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 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
Subjects treated with citalopram (Level 1). No placebo group
y: outcome vector - but y is not the only data
m: missing data indicator vector

MCAR

\[
P(m_i | y_{i, obs}^o, y_{i, mis}^m) = P(m_i) \tag{1}
\]

MAR

\[
P(m_i | y_{i, obs}^o, y_{i, mis}^m) = P(m_i | y_{i, obs}^o) \tag{2}
\]

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

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

\[ [y, m] = [y] [m|y], \quad \text{Selection modeling} \]  \hspace{1cm} (4)

\[ = [m] [y|m], \quad \text{Pattern – mixture modeling} \] \hspace{1cm} (5)

\[ = \sum_c [c] [m|c] [y|c] \quad \text{Shared – parameter modeling} \] \hspace{1cm} (6)

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
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)

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

Roy (2003) Biometrics: c on dropout time
Summary Of Mixture Modeling Of STAR*D Data Using Dropout Pattern Dummies

<table>
<thead>
<tr>
<th>Model</th>
<th>Loglikelihood</th>
<th>#par.s</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pattern-mixture</td>
<td>-44946</td>
<td>27</td>
<td>90117</td>
</tr>
<tr>
<td>Roy 2c</td>
<td>-44871</td>
<td>24</td>
<td>89942</td>
</tr>
<tr>
<td>Roy 3c</td>
<td>-44777</td>
<td>33</td>
<td>89828</td>
</tr>
<tr>
<td><strong>Roy 4c</strong></td>
<td><strong>-44728</strong></td>
<td><strong>42</strong></td>
<td><strong>89806</strong></td>
</tr>
<tr>
<td>Roy 5c</td>
<td>-44698</td>
<td>51</td>
<td>89820</td>
</tr>
</tbody>
</table>
An aside: BIC curves for Roy-GMM and Roy-LCGA

Number of Classes

GMM
LCGA

Bengt Muthén  Non-Ignorable Dropout  13/25
Depression Mean Curves Estimated Under MAR, Diggle-Kenward NMAR, And Roy NMAR
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.
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

\[ y_0 \quad y_1 \quad y_2 \quad y_3 \quad y_4 \quad y_5 \]

\[ i \quad s \quad q \]

\[ c_y \quad c_d \]

\[ d_1 \quad d_2 \quad d_3 \quad d_4 \quad d_5 \]
## Model Comparisons

<table>
<thead>
<tr>
<th>Model</th>
<th>Loglikelihood</th>
<th>#par.s</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pattern-mixture</td>
<td>-44946</td>
<td>27</td>
<td>90117</td>
</tr>
<tr>
<td>Roy 4c</td>
<td>-44728</td>
<td>42</td>
<td>89806</td>
</tr>
<tr>
<td>Muthen 4c-Roy 2c</td>
<td>-44662</td>
<td>44</td>
<td>89689</td>
</tr>
</tbody>
</table>
Comparing Trajectory Class Percentages Across Models

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

<table>
<thead>
<tr>
<th>Model</th>
<th>Response class</th>
<th>Temporary response class</th>
<th>Non-response class</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAR 4 classes</td>
<td>55 %</td>
<td>3 %</td>
<td>15 %</td>
</tr>
<tr>
<td>NMAR models:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roy 4 classes</td>
<td>43 %</td>
<td>18 %</td>
<td>28 %</td>
</tr>
<tr>
<td>Muthen 4c-Roy 2c</td>
<td>32 %</td>
<td>19 %</td>
<td>32 %</td>
</tr>
</tbody>
</table>
Selection models with latent classes: Generalized Diggle-Kenward and Beunckens et al.
Muthén-Roy model extended to include an ultimate outcome


Muthén et al. (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics*.


