Title
Uniform data set language measures for bvFTD and PPA diagnosis and monitoring.

Permalink
https://escholarship.org/uc/item/31v6d3qh

Journal
Alzheimer's & dementia (Amsterdam, Netherlands), 13(1)

ISSN
2352-8729

Authors
Staffaroni, Adam M
Weintraub, Sandra
Rascovsky, Katya
et al.

Publication Date
2021

DOI
10.1002/dad2.12148

Peer reviewed
Uniform data set language measures for bvFTD and PPA diagnosis and monitoring

Adam M. Staffaroni1 | Sandra Weintraub2 | Katya Rascovsky3 | Katherine P. Rankin1 | Jack Taylor1 | Julie A. Fields4 | Kaitlin B. Casaletto1 | Argye E. Hillis5 | Sladjana Lukic1 | Maria Luisa Gorno-Tempini1 | Hilary Heuer1 | Merilee A. Teylan6 | Walter A. Kukull6 | Bruce L. Miller1 | Bradley F. Boeve7 | Howard J. Rosen1 | Adam L. Boxer1 | Joel H. Kramer1

1 Department of Neurology, Memory and Aging Center, Weill Institute for Neurosciences, University of California at San Francisco (UCSF), San Francisco, California, USA
2 Mesulam Center for Cognitive Neurology and Alzheimer’s Disease and Department of Psychiatry and Behavioral Sciences, Feinberg School of Medicine, Department of Neurology, Northwestern University, Chicago, Illinois, USA
3 Penn Frontotemporal Degeneration Center, Department of Neurology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA
4 Department of Psychiatry and Psychology, Mayo Clinic, Division of Neurocognitive Disorders, Rochester, Minnesota, USA
5 Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
6 National Alzheimer’s Coordinating Center, Department of Epidemiology, University of Washington, Seattle, Washington, USA
7 Department of Neurology, College of Medicine, Mayo Clinic, Rochester, Minnesota, USA

Correspondence
Adam M. Staffaroni, Assistant Professor, University of California, San Francisco, Weill Institute for Neurosciences, Department of Neurology, Memory and Aging Center, 675 Nelson Rising Lane, Suite 190, San Francisco, CA, USA.
Email: adam.staffaroni@ucsf.edu

Funding information
National National Institute on Aging/National Institutes of Health, Grant/Award Numbers: AG061253, AG058752, AG032306, AG017586, AG054519, U01 AG016976; Advancing Research & Treatment for Frontotemporal Lobar Degeneration, Grant/Award Number: U54 NS092089; Longitudinal Evaluation of Familial Frontotemporal Dementia Subjects, Grant/Award Number: U01 AG045390; ARTFL LEFFTDS Longitudinal Frontotemporal Lobar Degeneration ALLFTD, Grant/Award Number: U19 AG063911; Larry L. Hillblom Foundation, Grant/Award Numbers: 2018-A-0-FEL, 2017-A-004-FEL; National Science Foundation, Grant/Award Number: 1815068.

Abstract

Introduction: The Frontotemporal Lobar Degeneration Module (FTLD-MOD) includes a neuropsychological battery designed to assess the clinical features of FTLD, although much is unknown about its utility. We investigated FTLD-MOD and Uniform Data Set 3.0 (UDS) language tests for differential diagnosis and disease monitoring.

Methods: Linear regressions compared baseline performances in 1655 National Alzheimer’s Coordinating Center participants (behavioral variant frontotemporal dementia (bvFTD, n = 612), semantic variant primary progressive aphasia (svPPA, n = 168), non-fluent/agrammatic variant PPA (nfvPPA, n = 168), logopenic variant PPA (lvPPA, n = 109), and controls (n = 581)). Sample sizes to detect treatment effects were estimated using longitudinal data.

Results: Among PPAs, the FTLD-MOD language tasks and UDS Multilingual Naming Test accurately discriminated svPPA. Number Span Forward best discriminated lvPPA; Phonemic:Semantic Fluency ratio was excellent for nfvPPA classification. UDS fluency and naming measures required the smallest sample size to detect meaningful change.

Discussion: The FTLD-MOD and UDS differentiated among PPA subtypes. UDS 3.0 measures performed best for longitudinal monitoring.
1 | BACKGROUND

Frontotemporal dementia (FTD) is a collection of clinical syndromes that present with cognitive and motor impairments. Diverse FTD syndromes often arise from a non-Alzheimer’s pathologic substrate, known as frontotemporal-lobar degeneration (FTLD), with most cases caused by FTLD-tau or FTLD with inclusions of transactive response DNA-binding protein (FTLD-TDP). However, the clinical syndromes often do not map consistently to specific pathologic entities. Behavioral variant FTD (bvFTD) is typified by early loss of social decorum and behavioral changes, with relative preservation of language, although language impairments can be present. Of the FTD clinical syndromes, pathological prediction is the most challenging in bvFTD. The FTD spectrum includes two language syndromes, classified as variants of primary progressive aphasia (PPA). Patients with the semantic variant (svPPA) exhibit a loss of semantic knowledge with relative preservation of speech production, grammar, and repetition. Most svPPA cases are associated with FTLD-TDP. The primary clinical features of the non-fluent/agrammatic variant (nfvPPA) are effortful spontaneous speech, agrammatism, and errors in speech production with spared semantic knowledge; an nfvPPA clinical syndrome is most predictive of FTLD-tau. A third PPA syndrome, logopenic variant PPA (lvPPA), presents with impaired repetition and single-word retrieval, with relative preservation of semantic knowledge and grammar. Unlike the other PPA variants, lvPPA has been associated with Alzheimer’s disease (AD) pathology in 80% to 90% of cases, although other studies have reported FTLD pathology in 44% to 60%. Thus correct syndromic classification can serve as a guide to predicting underlying pathology, albeit with probabilistic accuracy.

Differentiating bvFTD and the PPA subtypes can present a substantial diagnostic challenge. Although as common as AD in early onset neurodegenerative disease, FTD is still relatively rare and provides fewer opportunities for clinicians to sharpen their diagnostic acumen. Nevertheless, accurate diagnosis is important for clinical care and treatment planning. The value of differential diagnosis will continue to grow as therapeutic treatments become available. Unlike AD, there are currently no molecular biomarkers to differentiate FTLD-tau and FTLD-TDP; clinical syndromes, although imperfect, remain one of the best in vivo predictors of pathology.

The differential diagnosis of FTD and PPA requires not only a detailed clinical history, but also a comprehensive evaluation of speech, language, cognition, and behavior. Many FTD and PPA patients in the United States have been studied through the Alzheimer’s Disease Centers (ADC) program of the National Institute on Aging (NIA). These participants complete a battery of neuropsychological measures as part of the Uniform Data Set (UDS). A study of pathologically verified FTLD and AD patients, however, concluded that an earlier version of UDS did not adequately differentiate between FTLD and AD. In response, a battery of tests called the FTLD Module (FTLD-MOD) was designed to capture the core features of bvFTD and PPA. This battery comprises several measures of language, social cognition, and behavior. An initial investigation of biomarker-supported bvFTD and AD suggested that the FTLD-MOD improves discrimination of these groups. In a recent study, clinically diagnosed PPA participants performed worse than bvFTD on all FTLD-MOD language tasks. The utility of these tasks for discriminating among the various PPA subtypes, however, remains unknown. In addition, the newest version of the UDS (v3.0; henceforth, UDS) includes several language tests relevant to FTLD.

In addition to validating tests for differential diagnosis, there is an urgent need for measures that are sensitive to changes over time in PPA and bvFTD. These measures could serve as endpoints for treatment trials and help clinicians monitor cognition. Sensitive endpoints are especially important in rare conditions such as FTLD. Although several functional and cognitive endpoints have been identified in PPA and bvFTD, there remains room for improvement. FTLD-MOD and UDS tests may be candidates for longitudinal tracking. The present study investigated the utility of the FTLD-MOD and UDS language tests for differential diagnosis in PPA and longitudinal monitoring in bvFTD and PPA.

2 | METHODS

2.1 | Participants

The sample included participants from the National Alzheimer’s Coordinating Center (NACC) database who completed at least one
FTLD-MOD language task, between February 2012 and February 2020, at NIA-funded ADCs and affiliated centers. Beginning in 2014, many PPA and bvFTD participants were recruited into the Advancement of Research and Treatment in Frontotemporal Lobar Degeneration (ARTFL; U54 NS092089)22 +// Longitudinal Evaluation of Familial Frontotemporal Dementia Subjects (LEFFTDS: U01 AG045390)23 programs and co-enrolled into the respective site ADCs. Participants were excluded if their first language was not English. All patients were tested in English. All participants (or their proxies) provided written informed consent.

Controls were defined as those having normal cognition according to UDS protocol and a global Clinical Dementia Rating Scale plus NACC FTLD module (CDR®+NACC FTLD)24,25 score of 0 at all visits. PPA or bvFTD syndromes were diagnosed according to research criteria2,4 without considering FTLD-MOD performance. Patients with co-diagnosed bvFTD and PPA were diagnosed according to their PPA syndrome (svPPA = 79; nfvPPA = 21; lvPPA = 8). These criteria resulted in 1655 participants, including 612 participants with a primary clinical diagnosis of bvFTD, 185 with svPPA, 168 with nfvPPA, 109 with lvPPA, and 581 cognitively normal controls. The same pattern of results reported in this manuscript was observed after restricting the sample to cases without co-diagnoses.

2.2 | Measures

2.2.1 | CDR®+NACC-FTLD

The CDR®+NACC-FTLD is scored similarly to the traditional CDR®, a total score is calculated to categorize each patient as having mild features of neurodegenerative disease (CDR®+NACC-FTLD = 0.5) or clear features of an overt neurodegenerative syndrome (CDR®+NACC-FTLD = 1, 2, or 3). The CDR®+NACC-FTLD includes two additional domains germane to FTLD: Behavior and Language.25,26 The eight domain scores were summed to calculate the Sum of Boxes (range:0-24; higher scores indicate greater impairment).18,24

2.2.2 | FTLD-MOD language measures

The FTLD-MOD primarily comprises measures of language and social cognition/behavior. The social cognition measures are detailed in several published16,27–29 and ongoing studies. For FTLD-MOD documentation, visit: https://www.alz.washington.edu.30

Regular and irregular word reading

Participants first read a list of 15 regularly spelled words (ie, adhering to standard phonic rules) aloud. Participants then read 15 irregularly spelled words (eg, gnome, yacht). Irregular word reading requires semantic knowledge or familiarity. Making “regularization” errors when reading these words, termed surface dyslexia, is featured in the svPPA diagnostic criteria.4 Regular word reading was hypothesized to be more impaired in lvPPA and nfvPPA secondary to impaired phonological processing (lvPPA) or speech production (nfvPPA).

Sentence repetition and reading

Participants first repeat five sentences that are presented orally. Participants are later provided with these five sentences in a written format and asked to read them aloud. Repetition impairment is a core feature of lvPPA and relatively preserved in svPPA.4 Repetition may be impaired in nfvPPA due to impairments in motor speech, apraxia of speech, grammatical processing errors, or executive dysfunction.

Northwestern Anagram Test (NAT)

The NAT requires participants to organize words to create 10 grammatically correct sentences that describe a picture stimulus, allowing for assessment of sentence production independent of speech production, word-finding difficulties, or working memory capacity.31 This test was hypothesized to detect the grammatic impairments often seen in nfvPPA.

Semantic Word-Picture Matching Test

Participants hear a word and then choose the picture that matches this word from one of four semantically related stimuli. This 20-item task of single-word comprehension requires lexical-semantic knowledge, and similar tasks are consistently impaired in svPPA.18,32

Semantic Association Test

Participants are presented with two pairs of pictures (eg, squirrel-tree, squirrel balloon) and are asked to choose the pair with
semantically related objects (Northwestern Naming Battery; NNB, 16 trials). We expected svPPA patients to perform poorly on this task, as they consistently show impairment on similar tasks of semantic knowledge.

Noun and verb naming
In this NNB naming subtest, participants are first presented with 16 drawings of objects (ie, nouns) followed by 16 drawings of action items (ie, verbs). The outcome is total correct. Noun naming was hypothesized to require temporal lobe–mediated semantic knowledge and, therefore, svPPA patients were expected to show the greatest degree of impairment. Naming verbs has frontal and posterior parietotemporal correlates and was anticipated to be impaired in nfvPPA based on prior findings.

2.2.3 Other UDS language measures
Phonemic and semantic fluency
Phonemic fluency requires participants to generate as many words as possible in 60 seconds that begin with two letters (“F” and “L”). The outcome is total correct, summed across both trials. In semantic/category fluency, participants have 60 seconds to produce words belonging to two different semantic categories (“animals” and “vegetables”). We calculated the ratio of Phonemic:Semantic Fluency, as svPPA participants display relatively greater difficulty with semantic fluency compared to phonemic fluency, whereas nfvPPA participants evidence the opposite pattern.

Multilingual Naming Test (MINT)
This 32-item object picture naming task was designed to assess naming skills in speakers of multiple languages. Outcome was total score, including items named correctly with semantic, but not phonemic cues.

Number span forward
Participants are read increasingly longer sequences of numbers and asked to repeat them immediately in the order presented. This measure of phonological loop function was anticipated to help discriminate lvPPA.

2.3 Statistical analysis
Continuous demographic variables (ie, age, education) were compared between diagnostic groups using two separate linear models with a categorical predictor (ie, diagnostic group) and the demographic variable as the outcome, followed by pairwise group contrasts. A chi-square test was used to assess group differences in sex.

A cross-sectional comparison of FTLD-MOD performance among controls, bvFTD, and PPA variants was conducted using a multivariable linear model, with test performance as the outcome and diagnostic group as the categorical predictor of interest, adjusting for age, sex, education, and disease severity (Sum of Boxes). Predictors were chosen a priori and entered simultaneously. Post hoc group comparisons were made with Bonferroni multiple comparison correction. Disease severity was not a covariate in comparisons between patient groups and controls, as this would obscure the effect of interest.

In a follow-up, cross-sectional analysis, we estimated the potential use of these measures for differential diagnosis by fitting logistic regressions and receiver-operating characteristic (ROC) curves. We first selected promising measures for each of the three PPA variants based on results of pairwise group comparisons (ie, those differentiating groups after Bonferroni correction, or in the case of nfvPPA, measures in which nfvPPA performed the worst). We estimated the area under the curve (AUC) for differentiating each PPA variant from the other two PPA groups. We also compared lvPPA and nfvPPA directly, given that this is often the most challenging clinical comparison due to overlapping clinical features.

Longitudinal rates of decline were estimated by fitting linear mixed-effects models with random intercepts and slopes. Group differences in slopes were assessed by including a diagnosis-by-time interaction term with post hoc pairwise comparisons (Bonferroni corrected). Time was included as a continuous variable indicating years from baseline assessment. Models controlled for baseline age, sex, education, and severity (Sum of Boxes). For illustrative purposes, subject-specific slopes were extracted and plotted. Statistical tests of between-group differences were conducted using mixed-effects models, not by performing statistics on the subject-specific extracted slopes, as they are biased by shrinkage.

Sample size estimates to detect a 40% and 25% treatment effects (40% or 25% reduction in decline from timepoint 1 to 2, adjusted for 20% attrition) were derived based on the annualized change score between baseline and follow-up. For each measure, participants were included if they had received a second assessment of that measure within 2.5 years of their baseline visit.

3 RESULTS
Baseline demographics and clinical characteristics are displayed in Table 1. Groups differed on age and sex, but not education. Post hoc comparisons (all P-values Bonferroni-adjusted) showed that all four patient groups were older than controls (P < .001). LvPPA and nfvPPA were older than bvFTD (P-values < .05). NfvPPA were older than svPPA (P = .022), with no significant differences between svPPA and lvPPA (P = .196) or between lvPPA and nfvPPA (P = 1.0). Group differences were observed in CDR®+NACC FTLD Sum of Boxes, with svPPA and bvFTD having the greatest box scores.

3.1 Baseline performances
Ceiling effects were noted for controls on all measures (Table 2; Supplemental Figure S1). After Bonferroni correction, all diagnostic groups performed worse than controls on nearly every FTLD-MOD measure, with the exception of Word-Picture Matching. Similarly, on the
TABLE 1  Sample description, baseline demographics, and clinical characteristics: M (SD)

|                         | Controls | bvFTD   | svPPA   | nfvPPA  | lvPPA  | P-value |
|-------------------------|----------|---------|---------|---------|--------|---------|
| N                       | 581      | 612     | 185     | 168     | 109    |         |
| N with follow-up visits | 224      | 233     | 79      | 65      | 47     |         |
| Number of follow-up visits | 2.9 (1.2) | 2.9 (1.4) | 2.7 (0.9) | 2.7 (0.8) | 2.7 (1.0) |         |
| Length of follow-up (y) | 2.6 (1.6) | 2.3 (1.7) | 2.1 (1.2) | 1.8 (0.9) | 1.9 (1.1) |         |
| Age (y)                 | 50.2 (15.8) | 63.7 (8.5) | 65.0 (7.6) | 67.9 (8.2) | 67.8 (8.2) | <.001   |
| Education (y)           | 17.2 (10.6) | 17.3 (11.4) | 17.4 (10.8) | 17.7 (13.0) | 17.0 (8.3) | 0.27    |
| Sex (% male)            | 42.9     | 62.4    | 48.7    | 47      | 54.13  | <.001   |
| CDR® + NACC FTLD        |          |         |         |         |        |         |
| Sum of Boxes            | 0.0 (0.0) | 8.4 (4.2) | 7.6 (4.7) | 5.1 (4.0) | 5.1 (3.5) | <.001   |
| Global Score (N, %)     |          |         |         |         |        |         |
| 0                       | 581 (100)| –       | –       | –       | –      |         |
| 0.5                     | –        | 39 (6.4)| 35 (18.9)| 59 (35.1)| 37 (33.9)|         |
| 1                       | –        | 139 (22.7)| 52 (28.1)| 43 (25.6)| 38 (34.9)|         |
| 2                       | –        | 367 (60.0)| 80 (43.2)| 56 (33.3)| 29 (26.6)|         |
| 3                       | –        | 67 (11.0)| 18 (9.7)| 10 (6.0)| 5 (4.6) |         |

Note. Length of follow-up and number of follow-up visits calculated for participants with more than one time point.
Abbreviations: bvFTD = behavioral variant frontotemporal dementia; svPPA = semantic variant primary progressive aphasia; nfvPPA = nonfluent variant PPA; lvPPA = logopenic variant PPA; y = years; CDR® + NACC FTLD = Clinical Dementia Rating Scale plus National Alzheimer’s Coordinating Center FTLD Module

---

TABLE 2  Baseline FTLD-MOD subtest raw scores by diagnostic group

| Instrument                     | Controls | bvFTD   | svPPA   | nfvPPA  | lvPPA  |
|--------------------------------|----------|---------|---------|---------|--------|
| Word reading – regular         | Mean (SD)| Mean (SD)| Mean (SD)| Mean (SD)| Mean (SD) |
|                                 | 15.0 (0.1)| 14.7 (0.9)| 13.9 (2.4)| 13.8 (3.1)| 14.2 (2.4) |
| Word reading – irregular       | 14.5 (0.8)| 13.2 (2.4)| 8.9 (3.9)| 12.0 (3.6)| 11.5 (3.2) |
| Sentence repetition            | 4.5 (0.6)| 4.0 (1.1)| 3.5 (1.4)| 2.9 (1.7)| 2.2 (1.7) |
| Sentence reading               | 4.9 (0.4)| 4.4 (1.0)| 3.9 (1.3)| 3.4 (1.7)| 3.9 (1.4) |
| NW anagram total               | 9.3 (1.4)| 6.9 (2.7)| 6.9 (2.9)| 6.5 (3.0)| 5.8 (2.5) |
| Word-picture matching          | 20.0 (0.1)| 19.3 (1.8)| 17.2 (3.2)| 18.9 (3.1)| 19.4 (1.3) |
| Semantic associations          | 16.0 (0.2)| 14.4 (2.7)| 12.7 (3.0)| 14.9 (2.6)| 15.0 (2.3) |
| Noun naming                    | 15.9 (0.3)| 14.8 (2.3)| 8.5 (5.3)| 14.7 (2.8)| 13.1 (3.6) |
| Verb naming                    | 15.9 (0.3)| 14.2 (2.8)| 10.5 (4.5)| 13.3 (4.3)| 11.7 (4.3) |
| Other UDS measures             |          |         |         |         |        |
| Phonemic fluency               | 29.8 (7.7)| 15.9 (9.9)| 14.1 (7.4)| 11.0 (7.1)| 14.5 (6.6) |
| Semantic fluency               | 38.9 (7.9)| 19.3 (10.6)| 10.7 (7.4)| 18.3 (9.9)| 13.5 (7.6) |
| Phonemic: Semantic ratio       | 0.8 (0.2)| 0.9 (0.6)| 2.0 (2.6)| 0.6 (0.3)| 1.2 (0.9) |
| MINT                           | 30.3 (1.7)| 24.9 (6.9)| 8.6 (8.7)| 26.5 (6.7)| 19.9 (8.9) |
| Number span forward            | 9.0 (2.3)| 6.7 (2.4)| 6.5 (2.7)| 5.0 (2.3)| 3.8 (2.2) |

Note. Group comparisons include Bonferroni adjustments for multiple comparisons; P < .05.
All = all four diagnostic groups.
asvPPA was significantly higher than controls and bvFTD on Phonemic:Semantic ratio.
Abbreviations: bvFTD = behavioral variant frontotemporal dementia; svPPA = semantic variant primary progressive aphasia; nfvPPA = nonfluent variant PPA; lvPPA = logopenic variant PPA; NW = Northwestern; UDS = Uniform Data Set; MINT: Multilingual Naming Test
FIGURE 1 Uniform Data Set language measures for differential diagnosis and longitudinal monitoring of frontotemporal dementia (FTD) syndromes. Note. Figures 1A and B present the most promising cross-sectional (A) and longitudinal (B) measures for semantic variant primary progressive aphasia (PPA). Figures 1C and D present the most promising cross-sectional (C) and longitudinal (D) measures for non-fluent/agrammatic variant PPA. Figures 1E and F present the most promising cross-sectional (E) and longitudinal (F) measures for logopenic variant PPA. The most promising cross-sectional measures were determined based on receiver-operating characteristic curves, shown in Figure 2. The most promising longitudinal measures were determined based on estimates of the sample sizes needed to detect a 40% treatment effect (Table 4). The y-axis in 1C was truncated for illustrative purposes; four values were cut off from the svPPA group (10, 12, 14, and 19).

majority of measures, all PPA groups performed worse than bvFTD, with a few notable exceptions. On several measures (e.g., Noun Naming, Word-Picture Matching, Semantic Associations) only the svPPA group performed consistently worse than bvFTD. Only the lvPPA group performed worse than bvFTD on the NAT.

Several tests of semantic knowledge consistently differentiated svPPA from the other diagnostic groups. ROC analyses testing the discrimination of svPPA from the other two PPA variants suggested excellent accuracy for Noun Naming (area under the curve [AUC] = 0.82 [0.78, 0.86]), and acceptable discrimination for Semantic Associations
(AUC = 0.77 [0.73, 0.82]), Irregular Word Reading (AUC = 0.74 [0.69, 0.78]), and Word-Picture Matching (AUC = 0.73 [0.68, 0.78]). When all four predictors were added to the same logistic regression, Noun Naming (P < .001) and Semantic Association (P = .043) remained as independent predictors of diagnosis. Including both tasks in the ROC analysis, however, did not provide a clinically significant improvement in accuracy (AUC = 0.83 [0.79, 0.88]) compared to Noun Naming alone. The Multilingual Naming Test (MINT) (Figures 1A and 2A) outperformed all FTLD-MOD tasks in discriminating svPPA (AUC = 0.89 [0.85, 0.93]), although there was overlap in the 95% confidence intervals.

The non-fluent/agrammatic and logopenic variants can often be difficult to discriminate in clinical practice. Similarly, their performance on many measures were not statistically different at the group level. Sentence Reading appeared to be the most promising FTLD-MOD measure for differentiating nfvPPA from the other two PPA syndromes. ROC analysis, however, suggested minimal utility for this task in discriminating nfvPPA from svPPA and lvPPA (AUC = 0.55 [0.49, 0.61]) and nfvPPA from lvPPA (AUC = 0.55 [0.48, 0.62]). NfvPPA participants were the worst performing group on Phonemic Fluency, although their performance did not differ significantly from lvPPA. A ratio of Phonemic:Semantic Fluency (Figures 1C and 2B) provided excellent accuracy in classifying nfvPPA from svPPA and lvPPA (AUC = 0.84 [0.79, 0.89]), and from lvPPA only (AUC = 0.82, [0.75, 0.90]).

The FTLD-MOD test that performed best at discriminating lvPPA was Sentence Repetition, which was most impaired in lvPPA. ROC analysis revealed only minimal accuracy in separating lvPPA from the other two PPA syndromes (AUC = 0.67 [0.60, 0.72]) and from nfvPPA (AUC = 0.61 [0.54, 0.69]). Acceptable discrimination was observed for Number Span Forward in classifying lvPPA compared to nfvPPA and nfvPPA (Figures 1E and 2C; AUC = 0.72 [0.64, 0.79]), but only minimal accuracy for classifying lvPPA versus nfvPPA (AUC = 0.64 [0.55, 0.74]).

3.2 Longitudinal performances and sample size estimates

Annualized rates of decline for each diagnostic group are presented in Table 3, along with group comparisons. Similar to the cross-sectional results, the most consistent differences in longitudinal trajectories were observed for the svPPA cases (Table 3; Figure 1C and D). To understand the effect sizes of longitudinal decline, we calculated sample sizes for a planned clinical trial (Table 4). Of the FTLD-MOD tasks, the most encouraging results were observed for svPPA, with Noun and Verb Naming and Word-Picture Matching showing the most promise. Indeed, these measures suggest a trial in svPPA would require only 260 participants to detect a 40% treatment effect if Noun Naming was used, compared to 575 if Sum of Boxes was used. Of note, for all diagnoses, standard UDS language measures outperform the FTLD-MOD with regard to sample size (Table 4). Longitudinal trajectories for the best-performing measures in each variant, based on sample size estimates, are displayed in Figure 1B, D, and F.
## TABLE 3  Raw annualized change and adjusted group comparisons

| Instrument                        | Controls | bvFTD | svPPA | nfvPPA | lvPPA | Worse than controls | Worse than bvFTD | PPA comparison |
|-----------------------------------|----------|-------|-------|--------|-------|---------------------|------------------|---------------|
|                                   | Annual change (SE) | Annual change (SE) | Annual change (SE) | Annual change (SE) | Annual change (SE) |                     |                 |
| Word reading - regular           | 0.0003 (0.00) | -0.05 (0.04) | -0.87 (0.20) | -0.50 (0.22) | -0.38 (0.16) | bv, sv             | sv               |
| Word reading - irregular         | 0.019 (0.01) | -0.13 (0.04) | -1.32 (0.15) | -0.70 (0.20) | -1.16 (0.23) | all                | lv, sv           |
| Sentence repetition               | -0.0003 (0.01) | -0.08 (0.03) | -0.39 (0.09) | -0.36 (0.13) | -0.55 (0.08) | lv, sv, nfv        | lv, sv           |
| Sentence reading                  | 0.005 (0.01) | -0.07 (0.02) | -0.41 (0.09) | -0.30 (0.13) | -0.48 (0.12) | all                | lv, sv           |
| NW anagram total                  | 0.04 (0.02) | 0.00 (0.06) | -0.37 (0.15) | -0.61 (0.23) | -0.53 (0.16) | sv, nfv            | sv, nfv          |
| Word-picture matching             | 0.001 (0.00) | -0.35 (0.08) | -1.37 (0.14) | -0.28 (0.16) | -0.56 (0.23) | bv, lv, sv         | sv               |
| Semantic associations             | -0.003 (0.01) | -0.65 (0.11) | -0.86 (0.15) | -0.23 (0.10) | -0.53 (0.17) | bv, lv, sv         | none             |
| Noun naming                      | 0.004 (0.00) | -0.68 (0.11) | -2.32 (0.22) | -0.31 (0.21) | -1.36 (0.30) | bv, lv, sv         | sv               |
| Verb naming                       | 0.013 (0.01) | -1.01 (0.15) | -2.18 (0.23) | -0.80 (0.25) | -1.36 (0.26) | all                | sv               |

**Other UDS measures**

| instrument                        | Controls | bvFTD | svPPA | nfvPPA | lvPPA | worse than controls | worse than bvFTD | PPA comparison |
|-----------------------------------|----------|-------|-------|--------|-------|---------------------|------------------|---------------|
|                                   |          |       |       |        |       |                     |                  |
| Phonemic fluency                  | 0.59 (0.16) | -1.18 (0.23) | -1.46 (0.36) | -1.89 (0.46) | -2.10 (0.40) | all                | none            |
| Semantic fluency                  | -0.13 (0.13) | -2.38 (0.25) | -3.29 (0.27) | -3.32 (0.40) | -2.49 (0.57) | all                | none            |
| Phonemic Semantic ratio           | 0.02 (0.005) | 0.01 (0.01) | 0.23 (0.15) | 0.01 (0.02) | 0.001 (0.06) | sv                | sv               |
| MINT                              | 0.14 (0.04) | -1.51 (0.25) | -2.38 (0.46) | -1.28 (0.59) | -3.52 (0.68) | all                | lv               |
| Number span forward               | 0.03 (0.05) | -0.16 (0.06) | -0.59 (0.12) | -0.66 (0.19) | -0.60 (0.12) | all                | sv, nfv          |

**Note.** Annualized changes scores were derived from linear mixed-effects models.

Group comparisons among diagnostic groups control for baseline age, baseline CDR®+NACC FTLD Sum of Boxes, Education, and Sex.

Post hoc comparisons with controls include demographic covariates but do not control for disease severity (CDR®+NACC FTLD Sum of Boxes).

"Worse than controls" indicate more steeper longitudinal decline compared to controls.

"Worse than bvFTD" indicates a steeper longitudinal decline compared to bvFTD.

svPPA showed more rapid increases in the ratio of Phonemic:Semantic fluency compared to all other groups.

Abbreviations: SE = standard error; bvFTD = behavioral variant frontotemporal dementia; svPPA = semantic variant primary progressive aphasia; nfvPPA = nonfluent variant PPA; lvPPA = logopenic variant PPA; Sent = Sentence; NW = Northwestern; UDS = Uniform Data Set; MINT: Multilingual Naming Test.
| Instrument                        | Effect Size (%) | bvFTD n | Sample size | svPPA n | Sample size | nfvPPA n | Sample size | lvPPA n | Sample size |
|----------------------------------|-----------------|---------|-------------|---------|-------------|---------|-------------|---------|-------------|
| Word reading - regular           | 40              | 218     | 5540        | 78      | 1593        | 43      | 1883        | 45      | 2390        |
|                                  | 25              | 14173   |             | 4070    |             | 4815    |             | 6115    |             |
| Word reading - irregular         | 40              | 217     | 2968        | 77      | 565         | 43      | 1125        | 45      | 3260        |
|                                  | 25              | 7590    |             | 1443    |             | 2875    |             | 8340    |             |
| Sentence repetition              | 40              | 215     | 2300        | 72      | 1818        | 41      | 1495        | 42      | 918         |
|                                  | 25              | 5883    |             | 4650    |             | 3820    |             | 2345    |             |
| Sentence reading                 | 40              | 213     | 6385        | 72      | 1738        | 40      | 4753        | 42      | 5033        |
|                                  | 25              | 16338   |             | 4445    |             | 12160   |             | 12878   |             |
| NW anagram total                 | 40              | 173     | 4248        | 54      | 3195        | 48      | 5298        | 34      | 22613       |
|                                  | 25              | 10870   |             | 8170    |             | 13553   |             | 57880   |             |
| Word-picture matching            | 40              | 227     | 2943        | 75      | 405         | 55      | 5138        | 47      | 3050        |
|                                  | 25              | 7528    |             | 1030    |             | 13148   |             | 7800    |             |
| Semantic associations            | 40              | 209     | 1615        | 64      | 1305        | 53      | 4158        | 45      | 1650        |
|                                  | 25              | 4133    |             | 3338    |             | 10640   |             | 4223    |             |
| Noun naming                      | 40              | 219     | 1875        | 66      | 260         | 44      | 5515        | 38      | 770         |
|                                  | 25              | 4795    |             | 663     |             | 14110   |             | 1968    |             |
| Verb naming                      | 40              | 218     | 1790        | 66      | 310         | 44      | 8795        | 38      | 683         |
|                                  | 25              | 4575    |             | 785     |             | 22508   |             | 1738    |             |
| Other measures                   |                 |         |             |         |             |         |             |         |             |
| CDR®+NACC FTLD SB                | 40              | 227     | 500         | 78      | 575         | 58      | 775         | 48      | 388         |
|                                  | 25              | 1275    |             | 1465    |             | 1980    |             | 983     |             |
| Phonemic fluency                 | 40              | 153     | 898         | 44      | 2878        | 27      | 883         | 27      | 208         |
|                                  | 25              | 2293    |             | 7360    |             | 2255    |             | 528     |             |
| Semantic fluency                 | 40              | 208     | 673         | 69      | 323         | 43      | 518         | 40      | 980         |
|                                  | 25              | 1718    |             | 823     |             | 1320    |             | 2505    |             |
| Phonemic: Semantic ratio         | 40              | 148     | 44498       | 35      | 1833        | 24      | 4053        | 27      | 3965613     |
|                                  | 25              | 113908  |             | 4685    |             | 10370   |             | 10151965|             |
| MINT                             | 40              | 156     | 1025        | 34      | 198         | 32      | 1195        | 27      | 448         |
|                                  | 25              | 2620    |             | 500     |             | 3050    |             | 1138    |             |
| Number span forward              | 40              | 156     | 1715        | 46      | 2240        | 29      | 840         | 29      | 2590        |
|                                  | 25              | 4388    |             | 5730    |             | 2145    |             | 6628    |             |

Note. Estimates are for the total sample size required for a clinical trial (both arms) to detect “moderate” (40%) or “small” (25%) treatment effects, accounting for 20% expected attrition.

n is the number of participants that went into the estimate.

bvFTD, behavioral variant frontotemporal dementia; lvPPA, logopenic variant PPA; nfvPPA, non-fluent variant PPA; PPA, primary progressive aphasia; svPPA, semantic variant PPA; Sent = Sentence; NW = Northwestern; CDR®+NACC FTLD SB = Clinical Dementia Rating Scale plus National Alzheimer’s Coordinating Center FTLD Module Sum of Boxes; MINT: Multilingual Naming Test

4 | DISCUSSION

Clinical care and clinical trials in FTLD require measures that accurately differentiate FTD clinical syndromes and are sensitive to longitudinal change. The FTLD-MOD has shown promise for differentiating between the clinical syndromes of bvFTD and PPA.16 Our study extended prior work by investigating the utility of FTLD-MOD and UDS language tasks for cross-sectional differential diagnosis among the PPA subtypes and their potential for longitudinal monitoring using a large publicly available data set. We confirmed prior findings that PPA participants generally performed worse than bvFTD, and the bvFTD group also performed statistically worse than controls on nearly every measure.

Cross-sectional analyses indicated that a subset of FTLD-MOD language tests consistently differentiate svPPA from controls, bvFTD, and other PPA subtypes. Noun Naming showed the most promise for
differentiating svPPA from lvPPA and nfvPPA (AUC = 0.82). Word-Picture Matching (AUC = 0.73) and Semantic Association (AUC = 0.77) were also promising. Another naming test from the standard UDS, the MINT, evidenced excellent discrimination of svPPA from other PPA syndromes (AUC = 0.89); adding additional FTLD-MOD measures to the model did not strengthen prediction. The MINT was previously found to discriminate AD and amnestic mild cognitive impairment (MCI) from controls.17

FTLD-MOD measures showed less promise for discriminating lvPPA and nfvPPA. Although Sentence Repetition performance was worst in lvPPA, as expected, the ROC analysis suggested minimal diagnostic utility (AUC = 0.67). A similar finding was observed for nfvPPA: although Sentence Reading was most impaired in nfvPPA, it showed little promise as a diagnostic measure (AUC = 0.55). Patterns of performance on UDS language measures were consistent with the hypothesized directions, with Number Span Forward performing the best for classifying lvPPA (AUC = 0.72) and the ratio of Phonemic:Semantic Fluency for nfvPPA (AUC = 0.84); of interest, the ratio of Phonemic:Semantic Fluency also performed well at classifying nfvPPA versus lvPPA (AUC = 0.82).

The results of the longitudinal analysis paralleled the cross-sectional analysis in several ways. Most groups showed greater longitudinal decline on all measures compared to controls, and in many cases, compared to bvFTD. SvPPA emerged as the group that was most accurately discriminated and had the lowest sample size estimates needed to detect a 40% treatment effect. Noun Naming and Word-Picture Matching had reasonable sample size estimates (260 and 405, respectively). Consistent with the cross-sectional analysis, longitudinal UDS measures such as the MINT (n = 198) outperformed the FTLD-MOD measures in svPPA. We also replicated prior work showing promising sample sizes using Semantic Fluency for svPPA.18 Also analogous to the cross-sectional results, the FTLD-MOD tasks performed less well at longitudinally differentiating lvPPA and nfvPPA. Sample size estimates were unrealistic for all FTLD-MOD tasks in lvPPA and nfvPPA, with UDS measures performing best. Semantic Fluency in nfvPPA (n = 518) and Phonemic Fluency in lvPPA (n = 208) were the most promising for these subtypes. A prior study of nfvPPA showed a similar estimate (n = 507) to detect the same treatment effect with semantic fluency.18 Sum of Boxes was among the best measures for all groups, consistent with prior work.18,19

There are several potential explanations for the difficulty differentiating nfvPPA and lvPPA, and the high sample size estimates needed to detect meaningful change in these variants. For example, the lack of differentiation of the nfvPPA group from others on the NAT could result from the fact that the nfvPPA category can include individuals with either grammatical processing deficits or motor speech deficits.4,48 Although the NAT was designed to avoid speech output problems, some non-fluent/agrammatic cases may lack agrammatism. Furthermore, our finding that only the lvPPA group performed lower than the bvFTD group on the NAT is difficult to interpret. It may indicate that the NAT is sensitive to their known working memory deficits.

Many FTLD-MOD measures, although theoretically sound, have psychometric problems. Notably, controls performed at ceiling on most measures. In addition, several measures have only a few items, which reduces the reliability and variance of the tests. For example, repetition tasks are commonly used clinically to distinguish lvPPA from other diagnoses, but the FTLD-MOD repetition task includes only five items of familiar content. Recent work using a 20-item repetition task,49 with phrases varying in length, meaningfulness, and familiarity, evidenced 89% accuracy for classifying PPA subtypes,50 compared to 67% in the current study. Future studies should seek to evaluate repetition tasks with more items, longer, less frequent sentence structures, and greater phonemic complexity.50

The ability of a single task to discriminate among clinical syndromes is a high bar and does not map directly to clinical diagnostic practices. In the clinic, neuropsychological tests are only one component of a clinician’s diagnostic armamentarium and are not considered in isolation. In this research study; however, we wanted to limit circularity and focus on FTLD-MOD tasks, which were not used for syndromic diagnoses. Future studies could explore the utility of adding FTLD-MOD tasks to other aspects of the clinical workup. A second limitation is that we focused on total scores for these tasks. The nature of language disorders, however, is that poor performance on any given test could be the consequence of several different language impairments. Future work should include a granular analysis of additional test variables, such as types of errors, that might help improve diagnostic distinctions. Although one of this study’s strengths was the large sample size, there are important aspects of phenotyping (eg, identifying familial cases) that were not conducted in this sample. Finally, the samples that completed UDS measures (collection started in March 2015) were often smaller than those that completed the FTLD-MOD, and thus results warrant replication.

In summary, the FTLD-MOD tasks perform well at distinguishing PPA from bvFTD and distinguishing svPPA from other PPA subtypes. UDS tasks perform best at discriminating among all PPA subtypes. Longitudinal results paralleled the cross-sectional findings and suggested that naming tests are potentially useful as trial endpoints. SvPPA stands out as a strong indication for treatment trials, given the accuracy of diagnosis, the array of sensitive outcomes, and strong clinico-pathological correlation. These results may help clinicians with their test selection and highlight the need for developing and validating better endpoints for disease monitoring in lvPPA and nfvPPA.

ACKNOWLEDGMENTS

The National Alzheimer’s Coordinating Center (NACC) database is funded by National Institute on Aging/National Institutes of Health (NIA/NIH) Grant U01 AG16976. NACC data are contributed by the NIA-funded ADCs: National Science Foundation; P30 AG019610 (PI Eric Reiman, MD), P30 AG013846 (PI Neil Kowall, MD), P30 AG062428-01 (PI James Leverenz, MD), P50 AG08702 (PI Scott Small, MD), P50 AG025688 (PI Allan Levey, MD, PhD), P50 AG047266 (PI Todd Golde, MD, PhD), P30 AG010133 (PI Andrew Saykin, PsyD), P50 AG005146 (PI Marilyn Albert, PhD), P30 AG005146 (PI Eric Reiman, MD), P30 AG013846 (PI Neil Kowall, MD), P30 AG010133 (PI Andrew Saykin, PsyD), P50 AG005146 (PI Marilyn Albert, PhD), P30 AG062421-01 (PI Bradley Hyman, MD, PhD), P30 AG062422-01 (PI Ronald Petersen, MD, PhD), P50 AG005146 (PI Mary Sano, PhD), P30 AG008051 (PI Thomas Wiesniewski, MD), P30 AG013854 (PI Robert Vassar, PhD), P30...
AG008017 (PI Jeffrey Kaye, MD), P30 AG010161 (PI David Bennett, MD), P50 AG047366 (PI Victor Henderson, MD, MS), P30 AG010129 (PI Charles DeCarli, MD), P50 AG016573 (PI Frank LaFerla, PhD), P30 AG062429-01 (PI James Brewer, MD, PhD), P50 AG023501 (PI Bruce Miller, MD), P30 AG035982 (RI Russell Swerdlow, MD), P30 AG028383 (PI Linda Van Eldik, PhD), P30 AG053760 (PI Henry Paulson, MD, 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), P30 AG049638 (PI Suzanne Craft, PhD), P50 AG005136 (PI Thomas Grabowski, MD), P30 AG062715-01 (PI Sanjay Asthana, MD, FRCPI), P50 AG005681 (PI John Morris, MD), and P50 AG047270 (PI Stephen Strittmatter, MD, PhD). This work was supported in part by the National Institutes of Health/National Institute on Aging NIA/NIH (AG061253, AG058752, AG032306, AG017586, AG054519), including the Advancing Research & Treatment for Frontotemporal Lobar Degeneration (ARTFL: (U54 NS092089)) and Longitudinal Evaluation of Familial Frontotemporal Dementia Subjects (LEFFTDS: (U01 AG045390)) programs, which are now combined into the ARTFL LEFFTDS Longitudinal Frontotemporal Lobar Degeneration ALLFTD (U19 AG063911) program. This work was also supported in part by the Larry H. Hillblom Foundation (2018-A-0-FEL, 2017-A-004-FEL).

CONFLICT OF INTEREST

BB has served as an investigator for clinical trials sponsored by Axovant and Biogen. He receives royalties from the publication of a book entitled Behavioral Neurology of Dementia (Cambridge Medicine, 2009, 2017). He serves on the Scientific Advisory Board of the Tau Consortium. He receives research support from the NIH, the Mayo Clinic Dorothy and Harry T. Mangurian Jr. Lewy Body Dementia Program, and the Little Family Foundation. HJR has a consulting agreement Ionis Pharmaceuticals. JHK has provided consultation for Biogen. He helped develop the Delis–Kaplan Executive Function System and receives royalties from Pearson Education, Inc., for helping to develop the California Verbal Learning Test. ALB receives research support from NIH, the Tau Research Consortium, the Association for Frontotemporal Degeneration, Bluefield Project to Cure Frontotemporal Dementia, Corticobasal Degeneration Solutions, the Alzheimer’s Drug Discovery Foundation, and the Alzheimer’s Association. He has served as a consultant for Aeton, Abbvie, Alector, Amgen, Arkuda, Ionis, Iperian, Janssen, Merck, Novartis, Samumed, Toyama, and UCB; and has received research support from Avid, Biogen, BMS, C2N, CorTice, Eli Lilly, Forum, Genentech, Janssen, Novartis, Pfizer, Roche, and TauRx.

REFERENCES

1. Olney NT, Spina S, Miller BL. Frontotemporal Dementia. Neurol Clin. 2017;35(2):337-348.
2. Rascovsky K, Hodges JR, Knopman D, et al. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. Brain. 2011;134(9):2456-2477.
3. Perry DC, Brown JA, Possin KL, et al. Clinico pathological correlations in behavioural variant frontotemporal dementia. Brain. 2017;140(12):3329-3345.
4. Gorno-Tempini ML, Hillis AE, Weintraub S, et al. Classification of primary progressive aphasia and its variants. Neurology. 2011;76(11):1006-1014.
5. Spinelli EG, Mandelli ML, Miller ZA, et al. Typical and atypical pathology in primary progressive aphasia variants. Ann Neurol. 2017;81(3):430-443.
6. Santos-Santos MA, Mandelli ML, Binney RJ, et al. Features of patients with nonfluent/agrammatic primary progressive aphasia with underlying progressive supranuclear palsy pathology or corticobasal degeneration. JAMA Neurol. 2016;94(13):1-10.
7. Gorno-Tempini ML, Brambati SM, Ginex V, et al. The logopenic/phonological variant of primary progressive aphasia. Neurology. 2008;71(16):1227-1234.
8. Harris JM, Gali C, Thompson JC, et al. Classification and pathology of primary progressive aphasia. Neurology. 2013;81(21):1832-1839.
9. Mesulam M-M, Weintraub S, Rogalski EJ, Wieneke C, Geula C, Bigio EH. Asymmetry and heterogeneity of Alzheimer’s and frontotemporal pathology in primary progressive aphasia. Brain. 2014;137(Pt 4):1176-1192.
10. Mesulam M, Wicklund A, Johnson N, et al. Alzheimer and frontotemporal pathology in subsets of primary progressive aphasia. Ann Neurol. 2008;63(6):709-719.
11. Knopman DS, Roberts RO. Estimating the number of persons with frontotemporal lobar degeneration in the US population. J Mol Neurosci. 2011;45(3):330-335.
12. Morhardt D, Weintraub S, Khayum B, et al. The CARE pathway model for dementia: psychosocial and rehabilitative strategies for care in young-onset dementias. Psychiatr Clin North Am. 2015;38(2):333-352.
13. Weintraub S, Besser L, Dodge HH, et al. Version 3 of the Alzheimer Disease Centers’ Neuropsychological Test Battery in the Uniform Data Set (UDS). Alzheimer Dis Assoc Disord. 2018;32(1):10-17.
14. Ritter AR, Leger GC, Miller JB, Banks SJ. Neuropsychological testing in pathologically verified Alzheimer disease and frontotemporal dementia: how well do the uniform data set measures differentiate between diseases? Alzheimer Dis Assoc Disord. 2017;31(3):187-191.
15. Rascovsky K, Moran E, Baehr L, Irwin D, McMillan C, Grossman M. Utility and neuroanatomical correlates of the FTLD-NACC neuropsychology module in the differential diagnosis of behavioral variant frontotemporal dementia and Alzheimer’s disease. Neurology. 2015;84(SUPPL. 14). https://n.neurology.org/content/84/14_Supplement/P1.1220. Accessed July 26, 2020.
16. Gelen T, Teylan MA, Besser L, Pollner E, Moshkovich A, Weintraub S. Measurement and characterization of distinctive clinical phenotypes using the Frontotemporal Lobar Degeneration Module (FTLD-MOD). Alzheimer Dis Assoc Disord. 2020;16(6):918-925. https://doi.org/10.1002/alz.12098.
17. VandeVrede L, Ljubenkov PA, Rojas JC, Welch AE, Boxer AL. Four-repeat tauopathies: current management and future treatments. Neurorther J Am Soc Exp Neurother. 2020.
18. Staffaroni AM, Ljubenkov PA, Kornak J, et al. Longitudinal multimodal imaging and clinical endpoints for frontotemporal dementia clinical trials. Brain. 2019;142(2):443-459.
19. Knopman DS, Kramer JH, Boeve BF, et al. Development of methodology for conducting clinical trials in frontotemporal lobar degeneration. Brain. 2008;131(Pt 11):2957-2968.
20. Tsai RM, Boxer AL. Therapy and clinical trials in frontotemporal dementia: past, present, and future. J Neurochem. 2016;138:211-221.
21. Staffaroni AM, Bajorek L, Cazaletto KB, et al. Assessment of executive function declines in presymptomatic and mildly symptomatic familial frontotemporal dementia: nih-EXAMINERS as a potential clinical trial endpoint. Alzheimer Dis Assoc Disord. 2020;16(1):11-21.
22. Heuer HW, Wang P, Rascovsky K, et al. Comparison of sporadic and familial behavioral variant frontotemporal dementia (FTD) in a North American cohort. Alzheimer Dis Assoc Disord. 2020;16(1):60-70.
23. Boeve B, Bove J, Brannelly P, et al. The longitudinal evaluation of familial frontotemporal dementia subjects protocol: framework and methodology. *Alzheimer's Dement*. 2020;16(1):22-36.

24. Miyagawa T, Brushaber D, Syrjanen J, et al. Use of the CDR® plus NACC FTLD in mild FTLD: data from the ARTFL/LEFFTDS consortium. *Alzheimer's Dement*. 2020;16(1):79-90.

25. Miyagawa T, Brushaber D, Syrjanen J, et al. Utility of the global CDR® plus NACC FTLD rating and development of scoring rules: data from the ARTFL/LEFFTDS Consortium. *Alzheimer's Dement*. 2020;16(1):106-117.

26. Knopman DS, Weintraub S, Pankratz VS. Language and behavior domains enhance the value of the clinical dementia rating scale. *Alzheimers Dement*. 2011;7(3):293-299.

27. Toller G, Ranasinghe K, Cobigo Y, et al. Revised self-monitoring scale: a potential endpoint for frontotemporal dementia clinical trials. *Neurology*. 2020;94(22):e2384-e2395.

28. Rankin KP, Gorno-Tempini ML, Allison SC, et al. Structural anatomy of empathy in neurodegenerative disease. *Brain*. 2006;129(Pt 11):2945-2956.

29. Rankin KP, Kramer JH, Miller BL. Patterns of cognitive and emotional empathy in frontotemporal lobar degeneration. *Cogn Behav Neurol*. 2005;18(1):28-36.

30. National Alzheimer’s Coordinating Center. https://www.alz.washington.edu/. Accessed August 1, 2020.

31. Weintraub S, Mesulam M-M, Wierenga RP, et al. The longitudinal evaluation of familial frontotemporal dementia subjects protocol: framework and methodology. *Alzheimer's Dement*. 2012;11(6):545-555.

32. Mesulam M, Rogalski E, Wierenga RP, et al. Neurology of anoma in the semantic variant of primary progressive aphasia. *Brain*. 2009;132(9):2553-2565.

33. Thompson CK, Lukic S, King MC, Mesulam MM, Weintraub S. Verb and noun deficits in stroke-induced and primary progressive aphasia: the Northwestern Naming Battery. *Aphasiology*. 2011;26(5):632-655.

34. Klein LA, Buchanan JA. Psychometric properties of the Pyramids and Palm Trees Test. *J Clin Exp Neuropsychol*. 2009;31(7):803-808.

35. Moore K, Convery R, Bocchetta M, et al. A modified Camel and Cactus Test detects presymptomatic semantic impairment in genetic frontotemporal dementia within the GENFI cohort. *Appl Neuropsychol Adult*. 2020;1:8.

36. Vonk JMJ, Borghesani V, Battistella G, et al. Verbal semantics and the left dorsolateral anterior temporal lobe: a longitudinal case of bilateral temporal degeneration. *Aphasiology*. 2020;34(7):865-885.

37. Reilly J, Peelle JE, Antonucci SM, Grossman M. Anoma as a marker of distinct semantic memory impairments in Alzheimer’s disease and semantic dementia. *Neuropsychology*. 2011;25(4):413-426.

38. Montembeault M, Brambati SM, Joubert S, et al. Naming unique entities in the semantic variant of primary progressive aphasia and Alzheimer’s disease: towards a better understanding of the semantic impairment. *Neuropsychologia*. 2017;95:11-20.

39. Rogers SL, Friedman RB. The underlying mechanisms of semantic memory loss in Alzheimer’s disease and semantic dementia. *Neuropsychologia*. 2008;46(1):12-21.

40. Birn RM, Kenworthy L, Case L, et al. Neural systems supporting lexical search guided by letter and semantic category cues: a self-paced overt response fMRI study of verbal fluency. *Neuroimage*. 2010;49(1):1099-1107.

41. Baldo JV, Schwartz S, Wilkins D, Drønkers NF. Role of frontal versus temporal cortex in verbal fluency as revealed by voxel-based lesion symptom mapping. *J Int Neuropsychol Soc*. 2006;12(6):896-900.

42. Henry JD, Crawford JR. A meta-analytic review of verbal fluency performance following focal cortical lesions. *Neuropsychology*. 2004;18(2):284-295.

43. Hodges JR, Patterson K. Semantic dementia: a unique clinicopathological syndrome. *Lancet Neurol*. 2007;6(11):1004-1014.

44. Grossman M. The non-fluent/agrammatic variant of primary progressive aphasia. *Lancet Neurol*. 2012;11(6):545-555.

45. Gollan TH, Weissberger GH, Runnqvist E, Montoya RI, Cera CM. Self-ratings of Spoken Language Dominance: a Multi-Lingual Naming Test (MINT) and preliminary norms for young and aging spanish-english bilinguals. *Biling*. 2012;15(3):594-615.

46. Sakpal TV. Sample size estimation in clinical trial. *Perspect Clin Res*. 2010;1(2):67-69.

47. Ivanova I, Salmon DP, Gollan TH. The multilingual naming test in Alzheimer’s disease: clues to the origin of naming impairments. *J Int Neuropsychol Soc*. 2013;19(3):272-283.

48. Mesulam M-M, Rogalski EJ, Wierenga RP, et al. Primary progressive aphasia and the evolving neurology of the language network. *Nat Rev Neurol*. 2014;10(10):554-569.

49. Bayles KA, Tomoeda CK, Rein JA. Phrase repetition in Alzheimer’s disease: effect of meaning and length. *Brain Lang*. 1996;54(2):246-261.

50. Lukic S, Mandelli ML, Welch A, et al. Neurocognitive basis of repetition deficits in primary progressive aphasia. *Brain Lang*. 2019;194:35-45.

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section at the end of the article.