Methods Seminar Six: Hierarchical Models

August 2, 2002

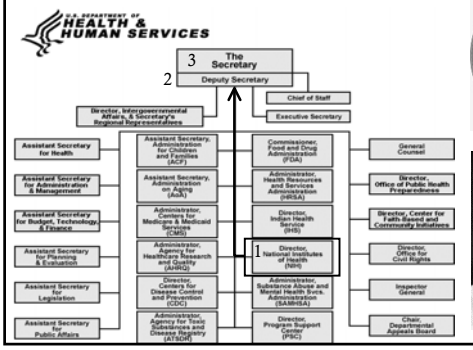
Thomas E. Love
tel3@po.cwru.edu

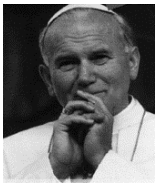
David Litaker
WrightLit@aol.com

Plan of Action

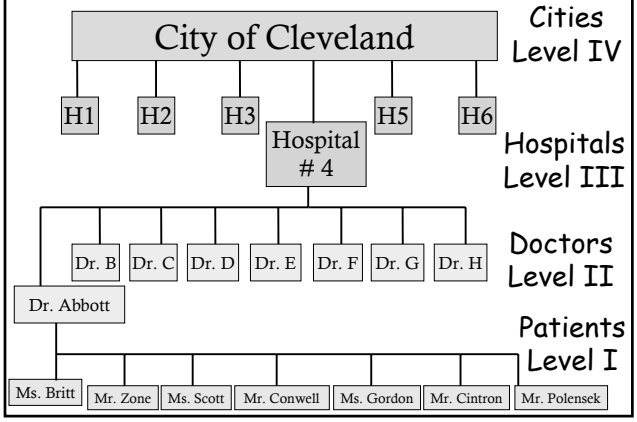
- Love
  - Introduction / Motivation / Fundamentals
- Litaker
  - Looking at A Real Example
  - Interpreting the Hierarchical Model
  - What Does Using Hierarchical Models Get Me?
- Love
  - Discussion / Summary

Hierarchy: a series of ordered groupings of people or things ranked one above the other within a system





Hierarchies in Health



Do I Really Need To Use Hierarchical Models?

- Goal: assess differences in mortality rates across doctors relative to a specific condition or procedure
- Data: random samples of patients nested within each doctor (2-level hierarchy)
- Adjustments for patient-level covariates (age, gender, severity of diagnosis, etc.)
- Adjustments for doctor-level covariates (time in practice, certification, # procedures, etc.)

"Naïve" Modeling Strategies: 1. The Aggregated Model

- Model Level I units (e.g. patients)
- Ignore Level II and up (e.g. doctors, hospitals, etc.)
- What are the limitations of doing this?

Limitations: The Aggregated Model

- Wrong std. errors due to ignoring clustering (can fix up with survey analysis software).
- Can't estimate effects of Level II (and higher) units (e.g. doctor effects, hospital effects, etc.)
- Within and Between-Level II effects are confounded…
  - Level I effect: "healthy patients treated better by each doctor" vs.
  - Level II effect: "good doctors have healthy patients"

"Naïve" Modeling Strategies: 2. The "Fixed Effects" Model

- Dummy variable for every Level II unit
- Benefits:
  - estimate within-Level II unit effects
- Limitations:
  - Lose all between Level II unit information
  - No estimates of Level II effects
  - Still no good standard errors

WrightLit@aol.com or
TEL3@po.cwru.edu
http://www.chrp.org/love

"Naïve" Modeling Strategies: 3. Model Level II Means

- Regress Level II mean outcome on predictors
  - e.g. regress success rate of surgeons on surgeon characteristics, mean patient characteristics
- Benefits: estimate between-Level-II unit effects
  - e.g. "better results obtained by Board-certified surgeons and surgeons with younger patients."

Limitations of Modeling Level II Means

- Poor use of information from within clusters
  - Cannot interpret "doctors with younger patients have better results" as "younger patients have better results"
- No estimation of individual level variation

Hierarchical Modeling

- Other names: Multi-level Models, Random effects models, Empirical Bayes models, Hierarchical Bayes models, Random coefficient models, Mixed models, Growth curve models.
- Basic idea: "random" and "fixed" (systematic) effects are explicitly modeled at each level of the hierarchy

What Do We Mean By "Random" Effects?

- We're not trying to model noise/chaos here.
- The covariate patterns we observe for the individuals in our study are just a sample of all possible covariate patterns.
- These covariates can vary according to a distribution (cannot be viewed as fixed) but can have a real, and large impact on the outcomes of interest.
- For example, a treatment in an RCT is fixed – in a quasi-experiment, there's not uniformity to any treatment – so the effect of the treatment can vary randomly according to a distribution.

Importance of Hierarchical Models

- Can explore the nature and extent of relationships within level II units and among level II units.
- Estimates based on aggregating level I units over level II may not be appropriate
  - What if the level II's are heterogeneous?
  - Regression under-estimates standard errors

Notation

- j = index of Level I unit (e.g. patient)
- i = index of Level II unit (e.g. doctor)
- y_{ij} = outcome measure for patient j of doctor i
- x_{ij} = characteristics of patient j of doctor i
- Ordinary (aggregated) regression model:

$$y_{ij} = x_{ij}\beta + e_{ij} \qquad e_{ij} \sim N(0, \sigma^2)$$

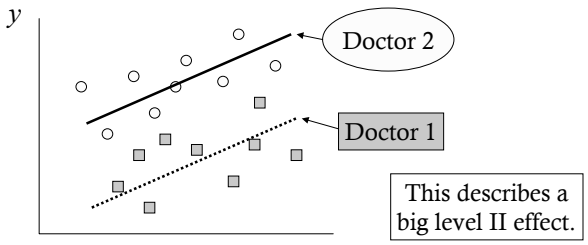
Random Intercept Model

- Model: $y_{ij} = x_{ij}\beta + u_i + e_{ij}$

$$e_{ij} \sim N(0, \sigma^2) \qquad u_i \sim N(0, \tau^2)$$

- u_i is Level II (doctor) random effect
- Estimate second variance component τ^2 for Level II effects

Random Intercept Model:
Effect of Age Same for Each Doctor
(Parallel regression)



x = Patient Age

Random Intercept and Slope

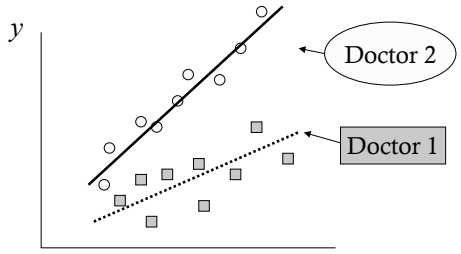
- Patients of different doctors have different slopes and different intercepts.
- Model: $y_{ij} = x_{ij}\beta_i + e_{ij}$

$$\beta_i = \beta_0 + u_i \qquad e_{ij} \sim N(0, \sigma^2)$$

$$u_i \sim N(0, \mathbf{T})$$

- T is now a covariance matrix – tells us how the intercept and slope are related

Random Intercept & Slope:
Effect of Age Different for Each Doctor



x = Patient Age

"Simple" Mixed Model Specification

We have J models for individual patients:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + \varepsilon_{ij}$$

To formulate the Level II models, treat the individual's β's as regression outcomes, using doctor covariates W_j …

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + v_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + v_{1j}$$

Mixed Model

$$y_{ij} = x_{ij}\beta_i + z_{ij}\alpha + e_{ij}$$

- β_i represents random part of coefficients (different coefficients for each doctor)
- α represents "fixed" (each doctor has the same coefficients) part

Output

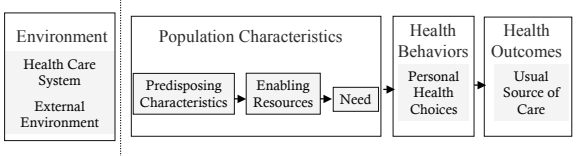
- Coefficients at various levels: α, β, γ
  - Interpret as effects of observed characteristics
- Variance components: σ^2, τ^2
  - Interpret as unexplained variation

Objectives

- Present research question, conceptual framework
- Consider traditional modeling approaches/limitations
- Describe use of multi-level strategy
- Highlight unique aspects of interpretation

Scenario and Conceptual Framework

- Data on:
  - patterns of access, insurance, chronic medical conditions for >15,000 Ohio residents
  - health system resources, economic status for all Ohio zips



Research Questions

- Do community characteristics explain variation in patterns of access, controlling for access-relevant characteristics of the individual?
- Is economic status of community important?

Also of interest:
Do uninsured residents of economically depressed areas have disproportionately greater problems obtaining access?

Variables - Level I (Individual Level)

- Dependent variable
  - Usual source of care (Yes/No)
- Independent variables
  - Continuously insured (Yes/No)
  - General Health status (1=excellent, 5=poor)
  - # of chronic medical conditions (0-3)

Variables - Level II (Zip/Community Level)

- Two continuous measures
  - Primary care physicians practicing in zip
  - Economic status/vitality of the zip

Aggregation

- Assign values for level II unit of analysis (zip) to each case at level I (individual) (n=15,613)

$$\text{NO\_RSOC} = \beta_0 + \beta_1(\text{CURR\_INS}) + \beta_2(\text{GEN\_HEALTH}) + \beta_3(\text{MED\_PROBS}) + \beta_4(\text{PCPS\_POP}) + \beta_5(\text{ECON\_STAT}) + r$$

Averaging

- Average data at level I (individual) for each level II (zip) unit of analysis (n=623)

$$\text{PROPORTION\_RSOC} = \beta_0 + \beta_1(\text{PROP\_CURR\_INS}) + \beta_2(\text{AVG\_HEALTH}) + \beta_3(\text{AVG\_MED\_PROBS}) + \beta_4(\text{PCPS\_POP}) + \beta_5(\text{ECON\_STAT}) + r$$

Problems

- Ignores correlation within groups
  - Assumption of independence violated
- Loss of power with averaging strategy
- Research questions not fully addressed

When is a Multi-level Model Better?

- If important variation exists at level II (between zips)
- If individuals within level II units (zips) share characteristics
- If alternative approaches do not address research questions

Does important variation exist at level II?

Determining Variance at Level II

Creating a "Null Model"

LEVEL I MODEL:

$$NO\_RSOC = \beta_{0j}$$

LEVEL II MODEL:

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

WrightLit@aol.com or
TEL3@po.cwru.edu
http://www.chrp.org/love

Do individuals within level II units share characteristics?

Intra-class correlation

$\rho = \tau^2 / (\tau^2 + \sigma^2)$

τ^2 = variance at level II
σ^2 = variance at level I

For binary outcomes:
$\rho = \tau^2 / (\tau^2 + \pi^2/3)$

Intra-class correlation

$\tau^2 = .09654$ $\rho = .029$

OK, is an intra-class coefficient of .029 big enough to worry about and should I use an approach that accounts for it?

Size of Intra-Class Correlation Inflates Alpha Level

	ρ		
N	*0.01*	*0.05*	*0.20*
10	0.06	0.11	0.28
25	0.08	0.19	0.46
50	0.11	0.30	0.59
100	0.17	0.43	0.70

Barcikowski, RS (1981) Journal of Educational Statistics, 6(3), 267-285

Do alternative approaches address research questions?

The Multi-level Approach

LEVEL I MODEL:

$$NO\_RSOC = \beta_{0j} + \beta_{1j}(CURR\_INS) + \beta_{2j}(GEN\_HEALTH) + \beta_{3j}(MED\_PROBS)$$

LEVEL II MODEL:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(PCPS\_POP) + \gamma_{02}(ECON\_STAT) + u_{0j}$$

Cross level interaction

$$\beta_{1j} = \gamma_{10} + \gamma_{12}(ECON\_STAT) + u_{1j}$$

Model Comparison

	OLS Model (Aggregated)	Multi-level Model
	Coefficient (Std error)	
GEN_HEALTH	.137 (.024)***	.135 (.025)***
CURR_INS	1.255 (.054)***	1.245 (.053)***
MED_PROBS	-.529 (.038)***	-.528 (.041)***
ECON_STAT	-.036 (.008)***	-.034 (.008)***
PCPS_POP	.003 (.002)	.003 (.002)
interaction	.009 (.017)	.012 (.016)

Is There Enough Power?

- Number of individuals, number of groups equally important
- Simulation studies suggest 5/group yields 90% power if large number of groups
- Freeware available for calculations at http://www.ssicentral.com/other/hlmod.htm

Finding Random Effects

- Examines data for sources of unexplained variability
- Analysis limited to groups with unique covariate patterns (unlike model for fixed effects)
- Provides estimate for variance component and significance

Random Effects
Estimation of Variance Components

Random Effect	Standard Deviation	Variance Component	df	χ^2	p
INTRCPT1, U0	0.347	0.121	514	552.8	.12
NO_INS slope, U1	0.220	0.048	516	520.5	.44
GEN_HLTH slope, U2	0.148	0.022	516	542.6	.20
MED_PROB slope, U3	0.234	0.055	516	470.9	>.50

- χ^2 statistics are based on only 517 of 623 units that had sufficient data for computation. Fixed effects and variance components are based on all of the data.

Unit-specific and Population Average Models

- Unit-specific: Intercept is expected log-odds of outcome for individual with level I and II characteristics=0 and random effect=0
- Population average: Intercept is expected log-odds of outcome for individual with level I and II characteristics=0

How To Decide If You Need Hierarchical Models

Before Data Collection

- Do the data form a hierarchy?
- Are characteristics at multiple levels of interest?
- Do you want to look at cross-level interactions?

After Data Collection

- Check intra-class correlation coefficient, looking for … ?
- What else should we look for in data?
  - Are random effects present? (Costs)
  - Can we simplify?

Applications in HSR?

- How have hierarchical models been applied in health services research?
  - Estimation of individual and contextual effects
  - Profiling of health care units
  - Meta-analysis
  - Attribution of variation in outcomes to levels of a multi-level health care system

Discussion Questions

- How might we apply hierarchical models to the health services research we are doing?
- Do hierarchical models open a previously sealed door, or is this just methodological muscle-flexing?
- Application to clinical trials? To surveys?

Need for Hierarchical Models in HSR

- Data clustered at multiple levels
  - Patients clustered by provider: physician, hospital, etc., or by study (meta-analysis)
  - Providers clustered by health care systems, market areas, geographic areas
- Cluster sizes may vary substantially
  - Can be a big problem – especially in profiling

Need for Hierarchical Models in HSR

- Covariates: patient-level, provider-level, etc.
- Estimates may be needed of:
  - Systematic and random components of variation
  - Cluster-specific measures of utilization, costs, outcomes
  - Covariate effects and decomposition of effects

Bibliography for Methods Seminar Six: Hierarchical Models By D. Litaker

•From Sage Publications:
–Kreft & DeLeeuw. *Introducing Multi-level Modeling.* (1998)
–Snijders & Bosker. *Multilevel Analysis: An introduction to basic and advanced multilevel modeling*. (1999)
–Bryk and Raudenbush. *Hierarchical Linear Models: Applications and data analysis methods.* (1992)

•Other references:
Barcikowski RS. Statistical power with group mean as the unit of analysis. *J Educ Statistics*. 1981;6(3):267-285.

Witte J et al. Multilevel modeling in Epidemiology with Glimmix. *Epidemiology* 2000;11(6):684-688

Guo G, Zhao H. Multilevel modeling for binary data. *Ann. Rev. Sociol* 2000;26:441-462

Localio AR et al. Adjustments for center in multicenter studies: An overview. *Ann. Intern Med.* 2001;135:112-123.

Sullivan LM, Dukes KA, Losina E. Tutorial in Biostatistics: An introduction to hierarchical linear modeling. *Statist. Med*. 1999;18:855-888.

Susser M. The logic in ecological: I. The logic of analysis. *Am. J. Public Health*. 1994;84(5):825-829.

Susser M. The logic in ecological: II. The logic of design. *Am. J. Public Health*. 1994;84(5):830-835.

Schwartz S. The fallacy of the ecological fallacy: The potential misuse of a concept and the consequences. *Am. J. Public Health*. 1994;84(5):819-824.

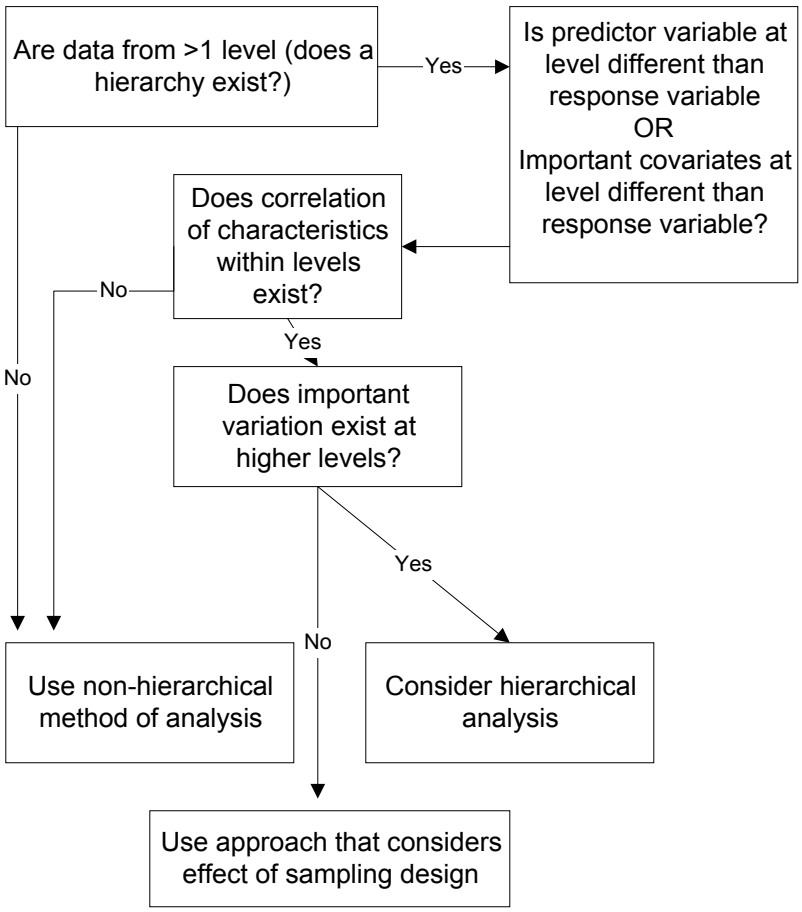
Diez-Roux, AV. Bringing context back into Epidemiology: Variables and fallacies in multilevel analysis. *Am. J. Public Health*. 1998;88(2):216-222.

Yen IH, Syme SL. The social environment and health: A discussion of the epidemiologic literature. *Annu. Rev. Public Health*. 1999;20:287-308.

•Software (free downloads)
http://www.ssicentral.com/other/download.htm (Website for HLM software developers)
http://www.mrc-bsu.cam.ac.uk/bugs/overview/contents.shtml (Website for WinBUGS software)
http://www.ioe.ac.uk/hgpersonal/papers_for_downloading.htm (Harvey Goldstein's personal website - includes papers in Adobe format discussing applications of multi-level modeling in meta-analysis, factor-analysis, and to medical data)

Table: Labels, measurement type, and definition of variables used in example

Level I characteristic	Measurement type	Definition
NO_RSOC **(dependent variable)**	Dichotomous	Identifies a usual source of care other than E.D., family member, friend (Y/N)
CURR_INS	Dichotomous	Continuously insured (Y/N)
GEN_HEALTH	Continuous (1= excellent, 5=poor)	General health status
MED_PROBS	Continuous (0-3)	Number of chronic medical conditions diagnosed by health care professional, lasting 12 months or more
Level II characteristic		
PCPS_POP	Continuous	Number of primary care physicians in zip per 10,000 residents
ECON_STAT **(independent variable)**	Continuous (z score)	Economic status of zip code; based on median family income, levels of educational attainment, distribution of employment categories



David Litaker (WrightLit@aol.com) and Thomas Love (tel3@po.cwru.edu)