

UC Davis

UC Davis Previously Published Works

Title

The Effectiveness and Unique Contribution of Neuropsychological Tests and the δ Latent Phenotype in the Differential Diagnosis of Dementia in the Uniform Data Set

Permalink

<https://escholarship.org/uc/item/4n65t5f9>

Journal

Neuropsychology, 30(8)

ISSN

0894-4105

Authors

John, Samantha E

Gurnani, Ashita S

Bussell, Cara

et al.

Publication Date

2016-11-01

DOI

10.1037/neu0000315

Peer reviewed



Published in final edited form as:

Neuropsychology. 2016 November ; 30(8): 946–960. doi:10.1037/neu0000315.

The Effectiveness and Unique Contribution of Neuropsychological Tests and the δ Latent Phenotype in the Differential Diagnosis of Dementia in the Uniform Data Set

Samantha E. John¹, Ashita S. Gurnani², Cara Bussell², Jessica L. Saurman², Jason W. Griffin², and Brandon E. Gavett²

¹ Department of Neurology and Alzheimer's Disease Research Center, Emory University, Atlanta, GA, USA

² Department of Psychology, University of Colorado Colorado Springs, Colorado Springs, CO, USA

Abstract

Objective—Two main approaches to the interpretation of cognitive test performance have been utilized for the characterization of disease: evaluating shared variance across tests, as with measures of severity, and evaluating the unique variance across tests, as with pattern and error analysis. Both methods provide necessary information, but the unique contributions of each are rarely considered. This study compares the two approaches on their ability to differentially diagnose with accuracy, while controlling for the influence of other relevant demographic and risk variables.

Method—Archival data requested from the NACC provided clinical diagnostic groups that were paired to one another through a genetic matching procedure. For each diagnostic pairing, two separate logistic regression models predicting clinical diagnosis were performed and compared on their predictive ability. The shared variance approach was represented through the latent phenotype δ , which served as the lone predictor in one set of models. The unique variance approach was represented through raw score values for the 12 neuropsychological test variables comprising δ , which served as the set of predictors in the second group of models.

Results—Examining the unique patterns of neuropsychological test performance across a battery of tests was the superior method of differentiating between competing diagnoses, and it accounted for 16-30% of the variance in diagnostic decision making.

Conclusion—Implications for clinical practice are discussed, including test selection and interpretation.

Keywords

differential diagnosis; neuropsychological tests; Alzheimer's disease; genetic matching algorithm; latent phenotype δ

Clinical neuropsychology has traditionally relied upon the use of cognitive tests to diagnose disorders and characterize their effect. Two main approaches to the interpretation of cognitive test performance have been utilized: 1) evaluating shared variance across tests, as through index scores (e.g., Randolph, 1998; Stern & White, 2003; Wechsler, 2008; 2009), latent variable modeling approaches (e.g., Crane et al., 2012; Gibbons et al., 2012; Mungas, Reed, & Kramer, 2003), and more recently through cognitive phenotypes of severity or burden (e.g., Royall & Palmer, 2012; Royall, Palmer, & O'Bryant, 2012), and 2) evaluating the unique variance across tests, as through interpretations of performance patterns (e.g., Reitan & Wolfson, 1993) and through use of process-oriented approaches such as the Boston Process approach (e.g., Ashendorf, Swenson, & Libon, 2013) and the Luria-Nebraska (e.g., Golden, Purisch, & Hammeke, 1991). Both methods provide us with necessary information for understanding the severity and the type of disease process implicated. The assessment and differential diagnosis of dementia benefit from both approaches, but the unique contributions of each method and the situations within which each is most useful are rarely considered. This study compares the two approaches for the purpose of differential diagnosis of dementia, while controlling for the influence of other relevant demographic and risk variables. In doing so, this study will provide support for the continued use of unique variance approaches as well as evidence for the continued improvement of measurement through novel approaches.

Dementia Etiologies and Clinical Neuropsychological Profiles

The term “dementia” describes a clinical syndrome or presentation that results in a patient's acquired inability to meet his or her occupational and social demands (McKhann et al., 2011). Several disorders fall under the umbrella term of “dementia,” and each represents a unique neurocognitive condition with its own etiology, symptom profile, and disease course. Differential diagnosis of these disorders is clarified through the use of standardized neuropsychological assessment (Simpson, 2014). The exact cause and disease mechanisms of each of the most common neurocognitive disorders remain unknown, but each of the dementia etiologies is associated with a slightly different set of contributing factors as well as its own prototypical presentation (Green, 2005).

Early diagnosis of dementia and its etiology is imperative for prognosis and for the application of disease specific treatments. The most common dementia subtypes, Alzheimer's disease (AD), Vascular dementia (VaD), Dementia with Lewy Bodies (DLB), and Frontotemporal dementia (FTD) are stage dependent and present with similar clinical manifestations in their advanced phases. To accurately identify contributing etiology, early elucidation of distinct patterns of clinical performance may be the best way to understand how impairment can fit into one of the several narrow, well-established courses of disease. Unfortunately, even a well-characterized dementia like AD can be mistaken for another etiology (Lippa et al., 2007; Karantzoulis & Galvin, 2012). The lack of a single *in vivo* biomarker that can distinguish AD from other dementia subtypes emphasizes the need for differential diagnosis based on neuropsychological test performance.

Alzheimer's disease (AD) is a progressive degenerative disease that most commonly affects the medial temporal, temporal, parietal, and frontal lobes, while sparing the primary motor

and sensory cortices and subcortical regions. Memory problems, including difficulty learning and recalling new information, are the most common clinical symptoms of the disease. Other prominent symptoms of AD include word finding difficulty, declines in semantic knowledge and language ability, as well as deficits in executive functioning, attention, and visuospatial processing (McKhann et al., 2011; Salmon & Bondi, 2009). Vascular dementia (VaD) is the second most prevalent form of dementia after AD (Jellinger, 2007) and is caused by vascular disease of the brain, which is characterized by infarcts, lacunes, hippocampal sclerosis, and white matter lesions (Jellinger, 2008). In contrast to AD's primary effects on memory, patients with VaD often present clinically with deficits in processing speed, executive functioning, attention, and language (Staekenborg et al., 2008; Smitts et al., 2015) as well as depression and gait disturbance (Staekenborg et al., 2008). In stark contrast to the primary cognitive deficits of AD and VaD, behavioral variant frontotemporal dementia (bvFTD) is marked by a progressive deterioration of personality and social functioning (Elfgren et al., 1994; Hou, Carlin, & Miller, 2004; Perry & Hodges, 2000). If memory impairment is found in bvFTD, it is typically related to a decline in the active strategies used for learning and retrieval, those related to executive functioning, as a result of frontal lobe atrophy (Perri et al., 2005). In addition to executive functioning and language deficits, bvFTD has an earlier age of onset than the other common dementia syndromes; the disorder presents with disinhibition, impaired judgment and stereotypy, higher levels of apathy, euphoria, emotional blunting, hyperorality, aberrant motor behaviors, and lower levels of empathy and personal hygiene (Barber, Snowden, & Craufurd, 1995; Elfgren et al., 1994; Liu et al., 2012; Mathias & Morphet, 2010). Finally, Dementia with Lewy Bodies (DLB) is differentiated from the other syndromes by its core criteria of fluctuating cognition, visual hallucinations, and extrapyramidal motor features. In fact, visual hallucinations occur more often in DLB (75%) than in any other neurodegenerative disease, and these continue to be a good indicator for clinical diagnosis (Karantzoulis & Galvin, 2011), in addition to measured declines and impairment in visuospatial processing and processing speed.

While each of the aforementioned neurodegenerative diseases can be characterized by a specific pattern of performance on neuropsychological tests, the diseases share neuropsychological features, may co-occur with one another, and/or present atypically. As a result, new developments and improvements in the measurement of disease type and burden are continually sought. In recent years, efforts to use existing measurement techniques in novel ways have been embraced.

δ as a Measure of Dementia Severity

A recent advancement in the field of early detection of dementia involves the development of the latent dementia phenotype, " δ ", by Royall and colleagues (Royall & Palmer, 2012; Royall, Palmer, & O'Bryant, 2012). This phenotype is an extension of Spearman's general intelligence factor " g ," which represented the shared variance across observed performance on cognitive and intelligence tests (Royall et al., 2012; Spearman, 1904). Likewise, δ is a shared variance approach to understanding cognition that is derived through multiple observed indicators of neuropsychological test performance. Because δ factor scores contain both cognitive and functional indicators of ability, the latent phenotype represents the shared

variance between cognitive and functional measures, capturing the most critical aspect of the clinical manifestation of dementia – functional decline. Given its ability to account for functional declines as well as cognitive performance, δ is thought to underlie one's ability to perform activities of daily living (ADL) and is therefore strongly associated with dementia severity (Gavett et al., 2015; Royall et al., 2007; Royall et al., 2012). Through the use of this latent construct, Royall and colleagues conceptualize the development of dementia as the simultaneous occurrence of cognitive and functional changes (Royall & Palmer, 2012; Royall et al., 2012). δ has consistently shown the ability to distinguish between those with and without clinically diagnosed dementia, and it has been validated in various samples (e.g., healthy controls, mixed clinical and healthy samples, ethnically diverse older adults, and samples with individuals suffering from multiple etiologies and differing levels of disease severity) and with a variety of assessment methods (Gavett et al., 2015; Gurnani, John, & Gavett, 2015; Royall & Palmer, 2012; Royall et al., 2012; Royall & Palmer, 2013). Thus, cross-sectional analyses provide consistent evidence that δ is an effective and efficient method for measuring dementia severity, regardless of the sample characteristics and specific neurocognitive tests used in its calculation.

Longitudinally, δ has been validated as a tool for predicting dementia severity. Utilizing the National Alzheimer's Coordinating Center Uniform Data Set (NACC UDS), Gavett and colleagues (2015) showed that individuals' latent dementia status changed in conjunction with changes in their Clinical Dementia Rating (CDR) Sum of Boxes (SOB) score over six annual visits, indicating that dementia-related neuropathological changes drive dementia-related clinical changes. δ was also strongly associated with 3-4 year changes on cognitive variables, further supporting its utility in longitudinal assessment (Royall & Palmer, 2012; Palmer & Royall, 2016). Given the accumulation of evidence supporting δ 's validity, there is growing interest in the examination of the δ construct for other clinical purposes, such as the more finely tuned estimate of disease course and severity as would be necessary for differentiating among the common dementia subtypes. Though δ has been shown to distinguish between AD, MCI, and healthy adults, its utility in distinguishing between different causes of dementia has not yet been explored. δ provides an estimate of disease severity by measuring the concomitant changes in cognitive and functional ability; however, the clinical diagnosis of a dementia also accounts for the contribution of several other risk factors and predictors of disease.

Other Predictors of Cognitive Decline

The presence of several different risk and protective factors is considered prior to the clinical diagnosis of a neurodegenerative disorder, given the known effects of these factors on disease development. An evaluation of the factors that mediate disease severity and differentially contribute to specific dementia subtypes can help with diagnosis and characterization when multiple disorders are considered plausible. These factors include age, apolipoprotein E (*APOE*) status, race/ethnicity, and years of education.

The age of onset of different dementia subtypes may vary by etiology, with bvFTD and familial AD occurring at earlier ages, as young as the 20s-50s, and sporadic AD, VaD, and DLB typically occurring more frequently in later life within the 70s and later (Rossor, Fox,

Mummery, Schott, & Warren, 2010). Increasing age is the most pronounced risk factor for the development of sporadic AD, but it may also implicate VaD and DLB. Older age is associated with changes in AD pathology, including increased levels of cerebrospinal fluid biomarkers of AD (Paternicò et al., 2012), as well as with total brain volume loss and the presence of cardiovascular disease pathology (Erten-Lyons et al., 2013; Toledo et al., 2013).

Polymorphism of the *APOE* gene can give rise to three different isoforms labeled *APOE* ϵ 2, *APOE* ϵ 3, and *APOE* ϵ 4. The presence of *APOE* ϵ 4 confers genetic risk for AD; however, there is evidence that *APOE* ϵ 4 interacts with age and each dementia subtype through the activation of different neurobiological pathways (Hauser & Ryan, 2013; Lei, Boyle, Leurgans, Schneider, & Bennett, 2012). Ethnoracial differences in risk factors and social conditions may help to explain the prevalence of specific types of dementia (Mayeda et al., 2013), particularly VaD (Howard et al., 2011). For example, ethnoracial differences in risk factors for other medical conditions (Edwards, Hall, Williams, Johnson, & O'Bryant, 2015; Johnson, et al., 2015) as well as differences in socioeconomic status and lifestyle choices may help to explain the apparent ethnoracial differences in disease development (Howard, 2013). Of those culturally determined risk factors, educational attainment, as measured through the number of years of formal schooling, is negatively correlated with dementia risk (Brayne et al., 2010; Xu et al., 2015) and appears to confer a protective advantage against the development of disease (Sharp & Gatz, 2011). As such, educational attainment provides a buffer that may delay the symptomatic presentation of changes to pathology (Jefferson et al., 2011; also see Reed et al., 2011).

The Present Study

Given the influence of risk and protective factors in the clinical diagnosis of dementia, clinical decision making typically involves the consolidation of evidence from patient demographics and medical history as well as neuropsychological test performance. Each piece of evidence in the decision is subjectively weighted according to a myriad of factors; however, clinical decision making is often flawed and difficult to quantify (Dawes, Faust, & Meehl, 1989). There is a need for improved measurement in differential diagnosis of dementia and for objective evidence to guide diagnostic outcomes. The present study will examine the differential diagnosis of the four most common dementia subtypes (AD, VaD, DLB, and FTD), while controlling for the influence of relevant demographic and risk variables, to measure the efficacy of two different approaches to disease characterization that utilize the same set of neuropsychological test variables: 1) δ , which extracts shared variance only, and 2) linear combinations, which emphasize unique variance. Logistic regression analyses will compare these two methods within pairs of diagnostic samples that have been matched to one another on demographic variables and risk and protective factors (i.e., those mentioned above, including sex, race, ethnicity, handedness, education, age, and number of *APOE* ϵ 4 alleles). Matched diagnostic samples will be attained through use of a genetic matching algorithm (described below) that will remove the confounding effects of these variables to more closely mimic an experimental design. This process will ensure that comparisons between δ and its indicators are unaffected by these covariates (the “purest” possible way to compare the two approaches). Previous research has suggested that genetic

matching algorithms are able to assist with the identification of causal relationships using observational data (Sekhon, 2011; Park et al., 2015).

The Uniform Data Set (UDS), maintained by the National Alzheimer's Coordinating Center (NACC), contains assessment data that examines both the cognitive and functional changes that occur within neurodegenerative disease processes (Morris et al., 2006; Weintraub et al., 2009). The UDS provides annual assessments that measure multiple cognitive domains and which are used for consensus diagnostic decisions (Weintraub et al., 2009). Clinical diagnoses of patients are made during a blind review of relevant patient characteristics, including medical history, demographic variables, and risk factors, the history and course of symptoms as obtained through a clinical interview with the patient and/or caregiver, and neuropsychological test performance. A multidisciplinary team of professionals – or, at some centers, a single clinician – comes to agreement about the clinical diagnosis of each patient based on all of the available evidence. While definitive diagnoses for any of the dementia syndromes cannot be obtained without autopsy confirmation, the assignment of clinical diagnoses through the consensus conference helps to reduce the possibility of misdiagnosis and is therefore used internationally by experts in aging, neurology, and neuropsychology (Ngo & Holroyd-Leduc, 2014).

The present study will utilize the NACC UDS sample to evaluate our hypothesis that a unique variance approach to differential diagnosis will be superior to a shared variance approach. We hypothesize that the pattern of test performance, rather than the overall measure of disease severity, will provide the most useful information in differentiating between two dementia subtypes and we believe that in showing this, we will provide evidence for the limitation of shared variance approaches to disease characterization. In addition, evaluation of odds ratios from the logistic regression analyses will provide useful information for both test selection and interpretation by providing some of the first quantifiable measures of clinical diagnostic judgment.

Method

We obtained archival data for this study through a request to NACC. The NACC database contains data from 34 past and present Alzheimer's Disease Centers (ADCs). Patient evaluations from initial (baseline) visits only were used in the current analyses. These visits were completed between January 2005 and February 2015. Patient cognitive and demographic variables, as well as diagnostic status, were used to identify the sample for analysis (see *Participants* and *Materials*).

Participants

The total sample size of all initial participant visits was 13,884; we excluded participants who were not tested in English ($n = 1,025$) and another 170 participants under the age of 50, leaving 12,708 baseline visits for possible analysis (19 participants met more than one of these exclusion criteria). We initially included all participants who had received a clinical diagnosis of one of the four common dementia syndromes (AD, VaD, FTD, and DLB) at their baseline visit and who possessed complete data on all demographic variables necessary for the analyses (specified below). These baseline diagnostic labels were assigned to

participants through consensus conferences or individual clinician diagnosis at each NACC visit that followed the full intake and baseline neuropsychological assessment.

Materials and Measures

Demographic variables for each participant were used in the analyses to identify matched samples within diagnostic pairings (see *Data Analysis* below). Participants were matched to one another according to age, sex, race, ethnicity, education, handedness, and number of APOE $\epsilon 4$ alleles. Performance variables from the following neuropsychological measures were used as predictors within the logistic regression models. In one set of models, the raw data values served as direct predictors of diagnosis, and in the other set of models, the raw data values were used to derive d factor scores that served as the predictor of diagnosis. All neuropsychological test variables are part of the annual assessment performed on NACC participants.

Mini-Mental State Examination—The Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) is a 30-point global screening measure used in the assessment of neurodegenerative disease. It is one of the most commonly administered global screening measures and provides a score indicative of dementia severity.

Semantic fluency—Two separate trials of semantic or category fluency are included in the NACC battery (animals and vegetables). These trials require participants to verbally generate words belonging to a specific semantic category within a timed interval. Semantic fluency is considered to be a measure of language ability.

Boston Naming Test—The Boston Naming Test (BNT 30-item version; Kaplan, Goodglass, & Weintraub, 1983; Jefferson et al., 2007) is a test of visual confrontation naming. This particular version, which uses the odd items from the 60-item BNT, measures naming ability, and serves as a test of language.

Logical Memory, Immediate and Delayed—The Logical Memory subtest is a multi-component instrument that provides both immediate and delayed measures of contextual memory (LM-I and LM-D, respectively; Wechsler, 1987). A short story is presented to participants and they are asked to freely recall the information presented to them, immediately after hearing it and following a 20-30 minute delay.

Digit Span Forward and Backward—The Digit Span test is thought to measure auditory attention and working memory (Wechsler, 1981). Participants are read a series of digits that progressively increase in span length and must repeat the digit sequence in either forward (DS-F) or reverse (DS-B) order.

Digit Symbol Coding—Digit Symbol coding (Wechsler, 1981) requires participants to fill in empty boxes with symbols below the numbers 1 through 9, based on a matching key. The number of correctly drawn matches completed in 90 s is used to measure visuomotor and graphomotor speed.

Trail Making Test Parts A and B (TMT-A and TMT-B; Reitan & Wolfson, 1993)—

The Trail Making Test is another popular and frequently administered test of executive functioning ability that involves rapid graphomotor sequencing of numbers (Part A) and then an alternating pattern of numbers and letters (Part B) visually presented to the participant. TMT-A provides a measure of visual attention and processing speed, whereas TMT-B measures those abilities plus cognitive flexibility and maintenance of mental set.

Functional Activities Questionnaire—The Functional Activities Questionnaire (FAQ; Pfeffer, Kurosaki, Harrah, Chance, & Filos, 1982) is an informant-report measure of a participant's ability to perform instrumental activities of daily living (IADLs). The FAQ is commonly used in dementia evaluations because of its reliability, validity, sensitivity, and specificity (Juva et al., 1997; Olazarán, Mouronte, & Bermejo, 2005; Teng, Becker, Woo, Cummings, & Lu, 2010).

Procedure

To compare the efficacy of shared versus unique variance approaches to differential diagnosis of dementia, participant variables regarding demographic factors, clinical diagnosis, and neuropsychological test performance were taken from the NACC dataset and utilized within competing logistic regression models. Clinical diagnostic groups were paired to one another through a genetic matching procedure (see below) to create dyad samples. For each diagnostic dyad, two separate logistic regression models predicting clinical diagnosis were performed, with each of the two models utilizing a different approach: δ factor scores were derived from the 12 neuropsychological test variables and served as the single predictor within logistic regression models that utilized a shared variance approach to differential diagnosis, and raw score values of the same 12 neuropsychological test variables served as the predictors within logistic regression models that utilized a unique variance approach to differential diagnosis. For the TMT-A, TMT-B, and FAQ tests, we recoded the variables to match the direction of all other variables (higher scores being indicative of better performance). All three scores were recoded by subtracting the observed score from the maximum score (150 for TMT-A, 300 for TMT-B, and 30 for FAQ).

Data Analysis

All analyses, with the exception of the latent variable modeling used to estimate δ factor scores, were performed in R version 3.0.2 (R Core Team, 2013). To estimate δ factor scores, Mplus version 6.11 (Muthen & Muthen, 1998-2010) was used for latent variable modeling using a robust maximum likelihood estimator; missing data were handled using full information maximum likelihood (FIML). We fixed the variance of the latent variables to 1.0 to scale the resulting factor scores as z-scores ($M = 0.0$, $SD = 1.0$). The model for δ utilizing NACC neuropsychological test variables was identified in a previous publication (Gavett et al., 2015). All previously identified factor loadings from that publication were fixed in this sample. We therefore did not estimate a new model for δ , but derived new δ factor scores for this sample of participants according to our previous results. The factor score determinacy for δ was estimated in MPlus to verify the strength of the factor scores' ability to represent the "true" latent trait of cognitive and functional disease severity.

Genetic Matching—Matching procedures identified our diagnostic dyads through both propensity scores as well as a genetic search algorithm (GenMatch; Sekhon & Mebane, 1998), according to the software description and procedures provided by Sekhon (2011). Genetic matching, so named for its use of an evolutionary search algorithm, is a statistical method that improves upon the more commonly used propensity score matching by performing an iterative search for imbalance among the covariates between groups (Sekhon, 2011). Matching techniques, broadly, are a popular and common method for achieving causal inference for observational data, when random assignment of patients to conditions cannot be performed. The genetic search algorithm utilized in genetic matching removes the limitations of propensity score matching by using the propensity score as a covariate in the match and imposing additional properties to reduce bias in the sample (Diamond & Sekhon, 2005; Sekhon & Grieve, 2011). The search algorithm *looks* for balance at each iteration and replicates the manual procedures proposed by Rosenbaum and Rubin (1984) for maximizing the balance of covariates between the two groups. A detailed description of the algorithm along with evidence of its superiority for achieving balanced matched samples can be found in the work of Diamond and Sekhon (2005; 2013).

For each matched pair of diagnoses, we began with the full sample of qualifying NACC participants (described above in *Participants*) who possessed a probable baseline visit diagnosis of the particular dementia subtypes in question (e.g., probable AD and probable DLB). As mentioned previously, participants from that full sample were excluded for having missing values on any of the covariates for which the samples were matched: sex, race, ethnicity, handedness, education, age, and number of e4 alleles. These covariates were selected for the matching procedure to control for their influence in diagnostic decision making, given that each of these typically contributes differentially to the likelihood of a particular etiology. A logistic regression model, regressing diagnostic outcome (e.g., AD vs. DLB status) onto the covariates was performed. In each case, the less commonly observed dementia was always the target (e.g., DLB). The predicted probability of DLB was thus derived from the logistic regression. This probability is the “propensity score,” which was subsequently used for genetic matching, along with the individual covariates themselves, described above. Genetic matching without replacement was used to balance out all of the covariates between the two groups. This process is “used to automatically find balance by the use of a genetic search algorithm which determines the optimal weight to give each covariate” (Sekhon, 2011, p. 8). Bootstrapping with 2000 replicates was used to facilitate a statistical test of how well balanced the groups were. Table 1 displays the characteristics of each matched dyad used in the final analyses.

Multiple Imputation—Following the creation of matched dyad samples, multiple imputation was used to impute missing data for the 12 neuropsychological test variables listed above. This procedure took into account the factor scores for δ and g' , as well as relevant demographic variables (age, race, ethnicity, education, and sex), and variables from the NACC dataset relevant to severity and presentation [Clinical Dementia Rating Sum of Boxes (CDR-SOB), Hachinski Ischemic Score, and the Uniform Parkinson's Disease Rating Scale]. A total of five imputed data sets were created using the *Hmisc* package in R, through

the use of additive regression, bootstrapping, and predictive mean matching (Harrell et al., 2015).

Logistic Regression—Two competing logistic regression models were analyzed for each diagnostic dyad, each using the same set of NACC neuropsychological test variables in one of two sets of regression predictors. Logistic regression was performed using the R package *rms*. (Harrell, 2015). The first logistic regression for each dyad utilized a shared variance approach as the only predictor of dementia diagnosis, the latent factor δ . The second logistic regression for each dyad utilized a unique variance approach through the use of raw test scores of all 12 of the neuropsychological tests as predictors of dementia diagnosis. As with the genetic matching procedure, within each dyad, the less commonly observed dementia was always the target in the regression. Regression analyses provided unstandardized parameter estimates, odds ratios, and model fit statistics [e.g., Akaike Information Criterion (AIC)], and classification accuracy data [e.g., area under the receiver operating characteristics (ROC) curve (AUC)]. An ROC curve with bootstrapped ($B = 2000$) 95% confidence intervals was plotted as a way of visually comparing the classification accuracy of the unique and shared variance approaches (Robin et al., 2011).

Results

Participant Characteristics

The demographics of the matched diagnostic dyad samples are provided in Table 1. The genetic matching procedure produced demographically similar pairs of samples across the seven demographic variables, with unique participants making up each diagnostic dyad, according to individual participant characteristics. The dyad samples with significantly different values on matched variables are indicated within the table. The overall NACC FTD sample proved to be the most difficult dementia subtype to match; the matching procedure did not produce corresponding samples of AD, DLB, and VaD participants that were similar across all variables. The FTD sample that matched to AD was statistically different on mean Global CDR score (FTD $M_{\text{CDR}} = 1.31$, AD $M_{\text{CDR}} = 1.08$). The FTD sample matched to both DLB (FTD $M_{\text{age}} = 74.04$, DLB $M_{\text{age}} = 68.34$) and VaD (FTD $M_{\text{age}} = 76.36$, VaD $M_{\text{age}} = 79.39$) was statistically different on mean age. Mean values for each of the predictors included within the logistic regression models are provided in Table 2, organized by matched dyads.

Logistic Regression Analyses and Model Fit

Two separate logistic regression models predicting clinical diagnosis of matched diagnostic dyad samples were performed on each dyad, with each of the two models utilizing a different approach (shared variance vs. unique variance) represented by the use of different predictors taken from the same NACC data (derived δ factor scores vs. the raw score values of 12 neuropsychological test variables). The logistic regression models, through evaluation of model fit statistics, indicated that the 12 neuropsychological test variables were superior predictors to δ at accurately identifying the clinical diagnosis within each matched dyad. Model fit was evaluated using four indicators, AUC, AIC, Bayesian Information Criterion (BIC), and R^2 . Table 3 displays the model fit statistics and R^2 values for both logistic

regression models for each dyad. In each logistic regression model, at least three of the four indicators of model fit demonstrated the superiority of neuropsychological predictors for the differential diagnosis of dementia. Figure 1 displays the AUC graphs for each dyad, directly comparing neuropsychological tests to δ factor scores and providing sensitivity and specificity cut-offs for each point along the curve. The factor score determinacy for δ was 0.93, 95% CI [0.904, 0.940]. The full unstandardized parameter estimates and standard errors of measurement for the logistic regression models can be seen in Table 4 for the models utilizing neuropsychological test scores and Table 5 for the models utilizing δ .

Odds Ratios

Given the superiority of neuropsychological test variables over δ factor scores in the differential diagnosis process, further evaluation of odds ratios from each logistic regression using neuropsychological test predictors by dyad can help to elucidate the benefits of specific tests for differentiating between different dementia subtypes. Table 6 presents the odds ratios and the 95% confidence intervals for the logistic regression models using the raw neuropsychological test scores as predictors. For comparison, Table 7 presents the same information using the δ factor scores as predictors.

Discussion

The current investigation sought to compare the efficacy of two different methods of dementia characterization on their ability to accurately differentiate between common dementia syndrome diagnoses (AD, VaD, DLB, FTD), when all other contributing factors were controlled. Utilizing the same set of NACC neuropsychological test variables, we compared a shared variance approach to characterization, the δ latent phenotype, to a unique variance approach to characterization, the raw scores from 12 neuropsychological tests of functioning. Results revealed the relative contribution of each test variable as well as their composite determination (i.e., δ , level of severity), which was previously unknown and understudied. Utilizing matched dyad samples within a logistic regression model of two dementia syndrome diagnoses eliminated the influence of risk factors and demographic variables, allowing for a true comparison between neuropsychological test variables and another method of dementia characterization, the latent factor δ . The results of this study suggest that δ and the neuropsychological variables of which it is composed possess different levels of accuracy in their ability to differentiate between two comparably similar dementia syndrome presentations. Although δ comprises all of the neuropsychological test variables, the patterns of performance across tests, rather than the level of severity characterized through δ , is better at classification. As previous studies have suggested, δ is most suitable as a measure of dementia severity, given its ability to account for both cognitive as well as functional changes (Gavett et al., 2015; Royall et al., 2007; Royall et al., 2012). Examining the unique patterns of neuropsychological test performance across a battery of tests is a superior method of differentiating between competing diagnoses, and it accounts for 16-30% of the variance in diagnostic decision making.

The current investigation supports the use of comprehensive neuropsychological test batteries for the accurate diagnosis of dementia syndrome subtypes because of the unique

variance available for interpretation across tests. The current study also highlights the limitations of shared variance methods of disease characterization, like the latent dementia phenotype δ or its less psychometrically advanced correlates of severity [e.g., CDR-SB, MMSE, Montreal Cognitive Assessment (MoCA), etc.]. These results also quantify the benefits of interpreting the unique variance of specific tests in certain diagnostic decisions. For example, evaluation of the odds ratios from the logistic regression models suggests that DLB is most clearly differentiated from AD by better performance on Logical Memory Delay (OR = 2.48), Logical Memory Immediate (OR = 2.03), and the MMSE (OR = 2.01). These results are not surprising, given the primary deficit in episodic memory that accompanies AD; however, our results can quantify the contribution of the tests within each comparison, providing a level of confidence for test pattern interpretation not previously established. In fact, across comparisons, AD is differentiated from other diagnoses by a worse MMSE score, which highlights the sensitivity of that measure to a particular dementia subtype and suggests the use of other screening tools [e.g., MoCA, St. Louis University Mental Status Examination (SLUMS)] when alternate subtypes are suspected. In other words, our results can be used to guide both test selection as well as test interpretation. Across all comparisons, VaD is characterized by worse performance on timed measures, as it is differentiated from AD by worse performance on Trails A, differentiated from DLB by worse performance on Animal fluency, and differentiated from FTD by worse performance on Digit Symbol Coding. In this example, our results provide a clinical correlation to suspected pathology, given the often slowed processing speed that is observed in VaD as compared to other subtypes. Certain tests, across all comparisons, failed to differentiate between competing diagnoses. Digit Span-Forward, for example, did not provide a significant benefit to any of the differential diagnoses conducted in our analyses. An absence of significant findings for a particular test suggests that a specific battery of cognitive tests might better aid certain diagnostic decisions, and that those considerations should be made in advance. Common neuropsychological practice would suggest the administration of both Digit Span-Forward and Backward (e.g., Rabin, Barr, & Burton, 2005; Rabin, Paolillo, & Barr, 2016); however, a consideration of our results in the design of a cognitive battery can maximize the information attained from cognitive evaluations while also decreasing their burden.

Implications, Strengths, and Limitations of the Study

There is accumulating evidence that the latent variable δ provides an accurate, and at times superior, characterization of disease severity when compared to a battery of neuropsychological tests (Koppara et al, 2016); however, this study demonstrates a limitation to the use of δ , differential diagnosis within the most common subtypes of dementia: AD, VaD, FTD, and DLB. It is important to note that δ typically possesses better than chance levels of accuracy in differentiating between two dementia subtypes, but the variables with which it is composed (i.e., those taken from neuropsychological tests) provide better accuracy for differential diagnosis and do so at substantially higher rates. This result is not surprising. As a latent factor score, δ accounts for the shared variance among test scores, providing an estimate of what is common across measures (typically, the presence of impairment). By its very nature, it cannot represent the unique patterns of performance that are often used clinically to characterize a patient's cognitive profile and subsequently

implicate a particular diagnosis. Because the different dementia subtypes typically present with patterns of impairment on specific tests or within specific domains of functioning, the preservation of test score patterns is necessary for identifying and characterizing the subtype. In other words, our head to head comparison roughly modeled two different approaches to diagnosis – use of a novel dementia characterization factor versus use of pattern analysis/unique contribution, a method akin to clinical decision making. While it may not be particularly surprising that δ is limited in its ability to differentiate among subtypes of dementia, this study has provided the first evidence of δ 's limitations, reinforcing its use in the measurement of severity and characterization of disease status (i.e., healthy vs. MCI vs. dementia), but documenting the boundaries of its extension into other areas of study.

Another unique aspect of this study is its use of statistically matched samples of clinical populations. The use of genetic algorithms and propensity score matching represents a novel approach to the investigation of neuropsychological predictors of disease, allowing us to isolate the predictive power of each of the neuropsychological tests, for multiple diagnoses, without the influence of demographic or other risk variables that are known to affect neuropsychological test performance (e.g., Gurnani, John, & Gavett, 2015). Use of the genetic matching algorithm allowed us to approximate the conditions of random assignment by using existing data. We were able to control for a large number of variables, mimicking the conditions necessary for the experimental investigation of δ 's and the neuropsychological tests' ability to differentiate between types of dementia. To our knowledge, genetic algorithms and propensity score matching have not previously been used in a study of neuropsychological predictors, and it has therefore not previously been possible to estimate the percentage of variance independently accounted for by neuropsychological test variables in diagnostic decision making. After accounting for a host of other variables that are important for differential diagnosis (e.g., age, APOE genotype), we found that a combination of 12 neuropsychological test scores was able to explain 16-30% of the variability in diagnostic outcomes. Presumably this number could be improved with a more comprehensive battery, as is typical of most outpatient neuropsychological assessments. McKhann et al. (2011) recommend “neuropsychological testing should be performed when the routine history and bedside mental status examination cannot provide a confident diagnosis.” Our study supports the conclusion that each variable of evidence obtained through testing and interviewing accounts for only a circumscribed amount of the overall variance in clinician decision making. While it is important to continue investigating novel approaches to diagnosis, the role of the clinical neuropsychologist and the comprehensive neuropsychological evaluation should not be overlooked. Traditional methods of clinical characterization continue to show efficacy and accuracy in determining diagnosis.

Given this study's attempt to quantify contributions to diagnostic decision making, it would have been ideal to use a sample of participants with pathologically verified diagnoses. Unfortunately, the sample size of pathologically confirmed NACC participants of each diagnostic subtype is limited and was therefore insufficient for our planned analyses. Instead, we relied upon clinical diagnostic labels, and this is a potential limitation to our study. In the absence of autopsy-confirmation of diagnoses, it is possible that the samples utilized in the present study contain patients who were clinically misdiagnosed. Very little

available research has been done to explore the accuracy of clinical diagnoses within the NACC sample specifically; however, where available, the literature suggests that clinical diagnoses of AD are more accurate than the other etiological categories and that all clinical diagnoses are at least moderately associated with final pathological diagnoses (Beach, Monsell, Phillips, & Kukull, 2012; Nelson et al., 2010). Of note, when other dementia syndromes and neurodegenerative disorders are clinically misdiagnosed as cases of AD, 39% of those misdiagnosed cases possess or exceed the minimum AD histopathological threshold used for diagnosis, suggesting simultaneous pathology (Beach et al., 2012). Outside of the NACC, there is evidence that the criteria for bvFTD have a sensitivity of 95% and specificity of 82% when differentiating it from AD (Harris et al., 2013). Taken together, these limited studies suggest that clinical diagnoses may at times be inaccurate and that steps should be taken to ensure longitudinal follow-up of diagnosed cases as well as the use of consensus criteria and judgment. These steps are taken for the NACC sample. Just as our study is limited by the absence of an autopsy confirmed sample, so to is the research literature on accuracy of clinical diagnoses. Future studies should aim to create a bank of autopsy and pathologically confirmed patients, and when possible, research on the different dementia syndromes should strive to use pathologically confirmed diagnoses.

Clinical diagnoses are assigned to each participant through a consensus team decision in which the neuropsychological data are used as evidence. Given that our predictors of diagnosis were already considered within the assignment of diagnosis in the first place, our analyses are at risk for criterion contamination. However, consensus diagnosis teams were not using the same data that we used in our analyses. Consensus teams did not have access to the logistic regression weights used here, nor did they have access to patients' δ scores, which can mitigate the contamination to some degree. And because δ is derived exclusively from the individual test scores, both methods are affected equally, which does not diminish the relative advantage of individual test scores over δ for differential diagnosis. Estimated latent factor scores are not "pure" measures of shared variance due to factor indeterminacy; however, our estimation of δ had a high determinacy value (0.93) and therefore represents a valid estimate of the true latent trait. To further support the accuracy and validity of estimated δ factor scores, future research should replicate our analysis procedures in independent, community-based samples.

It is also important to distinguish the decision making of the consensus team from the "decision making" conducted through our analyses. Our logistic regressions were set up to explicitly distinguish between two matched samples of different dementia subtypes (e.g., AD vs. DLB) without any consideration of demographic or genetic data. This process allowed us to quantify specific contributions of the decision making process, but it does not mimic the process through which consensus diagnosis is determined. The nature of our hypotheses that directed our approach ultimately provided us with several new findings specific to the benefits of neuropsychological tests for differential diagnosis (in the absence of the other typically considered factors). Another potential limitation within our study was the small number of VaD cases available, which kept our sample sizes small for each of the VaD dyads and their associated comparisons. As a result of the small sample size, these analyses were not well powered to differentiate the accuracy of δ versus the

neuropsychological tests or to provide evidence for the specific cognitive variables best suited for differentiating groups.

Given that our initial population consisted of mostly AD patients, our dyad samples may have included matched participants from differing diagnostic categories that more closely approximated one another than would be expected in a clinical setting. In other words, our matching procedure may have created dyad samples of participants that represented slightly atypical versions of their diagnostic categories (given the required similarity to the other diagnostic group to which they were matched). This may have created a somewhat artificial, but increasingly difficult diagnostic scenario in which prototypical presentations were not available. As a result of the matching procedure used, our dyad samples represented a unique group of participants for whom accurate differential diagnosis would be particularly difficult to achieve, and therefore, particularly important to investigate. Our dyads maximized the diagnostic ambiguity possible when subtypes present atypically and thus, our results indicate the strength of particular tests at differentiating between diagnostic groups. From this perspective, our study provided strong evidence for the use of neuropsychological tests for differential diagnosis between clinically similar presentations of different dementia subtypes. This may offer insight into clinical practice and real world conditions. Presentations of dementia can vary widely and many different etiologies can result in a similar presentation, especially during advanced stages of disease. Individual patients may present atypically or with limited information about the disease course and history. For these cases, our study offers evidence of the diagnostic power of particular neuropsychological tests for determining diagnosis. In situations in which differential diagnosis is not obvious (i.e., two different dementias look similar), a tool or set of likelihoods for detecting this difference is beneficial, strengthening the applicability of our research to the real world.

Overall, the data suggest that the critical component of clinical neuropsychological practice, the interpretation of test results within a battery of findings, is crucial for accurate differential diagnosis. Understanding the unique differences among dementia subtypes requires a method of interpretation that goes beyond shared variance approaches to characterization, highlighting brain-behavior relationships and the use of clinically correlated findings. While δ has been shown to effectively identify the presence of dementia and its stage, this study quantifies what is often a messy and difficult to determine clinical diagnostic judgment. Traditional assessment methods are important and necessary for differentiating between subtypes of dementia, and more importantly, should be scrutinized according to their unique contribution and utilized according to objective, quantifiable evidence. Through the use of genetic matching and the creation of similar diagnostic dyads, this study provides evidence of the unique contribution of neuropsychological tests in the diagnostic decision making process and provides further support for the use of specific tests according to suspected differentials.

Acknowledgments

This work was supported through the NACC database, which is funded by NIA/NIH Grant U01 AG016976. NACC data are contributed by the NIA funded ADCs: P30 AG019610 (PI Eric Reiman, MD), P30 AG013846 (PI Neil Kowall, MD), P50 AG008702 (PI Scott Small, MD), P50 AG025688 (PI Allan Levey, MD, PhD), P30 AG010133 (PI Andrew Saykin, PsyD), P50 AG005146 (PI Marilyn Albert, PhD), P50 AG005134 (PI Bradley Hyman, MD,

PhD), P50 AG016574 (PI Ronald Petersen, MD, PhD), P50 AG005138 (PI Mary Sano, PhD), P30 AG008051 (PI Steven Ferris, PhD), P30 AG013854 (PI M. Marsel Mesulam, MD), P30 AG008017 (PI Jeffrey Kaye, MD), P30 AG010161 (PI David Bennett, MD), P30 AG010129 (PI Charles DeCarli, MD), P50 AG016573 (PI Frank LaFerla, PhD), P50 AG016570 (PI David Teplow, PhD), P50 AG005131 (PI Douglas Galasko, MD), P50 AG023501 (PI Bruce Miller, MD), P30 AG035982 (PI Russell Swerdlow, MD), P30 AG028383 (PI Linda Van Eldik, PhD), P30 AG010124 (PI John Trojanowski, MD, PhD), P50 AG005133 (PI Oscar Lopez, MD), P50 AG005142 (PI Helena Chui, MD), P30 AG012300 (PI Roger Rosenberg, MD), P50 AG005136 (PI Thomas Montine, MD, PhD), P50 AG033514 (PI Sanjay Asthana, MD, FRCP), and P50 AG005681 (PI John Morris, MD).

References

- Ashendorf, L.; Swenson, R.; Libon, DJ. The Boston Process approach to neuropsychological assessment. Oxford University Press; New York, NY: 2013.
- Barber R, Snowden JS, Craufurd D. Frontotemporal dementia and Alzheimer's disease: Retrospective differentiation using information from informants. *Journal of Neurology, Neurosurgery, and Psychiatry*. 1995; 59:61–70. <http://dx.doi.org/10.1136/jnnp.59.1.61>.
- Beach TG, Monsell SE, Phillips LE, Kukull W. Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005–2010. *Journal of Neuropathology and Experimental Neurology*. 2012; 71:266–273. <http://dx.doi.org/10.1097/NEN.0b013e31824b211b>. [PubMed: 22437338]
- Brayne C, Ince PG, Keage HA, McKeith IG, Matthews FE, Polvikoski T, Sulkava R. Education, the brain and dementia: Neuroprotection or compensation? *Brain: A Journal of Neurology*. 2010; 133:2210–2216. <http://dx.doi.org/10.1093/brain/awq185>. [PubMed: 20826429]
- Crane PK, Carle A, Gibbons LE, Insel P, Mackin RS, Gross A, Mungas D. Development and assessment of a composite score for memory in the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Brain Imaging and Behavior*. 2012; 6:502–516. <http://dx.doi.org/10.1007/s11682-012-9186-z>. [PubMed: 22782295]
- Dawes RM, Faust D, Meehl PE. Clinical versus actuarial judgment. *Science*. Mar 31.1989 243:1668–1674. <http://dx.doi.org/10.1126/science.2648573>. [PubMed: 2648573]
- Edwards M, Hall J, Williams B, Johnson L, O'Bryant S. Molecular markers of amnesic mild cognitive impairment among Mexican Americans. *Journal of Alzheimer's Disease*. 2015; 49:221–228. <http://dx.doi.org/10.3233/JAD-150553>.
- Elfgren C, Brun A, Gustafson L, Johanson A, Minthon L, Passant U, Risberg J. Neuropsychological tests as discriminators between dementia of Alzheimer's type and frontotemporal dementia. *International Journal of Geriatric Psychiatry*. 1994; 9:635–642. <http://dx.doi.org/10.1002/gps.930090807>.
- Erten-Lyons D, Dodge HH, Woltjer R, Silbert LC, Howieson DB, Kramer P, Kaye JA. Neuropathologic basis of age-associated brain atrophy. *Journal of the American Medical Association Neurology*. 2013; 70:616–622. <http://dx.doi.org/10.1001/jamaneurol.2013.1957>. [PubMed: 23552688]
- Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research*. 1975; 12:189–198. [PubMed: 1202204]
- Gavett BE, Vudy V, Jeffrey M, John SE, Gurnani AS, Adams JW. The _ latent dementia phenotype in the uniform data set: Cross-validation and extension. *Neuropsychology*. 2015; 29:344–352. <http://dx.doi.org/10.1037/neu0000128>. [PubMed: 25151112]
- Gibbons LE, Carle AC, Mackin RS, Harvey D, Mukherjee S, Insel P, Crane PK. A composite score for executive functioning, validated in Alzheimer's Disease Neuroimaging Initiative (ADNI) participants with baseline mild cognitive impairment. *Brain Imaging and Behavior*. 2012; 6:517–527. <http://dx.doi.org/10.1007/s11682-012-9176-1>. [PubMed: 22644789]
- Golden, CJ.; Purisch, AD.; Hammeke, TA. Luria–Nebraska Neuropsychological Battery: Forms I and II. Western Psychological Services; Los Angeles, CA: 1991.
- Green, RC. Diagnosis and management of Alzheimer's disease and other dementias. Professional Communications, Inc.; 2005.

- Gurnani AS, John SE, Gavett BE. Regression based norms for a bi-factor model for scoring The Brief Test of Adult Cognition by Telephone (BTACT). *Archives of Clinical Neuropsychology*. 2015; 30:280–291. <http://dx.doi.org/10.1093/arclin/acv005>. [PubMed: 25724515]
- Harrell, FE. RMS: Regression modeling strategies: R package (Version 4.3–1) [Computer software]. 2015. Retrieved from http://CRAN.Rproject.org/package_rms
- Harrell, FE.; Dupont, C. R package (Version 3.17–0) [Computersoftware]. 2015. Retrieved from http://CRAN.R-project.org/package_Hmisc
- Harris JM, Gall C, Thompson JC, Richardson AM, Neary D, du Plessis D, Jones M. Sensitivity and specificity of FTDC criteria for behavioral variant frontotemporal dementia. *Neurology*. 2013; 80:1881–1887. <http://dx.doi.org/10.1212/WNL.0b013e318292a342>. [PubMed: 23596080]
- Hauser PS, Ryan RO. Impact of apolipoprotein E on Alzheimer's disease. *Current Alzheimer Research*. 2013; 10:809–817. <http://dx.doi.org/10.2174/15672050113109990156>. [PubMed: 23919769]
- Hou CE, Carlin D, Miller BL. Non-Alzheimer's disease dementias: Anatomic, clinical, and molecular correlates. *Canadian Journal of Psychiatry/La Revue Canadienne de Psychiatrie*. 2004; 49:164–171.
- Howard VJ. Reasons underlying racial differences in stroke incidence and mortality. *Stroke*. 2013; 44:S126–S128. <http://dx.doi.org/10.1161/STROKEAHA.111.000691>. [PubMed: 23709708]
- Howard VJ, Kleindorfer DO, Judd SE, McClure LA, Safford MM, Rhodes JD, Howard G. Disparities in stroke incidence contributing to disparities in stroke mortality. *Annals of Neurology*. 2011; 69:619–627. <http://dx.doi.org/10.1002/ana.22385>. [PubMed: 21416498]
- Jefferson AL, Gibbons LE, Rentz DM, Carvalho JO, Manly J, Bennett DA, Jones RN. A life course model of cognitive activities, socioeconomic status, education, reading ability, and cognition. *Journal of the American Geriatrics Society*. 2011; 59:1403–1411. <http://dx.doi.org/10.1111/j.1532-5415.2011.03499.x>. [PubMed: 21797830]
- Jefferson AL, Wong S, Gracer TS, Ozonoff A, Green RC, Stern RA. Geriatric performance on an abbreviated version of the Boston naming test. *Applied Neuropsychology*. 2007; 14:215–223. <http://dx.doi.org/10.1080/09084280701509166>. [PubMed: 17848132]
- Jellinger KA. The enigma of vascular cognitive disorder and vascular dementia. *Acta Neuropathologica*. 2007; 113:349–388. <http://dx.doi.org/10.1007/s00401-006-0185-2>. [PubMed: 17285295]
- Jellinger KA. The pathology of “vascular dementia”: A critical update. *Journal of Alzheimer's Disease*. 2008; 14:107–123.
- Johnson LA, Gamboa A, Vintimilla R, Cheatwood AJ, Grant A, Trivedi A, O'Bryant SE. Comorbid depression and diabetes as a risk for mild cognitive impairment and Alzheimer's Disease in elderly Mexican Americans. *Journal of Alzheimer's Disease*. 2015; 47:1–8.
- Juva K, Mäkelä M, Erkinjuntti T, Sulkava R, Yukoski R, Valvanne J, Tilvis R. Functional assessment scales in detecting dementia. *Age and ageing*. 1997; 26:393–400. [PubMed: 9351484]
- Kaplan, EF.; Goodglass, H.; Weintraub, S. *The Boston naming test*. 2nd Ed.. Lea & Febiger; Philadelphia, Pennsylvania: 1983.
- Karantzoulis S, Galvin JE. Distinguishing Alzheimer's disease from other major forms of dementia. *Expert Review of Neurotherapeutics*. 2011; 11:1579–1591. <http://dx.doi.org/10.1586/ern.11.155>. [PubMed: 22014137]
- Koppara A, Wolfsgruber S, Kleineidam L, Schmidtke K, Frölich L, Kurz A, Wagner M. The latent dementia phenotype _ is associated with cerebrospinal fluid biomarkers of Alzheimer's disease and predicts conversion to dementia in subjects with mild cognitive impairment. *Journal of Alzheimer's Disease*. 2015; 49:547–560. <http://dx.doi.org/10.3233/JAD-150257>.
- Lei Y, Boyle P, Leurgans S, Schneider J, Bennett D. Disentangling the effects of age and APOE on neuropathology and late life cognitive decline. *Neurobiology of Aging*. 2012; 29:997–1003.
- Lippa CF, Duda JE, Grossman M, Hurtig HI, Aarsland D, Boeve BF, Wszolek ZK. DLB and PDD boundary issues: Diagnosis, treatment, molecular pathology, and biomarkers. *Neurology*. 2007; 68:812–819. <http://dx.doi.org/10.1212/01.wnl.0000256715.13907.d3>. [PubMed: 17353469]
- Liu W, Miller BL, Kramer JH, Rankin K, Wyss-Coray C, Gearhart R, Rosen HJ. Behavioral disorders in the frontal and temporal variants of frontotemporal dementia. *Neurology*. 2004; 62:742–748. [PubMed: 15007124]

- Mathias JL, Morphet K. Neurobehavioral differences between Alzheimer's disease and frontotemporal dementia: A metaanalysis. *Journal of Clinical and Experimental Neuropsychology*. 2010; 32:682–698. <http://dx.doi.org/10.1080/13803390903427414>. [PubMed: 20063255]
- Mayeda ER, Karter AJ, Huang ES, Moffet HH, Haan MN, Whitmer RA. Racial/ethnic differences in dementia risk among older Type 2 diabetic patients: The diabetes and aging study. *Diabetes Care*. 2014; 37:1009–1015. <http://dx.doi.org/10.2337/dc13-0215>. [PubMed: 24271192]
- McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Knopman DS, Jack. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging—Alzheimer's Associations workgroups on diagnostic guidelines for Alzheimer's disease. *Journal of Alzheimer's Disease*. 2011; 7:263–269. <http://dx.doi.org/10.1016/j.jalz.2011.03.005>.
- Morris JC, Weintraub S, Chui HC, Cummings J, DeCarli C, Ferris S, Beekly D. The Uniform Data Set (UDS): Clinical and cognitive variables and descriptive data from Alzheimer Disease Centers. *Alzheimer Disease & Associated Disorders*. 2006; 20:210–216. [PubMed: 17132964]
- Mungas D, Reed BR, Kramer JH. Psychometrically matched measures of global cognition, memory, and executive function for assessment of cognitive decline in older persons. *Neuropsychology*. 2003; 17:380–392. <http://dx.doi.org/10.1037/0894-4105.17.3.380>. [PubMed: 12959504]
- Muthén, L.; Muthén, B. *Mplus User's Guide*. Muthén & Muthén; Los Angeles, CA.: 1998.
- Nelson PT, Jicha GA, Kryscio RJ, Abner EL, Schmitt FA, Cooper G, Markesbery WR. Low sensitivity in clinical diagnoses of dementia with Lewy bodies. *Journal of Neurology*. 2010; 257:359–366. <http://dx.doi.org/10.1007/s00415-009-5324-y>. [PubMed: 19795154]
- Ngo J, Holroyd-Leduc JM. Systematic review of recent dementia practice guidelines. *Age and Ageing*. 2015; 44:25–33. <http://dx.doi.org/10.1093/ageing/afu143>. [PubMed: 25341676]
- Olazarán J, Mouronte P, Bermejo F. Clinical validity of two scales of instrumental activities in Alzheimer's disease. *Neurología (Barcelona, Spain)*. 2005; 20:395–401.
- Paternicò D, Galluzzi S, Drago V, Bocchio-Chiavetto L, Zanardini R, Pedrini L, Frisoni GB. Cerebrospinal fluid markers for Alzheimer's disease in a cognitively healthy cohort of young and old adults. *Journal of Alzheimer's Disease*. 2012; 8:520–527. <http://dx.doi.org/10.1016/j.jalz.2011.10.003>.
- Perri R, Koch G, Carlesimo GA, Serra L, Fadda L, Pasqualetti P, Caltagirone C. Alzheimer's disease and frontal variant of frontotemporal dementia—a very brief battery for cognitive and behavioural distinction. *Journal of Neurology*. 2005; 252:1238–1244. <http://dx.doi.org/10.1007/s00415-005-0849-1>. [PubMed: 15906058]
- Perry RJ, Hodges JR. Differentiating frontal and temporal variant frontotemporal dementia from Alzheimer's disease. *Neurology*. 2000; 54:2277–2284. <http://dx.doi.org/10.1212/WNL.54.12.2277>. [PubMed: 10881252]
- Pfeffer RI, Kurosaki TT, Harrah CH, Chance JM, Filos S. Measurement of functional activities in older adults in the community. *Journal of gerontology*. 1982; 37:323–329. [PubMed: 7069156]
- Rabin LA, Barr WB, Burton LA. Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Archives of Clinical Neuropsychology*. 2005; 20:33–65. <http://dx.doi.org/10.1016/j.acn.2004.02.005>. [PubMed: 15620813]
- Rabin LA, Paolillo E, Barr WB. Stability in test-usage practices of clinical neuropsychologists in the United States and Canada over a 10-year period: A follow-up survey of INS and NAN members. *Archives of Clinical Neuropsychology*. 2016; 31:206–230. <http://dx.doi.org/10.1093/arclin/acw007>. [PubMed: 26984127]
- Randolph, C. *RBANS manual*. The Psychological Corporation; San Antonio, TX: 1998.
- R Core Team. *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing; Vienna, Austria: 2013. Retrieved from <https://www.R-project.org/>
- Reed BR, Dowling M, Tomaszewski Farias S, Sonnen J, Strauss M, Schneider JA, Mungas D. Cognitive activities during adulthood are more important than education in building reserve. *Journal of the International Neuropsychological Society*. 2011; 17:615–624. <http://dx.doi.org/10.1017/S1355617711000014>. [PubMed: 23131600]

- Reitan, RM.; Wolfson, D. The Halstead-Reitan neuropsychological test battery: Theory and clinical interpretation. 2nd ed.. Neuropsychology Press; Tuscon, AZ: 1993.
- Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M. pROC: An open-source package for R and S_ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011; 12:77. <http://dx.doi.org/10.1186/1471-2105-12-77>. [PubMed: 21414208]
- Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*. 1984; 79:516–524. <http://dx.doi.org/10.1080/01621459.1984.10478078>.
- Rossor MN, Fox NC, Mummery CJ, Schott JM, Warren JD. The diagnosis of young-onset dementia. *Lancet Neurology*. 2010; 9:793–806. [http://dx.doi.org/10.1016/S1474-4422\(10\)70159-9](http://dx.doi.org/10.1016/S1474-4422(10)70159-9). [PubMed: 20650401]
- Royall DR, Lauterbach EC, Kaufer D, Malloy P, Coburn KL, Black KJ. The cognitive correlates of functional status: A review from the Committee on Research of the American Neuropsychiatric Association. *The Journal of Neuropsychiatry and Clinical Neurosciences*. 2007; 19:249–265. [PubMed: 17827410]
- Royall DR, Palmer RF. Validation of a latent construct for dementia case-finding in Mexican-Americans. *J Alzheimers Dis*. 2013; 37:89–97. [PubMed: 23800829]
- Royall DR, Palmer RF. Getting past “g”: Testing a new model of dementing processes in persons without dementia. *The Journal of Neuropsychiatry and Clinical Neurosciences*. 2012; 24:37–46. <http://dx.doi.org/10.1176/appi.neuropsych.11040078>. [PubMed: 22450612]
- Royall, DR.; Palmer, RF.; Markides, KS. Exportation and Validation of Latent Constructs for Dementia Case Finding in a Mexican American Population– based Cohort.. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*. 2016. <http://dx.doi.org/10.1093/geronb/gbw004>
- Royall DR, Palmer RF, O'Bryant SE. Validation of a latent variable representing the dementing process. *Journal of Alzheimer's Disease*. 2012; 30:639–649.
- Royall DR, Palmer RF, Vidoni ED, Honea RA, Burns JM. The default mode network and related right hemisphere structures may be the key substrates of dementia. *Journal of Alzheimer's Disease*. 2012; 32:467–478.
- Salmon DP, Bondi MW. Neuropsychological assessment of dementia. *Annual Review of Psychology*. 2009; 60:257–282. <http://dx.doi.org/0.1146/annurev.psych.57.102904.190024>.
- Sekhon JS. Multivariate and propensity score matching software with automated balance optimization: The matching package for R. *Journal of Statistical Software*. 2011; 42:1–52. <http://dx.doi.org/10.18637/jss.v042.i07>.
- Sekhon, JS.; Diamond, A. Genetic matching for estimating causal effects (Unpublished Manuscript).. Paper Presented at the Annual Meeting of the Political Methodology; Tallahassee, FL.. 2005.
- Sekhon JS, Grieve RD. A matching method for improving covariate balance in cost-effectiveness analyses. *Health economics*. 2012; 21:695–714. [PubMed: 21633989]
- Sekhon JS, Mebane WR. Genetic optimization using derivatives. *Political Analysis*. 1998; 7:187–210. <http://dx.doi.org/10.1093/pan/7.1.187>.
- Sharp ES, Gatz M. Relationship between education and dementia: An updated systematic review. *Alzheimer Disease and Associated Disorders*. 2011; 25:289–304. <http://dx.doi.org/10.1097/WAD.0b013e318211c83c>. [PubMed: 21750453]
- Simpson JR. DSM-5 and neurocognitive disorders. *Journal of the American Academy of Psychiatry and the Law Online*. 2014; 42:159–164.
- Smits LL, van Harten AC, Pijnenburg YA, Koedam EL, Bouwman FH, Sistermans N, van der Flier WM. Trajectories of cognitive decline in different types of dementia. *Psychological Medicine*. 2015; 45:1051–1059. <http://dx.doi.org/10.1017/S0033291714002153>. [PubMed: 25229325]
- Spearman C. “General intelligence,” objectively determined and measured. *American Journal of Psychology*. 1904; 15:201–292.
- Staekenborg SS, van der Flier WM, van Straaten EC, Lane R, Barkhof F, Scheltens P. Neurological signs in relation to type of cerebrovascular disease in vascular dementia. *Stroke*. 2008; 39:317–322. <http://dx.doi.org/10.1161/STROKEAHA.107.493353>. [PubMed: 18096841]

- Stern, RA.; White, T. Neuropsychological Assessment Battery. Psychological Assessment Resources; Lutz, FL: 2003.
- Teng E, Becker BW, Woo E, Cummings JL, Lu PH. Subtle deficits in instrumental activities of daily living in subtypes of mild cognitive impairment. *Dementia and geriatric cognitive disorders*. 2010; 30:189–197. [PubMed: 20798539]
- Toledo JB, Arnold SE, Raible K, Brettschneider J, Xie SX, Grossman M, Trojanowski JQ. Contribution of cerebrovascular disease in autopsy confirmed neurodegenerative disease cases in the National Alzheimer's Coordinating Centre. *Brain: A Journal of Neurology*. 2013; 136:2697–2706. <http://dx.doi.org/10.1093/brain/awt188>. [PubMed: 23842566]
- Wechsler, D. Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV). Pearson Assessment; San Antonio, TX: 2008.
- Wechsler, D. Wechsler Memory Scale—Fourth Edition (WMS-IV) Technical and interpretive manual. Pearson Assessment; San Antonio, TX: 2009.
- Wechsler, D. WAIS-R manual: Wechsler Adult Intelligence Scale-Revised. Psychological Corporation; New York, NY.: 1981.
- Wechsler, D.; Stone, CP. Wechsler Memory Scale-Revised. Psychological Corporation; New York, NY.: 1987.
- Weintraub S, Salmon D, Mercaldo N, Ferris S, Graff-Radford NR, Chui H, Peskind E. The Alzheimer's disease centers' uniform data set (UDS): The neuropsychological test battery. *Alzheimer disease and associated disorders*. 2009; 23:91. [PubMed: 19474567]
- Xu W, Tan L, Wang H-F, Tan M-S, Tan L, Li J-Q, Yu J-T. Education and risk of dementia: Dose-response meta-analysis of prospective cohort studies. *Molecular Neurobiology*. 2016; 53:113–123.

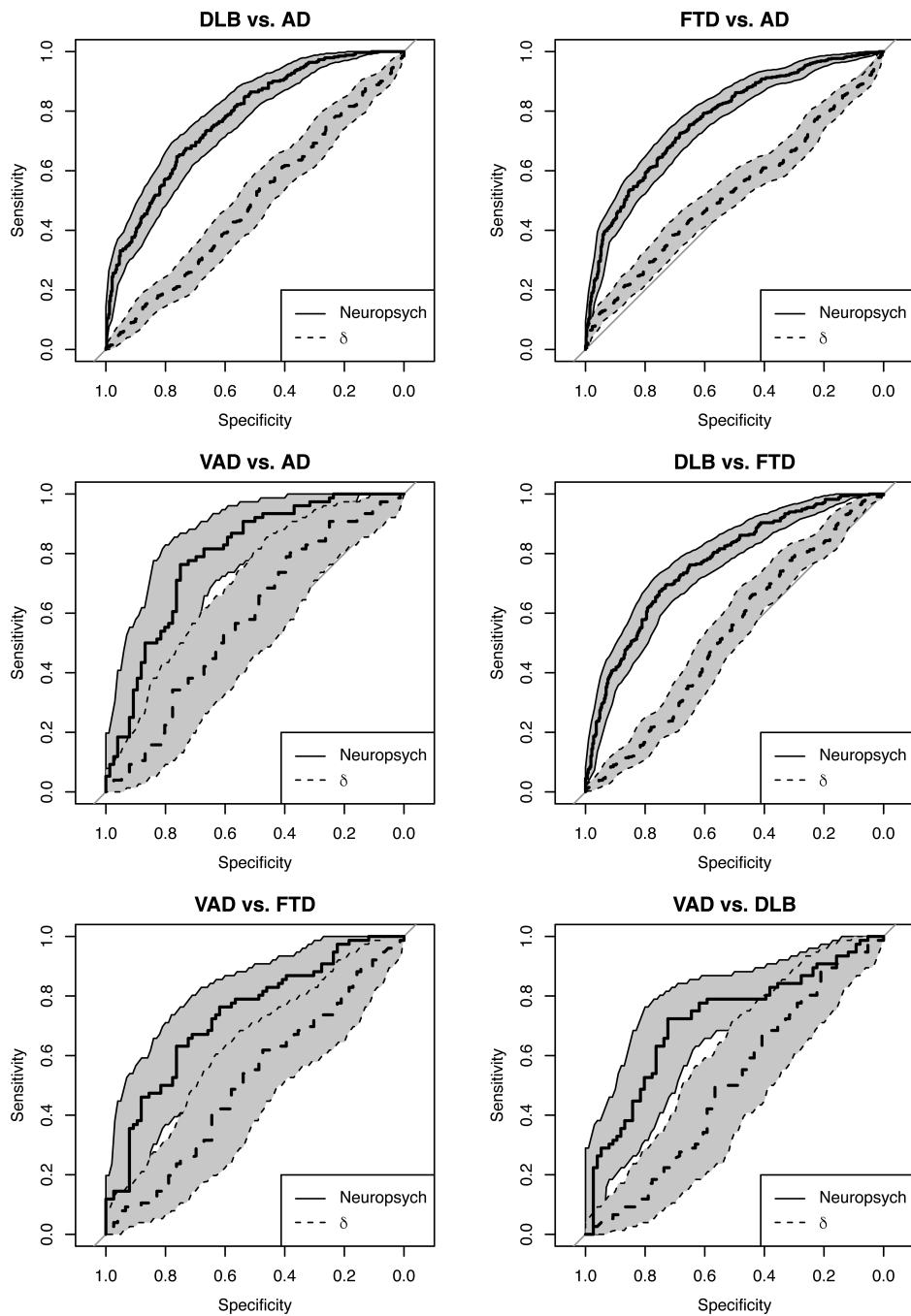


Figure 1.

Graphs show area under the receiver operating characteristics (ROC) curve (AUC) for each dementia disorder dyad, comparing the neuropsychological tests to factor scores.

Classification accuracy of the two methods can be evaluated by examining the sensitivity and specificity cut-offs for each point along the curve. For every dyad, neuropsychological tests were the more accurate predictor of diagnosis.

Table 1

Participant demographics in the matched samples.

	AD vs. DLB		AD vs. FTD		AD vs. VAD		FTD vs. DLB		DLB vs. VAD		FTD vs. VAD	
	AD	DLB	AD	FTD	AD	VAD	FTD	DLB	DLB	VAD	FTD	VAD
<i>n</i>	332	332	479	479	76	76	332	332	76	76	76	76
Age, years; <i>M</i> (<i>SD</i>)	74.05 (8.00)	74.04 (8.11)	65.18 (8.00)	64.97 (8.17)	79.39 (9.01)	79.39 (9.08)	74.04 ^a (8.11)	68.34 ^a (7.11)	78.34 (7.65)	79.39 (9.08)	76.36 ^b (6.50)	79.39 ^b (9.08)
Female sex, <i>n</i> (%)	239 (72.0%)	239 (72.0%)	305 (63.7%)	305 (63.7%)	38 (50.0%)	38 (50.0%)	239 (72.0%)	240 (72.3%)	43 (56.6%)	38 (50.0%)	44 (57.9%)	38 (50.0%)
Education, years; <i>M</i> (<i>SD</i>)	14.83 (3.28)	14.84 (3.61)	15.41 (2.79)	15.20 (3.16)	14.11 (3.19)	14.11 (3.16)	14.84 (3.61)	15.18 (3.18)	14.50 (3.06)	14.11 (3.16)	14.74 (3.03)	14.11 (3.16)
Caucasian Race, <i>n</i> (%)	307 (92.5%)	307 (92.5%)	462 (96.5%)	462 (96.5%)	55 (72.4%)	55 (72.4%)	307 (92.5%)	315 (94.9%)	58 (76.3%)	55 (72.4%)	64 (84.2%)	55 (72.4%)
Hispanic ethnicity, <i>n</i> (%)	7 (2.1%)	7 (2.1%)	5 (1.0%)	5 (1.0%)	3 (3.9%)	3 (3.9%)	7 (2.1%)	5 (1.5%)	3 (3.9%)	3 (3.9%)	3 (3.9%)	3 (3.9%)
Right handedness, <i>n</i> (%)	299 (90.1%)	299 (90.1%)	423 (88.3%)	424 (88.5%)	71 (93.4%)	71 (93.4%)	299 (90.1%)	297 (89.5%)	71 (93.4%)	71 (93.4%)	70 (92.1%)	71 (93.4%)
APOE $\epsilon 4$ alleles, <i>M</i> (<i>SD</i>)	0.59 (0.67)	0.59 (0.66)	0.36 (0.56)	0.36 (0.56)	0.37 (0.54)	0.37 (0.54)	0.59 (0.66)	0.45 (0.60)	0.38 (0.54)	0.37 (0.54)	0.38 (0.56)	0.37 (0.54)
Global CDR; <i>M</i> (<i>SD</i>)	1.14 (0.71)	1.28 (0.81)	1.08 ^c (0.68)	1.31 ^c (0.82)	1.25 (0.71)	1.28 (0.79)	1.28 (0.81)	1.36 (0.83)	1.39 (0.89)	1.28 (0.79)	1.53 (0.83)	1.28 (0.79)

Note. AD = Alzheimer's disease; DLB = dementia with Lewy Bodies; FTD = Frontotemporal dementia; VAD = Vascular dementia; APOE = apolipoprotein E.

^aThe mean age differs between DLB and FTD groups ($p < .05$).

^bThe mean age differs between FTD and VAD groups ($p < .05$).

^cThe frequencies of Global CDR scores differ between AD and FTD groups ($p < .05$).

Table 2

Means and standard deviations for neuropsychological tests and d scores by diagnostic group.

Test	AD vs. DLB			AD vs. FTD			AD vs. VAD			FTD vs. DLB			DLB vs. VAD			FTD vs. VAD		
	AD	DLB	FTD	AD	FTD	VAD	AD	VAD	DLB	FTD	DLB	DLB	VAD	DLB	VAD	FTD	FTD	VAD
MMSE	20.6 (6.31)	20.82 (6.93)	20.24 (6.44)	21.03 (7.91)	17.84 (7.03)	20.44 (6.62)	21.2 (7.83)	20.82 (6.93)	19.85 (7.66)	20.44 (6.62)	20.87 (6.61)	20.44 (6.62)	20.44 (6.62)	20.87 (6.61)	20.44 (6.62)	20.87 (6.61)	20.44 (6.62)	20.44 (6.62)
LM-I	3.84 (3.39)	5.77 (4.1)	4.16 (3.53)	6.34 (4.87)	3.86 (4.21)	5.27 (3.74)	6.17 (4.74)	5.77 (4.1)	5.55 (4.86)	5.27 (3.74)	4.58 (3.94)	5.27 (3.74)	5.27 (3.74)	4.58 (3.94)	5.27 (3.74)	4.58 (3.94)	5.27 (3.74)	5.27 (3.74)
LM-D	1.93 (2.83)	4.18 (3.92)	2.33 (3.15)	4.84 (4.74)	1.77 (2.75)	2.95 (3.91)	4.47 (4.36)	4.18 (3.92)	3.77 (4.11)	2.95 (3.91)	2.74 (3.44)	2.95 (3.91)	2.95 (3.91)	2.74 (3.44)	2.95 (3.91)	2.74 (3.44)	2.95 (3.91)	2.95 (3.91)
DS-F	6.75 (2.33)	6.68 (2.19)	6.63 (2.41)	6.62 (2.63)	6.87 (2.2)	6.68 (1.95)	6.68 (2.56)	6.68 (2.19)	6.63 (2.17)	6.68 (1.95)	6.72 (2.68)	6.68 (1.95)	6.68 (1.95)	6.72 (2.68)	6.68 (1.95)	6.72 (2.68)	6.68 (1.95)	6.68 (1.95)
DS-B	4.38 (2.11)	3.83 (1.83)	4.07 (2.02)	4.31 (2.5)	4.53 (2.06)	4.05 (1.87)	4.45 (2.39)	3.83 (1.83)	3.42 (1.81)	4.05 (1.87)	4.37 (2.48)	4.05 (1.87)	4.05 (1.87)	4.37 (2.48)	4.05 (1.87)	4.37 (2.48)	4.05 (1.87)	4.05 (1.87)
Animals	10.92 (5.73)	10.63 (5.44)	10.88 (5.43)	10.25 (6.17)	9.33 (5.13)	9.23 (4.97)	10.22 (5.99)	10.63 (5.44)	10.75 (6.32)	9.23 (4.97)	9.29 (4.52)	9.23 (4.97)	9.23 (4.97)	9.29 (4.52)	9.23 (4.97)	9.29 (4.52)	9.23 (4.97)	9.23 (4.97)
Vegetables	6.83 (4.09)	6.66 (3.57)	6.55 (4.01)	6.22 (4.19)	5.83 (3.7)	6.52 (3.64)	6.02 (3.9)	6.66 (3.57)	7.05 (4.28)	6.52 (3.64)	5.54 (3.28)	6.52 (3.64)	6.52 (3.64)	5.54 (3.28)	6.52 (3.64)	5.54 (3.28)	6.52 (3.64)	6.52 (3.64)
TMT-A ^a	77.49 (44.39)	54.07 (42.8)	77.1 (45.14)	84.84 (41.3)	74.24 (42.58)	54.2 (37.11)	84.56 (41.33)	54.07 (42.8)	53.31 (43.1)	54.2 (37.11)	68.22 (44.3)	54.2 (37.11)	54.2 (37.11)	68.22 (44.3)	54.2 (37.11)	68.22 (44.3)	54.2 (37.11)	54.2 (37.11)
TMT-B ^a	98.3 (87.11)	44.02 (71.36)	101.97 (91.5)	125.93 (93.17)	79.1 (79.77)	46.4 (69.3)	120.44 (91.03)	44.02 (71.36)	37.88 (64.78)	46.4 (69.3)	102.14 (88.21)	46.4 (69.3)	46.4 (69.3)	102.14 (88.21)	46.4 (69.3)	102.14 (88.21)	46.4 (69.3)	46.4 (69.3)
DSC	25.98 (14.59)	19.09 (11.21)	24.45 (14.98)	30.05 (14.84)	22.92 (13.69)	20.22 (12.58)	29.4 (14.22)	19.09 (11.21)	18.75 (10.33)	20.22 (12.58)	24.61 (13.47)	20.22 (12.58)	20.22 (12.58)	24.61 (13.47)	20.22 (12.58)	24.61 (13.47)	20.22 (12.58)	20.22 (12.58)
BNT	20.31 (7.44)	22.6 (5.91)	22.11 (6.95)	20.79 (8.52)	16.4 (8.2)	19.17 (6.84)	20.98 (8.43)	22.6 (5.91)	20.71 (7.56)	19.17 (6.84)	19.72 (7.7)	19.17 (6.84)	19.17 (6.84)	19.72 (7.7)	19.17 (6.84)	19.72 (7.7)	19.17 (6.84)	19.17 (6.84)
FAQ ^a	11.44 (8.45)	8.4 (8.11)	12.43 (8.28)	9.59 (8.34)	7.66 (7.53)	8.63 (8.59)	8.98 (8.29)	8.4 (8.11)	8.54 (8.45)	8.63 (8.59)	7.62 (8.13)	8.63 (8.59)	8.63 (8.59)	7.62 (8.13)	8.63 (8.59)	7.62 (8.13)	8.63 (8.59)	8.63 (8.59)
<i>d</i>	-1.84 (1.33)	-1.92 (1.45)	-1.89 (1.39)	-1.85 (1.64)	-2.47 (1.56)	-2.04 (1.33)	-1.84 (1.6)	-1.92 (1.45)	-2.14 (1.61)	-2.04 (1.33)	-1.99 (1.27)	-2.04 (1.33)	-2.04 (1.33)	-1.99 (1.27)	-2.04 (1.33)	-1.99 (1.27)	-2.04 (1.33)	-2.04 (1.33)

Note. AD = Alzheimer's disease; DLB = dementia with Lewy Bodies; FTD = Frontotemporal dementia; VAD = Vascular dementia; MMSE = Mini-Mental State Examination; LM-I = Logical Memory – Immediate; LM-D = Logical Memory – Delayed; DS-F = Digit Span – Forward; DS-B = Digit Span – Backward; TMT-A = Trail Making Test part A; TMT-B = Trail Making Test part B; DSC = Digit Symbol Coding; BNT = Boston Naming Test; FAQ = Functional Activities Questionnaire.

^aTMT and FAQ scores are recoded so that high scores are reflective of better performance.

Table 3

Model fit statistics for differential diagnosis of dementia using neuropsychological tests and δ factor scores.

Comparison	AUC		AIC		BIC		R^2	
	NP	δ	NP	δ	NP	δ	NP	δ
AD vs. DLB	.774	.506	779.42	924.08	837.90	933.07	.303	.001
AD vs. FTD	.770	.526	1124.84	1331.91	1188.08	1341.64	.284	.000
AD vs. VAD	.782	.579	195.53	211.42	234.84	217.46	.302	.029
FTD vs. DLB	.766	.526	787.68	924.14	846.16	933.14	.280	.001
DLB vs. VAD	.710	.497	213.51	214.56	252.82	220.61	.160	.001
FTD vs. VAD	.719	.510	211.27	214.65	250.58	220.70	.199	.001

Note. AUC = Area under the curve; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; NP = neuropsychological tests; AD = Alzheimer's disease; DLB = dementia with Lewy Bodies; FTD = Frontotemporal dementia; VAD = Vascular dementia.

Table 4

Unstandardized parameter estimates and standard errors for the six logistic regression models using raw neuropsychological test scores as predictors

Comparison	Intercept	MMSE	LM-I	LM-D	DS-F	DS-B	Animals	Vegetables	TMT-A	TMT-B	DSC	BNT	FAQ
DLB vs. AD													
<i>b</i>	0.056	0.087*	0.118*	0.181*	-0.068	-0.154*	-0.007	-0.038	-0.008*	-0.004*	-0.014	0.014	-0.060*
SE	0.333	0.028	0.046	0.050	0.056	0.067	0.031	0.040	0.003	0.002	0.014	0.018	0.016
FTD vs. AD													
<i>b</i>	0.234	0.063*	0.088*	0.194*	-0.066	0.029	-0.086*	-0.092*	-0.003	0.001	0.038*	-0.023	-0.088*
SE	0.265	0.022	0.038	0.038	0.041	0.052	0.023	0.032	0.003	0.002	0.010	0.013	0.013
VAD vs. AD													
<i>b</i>	-1.389*	0.112*	0.148	-0.060	-0.001	-0.147	-0.048	0.012	-0.025*	-0.005	0.035	0.059	-0.023
SE	0.685	0.053	0.105	0.104	0.116	0.137	0.068	0.081	0.009	0.004	0.029	0.034	0.034
DLB vs. FTD													
<i>b</i>	-0.059	0.005	0.006	-0.043	0.065	-0.050	0.067*	0.113*	-0.011*	-0.006*	-0.045*	0.003	0.043*
SE	0.357	0.028	0.048	0.048	0.053	0.072	0.029	0.042	0.004	0.002	0.014	0.018	0.013
VAD vs. DLB													
<i>b</i>	-0.650	-0.004	0.072	-0.125	-0.035	0.256	-0.116*	0.029	-0.007	0.001	0.050	0.031	-0.015
SE	0.684	0.050	0.085	0.085	0.098	0.132	0.059	0.070	0.007	0.004	0.026	0.030	0.027
VAD vs. FTD													
<i>b</i>	0.208	-0.048	0.189	-0.116	0.048	-0.016	-0.026	0.111	-0.012	-0.009*	0.012	-0.010	0.066*
SE	0.674	0.050	0.111	0.102	0.108	0.129	0.059	0.070	0.008	0.004	0.028	0.034	0.030

Note. MMSE = Mini-Mental State Examination; LM-I = Logical Memory – Immediate; LM-D = Logical Memory – Delayed; DS-F = Digit Span – Forward; DS-B = Digit Span – Backward; TMT-A = Trail Making Test part A; TMT-B = Trail Making Test part B; DSC = Digit Symbol Coding; BNT = Boston Naming Test; FAQ = Functional Activities Questionnaire; AD = Alzheimer's disease; DLB = dementia with Lewy Bodies; FTD = Frontotemporal dementia; VAD = Vascular dementia. Positive regression coefficients mean that the first diagnosis shown in the column labeled "Comparison" is associated with better scores on a test, whereas negative regression coefficients mean that the first diagnosis shown in that column is associated with worse scores on a test.

* $p < .05$

Table 5

Unstandardized parameter estimates and standard errors for the six logistic regression models using δ scores as predictors

Comparison	Intercept	δ
DLB vs. AD		
<i>b</i>	-0.069	-0.036
SE	0.131	0.056
FTD vs. AD		
<i>b</i>	0.032	0.017
SE	0.102	0.043
VAD vs. AD		
<i>b</i>	0.464	0.206
SE	0.306	0.116
DLB vs. FTD		
<i>b</i>	-0.057	-0.030
SE	0.123	0.051
VAD vs. DLB		
<i>b</i>	0.091	0.043
SE	0.283	0.111
VAD vs. FTD		
<i>b</i>	-0.066	-0.033
SE	0.301	0.126

Note. AD = Alzheimer's disease; DLB = dementia with Lewy Bodies; FTD = Frontotemporal dementia; VAD = Vascular dementia. Positive regression coefficients mean that the first diagnosis shown in the column labeled "Comparison" is associated with higher δ scores (less severe dementia), whereas negative regression coefficients mean that the first diagnosis shown in that column is associated with lower δ scores (more severe dementia).

* $p < .05$

Table 6

Odds ratios and 95% confidence intervals for the six logistic regression models using raw neuropsychological test scores as predictors

Comparison	MMSE	LM-I	LM-D	DS-F	DS-B	Animals	Vegetables	TMT-A	TMT-B	DSC	BNT	FAQ
DLB vs. AD												
OR	2.01	2.028	2.477	0.873	0.63	0.942	0.797	0.657	0.475	0.748	1.163	0.319
95% CI	[1.29, 3.12]	[1.18, 3.48]	[1.52, 4.04]	[0.7, 1.09]	[0.43, 0.94]	[0.58, 1.54]	[0.5, 1.28]	[0.47, 0.92]	[0.24, 0.92]	[0.44, 1.27]	[0.79, 1.71]	[0.18, 0.58]
FTD vs. AD												
OR	1.661	1.699	2.632	0.876	1.091	0.502	0.574	0.866	1.19	2.137	0.773	0.188
95% CI	[1.17, 2.35]	[1.09, 2.65]	[1.82, 3.81]	[0.75, 1.03]	[0.8, 1.48]	[0.35, 0.72]	[0.39, 0.84]	[0.64, 1.17]	[0.67, 2.1]	[1.47, 3.12]	[0.58, 1.03]	[0.12, 0.3]
VAD vs. AD												
OR	2.444	2.426	0.741	0.998	0.644	0.683	1.077	0.269	0.375	2	1.913	0.645
95% CI	[1.06, 5.62]	[0.71, 8.34]	[0.27, 2.06]	[0.63, 1.57]	[0.29, 1.44]	[0.23, 1.99]	[0.41, 2.79]	[0.1, 0.71]	[0.09, 1.54]	[0.64, 6.24]	[0.91, 4.01]	[0.18, 2.28]
DLB vs. FTD												
OR	1.041	1.039	0.805	1.139	0.86	1.705	1.964	0.554	0.331	0.41	1.032	2.253
95% CI	[0.67, 1.62]	[0.59, 1.84]	[0.51, 1.28]	[0.92, 1.4]	[0.56, 1.31]	[1.08, 2.7]	[1.19, 3.23]	[0.38, 0.8]	[0.17, 0.63]	[0.24, 0.7]	[0.71, 1.51]	[1.38, 3.68]
VAD vs. DLB												
OR	0.971	1.544	0.535	0.932	2.157	0.395	1.189	0.687	1.225	2.701	1.413	0.757
95% CI	[0.44, 2.13]	[0.57, 4.18]	[0.23, 1.23]	[0.63, 1.37]	[0.99, 4.7]	[0.16, 1]	[0.52, 2.71]	[0.35, 1.36]	[0.26, 5.71]	[0.99, 7.36]	[0.73, 2.72]	[0.27, 2.11]
VAD vs. FTD												
OR	0.683	3.115	0.56	1.101	0.953	0.811	1.945	0.536	0.188	1.262	0.896	3.484
95% CI	[0.31, 1.5]	[0.85, 11.44]	[0.21, 1.52]	[0.72, 1.68]	[0.45, 2.04]	[0.32, 2.06]	[0.85, 4.43]	[0.24, 1.21]	[0.05, 0.78]	[0.42, 3.77]	[0.43, 1.87]	[1.13, 10.7]

Note. MMSE = Mini-Mental State Examination; LM-I = Logical Memory – Immediate; LM-D = Logical Memory – Delayed; DS-F = Digit Span – Forward; DS-B = Digit Span – Backward; TMT-A = Trail Making Test part A; TMT-B = Trail Making Test part B; DSC = Digit Symbol Coding; BNT = Boston Naming Test; FAQ = Functional Activities Questionnaire; AD = Alzheimer's disease; DLB = dementia with Lewy Bodies; FTD = Frontotemporal dementia; VAD = Vascular dementia. Odds ratios greater than 1 mean that better scores on a test are associated with greater odds of the first diagnosis shown in the column labeled "Comparison," whereas odds ratios less than 1 mean that better scores on a test are associated with lower odds of the first diagnosis shown in that column.

Table 7Odds ratios and 95% confidence intervals for the six logistic regression models using δ scores as predictors

Comparison	δ
DLB vs. AD	
OR	0.933
95% CI	[0.76, 1.15]
FTD vs. AD	
OR	1.033
95% CI	[0.88, 1.21]
VAD vs. AD	
OR	1.481
95% CI	[0.96, 2.28]
DLB vs. FTD	
OR	0.944
95% CI	[0.78, 1.14]
VAD vs. DLB	
OR	1.086
95% CI	[0.72, 1.64]
VAD vs. FTD	
OR	0.94
95% CI	[0.59, 1.5]

Note. AD = Alzheimer's disease; DLB = dementia with Lewy Bodies; FTD = Frontotemporal dementia; VAD = Vascular dementia. Odds ratios greater than 1 mean that higher δ scores (less severe dementia) are associated with greater odds of the first diagnosis shown in the column labeled "Comparison," whereas odds ratios less than 1 mean that higher δ scores (less severe dementia) are associated with lower odds of the first diagnosis shown in that column.