Choosing a practical and valid Image-Based Meta-Analysis

Camille Maumet[1] and Thomas E. Nichols[2]

1. University of Rennes, Inria, CNRS, Inserm, IRISA, Rennes, France.
2. Oxford Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Population Health, University of Oxford, UK.

Introduction

While most **neuroimaging meta-analyses** are based on peak coordinate data, the best practice method is an **Image-Based Meta-Analysis (IBMA)** [1].

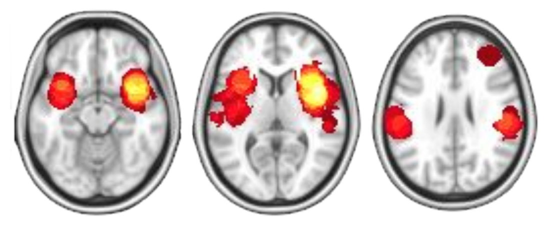
A **number of IBMA approaches have been proposed** combining:

- standardised statistics (Z's),
- just effect estimates (E's) or
- both effect estimates and their standard errors (E+SE's).

While using E+SE's and estimating between-study variance should be optimal, **the methods are not guaranteed to work for small number of studies**. Also, **often only standardised estimates are shared**, reducing the possible meta-analytic approaches. Finally, because the **BOLD signal is non-quantitative** care has to be taken in order to insure that E's are expressed in the same units [2,3].

Given the growing interest in data sharing in the neuroimaging community there is a need to **identify what is the minimal data to be shared in order to allow for future IBMAs**.

Coordinate-based meta-analysis (CBMA)

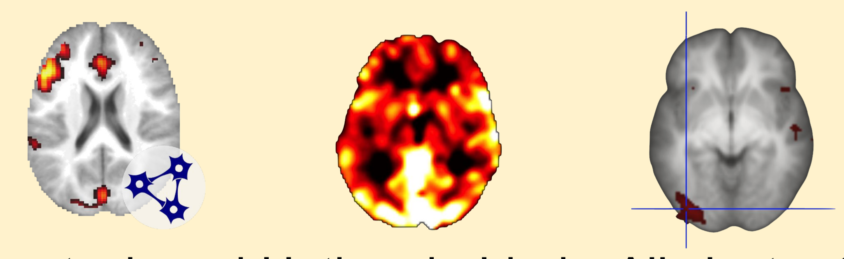


Input: paper

Thresholded statistics

Selected peak locations

Image-based meta-analysis (IBMA)



Input: shared data

Unthresholded maps

All cluster & peak locations

	Meta-analysic statistic	Nominal H_0 distrib.	Inputs	Assumptions
MFX	$\left(\sum \kappa_i \hat{\beta}_i\right) / \sqrt{\sum_{i=1}^k \kappa_i}$ with $\kappa_i = 1/(\hat{\tau}^2 + s_i^2)$	$\mathcal{T}_{k-1}$	$\hat{\beta}_i, s_i^2$	IGE, large sample.
RFX	$\left(\sum_{i=1}^k \frac{\hat{\beta}_i}{\sqrt{k}}\right) / \widehat{\sigma_C^2}$	$\mathcal{T}_{k-1}$	$\hat{\beta}_i$	IGE; $\tau^2 + \sigma_i^2 = \sigma_C^2\ \forall i$
Perm. E	$\left(\sum_{i=1}^k \frac{\hat{\beta}_i}{\sqrt{k}}\right) / \widehat{\sigma_C^2}$	Empirical	$\hat{\beta}_i$	ISE.
FFX	$\left(\sum_{i=1}^k \frac{\hat{\beta}_i}{s_i^2}\right) / \sqrt{\sum_{i=1}^k 1/s_i^2}$	$\mathcal{T}_{(\sum_{i=1}^k n_i - 1)-1}$	$\hat{\beta}_i, s_i^2$	$\tau^2 = 0$, large sample.
Fisher's	$-2\sum_i \ln P_i$	$\chi^2_{(2k)}$	Z_i	$\tau^2 = 0$
Stouffer's	$\sqrt{k} \times \frac{1}{k}\sum_i Z_i$	$\mathcal{N}(0,1)$	Z_i	$\tau^2 = 0$
Weighted Z	$\frac{1}{\sqrt{\sum_i n_i}} \sum_i \sqrt{n_i} Z_i$	$\mathcal{N}(0,1)$	Z_i, n_i	$\tau^2 = 0$
Perm. Z	$\left(\sum_{i=1}^k Z_i\right) / \sqrt{k}$	Empirical	Z_i	ISE.

Table 1. Statistics for one-sample meta-analysis tests and their sampling distributions under the null hypothesis H0.

Methods

We studied **8 IBMA methods** (Table 1) and investigated the validity of each estimator with **Monte Carlo simulations** under H_0 with:

- $k \in \{5,10,25,50\}$ studies; n = 20 subjects; also k = 25, n = 100.
- $\tau^2 = 0$ (**homogeneity**) or $\tau^2 = 1$ (**heterogeneity**);
- $\sigma^2_i = n \times \{0.25,0.5,1,2,4\}$ (**homoscedasticity**) or varying between 1 and $\alpha \in \{2, 4, 8, 16\}$ (**heteroscedasticity**),
- 10^6 realisations.

Notations τ^2: pure between-study variance, σ^2_i: ith study's variance, σ_C^2: usual one-sample variance. IGE=Independent Gaussian Errors, ISE=Independent Symmetric Errors. Note: $P_i=\Phi(-Z_i)$

Results

Fig. 2 presents method performance in terms of P-value distributions under different violations of model assumptions.

When the number of subjects is small (Fig. 2A), FFX is invalid regardless of the number of studies included in the meta-analysis. MFX is conservative for small number of studies and constant within-study variance. More surprisingly, MFX is invalid in the presence of large variations in the within-study variances, regardless of the number of subjects included in each study. **Under heteroscedasticity** (Fig. 2B)**, RFX and Perm. E appear robust.** For small P-values, Perm. E is conservative as expected due to the discrete nature of its distribution.

Under heterogeneity (Fig. 2C), all fixed-effects methods are invalid.

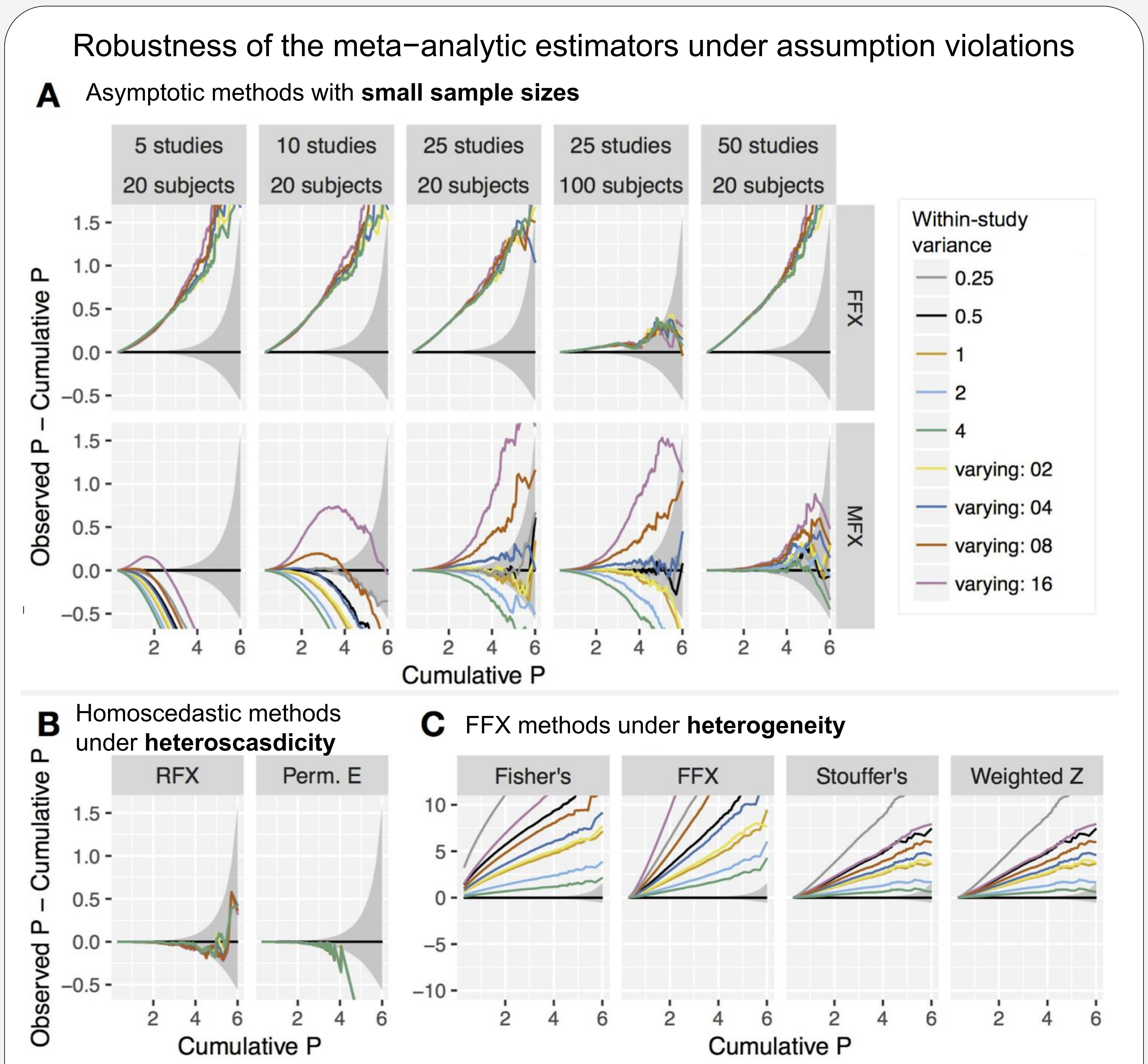


Fig. 2. Deviation of observed from theoretical P-values (difference of observed and Monte Carlo ('true') -log10 p-value distributions) in one-sample tests under violations of the underlying model assumptions. Positive deflections in Y-axis correspond to inflated false positive risk.

Conclusion

As expected, fixed-effects methods were invalid in the presence of heterogeneity. In line with fMRI literature [9], homoscedastic methods were robust to heteroscedasticity. More surprisingly, MFX was invalid in the presence of strong heteroscedasticity due to its approximations in small samples.

Given the still relatively small sample sizes that can be achieved **in IBMA as of today, we recommend using RFX, Perm. E, Z MFX or Perm. Z that do not rely on small sample approximations and are robust to both heterogeneity and heteroscedasticity**. Although they are suboptimal [10], until full metadata are routinely shared, we recommend Z-based methods that are insensitive to units.

Acknowledgments

This work was supported by the Wellcome Trust. Majority of the work was conducted while TEN and CM were at the University of Warwick and used the High Performance Computing cluster of the Department of Statistics, University of Warwick.

References

[1] Salimi-khorshidi, Neuroimage 2009. [2] Pernet, Front Neurosci. 2014 [3] Nichols, https://blog.nisox.org/2012/07/31/spm-plot-units/ 2017. [4] Nichols, Hum Brain Mapp. 2002. [5] Woolrich, Neuroimage. 2004. [6] Fisher 1932. [7] Stouffer 1949. [8] Zaykin J Evol Biol. 2011. [9]. Mumford, Neuroimage. 2009. [10] Cummings 2004.