Methods: Mind the Gap

Webinar Series

Why You Should Consider Using Randomization Tests To Analyze Your Stepped Wedge Trial

Presented by:

Jennifer Thompson, Ph.D.

London School of Hygiene & Tropical Medicine

National Institutes of Health

Office of Disease Prevention

Why you should consider using randomization tests to analyse your stepped wedge trial

Jennifer Thompson

International
Statistics &
Epidemiology
Group

The difficulty of trial analysis

The analysis of trials is, by design, usually simple

What makes it difficult is the need to pre-specify what we are going to do

Most analysis methods make some assumptions

We must anticipate what assumptions will be appropriate for our trial data before we see that data

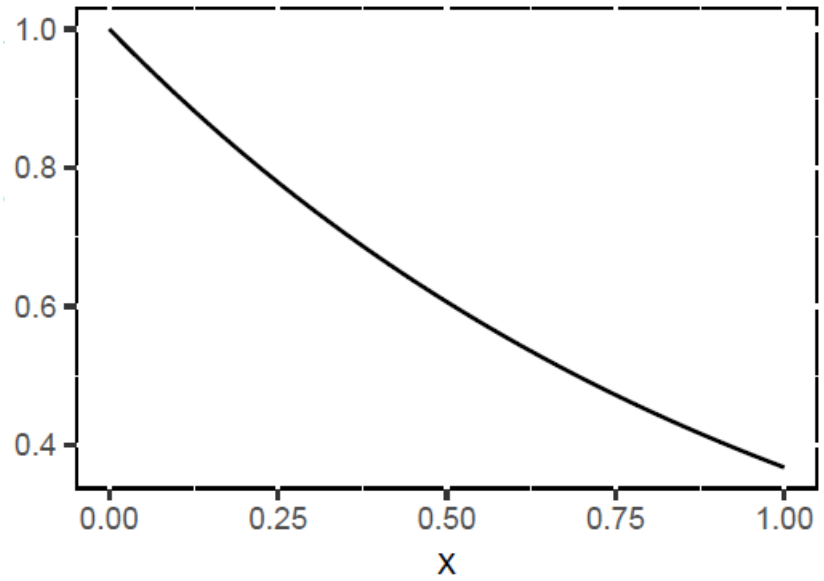
We can end up in a difficult situation if our chosen analysis methods turns out to be inappropriate for the data, potentially giving misleading results

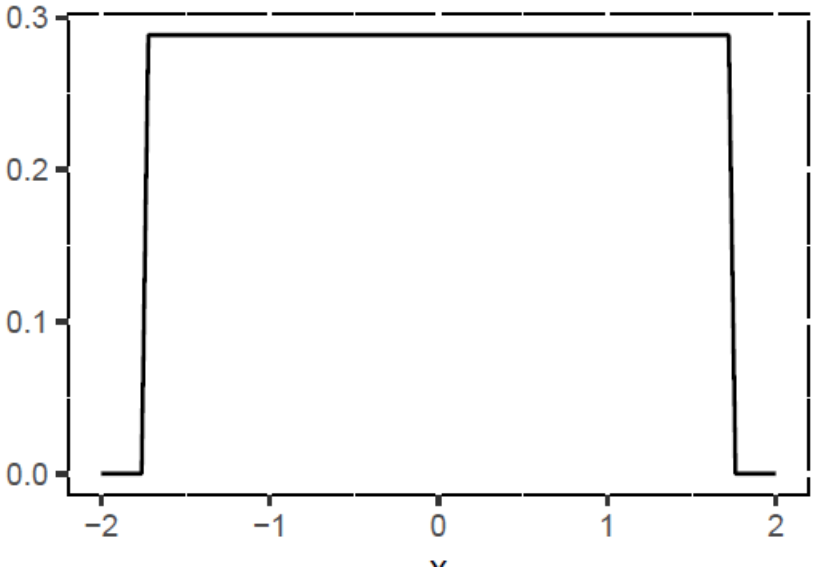
Simpler trial analysis

We can make this easier by choosing analysis methods that

- Do not make as many assumptions
- Are robust against the assumptions they do make

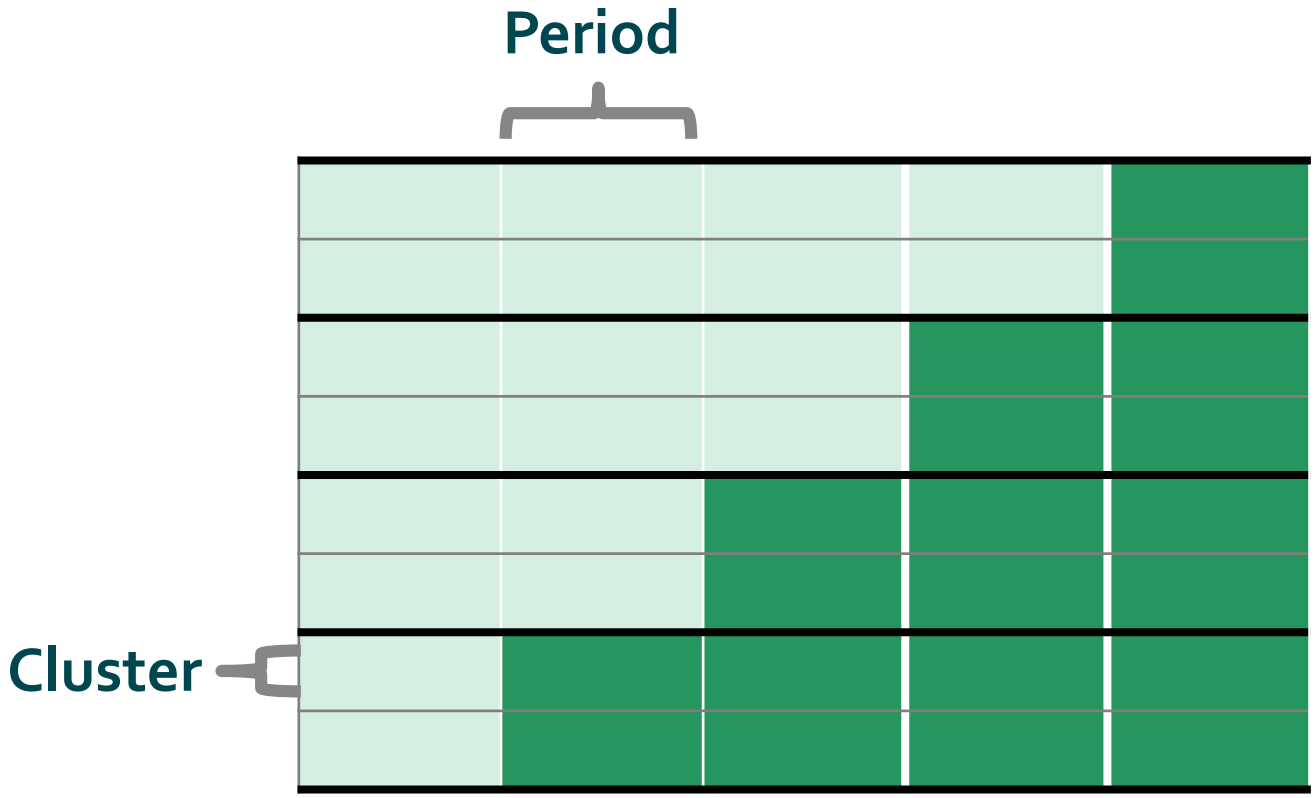
e.g. T-test





What is a stepped wedge cluster randomised trial

Stepped wedge trials are an example where pre-specification can be difficult



Example of a stepped wedge trial

Tuberculosis diagnostic test trial, 2012

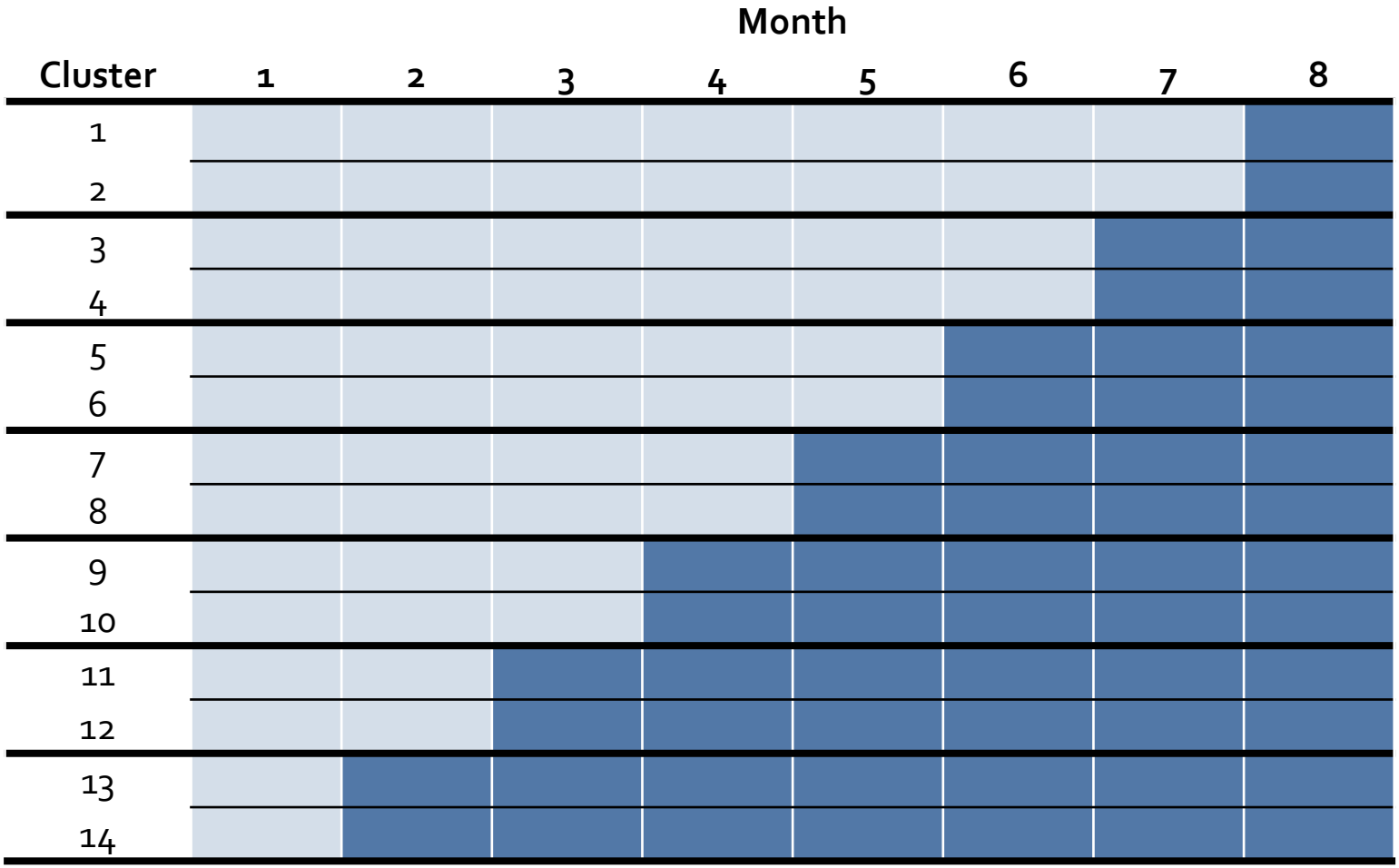
Switching from the standard test to a new, more sensitive diagnostic test for initial diagnosis of TB.

Impact on patient's response to TB treatment.

Trajman, Anete, et al. "Impact on patients' treatment outcomes of XpertMTB/RIF implementation for the diagnosis of tuberculosis: follow-up of a stepped-wedge randomized clinical trial." *Plos one* 10.4 (2015): e0123252.

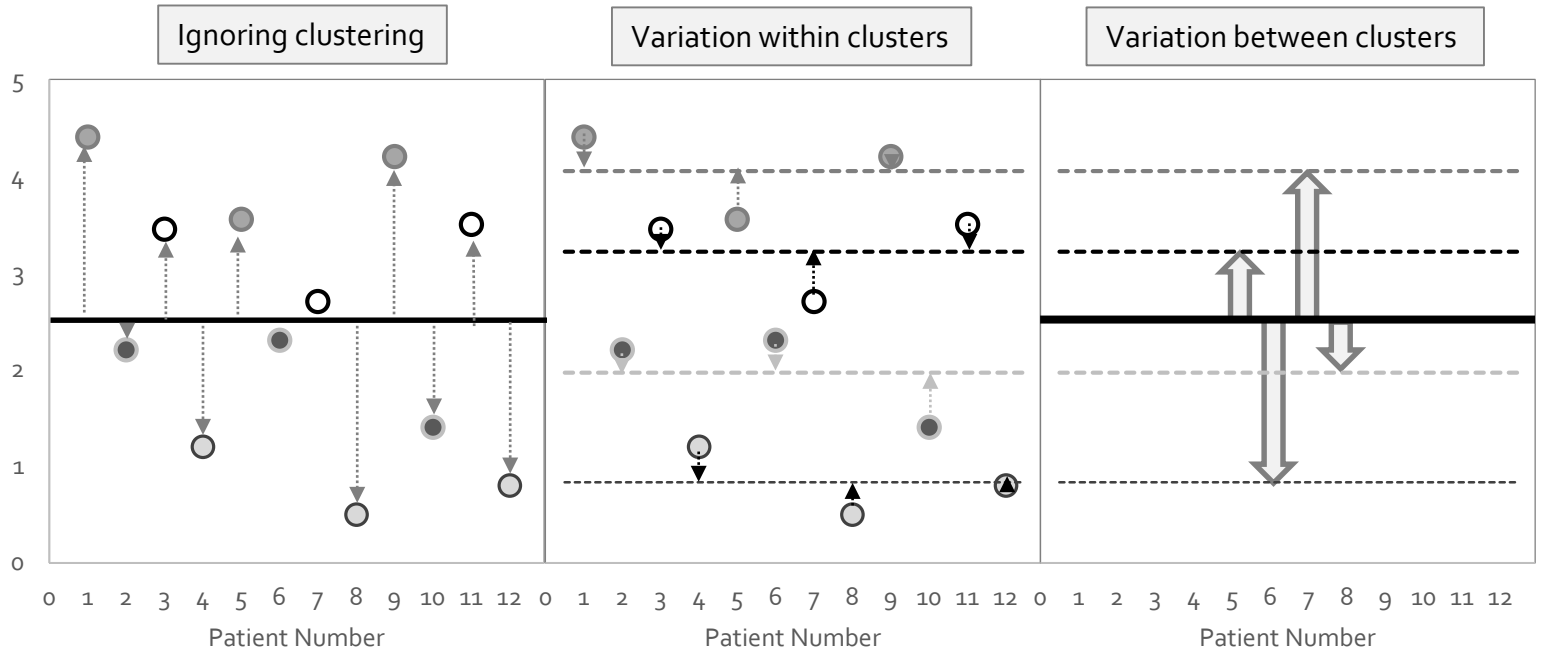
Testing a sample from a person with suspected TB in a GeneXpert machine, Bangassou Hospital, Central African Republic.

Example of a stepped wedge trial



The correlation problem

As with parallel cluster randomised trials, stepped wedge trial analysis must account for any between correlation observations

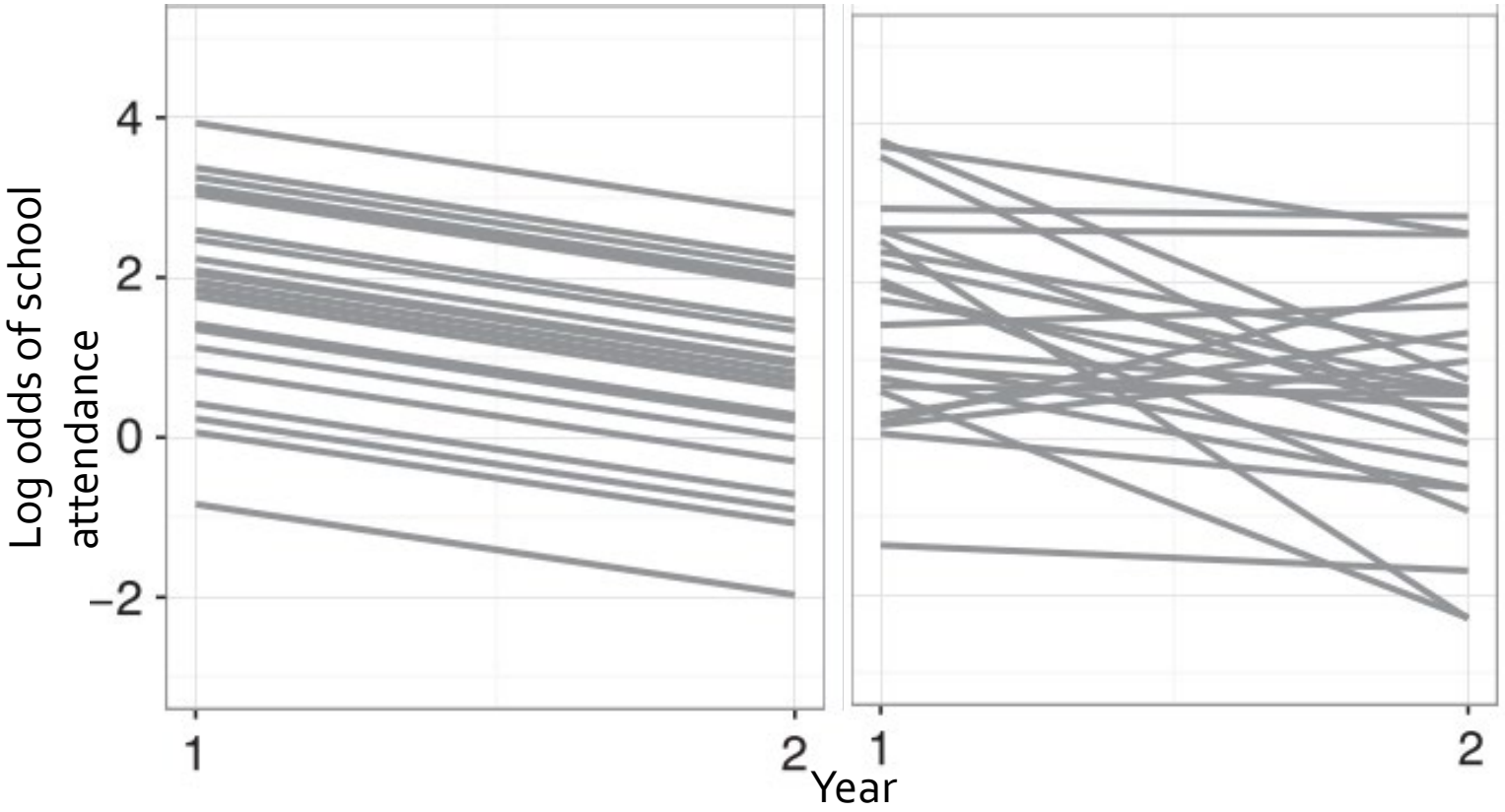


The correlation problem

Stepped wedge trial analysis must also account for changes to the correlations within clusters.

Same change over time for all clusters?

Or change over time varies between clusters?



The correlation problem

If change over time varies between clusters, after adjusting for the overall time trend, the correlation within clusters will reduce over time

Observations collected closer together in time will be more similar to one another than observations collected further apart in time

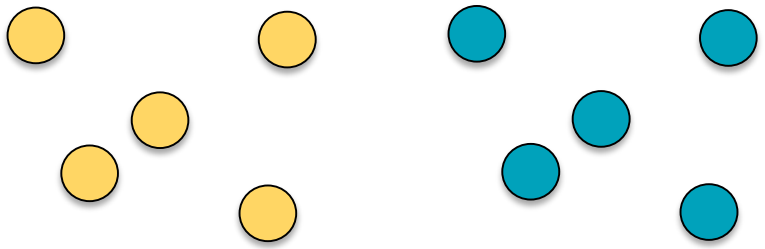
Randomization tests to the rescue

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

What is a randomization test?

Lets start with a simple individually randomised trial.

We have some observed data collected under the control condition and some collected under the intervention condition.



We use these to estimate the effect of the intervention

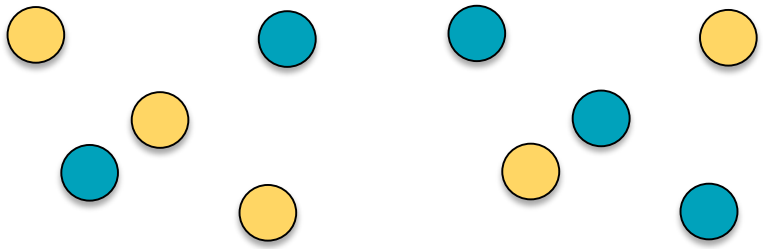
$$\hat{\theta}$$

What is a randomization test?

If the intervention had no effect, it wouldn't have mattered who received the control or intervention conditions: we would have expected to see a similar result.

What is a randomization test?

We can mix up the assignment of control and intervention:

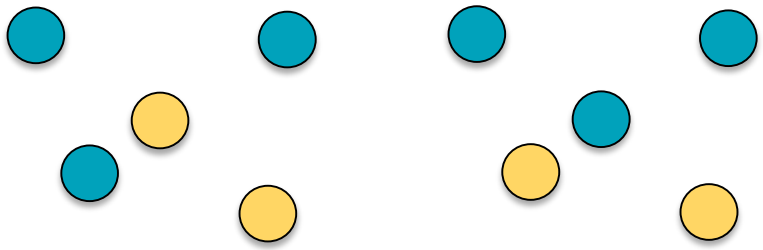


And re-estimate an effect of the intervention with this assignment:

$$\widehat{\boldsymbol{\theta}_1}$$

What is a randomization test?

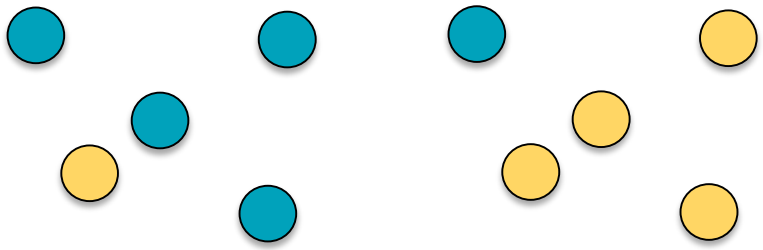
We do this many times:



$$\widehat{\boldsymbol{\theta}_2}$$

What is a randomization test?

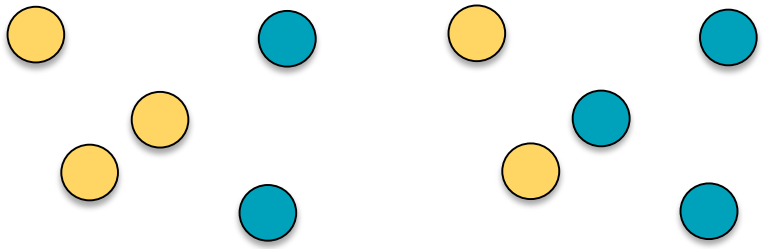
We do this many times:



$$\widehat{\boldsymbol{\theta}_3}$$

What is a randomization test?

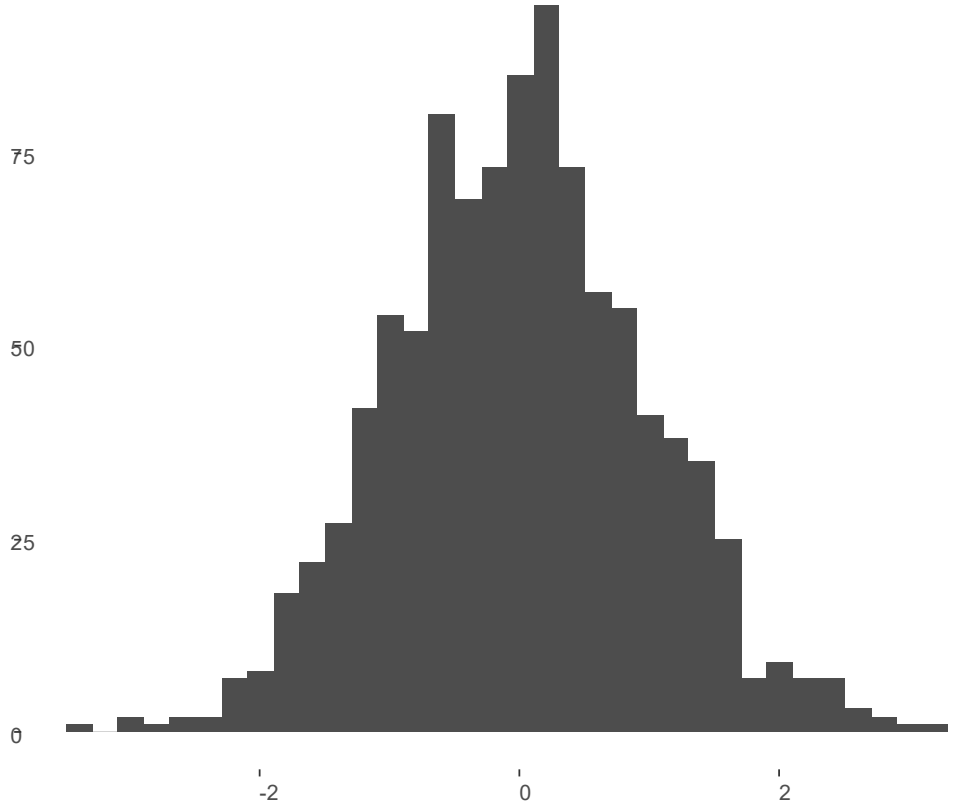
We do this many times:



$\widehat{\boldsymbol{\theta}_4}$

What is a randomization test?

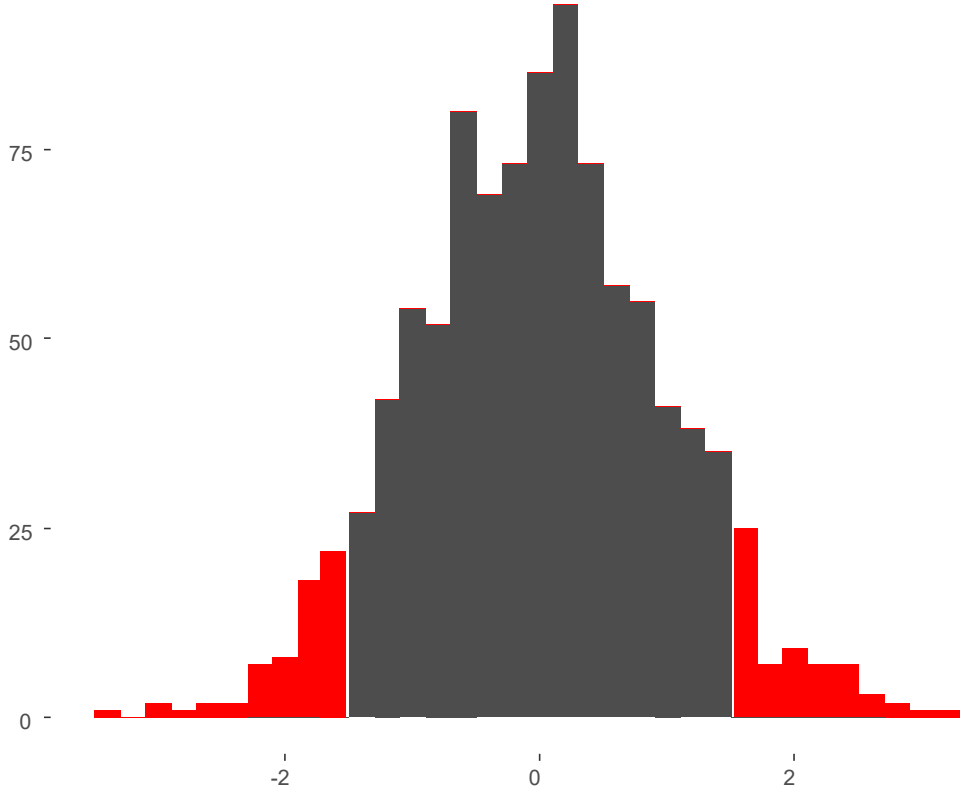
Distribution of intervention effect estimates under null hypothesis of no effect:



What is a randomization test?

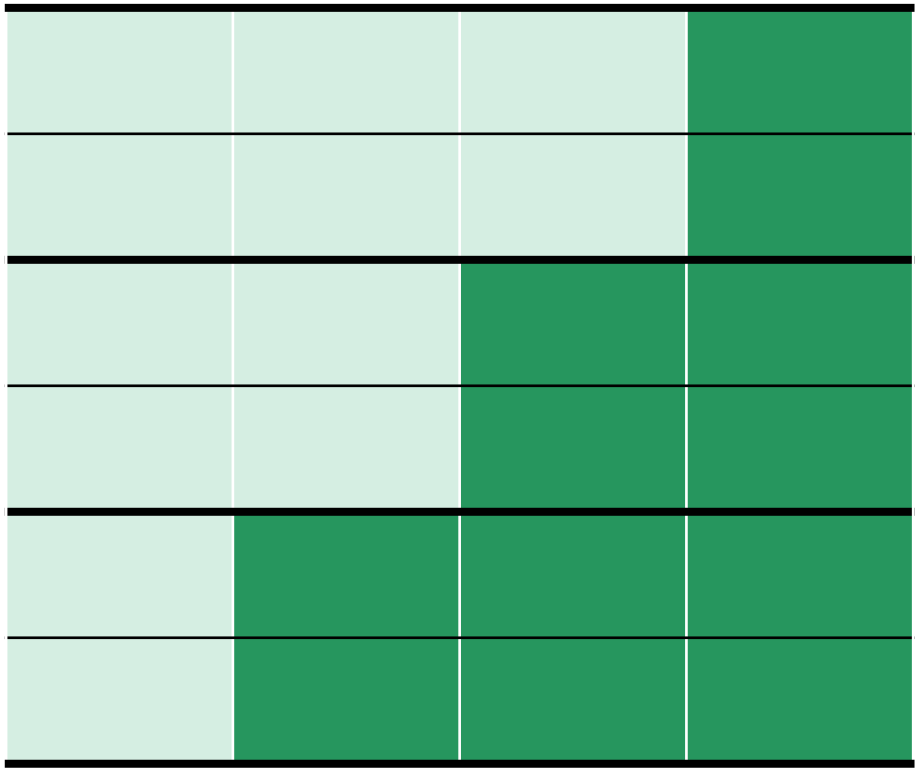
P value is the probability of observing the same or a more extreme results under the null hypothesis of no intervention effect

P=0.10



Randomization test in stepped wedge trial

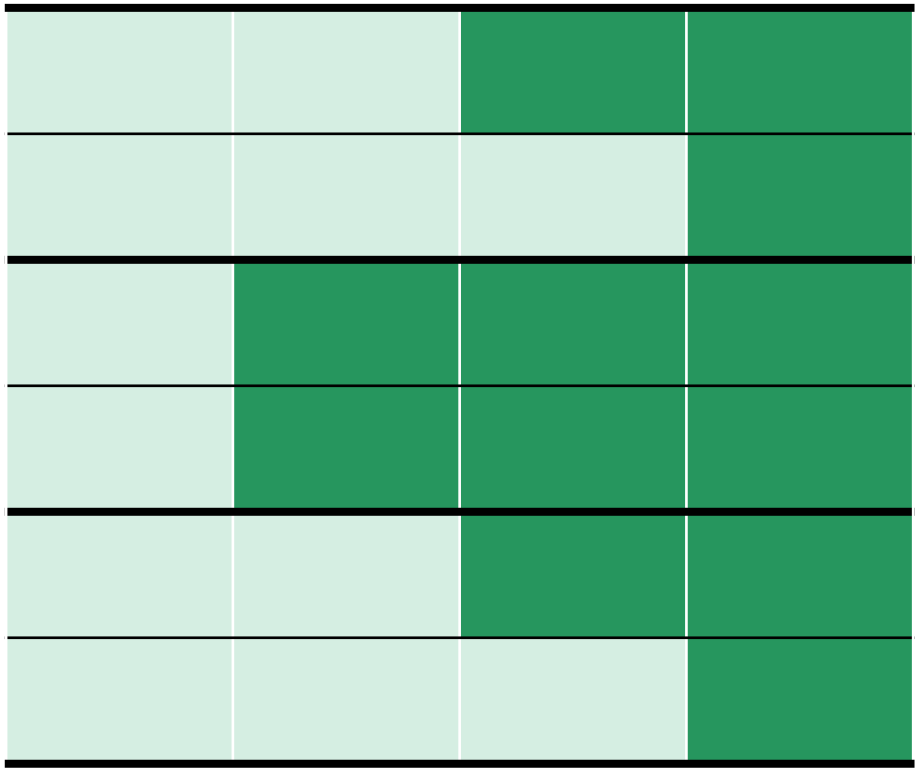
For a stepped wedge trial, we mix up the assignment of clusters to the times at which they switch to the intervention



$\hat{\theta}$

Randomization test in stepped wedge trial

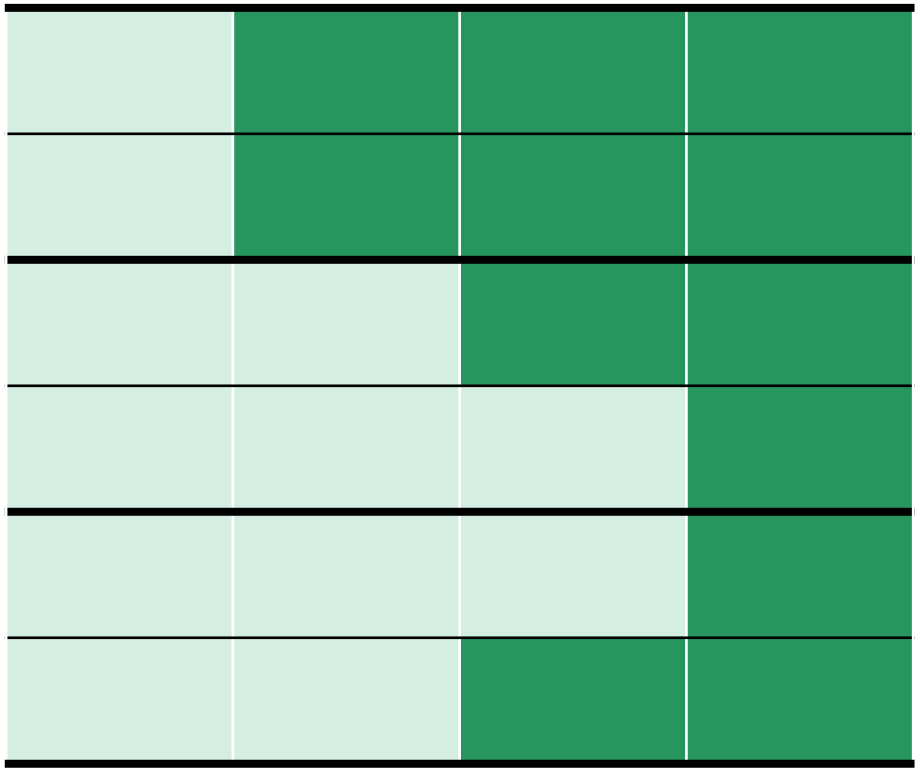
For a stepped wedge trial, we mix up the assignment of clusters to the times at which they switch to the intervention



$\widehat{\theta_1}$

Randomization test in stepped wedge trial

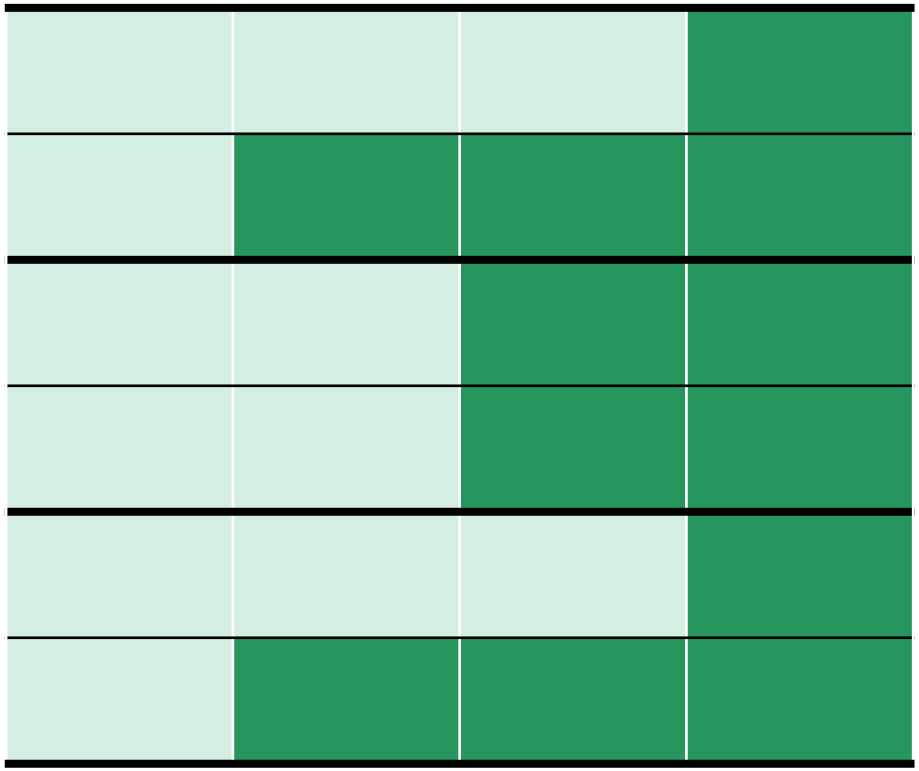
For a stepped wedge trial, we mix up the assignment of clusters to the times at which they switch to the intervention



$\widehat{\theta_2}$

Randomization test in stepped wedge trial

For a stepped wedge trial, we mix up the assignment of clusters to the times at which they switch to the intervention



$\widehat{\theta_3}$

Randomization test in stepped wedge trial

No specification of correlation structure within clusters.

Instead, we honour the structure of the data when we re-assign control and intervention conditions

What is the intervention effect and how do we estimate it?

Mixed effect model estimator

Mixed effect models model the correlation structure of the data to estimate the intervention effect and adjust for time

+ Very efficient

- Assume the correlation structure is correctly specified. If it is incorrect, the effect estimate could be biased. This makes them difficult to prespecify

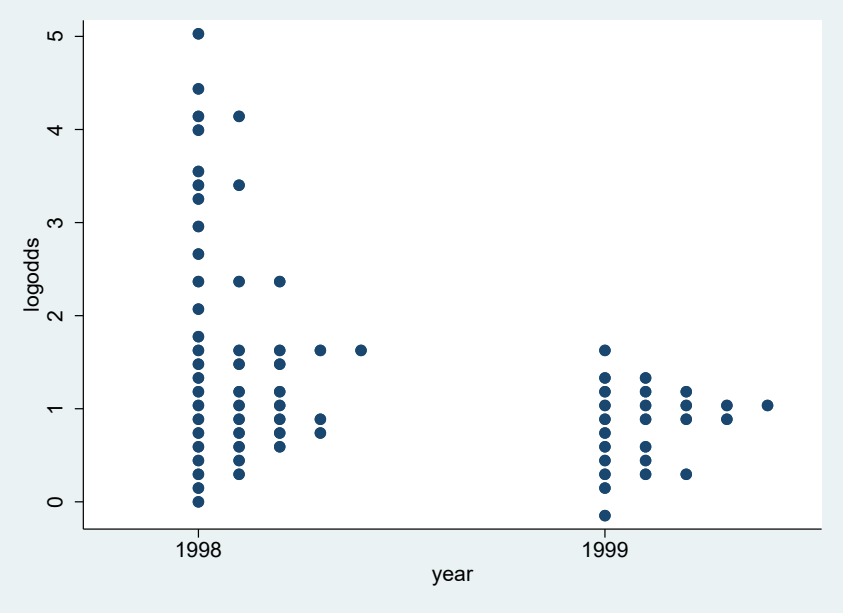
Mixed effect model estimator bias

Econometrica, Vol. 72, No. 1 (January, 2004), 159–217

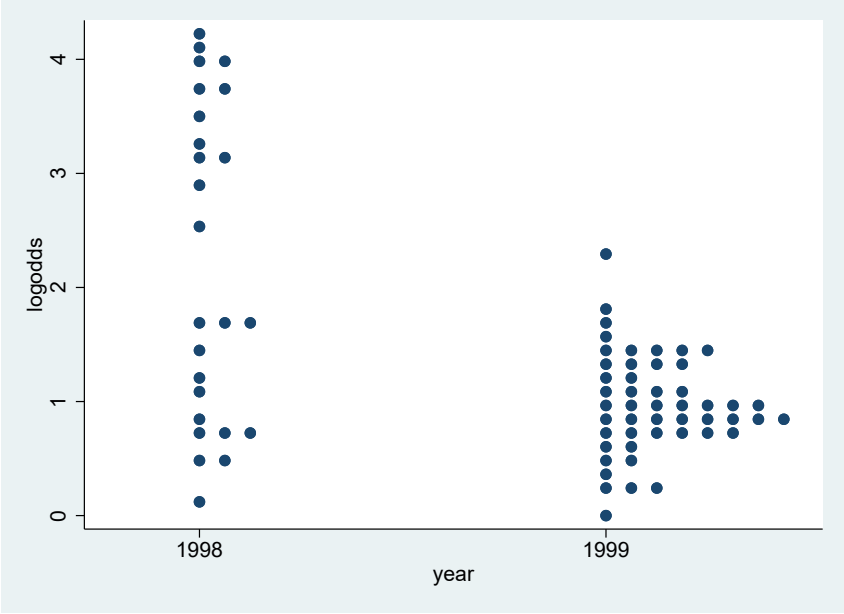
WORMS: IDENTIFYING IMPACTS ON EDUCATION AND HEALTH IN THE PRESENCE OF TREATMENT EXTERNALITIES

BY EDWARD MIGUEL AND MICHAEL KREMER[1]

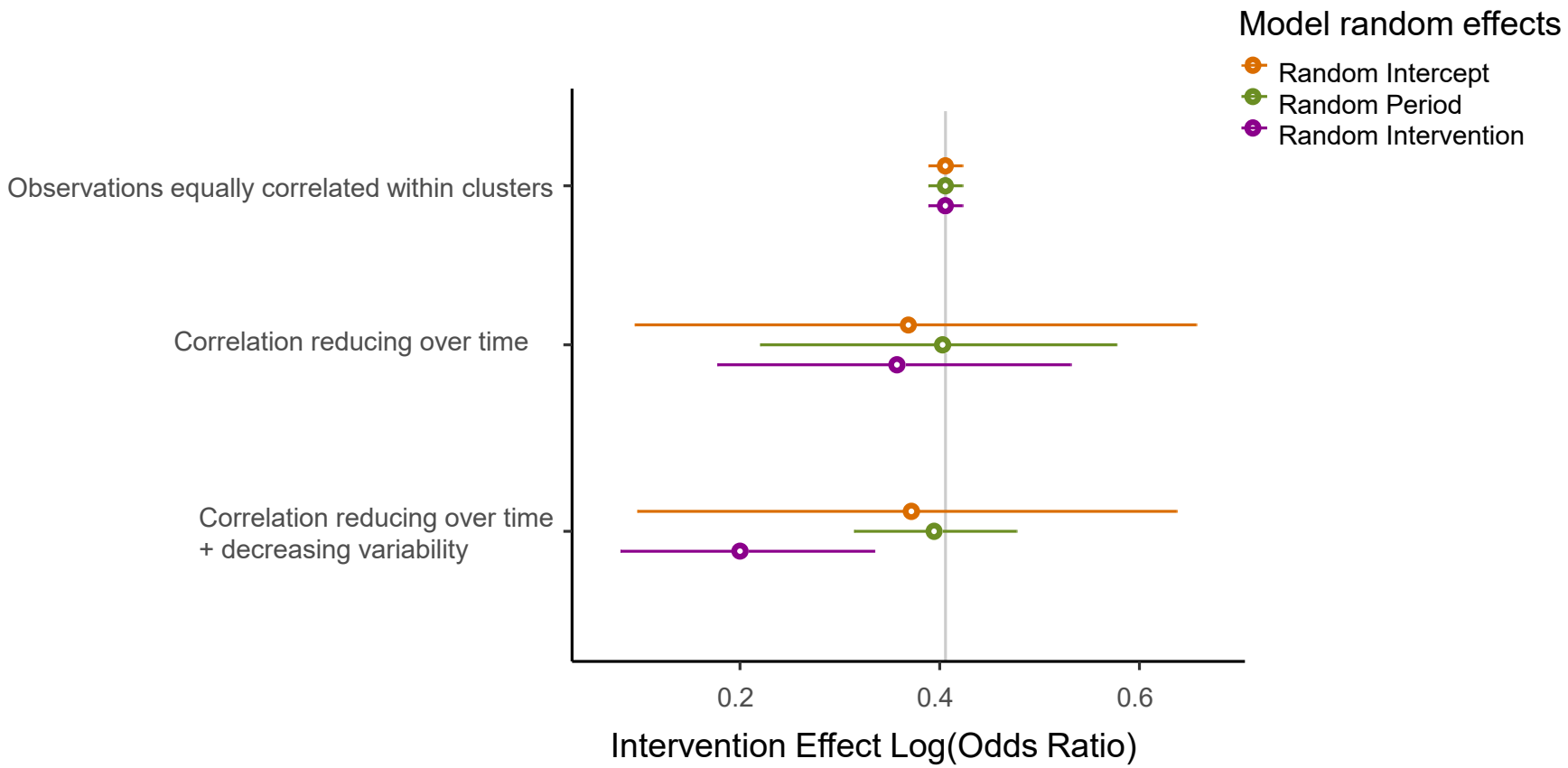
Control



Intervention



Mixed effect model estimator bias



Thompson, Jennifer A., et al. "Bias and inference from misspecified mixed-effect models in stepped wedge trial analysis." *Statistics in medicine* 36.23 (2017): 3670-3682.

Within period estimator

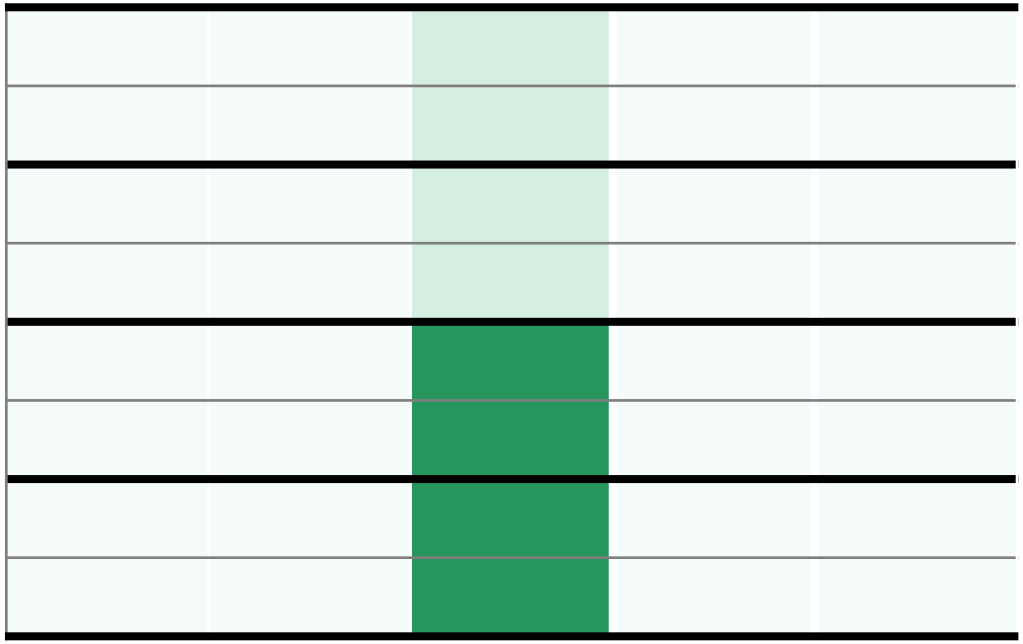
Within-period analysis more robust but less efficient. Different approaches. Focus on cluster-level analysis calculated in each period, then combined as a weighted average

Cluster level analysis within period

Trial can be split into "parallel" CRTs between clusters switching

Cluster level analysis within period

Trial can be split into "parallel" CRTs between clusters switching



Cluster level analysis within period

Summarise the outcome in each cluster in each period

p_{11}	p_{12}	p_{13}	p_{14}	p_{15}
p_{21}	p_{22}	p_{23}	p_{24}	p_{25}
p_{31}	p_{32}	p_{33}	p_{34}	p_{35}
p_{41}	p_{42}	p_{43}	p_{44}	p_{45}
p_{51}	p_{52}	p_{53}	p_{54}	p_{55}
p_{61}	p_{62}	p_{63}	p_{64}	p_{65}
p_{71}	p_{72}	p_{73}	p_{74}	p_{75}
p_{81}	p_{82}	p_{83}	p_{84}	p_{85}

Cluster level analysis within period

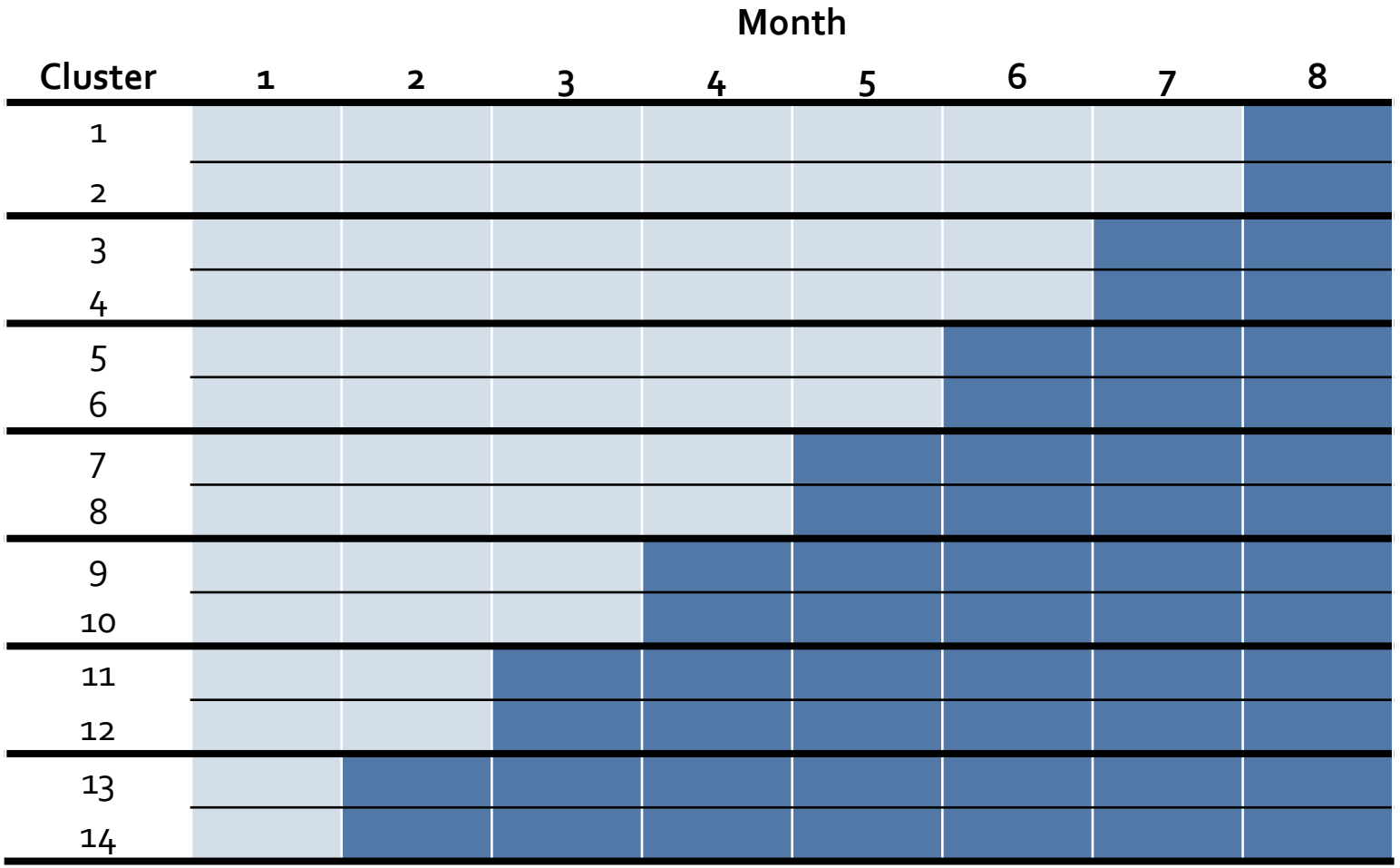
Use a t-test to compare control and intervention conditions within each period and estimate an intervention effect and variance

	$\hat{p}_{c2}$	$\hat{p}_{c3}$	$\hat{p}_{c3}$	
	$\hat{p}_{I2}$	$\hat{p}_{I3}$	$\hat{p}_{I3}$	
	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$	
	$\hat{V}(\theta_2)$	$\hat{V}(\theta_3)$	$\hat{V}(\theta_4)$	

$$\hat{\theta} = \sum \hat{V}^{-1}(\theta_i) \left[\frac{1}{\hat{V}(\theta_2)} \hat{\theta}_2 + \frac{1}{\hat{V}(\theta_3)} \hat{\theta}_3 + \frac{1}{\hat{V}(\theta_4)} \hat{\theta}_4 \right]$$

Applying it the tuberculosis diagnostic test example

TB diagnostic test example



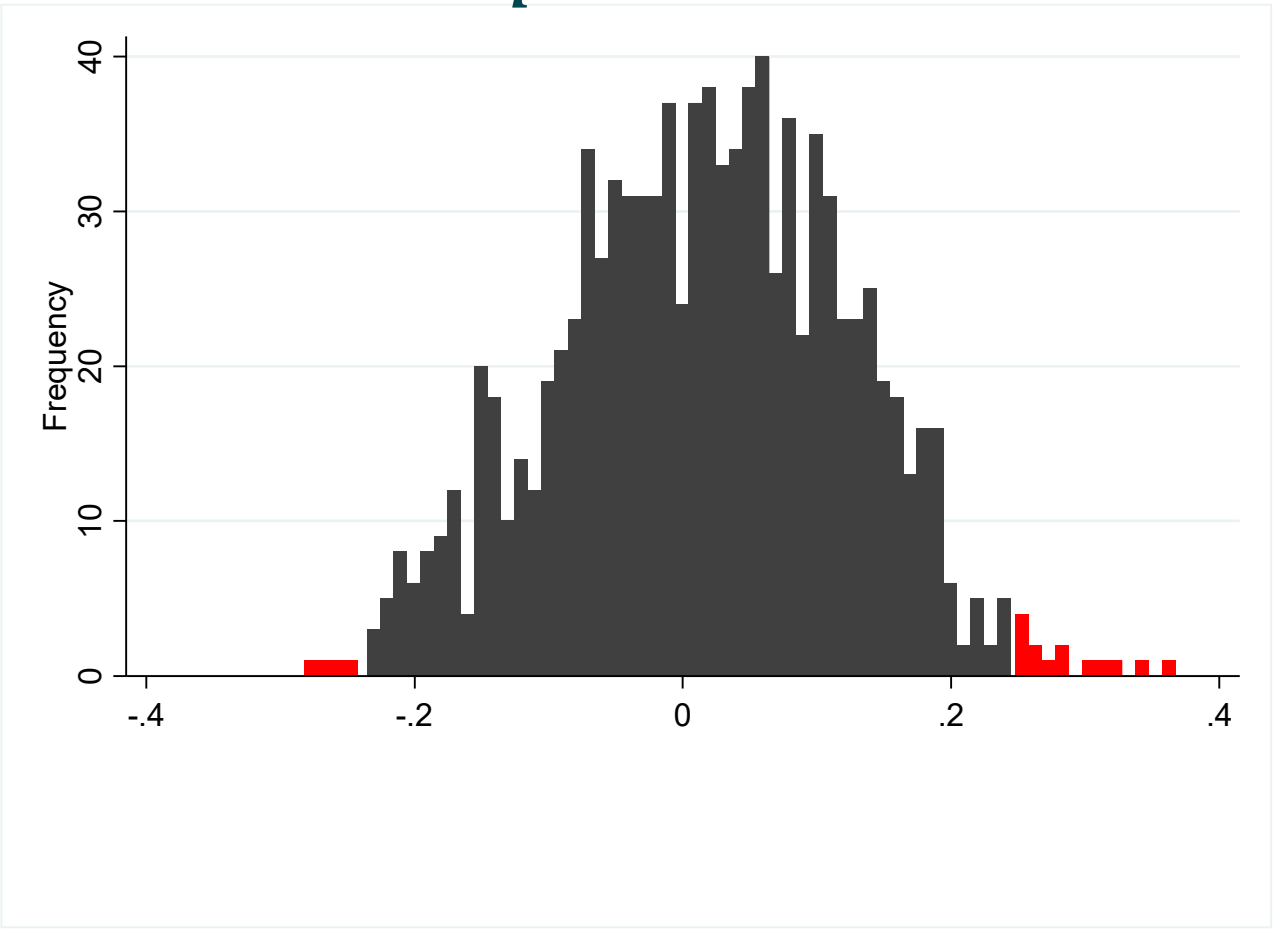
TB diagnostic test example

	Month						
	1	2	3	4	5	6	7	8
Log odds ratio		0.03	-0.26	-0.48	-0.09	-0.33	-0.17	
Odds ratio		1.03	0.77	0.62	0.91	0.72	0.85	
Weight		0.05	0.07	0.30	0.43	0.12	0.03	

Overall odds ratio: 0.78

TB diagnostic test example

$p = 0.02$



Do they work?

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

How robust are randomization tests?

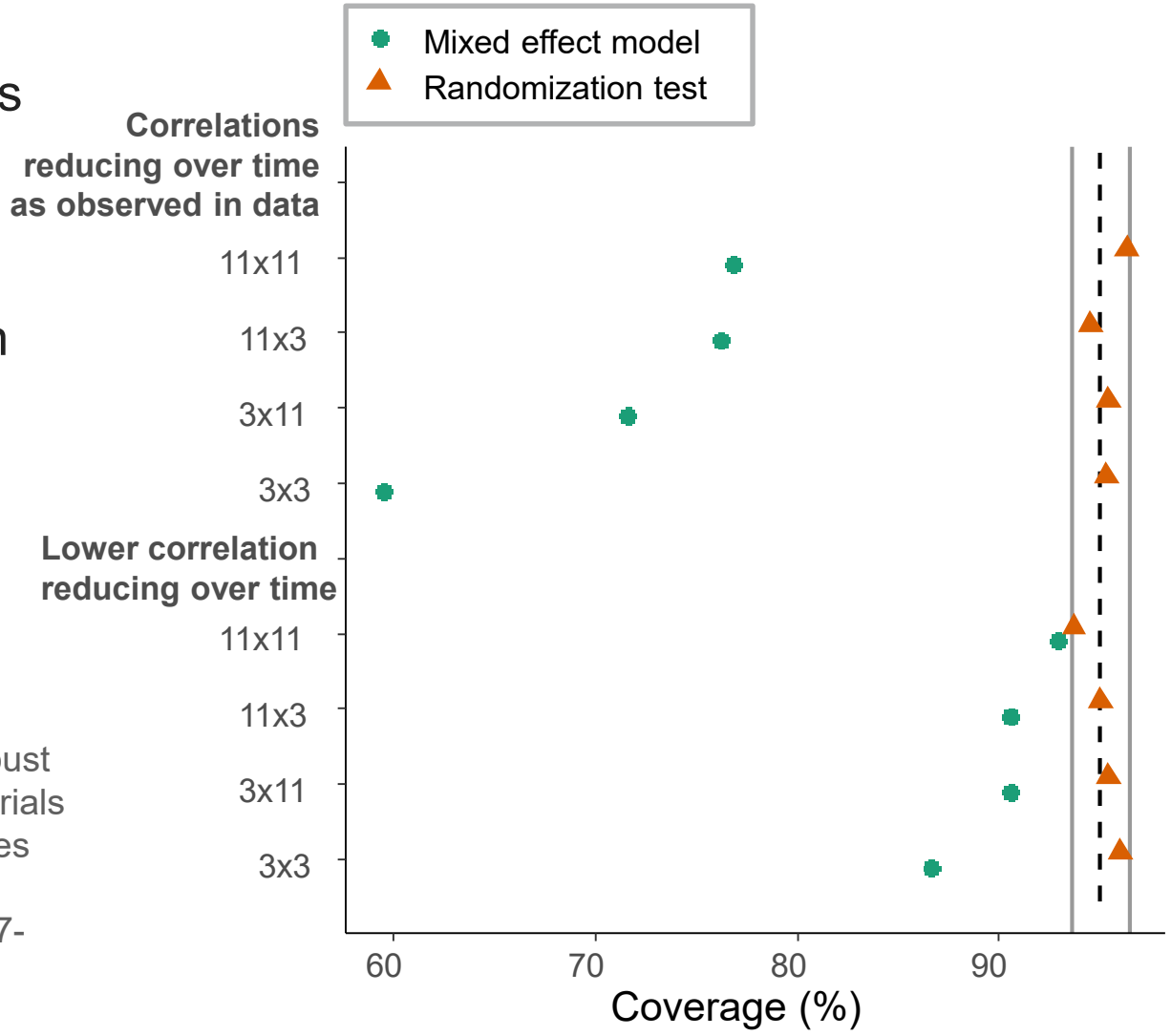
Simulation study

- Generated 1000 dataset to mimic a stepped wedge trial of an intervention aimed at increasing uptake of free health checks from the NHS
- Time trends varied between clusters so correlation reduced over time
- Looked at several different trial designs

Thompson, J. A., et al. "Robust analysis of stepped wedge trials using cluster-level summaries within periods." *Statistics in medicine* 37.16 (2018): 2487-2500.

How robust are randomization tests?

Randomization tests had good 95% confidence interval coverage with reducing correlation over time



Thompson, J. A., et al. "Robust analysis of stepped wedge trials using cluster-level summaries within periods." *Statistics in medicine* 37.16 (2018): 2487-2500.

Other research

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Research Article

Statistics in Medicine

Received 29 December 2015, Accepted 10 April 2017 Published online 2 May 2017 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.7329

The use of permutation tests for the analysis of parallel and stepped-wedge cluster-randomized trials

Rui Wang[a,b*†] and Victor De Gruttola[b]

Submitted to the Annals of Applied Statistics
arXiv: arXiv:0000.0000

RANDOMIZATION INFERENCE FOR STEPPED-WEDGE CLUSTER-RANDOMIZED TRIALS: AN APPLICATION TO COMMUNITY-BASED HEALTH INSURANCE

BY XINYAO JI *, GUNTHER FINK†, PAUL JACOB ROBYN‡ AND DYLAN S. SMALL*

Statistical power and validity of Ebola vaccine trials in Sierra Leone: a simulation study of trial design and analysis

Steven E Bellan, Juliet R C Pulliam, Carl A B Pearson, David Champredon, Spencer J Fox, Laura Skrip, Alison P Galvani, Manoj Gambhir, Ben A Lopman, Travis C Porco, Lauren Ancel Meyers, Jonathan Dushoff

Drawbacks

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Drawback: Efficiency

Randomization tests are as efficient as parametric methods

Efficiency depends on method of estimating the intervention effect

Within period methods are less efficient than methods that also utilise comparisons of control and intervention conditions within clusters (e.g. mixed effect models)

Drawback: Efficiency

	No. of switch points	ICC	Relative efficiency
Time trends varying between clusters	3	0.02	1.0
		0.08	0.9
	11	0.02	1.4
		0.08	1.9
Same time trends in all clusters	3	0.02	1.6
		0.08	3.5
	11	0.02	2
		0.08	6.0

Drawback: Common intervention effect

Randomization tests test a null hypothesis that the intervention effect has no impact on the outcome.

Parametric methods tend to only test the average effect of the intervention

Randomization test p-value will be small if

- the intervention has an effect on average

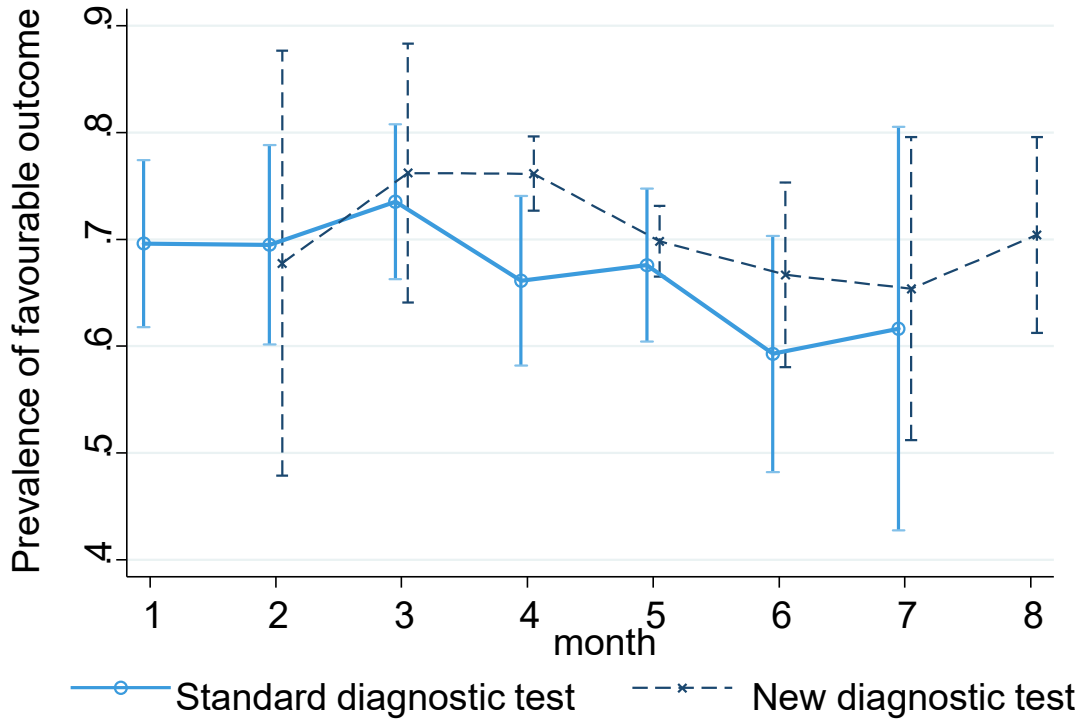
OR

- the intervention has a positive effect in some clusters but a negative effect in others.

Result is valid, just answering a different question

Unlikely to know whether this would be the case when writing the analysis plan

Drawback: Common intervention effect



How to do it?

The Stata Journal (2019)
19, Number 4, pp. 803–819 DOI: 10.1177/1536867X19893624

Permutation tests for stepped-wedge cluster-randomized trials

Jennifer Thompson
London School of Hygiene and Tropical Medicine
London, UK
jennifer.thompson@lshtm.ac.uk

Calum Davey
London School of Hygiene and Tropical Medicine
London, UK
calum.davey@lshtm.ac.uk

Richard Hayes
London School of Hygiene and Tropical Medicine
London, UK
richard.hayes@lshtm.ac.uk

James Hargreaves
London School of Hygiene and Tropical Medicine
London, UK
james.hargreaves@lshtm.ac.uk

Katherine Fielding
London School of Hygiene and Tropical Medicine
London, UK
katherine.fielding@lshtm.ac.uk

help swpermute ({})

Title

swpermute — Monte Carlo permutation tests for stepped wedge trial designs

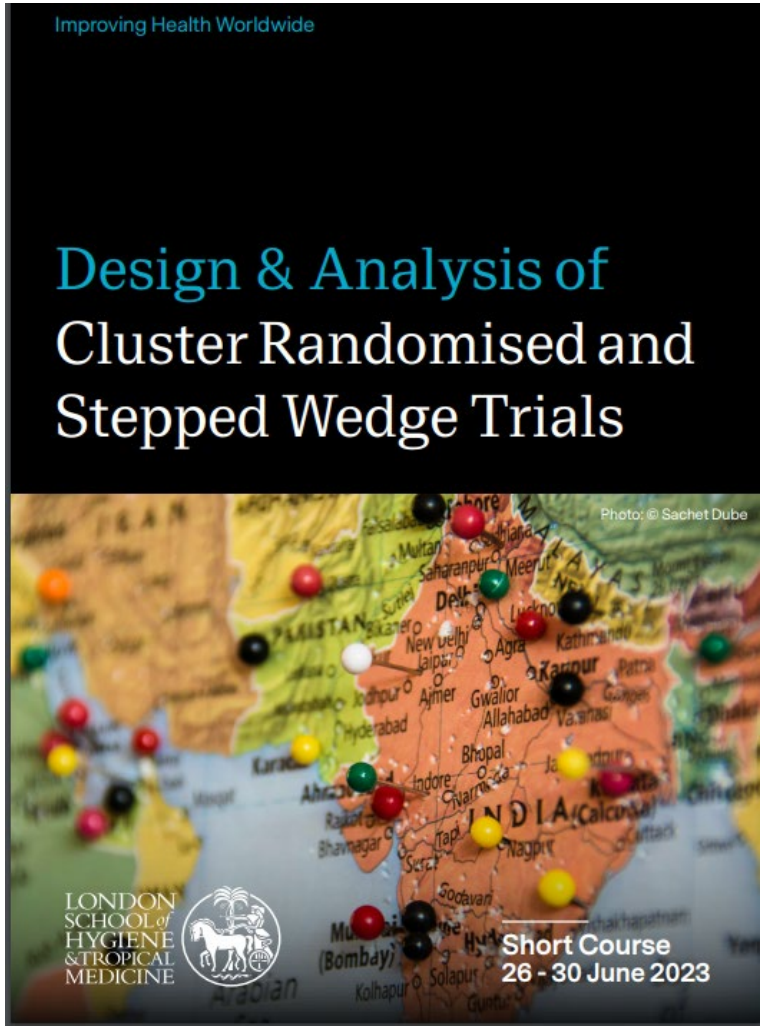
Syntax

swpermute *exp* **, cluster(***varname***) period(***varname***) intervention(***varname***)** [*options*] : *command*

Thank you

Acknowledgements: Calum Davey, Katherine Fielding, James Hargreaves, Richard Hayes

jennifer.thompson@lshtm.ac.uk



https://www.lshtm.ac.uk/study/courses/short-courses/cluster-randomised-trials

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

International Statistics & Epidemiology Group

P-value depends on random selection

Random selection of randomizations means you can get different p-values for the same data set:

- set a seed
- Use enough randomizations that the interpretation of the p-value is unambiguous.

statistic	obs_value	null	c	n	p	[95% Conf.	Interval]
_b[arm]	.4155579	0	2	1000	0.0020	.0002423	.0072058

Note: confidence interval is with respect to p
p-value is two-sided

Creating confidence intervals

Test different null intervention effects to find which fall within 5% evidence level.

Values	Trial arm	Testing diff = 4
1	0	1
2	0	2
3	0	3
4	0	4
5	1	1
6	1	2
7	1	3
8	1	4

Creating confidence intervals

Test different null intervention effects to find which fall within 5% evidence level.

How to create confidence intervals

$$p = \phi\left(\frac{\theta}{se(\theta)}\right)$$

$$se(\theta) = \frac{\theta}{\phi^{-1}(p)}$$

$$95\%\ CI \approx \theta \pm 1.96 * se(\theta)$$

Creating confidence intervals

TB diagnostic test trial:

$$\theta = \log(0.78)$$
$$p = 0.018$$
$$se \approx 0.104$$
$$95\%\ CI\ \approx (-0.45, -0.04)$$

Creating confidence intervals

Null log odds value	p	
-0.45	0.09	Inside CI
-0.5	0.04	Outside CI
-0.48	0.057	Inside CI
-0.49	0.045	Outside CI
CI lower bound = exp(-0.48) = 0.62		
CI upper bound = exp(-0.05) = 0.95		