British Journal of Clinical Psychology

Estimating verbal fluency and naming ability from the test of premorbid functioning and demographic variables: Regression equations derived from a UK sample

--Manuscript Draft--

Article Type:	Full Article
Manuscript Number:	BJCP.17.0081R1
Full Title:	Estimating verbal fluency and naming ability from the test of premorbid functioning and demographic variables: Regression equations derived from a UK sample
Corresponding Author:	Stephen James Evans, D.Clin.Psy, Msc., Bsc., PGdip. NHS N/A, N/A UNITED KINGDOM
Corresponding Author E-Mail:	stephen.evans@stees.nhs.uk
Order of Authors (with Contributor Roles):	Toni-Marie Jenkinson (Data curation; Formal analysis; Investigation; Methodology; Project administration; Writing – original draft; Writing – review & editing)
	Steven Muncer (Formal analysis; Methodology; Supervision; Writing – review & editing)
	Miranda Wheeler (Data curation; Project administration)
	Donald Brechin (Data curation)
	Stephen James Evans, D.Clin.Psy, Msc., Bsc., PGdip. (Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Supervision; Writing – review & editing)
Funding Information:	
Keywords:	Neuropsychology; Regressions; TBI; Meningioma; TOPF
Abstract:	Abstract Objectives. Neuropsychological assessment requires accurate estimation of an individual's premorbid cognitive abilities. Oral word reading tests, such as the Test of Premorbid Functioning (TOPF), and demographic variables, such as age, sex and level of education, provide a reasonable indication of premorbid intelligence, but their ability to predict other related cognitive abilities is less well understood. The current study aimed to develop regression equations, based on the TOPF and demographic variables, to predict scores on tests of verbal fluency and naming ability. Methods. A sample of 119 healthy adults provided demographic information and were tested using the TOPF, FAS, Animal Naming Test (ANT) and Graded Naming Test (GNT). Correlational analyses were used to explore relationships between the test and demographic variables. Multiple regression analyses, using the TOPF and demographics as predictor variables, were used to estimate verbal fluency and naming ability test scores. Change scores and cases of significant impairment were calculated using the method in Knight et al., (2006). Results. Demographic variables provided a significant contribution to the prediction of all verbal fluency and naming ability test scores; however, adding TOPF score to the equation considerably improved prediction beyond that afforded by demographic variables alone. The percentage of variance accounted for by demographic variables and/or TOPF score varied from 19 percent (FAS), 28 percent (ANT) and 41 percent (GNT). Change scores revealed significant differences in performance in the clinical groups, particularity the TBI group. Conclusions. Demographic variables, particularly education level, and scores on the TOPF should be taken into consideration when interpreting performance on tests of verbal fluency and naming ability.
Additional Information:	

Question	Response
If you have any potentially competing interests to declare, please enter them in the box below. If you have no interests to declare, please enter 'none'.	None
Does this submission have any links or overlap with any other submitted or published manuscripts, for this or any other publication? (For example; as part of a long-term project, using a shared data set, a response to, or extension of, earlier work.) If yes, please give brief details. If no, please enter 'none'. Any overlap not declared and later discovered will result in the manuscript being withdrawn from consideration.	No
Please specify the word count of your manuscript (excluding the abstract, tables, figures and references). Please note: papers exceeding the word limit will be returned to the author, unless they have received prior permission from the Editor for the submission of a longer manuscript.	4995

Dear reviewers,

We were very pleased that you saw potential in our research and we have been working to address the issues both reviewers raised with our paper in its previous form. We hope you are happy with the amendments we have made and see the work as fit for publication in the BJCP.

Reviewer 1.

(1) The sentence (p 3) … 'One approach to this is to use demographic variables such as age, sex, ethnicity and education within regression equations to generate estimates of premorbid ability based on the mean IQ of a demographically identical sample.' …seems ill-worded in is final part. The reference to '…based on the mean IQ…' is confusing and should be deleted.

We have reworded the sentence in question and have deleted "based on the mean IQ".

(2) Because the application of prediction equations for verbal fluency and naming ability to a clinical sample is the main focus of this investigation, it would be useful to expand a bit on this particular aim at the end of the introduction. What is currently referred to as 'utility of these equations' appears too vague.

We have expanded at the end of the introduction in order to clarify how we hope our equations will add to clinical assessment and haver references previous papers in more detail (such as Knight, Crawford etc..)

With respect to a revision of the Method section, the following points should be considered:

(1) The socio-demographic description of the normative sample is rather limited; more information should be provided about how the participants were selected, probably via opportunity sampling, their ethnicity, employment status and the geographical location. It is important that the reader gets a very clear idea which segment of the general population this sample is likely to represent. The information currently presented in Table 1 could be easily integrated into the description of the normative sample and would free up one table.

We have added Information on how participants were selected, their geographical location and ethnicity has been added under the section on the 'Normative Sample' in the 'Participants' section (para 1 method). We have decided to keep table 1. As we feel it depicts the differences between age and education between the normative sample and the clinical groups more clearly than if we try to describe it– we hope you agree. We were unable to comment on employment status as we did not collect this data. We have commented on this in the discussion as a limitation of the study.

(2) Report the sample sizes for the meningioma and TBI sub-samples.

We have now reported the sample sizes for meningioma and TBI under the "Participants" section.

(3) Report a reference for the FAS and ANT. It would also be useful to report the reliabilities of the total scores for the ability measures.

We have added references for the FAS and the ANT, and information relating to the reliability and validity of each measure has been added in the 'Measures' section.

(4) In the section Measures, add a brief explanation on the scoring of Level of Education. In the subsequent regression analyses you have used this variable as a linear predictor, but this requires that its scores are equidistant (ie an interval scale). Can this be justified?

We have added an explanation of the levels of education added under the 'Measures' section. The coding of educational level assumed a hierarchical order in which passing exams at a higher level scored more points. To some extent this is arbitrary, however, we also checked other methods of coding the educational level which assumed a categorical structure of different forms. The original coding was found to produce a stronger relationship to the predicted variables, even though this was not strong enough to enter the equation.

(5) The final sentence in Procedure (in Clinical Sample) relates to the statistical data analysis and is misplaced there. Rather, there should be a new section Data Analysis following Procedure where you provide relevant information about the statistical methods that were used (eg ANOVA, multiple regression) and how their assumptions where checked. More specifically, information should be given for how the classic assumptions of multiple regression (absence of multicollinearity, linearity of the relationship between the predictors and the criterion variable, normality and homoscedasticity of the residuals) have been checked. Relevant results should be provided either in this section or later in Results. This is very important since a violation of the linearity or the homogeneity assumption of the residuals would result in distorted prediction equations. A classic textbook to consult could be *Cohen, Cohen, West & Aiken (2003) Applied Multiple Correlation/Regression Analysis for the Behavioural Sciences*.

We have added a "Data Analysis" section, including details of the statistical method used and how the assumptions were checked, with results provided for skew and kurtosis.

(6) Finally, briefly explain how the sample size was determined. This could be done by reporting the results of a power calculation for a multiple regression with 4 predictors to detect a minimum effect size of interest. A sample size of 119 seems large enough to run a multiple regression equation with up to 4 predictors and produce fairly precise regression parameter estimates.

Details of power calculations are now contained in the 'Data Analysis' section.

My major criticism of the paper concerns the Results section. At present, important details are missing, notably the analysis of the predicted scores in the clinical sample seems incomplete and so the clinical utility of the prediction equations is not yet established. In the following, I am raising a number of points for you to consider when revising the results section.

(1) The first section Comparison of samples could be reduced considerably. Firstly, there is no explanation why you reported in Table 3 both results for ANOVAs and the Kruskal-Wallis test? If the assumptions for ANOVA where checked and not violated, the K-W results are superfluous. Alternatively, if the assumptions for ANOVA were not met, the K-W results should be reported only. To save on words, the results for age could be added to Table 3. Secondly, in this section you are treating Level of Education as a categorical variable whereas in the subsequent regression analyses it has been treated as an interval variable; this appears inconsistent. If Level of Education is a metric index variable, a mean comparison between the normative and clinical sample would be feasible instead of the reported crosstabulation analysis. In any case, comments on why the crosstable was collapsed should be taken out as this is irrelevant information. Finally, only the p-values have been reported, but no effect sizes of the mean differences that were statistically significant.

We agree with these helpful suggestions and have reduced the comparison of samples section and have reported the appropriate test (ANOVA/KW) in Table 2. We have removed the cross validation section and instead included comparative information between the normative and clinical samples. We have included effect sizes of the mean differences for the mean differences that were statistically significant.

(2) Add percentages and samples sizes to Table 2. If you added the result of the Chi-squared test of independence as a footnote, this would save you further words in the results section.

We have added percentages and sample sizes to table 2 and have added the Chi-squared results as a footnote to table 2 and have removed the text from the body of the report.

(3) In the section Correlation you mention non-normality of the distribution of some of the variables. Provide relevant descriptive statistics (ie skewness) for the measures so the reader gets an idea as to the extent of non-normality; this could be briefly presented in the section Data Analysis.

We have added a "Data Analysis" section, including details of the statistical method used and how the assumptions were checked, with results provided for normality, skew and kurtosis.

(4) The section Multiple Regression Analysis lacks any information about the statistical significance of a prediction equation (F-test) or the R^2-change by adding a predictor. Whilst this information could be kept to a minimum, it should not be missing. One option would be to present all results pertaining to the prediction equations in a table rather than in individual equations, and the respective F-tests could be stated as footnotes. An example for such a table would be Table 4 in *Knight et al. (2006) in BJCP*.

We thank the reviewer for the helpful reference to Knight et al, and have adopted his suggestion of putting the majority of the regression equation information in a table (see table 4)

(5) The abbreviation SE at the end of each equation is both confusion and wrong. I presume you meant to write SEE or SEest? This information could be reported in a separate column in a table showing the regression weights for the different predictors and criterion variables (see my previous point). Ideally, you would also want to report in this table for each regression weight its 95% - CI or at least its SE in order to convey its precision.

We have changed the abbreviation to SEest and have added it as a separate column to table 4.

(6) Unfortunately, there is no information on regression diagnostics to ensure that the statistical assumptions have been met and that the presented regression equations are robust. You therefore want to comment on the influence of any large outliers on the regression equation (eg Cook's distance), any problems with non-constant error variance or non-normality of the residuals and, very importantly, whether the predictors were linearly related with the criterion variable. Relevant information can be presented succinctly either in the results section or in the Data Analysis section in Methods.

We have now also included information on the residual checks, which include normality, homoscedasticity, outliers, checks for the influence of particular cases, multicollinearity etc. Specifically, Shapiro Wilks is used to check the normality of the residuals, Durbin Watson, Cook's and Mahalnobis statistics are given, as are VIF and Tolerance. Levene's test of equality of variances is used to check heteroscedasticity on the tertiles based on the predicted scores.

(7) The final section Cross Validation, presents results not only for the meningioma and TBI sample, but also results for these two samples combined. I could not understand the reason for this? In previous analyses the two clinical groups had been separated for obvious reasons and this should also be the case here. Results relating to the combined clinical sample should be removed.

We have removed any results relating to combined clinical samples.

(8) The reported correlations between the observed test scores and the predicted pre-morbid scores are of limited interest, whereas the reported mean differences between these two measures are important evidence for a reduced performance in the two clinical groups post injury. Again, in addition to reporting p-values, an effect size d should be reported to convey the practical significance of the findings.

See answer to question 9 we have removed the correlation section and the table from the study to refocus it and save words.

(9) Because a major aim of this study was to develop RBNs to estimate the premorbid level in verbal fluency and naming ability in the clinical sample, this aspect deserves more attention. In order to evaluate the clinical utility of these newly developed prediction equations properly, it would be important to carry out further analysis focussing on the size and distribution of the individual discrepancy scores between the predicted premorbid level and the actual performance score obtained post injury. These discrepancy scores can be standardized (using the SEE) and translated into a z-values enabling to identify those cases with a particular large (eg at the 10th percentile) deterioration in performance relative to their estimated pre-morbid level. It would also be possible to construct a lower confidence limit around the predicted pre-morbid performance scores and then identify those cases with test scores falling outside this interval. Relevant references on this topic to consult would be; *Crawford & Garthwaite (2006) Neuropsychology; Van Breukelen & Vlaeyen (2005) Psychological Assessment; Tesa et al. (2009) Journal of the International Neuropsychological Society*.

We thank the reviewers for helpful suggestions about the use of the equations for prediction. We have added the 10th percentile score to the regression table which helps define the size of the difference for impairment. We have also calculated the standard scores of the difference between predicted and obtained scores. Here we have calculated both the standard error of the estimate and standard error of a new individual. We have supplied the effect size values along with the t tests for differences between predicted and obtained scores for the two clinical samples. These calculations can be seen in the "Predicting scores for the clinical sample" section of the results.

(10) Obtaining the percentage of significant deterioration in each clinical group would be interesting to know and would take the analysis more into the area of clinical assessment where test scores of individual patients are at the centre. In addition, you could compare the discrepancy scores between the two clinical groups to find out whether there were more serious cases of impairment in the TBI as compared to the meningioma group. Finally, you could also consider contrasting a few serious cases of TBI or meningioma with mild cases in order to find out whether the discrepancy scores of the former are considerable larger in comparison with the latter. This would allow you to comment on the clinical utility of these prediction equations in distinguishing well between mild and severe clinical cases (discriminant validity).

The standard scores computed with the standard error of a new individual were used to define impairment with 1.29 being used as a cut off. We have then compared the two groups on the three measures for levels of impairment. There are some differences which we have commented on (para 3 "Predicting scores for the clinical sample" section).

With respect to the Discussion section, I have a few comments and suggestions to make:

(1) The discussion of the correlational analysis of the ability measures at the beginning should be shortened considerably as it appears somewhat redundant in view of the detailed presentation of this topic in the introduction. Also, no results should be reiterated in the discussion section unless there is a particular reason.

Paragraph 1 of the discussion has been shortened and some of the detail regarding correlation analysis has also been removed.

(2) Whilst I agree with you conclusion that the TOPF appears to be a valuable predictor for estimating pre-morbid levels of performance in naming ability and semantic fluency in addition to socio-demographic predictors, I fear your comments about the 'validity of the equations' (p 16) when applied to the clinical sample are as yet an overstatement. This is because your comments are based only on an average reduction in performance and so the extent of the variation of the size of the impairment in verbal fluency or naming ability for

individual patients is not clear. However, here you could comment on results relating to the discrepancy scores as well as the percentages of severe deterioration (impairment) in the two clinical groups.

We have changed this section and have included discrepancy and impairment scores between the two clinical groups.

(3) Also, in the interest of clarity, I think results for a 'mixed clinical sample' should be removed. Therefore, your comments about the GNT on (p 17) should be amended accordingly, and you could focus the discussion a potentially interesting differential finding for the TBI.

Any references to mixed clinical samples have now been removed and the discussion focuses on the TBI results as advised.

(4) Because the normative sample is crucial for the whole project, you should discuss in more detail limitations relating to how it was obtained. Is there evidence for any self-selection bias? Which type of the normal population does it represent? As it stands at the moment, I feel relating to it as a …UK sample … in the title of the paper, could be easily misunderstood.

We thank the reviewer for this helpful consideration. We have included more critique on the sample and have taken into consideration the regional bias, and looked at issues relating to self-selection bias and other salient demographic issues.

(5) I would be more appropriate to refer to the normative sample as a 'comparison group' and not a 'control group', the latter being a feature of an experimental design.

We have changed any references to control group and have used the term comparison group as advised.

(6) The scoring of Level of Education limits the use of the prediction equations to countries with a similar educational system. Luckily, if the TOPF is being used, then level of education would not be required as a predictor.

Many thanks for this, we have included this as a consideration in the limitations section.

(7) Comment on the effect sizes relating to mean differences between the groups or the pre-morbid to post injury comparisons.

We have commented on this in the penultimate paragraph.

(8) Comment on the reliability of the predictor TOPF as well as the robustness of the statistical findings. You should point out that the estimated pre-morbid scores notably for verbal fluency are only rough estimates in view of the fact that their explained variance by the predictors is only around 20% and at most 40% for the GNT.

We have commented on this in the penultimate paragraph and have made some comparisons to other published work.

Reviewer #2:

<u>Specific points</u>

1) Stepwise multiple regression was used to select models - why? there seems to be no rationale for this if you are generating a prediction equation you get best prediction from using all the predictors

2) Non-normality is mentioned but the regression methods ignore this and thus there is no attempt to improve prediction by using transformations or robust methods (and the nature of non-normality is not described)

   We have now also included information on the residual checks, which include normality, homoscedasticity, outliers, checks for the influence of particular cases, multicollinearity etc. Specifically, Shapiro Wilks is used to check the normality of the residuals, Durbin Watson, Cook's and Mahalnobis statistics are given, as are VIF and Tolerance. Levene's test of equality of variances is used to check heteroscedasticity on the tertiles based on the predicted scores.

3) Skew and kurtosis (or graphical summaries of the key variables) would be relevant if any variables are non-normal; graphical summaries would be superior particularly if there is bimodality etc.

   See answer to question 2 – this is covered in the "data analysis" section and the results.

4) The goal in creating a regression equation is to get good out of sample prediction (in the relevant population) so rather than select models based on statistical significance I'd consider a method such as AIC based around minimising entropy and hence prediction;

rather than present these stepwise in the text it would make more sense to present the models in a summary table

AIC is reported in the results section and we have updated table 4 to include this information.

5) For out of sample prediction R^2 is not sufficient in my view - at the very least I'd want to plot the predictions of one or more competing equations versus the actual scores and if possible look at specific methods commonly used for this purpose such as bland-altman plots

6) Are there any missing data? If so, how were missing data handled? If there were no missing data (e.g., because clinical sample were selected from complete cases) could this have biased sampling?

Many tanks for this comment – we have acknowledged this in the limitations section.

<u>Minor points</u>

1) Too many descriptive statistics are reported to spurious levels of precision (e.g., 46.99 years where 46.7 is more sensible; and never report the SD to more d.p. than the mean)

We have checked and amended this throughout the manuscript.

Estimating verbal fluency and naming ability from the test of premorbid functioning and demographic variables: Regression equations derived from a UK sample

Toni-Marie Jenkinson[1], Steven Muncer[2] , Miranda Wheeler[3], Don Brechin[3], & Stephen Evans[3]

[1]York Teaching Hospital NHS Foundation Trust
[2]Teesside University
[3] South Tees Hospital NHS Foundation Trust

Word count (exc. figures/tables): 4995

*Requests for reprints should be addressed to Stephen Evans, James Cook University Hospital, Marton Road, Middlesbrough, England, TS4 3BW (e-mail: Stephen.Evans@stees.nhs.uk).

Introduction

Neuropsychological assessment aims to identify areas of cognitive impairment resulting from neurological injury or disease. Results from the assessment can be used to inform diagnosis, aid treatment planning and evaluation, generate research data and play a key role in forensic legal proceedings (Lezak, 2004). Identifying whether genuine impairment has occurred is reliant upon obtaining accurate information regarding premorbid ability, which can be defined as "a status that pre-existed some intervening event" (Reynolds, 1997, p. 769). Assessment of premorbid ability helps to establish whether an individual's current level of cognitive functioning reflects a decrement from a previous, higher level of functioning. One approach is to use demographic variables such as age, sex, ethnicity and education to generate estimates of expected premorbid ability based on a demographically similar sample. This approach is based on the well-established relationship between demographic data and intellectual functioning (see Wechsler, 2008). An alternative approach is to use an individual's performance on an oral word reading test to predict IQ, as such tests are considered relatively resistant to the effects of neurological impairment (Nelson & McKenna, 1975). The Test of Premorbid Functioning (TOPF; Wechsler, 2009) is one such example, accounting for 72 percent of the variance in Full Scale IQ (Wechsler, 2009) on the Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV; Wechsler, 2008). Whilst the estimation of intelligence is certainly an important clinical issue it should not be assumed that an individual's functioning in non-intellectual domains is always consistent with their general intellectual level. Despite repeated recommendations to develop predictive methods specifically for cognitive variables other than intelligence (see Crawford, 2004; Franzen, Burgess & Smith-Seemiller, 1997; Vanderploeg & Schinka, 2004) attempts to do so remain elusive.

Tests of verbal fluency often form a central part of neuropsychological assessments, especially in cases involving damage to the left frontal lobe, which is known to be associated with phonemic fluency, or damage to temporal structures, which are known to be involved with semantic fluency (Henry & Crawford, 2004). Closely related to semantic fluency is the skill of naming ability, which relies on input from temporally based semantic networks and frontally based phonemic retrieval processes (Melrose et al., 2009). Many authors describe the influence of age (e.g. Acevedo et al., 2000; Barry, Bates & Labouvie, 2008; Brickman et al., 2005; Zimmermann, Parente, Joanette & Fonseca, 2014) and education (e.g. da Silva, Petersson, Faisca, Ingvar & Reis, 2004; Dursun, Robertson, Bird, Kutcher & Kutcher, 2002; Moraes et al., 2013; Ratcliff et al., 1998) on verbal fluency performance.

Several studies have reported significant correlations between verbal fluency and premorbid intellectual ability as measured by oral word reading. Crawford, Moore and Cameron (1992) found a highly significant correlation of .67 between phonemic fluency and the National Adult Reading Test (NART; Nelson, 1982). Similar to the TOPF, the NART requires the individual to read out loud words with atypical grapheme to phoneme translation, which, due to their irregular spelling, cannot be decoded phonologically using current cognitive resources, meaning that correct pronunciation is believed to rely upon the individual's premorbid knowledge of the word. Harnett, Godfrey and Knight (2004) reported a correlation of .47 between phonemic fluency and NART and a correlation of .33 between semantic fluency and NART. These correlations are unsurprising for two reasons; firstly, tests of verbal fluency and oral word reading are both dependent on verbal ability; and secondly, tests of verbal fluency are closely linked with verbal intelligence (Ardila, Pineda & Roselli, 2000).

Whilst some attempts to develop regression models for the prediction of verbal fluency and naming ability have been made, endeavours to do so remain limited. Initial efforts suggest that tests of oral word reading can account for around 19-22 percent of the variance in phonemic fluency, 11 percent in semantic fluency and 12 percent in naming ability (Harnett, Godfrey & Knight, 2004; Schretlen, Buffington, Meyer & Pearlson, 2005; Testa, Winicki, Pearlson, Gordon & Schretlen, 2009). These figures increase further when adding demographic variables to the equation.

While predictive regression modelling is important it is equally important that equations can be used to identify significant deficits in performance in clinical (cognitively impaired) populations. Crawford et al., (1992) used the NART scores of 142 participants without neurological injury and 38 with an identifiable neurological injury to generate regression based equations to estimate at what discrepancy between predicted and obtained scores would fall at a 0.5 level on a test of verbal fluency. The general method used to determine discrepancy is to identify what score in the normative sample was exhibited by less than 10 percent of the sample and then to compute the standard scores of the difference between obtained and predicted scores and compare this against the t statistic at the 90% confidence level. Knight et al., (2006) used this approach to generate predictive regression equations using the NART and several neuropsychology tests, comprising the Ray Auditory Verbal Learning Test (AVLT), the Trail Making Test (TMT), the Mini-mental state exam (MMSE) and measures of semantic fluency with a large sample of 272 healthy older individuals and provided a useful illustration of how these equations could be used in clinical practice. More recently Testa et al., (2009) developed some regression based norms using discrepant T scores as a way of trying to understand the value of

regression based norms as a means of improving diagnostic accuracy in patient samples. Testa et al., study involved 327 health controls who were tested using a comprehensive neuropsychology battery. Testa et al., reported that although estimated IQ scores improved most of the prediction models, it was unclear whether this would increase diagnostic accuracy.

Our aims are to develop regression equations using the most recent oral word reading test the Test of Premorbid Functioning (TOPF; Wechsler, 2009) to establish how well the TOPF score and typical demographics (age & education level) predicts relative performance on three commonly used verbal neuropsychology tests. We will then use a similar method to Crawford et al., (1992) and Knight at al., (2006) to explore whether the regression based equations can be used to determine impairment in clinical groups. A weakness of previous studies has been the lack of distinct clinical samples. We aim to rectify this by testing our regression equations on two groups of individuals with confirmed neurological diagnoses.

Method

Participants

Normative sample. A total of 119 participants (51 males, 68 females) were recruited from the local community. The participants were recruited via opportunity sampling, using a poster or word of mouth to advertise the project. Participants recruited from the hospital were typically family members of patients attending the neuropsychology department outpatient clinic for routine assessment. All participants resided in the North East of England and North Yorkshire.

The normative sample were free from neurological injury, disorder or disease (e.g. stroke, dementia, learning disability). Participants were excluded from the study if their primary language was not English or if they had an uncorrected visual impairment or hearing loss. All participants were required to be 18 years of age or above at the time of testing. The average age of the sample was 46.9 years (SD = 16.4) ranging from 19 to 87 years. As shown in Table 1 participants were most frequently educated to university level. All participants categorised their ethnicity as 'White British' except from one participant whose ethnicity was categorised as 'Pakistani'.

Clinical samples. A total of 83 patients (40 males, 43 females) were recruited from a local neuropsychology department. Patients were selected if they had been assessed using the Test of Premorbid Functioning (TOPF), FAS, Animal Naming Test (ANT) and Graded Naming Test (GNT) within the last three years and demographic data were available for them. All data was collected in routine clinical practice comprising a wider neuropsychological battery. In 63 percent of cases patients had a diagnosed meningioma (n = 52) and in the remaining 37 percent of cases patients had suffered a Traumatic Brain Injury (TBI; n = 31). The average age of the meningioma sample was 59.4 years (SD = 13.6) and ranged from 28 to 82 years. The average age of the TBI sample was 46.1 years (SD = 15.01) and ranged from 20 to 73 years.

Measures

Test of Premorbid Functioning (TOPF). The TOPF (Wechsler, 2009) provides an estimate of an individual's premorbid intelligence based on their ability to pronounce words with atypical grapheme to phoneme translations. Examination of the internal consistency between performances on each of the items revealed a

Cronbach's alpha of .95 and a split-half reliability of .95. Examination of the concurrent validity of the TOPF is also positive, with correlations ranging from .43 (processing speed) to .71 (verbal comprehension) and .72 (FSIQ) on the WAIS-IV (Wechsler, 2009).

FAS & Animal Naming Test (ANT). The FAS and ANT are both tests of verbal fluency. The FAS is a form of Controlled Oral Word Association Test (COWAT; Benton, Hamsher & Sivan, 1994) that assesses phonemic fluency through verbally producing as many words as possible beginning with a specific letter (F, A and S, respectively), allowing one minute per letter. The ANT is a test of semantic fluency where the examinee is provided with one minute to name out loud as many animals as possible (Goodglass & Kaplan, 1983). Examination of internal consistency between performance on the letters F, A and S reveals high item homogeneity, with a Cronbach's alpha of .83 (Tombaugh, Kozak & Rees, 1999). Test-retest reliability also tends to be high, with a reliability coefficient of .82 for the FAS after a short interval of one to eight weeks (Harrison, Buxton, Husain & Wise, 2000). In the case of the ANT test-retest reliability has been found to be slightly lower, with a reliability coefficient of .56 reported following a one-month interval (Bird, Popadopoulou, Ricciardelli, Rossor & Cipolotti, 2004).

Graded Naming Test (GNT). The GNT (McKenna & Warrington, 1980) is a measure of naming ability. Examinees are presented with 30 items graded in difficulty from well-known items such as "kangaroo" to more difficult items such as "retort".

Analysis of the psychometric properties of the GNT has revealed high levels of test-retest reliability. Using a sample of 188 normal adults Bird et al. (2004) documented a reliability coefficient of .92 ($p < .001$) following a one-month interval.

McKenna and Warrington (1980) documented a correlation of vocabulary and a correlation of .73 and .69 between the GNT and two separate tests of reading indicating good validity.

Level of Education. All participants from both normative and clinical samples were required to state their highest achieved level of education. Education levels could be classed as 'No Exams' at the lowest level followed by 'Certificate of Secondary Education', 'O-Level', 'A-Level', 'Further Education' through to 'Higher Education' at the highest level. Education was coded from 1-6 in hierarchical order from least (No Education) to most (Higher) education.

Procedure

Normative sample. Testing was conducted in private study rooms at a local library, at the neuropsychology outpatient clinic at the host hospital site, or at participants' homes. Participants were provided with a paper-based form collecting demographic information, including details of age, sex, and level of education. Testing was then completed in the following order: (1) TOPF, (2) FAS, (3) ANT, (4) GNT. This order was adhered to throughout data collection in order to match the data collection procedure in the clinical sample. Each participant was tested individually and no participant was paid for their involvement in the study.

Clinical sample. Clinical data was collected by a local neuropsychology department as part of routine clinical practice. Clinical data comprised demographic information (age, sex, level of education), raw test scores (TOPF, FAS, ANT & GNT), and details of the clinical group (Meningioma or TBI).

Data Analysis

In order to detect a medium effect size based on power of 0.8 and α of 0.05 a multiple regression using a set of four independent predictor variables requires a minimum of 84 participants (Cohen, 1992). All data were analysed using the Statistical Package for the Social Sciences Version 23.0 (SPSS; IBM Corp, 2015) and R (R Core Team, 2013). First, we conducted a series of stepwise multiple regression analyses with FAS, ANT and GNT acting as the dependent variable in turn and TOPF score, age, sex and level of education as the predictor variables. Outliers were checked against a criterion of over 3 for the standardised residuals. The influence of cases was assessed using Cook's distance and Mahalanobis distance. Multicollinearity was assessed by the variance inflation factor and the tolerance statistic. Independent errors were assessed using the Durbin-Watson test. Residual plots and the Shapiro-Wilks test were used to assess for normality. For the normative sample three variables were significantly different from normal on a Shapiro Wilks test; age (S-W =.964, $p <$ .01), TOPF (S-W = .960, $p < .01$) and educational record (S-W = .811, $p < .005$). Educational level had significant negative skew (z = -3.02, $p < .05$) and kurtosis problems (z= -2.41), TOPF was negatives skewed (z= -2.01, $p < .05$) and age was significantly platykurtic (z = -2.11, $p < .05$). Residual plots were also examined for heteroscedasticity, which was also checked by splitting the samples and examining the significance of Levene's test of equality of variance on the residuals. The regression models were then compared using Akaike's Information Criterion. Finally, we applied the developed regression equations to the clinical samples in order to examine the discrepancy between predicted and obtained test scores. Paired samples t-tests were used to carry out this comparison and Chi Square was used to compare clinical groups.

Results

Comparison of Samples

Comparison of age revealed a significant difference between the groups ($F(2,199) = 12.70$, $p < .001$; $r = .33$), with the meningioma sample containing older participants ($M = 59.4$) than the normative sample ($M = 46.9$) and the TBI sample ($M = 46.1$). The normative sample scored significantly higher than the clinical samples on educational level ($U = 3077.5$, $z = 4.40$, $p < .001$; $r = .31$), but there was no significant difference between the two clinical groups ($U = 762$, $z = .13$, $p = .89$). Results from a series of one-way independent Analyses of Variance (ANOVA) and Kruskal-Wallis tests showed significant differences ($p < .001$) between samples on each of the tests (see Table 2). Post hoc procedures showed that the normative sample scored significantly higher ($p < .001$) than both the meningioma ($r = .28$) and TBI samples ($r = .30$) on the TOPF. On the FAS the normative sample also scored significantly higher ($p < .001$) than the meningioma ($r = .38$) and TBI samples ($r = .40$). In the case of the ANT the normative sample again scored significantly higher ($p < .001$) than the meningioma ($r = .36$) and TBI samples ($r = .43$). Finally, on the GNT the normative sample also scored significantly higher ($p < .001$) than both the meningioma ($r = .11$) and TBI samples ($r = .25$). No significant differences were found between the meningioma and TBI samples except for on the GNT where the meningioma group scored significantly higher ($p < .001$, $r = .41$).

Multiple Regression Analyses

Stepwise regression is often used when developing predictive equations for predicting performance in clinical samples in order to reduce the number of variables necessary for prediction. In this case we used the backward selection method which

is less likely to suppress the effect of important predictor variables. The predictor variables were not transformed unless it was necessary because of the pattern of residuals. This makes it easier to use the resulting equations. The regression equations are presented in Table 3.

For the FAS the only variable to have a significant impact was the TOPF. The regression equation was significant ($F(1,117) = 28.84$, $p < .001$). There were no outliers using a criterion of over 3 for the standardized residuals, and less than 5% of cases had values over 2. The Durbin Watson statistic of 2.33 suggests that we can assume the errors are independent and there are no problems of autocorrelation. Multicollinearity as assessed by the variance inflation factor and tolerance was not a problem during any of the steps of the regression, and as only one variable was entered could not be a problem at the final stage. The highest Cook's distance was 0.07, and the highest Mahalanobis distance was 6.81, both figures are within acceptable levels. Residual plots were examined for normality and a Shapiro Wilks test was not significant ($SW(119) = .99$, $p = .61$), suggesting that they can be considered normally distributed. Residual plots were also examined for heteroscedasticity and appeared to be homoscedatic. This was also checked by splitting the sample into three groups based on their predicted score, and examining the significance of the Levene's test of equality of variance on the residuals (Levene $(2, 116) = .29$, $p = .75$), which supported the claim for homoscedasticity. Lastly a model with all the variables entered had an Akaike Information Criterion (AIC) score of 897.82 compared to the model with TOPF alone of 894.71, which suggests that the latter is very slightly better.

For the ANT two variables have a significant impact and remain in the regression equation; educational level and TOPF ($F(2,116) = 21.99$, $p < .001$). There

were no outliers over 3 and less than 5% had values of over 2. The variance inflation factor (1.00) and tolerance (.99) were both acceptable for the final model, and through all the steps of the regression which suggests that there are no collinearity problems. The Durbin-Watson statistic of 2.08 was acceptable. The highest Cook's distance was 0.1 and the highest Mahalnonis distance was 10.21, which are both within acceptable levels. Residual plots were examined for normality and the Shapiro Wilks was not significant (SW (119) = .99, p = .60). The Levene's test on the residuals compared across the tertiles of predicted score was not significant (Levene (2,116) = 2.46, p = .12). Lastly we compared models with just TOPF as a predictor, the current model and also a model with all of the variables entered. The AIC's were respectively, 713.12, 689.98 and 692.86, which suggests that the current model is the best fit of these.

For the GNT gender, age and TOPF score were entered in to the equation ($F(3,15) = 26.35$, $p < .001$), with TOPF being the most important variable. The residuals showed an acceptable pattern with no problems with outliers. There were also no collinearity problems with the highest variance inflation factor of 1.01. The Durbin Watson statistic of 2.16 was acceptable. The highest Cook's distance was .15 and although the Mahalanobis distance is higher than the others at 11.48, it is still acceptable for this sample size and number of predictors. Although the residuals appeared to be normal from the plot, the Shapiro Wilks would be significant at the .05 level (SW (119) = .98, $p = .04$). The Kolmogorov-Smirnoff test (KS = .07 $p = .18$) was not significant, however. To investigate the likely effect of the lack of normality of the residuals we conducted a regression with the square root of the reflected TOPF score, which converts it into a normally distributed variable. The residuals from this regression were normal according to the Shapiro Wilks test (SW(119) = .98, $p = .12$).

The effect on the predicted scores was minimal with the highest difference being only 1.63 and all of the other differences under 1. Accordingly, it was decided to keep the equation without transformed variables. The Levene's test of homogeneity of variance for the residuals from the original equation was not significant (Levene (2, 116) = 1.94, p = .15). The AIC for the current model was 611.78 which is better than a model with all predictors entered (612.04) and a model just using the TOPF score.

Predicting scores for the clinical sample

The regression equations were used to predict the scores of the clinical sample on GNT, ANT and FAS. For the TBI group the correlations between obtained and predicted scores on FAS, GNT and ANT were r = .5 (p < .01), r = .38 (p < .05) and r = .24 (p = .2) respectively. For the meningioma group the correlations were r = .4 (p < .01), r = .32 (p = .09) and r = .23 (p =.11) respectively. For the TBI group, FAS scores were significantly lower than the predicted scores (M = 7.98, t(30) = 4.44 p < .001; d = 0.80) as were the ANT scores (M = 4.49 (t(30) = 5.08, p < .001; d = 1.13). For the meningioma group both the FAS scores (M = 5.33, t(48) = 3.29, p < .01; d = 0.52) and the ANT scores (M = 1.81, t(48) = 2.47, p < .05; d = 0.44) were significantly lower than the predicted scores. The GNT scores, however, were not significantly different from the predicted scores in the meningioma group (M = 0.95; t(60) = 1.38, p = .17; d = - 0.18) but they were significantly lower for the TBI group (M = 2.5, t(30) = 2.31, p < .05; d = 0.46). The same pattern of results was found when the Wilcoxon test was used.

In Table 3, we have followed the Knight et al. (2006) example and included a score representing the magnitude of the difference between obtained and predicted scores which corresponds to the 10th percentile (one-tailed). This means,

for example, that an obtained performance more than 13.07 points worse than predicted score on the FAS lies outside of the 90% tolerance interval and suggests impairment. We also computed the standard scores of the difference between obtained and predicted scores and compared this against the t statistic at the 90% confidence level. In this case any score above a value of 1.29 would be a sign of impairment. It should be noted here that we calculated the standard score using the standard error for a new case, which was obtained by using the computer programme described in Crawford and Howell (1998). The standard error of a new individual for the FAS was 10.31 compared to the standard error of the estimate of 10.21; for the ANT it was 4.34 compared to 4.30 and for the GNT it was 3.13 compared to 3.09.

For each of the scales any score over +1.29 would be an indication of possible impairment. For the meningioma sample 12 of the 52 showed impairment and a similar proportion of the TBI group (7 of 31) showed impairment on the FAS. On the ANT significantly more of the TBI sample showed impairment (16/31) compared to the meningioma sample (10/52; χ^2 (1, 83) = 9.47, $p < .05$). A similar pattern was found for the GNT with 11 out of 31 showing impairment from the TBI sample and only 2 out of 52 showing impairment for the meningioma sample (χ^2 (1, 83) = 14.72, $p < .05$). Overall the TBI sample was more likely to show impairment. The use of the standard error for a new individual does not make any difference to the results, as Crawford and Howell (1998) noted, the effect is small with sample sizes over 100.

Discussion

This study aimed to bring attention to the need for developing predictive methods for cognitive abilities other than general intelligence, focusing specifically

on developing regression equations for the prediction of verbal fluency and naming ability using the TOPF as the predictive test. Overall the regression equations presented are similar on terms of the amount of variance accounted for to other equations that have been given for calculating premorbid abilities from other neuropsychological tests (for example, Hartnett et al, 2004; Knight et al., 2006).

Application of the equation to our clinical samples revealed a significant discrepancy between predicted and obtained FAS scores in both the meningioma sample and the TBI sample, with obtained scores significantly lower than scores predicted by the developed equation. In consideration of the number of participants showing severe deterioration, following the approach by Knight et al. (2006), we found 23% of the participants in the meningioma sample obtaining a score with the magnitude of difference between obtained and predicted score corresponding to the 10^{th} percentile (or below). A similar percentage of participants from the TBI sample showed impairment on the FAS with 23% of the sample falling at or below that level. In consideration of semantic fluency, ANT scores were significantly lower than predicted ANT scores in both the meningioma and TBI samples. The number of participants in the clinical samples showing severe deterioration on the ANT reveals that 52% of the participants in the TBI sample showed impairment compared to 19% of the meningioma sample. A similar pattern of results was found when the GNT results were examined, with 35% of the TBI sample showing impairment compared to just 4% of the meningioma sample. Application of the developed equations to our clinical samples revealed that overall the TBI group performed significantly worse than the comparison group and the meningioma group. This is likely due to meningioma differentially affecting naming ability compared to the TBI population where word finding and language impairments are more commonly seen.

Whilst findings from the current study are encouraging, it is important to recognise the limitations of our work. We must highlight that once the clinical sample was divided into subgroups (e.g. for the purpose of analysing clinical presentation, lateralisation of disease/injury, education etc.) the sizes of these sub-samples became small, meaning that deeper exploration was not possible. In order to enhance the power of statistical analysis additional clinical data are required. This would improve our ability to compare patients based on the specific localisation/lateralisation of their brain tumour/injury and draw more accurate conclusions regarding the sensitivity of our equations in identifying impairment amongst this sample rather than using broad clinical groupings as a proxy for the presence or absence of verbal fluency/naming ability impairment. Another limitation relates to the lack of ideal matching between the clinical groups and the comparison group. Examination of the demographics revealed that the meningioma group was not matched with the comparison group for age, education level or TOPF score, and the traumatic brain injury (TBI) group was only matched for age, and not education or TOPF score. The non-clinical sample was younger, better educated and had a higher predicted IQ and therefore predicted score comparisons with the clinical groups may not be fully accurate, although based on previous research and the strength of these findings, we suspect our results would be replicated in well matched groups. It should also be noted that the comparison group was recruited from the North East of England and comprised a combination of university students and individuals attending the hospital site where the study was conducted, generally comprising family members of individuals undertaking neuropsychological examinations. The sample can therefore not be considered fully representative of the UK and is better described as a regional comparison sample, of primarily White British ethnicity. We acknowledge the possibility of self-selection

bias as students or individuals attending the hospital may have an interest in neurological phenomenon and testing. We also acknowledge that data relating to employment status was not collected during the study which means we cannot be certain whether employment status was a factor in choosing to participate in the study. One final consideration is that scoring level of education may somewhat limit the sample to counties with a similar education system to the UK, although as TOPF does not require level of education as a predictor do not feel this unduly limits the used of our predictive equations.

In consideration of the robustness of the results, it is important to note that the predictive equations, although clearly increased the accuracy of the predicted scores over demographics alone, only accounted for 19 percent of the variance on the FAS, 28 on the ANT and 40 percent on the GNT. It is however, in line with previous findings by Knight and Crawford and we feel clinicians will still wish to use these equations to increase the accuracy of their predictions, even if they account for a smaller than ideal amount of variance and therefore cannot be said to provide perfectly accurate scientific prediction.

In conclusion, the findings from the current study provide further evidence of the utility and relative efficiency of using regression equations for predicting scores on tests other than IQ tests using the TOPF. Multiple regression analyses revealed that demographic variables can add a significant contribution to the prediction of verbal fluency and naming ability test scores; however, adding TOPF score to the equation considerably improved prediction beyond that afforded by any of the demographic variables alone. Despite some issues regarding the nature of our control sample, namely it being from a predominantly white background from the north east of the UK, we found that both clinical groups, and particularly the TBI group, comprised a

number of individuals who performed at an impaired level, in so much that their obtained score represented the magnitude of the difference between obtained and predicted scores corresponding to the 10th percentile or below. We believe this indicates that the equations have utility in clinical practice and can be used to discriminate between expected results and severe deterioration in performance. Computing the standard scores of the difference between obtained and predicted scores against the t statistic at the 90% confidence level revealed that a value of 1.29 would be a sign of impairment. Our aim was that these calculations will be used in clinical practice where the GNT, FAS and ANT ore often used alongside the TOPF. These will need to be considered alongside health status and the presence of any psychological disorder, as we were unable to factor all of the salient variables into our equations. Despite some of the shortcomings of the methods presented in this paper, we believe these equations will be of use to clinicians and we have developed an excel programme for clinicians to calculate significant discrepancy scores which can be requested from the authors.

References

Acevedo, A., Loewenstein, D. A., Barker, W. W., Harwood, D. G., Luis, C., Bravo, M., Hurwitz, D. A., Aguero, H., Greenfield, L., & Duara, R. (2000). Category fluency test: Normative data for english-and spanish-speaking elderly. *Journal of the International Neuropsychological Society, 6*(07), 760-769. doi: 10.1017/S1355617700677032

Ardila, A., Pineda, D., & Rosselli, M. (2000). Correlation between intelligence test scores and executive function measures. *Archives of Clinical Neuropsychology, 15*(1), 31-36. doi: 10.1016/S0887-6177(98)00159-0

Barry, D., Bates, M. E., & Labouvie, E. (2008). FAS and CFL forms of verbal fluency differ in difficulty: A meta-analytic study. *Applied Neuropsychology, 15*(2), 97-106. doi: 10.1080/09084280802083863

Benton, A. L., Hamsher, K., & Sivan, A. B. (1994). *Multilingual Aphasia Examination (3rd ed.).* San Antonio, TX: The Psychological Corporation.

Bird, C. M., Papadopoulou, K., Ricciardelli, P., Rossor, M. N., & Cipolotti, L. (2004). Monitoring cognitive changes: Psychometric properties of six cognitive tests. *British Journal of Clinical Psychology, 43*(2), 197-210. doi: 10.1348/014466504323088051

Brickman, A. M., Paul, R. H., Cohen, R. A., Williams, L. M., MacGregor, K. L., Jefferson, A. L., Tate, D. F., Gunstad, J., & Gordon, E. (2005). Category and letter verbal fluency across the adult lifespan: Relationship to EEG theta power. *Archives of Clinical Neuropsychology, 20*(5), 561-573. doi: 10.1016/j.acn.2004.12.006

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

Crawford, J. R. (2004). Psychometric foundations of neuropsychological assessment. In L. H. Goldstein & J. E. McNeil (Eds.), *Clinical Neuropsychology: A Practical Guide to Assessment and Management for Clinicians* (pp. 21-140). Chichester, UK: Wiley.

Crawford, J. R., & Howell, D. C. (1998). Regression equations in clinical neuropsychology: An evaluation of statistical methods for comparing predicted and obtained scores. *Journal of Clinical and Experimental Neuropsychology, 20*(5), 755-762. doi: 10.1076/jcen.20.5.755.1132

Crawford, J. R., Moore, J. W., & Cameron, I. M. (1992). Verbal fluency: A NART-based equation for the estimation of premorbid performance. *British Journal of Clinical Psychology, 31*(3), 327-329. doi: 10.1111/j.2044-8260.1992.tb00999.x

da Silva, C. G., Petersson, K. M., Faísca, L., Ingvar, M., & Reis, A. (2004). The effects of literacy and education on the quantitative and qualitative aspects of semantic verbal fluency. *Journal of Clinical and Experimental Neuropsychology, 26*(2), 266-277. doi: 10.1076/jcen.26.2.266.28089

Dursun, S. M., Robertson, H. A., Bird, D., Kutcher, D., & Kutcher, S. P. (2002). Effects of ageing on prefrontal temporal cortical network function in healthy volunteers as assessed by COWA: An exploratory survey. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 26*(5), 1007-1010. doi: http://dx.doi.org/10.1016/S0278-5846(01)00321-9

Franzen, M. D., Burgess, E. J., & Smith-Seemiller, L. (1997). Methods of estimating premorbid functioning. *Archives of Clinical Neuropsychology, 12*(8), 711-738. doi: 10.1016/S0887-6177(97)00046-2

Goodglass, H., & Kaplan, E. (1983). *Boston Diagnostic Aphasia Examination (BDAE).* Philadelphia: Lea & Febiger.

Harnett, M. A., Godfrey, H. P., & Knight, R. G. (2004). Regression equations for predicting premorbid performance on executive test measures by persons with traumatic brain injuries. *New Zealand Journal of Psychology, 33*(2), 78-87.

Harrison, J. E., Buxton, P., Husain, M., & Wise, R. (2000). Short test of semantic and phonological fluency: Normal performance, validity and test-retest reliability. *The British Journal of Clinical Psychology, 39*(2), 181-191. doi: 10.1348/014466500163202

Henry, J. D., & Crawford, J. R. (2004). A meta-analytic review of verbal fluency performance following focal cortical lesions. *Neuropsychology, 18*(2), 284-295. doi: 10.1037/0894-4105.18.2.284

IBM Corp. Released 2015. IBM SPSS Statistics for Windows, Version 23.p. Armonk, NY: IBM Corp.

Knight, R. G., McMahon, J., Green, T. J., & Skeaff, C. M. (2006). Regression equations for predicting scores of persons over 65 on the rey auditory verbal learning test, the mini-mental state examination, the trail making test and semantic fluency measures. *British Journal of Clinical Psychology, 45*(3), 393-402. doi: 10.1348/014466505X68032

Lezak, M. D. (2004). *Neuropsychological assessment.* New York: Oxford University Press.

McKenna, P., & Warrington, E. K. (1980). Testing for nominal dysphasia. *Journal of Neurology, Neurosurgery, and Psychiatry, 43*(9), 781-788. doi: 10.1136/jnnp.43.9.781

Melrose, R. J., Campa, O. M., Harwood, D. G., Osato, S., Mandelkern, M. A., & Sultzer, D. L. (2009). The neural correlates of naming and fluency deficits in alzheimer's disease: An FDG-PET study. *International Journal of Geriatric Psychiatry, 24*(8), 885-893. doi: 10.1002/gps.2229

Moraes, A. L., Guimarães, L. S., Joanette, Y., Parente, M. A. M. P., Fonseca, R. P., & Almeida, R. M. M. (2013). Effect of aging, education, reading and writing, semantic processing and depression symptoms on verbal fluency. *Psicologia: Reflexão e Crítica, 26*(4), 680-690. doi: 10.1590/S0102-79722013000400008

Nelson, H. E. (1982). *National Adult Reading Test: Test manual.* Windsor, UK: NFER-Nelson.

Nelson, H. E., & McKenna, P. (1975). The use of current reading ability in the assessment of dementia. *British Journal of Social and Clinical Psychology, 14*(3), 259-267.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Ratcliff, G., Ganguli, M., Chandra, V., Sharma, S., Belle, S., Seaberg, E., & Pandav, R. (1998). Effects of literacy and education on measures of word fluency. *Brain and Language, 61*(1), 115-122. doi: 10.1006/brln.1997.1858

Reynolds, C. R. (1997). Postscripts on premorbid ability estimation: Conceptual addenda and a few words on alternative and conditional approaches. *Archives of Clinical Neuropsychology, 12*(8), 769-778. doi: 10.1016/S0887-6177(97)00051-6

Schretlen, D. J., Buffington, A. L., Meyer, S. M., & Pearlson, G. D. (2005). The use of word-reading to estimate "premorbid" ability in cognitive domains other than intelligence. *Journal of the International Neuropsychological Society, 11*(6), 784-787. doi: 10.10170S1355617705050939

Testa, S. M., Winicki, J. M., Pearlson, G. D., Gordon, B., & Schretlen, D. J. (2009). Accounting for estimated IQ in neuropsychological test performance with regression-based techniques. *Journal of the International Neuropsychological Society, 15*(6), 1012-1022. doi: 10.1017/S1355617709990713

Tombaugh, T. N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Archives of Clinical Neuropsychology, 14*(2), 167-177. doi: http://dx.doi.org/10.1016/S0887-6177(97)00095-4

Vanderploeg, R. D., & Schinka, J. A. (2004). Estimation of premorbid cognitive abilities: Issues and approaches. In J. H. Ricker (Eds.), *Differential Diagnosis in Adult Neuropsychological Assessment* (pp. 27-65). New York: Springer Publishing Company, Inc.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale–Fourth Edition.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2009). *Advanced Clinical Solutions for the WAIS-IV and WMS-IV*. San Antonio, TX: The Psychological Corporation.

Zimmermann, N., Parente, M. A. M. P., Joanette, Y., & Fonseca, R. P. (2014). Unconstrained, phonemic and semantic verbal fluency: Age and education

effects, norms and discrepancies. *Psicologia: Reflexão e Crítica, 27*(1), 55-63. doi: 10.1590/S0102-79722014000100007

Table 1. Age & education levels across the normative and clinical samples

	Sample		
	Normative (*N* = 119)	Meningioma (*N* = 52**)	TBI (*N* = 31)
Level of Education**			
No Exams	24 (20.2%)	20 (38.5%)	10 (32.3%)
CSE*	4 (3.4%)	8 (15.4%)	3 (9.7%)
O-Level	10 (8.4%)	6 (11.5%)	11 (35.5%)
A-Level	13 (10.9%)	0	1 (3.2%)
Further Education	30 (25.2%)	7 (13.5%)	4 (12.9%)
Higher Education	38 (31.9%)	9 (17.3%)	2 (6.5%)
Age			
M	46.9	59.4	46.1
S.D.	16.4	13.6	15.0

* CSE = Certificate of Secondary Education, ** 2 cases missing

Table 2. Comparison of test scores obtained by the normative and clinical samples

	Normative (N = 119, 59%)		Meningioma (N = 52, 26%)		TBI (N = 31, 15 %)			
	M	S.D.	M	S.D.	M	S.D.	*F*	*H*
TOPF	48.37	12.41	40.44	12.71	37.52	14.70	-	21.69*
FAS	40.59	11.35	32.16	12.34	28.19	11.56	18.65*	-
ANT	21.29	5.01	16.96	5.18	15.52	4.79	23.45*	-
GNT	20.35	3.96	21.20	4.25	16.35	6.50	-	13.08*

* $p < .001$

No significant differences were found between the number of males and females in the normative and clinical samples ($\chi^2(1) = 0.56$, $p = .48$).

Table 3. Regression equations for predicting FAS, ANT and GNT scores

	Regression weights								
	Intercept	Age	Gender	Education	TOPF	R	Adj R^2	SEest	10pctile
FAS	20.91	-	-	-	.41	.45	.19	10.21	13.07
ANT	21.10	-	-	-0.13	0.13	.52	.28	4.30	5.50
GNT	9.36	0.09	-1.50	-	0.16	.64	.41	3.08	3.94