Multi-level modeling

Ellie Matthay, PhD
Postdoctoral Scholar
Department of Epidemiology and Biostatistics
Ellicott.Matthay@ucsf.edu

Epi 223
May 4, 2020

Slides credit:

Kelsey Holt, ScD
Assistant Professor
Department of Family and Community Medicine
Kelsey.Holt@ucsf.edu

Multi-level structures: strict hierarchy

Image credit: SV Subramanian

Implications

1. Technical
   – Individuals recruited from the same clinic will be more alike than a random sample
   – Incorrect estimates of precision; increased risk of finding relationships and differences where none exist
     • Robust standard errors also account for clustering but don't allow for addressing substantive questions
2. Substantive
   – Are patients in a clinic influenced by structural/policy factors at that clinic?
   – Is the variance in patient outcomes more attributable to differences at the individual or clinic level?

Multilevel analysis

- Simultaneous examination of effects of group- and individual-level factors on individual-level outcomes
- Examine both between group and within group variability

Multilevel models appropriate when…

1. Observations are correlated/clustered;
2. Causal processes are thought to operate at multiple levels; and/or
3. Intrinsic interest in modeling heterogeneity in the population (versus just controlling for higher level processes)

Motivations

1. Understand when variation at a cluster level (e.g., school, clinic, neighborhood) is due to composition versus context

Composition versus context

- Let’s say smoking rates are twice as high in San Francisco compared to Atlanta…
- Compositional effect: the difference people make to places
  - Say being young doubles the probability of smoking and San Francisco has twice the young people
- Contextual effect: the difference places make to people
  - Probability of smoking doubles when you move to San Francisco
- Multi-level models parse the variance in the outcome by level

Motivations

1. Understand when variation at a cluster level (e.g., school, clinic, neighborhood) is due to composition versus context
2. Model the association between cluster-level variables (i.e., attributes of schools, clinics, neighborhoods) and individual outcomes
3. Model the interaction between cluster-level attributes and individual-level attributes in relation to individual-level outcomes
4. Account for “nuisance” correlations among outcomes of individuals from the same cluster

Multi-level structures: cross-classification

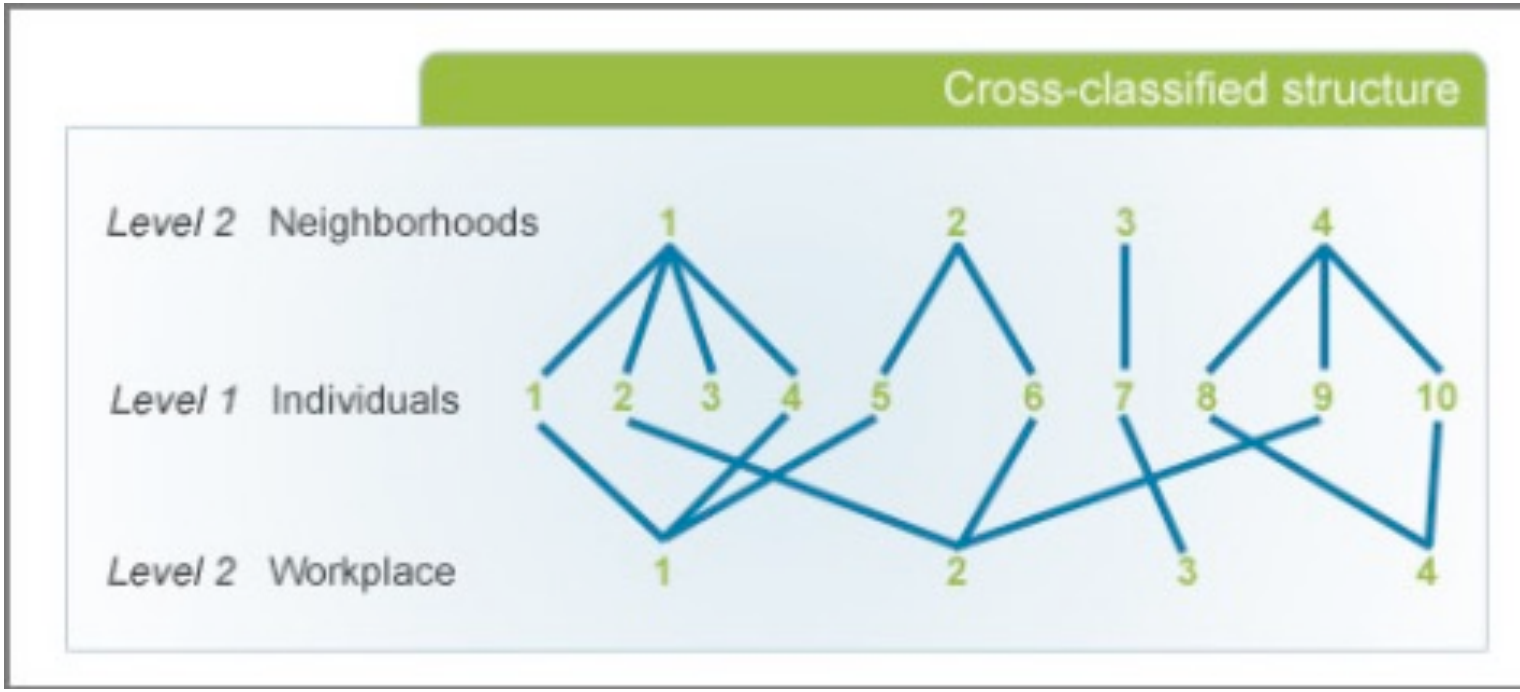


Image credit: SV Subramanian

Multi-level structures: repeated measures (longitudinal studies)

Examples from my research

Exposure	Outcome
Exposure to community violence	Stress-responsive illnesses and injuries
Malaria control policies	Child mortality
Local alcohol and cannabis policies	Self-directed and interpersonal violence
City firearm violence prevention program	Firearm and nonfirearm violent injuries
Gun shows	Firearm injuries
Alcohol outlet density	Binge drinking

Multi-level modeling

Ellie Matthay, PhD
Postdoctoral Scholar
Department of Epidemiology and Biostatistics
Ellicott.Matthay@ucsf.edu

Epi 223
May 4, 2020

Diez-Roux article

1. Multi-level modeling approaches aim to analyze contextual effects beyond individual characteristics. Briefly describe the benefits of conducting a multilevel analysis for health disparities research areas.
2. Draw from your own research or interest area to give an example of macro-level factors or group-level properties that influence individual health.
3. Define the ecological and atomistic fallacies and explain how these fallacies differ from psychologistic and sociologistic fallacies.

Subramanian article

1. Drawing on your area of work, what is an example of a potential individualistic fallacy and how could a multi-level modeling approach help lead to more meaningful analysis of individual level relationships?
2. Briefly describe the difference between fixed effects and random effects in the models in this paper. In broad stokes, in what circumstances would you use one or the other to examine a higher level variable?
3. What is a main takeaway you got from reading the authors' analysis of Robinson's original paper in historical context?

Gausman article

1. On page 3, discussing Table 2, the authors point out that "the fact that the variance parameters remain significantly different from zero indicates that the geographic variation at these levels is not entirely determined by differences in population composition." Describe what the authors would have missed if they had not conducted a multi level model with random effects for community and country.
2. Briefly describe the main findings and their implications for interventions to address childbearing before age 16 in low and middle income countries. How does the examination by country in Figure 1 change your answer, if at all?

Extra slides

In terms of equations…

Levels versus variables

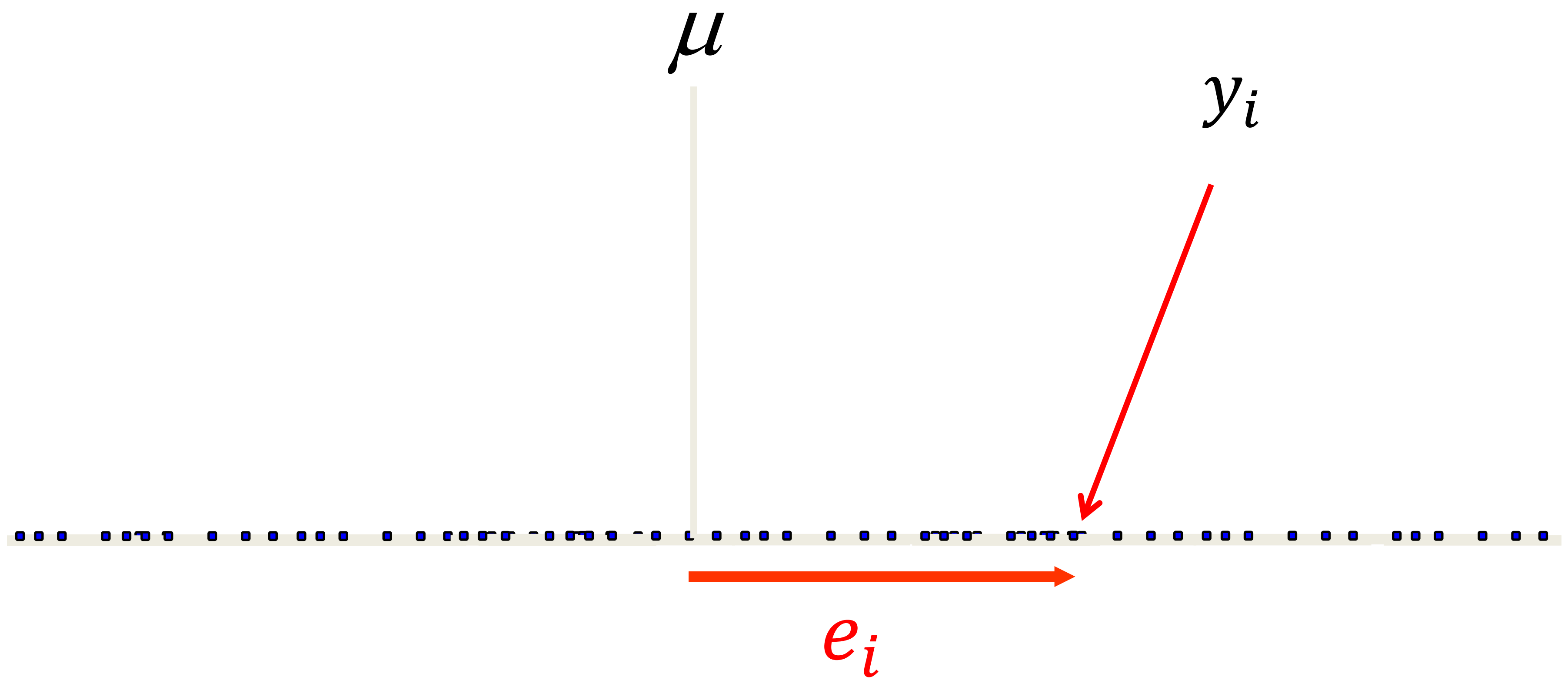
- A *level* is a unit of analysis
  - Clinics=population of units from which we have taken a random sample
- A *variable* is an attribute of a level
  - Gender is an attribute of individuals
  - Size is an attribute of clinics
  - Gender≠sample of all possible gender categories

Single Level Models

$$\underbrace{y_i}_{\textbf{Observed}} = \underbrace{\mu}_{\textbf{Fixed}} + \underbrace{e_i}_{\textbf{Random}}$$

Slide credit: Chuck Huber

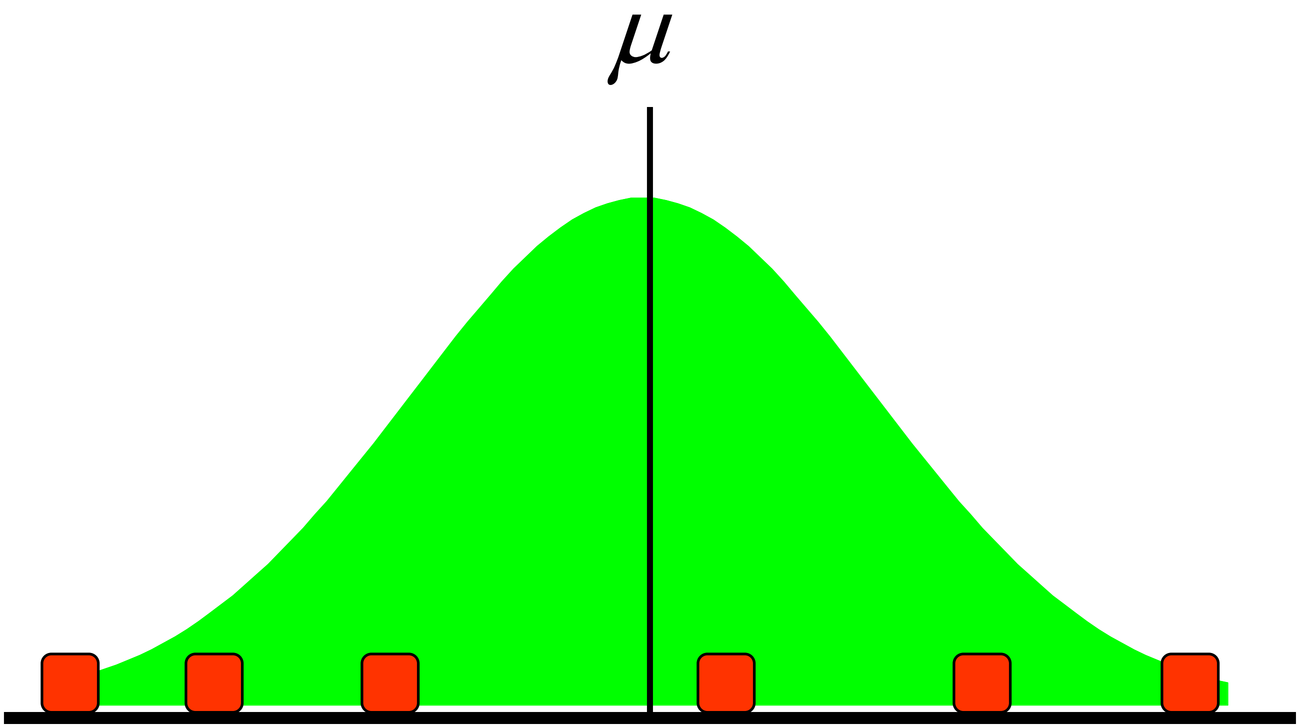
Single Level Models



$$y_i = \mu + e_i$$

Slide credit: Chuck Huber

$$y_i = \mu + e_i$$



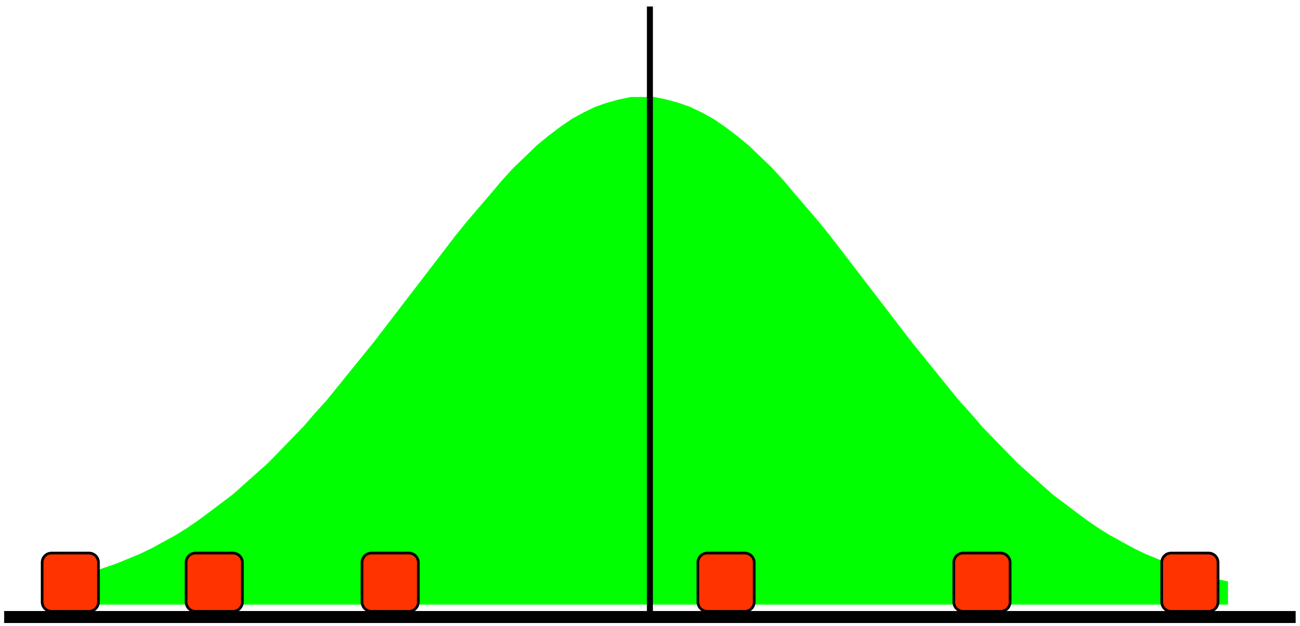
$$e_i \sim N(0, \sigma^2)$$

Slide credit: Chuck Huber

Levels can become variables

- When a group (e.g., clinic, state) is in the fixed part of a regression equation, as a variable:
  – Generalization to population of schools is not possible
- When group is treated as a level in the random part of a regression equation
  – Generalization possible to population of schools, in addition to identifying specific school effects

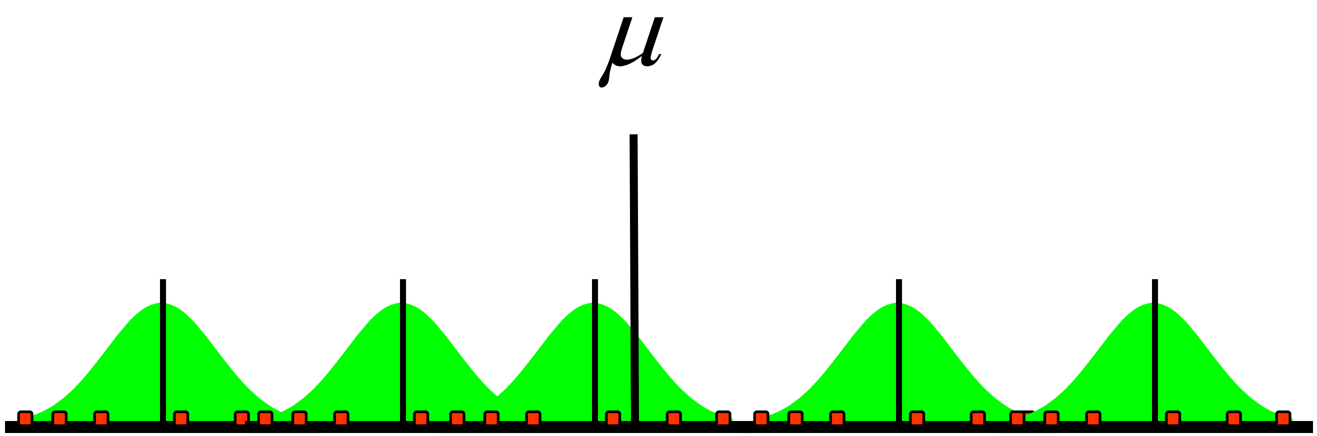
Two Level Models



Patient observations within physician practice will vary about each physician's mean.

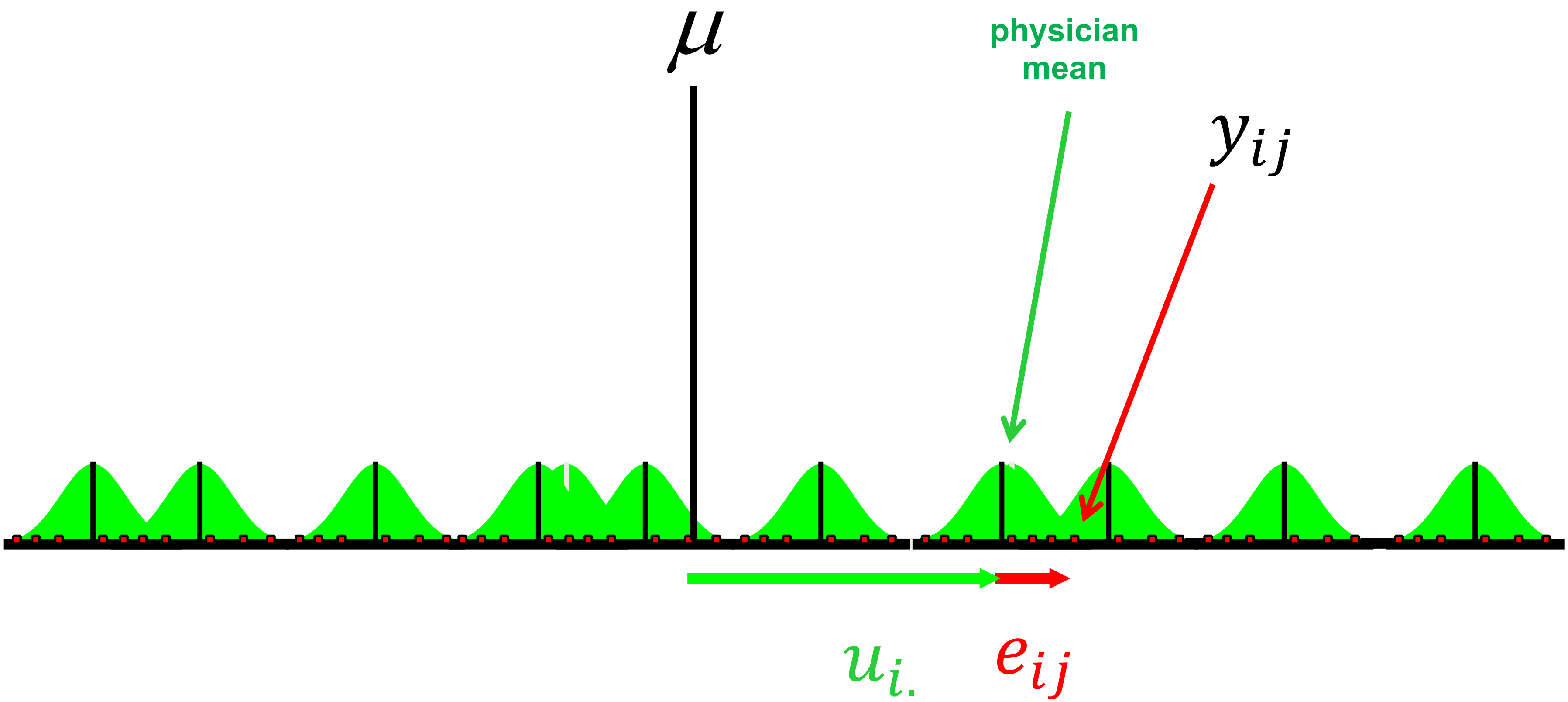
Slide credit: Chuck Huber

Two Level Models



Physicians' means will vary about the grand mean.

Slide credit: Chuck Huber



$$y_{ij} = \mu + u_{i.} + e_{ij}$$

Slide credit: Chuck Huber

Two Level Models

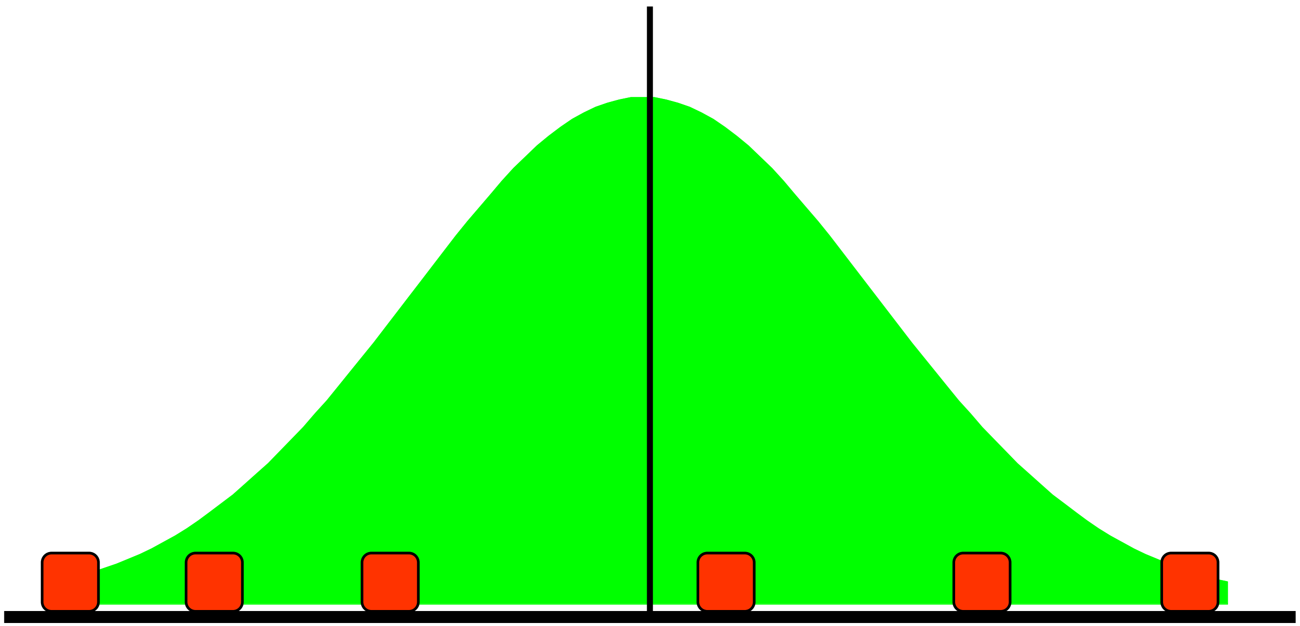
$$\underbrace{y_{ij}}_{\textbf{Observed}} = \underbrace{\mu}_{\textbf{Fixed}} + \underbrace{u_{i.} + e_{ij}}_{\textbf{Random}}$$

$$u_{i.} \sim N(0, \tau^2)$$

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

Slide credit: Chuck Huber

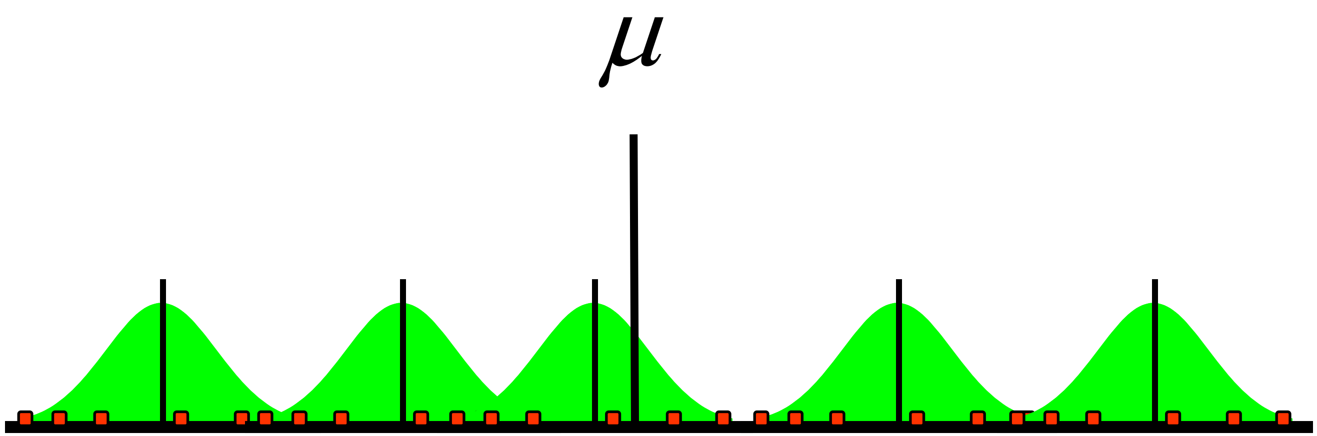
Longitudinal Models



Repeated observations within patients will vary about each patient's mean.

Slide credit: Chuck Huber

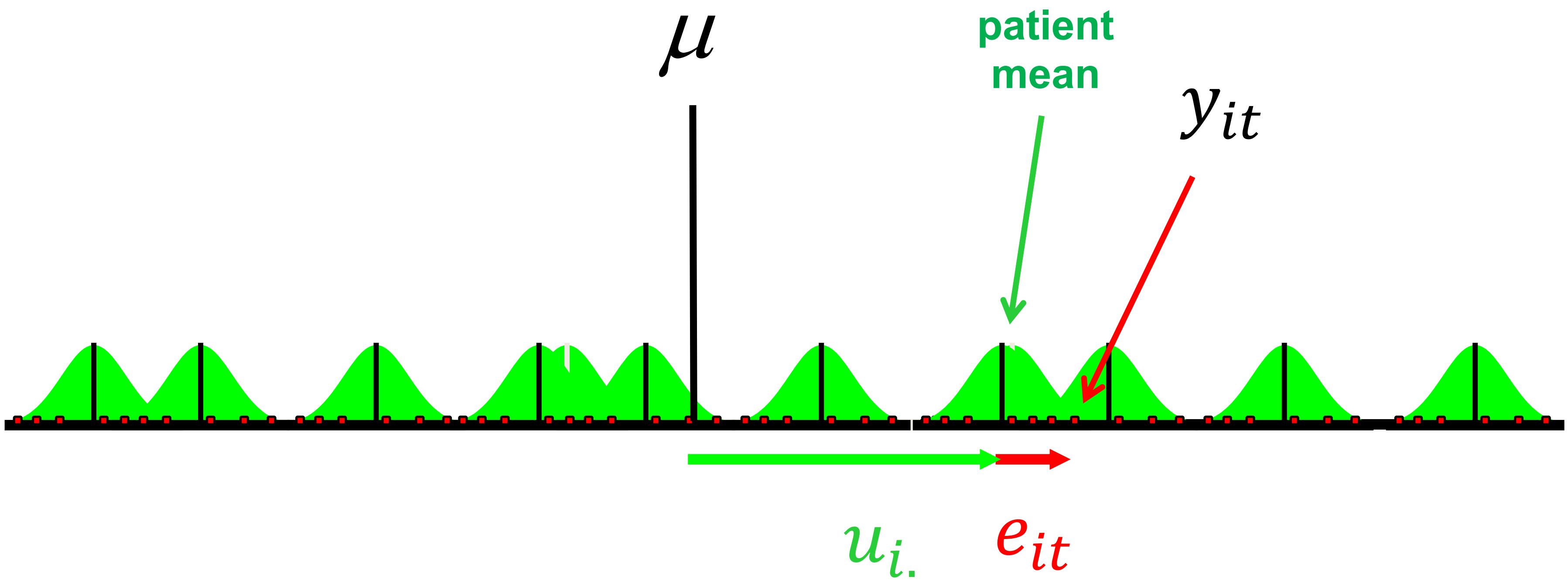
Longitudinal Models



Patients' means will vary about the grand mean.

Slide credit: Chuck Huber

Longitudinal Models



$$y_{it} = \mu + u_{i.} + e_{it}$$

Slide credit: Chuck Huber

Fixed versus random effects

- Do female patients report better experience with providers than male patients?
  - *Fixed part of model*
- Do patients in clinics with QI processes report better experiences?
  - *Fixed part of model*
  - If multi-level model, can draw inferences about the population of clinics
- Do females in clinics without QI processes report worse experiences?
  - Cross-level interaction
  - *Fixed part of model*
- How much of the variability in patient experience scores is due to differences between clinics versus differences between individuals?
  - *Random part of model*
- Is clinic X different from other clinics in the sample in its effect on patient experience scores?
  - *Fixed part of model*
- Are females more or less variable than males in their reported experience with providers?
  - *Random part of model*
- Is the gender gap in patient experience different across clinics?
  - *Random part of model*

		Exposure	
		x (measured at individual level)	X (measured at ecological level)
Outcome	y (measured at individual level)	{y, x} Traditional risk factor study	{y, X} Multilevel study
	Y (measured at ecological level)	{Y, x}*	{Y, X} Ecological study

Figure 5 Typology of Studies.[40] Asterisk represents that