Inference on multiverse meta-analysis

A multivariate permutation testing approach

Filippo Gambarota[1] Anna Vesely[2] Livio Finos[3]
Gianmarco Altoè[1]

[1]Department of Developmental Psychology and Socialization
University of Padova

[2]Department of Statistical Sciences
University of Bologna

[3]Department of Statistical Sciences
University of Padova

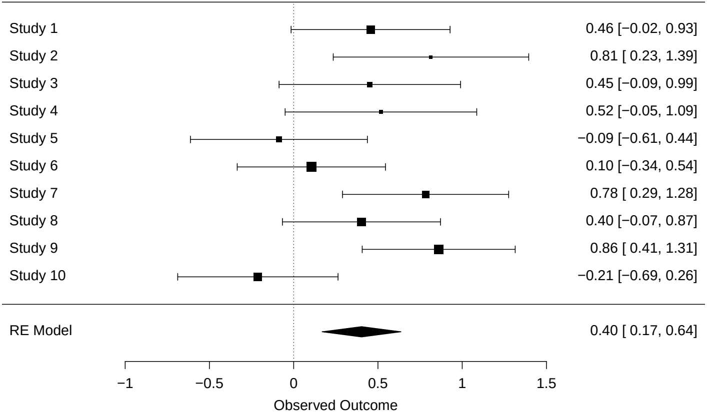
@META-REP
Munich, 2024

Contents

Meta-analysis

Meta-analysis in a nutshell

Meta-analysis is useful to combine information from multiple studies using an appropriate statistical model.



Meta-analysis model

We can define a (random-effects) meta-analysis model as:

$$y_i = \mu_\theta + \delta_i + \epsilon_i$$

$$\delta_i \sim \mathcal{N}(0, \tau^2)$$

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2_{\epsilon_i})$$

Where μ_θ is the average true effect, δ_i is the random-effect of the study i ($\theta_i = \mu_\theta + \delta_i$) and ϵ_i is the sampling error of the study i. When $\tau^2 = 0$ we have an equal-effects (or fixed-effect) model.

Inference on meta-analysis

Standard inference in meta-analysis can be done using a Wald test (Wald, 1943).

$$Z^* = \frac{\mu^*}{\sqrt{\sigma^2_{\mu^*}}}$$

$$\mu^* = \frac{\sum_{i=1}^k w_i^* y_i}{\sum_{i=1}^k w_i^*}$$

$$\sigma^2_{\mu^*} = \frac{1}{\sum_{i=1}^k w_i^*}$$

$$w_i^* = \frac{1}{\sigma^2_{\epsilon_i} + \tau^2}$$

Meta-analysis with permutations (Follmann & Proschan, 1999)

With k observed studies where y_i and $\sigma^2_{\epsilon_i}$ being the observed effect sizes and sampling variances:

1. Generate a random vector $\mathbf{s}$ of ± 1 of length k
2. Multiply the $\mathbf{y}$ vector with the $\mathbf{s}$ vector
3. Fit the meta-analysis model and calculate z^*_j (j for permuted)
4. Repeat 1-3 for a large number of times B. With small k we can do all the permutations $B = 2^k \times k$

The first permutation ($j = 1$) is the observed data. The p value can be computed as:

$$p = \frac{\#(|z^*_j| > |z^*_1|)}{B}$$

Is meta-analysis the perfect solution to everthing?

Is meta-analysis the perfect solution ?

- **garbage in, garbage out**: the quality of the meta-analysis results depends on the quality of input studies
- **uncontrolled heterogeneity**: the strength and clarity of meta-analysis results depends on the selection of studies and the research question
- **degrees of freedom**: conducting a meta-analysis requires making a lot of arbitrary choices

Meta-analysis, researcher degrees of freedom

Despite useful and very powerful, meta-analysis is characterized by several (arbitrary) choices. For example:

- Should the study x be excluded for theoretical or statistical (e.g., outliers) reasons?
- Should we use an equal or random-effects model?
- Which value should take the pre-post missing correlation?
- ...

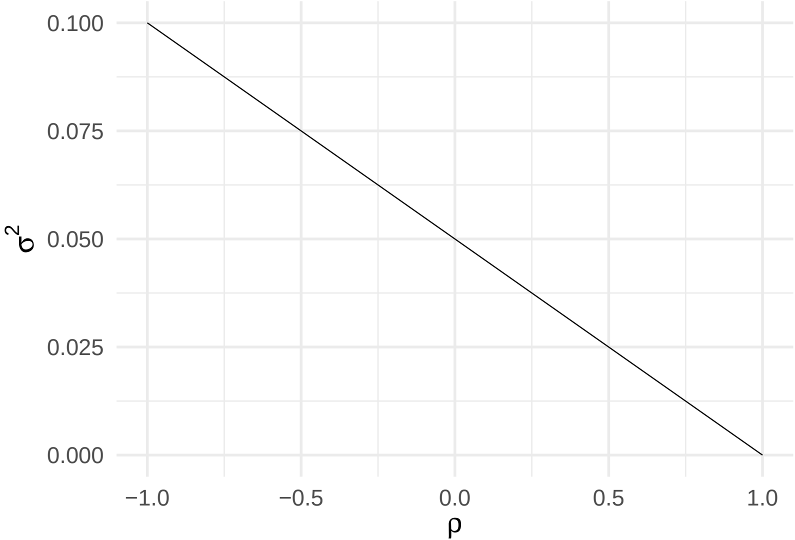
An example: Pre-post Cohen's d

With a pre-post Cohen's d we need the pre-post correlation ρ to calculate the sampling variance:

$$\sigma^2_{\epsilon_{pp}} = \frac{2(1-\rho)}{n} + \frac{d^2}{2n}$$

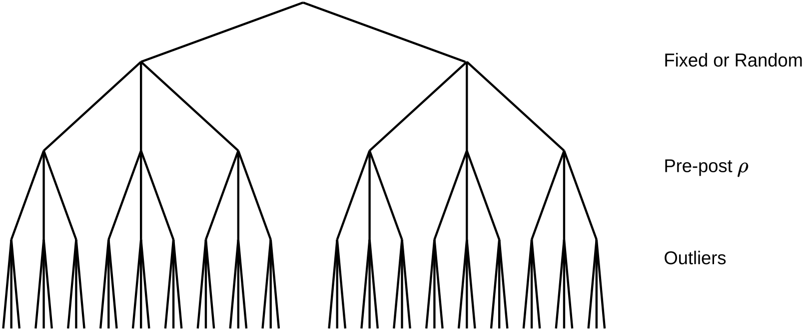
ρ is usually non reported and need to be chosen from previous literature or a plausible guess.

Pre-post Cohen's d



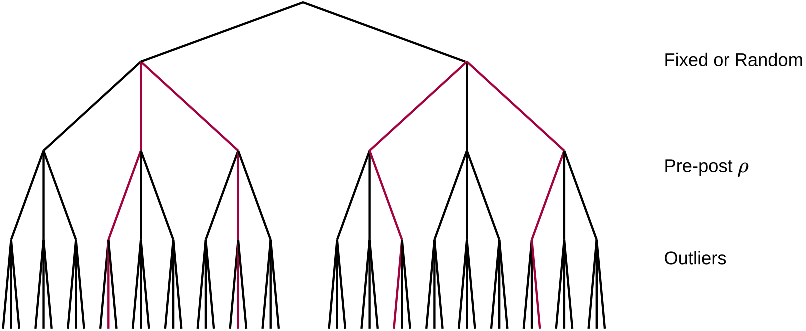
The garden of forking paths

With multiple choices, there is a tree of possibilities.



The garden of forking paths

With multiple choices, there is a tree of possibilities.



Only some of them produce a significant results and just one of them is usually reported in the final analysis and paper.

Multiverse Analysis

Multiverse (Steegen et al., 2016)

- Real-world data analysis involve **several choices at each step**
- There are **many plausible alternatives** to the chosen analysis
- The **impact of alternatives** is often neglected or strongly underrated

> The proposal!
>
> **Thus let's report all the plausible analysis with a given dataset!**

Inference on multiverse

- The increase in complexity after taking into account scenarios (hundreds or even thousands) is usually handled only descriptively
- The specification curve (Simonsohn et al., 2020) is the only inferential method but is not implemented for meta-analysis and do not provide and appropriate p-value adjustment

The problem...

There is a lack of a general and valid inferential framework for multiverse analysis

PIMA (Girardi et al., 2024)

PSYCHOMETRIKA—VOL. 89, NO. 2, 542–568
JUNE 2024
https://doi.org/10.1007/s11336-024-09973-6

POST-SELECTION INFERENCE IN MULTIVERSE ANALYSIS (PIMA): AN INFERENTIAL FRAMEWORK BASED ON THE SIGN FLIPPING SCORE TEST

PAOLO GIRARDI

CA' FOSCARI UNIVERSITY OF VENICE

ANNA VESELY

UNIVERSITY OF BOLOGNA

DANIËL LAKENS

EINDHOVEN UNIVERSITY OF TECHNOLOGY

GIANMARCO ALTOÈ AND MASSIMILIANO PASTORE

UNIVERSITY OF PADOVA

ANTONIO CALCAGNÌ

UNIVERSITY OF PADOVA

GNCS-INDAM RESEARCH GROUP

LIVIO FINOS

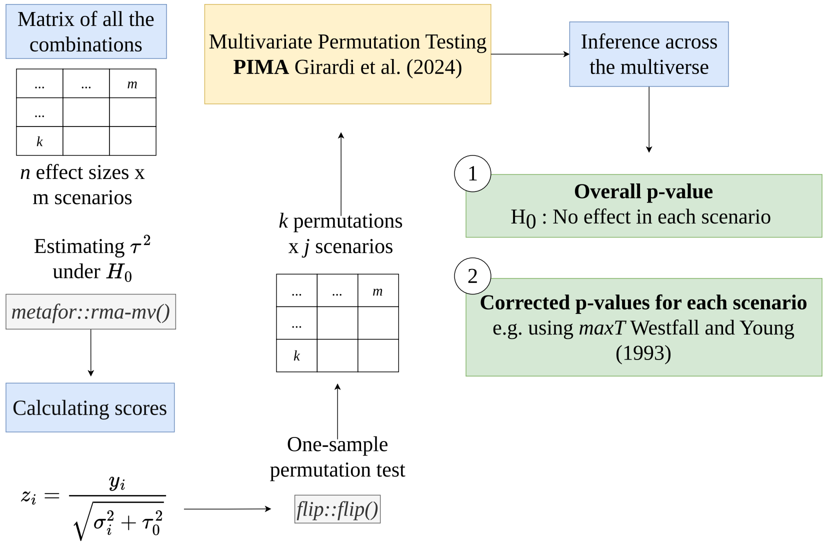
UNIVERSITY OF PADOVA

PIMA

- Use a multivariate extension of the sign-flip score test (Hemerik et al., 2020)
- Works on generalized linear models (and meta-analysis)
- Controls the family-wise error rate
- Provides an overall multiverse p-value and corrected p-values for each included scenario

Multiverse meta-analysis

General Workflow



Fast meta-analysis using permutations

- Meta-analysis using permutations requries recomputing τ^2 and μ_θ after each permutation.
- We proposed to estimate τ^2 under H_0 and use the value for the permutations (without re-estimating it)
- This is extremely fast especially for large datasets and several multiverse scenarios

Estimating τ^2 under H_0

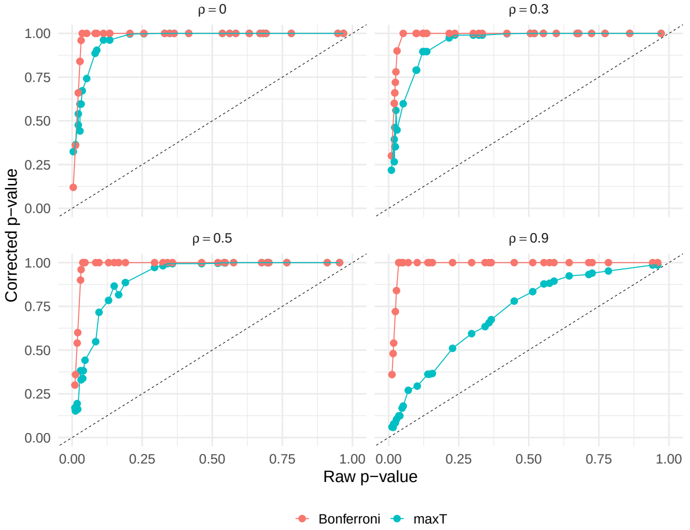
The crucial step is the point (1). This requires maximizing the log-likelihood fixing $\mu_\theta = 0$:

$$L(\mu_\theta, \tau^2|\mathbf{y}) = -\frac{1}{2}\sum_{i=1}^{k}\ln(\tau^2 + \sigma^2_{\epsilon_i}) - \frac{1}{2}\sum_{i=1}^{k}\frac{(y_i - \mu_\theta)^2}{\tau^2 + \sigma^2_{\epsilon_i}}$$

This can be done in R using some optimizer function (e.g., `optim`) or using directly the `metafor` package that allows fixing some parameters that are usually estimated.

The main advantage of PIMA

The power (i.e., the impact of the correction) increases as the correlation (likely to be high in a multiverse analysis) increase.



A simulated example

Simulation approach

The data structure: an outcome (e.g., depression) measured with multiple scales (e.g., different questionnaires) within each paper in a pre-post design:

```
#>    study outcome  ni    yi   vi
#> 1      1       1  25  0.65 0.09
#> 2      1       2  25  0.42 0.08
#> 3      1       3  25 -0.71 0.09
#> 4      2       1  88 -0.07 0.02
#> 5      2       2  88  0.15 0.02
#> 6    ...     ... ...   ...  ...
#> 7      9       3  72 -0.03 0.03
#> 8      9       4  72 -0.82 0.03
#> 9      9       5  72  0.13 0.03
#> 10    10       1 164  0.24 0.01
#> 11    10       2 164  0.58 0.01
```

`yi` is the pre-post effect size, `vi` is the sampling variance and `ni` the sample size.

Simulation approach

Let's make an example for a paper with $j = 3$ measures of the outcome:

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \end{bmatrix} = \begin{bmatrix} \mu_{\theta_1} \\ \mu_{\theta_2} \\ \mu_{\theta_3} \end{bmatrix} + \begin{bmatrix} \delta_{\theta_1} \\ \delta_{\theta_2} \\ \delta_{\theta_3} \end{bmatrix} + \begin{bmatrix} \epsilon_{\theta_{11}} \\ \epsilon_{\theta_{12}} \\ \epsilon_{\theta_{13}} \end{bmatrix}$$

$$\delta \sim \text{MVN} \begin{pmatrix} 0 & \tau_1^2 & & \\ 0 & \rho_{21}\tau_2\tau_1 & \tau_1^2 & \\ 0 & \rho_{31}\tau_2\tau_1 & \rho_{32}\tau_2\tau_1 & \tau_1^2 \end{pmatrix}$$

$$\epsilon \sim \text{MVN} \begin{pmatrix} 0 & \sigma_{\epsilon_1}^2 & & \\ 0 & \rho_{21}\sigma_{\epsilon_2}^2\sigma_{\epsilon_1}^2 & \sigma_{\epsilon_2}^2 & \\ 0 & \rho_{31}\sigma_{\epsilon_3}^2\sigma_{\epsilon_1}^2 & \rho_{32}\sigma_{\epsilon_3}^2\sigma_{\epsilon_2}^2 & \sigma_{\epsilon_3}^2 \end{pmatrix}$$

Simulation approach

We simulated individual participant data, thus:

1. Sampling the true values θ_{ij} for each study i and outcome j from the multivariate distribution
2. Generating n_i pre and post data with correlation ρ
3. Calculating the effect size (imputing the pre-post correlation)
4. Aggregating multiple outcomes within the same paper (imputing the correlation)
5. Fitting the meta-analyis model
6. Calculating the scores
7. Repeating 3-4 for each scenario
8. Using PIMA

Simulation approach

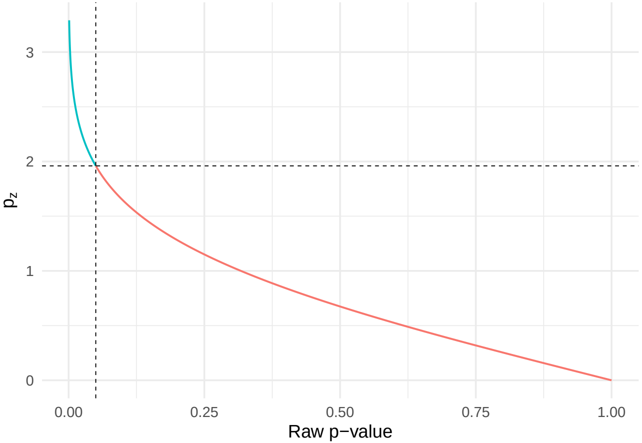
We simulated a relatively simple but plausible multiverse with:

- 4 pre-post correlations
- 4 correlations between multiple measures of the same outcome
- 2 meta-analysis models (fixed and random-effects)

For a total of 32 multiverse scenarios.

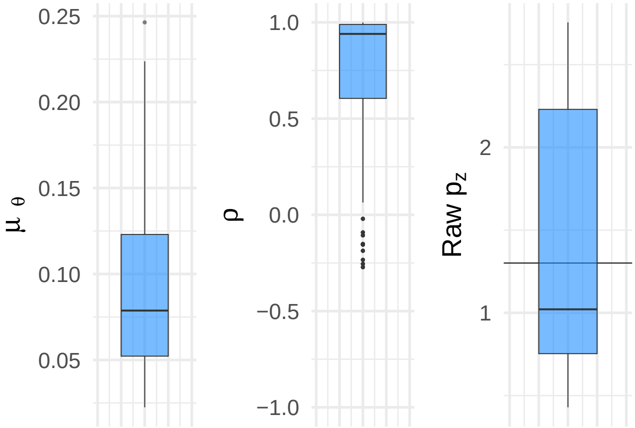
P-value transformation

For the sake of intepretability, we used a transformation of the p-value into pseudo Z scores as $p_z = \Phi^{-1}(1 - \frac{p}{2})$



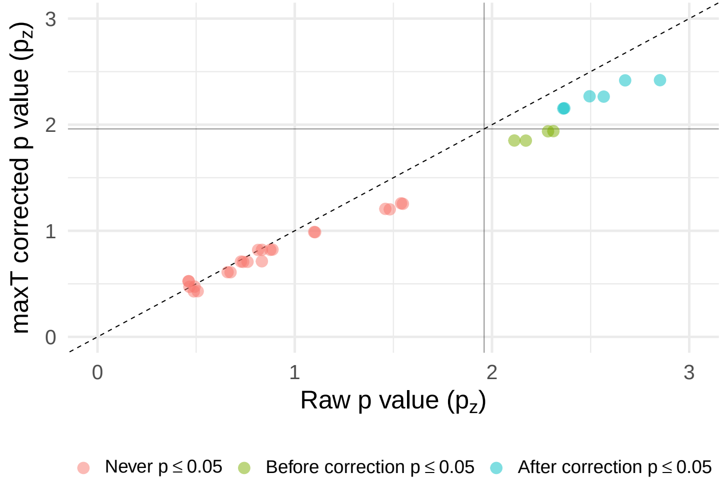
Multiverse results

The multiverse is associated with an overall p value of 0.016 [1].
Then we can describe the overall results:



[1] combined using the maxT method by Westfall & Stanley Young (1993)

Impact of multiplicity correction

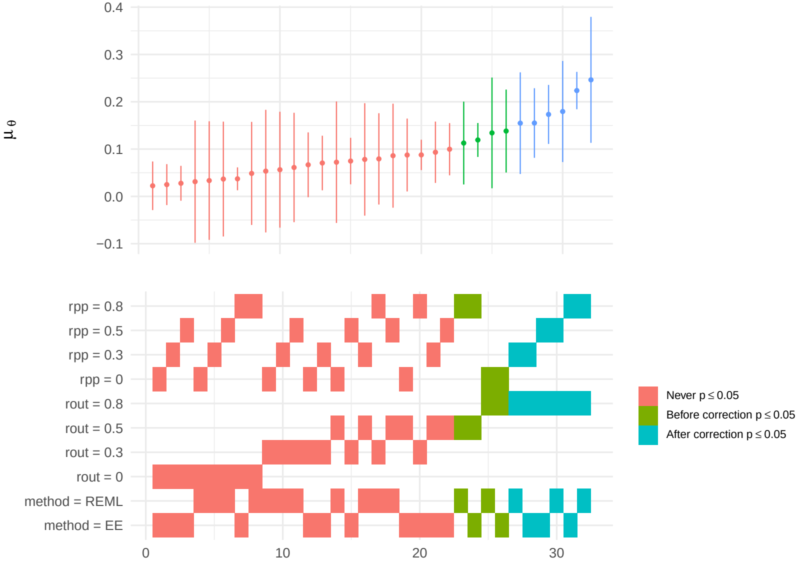


(valid) Post-hoc selective inference

Legal P-Hacking :)

After the overall test and p values correction, the survived scenarios (the blue dots) can be selectively commented, without inflating the type-1 error.

~ Specification Curve (Simonsohn et al., 2020)



Guidelines

Guidelines for multiverse meta-analysis

1. Multiverse meta-analysis must contain only **PLAUSIBLE** models. Including implausible models (e.g., assuming a pre-post correlation of 0.95) reduces the statistical power.

Guidelines for multiverse meta-analysis

1. Multiverse meta-analysis must contain only **PLAUSIBLE** models. Including implausible models (e.g., assuming a pre-post correlation of 0.95) reduces the statistical power.
2. As with any other inferential test, multiverse analysis should be **PLANNED** otherwise no control of type-1 error.

Guidelines for multiverse meta-analysis

1. Multiverse meta-analysis must contain only **PLAUSIBLE** models. Including implausible models (e.g., assuming a pre-post correlation of 0.95) reduces the statistical power.
2. As with any other inferential test, multiverse analysis should be **PLANNED** otherwise no control of type-1 error.
3. Like in standard meta-analysis, the quality of the conclusions is related to the input data and the choice of multiverse scenarios.

Future steps

Future steps

- extending to multilevel and multivariate meta-analysis (the permutation approach is not straightforward)
- create an R package for multiverse meta-analyses with ad-hoc functions to analyze, report, and visualize the results
- create a data simulation framework for simulating a plausible multiverse for power and design analysis

References

Follmann, D. A., & Proschan, M. A. (1999). Valid inference in random effects meta-analysis. *Biometrics*, *55*, 732–737. https://doi.org/10.1111/j.0006-341x.1999.00732.x

Girardi, P., Vesely, A., Lakens, D., Altoè, G., Pastore, M., Calcagnì, A., & Finos, L. (2024). Post-selection inference in multiverse analysis (PIMA): An inferential framework based on the sign flipping score test. *Psychometrika*. https://doi.org/10.1007/s11336-024-09973-6

Hemerik, J., Goeman, J. J., & Finos, L. (2020). Robust testing in generalized linear models by sign flipping score contributions. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, *82*, 841–864. https://doi.org/10.1111/rssb.12369

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour*, *4*, 1208–1214. https://doi.org/10.1038/s41562-020-0912-z

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *11*, 702–712. https://doi.org/10.1177/1745691616658637

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, *54*, 426–482. https://doi.org/10.1090/s0002-9947-1943-0012401-3

Westfall, P. H., & Stanley Young, S. (1993). *Resampling-based multiple testing: Examples and methods for p-value adjustment*. John Wiley & Sons. https://play.google.com/store/books/details?id=nuQXORVGI1QC

filippo.gambarota@unipd.it

filippogambarota.github.io