Latent growth curve analysis in perinatal and pediatric epidemiology

SPER Advanced Methods Workshop
Miami, FL
June 20, 2016

Janne Boone-Heinonen, PhD, MPH
Sheila Markwardt, MPH
Oregon Health & Science University
OHSU-PSU School of Public Health

SCHOOL OF PUBLIC HEALTH

Objectives

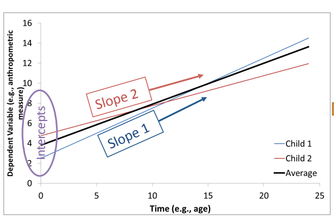
Participants will be able to:

- Describe the strengths and general approach of latent growth curve analysis in existing research
- Evaluate whether latent growth curve analysis is appropriate for their research
- Apply basic latent growth analysis in Mplus (example: infant growth)

2

Growth Curves

Predictors (e.g., SES)



Outcomes (e.g., diabetes)

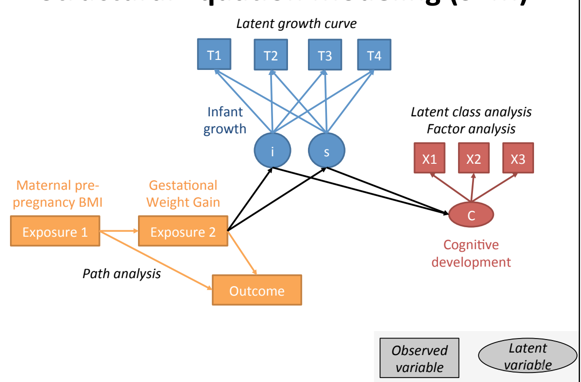
Generalized Estimating Equations (average trajectories)

Mixed effects models (individual trajectories)

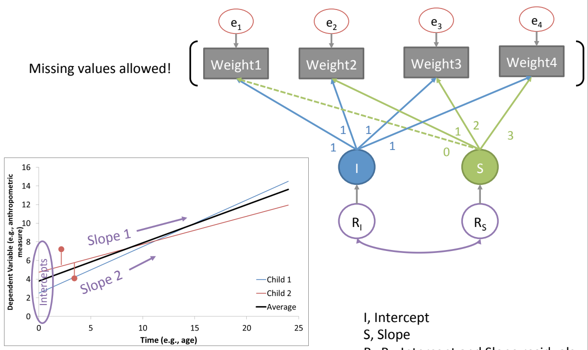
Latent growth curves (individual trajectories)

3

Structural Equation Modeling (SEM)



Latent Growth Curves

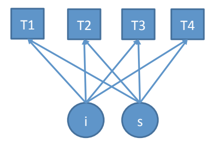


I, Intercept
S, Slope
R_I, R_S, Intercept and Slope residuals
e_1-e_4, error terms

5

Latent Growth Curves

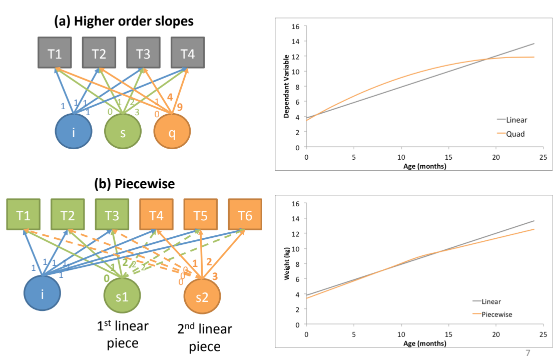
flexibility and options



1. Non-linear growth
2. Covariates
3. Subgroup differences
4. Scaling of time

6

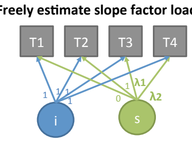
1. Non-linear growth (a, b)



7

1. Non-linear growth (c, d)

(c) Freely estimate slope factor loadings

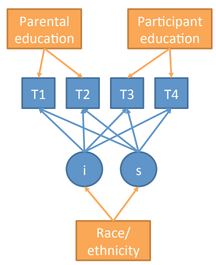


(d) Structured latent curves

e.g., exponential, Gompertz

8

2. Covariates



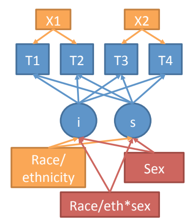
Time varying covariates

Time constant covariates

9

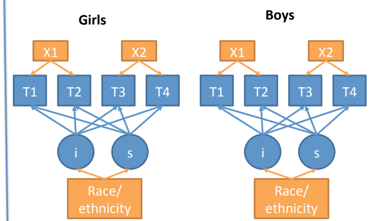
3. Subgroup differences

Covariate/interaction



Test: interaction term

Multiple groups



Test: overall model fit in models that (a) constrain parameter(s) to be <u>equal</u> in boys and girls and (b) parameter(s) to <u>differ</u> in boys in girls

10

4. Individually varying time points

Scenario 1

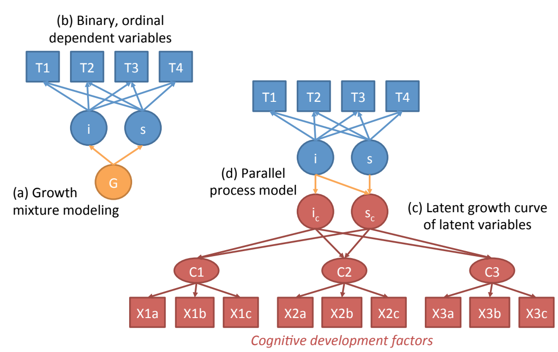
	Exam 1	Exam 2	Exam 3	Exam 4	Exam 5	Exam 6
Age (target)	Birth	3 mo	6 mo	9 mo	12 mo	18 mo
Age (actual)	Birth	2-4 mo	5-7 mo	8-10 mo	11-13 mo	14-19 mo

Scenario 2

	Exam 1	Exam 2	Exam 3	Exam 4	Exam 5	Exam 6
Child 1	Birth	2.9 mo	6.2 mo	9.8 mo	12.2 mo	19.7 mo
Child 2	Birth	1.2 mo	1.8 mo	4.4 mo	9.7 mo	12.1 mo

11

Other topics



12

Part II

PRACTICAL APPLICATION OF LATENT GROWTH CURVES

SEM software

- **Mplus**
- AMOS
- LISREL
- Stata
- R

1. Data management (**Stata**, SAS)

2. Export/Import

3. SEM program (**Mplus**)

Data processing

1. Data management (**Stata**, SAS)
   - Limit to key variables
   - 0/1 coding for binary variables
   - Create "dummy" variables
   - Create higher order terms (other than time)
   - Create interaction terms
   - 8-character variable names
   - "Wide" data structure
   - (Missing value code)

2. Export/Import
   - Stata: `stata2mplus` function
   - SAS: http://www.ats.ucla.edu/stat/mplus/faq/sas2mplus.htm

3. SEM program (**Mplus**)

2b. Import to Mplus

Stata: `stata2mplus` creates this Mplus code
SAS: analyst creates analogous code

```
Title:
  Stata2Mplus conversion for X:\SPH\Shared\Obesity\Infants.dta
  List of variables converted shown below
      …
Data:
  File is
    X:\SPH\Shared\Obesity\Infants.dat;

Variable:
  Names are id wt0-wt6 a0-a6;
  Missing are all (-9999);
Analysis:
  Type = basic ;
```

3. Mplus analysis

Example: Infant growth

Kaiser Permanente NW Region (KPNW)
Electronic Medical Record Data
MAT MORB (CDC-Funded, KPNW-based study)

Kaiser-based collaborators
Stephen P. Fortmann, MD
Kimberly Vesco, MD

N=21,899 live births in 2000-2007

- Dependent variable: **weight** (kg), 2 weeks to 24 months
  - In this example, selected measure within age intervals, closest to the following time points: birth, 3 months, 6 months, 9 months, 12 months, 18 months, 24 months
- Latent Growth Curve analysis (MPlus 7.4)

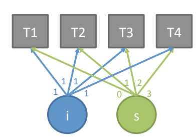
Mplus: basic syntax

```
Title:
  Linear LGM;
DATA:
  File is
  X:\SPH\Shared\Obesity\Infants.dat;
VARIABLE:
  Names are
     id wt0-wt6 a0-a6;

  Usevariables are wt0-wt6;

  Missing are all (-9999);

ANALYSIS: …
```

- Linear LGM; — Any (preferably informative) title
- X:\SPH\Shared\Obesity\Infants.dat; — File location
- id wt0-wt6 a0-a6; — ALL variables in dataset
- wt0-wt6; — Variables in the current analysis
- (-9999) — Missing data code assigned in export step

Latent growth curve: **linear**

[input]




```
Title:
  Linear LGM;
DATA:
  File is
  X:\SPH\Shared\Obesity\Infants.dat;
VARIABLE:
  Names are
     id wt0-wt6 a0-a6;
  Usevariables are wt0-wt6;
  Missing are all (-9999);
MODEL:
  i l | wt0@0 wt1@.3 wt2@.6 wt3@.9 wt4@1.2 wt5@1.8 wt6@2.4;
 OUTPUT:
  TECH1;
```

- wt0-wt6; — Weight variables at time 0 … 6
- Slope factor loadings (months/10)
- User-specified names:
  i = intercept term
  l = linear slope term

Note: FIML (for missing data) is the default

Latent growth curve: **linear**
[output (1)]

```
THE MODEL ESTIMATION TERMINATED NORMALLY

MODEL FIT INFORMATION
Loglikelihood
          H0 Value                         -174480.963
          H1 Value                         -108311.313

Information Criteria
          Akaike (AIC)                      348985.927
          Bayesian (BIC)                    349081.857
          Sample-Size Adjusted BIC          349043.721
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit
          Value                             132339.300
          Degrees of Freedom                        23
          P-Value                               0.0000
…
```

Useful for comparing models (BIC_{diff}>10)

21

Latent growth curve: **linear**
[output (2)]

```
…
RMSEA (Root Mean Square Error Of Approximation)
          Estimate                           0.513
          90 Percent C.I.                    0.510  0.515
          Probability RMSEA <= .05           0.000

CFI/TLI
          CFI                                0.000
          TLI                               -0.247

Chi-Square Test of Model Fit for the Baseline Model
          Value                          96903.981
          Degrees of Freedom                    21
          P-Value                           0.0000

SRMR (Standardized Root Mean Square Residual
      Value                             1.208
```

<=0.05 is good (RMSEA)

>0.95 Is good (CFI/TLI)

Usually sig (P-Value)

<=0.05 is good (SRMR)

This model has terrible fit

22

Latent growth curve: **linear**
[output (2)]

```
MODEL RESULTS
                                                  Two-Tailed
                     Estimate      S.E.  Est./S.E.    P-Value
 I        |
    WT0                1.000      0.000    999.000    999.000
    WT1                1.000      0.000    999.000    999.000
    WT2                1.000      0.000    999.000    999.000
    WT3                1.000      0.000    999.000    999.000
    WT4                1.000      0.000    999.000    999.000
    WT5                1.000      0.000    999.000    999.000
    WT6                1.000      0.000    999.000    999.000
 L        |
    WT0                0.000      0.000    999.000    999.000
    WT1                0.300      0.000    999.000    999.000
    WT2                0.600      0.000    999.000    999.000
    WT3                0.900      0.000    999.000    999.000
    WT4                1.200      0.000    999.000    999.000
    WT5                1.800      0.000    999.000    999.000
    WT6                2.400      0.000    999.000    999.000
```

Fixed at 1 (I loadings)

Fixed path loadings for each time point (L loadings)

Latent growth curve: **linear**
[output (2)]

```
MODEL RESULTS
                                                  Two-Tailed
                     Estimate      S.E.  Est./S.E.    P-Value

 L        WITH
    I                  0.139      0.004     33.084      0.000

 Means
    I                  3.847      0.009    411.701      0.000
    L                  4.097      0.009    456.723      0.000

 Intercepts
    WT0                0.000      0.000    999.000    999.000
    WT1                0.000      0.000    999.000    999.000
    WT2                0.000      0.000    999.000    999.000
    WT3                0.000      0.000    999.000    999.000
    WT4                0.000      0.000    999.000    999.000
    WT5                0.000      0.000    999.000    999.000
    WT6                0.000      0.000    999.000    999.000
```

Covariance between intercept and slope

Mean intercept and slope

Errors

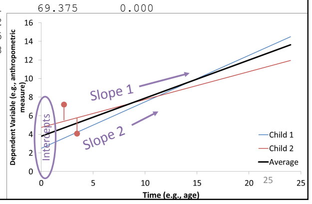
24

Latent growth curve: **linear**
[output (3)]

```
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
 Variances
    I                  0.067      0.004     16.916      0.000
    L                  0.440      0.010     46.095      0.000

Residual Variances
    WT0                0.380      0.009     40.987      0.000
    WT1                0.903      0.012     78.024      0.000
    WT2                2.467      0.034     73.090      0.000
    WT3                2.153      0.031     69.375      0.000
    WT4                1.466      0.022
    WT5                0.031      0.015
    WT6                2.097      0.038
```

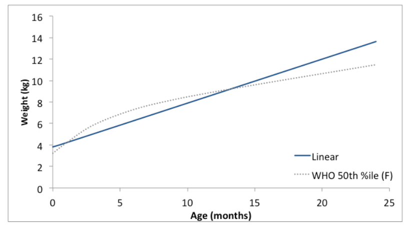
Variance of individual intercepts and slopes around the **average intercept/slope**

Errors at each time point (observed value vs. **individual predicted value**)



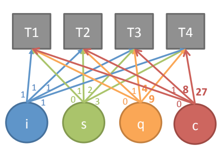
Linear growth?!

- (very) poor model fit
- Large, varying errors over time



Latent growth curve: **cubic**
[input]

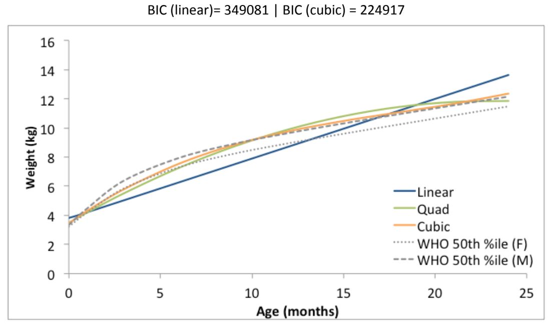
```
Title:
   Cubic LGM Binned;
DATA:
   File is
   X:\SPH\Shared\Obesity\Infants.dat;
VARIABLE:
   Names are
      id wt0 wt1 wt2 wt3 wt4 wt5 wt6 ;
   USEVAR = wt0-wt6;
   Missing are all (-9999);
 MODEL:
   i l q c |
      wt0@0 wt1@.3 wt2@.6 wt3@.9 wt4@1.2 wt5@1.8 wt6@2.4;
```



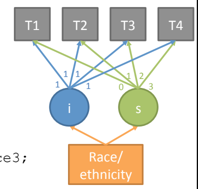
User-specified names:
- i = intercept term
- l = linear slope term
- **q = quadratic slope term**
- **c = cubic slope term**

Cubic growth?

Better model fit; further improves with subgroups, covariates
RMSEA=0.157 | CLI=0.917 | TLI=0.883 | SRMR 0.061
BIC (linear)= 349081 | BIC (cubic) = 224917



Latent growth curve: **predictor(s)**
[input]

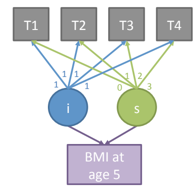



```
Title:
    Cube LGM, race as covariate;
DATA:
    File is
    X:\SPH\Shared\Obesity\Infants.dat;
VARIABLE:
    Names are
       id wt0-wt6 race1 race2 race3;
    Usevariables are wt0-wt6 race1 race2 race3;
    Missing are all (-9999) ;
MODEL:
    i l q c |
      wt0@0 wt1@.3 wt2@.6 wt3@.9 wt4@1.2 wt5@1.8 wt6@2.4;

    i on race1 race2 race3;
    l on race1 race2 race3;
    q on race1 race2 race3;
    c on race1 race2 race3;
```

Regress each growth term on race indicator variables

29

Latent growth curve: **outcomes**
[input]




```
Title:
    Cube LGM, BMI as outcome;
DATA:
    File is
    X:\SPH\Shared\Obesity\Infants_BMI.dat;
VARIABLE:
    Names are
       id wt0-wt6 bmi;
    Usevariables are wt0-wt6 bmi;
    Missing are all (-9999) ;
MODEL:
    i l q c |
      wt0@0 wt1@.3 wt2@.6 wt3@.9 wt4@1.2 wt5@1.8 wt6@2.4;

    bmi on i;
    bmi on l;
    bmi on q;
    bmi on c;
```

Latent growth curve: **subgroup differences**
[input]

```
Title:
    Cube LGM, by gender;
DATA:
    File is
    X:\SPH\Shared\Obesity\Infants_Gender.dat;
VARIABLE:
    Names are
       id wt0-wt6 bmi;
    Usevariables are wt0-wt6 gender;
    Missing are all (-9999) ;
    GROUPING = gender(0=Male 1=Female);
MODEL:
    i l q c |
      wt0@0 wt1@.3 wt2@.6 wt3@.9 wt4@1.2 wt5@1.8 wt6@2.4;
```

Test: overall model fit in models that (a) constrain parameter(s) to be equal in boys and girls and (b) parameter(s) to differ in boys in girls

Latent growth curve: **subgroup differences**
[input]

```
Title:
    Cube LGM, by gender w/ equal parameters;
DATA:
    File is
    X:\SPH\Shared\Obesity\Infants_Gender.dat;
VARIABLE:
    …
    GROUPING = gender(0=Male 1=Female);
MODEL:
    i l q c |
      wt0@0 wt1@.3 wt2@.6 wt3@.9 wt4@1.2 wt5@1.8 wt6@2.4;
Model:
    [i] (1);
    [l] (2);
    [q] (3);
    [c] (4);
Model Female:
    [i] (1);
    [l] (2);
    [q] (3);
    [c] (4);
```

Constraints:
Male intercepts = Female intercept
Male linear slope = Female linear slope
Male quadratic slope = Female quadratic slope
Male cubic slope = Female cubic slope

Intercept & slopes vary by gender: BIC 237906
Intercept & slopes equal: BIC 239834

Issues and challenges

- Conceptual challenges
  - Flexibility → more decisions, relies on strong theoretical framework
- Operational challenges:
  - Specialized software required (often Mplus)
  - Mplus learning curve
  - Difficulties in model convergence
    - Specify starting values
    - Increase number of iterations
    - Fixing variances, means, covariances

33

Questions?

Janne Boone-Heinonen, PhD, MPH
boonej@ohsu.edu

Sheila Markwardt, MPH
markward@ohsu.edu

OHSU
Portland State University
SCHOOL OF PUBLIC HEALTH