OPEN

Development of a Speech-based Composite Score for Remotely Quantifying Language Changes in Frontotemporal Dementia

Jessica Robin, PhD, Mengdan Xu, MS,* Liam D. Kaufman, MS,* William Simpson, PhD,*† Stella McCaughey, PhD,‡ Nadine Tatton, PhD,‡ Charles Wolfus, JD,‡ and Michael Ward, PhD‡*

Background: Changes to speech and language are common symptoms across different subtypes of frontotemporal dementia (FTD). These changes affect the ability to communicate, impacting everyday functions. Accurately assessing these changes may help clinicians to track disease progression and detect response to treatment.

Objective: To determine which aspects of speech show significant change over time and to develop a novel composite score for tracking speech and language decline in individuals with FTD.

Method: We recruited individuals with FTD to complete remote digital speech assessments based on a picture description task. Speech samples were analyzed to derive acoustic and linguistic measures of speech and language, which were tested for longitudinal change over the course of the study and were used to compute a novel composite score.

Results: Thirty-six (16 F, 20 M; M_{age} = 61.3 years) individuals were enrolled in the study, with 27 completing a follow-up assessment 12 months later. We identified eight variables reflecting different aspects of language that showed longitudinal decline in the FTD clinical syndrome subtypes and developed a novel composite score based on these variables. The resulting composite score demonstrated a significant effect of change over time, high test–retest reliability, and a correlation with standard scores on various other speech tasks.

Conclusion: Remote digital speech assessments have the potential to characterize speech and language abilities in individuals with FTD, reducing the burden of clinical assessments while providing a novel measure of speech and language abilities that is sensitive to disease and relevant to everyday function.

Key Words: frontotemporal dementia, speech, language, digital, composite

(*Cogn Behav Neurol* 2023;36:237–248)

bvFTD = behavioral variant frontotemporal dementia. **FTD** = frontotemporal dementia. **HC** = healthy controls. **ICC** = intraclass correlation. **PPA** = primary progressive aphasia. **WLA** = Winterlight Assessment.

Frontotemporal dementia (FTD) is a progressive neurodegenerative disease that is caused by atrophy of the frontal and temporal lobes, resulting in a set of clinical subtypes with differing behavioral, cognitive, and motor symptoms (Convery et al, 2019; Ljubenkov and Miller, 2016). Progressive speech and language deficits are a key clinical feature of FTD, with different symptoms being used to differentiate subtypes of primary progressive aphasia (PPA): nonfluent/agrammatic variant PPA, semantic variant PPA, logopenic variant PPA (Gorno-Tempini et al, 2011), and behavioral variant frontotemporal dementia (bvFTD) (Geraudie et al, 2021; Hardy et al, 2016).

Speech and Language Changes in FTD Subtypes

Nonfluent/agrammatic PPA is associated with reduced fluency of speech, word-finding difficulty, agrammatism, and apraxia of speech; *semantic PPA* is associated with a loss of word meanings, resulting in naming difficulties and a loss of semantic knowledge (Gorno-Tempini et al, 2011); and *logopenic PPA* is associated with decreased speech output and impairments in naming and repetition. Logopenic PPA has been associated with

Received for publication August 4, 2022; accepted April 7, 2023.
From *Winterlight Labs, Incorporated, Toronto, Ontario, Canada; †Department of Psychiatry and Behavioural Neuroscience, McMaster University, Hamilton, Ontario, Canada; and ‡Alector, Incorporated, San Francisco, California.
Portions of this work were presented at the Clinical Trials in Alzheimer's Disease Conference, November 2021, Boston, Massachusetts, and the annual meeting of the American Academy of Neurology, April 2022, Seattle, Washington.

J.R., M.X., L.D.K., and W.S. are full-time employees of Winterlight Labs, Inc. and hold financial interests in Winterlight Labs. S.M., N.T., C.W., and M.W. are or were employees of Alector LLC and may have an equity interest in Alector, Inc.
Correspondence: Jessica Robin, PhD, 46 Hayden Street, Suite 400, Toronto, Ontario, Canada M4Y 1V8 (email: jessica@winterlightlabs.com).
Supplemental digital content is available for this article. Direct URL citations are provided in the HTML and PDF versions of this article on the journal's website, www.cogbehavneurol.com.

Alzheimer's pathology (Etcheverry et al, 2012; Henry and Gorno-Tempini, 2010; Montembeault et al, 2018; Rohrer et al, 2013). Historically, bvFTD has been associated with mild or absent language impairments, but newer research has increasingly highlighted that speech and language deficits exist in bvFTD and in FTD–ALS (frontotemporal dementia with amyotrophic lateral sclerosis) spectrum disorder (Hardy et al, 2016; Luzzi et al, 2020; Nevler et al, 2017; Nevler, Ash, McMillan, et al, 2020; Rogalski et al, 2011; Saxon et al, 2017).

Speech and Language Assessments in FTD

Assessing speech and language abilities in individuals with FTD is important for determining disease subtype and for tracking disease progression. The addition of language and behavioral scales to the Clinical Dementia Rating Scale (Morris, 1993) has resulted in an improved clinical tool, the CDR plus NACC FTLD (Clinical Dementia Rating plus NACC frontotemporal lobar degeneration; Miyagawa et al, 2020), for identifying and characterizing FTD even in its early stages.

Automated speech analysis has been used to train and test classification models that are capable of differentiating individuals with FTD from healthy controls (HC), or between the subtypes of FTD, with accuracies of 80–90%, supporting the rationale that digital language tracking and analysis can aid in dementia diagnosis (Cho et al, 2020, 2022; Cho, Nevler, Ash, et al, 2021; Fraser et al, 2014, 2015; Jarrold et al, 2014; Parjane et al, 2021; Zimmerer et al, 2020). Recent studies have begun to use automated analysis tools to profile the specific speech and language changes in FTD subtypes in more detail (Cho, Nevler, Ash, et al, 2021; Nevler et al, 2019; Nevler, Ash, McMillan, et al 2020) and have identified changes such as decreased speech output and lower lexical diversity, which are common across subtypes (Cho, Nevler, Ash, et al, 2021). An ongoing study analyzing digital content-free speech data, language, and tone may provide equally important insights for diagnosis and care management (Tonn et al, 2020).

Longitudinal studies in FTD, such as the ALLFTD study (https://www.allftd.org/) and the GENFI (https://www.genfi.org/), indicate that language assessments, including naming tests, fluency tests, and reading tests, are useful for identifying language impairments in individuals with FTD and for differentiating subtypes (Ash et al, 2019; Rohrer et al, 2013; Staffaroni et al, 2021). The application of automated speech analysis tools in longitudinal data sets has been limited in previous research, although one study indicated that these methods may be useful for characterizing longitudinal changes in speech output and pause rate in individuals with bvFTD (Nevler, Ash, Cho, et al, 2020).

Assessing speech and language abilities in individuals with FTD is, therefore, an ideal area for the development of novel tools for remote assessment and monitoring of FTD in support of clinical assessment (Babrak et al, 2019; Coran et al, 2019; Coravos et al, 2019; Gold et al, 2018). Speech assessments can be completed quickly at home, with little instruction, via a smartphone or tablet, and advances in signal processing and natural language processing algorithms have the potential to provide greater insight into speech and language abilities using objective, automated tools.

Current Study

Building on this research that provides a proof-of-concept of how automated speech analysis can be used to characterize FTD, in the present study, we collected speech assessments remotely from individuals with FTD over the course of 12 months. We used an app-based speech assessment that included brief tasks based on clinical standards, including picture description, reading, fluency, and naming. We analyzed the speech recordings using an automated pipeline that generates both acoustic and linguistic variables representing different aspects of speech and language patterns. Using these data, we identified variables that showed longitudinal change and created a novel speech-based composite score to measure language abilities in individuals with FTD.

METHOD

Participants

We recruited individuals via the Association for Frontotemporal Degeneration, FTD researchers, and FTD clinics in Canada and the United States. Individuals who expressed interest in our study were contacted by a member of the Winterlight research team to confirm study eligibility. Inclusion criteria included male or female individuals between the ages of 45 and 90 years who were native or fluent English speakers; had a physician-confirmed diagnosis of bvFTD and/or nonfluent/agrammatic PPA, semantic PPA, or logopenic PPA within the last 12 months; and consented to have their physician release clinical information about their FTD diagnosis.

Additional inclusion criteria were the ability to consent to the study or provide consent from a legally authorized representative, have an available and willing relative or study partner to help administer the speech assessment, and have WiFi access in their residence and a private area where testing could take place. Exclusion criteria included the occurrence of a concussion or traumatic brain injury in the past 12 months and significant uncorrected hearing or vision impairments.

A member of the research team contacted each individual's physician, who provided written verification of the individual's diagnosis and identification of FTD subtype. Because individuals were recruited remotely without a visit to a clinical site, we were unable to perform additional diagnostic assessments.

We collected data from March 2019 to September 2020, with up to 12 months of data collection for each individual. No concurrent control individuals were recruited to participate in the study. However, healthy individuals who had been assessed using the same speech assessment in an independent study that was conducted by the same researchers were included as a comparison group. The HC consisted of community-dwelling older adults who had volunteered to take part in a normative speech assessment study that used the same assessment and stimuli at baseline.

The study protocol was reviewed and approved by the Advarra independent ethics review board, and all individuals and their study partners provided informed consent to participate.

Procedure

The study was a remotely conducted, observational, longitudinal study with a 12-month observation period. An individual's enrollment was confirmed upon acquisition of informed consent, release of medical information, and a completed clinical information form confirming a diagnosis of either bvFTD and/or PPA (all variants) by the individual's physician.

Depending on the individual's level of function, in most cases, a study partner was recruited to help administer and supervise the assessment. Upon enrollment, the study partner was sent a preconfigured, locked iPad. The individuals and their study partners were provided with training videos and written assessment instructions that described how to conduct the assessment at home.

After each individual and his or her study partner reviewed the material, they completed the first assessment with phone supervision and feedback from a member of the research team. Study partners were instructed that they could provide prompting and clarification to the individual during completion of the tasks, if necessary. If there were any difficulties in the assessment, such as technical issues using the app or clarifications needed to the instructions, the study partner was able to receive clarification from the team member and repeat the baseline assessment, if necessary. We also provided the individuals and their study partners with the email and phone contact information of the Winterlight technical support team for the duration of the study. Individuals were able to use the iPad for personal use and could keep the device at the end of the study as an incentive for participation.

The study consisted of periodic digital speech assessments that were administered via the Winterlight app and with study partner assistance, if needed. The individuals completed the speech assessment at a total of six time points: at baseline, 2 months, 3 months, 6 months, 9 months, and 12 months. No other clinical assessments were conducted.

Speech Assessment

The Winterlight Assessment (WLA) was developed to record and analyze speech via an app on a smartphone or tablet; it has been used in previous normative studies and clinical trials (Balagopalan et al, 2019; Robin et al, 2021; Simpson et al, 2019). The WLA is an investigational product that is used only for research purposes. It has not yet been approved by the US Food and Drug Administration.

The WLA consists of a series of speech tasks in which individuals are prompted to produce speech; this speech is recorded through the smartphone's or tablet's microphone. In the present study, the WLA included six speech tasks that took an average of 10–15 minutes to complete. All of the tasks were conducted in English. The included speech tasks are based on neuropsychological speech and language assessments, including picture description, paragraph reading and recall, phonemic fluency, semantic fluency, and object naming. The stimuli for each task varied across time points, but the assessment always contained all of the tasks in the same order. The individuals were encouraged to complete the assessment in one session but were allowed to take breaks if needed. Once an assessment was started, the speech task completion rate was very high, with 97% of all tasks being completed. The focus of the present research is on the picture description task.

Picture Description Task

In the picture description task, the instruction, "Please tell me everything you see in this picture" is presented visually and auditorily; then, a static line drawing of a scene is displayed on the screen (Figure 1). Individuals are expected to describe the picture in their own words, with no time limit, and are recorded. After the recording is completed, a second prompt asks, "Can you tell me anything else about this picture?" Tasks of this type have been shown to be good proxies for spontaneous discourse and are used in standard aphasia assessments (Borod et al, 1980; Giles et al, 1996).

The WLA includes six unique and proprietary pictures that were developed specifically for the WLA. They were designed to match the Cookie Theft picture (from the Boston Diagnostic Aphasia Examination; Goodglass et al, 2001) in lexico-syntactic complexity and amount of information content but depict more modern and varied scenes. Every picture depicts a black and white line drawing of a scene with a similar number of characters, salient actions, and objects.

At each time point in our study, the individuals were required to complete two picture descriptions using different picture stimuli. The picture stimuli were alternated across assessments to reduce any practice effects. Thus, two different stimuli were used at each of the assessments at months 1, 2, and 3, and then each set was repeated in the same order for the assessments at months 6, 9, and 12.

Because this task most closely approximates natural, unstructured speech, the focus of this paper is on speech analysis of the picture description task. There was no time limit for this task, allowing the individuals to speak naturally and to end the recording when they felt they had described the picture in as much detail as possible.

Data Analysis

Speech Variable Extraction

We used the Winterlight Labs speech analysis platform (Robin et al, 2021; Yeung et al, 2021) to analyze each individual's speech recordings. The speech recordings were

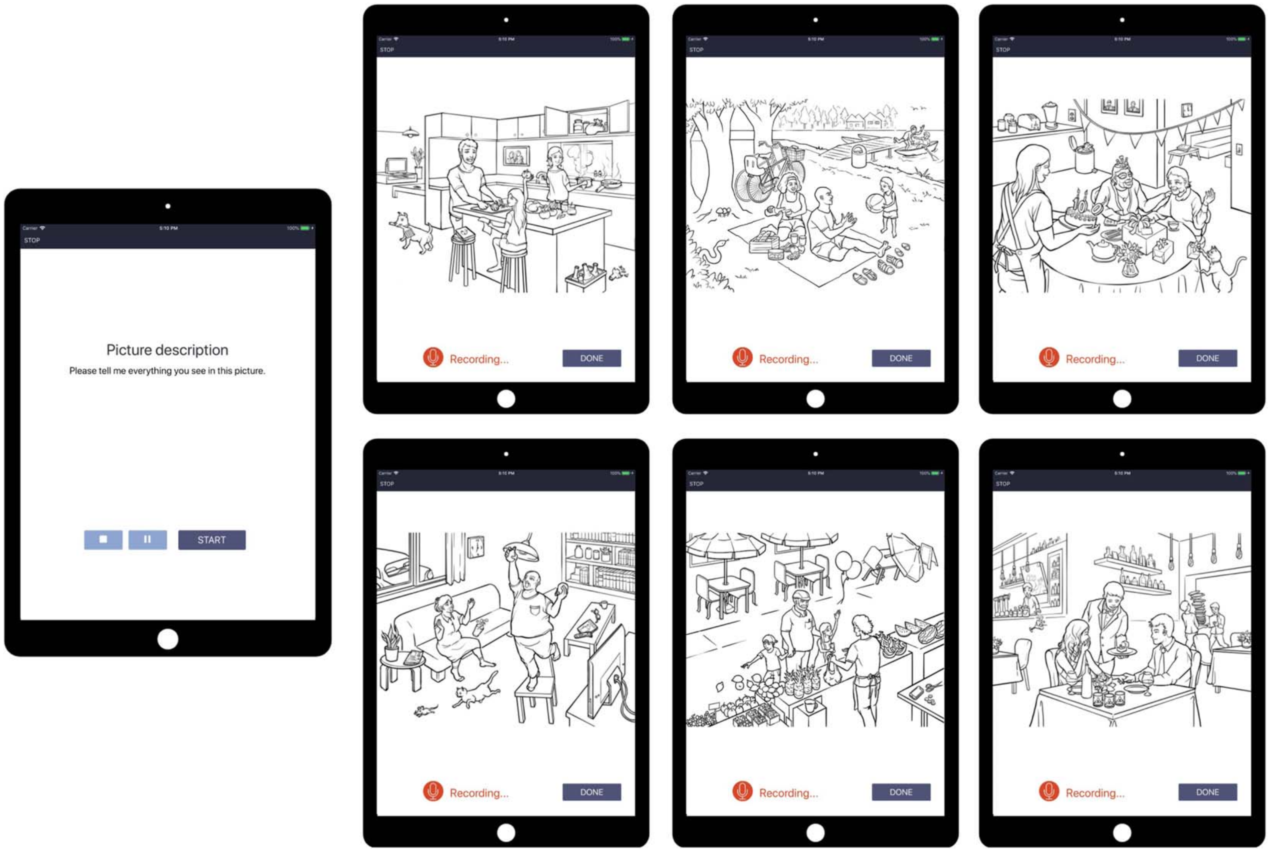

FIGURE 1. Schematic of the picture description task, part of the Winterlight Assessment, including all picture stimuli that were used in the study. Each assessment included two picture description tasks, using two distinct pictures. Picture stimuli are copyright of Winterlight Labs, Inc. and are used with permission for informational and descriptive purposes only. Reproduction of images for research or commercial use is prohibited.

first processed through a manual transcription platform. The generation of annotations such as speaker segmentation, transcription, and utterance segmentation was performed by trained raters using a standard protocol and custom software. An utterance is defined as an idea or unit of thought that can be equivalent to a phrase or a complete sentence (eg, *I saw a cat* is one utterance, whereas *I went into the house/because it was raining* is two utterances, with the boundary indicated by the slash). Raters listened to each recording and generated a text transcript of each individual's speech, including annotations for filled pauses (eg, *um, uh*), unfilled (silent) pauses, and unidentifiable words. Any speech from other speakers was identified and was removed from all of the recordings and transcripts.

The transcripts and preprocessed sound files were then used to generate speech variables using the Winterlight Labs pipeline (www.winterlightlabs.com), which uses Python-based acoustic and natural language processing libraries and custom code to compute >500 speech variables based on each speech recording and its accompanying transcript. An overview of all variables by category is provided in Table S1 of the supplementary digital content (SDC; http://links.lww.com/CBN/A132).

Next, each transcript was aligned to the audio recording, and both were used to produce speech variables, based on the audio recording, the transcript, or both. These variables reflect the acoustic (eg, properties of the sound wave, speech rate, number of pauses), lexical (eg, rates and types of words used, as well as their characteristics, such as frequency and imageability, which reflect how commonly words are used and how easy they are to picture, respectively), semantic (relating to the meaning of the words; eg, semantic relatedness of subsequent utterances, semantic relatedness of utterances to the items in the picture), and syntactic (relating to the grammar of the sentences; eg, syntactic complexity, use of different syntactic constructions) aspects of the speech sample.

Open source packages used to extract variables include SpaCy for parts-of-speech tagging and morphological variables, the Stanford NLP parser for syntactic variables (Chen and Manning, 2014), Praat and Parselmouth for acoustic variables (Jadoul et al, 2018), and GloVe and FastText models for semantic variables (Bojanowski et al, 2016; Pennington et al, 2014). The pipeline also uses custom code to compute additional variables based on the transcript and audio file using lexical norms from previous publications (Brysbaert and New, 2009; Kuperman et al, 2012; Stadthagen-Gonzalez and Davis, 2006; Warriner et al, 2013) or previously published models and variables (Mota et al, 2012). Custom variables for each speech task are also computed (eg, computing the number of correct items named in each picture for the picture description task).

Statistical Analysis

We used R Statistical Software version 4.1.1, with R packages tidyverse 1.3.1 (Wickham et al, 2019) for data cleaning and processing, lmerTest 3.1-3 (Kuznetsova et al, 2017) for linear mixed models, irr 0.84.1 for intraclass correlation (ICC) tests, and ggplot2 3.3.5 for visualizations for the statistical analyses.

Variable Identification

To identify the speech variables in the picture description task that demonstrated significant decline over time, we employed an exploratory, data-driven approach to test all of the extracted speech variables for significant linear changes over the time period of the observational study in the patient population. We first averaged the values for each variable across the two pictures at each time point and then standardized each speech variable with reference to the group mean and standard deviation.

Using linear mixed models with the fixed effects of time (in months), age at enrollment, sex, stimulus version (which varied across the time points), and random intercepts to account for baseline individual variations, we tested each variable for significant effects of time ($P < 0.05$), indicating significant change over the course of the study. Due to the exploratory nature of this step, we did not perform multiple comparison correction. Based on previous and internal unpublished research, we chose to control for the effects of age, sex, and stimulus version in the linear model because variations in these factors have been shown to affect speech and language patterns (Cho, Nevler, Shellikeri, et al, 2021). Due to incomplete data on years of education and time since diagnosis, we did not add these factors to the models.

Variable Verification

To further refine the set of speech variables, we selected those variables that showed consistent scores across the first three time points. To do so, we tested the ICC between the variable scores at months 1, 2, and 3. Variables that had significant ICCs >0.5 ($P < 0.05$) between the first three time points (which occurred within 2 months of one another) were selected as having moderate or higher test–retest reliability (Koo and Li, 2016). From this set of speech variables, we eliminated redundant variables based on whether ≥2 of the variables had very high correlations ($r > 0.9$) and measured similar constructs to one another. In cases where two variables were correlated ($r > 0.9$), we selected the variable that had a greater effect of time in the linear model and removed the other variable.

Composite Score Development

To combine the selected set of variables, we assigned equal positive or negative weights based on the direction of the time effect and summed the standardized variables to create a single composite score. To validate this score, we used linear mixed models to test for the effect of change over time and tested the test–retest reliability over the first three time points using the ICC.

Comparison With HC Group

We processed the speech data from the HC in the same way as the speech data from the FTD group. The values for each of the selected variables were averaged across the two pictures at the baseline assessment and were standardized with reference to the group means and standard deviations that we calculated for the FTD group, so that values were comparable across groups on the same scale. The composite score was computed in the identical way for the HC. The demographics of the HC were compared to the demographics of the FTD group using a two-sample t test for age and a χ^2 test for sex.

Comparison With Speech Task Scores

We developed custom variables for each speech task (ie, picture description, paragraph reading and recall, phonemic fluency, semantic fluency, and object naming) in order to provide an accuracy score for each task in the speech assessment. For picture description, scoring involves calculating the ratio of how many items in the picture were correctly named out of all of the items in each picture. For paragraph reading, scoring involves calculating the word error rate by dividing the number of deviations from the paragraph text by the total number of expected words. For paragraph recall, scoring involves calculating a recall score by assessing the proportion of story details that were remembered correctly. For phonemic and semantic fluency, scoring involves counting how many words were correctly generated in 60 seconds based on the prompt (ie, first letter for phonemic fluency, semantic category for semantic fluency). For object naming, scoring involves calculating a naming score by assessing the proportion of correctly named images out of the three.

All of these scores were automatically computed based on the speech transcripts using custom code. Correlations between the standardized speech composite score and the standardized speech task scores were computed at baseline using Pearson correlations.

RESULTS

Participants

We enrolled 36 individuals (16 female, 20 male) in the study. Enrollment began in December 2018 and continued until December 2019. The mean age at enrollment was 61.3 years (SD = 8.7, range = 45–81). For the individuals who provided information on years of education, the mean years of education was 16.1 years (SD = 2.1, range = 12–20; 7 individuals did not provide education data). Based on the clinical information forms that we obtained from the individuals' physicians, individuals had been diagnosed an average of 1.1 years before starting the study (SD = 1.2, range = 0–5 years; 3 individuals did not disclose time of diagnosis).

Twenty-one individuals (58.3%) were identified by their physician report as having bvFTD, and nine (25%) were identified as having a variant of PPA. Four of the cases of PPA were identified as having logopenic PPA, three as having semantic PPA, and two with the variant not specified by their physician. The other six individuals were identified as having both bvFTD and PPA, three of whom were identified as having semantic PPA, one as having nonfluent PPA, and two with no variant specified. For the purposes of our study, we labeled these individuals as *mixed* diagnosis. A summary of the number of individuals per FTD subtype is provided in Table 1.

All of the individuals and the study partners were able to operate the technology and understand how to complete the assessment; none of the individuals needed to repeat the assessment. Twenty-seven individuals (75%) completed the month 12 assessment. Of the nine individuals who did not complete the study, one individual did not complete any assessments after month 1 and did not respond to further contact attempts; three individuals withdrew due to severe decline in function and/or no longer being able to participate; three individuals were unable to complete the study due to factors relating to the COVID-19 pandemic, including not being able to have visitors in care homes in cases in which a care partner helped administer the assessment; and two individuals were late enrollments in the study (beginning the study in 2020) and therefore were unable to complete all of the assessments at the time of study termination.

The HC baseline comparison group included 41 individuals (25 female, 16 male, M_{age} = 62 years, range = 50–81 years) with a similar age range to the FTD group. All of the HC reported no neurologic or psychiatric diagnoses and completed the Montreal Cognitive Assessment (Nasreddine et al, 2005) screening test at baseline with a score ≥26 points, consistent with the threshold for no cognitive impairment.

Variable Selection

Based on linear mixed models including the factors of time, age at enrollment, sex, and stimulus version, 23 speech variables from the picture description task showed nominally significant ($P < 0.05$) effects of time at the group level. Of these 23 variables, 10 had moderate or higher ICCs (ICC > 0.5, $P < 0.05$) between the values at months 1, 2, and 3, suggesting good test–retest reliability.

After removing the redundant and very highly correlated ($r > 0.9$) speech variables, we identified eight speech variables (Table 2) as being sensitive to change over time and reliable across multiple assessments. (The variables with significant effects of time that were subsequently excluded from the final composite score are provided in SDC Table S2; http://links.lww.com/CBN/A132.) The eight variables were all reflective of the linguistic and timing properties of speech, including total words, unfilled pauses, noun frequency, use of noun phrases and prepositional phrases, a measure of information content, and measures relating to the graphical structure of the language produced.

Total words and *unfilled pauses* represent the proportion of speech to silence in the recordings. *Noun frequency* refers to how common, or frequently used, the nouns in the transcript were, on average. Two syntactic variables related to the use of *noun and prepositional phrases*, representing more complex sentence structures that contained multiple phrases and described relationships between items.

The *information units (subjects)* score, a custom computed score that counts the number of subjects (people or characters) that are accurately described for each picture, based on a list of key words for each picture stimulus, provides a measure of information richness and completeness of description. Finally, the graph measures are derived from using graph theoretical measures to map the co-occurrence patterns between successive words in the individual's description, and the *diameter* and *density* relate to the organization and repetition found in the participant's language (Mota et al, 2012, 2017).

TABLE 1. Number of Participants by FTD Subtype who Completed Each Assessment

	Assessment					
FTD Subtype	**Baseline (Month 1)**	**Month 2**	**Month 3**	**Month 6**	**Month 9**	**Month 12**
bvFTD	21 7 F/14 M M_{age} = 59.2 years	20	21	21	16	14
lvPPA	4 3 F/1 M M_{age} = 57.0 years	3	3	3	3	2
svPPA	3 1 F/2 M M_{age} = 69.7 years	3	3	3	3	3
Mixed/other	8 5 F/3 M M_{age} = 65.9 years	8	7	6	7	8

Mixed/other diagnosis includes PPA (unspecified variant) and bvFTD + PPA.
bvFTD = behavioral variant frontotemporal dementia. **lvPPA** = logopenic variant primary progressive aphasia. **svPPA** = semantic variant primary progressive aphasia.

TABLE 2. Effects of Time and Test–Retest Reliability for Selected Speech Variables Showing Significant Effects of Time in the Picture Description Task, Based on Linear Mixed Models, and the Resulting Composite Score

Speech Variable Name	Speech Variable Description	Change From Baseline to 12-month Endpoint M (SD)	Effect Time, Controlling for Age, Sex, and Stimulus (β)	ICC Between Months 1, 2, and 3
Total words	Ratio of the number of words to total annotations (eg, words, pauses, non-words) in the transcript	–0.181 (0.396)	–0.012, $P = 0.03$	0.65, $P < 0.001$
Unfilled pauses	Ratio of the number of silent pauses to total annotations (eg, words, pauses, non-words) in the transcript	0.203 (0.435)	0.015, $P = 0.01$	0.66, $P < 0.001$
Noun frequency	Average frequency score for all nouns in the transcript, representing the estimated frequency of usage of nouns; a measure of vocabulary complexity	0.200 (0.458)	0.013, $P = 0.03$	0.66, $P < 0.001$
Use of noun phrases containing a noun phrase and a prepositional phrase	Ratio of this syntactic construction to all syntactic constructions in the transcript; a measure of syntactic complexity	–0.060 (0.762)	–0.024, $P = 0.02$	0.50, $P < 0.001$
Use of prepositional phrases containing a preposition and a noun phrase	Ratio of this syntactic construction to all syntactic constructions in the transcript; a measure of syntactic complexity	–0.171 (0.593)	–0.017, $P = 0.03$	0.64, $P < 0.001$
Information units (subjects)	Ratio of distinct subjects described in the picture to the total number of subjects contained in the picture; a measure of information content	–0.713 (0.725)	–0.019, $P = 0.04$	0.55, $P < 0.001$
Graph diameter (undirected)	Length of the longest shortest path between all node pairs in the undirected graph representation of the transcript; a measure of language organization/complexity	0.115 (0.584)	0.019, $P = 0.01$	0.58, $P < 0.001$
Graph density	Number of edges in the graph representation of the transcript divided by the number of nodes squared; a measure of language organization/complexity	0.013 (0.111)	0.005, $P = 0.03$	0.54, $P < 0.001$
FTD speech composite score	Linear combination of the selected variable above, designed to measure progressive changes in language using a single score	–0.306 (0.449)	–0.026, $P < 0.001$	0.79, $P < 0.001$

FTD = frontotemporal dementia. **ICC** = intraclass correlation.

Composite Score Development

Based on the direction of the time effect, we assigned equal positive or negative weights to the selected variables, which were already standardized to be on the same, unit-free scale. Then, we summed the eight variables in order to create a single composite score that is designed to measure change in language patterns over time in individuals with FTD. The effect size of the change over time effect for the FTD speech composite score was numerically larger than the effect size of any of the single selected variables from the picture description task (Table 2). The composite score had a high ICC between the first three assessments, again higher than that of any single selected speech variable in the previous analyses. As shown in Figure 2, on average, the composite score followed a decreasing trajectory over the 12 months of the study despite some variability at the 2- and 3-month time points and in individual performance.

Comparison With HC

Because our study did not include a concurrent control group, we compared HC who had been assessed using the same speech assessment in an independent study to the FTD group at baseline. The HC group did not differ significantly from the FTD group based on age (t = 0.27, df = 70, P = 0.79) or sex (χ^2 = 1.49, df = 1, P = 0.22).

As shown in Figure 3, on average, the FTD group had lower composite scores than the HC at baseline (FTD group: M_{score} = 0.13, SD = 0.51; HC group: M_{score} = 0.67, SD = 0.29), representing increased speech impairment in the FTD group. A two-sample t test comparing the FTD and HC groups indicated a statistically significant group difference (t = 5.59, df = 55, P < 0.001). There was also more variation in the FTD group, with some of the lowest scores for individuals with logopenic PPA, and some of the individuals with bvFTD achieving scores that were consistent with more mild impairments.

Comparison With Speech Task Scores

To compare the novel FTD speech composite score with scores from the speech tasks in the WLA, we calculated correlations at baseline between the FTD speech composite score and the automatically computed scores for the picture description, paragraph reading, paragraph recall, phonemic fluency, semantic fluency, and object naming tasks (Figure 4). As described in the Method section, the standard scores are automated ways to score accuracy on each task (eg, number of words correct). The FTD speech composite score was positively correlated with all of the speech task scores, with the strongest correlation with the picture description task score (r = 0.73, df = 34, P < 0.001), followed by the object naming task score (r = 0.60, df = 34, P < 0.001), fluency task scores (phonemic fluency: r = 0.58, df = 34, P < 0.001; semantic fluency: r = 0.57, df = 34, P < 0.001), and paragraph recall task score (r = 0.51, df = 34, P = 0.002). Only the correlation between the FTD speech composite score and the paragraph reading score (r = 0.27, df = 34, P = 0.11) did not reach significance.

DISCUSSION

We demonstrated the feasibility of using a remote digital speech assessment to evaluate speech and language patterns in individuals with subtypes of FTD over the course of 12 months. Remote monitoring using digital tools provides new possibilities for clinical research, allowing assessment at higher frequency, with lower patient burden, without requiring in-person visits or lengthy clinical testing, which is of particular importance due to the debilitating nature of FTD (Coravos et al, 2019; Kourtis et al, 2019). Owing to its remote design and the use of a study partner, our study was able to maintain high study adherence and an individual retention rate of 75% during the COVID-19 pandemic while in-hospital assessments were limited or unavailable.

Using a data-driven analysis approach, we identified aspects of speech that showed significant changes over the course of the 12-month study, based on recordings of a naturalistic picture description task. Overall, we found that a number of characteristics of the picture descriptions changed over the course of the study period in the FTD group, including the amount of speech, the types of words and sentences used, and the content of the speech itself. These changes are consistent with previous research reporting shorter, simpler forms of speech being used by individuals with FTD and other dementias, with impaired

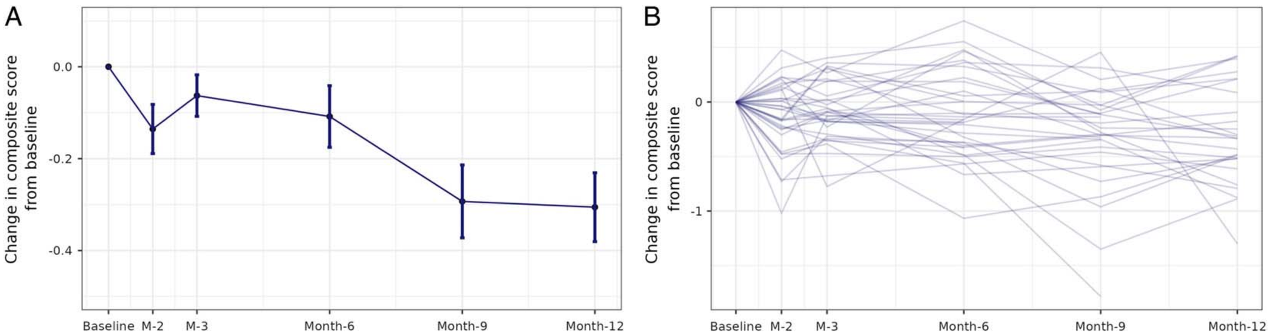

FIGURE 2. Change over time for the standardized FTD speech composite score. **A.** Group average change relative to baseline, with error bars showing the standard error of the mean change. **B.** Individual performance on the composite score for each individual in the study. Units represent the change from baseline. **FTD** = frontotemporal dementia.

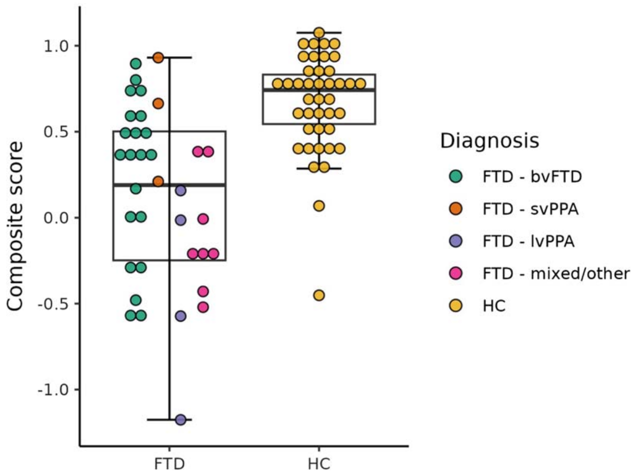

FIGURE 3. Baseline scores for the FTD speech composite score, based on the picture description task. For reference, FTD scores were compared with the scores from a group of HC with a similar age distribution to the FTD group. Dots represent each individual's score, color-coded by FTD subtype or HC, and the box plot indicates the median score and distribution (25% and 75% quantile) for the FTD and HC groups. Units are based on standardizing the component variables with reference to the mean and variation in the FTD group, with lower scores indicating increased language impairment. **bvFTD** = behavioral variant frontotemporal dementia. **FTD** = frontotemporal dementia. **HC** = healthy controls. **lvPPA** = logopenic variant primary progressive aphasia. **svPPA** = semantic variant primary progressive aphasia.

content and organization (Ash et al, 2006; Cho et al, 2020; Cho, Nevler, Ash, et al, 2021; Fraser et al, 2014; Gunawardena et al, 2010; Hardy et al, 2016; Nevler et al, 2017; Nevler, Ash, McMillan, et al, 2020; Poole et al, 2017).

We combined the variables that were identified as being most sensitive to change over time to form an exploratory composite measure to assess speech and language changes in individuals with FTD. The FTD speech composite score therefore captures multiple aspects of speech changes in individuals with FTD, including changes to the amount of speech and to the types and content of language used. The FTD speech composite score was sensitive to changes over 12 months in this heterogeneous group of individuals with FTD, had good test–retest reliability, and correlated with scores on other speech tasks. When compared with a group of HC, the FTD group had lower speech composite scores at baseline, suggesting that their speech impairments were present and detectable at study enrollment.

Although this composite score requires further clinical validation, our study serves as a first step in developing a novel digital measure for assessing longitudinal changes in speech and language symptoms in individuals with FTD, which could allow for more frequent, remote, low patient burden, and sensitive measurement of disease progression or improvement in individuals with FTD. As such, this novel digital measure illustrates how digital speech measures may be used to assess language ability across FTD subtypes and to complement current assessments by providing a naturalistic speech assessment that can be completed at home with a study partner.

Interpretation of Speech Changes

In order to interpret the aspects of speech that are captured by the FTD speech composite score, we discuss the component variables in turn. Over the duration of the study, the mean total number of words decreased and the mean number of unfilled (silent) pauses increased. These variables indicate that over 12 months, individuals with

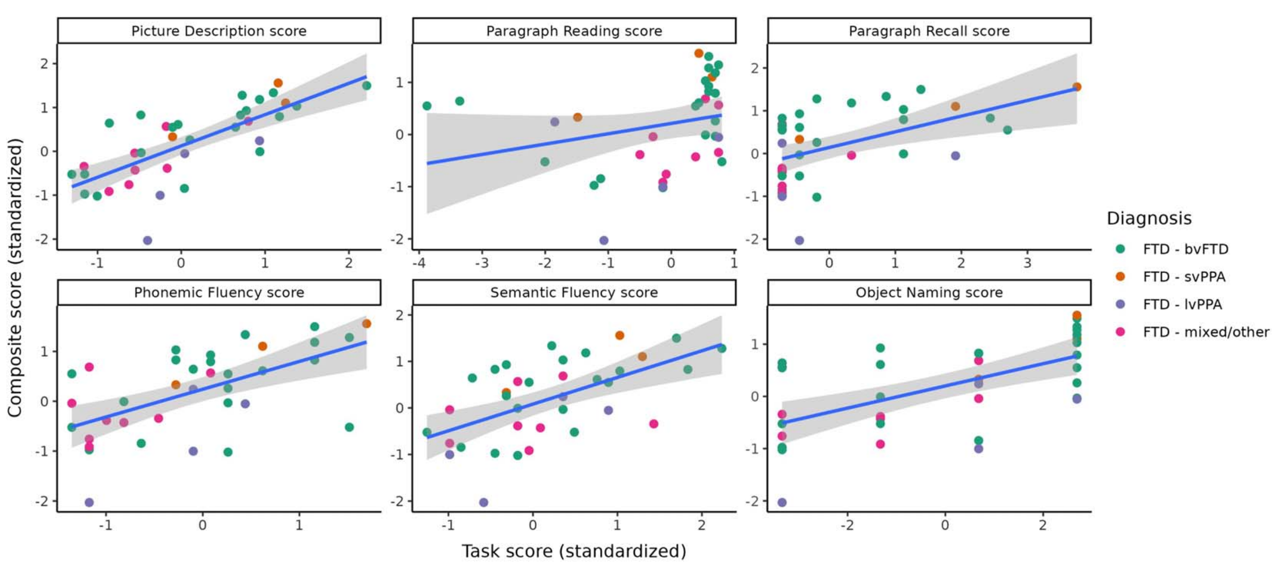

FIGURE 4. Correlations between baseline scores on the speech composite and baseline accuracy scores on the speech tasks included in the assessment. Scores have been standardized for comparison purposes, with higher scores indicating better performance. **bvFTD** = behavioral variant frontotemporal dementia. **FTD** = frontotemporal dementia. **lvPPA** = logopenic variant primary progressive aphasia. **svPPA** = semantic variant primary progressive aphasia.

FTD produced shorter descriptions of the pictures in the picture description task and paused more frequently during their descriptions, thereby replicating the findings from previous research reporting reduced speech output and increased pausing in FTD subtypes (including bvFTD) and progression of this pattern over time (Ash et al, 2019; Cho et al, 2020; Hardy et al, 2016; Nevler et al, 2017, 2019; Nevler, Ash, Cho, et al, 2020; Poole et al, 2017; Saxon et al, 2017; Yunusova et al, 2016).

A number of variables representing the content and structure of language were also found to change over time in the FTD group. The average frequency of nouns increased over the course of the study, indicating that more common nouns were employed as the disease progressed, which may signify declining vocabularies. In addition, the use of noun phrases including a prepositional phrase and the use of prepositional phrases (eg, *the boy in the kitchen, on the counter*) both declined over the course of the study. The decline in the use of these sentence structures may represent an impairment in the ability to form and describe relationships between entities in the picture. These changes in vocabulary and syntax use are consistent with previous research in both FTD and Alzheimer disease, in which higher frequency words and simpler syntaxes were used, resulting in simpler, less precise language (Ash et al, 2019; Cho, Nevler, Ash, et al, 2021; Cousins et al, 2016; Fraser et al, 2014, 2015; Hardy et al, 2016).

Other variables relating to the content of speech included an information unit score and graph metrics. The information unit subject score counted how many of the subjects in the picture were referred to, out of all subjects depicted. This score declined over the course of the study, indicating that the FTD group may have failed to notice the people in the picture or failed to comply with task instructions as the study progressed (Hardy et al, 2016).

We observed increases over the course of the study in graph metrics of discourse reflecting the diameter and density of the speech graph. In the context of the picture description task, increases in these scores may reflect individuals having reduced organization and increased repetition in their language output over the course of the study (Ash et al, 2006).

Study Limitations

Although the present results provide a proof-of-concept that longitudinal changes in speech and language patterns can be measured remotely in individuals with FTD using a digital speech assessment, there are several limitations to this work that will require further research and replication. First, due to the remote recruitment and assessment, we were limited in the depth and breadth of clinical information that we were able to collect from the individuals. As such, we were unable to verify FTD diagnoses and subtypes with in-depth clinical investigations and had to rely on the information provided by each individual's physician. In addition, because we had incomplete data on years of education and symptom duration, we did not add these factors to our models. We also were unable to collect other clinician-administered cognitive assessments that often serve as primary end points in clinical trials, biomarker data, or genetic information, which would have helped to better characterize the individuals and their disease progression using established methods.

Furthermore, the sample size was small, and the majority of the individuals were identified as the bvFTD subtype, which is typically characterized as having more minor language impairments compared with other subtypes (Hardy et al, 2016; Poole et al, 2017; Saxon et al, 2017), although other studies have indicated that language symptoms become more similar across subtypes with disease progression (Ash et al, 2019; Blair et al, 2007; Ranasinghe et al, 2016; Rogalski et al, 2011). Because of this imbalance in the subtypes, we were unable to identify different speech variables that could differentially track disease progression in individuals with different FTD subtypes. Nevertheless, by identifying speech variables that were affected across the FTD group, we hope to have developed a more general measure that is sensitive to FTD-related language impairments and is suitable for use across FTD subtypes.

Although we followed individuals over a 12-month period, we did not have comparable longitudinal data from a matched control group, and not all of the individuals were able to complete all of the assessments. Attrition may have been exacerbated by limitations relating to the COVID-19 pandemic, such as the inability for study partners to visit individuals living in care homes. Study dropout due to increased disease severity may have additionally biased the sample toward individuals who were in the earlier stage of the disease or with less impairment. Future studies with larger samples, more balanced subgroups, more in-depth clinical characterization, a well-matched control group, and a longer follow-up period will improve and expand on the current findings.

Limitations of our analytic approach include that this study employed an exploratory, data-driven method to identify speech variables that showed significant change over the course of the study, without correction for multiple comparisons. This method risks overfitting to the data and identifying false positive findings (Type I errors), thus requiring future replication studies to confirm the reproducibility and clinical significance of these findings in independent samples.

Both variation in the stimuli across assessments and the remote nature of the assessment may have contributed noise to the data, as evidenced by the variability in scores from the baseline assessment, which was supervised by the research team over the phone, and the month 2 assessment, which was not (Figure 2). Further work is needed to compare the results of remote and in-clinic assessments, though we note that the possible increased frequency of remote assessment may help to balance the increased variability that was introduced by at-home testing.

CONCLUSION

We conducted a pilot study of a remote digital speech assessment that was developed based on structured speech tasks to monitor speech and language changes in individuals with FTD. We demonstrated that remote tracking of speech and language abilities is feasible in the context of rapidly progressing dementia, with the use of tablet-based assessments and basic training of study partners to administer the assessments. Longitudinal assessment over a 12-month period suggested that automatically computed speech variables are reliable across repeated testing and may be sensitive to detecting disease progression.

Future work is needed to further develop speech-based tools for remote monitoring and to replicate and validate these findings in larger, more extensive clinically characterized samples. Larger cohorts will allow us to more clearly parse the differences between FTD subtypes and relate speech variables to standard clinical assessments and genetic markers. With the ubiquity of smartphones/tablets and high-quality microphones, and the prevalence of speech and language changes in other neurodegenerative disorders including Alzheimer disease, Parkinson disease, and amyotrophic lateral sclerosis, remote digital speech assessments have the potential to make disease detection and monitoring more accessible and less burdensome for clinicians and researchers.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the participants, their family members, and/or the study partners who participated in this study. The authors also thank Valerie Closson (Winterlight Labs) for her assistance with data collection.

REFERENCES

Ash S, Moore P, Antani S, et al. 2006. Trying to tell a tale: discourse impairments in progressive aphasia and frontotemporal dementia. *Neurology*. 66:1405–1413. doi:10.1212/01.wnl.0000210435.72614.38

Ash S, Nevler N, Phillips J, et al. 2019. A longitudinal study of speech production in primary progressive aphasia and behavioral variant frontotemporal dementia. *Brain Lang*. 194:46–57. doi:10.1016/j.bandl.2019.04.006

Babrak L, Menetski JP, Rebhan M, et al. 2019. Traditional and digital biomarkers: two worlds apart? *Digit Biomark*. 3:92–102. doi:10.1159/000502000

Balagopalan A, Yancheva M, Novikova J, et al. 2019. S44. Using acoustic and linguistic markers from spontaneous speech to predict scores on the Montreal Cognitive Assessment (MoCA). *Biol Psychiatry*. 85:S313. doi:10.1016/j.biopsych.2019.03.795

Blair M, Marczinski CA, Davis-Faroque N, et al. 2007. A longitudinal study of language decline in Alzheimer's disease and frontotemporal dementia. *J Int Neuropsychol Soc*. 13:237–245. doi:10.1017/S1355617707070269

Bojanowski P, Grave E, Joulin A, et al. 2016. Enriching word vectors with subword information. *TACL*. 5:135–146. doi:10.1162/tacl_a_00051

Borod JC, Goodglass H, Kaplan E. 1980. Normative data on the Boston Diagnostic Aphasia Examination, parietal lobe battery, and the Boston Naming Test. *J Clin Exp Neuropsychol*. 2:209–215. doi:10.1080/01688638008403793

Brysbaert M, New B. 2009. Moving beyond Kučera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav Res Methods*. 41:977–990. doi:10.3758/BRM.41.4.977

Chen D, Manning C 2014. A fast and accurate dependency parser using neural networks. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 740–750. doi:10.3115/v1/D14-1082

Cho S, Cousins KAQ, Shellikeri S, et al. 2022. Lexical and acoustic speech features relating to Alzheimer disease pathology. *Neurology*. 99:e313–e322. doi:10.1212/WNL.0000000000200581

Cho S, Nevler N, Ash S, et al. 2021. Automated analysis of lexical features in frontotemporal degeneration. *Cortex*. 137:215–231. doi:10.1016/j.cortex.2021.01.012

Cho S, Nevler N, Shellikeri S, et al 2020. Automatic classification of primary progressive aphasia patients using lexical and acoustic features. Paper presented at: *Language Resources and Evaluation Conference (LREC) 2020 Workshop on Resources and Processing of Linguistic, Para-Linguistic and Extra-Linguistic Data from People with Various Forms of Cognitive, Psychiatric and/or Developmental Impairments (RaPID-3)*; May 11–16, 2020. Accessed November 23, 2022. https://www.ldc.upenn.edu/sites/www.ldc.upenn.edu/files/lrec2020-automatic-classification-aphasia.pdf.

Cho S, Nevler N, Shellikeri S, et al. 2021. Lexical and acoustic characteristics of young and older healthy adults. *J Speech Lang Hear Res*. 64:302–314. doi:10.1044/2020_JSLHR-19-00384

Convery R, Mead S, Rohrer JD. 2019. Review: clinical, genetic and neuroimaging features of frontotemporal dementia. *Neuropathol Appl Neurobiol*. 45:6–18. doi:10.1111/nan.12535

Coran P, Goldsack JC, Grandinetti CA, et al. 2019. Advancing the use of mobile technologies in clinical trials: recommendations from the Clinical Trials Transformation Initiative. *Digit Biomark*. 3:145–154. doi:10.1159/000503957

Coravos A, Khozin S, Mandl KD. 2019. Developing and adopting safe and effective digital biomarkers to improve patient outcomes. *NPJ Digi Med*. 2:14. doi:10.1038/s41746-019-0090-4

Cousins KAQ, York C, Bauer L, et al. 2016. Cognitive and anatomic double dissociation in the representation of concrete and abstract words in semantic variant and behavioral variant frontotemporal degeneration. *Neuropsychologia*. 84:244–251. doi:10.1016/j.neuropsychologia.2016.02.025

Etcheverry L, Seidel B, Grande M, et al. 2012. The time course of neurolinguistic and neuropsychological symptoms in three cases of logopenic primary progressive aphasia. *Neuropsychologia*. 50:1708–1718. doi:10.1016/j.neuropsychologia.2012.03.028

Fraser KC, Meltzer JA, Graham NL, et al. 2014. Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. *Cortex*. 55:43–60. doi:10.1016/j.cortex.2012.12.006

Fraser KC, Meltzer JA, Rudzicz F. 2015. Linguistic features identify Alzheimer's disease in narrative speech. *J Alzheimers Dis*. 49:407–422. doi:10.3233/JAD-150520

Geraudie A, Battista P, García AM, et al. 2021. Speech and language impairments in behavioral variant frontotemporal dementia: a systematic review. *Neurosci Biobehav Rev*. 131:1076–1095. doi:10.1016/j.neubiorev.2021.10.015

Giles E, Patterson K, Hodges JR. 1996. Performance on the Boston Cookie theft picture description task in patients with early dementia of the Alzheimer's type: missing information. *Aphasiology*. 10:395–408. doi:10.1080/02687039608248419

Gold M, Amatniek J, Carrillo MC, et al. 2018. Digital technologies as biomarkers, clinical outcomes assessment, and recruitment tools in Alzheimer's disease clinical trials. *Alzheimers Dement*. 4:234–242. doi:10.1016/j.trci.2018.04.003

Goodglass H, Kaplan E, Barresi B. 2001. *Boston Diagnostic Aphasia Examination*, 3rd ed. Philadelphia, Pennsylvania: Lippincott Williams and Wilkins.

Gorno-Tempini ML, Hillis AE, Weintraub S, et al. 2011. Classification of primary progressive aphasia and its variants. *Neurology*. 76:1006–1014. doi:10.1212/WNL.0b013e31821103e6

Gunawardena D, Ash S, McMillan C, et al. 2010. Why are patients with progressive nonfluent aphasia nonfluent? *Neurology*. 75:588–594. doi:10.1212/WNL.0b013e3181ed9c7d

Hardy CJD, Buckley AH, Downey LE, et al. 2016. The language profile of behavioral variant frontotemporal dementia. *J Alzheimer's Dis*. 50:359–371. doi:10.3233/JAD-150806

Henry ML, Gorno-Tempini ML. 2010. The logopenic variant of primary progressive aphasia. *Curr Opin Neurol*. 23:633–637. doi:10.1097/WCO.0b013e32833fb93e

Jadoul Y, Thompson B, de Boer B. 2018. Introducing Parselmouth: a Python interface to Praat. *J Phon*. 71:1–15. doi:10.1016/j.wocn.2018.07.001

Jarrold W, Peintner B, Wilkins D, et al. 2014. Aided diagnosis of dementia type through computer-based analysis of spontaneous speech. *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, pages*. Baltimore, Maryland: Association for Computational Linguistics; 27–37. doi:10.3115/v1/W14-3204

Koo TK, Li MY. 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 15: 155–163. doi:10.1016/j.jcm.2016.02.012

Kourtis LC, Regele OB, Wright JM, et al. 2019. Digital biomarkers for Alzheimer's disease: the mobile/wearable devices opportunity. *NPJ Digit Med*. 2:9. doi:10.1038/s41746-019-0084-2

Kuperman V, Stadthagen-Gonzalez H, Brysbaert M. 2012. Age-of-acquisition ratings for 30,000 English words. *Behav Res Methods*. 44:978–990. doi:10.3758/s13428-012-0210-4

Kuznetsova A, Brockhoff PB, Christensen RHB. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *J Stat Softw*. 82:1–26. doi:10.18637/jss.v082.i13

Ljubenkov PA, Miller BL. 2016. A clinical guide to frontotemporal dementias. *Focus (Am Psychiatr Publ)*. 14:448–464. doi:10.1176/appi.focus.20160018

Luzzi S, Baldinelli S, Ranaldi V, et al. 2020. The neural bases of discourse semantic and pragmatic deficits in patients with frontotemporal dementia and Alzheimer's disease. *Cortex*. 128:174–191. doi:10.1016/j.cortex.2020.03.012

Miyagawa T, Brushaber D, Syrjanen J, et al. 2020. Use of the CDR® plus NACC FTLD in mild FTLD: data from the ARTFL/LEFFTDS Consortium. *Alzheimers Dement*. 16:79–90. doi:10.1016/j.jalz.2019.05.013

Montembeault M, Brambati SM, Gorno-Tempini ML, et al. 2018. Clinical, anatomical, and pathological features in the three variants of primary progressive aphasia: a review. *Front Neurol*. 9:692. doi:10.3389/fneur.2018.00692

Morris JC. 1993. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology*. 43:2412–2414. doi:10.1212/wnl.43.11.2412-a

Mota NB, Copelli M, Ribeiro S. 2017. Thought disorder measured as random speech structure classifies negative symptoms and schizophrenia diagnosis 6 months in advance. *NPJ Schizophr*. 3:18. doi:10.1038/s41537-017-0019-3

Mota NB, Vasconcelos NAP, Lemos N, et al. 2012. Speech graphs provide a quantitative measure of thought disorder in psychosis. *PLoS One*. 7:e34928. doi:10.1371/journal.pone.0034928

Nasreddine ZS, Phillips NA, Bédirian V, et al. 2005. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*. 53:695–699. doi:10.1111/j.1532-5415.2005.53221.x

Nevler N, Ash S, Cho S, et al. 2020. A longitudinal study of automated analysis of acoustic speech markers in FTD and PPA. *Alzheimers Dement*. 16:e045315. doi:10.1002/alz.045315

Nevler N, Ash S, Irwin DJ, et al. 2019. Validated automatic speech biomarkers in primary progressive aphasia. *Ann Clin Transl Neurol*. 6:4–14. doi:10.1002/acn3.653

Nevler N, Ash S, Jester C, et al. 2017. Automatic measurement of prosody in behavioral variant FTD. *Neurology*. 89:650–656. doi:10.1212/WNL.0000000000004236

Nevler N, Ash S, McMillan C, et al. 2020. Automated analysis of natural speech in amyotrophic lateral sclerosis spectrum disorders. *Neurology*. 95:e1629–e1639. doi:10.1212/WNL.0000000000010366

Parjane N, Cho S, Ash S, et al. 2021. Digital speech analysis in progressive supranuclear palsy and corticobasal syndromes. *J Alzheimers Dis*. 82:33–45. doi:10.3233/JAD-201132

Pennington J, Socher R, Manning C. 2014. GloVe: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages*. Doha, Qatar: Association for Computational Linguistics; 1532–1543. doi:10.3115/v1/D14-1162

Poole ML, Brodtmann A, Darby D, et al. 2017. Motor speech phenotypes of frontotemporal dementia, primary progressive aphasia, and progressive apraxia of speech. *J Speech Lang Hear Res*. 60: 897–911. doi:10.1044/2016_JSLHR-S-16-0140

Ranasinghe KG, Rankin KP, Lobach IV, et al. 2016. Cognition and neuropsychiatry in behavioral variant frontotemporal dementia by disease stage. *Neurology*. 86:600–610. doi:10.1212/WNL.0000000000002373

Robin J, Xu M, Kaufman LD, et al. 2021. Using digital speech assessments to detect early signs of cognitive impairment. *Front Digit Health*. 3:749758. doi:10.3389/fdgth.2021.749758

Rogalski E, Cobia D, Harrison TM, et al. 2011. Progression of language decline and cortical atrophy in subtypes of primary progressive aphasia. *Neurology*. 76:1804–1810. doi:10.1212/WNL.0b013e31821ccd3c

Rohrer JD, Caso F, Mahoney C, et al. 2013. Patterns of longitudinal brain atrophy in the logopenic variant of primary progressive aphasia. *Brain Lang*. 127:121–126. doi:10.1016/j.bandl.2012.12.008

Saxon JA, Thompson JC, Jones M, et al. 2017. Examining the language and behavioural profile in FTD and ALS-FTD. *J Neurol Neurosurg Psychiatry*. 88:675–680. doi:10.1136/jnnp-2017-315667

Simpson W, Kaufman LD, Detke M, et al. 2019. Utility of speech-based digital biomarkers for evaluating disease progression in clinical trials of Alzheimer's disease. *Alzheimers Dement*. 15:P1491–P1539. doi:10.1016/j.jalz.2019.08.089

Stadthagen-Gonzalez H, Davis CJ. 2006. The Bristol norms for age of acquisition, imageability, and familiarity. *Behav Res Methods*. 38: 598–605. doi:10.3758/BF03193891

Staffaroni AM, Weintraub S, Rascovsky K, et al. 2021. Uniform data set language measures for bvFTD and PPA diagnosis and monitoring. *Alzheimers Dement*. 13:312148. doi:10.1002/dad2.12148

Tonn P, Degani Y, Hershko S, et al. 2020. Development of a digital content-free speech analysis tool for the measurement of mental health and follow-up for mental disorders: protocol for a case-control study. *JMIR Res Protoc*. 9:e13852. doi:10.2196/13852

Warriner AB, Kuperman V, Brysbaert M. 2013. Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behav Res Methods*. 45:1191–1207. doi:10.3758/s13428-012-0314-x

Wickham H, Averick M, Bryan J, et al. 2019. Welcome to the tidyverse. *J Open Source Softw*. 4:1686. doi:10.21105/joss.01686

Yeung A, Iaboni A, Rochon E, et al. 2021. Correlating natural language processing and automated speech analysis with clinician assessment to quantify speech-language changes in mild cognitive impairment and Alzheimer's dementia. *Alzheimers Res Ther*. 13:109. doi:10.1186/s13195-021-00848-x

Yunusova Y, Graham NL, Shellikeri S, et al. 2016. Profiling speech and pausing in amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD. *PLoS One*. 11:e0147573. doi:10.1371/journal.pone.0147573

Zimmerer VC, Hardy CJD, Eastman J, et al. 2020. Automated profiling of spontaneous speech in primary progressive aphasia and behavioral-variant frontotemporal dementia: an approach based on usage-frequency. *Cortex*. 133:103–119. doi:10.1016/j.cortex.2020.08.027

