

Detecting Cognitive Impairment in FTD Spectrum Syndromes Using NLP-AI Analysis of Mobile Voice-Based Neuropsychological Testing

Bonnie Wong, PhD^{1,2}, Sam Murdock, BA¹, Erik Larsen, PhD³, Olivia Murton, PhD³, Lex Goldstein, BA¹, *Neguine Rezaii, MD¹, and *Bradford Dickerson, MD¹

¹Frontotemporal Disorders Unit, Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA;

²Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA; ³Sonde Health, Boston, MA, USA

*co-senior authors

BACKGROUND

- Frontotemporal dementias (FTDs) are heterogeneous neurodegenerative syndromes affecting language, behavior, and executive function. Early detection is challenging due to variable presentations and limited access to specialized evaluation.
- Digital cognitive assessments and NLP-AI tools can detect MCI and dementia^{1,2} and predict amyloid and tau accumulation and future decline^{3–5}, but their feasibility and accuracy across FTD spectrum disorders remain underexplored.
- This study evaluated the feasibility of a mobile app for voice-based neuropsychological testing and NLP-AI detection of cognitive impairment in a sample including FTD spectrum syndromes.

METHODS

- 34 patients with MCI or mild dementia (CDR ≤ 1) and 20 cognitively healthy controls completed a brief verbal neuropsychological battery on the Sonde One mobile app. 8 participants with un-scorable responses were excluded (analyzed: 27 patients, 19 controls).
- Patients were predominantly FTD spectrum syndromes (Figure 1). Given sample heterogeneity and small subgroup sizes, all patient subtypes were pooled for classification analyses.

Enrolled Patients by Clinical Syndrome (N = 34)

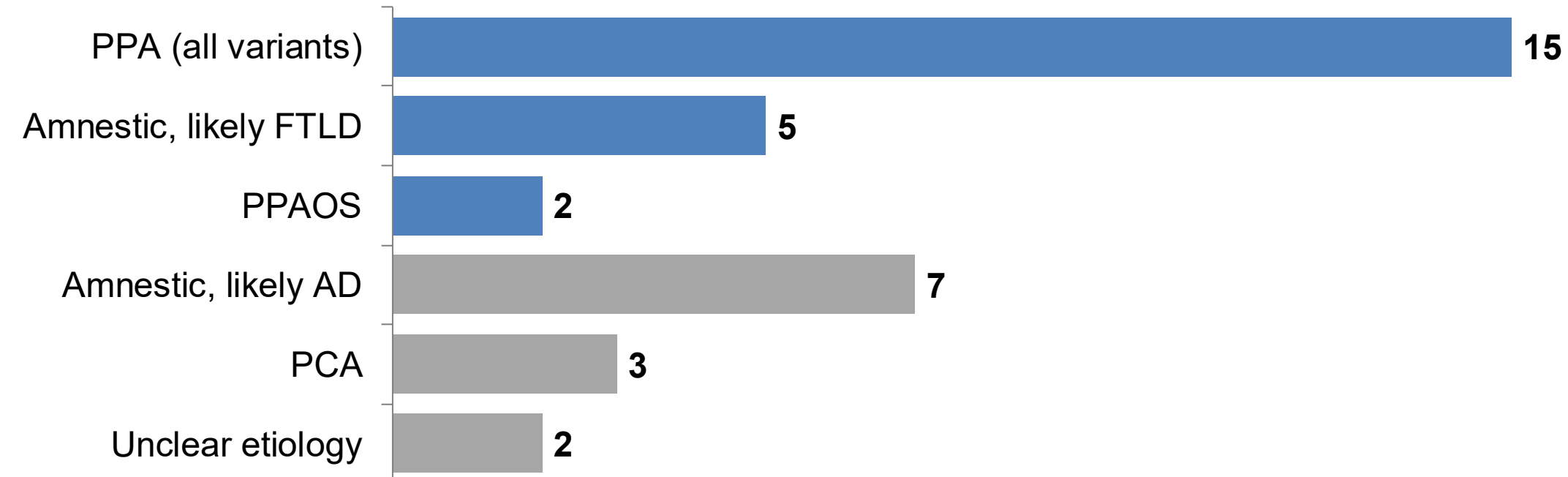
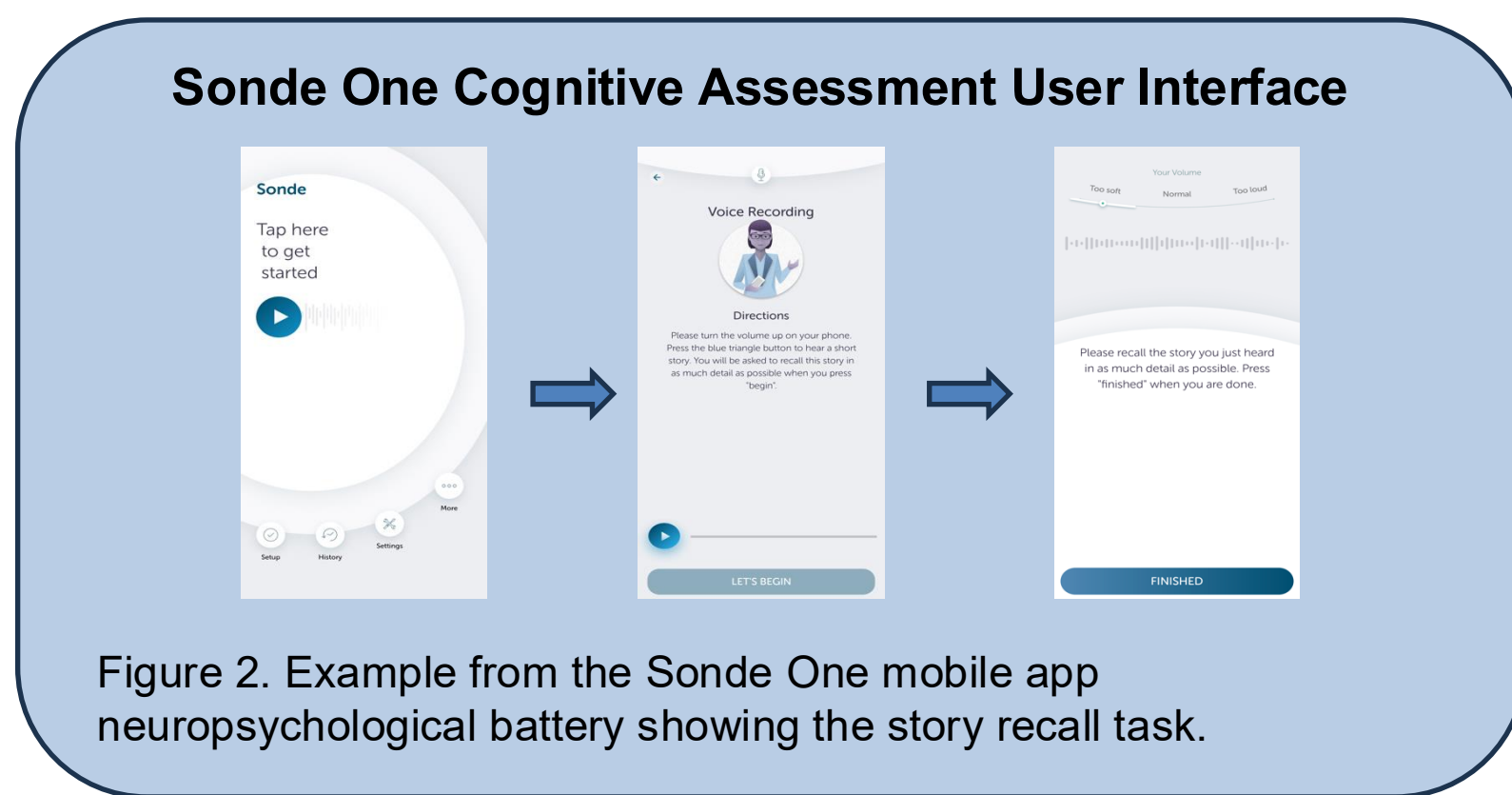


Figure 1. Blue: FTD spectrum syndromes (22 of 34, 65%); gray: other syndromes. Counts reflect enrollment, before 7 patients were excluded for un-scorable responses.

- Tests of language, memory, and executive function were completed at home, independently or with caregiver support (Figure 2).



- Baseline test scores were summed into a neuropsychological (NP) composite. A large language model-based algorithm computed the Language Informative Index (LII)^{6,7}, which measures the semantic similarity between immediate and delayed story recalls and the original story.
- ROC analysis assessed diagnostic accuracy. Post-hoc analyses related both measures to MoCA scores (N = 26, MoCA within 6 months).

RESULTS

Participant Demographics and Baseline Cognitive Performance

	Control (N=19)	Patient (N=27)
Age (years, Mean $\pm$ SD)	69.3 $\pm$ 4.5	66.3 $\pm$ 8.0
% female	58	41
Years of education		
12 - 15	n/a	3 (11%)
16 - 18	9 (47%)	13 (48%)
19 +	5 (26%)	4 (15%)
Unknown	5 (26%)	7 (26%)
% of mobile app tasks completed at baseline session (Mean $\pm$ SD)	93 $\pm$ 15	77 $\pm$ 26
Time to complete baseline session (mins, Mean $\pm$ SD)	11.0 $\pm$ 1.8	17.0 $\pm$ 9.3
Baseline NP composite score (Mean $\pm$ SD)	193.6 $\pm$ 23.1	100.0 $\pm$ 41.2
Story immediate recall (manual scoring, Mean $\pm$ SD)	20.6 $\pm$ 3.7	6.7 $\pm$ 5.3
Story delayed recall (manual scoring, Mean $\pm$ SD)	20.3 $\pm$ 4.3	6.1 $\pm$ 5.7
Story recall – LII (Immediate) (Mean $\pm$ SD, range 0 (least similar) - 1 (most similar))	0.93 $\pm$ 0.04	0.61 $\pm$ 0.25
Story recall – LII (Delayed) (Mean $\pm$ SD, range 0 (least similar) - 1 (most similar))	0.92 $\pm$ 0.06	0.57 $\pm$ 0.32
MoCA score (Mean $\pm$ SD) **	29.0 $\pm$ 1.0 (N = 5)	20.6 $\pm$ 4.5 (N = 21)

Table 1. The NP composite includes scores from the following tasks: category fluency, mental alternation (parts A and B), passage reading, serial subtraction, and story recall paraphrase scores (immediate and delayed). All participants completed at least five tasks. **Note: MoCA scores were available only for a subset of participants who completed the assessment through other research protocols.

ROC Analysis of Patient / Control Classification

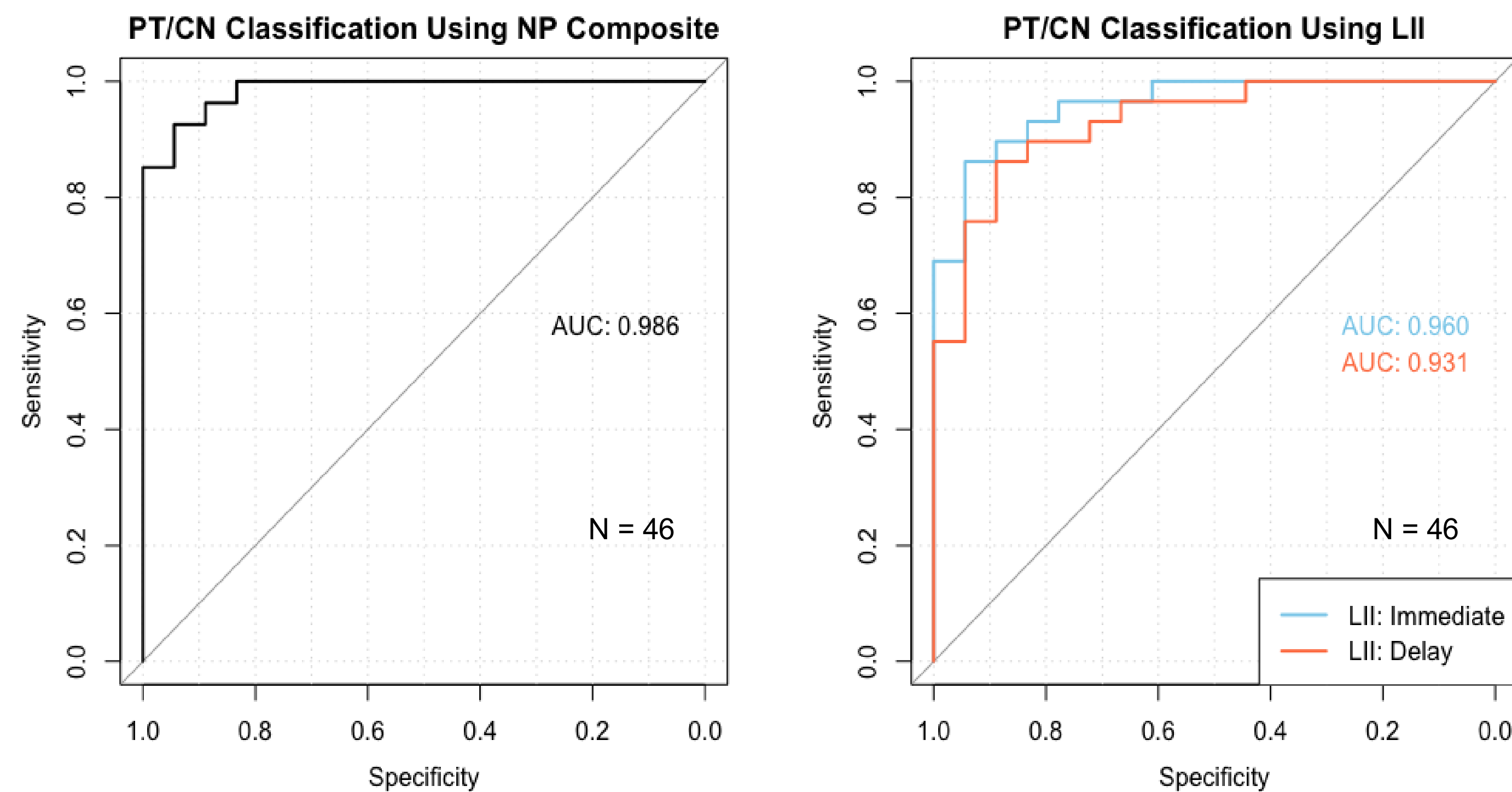


Figure 3. In the pooled patient sample vs. cognitively healthy controls, the NP composite (AUC = 0.99, 95% CI 0.96–1.00) and Story Recall-LII (Immediate: AUC = 0.96, 95% CI 0.91–1.00; Delayed: AUC = 0.93, 95% CI 0.86–1.00) classified participants with high accuracy.

RESULTS (CONT.)

Neuropsychological Composite and Story Recall-LII (Immediate) vs MoCA

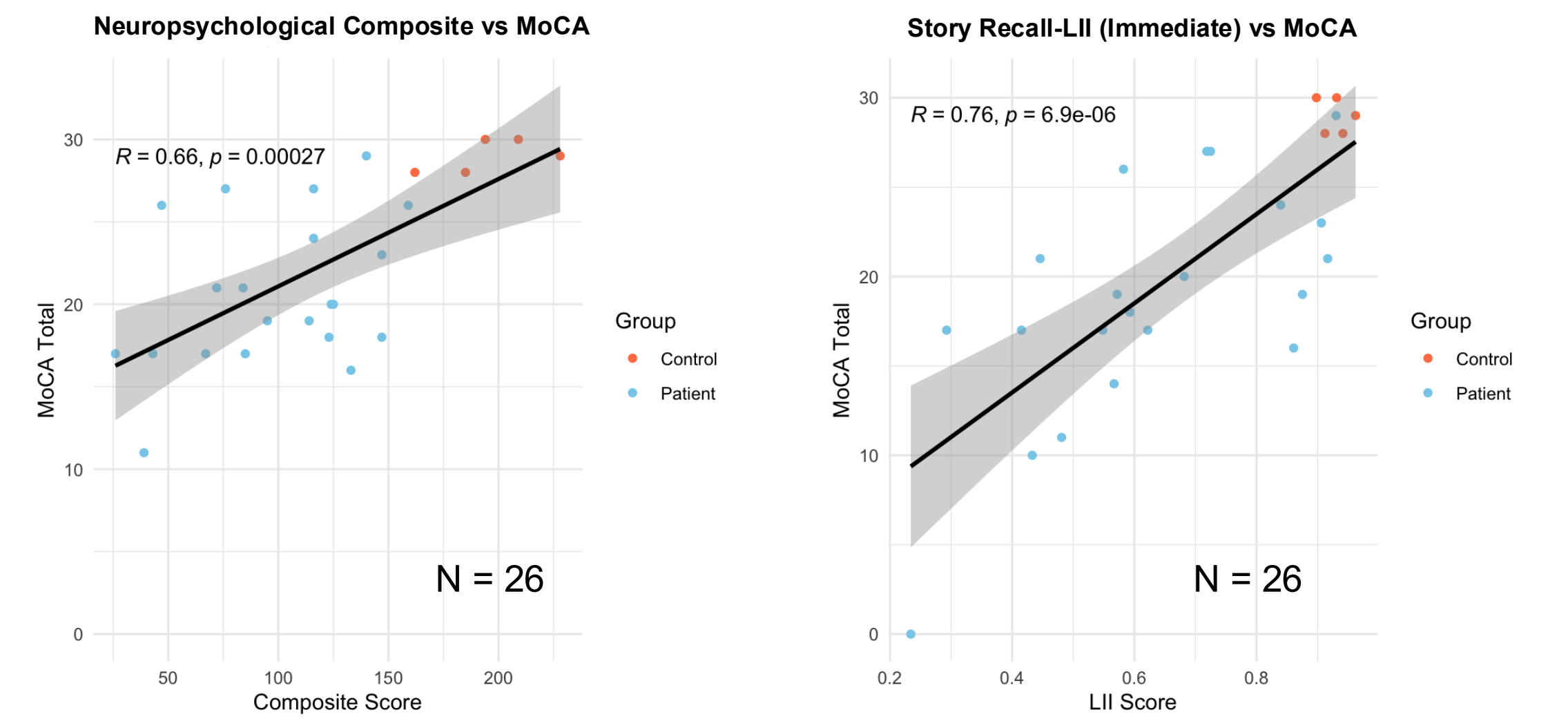


Figure 4. For a subset of participants who completed the Montreal Cognitive Assessment (MoCA) within six months of study enrollment, post-hoc analyses revealed a significant relationship ($p < 0.001$) between MoCA and both composite NP scores and Story-Recall LII scores.

CONCLUSIONS

- Remote, at-home mobile cognitive assessment was feasible in a sample predominantly composed of FTD spectrum syndromes, including patients with progressive aphasia and apraxia of speech.
- The NP composite and Story Recall-LII both distinguished patients from controls with high accuracy; LII-based analysis shows potential as an efficient, low-burden screening tool.
- Both measures tracked MoCA performance, with Story Recall-LII (Immediate) showing the stronger relationship.
- As sample sizes grow, analyses will examine whether NLP-AI approaches can differentiate among FTD syndromes and severity levels.

REFERENCES & ACKNOWLEDGEMENTS

¹Amini S, Hao B, Zhang L, Song M, Gupta A, Karjadi C, et al. Automated detection of mild cognitive impairment and dementia from voice recordings: A natural language processing approach. *Alzheimers Dement* 2023;19:946–55. <https://doi.org/10.1002/alz.12721>.

²Nyongesa CA, Hogarth M, Pa J. Artificial intelligence-driven natural language processing for identifying linguistic patterns in Alzheimer's disease and mild cognitive impairment: A study of lexical, syntactic, and cohesive features of speech through picture description tasks. *J Alzheimers Dis* 2025;106:120–38. <https://doi.org/10.1177/13872877251339756>.

³Balabin H, Tamm B, Spruyt L, Dusart N, Kabouche I, Eycken E, et al. Natural language processing-based classification of early Alzheimer's disease from connected speech. *Alzheimers Dement* 2025;21:e14530. <https://doi.org/10.1002/alz.14530>.

⁴Vanderlip, C. R., Stark, C. E. L., & Alzheimer's Disease Neuroimaging Initiative. (2024). Digital cognitive assessments as low-burden markers for predicting future cognitive decline and tau accumulation across the Alzheimer's spectrum. *Alzheimer's & Dementia*, 20(10), 6881–6895. <https://doi.org/10.1002/alz.14154>

⁵Amini S, Hao B, Yang J, Karjadi C, Kolachalama VB, Au R, et al. Prediction of Alzheimer's disease progression within 6 years using speech: A novel approach leveraging language models. *Alzheimers Dement* 2024;20:5262–70. <https://doi.org/10.1002/alz.13886>.

⁶Bayat S, Sanati M, Mohammad-Panahi M, Khodadadi A, Ghasimi M, Rezaee S, et al. Language abnormalities in Alzheimer's disease indicate reduced informativeness. *Ann Clin Transl Neurol* 2024;11:2946–57. <https://doi.org/10.1002/acn3.52205>.

⁷Rezaii N, Wong B, Aisen P, Beckett L, Dage JL, Eloyan A, et al. Voiceprints of cognitive impairment: analyzing digital voice for early detection of Alzheimer's and related dementias. *npj Dementia* 2025;1:35. <https://doi.org/10.1038/s4400-025-00040-0>.

Funding: The research reported in this poster was supported by an award from the Massachusetts AI and Technology Council for Connected Care in Aging & Alzheimer's Disease, Year 1 Pilot Projects. **Contact:** Bonnie.Wong@MGH.harvard.edu