British Journal of Clinical Psychology (1994), **33,** 65-69 *Printed in Great Britain*

The Mahalanobis Distance index of WAIS-R subtest scatter: Psychometric properties in a healthy UK sample

J. R. Crawford*

School of Psychology, Flinders University of South Australia, GPO Box 2100, Adelaide 5001, Australia

K. M. Allan

Department of Psychology, University of Aberdeen

Burgess (1991) has proposed a new index of subtest scatter for the WAIS-R which uses a test statistic, the Mahalanobis Distance (MD). When used with the WAIS-R, MD scores should be distributed as chi square with 11 degrees of freedom. The suitability of the MD index for UK clinical practice was assessed by examining its psychometric properties in a sample of 200 healthy subjects. The sample, which was representative of the adult UK population in terms of age, sex, and social class distribution, was administered a full-length WAIS-R. A goodness-of-fit test revealed that the sample distribution of MD scores did not deviate significantly from the chi square distribution. Furthermore, the percentage of subjects exceeding the critical value for significance at the .05 level (6.5 per cent) corresponded closely to the expected percentage (i.e. 5 per cent). It is concluded that the MD index is suitable for use in UK clinical practice. Demographic characteristics were only weakly related to MD scores which simplifies clinical interpretation.

Burgess (1991) has presented a new means of analysing subtest scatter on the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981). The method uses a test statistic, the Mahalanobis Distance (MD; see Huba, 1985) which, when used with the WAIS—R, is distributed as chi square with 11 degrees of freedom. The MD can be viewed as a multivariate measure of the distance of a single observation (i.e. an individual's test score profile) from the centre of the population from which the observation is drawn. In the case of the WAIS—R, the MD can be viewed as an index of the probability that a given individual could have come from the standardization sample (Burgess, 1991). Where the aim is to detect and quantify cognitive impairment, the MD index has two features which potentially make it preferable to many alternative methods of analysing subtest scatter. Firstly, it is concerned with the *abnormality* of the subtest scatter rather than whether the subtests are reliably different from one another (reliable differences between subtests are the norm rather than the exception in the *healthy* population; see Matarazzo, Daniel,

* Requests for reprints.

Prifitera & Herman, 1988). Secondly, with 11 subtests, 55 subtest comparisons are possible. As we rarely specify subtest comparisons *a priori,* it is necessary to have some means of taking account of these multiple comparisons or very serious Type 1 errors could occur. Correction for multiple comparisons is inherent in the MD procedure as it examines whether the subtest scatter *taken as a whole* is abnormal or not.

Because the WAIS-R has not been standardized in the UK, Burgess (1991) noted that the index should be used with caution by UK clinicians. Such advice is more than justified because of the large number of population parameters involved in the calculation of MD scores. Uncritical use of the MD in the UK (e.g. accepting that a significant MD score does indeed represent abnormal scatter or that conversely a non-significant score indicates that the scatter is within normal limits) involves making a number of important and largely untested assumptions about the UK psychometric properties of the WAIS—R: e.g. that all the subtest scores are normally distributed with means of 10 and 3 respectively and that the subtest intercorrelations are equivalent to the published values for the US.

The present study is aimed at establishing the basic statistical properties of the WAIS—R MD index in a sample of the UK general adult population. Specifically, the sample distribution of MD scores will be examined to determine whether or not it deviates significantly from the chi square distribution for 11 degrees of freedom. The percentage of scores at the tail of the distribution (i.e. that exceed the critical values for the MD) will also be examined because in clinical practice the principal concern will be with whether a client's scatter is abnormal or not.

A subsidiary aim was to examine the relationship between MD scores and demographic variables (e.g. age, sex, educational level and social class). The relationship between these variables is of practical importance. For example, take the case of an elderly client in social class 1 whose MD score achieves statistical significance at the .05 level. This would indicate that the subtest scatter is abnormal relative to the general population (i.e. less than 5 per cent of the population would be expected to obtain as large an MD score). However, if age and social class correlate with the MD index then the client's score may not be atypical for individuals with these demographic characteristics. Clinical interpretation of MD scores would obviously be easier if the scores prove to be unrelated to demographic characteristics. However, if it transpires that this is not the case, evaluation of MD scores would still be possible provided the clinician had a knowledge of the direction and strength of these relationships.

Method

The sample consisted of 200 subjects (104 females, 96 males) free of neurological, psychiatric or sensory disorder. Most received a small honorarium for their participation. Mean age was 44.3 (SD = 19.2) with a range of 16 to 83 years and mean years of education was 12.6 (SD = 3.0). Each subject's social class was coded from their occupation using the Classification of Occupations (OPCS, 1980).

The proportions of the sample in each social class were compared with the census-derived proportions for the adult UK population using a goodness-of-fit chi square test. This revealed that the sample and population proportions did not differ significantly ($X^2(4) = 5.66$, 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. The numbers in each band were as follows (the census-derived expected numbers are in parentheses): 16-17 = 6 (8.9); 18-19 = 7 (8.5); 20-24 = 22 (19.3); 25-34 = 45 (37.3); 35-44 = 29 (31.6); 45-54 = 24 (29.6); 55-64 = 23 (29.1); 65-69 = 16 (13.0); 70+ =28 (22.6). A goodness-of-fit test revealed that the observed and expected numbers did not differ significantly ($X^2(8) = 7.71$, n.s.). Finally the sample's sex distribution was also representative of the adult UK population ($X^2(l) = .01$, n.s.).

All subjects were administered a full-length WAIS-R (UK) according to standard procedures (Lea, 1986; Wechsler, 1981). For each subject, the raw scores on the 11 subtests were converted to age-graded scaled scores. These scores were then entered into a PASCAL program* which calculated the MD scores.

Results

The samples' mean FSIQ was 102.5 (SD = 13.12) with a range from 71 to 140. The distribution of MD scores in the sample was divided into four bands corresponding to the quartiles of the chi square distribution for 11 degrees of freedom (source: Hays, 1973, pp. 886—887). The number of scores in each band (going from lower to upper quartile) was as follows: 51, 54, 36, and 59. Although there was an underrepresentation of scores in the 3rd quartile and an overrepresentation in the 4th quartile, a goodness-of-fit test comparing the observed frequencies with the expected frequencies (i.e. 50 scores in each band given 200 scores) revealed that they did not differ significantly ($X^2(3) = 5.88$, n.s.). Turning attention to the tail of the distribution, the percentage of subjects exceeding the critical values of chi square for the commonly employed significance levels was as follows (the relevant critical values are presented in parentheses); 0.1 level (17.28) = 12 per cent, 0.05 (19.68) = 6.5 per cent, 0.01 (24.73) = 6 per cent, and 0.01 (31.26) = 1 per cent. A goodness-of-fit test was performed to compare the number of MD scores above and below the .05 level (187 and 13 respectively) with the expected frequencies (190 and 10 respectively, given 200 scores). This revealed that the observed and expected frequencies did not differ significantly ($X^2(l) = 0.95$, n.s.). The .05 level was chosen for this analysis because it is the significance level conventionally adopted and because the expected frequencies would have been too small had the .01 or .001 levels been chosen.

Examination of scatterplots and attempts to fit curvilinear models to the data indicated that the relationship between the MD scores and demographic variables did not depart markedly from linearity. The Pearson product moment correlations between MD scores and the demographic variables were as follows: age $r = —.09$ (n.s.); sex $r = —.11$ (n.s.), males were coded as 0, females as 1; years of education $r = .21$ ($p < .01$); social class $r = —.24$ ($p < .01$). The correlation between MD and full-scale IQ was .19 ($p < .01$).

Discussion

The distribution of MD scores in the sample broadly conformed to the chi square distribution for 11 degrees of freedom. Although the number of scores in the 3rd and 4th quartiles noticeably deviated from the expected frequencies, the goodness-

* Burgess (1991) made available the listing of a BASIC program for the computation of MD scores. For those not wishing to type in this listing, compiled versions of the above PASCAL program for IBM PCs and compatibles or Apple Macs can be obtained free of charge from the first author (specify 3.5 or 5.25 disk).

of-fit test indicated that such deviations were within normal limits for a sample of this size. Therefore, there are no grounds to reject the null hypothesis that the present sample's scores and the standardization sample's scores were drawn from the same population.

For the practising clinician the principal concern will be with whether a client's MD score can be viewed as abnormal or not. Analysis of the scores at the tail of the distribution suggest that, if the .05 level is adopted as the significance level, the percentage of the normal UK population that exceed the critical value (an MD score of 19.7) will be very close to the desired 5 per cent (i.e. 6.5 per cent of the present sample had MD scores which exceeded this critical value, a discrepancy which is well within normal error limits). For more stringent significance levels the results were less satisfactory. Only 1 per cent of the sample should have obtained MD scores that exceed the critical value for the .01 level; in fact 6 per cent did so. Thus, if a UK client's MD score exceeds the critical value for the .01 level, it would be advisable at present to interpret and report this as only exceeding the .05 level.

Reasonable confidence can be placed in the present results because the sample employed had a large *N* and was representative of the UK population in terms of age, sex and social class distribution. However, a cross-validation attempt in another UK sample would be welcome.

Given that a large number of the WAIS—R's population parameters are incorporated into the calculation of MD scores, the present results can also be viewed as providing an indication of the general robustness of the WAIS-R's psychometric properties in the UK. This is reassuring for UK users of the test, regardless of whether or not they use the MD index.

The correlations between the MD index and age and sex failed to achieve significance whereas the correlations between education and social class were highly significant. However, it should be noted that, even in the case of these latter variables, the relationships were weak; the high level of significance achieved was simply a function of the large sample size. In the case of the highest correlation, that between the MD and social class, less than 6 per cent of the variance was shared. The lack of a strong relationship between the MD index and any of the demographic variables simplifies clinical interpretation considerably. However, a knowledge of the relationships that do exist could be of some use in cases where the MD scores are borderline. The direction of the correlations was such that a large number of years of education and high social class is associated with larger MD scores (as is a high full-scale IQ). Thus in a borderline case the presence of the above characteristics would argue against a conclusion that the subtest scatter is abnormal. At the risk of stating the obvious, if FSIQ is examined in borderline cases some estimate of the clients *premorbid* IQ should be used rather than their currently obtained IQ.

It should be stressed that the present study only indicates that it is permissible to use the MD index in UK clinical practice; it does not address the issue of the sensitivity of the MD index to impairment of cognitive abilities. Because the index has only recently been developed, studies addressing this latter issue have not yet been conducted. The present authors agree with Burgess's (1991) suggestion that the index is liable to be of greatest value when used in combination with other established WAIS-R indices rather than being employed as an alternative to them.

We would therefore suggest that any studies designed to examine the sensitivity of the index should not evaluate it in isolation.

Finally, because the WAIS-R has not been standardized in the UK, studies such as the present one are necessary to determine the applicability of any procedures used to interpret WAIS—R test scores. However, such studies can never be as satisfactory as obtaining full UK standardization data. The WAIS—R is probably the most widely used cognitive measure in clinical practice. For example, a recent report from the EPS has recommended that the WAIS-R be used as the sole criterion for arriving at a legal definition of'mental impairment' (EPS, 1991). It is therefore to be hoped that the WAIS—R will eventually be standardized here. Were this to happen, indices such as the MD could be derived directly from the UK estimates of the relevant population parameters. The accurate detection and quantification of cognitive dysfunction in the individual case is problematic enough without the additional difficulties posed by using a test standardized elsewhere.

Acknowledgements

This research was partly supported by a grant from the UK Mental Health Foundation.

References

British Psychological Society (1991). Mental impairment and severe mental impairment. *The Psychologist,* 4, 373-376. Burgess, A. (1991). Profile analysis of the Wechsler Intelligence Scales: A new index of subtest scatter.

British journal of Clinical Psychology, 30, 257—263.

Hays, W. L. (1973). *Statistics for the Social Sciences,* 2nd ed. London: Holt, Rinehart & Winston.

Huba, G. J. (1985). How unusual is a profile of test scores? *journal of Psychoeducational Assessment,* 4, 321-325. Lea, M. (1986). *A British Supplement to the Manual of the Wechsler Adult Intelligence Scale-Revised.* San

Antonio: Psychological Corporation. Matarazzo, J. D., Daniel, M. H., Prifitera, A. & Herman, D. O. (1988). Inter-subtest scatter in the

WAIS-R standardisation sample, *journal of Clinical Psychology,* 44, 940-950.

Office of Population Censuses and Surveys (1980). Classification of Occupations. London: HMSO.

Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale-Revised.* New York: Psychological Corporation.

Received 3 October 1991; revised version received 23 September 1992