Using Multiple Group Modeling to Test Moderators in Meta-Analysis

Alexander M. Schoemann
Overview

- Meta-analysis
  - Meta-analysis with SEM
  - Meta-analysis with MLM
- Multiple group meta-analysis
  - Models
  - Advantages
  - Simulation Study
  - Real-world example
- Future directions
Meta-analysis

- A quantitative approach to summarizing and combining results from empirical literature
  - AKA: research synthesis, quantitative research synthesis

- Data points come from a number of primary studies
  - Data are effect sizes

- A goal of meta-analysis is to identify a mean and variance for effect sizes across studies
  - Hopefully the mean and variance represent some larger population of effect sizes
  - Effect size may be moderated by other variables
Meta-analysis

- Two types of models can be fit:
  - Fixed-effects
  - Random-effects
- This talk will focus on the random-effects model
Fixed-effects model

- Computes the mean effect size across studies:
  \[ y_i = \beta_0 + e_i \]

- Moderators can be included as predictors of \( y_i \)
  - Moderators can be either categorical or continuous
  \[ y_i = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p + e_i \]
Random-effects model

- Computes the mean and variance of the effect sizes across studies:

\[ y_i = \beta_0 + u_i + e_i \]

\[ u_i \sim N(0, \tau^2) \]

- Moderators can be included as predictors of \( y_i \)
  - Moderators can be either categorical or continuous

\[ y_i = \beta_0 + \beta_1 x_1 + ... + \beta_p x_p + u_i + e_i \]
Effect sizes are treated as a latent variable

- Mean and variance of the latent variable are estimated
- Effect sizes are regressed on a constant and the regression slope is treated as random
- Effect sizes are weighted by multiplying all effect sizes and predictors by the inverse of their standard errors (Cheung, 2008), or through the use of definition variables (Cheung, 2014)
- Moderators can be added as additional predictors of the effect size
Meta-analysis with SEM

Reproduced from Cheung (2008)
Meta-analysis with MLM

- Effect sizes are treated as the outcome variable
  - Mean and variance of effect size are estimated as a random intercept
  - Effect sizes are weighted by fixing $e_i$ to the sampling variance of each study
    - Moderators can be added as additional predictors of the effect size

$$y_i = \beta_0 + u_i + e_i$$

$$u_i \sim N(0, \tau^2)$$
Multiple group models involve splitting a sample into groups based on a categorical variable and simultaneously estimating models for each group.

- Model parameters can be equated across groups
- Multiple group models can be used to test if any model parameter differs across groups by comparing models
- Traditionally used in SEM contexts
Multiple group models: MLM

- Uses an approach similar to multivariate MLM (Goldstein, 1995)
- For each group an indicator variable is created
- Indicator variables are included in a model without an overall intercept
  - Fixed and random coefficients can vary across groups

\[ y_i = \beta_d d_1 + \beta_d d_2 + u_d d_1 + u_d d_2 + e_i \]

\[
\begin{bmatrix}
  u_{d1} \\
  u_{d2}
\end{bmatrix}
\sim N \left( \begin{bmatrix} \tau_{00}^2 & 0 \\ 0 & \tau_{11}^2 \end{bmatrix} \right)
\]
Multiple group models: MLM

- Differences between groups can be addressed with model comparisons
  - Comparison model is one with a single parameter estimated in across groups

\[ y_i = \beta_0 + u_{d1}d1 + u_{d2}d2 + e_i \]

\[
\begin{bmatrix}
  u_{d1} \\
  u_{d2}
\end{bmatrix}
\sim N \begin{pmatrix}
  \tau_{00}^2 & 0 \\
  0 & \tau_{11}^2
\end{pmatrix}
\]
Multiple group models

- Multiple group models provide a major advantage over traditional tests of moderators: the ability to allow any model parameter to differ across groups
  - Traditional tests for moderators only allow mean effect sizes to differ across groups
- Multiple group models allow meta-analysts to test for differences in the variance of effect sizes across groups
Multiple group models

- Differences in the variance of effect sizes may have important practical and theoretical implications:
  - Additional moderators that affect only one group
  - Effect size is more variable as a function of group membership
Multiple group models

Effect Size

Group 1

Group 2

Alexander M. Schoemann (ECU)

Using Multiple Group Modeling to Test Moderators in Meta-Analysis
Simulation Study: Data generation

- Two normally distributed populations (representing two levels of the moderator) of effect sizes were defined.

\[ \text{N} (\mu_1, \tau^2_1) \quad \text{N} (\mu_2, \tau^2_2) \]

- For each of \( k \) effect size pairs of independent random samples (\( n = 50 \)) were generated.

- Standardized mean difference and sampling error computed for each study.

- Group 1 means set to \( \mu = 0.2 \), Group 1 variance set to \( \tau^2 = 0.1 \)

- 500 replications per condition.

Data were generated using R v. 3.1.0.
Simulation Study: Data generation

- Two normally distributed populations (representing two levels of the moderator) of effect sizes were defined
  - \( N(\mu_1, \tau_1^2) \)
Simulation Study: Data generation

- Two normally distributed populations (representing two levels of the moderator) of effect sizes were defined
  - $N(\mu_1, \tau^2_1)$
  - $N(\mu_2, \tau^2_2)$

For each of $k$ effect size pairs of independent random samples ($n = 50$) were generated
- Standardized mean difference and sampling error computed for each study
- Group 1 means set to $\mu = 0.2$, Group 1 variance set to $\tau^2 = 0.1$

500 replications per condition

Data were generated using R v. 3.1.0

Alexander M. Schoemann (ECU)
Using Multiple Group Modeling to Test Moderators in Meta-Analysis
M3 2015 16 / 23
Simulation Study: Data generation

- Two normally distributed populations (representing two levels of the moderator) of effect sizes were defined
  - \( N(\mu_1, \tau_1^2) \)
  - \( N(\mu_2, \tau_2^2) \)
- \( k \) effect sizes were generated from each normal distribution

500 replications per condition

Data were generated using R v. 3.1.0
Simulation Study: Data generation

- Two normally distributed populations (representing two levels of the moderator) of effect sizes were defined
  - $N(\mu_1, \tau_1^2)$
  - $N(\mu_2, \tau_2^2)$

- $k$ effect sizes were generated from each normal distribution
  - For each of $k$ effect size pairs of independent random samples ($n = 50$) were generated

Data were generated using R v. 3.1.0

Alexander M. Schoemann (ECU)
Using Multiple Group Modeling to Test Moderators in Meta-Analysis
M3 2015 16 / 23
Simulation Study: Data generation

- Two normally distributed populations (representing two levels of the moderator) of effect sizes were defined
  - $N(\mu_1, \tau_1^2)$
  - $N(\mu_2, \tau_2^2)$

- $k$ effect sizes were generated from each normal distribution
  - For each of $k$ effect size pairs of independent random samples ($n = 50$) were generated
  - Standardized mean difference and sampling error computed for each study

Data were generated using R v. 3.1.0
Simulation Study: Data generation

- Two normally distributed populations (representing two levels of the moderator) of effect sizes were defined
  - $N(\mu_1, \tau_1^2)$
  - $N(\mu_2, \tau_2^2)$

- $k$ effect sizes were generated from each normal distribution
  - For each of $k$ effect size pairs of independent random samples ($n = 50$) were generated
  - Standardized mean difference and sampling error computed for each study
  - Group 1 means set to $\mu = 0.2$, Group 1 variance set to $\tau^2 = 0.1$
Simulation Study: Data generation

- Two normally distributed populations (representing two levels of the moderator) of effect sizes were defined
  - $N(\mu_1, \tau^2_1)$
  - $N(\mu_2, \tau^2_2)$
- $k$ effect sizes were generated from each normal distribution
  - For each of $k$ effect size pairs of independent random samples ($n = 50$) were generated
  - Standardized mean difference and sampling error computed for each study
  - Group 1 means set to $\mu = 0.2$, Group 1 variance set to $\tau^2 = 0.1$
- 500 replications per condition

Data were generated using R v. 3.1.0

Alexander M. Schoemann (ECU)
Using Multiple Group Modeling to Test Mod
Simulation Study: Data generation

- Two normally distributed populations (representing two levels of the moderator) of effect sizes were defined
  - \( N(\mu_1, \tau_1^2) \)
  - \( N(\mu_2, \tau_2^2) \)

- \( k \) effect sizes were generated from each normal distribution
  - For each of \( k \) effect size pairs of independent random samples (\( n = 50 \)) were generated
  - Standardized mean difference and sampling error computed for each study
  - Group 1 means set to \( \mu = 0.2 \), Group 1 variance set to \( \tau^2 = 0.1 \)

- 500 replications per condition
- Data were generated using R v. 3.1.0
Simulation Study: Conditions

- Three manipulated factors:
  - Number of studies ($k = 10$, $k = 50$)
  - Mean difference ($\mu_{\text{diff}} = 0$, $\mu_{\text{diff}} = .2$)
  - Variance difference ($\tau^2_{\text{diff}} = 0$, $\tau^2_{\text{diff}} = .5$)

- $2 \times 2 \times 2 = 8$ conditions
Simulation Study: Analysis

- Data analyzed using both SEM and MLM approach
  - SEM fit with Mplus v. 7.2
  - MLM fit with PROC MIXED in SAS 9.3
- Three models fit for each technique
  1. Means and variances allowed to differ across groups
  2. Means constrained to equality across groups, variances allowed to differ
  3. Variances constrained to equality across groups, means allowed to differ
- All analyses used FIML estimation
Simulation Study: Results

- All models converged in all simulation conditions
Simulation Study: Results

- All models converged in all simulation conditions
- MLM had slightly higher MSE for mean effect sizes than SEM
Simulation Study: Results

- All models converged in all simulation conditions
- MLM had slightly higher MSE for mean effect sizes than SEM
  - MSE did not differ between models where variances were allowed differ and models with variances constrained to be equal
- Power and Type I error rates for tests of mean differences did not differ between MLM and SEM
- Power for tests of variance differences did not differ between SEM and MLM
- MLM had lower type I error rates for tests of variance differences than SEM
Simulation Study: Results

- All models converged in all simulation conditions
- MLM had slightly higher MSE for mean effect sizes than SEM
  
  ▶ MSE did not differ between models where variances were allowed differ and models with variances constrained to be equal

- Power and Type I error rates for tests of mean differences did not differ between MLM and SEM
Simulation Study: Results

- All models converged in all simulation conditions
- MLM had slightly higher MSE for mean effect sizes than SEM
  - MSE did not differ between models where variances were allowed differ and models with variances constrained to be equal
- Power and Type I error rates for tests of mean differences did not differ between MLM and SEM
- Power for tests of variance differences did not differ between SEM and MLM
Simulation Study: Results

- All models converged in all simulation conditions
- MLM had slightly higher MSE for mean effect sizes than SEM
  - MSE did not differ between models where variances were allowed differ and models with variances constrained to be equal
- Power and Type I error rates for tests of mean differences did not differ between MLM and SEM
- Power for tests of variance differences did not differ between SEM and MLM
- MLM had lower type I error rates for tests of variance differences than SEM
Real-world example

- Data for this example were used by Field & Gillett (2010)
Data for this example were used by Field & Gillett (2010)

- Data originally come from a study by Else-Quest and colleagues (2006)
Real-world example

- Data for this example were used by Field & Gillett (2010)
  - Data originally come from a study by Else-Quest and colleagues (2006)
- Meta-analysis of relationship between child gender and temperament

30 Studies included in analyses
- 17 studies on infants, 13 studies on toddlers

Analyzed using both SEM and MLM
Real-world example

- Data for this example were used by Field & Gillett (2010)
  - Data originally come from a study by Else-Quest and colleagues (2006)
- Meta-analysis of relationship between child gender and temperament
  - Represented by a standardized mean difference
Real-world example

- Data for this example were used by Field & Gillett (2010)
  - Data originally come from a study by Else-Quest and colleagues (2006)
- Meta-analysis of relationship between child gender and temperament
  - Represented by a standardized mean difference
- Does child age moderate this relationship?
Real-world example

- Data for this example were used by Field & Gillett (2010)
  - Data originally come from a study by Else-Quest and colleagues (2006)
- Meta-analysis of relationship between child gender and temperament
  - Represented by a standardized mean difference
- Does child age moderate this relationship?
  - Child age was coded as infant or toddler (categorical)
Real-world example

- Data for this example were used by Field & Gillett (2010)
  - Data originally come from a study by Else-Quest and colleagues (2006)
- Meta-analysis of relationship between child gender and temperament
  - Represented by a standardized mean difference
- Does child age moderate this relationship?
  - Child age was coded as infant or toddler (categorical)
- 30 Studies included in analyses
Real-world example

- Data for this example were used by Field & Gillett (2010)
  - Data originally come from a study by Else-Quest and colleagues (2006)
- Meta-analysis of relationship between child gender and temperament
  - Represented by a standardized mean difference
- Does child age moderate this relationship?
  - Child age was coded as infant or toddler (categorical)
- 30 Studies included in analyses
  - 17 studies on infants, 13 studies on toddlers
Real-world example

- Data for this example were used by Field & Gillett (2010)
  - Data originally come from a study by Else-Quest and colleagues (2006)
- Meta-analysis of relationship between child gender and temperament
  - Represented by a standardized mean difference
- Does child age moderate this relationship?
  - Child age was coded as infant or toddler (categorical)
- 30 Studies included in analyses
  - 17 studies on infants, 13 studies on toddlers
- Analyzed using both SEM and MLM
Real-world example

- Results did not differ between SEM and MLM approaches
Real-world example

- Results did not differ between SEM and MLM approaches
- Mean effect size differed between infants \( (\mu = -0.08) \) and toddlers \( (\mu = 0.24) \), \( \chi^2(1) = 5.21, p = .022 \)
Real-world example

- Results did not differ between SEM and MLM approaches
- Mean effect size differed between infants ($\mu = -0.08$) and toddlers ($\mu = 0.24$), $\chi^2(1) = 5.21$, $p = .022$
- Variance of effect sizes differed between infants ($\tau^2 = 0.01$) and toddlers ($\tau^2 = .128$), $\chi^2(1) = 6.31$, $p = .012$
Real-world example

- Results did not differ between SEM and MLM approaches
- Mean effect size differed between infants ($\mu = -0.08$) and toddlers ($\mu = 0.24$), $\chi^2(1) = 5.21$, $p = .022$
- Variance of effect sizes differed between infants ($\tau^2 = 0.01$) and toddlers ($\tau^2 = .128$), $\chi^2(1) = 6.31$, $p = .012$
- Infants have very small, consistent gender differences in temperament
Real-world example

- Results did not differ between SEM and MLM approaches
- Mean effect size differed between infants ($\mu = -0.08$) and toddlers ($\mu = 0.24$), $\chi^2(1) = 5.21$, $p = .022$
- Variance of effect sizes differed between infants ($\tau^2 = 0.01$) and toddlers ($\tau^2 = .128$), $\chi^2(1) = 6.31$, $p = .012$
- Infants have very small, consistent gender differences in temperament
- Toddlers show a larger gender differences in temperament, but this difference is quite variable across studies
Real-world example

- Results did not differ between SEM and MLM approaches
- Mean effect size differed between infants ($\mu = -0.08$) and toddlers ($\mu = 0.24$), $\chi^2(1) = 5.21$, $p = .022$
- Variance of effect sizes differed between infants ($\tau^2 = 0.01$) and toddlers ($\tau^2 = .128$), $\chi^2(1) = 6.31$, $p = .012$
- Infants have very small, consistent gender differences in temperament
- Toddlers show a larger gender differences in temperament, but this difference is quite variable across studies
  - Perhaps other variables influence the relationship between gender and temperament in toddlers
Future directions

- Impact of heterogeneity on tests of mean effect size differences
  - Under what conditions can multiple group modeling provide more accurate tests of moderators?
- Investigate multiple group models with more advanced meta-analysis models
  - Multivariate meta-analysis
  - Multilevel meta-analysis
  - Bayesian meta-analysis
- Performance of REML estimation with multiple group models
- Models testing for differences in variance across continuous moderators
Thank you!

- Questions?
- email: schoemanna@ecu.edu
- Example code from today at:
  https://sites.google.com/site/alexandermschoemann/