Is Z enough? Impact of Meta-Analysis using only Z/T images in lieu of estimates and standard errors

Camille Maumet1, Gholamreza Salimi-Khorshidi2, Thomas E. Nichols1,3

1. University of Warwick, Warwick Manufacturing Group, Coventry, UK
2. FMRIB Centre, University of Oxford, Oxford, UK
3. University of Warwick, Dept. of Statistics, Coventry, UK

Introduction

While most neuroimaging meta-analyses are based on peak coordinate data, the best practice method is an Intensity-Based Meta-Analysis (IBMA) that combines the effect estimates and their standard errors (E+SE’s) [5].

![Intuition Based Meta-Analysis](image)

**Fig. 1:** Coorinde-based and Intensity-Based Meta-analysis.

There are various efforts underway to facilitate sharing of neuroimaging data to make such IBMA’s possible (see, e.g. [2]), but the emphasis is usually on sharing T-statistics. However, guidelines for (non-)imaging meta-analyses are clear that T-statistics-based meta-analysis is suboptimal and is to be discouraged [1]. But even if E+SE’s are shared, the units must be equivalent, and different software, models or contrasts can lead to incompatible units.

Here we compare the use of IMBA using only T-statistics to use of E+SE’s.

Theory

<table>
<thead>
<tr>
<th>Meta-analysis statistic</th>
<th>Nominal H0 dist.</th>
<th>Input Assumptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLAME MFX</td>
<td>$\sum_{i=1}^{n} \frac{Z_i}{\sqrt{1/2 + \tau^2}}$</td>
<td>Y, SE, IGE; $\tau^2 = s^2$</td>
</tr>
<tr>
<td>FLAME FFX</td>
<td>$\sum_{i=1}^{n} \frac{Z_i}{\sqrt{1/2 + \tau^2}}$</td>
<td>Y, SE; $\tau^2 = 0$</td>
</tr>
<tr>
<td>RFX GLM</td>
<td>$\sum_{i=1}^{n} \frac{Z_i}{\sqrt{\tau^2}}$</td>
<td>Empirical; Y, ISE</td>
</tr>
<tr>
<td>Contrast Perm</td>
<td>$\sum_{i=1}^{n} \frac{Z_i}{\sqrt{\tau^2}}$</td>
<td>Empirical; Z, ISE</td>
</tr>
<tr>
<td>Stouffer</td>
<td>$\sum_{i=1}^{n} \frac{Z_i}{\sqrt{\tau^2}}$</td>
<td>Y, IGE; $\tau^2 = 0$</td>
</tr>
<tr>
<td>Z MFX</td>
<td>$\sum_{i=1}^{n} \frac{Z_i}{\sqrt{\tau^2}}$</td>
<td>Z, IGE; $\tau^2 = 0$</td>
</tr>
<tr>
<td>Z Perm</td>
<td>$\sum_{i=1}^{n} \frac{Z_i}{\sqrt{\tau^2}}$</td>
<td>Z, ISE</td>
</tr>
<tr>
<td>Weighted-Z</td>
<td>$\sum_{i=1}^{n} \frac{Z_i}{\sqrt{1/2 + \tau^2}}$</td>
<td>Y, SE, IGE; $\tau^2 = s^2$</td>
</tr>
</tbody>
</table>

Table 1. Statistics for one-sample meta-analysis tests and their sampling distributions under the null hypothesis. IGE=Independent Gaussian Errors; ISE=Independent Symmetric Errors; for a study i: Y is the contrast estimate (E); SE; the contrast variance estimate (SE); $\tau^2$; the contrast variance; $\tau^2$ denotes the between-study variance; $\tau^2$ is the combined within-and-between-study variance.

Our reference approach is an IBMA based on a 3-level hierarchical model: level 1, subject FFx; level 2, study FFx; level 3: meta-analysis FFx (FLAME MFX) or FFx (FLAME FFX), using FSL’s FLAME method [6].

In the absence of E+SE’s, there are a number of methods to combine Z-scores [3]. We focused on three of them: Stouffer’s method [7], Weighted-Z [8,4], Z MFX [5] and Z Permutation.

We also investigated two alternative approaches using only the E’s: Random-Effects GLM (RFX GLM) and Contrast Permutation.

Experiments

First, we compared the Fixed-Effects (FFx) approaches. As results are usually presented as a thresholdmap, 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. Then, 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.

Second, we compared the z-scores obtained with the 7 meta-analyses approaches described in Table 1 to the reference FLAME MFX.

Results

Among fixed-effects meta-analytic methods, the weighted-Z approach demonstrated slightly better results than Stouffer’s as shown by the ROC curve in Fig. 2 and the dice similarity scores in Table 2. Unsurprisingly, Fixed-effects meta-analytic estimators seems overly liberal according to Fig. 3 advocating for the use of Random-Effects approaches. GLM RFX, Z Permutation and Contrast Permutation provide valid statistics.

Conclusion

We have compared seven meta-analytic approaches in the context of one-sample test. When only contrast estimates are available, RFX GLM was valid, closest to FLAME MFX reference. When only standardised estimates (i.e., $Z/\tau$) are available, permutation is the preferred option as the one providing the most faithful results. Further investigations are needed in order to assess the behaviour of these estimators in other configurations, including meta-analyses focusing on between-study differences.

Acknowledgments

We gratefully acknowledge the use of this data from the Tracey pain group, FMRIB, Oxford. We also acknowledge The Wellcome Trust for support of CM & TEN.

References


Fig. 2: ROC curves (a) for Stouffer’s (blue) and Weighted-Z (red). Ground truth (black). Table 2: Dice similarity score (c) with FLAME for three uncorrected thresholds.