Design-based theory for cluster rerandomization

Hanzhong Liu

Center for Statistical Science, Department of Industrial Engineering,
Tsinghua University

Joint work with Xin Lu, Tianle Liu, and Peng Ding

Overview

Cluster randomized experiments

- Cluster-randomized experiments assign the treatments at the cluster level, with units within a cluster receiving the same treatment or control condition
- It helps to avoid interference within clusters and is applicable when individual-level assignments are logistically infeasible
- Before experiments, researchers often collect covariates at the individual or cluster level
- Covariate imbalance after treatment assignments often occurs and complicates the interpretation of the experimental results

Rerandomization (Contrained randomization)

- Fisher (1926) proposed blocking, or stratification, for balancing discrete covariates
- Rerandomization is a more general approach to balance continuous covariates (e.g., Raab & Butcher, 2001; Morgan & Rubin, 2012, Li et al., 2016)
- The existing design-based theory for rerandomization assumes that the treatments are assigned at the individual level (Morgan & Rubin, 2012; Li et al., 2018)

Cluster randomized experiments

- N units grouped into M clusters, M_1 clusters assigned to the treatment arm and M_0 clusters to the control arm
- n_i: the cluster size of cluster i
- Treatment indicator for cluster and individual
  - Z_i: the treatment indicator for cluster i
  - Z_{ij}: the treatment indicator for unit j in cluster i
  - $Z_{ij} = Z_i$
- N_1 treated units, N_0 control units. N_1 and N_0 are **random** if the n_i's vary.
- Two potential outcomes for each unit
  - $Y_{ij}(1)$: if unit j in cluster i is assigned to the **treatment** arm
  - $Y_{ij}(0)$: if unit j in cluster i is assigned to the **control** arm

Cluster randomized experiments

- Stable Unit Treatment Value Assumption (Rubin, 1980)

$$Y_{ij} = Z_{ij} Y_{ij}(1) + (1 - Z_{ij}) Y_{ij}(0)$$

- Two types of covariates
  - $x_{ij} = (x_{ij1}, \ldots, x_{ijK})^{\mathrm{T}}$: individual-level covariates
  - $c_i = (c_{i1}, \ldots, c_{iK})^{\mathrm{T}}$: cluster-level covariates
- **A**verage **T**reatment **E**ffect (ATE)

$$\tau = N^{-1} \sum_{i=1}^{M} \sum_{j=1}^{n_i} \{ Y_{ij}(1) - Y_{ij}(0) \}$$

- Infer τ based on $\{Y_{ij}, Z_{ij}, x_{ij}, c_i\}$ for $i = 1, \ldots, M, j = 1, \ldots, n_i$

Design-based inference

- Design-based / randomization-based / model-assisted / model-free inference
  - $Y_{ij}(1)$, $Y_{ij}(0)$, x_{ij} and c_i are fixed quantities (finite population)
  - Randomness comes only from $Z = (Z_1, \cdots, Z_M)$
  - neither a fitted regression model nor a super-population model is assumed
- For a finite population $\{a_1, \ldots, a_M\}$
  - $\bar{a} = M^{-1} \sum_{i=1}^{M} a_i$
  - $\mathrm{var_f}(a) = (M-1)^{-1} \sum_{i=1}^{M} (a_i - \bar{a})^2$ denotes its finite-population variance
  - $\mathrm{cov_f}$ denotes the finite-population covariance
- $\mathrm{pr_a}$, $\mathrm{var_a}$ and $\mathrm{cov_a}$ denote the asymptotic probability, variance, and covariance, respectively

Two ATE estimators

- The difference-in-means estimator (Hajek estimator) (Su & Ding, 2021)

$$\hat{\tau}_{\mathrm{haj}} = N_1^{-1} \sum_{i=1}^{M} \sum_{j=1}^{n_i} Z_{ij} Y_{ij} - N_0^{-1} \sum_{i=1}^{M} \sum_{j=1}^{n_i} (1 - Z_{ij}) Y_{ij}$$

- Horvitz–Thompson estimator (Middleton & Aronow, 2015):

$$\hat{\tau}_{\mathrm{ht}} = (NM_1/M)^{-1} \sum_{i=1}^{M} Z_i \sum_{j=1}^{n_i} Y_{ij} - (NM_0/M)^{-1} \sum_{i=1}^{M} (1 - Z_i) \sum_{j=1}^{n_i} Y_{ij}$$

Two ATE estimators

- Scaled cluster total potential outcome:

$$\tilde{Y}_{i\cdot}(z) = \sum_{j=1}^{n_i} Y_{ij}(z) M/N$$

- Observed scaled cluster total potential outcome

$$\tilde{Y}_{i\cdot} = Z_i \tilde{Y}_{i\cdot}(1) + (1 - Z_i)\tilde{Y}_{i\cdot}(0)$$

- The Horvitz-Thompson estimator derives as

$$\hat{\tau}_{\mathrm{ht}} = M_1^{-1} \sum_{i=1}^{M} Z_i \tilde{Y}_{i\cdot} - M_0^{-1} \sum_{i=1}^{M} (1 - Z_i)\tilde{Y}_{i\cdot}$$

- Su & Ding (2021) showed that $M^{1/2}(\hat{\tau}_\star - \tau) \mathrel{\dot\sim} \mathcal{N}(0, V_{\star,\tau\tau})$ for $\star = \mathrm{ht}, \mathrm{haj}$

Two cluster reranomization schemes

- Define

$$\hat{\tau}_{\mathrm{ht},c} = M_1^{-1} \sum_{i=1}^{M} Z_i c_i - M_0^{-1} \sum_{i=1}^{M} (1 - Z_i) c_i$$

$$\hat{\tau}_{\mathrm{haj},x} = N_1^{-1} \sum_{i=1}^{M} Z_i \sum_{j=1}^{n_i} x_{ij} - N_0^{-1} \sum_{i=1}^{M} (1 - Z_i) \sum_{j=1}^{n_i} x_{ij}.$$

- Cluster rerandomization scheme based on cluster level covariates:

$$\mathcal{M}_c = \{\hat{\tau}_{\mathrm{ht},c}^{\mathrm{T}} \ \mathrm{cov}(\hat{\tau}_{\mathrm{ht},c})^{-1} \ \hat{\tau}_{\mathrm{ht},c} \leq a\}$$

- Cluster rerandomization scheme based on individual level covariates

$$\mathcal{M}_x = \{\hat{\tau}_{\mathrm{haj},x}^{\mathrm{T}} \mathrm{cov}_{\mathrm{a}}(\hat{\tau}_{\mathrm{haj},x})^{-1} \ \hat{\tau}_{\mathrm{haj},x} \leq a\}$$

Jointly asymptotic distribution

Proposition 2.1

Under regularity conditions,

$$M^{1/2}\begin{pmatrix} \hat{\tau}_{\text{haj}} - \tau \\ \hat{\tau}_{\text{haj},x} \end{pmatrix} \dot{\sim} \mathcal{N}\left(0, \begin{bmatrix} V_{\text{haj},\tau\tau} & V_{\text{haj},\tau x} \\ V_{\text{haj},x\tau} & V_{\text{haj},xx} \end{bmatrix}\right),$$

$$M^{1/2}\begin{pmatrix} \hat{\tau}_{\text{ht}} - \tau \\ \hat{\tau}_{\text{ht},c} \end{pmatrix} \dot{\sim} \mathcal{N}\left(0, \begin{bmatrix} V_{\text{ht},\tau\tau} & V_{\text{ht},\tau c} \\ V_{\text{ht},c\tau} & V_{\text{ht},cc} \end{bmatrix}\right).$$

Jointly asymptotic distribution

- The Mahalanobis distances based on $\hat{\tau}_{\mathrm{haj},x}$ and $\hat{\tau}_{\mathrm{ht},c}$ converge in distribution to χ^2_K
- We can choose a as the αth quantile of χ^2_K to ensure an asymptotic acceptance rate of α
- Morgan & Rubin (2012) suggested $\alpha = 0.001$ when the cluster numbers are moderate or large
- For small M, we can use Fisher randomization tests and choose the threshold a to ensure non-trivial powers (Johansson et al., 2021)

Asymptotic distributions under cluster rerandomization

- $L_{k,a} \sim D_1 \mid D^{\mathrm{T}} D \leq a$ where $D = (D_1, \ldots, D_k)^{\mathrm{T}}$ is a k-dimensional standard normal random vector

- ϵ: a standard normal random variable independent of $L_{k,a}$

- Squared multiple correlation (Li et al., 2018)

$$
\begin{aligned}
R_c^2 &= \mathrm{cov_a}(\hat{\tau}_{\mathrm{ht}}, \hat{\tau}_{\mathrm{ht},c})\mathrm{cov_a}(\hat{\tau}_{\mathrm{ht},c})^{-1}\mathrm{cov_a}(\hat{\tau}_{\mathrm{ht},c}, \hat{\tau}_{\mathrm{ht}})/\mathrm{var_a}(\hat{\tau}_{\mathrm{ht}}) \\
R_x^2 &= \mathrm{cov_a}(\hat{\tau}_{\mathrm{haj}}, \hat{\tau}_{\mathrm{haj},x})\mathrm{cov_a}(\hat{\tau}_{\mathrm{haj},x})^{-1}\mathrm{cov_a}(\hat{\tau}_{\mathrm{haj},x}, \hat{\tau}_{\mathrm{haj}})/\mathrm{var_a}(\hat{\tau}_{\mathrm{haj}})
\end{aligned}
$$

Theorem 1

Under regularity conditions,

$$
\begin{aligned}
M^{1/2}(\hat{\tau}_{\mathrm{haj}} - \tau) \mid \mathcal{M}_x &\dot{\sim} (V_{\mathrm{haj},\tau\tau})^{1/2}\{(1 - R_x^2)^{1/2}\epsilon + R_x L_{K,a}\}, \\
M^{1/2}(\hat{\tau}_{\mathrm{ht}} - \tau) \mid \mathcal{M}_c &\dot{\sim} (V_{\mathrm{ht},\tau\tau})^{1/2}\{(1 - R_c^2)^{1/2}\epsilon + R_c L_{K,a}\}.
\end{aligned}
$$

A comparison between two rerandomization schemes

Corollary 2

Under regularity conditions, if $c_i = (n_i, \tilde{x}_{i\cdot}^{\mathrm{T}})^{\mathrm{T}}$, *then*

$$V_{\mathrm{haj},\tau\tau}(1 - R_x^2) \geq V_{\mathrm{ht},\tau\tau}(1 - R_c^2).$$

- Parallel to the results of Su & Ding (2021): The regression-adjusted estimator based on scaled cluster totals dominates the regression-adjusted estimator based on individual-level data with properly defined covariates

Weighted Euclidean distance criterion

- The cluster rerandomization schemes using Mahalanobis distances view all covariates as equally important
- With prior knowledge about the relative importance of the covariates, a better choice is cluster rerandomization with the weighted Euclidean distance (Wight et al., 2002; Althabe et al., 2008; Li et al., 2016, 2017; Hayes & Moulton, 2017; Dempsey et al., 2018)
- Cluster rerandomization schemes:

$$\mathcal{D}_x(A_x) = \{M\hat{\tau}_{\text{haj},x}^{\mathrm{T}} A_x \, \hat{\tau}_{\text{haj},x} \leq a\}, \quad \mathcal{D}_c(A_c) = \{M\hat{\tau}_{\text{ht},c}^{\mathrm{T}} \, A_c \, \hat{\tau}_{\text{ht},c} \, \leq a\}$$

- Mahalanobis distance: $A_x = M^{-1}\text{cov}_{\text{a}}(\hat{\tau}_{\text{haj},x})^{-1}$ and $A_c = M^{-1}\,\text{cov}(\hat{\tau}_{\text{ht},c}\,)^{-1}$
- Weighted Euclidean distance: Diagonal A_x and A_c

Asymptotic distribution under weighted Euclidean distance criterion

Theorem 3

Under regularity conditions,

$$M^{1/2}(\hat{\tau}_{\mathrm{haj}} - \tau) \mid \mathcal{D}_x(A_x) \ \dot{\sim} \ V_{\mathrm{haj},\tau\tau}^{1/2}\{(1 - R_x^2)^{1/2}\epsilon + R_x \mu_x^{\mathrm{T}} \eta \mid \eta^{\mathrm{T}} V_{\mathrm{haj},xx}^{1/2} A_x V_{\mathrm{haj},xx}^{1/2} \eta \leq a\},$$

$$M^{1/2}(\hat{\tau}_{\mathrm{ht}} - \tau) \mid \mathcal{D}_c(A_c) \ \dot{\sim} \ V_{\mathrm{ht},\tau\tau}^{1/2}\{(1 - R_c^2)^{1/2}\epsilon + R_c \mu_c^{\mathrm{T}} \eta \mid \eta^{\mathrm{T}} V_{\mathrm{ht},cc}^{1/2} A_c V_{\mathrm{ht},cc}^{1/2} \eta \leq a\},$$

where $\eta = (\eta_1, \ldots, \eta_K)^{\mathrm{T}}$, $\epsilon, \eta_1, \ldots, \eta_K$ *are independent* $\mathcal{N}(0, 1)$,

$$\mu_x^{\mathrm{T}} = (V_{\mathrm{haj},\tau x} V_{\mathrm{haj},xx}^{-1} V_{\mathrm{haj},x\tau})^{-1/2} V_{\mathrm{haj},\tau x} V_{\mathrm{haj},xx}^{-1/2},$$

$$\mu_c^{\mathrm{T}} = (V_{\mathrm{ht},\tau c} V_{\mathrm{ht},cc}^{-1} V_{\mathrm{ht},c\tau})^{-1/2} V_{\mathrm{ht},\tau c} V_{\mathrm{ht},cc}^{-1/2}.$$

Properties of asymptotic distributions under cluster rerandomization

Proposition 2.2

Under regularity conditions, (i) the asymptotic distributions in Theorem 2 are symmetric around zero and unimodal, and (ii) $\mathrm{pr}_{\mathrm{a}}\{M^{1/2}|\hat{\tau}_{\mathrm{haj}} - \tau| < \delta \mid \mathcal{D}_x(A_x)\}$ is a non-decreasing function of R_x^2 and $\mathrm{pr}_{\mathrm{a}}\{M^{1/2}|\hat{\tau}_{\mathrm{ht}} - \tau| < \delta \mid \mathcal{D}_c(A_c)\}$ is a non-decreasing function of R_c^2 for any fixed $\delta > 0$.

- (i) ensures that the asymptotic distributions are both bell-shaped
- (ii) ensures that the asymptotic distributions are more concentrated at zero than those under standard cluster randomization

Comparing efficiency of different criteria

- We can compare their variance reductions given the same acceptance rate
- Let α denote the asymptotic acceptance rate:

$$\alpha = \text{pr}_\text{a}\{\mathcal{D}_x(A_x)\} = \text{pr}_\text{a}(M\hat{\tau}_{\text{haj},x}^\text{T} A_x \hat{\tau}_{\text{haj},x} \leq a),$$
$$\alpha = \text{pr}_\text{a}\{\mathcal{D}_c(A_c)\} = \text{pr}_\text{a}(M\hat{\tau}_{\text{ht},c}^\text{T} A_c \hat{\tau}_{\text{ht},c} \leq a)$$

- Let $\Gamma(\cdot)$ be the Gamma function and

$$p_K = \frac{2\pi}{K+2}\left\{\frac{2\pi^{K/2}}{K\Gamma(K/2)}\right\}^{-2/K}$$

Variance expansion

Theorem 4

Under regularity conditions,

$$\mathrm{var}_{\mathrm{a}}\{M^{1/2}(\hat{\tau}_{\mathrm{haj}} - \tau) \mid \mathcal{D}_x(A_x)\} = V_{\mathrm{haj},\tau\tau}\{(1 - R_x^2) + R_x^2 p_K \nu_x(A_x)\alpha^{2/K} + o(\alpha^{2/K})\},$$

$$\mathrm{var}_{\mathrm{a}}\{M^{1/2}(\hat{\tau}_{\mathrm{ht}} - \tau) \mid \mathcal{D}_c(A_c)\} = V_{\mathrm{ht},\tau\tau}\{(1 - R_c^2) + R_c^2 p_K \nu_c(A_c)\alpha^{2/K} + o(\alpha^{2/K})\},$$

for a small α, where

$$\nu_x(A_x) = \frac{V_{\mathrm{haj},\tau x} V_{\mathrm{haj},xx}^{-1} A_x^{-1} V_{\mathrm{haj},xx}^{-1} V_{\mathrm{haj},x\tau} \det(A_x)^{1/K} \det(V_{\mathrm{haj},xx})^{1/K}}{V_{\mathrm{haj},\tau x} V_{\mathrm{haj},xx}^{-1} V_{\mathrm{haj},x\tau}},$$

$$\nu_c(A_c) = \frac{V_{\mathrm{ht},\tau c} V_{\mathrm{ht},cc}^{-1} A_c^{-1} V_{\mathrm{ht},cc}^{-1} V_{\mathrm{ht},c\tau} \det(A_c)^{1/K} \det(V_{\mathrm{ht},cc})^{1/K}}{V_{\mathrm{ht},\tau c} V_{\mathrm{ht},cc}^{-1} V_{\mathrm{ht},c\tau}}.$$

Weighted Euclidean distance with optimal weights

Theorem 5

Under regularity conditions, if $V_{\text{haj},\tau x} V_{\text{haj},xx}^{-1} \xi_k$ *and* $V_{\text{ht},\tau c} V_{\text{ht},cc}^{-1} \xi_k$ *are nonzero for all* $k = 1, \ldots, K$*, then* $\nu_x\{\text{diag}(w_1, \ldots, w_K)\}$ *reaches minimum if* $w_k \propto (V_{\text{haj},\tau x} V_{\text{haj},xx}^{-1} \xi_k)^2$ *for* $k = 1, \ldots, K$*, and* $\nu_c\{\text{diag}(w_1, \ldots, w_K)\}$ *reaches minimum if* $w_k \propto (V_{\text{ht},\tau c} V_{\text{ht},cc}^{-1} \xi_k)^2$ *for* $k = 1, \ldots, K$*.*

- A_x^{opt} and A_c^{opt}: the optimal weighting matrices
- With orthogonalized covariates, the optimal weighted Euclidean distance better
- However, this conclusion does not hold if the covariates are not orthogonalized

Comparison with cluster rerandomization with tiers of covariates

- Morgan & Rubin (2015) proposed rerandomization with tiers of covariates as an alternative to rerandomization with the weighted Euclidean distance
- No comparison has been made between these two rerandomization schemes

Theorem 6

Under regularity conditions with orthogonalized covariates, rerandomization with the optimal weighted Euclidean distance is better than rerandomization with tiers of covariates.

Rerandomization and regression adjustment

- Rerandomization uses covariates in the design stage (Morgan & Rubin, 2012), and regression adjustment uses covariates in the analysis stage (Lin, 2013)
- Li & Ding (2020) showed that they could be used simultaneously
- Analogous results hold under cluster rerandomization but there are some differences

Regression adjustment under cluster randomized experiment

- Under $\mathcal{D}_x(A_x)$
  - coefficient of Z_{ij} in the least squares fit of Y_{ij} on $(1, Z_{ij}, x_{ij}, Z_{ij}x_{ij})$
  - cluster-robust standard error (Liang & Zeger, 1986)
- Under $\mathcal{D}_c(A_c)$
  - coefficient of Z_i in the least squares fit of $\tilde{Y}_{i\cdot}$ on $(1, Z_i, c_i, Z_i c_i)$
  - heteroskedasticity-robust standard error (Huber, 1967; White, 1980)
- Regression coefficient and variance estimator: $(\hat{\tau}^{\text{adj}}_{\text{haj}}, \hat{V}^{\text{adj}}_{\text{LZ}})$ and $(\hat{\tau}^{\text{adj}}_{\text{ht}}, \hat{V}^{\text{adj}}_{\text{HW}})$

Asymptotic results on cluster rerandomization combined with regression adjustment

Theorem 7

Assume regularity conditions hold.

(i) *Under $\mathcal{D}_c(A_c)$, the estimator $\hat{\tau}_{\text{ht}}^{\text{adj}}$ is consistent for τ and asymptotically normal, the probability limit of $M\hat{V}_{\text{HW}}^{\text{adj}}$ is larger than or equal to the true asymptotic variance of $M^{1/2}\hat{\tau}_{\text{ht}}^{\text{adj}}$, and the $1-\varsigma$ confidence interval*

$$\left[\hat{\tau}_{\text{ht}}^{\text{adj}} + (\hat{V}_{\text{HW}}^{\text{adj}})^{1/2} z_{\varsigma/2}, \hat{\tau}_{\text{ht}}^{\text{adj}} + (\hat{V}_{\text{HW}}^{\text{adj}})^{1/2} z_{1-\varsigma/2}\right]$$

has asymptotic coverage rate $\geq 1-\varsigma$;

Theorem 7 continued

(ii) *Under $\mathcal{D}_x(A_x)$, the estimator $\hat{\tau}_{\text{haj}}^{\text{adj}}$ is consistent for τ and its asymptotic distribution is a convolution of normal and truncated normal, the probability limit of $M\hat{V}_{\text{LZ}}^{\text{adj}}$ is larger than or equal to the true asymptotic variance of $M^{1/2}\hat{\tau}_{\text{haj}}^{\text{adj}}$, and the $1-\varsigma$ confidence interval*

$$\left[\hat{\tau}_{\text{haj}}^{\text{adj}} + (\hat{V}_{\text{LZ}}^{\text{adj}})^{1/2} z_{\varsigma/2}, \hat{\tau}_{\text{haj}}^{\text{adj}} + (\hat{V}_{\text{LZ}}^{\text{adj}})^{1/2} z_{1-\varsigma/2}\right]$$

has asymptotic coverage rate $\geq 1-\varsigma$;

(iii) *If $c_i = (n_i, \tilde{x}_{i\cdot}^{\mathrm{T}})^{\mathrm{T}}$, the asymptotic distribution of $\hat{\tau}_{\text{ht}}^{\text{adj}} \mid \mathcal{D}_c(A_c)$ is more concentrated at τ than $\hat{\tau}_{\text{haj}}^{\text{adj}} \mid \mathcal{D}_x(A_x)$, in the sense that for any $\delta > 0$, we have*

$$\text{pr}_{\text{a}}\{M^{1/2}|\hat{\tau}_{\text{haj}}^{\text{adj}} - \tau| < \delta \mid \mathcal{D}_x(A_x)\}$$
$$\leq \text{pr}_{\text{a}}\{M^{1/2}|\hat{\tau}_{\text{ht}}^{\text{adj}} - \tau| < \delta \mid \mathcal{D}_c(A_c)\}.$$

Simulation setup

- Potential outcomes model:

$$Y_{ij}(z) = g(n_i) + x_{ij}^{\mathrm{T}} \beta_{iz} + \varepsilon_{ij}(z)$$

- $M = 100,\ M_1 = M_0 = 50$
- size of each cluster is sampled uniformly from $\{m \in \mathbb{N} \mid 4 \leq m \leq 10\}$
- k: covariate dimension, ρ: correlation of covariates

Table: Parameters of four scenarios.

Scenario	k	ρ	$g(n_i)$
1	7	0.8	$(n_i - 7)/2$
2	7	-0.15	$(n_i - 7)/2$
3	12	0.4	6
4	10	-0.1	6

Estimators

- Three orthogonal axes,
  - individual-level (`X`) versus cluster-level (`C`)
  - the Mahalanobis distance (`M`) versus the optimal weighted Euclidean distance without orthogonalization (`W`)
  - using regression adjustment (`.adj`) or not
- Two baseline methods: Hajek (`Haj`) and Horvitz–Thompson (`HT`) estimators without using cluster rerandomization

Simulation results

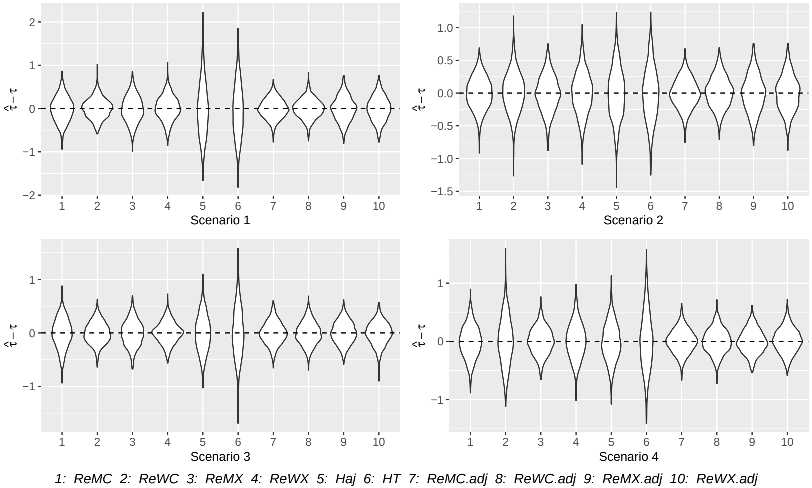

Figure: Comparison of methods in the simulated example.

Conclusion

- We study cluster rerandomization with both individual- and cluster-level covariates, and derive a design-based asymptotic theory for estimators either with or without regression adjustment
- We compare cluster rerandomization schemes based on weighted Euclidean distance and that based on Mahalanobis distance with tiers of covariates: for orthogonalized covariates, the former with optimal weights dominates the latter
- When M is small
  - Use a mixed-effects model by imposing modeling assumptions on the data generating process
  - Use Fisher randomization tests with studentized statistics (Zhao and Ding, 2021)

Thank you!