THE STEPPED-WEDGE CLUSTER RANDOMISED TRIAL:
RISKS OF BIAS AND LEVELS OF EVIDENCE

SECOND INTERNATIONAL CONFERENCE ON
STEPPED WEDGE TRIAL DESIGN
KARLA HEMMING
19TH MARCH, 2018; YORK
BRIEF HISTORY - THE FIRST SW-CRT

Gambia Hepatitis Intervention Study (GHIS):

- Nation-wide trial of the HBV vaccine
- Vaccine programme rolled-out in a phased way (random) over 4 years (1986 to 1990)
- Outcome: Hepato Cellular Carcinoma (HCC) ascertained from a national cancer registry. 30 years follow-up
- “This was because instantaneous universal vaccination in the country was impossible for logistic and financial reasons.”

Stepped wedge design

- 63,512 infants unvaccinated
- 61,065 infants vaccinated

WHAT HAS HAPPENED SINCE THE GAMBIA TRIAL

- Increasing uptake:
  - Protocols
  - Full trials
  - Methodological research

- Funding supported by many national funding streams such as:
  - NIHR (UK)
  - NIH (US)

METHODOLOGICAL PROGRESS – QUICK OVERVIEW

▶ Sample size methodology:
  • Hussey and Hughes
  • Design Effects / Software
  • Extended correlation structures
  • Statistical efficiency

▶ Design and analysis:
  • Incomplete designs / hybrid designs
  • Model misspecification
  • Methodological and quality of reporting reviews

https://clusterrcts.shinyapps.io/rshinyapp/
The SW-CRT is the epitome of the pragmatic trial;

Presents an unprecedented opportunity to increase the robustness with which healthcare-policy interventions are evaluated.

Questions:

• Under what circumstances is the estimate of treatment-effect from a SW-CRT as robust as that of a parallel-CRT?

• When should the SW-CRT be used in preference to a parallel CRT?
ISSUES TO CONSIDER

▶ How to design these studies so as those people (or data) that are included are not selectively different between those who receive different interventions/treatments (i.e. healthcare policies) – these are referred to as selection biases.

▶ The other issue is how to analyse the data from this complicated design so as to obtain a fair estimate of how the intervention is working – these are referred to as analysis biases.
AN EXAMPLE OF A MORE RECENT TRIAL: THE EPOCH STUDY

- Evaluation of an evidence based integrated care pathway (ICP)
- Setting:
  - 90 hospitals (15 geographical clusters); Emergency laparotomy
- Outcome:
  - 90 day mortality
- Sample size:
  - 27,500 patients
  - 90% power to detect a change from 25% to 22%
- Routinely collected outcome
- Funded by UK NIHR

Average **number of clusters**:  
Median 17 [IQR: 8 to 38]

Average **number of steps**:  
Median 4 [IQR: 2 to 6]

Average **study duration** (months):  
Median 16 [IQR: 8 to 24]
WHEN IS THE SW-CRT BEING USED….

- When no other randomised study design is feasible:
  - Gambia study
- When the alternative might be the parallel cluster randomised trial:
  - EPOCH study

WHY IS A RATIONALE IMPORTANT

▶ In the SW-CRT, this justification includes:
  • Cluster randomisation;
  • Roll-out of the intervention to all clusters;
  • Staggered roll-out of the intervention.

▶ Why this is important:
  • Cluster randomisation increases the sample size;
  • Roll-out to all clusters might also increase number of participants and clusters exposed;
  • Delayed exposure where intervention works.
The SW-CRT is more complicated in its design, analysis, and implementation than the parallel CRT.

Risks of bias in the SW-CRT may be higher than in a parallel CRT.
WHERE ON THE HIERARCHY OF EVIDENCE IS THE SW-CRT?

- Is there a distinction between the level of evidence provided from:
  - A parallel CRT;
  - A SW-CRT?

- Why this is important:
  - *Inform when the parallel CRT is preferential to the SW-CRT.*
Consider the march of progress from the parallel CRT to the SW-CRT

Look at some possible risks of bias in turn

Weigh up the likelihood of these risks by design

Recommendations for when the SW-CRT should be used
THE PARALLEL CLUSTER RANDOMISED TRIAL

- Typically have a small number of clusters.
  - Average 21 [IQR: 12-52]

- Finding of baseline imbalance is not uncommon.

Baseline imbalance

- Imbalance of key characteristics between participants exposed to intervention and control.
- Chance imbalance likely with a small numbers of clusters.
- Imbalance due to selection / recruitment biases likely when recruitment takes place after randomisation.

Perineal Assessment and Repair Longitudinal Study (PEARLS): a matched-pair cluster randomized trial Khaled M K Ismail, Christine Kettle, Sue E Macdonald, Sue Tohili, Peter W Thomas, Debra Bick BMJ 2014;348:g209.
MITIGATION STRATEGIES

Systematic imbalance

- Imbalance due to selection / recruitment biases likely when recruitment takes place after randomisation.

- Recruitment:
  - Before randomisation
  - Someone independent to study
  - Complete enumeration

Chance imbalance

- Chance imbalance likely with a small numbers of clusters.

- Restricted randomisation:
  - For known confounders
  - Not for unknown confounders


PARALLEL CRTS WITH BASELINE PERIOD

- Can adjust for baseline values (of outcome).

- Less than ideal?
  - Adjusted and unadjusted treatment effects might be discordant.
  - Statistical adjustment assumes a common relationship between baseline values and outcome in all clusters.

MARCH OF PROGRESS TO THE SW-CRT

- Includes within-cluster comparison;
- Every cluster contributes to control and intervention;
- Design should be less susceptible (compared to a parallel CRT) to risks of *chance baseline imbalance*.

TEMPORAL CONFOUNDING

- Observations under the control condition occur systematically earlier under the intervention condition.
- Outcomes might change over time.
- Time is therefore a potential confounder in a SW-CRT.
TEMPORAL CONFOUNDING IN PRACTICE

- SW-CRTs typically run over an extended duration (~ 2 years);
- External influences can affect outcomes of interest;
- Adjustment (modelling) therefore essential;

Temporal trend or effect of intervention?

POTENTIAL RISKS OF BIAS
IN THE SW-CRT
WHAT ARE THE POTENTIAL RISKS FOR BIAS IN A SW-CRT?

- Look at several common risks:
  - Temporal confounding
  - Selection bias
  - Within-cluster contamination
  - Delayed treatment effects
## RISK OF BIAS 1: TEMPORAL CONFOUNDING

<table>
<thead>
<tr>
<th>Detailed Description</th>
<th>Why this is important</th>
<th>Mitigating strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations collected under the control condition are, on average, from an earlier calendar time than observations collected under the intervention condition.</td>
<td>Changes external to the trial may create underlying secular trends. Where the same participants are repeatedly assessed, their health status might improve (or worsen) over the study. Because time is associated with both the treatment condition and the outcome, it means that time is a potential confounder.</td>
<td>Analysis and sample size should allow for the confounding effect of time.</td>
</tr>
</tbody>
</table>
### RISKS OF BIAS 2: SELECTION BIAS

<table>
<thead>
<tr>
<th>Detailed Description</th>
<th>Why this is important</th>
<th>Mitigating strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td>SW-CRTs can take a <strong>complete enumeration</strong> of the cluster, a random sample of individuals, or recruit participants into the trial. Furthermore, participants might be continuously recruited into the trial as they present; or all participants might be recruited at the beginning of the trial.</td>
<td>Information on how observations were sampled is important to elicit risks of bias. Studies which take a complete enumeration have <strong>lower risks of bias</strong> as do studies which recruit all participants at a fixed point in time before randomisation has occurred; studies which continuously recruit participants have higher potential for identification and recruitment biases.</td>
<td>Methods to reduce the risk of bias include taking a complete enumeration of the entire cluster-period; recruiting all participants before randomisation; or recruiting by <strong>someone independent</strong> to the study.</td>
</tr>
</tbody>
</table>
# RISK OF BIAS 3: WITHIN-CLUSTER CONTAMINATION

<table>
<thead>
<tr>
<th>Detailed Description</th>
<th>Why this is important</th>
<th>Mitigating strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td>In SW-CRTs, some or all of the clusters will be exposed to both the control and intervention conditions. Participants might only be exposed to one of the treatment conditions, or both.</td>
<td>Where the duration of exposure is long, it may be possible to sample the same participants under the control condition and later under the intervention condition.</td>
<td>In trials with long exposure delayed assessment of outcome should be avoided to prevent participants recruited under the control condition might later become exposed to the intervention condition.</td>
</tr>
</tbody>
</table>
### RISK OF BIAS: DELAYED TREATMENT EFFECTS

<table>
<thead>
<tr>
<th>Detailed Description</th>
<th>Why this is important</th>
<th>Mitigating strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sometimes the effect of the intervention is expected to materialise immediately, and sometimes there is a delay before its effect will be realised.</td>
<td>When there is a delay before the effect of the intervention is realised the estimate of effectiveness can be attenuated.</td>
<td>Where there is expected to be a delay before the effect of the intervention is materialised a transition period can be built into the design of the study.</td>
</tr>
</tbody>
</table>
SUSCEPTIBILITY OF BIAS
BY DESIGN
SUSCEPTIBILITY TO RISKS OF BIAS BY DESIGN

SW-CRT
- At risk of biases due to:
  - (Selection biases)
  - Temporal trends
  - (Within-cluster contamination)
  - (Delayed treatment effects)
- Protected in part from:
  - Chance imbalance

Parallel CRT
- At risk of biases due to:
  - (Selection biases)
  - Chance imbalance
- Protected from:
  - Temporal trends
  - (Within-cluster contamination)
  - (Delayed treatment effects)
WHERE ON THE HIERARCHY OF EVIDENCE IS THE SW-CRT?

- CRTs with a large number of clusters increase likelihood of:
  - Internal validity (protected from chance imbalance)
  - External validity

WHERE ON THE HIERARCHY OF EVIDENCE IS THE SW-CRT?

- CRTs and SW-CRTs which take a complete enumeration or recruit before randomisation will reduce risk of selection biases:
  - Routinely collected data; no patient recruitment.

WHERE ON THE HIERARCHY OF EVIDENCE IS THE SW-CRT?

- SW-CRTs with *careful design* can be protected against within-cluster contamination:
  - Transition periods
  - No delayed assessment of outcome with long treatment exposure.

WHERE ON THE HIERARCHY OF EVIDENCE IS THE SW-CRT?

- SW-CRTs and CRTs with baseline adjustment can be protected against chance imbalance:
  - Adjust for temporal trends

RECOMMENDATIONS
WHEN A SW-CRT IS PREFERABLE TO A PARALLEL CRT
LOGISTIC CONSTRAINTS (THE GAMBIA TRIAL)

- When a SW-CRT will almost certainly be preferable:
  - When the intervention will be rolled-out irrespective of any evaluation (CRT not feasible);
  - When logistical issues mean that the intervention must be phased into practice.
SMALL NUMBER OF CLUSTERS

- When there are only a small number of clusters the parallel CRT will be at risk of bias due to chance imbalance:
  - Add baseline assessment of outcome;
  - Use restricted randomisation;
- Consider a march of progress to a SW-CRT.
LARGE NUMBER OF CLUSTERS

- When there are a large number of clusters:
  - Important to consider the susceptibility of design to bias;
  - Some risks of biases can be minimised.

- Consider a parallel CRT to avoid over reliance on a model based analysis.
CONCLUSIONS

- The SW-CRT is the epitome of the pragmatic trial;
- Presents an unprecedented opportunity to increase the robustness with which healthcare-policy interventions are evaluated.

Questions:

- Under what circumstances is the estimate of treatment-effect from a SW-CRT is as robust as that of a parallel-CRT?
  - When there are only a small number of clusters
- When should the SW-CRT be used in preference to a parallel CRT?
  - When a parallel trial is not feasible
# Glossary of Terms

<table>
<thead>
<tr>
<th>Term</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster</td>
<td>The unit of randomisation.</td>
</tr>
<tr>
<td>Cluster-period</td>
<td>A grouping of observations by time of measurement and cluster.</td>
</tr>
<tr>
<td>Step</td>
<td>A planned point at which a cluster or group of clusters crosses from control to intervention.</td>
</tr>
<tr>
<td>Period</td>
<td>A grouping of observations by time of measurement.</td>
</tr>
<tr>
<td>Duration of period</td>
<td>Time (e.g. months) between each step.</td>
</tr>
<tr>
<td>Sequence of treatments (often abbreviated to sequence or allocated sequence)*</td>
<td>A sequence of codes defining the order of implementation of the treatment conditions for each cluster. More than one cluster can be allocated to each sequence.</td>
</tr>
<tr>
<td>Intervention condition*</td>
<td>The treatment under evaluation.</td>
</tr>
<tr>
<td>Control condition</td>
<td>The comparator treatment.</td>
</tr>
<tr>
<td>Transition period</td>
<td>The time needed to fully embed the intervention. A transition period may have the same or different duration than a measurement period.</td>
</tr>
<tr>
<td>Participant</td>
<td>A participant is someone on whom investigators seek to measure the outcome of interest.</td>
</tr>
<tr>
<td>Research participant</td>
<td>A research participant denotes a human research subject from the standpoint of ethical considerations.</td>
</tr>
</tbody>
</table>