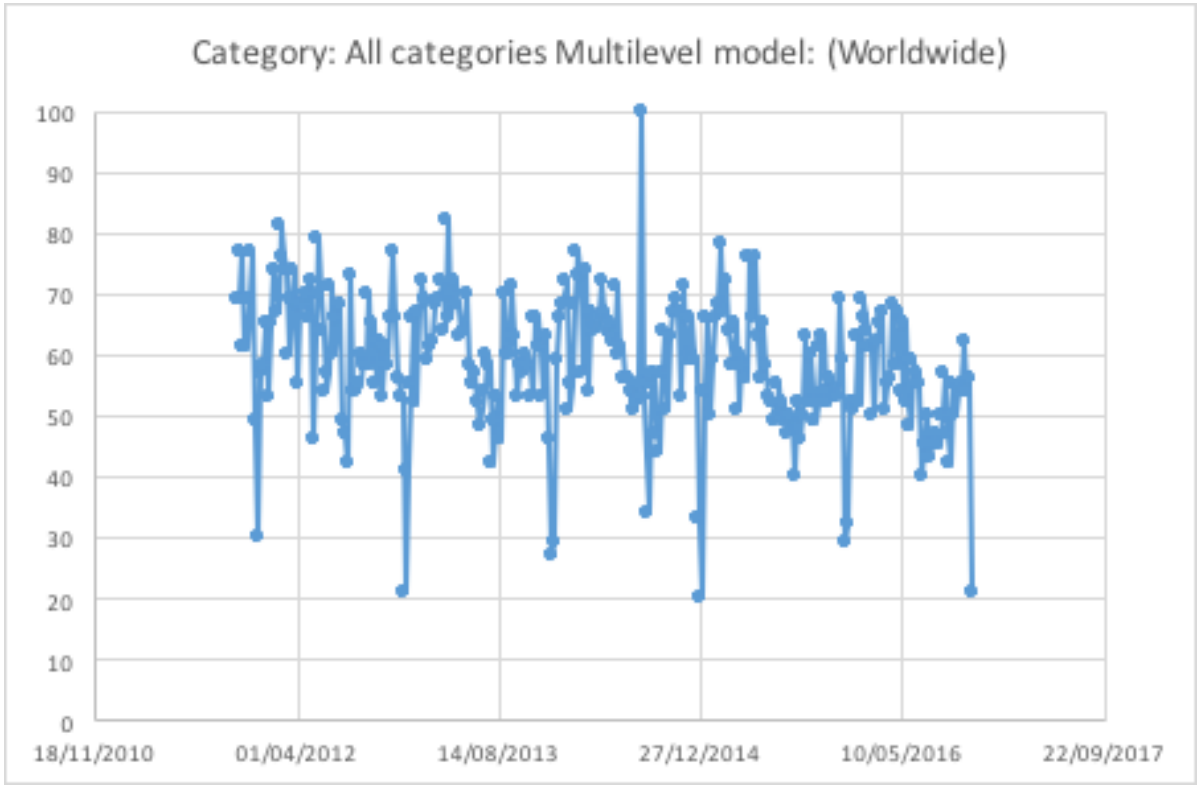
LIHS Mini Master Class
Multilevel Modelling

Robert M West

9 November 2016

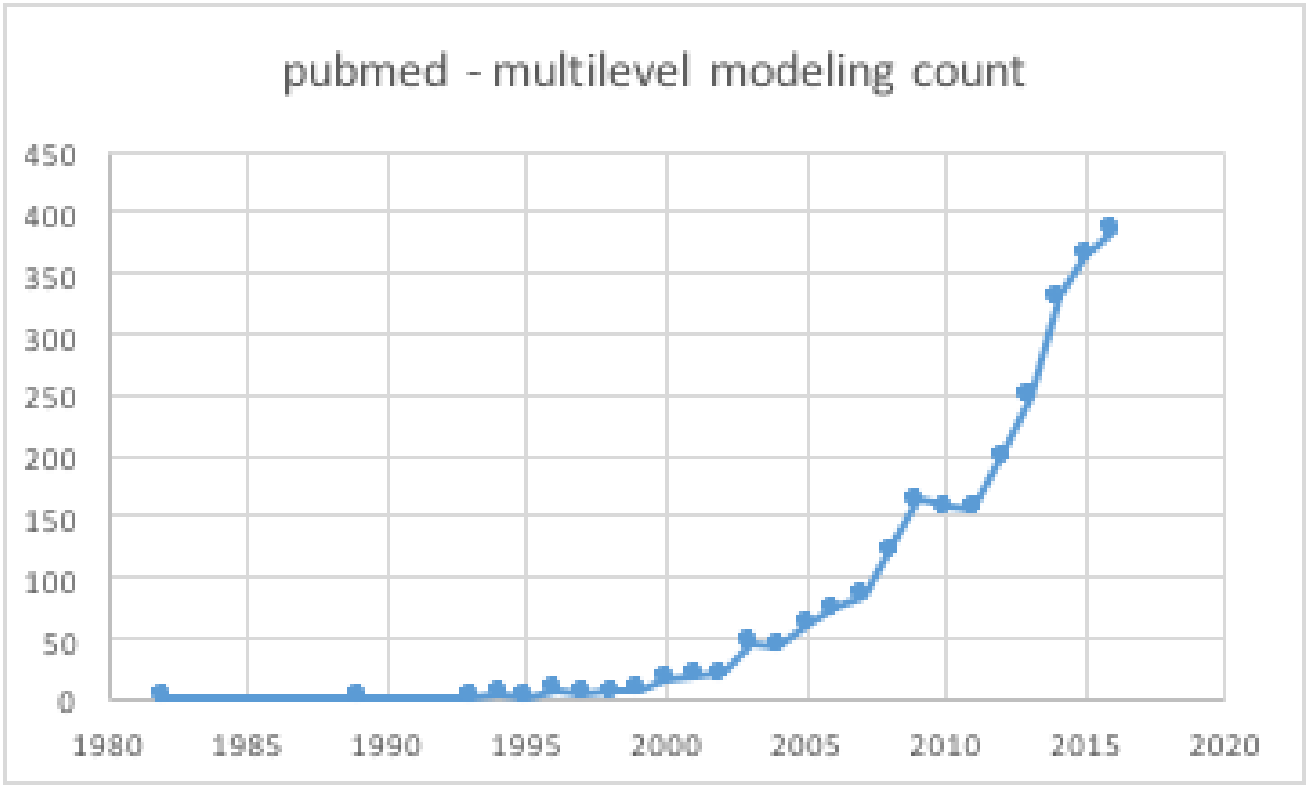
Outline

Steady interest in MLM – in general



A search with Google trends shows steady interest in recent years.

Publications within medicine



A search with PubMed shows a strongly increasing trend within biomedical disciplines.

Multilevel data

Data often comes as multilevel data or is collected in clusters

- Hierarchical or clustered data
- Cluster RCT where treatment allocated at cluster level
- Patients clustered with a general practice
- Multiple measurements made on patients

Rather than try to think of the general situation, think first about patients grouped within General Practices.

Can get the wrong answer

Ignoring the multilevel structure of data can lead to the analysis giving the wrong answer.

- Data are not independent: most statistical analysis assumes independence
- There should be research questions dependent on the data structure
  - Questions about individuals
  - Questions about groups

Aggregation

- It may be possible to simply aggregate the data to the highest level in the hierarchy.
- Then no special modelling needed
- May loose detail: group level only answers
- Beware the ecological fallacy: interpretations at group level may not be valid at individual level

Aggregation gives simple analysis but a limited range of answers and interpretation can be hard.

The structure is interesting

- Want to know about groups as well as individuals: GPs as well as patients
- Does it matter which GP a patient attends?

Random effects

The most common approach to multilevel modelling is to include random effects.

- Simplest model will include a random intercept
- Can include random slopes

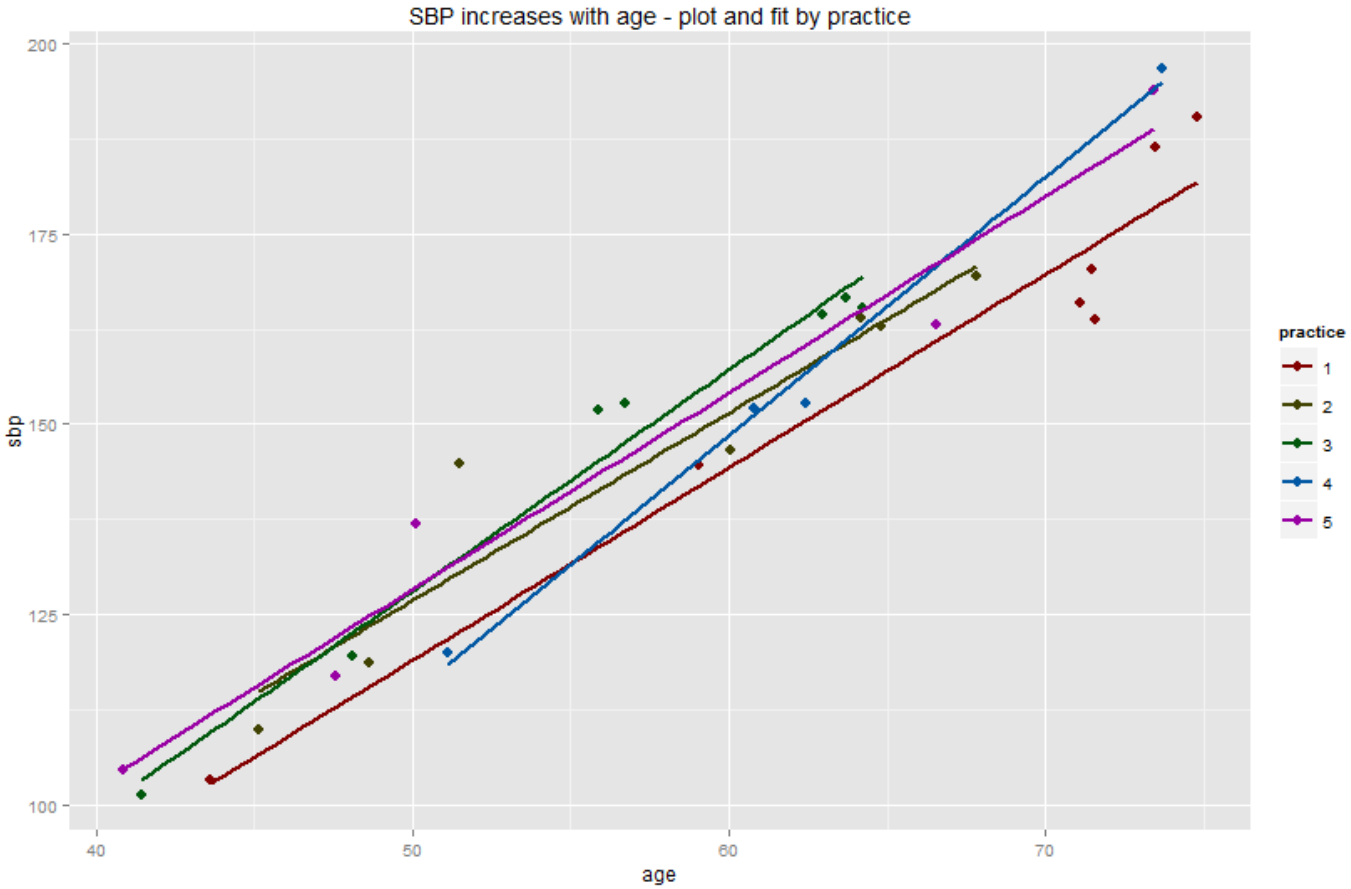
Consider an example.

Systolic blood pressure

- Patients aged 40–74: health check patients
- Know that SBP increases with age
- GPs have a role in controlling SBP of patients

I simulated some data.

Example – plotted



Comments

There are several points and questions which arise from seeing the plot

- An overall increase of SBP with age is seen across all patients
- There may be variation between practices – perhaps different intercepts, perhaps different slopes
- Can formalize these issues with modelling and test hypotheses

Some maths

The model might be

$$SBP_{i,j} = \alpha + \beta age_{i,j} + u_i + \epsilon_{i,j}$$

where

$$\epsilon \sim N(0, \sigma^2)$$

and

$$u \sim N(0, \sigma_1^2)$$

Some R code – naïve model

naive.model = lm(sbp ~ age, data=eg)
summary(naive.model)
Call: lm(formula = sbp ~ age, data = eg)

Coefficient	Estimate	Std error	t value	Pr(t)
(Intercept)	1.0784	8.1734	0.132	0.896
age	2.4999	0.1352	18.492	2e-16

Residual standard error: 7.701 on 28 df

R code – clustered model

```
library(lme4)
clustered.model = lmer(sbp ~ age + (1| practice), data=eg)
summary(clustered.model)
```

R output – clustered model

Random effects:

Groups	Name	Variance	Std.Dev.
practice	(Intercept)	17.01	4.124
Residual		45.54	6.748

Number of obs: 30, groups: practice, 5
Fixed effects:

Coefficient	Estimate	Std. Error	t value
(Intercept)	-6.1030	7.8358	-0.779
age	2.6216	0.1262	20.779

Correlation of Fixed Effects:
(Intr) age -0.959

Intraclass correlation coefficient – ICC

ICC = 17.01/ (17.01+45.54)
is 0.27, which is high.
There is considerable contribution from GPs.

Questions?

Questions, comments, feedback welcome