

# Assessing Long-Term Test-Retest Reliability of the CPT-IP in Schizophrenia

Eric Hahn<sup>1\*</sup>, Andrea Vollath<sup>1</sup>, Tam T. M. Ta<sup>1</sup>, Constanze Hahn<sup>2</sup>, Linn K. Kuehl<sup>1</sup>, Michael Dettling<sup>1</sup>, Andres H. Neuhaus<sup>1</sup>

**1** Department of Psychiatry and Psychotherapy, Charité-University Medicine, Campus Benjamin Franklin, Berlin, Germany, **2** Department of Biopsychology, Ruhr University, Bochum, Germany

## Abstract

**Background:** The Continuous Performance Test-Identical Pairs version (CPT-IP) is a well-established measure of sustained attention, and its more challenging versions are particularly suited to detect subtle processing deficits in patients with schizophrenia. However, while there are few longitudinal samples for the CPT-IP, no study has addressed stability for more than two months in patients with schizophrenia. Assessing long-term test-retest reliability of the CPT-IP would facilitate the ability of clinicians to draw conclusions from studies involving interventions as long term cognitive or pharmacological treatments. The present study assessed 12 month test-retest reliability for the two most challenging versions of CPT-IP (4-digit and shapes) in a matched sample of clinically stable schizophrenia outpatients and healthy controls.

**Methods:** Fifty clinically stable schizophrenia outpatients and 50 healthy controls were assessed with the CPT-IP for the 4-digit and shape conditions. From these, 40 patients and 47 controls were reassessed with an average interval of 12.3 months between test sessions. Test-retest reliability was analyzed with Pearson correlations and results were compared with previous data involving healthy controls and short-term studies in patients with schizophrenia.

**Results:** Especially  $d'$  and hit rate discriminated well between patients with schizophrenia and healthy controls for both CPT-IP conditions and at both test sessions. Healthy controls demonstrated sufficient long term test-retest correlations of  $d'$ , hit rate and reaction time for both the 4-digit and shape conditions. However, in schizophrenia patients, long-term reliability correlations were at best moderate for  $d'$  and hit rate only.

**Conclusions:** The current study provides further evidence that  $d'$  and hit rate yield consistent cross-sectional discrimination sensitivity. At best moderate long-term test-retest reliability of  $d'$  in schizophrenia outpatients may be not sufficient for practical use of this measure in long term clinical trials.

**Citation:** Hahn E, Vollath A, Ta TTM, Hahn C, Kuehl LK, et al. (2014) Assessing Long-Term Test-Retest Reliability of the CPT-IP in Schizophrenia. PLoS ONE 9(1): e84780. doi:10.1371/journal.pone.0084780

**Editor:** Yu-Feng Zang, Hangzhou Normal University, China

**Received:** June 7, 2013; **Accepted:** November 19, 2013; **Published:** January 8, 2014

**Copyright:** © 2014 Hahn et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** These authors have no support or funding to report.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: eric.hahn@charite.de

## Introduction

Attention deficits are well established symptoms of patients with schizophrenia that are associated with functional outcome in longitudinal studies and are considered valid predictors of impaired cognitive functions in schizophrenia [1,2]. Today the Continuous Performance Test (CPT) has emerged as the standard measure of deficits in sustained attention and such deficits are among the most robust cognitive measures that may predict the onset of schizophrenia spectrum disorders in at-risk individuals [3] and may differentiate between schizophrenia, depression and bipolar disorder [4,5].

The Identical Pairs CPT (CPT-IP) [4,6] especially in its more challenging versions was specifically designed to detect subtle processing deficits of sustained attention at early stages of schizophrenia and in at-risk individuals. Here subjects have to respond to a second stimulus in any pair of identical stimuli that may include 2, 3 or 4 digits and shapes, which activate the verbal

and spatial attentional neuronal systems [7]. Especially in its more challenging versions CPT-IP performance deficits were also found in prodromal stages of schizophrenia, in first episode neuroleptic-naive schizophrenic patients [8], in individuals within the schizophrenia spectrum, including schizotypal personality traits [9–11], and in asymptomatic first-degree relatives of patients with schizophrenia [12].

In contrast to strong evidence for CPT-IP task validity, data on its long-term reliability in independent samples is scarce. While good short and long-term test-retest reliability was assessed for healthy controls [6,13], studies investigating this important issue in schizophrenia have been conducted only for short-term, i.e. 2-week [14] or 4-week intervals [15]. A recent study aimed at standardization of the CPT-IP in schizophrenia research assessed the reliability in 43 patients with schizophrenia and 40 healthy controls in four different versions of the CPT-IP at three test sessions over five weeks. In this study, all four versions of the CPT-IP demonstrated good short-term reliabilities and discrimination

sensitivity [16]. Results of that study also indicated that the 4-digit CPT and the shapes CPT were the most challenging versions for schizophrenia patients [16].

Overall, while in healthy individuals relevant CPT-IP measures appeared to be psychometrically reliable for short and long-term intervals and while well designed studies demonstrated good short-term reliability in schizophrenia, studies assessing long-term reliability of CPT-IP measures in patients with schizophrenia are to date still lacking. In an attempt to fill this methodological gap, the current study aimed at investigating test-retest reliability for the most challenging versions (4-digit and shapes) of the CPT-IP in schizophrenia for a between test sessions interval of 12 month. Such long term reliability data of this standard measure of sustained attention can be useful when interpreting possible results of longitudinal effects of interventional studies aimed at improving attentional deficits in patients with schizophrenia.

## Methods

### Participants

All participants agreed to repeat the both CPT-IP tests approximately 12 months after the first session and gave written informed consent before participating in the study. The study protocol was approved by the ethics committee of the University Hospital Campus Benjamin Franklin of the Charité University Medicine Berlin, and the study was conducted in accordance with the Declaration of Helsinki. All participants were right-handed [17], reported normal or corrected-to-normal vision, and were of European descent.

### Patients with Schizophrenia

Fifty patients (20 female) meeting DSM-IV criteria for schizophrenia were initially enrolled at baseline. Patients met the following inclusion criteria: age 18–65; no major change in antipsychotic medications defined as switching to another antipsychotic class and no psychiatric hospitalization for at least six weeks prior to entering the study. Patients were excluded from the study when meeting any of the following criteria: psychiatric inpatient status between test sessions, history of severe medical or neurological disorder, substance abuse/dependence other than nicotine as evidenced by urinary drug screening, history of electroconvulsive therapy, current extra-pyramidal symptoms, current treatment with mood stabilizers; antidepressants or anticholinergic agents and benzodiazepine co-medication within six weeks prior to testing. Forty patients (15 female) remained clinically stable within one year, had no major change of antipsychotic medication and accepted reassessment (80% retention rate). Between both test sessions, all patients had monthly psychiatric appointments to ensure continuous clinical monitoring during the study. All patients were recruited from the outpatient unit of the Department of Psychiatry and Psychotherapy, Campus Benjamin Franklin, Charité-University Medicine Berlin in Germany.

### Psychopathology and Medications

During the initial screening and final assessment, patients were administered the Positive and Negative Symptom Scale (PANSS) as a measure of psychopathology severity [18]. We tried to assess test-retest reliability of CPT-IP measures independent of changes in psychopathology between test sessions and therefore included only clinically stable patients who had a severity rating of moderate or less ( $\leq 4$ ) on selected PANSS positive scale items (conceptual disorganization, hallucinations, hyperactivity and hostility) that most likely could interfere understanding or valid

execution of the CPT-IP testing [19]. PANSS ratings were performed by author EH at both days of CPT-IP testing. Symptom levels were mild, as reflected by a PANSS total score of 62.9 at test and 58.4 at retest sessions [19]. None of the patients had a PANSS score  $>4$  (moderate) for any single item on the positive subscale. Mean PANSS scores between test sessions did not differ significantly, as measured by the positive ( $13.33 \pm 4.6$  vs.  $13.18 \pm 3.5$ ;  $T_{39} = 0.179$ ;  $p = .86$ ), negative ( $16.70 \pm 4.5$  and  $15.35 \pm 4.9$ ;  $T_{39} = 1.546$ ;  $p = .13$ ) and general subscales ( $31.90 \pm 7.1$  and  $29.88 \pm 7.5$ ;  $T_{39} = 1.381$ ;  $p = .18$ ). All patients received oral second generation antipsychotic medication: amisulpride ( $N = 6$ ), aripiprazole ( $N = 9$ ), clozapine ( $N = 11$ ), olanzapine ( $N = 7$ ), paliperidone ( $N = 2$ ), quetiapine ( $N = 5$ ), risperidone ( $N = 14$ ), ziprasidone ( $N = 1$ ). Twenty-five patients received a second generation antipsychotic mono-therapy, and 15 patients received a second generation antipsychotic combination therapy. Calculation of chlorpromazine (CPZ) equivalents was performed following the suggestion of Andreasen et al. [20]. CPZ equivalents of oral second generation antipsychotic medication between baseline and retest session did not differ significantly ( $557.45 \pm 347.11$  vs.  $542.43 \pm 299.15$ ;  $T_{39} = 0.527$ ;  $p = .60$ ). Additionally, CPZ equivalents were highly correlated ( $r = .855$ ;  $p < .001$ ), and none of the patients was switched to a different antipsychotic medication class between test sessions, thus further ensuring equivalent psychopharmacological conditions within the patient group over time.

Demographic and clinical data are summarized in **Table 1**.

### Healthy Controls

Fifty healthy control participants (20 female) were recruited via advertisements in a local newspaper and on the department's homepage. Controls were initially matched for age, sex, smoking status and education years. Forty-seven control participants (19 female) were reassessed at follow-up (94% retention rate) and were entered into analysis. Three control participants had moved to another city and thus declined reassessment. Control participants were excluded if they met any of the following criteria: history of psychiatric axis I disorder according to DSM-IV, history of sustained abuse of alcohol or drugs other than tobacco smoking, severe medical or neurological condition and any psychopharmacological treatment in the past. A first-degree family history of psychiatric axis I disorder likewise led to exclusion from the study. All control participants were examined by a certified psychiatrist prior to inclusion using the Mini-International Neuropsychiatric Interview for DSM-IV [21].

### Continuous Performance Test–Identical Pairs Version (CPT-IP)

Sustained attention was measured using the two most challenging versions of the CPT-IP developed by Cornblatt et al. [6]. Both versions included 300 trials for the 4-digit and 300 trials for the shapes condition. All stimuli were presented in a distance of 60 cm on a 19-inch TFT monitor located in a sound- and light-attenuated testing room. Each stimulus was presented for 50 ms, followed by an inter-stimulus interval of 950 ms, resulting in a total trial time of 1000 ms. Subjects were asked to respond as fast and as accurately as possible (via a finger lift from a response key of their dominant right hand) as soon as two consecutive stimuli looked identical [3]. Following 50 practice trials with three-digit numbers and shapes,  $2 \times 300$  test trials were administered, divided into two successive blocks for both conditions in a counterbalanced order. In each condition, 20% target pairs and an equal percentage of “false alarm” pairs – i.e. catch trials – were presented. Another 60% of randomly presented stimuli served as

**Table 1.** Summary of demographic and clinical data at baseline and retest (mean  $\pm$  SD).

	Baseline Schizophrenia	Baseline Controls	Retest Schizophrenia	<i>p</i>
N (female/male)	40 (15/25)	47 (19/28)	–	n.s. <sup>a</sup>
Age [years]	34.4 $\pm$ 10.2	33.9 $\pm$ 9.5	–	n.s. <sup>b</sup>
Education [years]	13.7 $\pm$ 1.8	14.2 $\pm$ 1.8	–	n.s. <sup>b</sup>
Smokers/non-smokers	25/15	28/19	–	n.s. <sup>a</sup>
Nicotine [pack years]	7.48 $\pm$ 9.0	6.51 $\pm$ 8.1	–	n.s. <sup>b</sup>
Test interval [days]	377.7 $\pm$ 17.7	372.5 $\pm$ 7.6	–	n.s. <sup>b</sup>
DOI [months]	89.74 $\pm$ 68.6	–	–	–
N episodes	3.07 $\pm$ 1.9	–	–	–
PANSS positive scale	13.33 $\pm$ 4.6	–	13.18 $\pm$ 3.5	n.s. <sup>c</sup>
PANSS negative scale	16.70 $\pm$ 4.5	–	15.35 $\pm$ 4.9	n.s. <sup>c</sup>
PANSS general scale	31.90 $\pm$ 7.1	–	29.88 $\pm$ 7.5	n.s. <sup>c</sup>
CPZ equivalents [mg/d]	557.45 $\pm$ 347.1	–	542.43 $\pm$ 299.2	n.s. <sup>c</sup>

DOI, duration of illness; PANSS, Positive and Negative Syndrome Scale; CPZ, chlorpromazine.

<sup>a</sup> $\chi^2$  test;

<sup>b</sup>t-test for independent samples,

<sup>c</sup>t-test for paired samples.

doi:10.1371/journal.pone.0084780.t001

organized fillers. Participants did not have any information regarding the proportions of trial stimuli.

Outcome measures were calculated for the digits and shapes conditions separately. The primary outcome measure, *d'* assesses the ability of the participant to discriminate between signal and noise. Secondary outcome measures were hit rate (percent correct hits), mean reaction time (RT) for correct hits and calculated measures for the response bias ( $\ln \beta$ ) and (log random) as measure of responses to irrelevant stimuli.

### Statistical Analysis

Statistical calculations were conducted using SPSS for Windows 19.0 (IBM, Armonk, NY, US). All tests were performed as two-tailed tests with an alpha level set at  $p < .05$ . Demographic and clinical data were analyzed with  $\chi^2$  tests, t-tests for independent and for paired samples, as appropriate. Primary outcome measures, i.e. CPT-IP variables *d'*, hit-rate, mean reaction time,  $\ln \beta$ , and log random for 4-digit and shape stimuli, were submitted to a repeated measures analysis of variance (ANOVA). Time was entered as a within-subject factor and diagnostic group was entered a between-subject factor, thus resulting in a 2 $\times$ 2 ANOVA design, which was applied to every primary outcome measure separately. As this approach results in ten separate repeated measures ANOVAS, a Bonferroni correction was applied. To compare our results with those of previous studies, longitudinal data, i.e. test-retest reliability proper, were psychometrically analyzed computing Pearson's *r* correlation coefficients (Cook and Beckman, 2006). For both reliability measures, we initially defined the thresholds for reliability in our study protocol following Altman who considered  $r < 0.4$  as poor,  $r = 0.4$ – $0.6$  as moderate,  $r = 0.6$ – $0.8$  as good, and  $r > 0.8$ – $1.0$  as very good reliability [22].

### Results

The mean interval between test and retest was 12.3 months (range 11.8–13.7 months) for both groups. Mean intervals between test sessions did not differ significantly between healthy controls

(372.5 $\pm$ 7.6 days) and patients with schizophrenia (377.7 $\pm$ 17.2 days).

### Omnibus ANOVAs

Repeated measures ANOVAs demonstrated significant main effects of 'diagnostic group' for the outcome variables *d'* for both digits ( $F(1,85) = 23.87$ ;  $p < .001$ ) and shapes ( $F(1,85) = 32.18$ ;  $p < .001$ ); hit rate for both digits ( $F(1,85) = 28.65$ ;  $p < .001$ ) and shapes ( $F(1,85) = 31.71$ ;  $p < .001$ ); and log random for digits only ( $F(1,85) = 20.41$ ;  $p < .001$ ). No significant main effects of 'time' or significant interactions of 'time\*diagnostic group' for any outcome variable were found. Complete raw data is available on request by the corresponding author.

### Longitudinal Data

Test-retest reliability data are presented stratified by group and condition in **Table 2**. Pearson's *r* correlations indicated that *d'* (4-digit:  $r = .812$ ;  $p < .001$ ; shapes:  $r = .729$ ;  $p < .001$ ); hit rate (4-digit:  $r = .707$ ;  $p < .001$ ; shapes:  $r = .603$ ;  $p < .001$ ); and mean RT (4-digit:  $r = .739$ ;  $p < .001$ ; shapes:  $r = .732$ ;  $p < .001$ ) showed good long-term retest reliability in healthy controls. In patients with schizophrenia, *d'* (4-digit:  $r = .502$ ;  $p < .001$ ; shapes:  $r = .529$ ;  $p < .001$ ) and hit rate (4-digit:  $r = .459$ ;  $p < .001$ ; shapes:  $r = .410$ ;  $p < .01$ ) were only moderately correlated.

### Discussion

The primary aim of this study was to assess for the first time one year test-retest reliability of the most challenging 4-digit and shapes versions of the CPT-IP in outpatients with schizophrenia. We found that *d'* and - to a lesser degree - hit rate for correct responses revealed, only moderate long-term test-retest reliability in schizophrenia, while the same variables showed high test-retest reliability correlations in healthy controls. These high levels of long-term test retest scores in healthy controls are well in line with earlier studies. Cornblatt and colleagues (1988) evaluated test-retest reliability in 120 healthy participants over a long-term interval of 1.5 years and reported moderate to good Pearson's

**Table 2.** Test-retest correlations of CPT-IP performance measures (mean  $\pm$  SD).

Condition	CPT-IP measures	Baseline Controls	Retest Controls	Pearson r
<b>CPT-4-digit</b>	d'	1.73 $\pm$ 0.9	1.78 $\pm$ 0.8	0.812***
	hit rate (%)	75.60 $\pm$ 18.0	76.98 $\pm$ 18.5	0.707***
	RT hits [ms]	532.25 $\pm$ 61.5	537.06 $\pm$ 65.9	0.739***
	ln $\beta$	-.10 $\pm$ 0.8	-.09 $\pm$ 0.7	0.370**
	log random	0.34 $\pm$ 0.5	0.16 $\pm$ 0.3	0.286
<b>CPT-shapes</b>	d'	1.96 $\pm$ 0.7	2.28 $\pm$ 0.6	0.729***
	hit rate (%)	76.23 $\pm$ 16.3	81.15 $\pm$ 14.7	0.603***
	RT hits [ms]	512.49 $\pm$ 65.0	524.15 $\pm$ 68.5	0.732***
	ln $\beta$	0.09 $\pm$ 0.8	0.17 $\pm$ 1.1	0.636***
	log random	0.53 $\pm$ 0.6	0.64 $\pm$ 0.5	0.194
Condition	CPT-IP measures	Baseline Schizophrenia	Retest Schizophrenia	Pearson r
<b>CPT-4-digit</b>	d'	1.00 $\pm$ 0.7	1.10 $\pm$ 0.6	0.502***
	hit rate (%)	52.75 $\pm$ 23.3	57.45 $\pm$ 23.7	0.459***
	RT hits [ms]	550.41 $\pm$ 80.4	541.89 $\pm$ 116.6	0.281
	ln $\beta$	0.30 $\pm$ 0.7	0.23 $\pm$ 0.6	0.377*
	log random	0.80 $\pm$ 0.7	0.57 $\pm$ 0.6	0.253
<b>CPT-shapes</b>	d'	1.29 $\pm$ 0.9	1.31 $\pm$ 0.9	0.529***
	hit rate (%)	67.35 $\pm$ 24.5	67.12 $\pm$ 26.4	0.410**
	RT hits [ms]	525.53 $\pm$ 78.1	528.55 $\pm$ 118.8	0.317*
	ln $\beta$	0.44 $\pm$ 0.6	0.41 $\pm$ 0.7	0.285
	log random	0.70 $\pm$ 0.9	0.82 $\pm$ 0.7	0.380*

\* $p < .05$ ;\*\* $p < .01$ ;\*\*\* $p < .001$ .

doi:10.1371/journal.pone.0084780.t002

correlations for  $d'$  in digits and shapes condition [6]. Chen and Faraone (2000) reported overall good test-retest reliabilities of both  $d'$  and hit rate in healthy participants. Both studies agreed that other CPT-IP measures generally showed less satisfactory reliabilities [13]. Three studies investigated the important methodological issue of CPT-IP test-retest reliability in patients with schizophrenia but reported reliability estimates only for short-term tests intervals. Nuechterlein and co-workers (2008) assessed 167 clinically stable patients with schizophrenia (95% retention rate) and reported very good test-retest reliability for the mean  $d'$  value across 2-, 3-, and 4-digit conditions after 4 weeks [15]. Another large clinical trial investigated 323 clinically-stable outpatients with schizophrenia at 29 sites and confirmed good short-term test-retest reliability for the mean  $d'$  values across 2-, 3-, and 4-digit conditions of the CPT-IP [14]. The most recent publication assessed short-term reliability on the 2-digit, 3-digit, 4-digit and shapes condition of the CPT-IP at three test time points over a total of five weeks in a similarly large sample of 43 patients with schizophrenia and 40 healthy controls. The authors demonstrated again good short-term test-retest reliability in patients with schizophrenia for  $d'$  and hit rate for all 4 versions of the CPT-IP [16]. Overall and in contrast to our results, small differences in repeated short term test-retest reliability across three trials both for  $d'$  and hit rate were not considered meaningfully different between groups [16].

Our study replicates previous studies which revealed high short and long-term test-retest reliability for healthy controls for the most challenging CPT-IP conditions for both  $d'$  and hit rate. We found also consistent between-group differences for  $d'$  and hit rate

in our cross-sectional sample of healthy controls and schizophrenia patients, and thus our results replicate findings on excellent sensitivity of the CPT-IP. These results are also in line with previous findings of sustained attention deficits in individuals at risk for later development of schizophrenia-spectrum disorders and support the assertion that sustained attention impairment could be an indicator of a schizophrenia diathesis [3,13,23,24].

Inclusion of cognitive measures in DSM-5 as part of diagnostic criteria for schizophrenia has been carefully considered, since cognitive impairments are relatively independent of symptom severity and stage of illness, but their discriminative value to other "boundary" disorders as bipolar and schizoaffective disorder was not considered sufficient for their inclusion in DSM-5 [25,26]. Moreover, stability of cognitive processes varies according to assessed domains and test-retest intervals. In schizophrenia overall sufficient short term test-retest reliability was found for most cognitive measures assessing the domains: speed of processing, attention and vigilance, verbal learning, reasoning and problem solving. For other domains as working memory, visual learning and social cognition test-retest reliability varied with selected cognitive tests. For data of short term test-retest reliability for 36 included candidate cognitive tests, see Nuechterlein et al. [15].

Pietrzak et al. assessed stability of cognitive impairment in chronic schizophrenia over brief (i.e., hours) and intermediate retest intervals (i.e., one month) using computerized tests of the domains: psychomotor function (Detection Task), visual attention/information processing (Identification task), non-verbal learning (Visual Learning Task), and executive function (Groton Maze

Learning Test) and found slightly lower but good test-retest reliabilities in schizophrenia compared to healthy controls. For intermediate re-tests intervals the authors suggested that variability in cognitive performance may reflect more inherent characteristic of schizophrenia, rather than differences in test-retest reliability of cognitive measures [27].

Data on long-term neurocognitive stability in the course of illness is comparably scarce and needed in different schizophrenia patient samples, since the duration of most longitudinal studies was short to modest. A longitudinal investigation of cognitive function in schizophrenia over 1 year reported a decline in spatial recognition but not in pattern recognition or motor speed, using the Cambridge Neuropsychological Test Automated Battery (CANTAB) as part of the CULASS trials. In that study cognitive changes were present in schizophrenia patients but the magnitude of change was small compared with differences in cognitive measures that existed between patients [28]. A recent study assessed stability of cognition and its relation to functional outcome over a 1-year test-retest interval in 128 schizophrenia outpatients. Cognitive functioning was stable in most patients and deterioration was mainly observed for letter-number sequencing and semantic fluency tests. Estimates for 1 year test-retest reliabilities (Intra-Class Correlations, ICC), that were in our sample virtually identical to Pearson's  $r$ , varied between .50 for semantic fluency,  $>0.6$  for recall and intrusions of the California Verbal Learning Test-II,  $>.7$  for phonematic fluency, letter number sequencing and symbol search and up to ICC  $>.9$  for vocabulary subtests of the Wechsler Adult Intelligence Scale-III. ICCs for functional outcome status were substantially lower within a 1 year retest interval [29].

Another recently published study assessed the course of cognitive deficits in 78 first episode schizophrenia spectrum disorder patients over a 1-year and 3-year follow-up period. Six cognitive domains assessed in this study consisted of: verbal memory (Rey Auditory Verbal Learning Test), visual memory (Rey Complex Figure Test), motor dexterity (Grooved Pegboard Test), executive functions/speed of processing (Trail Making Tests A and B), WAIS III-Backward Digits and Digit Symbol), attention and impulsivity (Continuous Performance Test- degraded-stimulus). Interestingly while first episode patients and their controls increased their performance in all cognitive measures except for verbal and visual memory, the authors identified a subgroup of 34 patients that showed a cognitive decline, associated with negative symptoms and poor functional outcome [30]. Finally, assessment of neurocognitive performance and stability in a multiplex multigenerational study of schizophrenia using a computerized neurocognitive battery calculated test-retest reliabilities (ICCs) for accuracy and speed over a 5-year interval. Compared to unaffected family members, test-retest correlations were lower in schizophrenia patients for the domains abstraction and mental flexibility; verbal memory; face memory; spatial memory; language reasoning and emotion processing. Moreover schizophrenia patients were more impaired in relation to speed than in accuracy and showed higher across-task intra-individual variability in performance compared with unaffected family members [31].

Following the aim of the current study, the incremental value of our results is constituted by adding for the first time long-term reliability data in schizophrenia for the 4-digit and shapes CPT-IP conditions. In contrast to studies using healthy controls and in contrast to studies revealing high short-term reliability in different versions of the CPT-IP in schizophrenia, we could replicate these findings only in part for a substantially longer test-retest interval and found only moderate test-retest reliability in an independent sample of outpatients with schizophrenia with mild symptom

levels. Only moderate long-term reproducibility ( $r=.5$ ) of  $d'$  among patients with schizophrenia in our sample would imply that changes in CPT-IP performance may not be mainly attributed to hypothetical effects of clinical interventions, but also to random effects or noise. For clinical and practical significance and usefulness other authors proposed levels of observed agreement  $>.7$  as a minimal necessary threshold [32,33]. As a rule of thumb for clinical studies and in contrast to the thresholds of reliability proposed by Altman [22], Cicchetti considers reproducibility  $r<.7$  as poor,  $r=.7-.79$  as moderate,  $r=.8-.89$  as good, and  $r>.9$  as excellent [33,34].

Our study has several strengths but also limitations: All patients included in this study were known to the authors before entering the study and had approximately monthly appointments during and after the study. Although most patients exhibited residual psychopathological symptoms, the severity was mild to moderate and psychopathology did not differ significantly between test sessions. Additionally, all patients were on stable antipsychotic medication between test session, and changes in dosages were minor and not significant. While clinical stability and insignificant changes in medication ensured that our sample is well comparable to short term studies on reliability, we believe that our study design may have influenced results on long-term stability, since an interventional study could include more clinically varying patients and thus agreement between test-sessions may even be over-estimated. On the other hand when a test is given to a very similar (homogeneous) group, the resulting scores may be too closely clustered and the reliability coefficient might therefore be actually lower than in a more heterogeneous examinee group.

In our methodological approach, only moderate test-retest stability of individual differences on CPT-IP in schizophrenia where found in spite of consistent cross-sectional differences at both sessions and in spite of clinical and medication stability during a one year interval. Additionally although our sample-size is comparable to a recent study on short term reliability of CPT-IP in schizophrenia [16], sample size is another important factor for studies of test-retest reliability. Lower test-retest reliabilities of measures would imply that larger samples are needed to detect interventional effects in clinical studies. Finally we did not control for individual factors that may influence test-retest reliability as intra-individual variability, poor motivation, fatigue, insufficient sleep, food intake or cigarette smoking prior to CPT-IP test-sessions.

In conclusion, our results imply - in contrast to most studies assessing only short-term stability of  $d'$  in schizophrenia - at best moderate and for usage in clinical studies possibly not sufficient long-term temporal stability of  $d'$  and hit rate for the 4-digits and the less used shapes condition of the CPT-IP. It is noteworthy that our results is limited to the 4-digit and shapes version of the CPT-IP and has been assessed in one sample of patients with schizophrenia at one site only. Further research is needed to answer the question whether less challenging versions of the CPT-IP, in different samples of patients with schizophrenia or related psychiatric disorders or a variation of intervals between test-sessions may yield better, and clinically acceptable long-term reliability of CPT-IP. In that context it is interesting that the mean across the 2, 3, and 4 digit versions of the CPT-IP, due to its high test-retest reliability in short term intervals, was chosen for the MATRICS test battery as a measure for sustained attention [14,15].

Given our replication of consistent and stable cross-sectional differences within a one-year interval, we consider  $d'$  - at least for short term interventional studies - also in the most challenging versions of the CPT-IP still a valuable measure for sustained

attention in schizophrenia, while hit rate in both conditions could be considered a potentially valuable CPT-IP measure for the 4-digit and shapes condition. However our results of at best moderate test-retest reliability of  $d'$  and hit-rate for a 1 year test-retest interval in our sample of clinically stable outpatients with schizophrenia may point towards less practical use of these measures when evaluating long term clinical trials to improve sustained attention in schizophrenia.

## References

- Green MF, Kern RS, Heaton RK (2004) Longitudinal studies of cognition and functional outcome in schizophrenia: implications for MATRICS. *Schizophr Res* 72: 41–51.
- Shamsi S, Lau A, Lencz T, Burdick KE, DeRosse P et al. (2011) Cognitive and symptomatic predictors of functional disability in schizophrenia. *Schizophr Res* 126 (1–3): 257–264.
- Cornblatt BA, Malhotra AK (2001) Impaired attention as an endophenotype for molecular genetic studies of schizophrenia. *Am J Med Genet* 105(1): 11–5.
- Cornblatt BA, Lenzenweger MF, Erlenmeyer-Kimling L (1989) The continuous performance test, identical pairs version: II. Contrasting attentional profiles in schizophrenic and depressed patients. *Psychiatry Res* 29(1): 65–85.
- Kumar CT, Christodoulou T, Vyas NS, Kyriakopoulos M, Corrigan R et al (2010) Deficits in visual sustained attention differentiate genetic liability and disease expression for schizophrenia from Bipolar Disorder. *Schizophr Res* 124(1–3): 152–60.
- Cornblatt BA, Risch NJ, Faris G, Friedman D, Erlenmeyer-Kimling L (1988) The Continuous Performance Test, identical pairs version (CPT-IP): I. New findings about sustained attention in normal families. *Psychiatry Res* 26(2): 223–38.
- Cornblatt BA, Keilp JG (1994) Impaired attention, genetics, and the pathophysiology of schizophrenia. *Schizophr Bull* 20(2): 248.
- Salgado-Pineda P, Baeza I, Pérez-Gómez M, Vendrell P, Junqué C et al. (2003) Sustained attention impairment correlates to gray matter decreases in first episode neuroleptic-naïve schizophrenic patients. *Neuroimage* 19: 365–75.
- Roitman SE, Cornblatt BA, Bergman A, Obuchowski M, Mitropoulou V et al. (1997) Attentional functioning in schizotypal personality disorder. *Am J Psychiatry* 154(5): 655–60.
- Kurtz MM, Ragland JD, Bilker W, Gur RC, Gur RE (2001) Comparison of the continuous performance test with and without working memory demands in healthy controls and patients with schizophrenia. *Schizophr Res* 48: 307–316.
- Gooding DC, Matts CW, Rollmann EA (2006) Sustained attention deficits in relation to psychometrically identified schizotypy: evaluating a potential endophenotypic marker. *Schizophr Res* 82: 27–37.
- Franke P, Maier W, Hardt J, Hain C, Cornblatt BA (1994) Attentional abilities and measures of schizotypy: their variation and covariation in schizophrenic patients, their siblings, and normal control subjects. *Psychiatry Res* 54(3): 259–72.
- Chen WJ, Faraone SV (2000) Sustained attention deficits as markers of genetic susceptibility to schizophrenia. *Am J Med Genet* 97(1): 52–57.
- Keefe RS, Fox KH, Harvey PD, Cucchiaro J, Siu C et al. (2011) Characteristics of the MATRICS Consensus Cognitive Battery in a 29-site antipsychotic schizophrenia clinical trial. *Schizophr Res* 125(2–3): 161–168.
- Nuechterlein KH, Green MF, Kern RS, Baade LE, Barch D, et al (2008) The MATRICS Consensus Cognitive Battery, part 1: test selection, reliability, and validity. *Am J Psychiatry* 165(2): 203–213.
- Kahn PV, Walker TM, Williams TS, Cornblatt BA, Mohs RC et al. (2012) Standardizing the use of the Continuous Performance Test in schizophrenia research: A validation study. *Schizophr Res* 142: 153–158.
- Oldfield RC (1971) The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia* 9: 97–113.
- Kay SR, Fiszbein A, Opler LA (1987) The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull* 13(2): 261–276.
- Leucht S, Kane JM, Kissling W, Hamann J, Etschel E et al. (2005) What does the PANSS mean? *Schizophr Res* 79: 231–238.
- Andreasen NC, Pressler M, Nopoulos P, Miller D, Ho BC (2010) Antipsychotic dose equivalents and dose-years: a standardized method for comparing exposure to different drugs. *Biol Psychiatry* 67: 255–262.
- Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J et al. (1998) The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* 59: 22–33.
- Altman DG (1991) Practical statistics for medical research. Chapman and Hall, Padstow, Cornwall.
- Bergida H, Lenzenweger MF (2006) Schizotypy and sustained attention: confirming evidence from an adult community sample. *J Abnorm Psychol* 115 (3): 545–51.
- Thaden E, Rhinewine JP, Lencz T, Kester H, Cervellione KL et al. (2006). Early-onset schizophrenia is associated with impaired adolescent development of attentional capacity using the identical pairs continuous performance test. *Schizophr Res* 81(2–3): 157–66.
- Barch D, Keefe RS (2010) Anticipating DSM-V: Opportunities and challenges for cognition and psychosis. *Schizophr Bull* 36: 43–47.
- Rajiv T, Gaebel W, Barch DM, Bustillo J, Gur RE et al. (2013) Definition and description of schizophrenia in the DSM-5. *Schizophr Res* 150(1): 3–10.
- Pietrzak RH, Snyder PJ, Jackson CE, Olver J, Norman T et al. (2009) Stability of cognitive impairment in chronic schizophrenia over brief and intermediate retest intervals. *Hum Psychopharmacol Clin Exp* 24: 113–121.
- Barnett JH, Croudace TJ, Jaycock S, Blackwell C, Hynes F et al. (2007) Improvement and decline of cognitive function in schizophrenia over one year: a longitudinal investigation using latent growth modeling *BMC Psychiatry* 7: 16.
- Miles AA, Heinrichs RW, Ammari N, Hartman L, McDermid Vaz S et al. (2013) Stability and change in symptoms, cognition, and community outcome in schizophrenia. *Schizophr Res* Sep 27. doi: 10.1016/j.schres.2013.09.006.
- Rodríguez-Sánchez JM, Ayesa-Arriola R, Pérez-Iglesias R, Periañez JA, Martínez-García O et al. (2013) Course of cognitive deficits in first episode of non-affective psychosis: A 3-year follow-up study. *Schizophr Res* 150(1): 121–8.
- Roalf DR, Gur RC, Almasy L, Richard J, Gallagher RS et al. (2013) Neurocognitive Performance Stability in a Multiplex Multigenerational Study of Schizophrenia. *Schizophr Bull* 39(5): 1008–17.
- Nunnally JC, Bernstein IH (1994) Psychometric theory (third edition). New York, McGraw Hill.
- Cicchetti DV (1999) Sample size requirements for increasing the precision of reliability estimates: Problems and proposed solutions. *J Clin Exp Neuropsychol* 21: 567–570.
- Cicchetti DV (2001) The Precision of Reliability and Validity Estimates Re-Visited: Distinguishing Between Clinical and Statistical Significance of Sample Size Requirements. *J Clin Exp Neuropsychol* 23: 695–700.

## Author Contributions

Conceived and designed the experiments: AHN MD. Performed the experiments: EH TMT AV. Analyzed the data: EH AV TMT. Contributed reagents/materials/analysis tools: EH AHN. Wrote the paper: EH AV LKK CH AHN.