Scandinavian Journal of Psychology, 2023 DOI: 10.1111/sjop.12966

Empirical Article

Demographically adjusted Rey–Osterrieth Complex Figure Test norms in a Swedish and Norwegian cohort aged 49–77 years and comparison with North American norms

FREDRIK ÖHMAN,[1,2] MARIE ECKERSTRÖM,[3] ERIK HESSEN,[4,5] JACOB ESPENES,[6,7] INGVILD V. ELIASSEN,[4,5] INGRID M. LORENTZEN,[6] JACOB STÅLHAMMAR,[1] PETRONELLA KETTUNEN,[1] MICHAEL SCHÖLL,[1,2,8] TORMOD FLADBY,[5,9] ANDERS WALLIN[1] and BJØRN-EIVIND KIRSEBOM[6,7]

[1]*Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden*
[2]*Wallenberg Centre for Molecular and Translational Medicine, University of Gothenburg, Gothenburg, Sweden*
[3]*Department of Psychology, University of Gothenburg, Gothenburg, Sweden*
[4]*Department of Psychology, University of Oslo, Oslo, Norway*
[5]*Department of Neurology, Akershus University Hospital, Lørenskog, Norway*
[6]*Department of Psychology, Faculty of Health Sciences, The Arctic University of Norway, Tromsø, Norway*
[7]*Department of Neurology, University Hospital of North Norway, Tromsø, Norway*
[8]*Dementia Research Centre, Queen Square Institute of Neurology, University College London, London, UK*
[9]*Institute of Clinical Medicine, Campus Ahus, University of Oslo, Oslo, Norway*

Öhman, F., Eckerström, M., Hessen, E., Espenes, J., Eliassen, I. V., Lorentzen, I. M., Stålhammar, J., Kettunen, P., Schöll, M., Fladby, T., Wallin, A. & Kirsebom, B.-E. (2023). Demographically adjusted Rey–Osterrieth Complex Figure Test norms in a Swedish and Norwegian cohort aged 49–77 years and comparison with North American norms. *Scandinavian Journal of Psychology*.

Introduction
The Rey–Osterrieth Complex Figure Test (RCFT) is one of the most commonly used neuropsychological tests in Sweden and Norway. However, no publications provide normative data for this population. The objective of this study was to present demographically adjusted norms for a Swedish and Norwegian population and to evaluate these in an independent comparison group.

Methods
The RCFT was administrated to 344 healthy controls recruited from the Swedish Gothenburg MCI study, the Norwegian Dementia Disease Initiation study, and the Swedish Cardiopulmonary Bioimage Study. Age ranged from 49 to 77 years (mean = 62.4 years, *SD* = 5.0 years), and education ranged from 6 to 24 years (mean = 13.3 years, *SD* = 3.0 years). Using a regression-based procedure, we investigated the effects of age, sex, and years of education on test performance. We compared and evaluated our Swedish and Norwegian norms with North American norms in an independent comparison group of 145 individuals.

Results
In healthy controls, age and education were associated with performance on the RCFT. When comparing normative RCFT performance in an independent comparison group, North American norms generally overestimated immediate and delayed recall performance. In contrast, our Swedish and Norwegian norms appear to better take into account factors of age and education.

Conclusions
We presented demographically adjusted norms for the RCFT in a Swedish and Norwegian sample. This is the first normative study of the RCFT that presents normative data for this population. In addition, we showed that North American norms might produce inaccurate normative estimations in an independent comparison group.

Key words: Norms, cross-cultural neuropsychology, Swedish, Norwegian, RCFT, Scandinavian.

Fredrik Öhman, Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Wallinsgatan 6, 431 41, Mölndal, Gothenburg, Sweden. E-mail: fredrik.ohman@gu.se

INTRODUCTION

The Rey–Osterrieth Complex Figure Test (RCFT) is one of the most widely used neuropsychological tests (Rabin, Barr & Burton, 2005) in Swedish and Norwegian clinical practice (Berg & Billman, 2009; Ryder, 2021). Despite its common use in Sweden and Norway, no demographically adjusted norms have been published for this population. The complex figure task was originally developed in 1941 by André Rey (Rey, 1941) but was standardized in 1944 by Paul-Alexandre Osterrieth (Osterrieth, 1944), resulting in the RCFT. The initial purpose of the RCFT was to assess visuospatial constructional ability and visual memory. However, it has also proved helpful for assessing executive functions such as planning, organization, and fine motor skills. The time it takes to copy the figure may indicate reduced processing speed (Lezak, Howieson, Bigler & Tranel, 2012).

At the task's start, the person is asked to copy a complex geometric figure on a blank sheet of paper. Reproducing the figure is a challenging task involving organizing the figure into a meaningful perceptual unit. After copy completion, the stimulus and copy are removed from the patient's sight. Shortly after this,

the patient is asked to reproduce the previously copied figure from memory (immediate recall); then, after 30 min, a delayed recall is administrated. Philip Fasteneau (Fasteneau, 2003) and John Meyers (Meyers & Meyers, 1995a) expanded the task by introducing a recognition part. This study used the administration outlined in the professional manual (Meyers & Meyers, 1995a).

The RCFT has established sensitivity to cognitive impairment in a variety of clinical populations (Strauss *et al.*, 2006), including mild cognitive impairment (MCI) and dementia due to Alzheimer's disease (AD) (Kasai *et al.*, 2006), solvent exposure, traumatic head injury (Ashton, Donders & Hoffman, 2005), schizophrenia (Calev, Edelist, Kugelmass & Lerer, 1991), and vascular cognitive disease (Cherrier, Mendez, Dave & Perryman, 1999). Furthermore, the RCFT has demonstrated adequate psychometric properties, with an inter-rater reliability of copy ranging from 0.80 (Berry, Allen & Schmitt, 1991) to 0.96 (Meyers & Meyers, 1995a) and recall from 0.93 (Meyers & Meyers, 1995a) to 0.99 (Caffarra, Vezzadini, Dieci, Zonato & Venneri, 2002). Test–retest reliability has been reported as 0.76 (immediate recall), 0.88 (delayed recall), and 0.87 (recognition) after about 6 months (Meyers & Meyers, 1995a).

Several norms exist for older healthy populations from North America (Berry, Allen & Schmitt, 1991; Brauer Boone, Lesser, Hill-Gutierrez, Berman & D'Elia, 1993; Fastenau, Denburg & Hufford, 1999; Machulda *et al.*, 2007), but also for Czech (Drozdová, Štěpánková, Lukavský, Bezdíček & Kopeček, 2015), Spanish (Peña-Casanova *et al.*, 2009), Latin-American (Rivera *et al.*, 2015), Italian (Caffarra, Vezzadini, Dieci, Zonato & Venneri, 2002), Canadian (Tremblay *et al.*, 2015), Lebanese (Darwish, Zeinoun, Farran & Fares, 2018), and Greek (Tsatali *et al.*, 2022) populations. Among the most widely used norms are those included in the professional manual (Meyers & Meyers, 1995a).

Various studies have suggested that demographic variables (age, education, and sex) influence performance on the RCFT. A relationship between age and performance on the copy, immediate, and delayed recall tasks has been shown, indicating lower performance in higher age-groups, especially for memory performance (Chervinsky, Mitrushina & Satz, 1992; Meyers & Meyers, 1995a). Higher education is often associated with better scores (Peña-Casanova *et al.*, 2009; Tremblay *et al.*, 2015), but the effect is smaller compared with age (Fastenau, Denburg & Hufford, 1999). The effect of sex is equivocal; some studies report sex differences (Gallagher & Burke, 2007), but most report no difference in performance depending on sex (Brauer Boone, Lesser, Hill-Gutierrez, Berman & D'Elia, 1993; Peña-Casanova *et al.*, 2009; Tremblay *et al.*, 2015).

When assessing patients in clinical settings, appropriate normative data are used to interpret test scores correctly. Ideally, a comparison between a single person's scores and normative data uses comparison material derived from a group as similar to the person assessed as possible (e.g., in terms of age and education). Inversely, employing norms from populations differing in important aspects may lower the validity of the assessment (Lezak, Howieson, Bigler & Tranel, 2012). In Scandinavia, only Danish (Vogel, Stokholm & Jørgensen, 2012) norms of healthy populations have been published for this test. Also, the Danish norms are based on a relatively small and potentially biased sample, limiting their usefulness in clinical use in Sweden and Norway. Our first aim was, therefore, to present norms (hereinafter referred to as "proposed norms") for this region of Europe, using Norwegian and Swedish healthy older adults.

A problem for countries with small populations (such as Sweden and Norway) is collecting sufficient sample sizes to produce continuous norms. Commonly, traditional norming approaches entail estimating score distribution for different demographic subgroups resulting in the need for substantial sample sizes. Regression-based norming requires samples that are 2.5 to 5.5 times smaller than traditional norming samples to obtain equally precise norms (Oosterhuis, van der Ark & Sijtsma, 2016). Therefore, we employed regression-based norming procedures in the development of our norms.

Our second aim was to compare our proposed norms with the North American norms from the professional manual (Meyers & Meyers, 1995a) (hereinafter referred to as "original norms") commonly employed in Scandinavia. We evaluated our normative data in an independent sample of cognitively healthy participants. As norms may differ due to ethnic and cultural factors, we expected our proposed norms to better fit the independent comparison group compared with the original norms.

MATERIALS AND METHODS

Gothenburg MCI study, DDI study, and SCAPIS cohorts

Normative samples. Healthy control persons from the Swedish Gothenburg MCI study (G-MCI) (n = 112), the Norwegian Dementia Disease Initiation study (DDI) (n = 27), and the Swedish Cardiopulmonary Bioimage Study (SCAPIS) (n = 205) were included (see Fig. 1). Exclusion criteria of the studies were severe somatic or psychiatric disorders that may influence cognitive functions. Additionally, subjects with subjective or objective cognitive decline were excluded.

The G-MCI (Wallin *et al.*, 2016) is a single-center clinical longitudinal study based at the memory clinic of Sahlgrenska University Hospital, Gothenburg, Sweden. Healthy control persons were primarily recruited from senior citizen organizations and secondarily from relatives to clinical participants between 2001 and 2014. In the G-MCI study, the inclusion criteria for healthy control persons were ages 50 to 79 and an Mini-Mental State Examination (MMSE) score greater than 26. Licensed psychologists, or psychologists-in-training under supervision of licensed psychologists, performed all neuropsychological examinations.

The DDI (Fladby *et al.*, 2017) is a multicenter longitudinal study, including clinical participants from several sites in Norway. Healthy control persons were mainly recruited among partners of clinical participants between the years 2013 and 2018. In addition, there was recruitment through advertisements in local media and orthopedic wards. For inclusion in the DDI study, the criteria for healthy control persons were ages 40 through 80 and a native speaker of Norwegian, Danish, or Swedish with an MMSE score greater than 27. The neuropsychological examinations were administered by licensed psychologists, licensed study nurses, or psychologists-in-training under supervision of licensed psychologists.

The SCAPIS (Bergström *et al.*, 2015) is a large Swedish population–based study investigating cardiopulmonary disease in participants between the ages of 50 and 64 years. A smaller pilot study (SCAPIS pilot) was initially conducted in Gothenburg, Sweden, in 2012, recruiting participants who also completed additional neuropsychological examinations. Similarly, licensed psychologists, licensed study nurses, or psychologists-in-training under supervision of licensed psychologists performed all neuropsychological examinations.

Independent validation sample. We evaluated our proposed norms against the original norms in an independent sample of subjects similar to

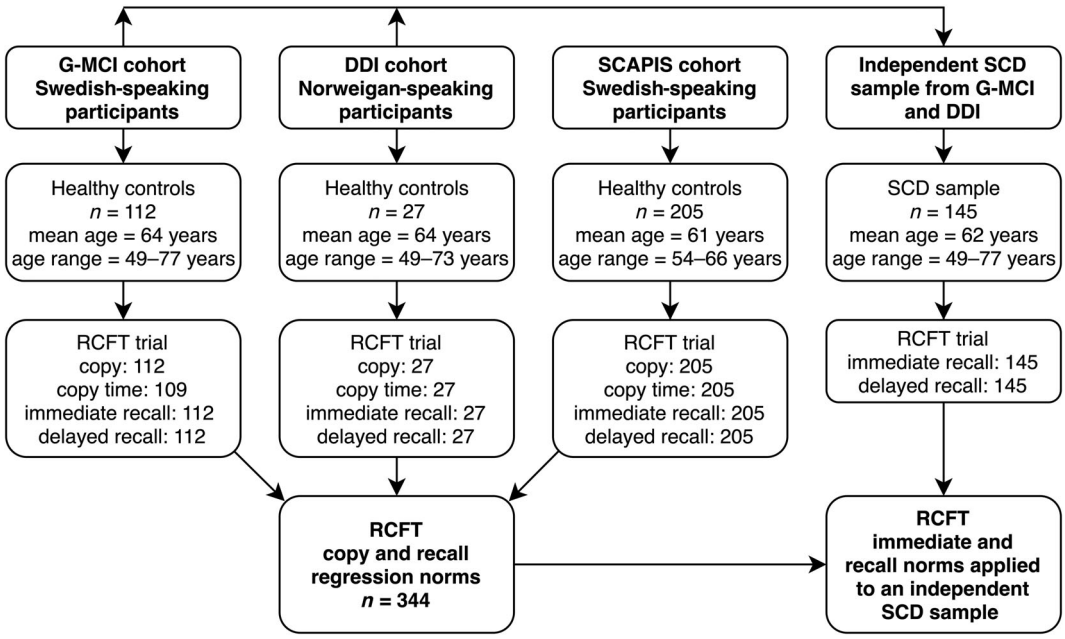

Fig. 1. Flowchart of the participant inclusion process from the G-MCI, DDI, and SCAPIS cohorts and workflow of the study. y = years; G-MCI = Gothenburg Mild Cognitive Impairment study; DDI = Dementia Disease Initiation; RCFT = Rey–Osterrieth Complex Figure Test; SCD = subjective cognitive decline

our healthy controls. This independent sample included subjects with subjective cognitive decline (SCD) (Jessen *et al.*, 2014) (N = 145) with performance within the normal range on neuropsychological tests. The SCD sample was sourced from both the G-MCI (N = 108) and the DDI (N = 37) (see Fig. 1). For SCD classification, all participants had subjectively reported cognitive decline but no objective cognitive impairment as determined by a battery of neuropsychological tests (Jessen *et al.*, 2014). The test battery comprised attention and executive function (Trail Making Test [TMT]-B) (Reitan & Wolfson, 1985) visuoperceptual ability (VOSP Silhouettes) (Warrington & James, 1991) phonemic verbal fluency (Controlled Oral Word Association Test [COWAT]/FAS) (Benton, Hamsher & Sivan, 1994), and verbal episodic memory (for the DDI, the Consortium to Establish a Registry for Alzheimer's Disease [CERAD] word list [Fillenbaum *et al.*, 2008] memory test was used; for the G-MCI, the Rey Auditory Verbal Learning Test [RAVLT] [Schmidt, 1996] was used). Demographically adjusted scores greater than −1.5 *SD* (>T35) on all of the aforementioned tests were deemed within the normal range (Eliassen *et al.*, 2020; Espenes *et al.*, 2020; Kirsebom *et al.*, 2019; Lorentzen *et al.*, 2021; Stricker *et al.*, 2020; Tallberg, Ivachova, Jones Tinghag & Östberg, 2008). Conversely, utilizing recommendations regarding the MCI (Albert *et al.*, 2011), we excluded subjects with a score of less than 1.5 *SD* (<T35) on the aforementioned cognitive tests as these would fulfill the criteria for MCI diagnosis.

RCFT administration and scoring

The original version of the RCFT (Meyers & Meyers, 1995a) was administrated to all subjects as part of a more extensive neuropsychological battery. The subjects were asked to copy the figure as accurately as possible. The subjects were not told that they would be asked to recall the figure from memory afterward. About 3 min following completion of the copy trial, subjects were asked to draw the figure from memory. In between, another brief neuropsychological test was administered. After 30 min, the participants were asked to reproduce the figure one last time. Finally, following the delayed recall, a recognition trial was administered. Also, the time to copy the figure was documented.

The Osterrieth scoring system was used, with a maximum possible score of 36 for the copy and recall trials. In this scoring system, the figure is separated into 18 different units. Two points were given if a unit was correctly reproduced. One point was given if the reproduction was distorted or incomplete but correctly placed, or was correctly reproduced but placed incorrectly. A score of 0.5 points was given if a unit was placed in an incorrect location and was distorted or incomplete. Unrecognizable units were scored zero.

Statistical analysis

Between-cohort comparisons. Raw scores and demographics for subjects in the G-MCI (n = 112), DDI (n = 27), and SCAPIS (n = 205) are compared in Table 1. To evaluate potential cohort differences between the normative samples, analysis of variance (ANOVA) models were fitted for age and years of education, whereas sex distributions were assessed using a chi-square test. Analysis of covariance (ANCOVA) models for each RCFT score (log-transformed copy trial scores, raw immediate and delayed recall trials) used age and education as covariates and cohort membership as the variable of interest (see Table 1). In addition, we applied independent sample t-tests for age and years of education to evaluate potential cohort differences within the independent SCD sample between G-MCI and DDI participants. All post hoc comparisons for ANOVA and ANCOVA were adjusted for demographics using the Tukey procedure. All analyses were conducted using R Statistical Software (version R 4.1.2; Core Development Team, 2020).

Regression-based norming procedure. We presented regression-based normative data using procedures similar to earlier work (Espenes *et al.*, 2020; Kirsebom *et al.*, 2019; Lorentzen *et al.*, 2021; Testa, Winicki, Pearlson, Gordon & Schretlen, 2009). First, using the Classical Test Theory Functions (CTT) R package (Willse, 2018), we normalized test scores by first computing the percentile ranks of the raw scores of the RCFT immediate and delayed recall and then used the normal cumulative quantile function (CDF) to produce standardized z-scores. These scores were then converted to scaled scores (S) with a mean of 10 and an *SD* of 3 ($S = z \times 3 + 10$). For RCFT copy time, the same procedure was applied but with the reverse percentile rank so that slower time to completion would produce lower scaled scores (see Table 2).

Subsequent, hierarchical regression models for the RCFT scaled scores were fitted by entering demographic variables (age, years of education, and sex) as well as squared terms to evaluate non-linear relationships

Table 1. *RCFT raw scores and demographics comparison between cohorts in the healthy controls sample*

	Comparison between normative subsamples				Comparison between normative total sample and independent SCD sample		
	G-MCI $N = 112$	DDI $N = 27$	SCAPIS $N = 205$	$\chi^2/F/H$, η^2/η_p^2, (P)	Total normative sample $N = 344$	SCD $N = 145$	$\chi^2/t/W/F$, d, (P)
Female *n* (%)	67 (59.8)	16 (59.3)	113 (55.1)	$\chi^2 = 0.71$, (0.699)	196 (57.0)	91 (63.0)	$\chi^2 = 1.18$, (0.277)
Age mean (*SD*) [range]	64.40 (6.19) [49–77]	63.60 (7.19) [49–73]	61.10 (3.30)[b] [54–66]	$F = 13.9$, (**<0.001**)	62.39 (5.01) [49–77]	62.25 (6.73) [49–77]	$t = -0.25$, (0.804)
Education mean (*SD*) [range]	12.40 (3.09)[c] [6–21]	14.90 (3.60) [8–22]	13.50 (2.79) [8–24]	$F = 19.4$, $\eta^2 = 0.05$ (**<0.001**)	13.27 (3.03) [6–24]	14.03 (3.22) [6–21]	$t = 2.47$, $d = -0.25$ (**0.014**)
RCFT copy median (IQR)	34 (4)	34 (3)	34 (3)	$H = 1.6$, (0.460)	34 (3)	34 (4)	$W = 22.912$, (0.150)
RCFT copy time[a] mean (*SD*)	229.06[d] (119.53)	195.26 (94.28)	181.23 (80.76)	$F = 7.21$, $\eta_p^2 = 0.04$ (**<0.001**)	197.67 (98.08)	216.40 (115.13)	$F = 3.31$, (0.070)
RCFT immediate recall[a] mean (*SD*)	16.74[d] (5.69)	20.0 (5.97)	19.30 (5.61)	$F = 9.10$, $\eta_p^2 = 0.05$ (**<0.001**)	18.53 (5.78)	18.49 (6.15)	$F = \sim 0$, (0.994)
RCFT delayed recall[a] mean (*SD*)	16.89 (5.49)	19.37 (5.81)	18.78 (5.60)	$F = 5.14$, $\eta_p^2 = 0.03$ (**0.006**)	18.23 (5.64)	18.52 (6.31)	$F = 0.29$, (0.589)

Note: Results are presented as mean, *SD*, and range, except sex, which is presented as male/female ratio. *n* = number of participants; *p* = *p*-value; χ^2 = chi-square; *H* = Kruskal–Wallis statistic; *F* = analysis of variance statistic; W = Wilcoxon signed rank test statistic; *t* = *t*-test statistic; *SD* = standard deviation; η^2 = eta-squared; η_p^2 = partial eta-squared; *d* = Cohen's *d*.

[a]ANCOVA models with age, years of education, and sex as covariates.

[b]Post hoc comparisons showed lower age in the SCAPIS as compared with the G-MCI and DDI (both $p < 0.001$).

[c]Post hoc comparison shows teh G-MCI with lower education as compared with the DDI and SCAPIS (both $p < 0.001$).

[d]Post hoc comparisons show the G-MCI with slightly worse performance in copy time and immediate recall as compared with the SCAPIS (both $p < 0.05$). No other significant between-group differences following post hoc comparisons in RCFT scores were found.

Table 2. *RCFT raw score to scaled score conversion table*

Scaled score	RCFT copy time (s)	RCFT immediate recall	RCFT delayed recall	Scaled score
2	≥596	≤ 3.5	≤ 4	2
3	524–595	4–4.5	4.5–5	3
4	450–523	5–6	5.5–7	4
5	372–449	6.5–9	7.5–9	5
6	300–371	9.5–10.5	9.5–10.5	6
7	250–299	11–13.5	11–13	7
8	219–249	14–15.5	13.5–15.5	8
9	187–218	16–17.5	16–17.5	9
10	161–186	18–20	18–19.5	10
11	141–160	20.5–21.5	20–21	11
12	125–140	22–23	21.5–23	12
13	114–124	23.5–24.5	23.5–24	13
14	100–113	25–26.5	24.5–26	14
15	89–99	27–28.5	26.5–27.5	15
16	75–88	29–29.5	28–29.5	16
17	60–74	30–31.5	30–30.5	17
18	≤59	≥32	≥31	18

Note: All scores are transformed to normal scaled scores using the cumulative percentile rank distribution of each score. Please note that for RCFT copy time, the rank distribution was reversed prior to conversion so that shorter copy time equals better scaled scores.

between test scores and continuous predictors (i.e., age^2 and years of education^2) and interaction terms (e.g., age × years of education). Age and education were mean centered in the equation when evaluating squared terms. We entered demographic variables into our analyses in accordance with their presumed importance. Most previous normative studies find prominent effects of age and education on test performance, while most do not find sex differences (Brauer Boone, Lesser, Hill-Gutierrez, Berman & D'Elia, 1993; Peña-Casanova *et al*., 2009; Tremblay *et al*., 2015). Therefore, age was inserted first, then age^2, years of education, years of education^2, and lastly, sex. Interactions between terms were only considered if more than one predictor was significantly associated with performance ($\alpha = 0.05$). Finally, we examined plots of predicted values and residuals (residuals vs. fitted plot, scale-location plot, and QQ plot) to ensure that the assumptions for regression analysis were met (e.g., linearity of the data, normality of residuals, homoscedasticity and independence of residuals error terms). We also assessed plots for outliers and influential cases (residuals vs. leverage plot and Cook's distance plot). The final regression models (see Table 3) comprised only statistically significant predictors and were subsequently used to calculate normative performance in our independent SCD group and compare with the original norms (see the section "Evaluation and comparison of our proposed norms with original norms in an independent sample").

Ethics

The procedures in the G-MCI, DDI, and SCAPIS studies were approved by the relevant ethical committees (local ethics committee in Sweden and regional committees for medical and health research ethics in Norway) and conducted in accord with the Helsinki declaration of 1964, revised in 2013. In the DDI, the Norwegian Health and Research Act was followed. All participants were informed about the details of their involvement in the study, including the right to withdraw and the potential risks and benefits involved (i.e., informed consent was received from all participants).

Table 3. *Normative regression model for RCFT measures from the G-MCI, DDI, and SCAPIS (N = 344)*

Variable	Predictor	*B*	Standard error *B*	*t*	*p*	R^2	Partial R^2	*SD* residual
RCFT copy time	Intercept	16.730	1.999	8.368	<0.001	0.030		2.9513
	Age	−0.108	0.032	3.365	<0.001		0.030	
RCFT immediate recall	Intercept	13.123	2.206	5.949	<0.001	0.036[a]		2.9427
	Age	−0.079	0.032	−2.474	0.014		0.018	
	Education	0.140	0.053	2.627	0.009		0.020	
RCFT delayed recall	Intercept	14.098	2.180	6.468	<0.001	0.050[a]		2.9082
	Age	−0.098	0.032	−3.100	0.002		0.028	
	Education	0.150	0.053	2.846	0.005		0.023	

Note: *b* = unstandardized regression coefficient; *t* = the *t*-test statistic; *SD* = standard deviation; *p* = *p*-value.
[a]Adjusted R^2 reported.

RESULTS

Comparison between normative subsamples

DDI and G-MCI participants were of similar age (mean difference [mdiff] = 0.85 years, *p* n.s.). However, the SCAPIS participants were younger than both the DDI (mdiff = 2.41 years, $p = 0.035$) and G-MCI (mdiff = 3.26 years, $p < 0.001$) samples. While the SCAPIS and DDI cohorts had a small, albeit non-significant difference in years of education (mdiff = 1.31 years, $p = 0.073$), the G-MCI sample had significantly shorter education as compared with both the DDI (mdiff = 2.45 years, $p < 0.001$) and SCAPIS (mdiff = 1.14 years, $p < 0.01$) samples. All samples had slightly more females than males (55–59% female), but distributions were similar between samples ($p = 0.699$). Some differences in test performance were noted between the samples. The G-MCI sample had slightly, albeit statistically significant poorer scores on the copy time (log median difference = −0.15, $p = 0.019$) and immediate recall measures (mdiff = 1.99, $p = 0.013$). For the delayed recall test, no significant between-samples differences were found following Tukey post hoc adjustment. See Table 1 for details.

Comparison between total normative sample and independent SCD sample

Apart from a small difference in educational attainment between the samples (the SCD sample had slightly more years of education [mdiff = 0.76 years, $p = 0.014$]), no other differences in demographics or test scores were observed. See Table 1 for details.

Demographic influence on test performance in the normative sample and development of regression-based norms

Older age was associated with slower RCFT copy time ($B = -0.108$, $R^2 = 0.030$, $p < 0.001$) and poorer RCFT immediate ($B = -0.079$, $R^2 = 0.018$, $p = 0.014$) and delayed recall ($B = -0.098$, $R^2 = 0.028$, $p = 0.002$). Conversely, more years of education were associated with better RCFT immediate ($B = 0.140$, $R^2 = 0.020$, $p = 0.009$) and delayed recall ($B = 0.150$, $R^2 = 0.023$, $p = 0.005$), but no effects of educational attainment on RCFT copy time performance were found. We found no sex differences, non-linear relationships, or interaction effects between age and education on performance. The final normative regression models are detailed in Table 3.

Demographic influence on RCFT copy trial and discrete norms

As the RCFT copy trial score distribution is positively skewed (ceiling effect) and thus not normally distributed, this subtest would not meet the requirements for linear regression modeling. To determine the demographic influence on test scores, non-parametric Spearman's rank correlations were performed between age and years of education, respectively. A small association was demonstrated between years of education and performance on the RCFT copy trial ($r = 0.124$, $p = 0.02$). The Wilcoxon signed-rank test was used to determine differences in scores between males and females. Following the results from these analyses, we computed discrete norms split by high (≥14 years) and low (≤13 years) levels of education using the cumulative frequency distribution and corresponding percentile ranks (see Table 4). The split was based on our normative sample's median (14 years of education).

Table 4. *Cumulative percentiles for the RCFT copy trial split by long and short education*

	≤13 years education *n* = 188	≥14 years education *n* = 156	Total *n* = 344
Years of education range (mean) [*SD*]	6–13 (11.0) [1.94]	14–24 (16.0) [1.65]	6–24 (13.3) [3.03]
Median (IQR)	33 (6)	33 (5)	33 (6)
Range	16–36	25–36	25–36
1%	≤24	≤26	≤25
2%	26	27	26
5%	27–28	28–29	27–28
9%	29	30	29
16%	30	31	31
25%	31–32	32	32
37%	33	33	33
50%	34	34	34
63%	34	35	35
75%	35	35	35
84%	35	36	36
91%	36	36	36

Note: *n* = number of participants; IQR = interquartile range.

Associations between the RCFT trials in the healthy control sample

We investigated the effect of copy time and copy score on RCFT test performance using regression analysis. The correlation between immediate and delayed recall was calculated using Pearson's correlation coefficient. Slower copy time was associated with poorer immediate recall ($B = -0.018$, $R^2 = 0.086$, $p < 0.001$) and delayed recall ($B = -0.016$, $R^2 = 0.077$, $p < 0.001$). Better copy score was associated with better immediate recall ($B = 0.784$, $R^2 = 0.138$, $p < 0.001$) and delayed recall ($B = 0.764$, $R^2 = 0.141$, $p < 0.001$). Immediate recall was strongly correlated with delayed recall ($r = 0.922$, $p < 0.001$).

Calculating normative performance using regression-based normative data

To calculate the normative effect of demographics (see Table 3) on the performance of a single individual, the following formula may be used: Intercept + (individual age × age coefficient) + (years of education × education coefficient). For example, for a 62-year-old man with 13 years of education, the regression equation for RCFT immediate recall would be (13.571) + (62 × –0.075) + (13 × 0.135) = 10.676. The participant's scaled score (Table 2) is subtracted from the regression equation predicted scaled score for each participant. The resulting discrepancy score is divided by the standard deviation of the healthy control group's regression residuals (Table 3) to yield a standardized *z*-score. In this case, the difference between the actual (10) and the predicted scaled score (10.676) is 0.676. Division by the standard deviation of the healthy control group residuals (2.74309) produces a *z*-score of −0.246. The following transformation will obtain *t*-scores: $t = z \times 10 + 50$, which is 47.5 ($t \sim 48$).

Evaluation and comparison of our proposed norms with original norms in an independent sample

First, regression normative data derived from the procedure above were used to estimate normative performance in the independent sample of $n = 145$ ($n = 37$, DDI; $n = 108$, G-MCI) cognitively healthy individuals experiencing SCD. Next, we calculated norms using the original norms from the professional manual (Meyers & Meyers, 1995a) in the independent sample. Lastly, multiple regression models were fitted to the *t*-scores from the respective norms to evaluate estimated means and demographic adjustment. The rationale behind this procedure is that a regression model yielding non-significant associations between demographic predictors and *t*-scores should indicate that the norms adjust adequately in the independent sample and vice versa (Table 5).

In our proposed norms, no significant associations were found between pertinent demographics and the *t*-scores for the RCFT subtests copy time, immediate recall, and delayed recall. Importantly, this indicates satisfactory adjustments of our proposed norms when applied to an independent sample of cognitively healthy adults. However, for the original norms, higher age was associated with higher *t*-scores for both immediate ($B = 0.333$, $R^2 = 0.029$, $p = 0.039$) and delayed recall ($B = 0.359$, $R^2 = 0.030$, $p = 0.037$), suggesting that the original age-adjusted norms presume a less detrimental effect of aging on normative performance than is evident in our sample. Moreover, higher education was associated with higher *t*-scores in the original norms for both immediate ($B = 0.770$, $R^2 = 0.036$, $p = 0.023$) and delayed recall ($B = 0.820$, $R^2 = 0.036$, $p = 0.023$). Note that these norms were stratified only for age, not education. The results are detailed in Table 5 and illustrated in Fig. 2.

As a last demonstration of the comparison between our proposed norms and the original norms, we conducted paired-samples *t*-tests to assess the mean *t*-scores in the independent sample involving both norms (see Fig. 3). For the immediate recall task, the independent sample exhibited a 4.796-point higher ($t = -13.265$, $p < 0.001$) estimated *t*-score in the original norms ($M = 54.566$, $SD = 13.210$) compared with our proposed norms ($SD = 49.836$, $SD = 10.346$). A similar result was found for delayed recall in which the original norms ($M = 54.731$, $SD = 14.025$) produced *t*-scores 4.397 higher ($t = -12.667$, $p < 0.001$) as compared with our proposed norms ($M = 50.334$, $SD = 11.383$).

DISCUSSION

In this study, we presented demographically adjusted norms for the RCFT at ages 49 to 77 years in a Swedish and Norwegian sample. Geographically local norms that evaluate performance

Table 5. *Multiple regression analysis on t-scores in the independent SCD sample using our proposed norms compared with the original norms*

		Proposed norms				Original norms			
Variable	Predictor	*B*	*p*	Adjusted R^2	Partial R^2	*B*	*p*	Adjusted R^2	Partial R^2
RCFT copy time	Intercept	43.73	<0.001	[a]		[a]	[a]	[a]	[a]
	Age	0.074	0.569	0	0	[a]	[a]		[a]
RCFT immediate recall	Intercept	47.76	<0.001	0	0	23.029	0.046	0.046	
	Age	0	0.992	0		0.333	0.039		0.029
	Years of education	0.142	0.599	0		0.770	0.023		0.036
RCFT delayed recall	Intercept	42.31	<0.001	0	0	20.896	0.087	0.047	
	Age	0.093	0.512	0		0.359	0.037		0.030
	Years of education	0.158	0.596	0		0.820	0.023		0.036

Note: *b* = unstandardized regression coefficient; *p* = *p*-value; partial R^2 = explained variance of predictor variable; Adj. R^2 = explained variance of combined predictor variables.
[a]No additional data needed.

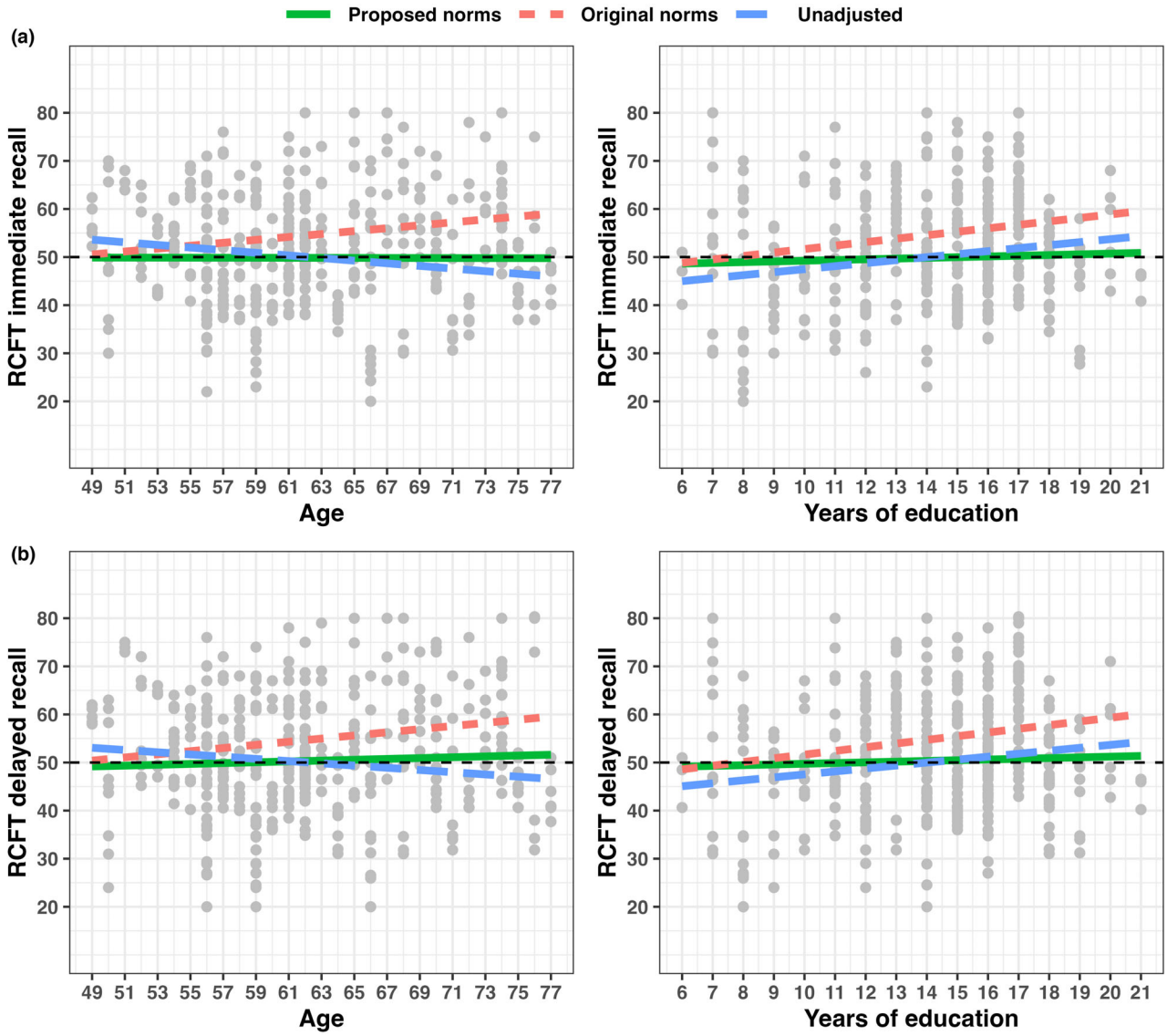

Fig. 2. Normative adjustment of our proposed norms and the original norms in the independent SCD sample. Panel A depicts relationships between RCFT immediate recall *t*-scores and age and education, respectively. Panel B similarly shows relationships for RCFT delayed memory. For all figures, the horizontal black dashed line at T50 is the ideal reference for normative adjustments at all age and education levels. Departures from the line suggest worse normative performance. The solid green line shows *t*-scores for our proposed norms, whereas the red short-dashed lines show original norm *t*-scores and the blue long-dashed lines are unadjusted *t*-scores added for visual comparison (data not shown).

based on relevant demographic variables are important for neuropsychological tests in clinical settings (International Test Commission, 2001). As expected, higher age was significantly associated with poorer RCFT performance. This pattern is supported by the fact that processing speed (Murman, 2015) and visuospatial memory (Klencklen, Després & Dufour, 2012) may drop during the life span. Surprisingly, we found that the magnitude of the effect differed compared with the original norms. The original norms demonstrated a much higher proportion of variance explained by age (recall = 19%), while the proportion of variance explained by education was insignificant. As for our normative data, sex was unrelated to RCFT scores for original norms. In our normative data, age explained only 3% of the variance for copy time and 2% to 3% of the variance for immediate and delayed recall. Education also slightly impacted the recall trials (but explaining only 2% of the variance).

The lower effect of age in our proposed norms may indicate that our sample was cognitively healthier. Previous reports have argued that the impact of age on cognitive test results, to a large degree, is driven by individuals with preclinical neurodegenerative disease (Borland, Stomrud, Van Westen, Hansson & Palmqvist, 2020; Harrington *et al.*, 2018). The weaker effect of age demonstrated in our normative data could also result from a narrower age range, as the original norms were based on individuals from 18 to 89 years. The Swedish and Norwegian participants were cognitively assessed during the 2000s, in contrast to the North American participants, who were tested at the earliest during the early 1990s. There might be cultural or generational differences between cohorts, such as differences in the educational system, education availability, or cerebrovascular disease incidence.

We compared the mean performance in the normative healthy control group with a comparable, independent sample of SCD

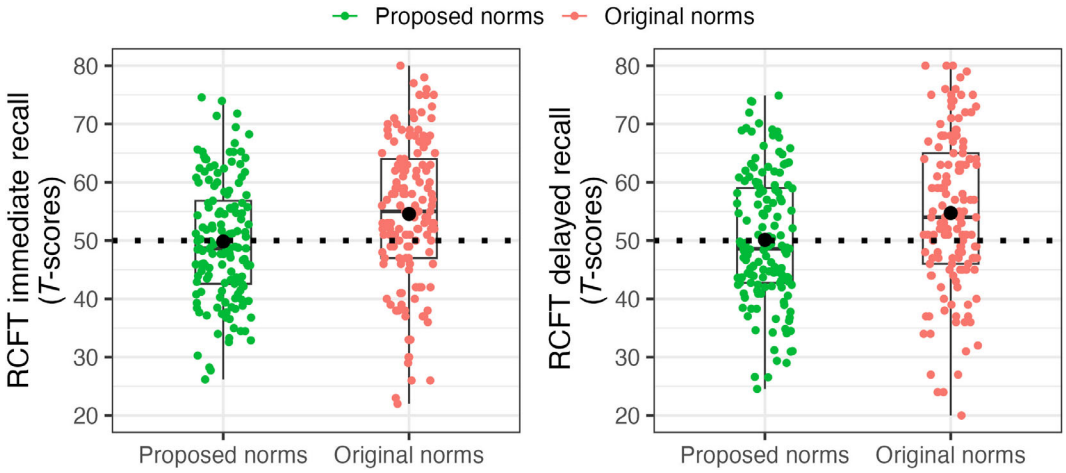

Fig. 3. The *t*-score distribution of our proposed norms and the original norms in the independent SCD sample. The left figure depicts the distribution of *t*-scores for RCFT immediate recall between our proposed norms (green dots) and the original norms (red dots), respectively. The right figure depicts the distribution of *t*-scores for RCFT delayed recall between our proposed norms (green dots) and the original norms (red dots), respectively. The black dots show the mean *t*-score.

participants. The absence of a significant difference in mean scores suggested that the SCD sample is comparable to the healthy control group and can be used to evaluate our normative data. As illustrated in Figs. 2 and 3, the original norms overestimated immediate and delayed recall performance for all significant results, indicating that our normative data better fit the independent sample. Failure to adjust for age and education is most evident at the higher end of the predictors, the oldest and those with the highest education. For example, the calculated *t*-score for a 57-year-old participant with 18 years of education scoring 15.50 on the RCFT immediate recall is T39 applying our normative data. By contrast, applying the original norms, the same individual would receive a *t*-score of T45. If a normative dataset incorrectly classifies a patient's scores as normal due to low expectations of higher scores, this may lead to incorrect diagnostic conclusions, for example the correct diagnosis of MCI (Albert *et al.*, 2011).

On average, longer copy time was associated with worse immediate and delayed recall performance (copy time explained 8–9% of the variance). An explanation for this is that individuals in need of longer copy time may have less effective organization strategies, resulting in less successful memory encoding and, thus, lower recall scores (Newman & Krikorian, 2001). Other studies (Tremblay *et al.*, 2015) have seen a reverse pattern for delayed recall, suggesting that prolonged exposure to the visual stimulus could be linked to better delayed recall. Studies manipulating allowed copy time are needed to illuminate the relationship between copy time and succeeding performance on the RCFT. However, our results indicate that individuals who spontaneously take longer to copy subsequently perform worse on both immediate and delayed recall. As expected, a better copy score was associated with better immediate and delayed recall performance (copy score explained 14% of the variance).

Similarly, as found in previous studies (Meyers & Meyers, 1995b; Tremblay *et al.*, 2015), the immediate and delayed recall were nearly perfectly positively correlated with each other ($r = 0.922$). This could be interpreted as delayed recall not being needed to get a valid visuospatial memory measurement. However, even though this may be true in a cognitively healthy population, this may not be true for individuals with memory impairment, such as in clinical groups (Strauss *et al.,* 2006). Therefore, caution must be exercised when assessing clinical groups and administering delayed recall if memory problems are suspected.

LIMITATIONS

Some study limitations are apparent. First, the size of the Norwegian sample is small ($N = 27$). Second, our sample does not include younger (under 49 years of age) or older subjects (over 77 years of age), meaning that our proposed norms are not representative of younger individuals or the oldest old. Third, our sample did not include bilingual participants who speak other languages, which might decrease the representativeness for individuals who have moved to Scandinavia during their lives for various reasons. Fourth, we used a cognitively healthy independent SCD sample to evaluate our normative data. Even though SCD is a common phenomenon in healthy aging and is generally considered a benign condition (Hessen *et al.*, 2017), it is still an established risk factor for neurodegenerative disease (Jessen *et al.*, 2014). The risk of progression from SCD to MCI is higher in groups with amyloid plaque deposition (Vogel, Stokholm & Jørgensen, 2012). As we lack imaging or fluid-based biomarkers for all our participants, we cannot consider potential underlying pathophysiological processes, such as preclinical AD. Fifth, even though we utilized standardized test administration and scoring, we have not been able to perform any inter-rater reliability measurements. As a result, some variability may exist between the examiners. However, previous studies applying the same standardized scoring methods have shown good inter-rater reliability (Berry, Allen & Schmitt, 1991; Caffarra, Vezzadini, Dieci, Zonato & Venneri, 2002; Meyers & Meyers, 1995a). Lastly, while our three cohorts (G-MCI, DDI, and SCAPIS) were similar in the effects of demographics on RCFT performance, there were still some small but significant differences. G-MCI performed worse on copy time (log mdiff = −0.15, $p = 0.019$)

and immediate recall (mdiff = 1.99, $p = 0.013$), even though adjusted for age and education. Nevertheless, our proposed norms correctly adjusted for demographics in the independent sample, thus indicating a better fit and better suitability in Swedish and Norwegian contexts.

CONCLUSION

In conclusion, this study has presented demographically adjusted norms for RCFT – one of Scandinavia's most commonly utilized neuropsychological tests. Further, we compared our proposed norms and original norms originating from North America. We demonstrated that our proposed norms might produce more accurate normative scores when employed in an independent comparison group. However, future studies need to assess our proposed norms in clinical settings and evaluate their usefulness for diagnosis and treatment.

ACKNOWLEDGMENTS

We thank each study's clinical teams (i.e., the memory clinics and the SCAPIS study in Gothenburg and various clinics in Norway) for their work with clinical examinations and support.

FUNDING INFORMATION

This work was supported by the Sahlgrenska University Hospital, the Swedish Research Council, Swedish Brain Power, the Swedish Dementia Foundation, the Swedish Alzheimer Foundation, Stiftelsen Psykiatriska forskningsfonden, Konung Gustaf V:s and Drottning Victorias Frimurarestiftelse, and Anna-Lisa och Bror Björnssons Stiftelse. Additional support was received from the University of Tromsø – the Arctic University of Norway; the Norwegian Research Council (Dementia Disease Initiation) under grant number 217780; Helse Sør-øst, NASATS (Dementia Disease Initiation) under grant number 2013131; and Helse Nord under grant number HNF1401-18. The funding sources were not involved in the drafting of this manuscript.

CONFLICT OF INTEREST

The authors declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are not publicly available due to the privacy of the research participants. Requests to access these datasets should be directed to Anders Wallin (anders.wallin@neuro.gu.se).

REFERENCES

Albert, M.S., DeKosky, S.T., Dickson, D., Dubois, B., Feldman, H.H., Fox, N.C. *et al.* (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: The Journal of the Alzheimer's Association*, *7*(3), 270–279. https://doi.org/10.1016/j.jalz.2011.03.008.

Ashton, V.L., Donders, J. & Hoffman, N.M. (2005). Rey Complex Figure Test performance after traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology*, *27*(1), 55–64. https://doi.org/10.1080/138033990513636.

Benton, A.L., Hamsher, K.D. & Sivan, A.B. (1994). Multilingual aphasia examination: Manual of instructions https://www.parinc.com/Products/Pkey/222%0A, http://www.springerreference.com/index/doi/10.1007/SpringerReference_183669.

Berg, O. & Billman, J. (2009). *Svenska neuropsykologers utredningsarbete: En kartläggning av praktik och instrumentanvändning*. Umeå: Umeå University.

Bergström, G., Berglund, G., Blomberg, A., Brandberg, J., Engström, G., Engvall, J. *et al.* (2015). The Swedish CArdioPulmonary BioImage Study: Objectives and design. *Journal of Internal Medicine*, *278*(6), 645–659. https://doi.org/10.1111/joim.12384.

Berry, D.T.R., Allen, R.S. & Schmitt, F.A. (1991). Rey-Osterrieth Complex Figure: Psychometric characteristics in a geriatric sample. *The Clinical Neuropsychologist*, *5*(2), 143–153. https://doi.org/10.1080/13854049108403298.

Borland, E., Stomrud, E., Van Westen, D., Hansson, O. & Palmqvist, S. (2020). The age-related effect on cognitive performance in cognitively healthy elderly is mainly caused by underlying AD pathology or cerebrovascular lesions: Implications for cutoffs regarding cognitive impairment. *Alzheimer's Research & Therapy*, *12*(1), 1–10. https://doi.org/10.1186/s13195-020-00592-8.

Brauer Boone, K., Lesser, I.M., Hill-Gutierrez, E., Berman, N.G. & D'Elia, L.F. (1993). Rey-Osterrieth complex figure performance in healthy, older adults: Relationship to age, education, sex, and IQ. *The Clinical Neuropsychologist*, *7*(1), 22–28. https://doi.org/10.1080/13854049308401884.

Caffarra, P., Vezzadini, G., Dieci, F., Zonato, F. & Venneri, A. (2002). Rey-Osterrieth complex figure: Normative values in an Italian population sample. *Neurological Sciences*, *22*(6), 443–447. https://doi.org/10.1007/s100720200003.

Calev, A., Edelist, S., Kugelmass, S. & Lerer, B. (1991). Performance of long-stay schizophrenics on matched verbal and visuospatial recall tasks. *Psychological Medicine*, *21*(3), 655–660. https://doi.org/10.1017/S0033291700022297.

Cherrier, M.M., Mendez, M.F., Dave, M. & Perryman, K.M. (1999). Performance on the Rey-Osterrieth Complex Figure Test in Alzheimer disease and vascular dementia. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*, *12*(2), 95–101.

Chervinsky, A.B., Mitrushina, M. & Satz, P. (1992). Comparison of four methods of scoring the Rey-Osterrieth Complex Figure Drawing Test on four age groups of normal elderly. *Brain Dysfunction*, *5*(5–6), 267–287.

Darwish, H., Zeinoun, P., Farran, N. & Fares, S. (2018). Rey Figure Test with recognition trial: Normative data for Lebanese adults. *The Clinical Neuropsychologist*, *32*(sup1), 102–113. https://doi.org/10.1080/13854046.2018.1480802.

Drozdová, K., Štěpánková, H., Lukavský, J., Bezdíček, O. & Kopeček, M. (2015). Normativní studie testu Reyovy-Osterriethovy komplexní fi gury v populaci českých seniorů. *Česká a slovenská neurologie a neurochirurgie*, *111*(5), 542–549.

Eliassen, I.V., Fladby, T., Kirsebom, B.E., Waterloo, K., Eckerström, M., Wallin, A. *et al.* (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. *et al.* (2020). Demographically adjusted trail making test norms in a Scandinavian sample from 41 to 84 years. *The Clinical Neuropsychologist*, *34*(S1), 110–126. https://doi.org/10.1080/13854046.2020.1829068.

Fastenau, P.S., Denburg, N.L. & Hufford, B.J. (1999). Adult norms for the rey-osterrieth complex figure test and for supplemental recognition and

matching trials from the extended complex figure test. *The Clinical Neuropsychologist*, *13*(1), 30–47. https://doi.org/10.1076/clin.13.1.30.1976.

Fasteneau, P.S. (2003). *Extended complex figure test manual*. Los Angeles: Western Psychological Services.

Fillenbaum, G.G., van Belle, G., Morris, J.C., Mohs, R.C., Mirra, S.S., Davis, P.C. *et al.* (2008). Consortium to Establish a Registry for Alzheimer's Disease (CERAD): The first twenty years. *Alzheimer's & Dementia*, *4*(2), 96–109.

Fladby, T., Palhaugen, L., Selnes, P., Waterloo, K., Brathen, G., Hessen, E. *et al.* (2017). Detecting at-risk Alzheimer's disease cases. *Journal of Alzheimer's Disease*, *60*(1), 97–105. https://doi.org/10.3233/JAD-170231.

Gallagher, C. & Burke, T. (2007). Age, gender and IQ effects on the Rey-Osterrieth complex figure test. *British Journal of Clinical Psychology*, *46*(1), 35–45. https://doi.org/10.1348/014466506X106047.

Harrington, K.D., Dang, C., Lim, Y.Y., Ames, D., Laws, S.M., Pietrzak, R.H. *et al.* (2018). The effect of preclinical Alzheimer's disease on age-related changes in intelligence in cognitively normal older adults. *Intelligence*, *70*, 22–29. https://doi.org/10.1016/j.intell.2018.07.004.

Hessen, E., Eckerström, M., Nordlund, A., Selseth Almdahl, I., Stålhammar, J., Bjerke, M. *et al.* (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. https://doi.org/10.1159/000454676.

International Test Commission. (2001). International guidelines for test use. *International Journal of Testing*, *1*(2), 93–114. https://doi.org/10.1207/s15327574ijt0102_1.

Jessen, F., Amariglio, R.E., Van Boxtel, M., Breteler, M., Ceccaldi, M., Chételat, G. *et al.* (2014). A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *10*(6), 844–852. https://doi.org/10.1016/j.jalz.2014.01.001.

Kasai, M., Meguro, K., Hashimoto, R., Ishizaki, J., Yamadori, A. & Mori, E. (2006). Non-verbal learning is impaired in very mild Alzheimer's disease (CDR 0.5): Normative data from the learning version of the Rey-Osterrieth Complex Figure Test. *Psychiatry and Clinical Neurosciences*, *60*(2), 139–146. https://doi.org/10.1111/j.1440-1819.2006.01478.x.

Kirsebom, B.E., Espenes, R., Hessen, E., Waterloo, K., Johnsen, S.H., Gundersen, E. *et al.* (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.

Klencklen, G., Després, O. & Dufour, A. (2012). What do we know about aging and spatial cognition? Reviews and perspectives. *Ageing Research Reviews*, *11*(1), 123–135. https://doi.org/10.1016/j.arr.2011.10.001.

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

Lorentzen, I.M., Espenes, J., Hessen, E., Waterloo, K., Bråthen, G., Timón, S. *et al.* (2023). Regression-based norms for the FAS phonemic fluency test for ages 40–84 based on a Norwegian sample. *Applied Neuropsychology. Adult*, *32*, 159–168. https://doi.org/10.1080/23279095.2021.1918128.

Machulda, M.M., Ivnik, R.J., Smith, G.E., Ferman, T.J., Boeve, B.F., Knopman, D. *et al.* (2007). Mayo's older Americans normative studies: Visual form discrimination and copy trial of the Rey-Osterrieth complex figure. *Journal of Clinical and Experimental Neuropsychology*, *29*(4), 377–384. https://doi.org/10.1080/13803390600726803.

Meyers, J.E. & Meyers, K.R. (1995a). *Rey complex figure and recognition trial: Professional manual*. Odessa: Psychological Assessment Resources.

Meyers, J.E. & Meyers, K.R. (1995b). Rey complex figure test under four different administration procedures. *The Clinical Neuropsychologist*, *9*(1), 63–67. https://doi.org/10.1080/13854049508402059.

Murman, D.L. (2015). The impact of age on cognition. *Seminars in Hearing*, *36*(3), 111–121. https://doi.org/10.1055/s-0035-1555115.

Newman, P.D. & Krikorian, R. (2001). Encoding and complex figure recall. *Journal of the International Neuropsychological Society*, *7*(6), 728–733. https://doi.org/10.1017/S1355617701766088.

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.

Osterrieth, P.A. (1944). Le test de copie d'une figure complexe; contribution à l'étude de la perception et de la mémoire. *Archives de Psychologie*, *30*, 206–356.

Peña-Casanova, J., Gramunt-Fombuena, N., Quiñones-Úbeda, S., Sánchez-Benavides, G., Aguilar, M., Badenes, D. *et al.* (2009). Spanish multicenter normative studies (NEURONORMA project): Norms for the rey-osterrieth complex figure (copy and memory), and free and cued selective reminding test. *Archives of Clinical Neuropsychology*, *24*(4), 371–393. https://doi.org/10.1093/arclin/acp041.

R Core Development Team. (2020). *A language and environment for statistical computing*, Vol. *2*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org.

Rabin, L.A., Barr, W.B. & Burton, L.A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA division 40 members. *Archives of Clinical Neuropsychology*, *20*(1), 33–65. https://doi.org/10.1016/j.acn.2004.02.005.

Reitan, R.M. & Wolfson, D. (1985). *The Halstead-Reitan neuropsychological test battery*. Tucson: Neuropsychology Press.

Rey, A. (1941). L'examen psychologique dans les cas d'encéphalopathie traumatique. *Archives de Psychologie*, *28*, 215–285.

Rivera, D., Perrin, P.B., Morlett-Paredes, A., Galarza-Del-Angel, J., Martínez, C., Garza, M.T. *et al.* (2015). Rey-Osterrieth complex figure-copy and immediate recall: Normative data for the Latin American Spanish speaking adult population. *NeuroRehabilitation*, *37*(4), 677–698. https://doi.org/10.3233/NRE-151285.

Ryder, T.M. (2021). Testkvalitetsprosjektet – del 1: Tester med behov for kvalitetstiltak. *Tidsskrift for Norsk Psykologforening*, *58*(1), 28–37.

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

Strauss, E., Sherman, E.M.S., &Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.,). New York: Oxford University Press.

Stricker, N.H., Christianson, T.J., Lundt, E.S., Alden, E.C., Machulda, M.M., Fields, J.A. *et al.* (2020). 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*, 211–226. https://doi.org/10.1017/s1355617720000752.

Tallberg, I.M., Ivachova, E., Jones Tinghag, K. & Östberg, P. (2008). Swedish norms for word fluency tests: FAS, animals and verbs. *Scandinavian Journal of Psychology*, *49*(5), 479–485. https://doi.org/10.1111/j.1467-9450.2008.00653.x.

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*(6), 1012–1022. https://doi.org/10.1017/S1355617709990713.

Tremblay, M.P., Potvin, O., Callahan, B.L., Belleville, S., Gagnon, J.F., Caza, N. *et al.* (2015). Normative data for the Rey-Osterrieth and the Taylor complex figure tests in Quebec-French people. *Archives of Clinical Neuropsychology*, *30*(1), 78–87. https://doi.org/10.1093/arclin/acu069.

Tsatali, M., Emmanouel, A., Gialaouzidis, M., Avdikou, K., Stefanatos, C., Diamantidou, A. *et al.* (2022). Rey complex figure test (RCFT): Norms for the Greek older adult population. *Applied Neuropsychology: Adult*, *29*(5), 958–966. https://doi.org/10.1080/23279095.2020.1829624.

Vogel, A., Stokholm, J. & Jørgensen, K. (2012). Performances on Rey auditory verbal learning test and Rey complex figure test in a healthy, elderly Danish sample: Reference data and validity issues.

Scandinavian Journal of Psychology, *53*(1), 26–31. https://doi.org/10.1111/j.1467-9450.2011.00909.x.

Wallin, A., Nordlund, A., Jonsson, M., Lind, K., Edman, Å., Göthlin, M. *et al*. (2016). The Gothenburg MCI study: Design and distribution of Alzheimer's disease and subcortical vascular disease diagnoses from baseline to 6-year follow-up. *Journal of Cerebral Blood Flow and Metabolism*, *36*(1), 114–131. https://doi.org/10.1038/jcbfm.2015.147.

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

Willse, J.T. (2018). Rey complex figure test (RCFT): Norms for the Greek older adult population https://CRAN.R-project.org/package=CTT.

Received 3 May 2023, Revised 24 July 2023, accepted 27 August 2023