Neurol Sci (2016) 37:365–372
DOI 10.1007/s10072-015-2428-2

ORIGINAL ARTICLE

An updated Italian normative dataset for the Stroop color word test (SCWT)

A. Brugnolo[1] · F. De Carli[2] · J. Accardo[1] · M. Amore[3] · L. E. Bosia[4] · C. Bruzzaniti[1] · S. F. Cappa[5] · L. Cocito[1] · G. Colazzo[6] · M. Ferrara[1] · L. Ghio[3] · E. Magi[3] · G. L. Mancardi[1] · F. Nobili[1] · M. Pardini[1] · R. Rissotto[4] · C. Serrati[4] · N. Girtler[1,7]

Received: 14 May 2015 / Accepted: 20 November 2015 / Published online: 30 November 2015

Abstract The Stroop color and word test (SCWT) is widely used to evaluate attention, information processing speed, selective attention, and cognitive flexibility. Normative values for the Italian population are available only for selected age groups, or for the short version of the test. The aim of this study was to provide updated normal values for the full version, balancing groups across gender, age decades, and education. Two kinds of indexes were derived from the performance of 192 normal subjects, divided by decade (from 20 to 90) and level of education (4 levels: 3–5; 6–8; 9–13; >13 years). They were (i) the correct answers achieved for each table in the first 30 s (word items, WI; color items, CI; color word items, CWI) and (ii) the total time required for reading the three tables (word time, WT; color time, CT; color word time, CWT). For each index, the regression model was evaluated using age, education, and gender as independent variables. The normative data were then computed following the equivalent scores method. In the regression model, age and education significantly influenced the performance in each of the 6 indexes, whereas gender had no significant effect. This study confirms the effect of age and education on the main indexes of the Stroop test and provides updated normative data for an Italian healthy population, well balanced across age, education, and gender. It will be useful to Italian researchers studying attentional functions in health and disease.

Keywords Italian normative dataset · Selective attention · Executive functions · SCWT

✉ A. Brugnolo
Andrea.Brugnolo@unige.it

1 Clinical Neurology, DINOGMI University of Genoa, Genoa, Italy

2 Institute of Bioimaging and Molecular Physiology, National Research Council, Genoa, Italy

3 Clinical Psychiatry, DINOGMI University of Genoa, Genoa, Italy

4 Clinical Neurology, IRCCS San Martino-IST, Genoa, Italy

5 Division of Neuroscience, IUSS Pavia, San Raffaele Scientific Institute, Milan, Italy

6 C.H. La Palmosa, Accueil de Jour Alzheimer, Centre Hospitalier La Palmosa, Menton Division of Neuroscience, Menton, France

7 Clinical Psychology and Psychotherapy Unit, IRCCS San Martino-IST, Genoa, Italy

Introduction

The Stroop color and word test (SCWT) [1] is widely used to evaluate selective attention, inhibition, and sustained attention.

The classical version of the SCWT is composed of three tables, showing color words, colored squares or circles, and color words printed in incongruent ink (i.e., red printed in blue ink), respectively. The Stroop effect consists of a delayed response when words have to be named according to the color of the ink, ignoring the meaning of the printed word.

The SCWT evaluates the reaction times to non-ambiguous stimuli (reading words in black ink or naming colors in painted forms), and to ambiguous stimuli (color word printed with incongruent ink), thus assessing the inhibition mechanisms that are crucial to executive functions. [2, 3].

Traditionally, the SWCT has been used to evaluate frontal function, although neuroimaging studies show that a distributed network is activated during the Stroop effect. Indeed, studies with fMRI demonstrate the activation not only of the dorsolateral prefrontal cortex (DLPFC) and anterior cingulate cortex (ACC), but also of the posterior parietal cortex (PPC) during the Stroop effect [4–6], even if the role of each structure is a subject of debate [4, 7].

Performance of the SWCT is affected by many pathological conditions. The test is frequently part of neuropsychological batteries for the evaluation of cognitive functions in Alzheimer's disease (AD) [8], fronto-temporal lobar degeneration (FTLD) [9, 10], dementia with Lewy body (DLB) [10], vascular cognitive impairment [11–13], depression [14], schizophrenia [15], and anorexia [16]. The reaction times increase physiologically with aging and some activation patterns are slightly different between elderly and young people [17, 18].

In this frame, updated norms for the complete form of SCWT (that uses five colors and 100 items for each table with respect to short versions that use less colors and less items) in the Italian language are lacking. Indeed, the main reference for the SCWT was published in 1998 and analyzed a sample with only 15 subjects between 70 and 80 years and without subjects older than eighty [19]. In a more recent study including subjects in the ninth decade [20], a short version (30 items and three colors) of SCWT was used but only one index [20] was evaluated, while in another study, finalized to the assessment of multiple sclerosis [21], older age groups were not assessed, and, again, only one index was considered.

The need to update the normative values is crucial in the evaluation of cognitive functions that are affected by demographic and cultural transformation. The changes in living conditions such as the increased life expectancy, (http://www.worldlifeexpectancy.com/country-health-profile/italy, http://www.istat.it/it/archivio/99464), and the increase of education modify the cultural background of the general population.

The aim of our study was to produce updated normal values for the SCWT in Italian language drawn from a sample balanced for age, gender, and education level.

Materials and methods

Subjects

At first, we planned to enroll 32 subjects for each decade between 20 and 90 years for a total of 224 subjects. This sample size has been established in order to provide reliable correction values, by applying power analysis for multiple regression [22] using the pwr package in R [23], with the following parameters: probability level (α): 0.05, desired statistical power ($1-\beta$): 0.80, effect size (Cohen's f^2): 0.05, number of linear predictors: 3. These 32 subjects would have included 16 subjects per gender, evenly divided according to four education levels (primary school, middle school, high school, and university degree). During the recruitment period it became progressively clear that subjects with a primary-school level of education could be found only among elderly subjects because current Italian legislation requires a minimum of 8 years of education starting from age 6 (law 1859/62). Thus, we decided to omit the enrollment of the 24 subjects aged 20–49 with primary school education. The new target population was formed by 200 subjects to be enrolled within 18 months.

With the exception of people under 50 years old with the lowest education level, now almost disappeared in Italian population, the plan of our study was to analyze a sample evenly distributed as for age, education, and gender. Such a choice was aimed at estimating the expected values of the indexes of Stroop test as function of significant predictors rather than to be representative of the whole population. Subjects were healthy volunteers checked by means of a general medical history, clinical, and neurological examination. The cognitive status was assessed by means of a clinical interview and MMSE. Depression was rated by means of the Montgomery–Åsberg depression rating scale (MADRS) [24].

The inclusion criteria were the following: (i) age 20–90; (ii) education from primary school to university degrees; (iii) signed informed consent; (iv) MMSE >27 (>25 for age >65); (v) a score <10 at MADRS of the depression evaluation [25].

The exclusion criteria were (i) diabetes mellitus, either treated or not; (ii) severe arterial hypertension not properly controlled by drug therapy (diastolic blood pressure >109 mmHg); (iii) history of a cerebrovascular accident; (iv) history of transient global amnesia in the last 3 years; (v) history of brain injury with a loss of consciousness longer than 30′; (vi) history of brain injury with a loss of consciousness of at least 10′ in the last 6 months; (vii) central nervous system diseases (Parkinson, epilepsy, migraine, etc.); (viii) major psychiatric disorders (psychosis, major depression, etc.); (ix) chronic use of benzodiazepines, neuroleptic, and other sedative drugs (stable low doses of benzodiazepine, SSRI, or other hypnotic drugs were allowed); (x) evidence of severe systemic pathology not properly controlled including, but not limited to, renal failure (creatinine level >2 mg/dL); liver failure (transaminase levels >3 × ULN); untreated thyroid disease; anemia (hemoglobin levels <10 mg/dL); cancer in the last 5 years; (xi) chronic or occasional use of illicit psychotropic substances in the last month; (xii) habitual consumption of >750 cc wine/die or equivalent for daily alcohol intake; (xiii) illiteracy or less of 2 years of education; (xiv) history of polychemotherapy or radiotherapy;

Table 1 Number of subjects enrolled arranged by age and education

Age	Primary School		Middle School		High School		University		Total
	M	F	M	F	M	F	M	F	
20–29	–	–	4	4	4	4	4	4	24
30–39	–	–	4	1	2	4	3	4	18
40–49	–	–	4	4	6	4	5	4	27
50–59	2	3	4	4	4	3	4	4	28
60–69	4	3	2	4	4	5	3	4	29
70–79	4	5	6	4	4	4	5	2	34
80–90	4	4	4	4	4	4	4	4	32
Total	14	15	28	25	28	28	28	26	192

(xv) infancy or adolescence development disorder; (xvi) sensory deficit (hypoacusia, visual deficit).

At the end of the recruitment period, 192 subjects fulfilled these inclusion and exclusion criteria, thus almost reaching the intended number, and were enrolled. The distribution of subjects by age, gender, and education is reported in Table 1.

Overall average values of the 192 subjects were as follows: age 57.3 ± 19.6 years (range 20–90), education 11.5 ± 4.3 years (range 5–19), MMSE score 29.5 ± 0.8 (range 27–30), MADRS score 2.5 ± 2.5 (range 0–9).

The Stroop color and word test

The present version of SCWT consists of three tables, with one hundred stimuli. The three tables, each arranged in 10 rows and 10 columns, are composed by: (1) color words printed in black ink, (2) colored squares, (3) color words printed with an incongruent ink. The measures of the tables are 420 mm width × 480 mm height, the font used for table one and three is Arial 24, the measures of the squares in Table 2 (color) are 20 mm × 20 mm. The colors used in this version are blue, green, red, brown, and purple, printed by the Organizzazioni Speciali (Florence, Italy) [26].The following parameters were obtained and further considered. The correct answers achieved in the first 30 s for each table, generating three scores, namely word items (WI), color items (CI), and color word items (CWI); the total time needed for reading each table, generating three more scores, labeled word time (WT), color time (CT), and color word time (CWT).

The study protocol was approved by the local ethics committee. According to the recommendations of the Helsinki Declaration of 1975, as revised in 2008, all subjects were informed about the objectives and methods of the research, and they agreed to take part in the study. The study was explained to all participants both orally and by written instructions.

Statistics

The preliminary analysis evaluated the distribution of raw scores for each of the six indexes (WT, WI, CT, CI, WCT, and WCI). Power analysis was preliminarily computed to fix the sample size to provide reliable correction values, but considering some missing values with respect to the planned number, power analysis was re-computed considering the effective sample size. Subsequently, multiple regression analysis was performed for each index, using the demographic variables (age, education and gender), as independent variables.

The rationale of this procedure was based on the methodology of equivalent scores originally described by Spinnler and Tognoni [27] and then applied to several normative studies [30], including those on the Stroop Test, in the Italian language [19–21]. Following this approach, we applied multiple regression analysis to study the effect of age, education, and gender on Stroop indexes. Moreover, in order to take not-linear effects into account we applied the same data transformation as suggested by Spinnler and Tognoni [27], namely, the square root of education and the logarithmic transformation of age [$\ln(100 - \text{age})$]. We then evaluated four regression models in which age and education were both included as raw values, or as transformed values, or one raw and the other one transformed values, alternatively. Among these four models, we choose that with the best R2 value. Dealing with not-nested models, we applied the Akaike information criterion (AIC) [28] for comparing the adequacy of different models [29]. For each model the AIC weights is reported and can be directly interpreted as the likelihood or relative probability of being the best model. The rates between AIC weights were also computed to compare the adequacy of couple of models.

At the following step, the equations to adjust scores for age and education were drawn from the best fit model for each of the six indexes. They were used to standardize all

Table 2 Comparison between regression models as fitted with or without transformation of predictors

	Education; age	$\sqrt{\text{Education}}$; age	Education; ln (100-age)	$\sqrt{\text{Education}}$; ln (100 − age)
WT				
R^2	0.179	0.185	0.191	0.197
AIC weight	0.067	0.137	0.263	0.535
AIC ratio	8.036	3.896	2.042	1
WI				
R^2	0.247	0.252	0.263	0.267
AIC weight	0.049	0.077	0.341	0.534
AIC ratio	10.994	6.967	1.567	1.000
CT				
R^2	0.323	0.329	0.357	0.362
AIC weight	0.002	0.005	0.316	0.676
AIC ratio	289.461	136.757	2.137	1.000
CI				
R^2	0.368	0.37	0.376	0.377
AIC weight	0.118	0.140	0.336	0.406
AIC ratio	3.426	2.909	1.208	1.000
CWT				
R^2	0.444	0.449	0.473	0.479
AIC weight	0.001	0.003	0.232	0.764
AIC ratio	690.181	222.113	3.298	1.000
CWI				
R^2	0.589	0.591	0.602	0.605
AIC weight	0.013	0.022	0.342	0.623
AIC ratio	47.350	27.917	1.824	1.000

For each model the following parameters are reported: R^2 coefficient of determination, representing the proportion of explained variance; *AIC weight* relative model adequacy as measured by the Akaike information criterion (AIC) expressing the likelihood that the current model is the best model; *AIC ratio* the ratio between AIC weight of the current model and the one of the best models

raw values and to build-up the tables reporting a correction value for each class of age and education, as computed for predictors at their central value. Gender was not considered as it was not significant at regression analysis. Reference limits were then computed by analyzing the whole sample of age- and education-corrected values. The corrected score was used to define the cut-off, in accordance with the system of equivalent scores adopted in many Italian normative [19–21, 27, 30]. The cut-off for each index was computed by the resolution of Wilks' integral equations [31] for 95 % tolerance limits at 95 % confidence level.

The cut-off value separates pathological performances from normal performances and defines the values corresponding to the equivalent score of zero. According to the method of equivalent scores, the scores were classified into five ranges corresponding to five categories (0, 1, 2, 3, 4). The equivalent score of 4 identifies the performances above the median value while the equivalent scores of 1, 2, 3 partition the intermediate range (between cut-off and median value) according to specific percentile ranks [27, 30].

Statistical analysis was performed by the statistic software SPSS 17.00 (SPSS Inc., Chicago, USA) and MatLab R2014a (MathWorks, Natick, Massachusetts, USA) and using the pwr package in R (http://CRAN.R-project.org/package=pwr).

Results

All subjects concluded the test with the following mean times: WT: 52.7 ± 21.5, CT: 79.1 ± 25.9, CWT: 151.9 ± 60.7, and with the following mean number of items read in the first 30 s early-items scores: WI: 64.92 ± 16.82, CI: 43.9 ± 11.8, CWI: 23.8 ± 8.1.

Power analysis was performed considering 3 predictors, a significance level of 0.05 and a power of 0.80 in the actual sample of 192 subjects. Results showed that the available number of participants allowed to detect a significant effect with an effect size equal to 0.0588, which is between the small (0.02; i.e., the best) and the moderate

(0.15) range [22] that is reasonable for a reliable regression analysis.

The best linear regression model for each index always included age and education as significant regressors, while gender never reached the statistical significance, the probability levels found for the effect of gender were, respectively: WT: $p = 0.152$; WI: $p = 0.071$; CT: $p = 0.704$; CI: $p = 0.931$; CWT: $p = 0.597$; CWT: $p = 0.235$. Gender was then excluded from regression models..

Using the Akaike information criterion (AIC) [28], we found that the ranking of AIC weights relevant to the four models was the same for all the indexes and the model based on transformed age and education was the best while the differences were mainly associated with the logarithmic transformation of age (see Table 2).

In all cases, the best fit was associated with transformed independent variables (i.e., natural logarithm of age and square root of education) and yielded significant models for WT ($R^2 = 0.197$, $F_{2,189} = 23.176$, $p < 0.001$), WI ($R^2 = 0.267$, $F_{2,189} = 34.369$, $p < 0.001$), CT ($R^2 = 0.362$, $F_{2,189} = 45.148$, $p < 0.001$), CI ($R^2 = 0.377$, $F_{2,189} = 57.204$, $p < 0.001$), CWT ($R^2 = 0.479$, $F_{2,189} = 86.908$, $p < 0.001$), CWI ($R^2 = 0.605$, $F_{2,189} = 144.726$, $p < 0.001$). Statistical data relevant to the effect of age and education on each index are reported in detail in Table 3.

The raw values of the six indexes were thus corrected according to the equations of the multiple regression models as shown in Table 4, which also reports the correction values arranged for age and education classes. The equivalent scores and the relevant score range for each index are reported in Table 5.

Discussion

The aim of our study was to produce updated normative values for one of the most used attention tests, the SCWT, in an Italian population by recruiting subjects balanced as for education, gender, and age. Such sampling arrangement was aimed at estimating the expected values of the indexes of Stroop test as a function of significant predictors. The sample did not reflect the age, education, and gender distribution of the general population, as in such case it could have introduced a bias, or lower precision, in the estimation of expected values for some predictor intervals, in particular for the highest age and education values.

To the best of our knowledge, only one among the three studies published so far about the Italian version of the Stroop has enrolled subjects up to the age of ninety (52.1 ± 19.56, range 20–90 years) [20] using the brief form of the SCWT (with thirty items and three colors). The main study that utilized the full version did not enroll subjects beyond the age of 81 (mean age: 49 ± 15.6, range 18–81 years) and was published more than 15 years ago [19], while in the third study, only one index was validated in a sample with a low mean age (40.9 ± 11) [21]. In the present study, we provide norms for an extended age range with balanced groups (mean age: 57.3 ± 19.6 years range 20–90), using the complete version of the SCWT (with one hundred items for each of the three tables and five colors) and analyzing several indexes.

The multiple regression analysis showed that age and education had statistically significant influence on the performances as evaluated by all the six indexes of the SCWT, whereas gender did not show significant effects.

The decline in performance with increasing age complies with previous Italian data both in the extended [19] and short forms [20, 21]. This effect has also been found in previous studies concerning SCWT implemented in other languages [32–36] and in experimental studies with fMRI [17, 37]. Education is a second variable that positively correlated with the SCWT performance. This effect is in keeping with two out of three previous Italian studies [19, 20]. The positive relationship between education and performance has also been found in other populations worldwide [32].

Table 3 Values of linear regression models

Indexes of Stroop test	ln (100 − age)			$\sqrt{\text{Education}}$		
	β	t	p	β	t	p
WT	−0.26	−3.97	<0.001	−0.30	−4.48	<0.001
WI	0.33	5.18	<0.001	0.32	5.12	<0.001
CT	−0.49	−8.30	<0.001	−0.24	−4.14	<0.001
CI	0.23	4.03	<0.001	0.51	8.72	<0.001
CWT	−0.24	−4.57	<0.001	−0.59	−11.00	<0.001
CWI	0.30	6.53	<0.001	0.64	13.80	<0.001

The table reports standardized beta values (β). t values and p values of the following regressors: logarithm of age ln (100 − age) and education transformed in its square root ($\sqrt{\text{Education}}$), for each index

Table 4 Correction table for the raw scores of the six indexes

Age	WT (word time)				WI (word item)			
	Primary School	Middle School	High School	University	Primary School	Middle School	High School	University
20–29	–	0.95	8.38	12.87	–	−1.58	−7.91	−11.68
30–39	–	−1.16	6.50	11.73	–	0.47	−6.09	−10.56
40–49	–	−2.43	5.05	9.59	–	1.71	−4.67	−8.49
50–59	−11.11	−5.04	2.65	7.46	9.46	4.25	−2.33	−6.42
60–69	−13.50	−7.51	0.44	5.24	11.78	6.65	−0.19	−4.26
70–79	−16.64	−10.25	−3.27	1.80	14.84	9.31	3.42	−0.91
80–90	−23.02	−15.93	−8.91	−3.91	21.05	14.83	8.90	4.64

Best multiple linear equation models of WT and WI

$WT = \text{Raw score} + (\sqrt{\text{Education}} - 3.4) \times 9.627 + \ln(100 - \text{age}) - 3.74] \times 10.88$

$WT = \text{Raw scores} + (3.4 - \sqrt{\text{Education}}) \times 8.207 + [3.74 - \ln(100 - \text{age})] \times 10.584$

Age	CT (color time)				CT (color item)			
	Primary School	Middle School	High School	University	Primary School	Middle School	High School	University
20–29	–	8.98	16.29	20.13	–	−4.45	−7.65	−9.30
30–39	–	4.25	12.08	17.55	–	−2.20	−5.66	−8.08
40–49	–	1.39	8.82	12.77	–	−0.85	−4.11	−5.82
50–59	−10.93	−4.46	3.42	7.99	4.80	1.93	−1.55	−3.55
60–69	−16.28	−10.00	−1.53	3.02	7.34	4.56	0.79	−1.19
70–79	−23.33	−16.14	−9.85	−4.70	10.68	7.46	4.74	2.47
80–90	−37.64	−28.87	−22.49	−17.52	17.47	13.51	10.74	8.55

Best multiple linear equation models of CT and CI

$CT = \text{Raw score} + (\sqrt{\text{Education}} - 3.4) \times 9.545 + \ln(100 - \text{age}) - 3.74] \times 24.414$

$CI = \text{Raw score} + (3.4 - \sqrt{\text{Education}}) \times 4.195 + [3.74 - \ln(100 - \text{age})] \times 11.584$

Age	CWT (color word time)				CWI (color word item)			
	Primary School	Middle School	High School	University	Primary School	Middle School	High School	University
20–29	–	27.69	44.68	53.13	–	−3.77	−6.58	−8.04
30–39	–	14.43	32.89	45.92	–	−1.84	−4.87	−6.99
40–49	–	6.43	23.78	32.53	–	−0.68	−3.55	−5.05
50–59	−25.44	−9.96	8.66	19.15	4.21	1.70	−1.35	−3.11
60–69	−40.44	−25.49	−5.21	5.22	6.39	3.96	0.66	−1.08
70–79	−60.18	−42.66	−28.51	−16.40	9.26	6.45	4.05	2.06
80–90	−100.25	−78.33	−63.90	−52.29	15.08	11.63	9.19	7.27

Best multiple linear equation models of CWT and CWI

$CWT = \text{Raw score} + (\sqrt{\text{Education}} - 3.4) \times 22.284 + [\ln(100 - \text{age}) - 3.74] \times 63.366$

$CWI = \text{Raw score} + (3.4 - \sqrt{\text{Education}}) \times 3.684 + [3.74 - \ln(100 - \text{age})] \times 9.929$

Age- and education-adjusted scores for the six indexes: WT, WI, CT, CI, CWT, and CWI (details of these indexes in the text). The values in the tables, arranged for age and education intervals, provide a mean adjustment to be added to the observed score to obtain an age- and education-independent score. Adjustment formulae are provided and can be applied to specific cases to obtain precise corrections

The correlation with gender did not reach the statistical significance for any index. The effect of gender on SCWT performance is controversial. Considering the previous Italian normative studies, only one reported a significant effect of gender using the classical ‘paper’ version but not in the computer version [19]. The remaining two studies [20, 21] are in keeping with our findings, even though partially different indexes were used. The international normative data do not describe the effect of gender in a Caucasian population [33], but only in an Asiatic sample [32]. Moreover, even in the original study of Stroop [1], data were inconclusive, because in the first experiment reaction times were similar between males and females, while in the third experiment females performed better than males The authors suggested that the difference could be ascribed to differences in educational background between males and females.

Table 5 Equivalent scores for six indexes (age- and education-corrected scores)

Equivalent score	WT range	WI range	CT range	CI range	CWT range	CWI range
0	≥105.78	≤38.2	≥129.2	≤30.08]	≥257.83]	≤15.80
1	(105.77–73.40]	(38.3–47.0]	(129.19–97.64]	(30.09–33.9]	(257.82–193.4]	(15.81–17.9]
2	(73.40–55.10]	(47.0–57.1]	(97.65–82.47]	(33.9–39.45]	(193.4–158.0]	(17.9–21.52]
3	(55.10–45.55]	(57.1–68.0]	(82.47–71.88]	(39.45–45.4]	(158.0–136.04]	(21.52–24.47]
4	<45.55	>68.00	<71.88	>45.4	<136.04	>24.47

Equivalent scores cut-off values for the six indexes of Stroop test to be applied to age- and education-adjusted scores. Square bracket indicates a right-closed (including endpoint) interval, while round bracket indicates a left-open interval

A possible limitation of this study arises from the regional composition of the sample, as subjects were enrolled in a single Northern Italian region, even if their origin was composite, due to inter-regional mobility. Further studies might clarify if additional regional or social factors affect the scoring.

This study introduced three indexes (WT, CT, CWT), computed as the total time needed for reading each table: these indexes were not described in the previous Italian studies and were considered here in order to specifically evaluate sustained attention through the reaction times to unambiguous and ambiguous stimuli. These indexes, together with the other three (WI,WI,CWI), could be useful in the identification of attentional profile both in brain pathologies of the third age, such as AD [8], FTLD [9, 10], and LBD [10] and in typical brain pathologies of the adulthood, such as brain injury [38] and multiple sclerosis [39].

In conclusion, we have collected an updated set of norms for the classical, full card version of the SCWT, in healthy subjects from 20 to 90 years old, balanced for age, education, and gender. Several indexes have been analyzed and cut-off values were derived, exploring different aspects of attention. This normal data set paves the way to further studies, exploring the usefulness of these indexes, either alone or taken together in various combinations, for the detection of attentional disorders in specific pathological conditions.

Compliance with ethical standards

Conflict of interest All authors disclose any actual or potential conflicts of interest including any financial, personal, or other relationship with other people or organizations that could inappropriately influence their work.

References

1. Stroop JR (1935) Studies of interference in serial verbal reactions. J Exp Psychol 18:643–662
2. Neill WT (1977) Inhibitory and facilitatory processes in selective attention. J Exp Psychol-Hum Percept Perform 3:444–450. doi:10.1037/0096-1523.3.3.444
3. Barkley RA (1997) Behavioral inhibition, sustained attention and executive functions: constructing a unifying theory of ADHD. Psychol Bull 121:65–94. doi:10.1037/0033-2909.121.1.65
4. Chen Z, Lei X, Ding C, Li H, Chen A (2012) The neural mechanisms of semantic and response conflicts: an fMRI study of practice-related effects in the Stroop task. Neuroimage 66C:577–584. doi:10.1016/j.neuroimage.2012.10.028
5. Roberts KL, Hall DA (2008) Examining a supramodal network for conflict processing: a systematic review and novel functional magnetic resonance imaging data for related visual and auditory stroop tasks. J Cogn Neurosci 20:1063–1078. doi:10.1162/jocn.2008.20074
6. Liu X, Banich MT, Jacobson BL, Tanabe JL (2004) Common and distinct neural substrates of attentional control in an integrated Simon and spatial Stroop task as assessed by event-related fMRI. Neuroimage 3:1097–1106
7. Roelofs A, van Turennout M, Coles MG (2006) Anterior cingulate cortex activity can be independent of response conflict in Stroop-like tasks. Proc Natl Acad Sci U S A 103:13884–13889
8. Amieva H, Phillips LH, Della Sala S, Henry JD (2004) Inhibitory functioning in Alzheimer's disease. Brain 2004(127):949–964
9. Dopper EG, Rombouts SA, Jiskoot LC, den Heijer T, de Graaf JR, de Koning I, Hammerschlag AR, Seelaar H, Seeley WW, Veer IM, van Buchem MA, Rizzu P, van Swieten JC (2014) Structural and functional brain connectivity in presymptomatic familial frontotemporal dementia. Neurology 83:e19–e26. doi:10.1212/WNL.0000000000000583
10. Johns EK, Phillips NA, Belleville S, Goupil D, Babins L, Kelner N, Ska B, Gilbert B, Inglis G, Panisset M, de Boysson C, Chertkow H (2009) Executive functions in frontotemporal dementia and Lewy body dementia. Neuropsychology 23:765–777. doi:10.1037/a0016792
11. Salvadori E, Poggesi A, Pracucci G, Inzitari D, Pantoni L (2015) Development and psychometric properties of a neuropsychological battery for mild cognitive impairment with small vessel disease: the VMCI-Tuscany Study. J Alzheimers Dis 43:1313–1323. doi:10.3233/JAD-141449
12. Kramer JH, Reed BR, Mungas D, Weiner MW, Chui HC (2002) Executive dysfunction in subcortical ischaemic vascular disease. J Neurol Neurosurg Psychiatry 72:217–220
13. Videbech P, Ravnkilde B, Gammelgaard L, Egander A, Clemmensen K, Rasmussen NA, Gjedde A, Rosenberg R (2004) The Danish PET/depression project: performance on Stroop's test linked to white matter lesions in the brain. Psychiatry Res 130:117–130
14. Westerhausen R, Kompus K, Hugdahl K (2011) Impaired cognitive inhibition in schizophrenia: a meta-analysis of the Stroop interference effect. Schizophr Res 133:172–181. doi:10.1016/j.schres.2011.08.025
15. Lallart E, Jouvent R, Herrmann FR, Perez-Diaz F, Lallart X, Beauchet O, Allali G (2014) Gait control and executive

dysfunction in early schizophrenia. J Neural Transm 121:443–450. doi:10.1007/s00702-013-1111-0
16. Ferro AM, Brugnolo A, De Leo C, Dessi B, Girtler N, Morbelli S, Nobili F, Rossi DS, Falchero M, Murialdo G, Rossini PM, Babiloni C, Schizzi R, Padolecchia R, Rodriguez G (2005) Stroop interference task and single-photon emission tomography in anorexia: a preliminary report. Int J Eat Disord 38:323–329
17. Zysset S, Schroeter ML, Neumann J, von Cramon DY (2007) Stroop interference, hemodynamic response and aging: an event-related fMRI study. Neurobiol Aging 28:937–946. doi:10.1016/j.neurobiolaging.2006.05.008
18. Mohtasib RS, Lumley G, Goodwin JA, Emsley HC, Sluming V, Parkes LM (2012) Calibrated fMRI during a cognitive Stroop task reveals reduced metabolic response with increasing age. Neuroimage 59:1143–1151. doi:10.1016/j.neuroimage.2011.07.092
19. Barbarotto L, Laiacona M, Frosio R, Vecchio M, Farinato A, Capitani E (1998) A normative study on visual reaction times and two Stroop colour-word tests. Ital J Neurol Sci 19:161–170
20. Caffarra P, Vezzadini G, Dieci F, Zonato A, Venneri A (2002) Una versione abbreviata del test di stroop: dati normativi nella popolazione italiana. Nuova Rivista di Neurologia 12:111–115
21. Amato MP, Portaccio E, Goretti B, Zipoli V, Ricchiuti L, De Caro MF, Patti F, Vecchio R, Sorbi S, Trojano M (2006) The Rao's Brief Repeatable Battery and Stroop Test: normative values with age, education and gender corrections in an Italian population. Mult Scler 12:787–793
22. Cohen J (1988) Statistical power analysis for the behavioral sciences, 2nd edn. Lawrence Erlbaum Associates, Hillsdale
23. Champely, S. (2015) pwr: Basic functions for power analysis. R package version 1.1-2. http://CRAN.R-project.org/package=pwr
24. Montgomery SA, Asberg M (1979) A new depression scale designed to be sensitive to change". Brit J Psychiat134(4): 382–89. doi:10.1192/bjp.134.4.382
25. Hawley CJ, Gale TM, Sivakumaran T, Hertfordshire Neuroscience Research group (2002) Defining remission by cut off score on the MADRS: selecting the optimal value. J Affect Disord 72:177–184
26. Venturini R, Lombardo Radice M, Imperiali MG (1983) Il "Colour-Word Test" o Test di Stroop. Organizzazioni Speciali, Florence
27. Spinnler H, Tognoni G (1987) Standardizzazione e Taratura Italiana di Test Neuropsicologici. Ital J Neurol Sci 6:8–21
28. Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Caski F (eds) Proceedings of the Second International Symposium on Information Theory. Akademiai Kiado, Budapest, pp 267–281
29. Wagenmakers EJ, Farrell S (2004) AIC model selection using Akaike weights. Psychon Bull Rev 11:192–196
30. Bianchi A, Dai Prà M (2008) Twenty years after Spinnler and Tognoni: new instruments in the Italian neuropsychologist's toolbox. Neurol Sci 29:209–217
31. Wilks SS (1941) Determination of sample size for setting tolerance limits. Ann Math Stat 12:91–96
32. Seo EH, Lee DY, Choo IH, Kim SG, Kim KW, Youn JC, Jhoo JH, Woo JI (2008) Normative study of the Stroop Color and Word Test in an educationally diverse elderly population. Int J Geriatr Psychiatry 23:1020–1027. doi:10.1002/gps.2027
33. Norman MA, Moore DJ, Taylor M, Franklin D Jr, Cysique L, Ake C, Lazarretto D, Vaida F, Heaton RK, the HNRC Group (2011) Demographically corrected norms for African Americans and Caucasians on the Hopkins Verbal learning test-revised, brief visuospatial memory test-revised, Stroop color and word test, and Wisconsin card sorting test 64-card version, 33:793–804. doi:10.1080/13803395.2011.559157
34. Comalli PE Jr, Wapner S, Werner H (1962) Interference effects of Stroop color-word test in childhood, adulthood, and aging. J Genet Psychol 100:47–53
35. Uttl B, Graf P (1997) Color-Word Stroop test performance across the adult life span. J Clin Exp Neuropsychol 19:405–420
36. Ludwig C, Borella E, Tettamanti M, de Ribaupierre A (2010) Adult age differences in the Color Stroop Test: a comparison between an item-by-item and a blocked version. Arch Gerontol Geriatr 51:135–142. doi:10.1016/j.archger.2009.09.040
37. Mohtasib RS, Lumley G, Goodwin JA, Emsley HC, Sluming V, Parkes LM (2012) Calibrated fMRI during a cognitive Stroop task reveals reduced metabolic response with increasing age. Neuroimage 59:1143–1151. doi:10.1016/j.neuroimage.2011.07.092
38. Shukla D, Devi BI, Agrawal A (2011) Outcome measures for traumatic brain injury. Clin Neurol Neurosurg 113:435–441. doi:10.1016/j.clineuro.2011.02.013
39. Patti F, Leone C, D'Amico E (2010) Treatment options of cognitive impairment in multiple sclerosis. Neurol Sci 31:S265–S269. doi:10.1007/s10072-010-0438-7