QUANTIFICATION OF BRAIN FUNCTION

Tracer Kinetics and Image Analysis in Brain PET

Proceedings of Brain PET '93 AKITA: Quantification of Brain Function, Akita, Japan, 29–31 May, 1993

Editors:

Kazuo Uemura
Akita Research Institute of
Brain and Blood Vessels
Akita, Japan

Niels A. Lassen
Department of Clinical Physiology
and Nuclear Medicine
Bispebjerg Hospital
Copenhagen, Denmark

Terry Jones
MRC Cyclotron Unit
Hammersmith Hospital
London, UK

Iwao Kanno
Akita Research Institute of
Brain and Blood Vessels
Akita, Japan

Co-editors:

Jun Hatazawa
Shuichi Miura
Matsutaro Murakami
Hidehiro Iida
Hideaki Fujita
Atsushi Inugami
Akita Research Institute of Brain and Blood Vessels
Akita, Japan

1993

Excerpta Medica, Amsterdam, London, New York, Tokyo

FESSLER

Quantification of Brain Function. Tracer Kinetics and Image Analysis in Brain PET
K. Uemura et al., editors.

Integrated and automated data analysis method for neuronal activation studies using ^{15}O-water PET

Satoshi Minoshima[1], Robert A. Koeppe[1], Jeffrey A. Fessler[1], Mark A. Mintun[1], Kevin L. Berger[1], Stephan F. Taylor[2] and David E. Kuhl[1]

[1]Division of Nuclear Medicine, Department of Internal Medicine, The University of Michigan, PET/Cyclotron Facility, 3480 Kresge III Bldg, Ann Arbor, MI 48109-1552; and [2]Department of Psychiatry, The University of Michigan, University Hospital 9C Box. 0120, Ann Arbor, MI 48109-0120, USA

Introduction

Neuronal activation studies using positron emission tomography (PET) have been reported to be a useful tool for examining functional localization and organization relating to specific tasks performed by the human brain. Fox and co-workers [1] initially proposed a novel method of data analysis for PET activation studies, in which cerebral blood flow (CBF) images from multiple subjects are spatially standardized and averaged to reveal consistent local CBF changes. When analyzing activation data in this manner, several factors such as individual neurophysiologic and neuroanatomic variations, as well as PET measurements, should be taken into account. To obtain objective, reproducible and accurate results, we have implemented and validated an integrated, automated method for analysis of ^{15}O-water PET activation studies. We expended particular effort on the development of automated image registration and anatomical standardization methods in three dimensions.

Methods and validations

Integrated data analysis: general procedure

The method assumes that repeated CBF measurements are acquired under specific stimulation and baseline conditions in multiple subjects. After accounting for an individual subject's head movement across serial scans by a three-dimensional image coregistration technique, stimulation and baseline image sets are averaged separately, creating a pair of average stimulation and baseline image sets (intrasubject coregistration). Each subject's image set is then transformed to stereotactic coordinates according to the intercommissural (AC-PC) line which is detected automatically in the PET image set. Pixel values in each image set are normalized to the mean global count. After linear correction of the brain size followed by 3D nonlinear regional anatomical standardization, the baseline image set is subtracted from the stimulation

image set, forming a subtraction image set for each subject. Subtraction image sets are smoothed with a 3D Gaussian filter and averaged over multiple subjects (inter-subject averaging). Statistical significance of activation foci is assessed by creating a t-statistics image set from the averaged subtraction image set using a pooled variance across the brain. Image smoothness, as defined by the total number of resolution elements, is estimated on the averaged subtraction image set [2]. A statistical threshold is derived from a 3D formula described by Worsley and co-workers [3] as well as a 3D extension of the stochastic formulation proposed by Friston and co-workers [2]. Peaks in the t-statistics image set are searched throughout the brain and reported with normalized CBF changes and stereotactic coordinates. Areas of significant change are also superimposed onto a standard magnetic resonance image for visual interpretation. We discuss details of each of these steps in the following sections.

Intrasubject coregistration

In a PET activation study, we typically obtain six to eight scans in each subject under multiple task conditions. The necessity for coregistering two or more PET scans arises from head movement during the study. Coregistration also is beneficial in many other occasions such as comparison of two scans obtained at different time points or comparison of two different types of functional images from the same subject. A common approach for coregistration is to maximize similarity between two image sets by rigid body transformation using a multidimensional iterative routine. To generalize coregistration techniques for many types of PET images, we have tested several criteria as indices for image similarity. Because two image sets are usually similar but not identical, criteria should be robust against differences between the image sets such as due to statistical noise, activation foci in one image set, or two different tracer distributions from the same individual. Venot and co-workers demonstrated the accuracy of image registration with different criteria in two dimensions and concluded that conventional criteria such as sum of squared differences and correlation coefficient were sensitive to image dissimilarity [4]. They proposed a stochastic sign change criterion, which counts the number of zero crossings in a subtraction image set. Mintun and Lee applied this criterion to PET image sets in three dimensions, showing a decreased standard deviation in a subtraction image set by using this coregistration technique [5]. Woods and co-workers alternatively have proposed a normalized standard deviation as a coregistration index [6]. We previously implemented the stochastic sign change criterion in three dimensions and found that the criterion was sensitive to small global changes of tracer distributions on two image sets [10]. For the purpose of generalization, we have introduced a new criterion which expresses image similarity as the number of pixels whose difference between image sets is less than a certain threshold.

The accuracy of the coregistration was assessed both visually, creating a smoothed subtraction image set between coregistered image sets, and through the use of phantom data. Stimulation and baseline image sets from eighteen subjects (short-term

memory) as well as paired dissimilar CBF image sets from eighteen subjects (carotid artery occlusion study) were examined. The coregistration technique clearly corrected head movement across serial scans in each subject. Without the coregistration, rim-like artifacts in subtraction images were observed in the eighteen simulation baseline pairs. These artifacts disappeared in all cases after the coregistration procedure. Average absolute values of head movements corrected by the algorithm in x (right-left), y (antero-posterior), z (ventro-dorsal) translations, and xy, yz, xz rotations were 0.4 ± 0.4, 0.5 ± 0.4, 0.8 ± 0.6 pixels, and 1.0 ± 0.8, 1.7 ± 1.8, 0.8 ± 0.6 degrees, respectively. The algorithm can correct for head movements as small as a few tenths of a pixel or degree. Furthermore, the algorithm worked well on the eighteen dissimilar paired image sets, which confirmed that the accuracy of the algorithm was insensitive to CBF alterations within the same subject.

Intersubject image registration

Prior to intersubject averaging, each PET brain image set needs to be realigned to a standard orientation, and anatomical variations of individual brains should be corrected. Therefore, intersubject image registration consists of two steps; stereotactic orientation and anatomical standardization. Anatomical standardization is subdivided further into brain size correction and regional anatomical standardization (warping).

Talairach and co-workers proposed a line passing through the anterior and posterior commissures (AC-PC) as a reference for stereotaxy [7], and this bicommissural stereotactic approach has been used for various radiological examinations. Fox and co-workers first applied the stereotactic approach to PET activation studies [8], and Friston and co-workers further developed a method for the AC-PC line estimation on PET image sets themselves [9]. To achieve accurate and objective estimation, we have proposed an automated method for the AC-PC line detection on PET image sets [11]. This method detects four landmarks in the brain: the frontal pole, the most ventral point of the anterior corpus callosum, the subthalamic point, and the posterior pole. These four landmarks were shown to estimate the AC-PC line accurately on both magnetic resonance images as well as PET images. In CBF activation data analysis, we first average all the scans obtained in a subject by using the intrasubject coregistration technique, then we apply the AC-PC line detection to the averaged coregistered image. The averaging procedure reduces noise in a single scan and, consequently, increases accuracy of the AC-PC line detection.

Following stereotactic alignment of a PET image set based on the AC-PC line, the gross brain size is scaled linearly to the Talairach atlas brain by an automated measurement of brain width, antero-posterior length, and brain height. We measure brain width and antero-posterior length directly from the PET image sets by using an edge detection technique, while the estimate of brain height is obtained by fitting the surface points from a midsagittal slice to those from an atlas brain using least-squares optimization. A proportional grid system proposed by Talairach and co-workers [12,13] provides the theoretical basis for the linear scaling.

In bicommissural stereotactic space, there obviously remain regional differences

in the shapes of individual brains. Regional anatomical standardization minimizes such differences. We have developed an automated standardization method by assuming multiple center points and corresponding surface landmarks in the stereotactic space based on the directions of major neuronal fibers in the brain. For each center point and its corresponding set of surface landmarks, the algorithm measures profile curves between the center and each landmark on an individual PET image set. The profile curves from an individual PET image set are then matched to a similar set of curves obtained from a standard stereotactic PET image set by linear scaling, thereby determining the locations of the individual's surface landmarks. These landmarks are then matched to standard stereotactic locations using a 3D thin-plate spline algorithm proposed by Bookstein [14], thus nonlinearly warping each individual brain to a standard stereotactic PET brain. The algorithm uses approximately 50 center points and 400 surface landmarks in the brain.

Effects of the regional anatomical standardization were evaluated in two ways. First, we visually inspected standard deviation (SD) images created on pixel-by-pixel basis from the CBF images of 18 subjects with and without the nonlinear standardization. Higher SD values were expected where mismatches of the regional gray matter exist. Second, vibratory and visual activation studies were conducted in five subjects, and CBF changes were measured in terms of t-statistics in the primary sensory cortex and visual cortex, respectively, with and without the standardization. The SD image created without regional anatomical standardization showed high SD values around the superior frontal and parietal cortices, the occipital cortex, the inferior temporal cortex, the cerebellum, and the inner surface of the caudate and thalamus. Those high SD values were reduced markedly by the standardization procedure, and regional SD values became more uniform throughout the brain. Peak t-values in the primary sensory and visual cortices were increased with the standardization. The increase in t-value obtained by using linear scaling plus nonlinear warping compared to using linear scaling only demonstrates nearly as great an improvement as that obtained by using linear scaling only compared to no correction (stereotactic orientation only). This suggests that regional nonlinear standardization is as important a step in analyzing PET activation studies as is correction by linear scaling.

Statistical assessment

To ensure reproducibility of an activation study, it is essential to assess the statistical significance of the activation foci that are identified. A variety of statistical inference methods have been reported; here we focus on the methods of Friston et al. [2] and Worsley et al. [3]. These two methods are based on the assumption that, under the null hypothesis of no difference between conditions, a normalized difference image can be modeled as a homogeneous Gaussian random field (GRF) with unit variance. As a simple omnibus test for significance, one could perform a hypothesis test by comparing the maximum value T_{max} of the field within a region of interest to a threshold t. To control the Type I error of such a test, one needs tail probabilities for

T_{max} so that the thresh
i.e., we need the prob

No exact formula
fields. For 2D homog
approximations to th

$$P(T_{max}>t) \approx P_F(t)$$
$$P(T_{max}>t) \approx P_W(t)$$

where p(t) is the tai
number of effective r
the 3D tail probabili
For simplicity, we f
Worsley presented e
results using N = 10
combinations of sca
probabilities (e.g., 0.
reliable validation. In
is large relative to t
approximations as a

We generated N =
standard normal var
FWHM = 2, 3, 4, 5 a
For each realization,
× 8,..., 128 × 128 w
widths 4, 8, 16, 32
cortical ribbon on th
FWHM and region
calculated as

$$P_M(t) = (\text{number o}$$

or a tail probability
about 15% or 7% re

Both P_F and P_W
pixels), but this is
resolutions. Therefo
× 4, 8 × 8 and 16 >
hence the tests may
× 128 square region
although only sligh
predicted by Worsle

T_{max} so that the threshold can be chosen to guarantee the level of significance desired, i.e., we need the probability $P(T_{max} > t)$ under H_0.

No exact formula for this tail probability is known even for homogeneous random fields. For 2D homogeneous GRFs, Friston and Worsley have presented the following approximations to the tail probability, respectively:

$$P(T_{max}>t) \approx P_F(t) = R(8\log 2)/[32\pi e^{t^2}p(t)]$$
$$P(T_{max}>t) \approx P_W(t) = R(4\log 2)(2\pi)^{-3/2}te^{-t^2/2} \quad (1)$$

where p(t) is the tail probability for the standard normal distribution and R is the number of effective resolution elements. Worsley also presented an approximation for the 3D tail probability, and we have generalized Friston's derivation to 3D as well. For simplicity, we focused on the 2D formulae in this study. Neither Friston nor Worsley presented extensive evaluations of their approximations; Friston presented results using N = 100 simulated images, and Worsley reported results from N = 70 combinations of scans from a pain study. Since we are interested in small tail probabilities (e.g., 0.01 or 0.05), a much larger N is required to ensure statistically reliable validation. In addition, both approximations assume that the number of voxels is large relative to the FWHM. One of the goals of our work was to examine the approximations as a function of region size.

We generated N = 4000 discrete GRFs on a 128 × 128 lattice using pseudorandom standard normal variates. Each image was smoothed using Gaussian kernels with FWHM = 2, 3, 4, 5 and 6 pixels, and normalized (analytically) to have unit variance. For each realization, the maxima T_{max} within square regions of dimensions 4 × 4, 8 × 8,..., 128 × 128 were tabulated. The maximum values within rim-like regions of widths 4, 8, 16, 32 and 48 pixels simulating different thicknesses of the cerebral cortical ribbon on the smoothed images were also recorded. For each combination of FWHM and region size, the Monte Carlo estimate of the tail probability was then calculated as

$$P_M(t) = (\text{number of } T_{max} \text{ values}) / 4000 \quad (2)$$

or a tail probability of 0.01 or 0.05, the coefficient of variation of this estimate is about 15% or 7% respectively.

Both P_F and P_W were poor approximations to P_M for the smaller FWHM (2, 3, 4 pixels), but this is probably due to the discrete lattice being too coarse for those resolutions. Therefore, in the remainder we focus on the FWHM = 6 case. For the 4 × 4, 8 × 8 and 16 × 16 square regions, both P_F and P_W were smaller than P_M (and hence the tests may not be conservative enough). For the 32 × 32, 64 × 64 and 128 × 128 square regions, P_W was very close to P_M, whereas P_F consistently exceeded P_M, although only slightly. The ratio P_F/P_W was approximately 0.8, close to the p/4 predicted by Worsley et al. Thus the tail probability approximation P_F is slightly too

small under the GRF assumption. For the "frame" shaped regions, again P_F was consistently slightly smaller than PM, whereas for the 16, 32 and 48 pixel width frames, P_W was very close to P_M. However, for the four and eight pixel width frames, even P_W was consistently about 10% too small, despite the fact that the frames do contain a large number resolution elements. This validation suggests that for accurate control of Type I error, the region of interest shape must be considered in addition to the total number of resolution elements. We currently calculate P_F and P_W in three dimensions in the integrated method. These thresholds need to be validated further using actual PET data.

Applications

We have implemented algorithms on a common workstation (SUN SPARC station, SUN Microsystems, Mountain View, CA). Since the method is fully automated, an investigator need only write a text file to specify a study design. Each step creates quality control images for visual inspection by an investigator. Although programs are not optimized, computational time is approximately 15 min for midsagittal slice and AC-PC line detection, 15 min for coregistration, and 90 min for nonlinear warping on a SPARC station 2. These computational times are important because a typical PET activation study involves large number of scans and subjects. We have applied the method to activation studies of vibratory, visual and thermal pain stimulations as well as cognitive tasks such as short-term memory, word processing, and stimulus response compatibility. PET images were acquired on a Siemens 931/08-12 scanner (CTI, Knoxville, TN) which simultaneously collects 15 slices with slice separation of 6.75 mm. Following a bolus injection of 50 to 66 mCi of ^{15}O-water, each image set was acquired over 60 s, starting at 5 s after the injected radioactivity bolus arrived in the brain. Images were reconstructed by Parzen filter with a cutoff frequency of 0.45 cycles per projection. Raw count images were used for the analysis.

In the vibratory activation study, the most significant CBF increase was observed in the contralateral postcentral gyrus. A peak t-value of 8.0 far exceeded the $p = 0.05$ significance threshold (multiple comparison adjusted) of 4.4. These t-values are dependent on the smoothing filter applied to subtraction image sets. A smoothing filter can compensate for anatomical variations to some extent but sacrifice image resolution. Although the filter function is not optimized yet, estimated smoothness is approximately 15 mm full width at half maximum in our usual setting. In the visual activation study, the most significant CBF increase was detected in the primary visual cortex. We also analyzed same-state subtraction paradigms (stimulation—stimulation and baseline—baseline) using the vibratory activation data and found that no false positive areas arose at levels of either P_F or P_W, although this should be validated further using larger number of subjects and studies. Results from other activation paradigms will be reported separately.

Summary and Cor

The integrated and a
and reproducible d
statistical assessmen
with other statistica
designs. These adv
activation studies. In
applicable to other t
should improve the r
localization of funct

References

1. Fox PT, Mintun MA
2. Friston KJ, Frith CD
3. Worsley KJ, Evans
4. Venot A, Lebruche
5. Mintun MA, Lee K
6. Woods RP, Cherry
7. Talairach J, De Ajur
8. Pox PT, Perlmutter
9. Friston KJ, Passingh
10. Minoshima S, Berge
11. Minoshima S, Koep
12. Talairach J, Szikla
    Masson & Cie, 196
13. Talairach J, Tourno
    Thieme, 1988.
14. Bookstein FL. IEEE

egions, again P_F was
2 and 48 pixel width
ht pixel width frames,
ıct that the frames do
gests that for accurate
onsidered in addition
ate P_F and P_W in three
o be validated further

SUN SPARC station,
is fully automated, an
ign. Each step creates
r. Although programs
n for midsagittal slice
90 min for nonlinear
e important because a
nd subjects. We have
ıal and thermal pain
ory, word processing,
quired on a Siemens
collects 15 slices with
50 to 66 mCi of ^{15}O-
5 s after the injected
ıcted by Parzen filter
unt images were used

ncrease was observed
exceeded the $p = 0.05$
ł. These t-values are
ge sets. A smoothing
ıt but sacrifice image
timated smoothness is
l setting. In the visual
l in the primary visual
mulation–stimulation
d found that no false
s should be validated
from other activation

Summary and Conclusions

The integrated and automated method described in this manuscript enables objective and reproducible data analysis for ^{15}O-water PET activation studies, although statistical assessment still needs to be further validated. The method can be combined with other statistical tests (i.e., correlation) and is applicable to a variety of study designs. These advantages should facilitate the application of PET to neuronal activation studies. Intrasubject and intersubject image registration techniques can be applicable to other types of PET images as well as SPECT images. These techniques should improve the reliability of comparisons across multiple subjects and anatomical localization of functional brain images.

References

1. Fox PT, Mintun MA, Reiman EM et al. J Cereb Blood Flow Metab 1988;8:642–653.
2. Friston KJ, Frith CD, Liddle PF et al. J Cereb Blood Flow Metab 1991;11:690–699.
3. Worsley KJ, Evans AC, Marrett S et al. J Cereb Blood Flow Metab 1992;12:900–918.
4. Venot A, Lebruches JF, Roucayrol JC. Comput Vis Graph Image Process 1984;28:176–184.
5. Mintun MA, Lee KS. J Nucl Med 1990;31:816.
6. Woods RP, Cherry SR, Mazziotta JC. J Comput Assist Tomogr 1992; 16:620–633.
7. Talairach J, De Ajuriaguerra J, David M. Press Med 1952;28:605–609.
8. Pox PT, Perlmutter JS, Raichle ME. J Comput Assist Tomogr 1985;9:141–153.
9. Friston KJ, Passingham RE, Nutt JG et al. J Cereb Blood Flow Metab 1989;9:690–695.
10. Minoshima S, Berger KL, Lee KS et al. J Nucl Med 1992;33:1579–1585.
11. Minoshima S, Koeppe RA, Mintun MA et al. J Nucl Med 1993;34:322–329.
12. Talairach J, Szikla G, Tournoux P et al. Atlas d'Anatomie Stéréotaxique du Télencéphale. Paris: Masson & Cie, 1967.
13. Talairach J, Tournoux P. Co-planar stereotaxic atlas of the human brain. Stuttgart, New York: Thieme, 1988.
14. Bookstein FL. IEEE Transactions on Pattern Analysis and Machine Intelligence 1989;11:567–585.

Discussion

Satoshi Minoshima and Robert A. Koeppe

Dr Evans: Could you give us the rational for the choice of control points that you used and do you have any information on how things change if you have more or less control points?

Dr Koeppe: Well we have generally added more control points to try and improve certain areas. We started with about 25 in white matter regions of the brain which are assumed not to vary very much. Those points themselves are not used in the 3D warping. It does not matter if they are off a little bit, as you have determined landmark points that go radially out from them and go in all dimensions, and you do not use that central point. After the first set of tests we know that certain areas of the brain improved more, some others were still not addressed. Dr. Minoshima did most of this work, and he continued to add points and adjust points to try and improve the variance in regions.

Dr Eberl: I have got a question regarding your cost function. It almost appears to me that you are accepting an unacceptable level of error by selecting a 10% threshold. And secondly, I would think that perhaps that threshold might be worth adjusting depending on the noise in the image. Have you adjusted for the noise in the image?

Dr Koeppe: We have tested quite a few different thresholds and it really does not seem to make a lot of difference. We have tried 5%, 10%. Theoretically yes, if you had noisy images you would probably want to use a larger threshold, but as far as the accuracy of the results, it seems to be fairly independent of the exact level. We have not tried anything like a 50% or a 1% threshold, however.

Dr Woods: One of the concerns that I have about these nonlinear methods is that they may give an image which looks good, but carries no biological validity. Your use of the t-statistic... it is impressive to see the increase, but I wonder whether there are factors which play into the t-statistic, particularly with regards to what the absolute values are in terms of whether there is grey matter or white matter at a particular location. This may make the t-statistic a suboptimal measure in terms of validating these techniques, and I wonder whether you had thought about any other types of measures that you might apply?

Dr Koeppe: For validation? We really have not at this time, that is as much validation as we have done. I think that is really a very important point, though. I

think a good gold standard of validating what improvement is, whether it is real or not, is important. We have looked at magnitudes, which I do not think is any better. For example, the magnitude of a mean stimulation goes up 10%.

Dr Kanno: I am wondering whether you have tried your warping technique between the different populations? For example, Japanese brains are different from European brains, they are very round and tall. If your subject group includes different populations, oriental and Americans, how about the result?

Dr Koeppe: We have looked at some oriental volunteers, but certainly not very many, so I do not have a good answer to that.