

A significance test of interaction in $2 \times K$ designs with proportions

George A. Michael
Université Lyon 2

When investigating the deficits of a single patient, psychologists usually compare his/her performance in one or more tests to the performance of a control group. This can be done for any kind of variables, provided (i) that the design does not require the investigation of interactions between two or more factors, (ii) that the comparison between two or more individuals is not desired, and (iii) that the collection of the control data is possible. Yet, researchers are constantly interested in assessing interactions in the performance of an individual, and in the comparison of two or more individuals for investigating double dissociations or the efficiency of different methods of therapy, etc. They also may desire to investigate cases where only extremely simple and easy tasks can be performed, where ceiling effects are observed in the performance of the controls, and thus the case-controls comparison is impossible. The available statistical tools for the analysis of intra-individual or inter-individual performance (mainly with proportions) do not offer the possibility to assess interaction, they are not appropriate when some cells may contain 0 or 1 proportions, and when the sample size is small. Here, we present the Q' test which may be used to test the hypothesis of equal proportions and proportion differences in $2 \times K$ designs, offering therefore the possibility for researchers to investigate the main effects and interaction. This test can be used for any sample size and even when the data contains extreme proportions. Finally, a procedure of multiple comparisons described in this paper may be used to locate statistically significant sources of variance and differences.

Damaged brains and disordered minds have been studied by psychologists mostly through single-case studies, where the performance of a patient is compared to the performance of a normative sample. In recent years, some statistical tests have been adapted to the single-case design with one, two, and K tests (Crawford & Garthwaite, 2002). However, the use of these otherwise remarkable tests is limited when investigating interactions between factors, when comparing two or more individuals for purposes of assessing double dissociations (a patient exhibits a deficit in task X but not in task Y , and a second patient exhibits exactly the opposite pattern of performance) or the

efficiency of different methods of therapy or reeducation, and, most interestingly, when investigating the performance of patients presenting with massive cognitive impairments. These patients can perform tasks the simplicity and ease of which do not allow to collect data from control groups, just because controls perform 100% correctly. The comparison to the normative data is thus impossible because of the absence of variance in the control group.

Cases where data is made of proportions and where intra-individual, inter-individual, or pooled group analyses are possible, may offer scientist's some exceptional opportunities to study human cognition and its breakdown, but also to study some phenomena in other domains of fundamental and applied research. At this aim, researchers frequently use the classical chi-square tests. Yet, these tests can be used with confidence only if the number of trials per condition is large ($N > 40$), and no low scores are observed

I would like to thank Louis Laurencelle for his useful comments and his commitment to helping making this a valuable contribution to the literature.

($r < 5$). Tests that assess the main effects and the interaction between at least two factors with proportions are rarely accessible to social, cognitive and other behavioral scientists. Marascuilo (1970) presented a test for the comparison of K independent sensitivity indexes, d -primes, of the signal detection theory (Green & Swets, 1966), and suggested its use either for the analysis of individual or pooled group data. This test is of most interest to our purpose, because, since a d -prime is the difference between two normalized proportions, the difference between K d -primes corresponds to the interaction in a $2 \times K$ design.

The direct transposition of this method to the analysis of non-normalized proportions is, however, not recommended. As a point of fact, Marascuilo (1970) uses a variance (Gourevitch & Galanter, 1967) based on the Wald variance of a proportion. This variance is well known and can be computed as follows,

$$\frac{\hat{p}(1-\hat{p})}{N} \quad (1)$$

where \hat{p} is the proportion, and N the sample size. The use of this variance seemingly leads to several kinds of anomalies (Newcombe, 1998), and its use is not recommended (Newcombe & Altman, 2000). The Q' test of $2 \times K$ interaction presented in this paper is a modified version of the Marascuilo test (1970) in which a different variance is introduced. This variance, can be computed as follows,

$$\frac{z^2 + 4N_{k1}\hat{p}_{k1}(1-\hat{p}_{k1})}{4(N_{k1} + z^2)^2} \quad (2)$$

where z is the $z_{1-\alpha/2}$ from the standard Normal distribution. This equation is extracted from the equation of the Wilson confidence interval of proportions (Newcombe & Altman, 2000; Brown et al., 2001):

$$IC = \frac{2N_x\hat{p}_x + z^2 \pm z\sqrt{z^2 + (4N_x\hat{p}_x(1-\hat{p}_x))}}{2N_x + z^2} \quad (3)$$

The Wald and the Wilson methods are both approximate but differ importantly in some points:

The Wald variance performs quite well when the sample size is large ($N > 40$), whilst the variance is inflated when the sample size is small. On the contrary, the Wilson variance is applicable for all samples and, interestingly, it performs just as well as the Wald variance for large samples.

The Wald variance should not be used for very low or very large observed proportions, and its value is 0 when the proportion is 0 or 1. No such restrictions exist for the Wilson variance, which can be applied in all cases and the value of which is not 0 for extreme proportions.

Finally, there are serious recommendations that the use of the Wald method requires that neither r nor $N-r$ is less than 5 (Newcombe & Altman, 2000). The Wilson method can

be applied in all cases.

These differences are visible in Figure 1, where the Wald and the Wilson variances are compared for proportions from 0 to 1, for a small ($N = 10$) and a large ($N = 60$) sample. The two methods perform equally well for the large sample, and this clearly establishes their direct link and equivalence. However, the Wilson variances are less inflated for the small sample than the Wald variances, and this allows a better interpretation of the results. As pointed out by Newcombe and Altman (2000), the Wald method leads to "too extreme an interpretation of the data, and sometimes do not make sense" (p. 46). Thus, unless the sample size is large and no extreme proportions are present (these conditions are rarely met in some research domains), the Wilson variance seems more appropriate and more adequate for the analysis of proportions and differences in proportions. For reasons of applicability to all samples and whichever the proportion, the Q' test presented here uses the Wilson variance, recommended for its applicability to any data (Newcombe, 1998; Agresti & Coull, 1998).

The Q' test: A $2 \times K$ test with proportions

In the $2 \times K$ design, the magnitude of difference between 2 proportions (signalled as 1 and 2) is compared in K (1, 2, ..., K) different conditions, where a condition may represent a group of observers (pooled group data), a single observer, or even a single test from individual data. The resulting test statistic has a χ^2 distribution with $\nu = (K-1)$ degrees of freedom. For each of the K conditions, let the estimates of a difference in two proportions be denoted by \hat{d}_k :

$$\hat{d}_k = \hat{p}_{k1} - \hat{p}_{k2} \quad (4)$$

where \hat{p}_1 and \hat{p}_2 denote the proportions to compare, and where $k = 1, 2, \dots, K$. Let the variance of the proportions be denoted by Var_{k1} and Var_{k2} , respectively:

$$Var_{k1} = \frac{z^2 + 4N_{k1}\hat{p}_{k1}(1-\hat{p}_{k1})}{4(N_{k1} + z^2)^2}$$

and

$$Var_{k2} = \frac{z^2 + 4N_{k2}\hat{p}_{k2}(1-\hat{p}_{k2})}{4(N_{k2} + z^2)^2}$$

(see equation 2) where N_{k1} and N_{k2} represent the sample size in the $k1^{th}$ and $k2^{th}$ condition, and where z is the $z_{1-\alpha/2}$ from the standard Normal distribution (i. e., 1,96). The variance of the proportion difference is:

$$\hat{D}_k = Var_{k1} + Var_{k2} \quad (5)$$

and the Q' is:

$$Q' = \sum_{k=1}^K \frac{(\hat{d}_k - \hat{d}_0)^2}{\hat{D}_k} \quad (6)$$

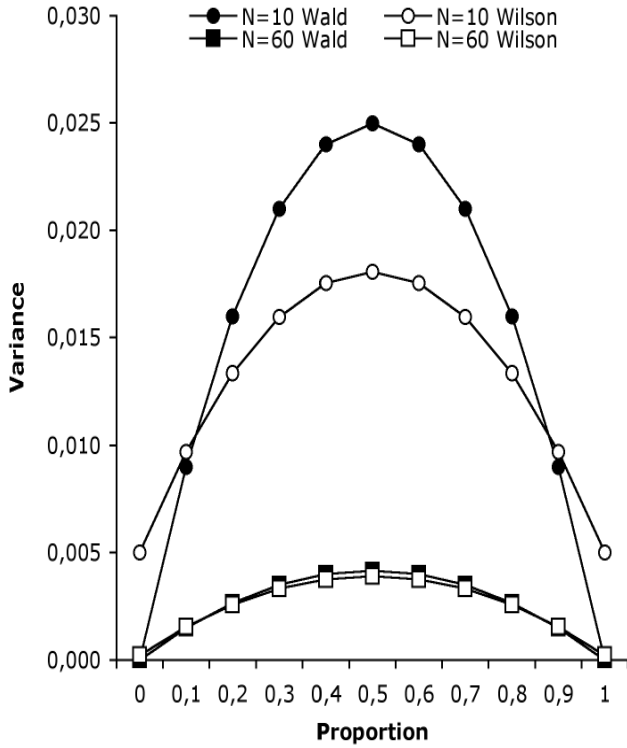


Figure 1: A direct comparison between the Wald and the Wilson variances for different proportions and for sample sizes of $N=10$ and $N=60$. The two variances are very similar for the large sample size. However, for the small sample size, there are important differences between the two variances. The greatest differences are observed for the extreme (0 and 1) and the intermediate proportions. The Wilson variance is thus applicable under the same conditions as the Wald variance, and it performs better when the sample size is not large, and when extreme proportions are observed. For these reasons, the Wilson method should be preferred to the Wald method.

where

$$\hat{d}_0 = \frac{\sum_{k=1}^K \hat{d}_k}{\sum_{k=1}^K \frac{1}{\hat{D}_k}} \quad (7)$$

As mentioned above, the Q' test statistic has a χ^2 distribution with $\nu = (K-1)$ degrees of freedom and the corresponding critical value above which a difference among the tested conditions is significant can be read in a χ^2 table.

A worked example with practical steps

The data must be first arranged in a $2 \times K$ table. Let's consider the following data, taken from a study where patient RR, suffering from progressive agnosia due to posterior cortical atrophy, was required to name pictures of usual objects. The null hypothesis is that patient RR's naming performance does not vary as a function of the picture color and semantic category of the objects. The

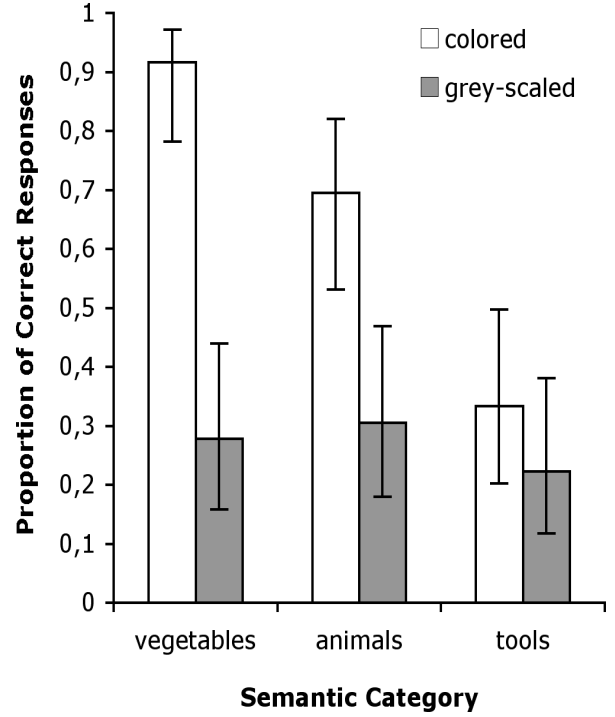


Figure 2: Graphic representation of the performance (proportion of correct responses) of patient RR in a picture naming task. The pictures were either colored or grey-scaled, and were representing usual objects of three semantic categories, vegetables, animals and tools. Error bars represent 95% Wilson confidence intervals for proportions (eq.3). The visual analysis of this graphic suggests that picture color and object semantic category interact, and this is confirmed by the Q' test ($Q'(2) = 15.82, P = 0.0004$).

pictures could be colored or grey-scaled (i.e., factor 1: color), and could represent (a) vegetables, (b) animals, or (c) tools (i.e., factor 2: semantic category). Thirty-six pictures per condition ($N = 36$) were presented. It is not obligatory for the N to be the same for all conditions. The results are presented in Table 1 and depicted graphically in Figure 2.

The $2 \times K$ interaction

There are seven steps of computations in order to assess the color X semantic category interaction, and they are represented in Table 1.

Step 1: compute the proportions for each condition, by dividing the observed score by the corresponding number of trials

$$\hat{p}_{a1} = \frac{r_{a1}}{N_{a1}} = \frac{33}{36} = 0,917$$

where r_{a1} is the score in the a_1^{th} cell, N_{a1} is the sample size for the a_1^{th} condition, and \hat{p}_{a1} is the proportion of the a_1^{th} condition and so forth.

Step 2: compute the variance for each proportion, for

Table 1. Scores (correct responses) obtained by patient RR in a naming task with colored (1) and grey-scaled (2) pictures of vegetables (a), animals (b) and tools (c), and the seven steps necessary for the analysis of the color X semantic category interaction with proportions. Thirty-six pictures were presented in each tested condition.

		group	a	b	c
Score (r)		1	33	25	12
		2	10	11	8
Step 1	Proportion (\hat{p})	1	0,917	0,694	0,333
		2	0,278	0,306	0,222
Step 2	Variance (Var)	1	0,0023	0,0054	0,0056
		2	0,0052	0,0054	0,0045
Step 3	Proportion Difference (\hat{d})		0,639	0,389	0,111
Step 4	Variance of Difference (\hat{D})		0,0075	0,0108	0,0102
Step 5	\hat{d} / \hat{D}		85,27	35,89	10,93
Step 6	$1 / \hat{D}$		133,47	92,30	98,33
Step 7	Contribution ($(\hat{d} - \hat{d}_0)^2 / \hat{D}$)		7,12	0,03	8,67

example:

$$Var_{\hat{p}_{a1}} = \frac{z^2 + 4N_{a1}\hat{p}_{a1}(1-\hat{p}_{a1})}{4(N_{a1} + z^2)^2} = \frac{1,96^2 + 4 \times 36 \times 0,917 \times (1-0,917)}{4(36 + 1,96^2)^2} = 0,0023$$

Step 3: For each condition of factor 2 (a, b, ..., k), compute the proportion differences between the two conditions of factor 1 (1, 2):

$$\begin{aligned}\hat{d}_a &= \hat{p}_{a1} - \hat{p}_{a2} = 0,917 - 0,278 = 0,639 \\ \hat{d}_b &= \hat{p}_{b1} - \hat{p}_{b2} = 0,694 - 0,306 = 0,389 \\ \hat{d}_c &= \hat{p}_{c1} - \hat{p}_{c2} = 0,333 - 0,222 = 0,111\end{aligned}$$

Step 4: Compute the variance of each difference:

$$\begin{aligned}\hat{D}_a &= Var_{\hat{p}_{a1}} + Var_{\hat{p}_{a2}} = 0,0023 + 0,0052 = 0,0075 \\ \hat{D}_b &= Var_{\hat{p}_{b1}} + Var_{\hat{p}_{b2}} = 0,0054 + 0,0054 = 0,0108 \\ \hat{D}_c &= Var_{\hat{p}_{c1}} + Var_{\hat{p}_{c2}} = 0,0056 + 0,0045 = 0,0102\end{aligned}$$

Step 5: divide each proportion difference computed in Step 3 by the corresponding variance of the difference computed in Step 4:

$$\begin{aligned}\frac{\hat{d}_a}{\hat{D}_a} &= \frac{0,639}{0,0075} = 85,27, \quad \frac{\hat{d}_b}{\hat{D}_b} = \frac{0,389}{0,0108} = 35,89, \text{ and} \\ \frac{\hat{d}_c}{\hat{D}_c} &= \frac{0,111}{0,0102} = 10,93\end{aligned}$$

Step 6: divide 1 by the variance of the proportion difference:

$$\frac{1}{\hat{D}_a} = \frac{1}{0,0075} = 133,47, \quad \frac{1}{\hat{D}_b} = \frac{1}{0,0108} = 92,30, \text{ and}$$

$$\frac{1}{\hat{D}_c} = \frac{1}{0,0102} = 98,33$$

The d_0 can be obtained by dividing the sum of the values obtained in Step 5 by the sum of the values obtained in Step 6:

$$\hat{d}_0 = \frac{85,27 + 35,89 + 10,93}{133,47 + 92,30 + 98,33} = \frac{132,09}{324,10} = 0,408$$

Step 7: for each proportion difference, compute the contribution to the interaction:

$$\begin{aligned}\frac{(\hat{d}_a - \hat{d}_0)^2}{\hat{D}_a} &= \frac{(0,639 - 0,408)^2}{0,0075} = 7,14 \\ \frac{(\hat{d}_b - \hat{d}_0)^2}{\hat{D}_b} &= \frac{(0,389 - 0,408)^2}{0,0108} = 0,03 \\ \frac{(\hat{d}_c - \hat{d}_0)^2}{\hat{D}_c} &= \frac{(0,111 - 0,408)^2}{0,0102} = 8,64\end{aligned}$$

The sum of the values obtained in Step 7 is the Q' value:

$$Q' = 7,14 + 0,03 + 8,64 = 15,82$$

If the Q' is equal or bigger than the value of χ^2 read in a statistical table, then the null hypothesis is rejected and the tested interaction is significant at the corresponding P level. For $df = 3-1 = 2$, the corresponding χ^2 value for $P = 0,05$ is 5,99. We can thus reject the null hypothesis. The interaction is significant, suggesting that patient RR's performance varies as a function of the color of the pictures and the object semantic category.

The main effects

One may use the Marascuilo procedure (Marascuilo,

1966) in order to derive a χ^2 for the main effect of each factor. However, the use of this method requires large samples, and scores that are bigger than 5. The method we describe below is very similar to the one used for the assessment of the interaction, and it can be used with any data. Let's consider the steps of the main effect of factor 2, in which 3 conditions were tested. The data is collapsed across the non-tested factor.

Step 1: The collapsed scores for a , b , and c are $r_a = 43$, $r_b = 36$ and $r_c = 20$, respectively, and the sample size for each condition is 72. The respective proportions are $\hat{p}_a = 0,597$, $\hat{p}_b = 0,500$ and $\hat{p}_c = 0,278$. The overall pooled proportion is

$$\hat{p}_{abc} = (r_a + r_b + r_c) / (N_a + N_b + N_c) = 0,458.$$

Step 2: The variance of each proportion, computed with the equation presented earlier (eq.2), is $Var_a = 0,0032$, $Var_b = 0,0033$, $Var_c = 0,0027$, and the variance of the pooled proportion is $Var_{abc} = 0,0011$.

Step 3: For each condition, the difference with the pooled proportion is derived:

$$\hat{d}_a = 0,597 - 0,458 = 0,139$$

$$\hat{d}_b = 0,500 - 0,458 = 0,042$$

$$\hat{d}_c = 0,278 - 0,458 = -0,180$$

Step 4: Compute the variance of the difference between each proportion and the pooled proportion:

$$\hat{D}_a = 0,0032 + 0,0011 = 0,0043$$

$$\hat{D}_b = 0,0033 + 0,0011 = 0,0044$$

$$\hat{D}_c = 0,0027 + 0,0011 = 0,0038$$

Step 5: divide each proportion difference by the corresponding variance of difference:

$$\frac{\hat{d}_a}{\hat{D}_a} = \frac{0,139}{0,0043} = 32,32, \quad \frac{\hat{d}_b}{\hat{D}_b} = \frac{0,042}{0,0044} = 9,49 \quad \text{and}$$

$$\frac{\hat{d}_c}{\hat{D}_c} = \frac{-0,180}{0,0038} = -47,33$$

Step 6: divide 1 by the variance of each proportion difference:

$$\frac{1}{\hat{D}_a} = \frac{1}{0,0043} = 232,16, \quad \frac{1}{\hat{D}_b} = \frac{1}{0,0044} = 225,95 \quad \text{and}$$

$$\frac{1}{\hat{D}_c} = \frac{1}{0,0038} = 262,63$$

Compute the \hat{d}_0 by dividing the sum of the values obtained in Step 5 by the sum of the values obtained in Step 6:

$$\hat{d}_0 = \frac{32,32 + 9,49 - 47,33}{232,16 + 225,95 + 262,63} = \frac{-5,521}{720,74} = -0,008$$

Step 7: for each proportion, compute the contribution of

each difference to the main effect:

$$\frac{(\hat{d}_a - \hat{d}_0)^2}{\hat{D}_a} = \frac{(0,139 - (-0,008))^2}{0,0043} = 5,03$$

$$\frac{(\hat{d}_b - \hat{d}_0)^2}{\hat{D}_b} = \frac{(0,042 - (-0,008))^2}{0,0044} = 0,565$$

$$\frac{(\hat{d}_c - \hat{d}_0)^2}{\hat{D}_c} = \frac{(-0,180 - (-0,008))^2}{0,0038} = 7,79$$

The sum of the values obtained in Step 7 is the Q' value of the main effect:

$$Q' = 5,03 + 0,565 + 7,79 = 13,39$$

As before, the Q' follows the χ^2 distribution with $df = k-1$ (here, $df = 3-1 = 2$). The critical χ^2 value for $P = 0,05$ when $df = 2$ is 5,99: the null hypothesis can be rejected. The main effect of semantic category is significant, suggesting that patient RR's naming performance differs as a function of the object semantic category.

The same procedure also applies to the other main effect, even if other tests exist, such as the z-score for the difference of two proportions. The use of the Q' test for the main effect of the 2-condition factor gains its interest in the fact that the two collapsed proportions are compared to a baseline represented by the overall pooled proportion, rendering the test more conservative.

Multiple comparisons

The presence of a significant interaction or a significant main effect does not really inform us on the reasons of rejection of the null hypothesis. Marascuilo and McSweeney (1967) developed a method of multiple comparisons, consisting in the comparison of two proportions through the confidence interval of their difference. Once again, the problem is that these formulas use the Wald confidence interval of proportions, rendering the results difficult to interpret when $N < 40$, as well as when $\hat{p} = 0$ or $\hat{p} = 1$. Here we propose an alternative procedure which can be used for any data.

As in the Marascuilo and McSweeney (1967) procedure, a critical value should be calculated:

$$critical = \sqrt{\chi^2_{k-1}(1-\alpha)} \quad (8)$$

For $df = 3-1 = 2$, the corresponding χ^2 value is 5,99, and the critical value is $\sqrt{5,99} = 2,45$. Then comes the computing of a value, ψ , for each desired comparison:

$$\psi_{ij} = \frac{\hat{d}_{ij}}{\sqrt{\hat{D}_{ij}}} \quad (9)$$

where \hat{d}_{ij} is the difference between the proportions to compare, i and j , and \hat{D}_{ij} is the variance of their differences, the equation of which was given in (eq.2) and (eq.5). Let's consider that, following a significant interaction we desire to

compare the conditions a_1 and a_2 :

$$\psi_{a_1a_2} = \frac{0,917 - 0,278}{\sqrt{0,0023 + 0,0052}} = 7,38$$

If $\psi_{ij} \geq \text{critical}$, then the difference is significant at $P = 0,05$. In our example the condition a_1 differs from the condition a_2 at $P = 0,05$. This procedure can be used for any other comparison, including the comparisons between the conditions of a single factor when its main effect is significant.

Other tests: two conditions, k conditions, and 2×2 design

The Q' test just described above has the advantage to be flexible and useful with three other designs. Even though other tests exist for the comparison of 2 proportions, K proportions and 2×2 designs, it is of most interest to use the same procedure and derive the same values for these tests. As a point of fact, this would render the scientific studies directly comparable. One has certainly noted that the computation of the main effects in the $2 \times K$ design corresponds to two distinct tests:

The main effect of the 2-condition factor is actually the comparison of two proportions derived from collapsed data. One can just compare two single proportions in the same way. As is the case for the main effect of factor 1, $df = 1$.

The main effect of the K -condition factor is in fact the comparison of K proportions. The computations of this main effect can be used in any study when the difference of K conditions is investigated.

Finally, even though the Q' test was primarily developed to test $2 \times K$ designs, it can also be used at its present form with 2×2 designs. In this case, $df = 1$ for both main effects, as well as for the interaction.

Use and abuse of the Q' tests

The original test by Marascuilo (1970) was designed for the assessment of differences among K *d-primes*. He recommended the use of the test for pooled group data, and in single-case studies. The Q' test, as a test deriving directly from the test of Marascuilo (1970), can thus be used in both cases, but it has not the disadvantages of more familiar tests. This makes the family of Q' tests most useful in single-case and small- N studies. Here are some recommendations for the use of the family of Q' tests, aiming at avoiding any possible abuse:

The analysis should be carried on original scores, never on percentage-transformed data. Even though the computations are carried on proportions, the variance is completely different when using scores than when using percentages because it is dependent on the N . Thus, if you have presented the patient with $N = 40$ trials in your original study, the percentage transformation would increase this

number to $N = 100$. This would allow Type I errors to be more frequent and the plausibility of the results would be strongly questioned.

The Q' tests can be used in a variety of situations and research domains, such as education and marketing, but they are also of great importance in neuropsychological studies involving single cases, especially when it is difficult to obtain any normative data from a control group because of ceiling effects. In this last case, the use of the Q' tests should be justified mostly by the impossibility to collect data from a control group. Whenever the collection of control group data is possible, the use of case-controls tests (Mycroft et al., 2002; Crawford & Garthwaite, 2002) should be preferred to the Q' tests, unless the comparison of at least two cases is desired (see below), or the design involves a 2×2 or $2 \times K$ interaction.

The key notion in cognitive neuropsychology is double dissociation. The existence of two independent cognitive systems or processes can be assumed when a patient exhibits a deficit in task X but not in task Y, and when a second patient exhibits exactly the opposite pattern of performance. Thus, the direct comparison of two cases is necessary for a double dissociation to be assessed. Yet, the single-case adapted tests do not allow the direct comparison of two patients. The Q' tests allow this kind of comparison. Precisely, the $2 \times K$ test allows the comparison of the performance of 2 individuals in K tests or conditions, or the comparison of the performance of K individuals in 2 tests or conditions. These possibilities render necessary the use of the Q' tests at least in neuropsychology.

Implementing the Q' tests

The equations presented in this article can be copied in a spreadsheet of any commonly used software (e.g., Excel) and be kept for further use. A free Excel file allowing the computation of main effects, interaction and multiple comparisons is available on the journal's web site.

Conclusions

The Q' tests family constitute an interesting and useful tool for the analysis of $2 \times K$ designs with proportions, and their use can improve the statistical inference in a variety of research domains and clinical contexts, such as education, marketing, neuropsychology and cognitive psychopathology. The analysis of more complex designs (e.g., $2 \times 2 \times K$) should be possible in the future by extending the Q' tests, and this exciting possibility should allow a more sophisticated, plausible and adequate analysis of data from intra-individual, inter-individual, and group studies.

References

Agresti A., Coull B.A. (1998). *Approximate is better than*

- "exact" for interval estimation of binomial proportions, *Am Stat*, 52, 119-126.
- Brown L.D., Cai T.T., DasGupta A. (2001). *Interval estimation for a binomial proportion*, *Statistical Science*, 16, 101-133.
- Crawford J.R., Garthwaite P.H. (2002). *Investigation of the single case in neuropsychology: confidence limits on the abnormality of test scores and test score differences*, *Neuropsychologia*, 40, 1196-1208.
- Green D., Swets J. (1966). *Signal detection theory and psychophysics*. New York, John Wiley & Sons.
- Gourevitch V., Galanter E.A. (1967). *A significance test for one parameter isosensitivity functions*, *Psychometrika*, 32, 25-33.
- Marascuilo L.A. (1966). *Large sample multiple comparisons*, *Psychological Bulletin*, 65, 280-290.
- Marascuilo L.A. (1970). *Extensions of the significance test for one-parameter signal detection hypotheses*, *Psychometrika*, 35, 237-243.
- Marascuilo L.A., McSweeney M. (1967). *Nonparametric post hoc comparisons for trend*, *Psychological Bulletin*, 67, 401-412.
- Mycroft R.H., Mitchell D.C., Kay J. (2002). *An evaluation of the statistical procedures for comparing an individual's performance with that of a group of controls*, *Cognitive Neuropsychology*, 19, 291-299.
- Newcombe R.G. (1998). *Interval estimation for the difference between independent proportions: comparison of eleven methods*, *Statistics in Medicine*, 17, 873-890.
- Newcombe R., Altman D. (2000). *Proportions and their differences*, In *Statistics with confidence*, 2nd Edition, Altman D., Machin D., Bryant T., Gardner M. (eds.), Bristol, BMJ Books, p.45-56.

Manuscript received August 30th, 2006
 Manuscript accepted December 3rd, 2006