Choosing a practical and valid Image-Based Meta-Analysis

Camille Maumet, Thomas E. Nichols

Choosing a practical and valid Image-Based Meta-Analysis

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

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, Oxford, UK.

Introduction

Meta-analysis provides a quantitative approach to summarise the rich functional Magnetic Resonance Imaging (fMRI) literature. When image data is available for each study, the optimal approach is to perform an Image-Based Meta-Analysis (IBMA) [1]. A number of IBMA approaches have been proposed including combination of standardised statistics (Z's), just effect estimates (E's) or both effect estimates and their standard errors (SE's). While using both E's & 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], especially when combining data from different software packages. 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. Here, we investigate the validity of 8 IBMA approaches.

Methods

We studied 8 IBMA methods: Random-effects (RFX) General Linear Model (GLM), contrast permutations (Perm. E) [4], mixed-effects GLM (MFX) [5], fixed-effects GLM (FFX), Fisher's [6], Stouffer's [7] and weighted Z [8] approaches, Permutations on Z's (Perm. Z). All methods are either approximate in small samples or rely on assumptions that might not be verified in the context of neuroimaging meta-analyses (Table 1).

We used Monte Carlo simulations to investigate the validity of each estimator. We simulated E's and SE's estimates under the null hypothesis with: $k \in \{5,10,25,50\}$ studies; n=20 subjects; between-study variance of 0 (homogeneity) or 1 (heterogeneity); within-study variance equal to n x $\{0.25,0.5,1,2,4\}$ (homoscedastic) or varying between 1 and $\alpha \in \{2, 4, 8, 16\}$ (heteroscedastic), 10^6 realisations. To simulate units mismatch, $\{20, 50\}\%$ of studies had (E, SE) pairs rescaled to simulate mismatched contrast vectors (4x) or data scaling algorithms (2x).

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, 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. RFX and Perm. E appear robust to heteroscedasticity for all settings studied. For small P-values, Perm. E is conservative as expected due to the discrete nature of its distribution. All fixed-effects methods are invalid under heterogeneity.

When different scaling algorithm are used, Perm. E has a behaviour that is close to nominal. RFX is valid but conservative. MFX is robust to the presence of unit mismatches when studies are homoscedastic. In the presence of strong heteroscedasticity, MFX is invalid (as when the units are matched). In the presence of slight heteroscedasticity, unit mismatches can cause invalidity of otherwise valid MFX. When the contrast are scaled differently, we observe a very similar pattern.

Conclusions [678 char]

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

Figures

	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 H_0. Empirical null distributions are determined using permutations with sign flipping. IGE=Independent Gaussian Errors, ISE=Independent Symmetric Errors. Note: $P_i = \Phi(-Z_i)$, τ^2 is pure between-study variance, σ_i^2 is the ith study's variance, and $\widehat{\sigma_C^2}$ is the usual one-sample variance.

Table 1. Statistics for one-sample meta-analysis tests and their sampling distributions under the null hypothesis H0. Empirical null distributions are determined using permutations with sign flipping. IGE=Independent Gaussian Errors, ISE=Independent Symmetric Errors. Note: Pi =Φ(−Zi), τ^2 is pure between-study variance, σ^2_i is the ith study's variance, and σ_C^2 is the usual one-sample variance.

References

1. Salimi-khorshidi G, Smith SM, Keltner JR, Wager TD, Nichols TE. Meta-analysis of neuroimaging data: a comparison of image-based and coordinate-based pooling of studies. Neuroimage. Elsevier B.V.; 2009;45: 810–823.

2. Pernet CR. Misconceptions in the use of the General Linear Model applied to functional MRI: a tutorial for junior neuro-imagers. Front Neurosci. 2014;8: 1–12.

3. Nichols, Thomas E. SPM plot units. In: NISOx [Internet]. 31 Jul 2012 [cited 11 Dec 2017]. Available: https://blog.nisox.org/2012/07/31/spm-plot-units/

4. Nichols TE, Holmes AP. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum Brain Mapp. 2002;15: 1–25.

5. Woolrich MW, Behrens TEJ, Beckmann CF, Jenkinson M, Smith SM. Multilevel linear modelling for FMRI group analysis using Bayesian inference. Neuroimage. 2004;21: 1732–1747.

6. Fisher, R.A. Statistical Methods for Research Workers. Edinburgh: Oliver and Boyd; 1932.

7. Stouffer S, DeVinney L, Suchmen E. The American Soldier: Adjustment During Army Life. Princeton, NJ: Princeton University Press; 1949.

8. Zaykin DV. Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis. J Evol Biol. 2011;24: 1836–1841.

9. Mumford JA, Nichols TE. Simple group fMRI modeling and inference. Neuroimage. Elsevier Inc.; 2009;47: 1469–1475.

10. Cummings P. Meta-analysis based on standardized effects is unreliable. Arch Pediatr Adolesc Med. 2004;158: 595–597.
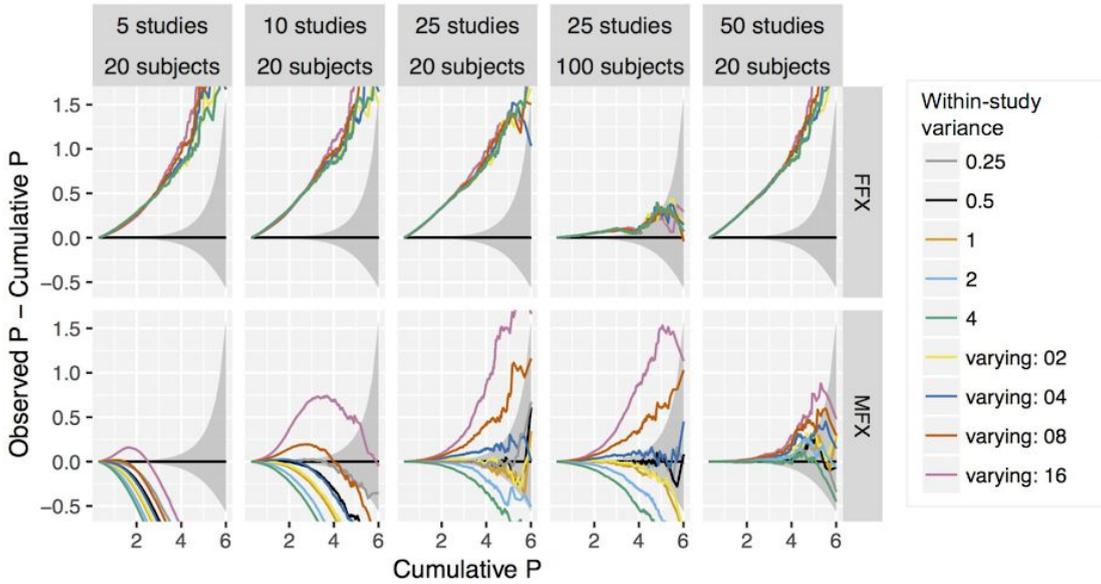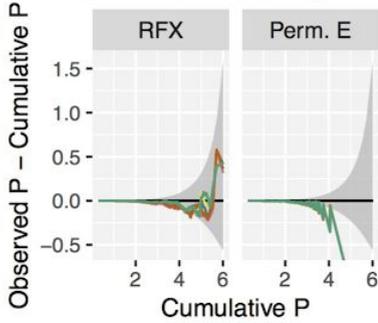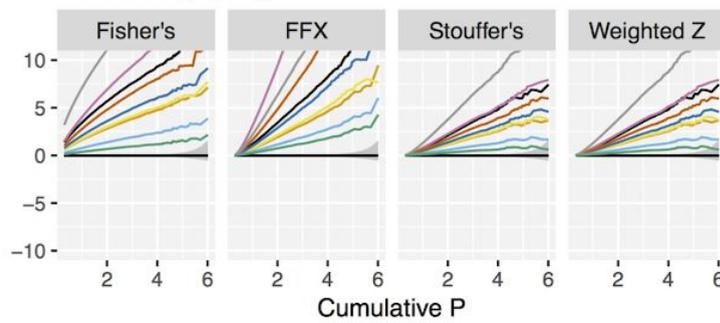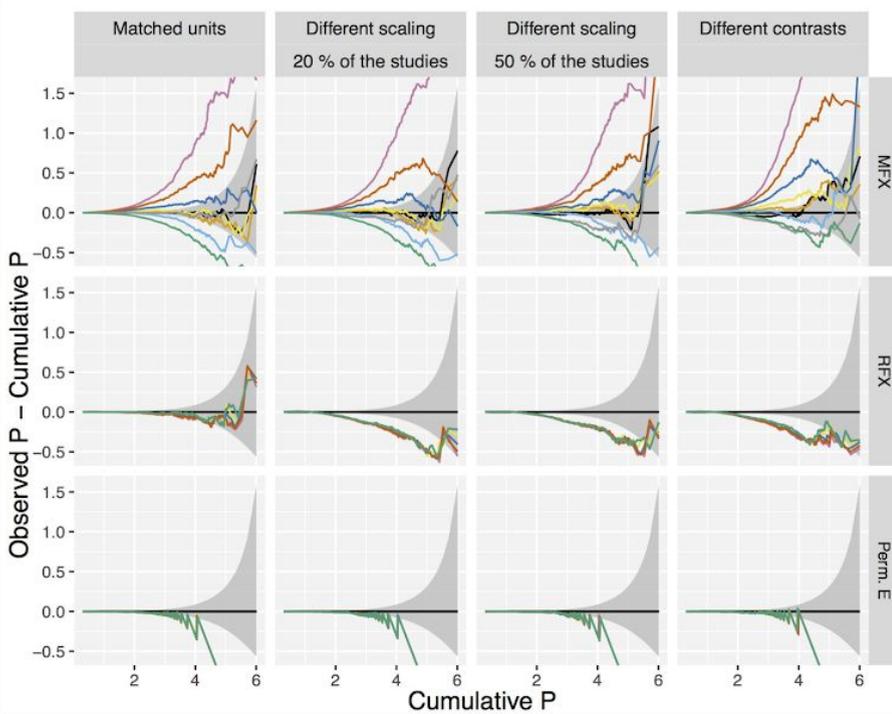
Fig. 2. Deviation of observed from theoretical P-values (difference of observed and Monte Carlo ('true') -log10 p-value distributions) in one-sample tests under violations of the underlying model assumptions: small sample sizes (A), heteroscedasticity (B), heterogeneity (C) unit mismatch (D). P-values are displayed using a negative log10 scale. Y-axis is the observed cumulative probability minus Monte Carlo cumulative probability for a given (X-axis) -log10 p-value; positive deflections correspond to inflated false positive risk.