilde{\boldsymbol{\theta}}_j - \widetilde{\boldsymbol{\theta}}_m = \varsigma(\boldsymbol{\theta}_j - \boldsymbol{\theta}_m)$, we have:

$$\widetilde{q}_4 = \lambda_1 \sum_{k_2=1}^{K_2} \sum_{j,m \in \mathcal{T}_{k_2}^*, j<m} \rho'(\|\widetilde{\boldsymbol{\theta}}_j - \widetilde{\boldsymbol{\theta}}_m\|_2)\|\boldsymbol{\theta}_j - \boldsymbol{\theta}_m\|_2.$$

Similar to (A.9), $\sup_j \|\breve{\boldsymbol{\theta}}_j - \widehat{\boldsymbol{\theta}}_j^o\|_2 \leqslant \sup_j \|\boldsymbol{\theta}_j - \widehat{\boldsymbol{\theta}}_j^o\|_2$. And note that $\breve{\boldsymbol{\theta}}_j = \breve{\boldsymbol{\theta}}_m$ for $j, m \in \mathcal{T}_{k_2}^*$. Then it can be shown that:

$$\begin{aligned}\sup_j \|\widetilde{\boldsymbol{\theta}}_j - \widetilde{\boldsymbol{\theta}}_m\|_2 &= \sup_j \|\widetilde{\boldsymbol{\theta}}_j - \breve{\boldsymbol{\theta}}_j + \breve{\boldsymbol{\theta}}_m - \widetilde{\boldsymbol{\theta}}_m\|_2 \leqslant 2\sup_j \|\widetilde{\boldsymbol{\theta}}_j - \breve{\boldsymbol{\theta}}_j\|_2 \leqslant 2\sup_j \|\boldsymbol{\theta}_j - \breve{\boldsymbol{\theta}}_j\|_2 \\ &\leqslant 2(\sup_j \|\boldsymbol{\theta}_j - \widehat{\boldsymbol{\theta}}_j^o\|_2 + \|\breve{\boldsymbol{\theta}}_j - \widehat{\boldsymbol{\theta}}_j^o\|_2) \leqslant 4\sup_j \|\boldsymbol{\theta}_j - \widehat{\boldsymbol{\theta}}_j^o\|_2 \leqslant 8\phi_n.\end{aligned}$$

Therefore, $\rho'(\|\widetilde{\boldsymbol{\theta}}_k - \widetilde{\boldsymbol{\theta}}_m\|_2) \geqslant \rho'(8\phi_n)$ by the concavity of $\rho(\cdot)$. So we have:

$$\widetilde{q}_4 \geqslant \sum_{k_2=1}^{K_2} \sum_{j,m \in \mathcal{T}_{k_2}^*, j<m} \lambda_1 \rho'(8\phi_n)\|\boldsymbol{\theta}_j - \boldsymbol{\theta}_m\|_2. \tag{A.15}$$

Similarly,

$$q_{31} \geqslant \sum_{k_1=1}^{K_1} \sum_{j,m \in \mathcal{G}_{k_1}^*, j<m} \lambda_2 \rho'(4\phi_n)\|\boldsymbol{\beta}_j - \boldsymbol{\beta}_m\|_2. \tag{A.16}$$

By (A.13), (A.14), (A.15), and (A.16), we have:

$$\begin{aligned}&Q(\boldsymbol{\beta}, \boldsymbol{\gamma}) - Q(\breve{\boldsymbol{\beta}}, \breve{\boldsymbol{\gamma}}) \\ &\geqslant \sum_{k_1=1}^{K_1} \sum_{j,m \in \mathcal{G}_{k_1}^*, j<m} \left[\lambda_2 \rho'(4\phi_n) - 2C_{D_1}|\mathcal{G}_{\min}|^{-1}\sqrt{\log n}\right] \|\boldsymbol{\beta}_j - \boldsymbol{\beta}_m\|_2 \\ &+ \sum_{k_2=1}^{K_2} \sum_{j,m \in \mathcal{T}_{k_2}^*, j<m} \left[\lambda_1 \rho'(8\phi_n) - 2C_{D_2}|\mathcal{T}_{\min}|^{-1}\sqrt{\log n}\right] \|\boldsymbol{\theta}_j - \boldsymbol{\theta}_m\|_2.\end{aligned}$$

Let $\phi_n = o(1)$, $\rho'(4\phi_n) \to 1$, and $\rho'(8\phi_n) \to 1$. Note that $|\mathcal{G}_{\min}| > |\mathcal{T}_{\min}| \gg \sqrt{(K_1 q + K_2 p)n \log n}$, $\lambda_1 \gg \tau_n'$, and $\lambda_2 \gg \tau_n'$. Then we have $\min(\lambda_1, \lambda_2) \gg |\mathcal{T}_{\min}|^{-1}\sqrt{\log n}$. Therefore, $Q(\boldsymbol{\beta}, \boldsymbol{\gamma}) - Q(\breve{\boldsymbol{\beta}}, \breve{\boldsymbol{\gamma}}) > 0$ for a sufficiently large n. So the result in (ii) is proved. This concludes the proof of Result 2.

B. Details of the computational algorithm

Recall that as iteration $t+1$, the updates are:

$$(\boldsymbol{\beta}^{(t+1)}, \boldsymbol{\gamma}^{(t+1)}) = \underset{\boldsymbol{\beta},\boldsymbol{\gamma}}{\operatorname{argmin}} \mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\upsilon}^{(t)}, \boldsymbol{\omega}^{(t)}, \boldsymbol{\xi}^{(t)}, \boldsymbol{\eta}^{(t)}), \tag{B.1}$$

$$(\boldsymbol{\upsilon}^{(t+1)}, \boldsymbol{\omega}^{(t+1)}) = \underset{\boldsymbol{\upsilon},\boldsymbol{\omega}}{\operatorname{argmin}} \mathcal{L}(\boldsymbol{\beta}^{(t+1)}, \boldsymbol{\gamma}^{(t+1)}, \boldsymbol{\upsilon}, \boldsymbol{\omega}, \boldsymbol{\xi}^{(t)}, \boldsymbol{\eta}^{(t)}), \tag{B.2}$$

$$\boldsymbol{\xi}_{jm}^{(t+1)} = \boldsymbol{\xi}_{jm}^{(t)} + \kappa(\boldsymbol{\beta}_j^{(t+1)} - \boldsymbol{\beta}_m^{(t+1)} - \boldsymbol{\upsilon}_{jm}^{(t+1)}), \tag{B.3}$$

$$\boldsymbol{\eta}_{jm}^{(t+1)} = \boldsymbol{\eta}_{jm}^{(t)} + \kappa(\boldsymbol{\gamma}_j^{(t+1)} - \boldsymbol{\gamma}_m^{(t+1)} - \boldsymbol{\omega}_{jm}^{(t+1)}). \tag{B.4}$$

For (B.1), it is equivalent to minimizing:

$$\begin{aligned}
&\frac{1}{2}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}_i - \boldsymbol{z}_i^\top \boldsymbol{\gamma}_i)^2 + \frac{\kappa}{2}\sum_{j<m}\|\boldsymbol{\beta}_j - \boldsymbol{\beta}_m - \boldsymbol{\upsilon}_{jm}^{(t)} + \frac{\boldsymbol{\xi}_{jm}^{(t)}}{\kappa}\|_2^2 + \frac{\kappa}{2}\sum_{j<m}\|\boldsymbol{\gamma}_j - \boldsymbol{\gamma}_m - \boldsymbol{\omega}_{jm}^{(t)} + \frac{\boldsymbol{\eta}_{jm}^{(t)}}{\kappa}\|_2^2 \\
&= \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{\gamma}\|_2^2 + \frac{\kappa}{2}\|\boldsymbol{H}(\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top - (\boldsymbol{\upsilon}^{(t)\top}, \boldsymbol{\omega}^{(t)\top})^\top + (\boldsymbol{\xi}^{(t)\top}, \boldsymbol{\eta}^{(t)\top})^\top/\kappa\|_2^2,
\end{aligned}$$

where $\boldsymbol{y} = (y_1, \cdots, y_n)^\top$, $\boldsymbol{X} = \operatorname{diag}(\boldsymbol{x}_1^\top, \cdots, \boldsymbol{x}_n^\top)$, $\boldsymbol{Z} = \operatorname{diag}(\boldsymbol{z}_1^\top, \cdots, \boldsymbol{z}_n^\top)$, $\boldsymbol{H} = \mathcal{E} \otimes \boldsymbol{I}_{p+q}$, $\mathcal{E} = \{(\boldsymbol{e}_j - \boldsymbol{e}_m, j < m)\}^\top$ with $\boldsymbol{e}_j$ being the vector whose jth element is 1 and the remaining ones are 0, $\boldsymbol{I}_{p+q}$ is a $(p+q) \times (p+q)$ identity matrix, and $\otimes$ is the Kronecker product. Then the update for $(\boldsymbol{\beta}^{(t+1)\top}, \boldsymbol{\gamma}^{(t+1)\top})^\top$ is:

$$\begin{aligned}
&(\boldsymbol{\beta}^{(t+1)\top}, \boldsymbol{\gamma}^{(t+1)\top})^\top \\
=&[(\boldsymbol{X}, \boldsymbol{Z})^\top(\boldsymbol{X}, \boldsymbol{Z}) + \kappa\boldsymbol{H}^\top\boldsymbol{H}]^{-1}[(\boldsymbol{X}, \boldsymbol{Z})^\top\boldsymbol{y} + \kappa\boldsymbol{H}^\top\{(\boldsymbol{\upsilon}^{(t)\top}, \boldsymbol{\omega}^{(t)\top})^\top - (\boldsymbol{\xi}^{(t)\top}, \boldsymbol{\eta}^{(t)\top})^\top/\kappa\}].
\end{aligned}$$

For (B.2), after removing the terms independent of $(\boldsymbol{\upsilon}^\top, \boldsymbol{\omega}^\top)^\top$, it is equivalent to minimizing:

$$\frac{\kappa}{2}\|(\boldsymbol{\upsilon}_{jm}^\top, \boldsymbol{\omega}_{jm}^\top)^\top - (\overline{\boldsymbol{\upsilon}}_{jm}^\top, \overline{\boldsymbol{\omega}}_{jm}^\top)^\top\|_2^2 + p(\sqrt{\|\boldsymbol{\upsilon}_{jm}\|_2^2 + \|\boldsymbol{\omega}_{jm}\|_2^2}, \lambda_1) + p(\|\boldsymbol{\upsilon}_{jm}\|_2, \lambda_2)$$

with respect to $(\boldsymbol{\upsilon}_{jm}^\top, \boldsymbol{\omega}_{jm}^\top)^\top$, where $\overline{\boldsymbol{\upsilon}}_{jm}^\top = \boldsymbol{\gamma}_j^{(t+1)} - \boldsymbol{\gamma}_m^{(t+1)} + \boldsymbol{\eta}_{jm}^{(t)}/\kappa$, and $\overline{\boldsymbol{\omega}}_{jm}^\top = \boldsymbol{\beta}_j^{(t+1)} - \boldsymbol{\beta}_m^{(t+1)} + \boldsymbol{\xi}_{jm}^{(t)}/\kappa$. The solution is a hierarchical groupwise thresholding operator corresponding to $p(\cdot, \lambda)$. Denote $\boldsymbol{u}_{jm} = (\boldsymbol{v}_{jm}^\top, \boldsymbol{\omega}_{jm}^\top)^\top$ and $(s)_+ = sI(s > 0)$. Then, it can be obtained that:

$$\boldsymbol{\omega}_{jm}^{(t+1)} = \overline{\boldsymbol{\omega}}_{jm}^\top \text{ and } \boldsymbol{v}_{jm}^{(t+1)} = \overline{\boldsymbol{v}}_{jm}^\top, \quad \text{if } \|\overline{\boldsymbol{u}}_{jm}^\top\|_2 > a\lambda_1 \text{ and } \|\overline{\boldsymbol{v}}_{jm}^\top\|_2 > a\lambda_2; \tag{B.5}$$

$$\begin{aligned}\boldsymbol{\omega}_{jm}^{(t+1)} &= \frac{[1-\lambda_1/(\kappa\|\overline{\boldsymbol{u}}_{jm}^\top\|_2)]_+}{1-1/(a\kappa)}\overline{\boldsymbol{\omega}}_{jm}^\top \text{ and } \boldsymbol{v}_{jm}^{(t+1)} = \frac{[1-\lambda_1/(\kappa\|\overline{\boldsymbol{u}}_{jm}^\top\|_2)]_+}{1-1/(a\kappa)}\overline{\boldsymbol{v}}_{jm}^\top,\\ &\quad \text{if } \|\overline{\boldsymbol{u}}_{jm}^\top\|_2 \leqslant a\lambda_1 \text{ and } \frac{[1-\lambda_1/(\kappa\|\overline{\boldsymbol{u}}_{jm}^\top\|_2)]_+}{1-1/(a\kappa)} \cdot \|\overline{\boldsymbol{v}}_{jm}^\top\|_2 > a\lambda_2;\end{aligned} \tag{B.6}$$

$$\begin{aligned}\boldsymbol{\omega}_{jm}^{(t+1)} &= \overline{\boldsymbol{\omega}}_{jm}^\top \text{ and } \boldsymbol{v}_{jm}^{(t+1)} = \frac{[1-\lambda_2/(\kappa\|\overline{\boldsymbol{v}}_{jm}^\top\|_2)]_+}{1-1/(a\kappa)}\overline{\boldsymbol{v}}_{jm}^\top,\\ &\quad \text{if } \|\overline{\boldsymbol{\omega}}_{jm}^\top\|_2^2 + \left[\frac{[1-\lambda_2/(\kappa\|\overline{\boldsymbol{v}}_{jm}^\top\|_2)]_+}{1-1/(a\kappa)}\right]^2 \cdot \|\overline{\boldsymbol{v}}_{jm}^\top\|_2^2 > (a\lambda_1)^2 \text{ and } \|\overline{\boldsymbol{v}}_{jm}^\top\|_2 \leqslant a\lambda_2;\end{aligned} \tag{B.7}$$

$$\boldsymbol{\omega}_{jm}^{(t+1)} = \frac{\overline{\boldsymbol{\omega}}_{jm}^\top}{1 + \frac{p'(\boldsymbol{u}_{jm}^{(t)}, \lambda_1)}{\kappa\|\boldsymbol{u}_{jm}^{(t)}\|_2}} \text{ and } \boldsymbol{v}_{jm}^{(t+1)} = \frac{\overline{\boldsymbol{v}}_{jm}^\top}{1 + \frac{p'(\boldsymbol{u}_{jm}^{(t)}, \lambda_1)}{\kappa\|\boldsymbol{u}_{jm}^{(t)}\|_2} + \frac{p'(\boldsymbol{v}_{jm}^{(t)}, \lambda_2)}{\kappa\|\boldsymbol{v}_{jm}^{(t)}\|_2}}, \quad \text{otherwise}, \tag{B.8}$$

where (B.8) is derived using the local quadratic approximation technique, which leads to an explicit solution at each iteration.

C. Additional numerical results

[Figure 1 about here.]

[Figure 2 about here.]

[Figure 3 about here.]

[Figure 4 about here.]

[Figure 5 about here.]

[Figure 6 about here.]

[Figure 7 about here.]

[Figure 8 about here.]

[Figure 9 about here.]

[Figure 10 about here.]

[Figure 11 about here.]

[Figure 12 about here.]

[Figure 13 about here.]

[Figure 14 about here.]

[Figure 15 about here.]

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

[Table 4 about here.]

[Table 5 about here.]

[Table 6 about here.]

[Table 7 about here.]

[Table 8 about here.]

[Table 9 about here.]

[Table 10 about here.]

[Table 11 about here.]

[Table 12 about here.]

[Table 13 about here.]

[Table 14 about here.]

[Table 15 about here.]

[Table 16 about here.]

[Table 17 about here.]

[Table 18 about here.]

[Table 19 about here.]

[Table 20 about here.]

[Table 21 about here.]

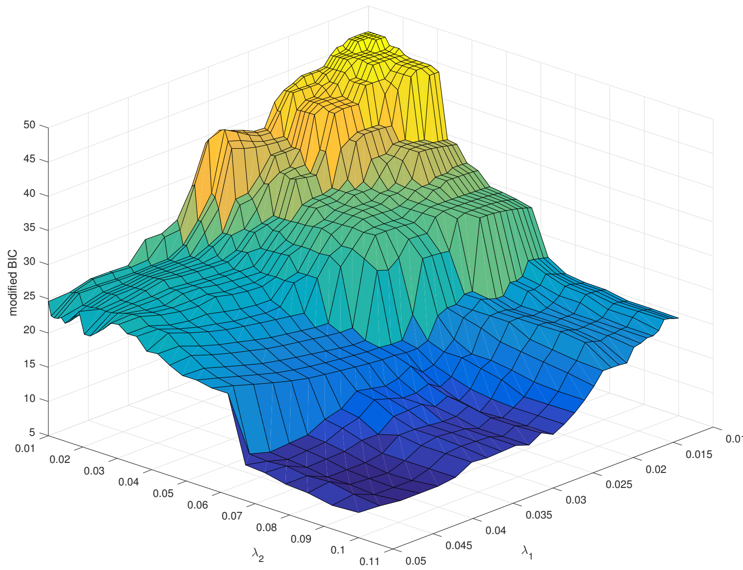

Figure C1 . Analysis of one simulated dataset: modified BIC as a function of λ_1 and λ_2. The optimal point has $(\lambda_1, \lambda_2) = (0.040, 0.085)$, which corresponds to $(\widehat{K}_1, \widehat{K}_2) = (2, 4)$.

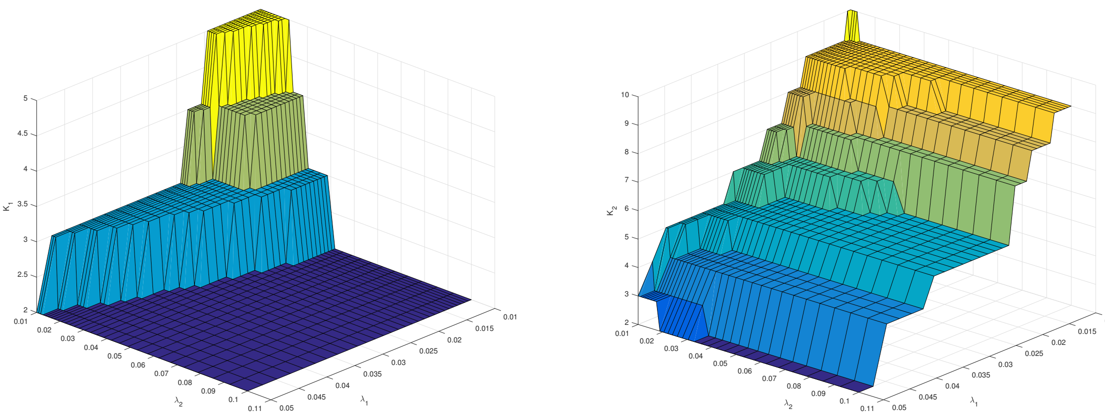

Figure C2 . Analysis of one simulated dataset: $\widehat{K}_1$ and $\widehat{K}_2$ as a function of λ_1 and λ_2. To improve presentation, results with extremely small λ_1 and λ_2 values are not shown.

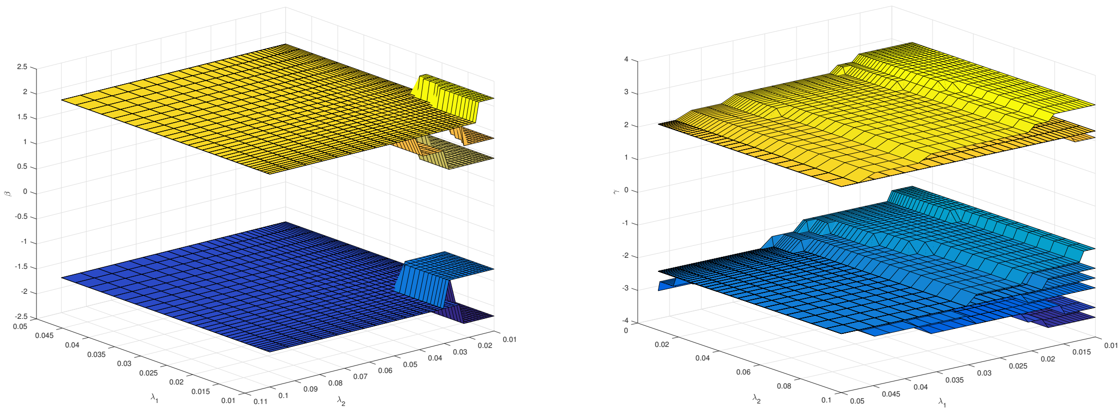

Figure C3 . Analysis of one simulated dataset: solutions paths for one component of β_i and one component of γ_i, as a function of λ_1 and λ_2. To improve presentation, results with extremely small λ_1 and λ_2 values are not shown.

Figure C4 . Simulation results for the scenario with correlated $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$, imbalanced design, and $\mu = 2$. Correlation structures from top to bottom: AR1, AR2, B1, and B2.

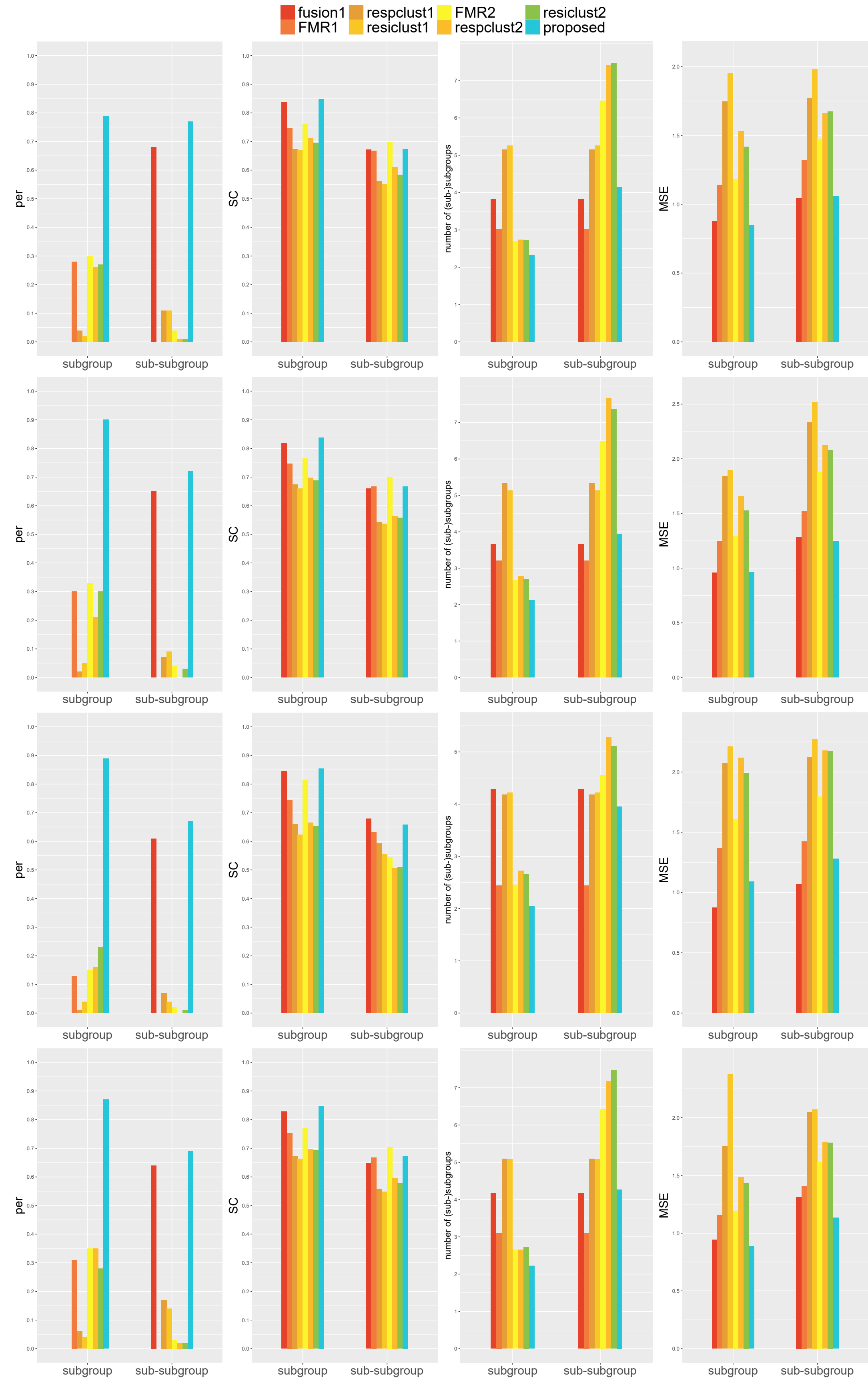

Figure C5 . Simulation results for the scenario with uncorrelated $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$, balanced design, and $\mu = 2$. Correlation structures from top to bottom: AR1, AR2, B1, and B2.

Figure C6 . Simulation results for the scenario with uncorrelated $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$, imbalanced design, and $\mu = 2$. Correlation structures from top to bottom: AR1, AR2, B1, and B2.

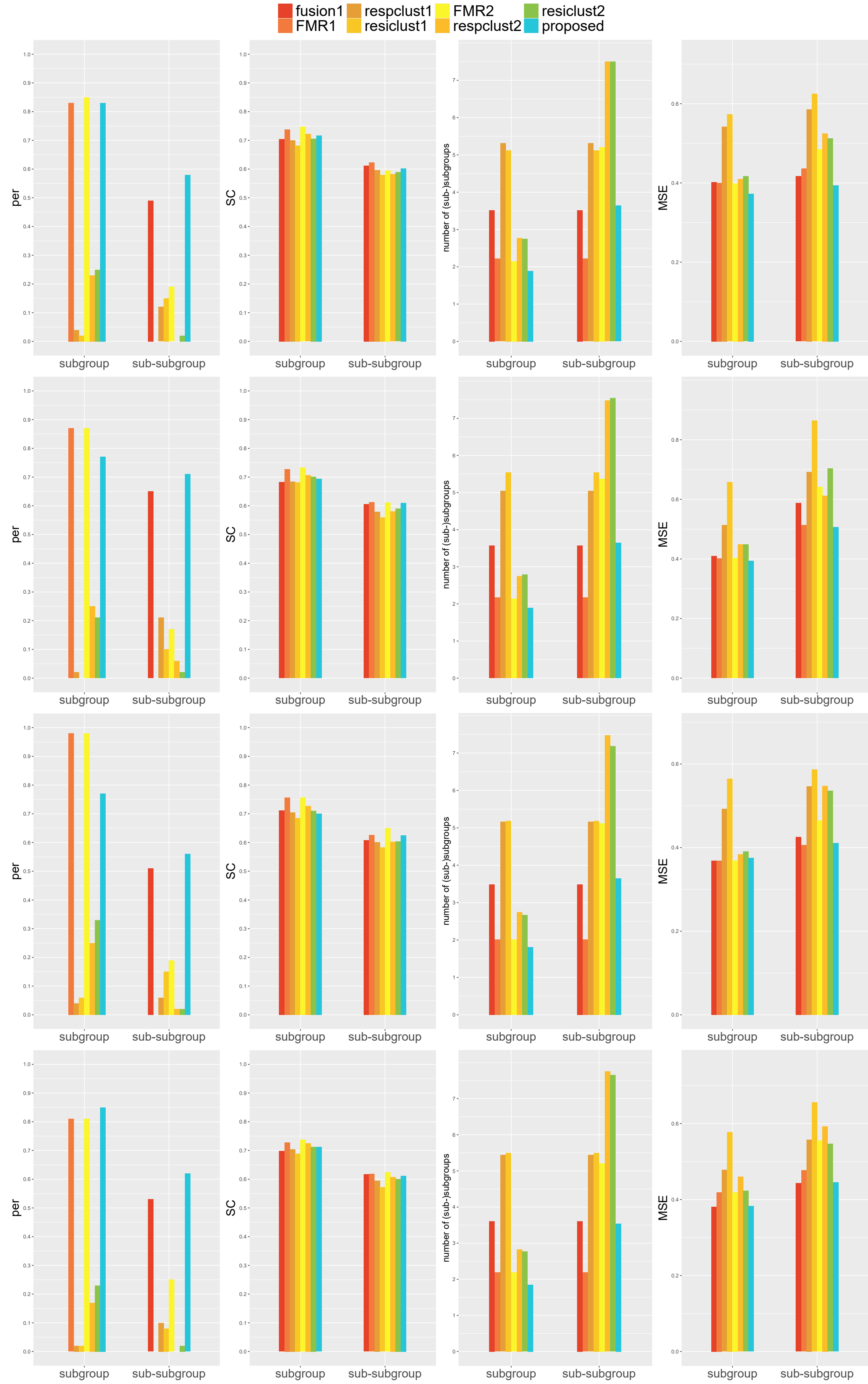

Figure C7 . Simulation results for the scenario with correlated $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$, balanced design, and $\mu = 0.6$. Correlation structures from top to bottom: AR1, AR2, B1, and B2.

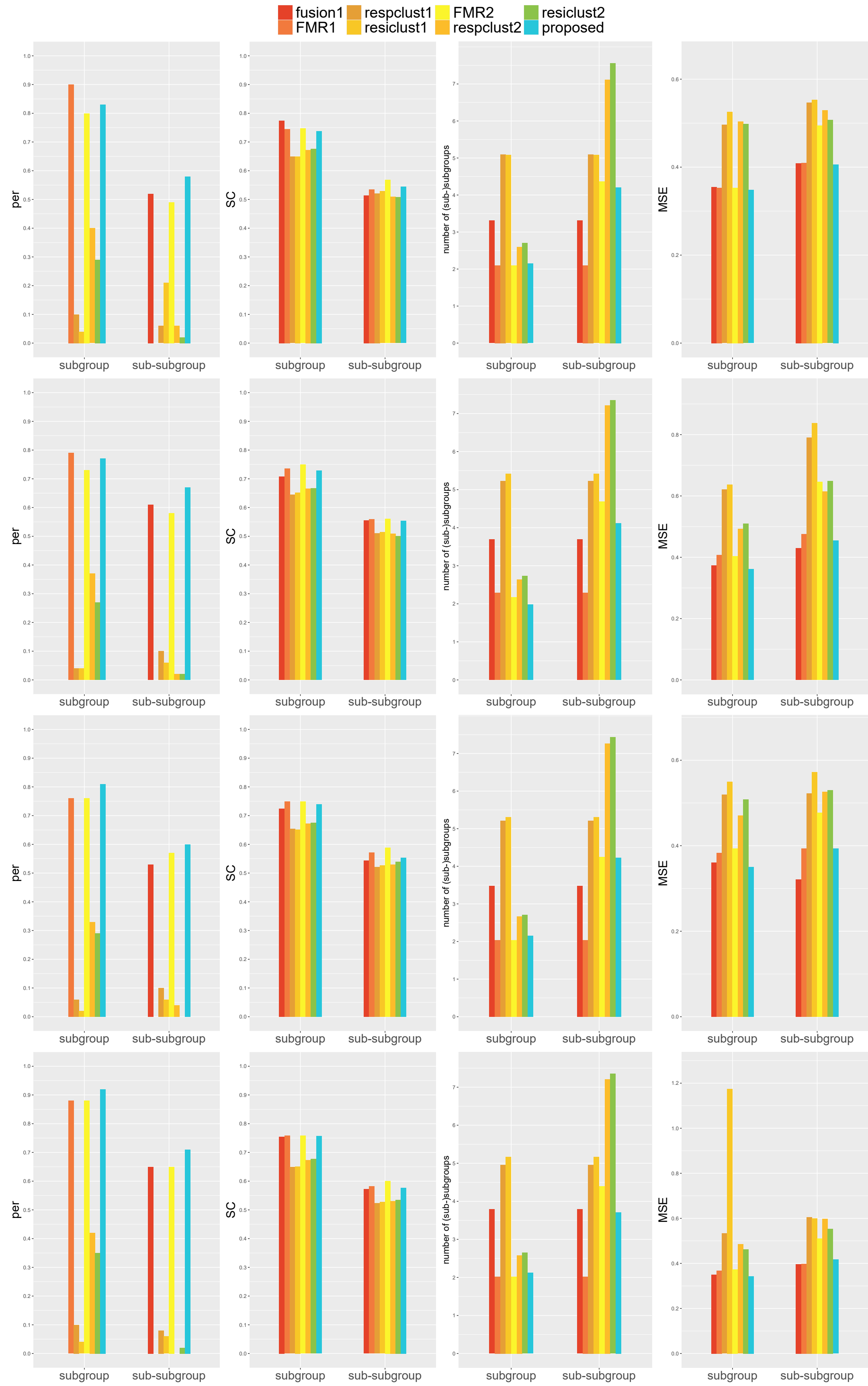

Figure C8 . Simulation results for the scenario with correlated $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$, imbalanced design, and $\mu = 0.6$. Correlation structures from top to bottom: AR1, AR2, B1, and B2.

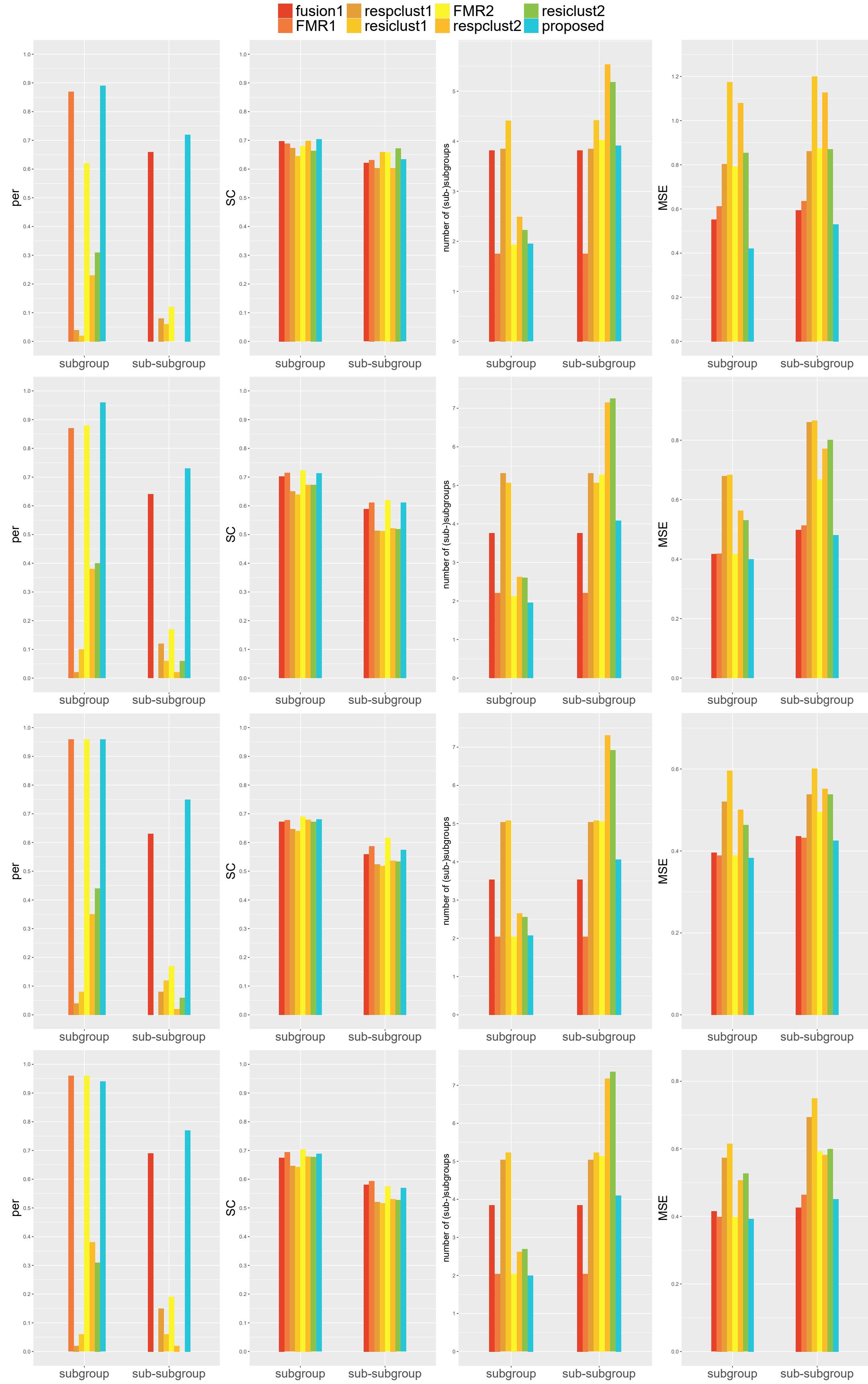

Figure C9 . Simulation results for the scenario with uncorrelated $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$, balanced design, and $\mu = 0.6$. Correlation structures from top to bottom: AR1, AR2, B1, and B2.

Figure C10 . Simulation results for the scenario with uncorrelated $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$, imbalanced design, and $\mu = 0.6$. Correlation structures from top to bottom: AR1, AR2, B1, and B2.

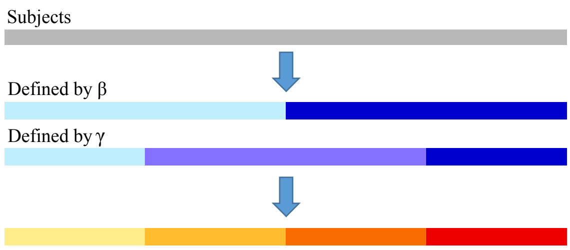

Figure C11 . Analysis scheme: when the heterogeneity hierarchy is violated. Middle panels: heterogeneity defined by $\boldsymbol{\beta}$ and by $\boldsymbol{\gamma}$. Lower panel: heterogeneity structure obtained with the proposed approach.

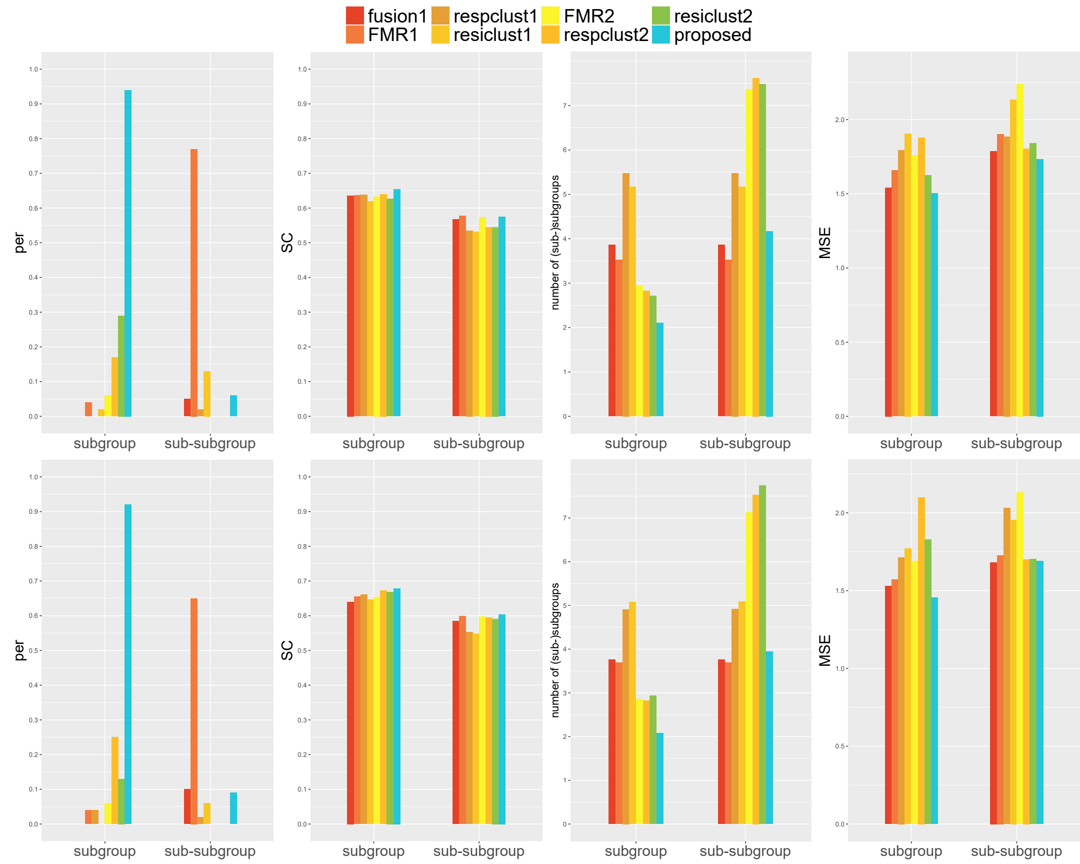

Figure C12 . Simulation with the heterogeneity hierarchy violated. Top/bottom: uncorrelated/correlated $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$.

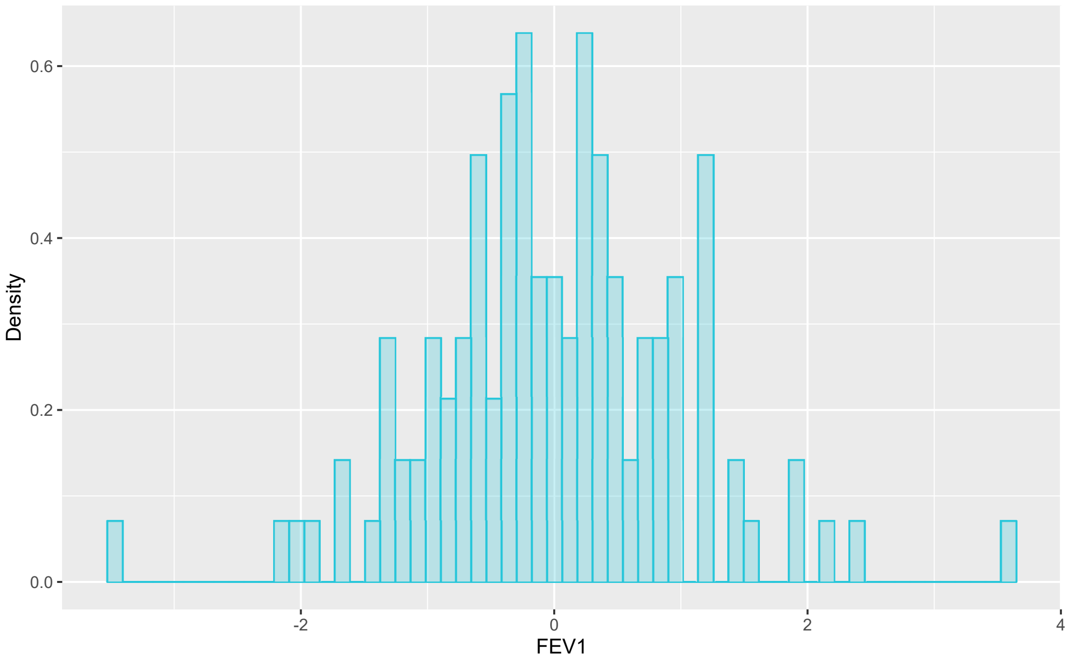

Figure C13 . Data analysis: histogram of the response variable.

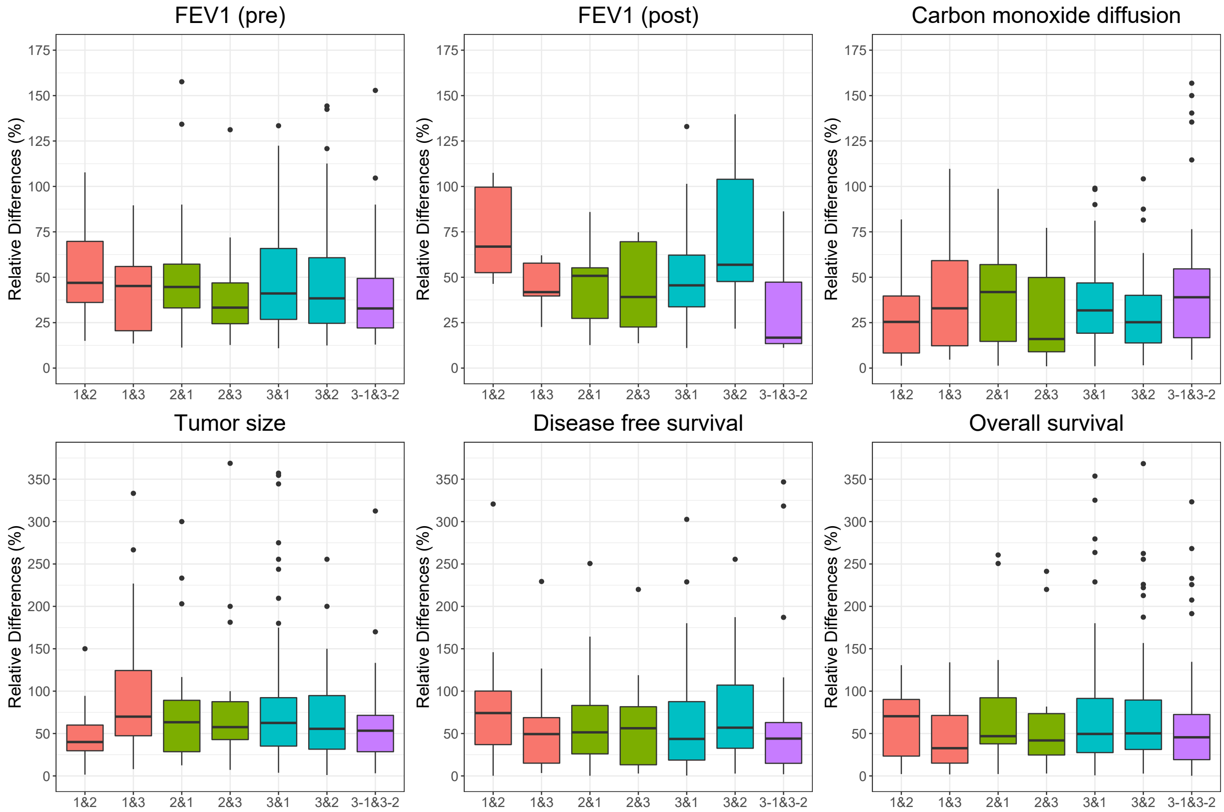

Figure C14 . Distributions of relative differences in clinical outcomes of the best matched subjects between different (sub-)subgroups. FEV1 (pre/post): percentage comparison to a normal value reference range of the volume of air that a patient can forcibly exhale from the lungs in one second pre/post-bronchodilator. Carbon monoxide diffusion: value measuring the amount of carbon monoxide detected in a patient's lungs. Disease free survival/Overall survival: disease free/overall survival time since initial treatment. 1&2: for subject i in Subgroup 1, identify subject i' in Subgroup 2 that is the closest if measured by $(\boldsymbol{X}, \boldsymbol{Z})$. For a clinical outcome y', the relative difference for subject i in Subgroup 1 and subject i' in Subgroup 2 is calculated as $|y'_i - y'_{i'}|/|y'_i|$. n_1 values are generated for this comparison, where n_1 is the size of Subgroup 1.

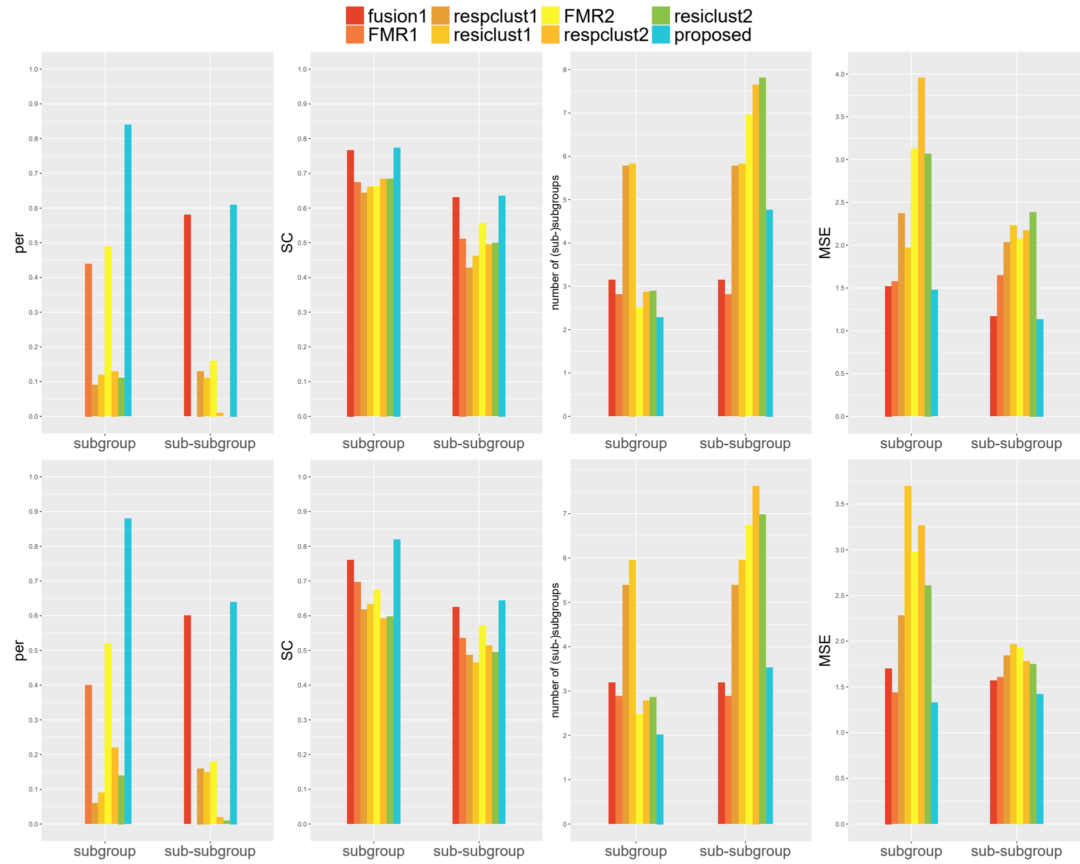

Figure C15 . Simulation results based on the LUAD data. Top/bottom: balanced/imbalanced design.

Table C1

Simulation with correlated $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$, balanced design, and $\mu = 2$. In each cell, mean(sd).

		$\boldsymbol{\beta}$				$\boldsymbol{\gamma}$			
Correlation ($\boldsymbol{z}_i$)	method	per	$\widehat{K_1}$	SC	MSE	per	$\widehat{K_2}$	SC	MSE
AR1	fusion1	0.00	4.14(1.25)	0.841(0.024)	0.861(0.114)	0.78	4.14(1.25)	0.677(0.012)	1.043(0.081)
	FMR1	0.51	2.49(0.50)	0.826(0.105)	1.197(0.473)	0.00	2.49(0.50)	0.700(0.020)	1.285(0.368)
	respclust1	0.17	4.64(1.45)	0.785(0.060)	1.334(0.690)	0.14	4.64(1.45)	0.682(0.034)	1.546(1.025)
	resiclust1	0.07	4.99(1.24)	0.732(0.034)	1.714(0.738)	0.10	4.99(1.24)	0.632(0.027)	1.752(0.698)
	FMR2	0.51	2.49(0.50)	0.834(0.096)	1.196(0.471)	0.08	6.10(1.07)	0.742(0.039)	1.511(0.375)
	respclust2	0.62	2.38(0.49)	0.930(0.718)	1.038(0.651)	0.06	6.47(1.53)	0.774(0.227)	1.689(0.980)
	resiclust2	0.32	2.68(0.47)	0.767(0.043)	1.198(0.406)	0.01	7.43(1.43)	0.691(0.042)	1.659(0.365)
	proposed	0.88	2.18(0.58)	0.923(0.051)	0.763(0.150)	0.80	4.40(1.48)	0.712(0.014)	0.973(0.090)
AR2	fusion1	0.00	3.76(1.34)	0.829(0.039)	0.970(0.181)	0.63	3.76(1.34)	0.669(0.019)	1.246(0.169)
	FMR1	0.46	2.54(0.50)	0.779(0.105)	1.430(0.472)	0.00	2.54(0.50)	0.677(0.026)	1.781(0.606)
	respclust1	0.00	5.16(1.02)	0.739(0.025)	1.617(0.971)	0.18	5.16(1.02)	0.650(0.036)	2.008(1.029)
	resiclust1	0.06	4.99(1.21)	0.712(0.034)	1.860(0.810)	0.16	4.99(1.21)	0.612(0.028)	2.586(1.728)
	FMR2	0.47	2.53(0.50)	0.801(0.085)	1.423(0.476)	0.01	6.31(0.93)	0.716(0.053)	2.127(0.559)
	respclust2	0.49	2.51(0.50)	0.798(0.057)	1.157(0.354)	0.06	6.78(1.59)	0.709(0.051)	1.853(0.671)
	resiclust2	0.49	2.51(0.50)	0.746(0.037)	1.310(0.680)	0.01	6.92(1.54)	0.674(0.051)	2.084(1.204)
	proposed	0.81	2.22(0.63)	0.854(0.112)	0.942(0.182)	0.69	4.38(1.53)	0.692(0.015)	1.263(0.157)
B1	fusion1	0.00	3.79(1.12)	0.845(0.028)	0.849(0.112)	0.60	3.79(1.12)	0.678(0.014)	1.025(0.070)
	FMR1	0.44	2.60(0.60)	0.805(0.106)	1.252(0.411)	0.00	2.60(0.60)	0.693(0.025)	1.390(0.335)
	respclust1	0.12	4.74(1.40)	0.772(0.052)	1.263(0.475)	0.17	4.74(1.40)	0.676(0.032)	1.411(0.354)
	resiclust1	0.14	4.62(1.47)	0.731(0.035)	1.592(0.795)	0.13	4.62(1.47)	0.636(0.028)	1.716(0.897)
	FMR2	0.45	2.55(0.50)	0.819(0.094)	1.240(0.405)	0.04	6.13(1.04)	0.729(0.044)	1.566(0.307)
	respclust2	0.51	2.49(0.50)	0.835(0.066)	1.020(0.610)	0.03	6.80(1.55)	0.742(0.037)	1.606(0.582)
	resiclust2	0.42	2.58(0.50)	0.768(0.045)	1.216(0.439)	0.01	7.10(1.44)	0.695(0.042)	1.728(0.741)
	proposed	0.87	2.89(1.210)	0.900(0.088)	0.815(0.186)	0.62	4.29(1.23)	0.707(0.012)	1.024(0.105)
B2	fusion1	0.00	3.97(1.24)	0.843(0.032)	0.881(0.137)	0.61	3.97(1.24)	0.677(0.015)	1.112(0.098)
	FMR1	0.45	2.55(0.50)	0.791(0.108)	1.319(0.504)	0.00	2.55(0.50)	0.684(0.031)	1.624(0.556)
	respclust1	0.13	4.98(1.36)	0.767(0.047)	1.383(0.676)	0.08	4.98(1.36)	0.662(0.028)	1.706(0.771)
	resiclust1	0.07	4.98(1.29)	0.719(0.035)	1.685(0.824)	0.08	4.98(1.29)	0.623(0.031)	2.039(0.774)
	FMR2	0.44	2.56(0.50)	0.805(0.091)	1.322(0.503)	0.02	6.39(1.05)	0.719(0.052)	1.902(0.493)
	respclust2	0.50	2.50(0.50)	0.820(0.058)	1.093(0.773)	0.03	6.88(1.49)	0.734(0.038)	1.725(0.535)
	resiclust2	0.43	2.57(0.49)	0.757(0.044)	1.238(0.659)	0.02	7.12(1.53)	0.681(0.045)	1.838(0.592)
	proposed	0.87	2.54(1.19)	0.873(0.112)	0.860(0.197)	0.67	4.46(1.10)	0.700(0.013)	1.127(0.133)

Table C2

Simulation with correlated $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$, imbalanced design, and $\mu = 2$. In each cell, mean(sd).

		$\boldsymbol{\beta}$				$\boldsymbol{\gamma}$			
Correlation ($\boldsymbol{z}_i$)	method	per	$\widehat{K_1}$	SC	MSE	per	$\widehat{K_2}$	SC	MSE
AR1	fusion1	0.00	3.83(1.22)	0.822(0.025)	0.888(0.118)	0.71	3.83(1.22)	0.629(0.018)	1.118(0.089)
	FMR1	0.46	2.56(0.56)	0.760(0.102)	1.217(0.379)	0.00	2.56(0.56)	0.636(0.029)	1.368(0.363)
	respclust1	0.06	5.15(1.18)	0.655(0.029)	1.599(0.486)	0.14	5.15(1.18)	0.526(0.023)	1.751(0.639)
	resiclust1	0.03	5.12(1.09)	0.659(0.029)	2.372(5.679)	0.15	5.12(1.09)	0.536(0.024)	1.950(1.169)
	FMR2	0.47	2.53(0.50)	0.776(0.087)	1.219(0.378)	0.26	5.62(1.26)	0.695(0.062)	1.515(0.359)
	respclust2	0.36	2.64(0.48)	0.676(0.022)	1.501(0.576)	0.03	7.24(1.49)	0.578(0.035)	1.675(0.356)
	resiclust2	0.27	2.73(0.45)	0.682(0.022)	1.529(0.671)	0.00	7.46(1.41)	0.584(0.038)	1.748(0.541)
	proposed	0.94	2.11(0.45)	0.862(0.032)	0.873(0.130)	0.81	4.27(1.52)	0.634(0.025)	1.056(0.090)
AR2	fusion1	0.00	4.03(2.01)	0.824(0.028)	0.980(0.137)	0.56	4.03(2.01)	0.606(0.020)	1.293(0.173)
	FMR1	0.43	2.61(0.60)	0.757(0.092)	1.327(0.466)	0.00	2.61(0.60)	0.631(0.031)	1.703(0.577)
	respclust1	0.02	5.24(0.97)	0.655(0.025)	1.823(0.683)	0.13	5.24(0.97)	0.509(0.022)	2.321(1.415)
	resiclust1	0.01	5.38(0.89)	0.661(0.023)	2.014(1.006)	0.09	5.38(0.89)	0.522(0.021)	2.402(0.765)
	FMR2	0.44	2.56(0.50)	0.773(0.075)	1.327(0.465)	0.24	5.93(1.51)	0.685(0.067)	2.392(1.547)
	respclust2	0.31	2.69(0.47)	0.673(0.018)	1.696(0.736)	0.04	7.24(1.46)	0.553(0.032)	2.113(1.085)
	resiclust2	0.30	2.70(0.46)	0.679(0.019)	1.761(0.983)	0.01	7.45(1.40)	0.572(0.038)	1.921(0.498)
	proposed	0.84	2.22(0.60)	0.849(0.041)	0.953(0.131)	0.64	4.20(1.21)	0.626(0.022)	1.276(0.142)
B1	fusion1	0.00	3.76(1.35)	0.846(0.027)	0.884(0.129)	0.54	3.76(1.35)	0.624(0.021)	1.007(0.103)
	FMR1	0.49	2.51(0.50)	0.768(0.102)	1.207(0.444)	0.00	2.51(0.50)	0.640(0.022)	1.365(0.366)
	respclust1	0.08	5.05(1.31)	0.654(0.031)	1.614(0.530)	0.08	5.05(1.31)	0.525(0.023)	1.770(0.883)
	resiclust1	0.03	4.91(1.22)	0.657(0.028)	1.885(0.988)	0.15	4.91(1.22)	0.539(0.028)	1.982(0.901)
	FMR2	0.50	2.50(0.50)	0.786(0.083)	1.203(0.444)	0.19	5.61(1.25)	0.697(0.057)	1.533(0.330)
	respclust2	0.30	2.70(0.46)	0.678(0.016)	1.555(0.809)	0.00	7.45(1.37)	0.572(0.037)	1.784(0.522)
	resiclust2	0.34	2.66(0.48)	0.682(0.024)	1.512(0.537)	0.01	7.23(1.48)	0.591(0.043)	1.776(0.627)
	proposed	0.89	2.18(0.54)	0.863(0.030)	0.867(0.120)	0.64	4.18(1.06)	0.635(0.022)	1.060(0.097)
B2	fusion1	0.00	3.55(1.68)	0.822(0.027)	0.887(0.113)	0.64	3.55(1.68)	0.610(0.019)	1.214(0.139)
	FMR1	0.32	2.70(0.52)	0.737(0.092)	1.333(0.423)	0.00	2.70(0.52)	0.636(0.029)	1.618(0.466)
	respclust1	0.04	5.31(1.08)	0.654(0.032)	1.767(0.752)	0.08	5.31(1.08)	0.515(0.021)	2.072(1.107)
	resiclust1	0.08	4.93(1.27)	0.652(0.036)	1.778(0.560)	0.14	4.93(1.27)	0.531(0.026)	2.121(0.654)
	FMR2	0.33	2.67(0.47)	0.758(0.075)	1.335(0.423)	0.15	6.15(1.36)	0.678(0.059)	1.910(0.528)
	respclust2	0.40	2.60(0.49)	0.674(0.021)	1.577(0.396)	0.02	7.13(1.51)	0.567(0.040)	1.938(0.623)
	resiclust2	0.37	2.63(0.49)	0.679(0.023)	1.569(0.558)	0.02	7.28(1.56)	0.580(0.041)	1.826(0.523)
	proposed	0.89	2.17(0.51)	0.856(0.031)	0.890(0.120)	0.71	4.15(0.86)	0.630(0.023)	1.184(0.118)

Table C3

Simulation with uncorrelated $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$, balanced design, and $\mu = 2$. In each cell, mean(sd).

		$\boldsymbol{\beta}$				$\boldsymbol{\gamma}$			
Correlation ($\boldsymbol{z}_i$)	method	per	$\widehat{K_1}$	SC	MSE	per	$\widehat{K_2}$	SC	MSE
AR1	fusion1	0.00	3.83(1.03)	0.839(0.034)	0.877(0.138)	0.68	3.83(1.03)	0.672(0.017)	1.047(0.080)
	FMR1	0.28	3.02(1.09)	0.746(0.082)	1.143(0.336)	0.00	3.02(1.09)	0.668(0.030)	1.318(0.293)
	respclust1	0.04	5.15(1.10)	0.674(0.027)	1.746(1.386)	0.11	5.15(1.10)	0.562(0.019)	1.771(1.006)
	resiclust1	0.02	5.26(1.04)	0.669(0.026)	1.953(1.298)	0.11	5.26(1.04)	0.551(0.017)	1.980(0.722)
	FMR2	0.30	2.70(0.46)	0.762(0.061)	1.181(0.316)	0.04	6.46(1.11)	0.698(0.047)	1.476(0.263)
	respclust2	0.26	2.74(0.44)	0.713(0.132)	1.532(0.756)	0.01	7.41(1.41)	0.611(0.145)	1.662(0.409)
	resiclust2	0.27	2.73(0.45)	0.696(0.019)	1.419(0.205)	0.01	7.47(1.38)	0.584(0.035)	1.673(0.335)
	proposed	0.79	2.32(0.67)	0.848(0.028)	0.851(0.116)	0.77	4.15(0.83)	0.673(0.014)	1.060(0.096)
AR2	fusion1	0.00	3.66(1.65)	0.818(0.023)	0.958(0.124)	0.65	3.66(1.65)	0.660(0.013)	1.284(0.113)
	FMR1	0.30	3.21(1.34)	0.747(0.073)	1.245(0.377)	0.00	3.21(1.34)	0.667(0.030)	1.523(0.454)
	respclust1	0.02	5.34(0.99)	0.674(0.027)	1.843(0.687)	0.07	5.34(0.99)	0.542(0.015)	2.337(1.274)
	resiclust1	0.05	5.13(1.18)	0.660(0.032)	1.897(0.392)	0.09	5.13(1.18)	0.537(0.015)	2.518(0.786)
	FMR2	0.33	2.67(0.47)	0.765(0.054)	1.294(0.355)	0.04	6.50(1.12)	0.700(0.046)	1.883(0.453)
	respclust2	0.21	2.79(0.41)	0.698(0.017)	1.657(0.475)	0.00	7.66(1.24)	0.563(0.028)	2.126(0.980)
	resiclust2	0.30	2.70(0.46)	0.688(0.020)	1.526(0.554)	0.03	7.36(1.53)	0.558(0.028)	2.079(0.925)
	proposed	0.90	2.13(0.42)	0.837(0.033)	0.963(0.128)	0.72	3.93(0.71)	0.667(0.017)	1.245(0.157)
B1	fusion1	0.00	4.28(0.98)	0.846(0.036)	0.876(0.103)	0.61	4.28(0.98)	0.679(0.017)	1.071(0.109)
	FMR1	0.13	2.44(1.29)	0.744(0.807)	1.368(0.751)	0.00	2.44(1.28)	0.634(0.225)	1.424(0.718)
	respclust1	0.01	4.18(2.23)	0.661(0.889)	2.075(1.746)	0.07	4.18(2.23)	0.593(0.121)	2.121(1.023)
	resiclust1	0.04	4.22(2.20)	0.623(0.890)	2.213(1.698)	0.04	4.22(2.20)	0.556(0.123)	2.275(1.073)
	FMR2	0.15	2.46(1.32)	0.815(0.484)	1.612(1.348)	0.02	4.56(2.61)	0.544(0.235)	1.797(1.065)
	respclust2	0.16	2.73(2.06)	0.665(0.152)	2.120(1.181)	0.00	5.28(3.24)	0.506(0.177)	2.180(0.124)
	resiclust2	0.23	2.66(2.06)	0.655(0.153)	1.994(1.146)	0.01	5.11(3.19)	0.510(0.177)	2.171(0.131)
	proposed	0.89	2.05(1.10)	0.854(0.848)	1.091(0.889)	0.67	3.95(1.97)	0.659(0.886)	1.281(0.858)
B2	fusion1	0.00	4.17(1.11)	0.828(0.032)	0.942(0.128)	0.64	4.17(1.11)	0.648(0.017)	1.313(0.121)
	FMR1	0.31	3.11(1.25)	0.753(0.084)	1.155(0.304)	0.00	3.11(1.25)	0.668(0.023)	1.406(0.374)
	respclust1	0.06	5.09(1.21)	0.672(0.028)	1.752(0.751)	0.17	5.09(1.21)	0.558(0.021)	2.049(1.241)
	resiclust1	0.04	5.08(1.20)	0.663(0.035)	2.379(6.268)	0.14	5.08(1.20)	0.548(0.021)	2.072(0.589)
	FMR2	0.35	2.65(0.48)	0.771(0.067)	1.194(0.312)	0.03	6.42(1.12)	0.704(0.050)	1.617(0.341)
	respclust2	0.35	2.65(0.48)	0.697(0.020)	1.484(0.396)	0.02	7.18(1.49)	0.595(0.042)	1.789(0.443)
	resiclust2	0.28	2.72(0.45)	0.694(0.020)	1.438(0.232)	0.02	7.48(1.43)	0.578(0.039)	1.785(0.468)
	proposed	0.87	2.22(0.60)	0.847(0.031)	0.887(0.128)	0.69	4.27(1.10)	0.672(0.016)	1.133(0.117)

Table C4

Simulation with uncorrelated $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$, imbalanced design, and $\mu = 2$. In each cell, mean(sd).

		$\boldsymbol{\beta}$				$\boldsymbol{\gamma}$			
Correlation ($\boldsymbol{z}_i$)	method	per	$\widehat{K_1}$	SC	MSE	per	$\widehat{K_2}$	SC	MSE
AR1	fusion1	0.00	3.79(1.04)	0.839(0.036)	0.908(0.133)	0.71	3.79(1.04)	0.632(0.024)	1.109(0.149)
	FMR1	0.56	2.46(0.56)	0.780(0.101)	1.118(0.353)	0.00	2.46(0.56)	0.638(0.021)	1.278(0.322)
	respclust1	0.01	5.25(1.02)	0.661(0.027)	1.537(0.377)	0.15	5.25(1.02)	0.531(0.021)	1.633(0.409)
	resiclust1	0.02	5.29(0.99)	0.667(0.023)	1.784(1.139)	0.11	5.29(0.99)	0.539(0.025)	1.822(0.765)
	FMR2	0.56	2.44(0.50)	0.792(0.084)	1.126(0.354)	0.41	5.23(1.21)	0.700(0.053)	1.387(0.313)
	respclust2	0.25	2.75(0.44)	0.684(0.017)	1.376(0.260)	0.01	7.52(1.34)	0.584(0.039)	1.674(0.432)
	resiclust2	0.15	2.85(0.36)	0.690(0.018)	1.463(0.454)	0.03	7.74(1.24)	0.590(0.039)	1.590(0.226)
	proposed	0.91	2.14(0.47)	0.863(0.029)	0.847(0.119)	0.78	4.11(0.72)	0.637(0.020)	1.031(0.085)
AR2	fusion1	0.00	4.38(1.31)	0.840(0.025)	0.967(0.110)	0.57	4.38(1.31)	0.625(0.018)	1.294(0.184)
	FMR1	0.54	2.48(0.56)	0.784(0.093)	1.197(0.427)	0.00	2.48(0.56)	0.640(0.026)	1.565(0.538)
	respclust1	0.03	5.36(0.99)	0.659(0.025)	1.720(0.671)	0.10	5.36(0.99)	0.513(0.017)	2.055(0.930)
	resiclust1	0.01	5.26(0.98)	0.659(0.025)	1.881(0.702)	0.14	5.26(0.98)	0.527(0.023)	2.434(1.055)
	FMR2	0.54	2.46(0.50)	0.797(0.078)	1.202(0.441)	0.41	5.33(1.31)	0.707(0.057)	1.979(0.698)
	respclust2	0.25	2.75(0.44)	0.678(0.020)	1.626(0.368)	0.00	7.55(1.35)	0.560(0.031)	2.088(0.859)
	resiclust2	0.28	2.72(0.45)	0.681(0.019)	1.616(0.492)	0.01	7.53(1.43)	0.582(0.041)	1.912(0.663)
	proposed	0.81	2.33(0.74)	0.862(0.025)	0.934(0.156)	0.66	4.19(0.98)	0.633(0.020)	1.240(0.153)
B1	fusion1	0.00	3.93(1.09)	0.853(0.041)	0.842(0.146)	0.66	3.93(1.09)	0.637(0.029)	1.037(0.106)
	FMR1	0.53	2.47(0.50)	0.784(0.094)	1.089(0.330)	0.00	2.47(0.50)	0.642(0.020)	1.261(0.291)
	respclust1	0.03	5.25(1.06)	0.658(0.026)	1.588(0.674)	0.10	5.25(1.06)	0.530(0.022)	1.730(0.755)
	resiclust1	0.02	5.20(1.09)	0.662(0.023)	1.651(0.587)	0.09	5.20(1.09)	0.541(0.025)	1.803(0.784)
	FMR2	0.53	2.47(0.50)	0.793(0.084)	1.089(0.330)	0.44	5.27(1.32)	0.697(0.047)	1.416(0.312)
	respclust2	0.23	2.77(0.42)	0.685(0.015)	1.382(0.379)	0.02	7.80(1.41)	0.580(0.032)	1.678(0.397)
	resiclust2	0.21	2.79(0.41)	0.689(0.019)	1.594(1.355)	0.00	7.84(1.31)	0.592(0.040)	1.645(0.401)
	proposed	0.89	2.14(0.43)	0.867(0.031)	0.840(0.121)	0.72	4.18(1.12)	0.639(0.021)	1.028(0.083)
B2	fusion1	0.00	3.52(1.11)	0.841(0.026)	0.894(0.127)	0.57	3.52(1.11)	0.626(0.019)	1.171(0.108)
	FMR1	0.56	2.44(0.50)	0.779(0.098)	1.146(0.395)	0.00	2.44(0.50)	0.636(0.026)	1.424(0.378)
	respclust1	0.00	5.42(0.89)	0.660(0.021)	1.652(0.505)	0.09	5.42(0.89)	0.521(0.022)	1.955(0.687)
	resiclust1	0.03	5.18(1.14)	0.660(0.026)	1.875(0.983)	0.09	5.18(1.14)	0.533(0.024)	2.187(1.310)
	FMR2	0.56	2.44(0.50)	0.792(0.082)	1.146(0.395)	0.44	5.27(1.30)	0.704(0.062)	1.648(0.467)
	respclust2	0.27	2.73(0.45)	0.678(0.017)	1.503(0.439)	0.00	7.53(1.30)	0.571(0.035)	1.821(0.649)
	resiclust2	0.26	2.74(0.44)	0.682(0.017)	1.672(1.197)	0.00	7.58(1.34)	0.587(0.035)	1.772(0.458)
	proposed	0.86	2.25(0.64)	0.859(0.032)	0.866(0.132)	0.66	4.34(1.46)	0.630(0.025)	1.160(0.127)

Table C5

Simulation with correlated $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$, balanced design, and $\mu = 0.6$. In each cell, mean(sd).

		$\boldsymbol{\beta}$				$\boldsymbol{\gamma}$			
Correlation ($\boldsymbol{z}_i$)	method	per	$\widehat{K_1}$	SC	MSE	per	$\widehat{K_2}$	SC	MSE
AR1	fusion1	0.00	3.52(1.53)	0.704(0.093)	0.402(0.132)	0.49	3.52(1.53)	0.612(0.118)	0.417(0.081)
	FMR1	0.83	2.22(0.54)	0.738(0.056)	0.400(0.094)	0.00	2.22(0.54)	0.623(0.023)	0.436(0.101)
	respclust1	0.04	5.31(1.08)	0.700(0.022)	0.542(0.357)	0.12	5.31(1.08)	0.597(0.021)	0.586(0.483)
	resiclust1	0.02	5.12(1.11)	0.682(0.026)	0.573(0.244)	0.15	5.12(1.11)	0.580(0.020)	0.625(0.206)
	FMR2	0.85	2.15(0.36)	0.747(0.041)	0.399(0.094)	0.19	5.21(0.85)	0.595(0.044)	0.485(0.100)
	respclust2	0.23	2.77(0.43)	0.722(0.017)	0.410(0.068)	0.00	7.50(1.37)	0.582(0.037)	0.525(0.113)
	resiclust2	0.25	2.75(0.44)	0.706(0.015)	0.417(0.075)	0.02	7.50(1.28)	0.589(0.034)	0.513(0.144)
	proposed	0.83	1.89(0.47)	0.717(0.099)	0.373(0.046)	0.58	3.64(1.42)	0.602(0.123)	0.394(0.090)
AR2	fusion1	0.00	3.57(1.13)	0.682(0.093)	0.409(0.079)	0.65	3.57(1.13)	0.606(0.061)	0.588(0.092)
	FMR1	0.87	2.17(0.51)	0.727(0.045)	0.402(0.077)	0.00	2.17(0.51)	0.613(0.024)	0.513(0.150)
	respclust1	0.02	5.04(1.03)	0.683(0.028)	0.514(0.193)	0.21	5.04(1.03)	0.579(0.022)	0.691(0.388)
	resiclust1	0.00	5.54(0.75)	0.681(0.017)	0.658(0.342)	0.10	5.54(0.75)	0.559(0.024)	0.864(0.436)
	FMR2	0.87	2.14(0.35)	0.733(0.035)	0.403(0.079)	0.17	5.37(0.91)	0.611(0.041)	0.642(0.154)
	respclust2	0.25	2.75(0.44)	0.706(0.024)	0.449(0.074)	0.06	7.48(1.50)	0.580(0.035)	0.612(0.199)
	resiclust2	0.21	2.79(0.41)	0.700(0.017)	0.449(0.082)	0.02	7.54(1.36)	0.590(0.033)	0.704(0.511)
	proposed	0.77	1.89(0.58)	0.694(0.102)	0.393(0.057)	0.71	3.65(0.91)	0.609(0.055)	0.507(0.071)
B1	fusion1	0.00	3.49(1.22)	0.712(0.056)	0.369(0.086)	0.51	3.49(1.22)	0.608(0.041)	0.425(0.107)
	FMR1	0.98	2.02(0.14)	0.756(0.035)	0.369(0.045)	0.00	2.02(0.14)	0.626(0.019)	0.406(0.037)
	respclust1	0.04	5.17(1.15)	0.705(0.025)	0.493(0.195)	0.06	5.17(1.15)	0.601(0.030)	0.547(0.170)
	resiclust1	0.06	5.19(1.14)	0.685(0.028)	0.565(0.195)	0.15	5.19(1.14)	0.583(0.024)	0.587(0.167)
	FMR2	0.98	2.02(0.14)	0.756(0.034)	0.369(0.045)	0.19	5.12(0.73)	0.650(0.023)	0.465(0.070)
	respclust2	0.25	2.75(0.44)	0.727(0.024)	0.384(0.073)	0.02	7.48(1.35)	0.603(0.046)	0.548(0.181)
	resiclust2	0.33	2.67(0.47)	0.710(0.020)	0.391(0.062)	0.02	7.19(1.39)	0.604(0.039)	0.536(0.167)
	proposed	0.77	1.81(0.45)	0.701(0.117)	0.375(0.057)	0.56	3.65(1.34)	0.625(0.018)	0.411(0.047)
B2	fusion1	0.00	3.61(1.42)	0.698(0.063)	0.381(0.076)	0.53	3.61(1.42)	0.617(0.091)	0.443(0.089)
	FMR1	0.81	2.19(0.40)	0.728(0.055)	0.419(0.095)	0.00	2.19(0.40)	0.618(0.023)	0.477(0.114)
	respclust1	0.02	5.44(0.96)	0.704(0.021)	0.478(0.118)	0.10	5.44(0.96)	0.595(0.026)	0.558(0.141)
	resiclust1	0.02	5.50(0.87)	0.689(0.025)	0.578(0.171)	0.08	5.50(0.87)	0.572(0.023)	0.656(0.148)
	FMR2	0.81	2.19(0.40)	0.737(0.041)	0.419(0.095)	0.25	5.21(0.85)	0.624(0.045)	0.555(0.115)
	respclust2	0.17	2.83(0.38)	0.725(0.020)	0.460(0.265)	0.00	7.75(1.15)	0.608(0.046)	0.593(0.199)
	resiclust2	0.23	2.77(0.43)	0.713(0.023)	0.423(0.074)	0.02	7.65(1.36)	0.601(0.045)	0.547(0.133)
	proposed	0.85	1.85(0.36)	0.712(0.098)	0.383(0.057)	0.62	3.54(1.28)	0.612(0.088)	0.445(0.091)

Table C6

Simulation with correlated $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$, imbalanced design, and $\mu = 0.6$. In each cell, mean(sd).

		$\boldsymbol{\beta}$				$\boldsymbol{\gamma}$			
Correlation ($\boldsymbol{z}_i$)	method	per	$\widehat{K_1}$	SC	MSE	per	$\widehat{K_2}$	SC	MSE
AR1	fusion1	0.00	3.31(1.32)	0.774(0.073)	0.355(0.063)	0.52	3.31(1.32)	0.514(0.231)	0.409(0.156)
	FMR1	0.90	2.10(0.30)	0.745(0.046)	0.353(0.061)	0.00	2.10(0.30)	0.535(0.031)	0.410(0.074)
	respclust1	0.10	5.10(1.36)	0.649(0.032)	0.496(0.258)	0.06	5.10(1.36)	0.521(0.022)	0.547(0.273)
	resiclust1	0.04	5.08(1.10)	0.650(0.023)	0.526(0.172)	0.21	5.08(1.10)	0.529(0.022)	0.553(0.106)
	FMR2	0.80	2.10(0.30)	0.747(0.041)	0.353(0.061)	0.49	4.37(0.79)	0.568(0.034)	0.495(0.097)
	respclust2	0.40	2.60(0.50)	0.672(0.020)	0.504(0.448)	0.06	7.12(1.53)	0.509(0.032)	0.529(0.163)
	resiclust2	0.29	2.71(0.46)	0.676(0.014)	0.498(0.307)	0.02	7.56(1.59)	0.508(0.036)	0.507(0.106)
	proposed	0.83	2.15(0.70)	0.738(0.068)	0.348(0.042)	0.58	4.21(1.29)	0.544(0.091)	0.406(0.049)
AR2	fusion1	0.00	3.69(1.23)	0.708(0.054)	0.373(0.052)	0.61	3.69(1.23)	0.555(0.156)	0.430(0.129)
	FMR1	0.79	2.29(0.67)	0.736(0.065)	0.407(0.201)	0.00	2.29(0.67)	0.559(0.029)	0.476(0.109)
	respclust1	0.04	5.23(1.12)	0.645(0.031)	0.621(0.444)	0.10	5.23(1.15)	0.510(0.019)	0.790(0.544)
	resiclust1	0.04	5.42(0.98)	0.651(0.030)	0.637(0.283)	0.06	5.42(0.98)	0.515(0.019)	0.837(0.375)
	FMR2	0.73	2.17(0.38)	0.749(0.048)	0.403(0.202)	0.58	4.69(1.09)	0.561(0.043)	0.646(0.165)
	respclust2	0.37	2.64(0.49)	0.665(0.022)	0.493(0.197)	0.02	7.21(1.53)	0.509(0.030)	0.615(0.172)
	resiclust2	0.27	2.73(0.45)	0.667(0.021)	0.510(0.217)	0.02	7.35(1.45)	0.500(0.029)	0.649(0.295)
	proposed	0.77	1.98(0.64)	0.729(0.086)	0.361(0.046)	0.67	4.12(1.22)	0.553(0.089)	0.455(0.056)
B1	fusion1	0.00	3.48(1.61)	0.724(0.071)	0.361(0.053)	0.53	3.48(1.61)	0.544(0.226)	0.321(0.153)
	FMR1	0.76	2.04(0.19)	0.750(0.037)	0.383(0.051)	0.00	2.04(0.19)	0.572(0.029)	0.394(0.036)
	respclust1	0.06	5.21(1.19)	0.655(0.025)	0.520(0.193)	0.10	5.21(1.19)	0.522(0.022)	0.522(0.131)
	resiclust1	0.02	5.31(1.02)	0.651(0.030)	0.550(0.195)	0.06	5.31(1.02)	0.527(0.020)	0.572(0.267)
	FMR2	0.76	2.04(0.19)	0.750(0.036)	0.394(0.051)	0.57	4.25(0.59)	0.588(0.038)	0.477(0.071)
	respclust2	0.33	2.67(0.47)	0.673(0.018)	0.471(0.259)	0.04	7.27(1.63)	0.530(0.038)	0.526(0.123)
	resiclust2	0.29	2.71(0.46)	0.675(0.017)	0.508(0.435)	0.00	7.44(1.41)	0.539(0.040)	0.530(0.154)
	proposed	0.81	2.15(0.67)	0.739(0.061)	0.350(0.050)	0.60	4.23(1.52)	0.554(0.071)	0.394(0.041)
B2	fusion1	0.00	3.79(0.92)	0.755(0.077)	0.350(0.047)	0.65	3.79(0.92)	0.572(0.246)	0.397(0.179)
	FMR1	0.88	2.02(0.14)	0.758(0.046)	0.368(0.059)	0.00	2.02(0.14)	0.582(0.033)	0.398(0.059)
	respclust1	0.10	4.96(1.37)	0.649(0.032)	0.533(0.218)	0.08	4.96(1.37)	0.523(0.022)	0.605(0.226)
	resiclust1	0.04	5.17(1.20)	0.651(0.034)	1.175(4.794)	0.06	5.17(1.20)	0.527(0.018)	0.600(0.114)
	FMR2	0.88	2.02(0.14)	0.758(0.044)	0.373(0.059)	0.65	4.40(0.77)	0.601(0.033)	0.511(0.085)
	respclust2	0.42	2.58(0.50)	0.673(0.021)	0.485(0.379)	0.00	7.21(1.55)	0.531(0.037)	0.598(0.202)
	resiclust2	0.35	2.65(0.48)	0.677(0.023)	0.463(0.231)	0.02	7.35(1.55)	0.535(0.036)	0.554(0.156)
	proposed	0.92	2.12(0.43)	0.757(0.049)	0.343(0.058)	0.71	3.71(0.75)	0.577(0.053)	0.417(0.058)

Table C7

Simulation with uncorrelated $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$, balanced design, and $\mu = 0.6$. In each cell, mean(sd).

		$\boldsymbol{\beta}$				$\boldsymbol{\gamma}$			
Correlation ($\boldsymbol{z}_i$)	method	per	$\widehat{K_1}$	SC	MSE	per	$\widehat{K_2}$	SC	MSE
AR1	fusion1	0.00	3.82(1.53)	0.697(0.087)	0.552(0.693)	0.66	3.82(1.53)	0.621(0.462)	0.594(0.102)
	FMR1	0.87	1.75(0.75)	0.689(0.054)	0.612(0.516)	0.00	1.75(0.75)	0.632(0.469)	0.636(0.506)
	respclust1	0.04	3.85(1.20)	0.674(0.075)	0.804(0.821)	0.08	3.85(2.20)	0.603(0.802)	0.862(0.807)
	resiclust1	0.02	4.41(1.29)	0.645(0.075)	1.175(1.762)	0.06	4.42(2.29)	0.659(0.766)	1.201(0.754)
	FMR2	0.62	1.93(1.06)	0.681(0.057)	0.792(1.118)	0.12	4.03(2.20)	0.658(0.064)	0.874(0.893)
	respclust2	0.23	2.49(1.70)	0.698(0.080)	1.080(1.829)	0.00	5.54(3.15)	0.604(0.819)	1.127(0.819)
	resiclust2	0.31	2.23(1.19)	0.664(0.061)	0.854(1.108)	0.00	5.18(3.11)	0.672(0.089)	0.870(0.897)
	proposed	0.89	1.95(0.25)	0.704(0.084)	0.421(0.904)	0.72	3.91(1.62)	0.634(0.868)	0.531(0.091)
AR2	fusion1	0.00	3.76(1.34)	0.702(0.045)	0.417(0.065)	0.64	3.76(1.34)	0.589(0.134)	0.498(0.096)
	FMR1	0.87	2.21(0.64)	0.715(0.045)	0.419(0.095)	0.00	2.21(0.64)	0.611(0.021)	0.513(0.133)
	respclust1	0.02	5.31(0.96)	0.650(0.026)	0.680(0.267)	0.12	5.31(0.96)	0.513(0.009)	0.861(0.483)
	resiclust1	0.10	5.06(1.35)	0.639(0.048)	0.684(0.188)	0.06	5.06(1.35)	0.511(0.010)	0.866(0.270)
	FMR2	0.88	2.12(0.32)	0.723(0.033)	0.417(0.091)	0.17	5.27(0.77)	0.619(0.037)	0.669(0.148)
	respclust2	0.38	2.62(0.49)	0.672(0.027)	0.563(0.260)	0.02	7.15(1.54)	0.521(0.018)	0.771(0.471)
	resiclust2	0.40	2.60(0.50)	0.673(0.028)	0.530(0.239)	0.06	7.25(1.68)	0.519(0.017)	0.801(0.984)
	proposed	0.96	1.96(0.19)	0.713(0.056)	0.399(0.066)	0.73	4.08(1.19)	0.611(0.021)	0.480(0.054)
B1	fusion1	0.00	3.54(1.75)	0.672(0.067)	0.396(0.065)	0.63	3.54(1.75)	0.559(0.163)	0.436(0.086)
	FMR1	0.96	2.04(0.19)	0.678(0.035)	0.389(0.051)	0.00	2.04(0.19)	0.587(0.025)	0.432(0.045)
	respclust1	0.04	5.04(1.27)	0.647(0.034)	0.520(0.154)	0.08	5.04(1.27)	0.524(0.010)	0.538(0.105)
	resiclust1	0.08	5.08(1.31)	0.641(0.039)	0.596(0.326)	0.12	5.08(1.31)	0.519(0.011)	0.601(0.143)
	FMR2	0.96	2.04(0.19)	0.691(0.027)	0.389(0.051)	0.17	5.06(0.67)	0.616(0.031)	0.495(0.065)
	respclust2	0.35	2.65(0.48)	0.680(0.021)	0.501(0.112)	0.02	7.31(1.52)	0.537(0.021)	0.552(0.185)
	resiclust2	0.44	2.56(0.50)	0.672(0.025)	0.464(0.076)	0.06	6.92(1.58)	0.534(0.018)	0.538(0.163)
	proposed	0.96	2.08(0.39)	0.681(0.035)	0.383(0.046)	0.75	4.06(1.04)	0.575(0.085)	0.425(0.073)
B2	fusion1	0.00	3.85(0.72)	0.675(0.084)	0.415(0.054)	0.69	3.85(0.72)	0.581(0.165)	0.426(0.185)
	FMR1	0.96	2.04(0.19)	0.694(0.039)	0.399(0.067)	0.00	2.04(0.19)	0.594(0.023)	0.464(0.080)
	respclust1	0.02	5.04(1.03)	0.646(0.033)	0.574(0.195)	0.15	5.04(1.03)	0.521(0.014)	0.694(0.390)
	resiclust1	0.06	5.23(1.17)	0.643(0.041)	0.615(0.201)	0.06	5.23(1.17)	0.517(0.013)	0.750(0.314)
	FMR2	0.96	2.04(0.19)	0.704(0.031)	0.399(0.067)	0.19	5.14(0.77)	0.575(0.032)	0.592(0.147)
	respclust2	0.38	2.62(0.49)	0.679(0.022)	0.507(0.105)	0.02	7.17(1.51)	0.531(0.027)	0.582(0.173)
	resiclust2	0.31	2.69(0.47)	0.677(0.020)	0.527(0.253)	0.00	7.35(1.37)	0.527(0.021)	0.600(0.206)
	proposed	0.94	2.00(0.34)	0.688(0.054)	0.393(0.051)	0.77	4.10(1.05)	0.569(0.108)	0.451(0.079)

Table C8

Simulation with uncorrelated $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$, imbalanced design, and $\mu = 0.6$. In each cell, mean(sd).

		$\boldsymbol{\beta}$				$\boldsymbol{\gamma}$			
Correlation ($\boldsymbol{z}_i$)	method	per	$\widehat{K_1}$	SC	MSE	per	$\widehat{K_2}$	SC	MSE
AR1	fusion1	0.00	3.34(1.23)	0.734(0.090)	0.365(0.067)	0.62	3.34(1.23)	0.549(0.059)	0.487(0.167)
	FMR1	0.60	2.46(0.70)	0.705(0.077)	0.416(0.114)	0.00	2.46(0.70)	0.580(0.028)	0.452(0.111)
	respclust1	0.02	5.31(0.96)	0.631(0.027)	0.629(0.292)	0.17	5.31(0.96)	0.479(0.011)	0.624(0.254)
	resiclust1	0.02	5.39(0.99)	0.636(0.026)	0.570(0.190)	0.08	5.39(0.99)	0.496(0.013)	0.617(0.232)
	FMR2	0.62	2.39(0.49)	0.727(0.054)	0.416(0.114)	0.37	5.08(1.06)	0.631(0.050)	0.517(0.122)
	respclust2	0.25	2.75(0.44)	0.658(0.019)	0.517(0.105)	0.02	7.64(1.36)	0.502(0.015)	0.565(0.165)
	resiclust2	0.17	2.83(0.38)	0.663(0.015)	0.520(0.130)	0.00	7.77(1.10)	0.523(0.021)	0.542(0.212)
	proposed	0.92	2.15(0.54)	0.758(0.029)	0.361(0.057)	0.67	4.52(1.04)	0.581(0.025)	0.414(0.053)
AR2	fusion1	0.00	3.48(1.15)	0.774(0.067)	0.436(0.056)	0.60	3.48(1.15)	0.584(0.061)	0.638(0.162)
	FMR1	0.37	2.69(0.61)	0.701(0.074)	0.488(0.162)	0.02	2.69(0.61)	0.594(0.036)	0.674(0.251)
	respclust1	0.04	5.19(1.05)	0.629(0.034)	0.695(0.256)	0.06	5.19(1.05)	0.479(0.012)	0.772(0.282)
	resiclust1	0.04	5.25(1.10)	0.633(0.032)	0.730(0.340)	0.21	5.25(1.10)	0.498(0.018)	0.894(0.418)
	FMR2	0.38	2.62(0.49)	0.725(0.054)	0.491(0.163)	0.10	5.91(1.05)	0.635(0.062)	0.827(0.226)
	respclust2	0.25	2.75(0.44)	0.658(0.022)	0.717(0.349)	0.00	7.52(1.50)	0.500(0.014)	0.785(0.566)
	resiclust2	0.17	2.83(0.38)	0.662(0.023)	0.685(0.432)	0.02	7.85(1.26)	0.529(0.025)	0.590(0.110)
	proposed	0.94	2.15(0.67)	0.779(0.032)	0.421(0.097)	0.67	4.15(1.09)	0.590(0.028)	0.607(0.137)
B1	fusion1	0.00	3.41(2.01)	0.756(0.151)	0.467(0.082)	0.55	3.41(2.01)	0.547(0.245)	0.803(0.160)
	FMR1	0.33	2.38(1.14)	0.697(0.697)	0.680(0.687)	0.00	2.38(1.14)	0.518(0.720)	0.717(0.675)
	respclust1	0.02	4.40(2.01)	0.528(1.605)	0.939(0.479)	0.13	4.40(2.01)	0.418(1.641)	0.975(0.469)
	resiclust1	0.02	4.26(2.01)	0.549(1.006)	0.889(0.932)	0.12	4.26(2.01)	0.447(1.040)	0.879(0.919)
	FMR2	0.37	2.31(1.04)	0.637(1.088)	0.762(0.993)	0.15	4.46(2.02)	0.538(1.106)	0.844(0.969)
	respclust2	0.29	2.48(1.43)	0.548(1.773)	0.909(0.519)	0.02	5.69(2.81)	0.529(1.805)	0.931(0.511)
	resiclust2	0.15	2.52(1.33)	0.590(1.548)	0.899(0.367)	0.00	6.10(3.08)	0.510(1.362)	0.986(0.586)
	proposed	0.91	1.99(1.17)	0.765(1.261)	0.456(0.125)	0.64	3.78(2.18)	0.574(1.089)	0.851(0.323)
B2	fusion1	0.00	3.41(1.16)	0.747(0.074)	0.463(0.066)	0.64	3.41(1.16)	0.616(0.240)	0.541(0.183)
	FMR1	0.29	2.77(0.65)	0.675(0.067)	0.492(0.124)	0.00	2.77(0.65)	0.590(0.035)	0.568(0.128)
	respclust1	0.06	5.04(1.28)	0.621(0.042)	0.659(0.349)	0.10	5.04(1.28)	0.482(0.012)	0.738(0.456)
	resiclust1	0.04	4.94(1.27)	0.625(0.043)	0.673(0.253)	0.13	4.94(1.27)	0.498(0.015)	0.789(0.271)
	FMR2	0.33	2.67(0.47)	0.706(0.046)	0.485(0.121)	0.12	5.83(1.13)	0.627(0.048)	0.660(0.139)
	respclust2	0.21	2.79(0.41)	0.661(0.016)	0.662(0.475)	0.00	7.64(1.30)	0.500(0.012)	0.668(0.298)
	resiclust2	0.15	2.80(0.49)	0.664(0.027)	0.540(0.123)	0.00	7.80(1.54)	0.523(0.022)	0.643(0.267)
	proposed	0.92	2.09(0.52)	0.793(0.173)	0.439(0.102)	0.67	4.22(1.11)	0.634(0.335)	0.504(0.086)

Table C9

Simulation with the heterogeneity hierarchy violated. In each cell, mean (sd).

		$\boldsymbol{\beta}$				$\boldsymbol{\gamma}$			
$\boldsymbol{x}_i$-$\boldsymbol{z}_i$	method	per	$\widehat{K_1}$	SC	MSE	per	$\widehat{K_2}$	SC	MSE
Uncorrelated	fusion1	0.00	3.86(1.86)	0.636(0.026)	1.542(0.250)	0.05	3.86(1.86)	0.567(0.030)	1.786(0.251)
	FMR1	0.04	3.52(1.20)	0.637(0.048)	1.658(0.477)	0.77	3.52(1.20)	0.577(0.034)	1.902(0.626)
	respclust1	0.00	5.48(0.78)	0.638(0.018)	1.793(0.558)	0.02	5.48(0.78)	0.535(0.015)	1.886(0.557)
	resiclust1	0.02	5.17(1.17)	0.619(0.020)	1.906(0.866)	0.13	5.17(1.17)	0.531(0.013)	2.133(1.129)
	FMR2	0.06	2.94(0.24)	0.633(0.021)	1.760(0.458)	0.00	7.37(0.91)	0.574(0.034)	2.240(0.529)
	respclust2	0.17	2.83(0.38)	0.640(0.016)	1.878(0.716)	0.00	7.62(1.21)	0.545(0.026)	1.804(0.575)
	resiclust2	0.29	2.71(0.46)	0.626(0.014)	1.625(0.893)	0.00	7.48(1.44)	0.545(0.021)	1.839(0.286)
	proposed	0.94	2.10(0.41)	0.654(0.025)	1.504(0.238)	0.06	4.17(0.71)	0.575(0.026)	1.734(0.308)
Correlated	fusion1	0.00	3.76(1.67)	0.640(0.029)	1.530(0.160)	0.10	3.76(1.67)	0.585(0.041)	1.682(0.242)
	FMR1	0.04	3.69(1.49)	0.655(0.108)	1.570(0.505)	0.65	3.69(1.48)	0.599(0.077)	1.727(0.533)
	respclust1	0.04	4.91(1.33)	0.661(0.206)	1.714(0.593)	0.02	4.92(1.31)	0.553(0.134)	2.032(0.933)
	resiclust1	0.00	5.08(1.18)	0.646(0.196)	1.771(0.317)	0.06	5.09(1.16)	0.549(0.136)	1.955(0.451)
	FMR2	0.06	2.86(0.46)	0.654(0.158)	1.689(0.483)	0.00	7.12(1.45)	0.598(0.101)	2.131(0.529)
	respclust2	0.25	2.82(0.87)	0.673(0.328)	2.100(1.696)	0.00	7.53(1.66)	0.595(0.341)	1.701(0.357)
	resiclust2	0.13	2.94(0.81)	0.669(0.330)	1.827(1.231)	0.00	7.74(1.53)	0.591(0.342)	1.704(0.530)
	proposed	0.92	2.08(0.34)	0.679(0.189)	1.454(0.298)	0.09	3.94(0.78)	0.604(0.199)	1.690(0.338)

Table C10

Sensitivity analysis on the number of Type 2 features: similarity measures for (sub-)subgrouping memberships under different scenarios.

	Subgroup			Sub-subgroup		
	$p = 5$	$p = 10$	$p = 15$	$p = 5$	$p = 10$	$p = 15$
$p = 5$	1.000	1.000	1.000	1.000	0.837	0.821
$p = 10$		1.000	1.000		1.000	0.984
$p = 15$			1.000			1.000

Analysis is conducted on all subjects, all Type 1 features, and 5, 10, and 15 top Type 2 features. Similarity is evaluated between the (sub-)subgrouping structures obtained under the three different sets (numbers) of Type 2 features. Perfect agreement is observed for subgrouping, and very high similarity is observed for sub-subgrouping, suggesting that the heterogeneity analysis results are not sensitive to the number of Type 2 features.

Table C11

Data analysis using the proposed approach: comparison of clinical variables between (sub-)subgroups.

Clinical variable (sample size)		Subgroup 1	Subgroup 2	Subgroup 3	p value
Smoking history (118)	1	0	6	7	
	2	7	5	23	
	3	10	6	16	
	4	6	6	26	0.088
		Sub-subgroup 3-1	Sub-subgroup 3-2		
ICD 10 (72)	C34.0	1	0		
	C34.1	30	10		
	C34.2	2	1		
	C34.3	15	7		
	C34.30	0	4		
	C34.8	1	1		0.041

p value: from Fisher's exact test. Smoking history: category describing self-reported current smoking status and smoking history. 1: lifelong non-smoker (less than 100 cigarettes smoked in lifetime); 2: current smoker (including daily smoker and non-daily smoker or occasional smoker); 3: current reformed smoker for greater than 15 years; 4: current reformed smoker for less than or equal to 15 years. ICD 10: tumor site (coded according to the 10th revision of the International Statistical Classification of Diseases and Related Health Problems). C34.00: malignant neoplasm of unspecified main bronchus; C34.1: upper lobe; C34.2: middle lobe; C34.3: lower lobe; C34.30: lower lobe, unspecified bronchus or lung; C34.8: overlapping sites.

Table C12

Data analysis using fusion1: estimated coefficients.

	Subgroup			
Type 1 imaging feature	1	2	3	4
LymphocytesPN	-3.7020	-1.1639	0.3662	2.9136
StromaPN	0.5128	-2.4217	0.4677	0.0239
TumorPN	-0.2276	1.7177	-0.2637	-0.5213
LymphocytesSN	2.0798	1.7328	-0.3230	-3.4996
StromaSN	-0.4788	1.0927	-0.2852	0.8715
TumorSN	-0.8897	-0.7060	-0.2215	-0.2184
Selected Type 2 imaging feature	1	2	3	4
AreaShape-Center-Y	-0.3052	0.8259	-0.0007	-0.0342
AreaShape-Zernike-8-2	-0.8348	1.1945	-0.2851	0.1510
Granularity-12-ImageAfterMath	-0.8728	0.6760	-0.1072	-0.1008
Texture-Contrast-maskosingray-3-03	-0.9639	-0.0197	0.4660	1.8763
Texture-Correlation-maskosingray-3-01	-0.6555	0.0108	0.1709	1.0714
Texture-DifferenceVariance-ImageAfterMath-3-03	-1.0194	0.1269	-0.1048	2.0625
Texture-SumVariance-ImageAfterMath-3-00	-0.0542	0.6458	-0.2483	0.4846
Texture-SumVariance-maskosingray-3-01	0.8556	-0.3614	-0.3454	-4.0015
Threshold-FinalThreshold-Identifyhemasub2	-0.0754	-1.0471	-0.3465	-2.3389
Threshold-SumOfEntropies-Identifyeosinprimarycytoplasm	-0.2844	-0.5317	-0.1523	-2.5887

Table C13
Data analysis using FMR1: estimated coefficients.

	Subgroup		
Type 1 imaging feature	1	2	3
LymphocytesPN	0.0000	0.0000	0.0000
StromaPN	0.0000	0.0000	0.0000
TumorPN	0.0000	0.0000	0.0000
LymphocytesSN	0.0098	0.7285	0.0496
StromaSN	-0.0821	-0.0007	-0.0181
TumorSN	0.0000	0.0000	0.0000
Selected Type 2 imaging feature	1	2	3
AreaShape-Center-Y	0.0000	0.0000	0.0000
AreaShape-Zernike-8-2	0.0000	0.0000	0.0000
Granularity-12-ImageAfterMath	0.0000	0.0000	0.0000
Texture-Contrast-maskosingray-3-03	0.0000	0.0000	0.0000
Texture-Correlation-maskosingray-3-01	0.0000	0.0000	0.0000
Texture-DifferenceVariance-ImageAfterMath-3-03	0.0000	0.0000	0.0000
Texture-SumVariance-ImageAfterMath-3-00	0.0000	0.0000	0.0000
Texture-SumVariance-maskosingray-3-01	0.0000	0.0000	0.0000
Threshold-FinalThreshold-Identifyhemasub2	0.0000	0.0000	0.0000
Threshold-SumOfEntropies-Identifyeosinprimarycytoplasm	0.0000	0.0000	0.0000

0.0000's are estimates with small but nonzero magnitudes.

Table C14

Data analysis using respclust1: estimated coefficients.

	Subgroup			
Type 1 imaging feature	1	2	3	4
LymphocytesPN	0.0017	-0.0072	-	-
StromaPN	-0.1003	-0.1493	-	-
TumorPN	0.1141	0.1127	-	0.2370
LymphocytesSN	-0.1017	-	-	-0.3991
StromaSN	-	-	-	0.4718
TumorSN	0.2150	-0.6564	-0.0051	-
Selected Type 2 imaging feature	1	2	3	4
AreaShape-Center-Y	-0.0291	-0.0051	-	0.2962
AreaShape-Zernike-8-2	0.0558	0.0014	-0.4136	0.6206
Granularity-12-ImageAfterMath	-0.0073	0.0366	0.4140	0.2339
Texture-Contrast-maskosingray-3-03	0.0949	0.3921	-0.1167	-1.2400
Texture-Correlation-maskosingray-3-01	-0.0122	0.4058	-	-0.2052
Texture-DifferenceVariance-ImageAfterMath-3-03	0.0604	-0.0072	-	-0.6074
Texture-SumVariance-ImageAfterMath-3-00	0.0588	0.2892	-	1.6819
Texture-SumVariance-maskosingray-3-01	-0.0367	-0.6213	-	-
Threshold-FinalThreshold-Identifyhemasub2	-0.0801	-0.1152	-0.0921	-
Threshold-SumOfEntropies-Identifyeosinprimarycytoplasm	-0.0329	-0.0186	0.1955	-0.8259

0.0000's are estimates with small but nonzero magnitudes. "-" represents exactly zero.

Table C15

Data analysis using resiclust1: estimated coefficients.

	Subgroup					
Type 1 imaging feature	1	2	3	4	5	6
LymphocytesPN	-0.2291	-0.2999	-0.3908	0.3494	-0.1772	-5.7874
StromaPN	-0.8162	-0.1535	-0.3074	-	-1.3838	-
TumorPN	0.5273	0.0163	0.1683	0.2393	0.6706	-
LymphocytesSN	0.4092	0.4906	0.4431	-	0.2244	-
StromaSN	0.2822	0.1078	0.0384	-	-	-
TumorSN	-	-	-	-	-3.2858	-
Selected Type 2 imaging feature	1	2	3	4	5	6
AreaShape-Center-Y	-0.0923	-0.0939	-0.0510	-	0.0519	-
AreaShape-Zernike-8-2	0.0368	-0.0945	-0.0632	-	0.0301	-
Granularity-12-ImageAfterMath	-0.0288	-0.0072	-0.1068	-0.3546	0.0422	-
Texture-Contrast-maskosingray-3-03	0.1925	0.1725	0.1058	-0.2954	0.7146	-
Texture-Correlation-maskosingray-3-01	-0.1384	0.0030	-0.1570	-	0.4110	-
Texture-DifferenceVariance-ImageAfterMath-3-03	0.0314	0.0980	-0.0129	-	-0.1170	-
Texture-SumVariance-ImageAfterMath-3-00	-0.0531	0.0260	0.1069	-	-0.3787	-
Texture-SumVariance-maskosingray-3-01	0.0128	-0.1915	-0.2124	-	-0.7294	-
Threshold-FinalThreshold-Identifyhemasub2	-0.1393	-0.3555	-0.3188	-	-0.2659	-
Threshold-SumOfEntropies-Identifyeosinprimarycytoplasm	-0.0661	-0.1879	-0.2282	-0.8972	-0.2816	-

0.0000's are estimates with small but nonzero magnitudes. "-" represents exactly zero.

Table C16
Data analysis using FMR2: estimated coefficients.

	Subgroup		
Type 1 imaging feature	1		2
LymphocytesPN	0.0000		0.0000
StromaPN	0.0000		0.0000
TumorPN	0.0000		0.0000
LymphocytesSN	0.0180		0.7786
StromaSN	-0.0534		-0.1337
TumorSN	0.0000		0.0000
	Sub-subgroup		
Selected Type 2 imaging feature	1-1	1-2	2-1
AreaShape-Center-Y	0.0780	-0.6078	0.0361
AreaShape-Zernike-8-2	0.0000	0.0160	-
Granularity-12-ImageAfterMath	-0.1551	0.0000	-
Texture-Contrast-maskosingray-3-03	0.0000	0.6475	-
Texture-Correlation-maskosingray-3-01	0.0000	0.0000	-
Texture-DifferenceVariance-ImageAfterMath-3-03	0.0000	0.0000	-0.1585
Texture-SumVariance-ImageAfterMath-3-00	0.0000	0.5024	-
Texture-SumVariance-maskosingray-3-01	0.0000	-0.8290	-
Threshold-FinalThreshold-Identifyhemasub2	-0.1841	0.0000	-0.0244
Threshold-SumOfEntropies-Identifyeosinprimarycytoplasm	0.0000	-0.4518	-

0.0000's are estimates with small but nonzero magnitudes. "-" represents exactly zero.

Table C17

Data analysis using respclust2: estimated coefficients.

	Subgroup	
Type 1 imaging feature	1	2
LymphocytesPN	-0.3273	-0.2720
StromaPN	-1.0007	0.0278
TumorPN	0.5415	0.0999
LymphocytesSN	0.0990	0.4509
StromaSN	0.4804	-0.0865
TumorSN	-	-

	Sub-subgroup				
Selected Type 2 imaging feature	1-1	1-2	1-3	2-1	2-2
AreaShape-Center-Y	0.0304	-	-	0.0402	0.8169
AreaShape-Zernike-8-2	0.0327	0.0598	-	0.0850	-0.3311
Granularity-12-ImageAfterMath	0.0197	0.3035	-	0.0553	-0.1551
Texture-Contrast-maskosingray-3-03	0.2558	-0.3609	-	0.2078	-0.6644
Texture-Correlation-maskosingray-3-01	0.0875	0.2616	-	0.2729	-1.5099
Texture-DifferenceVariance-ImageAfterMath-3-03	0.0666	-0.1094	-	-0.0306	-
Texture-SumVariance-ImageAfterMath-3-00	0.0886	-	-	0.1380	-
Texture-SumVariance-maskosingray-3-01	-0.2548	-	-	-0.1992	-
Threshold-FinalThreshold-Identifyhemasub2	-0.1543	0.1024	-	-0.0236	-
Threshold-SumOfEntropies-Identifyeosinprimarycytoplasm	-0.0998	-	-	-0.0251	-1.0401

0.0000's are estimates with small but nonzero magnitudes. "-" represents exactly zero.

Table C18

Data analysis using resiclust2: estimated coefficients.

		Subgroup	
Type 1 imaging feature	1	2	3
LymphocytesPN	-0.0020	0.2730	-1.7048
StromaPN	-0.1137	0.5963	-1.9215
TumorPN	0.1662	-0.3384	1.1271
LymphocytesSN	0.0336	0.0277	1.2224
StromaSN	-	-0.4428	0.8654
TumorSN	-0.1130	-	-

				Sub-subgroup				
Selected Type 2 imaging feature	1-1	1-2	2-1	2-2	2-3	3-1	3-2	3-3
AreaShape-Center-Y	-0.0483	-0.0065	-	-0.0262	-	-0.0274	0.0491	-
AreaShape-Zernike-8-2	0.0737	0.0253	-	0.0595	-0.1547	-0.1055	-0.2064	-
Granularity-12-ImageAfterMath	0.0250	-0.0197	0.1181	0.0988	-0.0135	0.1005	0.0422	-
Texture-Contrast-maskosingray-3-03	-0.1168	-0.1053	-	-0.6924	-	-0.1256	-	-
Texture-Correlation-maskosingray-3-01	-0.0973	-0.0812	-	-0.0626	-	-	-0.2468	-
Texture-DifferenceVariance-ImageAfterMath-3-03	0.0614	0.0011	-	1.0835	0.9192	-0.1561	-	-
Texture-SumVariance-ImageAfterMath-3-00	-0.0526	0.0523	-	0.1834	-	-	-	-
Texture-SumVariance-maskosingray-3-01	0.2419	0.0918	0.1505	0.5410	-	0.0289	-	-
Threshold-FinalThreshold-Identifyhemasub2	0.0703	0.0215	-	-0.0617	-	-	0.0499	-
Threshold-SumOfEntropies-Identifyeosinprimarycytoplasm	0.0667	0.0387	0.0281	0.0318	-	-	-	-

0.0000's are estimates with small but nonzero magnitudes. "-" represents exactly zero.

Table C19

Data analysis: concordance in subgrouping between different approaches.

	fusion1	FMR1	respclust1	resiclust1	FMR2	respclust2	resiclust2	proposed
fusion1	(1.000,1.000)	(0.594,0.596)	(0.654,0.646)	(0.681,0.668)	(0.514,0.568)	(0.616,0.658)	(0.657,0.671)	(0.837,0.975)
FMR1		(1.000,1.000)	(0.819,0.819)	(0.763,0.763)	(0.837,0.899)	(0.758,0.807)	(0.814,0.764)	(0.547,0.594)
respclust1			(1.000,1.000)	(0.857,0.857)	(0.695,0.762)	(0.908,0.958)	(0.837,0.856)	(0.555,0.654)
resiclust1				(1.000,1.000)	(0.634,0.704)	(0.808,0.869)	(0.902,0.946)	(0.555,0.681)
FMR2					(1.000,1.000)	(0.751,0.746)	(0.709,0.701)	(0.510,0.561)
respclust2						(1.000,1.000)	(0.813,0.877)	(0.530,0.670)
resiclust2							(1.000,1.000)	(0.562,0.687)
proposed								(1.000,1.000)

In each cell, (subgrouping consistency – as defined in Simulation – based on Type 1 imaging features, subgrouping consistency based on Type 2 imaging features).

Table C20

Simulation based on the LUAD data. In each cell, mean (sd).

		$\boldsymbol{\beta}$				$\boldsymbol{\gamma}$			
	method	per	$\widehat{K_1}$	SC	MSE	per	$\widehat{K_2}$	SC	MSE
balanced	fusion1	0.00	3.15(1.02)	0.767(0.035)	1.521(0.203)	0.58	3.15(1.02)	0.631(0.024)	1.167(0.593)
	FMR1	0.44	2.81(0.87)	0.674(0.110)	1.576(0.456)	0.00	2.81(0.87)	0.512(0.054)	1.648(0.438)
	respclust1	0.09	5.78(1.31)	0.644(0.056)	2.373(0.285)	0.13	5.78(1.31)	0.428(0.028)	2.033(0.394)
	resiclust1	0.12	5.83(0.96)	0.661(0.043)	1.969(0.364)	0.11	5.83(0.96)	0.462(0.037)	2.231(0.413)
	FMR2	0.49	2.51(0.50)	0.664(0.022)	3.127(0.582)	0.16	6.96(1.65)	0.556(0.031)	2.072(0.496)
	respclust2	0.13	2.87(0.34)	0.685(0.011)	3.957(0.644)	0.01	7.65(1.10)	0.495(0.023)	2.173(0.486)
	resiclust2	0.11	2.89(0.31)	0.684(0.013)	3.070(0.523)	0.00	7.81(1.01)	0.500(0.020)	2.386(0.512)
	proposed	0.84	2.28(0.89)	0.774(0.028)	1.479(0.179)	0.61	4.77(1.12)	0.636(0.019)	1.136(0.659)
imbalanced	fusion1	0.00	3.19(1.10)	0.760(0.041)	1.702(0.231)	0.60	3.19(1.10)	0.625(0.019)	1.567(0.489)
	FMR1	0.40	2.88(0.76)	0.697(0.085)	1.437(0.384)	0.00	2.88(0.76)	0.536(0.048)	1.608(0.386)
	respclust1	0.06	5.39(1.86)	0.618(0.086)	2.278(0.694)	0.16	5.39(1.86)	0.487(0.064)	1.844(0.468)
	resiclust1	0.09	5.95(1.92)	0.634(0.079)	3.697(0.596)	0.15	5.95(1.92)	0.465(0.052)	1.966(0.534)
	FMR2	0.52	2.48(0.50)	0.676(0.028)	2.974(0.538)	0.18	6.75(1.54)	0.572(0.042)	1.927(0.504)
	respclust2	0.22	2.78(0.42)	0.592(0.027)	3.263(0.463)	0.02	7.63(1.14)	0.514(0.053)	1.783(0.531)
	resiclust2	0.14	2.86(0.35)	0.598(0.034)	2.610(0.384)	0.01	6.98(1.24)	0.495(0.049)	1.747(0.497)
	proposed	0.88	2.02(0.43)	0.820(0.048)	1.326(0.205)	0.64	3.53(1.08)	0.644(0.044)	1.420(0.375)

Table C21
Data analysis based on PCA: estimated coefficients.

	Subgroup	
Type 1 imaging feature	1	2
LymphocytesPN	1.1635	-0.4779
StromaPN	0.8790	-0.4699
TumorPN	-0.2417	0.4039
LymphocytesSN	-2.1352	0.8060
StromaSN	-1.4939	0.2995
TumorSN	-1.6771	0.6515

	Sub-subgroup				
Type 2 imaging feature	1-1	2-1	2-2	2-3	2-4
PC 1	-0.0945	0.4132	-0.3529	0.4205	0.0341
PC 2	-1.3002	0.1978	-0.0058	0.5619	-0.0298
PC 3	-1.6011	-0.2111	-0.1239	0.0017	-0.2781
PC 4	-0.4473	-0.2193	0.1531	-0.0727	0.0434
PC 5	-0.1764	0.3298	0.0611	0.2125	-0.2703
PC 6	0.9042	0.3528	0.0270	-0.0535	0.0292
PC 7	-0.2324	-0.0515	-0.1611	0.0303	0.2007
PC 8	-0.5821	0.1377	-0.3605	0.6537	0.2847
PC 9	-0.1411	0.2147	-0.6851	0.7467	0.0023
PC 10	0.0145	0.0934	0.0950	0.1590	0.2564