ADNI Alzheimer's Disease Neuroimaging Initiative

Digital Cognitive Biomarkers: Quantifying Latent Cognitive Processes of Encoding and Retrieval with Hierarchical Bayesian Cognitive Processing Models

Jason R. Bock[1,2]; Junko Hara[1]; Dennis Fortier[1]; Ronald C. Petersen[3]; Steven M. Smith[3]; Jeffrey L. Cummings[4]; William R. Shankle[1]; Kaavya Shah[5]; Michael D. Lee[2]

[1]Embic Corporation; [2]Dept. of Cognitive Sciences, University of California at Irvine; [3]Alzheimer's Disease Research Center, Mayo Clinic; [4]School of Integrated Health Sciences, University of Nevada at Las Vegas; [5]School of Information, University of California at Berkeley

Summary

Embic's Digital cognitive biomarkers (**DCBs**) are quantified representations of the unobservable (latent) cognitive processes that underlie overall cognitive function. The DCBs are generated for various encoding and retrieval processes with a hierarchical Bayesian cognitive processing (**HBCP**) model that analyzes item response data from commonly used wordlist memory (**WLM**) tests of learning, recall, and recognition. The model generates seven base DCBs, each representing the probability of information processed through different encoding (N_1, N_2, N_3, or N_4) or retrieval (R_1, R_2, or R_3) pathways, to and from three distinct storage states (*pre-task*, *transient*, or *durable* storage states). There are three additional measures (M_1, M_2, and M_3) that quantify a person's probability of recall from transient storage on immediate recall tasks, durable storage on immediate recall tasks, and durable storage on delayed recall tasks, respectively. In contrast to purely observed behaviors (e.g., the number of words recalled), this generalized probability of recall, which includes the combination of encoding and retrieval processes that result in successful recall across WLM test tasks, is constructed from base DCBs subsequent to modeling.

This method document provides an overview of DCB generation. Accompanying this document is a datafile, ADASDCB.csv, containing all DCBs calculated for each ADAS-Cog WML assessment for 2,258 subjects (10,398 assessments) assessed under ADNI1, ADNI-GO, ADNI2, and ADNI3, date ranging from 5/17/2005 to 11/1/2021.

Background

Wordlist memory (**WLM**) tests are the most common measures of verbal episodic memory used in clinical and research settings [1,2]. Their total scores as observed behaviors are frequently used to screen individuals prior to neuroimaging or other assessments for cognitive impairment or dementia stages of Alzheimer's disease (**AD**) and to monitor progressive decline and treatment effects [3]. However, as AD research shifts its focus toward earlier or even asymptomatic or preclinical stages of the disease, the change at those stages may be very subtle and difficult to measure [4] and require more sophisticated approaches to maximize the information collected through WLM tests and achieve the greatest precision of measurements [5].

One such approach is the application of a hierarchical Bayesian cognitive processing (**HBCP**) model to item response data from commonly used WLM tests of learning, recall, and recognition. This approach not only improves the use of information collected through WLM tests, but also enables characterization of more subtle but distinct cognitive changes at the unobservable processing level that are not quantifiable using observed behaviors.

Developing the Hierarchical Bayesian Cognitive Processing Model for Generation of Digital Cognitive Biomarkers

Generating Digital Cognitive Biomarkers

To generate DCBs, an HBCP model is applied to item response data from commonly used WLM tests (e.g., ADAS-Cog, AVLT, CVLT, EVLT, and CERAD) with two or more immediate learning and recall tasks and at least one delayed recall task [6]. The model generates seven base DCBs, each representing the probability of information processed through different encoding (N_1, N_2, N_3, or N_4) or retrieval (R_1, R_2, or R_3) pathways, to and from three distinct storage states (*pre-task*, *transient*, or *durable* storage states). **Figure 1** is a graphical representation of how items on the WLM test move among storage states via the encoding DCBs and how they are recalled from them via the retrieval DCBs, across the tasks of the test. When data from a delayed recognition task is available, the model incorporates it for additional precision. Normative, prior information is provided to the model from evaluation of a large dataset of subjects, which were sampled from a database of nearly 2,000,000 individuals in the general population without diagnosis of cognitive impairment.

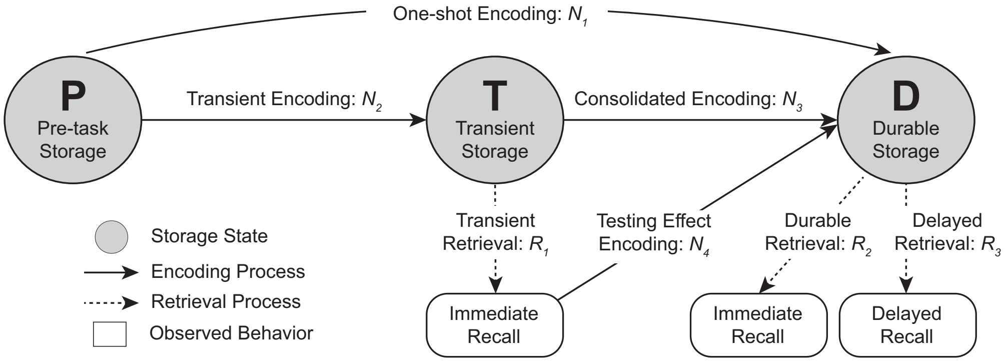

Figure 1. Hierarchical Bayesian Cognitive Processing Model. The model has three episodic memory storage states (P, T, and D), four processes of encoding into them (N_1, N_2, N_3, *and* N_4), and three processes of retrieval from them (R_1, R_2, *and* R_3).

Since the DCBs are generated by modeling latent processes which are used in every WLM test, the DCBs themselves represent absolute measures that can be compared across different WLM tests. However, it is important to note that a modified HBCP model is required for generating DCBs on WLM tests that shuffle the order of the word presentation following each recall trial or WLM tests that use multiple, non-equivalent wordlists. This is because each feature of WLM tests (e.g., word presentation, wordlists, the number of words) affects recall performance and those features need to be taken into consideration when generating the DCBs. More details on the different WLM test features and the importance of matching analytical method to those available features are discussed elsewhere [7].

HBCP Model Development

The HBCP model was developed from and expanded upon a multinomial processing tree (**MPT**) model, the Batchelder model [8], which calculates the probability for each possible pattern of recall across WLM test tasks for each presented word. These patterns are represented as binary tuples of 0s, for non-recall, and 1s, for recall. In a WLM test with four tasks (e.g., ADAS-Cog, EVLT), a word has one of 16 tuples: 0000, 0001, ..., 1111. The model-calculated probability that a word exhibits one of these recall patterns is derived from the DCBs, which are generated through Bayesian inference to accurately do so. For example, for the pattern 0000, indicating that an item was never recalled on any task, the probability is:

$$
\begin{aligned}
&\theta_{0000} = \\
&N_1(1-R_2)(1-R_2)(1-R_2)(1-R_3) + \\
&(1-N_1)N_2(1-R_1)N_3(1-R_2)(1-R_2)(1-R_3) + \\
&(1-N_1)(1-N_2)N_1(1-R_2)(1-R_2)(1-R_3) + \\
&(1-N_1)N_2(1-R_1)(1-N_3)(1-R_1)N_3(1-R_2)(1-R_3) + \\
&(1-N_1)(1-N_2)(1-N_1)N_2(1-R_1)N_3(1-R_2)(1-R_3) + \\
&(1-N_1)(1-N_2)(1-N_1)(1-N_2)N_1(1-R_2)(1-R_3) + \\
&(1-N_1)N_2(1-R_1)(1-N_3)(1-R_1)(1-N_3)(1-R_1) + \\
&(1-N_1)(1-N_2)(1-N_1)N_2(1-R_1)(1-N_3)(1-R_1) + \\
&(1-N_1)(1-N_2)(1-N_1)(1-N_2)(1-N_1)N_2(1-R_1) + \\
&(1-N_1)(1-N_2)(1-N_1)(1-N_2)(1-N_1)(1-N_2).
\end{aligned}
$$

The above equation shows that the probability of patterns of recall are the sums of more basic probabilities that correspond to specific ways in which the pattern could arise. If the item is never recalled because it was never encoded (remains in *pre-task* storage), the probability of this happening is $(1-N_1)(1-N_2)(1-N_1)(1-N_2)(1-N_1)(1-N_2)$. The item may also never be recalled because, while it was encoded into *durable* storage, it was never retrieved: $N_1(1-R_2)(1-R_2)(1-R_2)(1-R_3)$. Other rows in the equation correspond to other possible paths through the MPT that result in this pattern, and other patterns are constructed similarly [6].

The seven base DCBs that generate the response patterns through the Batchelder MPT are estimated through Bayesian inference with a Markov-chain Monte Carlo (**MCMC**) algorithm. Generation of DCBs for ADNI was done with the JAGS software program [9] and used 1,000 samples each from 8 independent MCMC chains evaluated for convergence with the $\hat{R}$ statistic.

Consideration of Wordlist Presentation Orders (Shuffled vs. Fixed)

Some WML tests employ a shuffled wordlist presentation (e.g., ADAS-Cog), whereby each word in the list is presented in a different position during each of the three learning tasks. Since shuffling the word presentation order eliminates the accumulative serial position effects of primacy and recency on encoding (which is a loss of available information), generating DCBs for this test protocol requires modeling the item-level responses with a distinct approach compared to that used for fixed-order tests. The modified HBCP model accounts for irregular differences in encoding and retrieval strength for various items in various positions across the wordlist.

Additional Consideration for Multiple Non-Equivalent Wordlists

Additionally, the ADAS-Cog WLM test in the Alzheimer's Disease Neuroimaging Initiative (**ADNI**) uses three different wordlists across different visits. Because individual item responses are used as the input for generating DCBs, the features of those items (e.g., word concreteness, valence, or semantic distance to other words) impact DCB values. As a result, multiple, non-equivalent wordlists can only be evaluated together after applying a normative adjustment for each word and DCB. To accommodate for differences across the ADAS-Cog wordlists used in ADNI, a normative adjustment was developed and applied to the three ADAS-Cog English wordlists. Three separate models were run to obtain this normative adjustment, one on each of the three wordlists, using only the first assessment on which a subject received a given wordlist (assessment n's = 2,286; 1,537; 673). These models inferred individual assessment DCBs in isolation and inferred per-DCB and -position penalties, which were applied to all assessments in the model, that account for the interactions between words on the wordlist, the positions they are presented in, and the DCBs used to encode and retrieve them. For example, the words in position 2 for wordlist 1 are *ARM, LETTER, LETTER*, and N_1 for this position is little penalized (-0.003); the words in the same position 2 for wordlist 2 are *POTATO, TEMPLE, TEMPLE*, and N_1 for this position is more strongly penalized (-0.418) to account for the different features of the words. This same penalty may now be applied to any assessment with that same ADAS-Cog wordlist entirely independently to adjust for word features.

Generalizability Across Different WML Tests

To evaluate the generalizability of DCBs across different WLM tests, values were generated using item response data from 2,456 subjects who were assessed with three different WLM tests (i.e., ADAS-Cog, EVLT, or AVLT) and who had diagnoses of cognitive normalcy (**CN**), amnestic mild cognitive impairment (**aMCI**), or dementia due to Alzheimer's disease (**AD**) [10]. The values and the changes in DCBs from CN to AD were consistent across the three different WLM tests and data sources (**Figure 2**). The results demonstrate that these DCBs are robust and generalizable, regardless of the underlying test protocol, and allow comparison of various studies conducted using different WLM tests.

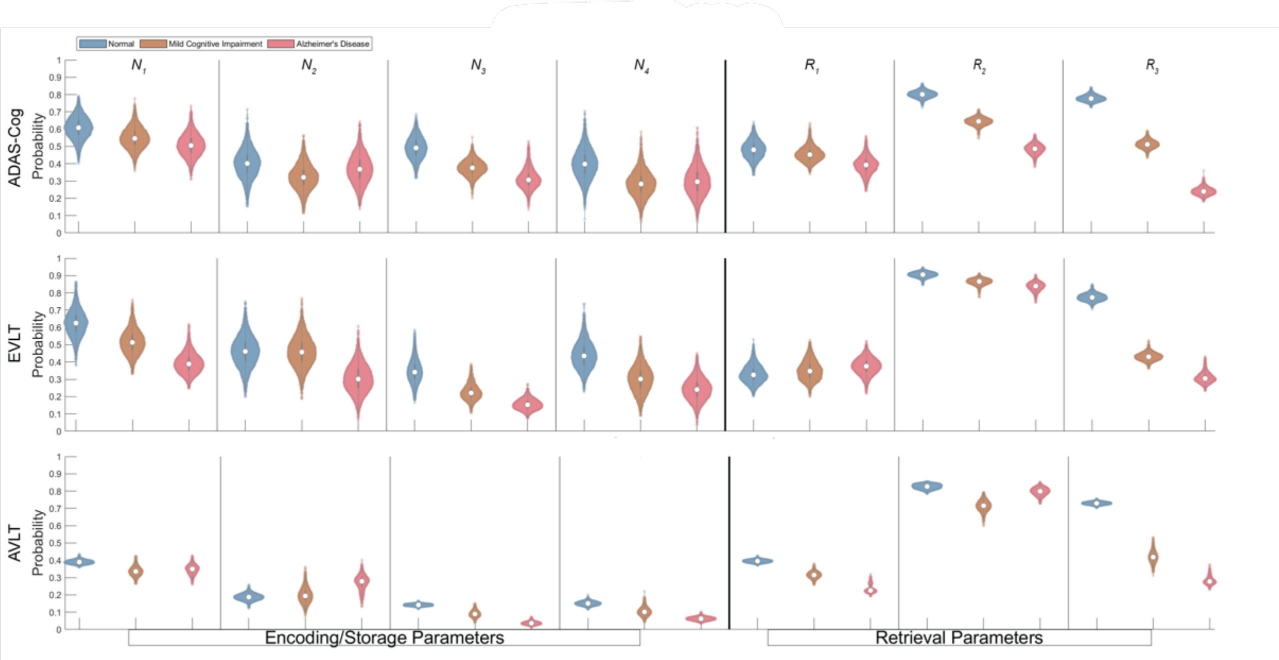

Figure 2. Distributions of DCBs by Disease Severity (CN, MCI, and AD dementia) and by Wordlist Memory Test (ADAS-Cog, EVLT, and AVLT). Across all three tests, highly similar values and slopes are observed for the distribution means of each group, providing strong evidence of DCB generalizability.

Relation to Functional Decline

To evaluate the DCBs against functional decline, DCBs were compared against the Functional Assessment Staging Test (**FAST**) for each of 14,096 EVLT test assessments from 3,635 patients at a cognitive disorders clinic. The results demonstrated characterization of functional decline from stage to stage across all DCBs (**Figure 3**), with noteworthy patterns of greater N_2 decline for words in primacy positions than in recency positions and a dramatic R_3 drop between stages 3 (MCI) and 4 (mild dementia) [6].

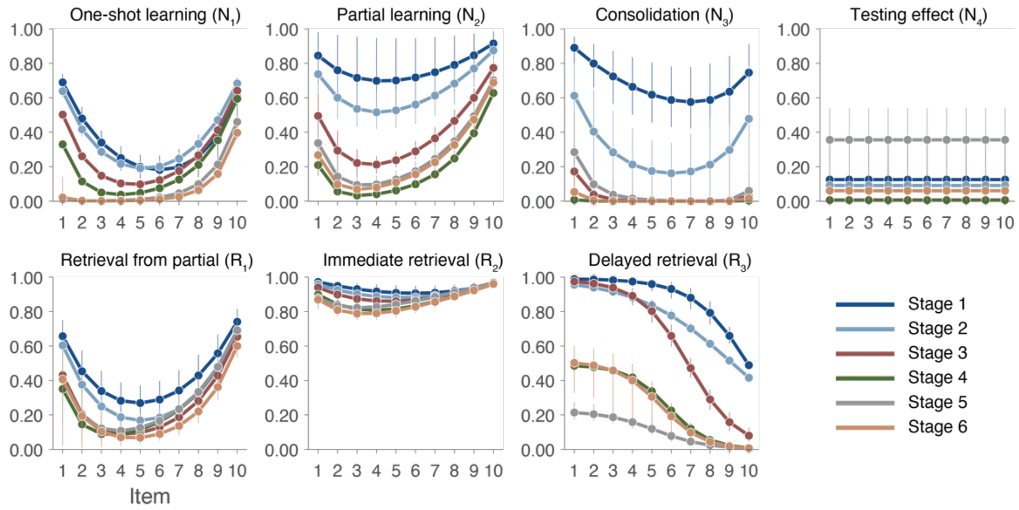

Figure 3. DCB means and 95% Credible Intervals by Each of the 10-word List in the EVLT and by FAST stage. Distinctive patterns of decline can be observed from less to more severe stages, including a faster decline for encoding words at the beginning of the list (primacy) than at the end of the list (recency). Of particular note is the pattern of decline for R_3. There is preserved delayed retrieval for list-beginning words through MCI (with declining retrieval for list-ending words) and a sharp drop in retrieval ability upon progression to mild dementia.

Generating Digital Cognitive Biomarkers for ADNI ADAS-Cog

ADNI ADAS-Cog Dataset

The dataset used for generation of DCBs for ADNI was downloaded on 11/2/2021 and includes assessments from ADNI1, ADNIGO, ADNI2, and ADNI3, ranging in date from 5/17/2005 to 11/1/2021. Assessments (n = 10,398 from 2,258 subjects) were included if there was no missing data for any of the following factors: item responses from the ADAS-Cog Word Recall test (immediate and delayed), specified ADAS-Cog Word Recall test wordlist, age, sex, and years of education. The sample was 47% female (n = 1,069), with a mean age at baseline of 73.05 years (*SD* = 7.27) and a mean education of 16.11 years (*SD* = 2.71). Diagnoses at baseline were 36% cognitively normal (n = 819), 45% mild cognitive impairment (n = 1,021), and 17% Alzheimer's dementia (n = 388).

Generating DCBs for ADNI ADAS-Cog Dataset

The HBCP model, including adjustment for word presentation position effects on each of the three ADAS-Cog English wordlists, was applied to all subject assessments independently of one another. Using Bayesian inference to include prior information of DCB distributions for a typical individual in the general population, the HBCP model updated the ADNI individual assessment DCBs according to ADAS-Cog WLM test performance. This results in DCBs that are informative regarding the probability of a subject's use of specific, latent cognitive processes, calculated from a single WLM test.

For ADNI ADAS-Cog, 7 base DCBs and 3 additional *M* measures were calculated (**Table 1**) for the ADNI database depository.

Table 1. Available Digital Cognitive Biomarkers for ADNI ADAS-Cog

DCBs	Correlate	Description
N_1	Encoding	Probability of encoding into the DURABLY LEARNED State
N_2	Encoding	Probability of encoding into the TRANSIENTLY LEARNED State
N_3	Encoding	Probability of encoding into the DURABLY LEARNED State, following previous TRANSIENT LEARNING (N2)
N_4	Encoding	Probability of encoding into the DURABLY LEARNED State, due to successful retrieval (R1) from the TRANSIENTLY LEARNED State
R_1	Retrieval	Probability of retrieving from the TRANSIENTLY LEARNED State
R_2	Retrieval	Probability of retrieving from the DURABLY LEARNED State
R_3	Retrieval	Probability of retrieving from the DURABLY LEARNED State after a 5-minute delay with distraction
M_1	Recall	Probability of immediate recall of a non-durably stored episodic memory
M_2	Recall	Probability of immediate recall of a durably stored episodic memory
M_3	Recall	Probability of delayed recall of a durably stored episodic memory

Evaluating ADNI ADAS-Cog DCBs

Relation to Disease Progression (Clinical Diagnosis)

To evaluate the DCBs against different disease stages, DCBs generated for each of 10,398 ADAS-Cog WLM tests from 2,258 subjects enrolled in the ADNI were compared against clinical diagnoses (CN, MCI, and AD) (**Figure 2**). After accounting for varying degrees of item effects in the shuffled-order test, the results show the expected decline from CN to MCI to AD across subject assessments for many of the DCBs. The *M* measures of generalized recall probability, summated from the DCBs, show yet more pronounced decline across clinical diagnoses.

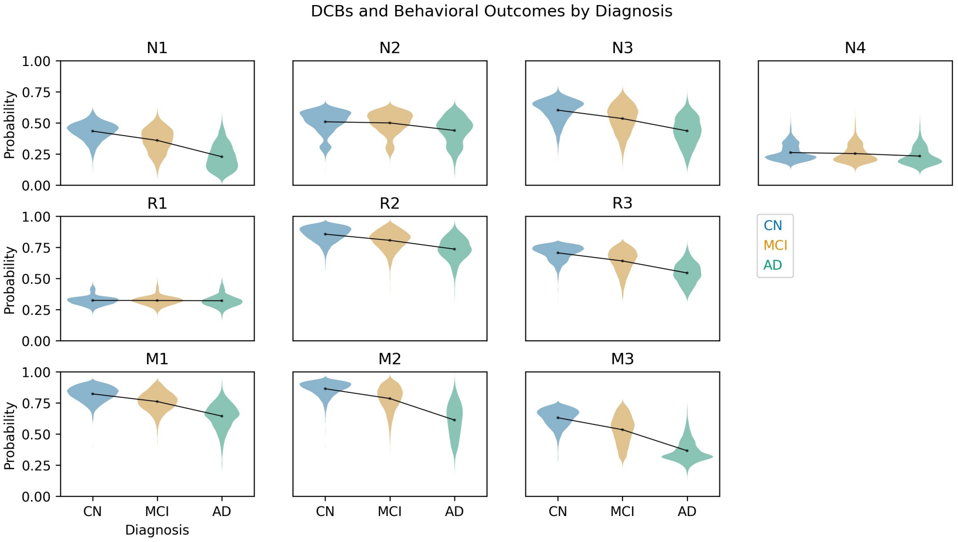

Figure 2. Distributions of DCBs (N_1 through R_3) and Behavioral Outcomes (M_1, M_2, and M_3) by Disease Severity (CN, MCI, and AD). Decline in many DCBs can be observed with increasing severity.

Relation to Observed Behavior and Clinical Outcomes

To evaluate the relationship between generalized recall probability of DCBs and the observed recall behavior, *M* measures (M_1, M_2, and M_3) generated from ADNI ADAS-Cog WLM DBCs were visually compared to total word recall for each immediate and delayed recall task and all tasks combined. Visual comparison of *M* parameters and total word recall showed a clear curvilinear relationship (**Figure 3**).

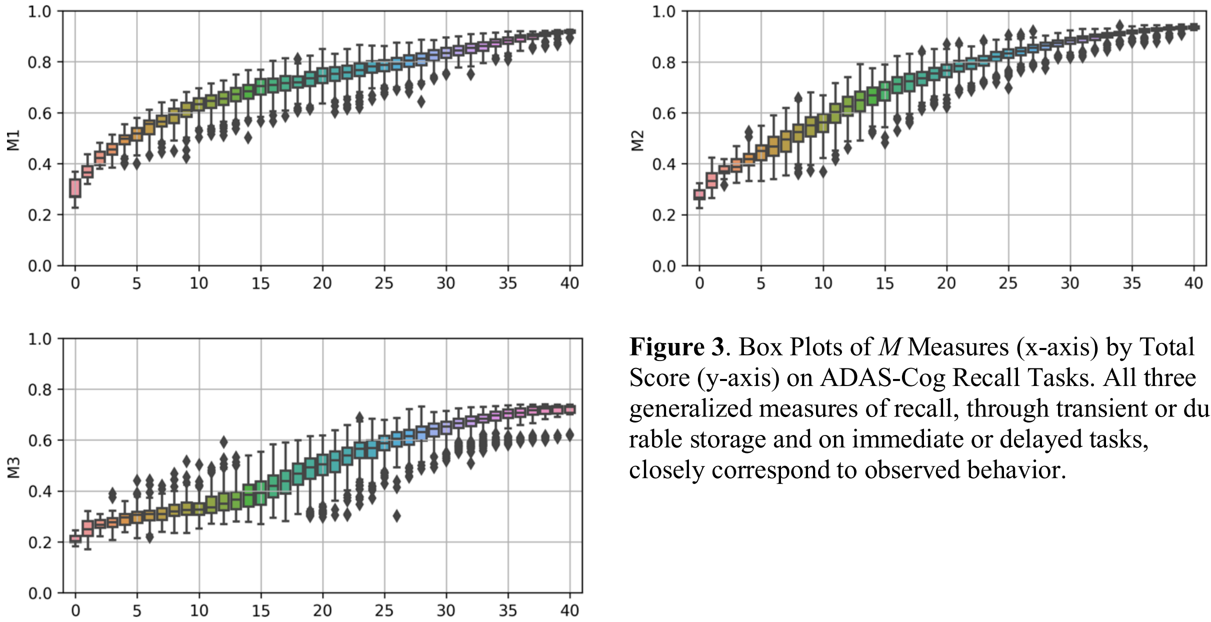

Figure 3. Box Plots of *M* Measures (x-axis) by Total Score (y-axis) on ADAS-Cog Recall Tasks. All three generalized measures of recall, through transient or durable storage and on immediate or delayed tasks, closely correspond to observed behavior.

To demonstrate the capability for classification of cognitive impairment with the *M* measures, these measures were included as predictors, along with demographics, in a Bayesian logistic regression model to classify MCI (against CN as the reference) with a sample of 7,445 assessments, age ≥ 65, from ADNI. The β coefficients of M_2 and M_3 showed extreme evidence for difference from the uninformed prior distribution centered on 0 (*BF*s > 1,000), with means = -6.99 (*SD* = 0.71) and -6.47 (*SD* = 0.48), respectively. The M_1 β coefficient distribution demonstrated strong evidence in favor of the uninformed prior distribution centered on 0 (*BF* = 24.7), with mean = -1.34 (*SD* = 0.67). Higher *M*s corresponded with reduced likelihood for MCI. A receiver operating characteristic (ROC) curve was generated with the model's predicted probability against observed severity (AUC = .79). A model with demographics only was analyzed with the same approach (AUC = .61) for comparison of data. Additionally, both models were analyzed with frequentist logistic regressions for comparison of methods, and both models and all predictors were significant (*p*s < .05), with pseudo R^2 = .20 and .03, respectively. A likelihood-ratio test found significantly greater prediction of the data under the model with *M* parameters than with demographics only, $\chi^2(3) = 1818.25$, $p < .001$.

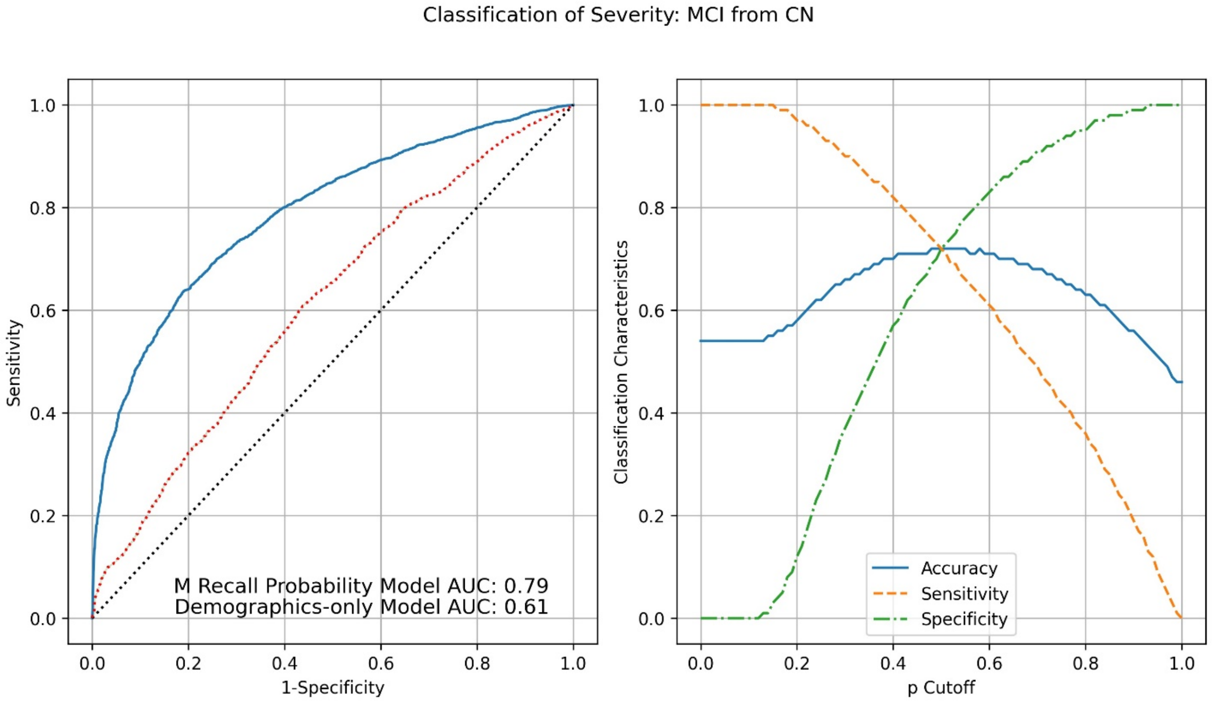

Figure 4. Left: Receiver Operating characteristic (ROC) curves, generated from logistic regression predicted probability of MCI (CN reference) with *M* measures (blue) and with demographics only (red). Right: Classification characteristics by cutoff value of *M* measures model predicted probability.

Dataset Information

This methods document applies to the following dataset(s) available from the ADNI repository:

Dataset Name	Date Submitted
ADASDCB.csv	July 27, 2022

Reference

1. Belleville S, Fouquet C, Hudon C, Zomahoun HTV, Croteau J. Neuropsychological Measures that Predict Progression from Mild Cognitive Impairment to Alzheimer's type dementia in Older Adults: a Systematic Review and Meta-Analysis. Neuropsychology Review. 2017;27(4):328–353. https://doi.org/10.1007/s11065-017-9361-5
2. Salmon DP, Bondi MW. Neuropsychological Assessment of Dementia. Annual Review of Psychology. 2009;60(1):257–282.
3. Lezak MD, Howieson DB, Bigler ED, Tranel D. (2012). Neuropsychological assessment (5th ed.). Oxford University Press.
4. Rafii MS, Aisen PS. Alzheimer's Disease Clinical Trials: Moving Toward Successful Prevention. CNS drugs. 2019;33(2): 99-106. https://doi.org/10.1007/s40263-018-0598-1
5. Aisen PS, Cummings J, Jack CR, Morris JC, Sperling R, Frölich L, Jones RW, Dowsett SA, Matthews BR, Raskin J, Scheltens P, Dubois B. On the path to 2025: Understanding the Alzheimer's disease continuum. Alzheimer's Research & Therapy. 2017;9(1):60-60. https://doi.org/10.1186/s13195-017-0283-5
6. Lee MD, Bock JR, Cushman I, Shankle WR. An application of multinomial processing tree models and Bayesian methods to understanding memory impairment. Journal of Mathematical Psychology. 2020; 95, Article 102328. https://doi.org/10.1016/j.jmp.2020.102328
7. Bock JR, Russell J, Hara J, Fortier D. Optimizing Cognitive Assessment Outcomes for Alzheimer's Disease by Matching Wordlist Memory Test Features to Scoring Methodology. Front. Digit. Health, 03 November 2021. https://doi.org/10.3389/fdgth.2021.750549
8. Alexander GE, Satalitch TA, Shankle WR, Batchelder WH. A Cognitive Psychometric Model for Psychodiagnostic Assessment of Memory Deficit Disorders. Psych Assessment. 2016;28(3):279-93.
9. Plummer, M. JAGS: A program for analysis of bayesian graphical models using Gibbs sampling. In K. Hornik, F. Leisch, & A. Zeileis (Eds.), Proceedings of the 3rd international workshop on distributed statistical computing. Vienna, Austria.
10. Shankle WR, Hara J, Bock JR, Fortier D, Mangrola T, Lee MD, Alexander GE, Batchelder WH, Petersen RC, Kremers W. Using Graphical Hierarchical Bayesian Cognitive Process Models Applied to Common Memory Tests to Predict AD Pathology within Normal Subjects. CTAD 2018. Poster Presentation. Barcelona, November 2018.

About the Authors

The authors Jason R. Bock, Junko Hara, Dennis Fortier, and William Shankle are employee of Embic Corporation. Bock serves as Director of Informatics at Embic Corporation, and is a research fellow at UC Irvine. Hara serves as Chief Science Officer, Shankle as Chief Medical Officer, and Fortier as Chief Executive Officer at Embic Corporation. Kaavya Shah is a consultant to Embic Corporation. Ronald C. Petersen is a professor in the department of Neurology at Mayo Clinic and directs the Mayo Clinic Alzheimer's Disease Research Center and the Mayo Clinic Study of Aging. Steven Smith is Sr. Program Coordinator, Alzheimer's Disease Research Center at Mayo Clinic. Jeffrey L. Cummings is Research Professor, Director, Chambers-Grundy Center for Transformative Neuroscience at University of Nevada, Las Vegas, School of Allied Health Sciences. Michael D. Lee is a professor in the department of Cognitive Sciences at UC Irvine.

For more information, please contact:
Jason R. Bock at jrbock@embic.us or
Junko Hara at junkoh@embic.us

Notice: This document is presented by the author(s) as a service to ADNI data users. However, users should be aware that no formal review process has vetted this document and that ADNI cannot guarantee the accuracy or utility of this document.