

Regions of interest analysis in pharmacological fMRI: How do the definition criteria influence the inferred result?

Georgios D. Mitsis,^a Gian Domenico Iannetti,^b Trevor S. Smart,^c
Irene Tracey,^{a,b} and Richard G. Wise^{a,d,*}

^aCentre for Functional Magnetic Resonance Imaging of the Brain (FMRIB), Department of Clinical Neurology, University of Oxford, Oxford, UK

^bDepartment of Physiology, Anatomy and Genetics, University of Oxford, Oxford, UK

^cPfizer Ltd., Sandwich, UK

^dCardiff University Brain Research Imaging Centre (CUBRIC), School of Psychology, Cardiff University, Cardiff, UK

Received 27 July 2007; revised 9 November 2007; accepted 14 November 2007

Available online 3 December 2007

Prior hypotheses in functional brain imaging are often formulated by constraining the data analysis to regions of interest (ROIs). In this context, this approach yields higher sensitivity than whole brain analyses, which could be particularly important in drug development studies and clinical decision making. Here we systematically examine the effects of different ROI definition criteria on the results inferred from a hypothesis-driven pharmacological fMRI experiment, with the aim of maximising sensitivity and providing a recommended procedure for similar studies. In order to achieve this, we compared different criteria for selecting both anatomical and functional ROIs. Anatomical ROIs were defined (i) specifically for each subject, (ii) at group level, and (iii) using a Talairach-like atlas, in order to assess the effects of inter-subject anatomical variability. Functional ROIs (fROIs) were defined, both for each subject and at group level, by (i) selecting the voxels with the highest Z-score from each study session, and (ii) selecting an inclusive union of significantly active voxels across all sessions. A single value was used to summarise the response within each ROI. For anatomical ROIs we used the mean of the parameter estimates (β values) of either all voxels or the top 20% active voxels. For fROIs we used the mean β value of all voxels constituting the ROI. The results were assessed in terms of the achieved sensitivity in detecting the experimental effect. The use of single-subject anatomical ROIs combined with a summary value calculated from the top 20% fraction of active voxels was the most reliable and sensitive approach for detecting the experimental effect. The use of fROIs from individual sessions introduced unacceptable biases in the results, while the use of union fROIs yielded a lower sensitivity than anatomical ROIs. For these reasons, fROIs should be employed with caution when it is not possible to make clear anatomical prior hypotheses.

© 2007 Elsevier Inc. All rights reserved.

Introduction

Regions of interest (ROIs) have been employed in functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) as a means of testing prior hypotheses about brain function, increase statistical power compared to whole-brain analyses and facilitate comparisons across multiple subjects or studies. These are particularly important in the case of drug discovery studies (Wise and Tracey, 2006), in which there is a need for strict hypothesis testing regarding drug action based on regulatory requirements as well as decision making about further compound development. In this context, the advantages offered by the ROI approach (e.g., increased statistical power) could be exploited in practical circumstances. Also, fMRI is increasingly used in clinical decision making (Jezzard and Buxton, 2006) and the ROI approach could prove useful by tying decisions to regional hypotheses.

In model-based, whole-brain fMRI analyses, the blood oxygen level dependent (BOLD) measurements in each voxel under different experimental conditions are fitted independently to a general linear model and the subsequently computed test statistics are compared to a threshold in order to assess activation. This procedure gives rise to a multiple testing problem whereby the number of test statistics are on the order of thousands. Moreover, these statistics are correlated across voxels, so corrections, such as Bonferroni-type corrections, random fields or permutation tests, are required to control for the number of false positives (Worsley et al., 1996; Nichols and Hayasaka, 2003). By constraining the area where one searches for effects, the number of multiple comparisons reduces dramatically and therefore statistical sensitivity is increased. Moreover, in the case of whole-brain analysis, allowance for specific mechanistic or spatial hypotheses is not normally made. On the other hand, strong prior hypotheses are required in order to take advantage of the increased statistical power offered by the ROI approach.

These prior hypotheses are often made on an anatomical basis. This has led to the development of ROI tools based on the

* Corresponding author. CUBRIC, School of Psychology, Park Place, Cardiff CF10 3AT, UK. Fax: +44 29 2087 0339.

E-mail address: wiserg@cardiff.ac.uk (R.G. Wise).

Available online on ScienceDirect (www.sciencedirect.com).

Talairach and Tournoux atlas (Talairach and Tournoux, 1988) and the Montreal Neurological Institute (MNI) space (Collins et al., 1994; Lancaster et al., 2000; Maldjian et al., 2003). One approach is to use anatomical landmarks, such as sulci or gyri, to identify ROIs in each subject's brain. The location of subcortical structures, such as basal ganglia or amygdala, on individual anatomical scans can be defined reliably in this way. However, in the case of cortical regions, clear anatomical landmarks are often lacking. Moreover, sulci or gyri may exhibit limited correspondence to finer structural organization such as cytoarchitecture. Since microstructural borders based on measurements of cytoarchitecture can only be defined post mortem, the use of probabilistic cytoarchitectonic maps, which provide information on the location and variability of cortical areas from human post mortem brains, has been proposed (Amunts and Zilles, 2001; Amunts et al., 2004; Eickhoff et al., 2005, 2006). In this context, each voxel of the reference space may be assigned to the most likely cytoarchitectonic area at this point (Eickhoff et al., 2005, 2006).

The normalization of certain cortical areas to a stereotaxic space results in non-consistent locations, by using both standard registration techniques, such as in the V1 area of the primary visual cortex (Saxe et al., 2006), and reconstructed cortical surfaces that respect sulcal locations, such as in category-selective areas in the temporal lobe (Spiridon et al., 2006). Different areas averaged across individuals may thus reduce the sensitivity of subsequent analyses (Brett et al., 2002). Moreover, ROIs defined by anatomical criteria only may not correspond to functional organization. Therefore, an alternative approach is to define ROIs based on functional criteria (fROIs), combining the aforementioned anatomical landmarks with consistent profiles of functional activation, in individual subjects. These profiles may be defined either within the main experimental design, i.e., as part of a factorial design, or from a separate session that constrains the search space in the main experiment (functional localisers). This approach has been used extensively in studies of the visual cortex, where fROIs such as the fusiform face area (Kanwisher et al., 1997) the MT/V5 (Tootell et al., 1995), the lateral occipital complex (Kourtzi and Kanwisher, 2001) and the anterior intraparietal cortex (Culham et al., 2003) have been shown to be reproducible across studies and have provided a rigorous way for between-study comparisons. The definition of fROIs from separate localising sessions or within the main experimental design has been the subject of debate recently (Friston et al., 2006; Saxe et al., 2006).

An alternative approach to define functionally distinct regions is based on the hypothesis that functional differences are reflected in the corresponding activity time courses of these regions, in an analogous manner to the hypothesis that functional differences are reflected in anatomical differences. Therefore, spatio-temporal profiles obtained from independent component analysis during natural conditions have been used to identify functionally specialized cortical subdivisions (Bartels and Zeki, 2005). Similarly, correlations between remote, functionally homogeneous areas defined on the basis of their temporal activity have been employed to identify functionally interconnected networks (Bellec et al., 2006). Finally, connectivity patterns, obtained from diffusion-weighted images and fMRI, have been used to define regions that exhibit similar connectivity profiles and their corresponding boundaries, where these profiles change. In this manner, regions connected to the motor and prefrontal cortex were identified, in agreement with the anatomical distinction between

the supplementary motor area (SMA) and pre-SMA, which were moreover correlated with activation profiles during suitably designed tasks (Johansen-Berg et al., 2004).

Once an ROI is defined, an overall summary measure may be calculated to assess its response as a whole rather than on a voxel-by-voxel basis, reducing the number of statistical inferences. The most straightforward way to do so is by taking the average response (i.e., parameter estimates or t/Z values) of either all or the top activated voxels within the ROI. In the latter case, the threshold should be such that signal is detected and noise is avoided. This approach is flexible and can be applied to any region type; however, its efficiency depends on the distribution of signal and noise within the latter.

This paper examines the choices made in a region of interest analysis of a human pain pharmacological fMRI experiment, by systematically assessing their influence on the inferred results, with the aim of maximising sensitivity in detecting drug effects. This pharmacological study of gabapentin, (Iannetti et al., 2005), was chosen for detailed reanalysis because of the growing need to adopt robust and sensitive analysis strategies in drug development (Wise and Tracey, 2006). Nociceptive inputs also elicit a brain response in a number of widely distributed regions, allowing us to assess the applicability of the described methods across the brain. Therefore, we aimed to develop a recommended procedure for ROI analyses, which would exploit the advantages and avoid the pitfalls of this approach, by considering some of the most common approaches that are applicable in practice. This is also important for new studies that could benefit from similar analyses.

Specifically, we considered both anatomical and functional, anatomically-constrained ROIs within the so-called "pain-matrix" (Apkarian et al., 2005; Tracey, 2005). The effect of inter-subject anatomical variability was examined by defining ROIs in individual anatomical space, group-averaging the former probabilistically in standard MNI space, as well as using a standard Talairach–Daemon based atlas (Maldjian et al., 2003). We also considered the effects of between-session functional variability and the bias introduced by defining fROIs from individual study sessions, both at first (individual) and group levels. Finally, fROIs were defined as the union of all significantly active voxels from all sessions, also at the first and group levels. The fMRI data were based on a hypothesis-driven study which demonstrated the increased sensitivity yielded by the ROI approach, compared to whole-brain analyses. Drug and sensitization effects within the 'pain matrix', which were not deemed significant by whole-brain analysis, were revealed by ROI analysis (Iannetti et al., 2005). Our analysis is conducted under the hypothesis that the administered drug (gabapentin) reduced pain-related brain activity in specific brain regions.

Methods

Experimental data

The original study aimed to assess the effects of gabapentin, a pain relieving drug, on brain processing of nociceptive stimuli during the normal state and the capsaicin-induced central sensitization state. Twelve right-handed male volunteers aged 25–30 years (mean, 27 ± 2.1) participated in the study. The subjects underwent four fMRI sessions (Fig. 1) after the administration of placebo or gabapentin in the absence (sessions P₁ and D₁) and

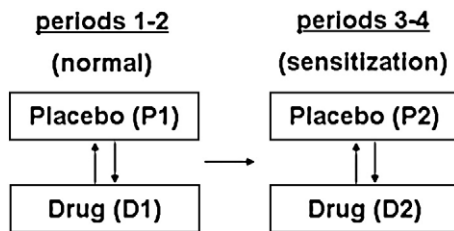


Fig. 1. Experimental design. Periods 1–2 (normal state) preceded periods 3–4 (sensitization), while the order of placebo or drug administration was randomized.

presence (sessions P_2 and D_2) of central sensitization. The order of the drug and placebo administration, which was 1800 mg *per os*, was randomized. Punctate mechanical stimuli were delivered on a 48-cm² target area on the shin area of both legs. Central sensitization was induced by combined thermal and chemical (0.075% capsaicin cream) stimulation on the right leg. Capsaicin application was always performed in periods 3–4 to avoid long-lasting effects (i.e., desensitization). The fMRI scans were performed 3 h after drug/placebo administration.

Data were collected at 3T with an Oxford Magnet Technology, 1m bore magnet which was driven by a Varian Unity Inova console. A whole-brain gradient-echo, echo-planar imaging sequence for functional scans (echo time=30 ms, 24 contiguous 6-mm-thick axial slices, field of view 256×192 mm, matrix 64×64) with a repetition time of 3 s over 500 volumes was used, corresponding to a total scan time of 25 min. During the scans, 40 identical, pseudorandomly jittered 1 s punctate stimuli, with an average interstimulus interval of 30 s (range 20–40 s), were delivered on the right or left leg, using a 60-g von Frey hair. Thirty seconds after the last mechanical stimulus, visual stimulation was delivered to control for a possible confounding effect of global modulation of BOLD signal by gabapentin. A T1-weighted, 1×1×3 mm structural image was collected during period 1. For more details on the experimental design, the reader is referred to Iannetti et al. (2005).

After data preprocessing (motion correction, spatial smoothing, demeaning and high-pass temporal filtering at a 50 s cutoff), parameter estimates for the fMRI signal in each voxel were obtained by a general linear model approach with local autocorrelation correction (Woolrich et al., 2001). Test statistics were calculated using FEAT (www.fmrib.ox.ac.uk/fsl). Group analysis was carried out using a mixed-effects approach (Beckmann et al., 2003), after registering the low-resolution functional images to high-resolution structural space (Jenkinson and Smith, 2001) and subsequently to standard MNI space (Collins et al., 1994). The group-level Z statistic images were thresholded at $Z>2.3$ using a cluster threshold of $P<0.01$ to correct for multiple comparisons (Forman et al., 1995). The whole-brain group analysis revealed a main gabapentin effect in reducing the stimulus-induced deactivations during central sensitization in response to right leg stimulation (periods 3–4). However, ROI analysis further revealed sensitization and drug effects within the “pain-matrix”; a significant increase in the fMRI signal during central sensitization in bilateral thalamus, anterior cingulate cortex and brainstem, a significant signal decrease during drug administration in the bilateral insula and secondary somatosensory cortex during the normal state (i.e., antinociceptive effect), as well as an interaction between central sensitization and gabapentin in the brainstem only (Iannetti et al., 2005).

ROI analysis

The analysis is illustrated in Fig. 2, where we summarise the ROI definition methods and the assessment metrics considered.

Anatomical ROIs

Individual ROIs were defined on the high-resolution structural images of each subject based on anatomical landmarks (Vogt et al., 1995; Duvernoy, 1999; Bense et al., 2001). These included the regions that define the ‘pain-matrix’ (Apkarian et al., 2005; Tracey, 2005), i.e., anterior cingulate cortex (ACC), brainstem (BRS), left and right insula (INSL and INSR), left and right primary (SIL, SIR) and secondary (SIIL, SIIR) somatosensory cortex and left and right thalamus (THL and THR), on which all subsequent analysis will be focused. The high-resolution structural images were registered to the corresponding four functional images for each subject using an affine transformation with 12 degrees of freedom (Jenkinson and Smith, 2001). Using this transformation matrix, the individual ROIs were registered to functional space.

In order to obtain a group-probabilistic anatomical map, the individually-drawn ROIs were registered to standard MNI space (Jenkinson and Smith, 2001) and summed across all subjects. Group-probabilistic ROIs of different sizes were subsequently obtained by thresholding this map at different levels by including, at the lowest threshold, all voxels that were part of each anatomical ROI in any 1 of the 12 subjects (1 subject threshold), those voxels common to at least 4 subjects and finally those common to at least 8 subjects (the 4 and 8 subject thresholds respectively). Inter-subject anatomical variability for each region was quantified by calculating the % spatial overlap, defined as the number of overlapping voxels divided by the average ROI size for each subject pair, between all possible 66 subject pairs. The same set of regions was finally defined from the standard Wake Forest University (WFU) Pickatlas toolbox, which generates ROI masks in standard MNI space based on the Talaraich Daemon (Lancaster et al., 2000) by using lobar, cortical and subcortical anatomy as well as Brodmann areas (Maldjian et al., 2003). In order to calculate the response within each region, the group-probabilistic and WFU atlas ROIs were registered to functional space.

This stimulus-induced BOLD response was assessed by computing a single summary value (SV) for each ROI. In the case of anatomical ROIs, both the regression parameter estimate values averaged over all and the Top 20% activated voxels, β_{mean} and $\beta_{\text{mean},20}$ were considered (Fig. 2), where the top activated voxels were selected on the basis of their Z scores, in order to account for possible noise effects. The latter measure (top fraction SV) provides a way to potentially increase signal-to-noise ratio and enhance sensitivity. Using a voxel fraction of 20% was found to yield optimal sensitivity in detecting drug-induced effects in preliminary analyses (data not shown), while keeping a significant number of voxels in each case even for the smallest ROIs. This approach can also account for spatial heterogeneities in the response within an ROI (i.e., simultaneous activations and deactivations), functional variability between subjects and allows the comparison of the same voxel number across sessions for each ROI. All SV calculations were performed in functional space.

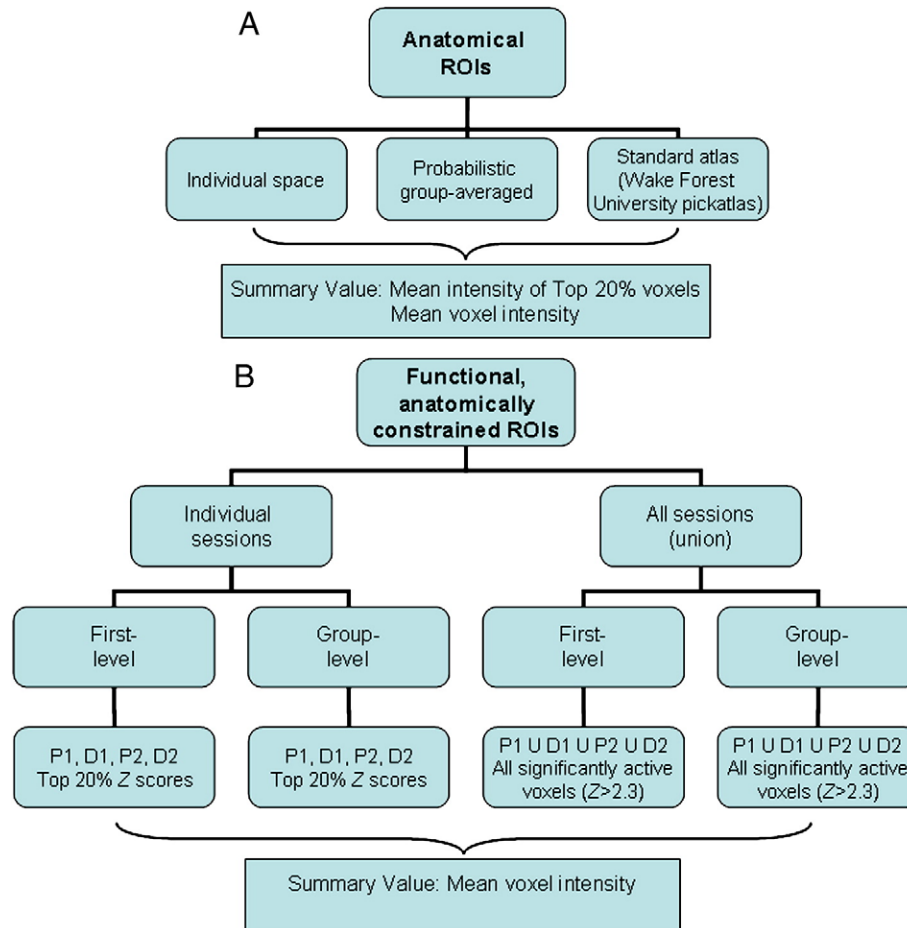


Fig. 2. (A) Anatomical ROI definition criteria and summary values (SVs). Individual, group-averaged probabilistic and standard atlas ROIs were considered. Two summary values were considered: (1) the mean intensity of the voxels with the Top 20% of Z scores and (2) the mean intensity of all voxels. (B) Functional ROIs were defined as the Top 20% activated voxels from individual sessions and as the union (U) of all significantly active voxels (corrected $Z > 2.3$) across all sessions, both from the first-level (individual subject) and group-level (all subjects) Z statistical images.

Functional ROIs

The individual-subject (first-level) Z statistical images from all four sessions were registered to each subject's individual structural space and constrained anatomically by the subject's corresponding anatomical ROIs. First-level, individual-session fROIs were defined as the Top 20% of Z-scores and were subsequently applied to all sessions (β images) to calculate summary values (e.g., activation during session P_1 was assessed by using fROIs defined from P_1 , D_1 , P_2 and D_2). Group-level, individual-session fROIs were defined similarly from the group-level Z statistical images of each session, i.e., as the Top 20% Z scores within each anatomically constrained region. These were again applied to all sessions (first-level β images) to calculate summary values in each region of interest. Finally, union fROIs were defined by including all significantly activated voxels (corrected $Z > 2.3$) across all four sessions. Two types of union fROIs were defined and applied to all sessions to calculate summary values. One type was defined from the inclusive union of the first-level (individual subjects) thresholded Z images. The other type was defined from the inclusive union of the group-level (all subjects) thresholded Z images.

In the case of fROIs, since we are considering only voxels that are deemed active beforehand, the mean β value of all voxels included in the fROI was used as a SV (Fig. 2).

Statistical analysis

After the SVs were calculated for all ROI definition methods, the sensitivity yielded by each method was assessed by paired *t*-tests between sessions that reflect drug effects, i.e., between sessions P_1 vs. D_1 ($P_1 - D_1$) and P_2 vs. D_2 ($P_2 - D_2$), for right leg stimulation. We chose to use *t*-tests without correcting for multiple comparisons because we were interested in the relative performance of each method. Since we have no ground truth, sensitivity was assessed on the basis of the statistical significance of what are plausible results for a positive drug effect. Therefore, we present the obtained *t*-scores and compare them to the uncorrected significant *t* values for a one-tailed *t*-test (i.e., 1.796 for $P = 0.05$ and 2.718 for $P = 0.01$), since we hypothesized that in general the drug reduced the signal in the pain-matrix ROIs (Iannetti et al., 2005).

Results

Anatomical variability

The group probabilistic anatomical maps for the regions that define the pain-matrix are shown in Fig. 3 in MNI space (top panel). We also show the same ROIs obtained from the standard WFU

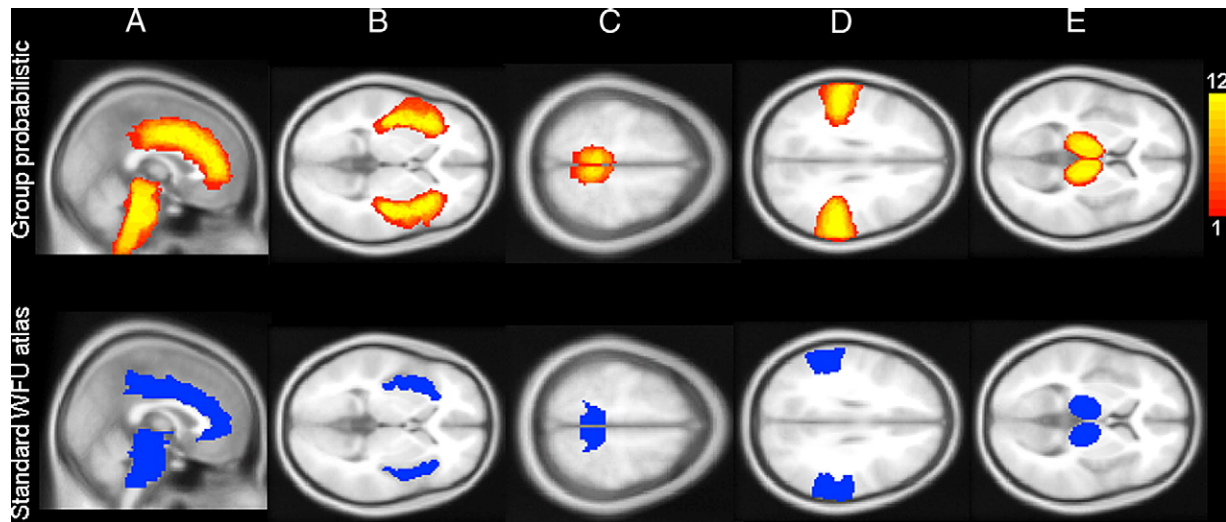


Fig. 3. Group probabilistic and WFU atlas anatomical ROIs within the pain matrix in MNI space. (A) Anterior cingulate cortex and brainstem, (B) bilateral insula, (C) bilateral primary somatosensory cortex SI, (D) bilateral secondary somatosensory cortex (SII), (E) bilateral thalamus.

atlas in the bottom panel of Fig. 3. The inter-subject anatomical variability, quantified by the % overlap for each ROI averaged between all possible subject pairs in standard space, is shown in Fig. 4 (mean \pm S.E.). Cortical regions such as SI, which are harder to define using anatomical criteria, exhibited the largest inter-subject anatomical variability. This implies that the definition of anatomical ROIs for such regions by group-averaged maps may compromise the reliability of subsequent analyses. The volumes of all anatomically defined ROIs are given in Table 1. When compared to individual ROIs, their WFU atlas counterparts were generally more extensive (with the exception of SII and left insula). The size of the group probabilistic ROIs, as obtained by thresholding at 1, 4 and 8 subjects, decreased by a factor of more than two for the most consistent regions across subjects (e.g. brainstem) to more than four for the least consistent regions (e.g. SI).

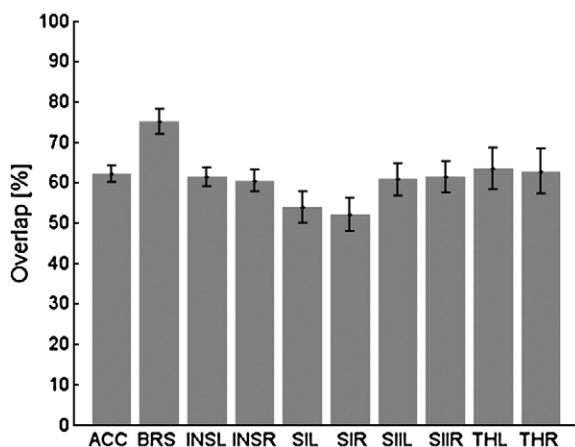


Fig. 4. Inter-subject anatomical variability for the pain-matrix ROIs defined on each subject's individual space, calculated as the mean percent overlap between all possible pairs in MNI space (the error-bars denote standard error). ACC: anterior cingulate cortex, BRS: brainstem, INSL–INSR: insula left and right, SIL–SIR: primary somatosensory cortex left and right, SIIL–SIIR: secondary somatosensory cortex left and right and THL–THR: thalamus left and right. SI exhibits the highest inter-subject variability.

In order to illustrate the signal distribution within a representative, single subject ROI, we show the β values in a single slice of the anterior cingulate cortex of one subject, defined on the corresponding structural image, in Fig. 5 (left panel: ROI mask, middle panel: β values). We also show the distribution of the Top 20% voxels, defined by their corresponding Z values, within the same region in the right panel of Fig. 5. The latter illustrates how the top fraction SV can account for heterogeneous activation responses within an ROI.

Anatomical ROIs

The t -scores between sessions P_1 – D_1 (top panels) and P_2 – D_2 (bottom panels), which reflect sensitivity in detecting hypothesized drug effects, are shown in Fig. 6 for all three anatomical ROI definition methods, when both $\beta_{\text{mean},20}$ and β_{mean} were used to assess activation (left and right panels respectively). The top fraction SV $\beta_{\text{mean},20}$ yielded considerably larger t -scores, i.e., more sensitivity, in all cases and revealed a significant drug-induced signal decrease in bilateral insula and SII during the non-sensitized

Table 1

ROI volumes (in cm^3) for individual, group-averaged probabilistic and standard WFU atlas anatomical ROIs

Region	Individual (Mean (S.E.))	ROI volume [cm^3]			WFU atlas
		$T=1$	$T=4$	$T=8$	
ACC	29.1(7.0)	80.0	42.2	21.8	54.9
BRS	31.6(6.7)	67.7	42.2	31.4	46.7
INSL	20.1(5.3)	58.5	28.3	15.1	14.2
INSR	18.5(4.5)	55.4	25.9	13.8	20.7
SIL	4.9(1.5)	15.0	8.9	3.6	7.0
SIR	4.8(1.5)	14.8	8.5	3.3	5.8
SIIL	19.5(4.2)	51.9	28.9	16.9	15.8
SIIR	17.0(4.0)	46.2	24.7	14.3	10.1
THL	7.2(1.4)	21.8	9.8	6.3	8.5
THR	7.4(1.6)	22.1	10.5	6.6	8.8

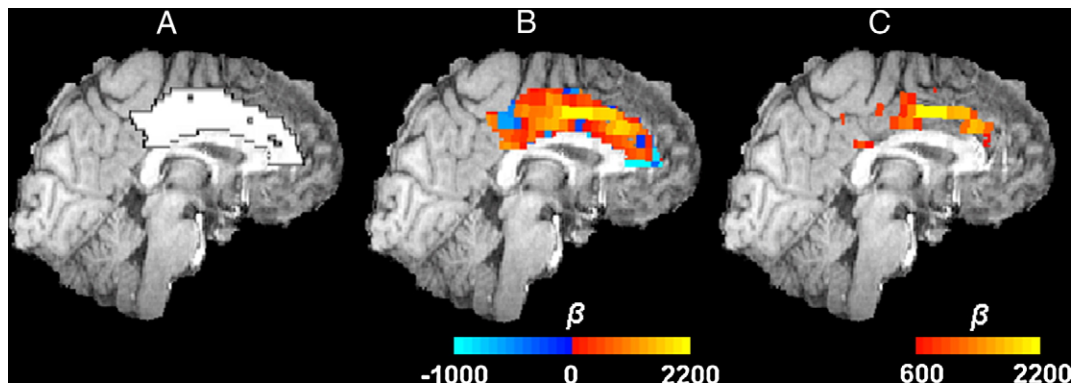


Fig. 5. Representative signal distribution (β values) in the ACC ROI of one subject in structural space (session P₂). (A) Individual anatomical ROI. (B) Distribution of β values; red–yellow scale corresponds to positive values, blue–light blue scale corresponds to negative values. (C) Top 20% voxels within ROI. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

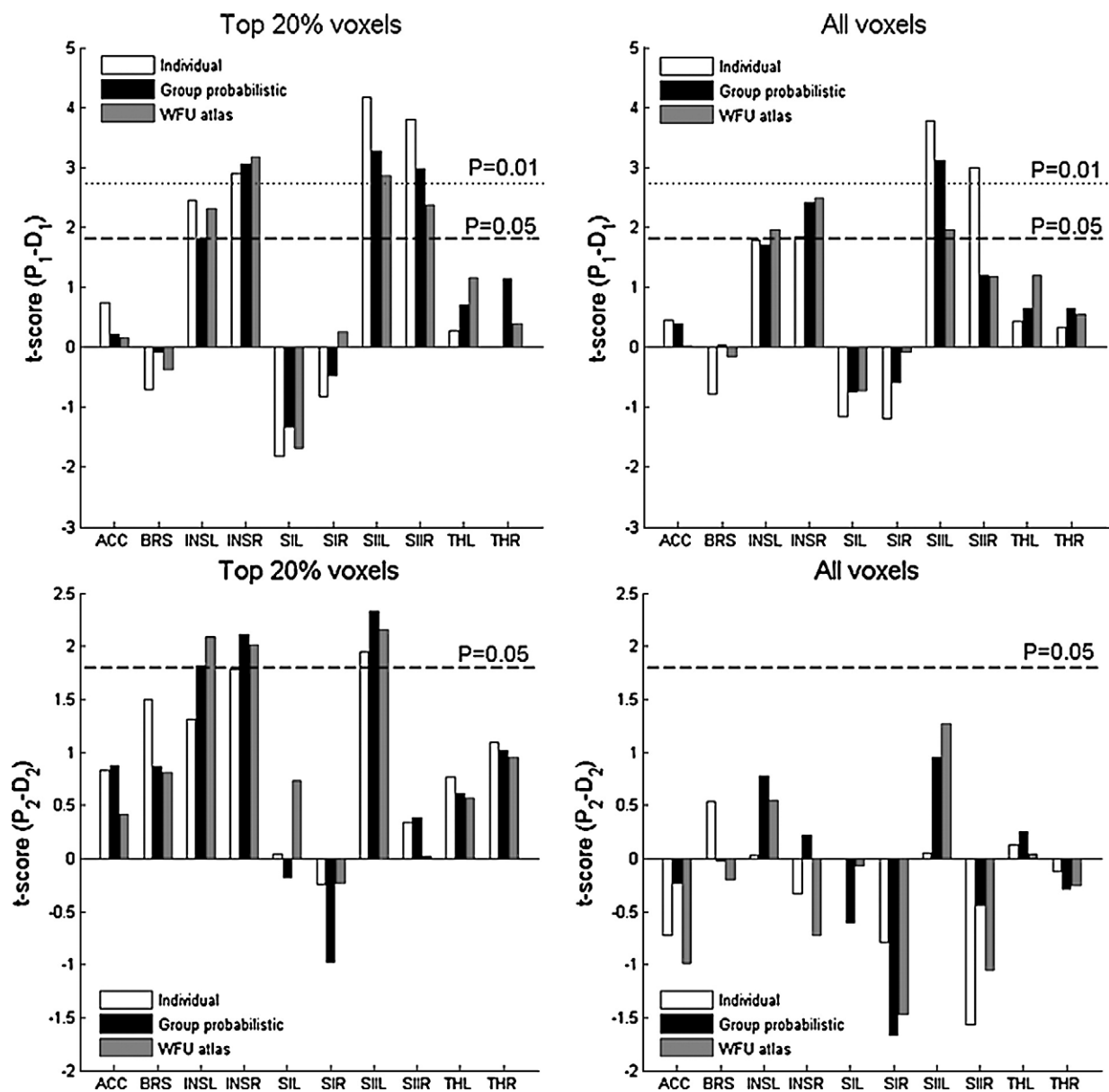


Fig. 6. Effect of anatomical ROI definition and SV choice on assessment of drug modulation: t -scores between sessions P₁–D₁ (top panels), P₂–D₂ (bottom panels) for both the top 20% ($\beta_{\text{mean},20}$ — left panels) and all voxels mean β value (β_{mean} — right panels) SVs. $\beta_{\text{mean},20}$ yields higher sensitivity in all cases.

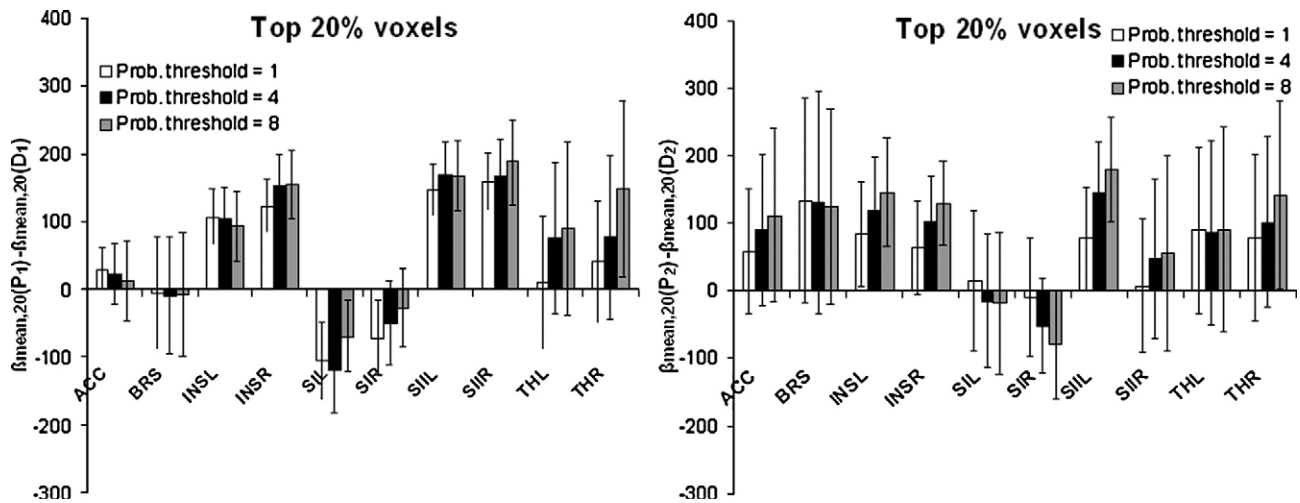


Fig. 7. Effect of probabilistic anatomical ROI size on assessment of drug modulation: Top 20% SV difference between sessions P_1-D_1 (left panels), P_2-D_2 (right panels) for three different thresholds — 1, 4 and 8 subjects, averaged over all subjects (the error bars denote standard error). The average SV difference between sessions generally increases for more smaller ROIs.

state, confirming its previously identified antinociceptive effects, as well as a significant signal decrease in bilateral insula and left SII during the sensitized state, reflecting its antihyperalgesic effects. It is also worth noting that the top fraction SV $\beta_{\text{mean},20}$ disclosed a significant signal increase in left SI, i.e. contralateral to stimulation site. The improvement achieved by using $\beta_{\text{mean},20}$ was more prominent for the sensitized state (i.e. between sessions P_2-D_2), for which the Top 20% results were close to statistical significance (t -score of 1.796, uncorrected one-tailed test, denoted by the dashed lines in Fig. 6), in contrast to β_{mean} , which yielded low t -score values. The three definition methods yielded similar results for most regions; however, more pronounced differences were observed in bilateral SI and SII (Fig. 6). For example, the t -score values between sessions P_1-D_1 calculated from the

individually-drawn ROIs were considerably higher for SII, particularly when compared to their WFU atlas counterparts (for which $P < 0.01$ only for SIIL — Fig. 6, top left panel). This suggests that using a standard atlas or group-averaging after normalization to standard space, even within the study subject group, may have a considerable detrimental effect on the obtained results for cortical regions that may not align well across subjects.

The effects of ROI size were examined by thresholding the group probabilistic anatomical maps of Fig. 3 at 1, 4 and 8 subjects. Group-probabilistic anatomical ROIs contained voxels that were common to any 1 of the 12 subjects (1 subject threshold), common to any 4 and common to any 8 subjects (4 and 8 subject thresholds respectively). This gave ROIs of decreasing size, allowing us to examine the effect of region volume on the results.

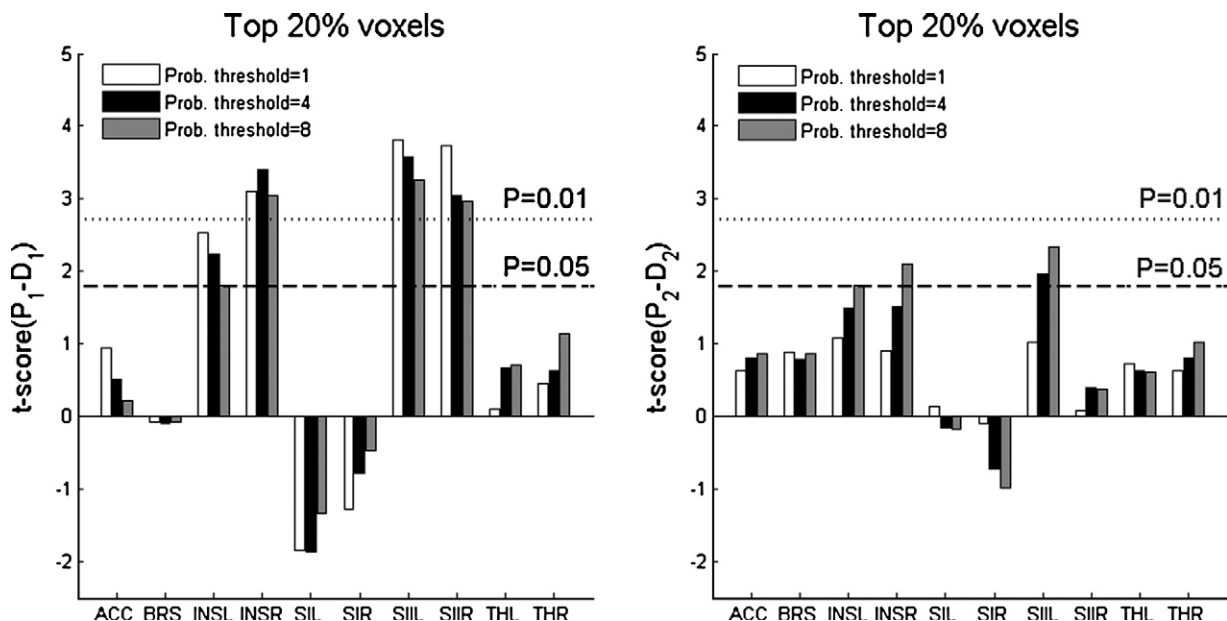


Fig. 8. Effect of probabilistic anatomical ROI size on assessment of drug modulation: t -scores between sessions P_1-D_1 (left panels), P_2-D_2 (right panels) for three different thresholds — 1, 4 and 8 subjects. More constrained regions do not always yield higher sensitivity (INSL, SIIL, SIIR between sessions P_1-D_1).

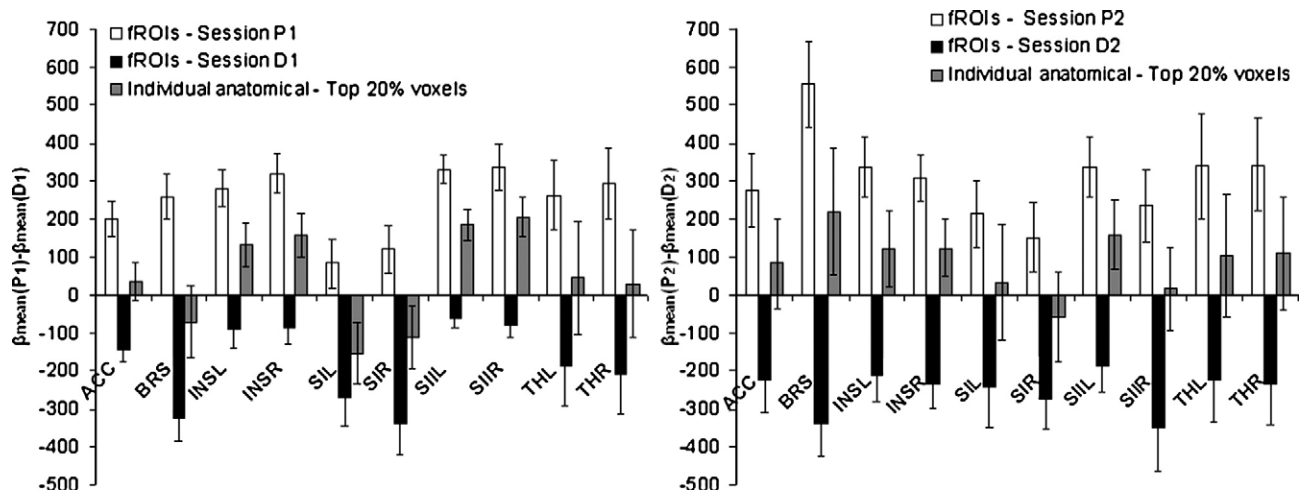


Fig. 9. Effect of first-level, individual-session fROIs on assessment of drug modulation: Top 20% SV difference between sessions P_1 – D_1 (left panel) and P_2 – D_2 (right panel), averaged over all subjects (the error bars denote standard error). Note the large bias introduced by selecting the fROI from within one of the individual sessions under comparison.

The difference in the top fraction SVs between sessions P_1 – D_1 and P_2 – D_2 (i.e., $\beta_{\text{mean},20}(P_i) - \beta_{\text{mean},20}(D_i)$ where $i=1,2$) are shown in Fig. 7, averaged over all subjects (mean \pm S.E.), and the corresponding t -scores are shown in Fig. 8. The averaged $\beta_{\text{mean},20}$ exhibited an increasing trend in most cases as the group-probabilistic ROI size was constrained (Fig. 7); however, this did not always result in increased sensitivity. The more constrained ROIs yielded increased t -scores between sessions P_2 – D_2 (bilateral insula and right SII), but not between sessions P_1 – D_1 (Fig. 8).

Functional ROIs

The effects of using first-level, individual-session fROIs were initially assessed by the SV difference between sessions P_1 – D_1 and P_2 – D_2 (i.e., $\beta_{\text{mean}}(P_i) - \beta_{\text{mean}}(D_i)$, $i=1,2$) after defining fROIs from all four sessions. Since considerable between-session

variability was observed in the location of the Top 20% voxels, defining fROIs using either of the sessions under comparison (e.g., P_1 or D_1 when comparing P_1 – D_1) biased the results in the corresponding direction. In order to illustrate the extent of this bias, we plot $\beta_{\text{mean}}(P_i) - \beta_{\text{mean}}(D_i)$ ($i=1,2$) for fROIs defined from P_1 , D_1 (left panel) and P_2 , D_2 (right panel), averaged over all subjects, in Fig. 9 (mean \pm S.E.), along with their individual anatomical ROI counterparts (Top 20% SV). Compared to the latter, fROIs affected the obtained SVs considerably. This is also evident in Fig. 10, where we show the corresponding t -scores from all four first-level individual session fROIs. For the fROIs considered in Fig. 9, the t -scores were affected to such an extent that in some cases statistical significance was achieved in opposite directions. Specifically, whereas individual anatomical ROIs yielded a statistically significant signal reduction between sessions P_1 – D_1 for bilateral insula and SII, fROIs defined from

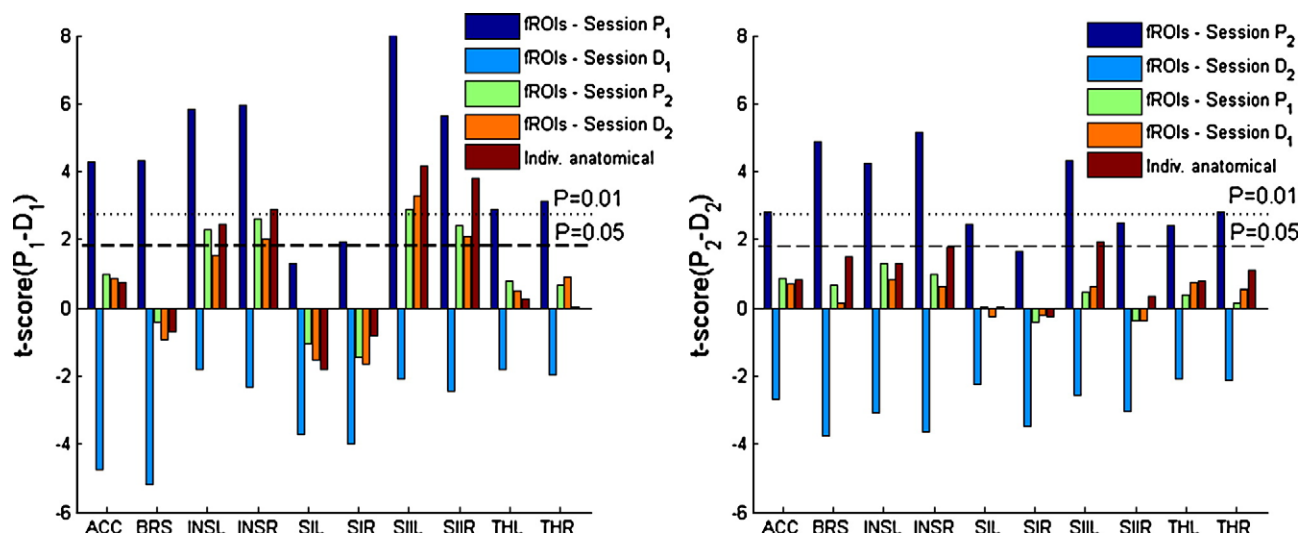


Fig. 10. Effect of first-level, individual-session fROIs on assessment of drug modulation: Note the large bias introduced by fROIs defined from P_1 , D_1 (left panel) and P_2 , D_2 (right panel) respectively. fROIs defined from the remaining sessions yield lower sensitivity than individual anatomical ROIs combined with the top 20% SV.

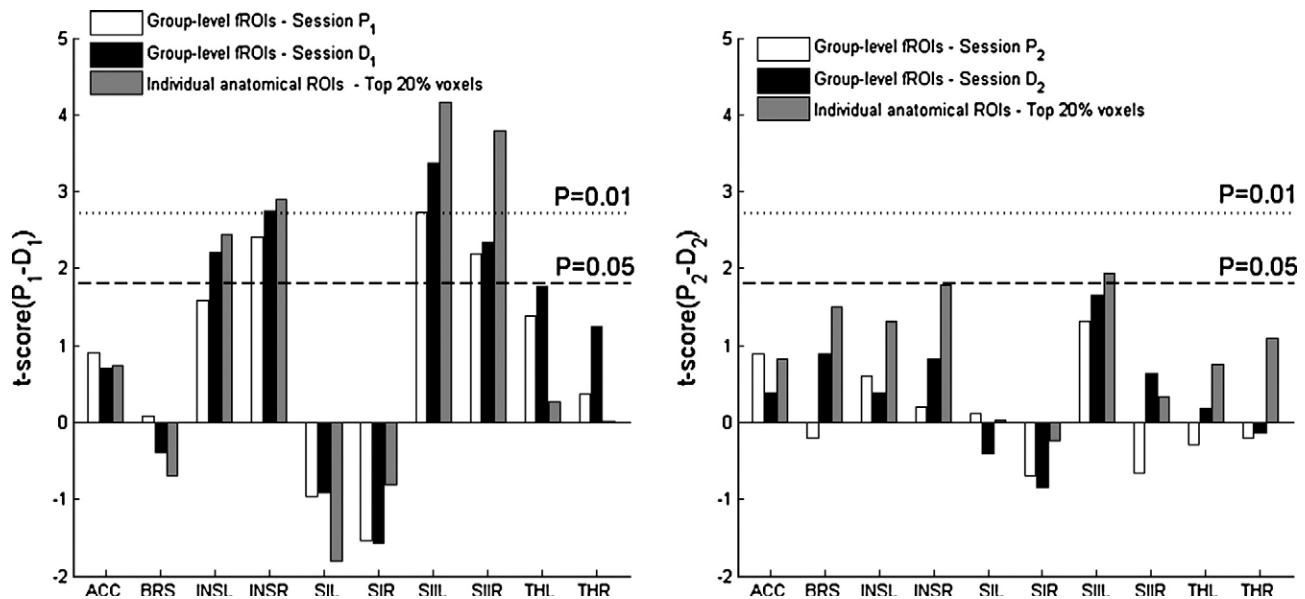


Fig. 11. Effect of group-level, individual-session fROIs on assessment of drug modulation: The bias resulting when selecting fROIs from the sessions under comparison is not evident due to between-subject functional variability. Individual anatomical ROIs combined with the Top 20% SV yield higher sensitivity.

session D_1 yielded a statistically significant increase for left SII (Fig. 10, left panel). On the other hand, no consistent bias effects were observed for fROI defined from the remaining sessions (e.g., P_2 or D_2 when comparing P_1 – D_1); however, reduced sensitivity was found when compared to individual anatomical ROIs (Fig. 10).

The t -scores obtained by using individual-session fROIs from the group-level statistical images are shown in Fig. 11 (left panel: sessions P_1 and D_1 , right panel: sessions P_2 and D_2), along with the individual anatomical ROI t -scores (Top 20% SV). In this case, the bias induced by using fROIs from either of the sessions under comparison is not evident, mainly because of the within-session, between-subject functional variability. In almost all cases, the group-level fROIs (e.g., SII) yielded lower sensitivity. Finally, we

show the results from the union fROIs, both first-level and group-level, in Fig. 12. The obtained t -scores between sessions P_1 – D_1 are similar for the first-level and group-level masks, however more differences are observed between sessions P_2 – D_2 . Overall, the individual anatomical ROIs yield more sensitivity in this case as well under our experimental hypothesis.

Discussion

We have examined the effects of different ROI definition criteria, both anatomical and functional, on the results inferred from a pharmacological fMRI experiment. When ROIs were defined by anatomical criteria, the use of the average intensity of the top 20% voxels ($\beta_{\text{mean},20}$) as a summary value (SV) increased

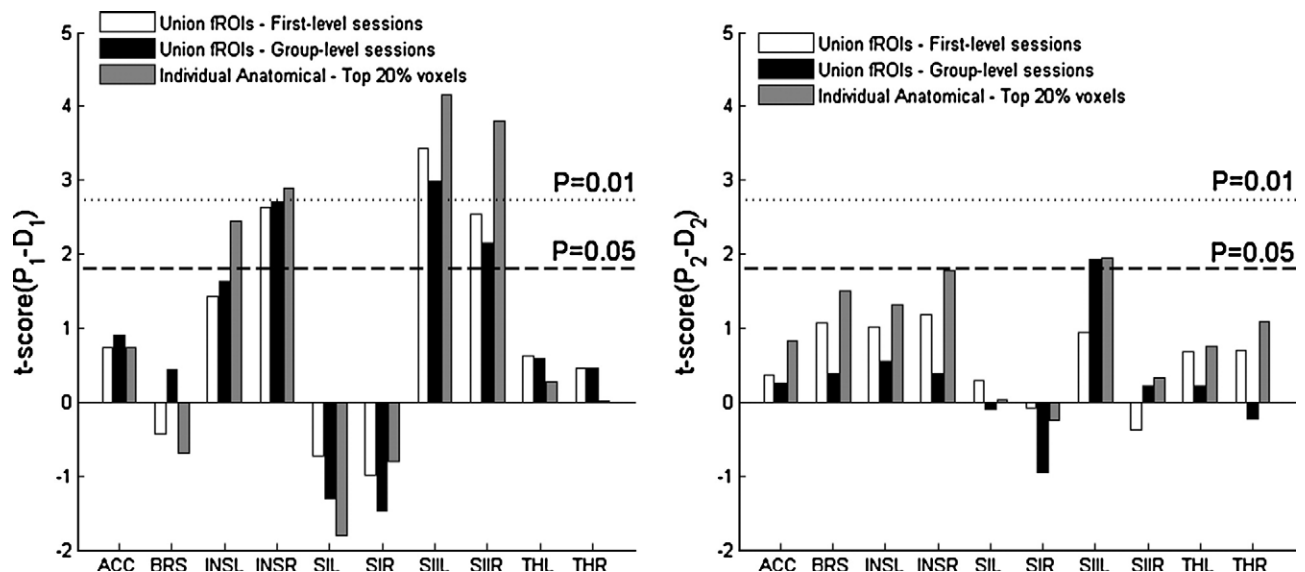


Fig. 12. Effect of first-level and group-level union fROIs on assessment of drug modulation: Individual anatomical ROIs combined with the Top 20% SV yield higher sensitivity.

sensitivity compared to the mean voxel intensity (β_{mean}). The results obtained using individually defined, group-probabilistic and standard-atlas anatomical ROIs were generally similar, except for regions that do not align well across subjects (e.g., SI, SII). Since individual ROIs account for between-subject variability in sulcal and gyral location, which may be averaged out by spatial normalization or by defining group-based ROIs, their use provides more accuracy when detecting an experimental effect (Brett et al., 2002). However, their definition may be demanding and time-consuming. An alternative, though less accurate approach would be to use group-based ROIs without constraining their size excessively, in order to include all possible activations by exploiting the approach of selecting the top fraction of active voxels; when using the latter approach, the size of anatomically-defined ROIs did not affect sensitivity in detecting drug effects to a large extent (Figs. 7 and 8). The use of functional ROIs from individual sessions, themselves included in the ROI comparisons, introduced unacceptable biases in the results (Figs. 9 and 10). Therefore, in cases where prior anatomical hypotheses are lacking, activation consistency between sessions should be established before using fROIs from individual sessions, either within the main experimental design or from a separate session (functional localisers).

Statistical sensitivity is crucial in drug development studies that employ functional imaging, since decisions for further compound development, based on regulatory procedures, may rely on the acceptance or rejection of prior hypotheses, formulated in the form of ROIs (Wise and Tracey, 2006). In the pharmacological fMRI data examined here, brain regions associated with nociceptive and hyperalgesic stimuli (Apkarian et al., 2005; Tracey, 2005) were of particular interest. By comparing the effect of different criteria to define these regions, one of our aims was to provide guidelines for future similar studies.

Summary values

The use of a single summary value, such as β_{mean} or $\beta_{\text{mean},20}$, provides a straightforward way to assess the response of an ROI as a whole and reduces the number of statistical inferences to one. Furthermore, a top fraction measure, when properly selected, increases signal-to-noise ratio and consequently sensitivity, and it can account for spatially heterogeneous responses within a region. Indeed, $\beta_{\text{mean},20}$ yielded considerably higher *t*-scores than β_{mean} (Fig. 6), while accounting for multiple within-region activations (Fig. 5) and between-subject functional variability. By using a fixed percentage of voxels (i.e. 20%) when calculating the SV, we included a significant fraction of the total voxels within an ROI, which may not be the case when an absolute threshold (e.g. *Z* value) is used. Moreover, ROI responses between different experimental conditions were directly comparable, since they were obtained from the same number of voxels. It should be noted that the choice of the percentage value in this approach, which was determined by preliminary analyses hereby, may depend on the specific experimental data and the resulting signal-to-noise distribution. This could be addressed by using mixture models for the intensity profile within a region; however, application of such models may be difficult for smaller ROIs due to the availability of a limited number of intensity values. Also, the application of a random field correction to the ROI voxels (small volume correction), in order to select the activation threshold, is more accurate when the region shape is convex (i.e. sphere or

ellipsoid) and may lead to higher thresholds than appropriate for complex non-convex shapes (Worsley et al., 1996). Finally, we used the voxel regression parameter estimates β , instead of *t* or *Z* values, to calculate the SVs, because β values reflect physiological effect size and its between-subject variability better. Note however that the top fraction voxels were selected on the basis of their *Z* values, in order to account for the effects of noise.

In addition to the aforementioned summary measures, other approaches commonly used include calculating the average response within a region of predetermined shape (e.g. spheres) around the voxel showing maximum activity, or extracting spatial components from the region data (e.g. by principal components analysis) (Nieto-Castanon et al., 2003; Friston et al., 2006). However, the top fraction approach is easier to interpret and is flexible in terms of ROI size and shape. For example, spherical ROIs around the maximum activity voxel may not be appropriate for regions with irregular or elongated shapes (Eickhoff et al., 2006) and may moreover fail to account for multiple activation sites within larger regions.

Because of the lack of ground truth for assessing the relative performance of different ROI criteria, we compared the obtained results by hypothesizing a general drug-induced signal decrease in the 'pain-matrix', as shown in Iannetti et al. (2005). Moreover, quantification of the extent to which the results were affected outside this working hypothesis is important. Although the top fraction approach concentrates solely on activations within a region, it is possible to apply a fraction-based measure to detect changes in deactivations within a region, by considering voxels that exhibit the lowest signal values.

Anatomical ROIs

Anatomical ROIs permit truly *a priori* hypotheses, in contrast to functional ROIs. Inter-subject anatomical variability and normalization to a standard template may influence the sensitivity of ROI analyses, making ROIs defined on an individual basis a more reliable choice (Brett et al., 2002). Our results did not show clear differences between the individual, group-probabilistic and standard WFU ROIs or an obvious correlation between inter-subject anatomical variability and sensitivity (Figs. 4 and 6); however, in certain cases (e.g. bilateral SII), individual ROIs yielded higher sensitivity in detecting drug effects (Fig. 6). The correspondence between the individual, group-probabilistic and WFU ROIs was better for subcortical areas such as the thalamus, compared to cortical areas such as SII (Fig. 3). For example, the ACC, as defined in the standard WFU atlas, was more extensive compared to its individually-drawn counterparts, which were defined with reference to Vogt et al. (1995). Note also that the WFU atlas is based on a single subject, and does not account for population variability.

The results obtained from group-probabilistic regions of different sizes revealed a lack of consistent discrepancies in sensitivity (Fig. 8), despite the wide range of ROI sizes examined (Table 1). This suggests that the top fraction approach was not affected by ROI size to a large extent, preserving statistical sensitivity. Varying the size of a group-based ROI may result in missing activations that are present in individual ROIs (i.e. voxels that are included in the Top 20%) or including activations that are not; however, this did not seem to influence the obtained results, probably due to averaging effects. This implies that when group or atlas-based ROIs are used in combination with a top fraction SV,

the size of the former need not be constrained excessively (Figs. 7 and 8). Between-subject ROI size variability is particularly relevant in studies where the subject age range may be broad, or where a clinical population may be heterogeneous.

Functional ROIs

fROIs are useful when clear anatomical hypotheses are lacking, which was not the case in the present study. In general, this approach also requires some anatomical constraints, unless whole brain activity is examined. We chose to use the individually defined anatomical ROIs as constraints, in order to focus on the effects of functional variability alone, rather than inducing additional effects due to anatomical variability. The use of fROIs in the present study did not improve the overall sensitivity beyond the combination of individual anatomical ROIs with the Top 20% SV (Figs. 10 and 12). Moreover, fROIs defined from first-level, individual sessions introduced unacceptable biases when defined from within either of the sessions under comparison (Figs. 9 and 10). This extreme case was considered in order to demonstrate the amount of bias that may result from an inappropriately defined fROI, rather than because it is an approach that would be recommended or applied in practice. The results reflect the fact that, even though the same stimulus was used in all sessions, administration of gabapentin and/or capsaicin induced significant functional variability. This is also relevant in drug (or clinical decision making) trials, where large variability in the response (or non-response) to a treatment may be observed. We used the Top 20% active voxels to define fROIs in order to allow direct comparisons with anatomical ROIs; however, using different voxel fractions (e.g. as many as Top 50% voxels) or absolute thresholds introduced similar biases (data not shown). Also, using different anatomical constraints is not expected to change the nature of the above observations.

The bias resulting from using individual, first-level fROI was reduced by pooling data across subjects and sessions (or both — Figs. 11 and 12). The fact that no clear biases were observed for individual-session, group-level fROIs (Fig. 11) is a consequence of between-subject variability within the same session, which resulted in group-level maps that were considerably different than their first-level counterparts. If the activation patterns were similar across subjects, we would expect results similar to those shown in Fig. 10. The use of fROIs defined on the group level has been also found to affect test–retest measures in fMRI reliability studies (Swallow et al., 2003). By considering all voxels that were activated during any condition (union fROIs), we allowed a fair chance of voxels from any session to be included. However, including voxels that were activated during any condition yielded lower sensitivity overall (Fig. 12), especially at the group-level, for which the resulting fROIs were larger in general, due to the increased statistical power of the group analysis. The use of an absolute threshold to define first-level union fROIs resulted in regions of widely varying sizes for each subject. Therefore, when prior hypotheses are difficult to formulate in an anatomical context and/or separate localising sessions are not feasible, a union fROI approach may be more reliable at the expense of lower sensitivity.

Inter-session variability is an important issue in the design of functional imaging experiments. It is questionable how well the results from a single session represent the response of a subject (McGonigle et al., 2000). Several factors may contribute to this

variability, such as physiological factors, scanner variance and first-level preprocessing methodology (McGonigle et al., 2000; Smith et al., 2005). In case functional localisers are used, careful experimental design should be made, taking into consideration prior evidence about this variability as well as context invariance of the stimulus, which may introduce confounds (Friston et al., 2006). For example, category-specific activations in the visual cortex, where functional localisers have been used extensively, were found to be consistent within and between sessions (Peelen and Downing, 2005). On the other hand, the lateral frontal and superior parietal regions are implicated across a wide range of task contexts (Jiang and Kanwisher, 2003). However, in many cases, such as in drug or clinical studies, only single observations are feasible. In the present study extra sessions were not desirable; therefore, we did not explore the use of a separate functional localiser session.

In conclusion, we have examined the effects of using different anatomical and functional criteria to define ROIs on the results inferred from a pharmacological fMRI experiment in a systematic manner. The results presented here provide guidelines regarding hypothesis-based analysis of similar data and should be further tested across a range of studies to evaluate their sensitivity as clinical and/or pharmacological measures.

Acknowledgments

This work was partly supported by Pfizer Ltd. RGW is supported by a UK Medical Research Council Career Development Award. GDM was supported by Pfizer Ltd, by the European Social Fund and National Resources, Operational Program Competitiveness — General Secretariat for Research and Development, Greece (Program ENTER04). GDI is a University Research Fellow of The Royal Society, UK.

References

- Amunts, K., Weiss, P.H., et al., 2004. Analysis of neural mechanisms underlying verbal fluency in cytoarchitectonically defined stereotaxic space — the roles of Brodmann areas 44 and 45. *NeuroImage* 22 (1), 42–56.
- Amunts, K., Zilles, K., 2001. Advances in cytoarchitectonic mapping of the human cerebral cortex. *Neuroimaging Clin. N. Am.* 11 (2), 151–169.
- Apkarian, A.V., Bushnell, M.C., et al., 2005. Human brain mechanisms of pain perception and regulation in health and disease. *Eur. J. Pain* 9 (4), 463–484.
- Bartels, A., Zeki, S., 2005. The chronoarchitecture of the cerebral cortex. *Philos. Trans. R. Soc. Lond., B Biol. Sci.* 360 (1456), 733–750.
- Beckmann, C.F., Jenkinson, M., et al., 2003. General multilevel linear modeling for group analysis in FMRI. *NeuroImage* 20 (2), 1052–1063.
- Bellec, P., Perlberg, V., et al., 2006. Identification of large-scale networks in the brain using fMRI. *NeuroImage* 29 (4), 1231–1243.
- Bense, S., Stephan, T., et al., 2001. Multisensory cortical signal increases and decreases during vestibular galvanic stimulation (fMRI). *J. Neurophysiol.* 85 (2.), 886–899.
- Brett, M., Johnsrude, I.S., et al., 2002. The problem of functional localization in the human brain. *Nat. Rev., Neurosci.* 3 (3), 243–249.
- Collins, D.L., Neelin, P., et al., 1994. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J. Comput. Assist. Tomogr.* 18 (2), 192–205.
- Culham, J.C., Danckert, S.L., et al., 2003. Visually guided grasping produces fMRI activation in dorsal but not ventral stream brain areas. *Exp. Brain Res.* 153 (2), 180–189.

- Duvernoy, H.M., 1999. *The Human Brain: Surface, Three-Dimensional Sectional Anatomy with MRI, and Blood Supply*. Springer-Verlag, New York, NY.
- Eickhoff, S.B., Stephan, K.E., et al., 2005. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage* 25 (4), 1325–1335.
- Eickhoff, S.B., Heim, S., et al., 2006. Testing anatomically specified hypotheses in functional imaging using cytoarchitectonic maps. *NeuroImage* 32 (2), 570–582.
- Forman, S.D., Cohen, J.D., et al., 1995. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn. Reson. Med.* 33 (5), 636–647.
- Friston, K.J., Rotshtein, P., 2006. A critique of functional localisers. *NeuroImage* 30 (4), 1077–1087.
- Iannetti, G.D., Zambreanu, L., et al., 2005. Pharmacological modulation of pain-related brain activity during normal and central sensitization states in humans. *Proc. Natl. Acad. Sci. U. S. A.* 102 (50), 18195–18200.
- Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* 5 (2), 143–156.
- Jezzard, P., Buxton, R.B., 2006. The clinical potential of functional magnetic resonance imaging. *J. Magn. Reson. Imaging* 23 (6), 787–793.
- Jiang, Y., Kanwisher, N., 2003. Common neural mechanisms for response selection and perceptual processing. *J. Cogn. Neurosci.* 15 (8), 1095–1110.
- Johansen-Berg, H., Behrens, T.E., et al., 2004. Changes in connectivity profiles define functionally distinct regions in human medial frontal cortex. *Proc. Natl. Acad. Sci. U. S. A.* 101 (36), 13335–13340.
- Kanwisher, N., McDermott, J., et al., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17 (11), 4302–4311.
- Kourtzi, Z., Kanwisher, N., 2001. Representation of perceived object shape by the human lateral occipital complex. *Science* 293 (5534), 1506–1509.
- Lancaster, J.L., Woldorff, M.G., et al., 2000. Automated Talairach atlas labels for functional brain mapping. *Hum. Brain Mapp.* 10 (3), 120–131.
- Maldjian, J.A., Laurienti, P.J., et al., 2003. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *NeuroImage* 19 (3), 1233–1239.
- McGonigle, D.J., Howseman, A.M., et al., 2000. Variability in fMRI: an examination of intersession differences. *NeuroImage* 11 (6 Pt 1), 708–734.
- Nichols, T., Hayasaka, S., 2003. Controlling the familywise error rate in functional neuroimaging: a comparative review. *Stat. Methods Med. Res.* 12 (5), 419–446.
- Nieto-Castanon, A., Ghosh, S.S., et al., 2003. Region of interest based analysis of functional imaging data. *NeuroImage* 19 (4), 1303–1316.
- Peelen, M.V., Downing, P.E., 2005. Within-subject reproducibility of category-specific visual activation with functional MRI. *Hum. Brain Mapp.* 25 (4), 402–408.
- Saxe, R., Brett, M., et al., 2006. Divide and conquer: a defense of functional localizers. *NeuroImage* 30 (4), 1088–1096.
- Smith, S.M., Beckmann, C.F., et al., 2005. Variability in fMRI: a re-examination of inter-session differences. *Hum. Brain Mapp.* 24 (3), 248–257.
- Spiridon, M., Fischl, B., et al., 2006. Location and spatial profile of category-specific regions in human extrastriate cortex. *Hum. Brain Mapp.* 27 (1), 77–89.
- Swallow, K.M., Braver, T.S., et al., 2003. Reliability of functional localization using fMRI. *NeuroImage* 20 (3), 1561–1577.
- Talairach, J., Tournoux, P., 1988. *Co-Planar Stereotaxic Atlas of the Human Brain*. Stuttgart, Thieme.
- Tootell, R.B., Reppas, J.B., et al., 1995. Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *J. Neurosci.* 15 (4), 3215–3230.
- Tracey, I., 2005. Nociceptive processing in the human brain. *Curr. Opin. Neurobiol.* 15 (4), 478–487.
- Vogt, B.A., Nimchinsky, E.A., et al., 1995. Human cingulate cortex: surface features, flat maps, and cytoarchitecture. *J. Comp. Neurol.* 359 (3), 490–506.
- Wise, R.G., Tracey, I., 2006. The role of fMRI in drug discovery. *J. Magn. Reson. Imaging* 23 (6), 862–876.
- Woolrich, M.W., Ripley, B.D., et al., 2001. Temporal autocorrelation in univariate linear modeling of FMRI data. *NeuroImage* 14 (6), 1370–1386.
- Worsley, K.J., Marrett, S., et al., 1996. A unified statistical approach for determining significant signals in images of cerebral activation. *Hum. Brain Mapp.* 4 (1), 58–73.