scientific reports

OPEN

Debiased inference for heterogeneous subpopulations in a high-dimensional logistic regression model

Hyunjin Kim, Eun Ryung Lee & Seyoung Park

Due to the prevalence of complex data, data heterogeneity is often observed in contemporary scientific studies and various applications. Motivated by studies on cancer cell lines, we consider the analysis of heterogeneous subpopulations with binary responses and high-dimensional covariates. In many practical scenarios, it is common to use a single regression model for the entire data set. To do this effectively, it is critical to quantify the heterogeneity of the effect of covariates across subpopulations through appropriate statistical inference. However, the high dimensionality and discrete nature of the data can lead to challenges in inference. Therefore, we propose a novel statistical inference method for a high-dimensional logistic regression model that accounts for heterogeneous subpopulations. Our primary goal is to investigate heterogeneity across subpopulations by testing the equivalence of the effect of a covariate and the significance of the overall effects of a covariate. To achieve overall sparsity of the coefficients and their fusions across subpopulations, we employ a fused group Lasso penalization method. In addition, we develop a statistical inference method that incorporates bias correction of the proposed penalized method. To address computational issues due to the nonlinear log-likelihood and the fused Lasso penalty, we propose a computationally efficient and fast algorithm by adapting the ideas of the proximal gradient method and the alternating direction method of multipliers (ADMM) to our settings. Furthermore, we develop non-asymptotic analyses for the proposed fused group Lasso and prove that the debiased test statistics admit chi-squared approximations even in the presence of high-dimensional variables. In simulations, the proposed test outperforms existing methods. The practical effectiveness of the proposed method is demonstrated by analyzing data from the Cancer Cell Line Encyclopedia (CCLE).

Significant efforts have been made in recent research to perform screening genetic profiling and drug testing in human cancer cell lines to explore how genomic backgrounds influence response to therapy[1]. These efforts have resulted in valuable cancer cell line (CCL) data resources, such as the Cancer Cell Line Encyclopedia (CCLE)[2]. CCLE data provide responses to 24 anticancer drugs in hundreds of cell lines across multiple tumor types, along with genomic information about these cell lines, such as the expression of nearly 20,000 genes. These data are often used to build computational models to predict drug response[3]. For example, to predict drug response using gene expression, researchers have used several methods, including Ridge regression[4], mixture regression[5], support vector machine[6], random forest[7], and neural networks[8].

Given that recent large public CCL datasets, such as CCLE, include multiple cancer types, there is a need to develop statistical inference to assess the heterogeneity of the effects of gene expression on a drug across multiple tumor types. Quantifying the heterogeneity of the effects of genes across different tumors can provide valuable information in drug response modeling, as cancer tissue heterogeneity must be considered when modeling drug response across different cancer types[9]. Different tumor types may be characterized by their own tumor-specific genes, and thus tumor type interactions may need to be considered. In addition, it is often observed that patients with different types of cancer have similar expression patterns of certain genes, but the efficacy of treatments varies. For example, HER2 overexpression has been observed in subsets of patients with cancers such as breast and gastric cancer, but a HER2 inhibitor, pertuzumab, is more effective in treating HER2-positive breast cancer compared to HER2-positive gastric cancer[10]. Therefore, interactions with tumor type may need to be considered

Department of Statistics, Sungkyunkwan University, Seoul 100190, South Korea. email: erlee@skku.edu; ishspsy@skku.edu

for more accurate statistical analysis. Several studies have attempted to develop statistical modeling methods for drug response that account for heterogeneity between different cancer types[11,12]. However, a common approach is still to analyze several tumor types simultaneously[2,13]. This approach assumes that there are no interaction effects with tumor-specific factors. Such inconsistencies in the current literature raises a critical research question regarding the uniformity of the effects of gene expression on drug response across different cancer types. To date, this question remains unanswered. This paper aims to fill this gap by developing a statistical test for homogeneity in the effects of gene expression on drug response across multiple cancer types. This task is particularly challenging due to the high-dimensional nature of CCLE data. In a different context, heterogeneity testing between studies in meta-analyses has been investigated[14,15].

In this study, we focus on statistical inference for the heterogeneity of the effects of gene expression on drug sensitivity across different cancer types representing distinct subpopulations. We consider a binary response setting, where the response for a cell line represents whether that cell line is sensitive or resistant to a drug, because determining whether a patient is sensitive or resistant to anticancer drugs is critical to treatment. More specifically, we focus on statistical inference in a high-dimensional logistic regression model where the population is stratified into heterogeneous subpopulations, i.e., cancer types. We consider the problem of testing whether a given covariate has the same effect on the binary response in different subpopulations in high-dimensional logistic regression settings as follows: for some $j \in \{1, \ldots, p\}$ and $c \in \mathbb{R}$,

$$H_0 : \beta_j^{(1)} = \cdots = \beta_j^{(G)} = c \quad \text{vs} \quad H_1 : \text{ not } H_0, \tag{1}$$

where $\beta_j^{(g)}$ is the underlying regression coefficient of the jth covariate for the gth subpopulation, and c can be specified, e.g. $c = 0$, or c can be unspecified. The null hypothesis with unspecified c indicates homogeneity, i.e., equal effects of the covariate. Testing (1) with unspecified c can provide valuable information in pharmacogenomics research, which studies how genes influence response to a drug. In addition, it may be helpful in understanding variations in drug response between cancer types in terms of gene expression. On the other hand, specifying c, e.g. $c = 0$, suggests that the covariate has zero effect, i.e., it is not significant. Testing (1) with $c = 0$ can provide candidates for gene expression markers of drug sensitivity that can be applied to multiple cancer types. Such versatile gene expressions are very useful for research and clinical settings[16]. While testing (1) can provide valuable insights, standard maximum likelihood estimation cannot be used for CCLE data analysis due to the high dimensionality of genes compared to the number of cell lines for tumor types. In recent years, much effort has been devoted to statistical inference for high-dimensional generalized linear models. Several studies have considered inference for either low-dimensional coefficients or a single coefficient in the presence of a large number of nuisance parameters. For example, Van de Geer et al.[17] studied the theoretical properties of a bias-corrected Lasso[18] estimator called desparsified Lasso or debiased Lasso. Theoretical properties of different types of bias-corrected estimators have also been studied under high-dimensional linear regression settings[19,20]. In addition, Ning and Liu[21] proposed a decorrelated score test that can be applied to generic penalized M estimators. Other researchers have considered interval estimation for a single coefficient[22,23]. More recently, Ma et al.[24] considered the multiple testing problem for high-dimensional logistic regression in two-sample settings.

Despite the nice theoretical properties of these methods in high-dimensional regimes, these methods were developed in classical single-population settings and thus may not be optimal for analyzing data consisting of heterogeneous subpopulations. In addition, the combination of limited sample sizes for cancer types and the high dimensionality of gene expression data poses challenges in obtaining accurate results when testing the homogeneity or significance of a gene's effect on a drug across different cancer types. Instead of considering inference based on Lasso penalized estimation, a common approach in existing high-dimensional inference, we propose statistical inference based on the fused group Lasso. Our proposed testing procedure consists of two steps. In the first step, we compute a suitable estimator for the underlying coefficients using variable selection and homogeneity detection. Our proposed objective function includes a negative log-likelihood, a group Lasso-type penalty[25], and a fusion-type penalty[26]. The group Lasso-type penalty controls overall sparsity in the model, i.e., removes irrelevant covariates in all subpopulations, while the fusion-type penalty promotes similarities between coefficients across subpopulations, which can improve estimation efficiency by clustering regression coefficients[27]. Thus, the combination of group Lasso-type and fusion-type penalties allows for sparsity control and integration of samples from different cancer types, effectively addressing the challenges posed by small sample sizes for cancer types. The effectiveness of our tests over Lasso-based approaches is demonstrated by a simulation model in section "Simultation study for an imbalanced design", where sample sizes for subpopulations are relatively limited. In the second step, we develop a bias correction procedure to correct the bias of our fused group Lasso estimator obtained in the first step. In essence, we extend the de-sparsified Lasso concept of Javanmard and Montanari[19] to the fused group Lasso method, thereby facilitating statistical inference in high-dimensional regression scenarios. This advancement requires rigorous theoretical investigations of the fused group Lasso technique. However, to the best of our current knowledge, there are limited theoretical investigations of fused group Lasso problems.

Note that Zhou et al.[28] considered fused sparse group Lasso in a multiple response linear regression model without rigorous convergence and inference analyses. The theoretical analysis of the combined penalty of group Lasso-type and fusion-type is highly non-trivial. In particular, the penalty function is not decomposable with respect to the support of the parameter, which means that the existing theory of decomposable regularizers[29] cannot be applied[30]. In addition, we provide a computationally efficient algorithm to address the computational difficulties arising from the logistic log-likelihood and the fused Lasso penalty. By integrating the principles of the proximal gradient method and the ADMM into our framework, we effectively handle the computational complexity associated with our settings. Furthermore, our theory and methods can be easily extended to general fused group lasso settings, such as those with discrete responses.

To test the hypothesis (1), existing high-dimensional inference methods based on ℓ_1 norm penalization, such as the Lasso, could be considered. A naive approach would be to test (1) by applying a separate penalized logistic regression for each subpopulation, using an existing ℓ_1 norm-based penalization method, such as the debiased Lasso[17]. However, when the sparsity patterns of the underlying coefficients are similar across different subpopulations—as is the case in the CCLE data example—the proposed fused group Lasso approach can combine subpopulations and increase the sample size, resulting in better performance. This was also observed in our simulation analysis and in different contexts[31].

The rest of the paper is organized as follows. Section "Method and theory" describes the proposed penalization method, its debiased version, and the test statistics with their theoretical properties. Section "Implementation" presents an algorithm that solves the proposed penalization method. Section "Simulation study" examines the finite sample performance of the proposed test along with other competing methods. Section "Application to the CCLE data" illustrates the application of our approach to the Cancer Cell Line Encyclopedia (CCLE) data[2]. Finally, Section "Conclusion" concludes the paper. Additional simulation results are presented in the Supplementary material.

Method and theory

In this section, we introduce the proposed penalization and test statistics. First, we introduce the notations that will be used throughout the paper. For any positive integer d, let $\boldsymbol{I}_d$ be the $d \times d$ identity matrix, and $\mathbf{1}_d$ and $\mathbf{0}_d$ be the $d \times 1$ vectors of all 1's and 0's, respectively. Let $\mathbf{0}_{d_1 \times d_2}$ and $\mathbf{1}_{d_1 \times d_2}$ be $d_1 \times d_2$ matrices whose entries are all 0's and 1's, respectively. When the size of the matrix is obvious, the subscript is sometimes omitted. For an index $1 \le l \le d$, let $\boldsymbol{e}_l$ denote the $p \times 1$ vector with one at the lth location and zero everywhere else. For any $d \times 1$ vector $\boldsymbol{a} = (a_1, \ldots, a_d)^\top$, let $\|\boldsymbol{a}\|_q := \left(\sum_{\ell=1}^d a_\ell^q\right)^{1/q}$ for $1 \le q < \infty$ and $\|\boldsymbol{a}\|_{\max} := \max_{1 \le \ell \le d} |a_\ell|$. For a set S, let $|S|$ denote the cardinality of S. For a vector $\boldsymbol{a}$ and an index set of elements, say S, let $\boldsymbol{a}_S$ be the $|S| \times 1$ sub-vector of $\boldsymbol{a}$ with elements in S. For a matrix $\boldsymbol{A} = (A_{ij})_{d_1 \times d_2}$, we let

$$\|\boldsymbol{A}\|_{\max} := \max_{i,j} |A_{ij}|, \quad \|\boldsymbol{A}\|_1 := \sum_{i,j} |A_{ij}|, \quad \|\boldsymbol{A}\|_F := \sqrt{\sum_{i,j} A_{ij}^2}, \quad |||\boldsymbol{A}|||_1 := \max_j \sum_i |A_{ij}|,$$

and use vec $(\boldsymbol{A})$ to represent vectorization by staking the columns of a matrix $\boldsymbol{A}$. For a symmetric matrix $\boldsymbol{A}$, let $\lambda_{\min}(\boldsymbol{A})$ be the minimum eigenvalue of $\boldsymbol{A}$. Given a matrix $\boldsymbol{A}$ and an index set of rows, say R, let $\boldsymbol{A}_{R,\cdot}$ denote sub-matrix of $\boldsymbol{A}$ with rows in R. We also use $\boldsymbol{A}_{i,\cdot}$ to represent the ith row of $\boldsymbol{A}$. For any vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ of equal length, let $\langle \boldsymbol{a}, \boldsymbol{b} \rangle := \sum_i a_i b_i$. For any matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ with equal dimensions, let $\langle \boldsymbol{A}, \boldsymbol{B} \rangle := \sum_{i,j} A_{ij} B_{ij}$. For non-negative sequences $\{a_n\}_{n=1}^\infty$ and $\{b_n\}_{n=1}^\infty$, we write $a_n \ll b_n$ or $a_n = o(b_n)$ if $b_n > 0$ and $a_n / b_n \to 0$. We also write $a_n = O(b_n)$ or $a_n \lesssim b_n$ if $a_n \le C b_n$ for some positive constant C. Let $a \vee b$ and $a \wedge b$ denote $\max(a, b)$ and $\min(a, b)$, respectively.

Setting and problem

Here we present a logistic regression model for heterogeneous subpopulations. Suppose the data come from G independent subpopulations, such as tumor sample groups generated independently in CCLE data. For each subpopulation $1 \le g \le G$, there exist n_g pairs $\left\{\mathbf{x}_i^{(g)}, y_i^{(g)}\right\}_{i=1}^{n_g}$, where $y_i^{(g)} \in \{0, 1\}$ represents a binary response (e.g., binary drug response) of the ith subject in the gth group (e.g., tumor group) and $\mathbf{x}_i^{(g)} = \left(x_{i1}^{(g)}, \ldots, x_{ip}^{(g)}\right)^\top \in \mathbb{R}^p$ represents a p-dimensional vector of covariates, e.g. gene expression variables. We consider the following logistic regression model

$$y_i^{(g)} \mid \mathbf{x}_i^{(g)} \sim \text{ Bernoulli} \left(\frac{\exp\left(\left[\mathbf{x}_i^{(g)}\right]^\top \boldsymbol{\beta}^{(g)}\right)}{1 + \exp\left(\left[\mathbf{x}_i^{(g)}\right]^\top \boldsymbol{\beta}^{(g)}\right)} \right) \tag{2}$$

for each group $g = 1, \ldots, G$ and samples $i = 1, \ldots, n_g$, where $\boldsymbol{\beta}^{(g)} := (\beta_1^{(g)}, \ldots, \beta_p^{(g)})^\top$ represents the underlying group-specific coefficient vector for the gth group. Let $\boldsymbol{\beta}_{(j)} := \left(\beta_j^{(1)}, \ldots, \beta_j^{(G)}\right)^\top$ denote the underlying coefficient vector for the jth covariate. We consider the high-dimensional setting where the number of covariates p increases with samples sizes $n_1, \ldots, n_G$. Let $n = \sum_{g=1}^G n_g$ be the total sample size. We assume that the groups are heterogeneous but share similar characteristics in the sense that most regression covariates have similar effects on the response across different groups. In this study, we consider the case where only a few covariates, e.g., a small number of gene expressions in the CCLE data example, are relevant to the response across different groups, i.e., the index set $S := \{j : \|\boldsymbol{\beta}_{(j)}\|_2 \ne 0\}$ is sparse in that $s := |S| \ll p$. We also assume that only a few pairs of groups have different covariate effects, i.e., $\Omega := \left\{(j, g, g') : \beta_j^{(g)} \ne \beta_j^{(g')}\right\}$ is sparse in that $\tilde{s} := |\Omega| \ll sG^2$.

For a matrix notation, let

$$\mathbf{y}^{(g)} = \left[y_1^{(g)}, \ldots, y_{n_g}^{(g)}\right]^\top \in \mathbb{R}^{n_g}$$

$$\mathbf{X}^{(g)} = \left[\mathbf{x}_1^{(g)}, \ldots, \mathbf{x}_{n_g}^{(g)}\right]^\top \in \mathbb{R}^{n_g \times p}$$

for $1 \le g \le G$, where each column of $\mathbf{X}^{(g)}$ has a zero mean and ℓ_2 norm $\sqrt{n_g}$. Let $\mathbf{y}$ be the binary response vector and $\mathbf{X}$ be the covariate matrix defined by

$$\mathbf{y} = \left[(\mathbf{y}^{(1)})^\top, \ldots, (\mathbf{y}^{(G)})^\top\right] \in \mathbb{R}^n$$
$$\mathbf{X} = \operatorname{diag}\left(\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(G)}\right) \in \mathbb{R}^{n \times pG},$$

where $\mathbf{X}$ is a block-diagonal matrix, consisting of $\mathbf{X}^{(g)}$'s. Define the coefficient matrix $\boldsymbol{B}$ as $\boldsymbol{B} = \left[\boldsymbol{\beta}^{(1)}, \ldots, \boldsymbol{\beta}^{(G)}\right] \in \mathbb{R}^{p \times G}$. Our main goal is to test the homogeneity of the effects of the lth covariate across the G groups, i.e., for $1 \le l \le p$,

$$H_0 : \beta_l^{(1)} = \cdots = \beta_l^{(G)} \quad \text{vs} \quad H_1 : \text{ not } H_0. \tag{3}$$

Another goal is to test whether the lth covariate is significant to at least one of the groups:

$$H_0 : \beta_l^{(1)} = \cdots = \beta_l^{(G)} = 0 \quad \text{vs} \quad H_1 : \text{ not } H_0. \tag{4}$$

The penalization method

We propose to test (3) or (4) based on the penalized method using the fused group Lasso. For any $y \in \{0, 1\}$ and $\nu \in \mathbb{R}$, define $\ell(y, \nu) = -y\nu + \log(1 + \exp(\nu))$. For the loss function $\ell(y, \nu)$, let $\dot{\ell}(y, \nu)$ and $\ddot{\ell}(y, \nu)$ denote its first and second derivatives with respect to ν, respectively. For a matrix $\boldsymbol{\Delta} = (\Delta_{jg})_{1 \le j \le p, 1 \le g \le G} = \left(\boldsymbol{\Delta}^{(1)}, \ldots, \boldsymbol{\Delta}^{(G)}\right) = \left(\boldsymbol{\Delta}_{(1)}^\top, \ldots, \boldsymbol{\Delta}_{(p)}^\top\right)^\top$, let

$$L_n(\boldsymbol{\Delta}) = \frac{1}{n} \sum_{g=1}^{G} \sum_{i=1}^{n_g} \ell\left(y_i^{(g)}, \left(\mathbf{x}_i^{(g)}\right)^\top \boldsymbol{\Delta}^{(g)}\right).$$

We propose $\hat{\boldsymbol{B}}$, which solves the following optimization problem:

$$\hat{\boldsymbol{B}} := \underset{\boldsymbol{\Delta} \in \mathbb{R}^{p \times G}}{\arg\min}\, L_n(\boldsymbol{\Delta}) + \lambda_1 \sum_{j=1}^{p} w_j \|\boldsymbol{\Delta}_{(j)}\|_2 + \lambda_2 \sum_{j=1}^{p} \sum_{g < g'} v_{j,gg'} |\Delta_{jg} - \Delta_{jg'}|, \tag{5}$$

where λ_1 and λ_2 are non-negative penalty parameters and w_j's and $v_{j,gg'}$'s are non-negative weights.

Let $\hat{\boldsymbol{B}} = \left(\hat{\beta}_{jg}\right)_{1 \le j \le p, 1 \le g \le G} = \left(\hat{\boldsymbol{\beta}}^{(1)}, \ldots, \hat{\boldsymbol{\beta}}^{(G)}\right) = \left(\hat{\boldsymbol{\beta}}^\top_{(1)}, \cdots, \hat{\boldsymbol{\beta}}^\top_{(p)}\right)^\top$. In (5), the two penalty terms are based on the sparsity assumption, as explained in section "Setting and problem": The group Lasso-type penalty promotes overall sparsity in the total coefficients, i.e., many $\hat{\boldsymbol{\beta}}_{(j)}$'s are the zero vector; while the fusion-type penalty promotes similarities among $\hat{\beta}_{jg}$'s for $1 \le g \le G$. For the weights w_j and $v_{j,gg'}$, we consider the following optimization problem:

$$\tilde{\boldsymbol{B}} := \underset{\boldsymbol{\Delta} \in \mathbb{R}^{p \times G}}{\arg\min}\, L_n(\boldsymbol{\Delta}) + \tilde{\lambda}_1 \sum_{j=1}^{p} \|\boldsymbol{\Delta}_{(j)}\|_2 + \tilde{\lambda}_2 \sum_{j=1}^{p} \sum_{g < g'} |\Delta_{jg} - \Delta_{jg'}|. \tag{6}$$

Let $\tilde{\boldsymbol{B}} = \left(\tilde{\beta}_{jg}\right)_{1 \le j \le p, 1 \le g \le G} = \left(\tilde{\boldsymbol{\beta}}^{(1)}, \ldots, \tilde{\boldsymbol{\beta}}^{(G)}\right) = \left(\tilde{\boldsymbol{\beta}}^\top_{(1)}, \ldots, \tilde{\boldsymbol{\beta}}^\top_{(p)}\right)^\top$. Following Zou[32], we set

$$w_j = 1/\|\tilde{\boldsymbol{\beta}}_{(j)}\|_2, \quad v_{j,gg'} = 1/|\tilde{\beta}_{jg} - \tilde{\beta}_{jg'}|.$$

The following theorem presents an estimation error bound for the initial estimator $\tilde{\boldsymbol{B}}$. For the details of the Conditions 1–2 assumed to derive Theorem 1, please refer to section S3.1 of the Supplementary materials.

Theorem 1 *Assume that Conditions* 1–2 *in the Supplementary materials hold, and* $\max_{1 \le g \le G} \|\boldsymbol{\beta}^{(g)}\|_1 \le \tilde{C}$ *for some absolute constant* $\tilde{C} > 0$. *Let the penalty parameters* $\tilde{\lambda}_1$ *and* $\tilde{\lambda}_2$ *be chosen so that*

$$\tilde{\lambda}_1 \ge \sqrt{\frac{16G(\log p + \log G)}{n}}, \quad \tilde{\lambda}_2 = G^{-3/2} \tilde{\lambda}_1.$$

Then, it holds that with probability at least $1 - 1/(pG)$,

$$\|\tilde{\boldsymbol{B}} - \boldsymbol{B}\|_F^2 \lesssim \tilde{\lambda}_1^2 s + \tilde{\lambda}_2^2 \tilde{s}.$$

Theorem 1 implies that if we take

$$\tilde{\lambda}_1 \asymp \sqrt{\frac{G(\log p + \log G)}{n}}, \quad \tilde{\lambda}_2 \asymp \sqrt{\frac{\log p + \log G}{nG^2}},$$

then the initial estimator $\tilde{\boldsymbol{B}}$ satisfies

$$\|\tilde{\boldsymbol{B}} - \boldsymbol{B}\|_F^2 = O_p\left(\left(sG + \frac{\tilde{s}}{G^2}\right)\frac{\log(p \vee G)}{n}\right).$$

The following theorem presents the theoretical properties of the estimator $\hat{\boldsymbol{B}}$. For the details of the additional Conditions 3–4 required to prove Theorem thm2, please see section S3.1 in the Supplementary materials.

Theorem 2 *Assume that the conditions of Theorem 1 and Conditions 3–4 in the Supplementary materials hold. Consider a minimizer $\hat{\boldsymbol{B}}$ with $\lambda_2 = G^{-3/2}\lambda_1$ and*

$$8\tilde{C}\sqrt{\frac{G(\log p + \log G)}{n}} \le \lambda_1 \asymp \sqrt{\frac{G(\log p + \log G)}{n}}.$$

Define the estimators of S and Ω by

$$\hat{S} := \{j : \ \|\hat{\boldsymbol{\beta}}_{(j)}\|_2 \neq 0\}, \ \hat{\Omega} := \{(j, g, g') : \hat{\beta}_{jg} \neq \hat{\beta}_{jg'}\}.$$

Then, $P(\hat{S} = S) \to 1$ and $P(\hat{\Omega} = \Omega) \to 1$.

Theorem 2 shows that the proposed fused group Lasso consistently selects relevant covariates and finds the same covariate effects across groups under additional conditions on the correlations between relevant and irrelevant covariates and on the minimum signal strengths. However, if a covariate of interest j is not selected in $\hat{\boldsymbol{B}}$, i.e., $j \notin \hat{S}$, statistical inference under the original model is impossible for $\hat{\boldsymbol{B}}$, while the debiased version of $\hat{\boldsymbol{B}}$ may give some useful information, e.g., asymptotic distributions and p-values.

Debiased test

In this subsection, we develop the debiased version of the fused group Lasso estimator $\hat{\boldsymbol{B}}$ for statistical inference. For $1 \le g \le G$, let

$$\boldsymbol{\Sigma}^{(g)} := E\left[n_g^{-1}\sum_{i=1}^{n_g}\ddot{\ell}\left(y_i^{(g)}, [\mathbf{x}_i^{(g)}]^\top \boldsymbol{\beta}^{(g)}\right)\mathbf{x}_i^{(g)}[\mathbf{x}_i^{(g)}]^\top\right]$$
$$\boldsymbol{M}^{(g)} := (\boldsymbol{\Sigma}^{(g)})^{-1}.$$

Let $\hat{\boldsymbol{\Sigma}}^{(g)}$ be an estimator of $\boldsymbol{\Sigma}^{(g)}$ defined by

$$\hat{\boldsymbol{\Sigma}}^{(g)} := n_g^{-1}\sum_{i=1}^{n_g}\ddot{\ell}\left(y_i^{(g)}, [\mathbf{x}_i^{(g)}]^\top \hat{\boldsymbol{\beta}}^{(g)}\right)\mathbf{x}_i^{(g)}[\mathbf{x}_i^{(g)}]^\top.$$

Using the main idea of Javanmard and Montanari[19], let $\hat{\boldsymbol{M}}^{(g)} := \left[\hat{\boldsymbol{m}}_1^{(g)}, \ldots, \hat{\boldsymbol{m}}_p^{(g)}\right]^\top \in \mathbb{R}^{p \times p}$ be the estimate of $\boldsymbol{M}^{(g)}$ defined by solving the quadratic programming

$$\hat{\boldsymbol{m}}_j^{(g)} = \mathop{\arg\min}_{\boldsymbol{m} \in \mathbb{R}^p} \frac{1}{2}\boldsymbol{m}^\top \hat{\boldsymbol{\Sigma}}^{(g)}\boldsymbol{m} \quad \text{subject to} \ \|\hat{\boldsymbol{\Sigma}}^{(g)}\boldsymbol{m} - \boldsymbol{e}_j\|_{\max} \le \mu_g \tag{7}$$

for each $1 \le j \le p$ and $1 \le g \le G$, where $\mu_g = C_1\sqrt{\log p/n_g}$ for some constant $C_1 > 0$. We propose a debiased estimator: for $g = 1, \ldots, G$,

$$\hat{\boldsymbol{b}}^{(g)} := \hat{\boldsymbol{\beta}}^{(g)} - \frac{\hat{\boldsymbol{M}}^{(g)}}{n_g}\sum_{i=1}^{n_g}\dot{\ell}\left(y_i^{(g)}, [\mathbf{x}_i^{(g)}]^\top \hat{\boldsymbol{\beta}}^{(g)}\right)\mathbf{x}_i^{(g)}.$$

In the debiased estimation approach, Javanmard and Montanari[19] considered only linear regression, while Van de Geer et al.[17] examined a generalized linear model. Van de Geer et al.[17] used nodewise regression[33] to estimate standard errors for the debiased Lasso estimator. While we use the bias correction technique developed by Javanmard and Montanari[19], our proposed method differs from existing methods. Specifically, we use the quadratic programming considered in Javanmard and Montanari[19] to estimate $\boldsymbol{M}^{(g)}$s instead of approximating the inverse of the sample covariance matrix for covariates. Compared to our method, most existing debiased inference methods for high-dimensional generalized linear models use nodewise regression (e.g. Ma et al.[24], Tian and Feng[34], Caner[35]). In addition, most of these existing methods were developed for statistical inference about a parameter in the classical single-population setting, and thus are Lasso to estimate parameters for a single population. In contrast, we develop tests based on our fused group Lasso for inference about parameters for multiple subpopulations.

To derive an asymptotic normality of the debiased Lasso[17], Van de Geer et al.[17] assumed that each row of $\boldsymbol{M}^{(g)}$ has a small number of non-zero entries. However, this assumption of exact sparsity may not hold in the generalized linear model, as pointed out by Xia et al.[36]. As demonstrated in Javanmard and Montanari[19], we can achieve asymptotic normality of the proposed debiased estimator without assuming exact sparsity for $\boldsymbol{M}^{(g)}$. Theorem 3 presents the theoretical properties of the proposed debiased estimator $\hat{\boldsymbol{b}}^{(g)}$. See section S3.2 of the Supplementary materials for the details of Condition 5, which is crucial for proving Theorem thm123.

Theorem 3 *Suppose conditions of Theorem 2 and Condition 5 in the Supplementary materials hold. Then, the debiased estimator satisfies*

$$\sqrt{n_g}\left(\hat{\boldsymbol{b}}^{(g)} - \boldsymbol{\beta}^{(g)}\right) = -\frac{\hat{\boldsymbol{M}}^{(g)}}{\sqrt{n_g}} \sum_{i=1}^{n_g} \dot{\ell}\left(y_i^{(g)}, \left[\mathbf{x}_i^{(g)}\right]^\top \boldsymbol{\beta}^{(g)}\right)\mathbf{x}_i^{(g)} + \boldsymbol{\Delta}_g,$$

where $\max_g \|\boldsymbol{\Delta}_g\|_{\max} = o_p(1)$.

Let $\hat{\boldsymbol{V}}^{(g)} := \hat{\boldsymbol{M}}^{(g)}\hat{\boldsymbol{\Sigma}}^{(g)}\left[\hat{\boldsymbol{M}}^{(g)}\right]^\top$ and $\hat{\boldsymbol{V}}_{(j)}$ be the $G \times G$ diagonal matrix with diagonal elements $\left\{\frac{1}{n_g}\hat{V}_{jj}^{(g)}\right\}_{g=1}^{G}$. Let $\hat{\boldsymbol{b}}_{(j)} := \left[\hat{b}_j^{(1)}, \ldots, \hat{b}_j^{(G)}\right]^\top$. Define $\boldsymbol{S}_j = \left[S_{j1}, \ldots, S_{jG}\right]^\top := \hat{\boldsymbol{V}}_{(j)}^{-1/2}\left(\hat{\boldsymbol{b}}_{(j)} - \boldsymbol{\beta}_{(j)}\right)$ and $\boldsymbol{S}_j^0 := \hat{\boldsymbol{V}}_{(j)}^{-1/2}\hat{\boldsymbol{b}}_{(j)}$. For a given significance level $0 < \alpha < 1$, we reject the null hypothesis H_0 in (4) when $\|\boldsymbol{S}_j^0\|_2^2 > \chi_\alpha^2(G)$, where $\chi_\alpha^2(G)$ represents the upper α-quantile of a central χ^2 distribution with G degrees of freedom. The notion of weak convergence is not well defined in our setting, since we consider the case where G diverges with n. Theorem 4 shows that our test for the hypothesis (4) is still valid in the χ^2 approximation.

Theorem 4 *Suppose conditions of Theorem 3 hold, and* $G^{7/2} = o(\min_{1\le g\le G} n_g)$. *Then, we have*

$$\sup_x |P\left(\|\boldsymbol{S}_j\|_2^2 \le x\right) - P(\chi^2(G) \le x)| \to 0.$$

The following Corollary implies $\hat{\boldsymbol{b}}_{(j)}^\top \hat{V}_{(j)}^{-1}\hat{\boldsymbol{b}}_{(j)} \xrightarrow{d} \chi_G^2$ under the null hypothesis $H_0 : \beta_j^{(1)} = \cdots = \beta_j^{(G)} = 0$ in (4).

Corollary 1 *Assume Conditions of Theorem 4. Then, under the null hypothesis H_0 in (4), for any significance level* $0 < \alpha < 1$, *we have*

$$\lim_{n\to\infty} P\left(\|\boldsymbol{S}_j^0\|_2^2 > \chi_\alpha^2(G)\right) = \alpha.$$

Based on this result, the corresponding chi-square test statistic for the hypothesis (4) is $\|S_j^0\|_2^2$, i.e., reject H_0 in (4) if $\|S_j^0\|_2^2 > \chi_\alpha^2(G)$.

Next, we consider testing the homogeneity hypothesis in (3), which can be rewritten as

$$H_0 : \boldsymbol{D}\boldsymbol{\beta}_{(j)} = 0 \quad \text{vs} \quad H_1 : \text{ not } H_0,$$

where $\boldsymbol{D}$ represents the $(G-1) \times G$ matrix such that $D_{\ell\ell} = 1$ and $D_{\ell,\ell+1} = -1$ for $\ell = 1, \ldots, G-1$. Define

$$\boldsymbol{K}_j := \left(\boldsymbol{D}\hat{\boldsymbol{V}}_{(j)}\boldsymbol{D}^\top\right)^{-1/2}\boldsymbol{D}(\hat{\boldsymbol{b}}_{(j)} - \boldsymbol{\beta}_{(j)})$$

$$\boldsymbol{K}_j^0 := \left(\boldsymbol{D}\hat{\boldsymbol{V}}_{(j)}\boldsymbol{D}^\top\right)^{-1/2}\boldsymbol{D}\hat{\boldsymbol{b}}_{(j)}.$$

Theorem 5 shows that the test procedure for the hypothesis (3) based on $\|\boldsymbol{K}_j\|_2^2$ admits a χ^2 approximation.

Theorem 5 *Assume Conditions of Theorem 4. Then, we have*

$$\sup_x \left|P\left(\|\boldsymbol{K}_j\|_2^2 \le x\right) - P\left(\chi^2(G-1) \le x\right)\right| \to 0.$$

Corollary 2 *Assume the conditions of Theorem 4. Then, under the null hypothesis H_0 in (3), for any significance level* $0 < \alpha < 1$, *we have*

$$\lim_{n\to\infty} P\left(\|\boldsymbol{K}_j^0\|_2^2 > \chi_\alpha^2(G-1)\right) = \alpha.$$

Corollary 2 implies that under $H_0 : \beta_j^{(1)} = \cdots = \beta_j^{(G)}$ in (3), it holds that

$$\hat{\boldsymbol{b}}_{(j)}^\top \boldsymbol{D}^\top(\boldsymbol{D}\hat{\boldsymbol{V}}_{(j)}\boldsymbol{D}^\top)^{-1}\boldsymbol{D}\hat{\boldsymbol{b}}_{(j)} \xrightarrow{d} \chi_{G-1}^2.$$

Based on this result, the corresponding chi-square test statistic for the hypothesis (3) is $\|\boldsymbol{K}_j^0\|_2^2$, i.e., reject H_0 in (3) if $\|\boldsymbol{K}_j^0\|_2^2 > \chi_\alpha^2(G-1)$.

Implementation

In this section we present the computational algorithm for solving (6). The algorithm for (5) can be obtained in a similar way. We use the proximal gradient method to solve (6). Let $\tilde{\boldsymbol{B}}^{(t)}$ be the tth update in the proximal gradient method. Then the $(t+1)$th update $\tilde{\boldsymbol{B}}^{(t+1)}$ is given by

$$\mathop{\arg\min}_{\boldsymbol{\Delta}\in\mathbb{R}^{p\times G}} \frac{1}{n} L_n\left(\tilde{\boldsymbol{B}}^{(t)}\right) + \left\langle \nabla L_n(\tilde{\boldsymbol{B}}^{(t)}), \boldsymbol{\Delta} - \tilde{\boldsymbol{B}}^{(t)} \right\rangle + \frac{\eta}{2}\|\boldsymbol{\Delta} - \tilde{\boldsymbol{B}}^{(t)}\|_F^2 + \tilde{\lambda}_1 \sum_{j=1}^{p} \|\boldsymbol{\Delta}_{j,\cdot}\|_2 + \tilde{\lambda}_2 \sum_{j=1}^{p} \sum_{g<g'} |\Delta_{jg} - \Delta_{jg'}|. \tag{8}$$

```
1. Initialize: $\tilde{\boldsymbol{B}}^{(0)} \leftarrow \mathbf{0}_{p\times G}$
2. For $t = 0, \ldots, T-1$, compute $\tilde{\boldsymbol{B}}^{(t+1)}$ via (3.1)-(3.3).
(3.1) $\boldsymbol{A}^{(0)} \leftarrow \mathbf{0}_{p\times G}$,   $\boldsymbol{U}^{(0)} \leftarrow \mathbf{0}_{p\times G}$,   $\boldsymbol{F}^{(0)} \leftarrow \mathbf{0}_{p\times \frac{G(G-1)}{2}}$,   $\boldsymbol{W}^{(0)} \leftarrow \mathbf{0}_{p\times \frac{G(G-1)}{2}}$
(3.2) For $s = 1 \ldots, S$, perform (3.2.1) - (3.2.5).
  (3.2.1) $\tilde{\boldsymbol{B}}^{(t+1,s)} \leftarrow \left(-\nabla L_n(\tilde{\boldsymbol{B}}^{(t)}) + \eta\tilde{\boldsymbol{B}}^{(t)} + \rho\boldsymbol{A}^{(s-1)} + (\rho\boldsymbol{F}^{(s-1)} - \boldsymbol{W}^{(s-1)})\boldsymbol{H}^\top - \boldsymbol{U}^{(s-1)}\right)\boldsymbol{R}$,
          $\boldsymbol{R} = \left(\rho\boldsymbol{H}\boldsymbol{H}^\top + (\eta+\rho)\,\boldsymbol{I}_G\right)^{-1}$ .
  (3.2.2) $\boldsymbol{A}_{j,\cdot}^{(s)} \leftarrow \max\left(0, 1 - \frac{\lambda_1/\rho}{\|\tilde{\boldsymbol{B}}_{j,\cdot}^{(t+1,s)} + \boldsymbol{U}_{j,\cdot}^{(s-1)}/\rho\|_2}\right) \cdot \left(\tilde{\boldsymbol{B}}_{j,\cdot}^{(t+1,s)} + \boldsymbol{U}_{j,\cdot}^{(s-1)}/\rho\right)$,   $j = 1, \ldots, p$.
  (3.2.3) For $1 \leqslant i \leqslant p$, $1 \leqslant j \leqslant G(G-1)/2$,
          $$\begin{aligned} \breve{F}_{ij}^{(s)} &\leftarrow [\tilde{\boldsymbol{B}}^{(t+1,s)}\boldsymbol{H} + \boldsymbol{W}^{(s-1)}/\rho]_{ij}, \\ F_{ij}^{(s)} &\leftarrow \max\left(0, \left|\breve{F}_{ij}^{(s)}\right| - \lambda_2/\rho\right) \cdot \operatorname{sgn}\left(\breve{F}_{ij}^{(s)}\right), \end{aligned} \quad (12)$$
          where $\operatorname{sgn}(x)$ represents the sign of a number $x$.
  (3.2.4) $\boldsymbol{W}^{(s)} \leftarrow \boldsymbol{W}^{(s-1)} + \rho(\tilde{\boldsymbol{B}}^{(t+1,s)}\boldsymbol{H} - \boldsymbol{F}^{(s)})$
  (3.2.5) $\boldsymbol{U}^{(s)} \leftarrow \boldsymbol{U}^{(s-1)} + \rho(\tilde{\boldsymbol{B}}^{(t+1,s)} - \boldsymbol{A}^{(s)})$
(3.3) $\tilde{\boldsymbol{B}}^{(t+1)} \leftarrow \tilde{\boldsymbol{B}}^{(t+1,S)}$
```

Algorithm 1. ADMM algorithm for the fused group Lasso logistic regression in (6)

We set $\eta = \sum_{g=1}^G \|\mathbf{X}^{(g)}\|_F^2/4n$ based on the convergence properties of the proximal gradient method[37]. Note that $\nabla L_n(\boldsymbol{\Delta})$ is Lipschitz continuous with Lipschitz parameter $\sum_{g=1}^G \|\mathbf{X}^{(g)}\|_F^2/(4n)$. Let $L_{n,\boldsymbol{A}}(\boldsymbol{\Delta})$ be the first-order approximation of $L_n(\boldsymbol{\Delta})$ at a matrix $\boldsymbol{A}$. Then, (8) can be rewritten as

$$\arg\min_{\boldsymbol{\Delta}} L_{n,\tilde{\boldsymbol{B}}^{(t)}}(\boldsymbol{\Delta}) + \frac{\eta}{2}\|\boldsymbol{\Delta} - \tilde{\boldsymbol{B}}^{(t)}\|_F^2 + \tilde{\lambda}_1\sum_{j=1}^p \|\boldsymbol{\Delta}_{j,\cdot}\|_2 + \tilde{\lambda}_2\sum_{j=1}^p\sum_{g<g'} |\Delta_{jg} - \Delta_{jg'}|. \tag{9}$$

To compute (9), we use the alternating direction method of multipliers (ADMM)[38]. Let $\boldsymbol{H}$ be the G by $G(G-1)/2$ matrix satisfying

$$\sum_{j=1}^p\sum_{g<g'} |\Delta_{jg} - \Delta_{jg'}| = \|\boldsymbol{\Delta}\boldsymbol{H}\|_1.$$

By introducing surrogate variables $\boldsymbol{A}$ and $\boldsymbol{F}$, (9) can be converted to solving the following optimization problem:

$$\begin{aligned} \min_{\boldsymbol{\Delta},\boldsymbol{A},\boldsymbol{F}} \; & L_{n,\tilde{\boldsymbol{B}}^{(t)}}(\boldsymbol{\Delta}) + \frac{\eta}{2}\|\boldsymbol{\Delta} - \tilde{\boldsymbol{B}}^{(t)}\|_F^2 + \tilde{\lambda}_1\sum_{j=1}^p \|\boldsymbol{A}_{j,\cdot}\|_2 + \tilde{\lambda}_2\|\boldsymbol{F}\|_1 \\ & \text{subject to } \boldsymbol{A} = \boldsymbol{\Delta} \text{ and } \boldsymbol{\Delta}\boldsymbol{H} = \boldsymbol{F}. \end{aligned} \tag{10}$$

The corresponding augmented Lagrangian is

$$\begin{aligned} K^{(t)}(\boldsymbol{\Delta}, \boldsymbol{A}, \boldsymbol{F}, \boldsymbol{W}, \boldsymbol{U}) :=& L_{n,\tilde{\boldsymbol{B}}^{(t)}}(\boldsymbol{\Delta}) + \frac{\eta}{2}\|\boldsymbol{\Delta} - \tilde{\boldsymbol{B}}^{(t)}\|_F^2 + \tilde{\lambda}_1\sum_{j=1}^p \left\|\boldsymbol{A}_{j,\cdot}\right\|_2 + \tilde{\lambda}_2\|\boldsymbol{F}\|_1 \\ & + \langle \boldsymbol{U}, \boldsymbol{\Delta} - \boldsymbol{A}\rangle + \langle \boldsymbol{W}, \boldsymbol{\Delta}\boldsymbol{H} - \boldsymbol{F}\rangle + \frac{\rho}{2}\|\boldsymbol{\Delta} - \boldsymbol{A}\|_F^2 + \frac{\rho}{2}\|\boldsymbol{\Delta}\boldsymbol{H} - \boldsymbol{F}\|_F^2, \end{aligned}$$

where $\boldsymbol{U}$ and $\boldsymbol{W}$ represent dual variables and $\rho > 0$ is a fixed parameter. Let $\tilde{\boldsymbol{B}}^{(t+1,s)}$ be the sth update in the ADMM to compute $\tilde{\boldsymbol{B}}^{(t+1)}$. Then, $\tilde{\boldsymbol{B}}^{(t+1)}$ for $t = 0, 1, 2, \ldots$ is obtained by iterating the following updates: starting with $\boldsymbol{A}^{(0)} = \boldsymbol{U}^{(0)} = \mathbf{0}_{p\times G}$ and $\boldsymbol{F}^{(0)} = \boldsymbol{W}^{(0)} = \mathbf{0}_{p\times\frac{G(G-1)}{2}}$, we repeat for $s = 1, \ldots, S$,

$$\begin{aligned}
\tilde{\boldsymbol{B}}^{(t+1,s)} &= \arg\min_{\boldsymbol{\Delta}} K^{(t)}\left(\boldsymbol{\Delta}, \boldsymbol{A}^{(s-1)}, \boldsymbol{F}^{(s-1)}, \boldsymbol{W}^{(s-1)}, \boldsymbol{U}^{(s-1)}\right)\\
\boldsymbol{A}^{(s)} &= \arg\min_{\boldsymbol{A}} K^{(t)}\left(\tilde{\boldsymbol{B}}^{(t+1,s)}, \boldsymbol{A}, \boldsymbol{F}^{(s-1)}, \boldsymbol{W}^{(s-1)}, \boldsymbol{U}^{(s-1)}\right)\\
\boldsymbol{F}^{(s)} &= \arg\min_{\boldsymbol{F}} K^{(t)}\left(\tilde{\boldsymbol{B}}^{(t+1,s)}, \boldsymbol{A}^{(s)}, \boldsymbol{F}, \boldsymbol{W}^{(s-1)}, \boldsymbol{U}^{(s-1)}\right) \qquad (11)\\
\boldsymbol{W}^{(s)} &= \boldsymbol{W}^{(s-1)} + \rho\left(\tilde{\boldsymbol{B}}^{(t+1,s)}\boldsymbol{H} - \boldsymbol{F}^{(s)}\right),\\
\boldsymbol{U}^{(s)} &= \boldsymbol{U}^{(s-1)} + \rho\left(\tilde{\boldsymbol{B}}^{(t+1,s)} - \boldsymbol{A}^{(s)}\right),
\end{aligned}$$

and $\tilde{\boldsymbol{B}}^{(t+1)}$ is updated as $\tilde{\boldsymbol{B}}^{(t+1)} = \tilde{\boldsymbol{B}}^{(t+1,S)}$, where the derivations of each update in (11) can be found in section "ADMM update". We set $S = 50$, $\rho = 1$, and the maximum iteration number $T = 200$ by analysis. The proposed ADMM algorithm is summarized in Algorithm 1. In our simulation and real data examples, it was observed that the algorithm achieves fast convergence. On average, it completes in less than one second, implemented on an Intel Xeon (2.20 GHz). See also section "Computational time comparison".

ADMM update

In this section, we include details of the ADMM update presented at (11).

Update for $\tilde{\boldsymbol{B}}^{(t+1,s)}$: For simplicity, let

$$K^{(t,s-1)}(\boldsymbol{\Delta}) := K^{(t)}\left(\boldsymbol{\Delta}, \boldsymbol{A}^{(s-1)}, \boldsymbol{F}^{(s-1)}, \boldsymbol{W}^{(s-1)}, \boldsymbol{U}^{(s-1)}\right).$$

By the convexity, $\tilde{\boldsymbol{B}}^{(t+1,s)}$ must satisfy

$$\frac{\partial K^{(t,s-1)}(\boldsymbol{\Delta})}{\partial \boldsymbol{\Delta}} \,|_{\boldsymbol{\Delta}=\tilde{\boldsymbol{B}}^{(t+1,s)}} = \boldsymbol{0}_{p\times G}.$$

Thus, it holds that

$$\tilde{\boldsymbol{B}}^{(t+1,s)}\left(\rho\boldsymbol{H}\boldsymbol{H}^\top + (\eta+\rho)\boldsymbol{I}_G\right) = \boldsymbol{V},$$

where $\boldsymbol{V} := -\nabla L_n\left(\tilde{\boldsymbol{B}}^{(t)}\right) + \eta\tilde{\boldsymbol{B}}^{(t)} + \rho\boldsymbol{A}^{(s-1)} + \rho\boldsymbol{F}^{(s-1)}\boldsymbol{H}^\top - \boldsymbol{U}^{(s-1)} - \boldsymbol{W}^{(s-1)}\boldsymbol{H}^\top$.

Update for $\boldsymbol{A}^{(s)}$: $\boldsymbol{A}^{(s)}$ is defined by

$$\boldsymbol{A}^{(s)} = \arg\min_{\boldsymbol{A}} \; -\langle \boldsymbol{U}^{(s-1)}, \boldsymbol{A}\rangle + \frac{\rho}{2}\|\tilde{\boldsymbol{B}}^{(t+1,s)} - \boldsymbol{A}\|_F^2 + \tilde{\lambda}_1\sum_{j=1}^{p}\|\boldsymbol{A}_{j,\cdot}\|_2,$$

which is separable with respect to j's. For each $1 \le j \le p$, let $\boldsymbol{s}_j$ be the subgradient of $\|\boldsymbol{x}\|_2$ at $\boldsymbol{x} = \boldsymbol{A}_{j.}^{(s)}$, i.e.,

$$\boldsymbol{s}_j = \begin{cases} \text{some } \boldsymbol{x} \in \mathbb{R}^{1\times G} \text{ with } \|\boldsymbol{x}\|_2 \le 1 & \text{if } \boldsymbol{A}_{j.}^{(s)} = \boldsymbol{0}_G^\top \\ \boldsymbol{A}_{j,\cdot}^{(s)}/\|\boldsymbol{A}_{j,\cdot}^{(s)}\|_2 & \text{otherwise.} \end{cases}$$

By the convexity, it holds that for $1 \le j \le p$,

$$-\rho\left(\tilde{\boldsymbol{B}}_{j,\cdot}^{(t+1,s)} - \boldsymbol{A}_{j,\cdot}^{(s)}\right) - \boldsymbol{U}_{j,\cdot}^{(s-1)} + \tilde{\lambda}_1\boldsymbol{s}_j = \boldsymbol{0}_G^\top.$$

By the definition of $\boldsymbol{s}_j$, we obtain that for $j = 1, \ldots, p$,

$$\boldsymbol{A}_{j,\cdot}^{(s)} = \max\left(0, 1 - \frac{\tilde{\lambda}_1/\rho}{\|\tilde{\boldsymbol{B}}_{j,\cdot}^{(t+1,s)} + \boldsymbol{U}_{j,\cdot}^{(s-1)}/\rho\|_2}\right)\cdot\left(\tilde{\boldsymbol{B}}_{j,\cdot}^{(t+1,s)} + \frac{\boldsymbol{U}_{j,\cdot}^{(s-1)}}{\rho}\right).$$

Update for $\boldsymbol{F}^{(s)}$: $\boldsymbol{F}^{(s)}$ can be derived using the definition of subgradient of ℓ_1 norm. For each $(i,j) \in \{1,\ldots,p\} \times \left\{1,\ldots,\frac{G(G-1)}{2}\right\}$, it must hold that

$$-W_{ij}^{(s-1)} - \rho\left[\tilde{\boldsymbol{B}}^{(t+1,s)}\boldsymbol{H} - \boldsymbol{F}^{(s)}\right]_{ij} + \tilde{\lambda}_2\zeta_{ij} = 0,$$

where ζ_{ij} is defined by

$$\zeta_{ij} = \begin{cases} 1 & \text{if } \; F_{ij}^{(s)} > 0 \\ -1 & \text{if } \; F_{ij}^{(s)} < 0 \\ \text{some } x \text{ with } |x| \le 1 & \text{if } \; F_{ij}^{(s)} = 0. \end{cases}$$

By the definition, $\boldsymbol{F}^{(s)} = (F_{ij}^{(s)})_{1\le i\le p,\, 1\le j\le \frac{G(G-1)}{2}}$ is given by

$$F_{ij}^{(s)} = \begin{cases} \left[\tilde{\boldsymbol{B}}^{(t+1,s)}\boldsymbol{H}\right]_{ij} + \frac{W_{ij}^{(s-1)}}{\rho} - \frac{\tilde{\lambda}_2}{\rho} & \text{if } \breve{F}_{ij}^{(s)} > \frac{\tilde{\lambda}_2}{\rho} \\ [\tilde{\boldsymbol{B}}^{(t+1,s)}\boldsymbol{H}]_{ij} + \frac{W_{ij}^{(s-1)}}{\rho} + \frac{\tilde{\lambda}_2}{\rho} & \text{if } \breve{F}_{ij}^{(s)} < \frac{-\tilde{\lambda}_2}{\rho} \\ 0 & \text{otherwise} \end{cases}$$

where $\breve{F}_{ij}^{(s)} = \left[\tilde{\boldsymbol{B}}^{(t+1,s)}\boldsymbol{H}\right]_{ij} + \frac{W_{ij}^{(s-1)}}{\rho}$.

Simulation study

In this section, we present empirical results through simulation analysis. A main objective of the simulation analysis is to investigate the finite sample performance of the proposed method (Debiased Fused Group Lasso; DFGL) in testing the following hypotheses:

$$H_0 : \beta_j^{(1)} = \cdots = \beta_j^{(G)} = 0 \text{ vs } H_1 : \text{ not } H_0, \tag{13}$$

$$H_0 : \beta_j^{(1)} = \cdots = \beta_j^{(G)} \text{ vs } H_1 : \text{ not } H_0. \tag{14}$$

To investigate the advantages of the proposed tests over existing ℓ_1 penalized approaches that do not use a fusion penalty, we compared the proposed method with the following methods:

- DL (Debiased Lasso) : Chi-squared test based on applying debiased Lasso[17] separately for each subpopulation
- DL-B: Bonferroni correction using p-values obtained by applying debiased Lasso[17] separately for each subpopulation
- DL-E (Debiased Lasso based on the Exact inverse of the information matrix): Chi-squared test based on applying debiased Lasso[36] separately for each subpopulation, where this bias-correction method is developed for the scenario where the sample size is greater than the number of regressors
- DL-E-B: Bonferroni correction using p-values obtained by applying debiased Lasso[36] separately for each subpopulation

We also compared the proposed method with DR-B, Bonferroni correction based on debiased Ridge[39]. R code (https://github.com/luxia-bios/DebiasedLassoGLMs) was used to implement DL-E, and R package hdi[40] was used to implement DL and DR-B. We also used R code (https://web.stanford.edu/~montanar/ssLasso/) to solve the quadratic programming (7). Following Javanmard and Montanari[19], we set μ_g in (7) as $\mu_g = c\sqrt{\log p/n_g}$ for some positive constant c. Specifically, we set $c = 0.7$ based on the results of the sensitivity analysis of c summarized in the Supplementary material. 5-fold cross-validation was used to determine the regularization parameters for the proposed penalization method.

Simulation setting

In our simulation study, we fix $G = 7$ and simulate $\mathbf{x}_i^{(g)}$ for all (g, i) from the p-dimensional multivariate normal distribution with mean $\mathbf{0}_p$ and covariance matrix $\boldsymbol{\Sigma}_x$. Specifically, we consider the following two different covariance matrices: (1) AR(1): $[\boldsymbol{\Sigma}_x]_{ij} = 0.5^{|i-j|}$; or (2) Block: $\boldsymbol{\Sigma}_x$ is a $p \times p$ block diagonal matrix consisting of $p/4$ identical blocks, where the 4×4 sub-block matrix, denoted by $\boldsymbol{\Sigma}_b$, is a Toeplitz matrix such that $[\boldsymbol{\Sigma}_b]_{ij} = 0.5^{|i-j|}, i = 1, \ldots, 4;\ j = 1, \ldots, 4$. Then, the response variables are generated independently as follows: for $1 \le g \le G,\ 1 \le i \le n_g$,

$$y_i^{(g)} \sim \text{Bernoulli}\left(\frac{\exp([\mathbf{x}_i^{(g)}]^\top \boldsymbol{\beta}^{(g)})}{1 + \exp([\mathbf{x}_i^{(g)}]^\top \boldsymbol{\beta}^{(g)})}\right),$$

where $\boldsymbol{\beta}_{(j)} = \left(\beta_j^{(1)}, \ldots, \beta_j^{(G)}\right)^\top \in \mathbb{R}^G$ is set as

$$\boldsymbol{\beta}_{(j)} := \begin{cases} (-0.6, -0.6, 0.6, 0.6, 0.6, 0.6, 0.6) & j = 1 \\ (0.6, 0.6, -0.6, -0.6, 0.6, 0.6, 0.6) & j = 2 \\ (0.6, 0.6, 0.6, 0.6, -0.4, -0.4, 0.6) & j = 3 \\ (-0.4, 0.6, 0.6, 0.6, 0.6, 0.6, -0.4) & j = 4 \\ (0.4, 0.4, 0.4, 0.4, 0.4, 0.4, 0.4) & j = 5 \\ (0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0) & j = 6, 7 \\ -(0.4, 0.4, 0.4, 0.4, 0.4, 0.4, 0.4) & j = 8 \\ (1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5) & j = 9, 10 \\ (0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0) & j = 11 \\ (2.5, 2.5, 2.5, 2.5, 2.5, 2.5, 2.5) & j = 12 \\ (0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0) & \mathit{otherwise} \end{cases}.$$

We set $n_1 = \cdots = n_G = m$ and consider the following two specifications of (m, p): $(m, p) = (200, 80)$ or $(m, p) = (300, 120)$. For each case, we simulate $M = 100$ Monte Carlo samples and summarize the results over 100 replications.

				Methods					
(m, p)	**j**	**$\min_g \beta_j^{(g)}$**	**$\max_g \beta_j^{(g)}$**	**DFGL**	**DL**	**DL-E**	**DR-B**	**DL-B**	**DL-E-B**
AR(1)									
(200, 80)	1	−0.6	0.6	0.93	0.84	0.54	0.00	0.00	0.00
	2	− 0.6	0.6	0.87	0.76	0.33	0.00	0.00	0.00
	3	− 0.4	0.6	0.62	0.44	0.21	0.00	0.00	0.00
	4	− 0.4	0.6	0.76	0.54	0.22	0.00	0.00	0.00
(300, 120)	1	− 0.6	0.6	1.00	0.97	0.78	0.00	0.00	0.00
	2	− 0.6	0.6	0.97	0.93	0.61	0.00	0.00	0.00
	3	− 0.4	0.6	0.90	0.75	0.44	0.00	0.00	0.00
	4	− 0.4	0.6	0.88	0.87	0.48	0.00	0.00	0.00
Block									
(200, 80)	1	− 0.6	0.6	0.95	0.84	0.57	0.00	0.02	0.01
	2	− 0.6	0.6	0.82	0.60	0.38	0.00	0.00	0.00
	3	− 0.4	0.6	0.66	0.44	0.20	0.00	0.00	0.00
	4	− 0.4	0.6	0.79	0.63	0.31	0.00	0.00	0.00
(300, 120)	1	− 0.6	0.6	1.00	0.97	0.76	0.00	0.00	0.00
	2	− 0.6	0.6	0.99	0.91	0.61	0.00	0.00	0.00
	3	− 0.4	0.6	0.88	0.78	0.40	0.00	0.00	0.00
	4	− 0.4	0.6	0.97	0.91	0.57	0.00	0.01	0.00

Table 1. Power for testing $H_0 : \beta_j^{(1)} = \cdots = \beta_j^{(G)}$ vs H_1 : not H_0 at $\alpha = 0.05$, where $n_1, \ldots, n_G$ are set as $n_1 = \cdots = n_G = m$.

Simulation results

First, we present simulation results when testing for homogeneity (14). Next, we present simulation results when testing for overall significance (13).

Testing homogeneity

We consider $j = 1, 2, 3, 4$ and $j = 5, 8, 9, 10, 12$ to measure powers and type I errors, respectively. Tables 1 and 2 record the powers and type I errors of different methods when the significance level α is set to $\alpha = 0.05$. As shown in Table 1, DFGL outperforms the other approaches in terms of higher power. DL ranks second in terms of higher power in most cases, but fails to control Type I errors when the covariates of interest have strong signals. Competing approaches based on a debiased Lasso using the exact inverse of the information matrix also produce type I errors higher than the nominal level when the covariate with the strongest signal is considered. In contrast to the despecified Lasso-based chi-squared test procedures, our proposed DFGL and other Bonferroni-corrected test procedures, including the Ridge-based DR-B, yield type I errors less than or close to the nominal level $\alpha = 0.05$, but these Bonferroni-corrected tests are conservative, as shown in Table 1. Note that the Ridge-based approaches are known to be conservative in various contexts[19,41].

Testing significance

In this subsection, we examine the performance of the proposed significance test. We consider $j = 1, 2, 3, 4, 5, 8, 9, 10, 12$ to measure power, while we consider $j = 6, 7, 11, 15, 20$ to measure type I error. Tables S5 and S6 in the Supplementary materials report the performance of each method in terms of power and Type I error, respectively, when $\alpha = 0.05$. The proposed method generally has higher power compared to the competing approaches. In particular, the proposed method is superior to the competing approaches in terms of higher power when the covariates of interest have relatively weak signals. This result was also observed in a previous study[41], which investigated debiased group Lasso for linear regression. When sample sizes are set as $n_1 = \ldots = n_G = 300$, DL and the proposed DFGL sometimes provide similar power. However, DL fails to control Type I errors when the covariate of interest is correlated with covariates with strong signals. The proposed DFGL and the other methods, except DL, have type I errors less than or close to the significance level in all cases considered, but the other methods are conservative.

Multiple testing

We evaluate the empirical performance of the proposed testing procedures in the context of multiple testing. We consider the following two multiple testing problems, (15) and (16), respectively:

$$H_{0,j} : \beta_j^{(1)} = \ldots = \beta_j^{(G)} \text{ vs } H_1 : \text{not } H_{0,j}, \quad j = 1, \ldots, p \tag{15}$$

$$H_{0,j} : \beta_j^{(1)} = \cdots = \beta_j^{(G)} = 0 \text{ vs } H_1 : \text{not } H_{0,j}, \quad j = 1, \ldots, p. \tag{16}$$

				Methods					
(m, p)	**j**	**$\min_g \beta_j^{(g)}$**	**$\max_g \beta_j^{(g)}$**	**DFGL**	**DL**	**DL-E**	**DR-B**	**DL-B**	**DL-E-B**
AR(1)									
(200, 80)	5	0.4	0.4	0.01	0.02	0.00	0.00	0.00	0.00
	8	− 0.4	− 0.4	0.01	0.03	0.01	0.00	0.00	0.00
	9	1.5	1.5	0.02	0.21	0.05	0.00	0.00	0.00
	10	1.5	1.5	0.05	0.16	0.04	0.00	0.00	0.00
	12	2.5	2.5	0.04	0.35	0.12	0.00	0.00	0.00
(300, 120)	5	0.4	0.4	0.04	0.02	0.02	0.00	0.00	0.00
	8	− 0.4	− 0.4	0.02	0.02	0.00	0.00	0.00	0.00
	9	1.5	1.5	0.03	0.11	0.05	0.00	0.00	0.00
	10	1.5	1.5	0.02	0.09	0.08	0.00	0.00	0.00
	12	2.5	2.5	0.00	0.22	0.14	0.00	0.00	0.00
Block									
(200, 80)	5	0.4	0.4	0.04	0.03	0.02	0.00	0.00	0.00
	8	− 0.4	− 0.4	0.02	0.01	0.00	0.00	0.00	0.00
	9	1.5	1.5	0.05	0.28	0.15	0.00	0.01	0.00
	10	1.5	1.5	0.04	0.15	0.04	0.00	0.00	0.00
	12	2.5	2.5	0.05	0.36	0.21	0.00	0.01	0.01
(300, 120)	5	0.4	0.4	0.01	0.00	0.00	0.00	0.00	0.00
	8	− 0.4	− 0.4	0.03	0.02	0.00	0.00	0.00	0.00
	9	1.5	1.5	0.02	0.26	0.14	0.00	0.00	0.00
	10	1.5	1.5	0.02	0.15	0.07	0.00	0.00	0.00
	12	2.5	2.5	0.03	0.33	0.19	0.00	0.01	0.00

Table 2. Type I error for testing $H_0 : \beta_j^{(1)} = \cdots = \beta_j^{(G)}$ vs H_1 : not H_0 at $\alpha = 0.05$, where $n_1, \ldots, n_G$ are set as $n_1 = \cdots = n_G = m$.

To control the familywise error rate (FWER), we adjust p-values using the Bonferroni-Holm (BH) procedure[42]. We also apply the BH procedure to p-values from DL and those from DL-E. We don't consider methods using Bonefrroni-correction, i.e., DL-B, DL-E-B, and DR-B, for multiple testing. This is because they are too conservative, as observed in sections "Testing homogeneity" and "Testing significance".

When considering multiple testing for significance (16), we measure FWER and power as follows:

- FWER: The percentage of the cases where $H_{0,j}$ is rejected for some $j \in S^c$,
- Power: Average of the empirical power $\sum_{j \in S} I\left(H_{0,j} \text{ is rejected }\right)/s$,

where $S = \{j : H_{0,j} \text{ is false }\}$ with cardinality $s = |S|$. Power and FWER are measured in the same way when considering homogeneity tests. Table 3 summarizes the results. The proposed DFGL has the highest power in all cases, while providing FWER below the nominal level $\alpha = 0.05$. Among the competing methods, DL has incorrect control of FWER when considering the homogeneity test (15) and DL-E has poor power, especially when considering the homogeneity test (15).

				Power			FWER		
Testing	**Covariate**	**m**	**p**	**DFGL**	**DL**	**DL-E**	**DFGL**	**DL**	**DL-E**
Homogeneity	AR(1)	200	80	0.408 (0.253)	0.170 (0.197)	0.015 (0.069)	0.000	0.190	0.000
		300	120	0.618 (0.229)	0.358 (0.208)	0.120 (0.172)	0.000	0.060	0.020
	Block	200	80	0.395 (0.291)	0.178 (0.231)	0.045 (0.097)	0.000	0.140	0.000
		300	120	0.690 (0.236)	0.380 (0.245)	0.085 (0.139)	0.010	0.140	0.000
Significance	AR(1)	200	80	0.649 (0.131)	0.518 (0.104)	0.352 (0.048)	0.000	0.030	0.000
		300	120	0.791 (0.092)	0.684 (0.090)	0.428 (0.088)	0.000	0.010	0.000
	Block	200	80	0.690 (0.141)	0.501 (0.110)	0.377 (0.067)	0.010	0.030	0.000
		300	120	0.841 (0.110)	0.677 (0.116)	0.421 (0.085)	0.010	0.020	0.000

Table 3. Performances of multiple testing at $\alpha = 0.05$, where $n_1, \ldots, n_G$ are set as $n_1 = \cdots = n_G = m$.

			Methods				
j	$\min_g \beta_j^{(g)}$	$\max_g \beta_j^{(g)}$	DFGL	DL	DPL	DR-B	DL-B
167	– 1.0	1.0	0.56	0.32	0.33	0.07	0.30
187	– 1.0	1.0	0.77	0.53	0.54	0.11	0.54
211	– 0.8	0.8	0.26	0.18	0.13	0.00	0.18
270	– 0.8	0.8	0.42	0.24	0.28	0.02	0.22

Table 4. Power for testing $H_0 : \beta_j^{(1)} = \cdots = \beta_j^{(G)}$ vs H_1 : not H_0 at $\alpha = 0.05$.

Simulation study for an imbalanced design

In this section, we perform a simulation analysis to assess the performance of DFGL when the sample sizes for the groups vary. Specifically, we consider the following model parameters: $G = 3$, $p = 300$, $n_1 = 90$, $n_2 = 70$, and $n_3 = 40$, reflecting the dimension of a subset of the CCLE data analyzed in section "Application to the CCLE data". For each group, we simulate the covariates $\mathbf{x}_i^{(g)}$ from p-dimensional multivariate normal distribution with mean $\mathbf{0}_p$ and covariance matrix $\mathbf{\Sigma}_x$ where $[\mathbf{\Sigma}_x]_{ij} = 0.5^{|i-j|}$. We set $s = |S| = 6$ where $S = \{j : \|\boldsymbol{\beta}_{(j)}\|_2 > 0\}$ and randomly draw elements of S from $\{1, \ldots, p\}$. As a result, we obtained $S = \{85, 129, 167, 187, 211, 270\}$, and we set $\boldsymbol{\beta}_{(j)} = (\beta_j^{(1)}, \cdots, \beta_j^{(G)})^\top \in \mathbb{R}^G$ as follows:

$$\boldsymbol{\beta}_{(j)} := \begin{cases} (1.5, 1.5, 1.5) & j = 85 \\ (2.0, 2.0, 2.0) & j = 129 \\ (1.0, 1.0, -1.0) & j = 167 \\ (-1.0, 1.0, 1.0) & j = 187 \\ (0.8, 0.8, -0.8) & j = 211 \\ (0.8, -0.8, 0.8) & j = 270 \\ (0.0, 0.0, 0.0) & otherwise \end{cases} .$$

Due to the relatively small sample sizes ($n_1 = 90$, $n_2 = 70$, and $n_3 = 40$), we consider DPL (Debiased Pooled-Lasso) as an additional competing approach. DPL refers to a chi-squared test based on applying bias correction[17] to Pooled-Lasso, which adapts to Lasso to analyze samples from different groups together, i.e., $\{\mathbf{y}, \mathbf{X}\}$. Here, $\mathbf{y} = \left[(\mathbf{y}^{(1)})^\top, \ldots, (\mathbf{y}^{(G)})^\top\right]^\top$, and $\mathbf{X} = \text{ diag } \left(\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(G)}\right)$ represents a block diagonal matrix consisting of $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(G)}$. Note that DL-E can't be used in this simulation analysis because $p > n_g$ for $g = 1, 2, 3$.

Testing homogeneity

We consider $j = 167, 187, 211, 270$ and $j = 1, 2, 3, 85, 129$ to measure powers and type I errors, respectively. Tables 4 and 5 show the powers and type I errors of different methods when $\alpha = 0.05$. DFGL outperforms the other methods in terms of higher power. In addition, DFGL successfully controls Type I errors; however, Lasso-based methods, including DL, DPL, and DL-B, produce Type I errors higher than the significance level when testing the null hypothesis $H_0 : \beta_{129}^{(1)} = \beta_{129}^{(2)} = \beta_{129}^{(3)}$ where $\beta_{129}^{(1)} = \beta_{129}^{(2)} = \beta_{129}^{(3)} = 2$.

Testing signficance

We consider $j = 85, 129, 167, 187, 211, 270$ and $j = 1, 2, 3$ to measure powers and type I errors, respectively. Tables 6 and 7 show the powers and type I errors of different methods when $\alpha = 0.05$. While all methods produce type I errors close to or below the significance level, DFGL has higher powers compared to all other methods.

The observed higher power of DFGL in testing for significance and homogeneity compared to Lasso-based approaches is attributed to the fused group Lasso regularization. This regularization method allows for increasing sample sizes in regression parameter estimation by combining subpopulations, leading to more accurate statistical inference. These results suggest the effectiveness of our DFGL in analyzing data characterized by limited sample sizes for groups, such as the CCLE data.

			Methods				
j	$\min_g \beta_j^{(g)}$	$\max_g \beta_j^{(g)}$	DFGL	DL	DPL	DR-B	DL-B
1	0.0	0.0	0.03	0.03	0.04	0.00	0.02
2	0.0	0.0	0.02	0.01	0.00	0.00	0.00
3	0.0	0.0	0.00	0.00	0.00	0.00	0.01
85	1.5	1.5	0.02	0.04	0.02	0.00	0.04
129	2.0	2.0	0.04	0.09	0.09	0.00	0.10

Table 5. Type I error for testing $H_0 : \beta_j^{(1)} = \cdots = \beta_j^{(G)}$ vs H_1 : not H_0 at $\alpha = 0.05$.

			Methods				
j	$\min_g \beta_j^{(g)}$	$\max_g \beta_j^{(g)}$	DFGL	DL	DPL	DR-B	DL-B
85	1.5	1.5	0.99	0.92	0.91	0.64	0.90
129	2.0	2.0	1.00	1.00	1.00	0.97	1.00
167	− 1.0	1.0	0.67	0.43	0.47	0.14	0.37
187	− 1.0	1.0	0.71	0.47	0.46	0.09	0.42
211	− 0.8	0.8	0.41	0.30	0.23	0.05	0.22
270	− 0.8	0.8	0.39	0.26	0.26	0.04	0.25

Table 6. Power for testing $H_0 : \beta_j^{(1)} = \cdots = \beta_j^{(G)}$ vs H_1 : not H_0 at $\alpha = 0.05$.

			Methods				
j	$\min_g \beta_j^{(g)}$	$\max_g \beta_j^{(g)}$	DFGL	DL	DPL	DR-B	DL-B
1	0.0	0.0	0.04	0.04	0.03	0.00	0.02
2	0.0	0.0	0.02	0.01	0.00	0.00	0.00
3	0.0	0.0	0.01	0.02	0.01	0.00	0.03

Table 7. Type I error for testing $H_0 : \beta_j^{(1)} = \cdots = \beta_j^{(G)}$ vs H_1 : not H_0 at $\alpha = 0.05$.

Multiple testing

In this section, we consider the two multiple testing problems (15) and (16), respectively. As in section Multiple testing, we use the BH procedure to control the familywise error rate. In this analysis, we don't consider DR-B because it is conservative as observed in sections "Testing homogeneity" and "Testing signficance". Table 8 summarizes the results at the $\alpha = 0.05$ significance level. Despite the limited sample size, which results in relatively low power, especially when testing for homogeneity, DFGL outperforms other methods in terms of higher power. All methods provide FWER below the nominal level $\alpha = 0.05$.

Computational time comparison

In this subsection, we discuss the computational cost of computing the DFGL estimate and compare it with the costs of computing three estimates used in DL, DL-E, and DR-B. To compute Lasso or Ridge penalized estimate, we use R package glmnet. For the penalty parameters, we consider 100 candidates for (λ_1, λ_2) for DFGL, and 100 candidates for the penalty parameter for DL, DL-E, and DR-B. The penalty parameters are determined using 5-fold Cross-Validation in all methods. Parallel computing is employed to compute estimates for each subpopulation separately in DL and DL-E.

In order to assess computational efficiency, we examine a simulation scenario with the AR(1) covariance matrix for the covariates, $n_1 = \cdots = n_G = 200$, and $p = 80$. Table 9 presents the computational times required to compute four estimates, determined from 100 Monte Carlo simulations and executed on an Intel Xeon (2.20 GHz) processor, are presented in Table 9. A noteworthy observation from Table 9 is that while the DFGL implementation requires slightly more time in comparison to the other methods, it remains significantly expeditious. This can be explained by the fact that DFGL involves addressing a more intricate large-scale problem, entailing a combined penalty of both group Lasso and fusion-type components, whereas the other methods involve simpler Lasso or Ridge techniques entailing a single penalty.

Application to the CCLE data

The dataset

In this section, we present real data analyses when our method is applied to the CCLE data. The CCLE data contains information on cancer treatment responses for 24 drugs on 504 cancer cell lines of 23 cancer types, where the transcription profile of each cell line is characterized by the measured expression levels of 19,177 probes. Cancer cell lines are widely used to understand cancer biology and test the efficacy of novel therapies[43], and are also used to identify predictive biomarkers for anticancer drug sensitivity[13,44]. We consider three cancer types that include at least 30 cancer cell lines: Lymphoid, Lung, and Skin in our CCLE data analysis. Our main objective is to check whether a specific gene is significant to binary drug response in at least one of these cancer types, and whether such significant genes have heterogeneous effects on a drug across the cancer types. Analyses similar to ours could be useful in two ways. First, examining the significance of the effects of a gene across different cancer types may lead to the identification of potential gene expression markers of drug response that can be used for multiple cancer types. Such versatile gene expression markers are valuable in research[16]. Second, studying the heterogeneity of a gene's effects across cancer types can provide insights into understanding differences in drug sensitivity across cancer types.

Following Park et al.[45], we classify cancer cell lines into two categories for each drug. If a drug response value (IC50) is less than 0.5, then the cancer cell line is assigned to the "sensitive" category; otherwise, it is assigned to the "resistant" category. Then, most cancer cell lines are either sensitive or resistant to most drugs in some of

	Power				FWER			
Testing	DFGL	DL	DPL	DL-B	DFGL	DL	DPL	DL-B
Homogeneity	0.120 (0.161)	0.020 (0.068)	0.015 (0.060)	0.025 (0.075)	0.040	0.000	0.000	0.000
Significance	0.408 (0.119)	0.227 (0.096)	0.227 (0.099)	0.133 (0.085)	0.010	0.010	0.000	0.010

Table 8. Performances of multiple testing at $\alpha = 0.05$.

DFGL	DL	DL-E	DR-B
59.58 (0.30)	23.13 (0.63)	1.89 (0.19)	16.98 (1.46)

Table 9. Average computational times (in seconds) for implementing estimates.

	Cancer type					
	Lung		Lymphoid		Skin	
Drug	Sensitive	Resistance	Sensitive	Resistance	Sensitive	Resistance
17-AAG	20	70	20	48	9	30
AZD6244	84	5	56	12	18	21
Irinotecan	4	46	15	30	12	18
PD-0325901	74	16	52	16	10	29
Topotecan	55	35	8	60	21	18

Table 10. The number of sensitive cell lines for each drug across cancer types.

the three cancer types. For example, all lymphoid cancer cell lines are resistant to Erlotinib, and only one of the lung cancer cell lines is sensitive to Panobinostat. After removing these imbalanced drugs in our analysis, we consider the following five drugs: 17-AAG, AZD6244, Irinotecan, PD-0325901, and Topotecan. Table 10 presents the number of sensitive cell lines for each of these five drugs.

The analysis of ultra-high-dimensional (UHD) data, such as CCLE data analysis, is accompanied by several challenges, including high collinearity, spurious correlation, noise accumulation, and a significant computational burden. To alleviate these difficulties inherent in UHD data, it is desirable to reduce the dimensionality of the feature space[46,47]. Similar to sure independence screening[46], we removed relatively irrelevant genes to each drug, respectively, before fitting models. The gene screening procedure is as follows:

1. We selected the top 3,000 genes with the largest sample variances.
2. For $j = 1, \ldots, 3000$, we applied logistic regression using each gene and two dummy variables indicating cancer types.
3. We selected the top 300 genes with the smallest p-values for the significance test.

These screening procedures have been used in the literature on high-dimensional regressions (e.g. Park et al.[48], Wang et al.[49], Li et al.[45]).

While we analyze a set of $p = 300$ genes obtained from the screening procedure described above, the number of genes still exceeds the sample sizes for the cancer types. As a result, standard maximum likelihood estimation couldn't be used for statistical inference in a model containing the $p = 300$ genes. Lasso-based approaches, including debiased Lasso[17], which do not incorporate a fusion penalty, can be considered to test for heterogeneity or significance of a gene's effects. However, it is worth noting that approaches based on lasso penalization may yield inaccurate results when applied to data characterized by limited sample sizes, as in the case of our CCLE dataset. This limitation was demonstrated in our simulation study in section "Simultation study for an imbalanced design". To achieve more accurate statistical inference, we use our DFGL based on a fusion penalty that combines subpopulations.

Results

First, we identify important genes using the proposed penalized estimation (5). Figure 1 shows the estimated cancer-specific coefficients for selected genes. We can see that for each drug, the penalized estimates of the coefficients for most genes are similar for at least two of the three cancer types. Next, we apply the proposed simultaneous significance test to investigate whether the genes detected through the penalization have significant effects on the drug when the effects of the other genes are adjusted. Figure 2 shows the debiased estimates corresponding to genes identified by the proposed significance test at the significance level $\alpha = 0.05$. Comparing Figs. 1 and 2, we found that most genes selected by the penalization are also significant at the significance level 0.05. However,

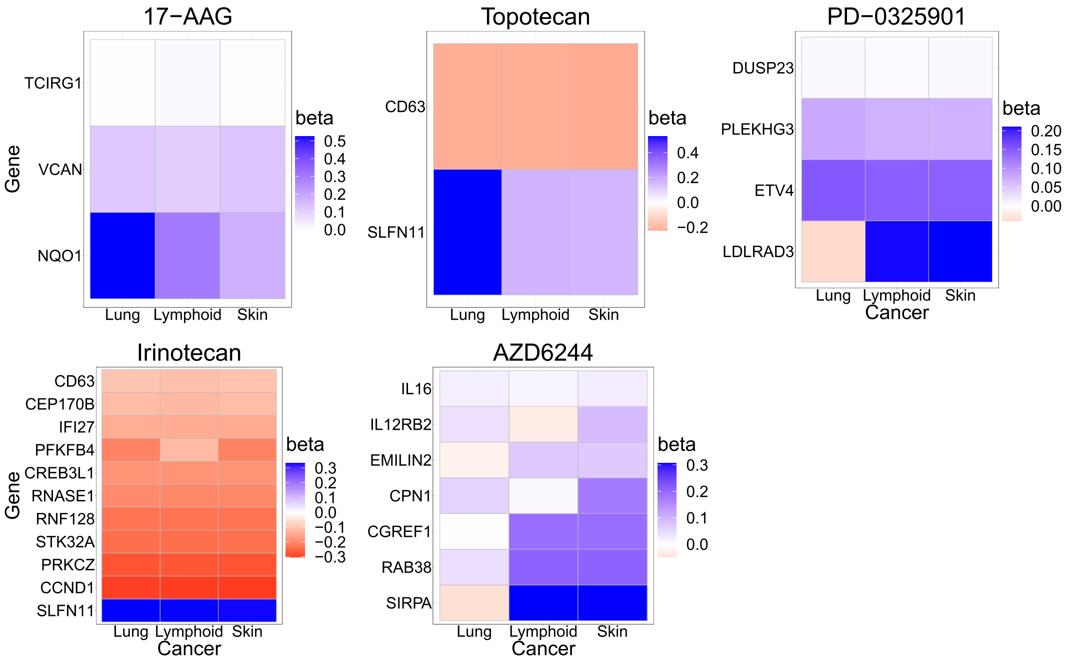


Figure 1. Heatmaps for fused group Lasso estimates $\hat{\beta}_j^{(g)}$. Genes with estimated regression coefficients of 0's are omitted when drawing the heatmap. The heatmap was created using the ggplot2 package[50] (version 3.4.3; https://cran.r-project.org/web/packages/ggplot2/index.html) in R software[51] (version 4.2.2 for Windows; https://cran.r-project.org/bin/windows/base/old/).

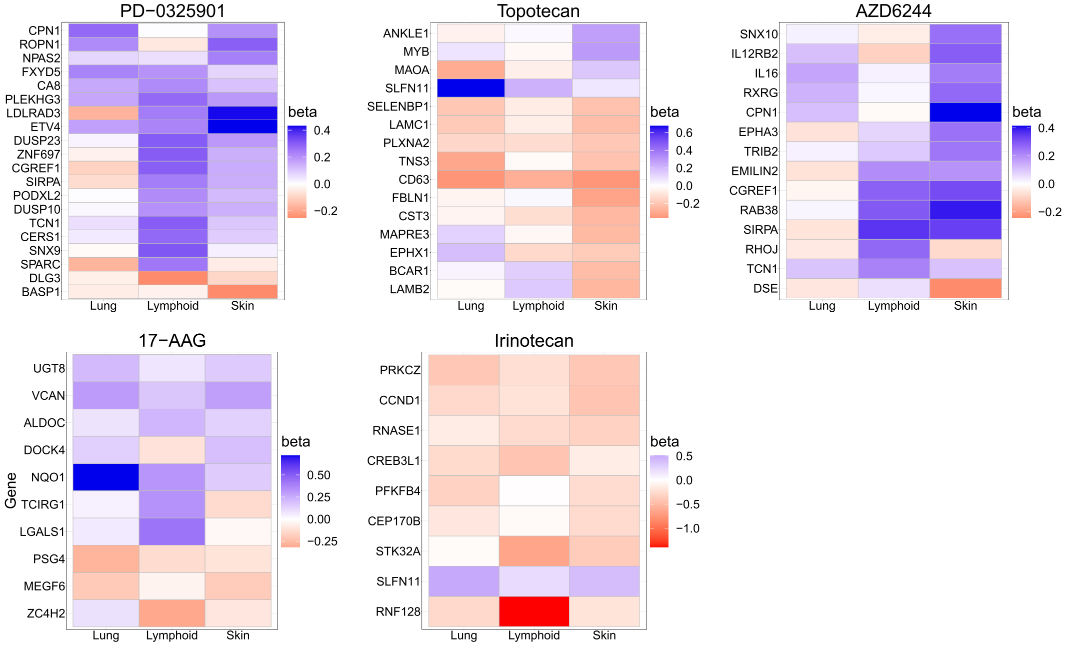
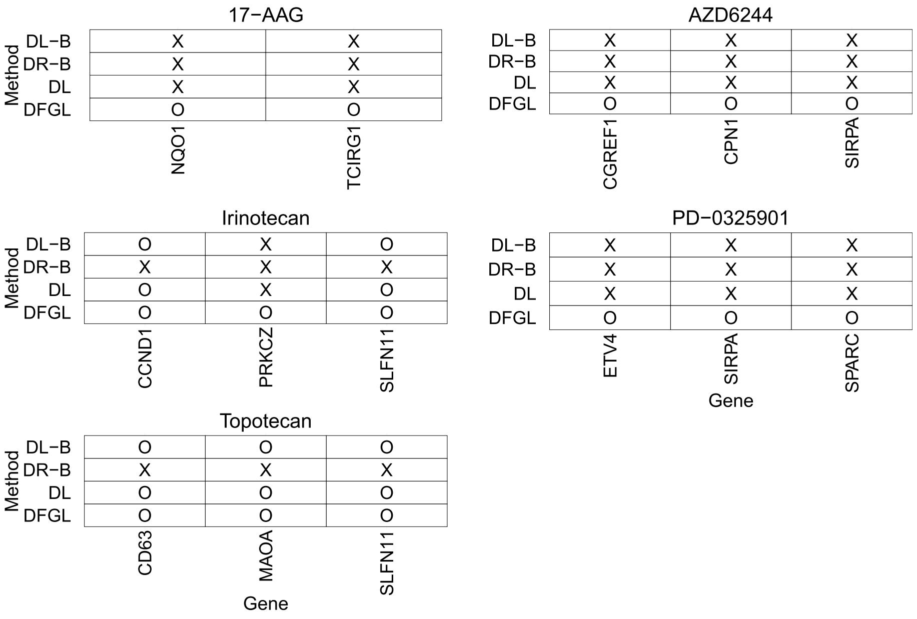


Figure 2. Heatmaps for debiased estimates $\hat{b}_j^{(g)}$. Genes with p-value for the simultaneous significance test greater than 0.05 were omitted when drawing the heatmap. The heatmap was created using the ggplot2 package[50] (version 3.4.3; https://cran.r-project.org/web/packages/ggplot2/index.html) in R software[51] (version 4.2.2 for Windows; https://cran.r-project.org/bin/windows/base/old/).



17−AAG

Method	NQO1	TCIRG1
DL−B	X	X
DR−B	X	X
DL	X	X
DFGL	O	O

AZD6244

Method	CGREF1	CPN1	SIRPA
DL−B	X	X	X
DR−B	X	X	X
DL	X	X	X
DFGL	O	O	O

Irinotecan

Method	CCND1	PRKCZ	SLFN11
DL−B	O	X	O
DR−B	X	X	X
DL	O	X	O
DFGL	O	O	O

PD−0325901

Method	ETV4	SIRPA	SPARC
DL−B	X	X	X
DR−B	X	X	X
DL	X	X	X
DFGL	O	O	O

Topotecan

Method	CD63	MAOA	SLFN11
DL−B	O	O	O
DR−B	X	X	X
DL	O	O	O
DFGL	O	O	O

Figure 3. The results of testing significance for some selective genes. The heatmap was created using the ggplot2 package[50] (version 3.4.4; https://cran.r-project.org/web/packages/ggplot2/index.html) in R software[51] (version 4.3.1 for Windows; https://cran.r-project.org/bin/windows/base/old/)..

some genes that were not detected by the penalization were found to be significant for those drugs. Such genes seem to have a significant effect in some of the cancer types or have weak signals across cancer types, as shown in Fig. 2. For example, the absolute value of the estimated effect of BASP1 on PD-0325901 is approximately 0.25 in the skin cancer and less than 0.04 in the other cancer types.

We also use DL, DL-B, and DR-B for comparison with DFGL. In the simulation in section "Simultation study for an imbalanced design", DPL shows similar performance to DL, so we do not consider DPL in the analysis. Note that DL-E can not be used in this analysis because the number of genes is larger than the sample sizes for the cancer types. Figure 3 compares the results of these methods in terms of testing significance for specific genes. We observe that DR-B seems to be too conservative, as observed in our simulation analysis. However, both the Lasso-based approaches and our method identify some common genes. For example, as shown in Fig. 3, all methods except for DR-B indicate the significance of the effects of SLFN11 on Irinotecan and Topotecan across cancer types. SLFN11 was previously identified as relevant to Irinotecan and Topotecan when a penalized mixture regression was applied to the CCLE data[5].

When performing a sensitivity analysis based on Bootstrap in section S4.4 in the Supplementary materials, we observe that the following drug-gene pairs are relatively frequently identified by our significance test: SLFN11-Topotecan, SLFN11-Irinotecan, NQO1-17AAG, SIRPA-AZD6244, and ETV4-PD-0325901. These results demonstrate the significance of SLFN11 for Topotecan and Irinotecan. Notably, the remaining three gene-drug pairs were not detected by Lasso-based approaches at $\alpha = 0.05$, as shown in Fig. 3. However, NQO1 was identified as an important gene for 17-AAG in the previous analyses of CCLE data[5,52], and ETV4 was also identified as a related gene to PD-0325901 in the CCLE data analysis performed by Liang et al.[52]. In addition, expression of ETV4, detected as a significant gene expression to PD-0325901, might modulate sensitivity to a MEK inhibitor trametinib[53]. Hayashi et al.[54] discovered that activation of MEK was induced by ligation of SIRPβ, while SIRPα (SIRPA) is significant to AZD6244. Despite the empirical evidence supporting our significant test results, the gene-drug pairs NQO1-17AAG, ETV4-PD-0325901, and SIRPα-AZD6244 were not identified by any of the other significance tests at $\alpha = 0.05$. These results suggest that DFGL may provide more accurate results for the significance of the association between a gene and a drug by combining different cancer types, as opposed to approaches based on Lasso or Ridge. Furthermore, as shown in Fig. 2, we expect the associations SIRPA-AZD6244 and ETV4-PD-0325901 to be relatively weak in specific cancer types. The relatively weak effects of SIPRA and ETV4 may result in Lasso-based approaches failing to detect them at the 0.05 significance level. This observation is consistent with our simulation results presented in section "Simultation study for an imbalanced design", which indicate that Lasso-based approaches have lower power compared to our approach, especially when testing the significance of effects for covariates with relatively weak signals in simulated data with small sample sizes.

Our equivalence test shows different results for different cancer types, as summarized in Table 11, i.e., some genes have heterogeneous coefficients depending on the cancer type at the significance level $\alpha = 0.05$. The estimated effect of SLFN11 on Topotecan is relatively large in lung cancer cell lines compared to other cancer types. The p-value for the equivalence test corresponding to SLFN11 was less than 0.05 when Topoptecan was considered. However, we note that other methods suggest a lack of significance for the heterogeneity of the effects of SLFN11 on Topotecan at the $\alpha = 0.05$ significance level. Further investigation is needed to determine

Drug	Target	Gene	# gene
Irinotecan	TOP I	RNF128	1
17-AAG	HSP90	TCIRG1, FUCA2, ZC4H2, LGALS1	4
AZD6244	MEK	CPN1, SIRPα, IL12RB2, RAB38, RHOJ, CGREF1	6
PD-0325901	MEK	LDLRAD3, SPARC, CGREF1, CYP27A1, BTBD19, PMP22, SNX9, BAMBI	8
Topotecan	TOP I	SLFN11, MAOA, MAPRE3, BCAR1, EPHX1, RCSD1, FBLN1, LAMB2	8

Table 11. Genes identified by the proposed homogeneity test at $\alpha = 0.05$. Genes are listed in order of p-value.

whether SLFN11 has heterogeneous effects on the response to Topotecan in patients with different cancer types. However, in the sensitivity analysis based on Bootstrap in Section S4.4 in the Supplementary materials, the association between SLFN11 and Topotecan appears to be the most significant among all gene-drug pairs in terms of heterogeneous effects across cancer types. In addition, in the previous analysis of CCLE data[5], a penalized mixture regression suggested that the effects of SLFN11 are different between some of the clusters. These results suggest that our method may uncover underlying heterogeneous effects of a gene across cancer types that are difficult to capture using Lasso or Ridge-based methods. We observed that some genes identified by the homogeneity test have positive estimated coefficients in specific cancer types. For Topotecan, the estimated effect of gene MAOA is only positive in skin cancer cell lines. Low expression of MAOA has been observed in melanoma skin cancer compared with normal samples[55], but high expression of MAOA has been observed in lung cancer tissues[56] and lymphoma[57].

Conclusion

In this paper, we propose two different tests: (1) testing the homogeneity of the effects of the covariate across different groups and (2) testing the significance of the covariate over groups. We develop non-asymptotic analyses for the proposed fused group Lasso and prove that the debiased test statistics admit chi-squared approximations even in the presence of high dimensional variables. The proposed tests generally outperform the existing bias-correction methods based on Lasso[17,36] or Ridge[39] in that it proves higher power, while it controls type I error quite well as shown in section "Simulation study". Through CCLE data analysis, we can observe that the proposed method can make significant scientific discoveries.

From a methodological point of view, there are some extensions to our method. First, our tests can be applied to generalized linear models, including linear regression and the Poisson regression model, although we focus on logistic regression. In addition, our theoretical analyses can be extended to the generalized linear regression (GLM) setting. Second, we expect that the performance of the proposed tests can be improved by simultaneously estimating the inverse of the information matrices across subpopulations, as in the joint estimation of precision matrices[58,59].

From the perspective of CCLE data analysis, there are several interesting directions for future research. Although our primary focus in CCLE data analysis was on the use of gene expression, which is known to be predictive of drug response[60], other omics features such as DNA copy number are available in the analysis of CCL data. It is of great interest to investigate which omics data are most predictive of drug response in a specific cancer type, or have heterogeneous effects on drug response across cancer types. Given the different characteristics of different types of omics data, it is expected that our method may have some limitations in the analysis of multi-omics data. Therefore, a sophisticated extension of our method in estimation and construction of test statistics will be needed for the analysis of multi-omics data. In CCLE data, there are cell lines with missing responses to a drug. Therefore, an extension of our method to include cell lines with missing drug responses would be beneficial for the analysis of CCLE data.

Data availability

All data generated or analyzed during this study are included in supplementary information files.

Received: 7 August 2023; Accepted: 30 November 2023
Published online: 11 December 2023

References

1. Caroli, J., Dori, M. & Bicciato, S. Computational methods for the integrative analysis of genomics and pharmacological data. *Front. Oncol.* **10**, 185 (2020).
2. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
3. Azuaje, F. Computational models for predicting drug responses in cancer research. *Brief. Bioinf.* **18**, 820–829 (2017).
4. Geeleher, P., Cox, N. J. & Huang, R. S. Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. *Genome Biol.* **15**, 782 (2014).
5. Li, Q., Shi, R. & Liang, F. Drug sensitivity prediction with high-dimensional mixture regression. *PloS one* **14**, e0212108 (2019).
6. Dong, Z. *et al.* Anticancer drug sensitivity prediction in cell lines from baseline gene expression through recursive feature selection. *BMC Cancer* **15**, 14 (2015).
7. Riddick, G. *et al.* Predicting in vitro drug sensitivity using random forests. *Bioinformatics* **27**, 220–224 (2011).
8. Choi, J., Park, S. & Ahn, J. Refdnn: A reference drug based neural network for more accurate prediction of anticancer drug resistance. *Sci. Rep.* **10**, 1861 (2020).

9. Huang, E. W., Bhope, A., Lim, J., Sinha, S. & Emad, A. Tissue-guided lasso for prediction of clinical drug response using preclinical samples. *PLoS Comput. Biol.* **16**, e1007607 (2020).
10. Oh, D. Y. & Bang, Y. J. HER2-targeted therapies—a role beyond breast cancer. *Nat. Rev. Clin. Oncol.* **17**, 33–48 (2020).
11. Zhao, Z., Wang, S., Zucknick, M. & Aittokallio, T. Tissue-specific identification of multi-omics features for pan-cancer drug response prediction. *iScience* **25**, 104767 (2022).
12. Rahman, R., Matlock, K., Ghosh, S. & Pal, R. Heterogeneity aware random forest for drug sensitivity prediction. *Sci. Rep.* **7**, 11347 (2017).
13. Garnett, M. J. *et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**, 570–575 (2012).
14. Zintzaras, E. & Ioannidis, J. P. A. Heterogeneity testing in meta-analysis of genome searches. *Genet. Epidemiol.* **28**, 123–137 (2005).
15. Lewis, C. M. & Levinson, D. F. Testing for genetic heterogeneity in the genome search meta-analysis method. *Genet. Epidemiol.* **30**, 348–355 (2006).
16. Martinez-Ledesma, E., Verhaak, R. G. & Treviño, V. Identification of a multi-cancer gene expression biomarker for cancer clinical outcomes using a network-based algorithm. *Sci. Rep.* **5**, 11966 (2015).
17. Van de Geer, S., Bühlmann, P., Ritov, Y. A. & Dezeure, R. On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Stat.* **42**, 1166–1202 (2014).
18. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B: Stat. Methodol.* **58**, 267–288 (1996).
19. Javanmard, A. & Montanari, A. Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.* **15**, 2869–2909 (2014).
20. Zhang, C. H. & Zhang, S. S. Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B: Stat. Method.* **76**, 217–242 (2014).
21. Ning, Y. & Liu, H. A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *Ann. Stat.* **45**, 158–195 (2017).
22. Cai, T. T., Guo, Z. & Ma, R. Statistical inference for high-dimensional generalized linear models with binary outcomes. *J. Am. Stat. Assoc.* **118**, 1319–1332 (2023).
23. Shi, C., Song, R., Lu, W. & Li, R. Statistical inference for high-dimensional models via recursive online-score estimation. *J. Am. Stat. Assoc.* **116**, 1307–1318 (2021).
24. Ma, R., Tony Cai, T. & Li, H. Global and simultaneous hypothesis testing for high-dimensional logistic regression models. *J. Am. Stat. Assoc.* **116**, 984–998 (2021).
25. Yuan, M. & Lin, Y. Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B: Stat. Method.* **68**, 49–67 (2006).
26. Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. & Knight, K. Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc. Ser. B: Stat. Method.* **67**(1), 91–108 (2005).
27. Tang, L. & Song, P. X. Fused lasso approach in regression coefficients clustering: Learning parameter heterogeneity in data integration. *J. Mach. Learn. Res.* **17**, 1–23 (2016).
28. Zhou, J., Liu, J., Narayan, V.A., & Ye, J. Modeling disease progression via fused sparse group lasso. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 1095–1103 (2012).
29. Negahban, S. N., Ravikumar, P., Wainwright, M. J. & Yu, B. A unified framework for high-dimensional analysis of m -estimators with decomposable regularizers. *Stat. Sci.* **27**, 538–557 (2012).
30. Cai, T. T., Zhang, A. R. & Zhou, Y. Sparse group lasso: Optimal sample complexity, convergence rate, and statistical inference. *IEEE Trans. Inf. Theory* **68**, 5975–6002 (2019).
31. Ollier, E. & Viallon, V. Regression modelling on stratified data with the lasso. *Biometrika* **104**, 83–96 (2017).
32. Zou, H. The adaptive lasso and its oracle properties. *J. Am. Stat. Assoc.* **101**, 1418–1429 (2006).
33. Meinshausen, N. & Bühlmann, P. High-dimensional graphs and variable selection with the lasso. *Ann. Stat.* **34**, 1436–1462 (2006).
34. Tian, Y. & Feng, Y. Transfer learning under high-dimensional generalized linear models. *J. Am. Stat. Assoc.* **2022**, 1–14 (2022).
35. Caner, M. Generalized linear models with structured sparsity estimators. *J. Econ.* **236**, 105478 (2023).
36. Xia, L., Nan, B. & Li, Y. Debiased lasso for generalized linear models with a diverging number of covariates. *Biometrics* **79**, 344–357 (2023).
37. Beck, A. & Teboulle, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2**, 183–202 (2009).
38. Boyd, S., Parikh, N., Chu, E., Peleato, B. & Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3**, 1–22 (2011).
39. Bülmann, P. Statistical significance in high-dimensional linear models. *Bernoulli* **819**, 1212–1242 (2013).
40. Dezeure, R., Bülmann, P., Meier, L. & Meinshausen, N. High-dimensional inference: Confidence intervals, p-values and R-software hdi. *Stat. Sci.* **30**, 533–558 (2015).
41. Mitra, R. & Zhang, C. H. The benefit of group sparsity in group inference with de-biased scaled group lasso. *Electron. J. Stat.* **10**, 1829–1873 (2016).
42. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6**, 65–70 (1979).
43. Sharma, S. V., Haber, D. A. & Settleman, J. Cell line-based platforms to evaluate the therapeutic efficacy of candidate anticancer agents. *Nat. Rev. Cancer* **10**, 241–253 (2010).
44. Nakatsu, N. *et al.* Chemosensitivity profile of cancer cell lines and identification of genes determining chemosensitivity by an integrated bioinformatical approach using cDNA arrays. *Mol. Cancer Therapeut.* **4**, 399–412 (2005).
45. Park, S., Lee, E. R. & Zhao, H. Low-rank regression models for multiple binary responses and their applications to cancer cell-line encyclopedia data. *J. Am. Stat. Assoc.* **2022**, 1–15 (2022).
46. Fan, J. & Lv, J. Sure independence screening for ultrahigh dimensional feature space. *J. R. Stat. Soc. Ser. B: Stat. Method.* **70**, 849–911 (2008).
47. Fan, J. & Lv, J. *Sure Independence Screening, Statistics Reference Online* (Wiley, 2018).
48. Wang, L., Wu, Y. & Li, R. Quantile regression for analyzing heterogeneity in ultra-high dimension. *J. Am. Stat. Assoc.* **107**, 214–222 (2012).
49. Li, Y., Nan, B. & Zhu, J. Multivariate sparse group lasso for the multivariate multiple linear regression with an arbitrary group structure. *Biometrics* **71**, 354–363 (2015).
50. Wickham, H. ggplot2. *Wiley Interdiscipl. Rev.: Comput. Stat.* **3**, 180–185 (2011).
51. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing (2021). https://www.R-project.org/.
52. Liang, F., Li, Q. & Zhou, L. Bayesian neural networks for selection of drug sensitive genes. *J. Am. Stat. Assoc.* **113**, 955–972 (2018).
53. Wang, B. *et al.* ATXN1L, CIC, and ETS transcription factors modulate sensitivity to MAPK pathway inhibition. *Br. J. Cancer* **18**, 1543–1557 (2017).
54. Hayashi, A. *et al.* Positive regulation of phagocytosis by SIRPβ and its signaling mechanism in macrophages. *J. Biol. Chem.* **279**, 29450–29460 (2004).
55. Rybaczyk, L. A., Bashaw, M. J., Pathak, D. R. & Huang, K. An indicator of cancer: Downregulation of monoamine oxidase-A in multiple organs and species. *BMC Genom.* **9**, 1–9 (2008).

56. Liu, F. *et al.* Increased expression of monoamine oxidase A is associated with epithelial to mesenchymal transition and clinicopathological features in non-small cell lung cancer. *Oncol. Lett.* **15**, 3245–3251 (2018).
57. Li, P. C. *et al.* Monoamine oxidase A is highly expressed in classical Hodgkin lymphoma. *J. Pathol.* **243**, 220–229 (2017).
58. Lee, W. & Liu, Y. Joint estimation of multiple precision matrices with common structures. *J. Mach. Learn. Res.* **16**, 1035–1062 (2015).
59. Cai, T. T., Li, H., Liu, W. & Xie, J. Joint estimation of multiple high-dimensional precision matrices. *Stat. Sin.* **27**, 445–464 (2016).
60. Costello, J. C. *et al.* A community effort to assess and improve drug sensitivity prediction algorithms. *Nat. Biotechnol.* **32**, 1202–1212 (2014).

Acknowledgements

Seyoung Park is supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT) (No. NRF-2022R1A2C4002150). Eun Ryung Lee is supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT) (No. NRF-2022R1A2C1012798).

Author contributions

E.R.L. and S.P. are corresponding authors who wrote the manuscript and derive the theoretical and computational aspects. H.K. derived algorithms and conduct numerical analyses. All authors reviewed the manuscript

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-48903-x.

Correspondence and requests for materials should be addressed to E.R.L. or S.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.