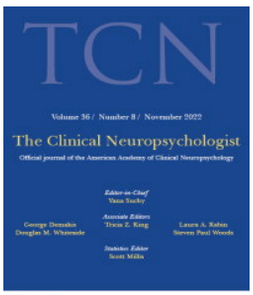

The Clinical Neuropsychologist

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ntcn20

Regression-based normative data for the Rey Auditory Verbal Learning Test in Norwegian and Swedish adults aged 49–79 and comparison with published norms

Jacob Espenes, Ingvild Vøllo Eliassen, Fredrik Öhman, Erik Hessen, Knut Waterloo, Marie Eckerström, Ingrid Myrvoll Lorentzen, Cecilie Bergland, Madelene Halvari Niska, Santiago Timón-Reina, Anders Wallin, Tormod Fladby & Bjørn-Eivind Kirsebom

To cite this article: Jacob Espenes, Ingvild Vøllo Eliassen, Fredrik Öhman, Erik Hessen, Knut Waterloo, Marie Eckerström, Ingrid Myrvoll Lorentzen, Cecilie Bergland, Madelene Halvari Niska, Santiago Timón-Reina, Anders Wallin, Tormod Fladby & Bjørn-Eivind Kirsebom (2022): Regression-based normative data for the Rey Auditory Verbal Learning Test in Norwegian and Swedish adults aged 49–79 and comparison with published norms, The Clinical Neuropsychologist, DOI: 10.1080/13854046.2022.2106890

To link to this article: https://doi.org/10.1080/13854046.2022.2106890

Published online: 13 Aug 2022.

THE CLINICAL NEUROPSYCHOLOGIST
https://doi.org/10.1080/13854046.2022.2106890

Regression-based normative data for the Rey Auditory Verbal Learning Test in Norwegian and Swedish adults aged 49–79 and comparison with published norms

Jacob Espenes[a,b], Ingvild Vøllo Eliassen[c,d], Fredrik Öhman[e,f], Erik Hessen[c,d], Knut Waterloo[a,b], Marie Eckerström[e], Ingrid Myrvoll Lorentzen[a], Cecilie Bergland[a], Madelene Halvari Niska[a], Santiago Timón-Reina[c,g], Anders Wallin[e], Tormod Fladby[c,h] and Bjørn-Eivind Kirsebom[a,b]

[a]Department of Psychology, Faculty of Health Sciences, The Arctic University of Norway, Tromsø, Norway; [b]Department of Neurology, University Hospital of North Norway, Tromsø, Norway; [c]Department of Neurology, Akershus University Hospital, Lørenskog, Norway; [d]Department of Psychology, University of Oslo, Oslo, Norway; [e]Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; [f]Wallenberg Centre for Molecular and Translational Medicine, University of Gothenburg, Gothenburg, Sweden; [g]Departamento de Inteligencia Artificial, Universidad Nacional de Educación a Distancia, Madrid, Spain; [h]Institute of Clinical Medicine, University of Oslo, Oslo, Norway

ABSTRACT

Objective: The Rey Auditory Verbal Learning Test (RAVLT) is a widely used measure of episodic verbal memory. To our knowledge, culturally adapted and demographically adjusted norms for the RAVLT are currently not available for Norwegian and Swedish adults, and imported North American norms are often used. We here develop regression-based norms for Norwegian and Swedish adults and compare our norms to North American norms in an independent sample of cognitively healthy adults. **Method:** Participants were 244 healthy adults from Norway and Sweden between the aged 49 and 79 years, with between 6 and 24 years of education. Using a multiple multivariate regression-based norming procedure, we estimated effects of age, sex, and years of education on basic and derived RAVLT test scores. The newly developed norms were assessed in an independent comparison group of cognitively healthy adults ($n = 145$) and compared to recently published North American regression-based norms. **Results:** Lower age, female sex and more years of education predicted higher performance on the RAVLT. The new norms adequately adjusted for age, education, and sex in the independent comparison group. The American norms corrected for demographics on all RAVLT trials except trials 4, 7, list B, and trials 1–5 total. Test-retest ($M = 2.55$ years) reliability varied from poor to good. **Conclusion:** We propose regression-based norms for the RAVLT adjusting for pertinent demographics. The norms may be used for assessment of Norwegian and Swedish adults between the aged of 49 and 79 years, with between 6 and 24 years of education.

ARTICLE HISTORY

Received 25 March 2022
Accepted 23 July 2022
Published online 16 August 2022

KEYWORDS

Normative; Rey Auditory Verbal Learning Test; Norway; Sweden; memory

CONTACT Jacob Espenes Johan.j.espenes@uit.no Department of Psychology, Faculty of Health Sciences, The Arctic University of Norway, Hansine Hansens veg 18, Tromsø, 9037, Norway.

Introduction

The Rey Auditory Verbal Learning Test (RAVLT) is a widely used measure of episodic verbal memory in the field of neuropsychology (Boake, 2000). It is a multi-trial, 15-item word list test that enables assessment of fundamental memory processes, including acquisition, interference effects, retention, and retrieval (Ivnik et al., 1992). The RAVLT is sensitive to learning and memory deficits in several clinical groups, including patients with mild cognitive impairment (MCI; Estévez-González et al., 2003), Alzheimer's Disease (AD; Ricci et al., 2012), left hemispheric brain pathology (Loring et al., 2008), and neuropathologies of various etiologies (Powell et al., 1991). RAVLT scores are good markers for progressive episodic memory deficits typical in common age-related conditions such as AD and MCI due to AD (Belleville et al., 2017). Delayed recall performance on the RAVLT has demonstrated adequate to excellent diagnostic accuracy for identifying which individuals with MCI will progress to AD dementia (Eckerström et al., 2013; Ewers et al., 2012).

Sociodemographic factors have been found to influence RAVLT performance. Age effects are consistently reported in middle-aged and older adults, showing declining performance with increasing age (Lavoie et al., 2018; Messinis et al., 2016; Stricker et al., 2021). Findings are somewhat less consistent regarding the influence of sex and educational attainment. Several studies indicate a clear female advantage on RAVLT performance (Asperholm et al., 2019; Lavoie et al., 2018; Stricker et al., 2021; Sundermann et al., 2016; 2017; Van Der Elst et al., 2005) while others find no significant influence of sex (Marqués et al., 2013; Messinis et al., 2016). In contrast, an older meta-analytic review of demographic influences on RAVLT performance suggests a male advantage on some trials, and otherwise no effects of sex on performance (Mitrushina et al., 2005). Individuals with more years of education often obtain higher scores on the RAVLT (Bezdicek et al., 2014; Lavoie et al., 2018; Messinis et al., 2016; Stricker et al., 2021; Van Der Elst et al., 2005). However, meta-analytic evidence has indicated no significant effect of education on performance (Mitrushina et al., 2005).

Linguistic and cultural differences may also contribute to systematic variation of RAVLT performance in different populations. Norms from different cultural groups are not necessarily interchangeable. Which norm set we choose to apply on an individual's scores may influence their likelihood of being classified as memory-impaired (Strauss et al., 2006). Local norms for the RAVLT have been developed for older adults with diverse cultural and linguistic backgrounds including north American (Stricker et al., 2021), Venezuelan (Correia & Osorio, 2014), French-Canadian (Lavoie et al., 2018), Greek (Messinis et al., 2016), Israeli (Vakil et al., 2010), and German (Boenniger et al., 2021).

To our knowledge, there are currently no demographically adjusted test norms for the RAVLT available for the Norwegian or Swedish middle-aged and older adults. Demographically adjusted and locally sourced normative material is needed to increase the likelihood for an accurate evaluation of memory function in this population. Thus, the first objective of this study was to develop normative data for the RAVLT for Norwegian and Swedish adults ages 49 to 79 years applying a multiple multivariate approach (Van der Elst et al., 2017). Secondly, clinicians in Norway and Sweden have several sets of norms available for use such as the newly developed

population-based regression-based norms from the Mayo Normative Study (Stricker et al., 2021) that may or may not be appropriate in a Scandinavian population. Thus, the study's second objective was to compare the currently proposed norms with published norms from Stricker et al. (2021) in an independent sample of cognitively healthy participants with subjective cognitive decline (SCD) from Norway and Sweden.

Methods and materials

Participants

The present study included 244 healthy control participants from three related research projects on early phases of dementia diseases conducted in Norway and Sweden; the Dementia Initiation study (DDI, $n = 70$); the Gothenburg Mild Cognitive Impairment (MCI) study ($n = 121$); and the Oslo MCI study ($n = 53$). Healthy controls included from DDI were assessed at the Akershus University Hospital or the University Hospital of Northern Norway between January 2013 and June 2020. The Oslo MCI study is the predecessor of the ongoing DDI study, and assessments were performed at the Akershus University hospital between 2005 and 2013. Participants included from the Gothenburg MCI study were assessed at the Sahlgrenska University Hospital, Sweden, between January 2001 and March 2014. Healthy controls from DDI- ($n = 70$) and the Oslo MCI- study ($n = 70$) were primarily recruited from spouses of symptom group participants and secondarily through advertisements in local media and from the orthopedic wards. Healthy controls included from the Gothenburg MCI study ($n = 121$) were primarily recruited through senior citizen organizations, and a small proportion were relatives of symptom group participants. All studies followed a similar standardized procedure for assessment that included neurological and psychical examination, neuropsychological assessment and self and informant-reported medical history. Most participants agreed to submit blood samples and cerebrospinal fluid samples. However, these were not analyzed for the purpose of this study. For a complete description of the Gothenburg MCI cohort, methods, and study procedures, see Wallin et al. (2016). For DDI see Fladby et al. (2017) and for Oslo MCI refer to Hessen et al. (2014).

Joint criteria for inclusion applied to all healthy controls employed in the normative analyses of the present study ($n = 244$) was aged 49 through 79, the absence of subjective symptoms of cognitive decline, mini mental state examination (MMSE) ≥ 26, and a native language of Norwegian or Swedish. The normative sample was split between 122 participants speaking Norwegian and 122 speaking Swedish. Two participants spoke Norwegian as the second language. Fifty participants were between aged 49 and 58; 122 were between 59 and 68 years; and 72 were between 69 and 79 years. Education ranged between 6 and 24 years of education. Every full year of formal education attained by the participants was counted, excluding degrees of the same level. Exclusion criteria were developmental disorders, neurological disease, intellectual disability, severe somatic disorders that might negatively influence cognitive performance, history of stroke, or severe psychiatric disorder, including major depression. Apart from MMSE, results on cognitive screening tests and neuropsychological measures were not used to verify cognitive normalcy or exclude participants as this potentially excludes normal healthy participants, thereby reducing variation

associated with normal aging, thus limiting the generalizability and validity of the norms. Scores from participants who did not complete the RAVLT or had missing scores on any RAVLT trial was excluded from the analysis. Thus, only participants with complete RAVLT administrations were included.

Between cohort comparisons of demographics and cognitive performance

Participants were recruited from three related research projects, and potential cohort effects were investigated. While the Gothenburg and -Oslo MCI cohort participants on average had fewer years of education, there were no cohort effects on RAVLT raw scores adjusted for age differences, years of education, and sex, except for trial 1. Scores on trial 1 were analyzed in a regression model, which included the predictor's cohort (dummy coded to account for three cohorts), age, years of education, and sex. Results showed that control participants recruited from the Oslo MCI study on average remembered 0.738 fewer words than the controls from DDI, adjusting for differences in education, age and sex ($b = -0.738$, 95% CI [−1.327, −.150], $p = .014$, $F(5, 238) = 10.246$, and $p = <.001$). There were no significant differences between the Oslo MCI cohort and participants from the Gothenburg MCI study.

Independent comparison group to assess norms

The DDI study and Gothenburg MCI study also include participants with subjective cognitive decline (SCD), and at the time of analysis, 145 cognitively healthy participants with SCD had available assessments on the RAVLT. These were included in a separate sample to evaluate and compare the current proposed norms with published norms from Stricker et al. (2021). All SCD participants underwent the same standardized procedure for assessment as previously described for healthy controls, including the general exclusion and inclusion criteria, and MMSE score ≥ 26. SCD participants were included via referrals from general practitioners to memory clinics, and self-referral following public advertisements aimed at individuals with memory complaints. As such, memory deficits were the main cognitive complaints. SCD was determined by self-report the following proposed guidelines in Jessen et al. (2014) and Molinuevo et al. (2017). All participants with SCD were subject to a clinical interview about the nature of progression since onset, experience of cognitive deficits in other domains, familiar history, and affective symptoms. To ensure cognitive normalcy and differentiate participants presenting SCD from MCI, recommendations from Albert et al. (2011) were applied and participants were excluded if they presented *objective* cognitive decline, operationalized as a score 1.5 standard deviation below the normative mean on at least one of the following neuropsychological tests (applied normative corrections in parenthesis); The Trail Making Test B (Espenes et al., 2020; Reitan & Wolfson, 1985), Controlled Oral Word Association test (COWAT, Heaton et al., 2004; Lorentzen et al., 2021), Silhouettes from Visual Object and Space Perception Battery (VOSP, Eliassen et al., 2020; Warrington & James, 1991). Participants with SCD from the DDI cohort were excluded on basis of the CERAD word list-delayed recall (Fillenbaum et al., 2008; Kirsebom et al., 2019). Participants from the Gothenburg MCI cohort did

not perform the CERAD word list delayed recall and were instead excluded based on the RAVLT trial 7 (Rey, 1958; Stricker et al., 2021).

To investigate if the SCD group would be suitable as an independent group for comparing normative adjustments, their RAVLT scores were compared to those of the controls (Table 1). Regression analyses indicated no significant differences between groups adjusting for differences in years of education, age, and sex, except for trial 4, where a minor difference was observed. The SCD group on average remembered 0.742 more words compared to the healthy controls ($b = 0.742$, $p = .032$, and 95% CI [0.063, 1.420]). The confidence interval suggests that this difference could be very small, possibly spurious, as there is no theoretical basis for trial 4 differing substantially from other parts of the RAVLT. We therefore conclude that the comparison group comprised of individuals presenting SCD had comparable scores to the healthy control group on the RAVLT, indicating that they were suitable as an independent comparison group.

RAVLT test version and administration

RAVLT assessments were performed by clinical psychologists or psychologists-in-training. Firstly, the participant is instructed to try to remember as many words as possible from a list of words that is about to be read aloud. Then, a list of 15 words (list A), is read aloud to the participant, to which the participant is required to recall as many words as possible directly after. This is repeated for a total of five trials, and the participant is required to freely recall as many words as possible after each

Table 1. Demographics, raw scores of the normative sample of healthy controls and the independent comparison group comprised of cognitively healthy participants with subjective cognitive decline (SCD).

Variables	Normative sample of healthy controls ($n = 244$)	Independent comparison group ($n = 145$)	t^a/χ^2	p
Age Mean (SD) [range]	64.3 (6.8) [49 – 79]	62.3 (6.7) [49 – 77]	2.952	.003
Female *n* (%)	138 (56.6 %)	91 (62.8 %)	1.444	*n.s.*
Years of education Mean (SD) [range]	12.7 (3.3) [6 – 24]	14.0 (3.2) [6 – 21]	–3.666	<.001
Trial 1 Mean (*SD*)	5.5 (1.7)	5.8 (1.9)	0.260	*n.s.*
Trial 2 Mean (*SD*)	8.3 (2.1)	8.7 (2.2)	–0.094	*n.s.*
Trial 3 Mean (*SD*)	9.7 (2.4)	10.5 (2.5)	1.306	*n.s.*
Trial 4 Mean (*SD*)	10.8 (2.4)	12.0 (4.5)	2.150	.032
Trial 5 Mean (*SD*)	11.4 (2.5)	12.0 (2.1)	0.397	*n.s.*
Trial 6 (immediate memory) Mean (*SD*)	9.3 (3.1)	10.3 (2.6)	1.263	*n.s.*
Trial 7 (delayed memory) Mean (*SD*)	9.0 (3.1)	10.1 (2.5)	1.522	*n.s.*
Trials A1–A5 total Mean (*SD*)	45.6 (9.6)	48.4 (9.2)	0.687	*n.s.*
List B Mean (*SD*)	5.4 (1.8)	5.6 (2.0)	–0.425	*n.s.*

Notes. n = Number of participants; $p = p$-value; $t = t$ statistic; *n.s.* = non-significant result ($p > .05$); Results are presented as mean (Standard deviation) [range] except for sex which is characterized by female percentage.

[a]for RAVLT scores, test statistics refer to mean difference between groups controlling for age, years of education, and sex. For age and years of education, independent samples *t*-tests with Welch correction were conducted; χ^2 = Chi Square test for 2 × 2 table.

presentation. After five consecutive trials, a distractor list (B) containing 15 separate words is presented, and the participant is asked to freely recall as many words as possible from this new list. Following immediately, without cues or renewed presentation, the participant is asked to recall list A again (trial 6). After a timed delay of 30 minutes, during which other neuropsychological tests with non-verbal stimuli were conducted, the participant is required to freely recall List A once more (trial 7), reflecting delayed verbal memory. On the RAVLT, the primary variables are correctly recalled words on learning trials (trial 1 to 5), list B, trial 6, and trial 7. In addition, derived measures (Table 2) are often computed to provide evaluations of learning (Ivnik et al., 1990), inhibition and interference effects, and retention (i.e., correctly recalled words after 30-minute delay relative to the number of words previously recalled).

Norwegian participants were administered a Norwegian translation of the RAVLT word list, available in English in Lezak et al. (2012). Likewise, the Swedish participants from the Gothenburg MCI cohort used a translated Swedish version. English, Norwegian, and Swedish versions of the RAVLT with standardized instructions and word lists A and B are presented in appendix A. Norwegian and Swedish versions of the RAVLT were not backtranslated or otherwise formally validated. The Swedish Gothenburg MCI study employed a different protocol for administering the recognition trial than the Norwegian cohorts from DDI and Gothenburg-Oslo MCI and we therefore do not present normative data for this part of the test.

Regression norming procedure

Following procedures described in Van der Elst et al. (2017), multivariate regression-based norms were developed based on the performance of the included healthy controls ($n = 244$) on all primary RAVLT measures. Exploratory analyses confirmed that all primary RAVLT measures were moderately to highly correlated (r = .289–.868), suggesting that

Table 2. Primary and the derived measures on the RAVLT.

RAVLT measures	Description
Primary measures	
Trial 1	Number of correctly recalled words from list A after first learning trial
Trial 2	Second learning trial
Trial 3	Third learning trial
Trial 4	Fourth learning trial
Trial 5	Fifth learning trial
List B	Free recall of list B
Trial 6	Recall of list A without renewed presentation
Trial 7	Thirty-minute delayed recall of list A
Derived measures	
Trials 1–5 total learning	∑ (Trial 1, Trial 2, Trial 3, Trial 4, Trial 5)
Learning over trials	(Trials 1–5 total—(Trial 1*5))
Learning rate[a]	(Trial 5—Trial 1)
Proactive inhibition[b]	(Trial 1—list B)
Retroactive inhibition[b]	(Trial 5—Trial 6)
Long-term percentage retention	(100 * (Trial 7/Trial 5))

Note: ∑ = sum; primary measures are reported in order of administration;
[a]Positive score on learning rate indicate that more words were repeated at Trial 5 than Trial 1.
[b]Positive score indicate inhibition effect, that is, more words were recalled in Trial 1 compared to List B, or more words recalled in Trial 5 compared to Trial 6.

primary RAVLT measures was suitable for multivariate analysis. Correlations between primary RAVLT measures and demographical variables are presented in Appendix A.2.1.

A preliminary multivariate regression model with predictors age, age^2, education, education^2, age*education interaction, sex, sex*age interaction, trial, trial*age interaction, trial*education interaction, trial*sex interaction, and a dummy coded variable accounting for cohort-effects was fitted. Age and education were mean-centered to avoid bias due to multicollinearity and improve interpretation of coefficients. Trial was dummy coded with 7 dummies and trial 1 as the reference category. The preliminary model was subsequently simplified and reduced by hierarchically dropping one covariate at a time in a stepwise manner and comparing log-likelihood ratios of models. The model selection process and associated test statistics are presented in Appendix A.2. Maximum likelihood estimation was used because this allows for classical likelihood ratio testing of nested models (i.e., directly comparing simpler models with complex models). If the simplified model with one reduced covariate did not significantly reduce log-likelihood, then the simplified model was preferred and subsequently used as reference model for further simplification. A nominal alpha-level criterion of $\alpha = .01$ was used. Once the mean structure of the model could not be simplified further without deterioration, the correlation structure of the model was attempted simplified using a homogenous/heterogeneous compound symmetry (CS) and a first-order autoregressive covariance structure (AR (1)). Results indicated that the default unstructured covariance matrix provided the best fit to the data. Once adequate mean structure and covariance structures were obtained, estimates were re-calculated using restricted maximum likelihood (REML), which may reduce small sample bias (Van der Elst et al., 2017; Verbeke & Molenberghs, 2009).

For the derived RAVLT variables, we fitted conventional univariate multiple regression models that were assessed for linear, nonlinear and interaction effects of age, education, and sex. These predictors were included if they significantly an improved model fit ($p < .05$). Histograms and QQ-plots of standardized residuals indicated slight deviations from normality for the measure long-term percentage retention (LTPR), and some caution is advised when interpreting extreme scores (e.g., $T < 30$) for this measure. Normative models for the secondary variables are provided in Table 4. We assessed all normative measures for influential cases and outliers that might disturb or unduly influence normative measures. Cases deemed highly influential and abnormal were excluded from analysis to ensure the validity of normative estimates. The variables proactive inhibition and retroactive inhibition were non-normally distributed and had no significant association with demographic variables. We therefore calculated the inverse cumulative distribution based on the performance of the entire normative sample ($n = 244$) for these measures. Raw scores and corresponding percentiles are provided in Table 5. All analyses were conducted using the Statistical Package for Social Sciences (SPSS) version 28, JASP version 0.16.1 (JASP Team, 2022), and R version 3.6.2 (R Core Team, 2020).

Calculating normative performance using regression-based norms

Three steps are required for calculating the normative performance: (1) estimating the predicted performance using regression coefficients, (2) subtracting the actual observed score from the predicted score, (3) standardization. Firstly, because age and

years of education was mean centered for all analyses, they must be calculated relative to the age ($M = 64.3$), and years of education ($M = 12.7$), in the normative sample (Table 1). Every full year of formal education is counted, excluding degrees of the same level. For instance, a participant could reach 24 years of education by 13 years basic schooling, a professional degree of 6 years and a Ph.D. position intended for 5 years. Then, predicted performance is calculated applying the coefficients in Tables 3 and 4. Regression coefficients from the multivariate regression model are applied using the following formula: [Intercept + (individual sex*sex coefficient) + (age centered*age coefficient) + (years of education centered*education coefficient) + (coefficient for Trial *n*) + (years of education centered * coefficient for education for Trial *n*) + (individual sex * sex coefficient for Trial *n*)]. This produces a predicted score based on individual demographics. The predicted score is then subtracted from the individual obtained score. Lastly, the normative score is standardized to the Z-scores following: [Obtained score − predicted scaled score/standard deviation of the residuals obtained from the regression = Z-score]. As customary, Z-scores were further converted to *T*-scores with a mean of 50 and standard deviation of 10 by [$T = Z * 10 + 50$].

As an example, suppose that a 70-year-old female with 15 years of education remembered 10 words on trial 2. Age centered equals 5.7 [= 70 − 64.3] and years of education centered is 2.3 [15 − 12.7]. Thus, the predicted score equals 9.1 [= (5.053 + (1 * 0.761) + (5.7 * −0.041) + (2.3 * 0.128) + 2.457 + (1 * 0.540) + (2.3 * 0.097))]. The standardized residual for Trial 2 is 1.813. So, the *T*-score is 55 [= ((((10 − 9.1)/1.813) *10)) + 50].

Table 3. Coefficients from multivariate regression for normative adjustments on the primary variables from the RAVLT based on 244 healthy adult participants.

Parameter	*b*	b 95 % CI [LL, UL]	*s.e.*	*t*	*p*	*SD residual*
Intercept	5.053	[4.750, 5.356]	0.155	32.643	<.001	1.589
Age	−0.041	[−0.065, −0.017]	0.012	−3.318	.001	
Education	0.128	[0.066, 0.189]	0.031	4.076	<.001	
Sex	0.761	[0.357, 1.166]	0.206	3.692	<.001	
Trial 2	2.457	[2.140, 2.773]	0.161	15.218	<.001	1.813
Trial 3	3.826	[3.458, 4.193]	0.187	20.422	<.001	2.068
Trial 4	5.099	[4.694, 5.504]	0.207	24.655	<.001	2.167
Trial 5	5.591	[5.183, 5.999]	0.208	26.876	<.001	2.119
Trial 6	3.282	[2.765, 3.798]	0.264	12.454	<.001	2.677
Trial 7	3.151	[2.643, 3.658]	0.259	12.171	<.001	2.695
List B	0.181	[−0.174, 0.536]	0.181	1.001	.317	1.673
Edu*Trial 2	0.097	[0.033, 0.160]	0.032	2.984	.003	
Edu*Trial 3	0.122	[0.049, 0.196]	0.038	3.253	.001	
Edu*Trial 4	0.120	[0.039, 0.202]	0.042	2.898	.004	
Edu*Trial 5	0.158	[0.076, 0.240]	0.042	3.777	<.001	
Edu*Trial 6	0.212	[0.109, 0.316]	0.053	4.010	<.001	
Edu*Trial 7	0.230	[0.128, 0.332]	0.052	4.425	<.001	
Edu*List B	0.053	[−0.018, 0.124]	0.036	1.458	.145	
Sex*Trial 2	0.540	[0.120, 0.961]	0.215	2.518	.012	
Sex*Trial 3	0.620	[0.132, 1.108]	0.249	2.489	.013	
Sex*Trial 4	0.318	[−0.221, 0.857]	0.275	1.155	.248	
Sex*Trial 5	0.571	[0.029, 1.113]	0.277	2.063	.039	
Sex*Trial 6	0.900	[0.213, 1.587]	0.350	2.569	.010	
Sex*Trial 7	0.712	[0.037, 1.387]	0.344	2.069	.039	
Sex*List B	−0.415	[−0.887, 0.057]	0.241	−1.722	.085	

Notes: Intercept represents reference category Trial 1; *b* = unstandardized beta coefficient; *s.e.* = standard error of the unstandardized beta coefficient; *SD* residual = standard deviation of the residuals; Sex was coded (0 = male, 1 = female); Age and Education were mean centered, thus Age = (calendar age—64.3); Education/Edu = (the number of years of education obtained—12.7).

Table 4. Coefficients from multiple regressions for derived RAVLT measures based on 244 healthy adult participants.

Parameter	*b*	b 95 % CI [LL, UL]	*s.e.*	*t*	*p*	*Partial* R^2	*Adj.* R^2	*SD residual*
Trials 1–5 total intercept	42.269	[40.752, 43.839]	0.783	53.988	<.001		.300	7.982
Trials 1–5 total age	−0.269	[−0.423, −0.116]	0.078	−3.458	<.001	.048		
Trials 1–5 total education	1.095	[0.782, 1.409]	0.159	6.888	<.001	.165		
Trials 1–5 total sex	5.854	[3.795, 7.913]	1.045	5.601	<.001	.116		
LTPR intercept	77.377	[75.289, 79.466]	1.060	72.983	<.001		.073	16.493
LTPR age	−0.427	[−0.741, −0.113]	0.159	−2.682	.008	.029		
LTPR education	1.054	[0.408, 1.700]	0.328	3.215	.001	.041		
LOT intercept	16.972	[15.685, 18.259]	0.653	25.977	<.001		.068	6.699
LOT education	0.479	[0.221, 0.738]	0.131	3.652	<.001	.052		
LOT sex	2.152	[0.440, 3.863]	0.869	2.477	.014	.025		
LR intercept	5.591	[5.181, 6.001]	0.208	26.877	<.001		.064	2.133
LR education	0.158	[0.076, 0.240]	0.042	3.777	<.001	.056		
LR sex	0.571	[0.026, 1.116]	0.277	2.063	.040	.017		

Notes: LTPR, long-term percentage retention (100 * (Trial 7/Trial 5)); LOT, Learning over trials (Trials 1–5 total—(Trial 1*5)); LR = learning rate (Trial 5—Trial 1); b = unstandardized beta coefficient; *s.e.* = standard error of the unstandardized beta coefficient; *SD* residual = standard deviation of the residuals; Sex was coded (0 = male, 1 = female); Age and Education were mean centered, thus Age = (calendar age—64.3); Education = (the number of years of education obtained—12.7).

Comparison of the proposed norms to published norms

T-scores on primary RAVLT measures and trials 1–5 total were calculated for the independent comparison group following the method described in the previous passage. Each participant in the independent comparison group was assigned two different demographically adjusted *T*-scores; one set of *T*-scores using our proposed norms; one set applying norms from Stricker et al. (2021). Multiple regression analyses on *T*-scores were performed to investigate if the predictors sex, age, or education explained variance in *T*-scores. Because *T*-scores should already be adjusted for differences in age, education, and sex a significant result implies that *T*-scores were not adequately corrected for these demographic variables. To reduce error due to chance capitalization, a nominal alpha criterion level of $\alpha = <.01$ for omnibus ANOVAs were used for all analyses. Coefficients related to significant ANOVAs were then interpreted following a conventional α-level criterion of $p < .05$. Plots comparing *T*-scores produced by norms for trial 7 and fitted lines based on predictors age, education and sex are presented in Figure 1.

Norm calculator

The proposed norms are available in a free web-based tool that computes the regression equations. To obtain normative *T*-scores for both RAVLT measures the user simply needs to enter valid demographic values (sex, age, and years of education) and raw-scores from the RAVLT trials. The tool is implemented as a self-contained HTML/Javascript webpage, available at (https://uit.no/ressurs/uit/cerad/ravlt-calc.html) and is released as open source at (https://github.com/DDI-NO/RAVLT-calc) under Apache License, version 2.0.

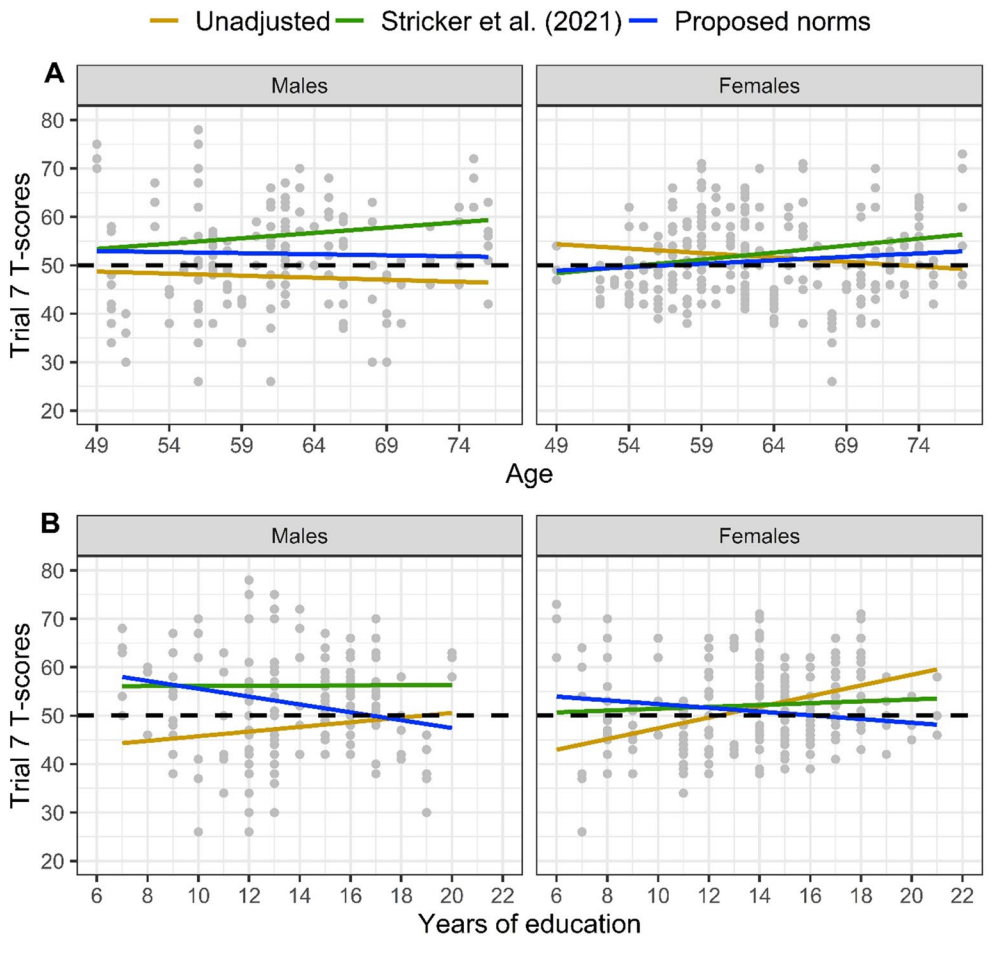

Figure 1. Linear plots of trial 7*T*-scores computed with Stricker et al. (2021) norms, unadjusted scores and proposed norms.

Test–retest reliability

A sub-set of the normative sample ($n = 98$) had available follow-up assessments on the RAVLT allowing for test–retest reliability analysis. The test–retest sample consisted of 65 women (66%) and 33 men (34%) with a mean age of 66.5 years old ($SD = 6.6$) and 12.5 ($SD = 3.2$) years of education. None of the included participants in the test–retest sample progressed to MCI, dementia or reported symptoms associated with SCD. The average time between assessments was 2.55 years ($SD = 0.53$). Intraclass correlation (ICC) estimates and 95% CIs were calculated based on a single rating, absolute-agreement two-way mixed-effects model. Values less than 0.5 are indicate poor reliability, 0.5 – 0.75 moderate reliability and 0.75–0.9 indicate good reliability (Koo & Li, 2016).

Ethics

The Norwegian Regional committees for medical and health research ethics (REK) approved the DDI project from which the current study draws upon. Guidelines in

Helsinki declaration of 1964 (revised 2013) and the Norwegian Health and Research Act were followed. The Gothenburg MCI study was approved by the local ethics committee and conducted in accordance with the Helsinki declaration. All participants gave written informed consents, including the right to withdraw and potential risks and rewards.

Results

Effects of demographics on the RAVLT test performance

The final multivariate model included significant effects of age, education, and sex across all RAVLT trials. As shown in Table 3, higher age was related to lower scores and more years of education were associated with more words recalled on the RAVLT. As expected, participants recalled more words with repeated presentation of the word list as reflected in the coefficients for trials 2–5, where the reference category is trial 1. Results showed that female participants on average recalled 0.76 more words compared to men on trial 1, adjusted for differences in age and education. Furthermore, the effects of education and sex, but not age, differed for subsequent trials of the RAVLT. The interaction term education * trial suggests that the effect of education increased for later parts of the RAVLT. The effect of education was strongest for trial 6 reflecting immediate memory recall ($b = 0.21$) and trial 7 which reflected 30 min delayed recall ($b = 0.23$). Similarly, the interaction term sex * trial indicate that the difference between men and women was most pronounced on trial 6 (b = 0.90) and trial 7 ($b = 0.71$), where women on average remembered 1.62 and 1.43 more words compared to men. List B, reflecting immediate recall of the novel word list B, did not differ significantly from the reference category trial 1. The effects of education and sex did not differ significantly on list B compared to trial 1.

On the derived measure trials 1–5 total (i.e., sum of words correctly recalled in trials 1–5), there were significant effects of age, sex, and education comparable to the observed effects on trials 1 to 5 separately. On the derived measure long-term percentage retention (LTPR), reflecting the amount of previously learned words on trial 5 retained after a 30-minute delay, lower age ($b = -0.43$), and higher education ($b = 1.05$) were significantly related to higher percent retained words (Table 4). We included two measures reflecting learning between trial 1 and trial 5 on the RAVLT, namely learning over trials (LOT) and learning rate (LR). On both measures, higher education predicted increased learning from trial 1 to trial 5 ($b = 0.48$; $b = 0.16$) and women attained significantly higher scores on learning measures than men ($b = 2.15$; $b = 0.57$). Lastly, on the measures proactive and retroactive inhibition, we found no significant effect of sex, age, or education. Because these variables followed a non-normal distribution, we report percentiles based on the inverse cumulative distribution of the normative sample (Table 5).

Adjustment of demographics using published norms

Results from multiple regression analysis on demographically adjusted *T*-scores applying norms from Stricker et al. (2021) indicated significant effects of age, education, and sex in the independent comparison group. Omnibus ANOVAs indicated that norms

Table 5. Raw scores to percentile ranks based on the inverse cumulative distribution of the normative sample ($n = 244$).

Percentile rank	Retroactive inhibition	Proactive inhibition
2	7	4
5	6	3
10	4–5	2
25	3	1
50	2	0
75	1	−1
90	0	−2
95	−1	−3

Note: Positive scores indicate inhibition effect, i.e., more words were recalled in Trial 1 compared to List B (proactive inhibition), or more words recalled in Trial 5 compared to Trial 6 (retroactive inhibition).

from Stricker et al. (2021) did not adequately adjust for demographics on trial 7 ($F(3, 141) = 5.563$, $p = .001$), trial 4 ($F(3, 139) = 6.517$, $p = <.001$), list B ($F(3, 140) = 4.690$, $p = .004$), and trials 1–5 total ($F(3, 140) = 3.379$, $p = .006$). As shown in Table 6, adjusted R^2 values indicated that age, sex, and education explained 10.4% of the total variance on trial 4, 8.7% on trial 7, 7.2% on list B, and 6.6% on trials 1–5 total. Stricker et al. (2021) norms did not adequately correct for the effect of age and sex on these trials. On trial 4, list B, trial 7, and trials 1–5 total, female participants were on average estimated 6.9, 4.6, 4.2, and 3.9, *T*-scores lower than males, respectively, and higher age predicted higher *T*-scores in all analyses. Omnibus ANOVAs indicated that the current proposed norms adequately adjusted for demographics in the independent comparison group on all measures. However, as shown in Figure 1, there was a tendency for faulty adjustment of education on trial 7, especially for males. In fact, the coefficient for education was significant ($b = -0.53$, $p = .021$), although omnibus ANOVAs indicate that the combined effect of predictors was not significant.

Test–retest reliability

Trials 6, 7, and trials 1–5 total, showed the best reliability in the follow-up sample, indicating moderate to good reliability. Trials 1 to 5 all had poor and poor-to-moderate reliability. Out of the derived measures, reliability estimates varied from poor to moderate for some trials, with retroactive inhibition and long-term percent retention showing the best reliability.

Discussion

Effects of demographics on the primary measures

We present normative data on the RAVLT based on the performance of a healthy control group from 49 to 79 years from Norway and Sweden ($n = 244$). The effect of age in this study stands out as small compared to some previous studies, which all have quoted age as the best predictor for performance (Bezdicek et al., 2014; Cavaco et al., 2015; Messinis et al., 2016; Stricker et al., 2021). On the combined measure trials 1–5 total, age explained merely 4.8% of the total variance in scores, compared

to 16.5% from education and 11.6% from sex. In other words, in this sample consisting of participants between 49 and 79 years from Norway and Sweden, we found less difference between the younger and elderly participants than expected from other studies. The weak effect of age might be due to the narrower age range comprised solely of middle-aged to elderly adults. Furthermore, the effect of age was the same for different trials on the RAVLT. The effect of age was the same for the initial learning trials as for the 30-minute delayed recall, which is consistent with some (Bezdicek et al., 2014; Messinis et al., 2016; Stricker et al., 2021), but not all studies (Boenniger et al., 2021; Cavaco et al., 2015; Lavoie et al., 2018). Weak effects of age in normative scores are not necessarily a weakness, as some studies have indicated that age-related deterioration might reflect undetected preclinical Alzheimer's disease and other pathological processes (Harrington et al., 2018; Yu et al., 2015).

Women outperformed men on all primary RAVLT trials, and the difference was greatest on trial 6 and trial 7, which has also been demonstrated in the previous studies (Sundermann et al., 2017, 2016). Stricker et al. (2021) argued for the necessity of demographically adjusted *T*-scores that incorporate sex. Their results demonstrated that women significantly outperformed men, and that failure to adjust for sex caused underestimation of amnestic mild cognitive impairment (aMCI) for women and overestimation for men. Previous studies of sex differences on the RAVLT specifically, and verbal memory in general, have found that women outperform men (Asperholm et al., 2019; Van Der Elst et al., 2005) even in samples with Alzheimer's pathology (Sundermann et al., 2017; 2016). As such, our results contribute to the collection of the previous studies that indicate the importance of adjusting for sex on the RAVLT. Despite women performing better than men on trial 7, we found no significant difference on long-term percent retention (LTPR). This suggests that the difference observed between men and women on trial 7 was mainly due to women successfully learning more words on the initial learning trials, and not better retainment of previously learned material per se. Indeed, women were able to learn more words on trial 1, reflecting attentional ability (Woodard, 2006, pp. 105–142), but also amassed more words over the subsequent trials, as reflected in the secondary measures learning over time (LOT) and total learning (TL, Ivnik et al., 1992; Vakil et al., 2010).

Previous studies in samples with comparable educational composition have generally found that education explained a substantial proportion of the variance in scores, but less so than observed in this Scandinavian sample (Bezdicek et al., 2014; Messinis et al., 2016; Stricker et al., 2021; Van Der Elst et al., 2005). This might be due to cultural differences between cohorts, possibly reflecting differences in the educational system and availability of education, or simply variation due to the estimation method. We entered education as a continuous predictor in all analyses and included participants with an extensive range of educational attainment. It is not feasible to provide a conclusive explanation for the difference between norms, particularly in terms of cultural differences, but this highlights the importance of locally sourced norms from a suitable sample that resembles the intended population. Of note, we have previously shown that on the Trail Making Test (TMT), Scandinavians with high education attainment were over-penalized (i.e., received too low *T*-scores) when applying norms from a North American sample by Heaton et al. (2004) and Espenes et al. (2020). On the other hand, Lorentzen et al. (2021) demonstrated that Scandinavians

with high educational attainment received too high *T*-scores compared to the expected normative mean on the controlled oral word association test (COWAT FAS), thus indicating that norms under-adjusted for the effects of education. In sum, we argue that local norms are necessary as the results from the current study suggests that education was more closely related to performance on the RAVLT; the effect of age was smaller; and previous investigations have found foreign norms to inadequately adjust for education when applied in a Scandinavian sample.

Evaluation of norms in an independent comparison group

A key objective of this study was to assess if the proposed norms sufficiently corrected for demographics in an independent comparison group and compare performance with norms from Stricker et al. (2021). Norms from Stricker et al. (2021) adequately corrected for demographics on all RAVLT trials, except trial 4, list B, trial 7 and trials 1–5 total (Table 6). The unstandardized coefficients for age were positive, indicating that increases in age were related to increased *T*-scores. Also, female participants were on average estimated about half a standard deviation lower *T*-scores than males. This suggests that both the generally unfavorable effect of higher age on RAVLT performance, and the difference between men and women, was exaggerated when applied in the independent comparison group. As shown in Table 6, the current proposed norms adequately adjusted for age, education, and sex on all RAVLT trials in the independent comparison group. Regarding education, norms from Stricker et al. (2021) adequately adjusted in all trials. From Figure 1 it is apparent that the current proposed norms produced *T*-scores on trial 7 that exhibited some under-adjustment, especially for males with low levels of education. Although the omnibus ANOVA indicate that the model was insignificant, the individual coefficient for education on trial 7 was significant (p = .021; Table 6). This likely stems from sample characteristics; the independent comparison group consisted of very few male participants with lower levels of education that displayed results that exceeded normative expectations. As such, we cannot guarantee the external validity of these results. However, we believe they provide some indication of the norms' ability to adjust in a Scandinavian sample and are valuable for direct comparison of normative adjustments.

Failure to adequately correct for demographics can lead to faulty estimates of the participants' performance, thus influencing the rate of correctly diagnosed patients with amnestic mild cognitive impairment (Stricker et al., 2021). In the normative sample and the independent comparison group females outperformed men on the RAVLT. Applying norms from Stricker et al. (2021) exaggerated the sex difference on the RAVLT such that males had higher *T*-scores than female participants. Over- or underestimation of performance on the RAVLT may result in missed treatment opportunities or unnecessary treatment, which may negatively affect quality of life (Stricker et al., 2021). Failure to adjust for age and education is most apparent in the end ranges of predictors, that is, for the youngest and oldest and individuals with either very low or very high levels of education. For example, a 68-year-old male with 19 years of formal education enrolled in the independent comparison group remembered 5 words on the 30-min delay on the RAVLT (trial 7). Applying norms from

Table 6. Results from multiple regression analysis on *T*-scores with predictors age, years of education, and sex in the independent comparison group ($n = 145$).

		Stricker et al.'s (2021) norms				Scandinavian norms			
Variable	Predictor	b	*p*	Partial R^2	Adj. R^2	b	*p*	Partial R^2	Adj. R^2
Trial 1	Intercept	53.573	<.001		.043	51.989	<.001		.015
	Age	0.247	.091	.020		0.205	.151	.015	
	Education	0.716	.020	.038		0.402	.181	.013	
	Sex	−2.815	.164	.014		−2.704	.173	.013	
Trial 2	Intercept	55.343	<.001		.022	52.417	<.001		.011
	Age	0.229	.115	.018		0.041	.776	<.001	
	Education	0.263	.386	.005		−0.195	.523	.003	
	Sex	−3.740	.065	.024		−4.072	.046	.028	
Trial 3	Intercept	54.799	<.001		.015	52.576	<.001		−.013
	Age	0.276	.060	.025		0.064	.650	.001	
	Education	0.302	.332	.007		−0.105	.721	<.001	
	Sex	−2.177	.283	.008		−1.758	.369	.006	
Trial 4	Intercept	58.029	<.001		.104*	53.888	<.001		−.020
	Age	0.262	.037	.031		0.039	.879	<.001	
	Education	0.229	.383	.006		−0.177	.741	<.001	
	Sex	−6.852	<.001	.10		−0.922	.796	<.001	
Trial 5	Intercept	53.499	<.001		.013	50.150	<.001		.028
	Age	0.242	.049	.027		0.065	.564	<.001	
	Education	−0.024	.925	<.001		−0.606	.012	.045	
	Sex	−1.673	.324	.007		−0.018	.991	<.001	
Trial 6	Intercept	54.875	<.001		.025	52.456	<.001		.006
	Age	0.158	.127	.017		−0.064	.572	.002	
	Education	0.246	.255	.009		−0.337	.157	.014	
	Sex	−2.731	.057	.026		−1.894	.229	.010	
Trial 7	Intercept	56.363	<.001		.087*	52.396	<.001		.027
	Age	0.267	.007	.051		0.048	.656	.001	
	Education	0.179	.384	.005		−0.527	.021	.037	
	Sex	−4.215	.002	.066		−1.566	.296	.008	
List B	Intercept	57.575	.001		.072*	50.539	<.001		.022
	Age	0.432	.003	.062		0.327	.025	.036	
	Education	0.253	.395	.005		−0.047	.879	<.001	
	Sex	−4.555	.022	.037		−2.238	.262	.009	
Trials 1–5 Total	Intercept	55.839	<.001		.066*	52.227	<.001		.015
	Age	0.377	.005	.055		0.229	.078	.022	
	Education	0.353	.205	.012		−0.104	.703	.001	
	Sex	−3.885	.035	.031		−2.505	.164	.014	

Notes: b = unstandardized regression coefficient; $p = p$-value; partial R^2 = explained variance of predictor variable; Adj. R^2 = explained variance of combined predictor variables
*Omnibus ANOVA was significant ($p = <.01$), and we report *F* and *P*-values for significant models in text; age and education was mean centered.

Stricker et al. (2021) the calculated *T*-score is $T = 43$ compared to $T = 30$ applying the current proposed norms. Thus, applying diagnostic criteria for amnestic mild cognitive impairment (aMCI, Albert et al., 2011; Bondi et al., 2014), this could have implications for correctly diagnosing aMCI and providing adequate treatment.

Test–retest reliability

Test–retest reliability is important for neuropsychological tests that are used to inform decisions on the cognitive status of patients at the time of assessments and their likely functioning in the future (Sherman et al., 2011). Trial 1, LOT and LR showed poor reliability (Table 7). This might be expected on attentional measures that typically show lower test–retest reliability as attention is considered a "changeable trait" (Sherman et al., 2011) compared to verbal memory, which may be regarded "trait-like" and stable in healthy participants. Thus, clinicians should exercise caution interpreting these measures in isolation. Instead, clinicians concerned with the reliability of test scores are recommended to use trials 1–5 total as a measure of acquisition, attention, and learning which showed moderate to good reliability. Both trial 6 and trial 7 showed moderate-to-good reliability, and LTPR showed poor-to-moderate reliability. The same pattern of test–retest reliability was reported by Stricker et al. (2021), though our reliability estimates were slightly lower overall. This is likely due to the longer test–retest interval in this study (2.5 years compared to 1.5 years) and small sample size ($n = 98$) for the follow-up group, thus inflating the associated 95% confidence intervals.

Effects of demographics on the derived measures

We provided norms for retroactive and proactive inhibition measures, which might have utility in specific clinical samples burdened with executive deficits. Proactive inhibition refers to the reduced ability to learn new material due to interference from

Table 7. Test–retest reliability of RAVLT measures based on a subset of the normative sample (n = 98).

Measure	ICC	95% CI [LL, UL]
Trial 1	0.324	[0.135, 0.491]
Trial 2	0.504	[0.335, 0.642]
Trial 3	0.511	[0.343, 0.647]
Trial 4	0.457	[0.279, 0.604]
Trial 5	0.560	[0.407, 0.682]
List B	0.549	[0.394, 0.674]
Trial 6	0.749	[0.646, 0.825]
Trial 7	0.712	[0.598, 0.797]
Trials 1–5 Total	0.659	[0.528, 0.759]
Learning over trials	0.174	[0.028, 0.363]
Learning rate	0.178	[−0.021, 0.364]
Proactive inhibition	−0.030	[−0.228, 0.17]
Retroactive inhibition	0.378	[0.193, 0.537]
Long-term percent retention	0.532	[0.372, 0.661]

Note: ICC, intraclass correlation coefficient.

the previously learned material, and is derived on the RAVLT by comparing performance on list B to trial 1. On the other hand, retroactive inhibition refers to the reduced ability to recall the previously learned material after inference has occurred (list B) and is measured by comparing performance on trial 5 with trial 6. We did not observe a significant difference between list B and trial 1 (Table 3), thus indicating no significant proactive inhibition in the normative sample on average. In line with the previous normative studies on cognitively healthy adults, we found no significant relationship between proactive inhibition and age (Boenniger et al., 2021; Vakil et al., 2010) or sex (Boenniger et al., 2021), or education. Proactive inhibition has been shown to be deficient in patients with frontal lobe lesions on a paired association test compared to healthy controls (Depue, 2012; Shimamura et al., 1995). Reduced performance on inhibition tasks has been associated with deficits in inhibition, response competition, deficits in source memory, and over-activation of irrelevant memory items (Vakil et al., 2010). As such, it may be expected to find significant deficits in inhibition (either proactive or retroactive) in clinical samples. However, as far as we are aware, there have been no studies comparing performance on these measures in samples with MCI or AD dementia on the RAVLT. Some retroactive inhibition appears to be normal, as participants on average remembered about 2 fewer words on trial 6 than trial 5 (Tables 1 and 3). Patients with schizophrenia have been shown to be susceptible to retroactive inhibition, owning to executive demands associated with retroactive inhibition. Specifically, the ability to; inhibit responses, verbal fluency to govern retrieval of target items; and memory of temporal order (Torres et al., 2001). Boenniger et al. (2021) and Vakil et al. (2010) found a small effect of age on retroactive inhibition, and Boenniger reported that men presented slightly more retroactive inhibition than women. Nevertheless, we found no significant effect of age, sex, or education on retroactive inhibition. Due to the lack of association to demographic variables, we simply report percentile ranks on these measures for clinicians to inform decisions on abnormal/normal performance.

Limitations

Some limitations of the current study are to be addressed. Firstly, participants were not formally screened for auditory deficits which might influence performance on the RAVLT. However, all participants with hearing aids were instructed to use these when applicable. The normative sample from which norms were computed was not a randomized sample of the Norwegian and Swedish population. We therefore cannot guarantee that this sample reflects the population in general. However, this limitation is not specific to this study. Still, it remains a common issue in the normative literature, with exceptions such as the Mayo normative study (Stricker et al., 2021) and the Rhineland study (Boenniger et al., 2021). Also, compared to some previous studies, the normative sample of the current study is relatively small, which influences the degree of certainty that a normative score reflects the true population parameters, especially for extreme scores (e.g., 1.5 *SD* below the sample mean, Crawford & Garthwaite, 2008; Oosterhuis et al., 2016). Lastly, while the experience of SCD is generally considered a normal and benign condition in an aging population (Bassett &

Folstein, 1993; Hessen et al., 2017), it is nevertheless a known risk-factor of neurodegenerative disease (Jessen et al., 2014). However, all included participants were cognitively healthy at the time of analysis, also supported by mean RAVLT scores being largely equivalent to our Normative sample (Table 1).

Conclusion

We propose regression-based test norms for the RAVLT based on a sample of healthy Swedish and Norwegian participants between 49 and 79 years old. A free online norm calculator is offered to improve availability of norms in clinical settings. Test–retest reliability analyses indicated that basic RAVLT trials showed poor-to-moderate reliability, while measures of total learning and verbal memory showed moderate-to-good reliability. Our results indicate that the current proposed norms successfully adjust for age, education, and sex in the independent comparison group. Norms from Stricker et al. (2021) overestimated the effect of age and difference between sexes on parts of the RAVLT. Notably, the failure to adequately adjust for demographical variables on the 30-min delayed recall (trial 7) might have implications for correctly diagnosing amnestic mild cognitive impairment (aMCI) in Scandinavian adults and elderly.

Acknowledgments

We thank Ragna Espenes and Ramune Grambaite for clinical examinations and help with the project.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the University of Tromsø - the Arctic University of Norway; the Norwegian Research Council, JPND/PMI-AD under grant number 311993; and Helse Nord under grant number HNF1540-20. Additional support was received from the Sahlgrenska University Hospital, the Swedish Research Council, Swedish Brain Power, the Swedish Dementia Foundation, the Swedish Alzheimer's Foundation, Stiftelsen Psykiatriska forskningsfonden, and Konung Gustaf V:s och Drottning Victorias Frimurarestiftelse, Demensförbundet, Helse Nord RHF, and EU Joint Programme - Neurodegenerative Disease Research. The funding sources were not involved in the drafting of this manuscript.

ORCID

Jacob Espenes http://orcid.org/0000-0002-2383-5348
Ingvild Vøllo Eliassen http://orcid.org/0000-0003-1288-6032
Fredrik Öhman http://orcid.org/0000-0002-6872-4481
Knut Waterloo http://orcid.org/0000-0003-3447-8312
Ingrid Myrvoll Lorentzen http://orcid.org/0000-0002-9942-7594
Tormod Fladby http://orcid.org/0000-0002-9984-9797
Bjørn-Eivind Kirsebom http://orcid.org/0000-0002-1413-9578

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy of the research participants.

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

Asperholm, M., Nagar, S., Dekhtyar, S., & Herlitz, A. (2019). The magnitude of sex differences in verbal episodic memory increases with social progress: Data from 54 countries across 40 years. *PloS One*, *14*(4), e0214945. https://doi.org/10.1371/journal.pone.0214945

Bassett, S. S., & Folstein, M. F. (1993). Memory complaint, memory performance, and psychiatric diagnosis: A community study. *Journal of Geriatric Psychiatry and Neurology*, *6*(2), 105–111. https://doi.org/10.1177/089198879300600207

Belleville, S., Fouquet, C., Hudon, C., Zomahoun, H. T. V., & Croteau, J, Consortium for the Early Identification of Alzheimer's, d.-Q. (2017). Neuropsychological measures that predict progression from mild cognitive impairment to Alzheimer's type dementia in older adults: A systematic review and meta-analysis. *Neuropsychology Review*, *27*(4), 328–353. Retrieved from https://doi.org/10.1007/s11065-017-9361-5

Bezdicek, O., Stepankova, H., Moták, L., Axelrod, B. N., Woodard, J. L., Preiss, M., Nikolai, T., Růžička, E., & Poreh, A. (2014). Czech version of Rey Auditory Verbal Learning Test: Normative data. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition*, *21*(6), 693–721. https://doi.org/10.1080/13825585.2013.865699

Boake, C. (2000). Edouard Claparede and the auditory verbal learning test. *Journal of Clinical and Experimental Neuropsychology*, *22*(2), 286–292. https://doi.org/10.1076/1380-3395(200004)22:2;1-1;FT286

Boenniger, M. M., Staerk, C., Coors, A., Huijbers, W., Ettinger, U., & Breteler, M. M. B. (2021). Ten German versions of Rey's auditory verbal learning test: Age and sex effects in 4,000 adults of the Rhineland Study. *Journal of Clinical and Experimental Neuropsychology*, *43*(6), 637–653. https://doi.org/10.1080/13803395.2021.1984398.

Bondi, M. W., Edmonds, E. C., Jak, A. J., Clark, L. R., Delano-Wood, L., McDonald, C. R., Nation, D. A., Libon, D. J., Au, R., Galasko, D., & Salmon, D. P. (2014). Neuropsychological criteria for mild cognitive impairment improves diagnostic precision, biomarker associations, and progression rates. *Journal of Alzheimer's Disease: JAD*, *42*(1), 275–289. https://doi.org/10.3233/JAD-140276

Cavaco, S., Gonçalves, A., Pinto, C., Almeida, E., Gomes, F., Moreira, I., Fernandes, J. & Teixeira-Pinto, A. (2015). Auditory Verbal Learning Test in a Large Nonclinical Portuguese Population. *Applied Neuropsychology. Adult*, *22*(5), 321–331. https://doi.org/10.1080/23279095.2014.927767 25580839

Correia, A. F., & Osorio, I. C. (2014). The Rey Auditory Verbal Learning Test: normative data developed for the Venezuelan population. *Arch Clin Neuropsychol*, *29*(2), 206–215. https://doi.org/10.1093/arclin/act070

Crawford, J. R., & Garthwaite, P. H. (2008). On the "optimal" size for normative samples in neuropsychology: Capturing the uncertainty when normative data are used to quantify the standing of a neuropsychological test score. *Child Neuropsychology : a Journal on Normal and Abnormal Development in Childhood and Adolescence*, *14*(2), 99–117. https://doi.org/10.1080/09297040801894709

Depue, B. E. (2012). A neuroanatomical model of prefrontal inhibitory modulation of memory retrieval. *Neuroscience and Biobehavioral Reviews*, *36*(5), 1382–1399.

Eckerström, C., Olsson, E., Bjerke, M., Malmgren, H., Edman, Å., Wallin, A., & Nordlund, A. (2013). A combination of neuropsychological, neuroimaging, and cerebrospinal fluid markers predicts

conversion from mild cognitive impairment to dementia. *Journal of Alzheimer's Disease: JAD*, *36*(3), 421–431. https://doi.org/10.3233/JAD-122440

Eliassen, I. V., Fladby, T., Kirsebom, B.-E., Waterloo, K., Eckerström, M., Wallin, A., Bråthen, G., Aarsland, D., & Hessen, E. (2020). Predictive and diagnostic utility of brief neuropsychological assessment in detecting Alzheimer's pathology and progression to dementia. *Neuropsychology*, *34*(8), 851–861. https://doi.org/10.1037/neu0000698

Espenes, J., Hessen, E., Eliassen, I. V., Waterloo, K., Eckerström, M., Sando, S. B., Timón, 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*(sup1), 110–126. https://doi.org/10.1080/13854046.2020.1829068

Estévez-González, A., Kulisevsky, J., Boltes, A., Otermín, P., & García-Sánchez, C. (2003). Rey verbal learning test is a useful tool for differential diagnosis in the preclinical phase of Alzheimer's disease: Comparison with mild cognitive impairment and normal aging. *International Journal of Geriatric Psychiatry*, *18*(11), 1021–1028.

Ewers, M., Walsh, C., Trojanowski, J. Q., Shaw, L. M., Petersen, R. C., Jack, C. R., Feldman, H. H., Bokde, A. L. W., Alexander, G. E., Scheltens, P., Vellas, B., Dubois, B., Weiner, M., & Hampel, H, North American Alzheimer's Disease Neuroimaging Initiative (ADNI). (2012). Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. *Neurobiology of Aging*, *33*(7), 1203–1214. e1202. https://doi.org/10.1016/j.neurobiolaging.2010.10.019

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 : The Journal of the Alzheimer's Association*, *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: JAD*, *60*(1), 97–105. https://doi.org/10.3233/JAD-170231

Harrington, K. D., Schembri, A., Lim, Y. Y., Dang, C., Ames, D., Hassenstab, J., Laws, S. M., Rainey-Smith, S., Robertson, J., Rowe, C. C., Sohrabi, H. R., Salvado, O., Weinborn, M., Villemagne, V. L., Masters, C. L., & Maruff, P, AIBL Research Group. (2018). Estimates of age-related memory decline are inflated by unrecognized Alzheimer's disease. *Neurobiology of Aging*, *70*, 170–179.

Heaton, R., Miller, S., 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*. Psychological Assessment Resources.

Hessen, E., Eckerström, M., Nordlund, A., Selseth Almdahl, I., Stålhammar, J., Bjerke, M., Eckerström, C., Göthlin, M., Fladby, T., Reinvang, I., & Wallin, A. (2017). Subjective cognitive impairment is a predominantly benign condition in memory clinic patients followed for 6 years: The Gothenburg-Oslo MCI study. *Dementia and Geriatric Cognitive Disorders Extra*, *7*(1), 1–14. Retrieved from https://doi.org/10.1159/000454676

Hessen, E., Reinvang, I., Eliassen, C. F., Nordlund, A., Gjerstad, L., Fladby, T., & Wallin, A. (2014). The combination of dysexecutive and amnestic deficits strongly predicts conversion to dementia in young mild cognitive impairment patients: A report from the Gothenburg-Oslo MCI study. *Dementia and Geriatric Cognitive Disorders Extra*, *4*(1), 76–85. https://doi.org/10.1159/000360282

Ivnik, R. J., Malec, J. F., Smith, G. E., Tangalos, E. G., Petersen, R. C., Kokmen, E., & Kurland, L. T. (1992). Mayo's older Americans normative studies: Updated AVLT norms for ages 56 to 97. *Clinical Neuropsychologist*, *6*(sup001), 83–104. https://doi.org/10.1080/13854049208401880

Ivnik, R. J., Malec, J. F., Tangalos, E. G., Petersen, R. C., Kokmen, E., & Kurland, L. T. (1990). The auditory-verbal learning test (AVLT): norms for ages 55 years and older. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, *2*(3), 304–312. https://doi.org/10.1037/1040-3590.2.3.304

JASP Team (2022). JASP (Version 0.16.3)[Computer software]..

Jessen, F., Amariglio, R. E., Van Boxtel, M., Breteler, M., Ceccaldi, M., Chételat, G., … Van Der Flier, W. M. (2014). A Conceptual Framework for Research on Subjective Cognitive Decline in Preclinical Alzheimer's Disease. *Alzheimer's & Dementia*, *10*(6), 844–852.

Kirsebom, B.-E., Espenes, R., Hessen, E., Waterloo, K., Johnsen, S. H., Gundersen, E., Botne Sando, S., Rolfseng Grøntvedt, G., Timón, S., & Fladby, T. (2019). Demographically adjusted CERAD wordlist test norms in a Norwegian sample from 40 to 80 years. *The Clinical Neuropsychologist*, *33*(sup1), 27–39. https://doi.org/10.1080/13854046.2019.1574902.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Lavoie, M., Bherer, L., Joubert, S., Gagnon, J.-F., Blanchet, S., Rouleau, I., Macoir, J., & Hudon, C. (2018). Normative data for the Rey Auditory Verbal Learning Test in the older French-Quebec population. *The Clinical Neuropsychologist*, *32*(sup1), 15–28. https://doi.org/10.1080/13854046.2018.1429670

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment*. (5th ed.). Oxford University Press.

Lorentzen, I. M., Espenes, J., Hessen, E., Waterloo, K., Bråthen, G., Timón, S., … 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*, 1–10.

Loring, D. W., Strauss, E., Hermann, B. P., Barr, W. B., Perrine, K., Trenerry, M. R., Chelune, G., Westerveld, M., Lee, G. P., Meador, K. J., & Bowden, S. C. (2008). Differential neuropsychological test sensitivity to left temporal lobe epilepsy. *Journal of the International Neuropsychological Society*, *14*(03), 394–400. https://doi.org/10.1017/S1355617708080582

Marqués, N. O., Caro, I. A., Uterga Valiente, J. M., & Rodríguez, S. M. (2013). Normative data for a Spanish version of the Rey Auditory-Verbal Learning Test in older people. *The Spanish Journal of Psychology*, *16*, E60. https://www.cambridge.org/core/journals/spanish-journal-of-psychology/article/abs/normative-data-for-a-spanish-version-of-the-rey-auditoryverbal-learning-test-in-older-people/6E3C9F4A3D9677B29C2AAE630E2DF1CF.

Messinis, L., Nasios, G., Mougias, A., Politis, A., Zampakis, P., Tsiamaki, E., Malefaki, S., Gourzis, P., & Papathanasopoulos, P. (2016). Age and education adjusted normative data and discriminative validity for Rey's Auditory Verbal Learning Test in the elderly Greek population. *Journal of Clinical and Experimental Neuropsychology*, *38*(1), 23–39. https://doi.org/10.1080/13803395.2015.1085496

Mitrushina, M., Boone, K. B., Razani, J., & D'Elia, L. F. (2005). *Handbook of normative data for neuropsychological assessment*. : Oxford University Press.

Molinuevo, J. L., Rabin, L. A., Amariglio, R., Buckley, R., Dubois, B., Ellis, K. A., Ewers, M., Hampel, H., Klöppel, S., Rami, L., Reisberg, B., Saykin, A. J., Sikkes, S., Smart, C. M., Snitz, B. E., Sperling, R., van der Flier, W. M., Wagner, M., & Jessen, F, Subjective Cognitive Decline Initiative (SCD-I) Working Group. (2017). Implementation of subjective cognitive decline criteria in research studies.)*Alzheimer's & Dementia : The Journal of the Alzheimer's Association*, *13*(3), 296–311. https://doi.org/10.1016/j.jalz.2016.09.012

Oosterhuis, H. E. M., van der Ark, L. A., & Sijtsma, K. (2016). Sample size requirements for traditional and regression-based norms. *Assessment*, *23*(2), 191–202. https://doi.org/10.1177/1073191115580638

Powell, J. B., Cripe, L. I., & Dodrill, C. B. (1991). R: A language and environment for statistical computing. *Archives of Clinical Neuropsychology*, *6*(4), 241–249. https://doi.org/10.1093/arclin/6.4.241

R Core Team (2020). Assessment of brain impairment with the Rey Auditory Verbal Learning Test: A comparison with other neuropsychological measures. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

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

Rey, A. (1958). L'examen clinique en psychologie. [The clinical examination in psychology]. Presses Universitaries De France.

Ricci, M., Graef, S., Blundo, C., & Miller, L. A. (2012). Using the Rey Auditory Verbal Learning Test (RAVLT) to differentiate Alzheimer's dementia and behavioural variant fronto-temporal dementia. *The Clinical Neuropsychologist*, *26*(6), 926–941.

Sherman, E., Brooks, B. L., Iverson, G. L., Slick, D. J., & Strauss, E. (2011). Reliability and validity in neuropsychology. In *The little black book of neuropsychology* (pp. 873–892): Springer.

Shimamura, A. P., Jurica, P. J., Mangels, J. A., Gershberg, F. B., & Knight, R. T. (1995). Susceptibility to memory interference effects following frontal lobe damage: Findings from tests of paired-associate learning. *Journal of Cognitive Neuroscience*, *7*(2), 144–152.

Strauss, E., Sherman, E. M., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary*. American Chemical Society.

Stricker, N. H., Christianson, T. J., Lundt, E. S., Alden, E. C., Machulda, M. M., Fields, J. A., Kremers, W. K., Jack, C. R., Knopman, D. S., Mielke, M. M., & Petersen, R. C. (2021). Mayo normative studies: Regression-based normative data for the auditory verbal learning test for ages 30-91 years and the importance of adjusting for sex. *Journal of the International Neuropsychological Society*, *27*(3), 211–226. https://doi.org/10.1017/S1355617720000752

Sundermann, E. E., Biegon, A., Rubin, L. H., Lipton, R. B., Landau, S., Maki, P. M., for the Alzheimer's, D., & Neuroimaging, I, Alzheimer's Disease Neuroimaging Initiative (2017). Does the female advantage in verbal memory contribute to underestimating Alzheimer's disease pathology in women versus men? *Journal of Alzheimer's Disease: JAD*, *56*(3), 947–957. https://doi.org/10.3233/JAD-160716

Sundermann, E. E., Biegon, A., Rubin, L. H., Lipton, R. B., Mowrey, W., Landau, S., Initiative, F., … t, A, s. D. N (2016). Better verbal memory in women than men in MCI despite similar levels of hippocampal atrophy. *Neurology*, *86*(15), 1368–1376. Retrieved from https://doi.org/10.1212/wnl.0000000000002570.

Torres, I. J., Flashman, L. A., O'Leary, D. S., & Andreasen, N. C. (2001). Effects of retroactive and proactive interference on word list recall in schizophrenia. *Journal of the International Neuropsychological Society : JINS*, *7*(4), 481–490. https://doi.org/10.1017/S1355617701744049

Vakil, E., Greenstein, Y., & Blachstein, H. (2010). Normative Data for Composite Scores for Children and Adults Derived from the Rey Auditory Verbal Learning Test. *The Clinical Neuropsychologist*, *24*(4), 662–677. https://doi.org/10.1080/13854040903493522

Van der Elst, W., Molenberghs, G., van Tetering, M., & Jolles, J. (2017). Establishing normative data for multi-trial memory tests: the multivariate regression-based approach. *The Clinical Neuropsychologist*, *31*(6/7), 1173–1187.

Van Der Elst, W., Van Boxtel, M. P., Van Breukelen, G. J., & Jolles, J. (2005). Rey's verbal learning test: normative data for 1855 healthy participants aged 24–81 years and the influence of age, sex, education, and mode of presentation. *Journal of the International Neuropsychological Society : JINS*, *11*(3), 290–302. https://doi.org/10.1017/S1355617705050344

Verbeke, G., & Molenberghs, G. (2009). *Linear mixed models for longitudinal data*. Springer Science & Business Media.

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 : official Journal of the International Society of Cerebral Blood Flow and Metabolism*, *36*(1), 114–131. https://doi.org/10.1038/jcbfm.2015.147

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

Woodard, J. L. (2006). Memory performance indexes for the Rey Auditory Verbal Learning Test. In *The quantified process approach to neuropsychological assessment*. (pp. 105–141). Taylor & Francis.

Yu, L., Boyle, P. A., Segawa, E., Leurgans, S., Schneider, J. A., Wilson, R. S., & Bennett, D. A. (2015). Residual decline in cognition after adjustment for common neuropathologic conditions. *Neuropsychology*, *29*(3), 335–343.

Appendix A

A.1. RAVLT test version and administration procedures in Norwegian and Swedish

Table A.1. Wordlists for RAVLT.

Norwegian		Swedish		English	
List A	List B	List A	List B	List A	List B
Tromme	Bord	Trumma	Skrivebord	Drum	Desk
Gardin	Jeger	Gardin	Polis	Curtain	Ranger
Måne	Fugl	Måne	Fågel	Bell	Bird
Kaffe	Sko	Kaffe	Sko	Coffee	Shoe
Skole	Ovn	Skola	Spis	School	Stove
Foreldre	Fjell	Bror	Berg	Parent	Mountain
Klokke	Briller	Klocka	Glas	Moon	Glasses
Hage	Håndkle	Trädgård	Penna	Garden	Towel
Hatt	Sky	Hatt	Moln	Hat	Cloud
Bonde	Båt	Bonde	Båt	Farmer	Boat
Nese	Lam	Nos	Lamm	Nose	Lamb
Kalkun	Pistol	Kalkon	Pistol	Turkey	Gun
Farge	Blyant	Färg	Handduk	Color	Pencil
Hus	Kirke	Hus	Kyrka	House	Church
Elv	Fisk	Flod	Fisk	River	Fish

Table A.1.2. Norwegian items for RAVLT recognition.

Klokke	Hjem	Håndkle	Båt	Briller
J N	J N	J N	J N	J N
Vindu	Fisk	Gardin	Varm	Strømpe
J N	J N	**J** N	J N	J N
Hatt	Måne	Blomst	Foreldre	Sko
J N	**J** N	J N	**J** N	J N
Låve	Tre	Farge	Vann	Laerer
J N	J N	**J** N	J N	J N
Jeger	Ballong	Bord	Bonde	Ovn
J N	J N	J N	**J** N	J N
Nese	Fugl	Gevaer	Rose	Rede
J N	J N	J N	J N	J N
Vaer	Fjell	Fargestift	Sky	Barn
J N	J N	J N	J N	J N
Skole	Kaffe	Kirke	Hus	Tromme
J N	**J** N	J N	**J** N	**J** N
Hånd	Mus	Kalkun	Fremmed	Karamell
J N	J N	**J** N	J N	J N
Blyant	Elv	Kilde	Hage	Lam
J N	J N	J N	**J** N	J N

Note. We do not provide normative data for the recognition trial. Correct items from list A are highlighted in bold text.

A.2. Swedish Administration procedures

ADMINISTRERING A1 - A6

En ordlista bestående av 15 ord läses upp för patienten.

Jag kommer att läsa upp en lista med ord för dig, och jag vill att du försöker lägga orden på minnet. När jag har läst listan klart, så vill jag att du säger de ord du kan minnas. Det är många ord, så du kommer inte att kunna minnas alla, men försök minnas så många du kan.

Läs listan i ett tempo av ungefärligen ett ord per sekund. När patienten återger, notera i protokollet vilken ordning orden återges i, samt eventuella upprepningar och konfabulationer. När patienten har varit tyst en stund, fråga om hon minns något mer. En del patienter ger upp snabbt vid denna uppgift och kan behöva uppmuntras att försöka tänka en liten stund till.

Nu kommer jag att läsa den här listan några gånger. Efter varje gång vill jag att du räknar upp de ord du minns. Du ska också ta med de ord som du har sagt tidigare.

Efter den femte retentionen läses en distraktionslista bestående av nya ord upp, och patienten ska återge ord från den nya listan.

Nu kommer jag att läsa en lista med helt nya ord. Även nu vill jag att du försöker att minnas dem, och sedan säga de ord du kan komma ihåg när jag har läst listan färdigt. Den här listan kommer jag bara att läsa en gång.

Patienten uppmanas därefter att återge vad hon nu minns från den första listan. Efter moment A6 går testledaren vidare i protokollet med övriga uppgifter i ca 30 minuter.

ADMINISTRERING A7

Efter ca 30 minuter ombeds patienten igen att dra sig till minnes den första listan.

A.3. Norwegian administration procedures

Administrering liste A, første presentasjon (trial 1).

Jeg vil nå lese opp en liste med ord. Hør nøye etter, for når jeg er ferdig vil jeg at du skal gjenta så mange som du kan huske. Rekkefølgen du sier det i har ingenting å si. Bare prøv å husk så mange du kan.

Liste A, andre presentasjon (trial 2).

Nå vil jeg lese den samme listen med ord igjen og på samme måte vil jeg at du skal gjenta så mange ord som du kan huske, inkludert de ordene du sa første gangen. Rekkefølgen som du sier ordene har ingenting å si, bare gjenta så mange ord du klarer uansett om du sa det første gang.

Gjenta instruksjonen ved behov for trial 3-5.

Direkte etter femte presentasjon skal liste B administreres. Si:

Jeg vil nå lese opp en ny liste med ord, og på samme måte som før skal du prøve huske så mange ord som mulig fra denne nye listen. Rekkefølgen du sier det i har ingenting å si.

Uten fornyet presentasjon av liste A skal testdeltager gjenkalle liste A på nytt. Si:

Kan du på nytt si alle ordene du husker fra den første listen?

Etter 30 minutter skal testdeltager gjenta liste A for siste gang. Si:

For litt siden leste jeg opp en liste med ord til deg flere ganger og du skulle forsøke laere disse ordene. Kan du gjenta disse ordene en gang til?

A.4. Regression norming procedure

Table A.2. Model selection procedure for the multivariate normative model (n = 244).

Model	Model structure	Cov. Structure	df	−2 log lik difference (G2)	*p*-value	Ref. model	Qualitative conclusion	BIC	AIC
1	Full	UN	74	−3550.00				7660.66	7247.99
2	Exclude Cohort	UN	72	1.71	.425	1	Exclude Cohort	7647.22	7245.70
3	Exclude edu^2	UN	71	0.004	.951	2	Exclude edu^2	7639.65	7243.71
4	Exclude age^2	UN	70	0.375	.549	3	Exclude age^2	7632.45	7242.08
5	Exclude age*edu	UN	69	0.18	.668	4	Exclude age*edu	7625.05	7240.27
6	Exclude age*sex	UN	68	0.19	.663	5	Exclude age*sex	7617.67	7238.46
7	Exclude trial*age	UN	61	8.23	.313	6	Exclude trial*age	7572.86	7232.69
8	Exclude trial*sex	UN	54	20.02	.005	7	Keep trial*sex	7539.85	7238.71
9	Exclude trial*edu	UN	54	20.87	.004	7	Keep trial*edu	7540.69	7239.56
10	Exclude age	UN	60	10.59	.001	7	Keep age	7575.88	7241.28
11	Same as model 7	AR(1) heterogeneous	34	355.58	<.001	7		7534.27	7723.87
12	Same as model 7	AR(1) homogeneous	27	502.63	<.001	7		7817.89	7667.32
13	Same as model 7	CS homogeneous	27	512.86	<.001	7		7828.12	7677.55
14	Same as model 7	CS heterogenous	34	381.46	<.001	7		7560.15	7560.15

Notes: Cov. Structure, covariance structure; BIC, Bayesian Information Criteron; AIC, Akaike Information Criterion; Edu, years of formal education; Sex was coded 0 = male and 1 = female; Edu and age were mean centered; AR(1), first-order autoregressive covariance structure; and CS, compound symmetry.

Table A.2.1. Pearson's correlations between RAVLT trial scores and the demographic variables ($n = 244$).

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Trial 7	1-5 total	List B	Age	Sex	Edu
Trial 1	1											
Trial 2	.624	1										
Trial 3	.572	.776	1									
Trial 4	.493	.682	.772	1								
Trial 5	.484	.688	.742	.792	1							
Trial 6	.412	.641	.740	.742	.788	1						
Trial 7	.446	.674	.883	.740	.788	.870	1					
1–5 total	.712	.878	.909	.883	.878	.769	.789	1				
List B	.433	.442	.478	.442	.416	.318	.292	.513	1			
Age	−.225	−.311	−.275	−.206	−.246	−.239	−.269	−.290	−.173	1		
Sex	.242	.324	.306	.240	.287	.284	.252	.330	.118	−.122	1	
Edu	.273	.372	.367	.360	.404	.383	.399	.410	.356	−.168	.010	1

Note: All coefficients are zero-order correlations; Edu = years of education; Sex was coded (0 = male, 1 = female); 1–5 total = trials 1–5 total.