Methodological issues in the design and analysis of cluster randomised trials

Clémence Leyrat

Department of Medical Statistics, LSHTM
In randomised trials, **different randomisation units** can be used (participants or clusters of participants).

The similarity of the observations within the same cluster is quantified by the **intraclass correlation coefficient** (ICC).
CRTs share characteristics with IRTs and observational studies:

- **Observational studies**
  - Selection bias
  - Real-life
  - Difficulty to combine evidence
  - « Complex » statistical analysis

- **CRTs**

- **IRTs**
  - Balance at baseline
  - Ideal-world
  - Meta-analyses
  - « Simple » statistical analysis
Biases in CRTs:

How to detect them in CRTs
Biases in CRTs

In CRTs, bias can arise from the design, according to:

- the chronology
- recruitment procedure
- blinding

Development of a graphical tool\(^1\): **Timeline cluster**

\(^1\) Caille et al.. Timeline cluster: a graphical tool to identify risk of bias in cluster randomised trials. BMJ. 2016;354:i4291
Timeline cluster

Clusters: GP practices in Australia
Intervention: Nurse training on coaching on glycaemic control of type 2 diabetes
Outcome: Glycated haemoglobin
Timeline cluster is a **qualitative tool** to identify the risk of bias

Can be adapted for more complicated designs such as cluster cross-over designs

This graph should be reported in protocols and publications
Pragmatism in CRTs:

Do CRT and IRT estimate the same effects?

Can we meta-analyse them together?
CRTs are thought to be more pragmatic than IRTs

How does it impact intervention effect estimates?

⇒ Disagreements in the literature

Meta-epidemiological study to compare intervention effect estimates in CRTs and IRTs:

- Inclusion of Cochrane systematic reviews
- 76 meta-analyses with a binary outcome:
  917 trials: 734 IRTs and 183 CRTs
- 45 meta-analyses with a continuous outcome:
  541 trials: 410 IRTs and 131 CRTs
Results

For binary outcomes:  
\[ \text{ROR} = 1.00 \ [0.93;1.08] \]

Similar result in subgroups:  
- objective v. subjective  
- pharmacological v. non pharmacological  
- active v. inactive control

For continuous outcomes:  
\[ \text{DSMD} = 0.13 \ [0.06;0.19] \]

High heterogeneity  
No difference when adjusting on sample size
Implications

From this study, no substantial differences between intervention effect estimates from IRTs and CRTs:

- They can be **meta-analysed together** IF clustering accounted for properly
- They estimate the “same” effect
Statistical analysis:

The intraclass correlation
3 main approaches to analyse CRTs: cluster-level analyses, mixed-models or GEEs

When only few clusters are randomised: inflated type I error rate for mixed-models and GEEs

Small-sample corrections available in standard software packages but:

- Not often implemented in practice\(^1\)
- Negative impact on power

\(^1\)Kahan et al. Increased risk of type I errors in cluster randomised trials with small or medium numbers of clusters: a review, reanalysis, and simulation study. Trials. 2016 Sep 6;17(1):438.
Small sample size

Cluster-level analysis

Mixed model

GEE

Background
CRTs
Challenges
Bias
Pragmatism
Statistical analysis
Small-sample
The ICC
Discussion
The ICC: an outcome?

The variation of the ICC could be useful in providing information about the **heterogeneity of the intervention effect**

⇒ Should this difference be reported along with the outcome?

For binary outcomes, the ICC depends on the prevalence

⇒ **Difficult to interpret** if there is a positive intervention effect

Ongoing work on the **rescaling of binary ICCs** to make them independent of the prevalence
**Challenges** in the design and analysis of CRTs not encountered in IRTs:

- Risk of selection bias
- Correlation in the data
- ...

However, the conclusions from CRTs are *similar* to those from IRTs whilst avoiding limitations in the implementation of IRTs.

A lot of unresolved questions...
Acknowledgements

Agnès Caille
Bruno Giraudieu
Elsa Tavernier

Sandra Eldridge
Brennan Kahan
Sally Kerry

Baptiste Leurent
Katy Morgan

Allan Donner
Cluster cross-over trial

**Clusters:** hospital wards

**Intervention:** medication reconciliation

**Outcome:** drug-related problem

Recruitment bias $\times$

Performance bias $\times$

Detection bias $\times$
Ongoing work on the **rescaling of binary ICCs** to make them independent of the prevalence

<table>
<thead>
<tr>
<th>Arm</th>
<th>Prevalence (%)</th>
<th>Binary ICC</th>
<th>Continuous ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malathion</td>
<td>85.0</td>
<td>0.44</td>
<td>0.74</td>
</tr>
<tr>
<td>Ivermectine</td>
<td>95.2</td>
<td>0.61</td>
<td>0.95</td>
</tr>
</tbody>
</table>