Test-retest Reliability and Minimal Detectable Change of Chu’s Attention Test in Persons with Chronic Schizophrenia

Posen Lee, Ph.D., O.T.R.¹, Chin-Hsuan Liu, M.Ed., O.T.R.²,³
Hung-Yu Lin, Ph.D., O.T.R.¹, Yen-Lin Chen, Ph.D.³
Wen-Shian Lu, M.S., O.T.R.⁴, Ching-Lin Hsieh, Ph.D., O.T.R.⁵,⁶

Objective: Attention is the most fundamental cognitive function in daily life. Chu’s Attention Test (CAT) is one of the most widely used measures for assessing attention function in clinical psychiatric settings in Taiwan. The test-retest reliability and the minimal detectable change (MDC) of this measure remain largely unknown for patients with schizophrenia, limiting the interpretability and applicability of this measure. The purpose of this study was to investigate the test-retest reliability and the MDC of the CAT in patients with schizophrenia. Methods: One hundred and forty-seven patients with schizophrenia were recruited in this study. The participants completed the CAT at two sessions spaced one week apart. The intra-class correlation coefficient (ICC) was used to examine the test-retest reliability, and the MDC was calculated from the standard error of measurement. In addition, the percentage MDC (MDC%) was calculated. Results: The test-retest reliability of the CAT was excellent (p < 0.001) with difference of mean 1.9; and ICC was 0.95. The MDC (MDC%) was 12.5 (22.8%). Conclusion: Based on the results of this study, we suggest that the CAT is reliable in monitoring the attention function in patients with schizophrenia. These findings could help clinicians and researchers judge whether a change in the CAT between two repeated sessions is real for patients with schizophrenia.

Key words: attention, reliability, Chu’s Attention Test, schizophrenia

Introduction

Attention is the most fundamental cognitive function in everyday life [1, 2]. Schizophrenia is widely thought to involve impaired attention [3-5]. For patients with schizophrenia, deficit in attention is an intermediate characteristic and may be a steady vulnerability indicator [6]. Impaired attention interferes severely with successful social and occupational functioning in individuals with this mental illness [7]. Deficits in attention are the central deficits and major obstacles to daily functioning in patients with schizophrenia [8]. Measuring the attention status of those patients regularly in a clinical setting is common and important to provide the most suitable treatment programs. Therefore, a reliable attention measure is crucial in quantifying the attention level of patients with schizophrenia in clinical intervention and research.

Test-retest reliability refers to the reproducibility of repeated trials of a measure or the absence of random errors [9-12]. The intra-class correlation (ICC), reflecting both the degree of correlation and level of agreement between measures [13], is widely accepted to better represent the reliability of a measure than a value of Pearson’s r [14-17]. Random measurement error of a test can be used to calculate the minimal detectable change (MDC) [18-20]. The MDC is defined as the minimal threshold beyond random measurement error at a certain level of confidence (usually 95%) [21]. Namely, when a difference between two successive measurements exceeds the MDC, the change is more likely to be viewed as a real difference, rather than merely random error. The MDC can also be used as a threshold to judge whether the change in score on a measure in an individual patient indicates real improvement (or deterioration) or is just due to random measurement error [22]. Once the MDC is determined on a particular test for a given population, clinicians can interpret whether the change score for their patients is at or above the minimal level of detectable change [23, 24]. Thus, the MDC is crucial for clinicians and researchers to determine real change in repeated measurements for an individual patient [25].

Chu’s Attention Test (CAT) is one of the most widely used measures for assessing attention function in clinical psychiatric settings in Taiwan [26, 27]. The test-retest reliability of the test has been examined in persons with mental illness [28]. But the test-retest reliability and the MDC of the CAT in patients with schizophrenia remain largely unknown, a shortcoming which limits the measure’s interpretability and applicability. The purpose of this study was to examine the test-retest reliability and the MDC of the CAT when administered to patients with schizophrenia.

Methods

Study subjects

We recruited participants from a chronic ward at Kaohsiung Municipal Kai-Syuan Psychiatric Hospital in southern Taiwan. All the participants fulfilled inclusion criteria of being: (A) diagnosed with schizophrenia according to the International Classification of Diseases, Ninth Revision, Clinical Modifications (ICD-9-CM), excluding schizoaffective disorders; (B) clinically stable with a history of consistent doses of antipsychotic medications for more than one month before the first testing sessions of the study; (C) able to follow instructions; and (D) absent of substance abuse or other neurological deficits (e.g., mental retardation, dementia, or developmental disability).
The study protocol was approved by the institutional review board of the study hospital with the need of having signed informed consents from the study subjects.

**Study procedures**

The CAT was administered at two sessions one week apart by a specially trained occupational therapist. All participants completed attention measures by themselves according to the therapist’s instructions. The therapist, who was blind to the purpose of the study during the study period, followed the standard procedure and gave the instructions according to the manual of the CAT [28]. The test was administered by group, and each group contained no more than five participants. We collected information of participants’ demographic characteristics from medical records. The psychological stability of the persons with schizophrenia was assessed using the Clinical Global Impression (CGI) scale.

**Instrument**

The CAT assesses sustained attention and is a standard evaluation tool frequently used in Taiwan [29]. The participants look at a series of scrambled codes and search for the symbol “*” randomly distributed among 00 to 99 items within 10 minutes [28]. The total numbers of correct symbol counts within a 10-min period were recorded for all participants [29].

**Statistical analysis**

Test-retest reliability was determined through calculation of the ICC_{(2,1)}, a two-way random-effects single-measure reliability (absolute agreement) [30]. The ICC is the ratio of the inter-subject component of variance to the total variance (inter-subject variance + within-subject variance) [17]. 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, excellent reliability; 0.80-0.89, good reliability; 0.70-0.79, fair reliability; 0.69 or below, poor reliability [31]. The ICC scores of CAT were computed for the test and re-test sessions [32].

The MDC was calculated based on SEM according to the equations below [22]:

$$\text{SEM} = \frac{\text{SD}}{\sqrt{2}} \times \sqrt{(1-r)}$$

$$\text{MDC} = 1.96 \times \sqrt{2} \times \text{SEM}$$

In these formulae, the 1.96 is the z-score at 95% confidence level, the $\sqrt{2}$ is used because of the underlying extra uncertainty during measurement at 2 time points, and r is the coefficient of the test-retest reliability, which was represented by ICC. In addition, we calculated the MDC% (= MDC/mean × 100%), which presents the relative amount of random measurement error. The ‘mean’ in this equation is the mean score of all trials. An MDC% of 30% or less is considered acceptable, and one of 10% or less, excellent [33].

We used a paired t-test in this study to examine whether systematic biases existed. The differences between the groups were considered significant if p-values were smaller than 0.05.

Bland-Altman plots were used to visually examine the agreement of a test by plotting the difference scores against the mean score of each pair of measurements [34]. Assuming the differences follow normal distribution, 95% of the differences (limits of agreement, LOA) should lie between $d \pm 1.96 \times \text{SD}$, where $d$ represents the mean difference between test and retest scores, and SD is the standard deviation of differences of each pair [35, 36]. In addition, we used Pearson’s r to examine the association between the absolute difference and the mean of each pair of repeated measure-
ments to examine the possibility of heteroscedasticity, meaning that a systematic trend (e.g., the higher the scores, the larger the differences) exists. If heteroscedasticity exists, the MDC should not apply for different levels (i.e., attention deficits in this study) of patients. According to Atkinson’s suggestions, if Pearson’s $r$ is greater than 0.3, the data are heteroscedastic [9].

Data were analyzed using Statistical Package for Social Science version 13.0 software (SPSS Inc., Chicago, Illinois, USA). The alpha level was set at 0.05 for all statistical tests and all $p$-values were two-tailed.

### Results

We recruited 154 eligible and voluntary patients in this study. But seven patients did not complete the whole process and were excluded from the data analysis. Table 1 shows the demographic data of the 147 remaining subjects. These participants were tested into the same CGI category at both sessions. Most participants’ CGI scores were either mild (score = 3, 49.0%) or borderline (score = 2, 28.6%). All participants confirmed that they had continued with their regular activities during the interval between the two sessions in this study. Table 2 lists the reliability indices of the CAT [33]. Figure 1 depicts the differences in scores between two successive sessions of the CAT, plotted against the mean scores shown in the Bland-Altman plot.

### Discussion

In a clinical setting, the MDC is crucial to determining whether a difference in an individual patient is real or merely due to random error, to

<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>Female</td>
<td>49</td>
<td>(33.3)</td>
</tr>
<tr>
<td>Male</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>

SD, standard deviation

Table 1. Demographic characteristics of the sample (n = 147)
Table 2. Reliability indices of Chu’s Attention Test (n = 147)

<table>
<thead>
<tr>
<th></th>
<th>First test Mean (SD)</th>
<th>Second test Mean (SD)</th>
<th>Difference Mean (SD)</th>
<th>ICC (95% CI)</th>
<th>SEM</th>
<th>MDC (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>53.6 (20.3)</td>
<td>55.5 (19.9)</td>
<td>1.9 (5.9)***</td>
<td>0.95 (0.93-0.97)</td>
<td>4.5</td>
<td>12.5§ (22.8)</td>
</tr>
</tbody>
</table>

*** Significantly different (p < 0.001, in paired t-test), indicating excellent test-retest reliability

§ Representing acceptable random measurement error [33]

ICC, intraclass correlation coefficient; CI, confidence interval; SEM, standard error or measurement; MDC, minimal detectable change

Figure 1. Bland-Altman method for plotting the difference in scores against the mean scores of Chu’s Attention Test. The 2 bold lines define the limits of agreement (mean of difference ± 1.96 SD). The LOA of the CAT ranged from 13.5 to -9.6. The association between the absolute difference and mean of each pair of repeated measurements were low (Pearson’s r = 0.15). SD, standard deviation.

Mean, 1.9; SD, 5.9; Mean + 1.96 SD, 13.5; Mean - 1.96 SD = -9.6
revise treatment programs. To our knowledge, this study is the first to report the MDC of the CAT when administrated to patients with schizophrenia. The attention function is closely related to one’s daily functioning abilities [8, 37, 38], and the CAT is popularly used in clinical psychiatric settings in Taiwan. Our results could help both clinicians and researchers monitor patients’ attention levels by judging individual differences between two successive measurements with the CAT more correctly and confidently.

As shown in Table 2, we found that the CAT had excellent test-retest reproducibility ($p < 0.001$, in paired $t$-test) with an ICC value of 0.95, which is higher than that in a previous study (Pearson’s $r$ = 0.84) [28]. In addition, the Bland-Altman plots (Figure 1) showed only negligible deviation from 0, indicating no salient systematic differences (e.g., practice effect) in scores between the two sessions of assessment. In addition, the Pearson’s $r$ (Figure 1) between the absolute difference score and the mean score of each pair of successive tests was 0.15, implying no existence of heteroscedasticity [9]. This study result suggests that we could use a fixed MDC value as the threshold to judge whether a difference between repeated measurements is real regardless of the different initial test score (attention deficits) [10]. These results support the reliability of the CAT in monitoring the changes of attention performances of patients with schizophrenia over time when administered by trained raters.

The MDC of the CAT in patients with schizophrenia was 12.5 points (Table 2). This finding means that in a clinical setting, it is necessary for the score of a patient with schizophrenia to increase (or decrease) by more than 13 points in successive administrations of the CAT to be interpreted as indicating real improvement (or deterioration) with 95% confidence. Thus, our results can help clinicians interpret a change in the CAT more reasonably.

From a statistical viewpoint, the MDC can also be seen as a threshold to determine a significant difference in an individual patient [37, 39]. Namely, if an individual patient improves by more than the MDC, then the improvement can be regarded as significant. Thus, the MDC can be used to determine if the individual patient has made significant improvement in a clinical setting.

Researchers can use the MDC as a threshold to report the proportion of a research sample that has attained real improvement, or deterioration. In general, a significant mean change score of a study sample does not guarantee that all the members in the sample achieved significant change. Thus, reporting the proportion of a study sample that achieved or exceeded the MDC can help researchers translate their study results to the clinical context and enhance the application of the study outcomes.

The MDC is an important characteristic for determining whether an individual change between two successive test administrations exceeds measurement error or is significant. But the minimal clinically important difference (MID) which represents a meaning difference to the patient, is also important in decision-making and treatment planning in clinical settings. To further enhance the use of the CAT, future research is warranted to investigate the MID of the test when administrated to patients with schizophrenia.

**Limitations of the study**

The readers are warned against over-interpreting the study results because this study has two limitations:

- The significance of the paired $t$-test indicates that systematic bias existed. This bias might have resulted from a practice effect, which
Reliability of Chu’s Attention Test

could have affected the accuracy of our results. We further calculated the effect size \( r \) between the two repeated measurements according to the following equation: 
\[
r = \frac{\text{mean}_2 - \text{mean}_1}{\sqrt{(s_1^2 + s_2^2)/2}}
\]
[40]. The value of the effect size between the successive tests in this study was 0.1, which is trivial, according to Cohen’s suggestion [40]. Thus, the systematic bias does not appear to threaten our results.

Because the participants were recruited by convenient sampling, nearly two-thirds were male patients who had chronic, relatively stable conditions. This homogeneity may limit the generalization of our study results. Future studies could recruit participants with more equal distributions of gender and illness condition to further validate our findings.

Summary

Our study was found that the CAT had excellent test-retest reliability and acceptable random measurement error, indicating that the CAT is reliable for monitoring the attention function in patients with schizophrenia. Our results could help both clinicians and researchers judge whether a difference in repeated measurements in the CAT is real to enhance the interpretability and applicability of the test.

Acknowledgements

We thank the staff at Kaohsiung Municipal Kai-Syuan Psychiatric Hospital for assistance with the data collection. All authors declare no potential conflict of interest in writing this report.

References


33. Smidt N, van der Windt DA, Assendelft WJ, et al.: Interobserver reproducibility of the assessment of se-


