Multivariate Meta-analysis in the Multiverse 🚀

Chiara Montuori, Filippo Gambarota, Gianmarco Altoè and Barbara Arfè

University of Padova @Psicostat

04/03/2022

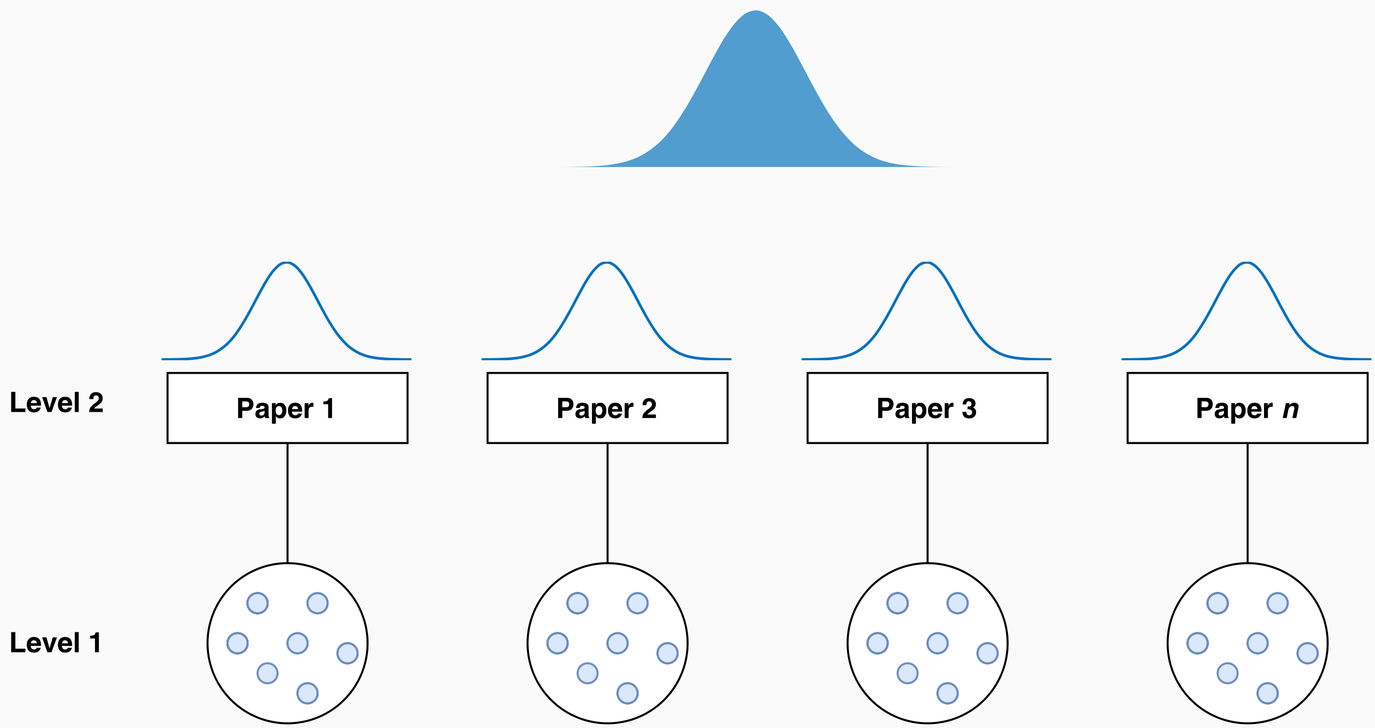
Outline

- Meta-analysis in 2 minutes 😱🕐
- The main problems
- Our modelling choice
- Our multiverse approach 🚀
- Reference and useful links

Meta-analysis in 2 minutes 😱 🕐

1. Changing the statistical unit

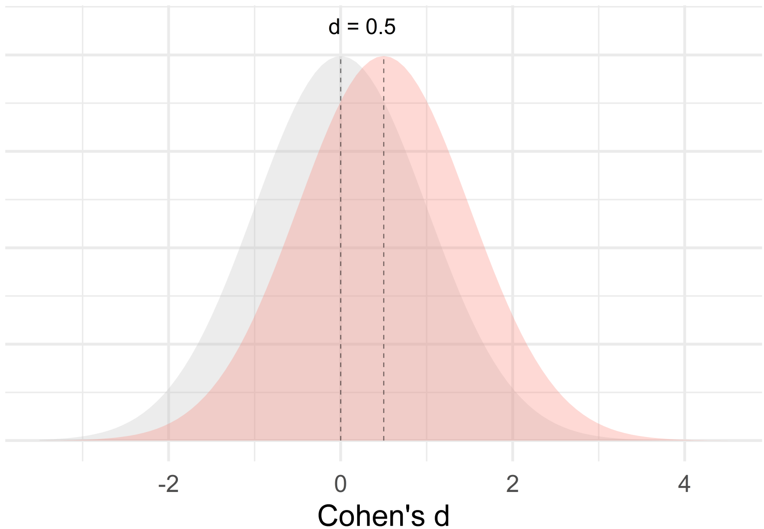
When we do a meta-analysis we are **switching the statistical unit** from e.g. participants to studies with multiple participants



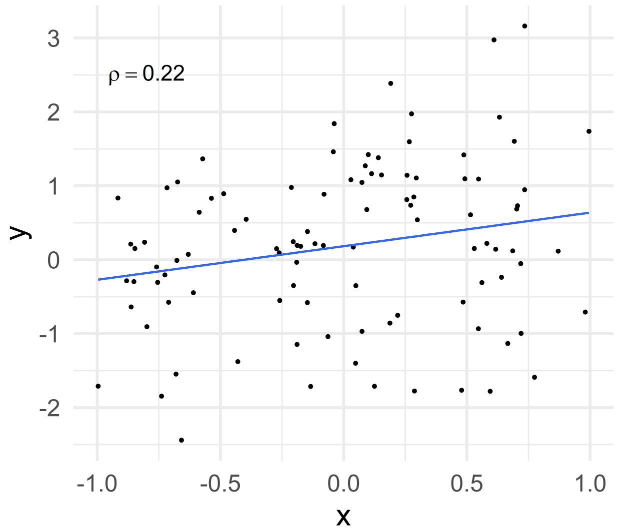
2. Summarizing with Effect Sizes

Usually (but not always) we use a standardized effect size measure (e.g., Cohen's *d* or Pearson Correlation) in order to compare studies with different designs, dependent measure (e.g., Accuracy and Reaction Times)

Cohen's *d*

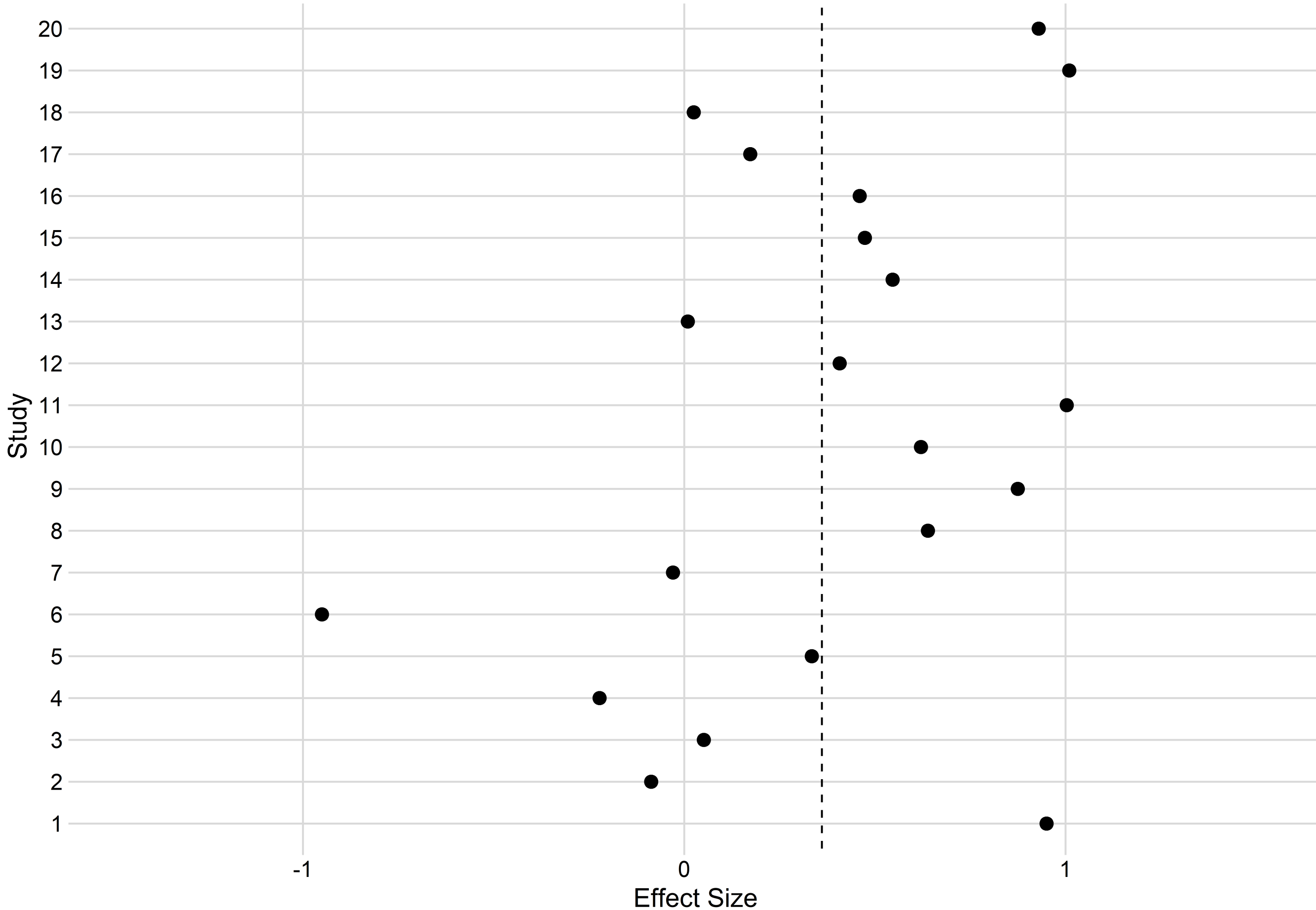


Correlation



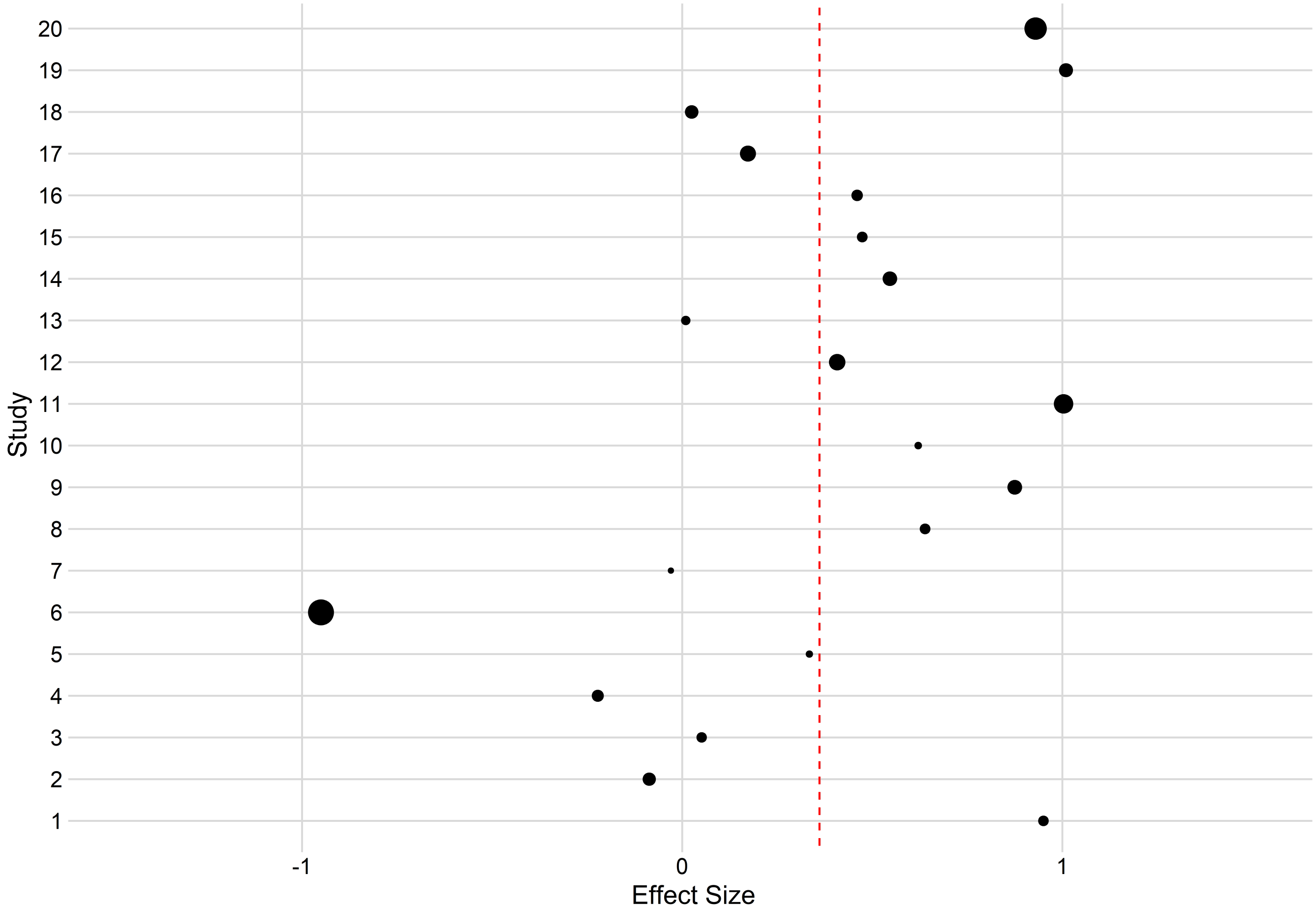
3. Weighting by precision

In order do a meta-analysis we need to pool together multiple studies taking into account that some studies should have more weight (e.g., higher sample size). In the simplest form, a meta-analysis is essentially a weighted average.



3. Weighting by precision

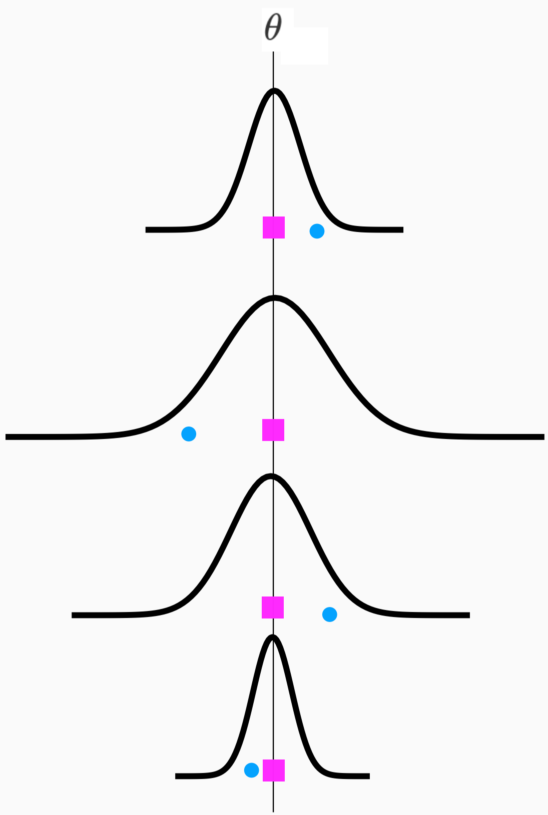
In order do a meta-analysis we need to pool together multiple studies taking into account that some studies should have more weight (e.g., higher sample size). In the simplest form, a meta-analysis is essentially a weighted average.



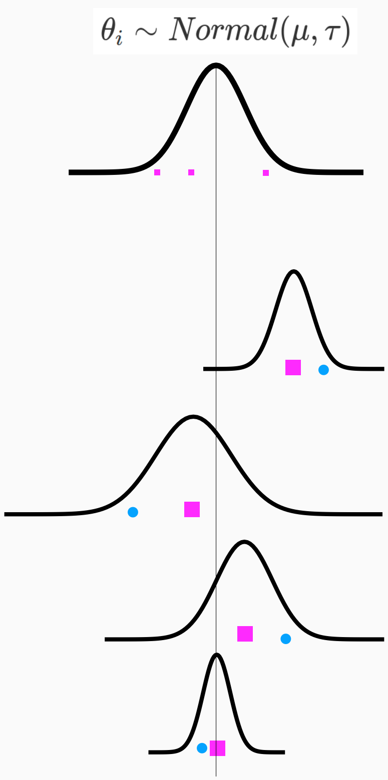
4. Fixed effect vs Random effect

This is an essential (*and often misunderstood*) step:

The **fixed-effect** model assume a single **population-level** effect/parameter to be estimated θ_{fixed}. Observed variability between effects is due to **sampling error** only.



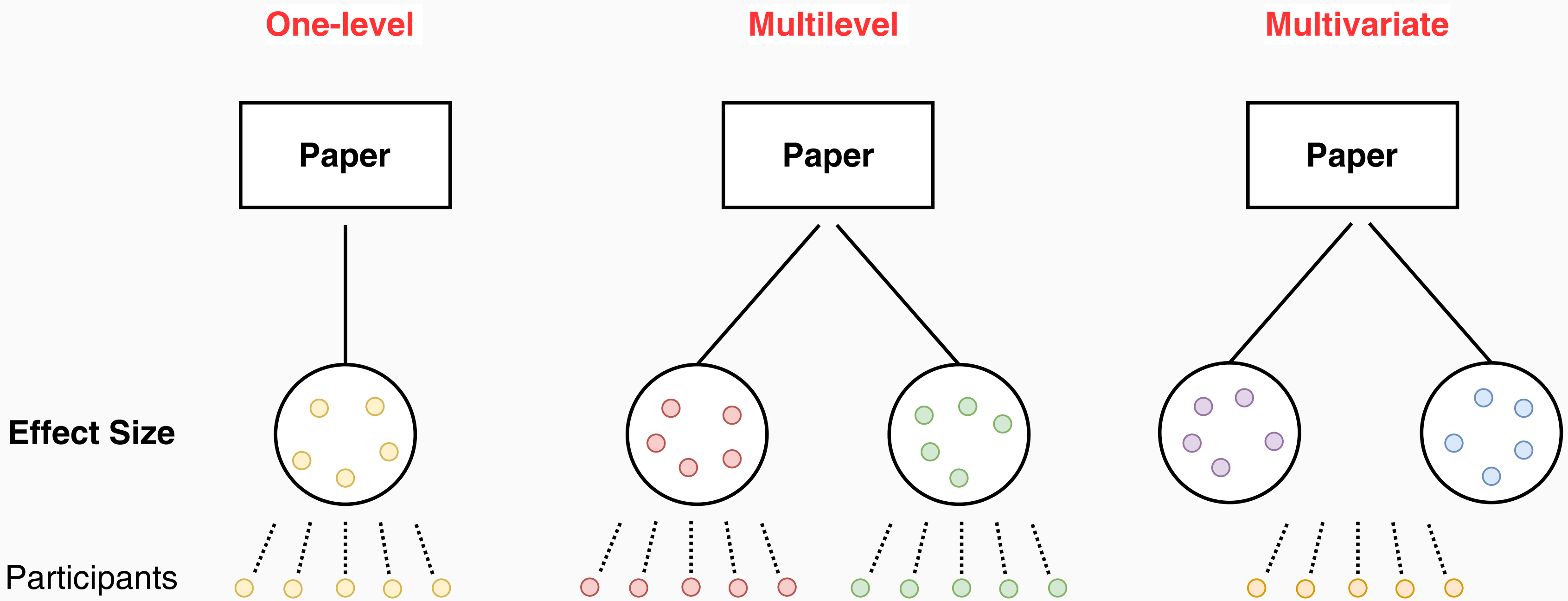
The **random-effect** model assume a distribution of **population-level** effects where the **true effect can vary**. We need to estimate the mean θ_{random} and the variance τ^2



5. Complex data structure

In some situations we need to take into account multilevel and/or multivariate situations:

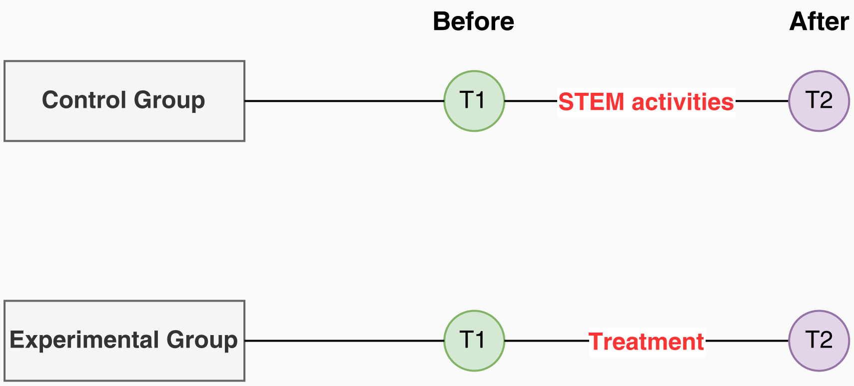
- multiple studies within the same paper (multilevel structure)
- multiple effects (dependent variables) measured on the same pool of participants (e.g., Accuracy and Reaction Times)



The present work

Coding and Executive Functions

The impact of **coding training** on children (~5-10 age) executive functions (**outcomes**). We selected only **randomized-control trials**.



First problem: **Effect size**

For PPC designs one of the mostly used effect size is the $dpcc$ by Morris (2008). In particular the $dpcc_2$:

$$d_{pcc_2} = c_p \frac{(M_{T,post} - M_{T,pre}) - (M_{C,post} - M_{C,pre})}{SD_{pooled,pre}}$$

With sampling variance:

$$\sigma^2(d_{pcc_2}) = c_p^2(1-\rho)(\frac{n_t + n_c}{n_t n_c})(\frac{n_t + n_c - 2}{n_t + n_c - 4})(\frac{1+\Delta^2}{2(1-\rho)(\frac{n_t+n_c}{n_t n_c})})$$

The critical component is the ρ i.e. the **pre-post** correlation that is often **not reported**!

Second problem: **Multiple Effect Sizes**

When measuring a certain cognitive function (e.g., **working memory**) different authors could use different measures. We decided to recode the **raw** test measure $y_1, y_2, \dots y_n$ into the **latent** psychological variable y_i. This create a situation where we have multiple y_i on the same paper.

Borenstein et al. (2009) and also the `metafor` package with the `metafor::aggregate.escalc()` function implemented a way to combine multiple dependent effect sizes:

Aggregate Multiple Effect Sizes or Outcomes Within Studies

Description

The function can be used to aggregate multiple effect sizes or outcomes belonging to the same study (or to the same level of some other clustering variable) into a single combined effect size or outcome.

Usage

```
## S3 method for class 'escalc'
aggregate(x, cluster, time, obs, V, struct="CS", rho, phi,
          weighted=TRUE, checkpd=TRUE, fun, na.rm=TRUE, subset, select, digits, ...)
```

Arguments

`x`	an object of class `"escalc"`.
`cluster`	vector to specify the clustering variable (e.g., study).
`time`	optional vector to specify the time points (only relevant when `struct="CAR"`, `"CS+CAR"`, or `"CS*CAR"`).
`obs`	optional vector to distinguish different observed effect sizes or outcomes measured at the same time point (only relevant when `struct="CS*CAR"`).
`V`	optional argument to specify the variance-covariance matrix of the sampling errors. If not specified, argument `struct` is used to specify the variance-covariance structure.
`struct`	character string to specify the variance-covariance structure of the sampling errors within the same cluster (either `"ID"`, `"CS"`, `"CAR"`, `"CS+CAR"`, or `"CS*CAR"`). See 'Details'.
`rho`	value of the correlation of the sampling errors within clusters (when `struct="CS"`, `"CS+CAR"`, or `"CS*CAR"`). Can also be a vector with the value of the correlation for each cluster.

Computing a combined effect across outcomes

Our notation will be to use Y_1, Y_2 etc. for effect sizes from different outcomes or time points within a study, and Y_j to refer to the j^{th} of these. Strictly, we should use Y_{ij}, for the j^{th} outcome (or time-point) in the i^{th} study. However, we drop the i subscript for convenience. The effect size for *Basic skills* is computed as the mean of the reading and math scores,

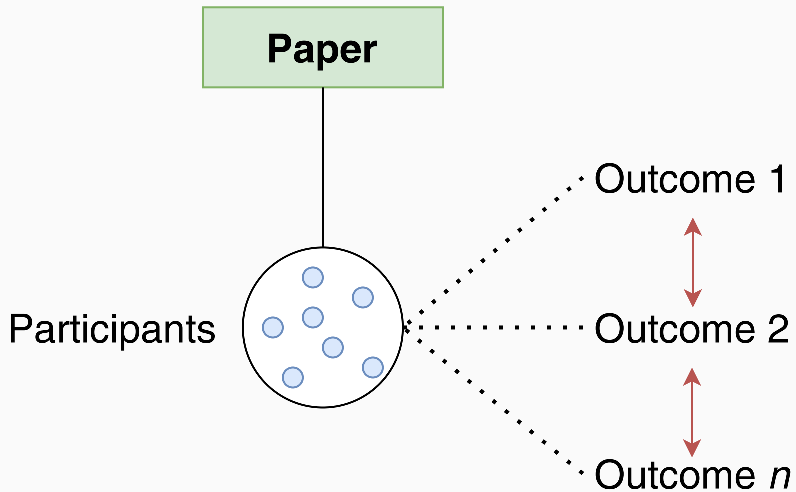
$$\overline{Y} = \frac{1}{2}(Y_1 + Y_2). \tag{24.1}$$

This is what we would use as the effect estimate from this study in a meta-analysis. Using formulas described in Box 24.1, the variance of this mean is

$$V_{\overline{Y}} = \frac{1}{4}\left(V_{Y_1} + V_{Y_2} + 2r\sqrt{V_{Y_1}}\sqrt{V_{Y_2}}\right) \tag{24.2}$$

Third problem: **Multiple Outcomes**

This is the classical **multivariate situation** where we need to take into account the correlation between different measures on the same pool of participants:



We need this matrix for each study, creating a **huge** variance-covariance matrix. But most importantly we need the **covariance between effects**!

Fourth problem: **Limited amount of studies**

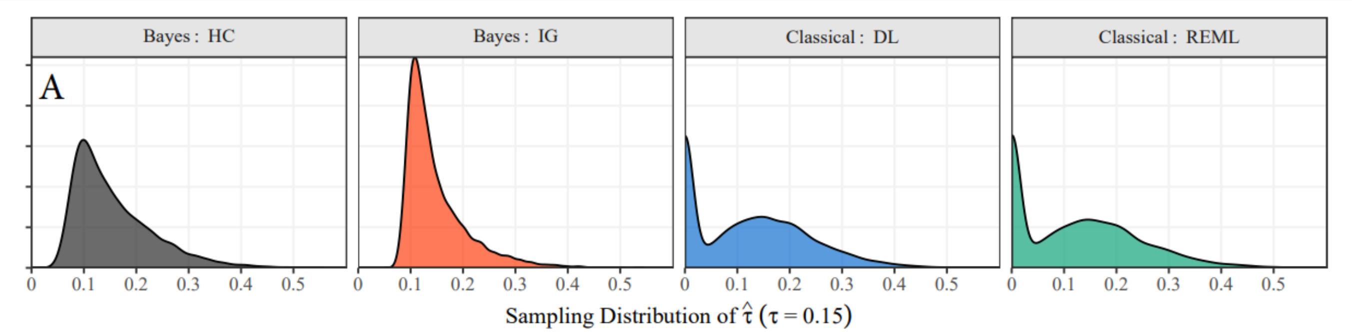
Often, for new area of research or not really widespread research topics the amount of available studies is limited. In particular according to our **strict** inclusion criteria we found **9 papers** with several effects within each paper:

Outcome	**n**
Cognitive Flexibility Acc.	2
Inhibition Acc.	5
Planning Acc.	3
Problem Solving	7
Working Memory Acc.	2

Why is a problem?

Depending on the model we need to estimate **one or several parameters**:

- Williams et al. (2018) clearly demonstraed the biased estimation of τ with a limited amount of studies impacting also the estimation of μ especially using the classical DerSimonian and Laird (1986) or REML estimators.
- With a multivariate model we estimate several μ and, in case of the random-effect model, several τ



Simulated sampling distribution of Tau from Williams et al. (2018)

Our solution? ...a Multiverse approach! 🚀

Why multiverse?

Perspectives on Psychological Science
2016, Vol. 11(5) 702–712

DOI: 10.1177/1745691616658637
pps.sagepub.com
SAGE

Increasing Transparency Through a Multiverse Analysis

Sara Steegen[1], Francis Tuerlinckx[1], Andrew Gelman[2], and Wolf Vanpaemel[1]
[1]KU Leuven, University of Leuven and [2]Columbia University

Review Article

Which Data to Meta-Analyze, and How?

A Specification-Curve and Multiverse-Analysis Approach to Meta-Analysis

Martin Voracek, Michael Kossmeier, and Ulrich S. Tran

Why multiverse?

> We suggest that instead of performing only one analysis, researchers could perform a multiverse analysis [...] A multiverse analysis offers an idea of **how much the conclusions change because of arbitrary choices in data construction** and gives pointers as to **which choices are most consequential** in the fragility of the result.

Our choice…Fixed-effect multivariate model!

Following the notation from Mavridis and Salanti (2013):

$$\begin{pmatrix} y_{i1} \\ \vdots \\ y_{ij} \end{pmatrix} \sim MVN\left(\begin{pmatrix} \mu_{i1} \\ \vdots \\ \mu_{ij} \end{pmatrix}, \begin{pmatrix} \sigma^2_{i1} & \dots & \sigma_{i1,ij} \\ \vdots & \ddots & \vdots \\ \sigma_{i1,ij} & \dots & \sigma^2_{ij} \end{pmatrix} \right)$$

Where each study y_i can have multiple outcomes j and come from a multivariate normal distribution with means the vector of effects and the variance-covariance matrix.

- Estimating an effect size for each outcome (as series of univariate analysis)
- No τ estimation (compared to the random-effect model)
- Takes into account the multivariate data structure (compared to univariate or multilevel analysis)
- More appropriate with a limited amount of studies (see Cai & Fan, 2020)

But our Multiverse...

- Fixed-effect or random-effect Model?
- Multivariate or Univariate?
- Which correlations to use?
  - A $\rho_{pre-post}$ of 0.5, 0.7 and 0.9
  - A ρ_{agg} of 0.3, 0.5, 0.7
  - A ρ_{multi} of 0.3, 0.5 and 0.7

We have a total of 108 meta-analysis to compute!

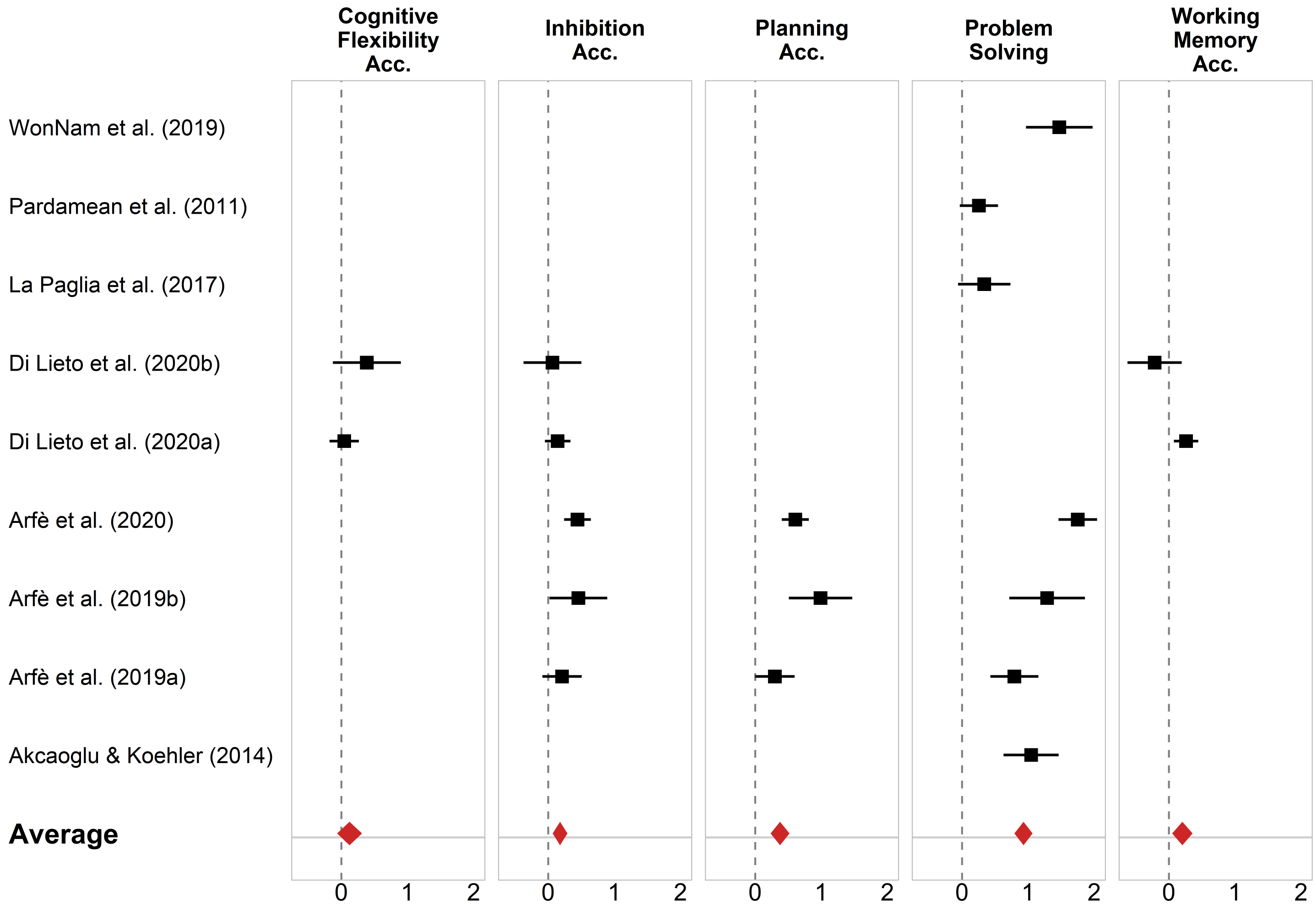
The main results...

The main results...

Outcome	β	SE	95% CI	z	*p*
Cognitive Flexibility Acc.	0.123	0.096	[-0.065, 0.311]	1.281	0.2
Inhibition Acc.	0.177	0.057	[0.065, 0.289]	3.098	0.002
Planning Acc.	0.377	0.073	[0.234, 0.519]	5.187	< 0.001
Problem Solving	0.929	0.070	[0.792, 1.066]	13.308	< 0.001
Working Memory Acc.	0.204	0.079	[0.049, 0.358]	2.583	0.01

Omnibus Test χ_5= 181.9 p < 0.001

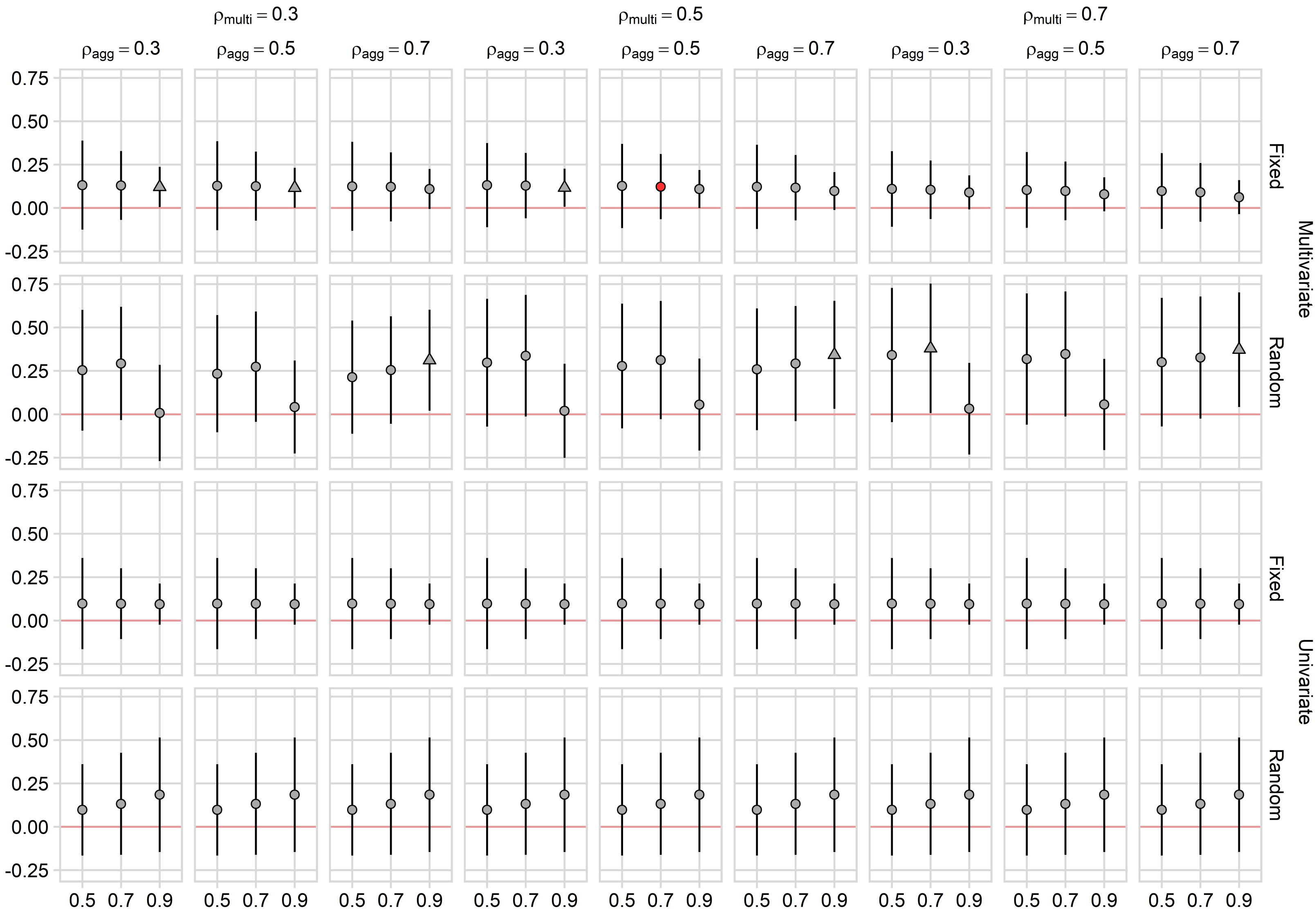
$\varrho_{\text{pre-post}}$ = 0.7, ϱ_{agg} = 0.5, ϱ_{multi} = 0.5



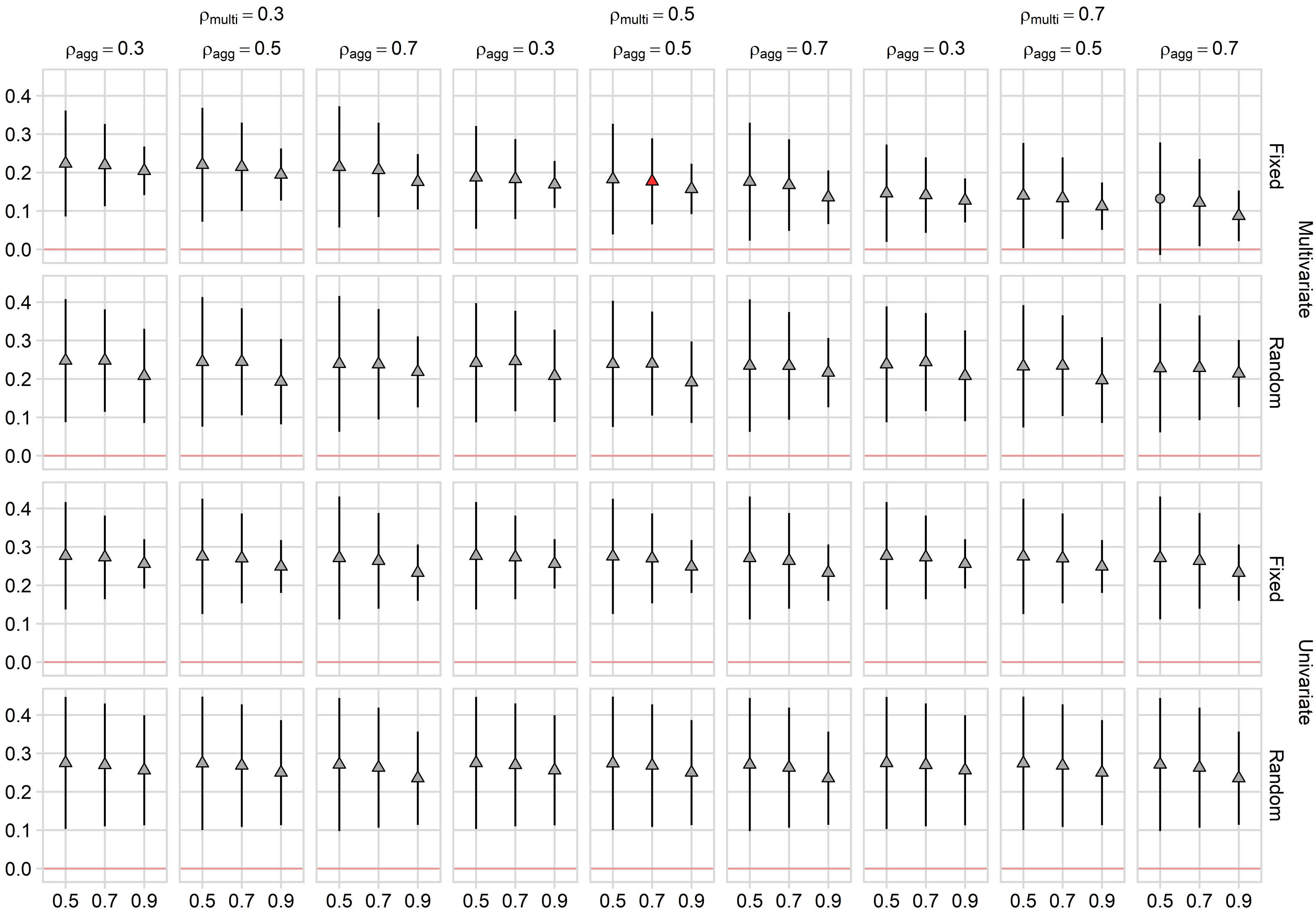
Cognitive Flexibility Acc.
Inhibition Acc.
Planning Acc.
Problem Solving
Working Memory Acc.
WonNam et al. (2019)
Pardamean et al. (2011)
La Paglia et al. (2017)
Di Lieto et al. (2020b)
Di Lieto et al. (2020a)
Arfè et al. (2020)
Arfè et al. (2019b)
Arfè et al. (2019a)
Akcaoglu & Koehler (2014)
Average
0 1 2

Our multiverse results!

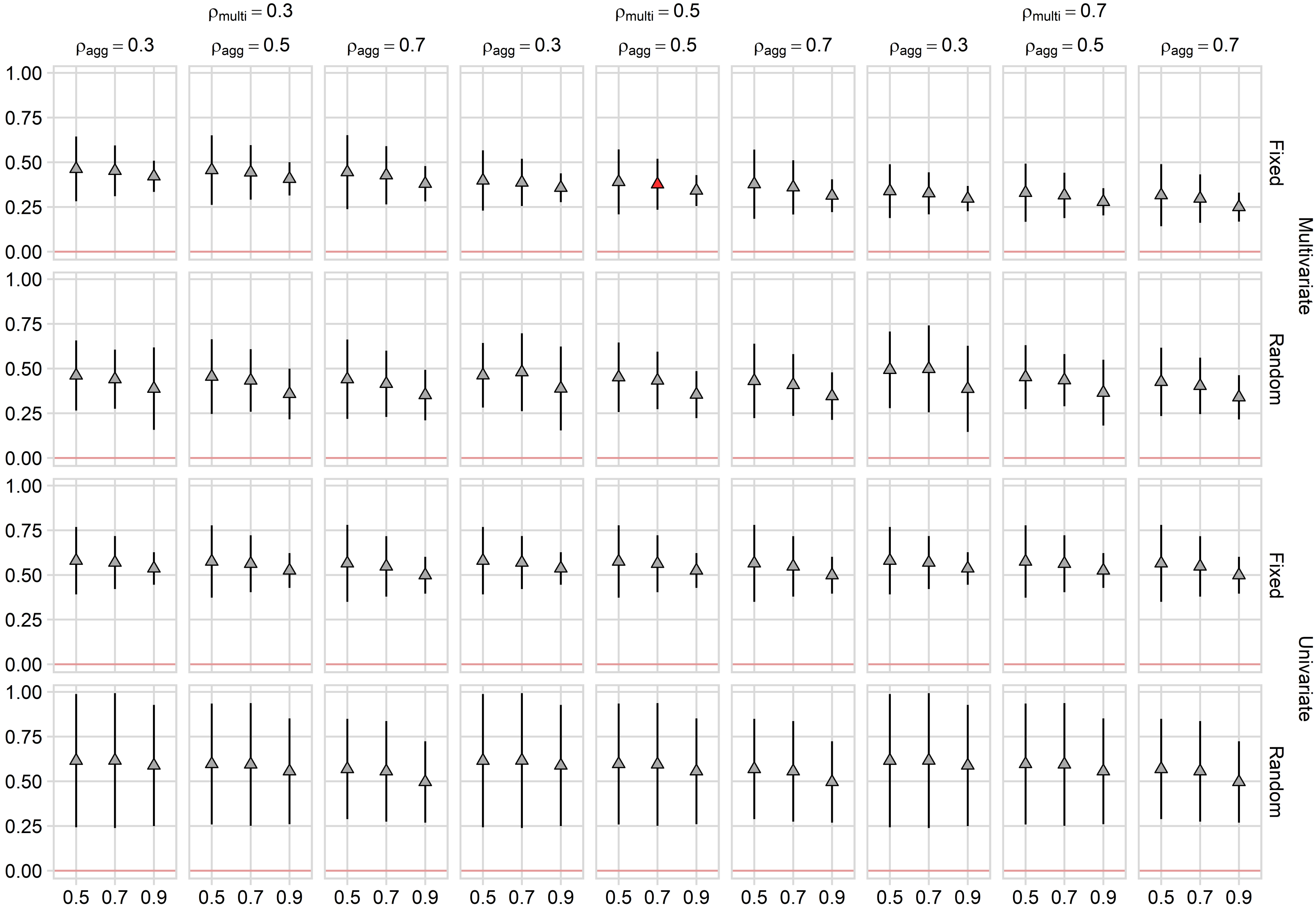
Cognitive Flexibility   Inhibition   Planning   Problem Solving   Working Memory



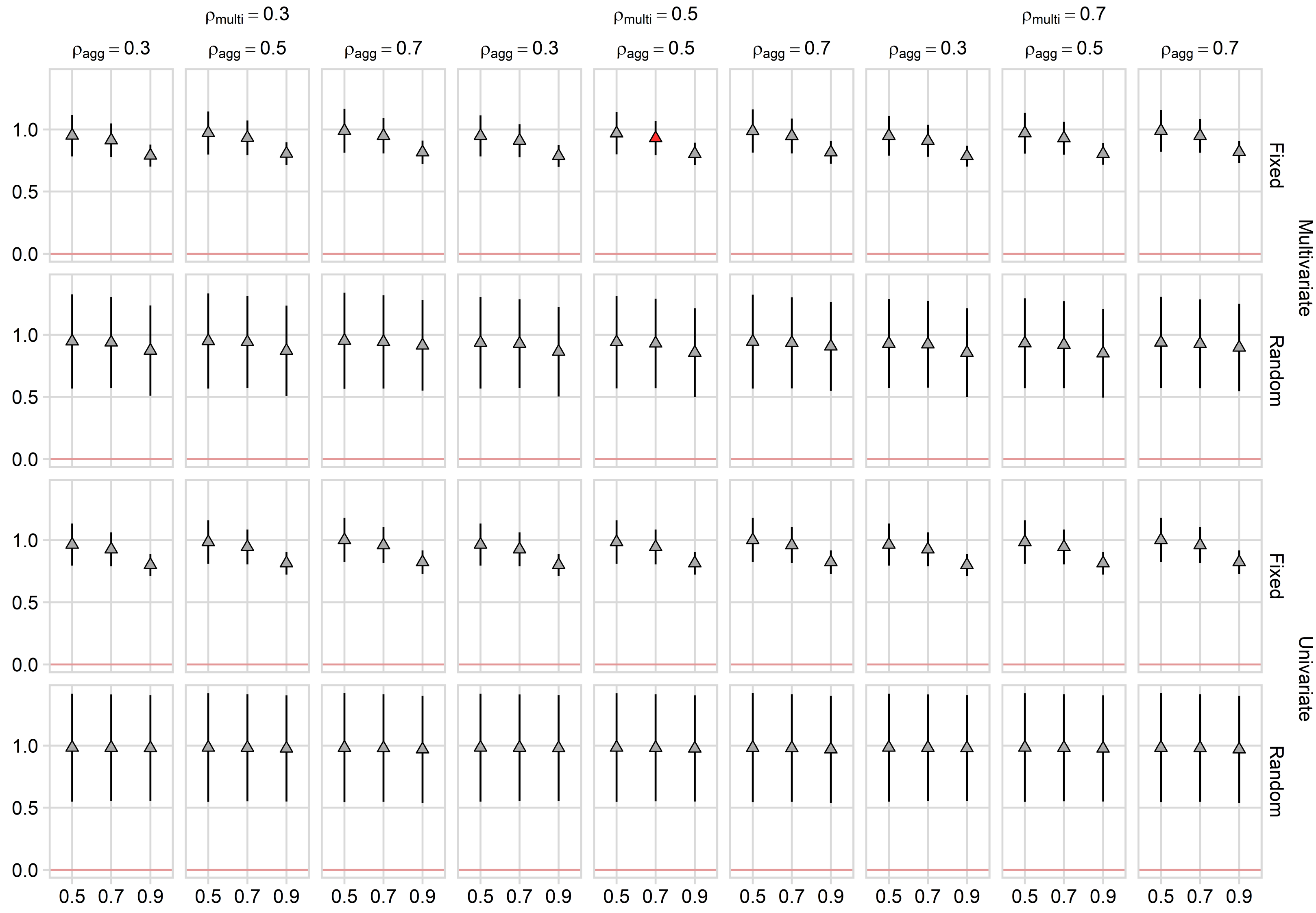
Cognitive Flexibility | Inhibition | Planning | Problem Solving | Working Memory



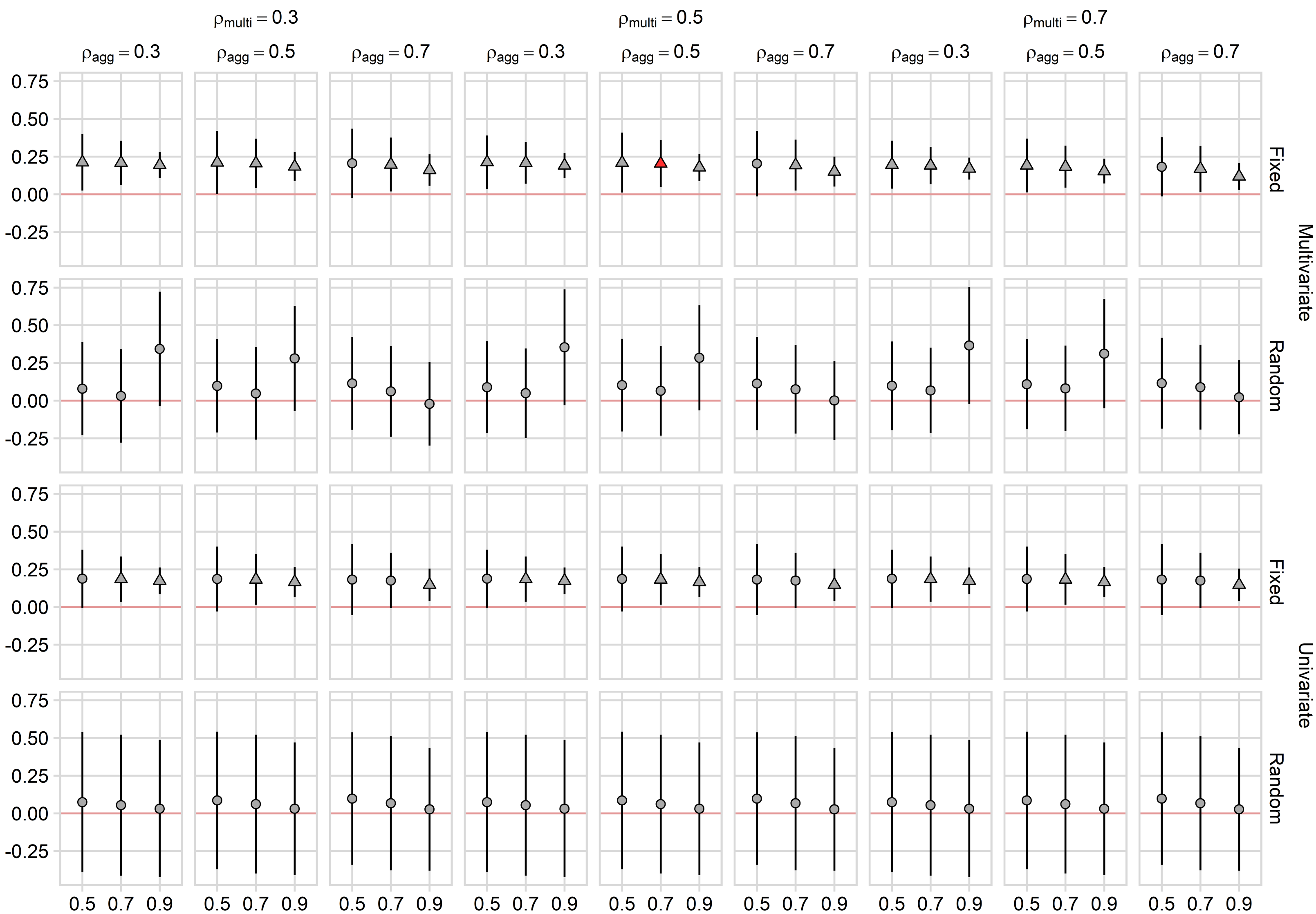
Cognitive Flexibility Inhibition Planning Problem Solving Working Memory



Cognitive Flexibility | Inhibition | Planning | **Problem Solving** | Working Memory



Cognitive Flexibility Inhibition Planning Problem Solving Working Memory



Take Home Message

Take Home Message

Data analysis is not **easy** and **cannot be oversimplified**. You have to take into account complex data structures and statistical dependence

You are **always** making a specific choice from **multiverse** of possibilities in terms of statistical models, values to impute, etc.

Doing one analysis is **FINE**. Doing Multiple analyses is **FUN** (and **useful** 😉)

filippo.gambarota@phd.unipd.it

@fgambarota

filippogambarota

Download PDF slides

References

Borenstein, M., L. V. Hedges, J. P. T. Higgins, et al. (2009). *Introduction to Meta-Analysis*. DOI: 10.1002/9780470743386.

Cai, Z. and X. Fan (2020). "A Comparison of Fixed-Effects and Random-Effects Models for Multivariate Meta-Analysis Using an SEM Approach". En. In: *Multivariate Behav. Res.* 55.6, pp. 839-854. ISSN: 0027-3171, 1532-7906. DOI: 10.1080/00273171.2019.1689348.

Mavridis, D. and G. Salanti (2013). "A practical introduction to multivariate meta-analysis". En. In: *Stat. Methods Med. Res.* 22.2, pp. 133-158. ISSN: 0962-2802, 1477-0334. DOI: 10.1177/0962280211432219.

Morris, S. B. (2008). "Estimating Effect Sizes From Pretest-Posttest-Control Group Designs". In: *Organizational Research Methods* 11.2, pp. 364-386. ISSN: 1094-4281. DOI: 10.1177/1094428106291059.

Steegen, S., F. Tuerlinckx, A. Gelman, et al. (2016). "Increasing Transparency Through a Multiverse Analysis". En. In: *Perspect. Psychol. Sci.* 11.5, pp. 702-712. ISSN: 1745-6916, 1745-6924. DOI: 10.1177/1745691616658637.

Viechtbauer, W. (2010). *Conducting meta-analyses in R with the metafor package*. DOI: 10.18637/jss.v036.i03.

Voracek, M., M. Kossmeier, and U. S. Tran (2019). "Which Data to Meta-Analyze, and How?" In: *Zeitschrift für Psychologie* 227.1, pp. 64-82. ISSN: 2190-8370. DOI: 10.1027/2151-2604/a000357.

Williams, D. R., P. Rast, and P. -. C. Bürkner (2018). "Bayesian meta-analysis with weakly informative prior distributions". DOI: 10.31234/osf.io/7tbrm.

Useful links

- Doing Meta-Analysis with R: A Hands-On Guide: Amazing resource
- Meta-analysis mailing list: A lot of Q&A
- Metafor: Not only the most important package for meta-analysis in R but also a collection of tutorial and practical solutions.
- Handbook of Meta-Analysis - 2020: The most complete and recent book on meta-analysis