



LJMU Research Online

Bruno, D, Jauregi Zinkunegi, A, Bock, JR and ADNI,

Predicting CDR status over 36 months with a recall-based digital cognitive biomarker

<http://researchonline.ljmu.ac.uk/id/eprint/23811/>

Article

Citation (please note it is advisable to refer to the publisher's version if you intend to cite from this work)

Bruno, D, Jauregi Zinkunegi, A, Bock, JR and ADNI, Predicting CDR status over 36 months with a recall-based digital cognitive biomarker. Alzheimer's & Dementia: The Journal of the Alzheimer's Association. ISSN 1552-5260 (Accepted)

LJMU has developed **LJMU Research Online** for users to access the research output of the University more effectively. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LJMU Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain.

The version presented here may differ from the published version or from the version of the record. Please see the repository URL above for details on accessing the published version and note that access may require a subscription.

For more information please contact researchonline@ljmu.ac.uk

<http://researchonline.ljmu.ac.uk/>

Title: Predicting CDR status over 36 months with a recall-based digital cognitive biomarker

Authors: Davide Bruno, PhD^a, Ainara Jauregi-Zinkunegi, PhD^a, Jason R. Bock, MA^{b*}, and The Alzheimer's Disease Neuroimaging Initiative[†]

[†]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

a: School of Psychology, Liverpool John Moores University
Tom Reilly building, Byrom St., Liverpool, L3 3AF, UK

b: Embic Corporation
3900 West Coast Highway, Suite 310, Newport Beach, CA, 92663, USA

Corresponding Author:

Jason R. Bock, M.A.
Embic Corporation
3900 West Coast Highway, Suite 310
Newport Beach, CA 92663, USA
Email: jrbock@gmail.com

Structured Abstract

INTRODUCTION: Word-list recall tests are routinely employed for cognitive assessment. Process scoring and latent modeling of these tests have shown potential to enhance accuracy without requiring test redesign. We examined whether a set of process-based digital cognitive biomarkers (DCBs) at baseline predicted Clinical Dementia Rating (CDR) longitudinally, and how they compared to standard metrics.

METHODS: Secondary data analyses were carried out with ADNI data. DCBs were derived from ADAS-Cog Word Recall test data. Participants were 330 (baseline mean age=71.4, SD=7.2). We conducted Bayesian and frequentist regression analyses using CDR at 36 months as outcome, controlled for demographics and genetic risk, and used ADAS-Cog traditional scores and DCBs as predictors.

RESULTS: Across tests, the best predictor of CDR score at 36 months appeared to be *M*, a DCB reflecting recall ability. This metric outperformed traditional ADAS-Cog scores. Diagnostic results suggest that *M* may be particularly useful to identify true negatives, i.e., individuals who are unlikely to decline over the follow-up period.

DISCUSSION: These results indicate that *M*, derived from word-list tests such as ADAS-Cog, may outperform some of the traditional metrics used for the purpose of identifying individuals who are at greater risk of disease progression. More research is needed, however, to determine these metrics' applicability and sensitivity with other word-list tests and across diverse populations.

1. Background

Word list recall tests are routinely employed in clinical practice for the assessment of verbal memory ability, including in individuals with Alzheimer's disease (AD) and other dementias [1]. Most commonly-employed neuropsychological tests of word-list recall were developed originally for the purposes of identifying individuals with well-defined memory loss predating dementia [2]. However, alongside the promise of disease-modifying drugs for AD, more emphasis is currently being put on identifying individuals who present *subtle* signs of underlying pathology at the earliest stages, as these subjects may benefit the most from pharmacological interventions. Therefore, together with advancements in neuroimaging and fluid biomarkers, there is also a critical need for development of more accurate neuropsychological testing, including with word lists, particularly as cognitive assessment typically is cost-effective and requires relatively little training, compared to neuroimaging and fluid biomarkers capture [3].

Process scoring and latent modeling of cognitive tests, which allow for the identification of underlying neurocognitive mechanisms of test performance, including with word list recall tests, have shown potential to enhance test accuracy without requiring test redesign [3-7]. For example, simple adjustments in the way popular word-list tests are scored have yielded enhanced sensitivity to cerebrospinal fluid biomarkers of AD [8], and post-mortem pathology [9].

One approach to latent modeling of word-list recall tests is with hidden Markov modeling (HMM), a class of established cognitive models [10,11]. In the context of word list recall tests, an HMM characterizes episodic memory for a list item as existing in one of a set of latent storage states upon each observation of recall or non-recall during the test. Retrieval parameters are associated with each storage state from which an item is capable of being

recalled, and encoding parameters describe the transitions among the storage states [12]. The sequence of encoding and retrieval transitions cannot be directly observed and are therefore “hidden” or “latent” parameters, quantifiable as probabilities that together produce the observed sequence of recall and non-recall.

One HMM that has been applied to word list recall tests is the hierarchical Bayesian cognitive processing (HBCP) model. The HBCP model posits that an item exists within one of three states upon each observation: Pre-task Storage (P), which contains only semantic memory of the non-novel word yet no episodic memory of its presentation during the word-list recall test; Transient Storage (T), which contains temporarily-stored episodic memory of the word presentation such that it can be retrieved on immediate free recall (IFR) tasks but not after a delay; and Durable Storage (D), which contains episodic memory of the word presentation that is capable of being retrieved on immediate or delayed free recall (DFR) tasks. Three retrieval parameters quantify the probability of recall from the latter two states: Transient Retrieval (R_1), from T on IFR tasks; Durable Retrieval (R_2), from D on IFR tasks; and Delayed Retrieval (R_3), from D on DFR tasks. Four encoding parameters quantify the probability of an item transitioning from one state to another during each task: One-shot Encoding (N_1), from P to D ; Transient Encoding (N_2), from P to T ; Consolidated Encoding (N_3), from T to D on a task subsequent to its encoding into T ; and Testing Effect Encoding (N_4), from T to D upon successful recall from T . In the HBCP, these parameters are employed as probabilities for each branch of a multinomial processing tree that reproduces the observed recall behavior [13]. For example, a particular word for a particular assessment that is not recalled on IFR 1 or 2 or DFR 1, but recalled on IFR 3, will express as increased probability of the encoding parameters that result in the word existing in the T or D state by IFR 3, and the retrieval parameters R_1 and R_2 , respectively. See Figure 1.

Subsequent to generating the encoding and retrieval parameters, measures of recall ability can be estimated by recombining specific subsets of the multinomial processing tree branches comprised of these parameters. Word recall through T (M_1) includes the probabilities for N_2 and R_1 ; word recall through D on IFR tasks (M_2) includes N_1, N_2, N_3, N_4 , and R_2 ; and word recall through D on DFR tasks (M_3) includes N_1, N_2, N_3, N_4 , and R_3 . In the above example of word recall on only IFR 3, M_1 is the summation of the subset of branches that results in retrieval through R_1 on IFR 3, including encoding through N_2 into T but excluding branches that encode through N_1, N_2, N_3 , and N_4 into D ; the reverse is true for the generation of M_2 via retrieval from D through R_2 . These memory (M) values represent generalized recall rate through the various processes across words and tests. For example, an individual with an M_3 of .71 is expected to recall 7.1 words out of 10 on average across delayed free recall tests. In unpublished findings, these M parameters have been used to distinguish individuals with mild cognitive impairment from those who were cognitively normal at time of assessment, with an AUC of .79. Collectively, the N , R , and M parameters are referred to as digital cognitive biomarkers (DCBs). In the present study, parameters are averaged together (e.g., average of M_1, M_2 , and $M_3 = M$) to further generalize recall ability across word-list recall test tasks.

In this paper, we aim to confirm previous, internal findings, by testing whether HBCP-derived DCBs (N , R , and M : indexing encoding, retrieval, and recall, respectively) are useful predictors of cognitive decline, as measured by the Clinical Dementia Rating (CDR) overall score. In particular, we want to ascertain whether baseline estimates of N , R , and M yield more predictive power than the traditional ADAS-Cog memory measures, such as immediate and delayed recall scores also derived from item response data.

2. Methods

2.1 Participants

Data were drawn from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, which has detailed methods [14] reported elsewhere [15,16]. The ADNI is a longitudinal study launched in 2003 with measures of cognitive impairment and AD including clinical and neuropsychological assessment. To be included in this secondary data analysis study, participants had to have: Alzheimer's Disease Assessment Scale–Cognitive subscale (ADAS-Cog) scores at baseline, from which traditional scores and DCBs were extracted; a CDR score at baseline and after a 36-month follow up; and a Polygenic Hazard Score (PHS) to determine genetic risk. The initial reference pool comprised of 3,418 participants, then reduced to 330 participants (mean age at baseline = 71.4, SD = 7.2), after applying the inclusion criteria above. Of these, 184 were males, and were 146 females. At baseline, 29 participants had a CDR score of 0; 300 scored 0.5; and one participant scored 1. At the 36-month follow-up visit, 64 participants had a CDR score of 0; 214 scored 0.5; 39 scored 1; 12 scored 2; and one participant scored 3. All activities for this study were approved by the ethics committees of the authors' universities and competed in accordance with the Declaration of Helsinki. All participants provided informed consent prior to testing.

2.2. Materials

The Alzheimer's Disease Assessment Scale–Cognitive subscale (ADAS-Cog) [17] is a neuropsychological test battery consisting of 13 subtests, including immediate and delayed word recall. In the word recall task, participants are visually and audibly presented a list of 10 words, then asked to recall as many words as possible over 3 tasks: the order of presentation is varied across tasks, and the score is calculated as the mean number of words *not* recalled across the three tasks. After a 10-minute delay with distraction, participants are once more asked to recall as many words as possible, this time without presentation, and the score is calculated as the number of words *not* recalled.

The Clinical Dementia Rating (CDR) scale [18] is a semi-structured clinical interview that assesses six cognitive areas: memory, orientation, judgement and problem-solving, community affairs, home and hobbies, and personal care. Each cognitive domain is then scored between 0 and 3, and to obtain a global score, the sum of each of the domains is calculated with equal weightings. The global CDR stages are 0, indicating normal cognition; 0.5, indicating mild cognitive impairment; and 1, 2, and 3, indicating mild, moderate, and severe dementia, respectively.

2.3. Digital Cognitive Biomarkers

Digital cognitive biomarkers scores were generated using ADAS-Cog Word Recall item response data with the HBCP model, using Bayesian inference with a Markov-chain Monte Carlo (MCMC) algorithm [13]. Each assessment's observed sequence of recall and non-recall was used to update prior information of DCB distributions for typical individuals in the general population who come from demographic groups (age, sex, and education level) specific to the participant who performed the assessment. The HBCP model additionally included adjustment for word presentation position effects on each of the three ADAS-Cog English word lists. These DCBs are proprietary (Embic Corporation).

2.4. Polygenic Hazard Score

The Desikan AD PHS was computed based on a Cox proportional hazard regression model combining 31 AD-associated single nucleotide polymorphisms (SNPs) with two *APOE* variants ($\epsilon 2/\epsilon 4$), trained with genetic data from an independent cohort. The PHS, composed of a weighted score of 33 risk- or protection-conferring SNPs, was calculated for each participant as previously described [19].

2.5. Analysis plan

First, we carried out a longitudinal Bayesian linear regression analysis with CDR score at 36 months as outcome. Predictors were N (as average of N_1, N_2, N_3, N_4), R (as average of R_1, R_2 and R_3), and M (as average of M_1, M_2 , and M_3), and the traditional ADAS-Cog memory scores (immediate and delayed recall), all measured at baseline, and control variables were years of education, gender, polygenic hazard score, age at baseline, and baseline CDR score. Credible intervals (CIs) were set to 95%. The prior was set to JZS, and the model prior was set to Uniform. One thousand MCMC simulations were conducted to determine parameters and compensate for possible violations of normality, but we also evaluated q-q plots of residuals to estimate normality. Following that, we carried out two sensitivity tests. We conducted a frequentist ordinal regression analysis with the same outcome and covariates but limited predictors to those that emerged from the initial Bayesian regression as strongest. Finally, to evaluate clinical validity of these predictive metrics, we conducted a frequentist logistic regression analysis: we used the increase in CDR score between baseline to 36 months of at least 0.5 as outcome and used the same predictor(s) and covariates (minus the baseline CDR score, already included in the change score) as in the ordinal analysis. Analyses were conducted using JASP (0.18.3) [20] and SPSS (29.0; IBM).

3. Results

Table 1 reports demographics, CDR, and memory scores in the cohort under examination. Results indicated that the CDR score at 36 months was best predicted by a model including only M (extreme evidence: $BF_{10} > 1$ billion, $BF_{inclusion} = 26.9$) — this model’s odds were about three times as high ($BF_M = 9.3$ vs. $BF_M = 3.2$) as the next best model, including M and R . In comparison to models including the traditional ADAS-Cog immediate and delayed recall scores, the model with M alone performed over 3.5 times better than the model with immediate recall and M combined ($BF_M = 9.2$ vs. $BF_M = 2.6$, respectively), and over 5 times

better than the model with delayed recall and M combined (whose BF_M was 1.7). The best model without DCBs (including both immediate and delayed recall) had a BF_M of 0.1: this finding indicates that the top model, with M alone, had model odds about 90 times greater than the best model with only traditional ADAS-Cog metrics.

The higher the M score was, the lower the CDR score at 36 months (mean coefficient = -2.38, SD = 0.79): a cross-sectional difference of 0.2 M points corresponds to a CDR difference of about 0.5 (credibility intervals: -4.52 to -1.05).

Given that the q-q plot for the analysis above displayed some degree of non-normality, we also carried out the same analysis on square-root transformed follow up CDR scores, which gave us more linear q-q plots. The overall pattern of results was unchanged.

The sensitivity ordinal (frequentist) regression confirmed the overall findings above. The model fit was significant ($\chi^2(6) = 122.8, p < .001$; Nagelkerke pseudo- $R^2 = 0.36$), as was M (coefficient estimate = -12.92, SE = 1.79, Wald coefficient = 52.32, $p < .001$).

Finally, the frequentist logistic regression (269 individuals did not increase their CDR score by 0.5 or more, and 61 did) showed that adding M to the model reduced the AIC from 293.95 to 245.24. M was a significant predictor in this analysis (unstandardized coefficient estimate = -14.94, standardized coefficient estimate = -1.23, Wald coefficient = 39.75, Odds Ratio < 0.001, $p < .001$). Setting M at approximately 0.58 yielded the following performance diagnostics: the area under the curve (AUC) was 0.84 (without M and only covariates the AUC was 0.71), negative predictive value was 0.87 (258 correct rejections vs. 38 misses), and positive predictive value was 0.68 (23 hits vs. 11 false alarm). Furthermore, specificity was very high (0.96; 258 correct rejections vs. 11 false alarms), while sensitivity was lower (0.38; 23 hits vs. 38 misses). Figure 2 displays the association between the baseline M score

and the probability of declining by at least 0.5 CDR points at follow-up (36 months), and Figure 3 reports the receiver operating characteristics (ROC) curve.

4. Discussion

In this analysis of ADNI data, we examined how HBCP-derived DCBs (N , R , and M , indexing encoding, retrieval, and recall, respectively) compared to the traditional ADAS-Cog assessment metrics (immediate and delayed recall scores) in predicting CDR score over a 36-month span. Our analysis included 330 individuals and showed that DCB M was the better overall predictor in the test. These findings are in line with recent efforts demonstrating the validity of process metrics for early detection of cognitive impairment [3-7,21-25].

One observation is that the M metric, comparably to other cognitive tools in recent literature [3,25,26], appears more useful to exclude false negatives than to identify targets correctly, as indexed by the high negative predictive value and specificity. In other words, individuals scoring at $M = 0.58$ or higher at baseline were unlikely to be declining after 36 months. While identification of positive cases appears more difficult with process scoring compared to, for example, fluid biomarkers [27,28], a high negative predictive value still yields great utility. Typically, cognitive assessment is cost- and resource-effective when compared to most biomarker assessments, as cognitive assessments are cheaper (many are non-proprietary), require less administrator training, and are less invasive for delivery. Therefore, especially in addressing global need where biomarker assessment is cost-restrictive, there is value in cognitive screening which may help exclude individuals who, despite possible subjective concerns, are unlikely to be on a disease trajectory. Future assessments of M and related DCBs should include direct comparisons to state-of-the-art fluid and imaging biomarkers.

Further research should also use the latest in process scoring and latent modeling. In the time since these analyses were performed, a second generation of DCBs was generated and included in the ADNI database as quantified cognitive processes (qCP). These qCPs account for additional differences in word features across the alternate lists of the ADAS-Cog Word Recall test and include alternate M parameters representative of recall on specific immediate and delayed tasks. Future analyses can be performed to evaluate the predictive capability of these qCPs.

This secondary and preliminary assessment of HBCP-derived DCBs has a definite limitation worth noting. The outcome (CDR score) is based upon clinical assessment of primarily cognitive function, and the predictors (N , R , M , and traditional immediate and delayed recall scores) are also measures of cognitive function, specifically memory, which risks issues of circularity. However, to note, 1) the memory scores are not contributors to the CDR score, and 2) M outperformed other measures of recall performance. In addition to measures of cognitive function, further confirmation of the utility of HBCP-derived DCBs should come from tests comparing this score to measures of pathology. Moreover, further studies should consider adding more demographic variety, such as including younger cohorts (as the present cohort was on average 70+ at baseline) and more ethnic diversity.

To conclude, word list memory tests are widely utilized for evaluating cognitive function, especially in Alzheimer's disease research and screening. These tests vary in their elements, such as list length, number of learning attempts, sequence of presentation across attempts, and inclusion of semantic categories. Traditionally, scoring techniques, such as overall scores and more recently composite scoring, have not adequately addressed differences among these elements or their impact on learning and memory during the test [29]. Recent advancements in process scoring and latent modeling offer promise in overcoming these limitations to

provide better ways to assess cognitive performance. In this study, we show a specific example in HBCP-derived DCBs, M .

References

- [1] De Simone MS, Perri R, Fadda L, et al. Different deficit patterns on word lists and short stories predict conversion to Alzheimer's disease in patients with amnesic mild cognitive impairment. *J Neurol*. 2017;264(11):2258-2267. doi:10.1007/s00415-017-8623-8
- [2] Mueller KD, Du L, Bruno D, et al. Item-level story recall predictors of amyloid-beta in late middle-aged adults at increased risk for Alzheimer's disease. *Front Psychol*. 2022;13:908651. doi:10.3389/fpsyg.2022.908651.
- [3] Bruno D, Jauregi Zinkunegi A, Pomara N, et al. Cross-sectional associations of CSF tau levels with Rey's AVLT: A recency ratio study. *Neuropsychology*. 2023;37(6):628-635. doi:10.1037/neu0000821
- [4] Bruno D, Reiss PT, Petkova E, Sidtis JJ, Pomara N. Decreased recall of primacy words predicts cognitive decline. *Arch Clin Neuropsychol*. 2013;28(2):95-103. doi:10.1093/arclin/acs116
- [5] Bruno D, Kosciak RL, Woodard JL, Pomara N, Johnson SC. The recency ratio as predictor of early MCI. *Int Psychogeriatr*. 2018;30:1883-1888. doi:10.1017/s1041610218000467
- [6] Mueller KD, Kosciak RL, Du L, et al. Proper names from story recall are associated with beta-amyloid in cognitively unimpaired adults at risk for Alzheimer's disease. *Cortex*. 2020;131:137-150. doi:10.1016/j.cortex.2020.07.008
- [7] Talamonti D, Kosciak R, Johnson S, Bruno D. Predicting early mild cognitive impairment with free recall: The primacy of primacy. *Arch Clin Neuropsychol*. 2020;35(2):133-142. doi:10.1093/arclin/acz013

- [8] Bruno D, Jauregi Zinkunegi A, Pomara N, et al. Cross-sectional associations of CSF tau levels with Rey's AVLT: A recency ratio study. *Neuropsychology*. 2023;37(6):628. doi:10.1037/neu0000821.
- [9] Bruno D, Gicas KM, Jauregi-Zinkunegi A, Mueller KD, Lamar M. Delayed primacy recall performance predicts post mortem Alzheimer's disease pathology from unimpaired ante mortem cognitive baseline. *Alzheimer's Dement*. 2024;16:e12524. doi:10.1002/dad2.12524
- [10] Greeno JG, Bjork RA. Mathematical Learning Theory and the new "mental forestry." *Annu Rev Psychol*. 1973;24:81-116. doi:10.1146/annurev.ps.24.020173.000501
- [11] Wickens TD. *Models for Behavior: Stochastic Processes in Psychology*. San Francisco, CA: Freeman and Company; 1982.
- [12] Alexander GE, Satalich TA, Shankle WR, Batchelder WH. A cognitive psychometric model for the psychodiagnostic assessment of memory-related deficits. *Psychol Assess*. 2016;28(3):279-293. doi:10.1037/pas0000163.
- [13] Lee MD, Bock JR, Cushman I, Shankle WR. An application of multinomial processing tree models and Bayesian methods to understanding memory impairment. *J Math Psychol*. 2020;95:102328. doi:10.1016/j.jmp.2020.102328.
- [14] <https://adni.loni.usc.edu> (last accessed April 2024)
- [15] Weiner MW, Veitch DP. Introduction to special issue: overview of Alzheimer's Disease Neuroimaging Initiative. *Alzheimers Dement*. 2015;11:730-733. doi:10.1016/j.jalz.2015.05.007

- [16] Weiner MW, Veitch DP, Aisen PS, et al. The Alzheimer's Disease Neuroimaging Initiative 3: Continued innovation for clinical trial improvement. *Alzheimers Dement*. 2017;13:561-571. doi:10.1016/j.jalz.2016.10.006
- [17] Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. *Am J Psychiatry*. 1984;141(11):1356-1364. doi:10.1176/ajp.141.11.1356
- [18] Morris JC. Clinical dementia rating. *Neurology*. 1994;44(10):1983. doi:10.1212/wnl.44.10.1983-b
- [19] Desikan RS, Fan CC, Wang Y, et al. Genetic assessment of age-associated Alzheimer disease risk: development and validation of a polygenic hazard score. *PLoS Med*. 2017;14:e1002258. doi:10.1371/journal.pmed.1002258
- [20] <https://jasp-stats.org/> (Last accessed April 2024)
- [21] Thomas KR, Edmonds EC, Eppig J, et al. Using neuropsychological process scores to identify subtle cognitive decline and predict progression to mild cognitive impairment. *J Alzheimers Dis*. 2018;64(1):195-204. doi:10.3233/JAD-180229.
- [22] Turchetta CS, De Simone MS, Perri R, et al. Forgetting rates on the recency portion of a word list predict conversion from mild cognitive impairment to Alzheimer's disease. *J Alzheimers Dis*. 2020;73(4):1295-1304. doi:10.3233/JAD-190509
- [23] Gicas KM, Honer WG, Wilson RS, et al. Association of serial position scores on memory tests and hippocampal-related neuropathologic outcomes. *Neurology*. 2020;95(24):e3303-3312. doi:10.1212/wnl.0000000000010952

- [24] Gicas KM, Honer WG, Leurgans SE, et al. Longitudinal change in serial position scores in older adults with entorhinal and hippocampal neuropathologies. *J Int Neuropsychol Soc.* 2023;29(6):561-571. doi:10.1017/s1355617722000595
- [25] Bruno D, Jauregi Zinkunegi A, Kollmorgen G, et al. A comparison of diagnostic performance of word-list and story recall tests for biomarker-determined Alzheimer's disease. *J Clin Exp Neuropsychol.* 2023;13:1-7. doi:10.1080/13803395.2023.2240060
- [26] Bruno D, Zinkunegi AJ, Kollmorgen G, et al. The recency ratio assessed by story recall is associated with cerebrospinal fluid levels of neurodegeneration biomarkers. *Cortex.* 2023;159:167-174. doi:10.1016/j.cortex.2022.12.004
- [27] Rissman RA, Langford O, Raman R, et al. Plasma A β 42/A β 40 and phospho-tau217 concentration ratios increase the accuracy of amyloid PET classification in preclinical Alzheimer's disease. *Alzheimers Dement.* 2024;20(2):1214-1224. doi:10.1002/alz.13542.
- [28] Barthélemy NR, Salvadó G, Schindler SE, et al. Highly accurate blood test for Alzheimer's disease is similar or superior to clinical cerebrospinal fluid tests. *Nat Med.* 2024;30:1085-1095. doi:10.1038/s41591-024-02869-z.
- [29] Bock JR, Russell J, Hara J, Fortier D. Optimizing cognitive assessment outcome measures for Alzheimer's disease by matching wordlist memory test features to scoring methodology. *Front Digit Health.* 2021;3:750549. doi:10.3389/fdgth.2021.750549.

Acknowledgments

We wish to thank all ADNI participants. ADNI data can be requested at:

<https://adni.loni.usc.edu/data-samples/access-data/>.

Conflicts

DB and AJZ have no conflicts of interest. JRB is an employee of Embic Corporation whose assessment outcomes were evaluated in this project.

Funding Sources

DB and AJZ have not received any funding from Embic, nor have any other relevant funding to disclose for this project. Data collection and sharing for the Alzheimer's Disease Neuroimaging Initiative (ADNI) is funded by the National Institute on Aging (National Institutes of Health Grant U19 AG024904). The grantee organization is the Northern California Institute for Research and Education. In the past, ADNI has also received funding from the National Institute of Biomedical Imaging and Bioengineering, the Canadian Institutes of Health Research, and private sector contributions through the Foundation for the National Institutes of Health (FNIH) including generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics.

Consent Statement

ADNI data collection was approved by the ethics committees of the participating institutions, and completed in accordance with the Declaration of Helsinki. All participants provided informed consent prior to testing.

Keywords

Digital cognitive biomarkers, Clinical Dementia Rating, Cognitive Assessment, ADAS-Cog

Figure Legends

Figure 1. Hierarchical Bayesian Cognitive Processing Model. The model has three episodic memory storage states (P , T , and D), four processes of encoding into them (N_1 , N_2 , N_3 , and N_4), and three processes of retrieval from them (R_1 , R_2 , and R_3) that episodic memory of a word may exist in or transition through during each immediate and delayed free recall task.

Figure 2. Conditional estimates plot with 95% confidence intervals (shaded area). The Y-axis represents the probability of CDR decline at follow-up (36 months). The X-axis represents M scores at baseline. Grey circles represent unique individuals' data points. Circles at the bottom represent individuals who did not show CDR decline at follow up, whereas circles at the top represent CDR decliners at follow-up.

Figure 3. Receiver Operating Characteristics (ROC) plot comparing false positive rate (Y-axis) to false negative rate (X-axis) in the frequentist logistic analysis.