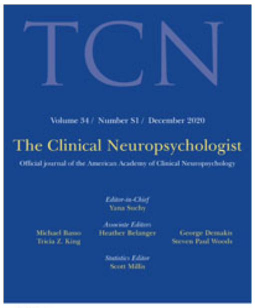

The Clinical Neuropsychologist

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

Demographically adjusted trail making test norms in a Scandinavian sample from 41 to 84 years

Jacob Espenes , Erik Hessen , Ingvild Vøllo Eliassen , Knut Waterloo , Marie Eckerström , Sigrid Botne Sando , Santiago Timón , Anders Wallin , Tormod Fladby & Bjørn-Eivind Kirsebom

To cite this article: Jacob Espenes , Erik Hessen , Ingvild Vøllo Eliassen , Knut Waterloo , Marie Eckerström , Sigrid Botne Sando , Santiago Timón , Anders Wallin , Tormod Fladby & Bjørn-Eivind Kirsebom (2020) Demographically adjusted trail making test norms in a Scandinavian sample from 41 to 84 years, The Clinical Neuropsychologist, 34:sup1, 110-126, DOI: 10.1080/13854046.2020.1829068

To link to this article: https://doi.org/10.1080/13854046.2020.1829068

Published online: 09 Oct 2020.

THE CLINICAL NEUROPSYCHOLOGIST
2020, VOL. 34, NO. S1, 110–126
https://doi.org/10.1080/13854046.2020.1829068

Demographically adjusted trail making test norms in a Scandinavian sample from 41 to 84 years

Jacob Espenes[a,b], Erik Hessen[c,d], Ingvild Vøllo Eliassen[c,d], Knut Waterloo[a,b], Marie Eckerström[e], Sigrid Botne Sando[f,g], Santiago Timón[c,h], Anders Wallin[e], Tormod Fladby[c,i] 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]Department of Neuromedicine and Movement Science, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway; [g]Department of Neurology and Clinical Neurophysiology, University Hospital of Trondheim, Trondheim, Norway; [h]Departamento de Inteligencia Artificial, Universidad Nacional de Educación a Distancia, Madrid, Spain; [i]Institute of Clinical Medicine, Campus Ahus, University of Oslo, Oslo, Norway

ABSTRACT

Objective: The trail making test (TMT) is one of the most widely used neuropsychological tests. TMT-A provides measures of visual scanning/visuomotor speed and TMT-B involves additional demands on executive functions. Derived scores TMT B-A and TMT B/A enhance measures of executive functioning. However, simple B-A subtraction may lead to false estimates of executive dysfunction in clinical samples. Norms for TMT have been published in several countries but are currently lacking for Scandinavia.

Methods: A total of 292 healthy controls between age 41 and 84 years were included from the Norwegian "Dementia Disease Initiation" (DDI) study ($n = 170$) and the Gothenburg Mild Cognitive Impairment (MCI) study ($n = 122$). We used a regression-based procedure to develop demographically adjusted norms for basic (TMT-A and TMT-B) and derived measures (TMT B-A and B/A). We also propose a regression-based alternative to the TMT B-A measure named "TMT-β". The proposed norms were compared to norms from Heaton et al. and Tombaugh.

Results: Due to differences in the estimated normative effects of demographics on performance, the proposed norms for TMT were better suited in the Scandinavian sample compared with published non-Scandinavian norms. The proposed TMT-β measure was highly correlated to TMT B-A ($r = 0.969$, $p < 0.001$).

Conclusion: We here propose demographically adjusted norms for the TMT for ages 41 through 84 years based on a Scandinavian

ARTICLE HISTORY
Received 15 May 2020
Accepted 19 September 2020
Published online 9 October 2020

KEYWORDS
Neuropsychological tests; norms; trail making test; TMT; Scandinavia; Norway; Sweden; cross-cultural neuropsychology

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 9019, Norway.

sample. We also present the regression-based derived measure TMT-β which may resolve issues with the conventional TMT B-A measure.

Introduction

Performance on the trail making test (TMT) is mediated through a set of global neural mechanisms (Moll et al., 2002) and TMT is sensitive to a variety of conditions with neurological deficits (Gonçalves et al., 2013). TMT is therefore suitable as a screening tool for neurological integrity and identification of individuals in need of cognitive assessment (Reitan & Wolfson, 2004). Basic task demands of TMT-A and TMT-B are visual search and/or visuomotor speed. TMT-B is a more difficult task, involving additional demands on executive functions including working memory and cognitive flexibility due to the alternation between numbers and letters (Sanchez-Cubillo et al., 2009). Derived measures of TMT have been suggested to highlight measurements of executive functions associated with TMT-B, primarily difference score TMT B-A (Lezak et al., 2012, p. 423) and ratio score TMT B/A (Arbuthnott & Frank, 2000; Lamberty et al., 1994).

Clinicians rely on published norms which aim to correct for demographics known to influence test performance. On TMT, increasing age is associated with decreased performance (Gonçalves et al., 2013; Goul & Brown, 1970; Kennedy, 1981; Stuss et al., 1988) and higher educational attainment relates to increases in performance especially on TMT-B (Heaton et al., 2004; Peña-Casanova et al., 2009; Periáñez et al., 2007; Tombaugh, 2004). Derived measures TMT B-A and TMT B/A are less affected by variations in age and education compared with the basic measures (Gonçalves et al., 2013). Most studies do not find sex differences on TMT (Mitrushina et al., 2005, p. 69). Normative studies investigating the effects of age and educational attainment on TMT scores show varying results due to differences in sample characteristics (e.g. range of educational attainment and age in the sample) and may limit the applicability of norms across different populations. Indeed, TMT norms have been shown to produce markedly diverging estimates when applied to different populations, ranging from 0.8 to 1.4 standard deviations (Fernandez & Marcopulos, 2008). In addition, cohort effects have been found on TMT, likely due to advancements in educational quality and health (i.e. a Flynn Effect; Dickinson & Hiscock, 2011; Dodge et al., 2014). To resolve these issues, local norms have been developed for the TMT in several countries (Abi Chahine et al., 2019; Cavaco et al., 2013; Gonçalves et al., 2013; Siciliano et al., 2019; St-Hilaire et al., 2018).

To our knowledge, test norms for TMT based on a Scandinavian sample have not been published. Thus, the first objective of this study was to investigate the influence of age, education, and sex on TMT scores in a sample of healthy Norwegians and Swedes between 41 and 84 years of age ($n = 292$) and develop norms for the basic and derived measures of TMT using a regression-based norming procedure. Second, we compare the current proposed norms with two sets of norms (Heaton et al. 2004; Tombaugh, 2004) frequently applied by clinicians and researchers and recommended by Norwegian health authorities in clinical use (Strobel et al., 2018). Third, we propose an alternate method for computing the conventional TMT B-A measure which might have applications in clinical populations.

A disadvantage of the simple subtraction method TMT B-A is that an elevated difference score is interpreted as difficulties with the additional task demands of TMT-B, indicating deficits in executive functions (cognitive flexibility and working memory). However, TMT-B is also more demanding than TMT-A on visual search and/or visuomotor abilities due to increased amount of connections to be drawn, and the distance between connections (Gaudino et al., 1995). Patients with general visuomotor and visual scanning deficits resulting in reduced performance on both TMT-A and TMT-B may therefore show a disproportionate increase in time to completion on TMT-B (Senior et al., 2018). Thus, a high TMT B-A difference could also be due to general visuomotor or visual scanning deficits rather than executive deficits. As shown by Senior et al. (2018), normative values on TMT B-A are based on mean values from the entire sample and do not accommodate this non-linear relationship by accounting for individual variability on TMT-A. We therefore propose an alternative method for the derived measure TMT B-A by regressing age and education along with scores from TMT-A on scores from TMT-B using multiple regression analysis. This approach resolves the issues with conventional B-A subtraction while simultaneously controlling for pertinent demographics. We have named this new measure "TMT-β" to avoid confusion with the conventional TMT B-A approach.

Methods and materials

Participants

We included healthy controls from the Norwegian Dementia Disease Initiation Study (DDI; $n = 170$) and the Swedish Gothenburg mild cognitive impairment (MCI) study ($n = 122$). DDI is a national multicenter longitudinal study aimed at early detection and diagnosis of common neurodegenerative diseases such as Alzheimer's disease (AD). Participants from DDI were recruited between January 2013 and October 2018. The Gothenburg MCI study started in 1999 and is an ongoing single-center study on early phases of AD and vascular dementia based in Sahlgrenska University Hospital in Sweden. Participants were recruited between January 2001 and March 2014.

Criteria for inclusion of healthy controls from the DDI study were ages 40 through 80, absence of subjective symptoms of cognitive decline and MMSE score >26 and a native language of Norwegian, Danish, or Swedish. Participants in the DDI cohort were recruited from all Norwegian health regions. Healthy controls were primarily recruited from spouses of symptom group participants and secondarily by self-referral through advertisements in local media and from orthopedic wards. All participants from the DDI study followed a standardized procedure for assessment following a Case Report Form (CRF) developed for DDI and is described in detail in Fladby et al. (2017). Briefly, this included standardized neurological and physical examinations by neurologist, brief neuropsychological assessment, and standardized interview involved taking a medical history from participants and informants. Licensed psychologists, neurologists, licensed study nurses, or psychologists-in-training under supervision from licensed psychologists performed cognitive assessments. Patients with history of stroke, severe psychiatric disorder including major depression, intellectual disability or developmental disorders, and severe somatic disorders that may influence cognitive functions were excluded.

Table 1. Demographics, raw scores, and *T*-scores of the healthy controls from the dementia disease initiation (DDI) and Gothenburg mild cognitive impairment (MCI) study ($n = 292$).

Variables	Test scores/demographics: DDI controls $n = 170$	Test scores/demographics: Gothenburg MCI $n = 122$	t/x^2	p
Age *M* (*SD*) [range]	62.0 (9.4) [41 – 84]	64.3 (6.5) [49 – 77]	$t = -2.39$	<0.05
Female *n* (%)	100 (58.8%)	74 (60.7%)	x^2 0.10	*ns*
Years of education *M* (*SD*) [range]	13.8 (3.3) [7 – 23]	12.4 (3.2) [6 – 24]	$t = 3.83$	$<.001$
TMT-A s *M* (*SD*)	35.0 (11.6)	34.8 (10.4)	$t = 0.19$	*ns*
TMT-B s *M* (*SD*)	82.6 (28.4)	82.2 (23.4)	$t = 0.13$	*ns*
TMT B-A raw score *M* (*SD*)	47.6 (25.6)	47.4 (18.5)	$t = 0.06$	*ns*
TMT B/A raw score *M* (*SD*)	2.5 (0.9)	2.4 (0.6)	$t = 0.53$	*ns*
TMT-A *T*-scores *M* (*SD*)	49.61 (10.2)	50.4 (9.7)	$t = -0.69$	*ns*
TMT-B *T*-scores *M* (*SD*)	49.2 (10.1)	51.0 (9.7)	$t = -0.16$	*ns*
TMT B-A *T*-scores *M* (*SD*)	49.7 (10.7)	50.3 (8.9)	$t = -0.05$	*ns*
TMT B/A *T*-scores *M* (*SD*)	49.8 (11.1)	50.3 (8.2)	$t = -0.41$	*ns*
TMT-β *T*-scores *M* (*SD*)	49.3 (10.8)	50.9 (8.7)	$t = -0.13$	*ns*

n, number of participants; *p*, *p*-value; *t*, *t* statistic; *ns*, non-significant result; x^2, Pearson Chi-Square. Results are presented as mean (standard deviation) [range] except for sex which is characterized by female percentage; *T*-scores adjusted for pertinent demographics applying current proposed norms (Table 3).

Healthy controls from the Gothenburg MCI study were primarily recruited through senior citizen organizations and a small proportion were relatives of symptom group participants. Inclusion criteria for healthy controls in the Gothenburg MCI study were age between 50 and 79, absence of subjective symptoms of cognitive decline and MMSE score >26. Exclusion criteria were severe somatic diseases and severe psychiatric disorders, which could potentially influence cognitive performance. Neuropsychological examinations including TMT-A and TMT-B were performed by licensed clinical psychologists or psychologist-in-training under supervision by a licensed clinical psychologist. For further description of the Gothenburg MCI study cohort, see Wallin et al. (2016).

Between cohort comparisons of demographics and cognitive performance

Demographics and raw scores on basic and derived measures for DDI ($n = 170$) and Gothenburg MCI study ($n = 122$) controls are compared in Table 1. Although participants from the Gothenburg MCI study were older ($p < 0.05$) and had less education ($p < 0.001$) compared to the DDI controls, no differences were observed between cohorts for basic or derived TMT raw scores or *T*-scores adjusted for pertinent demographics. Due to large differences in time of inclusion within the Gothenburg MCI cohort (i.e. participants included within a 13-year time frame), potential cohort effects were investigated by including a separate variable accounting for time of testing on TMT-A and TMT-B *T*-scores. Results from this analysis showed that time of testing was not a significant predictor of performance on TMT.

TMT administration

The TMT (Reitan & Wolfson, 1985) was administered following standardized instructions described in Strauss et al. (2006, pp. 656–657). Reitan and Wolfson (1985) version

Table 2. Raw score to scaled score conversions.

Scaled score	TMT-A	TMT-B	TMT B − TMT A	TMT B/TMT A	Scaled score
1					1
2	≥71	≥166	≥131	≥5.330	2
3	66–70	160–165	121–130	4.810–5.329	3
4	64–65	155–159	106–120	4.380–4.809	4
5	58–63	135–154	99–105	3.980–4.379	5
6	54–57	121–134	80–98	3.600–3.979	6
7	46–53	108–120	68–79	3.180–3.599	7
8	40–45	96–107	59–67	2.830–3.179	8
9	37–39	86–95	50–58	2.580–2.829	9
10	33–36	78–85	43–49	2.350–2.579	10
11	29–32	71–77	38–42	2.100–2.349	11
12	27–28	63–70	32–37	1.930–2.099	12
13	25–26	58–62	28–31	1.810–1.929	13
14	22–24	51–57	23–27	1.630–1.809	14
15	21	47–50	20–22	1.540–1.629	15
16	19–20	41–46	10–19	1.330–1.539	16
17	17–18	40–41	2–10	1.040–1.329	17
18	16	34–39	(−4)–1	0.890–1.039	18
19	≤15	≤33	≤(−5)	≤0.889	19

Conversions were performed to normalize TMT scores from healthy controls ($n = 292$). Normalized scaled scores were later used for development of normative models (Table 3).

of TMT is administered in two parts: In TMT-A, the participant is required to connect 25 encircled numbers from low to high, while in TMT-B, the participant must alternate between numbers and letters, from low to high (e.g. 1-A-2-B-3-C). Scoring criteria is time to completion, measured manually by digital stopwatch. In short, participants were asked to complete the task as quickly as they could without making mistakes and were presented with a rehearsal trial before the test. Participants were given a moment to familiarize with initial connections and finishing point. Time to completion was recorded between the initiation of the first pen stroke and terminated at completion of the task. In case of mistakes (e.g. connecting wrong number to letter), the participants were corrected by the administrator and promptly guided to the last correctly connected letter or number. Time was not paused during this correction. If a participant aborted TMT-B, maximum time to completion was set (300 s), although no participants in the healthy control groups achieved maximum time nor were reported to abort the assignment. In the normative sample $n = 1$ participant (0.34%) only had available data from TMT-A and was excluded from analysis.

Data analysis

Regression norming procedure

Following procedures outlined in Kirsebom et al. (2019) and Testa et al. (2009) regression-based norms were developed based on the normative performance of the included healthy controls ($n = 292$). To normalize measures of the TMT, we first determined the reverse cumulative frequency distribution for TMT raw scores (i.e. the scaled score distributions were reversed to ensure that higher times to completion was equal to lower performance in our normative models), and then converted raw scores into standardized scaled scores ($M = 10$, $SD = 3$). Multiple linear regression analyses were conducted on the standardized scaled scores (Table 2) from basic and

Table 3. Normative regression models for the TMT in healthy controls ($n = 292$).

Variable	Predictor	*b*	Standard error *b*	*t*	*p*	Partial r^2	Adjusted r^2	*SD* residual
TMT-A	Intercept	19.437	1.188	16.36	<0.001			2.675
	Age	−0.144	0.019	−7.70	<0.001	0.170	0.167	
TMT-B	Intercept	16.921	1.430	11.84	<0.001			2.645
	Age	−0.139	0.019	−7.40	<0.001	0.159		
	Education	0.170	0.047	3.62	<0.001	0.036	0.208	
TMT-B (age only)	Intercept	19.854	1.201	16.53	<0.001			2.704
	Age	−0.150	0.019	−7.93	<0.001	0.178	0.175	
TMT B–A	Intercept	13.844	1.496	9.26	<0.001			2.767
	Age	−0.091	0.020	−4.61	<0.001	0.069		
	Education	0.174	0.049	3.55	<0.001	0.042	0.116	
TMT B–A (age only)	Intercept	16.855	1.256	13.42	<0.001			2.827
	Age	−0.102	0.020	−5.15	<0.001	0.084	0.081	
TMT B/A	Intercept	8.705	0.705	12.34	<0.001			2.958
	Education	0.141	0.052	2.73	<0.01	0.025	0.022	
TMT-β	Intercept	8.475	1.600	5.30	<0.001			2.352
	Age	−0.075	0.018	−4.12	<0.001	0.055		
	Education	0.149	0.042	3.56	<0.001	0.042		
	TMT-A	0.453	0.052	8.73	<0.001	0.209	0.372	
TMT-β (age only)	Intercept	10.851	1.483	7.32	<0.001			2.403
	Age	−0.083	0.018	−4.50	<0.001	0.066		
	TMT-A	0.463	0.053	8.76	<0.001	0.210	0.346	

Regression analyses were performed on normalized scaled scores (Table 2). *b*, unstandardized regression coefficient; *t*, the *t*-test statistic; *SD* Residual, standard deviation of the residual; *p*, *p*-value; partial r^2, explained variance from individual predictor; adjusted r^2, combined explained variance from the model; standard error *b*, standard error of the unstandardized beta coefficient.

derived measures of the TMT in a healthy control group ($n = 292$) with age, sex, and education included as predictors. We included squared and interaction terms in our models to investigate potential non-linear effects of age (i.e. performance on TMT increasing at younger ages, then dropping off at older ages), and potential interaction effects between predictors such as between age and education, sex and education, as well as three-way interaction effects between age, sex, and education. For the proposed TMT-β measure, we included the normalized scaled scores for the TMT-A as a covariate.

All measures of basic and derived TMT scores were analyzed using a backwards regression method and only models with predictors that significantly contributed to the overall explained variance were selected. We found that the Gothenburg MCI study cohort was older and less educated (potentially due to differences in recruitment methods) and we therefore included a covariate to assess a potential difference between cohorts for the TMT measures. However, when controlling for demographics, no differences between cohorts were observed. There were no effects of sex on performance for any of the measures. For TMT-A, only age remained a significant predictor of test performance. For the TMT-B and derived measures TMT B-A and TMT-β, both age and education were significant predictors. For TMT B/A, only education significantly predicted performance. On the proposed measure TMT-β, age, education, and normalized scaled scores on TMT-A were significant predictors. None of the squared terms or interaction terms provided additional explained variance in the model. Education may not always be a relevant normative demographic for all target populations (e.g. low educational attainment while scoring above average on age adjusted measures of intelligence). Thus, we also provide regression-norms omitting education as a covariate. These norms may be applied to scores from individuals who did not have access to education but otherwise would have benefited from it, or when deemed appropriate by the clinician. Regression coefficients and partial r^2 values for the different predictors are presented in Table 3. For these models, we assessed plots of regression predicted values to residuals values to ensure that the assumption of homoscedasticity was not violated, and normality of the residuals were visually inspected with Q-Q plots. No collinearity between predictor variables were observed in the selected models (variance inflation factor <1.2).

Calculating normative performance using regression-based norms

The normative effects of demographics on performance are first determined using the regression coefficients obtained from the multiple regression analysis (Table 3) described above using the following formula (Intercept + [individual age*age coefficient] + [years of education * education coefficient]). For example, for a 60-year-old woman with 13 years of education, the resulting equation on TMT-B would be: ([16.921] + [60 * −0.139]) + (13 * 0.170). This formula produces an individual predicted scaled score for TMT-B. We then subtract the scaled score obtained by the individual (Table 2) from the demographically adjusted predicted scaled score and divide by the standard deviation of the regression model residuals (Table 3) which yields a standardized Z-score (Obtained scaled model score − predicted scaled score/standard deviation of the residuals obtained from the regression = Z-score). The resulting Z-score is the demographically adjusted normative score based on the healthy control's normative performance on the TMT. Z-scores may be converted to *T*-scores by the following transformation ($T = z * 10 + 50$).

Comparisons of proposed norms to published norms

As the published norms by Heaton et al. (2004) and Tombaugh (2004) are only provided for basic measures (TMT-A and TMT-B), comparisons with the current proposed norms did not include derived measures (TMT B-A, B/A, and β). Proposed norms with only age as a covariate was also not compared to published norms since neither Heaton et al. (2004) nor Tombaugh (2004) offer this option. Normative performance (*T*-scores) on the TMT measures was calculated for the control group ($n = 292$) following the method described in the previous passage. Next, *T*-scores were calculated using published norms from Heaton et al. (2004) and Tombaugh (2004). This resulted in three sets of demographically adjusted *T*-scores, which were compared using paired samples *t*-tests. The control group ($n = 292$) was then split based on the median level of education into a low education group (<13 years of education) and a high education group (≥13 years of education) and demographically adjusted *T*-scores were again compared with paired samples *t*-tests to investigate differences in normative estimations. Distribution of *T*-scores was assessed with Shapiro–Wilks test of normality and visual comparison with histograms. Norms from Tombaugh (2004) were calculated based on mean scores and standard deviations reported in Tombaugh (2004) and then transformed to *T*-scores. In some cases, this provided highly abnormal *T*-scores <0 due to narrow standard deviations in certain stratifications of age and education, and negative *T*-scores were in these cases set to 0.

Multiple regression analyses were conducted using the same predictors (age, sex, and education) on the *T*-scores derived using norms from Heaton et al. (2004), Tombaugh (2004) and the current proposed Scandinavian norms (Table 2). Reasoning that these *T*-scores should be adjusted for demographic variables (e.g. differences in age should already be corrected for), we expect that results will not be statistically significant ($p \geq 0.05$) if *T*-scores adequately adjust for the demographical variables. Significant effects of any predictor variable would suggest that norms did not adequately correct for the demographical variable when applied to the Scandinavian sample.

Lastly, we examined relationships (Pearson's *r*) between basic measures (TMT-A and TMT-B) and derived measures (TMT B-A and TMT B/A) to the new proposed derived measure TMT-β. All analyses were conducted using the Statistical Package for Social Sciences (SPSS) version 25 and RStudio version 1.2.5033.

Norm calculator

To facilitate the usability and adoption of the proposed regression norms in the clinic, we provide a free web-based tool that computes the regression equations. To obtain normative *T*-scores for both basic (TMT-A and TMT-B) and derived measures (TMT B-A, TMT B/A and TMT-β), the user simply needs to enter valid demographic values (age and years of education) and raw-scores from TMT-A and TMT-B. Except for the TMT B/A, *T*-score calculations are provided for both demographically adjusted norms (age and education) as well as age adjustment only. The tool is implemented as a self-contained HTML/Javascript webpage, available at https://uit.no/ressurs/uit/cerad/tmt-calc.html and is released as open source at https://github.com/DDI-NO/tmt-calc under Apache License, version 2.0.

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 right to withdraw and potential risks and rewards involved.

Results

Effects of demographics on TMT test performance in the healthy control group

Normative regression models and explained variance from predictors for basic and derived measures of the TMT are reported in Table 3. In the following section, improved performance refers to higher scaled scores (Table 2), that is, faster time to completion on basic measures (TMT-A and TMT-B) and reduced difference scores on derived measures (TMT B-A and TMT B/A). On the proposed measure TMT-β, improved performance refers to higher scaled scores on TMT-B adjusting for TMT-A scores.

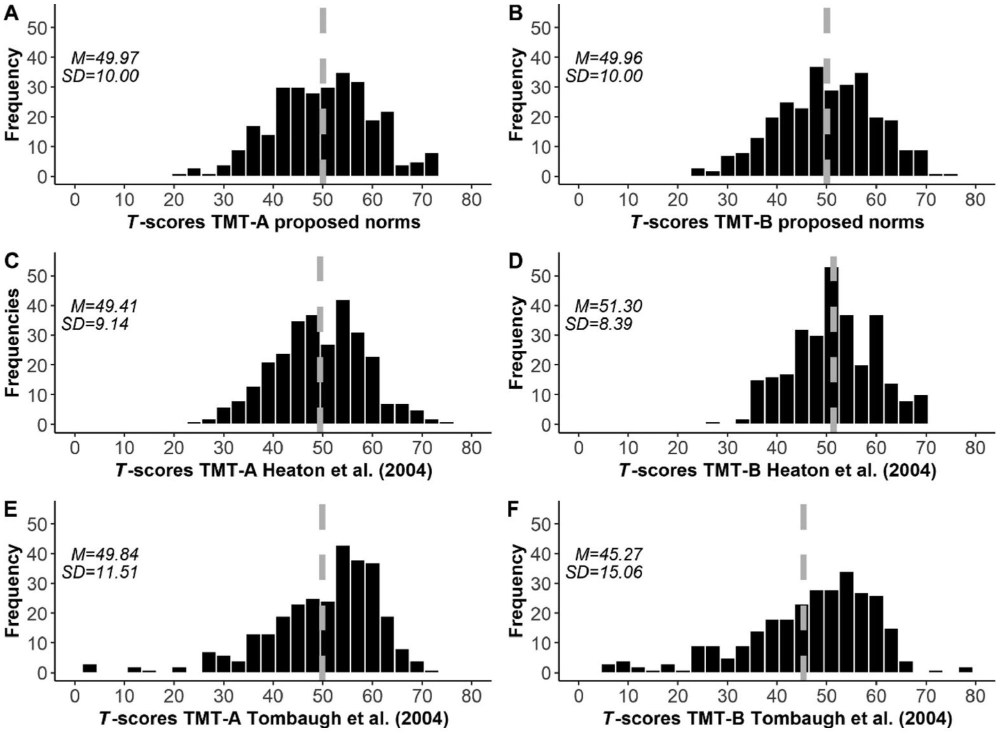

Figure 1. T-score distributions on TMT-A and TMT-B calculated using current proposed norms (A and B) and norms from Heaton et al. (2004) (C and D) and Tombaugh (2004) (E and F) in same control group ($n = 292$). The gray dashed line in each figure depicts the mean T-score for each norm. M and *SD* are mean and standard deviation, respectively.

Lower age and higher education predicted improved performance on TMT-B, TMT B-A, and TMT-β. On TMT-A, lower age was the only significant predictor for improved performance. Higher education was associated with improved performance on TMT-B, TMT B-A, TMT B/A, and TMT-β. Faster time to completion on TMT-A was associated with improved performance on the proposed measure TMT-β. When omitting education from the normative regression models explained variance from age increased slightly, but total explained variance from the model decreased on all measures.

Adjustment of demographics using published norms

Heaton et al. (2004) norms adequately adjusted for age ($b = -0.103$, $p = 0.101$) on TMT-A. However, Heaton et al. (2004) norms did not adequately correct for the effects of education ($b = -0.771$, partial $r^2 = 0.079$, $p < 0.001$; adjusted $r^2 = 0.076$, $F(3,288) = 8.950$, $p < 0.001$). A similar result was obtained for TMT-B where these norms did not adequately correct for effects of education ($b = -0.661$, partial $r^2 = 0.068$, $p < 0.001$), but adequately adjusted for the effects of age ($b = 0.016$, $p = 0.783$; adjusted $r^2 = 0.062$, $F(3,288) = 7.425$, $p < 0.001$). In contrast, norms from Tombaugh (2004) adequately adjusted for demographics on both TMT-A (adjusted $r^2 = -0.001$, $F(3,288) = 0.877$, $p = 0.453$) and TMT-B (adjusted $r^2 = 0.006$, $F(3,288) = 1.548$, $p = 0.202$).

Table 4. Comparison between normative estimates on the TMT in healthy controls ($n = 292$).

Variable	Test norms	*M* (*SD*)	*t*	*df*	*p*	*M*diff	95% CI Lower	95% CI Upper
TMT-A	Scandinavian	49.96 (10.00)						
	Tombaugh (2004)	49.86 (11.51)	0.42	291	0.676	0.12	−0.46	0.70
	Heaton et al. (2004)	49.41 (9.14)	2.25	291	0.025	0.56	0.07	1.04
TMT-B	Scandinavian	49.97 (10.00)						
	Tombaugh (2004)	45.27 (15.06)	9.26	291	<0.001	4.70	3.70	5.70
	Heaton et al. (2004)	51.30 (8.39)	−5.29	291	<0.001	1.33	−1.83	−0.84
TMT-A < 13 edu	Scandinavian	49.74 (10.15)						
	Tombaugh (2004)	50.37 (11.15)	1.54	142	0.125	0.62	−0.18	1.42
	Heaton et al. (2004)	51.85 (9.08)	6.74	142	<0.001	2.10	1.49	2.72
TMT-A ≥ 13 edu	Scandinavian	50.07 (10.05)						
	Tombaugh (2004)	49.30 (12.20)	−1.62	123	0.108	−0.77	−1.72	0.17
	Heaton et al. (2004)	46.95 (8.77)	−12.32	123	<0.001	−3.12	−3.62	−2.62
TMT-B < 13 edu	Scandinavian	49.60 (10.08)						
	Tombaugh (2004)	45.64 (14.61)	−5.56	142	<0.001	−3.96	−5.37	−2.55
	Heaton et al. (2004)	53.18 (7.90)	9.78	142	<0.001	3.58	2.86	4.30
TMT-B ≥ 13 edu	Scandinavian	49.99 (9.97)						
	Tombaugh (2004)	45.32 (15.40)	−6.04	123	<0.001	−4.67	−6.20	−3.14
	Heaton et al. (2004)	49.35 (8.68)	−2.41	123	0.017	−0.65	−1.18	−0.12

TMT Scores are *T*-scores adjusted for pertinent demographics. Tombaugh (2004) and Heaton et al. (2004) *T*-scores always compared to Scandinavian norms. *t*, the *t*-test statistic; *M*, mean; *SD*, standard deviation; *df*, degrees of freedom; *M*diff, mean difference; 95% CI, lower and upper confidence interval of the mean; *p*, p-value.

Distributions of T-scores using different norms

Visually comparing distributions of *T*-scores on TMT-A and TMT-B (Figure 1) showed differences in expected normal distributions. Distributions were normal and approximately similar between Heaton et al. (2004) and current proposed norms on TMT-A and TMT-B. In contrast, *T*-scores calculated using Tombaugh (2004) norms showed a non-normal distribution on TMT-A ($W(292) = 0.919$, $p < 0.001$) with a negative skew (−1.270) and leptokurtic kurtosis (kurtosis = 2.428). In addition, visually comparing the distribution showed a marked negative tail indicating an increased number of abnormal *T*-scores. This was also observed on TMT-B with Tombaugh (2004) *T*-scores (skew = −1.11, kurtosis = 1.27, $W(292) = 0.916$, $p < 0.001$).

Comparisons between mean normative estimates

Table 4 compares mean *T*-scores applying norms from Heaton et al. (2004) and Tombaugh (2004) with current proposed norms. On TMT-A, Tombaugh (2004) norms were not significantly different, but Heaton et al. (2004) norms produced lower mean *T*-scores. On TMT-B, Heaton et al. (2004) norms estimated higher mean *T*-scores and Tombaugh (2004) estimated considerably lower scores on TMT-B. Splitting the sample based on educational level showed that for individuals with less than 13 years of education, Heaton et al. (2004) norms produced higher *T*-scores and conversely produced lower *T*-scores for individuals with 13 or more years of education.

Correlations between TMT-β, TMT-A, TMT-B, B-A, and B/A

Correlations between all TMT *T*-score measures are shown in Table 5. A strong association was found between TMT-β and derived measure TMT B-A sharing 93.9% of the variance between measures. Both TMT B-A and TMT-β were highly correlated with

Table 5. Correlations between *T*-scores applying current proposed norms ($n = 292$).

TMT measures	TMT-A	TMT-B	TMT B-A	TMT-β
TMT-A	–			
TMT-B	0.457*	–		
TMT B-A	0.000^{ns}	0.864*	–	
TMT-β	−0.003^{ns}	0.886*	0.969*	–
TMT B/A	−0.522*	0.492*	0.830*	0.823*

*<0.001.
ns, non-significant result.

TMT-B sharing 74.7% and 78.5% of the variance, respectively. Both TMT B-A and TMT-β were associated with the TMT B/A measure sharing 68.9% and 67.7% of the variance, respectively. Neither TMT B-A nor TMT-β were associated with TMT-A indicating that performance on TMT-A had been adjusted for in both measures.

Discussion

In this study, we propose demographically adjusted test norms for basic and derived measures of TMT in a sample of Scandinavian adults between 41 and 84 years. We compared the proposed test norms to published norms from Heaton et al. (2004) and Tombaugh (2004) and assessed if these norms adequately adjust for demographics when applied to a Scandinavian sample. In addition, we propose a new regression-based approach for estimating the derived TMT B-A measure named TMT-β.

The effects of age on TMT-A and TMT-B were comparable to other regression-based norms with a similar age demographic (Peña-Casanova et al., 2009). Conversely, education accounted for much less variance on TMT-A and TMT-B (Gonçalves et al., 2013; Peña-Casanova et al., 2009). This discrepancy is likely due to differences in sample composition between the examined studies. For instance, Peña-Casanova et al. (2009) reported that about 20% of participants attained ≤5 years of education and over 20% attained ≥16 years. In contrast, the normative sample of Scandinavians employed in the current study had no participants with less than 6 years of education and generally a high level of education ($M = 13.21$, $SD = 3.34$; Table 1). Thus, the normative sample of Scandinavians had a restricted range of education compared to Peña-Casanova et al. (2009) which might explain why education accounted for less variance. While we believe the educational level observed in the Scandinavian sample is representative of the Scandinavian population (Eurostat, 2019), homogenic high levels of education limits the applicability of the norms to countries with a similar educational composition. Discrepancy between demographics of the initial normative sample and the target population where the norms are applied must be considered for reliable normative estimation, as argued by Heaton et al. (1999). Finally, sex did not contribute significantly to scores on any TMT measure which is consistent with most normative studies (Mitrushina et al., 2005, p. 69). As expected from earlier studies, derived measures TMT B-A and TMT B/A were less influenced by age and education than basic measures (Bezdicek et al., 2012; Gonçalves et al., 2013; Hester et al., 2005; Periáñez et al., 2007; Sanchez-Cubillo et al., 2009). As a result, adjusting for demographics on derived measures has less impact on normative estimations, but appropriate normative data based on a representative sample should still be used for reliable estimations.

Derived measures of TMT are employed to minimize the impact of visual search/visuomotor demands and subsequently enhance measurement of executive functioning associated with TMT-B. As an alternative approach to TMT B-A, we reasoned that we could regress TMT-A scores alongside pertinent demographics on TMT-B scores which would isolate the higher order executive functions associated with TMT-B. This new measure was named TMT-β to avoid confusion with the conventional TMT B-A approach. While the demographically adjusted TMT B-A and TMT-β *T*-scores were highly correlated in our sample (93.9% shared variance), TMT-β might still provide utility in clinical samples where both TMT-A and TMT-B is slow due to visual scanning and/or visuomotor deficits. This would result in an elevated difference score TMT B-A, thus giving the appearance of executive function deficits. Senior et al. (2018) showed that slow time to completion on both TMT-A and TMT-B occurred in 37% of cases in a clinical sample but when compared to others with similar TMT-A scores, 40% of these did not show a disproportionate increase in TMT-B, indicating that executive deficits were not the primary cause of the abnormal TMT B-A difference. Compared to the conventional TMT B-A measure, TMT-β should in these instances be able to discern the individuals who do not show a disproportionate increase in TMT-B completion times by adjusting scores based on their individual TMT-A completion time. As an example, a 75-year-old individual from a clinical sample with 9 years of education completing TMT-A in 71 s and TMT-B in 202 s estimates a demographically adjusted *T*-score of 25 on TMT B-A applying current proposed norms. In contrast, the same individual would receive a *T*-score of 35 on TMT-β. This indicates that TMT B-A may produce disproportionally low estimates of executive function as compared to the TMT-β when both TMT-A and TMT-B completion times are slow. TMT-β differs from the stratified approach used by Senior et al. (2018) as we employ multiple regression analysis to adjust for TMT-A completion time. This allows for the adjustment of TMT-A performance at a continuous level while at the same time correcting for normative effects of age and education. We have introduced TMT-β with some potential advantages discussed, but further research into criterion validity and clinical applications need to be established. Compared with the traditional TMT B-A measurement, we hypothesize that TMT-β should be better able to discern individuals with abnormal TMT B-A scores, and therefore correlate more strongly with cognitive flexibility and associated brain structures, particularly in clinical samples.

A key objective of this study was to compare norms from Heaton et al. (2004), Tombaugh (2004) and the current proposed norms in a Scandinavian sample. While the Heaton et al. (2004) norms produced apparently similar distributions of *T*-scores as current proposed norms (Figure 1), results from multiple regression analysis showed that significant effects of education were still evident on TMT-A (7.8%) and TMT-B (6.8%). The associated beta coefficients were negative, suggesting that the Heaton et al. (2004) norms generally overestimated the significance of education when applied in the Scandinavian sample. Individuals with lower educational attainment had significantly higher *T*-scores than expected while individuals with higher educational attainment had lower *T*-scores (Table 4). On TMT-A, Heaton et al. (2004) reported 10% explained variance from education, however no effects of education were evident in the Scandinavian sample on TMT-A. On TMT-B, Heaton et al. (2004) reported 16% on education compared with 4% in the Scandinavian sample. Thus, education accounted for larger amounts of variability in the initial normative sample employed in the Heaton et al. (2004) norms, providing a likely explanation for why norms overestimated the effects of education when applied in the Scandinavian sample. Education is generally considered more affordable and available to the public in Scandinavian countries which might be why education apparently has less impact on scores. Future normative studies in Scandinavia should compare the effects of demographic corrections to investigate if this applies to other neuropsychological measures as well.

T-scores from Tombaugh (2004) produced non-normal distributions with a negative skew and leptokurtic kurtosis (Figure 1) and subsequently lower mean scores on TMT-B (Table 4). This likely stems from narrow standard deviations of mean scores in certain stratifications of age and education in the Tombaugh (2004) sample, whereby slight deviation in scores result in highly abnormal *T*-scores for a substantial proportion of the Scandinavian sample. In terms of demographic corrections, however, results from multiple regression analysis showed that *T*-scores from Tombaugh (2004) adequately adjusted for age and education in the Scandinavian sample. Tombaugh (2004) also reported that age was the largest contributor to variance on TMT-A and TMT-B with only marginal effects of education. Results from multiple regression analysis suggested that normative estimates were comparable to the Scandinavian sample.

We provide normative regression models omitting education as a covariate. Education may not always be a relevant normative demographic for all target populations (e.g. low educational attainment while scoring above average on age adjusted measures of intelligence). The implications of using these norms for individuals with low educational attainment are slightly stricter normative corrections (i.e. lower *T*-scores). It can be appropriate to use these norms in instances where an individual did not have the opportunity for education that they otherwise would have benefited from. These norms should not be applied to individuals who lack education because they could not comprehend the material or otherwise were not eligible (Mitrushina et al., 2005, p. 31). Our results indicate that age accounted for slightly more variance in scores when omitting education as a covariate but overall explained variance in the models decreased (Table 3). Norms correcting for all pertinent demographics should therefore be used when appropriate.

Some limitations need to be addressed. First, an important limitation of this study was the lack of an independent sample of healthy controls to apply and assess our proposed norms. We therefore opted to compare current proposed norms to published norms within in the same sample ($n = 292$). Second, healthy controls enrolled in the normative sample were not screened for perceptual-motor deficits which might inhibit performance on the TMT prior to testing. Lastly, it is important to emphasize that the current proposed norms are not *better* than the published norms, but simply that there is an advantage to applying local norms, as shown when comparing current proposed norms to published norms in the Scandinavian sample. We also stress that the users of the current proposed norms should follow the same administration procedures on TMT for reliable estimates, which are described in Strauss et al. (2006).

Conclusions

We propose demographically adjusted regression-based norms for age 41 through 84 years on TMT-A and TMT-B and derived measures TMT B-A and TMT B/A based on healthy controls from the Norwegian DDI and Swedish Gothenburg MCI cohorts. We also propose a new measure named TMT-β developed using a regression-based procedure to improve on the conventional TMT B-A. Comparisons of norms from Heaton et al. (2004) and Tombaugh (2004) suggest that current proposed norms are better suited for use in a Scandinavian population. To ease the use and availability of the regression norms in clinical settings, a free online norm calculator is offered https://uit.no/ressurs/uit/cerad/tmt-calc.html.

Acknowledgments

We thank Svein Ivar Bekkelund, Kjell-Arne Arntzen, Kai Müller, Claus Albretsen, Mari Thoresen Løkholm, Ida Harviken, Line Saether, Ingrid Myrvoll Lorentzen, Erna Utnes, Marianne Wettergreen, Berglind Gisladottir, Marit Knapstad, Reidun Meling, Synnøve Bremer Skarpenes and Elin Margrethe Solli for clinical examinations and essential help with the project.

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.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by 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. Additional support was received from the Sahlgrenska University Hospital, the Swedish Research Council, Swedish Brain Power, the Swedish Dementia Foundation, the Swedish Alzheimer Foundation, Stiftelsen Psykiatriska forskningsfonden, and Konung Gustaf V:s and Drottning Victorias Frimurarestiftelse. 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
Knut Waterloo http://orcid.org/0000-0003-3447-8312
Tormod Fladby http://orcid.org/0000-0002-9984-9797
Bjørn-Eivind Kirsebom http://orcid.org/0000-0002-1413-9578

References

Abi Chahine, J., Rammal, S., Fares, Y., & Abou Abbas, L. (2019). Trail making test: Normative data for the Lebanese adult population. *The Clinical Neuropsychologist*, 4. https://doi.org/10.1080/13854046.2019.1701710

Arbuthnott, K., & Frank, J. (2000). Executive control in set switching: residual switch cost and task-set inhibition. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Experimentale*, *54*(1), 33–41. https://doi.org/10.1037/h0087328

Bezdicek, O., Motak, L., Axelrod, B. N., Preiss, M., Nikolai, T., Vyhnalek, M., Poreh, A., & Ruzicka, E. (2012). Czech version of the trail making test: normative data and clinical utility. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *27*(8), 906–914. https://doi.org/10.1093/arclin/acs084

Cavaco, S., Gonçalves, A., Pinto, C., Almeida, E., Gomes, F., Moreira, I., Fernandes, J., & Teixeira-Pinto, A. (2013). Trail making test: regression-based norms for the Portuguese Population. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *28*(2), 189–198. https://doi.org/10.1093/arclin/acs115

Dickinson, M. D., & Hiscock, M. (2011). The Flynn effect in neuropsychological assessment. *Applied Neuropsychology*, *18*(2), 136–142. https://doi.org/10.1080/09084282.2010.547785

Dodge, H. H., Zhu, J., Lee, C.-W., Chang, C.-C H., & Ganguli, M. (2014). Cohort effects in age-associated cognitive trajectories. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, *69*(6), 687–694. Retrieved from doi:https://doi.org/10.1093/gerona/glt181

Eurostat (2019). *Population by educational attainment level, sex and age (%)*. Retrieved from: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=edat_lfs_9903&lang=en&fbclid=IwAR1fKmi-U_BrKWdSr6v5yCPtAv3LgaJuLBS4lDnZqV2LLvKXhXvihB1WWRM

Fernandez, A. L., & Marcopulos, B. A. (2008). A comparison of normative data for the trail making test from several countries: equivalence of norms and considerations for interpretation. *Scandinavian Journal of Psychology*, *49*(3), 239–246. Retrieved from https://doi.org/10.1111/j.1467-9450.2008.00637.x

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

Gaudino, E. A., Geisler, M. W., & Squires, N. K. (1995). Construct validity in the trail making test: what makes part B harder? *Journal of Clinical and Experimental Neuropsychology*, *17*(4), 529–535. https://doi.org/10.1080/01688639508405143

Goul, W. R., & Brown, M. (1970). Effects of age and intelligence on trail making test performance and validity. *Perceptual and Motor Skills*, *30*(1), 319–326. https://doi.org/10.2466/pms.1970.30.1.319

Heaton, R. K., Avitable, N., Grant, I., & Matthews, C. G. (1999). Further crossvalidation of regression-based neuropsychological norms with an update for the Boston naming test. *Journal of Clinical and Experimental Neuropsychology*, *21*(4), 572–582. https://doi.org/10.1076/jcen.21.4.572.882

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.

Hester, R. L., Kinsella, G. J., Ong, B., & McGregor, J. (2005). Demographic influences on baseline and derived scores from the trail making test in healthy older Australian adults. *The Clinical Neuropsychologist*, *19*(1), 45–54. https://doi.org/10.1080/13854040490524137

Kennedy, K. J. (1981). Age effects on trail making test performance. *Perceptual and Motor Skills*, *52*(2), 671–675. https://doi.org/10.2466/pms.1981.52.2.671

Kirsebom, B. E., Espenes, R., Hessen, E., Waterloo, K., Harald Johnsen, S., Gundersen, E., … 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, DOI: https://doi.org/10.1080/13854046.2019.1574902

Lamberty, G. J., Putnam, S. H., Chatel, D. M., & Bieliauskas, L. A. &. (1994). Derived trail making test indices: a preliminary report. *Neuropsychiatry, Neuropsychology, & Behavioral Neurology*, *7*(3), 230–234.

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

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

Moll, J., Oliveira-Souza, R. d., Moll, F. T., Bramati, I. E., & Andreiuolo, P. A. (2002). The cerebral correlates of set-shifting: an fMRI study of the trail making test. *Arquivos de Neuro-Psiquiatria*, *60*(4), 900–905. https://doi.org/10.1590/s0004-282x2002000600002

Peña-Casanova, J., Quiñones-Ubeda, S., Quintana-Aparicio, M., Aguilar, M., Badenes, D., Molinuevo, J. L., Torner, L., Robles, A., Barquero, M. S., Villanueva, C., Antúnez, C., Martínez-Parra, C., Frank-García, A., Sanz, A., Fernández, M., Alfonso, V., Sol, J. M., & Blesa, R., NEURONORMA Study Team (2009). Spanish Multicenter Normative Studies (NEURONORMA Project): norms for verbal span, visuospatial span, letter and number sequencing, trail making test, and symbol digit modalities test. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *24*(4), 321–341. https://doi.org/10.1093/arclin/acp038

Periáñez, J. A., Ríos-Lago, M., Rodríguez-Sánchez, J. M., Adrover-Roig, D., Sánchez-Cubillo, I., Crespo-Facorro, B., Quemada, J. I., & Barceló, F. (2007). Trail making test in traumatic brain injury, schizophrenia, and normal ageing: sample comparisons and normative data. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *22*(4), 433–447. Retrieved from https://doi.org/10.1016/j.acn.2007.01.022

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

Reitan, R. M., & Wolfson, D. (2004). The trail making test as an initial screening procedure for neuropsychological impairment in older children. *Archives of Clinical Neuropsychology*, *19*(2), 281–288. https://doi.org/10.1016/S0887-6177(03)00042-8

Sanchez-Cubillo, I., Perianez, J. A., Adrover-Roig, D., Rodriguez-Sanchez, J. M., Rios-Lago, M., Tirapu, J., & Barcelo, F. (2009). Construct validity of the trail making test: role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. *Journal of the International Neuropsychological Society*, *15*(3), 438–450. https://doi.org/10.1017/s1355617709090626

Senior, G., Piovesana, A., & Beaumont, P. (2018). Discrepancy analysis and Australian norms for the Trail Making Test. *The Clinical Neuropsychologist*, *32*(3), 510–523. https://doi.org/10.1080/13854046.2017.1357756

Siciliano, M., Chiorri, C., Battini, V., Sant'Elia, V., Altieri, M., Trojano, L., & Santangelo, G. (2019). Regression-based normative data and equivalent scores for trail making test (TMT): an updated Italian normative study. *Neurological Sciences*, *40*(3), 469–477. https://doi.org/10.1007/s10072-018-3673-y

St-Hilaire, A., Parent, C., Potvin, O., Bherer, L., Gagnon, J.-F., Joubert, S., Belleville, S., Wilson, M. A., Koski, L., Rouleau, I., Hudon, C., & Macoir, J. (2018). Trail making tests A and B: regression-based normative data for Quebec French-speaking mid and older aged adults. *The Clinical Neuropsychologist*, *32*(sup1), 77–90. https://doi.org/10.1080/13854046.2018.1470675

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

Strobel, C., Johansen, H., Aga, O., Bekkhus-Wetterberg, P., Brierly, M., Egeland, J., Follesø, K., Rike, P., Schanke, A. (2018). Manual Norsk Revidert trail making test (TMT-NR3). Retrieved from https://aldring-og-helse-media.s3.amazonaws.com/documents/TMT-NR3_AoH_Manual_2018_4UhFcRC.pdf

Stuss, D. T., Stethem, L. L., & Pelchat, G. (1988). Three tests of attention and rapid information processing: an extension. *Clinical Neuropsychologist*, *2*(3), 246–250. https://doi.org/10.1080/13854048808520107

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. Retrieved from https://www.cambridge.org/core/article/accounting-for-estimated-iq-in-neuropsychological-test-performance-with-regressionbased-techniques/8D4BCB20747A14F10D656971A9F96160. https://doi.org/10.1017/S1355617709990713

Tombaugh, T. N. (2004). Trail making test A and B: normative data stratified by age and education. Archives of Clinical Neuropsychology, *19*(2), 203–214. Retrieved from https://doi.org/10.1016/S0887-6177(03)00039-8

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