

Methods: Mind the Gap
Webinar Series

Implications of Informative Cluster Size for the Design and Analysis of Cluster Randomized Trials

Presented by:
Brennan Kahan, Ph.D.
University College London



Outline

- **Background**
- **Informative cluster size**
- **Impact on estimands**
- **Impact on analysis**
- **What this means for CRTs**

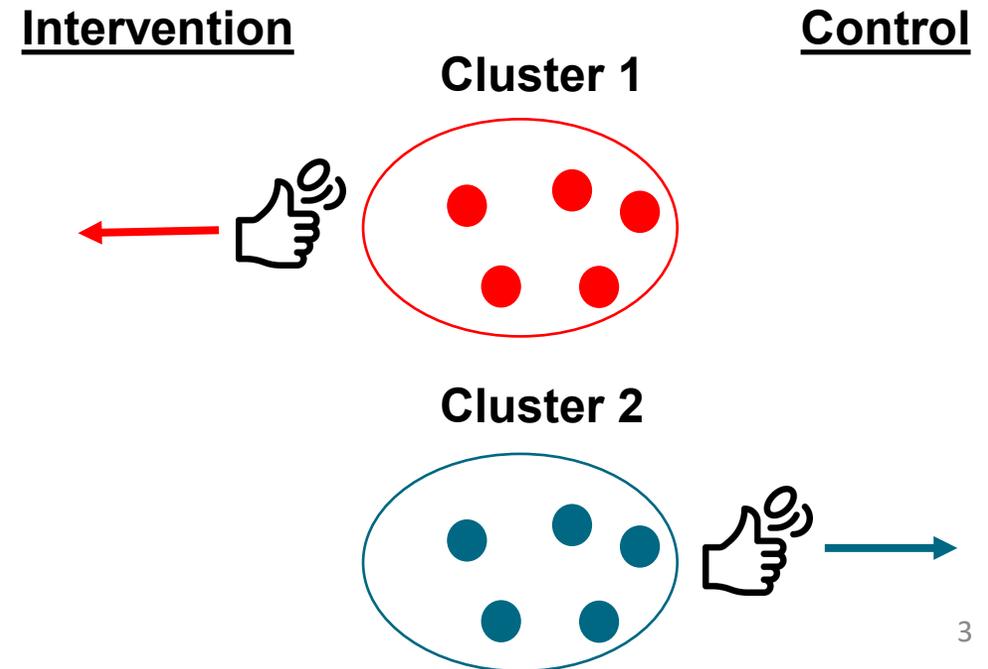
Cluster randomised trials

Groups of participants, such as hospitals, schools, or villages, are randomised

- All enrolled participants within a cluster receive the **same** intervention

Motivation for cluster randomisation:

- Intervention implemented at cluster level
- Avoid risk of contamination



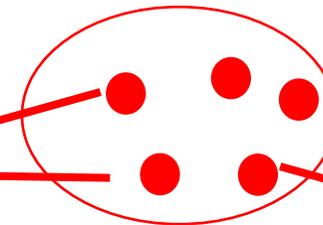
A key challenge (1 of 2)

Patients within a cluster are **correlated**

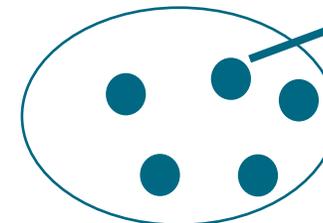
Within-cluster correlation

Two patients in the same cluster are more similar than...

Cluster 1



Cluster 2



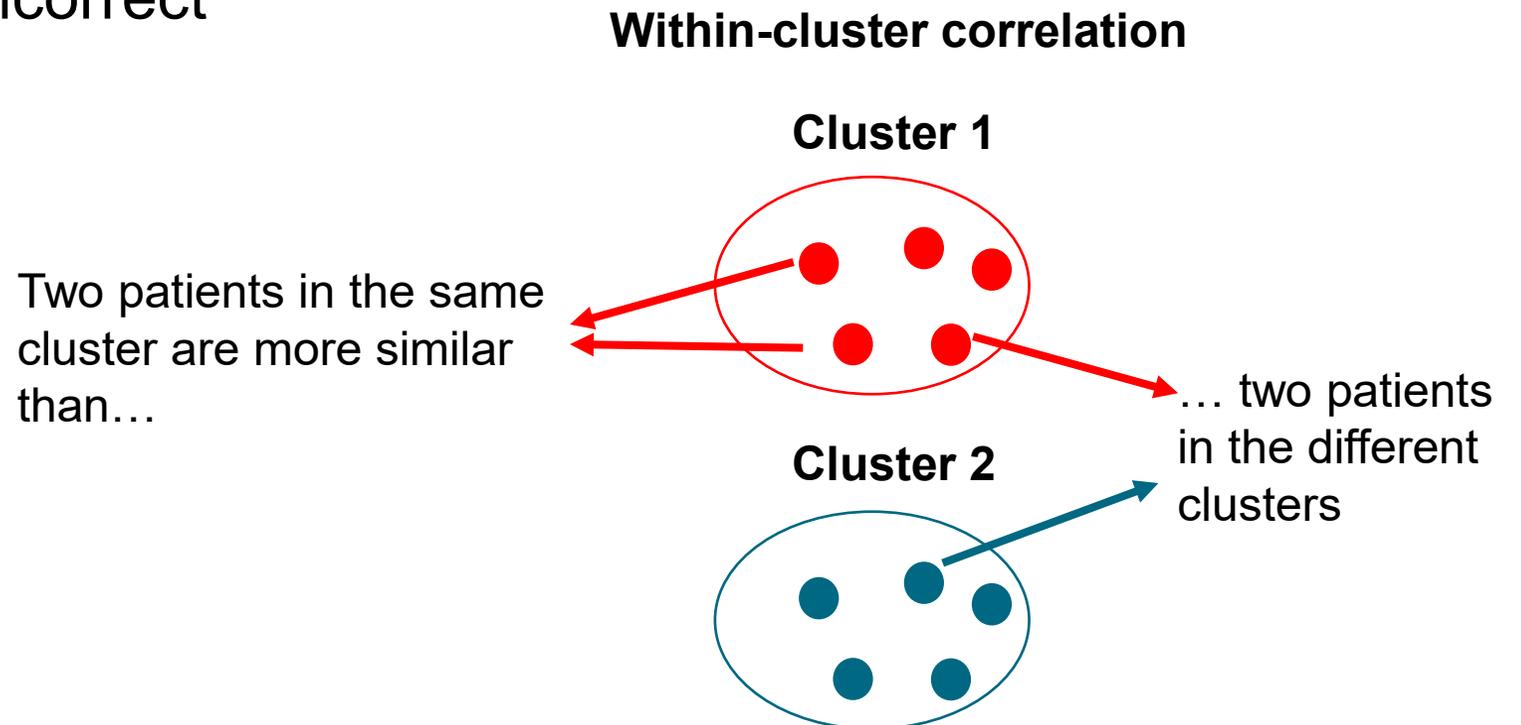
... two patients in the different clusters

A key challenge (2 of 2)

Patients within a cluster are **correlated**

Standard statistical assumption of independence is **violated**

- Need to account for this during analysis
- Otherwise results may be incorrect



Statistical methods for CRTs (1 of 6)

Many analysis methods which account for clustering¹

- GEE (exch correlation)
- Mixed-effects models
- Analysis of cluster-level summaries
- Independence estimating equations (IEEs)
- Many others

Statistical methods for CRTs (2 of 6)

Many analysis methods which account for clustering¹

- GEE (exch correlation)
 - Mixed-effects models
 - Analysis of cluster-level summaries
 - Independence estimating equations (IEEs)
 - Many others
- } Account for within-cluster correlation directly

Statistical methods for CRTs (3 of 6)

Many analysis methods which account for clustering¹

- GEE (exch correlation)
- Mixed-effects models
- Analysis of cluster-level summaries } Takes average outcome in each cluster
Analyses these “cluster-level” summaries
- Independence estimating equations (IEEs)
- Many others

Statistical methods for CRTs (4 of 6)

Many analysis methods which account for clustering¹

- GEE (exch correlation)
- Mixed-effects models
- Analysis of cluster-level summaries
- Independence estimating equations (IEEs) } Uses “working” assumption of independence
- Many others } Uses “cluster-robust” SEs

Statistical methods for CRTs (5 of 6)

Many analysis methods which account for clustering¹

- GEE (exch correlation)
- Mixed-effects models
- Analysis of cluster-level summaries
- Independence estimating equations (IEEs)
- Many others

For each, can use different **implementations** (e.g. different **weights**)

Statistical methods for CRTs (6 of 6)

Many analysis methods which account for clustering¹

- GEE (exch correlation)
- Mixed-effects models
- Analysis of cluster-level summaries
- Independence estimating equations (IEEs)
- Many others

For each, can use different **implementations** (e.g. different **weights**)

However, need to consider impact of **informative cluster size (ICS)**

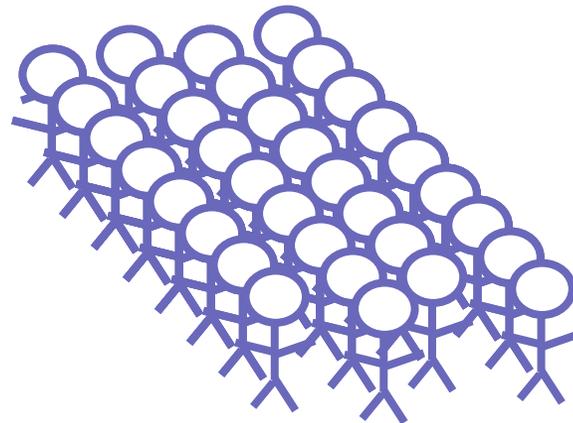
Informative cluster size (1 of 3)

The size of a cluster informs the outcomes/treatment effects with that cluster

Cluster 1



Cluster 2



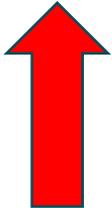
Informative cluster size (2 of 3)

The size of a cluster informs the outcomes/treatment effects with that cluster

For example:

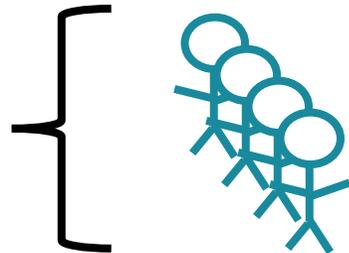
- Patients in larger clusters have worse outcomes than those in smaller clusters

Better outcomes

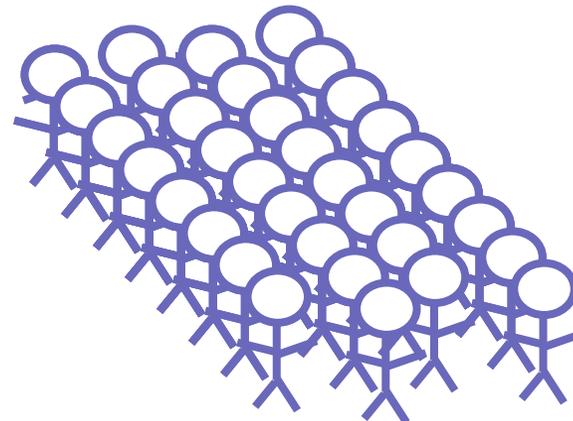


20% chance of event

Cluster 1



Cluster 2



Worse outcomes



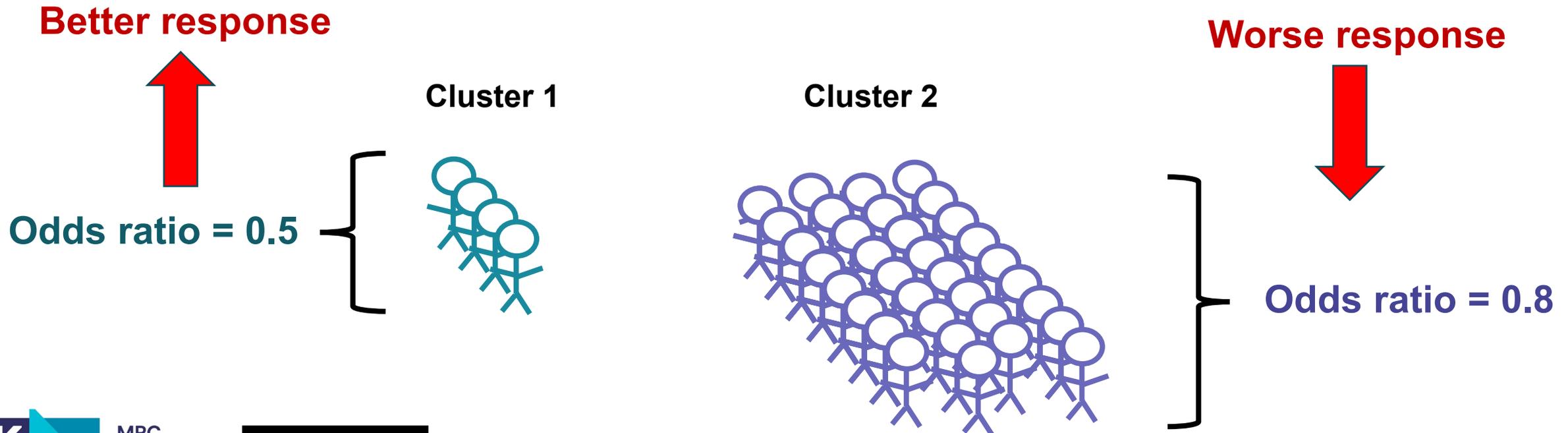
50% chance of event

Informative cluster size (3 of 3)

The size of a cluster informs the outcomes/treatment effects with that cluster

For example:

- Patients in larger clusters have worse outcomes than those in smaller clusters
- Patients in larger clusters respond better to treatment than those in smaller clusters



Causes of informative cluster size (1 of 2)

Can occur for different reasons

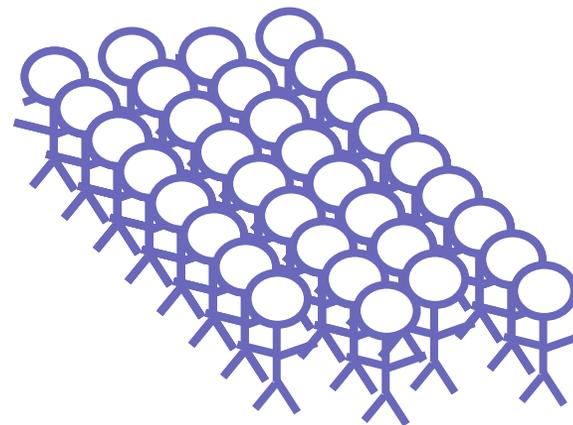
- Different types of patients present to larger vs. smaller clusters

**Higher
socioeconomic
status**

Cluster 1



Cluster 2

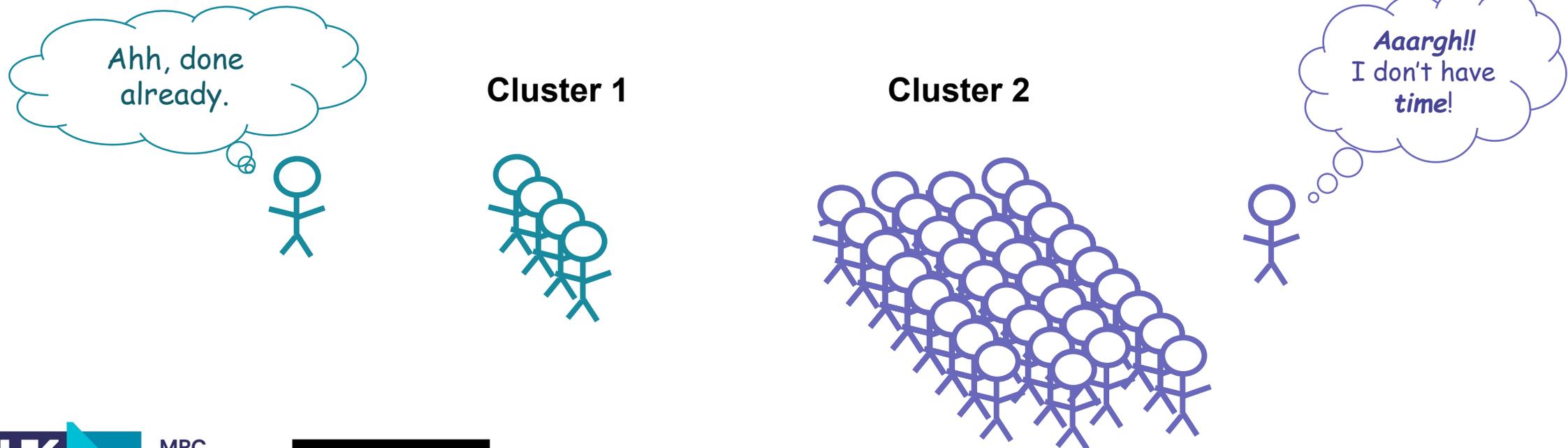


**Lower
socioeconomic
status**

Causes of informative cluster size (2 of 2)

Can occur for different reasons

- Different types of patients present to larger vs. smaller clusters
- Different implementation of intervention between larger vs. smaller clusters



Why informative cluster size matters (1 of 2)

Can change the **trial question** being addressed (estimand)

Can introduce **bias** for some analysis methods

Why informative cluster size matters (2 of 2)

Can change the **trial question** being addressed (estimand)

Can introduce **bias** for some analysis methods

Need to consider how ICS might be affecting trial **conclusions**

Example – the RESTORE trial (1 of 4)

CRT comparing protocolised sedation vs. usual care in critically ill children who were mechanically ventilated for acute respiratory failure

2449 participants enrolled from 31 clusters

- Range in cluster size: 12 to 272

One of safety outcomes was ‘post-extubation stridor’

- Denotes presence of inspiratory noise indicating narrowing of airway

Example – the RESTORE trial (2 of 4)

Based on analysis choice, results may show:

- Evidence of harm (OR=1.83, p=0.03); or
- Little effect (OR 1.38, p=0.17)

Method	Odds ratio (95% CI)	P-value
IEEs (unweighted)	1.65 (1.02 to 2.67)	0.04
GEE (exch. corr)	1.57 (0.98 to 2.50)	0.06
Cluster-level (weighted by cluster size)	1.83 (1.06 to 3.14)	0.03
Mixed-effects model	1.54 (0.97 to 2.44)	0.07
IEEs (weighted by 1/cluster size)	1.38 (0.87 to 2.19)	0.17
Cluster-level (unweighted)	1.51 (0.90 to 2.52)	0.11

Example – the RESTORE trial (3 of 4)

Based on analysis choice, results may show:

- Evidence of harm (OR=1.83, p=0.03); or
- Little effect (OR 1.38, p=0.17)

Informative cluster size is:

- Changing the target estimand
- Introducing bias for some methods

Method	Odds ratio (95% CI)	P-value
IEEs (unweighted)	1.65 (1.02 to 2.67)	0.04
GEE (exch. corr)	1.57 (0.98 to 2.50)	0.06
Cluster-level (weighted by cluster size)	1.83 (1.06 to 3.14)	0.03
Mixed-effects model	1.54 (0.97 to 2.44)	0.07
IEEs (weighted by 1/cluster size)	1.38 (0.87 to 2.19)	0.17
Cluster-level (unweighted)	1.51 (0.90 to 2.52)	0.11

Example – the RESTORE trial (4 of 4)

Based on analysis choice, results may show:

- Evidence of harm (OR=1.83, p=0.03); or
- Little effect (OR 1.38, p=0.17)

Informative cluster size is:

- Changing the target estimand
- Introducing bias for some methods

Or, to put it another way...

Under ICS, we need to think more carefully about:

- Our research question
- Our choice of analysis

Method	Odds ratio (95% CI)	P-value
IEEs (unweighted)	1.65 (1.02 to 2.67)	0.04
GEE (exch. corr)	1.57 (0.98 to 2.50)	0.06
Cluster-level (weighted by cluster size)	1.83 (1.06 to 3.14)	0.03
Mixed-effects model	1.54 (0.97 to 2.44)	0.07
IEEs (weighted by 1/cluster size)	1.38 (0.87 to 2.19)	0.17
Cluster-level (unweighted)	1.51 (0.90 to 2.52)	0.11

Impact on research question (estimand)

Estimands (1 of 2)

An estimand = precise description of our **research question**¹

- What is our study trying to estimate?

Attribute	Description
Population	Adults with uncontrolled persistent asthma
Treatment conditions	Dupilumab vs. placebo every 2 weeks
Endpoint	FEV ₁ at week 12
Summary measure	Difference in means
Intercurrent events	Stopping treatment early – hypothetical if hadn't stopped Receipt of rescue therapy – hypothetical if hadn't received rescue

¹ Kahan BC, Hindley J, Edwards M, Cro S, Morris TP. The estimands framework: a primer on the ICH E9(R1) addendum *BMJ* 2024; 384 :e076316 doi:10.1136/bmj-2023-076316

Estimands (2 of 2)

An estimand = precise description of our **research question**¹

- What is our study trying to estimate?

Aims

- **Clarity** - tell **others** what your estimating
- **Better methods** – ensure statistical methods **aligned** to research question

Attribute	Description
Population	Adults with uncontrolled persistent asthma
Treatment conditions	Dupilumab vs. placebo every 2 weeks
Endpoint	FEV ₁ at week 12
Summary measure	Difference in means
Intercurrent events	Stopping treatment early – hypothetical if hadn't stopped Receipt of rescue therapy – hypothetical if hadn't received rescue

¹ Kahan BC, Hindley J, Edwards M, Cro S, Morris TP. The estimands framework: a primer on the ICH E9(R1) addendum *BMJ* 2024; 384 :e076316 doi:10.1136/bmj-2023-076316

Estimands in CRTs (1 of 2)

CRTs can address **different** types of questions

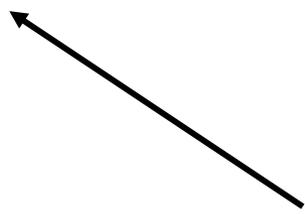
- Marginal vs. cluster-specific

Estimands in CRTs (2 of 2)

CRTs can address **different** types of questions

- Marginal vs. cluster-specific
- **Participant- vs. cluster-average**

Impacted by ICS



Participant- vs. cluster-average effects

(1 of 4)

Difference is in how individuals are **weighted**¹⁻³

	Cluster size	Avg Difference in Potential Outcomes
Cluster 1	10	2
Cluster 2	10	2
Cluster 3	100	5
Cluster 4	100	5

¹ Kahan BC, Li F, Copas AJ, Harhay MO, Estimands in cluster-randomized trials: choosing analyses that answer the right question, *International Journal of Epidemiology*, 2023.

² Kahan BC, Blette BS, Harhay MO, et al. Demystifying estimands in cluster-randomised trials. *Statistical Methods in Medical Research*. 2024.

³ Wang X, Turner EL, Li F et al. Two weights make a wrong: cluster randomized trials with variable cluster sizes and heterogeneous treatment effects. *Contemp Clin Trials* 2022.

Participant- vs. cluster-average effects

(2 of 4)

Participant-average treatment effect

- Each patients gets equal weight
- “How are test scores for my **child** likely to rise?”

$$\frac{(2)(10)(2) + (2)(100)(5)}{10 + 10 + 100 + 100} = 4.7$$

	Cluster size	Avg Difference in Potential Outcomes
Cluster 1	10	2
Cluster 2	10	2
Cluster 3	100	5
Cluster 4	100	5

Participant- vs. cluster-average effects

(3 of 4)

Cluster-average treatment effect

- Clusters get equal weight
- “How much are test scores at my **school** likely to rise?”

$$\frac{(2)(2) + (2)(5)}{4} = 3.5$$

	Cluster size	Avg Difference in Potential Outcomes
Cluster 1	10	2
Cluster 2	10	2
Cluster 3	100	5
Cluster 4	100	5

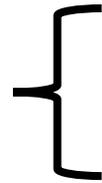
Participant- vs. cluster-average effects

(4 of 4)

When there is **ICS**, these effects are **different**

- Participant-average effect = 4.7
- Cluster-average effect = 3.5

Cluster-average effect -> more weight



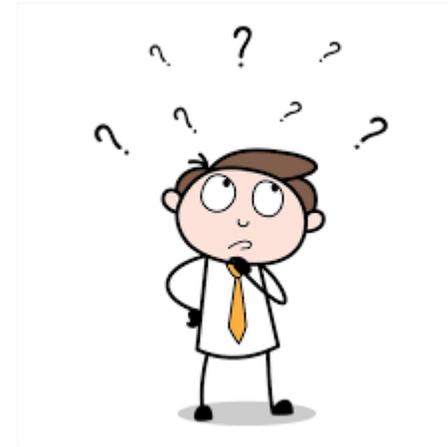
	Cluster size	Avg Difference in Potential Outcomes
Cluster 1	10	2
Cluster 2	10	2
Cluster 3	100	5
Cluster 4	100	5

Implication (1 of 4)

Many different methods of analysis for CRTs

- GEE (exch correlation)
- Mixed-effects models
- Analysis of cluster-level summaries
- Independence estimating equations (IEEs)
- Others

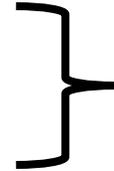
Which method is targeting which estimand?



Implication (2 of 4)

Many different methods of analysis for CRTs

- GEE (exch correlation)
- Mixed-effects models
- Analysis of cluster-level summaries*
- Independence estimating equations (IEEs)
- Others



Participant-average effects

Implication (3 of 4)

Many different methods of analysis for CRTs

- GEE (exch correlation)
 - Mixed-effects models
 - Analysis of cluster-level summaries
 - Independence estimating equations (IEEs)*
 - Others
- } **Cluster-average effects**

Implication (4 of 4)

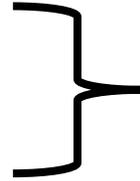
Many different methods of analysis for CRTs

- GEE (exch correlation)
 - Mixed-effects models
 - Analysis of cluster-level summaries
 - Independence estimating equations (IEEs)
 - Others
- } **What about these?**

Impact on bias

Impact of ICS on bias (1 of 2)

- GEE (exch correlation)
- Mixed-effects models



Do not give equal to **either** patients or clusters¹⁻³

(Typically) weight by **inverse variance**

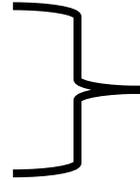
¹ Kahan BC, Li F, Copas AJ, Harhay MO, Estimands in cluster-randomized trials: choosing analyses that answer the right question, *International Journal of Epidemiology*, 2023.

² Kahan BC, Blette BS, Harhay MO, et al. Demystifying estimands in cluster-randomised trials. *Statistical Methods in Medical Research*. 2024.

³ Wang X, Turner EL, Li F et al. Two weights make a wrong: cluster randomized trials with variable cluster sizes and heterogeneous treatment effects. *Contemp Clin Trials* 2022.

Impact of ICS on bias (2 of 2)

- GEE (exch correlation)
- Mixed-effects models



Do not give equal to **either** patients or clusters¹⁻³

(Typically) weight by **inverse variance**

They **upweight** effects from smaller clusters and **down weight** effects from larger clusters

Unbiased if effects are the **same** in both small and large clusters (i.e. **no ICS**)

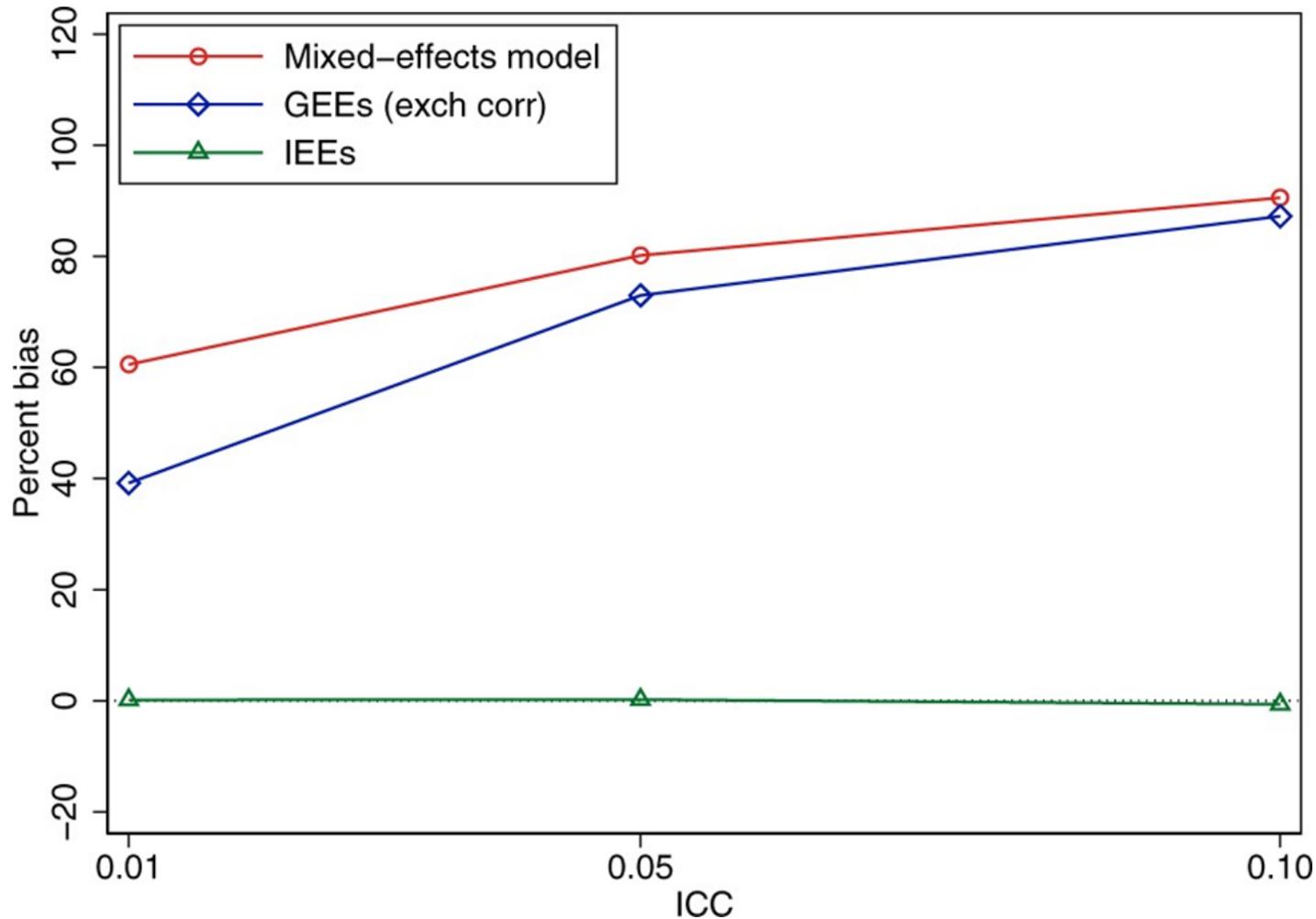
But **biased** under **ICS** for **both** the participant- and cluster-average effects

¹ Kahan BC, Li F, Copas AJ, Harhay MO, Estimands in cluster-randomized trials: choosing analyses that answer the right question, *International Journal of Epidemiology*, 2023.

² Kahan BC, Blette BS, Harhay MO, et al. Demystifying estimands in cluster-randomised trials. *Statistical Methods in Medical Research*. 2024.

³ Wang X, Turner EL, Li F et al. Two weights make a wrong: cluster randomized trials with variable cluster sizes and heterogeneous treatment effects. *Contemp Clin Trials* 2022.

Example of bias for participant-average effect



Simulated example based on:

- 30 clusters with N=100 (treatment effect = 1)
- 30 with N=10 (treatment effect = 5)

Kahan BC, Li F, Copas AJ, Harhay MO, Estimands in cluster-randomized trials: choosing analyses that answer the right question, *International Journal of Epidemiology*, Volume 52, Issue 1, February 2023, Pages 107–118.

Returning to the RESTORE trial (1 of 3)

Estimand	Method	Odds ratio (95% CI)	P-value
Participant-average (marginal)	IEEs (unweighted)	1.65 (1.02 to 2.67)	0.04
	GEE (exch. corr)	1.57 (0.98 to 2.50)	0.06
Participant-average (cluster specific)	Cluster-level (weighted by cluster size)	1.83 (1.06 to 3.14)	0.03
	Mixed-effects model	1.54 (0.97 to 2.44)	0.07
Cluster-average (marginal)	IEEs (weighted by 1/cluster size)	1.38 (0.87 to 2.19)	0.17
Cluster-average (cluster-specific)	Cluster-level (unweighted)	1.51 (0.90 to 2.52)	0.11

Marginal, participant-average odds ratio = 1.15 (0.56 to 2.36) in clusters of size <100, and 2.26 (1.49 to 3.43) in clusters of size ≥100. P-value for interaction between treatment and cluster size ≥100 = 0.21.

Returning to the RESTORE trial (2 of 3)

Estimand	Method	Odds ratio (95% CI)	P-value
Participant-average (marginal)	IEEs (unweighted)	1.65 (1.02 to 2.67)	0.04
	GEE (exch. corr)	1.57 (0.98 to 2.50)	0.06
Participant-average (cluster specific)	Cluster-level (weighted by cluster size)	1.83 (1.06 to 3.14)	0.03
	Mixed-effects model	1.54 (0.97 to 2.44)	0.07
Cluster-average (marginal)	IEEs (weighted by 1/cluster size)	1.38 (0.87 to 2.19)	0.17
Cluster-average (cluster-specific)	Cluster-level (unweighted)	1.51 (0.90 to 2.52)	0.11

Marginal, participant-average odds ratio = 1.15 (0.56 to 2.36) in clusters of size <100, and 2.26 (1.49 to 3.43) in clusters of size ≥100. P-value for interaction between treatment and cluster size ≥100 = 0.21.

Returning to the RESTORE trial (3 of 3)

Estimand	Method	Odds ratio (95% CI)	P-value
Participant-average (marginal)	IEEs (unweighted)	1.65 (1.02 to 2.67)	0.04
	GEE (exch. corr)	1.57 (0.98 to 2.50)	0.06
Participant-average (cluster specific)	Cluster-level (weighted by cluster size)	1.83 (1.06 to 3.14)	0.03
	Mixed-effects model	1.54 (0.97 to 2.44)	0.07
Cluster-average (marginal)	IEEs (weighted by 1/cluster size)	1.38 (0.87 to 2.19)	0.17
Cluster-average (cluster-specific)	Cluster-level (unweighted)	1.51 (0.90 to 2.52)	0.11

Marginal, participant-average odds ratio = 1.15 (0.56 to 2.36) in clusters of size <100, and 2.26 (1.49 to 3.43) in clusters of size ≥100. P-value for interaction between treatment and cluster size ≥100 = 0.21.

What does this mean for CRTs?

Implications of ICS

Aim: ensure CRTs are getting **good** answers to the **right** questions

ICS poses challenges for **both** parts of this

Choice of estimand

What is the right estimand?

- Participant- vs. cluster-average
- No one right answer; will depend on study objectives



Choice of analysis (1 of 6)

Need to analysis approach that aligns with our **estimand**

- Shouldn't use cluster-average estimator for participant-average effect

➡ Bias

Choice of analysis (2 of 6)

Need to analysis approach that aligns with our **estimand**

- Shouldn't use cluster-average estimator for participant-average effect
➡ Bias

Need to consider ICS in choice of analysis method

- Is it likely? ➡ Mixed-models/GEEs = biased

Choice of analysis (3 of 6)

Need to analysis approach that aligns with our **estimand**

- Shouldn't use cluster-average estimator for participant-average effect
➡ Bias

Need to consider ICS in choice of analysis method

- Is it likely? ➡ Mixed-models/GEEs = biased

Use IEEs/cluster-level summaries instead

Choice of analysis (4 of 6)

Need to analysis approach that aligns with our **estimand**

- Shouldn't use cluster-average estimator for participant-average effect
➡ Bias

Need to consider ICS in choice of analysis method

- Is it likely? ➡ Mixed-models/GEEs = biased
- Unlikely ➡ Mixed-models/GEEs = OK

Use IEEs/cluster-level summaries instead

Choice of analysis (5 of 6)

Need to analysis approach that aligns with our **estimand**

- Shouldn't use cluster-average estimator for participant-average effect
➡ Bias

Need to consider ICS in choice of analysis method

- Is it likely? ➡ Mixed-models/GEEs = biased
- Unlikely ➡ Mixed-models/GEEs = OK
- But can we **guarantee** no ICS at design stage?

Use IEEs/cluster-level summaries instead

Choice of analysis (6 of 6)

Need to analysis approach that aligns with our **estimand**

- Shouldn't use cluster-average estimator for participant-average effect
➡ Bias

Need to consider ICS in choice of analysis method

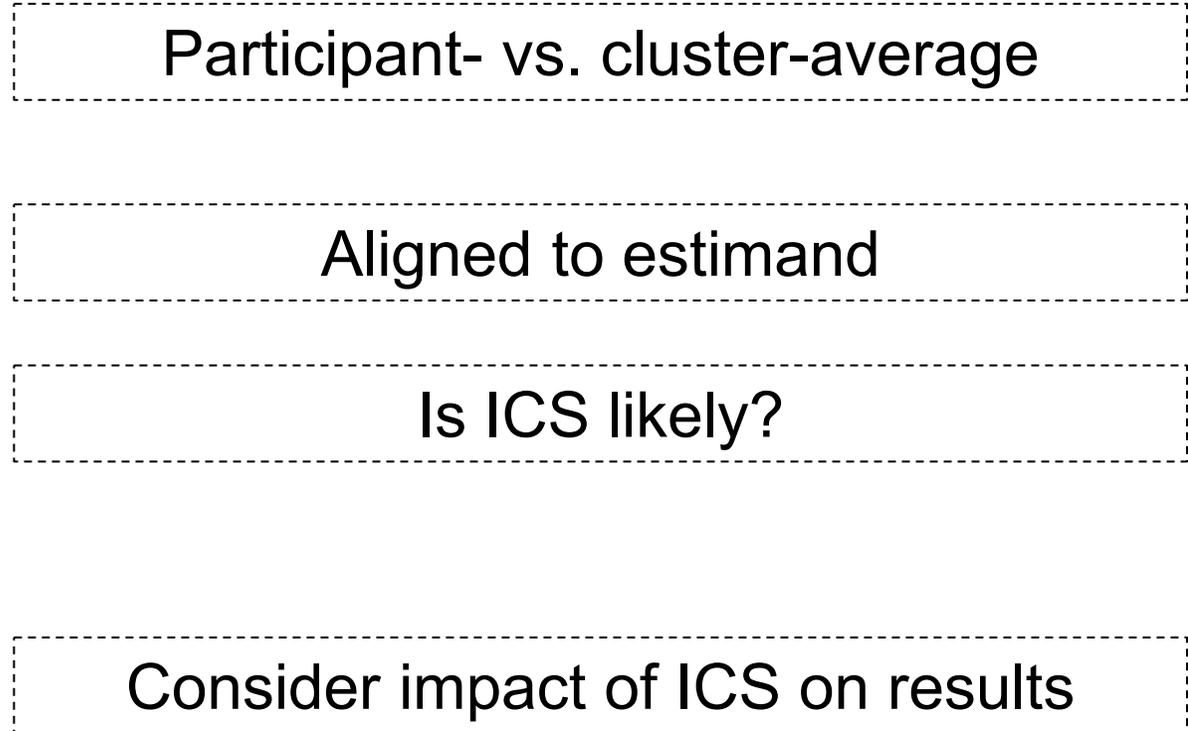
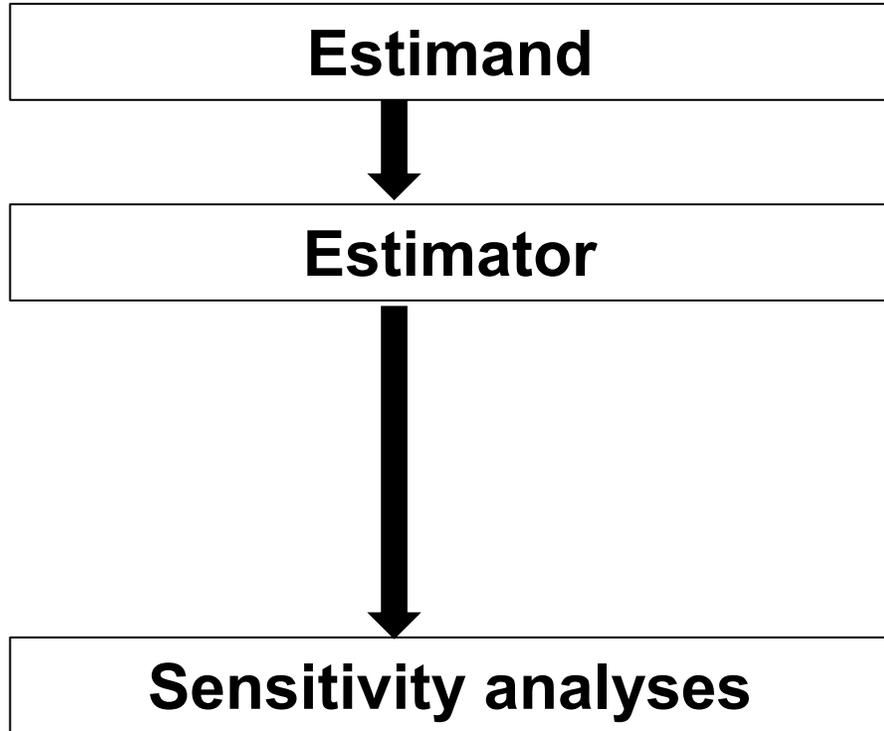
- Is it likely? ➡ Mixed-models/GEEs = biased
- Unlikely ➡ Mixed-models/GEEs = OK

- But can we **guarantee** no ICS at design stage?

Use IEEs/cluster-level summaries instead

Sensitivity analyses based on IEEs/cluster-level may be useful

Summary



Resources (1 of 2)

Describes informative cluster size:

- What it is
- What might cause it
- Its impact on analysis methods

Examples on how to choose between estimands

More on RCTs

Estimands in cluster-randomized trials: choosing analyses that answer the right question

Brennan C Kahan,^{1*} Fan Li ,^{2,3} Andrew J Copas¹ and Michael O Harhay ^{4,5}

¹MRC Clinical Trials Unit at UCL, Institute of Clinical Trials and Methodology, London, UK, ²Department of Biostatistics, Yale University School of Public Health, New Haven, CT, USA, ³Center for Methods in Implementation and Prevention Science, Yale University School of Public Health, New Haven, CT, USA, ⁴Clinical Trials Methods and Outcomes Lab, PAIR (Palliative and Advanced Illness Research) Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA and ⁵Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

*Corresponding author. MRC Clinical Trials Unit at UCL, Institute of Clinical Trials and Methodology, 90 High Holborn, London, WC1V 6LJ, UK. E-mail: b.kahan@ucl.ac.uk

Resources (2 of 2)

Helps match analysis method to estimand

- Covers each method from this talk
- Describes assumptions (including whether affected by ICS)

Original Research Article

Demystifying estimands in cluster-randomised trials

Brennan C Kahan¹ , Bryan S Blette² ,
Michael O Harhay^{1,3}, Scott D Halpern³, Vipul Jairath^{4,5},
Andrew Copas^{1,*} and Fan Li^{6,7,*} 

Abstract

Estimands can help clarify the interpretation of treatment effects and ensure that estimators are aligned with the study's objectives. Cluster-randomised trials require additional attributes to be defined within the estimand compared to individually randomised trials, including whether treatment effects are *marginal* or *cluster-specific*, and whether they are *participant-* or *cluster-average*. In this paper, we provide formal definitions of estimands encompassing both these attributes using potential outcomes notation and describe differences between them. We then provide an overview of estimators for each estimand, describe their assumptions, and show consistency (i.e. asymptotically unbiased estimation) for a series of analyses based on cluster-level summaries. Then, through a re-analysis of a published cluster-randomised trial, we demonstrate that the choice of both estimand and estimator can affect interpretation. For instance, the estimated odds ratio ranged from 1.38 ($p = 0.17$) to 1.83 ($p = 0.03$) depending on the target estimand, and for some estimands, the choice



Statistical Methods in Medical Research
2024, Vol. 33(7) 1211–1232
© The Author(s) 2024



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/09622802241254197
journals.sagepub.com/home/smm



Table 3. Overview of estimators for each estimand.

Estimand	Description	Estimator	Description	Assumptions
Marginal, participant-average	Average effect across participants. Potential outcomes are summarised under each treatment condition, and these summaries are contrasted.	IEEs (unweighted)	Individual participant outcomes are analysed using an independent working correlation structure in conjunction with cluster-robust standard errors. Can be implemented using (a) GEEs with a working independence correlation structure or (b) using maximum likelihood/least squares estimators (both options (a) and (b), using cluster-robust standard errors); importantly, the models are unweighted.	Standard assumptions ^a
		Cluster-level summaries (weighted)	A summary measure is taken in each cluster (e.g. the proportion), and an appropriate weighted generalised linear model is used for analysis, with the cluster-level summaries as the outcomes, and weights equal to the cluster size n_j . ^b For instance, a logit link could be used for the odds ratio; a family should be chosen such that the variance is independent of the mean (e.g. Gaussian).	Standard assumptions ^a
		GEEs with exchangeable correlation (unweighted) IEEs (weighted)	GEEs are applied to individual participant data, using an exchangeable working correlation structure, in conjunction with cluster-robust standard errors. As described above (under 'Marginal, participant average' estimand); however, participants are weighted by their	Standard assumptions ^a Non-informative cluster size
Marginal, cluster-average	Average effect across clusters Potential outcomes are summarised under each treat-			Standard assumptions ^a

Future work

We're at the beginning of the ICS journey in CRTs

We need more work to understand:

- How common it is in CRTs
- How much an impact it makes in practice
- How to test for and identify it robustly

Can we identify better estimators which maintain efficiency gains of mixed-models/GEEs, but are robust to ICS?

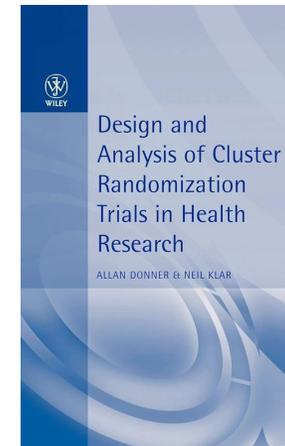
Acknowledgements

The content of this presentation is joint work with many others (including Fan Li, Andrew Copas, Michael Harhay, and others)

It builds on the decades of work that came before

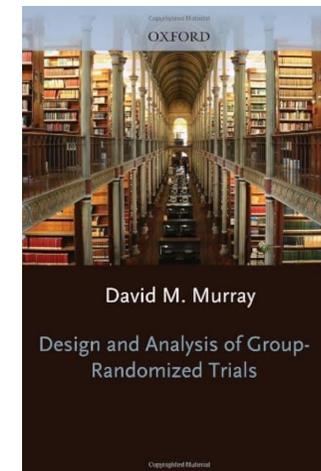
Thank you to Mind the Gap for the invitation

And thank you for listening!



SANDRA ELDRIDGE AND SALLY KERRY

A Practical Guide
to Cluster Randomised
Trials in Health
Services Research



Methods: Mind the Gap Webinar Series

- Updates on Research Methods Resources (researchmethodsresources.nih.gov)
 - Time-varying intervention effects in Stepped Wedge Designs
 - Group Regression Discontinuity Design section with background, key references, and a sample size calculator.