Performance of Digital Cognitive Assessment in Predicting Dementia Stages Delineated by Dementia Severity Rating Scale: Retrospective Study

Duong Huynh, Kevin Sun, Mary Patterson, Reza Hosseini Ghomi, Bin Huang

Submitted to: JMIR Aging
on: August 13, 2024

Table of Contents

Performance of Digital Cognitive Assessment in Predicting Dementia Stages Delineated by Dementia Severity Rating Scale: Retrospective Study

Duong Huynh[1] PhD; Kevin Sun[1]; Mary Patterson[1]; Reza Hosseini Ghomi[1, 2, 3] MD, MSE; Bin Huang[1] PhD

[1]BrainCheck, Inc. Austin US
[2]Frontier Psychiatry, PLLC Billings US
[3]Faculty, Department of Neurology Institute for Neuroengineering, & eScience Institute University of Washington Seattle US

Corresponding Author:
Bin Huang PhD
BrainCheck, Inc.
3000 E Cesar Chavez St 300
Austin
US

Abstract

Background: Dementia is characterized by impairments in an individual's cognitive and functional abilities. Digital cognitive assessments have been shown to be effective in detecting mild cognitive impairment and dementia but whether they can stage the disease remains to be studied.

Objective: In this study, we examined: (1) the correlation between scores obtained from BC-Assess, a digital cognitive assessment, and scores obtained from the Dementia Severity Rating Scale (DSRS), and (2) the accuracy of using the BC-Assess score to predict dementia stage delineated by the DSRS score. We also explored whether BC-Assess can be combined with information from the Katz Index of Independence in Activities of Daily Living (ADL) to obtain enhanced accuracy.

Methods: Retrospective analysis was performed on a BrainCheck dataset containing 1,751 dementia patients with different cognitive and functional assessments completed for cognitive care planning, including the DSRS, the ADL, and the BC-Assess. The patients were staged according to their DSRS total score (DSRS-TS): 982 mild (DSRS-TS: 10-18), 656 moderate (19-26), and 113 severe (37-54) patients. Pearson correlation was used to assess the associations between BC-Assess overall score (BC-OS), ADL total score (ADL-TS), and DSRS-TS. Logistic regression was used to evaluate the possibility of using patients' BC-OS and ADL-TS to predict their stage.

Results: We find moderate Pearson correlations between DSRS-TS and BC-OS (r = -.53), between DSRS-TS and ADL-TS (r = -.55), and a weak correlation between BC-OS and ADL-TS (r = .37). Both BC-OS and ADL-TS significantly decrease with increasing severity. BC-OS demonstrates to be a good predictor of dementia stages, with area under the ROC curve (ROC-AUC) of classification using logistic regression ranging from .733 to .917. When BC-Assess is combined with ADL, higher prediction accuracies are achieved, with ROC-AUC ranging from .786 to .961.

Conclusions: Our results suggest that BC-Assess could serve as an effective alternative tool to DSRS for grading dementia severity, particularly in cases where DSRS, or other global assessments, may be challenging to obtain due to logistical and time constraints.

(JMIR Preprints 13/08/2024:65292)
DOI: https://doi.org/10.2196/preprints.65292

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✔ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Performance of Digital Cognitive Assessment in Predicting Dementia Stages Delineated by Dementia Severity Rating Scale: Retrospective Study

Duong Huynh, PhD[1], Kevin Sun[1], Mary Patterson[1], Reza Hosseini Ghomi, MD, MSE[1,2,3], Bin Huang, PhD[1*]

[1] BrainCheck Inc., Austin, United States
[2] Frontier Psychiatry, PLLC, Billings, United States
[3] Faculty, Department of Neurology, Institute for Neuroengineering, & eScience Institute, University of Washington

*** Address correspondence to:** Bin Huang, PhD. Email: bin@braincheck.com

Abstract

Background

Dementia is characterized by impairments in an individual's cognitive and functional abilities. Digital cognitive assessments have been shown to be effective in detecting mild cognitive impairment and dementia but whether they can stage the disease remains to be studied.

Objective

In this study, we examined: (1) the correlation between scores obtained from BC-Assess, a digital cognitive assessment, and scores obtained from the Dementia Severity Rating Scale (DSRS), and (2) the accuracy of using the BC-Assess score to predict dementia stage delineated by the DSRS score. We also explored whether BC-Assess can be combined with information from the Katz Index of Independence in Activities of Daily Living (ADL) to obtain enhanced accuracy.

Methods

Retrospective analysis was performed on a BrainCheck dataset containing 1,751 dementia patients with different cognitive and functional assessments completed for cognitive care planning, including the DSRS, the ADL, and the BC-Assess. The patients were staged according to their DSRS total score (DSRS-TS): 982 mild (DSRS-TS: 10-18), 656 moderate (19-26), and 113 severe (37-54) patients. Pearson correlation was used to assess the associations between BC-Assess overall score (BC-OS), ADL total score (ADL-TS), and DSRS-TS. Logistic regression was used to evaluate the possibility of using patients' BC-OS and ADL-TS to predict their stage.

Results

We find moderate Pearson correlations between DSRS-TS and BC-OS (r = -.53), between DSRS-TS and ADL-TS (r = -.55), and a weak correlation between BC-OS and ADL-TS (r = .37). Both BC-OS and ADL-TS significantly decrease with increasing severity. BC-OS demonstrates to be a good predictor of dementia stages, with area under the ROC curve (ROC-AUC) of classification using logistic regression ranging from .733 to .917. When BC-Assess is combined with ADL, higher prediction accuracies are achieved, with ROC-AUC ranging from .786 to .961.

Conclusions

Our results suggest that BC-Assess could serve as an effective alternative tool to DSRS for grading

dementia severity, particularly in cases where DSRS, or other global assessments, may be challenging to obtain due to logistical and time constraints.

Keywords

Dementia stages, cognitive impairment, functional activities, cognitive assessment, BrainCheck

Introduction

Dementia is characterized by impairments in an individual's cognitive and everyday functional abilities. Although the pattern and advancement of these impairments vary among patients, the disease is usually considered as having three main stages: mild (early-stage), moderate (middle-stage), and severe (late-stage) [1]. The distinction between these categories lies in the extent to which a patient's physical, cognitive, and psychosocial well-being deteriorates. The rate of deterioration widely varies and progresses through subtle changes in daily functioning to complete loss of independence and the need for a caregiver [1]. Staging of dementia has important implications for practical decision making and research [2]. From the practical standpoint, knowledge of disease severity is helpful for selection of appropriate intervention options, for prognosis and communication with patients and their family about expectations, care needs, as well as early planning for the future [3–5]. From the research standpoint, staging patients is needed to determine their eligibility for participation, to achieve better clinical efficacy, and to obtain homogeneous sampling in research studies [4], particularly clinical trials.

As dementia is not a monolithic disease, both cognitive and functional abilities need to be measured to accurately assess its severity and progression. Standardized cognitive tests can provide objective measures of cognitive functioning in different domains such as memory and executive function. Brief cognitive screening tests such as the Montreal Cognitive Assessment (MoCA) and the Mini-Mental State Examination (MMSE), are widely used in clinical practices, but they are rarely used to grade dementia severity [6,7]. Formal neuropsychological tests provide a more comprehensive evaluation of cognitive functioning to support differential diagnoses [8,9]. However, these tests typically take hours and require administration by specialists. Digital cognitive assessments are emerging as an efficient solution due to their self-administration capability, remote accessibility, and automated scoring. Although these types of assessments have been shown to be effective in detecting mild cognitive impairment and dementia [10–12], their ability to stage dementia is not yet clear. Functional assessments quantify the ability to perform activities of daily living through questionnaires such as the Katz Index of Independence in Activities of Daily Living (ADL) [13,14]. Functional assessments are valuable for helping to evaluate dementia severity and also for providing proper guidance to patients and their caregivers.

In clinical practice, both cognitive and functional deficits can be measured using global staging scales. These scales typically come in the form of a questionnaire or interview, relying on subjective judgments and reports from patients or their knowledgeable informants. Compared with other scales, such as the Global Deterioration Scale (GDS) [15] or the Functional Assessment Staging (FAST) [16], the Clinical

Dementia Rating (CDR) [17–19] scale appears to be the most well-studied and best-evidenced for dementia staging [20,21]. However, the use of this instrument is often impractical in many situations where time and cost are concerns, due to its semi-structured nature, long administration time (30 - 60 minutes), and requirement of clinical judgment from a trained professional during administration and scoring [18,22]. In response to this, brief instruments have been developed to mirror the CDR, including the Dementia Severity Rating Scale (DSRS) [23]. The DSRS uses a multiple-choice format in which the caregiver rates the patient's cognitive and functional ability in 12 categories [23,24]. The DSRS has been shown to be effective in staging and determining the progression of dementia [23,25,26].

While cognitive tests have been shown to correlate with the above dementia staging tools [4,27,28], previous research primarily focused on traditional paper-based cognitive tests. The increasing adoption of digital solutions and tools in healthcare calls for the re-evaluation of dementia staging tools and digital cognitive assessments. The first goal of this study was to examine the correlation between scores obtained from BC-Assess, a digital cognitive assessment, and scores obtained from the DSRS, a global staging scale. The second goal of this study is to evaluate the accuracy of using the BC-Assess score to predict dementia stage delineated by the DSRS score. We also explored whether BC-Assess can be combined with information from ADL to obtain enhanced accuracy.

Methods

Data source

This retrospective study analyzed a real-world dataset of patients who received cognitive care planning services from their providers through BrainCheck Plan. These patients and their caregivers had completed a series of assessments including DSRS, ADL, and BC-Assess, and received a comprehensive and personalized cognitive care plan. Inclusion criteria for this study were: (1) patients 60 years of age or older; (2) assessments of cognitive care planning completed in English on an iPad; and (3) DSRS-TS $\geq$ 10. The criterion for DSRS-TS was to only include patients that were rated by DSRS to have mild, moderate and severe dementia [29].

Patients that showed evidence of moderate to severe depression, defined by a Geriatric Depression Scale score of 9 or above [30], were excluded. This is to avoid including reversible causes of dementia which may have poorly-correlated impact on cognitive and functional measures. Given depression is common among dementia patients [31–33], patients with mild depression were not excluded to avoid overfiltering of the data. For patients having multiple care plans, only the latest record was considered.

To reduce the impact of practice effect, only data from providers who had completed ≥20 cognitive care planning services for their patients were included. In total, data from 1,751 cognitive care planning services conducted between 02/2022 and 04/2024 across 48 providers were included for this analysis.

Measurements

The DSRS is a brief informant-based questionnaire made up of 12-items that measure functional abilities including memory, orientation to time/place, decision making, social interaction, home activities, personal care, eating, toileting, mobility, recognition, and speech and language. DSRS-TS is calculated by adding up scores across 12 functional areas, ranging from zero (no impairment) to 54 (extreme impairment) [25]. The patients could be categorized into three groups of different dementia severity levels based on their DSRS-TS. The majority 56% (N=982) were mild-stage patients (DSRS-TS 10–18), moderate-stage patients (DSRS-TS 19–26) accounted for 37.5% (N=656), and the remaining 6.5% (N=113) were severe-stage patients (DSRS-TS 37–54). The three severity levels serve as class labels in a logistic regression analysis in this study to predict patients' conditions based on their BC-Assess and ADL data.

The BC-Assess consisted of six individual cognitive assessments: Immediate Recognition and Delayed Recognition (memory), Digit Symbol Substitution (processing speed), Stroop (executive function), Trails Making Test A and Trails Making Test B (attention / mental flexibility). Detailed descriptions of these tests can be found in previous studies [12,34]. The raw score for each assessment is calculated using assessment-specific measurements based on accuracy or reaction time (Table S1). BC-Assess raw overall score is the average of raw scores from all completed assessments after each score has been transformed from its natural range into a common range [0,100] using the formula in table S1. A z-score is then calculated for each assessment score and for the overall score using the corresponding age- and device-specific mean and standard deviation values from the BrainCheck normative database. Assessment standardized scores and BC-Assess overall standardized score (BC-OS) are reported by rescaling the z-scores to have a mean of 100 and a standard deviation of 15.

The ADL is an informant-based 6-item survey that measures an individual's ability to independently perform basic activities of daily living, including bathing, dressing, going to the toilet, transferring, continence, and feeding. It is calculated by adding up scores from the 6 categories, each of which takes a value of 1 for independent and 0 for dependent, resulting in an ADL total score (ADL-TS) ranging from 0 to 6. A ADL-TS of 2 or less indicates severe functional impairment, 3-4 indicates moderate

impairment, and 5-6 indicates full function [13,35].

Statistical Analysis

Data analyses were performed using Python version 3.8.5. Descriptives were presented for demographics and each score. Chi-square test was used to examine whether the distribution of gender was similar in patients from the three groups. Kruskal-Wallis test was used to compare the mean age of patients across groups.

A 3-way Multivariate Analysis of Variance (MANOVA) was used to examine the joint variation of ADL-TS and BC-OS as a function of dementia stage, age group, and gender. Post-hoc analysis employed 1-way MANOVAs to compare these two scores across dementia stages for each individual age/gender group. Tests for statistical significance of the mean differences across stages, age groups, and genders were also performed separately for each score using 3-way ANOVAs. For these statistical comparisons, age is treated as a factor with three levels: 60-69, 70-79, and 80+.

Logistic regression was used to investigate the accuracy of using the patients' BC-OS and ADL-TS to predict their dementia stage, where age and gender served as covariates. In this analysis, age is treated as a continuous variable, and gender is treated as a binary variable: 1 for male and 0 for female. Although the three dementia stages form ordinal classes, separate binary logistic regressions were used to classify Mild vs. Moderate, Moderate vs. Severe, and Mild vs. Severe, because the proportional odds assumption was not satisfied for both BC-OS (P = .01) and ADL-TS (P < .001) from Brant's Wald test [36], suggesting ordered logistic regression was not appropriate.

The binary logistic regression model for predicting a patient's condition is:

$$logit(p) = \beta_0 + \beta_1 . st_{BC} + \beta_2 . st_{ADL} + \beta_3 . st_{Age} + \beta_4 . st_{Gender} \quad (1)$$

where p is the probability of predicting the patient as having a pre-specified positive class. In each of the above binary classifications, we chose the more severe stage to be the positive class. st_{BC}, st_{ADL}, st_{Age}, st_{Gender} are standardized values of predictor variables BC-OS, ADL-TS, Age, and Gender. The coefficients $\beta_i's$ (i = 1-4) represent the effects of the four predictors, and β_0 is the intercept.

Model fitting was based on weighted loss functions to take into consideration class imbalance across the three dementia groups. Model performance was evaluated using 5-fold cross validation with stratification, repeated 20 times (100 iterations), such that on each iteration, all training and testing subsets had roughly the same proportion of patients from each group as in the original dataset. We

compared four different models:

(1) full model that included all four predictors as in equation (1)

(2) reduced model 1 that included BC-OS and ADL-TS:

$$logit(p) = \beta_0 + \beta_1 . st_{BC} + \beta_2 . st_{ADL} \quad (2)$$

(3) reduced model 2 that included BC-OS, Age, and Gender:

$$logit(p) = \beta_0 + \beta_1 . st_{BC} + \beta_3 . st_{Age} + \beta_4 . st_{Gender} \quad (3)$$

(4) reduced model 3 that included only BC-OS:

$$logit(p) = \beta_0 + \beta_1 . st_{BC} \quad (4)$$

An ROC curve was generated for each model on each cross validation iteration. Paired-samples t-tests were used to compare areas under the ROC curves (ROC-AUC) across models.

Results

Demographics and Assessment Performance

Table 1 summarizes demographic characteristics of the patients in this study. Group sample size decreases with increasing severity for both genders. The distribution of gender is similar across the three groups (P = .84). Although the range of age is similar, mean age significantly increases with severity (P < .001; pairwise comparisons: P < .001 for Mild vs. Moderate and Mild vs. Severe, P = .02 for Moderate vs. Severe).

Table 1. Demographics and summary statistics of scores across dementia severity groups and the total sample.

Demographic characteristics	Mild (n = 982)	Moderate (n = 656)	Severe (n = 113)	Total (N = 1,751)
Gender: No. (%)[a]				
Female	589 (60%)	394 (60.1%)	71 (62.8%)	1,054 (60.2%)
Male	393 (40%)	262 (39.9%)	42 (37.2%)	697 (39.8%)
Age, years[b]				
Mean (SD)	78 (8.3)	80.3 (8.4)	80.9 (9.0)	79.0 (8.6)
Range	60-101	60-102	61-101	60-102

DSRS-TS: Mean (SD)	13.4 (2.5)	25.4 (4.8)	43.3 (5.0)	19.8 (9.2)
BC-OS: Mean (SD)	62.2 (35.1)	30.1 (38.9)	-5.0 (26.8)	45.8 (41.4)
ADL-TS: Mean (SD)	4.8 (1.7)	3.2 (2.1)	1.2 (1.5)	4.0 (2.1)

[a] $P = 0.84$ (Chi-square test): distribution of gender was not significantly different across groups
[b] $P < .001$ (Kruskal-Wallis test): mean age of patients was significantly different across groups

The means and standard deviations of DSRS-TS, ADL-TS, and BC-OS are provided in Table 1. Overall, the BC-OS and ADL-TS decrease with increasing severity delineated by the DSRS-TS. Figure 1 shows the distributions of BC-OS and ADL-TS across patients within each group. For both scores, the distribution is systematically skewed towards the high values for the mild group, towards the low values for the severe group, and more evenly distributed for the moderate group. Table 2 shows the number and percentage of cases in each group that fall in different BC-Assess and ADL-TS score intervals.

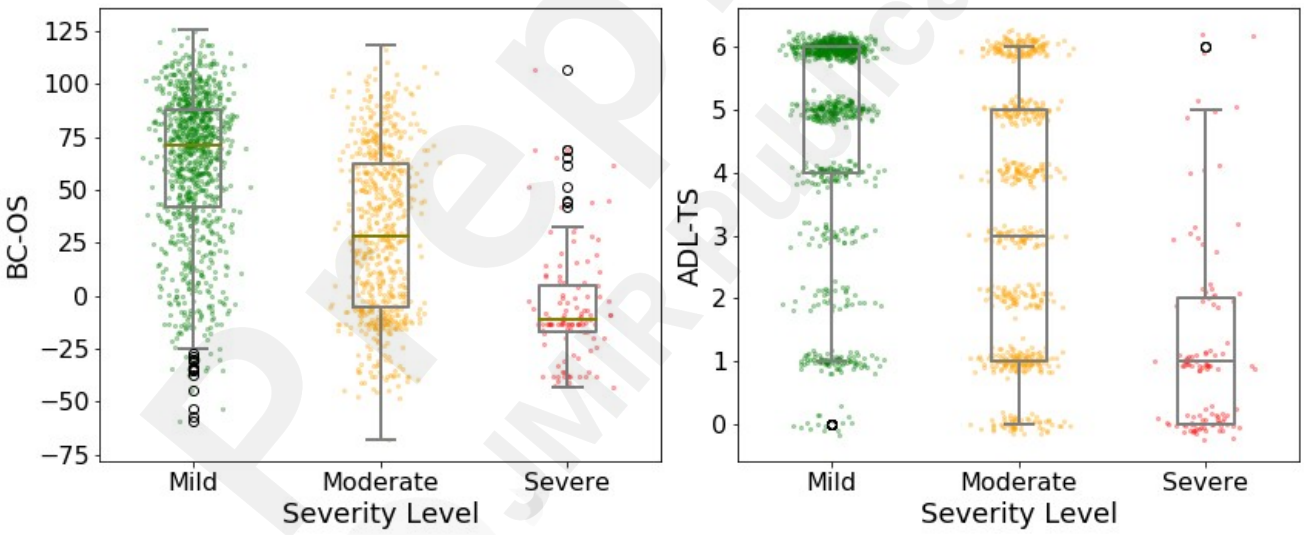

Figure 1. Box plots of the distributions of BC-OS (left) and ADL-TS (right) for each patient group: *Green* = Mild, *Yellow* = Moderate, and *Red* = Severe. Normal distributed random noise was used to add displacements along the x-axis for patients within each group (left and right), and along the y-axis at each ADL-TS value (right).

Table 2. ADL-TS and BC-OS distributions across dementia severity groups.

Score	Mild Group (n = 982)	Moderate Group (n=656)	Severe Group (n=113)
BC-OS: No. (%)			

Beyond 2 st.dev of normative mean[1]	480 (48.9%)	532 (81.1%)	112 (99.1%)
Within 2 st.dev of normative mean	502 (51.1%)	124 (18.9%)	1 (0.9%)
ADL-TS: No. (%)			
0-2 (Severe)	137 (14%)	275 (41.9%)	96 (85%)
3-4 (Moderate)	119 (12.1%)	142 (21.7%)	10 (8.8%)
5-6 (Full function)	726 (73.9%)	239 (36.4%)	7 (6.2%)

[1] Based on a BC-OS normative mean of 100 and a standard deviation of 15.

Correlations between Assessments

We find moderate Pearson correlations between DSRS-TS and BC-OS (r = -.53; $P < .001$), between DSRS-TS and ADL-TS (r = -.55; $P < .001$), and a weak correlation between BC-OS and ADL-TS (r = .37; $P < .001$). Since DSRS covers both cognitive and functional performance of a patient, moderate associations between DSRS-TS with BC-OS and ADL-TS are as expected. The weak correlation between ADL-TS and BC-OS suggests that cognitive and functional abilities are associated with each other but certain discrepancies exist between the two.

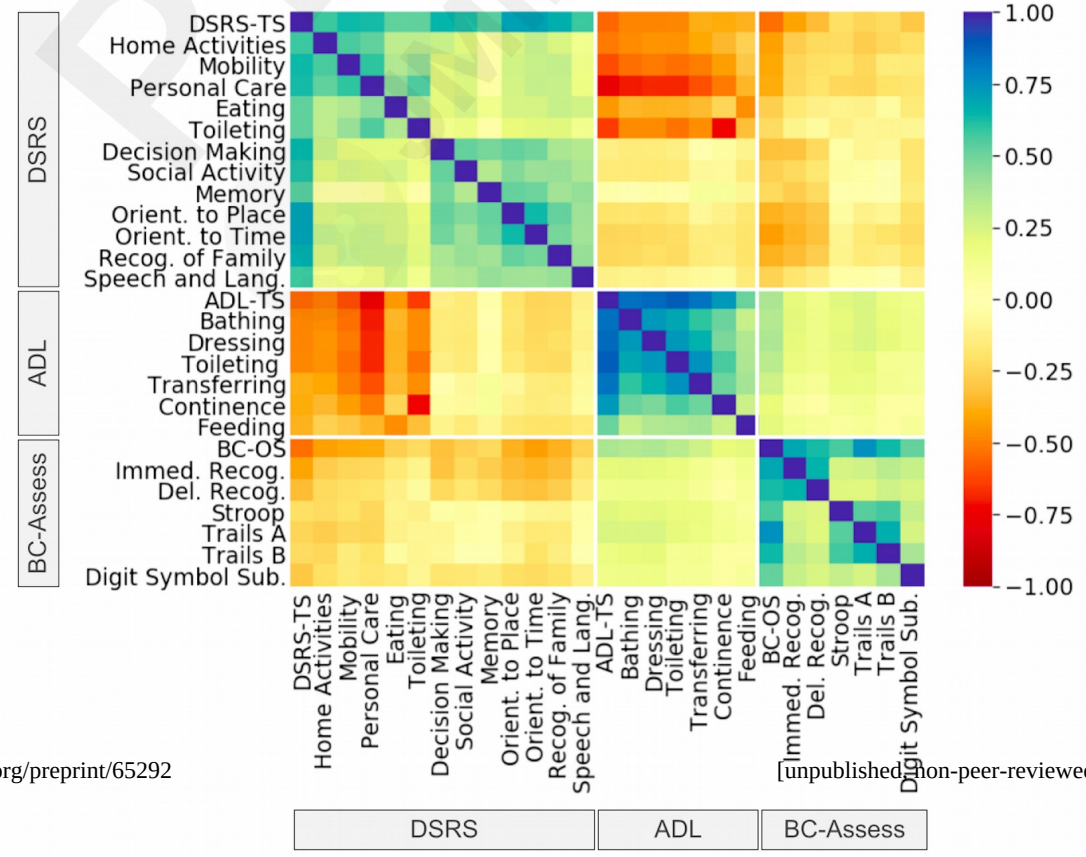

Figure 2. Correlations between BC-Assess, DSRS and ADL sub-scores.

The heatmap in Figure 2 plots Pearson correlations across DSRS, BC-Assess and ADL sub-scores. Compared with BC-Assess sub-scores, BC-OS shows stronger correlations with DSRS and ADL sub-scores. However, these correlations are weak. Among DSRS sub-scores, BC-OS is most associated with *home-activities* (r = -.41; $P < .001$), *mobility* (r = -.39; $P < .001$), *personal-care* (r = -.39; $P < .001$), *orientation-to-time* / *orientation-to-place* (r = -.43 / -.37; $P < .001$), and *recognition-of-family* (r = -.37; $P < .001$). Among ADL subscores, BC-OS is most associated with *dressing* (r = .36; $P < .001$), *bathing* (r = .35; $P < .001$), and *toileting* (r = .33; $P < .001$). Although weaker, a clear trend can be observed from Figure 2 for BC-Assess sub-scores. With regard to the DSRS, assessments of memory (*Immediate/Delayed Recognition*) are most associated with memory-demanding activities such as *memory, orientation-to-time*, *orientation-to-place*, *recognition-of-family*, and *decision-making* whereas assessments of executive function, attention, or mental flexibility (Stroop, Trails Making A/B) are most associated with *home-activities*, *mobility*, and *personal-care*. With regard to the ADL, BC-Assess sub-scores are more associated with *bathing*, *dressing*, and *toileting* than with categories that are more essential such as *feeding, continence, and transferring*. Between ADL and DSRS sub-scores, correlations mainly occurred within a subset of DSRS activities that are of the same types as those rated by the ADL such as *eating*, *home-activities*, *mobility*, *personal-care*, and *toileting*. Of these, strongest correlations were found between DSRS *toileting* and ADL *continence* (r = -.72; $P < .001$), and between DSRS *personal-care* and ADL *bathing* (r = -.70; $P < .001$), *dressing* (r = -.68; $P < .001$), *toileting* (r = -.68; $P < .001$).

Comparison of BC-OS and ADL-TS across Dementia Stages, Age Groups, Genders

Given the correlation between BC-OS and ADL-TS, we analyzed the differences in these two scores across dementia stages, age groups, and genders by running a 3-way Multivariate Analysis of Variance - MANOVA (*BC-OS + ADL-TS ~ Dementia Stage * Age Group * Gender*). Results based on the Pillais' Trace method show significant effect of Dementia Stage (Pillais' Trace = .046 , F(4,3466) = 20.5, $P < .001$) whereas the effects of Age/Gender and all interaction terms are not significant. Post-hoc one-way MANOVAs (*BC-OS + ADL-TS ~ Dementia Stage*) were run to compare BC-OS and ADL-TS combined between Mild vs. Moderate and between Moderate vs. Severe separately for each age/gender group. Except for the 60-69/Female group (N=136) showing a non-significant difference between Moderate vs. Severe, significant differences were observed for all other cases. We further performed 3-way ANOVAs where BC-OS (*BC-OS ~ Dementia Stage * Age Group * Gender*) and ADL-TS (*ADL-TS*

~ *Dementia Stage * Age Group * Gender*) were considered separately. For BC-OS, we found a significant main effect of Dementia Stage only (F(2,1733) = 270.31, *P* < .001). The insignificant differences in BC-OS across age groups are as expected as this score had been standardized to adjust for age differences. For ADL-TS, we found significant effects of Dementia Stage (F(2,1733) = 278.87, *P* < .001), Age Group (F(2,1733) = 4.77, *P* = .009), and Gender (F(1,1733) = 7.82, *P* = .005). For both scores, no interaction terms are significant.

Logistic Regression to Examine the Roles of BC-OS, ADL-TS, Age, and Gender in Predicting a Patient's Condition

ROC analysis (Fig.3) shows comparable performance between the full model (*BC-OS + ADL-TS + age/gender*: ROC-AUC = .787 for Mild vs. Moderate; .832 for Moderate vs. Severe; and .959 for Mild vs. Severe) and reduced model 1 (*BC-OS + ADL-TS*: ROC-AUC = .786 for Mild vs. Moderate; .836 for Moderate vs. Severe; and .961 for Mild vs. Severe), and between reduced model 2 (*BC-OS + age/gender*: ROC-AUC = .739 for Mild vs. Moderate; .765 for Moderate vs Severe; and .921 for Mild vs. Severe) and reduced model 3 (*BC-OS* only: ROC-AUC = .733 for Mild vs. Moderate; .767 for Moderate vs. Severe; and .917 for Mild vs. Severe). The small differences in ROC-AUC generated by the omission of age and gender suggest that they are not important predictors. Moreover, including these demographic factors appears to have led to some degree of overfitting where reduced model 1 performs slightly but significantly better than the full model (Mild vs. Severe: *P* = .002; Moderate vs. Severe: *P* < .001) and reduced model 3 performs slightly but significantly better than reduced model 2 (Mild vs. Moderate: *P* < .001; Moderate vs. Severe: *P* = .04; Mild vs. Severe: *P* < .001).

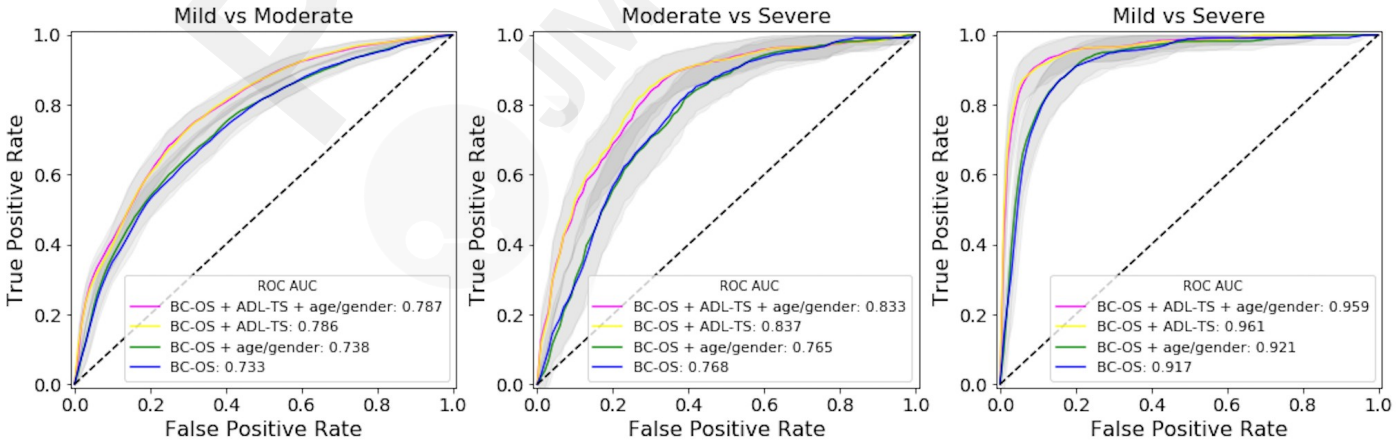

Figure 3. Model comparisons for the classification between Mild vs. Moderate (left) and between Moderate vs. Severe (right): ROC curves and mean AUCs for the full model (BC-OS + ADL-TS + age/gender; magenta), reduce model 1 (BC-OS + ADL-TS; yellow), reduced model 2 (BC-OS + age/gender;

green), and reduced model 3 (BC-OS only; blue). The shaded area along each curve represents the corresponding standard deviation across cross validation iterations. In each classification, the more severe condition was chosen to be the positive class.

The two models that include ADL-TS (full model and reduced model 1) perform significantly better than the two models without ADL-TS (reduced models 2 and 3) ($P < .001$ for each pairwise comparison), suggesting the important role of ADL. For all binary classifications, although BC-OS alone can serve as a fairly effective predictor with an ROC-AUC of at least .733, including ADL-TS in the model significantly improves prediction accuracy.

As age and gender are not important factors, we excluded them from further analysis. We examined the diagnostic performance of reduced models 1 (*BC-OS + ADL-TS*) and 3 (*BC-OS* only) at the optimal cut-points from their ROC curves. These are points on the ROC curves that maximize True Positive Rate (TPR) and minimize False Positive Rate (FPR). Table 3 shows sensitivity (TPR) and specificity (1-FPR) at the optimal cut-point for each model and classification. When included, ADL-TS improves the sensitivity/specificity by 3-5%. For reduced model 3 that includes only BC-OS, we find that the optimal cut-point corresponds to a BC-OS of 52 (roughly 3 standard deviations below the normative mean) for the classification between Mild vs. Moderate, and a BC-OS of 0 for the classification between Moderate vs. Severe.

Table 3. Sensitivity (Se) and specificity (Sp) by model and classification: mean (standard deviation) across cross validation iterations. In each classification, the more severe condition was chosen to be the positive class.

Classification	BC-OS + ADL-TS		BC-OS only	
	Se	Sp	Se	Sp
Mild vs Moderate	.74 (.04)	.72 (.04)	.69 (.05)	.68 (.05
Moderate vs Severe	.84 (.06)	.76 (.05)	.77 (.07)	.72 (.07
Mild vs Severe	.92 (.04)	.93 (.03)	.89 (.04)	.86 (.04

The fitted model coefficients are provided for reduced models 1 and 3 in Tables 4, respectively. The one-sided *P*-value obtained from bootstrapping for each coefficient is shown in the parentheses.

Table 4. Coefficients (*P*-values) of the fitted reduced models 1 (*BC-OS + ADL-TS*) and model 3 (*BC-OS* only)

	Regression Coefficients				
	Reduced Model 1			Reduced Model 3	
Classification	β_0 (Intercept)	β_1 (st_{BC})	β_2 (st_{ADL})	β_0 (Intercept)	β_1 (st_{BC})
Mild vs. Moderate	-0.071 (0.01)	-0.757 (< .001)	-0.717 (< .001)	-0.053 (0.002)	-0.890 (< .001)
Moderate vs. Severe	-0.819 (< .001)	-0.989 (0.03)	-1.059 (< .001)	-0.453 (< .001)	-1.215 (< .001)
Mild vs. Severe	-2.443 (< .001)	-1.611 (< .001)	-1.517 (< .001)	-1.557 (< .001)	-2.076 (< .001)

Discussion

By conducting a retrospective analysis of patient data in real-world clinical settings, this study sought to investigate the relationship between cognitive performance in a battery of digital cognitive assessments, BC-Assess, and dementia severity measured by the DSRS. We found a statistically significant moderate correlation between the BC-Assess overall score and the DSRS total score. This correlation is comparable with that between the ADL total score and the DSRS total score. Both BC-Assess overall score and ADL total score significantly decrease with increasing severity. BC-Assess demonstrated to be a good predictor of dementia severity, with ROC-AUC of classification using logistic regression ranging from .733 to .917. When BC-Assess is combined with ADL, higher prediction accuracies are achieved, with ROC-AUC ranging from .786 to .961.

Our results suggest that BC-Assess could serve as an alternative tool to DSRS for grading dementia severity, particularly in cases where DSRS, or other global assessments, may be challenging to obtain due to logistical and time constraints. Unlike DSRS, BC-Assess, as a brief digital cognitive assessment, offers the advantage of flexible choice of self-administration or administration by clinical support staff, runs on common consumer technology, and does not require availability of an informant. The significant improvement of prediction accuracies when BC-Assess is combined with ADL indicates the synergetic relationship between cognitive and functional measures in grading dementia severity. Previous studies have shown that patients' loss of independence to manage activities of daily living is non-linearly related with their cognitive decline [37], and their correlation weakens as the disease progresses [38]. This is consistent with the finding of relatively low correlation (r = .37) between the two measures in this study and elsewhere [39,40]. Together, these findings suggest that ADL carries

additional information of functional abilities that is partially independent of cognitive abilities measured by BC-Assess. When combined, the two measures provide a more comprehensive understanding of a patient's condition.

While the BC-Assess overall scores and the ADL total scores from the mild and severe groups separate well from each other, scores from the moderate group vary widely among patients and largely overlap with both the mild and severe groups. This is reflected in high sensitivity/specificity (.86 or higher) for the classification between the mild and severe groups, and moderate sensitivity/specificity (.83 or lower) for the classifications of the moderate group. Overall, however, a gradual change in the distribution of each score across stages can be observed. In line with current knowledge about the progression and stages of dementia [1,41], this pattern of results suggests that cognitive and functional declines do not happen in the same way across dementia patients, and that there might only be subtle differences in cognitive and/or functional performance across patients within two successive stages.

Implicit in this study is the assumption that the staging of dementia severity based on the DSRS-TS had accurately identified each patient's underlying degree of severity. Previous studies demonstrated that the DSRS has high test-retest and inter-rater reliability [25], and good concurrent validity with high correlations with the CDR, the Mini Mental Status Examination (MMSE), and the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) [23,25]. Other studies showed that the DSRS-TS can effectively differentiate between individuals with dementia, MCI, and healthy controls [42], and that it changes at a constant linear rate throughout the entire clinical course of dementia [25]. However, the psychometric properties of the DSRS in distinguishing between patients with mild, moderate, and severe dementia, have not been studied. The DSRS-TS cut-offs used for staging of dementia severity have been suggested based on the mapping of the DSRS-TS onto CDR stages where a CDR global score of 0, .5, 1, 2, and 3 represent no, questionable, mild, moderate, and severe dementia. However, it has been shown that there is a large variability in the degree of severity among patients placed in a particular CDR stage, and patients with the same degree of severity can be placed in different CDR stages [2,43,44]. Furthermore, the precision of severity grading depends on the scoring approach to the CDR, i.e. the item-response-theory approach is more precise than the sum-of-the-boxes approach which is more precise than the global score approach [43]. On top of that, the mapping of DSRS-TS to CDR global score was based on a limited sample of dementia patients that might not be representative of patients at different severity stages in general [23]. These limitations are possible contributing factors to the widespread distributions and overlaps of BC-OS and ADL-TS across patient groups delineated by the DSRS-TS in this study. To allow for more systematic investigations into the effectiveness of using

these scores in dementia staging, future research needs to address these limitations and establish more fine-grained and well-defined staging criteria as well as optimize the scoring methods for the DSRS, the CDR, and other assessments of dementia severity.

Suboptimal and inconsistent data quality in a dataset acquired outside typical clinical research settings is another factor that potentially causes large variabilities of scores observed in this study [45]. While real-world data may better represent diverse clinical environments, which is desired to obtain a generalized relationship between assessments, limited control over the data collection process and differences in clinical practices may result in reduced accuracy and consistency of the data. Inconsistencies exist across clinical sites and across units within each site due to differences in internal policies, staff training, workflow, and expertise. Inconsistencies also come from the different patient/caregiver population each site serves. Patients may also have different dementia pathologies, leading to significant differences in the pattern of scores.

The ADL and BC-Assess also have their own limitations that one should take into consideration when interpreting the current results. The ADL measures six basic activities of daily living and employs dichotomous scoring which allows only two possible scores for each functional category, i.e. 1 for *independent* and 0 for *dependent*. Therefore, it lacks the resolution to capture intermediate levels of dependences. Furthermore, as it is subjective ratings from informants, for cases with small changes in functional activities, patients may end up receiving substantially different scores depending on their caregivers' judgments, resulting in low inter-rater variability [46,47]. In this study, with only patients with dementia included (based on their DSRS-TS), we found high variability in the ADL-TS across patients within each group, especially for the moderate group. As for the BC-Assess, beside measurement errors that exist in any assessment and technical difficulties older adults may have when using smart devices for the assessment, it might have limited utility in severity staging because patients with extremely severe conditions may not be able to complete it, a common limitation of psychometric tests [48].

Our data shows a high imbalance in the number of patients across the three groups, with mild, moderate, and severe dementia accounting for 56%, 37.5%, and 6.5%, respectively. Although the trend is similar, higher proportions of patients estimated to be in the moderate (31%) and severe (21%) stages of Alzheimer's Disease were reported in a previous study [49]. As patients included in this study were those that received cognitive care planning services from their providers through the BrainCheck Plan platform within a 2-year period, our data do not necessarily reflect the prevalence of each stage. The fact that we only included patients with DSRS, ADL, and BC-Assess data also limited the number of

patients in later stages who might be too impaired to take the BC-Assess. Furthermore, given its main goals are to help patients and their family better understand the patients' condition and needs, to offer strategies to improve their overall quality of life, and to plan for the future when their condition gets worse, cognitive care planning is more meaningful for patients in early stages. Patients in the severe stage of dementia are completely dependent on their family or caregivers, and many of them require specialized care and attention in facilities. These institutionalized patients tend to have been diagnosed and given care plans tailored to their specific needs by the institution.

Despite the limitations, this study shows that BC-Assess could be a promising solution for measuring dementia severity. The use of BC-Assess for this purpose will be particularly useful in primary care settings, where DSRS or other comprehensive global assessments may pose implementation challenges. Due to its flexibility, efficiency, and ease of use, BC-Assess can help streamline the assessment process, supporting timely diagnosis and management of dementia. This, in turn, can improve patient outcomes and ease the burden on caregivers.

Acknowledgments

Research reported in this publication was supported by the National Institute On Aging of the National Institutes of Health under Award Number R44AG078006. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Authors' Contributions

DH and BH were responsible for the conceptualization and methodology of the study. DH was responsible for the formal analysis and wrote the original draft of the manuscript. All authors reviewed and edited the manuscript.

Conflicts of Interest

All authors report receiving salaries and stock options from BrainCheck.

Abbreviations

ADL: Katz index of independence in activities of daily living

ADL-TS: ADL total score

BC-Assess: BrainCheck standard battery of cognitive assessments

BC-OS: BC-Assess overall score

CDR: Clinical dementia rating

DSRS: Dementia severity rating scale

DSRS-TS: DSRS total score

ROC-AUC: area under the ROC curve

Supplementary Information

Table S1. Raw score (RS) metric and transformed score (TS) calculation for each assessment.

Assessment	Raw Score (RS) Metric	Transformed Score (TS)
Immediate/Delayed Recognition	Number of correct responses	TS = 100*RS/MAX[a]
Trails Making A	Median reaction time	TS = 100*(1-RS/MAX)
Stroop	Median reaction time	TS = 100*(1-RS/MAX)
Digit Symbol Substitution	Number of correct responses per second	TS = 100*RS/MAX

[a]MAX represents the population maximum score of the assessment across all individuals in the BrainCheck normative database.

References

1. Alzheimer's Disease and Dementia [Internet]. [cited 2024 Jun 25]. Stages of Alzheimer's. Available from: https://alz.org/alzheimers-dementia/stages
2. O'Bryant SE. Staging Dementia Using Clinical Dementia Rating Scale Sum of Boxes Scores: A Texas Alzheimer's Research Consortium Study. Arch Neurol. 2008 Aug 1;65(8):1091. Doi: 10.1001/archneur.65.8.1091
3. Horowitz FD. Developmental Theory, Prediction, and the Developmental Equation in Follow-up Research. In: Developmental Follow-up [Internet]. Elsevier; 1994 [cited 2024 Jun 25]. p. 27–44. Available from: https://linkinghub.elsevier.com/retrieve/pii/B9780122678554500064
4. Perneczky R, Wagenpfeil S, Komossa K, Grimmer T, Diehl J, Kurz A. Mapping Scores Onto Stages: Mini-Mental State Examination and Clinical Dementia Rating. Am J Geriatr Psychiatry. 2006 Feb;14(2):139–44. Doi: 10.1097/01.JGP.0000192478.82189.a8
5. Doody RS, Stevens JC, Beck C, Dubinsky RM, Kaye JA, Gwyther L, et al. Practice parameter: Management of dementia (an evidence-based review) [RETIRED]: Report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology. 2001 May 8;56(9):1154–66. Doi: 10.1212/WNL.56.9.1154
6. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. J Am Geriatr Soc. 2005 Apr;53(4):695–9. Doi: 10.1111/j.1532-5415.2005.53221.x
7. Vertesi A, Lever JA, Molloy DW, Sanderson B, Tuttle I, Pokoradi L, et al. Standardized Mini-Mental State Examination. Use and interpretation. Can Fam Physician Med Fam Can. 2001 Oct;47:2018–23. PMID: 11723596
8. Lezak MD, Howieson DB, Bigler ED, Tranel D. Neuropsychological assessment. Fifth edition. Oxford New York Auckland Cape Town: Oxford University Press; 2012. 1161 p. ISBN: 978-0-19-539552-5
9. Crawford JR, editor. A handbook of neuropsychological assessment. Nachdr. Hove: Erlbaum; 1994. 454 p. ISBN: 978-0-86377-274-0
10. Wild K, Howieson D, Webbe F, Seelye A, Kaye J. Status of computerized cognitive testing in aging: A systematic review. Alzheimers Dement. 2008 Nov;4(6):428–37. Doi: 10.1016/j.jalz.2008.07.003
11. Chan JYC, Yau STY, Kwok TCY, Tsoi KKF. Diagnostic performance of digital cognitive tests for the identification of MCI and dementia: A systematic review. Ageing Res Rev. 2021 Dec;72:101506. Doi: 10.1016/j.arr.2021.101506
12. Ye S, Sun K, Huynh D, Phi HQ, Ko B, Huang B, et al. A Computerized Cognitive Test Battery for Detection of Dementia and Mild Cognitive Impairment: Instrument Validation Study. JMIR Aging. 2022 Apr 15;5(2):e36825. Doi: 10.2196/36825
13. Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW. Studies of Illness in the Aged. The Index of ADL: a Standardized Measure of Biological and Psychosocial Function. JAMA. 1963 Sep 21;185:914–9. PMID: 14044222
14. Pashmdarfard M, Azad A. Assessment tools to evaluate Activities of Daily Living (ADL) and Instrumental Activities of Daily Living (IADL) in older adults: A systematic review. Med J Islam Repub Iran. 2020;34:33. PMID: 32617272
15. Reisberg B, Ferris SH, de Leon MJ, Crook T. The Global Deterioration Scale for assessment of primary degenerative dementia. Am J Psychiatry. 1982 Sep;139(9):1136–9. PMID: 7114305

16. Sclan SG, Reisberg B. Functional Assessment Staging (FAST) in Alzheimer's Disease: Reliability, Validity, and Ordinality. Int Psychogeriatr. 1992 Apr;4(3):55–69. Doi: 10.1017/S1041610292001157
17. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology. 1993 Nov;43(11):2412–4. PMID: 8232972
18. Morris JC. Clinical Dementia Rating: A Reliable and Valid Diagnostic and Staging Measure for Dementia of the Alzheimer Type. Int Psychogeriatr. 1997 Dec;9(S1):173–6. Doi: 10.1017/S1041610297004870
19. Fillenbaum GG, Peterson B, Morris JC. Estimating the validity of the clinical Dementia Rating Scale: the CERAD experience. Consortium to Establish a Registry for Alzheimer's Disease. Aging Milan Italy. 1996 Dec;8(6):379–85. PMID: 9061124
20. Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A new clinical scale for the staging of dementia. Br J Psychiatry J Ment Sci. 1982 Jun;140:566–72. PMID: 7104545
21. Rikkert MGMO, Tona KD, Janssen L, Burns A, Lobo A, Robert P, et al. Validity, reliability, and feasibility of clinical staging scales in dementia: a systematic review. Am J Alzheimers Dis Other Demen. 2011 Aug;26(5):357–65. PMID: 21914671
22. Duara R, Loewenstein DA, Greig-Custo MT, Raj A, Barker W, Potter E, et al. Diagnosis and staging of mild cognitive impairment, using a modification of the clinical dementia rating scale: the mCDR. Int J Geriatr Psychiatry. 2010 Mar;25(3):282–9. Doi: 10.1002/gps.2334
23. Clark CM, Ewbank DC. Performance of the dementia severity rating scale: a caregiver questionnaire for rating severity in Alzheimer disease. Alzheimer Dis Assoc Disord. 1996;10(1):31–9. PMID: 8919494
24. dsrs-scale.pdf [Internet]. [cited 2024 Jun 25]. Available from: https://www.alz.org/careplanning/downloads/dsrs-scale.pdf
25. Xie SX, Ewbank DC, Chittams J, Karlawish JHT, Arnold SE, Clark CM. Rate of decline in Alzheimer disease measured by a Dementia Severity Rating Scale. Alzheimer Dis Assoc Disord. 2009;23(3):268–74. PMID: 19812470
26. Weinstein JD, Gonzalez ER, Egleton RD, Hunt DA. A paradigm shift for evaluating pharmacotherapy for Alzheimer's disease: the 10-patient screening protocol. Consult Pharm J Am Soc Consult Pharm. 2013 Jul;28(7):443–54. PMID: 23835462
27. McCulloch K, Collins RL, Maestas KL, LeMaire AW, Pacheco V. Validation and diagnostic utility of the dementia rating scale in a mixed dementia population. Alzheimer Dis Assoc Disord. 2014;28(2):168–74. PMID: 24231456
28. Reisberg B, Ferris SH, Torossian C, Kluger A, Monteiro I. Pharmacologic treatment of Alzheimer's disease: a methodologic critique based upon current knowledge of symptomatology and relevance for drug trials. Int Psychogeriatr. 1992;4 Suppl 1:9–42. PMID: 1504291
29. Roalf DR, Moberg PJ, Xie SX, Wolk DA, Moelter ST, Arnold SE. Comparative accuracies of two common screening instruments for classification of Alzheimer's disease, mild cognitive impairment, and healthy aging. Alzheimers Dement. 2013 Sep;9(5):529–37. PMID: 23260866
30. de Craen AJM, Heeren TJ, Gussekloo J. Accuracy of the 15-item geriatric depression scale (GDS-15) in a community sample of the oldest old. Int J Geriatr Psychiatry. 2003 Jan;18(1):63–6. PMID: 12497557
31. Steffens DC, Otey E, Alexopoulos GS, Butters MA, Cuthbert B, Ganguli M, et al. Perspectives on Depression, Mild Cognitive Impairment, and Cognitive Decline. Arch Gen Psychiatry. 2006 Feb 1;63(2):130. Doi: 10.1001/archpsyc.63.2.130
32. Sultzer DL, Levin HS, Mahler ME, High WM, Cummings JL. A comparison of psychiatric symptoms in vascular dementia and Alzheimer's disease. Am J Psychiatry. 1993 Dec;150(12):1806–12. PMID: 8238634

33. Forsell Y, Winblad B. Major depression in a population of demented and nondemented older people: prevalence and correlates. J Am Geriatr Soc. 1998 Jan;46(1):27–30. PMID: 9434662
34. Yang S, Flores B, Magal R, Harris K, Gross J, Ewbank A, et al. Diagnostic accuracy of tablet-based software for the detection of concussion. Kobeissy FH, editor. PLOS ONE. 2017 Jul 7;12(7):e0179352. PMID: 28686616
35. Wallace M, Shelkey M, Hartford Institute for Geriatric Nursing. Katz Index of Independence in Activities of Daily Living (ADL). Urol Nurs. 2007 Feb;27(1):93–4. PMID: 17390935
36. Brant R. Assessing proportionality in the proportional odds model for ordinal logistic regression. Biometrics. 1990 Dec;46(4):1171–8. PMID: 2085632
37. Zidan M, Arcoverde C, Araújo NBD, Vasques P, Rios A, Laks J, et al. Motor and functional changes in different stages of Alzheimer's disease. Arch Clin Psychiatry São Paulo. 2012;39(5):161–5. Doi: 10.1590/S0101-60832012000500003
38. Monaci L, Morris RG. Neuropsychological screening performance and the association with activities of daily living and instrumental activities of daily living in dementia: baseline and 18- to 24-month follow-up. Int J Geriatr Psychiatry. 2012 Feb;27(2):197–204. PMID: 22223145
39. Louis ED. Functional correlates of lower cognitive test scores in essential tremor. Mov Disord. 2010 Mar 15;25(4):481–5. Doi: 10.1002/mds.22920
40. Lee G. Impaired Cognitive Function is Associated with Motor Function and Activities of Daily Living in Mild to Moderate Alzheimer's Dementia. Curr Alzheimer Res. 2020;17(7):680–6. PMID: 32811414
41. The progression, signs and stages of dementia | Alzheimer's Society [Internet]. 2021 [cited 2024 Jun 25]. Available from: https://www.alzheimers.org.uk/about-dementia/symptoms-and-diagnosis/how-dementia-progresses/progression-stages-dementia
42. Mitchell JC, Dick MB, Wood AE, Tapp AM, Ziegler R. The Utility of the Dementia Severity Rating Scale in Differentiating Mild Cognitive Impairment and Alzheimer Disease From Controls. Alzheimer Dis Assoc Disord. 2015 Jul;29(3):222–8. Doi: 10.1097/WAD.0000000000000057
43. Balsis S, Miller TM, Benge JF, Doody RS. Dementia staging across three different methods. Dement Geriatr Cogn Disord. 2011;31(5):328–33. PMID: 21555887
44. Chang YL, Bondi MW, McEvoy LK, Fennema-Notestine C, Salmon DP, Galasko D, et al. Global clinical dementia rating of .5 in MCI masks variability related to level of function. Neurology. 2011 Feb 15;76(7):652–9. PMID: 21321338
45. Bhatt A. Data quality – The foundation of real-world studies. Perspect Clin Res. 2023;14(2):92. Doi: 10.4103/picr.picr_12_23
46. Brorsson B, Asberg KH. Katz index of independence in ADL. Reliability and validity in short-term care. Scand J Rehabil Med. 1984;16(3):125–32. PMID: 6494836
47. Arik G, Varan HD, Yavuz BB, Karabulut E, Kara O, Kilic MK, et al. Validation of Katz index of independence in activities of daily living in Turkish older adults. Arch Gerontol Geriatr. 2015;61(3):344–50. PMID: 26328478
48. Auer SR, Sclan SG, Yaffee RA, Reisberg B. The neglected half of Alzheimer disease: cognitive and functional concomitants of severe dementia. J Am Geriatr Soc. 1994 Dec;42(12):1266–72. PMID: 7983290
49. Hebert LE, Scherr PA, Bienias JL, Bennett DA, Evans DA. Alzheimer disease in the US population: prevalence estimates using the 2000 census. Arch Neurol. 2003 Aug;60(8):1119–22. PMID: 12925369

Supplementary Files

Multimedia Appendixes

Table S1. Raw score (RS) metric and transformed score (TS) calculation for each assessment.
URL: http://asset.jmir.pub/assets/beef884c187801c6d67a26bf9f5f6571.docx