

# Minimal Data Needed for Valid & Accurate Image-Based fMRI Meta-Analysis

Camille Maumet, Thomas Nichols

*WMG, University of Warwick, Coventry, UK*

*Statistics Department, University of Warwick, Coventry, UK.*

---

## Abstract

Meta-analysis is a powerful statistical tool to combine results from a set of studies. When image data is available for each study, a number of approaches have been proposed to perform such meta-analysis including combination of standardised statistics, just effect estimates or both effects estimates and their sampling variance. While the latter is the preferred approach in the statistical community, often only standardised estimates are shared, reducing the possible meta-analytic approaches. 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 image-based meta-analysis. In this paper, we compare the validity and the accuracy of eight meta-analytic approaches on simulated and real data. In one-sample tests, combination of contrast estimates into a random-effects General Linear Model or non-parametric statistics provide a good approximation of the reference approach. If only standardised statistical estimates are shared, permutations of z-score is the preferred approach.

*Keywords:* Science, Publication, Complicated

---

## <sup>1</sup> 1. Introduction

<sup>2</sup> TODO: check number of meta-analytic approaches (8 or 9) and update  
<sup>3</sup> everywhere

<sup>4</sup> A growing literature is focusing on the lack of statistical power in neu-  
<sup>5</sup> roimaging studies (see, e.g. [2]), feeding the debate on the validity and re-  
<sup>6</sup> producibility of published neuroimaging results. Meta-analysis, by providing

7 inference based on the results of previously conducted studies, provides an  
8 essential method to increase power and hence confidence in neuroimaging.

9 A number of methods have been proposed for neuroimaging meta-analysis  
10 (see [11] for a review). As the results of neuroimaging studies are usually  
11 conveyed by providing a table of peak coordinate and statistics, most of  
12 these meta-analyses are restricted to combining coordinate-based informa-  
13 tion. Nevertheless the best practice method is an Intensity-Based Meta-  
14 Analysis (IBMA) that combines the effect estimates and their standard errors  
15 from each study [1].

16 In order for IBMA to be possible in neuroimaging, tools for sharing 3D  
17 volumes obtained as a result of a statistical analysis are needed. Various  
18 efforts are currently underway to facilitate sharing of neuroimaging data but  
19 emphasis is usually on statistical maps (see, e.g. [2]). There are three evident  
20 approaches to sharing summary data from each study  $i$ :

- 21 1. the contrast estimates  $\hat{\beta}_i$  and contrast variance estimates  $\hat{S}_i^2$ .
- 22 2. the contrast estimates  $\hat{\beta}_i$ .
- 23 3. the standardized statistical maps  $Z_k$ .

24 Depending on how much data is shared, different strategies can be used to  
25 combine the available results into a meta-analysis. While the first option is  
26 the best practice, leading to statistically optimal estimates [4], it requires the  
27 contrasts to be expressed with in the same units. In fMRI, units will depends  
28 on data, model and contrast vector scaling and are typically different across  
29 neuroimaging software due to different data scaling approaches [10].

30 TODO: units also depend on field strength: 3T/7T TODO: cite Gang's  
31 paper

32 Given the growing interest in data sharing in the neuroimaging commu-  
33 nity, and the relative easiness of sharing just (unitless) statistic maps, there  
34 is a need to identify what is the minimal data to be shared in order to allow  
35 for future IBMA.

36 Here we compare the use of IMBA using 9 meta-analytic approaches: 2  
37 approaches use  $\hat{\beta}_i$ 's and  $\hat{S}_i^2$ 's, 2  $\hat{\beta}_i$ 's only and 5  $Z_k$ 's. We compare the validity  
38 and the accuracy of the eight meta-analytic approaches on simulated and  
39 real data including 21 studies of pain in control subjects.

40 Section 2 describes the meta-analytic estimates along with the experi-  
41 ments undertaken on simulated and real data to assert their validity. The  
42 results are described in section 3. Finally, we conclude in section 4.

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

43 **2. Methods**

44 *2.1. Theory*

45 For study  $i = 1, \dots, k$  we have contrast estimate  $\hat{\beta}_i$ , its contrast variance  
 46 estimate  $\hat{S}_i^2$  (i.e. squared standard error), its statistic map  $Z_i$  and its sample  
 47 size  $n_i$ .

48 *Combining contrast estimates and their standard error.* The gold standard  
 49 approach is to fit contrast estimates and their standard error with a hierar-  
 50 chical general linear model (GLM) [4], creating a third-level (level 1: subject;  
 51 level 2: study; level 3: meta-analysis). The general formulation for the study-  
 52 level data is:

$$\vec{\hat{\beta}} = X\gamma + \epsilon \quad (1)$$

53 where  $\gamma$  is the meta-analytic parameter to estimate,  $\vec{\hat{\beta}} = [\hat{\beta}_1 \dots \hat{\beta}_K]^T$  is the  
 54 vector of contrast estimates,  $X$  is the  $k \times p$  study-level matrix (typically just  
 55 a column of ones for a one-sample test) and  $\epsilon \sim \mathcal{N}(0, W)$  is the residual error  
 56 term. Eq. (1) can be solved by weighted least squares giving:

57 TODO: add contrast

$$\hat{\gamma} = (X^T W^{-1} X)^{-1} X^T W^{-1} \vec{\hat{\beta}} \quad (2)$$

$$\text{Var}(\hat{\gamma}) = (X^T W^{-1} X)^{-1} \quad (3)$$

58 In a random-effects (RFX) meta-analysis, we have  $W = \text{diag}(\sigma_1^2 + \tau^2, \dots, \sigma_K^2 +$   
 59  $\tau^2)$  where  $\tau^2$  denotes the between-study variance. Approximating  $\sigma_i^2$  by  $\hat{S}_i^2$   
 60 and given  $\hat{\tau}^2$  an estimate of  $\tau^2$  we obtain the statistics detailed in Table 1  
 61 for one-sample tests. This reference approach will be referred to as **Mixed-**  
 62 **effects (MFX) GLM**. In a **fixed-effects (FFX) GLM** (i.e. assuming no  
 63 or negligible between-study variance), we have  $W = \text{diag}(\sigma_1^2 \dots \sigma_K^2)$  where  
 64  $\sigma_i^2$  denotes the contrast variance for study  $i$ .

65 TODO: need to state the assumptions (e.g. within-study is the true  
 66 within study var?)

67 TODO: We need a second table for two-sample tests

	Meta-analysis statistic	Nominal H <sub>0</sub> distrib.	Inputs	Assumptions
FFX GLM	$\left( \sum_{i=1}^K \frac{\hat{\beta}_i}{\hat{S}_i^2} \right) / \sqrt{\sum_{i=1}^K 1/\hat{S}_i^2}$	$\mathcal{T}_{(\sum_{i=1}^K n_i - 1) - 1}$	$\hat{\beta}_i, \hat{S}_i^2$	IGE; $\sigma_i^2$ cst; $\tau^2 = 0$
MFX GLM	$\left( \sum_{i=1}^K \frac{\hat{\beta}_i}{\hat{S}_i^2 + \hat{\tau}^2} \right) / \sqrt{\sum_{i=1}^K 1/(\hat{S}_i^2 + \hat{\tau}^2)}$	$\mathcal{T}_{K-1}$	$\hat{\beta}_i, \hat{S}_i^2$	IGE; $\tau^2 = \hat{\tau}^2$ .
RFX GLM	$\left( \sum_{i=1}^K \frac{\hat{\beta}_i}{\sqrt{K}} \right) / \widehat{\sigma}_C^2$	$\mathcal{T}_{K-1}$	$\hat{\beta}_i$	IGE; $\tau^2 + \sigma_i^2$ cst.
Ctrst Perm.	$\left( \sum_{i=1}^K \frac{\hat{\beta}_i}{\sqrt{K}} \right) / \widehat{\sigma}_C^2$	Empirical	$\hat{\beta}_i$	ISE.
Fisher's	$-2 \sum_k \ln P_k$	$\chi^2_{(2K)}$	$Z_k$	IGE; $\tau^2 = 0$ .
Stouffer's	$\sqrt{K} \times \frac{1}{K} \sum_k Z_k$	$\mathcal{N}(0, 1)$	$Z_k$	IGE; $\tau^2 = 0$ .
Wght Stouff.	$\frac{1}{\sqrt{\sum_k n_k}} \sum_k \sqrt{n_k} Z_k$	$\mathcal{N}(0, 1)$	$Z_k, n_i$	IGE; $\tau^2 = 0$ .
Z MFX	$\left( \sum_{i=1}^K Z_k \right) / \sqrt{K} \hat{\sigma}$	$\mathcal{T}_{K-1}$	$Z_k$	IGE; $1 + \tau^2 / \sigma_i^2$ cst
Z Perm.	$\left( \sum_{i=1}^K Z_k \right) / \sqrt{K}$	Empirical	$Z_k$	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:  $P_k = \Phi(-Z_k)$

68 *Combining contrast estimates.* If the  $\hat{S}_i^2$  are unavailable, the contrast es-  
69 timates  $\hat{\beta}_i$  can be combined by assuming that the within-study contrast  
70 variance  $\sigma_i^2$  is roughly constant ( $\sigma_i^2 \simeq \sigma^2$ ) or negligible in comparison to  
71 the between-study variance( $\sigma_i^2 \ll \tau^2$ ). Then  $W = \text{diag}(\sigma_C^2, \dots, \sigma_C^2)$  where  
72  $\sigma_C^2$  is the combined within and between-subject variance, i.e.  $\sigma_C^2 \simeq \tau^2$  or  
73  $\sigma_C^2 \simeq \tau^2 + \sigma^2$  (note, however, in this setting we do not separately estimate  
74  $\tau^2$  or  $\sigma^2$ ). Under these assumptions, Eq. (1) can be solved by ordinary least  
75 squares giving:

$$\hat{\gamma} = (X^T X)^{-1} X^T \hat{\beta} \quad (4)$$

$$\text{Var}(\hat{\gamma}) = (X^T X)^{-1} \sigma_C^2 \quad (5)$$

76 Given  $\hat{\sigma}_C^2$  an estimate of  $\sigma_C^2$  we obtain the statistics presented in Table 1  
77 for one sample tests. This approach will be referred to **RFX GLM** in the  
78 following.

79 As an alternative to parametric approaches, non-parametric inference [6,  
80 9] can be performed by comparing the RFX GLM T-statistic to the distri-  
81 bution obtained with “sign flipping”, i.e. randomly multiplying each study’s  
82 data by 1 or -1, justified by an assumption of independent studies and sym-  
83 metrically distributed random error. This approach will be referred to as  
84 **Contrast permutation.**

85 TODO: We should be able to do something if we have the sample sizes  
86 by assuming constant within subject variance

87 TODO: We should be able to do something if we have the sample sizes  
88 by assuming constant within subject variance

89 *Combining standardised statistics.* When only test statistic images are avail-  
90 able there are several alternate approaches available. **Fisher’s** meta-  
91 analysis provide a statistic to combine the associated p-values [5]. **Stouffer’s**  
92 approach combines directly the standardised statistic [14]. In [15] following  
93 [7], the author proposed a weighted method that weights each study’s  $Z_k$  by  
94 the square root of its sample size [3,7]. This approach will be referred to  
95 as **Weighted Stouffer’s**. All these meta-analytic statistics assumes no or  
96 negligible between-study variance and are suited only for one-sample tests.  
97 The corresponding statistics are presented in Table 1. As suggested in [1], to  
98 get a kind of MFX with Stouffer’s approach, the standardised statistical esti-  
99 mates  $Z_k$  can be combined in an OLS analysis. The corresponding estimate,  
100 referred as **Z MFX** is also provided in 1

101 With contrasts, non-parametric inference [6, 9] can be obtained by sign  
102 flipping on the  $Z_k$ ’s. This approach will be referred to as **Z permutation.**

103 *Approximations.* In practice, all of the methods based on contrast data have  
104 approximate parametric null distributions. The nominal distributions listed  
105 in Table 1 are under the (unrealistic) assumption of homogeneous standard  
106 errors over studies; even if all studies are ‘clean’ and conducted at the same  
107 center, variation in sample size will induce differences in  $\hat{S}_i^2$ ’s. Further, even  
108 under homoscedasticity, MFX GLM’s null is approximate due to iterative  
109 estimation of  $\hat{\tau}^2$ .

110 TODO: Clarify what is approximate in each + units issue + RFX/FFX  
111 choice

112 *2.2. Experiments*

113 *2.2.1. Simulations*

114 Due to the approximate nature of the sampling distributions, we conduct  
115 simulations to evaluate the validity of each estimator under inhomogeneity of  
116 contrast variances  $\hat{S}_i^2$  and under the presence of non-negligible between-study  
117 variance.

118 TODO: No longer at  $p < 0.05$

119 To verify the validity of each estimator under the null hypothesis we  
120 estimated the false positive rate at  $p < 0.05$  uncorrected. For each meta-  
121 analysis, we simulated  $\hat{\beta}_i$  and  $\hat{S}_i^2$  such as:

$$\hat{\beta}_i \sim \mathcal{N}\left(0, \frac{\sigma_i^2}{n_i} + \tau^2\right) \quad (6)$$

$$\hat{S}_i^2 \sim \frac{\sigma_i^2}{n_i - 1} \chi_{(n_i - 1)}^2 \quad (7)$$

122 where  $\sigma_i^2 \in [1/2, 1, 2, 4]$  is the within-study variance,  $\tau^2 \in [0, 1/20]$  is the  
123 between-study variance (fixed-effects models are strictly only appropriate  
124 for  $\tau^2 = 0$ ). For different number of studies per meta-analysis we used:  
125  $K \in [5, 10, 25, 50]$ , and set the number of subjects per studies  $n_i$  to vary  
126 across the common range of sample sizes in neuroimaging studies. In each  
127 simulated meta-analysis we simulated one study with exactly 20, 25, 10 and  
128 50 subjects. For the remaining studies 1/4 of the  $n_i$ 's were drawn from  
129  $\mathcal{U}(11, 20)$ , 1/4 from  $\mathcal{U}(26, 50)$  and the remaining from  $\mathcal{U}(21, 25)$ , where  $\mathcal{U}(a, b)$   
130 is the discrete uniform distribution on the integers  $a$  to  $b$  inclusive. A total  
131 of 32 parameter sets ( $4 \sigma_i^2 \times 2 \tau^2 \times 4 K$ ) was therefore tested and a total of  
132  $71^3$  realisations were created.

133 *2.2.2. Real data*

134 We then compared the 8 meta-analytic estimators to the reference ap-  
135 proach, MFX GLM, on a dataset of 21 studies of pain. Comparability of  
136 contrast estimates depends on equivalent scaling of the data, models, and  
137 contrast vectors. Data scaling was consistently performed by FSL, setting  
138 median brain intensity to 10,000; model were all created by FSL's Feat tool;  
139 and contrasts were constructed to preserve units, with sum of positive ele-  
140 ments equal to 1, sum of negative elements equal to -1.

141 To investigate the presence of between-study variation, we computed the  
142 ratio of the between-study variance (estimated using FSL's FLAME [13]) to

Figure 2: Histogram of the between-study variance to the sum of the between-subject variance and the mean within-study variance.

143 the total variance (sum of between- and within-study variances), as suggested  
144 in [3]. Here we use the average (across study) within-study variance as an  
145 estimate of within-study variance in the denominator:  $\hat{\tau}^2 / (\hat{\tau}^2 + \sum_{i=1}^K \hat{S}_i^2)$ .  
146 Using this metric, voxels with values close to 0 present negligible between-  
147 study variance and values close to 1 outline appreciable study heterogeneity  
148 and the importance of RFX models.

149 Then for each estimator we compared the standardised meta-analytic  
150 statistic to the z-statistic obtained with the reference approach. Overestima-  
151 tion of z-statistic leads to overly optimistic detections while underestimation  
152 outline a reduced sensitivity of the approach.

### 153 3. Results

#### 154 3.1. How bad is the units issue?

##### 155 3.1.1. Group meta-analysis

156 Fig. 3 presents the simulation results for a one-sample with  $\tau^2 = 1$  and  
157 a sample size  $K = 5, 25, 50$ . For the nominal case, i.e. when the units are  
158 matched across studies and contrasts, MFX GLM, RFX GLM and Contrast  
159 Permutation are all valid, as expected. For small sample sizes ( $K = 5$ ), MFX  
160 GLM and contrast permutation are both very conservative. For large values  
161 of Z, Contrast Estimation is conservative as expected due to the discrete  
162 nature of its distribution. More suprinsing, in the presence of a high within-  
163 subject variance, MFX GLM also appears to be conservative. RFX GLM  
164 displays the best behaviour with a pattern that is always within the 95%  
165 confidence interval of the theoretical Z.

166 When different scaling algorithm are used, i.e. with different neuroimag-  
167 ing software packages, Contrast Permutation still has a behaviour that is very  
168 similar to nominal. MFX GLM and RFX GLM display invalidity for small  
169 Z's and conservativness for large Z's (but less conservative than Contrast  
170 Permutation).

171 When the contrast are scaled differently, we observe a very similar pattern  
172 than for different scaling algoritm.

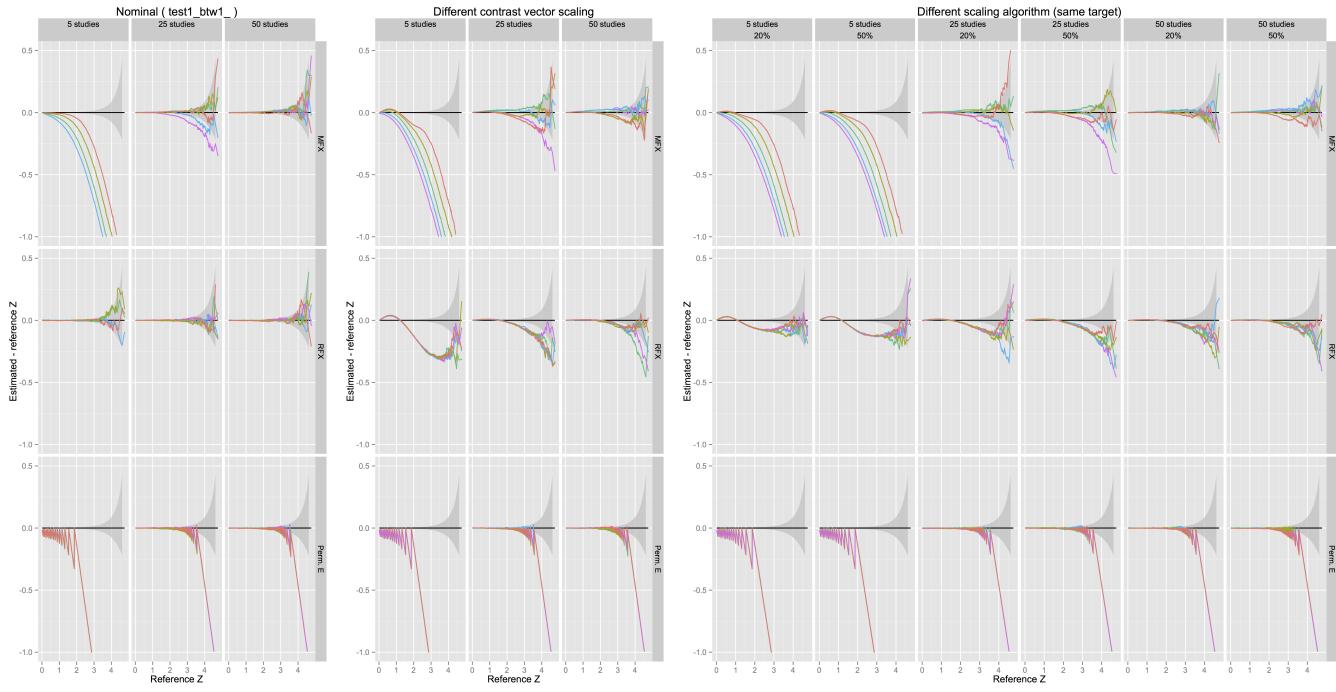


Figure 3: Deviation from theoretical Z in one-sample tests with  $\tau^2 = 1$  and  $K = 5, 25, 50$  with matched (“nominal”) or mismatched (“different scaling target”, “different scaling algorithm”, “different contrast vector scaling”) units.

### 173    3.1.2. Balanced between-group meta-analysis

174    Fig. ?? presents the simulation results for a two-sample meta-analyses  
 175    with  $\tau^2 = 1$  and a sample size  $K = 25$ . For the nominal case, GLM RFX,  
 176    GLM RFX and contrast estimation provide valid estimates. Contrast Per-  
 177    mutation is conservative for large Z values. Both RFX GLM and MFX GLM  
 178    display the best behaviour with a pattern that is within the 95% confidence  
 179    interval of the theoretical Z.

180    In the extreme case of different scaling target, contrast permutation is  
 181    always valid with a pattern very similar than its nominal behaviour. GLM  
 182    RFX is valid for Z values greater than 1.5, which is the area of interest in  
 183    detections, but display a strong conservativness, more pronounced than the  
 184    Contarst Permutation. GLM MFX is slightly invalid for all within-subject  
 185    variances except the largest one when 20% of the studies come from the  
 186    second software.

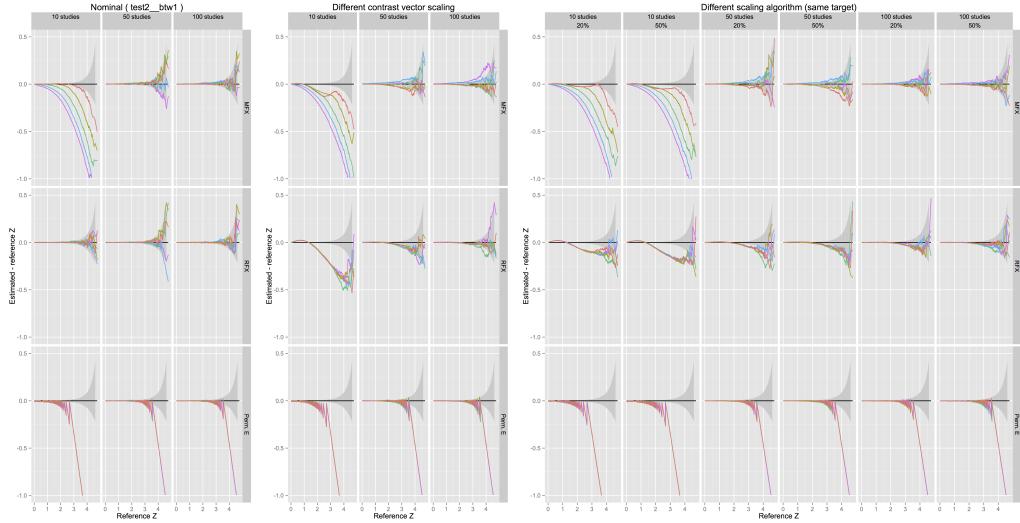


Figure 4: Deviation from theoretical Z in balanced two-sample tests with  $\tau^2 = 1$  and  $K = 25$  with matched (“nominal”) or mismatched (“different scaling target”, “different scaling algorithm”, “different contrast vector scaling”) units.

187     3.1.3. *Unbalanced between-group meta-analysis*

188     Fig. ?? presents the simulation results for unbalanced two-sample meta-  
 189     analyses with  $\tau^2 = 1$  and a sample size  $K = 25$ . For the nominal case,  
 190     MFX GLM, GLM RFX and contrast permutation provide valide estimate.  
 191     As expected due to the discrete nature of its ampling distribution, contrast  
 192     permutation is conservative for large Z value. GLM RFX is conservative.  
 193     RFX GLM is closest to the theoretical behaviour with Z-values that are  
 194     always within the 95% confidence interval.

195     In the extreme case of different scaling target, MFX GLM is always valid  
 196     but slightly conservative. RFX GLM is valid for Z values greater than 1.5  
 197     (area of interest in detections) but conservative. Similarly contrast permu-  
 198     tation is invalid for Z smaller than 1.5 and conservative otherwise. This can  
 199     be explained by the violation of the exchangeability condition.

200     When different scaling algorithm are used, (same paragraph as for one-  
 201     sample test)

202     When the contrast are scaled differently, we observe a very similar pattern  
 203     than for different scaling algorithm with higher varaince of the estimates.

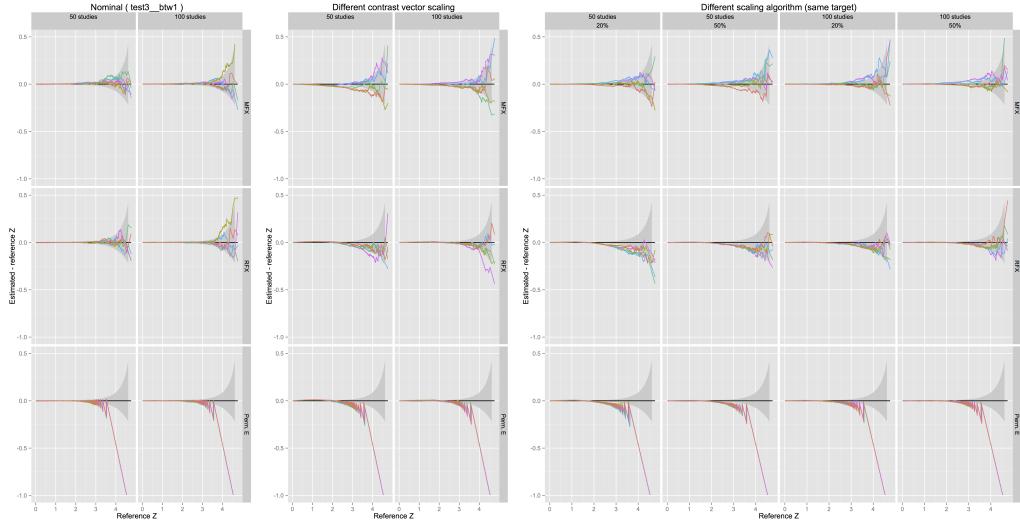


Figure 5: Deviation from theoretical Z in unbalanced two-sample tests with  $\tau^2 = 1$  and  $K = 25$  with matched (“nominal”) or mismatched (“different scaling target”, “different scaling algorithm”, “different contrast vector scaling”) units.

### 204 3.2. Simulations

205 Fig. 1 displays the false positive rate at  $p < 0.05$  obtained for the eight  
 206 estimators over all set of parameters in the absence and presence of between-  
 207 study variation. As expected, the fixed-effects meta-analytic summary statis-  
 208 tics, i.e. Fisher’s, Stouffer’s and weighted Stouffer’s estimates, are liberal in  
 209 the presence of study heterogeneity. The original Fisher’s approach is the  
 210 most invalid. More surprising, FFX GLM is also invalid with homogeneous  
 211 studies. The explanation is over-estimation of degrees-of-freedom (DF); while  
 212 DF is computed as  $(\sum n - 1)$ , under heteroscedasticity (from  $\sigma_i$  or  $n_i$ ) it  
 213 will be much lower [12]. Z MFX and GLM RFX provide valid estimates, and  
 214 the permutation estimates are valid but tend to be conservative with greater  
 215 variation in false positive rates.

216 The impact of the number of studies involved in the meta-analysis and of  
 217 the size of the within-study variance are investigated in Fig. ???. Permutation  
 218 inference is valid but conservative when 5 studies are used; this is because  
 219 there are only  $2^5 = 32$  possible permutations and thus  $1/32 = 0.03125$  is  
 220 largest attainable valid P-value. All approaches perform equally as soon as  
 221 10 or more studies are included in the meta-analysis.

Figure 6: Difference between the z-score estimated from each meta-analytic approach and the reference z-score from MFX GLM as a function of reference z-score.

222    *3.3. Real data*

223    The histogram of the ratio of between-subject variance to total variance is  
224    displayed in Fig. 2. From this graph it is clear that for most of the voxels the  
225    estimated between-study variance is greater than the within-study variance.  
226    We can therefore suppose the presence of study heterogeneity (non negligible  
227    between-study variance) in this collection of studies.

228    Fig. 6 plots the difference between the z-score estimated by each meta-  
229    analytic approach against the reference z-score computed with MFX GLM.  
230    All FFX statistics provide overly optimistic z-estimate suggesting, again, that  
231    study heterogeneity is present in the studied dataset. Among the RFX meta-  
232    analytic approaches, GLM RFX and contrast permutations provide z-scores  
233    estimate that are equal or smaller than the reference. Z permutation provides  
234    slightly larger z-scores between 1 and 3 (reference p-values between 0.16 and  
235    0.0013) but is mostly in agreement with the reference z-scores. On the other  
236    hand, Z MFX is more liberal than the reference for z-score ranging from 3  
237    to 5 (reference p-values between 0.0013 and 2.9e-07) and more stringent for  
238    z-scores smaller than 5.

239    **4. Conclusion**

240    We have compared eight meta-analytic approaches in the context of one-  
241    sample test. Through simulations, we found the expected invalidity of stan-  
242    dard FFX approaches in the presence of study heterogeneity, but also of  
243    FFX GLM even with no between-study variation. In a real dataset of 21  
244    studies of pain, there was evidence for substantial between-study variation  
245    that supports the use of RFX meta-analytic statistics. When only contrast  
246    estimates are available, RFX GLM was valid. This is in line with previous  
247    results on within-group one-sample t-tests studies [8]. When only standard-  
248    ised estimates are available, permutation is the preferred option as the one  
249    providing the most faithful results. Further investigations are needed in order  
250    to assess the behaviour of these estimators in other configurations, including  
251    meta-analyses focusing on between-study differences.

252    5. Acknowledgements

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

- 255    [1] Meta-analysis of neuroimaging data: a comparison of image-based and  
256    coordinate-based pooling of studies. *NeuroImage*, 45(3):810–23, 2009.
- 257    [2] Katherine S Button, John P a Ioannidis, Claire Mokrysz, Brian a Nosek,  
258    Jonathan Flint, Emma S J Robinson, and Marcus R Munafò. Power  
259    failure: why small sample size undermines the reliability of neuroscience.  
260    *Nature reviews. Neuroscience*, 14(5):365–76, 2013.
- 261    [3] G. Chen, Z. S. Saad, A. R. Nath, M. S. Beauchamp, and R. W. Cox.  
262    FMRI group analysis combining effect estimates and their variances.  
263    *NeuroImage*, 60(1):747–65, March 2012.
- 264    [4] P. Cummings. Meta-analysis based on standardized effects is unreliable.  
265    *Archives of pediatrics & adolescent medicine*, 158(6):595–7, 2004.
- 266    [5] R.A. Fisher. *Statistical Methods for Research Workers*. Oliver and Boyd,  
267    Edinburgh, 1932.
- 268    [6] A. P. Holmes, R. C. Blair, G. J.D. Watson, and I. Ford. Nonparamet-  
269    ric Analysis of Statistic Images from Functional Mapping Experiments.  
270    *Journal of Cerebral Blood Flow and Metabolism*, 16:7–22, 1996.
- 271    [7] T. Liptak. On the combination of independent tests. *Magyar Tud. Akad.*  
272    *Mat. Kutato Int. Kozl.*, 3:171–197, 1958.
- 273    [8] J. A. Mumford and T. E. Nichols. Simple group fMRI modeling and  
274    inference. *NeuroImage*, 47(4):1469–75, 2009.
- 275    [9] T. E. Nichols and A. P Holmes. Nonparametric permutation tests for  
276    functional neuroimaging: a primer with examples. *Human brain map-*  
277    *ping*, 15(1):1–25, 2002.
- 278    [10] Thomas E. Nichols. Spm plot units, 2012.
- 279    [11] J. Radua and D. Mataix-Cols. Meta-analytic methods for neuroimaging  
280    data explained. *Biology of mood & anxiety disorders*, 2(1):6, 2012.

- 281 [12] F. E. Satterthwaite. An approximate distribution of estimates of vari-  
282 ance components. *Biometrics Bulletin*, 2:110–114, 1946.
- 283 [13] S. Smith, P. R. Bannister, C. Beckmann, M. Brady, S. Glare, D. Flitney,  
284 P. Hansen, M. Jenkinson, D. Leibovici, B. Ripley, M. Woolrich, and  
285 Y. Zhang. FSL : New Tools for Functional and Structural Brain Image  
286 Analysis. (6):2001, 2001.
- 287 [14] S. Stouffer, L. DeVinney, and E. Suchmen. *The American Soldier:*  
288 *Adjustment During Army Life*, volume 1. Princeton University Press,  
289 Princeton, NJ, 1949.
- 290 [15] D V Zaykin. Optimally weighted Z-test is a powerful method for com-  
291 bining probabilities in meta-analysis. *Journal of evolutionary biology*,  
292 24(8):1836–41, 2011.