Lecture 9

Introduction to multilevel models

Social determinants

- People from the same area have more similar health status to each other than to people from other areas.
- …because of differing cultural, economic, political, climatic, historical, or geographical contexts
- clustering of **individual** health status within **areas**

Social epidemiology and public health

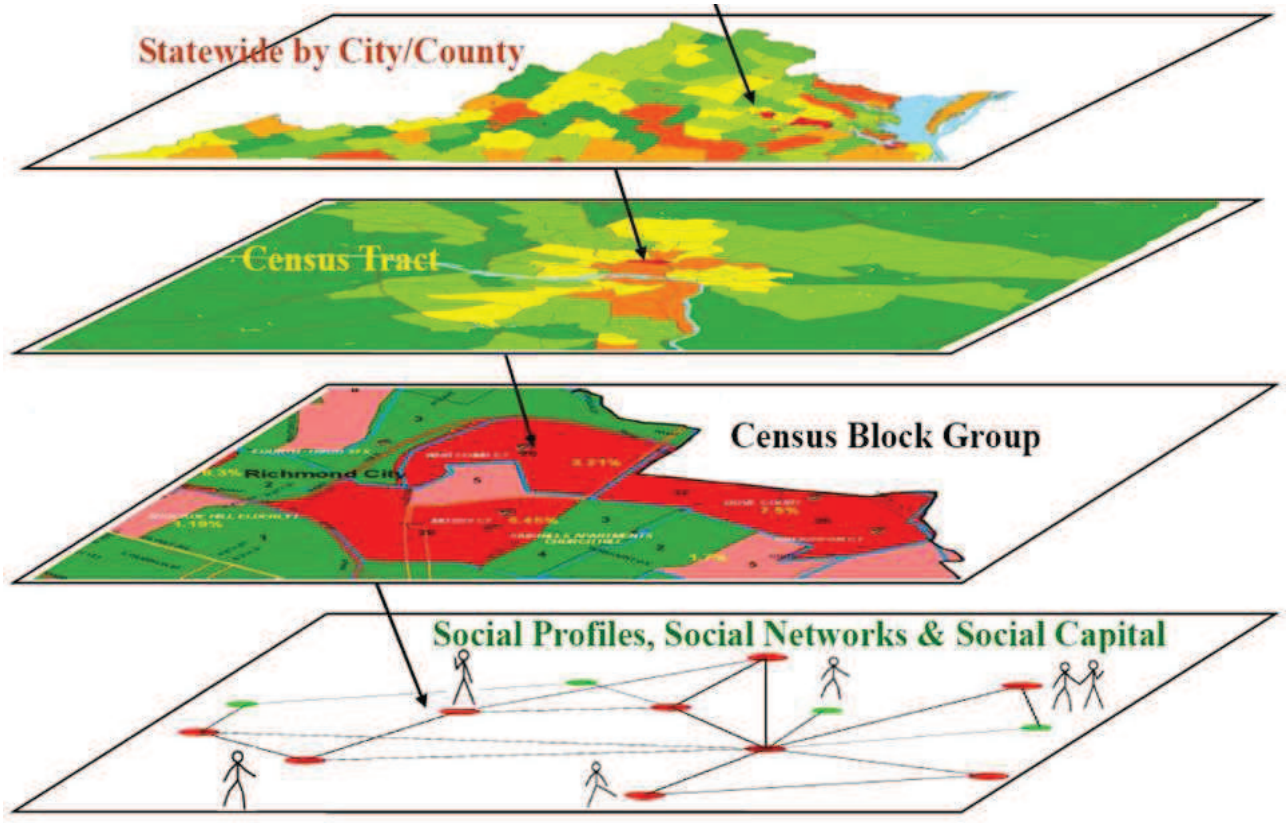
- Social epidemiology: focus interventions to reduce health inequalities on certain geographical areas rather than on specific people

- Public health: understand the significance of specific contexts for different individual health outcomes

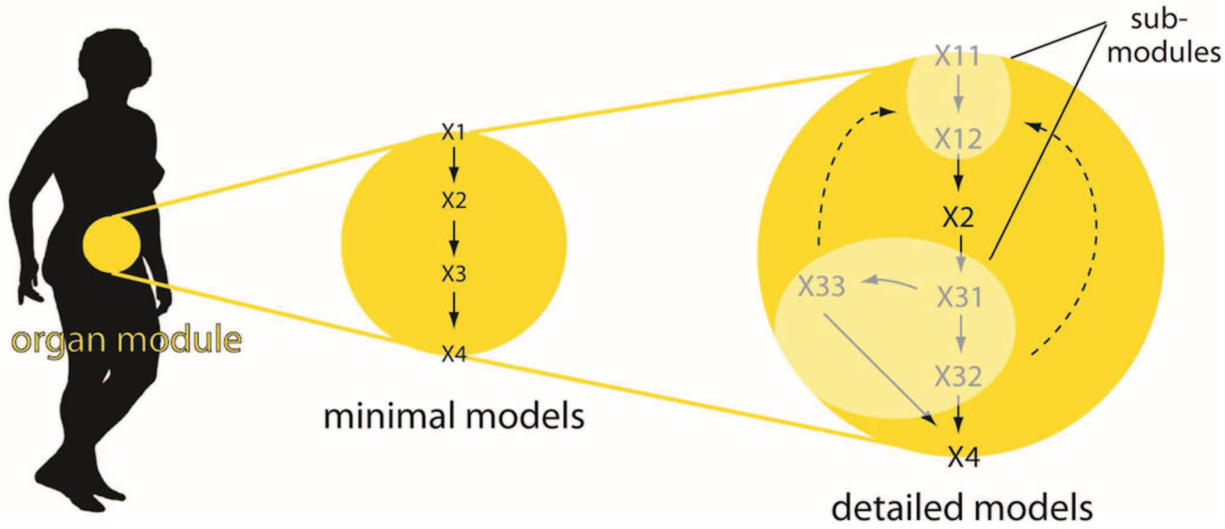
Multilevel units of analysis

Individuals are organised within a nearly infinite number of levels of organisation

- Individual up: families, schools, neighbourhoods, countries
- Individual down: body organs, cellular matrices, DNA
- Overlapping units: area of residence and work environment


Statewide by City/County
Census Tract
Census Block Group
Richmond City
Social Profiles, Social Networks & Social Capital


INTREC


sub-
modules
X11
X12
X2
X33
X31
X32
X4
X1
X2
X3
X4
organ module
minimal models
detailed models

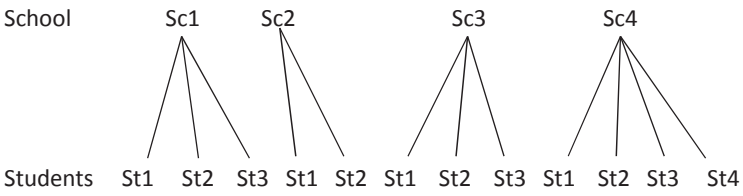
Levels, classifications and units

- A *level* (e.g. pupils, schools, households, areas) is made up of a number of individuals *units* (e.g. particular pupils, schools)
- The term *classification* can also be used but *level* implies a hierarchical relationship of units

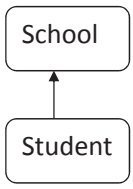
Two-level hierarchical structures

Students within schools

Unit diagram



Level diagram



Students within a school are more alike than a random sample of students. This is the 'clustering' effect of schools.

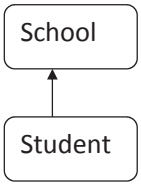
Data frame example

Classifications or levels		Response	Explanatory variables		
Student	*School*	*Student Exam score*	*Student previous Examination Score*	*Student gender*	*School Type*
1	1	75	56	M	State
2	1	71	45	M	State
3	1	91	72	F	State
1	2	68	49	F	Private
2	2	37	36	M	Private
3	2	67	56	M	Private
1	3	82	76	F	State

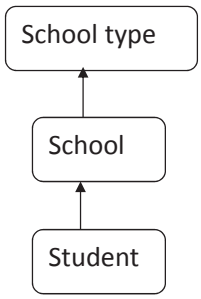
1. Do Males have better exam scores than Females?
2. Does the difference in Males and Females vary across schools?
3. Are Males more or less variable in their progress than Females?
4. Is School X (i.e. a specific school) different from other schools in the sample?

Variables, levels, fixed and random classifications

Given that school type (state or private) classifies schools, we could redraw our classification diagram



as



Do we now have a 3-level multilevel model?

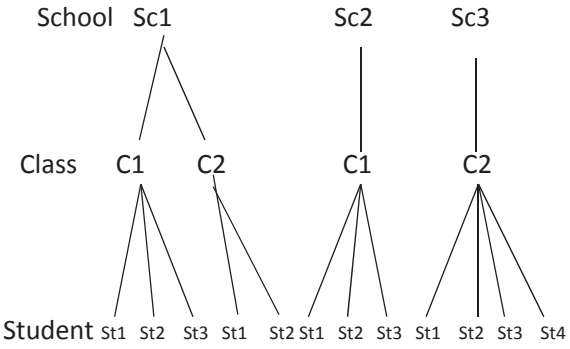
- We can divide classifications into two types: *fixed* classifications and *random* classifications.
- For a classification to be a level in a multilevel model it must be a random classification.
- School type is not a random classification!

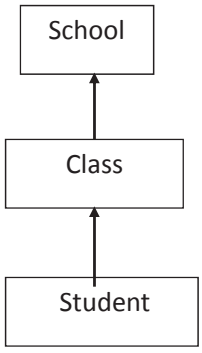
Random and Fixed Classifications

- A classification is a random classification if its units can be regarded as a *random* sample from a wider population of units.
- Traditional or single level statistical models have only one random classification which classifies the units on which measurements are made, typically people.
- Multilevel models have more than one random classification.

Three level structures

Students : classes : schools





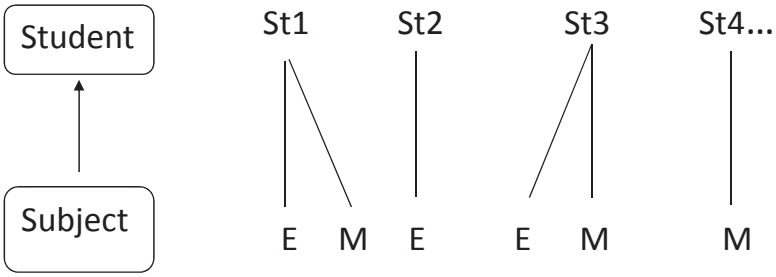
- Multilevel models allow a different number of students in each class and a different number of classes in each school.
- Can also be used in a repeated cross-sectional design

Repeated measures data

- When there are measurements on more than two occasions we can treat this as a level nested within individuals.
- Modelling between individual variation in growth, growth curves.
- In a multilevel repeated measures model data need not be balanced or equally spaced.
- Explanatory variables can be time invariant (gender) or time varying (age)

Repeated measures within individuals

- Example: we may wish to consider jointly English and Mathematics exam scores for students as two possibly related responses.

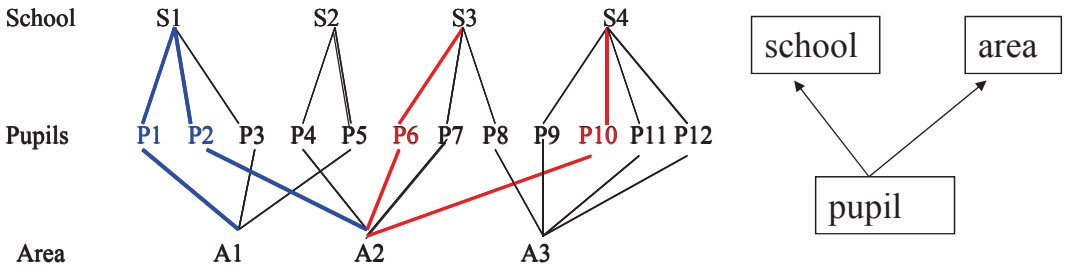


- A multilevel repeated measures model can estimate the correlation between responses and efficiently handle missing data.

Non-hierarchical structures

- So far all our examples have been exact nesting with lower level units nested in one and only one higher-level unit.
- …but social reality can be more complicated than this!
- There are two non-hierarchical structures which have been able to deal with all the different types of designs, realities and research questions
  - Cross-classified structures
  - Multiple membership structures

Cross-classified Model



In this structure schools are not nested within areas:

Pupils 2 and 3 attend school 1 but come from different areas

Pupils 6 and 10 come from the same area but attend different schools

School and area are cross-classified

Multilevel models

Multilevel methods consist of statistical procedures that apply when:

1. the observations being analysed are correlated or clustered;
2. the causal processes are thought to operate simultaneously at more than one level;
3. there is an interest in describing the variability in the population

(Diez-Roux 2002; Subramanian 2004a, 2004b; Subramanian et al. 2003).

The "empty" model

- Does not include any covariate but focuses only on how health differences are distributed between people and between areas.

- Example: study of 25,000 subjects, 35 to 64 years old, living in the 39 neighbourhoods of an imaginary city.
- The individual outcome variable is systolic blood pressure (SBP).

SBP example of empty model

$$SBP_I = SBP_C + E_{N\text{-}c} + E_{I\text{-}c}$$

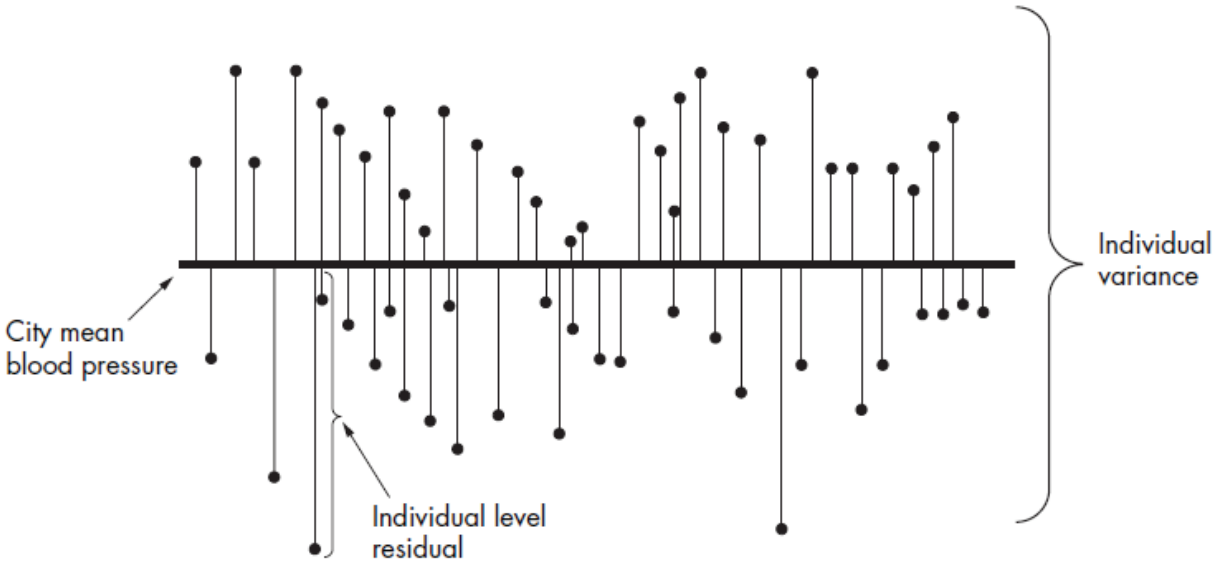
- SBP_I = SBP of an individual in a neighbourhood
- SBP_C = Mean SBP of the city
- $E_{N\text{-}c}$ = Difference between the city SBP mean and the neighbourhood SBP mean (also known as neighbourhood "shrunken residual")
- $E_{I\text{-}c}$ = Difference between the neighbourhood SBP mean and the individual SBP value (also known as "individual residual")

SBP example of empty model

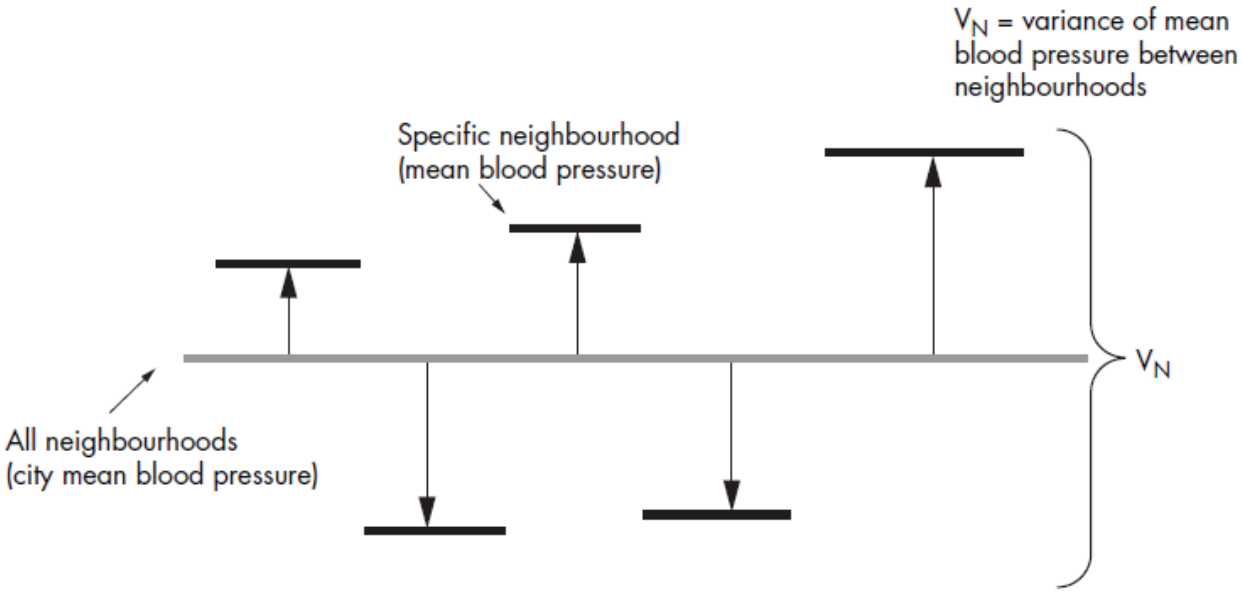
- The presence of neighbourhood and individual residuals in the empty multilevel model shows that SBP varies both at the individual and at the neighbourhood level.

- The main intent of the empty model is to partition the total variance in SBP in the city into a variance that occurs between neighbourhoods and a variance that occurs between people.

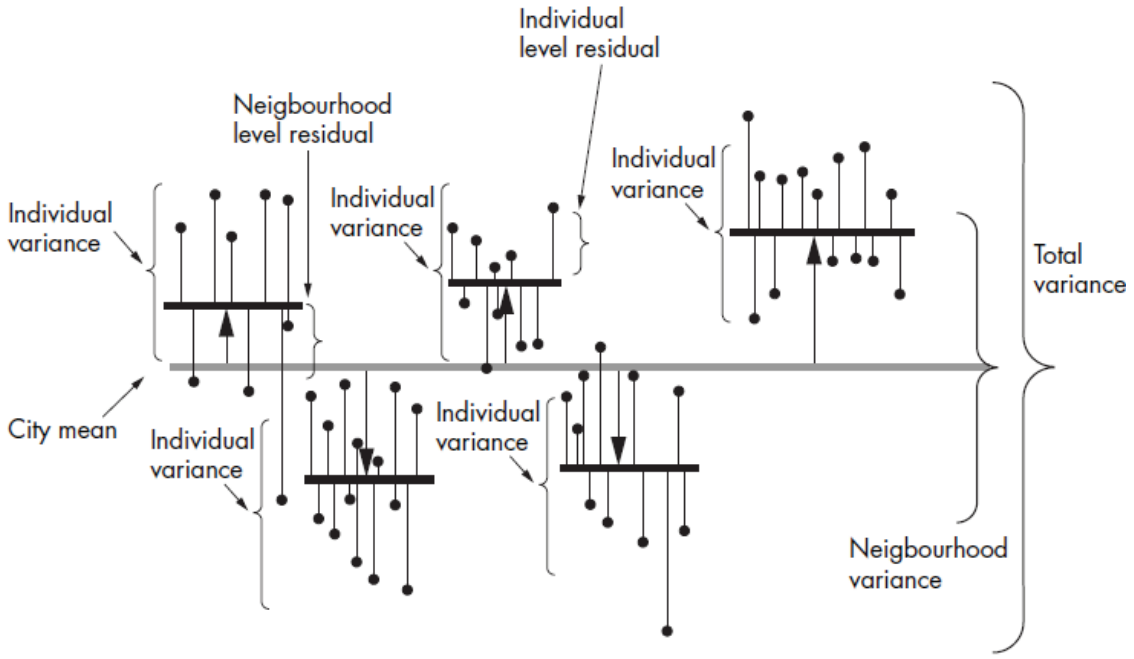
Individual variance



Neighbourhood variance



Total variance



Including individual co-variables

- What if we include age, BMI, and anti-hypertensive medication as co-variables?

- Fixed effects are used to model averages (for example, means or regression coefficients)
- Random effects are used to model differences (for example, neighbourhood variance).

Not so empty model…

$$SBP_I = SBP_C + \beta_1 \times age_I + \beta_2 \times BMI_I + \beta_3 \times AHM_I + (E_N + E_I)$$

Fixed effects

SBP in people not on AHM and with average age and BMI, and living in a neighbourhood with a shrunken residual equal to 0

Regression coefficient of the association between age and SBP

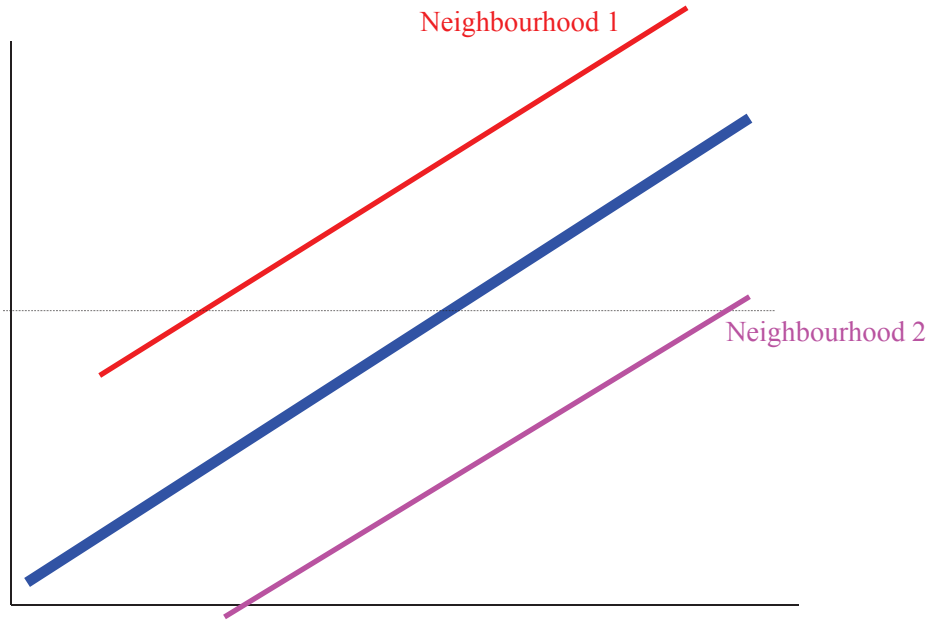
Regression coefficient of the association between BMI and SBP

Regression coefficient of the association between AHM and SBP

Random effects

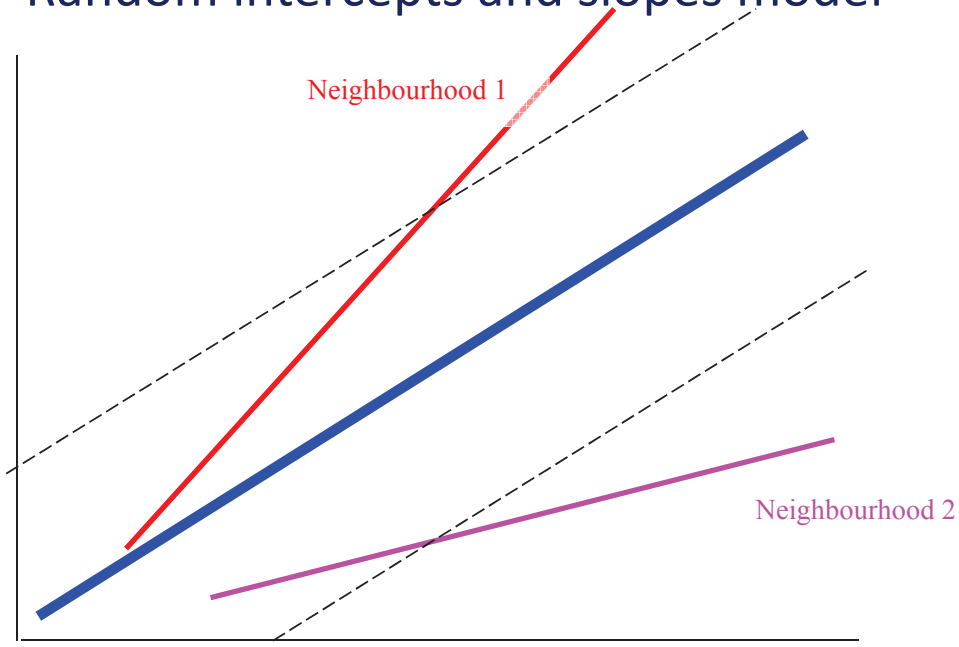
Neighbourhood shrunken residual + Individual residual

Random intercept models (parallel lines)



The neighbourhood lines have a different intercept but a similar slope. Therefore, differences between the neighbourhoods are constant.

Random intercepts and slopes model



As the slopes of the regression lines are different, the differences between the neighbourhoods are not constant

Example:

Van Minh *et al. BMC Geriatrics* 2010, **10**:7
http://www.biomedcentral.com/1471-2318/10/7

BMC Geriatrics

RESEARCH ARTICLE Open Access

Multilevel analysis of covariation in socioeconomic predictors of physical functioning and psychological well-being among older people in rural Vietnam

Hoang Van Minh[1*†], Dao Lan Huong[2†], Stig Wall[3†], Nguyen Thi Kim Chuc[1†], Peter Byass[3†]

Aims:

1. examine the effects of different socio-economic factors on physical functioning and psychological well-being among older adults in a rural community in northern Vietnam;
2. investigate the extent to which the two outcome variables co-vary within individuals.

2-level model of 17070 (two outcomes for each individual) at level 1 nested within 8535 individuals at level 2

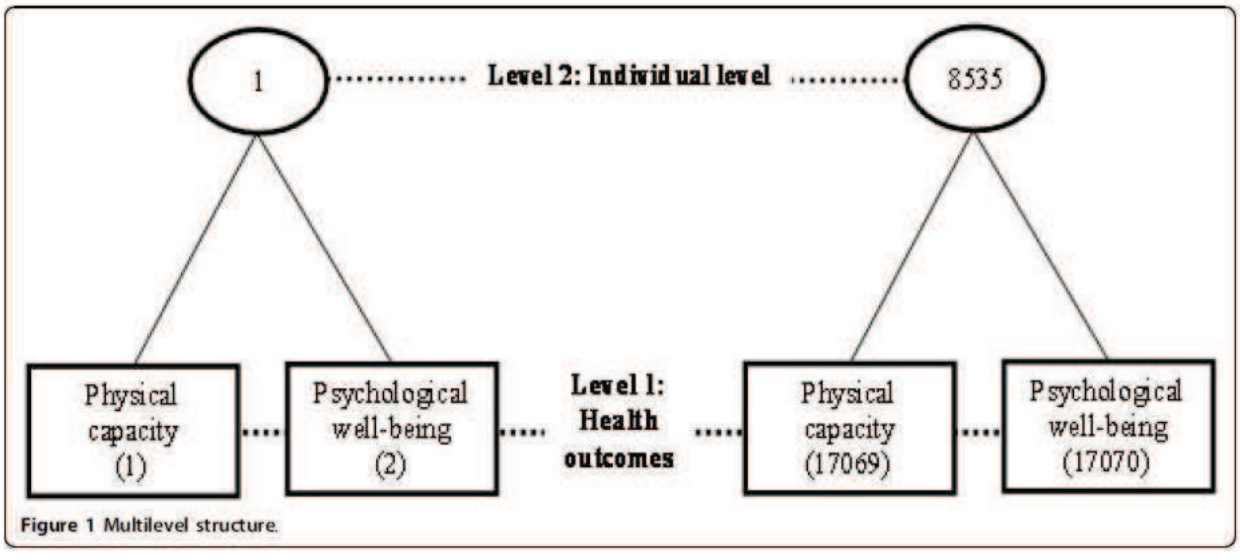


Figure 1 Multilevel structure.

Analysis

- The two-level model can be written as:

$$Y_{ij} = B_{01}Z_{1ij} + B_{02}Z_{2ij} + B_{11}Z_{1ij}X_j + B_{12}Z_{2ij}X_j + U_{1j} + U_{2j}$$

$$Z_{1ij} = \begin{cases} 1 \text{ if physical functioning} \\ 0 \text{ if psychological well-being} \end{cases}$$

$Z_{2ij} = 1 - Z_{1ij}$,
X_j = independent variables (gender, age, marital status, education, wealth status, place of residence, household size),
$\text{var}(U_{1j}) = \sigma^2_{u1}$ = variance in physical functioning,
$\text{var}(U_{2j}) = \sigma^2_{u2}$ = variance in psychological well-being,

Results: random part

Table 3 Random part of the multivariate response model, FilaBavi, Vietnam 2006

	Empty model		Full model		Change in estimate
	Estimate	SE	Estimate	SE	
Variance in Physical capacity	4.4	0.06	4.1	0.06	7%
Variance in Psychological well-being	48.4	0.74	36.6	0.56	24%
Covariance	5.0	0.17	4.3	0.14	14%

- The empty model indicates that there was significant variation in each outcome of interest.
- The variance in both physical functioning and psychological well-being became smaller after the socio-economic variables were included (Full model).
- The socioeconomic factors accounted for about 24% and 7% of variation in physical functioning and psychological well-being scores, respectively.

Results: fixed part

- Women were more likely to have poor physical functioning and low psychological well-being
- Age was a negative predictor of physical functioning, it had no effect on psychological well-being.
- People with higher educational levels better in both physical functioning and psychological well-being.
- Those in marital partnerships had better status of physical functioning and psychological well-being.
- Economic differentials were found for psychological wellbeing but not for physical functioning.
- Mountainous dwellers had lower levels of both physical functioning and psychological well-being.
- Household size had no important effect on the two outcomes variables.

Table 4 Fixed part of the multivariate response model, FilaBavi, Vietnam 2006

Socioeconomic variables	Physical capacity Coefficient [SE]	Psychological well-being Coefficient [SE]
Gender:		
- Male	Reference	Reference
- Female	0.34 [0.16]*	0.2 [0.05]*
Age:		
50-59	Reference	Reference
60-69	1.48 [0.17]*	0.08 [0.06]
70-79	4.05 [0.20]*	-0.02 [0.07]
80+	8.40 [0.27]*	-0.16 [0.09]
Education:		
- No schooling	Reference	Reference
- Less than six year of education	-2.20 [0.24]*	-0.35 [0.08]*
- Graduated from primary school and higher	-2.88 [0.29]*	-0.46 [0.10]*
Marital status:		
- Currently in marital partnership	Reference	Reference
- Currently single	0.58 [0.17]*	0.26 [0.06]*
Socio-economic:		
- 1st quintile [poorest]	Reference	Reference
- 2nd quintile	0.02 [0.24]	-0.34 [0.08]*
- 3rd quintile	-0.08 [0.24]	-0.48 [0.08]*
- 4th quintile	-0.16 [0.24]	-0.52 [0.08]*
- 5th quintile [richest]	-0.42 [0.25]	-0.72 [0.08]*
Family size:	0.02 [0.04]	-0.02 [0.01]
Place of residence:		
- Riverside/island	-0.61 [0.20]*	-0.58 [0.07]*
- Highland	-1.42 [0.15]*	-0.24 [0.05]*
- Mountainous area	Reference	Reference
Intercept:	15.15 [0.38]*	4.54 [0.13]*

* denotes a significant result ($p < 0.05$)

Discussion

- Some common significant socioeconomic predictors of both physical functioning and psychological well-being among the elderly
- Socioeconomic variables of interest had stronger associations with physical functioning (24% of variation) than psychological well-being (7% of total variation).
- Two health outcomes did not co-vary to a strong degree (adjusted correlation coefficient 0.35

More reading…

- The Centre for Multilevel Modelling (CMM) at the University of Bristol
- http://www.bristol.ac.uk/cmm/

- Also called mixed models, hierarchical linear models, or nested models