Methods: Mind the Gap
Webinar Series

Toward Causal Inference in Cluster Randomized Trials: Estimands and Reflection on Current Practice

Presented by:
Fan Li, Ph.D.
Yale School of Public Health

NIH National Institutes of Health
Office of Disease Prevention

Toward causal inference in cluster randomized trials: Estimands and reflection on current practice

Fan Li

Department of Biostatistics
Center for Methods in Implementation and Prevention Science (CMIPS)
Yale University School of Public Health

NIH Office of Disease Prevention, Methods: Mind the Gap Webinar Series
November 3, 2022

Cluster randomized trials

- Cluster randomized trials (CRTs) randomize entire clusters or groups of individuals to treatment conditions
  - cluster = hospital, school or village
  - often used when intervention is naturally implemented at the cluster-level or to avoid contamination
  - individuals receive the same treatment within same cluster
- Increasingly popular in pragmatic trials
  - Among 22 demonstration projects supported by NIH Pragmatic Trials Collaboratory, 16 cluster randomized trials, 4 individually randomized group treatment trials, 2 individually-randomized trials
- Multiple classic textbooks (Murray, 1998; Donner and Klar, 2000; . . .), with a consistent message: the design and analysis must account for intracluster correlation coefficient (ICC) of outcomes

Model-based treatment effect

- In analysis of cluster randomized trials, a primary interest is to estimate the treatment effect
  - outcomes frequently collected at the individual level
  - linear mixed model or generalized estimating equations (GEE) are two mainstream multilevel regression tools in this context
  - coefficient of the treatment variable, β, is simply taken as the "average" treatment effect
- While this is a common practice, a potential challenge in interpretation is that the treatment effect estimand is defined based on a model
  - may change with link function, covariate specification, or random-effects specification
  - if the model is NOT the data generating process, what treatment contrast does β represent?

Estimands

- Increasing interest in defining estimands at the outset
- For individually randomized trials

> **(Causal) Estimand**: *A precise description of the treatment effect investigators aim to estimate from a study. An estimand comprises five aspects: (i) population; (ii) treatment conditions; (iii) endpoint; (iv) summary measure (e.g. difference in means, risk ratio, etc.); and (v) how intercurrent events are to be handled.*
> ***(The Estimand Framework)***

- Unclear how current practice with interpreting regression coefficients in cluster trials fully address these 5 aspects
- We will primarily focus on the "population" concept in discussing estimands for cluster randomized trials

Averaging over what population?

(Kahan et al. 2022, IJE)

Consider a hypothetical cluster randomized trial with

Cluster	Cluster size N_i	True (constant) causal effect in cluster i
1	10	5
2	10	5
3	10	5
4	100	1
5	100	1
6	100	1

- treatment effect as an average across all individuals in the trial

$$\frac{(10)(5) + (10)(5) + (10)(5) + (100)(1) + (100)(1) + (100)(1)}{10 + 10 + 10 + 100 + 100 + 100} = 1.4$$

- treatment effect as an average across clusters in the trial

$$\frac{5 + 5 + 5 + 1 + 1 + 1}{6} = 3$$

- difference can be substantial depending on how treatment effect varies according to cluster size variation

Formalizing the estimand concept

- Individual j in cluster i has two counterfactual outcomes $\{Y_{ij}(1), Y_{ij}(0)\}$, and N_i the number of individuals per cluster
  - individual-average treatment effect

$$\text{i-ATE} = \underbrace{\frac{\sum_{i=1}^{I}\sum_{j=1}^{N_i}\{Y_{ij}(1)-Y_{ij}(0)\}}{\sum_{i=1}^{I}N_i}}_{\text{finite-population}} \Longleftrightarrow \underbrace{\frac{\mathbb{E}\ \sum_{j=1}^{N_i}\{Y_{ij}(1)-Y_{ij}(0)\}}{\mathbb{E}[N_i]}}_{\text{super-population}}$$

  - cluster-average treatment effect

$$\text{c-ATE} = \underbrace{\frac{1}{I}\sum_{i=1}^{I}\left[\frac{\sum_{j=1}^{N_i}Y_{ij}(1)-Y_{ij}(0)}{N_i}\right]}_{\text{finite-population}} \Longleftrightarrow \underbrace{\mathbb{E}\left[\frac{\sum_{j=1}^{N_i}Y_{ij}(1)-Y_{ij}(0)}{N_i}\right]}_{\text{super-population}}$$

- Emerging literature: Imai et al. (2009, Stat Sci); Su and Ding (2021, JRSSB) ; Balzer et al. (2019, SMMR); Wang et al. (2021, arXiv); Bugni et al. (2021, arXiv); Kahan et al. (2022, IJE); Wang et al. (2022, Cont Clin Trials); Wang et al. (2022, arXiv)
  - finite- v.s. super-population framework
  - focus on difference scale (absolute scale), can extend to ratio scale

Formalizing the estimand concept - cont'd

- Individual j in cluster i has two counterfactual outcomes $\{Y_{ij}(1), Y_{ij}(0)\}$, and N_i the number of individuals per cluster
- For risk ratio and odds ratio definitions, take the finite-population definition as an example
  - individual-average treatment effect (ratio scale)

$$\text{i-ATE}^{\text{RR}} = \frac{p_1}{p_0}, \qquad \text{i-ATE}^{\text{RR}} = \frac{p_1(1-p_0)}{p_0(1-p_1)}$$

$$p_1 = \frac{\sum_{i=1}^{I}\sum_{j=1}^{N_i} Y_{ij}(1)}{\sum_{i=1}^{I} N_i}, \qquad p_0 = \frac{\sum_{i=1}^{I}\sum_{j=1}^{N_i} Y_{ij}(0)}{\sum_{i=1}^{I} N_i}$$

  - cluster-average treatment effect (ratio scale)

$$\text{c-ATE}^{\text{OR}} = \frac{q_1}{q_0}, \qquad \text{c-ATE}^{\text{OR}} = \frac{q_1(1-q_0)}{q_0(1-q_1)}$$

$$q_1 = \frac{1}{I}\sum_{i=1}^{I}\left[\frac{\sum_{j=1}^{N_i} Y_{ij}(1)}{N_i}\right], \qquad q_0 = \frac{1}{I}\sum_{i=1}^{I}\left[\frac{\sum_{j=1}^{N_i} Y_{ij}(1)}{N_i}\right]$$

Interpretation

- The individual-average treatment effect answers

  "*What is the expected change in outcome associated with treatment for a typical patient?*"

  - **example**: for an intervention (delivered to patients) aiming to reduce mortality, then i-ATE may be of interest, as this represents the population impact of switching from the control to intervention

- The cluster-average treatment effect answers

  "*What is the expected change in outcome associated with treatment for a typical provider or clinic and its natural patient panel?*"

  - **example**: in a trial aiming to reduce unnecessary prescribing of antibiotics to patients, in which doctors act as the cluster, then a c-ATE may also be of interest, as this provides the intervention's effect on the clinician's prescribing habits

- When will they differ?

Informative cluster size

(Kahan et al. 2022, IJE)

- Key factor that drives such difference is whether cluster size is informative
  - Type 1: outcomes differ between small and large clusters but treatment effect is the same (**example**: baseline event rate is 10% and 20% in small and large clusters, but odds ratio is constant, 0.75)
  - Type 2: treatment effect differs between small and large clusters or cluster size is an effect modifier (**example**: odds ratio is 0.75 and 0.5 in small and large clusters)
  - can depend on effect measure
- Informative cluster size can occur when factors that differ between small and large clusters also affect the potential outcome or interact with the treatment group
  - **example**: differences in staff experience or levels of care between larger and smaller hospitals
  - **example**: differences in socioeconomic status between larger urban schools compared with smaller rural schools

Methods for estimand-aligned analysis

(Kahan et al. 2022, IJE)

- Give equal weight to each individual or each cluster?

Estimand	Example recipe
Individual-average treatment effect	**Cluster-level analysis** (1) Calculate cluster-level summaries; (2) Analyze cluster-level summaries using a weighted regression model (weights equal to N_i); (3) Huber-White SE. **Individual-level analysis** (1) GEE with a working independence correlation structure; (2) Cluster-robust SE
Cluster-average treatment effect	**Cluster-level analysis** (1) Calculate cluster-level summaries; (2) Analyze cluster-level summaries using regression model (unweighted); (3) Huber-White SE. **Individual-level analysis** (1) GEE with a working independence correlation structure, but with inverse cluster-size weights equal to $1/N_i$; (2) Cluster-robust SE.

- With covariates, see Su and Ding (2021, JRSSB) for a generalization

Example of different statistical decisions

- There are occasions where we face different modeling decisions. . .
- Take GEE as an example (again without covariates)

Method	Working Correlation Matrix		Cluster Size Weighting	
	Independence	Exchangeable	Inverse	No
IEE[a]	✓			✓
IEEW[b]	✓		✓	
EEE[c]		✓		✓
EEEW[d]		✓	✓	

[a] IEE: GEE with independence working correlation without cluster size weighting.

[b] IEEW: GEE with independence working correlation with inverse cluster size weighting.

[c] EEE: GEE with exchangeable working correlation without cluster size weighting.

[d] EEEW: GEE with exchangeable working correlation with inverse cluster size weighting.

Simulation evidence on bias

(Wang et al. 2022, Cont Clin Trials)

Statistical decisions can drive the estimand!

(Wang et al. 2022, Cont Clin Trials)

Method	Probability limit of treatment coefficient estimator	Interpretation
IEE	$\frac{1}{\mathbb{E}(N_i)}\mathbb{E}\left\{\sum_{j=1}^{N_i}[Y_{ij}(1)-Y_{ij}(0)\}\right]$	i-ATE
IEEW	$\mathbb{E}\left[\frac{1}{N_i}\sum_{j=1}^{N_i}\{Y_{ij}(1)-Y_{ij}(0)\}\right]$	c-ATE
EEE	$\mathbb{E}\left[\frac{\sum_{j=1}^{N_i}\{Y_{ij}(1)-Y_{ij}(0)\}}{1+(N_i-1)\rho^*}\right]\Big/\mathbb{E}\left[\frac{N_i}{1+(N_i-1)\rho^*}\right]$	?
EEEW	$\mathbb{E}\left[\frac{\sum_{j=1}^{N_i}\{Y_{ij}(1)-Y_{ij}(0)\}}{\{1+(N_i-1)\rho^*\}N_i}\right]\Big/\mathbb{E}\left[\frac{1}{1+(N_i-1)\rho^*}\right]$	?

- ρ^* is the limit of the ICC estimator under EEE and EEEW
- precision weight can introduce bias when the cluster size is correlated with the treatment contrasts within each cluster
- not an issue if the treatment effect is constant and does not depend on cluster size

An example context

(Kahan et al. 2022, IJE)

- **Context**: a cluster randomized trial comparing a quality improvement (QI) intervention to improve outcomes in patients undergoing emergency laparotomy
  - implementing hospital-wide improvement program in each cluster
- **Primary outcome**: mortality within 90 days
- **Secondary outcome**: whether a senior surgeon is present in the operating room (performing the surgery or supervising a junior surgeon in doing so)
  - intended to measure success of QI intervention in changing practice

An example context - estimand

- For the primary outcome, the interest lies in effect on individual patients
  - "*how many additional lives can be saved through the QI intervention*?"
  - individual-average treatment effect is relevant
- For the secondary outcome, the interest lies in measuring treatment success in changing practice
  - "*whether the QI intervention can improve the quality of health care delivery for an average hospital?*"
  - cluster-average treatment effect appears more relevant
- No uniform solution to all trials, and require a conversation between statistician and the study team to reflect on the scientific goal
  - maybe both are important!

An example context - analytical methods

- Suppose there is concern that the impact of treatment on outcome can differ between smaller and larger clusters
  - differing resource levels available across heterogeneous clusters
  - informative cluster size cannot be ruled out
- Suppose we are interested in performing unadjusted analysis without covariates
  - wish to estimate risk difference (absolute scale)
  - for primary outcome (mortality), can use independence GEE (identity link)
  - for secondary outcome (senior surgeon presence), can use independence GEE + inverse cluster-size weight
- This is a simplified discussion for unadjusted estimator to illustrate Estimand-aligned considerations

Connections to current practice

Despite the rigor of counterfactual outcome framework, it seems to raise more questions than answers, especially when we connect to current practice

- **Question 1**: When or to what extent informative cluster size is an issue?
  - informative cluster sizes concerns $\{Y_{ij}(1), Y_{ij}(0), j = 1, \ldots, N_i\}$ and N_i, not other covariates
- **Question 2**: Where do we place linear mixed model and exchangeable GEE under the counterfactual outcome framework?
  - Have we been doing things wrong for decades?
  - How do we give a counterfactual interpretation for the treatment coefficient?
- **Question 3**: Are the methods in Table 1 the only valid methods for estimand-aligned analyses?

Connections to current practice - cont'd

We have some answers for two-arm parallel cluster randomized trials!

- **Question 1**: When or to what extent is informative cluster size an issue?
  - needs more empirical evidence; ongoing work led by Brennan Kahan will provide such evidence
- **Question 2**: Where do we place linear mixed model and exchangeable GEE under the counterfactual outcome framework?
  - under some **structural conditions**, both are model-robust for estimating average causal effect (focus of the next part)
- **Question 3**: Are the methods in Table 1 the only valid methods for estimand-aligned analyses?
  - no, there can be improvement, e.g. by leveraging baseline covariates without compromising the clarity of estimands

Linear mixed model

- Look at a specific candidate analysis model not mentioned in Table 1
- By far the most commonly used approach in analyzing cluster randomized trials
  - the basis of many statistical development for designing cluster randomized trials
- Is linear mixed model estimating a valid average causal effect?
  - **Goal:** demystify the structural assumptions needed for causal inference

Notation and setup

- Start by conceptualizing a cluster randomized trial with I clusters, each of which include $N^* < \infty$ individuals in its source population
  - seemingly strong assumption, but when the source population varies in size, can imagine a representative within-cluster population of size N^* to equalize the source population
- Assume $N_i \in [2, N^*]$ individuals are enrolled in the study from the source population
- The treatment is assigned at the cluster level with $A_i = 1$ if treated and 0 otherwise
- For each individual $j = 1, \cdots, N^*$ in cluster i, we define Y_{ij} as the observed outcome, and under the *consistency* assumption, $Y_{ij} = A_i Y_{ij}(1) + (1 - A_i) Y_{ij}(0)$
- $\boldsymbol{X}_{ij}$ is a p-dimensional measured vector of baseline covariates, and $\boldsymbol{U}_{ij}$ is unmeasured covariates

Structural Assumption 1: Super-population cluster sampling

- Define the complete (but not fully observed) data vector for individual j in cluster i as $\boldsymbol{W}_{ij} = (Y_{ij}(1), Y_{ij}(0), \boldsymbol{X}_{ij}, \boldsymbol{U}_{ij})$, and the complete data vector for cluster i as $\boldsymbol{W}_i = (\boldsymbol{W}_{i1}, \ldots, \boldsymbol{W}_{iN^*})$

- **Assumption** 1. (Super-population cluster sampling)
  (a) $\boldsymbol{W}_i$, $i = 1, \ldots, I$ is a random draw from a super-population of clusters. (b) Within each cluster, the element $\boldsymbol{W}_{ij}$ follows a common marginal distribution. In other words, $(\boldsymbol{W}_{i1}, \ldots, \boldsymbol{W}_{iN^*})$ are marginally identically distributed, but $(Y_{ij}(1), Y_{ij}(0), \boldsymbol{X}_{ij}, \boldsymbol{U}_{ij}, j = 1, \ldots, N^*)$ can be arbitrarily correlated.

- Assumption 1 does not assume homogeneous treatment effects across clusters (e.g. allow for random intervention effect, cluster-covariate-dependent treatment effect)
- The conditional correlation structure among individuals of the same cluster can also vary across clusters

Structural Assumption 2: Random within-cluster sampling

- Denote $\boldsymbol{M}_i = (M_{i1}, \ldots, M_{iN^*})$ as the collection of sampling indicators.
- **Assumption** 2. (Random within-cluster sampling)
  $\{\boldsymbol{M}_i, i = 1, \ldots, I\}$ are independent, identically realizations of a common population distribution , and $(\boldsymbol{M}_1, \ldots, \boldsymbol{M}_I)$ is independent of $(\boldsymbol{W}_1, \ldots, \boldsymbol{W}_I)$ and $(A_1, \ldots, A_I)$.
- Assumption 2 implies that, within each cluster, the enrollment of individuals, as well as the cluster size N_i, is random and independent of the remaining data information (potential outcomes, treatment, measured or unmeasured covariates)
- Violated when informative enrollment of individuals by treatment conditions leads to post-randomization selection bias
  - need additional assumptions and tools to address enrollment bias (Li et al. 2021, Clin Trials; Li et al. 2022, Commun Stat Theory Meth)

A common estimand

- This set up rules out informative cluster size, which is what we implicitly assume in current practice
- Individual-average and cluster-average treatment effects coincide!
  - the within-cluster average treatment effect is $\tau_i = N_i^{-1} \sum_{j=1}^{N_i} \{Y_{ij}(1) - Y_{ij}(0)\}$
  - the individual-average treatment effect $\Delta_1 = \sum_{i=1}^{I} N_i \tau_i / \sum_{i=1}^{I} N_i$
  - the cluster-average treatment effect $\Delta_2 = \sum_{i=1}^{I} \tau_i / I$
- Under **Assumptions 1 & 2**

$$\mathbb{E}[\Delta_1] = \mathbb{E}[\Delta_2] = \Delta^* = \mathbb{E}\{Y_{ij}(1) - Y_{ij}(0)\}$$

  - defined based on the source population

Linear mixed ANCOVA

- A frequently used analysis model is given by (rarely adjust for cluster size)

$$Y_{ij} = \beta_0 + \beta_A A_i + \boldsymbol{\beta}_{\boldsymbol{X}}^T \boldsymbol{X}_{ij} + \delta_i + \epsilon_{ij}$$

  - $\delta_i \sim \mathcal{N}(0, \tau^2)$ is the random effect for cluster i
  - $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ is the independent residual error
  - ICC is given by $\tau^2/(\tau^2 + \sigma^2)$

- Consider maximum likelihood estimators of $(\beta_0, \beta_A, \boldsymbol{\beta}_{\boldsymbol{X}}, \sigma^2, \tau^2)$ based on the observed data

$$\{(Y_{ij}, A_i, \boldsymbol{X}_{ij}) : j = 1, \ldots, N_i; i = 1, \ldots, I\},$$

  and denote them by $(\widehat{\beta}_0, \widehat{\beta}_A, \widehat{\boldsymbol{\beta}}_X, \widehat{\sigma}^2, \widehat{\tau}^2)$

- An unadjusted model corresponds to removing $\boldsymbol{X}_{ij}$; results still apply

- Frequently, β_A is considered as the "treatment effect" (difference scale), and we use the model-based variance estimator $\widehat{\text{Var}}(\widehat{\beta}_A)$ for inference

Main result 1: unbiased point estimator

(Wang et al. 2022+, arXiv)

- **Result 1**. (Correct point estimator)
  Under Assumptions 1 & 2, the linear mixed ANCOVA estimator $\widehat{\beta}_A$ is consistent for Δ^*, i.e., $\widehat{\beta}_A$ converges in probability to Δ^* as $m \to \infty$, and is asymptotically normal, under arbitrary misspecification of its working model.

- **take-away**: as long as structural assumptions hold, $\widehat{\beta}_A$ is unbiased for average treatment effect even if the functional form of $\boldsymbol{X}_{ij}$ is incorrect, the random intercept specification is incorrect, and/or the residual normality assumption is incorrect

Main result 2: model-based variance estimator

(Wang et al. 2022+, arXiv)

- **Result 2**. (Correct variance estimator)
  Furthermore, under equal $1:1$ *randomization,* $m\widehat{Var}(\widehat{\beta}_A)$ *converges in probability to the true asymptotic variance* v*, and therefore the model-based variance estimator* $\widehat{Var}(\widehat{\beta}_A)$ *remains valid.*

- **take-away**: as long as structural assumptions hold, we can trust the standard error estimate and p-value from standard software output even if the linear mixed ANCOVA model is wrong

- **caveat**: if randomization is not $1:1$, model-based variance estimator is not guaranteed to be valid, but can use the robust sandwich variance estimator ("`clubSandwich`" package in R and `EMPIRICAL=` option in SAS GLIMMIX)

Simulation demonstration

Two example data generating processes (DGPs)

- **DGP 1**: (common treatment effect, nonlinear X)

$$Y_{ij}(A_i) = 0.5A_i + 0.2X_{ij}^3 + 0.5X_{ij}^2 + X_{ij} - n^{-1}\sum_{k=1}^{n} X_{ik} + \delta_i + \epsilon_{ij}$$

with $\delta_i \sim \mathcal{N}(0,1)$, $\epsilon_{ij} \sim \mathcal{N}(0,25)$, $X_{ij} \sim \mathcal{N}(0,4)$, $A_i \sim \text{Bernoulli}(0.5)$

- **DGP 2**: (individual effect, nonlinear X, random cluster treatment effect)

$$Y_{ij}(0) = 0.2X_{ij}^3 + 0.5X_{ij}^2 + X_{ij} + \delta_i + \epsilon_{ij}$$

$$Y_{ij}(1) = Y_{ij}(0) + 0.5 - n^{-1}\sum_{k=1}^{n} X_{ik} + b_i X_{ij} + \zeta_{ij}$$

$\delta_i \sim \text{Laplace}(\mu = 0, b = 1)$, $\epsilon_{ij} \sim \text{DiscreteUniform}(-5, 5)$, $X_{ij} \sim \mathcal{N}(0,4)$, $A_i \sim \text{Bernoulli}(0.5)$, $b_i \sim \text{DiscreteUniform}(-1, 1)$, $\zeta_{ij} \sim \text{Laplace}(\mu = 0, b = 5)$

- Fit $Y_{ij} = \beta_0 + \beta_A A_i + \beta_X X_{ij} + \delta_i + \epsilon_{ij}$ with/without X_{ij}
- Metric: bias, empirical standard error (Emp SE), averaged model-based standard error (ASE), coverage probability (CP), relative efficiency (RE)

Simulation evidence

$N^{\max} = 35$, $N_i \sim$ Uniformly between 4 and 12, under 3000 simulations

DGP	I	Estimator	Bias	Emp SE	ASE	CP	RE
1	20	unadjusted mixed model	-0.02	6.81	7.03	95.7%	1.00
		Linear mixed ANOVA	-0.01	6.85	7.02	95.7%	0.99
	50	unadjusted mixed model	0.06	4.19	4.35	96.1%	1.00
		Linear mixed ANOVA	0.12	4.19	4.31	96.1%	1.00
	100	unadjusted mixed model	-0.01	2.99	3.02	95.5%	1.00
		Linear mixed ANOVA	0.01	2.96	3.03	95.7%	1.02
	200	unadjusted	0.01	2.03	2.13	96.2%	1.00
		Linear mixed ANOVA	0.04	2.07	2.13	95.4%	0.96
2	20	unadjusted mixed model	0.22	8.84	9.15	96.2%	1.00
		Linear mixed ANOVA	-0.08	5.65	5.58	96.1%	2.45
	50	unadjusted mixed model	-0.08	5.55	5.66	95.3%	1.00
		Linear mixed ANOVA	-0.01	3.44	3.49	96.7%	2.60
	100	unadjusted mixed model	-0.02	3.88	3.95	95.1%	1.00
		Linear mixed ANOVA	0.04	2.45	2.47	95.4%	2.51
	200	unadjusted mixed model	-0.05	2.67	2.76	95.8%	1.00
		Linear mixed ANOVA	0.05	1.72	1.73	95.4%	2.41

Remarks

- Robustness of linear mixed ANCOVA models was demonstrated earlier in cluster randomized trials, but only through simulations
  - e.g., Murray et al. (2006) demonstrated that $\widehat{\beta}_A$ had correct type I error rate when the data were simulated from an ANCOVA model with non-normal random effect and/or residual errors
  - we provide a formal justification, and even provide a green light to the use of model-based variance estimator
  - linear mixed ANCOVA model application to CRTs is more robust than we originally thought
- Cannot formally prove results under REML, but simulations show no difference
- Elegant parallel to linear regression ANCOVA under individually randomized trials (Yang and Tsiatis, 2001; Lin, 2013; Wang et al., 2019 Biometrics)

Stratified randomization

(Wang et al. 2022+, arXiv)

- Let S_i be a categorical variable that encodes the randomization strata $\mathcal{S}$
  - $\mathcal{S} = \{\text{urban, rural}\}$, $S_i \in \mathcal{S}$
  - assume that fixed number of strata
  - randomization proportion within strata remains $\pi \in (0, 1)$
- **Result 3**: (Correct point estimator)

  *Under Assumptions 1 & 2 and stratified randomization, the linear mixed ANCOVA estimator $\widehat{\beta}_A$ is still consistent for Δ^* and asymptotically normal with a **smaller** asymptotic variance than without stratification, under arbitrary misspecification of its working model.*
- **Result 4**: (Correct variance estimator)

  If $\pi = 0.5$, adjusting for S_i in the linear mixed model as cluster-level dummy variables leads to a consistent model-based variance estimator.

Summary

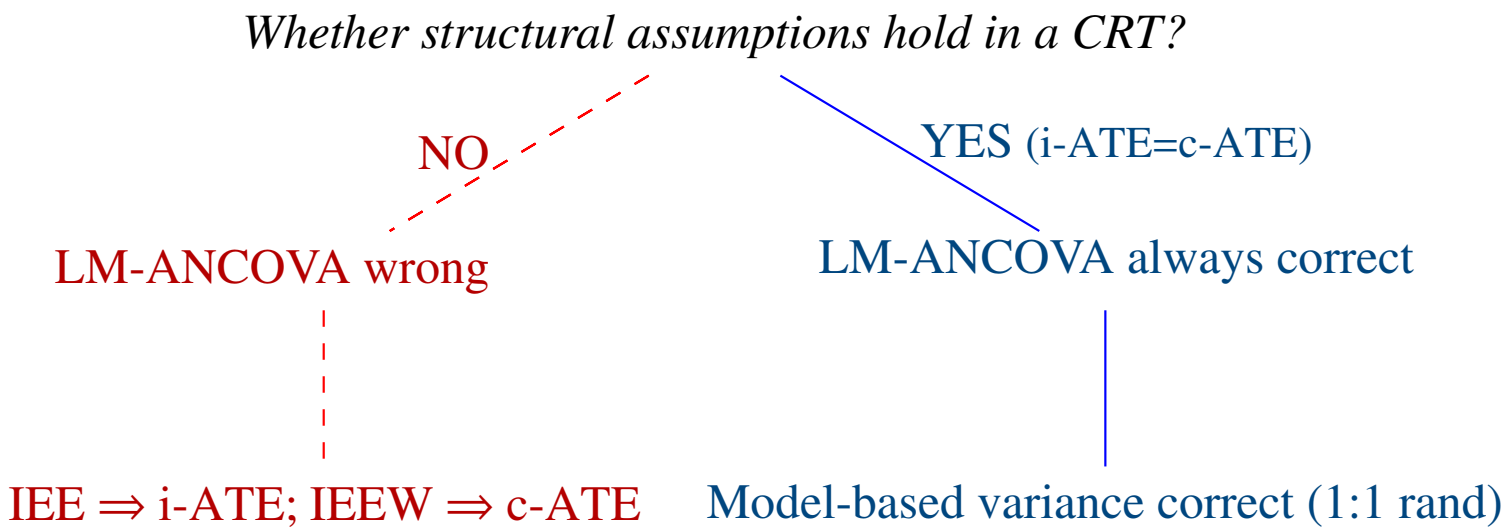
- **Message 1**: When informative cluster size arises, individual-average treatment effect and cluster-average treatment effect answer different questions in cluster randomized trials

> "*the target of inference in such studies (CRTs) could be at either the individual level or community level . . . These examples show the importance of investigators explicitly formulating and stating the hypothesis under test in cluster randomization trials.*" — p.13 in Donner and Klar (2000)

  - estimand should drive the statistical analysis, not vice versa
  - think through estimands and potential for informative cluster size
  - consider "sensitivity analysis"?
  - many open questions

Summary - Cont'd

- **Message 2**: Under **non-informative cluster size** and **random enrollment**, the two estimands **coincide**. In this case, standard linear mixed ANCOVA is consistent for the average treatment effect estimand under arbitrary model specification.
  - **caveat**: there are structural assumptions we implicitly make but rarely acknowledge



  - standard linear mixed model is very robust to modeling assumptions, but require the structural assumptions that we need to be aware of

Acknowledgement

- This talk is primarily based on
  - Kahan, BC, Li F, Copas AJ, Harhay MO. (2022). Estimands in cluster-randomized trials: choosing analyses that answer the right question. *International Journal of Epidemiology*.
  - Wang X, Turner EL, Li F, Wang R, Moyer J, Cook AJ, Murray DM, Heagerty PJ. (2022). Two weights make a wrong: Cluster randomized trials with variable cluster sizes and heterogeneous treatment effects. *Contemporary Clinical Trials*, 114, 106702.
  - Wang B, Harhay MO, Small DS, Morris TP, Li F. (2021+). On the mixed-model analysis of covariance in cluster-randomized trials. *preprint arXiv:2112.00832*.
- Research is supported by Patient-Centered Outcomes Research Institute
  - Award ME-2019C1-16146 (PI: F Li)
  - Award ME-2020C1-19220 (PI: MO Harhay)

  also grateful for support from NIH Pragmatic Trials Collaboratory & NIA IMPACT Collaboratory
- Thanks to Jiaqi Tong, Xueqi Wang, Bingkai Wang, Brennan Kahan, Andrew Copas, Monica Taljaard, Derek Ouyang and Michael Harhay for help and comments

Possible discussion

- **Question 1**: can we use linear mixed ANCOVA model to estimate i-ATE and c-ATE in the presence of informative cluster size?
  - yes but need adjustment(s) for cluster size; see Wang et al (2022, arXiv)
- **Question 2**: what about binary outcome and ratio estimands?
  - yes but need weighted g-computation formula (also referred to as standardization); see Wang et al (2022, arXiv)
- **Question 3**: what about generalized linear mixed models?
  - not quite, GLMM is in general not model-robust; see Wang et al (2022, arXiv)
  - Wang B, Park C, Small DS, Li F. (2022+). Model-robust and efficient inference for cluster-randomized experiments. preprint arXiv:2210.07324
- **Question 4**: how about non-random enrollment?
  - depends on the enrollment assumptions, see Li et al. (2021, Clin Trials); Li et al. (2022, Commun Stat Theory Meth)
- **Question 5**: what about other study designs?
  - ongoing work with promising preliminary evidence for stepped wedge trials

References

1. Donner A. and Klar N. (2000). Design and Analysis of Cluster Randomization Trials in Health Research. *Wiley*.

2. Murray, D. M. (1998). Design and analysis of group-randomized trials. *New York:Oxford University Press*.

3. Imai, K., King, G., Nall, C. (2009). The essential role of pair matching in cluster-randomized experiments, with application to the Mexican universal health insurance evaluation. *Statistical Science*, 24(1), 29-53.

4. Balzer LB, Zheng W, van der Laan MJ, Petersen ML. A new approach to hierarchical data analysis: Targeted maximum likelihood estimation for the causal effect of a cluster-level exposure. *Statistical Methods in Medical Research*. 2019;28(6):1761-1780.

5. Kahan, B. C., Li, F., Copas, A. J., Harhay, M. O. (2022). Estimands in cluster-randomized trials: choosing analyses that answer the right question. *International Journal of Epidemiology*.

6. Wang, X., Turner, E. L., Li, F., Wang, R., Moyer, J., Cook, A. J., Murray, D. M., Heagerty, P. J. (2022). Two weights make a wrong: Cluster randomized trials with variable cluster sizes and heterogeneous treatment effects. *Contemporary clinical trials*, 114, 106702.

7. Li, F., Tian, Z., Tian, Z., Li, F. (2022). A note on identification of causal effects in cluster randomized trials with post-randomization selection bias. *Communications in Statistics-Theory and Methods*, 1-13.

8.Li F, Tian Z, Bobb J, Papadogeorgou G, Li F. Clarifying selection bias in cluster randomized trials. *Clinical Trials*. 2022;19(1):33-41.

References - Cont'd

9. Lin, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining Freedmanâs critique. *The Annals of Applied Statistics*, 7(1), 295-318.

10. Wang, B., Ogburn, E. L., Rosenblum, M. (2019). Analysis of covariance in randomized trials: More precision and valid confidence intervals, without model assumptions. *Biometrics*, 75(4), 1391-1400

11. Yang L, Tsiatis AA. (2001). Efficiency study of estimators for a treatment effect in a pretest-posttest trial. The American Statistician, 55(4), 314-321.

12. Murray, D. M., Hannan, P. J., Pals, S. P., McCowen, R. G., Baker, W. L., and Blitstein, J. L. (2006). A comparison of permutation and mixed-model regression methods for the analysis of simulated data in the context of a group-randomized trial. *Statistics in Medicine*, 25(3):375-388.

13. Wang B, Harhay MO, Small DS, Morris TP, Li F. (2021+). On the mixed-model analysis of covariance in cluster-randomized trials. arXiv preprint arXiv:2112.00832.

14. Su, F. and Ding, P. (2021). Model-assisted analyses of cluster-randomized experiments. Journal of the Royal Statistical Society, Series B, 83(5):994-1015.

15. Bugni, F., Canay, I., Shaikh, A., and Tabord-Meehan, M. (2022). Inference for cluster randomized experiments with non-ignorable cluster sizes. arXiv preprint arXiv:2204.08356.

16. Wang B, Park C, Small DS, Li F. (2022+). Model-robust and efficient inference for cluster-randomized experiments. arXiv preprint arXiv:2210.07324