Evaluation on Long-Term Test–Retest Reliability of the Short-Form Childhood Trauma Questionnaire in Patients with Schizophrenia

Zhibiao Xiang1–3
Zhening Liu1–3
Hengyi Cao4,5
Zhipeng Wu1–3
Yicheng Long1–3

1Department of Psychiatry, The Second Xiangya Hospital, Central South University, Changsha, Hunan, People’s Republic of China; 2Mental Health Institute of Central South University, Changsha, Hunan, People’s Republic of China; 3China National Clinical Research Center on Mental Disorders, Changsha, Hunan, People’s Republic of China; 4Center for Psychiatric Neuroscience, Feinstein Institute for Medical Research, Hempstead, NY, USA; 5Division of Psychiatry Research, Zucker Hillside Hospital, Glen Oaks, NY, USA

Background: Many studies have reported an association between childhood trauma exposure and schizophrenia. Among these studies, the Short-form Childhood Trauma Questionnaire (CTQ-SF) is one of the most widely used measures of childhood trauma. However, little is known regarding the long-term reliability of the CTQ-SF, especially in patients with psychopathology.

Methods: The CTQ-SF was administered to 50 patients diagnosed with schizophrenia from a hospital in Changsha, Hunan, China. These patients were asked to re-complete the CTQ-SF when they were re-hospitalized or received outpatient treatments in the same hospital within 4 years of follow-up. Intraclass correlation coefficient (ICC) was used to assess test–retest reliability of the CTQ-SF over the intervals. Associations of the CTQ-SF with the Positive and Negative Syndrome Scale (PANSS) and Wechsler Adult Intelligence Scale (WAIS) were tested using Spearman correlation coefficients.

Results: Among the participants, 35 (70.0%) patients re-completed the CTQ-SF after an interval averaging 11.26 months. Excellent test–retest reliabilities (with ICC > 0.75) were found for the total CTQ-SF score (ICC = 0.772) as well as scores of the emotional abuse (ICC = 0.808), physical abuse (ICC = 0.756), sexual abuse (ICC = 0.877) and physical neglect (ICC = 0.751) subscales. Meanwhile, a moderate test–retest reliability was found for the emotional neglect subscale (ICC = 0.538). At both baseline and follow-up, no significant correlations (p > 0.05) were found between CTQ-SF scores and any other clinical assessments.

Conclusion: Our results suggest that CTQ-SF is reliable to assess childhood trauma exposures in schizophrenia over relatively long intervals, regardless of patients’ current symptoms and states of cognition.

Keywords: childhood trauma questionnaire, childhood adversity, childhood trauma, schizophrenia, test–retest reliability

Introduction

Schizophrenia is a serious mental disorder characterized by a range of symptoms including delusions and hallucinations,1 and is associated with alterations in brain structure2 and function.3,4 Several studies have reported an association between childhood trauma exposure and schizophrenia, in which patients with schizophrenia report higher levels of childhood trauma compared to the general population.5,6 Moreover, some of the brain alterations observed in schizophrenia have been associated with childhood trauma histories, including altered gray matter volume patterns (eg decreased hippocampus volumes)7–10 as well as disrupted brain
function (e.g., decreased connectivity and decreased stability within multiple brain areas) during rest or specific tasks. These results highlight the important role of childhood trauma in the development of schizophrenia, as a part of the integrative “bio-psycho-social” model.

In the field of childhood trauma research, it is common to assess childhood trauma by self-reported questionnaires. Among these questionnaires, the Short-form Childhood Trauma Questionnaire (CTQ-SF) is one of the most widely used and well-studied tools. The CTQ-SF consists of five subscales to assess five subtypes of childhood trauma: emotional abuse, physical abuse, sexual abuse, emotional neglect, and physical neglect. Previous studies have demonstrated the validity and overall good test-retest reliability of CTQ-SF in patients with schizophrenia when tested over a period of two or four weeks.

Despite these findings, it is possible that results of retrospective self-reports such as the CTQ-SF may be influenced by memory lapses and participants’ psychopathological status. Factors such as normal processes of memory loss in ageing, change of mood status, subsequent events, and a need to justify or understand mental illness may influence the reliability of such reports, especially during long-term follow-up. For example, previous studies have indicated that self-reported data can be biased by forgetting even with a short recall period of one month. Therefore, it is necessary to verify whether CTQ-SF scores are robust and reliable over longer time scales despite the influence of these various factors. However, research regarding the test-retest reliability of CTQ-SF in schizophrenia patients with long-term follow-up periods (i.e., two months or longer) is still lacking.

Therefore, the present study aimed to examine test-retest reliability of CTQ-SF in patients with schizophrenia over long terms (defined as over two months), in order to investigate if the CTQ-SF scores would be influenced by patients’ normal processes of forgetting, as well as changed psychopathological status such as the severity of psychiatric symptoms and cognitive deficits over time. The results were anticipated to provide evidence for the validity of CTQ-SF in the studies of schizophrenia and other mental disorders.

Materials and Methods

Participants

This study included 50 schizophrenia patients, who were recruited from the Department of Psychiatry, Second Xiangya Hospital of Central South University, Changsha, Hunan, China. The inclusion criteria for the participants were as follows: 1) met the diagnostic criteria for schizophrenia according to the Diagnostic and Statistical Manual of Mental Disorders, fifth edition (DSM-V); 2) between 14 and 35 years old; 3) no history of psychotropic substance abuse; and 4) able to understand and complete all questions of the CTQ-SF. All participants were followed-up from December, 2016 to January, 2021. During the follow-up, participants were excluded from the study if they had any of the following situations: 1) change in diagnosis; 2) contact was lost; and 3) participants had severe physical illnesses or extensive declines in cognition, making them unable to complete the assessments at follow-up.

Procedures and Assessments

All participants completed the CTQ-SF at baseline and were followed longitudinally. They were asked to re-complete the CTQ-SF when they were re-hospitalized or received outpatient treatments in the same hospital after a two-month or longer period within 4 years of follow-up. Besides the CTQ-SF, several other clinical assessments were also performed at both baseline and follow-up to evaluate the participants’ clinical symptoms and cognitions. These assessments are described as below:

1. CTQ-SF: the CTQ-SF consists of 25 clinical items which are grouped into five subscales (5 items each): emotional abuse, physical abuse, sexual abuse, emotional neglect, and physical neglect to assess such five types of childhood trauma. Each of the 25 items are 5-point Likert-type questions, rated from 1 (never) to 5 (very often) according to the frequency of each event. Thus, total scores range from 25 to 125, while the scores of each subscale range from 5 to 25. Higher scores indicate higher levels of childhood trauma exposures. In addition to the clinical items, all participants’ minimization/denial (MD) scores were also calculated referring to a previous work.

2. Positive and Negative Syndrome Scale (PANSS): for all patients, severity of their current clinical symptoms during the past two weeks was assessed using the PANSS. This was performed through face-to-face interviews with psychiatrists who were licensed as medical practitioners in China.

3. Wechsler Adult Intelligence Scale (WAIS): all participants finished the digit symbol (WAIS-DS) and information (WAIS-I) subtest of the WAIS Chinese version, which is one of the most widely used measure of
cognitive function in China,\textsuperscript{2,38,39} to assess their current cognitive function upon completion of the CTQ-SF. The WAIS-DS and WAIS-I are assessments of two important domains of cognition: measuring processing speed and verbal comprehension, respectively.\textsuperscript{2,37}

**Statistics**

Differences in demographic and clinical variables between different subsamples were tested using a \textit{t}-test (for age and years of education), Chi-square test (for sex) and the Mann–Whitney \textit{U}-test (for all clinical scales). Test–retest reliability of the CTQ-SF was evaluated by applying two-way random effects intra-class correlation coefficients (ICC).\textsuperscript{40} ICCs were calculated for both total score and each of the subscale scores of the CTQ-SF. Based on ICCs, the test–retest reliabilities can be interpreted as follows: poor reliability (ICC < 0.4), moderate reliability (0.4 ≤ ICC < 0.6), good reliability (0.6 ≤ ICC < 0.75), and excellent reliability (ICC ≥ 0.75).\textsuperscript{41,42}

An additional analysis was carried out to investigate if the CTQ-SF scores were influenced by participants’ current clinical symptoms or cognitions. Here, we tested possible associations between the CTQ-SF scores and PANSS/WAIS scores using the Spearman correlation coefficients. This analysis was performed at both baseline and follow-up.

All analyses were performed using the SPSS software (version 26.0) for MAC, and statistical significance was set at an alpha level of 0.05 (two-sided) for all tests.

**Exploratory Analysis**

An exploratory analysis was performed to investigate possible influences of follow-up interval length on results. Here, all participants were divided into two subgroups considering the distribution of follow-up intervals in the current sample (averaging 11.26 months): those with an interval shorter than 12 months (n = 23), and those with an interval longer than or equal to 12 months (n = 12); ICCs of the CTQ-SF scores were then calculated for each subgroup separately.

**Results**

**Participant Characteristics**

Among the initial sample of 50 patients, a total 15 participants were excluded during follow-up: 4 patients were excluded because their diagnoses were changed to depression or bipolar disorder; 6 patients were loss of contact; 5 patients had extensive declines in cognition which make them unable to re-complete the assessments. The age range of the participants who re-completed the CTQ-SF (n = 35) were from 14 to 28 years at baseline, with 15 males and 20 females (Table 1). The test–retest intervals ranged from 2 to 41 months (11.26 months on average). There were no significant differences in age, sex, education level, CTQ-SF total score, or the scores of PANSS and WAIS at baseline between those who were followed up and those who were not (Supplementary Table 1).

**Changes Over Time in Clinical Variables**

Comparisons of the results of the CTQ-SF and other clinical assessments for the followed participants at baseline and follow-up are summarized in Table 1. There were no significant differences in the CTQ-SF total score (\(p = 0.659\)), or scores of subscales including the MD score (all \(p > 0.05\)). No significant differences were found in the WAIS-DS/WAIS-I scores (both \(p > 0.05\)). Compared with baseline, the participants had significantly lower PANSS scores at follow-up (\(p = 0.045\)).

**Test–Retest Reliabilities**

As shown in Table 2, the Cronbach’s \(\alpha\) coefficient of the CTQ-SF total score was 0.871, and the \(\alpha\) coefficients for its five subscales ranged from 0.700 (physical neglect) to 0.934 (sexual abuse), suggesting a good internal consistency of the scale. The ICC of the CTQ-SF total score was 0.772, which suggests an excellent test–retest reliability. Excellent test–retest reliabilities (with ICC > 0.75) were also found for scores of the emotional abuse (ICC = 0.808), physical abuse (ICC = 0.756), sexual abuse (ICC = 0.877) and physical neglect (ICC = 0.751) subscales, while a moderate test–retest reliability was found for the emotional neglect subscale (ICC = 0.538). A poor test–retest reliability (ICC = 0.333) was found for the MD score.

**Relationships with Clinical Symptoms and Cognitions**

As shown in Tables 3 and 4, for both baseline and follow-up, no significant correlations (\(p > 0.05\)) were found between the CTQ-SF total score and any other clinical/cognitive assessments including the PANSS, WAIS-I and WAIS-DS. No significant correlations (\(p > 0.05\)) were found between any of the CTQ-SF subscale scores (including the MD score) and any other assessments.
Exploratory Analysis

As shown in Table 5, a higher test–retest reliability of CTQ-SF was shown in participants with a follow-up interval less than 12 months (ICC = 0.853) compared to those with an interval longer than 12 months (ICC = 0.504). The same trends of results (lower ICCs in those with a longer follow-up interval) were observed for all the subscales of CTQ-SF (see Table 5).
Table 4 Correlations Between the CTQ-SF Scores and Other Clinical/Cognitive Assessments at Follow-Up, with Spearman’s Rho and p values Reported

<table>
<thead>
<tr>
<th>CTQ-SF Scores</th>
<th>PANSS</th>
<th>WAIS-DS</th>
<th>WAIS-I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total score</td>
<td>Rho = 0.089, p = 0.613</td>
<td>Rho = −0.196, p = 0.258</td>
<td>Rho = −0.028, p = 0.875</td>
</tr>
<tr>
<td>Emotional abuse</td>
<td>Rho = −0.108, p = 0.538</td>
<td>Rho = −0.232, p = 0.180</td>
<td>Rho = −0.181, p = 0.299</td>
</tr>
<tr>
<td>Physical abuse</td>
<td>Rho = −0.015, p = 0.932</td>
<td>Rho = −0.209, p = 0.228</td>
<td>Rho = −0.181, p = 0.298</td>
</tr>
<tr>
<td>Sexual abuse</td>
<td>Rho = −0.106, p = 0.544</td>
<td>Rho = 0.157, p = 0.367</td>
<td>Rho = 0.318, p = 0.063</td>
</tr>
<tr>
<td>Emotional neglect</td>
<td>Rho = 0.086, p = 0.625</td>
<td>Rho = −0.002, p = 0.990</td>
<td>Rho = 0.181, p = 0.298</td>
</tr>
<tr>
<td>Physical neglect</td>
<td>Rho = 0.239, p = 0.167</td>
<td>Rho = 0.326, p = 0.056</td>
<td>Rho = −0.307, p = 0.073</td>
</tr>
<tr>
<td>MD score</td>
<td>Rho = −0.061, p = 0.729</td>
<td>Rho = 0.292, p = 0.089</td>
<td>Rho = 0.057, p = 0.745</td>
</tr>
</tbody>
</table>

Table 5 Exploratory Analysis on ICCs of the CTQ-SF Scores in Subsamples with Different Lengths of Follow-Up Intervals

<table>
<thead>
<tr>
<th>CTQ-SF Scores</th>
<th>ICC (with 95% Confidence Intervals)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Interval &lt; 12 Months (n = 23)</td>
</tr>
<tr>
<td>Total score</td>
<td>0.853* (0.68, 0.94)</td>
</tr>
<tr>
<td>Emotional abuse</td>
<td>0.890* (0.76, 0.95)</td>
</tr>
<tr>
<td>Physical abuse</td>
<td>0.713* (0.44, 0.87)</td>
</tr>
<tr>
<td>Sexual abuse</td>
<td>0.885* (0.75, 0.95)</td>
</tr>
<tr>
<td>Emotional neglect</td>
<td>0.659* (0.35, 0.84)</td>
</tr>
<tr>
<td>Physical neglect</td>
<td>0.827* (0.63, 0.92)</td>
</tr>
<tr>
<td>MD score</td>
<td>0.294 (−0.13, 0.62)</td>
</tr>
</tbody>
</table>

Notes: *Indicates significance at the 0.05 level.

Discussion

In this study, we investigated the long-term test–retest reliability of CTQ-SF in a sample of schizophrenia patients over a relatively long interval of 11 months on average. Generally, our results indicated excellent long-term test–retest reliability of the CTQ-SF total score as well as most of its subscales. Meanwhile, at both baseline and follow-up, no significant correlations were found between scores of the CTQ-SF and assessments of clinical symptoms or cognition. These results provide important evidence for validating the use of CTQ-SF in schizophrenia and may also point to its value in assessing other mental health populations.

As one of the most widely used assessments of childhood trauma in clinical research, previous studies have demonstrated that CTQ-SF is valid and has overall good long-term test–retest reliabilities in both healthy participants and multiple psychiatric populations. For example, in a study conducted by Cammack et al, the test–retest reliability of CTQ-SF was good when it was completed by 247 adult women twice with an interval between 18.1 and 369.9 weeks (ICCs > 0.7 for all subscales). In another study conducted by Shannon et al, the CTQ-SF was highly reliable (ICCs > 0.8 for 4 of the 5 subscales) when retested over 18 months in 60 patients with bipolar disorder. Clinical and neuroimaging studies have suggested the important role of childhood trauma in schizophrenia; however, little was known about the long-term test–retest reliability of CTQ-SF in schizophrenia patients. Our findings here may fill this knowledge gap by presenting the first investigation to our knowledge of CTQ-SF reliability in a long-term longitudinal schizophrenia population.

Stabilities of retrospective assessments, such as the CTQ-SF, generally rely on the accuracy of people’s memories. Schizophrenia is known to be associated with certain types of psychopathology including cognitive impairments and delusional symptoms which may lead to memory distortions. Our results showed, however, that no significant relationships existed between the reported CTQ-SF scores and current cognitive functions or psychotic symptoms in patients with schizophrenia (Tables 3 and 4). We thus preliminarily concluded that CTQ-SF is a temporally stable and state-independent assessment of childhood maltreatment in schizophrenic subjects, which is less likely to be influenced by patients’ current clinical symptoms and cognitions.

Among five subscales of the CTQ-SF, the sexual abuse and emotional neglect subscales showed the highest and
lowest test–retest reliability, respectively (Table 2). Our results are highly consist with a previous study conducted in healthy adults in the Netherlands: in such study, it was found that sexual abuse was most reliably reported during a 6-year follow-up; meanwhile, the emotional neglect subscale was found to be relatively unreliable, which may partly due to the impacts of participants’ emotional states such as depression as proposed by the authors.49 This may also support a cross-cultural generalizability of our results, indicating that contents of the emotional neglect dimension in CTQ-SF could be problematic in some ways. Nevertheless, the ICCs of all subscales were higher than 0.5 (Table 2), which suggest at least moderate reliability and generally consistent reporting.43

Similar to other studies,43,49,51 the assessments were done across varying follow-up intervals (ranging from 2 to 41 months) in the present study. To explore possible influences of interval length, we calculated ICCs in two subgroups with different intervals (<12 months and ≥12 months) separately, and lower ICCs were found in the subgroup with longer intervals (Table 5). This highlights the possibility that test–retest reliability of the CTQ-SF might decrease over longer time intervals. However, we should remain cautious in drawing such a conclusion considering the small sample size (n = 12 only) and wide confidence intervals for the subgroup with intervals longer than 12 months (see Table 5).

Another issue to note is that although most published studies reporting on the CTQ’s reliability did not mention the MD score,52–54 here we include the MD score in this study for its importance suggested by some recent work.35 We found that the MD score has a low test–retest reliability (ICC < 0.4, Table 2). However, such a conclusion may need to be verified in a larger sample considering the wide confidence interval for the MD score’s ICC (Table 2), too.

There are several minor points to note in the current study. First, as mentioned above, the sample size is relatively small and future studies are warranted to replicate out results in a larger sample. Second, during follow-up, patients were excluded when they had extensive declines in cognition and were unable to re-complete the CTQ-SF. This might cause a selection bias, and our conclusions may not be extrapolated to patients with poorer outcomes. Last, the follow-up interval is only 11 months on average in this study, and we are unable to know if the CTQ-SF results would be stable over a longer interval or even lifetime.

Conclusion
In conclusion, in this study, we investigated the long-term test–retest reliability of the CTQ-SF in patients with schizophrenia over an interval of 11 months on average. The results suggested an excellent reliability of the CTQ-SF total score, and at least moderate reliabilities for all of its subscales. Moreover, no significant associations were found between the CTQ-SF and assessments of clinical symptoms or cognitions. Our results suggest that CTQ-SF is reliable to assess childhood trauma exposures in schizophrenia over relatively long time intervals, regardless of patients’ current symptoms and cognitive status. However, future studies are warranted to replicate out results in a larger sample and to investigate whether the CTQ-SF would keep stable over longer time intervals.

Ethics Statement
The study was performed in accordance with the Declaration of Helsinki. The study was approved by the Ethics Committee of Second Xiangya Hospital in Changsha, and written informed consent was obtained from all participants or guardians.

Acknowledgments
This work was supported by the China Precision Medicine Initiative (grant number 2016YFC0906300) and the National Natural Science Foundation of China (grant numbers 81671335, 81701325, 81801353 and 82071506). We thank Brendan Ross from McGill University for his excellent help in polishing the English language.

Disclosure
The authors declare that they have no conflicts of interest in this work.

References


