OXFORD
UNIVERSITY PRESS

Archives of Clinical Neuropsychology 26 (2011) 314–321

Archives of CLINICAL NEUROPSYCHOLOGY

Reliability of Three Alternate Forms of the Trail Making Tests A and B

Stefanie Wagner [1,*], Isabella Helmreich [1], Norbert Dahmen [1,2], Klaus Lieb [1], André Tadić [1]

[1] *Department of Psychiatry and Psychotherapy, University Medical Centre, Mainz, Germany*
[2] *Clinic for Psychiatry and Psychotherapy, 56368 Katzenelnbogen, Germany*

*Corresponding author at: Department of Psychiatry and Psychotherapy, University Medical Centre Mainz, Untere Zahlbacher Strasse 8, D-55131 Mainz, Germany. Tel: +49-6131-17-2989; fax: +49-6131-176490.
E-mail address: stefwagn@uni-mainz.de (S. Wagner).

Accepted 22 March 2011

Abstract

The majority of patients with Major Depressive Disorder (MDD) suffer from significant executive dysfunctions. To investigate the time course of executive functions during antidepressant treatment, repeated measures of executive functions are necessary. In order to avoid practice effects, the assessment of alternate forms is suggested. The aim of this study was to compare the processing times of four alternate versions of the Trail Making Test (TMT) A and B in patients with MDD. Fifty-five subjects with DSM-IV MDD were included in the study. We analyzed mean processing times and retest reliability of the four versions of TMT A and B. Mean processing times did not differ between the four tested versions of TMT A and B. Retest reliability of TMT A and B was between 0.76 and 0.89 and between 0.86 and 0.94, respectively. Because of their identical difficulty and high reliability, the herein described versions of the TMT A and B are suitable for sequential testing of executive functioning.

Keywords: Major depression; Executive functioning; Trail Making Test A/B; Alternate versions

Introduction

Major Depressive Disorder (MDD) is one of the most prevalent psychiatric disorders in general population (Alonso, Angermeyer, & Lepine, 2004). Cognitive impairment in terms of diminished ability to think or to concentrate belongs to the defining criteria of MDD in ICD-10 (Dilling, Mombour, & Schmidt, 1991) and DSM-IV (American Psychological Association, 1994). Empirical evidence supports the existence of moderate but significant cognitive deficits in patients with MDD. With respect to cognitive domains, impairment has been reported for executive functioning in particular (Fossati, Ergis, & Allilaire, 2002; Harvey et al., 2004), whereas less significant deficits have been found for psychomotor speed (Sobin & Sackeim, 1997), attention (Christensen, Griffiths, Mackinnon, & Jacomb, 1997), and memory (Veiel, 1997).

Studies investigating the effectiveness of antidepressant treatment are nearly exclusively based on depression rating scales, for example, the Hamilton Depression Rating Scale (Hamilton, 1960), and usually do not contain measures for executive functioning. Because of their high frequency in subjects with MDD and their sensitivity to antidepressant treatment, executive functions could be an important phenotype regarding the overall outcome evaluation in the course of an antidepressant treatment. The term executive function describes a set of cognitive abilities that control and regulate other abilities and behaviors. Executive functions are necessary for goal-directed behavior. They include the ability to initiate and stop actions, to monitor and change behavior as needed, and to plan future behavior when faced with novel tasks and situations. Executive functions allow us to anticipate outcomes and adapt to changing situations (Bryan & Luszcz, 2000).

Until now, only a few studies with different methodological drawbacks have focused on the effects of antidepressant treatment on executive functioning. These studies reported a substantial improvement in cognitive deficits from the beginning to the

doi:10.1093/arclin/acr024

end of an antidepressant treatment, specifically in executive functions (Beblo, Baumann, Bogerts, Wallesch, & Herrmann, 1999; Biringer et al., 2005; De Groot, Nolen, Huijsman, & Bouvy, 1996), but they were limited to the comparison of executive functioning before and after antidepressant treatment.

The Trail Making Test (TMT) was originally developed as part of the Army Individual Test Battery (1944) and is one of the most popular tests in neuropsychological practice due to its high sensitivity in diagnosing brain impairment (Crowe, 1998; Gaudino, Geisler, & Squires, 1995). Furthermore, the TMT is a component of various test batteries like the CERAD-NP (Consortium to Establish a Registry for Alzheimer's Disease, 1986) or the Delis–Kaplan Executive Function System (Delis, Kaplan, & Kramer, 2001). Nevertheless, these test batteries did not provide alternate forms for the TMT (Swanson, 2005). The TMT provides information on visual search, scanning, speed of processing, mental flexibility, and executive functioning (Tombaugh, 2003). Previous studies reported that the TMT performance was affected by age, education, and general cognitive ability (Corrigan & Hinkeldey, 1987; Giovagnoli et al., 1996; Lamberty, Putnam, Chatel, Bieliauskas, & Adams, 1994). Men and women did not differ in their TMT performance (Giovagnoli et al., 1996). Emotional disturbances, severe anxiety, and psychotic symptoms may also affect TMT performance, especially in Part B (Giovagnoli et al., 1996). Furthermore, patients with MDD were found to be slower than healthy controls in both parts of the TMT.

Multiple studies reported significant improvement in both parts A and B of the TMT, when one test form was repeatedly administered to a patient. Because of this practice effect, previous researchers have attempted to develop alternate forms of the TMT. Lewis and Rennick (1979) developed four alternate forms of part B, but not of part A. Several investigators suggested the ratio between TMT A and B being an important index of cognitive flexibility and efficacy (Gaudino et al., 1995; Lamberty et al., 1994). Therefore, the use of the four alternate forms of trails B without alternate forms of part A should not be recommended for an appropriate investigation of the time course of executive functions. Franzen, Paul, and Price (1990) have developed alternate forms of the TMT called TMT C and D. The TMT C is an alternate form of part A and the TMT D an alternate form of part B. TMT C and D have the same relative position of each circle than TMT A and B but inverted the labeled sequences respective to their equivalents in TMT A and B. The high reliability and validity of these alternate forms is documented in previous studies (LoDasso, Rapport, Axelrod, & Reeder, 1998; McCracken & Franzen, 1992). In order to investigate the change of executive functions in the course of an antidepressant treatment, it seems to be sensible to assess the executive test performance repeatedly during an antidepressant treatment over several weeks. Therefore, a set of at least four alternate versions of the TMT part A and B is needed. Because Franzen and colleagues developed only one alternate form of part A and one of part B, these versions only allow an assessment of executive functioning before and after treatment. In order to describe the time course of executive functions in a more detailed fashion, it seems to be appropriate to assess the executive test performance more than two times in the course of treatment. In summary, in spite of their high reliability and validity, it seems not to be appropriate to use the alternate forms of Franzen and colleagues in order to investigate the time course of executive function during an antidepressant treatment.

To the best of our knowledge, there are no studies characterizing a repeated assessment of executive functions during antidepressant treatment as yet. In order to investigate the progress of executive functions in the course of an antidepressant treatment, two approaches seem to be appropriate. (a) To apply repeated assessments of one version of a neuropsychological test in MDD patients and in healthy controls. However, in the case of improved performance in both groups, it might be impossible to separate treatment effects from practice effects. Therefore, previous studies were not able to detect treatment effects because of the repeated assessment of one test version (Beblo et al., 1999; Reppermund, 2008). (b) The use of alternate test versions to avoid practice effects, which is the preferred procedure in literature (Amelang & Zielinski, 2002). In the case of using alternate forms, one has to consider that minor differences between the alternate versions could skew the results. By randomly distributing the alternate forms to the subjects, this bias can be controlled.

In order to provide a set of tests, which enables the attribution of changes in executive functioning to antidepressant treatment effects, we developed three alternate forms of the TMT A and B. The TMT provides information not only on executive functions but also on general attention, psychomotor speed, and cognitive efficiency. All these domains could be affected in patients with MDD (Christensen et al., 1997; Sobin & Sackeim, 1997). Therefore, one could suppose that an antidepressant medication affects not only the executive functions of MDD patients but attention, psychomotor speed, and cognitive efficiency as well. The aim of the present study was to compare the processing times of the three new versions of the TMT A and B with the original versions in patients with DSM-IV MDD (American Psychiatric Association, 1994). The main hypothesis of the study was that the processing times between the four versions of TMT A and B would not differ from each other. Additionally, we hypothesized that the TMT performance would decrease with age and increase with general cognitive ability.

Methods

Participants

All participants gave their written informed consent to participate in the study after a complete and extensive description. All study components were approved by the local ethical committee of the Landesärztekammer Rheinland-Pfalz (study code n°: 837.166.09 (6671)) and are compliant with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Fifty-five patients of Caucasian descent with DSM-IV Major Depression (American Psychiatric Association, 1994) were recruited at the Departments of Psychiatry and Psychotherapy in Mainz and Katzenelnbogen, Germany. Diagnosis according to DSM-IV was ascertained by at least one specialist in psychiatry and at least one resident physician in psychiatry. Patients were not included if one or more of the following criteria were present: (a) dementia or any other organic brain syndrome with obvious cognitive impairment, (b) doubts in the patients' ability to give informed consent (e.g., patients suffering from depression with psychotic symptoms), (c) a current diagnosis of a psychotic disorder, (d) the patient had a traumatic brain injury in the past, and (e) the patient consumed benzodiazepines (Lorazepam $\geq$ 1.5 mg/day).

Materials

The TMT (Reitan, 1979) consists of two parts. In the first part (TMT A), subjects had to trace a trail through a series of circles numbered from 1 to 25 variously spread over a sheet of paper. In the second part of the test (TMT B), subjects must perform an alternation between circles containing numbers and letters. The sequence proceeds from the first number to the first letter alphabetically followed by the second number and the second letter. TMT B contains 13 circles numbered 1–13, alternating with 12 circles lettered A to L. In both parts, the performance is the time needed to complete the trail correctly. In order to investigate the progress of executive functions in the course of an antidepressant treatment, we developed three alternate forms of the TMT A and B. The new versions of TMT A and B are mirror images of the original version. First, the original version was printed on an overhead transparency. Afterward, the overhead transparency was reflected three times about horizontal as well as vertical lines resulting in three different copies of the original version (Fig. 1).

In order to develop alternate forms of a neuropsychological test procedure, the new versions should have the same degree of difficulty compared with the original version. To achieve this aim, all versions should have comparable numbers of circles which have to be traced as well as comparable overall distances between all circles (Oswald & Roth, 1987). Table 1 shows these values for all versions of TMT A and B.

General cognitive ability was assessed by the Multiple Vocabulary Test (MWT-B; Lehrl, 1969), in which patients had to differentiate real German words from pseudowords. The correct words are summed up to a raw score (range: 0–37) and transferred into IQ values. Depression severity was assessed by the 30-items clinician-rated version of the Inventory of Depressive

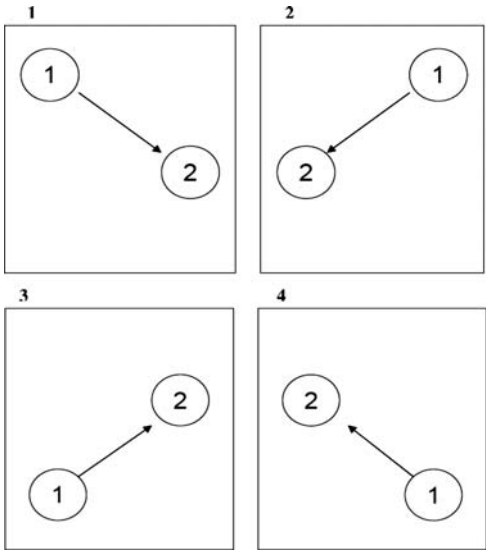

Fig. 1. Pattern of the development of three new versions of TMT A and B. A is the original version.

Table 1. Parameters of the alternate versions of the TMT A and B

	Version 1	Version 2	Version 3	Version 4
TMT A				
Number of circles	25	25	25	25
Overall distance (cm)	184.4	187.5	185.4	185.3
TMT B				
Number of circles	25	25	25	25
Overall distance (cm)	242.9	243.5	242.2	241.7

Note: TMT = Trail Making Test.

Symptomatology (IDS-C30; Rush, Gullion, Basco, Jarrett, & Trivedi, 1986). Each item of the IDS refers to a different depressive symptom. The severity of each symptom is expressed with a 4-point scale ranging from 0 to 3. The raw scores of each item were summed up. The IDS-C30 total score ranges from 0 to 84. Cutoff values have been established for different grades of depression severity (0–11 = none, 12–23 = mild, 24–36 moderate, 37–46 severe, 47–84 very severe; www.ids-qids.org).

Study procedure

First, subjects performed the Multiple Vocabulary Test. Afterward, the four versions of the TMT A and B were applied. In order to avoid any bias in processing times due to an identical sequence of test versions, the four versions were randomly distributed to the subjects, leading to the orders of versions 1, 2, 3, 4 ($N = 14$); versions 2, 3, 4, 1 ($N = 14$); versions 3, 4, 1, 2 ($N = 13$); and versions 4, 1, 2, 3 ($N = 14$). Finally, the IDS-C30 was applied.

Statistics

All analyses were done using SPSS 17.0. First, we calculated two composite scores averaging the mean processing times of all versions of TMT A and B, respectively. To identify the impact of potential covariates on the TMT performance, we performed correlation analysis between the composite scores of TMT A and B, respectively, and the covariates age, intelligence, gender, and depression severity. Age and intelligence were significantly correlated with the TMT performance in both parts, and gender and depression severity were not. Differences in mean processing times between the four versions of the TMT A and B were examined by repeated-measures ANCOVA with age and intelligence as covariates. The reliabilities of the four versions of TMT A and B were calculated by retest reliabilities of version 1 with each other version of TMT A and B. The association between the processing times of TMT versions 1 and 4 was analyzed by a multiple regression analysis with the TMT performance in version 4 as criterion and the TMT performance in version 1, age, intelligence, and depression severity as predictors. To investigate learning effects during the execution of the alternate forms of TMT A and B, we analyzed processing time of the alternate forms presented at positions 1, 2, 3, and 4 by repeated-measures ANCOVA with age and intelligence as covariates. Furthermore, mean processing times between the TMT A and B versions presented at positions 1 and 2, 2 and 3 as well as 3 and 4 were calculated by *t*-tests for paired variables. In order to separate treatment effects from practice effects in repeated analyses of executive functioning, we calculated a possible correction factor for practice effects. With this aim, we used three models: (1) we analyzed the mean differences between the versions of TMT A and B presented at positions 1 and 2, 2 and 3 as well as 3 and 4. As a result, we achieved three correction factors; (2) we built an average improvement score over the three correction factors calculated in model 1; and (3) we calculated the difference between the processing times of the TMT versions presented at positions 1 and 4. Significance was set at $p \leq .050$.

Results

Nineteen participants were men (35%), 36 (65%) were women. Mean age ($\pm SD$) was 46.2 $\pm$ 12.8 years (range: 20–65 years); mean ($\pm SD$) years of school education was 10.8 $\pm$ 1.83 (range: 9–13 years); mean general intelligence ($\pm SD$) was 102.5 $\pm$ 12.9 (range: 82–145). Subjects had a mean ($\pm SD$) IDS-C30 sum score of 28.8 points ($\pm$ 11.0; range: 12–54).

Correlation analyses revealed a significant correlation between age (TMT_A: $r = .461$; $p = .0001$; TMT_B: $r = .502$; $p = .0001$) and intelligence (TMT_A: $r = -.358$; $p = .007$; TMT_B: $r = -.445$; $p = .001$), but not between gender (TMT_A: $r = .066$; $p = .630$; TMT_B: $r = -.049$; $p = .722$) or depression severity (TMT_A: $r = .081$; $p = .558$; TMT_B: $r = .150$; $p = .275$) with the average processing time in TMT A and B.

Table 2. Mean ($\pm SD$) processing times and reliabilities of the four versions of TMT A and B

	Version 1	Version 2	Version 3	Version 4	Average score (1–4)	*p*-value*
Processing time ($M \pm SD$)						
TMT A	32.4 ± 13.9	32.8 ± 14.2	34.7 ± 14.9	33.7 ± 13.2	33.4 ± 12.2	.770
TMT B	77.4 ± 44.2	73.2 ± 30.6	76.6 ± 35.7	77.1 ± 39.8	76.1 ± 34.7	.948
TMT B/A ratio	2.39	2.23	2.21	2.29	2.28	
	Versions 1–2	Versions 1–3	Versions 1–4			
Retest reliability						
Original TMT A	0.81	0.82	0.76			
Original TMT B	0.86	0.89	0.86			

Notes: TMT = Trail Making Test. TMT A: $F_3 = 0.770$; $p = .516$; TMT B: $F_3 = 0.120$; $p = .948$.

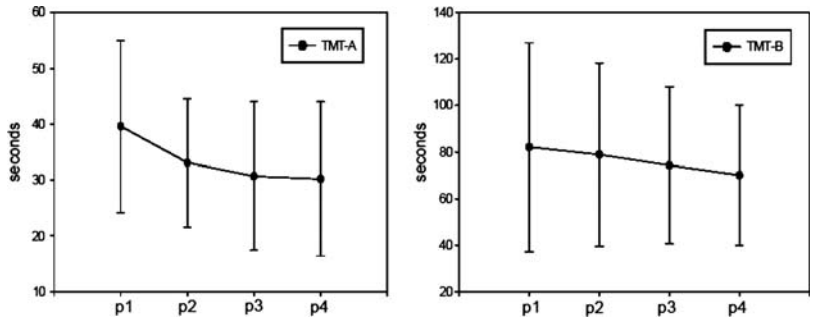

Fig. 2. Mean ($\pm SD$) processing times of the four versions of TMT A and B presented at positions 1, 2, 3, and 4. p1–4, positions 1–4.

ANCOVA revealed that mean processing times (Table 2) between the four versions of the TMT A and B, respectively, did not differ significantly (TMT A: $F_3 = 0.770$; $p = .516$; TMT B: $F_3 = 0.120$; $p = .948$). The covariates age and intelligence had significant effects on TMT performance (age: TMT A: $F_1 = 12.388$; $p = .001$; TMT B: $F_1 = 16.302$; $p = .0001$; intelligence: TMT A: $F_1 = 6.153$; $p = .016$; TMT B: $F_1 = 11.685$; $p = .001$). There were no interaction effects between the TMT performance and the covariates age and intelligence ($p > .219$). Table 2 presents mean ($\pm SD$) processing times of the four alternate forms of the TMT A and B, retest reliabilities as well as TMT B/A ratios.

Regression analysis revealed that TMT A version 1, age, intelligence, and depression severity accounted for 82.3% ($F_4 = 26.257$; $p = .0001$) of the variance of TMT A version 4. The performance in TMT A version 1 ($t = 6.928$; $p = .0001$) as well as age ($t = 3.778$; $p = .0001$) significantly predicted the performance in TMT A version 4; intelligence ($t = -0.882$; $p = .382$) and depression severity ($t = -0.097$; $p = .923$) were not associated with the performance in TMT A version 4. Regarding TMT B, the regression analysis showed that the performance in TMT B version 1, age, intelligence, and depression severity accounted for 88.1% ($F_4 = 43.461$; $p = .0001$) of the variance of TMT B version 4. The performance in TMT B version 1 ($t = 9.614$; $p = .0001$) significantly predicted the performance in TMT A version 4; age ($t = 1.662$; $p = .103$), intelligence ($t = -1.052$; $p = .298$), and depression severity ($t = 0.179$; $p = .859$) did not predict the performance in TMT B version 4.

Possible learning effects were analyzed by mean processing times of the alternate forms of the TMT A and B, which were presented at positions 1, 2, 3, and 4. Mean processing times of the four versions of the TMT A and B presented at positions 1, 2, 3, and 4 are shown in Fig. 2. Repeated-measures ANCOVA revealed a significant decrease in the processing times during the execution of the alternate forms of the TMT A and B presented at positions 1, 2, 3, and 4 (TMT A: $F_3 = 22.913$; $p = .0001$; TMT B: $F_3 = 5.157$; $p = .002$). Intelligence, but not age, had a significant effect on TMT performance (age: TMT A: $F_1 = 0.694$; $p = .411$; TMT B: $F_1 = 3.077$; $p = .090$; intelligence: TMT A: $F_1 = 7.537$; $p = .010$; TMT B: $F_1 = 6.817$; $p = .014$). A comparison of the processing times of the versions presented at positions 1 and 2, 2 and 3 as well as 3 and 4 by *t*-tests for paired variables revealed a significant difference between TMT A versions 1 and 2 ($T_{54} = 4.656$; $p = .0001$) and between TMT A versions 2 and 3 ($T_{54} = 2.524$; $p = .015$), but not between the versions presented at positions 3 and 4 ($T_{54} = 0.566$; $p = .574$). Regarding TMT B, there were no significant differences between the versions presented at positions 1 and 2 ($T_{54} = 0.953$; $p = .645$), at positions 2 and 3 ($T_{54} = 1.406$; $p = .165$) as well as between the versions presented at positions 3 and 4 ($T_{54} = 1.730$; $p = .089$).

In order to enable the separation of treatment effects from practice effects in future longitudinal analyses, we calculated correction factors for practice effects. In the first model, we analyzed the mean relative difference between the processing

Table 3. Correction factors in order to separate practice effects from treatment effects

	Position 1–2 (model 1)	Position 2–3 (model 1)	Position 3–4 (model 1)	Average Improvement (model 2)	Position 1–4 (model 3)
TMT A (%)	16.6	7.1	1.8	8.5	23.9
TMT B (%)	3.9	5.8	5.6	5.1	14.6

Note: TMT = Trail Making Test.

times of the version presented at positions 1 and 2, 2 and 3 as well as 3 and 4. In the second model, we calculated an average improvement factor; in model 3, we analyzed mean differences of the TMT versions presented at positions 1 and 4 as correction factor (Table 3). Results showed that the correction factors in TMT A were higher than in TMT B.

Discussion

In order to investigate the progress of executive functioning in the course of an antidepressant treatment, the same neuropsychological tests have to be performed several times. In the present study, the processing times of four alternate forms of the TMT A and B were compared in the target group of subjects with MDD. The three new versions of the TMT A and B did not differ in their mean processing times from the original versions. The TMT B/A ratios ranged from 2.21 to 2.39. These ratios are in line with the ratios of 2.0 and 2.5 which are suggested to be a representative normative TMT performance (Lamberty et al., 1994). Therefore, the difficulty of the new versions of both parts of the TMT seems to be comparable. The high retest reliability scores as well as the TMT B/A ratios additionally emphasize the high comparability and reliability of the new versions of the TMT A and B. Regression analyses revealed a very strong association between the TMT performance of the first and the last version in both part of the TMT. These results additionally point out the high comparability of the new forms of the TMT with the original versions.

In the present study, all four versions of the TMT A and B were presented consecutively in one session. We found that all subjects became faster in the course of executing the versions. Nevertheless, *a priori post hoc* tests showed that only the TMT A versions presented at positions 1 and 2 as well as 2 and 3 differed significantly in mean processing times. Because we anticipated such learning effects, we balanced the four sequences of the TMT versions across subjects and each version was presented at each position during the assessment virtually equally often (1234, 2341, 3412, 4123). With four versions of the TMT, there were 24 possible orders of these versions. Because of our sample size of 55 patients a distribution of these subjects among all 24 possible orders was not reasonable. Therefore, we limited the possible orders to the four above-mentioned sequences. With this procedure, which controlled for learning effects, we could demonstrate equivalent processing times for all versions of the TMT A and B, respectively. Because of their identical difficulty and high reliability, the four versions herein described are suitable for a sequential testing, for example, in order to track executive functions in MDD during an antidepressant treatment. Nevertheless, a replication study is recommended comparing the processing times of the four versions of the TMT A and B in the course of an antidepressant treatment, presenting the versions with 1 or 2 weeks in between.

We found that the processing times of all versions of TMT A and B increased with age and decreased with higher general cognitive ability. These results are fully in line with previous studies showing the TMT performance being affected by age, education, and general cognitive ability (Giovagnoli et al., 1996; Lamberty et al., 1994; Tombaugh, 2003). Men and women did not differ in their TMT performance. This result is also fully in line with previous studies showing no differences between men and women in their TMT performance (Giovagnoli et al., 1996). Our result that the TMT performance in all alternate forms was affected by the same covariates emphasizes the comparability and reliability of the new alternate forms with the original versions of the TMT A and B.

Depression severity was not associated with the TMT performance in our study. The regression analyses revealed that depression severity explained only a low percentage of the variance of the TMT performance. This result is contrary to previous studies reporting emotional disturbances affecting TMT performance (Giovagnoli et al., 1996). One explanation of this result might be that the depression severity of our patients was rather mild to moderate. Thus, it seems to be possible that our sample was too homogenous to detect clearer differences in the TMT performance between subjects with different depression severity scores. Furthermore, previous studies suggesting an association between emotional disturbances and TMT performance typically compared MDD patients with healthy controls. Studies investigating MDD patients with different depression severity scores are absent as yet. Thus, a replication study with larger sample sizes as well as with patients suffering from more severe depression seems to be necessary.

Our results showed that the use of alternate forms decreased practice effects. Nevertheless, the herein presented learning effects pointed out that—despite the use of alternate forms—the experience with the task improved the TMT performance.

In order to separate practice effects from treatment effects, it seems to be sensible to include a correction factor for the practice effect in repeated assessments of executive functions. We suggested three different models in order to calculate such a correction factor. The first model based on the analyses of multiple correction factors for each possible time point. The advantage of this model is that it is a very accurate method to deal with the data. Nevertheless, in the case of many follow-up assessments of executive functioning, it seems to be a very complex procedure. In the second model, we calculated an average improvement score defined as the mean improvement of the scores in model 1. In the third model, we analyzed the relative improvement in TMT performance from positions 1 to 4. In model 2, the correction factor is relatively low; thus, it seems to be possible that a practice effect will be misleadingly interpreted as treatment effect. Model 3 is very rigorous; therefore, there might be a high risk to misinterpret a treatment effect as practice effect and therefore underestimate the true treatment effect. In summary, the above-mentioned correction factors have to be validated in longitudinal clinical trials investigating repetitive measures of executive functions in the course of an antidepressant treatment as well as in healthy controls.

In conclusion, this study comparing mean processing times of three alternate forms of the TMT A and B with the original version showed that the processing times of the new versions of both parts of the TMT did not differ from the original version. As it has been shown for the original version, we found that the TMT performance was affected by age and general cognitive ability in all new versions of TMT A and B but not by sex or depression severity. These results suggest that this newly established set of four versions of both parts of the TMT is suitable to characterize in detail the time course of executive functioning during an antidepressant treatment. This might be very useful—for clinical as well as for research purposes—for a more comprehensive assessment of antidepressant treatment effects, which is currently exclusively based on changes in depression rating scale scores.

Conflict of Interest

None declared.

Acknowledgements

We are grateful to all participants as well as to the staff of the cooperating Clinic for Psychiatry and Psychotherapy in Katzenelnbogen, Germany, for study support.

References

Alonso, J., Angermeyer, M. C., & Lepine, J. P. (2004). Prevalence of mental disorders in Europe: Results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. *Acta Psychiatrica Scandinavica*, *109*, 21–27.

Amelang, M., & Zielinski, W. (2002). *Psychologische Diagnostik und Intervention*. Heidelberg: Springer-Verlag.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Press.

Beblo, T., Baumann, B., Bogerts, B., Wallesch, C., & Herrmann, M. (1999). Neuropsychological correlates of major depression: A short-term follow-up. *Cognitive Neuropsychiatry*, *4*, 333–341.

Biringer, E., Lundervold, K., Strodal, A., Mykletun, J., Efeland, R., Bottlender, R., et al.. (2005). Executive function improvement upon remission of recurrent unipolar depression. *European Archives of Psychiatry and Clinical Neuroscience*, *255*, 373–380.

Bryan, J., & Luszcz, M. A. (2000). Measurement of executive function: Considerations for detecting adult age differences. *Journal of Clinical and Experimental Neuropsychology*, *22*, 40–55.

Christensen, H., Griffiths, K., Mackinnon, A., & Jacomb, P. (1997). A quantitative review of cognitive deficits in depression and Alzheimer-type dementia. *Journal of the International Neuropsychological Society*, *3*, 631–651.

Corrigan, J. D., & Hinkeldey, N. S. (1987). Relationship between parts A and B of the Trail Making Test. *Journal of Clinical Psychology*, *4*, 402–409.

Crowe, S. F. (1998). The differential contribution of mental tracking, cognitive flexibility, visual search and motor speed to performance on parts A and B of the Trail Making Test. *Journal of Clinical Psychology*, *54*, 585–591.

De Groot, M. H., Nolen, W. A., Huijsman, A. M., & Bouvy, P. F. (1996). Lateralized neuropsychological functioning in depressive patients before and after drug therapy. *Biological Psychiatry*, *40*, 1282–1287.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan Executive Function System (D-KEFS)*. San Antonio, TX: The Psychological Corporation.

Dilling, H., Mombour, W., & Schmidt, M. H. (1991). *International classification of mental and behavioral disorders*. ICD-10, chap. V. Bern, Switzerland: Huber.

Fossati, P., Ergis, A. M., & Allilaire, J. F. (2002). Executive functioning in unipolar depression: A review. *Encephalcology*, *28*, 97–107.

Franzen, M. D., Paul, D. S., & Price, G. (1990). *Alternate form reliability of Trails A, B, C, and D*. Paper presented at the Ninth Annual Convention of the National Academy of Neuropsychology, Reno, NY.

Gaudino, E. A., Geisler, M. W., & Squires, N. K. (1995). Construct validity in the Trail Making Test: What makes part B harder? *Journal of Clinical and Experimental Neuropsychology*, *17*, 529–535.

Giovagnoli, A. R., Del Pesce, M., Mascheroni, S., Simoncelli, M., Laiacona, M., & Capitani, E. (1996). Trail Making Test: Normative values from 287 normal adult controls. *Italian Journal of Neurological Science*, *17*, 305–309.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, *23*, 56–62.

Harvey, P. O., Le Bastard, G., Pochon, J. B., Levy, R., Allilaire, J. F., Dubois, B., et al. (2004). Executive functions and updating of the contents of working memory in unipolar depression. *Journal of Psychiatric Research*, *38*, 567–576.

Lamberty, G. J., Putnam, S. H., Chatel, D. M., Bieliauskas, L. A., & Adams, K. M. (1994). Derived Trail Making Test Indices. *Neuropsychiatry, Neuropsychology and Behavioural Neurology*, *7*, 230–234.

Lehrl, S. (1969). *Mehrfachwahl-Wortschatz-Intelligenztest (MWT-B)*. Erlangen: Perimed Fachbuch Verlagsgesellschaft mbH.

Lewis, R. F., & Rennick, P. M. (1979). *Manual of the repeatable cognitive-perceptual-motor battery*. Grosse Pointe Park, MI: Axon.

LoDasso, G. L., Rapport, L. J., Axelrod, B. N., & Reeder, K. P. (1998). Intermanual and alternate-form equivalence on the Trail Making Tests. *Journal of Clinical and Experimental Neuropsychology*, *20*, 107–110.

McCracken, L. M., & Franzen, M. C. (1992). Principal-components analysis of the equivalence of alternate forms of the trail making test. *Psychological Assessment*, *4*, 235–238.

Oswald, W. D., & Roth, E. (1987). *Number Combination Test—Zahlen-Verbindungs-Test (ZVT)*. Goettingen: Hogrefe Verlag.

Reitan, R. M. (1979). *Trail Making Test—Manual for administration and scoring*. Tucson: Reitan Neuropsychology Laboratory.

Reppermund, S. (2008). *Neuropsychologische Charakterisierung depressiver Patienten im Verlauf unter besonderer Berücksichtigung exekutiver Funktionsstörungen*. Taunusstein: Verlag Dr. H.H. Driesen.

Rush, A. J., Gullion, C. M., Basco, M. R., Jarrett, R. B., & Trivedi, H. M. (1986). The inventory of Depressive Symptomatology (IDS). *Psychological Medicine*, *26*, 477–486.

Sobin, C., & Sackeim, H. A. (1997). Psychomotor symptoms of depression. *American Journal of Psychiatry*, *154*, 4–17.

Swanson, J. (2005). The Delis-Kaplan Executive Function System—a review. *Canadian Journal of School Psychology*, *20*, 117–128.

Tombaugh, T. N. (2003). Trail Making Tests A and B: Normative data stratified by age and education. *Archives of Clinical Neuropsychology*, *8*, 1–12.

Veiel, H. O. (1997). A preliminary profile of neuropsychological deficits associated with major depression. *Journal of Clinical and Experimental Neuropsychology*, *19*, 587–603.