ROI analysis of pharmafMRI data: an adaptive approach for global testing

Giorgos Minas, John A.D. Aston, Thomas E. Nichols and Nigel Stallard

Abstract Pharmacological fMRI (pharmafMRI) is a new highly innovative technique utilizing the power of functional Magnetic Resonance Imaging (fMRI) to study drug induced modulations of brain activity. FMRI recordings are very informative surrogate measures for brain activity but still very expensive and therefore pharmafMRI studies have typically small sample sizes. The high dimensionality of fMRI data and the arising high complexity requires sensitive statistical analysis in which often dimensionality reductions are crucial. We consider Region of Interest (ROI) analysis and propose an adaptive two-stage testing procedure for respectively formulating and testing the fundamental hypothesis as to whether the drug modulates the control brain activity in selected ROI. The proposed tests are proved to control the type I error rate and they are optimal in terms of the predicted chance of a true positive result at the end of the trial. Power analysis is performed by re-expressing the high dimensional domain of power function into a lower dimensional easily interpretable space which still gives a complete description of the power. Based on these results, we show under which circumstances our procedure outperforms standard single-stage and sequential two-stage procedures focusing on the small sample sizes typical in pharmafMRI. We also apply our methods to ROI data of a pharmafMRI study.

Giorgos Minas
Department of Statistics and Warwick Centre of Analytical Sciences, University of Warwick, UK e-mail: g.c.minas@warwick.ac.uk

John A.D. Aston
CRISM, Department of Statistics, University of Warwick, UK e-mail: j.a.d.aston@warwick.ac.uk

Thomas E. Nichols
Department of Statistics and Warwick Manufacturing Group, University of Warwick, UK e-mail: t.e.nichols@warwick.ac.uk

Nigel Stallard
Division of Health Sciences, Warwick Medical School, University of Warwick, UK e-mail: n.stallard@warwick.ac.uk

Key words: functional Magnetic Resonance Imaging, global testing, dimension reduction, adaptive designs, predictive power

1 Introduction

Pharmacological fMRI (pharmafMRI) is an exciting new technique employing functional Magnetic Resonance Imaging (fMRI) to study brain activity under drug administration. The so-called Blood Oxygenation Level Dependent (BOLD) fMRI contrast, often used in pharmafMRI studies, measures local blood flow changes known to be associated with changes in brain activity. While becoming more established, pharmafMRI faces a number of challenges of which some are statistical.

FMRI datasets are extremely high dimensional with enormous spatial resolution ($\approx 3mm$) and moderate temporal resolution ($\approx 3s$). The typical fMRI dataset produced by a single scanning session consists of BOLD recordings acquired during a relatively short period of time (few hundreds time points) from around 10^5 voxels (3-dimensional volume elements) throughout the brain. To handle such high dimensional datasets it is often appropriate to formulate specific regional hypotheses for the drug action and reduce accordingly the dimension of the data. The need for this type of analysis, which can provide regional summary measures of drug effect, is particularly acute in the typical pharmafMRI setting, in which due to the high cost of fMRI scans only a small number of subjects can be recruited.

Region of Interest (ROI) analysis can reduce an fMRI dataset into a relatively small number of ROI response summary measures expressing the local strength of treatment effect across the selected brain regions. If both the definition of ROI and the computation of the ROI response measures are cautiously conducted, a statistical analysis based on these ROI measures can potentially achieve high levels of sensitivity. We wish to go along this path and apply a multivariate test assessing the fundamental null hypothesis as to whether the new compound of interest changes the underlying brain activity in the selected ROI.

In previous work [5], we showed that tests based on a scalar linear combination of multivariate ROI responses can outperform fully multivariate methods, especially for the typically small sample sizes of fMRI studies. The decisive question for the former tests is the selection of the weights applied to ROI responses. In his seminal contribution O'Brien [6] use equal weights for all coordinates while Läuter [3] extract the weighting vector from the data sums of products matrix. In Minas et al. [5] the weights are optimally derived based on prior information and pilot data.

Here, we develop an adaptive two-stage procedure where a weighting vector, initially chosen based on prior information, is optimally adapted at a subsequent interim analysis based on the collected first stage data. The first and the second weighting vector are applied to the first and second stage responses, respectively, to produce the stage-wise linear combination test statistics. A combination function, combining the test statistics of the two studies, is used to perform the final analysis.

Both weighting vectors are optimal in terms of the predictive power [7] of this two-stage test which is analytically proved to control type I error rate.

Finally, we perform power analysis of the proposed tests and power comparisons to alternative methods. Note that the performance of a test with such a high dimensional domain of the power function can be hard to interpret. We tackle this problem by proving that the high dimensional power domain can be re-expressed into a lower dimensional easily interpretable space which still gives a complete description of the power. Using these results, our power analysis shows clearly those circumstances where our procedure outperforms standard single stage and two-stage sequential procedures. We also apply our methods to ROI data of a pharmafMRI study in which our tests are shown to be far more powerful than the latter methods.

2 Formulation

In this section we formally introduce our problem. We start by giving a brief description of the methods for extracting ROI measures from fMRI data.

ROI measures are typically extracted from mass univariate General Linear Models (GLMs) applied to the preprocessed series of 3-dim fMRI image at voxel-by-voxel resolutions (see figure 1). Estimates of the treatment effect in each voxel of each subject are first extracted from these GLMs and then averaged across the predefined, based on either brain anatomy or brain function, ROI. The coordinates of the produced multivariate outcomes correspond to representative measures of the treatment effect within each ROI of each subject.

In our methods, we assume that these ROI responses of the n_j subjects participating in stage j of the study are independent multivariate Normal random variables

$$Y_{ji} \sim N_K(\mu, \Sigma), \;\; i = 1,2,...,n_j, \;\; j = 1,2, \tag{1}$$

with mean μ and covariance matrix Σ. Normality is typically an acceptable assumption for modeling ROI linear measures in fMRI [2].

We summarise the ROI responses using scalar linear combinations

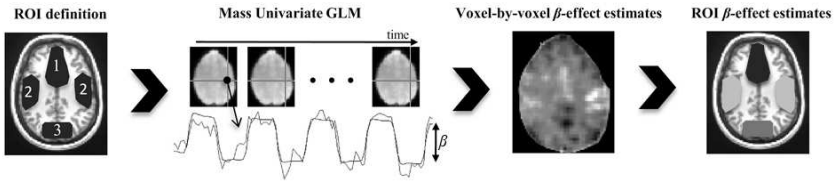

Fig. 1 Typical steps of fMRI data analysis producing a multivariate ROI outcome. The preprocessed series of fMRI images are modeled at voxel-by-voxel resolution using mass univariate GLMs. Suitable estimates of parameter values (β) expressing the treatment effect in each voxel are first extracted from the GLM and then averaged across the predefined ROI.

$$L_{ji} = \sum_{k=1}^{K} w_{jk} Y_{jik}, \tag{2}$$

where w_{jk} is the non-zero weight applied to the k-th ROI response, $k = 1,...,K$, of stage j. Using these linear combinations, we wish to test the global null hypothesis of no treatment effect across all ROI $H_0 : \mu = \underline{0}$ $(=(0,0,...,0)^T)$ against the two-sided alternative $H_1 : \mu \neq \underline{0}$.

The stage-wise test statistics in our design are the linear combination z and t statistics

$$Z_j = \frac{\overline{L}_j}{\sigma_j / n_j^{1/2}}, \qquad T_j = \frac{\overline{L}_j}{s_j / n_j^{1/2}} \tag{3}$$

for Σ known or unknown, respectively. Here, σ_j^2, $\overline{L}_j$, s_j^2 are the variance, sample mean and sample variance of the linear combination L_j, respectively. The two-sided p values, p_j, $j = 1,2$, may be obtained from the z or t statistics in (3). We use a **two-stage design** which instructs the investigators to:

1. stop the trial (after the first stage) and reject H_0 if $p_1 < \alpha_1$ or stop the trial without rejection if $p_1 > \alpha_0$,
2. continue to the second stage if $\alpha_1 \leq p_1 \leq \alpha_0$ and reject H_0 if $p_1 p_2 < c$.

Here, the Fisher's product combination function [1], $p_1 p_2$, is used for the final analysis. We also consider alternative functions including the Inverse Normal combination function [4]. Under this design, the **type I error rate** is controlled at the nominal α level if the rejection probability of the two-stage z or t test,

$$pr(p_1 < \alpha_1) + \int_{\alpha_1}^{\alpha_0} pr(p_1 p_2 < c \mid p_1)\, g(p_1) \mathrm{d}p_1, \quad g(\cdot) \text{ density of } p_1, \tag{4}$$

is under the null hypothesis H_0 equal to α.

We target on maximizing the **power** of the above two-stage tests, i.e. the rejection probability in (4) under H_1, with respect to the weighting vectors w_1, w_2, while controlling the type I error rate. In other words, we wish to find the optimal direction in which the projection of the treatment effect vector produces optimal power.

3 Methods

Here, we develop the proposed adaptive two-stage testing procedure. We start by providing the optimal weighting vector for the two-stage z and t tests described above.

Theorem 1. *Under the assumption in* (1)*, the power of the above two stage tests, i.e. the rejection probability in* (4) *under* H_1*, is* ***maximized*** *with respect to* w_1 *and* w_2 *if and only if the latter are both proportional to* $\omega = \Sigma^{-1}\mu$.

The optimal weighting vector ω is unknown and therefore we use the available information at the planning stage (prior) and at the interim stage (posterior) to select w_1, w_2.

Prior information D_0 elicited from previous studies and experts clinical opinion is used to inform the following Normal and Inverse-Wishart priors for μ and Σ, respectively,

$$(\mu \mid \Sigma, D_0) \sim N_K(m_0, \Sigma/n_0), \quad (\Sigma \mid D_0) \sim IW_{K\times K}\left(\nu_0, S_0^{-1}\right). \tag{5}$$

Here, m_0 represents a prior estimate for μ, n_0 the number of observations m_0 is based on; and ν_0, S_0 respectively represent the degrees of freedom and scale matrix of the inverse-Wishart prior.

Under this Bayesian model, the posterior distributions, given the prior information D_0 and the first stage data y_1, have the same form as the prior distributions

$$(\mu \mid \Sigma, D_0, y_1) \sim N_K(m_1, \Sigma/(n_0+n_1)), \quad (\Sigma \mid D_0, y_1) \sim IW_{K\times K}\left(\nu_0+n_1, S_1^{-1}\right). \tag{6}$$

where the posterior estimates

$$m_1 = \frac{n_0 m_0 + n_1 \bar{y_1}}{n_0+n_1}, \;\; S_1 = S_0 + (n_1-1)S_{y_1} + \frac{n_0 n_1}{n_0+n_1}(\bar{y_1}-m_0)(\bar{y_1}-m_0)^T \tag{7}$$

can be thought as "weighted averages" of the prior and first stage estimates of μ and Σ, respectively.

We wish to optimally select the weighting vectors of the two stages. Here optimality is defined in terms of the predictive power of the test. Predictive power expresses *"the chance, given the data so far, that the planned test rejects H_0 when the trial is completed"*. Given D_0, the predictive power $B_{z,1}$ and $B_{t,1}$ for the two-stage z and t tests, respectively, are defined as

$$pr(\, p_1 < \alpha_1 \mid D_0) + pr(p_1 \in [\alpha_1, \alpha_0],\; p_1 p_2 < c \mid D_0) \tag{8}$$

and if we continue to the second stage, the predictive power $B_{z,2}$ and $B_{t,2}$ given the prior information D_0 and the first stage data $\mathbf{y}_1$ are defined as

$$pr(p_1 p_2 < c \mid D_0, \mathbf{y}_1), \tag{9}$$

for p_j corresponding to either the z or t statistics in (3), respectively.

Theorem 2. *Under the assumptions (1) and (6), the first and second stage predictive power of the z test, $B_{z,1}$ and $B_{z,2}$ are maximized with respect to w_1, w_2, respectively, if the latter are proportional to $w_{z,1} = \Sigma^{-1}m_0$ and $w_{z,2} = \Sigma^{-1}m_1$, respectively.*

Further, for large ν_0, i.e. $\nu_0 \to \infty$, the weighting vectors $w_{t,1} = S_0^{-1}m_0$ and $w_{t,2} = S_1^{-1}m_1$ maximise the predictive power functions $B_{t,0}$ and $B_{t,1}$.

We can now describe the proposed adaptive two-stage z and t tests. These follow the two-stage design described earlier with the first and second stage weighting vectors of the stage-wise z and t statistics being equal to the vectors $w_{z,1}, w_{z,2}$ and

$w_{t,1}, w_{t,2}$, respectively. These tests are power optimal based on the collected information. We can also prove that they control the type I error rate.

4 Power analysis

The design variables that need to be considered for the analysis of the power function of the above z and t tests are: (i) the stopping boundaries α_0, α_1 and c, (ii) the sample sizes n_0, n_1 and n_2 (and v_0), (iii) the parameters μ and Σ and (iv) the prior estimate(s) m_0 (and S_0). While the variables in (i) and (ii) are scalar, those in (iii) and (iv) are high dimensional ($\mathbb{R}^K \times \mathbb{R}^{K\times K} \times \mathbb{R}^K$ ($\times \mathbb{R}^{K\times K}$)). Without any dimensionality reduction, it would be challenging to get a full picture and explain the power performance of our tests. However, we can prove that for the z test, (iii) and (iv) can be replaced by: (a) the Mahalanobis distance $(\mu\Sigma^{-1}\mu)^{1/2}$ of the null $N_K(\underline{0},\Sigma)$ to the alternative $N_K(\mu,\Sigma)$ distribution expressing the strength of the treatment effect and (b) the angle θ between the transformed optimal weighting vector $\tilde{\omega} = \Sigma^{-1/2}\mu$ and the transformed selected first stage weighting vector $\tilde{w}_{z,1} = \Sigma^{-1/2}m_0$ (both transformations correspond to left multiplication by $\Sigma^{1/2}$). Considering the t test the angular distance in (b) is replaced by one expressed in terms of easily interpretable vectors in $[0,\pi/2]^K \times [0,\pi/2]^K \times \mathbb{R}_+^K$. In figure 2, we illustrate how these results can be used to compare our procedure to standard testing procedures. For small sample sizes, the power of the single-stage t test is larger (smaller) than the power of the

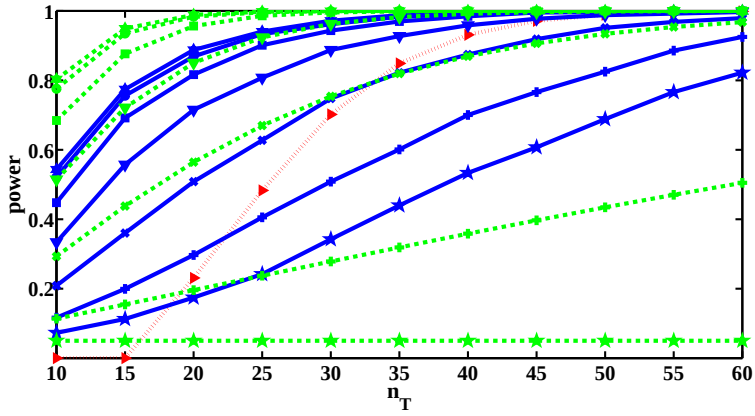

Fig. 2 Simulation-based approximation of the power, β_t, of the single-stage (green $--$) and adaptive (blue —) linear combination t test as well as the Hotelling's T^2 test (red $\cdots$) plotted against the total sample size n_T. The angle θ between $\tilde{\omega}$ and the transformed selected weighting vectors $\tilde{w}$ and $\tilde{w}_{z,1}$ of the single-stage t test and the first stage of the adaptive t test, respectively, are taken to be equal to 0° ($*$), 15° (●), 30° (■), 45° (▼), 60° ($\times$), 75° ($+$) and 90° ($\star$). Further, $\alpha_0 = 1$, $\alpha_1 = 0.01$, $c = 0.0087$ ($\alpha = 0.05$), $K = 15$, $n_0 = 5$, $v_0 = 4$, $f = n_1/n_T = 0.5$ and $D_1 = \Sigma^{-1/2}S_0\Sigma^{-1/2} = I$.

adaptive t test if the selected weighting vector is relatively close (distant) to the optimal weighting vector. For relatively large sample sizes, in contrast to the single stage test, the adaptive t test reaches high power levels even for first stage weighting vector orthogonal ($\theta = 90^\circ$) to the optimal. For increasing n_T and all other design variables remaining fixed, the angle $\ddot{\theta}$, for which the power of Hotelling's T^2 test (applicable only for $n_T > K$) is equal to the power of the t tests, is decreasing.

4.1 Application to a pharmafMRI study

We use the sample mean and sample covariance matrix (see table 1) of ROI data extracted from a GlaxoSmithKline pharmafMRI study ($K = 11$, $n_T = 13$) to perform power comparisons. As we can see in table 1, effect sizes differ across ROI and generally high correlations are observed. Further, the prior estimates presented are fairly poor resulting in angle θ between $\tilde{\omega}$ and $\tilde{w}_{t,1}$ equal to 67°. However, even for these prior estimates and such small sample sizes the adaptive t test might be considered as sufficiently powered ($\beta_t = 0.82$). This is in contrast to standard single stage tests, such as Hotelling's T^2, OLS [6], SS and PC [3] t tests ($\beta_{T^2} = 0.30$, $\beta_{OLS} = 0.13$, $\beta_{SS} = 0.13$, $\beta_{PC} = 0.14$) as well as their corresponding sequential two-stage versions ($\beta^s_{OLS} = 0.10$, $\beta^s_{SS} = 0.09$, $\beta^s_{PC} = 0.10$, sequential Hotelling's T^2 test not applicable for $n_T = 13$) which give very low power values. Note that for

Table 1 Means (line 1), variances (line 3) and correlations (upper triangle of matrix in lines $5-15$) and the corresponding prior estimates (lines 2, 4 and lower triangle of matrix in lines $5-15$) of ROI data of the sample ($n_T = 13$) of a GSK pharmafMRI study. The ROI are: *Anterior Cingulate (AC), Atlas Amygdala (A), Caudate (C), Dorsolateral Prefrontal Cortex (DLPFC), Globus Pallidus (GP),Insula (I), Orbitofrontal cortex (OFC), Putamen (P), Substantia Nigra (SA), Thalamus (T), Ventral Striatum (VS).*

	ROI	AC	A	C	DLPFC	GP	I	OFC	P	SA	T	VS
1	μ_k	−0.01	0.06	−0.08	−0.08	−0.14	−0.02	−0.08	−0.06	−0.10	−0.10	−0.13
2	$m_{0,k}$	0	0.10	−0.10	−0.10	−0.15	0	−0.15	0	−0.10	−0.10	−0.15
3	σ_k	0.11	0.11	0.03	0.05	0.11	0.08	0.13	0.15	0.10	0.11	0.10
4	$s_{0,k}$	0.15	0.10	0.02	0.10	0.10	0.10	0.15	0.15	0.10	0.10	0.10
5	AC	1	0.70	0.87	0.88	0.73	0.89	0.66	0.81	0.26	0.95	0.70
6	A	0.70	1	0.54	0.61	0.72	0.77	0.65	0.68	0.59	0.68	0.66
7	C	0.70	0.50	1	0.89	0.72	0.87	0.47	0.80	0.27	0.90	0.74
8	DLFPC	0.70	0.70	0.70	1	0.71	0.76	0.73	0.77	0.27	0.87	0.62
9	GP	0.70	0.70	0.70	0.70	1	0.86	0.51	0.90	0.54	0.70	0.90
10	I	0.70	0.70	0.70	0.70	0.70	1	0.45	0.85	0.46	0.86	0.84
11	OFC	0.50	0.50	0.50	0.70	0.50	0.50	1	0.44	0.09	0.65	0.30
12	P	0.70	0.70	0.70	0.70	0.70	0.70	0.50	1	0.49	0.82	0.89
13	SA	0.50	0.70	0.30	0.50	0.50	0.50	0.50	0.30	1	0.30	0.55
14	T	0.70	0.70	0.70	0.70	0.70	0.70	0.50	0.70	0.50	1	0.74
15	VS	0.70	0.50	0.70	0.70	0.70	0.70	0.50	0.70	0.50	0.70	1

improved prior estimates (smaller angles) the power of the adaptive t test can be increased further.

5 Discussion

The formulation of specific regional hypotheses for drug action and the associated dimensionality reductions are crucial for further establishment of pharmafMRI. As we illustrate in our methods, ROI analysis combined with multivariate methods can be successfully used to answer the fundamental question as to whether the drug modulates the brain activity over the regions of greatest interest for a particular study. We show that reduction of ROI responses to a scalar linear combination may substantially increase sensitivity compared to fully multivariate methods on ROI responses, without any cost in terms of specificity. For the latter reduction, we propose deriving the weights of the linear combination by exploiting the available prior information and allowing for data dependent adaptation at an interim analysis. These weights are optimal in terms of the predictive power given the available information at each selection time. Further, we show how the high dimensional power function domain space can be reduced to a lower dimensional easily interpretable space which allows us to show clearly under which circumstances the improvement over single stage and sequential designs is achieved. We finally show that our methods can outperform standard single stage and sequential two-stage multivariate tests in a pharmafMRI study.

References

1. Bauer, P., Kohne, K.: Evaluation of Experiments with Adaptive Interim Analyses. Biometrics **50**, 1029−−1041 (1994)
2. Friston, K., Ashburner, J., Kiebel, S., Nichols, T., and Penny, W.: Statistical parametric mapping: the analysis of funtional brain images. Elsevier/Academic Press, Amsterdam; Boston (2007)
3. L*ä*uter, J.: Exact t and F tests for analyzing studies with multiple endpoints. Biometrics **52**, 964−−970 (1996)
4. Lehmacher, W., Wassmer, G.: Adaptive Sample Size Calculations in Group Sequential Trials. Biometrics **55**, 1286−−1290 (1999)
5. Minas, G., Rigat, F., Nichols, T.E., Aston, J.A.D., Stallard, N.: A hybrid procedure for detecting global treatment effects in multivariate clinical trials: theory and applications to fMRI studies. Statist. Med. **31**, 253−−268 (2012)
6. OBrien, P. C.: Procedures for comparing samples with multiple endpoints. Biometrics **40**, 1079−−1087 (1984)
7. Spiegelhalter, D. J., Freedman, L. S., Blackburn, P.: Monitoring clinical trials: conditional or predictive power?. Control. Clin. Trials **7**, 8−−17 (1986)