Robustness and generalizability of a speech-based digital biomarker derived from recordings of the Clinical Dementia Rating (CDR) interview

Michael Spilka[1], Mengdan Xu[1], Bali Toth[2], Somaye Hashemifar[2], Rainier Amora[2], Jessica Robin[1], Edmond Teng[2], Cecilia Monteiro[2], & William Simpson[1]

[1]Winterlight Labs, Inc. (a division of Cambridge Cognition), Toronto, ON, Canada
[2]Genentech, Inc., South San Francisco, CA, USA

Contact: michael.spilka@camcog.com

Background

- Progressive language changes are established clinical characteristics of Alzheimer's disease (AD).
- Advances in Natural Language Processing (NLP) enable more objective, nuanced measurement of language, facilitating the development and validation of speech biomarkers for tracking longitudinal decline in language function.
- **Objective:** We examined the robustness and generalizability of our previously published speech-based digital biomarker score (Robin et al., 2023; doi: 10.1002/dad2.12445) in an independent clinical trial dataset.

Methods

- Participants were 165 English-speaking individuals with mild-to-moderate AD from the Lauriet Phase 2 trial of semorinemab (NCT03828747).
- CDR interview recordings were analyzed at screening, baseline, week 25, and week 49, focusing on participant speech from the "recent experience" section of the interview.
- Nine speech features (3 acoustic, 6 linguistic) were extracted and combined according to our previously published speech-based digital biomarker score (developed from a subset of patients with prodromal-to-mild AD from the Tauriel Phase 2 trial [NCT03289143]).
- Performance of the speech-based composite in the development dataset (Tauriel) and replication dataset (Lauriet) was compared on:
  - *Longitudinal trajectories* (using linear mixed effects models adjusting for participant age, gender, and level of education; the Lauriet analysis included the placebo group only [*n* = 81]).
  - *Baseline correlations with clinical endpoints* (performed on the full participant sample; screening and baseline visits were averaged, when available).
  - *Test-retest reliability* (using intraclass correlations of screening and baseline visit recordings).

Speech-based composite score features

Linguistic	Acoustic
Word length	MFCC 11 mean
Syntactic depth	MFCC 25 variance
Word frequency	MFCC 26 variance
Noun use	
Particle use	
Pronoun use	

Individual speech features were *z*-scored relative to the development dataset (using the *M* and *SD* across all data points) prior to being combined into the composite score.

Participant characteristics

- The speech-based composite confirmed the presence of more severe language dysfunction in the Lauriet vs. the development dataset, consistent with observed clinical endpoint differences.

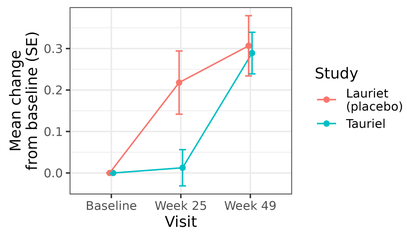
Baseline characteristics	Lauriet (replication sample)	Tauriel (development sample)	*p*-value
N	165	147	
Age (*M*, *SD*)	72.4 (8.3)	69.0 (7.8)	<.001
Sex (n, %)			.296
Female	106 (64%)	85 (52%)	
Male	59 (36%)	62 (42%)	
ADAS-Cog-11 Total (*M*, *SD*)	24.32 (6.43)	17.47 (5.74)	<.0001
ADAS-Cog-11 Language Composite (*M*, *SD*)	2.59 (2.46)	1.37 (1.46)	<.0001
CDR-SB (*M*, *SD*)	6.52 (1.84)	3.88 (1.43)	<.0001
ADCS-ADL Total (*M*, *SD*)	61.07 (8.70)	69.40 (6.00)	<.0001
MMSE Total (*M*, *SD*)	18.11 (2.03)	23.28 (2.74)	<.0001
Speech-based composite score	0.40 (0.46)	0.07 (0.45)	<.001

Note. MMSE = Mini Mental State Examination. ADAS-Cog-11 = Alzheimer's Disease Assessment Scale–Cognitive Subscale. CDR-SB = Clinical Dementia Rating Scale Sum of Boxes. ADCS-ADL = Alzheimer's Disease Cooperative Study – Activities of Daily Living Scale. The ADAS-Cog-11 Language Composite was the sum of the following five ADAS-Cog-11 scores: Spoken Language Ability (SLA), Word-Finding Difficulty (WFD), Comprehension of Spoken Language, Object and Finger Naming, Word Recognition.

Results: longitudinal trajectories

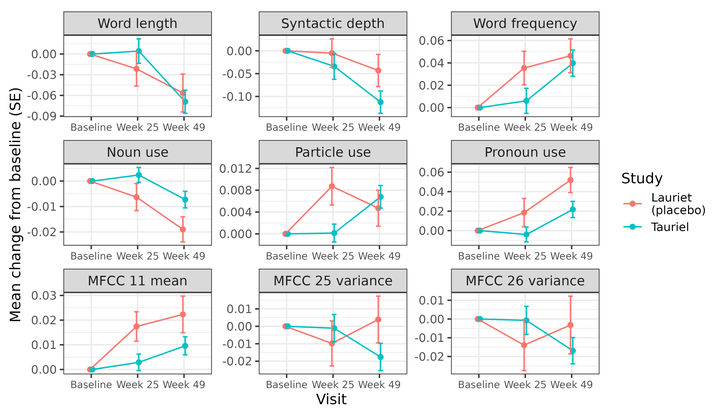
Composite score trajectories

- The biomarker showed a significant effect of change over time (β=0.26, *p*=.0003), in line with the development dataset (β=0.29, p<0.0001).



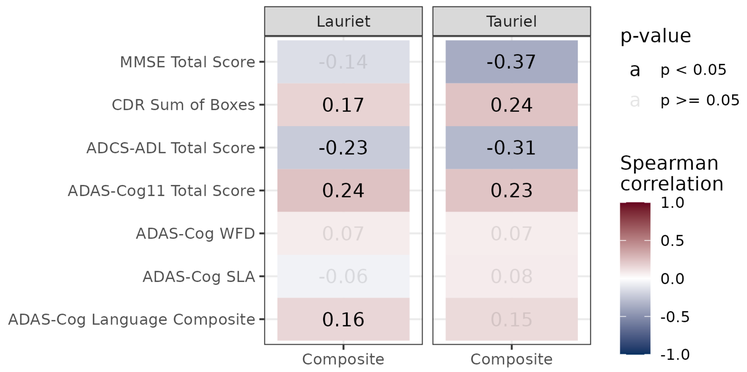
Individual feature trajectories

- 6 of 9 features had similar trajectories in both datasets, suggesting the generalizability of these features.



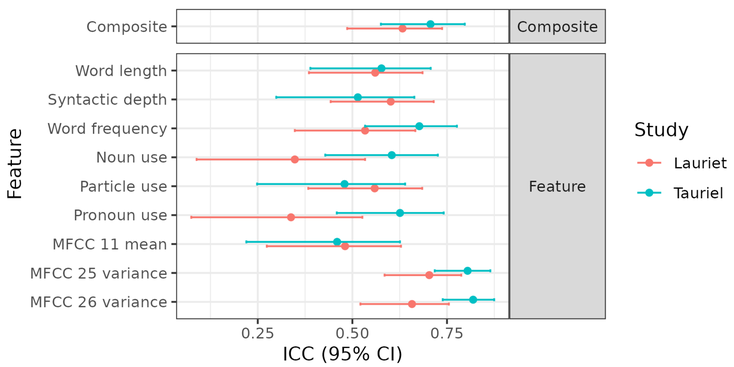
Results: correlations with clinical endpoints

- Baseline correlations with clinical endpoints were statistically significant and broadly similar between the two datasets for the CDR-SB, ADAS-Cog-11 Total, and ADCS-ADL scores.



Results: test-retest reliability

- Screening to baseline visit test-retest reliability of the composite score was similar between the two datasets (Lauriet ICC = 0.63; Tauriel ICC = 0.71). Test-retest reliability of individual features was more variable but generally comparable.



Conclusions

- We replicated prior speech-based digital biomarker score findings in an independent, more severe AD population, suggesting that the speech characteristics within this score are robust and aligned with clinical progression across disease stages.
- Additional validation work is ongoing, including the development of comparative biomarkers leveraging the combined data from both clinical trials.

SCAN HERE