Modeling Statistical Artefacts in Meta-Analysis

Brenton M. Wiernik

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

UNIVERSITY of SOUTH FLORIDA

https://psyarxiv.com/9mpbn/

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

Check for updates

ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Invited Forum: Meta-Analysis
General Article

Advances in Methods and Practices in Psychological Science
2020, Vol. 3(1) 94–123
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2515245919885611
www.psychologicalscience.org/AMPPS
SAGE

Obtaining Unbiased Results in Meta-Analysis: The Importance of Correcting for Statistical Artifacts

Brenton M. Wiernik[1] and Jeffrey A. Dahlke[2]
[1]Department of Psychology, University of South Florida, and [2]Human Resources Research Organization, Alexandria, Virginia

Abstract

Most published meta-analyses address only artifactual variance due to sampling error and ignore the role of other statistical and psychometric artifacts, such as measurement error variance (due to factors including unreliability of measurements, group misclassification, and variable treatment strength) and selection effects (including range restriction or enhancement and collider biases). These artifacts can have severe biasing effects on the results of individual studies and meta-analyses. Failing to account for these artifacts can lead to inaccurate conclusions about the mean effect size and between-studies effect-size heterogeneity, and can influence the results of meta-regression, publication-bias, and sensitivity analyses. In this article, we provide a brief introduction to the biasing effects of measurement error variance and selection effects and their relevance to a variety of research designs. We describe how to estimate the effects of these artifacts in different research designs and correct for their impacts in primary studies and meta-analyses. We consider meta-analyses of correlations, observational group differences, and experimental effects. We provide R code to implement the corrections described.

Keywords

psychometric meta-analysis, measurement error, reliability, range restriction, range variation, selection bias, collider bias, open materials

Received 11/6/18; Revision accepted 10/8/19

What’s the purpose of meta-analysis?

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

What’s the purpose of meta-analysis?

Widespread perspective:

To summarize and critique a body of studies

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

What’s the purpose of meta-analysis?

Alternative perspective:
To enhance studies

To facilitate the best possible inferences about their research questions

To identify and correct sources of bias

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

Factors that obscure study results

- Sampling error
  - Artificial variability and inconsistency across studies

- Confounding

- Poor measurement
  - Measurement error, measure contamination

- Biased sampling
  - Attenuated effects, collider bias

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

Statistical Artefacts in Meta-Analysis

- In recent years, a lot of attention has been paid to addressing confounding, collider bias, selection bias, measurement error in primary studies

- Meta-analysis can also be leveraged to estimate the impacts of these sources of bias
  - How much might effects have been biased by poor measurement?
  - How might biased sampling have impacted effects?
  - Can we pull information from other samples to remove confounding?

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

Meta-analysis as bias correction has a long history

Journal of Applied Psychology
1977, Vol. 62, No. 5, 529–540

Development of a General Solution to the Problem of Validity Generalization

Frank L. Schmidt
U. S. Civil Service Commission and
George Washington University

John E. Hunter
Michigan State University

Personnel psychologists have traditionally believed that employment test validities are situation specific. This study presents a Bayesian statistical model which allows one to explore the alternate hypothesis that variation in validity outcomes from study to study for similar jobs and tests is artifactual in nature. Certain outcomes using this model permit validity generalization to new settings without carrying out a validation study of any kind. Where such generalization is not justified, the procedure provides an improved method of data analysis and decision making for the necessary situational validity study. Application to four distributions of empirical validity coefficients demonstrated the power of the model.

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

UNIVERSITY of SOUTH FLORIDA

Connection with risk of bias assessment

- Methods for assessing risk of bias in studies have become routine
  - Especially outside of psychology

- Common sources of bias include
  - Lack of control groups, randomization
  - Lack of control for confounding factors
  - Poor quality measurement
  - Restricted sampling

- Many of these sources have estimable impacts on results and can be statistically adjusted

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

Connection with causal inference

- Causal modeling approaches are increasingly have spread widely to other social and biomedical sciences
  - Causal models, directed acyclic graphs (DAGs), generative models
  - Common practices from epidemiology and econometrics

- There are deep connections between causal models and, eg, corrections for measurement error and collider bias

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

Connection with model-based meta-analysis

- Model-based meta-analysis (Becker 2009, *HRSMA*)
- Meta-analytic structural equations modeling (Cheung,)

- Approaches to meta-analysis aimed at estimating full statistical models, rather than single parameters
  - Adjust for confounding
  - Might include measurement models

- Statistical adjustments to effect sizes use similar ideas, drawing on information from individual studies

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

Measurement error

- Artefact that causes observed (i.e., measured) values to deviate from the “true” values of underlying latent variables

- Also called
  - Unreliability
  - Observational error
  - Information bias
  - Low precision
  - Misclassification

- Examples
  - Fluctuations in scale responses
  - Treatment non-compliance
  - Rater idiosyncrasies
  - Instrument idiosyncrasies

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

Types of measurement error

Source	Description
Random response error	Truly random error specific to each item/response; unique to each measurement
Transient error	Error due the specific period or environment in which data are gathered; shared by measures completed within a meaningfully short time span
Content sampling / instrument error	Error due to specific content or features of the measure/instrument used; shared by measures with the same or highly similar content or features
Rater sampling / source error	Error due to the specific raters or sources used to gather data; shared by measures with the same rater

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

Systematic and random measurement error

- Systematic error (*bias*) affects each score in the same manner (e.g., consistent underestimates)
  - Mean error across persons
- Random error affects each score differently and refers to the *variance of the errors* across persons.

- Typically, only the error variance affects standardized effect sizes such as correlations and Hedges's *g*
  - Differential bias across groups or score ranges could also impact correlations or *g* values

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

Quantifying measurement error

- For any single measurement, we don't know the direction or size of measurement error
- But across many measurements, we can quantify how measurement error affects variability

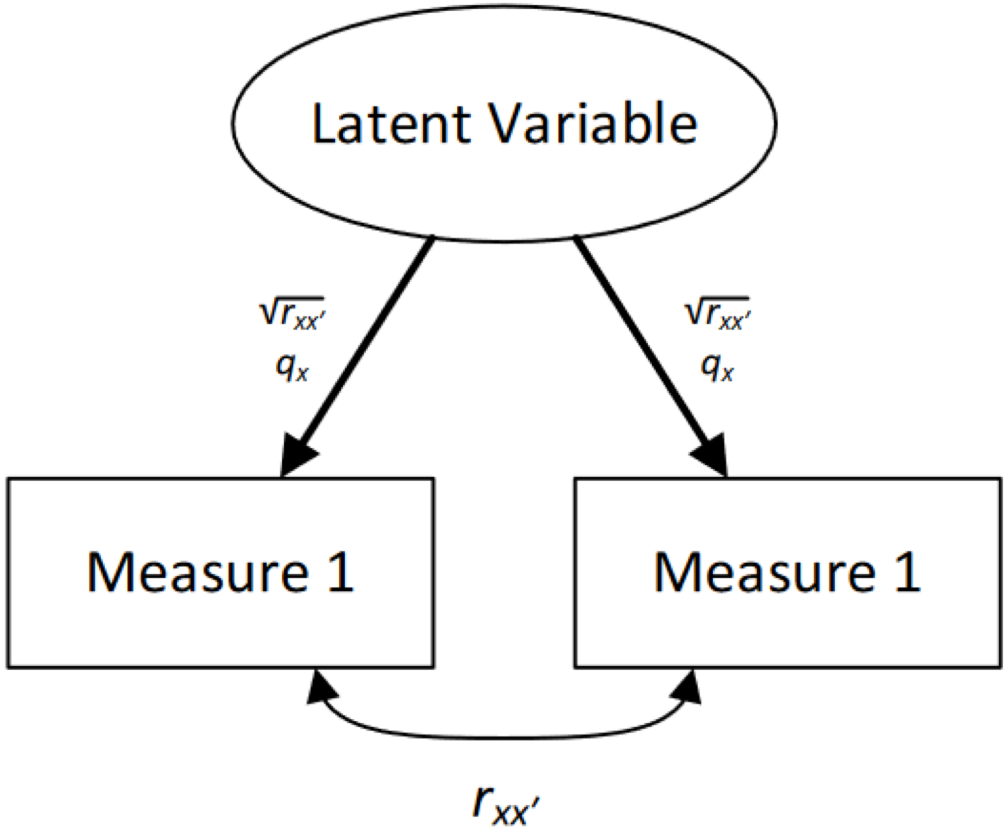
- We can conduct reliability studies to estimate the magnitude of measurement error variance
  - e.g., how consistent are measurements over time, across raters, across items in a scale?

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

Quantifying measurement error

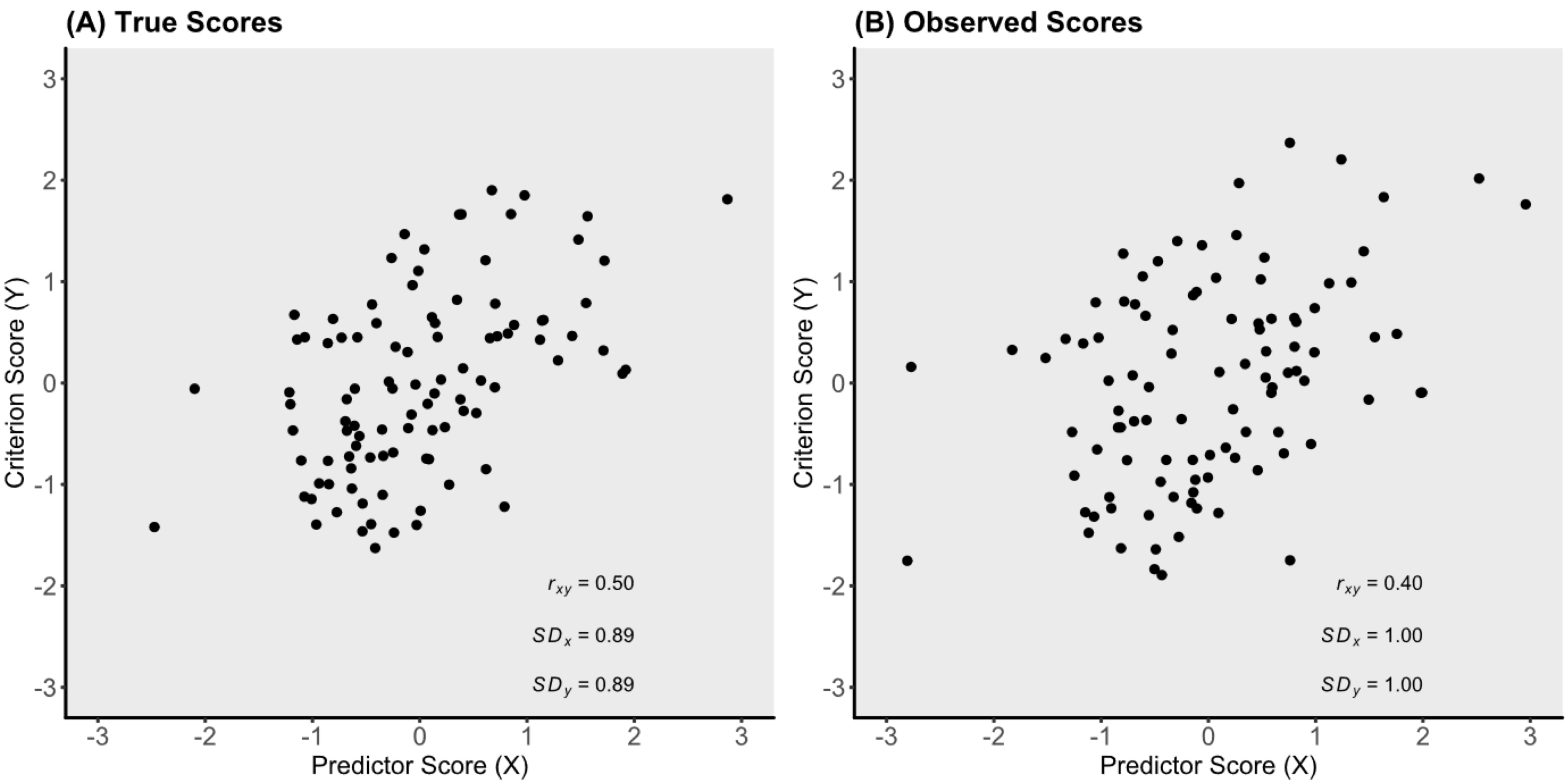
- Assume measurement error is independent of:
  - True latent variable
  - Other measurement errors

- Reliability coefficient
  - Correlation between measurements
  - $r_{xx'} = \frac{\sigma^2_{true}}{\sigma^2_{obs}} = 1 - \frac{\sigma^2_{error}}{\sigma^2_{obs}}$



Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

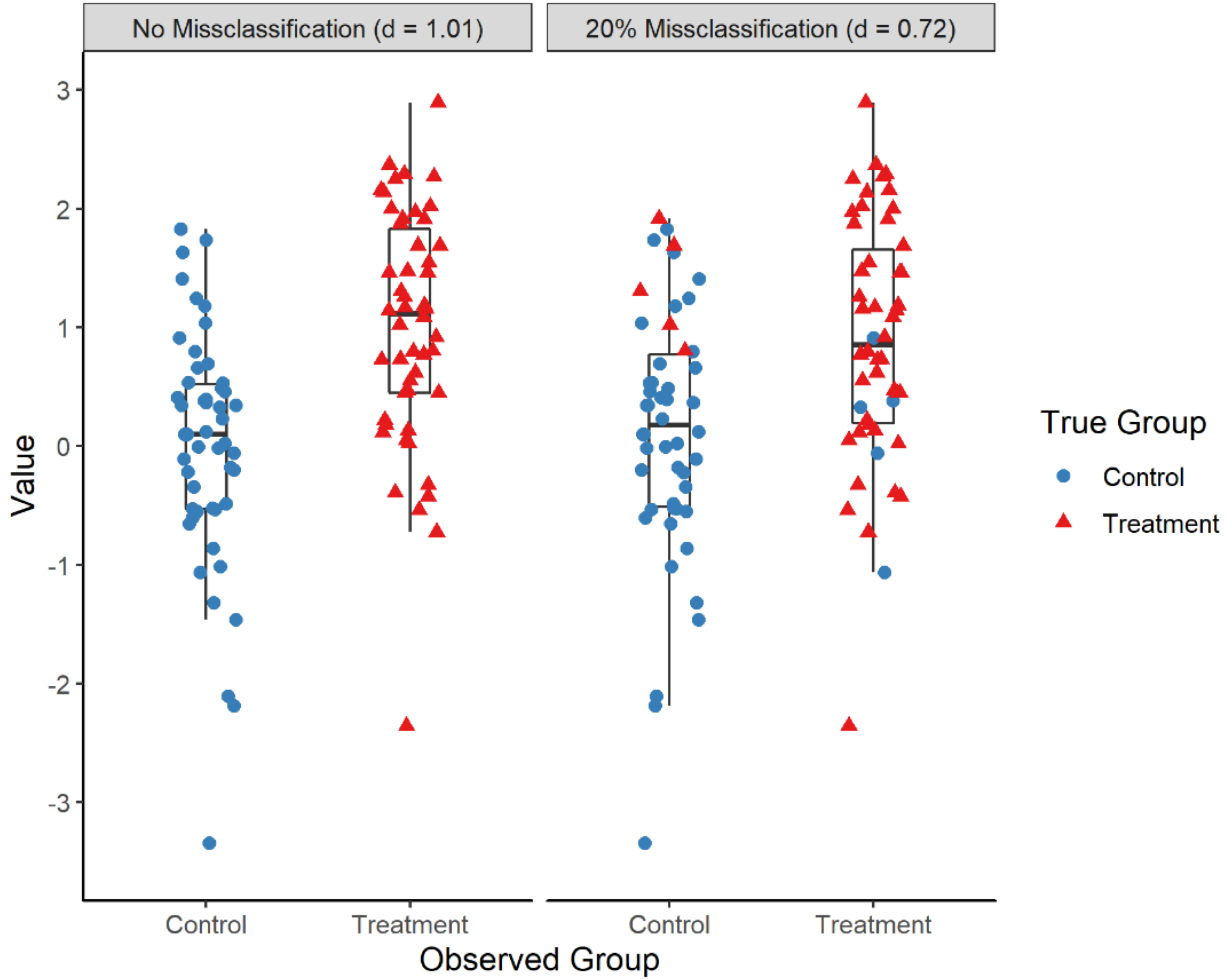
Effect of measurement error on results



Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

UNIVERSITY of SOUTH FLORIDA

Effect of measurement error on results



Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

Measurement error and meta-analysis

1. Mean effect size
   - Generally, systematic null-bias

$$\bar{r}_{xy} = \bar{\rho} \times \overline{\sqrt{r_{xx'}}} \times \overline{\sqrt{r_{yy'}}} = -.32 \times .89 \times .84 = -.24$$

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

Measurement error and meta-analysis

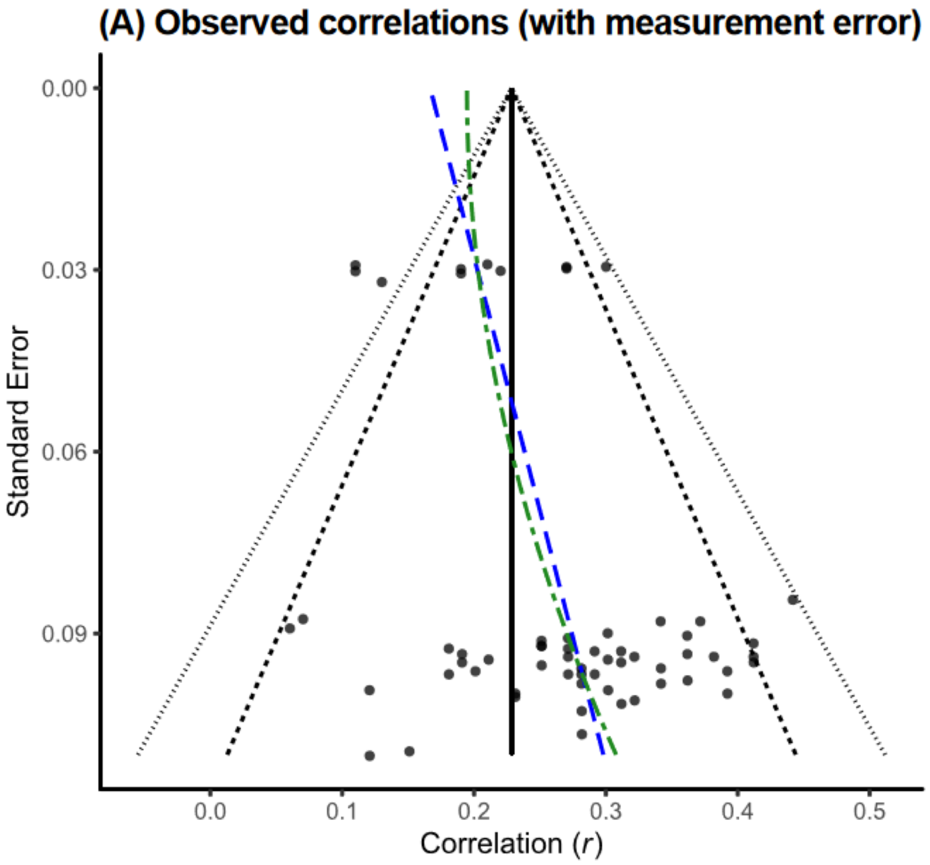
2. Heterogeneity and moderator analysis
   - Ignoring measurement error = overestimate heterogeneity τ
   - Differential measurement error can bias moderator analyses

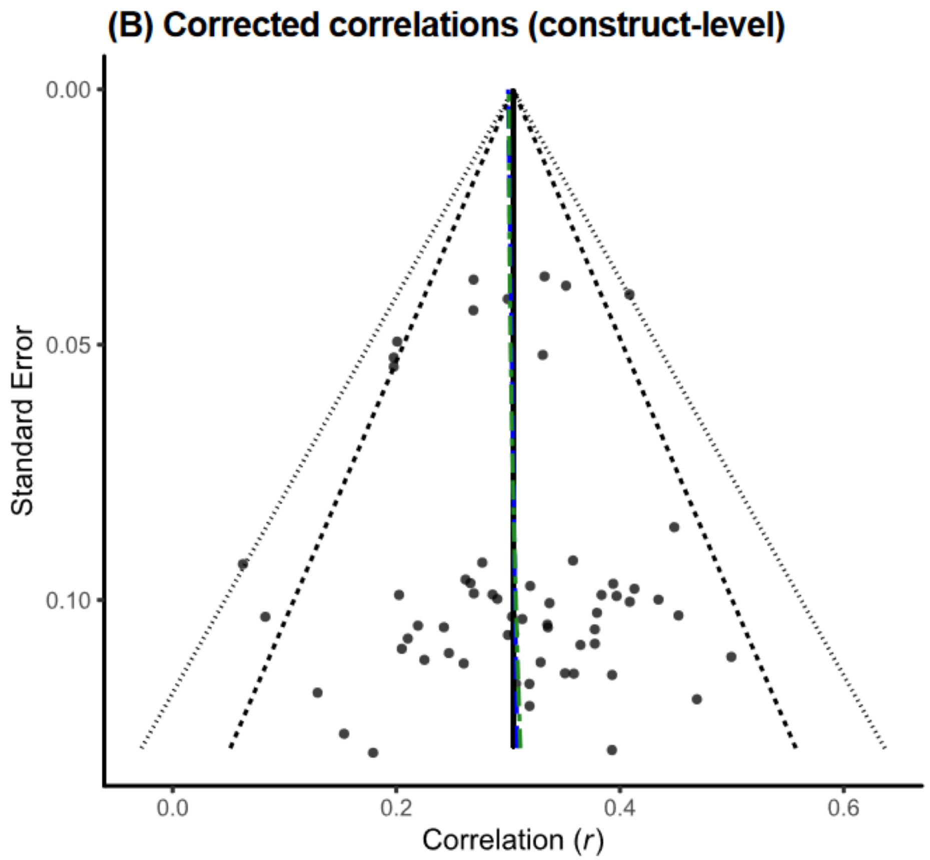
   - e.g., Comparing cognitive behavioral therapy (CBT) and mindfulness therapy for PTSD
   - Therapies are really equally effective ($\bar{\delta}$ = .40)
   - CBT studies measure PTSD more reliably ($\bar{r}_{yy'}$ = .90) than mindfulness studies ($\bar{r}_{yy'}$ = .60)
   - Observed effect sizes suggest moderation ($\bar{d}_{CBT}$ = .38 vs $\bar{d}_{Mind}$ = .31)

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

Measurement error and meta-analysis

3. Publication bias





Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

UNIVERSITY of SOUTH FLORIDA

Correcting for measurement error in measured variables

$$r_c = \frac{r_{obs}}{\sqrt{r_{xx'}}\sqrt{r_{yy'}}}$$

$$d_c = d_{obs} \Big/ \sqrt{r_{yy'_{pooled}}}$$

$$SE^2_{r_c} = SE^2_{r_{obs}} \times (r_c / r_{obs})^2$$

$$SE^2_{d_c} = SE^2_{d_{obs}} \times (d_c / d_{obs})^2$$

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

Correcting for measurement error (group misclassification)

r_{gG} = correlation of observed group with actual group

- eg, undiagnosed patients, misreporting

r_{gM} = correlation of observed group with manipulation check

- eg, differential response to manipulation

- 3 step procedure
  - Convert *d* and SE_d to point-biserial correlation metric
  - Correct r_{pb} for $\sqrt{r_{yy'}}$ and either r_{gG} or r_{gM}
  - Convert back to *d* metric
  - See paper for details

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

Correcting for measurement error in meta-analysis

- Can correct effect sizes individually
  - Using reliability information from the reports
  - Adjust by effect size and standard error
  - Impute missing reliability from other sources

- Can adjust meta-analysis model parameter post-hoc
  - “Artefact distribution method”
  - Assumes that artefacts and moderators aren’t correlated
  - Correct mean effect size using mean artefacts
  - $\tau_c^2 = \left[\tau^2 - Var_{art}\right]/b_{metric}^2$

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

When should you correct for measurement error?

- When you choose a reliability estimator that captures the important sources of error for a variable

- When measurement model assumptions are reasonable
  - Errors are uncorrelated with each other and true scores
  - When measurement models are complex, more sophisticated corrections could be applied

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

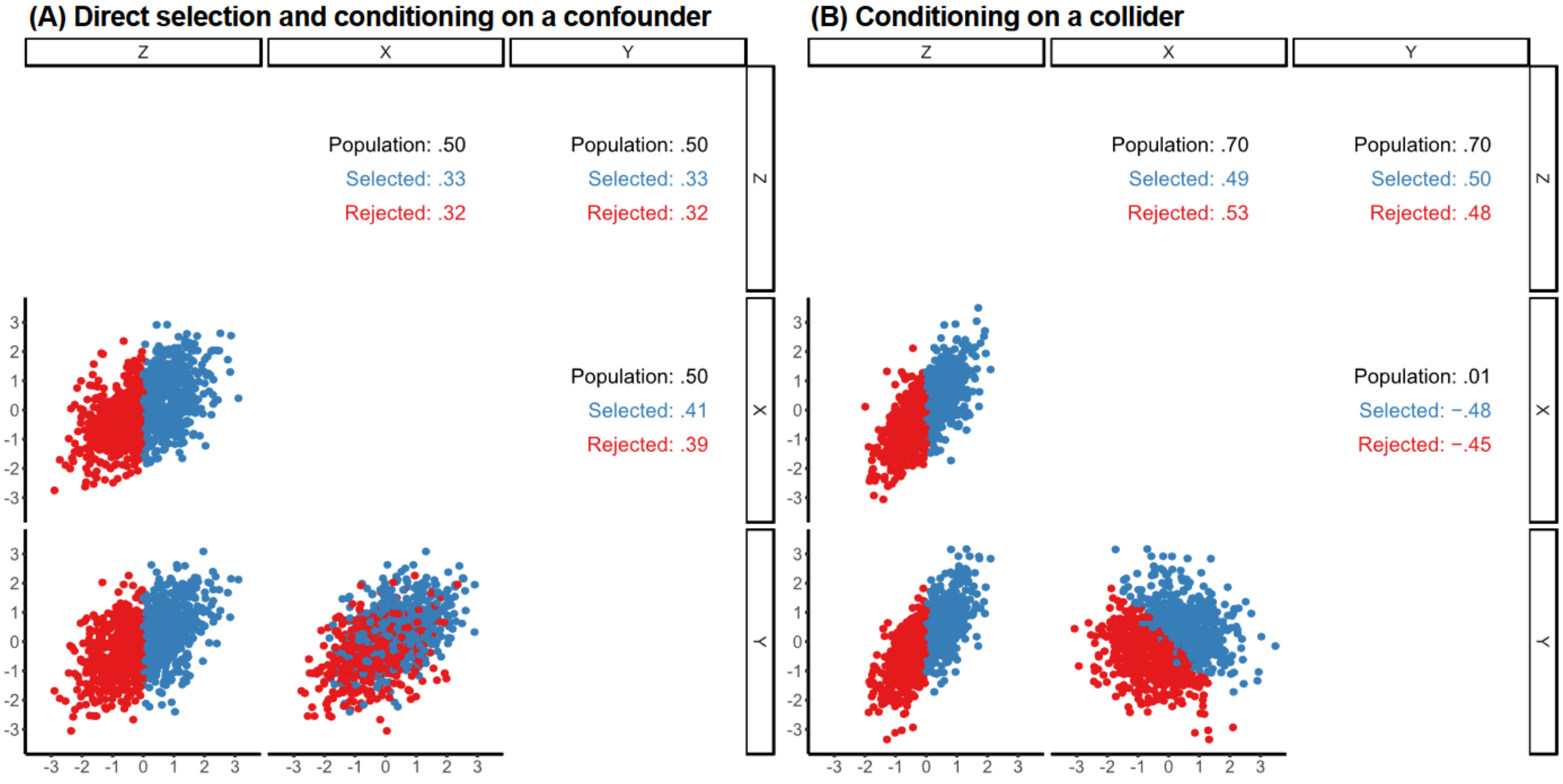
Selection effects

- When the sample where the effect size is computed does not reflect the population to which it should be generalized
  - Due to selection or conditioning on some variable

- e.g., range restriction, selection bias, collider bias

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

Impact of selection effects



In the population, X, Y, and Z are three distinct variables, each intercorrelated $r = .50$. Correlations and points in blue indicate correlations after Z has been selected on to retain only the top 50% of scorers on Z. Selected r_{xz} and r_{yz} are directly range-restricted; selected r_{xy} is indirectly range-restricted.

In the population, X and Y are uncorrelated. Z is a composite of X and Y plus a small amount of error. Correlations and points in blue indicate correlations after Z has been selected on to retain only the top 50% of scorers on Z. Selected r_{xz} and r_{yz} are directly range-restricted; selected r_{xy} is indirectly range-restricted.

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

Selection effects are ubiquitous!

- How much the sampling process condition on important covariates or other variables?

- These factors are present in most studies across many fields
  - Researchers should always consider how results might be affected by biased sampling

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

Selection effects and meta-analysis

- Selection effects can bias all 3 categories of results described above

- Mean effect size
- Heterogeneity and moderators
- Publication bias

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

Quantifying and correcting selection bias

$$u_x = SD_{x_{sample}} / SD_{x_{reference}}$$

- Depending on the nature of the selection mechanism and the specific information available, there are several correction models available

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

Example: Direct range restriction on *X*

$$r_c = \frac{r_{obs}}{u_x\sqrt{1 - u_x^2(1 - r_{xx'})}\sqrt{\left(\frac{1}{u_x^2} - 1\right)r_{obs}^2 + r_{yy'}}}$$

$$SE_{r_c}^2 = SE_{r_{obs}}^2 \times (r_c / r_{obs})^2$$

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

When should you correct for selection effects?

- Should always be considered and discussed

- Common selection effect corrections in *psychmeta* require
  - Linear relationships between predictor and outcome
  - Residual variances are equal in the selected sample and target population

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

Applying statistical corrections in *R*

```
correct_r(
  correction = "bvirr",
  rxyi = .40, n = 150,
  rxx = .80, ryy = .80,
  ux = .90, uy = .80
)
```

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

Applying statistical corrections in *R*

```
correct_d(
  correction = "uvirr_y",
  d = .40, n1 = 75, n2 = 75,
  rGg = .80, ryy = .80, uy = .80
)
```

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

Applying statistical corrections in *R*

```
ma_results_r <- ma_r(ma_method = "ic",
  rxyi = rxyi, n = n,
  rxx = rxxi, ryy = ryyi,
  ux = ux, uy = uy,
  data = data_r_bvirr
)
ma_results_d <- ma_d(ma_method = "ic",
  d = d, n1 = n1, n2 = n2, ryy = ryyi, construct_y = construct, data =
data_d_meas_multi)
```

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

Applying statistical corrections in *R*

```
es_data <- get_metafor(ma_results_d,
  analyses = list(construct_y = "Y"),
  ma_method = "ic",
  correction_type = "ts“
)
```

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

Applying statistical corrections in *R*

```
ma_results_r <- ma_r(ma_method = "ad",
  rxyi = rxyi, n = n,
  rxx = rxxi, ryy = ryyi,
  ux = ux, uy = uy, data = data_r_bvirr
)
```

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0

Conclusions

- Measurement error and selection effects are pervasive
- Detrimental impacts of these artefacts on the validity of research conclusions have been widely documented

- By applying carefully justified artefact corrections, our meta-analyses can better fulfill their research aims

- Meta-analyses should routinely present observed results and results adjusted for plausible measurement and sampling artefacts

Copyright 2021 | Brenton M. Wiernik | CC BY-NC-SA 4.0