Do the units matter? Validity of Intensity Based Meta-Analysis in the presence of unit mismatches

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

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].

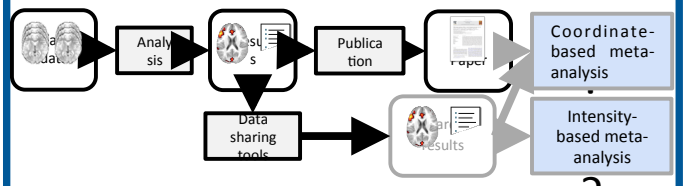

Fig. 1: Coordinate-based and Intensity-Based Meta-analysis.

Such analysis assumes that the **E + SE's** 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**:

	Meta-analysis statistic	Nominal H_0 distrib.	Inputs	Assumptions
MFX	$\left(\sum_{i=1}^{k} \frac{Y_i}{S_i^2+\hat{\tau}^2}\right) / \sqrt{\sum_{i=1}^{k} 1/(S_i^2+\hat{\tau}^2)}$	$\mathcal{T}_{k-1}$	Y_i, S_i^2	IGE; $\tau^2 = \hat{\tau}^2$.
RFX	$\left(\sum_{i=1}^{k} \frac{Y_i}{\sqrt{k}}\right) / \hat{\sigma}_C$	$\mathcal{T}_{k-1}$	Y_i	IGE; $\tau^2 + \sigma_i^2$ cst.
Perm. E	$\sum_{i=1}^{k} \frac{Y_i}{\sqrt{k}}$	Empirical	Y_i	ISE.
Perm. Z	$\left(\sum_{i=1}^{k} Z_i\right) / \sqrt{k}$	Empirical	Z_i	ISE.
Stouffer MFX	$\left(\sum_{i=1}^{k} Z_i\right) / \sqrt{k}\hat{\sigma}_Z$	$\mathcal{T}_{k-1}$	Z_i	IGE; $1 + \tau^2/\sigma_i^2$ cst.

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_i is the contrast estimate (E); S_i^2 the contrast variance estimate (SE^2), the contrast variance, $\hat{\sigma}_C^2$ denotes the between-study variance; is the combined within and between-study variance in data/contrast units, is same in Z units.

To check the **validity** of each estimator **under the null hypothesis** we plotted the difference observed 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.

References

[1] Fisher (1932) Statistical Methods for Research Workers. [2] Maumet et al. (2014) 20th annual meeting of OHBM. [3] http://blogs.warwick.ac.uk/nichols/entry/spm_plot_units. [4] Pernet (2014). Frontiers in Neuroscience, 8, 1–12. [5] Salimi-khorshidi et al. (2009) NeuroImage, 45(3):810–23. [6] Stouffer et al. (1949). Princeton University Press, Princeton, NJ. [7] Zaykin (2011) Journal of evolutionary biology, 24(8), 1836–41.

Results

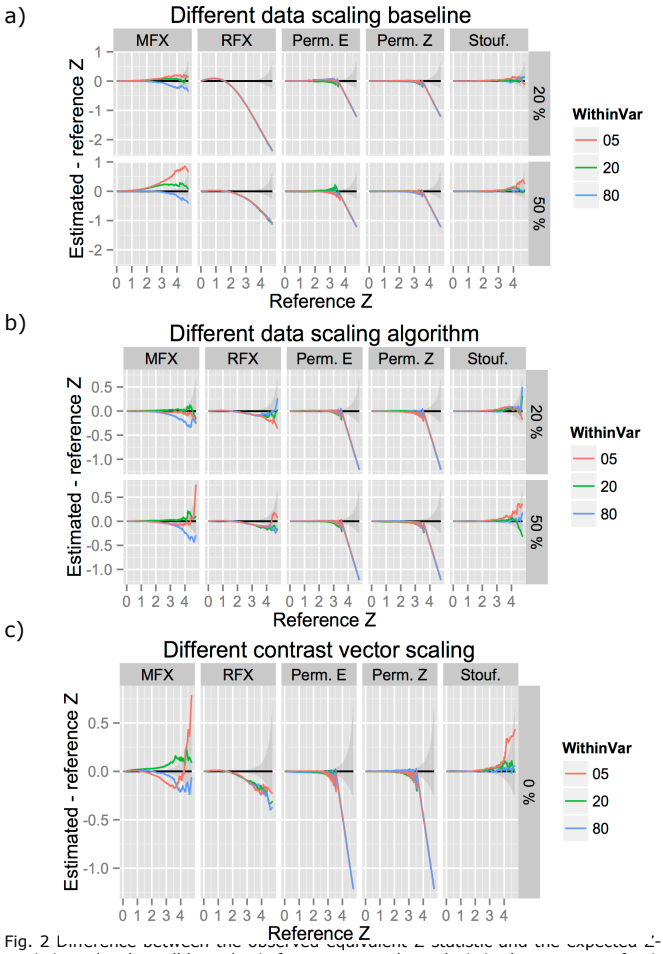

Fig. 2 Difference between the observed equivalent Z statistic and the expected Z-statistic under the null hypothesis for a one-sample analysis in the presence of unit mismatch due to different BOLD scaling algorithm (a), different target intensity for BOLD scaling (b) or different scaling of the contrast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