Modeling with Possible Non-Ignorable Dropout in Longitudinal Studies

Bengt Muthén
bmuthen@ucla.edu
www.statmodel.com

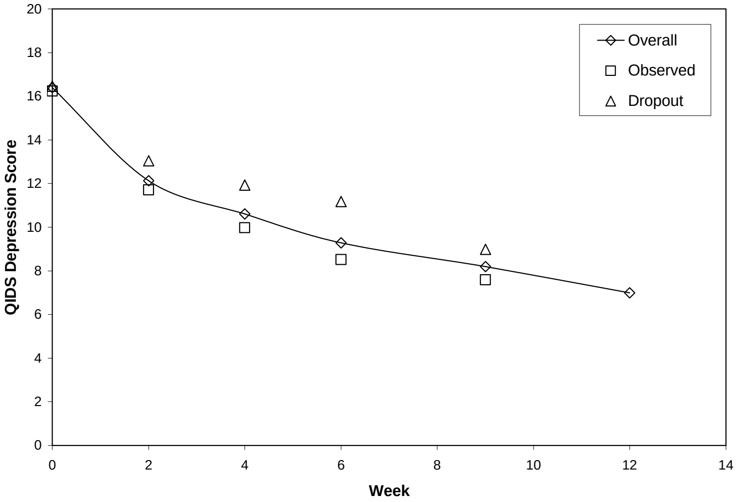
Presentation at Inserm Atelier de formation 205,
Saint-Raphael, June 3, 2010

Overview

- Overview of key missing data models presented in the general latent variable framework of Mplus
- Current approaches: Shared set of latent variables for the outcome and missing data processes
- New approach: Disentangling the outcome process and the missing data process

Example: Longitudinal Data From An Antidepressant Trial (STAR*D) $n = 4041$

Subjects treated with citalopram (Level 1). No placebo group



MCAR, MAR, And NMAR

- y: outcome vector - but y is not the only data
- m: missing data indicator vector
- MCAR

$$P(m_i | y_i^{obs}, y_i^{mis}) = P(m_i) \tag{1}$$

- MAR

$$P(m_i | y_i^{obs}, y_i^{mis}) = P(m_i | y_i^{obs}) \tag{2}$$

$$ML\,[y, m] = ML\,[y] \tag{3}$$

- NMAR (non-ignorable or informative missingness):
  None of the above. Missingness predicted by latent variables

NMAR Modeling: Joint Analysis Of The Outcome And Missing Data Processes

- y: outcome vector
- m: missing data indicator vector
  - Focus has typically been on dropout with m replaced by dropout indicators d. Often, d's are scored as discrete-time survival (event history) indicators. For example, dropout after 3rd visit: 0 0 0 1 999 999. But d's can also be dropout dummy variables.
  - Mplus creates m and d using Data Missing
- c: latent class variable

$$
\begin{aligned}
[y,m] &= [y]\,[m|y], \;\; \textit{Selection modeling} \qquad (4)\\
&= [m]\,[y|m], \;\; \textit{Pattern}-\textit{mixture modeling} \qquad (5)\\
&= \sum_c [c]\,[m|c]\,[y|c] \;\; \textit{Shared}-\textit{parameter modeling} \qquad (6)
\end{aligned}
$$

- Each NMAR model involves untestable assumptions - comparing results from several models gives a sensitivity analysis
- Overviews in Albert and Follman (2008), Little (2008)

The unifying theme of latent variables

Continuous Latent Variables

- Factors
- Random effects
- Frailties, liabilities
- Variance components
- Missing data
- Bayesian parameter priors

Categorical Latent Variables

- Latent classes
- Clusters
- Finite mixtures
- Missing data

The Mplus general latent variable modeling framework

- Mplus support from National Institutes of Health Small Business Innovation Research contracts and grants

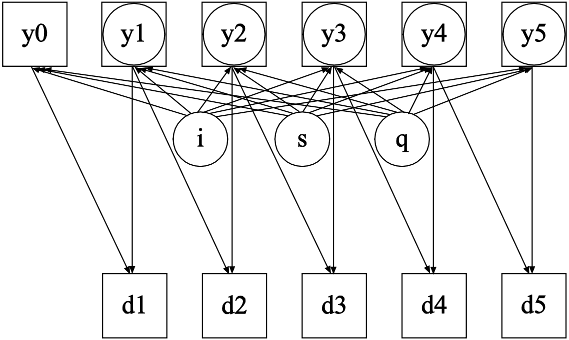
Several programs in one, fully integrated:

- Exploratory factor analysis
- Structural equation modeling
- Item response theory analysis
- Latent class analysis
- Latent transition analysis (Hidden Markov modeling)
- Survival analysis
- Growth modeling
- Multilevel analysis
- Complex survey data analysis
- Bayesian analysis using MCMC
- Monte Carlo simulation

NMAR analysis in Mplus

- A large set of NMAR models can be estimated in the general latent variable modeling framework of Mplus (www.statmodel.com)
- Maximum-likelihood estimation using EM in combination with FS and QN
- Number of classes informed by BIC and bootstrapped likelihood-ratio test
- Bayesian MCMC analysis available as well using DIC and PPC

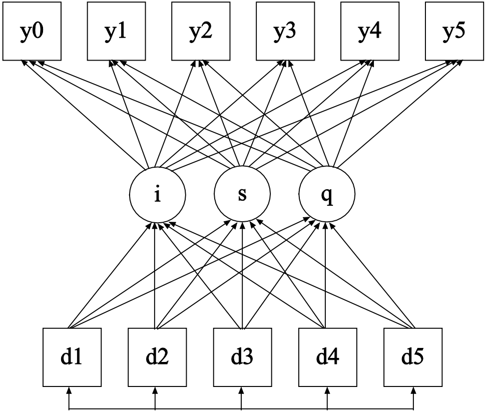
Diggle-Kenward Selection Model
(d's: dropout indicators in line with discrete-time survival)



Diggle-Kenward (1994) in Applied Statistics

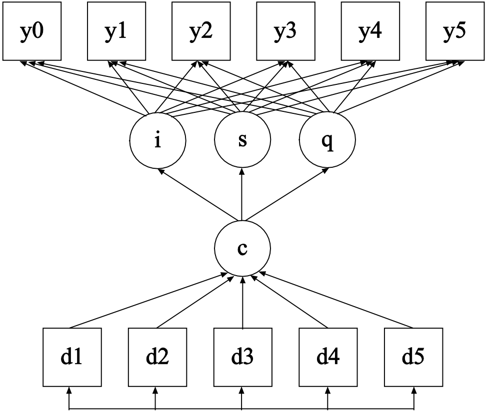
Logistic regression of d_t on observed y_{t-1} and observed/unobserved y_t

Pattern-Mixture Model
(d's: dropout pattern dummies)



Hedeker-Gibbons (1997) in Psychological Methods,
Demirtas-Schafer (2003) in Statistics in Medicine

Roy-Dantan Latent Class Dropout Model (d's: dropout pattern dummies)



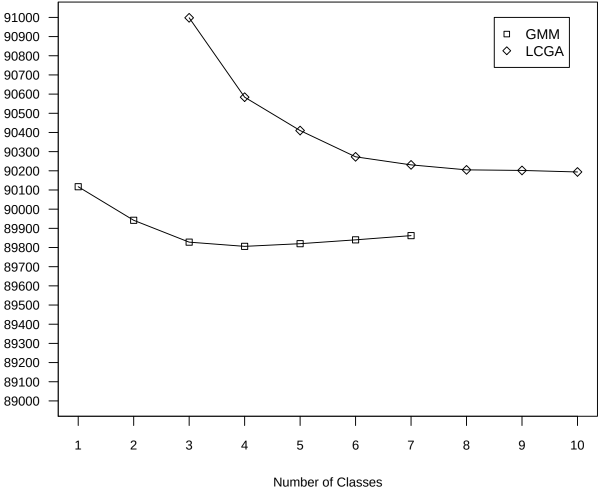
Roy (2003) Biometrics: c on dropout time

Dantan et al (2008) Int'l J. of Biostat: c on dropout dummies

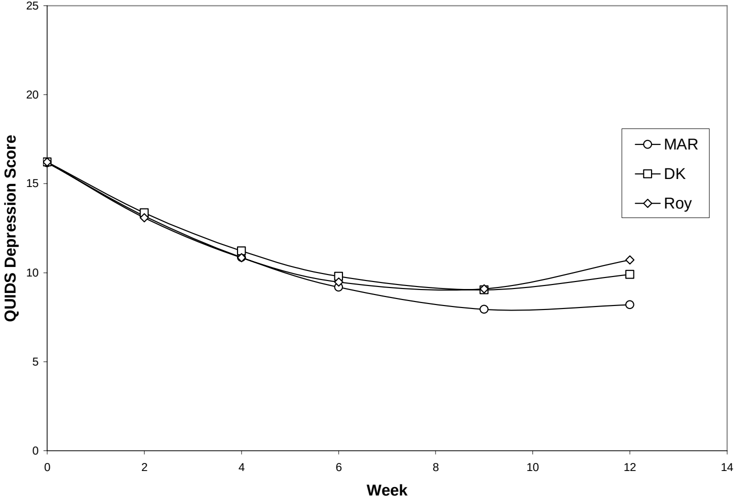
Summary Of Mixture Modeling Of STAR*D Data Using Dropout Pattern Dummies

Model	Loglikelihood	#par.s	BIC
Pattern-mixture	-44946	27	90117
Roy 2c	-44871	24	89942
Roy 3c	-44777	33	89828
Roy 4c	**-44728**	**42**	**89806**
Roy 5c	-44698	51	89820

An aside: BIC curves for Roy-GMM and Roy-LCGA



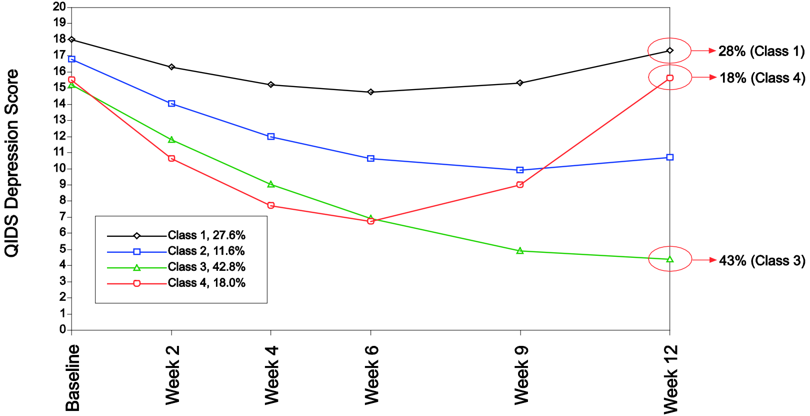
Depression Mean Curves Estimated Under MAR, Diggle-Kenward NMAR, And Roy NMAR



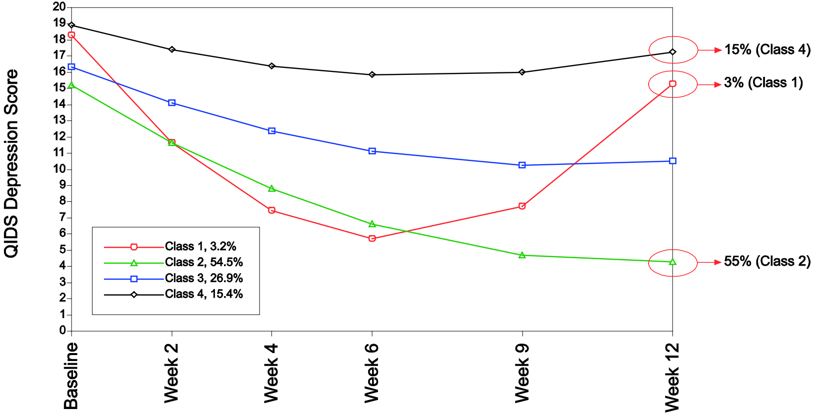
Disadvantages Of Pattern-Mixture And Roy Latent Class Dropout Modeling

- For both approaches the intention is to mix the parameter estimates over the patterns/classes to obtain an overall estimated growth curve
  - This mixing may hide substantively interesting trajectory classes

4-Class Roy-Dantan Latent Class Dropout Model



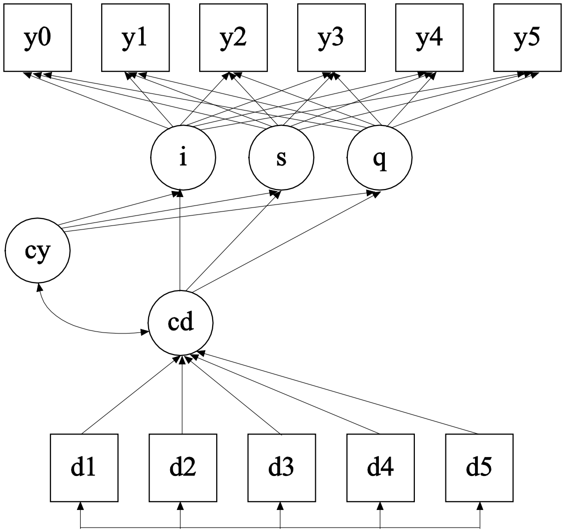
4-Class MAR Model (Y Outcomes Only)



Disadvantages Of Pattern-Mixture And Roy Latent Class Dropout Modeling

- Roy latent class dropout modeling forms classes based not only on the relationship between dropout and outcomes, but also based on the development of the outcomes over time
  - This may confound dropout classes with trajectory classes

Muthén-Roy Model Using Two Latent Class Variables



Model Comparisons

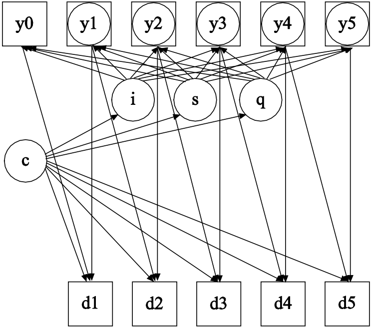
Model	Loglikelihood	#par.s	BIC
Pattern-mixture	-44946	27	90117
Roy 4c	-44728	42	89806
Muthen 4c-Roy 2c	**-44662**	**44**	**89689**

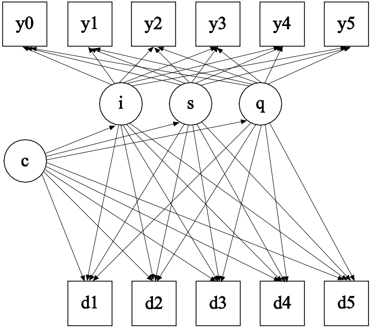
Comparing Trajectory Class Percentages Across Models

Adding dropout information gives a less favorable conclusion regarding drug response than assuming MAR

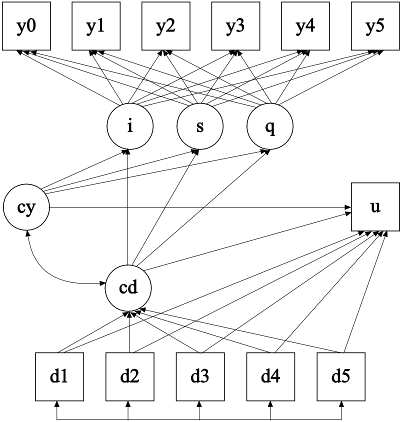
Model	Response class	Temporary response class	Non-response class
MAR 4 classes	55 %	3 %	15 %
NMAR models:			
Roy 4 classes	43 %	18 %	28 %
Muthen 4c-Roy 2c	32 %	19 %	32 %

Selection models with latent classes: Generalized Diggle-Kenward and Beunckens et al.





Muthén-Roy model extended to include an ultimate outcome



Technical Papers: See Papers at www.statmodel.com

- Muthén & Shedden (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. **Biometrics**.
- Muthén (2002). Beyond SEM: General latent variable modeling. **Behaviormetrika**.
- Muthén et al. (2002). General growth mixture modeling for randomized preventive interventions. **Biostatistics**.
- Muthén & Asparouhov (2009). Growth mixture modeling: Analysis with non-Gaussian random effects. In Fitzmaurice et al. (eds.), **Longitudinal Data Analysis**. Chapman & Hall.
- Muthén & Brown (2009). Estimating drug effects in the presence of placebo response: Causal inference using growth mixture modeling. **Statistics in Medicine**.
- Muthén et al. (2010). Growth modeling with non-ignorable dropout: Alternative analyses of the STAR*D antidepressant trial. Under review.

Applied Papers: See Papers at www.statmodel.com

- Kreuter & Muthén (2008). Analyzing criminal trajectory profiles: Bridging multilevel and group-based approaches using growth mixture modeling. Journal of Quantitative Criminology, 24, 1-31.
- Muthén (2001). Latent variable mixture modeling. In Marcoulides & Schumacker (eds.), New Developments and Techniques in Structural Equation Modeling (pp. 1-33). Lawrence Erlbaum Associates.
- Muthén (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (ed.), Handbook of quantitative methodology for the social sciences (pp. 345-368). Newbury Park, CA: Sage Publications.
- Muthén (2008). Latent variable hybrids: Overview of old and new models. In Hancock & Samuelsen (Eds.), Advances in latent variable mixture models, pp. 1-24. Charlotte, NC: Information Age Publishing, Inc.
- Muthén & Muthén (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. Alcoholism: Clinical and Experimental Research, 24, 882-891.