Test–Retest Reliability of Two Attention Tests in Schizophrenia

Posen Lee1, Ping-Chia Li 1,*, Chin-Hsuan Liu 2, Ching-Lin Hsieh 3

1Department of Occupational Therapy, I-Shou University, Kaohsiung 824, Taiwan, Republic of China
2Department of Occupational Rehabilitation, Kai-Suan Psychiatric Hospital, Kaohsiung 802, Taiwan, Republic of China
3School of Occupational Therapy, College of Medicine, National Taiwan University and Department of Physical Medicine and Rehabilitation, National Taiwan University Hospital, Taipei 100, Taiwan, Republic of China

*Corresponding author at: No. 8, Yida Road, Yan Chao District, Kaohsiung City 824, Taiwan. Tel.: +886-7-6151100; fax: +886-7-6155150.
E-mail address: pingchia@mail2000.com.tw (P.-C. Li).

Accepted 21 April 2011

Abstract

The Symbol Digit Modalities Test (SDMT) and Digit Vigilance Test (DVT), both well-recommended attention tests for schizophrenia, are measures of switching and sustained attention, respectively. The purpose of this study was to investigate the test–retest reliability of the two attention tests in schizophrenia. A rater administered both tests on 147 participants with schizophrenia twice at a 1-week interval. Test–retest reliability was determined through the calculation of the intra-class correlation (ICC) coefficient. We also carried out the Bland–Altman analysis, which include a scatter plot of the differences between test and retest against their mean. System biases were evaluated by use of a paired t-test. The ICC for the SDMT was 0.87 and that for the DVT was 0.83. The limits of agreement (LOAs) of the SDMT and DVT were 11.5 to 29.9 correct responses and 156.3 to 224.2 s, respectively. The mean difference scores of the SDMT and DVT were 1.5 (4.7% of the first session mean; p = 0.002) and 246.4 (7.6% of the first session mean; p < 0.001). The ICCs show that the SDMT and DVT are stable measures across assessment in different sessions in schizophrenia. However, the paired t-test indicates a practice effect, and the LOAs show large variations. Thus, the SDMT and DVT are reliable for a group of subjects but limited for individual subjects with schizophrenia in 1-week interval clinical trials.

Keywords: Attention; Digit Vigilance Test; Disability evaluation; Reproducibility of results; Schizophrenia; Symbol Digit Modalities Test

Introduction

Impaired attention is a significant cognitive deficit that has consistently been identified as a core feature of schizophrenia (Egeland, 2007; Gur et al., 2007; Luck & Gold, 2008; Mogami, 2007; Shimizu et al., 2007). Deficit in attention is an intermediate characteristic in schizophrenia and may be a steady vulnerability indicator of schizophrenia (Liu, Chen, Chang, & Lin, 2000). Impaired attention is also a primary characteristic deficit and a major obstacle to daily functioning, so it is a particularly important target for remediation (Harris, Minassian, & Perry, 2007; Mausbach et al., 2008; Meyer & Blechert, 2005).

Sustained and switching attention have both been found to be significant core factors during treatment of patients with schizophrenia (Kenny & Meltzer, 1991). Sustained attention refers to the ability to maintain the attentional focus or to sustain alertness while processing high priority signals over time (Chan, Yip, & Lee, 2004). Sustained attention is associated with global community functioning (Prouteau et al., 2004), and previous studies have consistently reported impairments in sustained attention in patients with schizophrenia (Mirsy, Anthony, Duncan, Ahearn, & Kellam, 1991; Nieuwenstein, Aleman, & de Haan, 2001). Switching attention is the ability to maintain attention to a particular stimulus characteristic, such as its color or its shape, and then to inhibit attention to that characteristic in order to attend to a new characteristic when it becomes appropriate (Amos, 2002). To manage and monitor sustained and switching attention deficits, clinicians have to reliably assess the two types of attention impairment in schizophrenia.

© The Author 2011. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com. doi:10.1093/arclin/acr038 Advance Access publication on 22 May 2011
Reliability generally refers to the consistency of scores. Reproducibility is one of the classic approaches of reliability (Anastasi & Urbina, 1997; Gomm, 2008; Laenen, Vangeneugden, Geys, & Molenberghs, 2006). Researchers must demonstrate that instruments are reproducible, which is fundamental to repeated assessments by clinicians at their daily clinics (Bjorkly, Hartvig, Heggen, Brauer, & Moger, 2009). In recent reliability studies, intra-class correlation (ICC) methods have become popular for statistical analysis for reproducibility (Brennan & Silman, 1992). However, the ICC should be complemented by the Bland–Altman analysis, which can be used to show variation (or the magnitude of difference) of repeated assessments (Rankin & Stokes, 1998). In addition, paired t-test can be used to detect a systematic change in patients’ mean performance values (Ageberg, Flenhagen, & Ljung, 2007). Thus, the ICC, the Bland–Altman analysis, and the paired t-test should be used together to comprehensively examine the reliability of attention measures used in persons with schizophrenia in clinical and research settings.

The Symbol Digit Modalities Test (SDMT) and Digit Vigilance Test (DVT) are two well-recommended attention tests for assessing switching and sustained attention, respectively, in schizophrenia (Chan et al., 2004; Kelland & Lewis, 1996; Lewis, 1995; Sheridan et al., 2006; Smith, 1995). For persons with multiple sclerosis (MS), the SDMT is highly reliable, shows minimal practice effects (Drake et al., 2010), and is one of the most suitable tests for use over long testing intervals (Portaccio et al., 2010). Although both the SDMT and the DVT have been commonly used in patients with schizophrenia, their test–retest reliabilities have not been examined. Thus, the purpose of this study was to examine the test–retest reliability for application of the SDMT and DVT in schizophrenia.

Methods

Subjects

Subjects were recruited from a clinical psychiatric center. Participants with the following characteristics were included: Diagnosed with schizophrenia based on the Diagnostic and statistical manual of mental disorders, 4th ed. (American Psychiatric Association, 1994; Lasser et al., 2007); clinically stable; and maintained on a stable dose of antipsychotic medication for at least 1 month prior to the research testing sessions. The last two criteria were used to support the assumption for examining test–retest reliability (i.e., the attention performance of participants was assumed to be stable between two measurements). All patients were offered the opportunity to join the study during a visit to the clinic and all gave written informed consent before participating in the study. Patients were excluded from the study (a) if they were medically unstable, had other major diseases, or had secondary diagnoses of mental retardation, dementia, or neurological impairment and/or (b) if they were suffering from a severe medical or neurological condition or another psychiatric disorder that required treatment, were participating in a clinical trial, were suffering an episode of major depression, or exhibited difficulties with reading and writing. The study protocol was reviewed and approved by the Institutional Review Board of the study hospital.

Procedure

The SDMT and DVT were each administered twice, with the tests conducted 1 week apart, by a specially trained occupational therapist. The occupational therapist was blinded to the research purposes during the study period. Participants completed the two tests randomly. Patients’ demographic data were collected from medical records. The psychopathological stability of the participants with schizophrenia was assessed with the Clinical Global Impression (CGI) scale.

Measures of Attention

Symbol Digit Modalities Test. The SDMT assesses switching attention and requires directed visual-shifting and -pairing of specific digits with a set of pre-specified symbols within 90 s. Participants were instructed to perform the written version of the SDMT in this study. The total number of correct responses within the 90 s period was recorded for all participants (Smith, 1995).

Digit Vigilance Test. The DVT assesses sustained attention and is a cancellation test that requires rapid visual tracking of the digits “6” or “9” from 59 rows of single digits presented randomly on two separate sheets of paper. Total completion time (in seconds) was recorded separately for the two stimulus sheets. We used the version with “6” for this study. Scores were calculated for total time in this study. If a participant required more than 400 s to complete the first page, the second page was not administered. For these subjects, the time score of the first page was doubled to estimate the total time (Lewis, 1995).
Statistical Analyses

Test–retest reliability was determined through calculation of the ICC\textsubscript{2,1}. We calculated ICC\textsubscript{2,1}, a two-way random-effects single-measure reliability (absolute agreement) (Rankin & Stokes, 1998) using SPSS 13.0 for Windows. The ICC is the ratio of the inter-subject component of variance to the total variance (inter-subject variance + within-subject variance). There is no universally agreed level for ICC values in relation to levels of reliability, but the following scheme has been previously reported as acceptable: 0.90–0.99, high reliability; 0.80–0.89, good reliability; 0.70–0.79, fair reliability; 0.69 or below, poor reliability (Arnall, Koumantakis, Oldham, & Cooper, 2002). However, for group comparisons, the accepted minimal standards for reliability are 0.70, whereas 0.90–0.95 is used for individual comparisons (Scientific Advisory Committee of the Medical Outcome Trust, 2002).

We also conducted the Bland–Altman analysis, which included a scatter plot of the differences between test and retest values against their mean with 95% limits of agreement (LOAs) (i.e., mean difference ± 1.96 standard deviations [SDs] of the difference) (Bland & Altman, 1986). The plot shows the difference between test sessions 2 and 1 (2 \(- 1\)) against the mean of the two test sessions for each subject (Bland & Altman, 2003). Thus, it provides visual interpretation for ease of inspection of the size and range differences between the two sessions (Liaw et al., 2008). The acceptable magnitude of the difference is not a statistical decision, but a clinical one—specifically, not whether it confirms to some absolute, arbitrary criterion, but whether the agreement is good enough for a particular purpose (Bland & Altman, 1990).

Systematic biases between test–retest measurements of the SDMT and DVT were evaluated by paired \textit{t}-test (Atkinson & Nevill, 1998). The differences between the means of the two test sessions were less than 7% of the mean at first session for the SDMT and DVT, suggesting that the practice effect might be negligible (Salinsky, Storzbach, Dodrill, & Binder, 2001).

Data were analyzed with SPSS 13.0 (SPSS Inc.). The alpha level was set at 0.05 for all statistical tests and all \textit{p}-values are two-tailed.

Results

Demographic and Clinical Characteristics

We recruited 154 patients who were eligible for the study. However, seven of the original subjects did not complete the second session because they were discharged from the hospital, withdrew, or were otherwise lost to follow-up. Thus, the final sample size was 147 patients. These subjects tested into the same CGI category at both sessions.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>41.2</td>
<td>10.1</td>
</tr>
<tr>
<td>Onset age</td>
<td>25.0</td>
<td>6.7</td>
</tr>
<tr>
<td>Psychiatric history in years</td>
<td>16.2</td>
<td>9.4</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>49</td>
<td>33.3%</td>
</tr>
<tr>
<td>Men</td>
<td>98</td>
<td>66.7%</td>
</tr>
<tr>
<td>Education status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graduate school</td>
<td>1</td>
<td>0.7%</td>
</tr>
<tr>
<td>College</td>
<td>19</td>
<td>12.9%</td>
</tr>
<tr>
<td>Senior high school</td>
<td>70</td>
<td>47.6%</td>
</tr>
<tr>
<td>Junior high school</td>
<td>43</td>
<td>29.3%</td>
</tr>
<tr>
<td>Elementary school</td>
<td>14</td>
<td>9.5%</td>
</tr>
<tr>
<td>Schizophrenia subtypes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disorganized</td>
<td>18</td>
<td>12.2%</td>
</tr>
<tr>
<td>Catatonic</td>
<td>1</td>
<td>0.7%</td>
</tr>
<tr>
<td>Paranoid</td>
<td>106</td>
<td>72.1%</td>
</tr>
<tr>
<td>Residual</td>
<td>16</td>
<td>10.9%</td>
</tr>
<tr>
<td>Undifferentiated</td>
<td>6</td>
<td>4.1%</td>
</tr>
</tbody>
</table>

Note: SD = Standard Deviation.
Table 1 displays demographic data for the entire sample of 147 participants. The mean age of participants with schizophrenia was 41.2 years. The average duration of illness in the participants was 16.2 (SD = 9.4) years. The study sample included 98 men (66.7%), and approximately half of the subjects were senior high-school graduates.

Test–Retest Reliability

Table 2 shows that the ICC for the SDMT was 0.87 (95% CI 0.81–0.90) and that for the DVT was 0.83 (95% CI 0.72–0.89). The mean difference score between the two test sessions for the SDMT was 1.5 correct responses (4.7% of the first session mean) and that for the DVT was −46.4 s (7.6% of the first session mean). There were significant differences in between-session tests of the SDMT ($t = −3.2, p = .002$) and DVT ($t = 5.6, p < .001$).

Bland and Altman plots, used for the difference of scores against the mean score of the SDMT and DVT, are shown in Figs. 1 and 2. The LOAs of the SDMT and DVT were 11.5 to −9.9 correct responses and 156.3 to −249.2 s, respectively. Therefore, the width of the LOA was 21.4 correct responses for the SDMT and 405.5 s for the DVT. For the SDMT, LOA values were 11.4 correct responses (34.8% of the average of the two test sessions); for the DVT, LOA values were 202.9 s (34.8% of the average of the two test sessions).

Discussion

We found that the ICC values of two tests (SDMT: 0.87; DVT: 0.83) were high, indicating good test–retest reliability. For the SDMT, two previous studies have also shown good reliability in a 1-week assessment schedule in MS (Benedict, 2005; Benedict et al., 2004). In addition, the test–retest reliability of the DVT is high for total time (1-week interval) in healthy adults (Kelland & Lewis, 1996). These two tests are shown to have stability of data across two sessions in a 1-week interval in a clinical setting. However, the ICC values show that when applying the SDMT and DVT for individualized use, care should be exercised in monitoring an individual’s score.

Both the SDMT and the DVT showed a systematic bias (improvement) between test–retest sessions ($p = .002$ and $p < .05$). These findings indicate that practice effects may occur across two sessions in a 1-week interval. However, the mean difference of the SDMT test and retest was small (4.7%), suggesting that the difference might be negligible. No previous studies have shown practice effects of test–retest reliability of the SDMT (1-week interval) either. On the other hand, the mean test–retest difference for the DVT was substantial (7.6%), leading us to conclude that a practice effect should be a concern. In a previous study, an 8.1% practice effect of the DVT was found in healthy adults (1-week interval) (Kelland & Lewis, 1996). Therefore, clinicians must be cautious when interpreting a better score on a second evaluation of the DVT.

The practice effects may be related to the fact that the participants explicitly remembered test items presented previously or had been familiarized with the procedures of testing. It is possible to minimize the practice effects by using alternate forms of a test instead of the same form (Lemay, Bedard, Rouleau, & Tremblay, 2004). Three alternate forms have been designed to match the original SDMT in terms of mirrored symbol pairings, and symbols for 1 and 8, 2 and 9, and 3 and 5 are mirrored in each form (Hinton-Bayre & Geffen, 2005; Hinton-Bayre, Geffen, & McFarland, 1997). For the alternate administration of the DVT, 9s could be substituted for 6s in the instructions and sample demonstration (Smith, 1995). In addition, practice effects may be reduced if a practice session is allowed before the formal testing begins (Flansbjer, Holmback, Downham, & Lexell, 2005). For further research, without considering the practice effects, it is not possible to distinguish a true change that reflects the occurrence or resolution of an intervention.

Although practice effects occur, the investigation of the test–retest reliability of an instrument assumes that the construct being measured does not change over time and that all testing conditions are as constant as possible (Portney & Watkins, 2000). However, scores from repeated assessments are hardly the same in reality (Portney & Watkins, 2000; Schoettke, Bartram, & Wiedl, 1993). In the study, we recruited participants with stable medical conditions to avoid instability of attention deficits. Furthermore, the change of attention performance occurs when intervention is continually applied during the administration of attention test on people with schizophrenia (Wiedl, 2003). Thus, unreliable scores in research and clinical settings

<table>
<thead>
<tr>
<th>Test</th>
<th>First session mean (SD)</th>
<th>Second session mean (SD)</th>
<th>Paired differences Mean (SD) (95% CI)</th>
<th>ICC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDMT</td>
<td>31.9 (11.5)</td>
<td>33.4 (11.6)</td>
<td>1.5 (5.8) (0.6–2.5)</td>
<td>0.87 (0.81–0.90)</td>
</tr>
<tr>
<td>DVT</td>
<td>606.9 (197.2)</td>
<td>560.5 (184.9)</td>
<td>−46.4 (103.5) (−63.3 to −29.6)</td>
<td>0.83 (0.72–0.89)</td>
</tr>
</tbody>
</table>

Notes: SDMT = Symbol Digit Modalities Test; DVT = Digit Vigilance Test; SD = Standard Deviation; ICC = Intra-class Correlation.
appear to be unavoidable due to repeated measurements over time. Further studies are warranted to confirm our findings and to examine the practice effects of repeatedly administered instruments over different intervals (e.g., 2 weeks or a month).

The LOA width represents the unstable range into which 95% of the sample fell between the two test sessions (Bland & Altman, 1986). The values of LOA were quite large for the SDMT and DVT (SDMT is 11.4 correct responses, 34.8% of

**Fig. 1.** Bland–Altman plot of the difference of scores against the mean scores of the SDMT. The two dotted lines define the LOAs (mean of difference ± 1.96 SD).

**Fig. 2.** Bland–Altman plot of the difference of scores against the mean scores of the DVT. The two dotted lines define the LOAs (mean of difference ± 1.96 SD).
the average of the two test sessions; DVT is 202.9 s, 34.8% of the average of the two test sessions), showing that the results of the two test sessions were unstable. Thus, it would not be easy to tell whether the difference is true change or unstable measurements when using the SDMT and DVT in clinical settings (Bruton, Conway, & Holgate, 2000; Monaghan, Delahunt, & Caulfield, 2007).

It is now understood that using only the ICC to evaluate test–retest reliability cannot satisfy the needs of judgments in clinical and research use. Because this sample had larger variability, the values of ICC are inevitably higher. Reliability will appear to be higher when the ICC is applied to data from a group of individuals demonstrating a wide range of the measured characteristic than when applied to a narrow range of the same characteristic (Bruton et al., 2000; Flansbjer, Holmback, Downham, Patten, & Lexell 2005). Thus, we used a paired t-test to test systematic bias and Bland–Altman plots to indicate the unstable range between the two test sessions (Grafton, Foster, & Wright, 2005; Santos, Delisle, Lariviere, Plamondon, & Imbeau, 2008). In this study, the values of ICC were between 0.7 and 0.9. Although these values indicate good test–retest reliability, they did not reach the standards for explaining individual scores (0.90–0.95) (Scientific Advisory Committee of the Medical Outcome Trust, 2002). In addition, the paired t-test showed systematic biases between the two test sessions of both the SDMT and DVT; the Bland–Altman plots revealed an instability between the two test sessions. The reliabilities of the SDMT and DVT were not as high as indicated by the results of ICC. Thus, the SDMT and DVT are more suitable for use in measuring between-subjects than within-subjects variance.

Because we used a relatively small sample size, we admit that our findings may not be representative of all patients and should be generalized to other populations with caution. More studies are needed to cross-validate our findings. Additionally, as with all measures, this technique requires extensive examination to further clarify its particular strengths and limitations.

In summary, the results of this study indicate that the DVT and the SDMT have appropriate test–retest reliability to allow researchers to use them to distinguish between groups. We found a negligible statistically significant bias between the two test sessions of the SDMT and a substantial but small bias of the DVT were found. The Bland–Altman plots with LOA were large, revealing that the two tests have limitations in clinical settings in 1-week retest assessment in persons diagnosed with schizophrenia. Thus, our results indicate that the SDMT and DVT are reliable for groups of subjects, but have limited reliability in clinical trials of individual subjects with schizophrenia.

Conflict of Interest

None declared.

References


