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

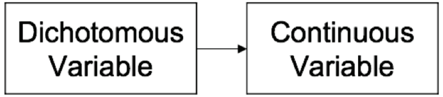
- To systematically synthesize all the empirical studies that are published

- 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-analysists 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)

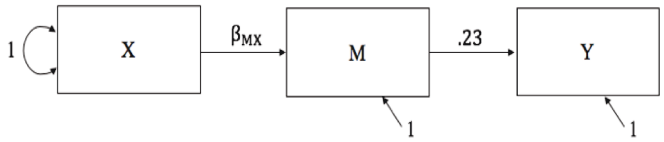
- Affect meta-analytic results in the same direction (Jacobs & Viechtbauer, 2017)

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.

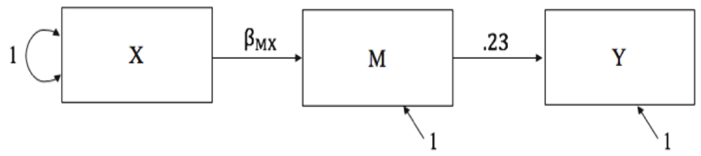
Simulation study

- Choices mainly based on typical situations in educational research
- Population model with fixed parameter values



- Systematically varied:
  - Size of β_{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 β_{MX}
  - **Point-biserial correlation:** −41.70% to −5.05%
    - β_{MX} seems systematically underestimated
  - **Biserial correlation:** −0.36% to 0.35%
    - No substantial bias in β_{MX}

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

- Relative percentage bias in standard errors of
  - **Point-biserial & Biserial:** both path coefficients < 10% in all conditions (Hoogland & Boomsma, 1998)
  - **Biserial** → β_{MX} and β_{MY} seems systematically negative
  - **Point-biserial** → β_{MY} seems systematically negative

Some possible causes

- Biserial correlation → negative bias in *SE* of β_{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

Conclusion

- 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

Becker, B. J. (1992). Using results from replicated studies to estimate linear models. *Journal of Educational Statistics, 17*, 341–362. doi:10.2307/1165128

Becker, B. J. (1995). Corrections to "using results from replicated studies to estimate linear models". *Journal of Educational and Behavioral Statistics, 20*, 100–102. doi:10.2307/1165390

Cheung, M. W. L. (2014). Fixed-and random-effects meta-analytic structural equation modeling: Examples and analyses in R. *Behavior Research Methods*, *46*, 29–40. doi:10.3758/s13428-013-0361-y

Cohen, J. (1983). The cost of dichotomization. *Applied psychological measurement*, *7*, 249-253. https://doi.org/10.1177/014662168300700301

de Jonge, H., & Jak, S. (2018, June). *A Meta-Meta-Analysis: Identifying Typical Conditions of Meta-Analyses in Educational Research.* Paper presented at the conference Research Synthesis 2018 of Leibniz Institute for Psychology Information, Trier, Germany. http://dx.doi.org/10.23668/psycharchives.853

Hafdahl, A. R. (2007). Combining correlation matrices: Simulation analysis of improved fixed-effects methods. *Journal of Educational and Behavioral Statistics*, *32*, 180–205. doi:10.3102/1076998606298041

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods & Research*, *26*, 329–367. doi:10.1177/0049124198026003003

Jacobs, P., & Viechtbauer, W. (2017). Estimation of the biserial correlation and its sampling variance for use in meta-analysis. *Research synthesis methods*, *8*, 161–180. doi:10.1002/jrsm.1218

Lev, J. (1949). The point biserial coefficient of correlation. *The Annals of Mathematical Statistics*, *20*, 125-126. doi:10.1214/aoms/1177730103

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

Sánchez-Meca, J., & Marín-Martínez, F. (2008). Confidence intervals for the overall effect size in random-effects meta-analysis. *Psychological Methods*, *13*, 31.doi:10.1037/1082-989X.13.1.31

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological methods*, *7*, 19–40. doi:10.1037//1082-989X.7.1.19

Tate, R. F. (1954). Correlation between a discrete and a continuous variable. Point-biserial correlation. *The Annals of mathematical statistics*, *25*, 603-607.

Tate, R. F. (1955). The theory of correlation between two continuous variables when one is dichotomized. *Biometrika*, *42*, 205–216. doi:10.2307/2333437

Sheng, Z., Kong, W., Cortina, J. M., & Hou, S. (2016). Analyzing matrices of meta-analytic correlations: current practices and recommendations. *Research synthesis methods*, *7*, 187–208. doi:10.1002/jrsm.1206

Soper, H. E. (1914). On the probable error of the bi-serial expression for the correlation coefficient. *Biometrika*, *10*, 384–390. doi:10.2307/2331789

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics*, *30*, 261-293.

Viswesvaran, C., & Ones, D. (1995). Theory testing: Combining psychometric meta-analysis and structural equations modeling. *Personnel Psychology, 48*, 865–885. doi:10.1111/j.1744-6570.1995.tb01784.x