



Trail Making Test errors in normal aging, mild cognitive impairment, and dementia

Lee Ashendorf^{a,b}, Angela L. Jefferson^b, Maureen K. O'Connor^{a,b},
Christine Chaisson^{b,c}, Robert C. Green^{b,c}, Robert A. Stern^{b,*}

^a Edith Nourse Rogers Memorial Veterans Hospital, Bedford, MA, United States

^b Alzheimer's Disease Clinical and Research Program, Department of Neurology,
Boston University School of Medicine, Boston, MA, United States

^c Boston University School of Public Health, Boston, MA, United States

Accepted 23 November 2007

Abstract

The objective of the present study was to provide normative data for Trail Making Test (TMT) time to completion and performance errors among cognitively normal older adults, and to examine TMT error rates in conjunction with time scores for pre-clinical and clinical Alzheimer's disease (AD) diagnostic decision-making. A sample of 526 individuals was classified into three diagnostic groups (normal controls, $N=269$; mild cognitive impairment, MCI, $N=200$; AD, $N=57$) by a multidisciplinary consensus conference. Results indicated that performance differed among the three groups for TMT A and B time scores as well as TMT B error rate. Diagnostic classification accuracy (i.e., sensitivity, specificity, and positive and negative predictive powers) is described for various combinations of the diagnostic groups. The findings show that TMT B time and errors are independently meaningful scores, and both therefore have clinical utility in assessing individuals referred for dementia evaluations.

© 2007 National Academy of Neuropsychology. Published by Elsevier Ltd. All rights reserved.

Keywords: Trail Making Test; Errors; Geriatrics; MCI; AD; Executive functioning

The Trail Making Test (TMT; Army Individual Test Battery, 1944) is among the most commonly used neuropsychological tests in clinical practice (Rabin, Barr, & Burton, 2005), in part because it is among the instruments most sensitive to brain damage (Reitan & Wolfson, 1994). The cognitive alternation required by Part B reflects executive functioning, although other cognitive abilities, such as psychomotor speed and visual scanning, are also required for the successful completion of the test (Lezak, Howieson, & Loring, 2004). Despite early research suggesting that TMT performance is independent of age (Boll & Reitan, 1973), newer recent studies using larger and more representative samples have reported declining performance with increasing age (e.g., Ivnik, Malec, Smith, Tangalos, & Petersen, 1996; Kennedy, 1981; Rasmusson, Zonderman, Kawas, & Resnick, 1998).

TMT normative data have been reported in many studies using a wide variety of sample demographics (e.g., Tombaugh, 2004). Larger-scale efforts examining performance include the Mayo Older Adults Normative Study (Ivnik et al., 1996) and the Revised Comprehensive Norms for an Expanded Halstead-Reitan Battery (Heaton,

* Corresponding author at: Alzheimer's Disease Clinical and Research Program, Boston University School of Medicine, Robinson Complex, Suite 7800, 715 Albany Street, Boston, MA 02118-2526, United States. Tel.: +1 617 638 5678; fax: +1 617 414 1197.

E-mail address: robstern@bu.edu (R. A. Stern).

Miller, Taylor, & Grant, 2004). Most published normative data for the TMT capture *time to completion*. Additional research has examined the clinical value of a ratio (B/A; Giovagnoli et al., 1996) or difference score (B–A; Arbuthnott & Frank, 2000) between the two TMT conditions. The ratio score in particular has been examined as a symptom validity indicator (O'Bryant, Hilsabeck, Fisher, & McCaffrey, 2003; Ruffolo, Guilmette, & Willis, 2000). *Frequency of errors*, while often recorded and reported clinically, has not been empirically evaluated in prior TMT normative studies. Some investigators have pointed out that the examiner's correction of errors adds additional time to the total score, thus accounting for difficulties reflected in the number of errors (Stuss et al., 2001).

The TMT error rate is difficult to interpret in isolation, particularly because errors are common among cognitively normal adults. In one study, 34.7% of control participants committed at least one error on TMT Part B (Ruffolo et al., 2000). Stuss et al. (2001) examined error rates in brain-lesioned patients and lesion location (i.e., frontal vs. nonfrontal). For a cutoff of >1 error, they found high positive predictive power (i.e., high error rate suggests presence of frontal lesion vs. other lesion) but poor negative predictive power (i.e., less than 2 errors does not necessarily implicate a nonfrontal lesion). Some have found that head-injured individuals are more likely to commit errors and less likely to correct them without prompting (e.g., Armitage, 1946), though other studies failed to distinguish head-injured participants from controls based on errors (Klusman, Cripe, & Dodrill, 1989).

Three different TMT error types have been identified (Mahurin et al., 2006; McCaffrey, Krahula, & Heimberg, 1989), including: (1) sequential or tracking errors (proceeding to the incorrect number or letter on Part A or B); (2) perseverative errors (failure to proceed from a number to a letter, or vice versa, on Part B); and (3) proximity errors (proceeding to an incorrect nearby circle on Part A or B). Mahurin et al. (2006) examined TMT errors, time to completion, and other neuropsychological measures among patients with schizophrenia and depression as well as healthy controls. While healthy controls completed both parts faster than the patients with depression (who completed each part faster than the patients with schizophrenia), elevated error rates were only observed in the schizophrenia group. Regression analyses implicated visual search in TMT A time performance, processing speed and mental tracking in TMT B time score, and mental tracking ability in TMT B error rate. However, a non-standard administration was used in this study, as errors were not corrected by the examiner.

The purpose of the present investigation was twofold: (1) to provide normative data for TMT total errors in cognitively normal older participants, and (2) to determine the diagnostic accuracy of TMT B errors in a sample consisting of cognitively normal elderly and individuals meeting diagnostic criteria for mild cognitive impairment (MCI) or Alzheimer's disease (AD).

1. Method

1.1. Participants

All study participants were enrolled in the Boston University Alzheimer's Disease Core Center (BU-ADCC) patient/control registry, which longitudinally follows older adults with and without memory problems. As previously reported (Jefferson et al., 2006; Jefferson, Wong, Gracer, Green, & Stern, 2007), participants undergo an annual neurological examination and neuropsychological evaluation. Inclusion criteria require that participants be community-dwelling and English-speaking with adequate hearing and visual acuity to participate in the examinations. All participants must have a study partner (to provide collateral information about functioning). Exclusion criteria include a history of major psychiatric illness (e.g., schizophrenia or bipolar disorder) or other significant central nervous system disorder (e.g., stroke, head injury with loss of consciousness). A multidisciplinary diagnostic review team consisting of neurologists, neuropsychologists, and research staff review evaluation results in order to reach a consensus diagnosis for each participant.

The current study was based on 622 participants evaluated between 2002 and 2006. Of these participants, 96 were unable to complete the TMT due to cognitive impairment ($N=74$), physical limitations ($N=8$), or refusal ($N=14$). These included individuals who could not attempt the task as well as individuals who discontinued during the task

norms), and (2) a Clinical Dementia Rating (CDR; Morris, 1993) global rating of 0. One hundred twelve individuals met criteria for “possible MCI.” Diagnostic criteria for this group required functional independence, as reported in the clinical interview, and objective cognitive impairment in one or more domains (i.e., neuropsychological performance falling ≥ 1.5 standard deviations below available normative data for at least one test in that domain). Eighty-eight individuals met criteria for “probable MCI,” which is similar to “possible MCI” but with the additional criterion of having a cognitive complaint by the participant and/or study partner. These participants (44% of total MCI sample) also met the Petersen (2004) workgroup criteria for MCI. All MCI participants (possible and probable) had a CDR = 0.5. The AD sample included 57 individuals meeting NINCDS-ADRDA criteria for probable ($N = 25$) or possible ($N = 32$) AD (McKhann et al., 1984) with CDR scores ≥ 1.0 .

Descriptive statistics for all demographic variables are provided in Table 1. The entire sample of 526 participants comprised 204 men and 322 women with a mean age of 73.2 (S.D. = 8.7) and education of 15.4 years (S.D. = 3.1). Non-Hispanic Caucasian individuals accounted for 77% of the sample, while African-American participants comprised the remaining 23%.

If the TMT A or B time to completion was >1.5 S.D. below the mean using Tombaugh's (2004) normative data, then the score was considered impaired. The TMT was used in the consensus conference for diagnostic decision-making (e.g., impairment on the TMT might contribute to a participant being labeled “MCI probable”). To avoid any tautological concerns, 14 participants with MCI were excluded from the study because their cognitive impairment was detected solely by TMT A and/or B.

1.2. Measures

Each participant underwent a neuropsychological protocol assessing global cognition, language, verbal and visuospatial memory, attention and processing speed, executive functioning, visuospatial skills, and motor skills (Jefferson et al., 2006, 2007). For the purpose of the present study, only the TMT, Mini Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975), and Wide Range Achievement Test-3rd Edition (WRAT-3; Wilkinson, 1993) Reading subtest are reported. The MMSE was included to provide a general estimate of cognitive functioning, while WRAT-3 Reading was included as an estimate of premorbid ability as well as literacy (Manly, Touradji, Tang, & Stern, 2003) to confirm that participants had adequate language skills to complete TMT B.

In administering the TMT, errors were defined as any incorrect line that reaches its target. As soon as an error was committed, the examinee was told only that he or she had made an error, and was directed to return to the last correct target. If an incorrect line was begun but not completed, the segment was not scored as an error. While a 5-min time limit is typically imposed on TMT B performance, participants in this study continued to completion; only individuals who were able to complete the task were included in this study.

Table 1
Descriptive statistics for the current sample

| | Control | MCI | AD | Total |
|--------------------------------|--------------|--------------|--------------|--------------|
| <i>N</i> | 269 | 200 | 57 | 526 |
| Age (years) | 72.4 (8.5) | 72.5 (8.6) | 79.7 (7.1) | 73.2 (8.7) |
| Education (years) | 16.1 (2.8) | 14.6 (3.3) | 14.6 (3.2) | 15.4 (3.1) |
| Sex (%) female | 66.9% | 59.0% | 42.1% | 61.2% |
| Ethnicity (%) African-American | 15.2% | 35.5% | 15.8% | 23.0% |
| Mini mental state exam | 29.2 (1.3) | 28.1 (1.9) | 24.2 (3.6) | 28.2 (2.4) |
| WRAT-3 reading scaled score | 111.3 (11.5) | 104.8 (13.8) | 106.1 (8.9) | 108.2 (12.6) |
| TMT-A time in seconds | 35.7 (12.8) | 48.1 (22.2) | 67.1 (31.0) | 43.8 (21.7) |
| TMT-A errors | 0.12 (0.35) | 0.18 (0.46) | 0.16 (0.45) | 0.15 (0.41) |
| TMT-B time in seconds | 81.5 (36.1) | 136.0 (78.0) | 190.8 (81.6) | 114.1 (71.0) |

1.3. Procedures

The local Institutional Review Board approved data collection for this study. All participants provided written informed consent prior to testing. Cognitive evaluations were conducted in a single session for all participants by trained research assistants.

1.4. Data analysis

The sample was classified by diagnostic category (i.e., Control, MCI, or AD) and between-group comparisons were made using analyses of variance (ANOVAs) or, when appropriate, analyses of covariance (ANCOVAs) for age, education, TMT A and B time to completion, and MMSE and WRAT-3 Reading standard scores. Pairwise multiple comparisons were conducted using Tukey HSD tests. Mann–Whitney tests were used to evaluate differences in error rate by age, sex, and education. The number of errors committed on TMT B was categorized into groups (0 errors, 1–2 errors, and 3 or more errors) and diagnostic group comparisons were made using a Kruskal–Wallis test.

Among the controls, cumulative percentages were calculated for errors made on TMT A and B. For TMT A and B time to completion, descriptive categories (i.e., superior, above average, etc.) were identified using percentile rankings within each of several demographically derived subgroups based on age (55–74, 75–98) and education (<16 years, ≥16 years). Finally, for all participants, classification accuracy statistics (i.e., sensitivity, specificity, and positive and negative predictive powers) were calculated using combinations of error and time to completion scores. Sensitivity refers to the proportion of the “disordered” group correctly identified by the test; specificity is the correctly classified proportion of the non-target sample. Positive predictive power (PPP) represents the rate at which a positive finding on the test accurately predicts diagnostic group membership, while negative predictive power (NPP) is the analogous statistic for negative findings on the test.

2. Results

Diagnostic between-group comparisons revealed a main effect for age, $F_{(2,523)} = 18.7, p < .001$. Post hoc comparisons revealed that the AD sample was significantly older than the Control and MCI groups, $p < .001$. No differences were observed between the Control and MCI groups. There was a main effect for MMSE score, $F_{(2,523)} = 168.3, p < .001$, and post-hoc comparisons revealed the three diagnostic groups differed in the expected direction (Controls > MCI > AD; $p < .001$). A main effect was also detected for WRAT-3 Reading scaled score, $F_{(2,495)} = 16.0, p < .001$. The Control group had a higher mean scaled score than the MCI ($p < .001$) and AD ($p = .020$) group; MCI and AD groups did not differ on this variable.

When demographic variables (age, education, and sex) were related to TMT time to completion scores among Control participants, age ($r = .451, p < .001$) and education ($r = -.216, p < .001$) were significantly correlated with TMT A time to completion. These variables were also correlated with TMT B time to completion (age: $r = .458, p < .001$; education: $r = -.329, p < .001$). Time to completion did not differ by sex for either TMT A or TMT B.

2.1. “Normative” data for TMT scores

Based on these demographic differences, normative data for TMT A and B time to completion score were presented for 4 groups categorized by age (55–74, 75–98) and education (<16 years and ≥16 years; see Tables 2 and 3).

In the normative (i.e., Control) sample, demographic variables were not related to TMT A error score. Though TMT B error rate did not vary by sex, it was significantly related to both age ($r = .170, p = .006$) and education ($r = -.268, p < .001$).

Participants who made no errors on TMT B had higher MMSE scores ($M = 29.3, S.D. = 1.1$) than those who committed at least one error ($M = 28.8, S.D. = 1.5$), $t_{(118.8)} = 2.86, p = .005$. No difference was found on WRAT-3 Reading scaled score between individuals with no errors ($M = 111.8, S.D. = 12.0$) and those with at least one error ($M = 110.0$,

Table 2

Normative (control) data for time to completion for Trail Making Test Part A, in seconds, by age and education

| Age | 55–74 | 75–98 | 55–74 | 75–98 | Performance descriptive ^a |
|----------------|---|---|---|---|---|
| Education | <16 years | <16 years | ≥16 years | ≥16 years | |
| Sample size | <i>N</i> = 46 | <i>N</i> = 50 | <i>N</i> = 106 | <i>N</i> = 67 | |
| Mean | <i>M</i> = 33.2 | <i>M</i> = 44.8 | <i>M</i> = 29.7 | <i>M</i> = 40.3 | |
| S.D. | S.D. = 13.1 | S.D. = 13.4 | S.D. = 7.8 | S.D. = 13.2 | |
| Time score (s) | <15 15–17 18–24 25–42 43–51 52–74 >74 | <25 25–28 29–33 34–51 52–64 65–78 >78 | <17 17–21 22–23 24–34 35–40 41–48 >48 | <21 21–27 28–30 31–46 47–57 58–80 >80 | Very superior (≥98th %ile) Superior (91st–97th %ile) High average (75th–90th %ile) Average (25th–74th %ile) Low Average (9th–24th %ile) Borderline (3rd–8th %ile) Extremely Low (≤2nd %ile) |

Note: *M*: mean, S.D.: standard deviation.^a Based on Wechsler (1997) description system.

Table 3

Normative (control) data for time to completion for Trail Making Test Part B, in seconds, by age and education

| Age | 55–74 | 75–98 | 55–74 | 75–98 | Performance descriptive ^a |
|----------------|---|---|---|---|--|
| Education | <16 years | <16 years | ≥16 years | ≥16 years | |
| Sample size | <i>N</i> = 46 | <i>N</i> = 50 | <i>N</i> = 106 | <i>N</i> = 67 | |
| Mean | <i>M</i> = 80.8 | <i>M</i> = 109.7 | <i>M</i> = 62.7 | <i>M</i> = 90.5 | |
| S.D. | S.D. = 30.4 | S.D. = 42.5 | S.D. = 20.6 | S.D. = 37.1 | |
| Time score (s) | <32 32–40 41–56 57–101 102–125 126–154 >154 | <45 45–65 66–77 78–128 129–190 191–205 >205 | <35 35–37 38–46 47–75 76–91 92–106 >106 | <42 42–50 51–64 65–111 112–144 145–192 >192 | Very Superior (≥98th %ile) Superior (91st–97th %ile) High Average (75th – 90th %ile) Average (25th – 74th %ile) Low Average (10th – 24th %ile) Below Average (3rd – 9th %ile) Impaired (≤2nd %ile) |

Note: *M*: mean; S.D.: standard deviation.^a Based on Wechsler (1997) description system.

Table 4

Trail Making Test part A error rates for control participants (*N* = 269)

| # of errors | Cumulative percentage |
|-------------|-----------------------|
| 0 | 100.0% |
| 1 | 11.2% |
| ≥2 | 0.7% |

Note: cumulative percentages represent the percent of individuals who committed at least that many errors (as in Stern & White, 2003).

and at least 16 years) in Table 5. TMT A error totals in this sample ranged from 0 to 2, while TMT B errors ranged from 0 to 6.

2.2. Diagnostic group differences in TMT scores

TMT A and B time to completion scores were compared across the Control, MCI, and AD groups. After controlling for age, education, and (for TMT B) literacy level, main effects were detected for both TMT A ($F_{(4,521)} = 54.91$, $p < .001$) and B ($F_{(5,520)} = 34.38$, $p < .001$) in the expected direction for both variables (i.e., Control < MCI < AD all

Table 5

Trail Making Test part B error rates for control participants ($N = 269$) (reported in cumulative percentages)

| Age | 55–74 | 75–98 | 55–74 | 75–98 |
|-----------|------------------------|------------------------|-------------------------------|------------------------------|
| Education | <16 years ($N = 46$) | <16 years ($N = 50$) | ≥ 16 years ($N = 106$) | ≥ 16 years ($N = 67$) |
| Errors | | | | |
| 0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 1 | 39.1 | 46.0 | 21.7 | 23.9 |
| 2 | 15.2 | 22.0 | 3.8 | 10.4 |
| 3 | 4.3 | 8.0 | 0.9 | 1.5 |
| ≥ 4 | 2.2 | 2.0 | 0.9 | |

Note: cumulative percentages represent the percent of individuals who committed at least that many errors (as in Stern & White, 2003).

Table 6

Frequency of errors and/or impaired time scores on Trail Making Test B, reported by percentage of individuals within each diagnostic group

| Errors/Time Score | Control | MCI | AD | Total |
|--------------------------------|---------|------|------|-------|
| No errors, intact time score | 66.9 | 31.0 | 19.3 | 48.1 |
| Any errors | 29.7 | 59.0 | 71.9 | 45.4 |
| Errors, intact time score | 24.2 | 26.0 | 28.1 | 25.3 |
| Impaired time score | 8.9 | 43.0 | 52.6 | 26.6 |
| Impaired time score, no errors | 3.3 | 10.0 | 8.8 | 6.5 |
| Impaired errors and time score | 5.6 | 33.0 | 43.9 | 20.2 |

Note 1: MCI: mild cognitive impairment; AD: Alzheimer's disease. Note 2: "Impaired" time score refers to any scores ≥ 1.0 standard deviation below the age-and-education-corrected mean according to Tombaugh (2004).

TMT B errors, Kruskal–Wallis $H = 62.1$, $p < .001$. Post hoc tests revealed that Control participants differed from MCI ($p < .001$) and AD ($p < .001$) participants, though the MCI group did not differ from the AD group.

In many individual cases, TMT B time to completion and error scores did not agree (see Table 6). For example, among individuals with AD, most (71.9%) committed TMT B errors, whereas approximately half (52.6%) had impaired time to completion scores. Only 5 AD participants (8.8%) had impaired time to completion scores without any errors, while 28% of the AD sample ($N = 16$) committed errors with normal time to completion scores. Given the frequent disagreement between these scores, it can be concluded that the two variables are independently meaningful.

2.3. Use of TMT in diagnostic group classification

The ability of error score or time to completion score to predict diagnostic classification (i.e., Control, MCI, or AD) was evaluated. This cannot be done using the time cutoff used in the diagnostic process (z -score < -1.5), as all participants in the Control sample had scores above this level by definition. The cutoff for TMT B time score impairment was therefore set to be a z -score < -1.0 , which is often accepted as a more liberal cutoff for impairment (e.g., Heaton et al., 2004). Based on a receiver-operating characteristic (ROC) curve analysis, it was determined that an error score ≥ 1 provided an optimal error cutoff for diagnostic classification of Control vs. AD group membership. Both the time and error cutoff scores were independently effective; however, in spite of a significant correlation between TMT B errors and TMT B time to completion ($r = .572$, $p < .001$), the scores were often not in agreement. Using the error score to predict impairment on the time to completion score resulted in a PPP of only 44.4% (meaning that only 44.4% of individuals who committed errors also had an impaired time score). Using time score to predict impairment on the error score resulted in a PPP of 75.7% (meaning that 75.7% of individuals whose time score was impaired also committed errors).

Because of the discrepancy in diagnostic utility of the two cutoff scores, a combination of the variables was examined by determining classification accuracy statistics for four separate possible combinations of error and time

Table 7

Diagnostic classification accuracy based on time and/or error scores on Trail Making Test Part B

| | Sensitivity | Specificity | PPP | NPP | Hit Rate |
|--|-------------|-------------|-------|-------|----------|
| MCI/AD (<i>N</i> = 257) vs. Control (<i>N</i> = 269) | | | | | |
| Errors ^a | 61.9% | 70.3% | 66.5% | 65.9% | 66.2% |
| Time ^b | 45.1% | 91.1% | 82.9% | 63.5% | 68.6% |
| Errors and time ^c | 35.4% | 94.4% | 85.8% | 60.5% | 75.5% |
| Errors and/or time ^d | 71.6% | 66.9% | 67.4% | 71.1% | 69.2% |
| MCI (<i>N</i> = 200) vs. Control (<i>N</i> = 269) | | | | | |
| Errors ^a | 59.0% | 70.3% | 59.6% | 69.7% | 65.5% |
| Time ^b | 43.0% | 91.1% | 78.2% | 68.2% | 70.6% |
| Errors and time ^c | 33.0% | 94.4% | 81.5% | 65.5% | 76.2% |
| Errors and/or time ^d | 69.0% | 66.9% | 60.8% | 74.4% | 67.8% |
| AD (<i>N</i> = 57) vs. MCI (<i>N</i> = 200) | | | | | |
| Errors ^a | 71.9% | 41.0% | 25.8% | 83.7% | 47.9% |
| Time ^b | 52.6% | 57.0% | 25.9% | 80.9% | 56.0% |
| Errors and time ^c | 43.9% | 67.0% | 27.5% | 80.7% | 53.0% |
| Errors and/or time ^d | 80.7% | 31.0% | 25.0% | 84.9% | 42.0% |

Note: MCI: mild cognitive impairment; AD: Alzheimer's disease; PPP: positive predictive power; NPP: negative predictive power. Identification of impairment within each condition was based on the following contrasts.

^a ≥ 1 error vs. 0 errors.

^b $z \leq -1.0$ vs. $z > -1.0$.

^c ≥ 1 error and $z \leq -1.0$ vs. 0 errors and $z > -1.0$.

^d ≥ 1 error and/or $z \leq -1.0$ vs. 0 errors and $z > -1.0$.

comparisons (see Table 7). Overall, the incorporation of both time and error scores led to comparable or slightly improved classification rates.

3. Discussion

The present study described TMT error rates in a sample of well-characterized, elderly, cognitively intact participants. Data were generated for age-and-education-based norms for TMT Parts A and B time to completion and cumulative percentages of errors committed. In addition, time to completion and error scores on TMT B were explored to determine their incremental utility in diagnostic classification of cognitively healthy older adults and patients with MCI or AD. While many studies have described TMT time performance in aging adults, the present study also provides error frequencies. The additional normative data provided in this study allow the clinician to use the error and time to completion variables in combination rather than in isolation.

In this study, error rate was less susceptible to subtle age differences than was time to completion, which is consistent with past work (Robins-Wahlin, Bäckman, Wahlin, & Winblad, 1996). This finding may reflect the well-known association between processing speed (a major component of successful TMT performance) and age (Salthouse & Fristoe, 1995). Errors may not increase substantially with aging, and may thus be a consistent measure of impairment across the lifespan.

In classifying individuals into diagnostic categories, the use of a combined error-and-time algorithm demonstrated a slightly higher specificity and PPP than the use of TMT B time to completion or error score alone. Also, the presence of errors demonstrated a greater sensitivity in each set of comparisons than the presence of an impaired time to completion score. While these findings did not demonstrate a strong advantage to the use of an algorithm toward diagnostic classification, they did show that errors and time to completion lack strong dependence on each other, and both should therefore be considered in interpreting TMT performance. These findings were not surprising based on work by Mahurin et al. (2006), in which much variance in time to completion was accounted for by variables related

Future studies may wish to examine error subtypes (i.e., sequential vs. perseverative; McCaffrey, Krahula, & Heimberg, 1989) in addition to total error rates, and explore error rates in neurologically and psychiatrically diverse populations. Different degrees of impairment and different degenerative disorders may be associated not only with varying error rates, but also with different types of errors. For instance, examining error rates in other dementias (e.g., vascular dementia, frontotemporal dementia), neurological conditions (e.g., Parkinson's disease, multiple sclerosis), or psychiatric disorders (e.g., schizophrenia, depression) would yield more specific information regarding the underlying neuroanatomic substrates associated with error types.

There are several limitations to the current work. First, the AD participants comprised a smaller sample size as compared with the MCI and control groups. This difference is largely due to cognitive limitations that prevented many AD participants from completing the TMT. Sample size differences significantly influence classification accuracy statistics, such as those used in this study, suggesting that our findings should be interpreted with due consideration of the base rates of AD and/or MCI in the population of comparison. Second, our neuropsychological protocol excludes data for those participants who are unable to establish mental set during TMT B. This exclusion is most commonly found among the AD participants, so those who had the greatest difficulty in overall functioning are not included; this likely reduces the overall performance differences between our MCI and AD groups and contributes to the diminished PPP in MCI versus AD classification. Third, sample characteristics may limit the generalizability of findings to certain populations; for example, the present sample is highly educated, and therefore the findings may not extend to populations of less-educated individuals. Last, all participants whose MCI diagnoses were based solely on TMT A or B performance ($N = 14$) were excluded from the present study. This exclusion was intended a priori to avoid tautological concerns, but it may have resulted in a subtle diagnostic bias or a bias in the overall TMT findings.

In summary, the present study presented geriatric normative data for TMT time to completion and error scores, and compared TMT performance among Control, MCI, and AD participants. The results demonstrate the clinical utility of TMT error scores, in addition to time to completion, in assessing individuals referred for dementia evaluations.

Acknowledgements

This research was supported by P30-AG13846 (Boston University Alzheimer's Disease Core Center), M01-RR00533 (General Clinical Research Centers Program of the National Center for Research Resources, NIH), R03-AG026610 (ALJ), R03-AG027480 (ALJ), K12-HD043444 (ALJ), and K23-AG030962 (ALJ).

References

- 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.
- Armitage, S. G. (1946). An analysis of certain psychological tests used for the evaluation of brain injury. *Psychological Monographs*, 60, 1–48.
- Boll, T. J., & Reitan, R. M. (1973). Effect of age on performance of the Trail Making Test. *Perceptual and Motor Skills*, 36, 691–694.
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12, 189–198.
- Giovagnoli, A. R., Del Pesce, M., Mascheroni, S., Simoncelli, M., Laiacina, M., & Capitani, E. (1996). Trail Making Test: normative values from 287 normal adult controls. *Italian Journal of Neurological Sciences*, 17, 305–309.
- Heaton, R. K., Miller, S. W., Taylor, M. J., & Grant, I. (2004). *Revised comprehensive norms for an expanded halstead-reitan battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults*. Lutz, FL: Psychological Assessment Resources, Inc.
- Ivnik, R. J., Malec, J. F., Smith, G. E., Tangalos, E. G., & Petersen, R. C. (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.
- Jefferson, A. L., Wong, S., Bolen, E., Ozonoff, A., Green, R. C., & Stern, R. A. (2006). Cognitive predictors of HVOT performance differ between individuals with mild cognitive impairment and normal controls. *Archives of Clinical Neuropsychology*, 21, 405–412.
- Jefferson, A. L., Wong, S., Gracer, T. S., Green, R. C., & Stern, R. A. (2007). Geriatric performances on the Boston Naming Test-30 item even version. *Applied Neuropsychology*, 14, 1–9.
- Kennedy, K. J. (1981). Age effects on Trail Making Test performance. *Perceptual and Motor Skills*, 52, 671–675.
- Klusman, L. E., Cripe, L. I., & Dodrill, C. B. (1989). Analysis of errors on the Trail Making Test. *Perceptual and Motor Skills*, 68, 1199–1204.
- Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological Assessment* (4th ed.). New York: Oxford University Press.

- McCaffrey, R. J., Krahula, M. M., & Heimberg, R. G. (1989). An analysis of the significance of performance errors on the Trail Making Test in polysubstance users. *Archives of Clinical Neuropsychology*, 4, 393–398.
- McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. (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.
- Morris, J. C. (1993). The clinical dementia rating (CDR): Current version and scoring rules. *Neurology*, 43, 2412–2414.
- O'Bryant, S. E., Hilsabeck, R. C., Fisher, J. M., & McCaffrey, R. J. (2003). Utility of the Trail Making Test in the assessment of malingering in a sample of mild traumatic brain injury litigants. *The Clinical Neuropsychologist*, 17, 69–74.
- Petersen, R. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, 256, 183–194.
- 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.
- 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. (1994). A selective and critical review of neuropsychological deficits and the frontal lobes. *Neuropsychology Review*, 4, 161–198.
- Robins-Wahlin, T.-B., Bäckman, L., Wahlin, Å., & Winblad, B. (1996). Trail Making Test performance in a community-based sample of healthy very old adults: Effects of age on completion time, but not on accuracy. *Archives of Gerontology and Geriatrics*, 22, 87–102.
- Ruffolo, L. F., Guilmette, T. J., & Willis, W. G. (2000). Comparison of time and error rates on the Trail Making Test among patients with head injuries, experimental malingerers, patients with suspect effort on testing, and normal controls. *The Clinical Neuropsychologist*, 14, 223–230.
- Salthouse, T. A., & Fristoe, N. M. (1995). Process analysis of adult age effects on a computer-administered Trail Making Test. *Neuropsychology*, 9, 518–528.
- Stern, R. A., & White, T. (2003). *Neuropsychological assessment battery*. Lutz, FL: Psychological Assessment Resources, Inc.
- 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.
- Tombaugh, T. N. (2004). Trail Making Test A and B: Normative data stratified by age and education. *Archives of Clinical Neuropsychology*, 19, 203–214.
- Wechsler, D. (1997). *Wechsler adult intelligence scale. III*. San Antonio, TX: The Psychological Corporation.
- Wilkinson, G. S. (1993). *The wide range achievement test: Manual*. Wilmington, DE: Wide Range, Inc.

