Clustering issues in trials: Cluster randomised trials, therapist effects and group treatments

Sabine Landau & Trudie Chalder
Biostatistics and Psychological Medicine, IoP

Presentation at King’s Trials Partnership 12th Dec 2012

Outline
1. How does clustering arise?
2. Implications of clustering for the design of clinical trials
3. Dealing with clustering in the statistical analysis of trials
4. External validity of trials subject to clustering
5. An example therapist trial

1. What is clustering?
- In a trials clustering typically refers to the situation where outcomes for different participants are not (statistically) independent;
  - i.e. by knowing the cluster membership and outcome of a participant one can to some extent predict the outcome of other participants from the same cluster
  - i.e. outcomes of participants from the same cluster are similar/positively correlated
- Statisticians tend to think of clusters as the levels/units at which variability arises.

Clusters and associated clustering

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Clustering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Centres in a multicentre trials</td>
<td>Pre and post randomisation patient outcome due to catchment area characteristics</td>
</tr>
<tr>
<td>Cluster unit of a cluster randomised trial</td>
<td>Pre and post randomisation patient outcome due to catchment area</td>
</tr>
<tr>
<td>Post treatment patient outcome if treatment implementation can vary between clusters (e.g. GP training)</td>
<td></td>
</tr>
<tr>
<td>Therapist (or GP, nurse, facilitator..)</td>
<td>Post treatment patient outcome due to therapist variability in intervention implementation</td>
</tr>
<tr>
<td>Group treatments/training programmes</td>
<td>Post treatment patient outcome due to group membership characteristics</td>
</tr>
</tbody>
</table>

Example: Cluster randomised trial (CRT)
- Interventions are randomised to groups of patients (clusters) rather than individually.
  - Done when the intervention is applied at the cluster level
    - e.g. new clinic procedures (cluster=clinic)
    - e.g. professional behaviour change (cluster=health professional)
  - Also done to avoid treatment contamination.
    - e.g. several therapists in clinic, some trained in new treatment
- Example: Priebe et al. (2007), BJPsych 191: 420-426:
  - Aim: To test a computer intervention structuring patient-clinician dialogue (DIALOG)
  - 134 key workers randomised to DIALOG or treatment as usual (TAU)
  - 507 people with schizophrenia participated in trial
  - Primary outcome: participant quality of life

Example: Therapist-led intervention
- Current trial proposal
- ACTIB (Assessing Cognitive behavioural Therapy in Irritable Bowel): A randomized controlled trial of clinical and cost effectiveness of therapist delivered cognitive behavioural therapy
- Participants population: People with refractory Irritable Bowel Syndrome (IBS)
- Three arms (two superiority hypotheses):
  - T-CBT (therapist-led delivery) plus treatment as usual (TAU)
  - LIBT (low intensity web-based programme, therapist involvement) plus TAU
  - TAU
- Two primary outcomes:
  - IBS symptom severity score (IBS-SSS, continuous) at 12 months
  - Subjective Global Relief of Symptoms (SGA, binary) at 12 months
Example: Group treatment

- The Incredible Years (IY) parenting training is an intervention for parents of children with conduct disorder.
- Parents were randomised to
  - SPOKES
  - or minimal support.
- The primary clinical outcome is child conduct.
- The parenting training intervention was a group treatment led by different therapists.
- Two levels of treatment induced clustering: post-treatment outcome variability due to different trainers as well as due to group memberships
- Larger group effects for parenting than for distal child outcomes.

2. Trial Design

- Questions that need addressing at the trial design stage:
  - How many patients to recruit?
  - How many clusters to include?
- Cluster units in cluster randomised trial
- Therapists for therapist-led interventions
- Groups for group treatment
- How to allocate participants to clusters?

Useful distinction

- Crossed clustering = cluster levels occur in both treatment arms
  - e.g. centre effects in multi-centre trials
  - e.g. clustering due to general therapist characteristics (not treatment-specific)
- Nested clustering = each cluster level only occurs in one of the treatment arms
  - e.g. treatment-specific therapist effects
  - e.g. group treatment effects
- (Note the distinction refers to the true clustering process not the trial design.)

Impact of clustering

- Clustering means that the data provide less information
  - loss of independent information units
  - effective sample size reduced
- The uncertainty of statistical inferences is increased compared to non-clustered data
  - wider confidence intervals
  - larger p-values
- (Some designs can allow for conditioning on clusters in the analysis and removal of the extra variance due to crossed clustering.)

Sample size requirement

- A sample size calculation for independent data:
  - specifies the significance level that will be used
  - sets the power that is required
  - identifies the test of the treatment effect that will be used
  - specifies the minimal clinical significant effect size
  - and determines the sample size n as function of these inputs
- The sample size can be inflates by the so-called design factor D to compensate for the loss of information due to clustering.
  - The new sample size is then \( n^* = nD > n \)

Design factor

- The design factor (>1) is the ratio of the number required under clustering to the number required assuming no clustering
  \[ D = 1 + \frac{k}{I(k-1)} \]
  where
  - I is the intra-cluster correlation coefficient (ICC)
  - k is the number of cluster members
- Here clusters are understood as levels within which participant outcomes are correlated.
- Thus the required sample size inflation increases with the strength of the clustering (the ICC) as well as with the size of the clusters (decreases with increasing number of clusters to give fixed n)
Intra-cluster correlation (ICC)

- The ICC is defined by the variance proportion:
  \[ 0 < \frac{\tau^2}{\tau^2 + \sigma^2} < 1 \]
  where
  - \( \tau^2 \) between-cluster variance in outcome
  - \( \sigma^2 \) within-cluster (patient) variance in outcome

- ICC=0 implies no variation between clusters
- ICC=1 implies all cluster members have the same outcome.
- Typically, the ICC takes values <0.05 in large samples.

How to get an ICC estimate?

- First of all include cluster memberships in measurement protocols! (Therapist IDs, group IDs etc., clinic etc.)
- Variances and associated ICC can be estimated using post treatment outcomes from comparable previous studies (variance component models).
- Some literature advice on likely ranges of ICCs:
  - Typical ICC=0.02 for therapist effects (Baldwin et al. 2011 Cogn Behav Ther 40: 15-33)
  - Range of ICCs in CRTs from 0.01-0.05 (Smeeth et al. 2002 Controlled Clin Trials 23: 409-421)
  - Median ICC=0.04 in review of CRTs (Eldrige et al. 2004 Clin Trials 1: 80-90)
  - Would expect ICCs for process variables targeted by treatment to be larger than those for distal clinical outcomes.

Example: ACTIB trial

For practical reasons the maximum number of therapists that could be recruited to the trial is 10 with a projected workload of 17 patients each.

We assumed therapist clustering of a typical strength ICC=0.02.

This gives a sample size inflation for clustering of D=1.32

Other sample size settings: power 90%, significance level 2.5% to adjust for two primary outcomes, test statistic of independent samples t-test

<table>
<thead>
<tr>
<th>Post difference</th>
<th>Effect size</th>
<th>Initial number per arm</th>
<th>Inflation for clustering per arm ( (\times 1.32) )</th>
<th>Inflation for 20% attrition per arm ( (\times 1.08) )</th>
<th>Deflation for using baseline values ( (\times 0.86) )</th>
<th>Total needed for three arms</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.33</td>
<td>201</td>
<td>267</td>
<td>261</td>
<td>232</td>
<td>300</td>
</tr>
<tr>
<td>20</td>
<td>0.4</td>
<td>161</td>
<td>213</td>
<td>206</td>
<td>223</td>
<td>269</td>
</tr>
<tr>
<td>15</td>
<td>0.53</td>
<td>142</td>
<td>197</td>
<td>147</td>
<td>155</td>
<td>199</td>
</tr>
<tr>
<td>10</td>
<td>0.59</td>
<td>122</td>
<td>161</td>
<td>130</td>
<td>135</td>
<td>166</td>
</tr>
<tr>
<td>5</td>
<td>0.72</td>
<td>93</td>
<td>120</td>
<td>120</td>
<td>101</td>
<td>124</td>
</tr>
</tbody>
</table>

Example: ACTIB trial cont’d

- What would have happened if we could have recruited twice as many therapists (20 therapists each seeing \( k \times 8 \) participants per therapy)?
  \( D = 1 + 0.02 (k-1) = 1.14 \)

- Using the same parameters as on the last slide total sample size \( n=427 \)
  - i.e. saving of 14% patient recruitment (68 participants)
  - (Though not necessarily a saving in cost as part-time employment of therapist may cost more.)

Better sample size calculations

- The design effect approach is somewhat simplistic in that it assumes that there are clusters in all arms and that the ICCs are constant across arms.
  - E.g. in the ACTIB trial therapists are not involved in TAU delivery and their effects in LBT may be less than in T-CBT
- But to deal with this more appropriately we would have to work out power of estimators based on complex statistical models
  - this gets more complicated...
- See Chris Robert’s ciampsi STATA command for sample size calculations when ICCs depend on trial arm, e.g. clustering is only present in one arm
  [http://www.medicine.manchester.ac.uk/healthmethodology/research/biostatistics/data/ciampsi/](http://www.medicine.manchester.ac.uk/healthmethodology/research/biostatistics/data/ciampsi/)

Allocation of participants to clusters

- In CRTs the cluster memberships (e.g. centre memberships) are known before the trial.
- For group treatments and therapist-led interventions there is a choice as to how participants are allocated to clusters (groups, therapists).
  - If the allocation is driven by patient characteristics (e.g. “best” therapist sees “worst” patients) then therapist effects cannot be separated from patient effects.
  - Not helpful if one wants to estimate therapist effects, e.g. for future sample size calculations.
  - Ideally – to ensure internal validity – the allocation of participants to clusters should be at random.
  - Then there can’t be any confounding of cluster effects.
Design recommendations

- Record the cluster information.
- Avoid creating clusters (or be prepared to make up for it in recruitment and costs…)
  - Do not cluster randomise merely for convenience.
  - Consider whether contamination could be overcome in other ways?
- Where this is an option (therapist trials, group treatments) increase number of clusters/decrease cluster sizes for fixed sample size.

Design recommendations cont’d

- Inflate sample size for clustering effects if want to generalise to population from which the clusters are taken.
  - Always applies in CRTs.
  - Also applies in phase III (HTA funded) therapist or group treatment trials.
  - Arguable that in phase II trials (EME funded) the limited objective is to establish efficacy for a specific therapists/group set up.
- Where feasible randomise participants to clusters that are related to the delivery of treatment (therapists, groups).

3. Statistical analysis

- Should we always take potential clustering into account?
- How to deal with clustering in the statistical analysis?
  - Multiple sources of clustering
  - Is the correlation structure always simple?
- How to deal with cluster membership changes (multiple memberships)?

Models for clusters

- Cluster effects should always be included in the statistical model and tested where the design allows this.
  - See extension of Consort Guidelines for non-Pharmacological Treatment Trials.
- Things to consider when trying to specify an adequate statistical analysis model:
  - Do we wish to generalise treatment effects to the cluster populations (e.g. therapists delivering CBT in NHS)?
  - Is the cluster structure hierarchical (e.g. classes in schools)?
  - Nested or crossed clustering (in terms of treatments)?

Analysis approaches for clustering

- Simple approach is to calculate a summary statistic for each cluster.
  - A mean for a continuous outcome or a proportion for a dichotomous outcome.
  - But the approach cannot allow for individual covariates.
- Thus we need to model participant level data.
- Two types of methods:
  - Accounting for the impact of clustering:
    - Use standard error estimates that are robust against clustering
    - Generalised estimating equations (GEE)
  - Explicitly modelling the cluster effects: Random effects models

GEE

- GEE stipulates a “working correlation structure” at the level of the largest clusters and then deals with departures from this structure by using robust standard errors.
- Approach is easy to implement (e.g. in STATA)
  - But not as powerful as fitting the full data generation model.
Random effects models

- Cluster effects should be modelled as random effects rather than fixed effects if the clusters (e.g. therapists) represent a sample from a wider population to which our treatment effect estimate should apply.
  - Though in practice this is only possible if there is a sufficient number of clusters; say at least 8 clusters
- Random effects models include random variables that vary at the level of the cluster(s).
  - Can model hierarchies.
  - Crossed clustering implies random effects at the cluster level as well as at the cluster level within each arm.
  - But model only identified under certain designs (crossed design)
  - Nested clustering implies treatment arm-specific random effects at the cluster level.
Implications of trial type

- The type of nested trial design has implications for whether variances of cluster-level random effects vary between trial arms.

<table>
<thead>
<tr>
<th>Trial type</th>
<th>CRT</th>
<th>Nested therapist trial</th>
<th>Group therapy trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between cluster variation (ICC)</td>
<td>Same in both arms</td>
<td>May differ between arms</td>
<td>May differ between arms</td>
</tr>
<tr>
<td>Cluster size</td>
<td>Similar due to randomisation</td>
<td>May differ between arms</td>
<td>May differ between arms</td>
</tr>
<tr>
<td>Cluster membership</td>
<td>Defined at randomisation</td>
<td>May be poorly defined</td>
<td>Should be well-defined for closed group treatments</td>
</tr>
</tbody>
</table>

Further analysis issues

- Does it matter if the cluster size varies between arms?
  - Not sure; see work by Chris Roberts, University of Manchester
- How to deal with cluster membership changes?
  - Difficult.
  - A possibility might be to fit multiple membership models.

External validity in therapist trials

- Where interventions cannot be randomised to clusters (therapist-led experimental treatment) selection of therapists has implications for external validity.
- Need to consider how representative therapists in trial are of non-trial therapists.
  - Inclusion and exclusion criteria for therapists helps to define target populations.
  - Broad therapist eligibility criteria (e.g. recruited directly from clinical practice) makes trial results more generalisable but this could increase the therapist-cluster correlation.

4. External validity

- Are the results generalisable with respect to the participants?
- Are the results generalisable with respect to the cluster units?
  - E.g. clinic clusters in CRTs
  - E.g. therapists delivering treatments

5. One example trial using therapists (PACE)

- Alliance rates high
- No difference between groups in proportions who received adequate treatment (85% or more)
- Expectations high for all therapies
- Most participants were satisfied with therapy
- High rates of fidelity
Therapy quality in PACE

<table>
<thead>
<tr>
<th></th>
<th>APT</th>
<th>CBT</th>
<th>GET</th>
</tr>
</thead>
<tbody>
<tr>
<td>N sessions</td>
<td>13</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>“Confident” before</td>
<td>72 %</td>
<td>57 %</td>
<td>70 %</td>
</tr>
<tr>
<td>“Satisfied” after</td>
<td>85 %</td>
<td>82 %</td>
<td>88 %</td>
</tr>
<tr>
<td>Alliance</td>
<td>6.5</td>
<td>6.5</td>
<td>6.5</td>
</tr>
<tr>
<td>Adherence</td>
<td>6</td>
<td>6</td>
<td>6.5</td>
</tr>
</tbody>
</table>

Integrity of therapy

- All therapies: based on theory and interventions evolved from theory
- Detailed manuals written
- Therapists expected to adhere to treatment model and interventions
- Therapists received close supervision
- Sessions taped and used in supervision and for assessing fidelity

PACE therapist effect

Intraclass correlation coefficients ranged from -0.02-0.11 for fatigue and -0.01 to 0.03 for physical function

(White et al. 2011; Lancet)

Therapist effects in routine clinical practice (CBT for CFS)

- Well defined sample of patients and therapists from an outpatient service which specialises in providing cognitive behaviour therapy (CBT) for patients with chronic fatigue syndrome (CFS).
- Therapy was provided in a highly specialised clinical setting by qualified CBT therapists with at least two years experience with this client group.
- Three hundred and seventy four patients with CFS and 12 Cognitive Behavioural Psychotherapists took part.

Analysis and results

- Therapist effects on the primary outcomes of fatigue and disability was investigated with multilevel random effects models.
- Different models were computed and compared.
- Results showed a reduction in fatigue and disability scores after therapy.
- Variance explained by therapists, when demographic covariates were accounted for, was 0% for fatigue and under 2% for disability.
PACE discussion

- A number of important factors may have played a significant role in minimizing therapist effects in our study. These are:
  - specialist setting
  - single centre
  - patients with same primary diagnosis
  - therapists of same orientation and training
  - shared environment and supervision
- Future studies may stress the importance of these factors in the investigation of the therapist effects in psychotherapy (Cella et al. (2011) Psychotherapy Research)

Case example – IBS (Reme et al. 2011)

- RCT comparing mebeverine alone versus mebeverine plus CBT
- Mebeverine plus CBT was more effective than meb alone
- Changes in behaviour and cognitions mediated change in all outcomes with models placing behaviour upstream of cognition having best fit
- Using multilevel modelling NO therapist effect

Therapist effects in psychotherapy

- A random-effects modelling of the National Institute of Mental Health Treatment of Depression Collaborative Research Program data.
- Therapist variability modeled in several different ways, indicated that about 6% of the variance in outcomes was attributable to therapists.
- When therapist effects were appropriately modelled, previously detected differences in efficacy between the two psychotherapy conditions for more severely depressed patients disappeared.

Conclusions

- The impact of clustering can be large, inflating Type I errors – and can't be ignored.
- This may not be obvious to researchers.
- Clustering needs to be addressed as part of the design and analysis of trials.
- Where possible steps should be undertaken at the design stage to reduce the impact of clustering.