

Running head: WAIS-IV Supplementary Methods

Journal of Neuropsychology, in press

Some supplementary methods for the analysis of WAIS-IV Index scores in
neuropsychological assessment

John R. Crawford

School of Psychology

University of Aberdeen, UK

Paul H. Garthwaite

Department of Mathematics and Statistics

The Open University, UK

R. Stewart Longman

Department of Psychology

Foothills Hospital, Canada

Abigail M. Batty

Pearson Assessment

London, UK

Some supplementary methods for the analysis of WAIS-IV Index scores in
neuropsychological assessment

Abstract

Objectives: To develop supplementary methods for the analysis of the WAIS-IV in neuropsychological assessment

Design and Methods: Psychometric

Results: The following methods are made available: (a) provision of traditional confidence intervals on Index scores, (b) expression of the endpoints of confidence intervals as percentile ranks; (c) quantification of the number of abnormally low Index scores exhibited by a case and accompanying estimate of the percentage of the normative population expected to exhibit at least this number of low scores; (d) quantification of the reliability and abnormality of Index score deviations from an individual's Index score mean (thereby offering an alternative to the pairwise approach to Index score comparisons available in the WAIS-IV manual); (e) provision of confidence intervals on an individual's deviation scores or pairwise difference scores, (f) estimation of the percentage of the normative population expected to exhibit at least as many abnormal deviations or abnormal pairwise differences as a case; and (g) calculation of a case's Mahalanobis Distance Index (MDI), thereby providing a multivariate estimate of the overall abnormality of an Index score profile. With the exception of the MDI, all the methods can be applied using tables provided in this paper. However, for ease and speed of application, and to reduce the possibility of clerical error, all the methods have also been implemented in a computer program

Conclusions: The methods are useful for neuropsychological interpretation of the WAIS-IV

Keywords: neuropsychological assessment; abnormality & reliability of differences; base rates; profile analysis

Introduction

Lezak (1988) has described the Wechsler Intelligence Scale as the “the workhorse of neuropsychological assessment” and noted that it is the single most utilized component of the neuropsychological repertory" (p. 53). The latest edition of the scale is the Wechsler Adult Intelligence Scale -Fourth Edition (WAIS-IV; Wechsler, 2008a). The main strengths of the scale lie in its large and representative standardization sample, and its excellent psychometric properties (for example, the Full Scale IQ has a reliability of 0.98; that is, 98% of the variance in FSIQ is true variance, and only 2% is measurement error variance). The scale also provides many useful quantitative methods to assist neuropsychologists with interpretation of a case's performance (Crowe, 2010). The aim of the present paper is to offer a range of additional quantitative methods for the analysis of WAIS-IV Index scores. Some of the methods offered are alternatives to the existing methods but the majority are designed to complement them.

Traditional confidence limits on WAIS-IV Index scores

Confidence limits on test scores are useful because they serve the general purpose of reminding neuropsychologists that test scores are fallible (they counter any tendencies to reify the scores obtained on testing) and serve the very specific and practical purpose of quantifying this fallibility (Crawford, 2004). In the WAIS-IV, confidence intervals for Index scores are a form of true score confidence intervals (based on a particular form of the standard error of estimation, rather than the standard error of measurement) and are centered on estimated true scores, rather than on a case's obtained scores (Glutting, McDermott, & Stanley, 1987). The estimated true score is obtained by regressing the obtained score towards the normative sample mean.

The traditional approach (Charter & Feldt, 2001) to obtaining confidence limits for true scores expresses the limits on an obtained score metric and are centered on a case's obtained score rather than estimated true score. The limits are obtained by multiplying the standard error of measurement for obtained scores (formula 1) by an appropriate value of z , a standard normal deviate, (e.g., $z = 1.96$ for 95% two-sided limits, and 1.645 for 90% two-sided limits). The formula for the standard error of measurement is simply

$$\text{SEM}_x = s_x \sqrt{1 - r_{xx}}, \quad (1)$$

where r_{xx} is the reliability coefficient for the test, and s_x is the test's standard deviation. The formula to then obtain the required confidence limits is

$$\text{CI} = X_0 \pm z(\text{SEM}_x). \quad (2)$$

These alternative confidence limits are offered here for a number of reasons. First, Charter and Feldt (2001) have criticised the Glutting et al. method; the arguments are technical but centre around the mixing of parameter estimates from different theories of measurement. Second, as Charter and Feldt (2001) point out, J.C. Stanley, the principal psychometric theorist on the Glutting et al. (1987) paper, appears to have reverted to the traditional approach in subsequent writings (Hopkins, Stanley, & Hopkins, 1990). Third, traditional limits are more transparent and easier to interpret. Fourth, it is questionable whether, in neuropsychological populations (in which cognitive impairment will be common), it is reasonable to centre confidence limits around an estimated true score formed by regressing an observed score towards the mean of the normative (unimpaired) population. For example, in the population of patients who have suffered a severe head injury, the mean score on the Processing Speed Index of the WAIS-IV will not be 100 and yet, if we centre the confidence limits for a severely head injured client on the estimated true score, we proceed as

though it were (that is, in all likelihood we will have regressed the observed score towards the wrong mean). For further discussion of the differences between obtained and estimated true score limits see Charter and Feldt (2001) and Crawford and Garthwaite (2009).

In summary, opinions differ over the virtues of the Glutting et al. method of setting confidence limits on WAIS-IV scores versus the use of traditional limits. Given that, it is useful for clinicians to have the option of using the latter limits, hence their inclusion in the present paper and in the accompanying computer program.

Confidence limits on Index scores expressed as percentile ranks

As noted, all authorities on psychological measurement agree that confidence intervals should accompany test scores (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). However, it remains the case that some neuropsychologists do not routinely record confidence limits, whether these be the Glutting et al. form of limits or traditional limits. There is also the danger that others will dutifully record the confidence limits but that, thereafter, these limits will play no further part in test interpretation (Crawford & Garthwaite, 2009). Thus, anything that serves to increase the perceived relevance of confidence limits should be encouraged. Crawford and Garthwaite (2009) have recently argued that expressing the endpoints of confidence intervals as percentile ranks will help to achieve this aim (they also provided such limits for the WAIS-III and WISC-IV).

Expressing confidence limits on a score as percentile ranks is very easily achieved: the standard score limits (obtained using the traditional approach as outlined in the previous section) need only be converted to z and the probability of z (obtained

from a table of areas under the normal curve or algorithmic equivalent) multiplied by 100. For example, suppose an individual obtains a score of 84 on the Verbal Comprehension Index (the score is therefore at the 14th percentile): using the traditional method of setting confidence limits, the 95% lower and upper limits on this score (78 and 90) correspond to z s of -1.47 and -0.67 . Thus the 95% confidence interval, with the endpoints expressed as percentile ranks, is from the 7th percentile to the 25th percentile.

The WAIS-IV manual does not report confidence intervals as percentile ranks. However, as Crawford and Garthwaite (2009) argue, such intervals are more directly meaningful than standard score intervals and offer what is, perhaps, a more stark reminder of the uncertainties involved in attempting to quantify an individual's level of cognitive functioning. At the risk of laboring the point, the lower limit on the percentile rank in the foregoing example (the lower limit is at the 7th percentile) is more tangible than the Index score equivalent (78) since this latter quantity only becomes meaningful when we know that 7% of the normative population is expected to obtain a lower score.

Estimating the percentage of the normative population that will exhibit j or more abnormally low Index scores on the WAIS-IV

Information on the rarity or abnormality of test scores is fundamental in interpreting the results of a neuropsychological assessment (Crawford, 2004; Strauss, Sherman, & Spreen, 2006). When attention is limited to a single test (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 5th percentile then, by definition, 5% of the population is expected to obtain a score that is lower (for example, in the case of

WAIS-IV Indexes, scores of 75 or lower are below the 5th percentile). However, there are four Index scores in total 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 considered in isolation, and knowledge of it is liable to guard against over inference; that is, concluding impairment is present on the basis of one “abnormally” low score when such a result is not at all uncommon in the general, healthy 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; Crawford, Garthwaite, & Gault, 2007).

One approach to this issue would be to tabulate the percentages of a test battery’s standardization sample exhibiting j or more abnormal scores; that is, the question could be tackled empirically. However, as yet, this form of base rate data has not been provided for the WAIS-IV. The alternative approach adopted here is to use a Monte Carlo method developed by Crawford, Garthwaite, and Gault (2007) to estimate¹ 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 full-length WAIS-III and WISC-IV (Crawford et al., 2007) and on short-form versions of 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, Testa, Winicki, Pearlson, & Gordon, 2008).

¹ Note that the alternative 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*.

Comparing Index scores against an individual's mean Index score

As noted above, comparison of an individual's test scores against normative data is a basic part of the assessment process. However, it is widely recognized that 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 on a given Index but this may still represent a serious decline for the individual concerned. Conversely, a case may score well below the normative mean but this may be entirely consistent with the case's premorbid ability. Index scores on the WAIS-IV are moderately to highly correlated in the general population, thus large discrepancies in a case's Index score profile may suggest an acquired impairment on those Indexes that are performed relatively poorly (Crawford, 1992).

The WAIS-IV manual (Tables B.1 and B.2) allows users to examine the reliability and abnormality of differences among a case's Index scores. The method adopted is that of *pairwise* comparisons (that is, each Index score is compared against every other Index score). An alternative to pairwise comparisons is to obtain the individual's mean Index score and compare each Index to this mean. The advantages of this approach, which was developed by Silverstein (1984) is most evident when there are a large number of tests in a battery as it serves to reduce the number of comparisons involved to manageable proportions. For example, if there are 10 tests in a battery there are 45 possible pairwise comparisons.

Even with a smaller number of components, such as is the case with the WAIS-IV's four Indexes, the Silverstein method, by providing a common

individualized comparison standard for each Index, has advantages. In arriving at a formulation, neuropsychologists have to integrate the information from a profile analysis of a given battery with other test data, and information from a host of other sources (i.e., the medical history, the clinical interview, behavioral observations etc). Anything that eases this burden is to be encouraged. Longman (2004) presented tables for use with the WAIS-III to allow psychologists to estimate the reliability and abnormality of the deviations of Index scores from patients' mean Index scores. Equivalent tables for the WISC-IV have been provided by Flanagan and Kaufman (2004). In the present study we apply this approach to WAIS-IV Indexes. Critical values have been provided by Grégoire, Coalson, and Zhu (2011), but these do not include the full range of options presented here.

Confidence intervals on Index score deviations and pairwise differences

Regardless of whether examination of Index score differences is conducted using the pairwise approach, or by examining the deviations from an individual's mean Index score, it would be useful (and in keeping with contemporary thinking in neuropsychological assessment), to accompany the observed pairwise difference or deviations with confidence intervals. That is, just as we should not reify a case's score on a single Index, neither should we treat the *difference* between a case's Index scores as fixed and known (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). Rather, we should quantify the uncertainty associated with the point estimate of the difference.

Base rates for Index score differences

Again, regardless of whether examination of Index score differences is conducted using the pairwise approach, or by comparing each Index score to an individual's mean, it would also be useful to have base rate data on the number of abnormal differences in the normative population (just as it would be useful to have base rate data on the number of abnormally low scores, as was noted in an earlier section).

For example, suppose that we choose to define an abnormal difference between a given pair of WAIS-IV Index scores as a difference that is exceeded by less than 5% of the normative population. By definition, less than 5% of the normative population is expected to exhibit such a difference for this *particular* pair of Indexes. However, there are six pairs of such differences and therefore we should not be surprised to find that substantially more than 5% of the normative population will exhibit at least one abnormal difference overall. Fortunately, the Monte Carlo method referred to earlier can also generate base rates for such differences; indeed it has already been used for this purpose with the WAIS-III and WISC-IV (Crawford et al., 2007) and short-form versions of these scales (Crawford et al., 2008; Crawford et al., 2010).

In the present paper this method is used to provide base rates for both the number of abnormal pairwise differences between WAIS-IV Index scores (the approach adopted in the WAIS-IV manual for analyzing discrepancies) and the number of abnormal deviations from an individual's mean Index score (the alternative approach offered here – see the previous section).

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

It would be useful to have a single measure of the overall abnormality of a case's profile of WAIS-IV Index scores; i.e., a multivariate index that quantifies how

unusual a particular combination of Index scores is. One such measure 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.

This 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). However, it can equally be applied to an individual's profile of *Index* scores; see Crawford, Allum and Kinion (2008) for its use with short-form WAIS-III Index scores and Crawford, Anderson, Rankin and MacDonald (2010) for its use with short-form WISC-IV Index scores. Indeed we consider this latter usage preferable given that research indicates that analysis at the level of Wechsler factors (i.e., Indexes) achieves better differentiation between healthy and impaired populations than does analysis of subtest profiles (Crawford, Johnson, Mychalkiw, & Moore, 1997).

To date the Mahalanobis Distance Index has not been applied to WAIS-IV Index scores and we therefore implement it in the present paper. Note that calculating the MDI by hand is not a practical proposition, nor is it all practical to provide tabled values for it as there is a huge range of possible combinations of Index scores. Therefore, the MDI for a case's profile of Index scores is provided only by the computer program that accompanies this paper.

The distinction between the abnormality of differences and the reliability of differences

The supplementary methods developed in this paper examine both the reliability of

differences and the abnormality of differences between Index scores. The distinction is crucial (Crawford, 2004) and, although it will be fully appreciated by many neuropsychologists, it is worth briefly setting it out. If a difference is reliable, this means it is unlikely to have arisen as a result of measurement error in the tests. That is, the observed difference is liable to reflect a genuine difference. However, reliable differences can be very common in the cognitively intact population, particularly if the tests have high reliability. Thus, on their own, reliable differences cannot serve as a basis for inferring acquired impairment on the Index that is performed more poorly.

Automating calculations using a computer program

With the exception of the Mahalanobis distance index, 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, a computer program provides a very convenient alternative for busy neuropsychologists as, on being provided with a case's Index scores, it could perform all of the necessary calculations and record the results. A computer program has the additional advantage that it will markedly reduce the likelihood of clerical error. Research shows that psychologists make many more simple clerical errors than we like to imagine (e.g., see Faust, 1998; Sherrets, Gard, & Langner, 1979; Sullivan, 2000). In view of the foregoing, we developed a computer program to accompany this paper.

Method and Results

Traditional confidence intervals on WAIS-IV Index scores expressed on an Index score metric and as percentile ranks

The method for obtaining traditional confidence limits on Index scores, and the means

whereby these can then be converted to confidence limits expressed as percentile ranks, has already been set out in the Introduction and so is not repeated here. To illustrate these methods Table 1 presents a range of scores for the VC and PS Indexes (chosen because they are the most and least reliable Indexes respectively) and records the 95% confidence limits in both forms. So for example, if a case obtains a score of 89 on VC then, using the traditional limits expressed on the standard Index score metric, the interval is from 83 to 95. It can also be seen that, when the limits are expressed as percentile ranks, the interval is from 13 (i.e., the lower limit is at the 13th percentile) to 37.

~Tables 1 and 2 about here~

Reliability of Index score deviations from an individual's Index score mean

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_{Dev}). The formula is

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

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 WAIS-IV technical and interpretive manual (Wechsler, 2008b) using the standard formula. The resultant SEM_{Dev} for each of the four WAIS-IV Indexes are reported in Table 2. Critical values are obtained by multiplying the SEM_{Dev} by the standard normal deviate (z) corresponding to the

required value of alpha. Thus, for example, the SEM_{Dev} is multiplied by 1.645 to obtain the critical value for a one-tailed test with alpha of 0.05 (for a two-tailed test with the same alpha a z value of 1.96 is used). Critical values for one and two-tailed tests and four values of alpha (0.15, 0.10, 0.05, and 0.01) are presented in Table 2.

To illustrate, suppose that a patient has a Working Memory Index score of 85, and a mean Index score of 92, and that a one-tailed test is required because such a weakness is considered typical of traumatic brain injury. The difference (which we will hereafter refer to as a deviation to avoid any potential confusion with pairwise differences between Indexes) is -7 and by referring to Table 2 it can be seen that this exceeds the critical value of 5.27 and is therefore significant at the 0.05 level.

A closely related alternative to the procedure just described is to divide the difference by its standard error and treat the resultant quantity as a standard normal deviate (z); a precise probability (one- or two-tailed) can then be obtained for this quantile. Thus to continue with the example, dividing the deviation score for the WMI by its standard error (of 3.204) yields a z of -2.185 and the one-tailed probability for this z is 0.014. This latter approach is implemented in the computer program that accompanies this paper.

Sequential Bonferroni correction for multiple comparisons when testing for reliable differences

Testing for reliable deviations between Indexes and an individual's mean Index score involves making multiple comparisons (as there are four Index scores, there are four possible comparisons). Thus, if all comparisons are made, there will be an inflation of the Type I error rate. Although neuropsychologists will often have an *a priori* hypothesis concerning such deviations it is also the case that there is often insufficient

prior information to form firm hypotheses. Moreover, should a neuropsychologist wish to attend to large, unexpected, deviations in a patient's profile then, for all intents and purposes, they should be considered to have made all possible comparisons.

One possible solution to the multiple comparison problem is to apply a standard Bonferroni correction to the p values. That is, if the family wise (i.e., overall) Type I error rate (α) is set at 0.05 then the p value obtained for an individual deviation would have to be less than $0.05/4 = 0.0125$ to be considered significant at the specified value of α . This, however, is a conservative approach that will lead to many genuine deviations 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 comparisons that were significant are set aside and the procedure is repeated with $k - l$ in the denominator rather than k , where l is 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 could easily be performed by hand but, for convenience, the computer program that accompanies this paper offers a sequential Bonferroni correction as an option. Note that, when this option is selected, the program does not produce exact p values but simply records whether the deviations are significant at the .05 level after correction.

The multiple comparison issue applies equally to the existing *pairwise* comparison of Index scores used in the WAIS-IV administration and scoring manual.

The application of a sequential Bonferroni correction to pairwise comparisons requires dividing alpha by 6 rather than 4 (as there are six possible pairwise comparisons between Indexes). Again, although this correction could be applied by hand, the computer program accompanying this paper also offers the option of applying a sequential Bonferroni correction for pairwise comparisons.

Confidence Intervals on Index score deviations and on pairwise differences

As noted in the Introduction, it would be useful to quantify the uncertainty introduced by measurement error when comparing a case's pairwise differences, or deviations from the Index score mean, by forming confidence intervals. These intervals are not offered in the WAIS-IV manual. In the case of pairwise differences, to obtain these intervals the standard error of measurement of the difference between a given pair of Indexes need only be multiplied by 1.96 and the result added (for the upper limit of the interval) and subtracted (for the lower limit) from the observed difference. Thus, if a case obtains a difference of, say, 23 points between the VC and PR Indexes ($SEM_D = 4.50$), then the 95% confidence interval on this difference is from 14 to 32.

Similarly, if comparisons of Index scores is conducted by comparing scores against the case's mean Index score, then the relevant SEM_{Dev} is multiplied by 1.96 and added and subtracted from the observed deviation. For example, if the deviation score for a case on the PR Index is, say, -13.8 (from Table 2, the SEM_{Dev} is 3.023) then the 95% CI on this deviation is from -19.7 to -7.8 .

Abnormality of Index score deviations from a case's Index score mean

The three foregoing sections were concerned with the *reliability* of differences between Indexes (either pairwise differences or deviations from an individual's mean

Index score). Although it is important to test whether deviations are reliable (i.e., are unlikely to have arisen solely from measurement error) it may be quite common for members of the healthy general adult population to exhibit reliable differences, particularly when, as is the case here, the measures involved have high reliabilities. That is, a difference may be genuine but not unusual. Therefore, it is important, particularly when attempting to determine if a case has an acquired impairment, to also address the question of whether deviations from a case's mean Index score are rare or abnormal. To that end we set the calculations required to estimate how large deviations have to be such that they exceed those expected to occur in various percentages of the normative population².

To estimate these percentages requires calculation of the standard deviation of the difference between an Index and individuals' mean Index scores. The formula required uses the Index score correlation matrix (which was obtained from Table 5.1 of the WAIS-IV technical and interpretive manual):

$$SD_{Dev_a} = s\sqrt{1 + \bar{\mathbf{G}} - 2\bar{\mathbf{h}}_a}, \quad (4)$$

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 all other tests (including test a itself; that is, the unity in the main diagonal is included in this row mean). As there are four Index scores, this formula is applied four times to obtain the standard deviations of the difference between each Index and the mean Index score.

The standard deviations of the difference between each Index score and the

² Note that it is not necessary to set out the equivalent calculations for the existing *pairwise* method of examining differences as the relevant percentages are already provided in the WAIS-IV manual.

mean Index score are presented in Table 3. To calculate the size of deviation required for a specified level of abnormality the standard deviation of the difference for each Index was multiplied by values of z (standard normal deviates). The deviations required to exceed the deviations exhibited by various percentages of the normative population are also presented in Table 3. Two sets of percentages are listed – the first column records the size of deviation required regardless of sign, the second column records the deviations required for a directional difference.

To illustrate, suppose a case obtains a VC Index score of 106 and a mean Index score of 91.5; the deviation is therefore +14.5 points. Ignoring the sign of the deviation, it can be seen from Table 3 that this deviation is larger than that required (13.93) to exceed all but 10% of the population but is not large enough to exceed all but 5% of the population (deviation required = 16.60 points). If the concern is with the percentage of the population expected to exhibit a deviation *in favour* of the VCI, it can be seen that this deviation is larger than that required (13.93) to exceed all but 5% of the population but is not large enough to exceed all but 1% (deviation required = 19.70 points).

A closely related alternative to the approach outlined to is to divide an observed deviation by its standard deviation and refer the resultant z (z_D) to a table of areas under the normal curve (or algorithmic equivalent) to obtain a precise estimate of the percentage of the population expected to exhibit this large a deviation. To continue with the current example, it is estimated that 8.69% of the population would exhibit a deviation of 14.5 points between the VC Index and the mean Index score regardless of the sign of the deviation and that 4.34% would exhibit a difference of 14.5 points in favour of VC. This latter approach is that used in the computer program that accompanies the present paper (as was the case for reliable deviations,

these data are not presented in the present paper because they would require voluminous tables).

~Tables 3 and 4 about here~

Base rates for j or more abnormally low Index scores

As noted in the Introduction, the Monte Carlo method developed by Crawford et al. (2007) is used in the present study to estimate base rate data for WAIS-IV Index scores. That is, the percentage of the normative population expected to exhibit j or more abnormally low Index scores is quantified. The technical details of this method are described in Crawford et al (2007) and so are not repeated here.

Table 4 records the percentage of the normative population expected to exhibit j or more abnormally low Index scores. Seven criteria for an abnormally low Index score are applied, ranging from a score that is exhibited by less than 25% of the normative population (a very liberal criterion) through to a score that is expected to be exhibited by less than 1% (a very stringent criterion). Our own preference is to define an abnormally low score as a score exhibited by less than 5% of the normative population and so the percentages for this criterion are set in bold (this criterion will also be used later in a worked example).

It can be seen from Table 4 that although, by definition, 5% of the normative population would be expected to exhibit an abnormally low score on any *single* WAIS-IV Index, a much larger percentage of the population (13.74%) is expected to exhibit one or more abnormally low Index scores out of the possible four Index scores. However, it can also be seen that the percentages fall off fairly steeply when moving to two or more abnormal scores and beyond; e.g., suppose a case exhibited 3 abnormally low scores, only 1.52% of the normative population are expected to exhibit this number or more of abnormally low scores.

It can be seen that, not surprisingly, the criterion for abnormally low scores exerts a considerable impact on the estimated percentages. For example, if the most liberal of the criterion was chosen (i.e., an abnormally low score was taken to be a score below the 25th percentile), then a majority of the normative population (50.12%) are expected to exhibit at least one such low score. Some readers may be puzzled as to why two of the criteria for abnormality are so close to each other in magnitude (i.e., < 15.9th percentile and < 15th percentile); the first of these criterion was offered because some neuropsychologists prefer to define an abnormally low score as a score that is more than one standard deviation below the normative mean, and that corresponds to the 15.9th percentile.

Estimated percentages of the population exhibiting j or more abnormally large pairwise differences between Index scores

The results of estimating the percentage of the normative population exhibiting j or more abnormally large pairwise differences between WAIS-IV Index scores are presented in Table 5. If an abnormal difference between a single pair of Indexes is defined as a difference exhibited by less than 5% of the population, then it can be seen that a reasonable percentage of the population are expected to exhibit one or more abnormal pairwise differences (20.17%). As was the case for the number of abnormally low scores, the percentages drop off fairly rapidly thereafter (e.g., 7.71% are expected to exhibit two or more abnormally large differences).

Estimated percentages of the normative population exhibiting j or more abnormally large deviation scores relative to their mean Index score

The estimated percentages of the normative population exhibiting j or more

abnormally large deviation scores on the WAIS-IV are presented in Table 6. As was the case for pairwise differences, when an abnormal deviation score for each individual Index is defined as one exhibited by less than 5% of the population, a sizeable percentage of the healthy population (16.71%) are expected to exhibit one or more of such abnormal deviations. Again however, the percentages fall off rapidly as j increases. For example, only 3.20% of the population is expected to exhibit two or more abnormal deviations.

~Tables 5 and 6 about here~

Applying Huba's Mahalanobis Distance Index (MDI) to WAIS-IV Index score profiles

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

$$(\mathbf{x} - \bar{\mathbf{x}})' \mathbf{W}^{-1} (\mathbf{x} - \bar{\mathbf{x}}), \quad (5)$$

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; the covariance matrix is easily obtained from the WAIS-IV correlation matrix by multiplying all elements of the matrix by the scalar quantity 225 ($=15 \times 15$). 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 ($k = 4$ for the present problem). 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 the Discussion section.

Computer program

A computer program for PCs, WAIS4_Supplementary_Analysis.exe, has been written

to accompany this paper. The program will also run on a Mac with PC emulation software installed. The program can be downloaded, either as an executable file, or as a zip file, from the following web page:

<http://www.abdn.ac.uk/~psy086/dept/WAIS4Supp.htm> .

DISCUSSION

Application of the present supplementary methods: A worked example

To illustrate the potential applications of the present methods we take the hypothetical case of a patient who had suffered a severe traumatic brain injury and had been administered the WAIS-IV as part of a more comprehensive assessment. The case's scores on the WAIS-IV Indexes were as follows: Verbal Comprehension (VC) = 118; Perceptual Reasoning (PR) = 107, Working Memory (WM) = 77; and Processing Speed (PS) = 68.

The results of applying the methods developed in the present paper to this case are presented in Figure 1. The outputs resemble those provided by the accompanying computer program. Note that, although we discuss pairwise differences between Indexes in this section, the figure (Fig. 1) gives results only for the deviation method; this is how we recommend the methods be used in practice, e.g., neuropsychologists would decide to use *either* the pairwise *or* deviation method, not both.

Furthermore, because examples of confidence intervals on scores (expressed as traditional intervals and as percentile ranks) have already been provided, as have examples of testing for *reliable* differences between Indexes, the focus of the discussion will be on the methods that quantify the *abnormality* of the case's Index scores and Index score differences.

Suppose we define an abnormally low score as one that falls below the 5th

percentile (Index scores of 75 or lower are below the 5th percentile). Then, as can be seen from Fig. 1a, one of the case's Index scores (PS) would be classified as abnormally low. Referring to Table 4 (or in this case simply consulting Fig. 1a) it can be seen that 13.74% of the population is expected to exhibit one or more abnormally low scores. Thus, although the case's PS score is unusually low, it is by no means highly unusual for a member of the general population to exhibit an abnormally low score on one of the four Indexes.

Turning to pairwise comparisons of the case's Index scores, suppose we define an abnormal pairwise difference as one that is exhibited by less than 5% of the population, regardless of the sign of the difference. By consulting Table B.2 of the WAIS-IV manual, it can be concluded that four of the case's six pairwise comparisons are abnormal (VC vs. WM, VC vs. PS, PR vs. WM, and PR vs. PS). For example, a discrepancy of 26 or more points between VC and WM is required to meet our chosen criterion for abnormality. That is, from Table B.2 of the WAIS-IV administration and scoring manual it can be seen 5.8% of the standardization sample exhibited a difference of 25 or more points between VC and WM, whereas only 4.7% exhibited a difference of 26 or more points. The discrepancy between the patient's scores on VC and WM is 40 points and therefore easily meets the criterion for abnormality. Referring to Table 5 we can see that only 0.17% of the population are expected to exhibit four or more abnormal pairwise differences.

Turning to examination of the deviations from the patient's mean Index score, suppose that, as previously, we use a deviation that is expected to be exhibited by less than 5% of the normative population as our criterion, regardless of sign. As can be seen from Fig. 1c, the mean Index score is 92.50 and, therefore the deviations from this mean are 25.50 for VC, 14.50 for PR, -15.50 for WM, and -24.50 for PS.

Referring to Table 3 (or simply referring to Fig. 1), two of the Index score deviations qualify as abnormally large according to our chosen criterion (VC is abnormally high, PS abnormally low). For example, for the PS Index, the (absolute) observed difference must exceed 19.25 to meet the chosen criterion for abnormality. Having determined how many of the case's deviation scores are abnormally large according to our chosen criterion, the next step is to refer to Table 6, it can be seen that only 3.20% of the population are expected to exhibit two or more abnormal deviations from their mean Index scores.

In this example the indications of abnormally large pairwise discrepancies and deviation scores survived the further scrutiny applied; that is, it is estimated that few healthy individuals would exhibit this number of abnormal score differences. Note also that the client's pattern of strengths and weaknesses is consistent with a head injury. This combination of profile and degree of abnormality gives a high degree of confidence in a conclusion that the case has suffered significant acquired impairment.

It will be appreciated, however, that this need not be the case. For example, suppose an individual exhibited only one abnormally large pairwise difference or one abnormally large deviation. Such results will not be uncommon in the healthy population, 20.17% are expected to exhibit one or more abnormally large pairwise differences (see Table 5) and 16.71% are expected to exhibit one or more abnormally large deviation scores (see Table 6). Even where the cognitive weakness in such a case is in line with clinical expectations (e.g., PS or WM in a head-injured client), more in the way of converging evidence from other sources would be required before one could be confident in inferring the presence of acquired impairment. In cases where there is little in the way of theory or prior empirical evidence to specify a likely pattern of impairment, a knowledge of the base rates for the number of abnormal

scores or score differences assumes a particular importance.

Returning to the case in hand, the conclusion that the case's profile of scores is abnormal is strongly supported by consulting the results of applying the Mahalanobis Distance Index to the case's scores (see Fig. 1d): the chi square value is highly significant ($p = 0.00184$): thus we can reject the null hypothesis that the case's profile is an observation from the profiles in the normative population. It is estimated that only 0.18% of the normative population would exhibit a more unusual profile.

Before concluding this example, it should be noted that making use of the data provided in the present paper adds little to the time taken to analyze a patient's profile. The time consuming aspects are those that precede use of the tables presented here and these will already form a part of many neuropsychologists' practice (that is, most neuropsychologists are aware that they should be concerned with the degree of abnormality of any scores or score differences present in a case's profile, and use both psychometric and informal methods to ascertain these). Moreover, as noted earlier, it is not suggested that neuropsychologists use both the pairwise *and* the deviation approach to examining discrepancies. Our own preference is for the approach of Longman (2004), see also Grégoire et al (2011), but we recognize that many may prefer, or at least be more familiar with, the pairwise method adopted in the WAIS-IV administration and scoring manual. Finally, if the computer program accompanying this paper is used, all calculations (including the determination of how many of a case's scores and score differences are abnormal) are automated and are available instantaneously (see next section and Fig. 1).

~Figures 1 and 2 about here~

Using the accompanying computer program

The program requires that users enter an individual's scores on the four Indexes in the

data fields provided, and then select their preferred analysis options using radio buttons. The program automates the process of determining whether each of an individual's Index scores are abnormally low, and (depending on which option has been selected) whether they exhibit abnormally large pairwise differences, or abnormally large deviations from the mean Index score. A screen capture of the input form for the program is presented as Fig 2; the input data and options selected are those used in the foregoing worked example, as recorded in Fig 1.

It should be noted that, if a user has opted for pairwise analysis of differences, the program estimates the abnormality of each pair *statistically* rather than using the *empirical* base rates as found in the relevant test manuals (e.g., Table B.2 of the WAIS-IV administration and scoring manual). The former approach is in keeping with the use of statistical rather than empirical methods to estimate the abnormality of deviations from individuals' mean Index scores in the present paper (formula 4) and in the tables provided by Longman (2004) and Flanagan and Kaufman (2004).

The result is that, on occasion, the number of pairwise differences or deviations estimated to be abnormal will differ depending on whether the program is used or the empirical base rate data in the administration and scoring manual. For example, take the earlier example in which a TBI case exhibited four abnormal pairwise differences between his Index scores. Suppose however that the case scored 103 rather than 107 on the Perceptual Reasoning Index. Using the statistical method (as implemented in the WAIS-IV program) the case is still classified as exhibiting four abnormal pairwise differences – the difference (of 26 points) between the PR Index and WM remains abnormal (it is estimated that 4.68% of the healthy population will exhibit this size of a difference or larger; this is less extreme than in the original example [2.18%] but nevertheless still meets the selected criterion for abnormality).

However, using the empirical base rates, this difference is not classified as abnormal (from Table B.2, 5.1% of the normative sample exhibited a difference of this magnitude or larger, compared to 3.1% for the difference of 30 points in the original example).

A number of factors contribute to differences between empirical and statistical base rates, for example statistical base rates are unaffected by the inevitable small “bumps and wiggles” that will occur in empirical distributions even when normative samples are large. However, the most important factor is that a number of people in the normative sample will obtain the same difference score as that obtained by the individual of interest. Thus, when empirical base rates are employed, typically, the percentage recorded and read off by the user is the percentage equalling or exceeding this difference. In contrast, because the statistical approach treats the data as continuous, it will, in essence, credit half those obtaining the same difference score as having a larger difference and the other half as obtaining a smaller difference, thus the percentages will normally be a little lower than those obtained from empirical rates. This is akin to the procedure used for forming standard percentiles, moreover the same holds when estimating the abnormality of deviation scores using existing tables as these too are derived statistically.

Conclusion

The aim of the present paper was to develop a package of supplementary quantitative methods to assist with interpretation of the WAIS-IV. 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.

Although we consider that all of the methods developed here are useful, they are not interdependent. Therefore it is perfectly possible for a neuropsychologist to pick and choose among them. That is, a particular neuropsychologist 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. Still others may find that expressing confidence limits on a case's score as percentile ranks helps them assimilate the degree of uncertainty attached to a case's score whereas others consider they already have a sufficient grasp of the uncertainties without requiring such additional support. Indeed in the case of the methods offered for analysis of differences (pairwise versus deviations), it is intended that only one of these would be used at any one time.

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 concrete examples (they are chosen to be extreme). Suppose that a case obtains a score of 71 on all four Indexes. This is a very poor level of performance: from Table 4 it is estimated that only 0.37% of the normative population will obtain scores below the 5th percentile on all four Indexes. Although very poor, the case's performance is remarkably consistent. For this example the chi square for the MDI is not significant ($\chi^2 = 5.643$ on 4 df, p value = 0.22747) underlining that the MDI is not sensitive to a case's absolute levels of performance.

In contrast, suppose that, as in the previous example, a case scored 71 on WM and PS, but obtained scores of 115 on the two remaining Indexes (VC and PR). In this scenario the MDI is highly significant: $\chi^2 = 20.810$ on 4 df, p value = 0.00035. The *profile* of scores is therefore *highly* unusual; very few individuals (0.035%) in the normative population would be expected to exhibit a more unusual profile of scores. In this latter example two of the case's Index scores are abnormally low; it is estimated that 4.4% of the normative population will exhibit this number of low scores (see Table 4); although this is unusual it is not nearly as unusual as the overall profile of performance as captured by the MDI. It can be seen then that the base rate data on low scores and the MDI are complimentary in the process of identifying cognitive difficulties. Needless to say if both methods converge to suggest either normal or abnormal performance then interpretation of the results is simplified and the neuropsychologist can have more confidence when arriving at a formulation.

Finally, the provision of additional quantitative methods of analysis by no means undermines the role of the neuropsychologist in decision making: rather, it should be viewed as an aid to such decision making. The neuropsychologist 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 person's cognitive strengths and weaknesses and thereafter develop the 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 neuropsychological assessment.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington DC: American Educational Research Association.
- 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.
- Charter, R. A., & Feldt, L. S. (2001). Confidence intervals for true scores: Is there a correct approach? *Journal of Psychoeducational Assessment*, *19*, 350-364.
- Crawford, J. R. (1992). Current and premorbid intelligence measures in neuropsychological assessment. In J. R. Crawford, D. M. Parker & W. W. McKinlay (Eds.), *A handbook of neuropsychological assessment* (pp. 21-49). London: Erlbaum.
- 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. (2009). Percentiles please: The case for expressing neuropsychological test scores and accompanying confidence limits as percentile ranks. *The Clinical Neuropsychologist, 23, 193-204.*

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., Morrice, N., & Duff, K. (in press). Some supplementary methods for the analysis of the RBANS. *Psychological Assessment.*

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

Crawford, J. R., Johnson, D. A., Mychalkiw, B., & Moore, J. W. (1997). WAIS-R performance following closed head injury: A comparison of the clinical utility of summary IQs, factor scores and subtest scatter indices. *The Clinical Neuropsychologist, 11, 345-355.*

- Crowe, S. F. (2010). *Evidence of absence: a guide to cognitive assessment in Australia*. Melbourne: Australian Academic Press.
- 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.
- Glutting, J. J., McDermott, P. A., & Stanley, J. C. (1987). Resolving differences among methods of establishing confidence limits for test scores. *Educational and Psychological Measurement, 47*, 607-614.
- Grégoire, J., Coalson, D. L., & Zhu, J. (2011). Analysis of WAIS-IV index score scatter using significant deviation from the mean index score. *Assessment, 18*, 168–177.
- Hopkins, K. D., Stanley, J. C., & Hopkins, B. R. (1990). *Educational and psychological measurement and evaluation*. Englewood Cliffs, NJ: Prentice Hall.
- Huba, G. J. (1985). How unusual is a profile of test scores? *Journal of Psychoeducational Assessment, 4*, 321-325.
- Larzelere, R. E., & Mulaik, S. A. (1977). Single-sample tests for many correlations. *Psychological Bulletin, 84*, 557-569.
- Lezak, M. D. (1988). IQ: R.I.P. *Journal of Clinical & Experimental Neuropsychology, 10*, 351-361.
- 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.

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.

Silverstein, A. B. (1984). Pattern analysis: The question of abnormality. *Journal of Consulting and Clinical Psychology*, 52, 936-939.

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.

Wechsler, D. (2008a). *WAIS-IV administration and scoring manual*. San Antonio TX: Pearson.

Wechsler, D. (2008b). *WAIS-IV technical and interpretive manual*. San Antonio TX: Pearson.

Table 1. Illustration of traditional confidence limits expressed on a standard Index score metric (SM) and as percentile ranks (PR) for a range of obtained scores on the Verbal Comprehension and Processing Speed Indexes of the WAIS-IV

Score	Verbal Comprehension (VC)					Processing Speed (PS)			
	PR	Lower		Upper		Lower		Upper	
		SM	PR	SM	PR	SM	PR	SM	PR
122	93 rd	116	86 th	128	97 th	113	80 th	131	98 th
114	82 nd	108	71 st	120	91 st	105	62 nd	123	94 th
102	55 th	96	40 th	108	70 th	93	31 st	111	77 th
89	23 rd	83	13 th	95	37 th	80	9 th	98	46 th
76	5 th	70	2 nd	82	11 th	67	1 st	85	16 th

Note: the values chosen for the Index scores may appear a little arbitrary but were constrained by the fact that they had to be obtainable from the sum of scaled scores for both VC and PS (these two Indexes were chosen as they differ appreciably in their reliability)

Table 2. Testing for a reliable difference between a WAIS-IV Index score and an individual's mean Index score: Standard errors of measurement of the difference (SEM_{Dev}) and one- and two-tailed critical values

Index	SEM_{Dev}	One-tailed critical values				Two-tailed critical values			
		.15	.10	.05	.01	.15	.10	.05	.01
VC	2.831	2.93	3.63	4.66	6.58	4.08	4.66	5.55	7.29
PR	3.023	3.13	3.88	4.97	7.03	4.35	4.97	5.93	7.79
WM	3.204	3.32	4.11	5.27	7.45	4.61	5.27	6.28	8.25
PS	3.843	3.98	4.93	6.32	8.94	5.53	6.32	7.53	9.90

Table 3. Abnormality of differences between a WAIS-IV Index score and an individual's mean Index score: Standard deviations of the difference (SD_{Dev}) and size of difference required to exceed various percentages of the normative population

Index	SD_{Dev}	Directional difference				Difference regardless of sign			
		15%	10%	5%	1%	15%	10%	5%	1%
VC	8.469	8.77	10.86	13.93	19.70	12.20	13.93	16.60	21.82
PR	8.130	8.42	10.42	13.37	18.91	11.71	13.37	15.93	20.94
WM	7.990	8.28	10.24	13.14	18.58	11.51	13.14	15.66	20.58
PS	9.822	10.18	12.59	16.16	22.85	14.14	16.16	19.25	25.30

Note. The computer program accompanying this paper provides the estimated percentage of the normative population

Table 4. Percentage of the normative population expected to exhibit at least j abnormally low Index scores on the WAIS-IV; increasingly stringent definitions of abnormality are used ranging from $< 15.9^{\text{th}}$ percentile (i.e., more than 1 SD below the mean) to below the 1^{st} percentile.

Criterion for abnormality	Percentage exhibiting j or more abnormally low			
	WAIS-IV Index scores			
	1	2	3	4
$<25^{\text{th}}$	50.12	28.62	15.35	6.06
$<15.9^{\text{th}}$	35.69	17.22	8.00	2.69
$<15^{\text{th}}$	34.18	16.14	7.37	2.44
$<10^{\text{th}}$	24.66	10.08	4.12	1.20
$<5^{\text{th}}$	13.74	4.43	1.52	0.37
$<2^{\text{nd}}$	6.10	1.47	0.41	0.08
$<1^{\text{st}}$	3.23	0.63	0.15	0.02

Table 5. Percentage of the normative population expected to exhibit j or more abnormal pairwise differences (regardless of sign) between Index scores on the WAIS-IV; 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 for abnormality	Percentage exhibiting j or more abnormal pairwise differences (regardless of sign) between WAIS-IV Indexes					
	1	2	3	4	5	6
<25%	65.58	47.69	28.04	7.56	1.06	0.01
<15.9%	49.09	29.79	13.60	2.48	0.19	0.00
<15%	47.22	28.04	12.42	2.16	0.15	0.00
<10%	35.10	17.69	6.34	0.81	0.04	0.00
<5%	20.17	7.71	1.99	0.17	0.00	0.00
<2%	9.14	2.42	0.44	0.02	0.00	0.00
<1%	4.88	0.99	0.14	0.00	0.00	0.00

Table 6. Percentage of the normative population expected to exhibit j or more abnormal deviations between WAIS-IV Index scores and individual's mean Index scores (regardless of sign); increasingly stringent definitions of abnormality are used ranging from a deviation exhibited by less than 25% of the population to a deviation exhibited by less than 1%.

Criterion for abnormality	Percentage exhibiting j or more abnormal deviation scores (regardless of sign) on the WAIS-IV			
	1	2	3	4
<25%	61.22	32.82	5.00	0.92
<15.9%	44.14	17.51	1.52	0.23
<15%	42.26	16.17	1.31	0.19
<10%	30.53	8.95	0.44	0.06
<5%	16.71	3.20	0.07	0.01
<2%	7.21	0.80	0.01	0.00
<1%	3.74	0.29	0.00	0.00

Figure Legends

Figure 1. Illustrative example of results obtained when the supplementary methods for the analysis of WAIS-IV Index scores are applied (note that analysis of differences between Indexes is based on comparing each Index to the case's mean Index score)

Figure 2. Screen capture of the input for the computer program written to accompany the present paper

Figure 1.

(a)

Index scores with confidence limits expressed on an Index score metric and as percentile ranks:

Index	Score	(95% CI)	PR	(95% CI on PR)
Verbal Comprehension Index	118	(112 to 124)	88.5	(79.0 to 94.4)
Perceptual Reasoning Index	107	(100 to 114)	68.0	(51.1 to 81.7)
Working Memory Index	77	(70 to 84)	6.3	(2.2 to 14.6)
Processing Speed Index	68	(59 to 77)	1.6	(0.3 to 6.5)

NUMBER of case's Index scores classified as abnormally low = 1

PERCENTAGE of normal population expected to exhibit this number or more of abnormally low scores: Percentage = 13.74%

(b)

RELIABILITY of Index score deviations from the case's mean Index score:

Case's mean Index score = 92.50

Index	Deviation	(95% CI for deviation)	Two-tailed <i>p</i>	One-tailed <i>p</i>
VC minus Mean:	25.5	(20.0 to 31.0)	0.000	0.000
PR minus Mean:	14.5	(8.6 to 20.4)	0.000	0.000
WM minus Mean:	-15.5	(-21.8 to -9.2)	0.000	0.000
PS minus Mean :	-24.5	(-32.0 to -17.0)	0.000	0.000

(c)

ABNORMALITY of Index score deviations from the case's mean Index score, i.e., percentage of population estimated to obtain a larger deviation in same direction (figure in brackets is percentage regardless of sign):

Case's mean Index score = 92.50

Index Pair	Deviation	Percentage of population with larger deviations	Percentage with larger deviations regardless of sign
VC minus Mean:	25.5	0.13	0.26
PR minus Mean:	14.5	3.72	7.45
WM minus Mean:	-15.5	2.62	5.24
PS minus Mean:	-24.5	0.63	1.26

NUMBER of case's deviation scores that meet criterion for abnormality = 2

PERCENTAGE of normal population expected to exhibit this number or more of abnormal deviations = 3.20%

(d)

MAHALANOBIS DISTANCE Index of the overall abnormality of the case's Index score profile:

Chi-square = 17.108 on 4 df, *p* value = 0.00184

Percentage of normative population expected to exhibit a more unusual profile = 0.184%

