# On Two Methods of Statistical Image Analysis

**J. Missimer,[1]* U. Knorr,[2] R.P. Maguire,[2] H. Herzog,[3] R.J. Seitz, L. Tellman,[3] and K.L. Leenders[1,4]**

[1]*PET Program, Paul Scherrer Institute, Villigen, Switzerland*
[2]*Department of Neurology, University of Düsseldorf, Germany*
[3]*Institute of Medicine Research Center, Jülich, Germany*
[4]*Department of Neurology, University of Zürich, Switzerland*

**Abstract:** The computerized brain atlas (CBA) and statistical parametric mapping (SPM) are two procedures for voxel-based statistical evaluation of PET activation studies. Each includes spatial standardization of image volumes, computation of a statistic, and evaluation of its significance. In addition, smoothing and correcting for differences of global means are commonly performed in SPM before statistical analysis. We report a comparison of methods in an analysis of regional cerebral blood flow (rCBF) in 10 human volunteers and 10 simulated activations. For the human studies, CBA or linear SPM standarization methods were followed by smoothing and computation of a statistic with the paired t-test of CBA or general linear model of SPM. No standardization, linear, and nonlinear SPM standardization were applied to the simulations. Significance of the statistic was evaluated using the cluster-size method common to SPM and CBA. SPM employs the theory of Gaussian random fields to estimate the cluster size distributions; simulations described in the Appendix provided empirical distributions derived from t-maps. The quantities evaluated were number and size of functional regions (FRs), maximum statistic, average resting rCBF, and percentage change. For the simulations, the efficiency of signal detection and rate of false positives could be evaluated as well as the distributions of statistics and cluster size in the absense of signal. The similarity of the results yielded by similar methods of analysis for the human studies and the simulated activations substantiates the robustness of the methods for selecting functional regions. However, the analysis of simulated activations demonstrated that quantitative evaluation of significance of a functional region encounters important obstacles at every stage of the analysis. *Hum. Brain Mapping 8:245–258, 1999.* 

**Key words:** positron emission tomography; PET; human brain; activation studies; spatial standardization; evaluation of significance

## INTRODUCTION

Determining regions of significant change in activation studies involving several individuals requires: (1) mapping the images into a standard reference volume, i.e., spatial standardization, and (2) statistical analysis consisting of computing a statistic for each voxel and evaluating its significance. The computerized brain atlas (CBA) [Seitz et al., 1990; Greitz et al., 1991] and statistical parametric mapping (SPM) [Friston et al., 1991a] were both designed to accomplish these tasks. Since mapping into a standard volume requires nonlinear as well as linear transformations of individual images, it is possible that the transformations influence the statistical evaluation. Moreover, two intermediate procedures between spatial standardization and statis-

*Correspondence to: J. Missimer, PET Program, Paul Scherrer Institute, 5232 Villigen PSI, Switzerland. E-mail: missimer@psi.ch
Received for publication 11 March 1999; accepted 18 June 1999

tical evaluation are commonly performed. Smoothing the transformed images and correcting for differences of the global means are commonly used in SPM to facilitate detection of local regions of significant change, or functional regions (FRs) as they are denoted in the following. Advocates of SPM originally suggested additional filters as large as 20 mm FWHM in plane [Friston et al., 1991a] in order to accommodate the resolution of the scanner as well as the known scale of individual variations in human brain anatomy [Steinmetz et al., 1989]. Proportional scaling and ANCOVA are the two options for correcting for differences in global means offered by SPM. Proponents of CBA employ no smoothing of images other than those performed at reconstruction [Roland et al., 1993; Wunderlich et al., 1997] and do not correct routinely for differences in global means.

The statistical models used to detect functional regions also differ. CBA employs a paired t-test, whereas SPM derives a t-test from the general linear model in which a block parameter for each individual is fitted to model the intersubject differences. To assess significance, two procedures and a conjunction of the two are employed. The method common to SPM and CBA, denoted spatial extent or cluster size, respectively [Roland et al., 1993; Friston et al., 1994; Worsley et al., 1996], follows from the probability that clusters of contiguous voxels exceeding a given search threshold occur by chance in the search volume. SPM employs the theory of Gaussian random fields to estimate the cluster size distributions from which assessment of significance is deduced. However, the distributions appropriate for cluster analyses arise from t-statistics for which there is not yet a complete theory, and the Gaussian approximation underestimates significance for the few degrees of freedom characteristic of PET activation studies. The only alternatives to the Gaussian random field theory are the empirical distributions derived by Roland et al. [1993], or the simulations described in the Appendix.

We present here a comparison of the procedures used by SPM and CBA in evaluating activation studies. The data consist of a set of 10 human studies for which partial results already have been presented [Missimer et al., 1996] and a set of 10 simulated studies that were derived from high resolution MR images of a human brain. The MR images were processed to yield the contrasts of PET blood flow images, furnished with activation regions, added to realistic noise images and smoothed to the resolution of reconstructed PET images. We divided the comparison into two stages, spatial standardization and statistical analysis, and investigated the effect of smoothing.

For the human studies, the CBA or SPM linear spatial standardization methods were followed by statistical analysis with the paired t-test or general linear model. The set was analysed in three ways for six filter sizes: the three procedures are denoted CBA-CBA, SPM-CBA, and SPM-SPM, the first acronym referring to the method of spatial standardization. The parameters evaluated were number and size of FRs, maximum and average statistic (t-statistic or z-score), average baseline and activation rCBF, and percentage change.

For the simulated studies, no spatial standardization (NOS) and SPM linear standardization were followed by the same statistical procedures as for the human studies. The set was analyzed in three ways, denoted NOS-CBA, SPM-CBA, and SPM-SPM, for six filter sizes. In addition to the parameters of the human studies, the efficiency of signal detection and rate of false positives could be evaluated as well as the distributions of statistics and cluster size in the absense of signal; this analysis was also performed for the SPM nonlinear standardization. Thus the simulated studies were essential to unravelling the effects of spatial standardization and statistical analysis in determining regions of significant change in activation studies.

The SPM analysis was done using the software SPM95 from the Welcome Department of Cognitive Neurology (London, UK) implemented in MATLAB (Mathworks, Sherborn, MA). The algorithms of SPM95 have not been essentially modified in succeeding versions.

## METHODS

### Human volunteers

Ten healthy, right-handed volunteers, 20–29 years of age, performed self-paced index flexions with their right hand at 2 Hz. Patients were blindfolded; the ambient noise of the scanner was the only aural stimulation. Activations were recorded by comparing the regional cerebral blood flow (rCBF) during activation with that at rest. Blood flow was measured using positron emission tomography (PET) with the tracer [150]-butanol. Arterial blood sampling permitted absolute quantification of the rCBF [Herzog et al., 1996]. The examinations were performed in accordance with the guidelines of the Declaration of Helsinki 1975 [Varga, 1984] and were approved by the Ethics Committee of Heinrich-Heine-University Düsseldorf.

All measurements were performed with a GE/SCANDITRONIX PC-4096/15WB PET tomograph. For attenuation correction, a transmission scan was obtained before the emission measurements using a rotating 68Ge pin source. All PET data sets had the following format: 15 slices, 6.50 mm slice distance, 128 × 128 image matrix with voxels of size 2 × 2 mm and a 16-bit gray scale. The images were reconstructed using filtered backprojection with a 6 mm FWHM Hanning filter providing an in-plane resolution of 9 mm FWHM and an axial resolution of 8 mm FWHM [Rota et al., 1990].

## Simulated activations

An MR image of a normal brain (256 × 256 × 128) with voxel size 1 $mm^3$ was segmented into gray and white matter compartments using simple thresholding. White matter was assigned a voxel value of 50 and gray matter a voxel value of 150. Ten rest image volumes were created by adding Gaussian-smoothed white noise to the noise-free images. For each volume, 128 planes of white noise with zero mean and a standard deviation of 300 were convolved in plane using a 2D Gaussian kernel with a standard deviation of 3 mm. Ten activation image volumes were generated by placing 3D Gaussian hot spots of maximum amplitude 200 at 14 locations in the gray matter of the original noise free image volume. The origins of two (cbr and cbl) were placed on opposite sides of the cerebellum in plane 92; four (il, ir, bgl, bgr) were placed on the insula and basal ganglia with origin in plane 74; and three (mcl, mcr, sma) were placed with origin in plane 34 corresponding to the left and right motor cortex and the supplementary motor area. To simulate individual differences of these activations, the origins were varied in each study by randomly placing them about a fixed location; the standard deviation of the variation was 3 mm. Four additional hot spots (bl, bc1, bc2, and fc) were placed with origins at fixed locations toward the center and one (br) on the right edge of the cortex in plane 54. Except for the hot spots cbl and cbr, the standard deviation of the widths was 5 mm; the standard deviation of the widths of cbl and cbr were 7 mm and 9 mm, respectively. Simulated PET images were constructed from the rest and activation volumes by convolving with a 3D Gaussian with a FWHM of 6.5 mm, corresponding to the reconstruction filter. Then, the images were resliced to dimensions: 128 × 128 × 15 corresponding to the voxel sizes: 2 × 2 × 6.5 mm. Accounting for the reduction of noise (1/30) and signal (.75) due to the final convolution yields an estimated t-statistic for the 10 resulting difference images of

$$\langle t \rangle \approx 8 \approx \frac{50 \cdot .75}{\frac{300}{30}} \cdot \left(\frac{10}{2}\right)^{1/2}$$

neglecting the deviations of location.

## Analysis

### *Spatial standardization*

Standardization of the images of human subjects using the CBA program of Bohm et al. [1986] proceeds by fitting the contour of the cerebral surface and the ventricular system of the standard atlas brain using linear and nonlinear transformations to the brain of each subject [Greitz et al., 1991], as presented on typically 17 transaxial MR images. The individual adaption parameters were subsequently used to transform the PET images of each subject into the standard brain anatomy of CBA. This process yielded 14 transaxial standard-shape images with a voxel side length of 2.55 mm and a distance between slices of 6.75 mm. The accuracy of the CBA standardization procedure has been documented elsewhere [Seitz et al., 1990]. The 20 volumes of images were smoothed in plane using a Gaussian filter with full widths at half-maximum (FWHM) of 0, 5, 10, 15, 20, 25, and 30 mm. Differences in global means were not corrected for before statistical evaluation.

The generation of all simulated image volumes from a single MRI provided the opportunity to perform statistical evaluation independently of possible distortions due to the transformations of spatial standardization. Since the locations of the activations were known and thus could be eliminated from statistical analysis, distributions of statistics and cluster size in the absense of activation could be verified. Smoothing was performed as specified above; differences in global means, due exclusively to the local activations, were not corrected for before statistical evaluation.

Before standardization of the human studies using SPM, the pairs of rest-activation images were aligned using a least-squares approach to the six parameter rigid body transformations. Stereotactic normalization, the designation for standardization in SPM, transforms each image into the standard space determined by the Talairach atlas [Talairach and Tournoux, 1988]. This transformation matches each scan in a least-squares sense to the template image, supplied by SPM,

which conforms to the standard space. In general, the matching involves a 12-parameter linear affine transformation and a nonlinear quadratic transformation in three dimensions followed by a 2-dimensional piecewise nonlinear matching in the transverse planes [Friston et al., 1995]. Only the linear transformations were applied to the human studies. The voxels of the 26 standardized planes in an image volume are 2 mm on a side and the distance between planes is 4 mm. The image volumes were smoothed uniformly in 3 dimensions using a Gaussian filter with FWHM of 0, 5, 10, 15, 20, 25, and 30 mm. Corrections for differences in global means were not made.

Spatial standardization of the simulated images differed from that of the human studies only in that no realignment was performed and all images were standardized using a single set of linear or nonlinear transformation parameters in accordance with the common origin of the images.

### *Statistical analysis*

From the spatially standardized PET images, the statistical analysis of CBA begins with the creation of mean images, in which each voxel value (V) represents the sum of the values in the same voxel of each subject divided by the number of subjects. Images of the mean rCBF in the rest and activation states as well as images of mean rCBF changes (ΔV) were calculated. Corresponding to each average image was an image showing the standard error of the mean (SEM) of each voxel value (V or ΔV). The mean rCBF changes were accepted as significant if they fulfilled the following criteria. First, a descriptive t-map was calculated as an omnibus test voxel-by-voxel according to:

$$t = \Delta V_{(Act - Rest)} / SEM_{(Act - Rest)}$$

and regions of significant change were determined by selecting voxels for which the t-statistic exceeded 2.82, corresponding to $P = .01$ for 9 degrees of freedom in one-tailed comparisons as in the human studies and simulated activations presented here. Second, the number of contiguous voxels fulfilling the threshold criterion must exceed a certain number, which depends on the total number of voxels considered (voxels outside the brain were excluded from the analysis of human studies), the effective FWHM of the image, and the voxel size. The cluster criterion bounds the probability that clusters of voxels exceeding a given search threshold occur by chance in the search volume; the level of significance used to determine the critical cluster sizes for both CBA and SPM analyses in the following analysis is $P = .05$. Introduced to assess the significance of focal activations in CBA [Knorr et al., 1993; Roland et al., 1993], the criterion was provided a theoretical basis for Gaussian statistical fields [Friston et al., 1994] and included in SPM. For t-statistical fields, no theoretical derivation has yet appeared, but the gamma distribution has been proposed [Knorr et al., 1993] to describe the cluster size distribution for two-dimensional fields. A summary of the criteria and extensive simulations performed to substantiate this proposition and deduce the critical cluster sizes are deferred to the Appendix. Analyses of the human studies employing the statistical analysis of CBA are denoted CBA-CBA and SPM-CBA, in which the first acronym refers to the method of spatial standardization. Similarly, analyses of the simulated activations are denoted NOS-CBA and SPM-CBA, NOS, indicating that no spatial standardization was performed.

The statistical analysis of SPM proceeds from a design matrix of a single study with two conditions. The condition and subject (block) effects were estimated according to the general linear model. The resulting set of voxel values for the contrast, activation-rest, constituted the statistical parametric map of the t statistic, SPM{t}. The SPM{t} was transformed to the unit normal distribution and thresholded at 2.33 corresponding to $P = .01$ as in the CBA analysis; the result was represented in a statistical parametric map, SPM{Z}. The distribution of cluster sizes for Gaussian fields in three dimensions was used to assess the significance of a cluster at the level $P = .05$. Analyses of the human studies and simulated activations employing the statistical analysis of SPM are denoted SPM-SPM. The quantification of the original blood flow images was preserved in the standardization and statistical analysis stages, i.e., the average of the global means was entered as the grand mean in SPM.

## RESULTS

### Human volunteers

At the required level of significance and with no additional smoothing (FWHM = 0 mm), CBA-CBA detected 2 FRs: the left motor cortex and right anterior cerebellum. Relaxing the critical size criterion, clusters were observed at the positions of the two FRs for all filter sizes. Two other regions, the supplementary motor area and the middle anterior cerebellum, barely failing significance and contributing to the FRs in SPM-SPM were included in the Table I and in the figures for comparison. The dependence of FR volume and detection rate on additional smoothing is shown

**TABLE I. Comparison of areas, rest rCBF, and percentage change in 10 human volunteers for three methods of analysis with additional smoothing of 5 mm FWHM and a search threshold of P = .01**

| | Volume (ml) | | | Rest rCBF (ml/100 g/ml) | | | Increase (%) | | |
|---|---|---|---|---|---|---|---|---|---|
| | cba-cba | spm-spm | spm-cba | cba-cba | spm-spm | spm-cba | cba-cba | spm-spm | spm-cba |
| region | | | | | | | | | |
| cbr | 2.2 | 2.5 | 1.6 | 49 | 49 | 47 | 28 | 30 | 29 |
| cbm | .88 | 1.6 | .88 | 54 | 50 | 49 | 28 | 26 | 25 |
| sma | .83 | 2.7 | 1.2 | 66 | 71 | 72 | 23 | 21 | 21 |
| mcl | 7.7 | 9.5 | 10.4 | 54 | 52 | 44 | 36 | 31 | 33 |

in Figure 1A for individual planes. The activation volumes rise smoothly to maxima at a filter width that depends on the relative size of the volume and then decrease. The increase in critical cluster size with additional smoothing shown by the dashed line is roughly quadratic; thus the number of detected FRs decreases to one at FWHM = 25 mm and none at FWHM = 30 mm. Figure 1B shows the dependence of the maximum t-statistics in the FRs on smoothing; they decrease smoothly from maximum values attained with no additional smoothing.

With linear SPM spatial standardization and no additional smoothing, the global mean (standard deviation) of the studies was 42 (10) ml/100 g/min in the rest state and 44 (12) ml/100 g/min in the activation state. SPM-SPM detected one FR, the left motor cortex, with no additional smoothing; three FRs, the left motor cortex, supplementary motor area, and right cerebellum with additional smoothing of 5 mm; and three FRs, one consisting of left motor cortex and supplementary motor area, one of left putamen and insula, and one of middle and right cerebellum with additional smoothing of 10 mm. With further additional smoothing, the number of FRs decreases to one for FWHMs of 15 mm and 20 mm and to none for FWHMs exceeding 20 mm. Four FRs could be identified by location of the local maximum z-score for all filters and the dependence of detection rate and FR volume on additional smoothing is shown in Figure 2A. The activation volumes measured by SPM-SPM increase by as much as a factor of four with increasing FWHM, whereas the increase for CBA-CBA was typically a factor of two. Figure 2B shows the dependence of the maximum z-score on the filter size; the z-scores decrease weakly with additional smoothing as the t-statistics do in CBA-CBA. The dashed line shows the critical intensity or peak height computed using the theory of Gaussian fields for a significance level of $P = .05$; this criterion eliminates all FRs from significance.

SPM-CBA detected the left motor cortex and supplementary motor area with no additional smoothing. For

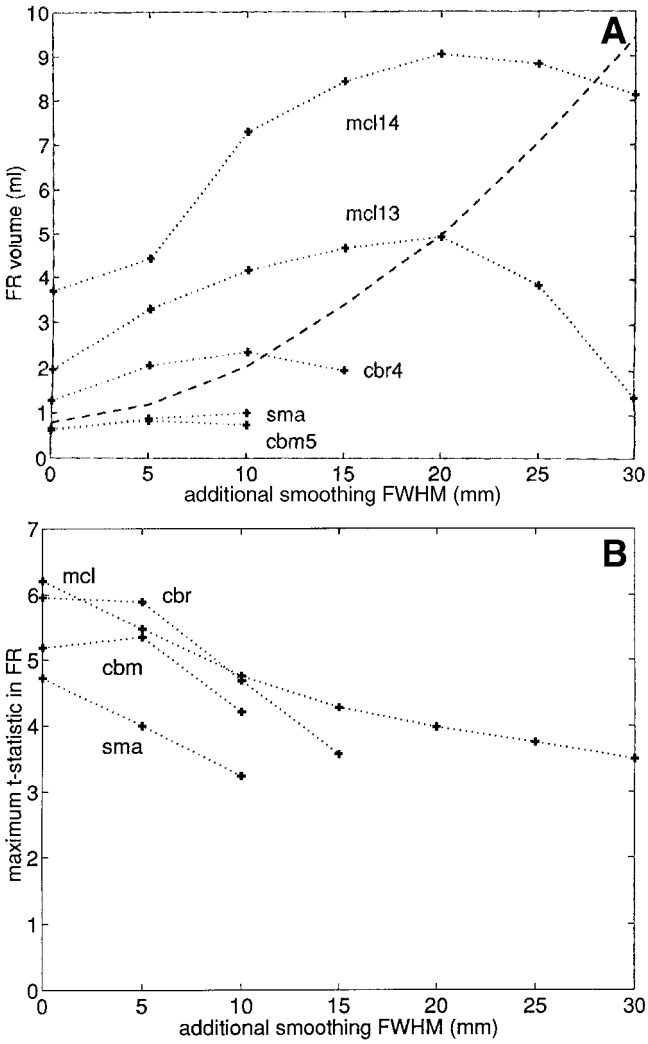



**Figure 1.**
Plane volumes (**A**) and maximum t-statistic of activated FRs (**B**) as a function of additional smoothing in human volunteers evaluated by CBA-CBA. In A, the plane is given by the suffix of the FR label. Dashed line shows volume corresponding to level of significance of .05 computed from cluster size criteria derived in the Appendix.

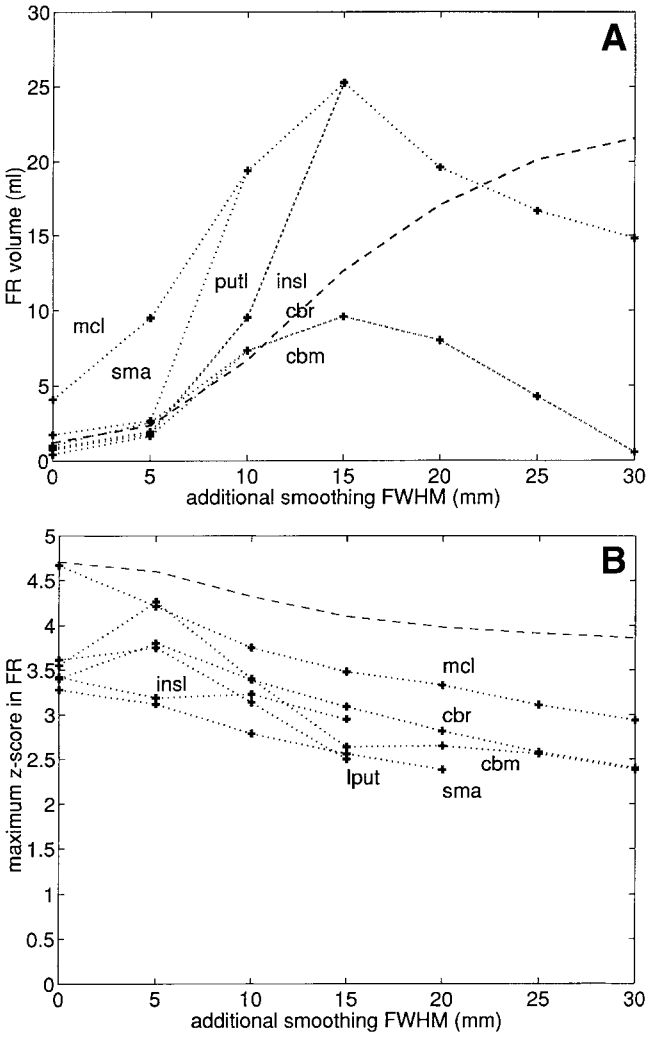



**Figure 2.**

Volumes (**A**) and maximum t-statistic of activated FRs (**B**) as a function of additional smoothing in human volunteers evaluated by SPM-SPM. In A, dashed line shows volume corresponding to level of significance of .05 computed from cluster size criteria deduced from the theory of Gaussian random fields. In B, dotted line shows z-score corresponding to level of significance of .05 computed from peak height criteria deduced from the theory of Gaussian random fields.

additional smoothing up to 20 mm, only the motor cortex was detected. The two cerebellum FRs could be identified by location for additional smoothing <20 mm FWHM. For FWHM >20 mm, no FR could be detected. As in CBA-CBA and SPM-SPM, the volumes rise smoothly to maxima at a filter width depending on the relative size of the FR and then decrease; the volumes varied by factors of two. The maximum t-statistics decrease smoothly with increasing filter size.

Table I presents a comparison of volumes, rest rCBF, and percentage change for the four FRs at FWHM = 5 mm for the three procedures: CBA-CBA, SPM-CBA, and SPM-SPM. Estimates of volume are consistent within 15% for the left motor cortex, but vary by as much as a factor of two for the smaller supplementary motor area and cerebellar regions. Estimates of the rest rCBF and of percent increase are consistent within 10% with a single exception.

### Simulated activations

Since the locations of activations in the simulated images were known, the associated clusters could be subtracted from the statistical t-maps and the distribution of statistic and cluster frequency in the remaining volumes analyzed. Since the activations were increases, negative t-statistics should occur purely by chance, providing additional distributions of statistic and cluster frequency. Thus the effect of spatial standardization on the frequency distributions of statistic and cluster could be investigated. Figure 3 shows distributions of t-statistic for the case of no spatial standardization (Fig. 3a), for linear SPM (Fig. 3b) and for nonlinear SPM (Fig. 3c) with no additional smoothing. Fewer voxels were included in the distributions derived from the standardizations because the search volumes were restricted to nontrivial voxels in 18 SPM planes containing no anomalies associated with the nonlinear standardization at the top and bottom of the brain. The distribution derived in the absense of spatial standardization reproduces the t-distribution very well. The distribution derived from the linear standardization deviates in two domains: an excess for $1 < t < 2.8$, which appears to be due to an unsubtracted residue of activation, and a deficit for $-4 < t < -2.8$. The deviations yielded by the nonlinear standardization are more extensive: the peak of the distribution is reduced and the tails extended relative to the expected t-distribution: an excess occurs for $|t| > 1$. Much, but not all, of the excess is associated with one cluster exceeding the critical size for $t > 2.8$ and four clusters for $t < -2.8$. Three of the clusters were distributed along the midline immediately above the cerebellum; the other two were located in the left parietal and right frontal lobes.

The clusters also distort the cluster frequencies summarized in Table II as evidenced most clearly in the large $\chi^2$. With increased smoothing (not shown), the distribution associated with no spatial standardization is unaltered, as expected. The distributions yielded by linear and nonlinear standardizations evidence excesses for $1 < t < 2.8$, which become increasingly

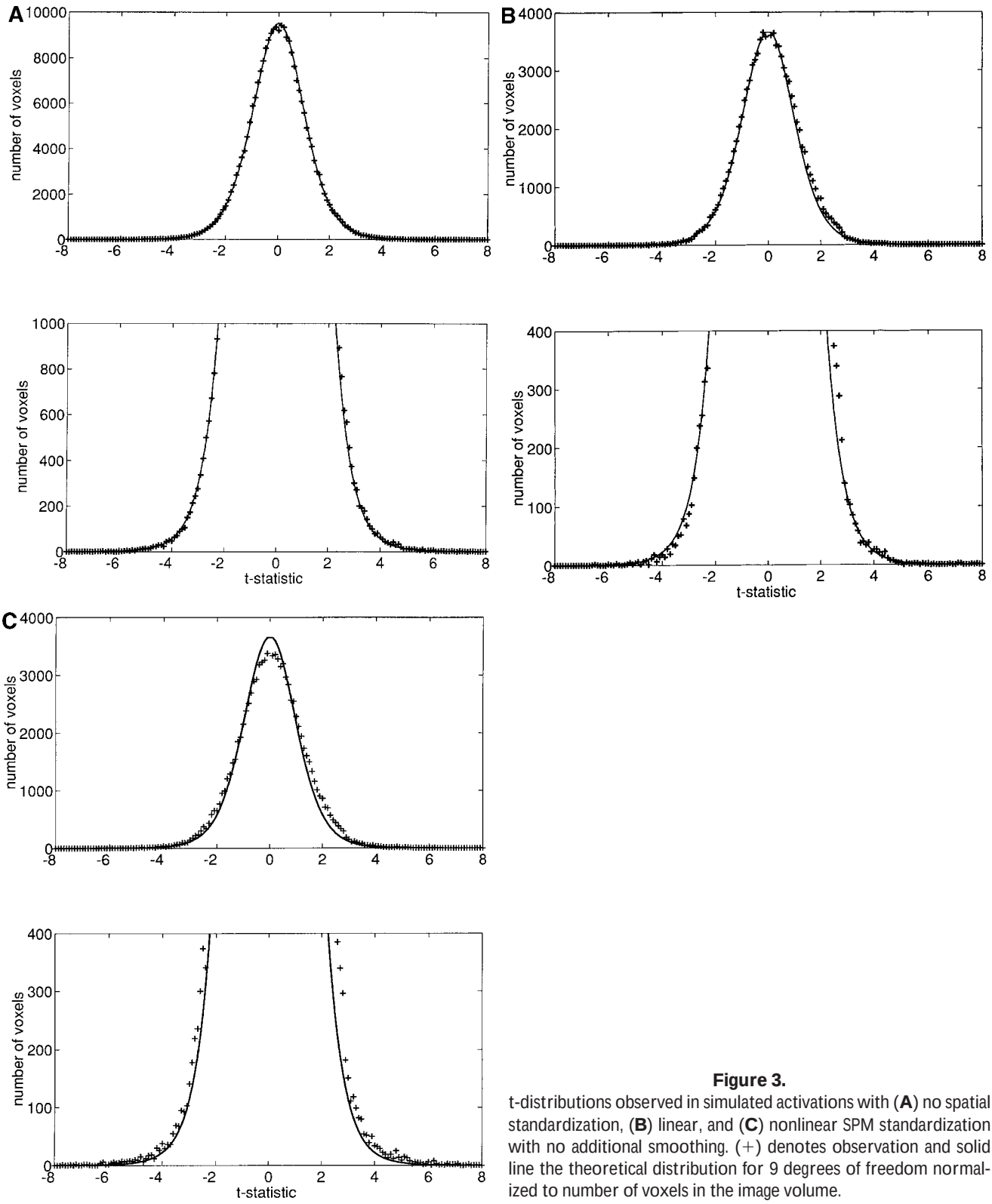



**Figure 3.**
t-distributions observed in simulated activations with (**A**) no spatial standardization, (**B**) linear, and (**C**) nonlinear SPM standardization with no additional smoothing. (+) denotes observation and solid line the theoretical distribution for 9 degrees of freedom normalized to number of voxels in the image volume.

**TABLE II. First two moments, E{N} and E{m}, and goodness of fit, $\chi^2$/dof, evaluated for 10 simulated activations using three procedures of spatial standardization: none, linear and nonlinear SPM, and three choices of additional smoothing[a]**

| FWHMs | 0 | 10 | 20 |
|---|---|---|---|
| no spatial standardization $t < -2.82$ | | | |
| E{m} | 48 | 16 | 5 |
| E{N} | 155 | 143 | 146 |
| $\chi^2$/dof[b] | .85 | .71 | .87 |
| $t > 2.82$ | | | |
| E{m} | 54 | 18 | 6 |
| E{N} | 178 | 205 | 177 |
| $\chi^2$/dof[b] | 1.69 | 1.01 | .81 |
| linear spm $t < -2.82$ | | | |
| E{m} | 40 | 18 | 15 |
| E{N} | 121 | 115 | 71 |
| $\chi^2$/dof[b] | 1.77 | 2.38 | 175 |
| $t > 2.82$ | | | |
| E{m} | 49 | 21 | 39 |
| E{N} | 158 | 128 | 114 |
| $\chi^2$/dof[b] | .89 | 4.84 | 1711 |
| nonlinear spm $t < -2.82$ | | | |
| E{m} | 51 | 18 | 23 |
| E{N} | 252 | 178 | 144 |
| $\chi^2$/dof[b] | 5.23 | 2.63 | 480 |
| $t > 2.82$ | | | |
| E{m} | 58 | 18 | 33 |
| E{N} | 237 | 161 | 107 |
| $\chi^2$/dof[b] | 5.75 | 4.03 | 1327 |
| plane simulations—Appendix | | | |
| E{m} | 50 | 17 | 6 |
| E{N} | 164 | 164 | 164 |

[a] Paired t-test computed for each voxel.
[b] Goodness of fit determined with reference to gamma distributions described in Appendix (Table IV); degrees of freedom (dof) given by number of nonvanishing cluster sizes.

pronounced with additional smoothing. The number of clusters exceeding the critical size decreases to one, for $t < -2.8$, with additional smoothing of 10 mm FWHM and to none for 20 mm, and the distributions for $|t| > 2.8$ approach the t-distribution. However, for additional smoothing of 20 mm, the cluster frequencies associated with both standardizations deviate substantially from those derived from the plane simulations. Analysis of the $\chi^2$ reveals that the dominant contributions arise from clusters of one or two voxels.

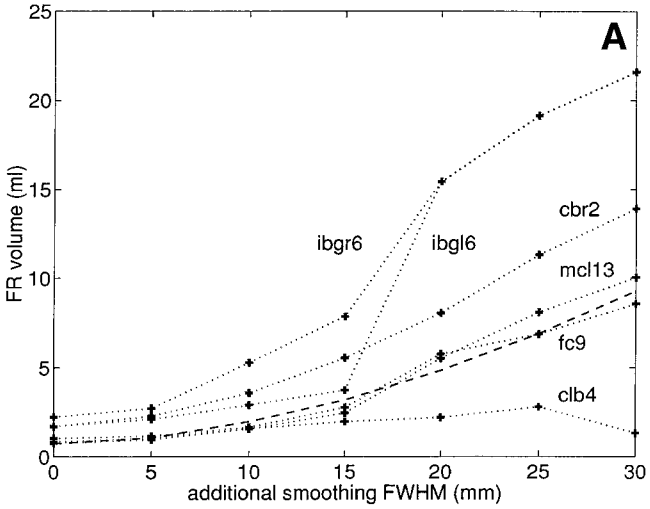
With SPM linear standardization and no additional smoothing, the global mean (standard deviation) of the simulated activations was 52.0 (1.2) ml/100g/min in the rest state and 54.3 (1.7) ml/100 g/min in the activation state. The increase in the activation state is completely due to local activations, so no correction for differences in global means would appear necessary. In fact, not correcting results in a distribution centered at positive values and yielding larger clusters, as evident in the maps of SPM{Z} or the statistics recorded in the file SPMt.mat. Since the misalignment is not found in the SPM-CBA analysis, it must be attributed to the fitting of block parameters in the general linear model. The proportional scaling option removed most of the asymmetry and is, therefore, used in the following comparisons.

No procedure discriminated between the fixed FRs denoted bc1 and bc2, separated by a mean distance of 5 mm, nor between the variable regions denoted il and bgl or ir and bgr, respectively, separated by mean distances of 12 mm. (Abbreviations were defined in Data-Simulations.) In the following, the first pair is denoted bc and the other two pairs, ibgl and ibgr. Furthermore, no procedure detected the region denoted sma. Of the remaining 10 FRs, the NOS-CBA procedure detected six: cbr, cbl, ibgl, ibgr, fc, and mcl at the required level of significance with no additional smoothing. The number of detected FRs decreased to three: cbr, a merged ibgl and ibgr, and mcl as the additional smoothing increased to 30 mm. The SPM-SPM procedure detected two FRs, the cbr and ibgr, at the same significance level with no additional smoothing. With additional smoothing of 10 mm, the cbr lost detectability and was replaced by ibgl, which merged with the ibgr at 20 mm of additional smoothing. The SPM-CBA procedure detected the same six FRs as NOS-SPM with no additional smoothing. At 20 mm, two regions, the mcl and one FR consisting of merged ibgl, ibgr, fc, were detected. Only the merged region remained with additional smoothing of 30 mm.

Relaxing the cluster criterion, all 10 FRs could be detected by all three procedures at a search threshold corresponding to $P = .01$. Figure 4A shows the dependence on additional smoothing of the volumes of the FRs for NOS-CBA. With no additional smoothing, the volumes are about a factor of two larger than those observed in the human studies. This factor is also valid for the ratio of maximum volumes. Only the smallest FRs exhibit maxima, whereas the larger volumes increase monotonically with filter size. Figure 4B shows the dependence of the maximum t-statistic on additional smoothing. With the exception of substantial variations in cbr and fc, the maxima vary weakly with

filter size. The magnitudes are consistent with the estimate given above. Figure 5A shows the dependence on additional smoothing of the volumes of the FRs for SPM-SPM. Except for the volume of ibgr, which merges first with fc and then with ibgl, the volumes of the FRs vary slowly with increasing filter size, reaching maxima between 10 mm and 25 mm of additional smoothing, and tend to be smaller than observed in NOS-CBA. Figure 5B shows the dependence of the maximum z-score on additional smoothing: the maxima vary weakly as long as the FRs are detectable, consistent with the variation observed in the human studies. The dashed line shows the critical intensity or peak height computed using the theory of Gaussian fields for a significance level of $P = .05$. This criterion classifies fc and cbr as detectable for intermediate smoothing while dismissing ibgl.

The SPM-CBA procedure exhibits volumes that are typically a factor of two larger than found in NOS-CBA, whereas the increases in volume with filter size showed a variation comparable to that of NOS-CBA. The maximum statistics varied more noticeably than for the other two procedures acquiring a maximum typically between 10 and 20 mm of additional smoothing and then decreasing weakly with the exception of igbr, which attains a prominent maximum at 20 mm. The ordering of the maxima according to FR is not consistent among the three procedures.



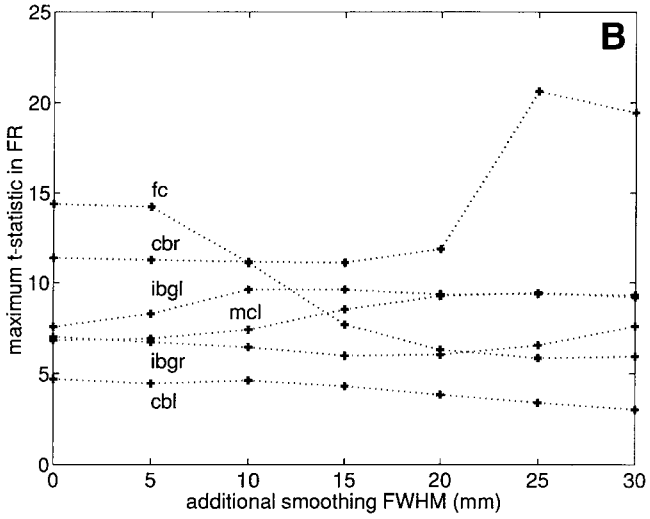
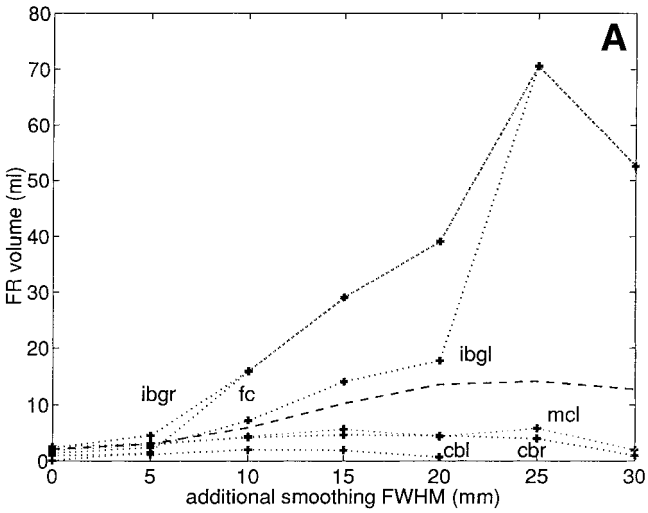




**Figure 4.**
In legend of Figure 1, plane volumes (**A**) and maximum t-statistic of activated FRs (**B**) as a function of additional smoothing in simulated activations evaluated by CBA-CBA.



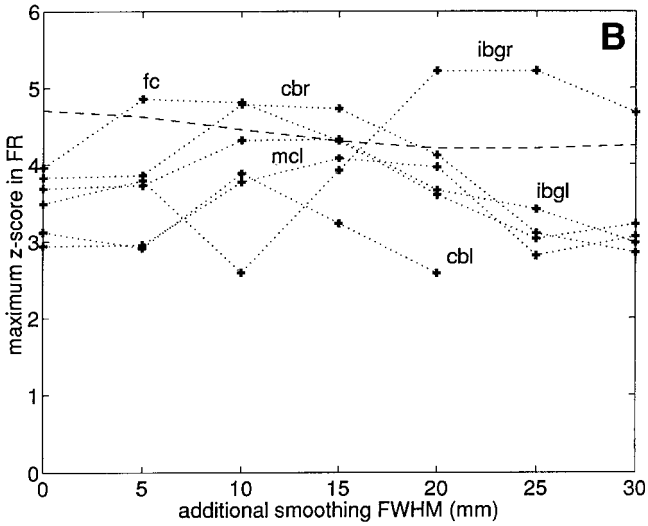




**Figure 5.**
In legend of Figure 2, volumes (**A**) and maximum z-score of activated FRs (**B**) as a function of additional smoothing in simulated activations evaluated by SPM-SPM.

**TABLE III. Comparison of areas, rest rCBF, and percentage change in 10 simulated activations for three methods of analysis with additional smoothing of 5 mm FWHM and search threshold of $P = .01$**

| | Volume (ml) | | | Rest rCBF (ml / 100 g / ml) | | | Increase (%) | | |
|---|---|---|---|---|---|---|---|---|---|
| | nos-cba | spm-spm | spm-cba | nos-cba | spm-spm | spm-cba | nos-cba | spm-spm | spm-cba |
| region | | | | | | | | | |
| cbr | 5.8 | 3.1 | 3.6 | 86 | 80 | 79 | 29 | 16 | 15 |
| cbl | 1.0 | 1.2 | 1.6 | 137 | 92 | 89 | 8.6 | 7.0 | 7.3 |
| ibgr | 4.4 | 4.5 | 8.8 | 82 | 72 | 60 | 24 | 12 | 15 |
| ibgl | 3.6 | 2.4 | 7.1 | 67 | 65 | 52 | 31 | 14 | 18 |
| fc | 2.2 | 1.6 | 2.8 | 78 | 62 | 52 | 24 | 17 | 20 |
| br | .29 | .30 | .94 | 93 | 69 | 66 | 13 | 7.1 | 7.7 |
| bc | .47 | .48 | .83 | 137 | 91 | 88 | 7.9 | 5.5 | 5.6 |
| bl | .78 | .80 | 1.1 | 90 | 68 | 67 | 16 | 9.6 | 9.4 |
| mcr | 1.1 | .85 | 1.4 | 86 | 59 | 57 | 15 | 11 | 11 |
| mcl | 2.6 | 2.8 | 3.4 | 91 | 63 | 61 | 16 | 13 | 13 |

Table III presents a comparison for the three procedures of volumes, rest rCBF and percentage change at FWHM = 5 mm for the 10 simulated FRs with a search threshold of $P = .01$. With one exception, the volumes found by SPM-CBA are larger than those found by the other two procedures; in four FRs, the discrepancy is a factor of two or more. For six of the 10 FRs, NOS-CBA and SPM-SPM yield comparable estimates; for three FRs the estimates of NOS-CBA exceed those of SPM-SPM by >30%. The estimates of rest rCBF determined by SPM-SPM are consistent with those of SPM-CBA within 20%, whereas NOS-CBA yields estimates that are consistently higher than those of SPM-SPM by as much as 50%. These differences are propagated into the estimates of percent increase of rCBF where NOS-CBA estimates are consistently larger than those of SPM-CBA and SPM-SPM by as much as a factor of two; these last two procedures are consistent within 25%.

## DISCUSSION AND CONCLUSIONS

Images of blood flow acquired from 10 human volunteers and 10 simulated activations were evaluated with methods of spatial standardization and statistical analysis employed by CBA and SPM. Analysis of the human subjects using comparable methods yielded similar results. CBA or SPM linear spatial standardization followed by smoothing in 2D or 3D, respectively, were performed: the paired t-test of CBA or the general linear model of SPM applied without correcting for differences in global mean; and the significance of the resulting statistic evaluated using the cluster size criteria derived from the plane simulations of the Appendix or Gaussian field theory. These methods detected two or three FRs; the detection efficiency of CBA was not improved by additional smoothing, whereas SPM appeared to benefit from filters between 5 mm and 10 mm FWHM. Using peak height to evaluate significance for SPM yielded poorer detection efficiency (Fig. 2b). The volumes of the FRs displayed maxima for additional smoothing between 10 mm and 20 mm FWHM (Figs. 1a, 2a), and the maximum statistics of the FRs decreased slightly with additional smoothing (Figs. 1b, 2b). The rest rCBF and the percentage change due to activation were comparable with additional smoothing of 5 mm (Table I).

With one exception, the largest volumes of the FRs were found using SPM-SPM, the smallest using CBA-CBA. The remarkable increase in FR volumes and increase in detection efficiency with additional smoothing for SPM found in Missimer et al. [1996] appear to have been associated with the correction for differences in global means yielding an off-center t-distribution. The methods employed in the analysis of the simulated activations deviated from those used in the analysis of human subjects in the choice of spatial standardization and correction for differences in global means. Since the simulated activations stemmed from a single MRI volume, the effects of spatial standardization on statistical analysis could be evaluated directly. Thus, SPM linear standardization was compared with no spatial standardization. Since the standardization method yielded distributions of statistic centered at positive values when not corrected for differences in global means, a result attributable to poor fitting of the block parameters, proportional scaling was used before applying the general linear model.

Applying the paired t-test and evaluating significance with simulated cluster frequencies (CBA) yielded a detection efficiency of 6 / 10 regardless of the method

**TABLE IV. Parameters of gamma distribution fitted to plane simulations described in Appendix, direct determinations of zeroth and first moments, E{N} and E{m}, from the simulated activations, and predictions of Worsley for smoothing used in CBA and SPM evaluations of significance**

| σ (vox) | $N_0$ | α | β | E{N} | E{m} | $E\{m\}_W$ |
|---|---|---|---|---|---|---|
| 1.08 | 166 | .4166 | .3925 | 164 | 73 | 97 |
| 1.38 | 78.5 | .5560 | .2592 | 164 | 50 | 60 |
| 1.74 | 43.7 | .6547 | .1728 | 164 | 35 | 37 |
| 1.99 | 32.3 | .6939 | .1358 | 164 | 27 | 29 |
| 2.53 | 19.3 | .7418 | .0852 | 167 | 17 | 18 |
| 2.72 | 16.5 | .7630 | .0765 | 164 | 15 | 15 |
| 3.47 | 9.98 | .8064 | .0488 | 165 | 9.4 | 9.4 |
| 4.30 | 6.71 | .7890 | .0321 | 165 | 6.4 | 6.1 |
| 4.47 | 6.32 | .7840 | .0304 | 164 | 6.1 | 5.7 |
| 5.11 | 5.76 | .6024 | .0210 | 165 | 5.3 | 4.3 |
| 5.48 | 6.07 | .4616 | .0169 | 165 | 5.3 | 3.8 |
| 6.00 | 7.80 | .2695 | .0127 | 164 | 5.5 | 3.1 |
| 6.52 | 15.6 | .1067 | .0100 | 172 | 6.2 | 2.7 |

of spatial standardization; the detection efficiency decreased with additional smoothing. The general linear model, preceeded by proportional scaling and followed by evaluation of significance with Gaussian random field theory (SPM), yielded a detection efficiency of 2/10 and no improvement with additional smoothing; using peak height to evaluate significance does not improve the detection efficiency (Fig. 5b). The CBA statistical analysis produced volumes of the FRs that increased but achieved a maximum only for cbl4 (Fig. 4a); slight maxima between 15 and 25 mm are observed with SPM (Fig. 5a). The maximum t-statistics of the FRs computed using CBA, except fc and cbr, varied little with smoothing (Fig. 4a); irregular variation with a tendency to decrease resulted from statistical analysis using SPM. The volumes of the FRs with additional smoothing of 5 mm (Table III) varied noticeably with the method of statistical evaluation: SPM-SPM yielded distinctly smaller values than SPM-CBA. With three exceptions, the volumes determined with NOS-CBA were the smallest of the three methods. The reduction of rest rCBF and percentage change resulting from spatial standardization might be attributed to the interpolation and rescaling of the procedure; the two methods of statistical analysis give comparable results.

A Gaussian model of activation explains qualitatively the differences between the human studies and simulated activations in terms of the larger t-statistics observed and larger FR volumes assumed in the simulated activations. This model implies that the dependence of FR volume and maximum t-statistic on smoothing observed in the human studies, as well as the magnitudes of the rest rCBF and percentage increases, could be reproduced in the simulated activations by decreasing the rest and activation rCBF to about one-half the values used, decreasing the noise to the difference of the two rCBFs, and assuming that the extent of the FRs including individual variation is given by a standard deviation between 3 mm and 4 mm.

The consistency between human studies and simulated activations shown in the model implies that the results are not artifacts of the particular activation employed in the human studies. In addition to verifying the results obtained for human subjects, the simulated activations permitted evaluation of statistical and cluster distributions in the absense of activation. For this evaluation, simulations described in the Appendix provided empirical cluster distributions appropriate to planar t-maps. Well represented by gamma distributions (Table IV) they provided a quantitative demonstration that the theory of Gaussian random fields underestimates significance for studies having few degrees of freedom. An important finding is that the method of spatial standardization affects the distributions. The nonlinear standardization yields excesses in the tails of the statistical distribution in the absense of additional smoothing (Fig. 3), and both methods of standardization show a residue of activation for $1 < t < 2.8$, which increases with additional smoothing. The excess in the tails can be partly, but not completely, attributed to five clusters exceeding the critical size. The residues might result from the interpolation and scaling performed in the standardization procedure. In

addition to distortions of the statistical distribution, both standardization methods perturb the cluster distributions with additional smoothing of 20 mm (Table II); an excess of clusters of few voxels is the dominant effect. Thus linear standardization appears to yield distributions consistent with the statistical assumptions underlying the evaluation of significance for additional smoothing of 10 mm or less, whereas nonlinear standardization is inadequate in this respect.

The similarity of the results yielded by similar methods of analysis for the human studies and the simulated activations substantiates the robustness of the methods for selecting functional regions. However, the analysis of simulated activations demonstrated that quantitative evaluation of significance of an FR encounters important obstacles at every stage of the analysis: spatial standardization, correction for differences in global activation, computation of the t-statistic, and evaluation of the cluster frequencies expected in absense of signal. A necessary component of any analysis is, therefore, a procedure for observing the complete distribution of t-statistic; faulty correction for global activation or poor fitting of model parameters produce massive shifts of the distribution, which are inconsistent with the number of voxels showing activation being a small fraction of the total for modest smoothing. The complete distribution of cluster sizes could be similarly useful. Finally, the creation of simulated activation studies incorporating more realistically the PET data acquisition process and individual anatomical variation could contribute substantially to refining methods of spatial standardization and statistical modelling employed in the evaluation of statistical significance.

## APPENDIX

The critical cluster size derives from the distribution, $f_n$, of clusters of n voxels occuring by chance in a statistical field. The distribution depends on the type of statistical field, its smoothness, the number of voxels in the field, and on the search threshold.

The distributions of cluster sizes for Gaussian fields in two dimensions decrease exponentially with the number of voxels in the cluster [Friston et al., 1994]:

$$f_n = N_0 \cdot \beta \cdot \exp(-\beta \cdot n)$$

so the distribution of cluster sizes is determined uniquely by the first two moments: the expected total number of clusters, E{m}, and the expected number of voxels exceeding the search threshold, E{N}:

$$E\{m\} = \sum_{n=1} f_n$$

$$E\{N\} = \sum_{n=1} n \cdot f_n$$

E{N} is simply the probability corresponding to the search threshold times the search volume. Like the distribution, E{m} depends on the type of statistical field, its smoothness, the search volume, and on the search threshold. A theoretical derivation of E{m} for Gaussian fields [Friston et al., 1991a, Worsley et al., 1992] was followed by extensions to t-, $\chi^2$-, and F-statistical fields in 1, 2, and 3 dimensions [Worsley et al., 1996].

For t-statistical fields, the gamma distribution has been proposed [Knorr et al., 1993] to describe two-dimensional fields, i.e.:

$$f_n = N_0 \cdot \frac{\beta}{\Gamma(a)} \cdot (\beta \cdot n)^{a-1} \cdot \exp(-\beta \cdot n)$$

For these fields, Worsley showed that given a sufficiently high search thresholds, $t_c$, the zeroth moment satisfies:

$$E\{m\} \rightarrow \frac{S}{\sigma_1 \cdot \sigma_2} \cdot \frac{1}{4 \cdot \pi} \cdot t_c \cdot \rho_1(t_c)$$

where S is the search volume, $\sigma_i$ characterizes the smoothness in dimension i, and $\rho_t(\tau)$ denotes the t-distribution:

$$\rho_t(\tau) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \cdot \frac{1}{\sqrt{\nu \cdot \pi}} \cdot \left(1 + \frac{\tau^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

The first moment of the distribution is given by the cumulative probability distribution:

$$E\{N\} = S \cdot P(t \geq t_c) = S \cdot \int_{t_c}^{\infty} \rho_t(\tau) \cdot d\tau$$

where $\rho_t(\tau)$ is the usual t-distribution. However, the two moments are insufficient to determine the parameters of a gamma distribution and for additional moments there is no theory. Therefore, the cluster distributions necessary to determine the critical cluster size must be simulated.

In the simulations, normally distributed random noise was generated in each voxel of a plane of

**TABLE V. Deviations of theory of Gaussian random fields from plane simulations of the Appendix[a]**

| FWHMs (mm) | 0 | 5 | 10 | 15 | 20 | 25 | 30 |
|---|---|---|---|---|---|---|---|
| $\sigma_{spm}$ | 1.38 | 1.74 | 2.53 | 3.47 | 4.47 | 5.48 | 6.52 |
| $S_{spm}$ | 51013 | 52396 | 55508 | 59432 | 63703 | 68230 | 72993 |
| $n_c(.05)$ | 28 | 41 | 76 | 123 | 186 | 265 | 357 |
| $P_g(n_c)$ | .0869 | .0991 | .1125 | .1430 | .1445 | .1318 | .1236 |
| $\sigma_{cba}$ | 1.08 | 1.37 | 1.99 | 2.72 | 3.50 | 4.30 | 5.11 |
| $S_{cba}$ | 30218 | 31038 | 32881 | 35205 | 37735 | 40417 | 43238 |
| $n_c(.05)$ | 18 | 27 | 47 | 78 | 114 | 161 | 215 |
| $P_g(n_c)$ | .0595 | .0636 | .1083 | .1189 | .1402 | .1342 | .1289 |

[a] $\sigma$ characterizes smoothness, S the search volume in voxels: spm denotes voxels of dimension $2 \times 2$ mm$^2$ and cba voxels of $2.55 \times 2.55$ mm$^2$. Critical cluster size for significance level of $P = .05$ determined by plane simulations is denoted $n_c(.05)$; significance level for that cluster size given by theory of Gaussian random fields $P_g(n_c)$.

dimension $128 \times 128$ voxels; the planes were then convolved with Gaussian kernels corresponding to those used in the analysis. Sets of 10 planes were averaged and the standard deviation calculated voxel-by-voxel to yield statistical t-maps. Applying a search threshold of $t = 2.82$, corresponding to $P = .01$, yielded a sample distribution of cluster sizes. Repeating the process 10,000 times for each convolution kernel yielded estimates of the parent distributions summarized in Tables IV and V. Table IV lists the standard deviation of the smoothing kernels expressed in voxels, the parameters of the fitted gamma distributions, the zeroth and first moments of the distributions, E[N] and E[m], and Worsley's estimate of E[m]. The parameters of the gamma distribution were found by maximizing the Poisson likelihood function; the Poisson character of the distribution about the mean frequency for each cluster size was verified empirically. The result that the parameter $\alpha$ is substantially less than one implies divergence of the distribution as n approaches zero, and a more rapid than exponential decrease for large n. In the tails of the distribution, essential to the determination of the critical cluster size, the gamma distribution provided a reliable representation of the simulated distributions. Fits to an exponential distribution yielded goodness-of-fits worse by factors of 4 to 10 times, and even worse overestimates of the number of clusters in the tails. For the small smoothing kernels, the parameters of the gamma distribution are consistent with those derived from fitting the distribution, Table II ($\sigma = 1.05$), of Roland et al. [1993] derived from PET data. Concerning the systematic deviations of E[m] from Worsley's estimates, a plausible explanation for $\sigma < 2$ is that the continuous approximation of the theory loses its validity; the deviations for $\sigma > 5$ can be attributed to the large volume approximation breaking down.

Table V summarizes the results for the critical cluster sizes used in the CBA statistical analysis. The standard deviation of the convolution kernel, $\sigma$, and the search volume, S, are given in voxels; voxel dimensions in mm are given as subscripts. The relation between the width of additional smoothing in mm, $FWHM_s$, and the total width is approximated:

$$FWHM \approx \sqrt{6.5^2 + FWHM_s^2}$$

where the reconstruction filter is assumed to have a FWHM of 6.5 mm. The standard deviation of the convolution kernel is related to the total width via:

$$\sigma = FWHM / \sqrt{8 \cdot \log(2)}$$

Division by the voxel dimensions, 2.55 mm or 2 mm for CBA or SPM spatial normalizations, respectively, yields the quantities in voxels. The critical cluster size for a significance level of .05 is given by the condition that at most one cluster of the critical number of voxels, $n_c$, or greater will occur by chance in 20 search volumes. Expressed in resolution elements (resels), which for two-dimensional fields are the number of voxels contained in $FWHM_1 \times FWHM_2$, the critical cluster sizes decrease from maxima of $<3$ for the smallest convolution kernels to a plateau of $\sim 1.5$ for kernels with $\sigma > 4$. Therefore, even for the few degrees of freedom simulated, a confidence level of $P = .05$ does not require critical cluster sizes of very many resels; activations exceeding three resels in extent are very unlikely to occur by chance for any amount of additional smoothing. $P_g(n_c)$ is the significance level of the critical cluster size as calculated by theory of Gaussian random fields. It exceeds most the confi-

dence level of $P = .05$ for the additional smoothing commonly used in PET studies, resulting in a considerable loss of significance.

## REFERENCES


Adler RJ, Hasofer AM. 1981. The Geometry of Random Fields. New York: Wiley-Liss, Inc.

Bohm C, Greitz T, Blomqvist G, Farde L, Forsgren PO, Kingsley D. 1986. Applications of a computerized adjustable brain atlas in positron emission tomography. Acta Radiol Suppl 369:449–452.

Friston KJ, Frith CD, Liddle PF, Dolan RJ, Lammertsma AA, Frackowiak RSJ. 1990. The relationship between global and local changes in PET scans. J Cereb Blood Flow Metab 10:458–466.

Friston KJ, Frith CD, Liddle PF, Frackowiak RSJ. 1991a. Comparing functional (PET) images: assessment of significant change. J Cereb Blood Flow Metab 11:690–699.

Friston KJ, Frith CD, Liddle PF, Frackowiak RSJ. 1991b. Plastic transformation of PET images. J Comput Assisted Tomography 15:634–639.

Friston KJ, Worsley KJ, Frackowiak RSJ, Mazziotta JC, Evans AC. 1994. Assessing the significance of focal activations using their spatial extent. Hum Brain Mapp 1:210–220.

Friston KJ, Ashburner J, Frith CD, Poline J-B, Heather JD, Frackowiak RSJ. 1995. Spatial registration and normalization of images. Hum Brain Mapp 2:165–189.

Greitz T, Bohm C, Holte S, Eriksson L. 1991. A computerized brain atlas: construction, anatomical content, and some applications. J Comput Assisted Tomography 15:26–38.

Herzog H, Seitz RJ, Tellmann L, Rota Kops E, Jülicher F, Schlaug G, Kleinschmidt A, Mueller-Gärtner HW. 1996. Quantification of regional cerebral blood flow with 15O-butanol and positron emission tomography in humans. J Cereb Blood Flow Metab 16:645–649.

Knorr U, Weder B, Kleinschmidt A, Wirrwar A, Huang Y, Herzog H, Seitz RJ. 1993. Identification of task-specific rCBF changes in individual subjects: validation and application for PET. J Comput Assisted Tomography 17:517–528.

Missimer J, Knorr U, Seitz RJ, Maguire RP, Schlaug G, Herzog H, Tellmann L, Leenders KL. 1996. Localization of rCBF increases in human frontal cortex and cerebellum in voluntary index flexion movements. In: Myers R, Cunningham B, Bailey D, Jones T (eds): Quantification of Brain Function using PET. San Diego: Academic Press, pp 393–397.

Roland PE, Levin B, Kawashima R, Åkerman S. 1993. Three-dimensional analysis of clustered voxels in 15O-butanol brain activation images. Hum Brain Mapp 1:3–19.

Roland PE, Graufelds CJ, Wåhlin J, Ingelman L, Andersson M, Ledberg A, Pederson J, Åkerman S, Dabringhaus A, Zilles K. 1994. Human brain atlas: for high-resolution functional and anatomical mapping. Hum Brain Mapp 1:173–184.

Rota Kops E, Herzog H, Schmid A, Holte S, Feinendegen LE. 1990. Performance characteristics of an eight-ring whole body PET scanner. J Comput Assisted Tomography 14:437–445.

Seitz RJ, Bohm C, Greitz T, Roland PE, Eriksson L, Blomqvist G, Rosenqvist, Nordell B. 1990. Accuracy and precision of the computerized brain atlas programme for localization and quantification in positron emission tomography. J Cereb Blood Flow Metab 10:443–457.

Steinmetz H, Fürst G, Freund HJ. 1989. Cerebral cortical localization: application and validation of the proportional grid system in MR Imaging. J Comput Assisted Tomography 13:10–19.

Talairach T, Tournoux P. 1988. Co-planar Stereotactic Atlas of the Human Brain. Stuttgart, Germany: Thieme.

Varga AC. 1984. Declaration of Helsinki (adopted by the 18th World Medical Assembly in Helsinki, Finland, and revised by the 29th World Medical Assembly in Tokyo, 1975.) In: The Main Issue in Bioethics, rev. ed. New York: Paulist Press.

Worsley KJ, Evans AC, Marrett S, Neelin P. 1992. A three-dimensional statistical analysis for CBF activation studies in human brain. J Cereb Blood Flow Metab 12:900–918.

Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC. 1996. A unified statistical approach for determining significant signals in images of cerebral activation. Hum Brain Mapp 4:58–73.

Wunderlich G, Knorr K, Stephan KM, Tellmann L, Azari NP, Herzog H, Seitz R. 1997. Dynamic scanning of $^{15}$O-butanol with positron emission tomography can identify regional cerebral activations. Hum Brain Mapp 5:364–378.