Sample size in cluster randomised trials

Sandra Eldridge
Professor of Biostatistics
Director of Pragmatic Clinical Trials Unit
Outline

• Introduction
• Background to trials in health services research/ primary health care
• Accounting for variation in cluster size
• Intra-cluster correlation coefficients
Sample size calculations in cluster randomised trials
- usual method
Sample size for individually randomised trial

To be able to detect a difference $\mu_1 - \mu_2$ at the $\alpha$ significance level with power $1-\beta$ if standard deviation of outcome is $\sigma$, requires a sample size of $N$ in each arm

$$N = \frac{2(z_{\alpha/2} + z_\beta)^2(\sigma^2)}{(\mu_1 - \mu_2)^2}$$
Sample size for cluster randomised trials

\[ N = \frac{2(z_{\alpha/2} + z_{\beta})^2(\sigma^2)(1 + (m-1)\rho)}{(\mu_1 - \mu_2)^2} \]

Cluster size

Intra-cluster correlation coefficient (ICC)

Inflation factor = Design effect
Derivation

\[ \rho = \frac{\sigma_b^2}{(\sigma_b^2 + \sigma_w^2)} \]

\[ m = \text{cluster size (assuming clusters are all the same size)} \]
How many trials take account of clustering?

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selected journals (1)</td>
<td></td>
</tr>
<tr>
<td>Primary prevention (2)</td>
<td></td>
</tr>
<tr>
<td>Decision support systems (3)</td>
<td></td>
</tr>
<tr>
<td>Primary care (4)</td>
<td></td>
</tr>
<tr>
<td>Sub-Saharan Africa (5)</td>
<td></td>
</tr>
<tr>
<td>Selected journals (6)</td>
<td></td>
</tr>
<tr>
<td>Selected journals (7)</td>
<td></td>
</tr>
<tr>
<td>Primary care (8)</td>
<td></td>
</tr>
<tr>
<td>Cancer (9)</td>
<td></td>
</tr>
<tr>
<td>Tropical parasitic diseases...</td>
<td></td>
</tr>
<tr>
<td>Maternal and child health...</td>
<td></td>
</tr>
<tr>
<td>Oral health (12)</td>
<td></td>
</tr>
</tbody>
</table>

Include trials published 1973 to 2002
Include trials published 1998 to 2009
Practical issues for investigators

What do I do if cluster sizes are likely to vary?
Use mean cluster size in place of m?

How do I estimate the intra-cluster correlation coefficient (ICC)?
Trials in health services research
Particularly in primary care
Frequently cluster randomised

Answer questions about the effect of

– Education to health professionals
– Changing organisational structure
– Addition of new staff
Example - ELECTRA

(East London randomised controlled trial for high risk asthma)

To determine whether asthma specialist nurses, using a liaison model of care, reduce unscheduled care in a deprived multiethnic setting

Setting: UK general (family) practices

Intervention:

1. patient review in asthma-liaison nurse led clinic
2. liaison with general practitioners and practice nurses, ongoing clinical support
3. educational outreach, promotion of guidelines for high risk asthma

(Griffiths et al, 2004)
Example - ELECTRA

- Recruitment

- 44 practices (clusters) involved
- Range of cluster sizes = 2 to 28
- Mean cluster size = 7.78
- Coefficient of variation of cluster size (sd/mean) = 0.64

Diagram:
- Start of trial
- 2 years:
  - Attended Accident & Emergency or admitted
- 1 year:
  - Attended Accident and Emergency or general practice
Example - ELECTRA

- Primary outcome = attendance for unscheduled care in trial period
- ICC used in sample size calculation = 0.05
- Actual ICC = -0.0056
- Negative ICC set to zero for analysis (ie assume no clustering)
Four reviews, one methods paper

- Trials in primary care *(Eldridge et al, 2004)*
- Trials in oral health *(Froud et al, in press)*
- Trials in residential facilities for older people *(Diaz-Ordaz, in preparation)*

- Sample size estimation methods - review *(Clare Rutterford, current PhD student)*

- Sample size estimation when cluster sizes vary *(Eldridge et al, 2006)*
Trials in primary care (1997-2000)
Number of clusters analysed (n=87)

ELECTRA: 44
Trials in primary care (1997-2000)

Average size of clusters (n=71)

ELECTRA: 7.78
Trials in primary care (1997-2000)
Unequal sized clusters (n=139)

- 27 (19%) attempted to have equal cluster sizes
- Usually recruiting
  - from registers
  - using incident cases
- Cluster size approx. proportional to total size of cluster
Variation in cluster size
Six trials in UK primary care (*Eldridge et al* 2006)
Coefficients of variation: 0.42, 0.61, 0.62, 0.64, 0.72, 0.75

CV for list size for UK general practices = 0.65
150 ICCs from trials randomising general (family) practices

Fraction of ICCs at this value

ICC value
-.2 0 .2 .4 .6

0
.1
.2
.3

ELECTRA : -0.0056
How many trials report ICCs observed in analyses?

<table>
<thead>
<tr>
<th>Area</th>
<th>Trials Reported</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary Care (1997-2000)</td>
<td>11/152</td>
<td>7%</td>
</tr>
<tr>
<td>Oral health (2005 - 2009)</td>
<td>8/23</td>
<td>34%</td>
</tr>
<tr>
<td>Residential facilities (1992 – 2010)</td>
<td>8/72</td>
<td>11%</td>
</tr>
</tbody>
</table>
## Estimated and observed ICCs

<table>
<thead>
<tr>
<th>Trial</th>
<th>Type of outcome</th>
<th>ICC used in sample size</th>
<th>Observed ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELECTRA (Griffiths BMJ 2004)</td>
<td>Binary</td>
<td>0.05</td>
<td>-0.0056</td>
</tr>
<tr>
<td>TB trial (Griffiths Lancet 2007)</td>
<td>Binary</td>
<td>0.05</td>
<td>-0.0313</td>
</tr>
<tr>
<td>Diabetes manual (Sturt Diab. Med. 2008)</td>
<td>Continuous</td>
<td>0.043</td>
<td>0.0256</td>
</tr>
<tr>
<td>IRIS (in press, Lancet)</td>
<td>Binary</td>
<td>0.03</td>
<td>0.003</td>
</tr>
<tr>
<td>Smoking cessation in schools (Resnicow Am. J. PH 2010)</td>
<td>Binary</td>
<td>0.02</td>
<td>0.118</td>
</tr>
</tbody>
</table>
Summary

• ELECTRA not untypical in terms of
  – Number of clusters
  – Cluster size
  – Unequal cluster sizes
  – Observed ICC not very close to ICC used in sample size calculation
Issues

• Adjusting sample size estimation to account for variable cluster size

• Need information about variability in cluster size

• Need information about ICC
Review of methods for sample size estimation (58 papers)

14 papers (24%) focus on methods for use when cluster sizes vary.
Papers focusing on sample size estimation when cluster sizes vary

• Actual cluster sizes known in advance (5 papers)
• Methods based on coefficient of variation of cluster size (cv) assuming analysis weights by cluster size (4 papers)
• Other methods (5 papers)
Method using cv assuming analysis weighting by cluster size

- Accounting for clustering
  Inflation factor = $1+(m-1)\rho$

- Accounting for variable cluster size
  Inflation factor = $1+(m(1+cv^2)-1)\rho$

- Appropriate for continuous and binary outcomes

No need to use adjustment if $cv<0.23$

Extra term
Other methods

• Assume more efficient analysis e.g. maximum likelihood
• Also result in approximation based on coefficient of variation of cluster size
• Result in smaller sample size required
• Strictly more appropriate because more likely to match analysis BUT....
Design effect by ICC
Mean cluster size = 10, cv = 0.65

If ICC and \( \bar{m} \) small, methods to account for variable cluster size give similar results.
# Use of methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Number of citations</th>
<th>Approximate number of citations by trial investigators (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual cluster sizes</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>Coefficient of variation – weighting by cluster size</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>Other methods</td>
<td>25</td>
<td></td>
</tr>
</tbody>
</table>

- Assuming weighting by cluster size much more common in practice

Not related to a cluster randomised trial
Information about variability in cluster size?

• Available from previous studies?

• Modelling
  – Distribution of whole cluster sizes in population
  – Sampling, recruitment, drop-out and non-response of clusters
    – Sampling, recruitment, drop-out and non-response of individuals

• Minimum and maximum cluster sizes
  – Standard deviation approx. \((\text{max-min})/4\)

• Other particular situations
Summary

• Most cluster randomised trials have variable cluster sizes
• Methods exist to account for variation in cluster size and should be used
• The most commonly used method is conservative for the most popular forms of analysis
• Methods rely on an estimate of the coefficient of variation of cluster size
• This can be approximated most easily using minimum and maximum cluster size values
Need to have estimate of ICC
Obtaining estimates of ICCs

- Guess

- Single estimate from previous study or pilot
  - popular method but......
Width of ICC confidence interval by number of individuals in study and number of clusters, ICC = 0.05
Other methods of obtaining ICCs

• Based on patterns in ICCs
  – Higher ICCs for ‘process’ than for clinical outcomes
    ICC for ‘blood pressure measured’ > ICC for blood pressure
  – Lower ICCs for clusters that are naturally larger
    ICCs for communities < ICCs for general practices < ICCs for households/families
  – For binary outcomes, higher ICCs if nearer to 50%

Adams et al 2004: The precise value of an ICC for a given outcome ‘can rarely be estimated in advance’
‘Studies should be designed with reference to the overall distribution of ICCs and with attention to features that increase efficiency’
Other methods of obtaining ICCs

• Based on simple combination of several estimates
• Modelling several estimates to produce distribution of ICC values (Turner et al 2005)
Example – IRIS

(Identification and referral to improve safety for women)

To test the effectiveness of a training and support programme for general practice teams targeting identification of women experiencing domestic violence and referral to specialist domestic violence advocates

Setting: UK general practices

Intervention:

1. Practice-based training sessions
2. Electronic prompts to ask about abuse, simple referral pathway to a named advocate in a specialist domestic violence agency, identification of an IRIS practice champion, feedback on referrals and reinforcement over the course of a year
### Example - IRIS

**Primary outcome = proportion of women identified in practice consultation**

<table>
<thead>
<tr>
<th>Sample size estimation assumptions</th>
<th>Method of analysis</th>
<th>ICC</th>
<th>Mean</th>
<th>cv</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis with cluster size weights</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actual</td>
<td>More efficient analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Key determinant: Trial overpowered. The only thing we might have been able to predict in advance was that we would use more efficient analyses. This would have reduced our estimate of number of clusters needed from 24 to 20........
Example – diabetes manual

<table>
<thead>
<tr>
<th>Sample size estimation assumptions</th>
<th>Method of analysis</th>
<th>ICC</th>
<th>Mean cluster size</th>
<th>cv</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis using cluster size weights</td>
<td></td>
<td>0.043</td>
<td>5</td>
<td>0.5</td>
</tr>
<tr>
<td>Actual</td>
<td>More efficient analysis</td>
<td>0.0256</td>
<td>4.2</td>
<td>0.65</td>
</tr>
</tbody>
</table>

Trial slightly underpowered because of higher cv and lower numbers of individuals recruited per practice.
Conclusions

• Sample size estimation challenging for these trials
• Simple methods now exist to account for variable cluster size; these should be used
• Useful to acquire as much knowledge as possible in particular area
• Information on patterns in ICCs is useful
• Need more information about cvs of cluster size