

Same data meta-analysis

Enter the multiverse...

Lefort-Besnard Jeremy¹, Nichols Thomas^{2*}, Maumet Camille^{1*}

¹ Inria, Univ Rennes, CNRS, Inserm, IRISA UMR 6074, Empenn ERL U 1228, Rennes, France

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

* Sharing last authorship

Problem

Researchers using fMRI data have a wide range of analysis tools to model brain activity. This diversity of analytical approaches means there are many possible variations of the same imaging result (Bowring et al., 2019). Analyzing a dataset with a single approach can thus be misleading.

Alternatively, a multiverse analysis can be used, where multiple sets of results are obtained from running different pipelines on the same single dataset. Such a setting produces several test statistic maps. Meta-analysis approaches can then help to effectively extract valid and robust results from these maps.

A required assumption for traditional meta-analyses is the independence among input datasets (Fig 1a). This assumption is no longer true in multiverse setting (Fig 1b), thus treating dependent studies as independent may lead to invalidity and thus decrease the generalizability of the result.

Therefore, we present a variety of same data meta-analysis (SDMA) models. The validity and accuracy of these models were assessed in a set of different simulations as well as on a real world dataset from "NARPS", a multiverse analysis with 70 different statistic maps originating from the same data. We also contrasted the outcomes of these SDMA models with those of a conventional meta-analysis approach.

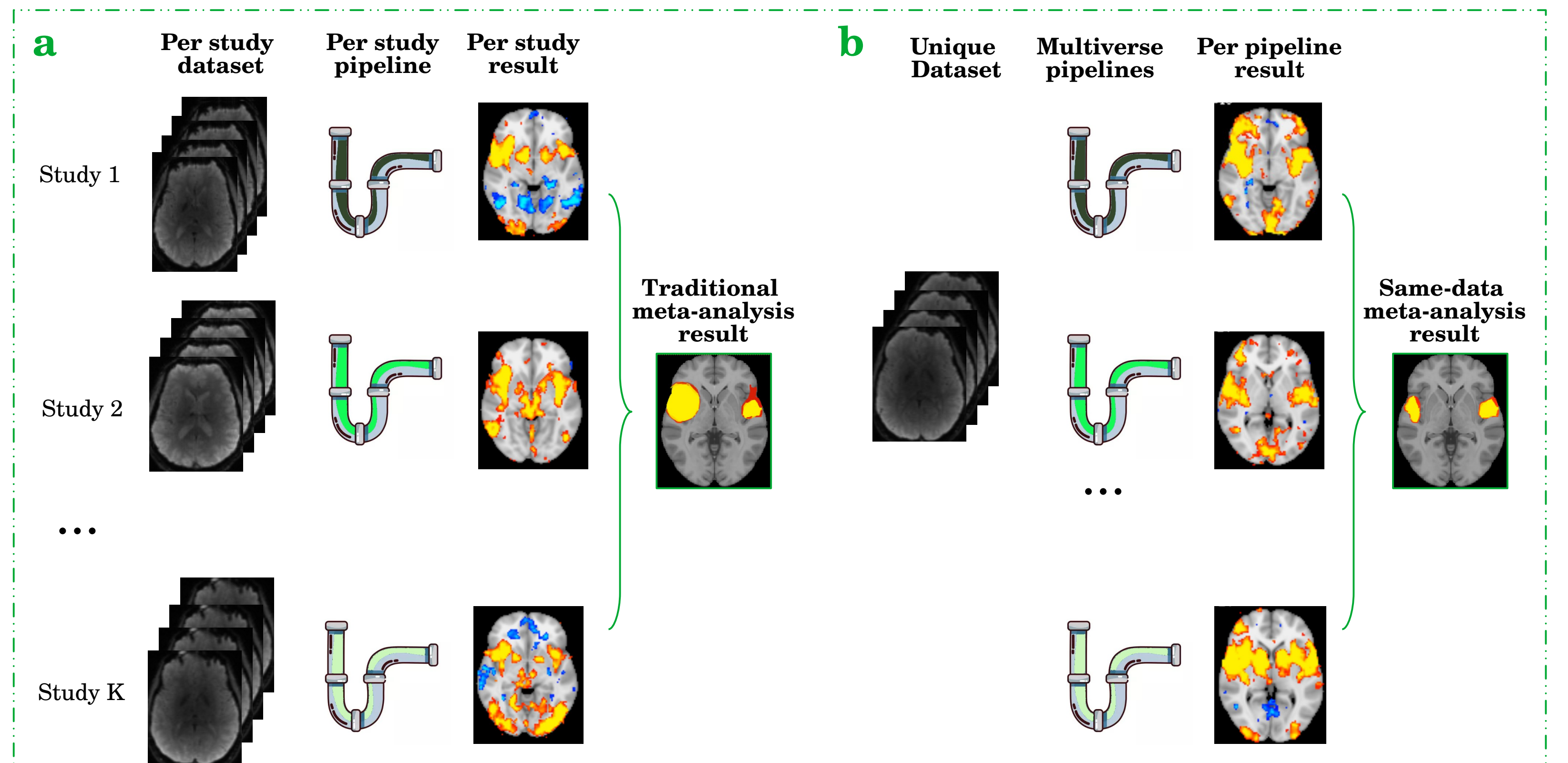


Fig 1: Traditional meta-analyses setting (a) and multiverse setting (b). The traditional approach assumes independence among images (K studies and thus K pipelines computed with N different participants each) when estimating summary effect. This underlying assumption is undermined in the multiverse setting where the outcomes are dependent over pipelines (One unique set of participants data for each pipeline).

Traditional and SDMA models

Average $\bar{Y}_j = \frac{1}{K} \sum_k Y_{kj}$	Dependence-Corrected Stouffer $\bar{Y}_j^{SC} = \frac{\bar{Y}_j}{\sqrt{1^T Q 1 / K^2}}$	GLS Stouffer $\bar{Y}_j^{SG} = \frac{\bar{Y}_j^G}{\sqrt{1^T Q^{-1} 1}^{-1}}$	Consensus Stouffer $\bar{Y}_j^{CS} = \frac{\bar{Y}_j - \mu_C}{\sqrt{1^T Q 1 / K^2}} + \mu_C$	Consensus Weighted Stouffer $\bar{Y}_j^{CWS} = \frac{\bar{Y}_j^*}{\sqrt{1^T Q 1 / K^2}} \sigma_C + \mu_C$	Consensus GLS Stouffer $\bar{Y}_j^{CSG} = \frac{\bar{Y}_j^G - \mu_C}{\sqrt{1^T Q^{-1} 1}^{-1}} + \mu_C$	Consensus Average $\bar{Y}_j^{CA} = \frac{\bar{Y}_j - \mu_C}{\sigma_Y^2} + \mu_C$
--	--	---	---	--	--	--

Stouffer
 $\bar{Y}_j^S = \sqrt{K} \bar{Y}_j$

With Y_{kj} being the test statistic for pipeline $k = 1, \dots, K$ and voxel $j = 1, \dots, J$; Q the interpipeline correlation matrix; μ_C and σ_C^2 respectively the average of the voxel-wise mean and the average of the voxel-wise variance. σ_Y^2 the variance of the average. \bar{Y}_j^G is defined as $1^T Q^{-1} Y_j$, down-weighting the influence of highly dependent pipeline. * means that data were standardized. Average and Stouffer methods are traditional meta-analysis models.

Model evaluation in generated data

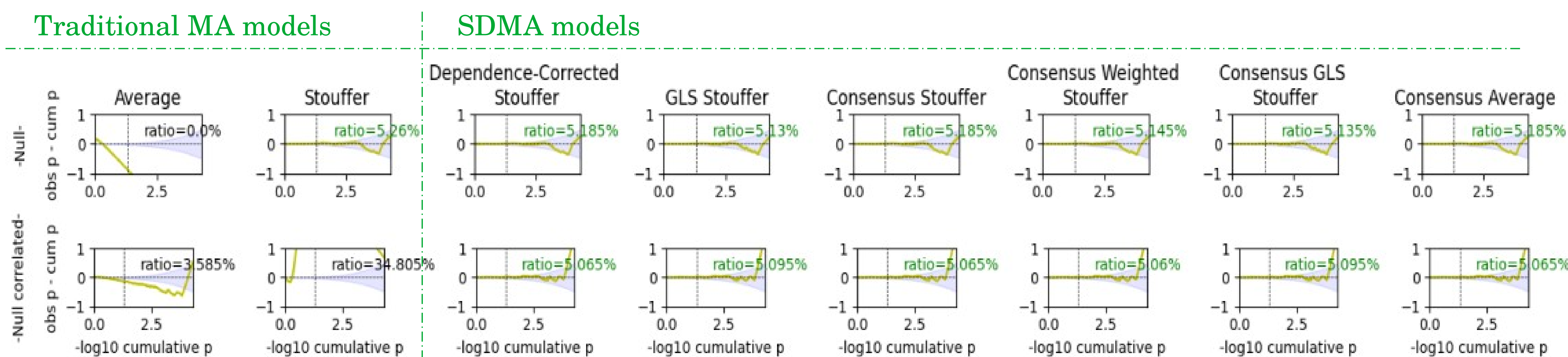


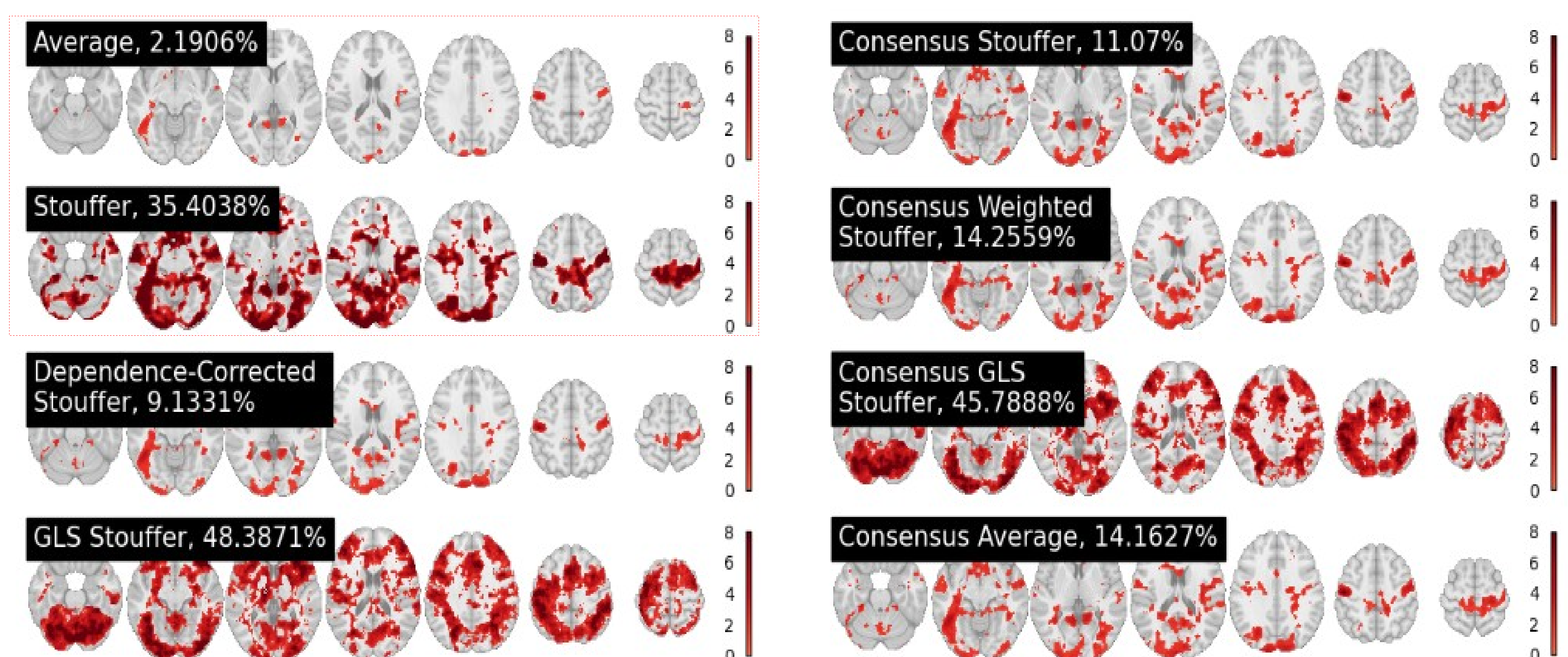
Fig 2: Comparative P-P plots for each meta-analysis estimator in the independent (upper row) and correlated data (lower row) simulations, where the y-axis is the difference in observed and expected sorted p-value, and the x-axis is the expected sorted p-value. The blue shadow depicts the theoretical confidence interval, and the observed false-positive rate at $p < 0.05$ is written on each plot, and appears in green when working as expected (i.e. within the confidence bounds).

Simulation results under the null setting of no effect shows that when pipelines are independent, most meta-analysis estimators are valid except for the simple mean method (Fig 2 top). However, when data are correlated, as in the multiverse setting, only the SDMA estimators worked properly (Fig 2). These results demonstrated the validity of the SDMA models in the setting of no signal.

Model evaluation in NARPS data

Fig 3: Demonstration of different traditional MA and SDMA methods using the statistical maps from the first hypothesis in the NARPS study (parametric effect of gains in a mixed gambles task). Maps were thresholded at $P < 0.05$ uncorrected to allow for direct comparison. Name of the MA model and percentage of significant voxels are displayed on top of each map

As expected from simulation, the Average method appears conservative and Stouffer's liberal; while the SDMA methods have different level of significance based on their specificities (Fig 3). For example, the GLS Stouffer and Consensus GLS Stouffer increase weights of correlated pipelines, while the the Dependence-Corrected Stouffer and the Consensus average put an equal weight on each pipeline.



Bibliographie

Bowring A, Maumet C, Nichols TE. Exploring the impact of analysis software on task fMRI results. Hum Brain Mapp. 2019

