

Running Head: RBANS supplementary methods

**Psychological Assessment, in press**

Journal Home Page: <http://www.apa.org/pubs/journals/pas/index.aspx>

Copyright American Psychological Association. This article may not exactly replicate the final version published in the APA journal. It is not the copy of record

Some Supplementary Methods for the Analysis of the RBANS

John R. Crawford<sup>1</sup>, Paul H. Garthwaite<sup>2</sup>, Nicola Morrice<sup>1</sup> and Kevin Duff<sup>3</sup>

School of Psychology  
University of Aberdeen  
Aberdeen, United Kingdom

Department of Mathematics and Statistics  
The Open University  
Milton Keynes, United Kingdom

Department of Neurology  
University of Utah  
Salt Lake City, Utah

---

Address for correspondence: Professor John R. Crawford, School of Psychology, College of Life Sciences and Medicine, King's College, University of Aberdeen, Aberdeen AB24 3HN, United Kingdom. E-mail: [j.crawford@abdn.ac.uk](mailto:j.crawford@abdn.ac.uk)

Acknowledgements: The first author (JRC) undertakes consultancy for Pearson Assessment / The Psychological Corporation (publishers of the RBANS).

### Abstract

Supplementary methods for the analysis of the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) are made available including: (a) quantifying the number of abnormally low Index scores and abnormally large differences exhibited by a case and accompanying this with estimates of the percentages of the normative population expected to exhibit at least this number of low scores and large differences; (b) estimating the overall abnormality of a case's Index score profile using the Mahalanobis Distance Index (MDI); (c) reporting confidence limits on differences between a case's Index scores; and (d) offering the option of applying a sequential Bonferroni correction when testing for reliable differences. With the exception of the MDI, all the methods can be obtained using the formulas and tables provided in this paper. However, for the convenience of clinicians, and to reduce the possibility of clerical error, the methods have also been implemented in a computer program. More importantly the program allows the methods to be applied when only a subset of the Indexes is available. The program can be downloaded from [www.abdn.ac.uk/~psy086/dept/RBANS\\_Supplementary\\_Analysis.htm](http://www.abdn.ac.uk/~psy086/dept/RBANS_Supplementary_Analysis.htm).

**Keywords:** profile analysis; base rates; single-case methods

## Introduction

The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS; Randolph, 1998) is a brief, individually administered test battery used to provide a concise evaluation of cognitive function in adults with neurological disorders. The battery consists of five Indexes: Attention, Language, Visuospatial / Constructional Abilities, Immediate Memory, and Delayed Memory. A Total Scale Index is also provided. The aim of the present paper is to provide supplementary quantitative methods to assist in the interpretation of RBANS Index scores. The following sections set out the rationale underlying each of the methods.

### Estimating the percentage of the normative population that will exhibit $j$ or more abnormally low RBANS Index scores

Information on the rarity or abnormality of test scores is fundamental in interpreting the results of a cognitive assessment (Crawford, 2004; Strauss, Sherman, & Spreen, 2006). When attention is limited to a single measure (an Index score in the present context), this information is immediately available; if an abnormally low score is defined as, say, one that falls below the 5<sup>th</sup> percentile then, by definition, 5% of the normative population is expected to obtain a score that is lower (for example, in the case of RBANS Index scores, scores of 75 or lower are below the 5<sup>th</sup> percentile).

However, if a full RBANS has been administered, there are five Index scores (ignoring the Total Scale Index which is, in essence, the average of the others) and the important question arises as to what percentage of the normative population would be expected to exhibit at least one abnormally low Index score. This percentage will be higher than that for any single Index score considered in isolation, and knowledge of it is liable to guard against over inference; that is, it guards against concluding that

impairment is present on the basis of one “abnormally” low score when such a result is not at all uncommon in the normative population. More generally, having observed the number of abnormally low scores exhibited by a case, it would be useful to know what percentage of the normative population would be expected to obtain at least as many abnormally low scores (Binder, Iverson, & Brooks, 2009; Brooks & Iverson, 2010; Crawford, Garthwaite, & Gault, 2007; Schretlen, Testa, Winicki, Pearlson, & Gordon, 2008) .

One approach to this issue would be to tabulate the percentages of the RBANS standardization sample exhibiting  $j$  or more abnormally low scores; that is, the question could be tackled empirically. However, as yet, this form of base rate data has not been provided for the RBANS. The alternative approach adopted here is to use a Monte Carlo method developed by Crawford, Garthwaite, and Gault (2007) to estimate<sup>1</sup> the required quantities. This method has been used to estimate the percentage of the normative population expected to exhibit  $j$  or more abnormally low Index scores on the WAIS-III and WISC-IV (Crawford et al., 2007) and for short-form versions of both these scales (Crawford, Allum, & Kinion, 2008; Crawford, Anderson, Rankin, & MacDonald, 2010); it has also been applied for similar purposes to other test batteries (Brooks & Iverson, 2010; Crawford, Garthwaite, Sutherland, & Borland, in press; Schretlen et al., 2008).

An important advantage of the Monte Carlo approach over the empirical approach lies in its flexibility: it can be used to generate base rates when only a subset of the five Index scores is available. Whether by choice or through necessity, only a subset of RBANS subtests may have been administered to a patient. For example, a

---

<sup>1</sup> Note that the empirical approach also only provides an estimate because the quantity of interest is the percentage of the normative *population* that will exhibit a given number of abnormally low scores, rather than the percentage among those who happened to make up the normative *sample*.

psychologist may be under time pressure, or they may have a specific hypothesis they want to test which requires only particular Indexes to be administered. Moreover, a patient may be easily fatigued, or may be suffering from physical or sensory disabilities that preclude administration of particular subtests.

The percentage of the normative population expected to exhibit a given number of abnormally low scores will vary markedly with the number of Index scores involved. Moreover, even with a fixed number of Index scores, the percentages will vary as a function of which particular subset of scores was selected (because the percentages are strongly determined by the magnitude of the correlations between scores, and these correlations vary). Thus an accurate estimate of these percentages requires that the base rate data are generated from the *particular* subset of Index scores obtained for the case. It will be appreciated that it is impractical to use the empirical approach to make such data available as voluminous sets of tables would be required (particularly as it would be useful for clinicians to be able to choose between different criteria for an abnormally low score). A subset of the five Index scores could consist of as few as two scores, or as many as four; there are therefore 25 unique combinations.

In the present paper we use Crawford et al's (2007) method to produce base rate tables for the full set of five Index scores. We also implement Crawford et al's method in a computer program that accompanies this paper. Because the program performs the required calculations in real time it is entirely flexible. That is, provision of base rate data is not limited to the case where the full set of five scores are available but rather can be calculated for any particular subset of Index scores.

### Number of abnormal differences between RBANS Indexes

Comparison of an individual's test scores against normative data is a basic part of the assessment process. However, in psychological assessment, such normative comparison standards should be supplemented with the use of *individual* comparison standards when attempting to detect and quantify the extent of any acquired impairments (Crawford, 2004; Lezak, Howieson, Loring, Hannay, & Fischer, 2004). For example, a patient of high premorbid ability may score at or close to the mean of a normative sample but this may still represent a serious decline for the individual concerned. Conversely, a patient may score well below the normative mean but this may be entirely consistent with the individual's premorbid ability. Because of this, emphasis is placed on the use of *individual* comparison standards: Like most tests used in psychological assessment, RBANS Indexes are at least moderately correlated in the general population, thus large discrepancies in a case's test profile suggest an acquired impairment on those tasks that are performed relatively poorly.

The RBANS manual (Table A.2) provides base rate data to allow users to quantify the abnormality of pairwise discrepancies between Indexes. Although these data provide invaluable information, an obvious issue arises: If a case's profile of strengths and weaknesses are examined then, by definition, multiple comparisons are involved. Therefore, it would be useful to know what percentage would exhibit  $j$  or more abnormal pairwise differences overall. This form of base rate data is not currently available for the RBANS but fortunately it is only a little more complicated to estimate the required percentages using Monte Carlo simulation methods than it is to estimate the percentage expected to exhibit a given number of abnormally low scores.

The RBANS manual treats the Total Scale Index like any other Index; that is

each Index can be compared pairwise with every other so that there are 15 possible pairwise comparisons with inclusion of the Total scale (rather than the 10 pairwise comparisons involved when it is excluded). Our view is that, given that the Total scale is essentially the average of the other Indexes, it should be treated differently. That is, comparisons of Index scores should either be conducted pairwise, excluding the Total Scale, *or* each Index should be compared against the case's average Index score. In either of these approaches, given that psychologists have to assimilate a large volume of information from multiple sources when arriving at a formulation, this reduces the comparisons to more manageable proportions. That is, there are either 10 comparisons (pairwise) or five comparisons (comparing against the case's mean Index score). Our own preference is to compare Index scores against the case's mean Index score. This is in keeping with the provision of this form of discrepancy analysis for the WAIS-III (Longman, 2004), WAIS-IV (Crawford, Garthwaite, Longman, & Batty, submitted), and WISC-IV (Flanagan & Kaufman, 2004).

#### A global measure of the abnormality of an individual's Index score profile

As noted, in psychological assessment much emphasis is placed on examining a case's profile of strengths and weaknesses (Crawford, 2004; Lezak et al., 2004; Strauss et al., 2006). The methods outlined in the previous section can help with this process. However, it would also be useful to have a single, continuous, multivariate, index of the *overall* abnormality of an individual's profile of Index scores; that is, an index that quantifies how unusual a particular overall combination of Index scores is.

One such index was proposed by Huba (1985) based on the Mahalanobis distance index (MDI). When the MDI is calculated for an individual's profile it yields a probability value. This value is an estimate of the proportion of the normative

population that will exhibit a more unusual combination of scores. The method has been used to examine the overall abnormality of an individual's profile of subtest scores on the WAIS-R (Burgess, 1991; Crawford, 1994), index score profiles for short-forms of the WAIS-III (Crawford et al., 2008) and WISC-IV (Crawford et al., 2010), and achievement score profiles on the D-KEFS (Crawford et al., in press).

In the present paper the MDI is implemented for RBANS Index scores. It is not a practical proposition to calculate the MDI by hand, nor is it at all practical to provide tabled values for it as there is a huge range of possible combinations of Index scores. A patient's scores can range from 40 to 154 on the Index scores so that there are a myriad of potential profiles (moreover, as noted earlier, only a subset of Index scores may have been obtained for a particular case leading to an additional combinatorial explosion). Therefore the MDI for a case's profile of Index scores is provided only by the computer program that accompanies this paper. Using the computer program, the MDI is calculated in real time, so the method is both fast and flexible. That is, provision of the MDI is not limited to the case where the full set of five scores are available but rather can be calculated for any particular subset of Index scores.

#### Reliability of differences between Index scores

Thus far the discussion of differences between RBANS Indexes has focused on the rarity or *abnormality* of such differences. The RBANS manual however also allows users to examine the *reliability* of differences between Indexes. Tables of critical values are provided that allow the user to determine if the difference between any pair of Indexes is reliable (i.e., unlikely to have arisen from measurement error). This could be supplemented in three ways: (a) an alternative to the use of discrete critical



values is to provide the exact  $p$  value for the difference between Indexes (i.e., the probability that the difference stems from measurement error), (b) the point estimate of the difference between Indexes could be supplemented with confidence intervals for the difference, and (c) the option of applying a Bonferroni correction when testing for reliable differences could be offered to reflect the fact that (excluding comparisons involving the Total Scale) there are ten pairwise comparisons among the five Indexes. With regard to this last point, although psychologists will often have an *a priori* hypothesis concerning a difference between two or more particular Index scores, it is also the case that often there is insufficient prior information to form firm hypotheses. Moreover, should a psychologist wish to attend to a large, unpredicted, difference in a case's profile then, for all intents and purposes, they should be considered to have made all possible comparisons.

### Method

This research was approved by the School of Psychology Ethics Committee, University of Aberdeen, and was conducted in accordance with the Helsinki Declaration.

#### The RBANS

The RBANS consists of 12 subtests that are used to form the five Indexes. The five Indexes measure: Attention (Digit Span and Coding subtests), Language (Picture Naming, Semantic Fluency), Visuospatial / Constructional Abilities (Figure Copy, Line Orientation), Immediate Memory (List Learning, Store Memory) and Delayed Memory (List Recall, List Recognition, Store Memory, Figure Recall).

The RBANS was originally developed as a test for the diagnosis and

characterisation of dementia among the elderly, however its use has since then been extended to a wide range of other clinical populations such as brain injury, depression, and stroke (Strauss et al., 2006). It is quick to administer, taking around 20-30 minutes, and is intended for use with adults. The test is portable, so as to allow for bedside examination, and there are four alternate forms available to allow for repeat testing (Randolph, 1998; Strauss et al., 2006).

The normative sample consists of 540 healthy adults aged between 20 and 89 years who were census matched according to age, sex and ethnicity so as to be representative of the US population (Randolph, 1998; Strauss et al., 2006).

The test has good reliability, with split half reliability for the Total Score Index being .94, and individual Index scores ranging from .82 for the Language Index to .88 for the Immediate Memory Index (Randolph, 1998). Across a one-year retest interval, stability coefficients ranged from .58 to .83 for the Index scores and from .51 to .83 for the subtest scores and practice effects were largely absent in a large cohort of community-dwelling elders (Duff et al., 2005a). Regression-based change formulas have also been developed for the subtests and Indexes of the RBANS to allow clinicians to evaluate reliable changes across time (Duff et al., 2004; Duff et al., 2005b). Studies of the reliability and validity of the RBANS in clinical populations have also been generally encouraging. Although originally designed as a screening tool for dementia, it has been shown to be sensitive to a number of clinical disorders, including Huntington's disease, Parkinson's disease, and HIV dementia (Randolph, 1998), as well as stroke (Larson, Kirschner, Bode, Heinemann, & Goodman, 2005), traumatic brain injury (MacKay, Casey, Wertheimer, & Fichtenberg, 2007), and schizophrenia (Gold, Queern, Iannone, & Buchanan, 1999). However, some have raised caution that the RBANS might be less useful at identifying milder cognitive

impairments (Duff, Hobson, Beglinger, & O'Bryant, 2010). It is useful for differentiating between the various dementing disorders, e.g. those with Alzheimer's disease tend to show poorer performance on language and delayed memory indices, whereas those with Parkinson's disease tend to show very poor performance on the Attention Index. It has also proven to be a generally useful screening tool for assessing the severity of cognitive impairment among clinical populations (Hobart, Goldberg, Bartko, & Gold, 1999; Strauss et al., 2006).

#### Reliability coefficients and correlation matrix for RBANS Index scores

The methods developed in the present paper require only the reliability coefficients for RBANS Index scores and the correlation matrix for the Indexes; these data are presented in the RBANS technical manual (Randolph, 1998); the reliability coefficients are presented in Table 3.6 of the test manual and the correlation matrix in Table 4.1. The reliability data were used to calculate the standard errors of measurement and standard errors of measurement of the difference between Indexes; the correlation matrix was used to obtain the RBANS covariance matrix.

#### Estimating the percentage of the normative population that will exhibit $j$ or more abnormally low RBANS Index scores

As noted, Crawford et al's (2007) Monte Carlo method was used to generate base rate data on the number of abnormally low Index scores. Full technical details of this method are provided in the aforementioned paper and thus are not repeated here. In essence, the method simulates observations (one million in the present application) from the normative population. To do this it requires only the covariance matrix of RBANS Index scores; the covariance matrix is easily obtained from the RBANS

correlation matrix by multiplying all elements by  $(15 \times 15 =) 225$ . For each simulated member of the normative population it records the number of scores classified as abnormally low according to a specified criterion (e.g., below the 5<sup>th</sup> percentile) and reports the estimated percentage of the normative population that will exhibit  $j$  or more abnormally low scores.

In the present study the Monte Carlo simulation was run by drawing observations from a multivariate normal distribution in which each of the marginal distributions had a mean of 100 and standard deviation of 15, and the covariance matrix was set equal to the covariance matrix of RBANS Index scores. Observations drawn from this distribution were then rounded to integers thereby simulating a vector of integer-valued Index scores. For each vector of observations (i.e., for each simulated member of the normative population) the number of scores meeting the specified criterion for an abnormally low score was recorded and used to determine what percentage of the normative population would be expected to exhibit  $j$  or more abnormally low scores.

Psychologists are liable to differ in their preferred definition of an abnormally low score. Therefore the base rate data were generated for a range of criteria for abnormality, ranging from the very liberal criterion of a score below the 25<sup>th</sup> percentile, through to the very stringent criterion of a score below the 1<sup>st</sup> percentile. The present authors' personal preference is to define scores below the 5<sup>th</sup> percentile as abnormally low and this is one of the intermediate criteria offered.

Estimating the percentage of the normative population that will exhibit  $j$  or more abnormally large pairwise discrepancies between RBANS Index scores

The Monte Carlo method outlined in the preceding section was also used to obtain

base rate data on pairwise discrepancies between Indexes. For each simulated member of the normative population, the difference between each pair of Indexes was calculated and divided by the standard deviation of the difference between the relevant pair of Indexes. This yielded a  $z$  score for the difference; if the probability for the absolute value of this  $z$  score (i.e.,  $|z|$ ) exceeded the value corresponding to the specified criterion for an abnormal difference, this was recorded (e.g., if an abnormal pairwise difference was defined as a difference, regardless of sign, exceeded by only 5% of the population, then an abnormal difference was recorded if  $|z|$  was  $> 1.96$ ). These data were then summed to record the percentage of the normative population expected to exhibit a given number of abnormal pairwise differences. For fuller technical details see Crawford et al. (2007).

Estimating the percentage of the normative population that will exhibit  $j$  or more abnormally large deviations from the mean

A similar procedure to that just described for pairwise differences was used to estimate the percentage of the normative population expected to exhibit  $j$  or more abnormally large deviations from their mean Index score. For each simulated member of the normative population the difference between each Index score and the simulant's mean Index score was divided by the standard deviation of these differences. The number of differences classified as abnormal was recorded and summed across stimulants to estimate the percentage of the normative population expected to exhibit a given number of abnormally large deviations. For the formula for the standard deviation of the difference between a component (Index) and the mean of a set of components (Indexes) including the component of interest see Appendix A; for fuller technical details of the procedure see Crawford et al. (2007).

### Calculating the Mahalanobis Distance Index (MDI) for RBANS Index score profiles

The formula for Huba's (1985) MDI of the abnormality of a case's profile of scores on  $k$  tests is

$$(\mathbf{x} - \bar{\mathbf{x}})' \mathbf{W}^{-1} (\mathbf{x} - \bar{\mathbf{x}}), \quad (1)$$

where  $\mathbf{x}$  is the vector of scores for the case on each of the  $k$  tests of a battery,  $\bar{\mathbf{x}}$  is the vector of normative means, and  $\mathbf{W}^{-1}$  is the inverse of the covariance matrix for the battery's standardization sample. When the MDI is calculated for an individual's Index score profile it is evaluated against a chi-square distribution on  $k$  degrees-of-freedom ( $k$  would be 5 if a full RBANS had been administered but will vary between 2 and 5 depending on how many Index scores a case has available). The probability obtained is an estimate of the proportion of the normative population that would exhibit a more unusual combination of Index scores. A case example of the use of the MDI is provided in a later section.

### Reliability of differences between RBANS Index scores

The RBANS manual provides critical values to allow users to test for reliable differences between Indexes. These critical values were obtained by multiplying the standard errors of measurement of the difference ( $SEM_D$ ) between each pair of Indexes by values of  $z$  (e.g., standard normal deviates). An alternative is to divide the difference between a given pair of Indexes by its corresponding  $SEM_D$  to obtain a  $z$  score and convert this quantile to a one- or two-tailed probability. For example, suppose that a case exhibits a difference of 16 points between the Immediate Memory and Language Indexes (IM minus La = -16). Dividing this difference by the  $SEM_D$  for this pairwise comparisons (7.79), yields a  $z$  score of -2.054. Thus the one-tailed  $p$

value is 0.020 and the two-tailed  $p$  value is 0.040. These calculations could easily be done by hand with the use of tables of areas of the normal curve but, for convenience, they are implemented in the computer program that accompanies this paper.

It would also be useful to capture the uncertainty over the difference between a case's Index scores using a (95%) confidence interval. Again, this is easily achieved. The  $SEM_D$  is multiplied by 1.96 and then added (for the upper limit) and subtracted (for the lower limit) from the observed difference. Thus, in the previous example (where the point estimate of the difference was  $-16$ ) the 95% confidence interval is from  $-31$  to  $-1$ .

Exactly the same procedures can be applied to obtain  $p$  values and confidence intervals when the comparisons are between the Index scores and the mean of a case's Index scores, except that the standard errors of measurement of the difference between an index score and the mean index score ( $SEM_M$ ) is used in place of the pairwise  $SEM_D$ . See Appendix B for the formula for the  $SEM_M$ .

As noted previously, when testing for reliable differences between Indexes there are ten pairwise comparisons. Alternatively, if Indexes are compared against the Index score mean, there are five comparisons. One possible solution to these multiple comparison issues would be to apply a standard Bonferroni correction to the  $p$  values obtained when testing for reliable differences. That is, if the family wise (i.e., overall) Type I error rate ( $\alpha$ ) is set at 0.05 then the  $p$  value obtained for an individual pairwise difference between two Indexes would have to be less than  $0.05/10 =$  to be considered significant at the specified value of alpha. This, however, is a conservative approach that will lead to many genuine differences being missed.

A better option is to apply a *sequential* Bonferroni correction (Larzelere & Mulaik, 1977). The first stage of this correction is identical to a standard Bonferroni

correction. Thereafter, any pairwise comparisons that were significant are set aside and the procedure is repeated with  $k - l$  in the denominator rather than  $k$ , where  $l =$  the number of comparisons recorded as significant at any previous stage. The process is stopped when none of the remaining comparisons achieve significance. This method is less conservative than a standard Bonferroni correction but ensures that the overall Type I error rate is maintained at, or below, the specified rate.

This sequential procedure can easily be performed by hand but, for convenience, the computer program that accompanies this paper offers a sequential Bonferroni correction as an option (it can be applied regardless of whether the comparisons are pairwise or against the mean). Note that, when this option is selected, the program does not produce exact  $p$  values but simply records whether the discrepancies between Indexes are significant at the .05 level after correction.

## Results and Discussion

### Estimated percentages of the normative population that will exhibit $j$ or more abnormally low RBANS scores

The results of applying Crawford et al's. Monte Carlo method to estimate the percentage of the normative population exhibiting  $j$  or more abnormally low scores are presented in Table 1. To illustrate, it can be seen from Table 1 that, if an abnormally low score is defined as below the 5<sup>th</sup> percentile, then 17.96% of the normative population are expected to exhibit at least one such abnormally low score. Using the same criterion for abnormality, 5.11% are expected to exhibit two or more such low scores. Thereafter the percentages fall with increasing rapidity; for example, only 1.51% are expected to exhibit three or more abnormally low scores. This criterion is our own preferred criterion for an abnormally low score (hence the



percentages for this criterion appear in bold in the tables and is also the default option in the accompanying computer program). It can be seen that the percentages vary markedly with the choice of criterion. For example, if the most liberal of the criteria is applied (below the 25<sup>th</sup> percentile) then a substantial majority of the normative population (60.53%) are expected to exhibit at least one abnormally score.

As discussed, the percentage of the normative expected to exhibit  $j$  or more abnormally low scores will also vary with the number of scores available (to a lesser extent it will also vary as a function of which particular combination of scores are available). Therefore, if a case has only been administered a subset of the Index scores, the computer program accompanying this paper should be used to obtain the relevant base rate data. (The program also does away with the need to count the number of abnormally low scores exhibited by a case as it applies the user's chosen criterion for abnormality and performs the count).

For purposes of illustration, suppose that only the first three Index scores had been obtained. Suppose also that an abnormally low score has been defined as a score below the 5<sup>th</sup> percentile, and that two of the case's scores meet this criterion. Using the program, it is estimated that 2.07% of the normative population will exhibit this number of abnormally low scores; this compares to 5.11% if all five scores had been used.

Estimated percentages of the normative population that will exhibit  $j$  or more abnormally large Index score differences

As noted, analysis of the abnormality of differences between Index scores can be conducted either using pairwise comparisons or by comparing each Index score to a case's mean Index score (the latter is our preferred option). The results of applying

Crawford et al's. Monte Carlo method to estimate the percentage of the normative population exhibiting  $j$  or more abnormally large pairwise difference are presented in Table 2. To illustrate, it can be seen from Table 2 that, if an abnormally large pairwise difference is defined as a difference exceeded by less than 5% of the normative population our preferred criterion), then 28.15% of the normative population are expected to exhibit at least one such abnormally large difference. Using the same criterion for abnormality, 13.67% are expected to exhibit two or more such differences. Thereafter the percentages fall with increasing rapidity.

The equivalent base rate data on the percentage of the normative population exhibiting abnormally large Index score deviations from their Index score means is presented in Table 3. This table is used in the same fashion as Table 2. For example, if a psychologist has defined an abnormally large difference between each Index and a case's mean Index score to be a difference exceeded by less than 5% of the normative population and finds that a case exhibits two such differences, then, referring to Table 3, it can be seen that it is estimated that only 4.02% of the normative population are expected to exhibit two such differences.

As was the case when considering the issue of abnormally low scores, the percentages expected to exhibit a given number of abnormally large pairwise differences, or abnormally large deviations from the mean Index score, will vary with the criterion used to define abnormality. Thus, if a psychologist had chosen to define an abnormal *pairwise* difference as one that would be exhibited by less than 25% of the normative population (a liberal criterion), then we would expect a very substantial majority of the normative population (77.5%; see Table 2) to exhibit at least one such difference.

Just as was the case for abnormally low scores, the percentage of the

normative population expected to exhibit  $j$  or more abnormally large differences (whether these be pairwise differences or differences from the mean Index score) will vary with the number of Indexes available. Therefore, if a case has only been administered a subset of the Indexes, the computer program accompanying this paper should be used to obtain the relevant base rate data for abnormal differences. (Again, as was the case for abnormally low scores, the program also does away with the need to count the number of abnormally large differences exhibited by a case as it applies the user's chosen criterion for abnormality and performs the count).

#### The MDI for RBANS Index score profiles

The application of the MDI is best illustrated with an example. Suppose that a full RBANS had been administered and that the Index scores obtained (presented in the standard order used in the manual and record form) were: 106, 105, 116, 79, and 75. The chi square value for this profile of scores is 12.187 (on 5 df) and is statistically significant,  $p = 0.032$ . Therefore we can reject the null hypothesis that this profile is an observation from the profiles in the normative population (i.e., it is unusual).

Multiplying this probability by 100 also provides us with the estimated percentage of the normative population that would exhibit an even more unusual profile than the case (3.23%). It can be seen then that the probability value serves both as a significance test and a point estimate of the abnormality of the profile (Crawford et al., 2008).

#### Use of the supplementary methods

Although we consider that all of the methods developed here are useful, they are not interdependent. Therefore it is perfectly possible for a psychologist to pick and

choose among them. That is, a particular psychologist may find the ability to generate base rate data on the number of abnormally low scores particularly useful but have reservations over the use of the Bonferroni correction when testing for reliable differences, whereas another may take the diametrically opposite view.

Although the methods are not interdependent it is worth noting that most of them can be used in a complementary fashion. For example, the estimate of the percentage of the normative population that will exhibit at least as many low scores as a case will potentially identify consistently poor performance. In contrast, the MDI is relatively insensitive to the *absolute* level of performance on each of the Index scores but is sensitive to the overall *profile* of performance. These contrasting features are best illustrated with a concrete example. Suppose that a case has been administered all five Indexes and obtains a score of 73 on all of these<sup>2</sup>. This is a very poor level of performance: from Table 1 it is estimated that only 0.06% of the normative population will obtain scores below the 5<sup>th</sup> percentile on all five Indexes. Although poor, the case's performance is remarkably consistent. For this example the chi square for the MDI is not significant ( $\chi^2 = 6.405$  on 5 df,  $p$  value = 0.269) underlining that the MDI is not sensitive to a case's absolute levels of performance.

In contrast, suppose that everything was the same as in the first example but that, on the first three of the Indexes, the case obtained scores of 115. In this scenario the MDI is highly significant:  $\chi^2 = 20.51$  on 5 df,  $p$  value = 0.001. The *profile* of scores is therefore *highly* unusual; very few individuals (0.1%) in the normative population would be expected to exhibit a more unusual profile of scores. In this

---

<sup>2</sup> We use this example and the example that follows in the interests of simplicity to represent generic examples of very consistent and very inconsistent performance; a case is unlikely to obtain exactly the same scores across all five indexes, for example in the 20-39 age group it is not possible to obtain an Attention Index score of 73 (the Attention Index scores jump from 72 to 75 in one step in this age group).

latter example two of the case's scores are abnormally low; it is estimated that 5.11% of the normative population will exhibit this number of low scores (see Table 1). It can be seen then that the base rate data on low scores and the MDI are complementary in the process of identifying cognitive difficulties. Needless to say, if both methods converge to suggest either abnormal or normal performance, then interpretation of the results is simplified and the clinician can have more confidence when arriving at a formulation.

The MDI and the base rate data on the difference between Indexes may also play a useful role in determining how much weight should be afforded to the Total Scale Index. For example, if the MDI or base rate data suggest that a combination of Index scores is highly unusual (i.e., there are sizeable discrepancies within the profile) then the Total Scale Index is clearly less useful as a summary of a case's abilities.

#### Computer program for supplementary analysis of RBANS Index scores

As referred to above, a compiled computer program for PCs (written in the Delphi programming language), RBANS\_Supplementary\_Analysis\_EXE, accompanies this paper. With the exception of the MDI, all the methods presented here can be applied either using the tables provided or by relatively simple calculations on the part of the user. However, the program provides a very convenient alternative for busy psychologists as, on being provided with a case's Index scores, it performs all of the necessary calculations and records the results. The computer program has the additional advantage that it will markedly reduce the likelihood of clerical error. Research shows that, when working with test scores, psychologists make many more simple clerical errors than we like to imagine (e.g., see Faust, 1998; Sherrets, Gard, & Langner, 1979; Sullivan, 2000).

To use the program the user need only select their preferred analysis options, and select their preferred criterion for a low score / large difference; this is done using radio buttons. Thereafter the case's Index scores can be entered. If only a subset of Index scores has been administered the data fields for the omitted Indexes need simply be left blank. There is also the option of adding user's notes (e.g., a Case ID, date of testing etc.) for future reference. A screen capture of the input form for the program is presented as Figure 1a.

The output of the program reproduces the case's Index scores and accompanies them with their confidence limits and their percentile ranks. The number of a case's scores that meet the user-selected criterion for a low score is recorded along with the percentage of the normative population expected to exhibit at least this number of low scores.

Next, the reliability of differences are recorded, either the reliability of pairwise differences, or the reliability of differences between the Indexes and the case's mean Index score, determined by the option selected by the user (the latter is the default option). The results consist of the differences, 95% confidence intervals on the differences, and the one- and two-tailed probability values for the differences. The abnormality of the differences is then presented: the results consist of the estimated percentage of the normative population that would exhibit a difference of the magnitude observed for the case (in the same direction as the case, and also regardless of the sign of the difference). The number of a case's differences that meet the selected criterion for abnormality is also recorded and this is accompanied by the estimated percentage of the normative population that would exhibit that number or more of such differences.

Finally, the results of applying the MDI to the case's score profile are

reported. These results consist of the chi square value and its associated probability; this probability is multiplied by 100 to provide the estimated percentage of the normative population that would exhibit a more unusual overall profile than the case. A screen capture of the results form, showing a portion of the output (the abnormality of pairwise differences between Indexes and the results of the MDI) is presented as Figure 1b.

The results from the program can be viewed on screen, saved to a file, or printed. Because the program performs a Monte Carlo simulation to obtain the multiple base rates there will typically be a delay of around 10 seconds before the results are available. The program can be downloaded, either as a raw executable or as a zip file, from the following web page:

[www.abdn.ac.uk/~psy086/dept/RBANS\\_Supplementary\\_Analysis.htm](http://www.abdn.ac.uk/~psy086/dept/RBANS_Supplementary_Analysis.htm).

### Conclusion

The aim of the present paper was to develop a package of supplementary quantitative methods to assist with interpretation of the RBANS. Although some of the underlying calculations required to implement these methods are complex, this is not an impediment to their adoption as the tabled values and, particularly, the accompanying computer program, makes this process both quick and reliable.

The provision of additional quantitative methods of analysis by no means undermines the role of the psychologist in decision making, rather it should be viewed as an aid to such decision making. The psychologist still needs to employ the uniquely human ability of combining quantitative results with the qualitative data obtained from interview and testing in order to arrive at a satisfactory formulation of a case's cognitive strengths and weaknesses and thereafter develop its implications for

management and/or intervention. Thus, although the focus of the present paper is firmly quantitative, it should not be taken as a plea for an actuarial / mechanistic approach to assessment.



## References

- Binder, L. M., Iverson, G. L., & Brooks, B. L. (2009). To err is human: “Abnormal” neuropsychological scores and variability are common in healthy adults. *Archives of Clinical Neuropsychology, 24*, 31-46.
- Brooks, B. L., & Iverson, G. L. (2010). Comparing actual to estimated base rates of ‘abnormal’ scores on neuropsychological test batteries: Implications for interpretation. *Archives of Clinical Neuropsychology, 25*, 14-21.
- Burgess, A. (1991). Profile analysis of the Wechsler Intelligence Scales: A new index of subtest scatter. *British Journal of Clinical Psychology, 30*, 257-263.
- Crawford, J. R. (2004). Psychometric foundations of neuropsychological assessment. In L. H. Goldstein & J. E. McNeil (Eds.), *Clinical neuropsychology: A practical guide to assessment and management for clinicians* (pp. 121-140). Chichester: Wiley.
- Crawford, J. R., & Allan, K. M. (1994). The Mahalanobis distance index of WAIS-R subtest scatter: Psychometric properties in a healthy UK sample. *British Journal of Clinical Psychology, 33*, 65-69.
- Crawford, J. R., Allum, S., & Kinion, J. E. (2008). An Index based short form of the WAIS-III with accompanying analysis of reliability and abnormality of differences. *British Journal of Clinical Psychology, 47*, 215-237.
- Crawford, J. R., Anderson, V., Rankin, P., & MacDonald, J. (2010). An index-based short-form of the WISC-IV with accompanying analysis of the reliability and abnormality of differences. *British Journal of Clinical Psychology, 49*, 235-258.
- Crawford, J. R., Garthwaite, P. H., & Gault, C. B. (2007). Estimating the percentage of the population with abnormally low scores (or abnormally large score

differences) on standardized neuropsychological test batteries: A generic method with applications. *Neuropsychology*, *21*, 419-430.

Crawford, J. R., Garthwaite, P. H., Longman, R. S., & Batty, A. (submitted). Some supplementary methods for the analysis of WAIS-IV index scores.

Crawford, J. R., Garthwaite, P. H., Sutherland, D., & Borland, N. (in press). Some supplementary methods for the analysis of the D-KEFS. *Psychological Assessment*.

Duff, K., Beglinger, L. J., Schoenberg, M. R., Patton, D. E., Mold, J., Scott, J. G., et al. (2005a). Test-retest stability and practice effects of the RBANS in a community dwelling elderly sample. *Journal of Clinical and Experimental Neuropsychology*, *27*, 565-575.

Duff, K., Hobson, V. L., Beglinger, L. J., & O'Bryant, S. E. (2010). Diagnostic Accuracy of the RBANS in Mild Cognitive Impairment: Limitations on Assessing Milder Impairments. *Archives of Clinical Neuropsychology*, *25*, 429-441.

Duff, K., Schoenberg, M. R., Patton, D., Mold, J., Scott, J. G., & Adams, R. L. (2004). Predicting change with the RBANS in a community dwelling elderly sample. *Journal of the International Neuropsychological Society*, *10*, 828-834.

Duff, K., Schoenberg, M. R., Patton, D., Paulsen, J. S., Bayless, J. D., Mold, J., et al. (2005b). Regression-based formulas for predicting change in RBANS subtests with older adults. *Archives of Clinical Neuropsychology*, *20*, 281-290.

Faust, D. (1998). Forensic assessment. In *Comprehensive Clinical Psychology Volume 4: Assessment* (pp. 563-599). Amsterdam: Elsevier.

Flanagan, D. P., & Kaufman, A. S. (2004). *Essentials of WISC-IV assessment*. New Jersey: Wiley.

- Gold, J. M., Queern, C., Iannone, V. N., & Buchanan, R. W. (1999). Repeatable battery for the assessment of neuropsychological status as a screening test in schizophrenia I: sensitivity, reliability, and validity. *American Journal of Psychiatry, 156*, 1944-1950
- Hobart, M. P., Goldberg, R., Bartko, J. J., & Gold, J. M. (1999). Repeatable Battery for the Assessment of Neuropsychological Status as a Screening Test in Schizophrenia, II: Convergent/discriminant validity and diagnostic group comparisons. *American Journal of Psychiatry, 156*, 1951-1957.
- Huba, G. J. (1985). How unusual is a profile of test scores? *Journal of Psychoeducational Assessment, 4*, 321-325.
- Larson, E. B., Kirschner, K., Bode, R., Heinemann, A., & Goodman, R. (2005). Construct and predictive validity of the Repeatable Battery for the Assessment of Neuropsychological Status in the evaluation of stroke patients. *Journal of Clinical and Experimental Neuropsychology, 27*, 16-32.
- Larzelere, R. E., & Mulaik, S. A. (1977). Single-sample tests for many correlations. *Psychological Bulletin, 84*, 557-569.
- Lezak, M. D., Howieson, D. B., Loring, D. W., Hannay, H. J., & Fischer, J. S. (2004). *Neuropsychological Assessment* (4th ed.). New York: Oxford University Press.
- Longman, R. S. (2004). Values for comparison of WAIS-III index scores with overall means. *Psychological Assessment, 16*, 323-325.
- MacKay, C., Casey, J. E., Wertheimer, J., & Fichtenberg, N. L. (2007). Reliability and validity of the RBANS in a traumatic brain injured sample. *Archives of Clinical Neuropsychology, 22*, 91-98.
- Randolph, C. (1998). *RBANS manual*. San Antonio TX: The Psychological

Corporation.

Schretlen, D. J., Testa, S. M., Winicki, J. M., Pearlson, G. D., & Gordon, B. (2008).

Frequency and bases of abnormal performance by healthy adults on neuropsychological testing. *Journal of the International Neuropsychological Society, 14*, 436–445.

Sherrets, F., Gard, G., & Langner, H. (1979). Frequency of clerical errors on WISC protocols. *Psychology in the Schools, 16*, 495-496.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms and commentary* (3rd ed.).

New York: Oxford University Press.

Sullivan, K. (2000). Examiners' errors on the Wechsler Memory Scale-Revised.

*Psychological Reports, 87*, 234-240.

Table 1. Percentage of the normative population expected to exhibit at least  $j$  abnormally low Index scores on the RBANS; increasingly stringent definitions of abnormality are used ranging from below the 25<sup>th</sup> percentile to below the 1<sup>st</sup> percentile.

Criterion	Percentage exhibiting $j$ or more abnormally low RBANS Index scores				
	1	2	3	4	5
<25 <sup>th</sup>	60.53	35.10	18.64	8.29	2.41
<15.9 <sup>th</sup>	44.68	21.05	9.34	3.47	0.83
<10 <sup>th</sup>	31.59	12.14	4.57	1.43	0.29
<5 <sup>th</sup>	<b>17.96</b>	<b>5.11</b>	<b>1.51</b>	<b>0.38</b>	<b>0.06</b>
<2 <sup>nd</sup>	8.05	1.57	0.35	0.06	0.00
<1 <sup>st</sup>	4.24	0.64	0.12	0.02	0.00

Note. The above figures assume that all five RBANS Indexes scores were obtained; when only a subset of the Indexes is available for a case the computer program accompanying this paper records the percentage of the population expected to exhibit at least as many abnormally low scores as the case

Table 2. Percentage of the normative population expected to exhibit at least  $j$  abnormally large pairwise differences between RBANS Indexes (regardless of sign); increasingly stringent definitions of abnormality are used ranging from a difference exhibited by less than 25% of the population to a difference exhibited by less than 1%.

Criterion	Percentage exhibiting $j$ or more abnormal pairwise differences (regardless of sign) between RBANS Indexes									
	1	2	3	4	5	6	7	8	9	10
<25%	77.50	64.16	48.78	34.11	15.81	7.67	1.49	0.24	0.01	0.00
<15%	61.64	44.40	28.47	16.27	5.47	2.09	0.20	0.02	0.00	0.00
<10%	46.30	28.65	15.40	7.17	1.79	0.57	0.03	0.00	0.00	0.00
<b>&lt;5%</b>	<b>28.15</b>	<b>13.67</b>	<b>5.66</b>	<b>2.01</b>	<b>0.34</b>	<b>0.09</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
<2%	13.42	4.74	1.44	0.36	0.04	0.01	0.00	0.00	0.00	0.00
<1%	7.38	2.06	0.51	0.10	0.01	0.00	0.00	0.00	0.00	0.00

Note. The above figures assume that all five RBANS Indexes scores were obtained; when only a subset of the Indexes is available for a case the computer program accompanying this paper records the percentage of the population expected to exhibit at least as many abnormally large pairwise differences as the case.

Table 3. Percentage of the normative population expected to exhibit  $j$  or more abnormal RBANS Index scores *relative to individuals' mean Index scores* (regardless of sign); increasingly stringent definitions of abnormality are used ranging from a difference exhibited by less than 25% of the population to a difference exhibited by less than 1%.

Criterion	Percentage exhibiting $j$ or more abnormal deviation scores (regardless of sign) on the RBANS				
	1	2	3	4	
<25%	70.35	40.47	11.16	2.79	0.14
<15%	52.53	22.09	3.91	0.71	0.02
<10%	37.03	11.37	1.34	0.18	0.00
<b>&lt;5%</b>	<b>20.64</b>	<b>4.02</b>	<b>0.26</b>	<b>0.02</b>	<b>0.00</b>
<2%	8.98	0.98	0.03	0.00	0.00
<1%	4.67	0.33	0.01	0.00	0.00

Note. The above figures assume that all five RBANS Indexes scores were obtained; when only a subset of the Indexes is available for a case the computer program accompanying this paper records the percentage of the population expected to exhibit at least as many abnormally large deviations as the case

## Appendix A

Calculation of the standard deviation of the difference between an Index score and the mean of Index scores

The formula for the standard deviation of the difference between an Index score and the mean of the Index scores is

$$SD_{M_a} = s\sqrt{1 + \bar{\mathbf{G}} - 2\bar{\mathbf{h}}_a}, \quad (2)$$

where  $s$  is the common standard deviation of the tests (15 in the present case),  $\bar{\mathbf{G}}$  is the mean of all elements in the full correlation matrix for the  $k$  tests contributing to the mean score, and  $\bar{\mathbf{h}}_a$  is the mean of the row (or equivalently the column) of correlations between test  $a$  and the other  $k$  tests (including test  $a$  itself; that is, the unity in the main diagonal is included in this row mean). This formula is applied separately for each Index to obtain the standard deviations of the difference between each Index and the mean Index score.

## Appendix B.

Calculation of the standard error of measurement of the difference between an Index score and the mean of Index scores

To test whether an Index score is reliably different from an individual's mean Index score requires calculation of the standard error of measurement for such a difference (here denoted as  $SEM_{M_i}$ ). The formula is

$$SEM_{M_i} = \sqrt{\left(\frac{k-2}{k}\right)s_i^2 + \frac{1}{k^2} \sum s_j^2}, \quad (3)$$

where  $k$  is the number of tests contributing to the mean,  $s_i^2$  is the square of the standard error of measurement (i.e., it is the variance of the errors of measurement)



for the index score of interest and the summation signs tells us to sum the squared standard errors of measurement ( $s_j^2$ ) for all  $k$  indexes, including the index of interest.

In the present case the required standard errors of measurement were obtained from the reliability coefficients reported in the RBANS manual using the standard formula.

Figure Legends

Figure 1. Screen capture of the computer program that accompanies this paper showing (a) the input form, and (b) the results form

(a)

RBANS\_Supplementary\_Analysis.EXE: Supplementary Analysis of RBANS Index Scores

This program accompanies the paper by Crawford, JR, Garthwaite, PH, Morrice, N, & Duff, K. Some supplementary methods for the analysis of RBANS Index scores. The program reproduces the RBANS Index scores entered by the user and accompanies these with confidence intervals (and the score's percentile rank). The program also records the number of the case's Index scores that are classified as abnormally low according to a criterion selected by the user. It then provides base rate data on the percentage of the normative population expected to

User's Notes: Example setup for program

Base analysis of differences on...

- PAIRWISE comparisons of Indexes
- DEVIATIONS from an individual's mean Index score

When testing for reliability of differences ...

- Do NOT apply sequential Bonferroni correction
- DO apply sequential Bonferroni correction

Criterion for abnormally low score / large difference...

- 25% of normative population
- 15.87% of normative population (corresponds to one SD)
- 15% of normative population
- 10% of normative population
- 5% of normative population
- 2% of normative population
- 1% of normative population

Immediate Memory Index score: 73

Visuospatial/Constructional Index score: 112

Language Index score: 117

Attention Index score: 68

Delayed Memory Index score: 60

Total Scale Index score:

Compute Clear Data Exit

(b)

Results viewer: RBANS\_Supplementary\_Analysis.EXE: Supplementary Analysis of RBANS Index Scores

Printer options...

ABNORMALITY of pairwise differences between Index scores, i.e., percentage of population estimated to exhibit a larger difference in the same direction; the final column records the percentage regardless of sign (the count of the case's number of abnormal differences is based on this latter column):

Index Pair	Difference	Percentage of population with larger differences	Percentage with larger differences regardless of sign
IM minus VC	-39	1.513	3.026
IM minus La	-44	0.238	0.476
IM minus At	5	38.045	76.091
IM minus DM	13	15.685	31.370
VC minus La	-5	38.750	77.501
VC minus At	44	0.595	1.189
VC minus DM	52	0.118	0.236
La minus At	49	0.143	0.286
La minus DM	57	0.023	0.047
At minus DM	8	31.868	63.735

NUMBER of case's pairwise differences that meet criterion for abnormality = 6  
 PERCENTAGE of normal population expected to exhibit this number or more of abnormal differences = 0.09%

MAHALANOBIS DISTANCE Index of the overall abnormality of the case's Index score profile:  
 Chi-square = 20.758 on 5 df, p value = 0.00090  
 Percentage of normative population expected to exhibit a more unusual profile = 0.090%

Save Output Clear Results Return to Worksheet Exit