Dementia screening accuracy is robust to premorbid IQ variation: Evidence from the Addenbrooke's Cognitive Examination-III and the Test of Premorbid Function

Joshua Stott[a*], Katrina Scior[a], William Mandy[a], Georgina Charlesworth[a],

[a] Address and affiliation of all authors:

Research department of clinical, educational and health Psychology

University College London,

London, UK,

WC1E 6BT

[*] Corresponding Author: Address as above; Telephone, 02076795950; Fax, 02079161989; E mail address, j.stott@ucl.ac.uk

Running title: Dementia screening accuracy and premorbid IQ

ABSTRACT

Background: Scores on cognitive screening tools for dementia are associated with premorbid IQ. It has been suggested that screening scores should be adjusted accordingly. However, no study has examined whether premorbid IQ variation affects screening accuracy.

Objective: To investigate whether the screening accuracy of a widely used cognitive screening tool for dementia, the Addenbrooke's cognitive examination-III (ACE-III), is improved by adjusting for premorbid IQ.

Methods: 171 UK based adults (96 memory service attendees diagnosed with dementia and 75 healthy volunteers over the age of 65 without subjective memory impairments) completed the ACE-III and the Test of Premorbid Function (TOPF). The difference in screening performance between the ACE-III alone and the ACE-III adjusted for TOPF was assessed against a reference standard; the presence or absence of a diagnosis of dementia (Alzheimer's disease, vascular dementia or others).

Results: Logistic regression and receiver operating curve analyses indicated that the ACE-III has excellent screening accuracy (93% sensitivity, 94% specificity) in distinguishing those with and without a dementia diagnosis. Although ACE-III scores were associated with TOPF scores, TOPF scores may be affected by having dementia and screening accuracy was not improved by accounting for premorbid IQ, age or years of education.

Conclusion: ACE-III screening accuracy is high and screening performance is robust to variation in premorbid IQ, age and years of education. Adjustment of ACE-III cut-

offs for premorbid IQ is not recommended in clinical practice. The analytic strategy used here may be useful to assess the impact of premorbid IQ on other screening tools.

Key words; sensitivity and specificity, dementia, screening, neuropsychology

INTRODUCTION

Individuals screened for dementia vary widely in premorbid IQ, with implications for the interpretation of performance on cognitive screening tests[1]. Premorbid IQ in dementia is hard to measure as it requires a test that is both highly associated with current IQ in individuals without dementia and robust to the effects of dementia[2]. Reading of irregular English words meets the first criterion[3] and despite some controversy over whether it meets the second[4, 5] it is the most established methodology for premorbid IQ measurement[2]. There are a number of English irregular word reading tests, of which the Test of Premorbid Function (TOPF)[6] is the most recent. The TOPF is perhaps the gold standard as it was developed and co-normed with the most widely used measure of IQ, the Wechsler Adult Intelligence Scale IV[7]

The association of estimated Premorbid IQ based on irregular word reading with one widely used and recommended[8] screening tool, the Addenbrooke's Cognitive Examination III (ACE-III)[9] is unknown. Premorbid IQ is, however, strongly associated with other screening tools such as the Mini Mental State Examination (MMSE)[1, 10] and Montreal Cognitive Assessment (MocA) [11, 12]. As a consequence, it has been suggested that scores on cognitive screening tests should be adjusted to account for premorbid IQ[11]. Despite this, the critical issue of whether adjustment increases the accuracy of screening tools and thus the clinical utility of such adjustment is unknown.

Consequently, the current study has three novel aims. The first aim is to evaluate the association of ACE-III with TOPF performance in individuals without dementia and

thus the potential influence of premorbid IQ on ACE-III performance. Given the controversy over the impact of dementia on premorbid IQ, the second aim is to evaluate the impact of dementia on TOPF performance and thus the utility of the TOPF as a measure of premorbid IQ in dementia. The third and primary aim is to evaluate whether adjusting ACE-III scores for TOPF premorbid IQ improves screening accuracy and is thus clinically useful.

MATERIALS AND METHODS

Design

The study assessed screening tests (index tests) against a criterion (reference standard). Consequently it is written up according to Standards for Reporting of Diagnostic Accuracy (STARD)[13] and Standards for Reporting of Diagnostic Accuracy in Dementia[14] guidelines.

Participants

Eligibility criteria

All participants were fluent in English and had capacity to consent. Exclusion criteria included a current diagnosis of a DSM-IV depressive disorder, anxiety disorder, bipolar disorder, substance related disorder, schizophrenia or other psychotic disorder, diagnosed intellectual disability and significant uncorrected sensory deficits. As this study was the first from a dataset which will also be used to examine the relationship of neuropsychological functioning to Cognitive Behavioural Therapy (CBT), participants reporting current or previous experience of CBT were excluded.

Recruitment procedures

Data on ACE-III and TOPF were prospectively collected. The dementia group constituted a consecutive referrals sample of 102 people with mild dementia (last MMSE score $\geq$ 24 or equivalent on other cognitive screen[15]). Dementia was diagnosed according to consensus criteria[16-19] by a psychiatrist led, multi-disciplinary memory clinic. All clients had cognitive assessment, with extent of assessment driven by client need as per British Psychological Society guidelines[20]. Results of cognitive screening tests (ACE-III, Addenbrooke's Cognitive Examination Revised (ACE-R) [21], MMSE, MoCA) from an initial diagnostic interview were discussed in the multidisciplinary team with interpretation guided by experienced clinical psychologists specialising in neuropsychology. Where diagnosis was unclear, a more extensive neuropsychological assessment was conducted by a clinical psychologist. In determining client functioning, client and informant report were used. Where informant report was unavailable and functional status was unclear, occupational therapy assessment was used to clarify. Diagnostic subtypes included Alzheimer's disease[17], vascular dementia[19], mixed dementia[17], dementia in Parkinson's disease[16] and Frontotemporal dementia[18]. When criteria were not met, a diagnosis of dementia not otherwise specified was made according to ICD-10 criteria[22]. Intermediate diagnoses such as possible Alzheimer's disease were not included. The control group constituted a convenience sample of 75 healthy volunteers over the age of 65 without a diagnosis of dementia (determined through self-report) and not reporting subjective memory problems. They were recruited by advertisement in community groups and from the Join Dementia Research database[23]. All participants from both dementia and control groups gave informed consent to participate in the study. Work was conducted in compliance with ethical guidelines on human experimentation[24] Ethical approval was given by NRES

Committee London – City Road & Hampstead (REC Reference 14/LO/0554). Sample size estimation for the main logistic regression analysis[25, 26], suggested a minimum of 100 participants in total was required. Data were collected between July 2014 and July 2016.

Test methods

Demographic information was obtained verbally from all participants and for the dementia group was verified via electronic healthcare records. The presentation order of the TOPF and the ACE-III was randomised across participants to minimise order effects. Measures were administered in clinics or participants' homes by five psychology graduates trained in administration and scoring of the TOPF and ACE-III and supervised by a clinical psychologist with a postgraduate diploma in clinical neuropsychology.

Test materials

Index tests

We compared two index tests in terms of screening performance; the ACE-III alone and the ACE-III adjusted for TOPF scores. The index tests were not incorporated within the reference standard so as to avoid diagnostic circularity.

Reference standard

Performance on index tests was evaluated against a reference standard of whether the participant belonged to the dementia group or the control group as defined above. Reference standard assessors were blind to index test results but not vice versa.

ACE-III

The ACE-III is a validated 'pen and paper' cognitive screening tool for dementia[9] covering five cognitive domains including memory, language, orientation and attention, verbal fluency and visuospatial abilities, with a maximum score of 100. The recommended cut-off score for screening cognitive impairments related to dementia is 87/88[9]. Given mixed findings with optimal cut-off varying from 76/77[27] to 87/88[9] our cut-off score was exploratory and devised using receiver operating characteristic (ROC) analysis to to calculate an optimal cut-off score for which the Youden's index (sensitivity+specificity-1) was maximized [28]

ACE-III adjusted for TOPF

The TOPF[6] involves reading up to 70 irregular English words. The raw score (total number correct) can be converted into two estimates of premorbid IQ. The unadjusted premorbid IQ is based on published tables developed through regression with TOPF alone as a predictor of IQ. The adjusted premorbid IQ is obtained through entering TOPF score along with age, gender and years of education into a regression equation. In support of construct validity, both unadjusted and adjusted TOPF premorbid IQ show strong associations with current measures of IQ, are reliable over time and robust to effects of brain injury[6].

In order to derive a single index test which optimally combined both the TOPF and ACE-III scores, we used the TOPF and ACE-III as predictors in a logistic regression model[29]. The performance of this model in correctly identifying those with dementia was compared against the ACE-III alone using recommended methodologies[29] as described below. As this index was novel, cut-offs were not

pre-defined, but based on the value of scores derived from the logistic regression that maximised the Youden index in a ROC analysis[30].

Statistical methods:

Data were analysed using R[31] and the pROC[32], BinomTools[33], QuantPsych[34] and BaylorEdPsych[35] packages. In examining between group demographic differences, all continuous variables were assessed for parametric assumptions. Where these were met, t-tests were conducted, where not; Mann Whitney U tests were used. For categorical variables Chi Squared or Fisher's exact test were used, depending on minimum cell counts[36] Where necessary to quantify uncertainty, 95% confidence intervals were calculated.

For aims one and two regarding the association of ACE-III with TOPF and the impact of having a dementia diagnosis on TOPF performance, we used hierarchical multiple linear regression to statistically control for potential confounding variables of age, gender and years of education, all of which can influence TOPF performance[6]. These potential confounders were entered in a first block with variables of interest added in a second block. For all regression models, outliers with undue influence on coefficients were investigated, and where necessary, removed. Where assumptions of regression[36] were not met, bootstrapped bias-corrected accelerated confidence intervals were generated to increase model robustness[36].

To assess the primary aim, we compared the prediction of dementia status using two binomial logistic regression models; one model with ACE-III alone as a predictor and the other with TOPF and ACE-III as predictors. We assessed performance of

individual predictors using beta coefficients, Wald statistics and odds ratios. We assessed overall model effect size using Nagelkerke's R2 and compared overall model performance using the Bayesian information criterion (BIC)[29]. To provide screening accuracy statistics for both models and thus an indication of clinical utility, we used the predicted probability of dementia from each model in ROC analyses to derive optimal cut-offs with corresponding numbers of true and false negatives and positives. We also obtained sensitivity, specificity and Area Under The Curve (AUC) values. The unadjusted TOPF premorbid IQ was used in all analyses due to multicollinearity[36] in regression equations containing the adjusted TOPF score. However, we re-ran the main analysis again with the adjusted TOPF score to ensure this did not alter results. Some participants were administered ACE-III (n=41) and others, the ACE-R (n=9) in determining their original diagnoses. This may have artificially inflated ACE-III sensitivity and specificity and could have an impact on our main results regarding TOPF utility in diagnosis. To assess this, we re-ran our main analyses twice, first excluding participants where ACE-III was used in diagnosis and second excluding participants where ACE-R or ACE-III were used in diagnosis.

RESULTS:

Participant flow

345 people were initially approached to take part in the study. Of 285 potentially eligible, 179 participants took part. In the control and dementia sample those potentially eligible did not differ from participants in gender (X^2 = 0.006 and 0.002 respectively, p>0.05) or age (t= 0.77, and -0.06 respectively p>0.05). Those in the dementia sample did not differ in diagnostic subtype (whether they had Alzheimer's disease or another dementia type) either (X^2 =0.2, p>0.05). Figure 1 shows a modified

STARD[13] flow diagram detailing flow of participants through the study and reasons for exclusion. Eight participants had missing data on one or more measures. Data were missing completely at random (Little's MCAR test $p<0.05$) and < 5%, so was removed listwise as recommended by Graham [37]. Thus, the final analysis included 171 participants (96 with dementia and 75 controls).

-Insert figure 1 about here –

Demographic and clinical characteristics

Table 1 shows summary statistics for demographics as well as ACE-III (which in the dementia group indicates level of impairment) and TOPF scores for both groups. The dementia group were significantly older, had significantly fewer years' education and lower ACE-III and TOPF scores than the control group. Gender and ethnicity did not significantly differ between groups.

- Insert table 1 about here -

Analyses for aims 1 and 2

All analyses were hierarchical multiple linear regressions. Standardised betas (β) with bootstrap bias corrected accelerated 95% confidence intervals (CI)s for all significant predictors (at $p<0.05$) *after adjustment for all other predictors are reported for variables of interest and potential confounders (age, gender and years of education).*

1. Association between TOPF and ACE-III scores in those without dementia

Our results suggest that ACE-III scores are highly and independently associated of with premorbid IQ scores: In controls, higher TOPF premorbid IQ scores independently predicted higher ACE-III scores (β=0.65, CI 0.38 to 0.81). This was the case even when controlling for gender, age and education, with older age (β = -0.24, CI -0.48 to -0.05) the only other significant independent predictor of higher ACE-III.

2a) Impact of dementia on TOPF performance

Our results suggest that TOPF performance is associated with and may be affected by dementia: In an analysis conducted on the whole sample, having a dementia diagnosis was a significant independent predictor of poorer TOPF performance (β=-0.45, CI -0.56 to -0.33). This was the case even when controlling for the significant prediction of TOPF performance by years of education (β=0.38, CI 0.25 to 0.48) and age (β=0.17, CI 0.06 to 0.29).

2b) Degree of cognitive impairment and TOPF performance

Our results indicate that TOPF performance may be lower in those with increased cognitive impairment (as measured by ACE-III scores) independent of any differences in age and education and consequently suggest that TOPF may be more affected with greater cognitive impairment : The impact of degree of cognitive impairment (ACE-III score) and dementia subtype on TOPF performance was assessed in the dementia sample. Lower ACE-III score was predictive of lower TOPF score (β=0.49, CI 0.31 -0.66) even when controlling for the fact that lower age (β=0.19, CI 0.04 -0.33) and fewer years of education (β=0.34, CI 0.18 -0.48) were

also significant predictors of lower TOPF scores. Type of dementia (Alzheimer's Disease vs. any other dementia subtype) was not associated with TOPF performance.

Analyses for primary aim: Does adjusting ACE-III scores by TOPF scores improve screening accuracy?

Preliminary analyses

No adverse events were reported in using index or reference standard tests. There were no indeterminate results on index testing. The index test was performed after the reference standard and median (IQR) time difference was 145 (323.5) days.

Logistic regression analysis

Our results suggest that ACE-III screening accuracy is robust to variance in premorbid IQ, age and years of education: In the binomial logistic regression model with ACE-III score alone as a predictor of dementia status, and the model with ACE-III and TOPF scores as predictors, lower ACE-III score was the only significant ($p<0.05$) predictor of increased odds of dementia (see table two for regression coefficients).

As age and education relate to premorbid IQ[6] and differed between the dementia and control groups we ran two further binomial logistic regression models, one with ACE-III, age and years of education as predictors and one with the TOPF in addition to these predictors. Across both models, ACE-III was the only significant predictor (see Table 2 for regression coefficients). The significance of ACE-III across models and the lack of significance of other predictors implies that if individuals have the same ACE-III score but differ in TOPF score, age, gender, or years of education,

there is no change in the odds of them having dementia. The increase in BIC and similarity of R^2 across models suggests that the predictive power of a model containing the ACE-III alone is not enhanced by the other variables included here (see Table 2 for model R2 (effect size) and BIC values).

- Insert table 2 about here -

Screening accuracy metrics for all models

Numbers correctly and incorrectly classified by each model are given in Table 3 ROC analyses indicated that the optimal ACE-III cut-off was 88.5 with an associated sensitivity of 93% and specificity of 94%. Discrimination metrics (sensitivity, specificity, area under the curve) from this and ROC analyses based on the predicted probabilities from all logistic regression models are given in Table 2. There was no difference between model AUCs ($p>0.05$), further supporting our conclusion that ACE-III is robust to variance in premorbid IQ and demographics.

- Insert table 3 about here -

Use of the recommended ACE-III cut-off and adjusted TOPF premorbid IQ

Our results suggest that using the recommended ACE-III cut-offs and the adjusted TOPF premorbid IQ index does not alter our findings: The optimal ACE-III cut-off in our sample was 88.5. We assessed whether using the previously recommended cut-off of 87/88[9] would change results, which it did not; results given in Tables 2 and 3 show overlap of confidence intervals with all other models. Similarly, re-running the TOPF and ACE-III model with the adjusted rather than unadjusted TOPF premorbid

IQ score as a predictor did not change results. ACE-III remained the only significant predictor (β=0.34, Z=-5.80, p<0.001, OR=0.71; Model R^2=0.81, BIC=92.42, AUC=0.97) and AUC test against other models was non-significant (p>0.05).

Analysis excluding those who had ACE-III or ACE-R as part of their diagnosis.
The main analysis was rerun twice, once excluding those who were administered ACE-III as part of their initial diagnostic process (dementia n = 61) and again excluding those administered the ACE-III or ACE R in diagnosis (dementia n =52). This produced very similar results to our original findings. The main finding of robustness of the ACE-III to variation in premorbid IQ, education, or age was replicated in both analyses and ACE-III was once again the only significant predictor of dementia status in both analyses with very similar odds ratios to the original analysis (odds ratios of 0.74, 95% CI 0.66-0.81 and 0.75, 95% CI 0.67 – 0.82 respectively). In terms of diagnostic accuracy metrics, sensitivity of diagnosis using ACE-III in the analyses was 93.1%, (95% CI, 82.75-98.28) and 92% (95% CI, 81.63-97.96) respectively and specificity was 93.33%, (95% CI 86.66-98.67) and 93.33% (95% CI, 85.33-98.67) respectively. These point values for sensitivity and specificity were around 0.5%-1.5% lower than in the original analyses. However, the overlap of 95% confidence intervals with original point values suggest no significant difference between these and original findings and as stated above the main finding of lack of TOPF utility in adding to screening accuracy was unchanged.

DISCUSSION

This study is the first to evaluate whether adjusting a dementia-screening tool for premorbid IQ improves screening accuracy. We found that a recommended and

widely used screening tool (ACE-III) had excellent screening accuracy and was robust to variation in premorbid IQ (as measured by TOPF), age, and years of education. Our results do not support the proposed clinical practice[11] of adjusting screening cut-offs for these factors.

A subsidiary aim of our study was to examine the association of ACE-III with premorbid IQ. It has been argued that screening cut-off scores should be adjusted as a corollary of a strong association between premorbid IQ and screening[11]. Our study mitigates against this given that we found such an association but no increase in screening accuracy. One reason for this apparent contradiction may have been the excellent screening performance of the ACE-III alone, which left little room for improvement.

A second subsidiary aim was to preliminarily evaluate the impact of dementia on a gold standard measure of premorbid IQ (TOPF) through comparison to controls. Although our design does not allow inference of causality, our results indicate that possibly TOPF scores may be affected by dementia. Thus one reason for lack of impact on screening accuracy could be that the TOPF is a poorer measure of premorbid IQ than other irregular word reading tests [2].We think this is unlikely as the TOPF is highly associated with the most widely used current IQ test[6] and there is evidence that any effect of dementia on the TOPF is much less than the effect of dementia on tests of current cognitive function[6]. In support of this, the TOPF in our sample appeared less associated with the ACE-III in those with dementia than those without. Furthermore, making TOPF more robust to dementia impact by using a derived premorbid IQ score adjusted for demographics did not change our results.

Additionally, our findings are similar to those for other premorbid IQ measures in indicating that impact of dementia may be mitigated by less severe cognitive impairment[5] albeit within the limited range afforded by our mild dementia sample. This is important as it is in mild cases and in the clarification of potential false negatives that premorbid IQ measurement is particularly clinically useful[2]. Finally, the lower TOPF in the dementia sample may reflect a difference in underlying IQ between those with and without mild dementia as suggested by the cognitive reserve hypothesis[38]. Consequently, while our results potentially indicate that dementia may have some effect on TOPF scores, they do not do so conclusively and this is unlikely to fully account for our main results, which would likely be similar for other measures of premorbid IQ.

Generalisability of our results to clinical practice is enhanced by the fact that, unlike many studies, having a carer was not an inclusion criterion. Our dementia sample was also very similar to UK estimated prevalence in diagnostic subtype breakdown[39]. Additionally, our control participants and those with dementia were similar in gender and age to eligible non-participants. Participants with dementia were also similar in diagnostic breakdown to eligible non-participants, suggesting no selection bias on these domains.

The robustness of our conclusions is increased by the fact that changing ACE-III cut-offs did not affect results. Thus while the optimal ACE-III cut-off in our sample was 88.5, we would suggest continued clinical use of the proposed cut-off of 87/88[9]. A further strength of the study was that we used recommended methods, for comparing the incremental value of adding new markers to a diagnostic model[30]

Limitations worth noting include the fact that the control sample was younger and more educated than the dementia sample. Although we controlled for this statistically and our main finding was unchanged when age and years of education were added to regression models, it should be addressed by sample matching in future. Severity of dementia may also have been inaccurately measured in our sample, where a third have a non-Alzheimer's dementia, as the ACE-R has not been validated for people with non-Alzheimer's dementias and early cognitive symptoms of such dementias may not be accurately assessed by the MMSE. Additionally, we did not record interventions occurring between reference standard and index assessment, which could have reduced dementia severity[14] and thus made dementia more difficult to detect. We did not record alternative diagnoses in either group, which may also have affected index test performance[14], potentially also making dementia more difficult to detect. Importantly, absence of dementia diagnosis in the control group was self-reported and not confirmed by cognitive tests, with consequent potential undiagnosed dementia in this group. This could inaccurately inflate the false positive rate of index tests. However, all of these issues should make screening performance worse and the high accuracy of ACE-III found in our study may indicate that they did not affect our results. Finally, index assessment was done some time after the reference standard and there could have been deterioration in participants' cognitive function[14], making it more severe and dementia easier to detect. This, as well as the fact that assessors were not blind to dementia status of the participants [40] perhaps accounts for the fact that our screening accuracy values for ACE-III were slightly higher than have been found before[9, 27]. Screening accuracy may also have been slightly inflated by the use, in some cases, of the ACE-III or ACE-R in the diagnostic process.

In summary, we found that screening accuracy of ACE-III was robust to variation in premorbid IQ and demographic variables and our study supports use of the recommended 87/88 cut-off[9]. The TOPF may not be sufficiently robust to dementia but this seems unlikely to fully account for our results and may not be different to other premorbid IQ measures. Importantly, association of screening tools with premorbid IQ is not sufficient to recommend adjusting screening cut-offs, at least in a high IQ sample using the ACE-III. To increase generalisability of findings, future research should assess the impact of premorbid IQ on screening accuracy of other tools (e.g. MMSE) in those with lower intellectual capacity with the analytic strategy used here.

Acknowledgements (including sources of support)

Funding: This study was completed as part of a fellowship awarded to Joshua Stott by the Alzheimer's society. Grant number 236 (AS-CTF-14-005). This funding source had no involvement in the study design, collection, analysis and interpretation of data, writing the manuscript and in the decision to submit the manuscript for publication.

Other Acknowledgements: The authors would like to acknowledge the research participants who generously gave their time to take part in this research and also Noor Habib, Lucy Gore, Els Chadwick, Joanne Sweeney and Catherine Bousfield for their role in data collection.

Conflict of interest: The authors have no conflict of interest to report.

References

[1] Starr JM, Lonie J (2007) The influence of pre‐morbid IQ on Mini‐Mental State Examination score at time of dementia presentation. *Int J Geriatr Psychiatry* **22**, 382-384.

[2] Crawford JR, Stewart L, Parker D, Besson J, Cochrane R (1989) Estimation of premorbid intelligence: Combining psychometric and demographic approaches improves predictive accuracy. *Pers Individ Dif* **10**, 793-796.

[3] Crawford JR, Deary IJ, Starr J, Whalley LJ (2001) The NART as an index of prior intellectual functioning: a retrospective validity study covering a 66-year interval. *Psychol Med* **31**, 451-458.

[4] Dykiert D, Deary IJ (2013) Retrospective validation of WTAR and NART scores as estimators of prior cognitive ability using the Lothian Birth Cohort 1936. *Psychol Assess* **25**, 1361.

[5] McFarlane J, Welch J, Rodgers J (2006) Severity of Alzheimer's disease and effect on premorbid measures of intelligence. *Br J Clin Psychol* **45**, 453-464.

[6] Wechsler D (2011) Test of Premorbid Functioning. UK Version (TOPF UK). *UK: Pearson Corporation*.

[7] Wechsler D, Coalson DL, Raiford SE (2008) *WAIS-IV: Wechsler adult intelligence scale*, Pearson San Antonio, TX.

[8] Larner AJ, Mitchell AJ (2014) A meta-analysis of the accuracy of the Addenbrooke's Cognitive Examination (ACE) and the Addenbrooke's Cognitive Examination-Revised (ACE-R) in the detection of dementia. *Int Psychogeriatr* **26**, 555-563.

[9] Hsieh S, Schubert S, Hoon C, Mioshi E, Hodges JR (2013) Validation of the Addenbrooke's Cognitive Examination III in frontotemporal dementia and Alzheimer's disease. *Dement Geriatr Cogn Disord* **36**, 242-250.

[10] Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* **12**, 189-198.

[11] Alves L, Simões MR, Martins C, Freitas S, Santana I (2013) Premorbid IQ influence on screening tests' scores in healthy patients and patients with cognitive impairment. *J Geriatr Psychiatry Neurol*, 0891988713484194.

[12] Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, Cummings JL, Chertkow H (2005) The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* **53**, 695-699.

[13] Meyer GJ (2003) Guidelines for reporting information in studies of diagnostic test accuracy: the STARD initiative. *J Pers Assess* **81**, 191-193.

[14] Noel-Storr AH, McCleery JM, Richard E, Ritchie CW, Flicker L, Cullum SJ, Davis D, Quinn TJ, Hyde C, Rutjes AW (2014) Reporting standards for studies of diagnostic test accuracy in dementia The STARDdem Initiative. *Neurology* **83**, 364-373.

[15] Law E, Connelly PJ, Randall E, McNeill C, Fox HC, Parra MA, Hudson J, Whyte LA, Johnstone J, Gray S (2013) Does the Addenbrooke's Cognitive

Examination‐revised add to the Mini‐Mental State Examination in established Alzheimer disease? Results from a national dementia research register. *Int J Geriatr Psychiatry* **28**, 351-355.
[16] Emre M, Aarsland D, Brown R, Burn DJ, Duyckaerts C, Mizuno Y, Broe GA, Cummings J, Dickson DW, Gauthier S (2007) Clinical diagnostic criteria for dementia associated with Parkinson's disease. *Mov Disord* **22**, 1689-1707.
[17] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R (2011) The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & dementia* **7**, 263-269.
[18] Neary D, Snowden JS, Gustafson L, Passant U, Stuss D, Black Sa, Freedman M, Kertesz A, Robert P, Albert M (1998) Frontotemporal lobar degeneration A consensus on clinical diagnostic criteria. *Neurology* **51**, 1546-1554.
[19] Román GC, Tatemichi TK, Erkinjuntti T, Cummings JL, Masdeu J, Garcia Ja, Amaducci L, Orgogozo J-M, Brun A, Hofman A (1993) Vascular dementia Diagnostic criteria for research studies: Report of the NINDS‐AIREN International Workshop. *Neurology* **43**, 250-250.
[20] Guss R, Middleton J, Beanland T, Slade L, Moniz-Cook E, Watts S, Bone A (2014) *Clinical Psychology in the Early Stage Dementia Care Pathway*, London: The British Psychological Society.
[21] Mioshi E, Dawson K, Mitchell J, Arnold R, Hodges JR (2006) The Addenbrooke's Cognitive Examination Revised (ACE‐R): a brief cognitive test battery for dementia screening. *Int J Geriatr Psychiatry* **21**, 1078-1085.
[22] World Health Organization (1993) *The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research*
[23] Join dementia research, https://www.joindementiaresearch.nihr.ac.uk/ (date last accessed 7/11/2016),
[24] Association WM (1975) *Declaration of Helsinki: recommendations guiding medical doctors in biomedical research involving human subjects*, World Medical Association.
[25] Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR (1996) A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* **49**, 1373-1379.
[26] Scott Long J (1997) Regression models for categorical and limited dependent variables. *Advanced quantitative techniques in the social sciences* **7**.
[27] Cheung G, Clugston A, Croucher M, Malone D, Mau E, Sims A, Gee S (2015) Performance of three cognitive screening tools in a sample of older New Zealanders. *Int Psychogeriatr* **27**, 981-989.
[28] Zhou X-H, McClish DK, Obuchowski NA (2009) *Statistical methods in diagnostic medicine*, John Wiley & Sons.
[29] Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW (2010) Assessing the performance of prediction

models: a framework for some traditional and novel measures. *Epidemiology (Cambridge, Mass.)* **21**, 128.
[30] Kattan MW (2003) Judging new markers by their ability to improve predictive accuracy. *J Natl Cancer Inst* **95**, 634-635.
[31] R Core Team (2013) R Foundation for Statistical Computing.
[32] Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M (2011) in *BMC Bioinformatics*, p. 1.
[33] Hansen MK, Haubo R, Christensen B (2011).
[34] Fletcher TD (2008) in *R package version*.
[35] Beaujean AA, Beaujean MAA (2012).
[36] Field A (2009) *Discovering statistics using SPSS*, Sage publications.
[37] Graham JW (2009) Missing data analysis: making it work in the real world. *Annu Rev Psychol* **60**, 549-576.
[38] Stern Y (2002) What is cognitive reserve? Theory and research application of the reserve concept. *J Int Neuropsychol Soc* **8**, 448-460.
[39] Prince M, Knapp M, Guerchet M, McCrone P, Prina M, Comas-Herrera A, Wittenberg R, Adelaja B, Hu B, King D (2014) in *Alzheimer's Society, London*.
[40] Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJM, Gavaghan DJ, McQuay HJ (1996) Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* **17**, 1-12.

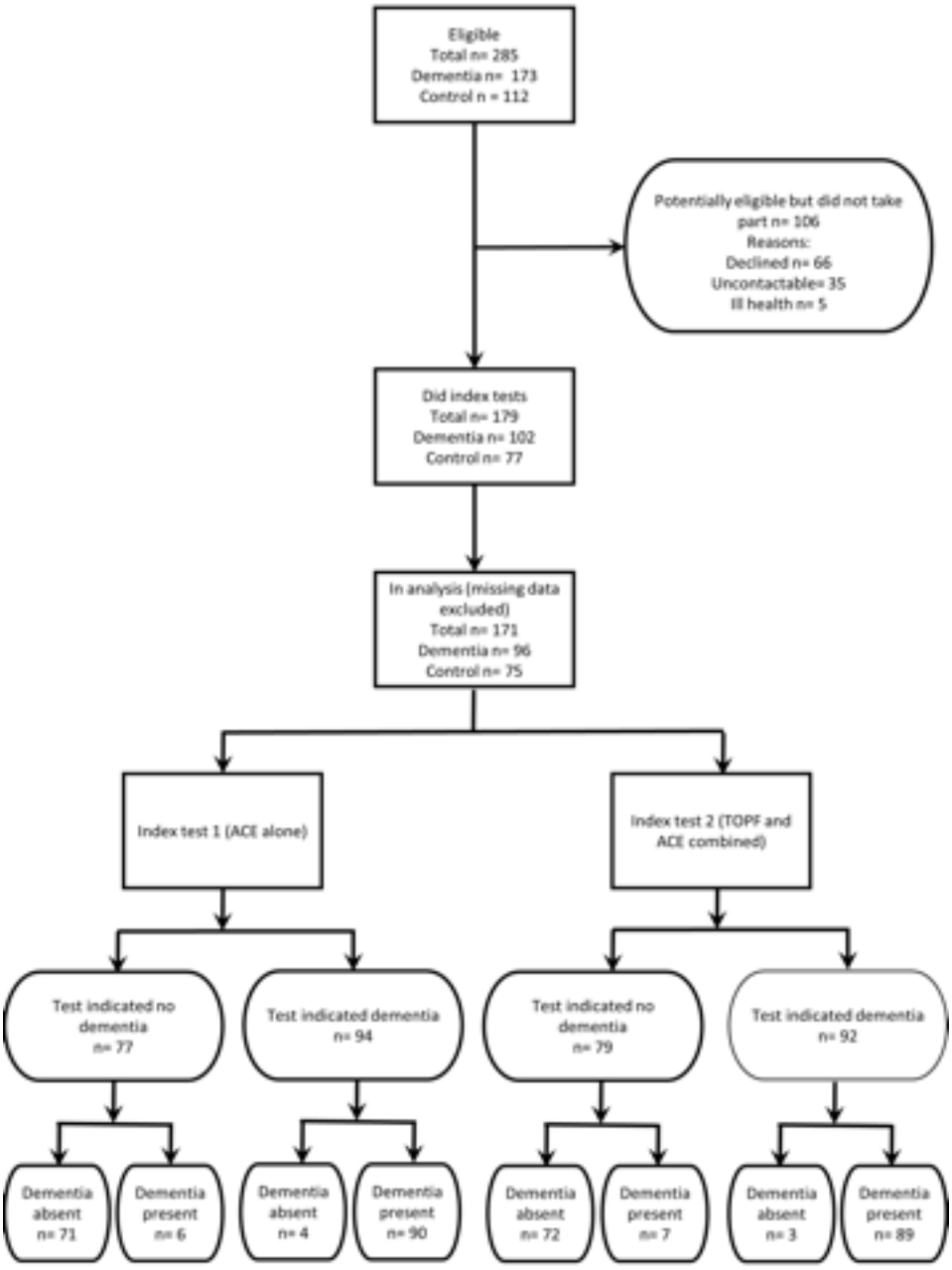

Figure 1 -
[1]Modified STARD flowchart showing flow of participants through the study

1

Table 1: Demographic variables and TOPF and ACE-III scores for dementia and control groups (Median ± IQR)

Variable	Control (n= 75)	Dementia (n= 96)	U/Fisher's Exact test
Age	72±8	81±10	1552***
Sex (M:F)	28:47	43:53	N/S
Ethnicity (White:Non-White)	1:74	8:88	N/S
Education (years)	16±6	12±5	2059.5***
Dementia type	-	60 AD:8 VaD:15 Mixed VaD and AD:13 Other	-
ACE-III	96± 6	74±16	232.5***
TOPF	119±8	104.5±22	1462***

Due to non-normal distributions, medians and Interquartile Ranges (IQRs) are reported rather than means and standard deviations.

***, Significant at P<0.001 using Mann-Whitney U test.

N/S, Non-significant using X^2 or Fisher's exact test.

AD, Alzheimer's Disease; VaD, Vascular Dementia; 'Other' dementia type included two with Dementia in Parkinson's disease, one with Frontotemporal Dementia and 10 with Unspecified Dementia.

Table 2 Logistic regression coefficients and diagnostic accuracy metrics for prediction of dementia by all diagnostic indexes

	Logistic regression coefficients			Diagnostic accuracy metrics		
Diagnostic index	Significant predictors in model (β, Wald Statistic, OR)	Model R2	Model BIC	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)
ACE-III alone (88.5 Cut-off)	ACE-III (β=-0.33, Z=-6.14***, OR= 0.72)	0.81	87.3	93.74% (86.45 to 97.92)	94.66% (88 to 98.67)	0.967 (0.94 to 0.99)
ACE-III and TOPF	ACE-III (β=-0.35 Z= -5.84***, OR=0.71)	0.81	92.11	92.7% (86.45 to 97.91)	96% (89.33 to 1)	0.97 (0.94 to 0.99)
ACE-III and demographics	ACE-III (β=-0.3 Z=-5.66***, OR=0.74)	0.82	92.82	94.79 (86.45 to 98.95)	94.67 (86.67 to 1)	0.97 (0.95 to 1)
ACE-III, TOPF, demographics	ACE-III (β=-0.3 Z= -5.1*** OR=0.74)	0.82	97.95	94.79% (85.41 to 98.96)	93.33% (86.67 to 98.67)	0.97(0.95 to 0.99)
ACE-III alone (87/88 cut-off)	-	-	-	94.79% (89.58 to 98.96)	92% (85.33 to 97.33)	-

*** P<0.001

$R^{2,}$ Nagelkerke's R^2 (an index of model effect size, with larger values indicating higher predictive power); BIC, Model Bayesian information Criterion (lower values indicate that the model accounts better for the data); CI, bootstrapped, bias corrected accelerated confidence interval; AUC, area under the ROC curve; β, β coefficient associated with a predictor in the logistic regression model; Z, value of the Wald Statistic for a predictor in the logistic regression model; OR, Odds Ratio; Demographics included gender, age and years of education.

Table 3 Cross-tabulation of dementia status of participants with results of each diagnostic index

	Diagnostic indices									
	ACE-III alone (cut-off 88/89)		ACE-III and TOPF		ACE-III and demographics		ACE-III, TOPF, demographics		ACE-III alone (cut-off 87/88)	
	+Ve	-Ve	+Ve	-Ve	+Ve	-Ve	+Ve	-Ve	+Ve	-Ve
Dementia n	90*	6 §	89*	7§	91*	5§	91*	5§	91*	5§
No dementia n	4‡	71¶	3‡	72¶	4‡	71¶	5‡	70¶	6‡	69¶

Demographics, included gender, age and years of education; +ve, dementia is present according to an index; -ve, dementia is absent according to an index.

*True positives on index compared to the reference standard.

‡False positives on index compared to the reference standard.

¶True negatives on index compared to the reference standard.

§False negatives on index compared to the reference standard.