Dealing with Artificially Dichotomized Variables in Meta-Analytic Structural Equation Modeling

H. de Jonge, S. Jak, & K.J. Kan
University of Amsterdam
H.deJonge@uva.nl
Research Synthesis 2019 Conference
Meta-analysis

To systematically synthesize all the empirical studies that are published

MASEM (Becker, 1992, 1995; Viswesvaran & Ones, 1995)
  - Testing a complete hypothesized model
  - Provides parameter estimates & overall model fit
  - **Stage 1:** Pooling correlation coefficients in a matrix
  - **Stage 2:** Fitting SEM on this pooled correlation matrix

Effect size: strength and direction of the association

In primary studies expressed in different ways depending on
  - The nature of the variables
  - The way the variables are measured or analyzed
Artificial dichotomization

- Meta-analyses

- Dichotomous variable
  - Natural or artificial

- Often argued against artificial dichotomization (e.g., Cohen, 1983; MacCallum et al., 2002)

- Meta-analysts frequently have to deal with artificially dichotomized variables in primary studies
To estimate a pooled correlation matrix

- Primary studies may report different kinds of effect sizes

- One needs to express the bivariate effect sizes as correlation coefficients

- Based on information provided in primary studies
  - The point-biserial and biserial correlation can be calculated
The (point-)biserial correlation

× Point-biserial correlation (Lev, 1949; Tate, 1954)
  × Association between natural dichotomous and continuous variable
  × Relationship between *artificially* dichotomized and continuous variable →
    Typically leading to an underestimation (e.g., Cohen, 1983; MacCallum et al., 2002)

× Biserial correlation (Pearson, 1909)
  × Assumes a continuous, normally distributed variable underlying the dichotomous variable
  × Relationship between *artificially* dichotomized and continuous variable →
    Should generally provide an unbiased estimate (Soper, 1914; Tate, 1955)

× Affect meta-analytic results in the same direction (Jacobs & Viechtbauer, 2017)
Aim

- Investigate the effects of using (1) the point-biserial correlation and (2) the biserial correlation for the relationship between an artificially dichotomized variable and a continuous variable on MASEM-parameters and model fit.
Simulation study

- Choices mainly based on typical situations in educational research

- Population model with fixed parameter values

- Systematically varied:
  - Size of $\beta_{MX}$ (.16, .23, .33) (de Jonge & Jak, 2018)
  - Percentage of dichotomization (25%, 75%, 100%)
  - Cut-off point of dichotomization (.5, .1)

- Number of primary studies: 44 (de Jonge & Jak, 2018)

- Within primary study sample sizes: randomly sampled from a positively skewed distribution (Hafdahl, 2007) with a mean of 421.75 (de Jonge & Jak, 2018)

- 39% missing correlations (Sheng, Kong, Cortina, & Hou, 2016)

- Random-effects two stage structural equation modeling (Cheung, 2014)
Estimation bias

× Relative percentage bias in $\beta_{MX}$
  × **Point-biserial correlation:** $-41.70\%$ to $-5.05\%$
  × $\beta_{MX}$ seems systematically underestimated
  × **Biserial correlation:** $-0.36\%$ to $0.35\%$
  × No substantial bias in $\beta_{MX}$

× Relative percentage bias in $\beta_{MY}$
  × **Point-biseral & Biserial:** $< 5\%$ in all conditions (Hoogland & Boomsma, 1998)
  × No substantial bias in $\beta_{MY}$

× Relative percentage bias in standard errors of
  × **Point-biserial & Biserial:** both path coefficients $< 10\%$ in all conditions (Hoogland & Boomsma, 1998)
  × **Biserial $\rightarrow$ $\beta_{MX}$** and **$\beta_{MY}$** seems systematically negative
  × **Point-biserial $\rightarrow$ $\beta_{MY}$** seems systematically negative
Some possible causes

× Biserial correlation → negative bias in $SE$ of $\beta_\text{MX}$
  × Used formulas for estimating the sampling (co)variances
  × Generally leads to an underestimation of the true sampling variance (Jacobs & Viechtbauer, 2017)

× Sampling (co)variances from the primary studies are treated as known in MASEM
  × Underestimation in standard errors in univariate random-effects meta-analysis (Sánchez-Meca & Marín-Martínez, 2008; Viechtbauer, 2005)

× Note → bias was within the limit of 10%

× Future research is needed
We advise researchers who want to apply MASEM and want to investigate mediation to convert the effect size between any artificially dichotomized predictor and continuous variable to a:

- Biserial correlation
Thank you!

Any questions?

✉️ H.deJonge@uva.nl
References


Pearson, K. (1909). On a new method of determining correlation between a measured character a, and a character b, of which only the percentage of cases wherein b exceeds (or falls short of) a given intensity is recorded for each grade of a. *Biometrika*, 7, 96-105. doi:10.2307/2345365


