The Journal of Prevention of Alzheimer's Disease 13 (2026) 100642

Contents lists available at ScienceDirect

The Journal of Prevention of Alzheimer's Disease

journal homepage: www.elsevier.com/locate/tjpad

Original Article

MIND GamePack: A novel digital health technology for assessment of participants with and without cognitive impairment

Jake A. Galler [a,1], Liuqing Yang [b,1], Chao-Yi Wu [a], Cathrine Young [a], Edmarie Guzmán-Vélez [c], Katherine W. Cropp [a], Anthony W. Bannon [b], Hiroko H. Dodge [a], Jessica A. Gerber [a], Steven E. Arnold [a,*]

[a] Department of Neurology, Massachusetts General Hospital, Harvard Medical School, 55 Fruit Street, Boston, MA, 02114, United States
[b] AbbVie Inc., 1 Waukegan Rd., North Chicago, IL, 60064, United States
[c] Department of Neurology, Rutgers Robert Wood Johnson Medical School, 112 Paterson Street, New Brunswick, NJ, 08901, United States

ARTICLE INFO

Keywords:
Digital health technologies
Remote cognitive assessment
Digital biomarkers
Neuropsychological testing
High-frequency sampling

ABSTRACT

Background: Digital biomarkers offer promising avenues for enhancing the identification, prognosis, and phenotyping of Alzheimer's Disease (AD). This study evaluates the feasibility and utility of the MIND GamePack©, a digital game platform designed to measure leisure cognitive activity over time.
Methods: Gameplay data were collected from 60 participants across two cohorts over 3–6 months: Cohort I (no cognitive complaints) and Cohort II (subjective complaints, mild cognitive impairment, mild dementia). Analyses focused on compliance, correlation with validated neuropsychological instruments (Montreal Cognitive Assessment [MoCA], Repeatable Battery for the Assessment of Neuropsychological Status [RBANS], and Trail Making Test [TMT]), test-retest reliability, and known groups comparison.
Results: The MIND GamePack© displayed high compliance and excellent test-retest reliability within 1 week (ICC=0.81–0.95) for most selected features. Several game features displayed moderate to strong correlations with standardized neuropsychological test performance. Features related to memory tasks across select games (Normalized Accuracy and Normalized Redundant Move Variability of Memory Match and Word Repetition Rate of Word Scramble) exhibited significant associations with RBANS Sum of Index Score, TMT Part A, and MoCA, respectively. Participants without cognitive impairment exhibited improvement in features over 3 months compared to those with cognitive impairment. Baseline game features differentiated cognitively normal [CN] from cognitively impaired [CI] participants across nearly all domains after controlling for age, sex, and education ($p<.01$).
Conclusions: The MIND GamePack© offers a sensitive, engaging approach to cognitive monitoring and screening, with potential use for dense tracking in longitudinal research and interventional clinical trials.

1. Introduction

Neurocognitive disorders of aging, including Alzheimer's disease (AD) and related disorders (ADRD) emerge and progress along varied biological and clinical trajectories [1–5]. Advances in imaging and biofluid biomarkers have yielded sensitive and disease-specific measures to detect brain disease pathology and track disease progression. Digital health technologies (DHTs) that detect and monitor cognitive symptom progression in ADRD have also rapidly advanced. These include devices, programs, and applications, as well as networks and online spaces, that have been developed to measure movement (e.g., steps, haptics), sound (e.g., voice), and physiology (e.g., pulse) [6,7] These have been applied to detect and monitor cognition, behavior, and even emotion in AD/ADRDs and other brain disorders.

Standardized neuropsychological testing, whether paper & pencil or computerized, remains the gold standard for evaluation of cognitive impairment and dementia due to AD/ADRDs [4]. These are traditionally administered in a controlled exam room or office setting once or periodically and may take an hour or more to complete. Remote web-based administrations, and even smart-phone applications have demonstrated

* Corresponding author at: MassGeneral Neurology, 149-13th Street, Charlestown, MA 02114, United States.
E-mail address: searnold@mgh.harvard.edu (S.E. Arnold).
[1] These authors contributed equally

https://doi.org/10.1016/j.tjpad.2026.100642
Received 25 March 2026; Received in revised form 28 May 2026; Accepted 20 June 2026
Available online 9 July 2026

good validity and have made neuropsychological testing more widely accessible and economically viable [8–12]. However, even these are burdensome and they are often experienced as challenging if not frustrating and unpleasant, leading to suboptimal compliance, effort, and reliability [13].

In longitudinal studies, high frequency and dense sampling of outcomes offers many advantages over occasional assessment for accurate and precise estimation of cognitive abilities and has been shown to be as reproducible and replicable as traditional neuropsychological test outcomes [14]. Thus, the development of more validated DHTs which capture high frequency data in a user-friendly and engaging format remains an active area of research. We previously introduced the Mass-General Institute for Neurodegenerative Disease (MIND) GamePack© iPadOS platform (**Figure S1**) and demonstrated its feasibility as an engaging platform in 26 generally healthy older adults with no cognitive complaints (NCC), supported by high compliance rates [91% ± 11%] and high weekly average playtimes [210 ± 132 min] [15].

The aims of the current study were to (1) further investigate compliance in people with a range of cognitive impairments (2) present neuropsychological assessment correlation results (3) assess test-retest reliability, and (4) perform known groups comparison of the MIND GamePack© platform in identifying cognitive impairment. Here, we present the diagnostic and psychometric properties for the MIND GamePack from this Cohort 1 study that was limited to participants with normal cognition, along with new data from a Cohort II study consisting of 34 participants with subjective cognitive decline (SCD), mild cognitive impairment (MCI), or mild dementia (MD).

2. Methods

2.1. Description of study population

Sixty-eight participants from Cohorts I-II were screened to determine eligibility ($n_{COHORT\ I} = 29$, $n_{COHORT\ II} = 39$). A CONSORT diagram is presented in **Figure S2A**. Participants were recruited from the metro-Boston area through community outreach efforts, clinician referrals, and social media. Key eligibility criteria (**Table S1A, Table S1B**) included: (1) age 55–90, (2) high school education or greater, (3) English proficiency, and (4) a Montreal Cognitive Assessment (MoCA) score $\geq$ 26 in Cohort I or $\geq$ 16 in Cohort II. Seven participants did not meet the eligibility criteria due to ineligible MoCA scores. One participant withdrew, citing financial concerns. The remaining 60 participants completed all study procedures. Individuals in the SCD group had a Cognitive Function Instrument (CFI) [16] score $\geq$2 and a Telephone Interview of Cognitive Status (TICS) [17] score $\geq$31. MCI and MD status was designated based on clinical diagnosis derived from medical record review by a board-certified neurologist specializing in neurodegenerative dementia. NCC and SCD participants were categorized as cognitively normal (CN), and MCI and MD were categorized as cognitively impaired (CI) based on performance on standardized cognitive testing at baseline and verified by both a neurologist and licensed neuropsychologist to establish consensus.

2.2. Procedure

Participants from both Cohort I and Cohort II were given an Apple iPad 8th/9th Generation at their Screening Visit (initiated between Day −21 and Day −14, concluding on Day −1). All 29 participants in Cohort I received Apple iPad 8th Generation devices, while 25/34 (73.5%) of Cohort II participants received Apple iPad 8th Generation devices and the remaining nine (26.5%) received Apple iPad 9th Generation devices. In benchmark comparability analyses, we noted no differences in distributions of response rates or data in gameplay. During the Screening Visit period, participants received standardized training on (1) basic hardware features of the study device (2) basic software features of the device operating system (i.e., iPadOS), and (3) the MIND GamePack© application. Participants were instructed to play a minimum of "5 min per game, per day, for 5 days a week" (100 min/week) to maximize all game play within session and minimize daily participant burden. Beyond the minimum, they were allowed to play ad lib. Both cohorts received compensation based on meeting weekly compliance requirements. Compliance for each cohort began at Baseline (Day 0) to allow for the capture of procedural learning effects and device troubleshooting as needed during the screening period. Cohort I participants (NCC) enrolled in a minimum 3-month (12-week) pilot feasibility study, with a telephone check-in at week 6. Cohort II participants (SCD, MCI, MD) enrolled in a 6-month (24-week) study, with telephone check-ins at weeks 3, 6, 12, and 18 to better assess learning plateaus and build upon the promising results of Cohort I.

2.3. MIND GamePack©

The MIND GamePack© is a digital platform with four familiar games: Memory Match (inspired by Concentration/Memory®), FreeCell (Solitaire), Word Scramble (Boggle™), and Block Drop (inspired by Tetris®). Memory Match is a digitization of a simple picture card memory game, where players utilize short-term recall to match cards in an untimed format. The user may select difficulty levels 1–26, respectively, ranging from 2 × 2 to 8 × 8 matrices of cards. FreeCell is an untimed Solitaire variant played with a standard deck of 52 playing cards, where players aim to arrange a pre-configured deal into "foundation" piles, organizing by suit in ascending order. Unlike Solitaire, players may utilize four *free cells* for temporary storage. Word Scramble is a 5-minute timed word puzzle like Boggle™ where players are tasked with spelling as many words as possible within a 5 × 5 randomized matrix of letters/phonemes (e.g. 'k', 'qu', 'e'). Block Drop is a 5-minute dynamic block puzzle like Tetris™ which tests both players' perceptual-motor abilities and visuospatial attention/awareness [18]. Each game captures raw data and calculates summary feature metrics on participants' individual game sessions. Data is transmitted to a secure internal (i.e. Massachusetts General Brigham [MGB]) Google Cloud Computing Platform (GCP) instance for query, analysis, and storage automatically upon completion of a game session.

2.4. Game features

Across all four games of the MIND GamePack©, 10 visit-level representative features were identified as analysis candidates (Table 1). For games with multiple difficulty levels (i.e. Memory Match (levels 1-26), Block Drop (levels 1-5), and FreeCell (levels 1-3), raw game features were normalized by computing the z-scores relative to normative data: first the mean and standard deviation for NCC participants at a specific difficulty level was calculated, then raw game features were standardized by subtracting and dividing the difficulty level-specific mean and standard deviation. At baseline visit (Day 0), game features were calculated by aggregating data within a 2-week window post-baseline. For Month 3 and Month 6 assessments, game feature aggregation occurred using a 30-day window post-timepoint to enhance data inclusion across subjects (i.e., Day 60–90 and Day 150–180, respectively). A minimum of five gameplay sessions per game within each aggregation window was required to maintain data quality. Game-specific quality filters were further applied to session-level data: Memory Match included levels above seven, as levels 1–7 were considered too easy (participants achieved a ceiling 100% accuracy in a short time); FreeCell session data required a winning outcome; Word Scramble required timer completion, no pauses, and at least one successful word; and Block Drop required completion of in-game timer or a "block out" event. Sample sizes following quality filtering are visualized in S2A and tabulated in S2B.

Table 1
Descriptions of game features and ICC (95% C.I.) of select features aggregated by different periods of data used.

Game		Feature	Game-level Description	Duration of Aggregation			
				1 Day	3 Days	1 Week	2 Weeks
Memory Match	1	Difficulty Level[†]	Level selected by player [Range: 1–26] Size of session matrix [Range: 2 × 2 – 8 × 8]	0.78 (0.76,0.8)	0.76 (0.73,0.79)	0.82 (0.78,0.85)	0.88 (0.85,0.91)
	2	Normalized Redundant Move Variability	Total number of same-pair redundant (i.e. non-match) cards selected	0.41 (0.37, 0.45)	0.6 (0.55,0.64)	0.83 (0.8,0.86)	0.8 (0.75,0.85)
	3	Normalized Accuracy[†]	Percentage of pair selections within a session which lead to a match	0.72 (0.69, 0.74)	0.84 (0.81,0.86)	0.90 (0.88,0.92)	0.95 (0.94,0.97)
Word Scramble	4	Word Repetition Rate	Number of repeated words divided by total word attempts per game	0.61 (0.58,0.64)	0.73 (0.7,0.77)	0.81 (0.77,0.84)	0.90 (0.86,0.92)
	5	Word Repetition Rate Variability					
	6	Total Score[†]	Total score obtained by the player during a game session	0.70 (0.67,0.72)	0.80 (0.77,0.82)	0.88 (0.86,0.9)	0.93 (0.9,0.95)
FreeCell	7	Normalized Total Move Variability	Total number of moves made by a player	0.29 (0.24,0.34)	0.1 (0.03,0.17)	0.25 (0.15,0.35)	0.1 (−0.05,0.25)
	8	Normalized Valid Move Rate[†]	Rate of game-progressing moves out of total moves	0.52 (0.49,0.56)	0.66 (0.61,0.7)	0.82 (0.78,0.85)	0.85 (0.8,0.89)
Block Drop	9	Hard Drop Count[†]	Number of times the 'Hard Drop' button (i.e. to immediately place a falling piece) was pressed	0.87 (0.85,0.88)	0.92 (0.91,0.93)	0.94 (0.92,0.95)	0.91 (0.86,0.93)
	10	Rotation Count[†]	Number of times a player rotationally manipulated a falling piece	0.81 (0.79,0.83)	0.88 (0.86,0.89)	0.91 (0.89,0.93)	0.93 (0.9,0.95)

[†] Higher value indicates better performance.

2.5. Neuropsychological assessments

Participants were administered standardized neuropsychological test batteries at baseline and end of study (Cohort I = Month 3, Cohort II = Month 6). **Table S2** provides detailed descriptions of all assessments. Here we describe the following assessments in detail due to high correlation with game features: Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) [19], Trail Making Test (TMT) [20] Parts A (TMT-A) and B (TMT-B), and the Montreal Cognitive Assessment (MoCA, Range: 0–30) [21]. Index scores of the RBANS including Sum of Index Score (SID) Immediate Memory Index (IMI), Delayed Memory Index (DMI)), Attention Index (AI), Language Index (LI), Visuospatial/Constructional Index (VI)], completion times of both TMT-A and TMT-B and MoCA score totals were compared with 10 visit-level features across Memory Match [*Normalized Accuracy, Normalized Redundant Move Variability, Difficulty Level*], Word Scramble [*Word Repetition Rate, Word Repetition Rate Variability, Total Score*], FreeCell [*Normalized Valid Move Rate, Normalized Total Move Variability*], and Block Drop [*Hard Drop Count, Rotation Count*]. While there was substantial overlap of tests for both cohorts, the specific batteries intentionally differed to optimize sensitivity of measurement for those without cognitive complaints (NCC) in Cohort I versus those with cognitive impairments (SCD, MCI, MD) and to reduce burden for those with cognitive complaints and/or impairments in Cohort II.

2.6. Statistical analysis

To assess compliance, each participant's average weekly gameplay time was calculated over 3 months for Cohort I participants and over 6 months for Cohort II participants. The median, first quartile (Q1), third quartile (Q3), and the mean of these subject-level weekly gameplay times were then summarized by diagnostic cohort, or by study cohort.

Pearson correlation coefficient was used to evaluate the association between game features and established neuropsychological test scores and demographic variables. For sex, the point-biserial correlation was computed. To examine group differences across game features, the Wilcoxon rank sum test was conducted. To assess whether group differences remained significant after controlling for demographic variables, analysis of covariance (ANCOVA) was performed with age, sex, and education as covariates.

Test-retest reliability of game features was evaluated using intraclass correlation coefficients (ICCs), calculated with a two-way random effects model with absolute agreement per subject. Data were aggregated across four different time intervals: (1) every day (2) every three days (3) every week and (4) every two weeks. Each test-retest was then completed, comparing each time interval across the time frame of the study (e.g., (1) every day - Day 1v2, Day 2v3… Day 89v90).

Known sensitivity to change of game features was assessed utilizing mixed models for repeated measures (MMRMs) to analyze changes in game features from baseline to Month 3 and Month 6. The statistical model included age, sex, diagnosis group, visit, game feature values at baseline, and diagnosis group-by-visit interaction as fixed effects and visit as a repeated effect within each individual. All analyses were conducted using R (Version 4.3.2). Statistical significance was set at $p < .05$, without adjustment for multiple comparisons.

3. Results

3.1. Demographics

A total of sixty participants were included in Cohorts I and II: age mean = 71.57, SD = 8.04; sex 38 females (63.3%), 22 males (36.7%); years of education mean = 17.12, SD = 2.68; race White 57/60 (95%) and ethnicity non-Hispanic/Latino 53/60 (88.3%). Cohort I was ethnically homogenous with no participants identifying as Hispanic/Latino. Cohort II participants exhibited higher total years of education (mean = 17.51, SD = 3.24) compared to Cohort I (mean = 16.59, SD = 1.61). As expected for the MoCA, the CN group ($n = 37$) scored higher (mean = 28.16, SD = 1.59; median = 28, IQR = 27–29) than the CI group ($n = 23$) (mean = 22.39, SD = 3.03, median = 22, IQR = 21–24.5), a difference that was statistically significant by both Wilcoxon and *t*-test ($p < .001$). Detailed cohort statistics are summarized in Table 2.

3.2. Compliance

NCC participants (Cohort I) achieved a median compliance of 165 min per week [Q1, Q3 = 140, 252] (**Table S3**). Participants who completed Cohort II achieved a median compliance of 210 min per week [Q1, Q3 = 171, 387]. MCI participants achieved the highest compliance, with a median of 308 min/week [Q1, Q3 = 193, 524], followed by mild dementia participants playing 264 min/week [Q1, Q3 = 194, 553], and then SCD participants with 181n min/week [Q1, Q3 = 133, 210].

Table 2
Demographic data for Cohort I and II.

		Cohort I (n = 26)	Cohort II (n = 34)	Total (n = 60)
	General			
	Age in years [mean (SD)]	71.87 (8.63)	71.32 (7.70)	71.57 (8.04)
	Female [n (%)]	19 (73.08)	19 (55.88)	38 (63.33)
	Gender identity as female [n (%)]	20 (76.92)	19 (55.88)	39 (65)
	Years of education [mean (SD)]	16.59 (1.61)	17.51 (3.24)	17.12 (2.68)
	Race [n (%)]			
	White-Caucasian/ European	24 (92.31)	31 (91.18)	55 (91.67)
	White-Arabic/North African	1 (3.85)	1 (2.94)	2 (3.33)
	East Asian	-	1 (2.94)	1 (1.67)
	African American	1 (3.85)	1 (2.94)	2 (3.33)
	Ethnicity [n (%)]			
	Hispanic/Latino	0 (0)	2 (5.88)	2 (3.33)
	Non-Hispanic/Latino	26 (100)	27 (79.41)	53 (88.33)
	Unknown	-	5 (14.71)	5 (8.33)
	Diagnostic Groups [n (%)]			
CN	NCC	26 (100)	-	26 (43.33)
	SCD	-	11 (32.36)	11 (18.33)
CI	MCI	-	15 (44.11)	15 (25)
	MD	-	8 (23.53)	8 (13.33)

3.3. Correlation with standardized neuropsychological assessments and demographics at baseline

The main correlation results are summarized in Table 3. Memory Match (N = 34) '*Normalized Accuracy*' feature showed the highest correlations with RBANS SID [r = 0.76, p <.001], RBANS DMI [r = 0.75, p <.001], and MoCA [r = 0.69, p <.001]. The '*Normalized Redundant Move Variability*' feature showed high-to-moderate positive correlations with TMT-B [r = 0.73, p <.001], TMT-A [r = 0.67, p <.001] and negative correlation with RBANS DMI [r = −0.58, p <.001], RBANS SID [r = −0.56, p <.001], and MoCA [r = −0.51, p <.01]. '*Difficulty Level*' of Memory Match showed moderate to low positive/negative correlations with RBANS SID [r = 0.36, p <.05] and RBANS DMI [r = 0.40, p <.05].

Word Scramble (N = 34) '*Word Repetition Rate*' feature showed high correlations to RBANS DMI [r = −0.73, p <.001], SID [r = −0.59, p <.001], and MoCA [r = −0.73, p <.001]. '*Word Repetition Rate Variability*' showed a moderate correlation to both TMT-B [r = 0.55, p <.001] and MoCA [r = −0.53, p <.001]. '*Total Score*' displayed significantly high/moderate correlations with RBANS SID [r = 0.65, p <.001], DMI [r = 0.54, p < 0.001], IMI [r = 0.54, p < 0.001], LI [r = 0.50, p < 0.01], TMT-B [r = −0.60, p <.001], and MoCA [r = 0.61, p <.001].

Block Drop '*Hard Drop Count*' displayed high to moderate negative correlation with TMT-A [r = −0.71, p <.001] and TMT-B [r = −0.61, p <.001]. Similarly, '*Rotation Count*' displayed high to moderate correlations with TMT-B [r = −0.65, p <.001] and TMT-A [r = −0.49, p <.01].

FreeCell features displayed low correlation to most neuropsychological metrics, but RBANS DMI was moderately correlated with both '*Normalized Valid Move Rate*' [r = 0.57, p <.01] and '*Normalized Total Move Variability*' [r = −0.46, p <.05].

We also examined correlations between game features and demographic variables (Table 3). Age was negatively correlated with Block Drop '*Rotation Count*' (r = −0.65, p < .001) and Memory Match '*Normalized Accuracy*' (r = −0.41, p < .05), indicating that older participants showed slower processing speed and lower memory accuracy. Education was positively correlated with Word Scramble '*Total Score*' (r = 0.36, p < .05) and negatively with '*Word Repetition Rate Variability*' (r = −0.43, p < .05). Sex showed no significant correlations with any game features.

All other results of correlation testing for Memory Match (**Figure S3A**), Word Scramble (**Figure S3B**), FreeCell (**Figure S3C**) and Block Drop (**Figure S3D**) were assessed and presented as heat maps.

3.4. Test-Retest reliability

ICCs for game features derived from daily median values over various durations were evaluated to assess feature consistency. The

Table 3
Baseline correlation between select game features and standard clinical test scores and demographics. Comparisons for Cohort II only.

	Memory Match (N = 34)			Word Scramble (N = 34)			FreeCell (N = 30)		Block Drop (N = 27)	
Clinical Scores	Normalized Accuracy	Normalized Redundant Move Variability	Difficulty Level	Word Repetition Rate	Word Repetition Rate Variability	Total Score	Normalized Valid Move Rate	Normalized Total Move Variability	Hard Drop Count	Rotation Count
RBANS - Sum of Index Scores	0.76***	−0.56***	0.3	−0.59***	−0.48**	0.65***	0.44*	−0.41*	0.44*	0.44*
RBANS - Immediate Memory Index	0.60***	−0.43*	0.36*	−0.44**	−0.42*	0.54***	0.26	−0.24	0.29	0.38*
RBANS - Delayed Memory Index	0.75***	−0.58***	0.40*	−0.73***	−0.32	0.54***	0.57**	−0.46*	0.39*	0.3
RBANS - Attention Index	0.51**	−0.32	−0.03	−0.26	−0.41*	0.46**	0.23	−0.33	0.15	0.32
RBANS - Language Index	0.38*	−0.39*	0.27	−0.22	−0.22	0.50**	0.11	−0.13	0.31	0.17
RBANS - Visuospatial/ Constructional Index	0.44**	−0.29	0.04	−0.38*	−0.38*	0.33	0.25	−0.17	0.49**	0.41*
Trail.Making.A	−0.58***	0.67***	0.01	0.06	0.44**	−0.37*	−0.18	0.33	−0.61***	−0.49**
Trail.Making.B	−0.65***	0.73***	−0.06	0.36*	0.55***	−0.60***	−0.40*	0.50**	−0.71***	−0.65***
MoCA	0.69***	−0.51**	0.28	−0.73***	−0.53***	0.61***	0.46*	−0.29	0.41*	0.39
Demographics										
Age	−0.41*	0.32	−0.06	0.31	0.16	−0.16	−0.15	0.25	−0.36	−0.65***
Years of Education	0.2	−0.07	0.04	−0.22	−0.43*	0.36*	0.15	−0.22	0.2	0.27
Sex	−0.03	0.1	−0.21	−0.04	−0.31	0	−0.31	−0.03	0.18	0.31

*** p < 0.001, ** p < 0.01, * p < 0.05.

findings indicate good to excellent reliability (ICCs = 0.85–0.95) across all four games when using two weeks of data per participant (Table 1). *'Normalized Total Move Variability'* of FreeCell showed greater variability in ICC testing and was less consistent than other game features.

3.5. Known groups comparison

All game features at baseline (after acclimatization gameplay during screening) significantly differentiated the CN group from the CI group [p <.012], except for the '*Difficulty Level*' within Memory Match [$p = .114$] (Fig. 1, **Table S4**). These group differences remained significant after adjusting for age, sex, and education using ANCOVA (all p < .01 except '*Difficulty Level*'), indicating that the discriminative ability of game features is independent of demographic differences between groups (Table S4). CI participants had worse median game performance such as lower '*Normalized Accuracy*' for Memory Match [CI = −1.25 (−1.83,−0.48), CN = −0.17 (−0.52,0.16); p <.001] higher median '*Word Repetition Rate*' for Word Scramble [CI = 0.17 (0.1,0.23), CN = 0.09 (0.07,0.11); $p = .002$], lower median '*Normalized Valid Move Rate*' for FreeCell [CI = −1.06 (−1.42,−0.57), CN = −0.01 (−0.64,0.45); p <.001] and lower median '*Hard Drop Count*' for Block Drop [CI = 27 (10.25,37), CN = 37 (−29,46); $p = .012$] compared to CN participants.

When examining subgroups, we observed that the Memory Match feature *Normalized Accuracy* improved in the NCC group at Month 3, but not in the SCD, MCI, or MD groups. Additionally, a significant difference was detected between the NCC and MCI groups in change from baseline (T-ratio = 3.07, $P = .003$, Fig. 2A). For '*Normalized Redundant Move Variability*', there was a statistically significant difference between the NCC and MD groups in change from baseline to 3 months (T-ratio = −2.07, $P = .043$, Fig. 2B). Regarding the self-selected '*Difficulty Level*' from Memory Match, NCC participants tended to choose increased difficulty levels at Month 3. There was a significant difference in change from baseline to 3 months between the NCC and MD groups (T-ratio = 2.57, $P = .013$, Fig. 2C).

For Word Scramble's '*Word Repetition Rate*', there was no difference in changes from baseline between the NCC and other groups. However, there was a significant within-group decrease in '*Word Repetition Rate*' at month 3 for the NCC and SCD groups (NCC LSmean = −0.012, SCD LSmean = −0.018, Fig. 2D). An opposite trend was observed for the MCI and MD groups at Month 3 and Month 6. The variation in '*Word Repetition Rate*' measured by '*Word Repetition Rate Variability*' also significantly decreased at Month 3 for the NCC and SCD groups (NCC LSmean = −0.013, SCD LSmean = −0.013, Fig. 2E). The difference in changes from baseline for this feature between the NCC and MCI/MD groups was statistically significant (NCC vs. MCI, T-ratio = −2.75, $P = .008$; NCC vs. MD, T-ratio = −4.35, P < .001, Fig. 2E).

We did not observe a significant difference in changes from baseline for the feature '*Total Score*' between the NCC and other groups, although all groups showed a significant within-group increase in this feature at Month 3 (Fig. 2F).

4. Discussion

The current study investigated the psychometric properties of MIND GamePack©, a platform that offers semi-structured gameplay to provide higher frequency, more sensitive data of cognitive and psychomotor functioning over time. Participants in all groups demonstrated high compliance with the MIND GamePack© in Cohort I and II studies, well beyond what was required (165% for Cohort I; 210% for Cohort II), suggesting maintained interest, motivation, and enjoyment. These results are highly reassuring as general compliance rates for other remote unsupervised technologies often yield significantly lower compliance rates (63 to 94%), depending on timescale length [22,23]. In a trial of similar length to Cohort II implementing scheduled self-assessments (six months), participants did adhere to thrice-weekly self-administration but were more likely to do so when compliance requirements were less stringent, strengthening the case for unsupervised use [24]. Although adherence to DHTs remains a significant challenge, gamification of DHTs has demonstrated utility in sustaining engagement and making e-Health activities both entertaining and engaging [25,26].

The MIND GamePack© demonstrated good to excellent test-retest reliability for all game features selected [ICC = 0.85 – 0.95] with two weeks interval, but lower when using less than a week interval [ICC = 0.52 – 0.87]. The lower ICCs at the one-day interval likely reflect both measurement noise and real day-to-day changes in the feature scores. Some variability is expected due to random fluctuations, but our previous work showed that self-reported cognitive ability is associated with daily gameplay behaviors suggesting that these short-term changes can reflect meaningful psychological differences [15]. In contrast, higher ICCs at longer intervals, such as two weeks, suggest greater stability over time as the short-term fluctuations average out. This highlights the importance of data granularity, as daily versus aggregated measures may be more suitable depending on whether the goal is diagnosing or continuous monitoring. We also observed that some features reached higher ICCs after one week, while others required two weeks of data to achieve similar levels of reliability. These differences reflect distinct characteristics of each gameplay task and highlight the importance of considering such variability when selecting features for use in digital assessments.

Longitudinal change of game features as measured between Baseline (Day 0) and Month 6 displayed unique trajectories between NCC, SCD, MCI, and MD groups. Notably, game features associated with memory showed significantly greater improvements in NCC participants compared to MCI and MD groups, consistent with expected learning/practice effects, while CI groups were flatter. Likewise, '*Normalized Redundant Move Variability*' [Memory Match], '*Word Repetition Rate*', '*Word Repetition Rate Variability*', and *Total Score* [Word Scramble] displayed similar differential trajectories between NCC and MD groups.

All features examined within the research displayed good to excellent differentiation between CN and CI metagroups. Importantly, these group differences were robust to demographic adjustment, suggesting that game performance reflects cognitive status rather than confounding factors such as age or education level. We further found '*Word Repetition Rate*', and '*Word Repetition Rate Variability*' features displayed the strongest differential trajectories. Intriguingly, '*Difficulty Level*' was sensitive at differentiating NCC from SCD, MCI, and MD participants. NCC participants may select more difficult boards of Memory Match, while those with cognitive deficiency (SCD, MCI, MD) tended to choose easier levels – potentially due to an awareness of their cognitive limitations or a desire to avoid potential frustration with more difficult boards.

Correlation of game features with traditional neuropsychological tests yielded significant results within the memory and learning domain (s); indices of the RBANS and total scores of the TMT showed the highest correlation with selected game features from Memory Match. Similar research investigating construct validity of a structured remote web-assessment (BRANCH) compared to gold-standard neuropsychological tests such as the MMSE (Mini Mental State Examination) and QDRS (Quick Dementia Rating System) yielded moderate Pearson coefficients [$r = 0.53$ (p <.001) and r = −0.40 (p<.001)] respectively while other mobile assessments (MyCog Mobile) displayed battery-wide construct validity [27,28]. Overall, DHTs are improving in their ability to validate against traditional pen-and-paper tests and future research may explore these correlations within different populations and employ expanded neuropsychological batteries.

Several limitations should be acknowledged for both the MIND GamePack© platform and the study herein. Small sample ($N = 60$) and subgroup sizes may minimize correlation with neuropsychological tests. Both Cohort I and Cohort II predominantly included educated, white, non-Hispanic participants. Therefore, caution must be exercised when generalizing these findings to more diverse populations. Additionally, the differing study designs of Cohort I and Cohort II present challenges

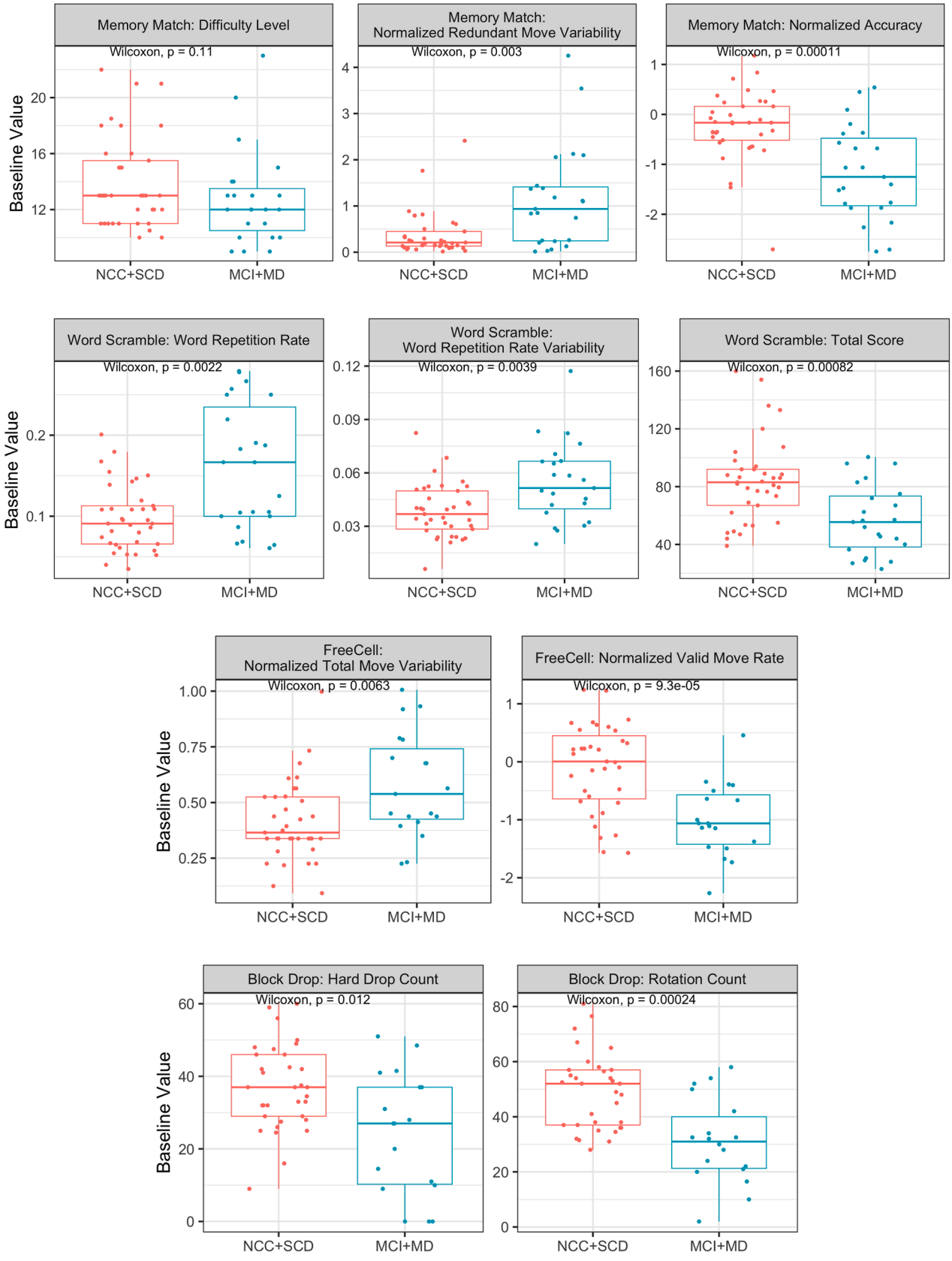
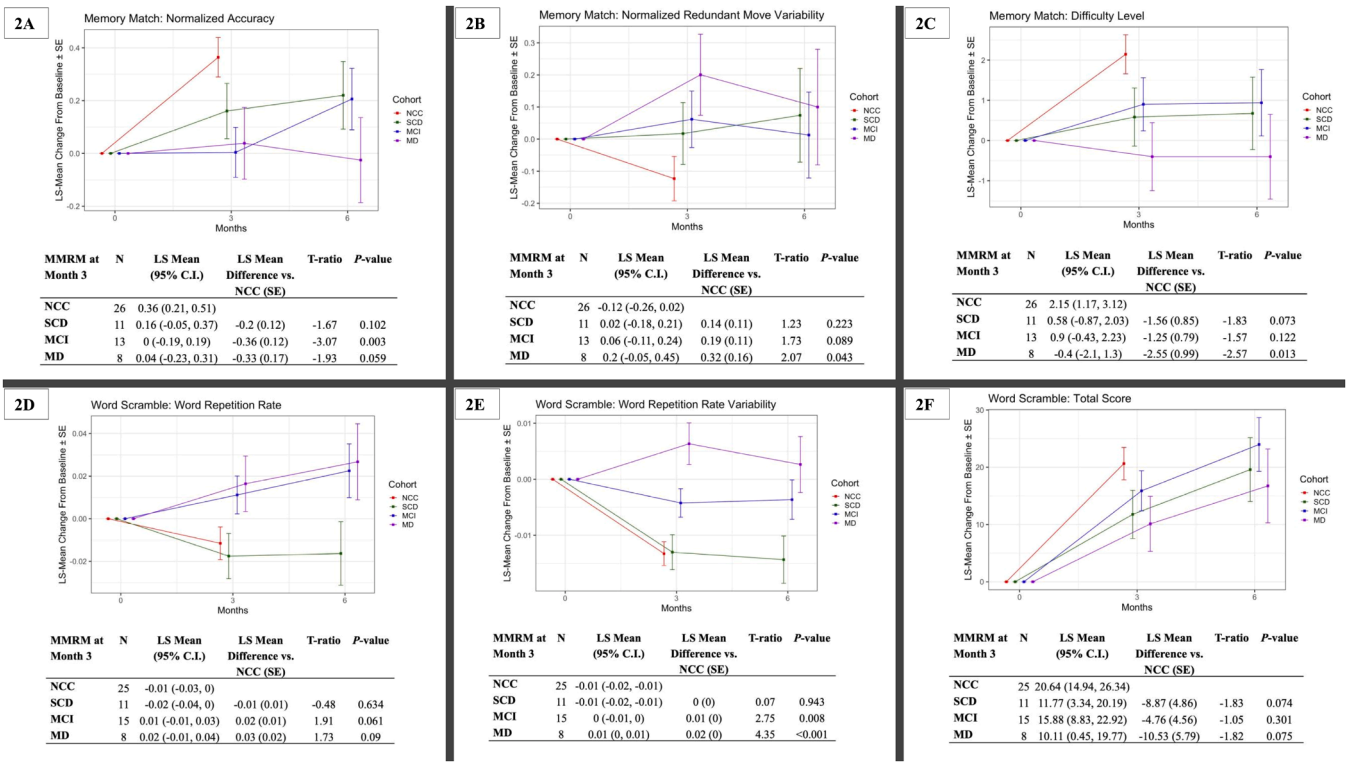


Fig. 1. Box plots for game features collected at baseline from four games by cognitive groups (NCC + SCD vs MCI + MD).



MMRM at Month 3	N	LS Mean (95% C.I.)	LS Mean Difference vs. NCC (SE)	T-ratio	*P*-value
NCC	26	0.36 (0.21, 0.51)			
SCD	11	0.16 (-0.05, 0.37)	-0.2 (0.12)	-1.67	0.102
MCI	13	0 (-0.19, 0.19)	-0.36 (0.12)	-3.07	0.003
MD	8	0.04 (-0.23, 0.31)	-0.33 (0.17)	-1.93	0.059

MMRM at Month 3	N	LS Mean (95% C.I.)	LS Mean Difference vs. NCC (SE)	T-ratio	*P*-value
NCC	26	-0.12 (-0.26, 0.02)			
SCD	11	0.02 (-0.18, 0.21)	0.14 (0.11)	1.23	0.223
MCI	13	0.06 (-0.11, 0.24)	0.19 (0.11)	1.73	0.089
MD	8	0.2 (-0.05, 0.45)	0.32 (0.16)	2.07	0.043

MMRM at Month 3	N	LS Mean (95% C.I.)	LS Mean Difference vs. NCC (SE)	T-ratio	*P*-value
NCC	26	2.15 (1.17, 3.12)			
SCD	11	0.58 (-0.87, 2.03)	-1.56 (0.85)	-1.83	0.073
MCI	13	0.9 (-0.43, 2.23)	-1.25 (0.79)	-1.57	0.122
MD	8	-0.4 (-2.1, 1.3)	-2.55 (0.99)	-2.57	0.013

MMRM at Month 3	N	LS Mean (95% C.I.)	LS Mean Difference vs. NCC (SE)	T-ratio	*P*-value
NCC	25	-0.01 (-0.03, 0)			
SCD	11	-0.02 (-0.04, 0)	-0.01 (0.01)	-0.48	0.634
MCI	15	0.01 (-0.01, 0.03)	0.02 (0.01)	1.91	0.061
MD	8	0.02 (-0.01, 0.04)	0.03 (0.02)	1.73	0.09

MMRM at Month 3	N	LS Mean (95% C.I.)	LS Mean Difference vs. NCC (SE)	T-ratio	*P*-value
NCC	25	-0.01 (-0.02, -0.01)			
SCD	11	-0.01 (-0.02, -0.01)	0 (0)	0.07	0.943
MCI	15	0 (-0.01, 0)	0.01 (0)	2.75	0.008
MD	8	0.01 (0, 0.01)	0.02 (0)	4.35	<0.001

MMRM at Month 3	N	LS Mean (95% C.I.)	LS Mean Difference vs. NCC (SE)	T-ratio	*P*-value
NCC	25	20.64 (14.94, 26.34)			
SCD	11	11.77 (3.34, 20.19)	-8.87 (4.86)	-1.83	0.074
MCI	15	15.88 (8.83, 22.92)	-4.76 (4.56)	-1.05	0.301
MD	8	10.11 (0.45, 19.77)	-10.53 (5.79)	-1.82	0.075

Fig. 2. Change from baseline in digital features estimated by MMRM. Modeled change from baseline means (± SE) at Month 3 and Month 6 were shown in plots for game features from Memory Match and Word Scramble games. Model outputs for Month 3 for LS means and mean difference from the NCC group were summarized in the table below each diagram.

for data harmonization, such as varying study timelines and distinct neuropsychological assessments. These differences hinder direct comparisons between the cohorts.

5. Conclusions

Future research should expand sample size, diversity, and diagnostic status and may also evaluate the performance of MIND GamePack© amongst individuals with various neurocognitive disorders beyond neurodegenerative dementias. In particular, the implementation of DHTs presents unique challenges for important non-English speaking populations, including accurate translation, cultural familiarity with stimuli and differences in digital accessibility [29]. Although assets of Memory Match, Block Drop, and FreeCell were selected to be culturally neutral, Word Scramble remains nativized for English-speaking users. Future work plans to expand the MIND GamePack© for deployment in other cultures; with initial plans to translate and validate first in Spanish speaking populations. The MIND GamePack© could provide researchers with high-density data on other neurological (e.g. static vascular, epilepsy, post-traumatic or other brain injuries) and psychiatric (e.g. mood, anxiety, psychotic, neurodevelopmental) disorders. The MIND GamePack© may be useful for researchers conducting drug or other interventional trials. The frequent and detailed gameplay measurements taken over time, both before and after drug initiation, may enable researchers to better understand cognitive changes within short or extended periods. In such trials, the MIND GamePack© could provide supplemental or complementary high-resolution data alongside more conventional outcome measures, while providing a longitudinal trial experience which encourages high rates of compliance. Lastly, current features and future updates to the MIND GamePack© platform may also show usefulness for clinical monitoring, both as a screening tool for change and as an ancillary tool clinicians can gather personal supplemental cognitive data from. The ease of deployment, low-cost of scalability, and high compliance by users across groups of varying levels of cognitive decline make the MIND GamePack© an encouraging platform for a versatile array of research applications.

Ethics statement

Study approval was obtained from the governing Institutional Review Board at Massachusetts General Brigham (MGB) (2021P000144 and 2021P003132). All participants provided informed consent.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used generative AI [Gemini 3.1 Pro] to perform minor copyediting of early drafts. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Data sharing statement

Datasets and associated code repositories associated with data analysis are available from the corresponding author upon reasonable request. Code, algorithms, and assets associated with the MIND GamePack© platform, infrastructure, and apparatus are considered proprietary and may not be shared upon request.

Funding sources

The MIND GamePack© was supported in part with private funding from both AbbVie, Inc and The Challenger Foundation.

CRediT authorship contribution statement

Jake A. Galler: Conceptualization, Methodology, Validation, Software, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Liuqing Yang:** Methodology, Validation, Software, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Chao-Yi Wu:** Methodology, Validation, Software, Formal analysis, Data curation, Writing – review & editing, Visualization. **Cathrine Young:** Writing – review & editing, Supervision, Project administration. **Edmarie Guzmán-Vélez:** Writing – review & editing, Supervision. **Katherine W. Cropp:** Investigation, Writing – review & editing. **Anthony W. Bannon:** Resources, Supervision, Project administration, Funding acquisition. **Hiroko H. Dodge:** Writing – review & editing, Supervision. **Jessica A. Gerber:** Supervision, Project administration, Funding acquisition. **Steven E. Arnold:** Conceptualization, Methodology, Investigation, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

J.A.Galler., C.Y.W., C.Y., K.W.C., H.H.D., and J.A.Gerber are employees of The General Hospital Corporation d/b/a Massachusetts General Hospital and declare no interests associated with the work detailed herein. E.G.V. is an employee of Robert Wood Johnson Medical School, operating as part of Rutgers University, and declares no conflicts of interest associated with the work detailed herein. L.Y. and A.W.B. are both employees of, and hold stock and/or options in, AbbVie, Inc. S.E.A. received funding in part from AbbVie Inc for this work. S.E.A. has contracted with/received grants from AC Immune, Alzheimer's Association, Athira, Challenger Foundation, Chromadex, Eli Lilly and Company, Fortrea, Gatehouse Bio, Janssen Pharmaceuticals, John Sperling Foundation, National Institutes of Health (NIH), Novartis AG, Seer Biosciences, Superfluid Dx, Inc., and Venture Well. S.E.A. has received consulting fees from Allyx Therapeutics, BioVie, Inc., Bob's Last Marathon, Merk, Jocasta Neuroscience, Sage Therapeutics, Sanofi, and Vandria. S.E.A. has received expert testimony fees from Foster & Eldredge and ProSelect Insurance Co.

Acknowledgements

Thank you to our research participants – collaborators in our work whose feedback and zeal have been instrumental in advancing the development of the MIND GamePack©. Our team would also like to thank both the departments of Digital Science and Precision Medicine-Neuroscience at AbbVie, Inc. and the Challenger Foundation for their support in this research.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.tjpad.2026.100642.

References

[1] Elhage A, Cohen S, Cummings J, et al. Defining benefit: clinically and biologically meaningful outcomes in the next-generation Alzheimer's disease clinical care pathway. Alzheimers Dement 2025;21:alz.14425. https://doi.org/10.1002/alz.14425. Publ online.

[2] Wheatley SH, Mohanty R, Poulakis K, et al. Divergent neurodegenerative patterns: comparison of [18F] fluorodeoxyglucose-PET- and MRI-based Alzheimer's disease subtypes. Brain Commun 2024;6(6):fcae426. https://doi.org/10.1093/braincomms/fcae426.

[3] Vermunt L, Sikkes SAM, Van Den Hout A, et al. Duration of preclinical, prodromal, and dementia stages of Alzheimer's disease in relation to age, sex, and APOE genotype. Alzheimers Dement 2019;15(7):888–98. https://doi.org/10.1016/j.jalz.2019.04.001.

[4] Nguyen TTT, Lee HH, Huang LK, et al. Heterogeneity of Alzheimer's disease identified by neuropsychological test profiling. Ginsberg sd, 18. PLOS ONE; 2023, e0292527. https://doi.org/10.1371/journal.pone.0292527. editor.

[5] Wilson RS, Segawa E, Boyle PA, Anagnos SE, Hizel LP, Bennett DA. The natural history of cognitive decline in Alzheimer's disease. Psychol Aging 2012;27(4):1008–17. https://doi.org/10.1037/a0029857.

[6] Gold M, Amatniek J, Carrillo MC, et al. Digital technologies as biomarkers, clinical outcomes assessment, and recruitment tools in Alzheimer's disease clinical trials. Alzheimers Dement Transl Res Clin Interv 2018;4(1):234–42. https://doi.org/10.1016/j.trci.2018.04.003.

[7] Kourtis LC, Regele OB, Wright JM, Jones GB. Digital biomarkers for Alzheimer's disease: the mobile/wearable devices opportunity. Npj Digit Med 2019;2(1):9. https://doi.org/10.1038/s41746-019-0084-2.

[8] Cheng Y, Ho E, Weintraub S, et al. Predicting brain amyloid status using the national institute of health toolbox (NIHTB) for assessment of neurological and behavioral function. J Prev Alzheimers Dis 2024;11:943–57. https://doi.org/10.14283/jpad.2024.77. Publ online.

[9] Jutten RJ, Rentz DM, Fu JF, et al. Monthly At-home computerized cognitive testing to detect diminished practice effects in preclinical Alzheimer's disease. Front Aging Neurosci 2022;13:800126. https://doi.org/10.3389/fnagi.2021.800126.

[10] Berron D, Olsson E, Andersson F, et al. Remote and unsupervised digital memory assessments can reliably detect cognitive impairment in Alzheimer's disease. Alzheimers Dement 2024;20(7):4775–91. https://doi.org/10.1002/alz.13919.

[11] Nicosia J, Aschenbrenner AJ, Balota DA, et al. Unsupervised high-frequency smartphone-based cognitive assessments are reliable, valid, and feasible in older adults at risk for Alzheimer's disease. J Int Neuropsychol Soc 2023;29(5):459–71. https://doi.org/10.1017/S135561772200042X.

[12] Welhaf MS, Wilks H, Aschenbrenner AJ, et al. Naturalistic assessment of reaction time variability in older adults at risk for Alzheimer's disease. J Int Neuropsychol Soc 2024;30(5):428–38. https://doi.org/10.1017/S1355617723011475.

[13] Lai JM, Hawkins KA, Gross CP, Karlawish JH. Self-reported distress after cognitive testing in patients with Alzheimer's disease. J Gerontol Biol Sci Med Sci 2008;63(8):855–9. https://doi.org/10.1093/gerona/63.8.855.

[14] Wu C, Beattie Z, Mattek N, Sharma N, Kaye J, Dodge HH. Reproducibility and replicability of high-frequency, in-home digital biomarkers in reducing sample sizes for clinical trials. Alzheimers Dement Transl Res Clin Interv 2021;7(1):e12220. https://doi.org/10.1002/trc2.12220.

[15] Schwab N, Wu CY, Galler J, et al. Feasibility of common, enjoyable game play for assessing daily cognitive functioning in older adults. Front Neurol 2023;14:1258216. https://doi.org/10.3389/fneur.2023.1258216.

[16] Li C, Neugroschl J, Luo X, et al. The utility of the cognitive function instrument (CFI) to detect cognitive decline in non-demented older adults. J Alzheimer 19s Dis 2017;60(2):427–37. https://doi.org/10.3233/JAD-161294.

[17] Brandt J, Spencer M, Folstein M. The Telephone Interview for cognitive status. Cogn Behav Neurol 1988;1(2).

[18] Berry JH. Zone In not out! the key to winning high-level tetris. Percept Mot Ski 2024;131(6):2304–23. https://doi.org/10.1177/00315125241289687.

[19] Thaler NS, Hill BD, Duff K, Mold J, Scott JG. Repeatable battery for the assessment of neuropsychological Status (RBANS) intraindividual variability in older adults: associations with disease and mortality. J Clin Exp Neuropsychol 2015;37(6):622–9. https://doi.org/10.1080/13803395.2015.1039962.

[20] Reitan RM. Validity of the trail making test as an indicator of organic brain damage. Percept Mot Ski 1958;8(7):271. https://doi.org/10.2466/PMS.8.7.271-276.

[21] Nasreddine ZS, Phillips NA, Bédirian V, et al. The montreal cognitive assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005;53(4):695–9. https://doi.org/10.1111/j.1532-5415.2005.53221.x.

[22] Chen L, Zhen W, Peng D. Research on digital tool in cognitive assessment: a bibliometric analysis. Front Psychiatry 2023;14:1227261. https://doi.org/10.3389/fpsyt.2023.1227261. Published 2023 Aug 23.

[23] Polk SE, Öhman F, Hassenstab J, et al. A scoping review of remote and unsupervised digital cognitive assessments in preclinical Alzheimer's disease. Npj Digit Med 2025;8(1):266. https://doi.org/10.1038/s41746-025-01583-5.

[24] Christianson K, Prabhu M, Popp ZT, et al. Adherence timescale impacts completion rates of high-frequency mobile cognitive assessments among older adults. Explor Med 2025;6. https://doi.org/10.37349/emed.2025.1001356. 10.37349/emed.2025.1001356.

[25] Figueiredo T, Alcântara L, Carrilho J, Paúl C, Costa E. Understanding adherence to digital health technologies: systematic review of predictive factors. J Med Internet Res 2025;27:e77362. https://doi.org/10.2196/77362. Published 2025 Nov 17.

[26] Sardi L, Idri A. Fernández-Alemán JL. A systematic review of gamification in e-Health. J Biomed Inf 2017;71:31–48. https://doi.org/10.1016/j.jbi.2017.05.011.

[27] Weizenbaum EL, Hsieh S, Molinare C, et al. Validation of the multi-day Boston remote assessment of neurocognitive health (BRANCH) among cognitively impaired & unimpaired older adults. J Prev Alzheimers Dis 2025;12(3):100057. https://doi.org/10.1016/j.tjpad.2025.100057.

[28] Young SR, Dworak EM, Byrne GJ, et al. MyCog Mobile smartphone-based cognitive screening system: a cross-sectional construct validation in a general population sample at multiple research facilities in the USA. BMJ Open 2025;15(1):e092063. https://doi.org/10.1136/bmjopen-2024-092063.