Standardisation Study of Cogtest

Neurocognitive Battery for Clinical Trials

TONMOY SHARMA, Clinical Neuroscience Research Centre, Dartford, Kent, United Kingdom

ROBERT BILDER, Department of Psychiatry, University of California at Los Angeles

ABSTRACT

Paper and pencil tests are increasingly being replaced by computerized batteries for cognitive testing in clinical trials. We sought to standardize and cross-validate Cogtest, an automated neurocognitive battery with electronic data capturing ability. 200 cognitively healthy individuals took part in two studies, one aimed at assessing concurrent validity with respect to paper-and-pencil measures and another to establish norms for individual Cogtest endpoints that comprise a Neurocognitive Global Score (NGS) for use in clinical trials. The overall design of the first study (N=75) involved parallel, counterbalanced administration of both the Cogtest computerised test battery and a battery of paper-and-pencil (P&P) tests (selected to approximate the testing done in current clinical trials). There were two test sessions, and both sets of instruments were presented in counterbalanced order at both sessions, with testing at 0 and 4 weeks (+/- 3 days) controlling for time of day. Four alternate forms of the Cogtest battery were counterbalanced so that each subject experienced two different forms. The 75 subjects were stratified by age over 4 decades from age 20 to 60. Age 20 subjects, equally distributed across age group and sex. The second study (N=120) involved two "baseline" sessions and one follow-up session, similar counterbalancing of alternate forms, and 20 participants (10 men and 10 women) in each of 6 age bands from age 13 to 69. The first study revealed sex differences with men generally faster and more accurate on spatial processing but women more accurate in face recognition memory, paralleling prior research. Correlations of Cogtest measures with analogous P&P tests revealed correlations in the range of r=.3 to r=.7. The second study revealed a "classic" curvilinear age effect on NGS and individual domains with lower scores in the 13 to 19 year age group, highest scores in the 20-29 year group, and then monotonically decreasing scores through the 6th decade. Both studies revealed that test-retest stability over up to 4 weeks was in keeping with published results for the paper-pencil tests, with typical test-retest reliability coefficients in the range of ~0.4 to ~0.9. This study shows that the Cogtest computerized testing battery has similar psychometric properties to paper-and-pencil tests that have been used in clinical trials and offer the added advantage of computerization, millisecond accuracy, electronic data capture and audit trail.

INTRODUCTION

Cognitive function is the process that underpins everyday behavior. It enables processing of information and organizes our thoughts and actions. It allows storage and retrieval of information and enables individuals to execute tasks. Cognitive performance can be described and measured in different domains including memory (working, declarative), attention, motor function, general executive functioning, and perpetual organization. Deficits in cognition are part of the etiology of several neuropsychological disorders, eg., schizophrenia, attention deficit disorder, Alzheimer's disease, etc. These deficits are often clinically meaningful in that they are related to the degree of difficulty that some individuals have with activities of daily living, as well as the ability to acquire skills in psychosocial rehabilitation.

OBJECTIVES

Study I: To standardize the Cogtest battery (Cogtest Inc, London; www.cogtest.com) against paper-and-pencil equivalent tests.

Study II: To establish means and standard deviations for the Cogtest endpoints on cognitively healthy individuals that will feed into the Neurocognitive Composite Score (NCS) to enable: a) Computation of a NCS from subject screening data which will serve as a baseline reference in future studies. b) Characterization of severity of cognitive deficit at baseline of the study sample; c) To serve as a reference for inter-study comparisons of NCS.; d) Interpretation of NCS=0.0 as "normal" on all tests.

METHODS

All participants were administered by the Cogtest battery of following 14 tests; to test-retest the reliability of the Cogtest battery:
Set Shifting Test, CPT – Flanker, CPT – AX, Tapping Speed Test – left/right, Face Matching immediate and delayed, Word List Memory immediate and delayed, Object Working Memory diffculty and delayed Spatial Working Memory 2s and 12 s. The average time to complete the battery was approximately 1 hr.

The overall design involved parallel, counterbalanced administration of both the Cogtest battery and a battery of paper-and-pencil (P&P) tests (selected to approximate the testing done in current clinical trials). There were two test sessions, with both sets of instruments presented at both sessions, with time between testing of 4 to 8 weeks (paralleling the duration of typical clinical trials). There were 4 alternate forms of each test in the Cogtest battery. Subjects were assigned one of the four forms at each session, so each subject will experience two different forms, and each of the 12 possible administration orders will be experienced by 20 subjects, equally distributed across age group and sex.

Paper and Pencil Tests for the Cogtest Standardization Study

Test	Description	Time (min.)
AVLT	Verbal list learning test	15 + 5 = 20
Digit Span (WAIS-III) Memory	Digit recall and working	5
Letter-Number Span (WAIS-III)	Working memory	5
Coding (WAIS-III) Speed	Attention and processing	4
WCST	Executive functions	20
TMT	Executive functions	8
Verbal Fluency (COWAT, Animal Naming)	Executive functions	6
Finger Tapping Test	Motor speed	5
Grooved Pegboard Test	Motor dexterity	5
Vocabulary (WAIS-III)	Word knowledge	10
MMSE – Mini-Mental State Exam	General ability	12
ADAS-Cog	General ability	20 (two highest age strataf only)

STUDY DESIGN

This methodology study followed a single center, parallel group design, and was conducted in male and female adult subjects. One hundred and twenty (6 groups of 20) subjects were required. Groups were defined as follows:

Group	Age Range (yrs)	Males	Females
1	13-19	10	10
2	20-29	10	10
3	30-39	10	10
4	40-49	10	10
5	50-59	10	10
6	60-69	10	10

RESULTS

Endpoints and Criteria for Evaluation: Collection of results for the following cognition tests in healthy normals:

- Cogtest Workstation Orientation
- Continuous Performance Test – AX Version
- Continuous Performance Test – Identical Pairs Version
- Continuous Performance Test Flanker Version
- Strategic Target Detection Test – 4-shape version
- Object Working Memory Test
- Spatial Working Memory Test
- Facial Memory Test (immediate and delayed)
- Word List Memory Test (immediate and delayed)
- Symbol Digit Substitution
- Competing Programs Test
- Auditory Digit Span
- Set Shifting Test
- Simple Go-No Go test
- Auditory Letter-Number Sequencing
- Emotion recognition test (PEAT)

Correlations of computerized test results with age were nonsignificant, except for WO reaction time, which was positively correlated with age, suggesting that slower reaction time was associated with increasing age (r=.27, p=.027, n = 68). It should be recognized that this was despite a trend for the older subjects to have higher scores on measures usually associated with higher levels of premorbid ability (Vocabulary: r=.47, p=.001, n=71; phonemic fluency: r=.44; p<.001; n=70; WRAT Reading: r=.29, p=.015, n=71; Mini-Mental State Exam: r=.26, p=.030, n=70). Partial correlations of age with computerized test scores, controlling for Vocabulary, revealed that older subjects were more accurate in target selection (WO distance: r=-.28, p=.025, df=68) but slower to detect these (WO reaction time: r=.32, p=.01, df=68), suggesting a more conservative response strategy. Older subjects also tended to be more accurate in the Flanker CPT (r=.28, p=.023, df=66).

Sex differences are often found for certain verbal measures (on which women perform better) and certain spatial and motor performance measures (on which men perform better and/or faster). Compatible findings were observed for some of the paper-and-pencil measures, in that men were faster on the Finger Tapping Test (both dominant and non-dominant hand), and women performed better on all indices of the Buschke Selective Reminding Test (see Table 2). Significant sex effects were also found on the computerized tests. Specifically, men were more accurate on the Flanker CPT, had faster reaction times on the Flanker CPT and Set Shifting Test, were faster on the Tapping Speed Test, and more accurate (i.e., were closer to the target after a delay) on the Spatial Working Memory Test. In contrast, women were more accurate on the Face Memory Test (delayed recall). Women also tended to be faster on the computerized Word List Memory test, but only a subset of the total sample took this test (female n = 28, male n = 12) so although the scores and score differences between the groups were similar, the comparison to the BSRT is not exact.

There was no significant age difference between the sexes (mean ± SD age for women and men was 36.35 ± 12.67 and 32.75 ± 11.86, respectively; t=1.17, df=73, p=.24). Their estimated IQ were females 102, (SD + 9.02) and males 104 (+ 9.98). None of the tests were correlated with IQ. Chronbach's alpha coefficient (also called an `intraclass correlation') was computed to assess internal consistency across tests and trials (Chronbach, 1970). This statistic provides information on how well scores on each trial correlates with the total score. Chronbach's alpha was 0.98 and 0.90 across all tests for test-retest. Reliabilities between two tests spaced 4 weeks apart were 0.85 and higher for the Cognitive test battery data.

SUMMARY OF RESULTS

There were no significant gender effects except for the total errors endpoint in the competing program test. In general, this test was found to be problematic as there were a number of individuals that failed to complete the practice and never progresses to the actual test. Significant practice effects were observed for many of the individual test endpoints and for all of the composites except for the executive function domain composite. The declarative memory score does not appear to follow the pattern of learning as the other domains. Although there is substantial learning from the first practice, there appears to continue to be subsequent learning at the second practice as well. Significant age differences were observed for many of the individual test endpoints and for all of the composites except for the declarative and working memory domain composites. Figure 2. shows the composite scores across age groups for the test session responses standardized to the overall means. A more negative score reflects a worse cognition response.

DISCUSSION

Study I: The test-retest stability data shows the range of correlations (r) for the tes-retest which ranged from 0.38 (Face matching) to 0.84 (word list memory), with an Overall Test Battery Mean (OTBM) test retest of 0.57.

Figure 2. Age Effect on Composite Scores

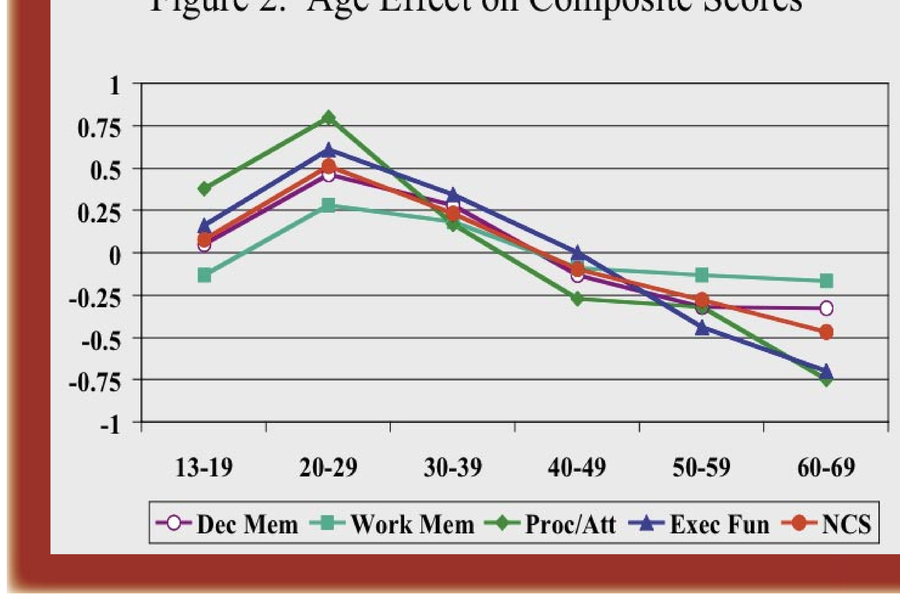

Table 1. Descriptive Statistics for Computerized Tests at Baseline

	N	Minimum	Maximum	Mean	SD
Workstation Orient, Dist (WODIST1)	68	6.50	22.50	14.01	3.37
Workstation Orientation, RT (WORT1)	68	481.00	1286.00	760.37	186.68
AX-CPT, Discriminability (AXPR1)	74	.28	1.00	0.86	0.13
Demographic Based IQ-Test (DC_FSIQ1)	68	84.01	120.39	102.88	9.32
Flanker CPT Accuracy (FL_ACC1)	71	25	48.00	41.72	4.45
Flanker CPT, Overall RT (session 1)	73	411.00	693.67	535.12	63.93
Flanker CPT, Cost (FLCOST1)	73	-172.00	58.00	-40.38	46.96
Face Matching Test, Initial (FMTI1)	74	60.00	95.00	79.01	7.52
Face Matching Test, Delay (FMTD1)	72	55.00	95.00	76.92	9.39
Obj Working Mem, Diff (OWMDIFF1)	72	3.00	15.80	7.37	3.12
Obj Working Mem, Delay (OWMDELA1)	72	8.00	20.00	15.40	2.71
Set Shifting, Basal RT (SSTBASR1)	74	233.00	581.00	334.72	62.74
Set Shifting, Imit Errs (SSTIERR1)	74	.00	37.00	4.08	5.78
Set Shifting, Rev Errs (SSTRERR1)	74	.00	23.00	3.76	4.35
SST, RT benefit both conds (session 1)	74	-31.50	200.50	46.45	43.22
SST, Shift Cost (session 1)	74	-180.00	149.00	30.45	52.81
Spatial Work Mem, 2 s (SWMD21)	74	15.00	48.00	28.97	7.42
Spatial Work Mem, 12 s SWMD121)	74	24.00	75.00	47.43	10.98
Tapping Speed, Right (TSTMR1)	73	136.23	238.59	175.78	21.56
Tapping Speed, Left (TSTML1)	73	142.75	282.50	193.86	28.68
Word List Memory Total (WLMTOT1)	40	28.00	67.00	50.58	10.35
Word List Mem, Trial-Trial (WLMTT1)	40	.23	.96	0.68	0.20

Table 2. Descriptive Statistics for Paper and Pencil Tests at Baseline

	N	Minimum	Maximum	Mean	SD
Mini-Mental State Exam (MMSE1)	70	20.00	30.00	29.13	1.53
Wide Range Ach Test– Reading (WRAT1)71	32	57.00	48.76	5,55	
WAIS-III Vocabulary (VOCB1)	71	11.00	72.00	41.79	12.75
WAIS-III Letter-Number Span (LNS1)	70	2.00	20.00	10.94	3.81
WAIS-III Digits Forward (DF1)	71	4.00	16.00	10.32	2.42
WAIS-III Digits Backward (DB1)	71	2.00	12.00	6.96	2.16
WAIS-III Digits Total (DIGTOT1)	71	10.00	25.00	17.28	3.98
WAIS-III Coding (CODG1)	71	46.00	114.00	76.54	15.60
Trail Making Part A time (TMTA1)	72	13.00	70.00	32.47	12.06
Trail Making Part B time (TMTB1)	72	27.00	153.00	63.50	22.26
Grooved Peg – Dom time (GPEGD1)	70	49.00	131.00	70.86	14.23
Grooved Peg – Nondom time GPEGND1	70	56.00	124.00	76.19	12.58
Finger Tapping – Dom (FTAPD1)	70	26.00	74.00	44.89	8.55
Finger Tapping – Nonom (FTAPND1)	70	18.00	61.00	39.63	7.48
Semantic Fluency (SFLU1)	70	10.00	28.00	19.89	4.05
Phonemic Fluency (P_FLU1)	70	17.00	64.00	39.84	12.20
Buschke SRT – Total (BSRTOT1)	70	27.00	77.00	55.13	11.09
Buschke SRT, CLTR (CLTR1)	75	00	74.00	32.13	19.52
Buschke SRT, Delayed Recall (BSRTD1)	69	1.00	16.00	10.51	3.93

There were exceptions of the object- and spatial delayed- working memory of showing 0.29. However, this suggest that this combination of tests, as a formal battery, has good reliability, both for individual tests and for the battery overall. Tabachnick and Fidell (1989) indicated that a correlation must be around 0.30 or higher to be interpretable. Our results suggest that the battery of cognitive tests has adequate consistency for test-retest comparison at about 4 weeks. In addition, none of the measures showed significant test-retest improvements indicating no prior exposure effects and change in motivation (Maruff. Darby, and McStephen, 2003).

Study II: It gives rise to two recommendations 1) Using the age-specific means for standardization in creating composite scores and 2) incorporating at least two practice sessions before establishing the study baseline for these tests.

REFERENCES

1. Button, E. (1996). Validation and invalidation. In J.W. Scheer & A. Catina (Eds.) Empirical Constructivism in Europe: The Personal Construct Approach. Giessen: Psychosozial-Verlag, pp. 142-148.
2. Habel, U., Gur, R.C., Mandal, M.K., Salloum, J.B., Gur, R.E., & Schneider,F. (2000) Emotional processing in schizophrenia across cultures:standardized measures of discrimination and experience. Schizophrenia Research 42:1, 57-66.
3. Kohler, C.G., Bilker, W., Hagendoorn, M., Gur, R.E., & Gur, R.C. (2000). Emotion recognition deficit in schizophrenia: Association with symptomatology and cognition. Biological Psychiatry 48: 127-136.
4. Poole, J.H., Tobias, F.C., & Vinogradav, S. (2000) The functional relevance of affect recognition errors in schizophrenia. Journal of the International Neuropsychological Society 6:6, 649-658.
5. Bilder, R.M., Turkel, E., Lipschutz-Broch, L., Lieberman, J.A. (1992). Antipsychotic medication effects on neuropsychological functions. Psychopharmacology Bulletin, 28, 353-366.
6. Lipschutz-Broch, L., Bilder, R.M., Turkel, E., Reiter, G., Lieberman, J.A. (1993, April). A chronometric analysis of executive deficits in acute andstable phases of first episode schizophrenia. IVth International Congress on Schizophrenia Research, Colorado Springs, CO.
7. Lipschutz-Broch, L., Bilder, R.M., Turkel, E., Reiter, G., Lieberman, J.A. (1993, May). Validation of new measures of executive dyscontrol. Fourth-Eighth Annual Convention of the Society for Biological Psychiatry, San Francisco, CA.

Cogtest Inc, 102 E Main Street,
Suite 305B, Newark, DE 19711, USA

Cogtest®

www.cogtest.com.