

WAIS—R subtest scatter: Base-rate data from a healthy UK sample

J. R. Crawford*

School of Psychology, Flinders University of South Australia, GPO Box 2WO, Adelaide SA 5001

K. M. Allan

St Johns Hospital, Edinburgh, UK

Base-rate data on subtest scatter for the WAIS-R was obtained from a sample of 200 healthy subjects recruited to match the adult UK population in terms of age, sex and social class distribution. Tables are presented which permit an assessment of the abnormality (i.e. rarity) of an individual's pattern of WAIS—R performance in terms of subtest range and subtest deviations from an individual's subtest mean. Guidance on the appropriate use of the tables is offered and the data are compared with data from the US WAIS—R standardization sample where appropriate. The distinction between reliable and abnormal differences is highlighted.

In clinical interpretation of WAIS—R performance, considerable emphasis is placed on the examination of subtest scatter, i.e. the subtest profile is used to identify a client's relative strengths and weaknesses (Crawford, 1992; Kaufman, 1990; Lezak, 1983). When conducting such an analysis it is important to be aware of the distinction between the *reliability* and *abnormality* of any subtest differences. A reliable difference is one that is unlikely to have arisen from measurement error and simply indicates that a client is genuinely superior on the abilities tapped by one measure over those tapped by another.

For many applications of the WAIS—R, determining whether there are reliable differences may be sufficient empirical analysis to guide interpretation. However, in clinical practice a common aim is to identify *impaired* areas of cognitive function whether this be for medico-legal, rehabilitation planning or diagnostic purposes. In these circumstances it is appropriate to examine also whether subtest discrepancies are *abnormal*, i.e. are of magnitude such that they exceed that shown by the majority of the general, healthy population.

One source of base-rate information for such purposes has been provided by Matarazzo, Daniel, Prifitera & Herman (1988). They calculated the subtest range for each subject in the WAIS—R US standardization sample (the range was simply the difference between an individual's highest and lowest subtest score). Ranges were

* Requests for reprints should be addressed to John R. Crawford, Department of Psychology, University of Aberdeen, AB9 2UB, UK.

calculated separately for the Verbal and Performance subtests and, for all 11 subtests combined; frequency tables were then constructed.

These tables demonstrate the large degree of w/ra-subject variability in abilities that can be expected in the healthy population and therefore provide a useful caution against overinference when working with clinical populations. For example the mean subtest range for the Full Scale was 6.66 scaled score points; for an individual's subtest range to be abnormal (operationally defined for present purposes as being of a sufficient magnitude such that it occurs in less than 5 per cent of the general population) it would have to exceed 11 scaled score points.

Although this base-rate information is useful and easy to use it does have limitations. First and most obvious, it utilizes only a small proportion of the available information contained within the subtest profile. Second, the tables of ranges are based on normal scaled scores. There is widespread agreement that any analysis of subtest profiles is best conducted after correcting for the effects of age (Crawford, 1992; Kaufman, 1990; Lezak, 1983). Therefore, tables of subtest ranges based on age-graded scaled scores might have been more useful and more in keeping with most clinician's existing practices.

Silverstein (1984) has developed an alternative table which considers the abnormality of a discrepancy between each of the 11 subtest scores and an *individual's* mean subtest score. This table has the disadvantage that its use requires more computation but has the advantage that it utilizes more of the subtest profile information. Furthermore, it is designed for use with age-graded scaled scores. As an illustration of the use of this table consider the example of an individual whose mean subtest score is 11.2 and whose Digit Symbol score is 6; a discrepancy of 5.2. Silverstein's table suggests that this size of discrepancy can be viewed as abnormal as it is estimated that it would be exhibited by less than 5 per cent of the standardization sample; to be highly abnormal (operationally defined for these purposes as being exhibited by less than 1 per cent of the standardization sample) the discrepancy would have to exceed 5.9.

The WAIS-R is the instrument most widely used in the UK for the assessment of cognitive abilities. Its importance has recently been highlighted by a British Psychological Society report which addressed the need to develop a definition of mental impairment for legal purposes (BPS, 1991). The recommended criterion was based solely on WAIS-R performance. However, despite its widespread use, the WAIS—R has not been standardized in the UK. Thus, although the type of base-rate information reviewed above is potentially of great value to UK clinicians, its applicability has yet to be determined.

The aim of the present study is to derive provisional base-rate information on subtest scatter for the UK from a sample of the healthy population recruited to be representative of the adult UK population in terms of age, sex and social class distribution. The approaches of Matarazzo *et al.* (1988) and Silverstein (1984) will be adopted and extended; the UK data will also be compared with US derived data where possible. Additionally, UK base-rate data will be provided for discrepancies between Verbal and Performance IQ.

Method

Participants

Two hundred subjects (104 female, 96 male), screened by interview for the absence of neurological or psychiatric disorder, participated in the present study. Subjects were recruited from a variety of sources including local and national companies, clubs (e.g. senior citizens clubs, angling clubs), community centres, etc. Most subjects were urban dwellers and most received a small honorarium for their participation.

The mean age of the sample was 44.3 years (SD = 19.2 years), with a range from 16 to 83 years. The mean number of years of education was 12.6 years (SD = 3.0 years) with a range from seven to 21 years. The social class of each subject was derived from their occupation, or former occupation, using the Classification of Occupations (OPCS, 1980).

The recruitment strategy was intended to obtain a sample that was broadly representative of the adult population in the UK, with respect to the distributions of social class, age and sex. To determine the extent to which this aim was met, the social class distribution in the present sample was compared with that of the UK adult population according to the 1981 Census (see Table 1).

Table 1. Social class distribution in the WAIS-R sample and in the adult UK population (the entries are percentages)

	Social Class				
	1	2	3	4	5
WAIS-R sample	7	27	42	17	8
General adult UK population	5	23	48	18	6

A chi-square goodness-of-fit test showed that the social class distribution in the present sample did not differ significantly from the population distribution ($\chi^2(4) = 5.66$, n.s.). A similar procedure was adopted to examine the representativeness of the sample in terms of age distribution. Nine age bands were formed, corresponding to those adopted for the WAIS-R standardization sample, with the exception that the 70-74 age band was replaced with a 70+ age band. A goodness-of-fit test revealed that the sample distribution did not differ significantly from the (census-derived) expected distribution ($\chi^2(8) = 7.71$, n.s.). Finally, a goodness-of-fit test revealed that the sex distribution did not differ significantly from the census-derived distribution ($\chi^2(1) = 0.01$, n.s.).

Procedure

All subjects were administered a full-length UK WAIS-R according to standard procedures (Lea, 1986; Wechsler, 1981). IQs, normal scaled scores and age-graded scaled scores were derived in the normal way. In a recent study of the basic psychometric properties of the WAIS-R, Crawford, Gray & Allan (1995) have reported that there are highly significant differences among WAIS-R subtests means in the UK. Because of this, these authors provided a table (Table 10) which adjusts US-derived age-graded scaled scores to have a common mean of 10 and SD of 3 in the UK. Given that the aim of the present study was to provide base-rate information on subtest scatter for the UK, all analysis of age-graded scaled scores was performed with the UK-adjusted scores.

Results

Summary statistics

Mean scores, SD and ranges for the three IQ scales are presented in Table 2.

Table 2. Summary statistics for the WAIS-R in a healthy UK sample ($N = 200$)

	Scale	Mean	SD	Range
	FSIQ	102.5	13.12	71–140
	VIQ	102.4	12.81	73–133
	PIQ	102.0	13.39	67–139

Subtest ranges

Subtest ranges were calculated for each subject based on their normal scaled scores. Cumulative frequency distributions were then formed and used to derive the critical values presented in Table 3 (the full frequency distributions can be obtained from the first author). This table shows the size of subtest range necessary for varying degrees of abnormality. For example, if the subtest ranges for the Verbal subtests are considered it can be seen that the mean subtest range was 5.09 and that a range of 10 would be necessary to be considered abnormal.

The equivalent critical values for the US standardization sample were extracted from Matarazzo *et al.*'s (1988) table and form the middle columns of Table 3. It can be seen that the critical values are very similar in the two samples. As most UK clinicians use age-graded scaled scores when examining subtest scatter, critical values for subtest ranges based on UK adjusted *age-graded* scores are also presented in Table 3.

Correlates of subtest ranges

Matarazzo *et al.* (1988) reported that in the US standardization sample the subtest range varied with FSIQ such that higher IQ subjects tended to have larger subtest ranges. To determine whether this would hold for the UK, the correlations between the subtest ranges and FSIQ were calculated. When the subtest ranges were based on normal scaled scores low, but significant correlations were obtained between FSIQ and the Performance and overall subtest ranges ($r = .23$ and $.20$ respectively, both $p < .01$). When the subtest ranges were based on age-graded scaled scores none of the correlations achieved significance. Matarazzo *et al.* (1988) also reported that age was unrelated to the subtest range in the standardization sample (as noted, the subtest ranges were based on normal scaled scores). In the present sample a low but significant correlation was obtained between age and the Performance subtest range when this was based on normal scaled scores ($r = .25$, $p < .01$). When the subtest ranges were based on age-graded scaled scores none of the correlations achieved significance.

Subtest deviations from the sub test means

To prepare a table based on Silverstein's (1984) method each subject's mean subtest scores (i.e. Verbal, Performance and overall mean subtest scores) were calculated and frequency distributions formed of the deviations of each of the 11 subtests from the

Table 3. W AIS-R subtest range necessary to *exceed* that exhibited by various percentages of a healthy UK sample (N = 200) and the US standardization sample (N = 1880)

Percentage	UK sample				Normal scaled scores				US standardization				Age-graded scaled scores						
	Verbal		Performance		All subtests		Verbal		Performance		All subtests		Verbal		Performance		All subtests		
	8	9	10	11	8	9	10	11	8	9	10	11	8	9	10	11	8	9	10
15	8	9	10	11	8	8	8	8	8	8	10	10	8	9	9	9	8	9	10
10	9	10	11	12	9	9	9	9	9	9	10	10	9	10	10	10	9	10	11
5	10	11	12	14	11	11	11	11	11	11	12	12	10	11	11	11	10	10	12
1	11	12	14	14	11	12	12	12	12	12	14	14	12	13	13	13	12	13	15
Mean	5.09	4.74	6.99	6.99	4.67	4.71	4.71	4.71	4.71	4.71	6.66	6.66	5.28	5.20	5.20	5.20	5.20	5.20	7.14
SD	1.93	2.17	2.03	2.03	1.87	2.10	2.10	2.10	2.10	2.10	2.08	2.08	2.04	2.15	2.15	2.15	2.04	2.15	2.24
Median	5	5	7	7	4	4	4	4	4	4	6	6	5	5	5	5	5	5	7

Table 4. Size of difference between each subtest and an individual's mean W AIS-R subtest score necessary to *exceed* the differences exhibited by various percentages of a healthy UK sample (regardless of sign)

Subtest	Verbal mean					Performance mean					Overall mean					
	15%		10%		5%	15%		10%		5%	15%		10%		5%	
	15%	10%	10%	5%	5%	15%	10%	10%	5%	5%	15%	10%	10%	5%	1%	
Information	2.51	3.01	3.01	3.51	4.84											
Digit Span	3.51	4.34	4.34	5.01	6.68											
Vocabulary	2.34	2.68	2.68	3.34	4.18											
Arithmetic	3.01	3.51	3.51	4.51	5.84											
Comprehension	2.51	2.68	2.68	3.34	4.01											
Similarities	2.51	2.84	2.84	3.51	6.01											
Picture Completion						2.81	3.01	3.01	3.41	3.41	5.01	5.01	2.83	3.19	4.01	5.92
Picture Arrangement						3.01	3.41	3.41	4.01	4.01	5.61	5.61	3.19	3.65	3.92	5.37
Block Design						2.61	3.21	3.21	3.61	3.61	5.01	5.01	3.19	3.56	4.46	5.74
Object Assembly						2.81	3.41	3.41	4.41	4.41	6.01	6.01	3.37	3.92	5.19	7.01
Digit Symbol						3.61	3.81	3.81	4.81	4.81	5.81	5.81	3.65	4.10	5.10	8.37

relevant subtest means. Critical values were then derived and are presented in Table 4. To illustrate the use of this table consider a client who has a mean Verbal subtest score of 14.2 age-graded scaled score points and a Digit Span score of 8. This yields a deviation score for Digit Span of 6.2 which is abnormal; less than 5 per cent of the present healthy sample exhibited this large a deviation (i.e. 6.2 is larger than the 5 per cent critical value of 5.01).

It will be noted that the critical values for the four degrees of abnormality vary with the subtest under consideration; larger discrepancies are necessary for abnormality with Digit Span, Object Assembly and Digit Symbol because they have relatively low correlations with other subtests.

Space constraints do not permit presentation of Silverstein's (1984) equivalent table derived from the US standardization sample for comparison purposes. The UK critical values tend to be larger than their US equivalents although the magnitude of these differences are relatively modest (see Silverstein, 1984, p. 938, Table 3).

Number of abnormal subtest deviations

If, in clinical practice, abnormal subtest deviations are present in an individual's profile a further issue arises; what percentage of the healthy population are liable to exhibit this particular number of cumulative deviations? Unfortunately, base-rate data relevant to this issue are not available for the US. However, UK base-rate data are provided in Table 5.

Table 5. Percentage of healthy subjects exhibiting abnormal subtest deviations from their Verbal, Performance or overall subtest means

	Number of abnormal subtest deviations					
	≥ 1	≥ 2	≥ 3	≥ 4	≥ 5	≥ 6
Verbal	22	2	0	0	0	0
Performance	14.5	5	1.5	0	0	0
Overall	37	7	2	0.5	0.5	0

Note. Abnormal is operationally defined as a subtest deviation which occurred in less than 5 per cent of the present healthy sample.

Tabulation of positive and negative subtest deviations

In line with Silverstein's table, Table 4 of the present study presented critical values for the subtest deviations regardless of sign (i.e. regardless of whether a subtest score was above or below the mean subtest score). It was considered that tabulating critical values separately for positive and negative subtest deviations would be of value in clinical practice. The critical values are presented in Table 6. Because of space constraints only critical values for the 15 per cent and 5 per cent level of abnormality are tabulated. The left-hand columns in Table 6 give the critical values when subtest scores are higher than the subtest mean, the right-hand columns the values when the subtest scores are lower.

Table 6. Size of difference between each subtest and an individual's mean WAIS-R subtest score necessary to exceed the differences exhibited by various percentages of a healthy UK sample; positive and negative discrepancies tabulated separately

Subtest	Verbal mean			Performance mean			Overall mean					
	S < M	5%	15%	S < M	5%	15%	S < M	5%	15%	S > M	5%	15%
Information	1.84	2.84	1.84									
Digit Span	2.18	4.84	2.51									
Vocabulary	1.68	2.51	1.51									
Arithmetic	2.31	3.51	2.18									
Comprehension	1.84	2.68	1.68									
Similarities	1.68	3.01	1.68									
Picture Completion							2.01	3.01	2.21	3.41		
Picture Arrangement				2.21	3.61	2.01	2.21	3.61	2.01	3.41		
Block Design				2.01	3.01	2.01	2.01	3.01	2.01	3.41		
Object Assembly				2.01	3.41	1.81	2.01	3.41	1.81	3.21		
Digit Symbol				2.61	3.81	2.61	2.61	3.81	2.61	3.81		
							1.83	3.10	2.10	3.92		
							2.56	5.10	2.74	4.19		
							1.83	2.74	1.92	3.19		
							2.65	3.74	2.37	4.28		
							2.10	2.92	1.92	3.01		
							1.92	2.74	1.92	3.10		
							2.19	2.83	2.01	3.92		
							2.37	3.74	2.10	3.46		
							2.01	3.28	2.01	3.65		
							2.83	4.01	2.56	3.92		
							2.56	3.83	2.83	4.37		

Verbal/Performance discrepancies

A frequency distribution of Verbal IQ/Performance IQ discrepancies was formed (available from the first author) and used to derive critical values for the same four levels of abnormality used in the subtest tables. The critical values are presented in Table 7; the first column gives values where the direction of the difference has been

Table 7. Verbal/Performance discrepancies necessary to *exceed* that exhibited by various percentages of a healthy UK sample

Percentage	UK Sample			US Sample ^b		
	V/P ^a	V > P	P > V	V/P ^a	V > P	P > V
15%	16	12	12	17	12	12
10%	18	14	15	20	15	15
5%	23	19	18	23–25	20	20
1%	28	28	28	30+	26–29	26–29

^a Regardless of sign.

^b Data derived from Matarazzo & Herman (1985, Table 3, pp. 916–917).

ignored, the second columns give the critical values when Verbal IQ is higher than Performance IQ, the final column the converse. For comparison purposes the equivalent critical values for the US standardization sample were extracted from tables given by Matarazzo & Herman (1985, pp. 916–917, Table 3). Some of the US values in this table could not be specified precisely as Matarazzo & Herman did not provide exact frequencies for extreme discrepancies; where this occurred the band within which the critical value lies is presented.

Reliable vs. abnormal VIQ/PIQ discrepancies

The WAIS-R manual (Wechsler, 1981) records the discrepancy necessary for a *reliable* (as opposed to abnormal) difference between the Verbal and Performance IQ

Table 8. Verbal/Performance discrepancy required for a statistically significant (i.e. reliable) difference and the percentage of subjects exhibiting this size of a difference or larger in the UK sample ($N = 200$) and US standardization sample ($N = 1880$)

p value	Discrepancy required ^a	% exhibiting in UK	% exhibiting in US ^b
.15	7	55.5	53.3
.10	8	50.5	47.4
.05	10	40.0	37.8
.01	13	25.5	24.3

^a .15 and .05 values obtained from the WAIS-R manual (Wechsler, 1981, Table 14, p. 35); .10 and .01 levels from Naglieri (1982; Table 2, p. 320).

^b Data derived from Matarazzo & Herman (1985, Table 3, pp. 916–917).

at the .15 and .05 probability levels. Naglieri (1982) has supplemented this information by calculating the discrepancy necessary for a significant difference at the .1 and .01 levels. To highlight the distinction between reliable and abnormal differences, the percentage of the present UK sample that exhibited these sizes of discrepancy or larger are presented in Table 8. The equivalent percentages from the US standardization sample are also presented; these were derived from the aforementioned table in Matarazzo & Herman (1985).

Discussion

The WAIS-R is widely used in the UK to assess cognitive abilities, particularly in work with neurological populations. The present study was motivated by the authors' unease over using a test in clinical decision making for which UK standardization data is lacking. The base-rate data presented here can never be as satisfactory as data from a UK standardization of the WAIS—R. However, the prospects of such a standardization are fairly remote. In addition, the care taken over matching the sample to the adult population in terms of age, sex and social distributions was such that considerably more confidence can be placed in the present data than could be placed in equivalent data derived from a sample of convenience.

Base-rate information, such as that presented here, is valuable because of the rationale underlying deficit measurement. In attempting to identify and quantify acquired deficits, comparison of an individual's test scores with the population mean is of limited usefulness because of the variability in cognitive ability in the general population. Because of this, much clinical interpretation is based on the analysis of discrepancies present in a client's profile. The assumption is that where an individual acquires cognitive impairment some cognitive ability measures will nevertheless be relatively preserved; these measures serve as an individualized comparison standard (Lezak, 1983). In determining the clinical significance of subtest discrepancies it is necessary to determine the frequency with which such discrepancies occur in the healthy, unimpaired population.

The importance of such empirical information is underscored by the fact that clinicians have normally had limited opportunity to test individuals drawn from the general healthy population. Because of this it is possible that we develop seriously distorted impressions of what degree of subtest scatter constitutes normal limits. On the basis of the present results, it could be argued that there is a particular danger of clinicians committing a Type I error, that is of inferring impairment from a subtest profile when the degree of subtest variability exhibited commonly occurs in the healthy population.

A dramatic illustration of the large degree of *nz*-subject variability that is to be expected in the general population is provided by the data in Table 3. On average a subtest range of 7 can be expected in the healthy population (mean subtest range in the UK sample = 6.99). If one operationally defines a subtest range as abnormal if it is exhibited by less than 5 per cent of the healthy sample, then it can be seen that the range would have to be 12 or greater to meet this criterion. Given that the subtests have an SD of 3, this represents a discrepancy of 4 standard deviations. It should also be noted that even if a less stringent definition of abnormality is adopted,

the discrepancies required are still massive (i.e. a range of 10 to exceed that exhibited by less than 15 per cent of the healthy sample).

The data presented in Table 3 were based on normal scaled scores. This analysis was performed in order to determine whether the large discrepancies which characterized the US standardization sample would also be observed in a UK sample; to permit a direct comparison the same method had to be adopted. Comparison of the UK and US data shows that the statistics are very similar and in fact verge on being identical. Thus, large ranges appear to be a consistent feature of healthy performance on the WAIS-R.

The supplementary columns in Table 3 present range data when age-graded scaled scores are used. As noted, we would suggest that this data should be used in preference to the normal scaled score data. In line with many authors, we consider it appropriate to base any analysis of subtest scatter on scores where the effects of normal ageing have been controlled for. The correlational analysis reported also indicates that the use of age-graded ranges simplifies clinical interpretation as they are independent of FSIQ and age.

The Silverstein (1984) method of analysing subtest scatter involves comparing each subtest against an *individual's* mean subtest score. Table 4 presents UK data which allows the clinician to evaluate the importance of these subtest deviations; where the magnitude of a subtest deviation is such that a large percentage of the general population equal or exceed it, then it would be inappropriate to accord it much clinical significance. The mechanics of using this table have been illustrated in the Results Section.

The Silverstein method has a number of features which commend it for use in clinical practice. First, in common with Matarazzo's approach, it is concerned with abnormal rather than reliable differences. Second, it employs age-graded scaled scores. Third, a problem with the analysis of subtest scatter is that the clinician is faced with an embarrassment of riches; with 11 subtests there are 55 possible subtest comparisons that could be made. As subtest scatter data has to be integrated with a host of other quantitative and qualitative information, some means of reducing this data to manageable proportions must be applied. Against this must be set the need to retain clinically significant attributes of the subtest profile. In the present authors' view the Silverstein method strikes an appropriate balance between these competing demands. The method lends itself readily to testing clinical hypotheses derived from other information sources or for generating hypotheses which can be followed up with further investigations. As these hypotheses will be directional, Table 6 was constructed to record critical values separately for positive and negative subtest deviations.

The presence of subtest scores which are abnormally higher than the subtest mean suggests that the client has suffered impairment across a relatively wide range of other cognitive functions. The hypothesis of a decline from a higher premorbid level could be integrated with information on occupational and educational history, qualitative aspects of the client's presentation and results on tests such as the NART (Nelson & Willison, 1992) to evaluate its likelihood. Where a subtest is abnormally lower than the subtest mean then, again, this information can be integrated with other pertinent information. For example, if Block Design is abnormally low then

this knowledge can be integrated with qualitative information on the client's performance on the subtest and with other test information to evaluate, for example, a clinical hypothesis of impaired planning vs. visuospatial dysfunction or a general slowing of information processing.

While it is true that the Silverstein method does not allow one to directly test hypotheses of a difference between *individual pairs* of subtests, such hypotheses are probably generated fairly rarely; where this does arise it should be possible to reformulate the hypotheses in terms of the subtests' deviations from the mean without doing too much violence to the underlying clinical reasoning.

It should be noted that unlike Matarazzo *et al.*, Silverstein (1984) did not have access to the US standardization sample raw data. Thus the critical values in his original table were *estimated* from a formula which incorporated the subtest means, SDs and correlation matrix rather than being constructed from a direct examination of the frequency of subtest deviations. In the present study the critical values were derived empirically from direct examination of the UK sample. Because of this it was a straightforward task to extend the analysis to determine the percentage of the healthy population that would be expected to exhibit a given number of abnormal subtest deviations (see Table 5). Theoretically it may be possible mathematically to derive similar data from the US standardization sample summary data but we are unaware of any existing procedure that would achieve this. The use of Silverstein's (1984) method has been discussed and endorsed by a number of authors (e.g. Crawford, 1992; Kaufman, 1990) but this issue does not appear to have been previously addressed. However, it can be seen from Table 5 that such data is an important adjunct to the basic information on subtest deviations. Take the example of a client whose profile contains three subtests which deviate from her/his overall subtest mean to the extent that they meet the criteria for abnormality (i.e. deviations this large are expected to occur in less than 5 per cent of the healthy population). Consulting Table 5 it can be seen that only 2 per cent of the present sample exhibited this number of abnormal subtest deviations; this would be consistent with a clinical hypothesis of acquired impairment of cognitive functioning.

In contrast, take the example of a client who exhibits only one abnormal subtest deviation. This subtest deviation would require an explanation as it occurs rarely in the healthy population. However, although less than 5 per cent of the healthy population would be expected to exhibit this *particular* subtest deviation, a relatively large percentage (37 per cent) would be expected to obtain at least one abnormal subtest deviation out of the total of 11 (see Table 5). It is clear that much more convergent information from other sources would be required to support a hypothesis of acquired impairment in this example than in the previous one.

The role of the Verbal/Performance discrepancy in clinical assessment is a matter of considerable contention. Matarazzo and his co-workers (e.g. Matarazzo & Herman, 1985) have made a case for viewing it as the best validated Wechsler index of dysfunction. Others are dismissive of its significance (e.g. Lezak, 1988) or argue that, if discrepancies between composite scores are to be used, the composites should be factorially derived (e.g. Canavan, Dunn & McMillan, 1986; Crawford, Allan, Stephen, Parker & Besson, 1989). For Kaufman (1990) the presence or absence of substantial Verbal/Performance discrepancies determines the type of analysis

subsequently performed on the subtest scatter (where such a discrepancy exists, Kaufman recommends that analysis of subtest discrepancies be restricted to subtests within the same scale). The present authors favour the use of factorially derived composites but have included Verbal/Performance base-rate information for the benefit of those UK clinicians who find their analysis useful.

Three features of the UK data on VIQ/PIQ discrepancies require comment. First, it can be seen (Table 7) that the distribution of extreme discrepancies is essentially symmetrical, e.g. extreme discrepancies in favour of VIQ are as likely as extreme discrepancies in favour of PIQ. Second, VIQ/PIQ discrepancies must exceed one SD before they can be considered abnormal. Third, although the discrepancies necessary for abnormality tend to be smaller in the UK sample, the similarities between the UK and US data are much more striking than the differences.

The important distinction between the reliability and the abnormality of a difference has been noted above. Table 8 was constructed to highlight this distinction using VIQ/PIQ discrepancies as an example. Table 8 shows that for a VIQ/PIQ discrepancy to be statistically significant (i.e. reliably different) at the .05 level it must exceed 10 IQ points. However, by examining the percentage of the healthy UK sample that exhibited a discrepancy of this size (or greater) it can be seen that it would be a serious error to view the presence of a reliable difference as suggestive of cognitive impairment; a very substantial percentage of the present sample (40 per cent) exhibited a significant difference between VIQ and PIQ. A similarly high percentage of the US standardization sample also exhibited such differences (37.8 per cent). If the less stringent .15 significance level is adopted, which has been suggested as appropriate when working with individuals (Wechsler, 1981) then it can be seen that the *majority* of the UK and US samples exhibit significant differences. These points have been illustrated with VIQ/PIQ discrepancies but they apply equally to the analysis of subtest discrepancies. Indeed the dangers of over-inference are potentially much greater for subtest analysis if some correction is not introduced for multiple comparisons (Crawford, 1992).

All of the base-rate information presented here has been derived from full-length WAIS—R data. It will therefore be appreciated that the validity of inferences made from it will be compromised if it is applied to scores obtained from a short-form administration. This is the case for all published base-rate data of which we are aware. It would be possible to generate equivalent tables for published short-forms (e.g. Britton & Savage, 1966; Crawford, Allan & Jack, 1992; Warrington, James & Maciejewski, 1986; Silverstein, 1982). However, many clinicians devise their own short-forms and, indeed, will vary the selection of subtests from client to client. The development of a computer program which would generate base-rate information for a user-defined selection of subtests (using the UK sample and US standardization sample data) is currently underway as is preparation of tables of base-rate information for commonly used short-forms.

Acknowledgements

This research was partly supported by the UK Mental Health Foundation and the Julia Farr Foundation, South Australia.

References

- British Psychological Society (1991). Mental impairment and severe mental impairment. *The Psychologist*, 4, 373-376.
- Britton, P. G. & Savage, R. D. (1966). A short form of the WAIS for use with the aged. *British journal of Psychiatry*, 112, 417-418.
- Canavan, A. G. M., Dunn, G. & McMillan, T. M. (1986). Principal components of the WAIS-R. *British journal of Clinical Psychology*, 25, 81-85.
- Crawford, J. R. (1992). Current and premorbid intelligence measures in neuropsychological assessment. In J. R. Crawford, W. McKinlay & D. M. Parker (Eds). *Handbook of Neuropsychological Assessment*, pp. 21-49. Hove, Sussex: Erlbaum.
- Crawford, J. R., Allan, K. M., Stephen, D. W., Parker, D. M. & Besson, J. A. O. (1989). The Wechsler Adult Intelligence Scale—Revised (WAIS-R): Factor structure in a UK sample. *Personality and Individual Differences*, 10, 1209-1212.
- Crawford, J. R., Allan, K. M. & Jack, A. M. (1992). Short-forms of the UK WAIS-R: Regression equations and their predictive accuracy in a general population sample. *British journal of Clinical Psychology*, 31, 191-202.
- Crawford, J. R., Gray, C. D. & Allan, K. M. (1995). The WAIS-R(UK): Basic psychometric properties in an adult UK sample. *British journal of Clinical Psychology*, 34, 237—250.
- Kaufman, A. S. (1990). *Assessing Adolescent and Adult Intelligence*. Boston, MA: Allyn & Bacon.
- Lea, M. (1986). *A British Supplement to the Manual for the Wechsler Adult Intelligence Scale—Revised*. San Antonio, TX: Psychological Corporation.
- Lezak, M. D. (1983). *Neuropsychological Assessment*, 2nd ed. New York: Oxford University Press.
- Lezak, M. D. (1988). IQ: R.I.P. *journal of Clinical and Experimental Neuropsychology*, 10, 351-361.
- Matarazzo, J. D., Daniel, M. H., Prifitera, A. & Herman, D. O. (1988). Inter-subtest scatter in the WAIS-R standardization sample. *journal of Clinical Psychology*, 44, 940-950.
- Matarazzo, J. D. & Herman, D. O. (1985). Clinical uses of the WAIS-R: Base rates of differences between VIQ and PIQ in the WAIS-R standardization sample. In B. Wolman (Ed.), *Handbook of Intelligence: Theories, Measurement and Applications*, pp. 899—932. New York: Wiley.
- Naglieri, J. A. (1982). Two types of tables for use with the WAIS—R. *journal of Consulting and Clinical Psychology*, 50, 319-321.
- Nelson, H. E. & Willison, J. (1991). *National Adult Reading Test Test Manual*, 2nd ed. Windsor: NFER-Nelson.
- Office of Population, Censuses and Surveys (1980). *Classification of Occupations*. London: HMSO.
- Silverstein, A. B. (1982). Two- and four-subtest short forms of the Wechsler Adult Intelligence Scale—Revised. *journal of Consulting and Clinical Psychology*, 50, 415-418.
- Silverstein, A. B. (1984). Pattern analysis: The question of abnormality. *journal of Consulting and Clinical Psychology*, 52, 936-939.
- Warrington, E. K., James, M. & Maciejewski, C. (1986). The WAIS as a lateralizing and localising diagnostic instrument: A study of 656 patients with unilateral cerebral lesions. *Neuropsychologia*, 24, 223-239.
- Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale—Revised*. New York: Psychological Corporation.

Received 16 May 1994; revised version received 6 October 1994