Combining fMRI studies in Image-Based Meta-analysis

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

[1] Warwick Manufacturing Group, The University of Warwick, Coventry, UK.
[2] Statistics Department, The University of Warwick, Coventry, UK.

Abstract. The abstract should summarize the contents of the paper using at least 70 and at most 150 words. It will be set in 9-point font size and be inset 1.0 cm from the right and left margins. There will be two blank lines before and after the Abstract. ...

Keywords: TODO

1 Introduction

2 Methods

2.1 Theory

Given a set of k studies, we denote for each study i: its contrast estimate by Y_i, its contrast variance estimate by V_{Y_i}, its standardized statistical map by Z_i and its sample size by n_i.

Combining contrast estimates and their standard error The gold standard approach to combine contrast estimates and their standard errors is to input them into a GLM [3], creating effectively the third-level of a hierarchical model (level 1: subject; level 2: study; level 3: meta-analysis). The general formulation is:

$$Y = X\beta + \epsilon \tag{1}$$

where β is the meta-analytic parameter to be estimated, $Y = [Y_1 \dots Y_k]^T$ is the vector of contrast estimates and $\epsilon \sim \mathcal{N}(0, W)$ is the residual term. Eq. (1) can be solved by weighted least square giving:

$$\hat{\beta} = (X^T W X)^{-1} X^T W Y \tag{2}$$

$$\mathrm{Var}(\hat{\beta}) = (X^T W X)^{-1} \tag{3}$$

In a random-effects model, we have $W = diag(\sigma_1^2 + \tau^2 \dots \sigma_k^2 + \tau^2)$ where τ^2 denotes the between-studies variance. Approximating σ_i^2 by V_{Y_i} and given $\hat{\tau^2}$ an estimate of τ^2 we obtain the statistics detailed in table 1 for a one sample test. This reference approach will be referred to as **Mixed-effects (MFX) GLM**.

In a fixed-effects model (i.e. assuming no between-study variances), we have $W = diag(\sigma_1^2 \dots \sigma_k^2)$ where σ_i^2 denotes the contrast variance for study i. This approach will be referred to as **Fixed-effects (FFX) GLM**.

	Statistic	Disbribution under H_0
FFX GLM	$\frac{1}{\sqrt{\sum_{i=1}^{k} 1/V_{Y_i}}} \sum_{i=1}^{k} \frac{Y_i}{V_{Y_i}}$	$\mathcal{T}_{(\sum_{i=1}^{k} n_i - 1) - 1}$
MFX GLM	$\frac{1}{\sqrt{\sum_{i=1}^{k} 1/(V_{Y_i} + \hat{\tau^2})}} \sum_{i=1}^{k} \frac{Y_i}{V_{Y_i} + \hat{\tau^2}}$	$\mathcal{T}_{k-1}$
RFX GLM	$\frac{1}{\widehat{\sigma}_C^2/\sqrt{k}} \sum_{i=1}^{k} \frac{Y_i}{k}$	$\mathcal{T}_{k-1}$
Contrast Permutation	$\frac{1}{\widehat{\sigma}_C^2/\sqrt{k}} \sum_{i=1}^{k} \frac{Y_i}{k}$	Determined through permutations with sign switching.
Fisher's	$-2 \sum_{i=1}^{k} \ln(\Phi(-Z_i)))$	$\chi^2_{(2k)}$
Stouffer's	$\frac{\sum_{i=1}^{k} Z_i}{\sqrt{k}}$	$\mathcal{N}(0,1)$
Optimally weighted-Z	$\frac{\sum_{i=1}^{k} \sqrt{n_i} Z_i}{\sqrt{\sum_{i=1}^{k} n_i}}$	$\mathcal{N}(0,1)$
Stouffer's MFX	$\frac{\sum_{i=1}^{k} Z_i}{\sqrt{k}\hat{\sigma}}$	$\mathcal{T}_{k-1}$
Z Permutation	$\frac{\sum_{i=1}^{k} Z_i}{\sqrt{k}}$	Determined through permutations with sign switching.

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

Combining contrast estimates In the absence of standard error, the contrast estimates Y_i can be combined by assuming that the within-study variance σ_i^2 is roughly constant ($\sigma_i^2 \simeq \sigma^2 \ \forall 1 \leq i \leq k$) or negligible by comparison to the between-study variance($\sigma_i^2 \ll \tau^2 \ \forall 1 \leq i \leq k$). Then $W = diag(\sigma_C^2 \dots \sigma_C^2)$ where σ_C^2 is the combined within and between-subject variance such as $\sigma_C^2 \simeq \tau^2$ or $\sigma_C^2 \simeq \tau^2 + \sigma^2$. Under these assumptions, eq. (1) can be solved by ordinary least square giving:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{4}$$

$$\mathrm{Var}(\hat{\beta}) = (X^T W X)^{-1} \tag{5}$$

Given $\hat{\sigma_C^2}$ an estimate of σ_C^2 we obtain the statistics presented in table 1 for one sample tests. This approach will be referred to **Random-effects (RFX) GLM** in the following.

As an alternative to parametric approaches, non-parametric statistics [5, 7] can be computed by comparing the RFX GLM T-statistic to the distribution obtained by permuting the sign of each sample included in the analysis. This approach will be referred to as **Contrast permutation**.

Combining standardised statistics In the presence of standardised statistical estimates, **Fisher's** meta-analysis provide a statistic to combine the associated p-values [4]. **Stouffer's** approach combines directly the standardised statistic [8]. In [10] following [2], the author proposed a weighted method that weights each study's Z_i by the square root of its sample size [3,7]. This approach will be referred to as **Optimally weighted-Z**. All these meta-analytic statistics assumes fixed-effects (no between-study variance) and are suited only for one-sample tests. The corresponding statistics are presented in table 1.

As suggested in [1], to get a kind of MFX with Stouffer's approach, the standardised statistical estimates Z_i can be combined in an OLS analysis. The corresponding estimate, referred as **Stouffer's MFX** is also provided in 1

As an alternative to parametric approaches, a non-parametric distribution [5, 7] can be estimated by permutation on the Z_i's. This approach will be referred to as **Z permutation**.

2.2 Experiments

Simulations To verify the validity of each estimator under the null hypothesis we estimated the false positive rate at $p < 0.05$ uncorrected. For each meta-analysis, we simulated a contrast estimate and a variance estimates such as:

$$Y_i \sim \mathcal{N}(0, \frac{\sigma_i^2}{n_i} + \tau^2) \tag{6}$$

$$V_{Y_i} \sim \frac{\sigma_i^2}{n_i - 1}\chi^2_{(n_i-1)} \tag{7}$$

where $\sigma_i^2 \in [1/2, 1, 2, 4]$ is the within-study variance, $\tau^2 \in [0, 1]$ is the between-study variance (fixed-effects if τ^2 is 0, random-effects otherwise). We simulated different number of studies per meta-analysis: $k \in [5, 10, 25, 50]$ and the number of subjects per studies n_i was selected such as we would have varying number of subjects per studies in given meta-analysis across the common range of subjects involved in neuroimaging studies. In each simulated meta-analysis we simulated one study with exactly 20, 25, 10 and 50 subjects. For the remaining studies the number of subjects were drawn from uniform distributions a quarter from $\mathcal{U}(11, 20)$, a quarter from $\mathcal{U}(26, 50)$ and the remaining from $\mathcal{U}(21, 25)$. A total of 32 parameter sets (4 σ_i^2 x 2 τ^2 x 4 k) was therefore tested, 71 repeats with 5041 samples per repeats were simulated.

Real data We first compared the Z-scores obtained by the three approaches using a Bland-Altman plot. Then, as results are usually presented as a thresholded map, we computed the dice similarity score between thresholded maps obtained with Stouffer's and weighted-Z FFX with FLAME FFX for three (uncorrected) thresholds: p ¡ 0.001, 0.01 and 0.05. Finally, as results are best reported using a multiple comparison correction, we defined ground truth activations as the FLAME FFX analysis FDR-corrected at a threshold of p¡0.05 and plotted

Receiver-Operating-Characteristics (ROC) curves of Stouffer's and weighted-Z FFX.

All plots were generated using ggplot [9].

3 Results

3.1 Simulations

Fig. 1 displays the false positive rate at $p < 0.05$ obtained for the eight estimators over all set of parameters in the absence and presence of random-effects. From this graph, it is clear that the fixed-effects meta-analytic summary statistics, i.e. Fisher's, Stouffer's and weighted-z estimates are overly liberal in the presence of random-effects. As expected the original Fisher's approach is the most invalid. Surprisingly, FFX GLM is also invalid under fixed-effects, maybe suggesting inaccurate degrees of freedoms (here set to $(\sum_{i=1}^{k} n_i - 1) - 1)$). Stouffer's MFX, GLM RFX and permutations of Y_i's or Z_i's provide valid estimates. The permutation estimates present the largest sampling variance.

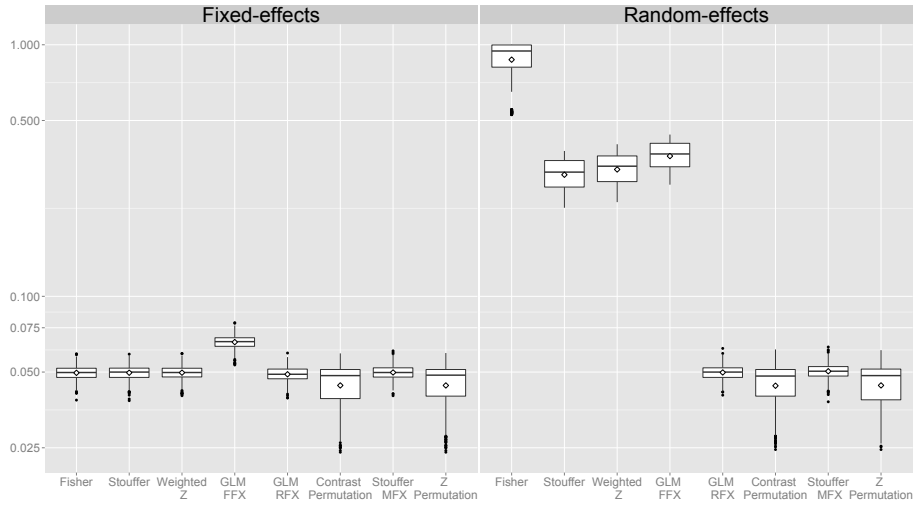

Fig. 1. False positive rates of the meta-analytic estimators under the null hypothesis for $p < 0.05$.

The impact of the number of studies involved in the meta-analysis and of the size of the within-study variance are investigated in fig. 2. The permutation estimates appears conservative (FPR $\simeq 0.03$) when 5 studies are involved. All approaches perform equally as soon as 10 or more studies are included in the meta-analysis.

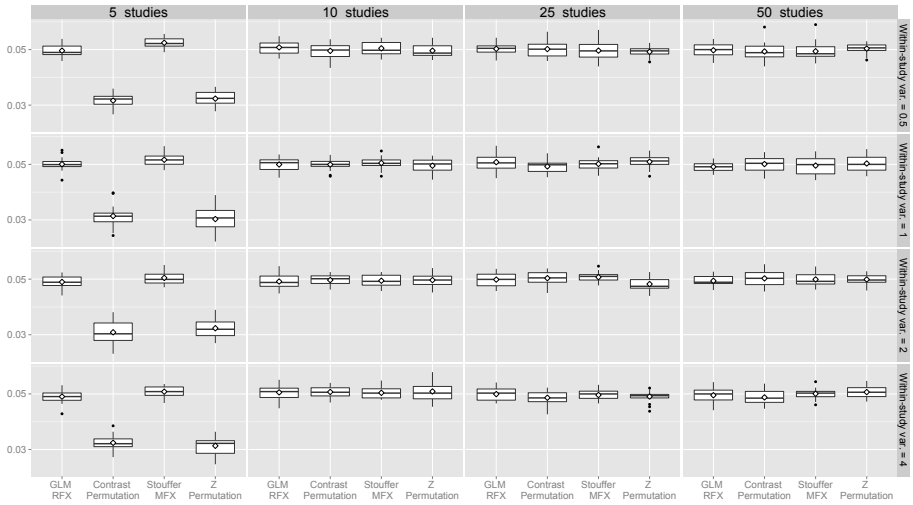

Fig. 2. False positive rates of the valid random-effects meta-analytic estimators under the null hypothesis for $p < 0.05$ as a function of the number of studies and the within-study variance.

3.2 Real data

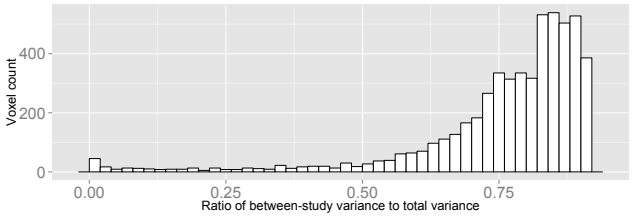

Fig. 3. Histogram of the between-study variance to the sum of the between-subject variance and the mean within-study variance.

The histogram of the ratio of between-subject variance onto total variance is displayed in fig. 3. From this graph it is clear that for most of the voxels the estimated between-study variance is greater than the within-study variance. We can therefore suppose the presence of random-effects (non zero between-study variance) in this dataset.

Fig. 4 and fig. 5 plots the difference between the z-score estimated by each meta-analytic approach and the reference z-score computed with MFX GLM for FFX and RFX approaches respectively. All FFX statistics provide overly

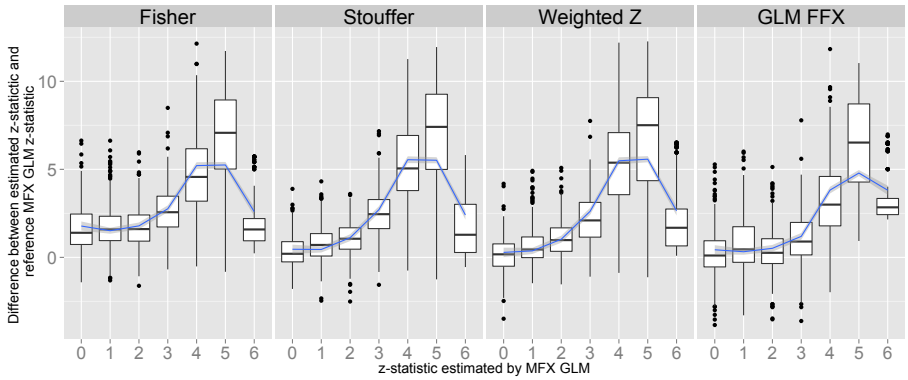

Fig. 4. Difference between the z-score estimated from each FFX meta-analytic approach and the reference z-score from MFX GLM as a function of reference z-score.

optimistic z-estimate suggesting, again, that random-effects are indeed present in the studied dataset.

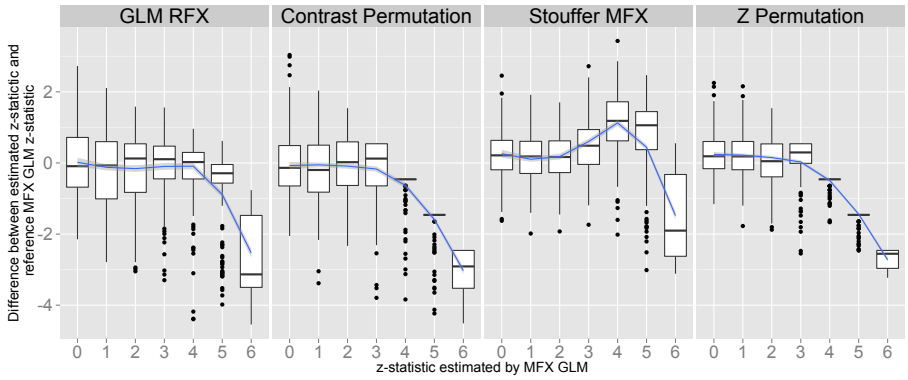

Fig. 5. Difference between the z-score estimated from each RFX meta-analytic approach and the reference z-score from MFX GLM as a function of reference z-score.

Among the RFX meta-analytic approaches presented in fig. 5, GLM RFX and contrast permutations provide z-scores estimate that are equal or smaller than the reference. Z permutation provides slightly larger z-scores between 1 and 3 (reference p-values between 0.16 and 0.0013) but is mostly in agreement with the reference z-scores. On the other hand, Stouffer's MFX is more liberal than the reference for z-score ranging from 3 to 5 (reference p-values between 0.0013 and 2.9e-07) and more stringent for z-scores smaller than 5.

4 Conclusion

We have compared eight meta-analytic statistic in the context of one-sample test. Through simulations, we outlined the invalidity of standard FFX approaches in the presence of random-effects. In a real dataset of 21 studies of pain, we outline the presence of random-effects advocating for the use of RFX meta-analytic statistics. When contrast estimates only are available, the RFX procedure was valid. This is in line with previous results on within-group one-sample t-tests studies [6]. When only standardised estimates are available, permutation is the preferred option as the one providing the most faithful results. Further investigations are needed in order to investigate the behaviour of these estimators in other configurations, including meta-analyses focusing on between-study differences.

5 Acknowledgements

We gratefully acknowledge the use of this data from the Tracey pain group, FMRIB, Oxford.

References

1. Meta-analysis of neuroimaging data: a comparison of image-based and coordinate-based pooling of studies. *NeuroImage*, 45(3):810–23, 2009.
2. Peter Cummings. On the combination of independent tests. *Magyar Tud. Akad. Mat. Kutato Int. Kozl.*, 3:171–197, 1958.
3. Peter Cummings. Meta-analysis based on standardized effects is unreliable. *Archives of pediatrics & adolescent medicine*, 158(6):595–7, 2004.
4. R.A. Fisher. *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh, 1932.
5. A. P. Holmes, R. C. Blair, G. J.D. Watson, and I. Ford. Nonparametric Analysis of Statistic Images from Functional Mapping Experiments. *Journal of Cerebral Blood Flow and Metabolism*, 16:7–22, 1996.
6. Jeanette a Mumford and Thomas E. Nichols. Simple group fMRI modeling and inference. *NeuroImage*, 47(4):1469–75, October 2009.
7. Thomas E. Nichols and Andrew P Holmes. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human brain mapping*, 15(1):1–25, January 2002.
8. S. Stouffer, L. DeVinney, and E. Suchmen. *The American Soldier: Adjustment During Army Life*, volume 1. Princeton University Press, Princeton, NJ, 1949.
9. H. Wickham. *ggplot2: elegant graphics for data analysis*. Springer New York, 2009.
10. D V Zaykin. Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis. *Journal of evolutionary biology*, 24(8):1836–41, 2011.