Evaluation of Public Health Interventions: Recent Developments in Cluster Randomized Trials and Related Designs

Department of Epidemiology & Biostatistics, GWU, March 26 2018

Liz Turner, PhD
Assistant Professor, Department of Biostatistics & Bioinformatics and Duke Global Health Institute (DGHI)
Director, Research Design and Analysis Core, DGHI
Collaborators

**Duke Global Health**
Wendy O’Meara, Dorothy Dow, Joe Egger, Larry Park, Tazeen Jafar, Lijing Yan, Eve Puffer, Eric Green

**Duke Biostatistics**
Liz Delong, Fan Li, John Gallis, Alyssa Platt, Ryan Simmons, Kingshuk Roy, Siyun Yan

**GWU**
Brandon Kohrt

**Harvard**
Melanie Prague (& University of Bordeaux), Vikram Patel

**NIH**
David Murray

**LSHTM**
Rhian Daniel, Helen Weiss, Simon Brooker, Kate Halliday, Liz Allen

**UNC Chapel Hill**
Joanna Maselko, Brian Pence, John Preisser

**Other affiliations**
Atif Rahman (Liverpool); Siham Sikander (HDRF, Pakistan); Hengshi Yu (Minnesota), and many others…….
Overview

1. Motivating example
2. Clustering
3. Small # clusters & baseline covariate imbalance
4. Stepped wedge designs
Cluster randomized trials

Motivating example
Background and motivation
Health and Literacy Intervention (HALI) cluster randomized trial (CRT)

• 101 schools: 51 intervention and 50 control
  ~ 5000 children → ~ 50/school

• Intervention: screen & treat 1/term for 2 years

• Primary endpoint: malaria (yes vs. no) at 24 months
Background and motivation
Health and Literacy Intervention (HALI) cluster randomized trial (CRT)

Hypothesis: screening and treating children for malaria will lead to reduced prevalence of malaria
Background and motivation
Health and Literacy Intervention (HALI) cluster randomized trial (CRT)

Hypothesis: screening and treating children for malaria will lead to reduced prevalence of malaria

Factors related to malaria: age, geographic location, bed-net use etc.
Background and motivation
Health and Literacy Intervention (HALI) cluster randomized trial (CRT)

Malaria screening and treatment

Factors related to malaria (e.g., age, bednet use)

Malaria

Hypothesis: screening and treating children for malaria will lead to reduced prevalence of malaria

Factors related to malaria: age, geographic location, bed-net use etc.
Background and motivation
Health and Literacy Intervention (HALI) cluster randomized trial (CRT)

Hypothesis: screening and treating children for malaria will lead to reduced prevalence of malaria

Factors related to malaria: age, geographic location, bed-net use etc.
Background and motivation
Health and Literacy Intervention (HALI) cluster randomized trial (CRT)

Hypothesis: screening and treating children for malaria will lead to reduced prevalence of malaria

Factors related to malaria: age, geographic location, bed-net use etc.
Background and motivation
Health and Literacy Intervention (HALI) cluster randomized trial (CRT)

Level 2: Randomization at clinic (i.e., cluster) level

- Malaria screening and treatment
- Factors related to malaria (e.g., age, bednet use)

Level 1: Individual-level outcomes nested in schools
Background and motivation
Health and Literacy Intervention (HALI) cluster randomized trial (CRT)

Level 2: Randomization at clinic (i.e., cluster) level

- Malaria screening and treatment
- Factors related to malaria (e.g., age, bednet use)

Malaria

Level 1: Individual-level outcomes nested in schools

- Child-level outcomes within same school expected to be correlated with each other (i.e., to cluster)
Background and motivation

Health and Literacy Intervention (HALI) cluster randomized trial (CRT)

Level 2: Randomization at clinic (i.e., cluster) level

- Malaria screening and treatment
- Factors related to malaria (e.g., age, bednet use)

Level 1: Individual-level outcomes nested in schools

- Child-level outcomes within same school expected to be correlated with each other (i.e., to cluster)

Reduces power to detect treatment effect if same sample size used as under individual randomization
Implications of using CRT design

• CRT (statistical) price to pay
  • Lower power for same total sample size under individual randomization
  • Harder to detect an intervention effect

• So why use CRT design?
  • Intervention at cluster level (e.g., pump in village)
  • To avoid treatment contamination under individual randomization (e.g., HALI trial)
  • Logistically easier to implement trial
HALI trial

Two published outcomes papers

**Impact of Intermittent Screening and Treatment for Malaria among School Children in Kenya: A Cluster Randomised Trial**

Katherine E. Halliday\(^1\)*, George Okello\(^2\), Elizabeth L. Turner\(^3\), Kiambo Njagi\(^4\), Carlos Mcharo\(^5\), Juddi Kengo\(^5\), Elizabeth Allen\(^6\), Margaret M. Dubeck\(^7\), Matthew C. H. Jukes\(^8\), Simon J. Brooker\(^1,9\)

---

**Improving Literacy Instruction in Kenya Through Teacher Professional Development and Text Messages Support: A Cluster Randomized Trial**

Matthew C. H. Jukes\(^a,b\), Elizabeth L. Turner\(^c\), Margaret M. Dubeck\(^a,b,d\), Katherine E. Halliday\(^e\), Hellen N. Inyega\(^f\), Sharon Wolf\(^g\), Stephanie Simmons Zuilkowski\(^h\), and Simon J. Brooker\(^e\)
HALI trial

Two published outcomes papers

Impact of Intermittent Screening and Treatment for Malaria among School Children in Kenya: A Cluster Randomised Trial

Katherine E. Halliday¹*, George Okello², Elizabeth L. Turner³, Kiambo Njagi⁴, Carlos Mcharo⁵, Juddy Kengo⁶, Elizabeth Allen⁶, Margaret M. Dubeck⁷, Matthew C. H. Jukes⁸, Simon J. Brooker¹,⁹

Note: no evidence of an effect of intervention on malaria prevalence
HALI trial

Two published outcomes papers

Evidence of an effect on literacy outcomes due to a teacher intervention evaluated in same trial

JOURNAL OF RESEARCH ON EDUCATIONAL EFFECTIVENESS
http://dx.doi.org/10.1080/19345747.2016.1221487

Improving Literacy Instruction in Kenya Through Teacher Professional Development and Text Messages Support: A Cluster Randomized Trial

Matthew C. H. Jukes\textsuperscript{a,b}, Elizabeth L. Turner\textsuperscript{c}, Margaret M. DuBek\textsuperscript{a,b,d}, Katherine E. Halliday\textsuperscript{e}, Hellen N. Inyega\textsuperscript{f}, Sharon Wolf\textsuperscript{g}, Stephanie Simmons Zuilkowski\textsuperscript{h}, and Simon J. Brooker\textsuperscript{e}
Cluster randomized trials
Design challenge: clustering
Baseline clustering: malaria prevalence by school

Halliday, Karanja, Turner et al. (2012), Tropical Medicine & International Health, 17(5): 532-549
Complete clustering (ICC = 1)

>1 child /school gives no more information than 1 child/school since every child in a given school has the same outcome
No clustering (ICC = 0)

20% prevalence of malaria in each school
No structure by school - more like a random sample of children

- Malaria
- No malaria
Some clustering ($0 < ICC < 1$)

A more typical situation: e.g., cluster-prevalence 0% - 80%
Clustering in CRTs

- Outcomes in same clusters more similar to each other than to those in other clusters
- Previous example
  - 50 children in 10 schools
  - Effective sample size between 10 – 50
- Implications for statistical inference
- Major challenge in design & analysis
Measure of clustering: ICC

Intra-cluster correlation coefficient (ICC, $\rho$)

- Most commonly used measure of clustering
- Ranges: 0-1; 0 = no clustering; 1 = total clustering
- Typically $< 0.2$, commonly around 0.01 - 0.05

ICC for continuous outcomes:

$$\rho = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2} = \frac{\sigma_B^2}{\sigma_{Total}^2}$$

- Involves both Between-cluster & Within-cluster variance
Clustering in CRTs: implications for analysis

- 5 schools each randomized to control and intervention
- 100 eligible participants per clinic measured

Example from Hayes & Moulton (2009)
Clustering in CRTs: implications for analysis

Example from Hayes & Moulton (2009)

- 5 schools each randomized to control and intervention
- 100 eligible participants per clinic measured

Overall malaria prevalence in each trial: 10% vs 6%

Question: is intervention effective?
Clustering in CRTs: implications for analysis

Which trial shows more evidence of benefit?

Example from Hayes & Moulton (2009)
Clustering in CRTs: implications for analysis

Malaria prevalence

Trial A

15%
10%
5%

C I

Trial B

C I

C=Control
I=Intervention

Study features

Example from Hayes & Moulton (2009)
Clustering in CRTs: implications for analysis

Study features

- **Trial A:**
  - Lower between-school variability
  - Little overlap of I & C clinic-level proportions

- **Trial B:** overlap of I & C school-level proportions

Example from Hayes & Moulton (2009)
Clustering in CRTs: implications for analysis

- If ignore clustering: p-value = 0.02 for both trials
- Comparison of 10% (50/500) vs 6% (30/500) by chi-sq. test

Example from Hayes & Moulton (2009)
Clustering in CRTs: implications for analysis

- Trial B p-value accounting for clustered design = ?
- If ignore clustering: p-value = 0.02

Example from Hayes & Moulton (2009)
Clustering in CRTs: implications for analysis

- Trial B p-value accounting for clustered design = 0.17
- If ignore clustering: p-value = 0.02

Example from Hayes & Moulton (2009)
Clustering in CRTs: implications for analysis

- Trial A p-value accounting for clustered design = ?
- If ignore clustering: p-value = 0.02

Example from Hayes & Moulton (2009)
Clustering in CRTs: implications for analysis

- Trial A p-value accounting for clustered design = 0.01
- If ignore clustering: p-value = 0.02

Example from Hayes & Moulton (2009)
Clustering in CRTs: implications for analysis

- Trial A p-value accounting for clustered design* = 0.01
- Trial B p-value accounting for clustered design* = 0.17

*By using a cluster-level analysis where the 10 cluster-level proportions (5 per arm) are treated as continuous variables and analyzed with Wilcoxon rank sum test

Example from Hayes & Moulton (2009)
Summary: clustering & analysis

• Two example trials
  • Analyzed with cluster-level analysis
  • Overall sample size (# schools/trial) =10

• Both trials had same signal (10% vs 6%)
  • Totally different conclusions from each trial
  • Between-cluster variability Trial A < Trial B
  • $P$-value Trial A < $P$-value Trial B

• Important: If ignore clustered design, could claim ‘significant’ when not (eg, Trial B)
Summary: clustering & analysis

- Cluster-level analysis rarely used
- Typically use regression methods
  - Random effects / mixed effects models
  - Generalized estimating equations (GEE)
- Analyze individual-level data
  - e.g., N=1000 participants/trial not N=10 schools
Recent examples from my research

CRT methods

Review of Recent Methodological Developments in Group-Randomized Trials: Part 1—Design

In 2004, Murray et al. reviewed Elizabeth L. Turner, PhD, Fan Li, MSc, John A. Gallis, ScM, Melanie Prague, PhD, and David M. Murray, PhD

Review of Recent Methodological Developments in Group-Randomized Trials: Part 2—Analysis

In 2004, Murray et al. reviewed methodological developments Elizabeth L. Turner, PhD, Melanie Prague, PhD, John A. Gallis, ScM, Fan Li, MSc, and David M. Murray, PhD
Recent examples from my research

**BMJ Open**
Innovative public–private partnership to target subsidised antimalarials: a study protocol for a cluster randomised controlled trial to evaluate a community intervention in Western Kenya

Jeremiah Laktabai, Adriane Lesser, Alyssa Platt, Elisa Maffioli, Manoj Mohanan, Diana Menya, Wendy Prudhomme O'Meara, Elizabeth L. Turner

**STUDY PROTOCOL**
Reducing stigma among healthcare providers to improve mental health services (RESHAPE): protocol for a pilot cluster randomized controlled trial of a stigma reduction intervention for training primary healthcare workers in Nepal

Cluster randomized trials

Design challenge: clustering

Solution: design & analyze accounting for it
Cluster randomized trials

Design challenge: baseline imbalance
Motivating example CRT
Health and Literacy Intervention (HALI)

Goal: randomization → baseline balance of covariates
Check: baseline tables for 101 clusters (schools)
### Table 1. Baseline characteristics of 5,233 study children in the 50 control and 51 IST intervention schools.

<table>
<thead>
<tr>
<th>Characteristics; n (%)</th>
<th>Measure/Subcharacteristic</th>
<th>Control 50 schools</th>
<th>Intervention 51 schools</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>School characteristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exam score</td>
<td>Mean (SD)</td>
<td>223.4 (27.7)</td>
<td>225.8 (29.0)</td>
</tr>
<tr>
<td>School size</td>
<td>Median (IQR [min, max])</td>
<td>505 (308, 961) [85, 4,891]</td>
<td>568 (389, 692) [27, 359]</td>
</tr>
<tr>
<td>Enrolled class 1</td>
<td>Mean (SD) [min, max]</td>
<td>24.4 (3.3) [10,30]</td>
<td>25.8 (1.5) [23,30]</td>
</tr>
<tr>
<td>Enrolled class 5</td>
<td>Mean (SD) [min, max]</td>
<td>26.0 (4.6) [8,30]</td>
<td>27.3 (3.3) [16,32]</td>
</tr>
<tr>
<td>School programmes</td>
<td>Feeding</td>
<td>22 (44.0)</td>
<td>27 (52.9)</td>
</tr>
<tr>
<td></td>
<td>De-worming</td>
<td>50 (100.0)</td>
<td>49 (96.1)</td>
</tr>
<tr>
<td></td>
<td>Malaria control</td>
<td>9 (18.4)</td>
<td>12 (23.5)</td>
</tr>
<tr>
<td><strong>Child characteristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>Mean (SD)</td>
<td>10.1 (2.8)</td>
<td>10.3 (2.8)</td>
</tr>
<tr>
<td></td>
<td>5–9</td>
<td>1,041 (41.2)</td>
<td>1,069 (39.5)</td>
</tr>
<tr>
<td></td>
<td>10–12</td>
<td>877 (34.8)</td>
<td>925 (34.1)</td>
</tr>
<tr>
<td></td>
<td>13–20</td>
<td>605 (24.0)</td>
<td>716 (26.4)</td>
</tr>
<tr>
<td>Sex</td>
<td>Male</td>
<td>1,257 (49.8)</td>
<td>1,319 (48.7)</td>
</tr>
<tr>
<td>Child sleeps under net</td>
<td>Usually</td>
<td>1,668 (67.3)</td>
<td>1,682 (63.1)</td>
</tr>
<tr>
<td></td>
<td>Treated net</td>
<td>1,357 (83.3)</td>
<td>1,308 (80.5)</td>
</tr>
<tr>
<td></td>
<td>Last night</td>
<td>1,606 (96.3)</td>
<td>1,609 (95.7)</td>
</tr>
</tbody>
</table>

Halliday (2014), PLOS Medicine, 11(1) e1001594
http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.1001594
### Table 1. Baseline characteristics of 5,233 study children in the 50 control and 51 IST intervention schools.

<table>
<thead>
<tr>
<th>Characteristics; n (%)</th>
<th>Measure/Subcharacteristic</th>
<th>Control 50 schools</th>
<th>Intervention 51 schools</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>School characteristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exam score</td>
<td>Mean (SD)</td>
<td>223.4 (27.7)</td>
<td>225.8 (29.0)</td>
</tr>
<tr>
<td>School size</td>
<td>Median (IQR) [min, max]</td>
<td>505 (308, 961) [85, 4,891]</td>
<td>568 (389, 692) [2,230]</td>
</tr>
<tr>
<td>Enrolled class 1</td>
<td>Mean (SD) [min, max]</td>
<td>24.4 (3.3) [10,30]</td>
<td>25.8 (1.5) [23,30]</td>
</tr>
<tr>
<td>Enrolled class 5</td>
<td>Mean (SD) [min, max]</td>
<td>26.0 (4.6) [8,30]</td>
<td>27.3 (3.3) [16,32]</td>
</tr>
<tr>
<td>School programmes</td>
<td>Feeding</td>
<td>22 (44.0)</td>
<td>27 (52.9)</td>
</tr>
<tr>
<td></td>
<td>De-worming</td>
<td>50 (100.0)</td>
<td>49 (96.1)</td>
</tr>
<tr>
<td></td>
<td>Malaria control</td>
<td>9 (18.4)</td>
<td>12 (23.5)</td>
</tr>
<tr>
<td><strong>Child characteristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>Mean (SD)</td>
<td>10.1 (2.8)</td>
<td>10.3 (2.8)</td>
</tr>
<tr>
<td>5–9</td>
<td>1,041 (41.2)</td>
<td>1,069 (39.5)</td>
<td></td>
</tr>
<tr>
<td>10–12</td>
<td>877 (34.8)</td>
<td>925 (34.1)</td>
<td></td>
</tr>
<tr>
<td>13–20</td>
<td>605 (24.0)</td>
<td>716 (26.4)</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>Male</td>
<td>1,257 (49.8)</td>
<td>1,319 (48.7)</td>
</tr>
<tr>
<td>Child sleeps under net</td>
<td>Usually</td>
<td>1,668 (67.3)</td>
<td>1,682 (63.1)</td>
</tr>
<tr>
<td></td>
<td>Treated net</td>
<td>1,357 (83.3)</td>
<td>1,308 (80.5)</td>
</tr>
<tr>
<td></td>
<td>Last night</td>
<td>1,606 (96.3)</td>
<td>1,609 (95.7)</td>
</tr>
</tbody>
</table>

http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.1001594
### Table 1. Baseline characteristics of 5,233 study children in the 50 control and 51 IST intervention schools.

<table>
<thead>
<tr>
<th>Characteristics; $n$ (%)$^a$</th>
<th>Measure/Subcharacteristic</th>
<th>Control 50 schools</th>
<th>Intervention 51 schools</th>
</tr>
</thead>
<tbody>
<tr>
<td>School characteristics$^b$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exam score</td>
<td>Mean (SD)</td>
<td>223.4 (27.7)</td>
<td>225.8 (29.0)</td>
</tr>
<tr>
<td>School size</td>
<td>Median (IQR) [min, max]</td>
<td>505 (308, 961) [85, 4,891]</td>
<td>568 (389, 692) [23,30]</td>
</tr>
<tr>
<td>Enrolled class 1</td>
<td>Mean (SD) [min, max]</td>
<td>24.4 (3.3) [10,30]</td>
<td>25.8 (1.5) [23,30]</td>
</tr>
<tr>
<td>Enrolled class 5</td>
<td>Mean (SD) [min, max]</td>
<td>26.0 (4.6) [8,30]</td>
<td>27.3 (3.3) [16,32]</td>
</tr>
<tr>
<td>School programmes</td>
<td>Feeding</td>
<td>22 (44.0)</td>
<td>27 (52.9)</td>
</tr>
<tr>
<td></td>
<td>De-worming</td>
<td>50 (100.0)</td>
<td>49 (96.1)</td>
</tr>
<tr>
<td></td>
<td>Malaria control</td>
<td>9 (18.4)</td>
<td>12 (23.5)</td>
</tr>
<tr>
<td>Child characteristics$^b$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age$^c$</td>
<td>Mean (SD)</td>
<td>10.1 (2.8)</td>
<td>10.3 (2.8)</td>
</tr>
<tr>
<td>5–9</td>
<td>1,041 (41.2)</td>
<td>1,069 (39.5)</td>
<td></td>
</tr>
<tr>
<td>10–12</td>
<td>877 (34.8)</td>
<td>925 (34.1)</td>
<td></td>
</tr>
<tr>
<td>13–20</td>
<td>605 (24.0)</td>
<td>716 (26.4)</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>Male</td>
<td>1,257 (49.8)</td>
<td>1,319 (48.7)</td>
</tr>
<tr>
<td>Child sleeps under net</td>
<td>Usually</td>
<td>1,668 (67.3)</td>
<td>1,682 (63.1)</td>
</tr>
<tr>
<td></td>
<td>Treated net$^d$</td>
<td>1,357 (83.3)</td>
<td>1,308 (80.5)</td>
</tr>
<tr>
<td></td>
<td>Last night$^d$</td>
<td>1,606 (96.3)</td>
<td>1,609 (95.7)</td>
</tr>
</tbody>
</table>

Good balance of age

http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.1001594
Table 1. Baseline characteristics of 5,233 study children in the 50 control and 51 IST intervention schools.

<table>
<thead>
<tr>
<th>Characteristics; n (%)(^a)</th>
<th>Measure/Subcharacteristic</th>
<th>Control 50 schools</th>
<th>Intervention 51 schools</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>School characteristics(^b)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exam score</td>
<td>Mean (SD)</td>
<td>223.4 (27.7)</td>
<td>225.8 (29.0)</td>
</tr>
<tr>
<td>School size</td>
<td>Median (IQR) [min, max]</td>
<td>505 (308, 961) [85, 4,891]</td>
<td>568 (389, 692) [239, 2,890]</td>
</tr>
<tr>
<td>Enrolled class 1</td>
<td>Mean (SD) [min, max]</td>
<td>24.4 (3.3) [10,30]</td>
<td>25.8 (1.5) [23,30]</td>
</tr>
<tr>
<td>Enrolled class 5</td>
<td>Mean (SD) [min, max]</td>
<td>26.0 (4.6) [8,30]</td>
<td>27.3 (3.3) [16,32]</td>
</tr>
<tr>
<td>School programmes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feeding</td>
<td></td>
<td>22 (44.0)</td>
<td>27 (52.9)</td>
</tr>
<tr>
<td>De-worming</td>
<td></td>
<td>50 (100.0)</td>
<td>49 (96.1)</td>
</tr>
<tr>
<td>Malaria control</td>
<td></td>
<td>9 (18.4)</td>
<td>12 (23.5)</td>
</tr>
<tr>
<td><strong>Child characteristics(^b)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age(^c)</td>
<td>Mean (SD)</td>
<td>10.1 (2.8)</td>
<td>10.3 (2.8)</td>
</tr>
<tr>
<td>5–9</td>
<td></td>
<td>1,041 (41.2)</td>
<td>1,069 (39.5)</td>
</tr>
<tr>
<td>10–12</td>
<td></td>
<td>877 (34.8)</td>
<td>925 (34.1)</td>
</tr>
<tr>
<td>13–20</td>
<td></td>
<td>605 (24.0)</td>
<td>716 (26.4)</td>
</tr>
<tr>
<td>Sex</td>
<td>Male</td>
<td>1,257 (49.8)</td>
<td>1,319 (48.7)</td>
</tr>
<tr>
<td>Child sleeps under net</td>
<td>Usually</td>
<td>1,668 (67.3)</td>
<td>1,682 (63.1)</td>
</tr>
<tr>
<td>Treated net(^d)</td>
<td></td>
<td>1,357 (83.3)</td>
<td>1,308 (80.5)</td>
</tr>
<tr>
<td>Last night(^d)</td>
<td></td>
<td>1,606 (96.3)</td>
<td>1,609 (95.7)</td>
</tr>
</tbody>
</table>

Some imbalance of bednet use

http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.1001594
Small # of clusters & baseline imbalance

• CRTs often enroll small # (<40) clusters
• Randomization may not balance baseline covariates
• Baseline imbalance threatens internal validity
• Could address with adjusted analysis
• Better to use design strategy: ‘Restricted randomization’
  • Pair-matching
  • Stratification
  • Covariate-constrained randomization
Baseline covariate imbalance
Example: 8 schools (clusters)

Baseline malaria prevalence

20%

0%
Baseline covariate imbalance
Example: 8 schools (clusters)

**Question:** Why do we care about getting balance between treatment arms on school-level malaria prevalence?

*It might be related to prevalence in future!*
Baseline covariate imbalance
Example: 8 schools (clusters)

Example of extreme baseline imbalance using simple (ie, regular) randomization
Baseline covariate imbalance
Possible design solution 1: pair-matching
Baseline covariate imbalance
Possible design solution 1: pair-matching

One example of pair-matched randomization to control & intervention arms

Important: account for paired design in the analysis (eg, paired t-test or Wilcoxon signed rank test for cluster-level analysis or matched regression model)
Pair-matching in practice
Example from my research: published CRT outcomes paper

Efficacy of iron-supplement bars to reduce anemia in urban Indian women: a cluster-randomized controlled trial

Rajvi Mehta, Alyssa C Platt, Xizi Sun, Mukesh Desai, Dennis Clements, and Elizabeth L Turner

Duke University School of Medicine, Departments of Biostatistics and Bioinformatics and Pediatrics, and Duke Global Health Institute, Duke University, Durham, NC; and Department of Hematology and Immunology, B.J. Wadia Hospital, Mumbai, Maharashtra, India

Baseline covariate imbalance
Example: 8 schools (clusters)
Baseline covariate imbalance
Possible design solution 2: stratification

Baseline malaria prevalence

Stratum 1
Stratum 2

0%
20%
Baseline covariate imbalance
Possible design solution 2: stratification

An example of stratified randomization to control & intervention arms

Important: account for stratified design in the analysis (eg, stratified permutation test or fixed effect for strata in model-based analysis)
Stratification in practice
Example from my research: published CRT protocol paper

Turner et al. Trials (2016) 17:442

The effectiveness of the peer delivered Thinking Healthy Plus (THPP+) Programme for maternal depression and child socio-emotional development in Pakistan: study protocol for a three-year cluster randomized controlled trial

Elizabeth L. Turner¹,², Siham Sikander³, Omer Bangash³, Ahmed Zaidi³, Lisa Bates⁴, John Gallis¹,², Nima Ganga¹, Karen O’Donnell¹, Atif Rahman⁵ and Joanna Maselko⁶
Baseline covariate imbalance
Possible design solution 3: Constrained randomization

• Previous examples – only one school-level covariate
  • i.e., baseline malaria prevalence

• Often have multiple school-level covariates
  • Categorical & continuous
  • Pair-matching & stratification cannot easily handle this

• Need more general form of restricted randomization
  • Covariate-constrained randomization
Baseline covariate imbalance
Possible design solution 3: Constrained randomization

Example: balance two continuous cluster covariates

Baseline malaria prevalence
Bednet use
Baseline covariate imbalance
Possible design solution 3: Constrained randomization

An example of simple randomization to control & intervention arms

Not well-balanced on baseline malaria prevalence but reasonable balance on bednet use
Baseline covariate imbalance
Possible design solution 3: Constrained randomization

Neither randomization has good balance of both covariates across trial arms.

Solution: only allow randomizations that are “balanced enough” as measured by a “balance score” i.e., use covariate-constrained randomization
Baseline covariate imbalance
Possible design solution 3: Constrained randomization

This randomization could be “balanced enough”

Must account for constrained randomization design in the analysis
Covariate constrained randomization
Example from my research - methods
Covariate constrained randomization
Example from my research – software implementation

The Stata Journal (yyyy)  vv, Number ii, pp. 1–23

cvcrand and cptest: Efficient design and analysis of cluster randomized trials

John A. Gallis
Duke University Department of Biostatistics
Duke Global Health Institute
Durham, NC
john.gallis@duke.edu

Fan Li
Duke University Department of Biostatistics
Durham, NC
frank.li@duke.edu

Hengshi Yu
University of Michigan Department of Biostatistics
Ann Arbor, MI
hengshi@umich.edu

Elizabeth L. Turner
Duke University Department of Biostatistics
Duke Global Health Institute
Durham, NC
liz.turner@duke.edu

---

cvcrand: Efficient Design and Analysis of Cluster Randomized Trials

Constrained randomization by Raab and Butcher (2001) (doi:10.1002/1097-0258(20010215)20:2<118::AID-CDR21>3.0.CO;2-6), suitable for cluster randomized trials (CRTs) with a small number of clusters (e.g., 20 or fewer). The intervention effect can be estimated through a clustered permutation test introduced by Gail, et al. (1996) (doi:10.1002/sim.17410), the package baseline values of some cluster-level covariates and cluster permutation test on the individual-level output.
Cluster randomized trials

Design challenge: baseline imbalance

Solution:
use restricted randomization
Cluster randomized trials
Stepped-wedge designs
Parallel CRT vs. SW-CRT

Examples with 8 clusters: 1-year intervention

- Parallel design
- Complete stepped-wedge design
- Incomplete stepped-wedge design

Parallel CRT vs. SW-CRT

Examples with 8 clusters: 1-year intervention

- **Parallel design**
- **Complete stepped-wedge design**
- **Incomplete stepped-wedge design**

CRT analysis: treatment effects

Estimated (primarily) using between-cluster information

Estimated using both **vertical** & **horizontal** (ie, within-cluster) information

SW-CRT design and analysis
Examples from my research

Sample size determination for GEE analyses of SW-CRTs
Li F, Turner EL, Preisser J. Under review.

Optimal allocation of clusters in cohort SW designs

Covariate constrained randomization for the design of parallel and SW-CRTs
- Invited session at Society of Clinical Trials Annual Meeting, May 2018
- Joint work with Karla Hemming (University of Birmingham), Andrew Copas (University College London) and Fan Li (Duke)
Summary

Evaluation of Public Health Interventions: Recent Developments in Cluster Randomized Trials and Related Designs
Summary

• Recent developments in CRTs

  1. Motivating example
  2. Clustering
  3. Small # clusters & baseline covariate imbalance
  4. Stepped wedge designs
References - Statistical

- Campbell MK, Grimshaw JM, Elbourne DR (2004). Intracluster correlation coefficients in CRT: empirical insights into how they should be reported *BMC Medical Research Methodology* 4:9
References – Motivating example


