Scaling global counts to a regional white matter reference volume for brain perfusion SPECT

Jeff S. Spence[a], Patrick S. Carmack[a], Wayne A. Woodward[a], Richard F. Gunst[a], William R. Schucany[a], and Robert W. Haley[*b]

June 15, 2004

[a]*Department of Statistical Science, Southern Methodist University, P.O. Box 750332, Dallas, TX 75275-0332, USA*

[b]*Department of Internal Medicine, Epidemiology Division, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Blvd., Dallas, TX 75390-8874, USA*

[*]Corresponding author. Robert W. Haley, M.D., Department of Internal Medicine, Epidemiology Division, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Blvd., Dallas, TX 75390-8874, USA

E-mail address: Robert.Haley@UTSouthwestern.edu (R.W. Haley).

Scaling voxel counts by a global mean is used in SPM of brain perfusion SPECT to remove subject-to-subject variability in analyses of regional blood flow. Global means, however, can be influenced by an experimental condition that elicits a widespread response and by extracranial counts such as facial activation, both of which may induce regional confounding. As an alternative, we evaluated scaling to a white matter reference volume, the centrum semiovale, since such regions have voxel counts proportional to cerebral blood flow but are less affected by experimental effects on gray matter and by extracranial activations. We compared the statistical power to detect regional changes after scaling to the global mean or to the median of the white matter count distribution in a repeated-measures experiment involving brain SPECT scans before and after a physostigmine challenge. Two white matter regions were considered, the centrum semiovale and the corpus callosum. With inherent power of 0.85 to detect a 2.5σ regional physostigmine effect, if the stimulus reduces the global mean by 2.5%, 5%, or 10%, power would decrease to 0.75, 0.63, or 0.37, respectively. Under white matter scaling, physostigmine reduced global mean counts in our study by an average of 1.15% ($P = 0.089$). Additionally, physostigmine reduced the correlation between global and coronal means from 0.93 to 0.67 ($P = 0.0005$). In an SPM2 two-sample t-test, a regional difference in physostigmine effect in the caudate head was 34 voxels in extent ($P = 0.067$) under global mean scaling and 70 voxels ($P = 0.008$) under white matter scaling. White matter scaling provides higher power than global mean scaling in perfusion SPECT studies where global means are affected by experimental conditions.

Keywords: tomography, emission-computed, single-photon; statistics; sensitivity and specificity; confounding factors; Persian Gulf syndrome; physostigmine

Introduction

Brain perfusion experiments involving single-photon emission computed tomography (SPECT) can be confounded by inter-individual differences in measured global cerebral blood flow (gCBF). These differences may have a physiological cause (e.g., metabolic and pharmacokinetic), or they may be non-physiological in nature (e.g., variation in radiotracer preparation and dosage). To detect an experimental change in regional cerebral blood flow (rCBF) across individuals or groups requires count normalization, a scaling of counts, to remove the global confounding.

A count normalization technique commonly employed is proportional scaling. This method is based on dividing the intensity of all intracranial volume elements (voxels) either by the mean of all voxels above a prescribed threshold or by the mean of all voxels that survive a pre-specified brain mask and then setting the scaled means equal to an arbitrary constant for all images. A mean calculated by either of these methods is referred to as a global mean. Since the global mean is proportional to measured cerebral blood flow, dividing all voxels by the global mean removes global confounding due to between-individual variability in gCBF. If the calculation of global mean includes voxels affected by an experimental condition, however, the scaling factor could be substantially correlated with an effect being measured in a specific region of interest (ROI), thus introducing regional confounding. This type of regional confound is mentioned by Frackowiak et al. (2004, p.620 and p.738). In particular, they point out that in this case it may be useful to find a different measure of gCBF, one that is based on regions unaffected by the pharmacologic stimulus. For example, if the effect of an experimental condition decreases gCBF and the global mean decreases as a result of the decrease in gCBF, then dividing by the global mean may diminish the statistical power to detect a local change in rCBF. The average of the scaled counts in the ROI may

consequently be similar across individuals or groups and not reveal the change in rCBF due to the experimental condition. The extent to which the correlation between a scaling factor and an experimental effect confounds a measured regional effect depends on the extent of experimental influence on gCBF. Larger areas of experimental influence will affect the global mean to a greater extent, resulting in greater regional confounding and a proportionally greater loss of statistical power.

An ideal scaling factor for removing artifactual global confounding is one that is proportional to measured global cerebral blood flow, but one that is not affected by experimentally evoked responses. The method proposed here takes advantage of the relative invariance of the gray matter to white matter activity ratio that is maintained during a known time interval and under normal conditions where ^{99m}Technetium-hexamethyl propyleneamine oxime (^{99m}Tc-HMPAO) is used to measure blood flow in brain perfusion SPECT (Leveille et al., 1992). Invariance of this ratio implies that a scaling factor based on the distribution of white matter voxel counts is proportional to gCBF. Thus, in the absence of white matter pathology, proportional scaling using a factor based on the white matter count distribution is suitable for the removal of global confounding. More importantly, since many brain pathologies affect gray matter regions and many experimental conditions target only gray matter ROIs, a scaling factor based on only white matter will, in these cases, be independent of experimentally induced changes in regional gray matter blood flow, and hence will not contribute to regional confounding.

In this paper we explore quantitatively the effect of scaling with reference to white matter regions and the effect of scaling with reference to the whole brain on the ability to detect experimental changes in rCBF. In addition, we address difficulties specific to brain perfusion

SPECT that arise from defining a global mean by thresholding methods. Finally, we compare the suitability of different white matter regions as the reference volume for global count normalization.

Methods

Global mean calculation

Effective global mean scaling requires the calculation of a mean for brain voxels only. Since low-resolution SPECT images retain relatively high counts near but outside brain tissue, an area commonly referred to as the "corona," a technique to discriminate between brain voxels and non-brain voxels is needed.

One automated procedure, designated thresholded global mean and implemented in software packages such as SPM2 (http://www.fil.ion.ucl.ac.uk/spm/spm2.html), defines brain voxels by thresholding counts in the following manner. For each image a grand mean is calculated based on all voxels within the bounding box of the SPECT scan. All voxels not exceeding the threshold, typically $1/8^{th}$ of the grand mean, are considered non-brain tissue and are disregarded. The thresholded global mean is defined as the arithmetic mean of voxels above this threshold.

Another approach in current use, designated masked global mean, defines brain voxels by applying a single, pre-specified brain mask to the spatially normalized SPECT image. This could be a binary mask or a probability mask, which down-weights counts within the corona. SPM2 provides a probability mask that can be used to calculate a masked global mean, defined as the weighted average of all voxels remaining after the application of the brain mask.

Regional white matter localization

For a region of white matter to be suitable for use as a standard in count normalization, it must contain a large volume of white matter well removed from gray matter, ventricles, and extracranial matter to exclude non-white matter voxels and minimize partial volume effects, a particular hazard of low-resolution SPECT images. Two white matter regions meeting these requirements were considered: the centrum semiovale and the corpus callosum.

The location of centrum semiovale voxels was isolated in the following way. First, a white matter probability mask, provided by SPM2 and derived from the Bayesian segmentation algorithm implemented in SPM2, was spatially normalized to the same stereotactic space as the SPECT images to be analyzed. This resulted in a bounding box of volume 79 x 95 x 69 voxels and a voxel size of 2 x 2 x 2 mm^3. Next, a binary mask was created from the original white matter probability mask by thresholding the weights at 0.93 and restricting the lower limit of the transverse plane to a value superior to the lateral ventricle. The voxels remaining were exclusively within the centrum semiovale with a volume of approximately 1,600 voxels (Figure 1A).

The corpus callosum coordinates were obtained first by a query to the Talairach Daemon (Lancaster et al., 2000) available online (http://ric.uthscsa.edu/projects/talairachdaemon.html). This is a digital atlas that provides the spatial coordinates of many anatomical landmarks based on the Talairach atlas. Due to the poor agreement between this Talairach coordinate space and the MNI space to which the SPECT scans are spatially normalized in SPM2, an affine transformation was used to transform the Talairach coordinates from the digital atlas to coordinates that map to the MNI space for deep brain regions near the lateral ventricle (Carmack et al., 2004). See Brett et al. (2001) for a similar transformation suitable for cortical regions of the brain

(http://www.mrc-cbu.cam.ac.uk/personal/matthew.brett/abstracts/MNITal/mniposter.pdf). The voxels labeled as corpus callosum were those obtained after the Talairach-to-MNI transformation; the resulting volume was approximately 1,800 voxels (Figure 1B).

For each white matter region we defined the scaling factor for count normalization as the median intensity of all voxels in the region. The median intensity was selected instead of the mean to avoid the disproportionate partial-volume influence from relatively few extreme voxels at the boundary of the white matter region.

Subjects

The procedures reported in this paper were developed for analysis of a brain perfusion SPECT study of 39 Gulf War veterans conducted at the University of Texas Southwestern Medical Center at Dallas. Twenty-two of these veterans were classified previously as having one of three well-documented syndrome variants (designated syndromes I - III), and seventeen were classified as either Gulf War-deployed or non-deployed controls (Haley et al., 2000). Informed consent was obtained from all subjects in accordance with the Declaration of Helsinki, and ethical approval was granted by the University's Institutional Review Board.

Data

SPECT scans were obtained at three different time points: a baseline session (T1) taken two years prior to the controlled experiment and two sessions taken 48 hours apart (T2 and T3). In this study we were interested only in T2 and T3. In session T2 the brain images were obtained after an

intravenous saline infusion and at session T3 after infusion of a short-acting cholinergic drug, physostigmine, in order to identify possibly damaged regions in the brain. A cholinergic stimulus generally slows regional cerebral metabolism and blood flow. A question of interest in the Gulf War study is whether a *regional* cholinergic effect is diminished in subjects with one of the identified syndrome variants compared to that of controls. Therefore the issue is whether the count normalized T2 minus T3 voxel differences in a ROI are larger for control groups than syndrome groups.

Calculation of statistical power curves

Our first approach to comparing the relative usefulness of the two scaling methods was to calculate their effects on the statistical power to detect a greater rCBF in T2 than T3 images within a given 125-voxel ROI, modeled as a normal random field. Statistical power is defined as the probability of identifying an experimental effect when, in fact, one exists. To specify a realistic intervoxel covariance structure over the field for use in power calculations, we estimated semivariograms (Cressie, 1993) within the left thalamus for each member of the non-deployed control group, those most likely to have normal brain function (Haley, 2000).

Specifically, let $\boldsymbol{X}$ be a multivariate normal random vector representing T2 minus T3 differences in a 125-voxel ROI that is of particular experimental interest. The null and alternative hypotheses are, respectively,

$$H_0: \boldsymbol{\mu} \leq 0$$

$$H_a: \boldsymbol{\mu} > 0$$

where $\boldsymbol{\mu}$ is the mean T2 minus T3 difference vector of length 125 within the ROI and $\boldsymbol{\mu} > 0$ indicates each element $\mu_i > 0$, $i = 1, \ldots, 125$. Under the hypothesis $\boldsymbol{\mu} = 0$, the statistic $T = \boldsymbol{X}^T \sum^{-1} \boldsymbol{X}$ has a central χ^2 distribution, where $\sum$ is the covariance matrix of the random vector $\boldsymbol{X}$ obtained from the semivariogram model above. If we define R to be the corresponding rejection region of the test, then the power is calculated as $P\{T \in R\}$, the probability that the test statistic, T, is in the rejection region. Under the alternative hypothesis, T has a non-central χ^2 distribution with non-centrality parameter $\lambda = \frac{1}{2}\boldsymbol{\mu}^T \sum^{-1} \boldsymbol{\mu}$.

SPM analyses

Our second approach to comparing the two scaling methods was to determine empirically how each affected the SPM analysis of a suspected difference in rCBF in the Gulf War study data. Specifically, the T2 minus T3 differences for syndrome II and syndrome III were analyzed in SPM2 using the two-sample t-test model. In the first analysis all SPECT images were scaled by their respective masked global means; in the second analysis all SPECT images were scaled by their respective centrum semiovale medians. The analyses were similar in every other respect. We were interested in testing whether the syndrome II differences were less than the syndrome III differences, and we based our inference on corrected *P*-values at the cluster level. The scaling methods were assessed by comparing the *P*-values as well as the expected false discovery rates (*E[FDR]*) that were obtained from the test. Procedure designed to control the false discovery rate are new multiple comparison techniques that control the expected proportion of the rejected hypotheses that are falsely rejected (Genovese et al. 2002).

Calculation of corona means

The calculation of corona means requires a clearly defined boundary separating brain tissue from non-brain regions. We defined this boundary based on the same brain mask in SPM2 that was used for the calculation of the masked global means. By visual inspection, we chose the mask that considered all voxels with weights of at least 0.4 to be brain voxels, while those that did not meet this criterion were considered voxels outside of the brain, i.e. within the corona. Corona means were then defined as the arithmetic mean of all voxels with weights between (but not including) 0 and 0.4, as given by the probability brain mask.

Results

Global mean scaling can reduce statistical power

Since the cholinergic effect of physostigmine to inhibit function is widespread in gray matter regions throughout the brain (Blin et al., 1997), we expected the scaled global means in session T3 (physostigmine) to be smaller on average than those in session T2 (saline). In fact, in the Gulf War study data the difference between a subject's masked global mean at session T2 and T3, when scaled by the medians of the centrum semiovale, averaged a 1.15% reduction ($P = 0.089$ by a paired t-test). This indicates that the pharmacologic effect of physostigmine on specific gray matter ROIs that we wish to study is also reflected in the masked global mean, which is used to normalize all voxels to a common standard. If so, normalizing all images by the masked global

mean would be expected to reduce the power to detect the effect of the physostigmine treatment in discrete gray matter regions.

To illustrate, consider the effect of the two normalization approaches on the esitmated T2-T3 difference of a hypothetical ROI in a normal control subject. Assuming that the median of counts in white matter is not affected measurably by physostigmine but that the global mean, which includes both gray and white matter, is affected by the drug, we posit the scaling factors for normalization shown in Table 1. Further positing that physostigmine treatment causes a decrease in mean counts of the hypothetical ROI from 300 at T2 to 225 at T3, 2.5σ decrease in standard deviation units (Table 1), normalizing by the median white matter counts has no effect on the estimated T2-T3 difference in the subject's ROI, still 75 units or a 2.5σ decrease (Table 2). Normalizing by the masked global mean, however, reduces the estimated difference to 69.2, only a 2.3σ difference (Table 2).

To see the effect of this problem on statistical power, we show a set of power curves for the case in which the power to detect the T2-T3 difference of 2.5σ in the hypothetical ROI when normalizing with the median white matter counts is 0.85 (Figure 2). If physostigmine reduces the masked global mean counts by 2.5%, as in the illustration above, the power drops to 0.75. In the Gulf War study data, we found several subjects with decreases in the masked global mean of 5%, and a few where it decreased by 10%, suggesting that such effects on the global mean are plausible. At a 5% decrease in the masked global mean, the power would fall to 0.63, and with a 10% decrease the power would drop to 0.37. In all of these cases scaling by median white matter counts gives power of 0.85, which is substantially larger.

Corona voxels confound global mean scaling by automated thresholding

In proportional scaling of images by the thresholded global mean (in contrast to the masked global mean), SPM attempts to eliminate extracranial voxels (the corona) by an automated thresholding routine that excludes all voxels with activation below a threshold. Although automated thresholding is attractive to software users, it often retains large proportions of the corona that may include areas of intense facial activation. The relatively low default threshold in SPM (1/8th of the grand mean) indeed allows a large proportion of extracranial voxels in the calculation of the mean. If the intensity of coronal activation varies among subjects, automated thresholding essentially produces a different brain mask for each image, so that differences in coronal (facial) activation confound the calculation of the thresholded global mean.

In the Gulf War study we found substantial inter-subject and inter-time period variation in the intensity of activation of the corona, particularly in its facial component. With the SPM threshold (1/8th of the grand mean), a substantial volume of the corona was not excluded by thresholding (Figure 3). Raising the threshold to eliminate more of the corona excluded brain voxels also. The potential for confounding was indicated by between-subject differences in the size of the corona and in the intensity of coronal (particularly facial) activation by physostigmine at session T3 (Figure 3).

The potential for differences in coronal activation to confound the treatment effect was also illustrated by the fact that coronal activation was significantly more strongly correlated with thresholded global means in session T2 than in session T3 (Figure 4).

An example involving the caudate nucleus

In practice we have encountered cases for which normalization by median white matter counts yielded greater power than normalization by the masked global mean to detect group differences in the physostigmine effect for certain ROIs. One such example, involving the left caudate nucleus, is shown in Figure 5. An SPM2 two-sample t-test was run in which the images were scaled alternatively by the masked global means and by the medians of the centrum semiovale counts. The SPM2 analysis isolated a cluster within the left caudate nucleus that consisted of 34 voxels (corrected cluster-level $P = 0.067$) using global mean normalization and of 70 voxels (corrected cluster-level $P = 0.008$) using normalization to the white matter median in the centrum semiovale. The expected false discovery rates are 0.28 and 0.10, respectively. The same height threshold (3.79) was used in both analyses. The greater cluster size, smaller corrected *P*, and smaller *E[FDR]* obtained using white matter normalization illustrate how this approach provides substantially greater power to detect regions of abnormal activation than either method using the global mean.

Comparison of two white matter regions

The centrum semiovale and corpus callosum are two relatively large regions of white matter (approximately 14 cm^3) that were used as scaling referents in count normalization. Overlaying the voxels comprising these structures on an individual T1-weighted MRI image suggested that partial voluming, particularly with lateral ventricle voxels, was a greater concern with corpus callosum than with centrum semiovale in SPECT imaging (Figure 1).

Partial voluming in general should increase the variability of the voxel count distribution and, when parial-volumed with the lateral ventricle, should reduce the median white matter counts. In the raw SPECT images from the Gulf War study, the average standared deviation in the centrum semiovale (30.3) was 11.5 units smaller than in the corpus callosum (41.8, $p<0.0001$). Moreover, the average of the median counts in the centrum semiovale (194 units) was 17 units (9.6%) larger than that of the corpus callosum (177 units, $P = 0.0002$). The heterogeneity of effects from partial voluming among subjects is illustrated in Figure 6. These findings indicate that the centrum semiovale provides a more precise and less biased measure of cerebral blood flow than the corpus callosum, although restricting the calculation to the more centrally located voxels in the corpus callosum might reduce partial voluming and provide a standard comparable to the centrum semiovale.

Discussion

Count normalization in brain perfusion SPECT is best achieved by scaling counts relative to a regional white matter distribution, particularly for SPECT experiments in which small regions of interest are isolated and compared but experimental effects are widespread. A strict white matter reference value maximizes statistical power because white matter regions generally are not appreciably influenced by experimental protocol. In the Gulf War study, for example, physostigmine affects a large proportion of gray matter, and thereby, the global mean. Consequently, small regional changes of interest can be subsumed by the more general experimental effect, resulting in loss of power with global mean scaling. The degree to which statistical power is decreased depends on the amount by which the global mean is affected by the

experiment. If the pharmacologic stimulus affects only small brain regions, it will not appreciably affect the global mean. In this case, global mean scaling and white matter scaling will produce similar results, absent thresholding effects in the global mean. When the pharmacologic stimulus appreciably affects the global mean counts, however, our findings show that scaling to a white matter reference region minimizes potentially important losses of statistical power.

Statistical power arguments remain valid whether the global mean is calculated by masking or by thresholding. Thresholded global means should be used cautiously in brain perfusion SPECT due to the added confounding introduced by coronal activation. This problem cannot generally be overcome by simply using a higher threshold value to exclude more of the corona. When there is facial activation the voxel counts within that region can be higher than the global mean. In such cases, no amount of thresholding will exclude the corona without excluding a substantial proportion of brain voxels. Another option for removing global confounding is to include a parameter for the effect of global cerebral blood flow as a nuisance parameter in the linear model. This method can be implemented in SPM2. However, in the SPM2 implementation, this parameter is estimated by the global mean, giving it the same disadvantage as scaling by the global mean. It also reduces the degrees of freedom of statistical tests, which is ill-advised for the usual SPECT study that already has few degrees of freedom. Proportional scaling to an appropriate white matter reference region is a simple technique that increases statistical power while preserving degrees of freedom.

The centrum semiovale is the preferred white matter region because it is relatively large, easy to identify, and within these voxels partial voluming is minimized. However, scaling counts using strictly white matter regions depends on the absence of white matter pathology. Diseases such as multiple sclerosis, in which white matter is affected by demyelinating processes, affected white matter regions should not be used for scaling purposes. If extensive demyelination occurs

selectively within the centrum semiovale, then one should consider using the corpus callosum. In general, restricting the selected voxels to those more interior to the particular white matter region can minimize partial voluming. However, such a procedure can be accompanied by an increase in variability of the white matter median if the number of white matter voxels is small. In addition, we have chosen the median of white matter counts by which to scale, because the median is robust to a moderate amount of partial voluming. One is also not restricted to bilateral estimates of the centrum semiovale distribution. In the event of localized demyelination, a good estimate can be obtained unilaterally from the unaffected side. Neither is one restricted to these two regions. There is much flexibility in defining white matter regions relatively safely while minimizing the partial volume effect. Use of the centrum semiovale, however, should be feasible in the majority of cases, experimentally and clinically. Therefore, we recommend this region as the reference for count normalization in brain perfusion SPECT.

Acknowledgments

This study was supported by the U.S. Army Medical Research and Material Command cooperative agreement numbers DAMD17-97-2-7025 and DAMD17-01-1-0741 through a consortium agreement with the University of Texas Southwestern Medical Center at Dallas. The content of this paper does not necessarily reflect the position or the policy of the U.S. government, and no official endorsement should be inferred.

The authors would also like to thank Dr. Michael Devous of the University of Texas Southwestern Medical Center at Dallas and Dr. Herve Abdi of the University of Texas at Dallas for their valuable advice and assistance during the life of this project.

References

Blin, J., Ivanoiu, A., Coppens, A., De Volder, A., Labar, D., Michel, C., Laterre, E.C. 1997. Cholinergic Neurotransmission Has Different Effects on Cerebral Glucose Consumption and Blood Flow in Young Normals, Aged Normals, and Alzheimer's Disease Patients. NeuroImage 6, 335-343.

Brett, M., Christoff, K., Cusack, R., Lancaster, J. 2001. Using the Talairach Atlas with the MNI Template. NeuroImage 13, S85.

Carmack, P., Spence, J., Gunst, R., Schucany, W., Woodward, W., Haley, R. 2004. Improved Agreement Between Talairach and MNI Coordinate Spaces in Deep Brain Regions. NeuroImage 22, 367-371.

Cressie, N. 1993. Statistics for Spatial Data, revised edition. J. Wiley and Sons, New York.

Frackowiak, R.,Friston, K., Frith, C., Dolan, R., Price, C., Zeki, S., Ashburner, J., Penny, W. 2004. Human Brain Function, 2nd Ed. Elsevier Academic Press, London.

Genovese, C., Lazar, N., Nichols, T. 2002. Thresholding of Statistical Maps in Functional Neuroimaging Using the False Discovery Rate. NeuroImage 15, 870-878.

Haley, R., Marshall, W., McDonald, D., Daugherty, M., Petty, F., Fleckenstein, J. 2000. Brain Abnormalities in Gulf War Syndrome: Evaluation with 1H MR Spectroscopy. Neuroradiology 215, 807-817.

Lancaster, J., Woldorff, M., Parsons, L., Liotti, M., Freitas, C., Rainey, L.,

Kochunov, P., Nickerson, D., Mikiten, S., Fox, P. 2000. Automated Talairach Atlas Labels for Functional Brain Mapping. Human Brain Mapping 10, 120-131.

Leveille, J., Demonceau, G., Walovitch, R.C. 1992. Intrasubject Comparison Between Technetium-99m-ECD and Technetium-99m-HMPAO in Healthy Human Subjects. J Nucl Med 33(4), 480-484.

Figure legends

Fig. 1. Tranverse sections of an individual T1-weighted MRI, provided in SPM96. The voxels within the region of the centrum semiovale (A) and the corpus callosum (B) are marked in color scale: red for count intensities within white matter, orange for count intensities within gray matter, and yellow for count intensities within cerebrospinal fluid. Homogeneity in color indicates homogeneity in count intensities.

Fig. 2. Power curves quantifying the ability to detect a specified mean difference between session T2 and session T3 for a 125-voxel ROI in a normal control subject from the Gulf War study using the hypothesis testing procedure described in the text. As a mean difference between T2 and T3 increases, the ability to detect an experimental change in the ROI increases; i.e., the larger the change in signal, the more likely it is to be detected. However, if the global mean is affected by the global response to physostigmine, then a larger regional change in signal would be required to achieve a given statistical power using global mean scaling that the change required using the median of white matter voxel counts as a scaling factor. As the effect of physostigmine decreases the global mean by 2.5%, 5%, and 10%, shown here, the statistical power to detect a true 2.5σ difference is reduced from 0.85 to 0.75, 0.63, and 0.37, respectively.

Fig. 3. Differences in coronal activation between two subjects at session T2 (saline) and T3 (physostigmine) after masking by SPM's automated thresholding (default threshold 1/8th of grand mean). Brain is in gray scale, and the corona is in heat scale. Areas excluded by the automated thresholding are in white. The intense yellow area in Subject A at T3 is facial activation in the corona which is absent in Subject B.

Fig. 4. Scatterplots of standardized means showing all subjects in the Gulf War study at session T2 (saline) and T3 (physostigmine). A significant decrease in correlation occurs from T2 to T3 (P = 0.0005), suggesting that the effect of corona on the thresholded global mean differs in the two sessions.

Fig. 5. Results of an SPM two-sample t-test in which the images were scaled by the masked global means (A) or by the medians of the centrum semiovale counts (B). The cluster size within the left caudate nucleus (arrow) is 34 voxels (A) and 70 voxels (B), and the corrected P -values are, respectively, 0.067 and 0.008. The corresponding expected false discovery rates are 0.28 (A) and 0.10 (B). These results suggest an increased power when using white matter normalization.

Fig. 6. Nonparametric density estimates (i.e. smoothed histograms) showing the distributions of the intensities for voxels in the centrum semiovale (solid) and the corpus callosum (dashed) for three subjects in the Gulf War study. These three subjects were selected to show representative relationships that were observed between the two regions of white matter used in the study: similar medians, different variability (subject 1); different medians, similar variability (subject 2); and similar medians, similar variability (subject 3). However, across all subjects, the centrum semiovale had smaller variability, higher average medians, and density estimators that were more consistent.

Table 1

Table 1

Scaling factors used in count normalization and ROI means for a non-ill subject

	Session	
	T2 (saline)	T3 (physostigmine)
White Matter Median Counts	150	150
Masked Global Mean Counts	200	195
Mean Counts in ROI	300	225

Note: This example assumes that global mean blood flow and blood flow in the hypothetical ROI are reduced from session T2 (saline) to session T3 (physostigmine), while blood flow in white matter is not.

Table 2

Table 2

Scaling calculation showing the adjustments to the ROI means in T2 and T3 for both scaling methods

	Session					
	T2 (saline)		T3 (physostigmine)		T2-T3 difference in scaled ROI mean	
Scaling method	Scaling calculation	Scaled ROI mean	Scaling calculation	Scaled ROI mean	In absolute counts	In SD units
White matter median	$\frac{300}{150}\times 150$	300	$\frac{225}{150}\times 150$	225	75.0	2.5σ
Masked global mean	$\frac{300}{200}\times 200$	300	$\frac{225}{195}\times 200$	230.8	69.2	2.3σ

Note: The scaled ROI means are calculated using (observed ROI mean / scaling factor x multiplier) where the scaling method is either the white matter median or the masked global mean. The multiplier is used to set the masked global mean equal to 200 and the median of the white matter to 150 in both sessions to maintain the original T2 reading of 300 for the mean in the ROI for each scaling method. Global mean scaling artificially reduces the mean ROI difference from 2.5σ to 2.3σ, thereby reducing the statistical power to detect the difference.

Figure 1

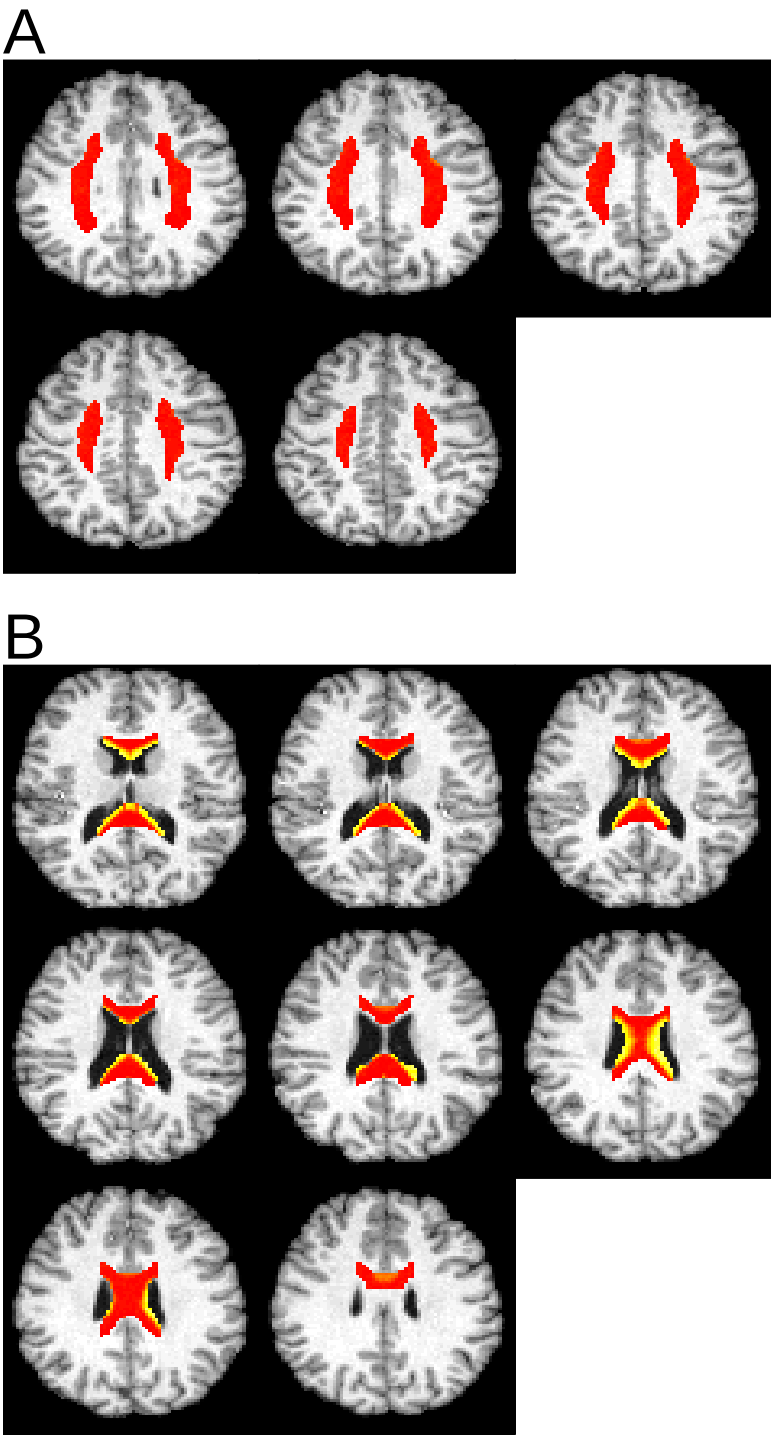


Figure 2

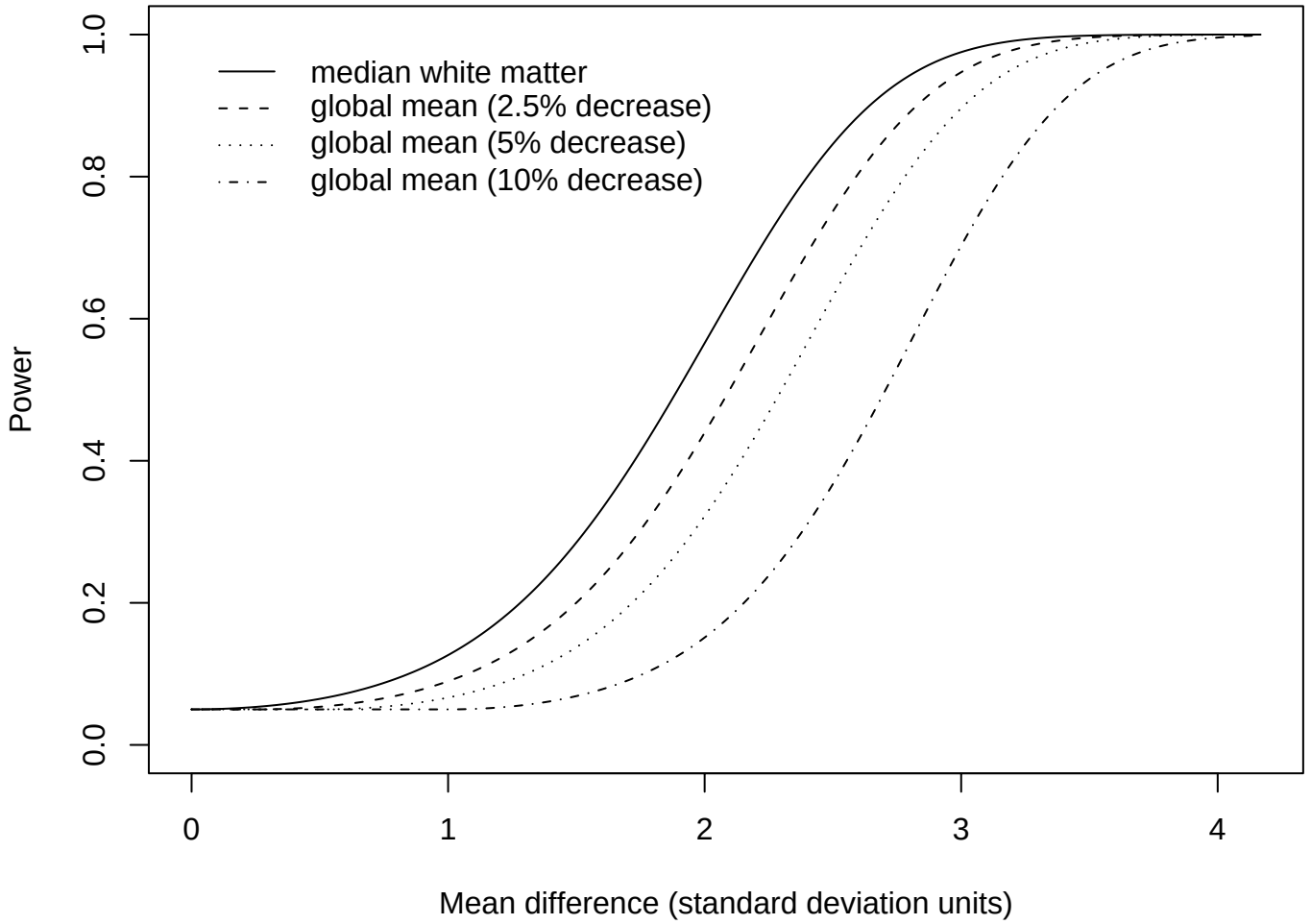

median white matter
global mean (2.5% decrease)
global mean (5% decrease)
global mean (10% decrease)
Power
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
Mean difference (standard deviation units)

Figure 3

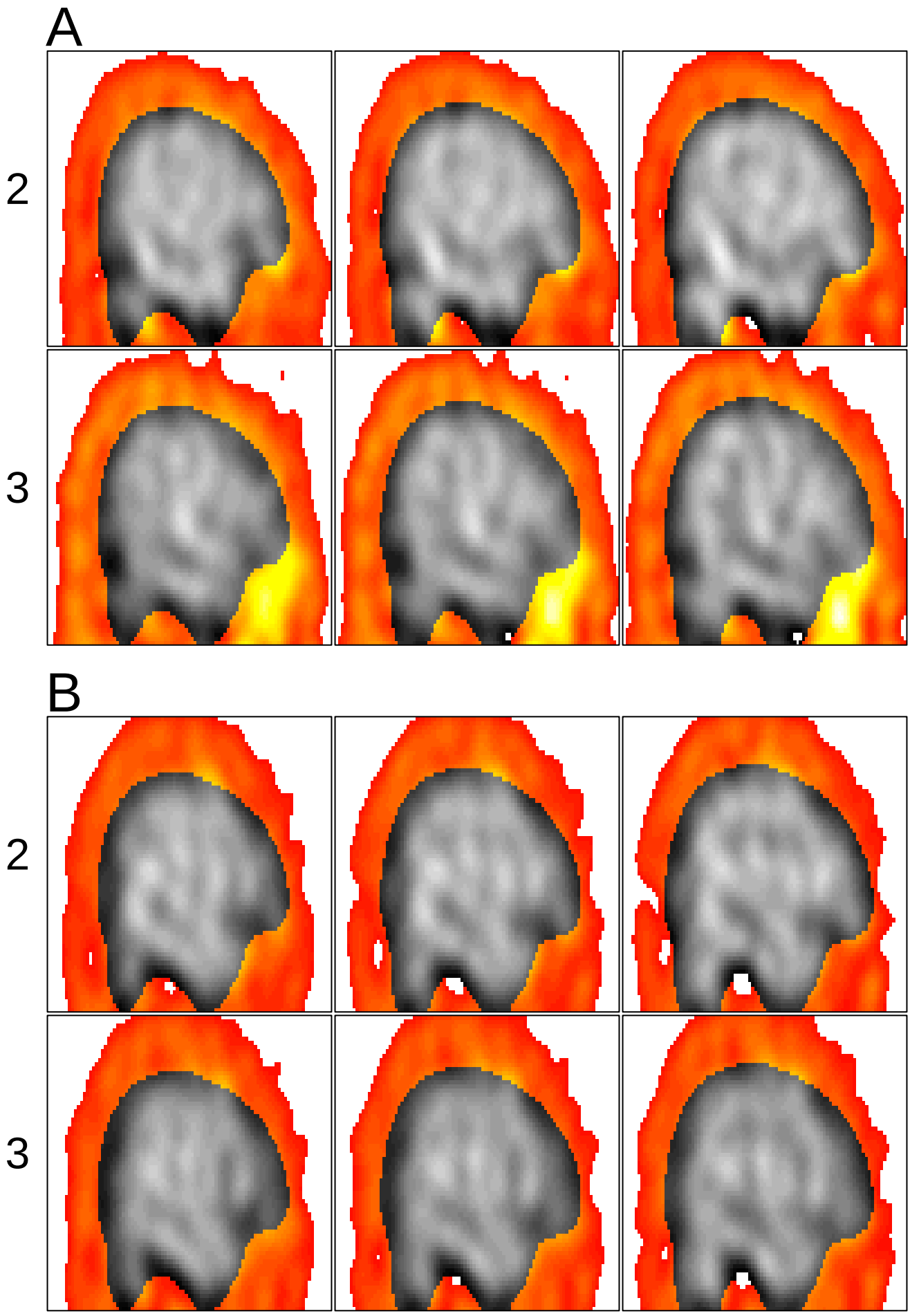


Figure 4

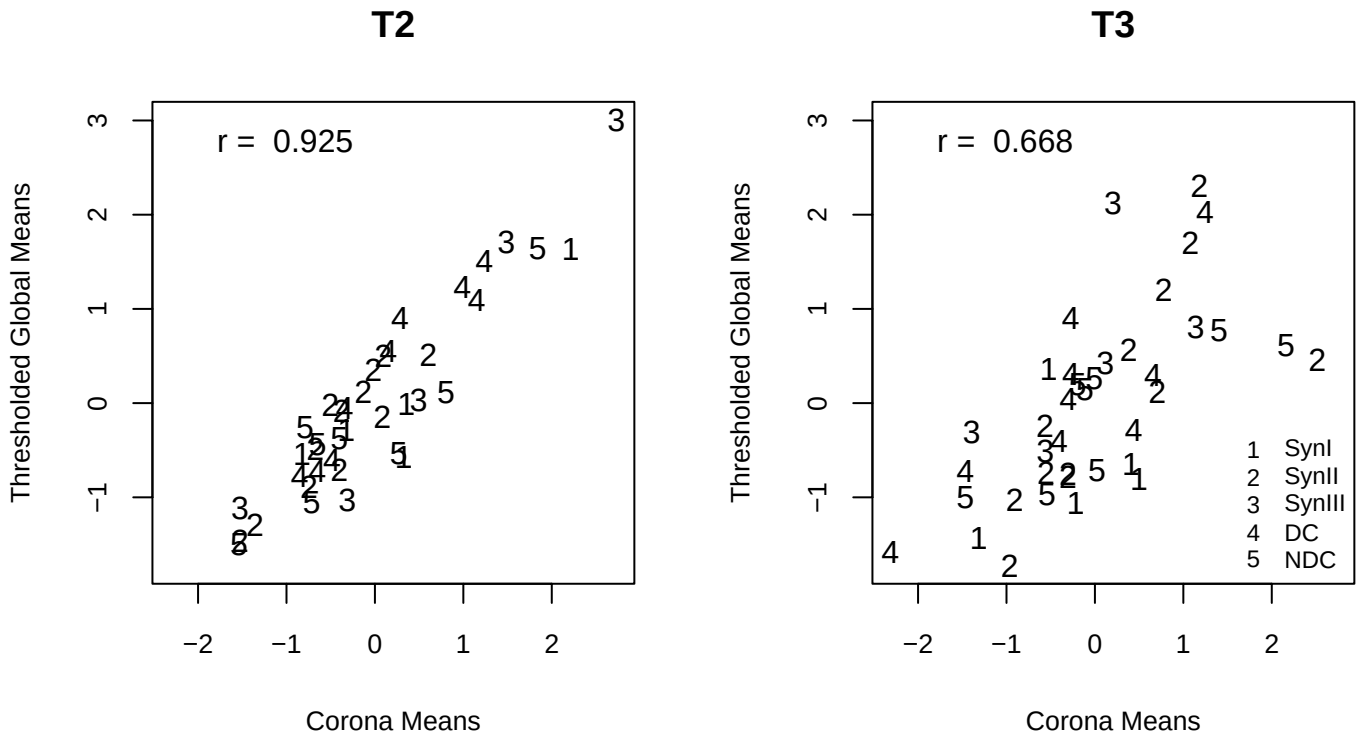
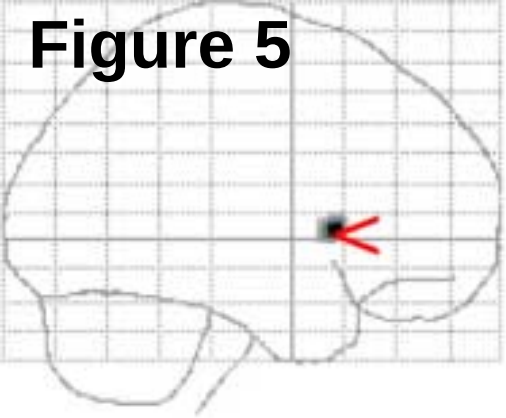
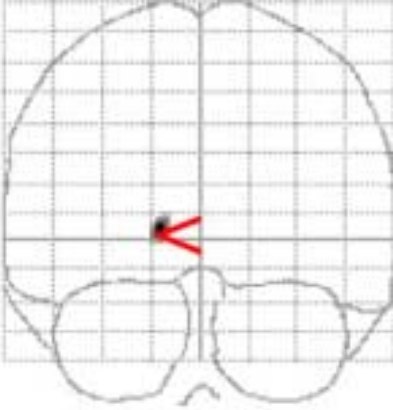
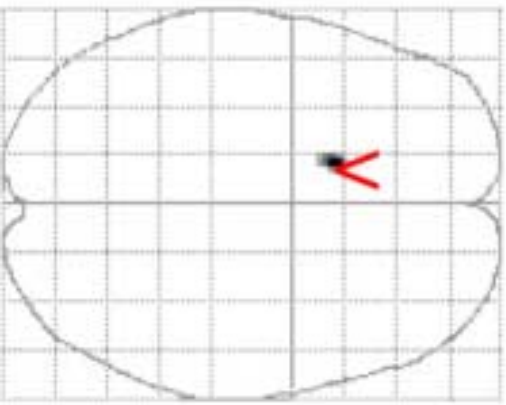

T2
r = 0.925
Thresholded Global Means
Corona Means
T3
r = 0.668
Thresholded Global Means
Corona Means
1 SynI
2 SynII
3 SynIII
4 DC
5 NDC

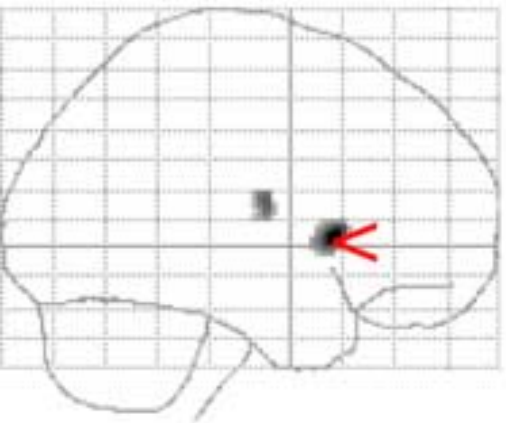
Figure 5

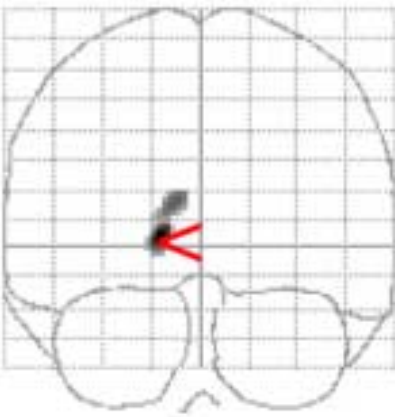


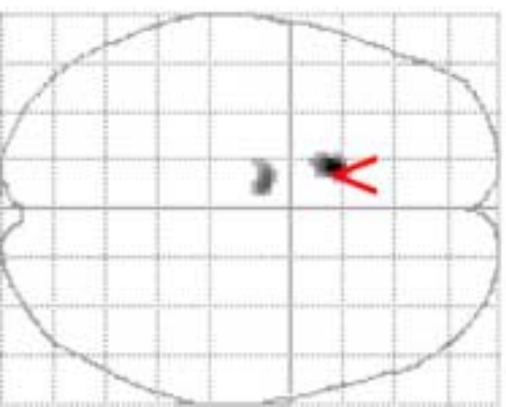




A

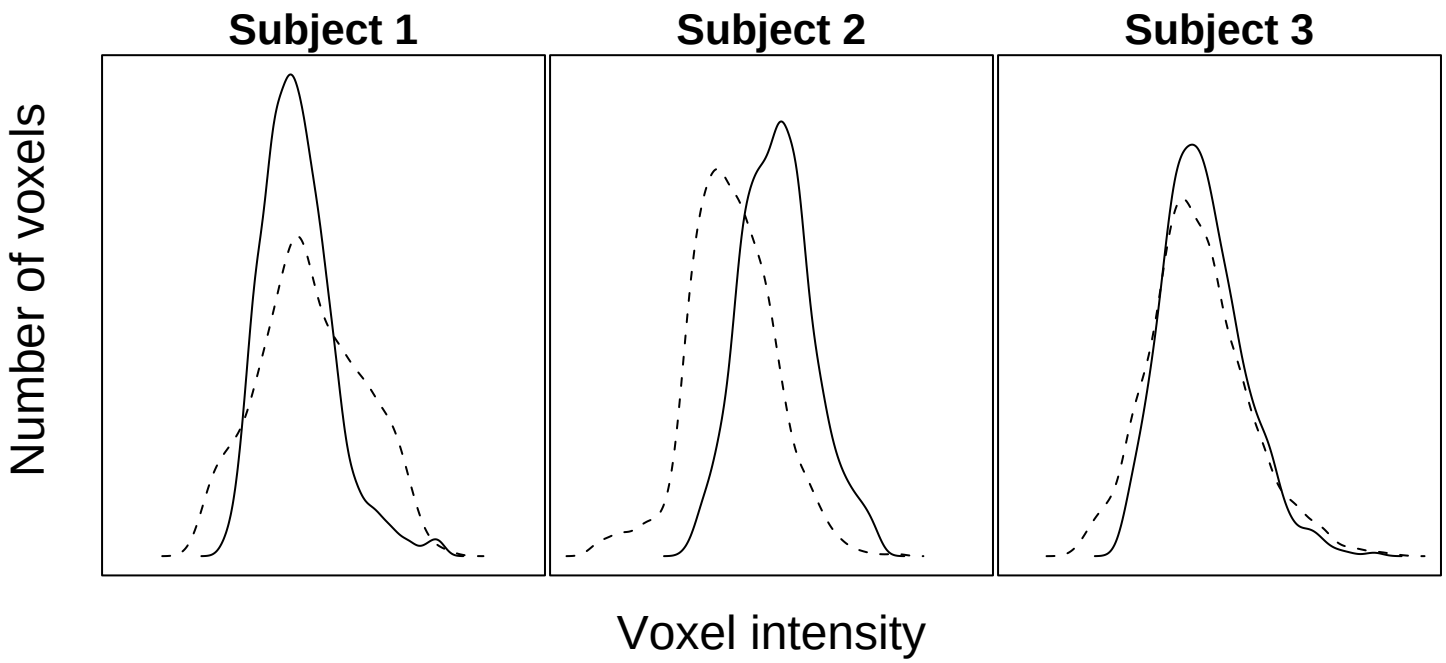






B

Figure 6


Subject 1
Subject 2
Subject 3
Number of voxels
Voxel intensity