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

Hannelies de Jonge, Suzanne Jak, & Kees Jan Kan
University of Amsterdam
H.deJonge@uva.nl

Psychoco 2020: International Workshop on Psychometric Computing

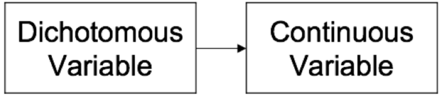
Meta-analysis

- 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:** Hypothesized model fitted to a pooled correlation matrix using SEM

- How to deal with primary studies in which variables have been artificially dichotomized?

Artificial dichotomization

- 

- 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

Estimating 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

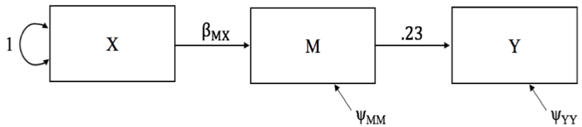
- Meta-analysist may not be aware of the difference
- **Point-biserial** correlation (Lev, 1949; Tate, 1954)
  - Association between natural dichotomous and continuous variable
- **Biserial** correlation (Pearson, 1909)
  - Assumes a continuous, normally distributed variable underlying the dichotomous variable
- Previous research

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 1: full mediation

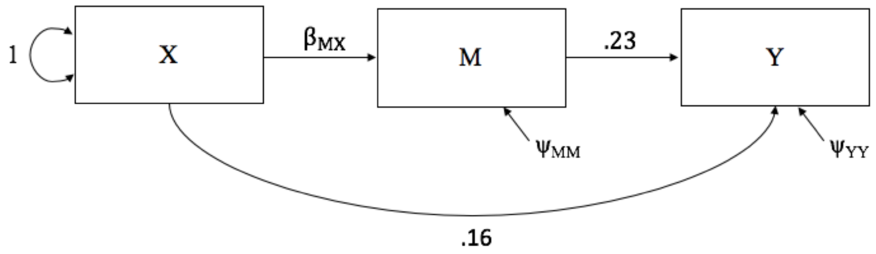
- Choices mainly based on typical situations in educational research
- Population model with fixed parameter values



- Systematically varied:
  - Percentage of dichotomization (25%, 75%, 100%)
  - Size of β_{MX} (.16, .23, .33) (de Jonge & Jak, 2018)
  - 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)
- In each condition, we generated 2000 meta-analytic datasets
- Random-effects two stage structural equation modeling (Cheung, 2014)

Simulation study 2: partial mediation

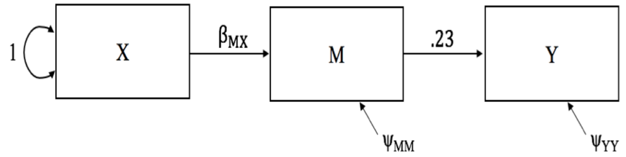
- Population model with fixed parameter values



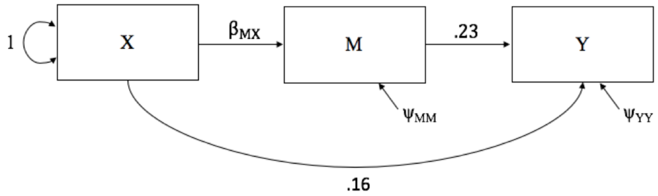
- Same conditions as in the first simulation study

Relative percentage bias in β_{MX}

Simulation study 1



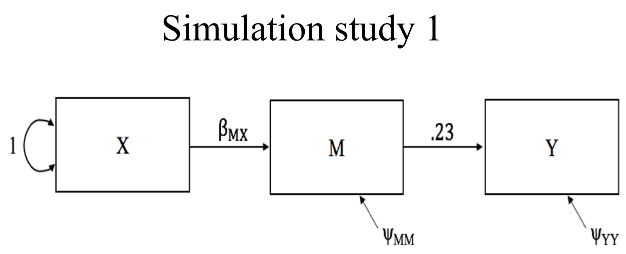
Simulation study 2

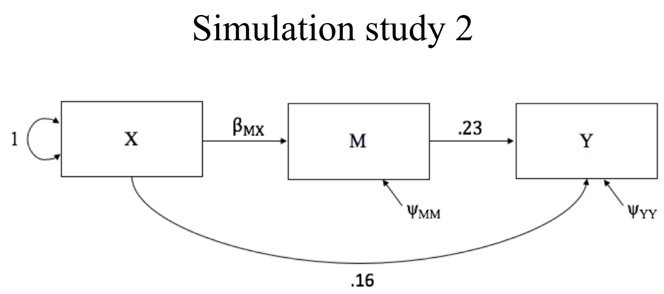


- **Point-biserial correlation:**
  - *Full mediation:* −41.70% to −5.05%
  - *Partial mediation:* −41.68% to −5.05%
    - > 5% (Hoogland & Boomsma, 1998) → β_{MX} seems systematically underestimated

- **Biserial correlation:**
  - *Full mediation:* −0.36% to 0.35%
  - *Partial mediation:* −0.42% to 0.25%
    - < 5% (Hoogland & Boomsma, 1998) → No substantial bias in β_{MX}

Relative percentage bias in β_{YM}



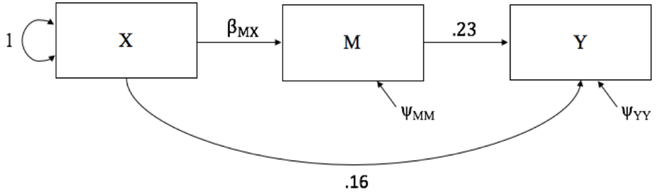


- Full mediation
  - **Point-biseral & Biserial:** < 5% (Hoogland & Boomsma, 1998)
    - No substantial bias in β_{YM}

- Partial mediation
  - **Point-biseral:** 1.17% to 15.56% (in 10 of the 18 conditions > 5%)
    - β_{YM} seems systematically overestimated
  - **Biserial:** −0.36% to 0.47%
    - < 5% (Hoogland & Boomsma, 1998) → No substantial bias in β_{YM}

Relative percentage bias in β_{YX}

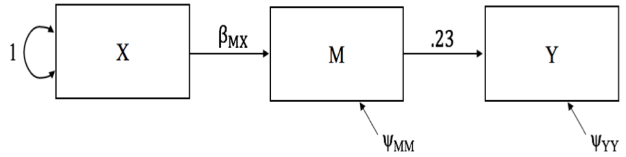
Simulation study 2



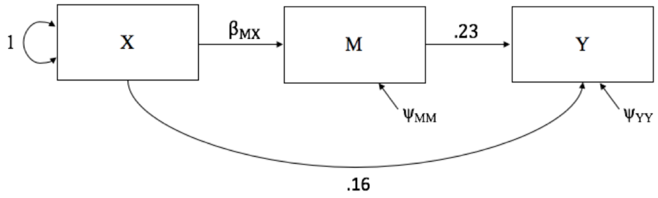
- **Point-biserial correlation:**
  - -45.85% to -5.30%
    - $> 5\%$ (Hoogland & Boomsma, 1998) → β_{YX} seems systematically underestimated
- **Biserial correlation:**
  - -0.54% to -0.80%
    - $<5\%$ (Hoogland & Boomsma, 1998) → No substantial bias in β_{YX}
- **Indirect effects**

Relative percentage bias in standard errors

Simulation study 1

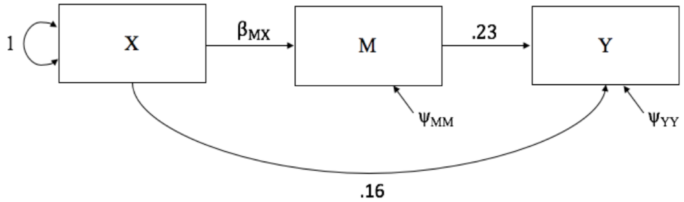


Simulation study 2



- **Point-biserial & Biserial:** β_{MX}, β_{YM}, and β_{YX} typically < 10% (Hoogland & Boomsma, 1998)
- **Biserial →** β_{MX} and β_{YM} seems systematically negative
- **Point-biserial →** β_{YM} 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)

- **Biserial & point-biserial** correlation → negative bias in *SE* of β_{YM}
  - When the data were not dichotomized at all
  - The *SE*s of the pooled correlation coefficients between M and Y in Stage 1
  - 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 typically within the limit of 10%

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

Do you want to read the whole article?

Please see https://osf.io/j6nxt/