MRC Clinical Trials Unit

Smarter Studies
Global Impact
Better Health

UCL

Estimands in Cluster-Randomized Trials: Choosing Analyses that Answer the Right Question

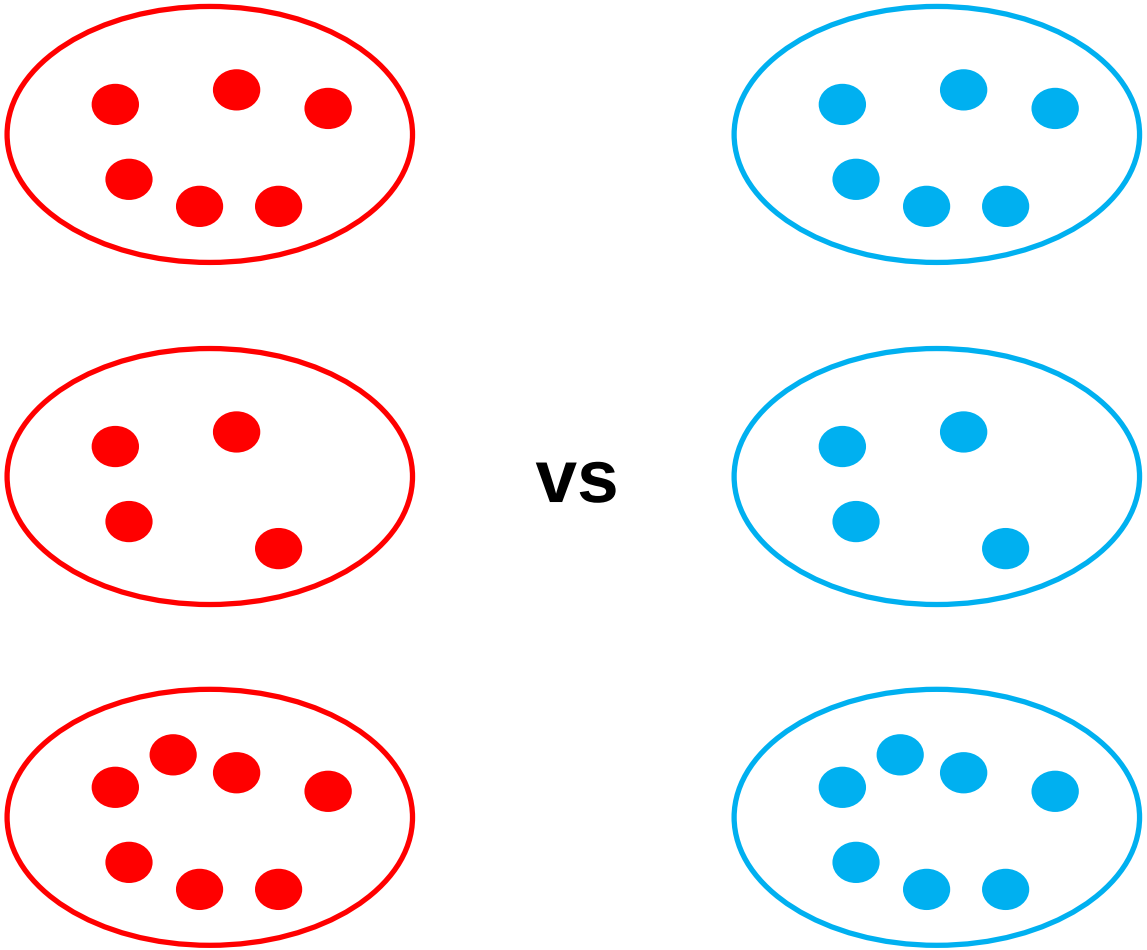
Brennan Kahan

Outline

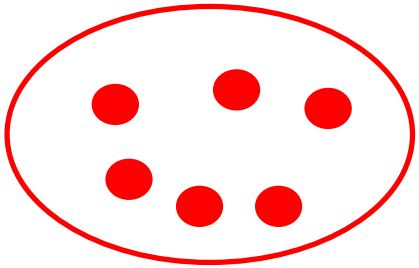
- Motivation
- Estimands
- Implications for analysis
- Conclusions

TRIGGER trial

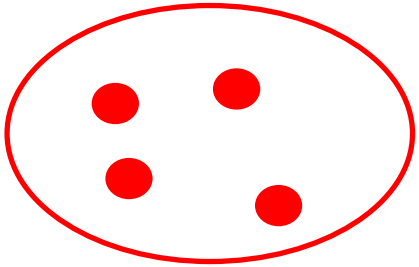
- Pragmatic CRT comparing two transfusion thresholds for acute upper gastrointestinal bleeding
- Feasibility trial – is a CRT approach feasible?
  - Adherence, etc



- How best to estimate differences in adherence etc with only 6 clusters?
- Methods such as GEEs/mixed-effects models tend to not do well with so few clusters

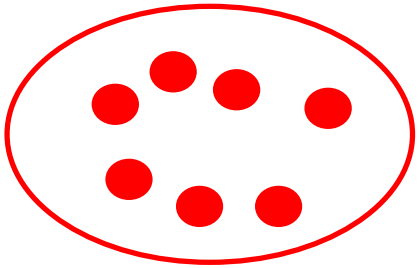


$= 13$

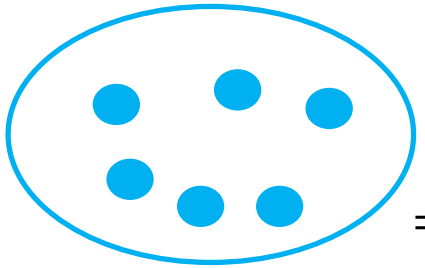


$= 7$

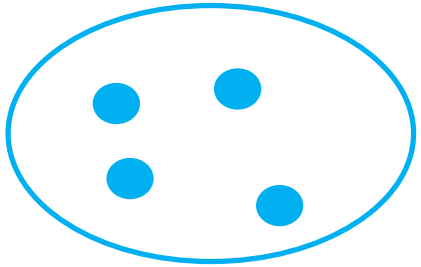
vs



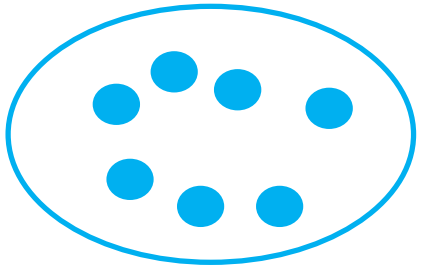
$= 21$



$= 16$



$= 10$



$= 25$

- Analysis of cluster level summaries:
  - Calculate mean outcome in each cluster
  - Apply a regression model to these cluster summaries

But what if we'd chosen a different analysis?

- As a statistician, often interesting to see what would have happened if we'd chosen differently
  - How much would standard errors *really* change?

But what if we'd chosen a different analysis?

- As a statistician, often interesting to see what would have happened if we'd chosen differently
  - How much would standard errors *really* change?

- I found something odd
  - I expected standard errors to change, but the treatment effects changed too
- Odds ratio:
  - Not accounting for clustering: 3.69
  - Cluster-level summaries: 4.85
  - GEEs (exch): 3.83
  - Mixed-effects model: 4.21

Estimands

- The treatment effect we want to estimate
- Popularised with publication of ICH-E9(R1) addendum
  - Though concepts were floating around much earlier

Estimands

- CRTs can be used to estimate *different* treatment effects
- Different odds ratio for adherence correspond to different questions

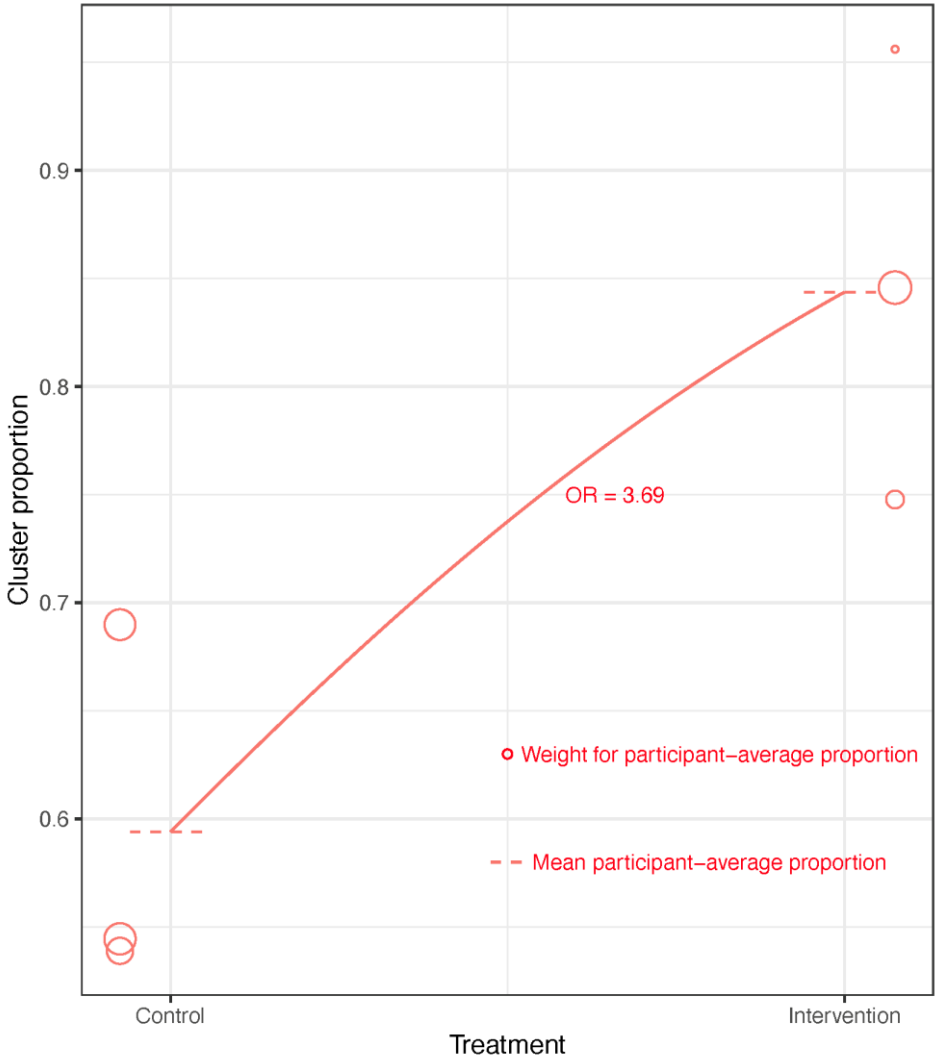
Estimands for CRTs

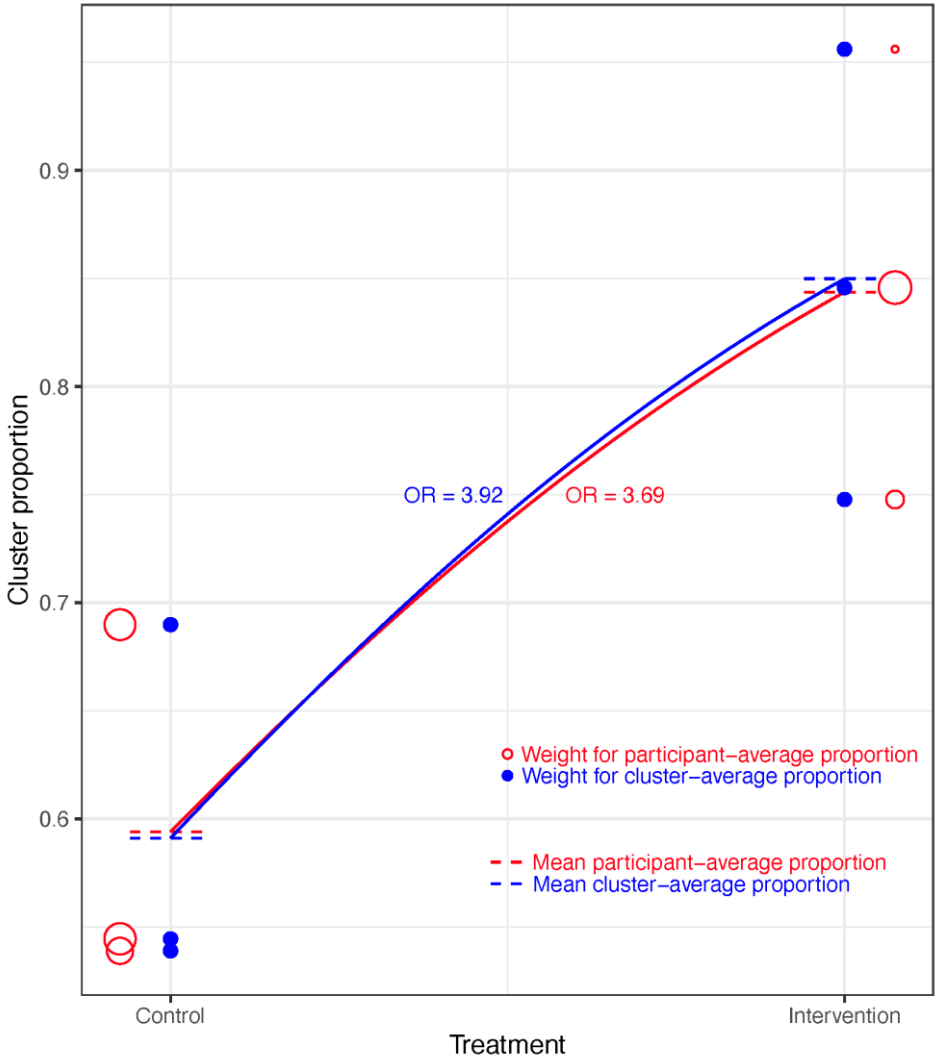
- Have additional considerations compared to individually randomised trials
  - Participant- vs. cluster-average effect
  - Marginal vs. cluster-specific*

Participant- vs. cluster-average effects

- Participant-average effect:
  - The average effect across participants
- Cluster-average effect:
  - The average effect across clusters

- Difference is in how data are weighted
  - Participant-average effect -> participants all get equal weight
  - Cluster-average effect -> clusters all get equal weight


0.9
0.8
0.7
0.6
Cluster proportion
OR = 3.69
Weight for participant−average proportion
Mean participant−average proportion
Control
Intervention
Treatment


0.9
0.8
0.7
0.6
Cluster proportion
OR = 3.92
OR = 3.69
Weight for participant−average proportion
Weight for cluster−average proportion
Mean participant−average proportion
Mean cluster−average proportion
Control
Intervention
Treatment

Informative cluster size (ICS)

- These two estimands will differ when there is informative cluster size (ICS)
- ICS:
  - Outcomes and/or treatment effects from large clusters differ to those from smaller clusters
  - E.g. patients experience better outcomes/treatment effects if they present to a large hospital compared to a small hospital (or vice versa)

Which estimand to use?

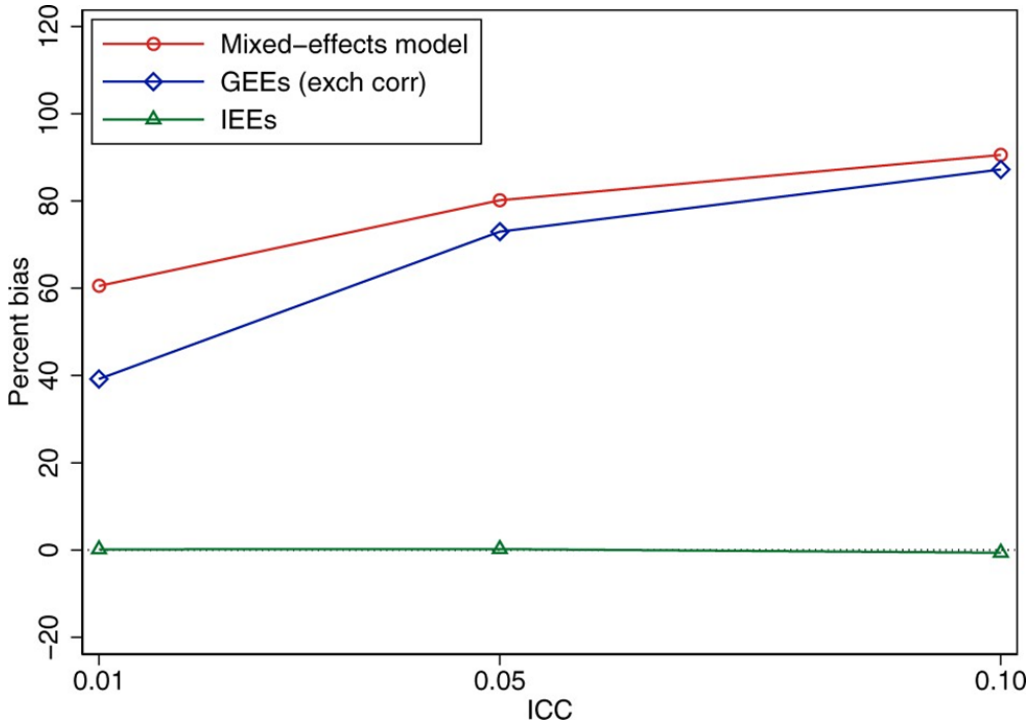
- Depends on the study question
- Participant-average
  - provides population-level effect of going from one intervention to the other
  - I.e. shows effect across patients
- Cluster-average
  - enables evaluation of intervention's impact directly on clusters
  - I.e. can show whether intervention modified behaviour of clusters

Implications for analysis

- Mixed-effects models/GEEs with an exchangeable correlation structure are the most common methods of analysis for CRTs
  - **Problem**: when ICS is present, both are biased

- The reason is to do with how these methods weight the data
  - For the PA effect we need to weight participants equally
  - For the CA effect we need to weight clusters equally
  - These methods do neither; weighting is based on efficiency
  - Biased for both PA and CA effects

Example of bias



- Simulated example based on 30 clusters with N=100 and 30 with N=10
- Bias is for PA effect

What is the alternative?

- Two options: Independence estimating equations (IEEs) and cluster-level summaries
- Both can be used to estimate either cluster- or participant-average effects
- Both unbiased under ICS

Independence estimating equations

- Use working independence correlation structure
  - This is to ensure proper weighting of data corresponding to our target estimand
- We know this assumption is likely false in practice
  - Use in conjunction with cluster-robust SEs to obtain correction confidence intervals/p-values

Independence estimating equations

- Can be implemented different ways
  - GEEs with working independence correlation structure
  - Maximum likelihood/least squares
  - Key thing is to ensure cluster-robust SEs

- Can be used to estimate either participant- or cluster-average effect
  - For PA effect -> implement as usual (i.e. unweighted)
  - For CA effect -> weight participants by $\frac{1}{n_j}$

Analysis of cluster-level summaries

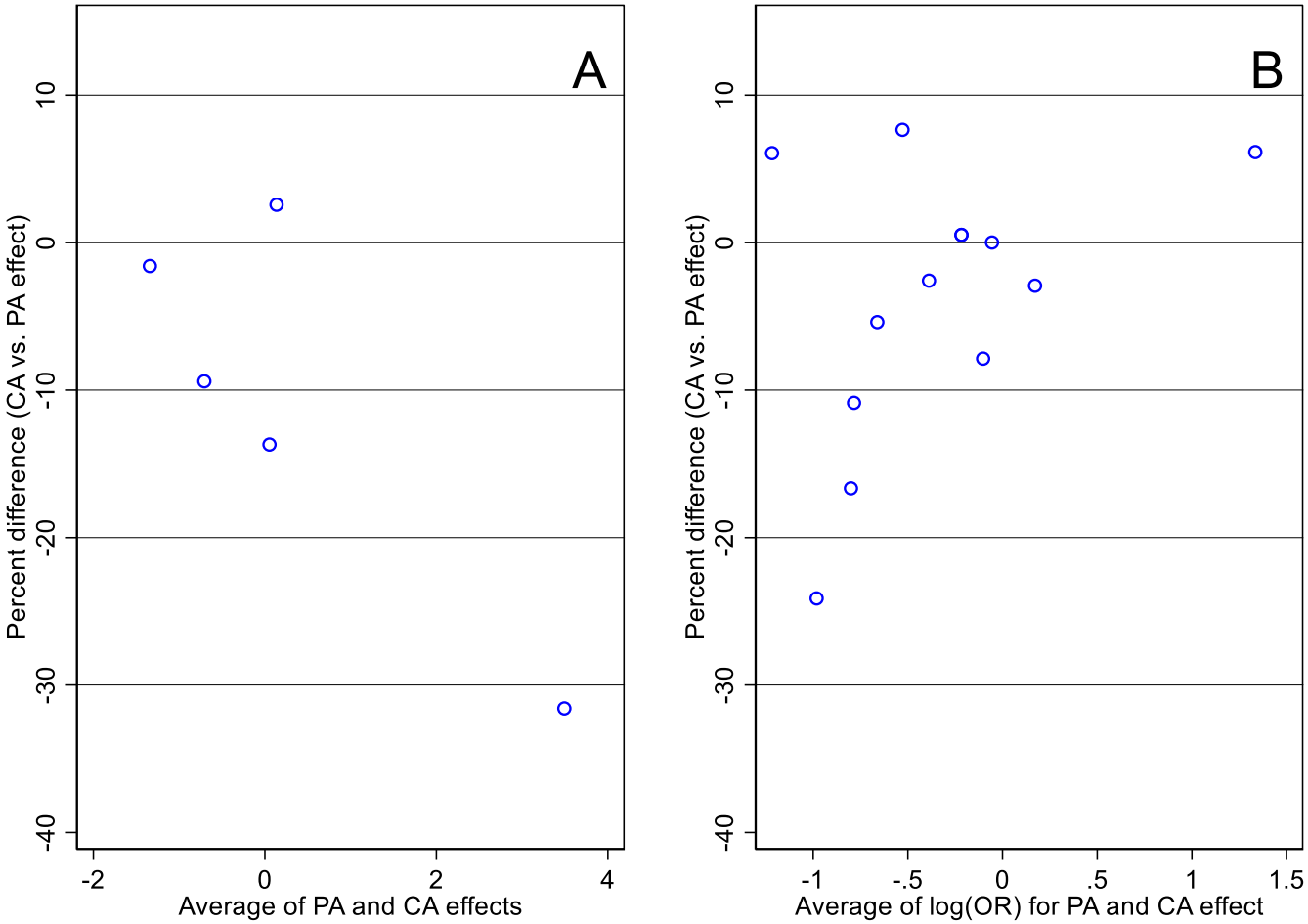
- Calculate mean outcome in each cluster
- Apply regression model to cluster-level summaries
  - Unweighted regression model for cluster-average effect
  - Weighted by n_j for participant-average effect

Application to TRIGGER

Estimand	Estimator	Odds ratio (95% CI)
Marginal participant-average		
	GEEs (exchangeable correlation structure)	3.83 (1.65 to 8.86)
	IEEs (unweighted)	3.69 (1.83 to 7.43)
	Cluster-level summaries (weighted)	3.69 (1.83 to 7.43)
Cluster-specific participant-average		
	Mixed-effects model	4.21 (1.86 to 9.51)
	Cluster-level summaries (weighted)	4.28 (1.11 to 16.48)
Marginal cluster-average		
	IEEs (weighted)	3.92 (1.59 to 9.64)
	Cluster-level summaries (unweighted)	3.92 (1.51 to 10.19)
Cluster-specific cluster-average		
	Cluster-level summaries (unweighted)	4.85 (0.85 to 27.53)

How common is ICS in practice?

- If unlikely, then means we could use our standard methods (mixed-effects models/GEEs with an exchangeable correlation structure) and not worry about it
- Occurrence of ICS has never (to our knowledge) been evaluated


A
B
Percent difference (CA vs. PA effect)
10
0
-10
-20
-30
-40
-2
0
2
4
Average of PA and CA effects
-1
-.5
0
.5
1
1.5
Average of log(OR) for PA and CA effect

Practical implications

- Need to think about estimand
  - Which question is most relevant for my study?
- Tailor analysis around chosen estimand

If ICS expected

- Use independence estimating equations/cluster-level summaries
  - Robust to ICS
  - Need to ensure appropriate weighting is used corresponding to desired estimand

If ICS not expected

- Could use mixed-effects models/GEEs(exch)
  - Increase precision compared to IEEs/cluster summaries
  - With IEEs/cluster-level summaries as a sensitivity analysis

- Could use IEEs/cluster-level summaries anyways
  - Ensures results robust even if you're wrong about ICS

Future work

- Evaluating ICS in other trial datasets
- Sample size calculations for when ICS is expected
- Evaluating performance of estimators with small number of clusters
- Extending to cluster-crossover/stepped wedge trials

International Journal of Epidemiology, 2022, 1–10
https://doi.org/10.1093/ije/dyac131
Original article

Original article

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

Brennan C Kahan,[1]* Fan Li ,[2,3] Andrew J Copas[1] and Michael O Harhay [4,5]

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

Received 25 October 2021; Editorial decision 6 June 2022; Accepted 7 June 2022

New articles coming soon(ish)

- Informative cluster size in cluster-randomised trials: A case study from the TRIGGER trial
  - Our results on ICS in TRIGGER

- Demystifying estimands in cluster randomised trials
  - More on PA vs. CA and marginal vs. cluster-specific effects, and estimation

Acknowledgements

- Fan Li
- Michael Harhay
- Andrew Copas
- Bryan Blette
- Vipul Jairath