Demographically adjusted CERAD wordlist test norms in a Norwegian sample from 40 to 80 years

Bjørn-Eivind Kirsebom, Ragna Espenes, Erik Hessen, Knut Waterloo, Stein Harald Johnsen, Elisabeth Gundersen, Sigrid Botne Sando, Gøril Rolfseng Grøntvedt, Santiago Timón & Tormod Fladby

Published online: 08 Mar 2019.

Submit your article to this journal ⬀

Article views: 371

View Crossmark data ⬀

Demographically adjusted CERAD wordlist test norms in a Norwegian sample from 40 to 80 years

Bjørn-Eivind Kirsebom[a,b], Ragna Espenes[a,b], Erik Hessen[c,d], Knut Waterloo[a,b], Stein Harald Johnsen[a,e], Elisabeth Gundersen[a,b], Sigrid Botne Sando[f,g], Gøril Rolfseng Grøntvedt[f,g], Santiago Timón[c,h] and Tormod Fladby[c,i]

[a]Department of Neurology, University Hospital of North Norway, Tromsø, Norway; [b]Department of Psychology Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway; [c]Department of Neurology, Akershus University Hospital, Lørenskog, Norway; [d]Department of Psychology, University of Oslo, Oslo, Norway; [e]Department of Clinical Medicine, Brain and Circulation Research Group, UiT The Arctic University of Norway, Tromsø, Norway; [f]Department of Neuromedicine and Movement Science Faculty of Medicine and Health, Sciences Norwegian University of Science and Technology, Trondheim, Norway; [g]Department of Neurology and Clinical Neurophysiology, University Hospital of Trondheim, Trondheim, Norway; [h]Departamento de Inteligencia Artificial Universidad Nacional de Educación a Distancia, Madrid, Spain; [i]Institute of Clinical Medicine, Campus Ahus, University of Oslo, Oslo, Norway

ABSTRACT

Background/Objective: In recent years, several slightly younger cohorts have been established in order to study the preclinical and prodromal phases of dementia. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD) wordlist memory test (WLT) is widely used in dementia research. However, culturally adapted and demographically adjusted test norms for younger ages are lacking.

Method: This paper investigates effects of age, gender and years of education on test performance and offers demographically adjusted norms for the CERAD WLT using a regression-based norming procedure for the age span 40–80 years based on healthy controls ($n = 227$) from the Norwegian "Dementia Disease Initiation" (DDI) ($n = 168$) and "Trønderbrain" ($n = 59$) cohorts. In order to evaluate normative performance, we apply the norms to an independent sample of persons diagnosed with mild cognitive impairment (MCI = 168) and perform multiple regression analyses to evaluate adjustment of pertinent demographics.

Results: CERAD WLT norms adjusted for effects of age, gender and educational level are proposed. The norms successfully adjusted for effects of age, gender and education in an independent sample of Norwegians with MCI.

Conclusion: Demographically adjusted norms for the CERAD WLT for ages 40–80 years based on a Norwegian sample are proposed. To our knowledge, this is the first normative study of this test to offer demographically adjusted norms for this age span.

CONTACT Bjørn-Eivind Kirsebom ✉ bjorn-eivind.kirsebom@unn.no 🖅 Department of Neurology, University Hospital of North Norway, UNN 9037 Tromsø, Norway.

Supplemental data for this article is available online at https://doi.org/10.1080/13854046.2019.1574902.

Introduction

The Consortium to Establish a Registry for Alzheimer's Disease (CERAD) was founded to standardize procedures for the evaluation and diagnosis of patients with Alzheimer's Disease (AD) (Morris et al., 1989). The instruments developed by CERAD have been widely used and have been translated into several languages and validated within different cultural contexts (Fillenbaum et al., 2008). Its clinical utility has mainly focused on detecting mild cognitive impairment (MCI) and AD with dementia. Normative data are primarily developed for elderly cohorts (Beeri et al., 2006; Fillenbaum et al., 2005; Schmidtke & Hermeneit, 2008; Sotaniemi et al., 2012; Welsh et al., 1994). However, it has been shown that AD develops over 10–15 years before clinical cognitive impairment is evident (Bateman et al., 2012; Jack et al., 2018). Thus, a major focus in dementia research has shifted to the asymptomatic or preclinical stages (Sperling et al., 2011). In order to capture disease events in these stages, several slightly younger cohorts have been established (Soldan et al., 2013; Weiner et al., 2015). To capture individual cognitive decline and significant treatment effects at these stages, more narrow and culturally adapted norms for cognitive tests including CERAD subtests may need to be established. Recently, Hankee et al. (2016) proposed norms for the CERAD Word list test (WLT) for younger and middle-aged adults based on an American sample. These norms are primarily provided for younger persons (<55 years) and are adjusted for either age or education. However, as learning and memory are influenced by age, education and gender (Beeri et al., 2006; Liu et al., 2011) correction for additional demographic factors may be necessary in order to avoid misclassification of cognitively normal and impaired individuals. In addition, CERAD WLT norms developed for Scandinavian countries (Danish, Swedish or Norwegian language) are lacking.

The use of discrete norms (e.g. capturing the normative performance of a certain demographic) requires an adequate sample size in order to ensure that the reference group is a representative sample of the population distribution. When adjusting for several demographic characteristics such as gender, age and education, the sample size requirement increases dramatically (Oosterhuis, van der Ark, & Sijtsma, 2016). Moreover, norm-based performance may increase substantially by moving from one age category to the next, due to distinct differences between normative reference groups (e.g. moving from a 54–59 year group to 60–65 year group) (Zachary & Gorsuch, 1985). Continuous norms employing regression-based norming procedures offer a possible solution to these issues by requiring 2.5–5.5 times smaller sample size (Oosterhuis et al., 2016) while offering the possibility for continuous adjustment of multiple demographic variables such as age, gender and years of education.

We propose norms adjusted for age, gender and years of education based on a regression-based norming procedure using the normative performance of healthy controls ($n = 227$) aged 40–80 years from two established prospective Norwegian cohorts, investigating preclinical and prodromal dementia. A primary utility of these norms is to detect cognitive decline not caused by normal aging or expected performance differences due to gender or educational attainment. Thus, to evaluate the regression-based norms, we calculate T scores in a group of Norwegian speaking patients ($n = 168$) aged 40–80 years previously diagnosed with MCI from the Dementia Disease

Initiation (DDI) cohort and fit regression models to confirm that the norms reliably adjust for demographic variables when applied to an independent sample.

Methods and materials

The DDI cohort employs a standardized protocol for participant selection and assessment and includes healthy controls, as well as participants with subjective cognitive decline (SCD) and MCI. Healthy control subjects were recruited primarily from spouses of symptomatic participants (SCD or MCI), and secondarily from volunteers responding to advertisements in media, newspapers, or news bulletins. The cohort was recruited between 2013 and 2018. Criteria for inclusion were age between 40 and 80 years and a native language of Norwegian, Swedish, or Danish. Exclusion criteria were brain trauma or disorder, including clinical stroke, dementia, severe psychiatric disorder, severe somatic disease that might influence cognitive functions, or intellectual disability or other developmental disorders. At the time of analysis, 168 fluent Norwegian speakers, 166 (98.8%) Norwegian native and 2 (1.2%) Swedish native healthy controls were included ($n = 168$). A subset of these controls ($n = 23$) were missing total sum scores on the CERAD WLT 20-item recognition subtest (total true positives and true negatives) due to the scoring of only true positives (10-item score). These subjects were removed from further analysis, leaving a total of $n = 145$ with total sum scores on this test. The healthy controls were all able to complete the CERAD WLT in accordance with test instructions (detailed below). In order to evaluate the regression-based norms in an independent sample, we included 168 fluent Norwegian speakers, 167 (99.4%) native Norwegian and 1 (0.6%) native Danish participants from the DDI cohort previously diagnosed with MCI. MCI cases from the Trønderbrain cohort was not included in this analysis since the sample was smaller compared to the MCI sample from the DDI cohort, and slightly different cognitive tests and diagnostic algorithms for classification of MCI diagnosis were used (Berge et al., 2016; Fladby et al., 2017). In the DDI cohort, MCI was determined according to published criteria (Albert et al., 2011; Petersen, 2004), and cases were classified as cognitively impaired when obtaining a score ≤ 1.5 standard deviation below the normative mean on CERAD word list delayed recall (using norms from Sotaniemi et al. (2012)), Visual Object and Space Perception Battery (VOSP) silhouettes (Warrington & James, 1991), Trail Making Test B (TMT-B) or Controlled Oral Word Association test (COWAT) (Heaton, Miller, Taylor, & Grant, 2004). Cognitive functioning was also assessed by the Clinical Dementia Rating scale (CDR) (Hughes, Berg, Danziger, Coben, & Martin, 1982). Participants with dementia were excluded if CDR was >0.5 (Petersen, 2004). For further description of the DDI cohort and methods, see Fladby et al. (2017).

The Trønderbrain cohort recruited participants with MCI, early AD dementia and healthy controls between 2009 and 2015. Healthy controls were recruited from societies for retired people in central Norway, or spouses of recruited MCI or early AD dementia participants. At the time of analysis, 59 healthy controls with Norwegian native language aged 57–80 years were included. They were all able to complete the CERAD WLT in accordance with test instructions (detailed below). The CERAD recognition subtest was not administered and normative data for this subtest is therefore
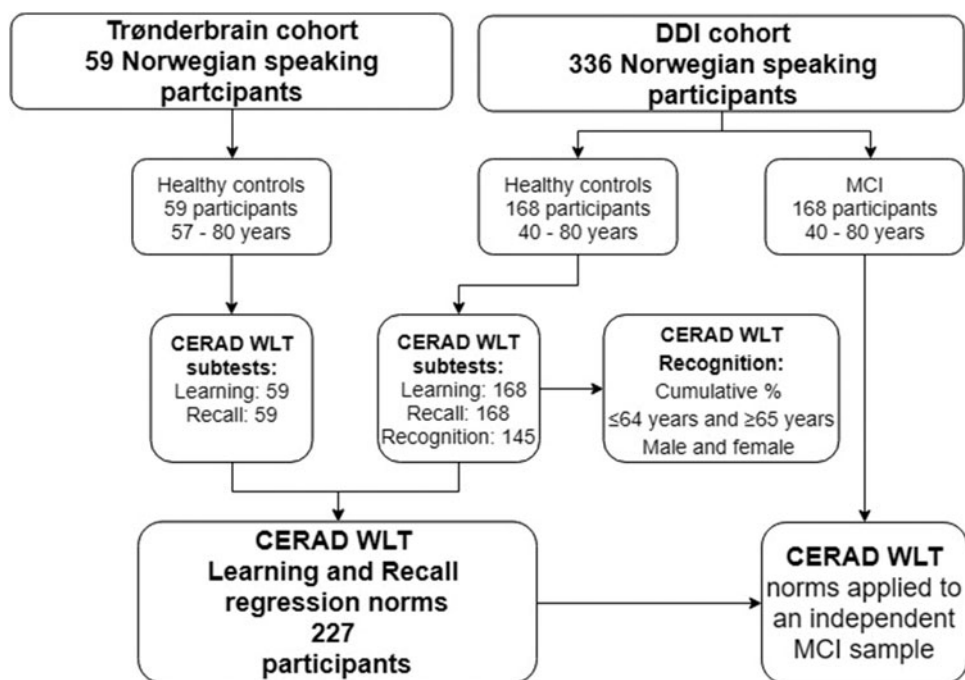
Figure 1. Flowchart depicting an outline of the participant inclusion process from the Trønderbrain and DDI cohorts and workflow of the paper.

only available from the DDI cohort. Exclusion criteria were a present psychiatric or malignant disease (i.e. currently undergoing treatment for cancer), use of anticoagulant medication, or high alcohol consumption. For further description of the Trønderbrain cohort and methods, see Berge et al. (2016). An outline of the participant inclusion process is depicted in Figure 1.

Participants were recruited and assessed from several university hospitals across Norway. This covered a representative sample of residents in cities and nearby rural areas from northern Norway (Troms and Finnmark, $n = 10$), mid-Norway ($n = 66$, Trondheim), south-east Norway ($n = 58$, Akershus/Oslo), south-west Norway ($n = 43$, Bergen and $n = 50$, Stavanger/Haugesund).

CERAD word list test version and administration

The CERAD WLT was translated to Norwegian by Liv Barnett in 2004 using the 10-item word list from the original CERAD test description. The Norwegian version was sourced from CERAD (Gerda Fillenbaum, PhD) at the Center for the Study of Aging and Human Development, USA (Fillenbaum et al., 2008). The following words (and Norwegian translation) were used: Queen (Dronning), Grass (Gress), Arm (Arm), Cabin (Hytte), Pole (Stokk), Shore (kyst), Butter (Smør), Engine (Motor), Ticket (Billett), Letter (Brev). The test was administered by healthcare professionals (medical doctors, nurses and psychologists) at the different sites. Beforehand, all had received training in the administration of the CERAD WLT by the DDI projects senior/chief neuropsychologist, prof. Erik Hessen.

The CERAD WLT learning score was obtained from the sum of three 10-word learning trials yielding a maximum score of 30. After 10 min, participants were asked to recall the words from the 10-word learning trials, yielding a maximum score of 10. We did not record intrusions or perseverations. Finally, a recognition trial comprising 20 words was administered where 10 of the words were distractors, and 10 words were target items from the 10-word learning list. This yielded a maximum score of 20 (10 true positive and 10 true negative responses). However, total sum of false positives was not recorded for the present project.

The following procedure was used when administering the CERAD WLT (here translated from Norwegian to English):

CERAD WLT trial 1

Verbal instruction to participant: "I will now show you 10 words consecutively. Read every word out loud. Afterward I will ask you to recall the presented words"

Instruction to the test administrator: Show every word for 2 s, if the participant is not able to read the word, please read the word out loud for the participant.

CERAD WLT trial 2–3

Verbal instruction to participant: "I will now show you 10 words consecutively once more. Read every word out loud. Afterward I will ask you to recall the presented words"

Instruction to the test administrator: Same as for trial 1 instruction detailed above.

CERAD WLT Recall trial (presented 10 min after trial 3 administration)

Verbal instruction to participant: "A little while ago, I asked you to learn and remember a list of words, which you read out loud for me one at a time. Now, I want you to recall as many of those 10 words as you can remember".

CERAD WLT Recognition trial

Verbal instruction to participant: "Now I will read out loud all 10 words from the list in addition to some other words that were not on the list. I want you to reply "Yes" if you recognize a word from the list you read out loud for me, and "No", if it is a word that did not belong to the list you read out loud".

Statistical analysis

Multiple linear regression analyses with age, gender and years of education as predictors were fitted to model CERAD WLT performance in healthy controls ($n = 227$). Models were also fitted with interaction terms (age squared) to test for non-linear relationships between age and test performance (i.e. improving with younger age, and declining with older age). However, the inclusion of this interaction term did not add to the overall explained variance (adjusted R^2) of the regression models. Thus, only linear terms were included in our models. Overall estimates of the models (adjusted R^2, F value, p value), and relative contributions for individual predictors (β, partial R^2, p value) are reported. Since the DDI and Trønderbrain cohorts employ slightly different

Table 1. Cumulative percentiles for the CERAD WLT recognition subtest.

	Non-geriatric (≤64 years)			Geriatric (≥65 years)		
	Male $n = 33$	Female $n = 52$	Total $n = 85$	Male $n = 26$	Female $n = 34$	Total $n = 60$
Age range (mean)	45–64 (57.8)	40–64 (54.8)	40–64 (55.9)	65–80 (70.4)	65–78 (69.6)	65–80 (69.9)
Mean (SD)	19.2 (1.1)	19.8 (0.6)	19.6 (0.9)	19.0 (1.7)	19.2 (1.2)	19.1 (1.4)
Range	16.0–20.0	17.0–20.0	16.0–20.0	13.0–20.0	15.0–20.0	13.0–20.0
2%	16	17	17	13	15	13
5%	17	19	17	14	17	15
10%	17	19	19	16	18	17
25%	19	20	19	19	19	19
50%	20	20	20	20	20	20
75%	20	20	20	20	20	20
90%	20	20	20	20	20	20

Notes. n: Number of participants; SD: standard deviation.

criteria for inclusion and exclusion, this variable was assessed in separate regression models to assess a potential between-cohort bias. However, no significant bias of cohort was found.

Norming procedure

Due to a marked ceiling effect, the CERAD recognition subtest failed to produce a normal distribution of test scores, which is required for the regression-based norming procedure. Our data suggest that age and gender are the strongest demographic contributors to test performance. Thus, percentiles split by gender are provided for non-geriatric (e.g. 40–64 years, $n = 85$) and geriatric (65–80 years, $n = 60$) groups (Table 1).

As shown in Figure 1, the CERAD learning and recall raw test scores from the healthy control group were used to develop demographically adjusted regression-based norms. Methods and rationale used for regression-based norming in this paper are similar to procedures employed by Heaton et al. (2004), Testa, Winicki, Pearlson, Gordon, and Schretlen (2009) and Parmenter, Testa, Schretlen, Weinstock-Guttman, and Benedict (2010). We first normalized the control groups raw test scores by retrieving the cumulative frequency distribution of both measures. The resulting distribution was converted into a standard scaled score with a mean of 10, and a standard deviation of 3 (Table 2). We regressed the resulting scaled scores on age, gender and education. Plots of standardized residuals predicted values were assessed to ensure that the assumption of homoscedasticity was not violated, and normality of the residuals was checked visually with Q-Q plots. To derive normative information and calculate demographically adjusted T scores for each participant in the MCI group, we used the multiple regression equations derived from this analysis (Table 3) to compute the participants predicted scaled scores. The participants scaled score, derived from the healthy control group's normal distribution (Table 2) was subtracted from the regression equation predicted scaled score for each participant. The resulting discrepancy score was divided by the standard deviation of healthy control group's residuals (Table 3) to yield a standardized z score, which was then converted to a T score.

Lastly, multiple linear regression analyses with age, gender and years of education as predictors were fitted to the DDI MCI group's CERAD WLT learning and recall T

Table 2. Raw score to scaled score conversions.

Scaled score	CERAD learning	CERAD recall
3		≤1
4	≤13	2
5	14–15	3–4
6	16–17	
7	18	5
8	19	6
9	20	
10	21–22	7
11	23	8
12	24	
13	25	9
14	26	10
15	27	
16	28	
17	≥29	

Table 3. Normative regression models for CERAD WLT Learning and Recall subtests.

Variable	Predictor	B	Standard error B	T	SD residual
CERAD Word list Learning					
	Constant	14.195	1.600	8.871	2.54678
	Age	−0.100	0.020	−5.024	
	Gender	1.060	0.354	2.992	
	Education	0.113	0.053	2.125	
CERAD Word list Recall					
	Constant	13.756	1.571	8.759	2.51947
	Age	−0.093	0.020	−4.752	
	Gender	1.176	0.350	3.359	
	Education	0.107	0.052	2.045	

Notes. *B*: unstandardized regression coefficient; *T*: the t test statistic; SD: standard deviation.

score distributions to confirm adequate adjustment of demographic variables in an independent sample. All analyses were performed in the Statistical Package for Social Sciences (SPSS) version 25.

Norm calculator implementation

To facilitate the adoption and usage of the proposed norms in the clinic, we have developed a norm calculating tool that computes the regression equations. The functionality is simple and straightforward. To obtain both learning and recall *T* Scores, the user needs to enter valid demographic (age, gender and years of education) values and respective raw scores obtained from the tests. The tool is implemented as a self-contained HTML/Javascript webpage, available at (https://uit.no/ressurs/uit/cerad/cerad-calc.html), and released as open source at https://bitbucket.org/apgem/cerad-calc under Apache License, version 2.0.

Ethics

Both DDI and the Trønderbrain projects had been approved by the regional medical research ethics committees. Before taking part in the study, participants gave their written informed consent. All further study conduct was in line with the guidelines

provided by the Helsinki declaration of 1964 (revised 2013) and the Norwegian Health and Research Act.

Results

Demographic characteristics in healthy controls compared to the MCI group

The demographic characteristics of the healthy control group ($n = 227$) are summarized in Table 4 and compared to pertinent demographics of the independent MCI sample ($n = 168$) using summary independent t tests. While the MCI group obtained significantly lower raw scores in both CERAD WLT learning ($p < .0001$) and recall subtests ($p < .0001$), the groups were similar with regards to mean age and years of education.

Impact of demographics on the CERAD WLT performance within the healthy control group

Multiple regression analysis showed advancing age to predict decline in performance ($\beta = -.310$, partial $R^2 = .101$, $p < .0001$), whereas years of education ($\beta = .126$, partial $R^2 = .018$, $p < .05$) and female gender ($\beta = .201$, partial $R^2 = .046$, $p < .001$) was associated with increased performance on the CERAD word list learning subtest (adjusted $R^2 = .164$, F (3,223 $= 15.751$, $p < .0001$). Similarly, younger age ($\beta = -.279$, partial $R^2 = .083$, $p < .0001$), higher education ($\beta = .139$, partial $R^2 = .022$, $p < .05$) and female gender ($\beta = .208$, partial $R^2 = .048$, $p < .01$) predicted better performance on the CERAD word list recall subtest (adjusted $R^2 = .151$, F (3,224) $= 14.418$, $p < .0001$). No collinearity was observed between the predictor variables in any of the models (Variance Inflation Factor < 1.1).

Regression-based norms and scoring instruction

Table 3 shows the regression models based on the healthy controls to derive norms for the CERAD learning and recall subtests. All models include coefficients to adjust for age, gender (male $= 0$, female $= 1$) and years of education.

Table 4. Demographics, raw scores and T scores of the healthy controls and MCI group.

Variables	Healthy controls $n = 227$ Test scores/demographics	MCI group $n = 168$ Test scores/demographics	Sig.
Age	63.1 (8.6) [40–80]	64.4 (9.4) [40–80]	n.s*
Gender (% females)	62	54	
Years of education	14.2 (3.2) [7–23]	13.6 (3.4) [7–22]	n.s*
CERAD WORD LIST (30) LEARNING			
Raw score	21.5 (3.3)	16.9 (4.3)	**p < .0001**[*]
T score	50.0 (10.0)	38.7 (10.1)	**p < .0001**[*]
CERAD WORD LIST (10) RECALL			
Raw score	7.2 (2.0)	4.3 (2.7)	**p < .0001**[*]
T score	50.0 (10.0)	37.4 (11.4)	**p < .0001**[*]

Notes. n: Number of participants; MCI: mild cognitive impairment; Sig.: significance tests; p: p value; T: T score; n.s.: non-significant result. Results are presented as mean (standard deviation) [range] except for gender which is characterized by female percentage. *Summary independent T tests. Significant p values are shown in bold.

The raw test scores of the MCI group ($n = 168$) were converted to T scores using the following stepwise procedure: (1) Look up the scaled score for a given subtest in Table 2. (2) Use the regression coefficients found in Table 3 to obtain a predicted scaled score [constant + individual age(coefficient for age)+individual gender(coefficient for gender)+individual years of education(coefficient for education)]. (3) Then, subtract the actual scaled score from the predicted scaled score and divide it by the standard deviation of the residual (Table 3) to obtain a standardized z score which may be converted to a T score [$T = z(10)+50$]. For example, the T score calculation for a 50-year-old male with 8 years of education with a scaled score of 10 (Table 2) on CERAD learning: $14.195 + 50(-0.100)+8(0.113)+0(1.060)=10.099$. The difference between actual (10) and the predicted scaled score (10.099) is -0.099. Divided by the standard deviation of the healthy control groups residuals (2.54678) gives a z score of -0.039 which equates to a T score of 49.61.

Evaluation of demographic adjustment in the MCI group

Multiple regression models with age, gender and years of education as predictors were non-significant in the MCI group for both regression derived normative CERAD learning T scores (adjusted $R^2=.009$, F (3,165 = 1.531, $p=.208$) and CERAD recall T scores (adjusted $R^2=.005$, F (3,165 = 1.293, $p=.279$), indicating adequate adjustment of pertinent demographics when norms are applied to an independent sample.

Discussion

In this study, we have developed demographically adjusted norms for the CERAD WLT aimed at ages 40–80 years in a Norwegian sample. To our knowledge, this is the first normative study offering CERAD WLT norms aimed at this age interval, adjusted for the effects of both age, gender and education. As expected, increasing age had the largest impact on CERAD word list performance, followed by smaller effects of education and gender. These findings are in line with previous studies showing declining performance with increasing age (Sotaniemi et al., 2012; Welsh et al., 1994), a positive effect of educational attainment on performance (Beeri et al., 2006) and a female advantage on tests of verbal list learning tests or vocabulary (Beeri et al., 2006; Heaton et al., 2004; Liu et al., 2011). Thus, the regression-based norms were developed adjusted for these demographics. Healthy controls were recruited from two different prospective Norwegian cohorts. No between-cohort bias on test performance was found.

Regression-based norming procedures require stringent methodological criteria to be met (Testa et al., 2009). However, when criteria are met, this method has several advantages over the conventional discrete norming approach. Since we are using the entire normative sample, regression norming allows for the adjustment of several covariates in a linear fashion, meaning that the estimation of normative performance is possible at yearly increases in age and education for both males and females. Moreover, this is achieved with a lower sample size than required by discrete norms (Oosterhuis et al., 2016). However, when assumptions of linear regression are violated (i.e. normal distribution of errors, homoscedasticity and linearity), this method may

produce biased and unreliable estimates (Oosterhuis et al., 2016). In this study, efforts were made to ensure that assumptions of homoscedasticity and normal distributions of residuals were met. Furthermore, age is non-linearly related to many cognitive functions, including memory performance, with increasing capacity in early life superseded by a slow decline in later life (Hartshorne & Germine, 2015). We accounted for non-linearity by introducing an age squared term in our regression models. However, non-linearity was not demonstrated in our data, possibly because learning and memory capacity is fully developed, or showing normal age-related decline in this age cohort (Hartshorne & Germine, 2015).

While the normative data provided by Hankee et al. (2016) provide age adjusted normative data for younger ages (primarily for ages 35–55 years), comparisons with the proposed regression-based norms would not be appropriate due to the insufficient coverage of older ages (40–80 years). Similarly, the norms by Sotaniemi et al. (2012) originally employed in the DDI study are based on an older cohort with lower educational level compared to the participants enrolled in the DDI study (Kirsebom et al., 2017). In summary, this prompted the need to provide adequate norms covering ages for both earlier, and later stages of disease development and progression. We therefore opted to assess normative performance in an independent sample of MCI cases drawn from the DDI study covering both younger and older patients. We found that the regression-based norms successfully adjusted for age, gender and years of education in an independent sample of MCI cases. Furthermore, estimated T scores in the MCI group reflected an impaired normative performance with mean scores below 1 SD compared to the healthy controls. Owing to the successful adjustment of pertinent demographics, impaired learning and memory recall on the CERAD WLT should therefore be due to factors largely independent of normal aging, gender differences and educational level.

Interestingly, while years of education did predict higher performance on both CERAD WLT learning and recall subtests, the explained variance was relatively low (about 2%) compared to gender (about 5%). The relatively low variance explained by this variable may be due to a high mean educational level in both the healthy control group (14.2 years) and in the independent MCI group (13.6 years). While these mean levels seem fairly high, they are consistent with Norwegian population statistics (Statistics Norway, 2018), which indicate that 37.4% of Norwegians have completed upper secondary school (12–13 years) and 33.4% of the population has obtained a university degree (bachelor's degree or higher) with more than 15 years of education in total. As such, the relatively high educational level observed in our study could be a cultural bias, which could influence estimated normative performance on neuropsychological tests (Hayden et al., 2014; Heaton et al., 2004). These norms were developed in a Norwegian sample. However, they should be adequate for other Scandinavian countries which share similarities in culture, education and language. While all of our healthy controls were fluent in Norwegian, two participants (0.9%) were Swedish natives, who had lived in Norway most of their adult lives. Similarly, one Norwegian speaking Danish native (0.4%) was included in the MCI sample. The English CERAD WLT items were translated to Norwegian using back translation procedure to ensure accuracy.

A noteworthy finding using the predictions offered from the regression norms is that younger people between the ages of 40–50, and especially women, generally do

very well on this test, and the estimated normative performance for these individuals is therefore truncated and skewed. This indicates that the CERAD WLT may be too easy for these individuals. Thus, in order to detect longitudinal change in cognitive proficiency due to degenerative brain disease, we recommend the addition of a more challenging wordlist test such as the Rey Auditory Verbal Learning Test (RAVLT) (Schmidt, 1996) for younger individuals.

A limitation of this study is the missing scores on the CERAD WLT recognition memory test. In addition, this subtest shows a marked ceiling effect, and does not produce a normal distribution of test scores required for regression-based norming. Our data indicate that age and gender had the highest influence on normative performance. Thus, normative performance on this test is shown by providing cumulative percentile ranks for geriatric ($\geq$65) and non-geriatric ($\leq$64) age groups, further split by gender. Secondly, we did not have a complete longitudinal record of our healthy controls to verify that they remained cognitively healthy within a reasonable timeframe. Thirdly, while the regression equations will mathematically estimate age, and educational effects beyond the age and education range in this study, estimates are not reliable beyond these ranges. Finally, regression-based norms may not be as easy and familiar to use for clinicians compared to conventional discrete norms. In order to overcome this, we offer a free web-based intuitive normative calculator (supplementary file 1/https://uit.no/ressurs/uit/cerad/cerad-calc.html).

Conclusion

We propose demographically adjusted regression-based norms for the CERAD WLT, based on healthy controls from the Norwegian DDI and Trønderbrain cohorts. The norms are linearly adjusted for the effects of gender, age and education between the ages of 40 and 80 years, with an educational attainment between 7 and 23 years.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author, B-E.K. The data are not publicly available due to restrictions e.g. their containing information that could compromise the privacy of research participants.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., … 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. doi: 10.1016/j.jalz.2011.03.008

Bateman, R. J., Xiong, C., Benzinger, T. L. S., Fagan, A. M., Goate, A., Fox, N. C., … Morris, J. C. (2012). Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *New England Journal of Medicine*, 367(9), 795–804. doi:10.1056/NEJMoa1202753

Beeri, M. S., Schmeidler, J., Sano, M., Wang, J., Lally, R., Grossman, H., & Silverman, J. M. (2006). Age, gender, and education norms on the CERAD neuropsychological battery in the oldest old. *Neurology*, 67(6), 1006–1010. doi:10.1212/01.wnl.0000237548.15734.cd

Berge, G., Sando, S. B., Albrektsen, G., Lauridsen, C., Møller, I., Grøntvedt, G. R., … White, L. R. (2016). Alpha-synuclein measured in cerebrospinal fluid from patients with Alzheimer's disease, mild cognitive impairment, or healthy controls: A two year follow-up study. *BMC Neurology*, 16(1), 180p.

Fillenbaum, G. G., McCurry, S. M., Kuchibhatla, M., Masaki, K. H., Borenstein, A. R., Foley, D. J., … White, L. (2005). Performance on the CERAD neuropsychology battery of two samples of Japanese-American elders: Norms for persons with and without dementia. *Journal of the International Neuropsychological Society*, 11(2), 192–201.

Fillenbaum, G. G., van Belle, G., Morris, J. C., Mohs, R. C., Mirra, S. S., Davis, P. C., … Heyman, A. (2008). Consortium to Establish a Registry for Alzheimer's Disease (CERAD): The first twenty years. Alzheimers Dement, 4(2), 96–109.

Fladby, T., Pålhaugen, L., Selnes, P., Waterloo, K., Bråthen, G., Hessen, E., … Aarsland, D. (2017). Detecting at-risk Alzheimer's disease cases. *Journal of Alzheimer's Disease*, 60(1), 97–105.

Hankee, L. D., Preis, S. R., Piers, R. J., Beiser, A. S., Devine, S. A., Liu, Y., … Au, R. (2016). Population normative data for the CERAD Word list and Victoria Stroop test in younger- and middle-aged adults: Cross-sectional analyses from the Framingham heart study. *Experimental Aging Research*, 42(4), 315–328. doi:10.1080/0361073X.2016.1191838

Hartshorne, J. K., & Germine, L. T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychological Science*, 26(4), 433–443. doi:10.1177/0956797614567339

Hayden, K. M., Makeeva, O. A., Newby, L. K., Plassman, B. L., Markova, V. V., Dunham, A., … of the behalf of the T.-D. S. G. (2014). A comparison of neuropsychological performance between US and Russia: Preparing for a global clinical trial. *Alzheimer's & Dementia*, 10(6), 760–768. doi: 10.1016/j.jalz.2014.02.008

Heaton, R. K., Miller, S. W., Taylor, M. J., & Grant, I. (2004). *Revised comprehensive norms for an expanded Halstead-Reitan battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults*. Lutz, FL: Psychological Assessment Resources.

Hughes, C. P., Berg, L., Danziger, W., Coben, L. A., & Martin, R. L. (1982). A new clinical scale for the staging of dementia. *British Journal of Psychiatry*, 140(6), 566–572. doi:10.1192/bjp.140.6.566

Jack, C. R., Jr., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., … Sperling, R. (2018). NIA-AA research framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia*, 14(4), 535–562. doi:10.1016/j.jalz.2018.02.018

Kirsebom, B.-E., Espenes, R., Waterloo, K., Hessen, E., Johnsen, S. H., Bråthen, G., … Fladby, T. (2017). Screening for Alzheimer's disease: Cognitive Impairment in self-referred and memory

clinic-referred patients. *Journal of Alzheimer's Disease*, *60*(4), 1621–1631. doi:10.3233/JAD-170385

Liu, K. P. Y., Kuo, M. C. C., Tang, K-C., Chau, A. W. S., Ho, I. H. T., Kwok, M. P. H., … Chu, L-W. (2011). Effects of age, education and gender in the Consortium to Establish a Registry for the Alzheimer's Disease (CERAD)-neuropsychological assessment battery for cantonese-speaking Chinese elders. *International Psychogeriatrics*, *23*(10), 1575–1581. doi:10.1017/S1041610211001153

Morris, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., van Belle, G., Fillenbaum, G., … Clark, C. (1989). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*, *39*(9), 1159–1165. doi:10.1212/WNL.39.9.1159

Oosterhuis, H. E., van der Ark, L. A., & Sijtsma, K. (2016). Sample size requirements for traditional and regression-based norms. *Assessment*, *23*(2), 191–202.

Parmenter, B. A., Testa, S. M., Schretlen, D. J., Weinstock-Guttman, B., & Benedict, R. H. B. (2010). The utility of regression-based norms in interpreting the minimal assessment of cognitive function in multiple sclerosis (MACFIMS). *Journal of the International Neuropsychological Society*, *16(1)*, 6–16. doi:10.1017/S1355617709990750

Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, *256*(3), 183–194.

Schmidt, M. (1996). Rey *auditory verbal learning test:* A handbook. Los Angeles, CA: Western Psychological Services.

Schmidtke, K., & Hermeneit, S. (2008). High rate of conversion to Alzheimer's disease in a cohort of amnestic MCI patients. *International Psychogeriatrics*, *20*(1), 96–108.

Soldan, A., Pettigrew, C., Li, S., Wang, M.-C., Moghekar, A., Selnes, O. A., … O'Brien, R. (2013). Relationship of cognitive reserve and cerebrospinal fluid biomarkers to the emergence of clinical symptoms in preclinical Alzheimer's disease. *Neurobiology of Aging*, *34*(12), 2827–2834. doi:10.1016/j.neurobiolaging.2013.06.017

Sotaniemi, M., Pulliainen, V., Hokkanen, L., Pirttilä, T., Hallikainen, I., Soininen, H., & Hänninen, T. (2012). CERAD-neuropsychological battery in screening mild Alzheimer's disease. *Acta Neurologica Scandinavica*, *125*(1), 16–23. doi:10.1111/j.1600-0404.2010.01459.x

Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., … Phelps, C. H. (2011). Toward defining the preclinical stages of 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), 280–292.

Statistics Norway. (2018, October 20). Befolkningens utdanningsnivå (Population educational level). Retrieved from https://www.ssb.no/utdanning/statistikker/utniv.

Testa, S. M., Winicki, J. M., Pearlson, G. D., Gordon, B., & Schretlen, D. J. (2009). Accounting for estimated IQ in neuropsychological test performance with regression-based techniques. *Journal of the International Neuropsychological Society*, *15*(06), 1012–1022. doi:10.1017/S1355617709990713

Warrington, E. K., & James, M. (1991). *The Visual Object and Space Perception Battery*. Bury St Edmunds, UK: Thames Valley Test Company.

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Cedarbaum, J., … Trojanowski, J. Q. (2015). 2014 Update of the Alzheimer's disease neuroimaging initiative: A review of papers published since its inception. *Alzheimer's & Dementia*, *11*(6), e1–e120. doi:10.1016/j.jalz.2014.11.001

Welsh, K. A., Butters, N., Mohs, R. C., Beekly, D., Edland, S., Fillenbaum, G., & Heyman, A. (1994). The Consortium to Establish a Registry for Alzheimer's disease (CERAD). Part V. A normative study of the neuropsychological battery. *Neurology*, *44*(4), 609–614. doi:10.1212/WNL.44.4.609

Zachary, R. A., & Gorsuch, R. L. (1985). Continuous norming: Implications for the WAIS-R. *Journal of Clinical Psychology*, *41*(1), 86–94. doi:10.1002/1097-4679(198501)41:1<86::AID-JCLP2270410115>3.0.CO;2-W