Archives of Clinical Neuropsychology Advance Access published September 30, 2012

OXFORD
UNIVERSITY PRESS

Archives
of
CLINICAL
NEUROPSYCHOLOGY

Czech Version of the Trail Making Test: Normative Data and Clinical Utility

Ondrej Bezdicek[1,*], Ladislav Motak[2,3], Bradley N. Axelrod[4], Marek Preiss[5], Tomas Nikolai[1], Martin Vyhnalek[6,7], Amir Poreh[8], Evzen Ruzicka[1]

[1]*Department of Neurology and Centre of Clinical Neuroscience, 1st Faculty of Medicine and General University Hospital in Prague, Charles University in Prague, Prague, Czech Republic*
[2]*Université de Lyon, Lyon, France*
[3]*IFSTTAR, LESCOT, Bron, France*
[4]*John D. Dingell Department of Veterans Affairs Medical Center, Detroit, MI, USA*
[5]*Prague Psychiatric Center, 3rd Faculty of Medicine, Charles University in Prague, Prague, Czech Republic*
[6]*Memory Disorders Clinic, Department of Neurology, 2nd Faculty of Medicine, Charles University in Prague, Prague, Czech Republic*
[7]*International Clinical Research Center, St. Anne's University Hospital Brno, Czech Republic*
[8]*Cleveland State University, Cleveland, OH, USA*

*Corresponding author at: Department of Neurology and Centre of Clinical Neuroscience, Charles University in Prague, First Faculty of Medicine and General University Hospital in Prague, Kateřinská 30, 128 21 Praha 2, Czech Republic. Tel.: +420-224-965-524/+420-731-478-794; fax: +420-224-922-678.
E-mail address: ondrejb@yahoo.com

Accepted 30 August 2012

Abstract

The Trail Making Test (TMT) comprises two psychomotor tasks that measure a wide range of visual-perceptual and executive functions. The purpose of this study was to provide Czech normative data and to examine the relationship between derived TMT indices and demographic variables. The TMT was administered to 421 healthy adults. Two clinical groups ($n = 126$) were evaluated to investigate the clinical utility of the TMT-derived scores: amnestic mild cognitive impairment ($n = 90$) and Alzheimer's disease ($n = 36$). Statistical analyses showed that age and education, but not gender, were significantly associated with TMT completion times and derived scores. Of all the indices, only the TMT ratio score was insensitive to age. We present normative values for the Czech version of the TMT, providing a reference for measuring individual performance in native Czech speakers. Moreover, we found that accuracy on the TMT was improved with the attenuation of age.

Keywords: Trail Making Test; Czech; amnestic mild cognitive impairment; Alzheimer's disease; normative data

Introduction

The Trail Making Test (TMT) is ranked 3 of the 40 most frequently used neuropsychological instruments in clinical practice in North America (Rabin, Barr, & Burton, 2005). The TMT consists of two parts, Part A and Part B (TMT-A and TMT-B). In TMT-A, the subject is instructed to connect 25 digits randomly distributed on a sheet of paper as quickly as possible by drawing lines between them in increasing order. In TMT-B, the task is to alternate in ascending and alphabetical order between both numbers and letters (e.g., 1–A–2–B, etc.). Based on the (Reitan & Wolfson, 1993) administration procedure, two scores are obtained that reflect the total time (in s) for task completion for both TMT-A and TMT-B. TMT-A is considered as a measure of psychomotor speed and visual attention, whereas TMT-B is considered as a measure of executive control, more specifically cognitive flexibility (Jacobson et al., 2011; Kortte, Horner, & Windham, 2002) and set shifting (Arbuthnott & Frank, 2000).

The TMT is considered to be sensitive to many dysfunctions in both adults and children (Reitan & Wolfson, 2004). For example, it is sensitive to cognitive deficits following the subarachnoid hemorrhage (Haug et al., 2007; Ogden, Mee, & Henning, 1993; Orbo et al., 2008) and Alzheimer's disease (AD; Amieva et al., 1998; Cahn et al., 1995; Chen et al., 2000;

doi:10.1093/arclin/acs084

Lafleche & Albert, 1995). However, there may be some limitations to its clinical utility. Although the TMT-B can differentiate executive dysfunction between different neuropsychiatric conditions (e.g., Parkinson's disease and progressive supranuclear palsy [Pillon et al., 1995]), neither the TMT-A nor the TMT-B can discriminate between the executive deficits of AD and vascular dementia (Barr, Benedict, Tune, & Brandt, 2004). This non-specificity of the TMT may be due to its scoring method, which implicitly combines the cognitive processes of processing speed, visual scanning, and cognitive flexibility.

Two other TMT scoring methods were developed (Lamberty & Axelrod, 2006) that allow for the measurement of executive control and set shifting in a manner that is independent of psychomotor speed and visual scanning. One method is the TMT difference score where the total time to complete TMT-A is subtracted from the total time to complete TMT-B (TMT-B − A; subtraction of TMT-A from TMT-B; Heaton, Nelson, Thompson, Burks, & Franklin, 1985). The other scoring method (Golden, Osmon, Moses, & Berg, 1981), the TMT ratio score, derives a ratio between the total time to task completion for TMT-B and TMT-A (TMT-B/A; TMT-B completion time divided by TMT-A completion time). A ratio score of <2.0 indicates relative impairment on TMT-A, whereas a ratio score of >3.0 indicates relative impairment on TMT-B. Prior research has provided inconsistent findings with regard to the relationship between the TMT-B/A ratio score, with some suggesting that there is no relationship (Horton & Roberts, 2001; Lamberty & Axelrod, 2006; Lamberty, Putnam, Chatel, Bielauskas, & Adams, 1994) and one indicating a relationship between age and education, especially in advanced age groups (Drane, Yuspeh, Huthwaite, & Klingler, 2002).

The interpretation of TMT performance is based on those normative variables that affect performance, which are stratified by age, education, intellectual ability, and ethnicity (Abe et al., 2004; Bornstein & Suga, 1988; Drane et al., 2002; Hester, Kinsella, Ong, & McGregor, 2005; Rasmusson et al., 1998; Soukup, Ingram, Grady, & Schiess, 1998; Tombaugh, 2004; Zalonis et al., 2008). Consequently, normative data across different countries are not equivalent. Therefore, the basis of normative comparisons for different populations should be derived from the culture in which they were obtained (Fernández & Marcopulos, 2008). Creating normative data for the TMT Czech version could help establish cultural-specific data thereby minimizing the interpretive impact of misapplying non-cultural-specific normative information (Manly, 2008; van de Vijver & Tanzer, 2004).

The principal goals of this study were to (a) provide normative data for the TMT in the Czech population stratified by age and education (Drane et al., 2002; Hester et al., 2005; Preiss & Preiss, 2006; Preiss, Rodriguez, Kawaciukowa, & Laing, 2007; Tombaugh, 2004; Zalonis et al., 2008), (b) determine if the TMT ratio score can better differentiate between different clinical groups than time to task completion (Arbuthnott & Frank, 2000; Lamberty et al., 1994; Lamberty & Axelrod, 2006; Martin, Hoffman, & Donders, 2003), (c) test whether the TMT ratio score is relatively resistant to demographic influences in various clinical neuropsychiatric groups (Lamberty & Axelrod, 2006), and (d) create a basis for cross-cultural comparison with existing normative studies from Western cultures (Fernández & Marcopulos, 2008).

Methods

Participants

The study enrolled a healthy control group and two clinical cohorts. The control group (Ctrl) consisted of 421 healthy subjects (161 men and 260 women; Table 1) who we recruited from community-dwelling volunteers. All were required to have no history of brain damage, psychiatric illness, chronic drug or alcohol abuse, or any medical illness that could affect neurocognitive function. Objective cognitive abilities were within normal limits (i.e., not more than 1 *SD* below age- and education-adjusted normative values in neuropsychological testing), and there were no reports of subjective memory complaints in the healthy control cohort.

The first clinical cohort consisted of 90 subjects (42 men and 48 women) diagnosed with amnestic mild cognitive impairment (aMCI) according to Petersen's criteria (2000) with memory test scores of -1.5 *SD*. The second clinical cohort comprised 36 subjects (8 men and 28 women) with probable AD according to the McKhann and colleagues criteria (1984). All patients with aMCI and probable AD underwent a standard examination protocol including magnetic resonance imaging, neurological, medical, and laboratory evaluation, a semi-structured clinical neuropsychological interview, and a neuropsychological test battery sensitive in detecting aMCI (Hort et al., 2007; Laczó et al., 2011, 2009).

Materials

We administered the paper–pencil versions of TMT-A and TMT-B to all study participants. Subjects were required to use a pencil to connect, with a line in proper ascending order, 25 encircled Arabic numerals that were randomly arranged on an A4 page (21 × 29 cm). For TMT-B, subjects connected stimuli that consisted of 25 encircled Arabic numerals and Latin letters in alternating order from A to K (not L in the Czech version). In Czech orthography, there is a "CH" digraph, a pair of characters

Table 1. Basic participant characteristics, including demographic factors, RAVLT scores, and also performance on TMT-A, TMT-B, and derived TMT scores (B − A, difference; B/A, ratio)

Variables	Ctrl ($n = 421$)		AD ($n = 36$)		aMCI ($n = 90$)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Gender ratio (men/women)	161/261*		8/28*		43/50	
Age (years)	47.69^{a}	17.28	75.06^{b}	7.16	72.72^{b}	9.28
Education (years)	14.18^{a}	3.31	11.78^{c}	2.80	13.87^{a}	3.29
RAVLT	53.82^{a}	10.24	20.75^{c}	7.56	30.85^{d}	8.78
TMT-A (s)	$32.85^{a,c}$	0.62	107.92^{b}	118.15	50.91^{c}	51.45
TMT-B (s)	81.59^{a}	2.16	368.33^{b}	147.61	222.67^{c}	151.95
B − A (s)	48.74^{a}	1.86	$260.42^{b,c}$	138.42	$171.75^{c,d}$	137.50
B/A	$2.54^{a,c}$	0.05	$4.82^{b,c}$	2.71	$4.71^{b,c}$	2.57

Notes: Groups: Ctrl = healthy controls; AD = Alzheimer's disease; aMCI = amnestic mild cognitive impairment. RAVLT = Rey Auditory Verbal Learning Test Czech version, sum of Trial 1 + 2 + 3 + 4 + 5 (Preiss et al., 2007); TMT = Trail Making Test Czech version (Preiss et al., 2007); *M* = mean (expected mean for TMT-related variables, based on the following theoretical values of covariates: $M_{age} = 52.44$, $M_{education} = 13.99$, $M_{RAVLT} = 45.39$); *SD* = standard deviation. Different alphabets across one line indicate significantly different mean values (TMT-related superscripts relate to statistical treatments performed on transformed data).
*All χ^2 (1) > 11.11, all $p \leq .001$.

used to write one phoneme (a distinct consonant /×/ in the International Phonetic Alphabet, e.g., "prach" [prax] "dust") that does not correspond to the normal values of the two characters combined (the Czech alphabet and a tabular, letter-by-letter comparison with the English alphabet, can be sent on request). This digraph is considered an individual letter, has its own place in the alphabet (after "H" and before "I"), and cannot be separated into constituent graphemes. The Czech version of the TMT is based on the original format of the English version (Preiss & Preiss, 2006; Reitan & Wolfson, 1993; U.S. War Department, Adjutant General's Office, 1944) and is part of the neuropsychological battery of the Prague Psychiatric Center (Preiss et al., 2007).

Procedure

All participants were instructed in the aims and procedures of the study, provided signed, informed consent, and completed a neuropsychological test battery. The study was approved by the local medical ethics committee. Healthy volunteers completed a neuropsychological test battery administered according to the standardized test procedures. The TMT was administered in a standardized manner as part of the neuropsychological evaluation and provided sufficient motivation and effort. The administration of the TMT followed the procedures outlined in Strauss, Sherman, and Spreen (2006, p. 656). The total score for TMT-A and TMT-B were measured as the total time in seconds required to complete both tasks (e.g., summation of total completion time of A and then separately of B). If subjects made an error(s), the examiner immediately called it to their attention, and then they had to proceed from the point at which the mistake occurred. Time did not stop during errors and correction of errors (Preiss et al., 2007; Strauss, Sherman, and Spreen, 2006), and errors therefore were reflected in the total completion time rather than as a separate index (Reitan & Wolfson, 1993). We computed mean total completion time (in s) for the TMT-A and TMT-B, and we derived the difference (TMT-B − A) and ratio (TMT-B/A) scores. Subjects also completed the Rey Auditory Verbal Learning Test (RAVLT), Beck Depression Inventory-Second Edition (BDI-II), and the 15-Item Geriatric Depression Scale (GDS-15). The tests were administered and scored following manualized instructions (Beck, Steer, & Brown 1996; Schmidt, 2004; Sheikh & Yesavage, 1986).

Statistical Analyses

Pearson's correlation coefficients were used to describe the strength and direction of the linear relationships between the TMT and demographic variables (age and education). As these variables were significantly correlated with TMT performance, they were used as covariates in the initial data analyses. Analyses of variance (ANOVA) and covariance (ANCOVA) models were applied to evaluate differences among the groups. Significant findings were followed with *post hoc* analyses using Tukey's HSD. Multiple comparisons were adjusted using the Sidak correction method. Effect sizes were reported as partial eta squared (η^2). In accordance with prior studies (Martin et al., 2003; Stuss et al., 2001), all TMT-related variables showed a positive skew and were submitted to logarithmic transformation ($\log_{10}$) prior to all analyses. However, to facilitate the comprehension of results, all values presented in the tables remain in standard, non-transformed units (Mitrushina, Boone,

Razani, & D'Elia, 2005). In the case of ANCOVA, the expected means and *p*-values are reported, and they are based on theoretical mean values of all considered covariates. The α-level was set at 0.05. All presented analyses were performed using IBM SPSS Statistics software (Version 19.0, SPSS Inc., Chicago, IL, USA).

Results

Psychometric Properties of the Normative Sample

Demographic data for the Czech normative sample are presented in Table 1. As age was significantly correlated with all TMT indices (TMT-A, $n = 420$, $r = .40$, $p < .001$; TMT-B, $n = 420$, $r = .46$, $p < .001$; TMT-B − A, $n = 420$, $r = .38$, $p < .001$; TMT-B/A, $n = 420$, $r = .24$, $p < .001$), accounting for 16%, 21%, 11%, and 5% of variance, respectively (Table 2), participants were assigned to seven different age-groups that each spanned 10 years. Education was also correlated with TMT scores (Horton & Roberts, 2001; Lamberty et al., 1994; TMT-A, $n = 420$, $r = -.17$, $p < .001$; TMT-B, $n = 420$, $r = -.26$, $p < .001$; TMT-B − A, $n = 420$, $r = -.19$, $p < .001$; TMT-B/A, $n = 420$, $r = -.20$, $p < .001$), accounting for TMT performance to a lesser extent than age (3%, 7%, 3% and 4%, respectively, Table 2). Education was dichotomized into two levels (≤12 years vs. ≥13 years) and was taken into account in further analyses.

A 2 × 7 ANOVA with education and age as the between-subject factor was performed separately on all TMT scores and revealed that participants with ≥13 years of education had shorter completion times (TMT-A and TMT-B) and lower composite scores (TMT-B − A and TMT-B/A) than participants with ≤12 years of education. Regarding age, it appeared that both TMT-A and TMT-B completion times were longer in participants of ≥55 years (Table 3), whereas there was no significant difference on either the TMT-B − A or TMT-B/A scores. The lack of any significant interaction indicated that the impact of education remains relatively constant, irrespective of participant's age. Normative data sampled in 10-year intervals for all four TMT-related indicators in a population of Czech native speakers, including the 95% confidence intervals, are presented in Table 3. The final prediction equations resulting from a regression-based approach (e.g., Uttl, 2002) are available in Appendix.

TMT Clinical Utility

ANCOVAs were performed to explore the clinical utility of the TMT. Although we were interested in the effect of group status (Ctrl vs. AD vs. aMCI) on TMT performance, the latter was significantly correlated with age ($n = 547$, $r = .27$ to .53,

Table 2. Main and interaction effects of ANOVA (*F*-values, preceded by corresponding degrees of freedom) performed on all TMT scores

Effect	*df*1, *df*2	TMT-A	TMT-B	TMT-B − A	TMT-B/A
Age	6,405	15.71***	5.61***	1.59	1.17
Education	1,405	30.57***	55.34***	39.59***	11.71***
Age × Education	6,405	1.11	1.46	0.58	0.73

Notes: TMT = Trial Making Test; *df* = degrees of freedom.
***$p \leq .001$.

Table 3. Demographic characteristics of the normative sample and age-adjusted normative data for all four TMT-related indicators in a population of Czech native speakers

Age (years)	*n*	Age		Education		Gender		TMT-A			TMT-B			TMT-B − A			TMT-B/A		
		M	*SD*	*M*	*SD*	Men	Women	*M*	*SD*	95% CI	*M*	*SD*	95% CI	*M*	*SD*	95% CI	*M*	*SD*	95% CI
20–24	35	21.86	1.44	14.43	2.39	15	20	27.74	10.45	24–31	67.09	37.42	54–80	39.34	30.60	29–50	2.41	0.81	2.13–2.69
25–34	95	29.53	2.78	15.18	4.09	49	46	28.20	9.93	26–30	72.31	34.00	65–79	44.28	29.88	38–50	2.66	1.00	2.46–2.89
35–44	56	38.70	2.92	13.71	2.82	25	31	27.21	7.13	25–29	70.57	26.55	63–78	43.36	25.21	37–50	2.67	1.01	2.40–2.94
45–54	71	49.28	2.79	14.04	3.40	26	45	32.38	9.70	30–35	83.41	53.46	71–96	51.03	47.67	40–62	2.53	1.02	2.29–2.77
55–64	79	59.58	2.84	13.66	3.19	21	58	35.56	13.23	33–39	88.44	54.00	76–101	52.89	46.83	42–63	2.50	0.95	2.29–2.72
65–74	64	69.25	2.87	13.95	2.87	21	43	40.27	12.39	37–43	90.44	35.29	82–99	50.17	33.16	42–58	2.36	0.92	2.13–2.56
75–84	19	79.11	2.60	13.32	2.61	4	15	46.79	20.88	36–58	115.74	54.56	86–146	68.95	45.57	44-93	2.53	0.84	2.15–2.91
Total sample	421	47.62	17.25	14.18	3.32	161	260	32.85	12.75	32–34	81.56	44.39	77–86	48.71	38.28	45–52	2.54	0.96	2.45–2.63

Notes: *n* = number; *M* = mean; *SD* = standard deviation; CI = confidence interval.

$p < .001$), education ($n = 547$, $r = -.20$ to $-.30$, $p < .001$), and RAVLT ($n = 547$, $r = -.46$ to $-.69$, $p < .001$) (Table 4). To determine whether the GDS-15 depression symptom severity total score should be included as a covariate (the ANCOVA), we correlated it with TMT performance. The GDS-15 scores did not correlate with any of the TMT-related scores ($n = 126$, all $r \leq .01$, all $p > .88$). There was a positive but only weak correlation between the BDI-II and the TMT-related scores ($n = 421$, all $r \leq .19$, all $p < .04$), and thus depression-related scores were not included in further analyses.

Comparisons were performed among the Ctrl, AD and aMCI groups to examine if there were group differences on TMT performance. Of note, the proportion of females was higher in both the Ctrl and AD groups (in both groups, $\chi^2 \geq 11.11$, $p < .001$). Yet, as gender did not impact any TMT indices, we collapsed it across all of the following analyses.

As can be seen in Table 4, a series of one-way ANCOVAs revealed that TMT scores differed between the Ctrl group and both the AD and aMCI groups. We provided a set of concrete cutoff scores to indicate cognitive deficit in the next section. However, the performance between the Ctrl and AD groups on TMT-A did not significantly differ.

Receiver Operating Characteristic Analyses

In order to provide cutoff scores useful in the clinical neuropsychological setting, receiver operating characteristic (ROC) analyses were performed on 30 healthy participants matched to 30 AD and 30 aMCI patients, so as to control for the effect of both age and education. The main sample characteristics are detailed in Table 5.

The ROC analyses (Table 6) resulted in robust cutoff scores between Ctrl and AD groups, and somewhat less robust but still acceptably accurate cutoff scores between the Ctrl and aMCI groups. More precisely, the TMT-B appeared to be the most accurate indicator (area under curve, AUC = 0.93) for differentiating between the Ctrl and AD groups and TMT-B/A for differentiating (AUC = 0.79) between Ctrl and aMCI. Thus, TMT performance appeared to be a reliable indicator of AD and aMCI diagnoses with respect to healthy controls. Moreover, TMT-A appears to be an acceptably accurate indicator for differentiating between the AD and aMCI groups (AUC = 0.80). In sum, TMT performance allowed for reliable distinction between healthy controls and both AD and/or aMCI pathologies.

Table 4. Effects of clinical groups on age, education, RAVLT and TMT scores

Variables	Ctrl ($n = 421$)		AD ($n = 36$)		aMCI ($n = 90$)		*F*	*df*1, *df*2
	M	*SD*	*M*	*SD*	*M*	*SD*		
Age	47.69[a]	17.28	75.06[b]	7.16	72.72[b]	9.28	132.13***	2, 548
Education	14.18[a]	3.31	11.78[a,b]	2.80	13.87[b]	3.28	9.00***	2, 548
RAVLT	53.82[a]	10.24	20.75[b]	7.56	30.85[c]	8.78	352.02***	2, 548
TMT-A	37.26[a]	2.12	89.14[b]	7.73	38.20[a]	4.98	23.79***	2, 545
TMT-B	95.90[a]	4.40	302.44[b]	16.07	183.29[c]	10.35	36.24***	2, 545
TMT-B − A	58.64[a]	4.03	213.30[b]	14.70	145.09[c]	9.47	48.22***	2, 545
TMT-B/A	2.67[a]	0.08	4.11[b]	0.30	4.39[b]	0.19	26.64***	2, 545

Notes: *M* = mean (expected value for ANCOVA); *SD* = standard deviation (expected value for ANCOVA); *F* = effect of clinical group; *df* = degrees of freedom. Ctrl = Healthy Control; AD = Alzheimer's disease; aMCI = amnestic mild cognitive impairment; RAVLT = Rey Auditory Verbal Learning Test. One-way ANOVA were performed for age, education and RAVLT, whereas one-way ANCOVA (with age, education, and RAVLT as covariates) were performed for TMT scores. Different superscripts in one line indicate significantly differing means in *post hoc* analyses.
***$p < .001$.

Table 5. Main characteristics and TMT performance of matched Ctrl, AD, and aMCI samples

Variable	Ctrl ($n = 30$)		AD ($n = 30$)		aMCI ($n = 30$)		*F*	*df*1, *df*2	*p*-value
	M	*SD*	*M*	*SD*	*M*	*SD*			
Age	73.43	3.44	73.83	5.31	73.90	5.12	0.09	2, 89	.92
Education	12.53	1.17	12.03	2.92	11.93	1.02	0.85	2, 89	.43
TMT-B	93.83	43.25	370.33	149.83	243.27	162.08			
TMT-B − A	52.17	34.65	256.03	144.125	187.80	150.35			
TMT-B/A	2.30	0.65	4.79	2.94	5.47	3.18			

Notes: *M* = mean; *SD* = standard deviation; *F* = effect of clinical group; *df* = degrees of freedom; Ctrl = Healthy Control; AD = Alzheimer's disease; aMC = amnestic mild cognitive impairment; TMT = Trail Making Test.

Table 6. Optimal cutoff scores among Ctrl, AD, and aMCI groups and their overall accuracy in respect of TMT-B, TMT-B − A, and TMT-B/A scores

	Ctrl versus AD	Ctrl versus aMCI	AD versus aMCI
TMT-B			
Optimal cutoff	135/136	111/112	×
Overall accuracy	0.92	0.80	0.50
TMT-B − A			
Optimal cutoff	96/97	75/76	×
Overall accuracy	0.90	0.82	0.50
TMT-B/A			
Optimal cutoff	2.78/2.79	3.48/3.49	×
Overall accuracy	0.85	0.87	0.52

Notes: Ctrl = Healthy Control; AD = Alzheimer's disease; aMCI = amnestic mild cognitive impairment; TMT = Trail Making Test.

Of note, the discriminant ability of all TMT performance scores between the AD and aMCI groups was very low (overall accuracy ≤ 0.52). Therefore, although TMT performance allowed for reliable distinction between healthy controls and both AD and/or aMCI pathologies, it did not allow for the same between AD and aMCI.

Discussion

Our results demonstrated that the influence of demographic variables on TMT-A and TMT-B performance using a Czech version of the TMT is comparable with prior research (Amodio et al., 2002; Drane et al., 2002; Hester et al., 2005; Horton & Roberts, 2001; Sherrill-Pattison, Donders, & Thompson, 2000; Zalonis et al., 2008). Demographic variables, specifically age and education, accounted for a significant amount of variance in both TMT-A and TMT-B performance in our healthy, adult sample from the Czech population. As we have shown, younger participants needed less time to complete the test than older participants, regardless of education level.

Analysis of the TMT-derived scores showed that the TMT-B/A ratio score was insensitive to age, a finding supported by other studies (Horton & Roberts, 2001; Lamberty et al., 1994). Moreover, education correlated with the TMT-B/A ratio and TMT-B − A difference scores, but the practical influence of education on TMT performance was rather low. Gender did not significantly correlate with TMT scores, a finding in accordance with previous studies (Lucas et al., 2005; Tombaugh, 2004) and partially in agreement with studies emphasizing its minimal but significant influence (Hester et al., 2005; Ivnik, Malec, Smith, & Tangalos, 1996; Lannoo & Vingerhoets, 1997; Mitrushina et al., 2005; Yeudall, Reddon, Gill, & Stefanyk, 1987).

This study provided age-corrected TMT and regression-based normative values (Appendix) for the Czech version of the TMT, which uses Arabic numbers and differs slightly from the English version. Fernández and Marcopulos (2008) argued that cultural differences between versions of the TMT make it imperative to use normative data collected for the respective test. In support of this argument, we observed that the Czech TMT-B data (Bezdicek et al., 2012) are inconsistent with the most comprehensive TMT data generated to date (Tombaugh, 2004). More precisely, it appeared that the Czech TMT-B completion times were significantly longer in all but the 65–74-year age group, even when we applied the Sidak correction for multiple comparisons. Given that this cannot be explained by differences in education, our study further stresses the need for culturally adjusted normative TMT data. Using non-culturally adjusted TMT normative data could possibly lead to critical diagnostic errors and inaccurate treatment recommendations.

Mitrushina and colleagues (2005, p. 70) outlined the following seven guidelines for future normative TMT studies: (1) sample size of at least 50 subjects per group, (2) description of sample composition including exclusionary criteria and demographic characteristics, (3–5) presentation of data by age intervals, IQ, and education levels, and (6) presentation of mean and standard deviation for total TMT-A and TMT-B time in seconds. Normative data from the present study did not precisely meet all of the guidelines suggested by Mitrushina, particularly with regard to the sample size and the presentation of data by IQ. Our sample consisted of a small number of individuals over 75 years of age ($n = 9$) with an education level of ≤ 12 years. Normative data for individuals over 85 years of age and IQ values for all age groups were unavailable in our study. However, these limitations can be compensated by the fact that this is the first broad, normative study on the Czech version of the TMT. The educational data were complete (Criterion 5), which is a variable that is highly correlated with IQ (Tombaugh, 2004). Considering Criterion 1, our study consisted of only two age groups (20–24 and ≥ 75 years) that fail to meet the criterion. Individuals of advanced age (in the present study, ≥ 75 years) with less education required more time to complete TMT-B; therefore, cautious interpretation of the results is warranted due to the restricted sample size. We also

must take into account that the general level of education in the present study was high ($M = 14.18$ years, $SD = 3.31$), although the level of education in the Czech population is generally high in comparison with other countries. This may be related to the Czech educational system where this mean value depicts a general level of education (for comparison, see the levels in clinical groups which were quite similar, means ranging from 13.03 to 14.16 years, with the exception of AD, $M = 11.78$ years, $SD = 2.80$).

In exploring the diagnostic potential of the TMT indices, all but the TMT-A scores appeared to differentiate between controls and patients. The TMT-B time score and TMT-B − A differences scores were found to accurately differentiate both Ctrl and AD and Ctrl and aMCI, respectively. Based on ROC analyses, these scores thus can be viewed as clinical indicators of both AD and aMCI. Neither the TMT-A, TMT-B, nor the B − A difference or B/A ratio scores allowed differentiation between patient with probable AD and aMCI. In this respect, our data further highlight the importance of relying on performance profiles, rather than just psychometrically based single indices (Schretlen, Munro, Anthony, & Pearlson, 2003).

In conclusion, the present study provided normative data for the Czech version of standard and derived TMT indices in young and elderly subjects. Our data will allow clinicians to evaluate individual TMT performance with greater precision, and importantly, attenuate the influence of age. Moreover, our data provide the basis for cross-cultural comparison with other normative TMT studies from Western and other cultures. Of importance, TMT indices, with the exception of TMT-A, may be useful clinical indicators in distinguishing AD and aMCI pathologies. A future direction for TMT research should concentrate on the comparison between large normative data across Western and non-Western countries, and study the non-equivalence of normative samples to specify more closely its sources.

Funding

The work was supported by grants from the Czech Ministry of Education (Research Program MSM0021620849, GAUK 251196 118410, and MSMT 1M0517) and by the European Regional Development Fund (Project FNUSA-ICRC CZ.1.05/1.1.00/02.0123).

Conflict of Interest

None declared.

Acknowledgements

The authors are grateful to Aaron Rulseh, MD, and Shawn M. McClintock, PhD, MSCS, for language revision and to anonymous reviewers for their comments on the manuscript.

Appendix

As both age and education were shown to have some effect on TMT performance, the following prediction equations included both variables. With available age and education level information, these equations can be used to determine the expected score of an individual. The corresponding standard error of estimate (SEE), a measure similar to the standard deviation, can be used to determine how far an individual score is from the expected value. Of note, the variance of TMT difference and ratio scores (B − A and B/A) is less explained by age and education (adjusted R^2 of 10% and 4%, respectively) than the variance of basic TMT scores (TMT-A and TMT-B; adjusted R^2 of 22% and 19%, respectively).

$$\text{TMT-A} = 27.040 + 0.420 \times \text{Age} - 0.185 \times \text{Education (SEE} = 11.465);\ R^2 = .22,\ F(2, 419) = 61.50,\ p < .001;$$
$$\text{TMT-B} = 95.280 + 0.272 \times \text{Age} - 0.316 \times \text{Education (SEE} = 41.314);\ R^2 = .19,\ F(2, 419) = 49.68,\ p < .001;$$
$$\text{TMT-B} - \text{A} = 477.529 + 0.124 \times \text{Age} - 0.278 \times \text{Education (SEE} = 36.521);\ R^2 = .10,\ F(2, 419) = 23.38,\ p < .001;$$
$$\text{TMT-B/A} = 3.524 - 0.090 \times \text{Age} - 0.207 \times \text{Education (SEE} = 0.935);\ R^2 = .04,\ F(2, 419) = 10.34,\ p < .001.$$

References

Abe, M., Suzuki, K., Okada, K., Miura, R., Fujii, T., Etsurou, M., et al. (2004). Normative data on tests for frontal lobe functions: Trail Making Test, Verbal fluency, Wisconsin Card Sorting Test (Keio version). *No To Shinkei*, *56*, 567–574.

Amieva, H., Lafont, S., Auriacombe, S., Rainville, C., Orgogozo, J. M., Dartigues, J. F., et al. (1998). Analysis of error types in the trial making test evidences an inhibitory deficit in dementia of the Alzheimer type. *Journal of Clinical and Experimental Neuropsychology*, *20*, 280–285. doi:10.1076/jcen.20.2.280.1161

Amodio, P., Wenin, H., Del Piccolo, F., Mapelli, D., Montagnese, S., Pellegrini, A., et al. (2002). Variability of trail making test, symbol digit test and line trait test in normal people. A normative study taking into account age-dependent decline and sociobiological variables. *Aging Clinical and Experimental Research*, *14*, 117–131.

Arbuthnott, K., & Frank, J. (2000). Trail making test, part B as a measure of executive control: Validation using a set-switching paradigm. *Journal of Clinical and Experimental Neuropsychology*, *22*, 518–528. doi:10.1076/1380-3395(200008)22:4;1-0;FT518

Barr, A., Benedict, R., Tune, L., & Brandt, J. (1992). Neuropsychological differentiation of Alzheimer's disease from vascular dementia. *International Journal of Geriatric Psychiatry*, *7*, 621–627. DOI: 10.1002/gps.930070903

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory-II*. San Antonio, TX: Harcourt Brace.

Bezdicek, O., Preiss, M., Motak, L., Axelrod, B. N., Nikolai, T., & Ruzicka, E. (2012). Czech version of the Trail Making Test normative data is not equivalent to the American version. *Fortieth Annual Meeting of the International Neuropsychological Society*, Montreal, 2012 (abstract 21).

Bornstein, R. A., & Suga, L. G. (1988). Education level and neuropsychological performance in healthy elderly subjects. *Developmental Neuropsychology*, *4*, 17–22. doi:10.1080/87565648809540386

Cahn, D. A., Salmon, D. P., Butters, N., Wiederholt, W. C., Corey-Bloom, J., Edelstein, S. L., et al. (1995). Detection of dementia of the Alzheimer type in a population-based sample: Neuropsychological test performance. *Journal of the International Neuropsychology Society*, *1*, 252–260. doi:10.1017/S1355617700000242

Chen, P., Ratcliff, G., Belle, S. H., Cauley, J. A., DeKosky, S. T., & Ganguli, M. (2000). Cognitive tests that best discriminate between presymptomatic AD and those who remain nondemented. *Neurology*, *55*, 1847–1853.

Drane, D. L., Yuspeh, R. L., Huthwaite, J. S., & Klingler, L. K. (2002). Demographic characteristics and normative observations for derived-trail making test indices. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*, *15*, 39–43.

Fernández, 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*, 239–246. doi:10.1111/j.1467-9450.2008.00637.x

Golden, C. J., Osmon, D. C., Moses, J. A., & Berg, R. A. (1981). *Interpretation of the Halstead-Reitan Neuropsychological Test Battery*. New York: Grune & Stratton.

Haug, T., Sorteberg, A., Sorteberg, W., Lindegaard, K. F., Lundar, T., & Finset, A. (2007). Cognitive outcome after aneurysmal subarachnoid hemorrhage: time course of recovery and relationship to clinical, radiological, and management parameters. *Neurosurgery*, *60*, 649–656.

Heaton, R. K., Nelson, L. M., Thompson, D. S., Burks, J. S., & Franklin, G. M. (1985). Neuropsychological findings in relapsing-remitting and chronic-progressive multiple sclerosis. *Journal of Consulting and Clinical Psychology*, *53*, 103–110. doi:10.1037//0022-006X.53.1.103

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*, 45–54. doi:10.1080/13854040490524137

Hort, J., Laczó, J., Vyhnálek, M., Bojar, M., Bures, J., & Vlcek, K. (2007). Spatial navigation deficit in amnestic mild cognitive impairment. *Proceedings of the National Academy of Sciences of the USA*, *104*, 4042–4047. doi:10.1073/pnas.0611314104

Horton, A. M., Jr, & Roberts, C. (2001). Derived trail making test indices in a sample of substance abusers: Demographic effects. *International Journal of Neuroscience*, *111*, 123–132. doi:10.3109/00207450108986557

Ivnik, R. J., Malec, J. F., Smith, G. E., & Tangalos, E. G. (1996). Neuropsychological tests' norms above age 55: COWAT, BNT, MAE Token, WRAT-R Reading, AMNART, STROOP, TMT, and JLO. *The Clinical Neuropsychologist*, *10*, 262–278. doi:10.1080/13854049608406689

Jacobson, S. C., Blanchard, M., Connolly, C. C., Cannon, M., & Garavan, H. (2011). An fMRI investigation of a novel analogue to the Trail-Making Test. *Brain and Cognition*, *77*, 60–70. doi:10.1016/j.bandc.2011.06.001

Kortte, K. B., Horner, M. D., & Windham, W. K. (2002). The trail making test, part B: Cognitive flexibility or ability to maintain set? *Applied Neuropsychology*, *9*, 106–109. doi:10.1207/S15324826AN0902_5

Laczó, J., Andel, R., Vlček, K., Macoška, V., Vyhnálek, M., Tolar, M., et al. (2011). Spatial Navigation and APOE in Amnestic Mild Cognitive Impairment. *Neurodegenerative Diseases*, *8*, 169–177. doi:10.1159/000321581

Laczó, J., Vlček, K., Vyhnálek, M., Vajnerová, O., Ort, M., Holmerová, I., et al. (2009). Spatial navigation testing discriminates two types of amnestic mild cognitive impairment. *Behavioral Brain Research*, *202*, 252–259. doi:10.1016/j.bbr.2009.03.041

Lafleche, G., & Albert, M. S. (1995). Executive function deficits in mild Alzheimer's disease. *Neuropsychology*, *9*, 313–320. doi:10.1037//0894-4105.9.3.313

Lamberty, G., & Axelrod, B. (2006). Derived adult Trail Making Test indices. In A. Poreh (Ed.), *The quantified process approach to neuropsychological assessment* (pp. 161–173). New York and London: Taylor & Francis.

Lamberty, G. J., Putnam, S. H., Chatel, D. M., Bielauskas, L. A., & Adams, K. M. (1994). Derived Trail Making Test indices: A preliminary report. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*, *7*, 230–234.

Lannoo, E., & Vingerhoets, G. (1997). Flemish normative data on common neuropsychological tests: Influence of age, education, and gender. *Psychologica Belgica*, *37*, 141–155.

Lucas, J. A., Ivnik, R. J., Smith, G. E., Ferman, T. J., Willis, F. B., Petersen, R. C., et al. (2005). Mayo's Older African Americans Normative Studies: Norms for Boston Naming Test, Controlled Oral Word Association, Category Fluency, Animal Naming, Token Test, Wrat-3 Reading, Trail Making Test, Stroop Test, and Judgment of Line Orientation. *The Clinical Neuropsychologist*, *19*, 243–269. doi:10.1080/13854040590945337

Manly, J. (2008). Critical issues in cultural neuropsychology: Profit from diversity. *Neuropsychology Review*, *18*, 179–183. doi:10.1007/s11065-008-9068-8

Martin, T. A., Hoffman, N. M., & Donders, J. (2003). Clinical utility of the trail making test ratio score. *Applied Neuropsychology*, *10*, 163–169. doi:10.1207/S15324826AN1003_05

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, *34*, 939–944.

Mitrushina, M., Boone, K. B., Razani, J., & D'Elia, L. F. (2005). *Handbook of normative data for neuropsychological assessment* (2nd ed.). New York: Oxford University Press.

Ogden, J. A., Mee, E. W., & Henning, M. (1993). A prospective study of impairment of cognition and memory and recovery after subarachnoid hemorrhage. *Neurosurgery*, *33*, 572–586.

Orbo, M., Waterloo, K., Egge, A., Isaksen, J., Ingebrigtsen, T., & Romner, B. (2008). Predictors for cognitive impairment one year after surgery for aneurysmal subarachnoid hemorrhage. *Journal of Neurology*, *255*, 1770–1776.

Petersen, R. C. (2000). Mild cognitive impairment: Transition between aging and Alzheimer's disease. *Neurologia*, *15*, 93–101.

Pillon, B., Gouider-Khouja, N., Deweer, B., Vidailhet, M., Malapani, C., Dubois, B., et al. (1995). Neuropsychological pattern of striatonigral degeneration: Comparison with Parkinson's disease and progressive supranuclear palsy. *Journal of Neurology, Neurosurgery and Psychiatry*, *58*, 174–179. doi:10.1136/jnnp.58.2.174

Preiss, M., & Preiss, J. (2006). *Test cesty [Trail making test. Manual in Czech]*. Bratislava: Psychodiagnostika.

Preiss, M., Rodriguez, M., Kawaciukowa, R., & Laing, H. (2007). *Neuropsychologická baterie Psychiatrického centra Praha [Neuropsychological battery of Prague Psychiatric Center. Clinical Assessment of Basic Cognitive Functions]* (2nd ed.). Praha: Psychiatrické centrum Praha.

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*, 33–65. doi:10.1016/j.acn.2004.02.005

Rasmusson, D. X., Zonderman, A. B., Kawas, C., & Resnick, S. M. (1998). Effects of age and dementia on the trail making test. *The Clinical Neuropsychologist*, *12*, 169–178.

Reitan, R. M., & Wolfson, D. (1993). *The Halstead-Reitan Neuropsychological Test Battery* (2nd ed.). Tucson, AZ: Neuropsychology Press.

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*, 281–288.

Schmidt, M. (2004). *Rey Auditory Verbal Learning Test. A handbook* (2nd ed.). Los Angeles, CA: Western Psychological Services.

Schretlen, D. J., Munro, C. A., Anthony, J. C., & Pearlson, G. D. (2003). Examining the range of normal intraindividual variability in neuropsychological test performance. *Journal of the International Neuropsychological Society*, *9*, 864–870. doi:10.1017/S1355617703960061

Sheikh, J. I., & Yesavage, J. A. (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. *The Clinical Gerontologist*, *5*, 165–173.

Sherrill-Pattison, S., Donders, J., & Thompson, E. (2000). Influence of demographic variables on neuropsychological test performance after traumatic brain injury. *The Clinical Neuropsychologist*, *14*, 496–503. doi:10.1076/clin.14.4.496.7196

Soukup, V. M., Ingram, F., Grady, J. J., & Schiess, M. C. (1998). Trail Making Test: Issues in normative data selection. *Applied Neuropsychology*, *5*, 65–73. doi:10.1207/s15324826an0502_2

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.

Stuss, D. T., Bisschop, S. M., Alexander, M. P., Levine, B., Katz, D., & Izukawa, D. (2001). The Trail Making Test: A study in focal lesion patients. *Psychological Assessment*, *13*, 230–239. doi:10.1037//1040-3590.13.2.230

Tombaugh, T. N. (2004). Trail Making Test A and B: Normative data stratified by age and education. *Archives of Clinical Neuropsychology*, *19*, 203–214. doi:10.1016/S0887-6177(03)00039-8

U.S. War Department, Adjutant General's Office. (1944). *Army Individual Test Battery. Manual of directions and scoring*. Washington, DC: Author.

Uttl, B. (2002). North American Adult Reading Test: Age norms, reliability, and validity. *Journal of Clinical and Experimental Neuropsychology*, *24 (8)*, 1123–1137.

van de Vijver, F. J. R., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *Revue européenne de psychologie appliquée*, *54*, 119–135.

Yeudall, L. T., Reddon, J. R., Gill, D. M., & Stefanyk, W. O. (1987). Normative data for the Halstead-Reitan neuropsychological tests stratified by age and sex. *Journal of Clinical Psychology*, *43*, 346–367. doi:10.1002/1097-4679(198705)43:3<346::AID-JCLP2270430308>3.0.CO;2-Q

Zalonis, I., Kararizou, E., Triantafyllou, N. I., Kapaki, E., Papageorgiou, S., Sgouropoulos, P., et al. (2008). A normative study of the trail making test A and B in Greek adults. *The Clinical Neuropsychologist*, *22*, 842–850. doi:10.1080/13854040701629301