My PHD Research
Margaret Rolfe

External Student QUT

Principal Supervisor
Professor Kerrie Mengersen
Commenced September 2006

Research Topic:

- Investigating hierarchical Bayesian latent variable models with respect to longitudinal data
- focusing on the impact and choice of priors and
- how the outcomes can assist future data collection issues such as experiment design and sampling plans,
- with application in the areas of biostatistics and statistical genetics.

Longitudinal Data

- Longitudinal data or repeated measures data has measurements of a response taken at several points in time on an individual subject.
- This data is hierarchical or multilevel in nature as response (over time) is clustered in or nested in subjects.
- The repeated measurements within a subject are expected to be more highly correlated than correlations between subjects
- Assumption of independence of observations fails

Latent Variables

Latent variables are unobserved variables and can be presented in a number of different forms

- As underlying constructs (often hypothesised) for complex multivariate data.
- As underlying parameters for a growth process
- As mixtures to represent subpopulations

Bayesian Analysis

The Bayesian approach to data analysis is derived from the work of Sir Thomas Bayes and utilises prior information and data to obtain a posterior distribution where

$$\text{posterior distribution} \propto \text{likelihood} \times \text{prior distribution}$$

The relative influence of the prior and data depends on the relative strength of both. Priors can be informative or non-informative of vague; and a large data sample would generally predominate the situation.

Posterior distributions of all parameters can be obtained.

Case Study of Longitudinal Data: Wesley Research Institute Cognition Study

Study Aim: To assess the impact of adjuvant chemotherapy on cognitive functioning in women with breast cancer.

Design: Longitudinal, prospective, repeated measures design with assessments conducted at four time points: before chemotherapy and at 1,6,18 months post-chemotherapy.

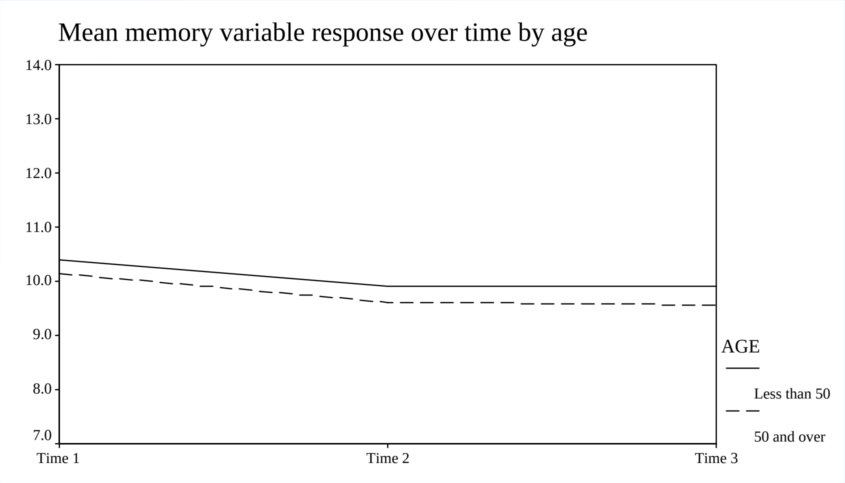
Participants:
Women aged 18-70 years;
Recruited from multiple hospitals throughout Brisbane;
Have histologically proven breast cancer treated by surgery and chemotherapy;
Proficiency in English plus other exclusion criteria

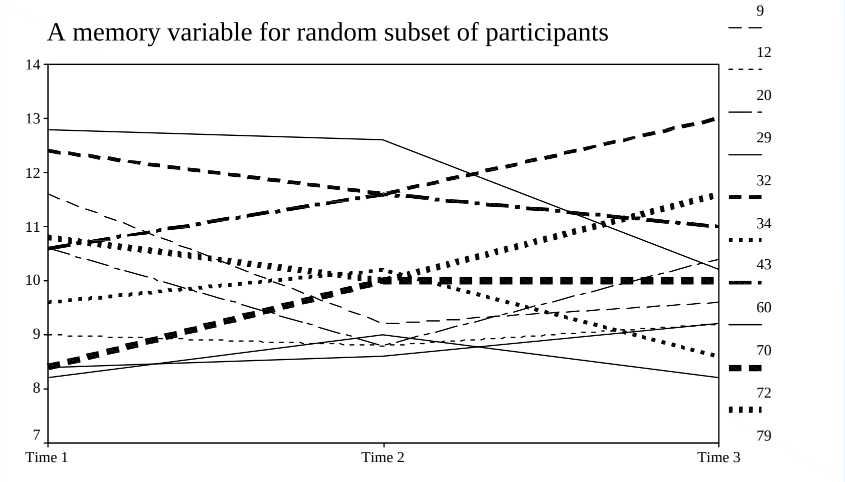
Outcome Measurements:

- Battery of 19 tests to measure differing aspects of cognition, many of these tests having numerous sub-items.
- Hospital Anxiety and Depression Scale (HADS) to measure mood.
- Subscales of Functional Assessment of Cancer Treatment Scale (general, breast, fatigue) to measure quality of life.

Cognitive Domain	Test
Verbal Learning and Memory	■ Rey Auditory Verbal Learning Test (RAVLT)
Visual Memory	■ WMS-III Visual Reproduction
Attention	■ WAIS-III Digit Span Forwards ■ Delis-Kaplan Executive Function System (DKEFS) Trail-Making Test – Letter Sequencing
Processing Speed	■ Stroop Word-Reading and Colour-Naming Trials ■ SymbolDigit Modalities Test - Oral Version ■ Speed and Capacity of Language Processing Test (SCOLP)
Psychomotor Speed	■ Purdue Pegboard ■ DKEFS Trail-Making Test – Motor Speed
Executive Functioning	
Attentional Switching	■ DKEFS Letter-Number Switching ■ Test of Everyday Attention (TEA) Visual Elevator
Working Memory	■ WAIS-III Digit Span Backwards
Verbal Fluency	■ FAS words ■ Animals
Inhibition of interference	■ Stroop Interference Trial
Multitasking (Dual Task)	■ TEA Telephone Search While Counting
Complex Reasoning/Planning	■ WMS-III Matrix Reasoning ■ DKEFS Card Sort Test ■ Tower of London - Dx

Patterns of Responses





Possible Methods for Analysis in both Bayesian and Non Bayesian frameworks

- Repeated Measures Analysis of Variance
- Multilevel or hierarchical Linear Regression
- Structural equation modelling approaches including
  - Latent growth curve models
  - Simplex or autoregressive models
  - Multiwave models on scores or differences of scores
- Mixture models
  - Latent class growth models
  - General growth mixture models
- General latent trait models

Multilevel Models in MLWIN

- Multilevel models are mixed effects models in that they involve both fixed and random components from hierarchically structured data.
- For repeated measures data the hierarchical structure is occasion nested in subject, with subjects as the level 2 units denoted by subscript i, and time occasions within subject as the level 1 units (subscript t)

For response y_{ti} for subject i on occasion t

$time_{ti}$ for subject i on occasion t eg 1,2,3 or -1,0,1 etc

β_0 , β_1 regression parameters

u_{0i} random intercept, e_{0ti} residual error term

σ^2_{u0} subject level variance, σ^2_{e0} residual variance

$$y_{ti} = \beta_0 + u_{0i} + e_{0ti}$$

Variance Component Model

No Fixed Effects Null Model

$$y_{ti} = \beta_0 + \beta_1 time_{ti} + u_{0i} + e_{0ti}$$

Random Intercept Model

Linear in Time

$$u_{0i} \sim N(0, \sigma^2_{u0}), \quad e_{0ti} \sim N(0, \sigma^2_{e0})$$

Multilevel Models in MLWIN

- Random effects (random intercept random slope) model for linear time fits individual linear regression for each participant and obtains variances for the intercept , slope and the intercept slope covariance

Random Effects Model

$$y_{ti} = \beta_0 + \beta_1 time_{ti} + u_{0i} + u_{1i} time_{ti} + +e_{0ti}$$

$$\begin{pmatrix} u_{0i} \\ u_{1i} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u0}^2 & \sigma_{u01} \\ \sigma_{u01} & \sigma_{u1}^2 \end{pmatrix}\right]$$

$$e_{0ti} \sim N(0, \sigma_{e0}^2)$$

Multilevel Models in MLWIN

Results of Multilevel analyses from models described previously

	Variance Component	Random Intercept Fixed Slope	Random Effects Random intercept & slope
	Null	Linear time	Linear time
Fixed			
Intercept	9.947 (0.135)	9.947 (0.135)	9.947 (0.135)
Linear Time		-0.264 (0.079)	-0.264 (0.080)
Random			
σ_{u0}^2	1.341 (0.257)	1.362 (0.257)	$\begin{pmatrix} 1.37[0.26] & \\ 0.20[0.11] & 0.03[0.12] \end{pmatrix}$
σ_e^2	1.251 (0.128)	1.188 (0.121)	1.158 (0.167)
Deviance	1018.9	1007.9	1004.3
Change in deviance		9.0 df=1	3.6 df=2
Sign of change		**0.0027**	0.165 ns improvement

Bayesian Multilevel Models in MLWIN

Results of Bayesian MCMC random intercept model

For 500 burnin and chain of length 5000, and with uniform prior distributions

Bayesian Deviance Information Criterion (DIC)

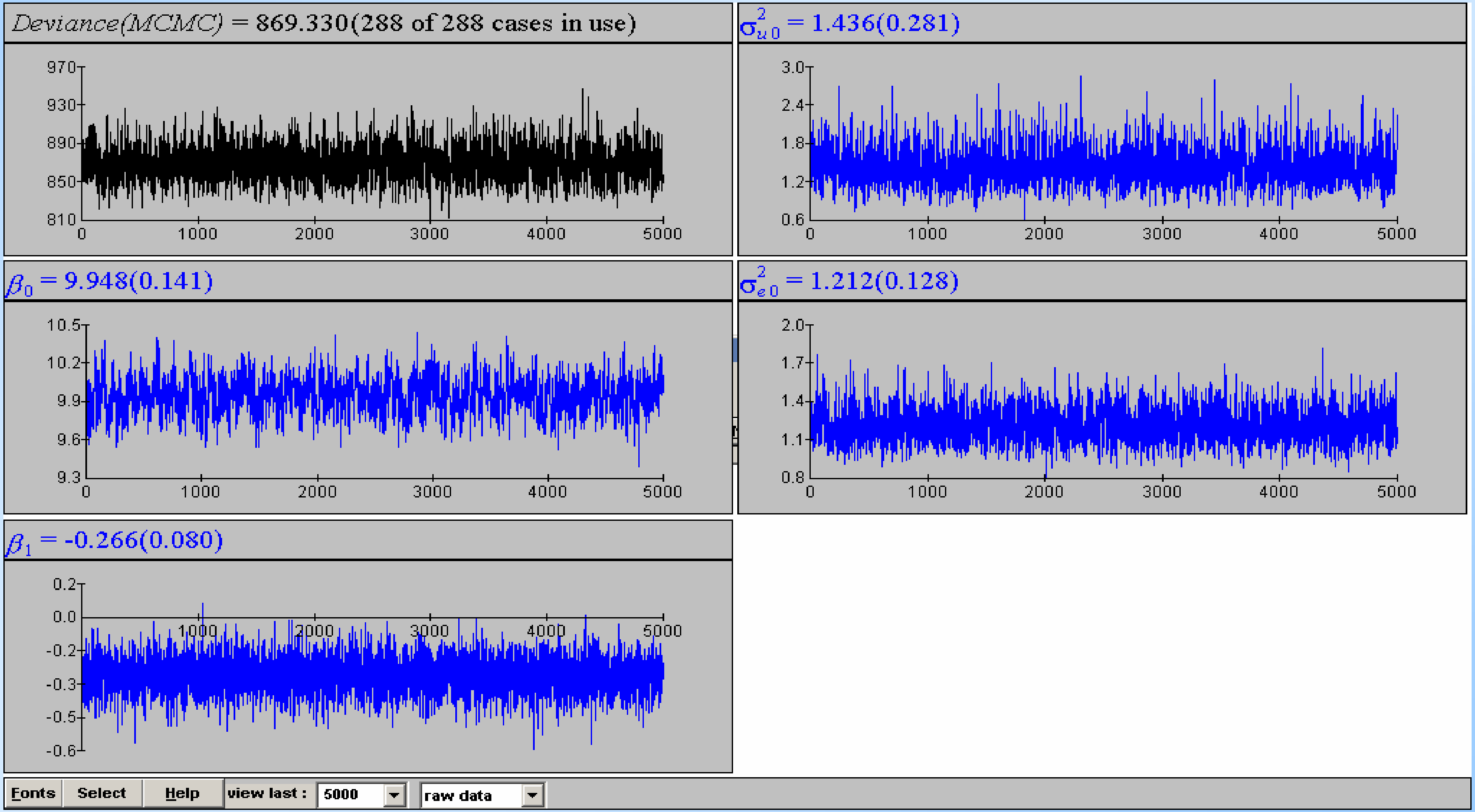
Dbar	D(thetabar)	pD	DIC
869.33	792.55	76.78	946.11

Parameter Estimates

Parameter	Posterior Mean	SD	Mode
β_0	9.948 (.0057)	0.141	1.381
β_1	-0.266 (0.0011)	0.080	-0.263
σ^2_{u0}	1.436 (.0063)	0.281	1.381
σ^2_{e0}	1.212 (.0026)	0.128	1.192

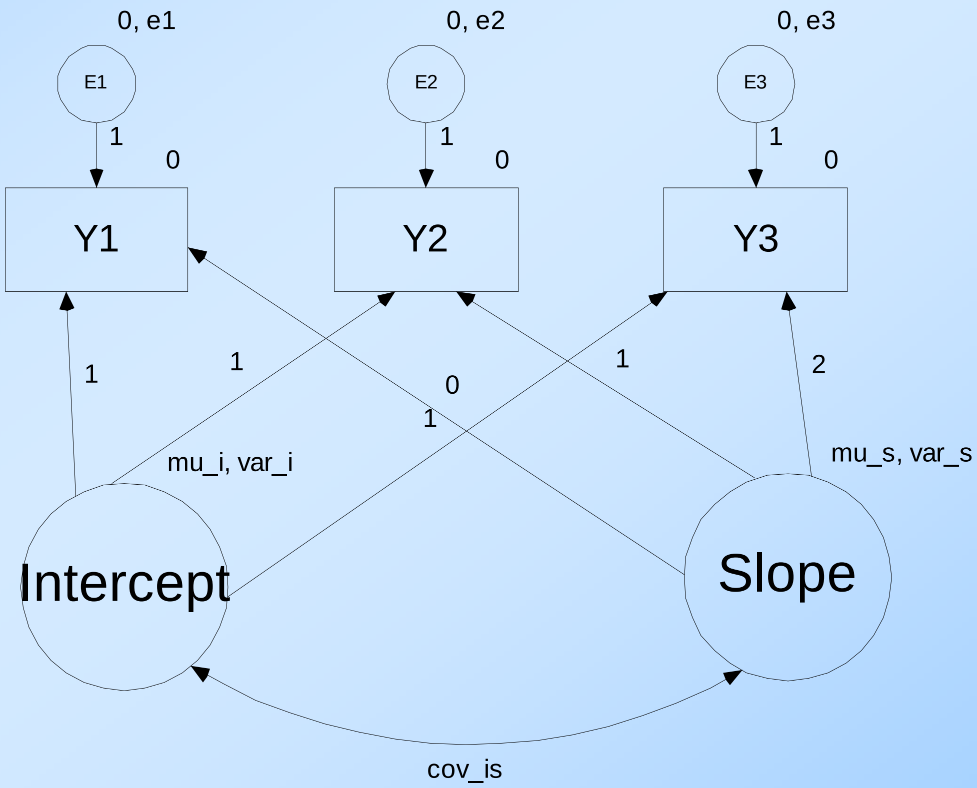
Bayesian Multilevel Models in MLWIN

Trajectories of Parameters by MCMC iterations



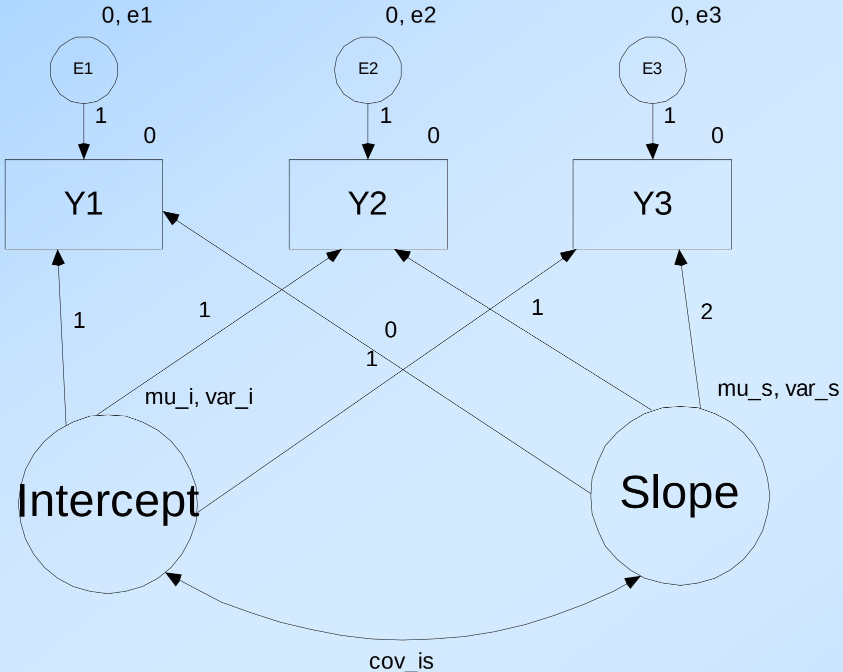
Latent Growth Curve Model

Trajectories of the observed data are driven by an underlying subject-level latent growth process, described by intercept and linear slope latent growth variables. Means are estimated and paths are fixed to 1 for intercept latent variable and 0,1,2.. for linear slope latent variable. Errors (E1,E2,E3) on measured variables can covary. Latent growth variables can also covary

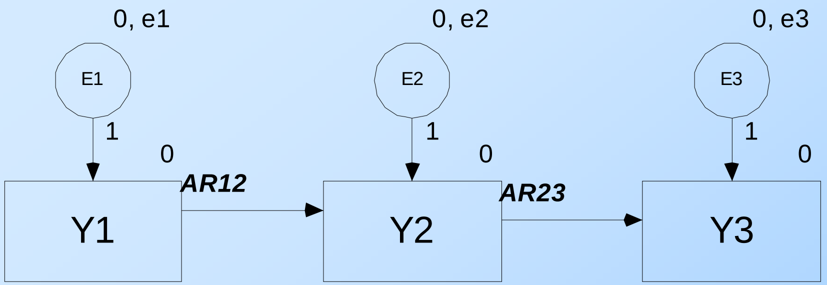


Some SEM Models

Latent Growth Curve Model

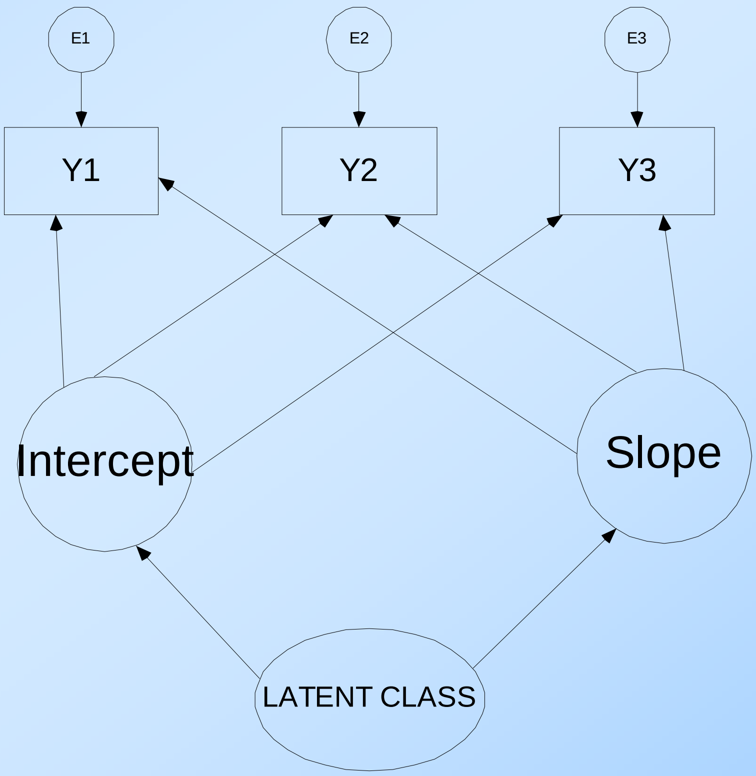


Simplex or autoregressive model



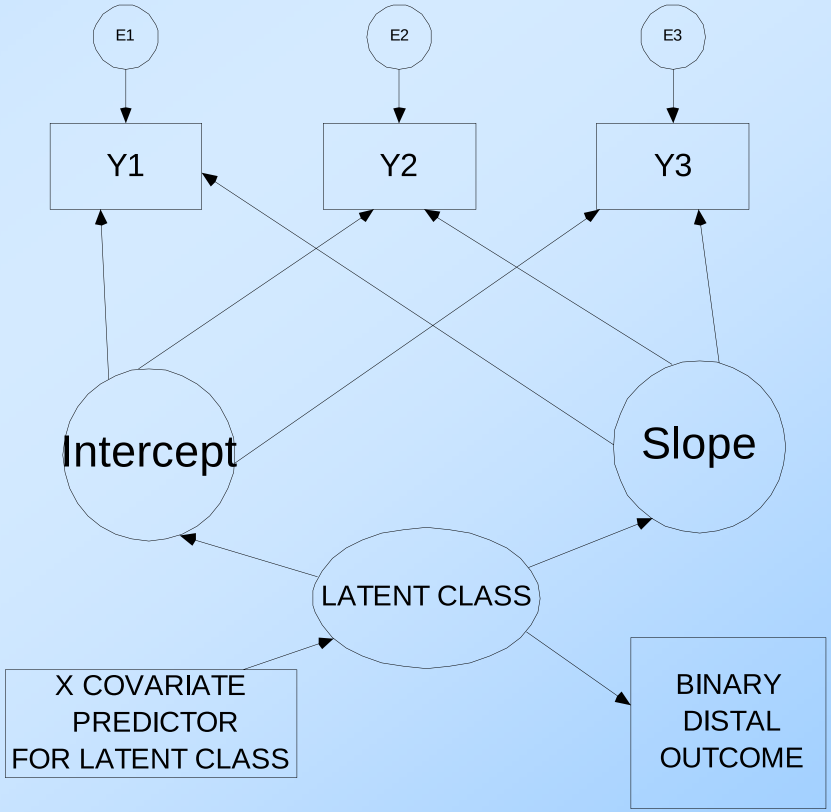
Latent Class Growth Model

Trajectories of the observed data are driven by an underlying subject-level latent growth process. The latent growth process depends on which of the K latent classes the subject belongs. That is the classes define different trends over time, means of growth factors change over classes, random effects are not generally included.



Growth Mixture Modeling

Growth mixture models have individual differences captured by random effects as seen in multilevel modeling and also have differences described by latent trajectory classes. These models can be extended to include covariates which influence classes and to have the classes predict a distal (binary) outcome and are known as general growth mixture models (GGMM).



Future Threat for the Ashes?