Modelling health-related behaviour

Professor Graham Moon
University of Southampton
2011 Seelye Charitable Trust Visiting Fellow

Acknowledgement: many slides used with permission of Kelvyn Jones and ESRC LEMMA project (University of Bristol); others by Liz Twigg (University of Portsmouth)

Focus

- Methodological: challenges in the application of MLM
- Substantive: health-related behaviour, particularly smoking and drinking
- Approach: overview and informal discussion

Multilevel Analysis

- Once you know that hierarchies exist you see them everywhere' (Kreft, 1990)
- For many (most?) research questions:
  - the real world has a complex structure **and/or**
  - we impose one during research design
- Ignoring structure leads to impoverished analyses and inferential error

Examples of real world hierarchies

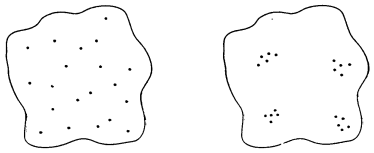
- **Education**
- pupils (1) in schools (2)
- pupils (1) in classes(2) in schools (3)
- **Geography**
- houses(1) in neighbourhoods(2) in regions(3) in countries(4)
- **Business**
- individuals(1) within teams(2) within organizations(3)
- **Psychology**
- individuals(1) within family(2)
- individuals(1) within twin sibling pair(2)
- **Economics**
- employees(1) within firms(2)
- **NB all are structures in the POPULATION** (ie exist in reality)

5

Hierearchy and sampling design

- for efficient collection of data
- most large-scale surveys are **not** SRS

a) Simple random sample b) Two stage sample



- Two-level structure imposed by design
- respondents nested within PSU's

6

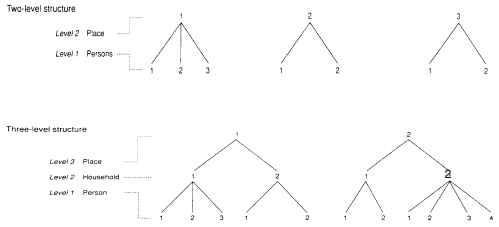
Multistage sampling designs

- Multistage designs (usually) generate dependent data
  - individuals living within the same PSU can be expected to be more alike than a random sample
- The 'design effect':
  - Inferential procedures (inflated SE's so problems with confidence limits, tests)
  - Type 1 errors: finding a relationship where none exists
- Multilevel models take account of this dependency and automatically correct for the 'design effect'

7

Traditional Analysis

Two-level structures and three-level structures
Unit Diagrams



8

Challenge I
Focus more on the random part

Understanding the basics of a two-level model

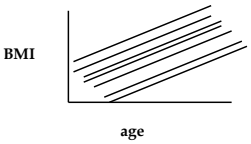
a) 500 people living in 7 communities

BMI vs age

all 500

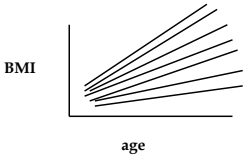
- age increases, BMI increases
- place does not matter – purely individual relationship
- NO CONTEXT; GEOGRAPHY UNIMPORTANT

b) Alternatively could have 7 regression lines for the communities



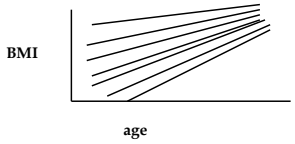
- BMI/age relationship same in each community
- People of same age in different communities have uniformly higher BMI than others
- PARALLEL LINES AND CONTEXT MATTERS (i.e. geography matters)
- Impact of individual and geographical influences

c) Alternatively …



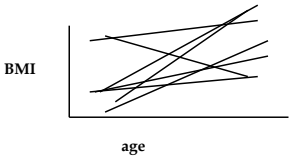
- steepness of lines varies from place to place
- place makes very little difference for the young
- but communities have very different BMI for elderly
- Geography more important for the elderly than the young

d) In contrast…



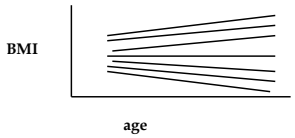
- again steepness varies from place to place
- large place-specific differentials for the young
- places similar for elderly

e) Or…

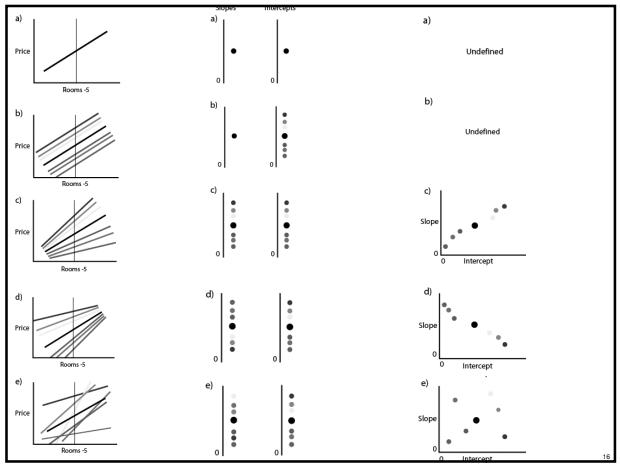


- here we have complex interactions between age and place COMPLEX HETEROGENEITY
- in some communities it is the young who have relatively high BMI, in others it is the old

f) And finally…



- here the young are similar in all communities
- the elderly are quite different – some communities have a high rate and some have a low rate for the elderly
- if we just looked across all communities it would appear that age does not matter



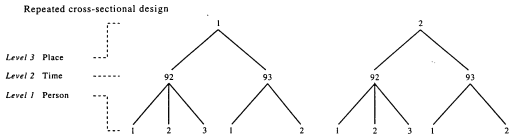
Challenge II
Look closely at interactions

Interactions

- Within Level
  - Smoking as a function of age
  - Smoking f sex
  - Smoking f age and sex
- Cross-level
  - Differential effects for younger people in deprived areas?
  - Differential effects for deprived people in deprived places?

Challenge III
Incorporate Time

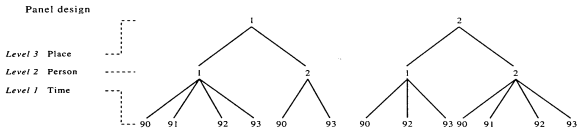
Repeated cross-sectional design



- **Example**
  - Level-3 is the area, level-2 is the year and level-1 is the individual. Level-2 represents repeated measurements on the area.
- **Handles imbalance:**
  - particular areas not included in particular years (level 2);
  - different number of individuals in each cohort in each area (level 1)
  - even if area is only measured once do NOT discard it, include it!
- **Research Question**
  - modeling changing area behaviour as different cohorts of individuals pass through the area
  - **EG** persistence of problem drinking in NE England

20

Repeated measures design: True Panels

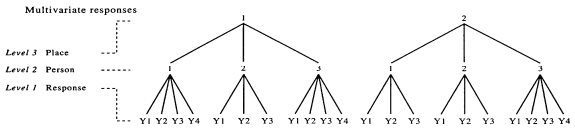


- **Structure of Example:**
  - Level-3 is the area, level-2 is the individual and level-1 is the occasion; PEOPLE are repeatedly measured ie the SAME individuals
- **Handles imbalance**
  - does not require a fixed set of repeated observations for all persons
  - both the number of observations per person and the spacing between observations may vary
- **Research Question**
  - what is the volatility of the outcome; how changeable are individuals, and how changeable are areas?
  - Classic longitudinal analyses of individual behaviour change over time

21

Challenge IV
Recognise co-behaviour / co-consumption

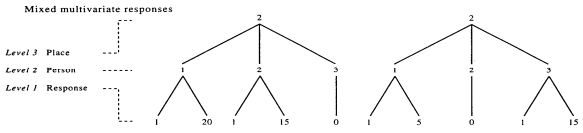
Multivariate multilevel structures



- **Structure:**
  - Multiple response variables, representing repeated measurements of distinctive but not unrelated outcome variables
  - SET of response variables (level-1) nested within individuals (level-2) nested within neighborhoods (level-3)
  - NB- imbalance: not all responses measured on all individuals
- **Research Question**
  - responses are health related behaviours (drinking, smoking exercise & diet)
  - how correlated are the behaviours at the individual level and community level (taking account of other characteristics)?

23

Mixed multilevel structures

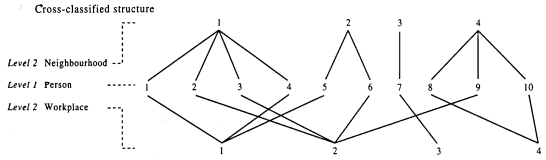


- **Structure:**
  - Level-3: Places; Level-2: Individuals; Level-1: Two responses:
    - smoke or not (a qualitative state)
    - how many cigarettes they consume in a day (a quasi-continuous measure)
    - Imbalance; data with a spike at zero
- **Question**
  - is social class related to whether you smoke or not; but unrelated to the amount you smoke?
  - Are places that have a high proportion of smokers also those that tend to have higher smoking intensity (ie more heavy smokers)?

24

Challenge V
Recognise non-hierarchical spatialised relationships

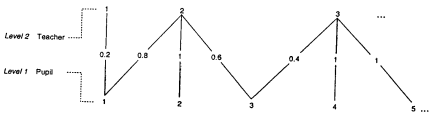
Non-Hierarchical structures
Cross-classified structures I



- **Structure:**
  - Individuals at level-1 in Workplaces at level-2 AND neighborhoods at level-2
  - Workplaces and neighborhoods are not nested but CROSSED.
  - Individuals are seen as occupying more than one set of contexts
- **Questions:**
  - Relative contribution of neighborhoods and workplaces to health-related behaviour
  - Where you live and which general practice you use

26

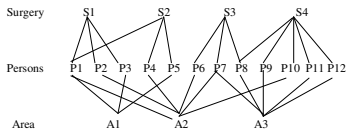
Non-Hierarchical structures
Multiple membership structures



- **Structure:**
  - lower level unit can belong simultaneously to more than one higher-level unit
  - Drinkers and bars; PA and parks
  - Some respondents use more than one facility
  - Structure includes a 'weight' based upon the proportion of time the respondent spends at each facility (sum to 1)

27

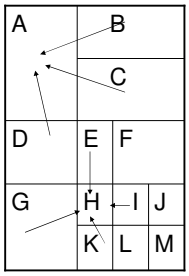
Combining structures: crossed-classifications and multiple membership relationships



- Person 1 moves in the course of the study from Area 1 to 2 and from Surgery 1 to 2
- Person 7 moves areas but stays with the same surgery
- Person 8 changes surgeries but does not move
- In addition to Surgeries being crossed with Areas, people are *multiple members* of both areas and surgeries (captured by weights).

28

Spatial Models as a combination of strict hierarchy and multiple membership



Person in A

Affected by A(SH) and B,C,D (MM)

Person in H

Affected by H(SH) and E,I,K,G (MM)

Multiple membership defined by common boundary; weights as function of inverse distance between centroids of areas; analogous to geographically-weighted regression

29

Scottish Lip Cancer spatial multiple-membership model

- Response: observed counts of male lip cancer for the 56 regions of Scotland (1975-1980)
- Predictor: % of workforce working in outdoor occupations (Agric;For; Fish) Expected count based on population size
- Structure areas and their neighbours defined as having a common border (up to 11); equal weights for each neighbouring region that sum to 1

Rate of lip cancer in each region is affected by both the region itself and its nearest neighbours after taking account of outdoor activity

- Model Log of the response related to fixed predictor, with an offset, Poisson distribution for counts;

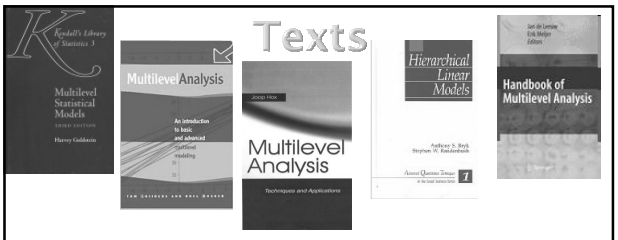
NB Two sets of random effects
1 area random effects; (ie unstructured; non-spatial variation);
2 multiple membership set of random effects for the neighbours of each region

30

Conclusions

- Stones left unturned
  - Group and Grand Mean Centring
  - Latent variables
  - Meta-analysis
  - Responses at different level
  - MCMC
  - SUTVA
- Realistically complex modeling
- The limits of data and data limitation

Texts



- Comprehensive but demanding! : Goldstein
- Thorough but a little dated: Snijders & Bosker
- Approachable : Hox
- Thorough (HLM) Raudenbush & Bryk
- Authoritative: de Leeuw & Meijer
- Applications: education, O'Connell & McCoach
- Applications: health, Leyland & Goldstein

http://www.cmm.bristol.ac.uk/learning-training/multilevel-m-support/books.shtml

32