Detecting and estimating the regions of activation in CBF activation studies in human brain

Keith J. Worsley[a], Alan C. Evans[b], Sean Marrett[b] and Peter Neelin[b]

[a]Department of Mathematics and Statistics, McGill University, 805 Sherbrooke Street, Montreal, Quebec, Canada H3A 2K6

[b]McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, 3801 University Street, Montreal, Quebec, Canada H3A 2B4

INTRODUCTION

Many studies of brain function with positron emission tomography (PET) involve the interpretation of a subtracted PET image, usually the difference between two images of cerebral blood flow (CBF) under baseline and activation conditions. In many cognitive studies, the activation is so slight (4-8%) that the experiment must be repeated on several subjects. The images are then mapped into a standardised coordinate space to account for differences in brain size and orientation, and the subtracted images averaged to improve the signal to noise ratio. The averaged ΔCBF image is then normalised to have unit variance and the resulting t-statistic image is searched for local maxima. If the between-subject variance is not demonstrably (significantly) different across all voxels then the normalisation can be based on a pooled estimate of the between subjects variance. If this is not so then a voxel-based normalisation must be used. We describe a method for determining if the population standard deviation image has regions of high or low values. If these are detected then we give an approximate P-value for the global maximum of the voxel-based t-statistic. We propose an estimator of the number of regions of high or low standard deviation, and the number of regions of activation in the voxel-based t-statistic image. The method uses the Euler characteristic, a concept borrowed from topology. We can thus determine not only if any activation has taken place, but also how many isolated regions of activation are present.

THEORY

t-statistic images

In a typical CBF activation study, PET image data are collected from n normal volunteers under an activation condition A and a baseline condition B. Let $A_i(x, y, z)$ and $B_i(x, y, z)$ be the cerebral blood flow (CBF) of subject i, $i = 1, \ldots, n$ at voxels with coordinates (x, y, z), under conditions A and B respectively, standardized to have the same mean of 100. The ΔCBF image for subject i is

$$\Delta_i(x, y, z) = A_i(x, y, z) - B_i(x, y, z),$$

and the average of the ΔCBF images, and their between subjects variance, are

$$\begin{aligned} M(x, y, z) &= \sum_{i=1}^{n} \Delta_i(x, y, z)/n, \\ S^2(x, y, z) &= \sum_{i=1}^{n} \{\Delta_i(x, y, z) - M(x, y, z)\}^2 /(n-1), \end{aligned}$$

respectively. If the variance is not significantly different across all voxels then we can obtain a better estimte of the underlying population variance by pooling the variance over the N voxels in the search volume to obtain

$$\bar{S}^2 = \sum_{x,y,z} S^2(x,y,z)/N.$$

The t-statistic based on this pooled standard deviation is

$$T^{\mathrm{P}}(x,y,z) = M(x,y,z)/\bar{S} \times \sqrt{n}.$$

If the ΔCBF images $\Delta_i(x,y,z)$ have a normal distribution, the population standard deviation is constant over all voxels, and the volume is large with respect to the effective FWHM then the degrees of freedom of $T^{\mathrm{P}}(x,y,z)$ is very large and its distribution can be well approximated by a normal distribution. Under the same conditions, the distribution of $(n-1)S^2(x,y,z)/\bar{S}^2$ can be well approximated by a χ^2 distribution with $n-1$ degrees of freedom. If there is strong evidence that the standard deviation is not constant over voxels, then we can use a voxel-based t-statistic with $n-1$ degrees of freedom

$$T^{\mathrm{V}}(x,y,z) = M(x,y,z)/S(x,y,z) \times \sqrt{n},$$

which has a t-distribution with $n-1$ degrees of freedom when no activation is present.

P-value of the maximum

We define R, the number of 'resels', to be the search volume V divided by the product of the full width at half maxima in the x, y and z directions of the PET camera:

$$R = V/(\mathrm{FWHM}_x \times \mathrm{FWHM}_y \times \mathrm{FWHM}_z).$$

If there is no population difference in normalised CBF between the activation condition and the baseline condition then the P-value of the maximum pooled t-statistic $T^{\mathrm{P}}_{\max}$ is approximately

$$\mathrm{P}(T^{\mathrm{P}}_{\max} \geq t) \approx R(4\log_e 2)^{\frac{3}{2}}(2\pi)^{-2}(t^2-1)e^{-\frac{1}{2}t^2}, \qquad (1)$$

for $R > 30$ and $t > 3$ (Adler, 1981, and Worsley et al., 1992). This result is valid provided $M(x,y,z)$ is normally distributed, the effective reconstructed resolution is spatially invariant, and most importantly, the population standard deviation is constant over all voxels. To test this last assumption, let $S_{\max}$ and $S_{\min}$ be the maximum and minimum of the standard deviation image inside the search volume. Then if the population standard deviation is constant over all voxels

$$\mathrm{P}(S_{\max} \geq s) \approx R(4\log_e 2)^{\frac{3}{2}} \frac{u^{\frac{1}{2}(n-4)}e^{-\frac{1}{2}u}}{(2\pi)^{\frac{3}{2}}2^{\frac{1}{2}(n-3)}\Gamma\left(\frac{n-1}{2}\right)}\{u^2-(2n-3)u+(n-2)(n-3)\}, \qquad (2)$$

for large R and s, where $u = (n-1)s^2/\bar{S}^2$ (Worsley, 1994). If $n \geq 5$ then $\mathrm{P}(S_{\min} \leq s)$ is approximated by the right hand side of (2) for large R and small s (Worsley, 1994). If the

data fails this test then we can use the maximum voxel-based t-statistic $T^{\mathrm{V}}_{\mathrm{max}}$. If $n \geq 5$ and R and t are large then its approximate P-value is

$$P(T^{\mathrm{V}}_{\mathrm{max}} \geq t) \approx R(4\log_e 2)^{\frac{3}{2}}(2\pi)^{-2}\left(\frac{n-2}{n-1}t^2 - 1\right)\left(1 + \frac{t^2}{n-1}\right)^{-\frac{1}{2}(n-2)}. \qquad (3)$$

For a typical 1000cc search volume with FWHM=20mm in each direction, so that the number of resels is $R = 125$, some threshold t-values calculated by equating the right hand side of (3) to α and solving for t, are shown in Table 1 for typical false positive rates $\alpha = 0.1$, 0.05 and 0.01. The increased noise from the voxel standard deviation greatly increases the threshold values as the number of subjects decreases. Note however that both P-values (2) and (3) are sensitive to the assumption of a normal distribution for ΔCBF values. However the P-value (1) is less sensitive to this, since the Central Limit Theorem gives some assurance that the average ΔCBF, $\Delta(x, y, z)$, will be approximately normal.

TABLE 1. *Approximate threshold t values of $T^{\mathrm{V}}_{\mathrm{max}}$ from (3) for $R = 125$ resels.*

No. of subjects	False positive rate α		
n	0.1	0.05	0.01
< 5	∞	∞	∞
5	877	1754	8771
6	54.0	76.4	171.0
7	21.8	27.6	47.4
8	14.0	16.8	25.4
9	10.8	12.5	17.6
10	9.1	10.3	13.8
∞	3.91	4.11	4.53

Small sample sizes

A curious phenomenon happens when the number of subjects is less than five: the standard deviation is very likely to be *exactly* zero somewhere in the image, which makes the voxel t-image infinite. This has nothing to do with statistics *per se*, and all to do with *continuity*, the property that the difference between ΔCBF values at adjacent voxels tends to zero as the voxel size tends to zero. To see why this is so, suppose there are four subjects and consider the four residual images formed by subtracting the average ΔCBF image from each subject's own ΔCBF image; note that the sum of squares of these four residual images equals three times the variance image. Suppose the images are sampled very finely, with very small voxel sizes, then the set of voxels where a residual image is zero forms a continuous surface or contour in the brain. The contours from two residual images will intersect in curves, and three will intersect in points. If three residual images are zero at a point, the fourth must be zero since by definition the sum of the residuals over subjects is zero. This implies that the variance map will be exactly zero at one or more points. The expected number of such zero points is $R(4\log_e 2)^{\frac{3}{2}}/\pi^2$ (Worsley, 1994), which equals 58.4 for a V =1000cc brain volume and a 20mm FWHM in each direction.

This of course has a strong effect on the voxel-based t-statistic image; if the number of subjects is less than five then $\mathrm{P}(T^{\mathrm{V}}_{\mathrm{max}} = \infty)$ approaches one as the resels becomes large. In

practice, with finite size voxels, these zero points almost surely lie between the voxels, but interpolation and re-sampling the image will produce sharp fluctuations in peak values as the voxels approach the zero points of the variance image. This particular problem disappears for five or more subjects but the variance image is still prone to large fluctuations which affect the stability of the location and intensity of voxel-based t-statistic images (Taylor et al, 1992).

The number of regions of activation

Our estimator is based on the topological concept of the Euler characteristic of the excursion set, the set of voxels for which the image is greater than a fixed threshold. The Euler characteristic essentially counts the number of isolated parts of the excursion set, irrespective of their shape, minus the number of 'holes'. It is easily calculated only using the values of adjacent voxels (Worsley et al., 1992). The exact expected Euler characteristic under no activation is given by the right hand sides of (1), (2) and (3) for the pooled t, standard deviation and voxel-based t images, respectively. We first equate the expected Euler characteristic to a small value, say 0.05, and solve for the threshold value. The observed Euler characteristic of the excursion set above this threshold is then our estimator of the number of regions of activation. Thus our estimator picks out the isolated peaks of activation in the signal, while protecting against finding any false positive regions in the unactivated parts of the image.

Results

Data acquisition

We shall illustrate the theory using data from a study of activation by a painfull activation (Talbot et al., 1991, and Worsley et al., 1992). PET scans were obtained using the Scanditronix PC-2048 system which produces 15 image slices 6.5mm apart with a transverse image resolution of 4.6-6.4mm and an axial resolution of 5.4-7.1mm. Using the bolus $H_2^{15}O$ methodology without blood sampling, the relative distribution of cerebral blood flow was measured in baseline and activation conditions. The 8 subjects underwent a procedure wherein a thermistor was applied to the forearm at both warm (42^oC) and high (48^oC) states, each condition being studied twice in each subject. For the present work, we analysed the difference images of the two warm conditions ('no activation') as a dataset which should have an expectation of zero throughout. We also analysed the difference between the average of the two hot conditions and the average of the two warm conditions ('activation') from each subject to search for activation due to the painful activation. One subject was scanned only once in the hot condition so this subject was dropped leaving 7 subjects for this dataset. Each subject in both studies also had an MRI scan containing 64 2mm-thick T_1-weighted multi-slice spin-echo images ($T_R = 550$msec ; $T_E = 30$msec) for later use in the 3-D analysis.

To overcome residual anatomical variability in subsequently transformed stereotactic images, a 20mm FWHM Hanning reconstruction filter was used. This has the effect of increasing signal to noise in the averaged image at the expense of resolution. In the experimental data sets studied here no axial filtering was used. Using a volumetric image registration procedure the MRI volume from each subject was aligned with the corresponding PET volume.

An orthogonal coordinate frame was then established based on the anterior commissure - posterior commissural (AC-PC) line as identified in the MRI image volume. These anatomical frame coordinates were used to apply a trilinear re-sampling of each matched pair of MRI and PET datasets into a standardized stereotactic coordinate system, $128 \times 128 \times 80$ voxels in extent and sampled at approximately 1.4mm, 1.7mm and 1.5mm along the x, y and z axes respectively.

Application to the 'no activation' dataset

The number of subjects was $n = 8$ and the standard deviation pooled over the search volume of $V = 1090\text{cm}^3$ was $\bar{S} = 5.93\%$. The resolutions of the image were $\text{FWHM}_x = 20\text{mm}$, $\text{FWHM}_y = 20\text{mm}$ and $\text{FWHM}_z = 7.6\text{mm}$ (the axial resolution was corrected for resampling and linear interpolation between the PET slices (Worsley et al., 1992)), which gives $R = 340$ resels. The observed maxima and minima of these fields, together with their P-values calculated from (1), (2) and (3) are given in Table 2. There is no evidence against the null hypotheses, apart from some evidence of regions of high standard deviation ($S_{\max} = 16.2$, $P < 0.01$).

TABLE 2. *P-values of maxima and minima for the pain study*

Dataset	Image	Minimum	P-value	Maximum	P-value
no activation	$T^{\text{P}}(x, y, z)$	–3.47	> 0.2	4.16	0.121
	$S(x, y, z)$	0.64	> 0.2	16.2	0.00013
	$T^{\text{V}}(x, y, z)$	–17.9	0.105	7.4	> 0.2
activation	$T^{\text{P}}(x, y, z)$	–4.32	0.066	4.99	0.0039
	$S(x, y, z)$	0.34	> 0.2	11.0	0.012
	$T^{\text{V}}(x, y, z)$	–42.6	0.037	15.6	0.72

The observed Euler characteristic of the excursion sets for several values of the threshold level, together with their expected values, are given in Figures 1, 3 and 5. To estimate the number of regions of activation of all images, the expected Euler characteristics given by the right hand sides of (1), (2) and (3) were equated to 0.05 and solved for the thresold values. The threshold values, together with the observed Euler characteristic of the images, are given in Table 3. The observed Euler characteristic of the standard deviation image above the upper tail threshold 13.7 was 2, indicating two regions of high standard deviation. These were located in the superior frontal region adjacent to the inter-hemispheric fissure. In all other cases the observed and expected Euler characteristic are in reasonable agreement.

TABLE 3. *Observed Euler characteristics at the 0.05 expected threshold*

Dataset	Image	Lower 0.05 threshold	Observed Euler	Upper 0.05 threshold	Observed Euler
no activation	$T^{\text{P}}(x, y, z)$	–4.39	0	4.39	0
	$S(x, y, z)$	0.30	0	13.7	2
	$T^{\text{V}}(x, y, z)$	–21.7	0	21.7	0
activation	$T^{\text{P}}(x, y, z)$	–4.39	0	4.39	1
	$S(x, y, z)$	0.10	0	10.5	1
	$T^{\text{V}}(x, y, z)$	–38.6	1	38.6	0

Application to the 'activation' dataset

One subject had missing values so $n = 7$, and the standard deviation pooled over the same search volume of $V = 1090$ cm^3 was $\bar{S} = 4.33\%$. Note that the pooled standard deviation is reduced by a factor of $0.73 \approx 1/\sqrt{2}$ from the 'no activation' study, since here two baseline and two activation images are averaged. The test statistics are given in Table 2, and they indicate that there has been an increase in mean activation ($T^{\mathrm{P}}_{\mathrm{max}} = 4.99$, $P < 0.01$) and at least one region of high standard deviation ($S_{\mathrm{max}} = 11.0$, $P < 0.02$). The test based on $T^{\mathrm{V}}_{\mathrm{max}}$ failed to detect an increase in blood flow, but did detect a decrease in blood flow ($T^{\mathrm{V}}_{\mathrm{max}} = -42.6$, $P < 0.05$).

Plots of the Euler characteristics, shown in Figures 2, 4 and 6, show considerable discrepancies between the observed and expected values of $T^{\mathrm{P}}(x, y, z)$ and $T^{\mathrm{V}}(x, y, z)$. The 0.05 threshold values of the expected Euler characteristics, together with the observed Euler characteristic of the images at these thresholds, are given in Table 3. Based on $T^{\mathrm{P}}(x, y, z)$ there is some evidence for one region of activation at the 0.05 level and three regions at the 0.1 level, identified in the anterior cingulate and somatosensory regions (Talbot et al., 1991). One region of significantly high standard deviation was located near the mesial edge of the thalamus at least 36mm from any of the three regions of high ΔCBF. The one region of 'de-activation' found using the voxel-based t-statistic is in the left frontal. Note that the 0.05 threshold value for the voxel-based t-statistic is much larger for 7 subjects (38.6) than for the 8 subjects (21.7) of the no activation data set. This may explain why no regions of activation are detected using $T^{\mathrm{V}}(x, y, z)$, despite the fact that there are some discrepancies between observed and expected Euler characteristic in the range $-5 < t < 5$.

Discussion

If we use the pooled t-statistic, and contrary to our assumptions there are some regions of large standard deviation then they will produce more false positives in the $T^{\mathrm{P}}_{\mathrm{max}}$ test (1); conversely regions of low standard deviation will make the $T^{\mathrm{P}}_{\mathrm{max}}$ test less sensitive. The voxel-based $T^{\mathrm{V}}_{\mathrm{max}}$ test (3) should in theory correct for this. However there is a high price to pay. First, there must be at least five subjects. Second, the test (3) is far more sensitive to normality of the subject ΔCBF vales. Third, the critical values are much higher, with a consequent loss in power at detecting activation, if in fact the population standard deviation is constant or approximately so. The critical issue in deciding between the pooled and the voxel-based t-statistic is whether there is any structure in the population standard deviation. Our test based on the maximum and minimum standard deviation (2) attempts to address this. However, failing to detect structure in the standard deviation image using test (2) does *not* mean that no structure is present, merely that we have failed to detect it. Following this with the pooled t-statistic test (1) does not guarantee a valid test. Nevertheless, it does indicate that a pooled t-statistic approach is no less justified than a voxel t-statistic approach. The consequent gain in detection power then favours the pooled t-statistic approach. As has been pointed out (Taylor et al., 1992) the use of voxel t-statistic methods also introduces peak localization difficulties since the location of minima in the standard deviation image is dependent on the choice of resolution.

We conclude with some notes of caution. This paper has proposed a test for regions of activation based on the maximum value. It suffers from the same problem of over-correction

as all tests, such as a Bonferroni-corrected test, which correct for searching over a volume of possible locations: the critical values of our test and estimator are much higher than the critical value for a pre-determined point in the image. The price paid for not knowing the location of the region of activation is therefore quite severe. In practice the investigator knows much more about which regions are likely to be activated. Experimenters using this test are thus faced with declaring that no activation has taken place in regions where activation seems 'meaningful' based on prior knowledge of brain function, but where the maximum of the image falls below the critical value. It is possible to overcome this by reducing the search volume to a smaller region containing just these likely regions, thus lowering the critical value, but these critical values do not decrease appreciably.

References

[1] Adler RJ. *The Geometry of Random Fields.* Wiley, New York, 1981.

[2] Worsley KJ, Evans AC, Marrett S, Neelin, P. A three-dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow and Metabolism*, 1992: 12: 900-918.

[3] Worsley KJ. Local maxima and the expected Euler characteristic of excursion sets of χ^2, F and t fields. *Advances in Applied Probability*, 1994: (in press).

[4] Taylor SF, Koeppe RA and Minoshima S. Regional variance in statistical maps can affect localization of cerebral blood flow activation foci. *Journal of Nuclear Medicine*, 1992: 33 (5) Supplement: 1008.

[5] Talbot JD, Marrett S, Evans AC, Meyer E, Bushnell MC, Duncan GH. Multiple representations of pain in human cerebral cortex. *Science*, 1991: 251: 1355-1358.

Figure 1. Pooled t-statistic, no activation

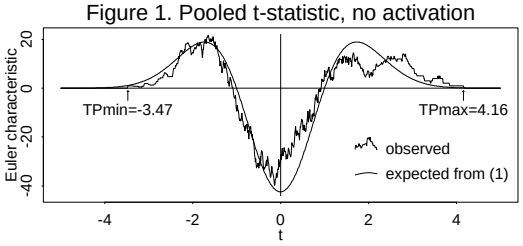

Figure 2. Pooled t-statistic, activation

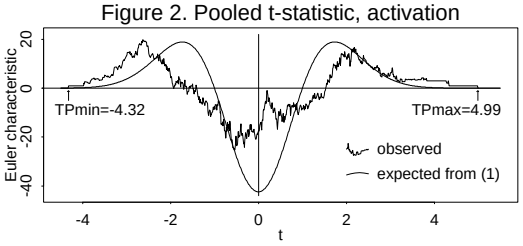

Figure 3. Standard deviation, no activation

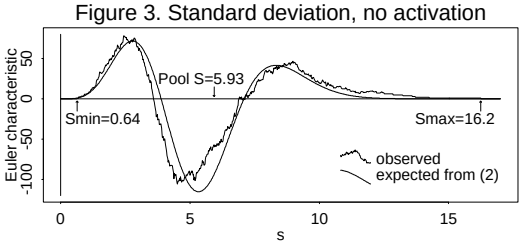

Figure 4. Standard deviation, activation

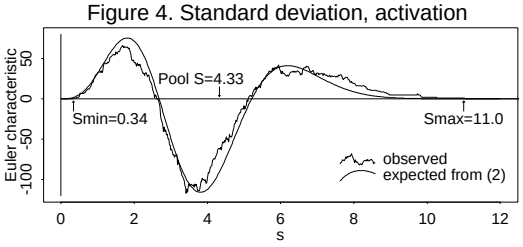

Figure 5. Voxel t-statistic, no activation

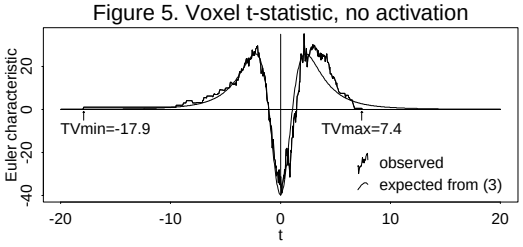

Figure 6. Voxel t-statistic, activation