Mplus Short Courses
Topic 4

Growth Modeling With Latent Variables Using Mplus: Advanced Growth Models, Survival Analysis, And Missing Data

Linda K. Muthén
Bengt Muthén

www.statmodel.com
06/14/2010

Table Of Contents

Mplus Background

- Inefficient dissemination of statistical methods:
  - Many good methods contributions from biostatistics, psychometrics, etc are underutilized in practice
- Fragmented presentation of methods:
  - Technical descriptions in many different journals
  - Many different pieces of limited software
- Mplus: Integration of methods in one framework
  - Easy to use: Simple, non-technical language, graphics
  - Powerful: General modeling capabilities

Mplus Background

- Mplus versions
  - V1: November 1998
  - V2: February 2001
  - V3: March 2004
  - V4: February 2006
  - V5: November 2007
  - V5.21: May 2009
  - V6: April, 2010
- Mplus team: Linda & Bengt Muthén, Thuy Nguyen, Tihomir Asparouhov, Michelle Conn, Jean Maninger

Statistical Analysis With Latent Variables
A General Modeling Framework

Statistical Concepts Captured By Latent Variables

Continuous Latent Variables

- Measurement errors
- Factors
- Random effects
- Frailties, liabilities
- Variance components
- Missing data

Categorical Latent Variables

- Latent classes
- Clusters
- Finite mixtures
- Missing data

Statistical Analysis With Latent Variables
A General Modeling Framework (Continued)

Models That Use Latent Variables

Continuous Latent Variables

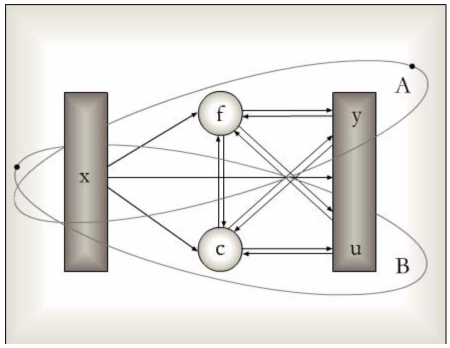
- Factor analysis models
- Structural equation models
- Growth curve models
- Multilevel models

Categorical Latent Variables

- Latent class models
- Mixture models
- Discrete-time survival models
- Missing data models

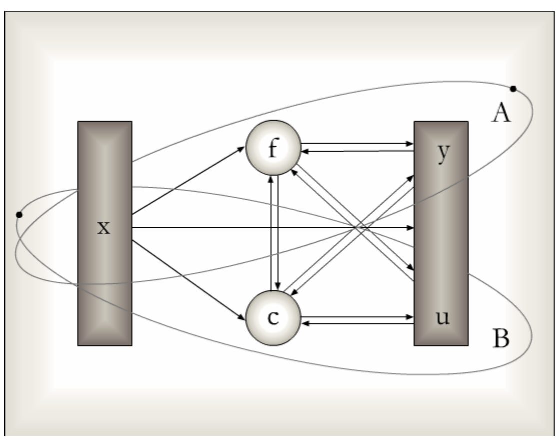
Mplus integrates the statistical concepts captured by latent variables into a general modeling framework that includes not only all of the models listed above but also combinations and extensions of these models.

General Latent Variable Modeling Framework

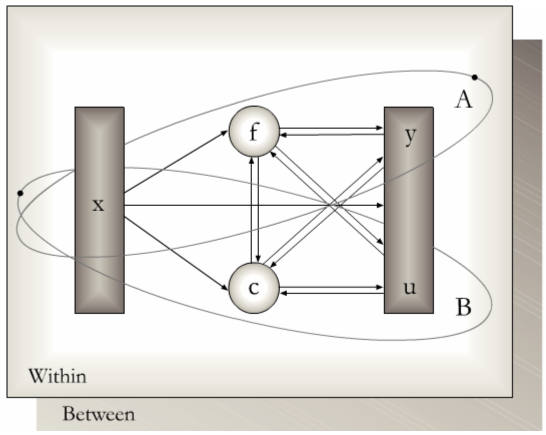


- Observed variables
  - x background variables (no model structure)
  - y continuous and censored outcome variables
  - u categorical (dichotomous, ordinal, nominal) and count outcome variables
- Latent variables
  - f continuous variables
    - – interactions among f's
  - c categorical variables
    - – multiple c's

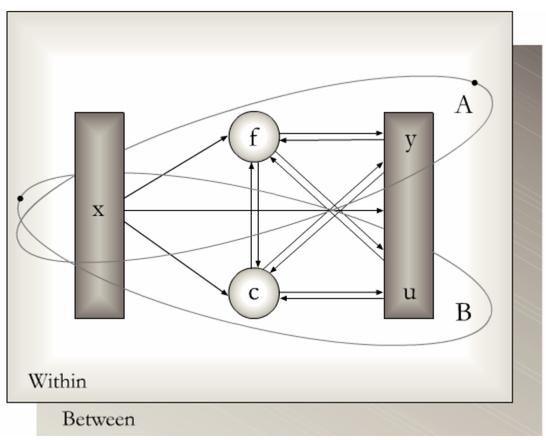
General Latent Variable Modeling Framework



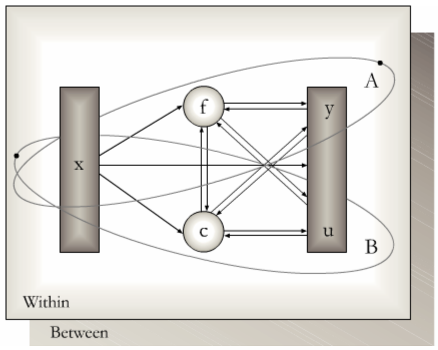
General Latent Variable Modeling Framework



General Latent Variable Modeling Framework



General Latent Variable Modeling Framework



- Observed variables
  - x background variables (no model structure)
  - y continuous and censored outcome variables
  - u categorical (dichotomous, ordinal, nominal) and count outcome variables
- Latent variables
  - f continuous variables
    - – interactions among f's
  - c categorical variables
    - – multiple c's

Mplus

Several programs in one

- Exploratory factor analysis
- Structural equation modeling
- Item response theory analysis
- Latent class analysis
- Latent transition analysis
- Survival analysis
- Growth modeling
- Multilevel analysis
- Complex survey data analysis
- Monte Carlo simulation

Fully integrated in the general latent variable framework

Overview Of Mplus Courses

- **Topic 1.** August 20, 2009, Johns Hopkins University: Introductory - advanced factor analysis and structural equation modeling with continuous outcomes
- **Topic 2.** August 21, 2009, Johns Hopkins University: Introductory - advanced regression analysis, IRT, factor analysis and structural equation modeling with categorical, censored, and count outcomes

- **Topic 3.** March 22, 2010, Johns Hopkins University: Introductory and intermediate growth modeling
- **Topic 4.** March 23, 2010, Johns Hopkins University: Advanced growth modeling, survival analysis, and missing data analysis

Overview Of Mplus Courses (Continued)

- **Topic 5.** August 16, 2010, Johns Hopkins University: Categorical latent variable modeling with cross-sectional data
- **Topic 6.** August 17, 2010, Johns Hopkins University: Categorical latent variable modeling with longitudinal data
- **Extra Topic.** August 18, 2010, Johns Hopkins University: What's new in Mplus version 6?

- **Topic 7.** March, 2011, Johns Hopkins University: Multilevel modeling of cross-sectional data
- **Topic 8.** March, 2011, Johns Hopkins University: Multilevel modeling of longitudinal data

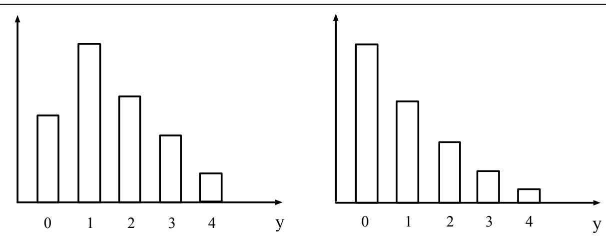
Advanced Growth Modeling

Modeling With Zeroes

Advantages Of Growth Modeling In A Latent Variable Framework

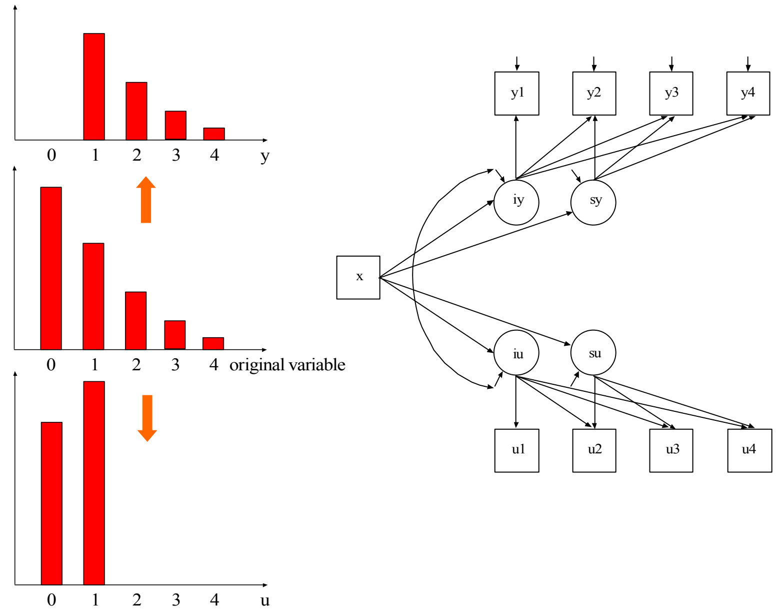
- Flexible curve shape
- Individually-varying times of observation
- Regressions among random effects
- Multiple processes
- **Modeling of zeroes**
- Multiple populations
- Multiple indicators
- Embedded growth models
- Categorical latent variables: growth mixtures

Modeling With A Preponderance Of Zeroes

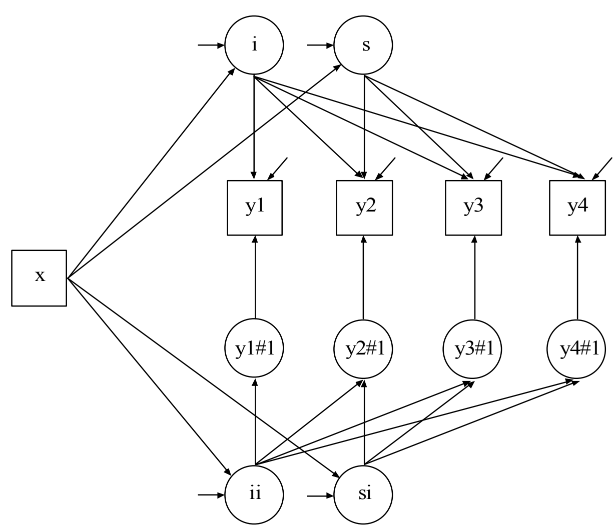


- Outcomes: non-normal continuous – count – categorical
- Censored-normal modeling
- Two-part (semicontinuous modeling): Duan et al. (1983), Olsen & Schafer (2001)
- Mixture models, e.g. zero-inflated (mixture) Poisson (Roeder et al., 1999), censored-inflated, mover-stayer latent transition models, growth mixture models
- Onset (survival) followed by growth: Albert & Shih (2003)

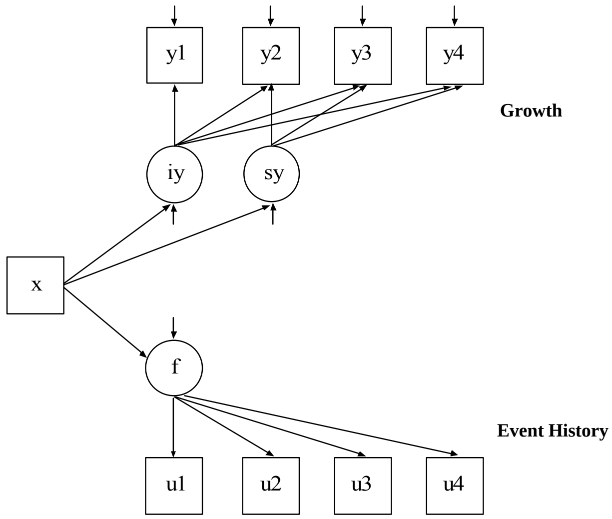
Two-Part (Semicontinuous) Growth Modeling



Inflated Growth Modeling (Two Classes At Each Time Point)



Onset (Survival) Followed By Growth

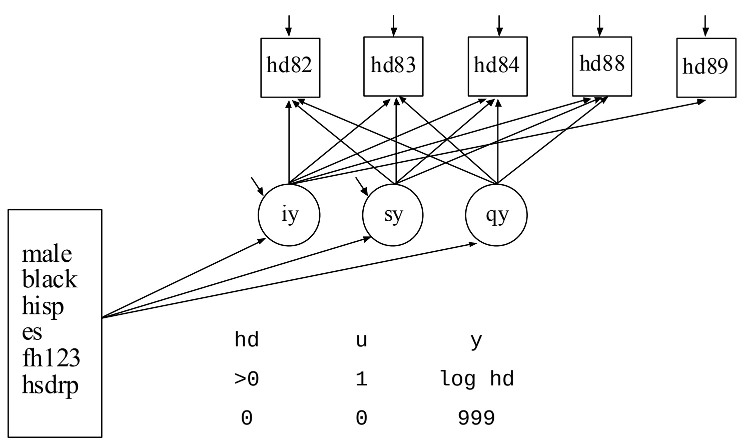


Two-Part Growth Modeling

NLSY Heavy Drinking Data

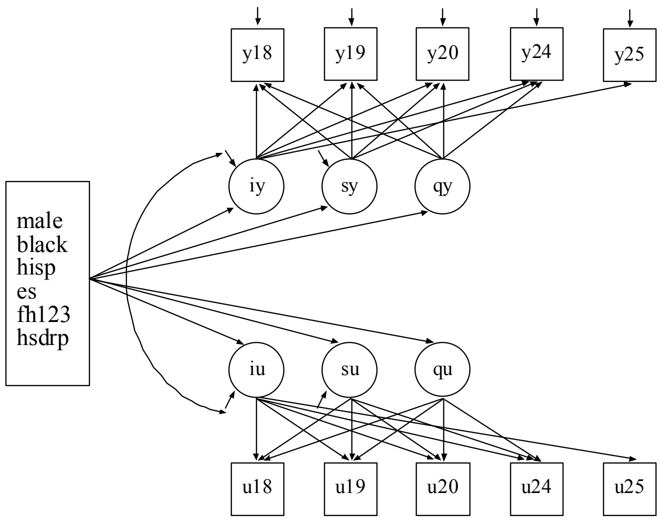
- The data are from the National Longitudinal Study of Youth (NLSY), a nationally representative household study of 12,686 men and women born between 1957 and 1964.
- There are eight birth cohorts, but the current analysis considers only cohort 64 measured in 1982, 1983, 1984, 1988, 1989, and 1994 at ages 18, 19, 20, 24, and 25.
- The outcome is heavy drinking, measured by the question: How often have you had 6 or more drinks on one occasion during the last 30 days?
- The responses are coded as: never (0); once (1); 2 or 3 times (2); 4 or 5 times (3); 6 or 7 times (4); 8 or 9 times (5); and 10 or more times (6).
- Background variables include gender, ethnicity, early onset of regular drinking (es), family history of problem drinking, high school dropout and college education

NLSY Heavy Drinking Data



hd	u	y
>0	1	log hd
0	0	999
999	999	999

NLSY Heavy Drinking Data



Input For NLSY Heavy Drinking

```
TITLE:      nlsy36425x25dep.inp
            cohort 64
            centering at 25
            hd82-hd89 (ages 18 - 25)
            log age scale: x_t = a*(ln(t-b) - ln(c-b)), where t is
            time, a and b are constants to fit the mean curve
            (chosen as a = 2 and b = 16), and c is the centering
            age, here set at 25.
DATA:       FILE = big.dat;
            FORMAT = 2f5, f2, t14, 5f7, t50, f8, t60, 6f1.0, t67,
            2f2.0, t71, 8f1.0, t79, f2.0, t82, 4f2.0;
DATA TWOPART:
            NAMES = hd82-hd89;
            BINARY = u18 u19 u20 u24 u25;
            CONTINUOUS = y18 y19 y20 y24 y25;
```

Input For NLSY Heavy Drinking (Continued)

```
VARIABLE:  NAMES = id houseid cohort weight82 weight83 weight84
           weight88 weight89 weight94 hd82 hd83 hd84 hd88 hd89
           hd94 dep89 dep94 male black hisp es fh1 fh23 fh123
           hsdrp coll ed89 ed94 cd89 cd94;
           USEOBSERVATIONS = cohort EQ 64 AND (coll GT 0 AND coll
           LT 20);
           USEV = male black hisp es fh123 hsdrp coll u18-u25
           y18-y25;
           CATEGORICAL = u18-u25;
           MISSING = .;
           AUXILIARY = hd82-hd89;
DEFINE:    CUT coll (12.1);
ANALYSIS:  ESTIMATOR = ML;
           ALGORITHM = INTEGRATION;
           COVERAGE = 0.09;
```


Input For NLSY Heavy Drinking (Continued)

```
MODEL:     iu su qu | u18@-3.008 u19@-2.197 u20@-1.621
           u24@-.235 u25@.000;
           iy sy qy | y18@-3.008 y19@-2.197 y20@-1.621 y24@-.235
           y25@.000;
           iu-qy on male black hisp es fh123 hsdrp coll;
OUTPUT:    TECH1 TECH4 TECH8 STANDARDIZED;
PLOT:      TYPE = PLOT3;
           SERIES = y18-y25(sy) | u18-u25(su);
```

Regular Growth Modeling of NLSY Heavy Drinking

Parameter	Estimate	S.E.	Est./S.E.
Regular growth modeling, treating outcome as continuous. Non-normality robust ML (MLR)			
i ON			
male	0.769	0.076	10.066
black	-0.336	0.083	-4.034
hisp	-0.227	0.103	-2.208
es	**0.291**	**0.128**	**2.283**
fh123	0.286	0.137	2.089
hsdrop	-0.024	0.104	-0.232
coll	-0.131	0.086	-1.527

Output Excerpts For Two-Part Growth Modeling of NLSY Heavy Drinking

Parameter	Estimate	S.E.	Est./S.E.
Two-part growth modeling			
iy ON			
male	0.329	0.058	5.651
black	-0.123	0.062	-1.986
hisp	-0.143	0.069	-2.082
es	**0.096**	**0.062**	**1.543**
fh123	0.219	0.076	2.894
hsdrop	0.093	0.063	1.466
coll	-0.030	0.056	-0.526

Output Excerpts For Two-Part Growth Modeling of NLSY Heavy Drinking (Continued)

Parameter	Estimate	S.E.	Est./S.E.
iu ON			
male	1.533	0.164	9.356
black	-0.705	0.172	-4.092
hisp	-0.385	0.199	-1.934
es	**0.471**	**0.194**	**2.430**
fh123	0.287	0.224	1.281
hsdrop	-0.191	0.183	-1.045
coll	-0.325	0.161	-2.017

NLSY Heavy Drinking Conclusions

As an example of differences in results between regular growth modeling and two-part growth modeling, consider the covariate es (early start, that is, early onset of regular drinking scored as 1 if the respondent had 2 or more drinks per week at age 14 or earlier):

Regular growth modeling says that es has a significant, positive influence on heavy drinking at age 25, increasing the frequency of heavy drinking.

Two-part growth modeling says that es has a significant, positive influence on the probability of heavy drinking at age 25, but among those who engage in heavy drinking at age 25 there is no significant difference in heavy drinking frequency with respect to es.

Output Excerpts For Two-Part Growth Modeling of NLSY Heavy Drinking (Continued)

Correlations

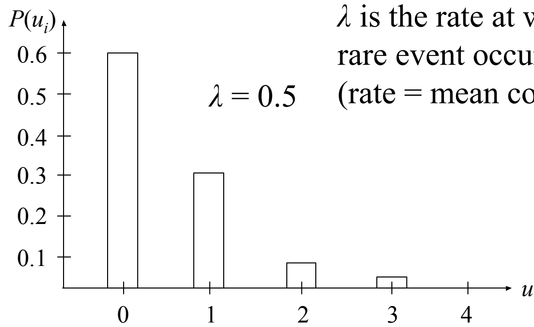
	iu	su	qu	iy	sy	qy
iu	1.000					
su	0.572	1.000				
qu	0.511	0.923	1.000			
iy	0.945	0.524	0.454	1.000		
sy	-0.140	0.472	0.440	-0.041	1.000	
qy	-0.155	0.375	0.539	-0.101	0.847	1.000

Regression With A Count Dependent Variable

Poisson Regression

A Poisson distribution for a count variable u_i has

$$P(u_i = r) = \frac{\lambda_i^r e^{-\lambda_i}}{r!}, \text{ where } u_i = 0, 1, 2, \ldots$$



λ is the rate at which a rare event occurs (rate = mean count)

Regression equation for the log rate:

$$^{e}\log \lambda_i = \ln \lambda_i = \beta_0 + \beta_1 x_i$$

Zero-Inflated Poisson (ZIP) Regression

A Poisson variable has mean = variance.

Data often have variance > mean due to preponderance of zeros.

$\pi = P$ (being in the zero class where only $u = 0$ is seen)

$1 - \pi = P$ (not being in the zero class with u following a Poisson distribution)

A mixture at zero:	ZIP mean count:	ZIP variance of count:
$P(u = 0) = \pi + \underbrace{(1 - \pi)\, e^{-\lambda}}_{\text{Poisson part}}$	$\lambda (1 - \pi)$	$\lambda (1 - \pi)(1 + \lambda * \pi)$

The ZIP model implies two regressions:

$$\text{logit}(\pi_i) = \gamma_0 + \gamma_1 x_i,$$

$$\ln \lambda_i = \beta_0 + \beta_1 x_i$$

Negative Binomial Regression

Unobserved heterogeneity ε_i is added to the Poisson model

$$ln\,\lambda_i = \beta_0 + \beta_1 x_i + \varepsilon_i \text{ , where } exp\ (\varepsilon) \sim \Gamma$$

Poisson assumes

$$E(u_i \mid x_i) = \lambda_i$$
$$V(u_i \mid x_i) = \lambda_i$$

Negative binomial assumes

$$E(u_i \mid x_i) = \lambda_i$$
$$V(u_i \mid x_i) = \lambda_i (1 + \lambda_i\ \alpha) \quad \text{NB-2}$$

NB with $\alpha = 0$ gives Poisson. When the dispersion parameter $\alpha > 0$, the NB model gives substantially higher probability for low counts and somewhat higher probability for high counts than Poisson.

Further variations are zero-inflated NB and zero-truncated NB (hurdle model or two-part model).

Zero-Inflated Poisson (ZIP) Growth Modeling Of Counts

$$u_{ti} = \begin{cases} 0 & \text{with probability } \pi_{ti} \\ \text{Poisson } (\lambda_{ti}) & \text{with probability } 1 - \pi_{ti} \end{cases}$$

$$ln\,\lambda_{ti} = \eta_{0i} + \eta_{1i} a_{ti} + \eta_{2i} a_{ti}^2$$
$$\eta_{0i} = \alpha_0 + \zeta_{0i}$$
$$\eta_{1i} = \alpha_1 + \zeta_{1i}$$
$$\eta_{2i} = \alpha_2 + \zeta_{2i}$$

In Mplus, $\pi_{ti} = P(u\#_{ti} = 1)$, where $u\#$ is a binary latent inflation variable

Philadelphia Crime Data ZIP Growth Modeling

- 13,160 males ages 4 - 26 born in 1958 (Moffitt, 1993; Nagin & Land, 1993)
- Annual counts of police contacts
- Individuals with more than 10 counts in any given year deleted (n=13,126)
- Data combined into two-year intervals

Input Excerpts Philadelphia Crime Data

```
DATA:        FILE = phillywide_zero_ln2006_del.dat;
VARIABLE:    NAMES = cohortid erace sesdummy sescomp1 juvtot
             adttot
             y10 y12 y14 y16 y18 y20 y22 y24
             sex race;
             MISSING = ALL (-9999);
             USEVAR = y10 y12 y14 y16 y18 y20 y22 y24;
             !y10 is ages 10-11, y12 is ages y12-13, etc
             COUNT = y10-y24(i);
             IDVAR = cohortid;
             USEOBS = y10 LE 10 AND y12 LE 10 AND y14 LE 10 AND
             y16 LE 10 AND y18 LE 10 AND y20 LE 10 AND y22 LE 10
             AND y24 LE 10;
```

Input Excerpts Philadelphia Crime Data (Continued)

```
ANALYSIS:
!           algorithm = integration;
            PROCESS = 4;
            INTERACTIVE = control.dat;

MODEL:      i s q | y10@0 y12@.1 y14@.2 y16@.3 y18@.4 y20@.5
            y22@.6 y24@.7;

OUTPUT:     TECH1 TECH10;

PLOT:       TYPE = PLOT3;
            SERIES = y10-y24(s);
```


Output Excerpts Philadelphia Crime Data

```
TESTS OF MODEL FIT

Loglikelihood
         H0 Value                          -40607.007
         H0 Scaling Correlation Factor          0.931
           for MLR

Information Criteria
         Number of Free Parameters                 17
         Akaike (AIC)                       81248.155
         Bayesian (BIC)                     81375.355
         Sample-Size Adjusted BIC           81321.330
           (n* = (n + 2) / 24)
```

Output Excerpts Philadelphia Crime Data (Continued)

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Means				
I	-4.689	0.145	-32.402	0.000
S	14.003	0.749	18.700	0.000
Q	20.036	0.913	21.942	0.000
Y10#1	0.768	0.123	6.268	0.000
Y12#1	-0.557	0.119	-4.679	0.000
Y14#1	-1.763	0.156	-11.322	0.000
Y16#1	-3.023	0.310	-9.746	0.000
Y18#1	-0.284	0.061	-4.638	0.000
Y20#1	-0.319	0.074	-4.293	0.000
Y22#1	-1.521	0.166	-9.156	0.000
Y24#1	-13.723	9.974	-1.376	0.169

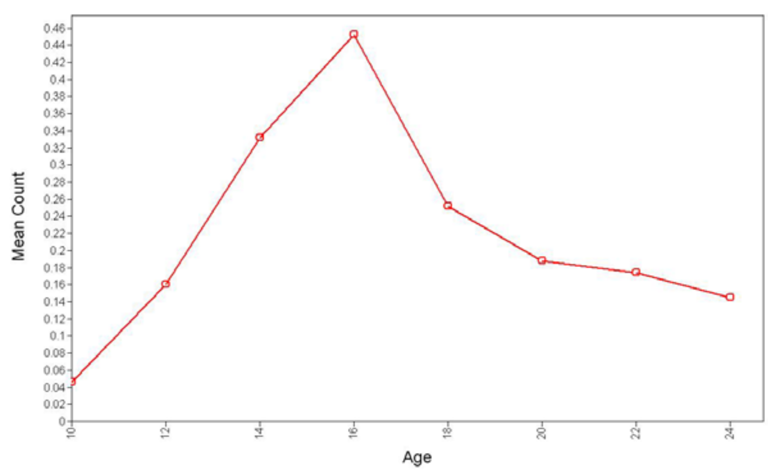
Output Excerpts Philadelphia Crime Data (Continued)

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Intercepts				
Y10	0.000	0.000	999.000	999.000
Y12	0.000	0.000	999.000	999.000
Y14	0.000	0.000	999.000	999.000
Y16	0.000	0.000	999.000	999.000
Y18	0.000	0.000	999.000	999.000
Y20	0.000	0.000	999.000	999.000
Y22	0.000	0.000	999.000	999.000
Y24	0.000	0.000	999.000	999.000
Variances				
I	5.509	0.345	15.960	0.000
S	32.931	4.568	7.206	0.000
Q	59.745	7.603	7.858	0.000

Output Excerpts Philadelphia Crime Data (Continued)

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
S	WITH				
I		-8.320	1.206	-6.896	0.000
Q	WITH				
I		5.864	1.358	4.318	0.000
S		-35.766	5.594	-6.394	0.000

Estimated Mean Counts For Philadelphia Crime Data



Philadelphia Crime Data Model Fit To Counts For Most Frequent Response Patterns

Pattern	Observed	Estimated	Z Score
00000000	8021	7850	**3.04**
00010000	572	673	**-4.00**
00100000	378	433	**-2.72**
00001000	292	354	**-3.32**
00000010	203	233	-1.95
00000100	201	266	**-4.03**
20000000	181	173	0.60
00000001	141	157	-1.27
00110000	117	112	0.50
00020000	107	95	1.28

Growth Modeling With Multiple Populations

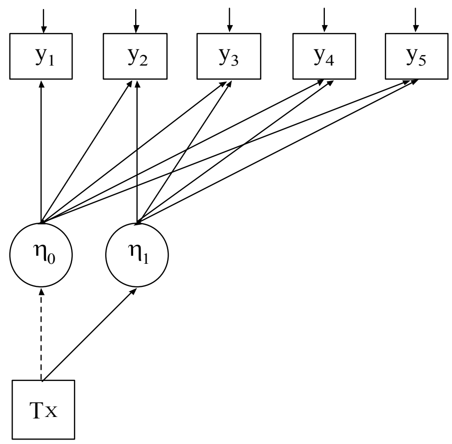
Advantages Of Growth Modeling In A Latent Variable Framework

- Flexible curve shape
- Individually-varying times of observation
- Regressions among random effects
- Multiple processes
- Modeling of zeroes
- **Multiple populations**
- Multiple indicators
- Embedded growth models
- Categorical latent variables: growth mixtures

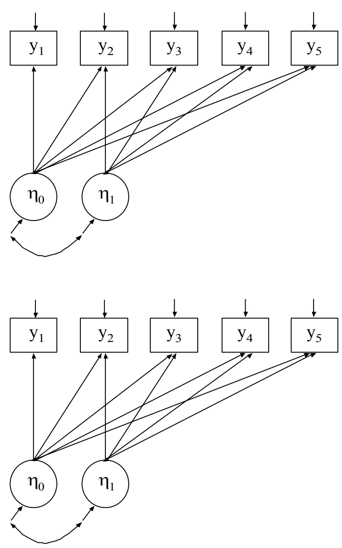
Multiple Population Growth Modeling

- Group as a dummy variable
- Multiple-group analysis
- Multiple-group analysis of randomized interventions

Group Dummy Variable As A Covariate



Two-Group Model



Multiple Population Growth Modeling Specifications

Let y_{git} denote the outcome for population (group) g, individual i, and timepoint t,

$$\textit{Level 1}: y_{gti} = \eta_{g0i} + \eta_{g1i}\, x_t + \varepsilon_{gti}, \quad (65)$$
$$\textit{Level 2a}: \eta_{g0i} = \alpha_{g0} + \gamma_{g0}\, w_{gi} + \zeta_{g0i}, \quad (66)$$
$$\textit{Level 2b}: \eta_{g1i} = \alpha_{g1} + \gamma_{g1}\, w_{gi} + \zeta_{g1i}, \quad (67)$$

Measurement invariance (level-1 equation): time-invariant intercept 0 and slopes 1, x_t

Structural differences (level-2): α_g , γ_g , $V(\zeta_g)$

Alternative parameterization:

$$\textit{Level 1}: y_{gti} = \nu + \eta_{g0i} + \eta_{g1i}\, x_t + \varepsilon_{gti}, \quad (68)$$

with α_{10} fixed at zero in level 2a.

Analysis steps:

1. Separate growth analysis for each group
2. Joint analysis of all groups, free structural parameters
3. Join analysis of all groups, tests of structural parameter invariance

NLSY: Multiple Cohort Structure

Birth Year Cohort	Age[a] 18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
57								82	83	84	85	86	87	88	89	90	91	92	93	94
58							82	83	84	85	86	87	88	89	90	91	92	93	94	
59						82	83	84	85	86	87	88	89	90	91	92	93	94		
60					82	83	84	85	86	87	88	89	90	91	92	93	94			
61				82	83	84	85	86	87	88	89	90	91	92	93	94				
62			82	83	84	85	86	87	88	89	90	91	92	93	94					
63		82	83	84	85	86	87	88	89	90	91	92	93	94						
64	82	83	84	85	86	87	88	89	90	91	92	93	94							

[a] Non-shaded areas represent years in which alcohol measures were obtained

Three Approaches To Cohort Structures

- Single group - using y18 - y37
- Single group - using 7 y's and AT with TSCORES to capture varying ages
- Multiple group - using 7 y's and 8 cohorts with s@xt

Multiple Group Modeling Of Multiple Cohorts

- Data – two cohorts born in 1961 and 1962 measured on the frequency of heavy drinking in the years 1983, 1984, 1988, and 1989
- Development of heavy drinking across chronological age, not year of measurement, is of interest

Cohort/Year	1983	1984	1988	1989
1961 (older)	**22**	23	**27**	28
1962 (younger)	21	**22**	26	**27**

Cohort/Age	21	22	23	24	25	26	27	28
1961 (older)		**83**	**84**				**88**	**89**
1962 (younger)	**83**	**84**				**88**	**89**	

Multiple Group Modeling Of Multiple Cohorts (Continued)

- Time scores calculated for age, not year of measurement

Age	**21**	**22**	**23**	24	25	**26**	**27**	**28**
Time score	**0**	**1**	**2**	3	4	**5**	**6**	**7**

Cohort 1961 time scores 1 2 6 7
Cohort 1962 time scores 0 1 5 6

- Can test the degree of measurement and structural invariance
  - Test of full invariance
    - Growth factor means, variances, and covariances held equal across cohorts
    - Residual variances of shared ages held equal across cohorts

Input For Multiple Group Modeling Of Multiple Cohorts

```
TITLE:      Multiple Group Modeling Of Multiple Cohorts
DATA:       FILE IS cohort.dat;
VARIABLE:   NAMES ARE cohort hd83 hd84 hd88 hd89;
            MISSING ARE *;
            USEV = hd83 hd84 hd88 hd89;
            GROUPING IS cohort (61 = older 62 = younger);
MODEL:      i s | hd83@0 hd84@1 hd88@5 hd89@6;
            [i] (1);
            [s] (2);
            i (3);
            s (4);
            i WITH s (5);
```

Input For Multiple Group Modeling Of Multiple Cohorts (Continued)

```
MODEL older:
              i s | hd83@1 hd84@2 hd88@6 hd89@7;
              hd83 (6);
              hd88 (7);
MODEL younger:
              hd84 (6);
              hd89 (7);

OUTPUT:       STANDARDIZED;
```


Output Excerpts Multiple Group Modeling Of Multiple Cohorts

Tests Of Model Fit

```
Chi-Square Test of Model Fit
          Value                               68.096
          Degrees of Freedom                      17
          P-Value                              .0000

Chi-Square Contributions From Each Group
          OLDER                               39.216
          YOUNGER                             28.880

Chi-Square Test of Model Fit for the Baseline Model
          Value                             3037.930
          Degrees of Freedom                      12
          P-Value                             0.0000

CFI/TLI
          CFI                                  0.983
          TLI                                  0.988
```

Output Excerpts Multiple Group Modeling Of Multiple Cohorts (Continued)

```
Loglikelihood
          H0 Value                      -18544.420
          H1 Value                      -18510.371

Information Criteria
          Number of Free Parameters             11
          Akaike (AIC)                   37110.839
          Bayesian (BIC)                 37175.770
          Sample-Size Adjusted BIC       37140.820
            (n* = (n + 2) / 24)

RMSEA (Root Mean Square Error Of Approximation)
          Estimate                           0.047
          90 Percent C.I.                    0.036  0.059

SRMR (Standardized Root Mean Square Residual)
          Value                              0.033
```


Output Excerpts Multiple Group Modeling Of Multiple Cohorts (Continued)

Model Results

	Estimates	S.E.	Est./S.E.	Std	StdYX
Group OLDER					
I WITH					
S	-.111	.010	-11.390	-.537	-.537
Residual Variances					
HD83	1.141	.046	24.996	1.141	.445
HD84	1.062	.057	18.489	1.062	.453
HD88	1.028	.041	25.326	1.028	.455
HD89	.753	.053	14.107	.753	.358
Variances					
I	1.618	.068	23.651	1.000	1.000
S	.026	.002	13.372	1.000	1.000
Means					
I	1.054	.030	35.393	.828	.828
S	-.032	.005	-6.611	-.200	-.200

Output Excerpts Multiple Group Modeling Of Multiple Cohorts (Continued)

```
GROUP YOUNGER
Residual Variances
    HD83           1.049    .066   15.916    1.049    .393
    HD84           1.141    .046   24.996    1.141    .445
    HD88           1.126    .056   19.924    1.126    .491
    HD89           1.028    .041   25.326    1.028    .455
```


Preventive Interventions Randomized Trials

Prevention Science Methodology Group (PSMG)

Developmental Epidemiological Framework:

- Determining the levels and variation in risk and protective factors as well as developmental paths within a defined population in the absence of intervention
- Directing interventions at these risk and protective factors in an effort to change the developmental trajectories in a defined population
- Evaluating variation in intervention impact across risk levels and contexts on proximal and distal outcomes, thereby empirically testing the developmental model

Aggressive Classroom Behavior: The GBG Intervention

Muthén & Curran (1997, Psychological Methods)

The Johns Hopkins Prevention Center carried out a school-based preventive intervention randomized trial in Baltimore public schools starting in grade 1. One of the interventions tested was the Good Behavior Game intervention, a classroom based behavior management strategy promoting good behavior. It was designed specifically to reduce aggressive behavior of first graders and was aimed at longer term impact on aggression through middle school.

One first grade classroom in a school was randomly assigned to receive the Good Behavior Game intervention and another matched classroom in the school was treated as control. After an initial assessment in fall of first grade, the intervention was administered during the first two grades.

Aggressive Classroom Behavior: The GBG Intervention (Continued)

The outcome variable of interest was teacher ratings (TOCA-R) of each child's aggressive behavior (breaks rules, harms property, fights, etc.) in the classroom through grades 1 – 6. Eight teacher ratings were made from fall and spring for the first two grades and every spring in grades 3 – 6.

The most important scientific question was whether the Good Behavior Game reduces the slope of the aggression trajectory across time. It was also of interest to know whether the intervention varies in impact for children who started out as high aggressive versus low aggressive.

Analyses in Muthén-Curran (1997) were based on data for 75 boys in the GBG group who stayed in the intervention condition for two years and 111 boys in the control group.

The GBG Aggression Example: Analysis Results

Muthén & Curran (1997):

- Step 1: Control group analysis
- Step 2: Treatment group analysis
- Step 3: Two-group analysis w/out interactions
- Step 4: Two-group analysis with interactions
- Step 5: Sensitivity analysis of final model
- Step 6: Power analysis

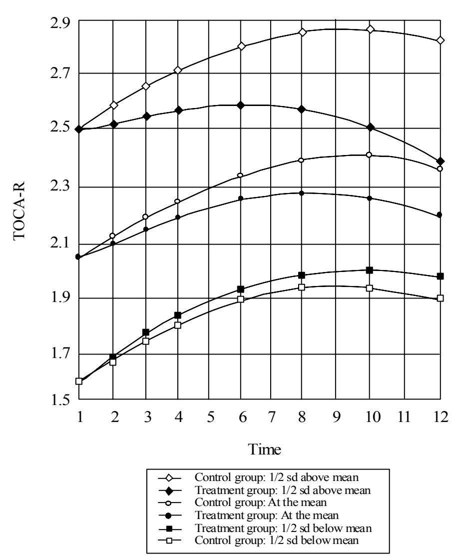
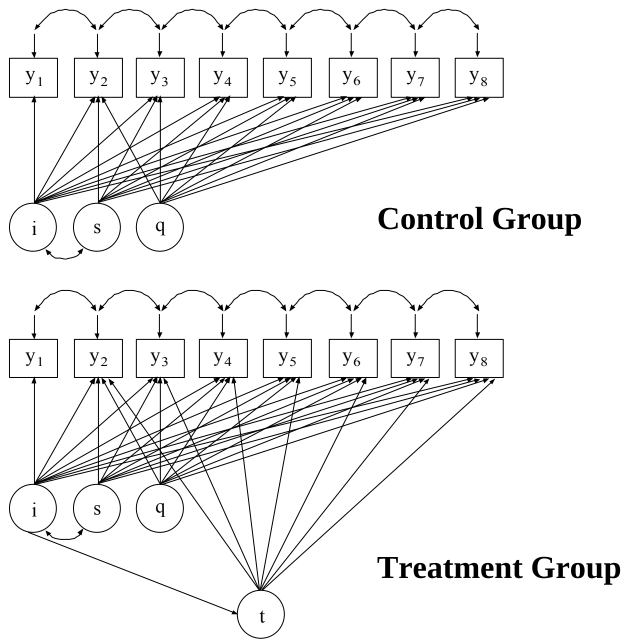


Figure 15. Model implied growth trajectories of Teacher Observation of Classroom Behavior—Revised (TOCA-R) scores as a function of initial status. Each timepoint represents one 6-month interval.



Input Excerpts For Aggressive Behavior Intervention Using A Multiple Group Growth Model With A Regression Among Random Effects

```
TITLE:          Aggressive behavior intervention growth model
                n = 111 for control group
                n = 75 for tx group
MODEL:          i s q | y1@0 y2@1 y3@2 y4@3 y5@5 y6@7 y7@9 y8@11;
                i t | y1@0 y2@1 y3@2 y4@3 y5@5 y6@7 y7@9 y8@11;
                [y1-y8] (1);   !alternative growth model
                [i@0];         !parameterization
                i (2);
                s (3);
                i WITH s (4);
                [s] (5);
                [q] (6);
                t@0 q@0;
                q WITH i@0 s@0 t@0; y1-y7 PWITH y2-y8;
                t ON i;
```

Input Excerpts For Aggressive Behavior Intervention Using A Multiple Group Growth Model With A Regression Among Random Effects (Continued)

```
MODEL control:
                t ON i@0;
                [t@0];
```


Output Excerpts Aggressive Behavior Intervention Using A Multiple Group Growth Model With A Regression Among Random Effects

Tests Of Model Fit

```
Chi-Square Test of Model Fit

          Value                         64.553
          Degrees of Freedom                50
          P-Value                        .0809

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                        .056
          90 Percent C.I.                 .000   .092
```

Output Excerpts Aggressive Behavior Intervention Using A Multiple Group Growth Model With A Regression Among Random Effects (Continued)

Group Control		Group Tx	
Observed Variable	R-Square	Observed Variable	R-Square
Y1	.644	Y1	.600
Y2	.642	Y2	.623
Y3	.663	Y3	.568
Y4	.615	Y4	.464
Y5	.637	Y5	.425
Y6	.703	Y6	.399
Y7	.812	Y7	.703
Y8	.818	Y8	.527

Output Excerpts Aggressive Behavior Intervention Using A Multiple Group Growth Model With A Regression Among Random Effects (Continued)

Group Control	Estimates	S.E.	Est./S.E.	Std	StdYX
T ON					
I	**.000**	**.000**	**.000**	**999.000**	**999.000**
Residual Variances					
Y1	.444	.088	5.056	.444	.356
Y2	.449	.079	5.714	.449	.358
Y3	.414	.069	6.026	.414	.337
Y4	.522	.080	6.551	.522	.385
Y5	.512	.079	6.469	.512	.363
Y6	.422	.074	5.677	.422	.297
Y7	.264	.083	3.186	.264	.188
Y8	.291	.094	3.097	.291	.182
T	.000	.000	.000	999.000	999.000
Variances					
I	.803	.109	7.330	1.000	1.000
S	.004	.001	3.869	1.000	1.000
Q	.000	.000	.000	999.000	999.000

Output Excerpts Aggressive Behavior Intervention Using A Multiple Group Growth Model With A Regression Among Random Effects (Continued)

	Estimates	S.E.	Est./S.E.	Std	StdYX
Means					
I	.000	.000	.000	.000	.000
S	.086	.021	4.035	1.285	1.285
Q	-.005	.002	-3.005	999.000	999.000
Intercepts					
Y1	2.041	.078	26.020	2.041	1.828
Y2	2.041	.078	26.020	2.041	1.823
Y3	2.041	.078	26.020	2.041	1.841
Y4	2.041	.078	26.020	2.041	1.753
Y5	2.041	.078	26.020	2.041	1.718
Y6	2.041	.078	26.020	2.041	1.711
Y7	2.041	.078	26.020	2.041	1.724
Y8	2.041	.078	26.020	2.041	1.612
T	.000	.000	.000	999.000	999.000

Output Excerpts Aggressive Behavior Intervention Using A Multiple Group Growth Model With A Regression Among Random Effects (Continued)

Group Tx

	Estimates	S.E.	Est./S.E.	Std	StdYX
T ON					
I	**-.052**	**.015**	**-3.347**	**-1.000**	**-1.000**
Residual Variances					
Y1	.535	.141	3.801	.535	.400
Y2	.439	.122	3.595	.439	.377
Y3	.501	.108	4.653	.501	.432
Y4	.701	.132	5.332	.701	.536
Y5	.736	.133	5.545	.736	.575
Y6	.805	.152	5.288	.805	.601
Y7	.245	.104	2.364	.245	.297
Y8	.609	.182	3.351	.609	.473
T	.000	.000	.000	.000	.000
Variances					
I	.803	.109	7.330	1.000	1.000
S	.004	.001	3.869	1.000	1.000
Q	.000	.000	.000	999.000	999.000

Output Excerpts Aggressive Behavior Intervention Using A Multiple Group Growth Model With A Regression Among Random Effects (Continued)

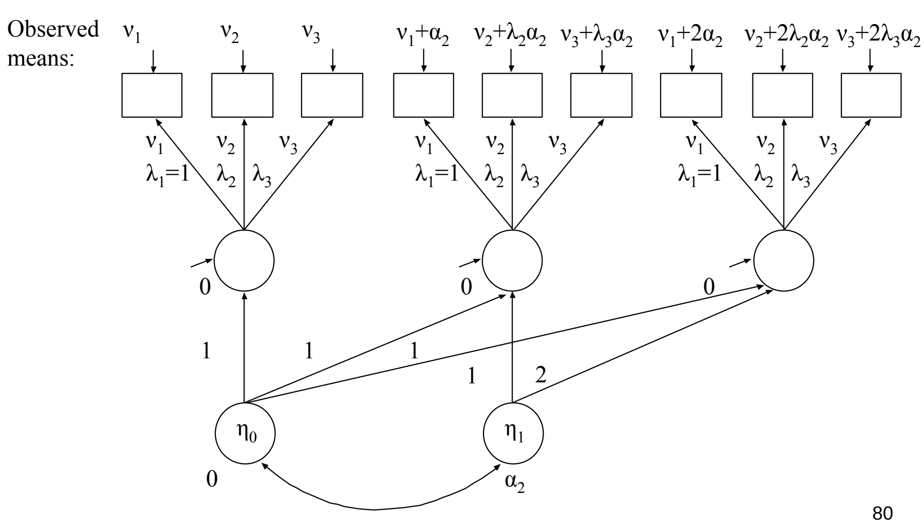
	Estimates	S.E.	Est./S.E.	Std	StdYX
Means					
I	.000	.000	.000	.000	.000
S	.086	.021	4.035	1.285	1.285
Q	-.005	.002	-3.005	999.000	999.000
Intercepts					
Y1	2.041	.078	26.020	2.041	1.764
Y2	2.041	.078	26.020	2.041	1.893
Y3	2.041	.078	26.020	2.041	1.895
Y4	2.041	.078	26.020	2.041	1.785
Y5	2.041	.078	26.020	2.041	1.805
Y6	2.041	.078	26.020	2.041	1.764
Y7	2.041	.078	26.020	2.041	2.248
Y8	2.041	.078	26.020	2.041	1.799
T	-.016	.013	-1.225	-.341	-.341

Growth Modeling With Multiple Indicators

Advantages Of Growth Modeling In A Latent Variable Framework

- Flexible curve shape
- Individually-varying times of observation
- Regressions among random effects
- Multiple processes
- Modeling of zeroes
- Multiple populations
- **Multiple indicators**
- Embedded growth models
- Categorical latent variables: growth mixtures

Growth Of Latent Variable Construct Measured By Multiple Indicators



Multiple Indicator Growth Modeling Specifications

Let y_{jti} denote the outcome for individual i, indicator j, and timepoint t, and let η_{ti} denote a latent variable construct,

Level 1*a (measurement part):*

$$y_{jti} = \nu_{jt} + \lambda_{jt}\,\eta_{ti} + \varepsilon_{jti}, \qquad (44)$$

$$\textit{Level } 1b: \eta_{ti} = \eta_{0i} + \eta_{1i}\; x_t + \zeta_{ti}, \qquad (45)$$

$$\textit{Level } 2a: \eta_{0i} = \alpha_0 + \gamma_0\, w_i + \zeta_{0i}, \qquad (46)$$

$$\textit{Level } 2b: \eta_{1i} = \alpha_1 + \gamma_1\, w_i + \zeta_{1i}, \qquad (47)$$

Measurement invariance: time-invariant indicator intercepts and slopes:

$$\nu_{j1} = \nu_{j2} = \ldots \nu_{jT} = \nu_j, \qquad (48)$$

$$\lambda_{j1} = \lambda_{j2} = \ldots \lambda_{jT} = \lambda_j, \qquad (49)$$

where $\lambda_1 = 1$, $\alpha_0 = 0$. $V(\varepsilon_{jti})$ and $V(\zeta_{ti})$ may vary over time. Structural differences: $E(\eta_{ti})$ and $V(\eta_{ti})$ vary over time.

Steps In Growth Modeling With Multiple Indicators

- Exploratory factor analysis of indicators for each timepoint
- Determine the shape of the growth curve for each indicator and the sum of the indicators
- Fit a growth model for each indicator—must be the same
- Confirmatory factor analysis of all timepoints together
  - Covariance structure analysis without measurement parameter invariance
  - Covariance structure analysis with invariant loadings
  - Mean and covariance structure analysis with invariant measurement intercepts and loadings
- Growth model with measurement invariance across timepoints

Advantages Of Using Multiple Indicators Instead Of An Average

- Estimation of unequal weights
- No need to assume full measurement invariance:
  - Partial measurement invariance—changes across time in individual item functioning
- No confounding of time-specific variance and measurement error variance
- Smaller standard errors for growth factor parameters (more power)

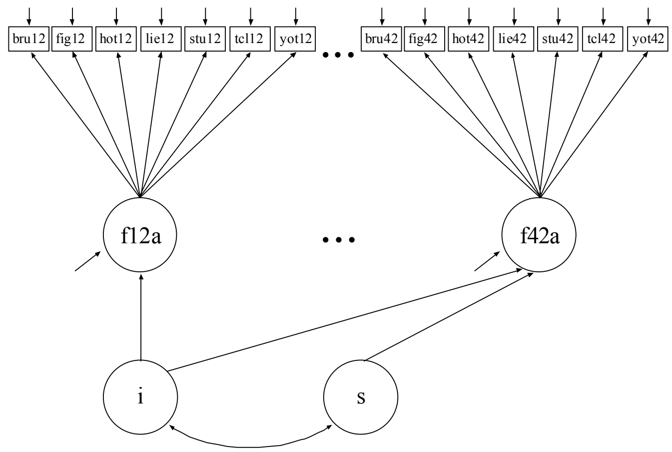
Classroom Aggression Data (TOCA)

The classroom aggression data are from an intervention study in Baltimore public schools carried out by the Johns Hopkins Prevention Research Center. Subjects were randomized into treatment and control conditions. The TOCA-R instrument was used to measure 10 aggression items at multiple timpoints. The TOCA-R is a teacher rating of student behavior in the classroom. The items are rated on a six-point scale from almost never to almost always.

Data for this analysis include the 342 boys in the control group. Four time points are examined: Spring Grade 1, Spring Grade 2, Spring Grade 3, and Spring Grade 4.

Seven aggression items are used in the analysis:

- Break rules
- Fights
- Harms others
- Lies
- Stubborn
- Teasing classmates
- Yells at others



Input Excerpts For TOCA Data Multiple Indicator CFA With No Measurement Invariance

```
TITLE:      Multiple indicator CFA with no measurement invariance
.
.
.
MODEL:      f12a BY bru12-yot12;
            f22a BY bru22-yot22;
            f32a BY bru32-yot32;
            f42a BY bru42-yot42;
```

Input Excerpts For TOCA Data Multiple Indicator CFA With Factor Loading Invariance

```
TITLE:       Multiple indicator CFA with factor loading invariance
.
.
.
MODEL:       f12a BY bru12
                     fig12-yot12 (1-6);
             f22a BY bru22
                     fig22-yot22 (1-6);
             f32a BY bru32
                     fig32-yot32 (1-6);
             f42a BY bru42
                     fig42-yot42 (1-6);
```


Input Excerpts For TOCA Data Multiple Indicator CFA With Factor Loading And Intercept Invariance

```
TITLE:       Multiple indicator CFA with factor loading and intercept
             incariance
.
.
.
MODEL:       f12a BY bru12
                     fig12-yot12 (1-6);
             f22a BY bru22
                     fig22-yot22 (1-6);
             f32a BY bru32
                     fig32-yot32 (1-6);
             f42a BY bru42
                     fig42-yot42 (1-6);
```

Input Excerpts For TOCA Data Multiple Indicator CFA With Factor Loading And Intercept Invariance (Continued)

```
[bru12 bru22 bru32 bru42] (7);
[fig12 fig22 fig32 fig42] (8);
[hot12 hot22 hot32 hot42] (9);
[lie12 lie22 lie32 lie42] (10);
[stu12 stu22 stu32 stu42] (11);
[tcl12 tcl22 tcl32 tcl42] (12);
[yot12 yot22 yot32 yot42] (13);

[f12a@0 f22a f32a f42a];
```


Input Excerpts For TOCA Data Multiple Indicator CFA With Factor Loading Invariance And Partial Intercept Invariance

```
TITLE:      Multiple indicator CFA with factor loading and partial
            intercept invariance

MODEL:      f12a BY bru12
                    fig12-yot12 (1-6);
            f22a BY bru22
                    fig22-yot22 (1-6);
            f32a BY bru32
                    fig32-yot32 (1-6);
            f42a BY bru42
                    fig42-yot42 (1-6);
```

Input Excerpts For TOCA Data Multiple Indicator CFA With Factor Loading Invariance And Partial Intercept Invariance (Continued)

```
[bru12 bru22 bru32 bru42] (7);
[fig12 fig22 fig32 fig42] (8);
[hot12 hot22 hot32      ] (9);
[lie12 lie22 lie32 lie42] (10);
[stu12 stu22            ] (11);
[tcl12 tcl22 tcl32      ] (12);
[yot12 yot22 yot32 yot42] (13);

[f12a@0 f22a f32a f42a];
```


Summary of Analysis Results For TOCA Measurement Invariance Models

Model	Chi-Square (d.f.)	Difference (d.f. diff.)
Measurement non-invariance	567.08 (344)	
Factor loading invariance	581.29 (362)	14.21 (18)
Factor loading and intercept invariance	654.59 (380)	73.30* (18)
Factor loading and partial intercept invariance	606.97 (376)	25.68* (14)
Factor loading and partial intercept invariance with a linear growth structure	614.74 (381)	7.77 (5)

Summary of Analysis Results For TOCA Measurement Invariance Models (Continued)

Explanation of Chi-Square Differences

Factor loading invariance (18)	6 factor loadings instead of 24
Factor loading and intercept invariance (18)	7 intercepts plus 3 factor means instead of 28 intercepts
Factor loading and partial intercept invariance (14)	4 additional intercepts
Factor loading and partial intercept invariance with a linear growth structure (5)	1 growth factor mean instead of 3 factor means 2 growth factor variances, 1 growth factor covariance, 4 factor residual variances instead of 10 factor variances/covariances

Input Excerpts For TOCA Data Multiple Indicator CFA With Factor Loading And Intercept Invariance With A Linear Growth Structure

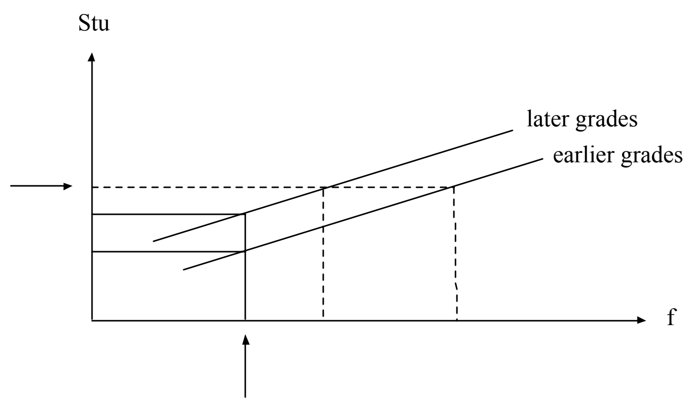
```
MODEL:      f12a BY bru12
                    fig12-yot12 (1-6);
            f22a BY bru22
                    fig22-yot22 (1-6);
            f32a BY bru32
                    fig32-yot32 (1-6);
            f42a BY bru42
                    fig42-yot42 (1-6);
```

Input Excerpts For TOCA Data Multiple Indicator CFA With Factor Loading And Intercept Invariance With A Linear Growth Structure (Continued)

```
[bru12 bru22 bru32 bru42] (7);
[fig12 fig22 fig32 fig42] (8);
[hot12 hot22 hot32      ] (9);
[lie12 lie22 lie32 lie42] (10);
[stu12 stu22            ] (11);
[tcl12 tcl22 tcl32      ] (12);
[yot12 yot22 yot32 yot42] (13);

i s | f12a@0 f22a@1 f32a@2 f42a@3;



Alternative language:

i BY f12a-f42a@1;
s BY f12a@0 f22a@1 f32a@2 f42a@3;
[f12a-f42a@0 i@0 s];
```


Output Excerpts For TOCA Data Multiple Indicator CFA With Factor Loading And Intercept Invariance With A Linear Growth Structure

	Estimates	S.E.	Est./S.E.	Std	StdYX
F12A BY					
BRU12	1.000	.000	.000	.190	.786
FIG12	1.097	.039	28.425	.208	.868
HOT12	.986	.037	26.586	.187	.811
LIE12	.967	.041	23.769	.184	.742
STU12	.880	.041	21.393	.167	.667
TCL12	1.034	.039	26.206	.196	.786
YOT12	.932	.039	23.647	.177	.709
Intercepts					
STU12	.331	.013	25.408	.331	1.324
STU22	.331	.013	25.408	.331	1.231
STU32	.417	.017	24.345	.417	1.592
STU42	.390	.017	23.265	.390	1.496

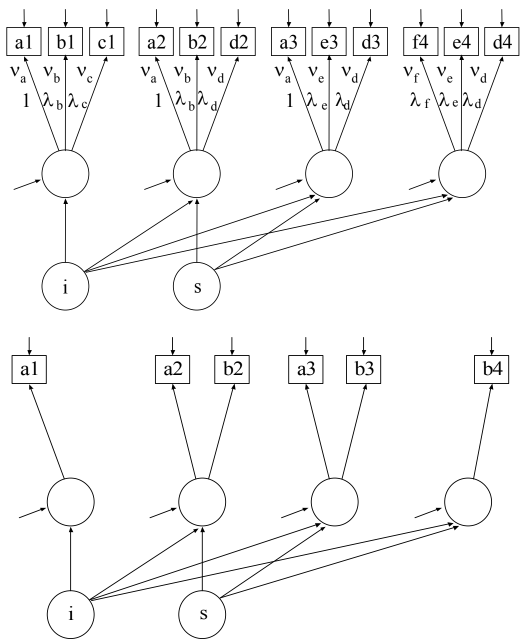


97

Degrees Of Invariance Across Time

- Case 1
  - Same items
  - All items invariant
  - Same construct
- Case 2
  - Same items
  - Some items non-invariant
  - Same construct
- Case 3
  - Different items
  - Some items invariant
  - Same construct
- Case 4
  - Different items
  - Some items invariant
  - Different construct

98

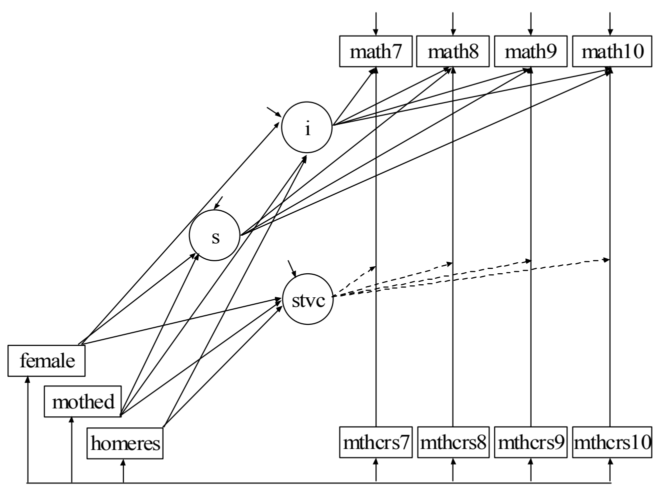


Embedded Growth Models

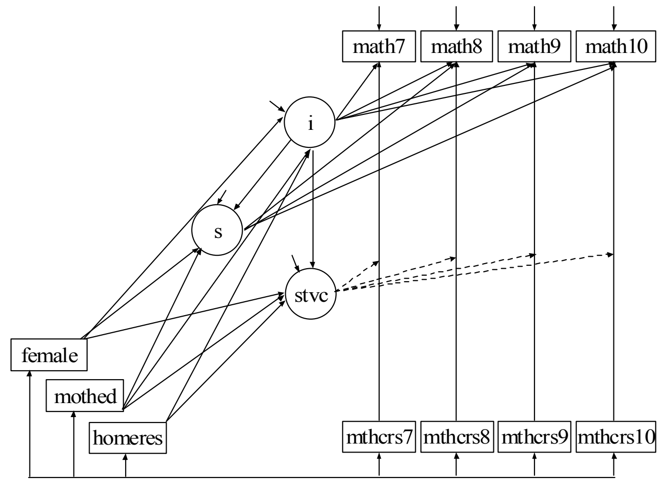
Advantages Of Growth Modeling In A Latent Variable Framework

- Flexible curve shape
- Individually-varying times of observation
- Regressions among random effects
- Multiple processes
- Modeling of zeroes
- Multiple populations
- Multiple indicators
- **Embedded growth models**
- Categorical latent variables: growth mixtures

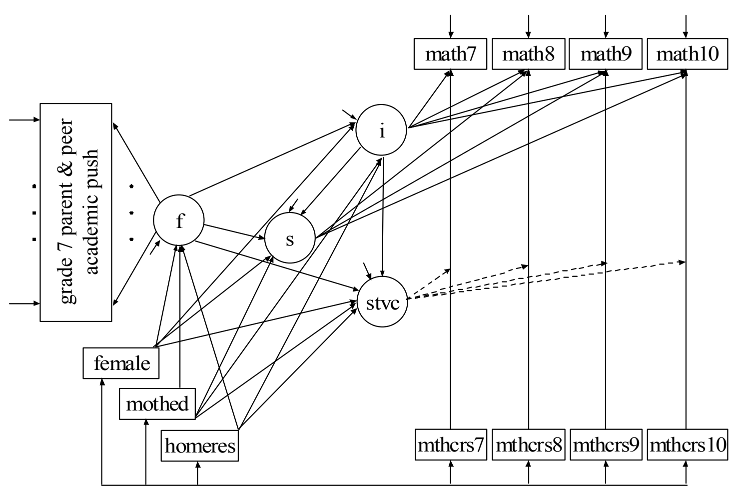
Growth Modeling With Time-Varying Covariates



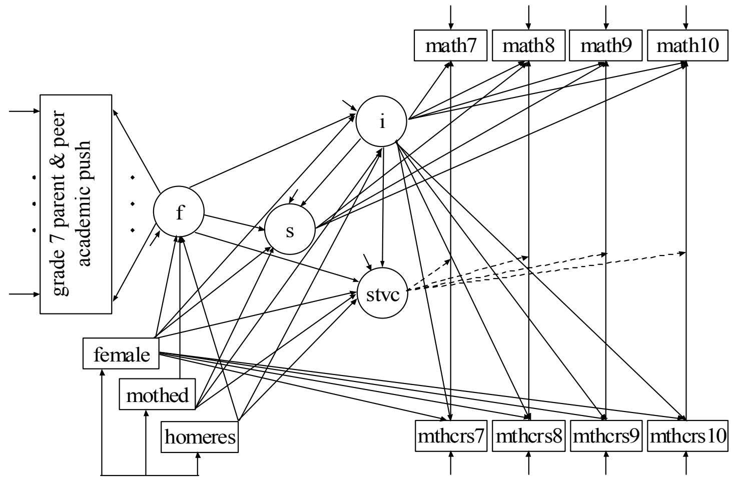
A Generalized Growth Model



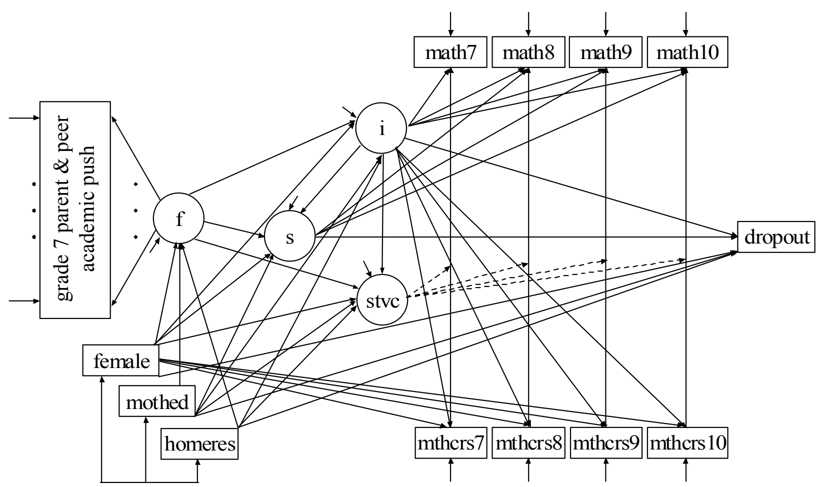
A Generalized Growth Model



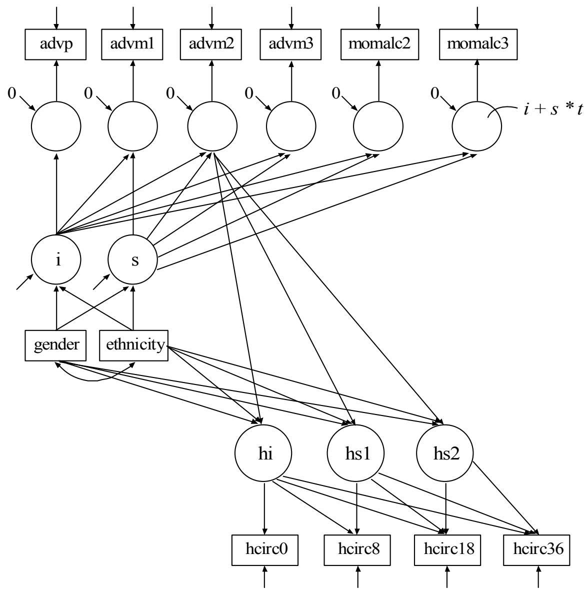
A Generalized Growth Model



A Generalized Growth Model



Two Linked Processes



Input Excerpts For Two Linked Processes With Measurement Error In The Covariates

```
TITLE:     Embedded growth model with measurement error in the
           covariates and sequential processes
           advp: mother's drinking before pregnancy
           advm1-advm3: drinking in first trimester
           momalc2-momalc3: drinking in 2nd and 3rd trimesters
           hcirc0-hcirc36; head circumference
MODEL:     fadvp     BY  advp;       fadvp@0;
           fadvm1    BY  advm1;      fadvm1@0;
           fadvm2    BY  advm2;      fadvm2@0;
           fadvm3    BY  advm3;      fadvm3@0;
           fmomalc2 BY  momalc2;  fmomalc2@0;
           fmomalc3 BY  momalc3;  fmomalc3@0;
           i BY fadvp-fmomalc3@1;
           s BY fadvp@0  fadvm1@1  fadvm2*2  fadvm3*3
                  fmomalc2-fmomalc3*5 (1);
           [advp-momalc3@0 fadvp-fmomalc3@0 i s];
```

Input Excerpts For Two Linked Processes With Measurement Error In The Covariates (Continued)

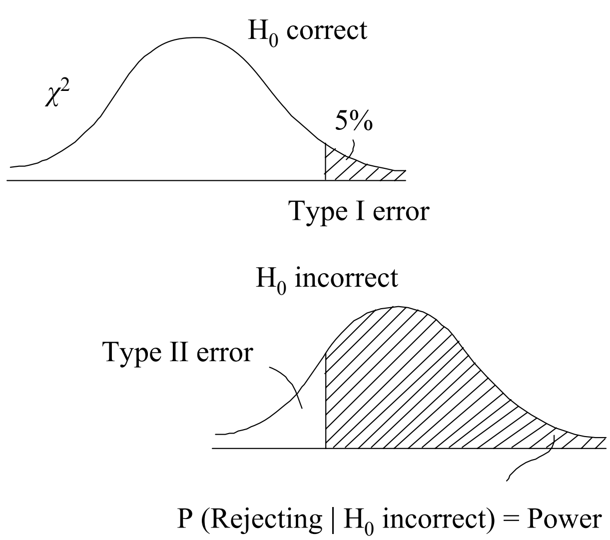
```
advp WITH advm1; advm1 WITH advm2; advm3 WITH advm2;

i s ON gender eth; s WITH i;


hi   BY  hcirc0-hcirc36@1;
hs1  BY  hcirc0@0 hcirc8@1.196 hcirc36@1.196 hcirc36@1.196;
hs2  BY  hcirc0@0 hcirc8@0 hcirc18@1 hcirc36*2;

[hcirc0-hcirc36@0 hi*34 hs1 hs2];

hs1 WITH hs2@0; hi WITH hs2@0; hi WITH hs1@0;
hi  WITH i@0; hi   WITH s@0; hs1  WITH i@0;
hi1 WITH s@0; hs2  WITH i@0; hs2  WITH s@0;

hi-hs2 ON gender eth fadvm2;
```


Power For Growth Models

Designing Future Studies: Power

- Computing power for growth models using Satorra-Saris (Muthén & Curran, 1997; examples)
- Computing power using Monte Carlo studies (Muthén & Muthén, 2002)
- Power calculation web site – PSMG
- Multilevel power (Miyazaki & Raudenbush, 2000; Moerbeek, Breukelen & Berger, 2000; Raudenbush, 1997; Raudenbush & Liu, 2000)
- School-based studies (Brown & Liao, 1999: Principles for designing randomized preventive trials)
- Multiple- (sequential-) cohort power
- Designs for follow-up (Brown, Indurkhia, & Kellam, 2000)

Designing Future Studies: Power



Power Estimation For Growth Models Using Satorra & Saris (1985)

- Step 1: Create mean vector and covariance matrix for hypothesized parameter values
- Step 2: Analyze as if sample statistics and check that parameter values are recovered
- Step 3: Analyze as if sample statistics, misspecifying the model by fixing treatment effect(s) at zero
- Step 4: Use printed x^2 as an appropriate noncentrality parameter and computer power.

Muthén & Curran (1997): Artificial and real data situations.

Input For Step 1 Of Power Calculation

```
TITLE:     Power calculation for a growth model
           Step 1: Computing the population means and
           covariance matrix
DATA:      FILE IS artific.dat;
           TYPE IS MEANS COVARIANCE;
           NOBSERVATIONS = 500;
VARIABLE:  NAMES ARE y1-y4;
MODEL:     i s | y1@0 y2@1 y3@2 y4@3;
           i@.5;
           s@.1;
           i WITH s@0;
           y1-y4@.5;
OUTPUT:    STANDARDIZED RESIUDAL;
```

Data For Step 1
Of Power Calculation (Continued)

```
0  0  0  0
1
0  1
0  0  1
0  0  0  1
```


Input For Step 2
Of Power Calculation

```
TITLE:     Power calculation for a growth model
           Step 2: Analyzing the population means and
           covariance matrix to check that parameters are
           recovered
DATA:      FILE IS pop.dat;
           TYPE IS MEANS COVARIANCE;
           NOBSERVATIONS = 500;
VARIABLE:  NAMES ARE y1-y4;
MODEL:     i s | y1@0 y2@1 y3@2 y4@3;
OUTPUT:    STANDARDIZED RESIUDAL;
```

Data For Step 2
Of Power Calculation (Continued)

Data From Step 1 Residual Output

```
0   .2   .4   .6
1
.5   1.1
.5   .7   1.4
.5   .8   1.1   1.9
```


Input For Step 3
Of Power Calculation

```
TITLE:     Power calculation for a growth model
           Step 3: Analyzing the population means and
           covariance matrix with a misspecified model
DATA:      FILE IS pop.dat;
           TYPE IS MEANS COVARIANCE;
           NOBSERVATIONS = 50;
VARIABLE:  NAMES ARE y1-y4;
MODEL:     i s | y1@0 y2@1 y3@2 y4@3;
OUTPUT:    STANDARDIZED RESIUDAL;
```

Step 4 Of Power Calculation

Output Excerpt From Step 3

```
Chi-Square Test of Model Fit

          Value                          9.286
          Degrees of Freedom                 6
          P-Value                        .1580
```

Power Algorithm in SAS

```
DATA POWER;
DF=1; CRIT=3.841459;
LAMBDA=9.286;
Power=(1 – (PROBCHI(CRIT, DF, LAMBDA)));
RUN;
```


Step 4 Of Power Calculation (Continued)

Results From Power Algorithm

SAMPLE SIZE	POWER
44	0.80
50	0.85
100	0.98
200	0.99

Note: **Non-centrality parameter =**
printed chi-square value from Step 3 =
2*sample size*F

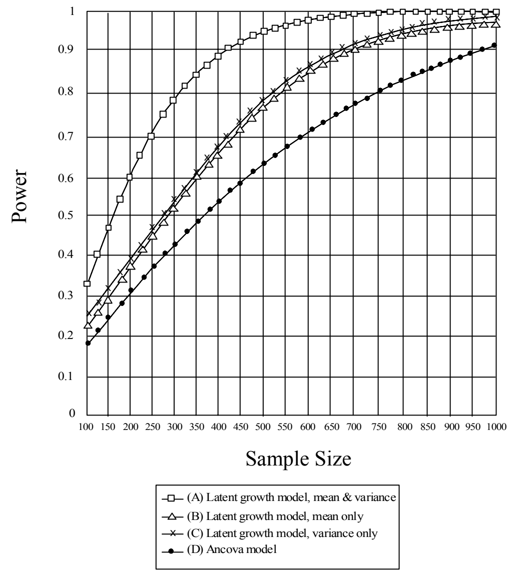


Figure 6. Power to detect a main effect of *ES* = .20 assessed at Time 5.

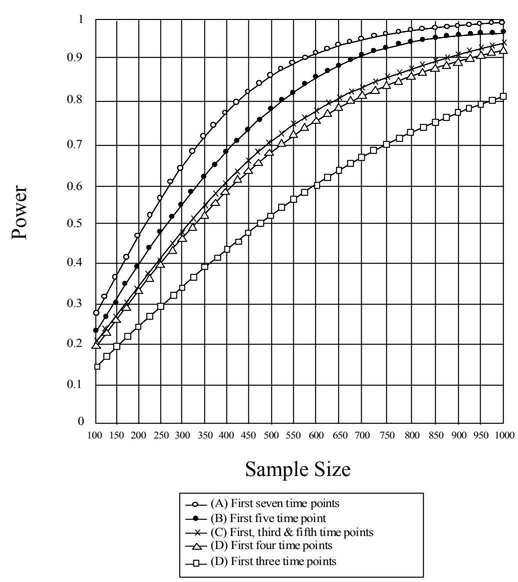


Figure 7. Power to detect a main effect of *ES* = .20 assessed at Time 5 varying as a function of total number of measurement occasions.

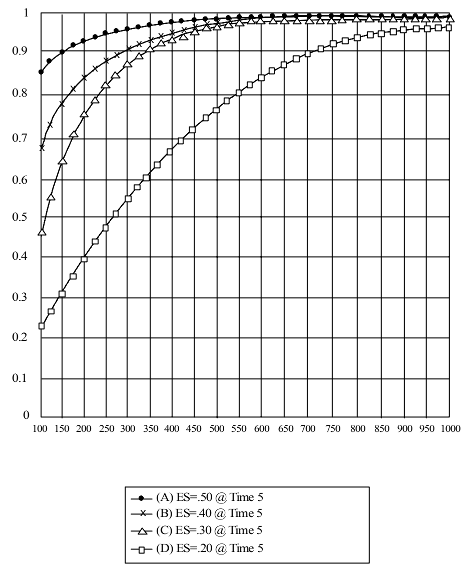


Figure 8. Power to detect various effect sizes assessed at Time 5 based on the first five measurement occasions

Power Estimation For Growth Models Using Monte Carlo Studies

Muthén, L.K. and Muthén, B.O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. <u>Structural Equation Modeling</u>, 4, 599-620.

Input Power Estimation For Growth Models Using Monte Carlo Studies

```
TITLE:        This is an example of a Monte Carlo
              simulation study for a linear growth model
              for a continuous outcome with missing data
              where attrition is predicted by time-
              invariant covariates (MAR)
MONTECARLO:   NAMES ARE y1-y4 x1 x2;
              NOBSERVATIONS = 500;
              NREPS = 500;
              SEED = 4533;
              CUTPOINTS = x2(1);
              MISSING = y1-y4;
```


Input Power Estimation For Growth Models Using Monte Carlo Studies (Continued)

```
MODEL POPULATION:     x1-x2@1;
                      [x1-x2@0];
                      i s | y1@0 y2@1 y3@2 y4@3;
                      [i*1 s*2];
                      i*1; s*.2; i WITH s*.1;
                      y1-y4*.5;
                      i ON x1*1 x2*.5;
                      s ON x1*.4 x2*.25;
MODEL MISSING:        [y1-y4@-1];
                      y1 ON x1*.4 x2*.2;
                      y2 ON x1*.8 x2*.4;
                      y3 ON x1*1.6 x2*.8;
                      y4 ON x1*3.2 x2*1.6;
```

Input Power Estimation For Growth Models Using Monte Carlo Studies (Continued)

```
MODEL:       i s | y1@0 y2@1 y3@2 y4@3;
             [i*1 s*2];
             i*1; s*.2; i WITH s*.1;
             y1-y4*.5;
             i ON x1*1 x2*.5;
             s ON x1*.4 x2*.25;
OUTPUT:      TECH9;
```


Output Excerpts Power Estimation For Growth Models Using Monte Carlo Studies

Model Results

		ESTIMATES		S.E.	M. S. E.	95%	%Sig
	Population	Average	Std. Dev.	Average		Cover	Coeff
I ON							
X1	1.000	1.0032	0.0598	0.0579	0.0036	0.936	1.000
X2	0.500	0.5076	0.1554	0.1570	0.0241	0.952	0.908
S ON							
X1	0.400	0.3980	0.0366	0.0349	0.0013	0.936	1.000
X2	0.250	0.2469	0.0865	0.0877	0.0075	0.938	0.830

Cohort-Sequential Designs and Power

Considerations:

- Model identification
- Number of timepoints needed substantively
- Number of years of the study
- Number of cohorts: More gives longer timespan but greater risk of cohort differences
- Number of measurements per individual
- Number of individuals per cohort
- Number of individuals per age

Tentative conclusion:

Power most influenced by total timespan, not the number of measures per cohort

Survival Analysis

Survival Analysis

- Discrete-time
  - Infrequent measurement (monthly, annually)
  - Limited number of time periods
- Continuous-time
  - Frequent measurement (hourly, daily)
  - Large number of time points

Discrete-Time Survival Analysis

Discrete-Time Survival Analysis

Other terms: event history analysis, time-to-event.

References: Allison (1984), Singer & Willet (1993), Vermunt (1997).

- Setting
  - Discrete time periods (e.g. grade), non-repeatable event (e.g. onset of drug use)
  - Uncensored and censored individuals
  - Time-invariant and time-varying covariates
- Aim
  - Estimate survival curves and hazard probabilities
  - Relate survival to covariates
- Generalized models using multiple latent classes of survival
  - Long-term survivors with zero hazard
  - Growth mixture modeling in combination with survival analysis
- Application: School removal and aggressive behavior in the classroom (Muthén & Masyn, 2005). Grade 1 sample, n = 403 control group children.

Data For Discrete-Time Survival Analysis

- Single non-repeatable event – data collection ends for individual i when the event has been observed, where j_i is the last time period of data collection for individual i
- u_j (j = 1, 2,…, j_i) are binary 0/1 event history indicators, where u_{ij} = 1 if individual i experiences an event in time period j

Event history information entered into an r x 1 data vector $\mathbf{u}_i$ where r denotes the maximum value of j_i over all individuals and where u = 999 denotes missing data.

Data For Discrete-Time Survival Analysis (Continued)

Examples:

- An individual who is censored after time period five ($j_i = 6$)

$$(\ 0\ \ 0\ \ 0\ \ 0\ \ 0\ ),$$

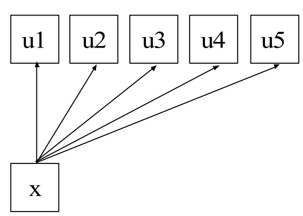
- An individual who experiences the event in period four ($j_i = 4$)

$$(\ 0\ \ 0\ \ 0\ \ 1\ \ 999\ ),$$

- An individual who drops out after period three, i.e. is censored during period four before the study ends ($j_i = 4$)

$$(\ 0\ \ 0\ \ 0\ \ 999\ \ 999\ ).$$

Model For Discrete-Time Survival Analysis



Hazard, Survival, Likelihood

The hazard is the probability of experiencing the event in the time period j given that it was not experienced prior to j. Letting the time of the event for individual i be denoted T_i, the logistic hazard function with q covariates x is

$$P(u_{ij}=1) = P(T_i = j \mid T_i \geq j) = h_{ij} = \frac{1}{1 + e^{-(-\tau_j + \kappa_j x_i)}}, \tag{49}$$

where a proportional-odds assumption is obtained by dropping the j subscript for κ_j. The survival function is

$$S_{ij} = \prod_{k=1}^{j} (1 - h_{ik}). \tag{50}$$

Hazard, Survival, Likelihood (Continued)

The likelihood $L = \prod_{i=1}^{n} l_i$, where

$$l_i = \prod_{j=1}^{j_i} h_{ij}^{u_{ij}} (1 - h_{ij})^{1-u_{ij}}. \tag{51}$$

A censored individual is observed with probability

$$l_i = \prod_{j=1}^{j_i} (1 - h_{ij}). \tag{52}$$

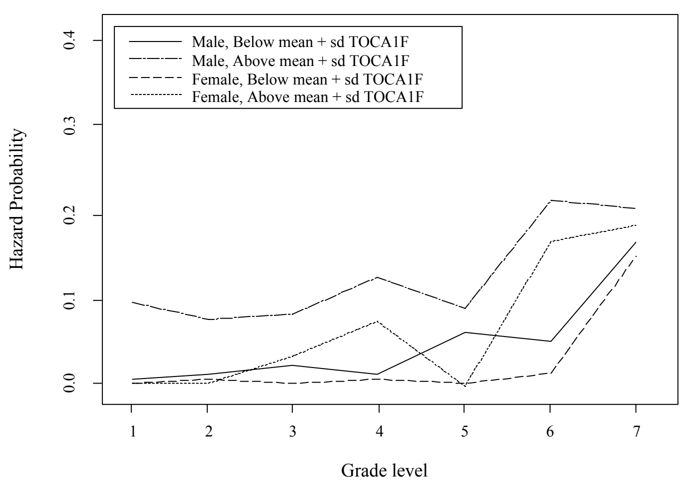
An uncensored individual experiences the event in time period j_i with probability

$$l_i = h_{ij_i} \prod_{j=1}^{j_i - 1} (1 - h_{ij}). \tag{53}$$

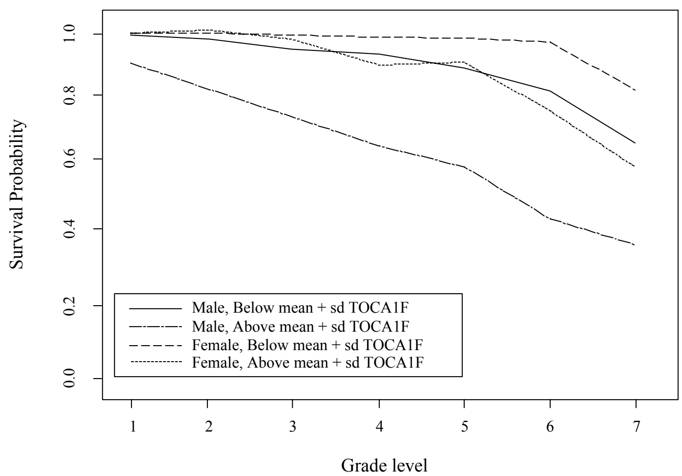
School Removal Data (n = 403)

Gender	Grade	No School removal	At least one school removal	Sample Hazard
Male	1	0	4	4/200 = 0.02
	2	0	5	5/196 = 0.03
	3	0	7	7/191 = 0.04
	4	0	6	6/184 = 0.03
	5	0	14	14/178 = 0.08
	6	0	14	14/164 = 0.08
	7	122	28	28/150 = 0.19
	Total	122	78	200
Female	1	0	0	0/203 = 0.00
	2	0	1	1/203 = 0.005
	3	0	1	1/202 = 0.005
	4	0	3	3/201 = 0.01
	5	0	1	1/198 = 0.005
	6	0	9	9/197 = 0.05
	7	157	31	31/188 = 0.16
	Total	157	46	203

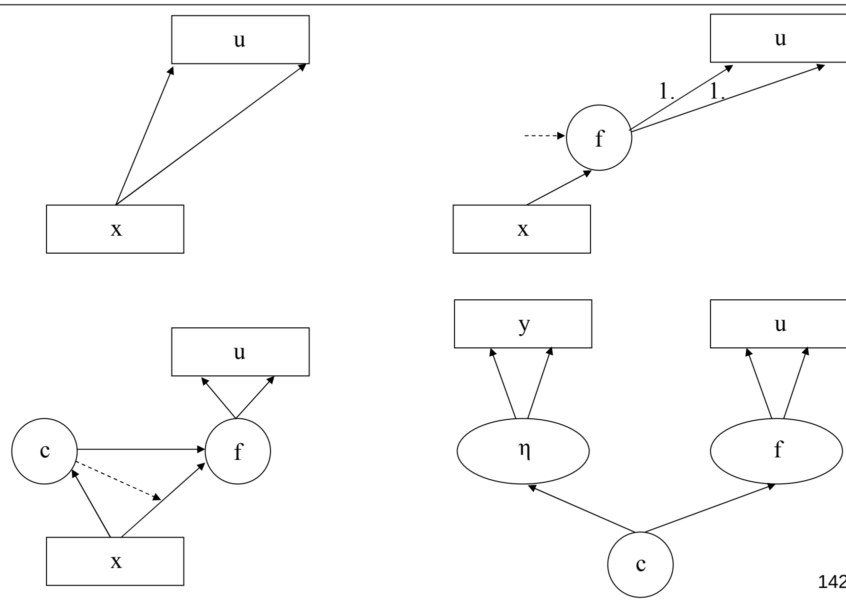
Hazard Sample Estimates



Survival Sample Estimates



Discrete-Time Survival Models



Input For A Discrete-Time Survival Analysis

```
TITLE:        A discrete-time survival analysis
DATA:         FILE IS survival.dat;
VARIABLE:     NAMES ARE u1-u7 race lunch cavtoca cavlunch
              cntrlg y1 gender;
              MISSING are all (999);
              CATEGORICAL ARE u1-u7;
ANALYSIS:     ESTIMATOR = MLR;
MODEL:        f BY u1-u7@1;
              f ON race-gender;
              f@0;
OUTPUT:       TECH1 TECH8;
```


Output Excerpts
A Discrete-Time Survival Analysis

Tests Of Model Fit

```
Loglikelihood
              H0 Value                        -388.074

Information Criteria
              Number of Free Parameters             14
              Akaike (AIC)                     804.147
              Bayesian (BIC)                   860.132
              Sample-Sized Adjusted BIC        815.709
               (n* = (n + 2)/24)
```

Output Excerpts
A Discrete-Time Survival Analysis (Continued)

Model Results

	Estimates	S.E.	Est./S.E.
Thresholds			
U1$1	4.707	0.694	6.782
U2$1	4.118	0.782	5.269
U3$1	3.764	0.658	5.725
U4$1	3.588	0.648	5.537
U5$1	2.958	0.677	4.371
U6$1	2.382	0.625	3.809
U7$1	1.048	0.609	1.721
F ON			
RACE	-0.449	0.379	-1.183
LUNCH	-0.136	0.268	-0.506
CAVTOCA	-1.104	0.295	-3.738
CAVLUNCH	1.571	0.476	3.302
CNTRLG	-0.336	0.213	-1.578
Y1	0.783	0.119	6.566
GENDER	-0.700	0.206	-3.402

Multiple Latent Classes

Muthén & Masyn (2005)

Unobserved heterogeneity in hazard and survival

- Long-term survivors (one class has zero hazards, non-zero long-term survival probability)
- Latent classes of survival
- Growth mixtures and survival

Example: Long-term survivors

Individuals who are not censored, i.e. who experience the event within the observation period, are not long-term survivors (known latent class membership).

Two different latent classes of censored individuals:

$$\textit{Eventually experiences the event} : 0 \ldots 0 | 0 \ldots 0\ 1$$
$$\textit{Long-term survivor} : 0 \ldots 0 | 0 \ldots 0$$

Input For A Two-Class Discrete-Time Survival Analysis

```
TITLE:      A 2-class discrete-time survival analysis in a
            mixture modeling framework including long-term
            survivors
DATA:       FILE IS long.sav;
VARIABLE:   NAMES ARE u1-u7 race lunch cavtoca cavlunch
            cntrlg y1 gender t1 t2;

            MISSING ARE ALL (999);

            CATEGORICAL ARE u1-u7;
            CLASSES = c(2);

            TRAINING = t1 t2;
ANALYSIS:   TYPE = MIXTURE;
```


Input For A Two-Class Discrete-Time Survival Analysis (Continued)

```
MODEL:       %OVERALL%

             f BY u1-u7@1;
             f ON race-gender;
             [f@0];
             c#1 ON race-gender;

             %c#1% ! class of non-long-term survivors

             [u1$1*4 u2$1*3 u3$1*3 u4$1*3.5 u5$1*2.5 u6$1*4 u7$1*1];
             [f@0];
             f ON race-gender;

             %c#2% ! class of long-term survivors

             [u1$1-u7$1@10];
             f ON race-gender@0;
OUTPUT:      PATTERNS TECH1 TECH8;
```

Output Excerpts A Two-Class Discrete-Time Survival Analysis

Tests Of Model Fit

```
Loglikelihood
             H0 Value                          -375.951

Information Criteria
             Number of Free Parameters               22
             Akaike (AIC)                       795.903
             Bayesian (BIC)                     883.879
             Sample-Sized Adjusted BIC          814.071
              (n* = (n + 2)/24)
             Entropy                              0.644
```


Output Excerpts A Two-Class Discrete-Time Survival Analysis (Continued)

Classification Information

```
FINAL CLASS COUNTS AND PROPORTIONS OF TOTAL SAMPLE SIZE

  Class 1          201.96062         0.50114
  Class 2          201.03938         0.49886

CLASSIFICATION OF INDIVIDUALS BASED ON THEIR MOST LIKELY CLASS
MEMBERSHIP

Class Counts and Proportions

  Class 1                190         0.47146
  Class 2                213         0.52854

Average Latent Class Probabilities for Most Likely Latent Class
Membership (Row) by Latent Class (Column)

                1           2

  Class 1      0.916     0.084
  Class 2      0.131     0.869
```

Output Excerpts A Two-Class Discrete-Time Survival Analysis (Continued)

Model Results

	Estimates	S.E.	Est./S.E.
Class 1			
Thresholds			
U1$1	4.590	0.809	5.673
U2$1	4.060	0.928	4.376
U3$1	3.654	0.804	4.542
U4$1	3.402	0.756	4.498
U5$1	2.656	0.764	3.477
U6$1	1.866	0.708	2.634
U7$1	-0.026	0.841	-0.030

Output Excerpts A Two-Class Discrete-Time Survival Analysis (Continued)

C#1 ON			
RACE	-1.099	0.552	-1.990
LUNCH	0.446	0.712	0.627
CAVTOCA	-2.459	0.824	-2.983
CAVLUNCH	2.907	1.470	1.977
CNTRLG	-0.101	0.498	-0.204
Y1	0.913	0.301	3.036
GENDER	0.150	0.665	0.226
Intercepts			
C#1	1.773	1.411	1.257
Class 1			
F ON			
RACE	0.882	0.511	1.728
LUNCH	-0.679	0.550	-1.236
CAVTOCA	-0.234	0.642	-0.365
CAVLUNCH	0.540	0.809	0.667
CNTRLG	-0.441	0.355	-1.242
Y1	0.605	0.218	2.774
GENDER	-1.141	0.510	-2.237

Further Readings On Discrete-Time Survival Analysis

Allison, P.D. (1984). Event history analysis. Regression for Longitudinal Event Data. Quantitative Applications in the Social Sciences, No. 46. Thousand Oaks: Sage Publications.

Masyn, K. E. (2008). Modeling measurement error in event occurrence for single, non-recurring events in discrete-time survival analysis. In Hancock, G. R., & Samuelsen, K. M. (Eds.), Advances in latent variable mixture models, pp. 105-145. Charlotte, NC: Information Age Publishing, Inc.

Muthén, B. & Masyn, K. (2005). Discrete-time survival mixture analysis. Journal of Educational and Behavioral Statistics, 30, 27-28.

Singer, J.D., & Willett, J.B. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. Journal of Educational Statistics, 18(2), 155-195.

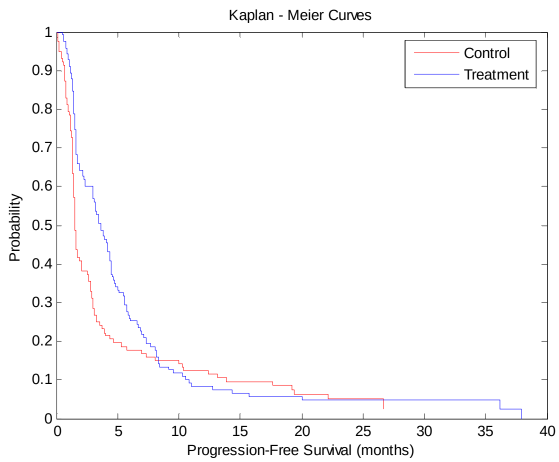
Singer, J. D. & Willett, J. B. (2003). Applied longitudinal analysis.

Modeling change and event occurrence. Oxford, UK: Oxford University Press.

Vermunt, J.K. (1997). Log-linear models for event histories. Advanced quantitative techniques in the social sciences, vol 8. Thousand Oaks: Sage Publications.

Continuous-Time Survival Analysis

Kaplan-Meier Survival Curves



Continuous-Time Survival Analysis

- T_0: time-to-event such as time to death
- I: time of censoring
- The survival variable T and the censoring indicator c are defined by

$$T = \min\{T_0, I\} \tag{1}$$

$$c = \begin{cases} 1 & \text{if } T_0 > I \\ 0 & \text{if } T_0 \le I \end{cases} \tag{2}$$

For example, $c = 1$ implies $T = I$, the time the individual leaves the sample

The Proportional Hazard Model

The proportional hazard (PH) model specifies that the hazard function is proportional to the baseline hazard function,

$$h(t) = \lambda(t)\, Exp(\beta X) \qquad (3)$$

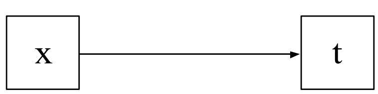
$$\log h(t) = \log \lambda(t) + \beta X \qquad (4)$$

Two proportional hazard models:

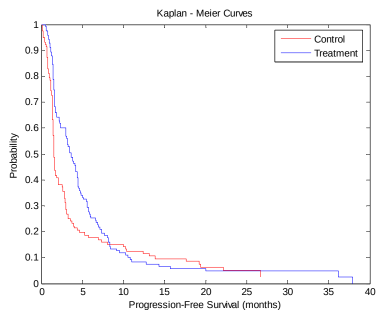
- Nonparametric shape for the baseline hazard function $\lambda(t)$: Cox regression
- Parametric model for the baseline hazard function $\lambda(t)$: parametric PH model

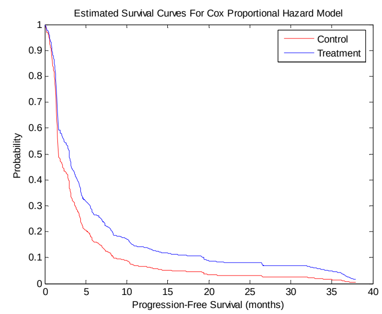
Example 6.21: Continuous-Time Survival Analysis Using The Cox Regression Model

```
TITLE:     this is an example of a continuous-time survival
           analysis using the Cox regression model
DATA:      FILE = ex.6.21.dat;
VARIABLE:  NAMES = t x tc;
           SURVIVAL = t (ALL);
           TIMECENSORED = tc (0 = NOT 1 = RIGHT);
MODEL:     t ON x;
```

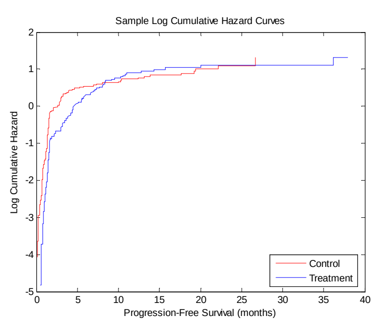


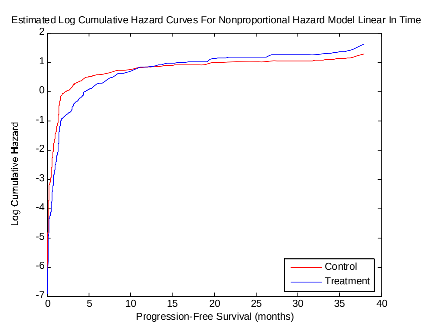
Survival Curves For Kaplan-Meier Versus Cox Proportional Hazard Model



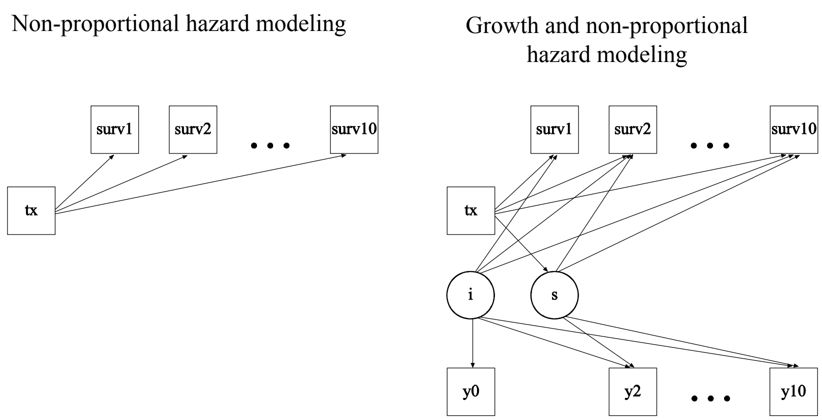


Log Cumulative Hazard Curves For Kaplan-Meier Versus Nonproportional Hazard Model





Survival Analysis With Non-Proportional Hazard Models



Continuous-Time Survival Data

t	x	c
7.330493	-0.378137	1.000000
0.894182	-0.880031	1.000000
1.219113	0.369423	0.000000
0.134073	1.886903	0.000000
0.598567	1.118025	0.000000
0.725646	0.642068	0.000000
1.637967	-0.324017	0.000000
5.534057	-0.760867	0.000000
3.316749	0.194822	1.000000
4.176435	-0.311791	1.000000

Event occurred at time 1.219113

Translating Continuous-Time Survival Data To Discrete-Time Survival Data

```
VARIABLE:   NAMES = t x c;
            ! t =time of death or censoring
            ! c = not censored (0), censored (1)
            CATEGORICAL = u1-u8(*);
            USEVAR = x u1-u8;
DATA:       FILE = surveq1.dat;
            VARIANCE = NOCHECK;
DEFINE:     IF (t>2) THEN u1=0;
            IF ((t>0) .AND. (t<2)) THEN u1=1-c;
            ! u1 = 0 if c = 1, i.e. censoring time between 0 and 2
            ! u1 = 1 if person died then

            IF (t>4) THEN u2=0;
            IF ((t>2) .AND. (t<4)) THEN u2=1-c;
            IF (t<2) THEN u2=_missing;
            ! u2 is missing either because u1 = 1 or because
            ! u1 = 0 and c = 1
```


Translating Continuous-Time Survival Data To Discrete-Time Survival Data (Continued)

```
            IF (t>6) THEN u3=0;
            IF ((t>4) .AND. (t<6)) THEN u3=1-c;
            IF (t<4) THEN u3=_missing;

            IF (t>8) THEN u4=0;
            IF ((t>6) .AND. (t<8)) THEN u4=1-c;
            IF (t<6) THEN u4=_missing;

            IF (t>10) THEN u5=0;
            IF ((t>8) .AND. (t<10)) THEN u5=1-c;
            IF (t<8) THEN u5=_missing;

            IF (t>12) THEN u6=0;
            IF ((t>10) .AND. (t<12)) THEN u6=1-c;
            IF (t<10) THEN u6=_missing;
```

Translating Continuous-Time Survival Data To Discrete-Time Survival Data (Continued)

```
           IF (t>14) THEN u7=0;
           IF ((t>12) .AND. (t<14)) THEN u7=1-c;
           IF (t<12) THEN u7=_missing;

           IF (t>16) THEN u8=0;
           IF ((t>14) .AND. (t<16)) THEN u8=1-c;
           IF (t<14) THEN u8=_missing;
MODEL:     u1-u8 ON x*1 (1);
ANALYSIS:  ESTIMATOR = MLR;
           TYPE = MISSING;
```


Further Readings For Continuous-Time Survival Analysis

Asparouhov, T., Masyn, K. & Muthen, B. (2006). Continuous time survival in latent variable models. Proceedings of the Joint Statistical Meeting in Seattle, August 2006. ASA section on Biometrics, 180-187.

Larsen, K. (2004). Joint analysis of time-to-event and multiple binary indicators of latent classes, Biometrics, 60(1), 85–92.

Larsen, K. (2005). The Cox proportional hazards model with a continuous latent variable measured by multiple binary indicators, Biometrics, 61(4), 1049–1055.

Singer, J.D., & Willett, J.B. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. Journal of Educational Statistics, 18(2), 155-195.

Singer, J. D. & Willett, J. B. (2003). Applied longitudinal analysis. Modeling change and event occurrence. Oxford, UK: Oxford University Press.

Analysis With Missing Data

Analysis With Missing Data

Used when individuals are not observed on all outcomes in the analysis to make the best use of all available data and to avoid biases in parameter estimates, standard errors, and tests of model fit.

Types of Missingness

- MCAR -- missing completely at random
  - Variables missing by chance
  - Missing by randomized design
  - Multiple cohorts assuming a single population
- MAR -- missing at random
  - Missingness related to observed variables
  - Missing by selective design
- Non-Ignorable (NMAR)
  - Missingness related to values that would have been observed
  - Missingness related to latent variables

Estimation With Missing Data

Types of Estimation (Little & Rubin, 2002)

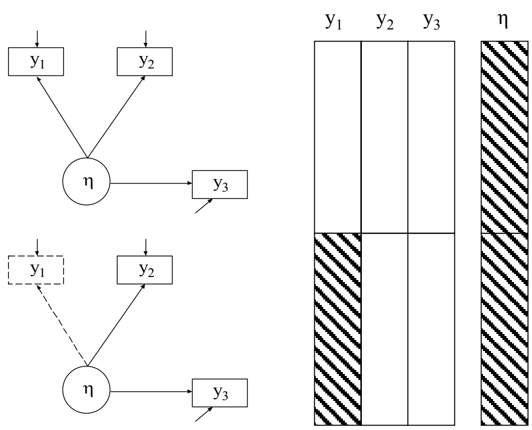
- Estimation using listwise deleted sample
  - When MCAR is true, parameter estimates and s.e.'s are consistent but estimates are not efficient
  - When MAR is true but not MCAR, parameter estimates and s.e.'s are not consistent
- Maximum likelihood using all available data
  - When MCAR or MAR is true, parameter estimates and s.e.'s are consistent and estimates are efficient
- Selection and pattern-mixture modeling – used for non-ignorable missingness
- Imputation
  - Mean and regression imputation – underestimation of variances and covariances
  - Multiple imputation using all available data – a Bayesian approach – credibility intervals are Bayesian justifiable under MCAR and MAR

Weighted Least Squares Estimation With Missing Data

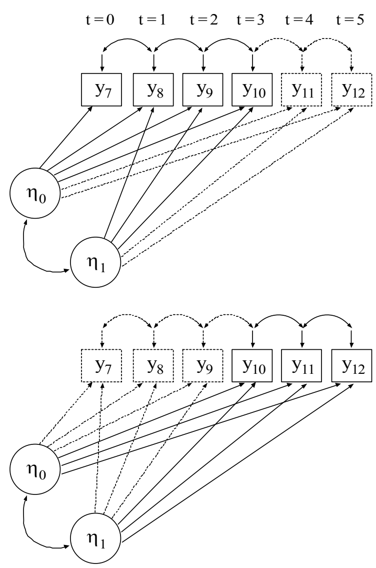
Weighted least squares for categorical and censored outcomes

- Assumes MCAR when there are no covariates
- Allows MAR when missingness is a function of covariates

MCAR: Missing By Design

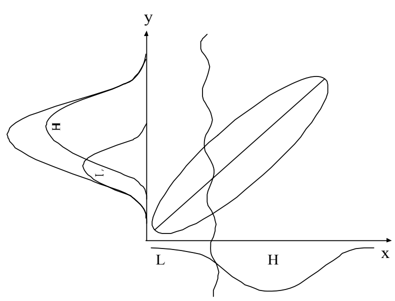


Two-Cohort Growth Model



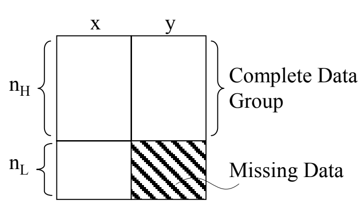
MAR

MAR: Bivariate, Monotone Missing Case



$y_i = \alpha + \beta x_i + \zeta_i$
$E(\zeta) = 0,\ V(\zeta) = \sigma_\zeta^2$
$E(x) = \mu_x,\ V(x) = \sigma_x^2$

- Data Matrix:



Missing At Random (MAR): Missing On y In Bivariate Normal Case

$$\hat{\mu}_x = \sum_{i=1}^{n_L + n_H} x_i / (n_L + n_H) = \frac{n_L \bar{x}_L + n_H \bar{x}_H}{n_L + n_H}, \qquad (52)$$

$$\hat{\sigma}_{xx} = \sum_{i=1}^{n_L + n_H} (x_i - \hat{\mu}_x)^2 / (n_L + n_H). \qquad (53)$$

Missing At Random (MAR): Missing On y In Bivariate Normal Case (Continued)

Consider the regression

$$y_i = \alpha + \beta x_i + \zeta_i \qquad (54)$$

estimated by the complete-data (listwise present) sample (sample size n_H)

$$\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x}, \qquad (55)$$

$$\hat{\beta} = s_{yx} / s_{xx}, \qquad (56)$$

$$\hat{\sigma}_{\zeta\zeta} = s_{yy} - s_{yx}^2 / s_{xx}. \qquad (57)$$

This gives the ML estimates of μ_y and σ_{yy}, adjusting the complete-data sample statistics:

$$\hat{\mu}_y = \hat{\alpha} + \hat{\beta} \hat{\mu}_x = \bar{y} + \hat{\beta} (\hat{\mu}_x - \bar{x}), \qquad (58)$$

$$\hat{\sigma}_{yy} = \hat{\sigma}_{\zeta\zeta} + \hat{\beta}^2 \hat{\sigma}_{xx} = s_{yy} + \hat{\beta}^2 (\hat{\sigma}_{xx} - s_{xx}). \qquad (59)$$

Correlates Of Missing Data

- MAR is more plausible when the model includes covariates influencing missing data
- Correlates of missing data may not have a "causal role" in the model, i.e. not influencing dependent variables, in which case including them as covariates can bias model estimates
  - Multiple imputation (Bayes; Schafer, 1997) with two different sets of observed variables
    - Imputation model
    - Analysis model
  - Modeling (ML)
    - Including missing data correlates not as x variables but as "y variables," freely correlated with all other observed variables

Recent overview in Schafer & Graham (2002).

Missing On X

- Regular modeling concerns the conditional distribution

$$[y \mid x] \qquad (1)$$

that is, as in regular regression the marginal distribution of $[x]$ is not involved. This is fine if there is no missing on x in which case considering

$$[y \mid x]$$

gives the same estimates as (Joreskog & Goldberger, 1975) considering the joint distribution

$$[y, x] = [y \mid x] \; [x]$$

Missing On X (Continued)

- With missing on x, ML under MAR must make a distributional assumption about $[x]$, typically normality. The modeling then concerns

  $$[y, x] = [y \mid x]\ [x] \tag{2}$$

  which with missing on $[x]$ is an expanded model that makes stronger assumptions as compared to (1).

- The LHS of (2) shows that y and x are treated the same - they are both "*y* variables" in Mplus terminology. This is the default in Mplus when all *y*'s are continuous. In other cases, *x*'s can be turned into "*y*'s" e.g. by the model statement

  x1-xq;

Technical Aspects Of Ignorable Missing Data: ML Under MAR

$$\textit{Likelihood:}\ \sum_{i=1}^{n} \log\,[\mathbf{y_i} \mid \mathbf{x_i}]. \tag{87}$$

With missing data on $\mathbf{y}$, the i^{th} term of (87) expands into

$$[\mathbf{y}_i^{obs}, \mathbf{y}_i^{mis}, \mathrm{m}_i \mid \mathbf{x}_i], \tag{88}$$

where $\mathbf{m}_i$ is a 0/1 indicator vector of the same length as $\mathbf{y}_i$.

The likelihood focuses on the observed variables,

$$[\mathbf{y}_i^{obs}, \mathbf{m}_i \mid \mathbf{x}_i] = \int [\mathbf{y}_i^{obs}, \mathbf{y}_i^{mis} \mid \mathbf{x}_i]\ [\mathbf{m}_i \mid \mathbf{y}_i^{obs}\ \mathbf{y}_i^{mis}, \mathrm{x}_i]\ d\mathbf{y}_i^{mis}, \tag{89}$$

which, when assuming that missingness is not a function of $\mathbf{y}_i^{mis}$ (that is, assuming MAR),

Technical Aspects Of Ignorable Missing Data: ML Under MAR (Continued)

$$= \int [\mathbf{y}_i^{obs}, \mathbf{y}_i^{mis} \mid \mathbf{x}_i] \; d\mathbf{y}_i^{mis} \; [\mathbf{m}_i \mid \mathbf{y}_i^{obs}, \mathbf{x}_i], \qquad (90)$$

$$= [\mathbf{y}_i^{obs} \mid \mathbf{x}_i] \, [\mathbf{m}_i \mid \mathbf{y}_i^{obs}, \mathbf{x}_i]. \qquad (91)$$

With distinct parameter sets in (91), the last term can be ignored and maximization can focus on the $[\mathbf{y}_i^{obs} \mid \mathbf{x}_i]$ term. This leads to the standard MAR ignorable missing data procedure.

AMPS Data

The data are taken from the Alcohol Misuse Prevention Study (AMPS). Forty-nine schools with a total of 2,666 students participated in the study. Students were measured seven times starting in the Fall of Grade 6 and ending in the Spring of Grade 12.

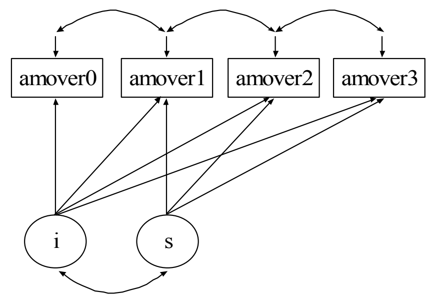
Data for the analysis include the average of three items related to alcohol misuse:

During the past 12 months, how many times did you

drink more than you planned to?
feel sick to your stomach after drinking?
get very drunk?

Responses: (0) never, (1) once, (2) two times, (3) three or more times

Four of the seven timepoints are studied: Fall Grade 6, Spring Grade 6, Spring Grade 7, and Spring Grade 8.



Input For AMPS Growth Model With Missing Data

```
TITLE:     AMPS growth model with missing data
DATA:      FILE IS amps.dat;
VARIABLE:  NAMES ARE caseid
           amover0 ovrdrnk0 illdrnk0 vrydrn0
           amover1 ovrdrnk1 illdrnk1 vrydrn1
           amover2 ovrdrnk2 illdrnk2 vrydrn2
           amover3 ovrdrnk3 illdrnk3 vrydrn3
           amover4 ovrdrnk4 illdrnk4 vrydrn4
           amover5 ovrdrnk5 illdrnk5 vrydrn5
           amover6 ovrdrnk6 illdrnk6 vrydrn6;
           USEV = amover0 amover1 amover2 amover3;
           MISSING = ALL (999);
```

Input For AMPS Growth Model With Missing Data (Continued)

```
MODEL:     i s | amover0@0 amover1@1 amover2@3 amover3*5;
           amover1-amover3 PWITH amover0-amover2;

OUTPUT:    PATTERNS SAMPSTAT MODINDICES STANDARDIZED;
```


Output Excerpts AMPS Growth Model With Missing Data

Summary of Data

Number of patterns 15

SUMMARY OF MISSING DATA PATTERNS

MISSING DATA PATTERNS

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
AMOVER0	x	x	x	x	x	x	x	x							
AMOVER1	x	x	x	x					x	x	x	x			
AMOVER2	x	x			x	x			x	x			x	x	
AMOVER3	x		x		x		x		x		x		x		x

MISSING DATA PATTERN FREQIENCIES

Pattern	Frequency	Pattern	Frequency	Pattern	Frequency
1	685	6	29	11	104
2	143	7	11	12	237
3	73	8	64	13	6
4	164	9	866	14	1
5	65	10	208	15	3

Output Excerpts AMPS Growth Model With Missing Data (Continued)

```
COVARIANCE COVERAGE OF DATA
Minimum covariance coverage value   0.100
PROPORTION OF DATA PRESENT
```

Covariance Coverage

	AMOVER0	AMOVER1	AMOVER2	AMOVER3
AMOVER0	0.464			
AMOVER1	0.401	0.933		
AMOVER2	0.347	0.715	0.753	
AMOVER3	0.314	0.650	0.610	0.682

Output Excerpts AMPS Growth Model With Missing Data (Continued)

Tests Of Model Fit

```
Chi-square Test of Model Fit

           Value                          0.011
           Degrees of Freedom                 1
           P-Value                       0.9177

RMSEA (Root Mean Square Error Of Approximation)

           Estimate                       0.000
           90 Percent C.I.                0.000  0.019
           Probability RMSEA <= .05        0.997
```

Output Excerpts AMPS Growth Model With Missing Data (Continued)

Model Results

	Estimates	S.E.	Est./S.E.	Std	StdYX
I |					
AMOVER0	1.000	0.000	0.000	0.426	0.921
AMOVER1	1.000	0.000	0.000	0.426	0.774
AMOVER2	1.000	0.000	0.000	0.426	0.645
AMOVER3	1.000	0.000	0.000	0.426	0.529
S |					
AMOVER0	0.000	0.000	0.000	0.000	0.000
AMOVER1	1.000	0.000	0.000	0.109	.198
AMOVER2	3.000	0.000	0.000	0.327	0.494
AMOVER3	6.244	0.426	14.645	0.680	0.843
S I					
WITH	-0.007	0.003	-2.278	-0.146	-0.146
AMOVER1 WITH					
AMOVER0	-0.022	0.011	-2.010	-0.022	-0.085
AMOVER2 WITH					
AMOVER1	0.017	0.007	2.505	0.017	0.047
AMOVER3 WITH					
AMOVER2	-0.001	0.027	-0.050	-0.001	-0.003

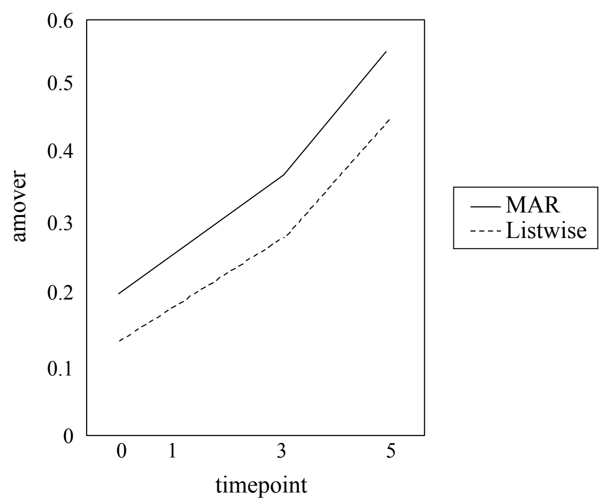
Output Excerpts AMPS Growth Model With Missing Data (Continued)

Residual Variances					
AMOVER0	0.033	0.013	2.509	0.033	0.152
AMOVER1	0.123	0.011	10.950	0.123	0.406
AMOVER2	0.190	0.017	11.461	0.190	0.433
AMOVER3	0.091	0.068	1.340	0.091	0.140
Variances					
I	0.182	0.014	12.891	1.000	1.000
S	0.012	0.002	5.378	1.000	1.000
Means					
I	0.200	0.010	19.391	0.469	0.469
S	0.057	0.005	11.858	0.520	0.520
Intercept					
AMOVER0	0.000	0.000	0.000	0.000	0.000
AMOVER1	0.000	0.000	0.000	0.000	0.000
AMOVER2	0.000	0.000	0.000	0.000	0.000
AMOVER3	0.000	0.000	0.000	0.000	0.000

Output Excerpts AMPS Growth Model With Missing Data (Continued)

```
R-SQUARE

  Observed
  Variable        R-Square

   AMOVER0           0.848
   AMOVER1           0.594
   AMOVER2           0.567
   AMOVER3           0.860
```


AMPS: Estimated Growth Curves



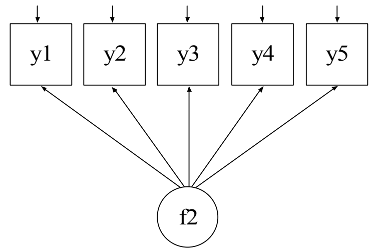
Missing Data Correlates Using ML

Missing Data Correlates

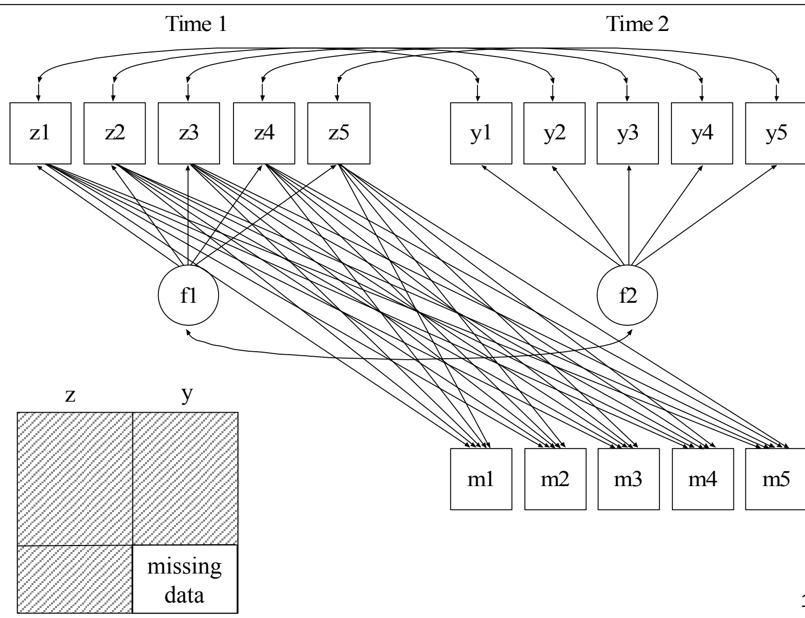
Handling missing-data-related variables that are different from analysis variables:

- Multiple imputation
- ML using missing data correlates
- References:
  - Collins, Schafer, Kam (2001) in Psych Methods
  - Graham (2003) in SEM
  - Enders and Peugh (2004) SEM
  - Savalei and Bentler (2007)

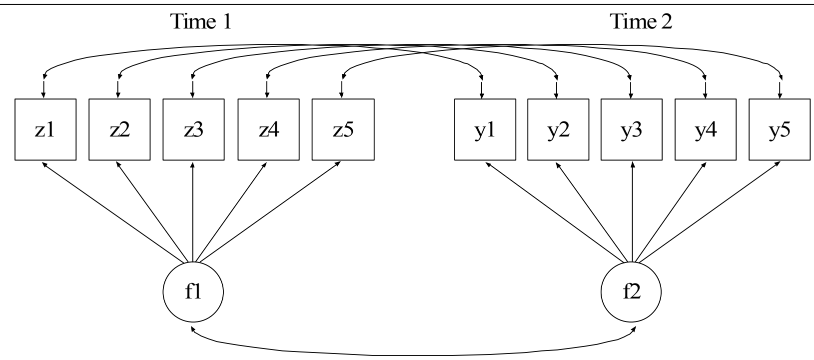
Model Of Interest

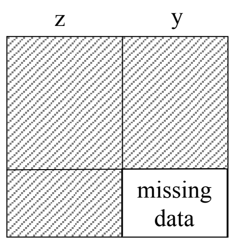


Data-Generating Model

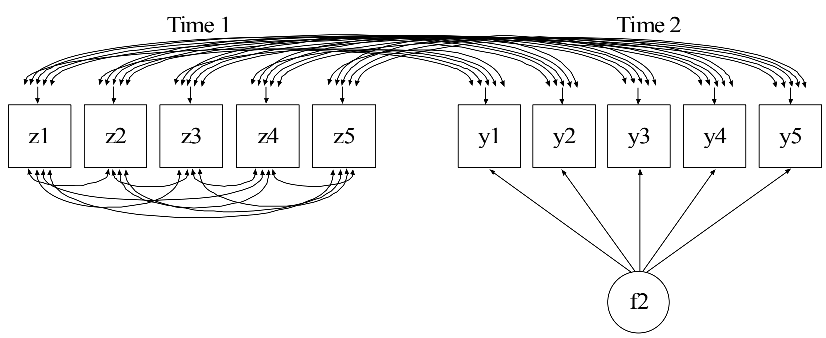


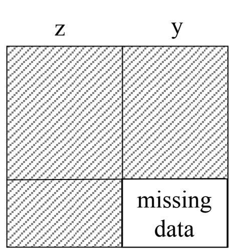
Full Analysis Model (ML Under MAR)





Simple Analysis Model (ML Under MAR)






```
VARIABLE:  NAMES = y1-y5 z1-z5;
           USEV = y1-y5;
           AUXILIARY = (M) z1-z5;
           MISSING = ALL(999);
MODEL:     f2 BY y1-y5;
```

Monte Carlo Simulation Results (N = 200)

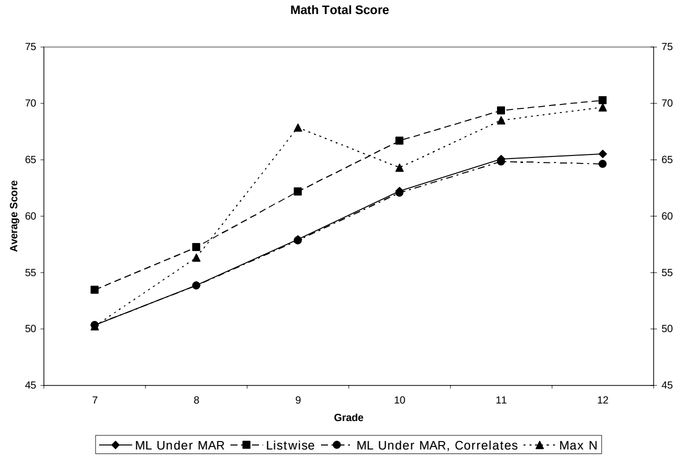
Y with z as aux

		ESTIMATES		S.E.	M.S.E.	95%
	Population	Average	Std. Dev.	Average		Cover
F2 BY						
Y1	0.7000	0.6914	0.0820	0.0766	0.0067	0.9400
Y2	0.7000	0.6854	0.0829	0.0766	0.0070	0.9300
Y3	0.7000	0.6947	0.0650	0.0767	0.0042	0.9600
Y4	0.7000	0.6934	0.0719	0.0767	0.0052	0.9600
Y5	0.7000	0.6911	0.0804	0.0761	0.0065	0.9300
Y alone						
F2 BY						
Y1	0.7000	0.6482	0.0869	0.0797	0.0102	0.890
Y2	0.7000	0.6392	0.0861	0.0803	0.0110	0.850
Y3	0.7000	0.6450	0.0751	0.0802	0.0086	0.900
Y4	0.7000	0.6474	0.0741	0.0802	0.0082	0.910
Y5	0.7000	0.6463	0.0825	0.0796	0.0096	0.890

References

Asparouhov & Muthen (2008). Auxiliary variables predicting missing data. Technical report. www.statmodel.com.

LSAY Estimated Means



MAR sample size = 3102
Listwise sample size = 782
Max N sample size = 3065, 2581, 2241, 2040, 1593, 1168

Input Excerpts LSAY Math Mean Using Missing Data Correlates

```
          USEV = math7 math8 math9 math10 math11 math12;
          AUXILIARY = (M) female mothed homeres expel arrest
          hisp black hsdrop
          expect droptht7
          lunch mstrat;
DEFINE:   lunch = lunch/100;
          mstrat = mstrat/1000;
MODEL:    math7 WITH math8-math12;
          math8 WITH math9-math12;
          math9 WITH math10-math12;
          math10 WITH math11-math12;
          math11 WITH math12;
PLOT:     TYPE = PLOT3;
          SERIES = math7-math12(*);
```

Non-Ignorable Missing Data

Non-Ignorable Missing Data
Modeling Approaches And References

Selection modeling: [y | x] [m | y, x]. Different approaches to [m | y, x]:

- Little & Rubin (2002) book, Little (2008) handbook chapter: overview
- Diggle & Kenward (1994) in Applied Statistics:
  - predicting from y, y* (y* is non-ignorable dropout)
- Wu & Carroll (1988), Wu & Bailey (1989) in Biometrics:
  - predicting from the slope s

Pattern-mixture modeling: [m | x] [y | m, x]. Different approaches to [y | m, x]:

- Little & Rubin (2002), Little (2008): overview
- Conventional approach: predicting i, s from m
- Roy (2003) in Biometrics:
  - predicting i, s from a latent class variable c (missing data patterns)

Non-Ignorable Missing Data Modeling Approaches And References (Continued)

Shared-parameter modeling: $[y, m] = \sum_{c} [c]\ [m \mid c]\ [y \mid c]$:

Albert & Follman (2008) handbook chapter: overview
Beunckens et al (2008) in Biometrics: using a latent class variable c and i, s

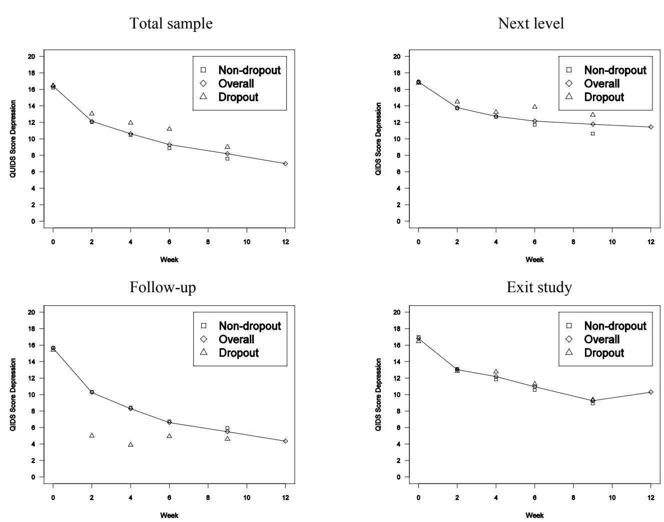
Non-Ignorable Missing Data Modeling In Longitudinal Studies

- Intermittent missingness versus dropout
- Muthén, B., Asparouhov, T., Hunter, A. & Leuchter, A. (2010). Growth modeling with non-ignorable dropout: Alternative analyses of the STAR*D antidepressant trial. Submitted for publication.
- Applied to STAR*D antidepressant trial data (n=4041)

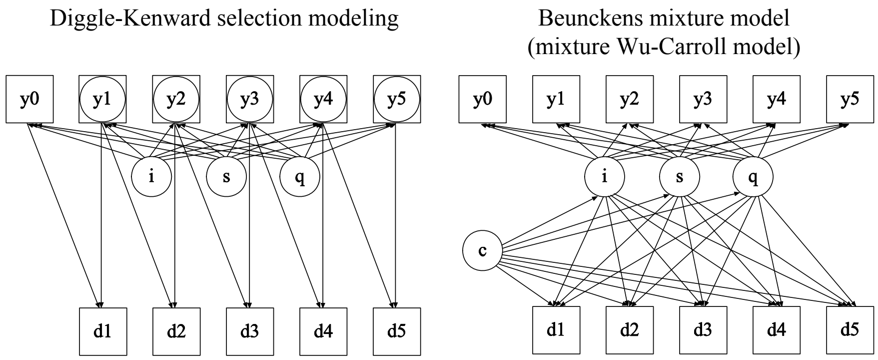
STAR*D Antidepressant Trial

- STAR*D multi-site NIMH antidepressant trial with $n = 4041$
- Subjects treated with citalopram (Level 1). No placebo group
- Clinician-rated QIDS depression score measured at baseline and weeks 2, 4, 6, 9, and 12
- 25% have complete data, 60% have monotone (dropout) missing data patterns, 14% have non-monotone missing data patterns
- Coverage at baseline and weeks 2, 4, 6, 9, and 12: 1.00, 0.79, 0.69, 0.68, 0.57, and 0.39
- Reasons for leaving Level 1: Remission (moved to follow-up), medication inefficient or not tolerated (moved to Level 2), study exit

STAR*D Antidepressant Trial



Selection Modeling



Selection modeling (d's are survival indicators)

Input Excerpts Diggle-Kenward Selection Modeling

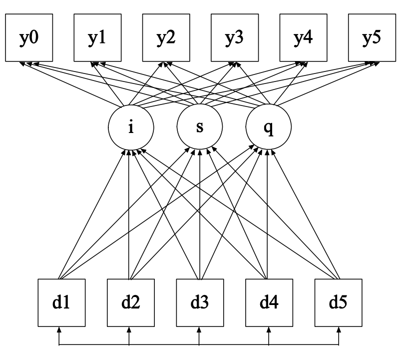
```
DATA:           FILE = StarD Ratings 1-23-09.dat;
VARIABLE:       NAMES = …
                … ;
                MISSING = ALL (-9999);
                USEV = y0 y1 y2 y3 y4 y5
                d1 d2 d3 d4 d5;
                CATEGORICAL = d1 d2 d3 d4 d5;
DATA MISSING:   NAMES = y1 y2 y3 y4 y5;
                TYPE = SDROPOUT;
                BINARY = d1-d5;
ANALYSIS:       PROCESS = 4;
                ALGORITHM = INTEGRATION;
                INTEGRATION = MONTECARLO;
                INTERACTIVE = control.dat;
```

Input Excerpts Diggle-Kenward Selection Modeling (Continued)

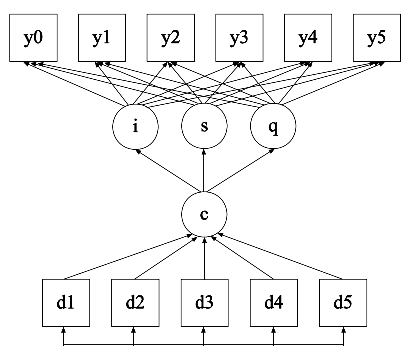
```
MODEL:          i s q | y0@0 y1@.2 y2@.4 y3@.6 y4@.9 y5@1.2;
                d1 ON y0 (beta2)
                y1 (beta1);
                d2 ON y1 (beta2)
                y2 (beta1);
                d3 ON y2 (beta2)
                y3 (beta1);
                d4 ON y3 (beta2)
                y4 (beta1);
                d5 ON y4 (beta2)
                y5 (beta1);
PLOT:           TYPE = PLOT3;
                SERIES = y0-y5 (s);
OUTPUT:         TECH1 SAMPSTAT RESIDUAL STANDARDIZED;
```


Pattern-Mixture Modeling

Conventional pattern-mixture modeling



Roy latent class dropout modeling



Pattern-mixture modeling (d's are dropout dummy variables)

Input Excerpts Pattern-Mixture Modeling

```
TITLE:
DATA:           FILE = StarD Ratings 1-23-09.dat;
VARIABLE:       NAMES = …
                … ;
                MISSING = ALL (-9999);
                USEV = y0 y1 y2 y3 y4 y5
                d1 d2 d3 d4 d5;
DATA MISSING:   NAMES = y1 y2 y3 y4 y5;
                TYPE = DDROPOUT;
                BINARY = d1-d5;
```


Input Excerpts Pattern-Mixture Modeling (Continued)

```
MODEL:          i s q | y0@0 y1@.2 y2@.4 y3@.6 y4@.9 y5@1.2;
                i-q ON d4-d5;
                i ON d1 d2 d3;
                s ON d3;
                s ON d1 (1);
                s ON d2 (1);
                q ON d1 (2);
                q ON d2 (2);
                q ON d3 (2);
PLOT:           TYPE = PLOT3;
                SERIES = y0-y5 (s);
OUTPUT:         TECH1 SAMPSTAT RESIDUAL STANDARDIZED;
```

Input Excerpts Roy 4-Class Latent Dropout Modeling

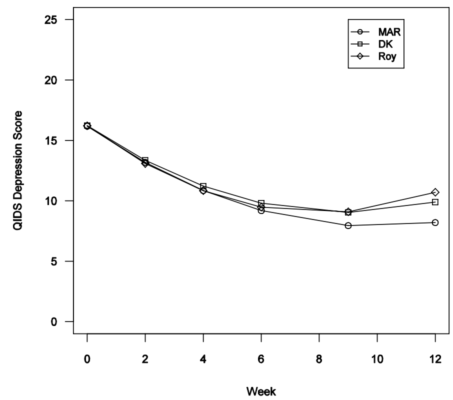
```
TITLE:
DATA:           FILE = StarD Ratings 1-23-09.dat;
VARIABLE:       NAMES = …
                … ;
                MISSING = ALL (-9999);
                USEV = y0 y1 y2 y3 y4 y5
                d1-d5;
                CLASSES = c(4);
DATA MISSING:   NAMES = y1 y2 y3 y4 y5;
                TYPE = DDROPOUT;
                BINARY = d1-d5;
ANALYSIS:       TYPE = MIXTURE;
                PROCESS = 4(STARTS);
                INTERACTIVE = control.dat;
                STARTS = 200 40;
```


Input Excerpts Roy 4-Class Latent Dropout Modeling (Continued)

```
MODEL:          %OVERALL%
                i s q | y0@0 y1@.2 y2@.4 y3@.6 y4@.9 y5@1.2;
                c ON d1-d5;
PLOT:           TYPE = PLOT3;
                SERIES = y0-y5 (s);
OUTPUT:         TECH1 SAMPSTAT RESIDUAL STANDARDIZED;
```

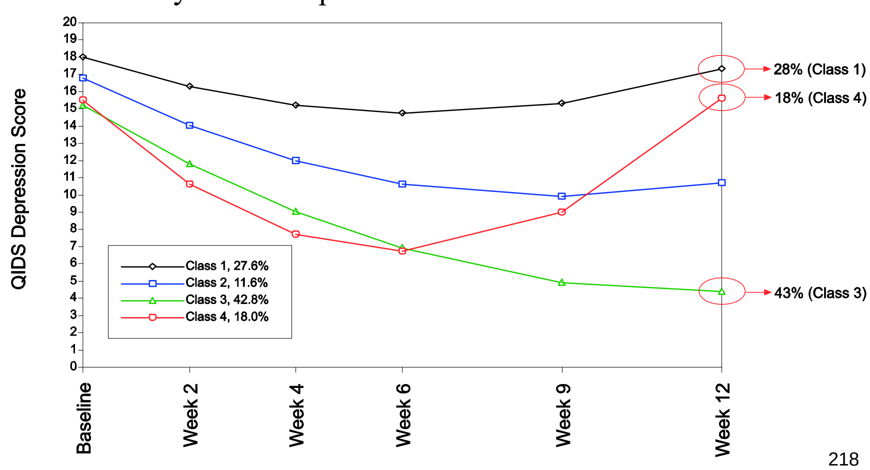
STAR*D Estimated Mean Growth Curves

Depression mean curves estimated under MAR, Diggle-Kenward selection modeling, and Roy latent dropout modeling

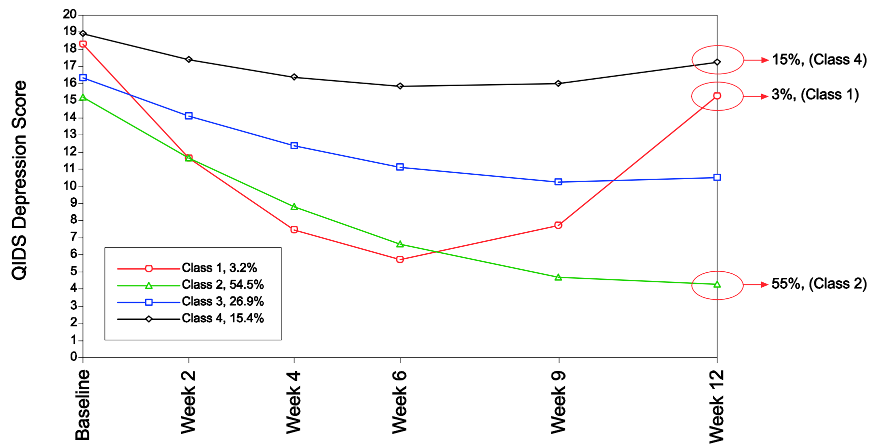


STAR*D Estimated Mean Growth Curves Under 4-Class Roy Model

4-class Roy latent dropout model



STAR*D Estimated Mean Growth Curves Under 4-Class MAR



Multiple Imputation Analysis In Mplus

- Multiply imputed data an input data alternative
- Estimates and SEs aggregated over the analyses
- Model testing

Further Readings On Missing Data Analysis

Albert, P.S. & Follman, D.A. (2008). Shared-parameter models. In Fitzmaurice, G., Davidian, M., Verbeke, G. & Molenberghs, G. (eds.), Longitudinal Data Analysis, pp. 433-452. Boca Raton: Chapman & Hall/CRC Press.

Beunckens, C., Molenberghs, G., Verbeke, G., and Mallinckrodt, C. (2008). A latent-class mixture model for incomplete longitudinal Gaussian data. Biometrics, 64, 96-105.

Dantan, E., Proust-Lima, C., Letenneur, L., & Jacqmin-Gadda, H. (2008). Pattern mixture models and latent class models for the analysis of multivariate longitudinal data with informative dropouts. The International Journal of Biostatistics, 4(1), 1-26.

Diggle, P.D. and Kenward, M.G. (1994). Informative dropout in longitudinal data analysis (with discussion). Applied Statistics, 43, 84.

Graham, J.W. (2003). Adding missing-data-relevant variables to FIML-based structural equation models. Structural Equation Modeling, 10(1), 80-100.

Further Readings On Missing Data Analysis

Hedeker, D. & Rose, J.S. (2000). The natural history of smoking: A pattern-mixture random-effects regression model. Multivariate applications in substance use research, J. Rose, L. Chassin, C. Presson & J. Sherman (eds.), Hillsdale, N.J.: Erlbaum, pp. 79-112.

Little, R.J., & Rubin, D.B. (2002). Statistical analysis with missing data. 2nd edition. New York: John Wiley & Sons.

Little, R.J. (2008). Selection and pattern-mixture models. In Fitzmaurice, G., Davidian, M., Verbeke, G. & Molenberghs, G. (eds.), Longitudinal Data Analysis, pp. 409-431. Boca Raton: Chapman & Hall/CRC Press.

Morgan-Lopez, A.A. & Fals-Stewart, W. (2007) Analytic methods for modeling longitudinal data from rolling therapy groups with membership turnover. Journal of Consulting and Clinical Psychology, *75*, 580-593.

Muthén, B., Asparouhov, T., Hunter, A. & Leuchter, A. (2010). Growth modeling with non-ignorable dropout: Alternative analyses of the STAR*D antidepressant trial. Submitted for publication.

Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. Psychometrika, 42, 431-462. (#17)

Further Readings On Missing Data Analysis (Continued)

Muthén, B., Jo, B. & Brown, H. (2003). Comment on the Barnard, Frangakis, Hill & Rubin article, Principal stratification approach to broken randomized experiments: A case study of school choice vouchers in New York City. Journal of the American Statistical Association, 98, 311-314.

Roy, J. (2003). Modeling longitudinal data with nonignorable dropouts using a latent dropout class model. Biometrics, 59, 829-836.

Schafer, J.L. (1997). Analysis of incomplete multivariate data. London: Chapman & Hall.

Schafer, J.L & Graham, J. (2002). Missing data: Our view of the state of the art. Psychological Methods, 7, 147- 177.

Wu, M.C. & Carroll, R.J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. Biometrics, 44, 175-188.

References

(To request a Muthén paper, please email bmuthen@ucla.edu.)

Analysis With Longitudinal Data
Introductory

Bollen, K.A. & Curran, P.J. (2006). Latent curve models. A structural equation perspective. New York: Wiley.

Collins, L.M. & Sayer, A. (Eds.) (2001). New methods for the analysis of change. Washington, D.C.: American Psychological Association.

Curran, P.J. & Bollen, K.A. (2001). The best of both worlds: Combining autoregressive and latent curve models. In Collins, L.M. & Sayer, A.G. (Eds.) New methods for the analysis of change (pp. 105-136). Washington, DC: American Psychological Association.

Duncan, T.E., Duncan, S.C., Strycker, L.A., Li, F., & Alpert, A. (1999). An introduction to latent variable growth curve modeling: concepts, issues, and applications. Mahwah, NJ: Lawrence Erlbaum Associates.

Goldstein, H. (1995). Multilevel statistical models. Second edition. London: Edward Arnold.

References (Continued)

Jennrich, R.I. & Schluchter, M.D. (1986). Unbalanced repeated-measures models with structured covariance matrices. Biometrics, 42, 805-820.

Laird, N.M., & Ware, J.H. (1982). Random-effects models for longitudinal data. Biometrics, 38, 963-974.

Lindstrom, M.J. & Bates, D.M. (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. Journal of the American Statistical Association, 83, 1014-1022.

Littell, R., Milliken, G.A., Stroup, W.W., & Wolfinger, R.D. (1996). SAS system for mixed models. Cary NC: SAS Institute.

McArdle, J.J. & Epstein, D. (1987). Latent growth curves within developmental structural equation models. Child Development, 58, 110-133.

McArdle, J.J. & Hamagami, F. (2001). Latent differences score structural models for linear dynamic analyses with incomplete longitudinal data. In Collins, L.M. & Sayer, A. G. (Eds.), New methods for the analysis of change (pp. 137-175). Washington, D.C.: American Psychological Association.

References (Continued)

Meredith, W. & Tisak, J. (1990). Latent curve analysis Psychometrika, 55, 107-122.

Muthén, B. (1991). Analysis of longitudinal data using latent variable models with varying parameters. In L. Collins & J. Horn (Eds.), Best methods for the analysis of change. Recent advances, unanswered questions, future directions (pp. 1-17). Washington D.C.: American Psychological Association.

Muthén, B. (2000). Methodological issues in random coefficient growth modeling using a latent variable framework: Applications to the development of heavy drinking. In Multivariate applications in substance use research, J. Rose, L. Chassin, C. Presson & J. Sherman (eds.), Hillsdale, N.J.: Erlbaum, pp. 113-140.

Muthén, B. & Khoo, S.T. (1998). Longitudinal studies of achievement growth using latent variable modeling. Learning and Individual Differences. Special issue: latent growth curve analysis, 10, 73-101.

Muthén, B. & Muthén, L. (2000). The development of heavy drinking and alcohol-related problems from ages 18 to 37 in a U.S. national sample. Journal of Studies on Alcohol, 61, 290-300.

References (Continued)

Muthén, B. & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. Biometrics, 55, 463-469.

Rao, C.R. (1958). Some statistical models for comparison of growth curves. Biometrics, 14, 1-17.

Raudenbush, S.W. & Bryk, A.S. (2002). Hierarchical linear models: Applications and data analysis methods. Second edition. Newbury Park, CA: Sage Publications.

Singer, J.D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. Journal of Educational and Behavioral Statistics, 23, 323-355.

Singer, J.D. & Willett, J.B. (2003). Applied longitudinal data analysis. Modeling change and event occurrence. New York, NY: Oxford University Press.

Snijders, T. & Bosker, R. (1999). Multilevel analysis: An introduction to basic and advanced multilevel modeling. Thousand Oakes, CA: Sage Publications.

Tucker, L.R. (1958). Determination of parameters of a functional relation by factor analysis. Psychometrika, 23, 19-23.

References (Continued)

Advanced

Albert, P.S. & Shih, J.H. (2003). Modeling tumor growth with random onset. Biometrics, 59, 897-906.

Brown, C.H. & Liao, J. (1999). Principles for designing randomized preventive trials in mental health: An emerging development epidemiologic perspective. American Journal of Community Psychology, special issue on prevention science, 27, 673-709.

Brown, C.H., Indurkhya, A. & Kellam, S.K. (2000). Power calculations for data missing by design: applications to a follow-up study of lead exposure and attention. Journal of the American Statistical Association, 95, 383-395.

Collins, L.M. & Sayer, A. (Eds.), New methods for the analysis of change. Washington, D.C.: American Psychological Association.

Duan, N., Manning, W.G., Morris, C.N. & Newhouse, J.P. (1983). A comparison of alternative models for the demand for medical care. Journal of Business and Economic Statistics, 1, 115-126.

References (Continued)

Ferrer, E. & McArdle, J.J. (2004). Alternative structural models for multivariate longitudinal data analysis. Structural Equation Modeling, 10, 493-524.

Fitzmaurice, G., Davidian, M., Verbeke, G. & Molenberghs, G. (2008). Longitudinal Data Analysis. Chapman & Hall/CRC Press.

Klein, A. & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. Psychometrika, 65, 457-474.

Khoo, S.T. & Muthén, B. (2000). Longitudinal data on families: Growth modeling alternatives. In Multivariate applications in substance use research, J. Rose, L. Chassin, C. Presson & J. Sherman (eds), Hillsdale, N.J.: Erlbaum, pp. 43-78.

Miyazaki, Y. & Raudenbush, S.W. (2000). A test for linkage of multiple cohorts from an accelerated longitudinal design. Psychological Methods, 5, 44-63.

Moerbeek, M., Breukelen, G.J.P. & Berger, M.P.F. (2000). Design issues for experiments in multilevel populations. Journal of Educational and Behavioral Statistics, 25, 271-284.

References (Continued)

Muthén, B. (1996). Growth modeling with binary responses. In A.V. Eye & C. Clogg (eds), Categorical variables in developmental research: methods of analysis (pp. 37-54). San Diego, CA: Academic Press.

Muthén, B. (1997). Latent variable modeling with longitudinal and multilevel data. In A. Raftery (ed), Sociological Methodology (pp. 453-380). Boston: Blackwell Publishers.

Muthén, B. (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, B. & Curran, P. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. Psychological Methods, 2, 371-402.

Muthén, B. & Muthén, L. (2000). The development of heavy drinking and alcohol-related problems from ages 18 to 37 in a U.S. national sample. Journal of Studies on Alcohol, 61, 290-300.

References (Continued)

Muthén, L.K. and Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. Structural Equation Modeling, 4, 599-620.

Olsen, M.K. & Schafer, J.L. (2001). A two-part random effects model for semicontinuous longitudinal data. Journal of the American Statistical Association, 96, 730-745.

Raudenbush, S.W. (1997). Statistical analysis and optimal design for cluster randomized trials. Psychological Methods, 2, 173-185.

Raudenbush, S.W. & Liu, X. (2000). Statistical power and optimal design for multisite randomized trials. Psychological Methods, 5, 199-213.

Roeder, K., Lynch, K.G., & Nagin, D.S. (1999). Modeling uncertainty in latent class membership: A case study in criminology. Journal of the American Statistical Association, 94, 766-776.

Satorra, A. & Saris, W. (1985). Power of the likelihood ratio test in covariance structure analysis. Psychometrika, 51, 83-90.

References (Continued)

Missing Data Analysis

Albert, P.S. & Follman, D.A. (2008). Shared-parameter models. In Fitzmaurice, G., Davidian, M., Verbeke, G. & Molenberghs, G. (eds.), Longitudinal Data Analysis, pp. 433-452. Boca Raton: Chapman & Hall/CRC Press.

Beunckens, C., Molenberghs, G., Verbeke, G., and Mallinckrodt, C. (2008). A latent-class mixture model for incomplete longitudinal Gaussian data. Biometrics, 64, 96-105.

Diggle, P.D. and Kenward, M.G. (1994). Informative dropout in longitudinal data analysis (with discussion). Applied Statistics, 43, 84.

Graham, J.W. (2003). Adding missing-data-relevant variables to FIML-based structural equation models. Structural Equation Modeling, 10, 80-100.

Hedeker, D. & Rose, J.S. (2000). The natural history of smoking: A pattern-mixture random-effects regression model. Multivariate applications in substance use research, J. Rose, L. Chassin, C. Presson & J. Sherman (eds.), Hillsdale, N.J.: Erlbaum, pp. 79-112.

References (Continued)

Little, R.J., & Rubin, D.B. (2002). Statistical analysis with missing data. 2nd edition. New York: John Wiley & Sons.

Little, R.J. (2008). Selection and pattern-mixture models. In Fitzmaurice, G., Davidian, M., Verbeke, G. & Molenberghs, G. (eds.), Longitudinal Data Analysis, pp. 409-431. Boca Raton: Chapman & Hall/CRC Press.

Morgan-Lopez, A.A. & Fals-Stewart, W. (2007) Analytic methods for modeling longitudinal data from rolling therapy groups with membership turnover. Journal of Consulting and Clinical Psychology, *75*, 580-593.

Muthén, B., Asparouhov, T., Hunter, A. & Leuchter, A. (2010). Growth modeling with non-ignorable dropout: Alternative analyses of the STAR*D antidepressant trial. Submitted for publication.

Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. Psychometrika, 42, 431-462. (#17)

Muthén, B., Jo, B. & Brown, H. (2003). Comment on the Barnard, Frangakis, Hill & Rubin article, Principal stratification approach to broken randomized experiments: A case study of school choice vouchers in New York City. Journal of the American Statistical Association, 98, 311-314. (#99)

References (Continued)

Roy, J. (2003). Modeling longitudinal data with nonignorable dropouts using a latent dropout class model. Biometrics, 59, 829-836.

Schafer, J.L. (1997). Analysis of incomplete multivariate data. London: Chapman & Hall.

Schafer, J.L & Graham, J. (2002). Missing data: Our view of the state of the art. Psychological Methods, 7, 147- 177.

Wu, M.C. & Carroll, R.J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. Biometrics, 44, 175-188.

References (Continued)

Survival Analysis

Allison, P.D. (1984). Event History Analysis. Regression for Longitudinal Event Data. Quantitative Applications in the Social Sciences, No. 46. Thousand Oaks: Sage Publications.

Asparouhov, T., Masyn, K. & Muthen, B. (2006). Continuous time survival in latent variable models. Proceedings of the Joint Statistical Meeting in Seattle, August 2006. ASA section on Biometrics, 180-187.

Larsen, K. (2004). Joint analysis of time-to-event and multiple binary indicators of latent classes, Biometrics, 60(1), 85–92.

Larsen, K. (2005). The Cox proportional hazards model with a continuous latent variable measured by multiple binary indicators, Biometrics, 61(4), 1049–1055.

Lin, H., Turnbull, B.W., McCulloch, C.E. & Slate, E. (2002). Latent class models for joint analysis of longitudinal biomarker and event process data: application to longitudinal prostate-specific antigen readings and prostate cancer. Journal of the American Statistical Association, 97, 53-65.

References (Continued)

Masyn, K. E. (2008). Modeling measurement error in event occurrence for single, non-recurring events in discrete-time survival analysis. In Hancock, G. R., & Samuelsen, K. M. (Eds.), Advances in latent variable mixture models, pp. 105-145. Charlotte, NC: Information Age Publishing, Inc.

Muthén, B. & Masyn, K. (2005). Discrete-time survival mixture analysis. Journal of Educational and Behavioral Statistics, 30, 27-28.

Muthén, B. & Muthén, L. (2000). Integrating person-centered and variable-centered analysis: growth mixture modeling with latent trajectory classes. Alcoholism: Clinical and Experimental Research, 24, 882-891. (#85)

Singer, J.D., & Willett, J.B. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. Journal of Educational Statistics, 18(20), 155-195.

Singer, J. D. & Willett, J. B. (2003). Applied longitudinal analysis. Modeling change and event occurrence. Oxford, UK: Oxford University Press.

Vermunt, J.K. (1997). Log-linear models for event histories. Advanced quantitative techniques in the social sciences, vol. 8. Thousand Oaks: Sage Publications.