

# *Wellcome Department of Cognitive Neurology*

Tel: +44 (0)171 833-7472  
Direct: +44 (0)171 833-7490  
Fax: +44 (0)171 813-1420  
email: andrew@fil.ion.ucl.ac.uk

Functional Imaging Laboratory  
12 Queen Square,  
London. WC1N 3BG  
2<sup>nd</sup> February 1998

Dr James McCulloch,  
Editor, Journal of Cerebral Blood Flow & Metabolism.  
Wellcome Surgical Institute & Hugh Fraser Neuroscience Laboratories,  
University of Glasgow,  
Garscube Estate,  
Bearsden Road,  
Glasgow. G61 1QH  
Scotland, UK

**Re: Holmes & Watson, on ‘Sherlock’**

Dear Sir

We welcome the interest and enthusiasm of Halber *et al.* (*JCBFM* **17**:1033-1039) in the non-parametric randomisation test that we described (Holmes *et al.*, *JCBFM* **16**:7-22), not least because they called their program ‘Sherlock’. However, as the great detective would surely have pointed out, there are a number of flaws in the paper that have led the authors to an incorrect primary conclusion, that our test “is a less sensitive procedure [than] a standard parametric test”.

First, “we must look for consistency” [AC Doyle, *The Problem of Thor Bridge*]. The principal flaw is that the ‘standard’ parametric test was used by Halber *et al.* in a non-standard way. The parametric statistic images were thresholded at  $Z = 1.64$ , or a  $p \leq 0.05$ , uncorrected for multiple comparisons. Thus the false positive or Type I error rate is 0.05 per voxel. So we expect that 5% of voxels to appear above the threshold by chance alone - exploiting inappropriately Holmes’ adage that “it is a question of cubic capacity. A man with so large a brain must have something in it” [AC Doyle, *The Adventure of the Blue Carbuncle*]. In contrast, the non-parametric randomisation test is constructed to control strongly for Type I error at the image level, giving a Type I error on average once for every twenty applications of the test. For an equitable comparison of test sensitivities we must “balance probabilities” [AC Doyle, *The Hound of the Baskervilles*]. The only valid comparison is with a parametric test that utilises the same statistical model and also controls Type I error strongly for the family of voxel tests, which invariably would lead to a much higher threshold than 1.64. Our experience is that with such a comparison the two approaches have remarkably similar sensitivities, and by using variance smoothing the non-parametric test is more powerful at low degrees of freedom.

There is a similar lack of consistency in the visual comparison that the reader is asked to make between Figures 2 and 3 in the paper of Halber *et al.* The maximum intensity projection method for the whole brain used in Figure 3 almost invariably makes the extent of activations appear larger than when they are shown in a set of orthogonal

views through a single focus, as used in Figure 2. “It is, of course, a trifle but there is nothing so important as trifles” [AC Doyle, *The Man with the Twisted Lip*].

In conclusion, we are concerned that Halber *et al.* have presented our method poorly through inappropriate comparisons, but are nonetheless pleased to see the interest in the method, for we “cannot live without brainwork. What else is there to live for?” [AC Doyle, *The Sign of Four*].

Yours faithfully,

*Andrew P Holmes, John DG Watson, Thomas E Nichols*