Novel Statistical Methods on Identifying Subgroups and Predicting Individualized Treatment Effects with Clustered/Longitudinal Data

Zhikuan Quan

Advisor: Shuai Chen

June 13, 2023

Introduction: Personalized Medicine

- A major challenge in the domain of medical science and healthcare is to evaluate the effect of an intervention or exposure (referred as "treatment") on the outcome.
- Traditional treatment guidelines are based on the average treatment effect (ATE) on the entire population.

Introduction: Personalized Medicine

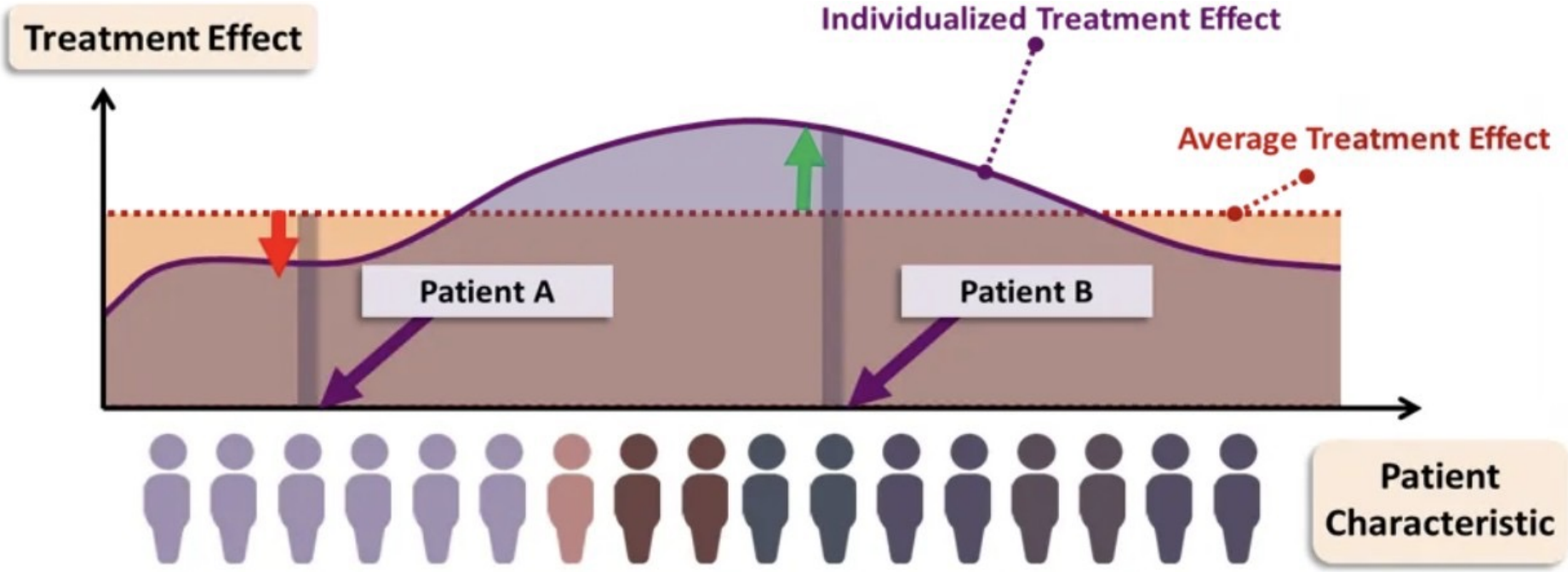


Figure: *Transit from ATE to ITE*

Introduction: Personalized Medicine

Individualized Treatment Effect (ITE)

- Goal: Novel statistical methods to estimate ITE
- Identify the subgroups that have heterogeneous treatment effects
- Predict the individualized treatment effects for new subjects

Introduction: Motivating Example

- Motivating Example: **Maternal Immune Activation (MIA) Study**
  - MIA during pregnancy alters postnatal brain growth and cognitive development in nonhuman primate offspring. *[Vlasova et al., 2021]*
  - High maternal status for vitamin D, iron, zinc, or choline could promote resilience to the effects of MIA. *[Meyer, 2019]*

Introduction: Motivating Example

- Motivating Example: **Maternal Immune Activation (MIA) Study**
  - MIA during pregnancy alters postnatal brain growth and cognitive development in nonhuman primate offspring. *[Vlasova et al., 2021]*
  - High maternal status for vitamin D, iron, zinc, or choline could promote resilience to the effects of MIA. *[Meyer, 2019]*
- MIA causes aberrant outcomes in only a subset of pregnancies.
  $\rightarrow$ **How to predict whether a pregnancy is susceptible to MIA?**

Introduction: Motivating Example

- Motivating Example: **Maternal Immune Activation (MIA) Study**
  - MIA during pregnancy alters postnatal brain growth and cognitive development in nonhuman primate offspring. *[Vlasova et al., 2021]*
  - High maternal status for vitamin D, iron, zinc, or choline could promote resilience to the effects of MIA. *[Meyer, 2019]*
- MIA causes aberrant outcomes in only a subset of pregnancies.
  $\rightarrow$ **How to predict whether a pregnancy is susceptible to MIA?**
- Goal: Estimate ITE of MIA (ie, individualized MIA effect)
  - Identify the subgroups that are resilient or susceptible to MIA using baseline information during pregnancy
  - Facilitate the intervention for high-risk mothers during pregnancy and early intervention for high-risk offspring.

Introduction: Literature Review

Statistical approaches to estimate ITE:

- Naive Full Regression Model
  - Need strong assumptions in model specification.

Introduction: Literature Review

Statistical approaches to estimate ITE:

- Naive Full Regression Model
  - Need strong assumptions in model specification.
- Robust methods bypassing the modeling of main effects: General Framework of Subgroup Identification: *[Chen et al., 2017]*
  - A-Learning: Model the treatment-covariate interaction with pre-estimated propensity score *[Murphy et al., 2003]*
  - Weighting Approaches: Inverse probability weighted estimator
    - Outcome Weighted Learning *[Zhao et al., 2012]*
    - D-Learning *[Tian et al., 2014]*

Statistical approaches to estimate ITE:

- Naive Full Regression Model
  - Need strong assumptions in model specification.
- Robust methods bypassing the modeling of main effects: General Framework of Subgroup Identification: *[Chen et al., 2017]*
  - A-Learning: Model the treatment-covariate interaction with pre-estimated propensity score *[Murphy et al., 2003]*
  - Weighting Approaches: Inverse probability weighted estimator
    - Outcome Weighted Learning *[Zhao et al., 2012]*
    - D-Learning *[Tian et al., 2014]*
- Recent extensions to the robust methods:
  - Residual Weighted Learning: Use residual as outcome to reduce the variance of the estimator *[Liu et al., 2018]*
  - Doubly Robust Direct Learning: Double robustness with possibly mis-specified main effect and propensity score models *[Meng et al., 2022]*

Introduction: Literature Review

Statistical approaches to estimate ITE:

- **However, most of current robust statistical approaches are only for single-outcome data.**
  - Cannot handle clustered/longitudinal outcomes

Introduction: New Challenge

New challenges in complicated clustered/longitudinal data:

- The correlation of outcomes is common in health studies.
  - Longitudinal data: e.g. repeated measures of cytokines level over time
  - Clustered data: e.g. multiple offspring within the same dam
  - Multi-leveled data: e.g. repeated outcomes over time for each offspring, and multiple offspring from same dam
- The increasing availability and complexity of observational data
  - High-dimensional Data: e.g. EHR, genetics information
  - Non-linear relationships

Introduction: Recent Development for Correlated Data

Type	Method	Robust to main effect	Robust to propensity score	Subgroup identification	Estimation of ITE	Comments
Linear Mixed Model	Two-stage Method [Cho et al., 2017]			✅		• Strong assumption in modeling treatment effect as slope of linear time
Generalized Weighting Method	Huling's Method [Huling et al., 2019]	✅		✅	✅	• Did not account for the serial correlation • Not applicable in clustered data
Tree-based Algorithm	Interaction Tree [Wei et al., 2020]			✅		• Need large sample size • Only 0 cut-off to identify subgroup

Introduction: Our Contribution

We propose a novel statistical framework for clustered/longitudinal data, with following advantages:

- Account for the correlation in data
- Directly estimate the ITE in both randomized and observational data
- Identify subgroups with heterogeneous intervention effects
- Doubly robust property with respect to mis-specification of main effect or propensity score
- Allow regularization approach to handle high-dimensional data
- Allow flexible modeling of ITE using flexible function space or machine learning techniques

Methodology

Notations and Assumptions

Data: $\{(\boldsymbol{Y_i}, T_i, \boldsymbol{X_i}) : i = 1, ..., n, j = 1, ..., k_i\}$

- Outcome: $\boldsymbol{Y_i} = (Y_{i1}, ..., Y_{ik_i})'$ for i-th subject
  - Time: $\boldsymbol{t_i} = (t_1, ..., t_{k_i})'$
  - $\boldsymbol{Y_i}(t_j) := Y_{ij}$ is the j-th observation for subject i at time t_j.
- Treatment: $T_i \in \mathcal{T} = \{1, -1\}$ is assigned at baseline
- Baseline Covariates: $\boldsymbol{X_i} := (X_{i,1}, ..., X_{i,p})' \in \mathcal{X} \subseteq \mathbb{R}^p$

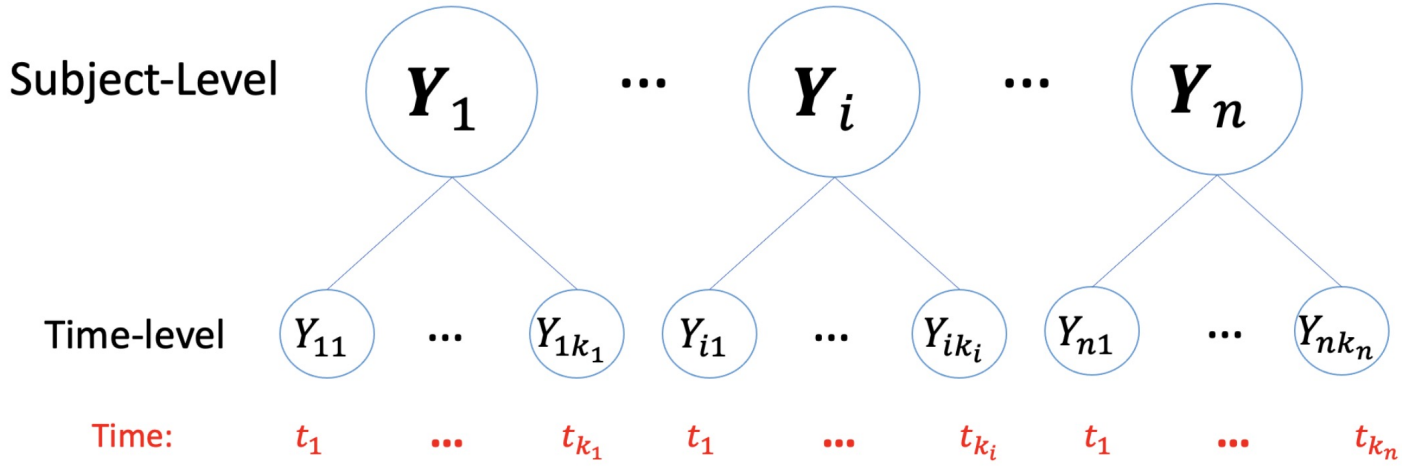


Figure: *Longitudinal Data*

Notations and Assumptions

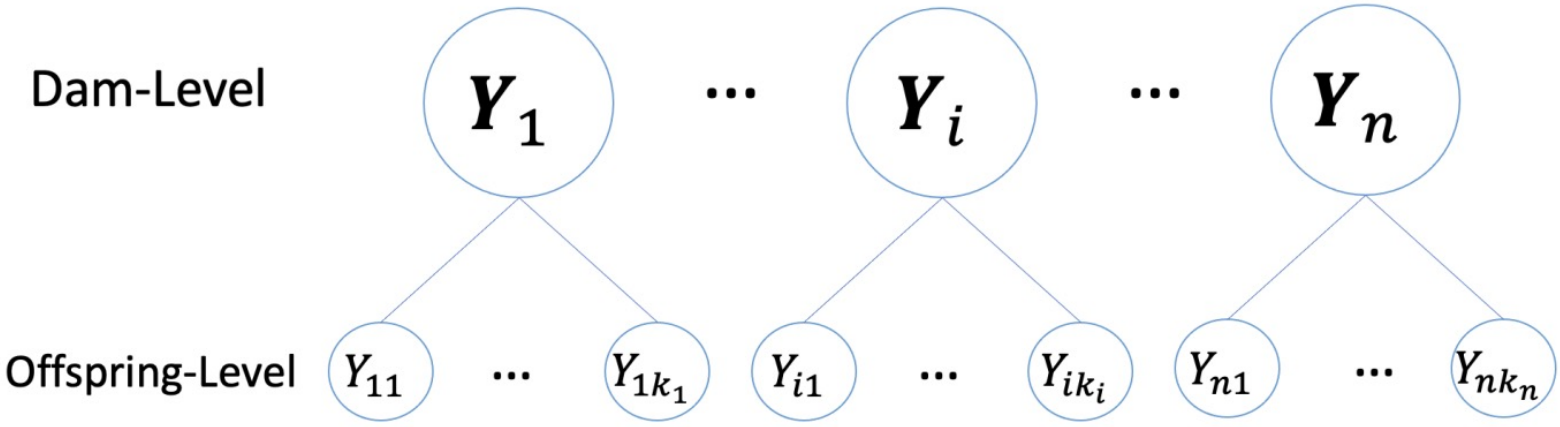


Figure: *Clustered Data in MIA Study*

Notations and Assumptions

- Potential Outcome: $\boldsymbol{Y_i^{(T_i)}}, T_i \in \{1, -1\}$
- **Causal Inference Framework**
  - Consistency Assumption:

$$\boldsymbol{Y_i} = I\{T_i = 1\}\boldsymbol{Y_i^{(1)}} + I\{T_i = -1\}\boldsymbol{Y_i^{(-1)}}$$

  - Unconfoundedness Assumption:

$$(\boldsymbol{Y_i^{(1)}}, \boldsymbol{Y_i^{(-1)}}) \perp\!\!\!\perp T_i|\boldsymbol{X_i}$$

  - Positivity Assumption:

$$\pi_1(\boldsymbol{X_i}) := P(T_i = 1|\boldsymbol{X_i}) \in (0, 1) \text{ and } \pi_{-1}(\boldsymbol{X_i}) = 1 - \pi_1(\boldsymbol{X_i})$$

Notations and Assumptions

We can decompose the continuous outcome into:

$$\boldsymbol{Y_i} = \boldsymbol{m}(\boldsymbol{X_i}, \boldsymbol{t_i}) + T_i \boldsymbol{\delta}(\boldsymbol{X_i}, \boldsymbol{t_i})/2 + \boldsymbol{\epsilon_i} \tag{1}$$

Notations and Assumptions

We can decompose the continuous outcome into:

$$\boldsymbol{Y}_i = \boldsymbol{m}(\boldsymbol{X_i}, \boldsymbol{t_i}) + T_i \boldsymbol{\delta}(\boldsymbol{X_i}, \boldsymbol{t_i})/2 + \boldsymbol{\epsilon_i} \tag{1}$$

- Main Effect is characterized by

$$\begin{aligned} \boldsymbol{m}(\boldsymbol{X_i}) :&= \mathbb{E}\left[(\boldsymbol{Y_i}^{(1)} + \boldsymbol{Y_i}^{(-1)})|\boldsymbol{X_i}\right]/2 \\ &= \{\mathbb{E}(\boldsymbol{Y_i}|T_i = 1, \boldsymbol{X_i}) + \mathbb{E}(\boldsymbol{Y_i}|T_i = -1, \boldsymbol{X_i})\}/2 \end{aligned}$$

where $\boldsymbol{m}(\boldsymbol{X_i}, \boldsymbol{t_i}) = (m(\boldsymbol{X_i}, t_1), ..., m(\boldsymbol{X_i}, t_{k_i}))'$

Notations and Assumptions

We can decompose the continuous outcome into:

$$\boldsymbol{Y}_i = \boldsymbol{m}(\boldsymbol{X}_i, \boldsymbol{t}_i) + T_i \boldsymbol{\delta}(\boldsymbol{X}_i, \boldsymbol{t}_i)/2 + \boldsymbol{\epsilon}_i \tag{1}$$

- Main Effect is characterized by

$$\begin{aligned} \boldsymbol{m}(\boldsymbol{X}_i) :&= \mathbb{E}\left[(\boldsymbol{Y}_i^{(1)} + \boldsymbol{Y}_i^{(-1)})|\boldsymbol{X}_i\right]/2 \\ &= \{\mathbb{E}(\boldsymbol{Y}_i|T_i = 1, \boldsymbol{X}_i) + \mathbb{E}(\boldsymbol{Y}_i|T_i = -1, \boldsymbol{X}_i)\}/2 \end{aligned}$$

  where $\boldsymbol{m}(\boldsymbol{X}_i, \boldsymbol{t}_i) = (m(\boldsymbol{X}_i, t_1), ..., m(\boldsymbol{X}_i, t_{k_i}))'$

- The individualized treatment effect (ITE) is represented by:

$$\boldsymbol{\delta}(\boldsymbol{X}_i, \boldsymbol{t}_i) := \mathbb{E}\left[(\boldsymbol{Y}_i^{(1)} - \boldsymbol{Y}_i^{(-1)})|\boldsymbol{X}_i\right]$$

  where $\boldsymbol{\delta}(\boldsymbol{X}_i, \boldsymbol{t}_i) = (\delta(\boldsymbol{X}_i, t_1), ..., \delta(\boldsymbol{X}_i, t_{k_i}))'$

Notations and Assumptions

We can decompose the continuous outcome into:

$$\boldsymbol{Y}_i = \boldsymbol{m}(\boldsymbol{X}_i, \boldsymbol{t}_i) + T_i \boldsymbol{\delta}(\boldsymbol{X}_i, \boldsymbol{t}_i)/2 + \boldsymbol{\epsilon}_i \tag{1}$$

- Main Effect is characterized by

$$\begin{aligned} \boldsymbol{m}(\boldsymbol{X}_i) :&= \mathbb{E}\left[(\boldsymbol{Y}_i^{(1)} + \boldsymbol{Y}_i^{(-1)})|\boldsymbol{X}_i\right]/2 \\ &= \{\mathbb{E}(\boldsymbol{Y}_i|T_i = 1, \boldsymbol{X}_i) + \mathbb{E}(\boldsymbol{Y}_i|T_i = -1, \boldsymbol{X}_i)\}/2 \end{aligned}$$

  where $\boldsymbol{m}(\boldsymbol{X}_i, \boldsymbol{t}_i) = (m(\boldsymbol{X}_i, t_1), ..., m(\boldsymbol{X}_i, t_{k_i}))'$

- The individualized treatment effect (ITE) is represented by:

$$\boldsymbol{\delta}(\boldsymbol{X}_i, \boldsymbol{t}_i) :\ = \mathbb{E}\left[(\boldsymbol{Y}_i^{(1)} - \boldsymbol{Y}_i^{(-1)})|\boldsymbol{X}_i\right]$$

  where $\boldsymbol{\delta}(\boldsymbol{X}_i, \boldsymbol{t}_i) = (\delta(\boldsymbol{X}_i, t_1), ..., \delta(\boldsymbol{X}_i, t_{k_i}))'$

- Random Error $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, ..., \epsilon_{ik_i})'$ with $\mathbb{E}(\boldsymbol{\epsilon}_i) = \boldsymbol{0}_{k_i}$ and invertible $\text{Var}(\boldsymbol{\epsilon}_i) = \boldsymbol{V}_i$

Notations and Assumptions

We can decompose the continuous outcome into:

$$\boldsymbol{Y}_i = \boldsymbol{m}(\boldsymbol{X}_i, \boldsymbol{t}_i) + T_i \boldsymbol{\delta}(\boldsymbol{X}_i, \boldsymbol{t}_i)/2 + \boldsymbol{\epsilon}_i \tag{1}$$

- Main Effect is characterized by

$$\begin{aligned} \boldsymbol{m}(\boldsymbol{X}_i) :&= \mathbb{E}\left[(\boldsymbol{Y}_i^{(1)} + \boldsymbol{Y}_i^{(-1)})|\boldsymbol{X}_i\right]/2 \\ &= \{\mathbb{E}(\boldsymbol{Y}_i|T_i = 1, \boldsymbol{X}_i) + \mathbb{E}(\boldsymbol{Y}_i|T_i = -1, \boldsymbol{X}_i)\}/2 \end{aligned}$$

  where $\boldsymbol{m}(\boldsymbol{X}_i, \boldsymbol{t}_i) = (m(\boldsymbol{X}_i, t_1), ..., m(\boldsymbol{X}_i, t_{k_i}))'$

- The individualized treatment effect (ITE) is represented by:

$$\boldsymbol{\delta}(\boldsymbol{X}_i, \boldsymbol{t}_i) :&= \mathbb{E}\left[(\boldsymbol{Y}_i^{(1)} - \boldsymbol{Y}_i^{(-1)})|\boldsymbol{X}_i\right]$$

  where $\boldsymbol{\delta}(\boldsymbol{X}_i, \boldsymbol{t}_i) = (\delta(\boldsymbol{X}_i, t_1), ..., \delta(\boldsymbol{X}_i, t_{k_i}))'$

- Random Error $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, ..., \epsilon_{ik_i})'$ with $\mathbb{E}(\boldsymbol{\epsilon}_i) = \mathbf{0}_{k_i}$ and invertible $\text{Var}(\boldsymbol{\epsilon}_i) = \boldsymbol{V}_i$

For clustered data, the time $\boldsymbol{t}_i$ can be excluded. (reduced model)

Review: A General Framework for Single Outcome

- For single outcome model with $\{(Y_i, T_i, \boldsymbol{X}_i) : i = 1, ..., n\}$:

$$\hat{\delta} := \underset{f \in \{\mathcal{X} \to \mathbb{R}\}}{\text{argmin}} \frac{1}{n} \sum_{i=1}^{n} \frac{M(Y_i, T_i f(\boldsymbol{X}_i)/2)}{\pi_{T_i}(\boldsymbol{X}_i)}$$

  where $M(.,.)$ is pre-specified loss function that characterizes the goodness of fit. *[Chen et al., 2017]*

- e.g. $M(a, b) = (a - b)^2$ for continuous outcome.

New Method for Correlated Data

- Our new method uses loss function:

$$M(\boldsymbol{a}, \boldsymbol{b}) = (\boldsymbol{a} - \boldsymbol{b})' \boldsymbol{V}^{-1} (\boldsymbol{a} - \boldsymbol{b})$$

- The ITE δ can be estimated by:

$$\hat{\delta} := \underset{f \in \mathcal{F}}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\pi_{T_i}(\boldsymbol{X_i})} \{\boldsymbol{Y_i} - T_i \boldsymbol{f}(\boldsymbol{X_i}, \boldsymbol{t_i})/2\}' \boldsymbol{V_i}^{-1} \{\boldsymbol{Y_i} - T_i \boldsymbol{f}(\boldsymbol{X_i}, \boldsymbol{t_i})/2\}$$

  - For longitudinal data: $\boldsymbol{f}(\boldsymbol{X_i}, \boldsymbol{t_i}) = (f(\boldsymbol{X_i}, t_1), ..., f(\boldsymbol{X_i}, t_{k_i}))'$.
  - For clustered data: $\boldsymbol{f}(\boldsymbol{X_i}) = (f(\boldsymbol{X_i}), ..., f(\boldsymbol{X_i}))'$

New Method for Correlated Data

$$\hat{\delta} := \underset{f \in \mathcal{F}}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\pi_{T_i}(\boldsymbol{X_i})} \{\boldsymbol{Y_i} - T_i \boldsymbol{f}(\boldsymbol{X_i}, \boldsymbol{t_i})/2\}' \boldsymbol{V_i}^{-1} \{\boldsymbol{Y_i} - T_i \boldsymbol{f}(\boldsymbol{X_i}, \boldsymbol{t_i})/2\}$$

- In longitudinal data, one example is to apply AR(1) or other correlation structure for $\boldsymbol{V_i}$;
- In clustered data, one example is to use exchangeable correlation structure for $\boldsymbol{V_i}$;

- Our method can be also applied to multi-level data.

Consistency

Theorem 1: Consistency

Under the assumptions in causal inference framework with model (1), for the working model of propensity score $\hat{\pi}_1(\boldsymbol{x})$, if $\hat{\pi}_1(\boldsymbol{x}) = \pi_1(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathcal{X}$ almost surely, we have

$$\boldsymbol{\delta} \in \underset{f \in \mathcal{F}}{\operatorname{argmin}}\, \mathbb{E}\left[\frac{1}{\hat{\pi}_{T_i}(\boldsymbol{X_i})}\{\boldsymbol{Y_i} - T_i \boldsymbol{f}(\boldsymbol{X_i}, \boldsymbol{t_i})/2\}' \boldsymbol{V_i}^{-1}\{\boldsymbol{Y_i} - T_i \boldsymbol{f}(\boldsymbol{X_i}, \boldsymbol{t_i})/2\}\right]$$

- Even modeling of main effects is by-passed, the $\hat{\delta}$ is consistent if the propensity score is consistent.
  - There are often many covariates in main effects, but far fewer intervention-moderators that alter intervention effects
  - We model intervention-moderators only $\rightarrow$ robust to model mis-specification of main effect

Efficiency Augmentation

- As in Residual Weighted Learning for single outcome, the variance of $\hat{\delta}$ can be reduced when the outcome is replaced by augmented outcome $Y - a(X)$. *[Liu et al., 2018]*
- Following this idea in our method with augmented outcome $\boldsymbol{Y_i} - \boldsymbol{a}(\boldsymbol{X_i}, \boldsymbol{t_i})$, we can prove that the optimal augmentation (with smallest variance) is:

$$\boldsymbol{a}(\boldsymbol{X_i}, \boldsymbol{t_i}) = \boldsymbol{m}(\boldsymbol{X_i}, \boldsymbol{t_i}) + \{1 - 2\pi_1(\boldsymbol{X_i})\}\delta(\boldsymbol{X_i}, \boldsymbol{t_i})$$

Efficiency Augmentation

- As in Residual Weighted Learning for single outcome, the variance of $\hat{\delta}$ can be reduced when the outcome is replaced by augmented outcome $Y - a(X)$. *[Liu et al., 2018]*
- Following this idea in our method with augmented outcome $\boldsymbol{Y_i} - \boldsymbol{a}(\boldsymbol{X_i}, \boldsymbol{t_i})$, we can prove that the optimal augmentation (with smallest variance) is:

$$\boldsymbol{a}(\boldsymbol{X_i}, \boldsymbol{t_i}) = \boldsymbol{m}(\boldsymbol{X_i}, \boldsymbol{t_i}) + \{1 - 2\pi_1(\boldsymbol{X_i})\}\delta(\boldsymbol{X_i}, \boldsymbol{t_i})$$

- In randomized trial with $\pi_1(\boldsymbol{X_i}) = 0.5$, the optimal efficiency augmentation is

$$\boldsymbol{a}(\boldsymbol{X_i}, \boldsymbol{t_i}) = \boldsymbol{m}(\boldsymbol{X_i}, \boldsymbol{t_i})$$

Efficiency Augmentation

- As in Residual Weighted Learning for single outcome, the variance of $\hat{\delta}$ can be reduced when the outcome is replaced by augmented outcome $Y - a(X)$. *[Liu et al., 2018]*
- Following this idea in our method with augmented outcome $\boldsymbol{Y_i - a(X_i, t_i)}$, we can prove that the optimal augmentation (with smallest variance) is:

$$\boldsymbol{a(X_i, t_i) = m(X_i, t_i) + \{1 - 2\pi_1(X_i)\}\delta(X_i, t_i)}$$

- In observational study with sparse high-dimensional data:
  - We often expect the main effect is much larger than the interaction part (most covariates contributes to main effects m but not in δ)

  Thus, the optimal efficiency augmentation is approximated by

$$\boldsymbol{a(X_i, t_i) \approx m(X_i, t_i)}$$

Efficiency Augmentation

Main effect estimation **for efficiency augmentation**:

$$\hat{\boldsymbol{m}} := \underset{\boldsymbol{g}\in\mathcal{G}}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\pi_{T_i}(\boldsymbol{X_i}, \boldsymbol{t_i})} \{\boldsymbol{Y_i} - \boldsymbol{g}(\boldsymbol{X_i}, \boldsymbol{t_i})\}' \boldsymbol{V_i}^{-1} \{\boldsymbol{Y_i} - \boldsymbol{g}(\boldsymbol{X_i}, \boldsymbol{t_i})\}$$

- It uses all the data units all at once to estimate the main effect.
- It can be easily generalized to other regression methods or flexible models using machine learning techniques.
- If the propensity score is known, the main effect estimator is consistent if $\boldsymbol{m} \in \mathcal{G}$.
- If the propensity score is unknown, one can estimated it by simple logistic regression with all baseline covariates before intervention.
- After obtaining $\hat{\boldsymbol{m}}$, we can plug it in outcome augmentation to estimate δ

Augmented New Method for Correlated Data

- STEP 1: Estimate the propensity score model $\hat{\pi}_{T_i}(\boldsymbol{X_i})$ and main effect model $\hat{\boldsymbol{m}}(\boldsymbol{X_i}, \boldsymbol{t_i})$ for efficiency augmentation
- STEP 2: Estimate ITE model $\hat{\boldsymbol{\delta}}(\boldsymbol{X_i}, \boldsymbol{t_i})$ by minimizing the loss function:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\hat{\pi}_{T_i}(\boldsymbol{X_i})}\left[\{\boldsymbol{Y_i}-\hat{\boldsymbol{m}}(\boldsymbol{X_i},\boldsymbol{t_i})\}-T_i\boldsymbol{f}(\boldsymbol{X_i},\boldsymbol{t_i})/2\right]'\boldsymbol{V_i}^{-1}$$
$$\left[\{\boldsymbol{Y_i}-\hat{\boldsymbol{m}}(\boldsymbol{X_i},\boldsymbol{t_i})\}-T_i\boldsymbol{f}(\boldsymbol{X_i},\boldsymbol{t_i})/2\right]$$

Double Robustness

Theorem 2: Double Robustness

Under the assumptions in causal inference framework with model (1), for the working model of propensity score $\hat{\pi}_1(\boldsymbol{x})$ and main effect $\hat{\boldsymbol{m}}(\boldsymbol{x}, \boldsymbol{t})$, if either $\hat{\pi}_1(\boldsymbol{x}) = \pi_1(\boldsymbol{x})$ or $\hat{\boldsymbol{m}}(\boldsymbol{x}, \boldsymbol{t}) = \boldsymbol{m}(\boldsymbol{x}, \boldsymbol{t})$ for $\boldsymbol{x} \in \mathcal{X}$ and all $\boldsymbol{t}$ almost surely, we have

$$\boldsymbol{\delta} \in \underset{f \in \mathcal{F}}{\operatorname{argmin}} \mathbb{E}\left[\frac{1}{\hat{\pi}_{T_i}(\boldsymbol{X_i})}\{\boldsymbol{Y_i} - \hat{\boldsymbol{m}}(\boldsymbol{X_i}, \boldsymbol{t_i}) - T_i \boldsymbol{f}(\boldsymbol{X_i}, \boldsymbol{t_i})/2\}' \boldsymbol{V_i}^{-1} \right.$$
$$\left. \{\boldsymbol{Y_i} - \hat{\boldsymbol{m}}(\boldsymbol{X_i}, \boldsymbol{t_i}) - T_i \boldsymbol{f}(\boldsymbol{X_i}, \boldsymbol{t_i})/2\}\right]$$

- For randomized study, the proposed method always leads to consistent ITE even main effects is mis-specified
- For observational study, the proposed method double the chances to obtain consistent ITE

Implementation

- Directly optimize the function among all functional spaces is not feasible $\rightarrow$ Need assumptions on the function space $f \in \mathcal{F}$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\hat{\pi}_{T_i}(\boldsymbol{X_i})}\left[\{\boldsymbol{Y_i}-\hat{\boldsymbol{m}}(\boldsymbol{X_i},\boldsymbol{t_i})\}-T_i\boldsymbol{f}(\boldsymbol{X_i},\boldsymbol{t_i})/2\right]'\boldsymbol{V_i}^{-1}$$
$$\left[\{\boldsymbol{Y_i}-\hat{\boldsymbol{m}}(\boldsymbol{X_i},\boldsymbol{t_i})\}-T_i\boldsymbol{f}(\boldsymbol{X_i},\boldsymbol{t_i})/2\right]$$

Implementation

- Linear case: $f_{\text{lin}}(\boldsymbol{X_i}, t_j) = \tilde{\boldsymbol{X}}'_{\boldsymbol{ij}}\boldsymbol{\beta}$ where $\boldsymbol{\beta} = (\beta_0, \beta_T, \beta_1, ..., \beta_p)'$ and $\tilde{\boldsymbol{X}}_{\boldsymbol{ij}} = (1, t_j, \boldsymbol{X_i})'$, then the loss function $L_{\text{lin}}(\boldsymbol{\beta})$ is

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\hat{\pi}_{T_i}(\boldsymbol{X_i})}\left[\{\boldsymbol{Y_i} - \hat{\boldsymbol{m}}(\boldsymbol{X_i}, \boldsymbol{t_i})\} - \{(T_i\tilde{\boldsymbol{X}}_{\boldsymbol{i}}/2)'\boldsymbol{\beta}\}\right]' \boldsymbol{V_i}^{-1}$$
$$\left[\{\boldsymbol{Y_i} - \hat{\boldsymbol{m}}(\boldsymbol{X_i}, \boldsymbol{t_i})\} - \{(T_i\tilde{\boldsymbol{X}}_{\boldsymbol{i}}/2)'\boldsymbol{\beta}\}\right]$$

  where $\tilde{\boldsymbol{X}}_{\boldsymbol{i}} = (\tilde{\boldsymbol{X}}_{\boldsymbol{i1}}, ..., \tilde{\boldsymbol{X}}_{\boldsymbol{ik_i}})$

  - The minimization can be implemented within linear mixed model or GEE method by specifying the correlation structure $\boldsymbol{V_i}$.

- Non-Linear case: $f_{\text{non}}(\boldsymbol{X_i}, t_j) = \beta_0 + \beta_T t_j + \sum_{q=1}^{p} B(X_{i,q})\beta_q$ where $B(.)$ is the B-spline based function in the additive model

Regularization in High-dimensional Data

- For high-dimensional data:
  - The number of covariates is large.
  - Often we expect only a small subset of the features is associated with the subgroup identification (ie, intervention-moderators).
- We can add Lasso penalty *[Tibshirani et al., 1996]* in our loss function, e.g.

$$L^*_{\text{lin}}(\boldsymbol{\beta}) = L_{\text{lin}}(\boldsymbol{\beta}) + \lambda||\boldsymbol{\beta}||_1$$

  where $||\boldsymbol{\beta}||_1 = |\beta_T| + \sum_{i=1}^{p} |\beta_i|$ and the tuning parameter $\lambda > 0$.
- Different regularization method is also applicable in our framework, but Lasso has better interpretation in application.

Simulation Study

Data Generation

For longitudinal data $\{(Y_{ij}, T_i, \boldsymbol{X}_i), i = 1, ..., n; j = 1, ..., K\}$ with baseline covariates only and observed time $\{t_j = j : j = 1, ..., K\}$, the continuous response was generated by:

$$Y_{ij} = m(\boldsymbol{X}_i, t_j) + T_i\delta(\boldsymbol{X}_i, t_j)/2 + \alpha_i + e_{ij}$$

with random intercept $\alpha_i \sim N(0, \sigma_\alpha^2 = 1)$ and iid $e_{ij} \sim N(0, \sigma_e^2 = 1)$

- Treatment: $T_i \in \{1, -1\}$ by Bernoulli(0.5)
- Estimating δ in the training set with $n = 100, K = 5$
- Evaluation in the independent testing set with $n_t = 10000, K = 5$
- Number of simulation replications $N = 500$

Data Generation

$$(1, \overbrace{X_{i,1}, \ldots, X_{i,5}}^{\textit{Continuous}}, \overbrace{X_{i,6}, \ldots, X_{i,10}}^{\textit{Binary}}, \overbrace{X_{i,11}, \ldots, X_{i,20}}^{\textit{Continuous}}, \overbrace{X_{i,21}, \ldots, X_{i,30}}^{\textit{Binary}})'$$

Not included in the outcome model

- 15 Continuous covariates: $(X_{i,1}, ..., X_{i,5}, X_{i,11}, ..., X_{i,20}) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_X)$

$$\mathbf{\Sigma}_X = \begin{pmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{14} \\ \rho & 1 & \rho & \ldots & \rho^{13} \\ \rho^2 & \rho & 1 & \ldots & \rho^{12} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \rho^{14} & \rho^{13} & \rho^{12} & \ldots & 1 \end{pmatrix}$$

  where $\rho = 0$ for independent case and $\rho = 0.6$ for correlated case.
- 15 Binary covariates: $(X_{i,6}, ..., X_{i,10}, X_{i,21}, ..., X_{i,30}) \sim$ Bernoulli(0.5)

Scenario Settings

Scenario 1: the validation of the new methods

- The response is generated by

$$
\begin{aligned}
Y_{ij} = \beta_0 + \beta_T t_j + \sum_{q=1}^{10} \beta_q X_{i,q} \\
+ T_i \left(\gamma_0 + \gamma_T t_j + \sum_{q=1,2,8,10} \gamma_q X_{i,q} \right) /2 \\
+ \alpha_i + \epsilon_{ij}
\end{aligned}
$$

Scenario Settings

Scenario 2: the robustness against mis-specification of main effect

- Other data generation process is the same with scenario 1, except

$$Y_{ij} = \beta_0 + \beta_T t_j + \sum_{q=1}^{10} \beta_q X_{i,q}^2 + \sum_{q=1}^{10} \cos(\beta_q X_{i,q}) + T_i \left(\gamma_0 + \gamma_T t_j + \sum_{q=1,2,8,10} \gamma_q X_{i,q} \right) /2 + \alpha_i + \epsilon_{ij}$$

- Main effect is mis-specified if using linear model

Scenario Settings

Scenario 3: the robustness against mis-specification of propensity score

- The treatment assignment is generated by the propensity score model:

$$Pr(T_i = 1|X) = \frac{2}{2 + \exp(X_1 + X_6 + X_7)}$$

- Propensity score is mis-specified if assuming randomized intervention with $\hat{\pi}_{T_i}(\boldsymbol{X}_i) = 0.5$
- Compare the results in both linear and non-linear main effect cases

Scenario Settings

For all scenarios, the parameters are:

- Interaction effects: $(\gamma_1, \gamma_2, \gamma_8, \gamma_{10}) = (8, -8, 8, -8)$; $\gamma_T = 2, \gamma_0 = 2$
- **Small main effect**: $\beta_T = 0.1$ and

$$(\beta_0, ..., \beta_{10}) = (0.3, 0.5, 0.4, 0.6, -0.3, -0.6, 0.3, 0.1, -0.2, -0.1, 0.2)$$

- **Big main effect**: $\beta_T = 0.4$ and

$$(\beta_0, ..., \beta_{10}) = (1.2, 2, 1.6, 2.4, -1.2, -2.4, 1.2, 0.4, -0.8, -0.4, 0.8)$$

Estimation Methods

- Model 1: Full Mixed Effect Model with Lasso penalty and exchangeable correlation structure.
- Model 2: Huling's Method using square loss with fused lasso in time-varying coefficients.
- Model 3: New Method with Lasso penalty and exchangeable correlation structure.

Estimation Methods

- Model 1: Full Mixed Effect Model with Lasso penalty and exchangeable correlation structure.
- Model 2: Huling's Method using square loss with fused lasso in time-varying coefficients.
- Model 3: New Method with Lasso penalty and exchangeable correlation structure.

The following statistics are obtained for Model h

- ITE over time for i-th subject:

$$\hat{\boldsymbol{\delta}}_{\boldsymbol{h}}(\boldsymbol{X_i}, \boldsymbol{t_i}) = (\hat{\delta}_h(\boldsymbol{X_i}, t_1), ..., \hat{\delta}_h(\boldsymbol{X_i}, t_K))'$$

- Time-average ITE for i-th subject: $\bar{\delta}_h(\boldsymbol{X_i}) = \frac{1}{K}\sum_{j=1}^{K} \hat{\delta}_h(\boldsymbol{X_i}, t_j)$

Model Evaluation in Independent Testing Data

- Accuracy of subgroup identification for model h:

$$ACC_h = \frac{1}{n_t} \sum_{i=1}^{n_t} I\left\{\text{sign}\{\bar{\delta}_h(\boldsymbol{X_i})\} = \text{sign}\{\bar{\delta}_0(\boldsymbol{X_i})\}\right\}$$

$\bar{\delta}_0(\boldsymbol{X_i}) = \frac{1}{K}\sum_{j=1}^{K} \delta(\boldsymbol{X_i}, t_j)$: true time-average ITE of i-th subject.

Model Evaluation in Independent Testing Data

- Accuracy of subgroup identification for model h:

$$ACC_h = \frac{1}{n_t} \sum_{i=1}^{n_t} I\left\{\text{sign}\{\bar{\delta}_h(\boldsymbol{X_i})\} = \text{sign}\{\bar{\delta}_0(\boldsymbol{X_i})\}\right\}$$

$\bar{\delta}_0(\boldsymbol{X_i}) = \frac{1}{K}\sum_{j=1}^{K} \delta(\boldsymbol{X_i}, t_j)$: true time-average ITE of i-th subject.

- Spearman's rank correlation coefficient (denoted by SCC_h for model h) between true time-average ITE and estimated time-average ITE.
  - To compare the ability of recovering the rank of time-average ITE.

Model Evaluation in Independent Testing Data

- Accuracy of subgroup identification for model h:

$$ACC_h = \frac{1}{n_t} \sum_{i=1}^{n_t} I\left\{\text{sign}\{\bar{\delta}_h(\boldsymbol{X_i})\} = \text{sign}\{\bar{\delta}_0(\boldsymbol{X_i})\}\right\}$$

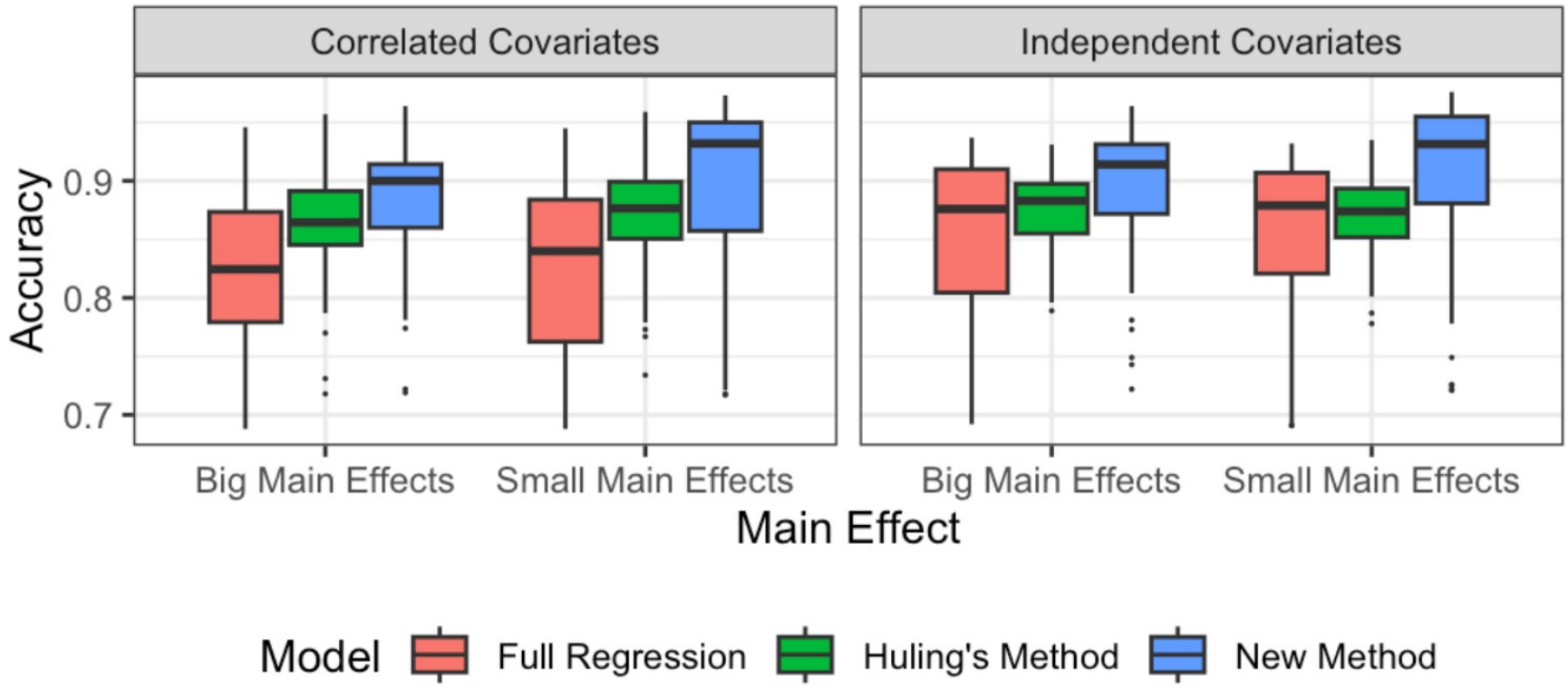
$\bar{\delta}_0(\boldsymbol{X_i}) = \frac{1}{K}\sum_{j=1}^{K} \delta(\boldsymbol{X_i}, t_j)$: true time-average ITE of i-th subject.

- Spearman's rank correlation coefficient (denoted by SCC_h for model h) between true time-average ITE and estimated time-average ITE.
  - To compare the ability of recovering the rank of time-average ITE.
- Average prediction error for model h:

$$APE_h = \frac{1}{n_t} \sum_{i=1}^{n_t} ||\hat{\boldsymbol{\delta}}_h(\boldsymbol{X_i}, \boldsymbol{t_i}) - \boldsymbol{\delta}_0(\boldsymbol{X_i}, \boldsymbol{t_i})||_2$$

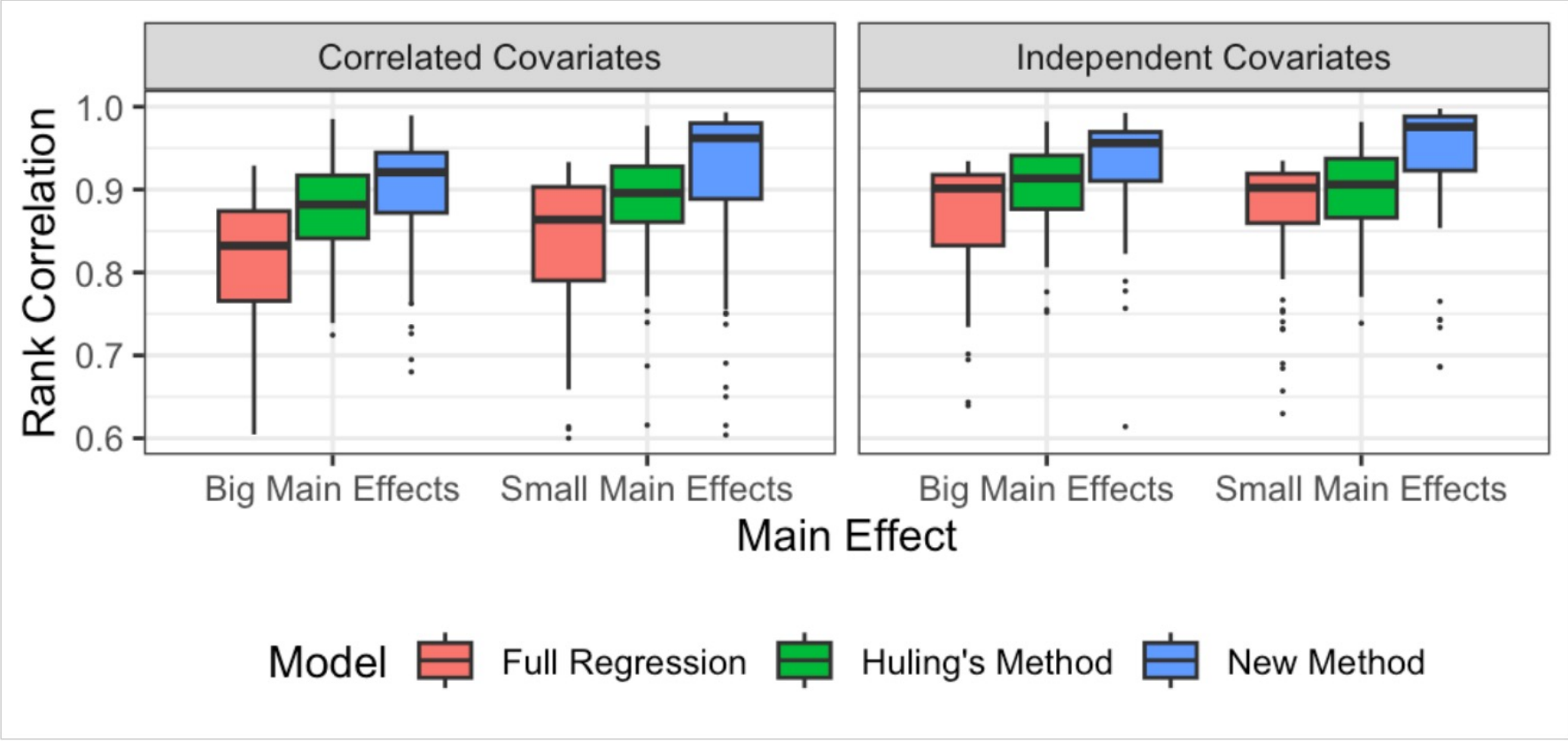
where $\boldsymbol{\delta}_0(\boldsymbol{X_i}, \boldsymbol{t_i}) = (\delta(X_i, t_1), ..., \delta(X_i, t_K))'$ and $||.||_2$ is L_2 norm.

Scenario 1: Estimation Performance



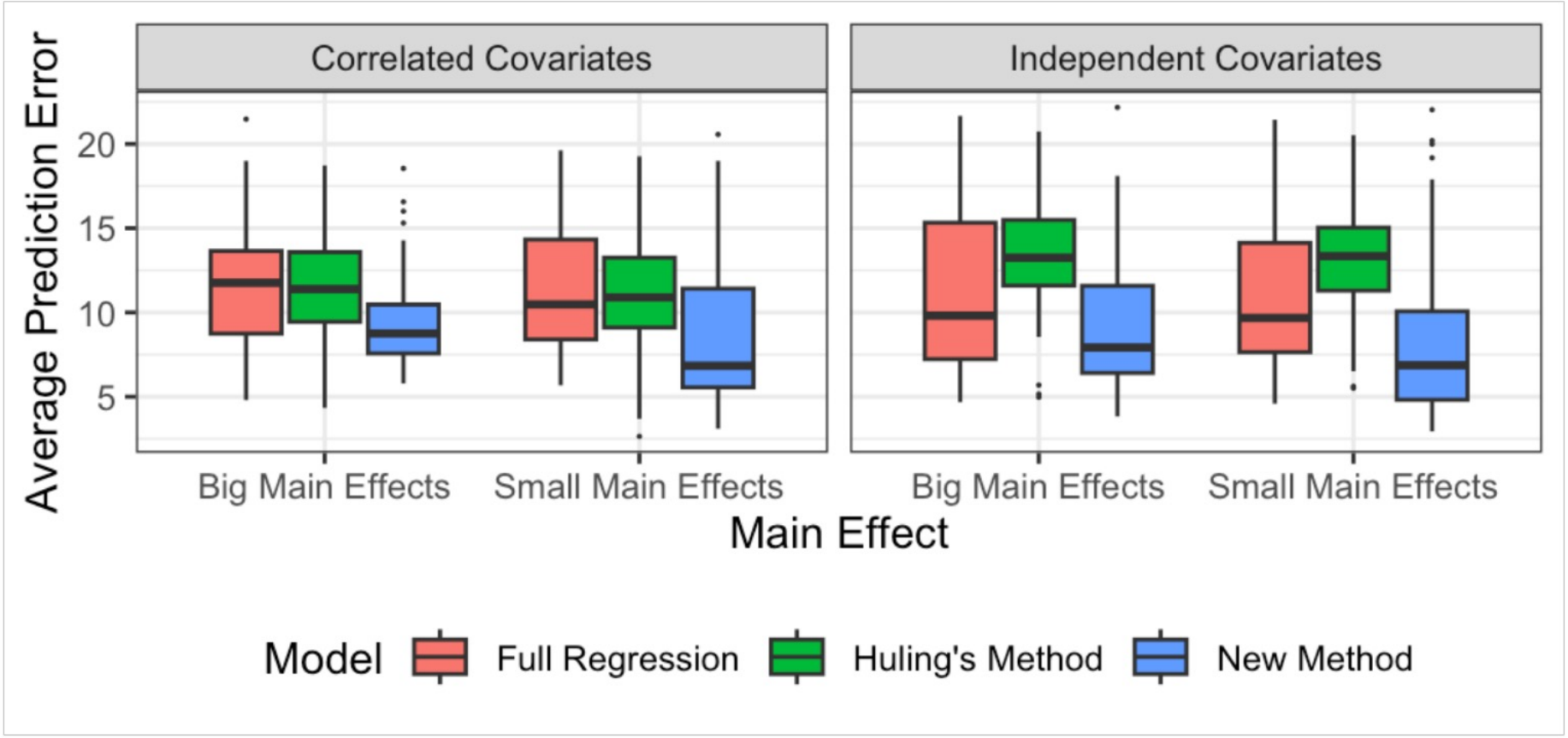
- New method can identify subgroups more precisely.

Scenario 1: Estimation Performance



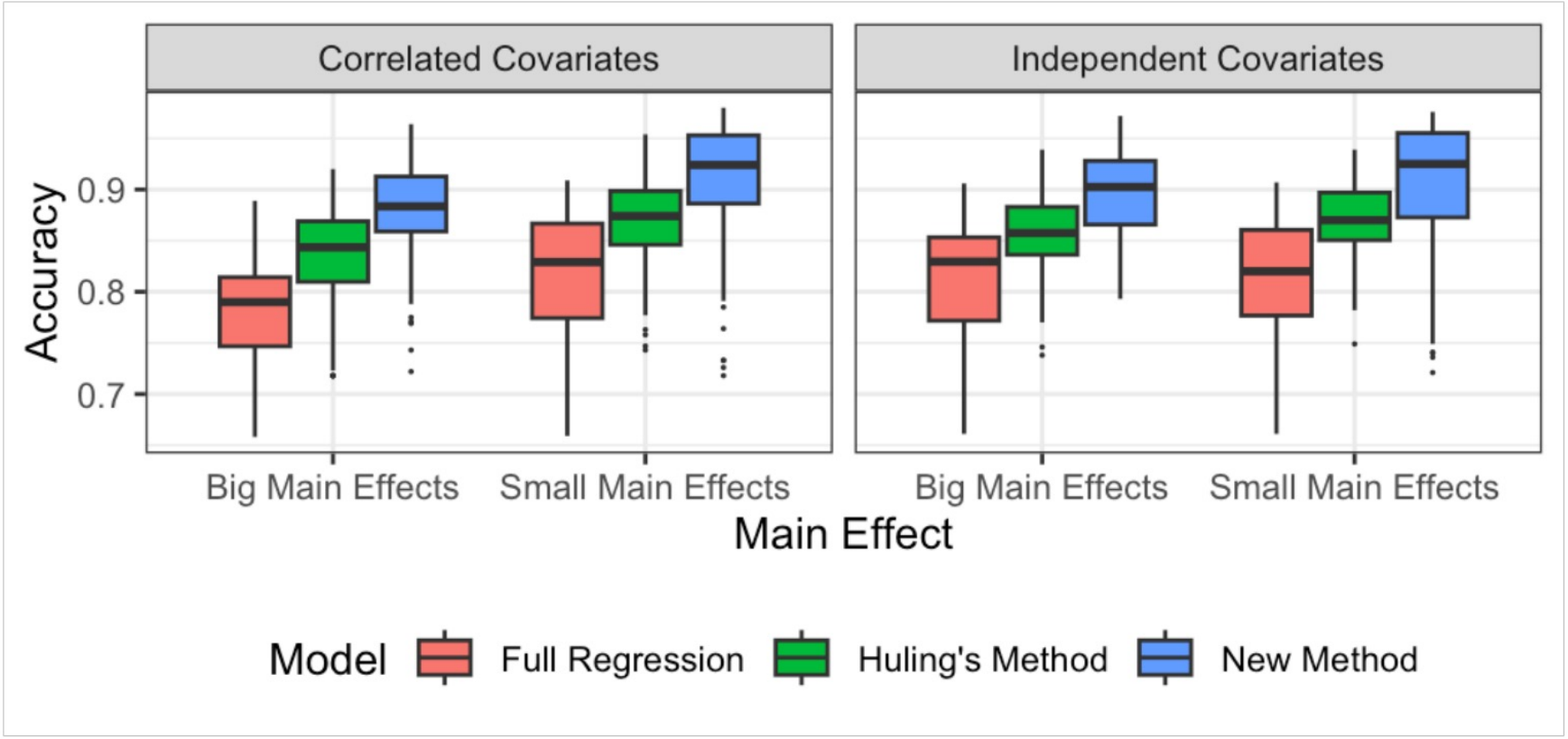
- New method can recover the rank of individualized treatment effects better.

Scenario 1: Estimation Performance



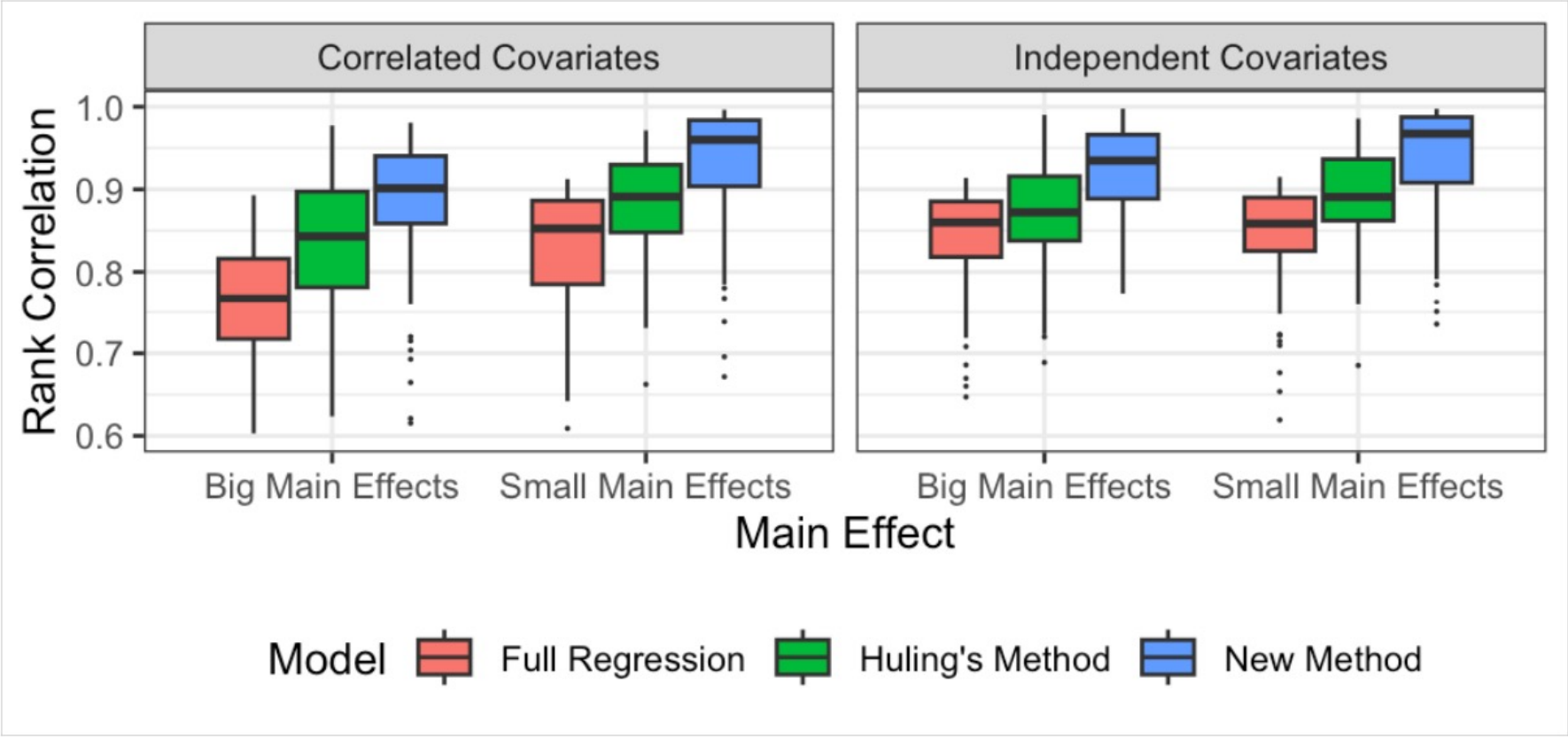
- New method can predict individualized treatment effects more precisely.

Scenario 2: Double Robustness to Mis-specification of Main Effect



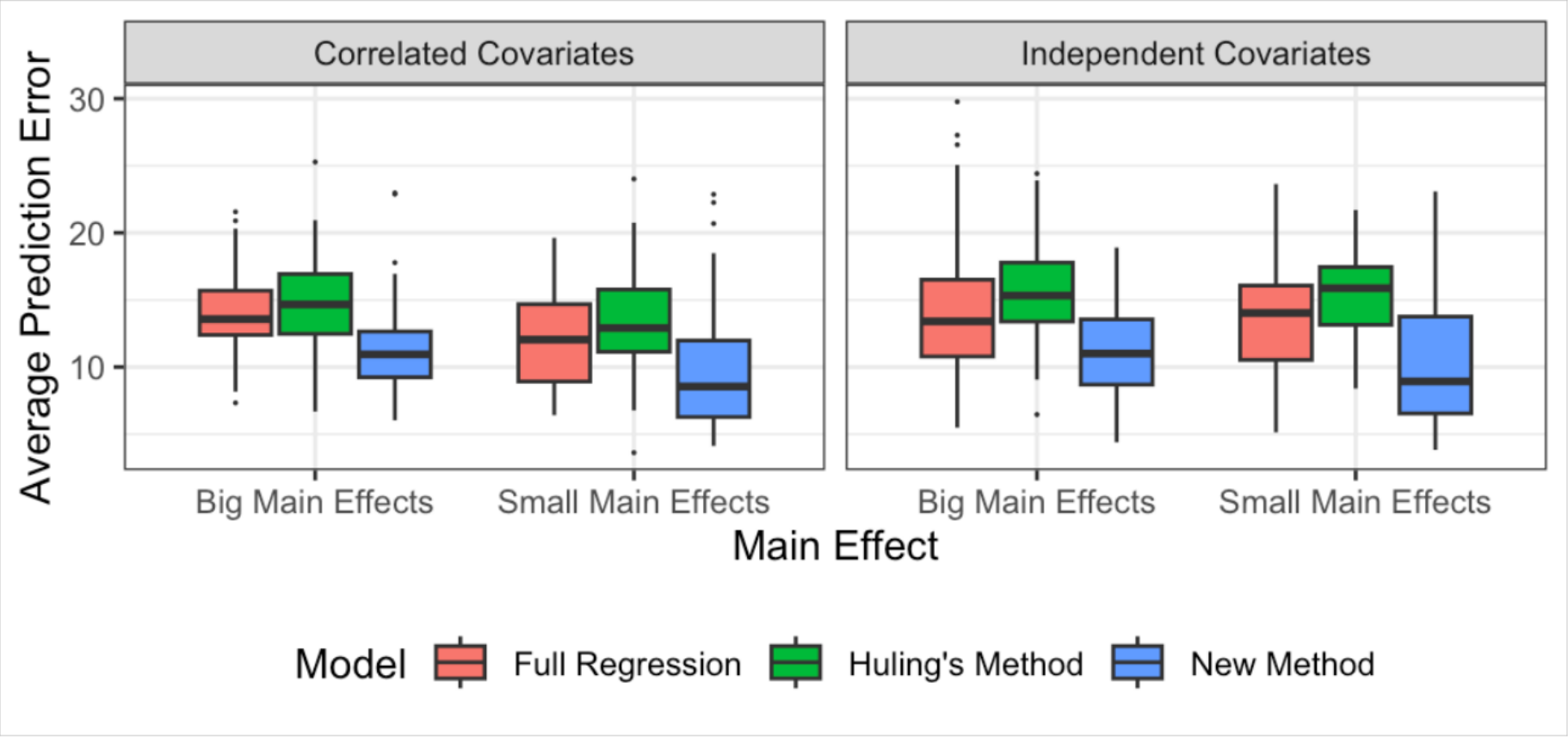
- Full regression model requires correctly specified main effects, leading to worse performance due to mis-specified main effects
- New method can identify subgroups precisely even if the main effect is mis-specified.

Scenario 2: Double Robustness to Mis-specification of Main Effect



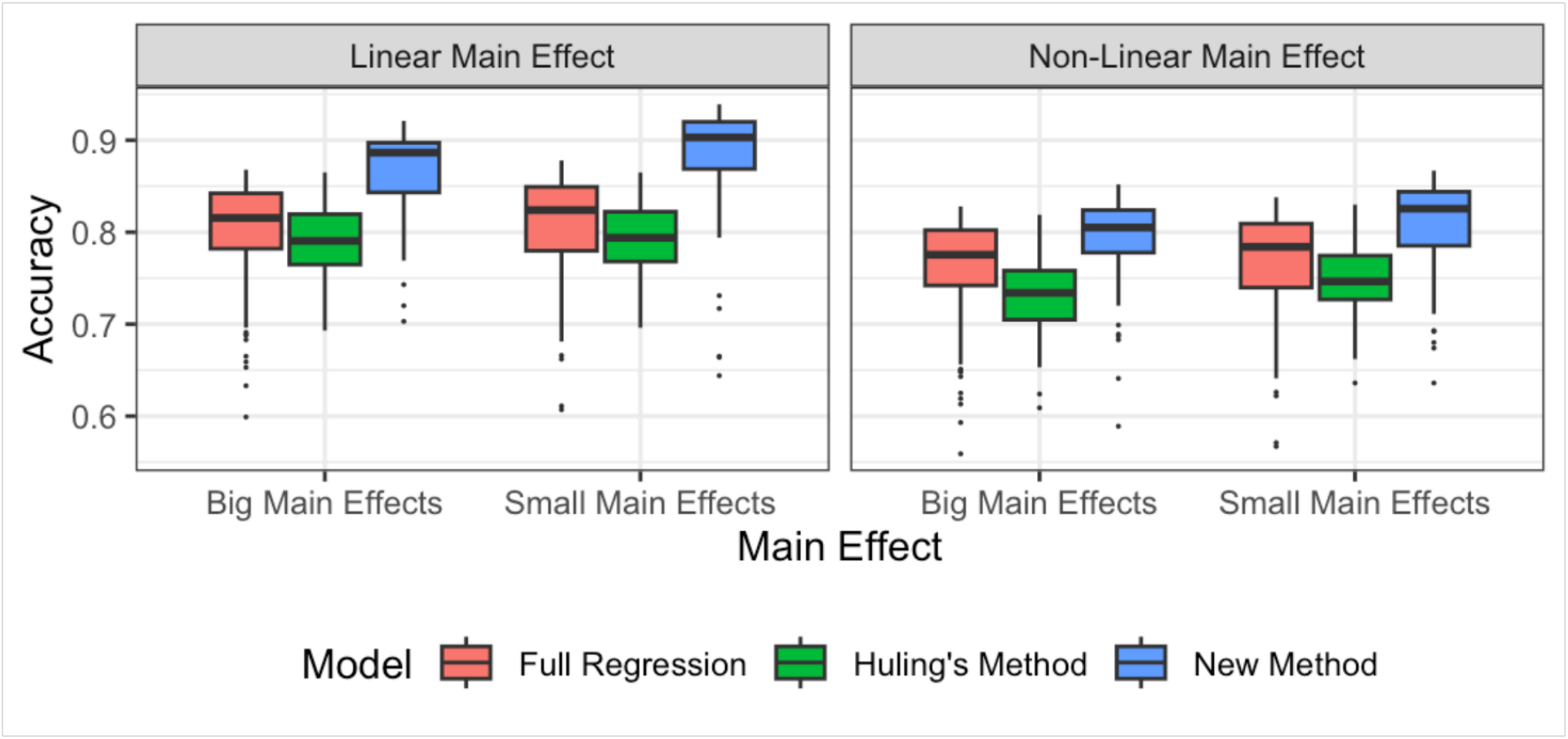
- The ability of recovering the rank is consistent with the accuracy of subgroup identification.

Scenario 2: Double Robustness to Mis-specification of Main Effect



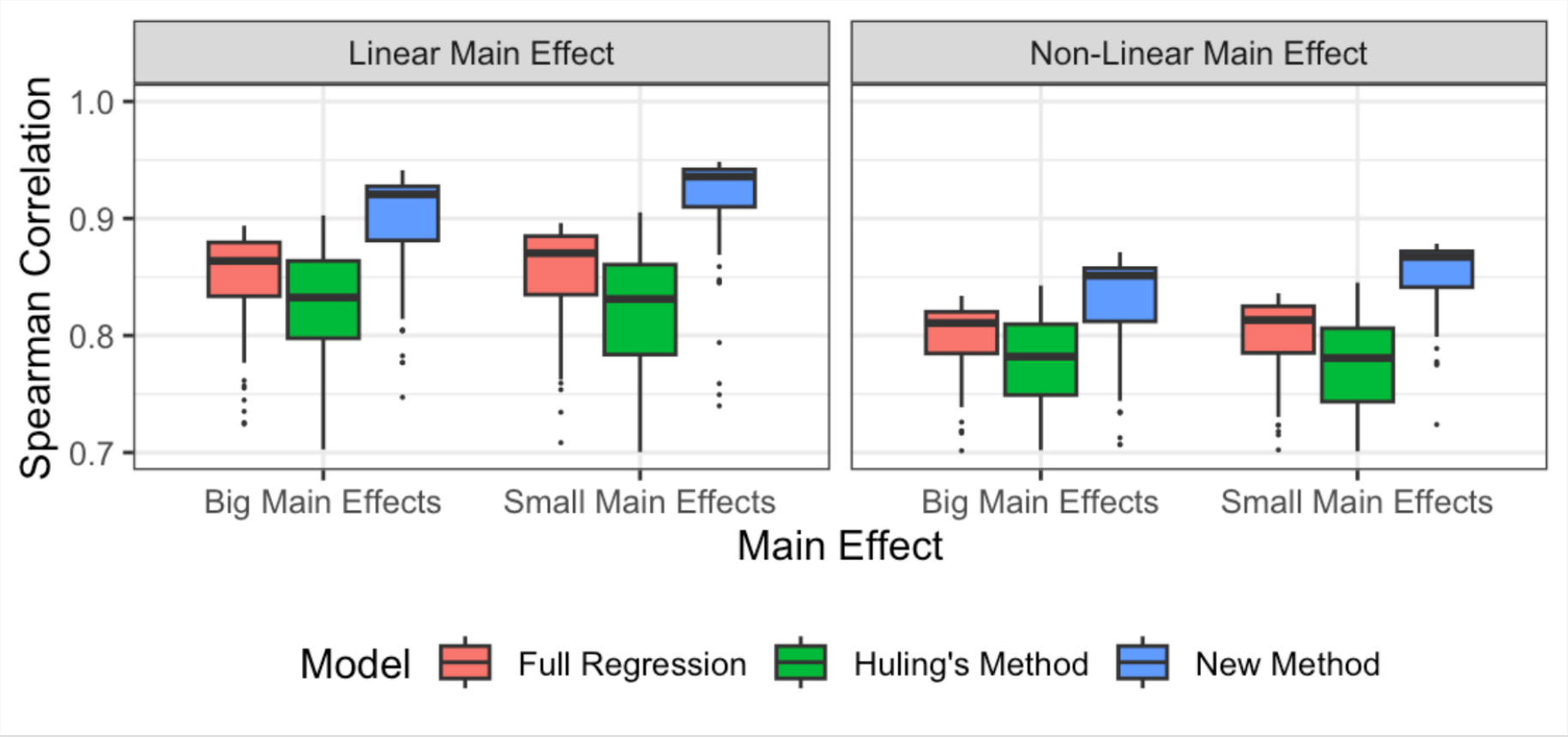
- New method can predict individualized treatment effects precisely even if the main effect is mis-specified.

Scenario 3: Double Robustness to Mis-specification of Propensity Score



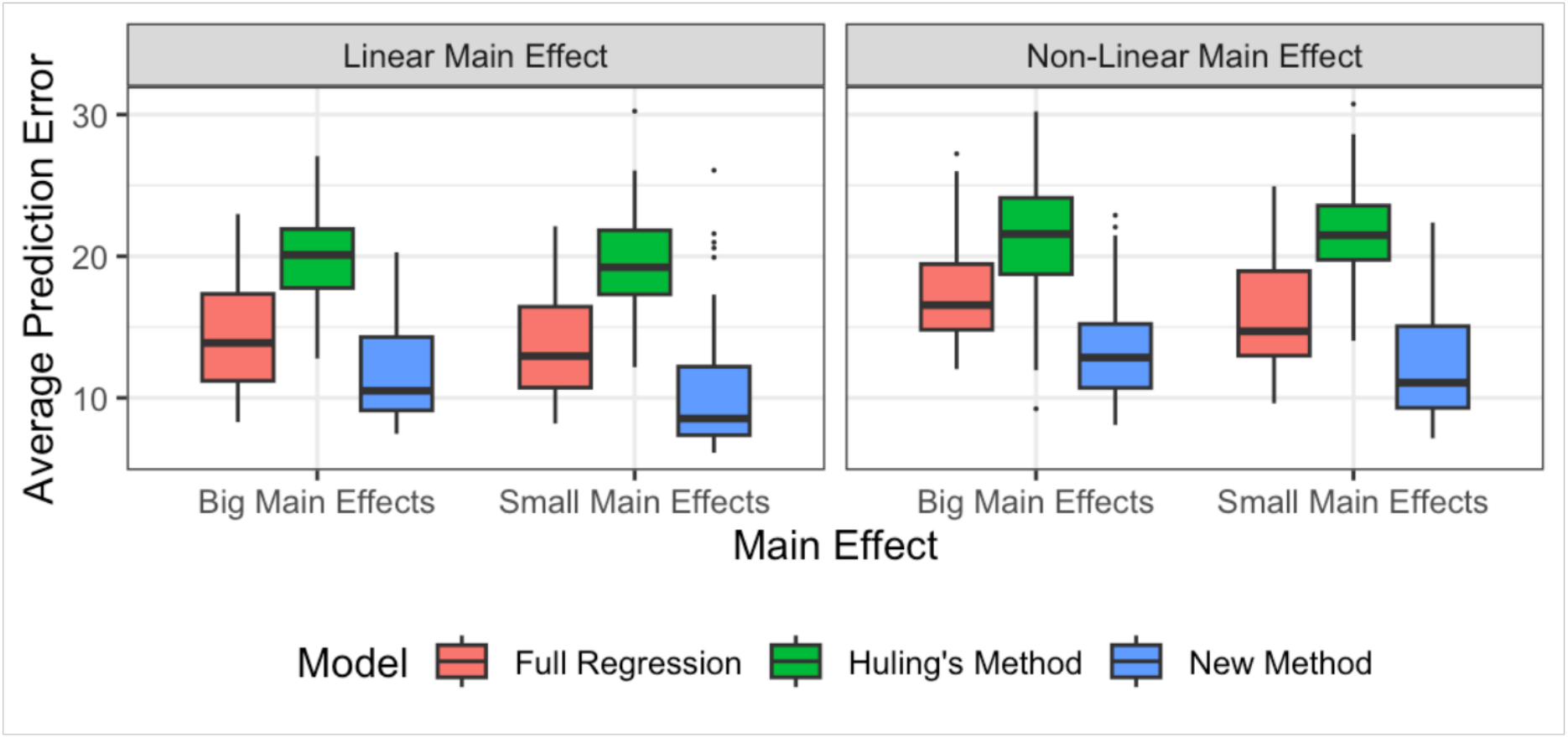
- Huling's method requires correct propensity score model, leading to worse performance.
- New method can identify subgroups precisely if only the propensity score is mis-specified.

Scenario 3: Double Robustness to Mis-specification of Propensity Score



- The ability of recovering the rank is consistent with the accuracy of ITR.

Scenario 3: Double Robustness to Mis-specification of Propensity Score



- New method can predict individualized treatment effects precisely if only the propensity score is mis-specified.

Additional Simulation Results

- Simulations for clustered data showed similar findings for the three scenarios.
  - For non-linear ITE:

$$
\begin{aligned}
Y_{ij} &= \beta_0 + \beta_T t_j + \sum_{q=1}^{10} \beta_q X_{i,q} \\
&+ T_i \left(\gamma_0 + \gamma_T t_j + \sum_{q=1,2,8,10} \gamma_q X_{i,q} + 2X_{i,8}^2 - 4X_{i,10}^3 \right) /2 \\
&+ \alpha_i + \epsilon_{ij}
\end{aligned}
$$

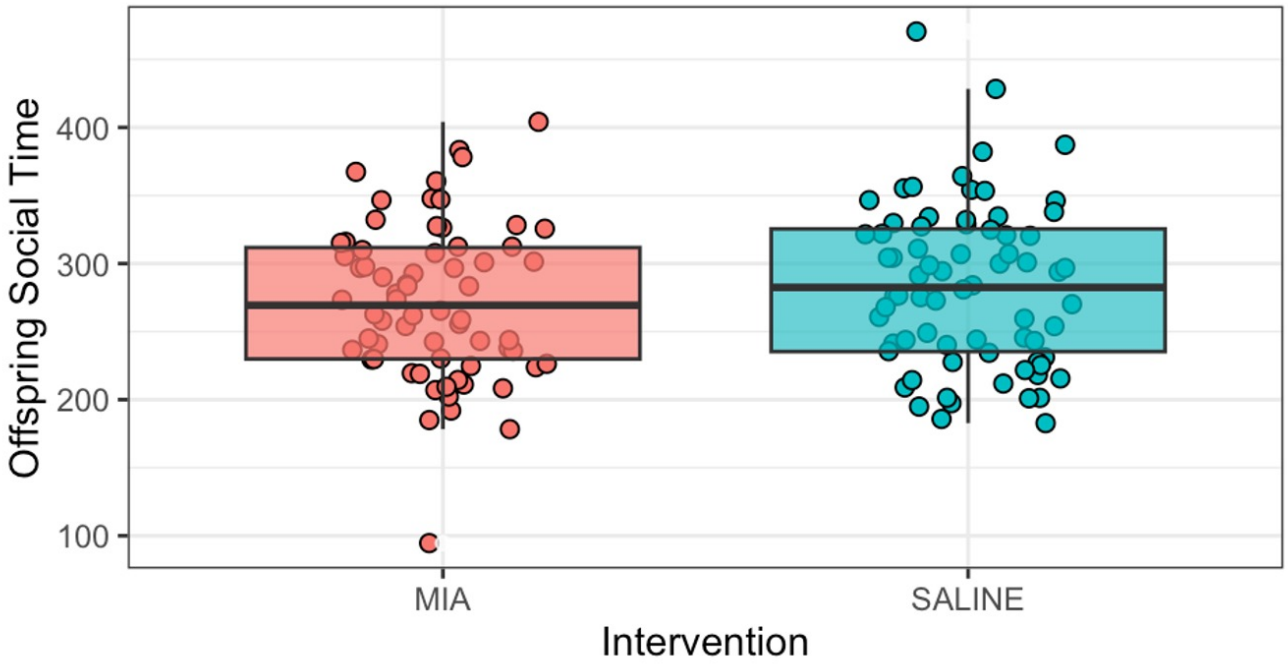
- Simulations show that the proposed method with B-spline based additive model performs better.

Real Data Analysis: MIA Study

MIA Study

- This is a randomized study in rat by the UC Davis Conte Center for studying effects of maternal immune activation on brain, behavior, and other development in offspring.
- Binary interventions at each mother
  - MIA: inject 50 LPS in dam to induce MIA
  - Saline: Control group
- Sample size: 138 offspring from 21 dams (9 MIA vs. 12 Saline)
- Outcome: offspring social investigation time
- Covariates: 13 cytokines for each mother before intervention.

MIA Study



- Average intervention effect for entire population is not significant.
- How about **individualized MIA effect**?
  - potential MIA-resilient group and MIA-susceptible group?

MIA Study

- Model Comparison by 100 random splits for 50% training set and 50% testing set at dam-level:
  - Method 1: Traditional full linear mixed model
  - Method 2: New method without main effect estimation
  - Method 3: New method with efficiency augmentation
- All methods use Lasso penalty to select variables (tuning parameter chosen by least MSE).

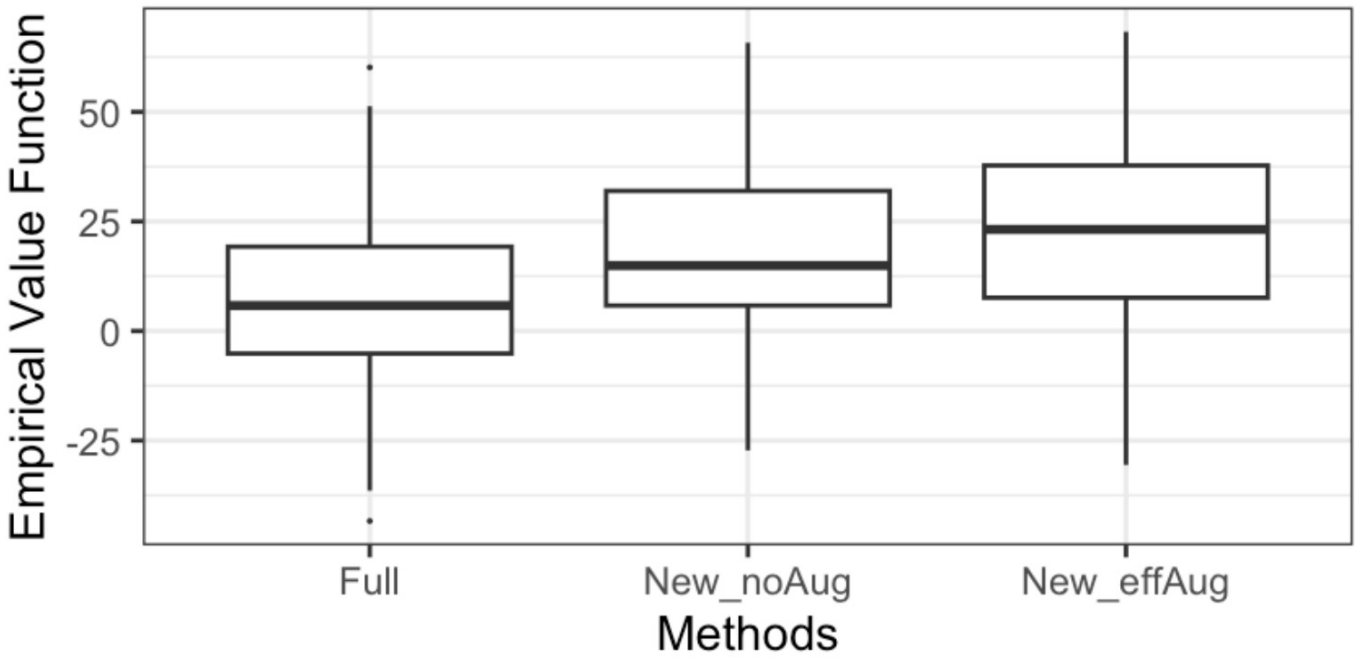
MIA Study

- Model Comparison by 100 random splits for 50% training set and 50% testing set at dam-level:
  - Method 1: Traditional full linear mixed model
  - Method 2: New method without main effect estimation
  - Method 3: New method with efficiency augmentation
- All methods use Lasso penalty to select variables (tuning parameter chosen by least MSE).
- Model can be evaluated by **Empirical Value Function** under :

$$\text{EVF} := E[Y_{ij}|\hat{D}(\boldsymbol{X_i}) = T_i] - E[Y_{ij}|\hat{D}(\boldsymbol{X_i}) \neq T_i]$$

where $\hat{D}(\boldsymbol{X_i}) := \text{sign}(\hat{\delta}(\boldsymbol{X_i}))$. The higher the EVF, the better the model to differentiate the subgroups.

MIA Study



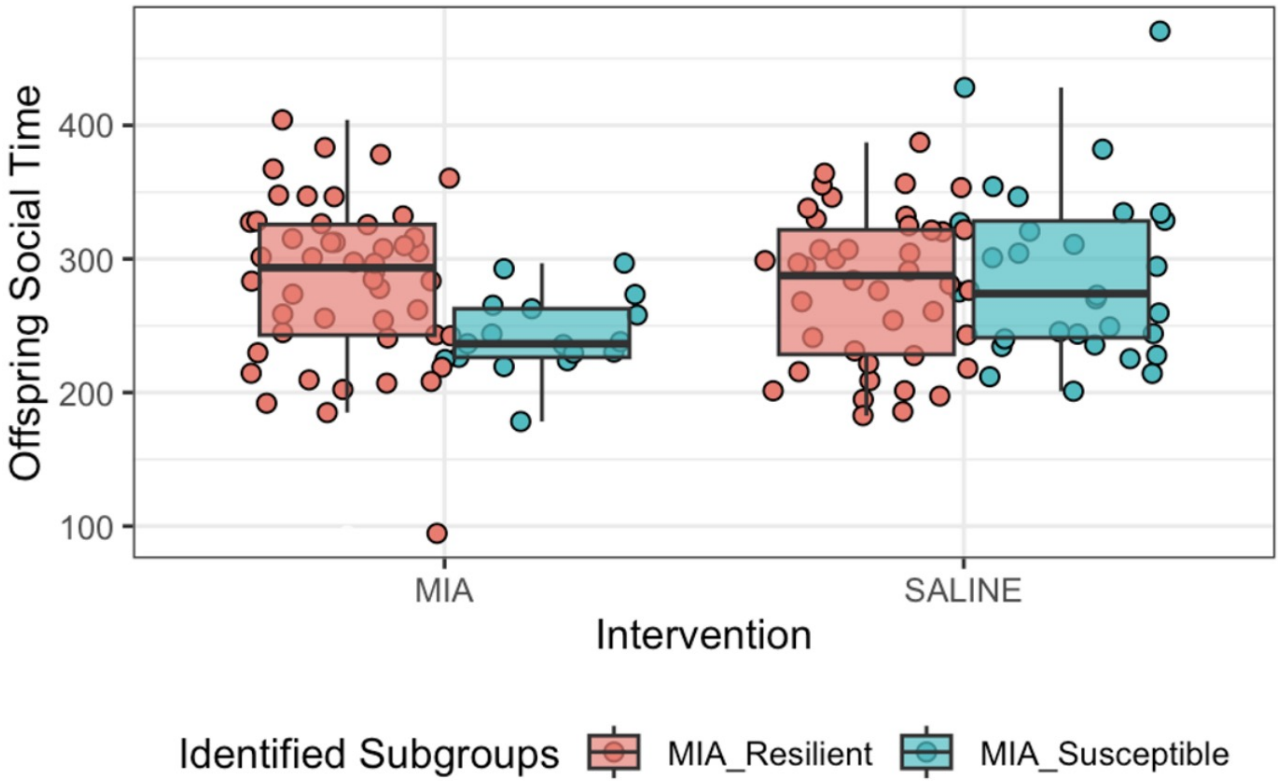
- New method with efficiency augmentation yields the largest value, which means subgroups can be differentiate better based on the ITE estimated by our method.

MIA Study

- Apply the proposed new method with augmentation, which selected 4 biomarkers for predicting individualized MIA effects.
- Mothers with high level baseline (pre-intervention) of GM-CSF and IL-1α, low level of IFN-γ and IL-5, are more susceptible to the effect of maternal immune activation. (ie, MIA lowers social time compared to control among their offspring)

Variable	GM-CSF	IFN-γ	IL-1α	IL-5
Coefficients	-0.45	0.29	-1.65	0.63

MIA Study



	MIA-Resilient Group	MIA-Susceptible Group
MIA	49 Offspring from 9 Dams	17 Offspring from 3 Dams
SALINE	42 Offspring from 5 Dams	30 Offspring from 4 Dams

Conclusion and Discussion

- New method can identify subgroups and predict individualized treatment effects more precisely than existing methods.
- New method shows doubly robust property with respect to main effect and propensity score mis-specification.
  - For randomized study, the proposed method always leads to consistent ITE even main effects is mis-specified
  - For observational study, the proposed method double the chances to obtain consistent ITE
- Allow regularization approach to handle high-dimensional data
- Allow flexible modeling of ITE using flexible function space or machine learning techniques

Future Work

- Extension to multiple treatments case
  - e.g. incorporated with the angle-based method *[Qi et al., 2020]*
- Extension to different types of outcome
  - e.g. binary outcome(with different loss function)
- Extension to involving post-MIA characteristics in identifying subgroups
  - e.g. following the idea of *[Barbosa et al., 2020]*
- Application with flexible function space to predict complicated ITE
  - e.g. more machine learning techniques (random forest, etc.) and semi-parametric method as in *[Liang et al., 2022]*
- Application to more MIA datasets and other real data examples

Reference

[1] Bankole A Johnson, Chamindi Seneviratne, Xin-Qun Wang, Nassima Ait-Daoud, and Ming D Li. Determination of genotype combinations that can predict the outcome of the treatment of alcohol dependence using the 5-ht3 antagonist ondansetron. *American Journal of Psychiatry*, 170(9):1020–1031, 2013.

[2] Laura H Goetz and Nicholas J Schork. Personalized medicine: motivation, challenges, and progress. *Fertility and sterility*, 109(6):952–963, 2018.

[3] Noura S Abul-Husn and Eimear E Kenny. Personalized medicine and the power of electronic health records. *Cell*, 177(1):58–69, 2019.

[4] Min Qian and Susan A Murphy. Performance guarantees for individualized treatment rules. *Annals of statistics*, 39(2):1180, 2011.

[5] Susan A Murphy. Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):331–355, 2003.

[6] Yingqi Zhao, Donglin Zeng, A John Rush, and Michael R Kosorok. Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association*, 107(499):1106–1118, 2012.

[7] Tianxi Cai, Lihui Tian, Hajime Uno, Scott D Solomon, and LJ Wei. Calibrating parametric subject-specific risk estimation. *Biometrika*, 97(2):389–404, 2010.

[8] James Edward Signorovitch. *Identifying informative biological markers in high-dimensional genomic data and clinical trials*. PhD thesis, Harvard University, 2007.

[9] Lu Tian, Ash A Alizadeh, Andrew J Gentles, and Robert Tibshirani. A simple method for estimating interactions between a treatment and a large number of covariates. *Journal of the American Statistical Association*, 109(508):1517–1532, 2014.

Reference

[10] Xin Zhou, Nicole Mayer-Hamblett, Umer Khan, and Michael R Kosorok. Residual weighted learning for estimating individualized treatment rules. *Journal of the American Statistical Association*, 112(517):169–187, 2017.

[11] Shuai Chen, Lu Tian, Tianxi Cai, and Menggang Yu. A general statistical framework for subgroup identification and comparative treatment scoring. *Biometrics*, 73(4):1199–1209, 2017.

[12] Marjolein Fokkema, Niels Smits, Achim Zeileis, Torsten Hothorn, and Henk Kelderman. Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees. *Behavior research methods*, 50:2016–2034, 2018.

[13] Yishu Wei, Lei Liu, Xiaogang Su, Lihui Zhao, and Hongmei Jiang. Precision medicine: Subgroup identification in longitudinal trajectories. *Statistical methods in medical research*, 29(9):2603–2616, 2020.

[14] Wei-Yin Loh, Luxi Cao, and Peigen Zhou. Subgroup identification for precision medicine: A comparative review of 13 methods. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(5):e1326, 2019.

[15] Hyunkeun Cho, Peng Wang, and Annie Qu. Personalize treatment for longitudinal data using unspecified random-effects model. *Statistica Sinica*, pages 187–206, 2017.

[16] Nichole Andrews and Hyunkeun Cho. Validating effectiveness of subgroup identification for longitudinal data. *Statistics in medicine*, 37(1):98–106, 2018.

[17] Juan Shen and Annie Qu. Subgroup analysis based on structured mixed-effects models for longitudinal data. *Journal of Biopharmaceutical Statistics*, 30(4):607–622, 2020.

[18] Francisco J Diaz. Measuring the individual benefit of a medical or behavioral treatment using generalized linear mixed-effects models. *Statistics in medicine*, 35(23):4077–4092, 2016.

Reference

[19] Mu Yue and Lei Huang. A new approach of subgroup identification for high-dimensional longitudinal data. *Journal of Statistical Computation and Simulation*, 90(11):2098–2116, 2020.

[20] Roza M Vlasova, Ana-Maria Iosif, Amy M Ryan, Lucy H Funk, Takeshi Murai, Shuai Chen, Tyler A Lesh, Douglas J Rowland, Jeffrey Bennett, Casey E Hogrefe, et al. Maternal immune activation during pregnancy alters postnatal brain growth and cognitive development in nonhuman primate offspring. *Journal of Neuroscience*, 41(48):9971–9987, 2021.

[21] Urs Meyer. Neurodevelopmental resilience and susceptibility to maternal immune activation. *Trends in neurosciences*, 42(11):793–806, 2019.

[22] Tianxi Cai, Lu Tian, Peggy H Wong, and LJ Wei. Analysis of randomized comparative clinical trial data for personalized treatment selections. *Biostatistics*, 12(2):270–282, 2011.

[23] Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

[24] Haomiao Meng and Xingye Qiao. Augmented direct learning for conditional average treatment effect estimation with double robustness. *Electronic Journal of Statistics*, 16(1):3523–3560, 2022.

[25] Michael Unser, Akram Aldroubi, and Murray Eden. B-spline signal processing. i. theory. *IEEE transactions on signal processing*, 41(2):821–833, 1993.

[26] Baqun Zhang, Anastasios A Tsiatis, Eric B Laber, and Marie Davidian. A robust method for estimating optimal treatment regimes. *Biometrics*, 68(4):1010–1018, 2012.

[27] Chengchun Shi, Rui Song, and Wenbin Lu. Robust learning for optimal treatment decision with np-dimensionality. *Electronic journal of statistics*, 10:2894, 2016.

[28] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

Reference

[29] Sahand N Negahban, Pradeep Ravikumar, Martin J Wainwright, and Bin Yu. A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers. 2012.

[30] Hui Zou. The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429, 2006.

[31] Chong Zhang and Yufeng Liu. Multicategory angle-based large-margin classification. *Biometrika*, 101(3):625–640, 2014.

[32] Chong Zhang, Jingxiang Chen, Haoda Fu, Xuanyao He, Ying-Qi Zhao, and Yufeng Liu. Multicategory outcome weighted margin-based learning for estimating individualized treatment rules. *Statistica sinica*, 30:1857, 2020.

[33] James M Robins, Andrea Rotnitzky, and Lue Ping Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association*, 89(427):846–866, 1994.

[34] Jared D Huling, Menggang Yu, and Maureen Smith. Fused comparative intervention scoring for heterogeneity of longitudinal intervention effects. 2019.

[35] David Ruppert, Matt P Wand, and Raymond J Carroll. *Semiparametric regression*. Number 12. Cambridge university press, 2003.

Thank You

Appendix: Continuous Outcome Model for Clustered Data

We can decompose the continuous outcome into:

$$\boldsymbol{Y}_i = \boldsymbol{m}(\boldsymbol{X}_i) + T_i\boldsymbol{\delta}(\boldsymbol{X}_i)/2 + \boldsymbol{\epsilon}_i$$

- Main Effect is characterized by

$$\boldsymbol{m}(\boldsymbol{X}_i) = \{\mathbb{E}(\boldsymbol{Y}_i|T_i = 1, \boldsymbol{X}_i) + \mathbb{E}(\boldsymbol{Y}_i|T_i = -1, \boldsymbol{X}_i)\}/2$$

  where $\boldsymbol{m}(\boldsymbol{X}_i) = (m(\boldsymbol{X}_i), ..., m(\boldsymbol{X}_i))'$
- The individualized treatment effect (ITE) is represented by:

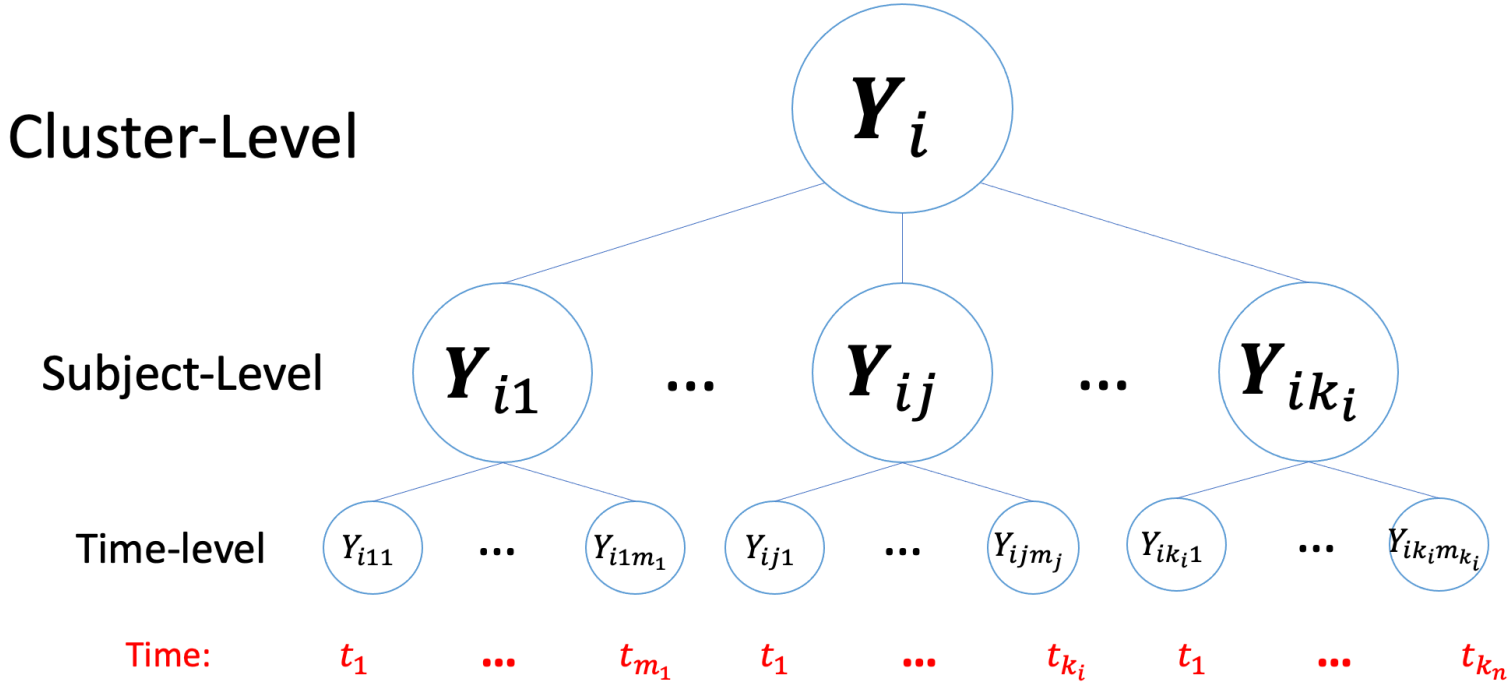
$$\boldsymbol{\delta}(\boldsymbol{X}_i) := \mathbb{E}\left[(\boldsymbol{Y}_i^{(1)} - \boldsymbol{Y}_i^{(-1)})|\boldsymbol{X}_i\right]$$

  where $\boldsymbol{\delta}(\boldsymbol{X}_i) = (\delta(\boldsymbol{X}_i), ..., \delta(\boldsymbol{X}_i))'$
- Random Error $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, ..., \epsilon_{ik_i})'$ with $\mathbb{E}(\boldsymbol{\epsilon}_i) = \mathbf{0}_{k_i}$ and invertible $\text{Var}(\boldsymbol{\epsilon}_i) = \boldsymbol{V}_i$

Appendix: Multi-leveled Data

Data: $\{(\boldsymbol{Y_i}, T_i, \boldsymbol{X_i}) : i = 1, ..., n; j = 1, ..., k_i; k = 1, ..., m_j\}$

- Outcome: $\boldsymbol{Y_i} = (\boldsymbol{Y_{i1}}, ..., \boldsymbol{Y_{ik_i}})'$ for i-th cluster
- $\boldsymbol{Y_{ij}} = (Y_{ij1}, ..., Y_{ijm_j})'$ for j-th subject in i-th cluster
- $\boldsymbol{Y_{ij}}(t_k) := Y_{ijk}$ is the k-th observation for j-th subject in i-th cluster at time t_k



Appendix: Model 1 in Simulation Study

- Model 1: Full Mixed Effect Model with Lasso penalty and exchangeable correlation structure:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\hat{\pi}_{T_i}(\boldsymbol{X_i})}\left[\boldsymbol{Y_i}-\{\tilde{\boldsymbol{X}}_{\boldsymbol{i}}'\beta+(T_i\tilde{\boldsymbol{X}}_{\boldsymbol{i}}/2)'\gamma\}\right]'\boldsymbol{V_i}^{-1}$$
$$\left[\boldsymbol{Y_i}-\{\tilde{\boldsymbol{X}}_{\boldsymbol{i}}'\beta+(T_i\tilde{\boldsymbol{X}}_{\boldsymbol{i}}/2)'\gamma\}\right]$$
$$+\lambda\left(\sum_{q=1}^{p}|\beta_q|+\sum_{q=1}^{p}|\gamma_q|+|\beta_T|+|\gamma_T|\right)$$

where $\tilde{\boldsymbol{X}}_{\boldsymbol{i}}=(\tilde{\boldsymbol{X}}_{\boldsymbol{i1}},...,\tilde{\boldsymbol{X}}_{\boldsymbol{iK}})$, $\tilde{\boldsymbol{X}}_{\boldsymbol{ij}}=(1,t_j,\boldsymbol{X_i})'$ and $\beta=(\beta_0,\beta_T,\beta_1,...,\beta_p)'$, $\gamma=(\gamma_0,\gamma_T,\gamma_1,...,\gamma_p)'$

- $\hat{\pi}_{T_i}(\boldsymbol{X_i})=0.5$ in randomized trial
- ITE over time: $\hat{\boldsymbol{\delta}}_1(\boldsymbol{X_i},\boldsymbol{t_i})=(\tilde{\boldsymbol{X}}_{\boldsymbol{i1}}\hat{\gamma},...,\tilde{\boldsymbol{X}}_{\boldsymbol{iK}}\hat{\gamma})'$
- Average ITE for i-th subject: $\bar{\delta_1}(\boldsymbol{X_i})=\frac{1}{K}\sum_{j=1}^{K}\tilde{\boldsymbol{X}}_{\boldsymbol{ij}}\hat{\gamma}$

Appendix: Model 2 in Simulation Study

- Model 2: Huling's Method using square loss with fused lasso in time-varying coefficients *[Huling et al., 2019]*:

$$(\hat{\gamma}_{(1)}, \ldots, \hat{\gamma}_{(K)}) := \underset{(\gamma_{(1)},\ldots,\gamma_{(K)})}{\operatorname{argmin}} \frac{1}{K}\sum_{t=1}^{K}\frac{1}{n}\sum_{i=1}^{n}\frac{(Y_{it} - T_i\tilde{\boldsymbol{X}}_i\gamma_{(t)}/2)^2}{\hat{\pi}_{T_i}(\boldsymbol{X}_i)} + \lambda_1\sum_{q=1}^{p}\sum_{t=2}^{K}|\gamma_{t,q} - \gamma_{t-1,q}| + \lambda_2\sum_{q=1}^{p}\sum_{t=1}^{K}|\gamma_{t,q}|$$

- $\tilde{\boldsymbol{X}}_i = (1, \boldsymbol{X}_i')'$ and $\gamma_{(t)} = (\gamma_{(t),0}, \gamma_{(t),1}, \cdots, \gamma_{(t),p})$
- ITE over time: $\hat{\boldsymbol{\delta}}_2(\boldsymbol{X}_i, \boldsymbol{t}_i) = (\tilde{\boldsymbol{X}}_i\hat{\gamma}_{(1)}, \ldots, \tilde{\boldsymbol{X}}_i\hat{\gamma}_{(K)})'$
- Average ITE for i-th subject: $\bar{\delta}_2(\boldsymbol{X}_i) = \frac{1}{K}\sum_{t=1}^{K}\tilde{\boldsymbol{X}}_i\hat{\gamma}_{(t)}$

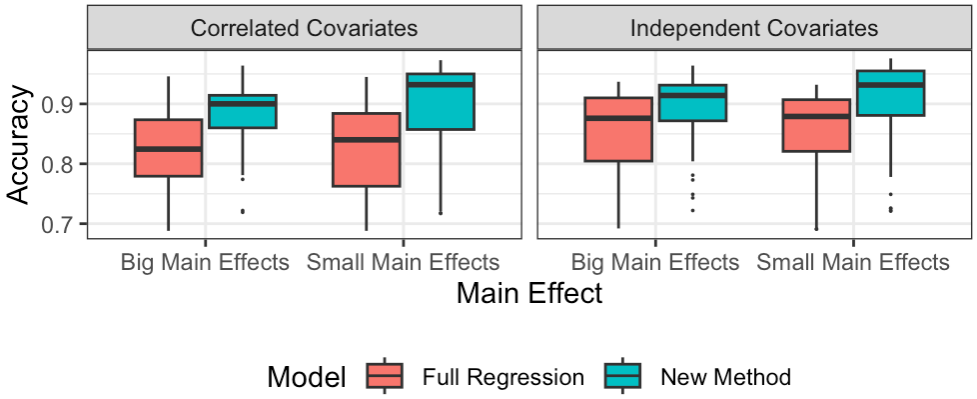
Appendix: Model 3 in Simulation Study

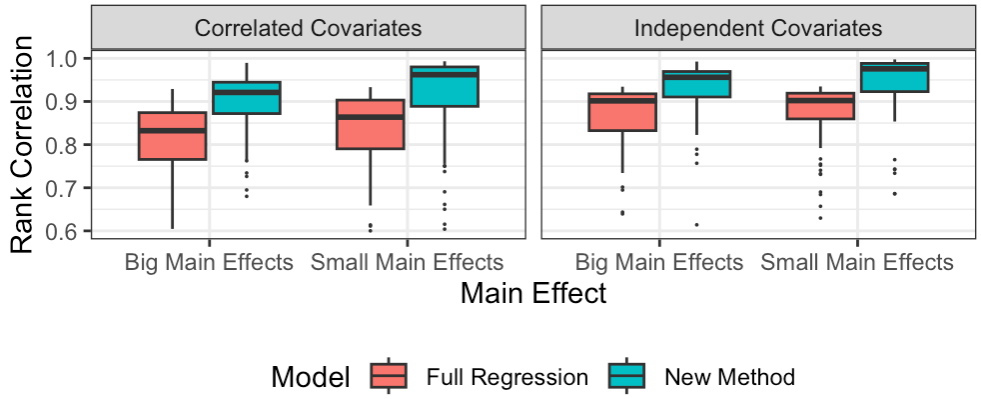
- Model 3: New Method with Lasso penalty and exchangeable correlation structure:

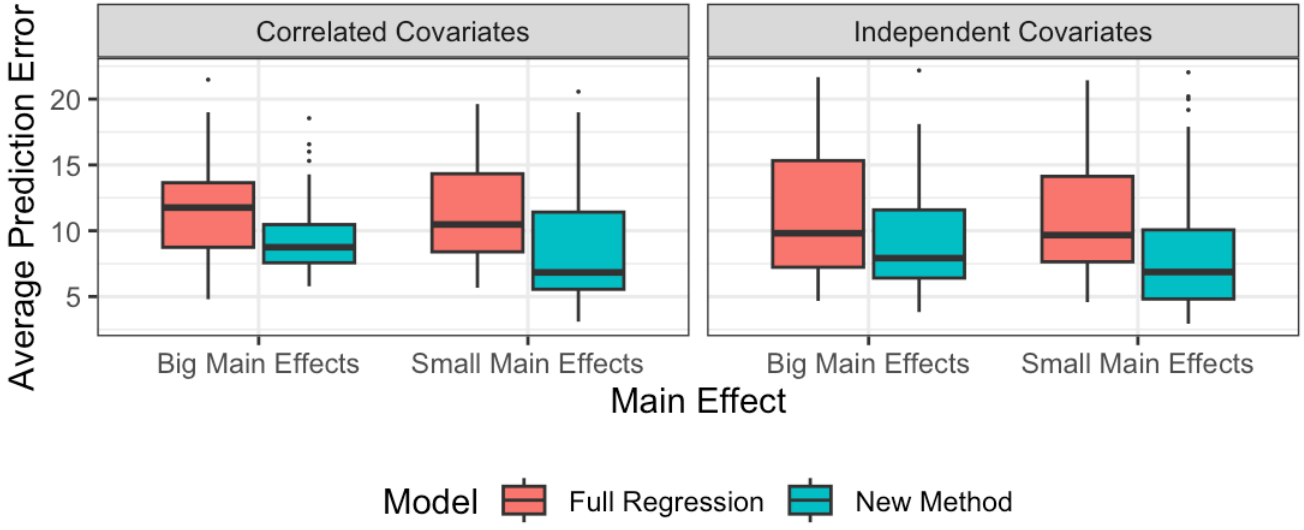
$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\hat{\pi}_{T_i}(\boldsymbol{X_i})}\left[\{\boldsymbol{Y_i}-\hat{\boldsymbol{m}}(\boldsymbol{X_i},\boldsymbol{t_i})\}-\{(T_i\tilde{\boldsymbol{X}}_i/2)'\gamma\}\right]'\boldsymbol{V_i}^{-1}$$
$$\left[\{\boldsymbol{Y_i}-\hat{\boldsymbol{m}}(\boldsymbol{X_i},\boldsymbol{t_i})\}-\{(T_i\tilde{\boldsymbol{X}}_i/2)'\gamma\}\right]$$
$$+\lambda\left(\sum_{q=1}^{p}|\gamma_q|+|\gamma_T|\right)$$

  where $\tilde{\boldsymbol{X}}_i=(\tilde{\boldsymbol{X}}_{i1},...,\tilde{\boldsymbol{X}}_{iK})$, $\tilde{\boldsymbol{X}}_{ij}=(1,t_j,\boldsymbol{X_i})'$, $\gamma=(\gamma_0,\gamma_T,\gamma_1,...,\gamma_p)'$
- $\hat{\pi}_{T_i}(\boldsymbol{X_i})=0.5$ in randomized trial
- $\hat{\boldsymbol{m}}(\boldsymbol{X_i},\boldsymbol{t_i})$ is estimated by linear mixed model with all covariates and time for efficiency augmentation.
- ITE over time: $\hat{\boldsymbol{\delta}}_{\boldsymbol{3}}(\boldsymbol{X_i},\boldsymbol{t_i})=(\tilde{\boldsymbol{X}}_{i1}\hat{\gamma},...,\tilde{\boldsymbol{X}}_{iK}\hat{\gamma})'$
- Average ITE for i-th subject: $\bar{\delta}_3(\boldsymbol{X_i})=\frac{1}{K}\sum_{j=1}^{K}\tilde{\boldsymbol{X}}_{ij}\hat{\gamma}$

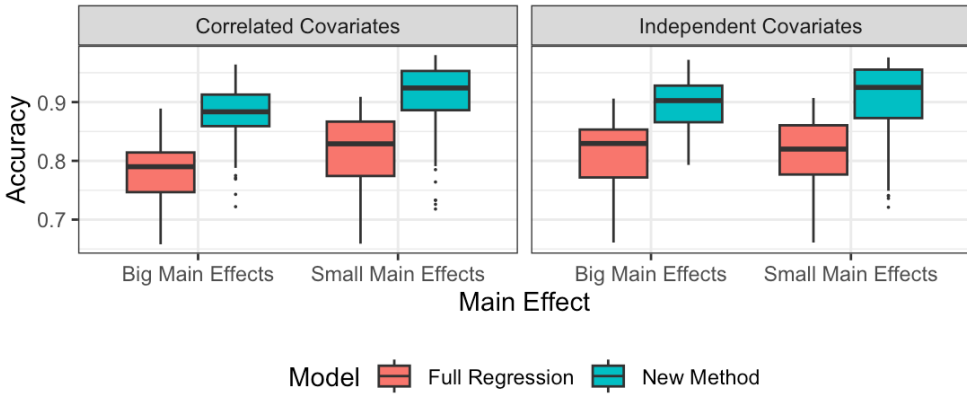
Appendix: Result of Clustered Data: S1

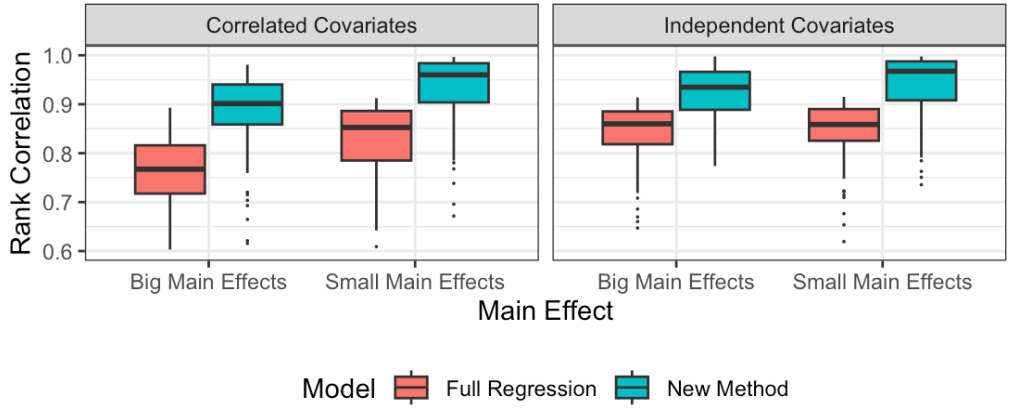


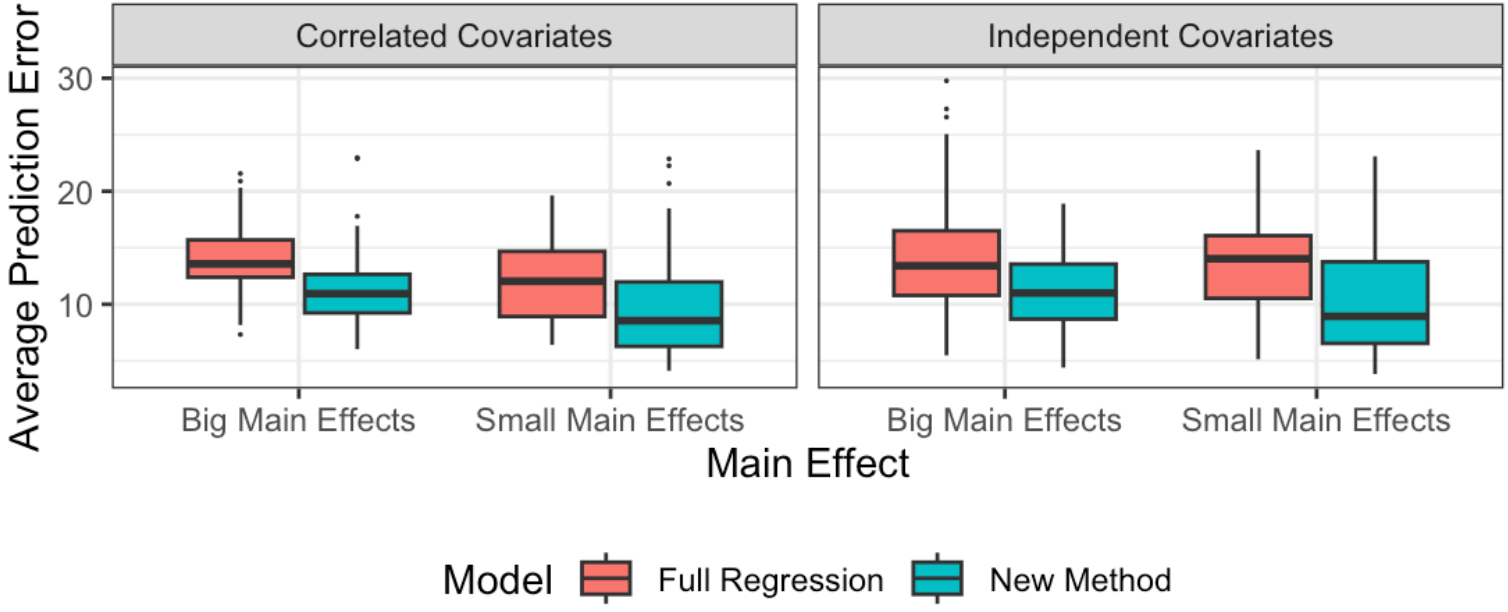




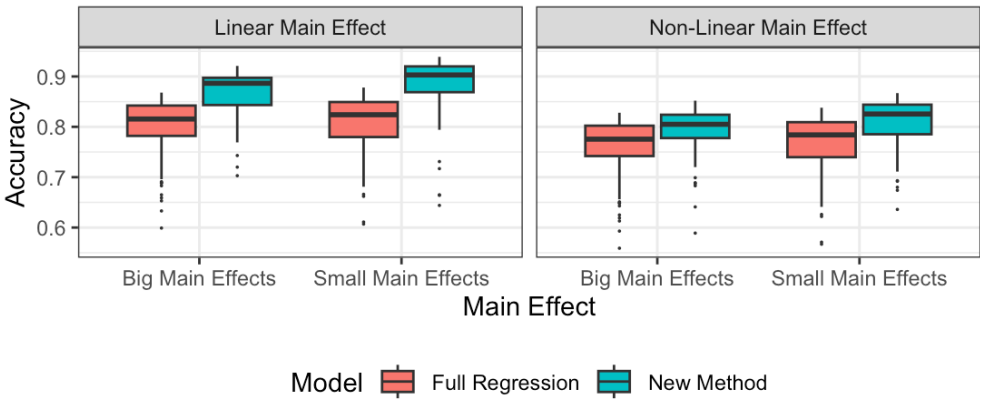
Appendix: Result of Clustered Data: S2

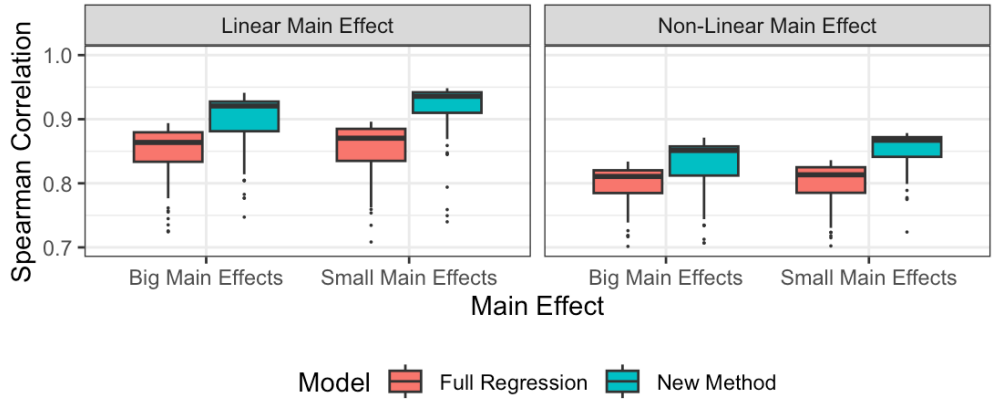


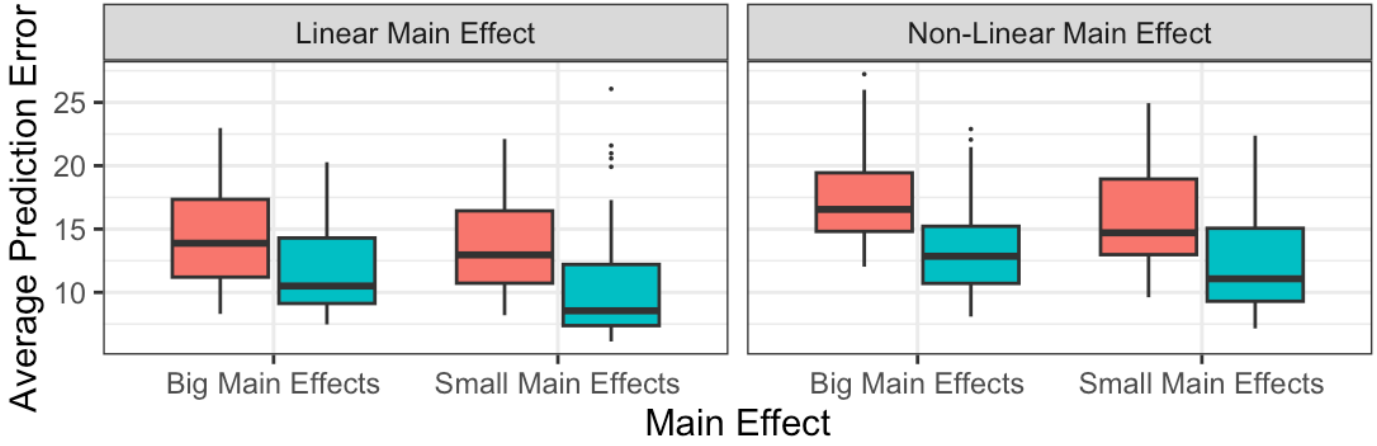


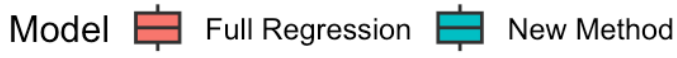


Appendix: Result of Clustered Data: S3

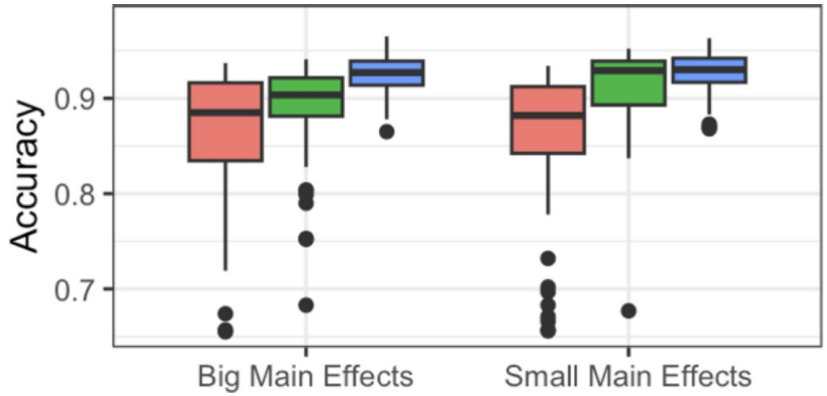




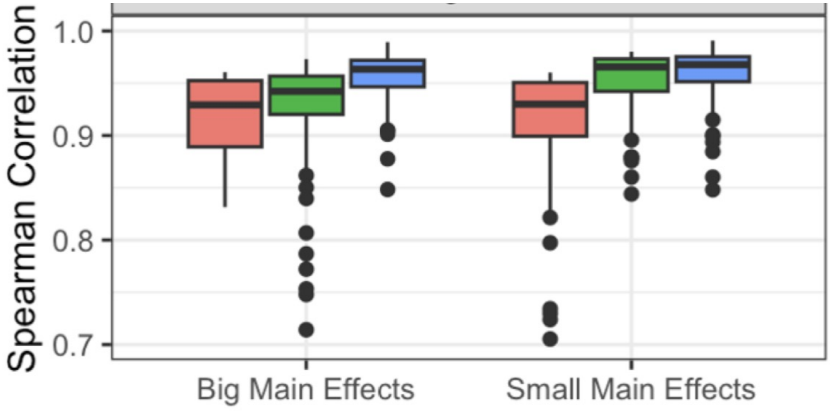


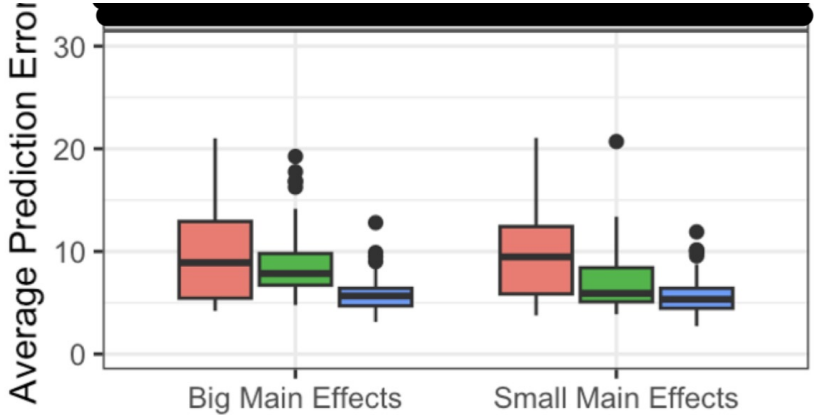


Appendix: Simulation of Non-Linear case



$$Y_{ij} = \beta_0 + \beta_T t_j + \sum_{q=1}^{10} \beta_q X_{i,q} + T_i \left(\gamma_0 + \gamma_T t_j + \sum_{q=1,2,8,10} \gamma_q X_{i,q} + 2X_{i,8}^2 - 4X_{i,10}^3 \right) /2 + \alpha_i + \epsilon_{ij}$$





Full Regression New Method New Method with B-Spline

- Non-Linear case: $f_{\text{non}}(\boldsymbol{X_i}, t_j) = \beta_0 + \beta_T t_j + \sum_{q=1}^{p} B(X_{i,q})\beta_q$ where $B(.)$ is the B-spline based function in the additive model