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

**Camille Maumet[1], Gholamreza Salimi-Khorshidi[2], Thomas E. Nichols[1,3]**

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].

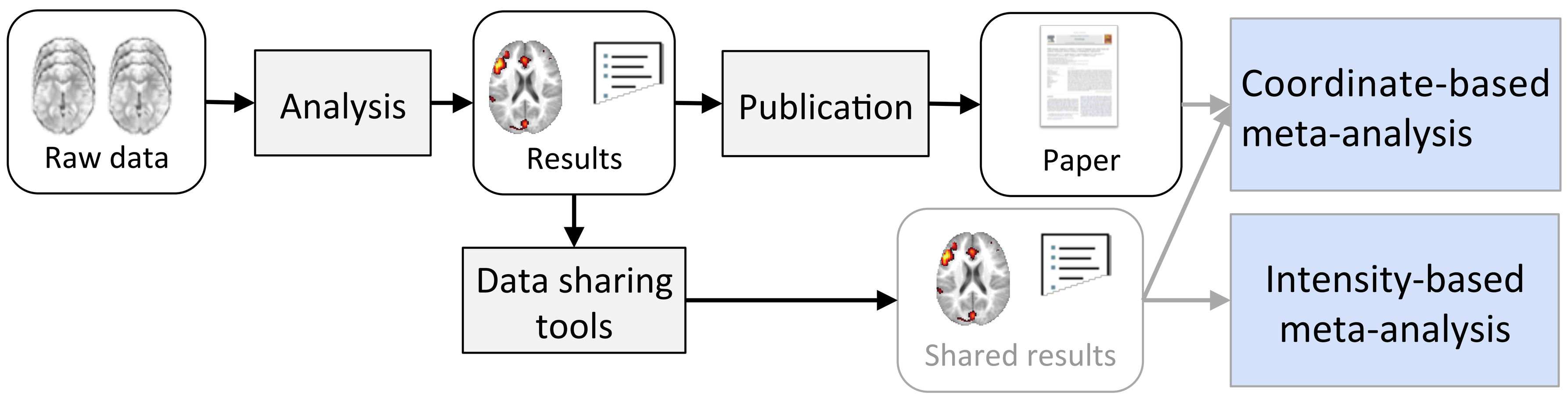



Fig. 1: Coordinate-based and Intensity-Based Meta-analysis.

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-analysis are clear that T-statistic-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

| | Meta-analysis statistic | Nominal $H_0$ distrib. | Inputs | Assumptions |
|---|---|---|---|---|
| FLAME 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; $\tau^2=\hat{\tau}^2$. |
| FLAME FFX | $\left(\sum_{i=1}^k \frac{Y_i}{S_i^2}\right)/\sqrt{\sum_{i=1}^k 1/S_i^2}$ | $\mathcal{T}_{(\sum_{i=1}^k n_i-1)-1}$ | $Y_i, S_i^2$ | IGE; $\tau^2=0$; $\sigma_i^2$ cst. |
| RFX GLM | $\left(\sum_{i=1}^k \frac{Y_i}{\sqrt{k}}\right)/\hat{\sigma}_C^2$ | $\mathcal{T}_{k-1}$ | $Y_i$ | IGE; $\tau^2+\sigma_i^2$ cst. |
| Contrast Perm. | $\sum_{i=1}^k \frac{Y_i}{\sqrt{k}}$ | Empirical | $Y_i$ | ISE. |
| Stouffer's | $\left(\sum_{i=1}^k Z_i\right)/\sqrt{k}$ | $\mathcal{N}(0,1)$ | $Z_i$ | IGE; $\tau^2=0$. |
| Z MFX | $\left(\sum_{i=1}^k Z_i\right)/\sqrt{k}\hat{\sigma}$ | $\mathcal{T}_{k-1}$ | $Z_i$ | IGE; $1+\tau^2/\sigma_i^2$ cst. |
| Z Perm. | $\left(\sum_{i=1}^k Z_i\right)/\sqrt{k}$ | Empirical | $Z_i$ | ISE. |
| Weighted-Z | $\left(\sum_{i=1}^k \sqrt{n_i}Z_i\right)/\sqrt{\sum_{i=1}^k n_i}$ | $\mathcal{N}(0,1)$ | $Z_i, n_i$ | IGE; $\tau^2=0$. |

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 ($SE^2$), $\sigma^2_i$ the contrast variance; $\tau^2$ denotes the between-study variance; $\sigma^2_C$ 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 MFX; level 3: meta-analysis MFX (**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 **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. 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.

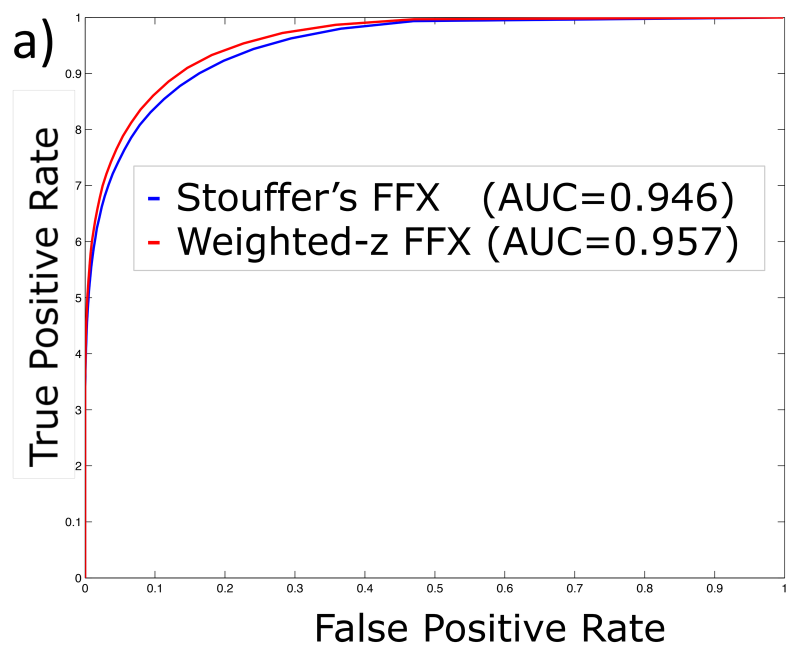
Second, we **compared the z-scores** obtained with the 7 meta-analyses approaches described in table 1 to the reference FLAME MFX.

## References

[1] Cummings (2004) Archives of pediatrics & adolescent medicine, 158(6), 595–7. [2] Gorgolewski et al. (2013) 19th Annual Meeting of the OHBM. [3] Lazar et al. (2002) NeuroImage, 16(2), 538–50. [4] Liptàk (1958) Magyar Tud. Akad. Mat. Kutato Int. Kozl 3: 171–196. [5] Salimi-khorshidi et al. (2009) NeuroImage, 45(3):810–23. [6] Smith et al. (2001) NeuroImage 13 (6), 249. [7] Stouffer et al. (1949) Princeton University Press, Princeton, NJ. [8] Zaykin (2011) Journal of evolutionary biology, 24(8), 1836–41.

## Fixed-Effects Meta-analyses approaches

a)




b)

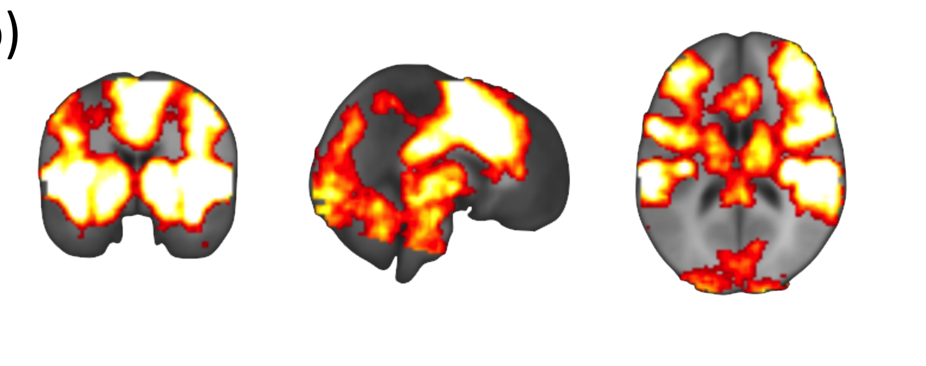


Fig. 2: ROC curves (a) for Stouffer's (blue) and Weighted-Z (red). Ground truth detections (b).

| | $p < 0.001$ | $p < 0.01$ | $p < 0.05$ |
|---|---|---|---|
| Stouffer's | 0.84 | 0.87 | 0.89 |
| Weighted-Z | 0.86 | 0.88 | 0.90 |

Table 2: Dice similarity score (c) with FLAME for three uncorrected thresholds.

## Fixed-Effects and Mixed-Effects approaches

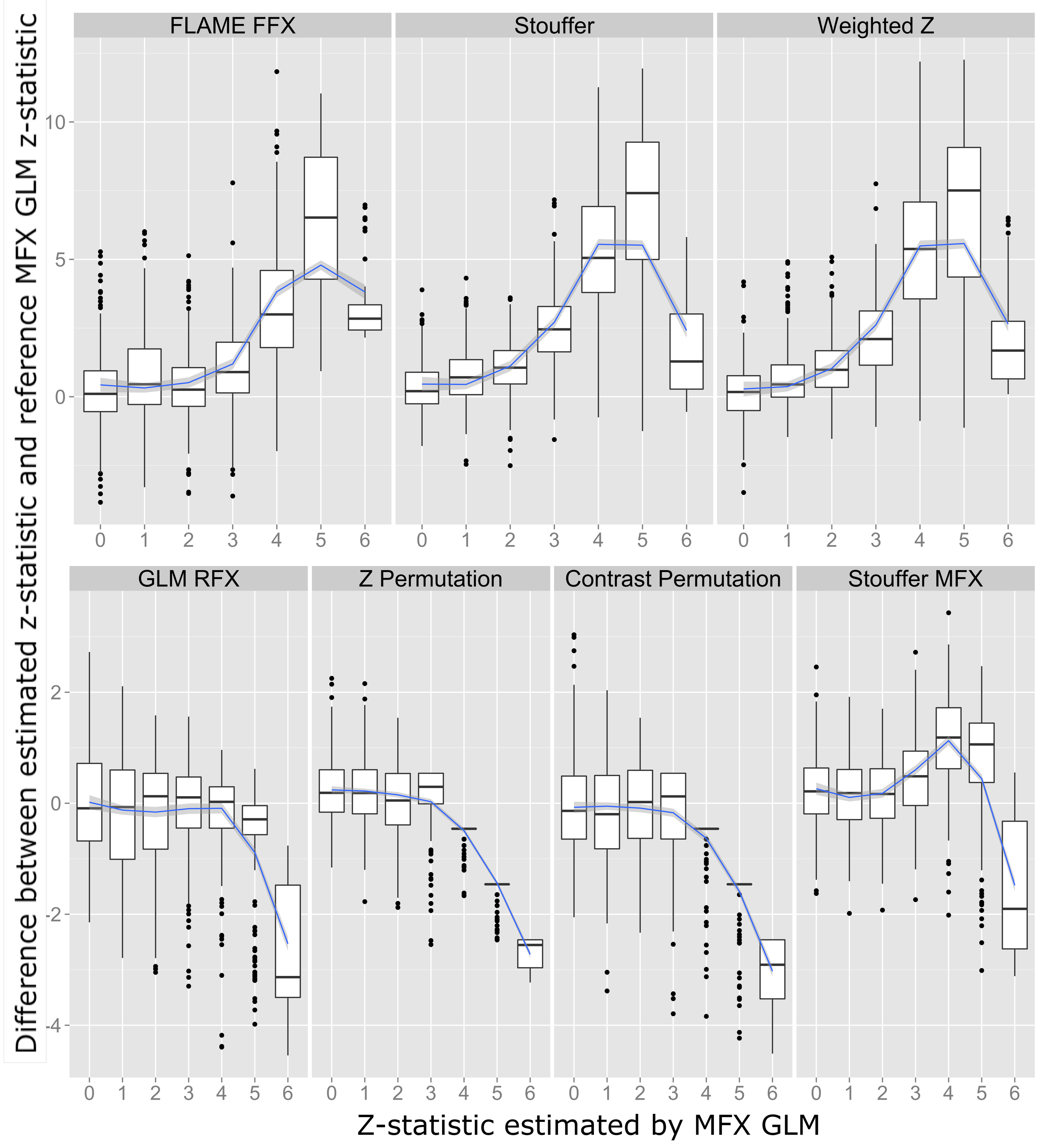



Fig 3: Difference between the z-score estimated from each meta-analytic approach and the reference z-score from MFX GLM as a function of reference z-score.

## 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/T's) 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.