

# Validity of summary statistics-based mixed-effects group fMRI

Camille Maumet<sup>1</sup> and Thomas E. Nichols<sup>2</sup>

1. University of Rennes, Inria, CNRS, Inserm, IRISA, Rennes, France.

2. Oxford Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Population Health, University of Oxford, UK.

## Introduction

Statistical analysis of multi-subject functional Magnetic Resonance Imaging (fMRI) data can be represented as a mixed-effect general linear model (GLM) and is traditionally solved using either:

1. **Feasible Generalised Least Squares (FGLS)** (sometimes referred to as mixed-effects GLM) where within-subject variance estimates are used and incorporated into per-subject weights or
2. **Ordinary least squares (OLS)** (sometimes referred to as random-effects GLM) where within-subject variance estimates are not used.

Both approaches are implemented and available in major neuroimaging software packages including: SPM (MFX analysis; 2nd-Level statistics), FSL (FLAME; OLS) and AFNI (3dMEMA; 3dttest++).

While **FGLS** provides the most efficient statistical estimate, its **properties are only guaranteed in large samples**, and it has been shown that OLS is a valid alternative for one-sample group analyses in fMRI [1].

We recently showed that FGLS for image-based meta-analysis could lead to invalid results in small-samples. As FGLS meta-analyses uses the same model, here we investigate, whether this issue also affects group fMRI.

	Group statistic	Assumptions
<b>FGLS</b>	$\left(\sum \kappa_i \hat{\beta}_i\right) / \sqrt{\sum_{i=1}^k \kappa_i}$ with $\kappa_i = 1/(\hat{\tau}^2 + s_i^2)$	IGE, large sample.
<b>OLS</b>	$\left(\sum_{i=1}^k \frac{\hat{\beta}_i}{\sqrt{k}}\right) / \hat{\sigma}_C^2$	IGE; $\tau^2 + \sigma_i^2 = \sigma_C^2 \forall i$

**Table 1.** Statistics for group fMRI tests and their underlying assumptions.

$\tau^2$ : pure between-study variance,  $\sigma_i^2$ : *i*th study's contrast variance,  $s_i^2$ : *i*th study's sampling contrast variance,  $\sigma_C^2$ : usual one-sample variance. IGE=Independent Gaussian Errors.

## Methods

GLM can be expressed with:  $Y = X\beta + \epsilon$ , where  $Y$  is the  $N$ -vector of subject-level contrast estimates,  $X$  the design matrix,  $\beta$  the group parameter to estimate and  $\epsilon$  the random error. In group fMRI, the error term has two contributions, from within- and between-subject variance.

**FGLS.** Using within-subject variance estimates requires a weighted least squares (WLS) approach, where the group parameter  $\beta$  is a weighted average of the subject-level contrasts. The weights are inversely proportional to the sum of the within- and between-subject variances.

But in practice, those weights are unknown and have to be estimated from the data leading to a Feasible Generalised Least Squares (FGLS). FGLS is asymptotically efficient but its finite sample properties are unknown [2]. We used FSL's 'FLAME 1' FGLS that uses maximum likelihood to estimate between-subject variance, computing a T-statistic compared to a Student distribution with  $N-1$  degrees of freedom (DF) [3].

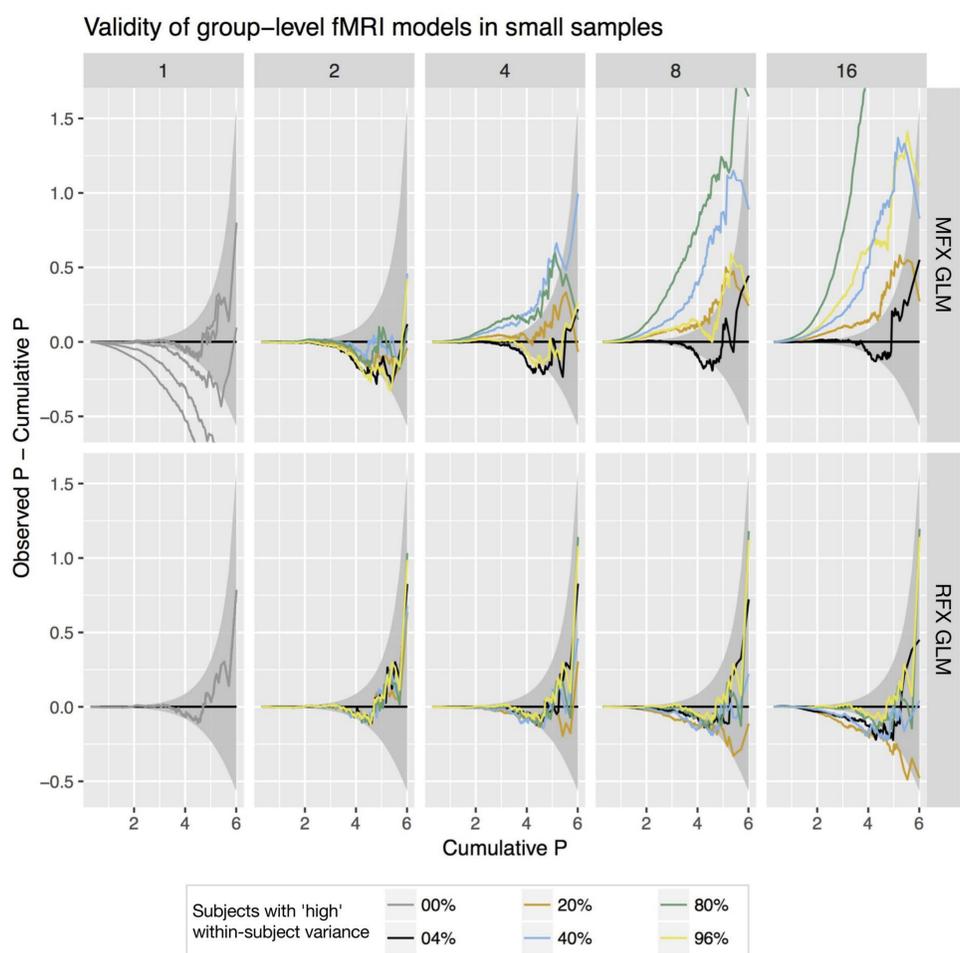
**OLS.** Under the assumption that the within-subject variance is constant or negligible in comparison to the between-study variance, the weights above are equal and the GLM can be estimated with Ordinary Least Squares (OLS),  $\beta$  estimated as the average of the subject-level contrasts. We used SPM's 2nd level one-sample model, computing a T statistic also compared to a Student distribution with  $N-1$  DF [4]. OLS p-values are exact for any sample size, in contrast to FGLS which are only asymptotically valid [2].

We used **Monte Carlo simulations** to investigate the validity of MFX and RFX GLMs under varying degrees of within-subject variance heteroscedasticity.

**Within-subject variances** took on 2 values, a 'good' value and a 'high', where the 'high' value was one of 2, 4, 8 & 16x good values; we considered 4%, 20%, 40%, 80% or 96% of the subjects to have the high values.

We fixed the mean within-subject standard error to be equal to the between-subject variance. We assumed 25 subjects per group and 1000 independent time points per subject.

Accuracy was assessed by comparing FSL & SPM distributions of  $-\log_{10}$  P-values to Monte Carlo  $-\log_{10}$  P-values based on  $10^6$  realisations.



**Fig. 1.** Deviation of observed from theoretical P-values (difference of observed and Monte Carlo ('true')  $-\log_{10}$  p-value distributions) for one-sample tests in the presence of varying percentages of subjects with outlying within-subject variances, high-variance factor 2, 4, 8 or 16, (columns), MFX GLM and RFX GLM (rows). Y-axis is the observed cumulative probability minus Monte Carlo cumulative probability for a given (X-axis)  $-\log_{10}$  p-value; positive deflections correspond to inflated false positive risk.

## Results

Fig. 1 presents deviation from theoretical P-values with varying percentage and intensity of high intra-subject variance. For low intensity heteroscedasticity ( $\leq 2x$ ), FGLS GLM is valid but becomes increasingly invalid in the presence strong and prevalent high variance subjects. OLS GLM is valid with all settings but displays some conservativeness in the presence of strong heteroscedasticity.

## References

- [1] Mumford, Neuroimage. 2009. [2] Greene 2003. [3] Woolrich, Neuroimage. 2004. [4] Penny 2011.

## Acknowledgments

This work was supported by the Wellcome Trust. We gratefully acknowledge the use of the pain dataset from the Tracey pain group, FMRIB, Oxford. The majority of the work was conducted while TEN and CM were at the University of Warwick and used the High Performance Computing cluster of the Department of Statistics, University of Warwick.