Do the units matter? Validity of Intensity Based Meta-Analysis in the presence of unit mismatches
Camille Maumet, Thomas Nichols

To cite this version:
Camille Maumet, Thomas Nichols. Do the units matter? Validity of Intensity Based Meta-Analysis in the presence of unit mismatches. 21st Annual Meeting of the Organization for Human Brain Mapping (OHBM), Jun 2015, Honolulu, United States. 2015. <inserm-01149475>

HAL Id: inserm-01149475
http://www.hal.inserm.fr/inserm-01149475
Submitted on 24 Aug 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Do the units matter? Validity of Intensity Based Meta-Analysis in the presence of unit mismatches

Camille Maumet1, Thomas E. Nichols1,2
1. University of Warwick, Warwick Manufacturing Group, Coventry, UK
2. 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 that combines the voxel-wise effect estimates and their standard errors (E + SE’s) from each study [5].

Such analysis assumes that the E + SEs are expressed in the same units across studies. In practice the observed units of the E + SE’s will depend on:
1. the method used to scale raw T2* data;
2. the scaling of the explanatory variables in the design matrix;
3. the scaling of the contrast vector [4].

However those meta-data are rarely shared, leaving the researcher with E + SE’s with unknown scaling.

Here, we investigate the validity of 5 random-effects meta-analytic approaches in the presence of imperfectly scaled contrast estimates.

Methods

We studied 5 meta-analytic approaches:

<table>
<thead>
<tr>
<th>Meta-analysis statistic</th>
<th>Nominal Hδ distrib.</th>
<th>Inputs</th>
<th>Assumptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFX (Σi=1 to n(yi – 1/Σj=1 to n(yj)) 2 ) 1/√(Σi=1 to n(yi – 1/Σj=1 to n(yj)) 2 )</td>
<td>Perm. E</td>
<td>Y, S</td>
<td>IGE; τ² = σ²</td>
</tr>
<tr>
<td>RFX (Σi=1 to n(yi – 1/Σj=1 to n(yj)) 2 ) 1/√(Σi=1 to n(yi – 1/Σj=1 to n(yj)) 2 )</td>
<td>Perm. Z</td>
<td>Y, S</td>
<td>ISE</td>
</tr>
<tr>
<td>Perm. E (Σi=1 to n(yi – 1/Σj=1 to n(yj)) 2 ) 1/√(Σi=1 to n(yi – 1/Σj=1 to n(yj)) 2 )</td>
<td>Empirical</td>
<td>Y</td>
<td>ISE</td>
</tr>
<tr>
<td>Stouffer MFX (Σi=1 to n(yi – 1/Σj=1 to n(yj)) 2 ) 1/√(Σi=1 to n(yi – 1/Σj=1 to n(yj)) 2 )</td>
<td>Empirical</td>
<td>Z</td>
<td>IGE; 1 + τ²/σ²</td>
</tr>
</tbody>
</table>

To check the validity of each estimator under the null hypothesis we plotted the observed difference and expected test statistic vs. expected, after converting all observed statistics to equivalent z-scores (via their P-values).

For each meta-analysis we simulated a given number of study-level E + SE’s under the null hypothesis. Simulated data were generated assuming: 25 studies in total; 20 subjects per study; a between-study variance of 1 and; a within-study variance of 20[0.25; 1; 4]. We studied one-sample t-tests.

In a first set of experiments, we focused on variations in the units that are observed across neuroimaging software packages (due to different data scaling strategies). We considered that 50% or 80% of the studies were generated by different software, for each introducing:

- A factor of 100 to simulate variations due to the use of a different baseline (100 for SPM, 10000 for FSL).
- A factor of 2 to simulate variations due to different data scaling algorithms [3].

In a second set of experiments, we focused on units variations that are observed when the design matrix or the contrast vector are scaled differently across studies. We introduced a different scaling factor for each study ranging from 0.4 to 1.6.

Results

Fig. 2 Difference between the observed equivalent Z-statistic and the expected Z-statistic under the null hypothesis for a one-sample analysis in the presence of unit mismatch due to different BOLD scaling algorithm (a), different target intensity for BOLD scaling (b) or different scaling of the contrast vectors (c).

In a one-sample analysis, unit mismatch induced by...

- Different baseline across neuroimaging software (Fig. 2.a) has drastic effects on both MFX and RFX meta-analysis. While RFX analyses becomes overly conservative, MFX analyses are conservative or invalid depending on the within-subject variance. The smaller the within-subject variance the more important the invalidity. Perm. E is also slightly invalid.
- Different data scaling algorithm (fig. 2.b) induces some conservativeness for both MFX and RFX. Perm. E is valid.
- Differences in scaling of the contrast vectors (fig. 2.c) induces conservativeness for RFX and conservativeness or invalidity for MFX (depending on within-study variance as was observed for differences in baselines). Perm. E is valid.

Conclusion

When there is uncertainty of data scaling, our evaluations have shown that, for one-sample tests, permutation is a good alternative: both RFX and MFX meta-analysis can lead to conservativeness (and even invalidity for MFX) in the presence of improperly scaled E+SE’s. Another alternative would be to use the Z-based meta-analytic approaches which are not sensitive to unit differences [2].

We acknowledge the Wellcome Trust for support of CM & TEN.

References