Estimating premorbid IQ from demographic variables: A comparison of a regression equation vs. clinical judgement

John R. Crawford*, Jane Millar and Alan B. Milne

Department of Psychology, University of Aberdeen, UK

Objectives. To assess the accuracy with which clinicians estimate premorbid IQ from demographic variables and compare it with a regression equation which uses the same information.

Design. Repeated measures and correlational.

Methods. Sixty participants were administered the WAIS-R and had their demographic variables recorded (age, sex, years of education and occupation). Eight clinical psychologists estimated the participants’ IQs from the demographic variables. Estimated IQs were also obtained using a regression equation developed by Crawford and Allan (1997).

Results. The correlation between obtained IQ and the equation-based estimate was significantly higher than the correlation between obtained IQ and the clinicians’ estimates. Further, mean estimated IQ from the regression equation did not differ significantly from obtained IQ whereas the means for four of the eight clinicians’ estimates did differ significantly.

Conclusions. Demographic-based regression equations can provide unbiased and useful estimates of premorbid IQ; these estimates can be modified in the light of the additional qualitative information available to the clinician.

Introduction

The detection and quantification of cognitive impairment in the individual case is problematic. Because of the substantial individual differences in cognitive ability in the general population, simply comparing a client’s current test performance with the relevant test norms will be of limited value. A particular test score can represent an entirely normal level of functioning for one individual and yet a serious decline for another (Crawford, Venneri, & O’Carroll, 1998; Deary, 1995). Therefore, it is necessary to compare current performance against an individualized comparison standard (Lezak, 1995). As previous test results are rarely available, the clinician must resort to some method of estimating a client’s ‘expected’ or ‘premorbid’ level of performance.

* Requests for reprints should be sent to Professor John R. Crawford, Department of Psychology, King’s College, University of Aberdeen, Aberdeen AB24 2UB, UK (e-mail: j.crawford@abdn.ac.uk).
The most common approach to obtaining an estimate of general premorbid functioning is to use tests of current functioning that: (a) have adequate reliability; (b) correlate highly with IQ in the general population; and (c) are resistant to the effects of neurological and psychiatric disorder (Crawford, 1992). However, it is highly questionable whether any existing test fully meets this second criterion. Encouraging results have been obtained using the National Adult Reading Test (Nelson, 1982; Nelson & Willison, 1991) but even for this test there are indications that a degree of decline takes place in some neuropsychiatric conditions (Crawford, Parker, & Besson, 1988; O’Carroll, Moffoot, Ebmeier, & Goodwin, 1992; Patterson, Graham, & Hodges, 1994; Stebbins, Wilson, Gilley, Bernard, & Fox, 1990).

An alternative approach is to take advantage of the well-established relationship between demographic variables (e.g. education, occupation) and IQ (Kaufman, 1990; Matarazzo, 1972). Wilson, Rosenbaum, Brown, Rourke, Whitman, and Grissell (1978) used the USA standardization sample to build regression equations for the prediction of WAIS IQ from five demographic variables. These equations predicted 54% of the variance in Full Scale (FSIQ). Subsequently the use of demographic regression equations to estimate premorbid IQ has been extended to countries outside the USA (Crawford et al., 1989a) and equations have been developed for use with the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981); e.g. see Barona, Reynolds, and Chastain (1984) and Crawford and Allan (1997). Crawford and Allan (1997) built regression equations using a healthy sample (N = 200) which was representative of the adult UK population in terms of the distributions of age, sex, and social class. In this study the three demographic variables of age, years of education and social class (the latter being coded from an individual’s occupation or former occupation) accounted for 53% of the variance in WAIS-R FSIQ.

In attempting to detect and quantify acquired cognitive deficits, a client’s demographic details (e.g. educational and occupational history) are used informally by clinicians to estimate her/his expected or premorbid level of performance (Crawford, 1996; Lezak, 1995; Walsh, 1991). This raises the question as to whether the use of formal regression equations to estimate premorbid ability from demographic variables are any more accurate than clinicians’ informal estimates based on the same information. In the present study this issue is examined empirically by gathering WAIS-R data from a healthy sample, asking experienced clinical neuropsychologists to predict participants’ IQs from demographic variables and comparing the accuracy of prediction with Crawford and Allan’s (1997) regression equation. It will be appreciated that any attempt to evaluate the accuracy (i.e. criterion validity) of demographic estimates of premorbid IQ, whether these are derived from regression equations or clinicians’ ratings, must necessarily use healthy participants as the criterion variable is no longer available in individuals suffering from neurological or psychiatric disorder; see Crawford (1989) for further discussion.

An additional aim of the present study was to explore whether clinicians show any systematic, implicit, biases in their estimates of IQ. For example, it may be that some clinicians reveal an implicit ageism or sexism in that the weight they attach to age or sex in arriving at their estimates is at variance with the empirical evidence.
Method

Participants

The WAIS-R sample consisted of 60 individuals (33 females, 27 males) free of known neurological, psychiatric or sensory disorder. Most received a small honorarium for their participation. Participants were recruited from a wide variety of sources, i.e. local and national businesses, clubs (e.g., senior citizens clubs, angling clubs), community centres etc. Mean age was 46.0 years (SD = 19.10) with a range of 16–82 years. Mean total years of education was 12.5 (SD = 3.28) with a range of 7–20 years. To conform to the procedure followed by Crawford and Allan (1997) participants were credited with 0.5 of a year for every year spent in part-time education. Part-time education was defined to include ‘day-release courses and evening classes provided they had led, or were leading to, a qualification’ (Crawford & Allan, 1997, p. 193). Each participant’s social class was coded from their occupation using the Office of Population Censuses and Surveys (1980) Classification of Occupations. The UK system for coding social class consists of five categories which can be defined broadly as follows: 1 = professional, 2 = intermediate, 3 = skilled, 4 = semi skilled, 5 = unskilled. To conform to Crawford and Allan’s (1997) procedure, retired participants were coded by their previous occupations as were those currently unemployed. Those who had never worked were coded as social class 5. Individuals describing themselves as househusbands/housewives were coded by their previous occupations.

The percentages of the sample in each social class band are presented in Table 1 together with the census-derived percentages for the adult UK population. A goodness-of-fit chi-squared test revealed that the sample and population proportions did not differ significantly, $\chi^2(4, N = 60) = 3.65, p > .05$. A similar procedure was used to examine the representativeness of the sample in terms of age distribution. The sample and UK census-derived percentages in each of three age bands are presented in Table 2. A goodness-of-fit test revealed that the observed and expected proportions did not differ significantly, $\chi^2(2, N = 60) = 5.42, p > .05$. Finally, the sample’s sex distribution was also representative of the adult UK population, $\chi^2(1, N = 60) = 0.22, p > .05$.

<table>
<thead>
<tr>
<th>Social class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>WAIS-R sample</td>
</tr>
<tr>
<td>General adult UK population</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age distribution (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>16–35</td>
</tr>
<tr>
<td>WAIS-R sample</td>
</tr>
<tr>
<td>General population</td>
</tr>
</tbody>
</table>

Eight raters were recruited (four females, four males) all of whom were qualified clinical psychologists with extensive experience of neuropsychological assessment. Two of the raters were NHS grade A at the time the study was conducted, the remainder were Grade B. All regularly conducted neuropsychological assessments for medico-legal purposes. All those approached to act as raters agreed to take part.
Procedure

All participants completed a full-length WAIS-R (UK) according to standardized procedures (Wechsler, 1981; Lea, 1986). The raters were presented with a pack consisting of information on the purpose of the study (available on request from the first author) and the following demographic details for each of the 60 participants: age, sex, occupation and accompanying social class code, and total years of education. Education was also broken down into years of schooling, years of further/higher education and years of part-time education. Each rater entered an estimated WAIS-R FSIQ for each participant. For each participant, the relevant demographic data (age, total years of education, and social class code) were also entered into Crawford and Allan’s (1997) regression equation for the estimation of WAIS-R FSIQ. The equation takes the following form:

\[
\text{Predicted FSIQ} = 87.14 - (5.21 \times \text{class}) + (1.78 \times \text{years of education}) + (0.18 \times \text{age})
\]

Results

Summary statistics (means, SDs and ranges) for obtained IQ, and estimated IQs from the demographic regression equation and eight raters are presented in Table 3. It can be seen that there is a close correspondence between the mean obtained IQ and the mean estimated IQ from the regression equation. In contrast, many of the mean estimates derived from the clinical raters diverge from the obtained IQ; the largest discrepancy (rater 3) amounts to approximately half an IQ SD. With one exception (rater 5), the clinicians ratings overestimate the obtained IQs. Dunnett’s test (Dunnett, 1955) was used to compare each of the nine mean estimated IQs with the mean obtained IQ; this test controls for inflation of the Type I error rate but is more powerful than any alternative procedures (Howell, 1997). The results of Dunnett’s test are presented in columns 4 and 5 of Table 3. It can be seen that the mean equation-based estimate did not differ significantly from the mean obtained IQ but that four of the eight mean estimated IQs derived from the clinical raters did differ significantly.

Table 3. Summary statistics for obtained WAIS-R IQs and IQs estimated by the demographic regression equation and eight clinical raters; also shown are the results of Dunnett’s test comparing the mean estimated IQs with the mean obtained IQ

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Range</th>
<th>(t_D)</th>
<th>(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obtained IQ</td>
<td>102.8</td>
<td>13.92</td>
<td>71–140</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Regression equation</td>
<td>102.5</td>
<td>10.01</td>
<td>82–126</td>
<td>0.33</td>
<td>n.s.</td>
</tr>
<tr>
<td>Rater 1</td>
<td>104.9</td>
<td>11.32</td>
<td>85–129</td>
<td>-2.06</td>
<td>n.s.</td>
</tr>
<tr>
<td>Rater 2</td>
<td>106.8</td>
<td>12.51</td>
<td>78–132</td>
<td>-3.88</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Rater 3</td>
<td>110.3</td>
<td>12.94</td>
<td>85–140</td>
<td>-7.32</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Rater 4</td>
<td>103.9</td>
<td>12.86</td>
<td>80–140</td>
<td>-1.07</td>
<td>n.s.</td>
</tr>
<tr>
<td>Rater 5</td>
<td>99.3</td>
<td>10.35</td>
<td>80–125</td>
<td>-3.45</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>Rater 6</td>
<td>103.6</td>
<td>9.55</td>
<td>90–148</td>
<td>-0.76</td>
<td>n.s.</td>
</tr>
<tr>
<td>Rater 7</td>
<td>107.0</td>
<td>12.87</td>
<td>90–148</td>
<td>-4.09</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Rater 8</td>
<td>103.3</td>
<td>13.87</td>
<td>80–130</td>
<td>-0.44</td>
<td>n.s.</td>
</tr>
</tbody>
</table>
The correlations between the obtained IQs and IQs estimated by the regression equation and clinical raters are presented in the second column of Table 4. The average correlation between the raters’ estimates and obtained IQs are presented at the foot of this column. All correlations were significant at the .01 level. However, it can be seen that the equation-based estimate has a higher correlation with obtained IQ than any of the raters’ estimates. To test if the former correlation was significantly higher than the latter correlations, a procedure developed by Meng, Rosenthal, and Rubin (1992) was used. This procedure tests the significance of contrasts between competing predictors and a criterion and yields a $z$ statistic that can then be evaluated for significance by referring to the area under the normal curve. To conduct the desired comparison the equation-based correlation was assigned a contrast weight of 8 and each of the raters’ correlations a contrast weight of −1 (the contrast weights must sum to zero). This test yielded a $z$ of 2.86 and a consequent two-tailed $p$ of .004. Thus, the equation-based estimate of the participants’ FSIQs had a significantly higher correlation with the participants’ obtained IQs than did the estimates derived from the clinical raters.

Table 4. Pearson correlations between obtained IQs and IQs estimated by the demographic regression equation and eight clinical raters; also shown are the correlations between demographic variables and estimated IQs/obtained IQ

<table>
<thead>
<tr>
<th></th>
<th>$r_{IQ}$</th>
<th>$r_{age}$</th>
<th>$r_{sex}$</th>
<th>$r_{education}$</th>
<th>$r_{class}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obtained IQ</td>
<td></td>
<td>.02</td>
<td>−.16</td>
<td>.65</td>
<td>−.72</td>
</tr>
<tr>
<td>Regression equation</td>
<td>.76</td>
<td>.03</td>
<td>−.18</td>
<td>.88</td>
<td>−.89</td>
</tr>
<tr>
<td>Rater 1</td>
<td>.68</td>
<td>−.33</td>
<td>−.21</td>
<td>.87</td>
<td>−.90</td>
</tr>
<tr>
<td>Rater 2</td>
<td>.66</td>
<td>−.28</td>
<td>−.13</td>
<td>.85</td>
<td>−.82</td>
</tr>
<tr>
<td>Rater 3</td>
<td>.70</td>
<td>−.29</td>
<td>−.18</td>
<td>.83</td>
<td>−.85</td>
</tr>
<tr>
<td>Rater 4</td>
<td>.63</td>
<td>−.29</td>
<td>−.16</td>
<td>.81</td>
<td>−.76</td>
</tr>
<tr>
<td>Rater 5</td>
<td>.67</td>
<td>−.22</td>
<td>−.25</td>
<td>.81</td>
<td>−.77</td>
</tr>
<tr>
<td>Rater 6</td>
<td>.63</td>
<td>−.33</td>
<td>−.28</td>
<td>.90</td>
<td>−.84</td>
</tr>
<tr>
<td>Rater 7</td>
<td>.64</td>
<td>−.27</td>
<td>−.22</td>
<td>.87</td>
<td>−.81</td>
</tr>
<tr>
<td>Rater 8</td>
<td>.64</td>
<td>−.34</td>
<td>−.28</td>
<td>.84</td>
<td>−.81</td>
</tr>
<tr>
<td>Averaged across eight raters</td>
<td>.64</td>
<td>−.29</td>
<td>−.21</td>
<td>.85</td>
<td>−.83</td>
</tr>
</tbody>
</table>

Note. The mean correlation for raters was obtained by converting each correlation to Fisher’s $z$, calculating the mean, and converting back to $r$.

The correlations between demographic variables (age, sex, years of education, and social class) and the obtained and estimated IQs are presented in Table 4. It can be seen that both years of education and social class are highly correlated with IQs estimated by the clinical raters; thus raters gave a substantial weight to these variables when forming their estimates. This is appropriate as the correlations between these variables and obtained IQs are also high and statistically significant. However, the correlation between age and obtained IQs in the present sample was essentially zero ($r = .02$). Therefore, it would be undesirable for the estimates of IQ to correlate with age, as this would represent a systematic bias. It can be seen from
Table 4 (column 3) that IQ estimated by the regression equation has a near zero correlation with age \((r = .03)\), in contrast with the raters’ estimated IQs, for which the average correlation is \(-.29\). To test if the differences between the two methods were significant, the previously outlined procedure of Meng et al. (1992) was used; age was the criterion variable in this analysis and weights of 1 were assigned to each of the rater correlations with age, and \(-8\) assigned to the equation-based correlation. This yielded a \(z\) of 5.72 \((p < .0001)\). Thus, overall, the raters’ estimates of IQ were significantly more highly correlated with age than the equation-based estimate.

As shown in Table 4, the point–biserial correlation between obtained IQ and the sex of the participants (females coded as 1 males as 0) was low and non-significant \((r = -.16)\) as was the correlation between IQ estimated by the regression equation \((r = -.18)\). The averaged correlation between IQs estimated by the clinical raters and sex of the participants was \(-.21\); the averaged correlation for female raters was \(-.19\) vs. \(-.24\) for male raters. Thus, raters did not exhibit a sizeable bias in the weightings given to the sex of the participants.

**Discussion**

From one perspective the performance of the clinical raters in the present study was impressive. The correlations between the estimated and obtained IQs were substantial and all were highly significant. Nevertheless, the regression equation outperformed the clinical raters in that the estimated IQs derived from it were significantly more highly correlated with the obtained IQs than were the rater-derived estimates. In addition, the mean estimated IQs from the regression equation closely corresponded to the mean obtained IQ, whereas the mean estimated IQs derived from four of the eight raters were significantly different. These latter results are as important as the correlation results because estimates of premorbid IQ are used to provide an individual comparison standard for current performance. Therefore, the absolute value of the estimated IQ is crucial.

We suggest that the superior performance of the regression equation can be attributed to three related factors. Firstly, the equation applies the optimal weights to the demographic predictor variables, thereby maximizing accuracy. Secondly, as the concern is with estimating *premorbid* IQ, experience with clinical cases (even if extensive) has limitations as a means of developing an accurate internalized model of IQ–demographic relationships. Experience with testing large numbers of individuals drawn from the healthy population would help with this process but few of us have had such opportunities. Therefore, clinical estimates may often be based on vague or distorted impressions of IQ–demographic relationships. Thirdly, empirical studies of expert judgment in other fields indicate that humans are poor at dealing with multiple predictors (Hogarth, 1975; Slovic & Lichtenstein, 1971).

Although the regression equation performed significantly better than clinicians, we do not wish to use these results to argue for actuarial prediction (Meehl, 1954) over clinical prediction. Instead, we suggest that it is appropriate to use regression equations to provide an *initial* estimate of premorbid IQ rather than to rely on a clinical estimate derived from the same information. This initial estimate can then be modified by an evaluation of the additional information available to the clinician.
That is, we believe it is best to use formal quantitative methods with information that can be readily quantified, whilst at the same time acknowledging an important role for qualitative data. Clinicians in the present study were provided with information over and above that used in the regression equation (i.e. the participants’ occupations in addition to their occupational codes, and a breakdown of the participants’ years of formal education). However, in clinical practice there is a wealth of additional sources of information that can also be incorporated into clinical judgments. These include the interview with the client and their relatives, and observations made during formal testing. For example, the clinical interview can be used to ascertain whether there were specific psychosocial or economic factors in the client’s past that may have prevented them achieving their full educational or occupational potential. Conversely, there may be factors that permitted a client to achieve educational or occupational goals despite relatively modest premorbid cognitive abilities.

IQ scores are age-corrected and thus the near zero correlation between age and obtained FSIQ in the present sample was to be expected. However, although clinicians are aware of this correction, all clinical raters’ estimates of IQ were negatively correlated with age. This systematic bias may reflect the direct effects of implicit ageism whereby clinicians overestimated the level of decline in cognitive abilities that occurs with ageing. Alternatively, it may be that, in weighting education, the raters gave insufficient regard to the fact that educational opportunities were generally more limited (and access was more variable and less equitable) when the late middle-aged and elderly of today were in their youth.

Regardless of the reason for this bias it is a finding that should be noted by all of us who work clinically with elderly populations as it was highly consistent across the clinical raters. It can also be seen from Table 4 that the regression equation shows no such bias. The latter result may seem at odds with the fact that age is a term in Crawford and Allan’s (1997) regression equation (and made a statistically significant contribution to prediction). However, inspection of their correlation matrix (Crawford & Allan, 1997, Table 1) reveals that age is not correlated with FSIQ in the sample used to generate the equations. Therefore, although this is not remarked upon by the authors, age acts as what is termed as a ‘suppressor’ variable (Darlington, 1990) in their regression model; i.e., it does not correlate with the criterion (IQ) but does correlate (negatively) with one of the other predictors (years of education). Thus the relationship between education and IQ would be attenuated if age was not controlled for in the model; presumably for the reasons outlined earlier, i.e., changes in educational provision.

The most common formal alternative to the demographic approach to estimating premorbid IQ is to use present ability measures, such as the NART, that are relatively resistant to many forms of cerebral dysfunction. The advantage of this latter approach is that it is liable to be a more powerful predictor of the criterion variable (i.e., IQ test performance). For example, the healthy sample used to generate the UK demographic equations for the WAIS (Crawford et al., 1989a) was also used to cross-validate the NART (Crawford, Stewart, Parker, Besson, & De Lacey, 1989b). Therefore, it is possible to compare directly the two methods in the same sample. The NART predicted 66% of FSIQ variance in comparison with 50% for demographic variables.
The lesser predictive power of demographic methods is offset by their major advantage, namely that the estimates they provide are entirely independent of an individual’s current cognitive status. As noted previously, there is increasing evidence for a degree of impairment of NART performance in some neuropsychiatric disorders. Therefore, the NART will tend to systematically underestimate premorbid ability; see Crawford (1992), Franzen, Burgess, and Smith-Seemiller (1997), and O’Carroll (1995) for reviews. Furthermore, the demographic approach can be used with clients in whom the NART would clearly be inappropriate, i.e., dysarthric or dyslexic cases. However, caution should be exercised when using demographic-based estimates with cases suffering from disorders that may have a prodromal phase. In such cases the prodromal phase may have resulted in a failure to achieve their educational and occupational potential. A similar degree of caution must be exercised when using reading-based estimates of premorbid IQ with such cases.

With the publication of the third edition of the Wechsler Adult Intelligence Scale (WAIS-III; Wechsler, 1997; Wechsler, Wycherley, Benjamin, Crawford, & Mockler, 1998) there will be a need to develop new regression equations for the estimation of premorbid IQ. Presumably, the NART, or alternative reading tests (Baddeley, Emslie, & Nimmo-Smith, 1993; Beardsall & Huppert, 1994), will be restandardized against the WAIS-III. The present results, and those of Crawford and Allan (1997), suggest that it would also be worthwhile to develop demographic-based estimates of premorbid WAIS-III IQs or Indexes. Indeed, we do not consider that clinicians should use either a reading-based approach or a demographic approach. Instead, given that neither method is time consuming, we view them as complementary.

References


Matarazzo, J. D. (1972). Wechsler’s measurement and appraisal of adult intelligence (5th ed.). Baltimore: Williams & Wilkins.


Received 11 November 1999; revised version received 11 April 2000