2018

Healthy Adolescent Performance With Standardized Scoring Tables for the MATRICS Consensus Cognitive Battery: A Multisite Study.

R. E. Smelror

K. N. Jørgensen

V. Lonning

I. Kelleher

M. Cannon

See next page for additional authors

Recommended Citation

Authors

R. E. Smelror, K. N. Jørgensen, V. Lonning, I. Kelleher, M. Cannon, P. DeRosse, A. K. Malhotra, K. H. Karlsgodt, O. A. Andreassen, I. Agartz, and +6 additional authors

Healthy Adolescent Performance With Standardized Scoring Tables for the MATRICS Consensus Cognitive Battery: A Multisite Study

Runar Elle Smelror*[1,2], Kjetil Nordbø Jørgensen[1,2], Vera Lonning[1,2], Ian Kelleher[3], Mary Cannon[3], Pamela DeRosse[4–6], Anil K. Malhotra[4–6], Katherine H. Karlsgodt[7,8], Ole A. Andreassen[1,9], Mathias Lundberg[10,11], Tobias Edbom[10], Neil Cleland[10], Torill Ueland[1,9,12], Anne Margrethe Myhre[9], Bjørn Rishovd Rund[12,13], and Ingrid Agartz[1,2,10]

[1]NORMENT, KG Jebsen Centre for Psychosis Research, Institute of Clinical Medicine, University of Oslo, Oslo, Norway; [2]Department of Psychiatric Research, Diakonhjemmet Hospital, Oslo, Norway; [3]Department of Psychiatry, Royal College of Surgeons in Ireland, Dublin, Ireland; [4]Department of Psychiatry, The Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY; [5]Center for Psychiatric Neuroscience, The Feinstein Institute for Medical Research, Manhasset, NY; [6]Division of Psychiatry Research, The Zucker Hillside Hospital, Division of Northwell Health, Glen Oaks, NY; [7]Department of Psychology, University of California, Los Angeles, CA; [8]Department of Psychiatry, University of California, Los Angeles, CA; [9]Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway; [10]Centre for Psychiatric Research, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden; [11]Department of Clinical Science and Education, Södersjukhuset, Internal medicine, Karolinska Institutet, Stockholm, Sweden; [12]Department of Psychology, University of Oslo, Oslo, Norway; [13]Research Department, Vestre Viken Hospital Trust, Drammen, Norway

*To whom correspondence should be addressed; Department of Psychiatric Research, Diakonhjemmet Hospital, University of Oslo, PO Box 85, Vinderen, 0319 Oslo, Norway; tel: +47-95744029, fax: +47-22495862, e-mail: runar.smelror@medisin.uio.no

Objective: The aim of this study was to develop standardized scores and scoring tables for test performance in healthy adolescents for the Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) Consensus Cognitive Battery (MCCB) for each year from 11 to 19 years of age, by sex, with *T* scores and percentile ranks. *Methods:* A total of 502 healthy participants (aged 11–19 years) from 7 cohorts from Ireland, Norway, Sweden, and United States, were included in this multisite study. Regression-predicted means for the MCCB tests, except the social cognition subtest, were calculated using the MCCB test scores as outcome variables and *age*, *age²*, *sex*, *age × sex* as predictors. The regression-predicted means for each combination of age and sex were added with the residuals from the entire cohort to yield the expected distribution of that group. Age effects were examined using regression models with *age* and *age²* as predictors. Sex differences were examined using Student's *t*-tests. *Results:* Significant positive age effects were found for all tests, except for the Brief Visuospatial Memory Test, revised (BVMT-R; measure of visual learning). Females performed significantly better than males on BACS Symbol coding (measure of speed of processing) and BVMT-R, while males performed significantly better than females on NAB Mazes (measure of reasoning and problem solving). Based on the regression-predicted distributions of scores, 19 standardized scoring tables for each test and domain were created. *Conclusions:* With the results from this study,

we have developed an accessible standardized data set of healthy adolescent test performance for the MCCB.

Key words: early-onset/psychosis/schizophrenia/neurocognition/neuropsychology

Introduction

The Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) Consensus Cognitive Battery (MCCB) is a hybrid cognitive test battery developed for use in clinical treatment trials for schizophrenia.[1] The MCCB is a result of the MATRICS initiative from the US National Institute of Mental Health (NIMH).[2] It includes 10 tests divided into 7 domains: Speed of processing, Attention/vigilance, Working memory, Verbal learning, Visual learning, Reasoning and problem solving, and Social cognition. In adults, the MCCB has shown good reliability and validity, and minimal practice effects.[2] The MCCB has been standardized and co-normed for adults aged 20–59 years by the developers of the battery.[1] The battery has also been used for assessment of cognition in children[3] and adolescents.[4] However, despite the success and popularity of the MCCB to date, norms have not yet been determined for children or adolescents.

Adolescence is a critical period for maturation of neurobiological processes that underlie higher cognitive

functions, and social and emotional behavior.[5,6] The expansion of standardized MCCB scores to adolescence is important because this period of life is associated with increased incidence of mental disorders, including schizophrenia,[7] and because it provides a method to examine patterns of performance between tests and domains included in the MCCB. Previous reports of healthy adolescent test performance using the MCCB[3,8–10] have not provided sufficient differentiation between age and sex strata to develop standardized scoring tables. As these studies have used age strata of 2 or more years, have relatively small sample sizes in each age group and lack differentiation by sex, their usefulness as norms for scoring of individual test results is limited.

The aim of this study was to develop standardized scores and scoring tables for each test and domain in the MCCB, with associated *T* scores and percentile ranks, based on regression-predicted scores of healthy test performance, within each year, stratified by sex, in the adolescent period from 11 to 19 years of age.

Methods

Relevant research groups were identified by database searching and by contacting researchers that could be aware of other groups with MCCB data for healthy children and adolescents. Principal investigators from 11 research groups that had collected MCCB data from healthy adolescents were contacted and invited to participate in this multisite study.

Participants

A total of 502 healthy participants between the ages 11 and 19 years were included. These comprised 7 cohorts from Ireland (*n* = 131), Norway (*n* = 178), Sweden (*n* = 26), and the United States (*n* = 167). In 5 out of 7 cohorts, the participants had been recruited as healthy controls (HC) to match clinical psychosis samples within their respective studies. Thus, the combined cohort of 502 participants may be regarded as a multinational healthy control sample rather than a population-based normative sample. Exclusion criteria for mental disorders were part of the protocols and are described below. All participants were screened for Axis I disorders, according to the *Diagnostic and statistical manual of mental disorders, fourth edition text revision (DSM-IV-TR)*. In each of the cohorts, written informed consent was obtained from the participants, parents, or guardians, according to the guidelines in the respective countries. All minor participants provided assent alongside parental written informed consent. All studies were approved by their local ethical committees and conducted in accordance with the Helsinki Declaration. For detailed information about the independent cohorts and the combined cohort, see table 1.

Study 1: The Thematically Organized Psychosis Study for Youth (YTOP), University of Oslo, Norway

Sixty-seven participants were recruited as HC to the study. The participants were recruited from the Norwegian population register living in the Oslo area and matched for age and sex to ensure similar distribution as the included patients. Participants were excluded if they currently met criteria for, or previously had received treatment for an Axis I disorder, illicit substance use or alcohol dependence, IQ < 70, head injury with loss of consciousness ≥10 minutes, a medical illness that could affect brain functioning, or first-degree relatives with a psychotic or bipolar disorder.

Study 2: The Early-Onset Study (EOS), University of Oslo, Norway

Eighty-three participants were recruited as HC to the study. The participants were recruited from the Norwegian population register living in the Oslo area and from selected schools. They were matched for age and sex to ensure similar distribution as the included patients. Participants were excluded if they currently or previously had met criteria for an Axis I disorder, illicit substance use or alcohol dependence, IQ < 70, head injury with loss of consciousness ≥30 minutes, a medical illness that could affect brain functioning, or first-degree relatives with a psychotic or bipolar disorder.

Study 3: The Thematically Organized Psychosis (TOP) Study, Oslo University Hospital and University of Oslo, Norway

Twenty-eight participants were recruited as HC to the study. The participants were recruited by random sampling from the Norwegian population register living in the Oslo area. Participants were excluded if they currently or previously had met criteria for an Axis I disorder, illicit substance use or alcohol dependence, IQ < 70, hospitalized head injury, a medical illness that could affect brain functioning, or first-degree relatives with a major depressive disorder, psychotic, or bipolar disorder.

Study 4: The Stockholm Child and Adolescent Psychosis Study (SCAPS), Karolinska Institutet, Sweden

Twenty-six participants were recruited as HC to the study. The participants were recruited from the Swedish population register living in the Stockholm area and matched for age to ensure similar distribution as the included patients. Participants were excluded if they currently met criteria for an Axis I disorder, or previously had met criteria for a psychotic or bipolar disorder, illicit substance use or alcohol dependence, head injury with loss of consciousness ≥3 minutes, a medical illness that could affect brain functioning, or first-degree relatives with a psychotic or bipolar disorder. All participants attended regular schools without any support measures and were therefore

Table 1. Demographic Characteristics and Test Statistics of the Participants by Cohort

Study	Age Range, Years	Mean Age (SD)	Year of Inclusion	Female (%)	Hand Dominance, Right (%)*	Ethnicity, Caucasian (%)*	IQ (SD)*	IQ Assessment	Symptom Assessment	Global Cog., Mean T Score (SD)*
YTOP (n = 67, 13%)	12–18	15.3 (1.57)	2013–2017	37 (55)	61 (92)	62 (93)	105 (12.4)	WASI[1]	K-SADS-PL[4]	48.7 (9.0)
EOS (n = 83, 17%)	12–19	16.0 (1.94)	2005–2007	43 (52)	76 (92)	77 (93)	108 (14.1)	WASI[1]	MINI[5]	53.3 (10.7)
TOP (n = 28, 6%)	17–19	18.3 (.61)	2010–2016	12 (43)	22 (79)	27 (96)	107[a] (9.4)	WASI[1]	PRIME-MD[6]	49.7 (9.8)
SCAPS (n = 26, 5%)	12–18	16.3 (1.37)	2015	21 (81)	24 (92)	N/A	N/A	N/A	K-SADS-PL[4]	52.4 (9.2)
ABD (n = 131, 26%)	11–13	11.6 (.56)	2009–2010	63 (48)	116 (91)	N/A	104[b] (15.3)	WRAT-4[2]	K-SADS-PL[4]	51.0 (9.6)
CANDY (n = 120, 24%)	11–19	15.5 (2.25)	2010–2013	58 (48)	N/A	66 (55)	105[c] (10.5)	WASI[1]/WRAT-3[3]	K-SADS-PL[4]/SCID-NP[7]	47.5 (10.1)
MEND (n = 47, 9%)	11–19	16.4 (2.53)	2013–2016	21 (45)	N/A	25 (53)	104 (10.4)	WASI[1]	K-SADS-PL[4]/SCID-NP[7]	N/A
Combined (N = 502, 100%)	11–19	14.8 (2.66)	2005–2017	255 (51)	299[e] (91)	257[f] (75)	105[d] (12.4)	N/A	N/A	50.0[g] (10.0)
Test statistics	N/A	$F_{(6, 495)} = 117.9$, $P < .001$**	N/A	$\chi^2_{(6)} = 12.0$, $P = .06$	$\chi^2_{(8)} = 9.2$, $P = .32$	$\chi^2_{(4)} = 68.4$, $P < .001$**	$F_{(5, 4)} = 1.3$, $P = .25$	N/A	N/A	$F_{(5, 405)} = 3.6$, $P = .003$**

*ns vary by cohort. Actual n indicated by letter superscript: [a]27; [b]64; [c]114; [d]402; [e]331; [f]345; [g]411.

**Post hoc Bonferroni: mean age: TOP > other cohorts: $P < .001$, ABD < other cohorts: $P < .001$. YTOP < MEND: $P = .02$; Ethnicity, Caucasian: YTOP, EOS, TOP >
CANDY, MEND: $P < .03$[11]; Global cognition: EOS > CANDY: $P = .002$. No other group differences were statistically significant.

[1]Wechsler Abbreviated Scale of Intelligence.[12]
[2]Wide Range Achievement Test, 4th ed.[13]
[3]Wide Range Achievement Test, 3rd ed.[14]
[4]Schedule for Affective Disorders and Schizophrenia for School Aged Children-Present and Lifetime Version.[15]
[5]MINI-International Neuropsychiatric Interview, Screening module.[16]
[6]Primary Care Evaluation of Mental Disorders interview.[17]
[7]Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders, non-patient Version.[18]

considered to be within the normal IQ range. No other intelligence assessment was performed.

Study 5: The Adolescent Brain Development (ABD) Study, Royal College of Surgeons in Ireland

One hundred thirty-one community-based participants were recruited from schools in Dublin and surrounding counties.[3] The purpose of the study was to investigate the prevalence of psychotic symptoms and general psychopathology, as well as associated cognitive and neurobiological abnormalities in an epidemiological sample. For the purposes of the current analyses, participants were excluded if they currently met criteria for an Axis I disorder, illicit substance use or alcohol dependence, or previously had met criteria for a psychotic or bipolar disorder, a medical illness that could affect brain functioning, or first-degree relatives with a psychotic or bipolar disorder.

Study 6: The Comprehensive Assessment of Neurodevelopment in Youth (CANDY) Study, Zucker Hillside Hospital, NY, United States

One hundred twenty participants were recruited to the study from the general population of a region bridging urban New York City with suburban Long Island. Participants were recruited through newspaper and Internet advertisements as well as posted flyers seeking healthy children and adolescents between the ages of 8 and 18 for a study aimed to understand normal brain development.[19] Participants were excluded if they currently or previously had met criteria for an Axis I disorder, illicit substance use or alcohol dependence, IQ < 70, head injury with loss of consciousness (for any amount of time), a medical illness that could affect brain functioning, taking medications with known cognitive effects, or first-degree relatives with a major depressive disorder, psychotic, or bipolar disorder.

Study 7: The Multimodal Assessment of Neurodevelopmental Disorders (MEND) Study, Zucker Hillside Hospital, NY, United States

Forty-seven participants were recruited as HC to the study. The participants were recruited from the general population via advertisement, recommendations from other participants and through outreach activities to educational organizations in the New York City area. Participants were excluded if they currently or previously had met criteria for an Axis I disorder, illicit substance use or alcohol dependence, IQ < 70, head injury with loss of consciousness (for any amount of time), a medical illness that could affect brain functioning, taking medications with known cognitive effects, conditions that were contraindications of MRI (claustrophobia, pregnancy, etc.), or first-degree relatives with a major depressive disorder, psychotic, or bipolar disorder. The CPT-IP test was not included in this study.

Neurocognitive Measures

Nine out of 10 tests from the MCCB were included. The *Social cognition* domain (measured with the MSCEIT Managing emotions test[20]) was not part of the assessment in any of the cohorts and was therefore excluded from this study. The 9 included tests covered 6 neurocognitive domains: (1) *Speed of processing*, measured with the BACS Symbol coding,[21] Trail making test, part A (TMT-A),[22] and Category fluency: Animal naming,[23] (2) *Attention/vigilance*, measured with the Continuous performance test, identical pairs (CPT-IP),[24] (3) *Working memory*, measured with the WMS-III Spatial span[25] and Letter-number span,[26] (4) *Verbal learning*, measured with the Hopkins verbal learning test, revised (HVLT-R),[27] (5) *Visual learning*, measured with the Brief visuospatial memory test, revised (BVMT-R),[28] and (6) *Reasoning and problem solving*, measured with the NAB Mazes.[29] The HVLT-R test was originally validated for ages ≥16. However, as the stimulus list consists of simple words and previously has been used in children and adolescents,[10,30,31] we considered the original test appropriate for our age groups. In the Norwegian and Swedish cohorts, licensed translated versions of the MCCB were used (see[32] and www.matricsinc.org/mccb).

Statistical Analyses

The main strategy for calculating regression-predicted scores for the age and sex groups is described in Iverson et al.[33] The general principle is that for each test, the variation in test scores present in the total age range is used to predict the distribution of scores for each combination of age (ie, 9 age groups from 11 to 19 years) and sex, resulting in a total of 18 groups. To achieve this, regression models are first fitted to the entire cohort. The regression-predicted means for each age and sex group are then summated with the residuals from the entire cohort, yielding the regression-predicted distribution of scores for that specific group.

Validation of Statistical Assumptions

When multiple regression is used for prediction of scores, it is crucial that the statistical assumptions are carefully investigated.[34,35] Thus, we initially examined the normality of whole-sample distributions for the raw scores of each of the 9 MCCB tests before analysis. If the distributions showed an absolute value of skewness of more than 0.8 they were transformed. Furthermore, we examined homogeneity of variance using one-way ANOVA models with Levene's test. Tests showing significant group differences in scores were transformed to achieve equal variances across age groups. Lastly, for each regression model under analysis we examined outliers, defined as ±3 standard deviations (SD) from the mean, and their potential influence on the model to ensure good model fit. Influence were evaluated using values for leverage

(≤0.2 considered to be safe[36]) and Cook's distance (<1 considered to be safe[37]).

Effects of Age and Sex

Previous studies have shown nonlinear development of cognitive performance during childhood and adolescence.[9,10,38] Therefore, regression models including *age* and age^2 as predictors were used to describe the effects of age on test performance. Sex differences were investigated using Student's *t*-tests. All tests were 2-tailed.

Standardization of Score Distributions

The following standardization procedure was performed: (1) Curve estimations (linear, logarithmic, and quadratic) of the age slope for each test were performed to identify the best fit for the data. Including quadratic terms in the model provided the best fit for all tests. (2) Multiple regression analyses were then performed with the MCCB raw or transformed scores as outcome variable and *age*, age^2, *sex, age × sex* as predictor variables. The unstandardized residuals and the unstandardized predicted scores (representing the regression-predicted mean for that age and sex) from each regression model were saved. (3) The regression-predicted means for each of the 18 groups were summated with the unstandardized residuals from the entire cohort (representing the variation relative to each group's mean score), resulting in the regression-predicted distributions of scores for each age group. These distributions were thereafter validated by comparing them with the MCCB raw scores to ensure that the predicted models were adequate fits of the original raw data. (4) The regression-predicted distributions were transformed to standard scores (*z* scores) using the SPSS standardization function, and further transformed to *T* scores using the formula 50 + (10 × [*z* score]). For the TMT-A test the scores were reversed so that higher scores were equal to better performance. (5) In domains containing more than 1 subtest (ie, speed of processing and working memory) and for the global cognition score (based on the 9 included tests), composite scores were calculated by summating the *T* scores of relevant tests and transforming the sum score to standard scores using the same procedure as described above. (6) The regression-predicted scores were rounded to nearest integer. Based on these distributions 19 detailed scoring tables were developed with associated *T* scores and percentile ranks. Italicized test scores in the scoring tables indicate that the scores are based on the regression-predicted means rather than based on individual scores from the participant distributions. All statistical analyses were performed in IBM SPSS Statistics for Windows, Version 24.

Results

Examining the normality of distributions of the MCCB raw scores, 3 tests were regarded as skewed and were transformed using formulas with best fit to the data. The BVMT-R and NAB Mazes tests had negatively skewed distributions and were transformed using reflect and square root. The TMT-A test had a positively skewed distribution and was transformed using natural logarithm. No notable deviation from normality was observed after completed transformations. Examination of homogeneity of variance indicated unequal variance between age groups for the WMS-III Spatial span, $F(8, 493) = 1.99$, $P = .046$, which was transformed using natural logarithm. Examination of outliers showed less than 1% outliers for all tests, except for Animal naming (1.2%). Leverage values were below .03 and Cook's distance values were below .07 for all tests, indicating little influence of extreme cases over the parameters of the model.

Age and Sex Effects

Examination of age effects on test performance was performed using regression models with *age* and age^2 as predictors. The overall models showed significant positive age effects for all tests ($P < .001$), except for the BVMT-R. Examination of the separate predictors revealed quadratic associations with age for 6 tests (HVLT-R, NAB Mazes, WMS-III Spatial span, TMT-A, BACS Symbol coding, Animal naming, all *P*-values for $age^2 < .005$), linear associations with age for 2 tests (Letter-number span, CPT-IP, all *P*-values for age < .001) and no significant association with age for one test (BVMT-R). We found significant sex differences for 3 tests, using independent samples *t*-tests. Females performed significantly better than males on BACS Symbol coding ($M_F = 58.6$, $SD = 10.5$ vs $M_M = 56.2$, $SD = 11.9$), $t(499) = 2.4$, $P = .02$, and BVMT-R ($M_F = 28.5$, 95% CI = [27.9, 29.1] vs $M_M = 27.3$, 95% CI = [26.6, 28.0]), $t(495) = 2.57$, $P = .01$. Males performed significantly better than females on NAB Mazes ($M_M = 21.4$, 95% CI = [20.8, 21.9] vs $M_F = 20.3$, 95% CI = [19.7, 21.0]), $t(497) = 2.39$, $P = .02$.

Regression-Predicted and Unadjusted Means

The multiple regression models using *age*, age^2, *sex, age × sex* as predictors, which were included for all tests, statistically significantly predicted performance on 8 out of 9 tests (BACS Symbol coding, $F(4, 496) = 53.42$, $P < .001$, adj. $R^2 = .295$, standard error of the estimate (SEE) = 9.43; Animal naming, $F(4, 481) = 21.84$, $P < .001$, adj. $R^2 = .147$, SEE = 5.17; TMT-A, $F(4, 479) = 57.45$, $P < .001$, adj. $R^2 = .319$, SEE = .31; CPT-IP, $F(4, 417) = 36.33$, $P < .001$, adj. $R^2 = .251$, SEE = .60; WMS-III Spatial span, $F(4, 497) = 14.96$, $P < .001$, adj. $R^2 = .100$, SEE = .19; Letter-number span, $F(4, 497) = 14.61$, $P < .001$, adj. $R^2 = .098$, SEE = 2.94; HVLT-R, $F(4, 497) = 6.35$, $P < .001$, adj. $R^2 = .041$, SEE = 4.13; NAB Mazes, $F(4, 494) = 29.73$, $P < .001$, adj. $R^2 = .187$, SEE = .88). No significant association was

found for the BVMT-R, $F(4, 492) = 1.85$, $p < .12$, adj. $R^2 = .007$, SEE = .83). The regression models are shown in figure 1. The regression-predicted means and SD for males and females are presented in table 2.

To compare the regression-predicted means with the original raw scores, we calculated unadjusted means and SD for each age and sex. The unadjusted means and SD for males and females are presented in tables 3 and 4.

Based on the regression-predicted distributions of scores, a total of 19 standardized scoring tables were made. These tables are found in the supplemental material. They include regression-predicted distributions of test scores and domain scores, with associated T scores and percentile ranks, for each age and sex.

Discussion

We present regression-predicted and unadjusted means of healthy adolescent performance on the MCCB, for each year from 11 to 19 years, stratified by sex. Based on the regression-predicted distributions of scores we present 19 standardized scoring tables with associated T scores and

percentile ranks. To date, all studies of the MCCB in healthy adolescents have reported on just one or, at most, 2 samples. This study represents the largest analysis of MCCB in healthy adolescents to date, combining data from 7 cohorts with more than 500 participants. Although our resulting scoring tables do not represent population-based norms, expansion of standardized scores to adolescence is essential as this stage in life is an important period for cognitive development[5,6] and for the emergence of mental disorders.[7]

We found significant quadratic age effects in 6 out of 9 tests, covering the Speed of processing, Verbal learning and Reasoning and problem solving domains, and for the WMS-III Spatial span test, included in the Working memory domain. This indicates that performance on these tests improves at the onset of puberty and levels off during late adolescence. We found significant linear age effects for the Attention/vigilance domain and the Letter-number span test, included in the Working memory domain, indicating a steady increase in test performance with higher age. We did not find significant age effects for the test used to measure Visual learning, suggesting that performance on this test is stable across the adolescent period.
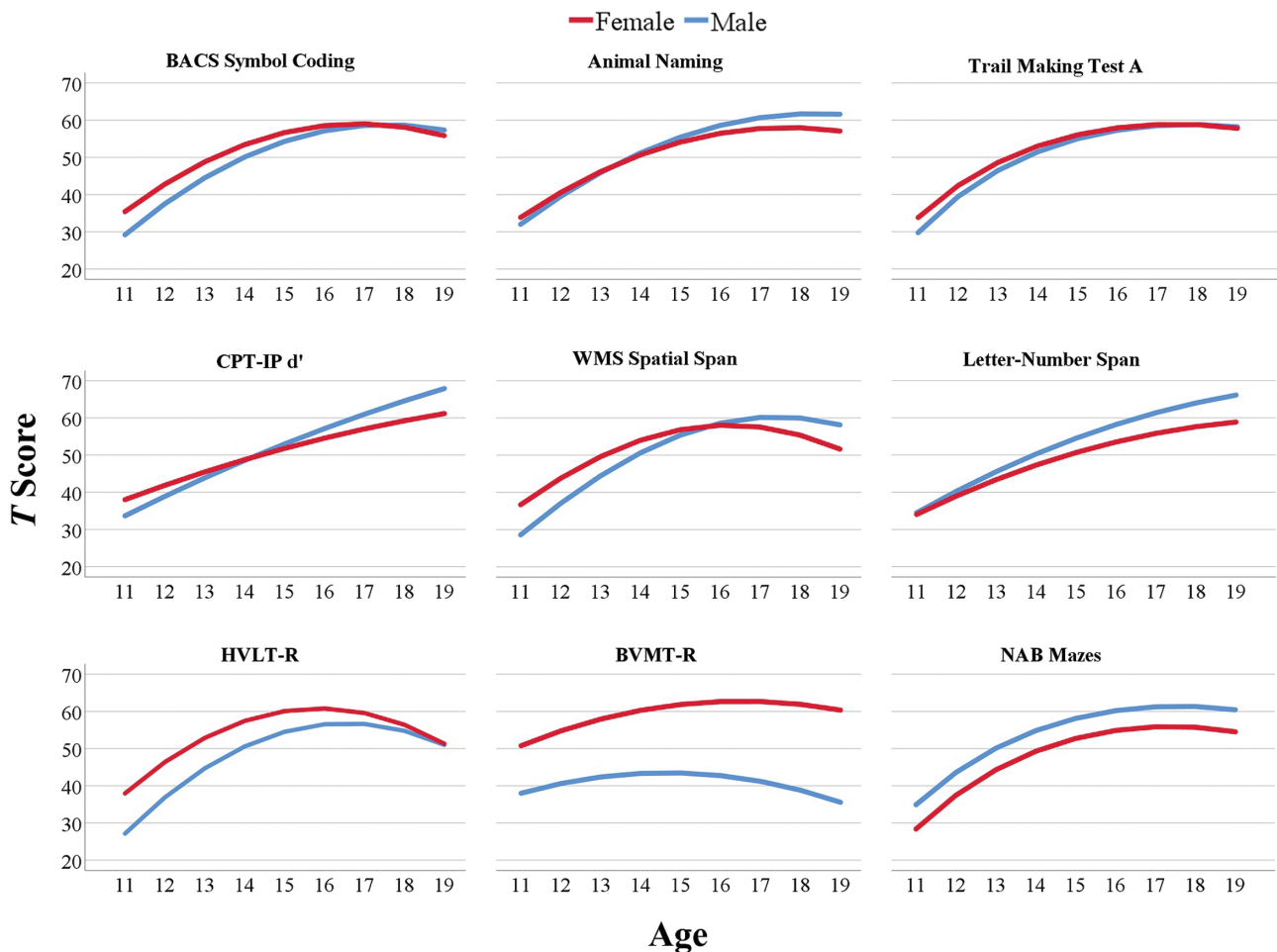

Fig. 1. Regression models of test performance for each neurocognitive test included in the MCCB from 11 to 19 years of age, separated by sex, and standardized to T scores.

Table 2. MCCB Subtest Regression-Predicted Means by Age and Sex

MCCB Tests	Sex	Age: Regression-Predicted Means (SD)								
		11	12	13	14	15	16	17	18	19
BACS Symbol coding	Male	44.63 (9.39)	49.77 (9.39)	54.06 (9.39)	57.50 (9.39)	60.10 (9.39)	61.84 (9.39)	62.73 (9.39)	62.78 (9.39)	61.97 (9.39)
(N = 501)	Female	48.47 (9.39)	53.01 (9.39)	56.71 (9.39)	59.55 (9.39)	61.55 (9.39)	62.70 (9.39)	62.99 (9.39)	62.44 (9.39)	61.04 (9.39)
Animal naming	Male	18.99 (5.15)	20.63 (5.15)	22.04 (5.15)	23.20 (5.15)	24.13 (5.15)	24.82 (5.15)	25.28 (5.15)	25.50 (5.15)	25.48 (5.15)
(N = 486)	Female	19.40 (5.15)	20.86 (5.15)	22.09 (5.15)	23.09 (5.15)	23.84 (5.15)	24.36 (5.15)	24.64 (5.15)	24.69 (5.15)	24.50 (5.15)
Trail making test A	Male	46.06 (14.90)	39.17 (12.67)	34.17 (11.05)	30.59 (9.90)	28.09 (9.09)	26.47 (8.56)	25.59 (8.28)	25.38 (8.21)	25.83 (8.36)
(N = 484)	Female	43.18 (13.97)	37.06 (11.99)	32.64 (10.56)	29.50 (9.54)	27.35 (8.85)	26.01 (8.41)	25.38 (8.21)	25.42 (8.22)	26.11 (8.45)
CPT-IP d′	Male	1.50 (.60)	1.69 (.60)	1.86 (.60)	2.03 (.60)	2.18 (.60)	2.33 (.60)	2.46 (.60)	2.59 (.60)	2.70 (.60)
(N = 422)	Female	1.65 (.60)	1.79 (.60)	1.92 (.60)	2.03 (.60)	2.14 (.60)	2.24 (.60)	2.32 (.60)	2.40 (.60)	2.47 (.60)
WMS-III Spatial span	Male	14.81 (2.73)	15.75 (2.91)	16.58 (3.06)	17.26 (3.18)	17.79 (3.28)	18.15 (3.35)	18.32 (3.38)	18.31 (3.38)	18.10 (3.34)
(N = 502)	Female	15.72 (2.90)	16.50 (3.04)	17.15 (3.16)	17.64 (3.25)	17.96 (3.31)	18.09 (3.34)	18.03 (3.33)	17.79 (3.28)	17.38 (3.21)
Letter-number span	Male	13.01 (2.93)	13.59 (2.93)	14.12 (2.93)	14.60 (2.93)	15.02 (2.93)	15.39 (2.93)	15.71 (2.93)	15.97 (2.93)	16.18 (2.93)
(N = 502)	Female	12.96 (2.93)	13.46 (2.93)	13.90 (2.93)	14.29 (2.93)	14.63 (2.93)	14.92 (2.93)	15.15 (2.93)	15.33 (2.93)	15.45 (2.93)
HVLT-R	Male	24.49 (4.12)	25.40 (4.12)	26.12 (4.12)	26.67 (4.12)	27.04 (4.12)	27.23 (4.12)	27.23 (4.12)	27.06 (4.12)	26.71 (4.12)
(N = 502)	Female	25.50 (4.12)	26.28 (4.12)	26.89 (4.12)	27.31 (4.12)	27.56 (4.12)	27.62 (4.12)	27.51 (4.12)	27.21 (4.12)	26.74 (4.12)
BVMT-R	Male	26.50 (5.38)	26.66 (5.34)	26.77 (5.31)	26.82 (5.30)	26.83 (5.30)	26.79 (5.31)	26.69 (5.33)	26.55 (5.37)	26.35 (5.42)
(N = 497)	Female	27.28 (5.18)	27.52 (5.11)	27.72 (5.06)	27.86 (5.01)	27.96 (4.99)	28.01 (4.97)	28.01 (4.97)	27.96 (4.99)	27.87 (5.01)
NAB Mazes	Male	16.53 (5.70)	18.48 (5.13)	19.93 (4.65)	21.00 (4.27)	21.73 (3.99)	22.20 (3.79)	22.43 (3.69)	22.45 (3.69)	22.25 (3.77)
(N = 499)	Female	15.07 (6.10)	17.10 (5.54)	18.63 (5.08)	19.75 (4.72)	20.53 (4.45)	21.01 (4.27)	21.22 (4.19)	21.20 (4.20)	20.93 (4.30)

Note: Different number of participants in the tests is due to missing data for individual participants, except for the CPT-IP test which was not included in one of the studies.

Our results agree with Waber et al,[38] who found quadratic age effects in most cognitive tests included in the US National Institute of Health's study of normal brain development from age 6 to 18 years. Our results are also in agreement with Gur et al,[39] who found significant age effects for the cognitive tests in their population-based sample of children from age 8 to 21 years, except for the visual and verbal memory tests. They suggested that the missing age effects in these tests could be a result of early maturation in brain structures associated with learning and memory. Gur et al[39] also suggested that ceiling effects might have occurred in the verbal test included in their test battery due to modifications in the word readability level for the youngest age group, which can potentially explain the difference from our findings on verbal learning.

We found significant sex differences in the Reasoning and problem solving domain, with males performing better than females. Females performed better in the Visual learning domain and on the BACS Symbol coding test, included in the Speed of processing domain. These findings are in keeping with Waber et al,[38] who showed that, as a group, males performed better on perceptual tests, while females performed better on processing speed tests. Our results are also partly in keeping with data from Kern et al[1] on healthy adult MCCB performance, which showed that males performed significantly better than females in the Reasoning and problem solving and Working memory domains, while females performed better in the Verbal learning domain.

Visual inspection of figure 1 shows that for most tests females perform better than males during early

Table 3. MCCB Subtest Unadjusted Means by Age for Males

MCCB Tests	Age: Unadjusted Means (SD)								
	11	12	13	14	15	16	17	18	19
	$n = 42$	$n = 41$	$n = 15$	$n = 17$	$n = 22$	$n = 29$	$n = 26$	$n = 35$	$n = 20$
BACS Symbol coding	46.26	46.73	56.40	60.75	58.18	61.72	63.04	63.23	61.50
($N = 246$)	(7.64)	(9.09)	(10.43)	(7.74)	(8.06)	(11.65)	(10.96)	(11.13)	(8.61)
Animal naming	18.67	21.02	24.13	23.35	22.32	24.08	24.38	27.24	24.85
($N = 240$)	(3.77)	(4.68)	(6.00)	(4.74)	(5.61)	(5.09)	(7.87)	(5.44)	(5.41)
Trail making test A	44.00	41.31	30.99	26.45	30.98	27.37	27.82	25.26	24.84
($N = 237$)	(12.84)	(12.11)	(9.97)	(6.32)	(9.33)	(9.42)	(9.32)	(9.21)	(13.64)
CPT-IP d′	1.46	1.73	1.83	2.17	2.10	2.32	2.44	2.56	2.78
($N = 208$)	(.77)	(.58)	(.91)	(.35)	(.51)	(.78)	(.46)	(.63)	(.44)
WMS-III Spatial span	14.79	15.56	17.60	17.47	17.18	17.90	18.42	18.60	18.25
($N = 247$)	(2.35)	(2.96)	(3.54)	(3.50)	(3.65)	(3.15)	(3.64)	(3.23)	(4.90)
Letter-number span	12.55	13.90	15.40	14.53	14.23	15.55	16.04	15.77	16.15
($N = 247$)	(2.39)	(3.03)	(4.64)	(2.50)	(2.58)	(2.34)	(2.68)	(2.97)	(3.22)
HVLT-R	23.69	26.46	26.27	27.18	25.50	26.69	28.77	27.31	25.70
($N = 247$)	(5.03)	(3.61)	(5.32)	(2.88)	(4.77)	(3.75)	(4.78)	(3.29)	(4.09)
BVMT-R	25.05	27.38	28.27	29.67	26.25	26.79	25.81	26.29	26.40
($N = 242$)	(6.30)	(4.73)	(5.82)	(3.37)	(6.15)	(4.66)	(5.11)	(5.93)	(5.72)
NAB Mazes	16.71	18.34	22.47	20.00	19.82	21.45	22.96	23.09	22.75
($N = 245$)	(5.47)	(5.22)	(4.00)	(3.67)	(5.41)	(4.65)	(3.49)	(3.13)	(2.51)

Note: Different number of participants in the tests is due to missing data for individual participants, except for the CPT-IP test which was not included in one of the studies.

Table 4. MCCB Subtest Unadjusted Means by Age for Females

MCCB Tests	Age: Unadjusted Means (SD)								
	11	12	13	14	15	16	17	18	19
	$n = 29$	$n = 40$	$n = 15$	$n = 31$	$n = 23$	$n = 27$	$n = 40$	$n = 31$	$n = 19$
BACS Symbol coding	49.38	51.33	56.47	60.77	64.70	61.04	61.85	63.35	60.84
($N = 255$)	(8.75)	(8.49)	(6.09)	(11.08)	(9.78)	(10.35)	(8.28)	(9.06)	(9.34)
Animal naming	19.38	21.13	20.47	22.55	24.65	25.65	24.31	25.16	23.39
($N = 246$)	(4.59)	(5.07)	(3.56)	(5.23)	(5.57)	(4.91)	(5.61)	(4.95)	(3.53)
Trail making test A	44.50	37.87	29.21	29.29	22.89	27.66	25.42	26.20	26.60
($N = 247$)	(13.90)	(9.18)	(9.13)	(9.66)	(5.83)	(8.83)	(7.18)	(8.88)	(9.56)
CPT-IP d′	1.70	1.76	1.83	2.02	2.27	2.19	2.37	2.33	2.49
($N = 214$)	(.58)	(.64)	(.73)	(.58)	(.54)	(.51)	(.54)	(.52)	(.43)
WMS-III Spatial span	15.72	16.18	17.93	17.77	18.30	17.93	17.78	17.48	17.84
($N = 255$)	(2.91)	(3.15)	(3.06)	(3.01)	(3.28)	(3.14)	(2.55)	(3.49)	(3.66)
Letter-number span	13.31	13.45	14.07	13.48	14.48	15.30	15.25	15.58	15.16
($N = 255$)	(2.55)	(3.21)	(2.76)	(3.45)	(3.36)	(3.59)	(2.47)	(2.85)	(2.67)
HVLT-R	26.55	26.05	25.27	26.19	28.48	28.00	27.55	27.84	25.95
($N = 255$)	(3.65)	(4.47)	(3.97)	(3.47)	(4.45)	(4.12)	(3.26)	(3.98)	(4.64)
BVMT-R	27.48	27.75	27.13	27.55	28.65	26.67	27.93	29.03	28.05
($N = 255$)	(6.14)	(6.10)	(3.46)	(3.82)	(4.07)	(5.70)	(4.50)	(3.83)	(5.52)
NAB Mazes	13.69	16.90	19.60	21.87	21.39	20.19	20.28	20.70	21.84
($N = 254$)	(5.18)	(4.95)	(5.47)	(4.09)	(4.15)	(4.39)	(4.98)	(4.24)	(4.48)

Note: Different number of participants in the tests is due to missing data for individual participants, except for the CPT-IP test which was not included in one of the studies.

adolescence, and that this sex difference evens out by mid/late adolescence. For the CPT-IP and the WMS Spatial span tests, males showed continued increase in performance. This pattern may reflect the fact that females, on average, enter puberty earlier than males,[40] a factor which is known to impact upon brain and cognitive development.[41,42] Females, eg, typically reach peak gray matter thickness 1–2 years earlier than males.[43] During puberty,

the prefrontal cortex, associated with attention and working memory,[5] goes through notable change.[44,45] Earlier cognitive maturation in females may therefore explain part of the observed age trajectories in figure 1.

Strengths and Limitations

The main strength of this study is the large sample size of more than 500 adolescents from 7 cohorts in 4 countries. With this sample size, we were able to make standardized scores for each year between the ages of 11–19 instead of using age strata of 2 or more years, as has previously been done. Second, using data from multiple independent cohorts decreases the chances of confounding effects from one site. A third strength is the use of regression-predicted means in combination with the unstandardized residuals from the total sample in the development of the standardized scoring tables. Using this method, the distribution of scores for each year from 11 to 19 years of age, stratified by sex, were built on the substantial size of the entire cohort, reducing the influence of extreme scores while upholding the broad range in variation.

The main limitation of this study is the possible site differences between the cohorts. The combined cohort does not represent a particular country or demographic population strata. Instead it represents a multinational control group for the MCCB. In this regard, several factors need consideration: The ABD cohort was the only cohort recruited in an epidemiological setting, whereas the other cohorts were recruited as part of clinical studies, using various convenience methods (see cohort descriptions for more information). Recruitment of healthy controls might thus have been influenced by factors such as housing location, subscription to newspapers, and friends, in which potential sampling biases cannot be ruled out. Other demographic differences in and between sites (eg, socioeconomic status, language, culture, ethnicity, education level, school systems, or teaching methods) were not specifically controlled for in our analyses. It was not possible to fully determine how site differences might have affected variation in cognitive performance since age and other inclusion or exclusion criteria differed between the cohorts. When comparing global cognition between the cohorts we found a significant difference between the EOS and the CANDY cohorts (table 1). However, when investigating homoscedasticity, there were no significant differences in variance for 7 of the 9 tests, indicating that the variability in cognitive performance for most tests was equal between the cohorts. To our knowledge, site differences that would be expected to impact cognitive performance substantially and thereby to cause systematic bias, have not been identified for any of the cohorts. A second limitation is the screening and exclusion of participants with current mental disorders. This might have resulted in an above average healthy group of participants which is not representative of the

adolescent population at large. The average IQ in the combined cohort was 105, slightly above the normative average of 100,[46] which may potentially have caused a bias toward elevated scores. A third limitation is that illicit substance use and alcohol dependence were investigated by interviews without confirmation from urine or hair samples. Consequently, we cannot rule out that drugs might have influenced test performance for some of the participants. A fourth limitation is differences in test administration. Due to the multisite nature of the study, it was not possible to ensure that the tests were administered in the recommended test order[2] or whether additional tests, not part of the MCCB, were included. Such factors may conceivably have caused differences between sites in test performance due to fatigue, interference, or reduced motivation. A fifth limitation is that the cohorts had different number of participants, in which larger cohorts have had greater influence on the results. A sixth caveat is that all participants were recruited from Western countries. It is not known how our findings relate to healthy cognitive performance in non-Western countries.

In sum, the studies contributing with participants to this study were not designed specifically to obtain standardized scores and normative data of the MCCB for adolescents. The scoring tables may be considered as a useful tool for clinicians and the scientific community, but the limitations discussed above should be considered when using these scores on individual patients from diverse backgrounds.

Conclusion

With this study, we have developed an easily accessible and standardized data set for healthy adolescent MCCB performance, which we hope will facilitate the manual scoring of adolescent protocols and help expand the use of the MCCB among children and adolescents in future research.

Supplementary Material

Supplementary data are available at *Schizophrenia Bulletin* online.

Funding

Acknowledgments

References

1. Kern RS, Nuechterlein KH, Green MF, et al. The MATRICS Consensus Cognitive Battery, part 2: co-norming and standardization. *Am J Psychiatry.* 2008;165:214–220.

2. Nuechterlein KH, Green MF, Kern RS, et al. The MATRICS Consensus Cognitive Battery, part 1: test selection, reliability, and validity. *Am J Psychiatry.* 2008;165:203–213.

3. Kelleher I, Clarke MC, Rawdon C, Murphy J, Cannon M. Neurocognition in the extended psychosis phenotype: performance of a community sample of adolescents with psychotic symptoms on the MATRICS neurocognitive battery. *Schizophr Bull.* 2013;39:1018–1026.

4. Holmén A, Juuhl-Langseth M, Thormodsen R, Melle I, Rund BR. Neuropsychological profile in early-onset schizophrenia-spectrum disorders: measured with the MATRICS battery. *Schizophr Bull.* 2010;36:852–859.

5. Blakemore SJ, Choudhury S. Development of the adolescent brain: implications for executive function and social cognition. *J Child Psychol Psychiatry.* 2006;47:296–312.

6. Blakemore SJ. Imaging brain development: the adolescent brain. *Neuroimage.* 2012;61:397–406.

7. Paus T, Keshavan M, Giedd JN. Why do many psychiatric disorders emerge during adolescence? *Nat Rev Neurosci.* 2008;9:947–957.

8. Mohn C, Sundet K, Rund BR. The Norwegian standardization of the MATRICS (Measurement and Treatment Research to Improve Cognition in Schizophrenia) Consensus Cognitive Battery. *J Clin Exp Neuropsychol.* 2012;34:667–677.

9. Nitzburg GC, Derosse P, Burdick KE, Peters BD, Gopin CB, Malhotra AK. MATRICS Cognitive Consensus Battery (MCCB) performance in children, adolescents, and young adults. *Schizophr Res.* 2014;152:223–228.

10. Stone WS, Mesholam-Gately RI, Giuliano AJ, et al. Healthy adolescent performance on the MATRICS Consensus Cognitive Battery (MCCB): developmental data from two samples of volunteers. *Schizophr Res.* 2016;172:106–113.

11. García-pérez MA, Núñez-antón V. Cellwise residual analysis in two-way contingency tables. *Educ Psychol Meas.* 2003;63:825–839.

12. WASI. *Wechsler Abbreviated Scale of Intelligence.* Stockholm, Sweden: Harcourt Assessment, Inc.; 2007.

13. Wilkinson G, Robertson G. *Wide Range Achievement Test.* 4th ed. Lutz, FL: Psychological Assessment Resources, Inc; 2005.

14. Wilkinson G. *The Wide Range Achievement Test–1993 Edition (WRAT3): Administration Manual.* Wilmington, DE: Wide Range; 1993.

15. Kaufman J, Birmaher B, Brent D, et al. Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime version (K-SADS-PL): initial reliability and validity data. *J Am Acad Child Adolesc Psychiatry.* 1997;36:980–988.

16. Sheehan DV, Lecrubier Y, Sheehan KH, et al. 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.* 1998;59 (suppl 20):22–33;quiz 34–57.

17. Spitzer RL, Williams JB, Kroenke K, et al. Utility of a new procedure for diagnosing mental disorders in primary care. The PRIME-MD 1000 study. *JAMA.* 1994;272:1749–1756.

18. First M, Spitzer R, Williams J, Gibbon M. *Structured Clinical Interview for DSM-IV-Non-patient Edition, Version 1.0.* Washington, DC: American Psychiatric Association; 1995.

19. DeRosse P, Ikuta T, Karlsgodt KH, et al. White matter abnormalities associated with subsyndromal psychotic-like symptoms predict later social competence in children and adolescents. *Schizophr Bull.* 2017;43:152–159.

20. Mayer JD, Salovey P, Caruso D. *Mayer–Salovey–Caruso Emotional Intelligence Test (MSCEIT© V2.0).* Toronto, Canada: Multi-Health Systems; 2002.

21. Keefe RS, Goldberg TE, Harvey PD, Gold JM, Poe MP, Coughenour L. The Brief Assessment of Cognition in Schizophrenia: reliability, sensitivity, and comparison with a standard neurocognitive battery. *Schizophr Res.* 2004;68:283–297.

22. Army Individual Test Battery. *Manual of Directions and Scoring.* Washington, DC: War Department; 1944.

23. Spreen O, Strauss E. *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary.* 2nd ed. New York, NY: Oxford University Press; 1998.

24. Cornblatt BA, Risch NJ, Faris G, Friedman D, Erlenmeyer-Kimling L. The Continuous Performance Test, Identical Pairs version (CPT-IP): I. New findings about sustained attention in normal families. *Psychiatry Res.* 1988;26:223–238.

25. Wechsler D. *WMS-III: Wechsler Memory Scale Administration and Scoring Manual.* London, UK: The Psychological Corporation; 1997.

26. Gold JM, Carpenter C, Randolph C, Goldberg TE, Weinberger DR. Auditory working memory and Wisconsin card sorting test performance in schizophrenia. *Arch Gen Psychiatry.* 1997;54:159–165.

27. Brandt J, Benedict RHB. *Hopkins Verbal Learning Test-Revised: Professional Manual.* Lutz, FL: Psychological Assessment Resources; 2001.

28. Benedict RHB. *Brief Visuospatial Memory Test—Revised.* Odessa, FL: Psychological Assessment Resources; 1997.

29. Stern RA, White T. *Neuropsychological Assessment Battery (NAB).* Lutz, FL: Psychological Assessment Resources; 2003.

30. Baillargeon A, Lassonde M, Leclerc S, Ellemberg D. Neuropsychological and neurophysiological assessment of sport concussion in children, adolescents and adults. *Brain Inj.* 2012;26:211–220.

31. Nagle AM, Everhart DE, Durham TW, McCammon SL, Walker M. Deception strategies in children: examination of forced choice recognition and verbal learning and memory techniques. *Arch Clin Neuropsychol.* 2006;21:777–785.

32. Nuechterlein KH, Green MF. *MATRICS Consensus Cognitive Battery, Norwegian Version.* Rund BR and Sundet KS, trans. Los Angeles, CA: MATRICS Assessment, Inc; 2009.

33. Iverson GL, Woodward TS, Iverson AM. Regression-predicted age norms for the children's orientation and amnesia test. *Arch Clin Neuropsychol.* 2002;17:131–142.

34. Cohen J, Cohen P, West SG, Aiken LS. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences.* 3rd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 2003.

35. Fastenau PS. Validity of regression-based norms: an empirical test of the comprehensive norms with older adults. *J Clin Exp Neuropsychol.* 1998;20:906–916.

36. Huber PJ, Ronchetti EM. *Robust Statistics*. 2nd ed. Hoboken, NJ: John Wiley & Sons, Inc; 2009.

37. Cook RD, Weisberg S. *Residuals and Influence in Regression*. London: Chapman & Hall; 1982.

38. Waber DP, De Moor C, Forbes PW, et al.; Brain Development Cooperative Group. The NIH MRI study of normal brain development: performance of a population based sample of healthy children aged 6 to 18 years on a neuropsychological battery. *J Int Neuropsychol Soc.* 2007;13:729–746.

39. Gur RC, Richard J, Calkins ME, et al. Age group and sex differences in performance on a computerized neuro-cognitive battery in children age 8–21. *Neuropsychology*. 2012;26:251–265.

40. Blakemore SJ, Burnett S, Dahl RE. The role of puberty in the developing adolescent brain. *Hum Brain Mapp.* 2010;31:926–933.

41. Sisk CL, Zehr JL. Pubertal hormones organize the adolescent brain and behavior. *Front Neuroendocrinol.* 2005;26:163–174.

42. Goddings AL, Mills KL, Clasen LS, Giedd JN, Viner RM, Blakemore SJ. The influence of puberty on subcortical brain development. *Neuroimage*. 2014;88:242–251.

43. Giedd JN, Clasen LS, Lenroot R, et al. Puberty-related influences on brain development. *Mol Cell Endocrinol.* 2006;254–255:154–162.

44. Huttenlocher PR. Synaptic density in human frontal cortex - developmental changes and effects of aging. *Brain Res.* 1979;163:195–205.

45. Petanjek Z, Judaš M, Šimic G, et al. Extraordinary neoteny of synaptic spines in the human prefrontal cortex. *Proc Natl Acad Sci USA.* 2011;108:13281–13286.

46. PsychCorp. *WASI, Wechsler Abbreviated Scale of Intelligence Manual*. San Antonio, TX: Harcourt Assessment, Inc.; 1999.