AMERICAN PSYCHOLOGICAL ASSOCIATION

Neuropsychology

ISSN: 0894-4105

2023, Vol. 37, No. 1, 32–43
https://doi.org/10.1037/neu0000853

Regression-Based Cognitive Change Norms Applied in Biochemically Defined Predementia Alzheimer's Disease

Ingvild Vøllo Eliassen[1, 2], Bjørn-Eivind Kirsebom[4, 5], Tormod Fladby[1, 2, 3], Knut Waterloo[5, 4], Sigrid Botne Sando[6, 7], Mathilde Suhr Hemminghyth[8, 9], Dag Aarsland[1, 10, 11], Santiago Timón-Reina[1, 12], Anders Wallin[13, 14], Fredrik Öhman[13, 14, 15], Marie Eckerström[13, 14], and Erik Hessen[1, 2]

[1] Department of Neurology, Akershus University Hospital, Lørenskog, Norway
[2] Department of Psychology, University of Oslo
[3] Institute of Clinical Medicine, University of Oslo
[4] Department of Psychology, Faculty of Health Sciences, UiT The Arctic University of Norway
[5] Department of Neurology, University Hospital of North Norway, Tromsø, Norway
[6] Department of Neuromedicine and Movement Science, Faculty of Medicine and Heath Sciences, Norwegian University of Science and Technology
[7] Department of Neurology and Clinical Neurophysiology, University Hospital of Trondheim, Trondheim, Norway
[8] Department of Research and Innovation, Research Group for Age-Related Medicine, Helse Fonna, Haugesund Hospital, Haugesund, Norway
[9] Department of Neuropsychology, Helse-Fonna, Haugesund Hospital, Haugesund, Norway
[10] Department of Old Age Psychiatry, Institute of Psychiatry, Psychology and Neuroscience, King's College London
[11] Center for Age-Related Diseases, Stavanger University Hospital, Stavanger, Norway
[12] Departamento de Inteligencia Artificial, Universidad Nacional de Educación a Distancia
[13] Institute of Neuroscience and Physiology, Department of Psychiatry and Neurochemistry, University of Gothenburg, Mölndal, Sweden
[14] Memory Clinic at Department of Neuropsychiatry, Sahlgrenska University Hospital, Mölndal, Sweden
[15] Wallenberg Centre for Molecular and Translational Medicine, University of Gothenburg

Objective: We aim to develop 2-year cognitive change norms for adults ages 41–84 for six cognitive tests, and to evaluate these norms in groups with AD biomarkers. ***Background:*** Practice effects are common in repeated neuropsychological testing. Not accounting for practice effects may obscure cognitive decline in early Alzheimer's disease (AD). ***Method:*** We developed standardized regression-based change norms from normative samples consisting of healthy controls from the Dementia Disease Initiation study ($n = 125$), the Trønderbrain study ($n = 57$), and the Gothenburg mild cognitive impairment (MCI) study ($n = 65$). Norms

This article was published Online First September 22, 2022.

Ingvild Vøllo Eliassen https://orcid.org/0000-0003-1288-6032

The authors thank all collaborators for contributions in the Dementia Disease Initiation project, the Trønderbrain project, and the Gothenburg mild cognitive impairment study.

Dag Aarsland has received research support and/or honoraria from Astra-Zeneca, H. Lundbeck, Novartis Pharmaceuticals and General Electric Healthcare (GE) Health, and served as paid consultant for H. Lundbeck, Eisai, Heptares, Sanofi, Mentis Cura. Dag Aarsland is partly funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley National Health Service (NHS) Foundation Trust and King's College London. The views expressed are those of the author and not necessarily those of the NIHR or the Department of Health and Social Care. Ingvild Vøllo Eliassen received funding from Grant HSØ 2018004 from the Helse Sør-Øst RHF. Tormod Fladby received funding from Grant NRC 311993 JPND-EU from the Precision Medicine Interventions in Alzheimer's Disease. Bjørn-Eivind Kirsebom served as a consultant for Biogen. Anders Wallin was supported by the Swedish state under the agreement between the Swedish government and the county councils, the avtal om läkarutbildning och forskning (ALF) agreement (Grant ALFGBG-720661).

Data are stored at services for sensitive data at the University of Oslo and are publicly unavailable, but code and anonymized data may be made available upon reasonable request to the corresponding author.

Ingvild Vøllo Eliassen played lead role in formal analysis and writing of original draft and equal role in conceptualization, investigation, methodology and writing of review and editing. Bjørn-Eivind Kirsebom played lead role in conceptualization and methodology and equal role in investigation and writing of review and editing. Tormod Fladby played supporting role in conceptualization and equal role in funding acquisition, resources, supervision and writing of review and editing. Knut Waterloo played supporting role in conceptualization and equal role in resources, supervision and writing of review and editing. Sigrid Botne Sando played equal role in investigation, resources and writing of review and editing. Mathilde Suhr Hemminghyth played equal role in investigation and writing of review and editing. Dag Aarsland played equal role in funding acquisition, investigation, resources, supervision and writing of review and editing. Santiago Timón-Reina played lead role in software and equal role in data curation and writing of review and editing. Anders Wallin played equal role in funding acquisition, resources, supervision and writing of review and editing. Fredrik Öhman played equal role in data curation, investigation and writing of review and editing. Marie Eckerström played lead role in writing of review and editing and equal role in investigation. Erik Hessen played equal role in conceptualization, supervision and writing of review and editing.

Correspondence concerning this article should be addressed to Ingvild Vøllo Eliassen, Department of Neurology, Akershus University Hospital, P.O. Box 1000, 1478 Lørenskog, Norway. Email: ingvivel@gmail.com

were applied in a sample with cognitive symptoms (subjective cognitive decline or MCI) and AD cerebrospinal fluid (CSF) biomarkers (n = 246), classified according to the A/T/N system. ***Results:*** The change norms adjusted for pertinent demographics and practice effects. The group with cognitive complaints displayed a trend toward cognitive decline compared to the normative group, with the A+T/N+ subgroup showing the most marked decline. This was observed in tests of episodic memory and cognitive flexibility/divided attention. ***Conclusions:*** We present 2-year cognitive change norms for adults between 41 and 84 years, adjusted for practice and demographics. A web-based change norm calculator is provided.

Key Points

Question: When we statistically adjust for relevant factors, how is cognitive change associated with Alzheimer's disease biochemical pathology? ***Findings:*** Using the norms for change developed in this study, we found that biomarkers indicating Alzheimer's disease were associated with cognitive decline in tests of memory and executive function. ***Importance:*** Our change norms can be used in the clinic or in research to detect cognitive decline that may be caused by early Alzheimer's disease or other relevant conditions. ***Next Steps:*** Future research could explore the usefulness of cognitive change norms when diagnosing mild cognitive impairment.

Keywords: Alzheimer's disease, neuropsychological change norms, practice effects, cerebrospinal fluid biomarkers

Supplemental materials: https://doi.org/10.1037/neu0000853.supp

Repeated neuropsychological testing is a common and valuable method for assessing cognitive change (Heilbronner et al., 2010). In progressive neurodegenerative disease such as Alzheimer's disease (AD), decline detected by serial testing may aid diagnosis and point to disease development (Albert et al., 2011). Cognitive improvement or lack of decline may indicate treatment effects in intervention studies. Such "true" cognitive change, attributable to disease progression, recovery, or intervention benefits, is arguably of most interest to clinicians and researchers. However, the observed change in test scores can also be influenced by other factors such as measurement error, regression to the mean, and practice effects (Duff, 2012).

Practice effects, defined as test score improvements in retesting due to prior exposure to the method and material, are a central challenge in serial testing (Duff, 2012; Heilbronner et al., 2010). Practice effects are documented in a wide range of neuropsychological tests, cognitive domains, and in several clinical groups (Calamia et al., 2012), including in preclinical (Machulda et al., 2013, 2017) prodromal and dementia phases of AD (Goldberg et al., 2015; Wang et al., 2020). Meta-analytic findings indicate that patient-related variables (e.g., demographics, clinical conditions), test characteristics (e.g., construct measured, novelty), and the test–retest interval length can moderate the magnitude of practice effects (Calamia et al., 2012). Through inflating follow-up performance, practice effects can obscure cognitive decline. This may lead to delayed or underdiagnosis of mild cognitive impairment (MCI; Elman et al., 2018) or overestimating treatment effects in clinical trials (Goldberg et al., 2015).

Attenuated practice effects have been proposed as an early cognitive marker of AD dementia (e.g., Hassenstab et al., 2015), and recent systematic reviews have documented that in AD and MCI, smaller practice effects are linked to greater cognitive decline over time (Jutten et al., 2020; Wang et al., 2020). Smaller practice effects are also associated with the presence of AD pathology, demonstrated by the presence of cerebrospinal fluid (CSF) AD biomarkers (Jutten et al., 2020). This includes the three hallmark AD biomarkers used in the A/T/N classification system (Jack et al., 2018): β-amyloid (Aβ) plaques (A), phosphorylated tau (T) and neurodegeneration/neuronal injury (N; Jutten et al., 2020). Classifying each biomarker as normal/abnormal (−/+), the A/T/N system is a unified framework for defining biomarker profiles and categories (Jack et al., 2018). Abnormal Aβ (A+) is required to classify an individual in the AD pathological continuum. Studies of clinically asymptomatic older adults with reexaminations approximately 1 year apart indicate that groups with normal Aβ and neurodegeneration markers (A−N−) have larger performance gains than groups with both pathological markers (A+N+), while results regarding the association between only Aβ pathology (A+N−) and practice effects are mixed (Machulda et al., 2017; Mormino et al., 2014). Thus, in the context of biomarker-defined AD, reduced practice effects can represent a subtle change in cognition that might be hard to detect without adequate methods, especially in cases without marked cognitive decline.

In their official position article on serial neuropsychological assessments, The American Academy of Clinical Neuropsychology advises neuropsychologists to be conscious of the influence of practice effects on repeat testing (Heilbronner et al., 2010). Evidence-based tools may help to determine whether measured discrepancies across testing sessions are clinically meaningful. While neuropsychological tests and performance norms are frequently used to evaluate the current cognitive capacity of an individual, norm-referenced scores for intraindividual change over time are less common (Attix et al., 2009). Such data have been published for some common neuropsychological batteries and tests (Brooks et al., 2016; Hammers et al., 2020; Kiselica et al., 2020). Statistical methods for cognitive change measurement may allow us to view practice effects more as a measurable and informative construct than as a confound (Heilbronner et al., 2010).

Standardized regression-based (SRB) methods are one approach for assessing reliable cognitive change in individual test scores (see Duff, 2012, for review of methods; McSweeny et al., 1993). By including baseline performance and other relevant variables in predicting follow-up performance, this method allows us to control

for practice effects, regression to the mean, and the effect of demographics on change (Duff, 2012). The final product in SRB norm development is a standardized change score, statistically stripped for relevant influences. This score may give a closer estimate of the individual's true cognitive change.

In the present study, we aim to develop cognitive change norms for older adults for six cognitive tests, adjusted for pertinent demographics, and practice effects. Norms are validated in a sample with self-reported cognitive symptoms (with either subjective cognitive decline [SCD] or MCI) and AD CSF biomarkers. Based on previous findings, we hypothesize that (a) Patients with self-reported cognitive symptoms at baseline will show cognitive decline compared with healthy control participants; (b) subgroups with biomarkers on the AD continuum (abnormal Aβ, with or without tau and neurodegeneration markers, as defined by Jack et al., 2018) will show decline in cognitive change scores; (c) amyloid positive individuals with phosphorylated tau and/or neurodegeneration markers (A+T/N+) will show larger cognitive decline than those without (A+T−N−).

Method

Participants and Procedure

Participants were mainly included from the Dementia Disease Initiation (DDI) study, a Norwegian observational study of early AD markers. To obtain larger sample sizes for the regression norming (normative sample; total $n = 247$), we also included healthy control participants from two additional observational studies: the Gothenburg MCI study and the Trønderbrain study. All participants in the symptom group sample ($n = 246$) were from the DDI study.

The DDI

The DDI is an observational multicenter study that recruits healthy controls and individuals with cognitive complaints for longitudinal multimodal evaluation. Inclusion criteria are age between 40 and 80 years and a native language of Norwegian, Swedish, or Danish. Brain trauma or disorder (including stroke, dementia, severe psychiatric disease, or developmental disorder) are exclusion criteria. Participants with cognitive complaints are recruited from hospital memory clinics and media advertisements. Healthy control participants are primarily spouses of symptom group participants, and smaller portions are recruited from media advertisements and the hospital orthopedic ward. The study was approved by the regional medical research ethics committee, and all participants provided written informed consent. In the present study, we included one sample of healthy controls ($n = 125$) and one sample consisting of participants with cognitive complaints (symptom group sample, $n = 246$). The symptom group sample was compiled of patients with SCD ($n = 117$), defined by criteria presented by Jessen et al. (2014); and patients with MCI ($n = 129$), defined according to the National Institute on Aging–Alzheimer's Association (NIA-AA) workgroup criteria (Albert et al., 2011). All participants had completed baseline and one follow-up assessment after approximately 2 years.

All participants completed a standardized case report form (CRF) at baseline and 2-year follow-up. The CRF includes medical history reports from the patient and an informant, a neurological examination, and a brief neuropsychological test battery.

The administration time of the test battery was approximately 30 min. The battery included the following tests: (a) The Consortium to Establish a Registry for Alzheimer's Disease (CERAD) Word List Learning Test (verbal learning; Fillenbaum et al., 2008) and (b) the CERAD Word List Recall Test (delayed verbal recall; Fillenbaum et al., 2008), which consist of 10 words presented to the participant. The learning score is the sum of three consecutive learning trials, and the recall score is the number of words recalled after 10 min; (c) the Visual Object and Space Perception Battery (VOSP) Silhouettes Test (visuoperceptual ability; Warrington & James, 1991), where the participant is shown 30 pictures of silhouettes of animals and objects and asked to identify them; (d) the Trail Making Test A (TMT-A; psychomotor speed; Reitan & Wolfson, 1985), a timed paper and pencil task where the participant is asked to connect, in ascending order, circles containing numbers 1–25 that are distributed out of order across the sheet; (e) the Trail Making Test B (TMT-B; cognitive flexibility/divided attention; Reitan & Wolfson, 1985), where circles containing numbers 1–13 and letters A–L are connected in ascending order, alternating between numbers and letters; and (f) the Controlled Word Association Test (COWAT)/FAS (phonemic word fluency; Benton & Hamsher, 1978), where the participant is asked to verbally produce as many words as possible starting with the letters F, A, and S, respectively, across three 1-min sessions.

The standardized protocol for a lumbar puncture is described in Fladby et al. (2017). CSF concentrations of total tau (T-tau), phosphorylated tau (P-tau), and β-amyloid ($A\beta_{42/40\ ratio}$) were analyzed using Meso Scale (β-amyloid), and ELISA; Innotest Phospho-Tau (181P) and Innotest h-Tau Ag, Fujirebio, Ghent Belgium.

Abnormality in CSF biomarkers was determined by the following cutoff values: β-amyloid 42/40 ratio ≥ 0.077 pg/ml (Siafarikas et al., 2021); p-tau ≥ 80 pg/ml; and t-tau > 300 pg/ml for age < 50 years, > 450 pg/ml for age 50–69 years, and > 500 pg/ml for age ≥70 years (modified from Sjögren et al., 2001). We based biomarker classification on the A/T/N system, proposed by the NIA-AA research framework for biochemical AD (Jack et al., 2018). The letters A, T, and N represent the three hallmark AD biomarkers described above: A = Aβ 42/40 ratio, T = p-tau, and N = t-tau. From the binary classification of the three CSF biomarkers according to the described cutoff values, individuals were assigned to one of three profiles at baseline: (a) normal AD biomarkers (A−T−N−); (b) Alzheimer's pathological change (A+T−N−); and (c) AD/ evidence of amyloidosis and neurodegeneration (A+T+N+/A+T+N−/A+T−N+). Due to small sample sizes, probably related to baseline exclusion of participants with known brain pathology, we did not include individuals with biomarker evidence of suspected non-AD pathological change in analyses (A−T+N+/A−T+N−/A−T−N+).

Gothenburg MCI Study

The Gothenburg MCI study is based at Sahlgrenska University Hospital in Sweden. It is a longitudinal observational study on patients seeking help for cognitive complaints at a memory clinic. Healthy controls are also included and repeatedly followed up like the patient groups. Participants included in the present study ($n = 65$) are cognitively normal healthy controls assessed between April 2000 and November 2015. They were mainly recruited through senior citizen organizations, and some were relatives of participants in the symptom group. Inclusion criteria for healthy controls were age between 50 and 79, no subjective or objective cognitive decline, and an Mini-Mental State examination (MMSE) score of > 26. Severe somatic or psychiatric disease that may cause cognitive

impairment lead to exclusion. The neuropsychological assessment consisted of a comprehensive test battery including TMT-A, TMT-B, and VOSP silhouettes. Data from these three tests are included in the present study. Study participants were assessed at baseline and after 2 years. The study was approved by the local ethics committee, and all participants gave written informed consent. For a more detailed account of the Gothenburg MCI study, see Wallin et al. (2016).

Trønderbrain

The Trønderbrain study included older adults with MCI, early AD, or normal cognition between 2009 and 2015 in central Norway. Healthy controls were recruited from societies for retired people organizations and spouses of symptom group participants. Exclusion criteria were insufficient sight or hearing to complete cognitive testing, psychiatric or malignant disease, use of anticoagulant medication, neurological disorder, and alcohol or drug abuse. All participants or their proxies provided written informed consent, and the study was approved by the regional committee for medical research ethics. In the present study, we included 57 healthy control participants who had completed baseline and 2-year follow-up assessments, including cognitive testing with CERAD word list learning and recall. See Grøntvedt et al. (2020), for further description.

Statistics

Statistical analysis was performed using R (R Core Team, 2020). We used the Statistical Package for Social Sciences (SPSS) Version 26 for data management. Descriptive statistics of baseline demographics and baseline and follow-up cognitive characteristics were calculated.

Norming Procedure

We used an SRB norming method to develop cognitive change norms (Duff, 2012; McSweeny et al., 1993). For each cognitive test, we developed 2-year SRB change norms from the normative sample. Using multiple linear regression analysis, follow-up cognitive test scores were modeled using baseline test score, number of months between assessments, and pertinent demographics (age, sex, years of education) as predictors. Plots of residuals versus predicted values were obtained to check the assumptions of independence of errors and homoscedasticity. Linearity was assessed through residual plots. The assumption of normality of residuals was assessed visually with qq plots and histograms, and this inspection indicated nonnormality. To normalize raw test scores, we obtained their cumulative frequency distributions and converted them into standardized scaled scores ($M = 10$, $SD = 3$). We used the Bayesian information criterion to guide model selection among possible models describing the relationship between the predictors.

Regression equations from each test were used to produce predicted scaled scores for all participants by applying individual baseline test score and demographic information as in the following example: Predicted follow-up scaled test score = Intercept + (Individual Baseline Scaled Test Score × $\beta_{\text{baseline test score}}$) + (Individual Sex × β_{sex}) + (Individual Age × β_{age}) + (Years of Education × $\beta_{\text{education}}$).

Next, we used the discrepancy between the predicted and actual test results to produce a standardized Z score: Z = (Obtained scaled test score − Predicted follow-up scaled test score)/SD of regression residuals. This value indicates whether individual performance after 2 years deviates from the predicted performance, controlling for baseline variables.

Norm Calculator

To facilitate the usage of the proposed norms, we have developed a web-based calculator, freely accessible at https://contattafiles.s3.us-west-1.amazonaws.com/tnt30503/bFttvc4M25RAmvM/cogchange.html. To obtain the Z scores, the user must enter valid demographic data (age, years of education, and sex) and the raw score values obtained from the administered tests. The application is implemented using HTML5 and client-side web technologies, and its graphical user interface follows a responsive design to provide a good experience across devices and platforms. The source code is released with opensource Apache License, Version 2.0, at https://github.com/DDI-NO/cogchange-calc.

Applying Norms in Symptom Group Sample

We applied the change norms on test scores in the symptom group sample. Using analyses of variance (ANOVAs), cognitive change scores were compared between the control group, participants with SCD, and participants with MCI. Planned comparisons were carried out with no corrections. Due to heterogeneity of variance in TMT-B change scores, these data were analyzed using Welch's ANOVA, and pairwise Welch's unequal variances t test was used for planned comparisons. We used the Kruskal–Wallis test to compare scores for CERAD recall and VOSP silhouettes, which showed nonnormal distributions. Here, pairwise Wilcoxon rank-sum tests were used for planned comparisons. Dummy regression analyses were employed to compare mean cognitive change in the different biomarker groups A−T−N−($n = 160$), A+T−N−($n = 44$), and A+T/N+($n = 42$). Data visualization was done using the R package "ggplot2" (Wickham, 2016).

Transparency and Openness

We report how we determined sample sizes and data exclusions. Baseline characteristics of nonanalyzed participants with incomplete follow-up or missing data, as well as of analyzed and nonanalyzed groups combined, are provided as Supplemental Material (Table S1) for evaluation of potential biases due to attrition and missing data. Citations for code and study materials are provided. Data are stored at services for sensitive data (TSD) at the University of Oslo (UiO) and are publicly unavailable, but code and anonymized data may be made available upon reasonable request to the corresponding author. The design and analysis of this study were not preregistered.

Results

Table 1 presents descriptive statistics of demographics and cognitive raw scores in the normative and symptom group.

Regression Norming

The normative groups included in total 247 control participants who completed baseline and follow-up evaluations. One participant was diagnosed with MCI at follow-up and was, therefore, excluded from the analyses. The mean time interval between assessments was

Table 1
Baseline Demographics and Cognitive Raw Test Scores at Both Assessments in the Normative and Symptom Group Samples

Variable	Normative group (n = 247)					Symptom group (n = 246)			
	M	*SD*	Range	*n*	*n* (%)	*M*	*SD*	Range	*n* (%)
Age	64.3	8.3	41–84	247		63.5	9.3	40–81	
Sex, female				247	148 (59.9%)				132 (53.7%)
Years of education	13.5	3.4	6–24	247		13.9	3.1	7–22	
Retest interval (months)	28.1	5.6	13–48	247		25.3	4.8	9–51	
CERAD learning baseline	21.3	3.3	13–29	178		19.1	4.6	2–28	
CERAD learning follow-up	22.7	3.8	12–30	178		20.0	5.4	3–30	
CERAD recall baseline	7.3	1.8	1–10	179		5.6	2.7	0–10	
CERAD recall follow-up	7.5	1.9	2–10	179		5.9	3.0	0–10	
TMT-A baseline	35.7	11.2	16–66	182		40.9	17.3	13–142	
TMT-A follow-up	35.0	11.6	15–83	182		41.4	27.1	13–300	
TMT-B baseline	84.0	27.6	34–191	180		108.0	57.5	25–300	
TMT-B follow-up	84.2	31.0	38–240	180		116.6	68.3	35–300	
FAS baseline	40.8	11.4	20–70	122		39.2	12.3	10–85	
FAS follow-up	42.8	11.9	17–84	122		39.7	13.3	4–95	
VOSP silhouettes baseline	22.6	4.0	11–30	179		21.6	4.6	7–30	
VOSP silhouettes follow-up	22.9	3.7	11–30	179		22.0	4.7	7–30	

Note. Follow-up assessment approximately 2 years after baseline. CERAD = The Consortium to Establish a Registry for Alzheimer's Disease; TMT = Trail Making Test; VOSP = Visual Object and Space Perception Battery.

28 months (SD = 5.6, range = 13–48). Raw score to scaled score conversions is shown in Table 2.

Results from the normative regression analyses are presented in Table 3. Baseline scores explained most of the variance for all cognitive tests, with higher baseline performance predicting higher performance in subsequent testing. For all tests except TMT-B and FAS, one or more demographic factors contributed to the prediction of follow-up performance in the selected model. Even after normalization, residuals for VOSP silhouettes showed some nonnormality. One outlier was removed from the analysis of CERAD recall, and two outliers were removed from the regression norming of CERAD learning. As expected, Z scores for all tests had a mean of approximately 0 and a standard deviation close to 1.

Change Norms Applied in Symptom Group Sample

Comparing Cognitive Change Between Controls, SCD, and MCI

Clear group differences were observed for CERAD learning, $F(2, 421) = 29.32$, $p < .001$, $\eta^2 = .12$, CERAD recall, $H(2) = 48.82$, $p < .001$, $\eta^2(H) = .10$, and TMT-B, $F(2, 251.59) = 10.711$, $p < .001$, $\eta^2 = .06$. Smaller and less certain differences were found for

Table 2
Scaled Score Conversion Table for Cognitive Test Raw Scores at Baseline and Follow-Up[a]

Scaled	CERAD learning		CERAD recall		TMT-A		TMT-B		FAS		VOSP		Scaled
	Baseline	Follow-up	Baseline	Follow-up	Baseline	Follow-up	Baseline	Follow-up	Baseline	Follow-up	Baseline	Follow-up	
1		≤12			≥66	≥82	≥186	≥229		≤17	≤12	≤12	1
2			≤1			81	171–185	197–228	≤20	18–19	13	13–14	2
3	≤13	13	2–3	≤2	65	75–80	161–170	174–196	21		14	15	3
4	14	14		3	63–64	65–74	159–160	161–173	22	20–21	15	16	4
5	15–16	15	4		58–62	56–64	139–158	139–160	23–24	22–25	16	17	5
6	17	16–17	5	4	54–57	49–55	119–138	130–138	25–26	26–29	17	18	6
7	18	18–19		5	47–53	45–48	108–118	112–129	27–28	30–31	18–19	19–20	7
8	19	20	6	6	41–46	41–44	98–107	97–111	29–31	32–35	20	21	8
9	20	21–22		7	37–40	38–40	88–97	87–96	32–38	36–39	21	22	9
10	21	23	7		34–36	34–37	80–87	79–86	39–41	40–42	22–23	23	10
11	22–23	24	8	8	30–33	30–33	72–79	71–78	42–44	43–45	24	24	11
12	24				27–29	26–29	65–71	62–70	45–48	46–49	25–26	25–26	12
13	25	25–26	9	9	26	24–25	60–64	56–61	49–51	50–54	27	27	13
14		27	10	10	22–25	22–23	52–59	51–55	52–56	55–58			14
15	26				21	21	48–51	45–50	57–62	59–65	28	28	15
16		28			20		43–47	41–44	63–67	66–70	≥29	≥29	16
17	27				19	18–20	41–42	40	68	71–76			17
18	28	29			17–18	16–17	38–40	39	69	77–81			18
19	≥29	30			≤16	≤15	35–37	≤38	≥70	82–83			19
20							≤34			≥84			20

Note. CERAD = The Consortium to Establish a Registry for Alzheimer's Disease; TMT = Trail Making Test; VOSP = Visual Object and Space Perception Battery.
[a] Follow-up assessment approximately 2 years after baseline. Test scores to be converted are raw scores.

Table 3
Normative Regression Models for Cognitive Tests

Variable	*N*	Predictor	β	*SE* β	*T*	*p*	Adj. R^2	Adj. partial R^2	*SD* residual
CERAD learning	178	Intercept	9.02993	2.08172	4.338	<.001	0.3819		2.562279
		Baseline score	0.40935	0.06903	5.930	<.001		0.164	
		Age	−0.09780	0.02315	−4.225	<.001		0.088	
		Education	0.22129	0.05954	3.717	<.001		0.069	
		Sex	0.75077	0.39735	1.889	.061		0.015	
CERAD recall	179	Intercept	6.87192	2.05756	3.340	.001	0.3171		2.45876
		Baseline score	0.42748	0.07211	5.928	<.001		0.163	
		Age	−0.05722	0.02250	−2.544	.012		0.030	
		Education	0.16019	0.05735	2.793	.006		0.037	
		Sex	0.90749	0.37496	2.420	.017		0.027	
TMT-A	182	Intercept	11.69765	1.71643	6.815	<.001	0.4109		2.346442
		Baseline score	0.49043	0.06243	7.856	<.001		0.251	
		Age	−0.10151	0.02160	−4.699	<.001		0.105	
TMT-B	180	Intercept	3.84766	0.61093	6.298	<.001	0.4113		2.336511
		Baseline score	0.62914	0.05604	11.227	<.001			
FAS	122	Intercept	2.39177	0.68416	3.496	<.001	0.5447		2.146246
		Baseline score	0.76805	0.06362	12.073	<.001			
VOSP silhouettes	179	Intercept	5.98706	1.33683	4.479	<.001	0.5933		2.007496
		Baseline score	0.72789	0.04899	14.727	<.001		0.550	
		Age	−0.05498	0.01782	−3.085	.002		0.046	

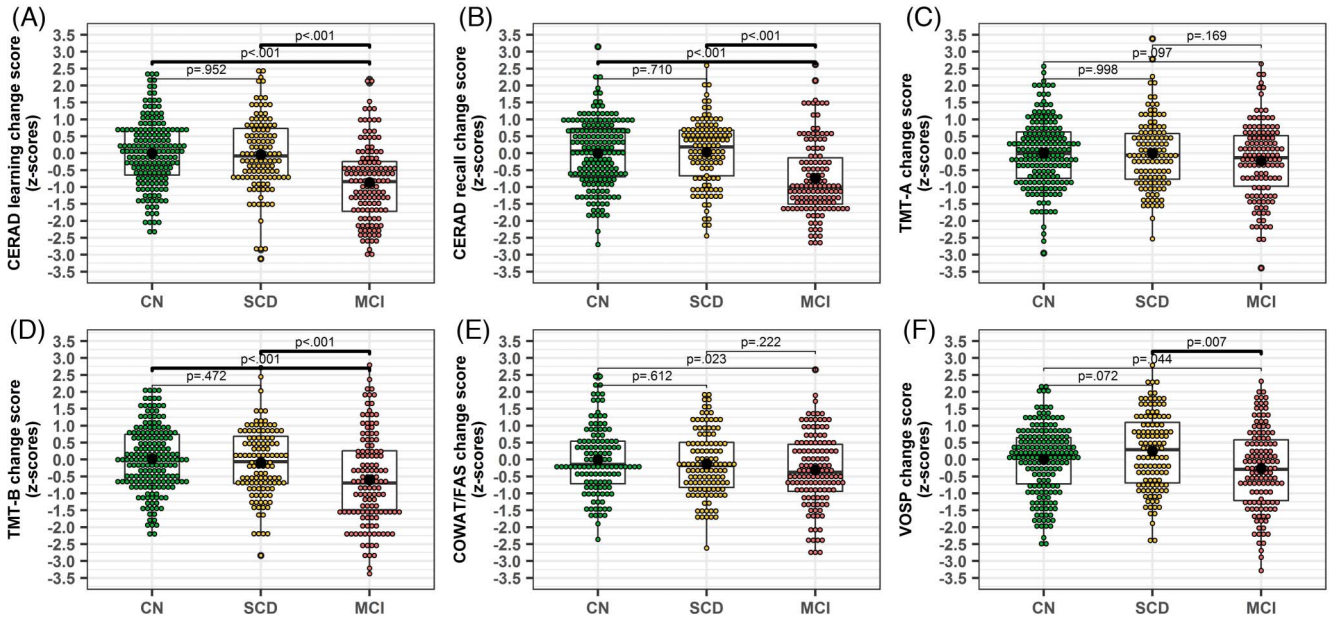
Note. The table presents coefficients for calculation of a predicted 2-year follow-up test score. β = unstandardized regression coefficient; *T* = the *t*-test statistic; Adj. = adusted; *SE* = standard error; CERAD = The Consortium to Establish a Registry for Alzheimer's Disease; TMT = Trail Making Test; VOSP = Visual Object and Space Perception Battery.

COWAT/FAS, $F(2, 365) = 3.602$, $p = .028$, $\eta^2 = .02$, VOSP silhouettes, $H(2) = 10.52$, $p = .005$, $\eta^2(H) = .02$, and TMT-A, $F(2, 425) = 2.494$, $p = .084$, $\eta^2 = .01$. Results from planned comparisons (Figure 1) show that for memory and executive tests, individuals diagnosed with MCI display substantial decline on cognitive measures, compared to participants in the normative group or with SCD.

Comparing Cognitive Change in Biomarker Profile Groups

Dummy regression models (Figure 2) with normal AD biomarker profile (A−T−N−) as reference group indicate negative cognitive change for the group with AD pathology (A+T/N+). Results show

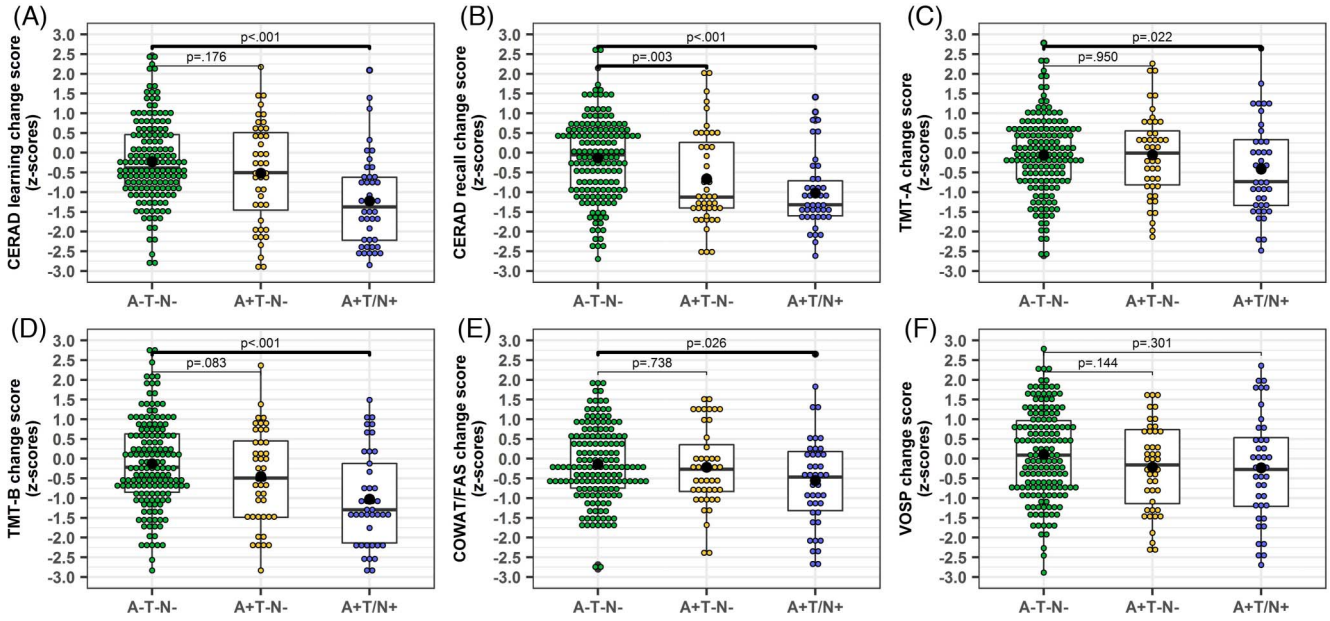
Figure 1
Comparison of Cognitive Change in Normative Controls, Subjective Cognitive Decline, and Mild Cognitive Impairment



Note. Box plots displaying the distribution of change scores (*z* scores) for each cognitive test (Panels A–F) in subgroups with different clinical profiles. CN = cognitively normal; SCD = subjective cognitive decline; MCI = mild cognitive impairment; TMT = Trail Making Test; COWAT = Controlled Word Association Test. CERAD = The Consortium to Establish a Registry for Alzheimer's Disease; VOSP = Visual Object and Space Perception Battery. The line within the box displays the median value, the dot displays the mean. See the online article for the color version of this figure.

Figure 2

Comparison of Cognitive Change in Subgroups With Different Biomarker Profiles



Note. Box plots displaying the distribution of change scores (*z* scores) for each cognitive test (Panels A–F) in subgroups with different CSF biomarker profiles. A = Aβ42/40 ratio; T = p-tau; N = t-tau; TMT = Trail Making Test; CSF = cerebrospinal fluid; COWAT = Controlled Word Association Test; CERAD = The Consortium to Establish a Registry for Alzheimer's Disease; VOSP = Visual Object and Space Perception Battery. The line within the box displays the median value, the dot displays the mean. See the online article for the color version of this figure.

normative decline in the A+T/N+ group for CERAD learning, β = −0.960, *SE* = 0.194, *t*(243) = −4.942, 95% CI [−1.343, −0.578], CERAD recall, β = −0.890, *SE* = 0.182, *t*(243) = −4.894, 95% CI [−1.249, −0.532], and TMT-B, β = −0.908, *SE* = 0.209, *t*(243) = −4.336, 95% CI [−1.321, −0.495]. The point estimates also indicate negative change for TMT-A, β = −0.438, *SE* = 0.190, *t*(243) = −2.305, 95% CI [−0.811, −0.064], and COWAT/FAS, β = −0.397, *SE* = 0.177, *t*(243) = −2.241, 95% CI [−0.745, −0.048], in the A+T/N+ group, but confidence intervals suggest estimates close to no group differences are also compatible with the data. VOSP change scores also indicated decline in the A+T/N+ group, β = −0.218, *SE* = 0.210, *t*(243) = −1.036, 95% CI [−0.632, 0.196], but the uncertainty around the point estimate means the true change could also be zero or positive.

In the group with only amyloid pathology (A+T−N−), we observed a slight negative change compared to the reference group for CERAD recall, β = −0.530, *SE* = 0.179, *t*(243) = −2.966, 95% CI [−0.882, −0.178]. Change scores showed a negative trend in the A+T−N− group for the CERAD learning, TMT-B, and VOSP (range β = −0.359, −0.259), but uncertainty around the point estimates indicates that no or positive change is also possible for our data. Regression diagnostic plots indicated some heteroscedasticity for TMT-A and TMT-B, which may influence inference from results from these tests.

Discussion

Summary and Main Findings

In the present study, we have developed change norms for adults 41–84 years, with a test–retest interval of 2 years, for the six cognitive tests CERAD word list learning and delayed recall, TMT-A, TMT-B, COWAT, and VOSP silhouettes. The norms are adjusted for relevant demographics and practice effects. Results from subsequent analyses applying these norms essentially support the initially presented hypotheses: (a) patients with self-reported cognitive symptoms at baseline showed cognitive decline compared with healthy control participants. However, only the MCI group showed differential cognitive change from the control group. This was observed on tests of episodic learning and memory (CERAD word list) and cognitive flexibility/divided attention (TMT-B); (b) subgroups with biomarkers on the AD continuum (abnormal Aβ, with or without tau and neurodegeneration markers) showed a trend toward negative cognitive change scores; (c) among amyloid positive participants, the group with tau and/or neurodegeneration markers (A+T/N+) showed larger cognitive decline than those without (A+T−N−). In the group with isolated amyloid pathology (A+T−N−), only verbal recall change scores showed significant decline compared to the biomarker negative (A−T−N−) group. In contrast, all cognitive domains except visuospatial ability appeared to decline in the A+T/N+ group, with larger (between 0.8 and 0.9 *SD*) and more certain change score estimates for tests of episodic learning and memory and cognitive flexibility/divided attention.

SRB Change Norms

By incorporating moderating variables into the calculation of cognitive change, regression-based change scores can strengthen or weaken hypotheses regarding true cognitive change. For example, as expected we found an association between negative change scores and CSF biomarkers. Since demographics, practice effects,

and regression to the mean are controlled for in these scores, we can be more confident that the observed cognitive decline reflects disease processes in the brain associated with AD.

Traditionally, alternative approaches to regression-based methods for evaluating change are based on the reliable change index (RCI) formula, calculated by dividing the discrepancy of the baseline and follow-up scores by a measure reflecting the error term of the difference (Duff, 2012). One variation of the RCI also controls for practice effects (Chelune et al., 1993). Typically employing a *z*-score cutoff of ±1.645 to demarcate reliable change, several investigators have compared the regression-based approach with the RCI methodology. Some find that these formulae are largely comparable in their overall ability to identify cognitive change (e.g., Temkin et al., 1999), while others report differences between them, mainly indicating that the SRB methods are preferable (e.g., Maassen et al., 2009).

Currently, there is no consensus on a preferred neuropsychological approach for evaluating cognitive change. A proposed resolution is validating change formulae against clinically relevant external criteria, such as biomarkers (Duff et al., 2019). In a recent article, Duff et al. (2019) found that short-term SRB change formulae showed the strongest relation to AD neuroimaging biomarkers of hippocampal volume and amyloid deposition, compared to RCIs. This finding supports the use of regression-based methods for the present study, where AD biomarkers are an outcome of interest.

Predictors of Change

We investigated the influence of several factors on test scores at follow-up. Higher baseline scores consistently predicted higher follow-up scores and explained the most variance across tests. Individuals starting with more cognitive resources seem to benefit more from prior exposure to the material, as suggested in previous research showing that the "rich get richer" in the case of repeated intelligence testing (Rapport et al., 1997).

For verbal learning and memory tests, lower age, female sex, and more years of education predicted higher performance when retested. We did not find other published regression-based change norms for the CERAD word list tasks used in our study, but studies using other list learning tests report effects of age (Duff et al., 2010; Hammers et al., 2021) and some report no influence of demographics (Attix et al., 2009; De Simone et al., 2020). Past work on the CERAD word list using a slightly different change score methodology (prediction of difference between testing sessions, rather than prediction of follow-up scores), did however use a broad set of demographic variables in regression analyses (Zehnder et al., 2007). Female advantage on verbal tasks is common (e.g., Munro et al., 2012), and our use of this modality for the memory tasks may have influenced findings.

In line with previous findings, we found that older age predicted lower scores at follow-up for most of the remaining tests (Attix et al., 2009; Kiselica et al., 2020; Salthouse, 2010). However, in contrast to our results, other studies using regression-based change norms report no age effects for TMT-A (Duff et al., 2010; Hammers et al., 2021; Sánchez-Benavides et al., 2016) and VOSP silhouettes (in a sample of younger adults; Conradi et al., 2020). Our TMT-B and COWAT/FAS change norms included no demographic predictors. In agreement with our findings, several studies report no demographic predictors of change in letter fluency tasks (Attix et al., 2009; De Simone et al., 2020; Sánchez-Benavides et al., 2016), or TMT-B (Sánchez-Benavides et al., 2016). However, some reports point to broader influence of demographics on TMT-B (Attix et al., 2009; Duff et al., 2010; Hammers et al., 2021; Kiselica et al., 2020) and an education effect in letter fluency (Duff et al., 2010; Hammers et al., 2021; Kiselica et al., 2020).

Evidently, there is no conclusive set of predictors of change for the tests in our battery. The presented variability might be related to factors such as different sample sizes, participant characteristics, and retest intervals of the studies mentioned. These aspects can determine the actual influence of demographics on cognitive change, as well as the likelihood of detecting these effects.

Culture and language may also influence change norms, as with traditional neuropsychological performance norms that can be less accurate when applied in cultural or linguistic groups that differ from the normative sample they were developed in (e.g., Espenes et al., 2020; Lorentzen et al., 2021; Raudeberg et al., 2019). Although we are not aware of studies suggesting cross-cultural influence on change norms, the factors that may contribute to differences in performance between cultures (e.g., language and access to education), could potentially cause different patterns of normative change as well. Therefore, our change norms are likely most adequate for use in Scandinavian (Norwegian and Swedish) adults. As we cannot assume universality without empirical basis, the norms should be applied with caution for individuals with different backgrounds.

Cognitive Change, Clinical Condition, and AD Biomarkers

Our results indicate that broader biomarker pathology in the symptom group is related to decline in several cognitive processes. While only delayed recall scores showed an evident negative change in the amyloid positive group (A+T−N−), the amyloid and tau/neurodegeneration (A+T/N+) group displayed marked declines in CERAD learning and delayed recall, as well as TMT-B scores. Since amnestic symptoms typically occur early in the AD disease trajectory, followed by reductions in other cognitive domains, the broader pattern of cognitive decline in the A+T/N+ group could reflect disease progression. The TMT-B has high cognitive processing demands, engaging basic visuomotor and search skills with more complex executive functions, and is known as a useful clinical indicator of brain dysfunction (Larrabee et al., 2008). These features could make the TMT-B more sensitive to change at high rather than at low levels of cognition (Mura et al., 2014).

Our findings also indicate that clinical disease progression is linked to degree of decline, as significant negative change was present only in the MCI group, not in SCD. Of note, asymptomatic or subjectively impaired individuals may also have an A+T/N+ biomarker profile, but represent a different stage of the clinical disease (preclinical AD) than cognitively impaired individuals (Dubois et al., 2016; Jack et al., 2018).

Our results and previous findings show that neurodegeneration is associated with more prominent decline, while cognitive decline with Aβ pathology alone is more limited. Although evidence supports that Aβ accumulation contributes to development of AD, researchers generally find a low cross-sectional correlation between severity of clinical symptoms and amyloid plaque burden (Bjorkli et al., 2020). The literature indicates that tau and

neurodegeneration often correlate stronger with cognitive symptoms/impairment than Aβ pathology does (e.g., Nelson et al., 2012). These findings may be explained as a result of a synergistic Aβ-tau interplay, where neurotoxic effects of Aβ are mediated by tau pathology (Bjorkli et al., 2020; Hane et al., 2017; Mattsson-Carlgren et al., 2021; Timmers et al., 2019).

Limitations

Limitations of this study include limited sample sizes and inclusion of participants from several cohorts. A more varied normative group could, however, also represent a strength. As with traditional performance norms, the change norms are valid in settings and samples that match the normative sample. Our samples comprise mainly native Norwegians and Swedes, and demographics are influenced by language requirements for study participation. These norms may not be valid for individuals with different linguistic or geographical backgrounds and are not applicable for more than one retest. Moreover, since the change norms are based on healthy control samples, they are most useful for evaluating whether performance is within normative boundaries of cognitively healthy persons or not. They are not developed for assessing expected change within clinical groups.

The mean time interval between testing sessions was shorter in the symptom group (25 months) than in the normative group (28 months). Moreover, both groups displayed variability in retest intervals, and the norms are based on retest intervals spanning 1 to over 4 years. Since the length of the retest interval is a known moderator of cognitive change and practice effects (Calamia et al., 2012; Duff, 2012), this variability in the data could affect the accuracy of our results. However, the retest interval variable was not useful in the regression analyses, suggesting that the time elapsed between assessments did not systematically influence change within the normative sample. The application of these norms on symptom group individuals with atypical retest intervals could still have influenced the estimation of change (e.g., shorter intervals may inflate results in the positive direction, while longer intervals could have a negative effect on change scores).

Biomarker data were not available for all participants, so the normative group might include some individuals with AD biomarkers. Since pathological AD biomarkers in cognitively normal individuals are relatively common, one could argue that the possible inclusion of such persons contributes to a more naturalistic normative group (Kern et al., 2018; Vos et al., 2013). On the other hand, evidence suggests that excluding these individuals from change norm development can improve detection of healthy persons who later progress to MCI or AD (De Santi et al., 2008). Moreover, other brain pathologies besides AD (such as vascular and α-synuclein-related pathologies) are prevalent in aging populations, and mixed pathologies are frequently observed in community-based studies (Rahimi & Kovacs, 2014). Borland et al. (2020) found that excluding all participants with measurable neuropathology (mostly cerebrovascular pathologies) from a traditional normative sample, to a large degree eliminated age-related cognitive decline on several tests, with the exception of TMT-A and TMT-B. This robust norm set provided stricter cutoffs for cognitive impairment. Our study participants may have undetected conditions that could have influenced cognition.

Participants reporting cognitive decline were included in the symptom group, regardless of meeting criteria for objective cognitive impairment or not. Since AD was defined biochemically, this approach allowed us to investigate cognitive change in different biomarker groups irrespective of performance level (e.g., participants with SCD and positive biomarkers may present with normal cognitive performance at both time points, but the change score could detect deviance from expected performance). Including all symptomatic participants increased the sample size and may help us detect nuances in cognitive change. However, including individuals in different stages of disease could have influenced our findings toward small/modest change. The comparisons of the normative group and the MCI and SCD symptom groups show that the SCD participants overall show no difference in change from the normative sample.

Symptom group participants were only from the DDI study, not from the two other normative cohorts. This was done because the DDI study uses CSF $A\beta_{42/40}$ ratio biomarkers, which are preferred for identifying AD (Hansson et al., 2019). Moreover, this sample also gives a complete data set, where all participants have completed all tests. After 2 years, nineteen individuals in the symptom group were diagnosed with probable or possible dementia. The majority ($n = 17$) were diagnosed with AD dementia, one with probable Lewy Body dementia, and one with possible Parkinson's disease dementia. The scores from these individuals may have had an influence on the observed cognitive change. Of note, we do not know how many of the symptom group participants with different A/T/N profiles who over time progress to get a clinical diagnosis of AD dementia.

Clinical Implications and Future Directions

The change norms we present in this study are a helpful tool for assessing whether an individual's cognitive change over 2 years deviates from what is normatively expected, and if this change is considered meaningful. They can be directly applied to test results from Scandinavian patients with similar demographics and retest intervals as the normative group. The web-based calculator facilitates the application of the change norms. The norms are validated in groups with biochemically defined AD, but are also applicable for clinicians and researchers in other fields where longitudinal monitoring or assessment of cognition is relevant. Not accounting for practice effects when retesting is likely to mask decline and might delay detection of clinically significant cognitive change.

By using biomarker signatures when validating the change norms, our study contributes to shedding light on the links between cognitive symptoms and neuropathology in preclinical and prodromal AD. We found negative memory change scores in individuals with positive biomarkers. The observation that scores were only markedly reduced in the A+T/N+ group could indicate that change scores are not sensitive to isolated amyloid pathology, which often precedes tau-related abnormalities in AD.

As mentioned, it is common to use a z-score cutoff of ±1.645 for reliable change (Duff, 2012). Within such a framework, our symptom group sample could be considered cognitively stable. Yet, we believe these more minor observations of change are of interest despite not reaching such a significance cutoff, especially in the context of confirmed biomarker pathology. Previous findings of associations between AD pathology and practice effects (Jutten et al., 2020) point to that one potential area of use for change norms

could be to indicate early AD. To illustrate how subtle cognitive change due to reduced practice effects may present in the clinic, we use an example of two individuals with subjective cognitive symptoms, who have identical demographics and baseline scores, but different biomarker profiles. The first person is biomarker negative and remembers eight words on CERAD recall at baseline, and nine words when retested—possibly due to a practice effect. The second person has both amyloid and tau/neurodegeneration biomarker pathology (A+T/N+) and presents a seemingly stable performance at eight words at both time points. In effect, this latter result might represent subtle cognitive decline masked by a practice effect. This supports the relevance of accounting for practice effects, for instance by using change norms.

Further research on the validity and utility of the change norms in AD, as well as other clinical conditions, would be of interest. For example, if they can contribute to more accurate MCI diagnosis or prediction of decline. One previous study has demonstrated the added value of using both baseline performance and change norms to predict AD dementia (De Santi et al., 2008). Future research could also address cognitive change within clinical groups, modeling trajectories of change in illness progression (Attix et al., 2009). Research on the diagnostic and prognostic value of practice effects in AD may increase the usefulness and validity of using cognitive tests to evaluate change. Moreover, comparisons of change between geographic, linguistic, and/or cultural backgrounds would be an interesting topic for further study, as such investigations could determine the need for local change norms.

Conclusion

The newly developed change norms for CERAD word list learning and delayed recall, TMT-A, TMT-B, COWAT, and VOSP silhouettes can be clinically useful to determine meaningful change, as they account for pertinent demographics and practice effects. An online change norm calculator is freely available for use. The norms are validated in a group with cognitive complaints and different AD biomarker profiles.

References

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., Gamst, A., Holtzman, D. M., Jagust, W. J., Petersen, R. C., Snyder, P. J., Carrillo, M. C., Thies, B., & Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, *7*(3), 270–279. https://doi.org/10.1016/j.jalz.2011.03.008

Attix, D. K., Story, T. J., Chelune, G. J., Ball, J. D., Stutts, M. L., Hart, R. P., & Barth, J. T. (2009). The prediction of change: Normative neuropsychological trajectories. *The Clinical Neuropsychologist*, *23*(1), 21–38. https://doi.org/10.1080/13854040801945078

Benton, A., & Hamsher, K. (1978). *Multilingual aphasia examination manual*. University of Iowa.

Bjorkli, C., Sandvig, A., & Sandvig, I. (2020). Bridging the gap between fluid biomarkers for Alzheimer's disease, model systems, and patients. *Frontiers in Aging Neuroscience*, *12*, Article 272. https://doi.org/10.3389/fnagi.2020.00272

Borland, E., Stomrud, E., van Westen, D., Hansson, O., & Palmqvist, S. (2020). The age-related effect on cognitive performance in cognitively healthy elderly is mainly caused by underlying AD pathology or cerebrovascular lesions: Implications for cutoffs regarding cognitive impairment. *Alzheimer's Research & Therapy*, *12*(1), Article 30. https://doi.org/10.1186/s13195-020-00592-8

Brooks, B. L., Holdnack, J. A., & Iverson, G. L. (2016). To change is human: "Abnormal" reliable change memory scores are common in healthy adults and older adults. *Archives of Clinical Neuropsychology*, *31*(8), 1026–1036. https://doi.org/10.1093/arclin/acw079

Calamia, M., Markon, K., & Tranel, D. (2012). Scoring higher the second time around: Meta-analyses of practice effects in neuropsychological assessment. *The Clinical Neuropsychologist*, *26*(4), 543–570. https://doi.org/10.1080/13854046.2012.680913

Chelune, G. J., Naugle, R. I., Lüders, H., Sedlak, J., & Awad, I. A. (1993). Individual change after epilepsy surgery: Practice effects and base-rate information. *Neuropsychology*, *7*(1), 41–52. https://doi.org/10.1037/0894-4105.7.1.41

Conradi, N., Behrens, M., Hermsen, A. M., Kannemann, T., Merkel, N., Schuster, A., Freiman, T. M., Strzelczyk, A., & Rosenow, F. (2020). Assessing cognitive change and quality of life 12 months after epilepsy surgery-development and application of reliable change indices and standardized regression-based change norms for a neuropsychological test battery in the German language. *Frontiers in Psychology*, *11*(2679), Article 582836. https://doi.org/10.3389/fpsyg.2020.582836

De Santi, S., Pirraglia, E., Barr, W., Babb, J., Williams, S., Rogers, K., Glodzik, L., Brys, M., Mosconi, L., Reisberg, B., Ferris, S., & de Leon, M. J. (2008). Robust and conventional neuropsychological norms: Diagnosis and prediction of age-related cognitive decline. *Neuropsychology*, *22*(4), 469–484. https://doi.org/10.1037/0894-4105.22.4.469

De Simone, M. S., Perri, R., Rodini, M., Fadda, L., De Tollis, M., Caltagirone, C., & Carlesimo, G. A. (2020). A lack of practice effects on memory tasks predicts conversion to Alzheimer Disease in patients with amnestic mild cognitive impairment. *Journal of Geriatric Psychiatry and Neurology*, *34*(6), 582–593. https://doi.org/10.1177/0891988720944244

Dubois, B., Hampel, H., Feldman, H. H., Scheltens, P., Aisen, P., Andrieu, S., Bakardjian, H., Benali, H., Bertram, L., Blennow, K., Broich, K., Cavedo, E., Crutch, S., Dartigues, J. F., Duyckaerts, C., Epelbaum, S., Frisoni, G. B., Gauthier, S., Genthon, R., … the Proceedings of the Meeting of the International Working Group (IWG) and the American Alzheimer's Association on "The Preclinical State of AD"; July 23, 2015; Washington DC, USA. (2016). Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. *Alzheimer's & Dementia*, *12*(3), 292–323. https://doi.org/10.1016/j.jalz.2016.02.002

Duff, K. (2012). Evidence-based indicators of neuropsychological change in the individual patient: Relevant concepts and methods. *Archives of Clinical Neuropsychology*, *27*(3), 248–261. https://doi.org/10.1093/arclin/acr120

Duff, K., Beglinger, L. J., Moser, D. J., Paulsen, J. S., Schultz, S. K., & Arndt, S. (2010). Predicting cognitive change in older adults: The relative contribution of practice effects. *Archives of Clinical Neuropsychology*, *25*(2), 81–88. https://doi.org/10.1093/arclin/acp105

Duff, K., Suhrie, K. R., Dalley, B. C. A., Anderson, J. S., & Hoffman, J. M. (2019). External validation of change formulae in neuropsychology with neuroimaging biomarkers: A methodological recommendation and preliminary clinical data. *The Clinical Neuropsychologist*, *33*(3), 478–489. https://doi.org/10.1080/13854046.2018.1484518

Elman, J. A., Jak, A. J., Panizzon, M. S., Tu, X. M., Chen, T., Reynolds, C. A., Gustavson, D. E., Franz, C. E., Hatton, S. N., Jacobson, K. C., Toomey, R., McKenzie, R., Xian, H., Lyons, M. J., & Kremen, W. S. (2018). Underdiagnosis of mild cognitive impairment: A consequence of ignoring practice effects. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *10*(1), 372–381. https://doi.org/10.1016/j.dadm.2018.04.003

Espenes, J., Hessen, E., Eliassen, I. V., Waterloo, K., Eckerström, M., Sando, S. B., Timon, S., Wallin, A., Fladby, T., & Kirsebom, B.-E. (2020). Demographically adjusted trail making test norms in a Scandinavian sample from 41 to 84 years. *The Clinical Neuropsychologist*, *34*(Suppl. 1), 110–126. https://doi.org/10.1080/13854046.2020.1829068

Fillenbaum, G. G., van Belle, G., Morris, J. C., Mohs, R. C., Mirra, S. S., Davis, P. C., Tariot, P. N., Silverman, J. M., Clark, C. M., Welsh-Bohmer, K. A., & Heyman, A. (2008). Consortium to establish a registry for Alzheimer's Disease (CERAD): The first twenty years. *Alzheimer's & Dementia*, *4*(2), 96–109. https://doi.org/10.1016/j.jalz.2007.08.005

Fladby, T., Pålhaugen, L., Selnes, P., Waterloo, K., Bråthen, G., Hessen, E., Almdahl, I. S., Arntzen, K. A., Auning, E., Eliassen, C. F., Espenes, R., Grambaite, R., Grøntvedt, G. R., Johansen, K. K., Johnsen, S. H., Kalheim, L. F., Kirsebom, B. E., Müller, K. I., Nakling, A. E., … Aarsland, D. (2017). Detecting at-risk Alzheimer's Disease cases. *Journal of Alzheimer's Disease*, *60*(1), 97–105. https://doi.org/10.3233/JAD-170231

Goldberg, T. E., Harvey, P. D., Wesnes, K. A., Snyder, P. J., & Schneider, L. S. (2015). Practice effects due to serial cognitive assessment: Implications for preclinical Alzheimer's disease randomized controlled trials. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *1*(1), 103–111. https://doi.org/10.1016/j.dadm.2014.11.003

Grøntvedt, G. R., Lauridsen, C., Berge, G., White, L. R., Salvesen, Ø., Bråthen, G., & Sando, S. B. (2020). The amyloid, tau, and neurodegeneration (A/T/N) classification applied to a clinical research cohort with long-term follow-up. *Journal of Alzheimer's Disease*, *74*(3), 829–837. https://doi.org/10.3233/JAD-191227

Hammers, D. B., Porter, S., Dixon, A., Suhrie, K. R., & Duff, K. (2021). Validating 1-year reliable change methods. *Archives of Clinical Neuropsychology*, *36*(1), 87–98. https://doi.org/10.1093/arclin/acaa055

Hammers, D. B., Suhrie, K. R., Porter, S. M., Dixon, A. M., & Duff, K. (2020). Generalizability of reliable change equations for the RBANS over one year in community-dwelling older adults. *Journal of Clinical and Experimental Neuropsychology*, *42*(4), 394–405. https://doi.org/10.1080/13803395.2020.1740654

Hane, F. T., Lee, B. Y., & Leonenko, Z. (2017). Recent progress in Alzheimer's disease research, part 1: Pathology. *Journal of Alzheimer's Disease*, *57*(1), 1–28. https://doi.org/10.3233/JAD-160882

Hansson, O., Lehmann, S., Otto, M., Zetterberg, H., & Lewczuk, P. (2019). Advantages and disadvantages of the use of the CSF Amyloid β (Aβ) 42/40 ratio in the diagnosis of Alzheimer's Disease. *Alzheimer's Research & Therapy*, *11*(1), Article 34. https://doi.org/10.1186/s13195-019-0485-0

Hassenstab, J., Ruvolo, D., Jasielec, M., Xiong, C., Grant, E., & Morris, J. C. (2015). Absence of practice effects in preclinical Alzheimer's disease. *Neuropsychology*, *29*(6), 940–948. https://doi.org/10.1037/neu0000208

Heilbronner, R. L., Sweet, J. J., Attix, D. K., Krull, K. R., Henry, G. K., & Hart, R. P. (2010). Official position of the American Academy of Clinical Neuropsychology on serial neuropsychological assessments: The utility and challenges of repeat test administrations in clinical and forensic contexts. *The Clinical Neuropsychologist*, *24*(8), 1267–1278. https://doi.org/10.1080/13854046.2010.526785

Jack, C. R., Jr., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., Holtzman, D. M., Jagust, W., Jessen, F., Karlawish, J., Liu, E., Molinuevo, J. L., Montine, T., Phelps, C., Rankin, K. P., Rowe, C. C., Scheltens, P., Siemers, E., Snyder, H. M., … Silverberg, N. (2018). NIA-AA research framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia*, *14*(4), 535–562. https://doi.org/10.1016/j.jalz.2018.02.018

Jessen, F., Amariglio, R. E., van Boxtel, M., Breteler, M., Ceccaldi, M., Chételat, G., Dubois, B., Dufouil, C., Ellis, K. A., van der Flier, W. M., Glodzik, L., van Harten, A. C., de Leon, M. J., McHugh, P., Mielke, M. M., Molinuevo, J. L., Mosconi, L., Osorio, R. S., Perrotin, A., … the Subjective Cognitive Decline Initiative (SCD-I) Working Group. (2014). A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. *Alzheimer's & Dementia*, *10*(6), 844–852. https://doi.org/10.1016/j.jalz.2014.01.001

Jutten, R. J., Grandoit, E., Foldi, N. S., Sikkes, S. A. M., Jones, R. N., Choi, S. E., Lamar, M. L., Louden, D. K. N., Rich, J., Tommet, D., Crane, P. K., & Rabin, L. A. (2020). Lower practice effects as a marker of cognitive performance and dementia risk: A literature review. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *12*(1), Article e12055. https://doi.org/10.1002/dad2.12055

Kern, S., Zetterberg, H., Kern, J., Zettergren, A., Waern, M., Höglund, K., Andreasson, U., Wetterberg, H., Börjesson-Hanson, A., Blennow, K., & Skoog, I. (2018). Prevalence of preclinical Alzheimer disease: Comparison of current classification systems. *Neurology*, *90*(19), e1682–e1691. https://doi.org/10.1212/WNL.0000000000005476

Kiselica, A. M., Kaser, A. N., Webber, T. A., Small, B. J., & Benge, J. F. (2020). Development and preliminary validation of standardized regression-based change scores as measures of transitional cognitive decline. *Archives of Clinical Neuropsychology*, *35*(7), 1168–1181. https://doi.org/10.1093/arclin/acaa042

Larrabee, G. J., Millis, S. R., & Meyers, J. E. (2008). Sensitivity to brain dysfunction of the Halstead-Reitan vs an ability-focused neuropsychological battery. *The Clinical Neuropsychologist*, *22*(5), 813–825. https://doi.org/10.1080/13854040701625846

Lorentzen, I. M., Espenes, J., Hessen, E., Waterloo, K., Bråthen, G., Timón, S., Aarsland, D., Fladby, T., & Kirsebom, B.-E. (2021). Regression-based norms for the FAS phonemic fluency test for ages 40–84 based on a Norwegian sample. *Applied Neuropsychology. Adult*. Advance online publication. https://doi.org/10.1080/23279095.2021.1918128

Maassen, G. H., Bossema, E., & Brand, N. (2009). Reliable change and practice effects: Outcomes of various indices compared. *Journal of Clinical and Experimental Neuropsychology*, *31*(3), 339–352. https://doi.org/10.1080/13803390802169059

Machulda, M. M., Hagen, C. E., Wiste, H. J., Mielke, M. M., Knopman, D. S., Roberts, R. O., Vemuri, P., Lowe, V. J., Jack, C. R., Jr., & Petersen, R. C. (2017). Practice effects and longitudinal cognitive change in clinically normal older adults differ by Alzheimer imaging biomarker status. *The Clinical Neuropsychologist*, *31*(1), 99–117. https://doi.org/10.1080/13854046.2016.1241303

Machulda, M. M., Pankratz, V. S., Christianson, T. J., Ivnik, R. J., Mielke, M. M., Roberts, R. O., Knopman, D. S., Boeve, B. F., Petersen, R. C., & the Incident Mild Cognitive Impairment and Dementia in The Mayo Clinic Study of Aging. (2013). Practice effects and longitudinal cognitive change in normal aging vs. incident mild cognitive impairment and dementia in the Mayo Clinic study of aging. *The Clinical Neuropsychologist*, *27*(8), 1247–1264. https://doi.org/10.1080/13854046.2013.836567

Mattsson-Carlgren, N., Janelidze, S., Bateman, R. J., Smith, R., Stomrud, E., Serrano, G. E., Reiman, E. M., Palmqvist, S., Dage, J. L., Beach, T. G., & Hansson, O. (2021). Soluble P-tau217 reflects amyloid and tau pathology and mediates the association of amyloid with tau. *EMBO Molecular Medicine*, *13*(6), Article e14022. https://doi.org/10.15252/emmm.202114022

McSweeny, A. J., Naugle, R. I., Chelune, G. J., & Lüders, H. (1993). "T scores for change": An illustration of a regression approach to depicting change in clinical neuropsychology. *Clinical Neuropsychologist*, *7*(3), 300–312. https://doi.org/10.1080/13854049308401901

Mormino, E. C., Betensky, R. A., Hedden, T., Schultz, A. P., Amariglio, R. E., Rentz, D. M., Johnson, K. A., & Sperling, R. A. (2014). Synergistic effect of β-amyloid and neurodegeneration on cognitive decline in clinically normal individuals. *JAMA Neurology*, *71*(11), 1379–1385. https://doi.org/10.1001/jamaneurol.2014.2031

Munro, C. A., Winicki, J. M., Schretlen, D. J., Gower, E. W., Turano, K. A., Muñoz, B., Keay, L., Bandeen-Roche, K., & West, S. K. (2012). Sex differences in cognition in healthy elderly individuals. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition*, *19*(6), 759–768. https://doi.org/10.1080/13825585.2012.690366

Mura, T., Proust-Lima, C., Jacqmin-Gadda, H., Akbaraly, T. N., Touchon, J., Dubois, B., & Berr, C. (2014). Measuring cognitive change in subjects with prodromal Alzheimer's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, *85*(4), 363–370. https://doi.org/10.1136/jnnp-2013-305078

Nelson, P. T., Alafuzoff, I., Bigio, E. H., Bouras, C., Braak, H., Cairns, N. J., Castellani, R. J., Crain, B. J., Davies, P., Del Tredici, K., Duyckaerts, C.,

Frosch, M. P., Haroutunian, V., Hof, P. R., Hulette, C. M., Hyman, B. T., Iwatsubo, T., Jellinger, K. A., Jicha, G. A., … Beach, T. G. (2012). Correlation of Alzheimer disease neuropathologic changes with cognitive status: A review of the literature. *Journal of Neuropathology and Experimental Neurology*, *71*(5), 362–381. https://doi.org/10.1097/NEN.0b013e31825018f7

R Core Team. (2020). *R: A language and environment for statistical computing*. https://www.R-project.org/

Rahimi, J., & Kovacs, G. G. (2014). Prevalence of mixed pathologies in the aging brain. *Alzheimer's Research & Therapy*, *6*(9), Article 82. https://doi.org/10.1186/s13195-014-0082-1

Rapport, L. J., Brines, D. B., Theisen, M. E., & Axelrod, B. N. (1997). Full scale IQ as mediator of practice effects: The rich get richer. *The Clinical Neuropsychologist*, *11*(4), 375–380. https://doi.org/10.1080/13854049708400466

Raudeberg, R., Iverson, G. L., & Hammar, Å. (2019). Norms matter: U.S. normative data under-estimate cognitive deficits in Norwegians with schizophrenia spectrum disorders. *The Clinical Neuropsychologist*, *33*-(Suppl. 1), 58–74. https://doi.org/10.1080/13854046.2019.1590641

Reitan, R. M., & Wolfson, D. (1985). *The Halstead-Reitan neuropsychological test battery: Theory and clinical interpretation* (Vol. 4). Neuropsychology Press.

Salthouse, T. A. (2010). Influence of age on practice effects in longitudinal neurocognitive change. *Neuropsychology*, *24*(5), 563–572. https://doi.org/10.1037/a0019026

Sánchez-Benavides, G., Peña-Casanova, J., Casals-Coll, M., Gramunt, N., Manero, R. M., Puig-Pijoan, A., Aguilar, M., Robles, A., Antúnez, C., Frank-García, A., Fernández-Martínez, M., Blesa, R., Oliva, R., Gómez-Ansón, B., Monte, G., Alayrach, E., Sainz, A., CAprile, C., Sánchez-Benavides, G., … the NEURONORMA Study Team. (2016). One-year reference norms of cognitive change in spanish old adults: Data from the NEURONORMA sample. *Archives of Clinical Neuropsychology*, *31*(4), 378–388. https://doi.org/10.1093/arclin/acw018

Siafarikas, N., Kirsebom, B.-E., Srivastava, D. P., Eriksson, C. M., Auning, E., Hessen, E., Selbaek, G., Blennow, K., Aarsland, D., & Fladby, T. (2021). Cerebrospinal fluid markers for synaptic function and Alzheimer type changes in late life depression. *Scientific Reports*, *11*(1), Article 20375. https://doi.org/10.1038/s41598-021-99794-9

Sjögren, M., Vanderstichele, H., Agren, H., Zachrisson, O., Edsbagge, M., Wikkelsø, C., Skoog, I., Wallin, A., Wahlund, L. O., Marcusson, J., Nägga, K., Andreasen, N., Davidsson, P., Vanmechelen, E., & Blennow, K. (2001). Tau and Abeta42 in cerebrospinal fluid from healthy adults 21–93 years of age: Establishment of reference values. *Clinical Chemistry*, *47*(10), 1776–1781. https://doi.org/10.1093/clinchem/47.10.1776

Temkin, N. R., Heaton, R. K., Grant, I., & Dikmen, S. S. (1999). Detecting significant change in neuropsychological test performance: A comparison of four models. *Journal of the International Neuropsychological Society*, *5*(4), 357–369. https://doi.org/10.1017/S1355617799544068

Timmers, M., Tesseur, I., Bogert, J., Zetterberg, H., Blennow, K., Börjesson-Hanson, A., Baquero, M., Boada, M., Randolph, C., Tritsmans, L., Van Nueten, L., Engelborghs, S., & Streffer, J. R. (2019). Relevance of the interplay between amyloid and tau for cognitive impairment in early Alzheimer's disease. *Neurobiology of Aging*, *79*, 131–141. https://doi.org/10.1016/j.neurobiolaging.2019.03.016

Vos, S. J. B., Xiong, C., Visser, P. J., Jasielec, M. S., Hassenstab, J., Grant, E. A., Cairns, N. J., Morris, J. C., Holtzman, D. M., & Fagan, A. M. (2013). Preclinical Alzheimer's disease and its outcome: A longitudinal cohort study. *Lancet Neurology*, *12*(10), 957–965. https://doi.org/10.1016/S1474-4422(13)70194-7

Wallin, A., Nordlund, A., Jonsson, M., Lind, K., Edman, Å., Göthlin, M., Stålhammar, J., Eckerström, M., Kern, S., Börjesson-Hanson, A., Carlsson, M., Olsson, E., Zetterberg, H., Blennow, K., Svensson, J., Öhrfelt, A., Bjerke, M., Rolstad, S., & Eckerström, C. (2016). The Gothenburg MCI study: Design and distribution of Alzheimer's disease and subcortical vascular disease diagnoses from baseline to 6-year follow-up. *Journal of Cerebral Blood Flow and Metabolism*, *36*(1), 114–131. https://doi.org/10.1038/jcbfm.2015.147

Wang, G., Kennedy, R. E., Goldberg, T. E., Fowler, M. E., Cutter, G. R., & Schneider, L. S. (2020). Using practice effects for targeted trials or subgroup analysis in Alzheimer's disease: How practice effects predict change over time. *PLOS ONE*, *15*(2), Article e0228064. https://doi.org/10.1371/journal.pone.0228064

Warrington, E. K., & James, M. (1991). *The visual object and space perception battery*. Thames Valley Test Company.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.

Zehnder, A. E., Bläsi, S., Berres, M., Spiegel, R., & Monsch, A. U. (2007). Lack of practice effects on neuropsychological tests as early cognitive markers of Alzheimer disease? *American Journal of Alzheimer's Disease and Other Dementias*, *22*(5), 416–426. https://doi.org/10.1177/1533317507302448

Received January 27, 2022
Revision received June 8, 2022
Accepted July 2, 2022 ■