ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 25 (2005) 653–660

Valid conjunction inference with the minimum statistic

Thomas Nichols,[a,*] Matthew Brett,[b] Jesper Andersson,[c] Tor Wager,[d] and Jean-Baptiste Poline[e]

[a]*Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA*
[b]*Department of Psychology, University of California, Berkeley, CA 94720, USA*
[c]*MR Centre, Karolinska Institute, Stockholm, Sweden*
[d]*Department of Psychology, Columbia University, New York, MI 10027, USA*
[e]*SHFJ/CEA/INSERM, Orsay, France*

Received 11 August 2004; revised 30 November 2004; accepted 1 December 2004
Available online 3 March 2005

In logic a conjunction is defined as an AND between truth statements. In neuroimaging, investigators may look for brain areas activated by task A AND by task B, or a conjunction of tasks (Price, C.J., Friston, K.J., 1997. Cognitive conjunction: a new approach to brain activation experiments. NeuroImage 5, 261–270). Friston et al. (Friston, K., Holmes, A., Price, C., Büchel, C., Worsley, K., 1999. Multisubject fMRI studies and conjunction analyses. NeuroImage 10, 85–396) introduced a minimum statistic test for conjunction. We refer to this method as the minimum statistic compared to the global null (MS/GN). The MS/GN is implemented in SPM2 and SPM99 software, and has been widely used as a test of conjunction. However, we assert that it does not have the correct null hypothesis for a test of logical AND, and further, this has led to confusion in the neuroimaging community. In this paper, we define a conjunction and explain the problem with the MS/GN test as a conjunction method. We present a survey of recent practice in neuroimaging which reveals that the MS/GN test is very often misinterpreted as evidence of a logical AND. We show that a correct test for a logical AND requires that all the comparisons in the conjunction are individually significant. This result holds even if the comparisons are not independent. We suggest that the revised test proposed here is the appropriate means for conjunction inference in neuroimaging.

Keyword: Conjunctions

Introduction

Many neuroimaging studies look for brain regions that respond to all of a set of different conditions. For example, researchers may be interested in whether a brain region responds generally to tasks requiring attentional control, or whether the area is only activated in specific attentional tasks. To address this issue, they may test participants using three attention-demanding tasks and ask, "Which brain regions are active in all three tasks?" This is referred to as a *conjunction*, and a positive conjunction test implies that the region is commonly activated across the tasks. A similar logic has been applied to inferences across individual subjects. Researchers are interested in whether all individual subjects show activation of a particular region.

The most commonly used test for conjunction is the minimum statistic method proposed by Friston et al. (1999a). For reasons that will become clear, we refer to the test described in that paper as the Minimum Statistic compared to the Global Null (MS/GN). Below we will argue that the MS/GN is not a valid test for conjunction in the sense that it is usually understood. Based on our own experience and a formal analysis of recent practice in neuroimaging, we find that this has caused considerable confusion. Many authors have used the MS/GN as evidence of a conjunction of effects when the nature of the test does not allow this conclusion. In this paper we set out the standard definition of a conjunction from logic, and derive a simple and valid alternative method based on the minimum statistic. We refer to our method as the Minimum Statistic compared to the Conjunction Null (MS/CN). Further, we show that the MS/CN method is valid under dependence between the tests. Finally, we document the confusion surrounding the interpretation of the SPM MS/GN test with an analysis of abstracts from the 9th International Conference on Functional Mapping of the Human Brain, June 18–22, 2003, New York.

Conjunction is simply defined in logic. If we have two truth statements A and B, then the conjunction of A and B is true if and only if both A AND B are true (Mendelson, 1987). In neuroimaging terms, the statements A and B are statements about the presence of an effect for a particular comparison. For example, say we have a binary image identifying the areas where an effect of task A is truly present; this image contains a 1 in voxels where there is a real effect for task A and zeros elsewhere. We have a similar binary image for task B. Let us call these images M_A and M_B. The *conjunction* map of M_A AND M_B will contain 1 for voxels where there is activation for task A and activation for task B, with zeros elsewhere. That is, if either M_A or M_B contains a zero (false), then the conjunction is false. (See Fig. 1).

* Corresponding author. Fax: +1 734 763 2215.
E-mail address: nichols@umich.edu (T. Nichols).
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2004.12.005

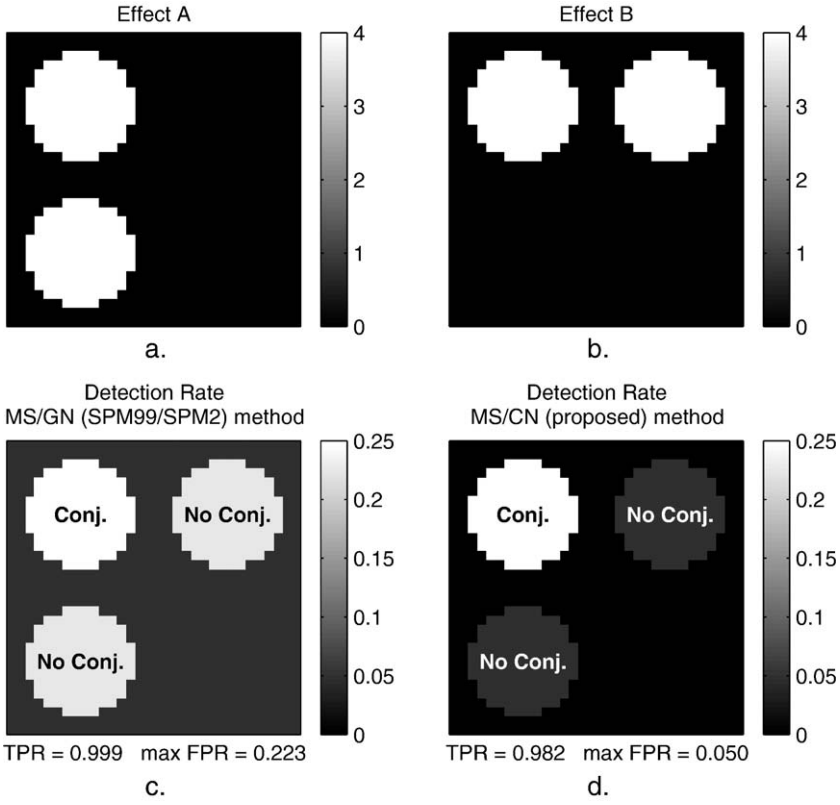

Fig. 1. Illustration of conjunctions and conjunction inference methods. Panels a and b are effects A and B to be conjoined. Panels c and d show the results of the SPM99/2 (MS/GN) conjunction and the method we propose (MS/CN). Labels in panels c and d indicate where a conjunction does and does not exist: in panel c, the false positive rate in the 'No Conj.' regions in 0.223, in excess of the nominal rate; in panel d, the false positive rate is exactly 0.05 in the 'No Conj.' regions, yet the true positive rate (TPR) is almost high as in MS/GN method (in panel c). See univariate results below for more detail.

To construct a statistical test we must specify a null hypothesis. The conjunction null hypothesis is the state of no conjunction of effects. If the conjunction hypothesis is M_A AND M_B, then the conjunction null hypothesis is: (not M_A) OR (not M_B) (c.f. Eq. (1)). The null hypothesis for a particular voxel i is true (and there is no conjunction), when there is either: no activation in i for M_A OR there is no activation for i in M_B.

Price and Friston (1997) were the first to describe conjunction in neuroimaging. They presented a statistical method to find voxels with conjoint effects which we will call *interaction masking*. The idea behind interaction masking is to find voxels where there is an average response across the effects, and all the effects are about the same size. Consider two comparisons, A and B. Say comparison A is the difference between a verbal working memory task and a matched baseline task; let comparison B be the difference between a spatial working memory task and a matched baseline. First we find a map identifying areas of signal change due to the *main effect* of $A + B$. This is an image for the effect of $A + B$, thresholded to give 1 in areas where there is a reliable effect of $A + B$, and zeros elsewhere. This map will contain areas where effects A and B are truly present, but can also contain areas where, for example, A is present but B is not. To restrict the conjunction map to areas where effect A is similar to effect B, we create a map of the *interaction effect*, which expresses the difference between the comparisons. In general the interaction is assessed with an F test, but here the interaction is equivalently assessed by a two-tailed test of $B - A$. We remove voxels from the main effect map that are significant in the interaction map and label all remaining voxels as positive for the conjunction. This is the conjunction algorithm implemented in SPM96.

The problem with interaction masking is that we are using a statistical test to define areas where there is *no* interaction. As usual in hypothesis testing, we cannot use the lack of significance to accept the null hypothesis. In this case, we cannot assume that there is no interaction if the interaction effect is not significant. A feature of the test that differs from the standard idea of a conjunction is that it can reject an area in which all the comparisons show large effects, but where the effect sizes differ. For example, if there is a voxel where effect A is very large, and effect B is large, but smaller than A, there may be a significant difference between A and B, and interaction masking can reject this voxel from the conjunction (see Caplan and Moo, 2004, for discussion).

Friston et al. (1999a, 1999b) proposed the MS/GN test for conjunction. The test uses the minimum t statistic across several comparisons, and is based on the following logic: Imagine a voxel where effect A gives a t statistic of 0.8 and effect B gives a t statistic of 1.6. Alone, neither t value is convincing, but the fact that *both* values are well above zero suggests that there may be a real effect. This intuition can be formalized by a test on the minimum t value from these two comparisons. If there is in fact no effect of A or B then both these t statistics will be drawn from a random (null) t distribution. Assuming independence between the tests, one can find uncorrected and corrected thresholds for a minimum of two or more t statistics (Worsley and Friston, 2000). We then compare the observed minimum t value to the null

minimum t distribution to see if the observed value is unlikely to have come about by chance. In our example, the minimum t from A and B is 0.8. In fact, 0.8 falls in the top 5% of the expected distribution for the minimum of two null t values, so we can conclude that this pair of values was unlikely to have come about by chance. This is the conjunction method implemented in SPM99 and SPM2.

In our example, the MS/GN conjunction method compares the observed minimum t statistic for A and B to the null distribution of a minimum t statistic. This null distribution assumes that there is no effect for A and there is no effect for B. Recall that our definition of a conjunction null hypothesis was (not M_A) OR (not M_B). The MS/GN conjunction tests the null hypothesis (not M_A) AND (not M_B) (c.f. Eq. (2)). In general the method tests against the null hypothesis of no effect in *any* of the comparisons, which is why we call this hypothesis the *global null hypothesis*. Note that "global" here means that the null is across all effects, not across all voxels.

The problem with the MS/GN method is that it does not test for an AND conjunction. As we have already noted, the correct null hypothesis for an AND conjunction is that one or more of the comparisons has not activated. As stated in Friston et al. (1999a), the MS/GN test has a different null hypothesis, which is that *none* of the comparisons have activated. This last null hypothesis can be refuted if *any* comparison has activated. This problem leads to situations where the MS/GN gives a result that is clearly wrong if we expect an AND conjunction.

Consider the following pharmaceutical parable. Three drug companies have each made a drug which they hope will reduce blood pressure. Each company has run a study comparing their own drug to placebo in people with high blood pressure. The three drugs are A, B, and C and the three studies have yielded t values of 0.5, 1.1 and 1.3 respectively when comparing drug to placebo. Thus, none of the individual compounds had a "statistically significant" effect on blood pressure. This was painful for the manufacturers of drug A because the drug had been expensive to develop. The mood was despondent until a company statistician remembered having read a neuroimaging paper on "conjunctions". He suggested that instead of testing the drugs individually, they should test if *all* of the drugs had an effect. The MS/GN threshold for the minimum of 3 t values is 0.34, so the MS/GN test is highly significant. If the drug company interprets this test as a logical AND, they would think they had hard statistical evidence that their drug was effective, when this is clearly not the case.

Fig. 2 illustrates the drug company's problem. The t statistic for each drug could well have come about by chance; particularly the t value for A; a t value of 0.5 or higher will occur about 1 time in 3 if the data is random. However, the fact that all three values are reasonably positive *is* unlikely if we had drawn *all* of the three t values from a null t distribution—shown by the distribution of the minimum of 3 null t values. So, we have evidence for a real effect somewhere across these three drugs, but the test statistic is perfectly compatible with no effect for A or no effect for B or no effect for C.

Exactly the same problem of interpretation arises in neuroimaging. Imagine we have four tasks testing different aspects of working memory. Each of the four working memory tasks strongly activates a particular voxel in the prefrontal cortex (PFC). We now add a new task, which is looking at a flashing checkerboard. Let us say that the PFC t values for the four working memory tasks are all higher than 3. As expected, the flashing checkerboard does not

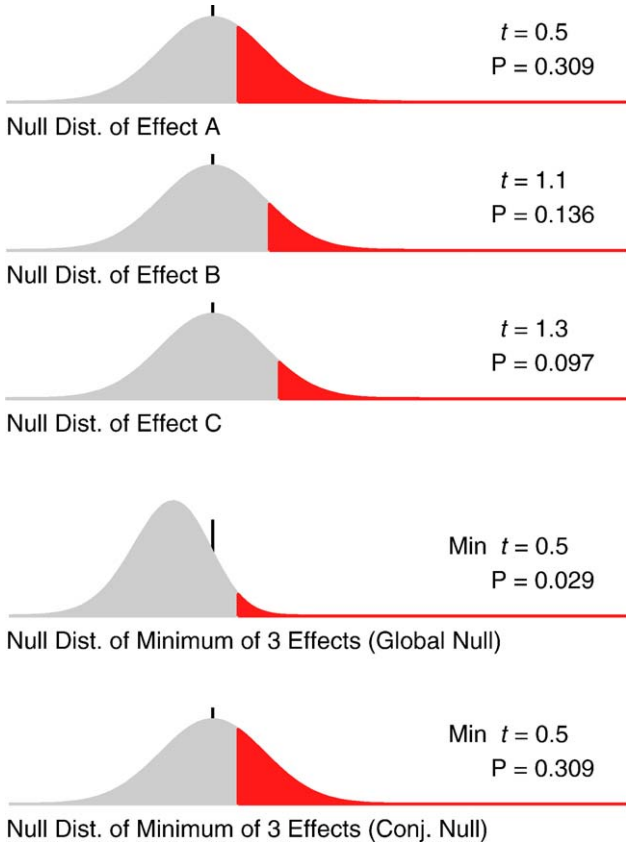

Fig. 2. Illustration of conjunctions and conjunction inference methods with drug study example. Each of the 3 distributions shown correspond to the t statistics for each of the 3 drugs.

activate the PFC, and gives us a t value of −0.1. For a MS/GN test on these 5 tasks, we assess the minimum t value under the global null (no activation in any of the 5 comparisons). The 5% uncorrected threshold for the minimum of 5 t values with a large number of degrees of freedom is −0.12. This means that the MS/GN test is significant even if the minimum t value is less than 0. If we try to interpret the MS/GN as a test of AND, we must conclude that the flashing checkerboard activates prefrontal cortex, when this is obviously false.

Note that the MS/GN test *is* valid in the situation where we really want to test against the global null. An example would be a test for *any* effect of a particular task across subjects. Here our null hypothesis should be that there is no effect in any subject. In this case we are using MS/GN for a meta analysis which combines evidence across statistic values to look for an overall effect. It is worth noting that the MS/GN test is one of the least powerful approaches to meta analysis; for a full comparison of meta-analytic methods, see Lazar et al. (2002). The use of MS/GN for meta analysis across subjects is the application described in Friston et al. (1999a) and Worsley and Friston (2000). The interpretation of a low probability from an MS/GN test across subjects is still not a logical AND; we can only conclude that *at least one* subject shows the effect. This use of the MS/GN also has the limitation that it is a fixed effects analysis and can only be used to make a statement about the cohort studied. However, the minimum statistic value can be correctly used for population inference; the primary result from the Friston et al. (1999a) paper was the use of the minimum t statistic to give a confidence interval on γ, the population

prevalence of individuals who would show activation at a given threshold. Although this application is valid, it has not been widely used.

Based on the derivations and results below, we feel the MS/GN should not be identified as a test of conjunction, but only as a meta-analytic method. Although the global null hypothesis is clearly defined in Friston et al. (1999a), our experience suggests that most authors using MS/GN are not aware that it does not test for an AND conjunction. To assess this impression, we took a sample of recent practice in neuroimaging by analyzing abstracts from the 2003 conference of the Organization for Human Brain Mapping (OHBM). OHBM is the primary conference for methods in neuroimaging, so we would expect conference abstracts to have a reasonable level of methodological sophistication. We assessed each abstract that used the MS/GN method to see if the authors intended to test for an overall effect (which would be valid) or an AND conjunction (which would be invalid).

We have argued that existing methods do not provide a valid test for a logical AND of effects. In the following sections we derive the minimum statistic to test the null hypothesis that one or more of the comparisons have not activated, the conjunction null (MS/CN). The result is straightforward; the valid test simply requires that all comparisons are individually significant at the usual level.

Methods

Let H_i^k denote the state of the null hypothesis for test k voxel i, $k = 1,\ldots,K$, $i = 1,\ldots,V$. The following definitions are for one given voxel, and so from here on we suppress the i subscript. $H^k = 0$ indicates that the null is true, $H^k = 1$ that the null is false and an effect is present. A conjunction of effects is $\cap_k\{H^k = 1\}$, and the conjunction null hypothesis is its complement

$$\mathcal{H}^{\mathrm{C}} = \bigvee_k \{H^k = 0\}. \tag{1}$$

The global null hypothesis, as used with MS/GN, is that all K tests are null

$$\mathcal{H}^{\mathrm{G}} = \bigwedge_k \{H^k = 0\}. \tag{2}$$

Write the minimum statistic as

$$M = \min_k T^k. \tag{3}$$

MS/GN conjunction inference

Before stating the attributes of the MS/GN test, we review the basic definitions of hypothesis testing (see, e.g. Schervish, 1995). The *size* of a hypothesis test is the greatest probability of a false positive, searched over the null hypothesis parameter space. Note that the null hypotheses are typically *simple*, and specify exactly one point in the parameter space (usually zero), but the conjunction null is *composite*, and corresponds to a set of parameters. The *level* is the desired or nominal false positive rate set by the user. A test is *valid* if the size is less than or equal to the level. A test is *invalid* if the size exceeds the level.

In Appendix A we show that assessing M under the global null does not control conjunction false positives. With the MS/GN test, the chance of a conjunction false positive depends on the particular configuration of null and non-null effects. When the level is α_0, the size of the MS/GN test, the worst-case false positive risk, is $\alpha_G = \alpha_0^{1/K}$. This is greater than α_0 and hence the test is invalid.

MS/CN conjunction inference

The solution to this problem is to find the worst-case configuration of null and non-null effects that comprise the conjunction null hypothesis. In Appendix A we show that the worst-case configuration is exactly that of one null effect and $K - 1$ arbitrarily large, non-null effects. Under this setting, the valid uncorrected threshold for M under the conjunction null hypothesis $\mathcal{H}^{\mathrm{C}}$ is u_{α_C}, $\alpha_{\mathrm{C}} = \alpha_0$, the usual level-$\alpha_0$ threshold for a single test. This yields our MS/CN method. Further, we show that independence between test statistics need not be assumed. (The MS/GN result assumes that, at each voxel, the K statistics $T^1, T^2, \ldots, T^K$ are independent.)

Usual corrected thresholds based on either level α_{C} or α_{G} can be used to control multiple comparisons under $\mathcal{H}^{\mathrm{C}}$ and $\mathcal{H}^{\mathrm{G}}$ respectively. For example, the Bonferroni corrected P value thresholds are α_0/V and $(\alpha_{\mathrm{G}}/V)^{1/K}$ for MS/CN and MS/GN respectively.

Calculations

We characterized degree of MS/GN's anticonservativeness by computing the conjunction error rate for different settings; in all of the settings considered the conjunction null hypothesis was true, that is, the correct action is to detect no conjunction. For a single univariate test we computed the conjunction Type I error rate for the case of $K = 2$ conjunctions. Note that the difference between the MS/CN and MS/GN thresholds will increase with K, so this is the case where the anticonservativeness of MS/GN is the least severe. We considered 5 settings, where one test's null hypothesis was true, the other test had activation magnitudes of 0, 2, 3, 4 and 6 (conjunction null true for all cases). The 5% thresholds are 0.7601 and 1.6449 for the MS/GN and MS/CN methods, respectively.

For the massively univariate imaging case, we computed the familywise conjunction Type I error rate. We used a $32 \times 32 \times 32$, $V = 32{,}768$ voxels image filled with independent, unit variance Gaussian noise. To these images we added activations in a configuration such that $\mathcal{H}^{\mathrm{C}}$ was true for all voxels but $\mathcal{H}^{\mathrm{G}}$ was false in two regions, one in each image (note that this differs from Fig. 1 panels a and b, where a conjunction exists for some voxels). We considered spherical regions of radius 1, 2, 4, 6 and 8 voxels, each consisting of 8, 32, 280, 912 and 2167 voxels respectively, crossed with activation magnitudes of 0, 2, 3, 4 and 6. We assumed independence between voxels in space; we did not consider smooth noise since smoothness will only change the precise corrected threshold used, not the qualitative interpretation of the final results.

Under independence, level α_0 familywise error (FWE) corrected P value thresholds are directly obtained as

$$\alpha_{C-\mathrm{FWE}} = 1 - (1 - \alpha_0)^{1/V} \tag{4}$$

under the conjunction null $\mathcal{H}^{\mathrm{C}}$, and

$$\alpha_{\mathrm{G}-\mathrm{FWE}} = \left(1 - (1 - \alpha_0)^{1/V}\right)^{1/K} \tag{5}$$

under the global null $\mathcal{H}^G$. For $\alpha_0 = 0.05$ and the setting considered these are 0.0000015 and 0.0012511, corresponding to Z thresholds of 4.662 and 3.023.

For both univariate and massively univariate cases the error rates have simple closed forms (see Appendix C).

Conjunction inference misuse

To gauge the typical use of the MS/GN method, we searched the abstracts presented at the 9th International Conference on Functional Mapping of the Human Brain, June 18–22, 2003, New York City (Available on CD-ROM in NeuroImage, Vol. 19, No. 2). All abstracts that contained the string "conjunction" or "conjoin" (case-insensitive) were reviewed; from these we selected all abstracts reporting results of an imaging conjunction. For each abstract we recorded the software used for the analysis. For abstracts using SPM99 or SPM2 we classified the abstracts on two dimensions. The first dimension was whether the conjunction was across subjects or across effects. The second dimension was whether the conjunction was correctly interpreted. We classified the abstract as: "Incorrect" if the abstract seemed to assume that the conjunction was evidence of a logical AND; "Correct" if the interpretation compatible with the effect not being present for every comparison, and "Unclear" if we could not judge which interpretation was being used.

Results

Univariate conjunction error rates

Table 1 shows the univariate conjunction Type I error rates for different effect magnitudes. Only for the 0 magnitude case (global null true) is the MS/GN method valid. When one effect is zero and the other effect has size 2 and larger the false positive rate is approximately 20%, 4 times the nominal 5% level. When one effect is arbitrarily large, the MS/GN conjunction false positive rate is the chance that the null effect exceeds the MS/GN threshold; for the threshold of 0.7601 this probability is $1 - \Phi^{-1}(0.7601) = 0.2236$.

The MS/CN method has a variable conjunction false positive rate, varying from $\alpha_0^{1/2} = 0.0025$ under the global null, to $\alpha_0 = 0.05$ for arbitrarily large effects. While the false positive rate varies, it never exceeds 0.05, demonstrating its validity.

Familywise conjunction error rates

Tables 2 and 3 show the familywise conjunction error rates under the different conditions. They record the probability of one or more conjunction false positives, searching over the minimum statistic image, for each thresholding method. While the MS/GN conjunction method controls FWER under the global null (where both effects are absent), for all other configurations the FWER was significantly greater than 0.05.

Table 3 shows the same results for the our conjunction method. For all configurations considered our method controls the FWER well below the nominal level of 0.05. Our method has a variable false positive rate, which is due to the composite nature of the conjunction null hypothesis $\mathcal{H}^C$. Our method must protect against the worst case, and thus false positive rate will not equal the size in non-worst-case settings. (In this situation, the worst-case setting is one statistic image having arbitrarily large effects at every voxel, the other image filled with null statistic values. Hence to make the greatest false positive rate in Table 3 approach 0.05, we would need to use a effect spatial extent corresponding to an whole-image activation.)

Conjunction inference misuse

There were 68 abstracts that contained the string "conjoin" or "conjunction", of which 42 reported the results of an imaging conjunction. 33 abstracts used SPM99 or SPM2; we describe these in more detail below. The other software packages were BrainVoyager (2 abstracts), AFNI (1) and MEDx (1). For the remaining 5 abstracts we could not be sure which algorithm had been used.

We record the classification of the SPM99/SPM2 abstracts in Table 4. Three quarters of abstracts using conjunctions across tasks were incorrectly using MS/GN conjunctions as evidence of logical AND. For example, one abstract used MS/GN "to reveal the brain regions *activated by all three* sensory modality conditions." Another abstract used MS/GN "...to find common areas *activated during both* retrieval and articulation processes ..." (all italics our own). Only 1 of 25 abstracts using conjunctions across tasks

Table 1
Univariate conjunction error rate, comparing original MS/GN and proposed MS/CN method

		Effect magnitude				
		0	2	3	4	6
Method	MS/GN	0.0500	0.1996	0.2208	0.2235	0.2236
	MS/CN	0.0025	0.0319	0.0456	0.0495	0.0500

For both methods a level 0.05 threshold is used on $K = 2$ tests. The error rates shown are the probability of each test rejecting when the conjunction null is true. For all settings one effect is zero, and the other effect has the specified magnitude.

Table 2
Familywise conjunction error rate, original MS/GN method

	Radius	Effect magnitude				
		0	2	3	4	6
Effect spatial extent	1	0.0500	0.0529	0.0593	0.0657	0.0688
	2		0.0615	0.0865	0.1114	0.1230
	4		0.1459	0.3259	0.4707	0.5279
	6		0.3283	0.6893	0.8587	0.9026
	8		0.5845	0.9340	0.9899	0.9959

The MS/GN's $\alpha_0 = 0.05$ FWER threshold of 4.332 is used on the minimum of $K = 2$ images. Note that the conjunction FWER is not controlled, and approaches 1 for large effect magnitudes and radii.

Table 3
Familywise conjunction error rate, proposed MS/CN method

	Radius	Effect magnitude				
		0	2	3	4	6
Effect spatial extent	1	0.0000	0.0000	0.0000	0.0000	0.0000
	2		0.0000	0.0000	0.0000	0.0001
	4		0.0000	0.0000	0.0002	0.0008
	6		0.0000	0.0001	0.0007	0.0026
	8		0.0000	0.0003	0.0017	0.0062

The MS/CN's $\alpha_0 = 0.05$ FWER threshold of 4.662 used on minimum of $K = 2$ images. Note that the conjunction FWER is always controlled below the nominal 5% level.

Table 4
Classification of OHBM abstracts using SPM for imaging conjunction

	Correct	Unclear	Incorrect
Subjects	3	1	4
Tasks	1	5	19

appeared to interpret the MS/GN correctly. Conjunctions across subjects were more likely to be interpreted correctly; only half were incorrect, and 3 of 8 appeared to be correct.

Discussion

We have described the MS/GN method used by SPM99 and SPM2, and shown that this is not a valid test for conjunction inference. If we have two comparisons A and B, then a valid conjunction test should allow us to draw the following conclusion: "I can be reasonably confident that there is an effect in both A and B." In contrast, the correct reporting of a MS/GN statistic would read something like "I can be reasonably confident that there is some effect in A, or B, or both." It is therefore unfortunate that the MS/GN has been labeled a conjunction test. This label has led to great confusion in the imaging literature. Our analysis of recent OHBM abstracts shows that most users of SPM wrongly assume that the MS/GN is a test of conjunction.

We have proposed the MS/CN method as a conjunction test. Unlike the MS/GN, the MS/CN test does control the false positive (type I) error for conjunction inference. The use of our MS/CN method in a typical data analysis is very straightforward. To test for common activations across subjects or tasks, one simply creates the intersection of statistical maps thresholded at a specified alpha rate (e.g., 0.05 FWE-corrected, from Random Field Theory or permutation).

We stress that our MS/CN result is valid under any type of dependence. (While modified MS/GN methods which account for dependence have been proposed (Taylor and Worsley, 2003), they are not yet available for routine use.) Dependent effects routinely arise in practice, such as when different tasks are compared to a common control. For example, if tasks A and B must be contrasted with control C to be meaningful, a conjunction of the effects $A - C$ and $B - C$ poses no difficulties for our method.

A potential weakness of the MS/CN method is its variable (though valid) false positive rate. The reason for this is the composite nature of the conjunction null, which causes the conjunction false positive rate to depend on the true state of affairs, that is, on how many null hypothesis are actually true, of which we can have no knowledge.

The MS/GN test will give the correct rate of false positives for a conjunction test only if all brain voxels fall exclusively into two categories, one where there is no activation in any contrast (the MS/GN null) and one where there is activation in every contrast (a conjunction of effects). Clearly this will never be the case, and the MS/GN test will thus always be inexact and invalid (i.e. anti-conservative).

The MS/CN test has the nominal false positive rate only if the voxels in the brain fall exclusively into two categories, one where there is activation in all but one contrast, and one where there is activation in every contrast (no false positive possible). This is of course never going to be the case either and the MS/CN test will in practice have variable *but* valid false positive rate (i.e. possibly conservative).

So, in practice, neither test will have the nominal false positive rate, but only one will be valid. The "true" threshold, that is, the one that will give us an exact test, will always lie somewhere between these two extremes but its value will depend on the true state of affairs, that is it will be different from study to study and always unknown to us. Given that a majority of voxels in the brain is NOT activated in any representative study it would be fair to say that the MS/GN threshold will typically be closer to the "true" (that which will give an exact test) threshold. However it will always be an underestimation, and it will always be anticonservative.

Recently, Friston et al. (2005) have pointed out that we can use a minimum statistic test with a null hypothesis that is between the global null and conjunction null. Let us imagine that we have 3 comparisons to compare—say 3 tasks. The MS/GN test allows us to conclude that at least one comparison shows an effect. The MS/CN (conjunction) test allows us to conclude that all 3 of the comparisons show an effect. However, we can also use an intermediate null hypothesis. Here the intermediate null (IN) would be that two or more of the comparisons do not have an effect. If we refute this null, we can conclude that at least 2 of the comparisons show an effect. We term this test the Minimum Statistic compared to Intermediate Null (MS/IN).

The motivation for this test is best given by example. Let us imagine that our 3 tasks are oddball detection tasks: one using visual stimuli, one using auditory stimuli, and a third using somatosensory stimuli. Our assumption is that the only process that that can be shared by *any* 2 of these 3 tasks is oddball detection. We apply the MS/IN test at a certain voxel, and find that it gives a small P value. This allows us to conclude that at least 2 tasks have an effect in this area. If we assume that the only process in common across any two of these tasks is oddball detection, then we can conclude that this area is involved in detecting oddballs. However, we have to be careful not to draw the conclusion that we have found an area that is involved in oddball detection regardless of modality. For example, we might have found an area that responds only to visual and auditory oddballs, but not to somatosensory oddballs. In general, the test does not allow us to conclude that the voxel responds to *all* oddballs, unless we make the following assumption: "if an area responds to two of the types of oddballs, it also responds to the third." This assumption is obviously difficult to justify. We also note that, unlike the MS/GN, the use of the MS/IN requires an assumption of independence for validity.

This work has focused solely on the minimum statistic for assessing conjunctions. While well-motivated, there are other possible methods to measure evidence of conjunctions. One approach, which uses Bayesian posterior probabilities of the conjunction null, may be fruitful (Nichols and Wager, 2004). Other, non-minimum statistic based methods are an open area of research.

In this article, we have tried to address the technical problems of finding a valid test to test for a conjunction of effects. We have not addressed the conceptual validity of conjunction to answer problems in functional imaging; see Caplan and Moo (2004) for a detailed discussion.

We hope this work clarifies the interpretation of conjunction inference based on the minimum statistic. Published results that use P values based on MS/GN tests must be carefully considered, as they can only be regarded as statistical evidence of one or more effects being active.

Acknowledgments

The authors wish to thank Karl Friston, Will Penny and Daniel Glaser for valuable feedback on this work.

Appendix A. Anticonservativeness of the MS/GN method

Consider K Gaussian random variables with mean vector μ and, without loss of generality, unit variance. We presently assume that they are independent, but we return to the dependent case below. Let the test statistic for effect k be

$$T^k \sim \mathcal{N}(\mu^k, 1), \tag{6}$$

where μ^k is the effect size, with $\mu^k = 0$ if $H^k = 0$. We consider only a single voxel here, and so suppress the spatial index for this appendix. We write $\Omega_0 \subset \mathbb{R}^K$ for the set of parameters μ satisfying the conjunction null hypothesis.

The MS/GN P value threshold $\alpha_G = \alpha_0^{1/K}$ controls false positives under $\mathcal{H}^G$ at level α_0. To see this, let $u_G = \Phi^{-1}(1 - \alpha_G)$ be the corresponding statistic threshold, where Φ is the CDF of a standard Gaussian. Then

$$\mathbf{P}\left(\min_k T^k \geq u_G | \mathcal{H}^G\right) = \prod_k \mathbf{P}\left(T^k \geq u_G | H^k = 0\right) \tag{7}$$

$$= (1 - \Phi(u_G))^K \tag{8}$$

$$= \alpha_G^K \tag{9}$$

$$= \alpha_0 \tag{10}$$

Since $\mathcal{H}^G$ corresponds to a single element or vector of parameters it is simple, while $\mathcal{H}^C$ is composite, as it corresponds to a family of parameters (Schervish, 1995). The size of a test with a composite null is the supremum of rejection probability over the null:

$$\alpha = \sup_{\mu \in \Omega_0} \mathbf{P}_\mu\left(\min_{k=1,..,K} T^k > u_G\right). \tag{11}$$

The notation $\mathbf{P}_\mu$ reminds us that the probability depends on a varying true mean μ. To evaluate this, consider just the case exactly of j nulls being true:

$$A_j = \sup_{\mu \in \Omega_0^j} \mathbf{P}_\mu\left(\min_{i=1,..,K} T^k > u_G\right) \tag{12}$$

where Ω_0^j is the subset of Ω_0 for which exactly j of the K null hypotheses are true. Note that the $\{\Omega_0^j\}$ partition Ω_0, so that $\alpha = \max A_j$.

By symmetry we can just consider the first j nulls to be true. Since we are interested in the supremum, and since the minimum operator is nondecreasing in its operands, it is sufficient to consider extreme behavior of the $K - j$ unconstrained μ^ks. In particular, as any of the unconstrained μ^k go to infinity the minimum can only get larger (letting them go to negative infinity is not of interest since this will only reduce the minimum). For each $j \geq 1$,

$$A_j = \sup_{\mu \in \Omega_0^j} \mathbf{P}_\mu\left(\min_{i=1,..,K} T^k \geq u_G\right) \tag{13}$$

$$= \limsup_{\mu_k \to \infty, k > j} \mathbf{P}_\mu\left(\min_{i=1,..,K} T^k \geq u_G | \mu^{k'} = 0, 1 \leq k' \leq j\right) \tag{14}$$

$$= \mathbf{P}\left(\min_{k=1,..,j} T^k \geq u_G | \mu^{k'} = 0, 1 \leq k' \leq j\right) \tag{15}$$

$$= \prod_{k=1}^{j} \mathbf{P}\left(T^k \geq u_G | \mu^k = 0\right) \tag{16}$$

$$= \alpha_G^j \tag{17}$$

$$= \alpha_0^{j/K} \tag{18}$$

Then we have the final result by finding the size as the maximum of the A_j

$$\alpha = \max_{1 \leq j \leq K} A_j \tag{19}$$

$$= \alpha_0^{1/K} \tag{20}$$

$$> \alpha_0 \tag{21}$$

Hence the MS/GN threshold does not control the false positive rate for the conjunction null $\mathcal{H}^C$. We must force the worst (greatest) A_j, $A_1 = \alpha_0^{1/K}$, to be α_0. But $A_1 = \alpha_G$, so α_G must be changed to be precisely α_0. Hence the appropriate threshold to control conjunction false positives is

$$u_C = \Phi^{-1}(1 - \alpha_0), \tag{22}$$

the usual level α_0 threshold.

To account for dependence, note that for the previous result we only relied on independence to express Eq. (16) as a product of j probabilities. Without independence, we can instead use a simple probability inequality to show that $A_j \leq A_1$, the worst case, and so the proof holds as is.

Note that the event $\{\min_k T^k > u_G\}$ is equivalent to $\cap_k \{T^k > u_G\}$, and that a probability of an intersection is smaller than the probability of the individual events (for events $\{E_k\}$, $\mathbf{P}(\cap_k E_k) \leq \mathbf{P}(E_{k'})$, for any k'). Then

$$A_j = \mathbf{P}\left(\min_{k=1,..,j} T^k \geq u_G | \mu^{k'} = 0, 1 \leq k' \leq j\right) \tag{23}$$

$$= \mathbf{P}\left(\underset{k=1,..,j}{u} \left\{T^k \geq u_G\right\} | \mu^{k'} = 0, 1 \leq k' \leq j\right) \tag{24}$$

$$\leq \min_{k=1,..,j} \mathbf{P}\left(T^k \geq u_G | \mu^{k'} = 0, 1 \leq k' \leq j\right) \tag{25}$$

$$= \alpha_G \tag{26}$$

$$= A_1 \tag{27}$$

We did not use the result in the first place, since it is useful to observe that how the false positive rate varies with j.

Appendix B. Valid familywise conjunction error rate control

Let M_i be the minimum over the K effects at voxel i. To control the FWER on a minimum statistic image $\{M_i\}$, a threshold must satisfy

$$\mathbf{P}\left(\max_i \; M_i > u_{\mathrm{FWE}}\right) \leq \alpha_0, \tag{28}$$

where the maximum is over space (not conjunctions). To control FWER under the conjunction null $\mathcal{H}^{\mathrm{C}}$, this expression must hold for the worst-case configuration, when M_i is as large as possible. This is the case when all but one of the K nulls is true at each voxel i. That is, as in the previous section,

$$\alpha_{\mathrm{FWE}} = \sup \mathbf{P}\left(\max_i \; M_i > u_{\mathrm{FWE}}\right) \tag{29}$$

$$= \mathbf{P}\left(\max_i \; T_i > u_{\mathrm{FWE}}\right). \tag{30}$$

Hence any standard FWER method valid for a single statistic image $\{T_i\}$ can be used to threshold a minimum statistic image $\{M_i\}$.

Appendix C. Univariate and FWER thresholds and rejection rates

Assuming independence across space, the FWER for threshold u applied to $\{M_i\}$ is

$$\alpha_{\mathrm{FWE}} = 1 - \prod_i \left(1 - \prod_k \mathbf{P}\left(T_i^k \geq u\right)\right). \tag{31}$$

To find the FWER global null P value threshold, assume that $\mathcal{H}^{\mathrm{G}}$ holds for all i and set the previous expression α_0. Solving for $\alpha = \mathbf{P}(T \geq u)$,

$$\alpha_{\mathrm{G\text{–}FWE}} = \left(1 - (1 - \alpha_0)^{1/V}\right)^{1/K}. \tag{32}$$

Setting $K = 1$ gives the P value threshold for the conjunction null,

$$\alpha_{\mathrm{C\text{–}FWE}} = 1 - (1 - \alpha_0)^{1/V}. \tag{33}$$

The chance of a familywise conjunction error can be found by similar computation. For statistic threshold u and configuration of true means $\mu = \{\mu_{ki}\}$, the probability of rejecting is

$$\phi_\mu(u) = 1 - \prod_i \left(1 - \prod_k \mathbf{P}_{\mu_{ki}}\left(T_i^k \geq u\right)\right). \tag{34}$$

The method used (MS/GN or MS/CN) determines the threshold u, and the null hypothesis of interest (GN or CN) dictates the means. Setting $V = 1$ gives the univariate case

$$\phi_\mu(u) = \prod_k \mathbf{P}_{\mu_k}\left(T^k \geq u\right). \tag{35}$$

References

Caplan, D., Moo, L., 2004. Cognitive conjunction and cognitive functions. NeuroImage 21, 751–756.

Friston, K., Holmes, A., Price, C., Büchel, C., Worsley, K., 1999a. Multisubject fMRI studies and conjunction analyses. NeuroImage 10, 385–396.

Friston, K.J., Holmes, A.P., Worsley, K.J., 1999b. Comments and controversies: how many subjects constitute a study. NeuroImage 10, 1–5.

Friston, K.J., Penny, W.D., Glaser, D.E., 2005. Conjunction revisited. NeuroImage 25, 661–667, doi:10.1016/j.neuroimage.2005.01.013 (this issue).

Lazar, N.A., Luna, B., Sweeney, J.A., Eddy, W.F., 2002. Combining brains: a survey methods for statistical pooling of information. NeuroImage, 538–550.

Mendelson, Elliot, 1987. Introduction to Mathematical Logic, (3rd ed.) Wardsworth and Brooks/Cole.

Nichols, T.E., Wager, T.D. 2004. Conjunction inference using the Bayesian interpretation of the positive false directory rate. Presented at the 10th international conference on functional mapping of the human brain, June 14-17, 2004, Budpest Hungary. Available on CD-Rom in NeuroImage, 22.

Price, C.J., Friston, K.J., 1997. Cognitive conjunction: a new approach to brain activation experiments. NeuroImage 5, 261–270.

Schervish, M.J., 1995. Theory of Statistics. Springer-Verlag.

Taylor, J.E., Worsley, K.J., 2003. Correlated conjunctions. Presented at the 9th International Conference on Functional Mapping of the Human Brain, June 18–22, 2003, New york. Available on CD Rom in NeuroImage, 19 (2).

Worsley, K.J., Friston, K.J., 2000. A test for conjunction. Stat. Probab. Lett. 47 (2), 135–140.