Journal of Clinical and Experimental Neuropsychology
2000, Vol. 22, No. 3, pp. 305-315

1380-3395/00/2203-305$15.00

Criterion Validation of Premorbid Intelligence Estimation in Persons with Traumatic Brain Injury: ‘‘Hold/Don’t Hold’’ versus ‘‘Best Performance’’ Procedures

Dan Hoofien[1,3], Eli Vakil[2], and Assaf Gilboa[3]
[1]The National Institute for the Rehabilitation of the Brain Injured, Tel Aviv, Israel, [2]Bar Ilan University, Ramat Gan, Israel, [3]The Hebrew University of Jerusalem, Israel

ABSTRACT

The goal of the present study was to validate previously suggested regression equations for the estimation of premorbid ability against a real premorbid intellectual criterion. Fifty-four patients with traumatic brain injuries, for whom a premorbid military Primary Psychometric Rating (PPR) was available, participated in the study. Two prediction procedures were validated: (a) ‘‘BEST-10’’, which generates a predicted score from the highest observed score of 10 WAIS-R subtests, according to the ‘‘best performance’’ estimation principle. (b) ‘‘BEST-2’’, which generates the predicted score from the higher of two subtests considered a priori resistant to neurological damage according to the ‘‘hold/don’t hold’’ principle. The two procedures showed similar correlation with the premorbid criterion. However, the BEST-10 method provided a more accurate estimation, generating a non-significant 2-point underestimation. The results support the application of previously proposed equations for estimating premorbid intelligence, and suggest that the use of the best performance principle is preferable as compared to the hold/don’t hold principle.

The estimation of the premorbid level of intellectual functioning is an essential component in the evaluation of cognitive decline suffered by survivors of traumatic brain injury (TBI). Real data on premorbid intelligence can sometimes be obtained from patients’ scores on standardized tests taken before the injury. These may include previously taken intelligence tests, various standardized aptitude tests or military selection examinations (Vanderploeg, 1994). However, such information is only rarely available to clinicians, thus requiring an alternative procedure for estimating patients’ premorbid level of intellectual functioning. Commonly used methods include: (a) the ‘‘Best performance method’’ (Lezak, 1995), which employs the highest level of functioning in test performance, academic or vocational achievement as an indicator for premorbid ability; (b) tests considered to ‘‘hold up’’ after TBI, such as the Information, Vocabulary and Picture Completion subtests of the Wechsler Adult Intelligence Scale – Revised (WAIS-R; Wechsler, 1981); and (c) estimates of premorbid intelligence based on regression equations of demographic variables, such as the Wilson Index for the WAIS (Wilson et al., 1978) or the Barona Index for the WAIS-R (Barona, Reynolds, & Chastain, 1984).

A review of the literature reveals that the latter two methods, or combinations thereof, are

---

Address correspondence to: Dan Hoofien, Department of Psychology, The Hebrew University of Jerusalem, Mount Scopus, Jerusalem 91905, Israel. E-mail: mshoofi@mscc.huji.ac.il
Accepted for publication: October 8, 1999.

the focus of research at present (for detailed reviews, see Crawford, 1992; Vanderploeg, 1994). One such combination recently suggested is the use of both demographic variables and current performance on hold measures in order to yield an improved regression equation (Crawford, Stewart, Parker, Besson, & Cochrane, 1989; Krull, Scott, & Sherer 1995). Compared to demographically based equations, these equations account for about double the variance of the actual IQ scores.

Nevertheless, the basic idea of combining demographic data with current performance in tests considered to hold up has a potential impediment in cases where specific brain damage does affect the cognitive functioning upon which the estimates depend (Vanderploeg & Schinka, 1995). To overcome this obstacle, Krull et al. (1995) suggested implementing a best performance criterion as part of their hold method for clinical populations. This procedure was later used with clinical samples (Scott, Krull, Williamson, Russel, & Iverson, 1997), combining elements from all three procedures: demographics, current hold measures and the addition of a best performance selection among these hold measures.

Within a similar conceptual framework, Vanderploeg and Schinka (1995) developed 33 regression formulae based on each of the 11 WAIS-R subtests for the estimation of the Verbal Intelligence Quotient (VIQ), the Performance Intelligence Quotient (PIQ) and the Full Scale Intelligence Quotient (FSIQ). Their reasoning was that performance on any of the WAIS-R subtests might be impaired following brain injury and no subtest was considered a priori to be a hold measure. They provide several guidelines as to which subtest is likely to remain the hold measure for different etiologies, and leave the choice of equation to the clinician.

Vanderploeg, Schinka, and Axelrod (1996) attempted to provide an alternative to the subjectivity inherent in clinical judgments. In a validation study of the 11 FSIQ regression equations they introduced two new standardized methods of choice among the equations. In the first method, the highest score of all 11 equations was used to generate one estimate of FSIQ (BEST-11), whereby the best performance was considered to simultaneously represent the hold ability. In the second method an estimate was generated using the best of three a priori determined hold measures (Information, Vocabulary and Picture Completion subtests) that are also known to exhibit high reliability coefficients (BEST-3), therefore using hold measures in a best performance manner. Applying the procedures to the WAIS-R standardization sample produced similar high correlation with the actual IQ scores (.84 and .85 for BEST-3 and BEST-11, respectively). However, the BEST-11 overestimated the FSIQ by about 10 points, while the BEST-3 generated a mean overestimation of only 5 IQ points.

Indeed, one of the main criticisms concerning the use of the best performance method is its substantial overestimation of real premorbid ability (Mortensen, Gade, & Reinich, 1991). Rebutting this criticism, Lezak (1995) points to the lack of a real criterion related validation (i.e. an actual index of premorbid ability). Most studies examining the predictive accuracy of the various equations use neurologically intact populations and concurrent test data as predictor and predicted variables. Following Lezak's comments, the problem with such cross-validation is two-fold. On the one hand it provides the *concurrent validity* (Crocker & Algina, 1986), whereas one is really interested in the *predictive validity* of the estimates, that is, the degree to which current information predicts criterion measurements at a different (prior) point in time. On the other hand it uses current test data as predictor variables, while at least some intellectual deficits are expected in the brain-injured but not in the neurologically intact.

Relatively few studies have attempted to test the validity of estimation methods on clinical populations. To the best of our knowledge all of them used concurrent test data, testing the concurrent but not the predictive validity of the procedures.

One problem with concurrent validation used with clinical samples is that due to the cognitive decline, the obtained FSIQ will always be lower than the predicted one. Two methods have been used in an attempt to by-pass this problem. In the first method, the means, standard deviations and ranges generated by the prediction equations are compared with those of the original WAIS-R standardization sample (Scott et al., 1997). A resemblance is considered indicative of the accuracy of the method. The second method involves comparison of correlations, wherein the accuracy of the estimation methods is measured in terms of the correlation between the predicted and obtained scores. This method was also applied by comparison to the WAIS-R results obtained from a matched sample of normal controls (Mortensen et al., 1991; Vanderploeg et al., 1996). A significant difference between predicted and obtained IQ scores only for the brain-injured group was interpreted by Vanderploeg et al. (1996) as an indication of the validity of the estimation methods. A further analysis tested the correlation between discrepancy scores (predicted minus obtained) and group membership, utilizing the matched-sample procedure. This figure was used for rebutting the criticism that the high correlations may merely be an indication of the relationship between the predicted scores and current test performance on which they are based.

Apparently Vanderploeg et al. (1996) met the demand for validation of the predicting methods against a real criterion of clinical group membership (controls vs. TBI) in a between-subjects design. However, as already mentioned, a complete validation of these methods requires comparison of observed versus predicted ability scores in a within-subject follow-up design. Such a validation corresponds better to the clinical application of the estimation procedures and takes full account of temporal as well as clinical dimensions of the validity criteria.

The purpose of the present study was to validate the method of previously studied regression equations against a real psychometric premorbid intellectual criterion in a within-subject clinical follow-up design. Our sample of army veteran patients with TBI, drawn from a large long-term follow-up study (Hoofien, Vakil, Donovick, & Rolnick, 1993), is unique in that it enabled us to test the criterion-based predictive validity of those estimates. A premorbid military Primary Psychometric Rating (PPR) was available for all of the subjects, allowing a close approximation of subjects' premorbid intelligence. Two estimation methods were studied to examine the difference between the best performance method with no a priori assumptions, and the best performance combined with a hold method. Due to the cognitive decline caused by the brain injury, we assumed that both procedures, when validated against a real premorbid criterion, would generate a much smaller overestimation than previously reported.

METHOD

Subjects

Fifty-four army veterans (7 females, 47 males) participated in the study. This sample was drawn from a larger sample (*n* = 99) of persons with severe TBI who participated in an extensive long-term neuropsychological follow-up study. Inclusion in the present study was based on the availability of data from the army's pre-enlistment screening, which all participants passed at the age of 18 (time 1, before the injury). This information was obtained by informed consent of the patient and by the army's special permit.

At the time of the current follow-up (time 2, 1993) the subjects were 39.5 years old on average (*SD* = 8.45) and 13.48 years post-injury on average (*SD* = 5.25). Forty-seven (87%) of the participants were right handed. The veterans had an average of 12.0 years of premorbid education (*SD* = 2.55); 15 (26%) of them reported additional education after the injury, that averaged 3.4 years (*SD* = 2.11; Range 1-7 years).

Forty-three (79.6%) of the individuals participated in post-acute neuropsychological rehabilitation programs. Out of the remaining 11 veterans, 6 reported to having participated in vocational rehabilitation programs and 5 did not report having been involved in any formal rehabilitation program. No significant difference in WAIS-R FSIQ was found between the group of 43 neuropsychologically treated individuals and the group of 11 untreated veterans.

Thirty-two (59%) participants had cerebral cranial injury; 17 (32%) closed head injuries; 4 (7%) cerebral vascular accidents; and 1 (2%) anoxia. Forty-seven (88%) had diffuse or multifocal damage; 2 (4%) right hemisphere damage; 2 (4%) left hemisphere damage; and 3 (6%) frontal lobe damage. Sixteen (30%) participants were comatose for more than 30 days; 15 (28%) between 1 week to 1 month; 13 (24%) between 1 day to 1 week; 7 (13%) less than 1 day and 3 (5%) had no history of coma at all. None of the veterans had a pre-injury history of psychiatric symptoms, drug addiction or neurological disease.

At the time of the follow-up, the intellectual functioning of the participants was within the low average range (Table 1). Memory and learning were evaluated by the Hebrew version of the Logical Memory and the Visual Reproduction subtests of the Wechsler Memory Scale – Revised (WMS-R, Wechsler, 1987), and the Hebrew version of the Rey Auditory Learning Test (Rey AVLT, Vakil & Blachstein, 1993). The participants performance on these tests ranged between average and low-average levels.

Tests and Procedures

Premorbid ability (Time 1)

Premorbid data was obtained from the participants' army files after informed consent. At the age of eighteen, as part of the standard pre-enlistment screening for compulsory military service, each recruit is assigned a PPR score (Gal, 1986). This composite score is derived from the conscript's performance on the Raven's Standard Progressive Matrices and on an Otis-type verbal test (a version of the Army Alpha Test). Both tests are highly correlated with the WAIS and WAIS-R scores and have been previously used as intelligence tests. Swiercinsky and Patterson (1978), for example, report a correlation of .81 between full scale WAIS IQ and Army Alpha predicted IQ scores. Similarly, O'Leary, Kathleen, and Guastello (1991) found high correlations between WAIS-R scores and Raven's Standard Progressive Matrices. They report age-stratified correlations of .74-79 for the 16-44 group. The PPR is considered a highly valid measurement of intelligence (Reeb, 1976; Davidson et al., in press). As part of its construction it was validated against the WAIS FSIQ scores of a large standardization sample and then was standardized to match the mean and standard deviation of the WAIS. The resulting scale of the PPR score ranges from 10 to 90, with a structured mean of 50 and a standard deviation of 20 points, with 90 being equal to an IQ of 135 and above. As all conscripts are drafted at the same age, no age correction had to be used (Gal, 1986). For the present study these scores were linearly transformed to match the mean and standard deviation of the WAIS-R. Thus, each participant was assigned a PPR transformed IQ score, based on his or her PPR score. These transformed scores of the sample ranged from 77.5 to 130 points, with a mean of 108.67 (*SD* = 13.07).

Current intelligence and demographic data (Time 2)

An extensive battery of neuropsychological tests, interviews and questionnaires was administered as part of a long-term neuropsychological follow-up study (Hoofien et al., 1993). In the present study, only the results of the WAIS-R (Hebrew version) and data from a demographic questionnaire specifically designed for the study were analyzed.

Based on the results of 10 of the 11 WAIS-R subtests and on the personal demographic data, each subject was assigned 10 predicted FSIQ scores. These scores were generated using Vanderploeg and Schinka's (1995) equations (see Appendix). Due to translation difficulties, the Hebrew version of the WAIS-R does not include the Vocabulary subtest. The Verbal IQ was calculated by multiplying by 6/5 the sum of the other five subtests. Hence, no FSIQ predicted score was available for this subtest.

Of the predicted scores, two final scores were

Table 1. Means and Standard Deviations of WAIS-R IQ Scores and WMS-R and Rey AVLT Percentile Ranks.

Test		*Mean*	(*SD*)
WAIS-R	FSIQ	91.29	(12.70)
	PIQ	87.45	(12.54)
	VIQ	95.43	(14.14)
WMS-R[a]	LM Immediate recall	42.66	(29.28)
	LM Delayed recall	38.80	(26.74)
	VR Immediate recall	52.28	(32.35)
	VR Delayed recall	36.50	(33.21)
Rey AVLT[a]	Total learning	14.21	(22.49)
	Delay trial	12.54	(21.15)

Note. LM = Logical Memory; VR = Visual Reproduction. Total Learning = sum of words recalled on trials 1-5.
[a] Age-corrected percentile rank score.

derived for each participant: the BEST-10 score, based on the highest of all 10 predicted scores and the BEST-2 score, based on the higher of the Information and Picture Completion predicted scores. These two scores are the equivalents of Vanderploeg et al.'s (1996) BEST-11 and BEST-3 scores, respectively.

The demographic variables used were age, gender, race, premorbid occupation and premorbid education. The last two variables were combined in the same manner as proposed by Vanderploeg and Schinka (1995) to form one variable, premorbid socioeconomic status (SES). In this procedure, the WAIS-R manual's occupational level coding was reversed to form a new coding which was then summed together with the educational level coding. For the subjects who were unemployed, the coded education value was used as an estimate of occupational level. This transformation was found particularly useful for our sample, as a large portion of our participants were either soldiers or high school students prior to their injury. The number of years of education before the injury was used for coding educational level (Vanderploeg and Schinka, 1995). In an attempt to estimate premorbid, as opposed to current abilities, additional postmorbid occupational and educational achievements were not included. A detailed description of the demographic variables used to generate the predicted FSIQ for each participant is provided in the Appendix.

RESULTS

To compare the predictive validity of the BEST-2 and BEST-10 methods, Pearson correlations were calculated between the premorbid PPR transformed IQ and three current postmorbid intelligence parameters: obtained WAIS-R FSIQ (time 2), BEST-2 estimated FSIQ and BEST-10 estimated FSIQ. These correlations are presented in the upper row of Table 2.

The obtained WAIS-R FSIQ (time 2), the BEST-2 and BEST-10 estimates had very similar and statistically significant correlations with premorbid PPR transformed IQ. There was no significant difference between the correlations of the two estimation methods with the PPR score (t (46) = .80, $p > .1$). Nor were significant differences found between the correlations of the post-injury WAIS-R score and of the BEST-2 and BEST-10 estimates with the PPR score (t (46) = 0.79, $p > .1$ and t (46) = 0.10, $p > .1$, respectively).

The concurrent validity of the estimation procedures was also examined by calculating Pearson correlations between postmorbid-obtained WAIS-R FSIQ (time 2) and the two estimations (Table 2, bottom row). As expected, these correlations were higher than the predictive validity correlations and were statistically significant.

Further analysis was conducted to compare the predictive accuracy of the two methods. Paired differences between premorbid PPR transformed IQ and (a) postmorbid WAIS-R FSIQ (time 2), (b) BEST-2 estimated FSIQ, and (c) BEST-10 estimated FSIQ were calculated. The means of the paired differences are presented in Table 3, as well as the results of paired t tests and 95% confidence interval for the mean IQ points difference for each of the pairs.

A significant difference of 19.04 IQ points was found between the premorbid PPR transformed IQ and the postmorbid WAIS-R FSIQs ($p < .001$). With regard to the predictive accuracy of the BEST-2 and BEST-10 estimation methods, the findings show a significant difference of 5.39 IQ points between premorbid PPR transformed IQ and the BEST-2 estimated FSIQ ($p < .002$). There is a non-significant difference of 2.07 IQ points between premorbid PPR trans-

Table 2. Correlations Between Premorbid PPR Transformed IQ, Current WAIS-R FSIQ (Time 2) and the BEST-10 and BEST-2 Estimates.

	Postmorbid WAIS-R FSIQ	BEST-2 Estimated FSIQ	BEST-10 Estimated FSIQ
Premorbid PPR Transformed IQ	.628**	.583**	.622**
Postmorbid WAIS-R FSIQ	–	.858**	.850**

**$p < .01$ (2-tailed).

Table 3. Paired Samples T-Test Between the Means of the Differences in IQ Points Between Premorbid PPR Transformed IQ and Obtained WAIS-R FSIQ (time 2), BEST-10 and BEST-2 Estimated FSIQs.

	Premorbid PPR Transformed IQ		Postmorbid WAIS-R FSIQ		BEST-2 Estimated FSIQ		BEST-10 Estimated FSIQ	
	M	(*SD*)	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)
Mean (SD)	109.11	(12.92)	91.29	(12.70)	104.73	(11.89)	107.84	(10.41)
IQ Point Difference From PPR			–19.04	(11.27)	–5.39	(11.47)	–2.07	(10.48)
95% Confidence Interval of the Difference			[–22.43]-[–15.66]		[–8.68]-[–2.09]		[–5.05]-[0.91]	
t value			(df = 44) –11.34		(df = 48) –3.29		(df = 49) –1.40	
p value (2-tailed)			.001		.002		.169	

formed IQ and the BEST-10 estimated FSIQ (*p* < .169). The accuracy of the estimate procedures is also evident from the 95% confidence interval. The upper boundary for the IQ point difference between estimated and obtained premorbid IQ is 0.91 for the BEST-10 procedure and –2.09 for the BEST-2 procedure.

The frequencies of absolute IQ point differences from the premorbid PPR transformed IQ for the postmorbid WAIS-R FSIQ and for the two predicted BEST-2 and BEST-10 IQ's are presented in Table 4. Frequencies, percentages and cumulative percentages of subjects falling within each range of IQ point difference are presented and may be utilized as an indication for the probability of an estimated score to fall within certain ranges of the actual score.

This frequency distribution gives another indication for the relative predictive accuracy of the estimation methods. The difference in accuracy between the BEST-10 and the BEST-2 method is again manifested. For example, in the BEST-10 estimation procedure 84% of the patients fell within one standard deviation of their premorbid PPR transformed IQ, compared to 73.5% for the BEST-2. For the postmorbid obtained WAIS-R FSIQ, only 35.6% of the patients fell within that range, once again reflecting the postmorbid intellectual compromise.

DISCUSSION

Following Vanderploeg et al.'s (1996) findings and Lezak's (1995) critical recommendations, the present study examined the predictive validity of premorbid intelligence estimation procedures. To overcome the inherent weaknesses of previous concurrent between-subjects validity findings, we validated the procedures against an actual premorbid criterion, in a within-subject, long-term follow-up design.

Two methods, representing two conceptual frames of reference were studied: (a) the BEST-2 procedure in which the one and the best of two subtests considered a priori resistant to neurological damage was used as the predictor and (b) the BEST-10 procedure where no such a priori assumption was made. The first (BEST-2) relies conceptually on the idea of hold/don't hold (Wechsler 1953; Lezak, 1995) predetermined ''neurologically resistant'' abilities. The second (BEST-10) relies more on the concept of best performance (Lezak, 1995), with no predetermined assumptions regarding the degree of resistance of specific abilities.

Table 4. Frequency Distribution of Absolute IQ Point Difference Between Premorbid PPR Transformed IQ and Postmorbid WAIS-R FSIQ, BEST-2 and BEST-10 Estimated FSIQ

Absolute IQ Point Difference From Premorbid PPR Transformed IQ.		Postmorbid WAIS-R FSIQ	BEST-2 Estimated FSIQ	BEST-10 Estimated FSIQ
0-5	Frequency	4	15	18
	%	8.9	30.6	36.0
	Cumulative %	8.9	30.6	36.0
5-10	Frequency	6	13	12
	%	13.3	26.5	24.0
	Cumulative %	22.2	57.1	60.0
10-15	Frequency	6	8	12
	%	13.3	16.3	24.0
	Cumulative %	35.6	73.5	84.0
15-20	Frequency	5	7	6
	%	11.1	14.3	12.0
	Cumulative %	46.7	87.8	96.0
20+	Frequency	24	6	2
	%	53.3	12.2	4.0
	Cumulative %	100.0	100.0	100.0

The results provide evidence for the predictive validity of both the BEST-2 and BEST-10 estimation procedures. However, the BEST-10 procedure was found to have higher predictive accuracy, lending support to the best performance approach with no a priori hold assumptions.

Our findings also provide evidence for the concurrent validity of the two methods. The high correlations between postmorbid WAIS-R FSIQ and the results of our two estimation procedures are almost identical to the correlations of .85 and .84 between the BEST-11 and BEST-3 formulas, respectively, reported by Vanderploeg et al. (1996). Such high correlations are to be expected, since the WAIS-R FSIQ in this analysis is concurrently predicted by a regression formula in which one of its subtests is used as a predictor.

However, our findings of the predictive validity correlations between premorbid PPR transformed IQ and either the postmorbid WAIS-R FSIQ or the two predicted IQs, are not as high. This may be the result of several factors. The first is the nature of the premorbid criterion – the PPR transformed IQ – which, as mentioned above, is very highly correlated with the WAIS-R FSIQ, but is nevertheless not the same test. A second possible factor is related more to the essence of premorbid intelligence estimation. An underlying assumption inherent in the two estimation methods is that intelligence, as a human attribute, is relatively reliable and stable over time. That is, an individual who has an average IQ at age 18 is expected to have the same average range of intelligence 15 and 20 years later. Although in general this assumption is true, it is also true that the correlation between two identical measurements of intelligence across lengthy gaps in time is never a full correlation.

In a meta-analysis, Schuerger and Witt (1989) examined the temporal stability of five intelligence tests, based on test-retest reliabilities. Two important correlates of test stability were identified, namely the time interval between testing and the age at the time of the first testing. These two variables accounted for more than 50% of the variance of the test-retest reliability. The overall trend was a decrease in reliability the longer the time interval, and an increase in reliability with progressing age.

The authors provide a table of expected stability of intelligence scores for age groups and test-retest intervals. For our sample, that was first tested around the age of 18 and reevaluated an average of 12 years later, the expected stability of the intelligence scores would be .79. According to their data 10% of the sample would be expected to have a time 1 to time 2 difference of 15 IQ points or more. This test-retest reliability coefficient may be regarded as the uppermost limit of criterion validation that may be expected. Hence, within this limit we found a long-term predictive validity of $r = .628$ for the BEST-10 method. Therefore, it seems that the previously reported high validity correlations, with the WAIS-R standardization sample as concurrent criterion for the generation of the formulas, should be considered cautiously. Besides being validated against a concurrent criterion, the correlations do not take into account the important influence of time.

Analysis of the predictive accuracy of estimation methods is an important clinical derivative of examining their validity. The advantage of testing predictive accuracy against an actual premorbid measure reflecting the cognitive decline of a clinical sample of real patients is even more apparent here. We found that both prediction models resulted in some underestimation of premorbid PPR transformed IQ scores. The BEST-2 method resulted in a significant underestimation by 5.39 points on the average. On the other hand, the BEST-10 generated a non-significant average underestimation of 2.07 IQ points. The same trend is also apparent in the probability distribution of the differences between estimated and obtained scores. For the BEST-2 procedure 26% of the subjects were assigned estimated IQ scores that were more than one standard deviation apart from their PPR transformed IQ score. For the BEST-10 procedure, only 16% were estimated to have FSIQs beyond that range. As already mentioned, Schuerger and Witt's (1989) table of expected stability assigns to 10% of our sample an expected difference of 15 IQ points as a result of their age and the test-retest interval.

These results are quite different from those reported by Vanderploeg et al. (1996). They found overestimations of approximately 5 and 10 points using the BEST-3 and BEST-11 methods, respectively, on the WAIS-R standardization sample. Here again, the disadvantage of validating these procedures on a neurologically intact sample is evident. Indeed, when they applied the same procedures of concurrent validation to a clinical sample, the average underestimation shrunk to 7 FSIQ points, even though this, too, was not a full within-subject follow-up.

Although more accurate, the best performance method (or BEST-11 for that matter) applied in a mechanical manner has the potential of generating erroneous predictions in cases where certain skills or isolated cognitive abilities lead to an unjustified high prediction. The inclusion of demographic data in the equation may have a balancing effect in such cases. It is again recommended to use clinical judgment in applying the formulas (Lezak, 1995). Judging from the probability distribution presented earlier, such cases are the exception rather than the rule. Hence, only when sufficient evidence of such a case is present should the use of the best result be rejected.

In summary, the addition of current performance data to demographic regression equations had been found to improve the accuracy of the prediction (Crawford et al., 1989; Krull et al., 1995; Vanderploeg et al., 1996; Scott et al., 1997). Methodological difficulties, however, left open the question of predictive validity and the accuracy of estimations. Our results indicate that a more accurate estimate of premorbid cognitive ability is generated by a purely best performance decision criterion with no a priori assumptions. Combined with demographic data, this method generates an estimate that is specific to each individual with a negligible, if any, systematic bias.

A methodological shortcoming of the present study is the absence of the prediction equation based on the Vocabulary subtest. Although the remaining Best-2 subtests (Information and Picture Completion) represent both verbal and performance aspects of intelligence, it is still recommended that our results be regarded as indicative of the difference between the prediction methods. A replication of the within-subject paradigm with all 11 subtests as generators of the equations would have more decisive power.

REFERENCES

Barona, A., Reynolds, C. R., & Chastain, R. (1984). A demographically based index of premorbid intelligence for the WAIS-R. *Journal of Consulting and Clinical Psychology, 52,* 885-887.

Crawford, J. R., (1992). Current and premorbid intelligence measures in neuropsychological assessment. In: Crawford, J. R., Parker, D. M. & McKinlay, W. W. (Eds.) *A Handbook of Neuropsychological Assessment* (pp. 21-49). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Crawford, J. R., Stewart, R. T., Parker, D. M., Besson, J. A. O., & Cochrane, R. H. B. (1989). Estimation of premorbid intelligence: Combining psychometric and demographic approaches improves predictive accuracy. *Personality and Individual Differences, 10,* 793-796.

Crocker, L. & Algina, J. (1986). *Introduction to Classical and Modern Test Theory.* (pp. 217-263). New York: Holt, Reinhart and Winston.

Davidson, M. Reichenberg, A. Rabinowitz, J. Weiser, M. Kaplan, Z.& Mark M. (1999). Cognitive and behavioral markers for schizophrenia in a population of apparently healthy male adolescents. *American Journal of Psychiatry*, In press.

Gal, R. (1986) *A Portrait of the Israeli Soldier.* (pp. 76-96). New York: Greenwood Press.

Hoofien, D., Vakil, E., Donovick, P., and Rolnick, A. (1993) *Final report of a comprehensive long term follow-up study on persons with brain injuries* (Hebrew). Israel: The National Institute for the Rehabilitation of the Brain Injured.

Krull, K. R., Scott, J. G., & Sherer, M. (1995). Estimation of premorbid intelligence from combined performance and demographic variables. *The Clinical Neuropsychologist. 9 (1),* 83-88.

Lezak, M. D. (1995). *Neuropsychological Assessment.* Oxford: Oxford University Press.

Mortensen, E. L., Gade, A. & Reinich, J. M. (1991). Best Performance Method in clinical neuropsychology. *Journal of Clinical and Experimental Neuropsychology, 13,* 361-371.

O'Leary, U. M., Kathleen, K. M., & Guastello, S. J. (1991). Estimating age-stratified WAIS-R IQs from scores on the Raven's Standard Progressive Matrices. *Journal of Clinical Psychology. 47 (2),* 277-284.

Reeb, M. (1976). Differential test validity for ethnic groups in the Israeli army and the effects of educational level. *Journal of Applied Psychology. 61 (3*), 253-261.

Schmand, B., Geerlings, M. I., Jonker, C., & Lindeboom, J. (1998). Reading ability as an estimator of premorbid intelligence: Does it remain stable in emergent dementia? *Journal of Clinical and Experimental Neuropsychology. 20 (1),* 42.

Schuerger, J. M., & Witt, A. C. (1989). The temporal stability of individually tested intelligence. *Journal of Clinical Psychology. 45 (2),* 294-30.

Scott, J. G., Krull, K. R., Williamson, D. J. G. , Russell, L. A., & Iverson, G. L. (1997). Oklahoma premorbid Intelligence Estimation (OPIE): utilization in clinical samples. *The Clinical Neuropsychologist. 11 (2),* 146-154.

Swiercinsky, D. P., & Patterson, T. W. (1978). The revised army alpha examination as a predictor of intelligence and brain impairment. *Psychological Report, 42,* 601-602.

Vakil, E., & Blachstein, H. (1993). Rey Auditory Verbal Learning Test: Structure analysis. *Journal of Clinical Psychology, 50,* 240-245.

Vanderploeg, R. D. (1994). Estimating premorbid level of functioning. In R. D. Vanderploeg (Ed.), *Clinician's guide to neuropsychological assessment.* New Jersey: Lawrence Erlbaum Associates.

Vanderploeg, R. D., Schinka, J. A., & Axelrod, B.N. (1996). Estimation of WAIS-R premorbid intelligence: current ability and demographic data used in a best-performance fashion. *Psychological Assessment, 8 (4),* 404-411.

Vanderploeg, R. D., & Schinka, J. A. (1995). Predicting WAIS-R IQ premorbid ability: combining subtest performance and demographic variable predictors. *Archives of Clinical Neuropsychology, 10 (3),* 225-239.

Wechsler, D. (1958) *The measurement and appraisal of adult intelligence* (4th ed.). Baltimore: Williams and Wilkins.

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

Wechsler, D. (1987) *Manual for the Wechsler Memory Scale-Revised (WMS-R)* San Antonio, TX: The Psychological Corporation.

Wilson, R. S., Rosenbaum, G., Brown, G., Rourke, D., Whitman, D.,& Grisell, J. (1978). An index of premorbid intelligence. *Journal of Consulting and Clinical Psychology, 46,* 1554-1555.

Appendix A

Vanderploeg and Schinka (1995) FSIQ estimation worksheet.

$\text{WAIS-R FSIQ} = 3.55(\text{Info}) + 1.00(\text{SES}) + 58.70$
$\text{WAIS-R FSIQ} = 2.56(\text{Dig Spn}) + 2.11(\text{SES}) + 59.39$
$\text{WAIS-R FSIQ} = 3.78(\text{Vocab}) + 0.70(\text{SES}) + 59.09$*
$\text{WAIS-R FSIQ} = 3.28(\text{Arith}) + 1.39(\text{SES}) + 58.32$
$\text{WAIS-R FSIQ} = 3.31(\text{Compre}) + 1.14(\text{SES}) + 59.60$
$\text{WAIS-R FSIQ} = 3.08(\text{Simil}) + 1.23(\text{SES}) + 0.84(\text{AGE}) + 5.61(\text{RACE}) + 53.60$

$\text{WAIS-R FSIQ} = 2.94(\text{Pic Com}) + 2.13(\text{SES}) + 1.62(\text{AGE}) + 49.41$
$\text{WAIS-R FSIQ} = 2.61(\text{Pic Arr}) + 2.17(\text{SES}) + 1.56(\text{AGE}) + 7.00(\text{RACE}) + 46.60$
$\text{WAIS-R FSIQ} = 3.20(\text{Blk Dsgn}) + 2.00(\text{SES}) + 1.81(\text{AGE}) + 47.62$
$\text{WAIS-R FSIQ} = 2.69(\text{Obj Asm}) + 2.58(\text{SES}) + 1.59(\text{AGE}) + 48.61$
$\text{WAIS-R FSIQ} = 2.21(\text{SES}) + 2.44(\text{Dig Sym}) + 2.16(\text{AGE}) - 4.38(\text{SEX}) + 58.17$

* This equation was not used in the present study, as there is no Vocabulary subtest in the Hebrew version of the WAIS-R.

Note: For each subject the highest predicted score was used to form the BEST-10 score. The highest of Information and Picture Completion was used to form the BEST-2 score.

Coding Variables

Sex	Male = 1, Female = 2
Race*	White = 1, Other ethnicity = 0
Age	16-17 years = 1, 18-19 = 2, 20-24 = 3, 25-34 = 4, 35-44 = 5, 45-54 = 6, 55-64 = 7, 65-69 = 8, 70-74 = 9
Education	0-7 years = 1, 8 = 2, 9-11 = 3, 12 = 4, 13-15 = 5, 16+ = 6.
Occupation	Unemployed = 1; Farm Laborers, Farm Foremen, & Laborers (unskilled) = 2; Operatives, Service Workers, Farmers & Farm Managers (semiskilled) = 3; Craftsmen, & Foremen (skilled workers) = 4; Managers, Officials, Proprietors, Clerical & Sales Workers = 5; Professional & Technical = 6
SES	Sum of Education Code and Occupation Code (if unemployed, SES = 2 X Education)

* Due to Israeli demographic characteristics, the following coding was used in the present study: European Origin = 1 , Other Origin = 0.

Note. Reprinted from: *Archives of Neuropsychology, Vol. 110*, *No. 3*, Vanderploeg R.D. and Schinka, J.A., Predicting WAIS-R IQ premorbid ability: Combining subtest performance and demographic variable predictors. Pp. 232, Copyright © 1995, with permission from Elsevier Science.