Comparing Methods for Estimating Premorbid Intellectual Functioning

Bradley N. Axelrod

Department of Veterans Affairs Medical Center

Rodney D. Vanderploeg and John A. Schinka

James A. Haley Veterans Hospital

The interpretation of data from a neuropsychological evaluation is based, in part, on examining performance strengths and weaknesses across measures of cognitive skills, especially as they relate to estimates of premorbid ability. There are numerous premorbid estimation approaches advocated in the literature, almost all of which attempt to estimate premorbid Full Scale Intelligence Quotient (FSIQ) for the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955) and its revised version (WAIS-R; Wechsler, 1981). These include: (a) demographic-based regression approaches (e.g., Barona, Reynolds, & Chastain, 1984; Wilson, Rosenbaum, Brown, Rourke, Whitman, & Grisell, 1978); (b) current ability (“hold”) approaches based on independent reading tests (Blair & Spreen, 1989; Berry et al., 1994; Grober, Sliwinski, Schwartz, & Saffran, 1989; Nelson, 1982; Nel-
son & McKenna, 1975; Nelson & O’Connell, 1978; Schwartz & Saffran, 1987), reading subtests from academic achievement batteries (Ritchie, Lam, & Rankin, 1996; Wiens, Bryan, & Crossen, 1993), word-identification tasks, such as the Peabody Picture Vocabulary Test-Revised (PPVT-R, Dunn & Dunn, 1981; Basso, Bielaiskas, & Roper, 1994); or specific Wechsler intelligence scale subtests (McFie, 1975; Wechsler, 1958; Yates, 1956); and (c) approaches that use a combination of performance and demographic data (Krull, Scott, & Sherer, 1995; Vanderploeg & Schinka, 1995). A review of these various approaches can be found in Vanderploeg (1994) and Vanderploeg, Schinka, and Axelrod (1996). Current literature suggests that approaches that combine demographic data and scores from current ability measures are most highly correlated with actual WAIS-R IQ in both normal and clinical samples (Krull et al., 1995; Vanderploeg & Schinka, 1995; Williamson, Krull, & Scott, 1996) and that the difference score (predicted – obtained) from these combined approaches are superior at discriminating normal versus brain-damaged group membership (Vanderploeg, Schinka, & Axelrod, 1996).

Vanderploeg and Schinka (1995) developed regression equations that include demographic information as well as performance scores on each of the subtests of the WAIS-R. They created 11 regression equations (one based on each subtest) to estimate premorbid FSIQ. Taking into account WAIS-R subtest “hold” characteristic and reliabilities, it was proposed that the equations using scores from Information, Vocabulary, and Picture Completion subtests would serve as particularly good estimates of premorbid intelligence. Hold tests are subtests that were thought by Wechsler to withstand the effects of age and brain damage (Lezak, 1995). Vanderploeg, Schinka, and Axelrod (1996) subsequently recommended that the best estimate of premorbid IQ would be whichever estimate of these three equations generated the highest score (BEST-3). They found that in a clinical sample, the BEST-3 approach estimated premorbid functioning better than the demographically based approach of Barona et al. (1984).

Independently, Krull et al. (1995) developed a regression equation to predict premorbid FSIQ that was based on WAIS-R Vocabulary and Picture Completion subtest performance and demographic information. Their approach differed from that originally proposed by Vanderploeg and Schinka in that they used a different coding system for demographic variables and employed subtest raw scores rather than scaled scores. This method has come to be known as the Oklahoma Premorbid Intelligence Estimate (OPIE) approach. When there is reason to suspect that either Vocabulary or Picture Completion is impaired in a clinical patient, the authors recommended using the higher score from two alternative equations—one that included only Vocabulary with demographic variables or the other that included only Picture Completion with demographic variables. For the purposes of the present study this approach will be referred to as the OPIE-2 (highest score from either of these two equations). More recently, these investigators (Williamson et al., 1996) recommended using the Vocabulary/demographic equation if the age-scaled score for Vocabulary was 4 or more points higher than Picture Completion, and using the Picture Completion/demographic equation if the reverse were true. Otherwise, they recommended using the original combined OPIE equation. This final approach was referred to as the OPIE-revised (OPIE-R).

Given the variety of competing approaches to premorbid prediction, Vanderploeg et al. (1996) suggested that a study simultaneously comparing these approaches in a clinical sample with a carefully matched normal control group would help clarify their relative merits. Here we describe such a study in which we compared premorbid estimation approaches using a neurologically impaired clinical sample and a matched control sample in which participants were yoked on the basis of age, education, and, when possible, on occupation and ethnic origin. Five approaches to premorbid estimation were compared:
Barona et al. (1984), BEST-3 (Vanderploeg & Schinka, 1995), OPIE (Krull et al., 1995), OPIE-2 and OPIE-R (Williamson et al., 1996).

**METHOD**

**Participants**

The clinical sample was composed of 104 patients with neurological disorders who were referred to the neuropsychology clinic of an urban veterans hospital. The sample had a mean age of 43.4 (SD = 11.2) and mean years of reported education of 12.1 (SD = 1.8). All patients were men, 84% were right-handed by report. The ethnic origin of the sample was 53% White, 44% Black, and 3% Hispanic. Diagnoses included closed (n = 42) and open (n = 7) head injury, stroke (n = 10), vascular dementia (n = 3), seizure disorder (n = 15), alcohol-related dementia (n = 17), multiple sclerosis (n = 3), HIV infection (n = 4), and cerebral aneurysm (n = 3).

The normal participants were drawn from the standardization sample of the WAIS-R. Participants were individually selected to match a specific subject in the clinical sample on the basis of age group, education group, and, when possible, on occupational group and ethnic origin. When more than one participant met the matching criteria, the selection was made randomly. The clinical and control samples each contained 1, 1, 29, 41, 27, and 5 participants who received less than 8, 8, 9 to 11, 12, 13 to 15, and more than 15 years of education, respectively. The two samples contained equal numbers of participants in each of the following age ranges: 20 to 24 (n = 6), 25 to 34 (n = 16), 35 to 44 (n = 32), 45 to 54 (n = 31), 55 to 64 (n = 15), 65 to 69 (n = 2), 70 to 74 (n = 2). The frequencies for each of the unskilled, semi-skilled, unemployed, skilled, managerial, and professional occupation categories for the normal group (ns for each category are 2, 36, 12, 27, 22, and 5, respectively) did not differ those of the patient group (ns for each category are 4, 37, 10, 26, 19, and 8, respectively), $\chi^2 = 1.8$, $p = .88$. Groups did differ significantly on ethnic origin. The clinical sample had a very large proportion of minority participants and there were insufficient minority participants in the WAIS-R standardization to match all clinical participants. This resulted in the clinical sample having significantly more minority participants than the control sample (47% vs. 23%), $\chi^2 = 33.8$, $p < .001$. The potential impact this difference may have on WAIS-R performance was noted in an elderly sample to be as great as 5.4 points (Boekamp, Strauss, & Adams, 1995). If one were to assume that the difference would be the same for this younger group of participants, then a net difference of only 1.3 (24% of 5.4) should be expected between the groups.

**Procedure**

The clinical participants were administered the WAIS-R as part of an extensive neuropsychological evaluation. Published formulas were used to calculate estimated premorbid WAIS-R FSIQ scores per the Barona et al. (1984), BEST-3, OPIE, OPIE-2, and OPIE-R methods.

**RESULTS**

The mean WAIS-R FSIQ score for the patient sample was 87.4 (SD = 9.9). Estimated premorbid WAIS-R FSIQ scores for the five methods appear in Table 1. To investigate whether estimates of premorbid FSIQ scores differed from obtained WAIS-R FSIQ, a
repeated measures analysis of variance (ANOVA) with simple contrast was conducted for the clinical sample. Estimated FSIQ scores from all methods were significantly higher than obtained FSIQ, \(F(5, 99) = 69.4, p < .001, \eta^2 = .83\). FSIQ for the control group drawn from the WAIS-R standardization sample averaged 97.9 (SD = 13.9). To investigate if estimates of premorbid FSIQ scores in the clinical sample differed from the matched control group’s obtained FSIQ, a series of five sample \(t\)-tests were conducted. A Bonferroni correction of \(p < .01\) was adopted. The estimated FSIQ scores for the different approaches did not differ significantly from the control group FSIQ; \(t\)-values ranging from 0.17 to 2.46, \(p > .01\) in all cases. The Barona et al. (1984), BEST-3, OPIE, OPIE-2, and OPIE-R methods were all equally effective in estimating mean FSIQ scores for this neurologically impaired clinical sample.

Correlation coefficients between estimated and obtained FSIQ appear in Table 1 for the control group. It was predicted that the estimation methods that included performance data from the WAIS-R would generate higher correlations with FSIQ, resulting from redundant of information across the tasks. Coefficients generated by the estimation methods are uniformly greater than .80, with the exception of the coefficient for Barona equation \((r = .62)\), which was significantly lower than the coefficients produced by the other methods, \(t\)-values for difference between correlation coefficients from related samples were \(\geq 5.06, p < .01\).

Logistic regression analyses were performed to evaluate the respective ability of the premorbid estimation methods to effectively differentiate patients from controls. The criterion was group membership (brain-damaged vs. normals) and the predictors were difference scores (estimated FSIQ – obtained FSIQ) for Barona, BEST-3, OPIE, OPIE-2, and OPIE-R. Each of the five estimation methods differentiated the groups better than chance \((p < .001)\) with an equal degree of overall classification accuracy (binomial tests: \(p > .33\) in all cases). The classification rates for the patients, normals, and entire sample appear in Table 2.

**DISCUSSION**

Three sets of analyses were performed to evaluate the effectiveness of different WAIS-R FSIQ premorbid prediction methods. The methods were evaluated on the basis of (a) the comparison of mean predicted FSIQ of a neurologically impaired group to the mean actual FSIQ of a matched control sample, (b) the magnitude of correlations

<table>
<thead>
<tr>
<th>FSIQ Estimation Type</th>
<th>Clinical Sample</th>
<th>Control Sample</th>
<th>(r) with Actual FSIQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual FSIQ</td>
<td>87.4</td>
<td>97.9</td>
<td>9.9</td>
</tr>
<tr>
<td>Barona et al. (1984)</td>
<td>98.1</td>
<td>100.3</td>
<td>8.6</td>
</tr>
<tr>
<td>BEST-3</td>
<td>99.9</td>
<td>105.7</td>
<td>8.4</td>
</tr>
<tr>
<td>OPIE</td>
<td>93.5</td>
<td>100.2</td>
<td>11.3</td>
</tr>
<tr>
<td>OPIE-2</td>
<td>98.9</td>
<td>104.2</td>
<td>9.3</td>
</tr>
<tr>
<td>OPIE-R</td>
<td>95.2</td>
<td>101.8</td>
<td>11.3</td>
</tr>
</tbody>
</table>

FSIQ = Full Scale Intelligence Quotient; OPIE = Oklahoma Premorbid Intelligence Estimate.
Estimates of Premorbid Intelligence

between estimates from different methods and obtained FSIQ in the control sample, and (c) the group classification accuracy (neurologic vs. normal control) using the difference of obtained minus estimated FSIQ scores as the predictors. All prediction approaches resulted in equivalent FSIQ scores between clinical and control samples. The four methods that combined demographic and performance data (BEST-3, OPIE, OPIE-2, and OPIE-R) produced acceptably high (.74) correlations with obtained FSIQ in the control sample, while the correlation for the Barona method was lower. Finally, the Barona, BEST-3, OPIE, OPIE-2, and OPIE-R methods were equally effective in classifying participants into their respective groups, neurologically impaired and control. In summary, the four methods that combined demographic and performance data (BEST-3, OPIE, OPIE-2, and OPIE-R) were comparable to each other and to the demographic only method proposed by Barona et al. (1984).

For the BEST-3 and Barona approaches, the results of the present study were similar to those for a different clinical sample described by Vanderploeg and colleagues (1996). In both studies, the Barona demographic-based method was less highly correlated with current FSIQ in the normal control samples. However, in contrast to the prior study, the logistic regression analyses of the present study were significantly less effective in differentiating controls from patients for all premorbid prediction approaches. In addition, results from the two studies differ in that the BEST-3 approach significantly outperformed the Barona approach in differentiating normals from patients in the prior study (Vanderploeg et al., 1996), whereas both methods were comparable in the current study. The discrepancy in the findings is most likely due to unidentified differences in the clinical populations. An additional concern that should be again noted is the difference between the groups of this study with regard to ethnic/racial composition. As noted by Boekamp, Strauss, and Adams (1995), predicted and obtained FSIQ might be altered by as much as one third of a standard deviation by this demographic difference.

Summarizing the results of the previous (Vanderploeg et al., 1996) and current studies together, it appears that premorbid estimation approaches that combine demographic and performance data (i.e., BEST-3, OPIE, OPIE-2, and OPIE-R) are equally effective approaches for premorbid prediction. Less well understood is the efficacy of using demographics alone (e.g., Barona et al., 1984).

Premorbid prediction research to date has used normative samples to develop the different prediction approaches and clinical samples to evaluate the validity of these methods. A more revealing study would use a prospective data collection design in which estimates of premorbid functioning were obtained via direct examination prior to the neurological event. In the interim, the BEST-3, OPIE, OPIE-2, and OPIE-R appear to be equally effective premorbid prediction methods.

### TABLE 2
Classification Accuracy of Patients and Controls Using Logistic Regression Analysis with Estimation Methods

<table>
<thead>
<tr>
<th>FSIQ Estimation Type</th>
<th>Patients (%)</th>
<th>Controls (%)</th>
<th>Overall Classification (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barona et al. (1984)</td>
<td>66.4</td>
<td>66.4</td>
<td>66.4</td>
</tr>
<tr>
<td>BEST-3</td>
<td>61.5</td>
<td>59.6</td>
<td>60.6</td>
</tr>
<tr>
<td>OPIE</td>
<td>61.5</td>
<td>64.4</td>
<td>63.0</td>
</tr>
<tr>
<td>OPIE-2</td>
<td>65.4</td>
<td>63.5</td>
<td>64.4</td>
</tr>
<tr>
<td>OPIE-R</td>
<td>62.5</td>
<td>61.5</td>
<td>62.0</td>
</tr>
</tbody>
</table>

FSIQ = Full Scale Intelligence Quotient; OPIE = Oklahoma Premorbid Intelligence Estimate.
REFERENCES


