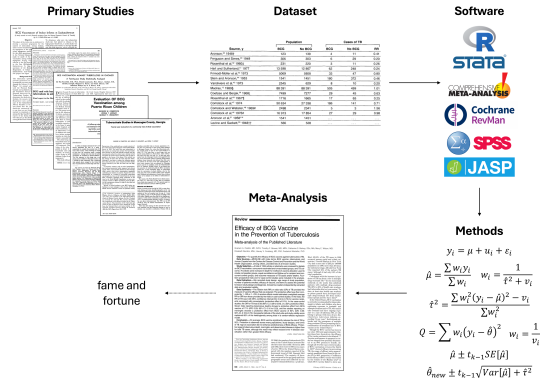
# Meta-Analysis: Structurally Speaking

Colloquia Treverensia
University of Trier
November 25, 2026

Wolfgang Viechtbauer
Maastricht University



## Meta-Analysis in 5 Easy Steps






## Major Advances in Statistical Methods for Meta-Analysis

- standardization of effect size measures (1970s-1980s)
- fixed/equal- and random-effects models (1980s)
- methods for assessing heterogeneity (1980s-2000s)
- meta-regression and moderator analyses (1980s-1990s)
- publication bias and selection/reporting biases (1980s-present)
- improved inference methods (1990s-2010s)
- multivariate and multilevel meta-analysis (1990s-2010s)
- network meta-analysis (2000s-2010s)
- Bayesian meta-analysis (1990s-present)
- robust variance estimation (2010s-present)
- individual participant data (IPD) meta-analysis (1990s-present)
- flexible/nonstandard meta-analytic models (2000s-present)
- living/large-scale meta-analysis and automation (2010s-present)



## But What About the Data?

- more attention: how meta-analytic data should be **analyzed**
- less attention: how the data should be **structured**
- important decisions are made before the statistical analysis



## Why the Structure Matters

- determines the unit of analysis
- identifies samples, groups, outcomes, and time points
- indicates which estimates 'belong together'
- determines which estimates are dependent
- helps to identify redundant or nearly redundant information
- determines which research questions can be addressed
- directly impacts modeling decisions



## Starting With the Basics: A Tidy Dataset

- one header row, with unique and meaningful variable names
- don't do this:

| | treatment | | | control | | |
|---|---|---|---|---|---|---|
| study | mean | sd | n | mean | sd | n |
| Smith, 2014 | 34.2 | 8.9 | 22 | 38.4 | 7.5 | 20 |
| Jones & Miller, 2021 | 5.7 | 2.1 | 48 | 6.2 | 1.8 | 45 |
| Williams et al., 1998 | 74.2 | 16.4 | 78 | 80.9 | 18.3 | 80 |

- do this:

| author | year | trt.mean | trt.sd | trt.n | ctrl.mean | ctrl.sd | ctrl.n |
|---|---|---|---|---|---|---|---|
| Smith | 2014 | 34.2 | 8.9 | 22 | 38.4 | 7.5 | 20 |
| Jones & Miller | 2021 | 5.7 | 2.1 | 48 | 6.2 | 1.8 | 45 |
| Williams et al. | 1998 | 74.2 | 16.4 | 78 | 80.9 | 18.3 | 80 |

## Starting With the Basics: A Tidy Dataset

- one variable per column
- one observation per row
- one value per cell

| study | dose |
|---|---|
| | 50mg |
| | 20-60mg |

→

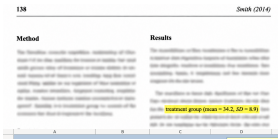
| study | dose | dose.min | dose.max | units |
|---|---|---|---|---|
| | 50 | | | mg |
| | | 20 | 60 | mg |

- use consistent coding of missing values
- avoid formatting that is meaningful only to a human reader
- see also [1, 2] for some useful advice



## Linking Information / Documenting Decisions

- there should be a clear link between the information in the article and what is stored in the dataset ('data provenance')
- comments in the spreadsheet is a simple approach [3]
- highlight the corresponding part in the article




| | author | year | trt.mean | | | | trl.sd | ctrl.n |
|---|---|---|---|---|---|---|---|---|
| 2 | Smith | 2014 | 34.2 | | | | 7.5 | 20 |
| 3 | Jones & Miller | 2021 | 5.7 | 2.1 | 48 | 6.2 | 1.8 | 45 |
| 4 | Williams et al. | 1998 | 74.2 | 16.4 | 78 | 80.9 | 18.3 | 80 |

WV, 11/25/2026 10:51
page 138, right column, second paragraph

- document extraction decisions/choices/steps [4, 5]



## What Does Each Row Represent?

- an important decision when constructing a dataset
- possible observational units:
  - a study or experiment
  - a sample
  - a treatment or comparison group
  - a contrast between groups
  - an outcome at a particular time point for a particular group
- the appropriate unit depends on the information available and the research questions
- different representations of the same information are possible



## Estimate-Level Identifier

- every estimate should have a **unique identifier**
- seems obvious, but this is surprisingly often missing
- not the same as the study identifier when studies can contribute multiple estimates
- the estimate-level identifier:
  - uniquely identifies each observation/row
  - facilitates checking, merging, and restructuring
  - can be used to specify an estimate-level random effect
- other identifiers describe how estimates are related



## Simple Structure and the Random-Effects Model

- the most basic/simple structure:

| study | estimate | variance |
|---|---|---|
| A | 0.54 | 0.0296 |
| B | 0.27 | 0.0174 |
| | | |

- the standard 'random-effects model' for meta-analysis:

$$\theta_i \sim N(\mu, \tau^2) \qquad y_i|\theta_i \sim N(\theta_i, v_i)$$

- heterogeneity in the true effects is modeled with an **estimate-level random effect**
- this is typically done automatically/implicitly by the software



## Multilevel Meta-Analysis

- when some studies provide multiple (independent) estimates, can use a **multilevel random-effects model**
- a common specification includes a random effect for studies, but omits a random effect at the estimate level
- assumes that estimates within a study share the same true effect

| study | estimate | variance |
|---|---|---|
| A | 0.54 | 0.02961 |
| A | 0.43 | 0.02921 |
| B | 0.27 | 0.01740 |
| | | |

→

| study | obs | estimate | variance |
|---|---|---|---|
| A | A1 | 0.54 | 0.02961 |
| A | A2 | 0.43 | 0.02921 |
| B | B1 | 0.27 | 0.01740 |
| | | | |

- adding the estimate-level random effect allows for heterogeneity of effects within studies

## Implicit Nesting

- in the example, the nesting of 'observations within studies' is **implicit** through the coding ('obs' is nested in 'study')

| study | obs | estimate | variance |
|---|---|---|---|
| A | A1 | 0.54 | 0.02961 |
| A | A2 | 0.43 | 0.02921 |
| B | B1 | 0.27 | 0.01740 |
| | | | |

- the two random effects can then be added to the model with:[1]

```
random = list(~ 1 | study, ~ 1 | obs)
```

[1] Code snippets are for the `rma.mv()` function in the `metafor` package for R.



## Explicit Nesting

- can also use a different coding scheme and make the nesting **explicit** via the way the model is specified

| study | obs | estimate | variance |
|---|---|---|---|
| A | 1 | 0.54 | 0.02961 |
| A | 2 | 0.43 | 0.02921 |
| B | 1 | 0.27 | 0.01740 |
| | | | |

- the two random effects can then be added to the model with:

```
random = ~ 1 | study / obs
```

- `random = ~ 1 | study` ignores within-study heterogeneity
- `random = list(~ 1 | study, ~ 1 | obs)` would add crossed random effects which is nonsense here



## Different Types of Identifiers

- identifiers can describe different levels and relationships:
  - author / research group
  - publication
  - study or experiment
  - sample
  - group or treatment arm
  - pairing or comparison unit
  - construct or outcome
  - measurement occasion
  - individual estimate
- these are not necessarily interchangeable
- some relationships are hierarchical; others are crossed
- the identifiers should reflect how the data were generated



## Contrast- vs Arm-Based Structure

- a major decision on how the data are structured
- contrast-based structure:

| | author | year | trt.x | trt.n | ctrl.x | ctrl.n |
|---|---|---|---|---|---|---|
| 2 | Smith | 2014 | 6 | 22 | 9 | 20 |
| 3 | Jones & Miller | 2021 | 12 | 48 | 18 | 45 |
| 4 | Williams et al. | 1998 | 27 | 78 | 31 | 80 |

- arm-based structure:

| | author | year | study | group | x | n |
|---|---|---|---|---|---|---|
| 2 | Smith | 2014 | 1 | trt | 6 | 22 |
| 3 | Smith | 2014 | 1 | ctrl | 9 | 20 |
| 4 | Jones & Miller | 2021 | 2 | trt | 12 | 48 |
| 5 | Jones & Miller | 2021 | 2 | ctrl | 18 | 45 |
| 6 | Williams et al. | 1998 | 3 | trt | 27 | 78 |
| 7 | Williams et al. | 1998 | 3 | ctrl | 31 | 80 |



## Arm-Based Structure

- the 'study' identifier reflects the natural pairing of the data
- need to account for the multilevel nature of the data
- various options:
  - random study effects (multilevel)
  - correlated random effects within studies (multivariate)
  - fixed study effects
- we also add 'group' as a predictor to the model



## Predictors in Contrast- vs Arm-Based Structures

- in the contrast-based structure, predictors can only moderate the size of the difference between the two groups
- in the arm-based structure, predictors can enter as 'main effects' and also interact with the 'group' variable
  - e.g., is the risk higher for older samples vs is the risk difference higher for older samples?
- opens up novel research questions [6–8]

## Dependency Due to Shared Control Groups

- in the contrast-based structure, shared control groups induce dependency in the estimates

| author | year | trt.x | trt.n | ctrl.x | ctrl.n |
|---|---|---|---|---|---|
| Smith | 2014 | 6 | 22 | 9 | 20 |
| Jones & Miller | 2021 | 12 | 48 | 18 | 45 |
| Williams et al. | 1998 | 27 | 78 | 31 | 80 |
| Williams et al. | 1998 | 25 | 83 | 31 | 80 |

- this is avoided in the arm-based structure

| author | year | study | group | x | n |
|---|---|---|---|---|---|
| Smith | 2014 | 1 | trt | 6 | 22 |
| Smith | 2014 | 1 | ctrl | 9 | 20 |
| Jones & Miller | 2021 | 2 | trt | 12 | 48 |
| Jones & Miller | 2021 | 2 | ctrl | 18 | 45 |
| Williams et al. | 1998 | 3 | trt1 | 27 | 78 |
| Williams et al. | 1998 | 3 | trt2 | 25 | 83 |
| Williams et al. | 1998 | 3 | ctrl | 31 | 80 |



## Coding Groups in Arm- vs Contrast-Based Structures

- in the arm-based structure, unique values (within studies) for the 'group' identifier indicate independent groups
- in the contrast-based structure, we need two 'group' identifiers
- how we code 'treatments' is separate from the 'group' identifiers

| paper | grp1 | grp2 | trt1 | trt2 |
|---|---|---|---|---|
| A | 1 | 2 | cbt | ctrl |
| B | 1 | 3 | cbt | ctrl |
| B | 2 | 3 | act | ctrl |
| C | 1 | 2 | cbt | attention |
| C | 1 | 3 | cbt | waitlist |
| D | 1 | 2 | cbt | act |
| | | | | |



## Papers, Studies, and Samples

- a paper may report multiple experiments or independent samples
- multiple papers may report results from the same sample
- the same sample may contribute multiple effect size estimates
- paper, study, and sample identifiers should be distinguished
- the relevant unit of independence is not necessarily the paper

| paper | study | sample | time |
|---|---|---|---|
| A | 1 | 1 | post-trt |
| A | 2 | 2 | post-trt |
| B | 3 | 3 | post-trt |
| C | 1 | 1 | follow-up |
| C | 2 | 2 | follow-up |
| D | 4 | 4 | post-trt |
| D | 4 | 4 | follow-up |
| | | | |



## Between- and Within-Subject Designs

- studies may use a mix of different experimental designs
- can use 'group' and 'time' identifiers for clarification

| paper | grp1 | grp2 | time1 | time2 |
|---|---|---|---|---|
| A | 1 | 2 | 1 | 1 |
| B | 1 | 3 | 1 | 1 |
| B | 2 | 3 | 1 | 1 |
| C | 1 | 1 | 2 | 1 |
| C | 1 | 1 | 3 | 1 |
| | | | | |

- **B:** between-subjects design involving three groups where two estimates were computed contrasting groups 1 and 2 versus group 3 at a single time point
- **C:** within-subjects design where two estimates were computed contrasting time points 2 and 3 versus time point 1 in a single group



## The Importance of the Pairing Identifier

- estimates need to be grouped into meaningful comparison units
- the pairing must be respected when estimating treatment effects
- important when studies contribute multiple comparisons



## Example: The Generation Effect

- McCurdy et al. (2020) [9]: meta-analysis of the 'generation effect' (better memory for self-generated than read information)
- a study might ask participants to recall words that were read vs words self-generated from word fragments
- studies often report multiple experiments and examine various conditions using a mix of within- and between-subject designs
- the dataset included:
  - 126 articles
  - 310 experiments
  - 582 samples
  - 804 pairings
  - 1653 recall estimates (proportion of words/items recalled)
- the pairing identifier was essential for defining the comparisons

## Example: Generation Effect

| Article | Experiment | Sample | Pairing | Condition | Id | Recall |
|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | read | 1 | 0.31 |
| 1 | 1 | 2 | 1 | generate | 2 | 0.45 |
| 2 | 1 | 3 | 2 | read | 3 | 0.45 |
| 2 | 1 | 3 | 2 | generate | 4 | 0.61 |
| 3 | 1 | 4 | 3 | read | 5 | 0.61 |
| 3 | 1 | 4 | 3 | generate | 6 | 0.72 |
| 3 | 1 | 4 | 3 | generate | 7 | 0.95 |
| 4 | 1 | 5 | 4 | read | 8 | 0.28 |
| 4 | 1 | 5 | 4 | generate | 9 | 0.38 |
| 4 | 1 | 6 | 5 | read | 10 | 0.36 |
| 4 | 1 | 6 | 5 | generate | 11 | 0.55 |
| ... | ... | ... | ... | ... | ... | ... |



## Why Pairings Matter

- ignoring pairings would amount to:
  - pooling all read-condition estimates
  - pooling all generate-condition estimates
  - taking the difference between these pooled estimates
- does not recover the average of the within-pairing effects
- pairing identifier defines which estimates form a comparison unit
- model included random effects for articles, experiments, samples, and estimates and a crossed random effect for pairings

```
rma.mv(Recall, Var, data = dat,
       mods = ~ Condition,
       random = list(~ 1 | article/exp/sample/id,
                     ~ 1 | pairing))
```



## Different Sources of Dependence

- two distinct sources of dependence:
  - **sampling errors:** estimates share information from the same participants, groups, or observations
  - **underlying true effects:** effects from the same study may be more similar than effects from different studies
- these are different sources of dependence
- accounting for one does not automatically account for the other



## Sampling Error Dependencies

- **rough rule:** sampling errors are dependent when there is overlap in the observations contributing information to the estimates
- examples:
  - comparisons sharing a control group
  - multiple outcomes measured in the same participants
  - repeated measurements over time
  - overall and subgroup estimates
- accounting for such dependencies may require computing or approximating sampling error covariances



## Multivariate Structures

- multiple estimates may represent different constructs (depression, anxiety), different observations of the same construct (BDI, HAM-D, CES-D), and/or different time points
- construct, observation, and time point must be explicitly identified
- allows construct-specific effects, time-specific effects, and relationships between effects to be examined

| paper | construct | scale | weeks |
|---|---|---|---|
| A | depression | BDI | 12 |
| B | depression | HAM-D | 8 |
| B | anxiety | STAI | 8 |
| C | depression | BDI | 12 |
| C | depression | CES-D | 12 |
| C | depression | BDI | 24 |
| C | depression | CES-D | 24 |
| | | | |



## Do We Always Need a Multivariate Model?

- if the goal is to estimate effects for depression and anxiety:
  - separate meta-analyses may be sufficient
- a multivariate model can offer advantages:
  - borrowing strength across outcomes / time points
  - directly testing differences between outcomes / time points
  - estimating correlations between true effects
- additional complexity should serve a substantive purpose

## (Near) Redundancies

- not all extractable estimates provide distinct information
- some may be exact or nearly exact functions of others
- examples:
  - multiple contrasts (e.g., trt1 - ctrl, trt2 - ctrl, trt1 - trt2)
  - subgroup estimates together with an overall estimate
  - subscale estimates together with an overall estimate
  - alternative adjusted estimates from the same data



## Why (Near) Redundancies Matter

- highly correlated estimates may add little information
- exact redundancies provide no additional information
- possible consequences:
  - singular or nearly singular covariance matrices
  - numerical instability
  - difficulties with model estimation
  - misleading precision when dependencies are ignored
- careful extraction and structuring can help identify these problems before modeling



## Keep the Original Information

- retain the information from which effect sizes were computed
- examples:
  - group means, standard deviations, and sample sizes
  - event counts and group sizes
  - correlations and sample sizes
  - reported adjusted estimates and their adjustment sets
- this facilitates:
  - checking effect size calculations
  - computing alternative measures
  - computing sampling-error covariances



## Harmonizing Different Sources of Information

- studies may report information in different forms:
  - means and standard deviations
  - change scores
  - event counts
  - odds ratios or regression coefficients
- obtaining comparable effect sizes may require transformations and assumptions
- record which information was originally reported
- document how the information was transformed



## Practical Recommendations

1. use a tidy, machine-readable dataset with unique variable names
2. link the extracted information to their sources
3. assign a unique identifier to every estimate
4. distinguish publications, studies, experiments, and samples
5. identify groups, pairings, outcomes, and time points
6. retain extracted information and document transformations
7. check for overlapping information and (near) redundancies
8. establish clear rules for which estimates to extract and retain
9. choose the representation that serves the research questions
10. consider dependencies before choosing the statistical model



## Conclusions

- data structuring is not merely a preliminary step before the 'real' statistical analysis
- identifiers encode relationships that are essential for meaningful comparisons
- the same information can sometimes be represented in different ways (with advantages/disadvantages)
- not every available estimate needs to be included
- some statistical complications can be avoided through thoughtful structuring and selection
- **the model should reflect the data structure and the data structure should reflect the research questions**

## References [1]


1. Wickham, H. (2014). Tidy data. *Journal of Statistical Software*, *59*(10), 1–23. doi:10.18637/jss.v059.i10
2. Broman, K. W., & Woo, K. H. (2018). Data organization in spreadsheets. *The American Statistician*, *72*(1), 2–10. doi:10.1080/00031305.2017.1375989
3. Shokraneh, F., & Adams, C. E. (2017). Increasing value and reducing waste in data extraction for systematic reviews: Tracking data in data extraction forms. *Systematic Reviews*, *6*, 153. doi:10.1186/s13643-017-0546-z
4. Lakens, D., Hilgard, J., & Staaks, J. (2016). On the reproducibility of meta-analyses: Six practical recommendations. *BMC Psychology*, *4*(1), 24. doi:10.1186/s40359-016-0126-3
5. Ivimey-Cook, E. R., Noble, D. W. A., Nakagawa, S., Lajeunesse, M. J., & Pick, J. L. (2023). Advice for improving the reproducibility of data extraction in meta-analysis. *Research Synthesis Methods*, *14*(6), 911–915. doi:10.1002/jrsm.1663
6. Bruin, M. de, Viechtbauer, W., Schaalma, H. P., Kok, G., Abraham, C., & Hospers, H. J. (2010). Standard care impact on effects of highly active antiretroviral therapy adherence intervention: A meta-analysis of randomized controlled trials. *Archives of Internal Medicine*, *170*(3), 240–250. doi:10.1001/archinternmed.2009.536




## References [2]


7. Black, N., Eisma, M. C., Viechtbauer, W., Johnston, M., West, R., Hartmann-Boyce, J., … Bruin, M. de. (2020). Variability and effectiveness of comparator group interventions in smoking cessation trials: A systematic review and meta-analysis. *Addiction*, *115*(9), 1607–1617. doi:10.1111/add.14969
8. Kraiss, J., Viechtbauer, W., Black, N., Johnston, M., Hartmann-Boyce, J., Eisma, M., … Bruin, M. de. (2023). Estimating the true effectiveness of smoking cessation interventions under variable comparator conditions: A systematic review and meta-regression. *Addiction*, *118*(10), 1835–1850. doi:10.1111/add.16222
9. McCurdy, M. P., Viechtbauer, W., Sklenar, A. M., Frankenstein, A. N., & Leshikar, E. D. (2020). Theories of the generation effect and the impact of generation constraint: A meta-analytic review. *Psychonomic Bulletin & Review*, *27*(6), 1139–1165. doi:10.3758/s13423-020-01762-3




# Thank You for Your Attention!

Questions, Comments, Suggestions?


wolfgang.viechtbauer@maastrichtuniversity.nl

https://www.wvbauer.com

https://www.metafor-project.org

@wviechtb @wviechtb