Cluster Randomised Controlled Trials

STATS 773: DESIGN AND ANALYSIS OF CLINICAL TRIALS

Dr Yannan Jiang
Department of Statistics

May 14th 2012, Monday, 09:00-10:30
Cluster Randomised Trial

A group of individuals is randomly allocated to receive either the intervention or the control.
Cluster Randomised Trials (CRT)

• Unit of randomisation is a cluster (i.e. a group of individual subjects)
  - GP clinic, hospital, school, geographic area
• Data collection is often at the individual level
  - Blood pressure, BMI
  - Number of hospital admissions
  - Time to healing
• Subjects in the same cluster are more like one another than those in different clusters
  - Patients registered at the same clinic
  - Students who attend the same school
Why do CRT?

• Intervention at the level of a cluster
  – All individuals in the same cluster will receive the same intervention
  – Administrative convenience, e.g. new hypertension screening implemented at a GP clinic

• Possibility of contamination
  – Individuals in the same cluster may influence each other and dilute the treatment effect

• Potential ethical issues
  – Seems unethical to treat patients in the same clinic differently
Complexity with CRT

• Randomisation is limited by the number of clusters recruited to the study
  – Often small due to logistic and budget
  – Clusters may vary, and hard to match or stratify

• Reduced effective sample size than individual randomized trials
  – Correlation among individuals in the same cluster need to be taken into account
  – The design effect is related to Intra-cluster Correlation Coefficient (ICC) and the number of individuals per cluster (m)

• Methods ignoring clustering may mislead data analysis
  – Assume independence among correlated observations within the same cluster
  – Standard errors and confidence intervals tend to be too small
  – Small P-values introduce false positive results
A Simulation Study

• Two groups, with two clusters per group and 10 subjects per cluster

• Generate four cluster means $\bar{Y}_j$ and a random sample within each cluster
  
  – $\bar{Y}_j \sim$ Normal (mean=10, SD=2)
  – $Y_{ij} \sim$ Normal (mean=0, SD=1)

• TRUE null hypothesis: No difference between the means in the two groups

• Conduct simple two-sample t-test using:
  
  – Individual sample data ($Y_{ij}$) ignoring the clustering
  – Four group means ($\bar{Y}_j$) only
1000 Simulation Results

• What we expect to see:
  – With a valid test, we expect
    • 50 (5%) significant differences with \( p \)-value < 0.05
    • 10 (1%) significant differences with \( p \)-value < 0.01

• What we actually see in simulations:
  – With a two-sample t-test using individual data ignoring the clustering
    • 600 (60%) significant differences with \( p \)-value < 0.05
    • 502 (50.2%) significant differences with \( p \)-value < 0.01
  – With a two-sample t-test using four cluster means only
    • 53 (5.3%) significant differences with \( p \)-value < 0.05
    • 14 (1.4%) significant differences with \( p \)-value < 0.01
Intra-cluster Correlation Coefficient (ICC)

- A measure of the relatedness of clustered data by comparing the variance within clusters with the variance between clusters
- Two sources of variability with clustered data:
  - Variance within the cluster ($\sigma_w^2$)
  - Variance between the clusters ($\sigma_b^2$)
- ICC is calculated as:
  $$\rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2}$$
- ICC ranges from 0 to 1
  - $\rho=0$ indicates no correlation between subjects within a cluster, i.e. same as a simple random sample
  - $\rho=1$ indicates all subjects within a cluster are identical, i.e. $m=1$ (the worse case scenario)
Estimation of ICC (1)

- We can estimate ICC from the analysis of variance (ANOVA) table with a continuous outcome

One-way Analysis of Variance:

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>DF</th>
<th>Mean Square</th>
<th>F-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between classes</td>
<td>BSS</td>
<td>BDF=g-1</td>
<td>BMS=BSS/BDF</td>
<td>BMS/WMS</td>
</tr>
<tr>
<td>Within class</td>
<td>WSS</td>
<td>WDF=g(m-1)</td>
<td>WMS=WSS/WDF</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>BSS+WSS</td>
<td>BDF+WDF</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ICC = (BMS − WMS) / [BMS + (m-1)WMS], or equivalently,
ICC = (F − 1) / (F + m − 1)

Here $g$ is the number of classes, $m$ the number of subjects per class
Estimation of ICC (2)

• For binomial responses, e.g. the success rate, an analogous measure to ICC is the kappa coefficient ($\kappa$)
  – Only work for small clusters (e.g. $m=2$)
• Let $p^*$ the proportion of the control group with concordant clusters
  – A concordant cluster with $\kappa=1$ is the one where all responses within a cluster are identical
• Let $p_c$ the underlying success rate in the control group
• The kappa coefficient is estimated as:
  $$\frac{p^* - [p_c^m + (1-p_c)^m]}{1 - [ p_c^m + (1-p_c)^m ]}$$
Sources of ICC

• The following sources may be used to estimate the ICC for a new trial:
  – Pilot study data
  – Other trials with similar clusters and outcome variable

• Many trial reports do not give the ICC!

CONSORT statement: extension to cluster randomised trials
Marion K Campbell, Diana R Elbourne and Douglas G Altman
BMJ 2004;328:702-708
doi:10.1136/bmj.328.7441.702
Design Effect and Sample Size

• The design effect for a CRT with equal sized cluster is:
  \[ Deff = 1 + (m - 1) \times ICC \]
  – \( m \) is the number of subjects per cluster
  – ICC is the intra-class correlation coefficient

• The required sample size for a CRT is estimated as:
  \[ N_c = N_s \times Deff, \]
  – \( N_c \) is the sample size for a cluster randomised trial
  – \( N_s \) is the sample size for a standard randomised controlled trial
  – Same when \( Deff = 1 \) (i.e. \( m = 1 \) or ICC=0)

• With unequal sized cluster, \( m \) is replaced by \( \tilde{m} \) where

  \[
  \tilde{m} = \frac{\sum m_i^2}{\sum m_i}
  \]
Size of Clusters

- Unequal cluster sizes increase the design effect

\[
\text{Var}(m_i) = \frac{1}{k-1} \left( \sum m_i^2 - \left( \frac{\sum m_i}{k} \right)^2 \right) > 0 \text{ if } m_i \text{ not equal (k clusters)}
\]

\[
\sum m_i^2 - \left( \frac{\sum m_i}{k} \right)^2 > 0
\]

\[
\sum m_i^2 > \left( \frac{\sum m_i}{k} \right)^2
\]

\[
\tilde{m} = \frac{\sum m_i^2}{\sum m_i} > \frac{\sum m_i}{k} = \bar{m}
\]

Hence \( \tilde{m} > \bar{m} \) if \( m_i \) not equal.
Number of Clusters

• The design effect depends on variances both within and between the clusters
• With small number of clusters per group, $\sigma_b^2$ has few degrees of freedom and the t-distribution needs to be used rather than the Normal
• In large sample approximation, sample size calculation has the multiplier $(z_{\alpha/2} + z_\beta)^2$, where $Z_\beta$ is now replaced by t-statistics with a small $df$
  – This will reduce the effective sample size even more!
What Increase the Sample Size?

• The following parameters may inflate the design effect and sample size in a CRT:
  – Large ICC
  – Large cluster size
  – Varying cluster sizes
  – Small number of clusters

• If we have enough number of clusters with reasonably small cluster size and small ICC, $Deff \sim 1$
  – Little effect if the clustering is ignored
Statistical Analysis

• A method that accounts for the cluster effect will be valid and efficient, compared to the naïve method ignoring correlated observations within a cluster

• Several analysis approaches can be used:
  – Simple analysis adjusting for SEs with the design effect
  – Summary statistics at the cluster-level
  – Robust variance estimates
  – Multilevel models
  – Generalized estimating equations (GEE)
  – Others, e.g. Bayesian hierarchical models
Simple Adjustment of Standard Tests

• Carry out standard tests on individual data assuming there is no clustering
• Adjust the standard error (SE) by multiplying with the square root of the design effect
• Calculate the associated p-value and confidence interval using the adjusted SE

➢ No longer recommended with other advanced methods widely available...
Simple Adjustment of Standard Tests

Advantages
• Simple
• Quick

Disadvantages
• Crude adjustment
• No adjustment for covariates
• Only valid when clusters are of equal size
Cluster-Level Analysis

• A two-stage process
  1. A summary measure of study endpoint is obtained for each cluster, with $c_i$ estimates per group
  2. Statistical inference is made on cluster-level summaries

• Point estimates are obtained for each group using
  – cluster summaries $\bar{r}_i = \frac{1}{c_i} \sum r_{ij}$
  – individual values to account for unequal cluster sizes $p_i = \frac{\sum m_{ij} p_{ij}}{\sum m_{ij}}$
  – log transformation to account for skewed distribution of summary data

• Statistical inferences
  – Parametric T-test
    • Reasonably robust to departure from the underlying assumptions
  – Nonparametric Cluster Permutation test
    • An exact test permuting the treatment assignment to estimate the distribution of the test statistics under the null hypothesis of no treatment effect
    • Use clusters of observations rather than individual observations
    • Make no distributional assumptions
Cluster-Level Analysis

Advantages
• Account well for clustering

Disadvantages
• Does not use the richness of the data
• Lower power with only one point per cluster
• Can only adjust for covariates at the cluster level
Robust Standard Error Estimate

• Huber-White standard errors are used to allow the fitting of a model that contains heteroscedastic residuals in a clustered sampling

• The sandwich estimator yields asymptotically consistent and robust covariance matrix estimates
  – without making distributional assumptions
  – valid even if the assumed model is incorrect

• Can be implemented using SPSS, STATA, SAS
Huber-White Robust Estimate

Advantages
• Take into account ICC in data analysis
• Can adjust for individual-level covariates

Disadvantages
• Does not work well if the number of clusters is small, i.e. < 10 clusters
  – Normally requires 30 or more clusters
Multilevel Models

• Special form of mixed model regression where the data has a hierarchical structure
  – Cluster and individual levels
  – Can handle more than two levels, e.g. repeated observations on each individual

 Advantages:
  – Allows complex correlation structure and additional clustering
  – Adjust for both cluster and individual level covariates

 Disadvantages:
  – Computationally intensive, and may have convergence problem
  – Does not work well with small number of clusters
Fixed and Random Effects

• Fixed effect
  – A factor is regarded as a fixed effect if the categories that were studied are the only relevant categories
  – This is unlikely

• Random effect
  – The categories that were actually studied represent only a sample of those potentially available for study from the population of interest
  – More plausible
A Simple Mixed Model

• Fixed effects model:
  \[ \text{Outcome} = \text{Mean} + \text{Treat} + \text{Cluster} + \text{error} \]
  – The cluster effect is estimated for each cluster (df = g - 1)
  – Random error only

• Random effects model:
  \[ \text{Outcome} = \text{Mean} + \text{Treat} + \text{Cluster} + \text{error} \]
  – The cluster variable has 0 mean and unknown variance (df = 1)
  – Random cluster effect plus error
A Simple Mixed Model

• Model assumptions
  – Both random cluster effect and random error have Normal distributions with mean zero and uniform variances
  – Independent fixed and random effects
  – Uniform correlations between pairs of subjects in a cluster

• Model fit
  – Iteratively using Restricted Maximum Likelihood (REML) and the Expectation Maximization (EM) algorithm
  – ICC can be directly estimated from the cluster ($\sigma_b^2$) and residual ($\sigma_w^2$) variances
Generalized Estimating Equations (GEE)

Steps in fitting a GEE model:

1. Choose an appropriate correlation structure for the clustered data
2. Fit a standard regression model
3. Specify a working correlation matrix
4. Estimate the residuals from regression model
5. Use the empirical distribution of residuals to estimate the correlation parameters
6. Refit model using a modified algorithm incorporating correlation matrix estimated in Step 5
7. Re-estimate the residuals
8. Repeat steps 4-7 until the estimate is converged
# Working Correlation Structure

<table>
<thead>
<tr>
<th></th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_1$</td>
<td>$[- \ 0 \ 0]$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_2$</td>
<td>$[0 \ - \ 0]$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_3$</td>
<td>$[0 \ 0 \ -]$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Independent**

\[
\begin{bmatrix}
- \rho & \rho \\
\rho & - \rho \\
\rho & \rho & - \\
\end{bmatrix}
\]

**Autoregressive**

\[
\begin{bmatrix}
- \rho & \rho^2 & \rho^3 \\
\rho & - \rho & \rho^2 \\
\rho^2 & \rho & - \rho \\
\rho^3 & \rho^2 & \rho & - \\
\end{bmatrix}
\]

**Exchangeable**

\[
\begin{bmatrix}
- \rho & \rho \rho \\
\rho & - \rho \\
\rho & \rho & - \\
\end{bmatrix}
\]

**Unstructured**

\[
\begin{bmatrix}
- \rho_1 & \rho_2 & \rho_3 \\
\rho_1 & - \rho_5 & \rho_4 \\
\rho_2 & \rho_5 & - \rho_6 \\
\rho_3 & \rho_4 & \rho_6 & - \\
\end{bmatrix}
\]
GEE Models

Advantages
• Do not require a Normal distribution
• Provides robust standard error even if correlation structure is incorrectly specified
• Allow for both cluster and individual level covariates

Disadvantages
• Treat cluster effect as a nuisance parameter (only if interested)
• Do not extend easily to more than two levels
• Standard errors can be unreliable with small number of clusters
Subject-Specific vs Population-Averaged

• **Subject-specific approach**
  – The random effects are modelled
  – Estimates the change in expected response for an individual cluster
  – Appropriate when prediction is required for the cluster

• **Population-averaged approach**
  – Fixed effects are modelled
  – Estimates the expected response for the overall population of clusters
GEE and Multilevel Models

• GEE fits marginal models
  – The treatment effect is estimated as the population average from which the sample is chosen

• Multilevel models are subject-specific
  – It estimates the treatment effect given a subject’s values of the other covariates

• GEE estimates are
  – More sensitive to missing data
  – Generally smaller for non-Normal data

• Both are more efficient than cluster-level approaches with
  – Large numbers of clusters
  – Both cluster and individual level covariates
  – Hierarchical data structure
Stepped-Wedge Design in CRT

- Stepped-wedge cluster randomised trial (SWCRT) involves sequential roll-out of an intervention to clusters of individuals over a number of time periods.
- The order in which a proportion of clusters switch from control to intervention is determined randomly by sequences.
- All clusters receive the intervention eventually and ideally stay on.
- Individual level data are collected repeatedly at each time period.
The Design Matrix

**SWCRT**

<table>
<thead>
<tr>
<th>Sequence 1</th>
<th>Before</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sequence 2</th>
<th>Before</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sequence 3</th>
<th>Before</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sequence 4</th>
<th>Before</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Standard CRT**

(Sequence doesn’t matter)

<table>
<thead>
<tr>
<th>Sequence 1</th>
<th>Before</th>
<th>T1</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sequence 2</th>
<th>Before</th>
<th>T1</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sequence 3</th>
<th>Before</th>
<th>T1</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sequence 4</th>
<th>Before</th>
<th>T1</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Shaded cells represent intervention periods; Blank cells represent control periods.
Population-Based Interventions

• General population of interest
  – Multi-centre trial across a number of cities, provinces, countries, etc.
  – Logistically impossible to implement the intervention simultaneously.
  – Practical and financial constraints

• Policy and ethical considerations
  – Intervention is likely to do more good than harm, but require population-based evidence
  – Unethical to withhold or withdraw intervention from participants
  – Roll out the intervention during and after the study
The Gambia Hepatitis Intervention Study*

• A large-scale vaccination project to introduce the national hepatitis B (HBV) vaccination to young infants progressively over a 4-year period, initial in July 1986.

• Rationale for the study summarised in the report of WHO meeting on the prevention of liver cancer:

  "... it was recognized that it would be undesirable to consider immunization on a widespread scale without, at the same time, organizing studies that were designed to establish whether or not the vaccine was producing a reduction in the incidence of the carrier state and subsequently cancer ... The best method of ensuring that a protective effect of immunization could be measured would be to use randomized controlled trials ... Such a randomized trial would present ethical problems if sufficient vaccine were available to treat all newborn children. At present, and probably for the next several years, however, this is not the position as the vaccine is expensive and supplies are limited."

* The Gambia Hepatitis study group, Cancer Research, 47:5782-87, 1987
The Gambia Hepatitis Intervention Study

Points to consider in the design of an appropriate RCT:

– Expense of the vaccine and its limited availability prohibiting immediate universal HBV vaccination
– Desirability of having comparison groups available from same time period
– Severe logistic and ethical difficulties with randomisation at individual level (team/field conditions, multiple doses p.p., infants at the same clinic treated differently, etc.)
– At the end of the study, the HBV vaccine should be widely available with a nationwide delivery system in place

These considerations lead to a “Stepped Wedge” design!
The Gambia Hepatitis Intervention Study

- 17 define country areas
- 17 teams to initiate intervention in a random order
- 3 months per step
- 4 years in total

Fig. 2. Phased introduction of hepatitis B vaccination in The Gambia.
Differences in SW Design

• Intervention is delivered in a stepped-wedge fashion where a rigorous evaluation of the effectiveness of intervention is desirable
  – Clusters act as their own control, with repeated individual data collected from both control and intervention “wedge”
  – Evaluation of the intervention effect is conducted by comparing the data between the control and intervention wedges

• All participants will receive the intervention in a random order, during and potentially after the trial period
  – Different from parallel and cross-over designs
Different Study Designs

<table>
<thead>
<tr>
<th>Parallel</th>
<th>Crossover</th>
<th>Stepped Wedge</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time</strong></td>
<td><strong>Time</strong></td>
<td><strong>Time</strong></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Cluster</td>
<td>Cluster</td>
<td>Cluster</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

1 = intervention phase; 0 = control phase
Sample size in SWCRT

• More complicated and different from standard CRT
  ➢ Basic parameters in CRT:
    – Study power (1-β) and significance level (α)
    – Size of expected treatment effect (ε)
    – ICC, or similarly the Coefficient of Variation (CV)
      • Ratio of between-cluster standard deviation over the mean prevalence
    – Number of clusters
    – Number of individuals per cluster
  ➢ Other parameters:
    – Number of sequences (steps) and time periods (T) in the wedge
    – Number of control (0) and intervention (1) phases in the design
Design of SWCRT

• Randomisation
  – One or more clusters will be randomly allocated to each sequence of treatment
    • E.g. 40 hospitals (clusters) with 4 groups of 10s were randomly allocated to 4 sequences (steps)

• Data collection
  – All participants in each of the clusters will ideally provide individual level data repeatedly at the end of each time period
    • E.g. clinical data on all patients in 40 hospitals will be collected at scheduled visits when the intervention rolls out sequentially
Analysis of SWCRT

• Random effects are commonly used to model the correlation between individuals within the same cluster

• Statistical modelling appropriate for SWCRT design needs to take into account:
  – Cluster effect (individuals more correlated within a cluster)
  – Repeated measures on the same individual
  – Secular (time) trend

• The following analysis approaches can be used:
  – Multilevel models
    • Linear Mixed models (LMM)
    • Generalized Linear Mixed models (GLMM)
  – Generalised estimating equations (GEE)
The Statistical Model

For a design with $I$ clusters, $T$ time points, and $N$ individuals sampled per cluster per time interval.

Let $Y_{ijk}$ be the outcome of individual $k$ at time $j$ from cluster $i$, and $Y_{ij}$ be the mean of cluster $i$ at time $j$.

Define $\mu_{ij} = \mu + \alpha_i + \beta_j + X_{ij}\theta$, where

- $\alpha_i$ is a random effect for cluster $i$, with $\alpha_i \sim N(0, \tau^2)$
- $\beta_j$ is a fixed effect at time $j$
- $X_{ij}$ is an indicator of treatment phase (1 vs 0) in cluster $i$ at time $j$
- $\theta$ is the treatment effect
The Statistical Model

• A model for individual level outcome is:

\[ Y_{ijk} = \mu_{ij} + e_{ijk} \quad \text{where } e_{ijk} \sim \text{iid } N(0, \sigma_e^2) \]
\[ \text{Var}(Y_{ijk}) = \tau^2 + \sigma_e^2 \quad \rho = \tau^2/(\tau^2 + \sigma_e^2) \]

• A model for the cluster means is:

\[ Y_{ij} = \mu_{ij} + e_{ij} \quad \text{where } e_{ij} = \sum_k e_{ijk}/N \sim \text{iid } N(0, \sigma^2) \]
\[ \text{Var}(Y_{ij}) = \tau^2 + \sigma^2 = \frac{\tau^2 + \sigma_e^2}{N} [1 + (N - 1)\rho] \]
\[ \sigma^2 = \sigma_e^2/N. \]

Design and analysis of stepped wedge cluster randomized trials, Hussey and Hughes (2007)
Power Calculation

• Provide statistical justification of the study power for a selected sample size
• We want to test the following hypothesis:
  
  \[ H_0: \theta = 0 \] versus \[ H_A: \theta = \theta_A \]

• The approximate power for a two-sided test at \( \alpha \) level of significance is given as:

\[
\text{power} = \Phi \left( \frac{\theta_A}{\sqrt{\text{Var}(\hat{\theta})}} - Z_{1-\alpha/2} \right)
\]

- where \( \Phi \) is cumulative standard Normal distribution function
Power Calculation

• Assuming equal N per cluster per time interval,

\[ \text{Var}(\hat{\theta}) = \frac{I \sigma^2 (\sigma^2 + T \tau^2)}{(IU - W) \sigma^2 + (U^2 + ITU - TW - IV) \tau^2} \]

where \( U = \sum_{ij} X_{ij} \), \( W = \sum_j (\sum_i X_{ij})^2 \), and \( V = \sum_i (\sum_j X_{ij})^2 \)

• For a binary outcome where the cluster level response is a proportion, it is reasonable to assume that \( \sigma_e^2 = \mu*(1-\mu) \)
A Case Study – BISkIT Trial

- A SWCRT to determine whether provision of a free school breakfast improves attendance, academic performance, nutrition and food security in children attending primary schools in deprived areas in New Zealand
  - 14 schools in low socioeconomic resource areas were randomised
  - 424 Year 1-8 school students participated in the study
  - Primary outcome: Children’s school attendance, measured as
    - Achievement of a school attendance rate of 95% or higher in a term
    - School attendance rate (%)
  - Secondary outcomes: academic achievement, self-reported grades, sense of belonging at school, short-term hunger, food security
Recruit
Schools

Randomise schools

Recruit students

Four sequences

Term 1
Demographics & Assess

Intervention  Control  Control  Control

Term 2
Assess

Intervention  Intervention  Control  Control

Term 3
Assess

Intervention  Intervention  Intervention  Control

Term 4
Assess

Intervention  Intervention  Intervention  Intervention

Process evaluation
Study Design and Analysis

• The target sample size was 16 schools (4 schools per sequence) with on average 25 students per school, a total of 400 participants followed over 4 terms in a school year.

• Assuming an intra-cluster coefficient (ICC) of 0.05, this sample size would provide at least 85% power at 5% level of significance to detect a 10% absolute increase in the proportion of children with an attendance rate of 95% or higher.

• Generalized Linear Mixed Model (GLMM) was used to evaluate the main treatment effect between intervention and control phases, controlling for individual covariates and the school term.
SAS Example Code

PROC GLIMMIX;
   CLASS Cluster ID Treatment Time;
   MODEL Y1 = X1 Treatment Time / SOLUTION DDFM=satterth
       ODDSRATIO DIST=binary LINK=logit;
   RANDOM Cluster;
   RANDOM _residual_ / SUBJECT=ID TYPE=CS;
RUN;

PROC MIXED;
   CLASS Cluster ID Treatment Time;
   MODEL Y2 = X2 Treatment Time / SOLUTION DDFM=satterth;
   RANDOM Cluster;
   REPEATED Time / SUBJECT=ID TYPE=UN;
RUN;
Systematic Review of SWCRT (1)

BMC Medical Research Methodology

Research article

The stepped wedge trial design: a systematic review
Celia A Brown* and Richard J Lilford

Address: Department of Public Health and Epidemiology, The University of Birmingham, Birmingham, UK
Email: Celia A Brown* - c.a.brown@bham.ac.uk; Richard J Lilford - r.j.lilford@bham.ac.uk
* Corresponding author

Published: 08 November 2006
BMC Medical Research Methodology 2006, 6:54 doi:10.1186/1471-2288-6-54
Received: 10 August 2006
Accepted: 08 November 2006
**Methods:** Comprehensive literature review of studies or protocols using a stepped wedge design. Data were extracted from the studies in three categories for subsequent consideration: study information (epidemiology, intervention, number of participants), reasons for using a stepped wedge design and methods of data analysis.

**Results:** The 12 studies included in this review describe evaluations of a wide range of interventions, across different diseases in different settings. However the stepped wedge design appears to have found a niche for evaluating interventions in developing countries, specifically those concerned with HIV. There were few consistent motivations for employing a stepped wedge design or methods of data analysis across studies. The methodological descriptions of stepped wedge studies, including methods of randomisation, sample size calculations and methods of analysis, are not always complete.

**Conclusion:** While the stepped wedge design offers a number of opportunities for use in future evaluations, a more consistent approach to reporting and data analysis is required.
Systematic review of stepped wedge cluster randomized trials shows that design is particularly used to evaluate interventions during routine implementation

Noreen D. Mdege\textsuperscript{a,\*}, Mei-See Man\textsuperscript{a}, Celia A. Taylor (nee Brown)\textsuperscript{b}, David J. Torgerson\textsuperscript{a}

\textsuperscript{a}York Trials Unit, Department of Health Sciences, University of York, Heslington, York YO10 5DD, UK
\textsuperscript{b}College of Medical and Dental Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

Accepted 3 December 2010
Mdege et al. (2011)

Compared with the first SR,

– Expanded search to non-health care trials
– Focus on randomised studies and cluster allocations
– Aim to describe the application of SWCRT design in evaluating the effectiveness of interventions

• Areas of research involved
• Motivation for using the design
• General characteristics of the design
• Method of data analysis
• Quality of reporting
Note: Heading and subheadings adapted from the CONSORT statement for reporting cluster randomised trials.

<table>
<thead>
<tr>
<th>Quality of reporting</th>
<th>Methodological descriptor</th>
<th>Yes</th>
<th>No/unclear</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Heading/subheading</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Title and abstract</td>
<td>Phrase “step or stepped wedge” used</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Phrase “randomized” used</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>Introduction</td>
<td>Rationale for stepped wedge design</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Rationale for clustering</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td><strong>Methods</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Design</td>
<td>Diagram of design</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>Intervention</td>
<td>Adequately described</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td>Sample size</td>
<td>Sample size calculation reported</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Use of intracluster correlation coefficient</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>Randomization</td>
<td>Sequence generation stated</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Allocation concealment</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Restrictions used</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Blinding</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Use of intention-to-treat principle</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td><strong>Results</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participant flow</td>
<td>Flow diagram</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Number of clusters analyzed</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Number of individuals analyzed</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Main results</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Baseline data reported</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Summary of primary outcome</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Effect size stated</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Precision estimation stated</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>Adverse effects</td>
<td>Side effects reported</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>Discussion</td>
<td>Generalizability</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Limitations</td>
<td>14</td>
<td>1</td>
</tr>
</tbody>
</table>
An MSc Student Project

293 records identified through main database-search (Feb. 2011)

285 records after excluding studies previously identified by Brown and Lilford

46 full-text articles assessed for eligibility

15 studies included in systematic review

9 studies included after excluding the studies identified by Møege et al.

14 new studies identified and included in this dissertation

28 records identified through supplement database-search (Jun. 2011)

14 records after excluding studies previously identified by Brown and Lilford and by Møege et al.

9 full-text articles assessed for eligibility

5 studies included in systematic review

4 full-text articles excluded
Systematic Review n=2
Methodology n=2

Reported by previous systematic reviews n=14

5 records excluded:
Non-English n=3
Irrelevant n=2

239 records excluded:
Duplication n=22
Non-English n=13
Irrelevant n=203
No Full-text n=1

31 full-text articles excluded:
Not Trial or Protocol n=17
Not Stepped-Wedge n=13
Referring to Same Trial n=1

8 new protocols
6 new completed trials

Lennex Yu (2011)
14 New Studies Identified

<table>
<thead>
<tr>
<th>Lead Author</th>
<th>Publication Date</th>
<th>Subject Area</th>
<th>Disease/Condition</th>
<th>Intervention</th>
<th>Country</th>
<th>Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brown</td>
<td>2006</td>
<td>Social Policy</td>
<td>Youth suicide</td>
<td>School-based gatekeeper training program for suicide prevention</td>
<td>USA</td>
<td>Middle and High Schools</td>
</tr>
<tr>
<td>Andersson</td>
<td>2010</td>
<td>Social Policy</td>
<td>Domestic violence in aboriginal communities</td>
<td>Community-led prevention</td>
<td>Canada</td>
<td>Aboriginal Communities/Shelters</td>
</tr>
<tr>
<td>Bacchieri</td>
<td>2010</td>
<td>Social Policy</td>
<td>Traffic accidents of cyclists</td>
<td>Educational program</td>
<td>Brazil</td>
<td>Census sectors</td>
</tr>
<tr>
<td>Bhutta</td>
<td>2010</td>
<td>Health</td>
<td>Childhood malnutrition</td>
<td>Ready-to-use therapeutic food</td>
<td>Pakistan</td>
<td>Community based</td>
</tr>
<tr>
<td>Kao</td>
<td>2010</td>
<td>Health</td>
<td>Non-compliance with surgical site infection prophylaxis guidelines</td>
<td>A package of quality improvement interventions</td>
<td>USA</td>
<td>Hospitals</td>
</tr>
<tr>
<td>Killam</td>
<td>2010</td>
<td>Health</td>
<td>Viral Infection in HIV-infected pregnant women</td>
<td>Antiretroviral therapy integrated in antenatal care</td>
<td>Zambia</td>
<td>Hospitals</td>
</tr>
<tr>
<td>Stern</td>
<td>2010</td>
<td>Health</td>
<td>Pressure ulcers</td>
<td>Multidisciplinary wound care teams</td>
<td>Canada</td>
<td>Long term care facilities</td>
</tr>
<tr>
<td>van den Broek</td>
<td>2010</td>
<td>Health</td>
<td>Chlamydia infection</td>
<td>Screening implementation</td>
<td>Netherland</td>
<td>Neighbourhoods</td>
</tr>
<tr>
<td>Gerritsen</td>
<td>2011</td>
<td>Health</td>
<td>Mental depression in nursing home residents of somatic and dementia special care units</td>
<td>Multidisciplinary, evidence-based care program</td>
<td>Netherland</td>
<td>Somatic and dementia special care units</td>
</tr>
<tr>
<td>Lilly</td>
<td>2011</td>
<td>Health</td>
<td>Various critical conditions</td>
<td>Tele-Intensive care unit</td>
<td>USA</td>
<td>ICUs</td>
</tr>
<tr>
<td>Mouchoux</td>
<td>2011</td>
<td>Health</td>
<td>Postoperative delirium in the elderly</td>
<td>Multidisciplinary prevention program</td>
<td>France</td>
<td>Surgical wards</td>
</tr>
<tr>
<td>Ni Mburuchu</td>
<td>2011</td>
<td>Education</td>
<td>Lack of breakfast in children living in deprived areas</td>
<td>Free daily before-school breakfast program</td>
<td>New Zealand</td>
<td>Primary Schools</td>
</tr>
<tr>
<td>Turner</td>
<td>2011</td>
<td>Health</td>
<td>Psychosocial stress due to cancer</td>
<td>Personalized psychosocial intervention integrated into routine care</td>
<td>Australia</td>
<td>Clinical sites</td>
</tr>
<tr>
<td>Zwischen</td>
<td>2011</td>
<td>Health</td>
<td>Behavioural problems in nursing home residents with dementia</td>
<td>Multidisciplinary prevention program</td>
<td>Netherlands</td>
<td>Dementia special care units</td>
</tr>
</tbody>
</table>

*The protocols are starred in the first column
Overall Summary

• An increasing number of SWCRT have been reported in recent years
• More studies are published from the developed than developing countries
• The methodological reporting is still sub-optimal in some studies and may lack appropriate information on sample size calculation, randomization method, data collection and analysis
Published Studies by Year and Country

Number of Publications

Year

Developing Countries
Developed Countries

80s 90s 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
A World Map of Published Studies by Trial Location