METHODOLOGICAL PAPER

STATISTICAL METHODS FOR SINGLE-CASE STUDIES IN NEUROPSYCHOLOGY: COMPARING THE SLOPE OF A PATIENT'S REGRESSION LINE WITH THOSE OF A CONTROL SAMPLE

John R. Crawford[1] **and Paul H. Garthwaite**[2]

([1]School of Psychology, University of Aberdeen; [2]Department of Statistics, The Open University)

ABSTRACT

Performance on some neuropsychological tests is best expressed as the slope of a regression line. Examples include the quantification of performance on tests designed to assess the accuracy of time estimation or distance estimation. The present paper presents methods for comparing a patient's performance with a control or normative sample when performance is expressed as slope. The methods test if there is a significant difference between a patient's slope and those obtained from controls, yield an estimate of the abnormality of the patient's slope, and provide confidence limits on the level of abnormality. The methods can be used with control samples of any size and will therefore be of particular relevance to single-case researchers. A method for comparing the difference between a patient's scores on two measures with the differences observed in controls is also described (one or both measures can be slopes). The methods require only summary statistics (rather than the raw data from the normative or control sample); it is hoped that this feature will encourage the development of norms for tasks that use slopes to quantify performance. Worked examples of the statistical methods are provided using neuropsychological data and a computer program (for PCs) that implements the methods is described and made available.

Key words: single-case studies; statistical methods; regression

INTRODUCTION

The assessment of neuropsychological deficits in the individual case normally involves comparing a patient's score on a neuropsychological test (e.g., number of items correct) with the distribution of scores obtained from an appropriate control or normative sample. However, for some neuropsychological constructs and their related measurement procedures, it is necessary, or at least preferable, to quantify performance using the slope of a regression line. Evaluation of an individual patient's performance is made by comparing the slope obtained from the patient with those of a normative or control sample. The present paper develops inferential statistical methods for such comparisons.

Examples of Neuropsychological Measures where Performance is expressed as a Slope

An obvious example of the use of slopes to quantify performance is provided by time estimation tasks. In such tasks the accuracy of time estimation is assessed by examining the magnitude of the slope relating the length of the actual time intervals and the estimated time intervals. There is evidence that, particularly when the time intervals exceed 20 seconds (Richards, 1973; Sherwin and Effron, 1980), amnesic patients exhibit deficits in time estimation as indicated by a weakening of the association between actual and estimated time (Nichelli et al., 1993; Venneri et al., 1998). Indeed, in severe cases, the direction of the relationship can be reversed (e.g., the slope of the regression line can be negative), thereby indicating that shorter intervals are estimated as longer.

As another example, in the area of motor control, there is a robust relationship between maximum grip aperture when reaching for objects and the size of the objects (Carey, 1996). An investigator may wish to determine whether this relationship breaks down in an individual with neurological disease. This could be examined by comparing the slopes of the intra-individual regression lines (relating object size and aperture size) obtained from healthy participants with the slope of the regression line obtained from the patient. Relatedly, there is a positive relationship between the peak velocity of manual reaching and the distance from a target; the further away the target, the faster the peak velocity (Jeannerod, 1984). Again, an investigator may wish to examine whether this relationship breaks down in cases with neurological disease; see Carey et al. (1998) for an example.

As a fourth example, in the area of object recognition, it has been found that there is a robust intra-individual relationship between the latency with which an object is named and the degree to which it is rotated away from its prototypical orientation (Turnbull et al., 1997). Failure to find such an association in a patient would suggest that mental rotation was not being used to achieve object identification. This can be assessed by comparing the slope of the regression line relating degree of rotation to latency of recognition for the patient with the slopes obtained for controls

(Turnbull et al., 2002). Distance estimation tasks provide yet another example from the area of visual perception. Here the issue would be whether a patient exhibits an attenuation of the expected relationship between visually estimated distance and actual distance (Carey et al., 1998).

Finally, slopes can be used to quantify the extent to which a patient exhibits a temporal gradient in their recall of material from the past (this material can be either autobiographical or relate to public events). Temporal gradients in recall (such that events from the distant past are better recalled than events that occurred in more recent decades; i.e., the Ribot effect) have been observed in a variety of conditions associated with amnesic problems (Greene et al., 1995; Squire and Alvarez, 1995). For example, suppose a measure of recall of past events covers five decades, then the number of events correctly recalled by a patient could be regressed on the decades (or even the year) in which the events occurred and the resultant slope compared to the slopes obtained from a control sample. It should be noted that a reversal of the Ribot effect has been reported in cases of semantic dementia, i.e., these cases show better recall for the more recent of a series of events (Hodges and Graham, 1998); the use of slopes is just as applicable when attempting to capture this pattern as it is for the pattern more commonly seen in amnesic patients.

Statistical Methods for Comparing an Individual with a Control or Normative Sample

In all of the foregoing examples the slope of the regression lines were treated as *data*. It is not uncommon for slopes to be used in this way. A particularly pertinent example is provided by Venneri et al.'s. (1998) study of time estimation in amnesic and control samples. In this study an ANOVA was used to compare the means of the slopes (relating actual elapsed time to estimated elapsed time) obtained from the amnesic and control participants.

In the foregoing example, inferential statistics were used to test differences between two *samples*. In contrast, both academic neuropsychologists who study single cases, and clinical neuropsychologists, are concerned with comparing an *individual* with a normative or control sample. However, just as the group researcher is concerned with whether group differences are statistically significant, so single-case researchers or clinicians would wish to determine whether the observed difference between their patient and a normative or control sample was statistically significant. More generally, neuropsychologists have an interest in estimating the *abnormality* or *rarity* of a patient's performance; that is, they wish to estimate the proportion of the healthy population that would be more extreme than their patient. The remainder of this paper is concerned with developing statistical methods to address these needs.

The range of potential solutions to these problems is constrained by the fact that the control or normative samples, against which an individual is to be compared, will often be modest in size. Among the reasons for this is the fact that theoretical advances in neuropsychology continue to occur at a rapid rate, whereas the collection of large-scale normative data is a time consuming and often arduous process (Crawford, 1996). Thus, neuropsychologists may continue to have access only to provisional normative data long after a new measure has been developed; these norms may be no more than control sample data from an experimental study. Secondly, when performance on a neuropsychological measure is best expressed as a slope, a specific factor that may have discouraged collection of norms is the lack of explicit, practical guidance on how to analyse and interpret an individual's score when data are in this form. It is hoped that the methods presented here will help remove this obstacle.

The need to develop methods that are suitable for use with small control or normative samples is perhaps most apparent when one considers single-case research. Within academic neuropsychology there has been a resurgence of interest in single-case studies and this has led to significant advances in our understanding of normal and pathological cognitive function (Caramazza and McCloskey, 1988; Code et al., 1996; Humphreys, 1999; McCarthy and Warrington, 1990; Shallice, 1988; Ellis and Young, 1996). In many of these studies the theoretical questions posed cannot be addressed using existing instruments and therefore novel instruments are designed specifically for the study (Shallice, 1979). The sample size of the control or normative group recruited for comparison purposes in such studies is typically < 10 and often < 5. In passing, it may be noted that the control group need not be healthy participants; for example, one can envisage many hypotheses that state that the slope of a particular patient will be significantly lower than those obtained from a sample of patients that have other clinical features in common.

For the reasons outlined above it is clear that statistical methods that treat the control sample statistics (i.e., the mean and SD of the control sample's slopes) as parameters (i.e., treat the normative sample as if it were a population rather than a sample) would have limited applicability. In this respect, the need for an appropriate method of dealing with slopes, is directly analogous to the simpler case where the researcher or clinician wishes to compare a conventional test score for a patient (e.g., number of items passed on a memory test) with a control or normative sample.

The 'standard' procedure for statistical inference in this latter situation is well known. When it is

reasonable to assume that scores are normally distributed, the patient's score is converted to a *z* score, based on the mean and standard deviation in the normative sample, and evaluated using tables of the area under the normal curve (Howell, 1997; Ley, 1972). Thus, if the researcher or clinician has formed a directional hypothesis concerning the patient's score prior to testing (e.g., that the score will be below the mean), then a *z* score which fell below –1.64 would be considered statistically significant (using the conventional 5% level). More generally, the procedure provides the neuropsychologist with information on the rarity or abnormality of the individual's score. This method treats the normative sample statistics as if they were parameters. When the *N* of the normative sample is large this is not problematic. However, it is problematic when, for example, the sample consists of only 10 persons.

Drawing on work by Sokal and Rohlf (1995), Crawford and Howell (1998) presented a method of comparing an individual's score with a normative sample in which the sample statistics are used as sample statistics rather than treated as population parameters. The method is a modified independent samples *t*-test in which the individual's score does not contribute to the estimate of the within-group variance. The formula is

$$t = \frac{X_1 - \bar{X}_2}{\hat{\sigma}_2 \sqrt{\frac{N_2 + 1}{N_2}}}, \qquad (1)$$

where X_1 = the individual's score, $\bar{X}_2$ = the mean score of the normative sample, $\hat{\sigma}_2$ = the standard deviation of the scores in the normative sample, and N_2 = the sample size. The degrees of freedom for *t* are $N_2 + N_1 - 2$ which reduces to $N_2 - 1$. This method can be used to determine if an individual's score is significantly different from that of the normative or control sample. More generally, it provides an *unbiased* estimate of the abnormality of the individual's score; i.e., if the *p* value (one-tailed) for *t* was calculated to be 0.03 then it can be estimated that only 3% of the healthy population would exhibit a score lower than that observed for the individual.

Crawford et al. (1998) extended this approach to cover circumstances where the neuropsychologist wishes to compare the difference between a pair of test scores (e.g., scores on verbal versus spatial short-term memory tasks) observed for an individual, with the distribution of differences observed in a control or normative sample. Crawford and Garthwaite (2002) also extended it to permit comparison of the differences between an individual's scores on each of *k* tests and the individual's mean score on the *k* tests with the differences between these quantities in a control or normative sample.

Most recently, Crawford et al. (2003) have extended the methods to cover situations where performance is quantified by parametric or non-parametric correlation coefficients (e.g., when assessing temporal order memory, an individual's performance is usually assessed by computing the rank order correlation between the reported order and the actual order in which stimuli were presented).

In the present paper we further extend this approach to cover circumstances where the clinician or researcher wishes to compare the slope of a regression line obtained from a patient with a normative or control sample.

Before presenting these methods it should be noted that an additional consideration in developing them was that they should only require summary statistics from the normative or control sample and the patient, rather than the raw data. This was motivated by three considerations. Firstly, the summary statistics required are easily obtained from any standard statistical package. Secondly, working with the summary statistics is less time consuming for the user. Thirdly, by requiring only summary data, this should encourage the development of norms for neuropsychological measures for which performance is best expressed as a slope. That is, publication of the summary statistics from a normative or control sample would be sufficient for independent researchers or clinicians to use the norms with their own patients.

A potential alternative means of testing the difference between the patient's slope and the mean slope in the controls would be simply to convert the patient's slope to *z*, based on the mean and *SD* of the controls, and refer this *z* to a table of the area under the normal curve. However, just as is the case with comparing a patient's performance on a conventionally scored test with controls, this method is inappropriate, as it treats the control sample statistics as if they were population parameters.

The practical effect of using this alternative method would be to exaggerate the abnormality of the patient's performance and to spuriously inflate the chance of finding statistically significant effects. Furthermore, when the performance of patient and controls is quantified using a slope there is an additional consideration, because a regression analysis yields not only a slope estimate, but also the (estimated) variance of the slope estimate. The appropriate test for comparing the slopes of a patient and controls depends upon the homogeneity of these variances for the controls, and whether the patient's slope estimate has a similar variance.

Testing Equality of Slopes

We have *N* individuals in a normative or control sample who each perform an identical set of *k* trials or items on a given neuropsychological

task. Let x denote the independent variable for the jth trial or item and let y denote the ith individual's response for that trial or item. For example, in a single case study aimed at determining whether the relationship between object size and maximum grip aperture has broken down in a patient, x would represent object size and y maximum grip aperture. If six trials were run for each of ten object sizes then k would equal 60. We assume that the responses for each individual follow a simple linear regression,

$$y_{ij} = \alpha_i + \beta_i x_j + \varepsilon_{ij}, \quad (2)$$

where ε_{ij} is random error. We assume that errors for the ith individual are normally distributed with variance τ_i^2 and that $\beta_1, \beta_2, \ldots, \beta_N$ are values from a normal distribution with mean b and variance σ^2, i.e., N (b, σ^2). A further individual (the patient) also performs the set of k tasks and his/her responses are assumed to come from the linear regression

$$y_{(N+1)j} = \alpha_{N+1} + \beta_{N+1} x_j + \varepsilon_{(N+1)j}, \quad (3)$$

where the $\varepsilon_{(N+1)j}$ are normally distributed with variance τ_{N+1}^2. We wish to test whether the slope of the regression line obtained from the patient's data (β_{N+1}) is significantly different from those of the control or normative sample; i.e., we wish to test if β_{N+1} is a value from the distribution N (b, σ^2). More generally, we are interested in estimating the rarity or abnormality of the patient's slope.

Let $\hat{\beta}_i$ denote the estimate of β_i given by the data for the ith individual and let ϕ_i^2 denote the variance of $(\hat{\beta}_i - \beta_i)$. Then

$$\phi_i^2 = \frac{\tau_i^2}{\sum_{j=1}^{k} (x_j - \bar{x})^2} \quad (4)$$

where $\bar{x} = \sum x_j / k$. Also,

$$\hat{\beta}_i \sim \text{N}\ (b,\ \sigma^2 + \phi_i^2). \quad (5)$$

The data consist of the values $\hat{\beta}_1, \ldots, \hat{\beta}_{N+1}$ and we would like to treat these quantities as coming from identical distributions. From (5), this will effectively be the case if either
I) ϕ_i^2 is small relative to σ^2 for $i = 1, \text{K}, N + 1$; or
II) the ϕ_i^2 are equal for $i = 1, \ldots, N + 1$.

Note that the value of $\sum_{j=1}^{k} (x_j - \bar{x})^2$ is the same for all individuals so the ϕ_i^2 are equal if the τ_i are equal.

Before we can move to compare these slopes, we must examine whether it is reasonable to believe that (I) or (II) hold. Testing is simplest if (I) holds, so we suggest it is checked first. If it holds then we proceed directly to a test (Test *c*) in which slopes are compared. If (I) holds for the controls but not for the patient, then we formally test whether the variance for the patient (σ_{N+1}^2) differs from those for the controls ($\sigma_1^2, \ldots, \sigma_N^2$). This is Test *b*. If (I) does not hold for the controls, then we first test if it is reasonable to assume σ_i^2, K, σ_N^2 are equal (Test *a*) and, if it is, we again move to Test *b*. If after performing Test *b* it is reasonable to assume that the σ_i^2 are either small relative to σ^2 [so that (I) holds] or equal [so that (II) holds], then we compare slopes using Test *c*; while if Test *b* suggests the variance for the patient differs from that of the controls, then we compare slopes using a different test (Test *d*). The sequence of tests and decision points are illustrated in the flow diagram presented as Figure 1. The next section presents the rationale for this sequence of tests, their derivations, and the relevant computational formulae. Some technical details of the tests are given in Appendix 1.

The sequence of tests and their computational formula may appear complicated. However, it should be noted that frequently scenario (I) will apply and the procedure is very straightforward. Furthermore, a computer program designed to accompany this paper automates the process (see later section). Thus, although it is important that the basis of the methods are set out formally, in practice the neuropsychologist need never carry out the computations.

Sequence of Tests for Slopes

To examine (I) requires estimates of σ_i^2 ($i = 1, \ldots, N + 1$) and σ^2. The regression analysis that determines the estimate $\hat{\beta}_i$ from the data for the ith person also gives a standard error of this estimate[1], which we will designate as s_i. This statistic is routinely provided by statistics packages that perform regression. An unbiased estimate of σ_i^2 is s_i. To estimate σ^2, we calculate

$$\bar{\beta} = \sum_{i=1}^{N} \hat{\beta}_i / N \quad (6)$$

and

$$\bar{s}^2 = \sum_{i=1}^{N} s_i^2 / N. \quad (7)$$

In words, we sum the slope estimates of each individual in the control sample and divide by the sample size to obtain their mean, and similarly we square each of the standard errors to form variances (s_i^2) and calculate the average of these variances. Then we put

[1]It is crucial that this statistic is not confused with what, in psychological statistics, is referred to as the standard error of estimate. The standard error of estimate is a measure of variability of obervations about the regression line and, in mainstream statistics would be referred to as the residual standard deviation.

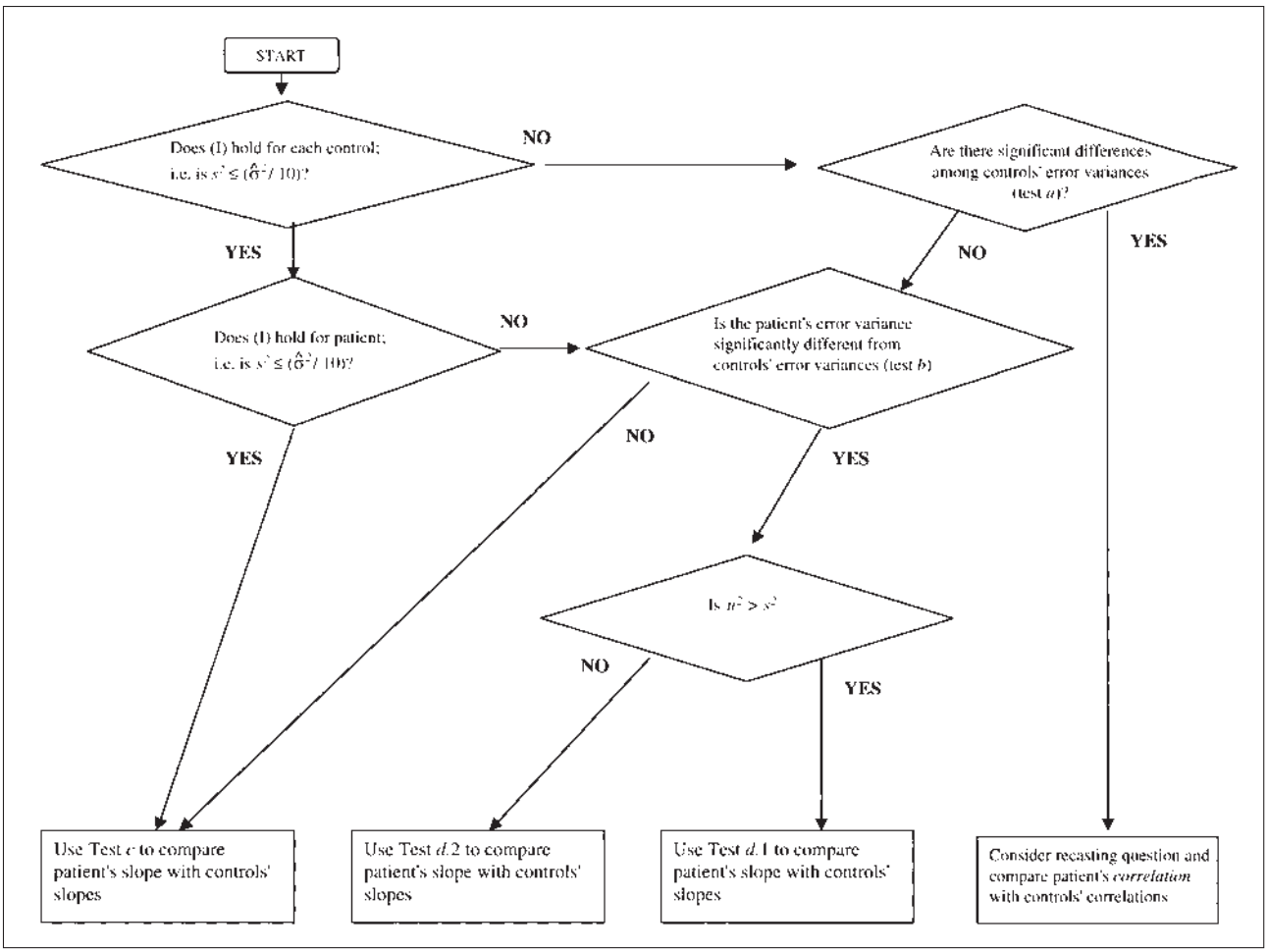

Fig. 1 – *Flow diagram for the sequence of tests used to compare a patient's slope with those of a normative or control sample (the computer program that accompanies this paper fully automates this process).*

$$u^2 = \frac{\sum_{i=1}^{N} (\hat{\beta}_1 - \bar{\beta})^2}{N-1} \tag{8}$$

and an unbiased estimate of σ^2 is

$$\hat{\sigma}^2 = u^2 - \bar{s}^2. \tag{9}$$

In theory, the variance of $\hat{\beta}_i$ about b [i.e. var $(\hat{\beta}_i - b)$] is at least as large as the variance of $\hat{\beta}_i$ about β_i. Sometimes estimates will not reflect this because of random variation, but otherwise the estimate of σ^2 given in (9) will be non-negative.

If the largest of the s_i^2 (among both the control group and the patient) is much smaller than $\hat{\sigma}^2$, say $s_i^2 \leq \hat{\sigma}^2/10$ for $i = 1, \ldots, N + 1$, then it is reasonable to treat the variances of the $\hat{\beta}_i$ as being equal without further testing, and we move on to Test *c*. If $s_i^2 \leq \hat{\sigma}^2/10$ for each of the controls but $s_{N+1}^2 > \hat{\sigma}^2/10$ for the patient, then we go on to Test *b*. Otherwise (i.e. when $s_i^2 > \hat{\sigma}^2/10$ for at least one individual in the control group), we go on to apply Test *a*. We also apply Test *a* if $\hat{\sigma}^2$ is negative.

Test a (Testing for Equal Variances in the Control Sample)

If $s_i^2 > \hat{\sigma}^2/10$ for at least one individual in the control group, then we cannot say that the ϕ_i^2 are sufficiently small for their exact values to be unimportant, and we must first test whether it is reasonable to assume that their value is the same for everyone in the normative or control sample. Thus the null hypothesis for this test (Test *a*) is H_0: $\phi_1^2 = \phi_2^2 = \ldots = \phi_N^2$, and the alternative hypothesis is that at least two of these variances differ.

Various homogeneity of variance tests have been proposed. Most were developed for the case where estimates of variances are derived from simple random samples, rather than from regression models, and they often require kurtosis estimates or the raw data, rather than only requiring estimated variances. However, these restrictions do not apply to the "standard" test for homogeneity of variances first suggested by Bartlett (1937). For this test, put

$$g = 1 + \frac{N+1}{3N(k-2)}, \tag{10}$$

where, as noted, k is the number of trials or items. The test statistic is

$$(k-2)\left\{N\ln\bar{s}^2-\sum_{i=1}^{N}\ln s_i^2\right\}/g. \quad (11)$$

This statistic has approximately a χ^2 distribution on $N-1$ degrees of freedom. H_0 is rejected for large values of the test statistic and if it is rejected at, say the 5% level of significance, then it is unreasonable to treat the variances as being equal. If H_0 is not rejected we proceed to carry out Test *b*. Otherwise, more computationally demanding methods are required, such as Bayesian techniques which make use of the BUGS software package (Spiegelhalter et al., 1996). Alternatively, one can consider expressing the patient and control data as correlation coefficients, rather than slopes, and use the methods recently developed for correlation coefficients by Crawford et al. (2003).

Test b (Comparing the Variance of the Patient with those of the Control Sample)

If it is reasonable to treat the variances of the slope estimates ($\sigma^2 + \phi_i^2$) as being equal for the control or normative sample, the next stage is to compare these variances with that of the patient. There are a number of alternatives available to make this comparison but the method selected here is the standard *F*-test for comparing two variances. The *F*-test has been criticised (Howell, 2002) because it is sensitive to nonnormality of the data (with non-normal data it yields inflated Type I errors). However, there are two reasons for its use in the present context. Firstly, the test can be performed when only summary statistics rather than raw data are available. This is in keeping with our aims of producing a practical method for testing hypotheses, and also it means that researchers or clinicians can compare their results for an individual with normative or control data from a third party. Secondly, in the present context, the test comparing the variances in the individual and control sample is not the primary test of interest (although it *may* have theoretical or clinical implications in its own right, see below). Instead it is used simply to select the most appropriate procedure to employ for the critical test of whether the index of association (i.e., the slope of the regression line) is significantly different from the control values.

Let ϕ^2 denote the common value of $\phi_1^2, \ldots, \phi_n^2$. To perform the *F*-test we form the ratio between the patient's variance (s_{N+1}^2) and the mean variance of the control or normative sample ($\bar{s}^2$), putting the larger of these two quantities in the numerator. That is, we form the ratio $s_{N+1}^2/\bar{s}^2$, or $\bar{s}^2/s_{N+1}^2$. Under the null hypothesis (H_0: $\phi_{N+1}^2 = \phi^2$), this ratio has an F-distribution on $[k-2, N(k-2)]$ degrees of freedom if $s_{N+1}^2 > \bar{s}^2$, or an *F*-distribution on $[N(k-2), k-2]$ degrees of freedom if $\bar{s}^2 > s_{N+1}^2$. Of course, if H_0 is rejected then we have evidence that the case differs from the controls. In some circumstances this in itself would be a theoretically and/or clinically important finding. (So we might choose to perform this test even when $s_1^2 \leq \hat{\sigma}^2/10$ for $i = 1, \ldots, N+1$). If s_{N+1}^2 is significantly greater than $\bar{s}^2$ then the patient's individual responses are further from her/his regression line (i.e., more erratic) than the responses of the controls.

Test c (Comparing Slopes whose Variances are the same for Patient and Controls)

This test is performed if Test *a* was unnecessary, or if Test *b* was performed and its null hypothesis was not rejected. In either case, it is reasonable to treat $\hat{\beta}_1, \ldots, \hat{\beta}_N, \hat{\beta}_{N+1}$ as a simple random sample from a normal distribution. Thus, to test the hypothesis H_0: $\beta_{N+1} \sim N(b, \sigma^2)$ (i.e., to test whether the slope for the patient differs from those of the control or normative sample), we use the result given by Crawford and Howell (1998) for sampling from a normal distribution. The mean ($\bar{\beta}$) and standard deviation (u) of the slopes for the control sample have been calculated using equations (6) and (8). If the patient's regression coefficient ($\hat{\beta}_{N+1}$) is not significantly different from the control sample regression coefficients ($\hat{\beta}_1, \ldots, \hat{\beta}_N$) then

$$\frac{\hat{\beta}_{N+1}-\bar{\beta}}{u\sqrt{\frac{N+1}{N}}} \quad (12)$$

has a *t* distribution on $N-1$ degrees of freedom.

It can readily be appreciated that formula (12) is directly equivalent to formula (1), which was employed by Crawford and Howell (1998). In (1) the difference between a patient's score on a neuropsychological test (e.g., number of items passed) and the mean score of the control or normative sample is divided by the standard error of the difference. In the present case, where performance on the neuropsychological task is represented not by a conventional score but by a slope, the difference between the patient's slope and the mean of the slopes in the control sample is divided by the standard error of that difference.

Test d (Assuming Equal Error Variances for the Control Sample but not for the Patient)

The hypothesis to be tested is that the slope for the patient (β_{N+1}) is not significantly different from the slopes for the control sample ($\beta_1, \ldots, \beta_N$). Two situations must be considered, those where $u^2 > \bar{s}^2$ and those where $u^2 \leq \bar{s}^2$. In theory, the variance of $\hat{\beta}_i$ about b [i.e. var $(\hat{\beta}_i - b)$] is at least as large as the variance of $\hat{\beta}_i$ about β_i. Sometimes estimates will not reflect this because of random variation, but otherwise $u^2 > \bar{s}^2$. When this holds then the following test statistic is appropriate (Test *d*.1)

$$\frac{\hat{\beta}_{N+1}-\bar{\beta}}{\sqrt{u^2\left(\frac{N+1}{N}\right)-\bar{s}^2+s^2_{N+1}}}. \quad (13)$$

Using Satterthwaite's approximation (Satterthwaite, 1946) this statistic (which we designate as *t*') has approximately a *t*-distribution with m_a degrees of freedom, where

$$m_a=\frac{\left[u^2\left(\frac{N+1}{N}\right)-\bar{s}^2+s^2_{N+1}\right]^2}{\left\{\frac{1}{N-1}\left[u^2\left(\frac{N+1}{N}\right)\right]^2+\frac{\bar{s}^4}{N(k-2)}+\frac{s^4_{N+1}}{k-2}\right\}}. \quad (14)$$

If the case where $u^2 \leq \bar{s}^2$ arises, then the Welch-Aspin test (Aspin, 1949) is appropriate. We designate this as Test *d*.2. The test statistic is

$$\frac{\hat{\beta}_{N+1}-\bar{\beta}}{\sqrt{s^2_{N+1}+\frac{\bar{s}^2}{N}}}. \quad (15)$$

This statistic has approximately a *t*-distribution with m_b degrees of freedom, where

$$m_b=\frac{\left[s^2_{N+1}+\frac{\bar{s}^2}{N}\right]^2}{\left[\frac{s^4_{N+1}}{k-2}+\frac{\bar{s}^4}{N^3(k-2)}\right]}. \quad (16)$$

Obtaining a Point Estimate of the Abnormality of a Patient's Slope

The above method is designed to test whether a patient's slope is significantly different from controls. However, it would also be desirable to obtain a point estimate of the percentage of the population that will perform more poorly than the patient (i.e., it would be informative to have an estimate of the abnormality or rarity of the slope observed for the patient). When Test *c* is applied the *p* value obtained not only tells us if the patient's slope is significantly different from controls but it also simultaneously provides an unbiased estimate of this percentage. That is, if the one-tailed *p* value for *t* obtained from Test *c* is 0.023 then this is an estimate of the proportion of the population that would obtain a slope lower than that obtained by the patient.

However, when Test *d*.1 or *d*.2 have to be applied then the *p* value for the significance test is *not* an estimate of the percentage of the population that would perform more poorly. These significance tests factor in the difference between the error variances of the patient and controls but this difference is not relevant when estimating the percentage (it is only the distribution of slopes in the controls that is relevant when estimating this quantity). Fortunately, even when Test *d*.1 or *d*.2 have been used to test for significance, applying Test *c* yields a consistent estimate of the percentage of the population that would obtain a slope lower than that observed for the patient (see second worked example).

Obtaining Confidence Limits on the Abnormality of a Slope

In addition to having a *point* estimate of the percentage of the population that would perform more poorly than the patient, it would also be useful to generate confidence limits on this percentage. This is in keeping with the contemporary emphasis in statistics, psychometrics, and biometrics on the use of confidence limits (American Psychological Association, 2001; Daly et al., 1995; Gardner and Altman, 1989; Zar, 1996). Gardner and Altman (1989) for example, in discussing the general issue of the error associated with sample estimates note that, "these quantities will be imprecise estimates of the values in the overall population, but fortunately the imprecision itself can be estimated and incorporated into the findings" (p. 3).

To generate confidence limits on the abnormality of the slope we use a method given by Crawford and Garthwaite (2002). The method may be applied when it is reasonable to assume that the slope estimates for the controls are identically distributed from a normal distribution. That is, if s_i^2 is much less than $\hat{\sigma}^2$ (say $s_i^2 \leq \hat{\sigma}^2/10$ for i = 1, …, N, or if the null hypothesis of *Test a* is not rejected. Hence, the method may be used with Tests *c*, *d.1* and *d.2*. No assumptions are needed about the distribution of the slope estimate of the patient ($\hat{\beta}_{N+1}$) as only the observed value of $\hat{\beta}_{N+1}$ effects the confidence interval.

Let *P* denote the percentage of the population that will fall below a given individual's estimated slope ($\hat{\beta}_{N+1}$), we suppose we require a 100(1-α)% confidence interval for *P*. If we put

$$c=\frac{\hat{\beta}_{N+1}-\bar{\beta}}{u}, \quad (17)$$

where $\bar{\beta}$ is the mean slope in the controls (equation (7)) and *u* is the standard deviation of the slopes in the controls (i.e., the square root of the variance given by equation (9)), then *c* is an observation from a non-central *t*-distribution on $N - 1$ degrees of freedom. Non-central *t*-distributions have a non-centrality parameter that affects their shape and skewness. We find a value of this parameter, δ_U, such that the resulting non-central *t*-distribution has $c\sqrt{N}$ as its 100 α/2 percentile. Then we find the value δ_L such that the resulting distribution has $c\sqrt{N}$ as its 100(1 – α/2) percentile. From tables for a standard normal distribution we obtain $\Pr(Z < \delta_L/\sqrt{N})$ and $\Pr(Z < \delta_U/\sqrt{N})$. These

TABLE I
Estimated Slopes of the Regression Lines ($\hat{\beta}_i$) and their Standard Errors (S_i) for 10 Control Participants on a Time Estimation Task (the data are artificial)

Control case	$\hat{\beta}_i$	S_i	S_i^2
Control 1	0.884	0.0109	.00012
Control 2	0.866	0.0140	.00020
Control 3	0.924	0.0154	.00023
Control 4	0.966	0.0120	.00014
Control 5	0.804	0.0126	.00016
Control 6	0.724	0.0231	.00054
Control 7	0.914	0.0257	.00066
Control 8	0.687	0.0134	.00018
Control 9	0.820	0.0191	.00037
Control 10	0.821	0.0174	.00030
Control 11	0.979	0.0234	.00055
Control 12	0.713	0.0126	.00016
Control 13	0.770	0.0223	.00050

probabilities depend upon α, c and N and we denote them by $h\ (\alpha/2;\ c;\ N)$ and $h\ (1 - \alpha/2;\ c;\ N)$, respectively. Then a $100(1 - \alpha)\%$ confidence interval for P may be written as

$$[h\ (\alpha/2;\ c;\ N),\ h\ (1 - \alpha/2;\ c;\ N)] \qquad (18)$$

Details of the derivation of h are given in Crawford and Garthwaite (2002) and a worked example of obtaining 95% confidence limits on the rarity of a slope is provided in a later section.

A Worked Example

To illustrate the methods we begin with the example of an amnesic patient's performance on time estimation. Suppose a patient was administered a computerised time estimation task resembling that described by Venneri et al. (1998). Briefly, in one particular version of this task participants are presented with 8 trials of each of 14 time intervals in a randomised order (the intervals ranged from 15 to 54 seconds). Suppose that 13 healthy participants were recruited to serve as a control sample. In Venneri et al.'s. (1998) group study comparing amnesic and healthy samples the performance of each individual was quantified by regressing their actual elapsed times on their estimates of elapsed times; suppose the slope ($\hat{\beta}_{N+1}$) obtained for our patient was 0.60 and its standard error (s_{N+1}) was 0.0143. The slopes and accompanying standard errors for each of the controls are presented in Table I (the data are artificial).

The first step is to calculate $\bar{s}^2$, u^2 and $\hat{\sigma}^2$. From Table I, $\bar{s}^2 = (.00012 + \ldots + .00050)/13 = 0.0003162$, and u^2 (the sample variance of $\hat{\beta}_i$ values) = 0.009176. Hence from equation (9),

$$\hat{\sigma}^2 = .009176 - .0003162 = 0.00886.$$

It can be seen from Table I that s_i^2 is less than $\hat{\sigma}^2/10 = 0.0008886$ for each control, and the patient's error variance ($s_{N+1}^2 = 0.0002$) is also less than this quantity. Therefore, in this example, we can, as will frequently be the case, proceed directly to testing for a difference between the slope of the patient and those of the controls using Test *c*. The mean slope in the controls was 0.8363 with a SD of 0.0958. Entering these figures into formula (12) we obtain the following,

$$t = \frac{0.60 - 0.8363}{0.0958\sqrt{\frac{13+1}{13}}} = \frac{-0.2363}{0.0958\sqrt{1.0769}} =$$

$$= \frac{-0.2363}{0.09942} = -2.377.$$

As the hypothesis tested by the researcher or clinician in this example is directional, i.e., that the patient's performance on time estimation will be significantly lower than matched controls, a one-tailed test is applicable. The one-tailed critical value for t at the 5% level on 12 degrees of freedom is 1.78. The individual's slope is, therefore, significantly different from the controls at the 5% level. The exact one-tailed probability for t in this example is 0.0175 and so the expectation is that only 1.75% of individuals in the population from which the normative sample was drawn would obtain a score as low as that observed for the patient.

To obtain 95% confidence limits on this percentage we proceed as follows:

$$c = \frac{0.60 - 0.8363}{0.0958} = -2.467$$

$$c\sqrt{N} = -2.467\sqrt{13} = -8.895.$$

We want a non-central t-distribution on $N - 1 = 12$ degrees of freedom that has -8.895 as its 0.975 quantile. This determines the non-centrality parameter to be -12.875 so we put $\delta_L = -12.875$. We also want a non-central t-distribution on 12 df that has -8.895 as its 0.025 quantile. This gives $\delta_U = -4.830$. Then,

$$\Pr\left(Z < \frac{-12.875}{\sqrt{13}}\right)\cdot 100 =$$

$$= \Pr\,(Z < -3.57088)\cdot 100 =$$

$$= 0.01778 = 0.002 \text{ (rounded)},$$

TABLE II
Estimated Slopes of the Regression Lines ($\hat{\beta}_i$) and their Standard Errors (S_i) for 8 Control Participants on a Distance Estimation Task (the data are artificial)

Control case	$\hat{\beta}_i$	S_i	S_i^2	$l_n\,(S_i^2)$
Control 1	0.978	0.04271	0.001824	– 6.307
Control 2	0.928	0.04191	0.001756	– 6.345
Control 3	0.950	0.03943	0.001555	– 6.466
Control 4	0.985	0.03882	0.001507	– 6.498
Control 5	0.835	0.04818	0.002321	– 6.066
Control 6	0.940	0.04113	0.001692	– 6.382
Control 7	0.992	0.03035	0.000921	– 6.990
Control 8	0.860	0.04709	0.002218	– 6.111

and

$$\Pr\left(Z < \frac{-4.830}{\sqrt{13}}\right)\cdot 100 =$$
$$= \Pr\,(Z < -1.3396)\cdot 100 = 9.02.$$

Hence the 95% lower confidence limit for *P* is 0.02% and the upper limit is 9.02%. To summarise the results for this case: the patient's time estimation was significantly poorer ($p < 0.05$) than controls and it is estimated that only 1.75% of the population would exhibit a score poorer than that observed; the 95% confidence interval on this percentage is from 0.02% to 9.02%.

An Alternative to the Present Method

As noted in a previous section, a potential alternative means of testing the difference between the slopes of the patient and controls would be simply to convert the patient's slope to *z* and refer this to a table of the areas under the normal curve. It is informative to compare this alternative with the proposed method for the present worked example. The patient's slope expressed as a *z* score from a normal distribution with a mean of 0.8363 and an SD of 0.0958 is – 2.47. Referring this *z* to a table of the normal curve reveals that the estimated percentage of the population that would obtain a slope lower than this is 0.68%. This exaggeration of the abnormality of the patient's performance would be even more pronounced with smaller control samples. Furthermore, in the present example, the conclusion from application of both the *t*-test and *z* is that the patient is significantly impaired ($p < 0.05$). However, obviously these methods need not be in agreement. For example, if the patient's slope had been 0.676, then *z* (– 1.67) would be significant ($p < .05$) but this is a spurious result arising from treating the sample as a population; the *t*-test would not be significant ($t = -1.612$, $p > .05$).

Furthermore, in the present example it was not necessary to be concerned about differences between the error variances of controls and the patient. However, in other cases (such as the worked examples that follow) where the error variances are too large to be ignored and differ between controls and the patient, tests that are more conservative than test *c* (i.e. tests *d*.1 and *d*.2) need to be applied. When this is the case, the use of *z* as an alternative would lead to even greater inflation of the Type I error rate: not only would the statistics of the controls be treated as parameters but the difference in error variances would be ignored.

Further Worked Examples

In the foregoing example it was possible to move directly to a test on the difference between slopes using Test *c*. In the next example we have intentionally chosen data values such that the criteria for applying Test *c* is not fulfilled. The example uses artificial data based on the study by Carey et al., 1998 referred to earlier. One of the issues examined in this study was whether there was a deficit in the visual estimation of distance in a patient (DF) with a posterior cortical lesion. The distance estimation task required participants to estimate distances ranging between 16 and 40cm; there were ten trials for each of the five distances (therefore *k* in this example = 50) under monocular conditions (participants were also tested under binocular conditions but this will be ignored at this point). The accuracy of distance estimation was quantified using the slope relating estimated and actual distance.

Taking this study as the basis for our example, suppose that the distance estimation task had been administered to a patient and eight controls and that, for each participant, estimated distance was regressed against actual distance to obtain their individual slopes and standard errors of the slopes. Let us suppose that the patient's slope was 0.60 and that the standard error of this slope was 0.1043; the equivalent data for the controls are presented in Table II.

As in the first example, we first calculate $\bar{s}^2$, u^2 and $\hat{\sigma}^2$. From the data in Table II, $\bar{s}^2 = .0017243$, and $u^2 = 0.003358$. Hence from equation (9), $\hat{\sigma}^2 = 0.001634$. As s_i^2 is not less than $\hat{\sigma}^2/10 = 0.0001634$ for each control (in fact, Table II shows that s_i^2 approximately equals $\hat{\sigma}^2$ for most controls), it is clear that the ϕ_i^2 are not small relative to σ^2. Hence

TABLE III
Estimated Slopes of the Regression Lines ($\hat{\beta}_i$) and their Standard Errors (S_i) for 6 Control Participants on an Event Dating Task

Control case	$\hat{\beta}_i$	S_i
Control 1	0.492	0.106
Control 2	0.559	0.108
Control 3	0.630	0.116
Control 4	0.627	0.065
Control 5	0.674	0.105
Control 6	0.538	0.107

we must apply Test *a* to determine whether there are significant differences among the error variances of the controls. We calculate *g* from formula (10)

$$g = 1 + \frac{8+1}{3 \times 8\,(50-2)} = 1 + \frac{9}{1152} = 1.0078.$$

Formula (11) requires us to take the log of $\bar{s}^2$ (the mean of the control's error variances), $\bar{s}^2$ = 0.001724 and ln (0.001724) = – 6.3631. We also require the sum of the logs of the control's error variances, the logs of these individual error variances are presented in Table III and their sum is – 51.165. Entering these quantities into formula (11),

χ^2 = (50 – 2) [(8 × 6.3631) – (– 51.165)]/1.0078 =
= (48) [– 50.905 + 51.165]/1.0078 =
= (48) [0.260]/1.0078 = 12.383.

As the result does not exceed the critical value for χ^2 on 7 degrees of freedom (14.07), we treat the error variances among the controls as equal.

As the differences among the controls were not significant, we proceed to test whether the error variance of the patient is significantly different from those of the controls. That is, we perform an *F* test (Test *b*). Because, as noted, the calculations are straightforward when the error variance of the patient does not differ from controls (i.e., Test *c* is applied), we have chosen values for the error variances such that the *F* test *is* significant. This allows us to demonstrate the full sequence of tests and provide a worked example of the more complex formulae (of course, with actual data, the error variances may not differ). In this example, the standard error of the patient's slope estimate (s_{n+1}) = 0.1043 and thus s^2_{n+1} = 0.01088. As this is larger than the mean of the controls' variances ($\bar{s}^2$ = 0.001724), it forms the numerator of the ratio for the *F*-Test; i.e., $s^2_{n+1}/\bar{s}^2$ = 0.01088/0.001724 = 6.311. This ratio exceeds the critical value of 1.39 for *F* on [k – 2, N (k – 2)] = [48, 384] degrees of freedom.

As the patient's error variance differs significantly from the controls, we next compare u^2 (= 0.003357) with $\bar{s}^2$ (= 0.001724). As $u^2 > \bar{s}^2$ we perform Test *d*.1. Substituting in equation (13) gives

$$\frac{0.60 - 0.934}{\sqrt{0.003357\left(\frac{9}{8}\right) - 0.001724 + 0.01088}} =$$

$$= \frac{-0.334}{\sqrt{0.00377625 - 0.001724 + 0.01088}} =$$

$$= \frac{-0.334}{\sqrt{0.01293}} = -2.94.$$

The degrees of freedom for *t*' are calculated using formula (14) (note that the value in the numerator has already been calculated above),

$$m_a = \frac{0.01293^2}{\left[\frac{1}{7}(0.00377625)^2 + \frac{0.001724^2}{8\,(48)} + \frac{0.01088^2}{48}\right]} =$$

$$= 37.06 \simeq 37.$$

The one-tailed critical value for t' at the 5% level on 37 degrees of freedom is 1.687. The individual's slope is, therefore, significantly different from the controls at the 5% level (the precise *p* is 0.0028). As noted, when tests *d*.1 or *d*.2 are used to test whether the patient's slope is significantly different from controls, we also need to run test *c* if we want to obtain an estimate of the percentage of the population that would obtain a lower slope than the patient (i.e., if we want to estimate the abnormality of the slope). Therefore, applying formula (12),

$$t = \frac{0.60 - 0.934}{0.0579\sqrt{\frac{8+1}{8}}} = \frac{-0.334}{0.0579\sqrt{1.125}} =$$

$$= \frac{-0.334}{0.0614} = -5.439.$$

The one-tailed p value for a *t* of 5.439 on N – 1 = 7 df is 0.00048. Multiplying this figure by 100 gives us the estimated percentage of the population that would obtain a score lower than the patient (rounded = 0.05%). To obtain confidence limits on the percentage we would proceed exactly as in the first worked example; i.e., we would use formula (18). The 95% CI on the percentage that this leads to is from the vanishingly small (0.00000005%) to 0.31%.

To summarise the results for this case; the patient's slope is significantly lower (p < .01) than controls. Furthermore, the slope is highly abnormal; it is estimated that only 0.05% of the population would obtain a lower score (upper 95% confidence limit = 0.31%).

The final example uses data obtained from a patient with Parkinson's disease and six age-matched controls on a task requiring participants to date past public events. The raw data were provided by Annalena Venneri (University of Hull). The dating task consisted of 25 events that occurred between 1966 and 1990 (Venneri et al., 1997). Performance for the patient and each of the controls was quantified by regressing the reported

years of occurrence on the actual years the events occurred. The slopes and accompanying standard errors for the controls are presented in Table III; the patient's slope was 0.247 with a standard error of 0.069. To avoid repetition we do not fully work this example (the data in Table III are sufficient for readers to verify the results by hand calculation or by using the accompanying computer program).

As in the previous example, s_i^2 is not less than $\hat{\sigma}^2/10$ for each control and therefore we must apply Test *a* to test for differences among the controls' error variances. This yields a value of 8.13 which does not exceed the critical value for χ^2 on 5 degrees of freedom (11.07), and we therefore treat the error variances among the controls as equal. We next apply test *b*. This yielded a significant result [F (138, 23) = 2.21, p = 0.0148] and therefore the patient's error variance differed from the error variances of controls. In contrast to the previous example, in the present example $u^2 < \bar{s}^2$ (u^2 = 0.004645, $\bar{s}^2$ = 0.010509) and therefore we perform test *d*.2 (i.e., the Welch-Aspin test) rather than *d*.1 to compare the patient's slope with those of the controls,

$$t' = \frac{0.247 - 0.5867}{\sqrt{0.004761 + \frac{0.010509}{6}}} = \frac{-0.3397}{\sqrt{0.0065125}} = 4.209.$$

The degrees of freedom for *t*' from formula (16) are $\simeq$ 42 and the one-tailed probability is 0.00066. Therefore the patient's performance on this task was significantly poorer than controls. As in the previous example we must also apply test *c* to obtain an estimate of the abnormality of the patient's slope (because of the difference between the error variances); the patient's slope is highly unusual, it is estimated that only 0.288% of the control population would obtain a lower slope (95% CI = 0.00000005% to 2.84%).

Detecting Dissociations when Performance on one or both of the Tasks is Expressed as a Slope

Up to this point we have been concerned with methods of testing for a significant deficit on a single task. Although the ability to identify a deficit in the individual case is fundamental, the presence of a deficit in a given cognitive function often only acquires theoretical importance when it is accompanied by the absence of a deficit in other related functions. That is, a central aim in many neuropsychological case studies is to fractionate the cognitive system into its constituent parts. This aim is pursued by attempting to establish the presence of dissociations of function.

Typically, if a patient obtains a score in the impaired range on a test of a particular function and is within the normal range on a test of another function, this is regarded as evidence of a dissociation. However, this evidence in isolation may not be at all convincing (Crawford and Garthwaite, 2002). For example, a patient's score on the "impaired" task could lie just below the cut-point for defining impairment and the performance on the other test lie just above it. Therefore, a more stringent test for the presence of a (classical) dissociation would also involve a comparison of the *difference* between tests observed for the patient with the distribution of differences between these same tests in the control sample (Crawford et al., 2003).

Crawford et al. (1998) devised a method that can be used to test whether the difference between an individual's score on two tasks is significantly different from the differences observed in a control sample. This method can, therefore, provide the additional test for the presence of a dissociation. It can also be employed when a patient's scores are within the impaired range on *both* tasks; i.e. when testing for what Shallice (1988) refers to as a "strong dissociation" (Crawford et al., 2003).

The method was primarily developed for use with tasks in which performance is quantified by conventional means (e.g., number of items correct). For example, Crawford et al. (1998) use the example of testing whether the difference between a patient's performance on a verbal short-term memory task and a spatial short-term memory was significantly larger than the differences in a control sample. However, their method can be just as applicable when performance on one or both of the tasks is expressed as the slope of a regression line with one *proviso*. In order to apply the test the individuals' error variance(s) must either be small relative to the between-subject variance of the slopes, i.e. (I) holds, or, failing that, the error variance(s) of the patient must not differ significantly from those of controls, i.e. (II) holds. In other words, one of the criteria for application of Test *c* must have been fulfilled.

The formula for this test for a dissociation, which is essentially a modified paired samples *t*-test, is

$$t = \frac{Z_X - Z_Y}{\sqrt{(2 - 2r_{xy})\left(\frac{N_2 + 1}{N_2}\right)}}, \tag{19}$$

where Z_X and Z_Y are the scores of an individual on Test X and Test Y expressed as z scores formed using the means and SDs of the normative sample, r_{xy} is the correlation between Tests X and Y in the normative sample and N_2 is the number of participants in the control sample. The test statistic follows a *t*-distribution on $N_2 - 1$ degrees of freedom. Multiplying the one-tailed probability of *t* by 100 gives the point estimate of the abnormality of the individual's score. A derivation for the formula can be found in Appendix 1 of Crawford et al. (1998).

The use of this method is best illustrated with an example. Let us suppose that the amnesic patient and

control cases whose time estimation performance was used in the first worked example had also been administered a test of abstract reasoning.

Suppose that the mean score for the controls on the abstract reasoning task (which was scored as number of items passed) was 34.0 with a SD of 8. Suppose also that the patient's score on the reasoning task was 30. As recorded earlier, the patient's slope on the time estimation task was 0.60. Also as reported earlier, the mean and SD of the slopes in the control sample was 0.8363 with a SD of 0.0958. The only remaining statistic required to test for a dissociation between time estimation and abstract reasoning is the correlation between performance on the two tasks in the control sample; let us suppose this (Pearson) correlation was 0.64.

Using the means and SDs of the controls, the patient's score on the time estimation task expressed as a *z* score is – 2.467 and the *z* score for abstract reasoning is – 0.50. We will designate the abstract reasoning task as Test *X* and the time estimation task as Test *Y* (the choice is arbitrary). Entering these data into formula (19) we obtain

$$t = \frac{|-0.50-(-2.467)|}{\sqrt{(2-1.28)\left(\frac{14}{13}\right)}} = \frac{1.967}{\sqrt{0.72\times 1.0769}} =$$

$$= \frac{1.967}{\sqrt{0.7754}} = 2.234.$$

The two-tailed probability for a *t* of 2.234 on 12 degrees of freedom is 0.0454. We would conclude, therefore, that the patient's performance on the time estimation task differs significantly from his performance on abstract reasoning, the latter being the better; i.e., there is evidence of a dissociation between time estimation and abstract reasoning. By multiplying the one-tailed *p* value (0.027) by 100 we also have an estimate of the percentage of the healthy population that would exhibit a discrepancy in favour of abstract reasoning larger than that observed for the patient (2.27%); i.e., discrepancies of this magnitude and direction are fairly rare. A confidence interval on this percentage can be obtained using a method devised by Crawford and Garthwaite (2002). In the interests of brevity we do not provide a worked example here but the 95% confidence interval for this example is from 0.04% to 10.7%.

The foregoing example involved comparing the difference between a patient's performance on two tasks with the differences in controls when performance on only one of the tasks was expressed as a slope. However, as noted, Crawford et al.'s. (1998) method is just applicable when performance on both tasks is expressed as a slope. This would be useful when there is a need to examine performance under different experimental conditions (i.e. comparison of a patient's slope obtained under condition A versus the slope obtained under condition B).

An obvious specific example is provided by DF, the case that formed the basis for our second worked example. Milner et al. (1991) hypothesised that cases such as DF should be markedly more impaired when distance is estimated using monocular versus binocular vision. Crawford et al.'s method could be used to test this in a patient by comparing the difference between the slopes under monocular versus binocular conditions against the differences between the slopes observed in controls. The indications are that, in the case of DF, application of such a test would reveal a significant difference. Carey et al. (1998) found that the slope relating actual and estimated distance for DF was markedly higher under binocular conditions, whereas the slopes of two controls were similar under monocular and binocular conditions.

As it happens, in the earlier worked example based on DF, the error variances were not small relative to the variance of the slopes and the patient's error variance was significantly different from controls (i.e., Test *c* could not be applied). Therefore, this would provide an example where it would *not* be possible to test for a dissociation. However, as noted, these were artificial data chosen intentionally to have these features (so that the full sequence of tests could be illustrated). In practice it will be common for the conditions to be met.

Caveats on the Use of the Foregoing Methods

The tests developed in the present paper are all modified *t*-tests. One of the assumptions underlying any form of *t*-test is that the data are normally distributed. Monte Carlo simulations have revealed that t-tests are surprisingly robust in the face of moderate violation of this assumption (Boneau, 1960). However, especially given the small *N*s with which we were concerned, these procedures are best avoided when it is known or suspected that the control or normative data are markedly skewed or platykurtic/leptokurtic.

It should be noted that the possible alternative method discussed in a previous section (i.e., treating the sample as a population and using *z* to evaluate a patient's performance) makes the same assumption of normality and is equally compromised by nonnormality. When the control or normative samples are small, the neuropsychologist should also be particularly alert to the presence of outliers. For example, in elderly control or normative samples it is not uncommon to observe occasional cases who perform very poorly despite the absence of any other evidence that suggests the presence of a brain pathology (e.g., early stage dementia).

When testing for differences between slopes, a crucial assumption is that relationships between the *X* and *Y* variables are linear. A visual check of this assumption can easily be made by plotting separate graphs of *Y* against *X* for each individual. If non-

linearity is apparent, the neatest solution, if it works, is to transform the *Y* and/or *X* variables. Transforming the *Y*-variable is more common, often by taking its logarithm, square-root or reciprocal, although transforming the *X*-variable can also be beneficial. Transformation of variables is discussed in most practical textbooks of regression (Weisberg, 1985).

A further assumption underlying the tests is that errors in *Y* are homoscedastic, that is, their variance does not vary with *X*. The most common way of examining this assumption is to plot residual errors against predicted values of *Y*. The plots should show no pattern if the assumptions of linearity and homoscedasticity are satisfied. Transformations are also the most common way of trying to cure homoscedasticity. If the relationship between *Y* and *X* is linear but errors are homoscedastic, then linearity can be preserved by applying the same transformation to both $Y - \hat{\alpha}$ and *X*, where $\hat{\alpha}$ is the constant term in the regression. If a transformation is applied it is of course essential that it is applied to all cases (i.e., to the data of all control cases *and* the patient).

Finally, it will be appreciated that the statistical power of any method of statistical inference will decline as sample size decreases. Thus with the small *N*s with which we are concerned it is inevitable that power will be low. The most obvious way of increasing power is to increase the size of the control or normative sample against which the individual's score is to be compared. Power can also be increased by adopting a more liberal significance level e.g. 15% rather than 5%, but although this more liberal strategy will increase Type I errors (false positives), it will decrease Type II errors (false negatives). The decision to depart from the conventional 5% level should be based on the relative risks the researcher or clinician attaches to the occurrence of these two types of errors. The reasons for departing from the 5% level must be strong, as the 5% level has proved a good choice in general.

Computer program for Evaluating Slopes in the Single Case

The calculations involved in comparing slopes can be tedious and are liable to be error prone. Therefore we have written a computer program (*SINGSLOPE.EXE*) for PCs to accompany this paper[2]. The output from this program can be viewed on the screen, printed, or saved to a file. The program performs the sequence of tests outlined in Figure 1. It prompts for the number of individuals making up the control or normative sample (*N*), the number of trials or items administered (*k*), and the slope and its standard error for the patient and each of the control cases. Standard errors are requested rather than error variances because the former are standard output from all statistical packages that perform regression; for example, the slope (*b*) and the standard error of *b* appear side by side in SPSS output). As previously noted, it is crucial that this latter statistic is not confused with what is commonly termed the standard error of estimate (which measures the variability of observations about the regression line).

The program informs the user of the results of the pre-tests (e.g., test *a* and *b* if it was necessary to apply them) and then lists the mean, standard deviation and standard error of the slopes in the control or normative sample; the *t* value obtained from the appropriate test (e.g., Test *c*, *d*.1 or *d*.2), and its associated one- and two–tailed probability. It also provides the point estimate of the abnormality of the slope and the 95% CI for this percentage.

If a researcher wishes to test for a dissociation between two tasks when performance on one or both of the tasks is expressed as a slope then a program previously provided by Crawford and Garthwaite (2002) can be used; this program implements the method developed by Crawford et al. (1998). However, as noted, this method and the accompanying program, was originally designed for use with conventionally scored tests. Therefore, if it is used to test for a dissociation involving a slope it is necessary that the individuals' error variances are either small relative to the between-subject variance of the slopes or that the patient's error variance is not significantly different from controls. The easiest way to establish whether this is the case is to run *SINGSLOPE.EXE* as it records whether either of these criteria are met. In most circumstances a researcher would want to run this program first in any case; i.e., they would want to test whether a patients slope was significantly different from controls before subsequently seeking evidence for a dissociation.

CONCLUSION

The single case approach in neuropsychology has made a significant contribution to our understanding of the architecture of human cognition (Caramazza and McCloskey, 1988; Code et al., 1996; Humphreys, 1999; McCarthy and Warrington, 1990; Shallice, 1988; Ellis and Young, 1996). However, as Caramazza (1988) notes, if advances in theory are to be sustainable they "… must be based on unimpeachable methodological foundations" (p. 619). The statistical analysis of single case data is an aspect of methodology that has been relatively neglected. This is to be regretted. Other methodological (and logical) considerations may have compelled many researchers to abandon group-based research, but it

[2]A compiled version of this program can be downloaded from the following web site address: www.abdn.ac.uk/~psy086/dept/singslope.htm

is clear that the statistical problems associated with drawing inferences from single cases significantly exceed those of the former approach.

Very useful and elegant methods have been devised for drawing inferences concerning an individual patient's performance on fully standardized neuropsychological tests; i.e., on tests that have been normed on a large representative sample of the population (Capitani, 1997; Willmes, 1985). However, in cognitive neuropsychology, new tests are constantly being devised to measure new theoretical constructs. Understandably, these tests are not fully standardized when employed with single cases; instead they are administered to a control sample that, typically, has a very modest *N*. Therefore, methods that treat the control sample statistics in such studies as population parameters are not appropriate.

Although there remains much to do, we believe that the methods presented here make a useful contribution to the process of developing valid, optimal, and practical statistical methods for single case research. To our knowledge the specific problems addressed in the present paper are not covered in any existing textbooks or papers in neuropsychology or psychological statistics. However, it is clear from the examples provided (which are by no means exhaustive) that the use of slopes to quantify performance has a wide range of applications in neuropsychological research and practice. Finally, we hope that these methods will encourage the development of normative data for tasks in which performance is best expressed as a slope.

Acknowledgements. We are grateful to Dr Sytse Knypstra of the Department of Econometrics, University of Groningen, The Netherlands, for providing an algorithm that finds the non-centrality parameter of a non-central *t* distribution given a quantile, its associated probability, and the degrees of freedom. The algorithm is incorporated into the computer program that accompanies this paper. We are also grateful to Annalena Venneri (Department of Psychology, University of Hull) for providing patient and control data for the worked example on dating of remote events.

References

American Psychological Association. *Publication Manual of the American Psychological Association*. Washington DC: Author, 2001.

Aspin AA. Tables for use in comparisons whose accuracy involves two variances, separately estimated. *Biometrika, 36:* 290-296, 1949.

Bartlett MS. Properties of sufficiency and statistical tests. *Proceedings of the Royal Society, A160:* 268-282, 1937.

Boneau CA. The effect of violation of assumptions underlying the t-test. *Psychological Bulletin, 57:* 49-64, 1960.

Capitani E. Normative data and neuropsychological assessment. Common problems in clinical practice and research. *Neuropsychological Rehabilitation,* 7: 1997.

Caramazza A and McCloskey M. The case for single-patient studies. *Cognitive Neuropsychology, 5:* 517-528, 1988.

Carey DP. Visuomotor sensitivity for shape and orientation in a patient with visual form agnosia. *Neuropsychologia, 34:* 329-337, 1996.

Carey DP, Dijkerman HC and Milner AD. Perception and action in depth. *Consciousness and Cognition, 7:* 438-453, 1998.

Code C, Wallesch C, Joanette Y and Lecours AR (Eds), *Classic Cases in Neuropsychology*. Hove, UK: Psychology Press, 1996.

Crawford JR. *Assessment*. In JG Beaumont, PM Kenealy and MJ Rogers (Eds), *The Blackwell Dictionary of Neuropsychology*. London: Blackwell, 1996, pp. 108-116.

Crawford JR and Garthwaite PH. Investigation of the single case in neuropsychology: Confidence limits on the abnormality of test scores and test score differences. *Neuropsychologia, 40:* 1196-1208, 2002.

Crawford JR, Garthwaite PH and Gray CD. Wanted: Fully operational definitions of dissociations in single-case studies. *Cortex, 39:* 357-370, 2003.

Crawford JR, Garthwaite PH, Howell DC and Venneri A. Intra-individual measures of association in neuropsychology: Inferential methods for comparing a single case with a control or normative sample. *Journal of the International Neuropsychological Society, 9:* 989-1000, 2003.

Crawford JR and Howell DC. Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist, 12:* 482-486, 1998.

Crawford JR, Howell DC and Garthwaite PH. Payne and Jones revisited: Estimating the abnormality of test score differences using a modified paired samples t-test. *Journal of Clinical and Experimental Neuropsychology, 20:* 898-905, 1998.

Daly F, Hand DJ, Jones MC, Lunn AD and McConway KJ. *Elements of Statistics*. Wokingham, England: Addison-Wesley, 1995.

Ellis AW and Young AW. *Human Cognitive Neuropsychology: A Textbook With Readings*. Hove, UK: Psychology Press, 1996.

Gardner MJ and Altman DG. *Statistics With Confidence-Confidence Intervals and Statistical Guidelines*. London: British Medical Journal, 1989.

Glaser RE. *Bartlett's test of homogeneity of variances*. In S Kotz, and NL Johnson (Eds), *Encyclopedia of Statistical Sciences*. New York: Wiley, 1983, pp. 189-191.

Greene JD, Hodges JR and Baddeley AD. Autobiographical memory and executive function in early dementia of Alzheimer type. *Neuropsychologia, 33:* 1647-1670, 1995.

Hodges JR and Graham KS. A reversal of the temporal gradient for famous person knowledge in semantic dementia: Implications for the neural organisation of long term memory. *Neuropsychologia, 36:* 803-825, 1998.

Howell DC. *Statistical Methods for Psychology*. Belmont, CA: Duxbury Press, 1997.

Howell DC. *Statistical Methods for Psychology*. Belmont, CA: Duxbury Press, 2002.

Humphreys GW (Ed). *Case Studies in the Neuropsychology of Vision*. Hove, UK: Psychology Press, 1999.

Jeannerod M. The timing of natural prehension movements. *Journal of Motor Behavior, 16:* 235-254, 1984.

Ley P. *Quantitative Aspects of Psychological Assessment*. London: Duckworth, 1972.

McCarthy RA and Warrington EK. *Cognitive Neuropsychology: A Clinical Introduction*. San Diego, CA: Academic Press, 1990.

Milner AD, Perret DI, Johnston RS, Benson PJ, Jordan TR, Heeley DW, Bettucci D, Mortara F, Mutani R, Terazzi E and Davidson DLW. Perception and action in visual form agnosia. *Brain, 114:* 405-428, 1991.

Nichelli P, Venneri A, Molinari M, Tavani F and Grafman J. Precision and accuracy of subjective time estimation in different memory disorders. *Cognitive Brain Research, 1:* 87-93, 1993.

Richards W. Time production by H.M. *Acta Psychologica, 37:* 279-282, 1973.

Satterthwaite FE. An approximate distribution of estimates of variance components. *Biometrics Bulletin, 2:* 110-114, 1946.

Shallice T. Case study approach in neuropsychological research. *Journal of Clinical Neuropsychology, 3:* 183-211, 1979.

Shallice T. *From Neuropsychology to Mental Structure*. Cambridge, UK: Cambridge University Press, 1988.

Sherwin I and Effron R. Temporal ordering deficits following anterior temporal lobectomy. *Brain and Language, 11:* 195-203, 1980.

Sokal RR and Rohlf JF. *Biometry*. San Francisco, CA: W.H. Freeman, 1995.

Spiegelhalter DJ, Thomas A, Best NG and Gliks WR. *BUGS: Bayesian Inference Using Gibbs Sampling*. Cambridge: Medical Research Council Biostatistics Unit, 1996.

Squire LR and Alvarez P. Retrograde amnesia and memory

consolidation: A neurobiological perspective. *Current Opinion in Neurobiology, 5:* 169-177, 1995.
TURNBULL OH, CAREY DP and MCCARTHY RA. The neuropsychology of object constancy. *Journal of the International Neuropsychological Society, 3:* 288-298, 1997.
TURNBULL OH, DELLA SALA S and BESCHIN N. Agnosia for object orientation: Naming and mental rotation evidence. *Neurocase, 8:* 296-305, 2002.
VENNERI A, NICHELLI P, MODONESI G, MOLINARI MA, RUSSO R and SARDINI C. Impairments in dating and retrieving remote events in patients with early Parkinson's disease. *Journal of Neurology, Neurosurgery and Psychiatry, 62:* 410-413, 1997.
VENNERI A, PESTELL S, GRAY CD, DELLA SALA S and NICHELLI P. Memory, attention and estimation of time. *Brain and Cognition, 37:* 169-172, 1998.
WEISBERG S. *Applied Linear Regression*. New York: Wiley, 1985.
WILLMES K. An approach to analyzing a single subject's scores obtained in a standardized test with application to the Aachen Aphasia Test (AAT). *Journal of Clinical and Experimental Neuropsychology, 7:* 331-352, 1985.
ZAR JH. *Biostatistical Analysis*. London: Prentice-Hall, 1996.

Professor John R. Crawford, School of Psychology, King's College, University of Aberdeen, Aberdeen AB24 3HN, UK. e-mail: j.crawford@abdn.ac.uk

(Received 27 September 2002; reviewed 2 December 2002; revised 20 January 2003; accepted 28 January 2003; Action Editor Carlo Umiltà)

Appendix 1

Notes on Tests

Test a:

The standard errors of the slopes are s_1^2, …, s_{N+1}^2. Formulae (10) and (11) are derived from Glaser (1983) when each s_i^2 follows a χ^2 distribution on $k - 2$ degrees of freedom.

Test b:

From equation (5), for each i

$\hat{\beta}_i \sim \text{N}(b, \sigma^2 + \phi_i^2)$.

Test b assumes the variances $\sigma^2 + \phi_i^2$ can be treated as equal for $i = 1, \ldots, N + 1$, and u^2 is the estimate of their common value.

Test d.1:

The estimated variance of $\hat{\beta}_{N+1} - \bar{\beta}$ is $\hat{\sigma}^2 + s_{N+1}^2 + (\hat{\sigma}^2 + \bar{s}^2)/N$, as the estimated variances of $\hat{\beta}_{N+1}$ and $\bar{\beta}$ are $\hat{\sigma}^2 + s_{N+1}^2$ and $(\hat{\sigma}^2 + \bar{s}^2)/N$, respectively. From equation (9),

$$\hat{\sigma}^2 + s_{N+1}^2 + \frac{(\hat{\sigma}^2 + \bar{s}^2)}{N} = u^2\left(\frac{N+1}{N}\right) - \bar{s}^2 + s_{N+1}^2$$

which yields the denominator in equation (13). This denominator has been expressed in terms of u^2, $\bar{s}^2$ and s_{N+1}^2 because these quantities are independently distributed: s_{N+1}^2 is clearly independent of u^2 and $\bar{s}^2$; u^2 and $\bar{s}^2$ are independent because u^2 is calculated from $\hat{\beta}_i$, …, $\hat{\beta}_N$ while $\bar{s}^2$ is calculated from s_1^2, …, s_N^2.

Test d.2

Assuming $\sigma^2 = 0$, we have that $\hat{\beta}_i \sim \text{N}(b, \Upsilon^2)$ for $i = 1, \ldots, n$. Hence the estimated variances of $\hat{\beta}_{N+1}$ and $\bar{\beta}$ are s_{N+1}^2 and s^2/n, which have $k - 2$ and $n(k - 2)$ degrees of freedom respectively.