BRIEF COMMUNICATION

Limitations of the Trail Making Test Part-B in Assessing Frontal Executive Dysfunction

Edgar Chan,[1] Sarah E. MacPherson,[1,2] Gail Robinson,[1,3] Martha Turner,[1,4] Francesca Lecce,[1,4]
Tim Shallice,[4,5] AND Lisa Cipolotti[1,6]

[1]National Hospital for Neurology and Neurosurgery, London, United Kingdom
[2]Department of Psychology, University of Edinburgh, Scotland, United Kingdom
[3]School of Psychology, University of Queensland, Brisbane, Australia
[4]Institute of Cognitive Neuroscience, University College, London, United Kingdom
[5]International School for Advanced Studies (SISSA), Trieste, Italy
[6]Dipartimento di Psicologia, University of Palermo, Italy

Abstract

Part B of the Trail Making Test (TMT-B) is one of the most widely used neuropsychological tests of "executive" function. A commonly held assumption is that the TMT-B can be used to detect frontal executive dysfunction. However, so far, research evidence has been limited and somewhat inconclusive. In this retrospective study, performance on the TMT-B of 55 patients with known focal frontal lesions, 27 patients with focal non-frontal lesions and 70 healthy controls was compared. Completion time and the number of errors made were examined. Patients with frontal and non-frontal lesions performed significantly worse than healthy controls for both completion time and the number of errors. However, there was no significant difference for both completion time and the number of errors when patients with frontal and non-frontal lesions were compared. Performance was also not significantly different between patients with focal lesions within different regions of the frontal lobe (orbital, left lateral, right lateral, medial). Our findings suggest that the TMT-B is a robust test for detection of brain dysfunction. However, its capacity for detecting frontal executive dysfunction appears rather limited. Clinicians should be cautious when drawing conclusions from performance on the TMT-B alone. (*JINS*, 2015, *21*, 169–174)

Keywords: Frontal lobes, Focal lesions, Executive function, Neuropsychology, Neuropsychological tests, Brain diseases

INTRODUCTION

The Trail Making Test (TMT; Army Individual Test Battery, 1944) is one of the most widely used neuropsychological tests in the clinical setting, with Part B of the TMT (TMT-B) the most commonly administered subtest. It is quick and easy to administer and has been shown to be highly sensitive to brain dysfunction in a variety of neurological disorders such as Traumatic Brain Injury and Alzheimer's Disease (e.g., Lange, Iverson, Zakrzewski, Ethel-King, & Franzen, 2005; Rasmusson, Zonderman, Kawas, & Resnick, 1998). TMT-B requires subjects to connect a series of 25 encircled numbers and letters pseudo-randomly arranged on a page in ascending order, alternating between number and letter (e.g., 1-a-2-b),

as quickly as possible. Completion time is the most frequently used dependent measure for performance. However, assessing the number of errors made has also been suggested to be a useful measure (Mahurin et al., 2006; Stuss et al., 2001). Since its original conception, many different versions of the task have been developed to accommodate and account for verbal/visual difficulties, physical limitations, and/or age differences (see Bowie & Harvey, 2006 and Spreen & Strauss, 1998).

TMT-B is generally regarded as a test of higher-order executive function (Spreen & Strauss, 1998). Accurate detection of executive impairment is important for the clinical management of many common neurological disorders such as stroke and dementia. A general assumption is that poor performance on TMT-B can be used as a marker for frontal lobe dysfunction. Indeed, the TMT-B is often used clinically for identifying patients with frontal lobe disturbances

compared with those with non-frontal disturbances (e.g., Delis, Kaplan, & Kramer, 2001; Ettlin et al., 2000).

Surprisingly, research evidence so far for the specificity of the TMT-B in detecting frontal executive dysfunction has been relatively limited and results have been mixed. The majority of studies have mainly compared the performance of patients with frontal lobe lesions with a healthy control group only (e.g., Davidson, Gao, Mason, Winocur, & Anderson, 2008).

Few studies have directly compared performance between patients with frontal and non-frontal lesions. In particular, a seminal study by Stuss and colleagues (2001) found that patients with frontal lobe lesions were slower at completing TMT-B compared with healthy controls and patients with non-frontal brain pathology. Within the frontal lobe patient group, those with dorsolateral damage were most impaired while those with inferior-medial damage were least affected. This is consistent with fMRI findings which show greater activity in the left dorsolateral prefrontal region during TMT-B performance (Moll et al., 2002). Also of interest, they noted that in their sample only patients with frontal lobe lesions made more than one error. Stuss and colleagues concluded that the TMT-B is useful for assessing frontal lobe function and in particular that error analysis may be a more informative measure to categorize performance than completion time. However, the sample of frontal patients included over 35% of patients with bilateral lesions as well as traumatic lesions. In contrast, the sample of non-frontal patients did not include bilateral or traumatic lesions, and the sample was relatively small (n = 13). These factors may have contributed to the heightened difference in performance found between the two lesion groups.

In contrast, other studies have not found a difference in performance between frontal and non-frontal patients on the TMT-B when comparing completion time data (for a meta-analysis, see Demakis, 2004). In a study by Reitan and Wolfson (1995), performance of four patient groups of equal size that differed in lesion location (frontal *vs.* non-frontal) and lateralization (left *vs.* right) on the TMT-B was compared. Completion time on the TMT-B was not significantly different for frontal *versus* non-frontal patients or between left and right lesioned patients. Similarly, in a relatively large retrospective series of acute stroke patients, no significant difference in performance was observed between frontal and non-frontal stroke patients (Tamez et al., 2011). However, neither study investigated error frequency in their analysis or looked at more specific neuroanatomical sub-regions within the frontal lobe (e.g., medial *vs.* dorsolateral) which Stuss and colleagues proposed as important discriminating factors. In addition, neither study included a healthy control group for comparison.

The aim of our retrospective study was to compare performance on the TMT-B in a large sample of patients with focal frontal and non-frontal lesions, and healthy controls. Importantly, both completion time and error frequency was examined for any differences in performance between groups.

METHODS

Participants

Patients assessed in the Neuropsychology Department of the National Hospital for Neurology and Neurosurgery (London, UK) were retrospectively screened for study eligibility. Inclusion criteria for the study were (a) the presence of a single focal lesion confined to the frontal or non-frontal brain region due to a cerebrovascular accident, low grade tumor or meningioma (b) availability of neuropsychological data which must include TMT Part-B (c) aged between 18 and 80 years (d) no gross perceptual disturbances (i.e., above the cutoff on the Incomplete Letters subtest of the VOSP), and (e) absence of psychiatric disorders or previous neurological disorders. A total of 82 patients with focal unilateral lesions who met the inclusion criteria were identified for the study (55 frontal patients, 27 non-frontal patients). All stroke and tumor diagnosis was confirmed by clinical neurological investigations including neuroimaging (MRI or CT). All tumor patients had undergone resection before neuropsychological assessment. We have previously demonstrated that the grouping together of patients with focal lesions caused by stroke and various tumors is methodologically justifiable (Cipolotti et al., 2014). Frontal lesions were reviewed by two independent neurologists from available MRI (n = 39; T1 weighted images on 1.5T scanners) and CT scans (n = 16). Lesions were then classified by standard laterality (left: n = 28, right: n = 27) and by a more refined anatomical sub-division of four main subgroups: orbital (n = 8), left lateral (n = 12), right lateral (n = 12), and medial (n = 23). The procedure for classifying lesion location is described in detail elsewhere (Murphy et al., 2013). In addition, seventy healthy controls who had no prior history of neurological or psychiatric disorders were also included for comparison. The study was approved by the local clinical governance and ethics committees. Some of the data from these patients were gathered as part of a larger study of frontal lobe lesions and have been included in previous published studies (e.g., Robinson, Shallice, Bozzali, & Cipolotti, 2012).

Neuropsychological Assessment

As this was a retrospective study, all participants were administered a series of cognitive tests at the time of their neurological investigation. An estimate of pre-morbid optimal level of functioning was obtained using the National Adult Reading Test (NART) and the Incomplete Letters of the Visual Object and Space Perception (VOSP) test was used to assess visual-perceptual functioning. TMT-B was administered as a part of the assessment, following the standard protocol (Bowie & Harvey, 2006). Part A of the TMT was not administered. The two main dependent variables were completion time in seconds and number of errors. Completion time was recorded at the time of administration and the number of errors made was classified and recorded retrospectively.

Statistical Analyses

For analyses of completion time data, an analysis of covariance (ANCOVA) was used with group (frontal, non-frontal, healthy control) as the main between subjects factor. Age and NART scores were entered as covariates of no interest in the analyses given that age and intellectual functioning are known to mediate neuropsychological test performance. The raw completion time measure was transformed using a natural log transformation before analyses to deal with the significant skew in the data. Error data were analyzed using the χ^2 test. Where comparisons contained cells of size less than 5, the Fisher's exact test was used. We divided participants of each group into those who did not make an error and those who made one or more error to assess whether one group is more likely to make an error than another. In view of the finding by Stuss et al. (2001) that only patients with frontal lobe lesions made greater than one error on the TMT, we ran an additional analysis dividing participants into those who made one error and those who made more than one error.

We conducted three main analyses. First, we compared TMT-B performance for each of the two patient groups with the healthy control group separately. This analysis investigated whether TMT-B could distinguish individuals with brain impairment from those without, irrespective of lesion location. Second, we compared performance between the two patient groups directly to examine whether performance on the TMT-B is significantly different between them. Finally, we examined possible lateralization or localization effects in performance for patients with focal lesions in different regions within the frontal lobe.

RESULTS

All groups were well-matched for age, gender and years of education ($p > .1$). Table 1 provides a summary of group demographic information. There was no apparent difference between the frontal and non-frontal patient groups in terms of chronicity or side of injury ($p > .1$). The time since injury at assessment was also not significantly correlated with TMT-B completion time ($p > .1$). On neuropsychological assessment, there was no apparent difference between groups on the NART ($p > .1$) or on a test of visual perception (VOSP, Incomplete letters, $p > .1$).

Table 1. *Demographic and Neuropsychological Data: Patients and Healthy Controls*

	N	Mean (SD)
Age (years)		
Frontal patients	55	46.45 (14.10)
Non-frontal patients	27	45.33 (14.99)
Healthy controls	70	48.44 (14.50)
Gender (M/F)		
Frontal patients	30/25 (55%/45%)	
Non-frontal patients	14/13 (52%/48%)	
Healthy controls	39/31 (56%/44%)	
Education (years)		
Frontal patients	46	13.43 (3.00)
Non-frontal patients	26	14.62 (3.44)
Healthy controls	52	13.87 (3.03)
Side of injury (left/right)		
Frontal patients	28/27 (51%/49%)	
Non-frontal patients	18/9 (67%/33%)	
Chronicity (months)		
Frontal patients	54	15.86 (26.55)
Non-frontal patients	27	12.11 (22.51)
Etiology (low grade tumour/cerebrovascular accident/meningioma)		
Frontal patients	22/18/15 (40%/33%/27%)	
Non-frontal patients	14/9/4 (52%/33%/15%)	
Full Scale IQ (NART)		
Frontal patients	53	108.72 (11.13)
Non-frontal patients	27	105.78 (12.06)
Healthy controls	70	110.50 (9.76)
VOSP incomplete letters		
Frontal patients	54	19.57 (0.60)
Non-frontal patients	14	19.50 (0.94)

Note. Despite absolute differences in means, no statistical difference was found between groups for all demographic and neuropsychological variables.
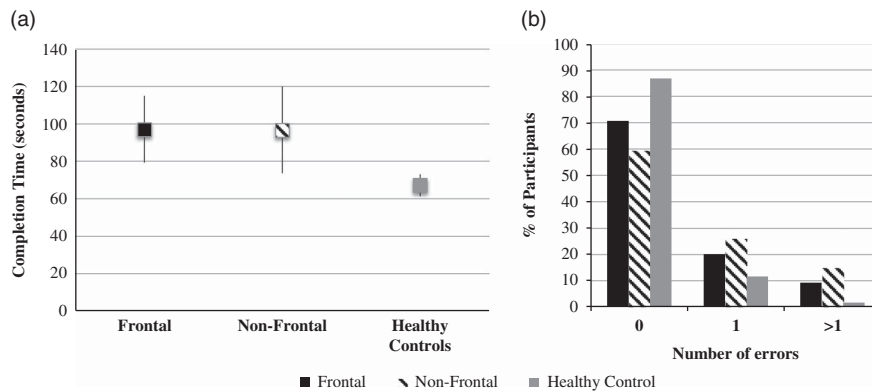
Fig. 1. Performance on the TMT-B for patients and healthy controls. (a) Mean completion time in seconds and 95% CI, and (b) the number of errors.

Comparison between frontal/non-frontal patients with healthy controls

For completion time, both frontal ($\overline{x} = 90.31$ s; $SD = 55.82$) and non-frontal ($\overline{x} = 96.70$ s; $SD = 58.25$) patients were significantly slower on completing the TMT-B compared with healthy controls (($\overline{x} = 67.24$ s; $SD = 24.44$); Frontal *vs.* Control – $F(1,119) = 11.48$; $p < .01$; $\eta_p^2 = 0.09$; Non-frontal *vs.* Control – $F(1,93) = 11.48$; $p < .01$, $\eta_p^2 = 0.11$; see Figure 1a).

Similarly, both frontal and non-frontal patients were significantly more likely to make an error than healthy controls (Frontal *vs.* Control – $\chi^2(1) = 5.07$; $p < .05$; $\Phi = -0.20$; Non-frontal *vs.* Control – $\chi^2(1) = 9.26$; $p < .05$; $\Phi = -0.31$). Figure 1b shows the percentage of participants in each error frequency category for the three groups. As shown, although the majority of healthy controls made no errors (87%), 13% did make an error with one participant (of the total 70) making two errors. Therefore, making an error in itself does not necessarily discriminate between those with and without a focal lesion.

Comparison between frontal and non-frontal patients

Comparison between the frontal and non-frontal patient group on completion time for the TMT-B revealed no significant difference ($F(1,76) = 0.39$; $p = .53$). In addition to conventional null-hypothesis significance testing, we also subjected completion time data to Bayesian analysis which is thought to be able to provide evidence in support of the null-hypothesis (Gallistel, 2009); in this instance, that completion time performance for the two patient groups are equivalent. Analyses yielded odds of 31:1 (weight: 1.50) in favor of the null-hypothesis. These odds are considered *strong* and weights are considered *heavy* (c.f., Gallistel, 2009).

With regard to errors, there was no significant difference between the frontal and non-frontal groups in the likelihood of making an error and making no errors ($\chi^2(1) = 1.11$; $p = .29$). Furthermore, there was also no difference when we compared the likelihood of these two groups in making one

or more than one error ($p = 1.00$). In our sample, five patients in the frontal group (8%) and four patients in the non-frontal group (15%) made more than one error.

Comparison between lesion locations for frontal patients

To examine whether lesion location within the frontal lobe had an effect on performance on the TMT-B, frontal patients were divided into separate subgroups based on standard lesion lateralization (left, right) and by more specific neuro-anatomical sub-regions (orbital, left lateral, right lateral, medial).

Comparison of completion time for the TMT-B between left ($\overline{x} = 89.47$ s; $SD = 30.34$) and right ($\overline{x} = 91.25$ s; $SD = 70.78$) frontal patients revealed no significant difference between groups ($F(1,49) = 0.01$; $p = .36$). Similarly, no significant difference in completion time was found between the different frontal sub-regions (orbital: $\overline{x} = 81.73$ s; $SD = 39.77$; left lateral: $\overline{x} = 93.91$ s; $SD = 47.00$; right lateral: $\overline{x} = 102.86$ s; $SD = 91.90$; medial: $\overline{x} = 85.95$ s; $SD = 46.67$; $F(3,47) = 0.67$; $p = .58$). There was also no significant difference in the likelihood of making an error between the left and right frontal patients ($\chi^2(1) = 0.26$; $p = .61$) or between the different sub-regions ($p = .88$). We ran an additional analysis to compare patients with "lateral" *versus* "medial" frontal lesions by combining the patients with left and right lateral lesions into one group. This did not alter the pattern of results.

DISCUSSION

Our findings show that the TMT-B is a reliable test for detecting brain impairment. Both patients with frontal and non-frontal lobe lesions performed worse than the neuro-logically intact participants with regard to both efficiency and accuracy. Critically however, completion time performance on the TMT-B was not significantly different between the two patient groups. The presence or number of errors performed during the task did not appear to provide any

additional information. Non-frontal patients were equally likely to make an error or more than one error on the task compared with frontal patients. Furthermore, a small proportion of neurologically intact participants also made an error on the task. These results do not corroborate with that of Stuss and colleagues (2001) who suggested that error analysis could distinguish frontal from non-frontal lesions. This could be due to the different composition of patient groups between the two studies; the frontal patient group in the previous study contained a large proportion of patients with bilateral lesions. Our study corroborates with previous studies that have shown no difference in performance between frontal and non-frontal patients in TMT-B efficiency (Demakis, 2004; Reitan & Wolfson, 1995; Tamez et al., 2011) and extends this by demonstrating a lack of effect similarly when examining the number of errors made.

We also found that TMT-B performance could not distinguish between different frontal lesion locations. Contrary to previous suggestions that TMT-B performance might be left lateralized (e.g., Moll et al., 2002) or localized to the dorsolateral region (e.g., Stuss et al., 2001), we found no significant differences in performance between patients with left and right frontal lesions or between patients with dorsolateral lesions compared with medial or orbital lesions. Our findings suggest that TMT-B performance does not depend on any specific frontal region.

Although our findings show that TMT-B performance cannot distinguish patients with frontal and non-frontal focal lesions, or between different sub-regions of the frontal lobe, we are not suggesting that the TMT is a task that does not require frontal lobe involvement. Rather, we suggest that performance on the TMT-B most likely relies on a distributed network involving both frontal and non-frontal regions. Functional imaging has been one useful way of elucidating underlying brain networks involved in performing the TMT-B, although translation of the task into a design optimized for imaging has its limitations. Nevertheless, recent studies have consistently shown that successful performance on the TMT-B involves not only frontal regions, but also posterior and subcortical regions (Jacobson, Blanchard, Connolly, Cannon, & Garavan, 2011; Zakzanis, Mraz, & Graham, 2005).

From a clinical perspective, the TMT-B can offer many insights regarding a patient's cognitive abilities. However, our findings suggest it may have limited utility as a tool in detecting frontal executive dysfunction specifically as task performance most likely depends upon a range of cognitive processes, some of which require non-frontal brain regions. As such, caution should be used when drawing conclusions from TMT-B performance alone. Further work is needed to establish whether current findings also extend to the various alternate versions of the task (e.g., Oral or Color TMT).

ACKNOWLEDGMENTS

REFERENCES

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

Bowie, C.R., & Harvey, P.D. (2006). Administration and interpretation of the trail making test. *Nature Protocols*, *1*(5), 2277–2281.

Cipolotti, L., Healy, C., Chan, E., Bolsover, F., Lecce, F., White, M., … Shallice, T. (2014). The impact of different aetiologies on the performance of frontal patients on cognitive tasks. *Neuropsychologia*, doi: 10.1016/j.neuropsychologia. 2014.12.025

Davidson, P.S.R., Gao, F.Q., Mason, W.P., Winocur, G., & Anderson, N.D. (2008). Verbal fluency, trail making, and Wisconsin Card Sorting Test performance following right frontal lobe tumor resection. *Journal of Clinical and Experimental Neuropsychology*, *30*(1), 18–32.

Delis, D.C., Kaplan, E., & Kramer, J.H. (2001). *Delis-Kaplan executive function system (D-KEFS)*. San Antonio, TX: Psychological Corporation.

Demakis, G.J. (2004). Frontal lobe damage and tests of executive processing: A meta-analysis of the category test, stroop test, and trail-making test. *Journal of Clinical and Experimental Neuropsychology*, *26*(3), 441–450.

Ettlin, T.M., Kischka, U., Beckson, M., Gaggiotti, M., Rauchfleisch, U., & Benson, D.F. (2000). The frontal lobe score: Part 1: Construction of a mental status of frontal systems. *Clinical Rehabilitation*, *14*(3), 260–271.

Gallistel, C.R. (2009). The importance of proving the null. *Psychological Review*, *116*(2), 439–453.

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.

Lange, R.T., Iverson, G.L., Zakrzewski, M.J., Ethel-King, P.E., & Franzen, M.D. (2005). Interpreting the trail making test following traumatic brain injury: comparison of tradition time scores and derived indices. *Journal of Clinical and Experimental Neuropsychology*, *27*, 897–890.

Mahurin, R.K., Velligan, D.I., Hazleton, B., Davis, J.M., Eckert, S., & Miller, A.L. (2006). Trail making test errrors and executive function in schizophrenia and depression. *The Clinical Neuropsychologist*, *20*, 271–288.

Moll, J., de Oliveira-Souza, R., 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 Neuropsiquiatria*, *60*(4), 900–905.

Murphy, P., Shallice, T., Robinson, G., MacPherson, S.E., Turner, M., Woollett, K., … Cipolotti, L. (2013). Impairments in proverb interpretation following focal frontal lobe lesions. *Neuropsychologia*, *51*(11), 2075–2086.

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*(2), 169–178.

Reitan, R.M., & Wolfson, D. (1995). Category test and trail making test as measures of frontal lobe functions. *The Clinical Neuropsychologist*, *9*(1), 50–56.

Robinson, G., Shallice, T., Bozzali, M., & Cipolotti, L. (2012). The differing roles of the frontal cortex in fluency tests. *Brian*, *135*(7), 2202–2214.

Spreen, O., & Strauss, E. (1998). *A compendium of neuropsychological tests: Administration, norms, and commentary* (2nd 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*(2), 230–239.

Tamez, E., Myerson, J., Morris, L., White, D.A., Baum, C., & Connor, L.T. (2011). Assessing executive abilities following acute stroke with the trail making test and digit span. *Behavioural Neurology*, *24*, 177–185.

Zakzanis, K.K., Mraz, R., & Graham, S.J. (2005). An fMRI study of the trail making test. *Neuropsychologia*, *43*, 1878–1886.