Plausible Values of Latent Variables: A Useful Approach of Data Reduction for Outcome Measures in Pediatric Studies

Jichuan Wang, Ph.D.

Children’s National Health System
The George Washington University School of Medicine
Washington, DC

Abstract

Multi-item scales are widely used to measure outcomes in pediatric studies. The often used data reduction approaches are total scale scores or estimated factor or IRT scores. However, the total score does not take into account of measurement errors; and using factor scores or IRT scores as dependent variables in secondary analysis gives biased slope coefficients. This presentation introduces a better approach -- plausible values of latent variables -- for data reduction. Real data are used to demonstrate how to estimate and apply plausible values to analyze multi-item outcome measures.

Multi-item outcome measures

- Mental disorders
- Quality of life
- Symptom domains
- Functional domains
- …

Challenge in using multi-item scales

- Too many variables are involved in a model, thus data reduction is needed.

Measures of outcome scales often used in data analysis

- Total scores.
- Factor scores, IRT scores, or latent class membership (categorical latent variable).
- Plausible values of latent variables (continuous or categorical).

What are plausible values of latent variable?

- A set of generated values of latent variables using multiple imputations.

Why using plausible values of latent variables?

- Total scores do not take into account of measurement errors.
- Factor analysis model often encounters estimation problem due to too many indicator (items) variables in a structural equation model.
- Using factor scores as dependent variables in secondary analysis gives biased slope coefficient estimates (Skrondal & Laake, 2001).
- Using plausible values can alleviate the bias.

How to use plausible values for secondary analysis

- Save the plausible values as “observed” variables and merge them with the original data set.
- When using plausible values in secondary statistical analysis, multiple data sets (e.g., 5) are needed just like multiple imputation (MI) data analysis using Rubin’s (1987) method.

Demonstration

Sample: Drug users (N=303) in Changsha, China recruited using RDS, 2012

Outcome measures: BSI-18

- Somatization: 6 items
- Depression: 6 items
- Anxiety: 6 items

Predictors in SEM model:

- Age
- Education
- Marital status
- Employment status
- Meth use in the past 30 days

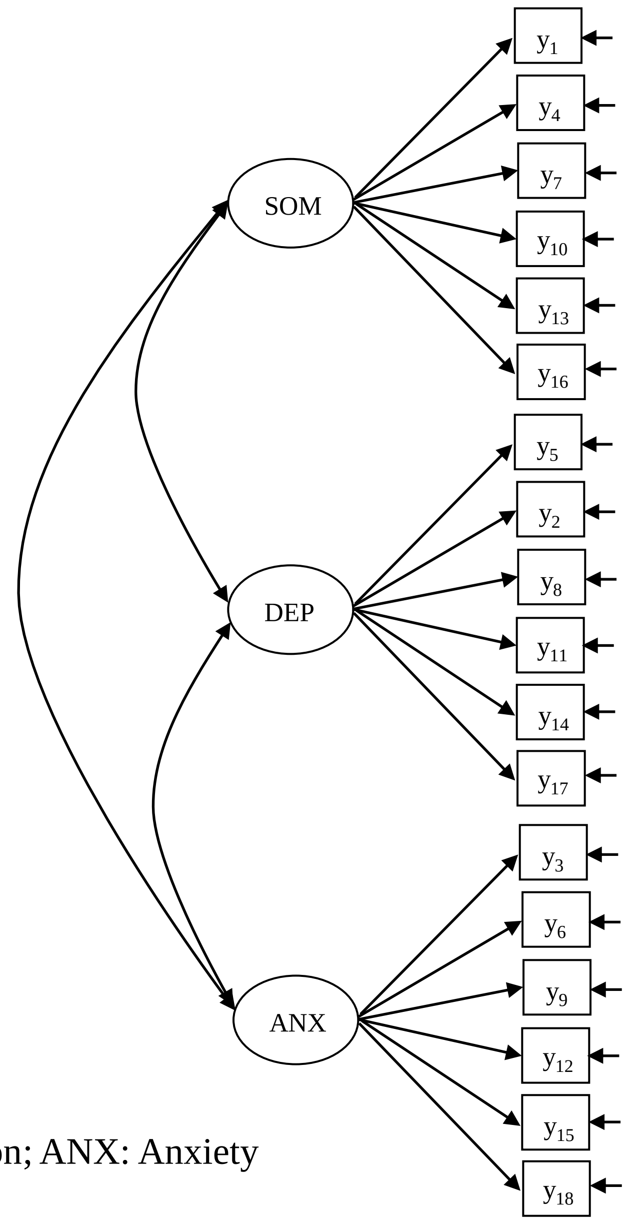

Figure 1. 3-factor CFA for BIS-18:
SOM: Somatization; DEP: Depression; ANX: Anxiety

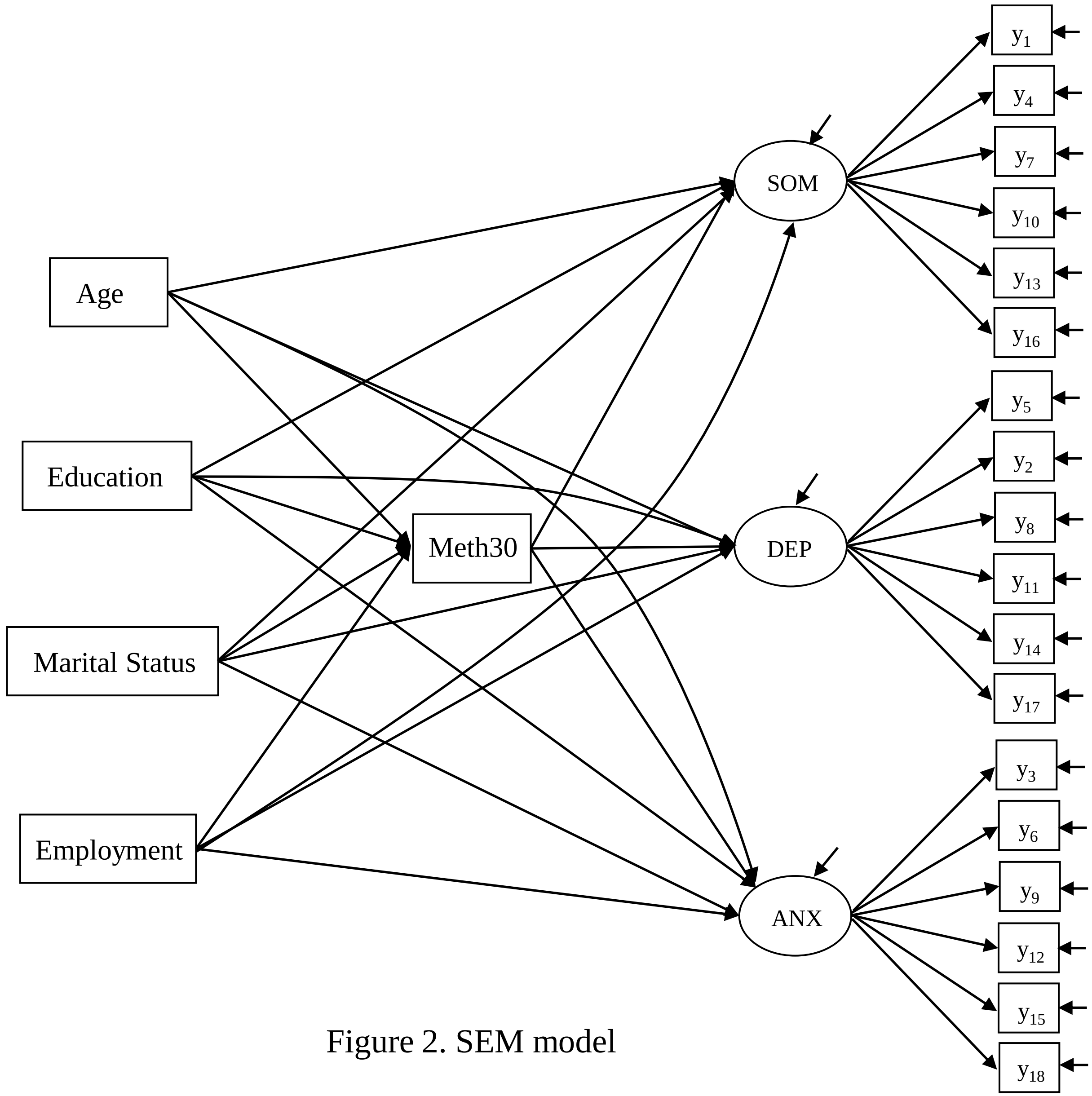

Figure 2. SEM model

WARNING: THE RESIDUAL COVARIANCE MATRIX (THETA) IS NOT POSITIVE DEFINITE. THIS COULD INDICATE A NEGATIVE VARIANCE/RESIDUAL VARIANCE FOR AN OBSERVED VARIABLE, A CORRELATION GREATER OR EQUAL TO ONE BETWEEN TWO OBSERVED VARIABLES, OR A LINEAR DEPENDENCY AMONG MORE THAN TWO OBSERVEDVARIABLES. CHECK THE RESULTS SECTION FOR MORE INFORMATION.PROBLEM INVOLVING VARIABLE Y10.

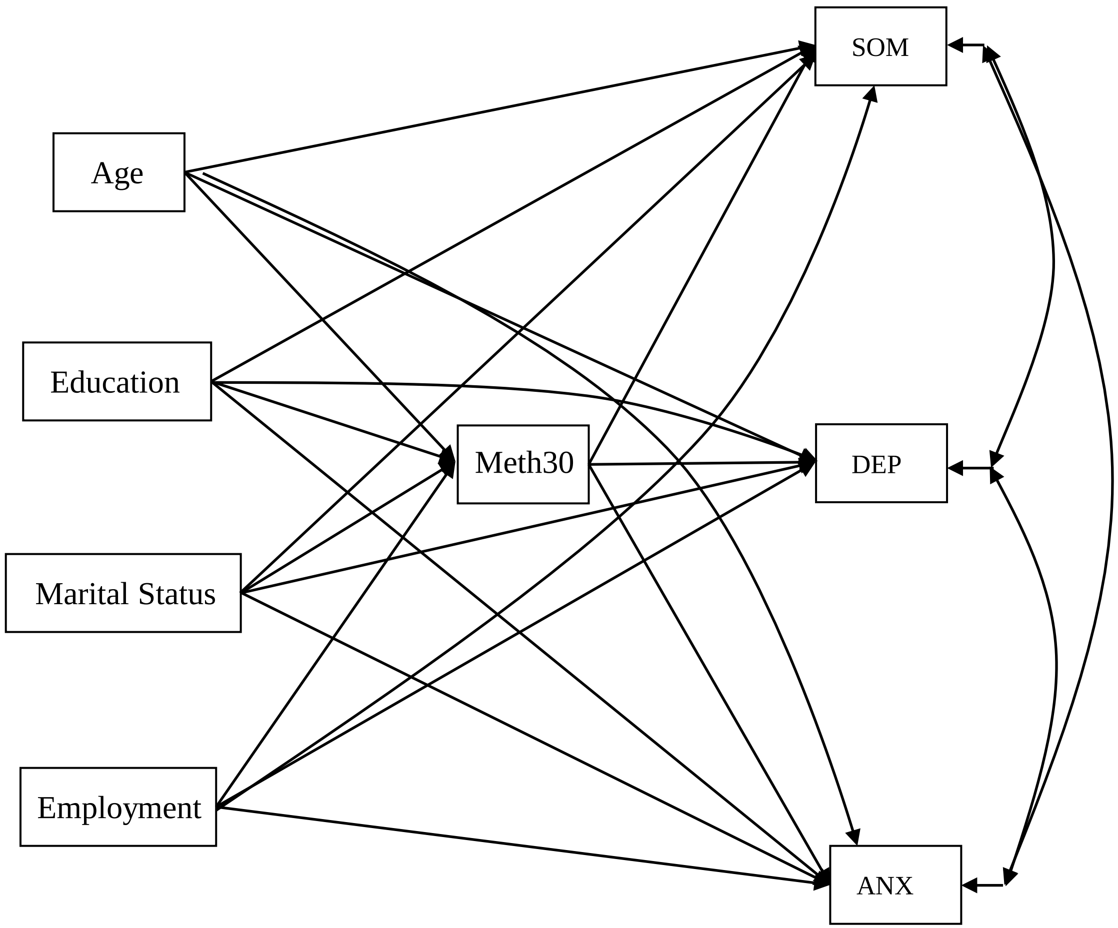

Figure 3. Path analysis model: SOM, DEP, and ANX are either total score or traditional point estimates of latent variables in CFA or IRT.

- Cross-loadings are specified; error covariance could be specified as well.
- Non-informative priors using a normal distribution with a mean of zero and a small variance (Muthén and Asparouhov, 2012).
- Model fit: Posterior Predictive P-Value (PPP) = **0.453.**
- Five sets of plausible values are imputed for each latent variable and saved for further analysis.

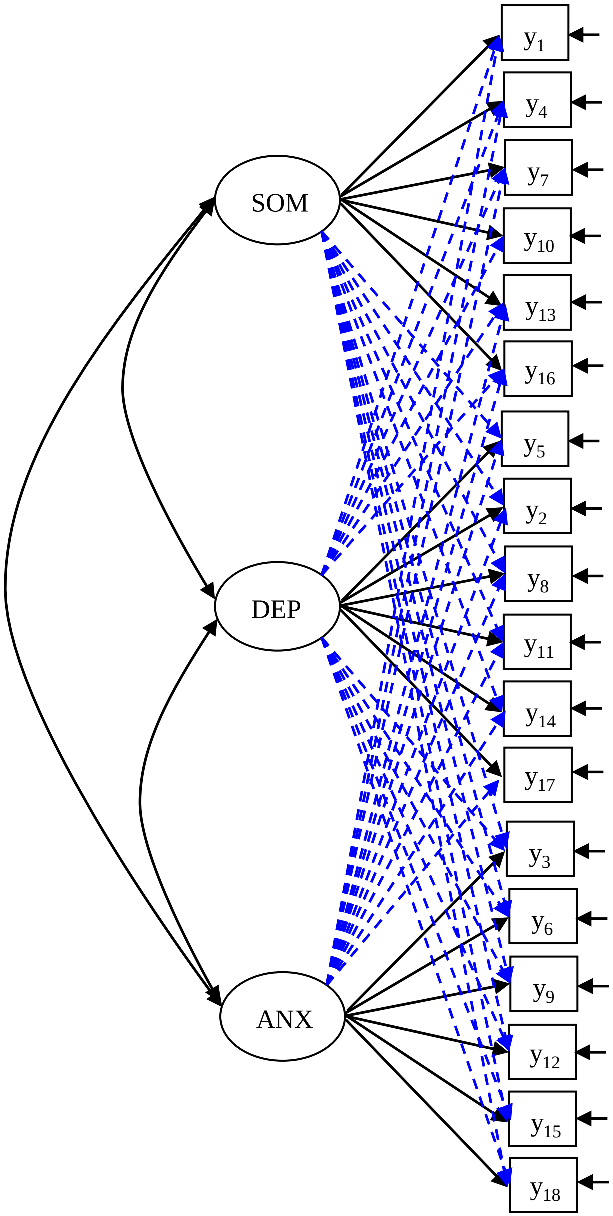

Figure 4. Bayesian CFA

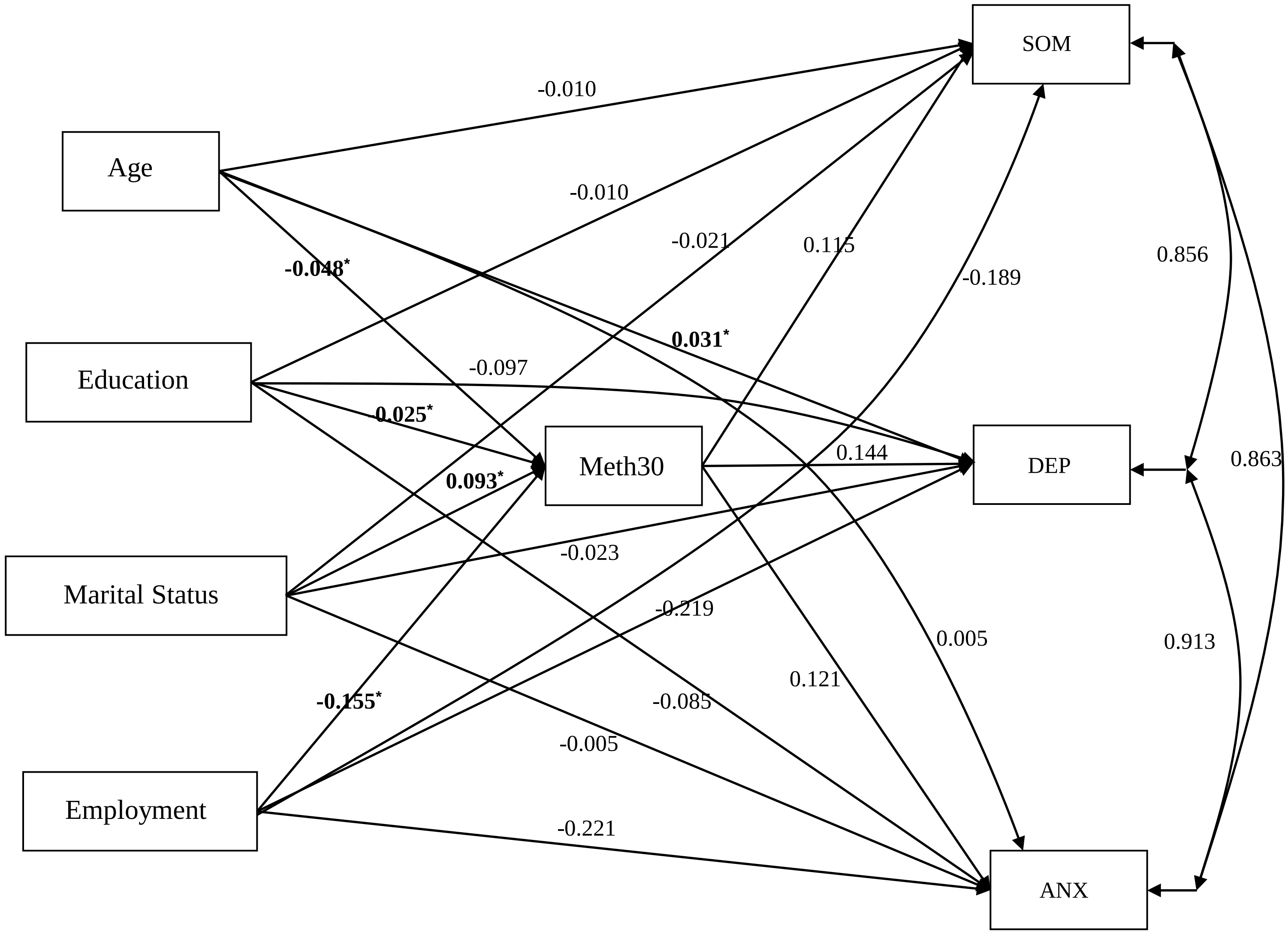

Figure 5. Example of path analysis model using five plausible values data sets.

References

Asparouhov, T. & Muthén, B. 2010. Plausible values for latent variables using Mplus.Technical Report.

Mislevy, R. J. 1991. Randomization-based inference about latent variables from complex samples. Psychometrika, 56, 177-196.

Muthen, B. & Asparouhov, T. 2012. Bayesian SEM: A more flexible representation of substantive theory. Psychological Methods, 17, 313-335.

Rubin, D. 1987. Multiple Imputation for Nonresponse in Survey. New York: Wiley.

Skrondal, A. and Laake, P. 2001. Regression among factor scores. Psychometrika, 66, 563-575.