Classical Multilevel Models for the analysis of clustered randomised trial data - Gaussian outcome

Ulrich Mansmann
IBE, LMU Munich
Outline

Advantages/ disadvantages of models relative to summary statistic method

Fitting a simple classical multilevel model using data from cluster randomized trials when outcome is Gaussian/Normal

Interpretation

Methods for parameter estimation and significance tests

Extension of the simple model

Estimation of measures of precision for fixed and random parameters
Advantages / disadvantages of models relative to summary statistic methods

Summary statistic methods:

- Simple to use and interpret
- Particularly suitable for trials with small no. of clusters
- Collapses all measurements into a single summary measure
- Cannot incorporate subject level covariates easily
- Do not extend easily to allow for explanatory variables
- Problems in choosing appropriate weighting for unbalanced clusters
- Maybe inefficient
Advantages / disadvantages of models relative to summary statistic methods

Models:

• Handle multiple covariates easily

• Can include both subject & cluster level covariates simultaneously

• Can model complex correlation structure

• Multilevel Models easily extend to accommodate additional clustering & complex structure

• Computationally intensive, can have convergence problems

• May not be suitable for studies with small number of clusters
Example of a cluster randomised trial:  
Effect of intervention on BMI  
(Kinmonth et al, BMJ, 1998)

**Aim:** to assess effect of additional training of practice nurses & GPs in patient centred care on lifestyle

**Subjects:** General practices randomised to ‘patient centred care’ or not, 18 practices in intervention & 20 in control, 240 patients diagnosed with type 2 diabetes
Cluster sizes range from 1 to 18

**Outcome:** BMI after one year.

Explanatory variables such as gender, age at patient level are available
Nested data structure

I+R → P → O

I+R → C → P → O

Intervention
Randomisation
Cluster
Patient
Observation

Individuals nested in groups
Ordinary regression models

Ordinary regression methods with single level:

\[ Y_i = \alpha + \beta \cdot Group + \varepsilon_i \]

Limitations:

- Ignores clustered data structure
- Violation of data independence assumption
Multilevel Models

Consider a 2-level model for this example: patients nested within practices.

Adopt standard regression tools, relating response to explanatory variables, but account for within subject correlation.

Regression coefficient in a standard regression model is assumed to take the same values for all practices.

In multilevel models correlation among responses for a practice is assumed to arise from a natural heterogeneity in regression coefficients across practices.

Practices are regarded as a random sample from a population of practices. Response measured at lowest level, covariates measured at all levels.
Multilevel Models

In ordinary regression models, regression coefficients are fixed.

Multilevel models have both fixed & random coefficients.

Regression coefficients are allowed to remained fixed or to vary from practice to practice.

Often termed as fixed, random or mixed effects models also known as:

- Random effects models
- Hierarchical models
Fitting a simple classical multilevel model

Initially we consider a simple multilevel model including:

- a single explanatory variable representing intervention
- a simple variance structure - allowing the intercept to vary between practices
- implies all patients within a practice have same correlation
Fitting a simple classical multilevel model

A fixed effects model
A random intercept model
Mixed effects Models

The model can be formulated as follows

\[
BMI_{ij} = a + u_i + b \text{ group}_i + e_{ij} \quad (1)
\]

\(i\) - refers to practice, \(j\) refers to patient

The group variable takes 0 for control and 1 for intervention groups

**Fixed effects:**

- \(a\) (intercept) - mean BMI for all patients in the control group of Practices
- \(b\) (intervention effect) - difference in mean BMI between control & intervention groups

**Random effects:**

- \(u_i\) (cluster effect)
- \(e_{ij}\) (error term)

Intercept is random, often termed as a random intercept model
Underlying assumption about variance/correlation structure

A single random intercept model assumes equal pairwise correlation between all subjects.

e.g. Correlation between subjects within a cluster

```
1   2   3   ..........   m
1   1   \rho   \rho   \rho   \rho
2   \rho   1   \rho   \rho   \rho
3   \rho   \rho   1   \rho   \rho
.   .   .   .   .   .   .
.   .   .   .   .   .   .
.   .   .   .   .   .   .
m   \rho   \rho   \rho   \rho   1
```
Underlying model assumptions

Assumes that there is some overall intercept $a$ for the population of practices

$u_i$ is the discrepancy between $a$ and the true intercept in the $i^{th}$ practice termed as the practice level residuals

$u_i$ is generated from a Normal distribution with a mean 0 and variance $s_u^2$

$$u_i \sim N(0, s_u^2)$$

$s_u^2$: Between practice variance

$e_{ij} \sim N(0, s_e^2)$, termed as subject level residuals

$s_e^2$: Within Subject variance (between different patients within a single practice)
Underlying model assumptions

Within and between-cluster random effects are assumed to be mutually independent.

Homogeneity of subject level residuals ($e_{ij}$)

Can use residual and scatter plots to check these assumptions.
One can estimate the Intra cluster correlation coefficient ICC from the variance components:

\[ \text{ICC} = \frac{s^2_u}{s^2_u + s^2_e} \]

Correlation between patients within practices

It is expressed as the variation between clusters as a proportion of total variance.
Mixed Effects Models

Method of Estimation

Can use method of maximum likelihood (ML) methods to estimate coefficients

The **maximum likelihood** estimator of the variance components can be biased downwards, particularly if number of clusters is small

Restricted maximum likelihood (REML) methods proposed as alternative

They are known to provide unbiased estimates
Mixed Effects Models

**Significance tests:**
Possible to perform both Likelihood ratio and Wald tests explanatory variables to assess significance of when using ML estimation.
Can only perform Wald tests when using REML estimation,
Tests based on REML may not be valid to assess significance of fixed effects (Welham & Thompson)
Goldstein suggests use of ML for model exploration and hyp. tests, & REML for estimation if required.
### Example: BMI

**Multilevel model**

<table>
<thead>
<tr>
<th></th>
<th>Estimates</th>
<th>SE</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention</td>
<td>1.90</td>
<td>0.83</td>
<td>0.022</td>
</tr>
<tr>
<td>Constant</td>
<td>29.37</td>
<td>0.62</td>
<td></td>
</tr>
</tbody>
</table>

Between cluster variance 0.92
Within cluster variance 32.60

**Summary statistic method**

<table>
<thead>
<tr>
<th>Method</th>
<th>Intervention effect</th>
<th>SE</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unweighted Means</td>
<td>2.47</td>
<td>0.88</td>
<td>0.008</td>
</tr>
<tr>
<td>Weighted Means¹</td>
<td>1.93</td>
<td>0.82</td>
<td>0.024</td>
</tr>
<tr>
<td>Weighted Means²</td>
<td>1.82</td>
<td>0.81</td>
<td>0.031</td>
</tr>
<tr>
<td>Ignoring Clustering</td>
<td>1.82</td>
<td>0.76</td>
<td>0.018</td>
</tr>
</tbody>
</table>

1 Weighted by inverse of estimated variance
2 Weighted by practice size
Example: BMI

One can estimate the ICC from the variance components:

\[ r = \frac{s^2_u}{s^2_u + s^2_e} \]

\[ = 0.028 \]

Multilevel models allow handling of clusters of unequal size
Example: BMI

Use of residual plots to check model assumptions
Checking normality of residuals at practice level:
Example: BMI

Checking normality of residuals at patient level
Example: BMI

Checking dependency between the cluster & patient level residuals

\[ u_{\text{praccode}} \]
\[ e_{\text{praccode},t} \]

-10 0 10

-10 -5 0 5

-10 0 10 20
Example: BMI

Extension of the model

Possible to include multiple covariates both at practice & patient levels

\[ BMI_{ij} = a + u_i + b_1 \text{group}_{ij} + b_2 \text{age}_{ij} + b_3 \text{gender}_{ij} + e_{ij} \]

- **Results**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Est.</th>
<th>SE.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention</td>
<td>1.88</td>
<td>0.77</td>
<td>0.014</td>
</tr>
<tr>
<td>Age</td>
<td>-0.11</td>
<td>0.04</td>
<td>0.003</td>
</tr>
<tr>
<td>Gender (male vs. female)</td>
<td>3.15</td>
<td>0.72</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Constant (intercept)</td>
<td>34.34</td>
<td>2.21</td>
<td></td>
</tr>
</tbody>
</table>

Between cluster variance 0.59
Within cluster variance 29.49
### Example: BMI

<table>
<thead>
<tr>
<th></th>
<th>Estimates</th>
<th>SE</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention</td>
<td>1.90</td>
<td>0.83</td>
<td>0.022</td>
</tr>
<tr>
<td>Constant (Intercept)</td>
<td>29.37</td>
<td>0.62</td>
<td></td>
</tr>
<tr>
<td>Between cluster variance</td>
<td>0.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within cluster variance</td>
<td>32.60</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Further model extensions

Possible to incorporate more complex correlation/variance structure by extending model adding variance components

Allow intercept and slopes to vary: Lines are no longer parallel
Precision of the parameter estimates

Estimates of SE obtained for random parameters may not always be reliable.

Estimates of intervention effects are obtained assuming that random parameters (variance components) are known and not estimated.

This may result in too narrow CI for intervention effects.

May need to use bootstrapping (re-sampling method) to provide confidence intervals for random parameters.
Consider the following model

\[ BMI_{ij} = a + u_i + b \text{ group}_i + e_{ij} \]

Select random samples of \(u_i\)s from \(N(0, \sigma_u^2)\) and \(e_{ij}\)s from \(N(0, \sigma_e^2)\) where \(\sigma_u^2\) and \(\sigma_e^2\) are estimated from the model

Add these to the fixed part of the model

Generate a new set of responses

Fit model again and estimate for \(a\), \(b\), \(\sigma_u^2\) and \(\sigma_e^2\) for each sample

Generate 1000 random samples

Confidence intervals can be obtained by assuming normality or non-parametrically from percentiles of empirical bootstrap values
General remarks

Possible to incorporate imprecision associated with random parameter estimates into estimation of fixed effects.

Random parameters may be poorly estimated if no. of clusters is small.

Sample size calculation as before.

However, if adjustment for explanatory variable is necessary in the model, that will increase the sample size requirement.


