Journal of Athletic Training 2007;42(4):515–523
© by the National Athletic Trainers' Association, Inc
www.journalofathletictraining.org

original research

Athlete Characteristics and Outcome Scores for Computerized Neuropsychological Assessment: A Preliminary Analysis

Cathleen N. Brown, PhD, ATC*; Kevin M. Guskiewicz, PhD, ATC, FACSM†; Joseph Bleiberg, PhD‡

*University of Georgia, Athens, GA; †University of North Carolina, Chapel Hill, NC; ‡National Rehabilitation Hospital, Washington, DC

Context: Computerized neuropsychological testing is used in athletics; however, normative data on an athletic population are lacking.

Objective: To investigate factors, such as sex, SAT score, alertness, and sport, and their effects on baseline neuropsychological test scores. A secondary purpose was to begin establishing preliminary reference data for nonsymptomatic collegiate athletes.

Design: Observational study.

Setting: Research laboratory.

Patients or Other Participants: The study population comprised 327 National Collegiate Athletic Association Division I athletes from 12 men's and women's sports.

Main Outcome Measure(s): Athletes were baseline tested before their first competitive season. Athletes completed demographics forms and self-reported history of concussion (1 or no concussion and 2 or more concussions) and SAT scores (<1000, 1000 to 1200, and >1200). The 108 women had a mean age of 18.39 ± 0.09 years, height of 167.94 ± 0.86 cm, and mass of 62.36 ± 1.07 kg. The 219 men had a mean age of 18.49 ± 0.07 years, height of 183.24 ± 1.68 cm, and mass of 88.05 ± 1.82 kg. Sports participation included women's soccer, lacrosse, basketball, and field hockey; men's football, soccer, lacrosse, and wrestling; and women's and men's track and cheerleading. We used the Automated Neuropsychological Assessment Metrics (Army Medical Research and Materiel Command, Ft Detrick, MD) and measured throughput scores (the number of correct responses per minute) as the dependent variable for each subtest, with higher scores reflecting increased speed and accuracy of responses. Subsets included 2 simple reaction time (SRT) tests, math processing (MTH), Sternberg memory search (ST6), matching to sample pairs (MSP), procedural reaction time (PRO), code digit substitution (CDS), and the Stanford sleep scale Likert-type score.

Results: Women scored better than men on the ST6 ($P < .05$), while men scored significantly better than women on the SRT and MSP tests. The highest-scoring SAT group performed better than other SAT groups on selected subtests (SRT, MTH, ST6, MSP, and CDS) ($P < .05$), and athletes tested during their season were more likely to score lower on the alertness scale ($\chi^2_2[n = 322] = 11.32$, $P = .003$). The lowest alertness group performed worse on the MSP and CDS subtests ($P < .05$). No differences were found between the group with a history of 1 or no concussion and the group with a history of 2 or more concussions ($P > .05$).

Conclusions: Performance on computerized neuropsychological tests may be affected by a number of factors, including sex, SAT scores, alertness at the time of testing, and the athlete's sport. To avoid making clinical misinterpretations, clinicians should acknowledge that individual baselines vary over time and should account for this variation.

Key Words: concussion, cognitive testing

Key Points

- Sex, SAT scores, alertness at the time of testing, and sport may influence performance on computerized neuropsychological tests.
- We observed no differences on any Automated Neuropsychological Assessment Metrics subtest scores based on history of concussion.
- Individual baseline testing is needed and scores may vary over time with fatigue, motivation, intensity of training, sex, standardized test scores, and sport.

Mild head injury in athletes participating in recreational and organized sports has evolved into a public health concern in recent years, particularly in collegiate athletes.[1–3] Sosin et al[4] reported in 1996 that approximately 300 000 sport-related concussions occur annually in the United States; however, authors[5,6] of more recent studies on unreported concussions have suggested that the actual value could be more than double that number. The second author (K.M.G.) and colleagues[7] estimated that at least 5% of all collegiate football players annually will sustain a concussion. In addition, the incidence of concussive injury in soccer, lacrosse, and hockey is rising.[8] Concussion has been defined as "an injury resulting from a blow to the head that caused an alteration in mental status and 1 or more of the following symptoms: headache, nausea, vomiting, dizziness/balance problems, fatigue, difficulty sleeping, drowsiness, sensitivity

to light or noise, blurred vision, memory difficulty, and difficulty concentrating."[9]

In an attempt to validate more-objective tests for assessing concussion and making safer return-to-play decisions, researchers[1,9–20] have conducted several studies using neuropsychological (NP) testing. Sports medicine personnel have used NP testing to identify cognitive deficits and track recovery after concussion.[10] Computerized NP testing has been advocated over the traditional pencil-and-paper versions because it offers more precise timing measures and provides instant scoring and multiple forms.[17,21] Computerized NP testing is more efficient in a sports medicine setting because the practitioner does not need to be present constantly,[22] and large numbers of athletes may be tested at one time for increased efficiency.[16,17,23] Ideally, a baseline battery assessing many neurocognitive functions would be performed before injury, but this model is not always possible.[10,16,23] Time and monetary constraints may prevent sports medicine personnel from baseline testing or may enable only the testing of athletes participating in a high-risk sport, such as football. Even if athletes are baseline tested before their competitive seasons, history of concussion may affect results.[11]

Normative data for collegiate athletes are not available currently for most of the computerized NP test batteries. One of the many computerized NP test batteries[24–28] is the Automated Neuropsycholgical Assessment Metrics (ANAM; Army Medical Research and Materiel Command, Ft Detrick, MD), which is not yet commercially available. It consists of a battery of subtests that assess standard NP constructs, such as processing speed, short-term memory, working memory, and resistance to interference.[22,29] Bleiberg et al[15] used ANAM to detect concussions in a military population. While ANAM data are available on large military samples,[15,29] limited civilian collegiate data are available.[30] As mentioned, access to individual baseline data is not always feasible; therefore, access to normative data may help the clinician.[23,31] Additionally, previous ANAM studies[29] did not encompass sex, ethnicity, or education, and researchers who studied athletes focused on football athletes and did not provide information on female athletes. Little information is available for differences based on age, sex, or learning disorder.[11,32] Before clinically interpreting the results of ANAM or other computerized NP tests, the clinician must document reference data and demographic factors to distinguish between decreased NP function and expected performance.[29] No one has made recommendations about the best time of year (ie, while the athlete is in season or out of season) to perform baseline testing or about the extent to which baseline scores vary over time. Therefore, the purpose of our study was to investigate the effects of factors, such as sex, SAT score, alertness, and sport, on baseline ANAM scores. Our secondary purpose was to establish preliminary reference data for nonsymptomatic collegiate athletes in various sports.

METHODS

Subjects

Our National Collegiate Athletic Association Division I institution employs a protocol that calls for conducting a baseline computerized NP test for all incoming freshmen and transfer athletes in the following sports: women's soccer, lacrosse, basketball, and field hockey; men's football, soccer, lacrosse, and wrestling; and men's and women's track (high jumpers and pole vaulters) and cheerleading. We collected data each July through November from 2001 through 2003 on incoming athletes. Of those tested during this period, 327 (219 male, 108 female) healthy collegiate athletes agreed to participate in our study. Subjects signed informed consent forms. The university's institutional review board approved the study.

When they were tested, some athletes, such as those participating in the 4 fall sports, were involved in heavy preseason training with 2 practices per day. These athletes were considered in-season athletes. Other athletes participating in the 8 winter and spring sports were not practicing as a team at the time of assessment. These athletes were considered out-of-season athletes. Subjects were tested individually in quiet rooms with minimal distractions. They were asked to complete demographics forms (sex, age, height, mass, sport, and position played) and to self-report history of concussion (number of concussions sustained in the 7 years before enrollment) and SAT scores.

The women had a mean age of 18.39 ± 0.09 years (range, 18–22 years), height of 167.94 ± 0.86 cm (range, 150–196 cm), and mass of 62.36 ± 1.07 kg (range, 43–114 kg). The men had a mean age of 18.49 ± 0.07 years (range, 18–24 years), height of 183.24 ± 1.68 cm (range, 152–198 cm), and mass of 88.05 ± 1.82 kg (range, 59–150 kg). In the women's sports, 23 participants were soccer athletes; 26, lacrosse; 14, basketball; 23, field hockey; 7, track; and 15, cheerleading. In the men's sports, 84 participants were football athletes; 33, soccer; 45, lacrosse; 37, wrestling; 10, track; and 10, cheerleading. Position played was recorded for only football players but was not analyzed. Race and ethnicity were not recorded.

We stratified subjects into 2 groups based on history of concussion: limited to low (1 or no concussion) and significant (2 or more concussions).[9,11] Of the 327 subjects, 17 (5%) reported 2 or more previous concussions. The remaining 310 subjects (95%) reported 1 or no concussion. Of the 327 athletes participating, 246 self-reported SAT scores (approximately 75%). The lowest-scoring group (<1000) represented 20% of those reporting scores (41 men, 8 women); the middle-scoring group (1000 to 1200), 44% (68 men, 41 women); and the highest-scoring group (>1200), 36% (61 men, 27 women).

Instruments

The ANAM consists of a series of subtests evaluating different neurocognitive functions. Among others, researchers at the National Rehabilitation Hospital in Washington, DC, have developed it over the last 20 years as a computerized NP test battery for military personnel.[22] Investigators[33,34] have shown that the ANAM is valid and reliable. The version of ANAM used in our study consisted of the following subtests: simple reaction time 1 (SRT1), math processing (MTH), Sternberg memory search (ST6), matching to sample pairs (MSP), procedural reaction time (PRO), code digit substitution (CDS), simple reaction time 2 (SRT2), and the Stanford sleep score (SLP).[35] The SRT subtest was presented twice, with the second session at the end of the battery to test reaction time after some cognitive fatigue was present. In previous research, we and our colleagues[36] indicated that the second SRT test was more sensitive to the effects of concussion after cognitive fatigue. Instructions for each ANAM subtest appeared on a personal computer screen; stimuli also appeared on the screen. Data were collected, processed, and stored on a personal computer as the ANAM battery was completed. Throughput scores

combining speed and accuracy were calculated and expressed in milliseconds.[29] Each subtest required a 1-button or 2-button mouse response. To decrease anticipatory responses, a variable interstimulus interval (ISI) was used for individual stimuli on each subtest.[29] The ISI is the amount of time between successive stimuli in a single subtest. When the ISI varies, participants are less able to anticipate when a stimulus will be presented. We report the ISI with the description of each subtest.

The SRT was a traditional reaction-time test that measured response time to an asterisk-like symbol stimulus appearing in the center of the computer screen. Subjects responded to the stimuli by clicking the left button on the mouse. A total of 25 trials were collected. The maximal response time was 9000 milliseconds, and the ISI ranged from 500 to 1800 milliseconds.[29]

In the MTH subtest, a 2-step arithmetic problem involving addition or subtraction, or both, was presented. Each subject had to solve the problem, then click the right button on the mouse if the answer was more than 5 or the left button if the answer was less than 5. For the 20 trials, the probability of a right click was greater than the probability of a left click. Up to 15 000 milliseconds were allowed for a response, with the ISI ranging from 950 to 1200 milliseconds.[29]

The ST6 subtest required the subjects to memorize a set of 6 capital letters presented on the screen. The letters disappeared after subjects indicated that they had memorized the letters. A series of individual letters appeared on the screen, and the subjects clicked the right button of the mouse if the appearing letters were in the set of the original 6 letters or the left button if they were not. Thirty trials were collected. The maximal response time was 9000 milliseconds. The ISI ranged from 950 to 1200 milliseconds. Left-button and right-button responses were represented equally.[29]

For the MSP scores, a checkerboard matrix was presented for 3 seconds while the subject memorized it. Next, 2 matrices were presented side by side after a 5-second delay, and the subject clicked the right or left mouse button to indicate which matrix was the initial matrix presented. The chance of a right or left display being the correct match was 50%. Fifteen trials were performed. The upper-limit response time was 8600 milliseconds. The ISI ranged from 750 to 1200 milliseconds.[29]

The PRO subtest measured reaction time with 20 items. The number 2, 3, 4, or 5 was presented on the screen as the stimulus, and the participant pressed the left mouse button if it was a 2 or a 3 and the right mouse button if it was a 4 or a 5. The stimulus was presented for 1000 milliseconds, and the ISI ranged from 650 to 1000 milliseconds. Up to 1400 milliseconds was allowed for response.[37]

In the CDS subtest, a key with 9 different symbols and 9 paired digits was presented in the upper portion of the screen. A single digit-symbol pairing appeared beneath the key, and subjects clicked the left button on the mouse if the single pairing correctly represented the key or the right button if it did not. Thirty-six trials were presented. The maximal response time was 9500 milliseconds, and the ISI ranged from 850 to 950 milliseconds.[29]

The SLP subtest presented a 7-point Likert-type scale with verbal descriptions of how alert or fatigued the participant was at the time of the test. Each participant entered the digit that best corresponded to his or her level of alertness. A 1 indicated *feeling very alert, wide awake, and energetic*; 2 indicated *able to concentrate but not quite at peak*; 3 indicated *relaxed and awake but not fully alert*; 4 represented *a little tired and having difficulty concentrating*; 5 represented *feeling tired and struggling to concentrate*; 6 indicated *sleepy and want to lie down*; and 7 indicated *very sleepy and cannot stay awake much longer*. Thus, lower scores indicated greater levels of alertness.[35]

Throughput scores (ie, the number of correct responses per minute) were recorded for each subtest and represented a combination of speed and accuracy.[15]

Procedures

Standardized ANAM instructions were provided for each subject by experienced examiners, emphasizing the importance of speed and accuracy in completing the examination. Instructions for each subtest and a short practice test were presented on the screen before administration of the scored subtest. These scores were retrieved using NRHReview (version 2001; National Rehabilitation Hospital, Washington, DC) for each subtest.

Statistical Analyses

Before analysis, the data set of subtest throughput scores was cleaned for statistical outliers or scores that were more than 3 SDs from the mean calculated from all subjects. This cleaning resulted in the removal of 51 univariate outlying scores, or approximately 2% of the data set. These scores were removed because it became obvious during the data-cleaning process that a very small group of athletes failed to follow the test instructions, resulting in accuracy scores well below an acceptable range. These subjects may not have understood the instructions, so they either scored no better than chance or scored very poorly because they reversed their responses. The validity of these scores was then questionable. No multivariate outliers were identified using the Mahalanobis distance, so we removed only univariate outliers. Reeves et al[29] demonstrated less than 4% of a healthy sample scored more than 2 SDs from the mean on any subtest (from 2 to 84 participants out of 2371). We believed that removing 2% of the data was an acceptable step.

Independent-samples t tests with an a priori α level set at .05 were used to compare subtest scores between the following groups: males and females, history of 1 or no concussion (limited to low) and history of 2 or more concussions (significant), and in-season and out-of-season athletes. We chose the 2 groupings for history of concussion based on evidence[9,11] that a history of more than 2 concussions has been demonstrated to result in long-term cognitive and symptomatic problems. We made no attempt to control for severity of concussion or time of injury within the 7 years before enrollment. One-way analysis of variance (ANOVA) was used to compare subtest throughput scores among 3 SAT-score groups ($<$1000, 1000 to 1200, and $>$1200), among 3 self-reported alertness-level groups (1 to 2 indicated high alertness, 3 to 4 indicated moderate alertness, and 5 to 7 indicated low alertness), and among sports. All statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS version 11.0; SPSS Inc, Chicago, IL). Post hoc analyses were conducted with the Tukey test. We also used univariate analysis of covariance to retrospectively analyze SAT scores. In addition, we performed secondary data analysis using the χ^2 test to check observed versus expected counts in the in-season and out-of-season groups across the 3 levels of alertness groups.

Table 1. Throughput Mean (SD) Scores on Automated Neuropsychological Assessment Metrics Subtests* by Sex

Subtest	Men (n = 219)	Women (n = 108)	*t* (df)	*P*	Effect Size	95% Confidence Interval Lower	95% Confidence Interval Upper
Simple reaction time test 1	224.7 (26.8)	212.5 (28.3)	3.74 (311)	.001	0.43	5.8	18.6
Math processing	21.5 (5.9)	21.4 (6.1)	0.04 (317)	.968	0.01	−1.4	1.4
Sternberg memory search	76.0 (15.3)	80.5 (14.5)	−2.54 (322)	.012	0.30	1.0	8.0
Matching samples pairs	40.4 (11.9)	37.5 (9.3)	1.03 (314)	.028	0.25	0.5	5.3
Procedural reaction time	85.5 (16.9)	83.6 (17.0)	0.91 (318)	.361	0.11	−2.1	5.8
Code digit substitution	49.8 (8.3)	48.8 (8.2)	1.03 (314)	.304	0.04	−0.9	3.0
Simple reaction time test 2	225.8 (26.1)	215.6 (30.2)	3.10 (308)	.002	0.34	3.7	16.7
Stanford sleep scale	3.0 (1.3)	3.1 (1.3)	−0.31 (320)	.760	0.04	−0.4	0.3

*Power was more than .80 on all subtests except math processing, procedural reaction time, code digit substitution, and Stanford sleep scale.

Table 2. The 95% Confidence Intervals and Effect Sizes for History of Previous Concussion

Subtest	95% Confidence Intervals for Group Difference Lower	95% Confidence Intervals for Group Difference Upper	Effect Size
Simple reaction time test 1	−15.6	12.6	0.05
Math processing	−5.3	0.6	0.33
Sternberg memory search	−4.3	10.6	0.18
Matching samples pairs	−10.0	1.2	0.29
Procedural reaction time	−6.3	10.4	0.12
Code digit substitution	−3.4	5.0	0.10
Simple reaction time test 2	−15.1	12.3	0.05
Stanford sleep scale	−0.7	0.6	0.03

Finally, we determined the 95% confidence intervals (CIs) for the differences between sexes and between concussion groups. The CIs were determined using the estimated SE produced by SPSS.[38] Based on the width of the intervals, they described the uncertainty and precision of the results.[39,40] They can provide plausible values for the mean differences between groups on the variable being estimated.[39]

Based on results from the sport-comparison ANOVA, a retrospective analysis comparing SAT scores between sports was performed using a 1-way ANOVA and Tukey HSD post hoc testing with the α set at .05.

RESULTS

Throughput scores for each sex and the effect sizes and 95% CIs for the differences between sexes are reported in Table 1. Independent-samples *t* tests revealed significant differences between sexes, with men scoring higher than women on SRT1, SRT2, and MSP. Women scored significantly higher than men on ST6. No other subtest scores revealed differences between sexes (Table 1).

Independent-samples *t* tests demonstrated no difference between any ANAM subtest scores based on number of previous concussions ($t_{308–322} = -1.558–0.833$, $P = .120–.912$). We report the ranges of the results of these tests because group sizes were unequal and because the data could be misused as cut-off scores between injured and uninjured athletes. We are continuing to investigate and hope to publish these data once the group sizes are more equal. The 95% CIs and effect sizes are presented in Table 2.

Independent-samples *t* tests revealed that out-of-season athletes scored significantly better than the in-season athletes on all subtests except ST6 and PRO, for which scores were not significantly different, and SLP (Table 3). The out-of-season group scored significantly lower than the in-season athletes on the SLP, indicating lower levels of fatigue and increased alertness in the former group. To ensure that the large number of football players in our sample was not skewing our results, we randomly selected 27 football players from the sample (the average number of athletes from every other sport) and conducted the same analysis with the other subjects for in season or out of season. The independent-samples *t* test revealed similar results, suggesting that the number of football players did not influence the results.

The results revealed significant differences between the SAT groups on most of the tests (Table 4). The highest-scoring SAT group performed better on most of the ANAM subtests, with the exception of the SRT1, PRO, and SLP (Table 4). This group performed significantly higher than the middle-scoring SAT group on MTH ($P < .001$), ST6 ($P < .001$), MSP ($P =$

Table 3. Throughput Mean (SD) Scores on Automated Neuropsychological Assessment Metrics Subtests by Season

Subtest	In Season (n = 163)	Out of Season (n = 164)	*t* (df)	*P*	Effect Size	95% Confidence Interval Lower	95% Confidence Interval Upper
Simple reaction time test 1	215.1 (30.0)	226.0 (24.5)	−3.51 (311)	.001	0.37	4.8	17.1
Math processing	20.7 (5.7)	22.2 (6.2)	−2.29 (317)	.023	0.25	0.2	2.8
Sternberg memory search	76.1 (15.7)	78.8 (14.5)	−1.63 (322)	.104	0.18	−0.6	6.0
Matching samples pairs	37.2 (10.6)	41.6 (11.3)	−3.55 (314)	.001	0.41	2.0	6.8
Procedural reaction time	83.0 (16.6)	86.6 (17.1)	−1.90 (318)	.059	0.21	−0.1	7.3
Code digit substitution	48.5 (8.6)	50.4 (7.9)	−1.97 (314)	.050	0.21	0.0	3.7
Simple reaction time test 2	217.2 (28.6)	227.2 (26.4)	−3.19 (308)	.002	0.35	3.8	16.1
Stanford sleep scale	3.2 (1.4)	2.8 (1.3)	2.77 (320)	.006	0.28	0.1	0.7

Table 4. Throughput Mean (SD) Scores on Automated Neuropsychological Assessment Metrics Subtests by SAT Score

Subtest	<1000 (n = 49)	1000–1200 (n = 109)	>1200 (n = 88)	F	*P*
Simple reaction time test 1	218.43 (30.76)	220.63 (27.47)	226.64 (23.11)	1.81	.20
Math processing	19.14 (4.74)	20.77 (5.81)	24.14 (5.59)	14.92	<.001
Sternberg memory search	71.88 (15.55)	75.00 (13.58)	83.88 (14.63)	13.92	<.001
Matching samples pairs	32.90 (8.15)	40.02 (10.85)	44.15 (11.4)	16.96	<.001
Procedural reaction time	81.62 (17.04)	85.72 (17.14)	87.68 (16.13)	2.00	.14
Code digit substitution	46.32 (8.27)	49.15 (7.83)	52.99 (7.55)	11.86	<.001
Simple reaction time test 2	215.64 (25.52)	219.7 (26.64)	228.46 (28.18)	4.10	.02
Stanford sleep scale	3.52 (1.68)	2.97 (1.14)	2.8 (1.09)	5.26	.01

Table 5. Throughput Mean (SD) Scores on Automated Neuropsychological Assessment Metrics Subtests by Alertness Level

Subtest	High Alertness (n = 133)	Moderate Alertness (n = 144)	Low Alertness (n = 45)	F	*P*
Simple reaction time test 1	223.92 (26.14)	219.76 (28.28)	214.65 (31.01)	1.95	.144
Math processing	21.29 (6.19)	21.90 (5.91)	20.65 (5.63)	0.83	.435
Sternberg memory search	79.03 (15.16)	77.37 (14.64)	72.76 (16.22)	2.86	.059
Matching samples pairs	40.12 (11.15)	40.55 (11.13)	34.70 (9.99)	4.95	.008
Procedural reaction time	85.28 (17.19)	85.66 (16.45)	81.22 (18.07)	1.18	.309
Code digit substitution	50.98 (8.20)	49.22 (8.10)	45.91 (7.97)	6.49	.002
Simple reaction time test 2	225.31 (27.66)	221.98 (27.55)	214.76 (28.96)	2.26	.106

.02), and CDS ($P = .003$) and significantly higher than the lowest-scoring SAT group on MTH ($P < .001$), ST6 ($P < .001$), MSP ($P < .001$), CDS ($P < .001$), and SRT2 ($P = .029$). The middle-scoring SAT group scored significantly higher than the lowest-scoring SAT group on MSP ($P < .001$). On SLP, the lowest-scoring SAT group scored significantly higher than the middle-scoring ($P = .031$) and the highest-scoring ($P = .004$) SAT groups, indicating increased fatigue.

At the end of the ANAM battery, athletes self-reported their level of alertness using a 7-point Likert scale. The high-alertness group included 41% (n = 133) of the sample; the moderate-alertness group, 45% (n = 144); and the low-alertness group, 14% (n = 45). No subjects reported a score of 7, and 5 subjects did not report a score. The results revealed significant differences among alertness-level groups on the MSP and CDS subtests (Table 5). Post hoc Tukey tests revealed that the group reporting low alertness scored significantly lower than the high-alertness group on MSP ($P = .015$) and CDS ($P = .001$) and scored significantly lower than the moderate-alertness group on the MSP ($P = .007$) and CDS ($P = .052$) (Table 5). The omnibus F was not significant on the ST6 subtest, but the post hoc Tukey test revealed that the low-alertness group scored lower than the high-alertness group ($P = .046$). No other differences were observed in the subtests.

We checked observed versus expected counts in the in-season and out-of-season groups across the 3 alertness-level groups. Five of the 327 athletes were missing some scores and were excluded from this analysis. For the 322 remaining athletes, we found that significantly more in-season than out-of-season athletes were in the low-alertness group and that more out-of-season than in-season athletes were in the high-alertness group ($\chi^2_2 = 11.32$, $P = .003$) (Table 6). A relationship existed between involvement in heavy in-season training and reporting lower levels of alertness.

The 1-way ANOVAs revealed significant differences among the 10 different sports on all subtest scores except SLP ($P = .146$) (Table 7). Post hoc Tukey testing revealed numerous differences among sports, with athletes participating in football and women participating in soccer tending to score lower on most subtests and with men participating in lacrosse and wrestling tending to score higher. For SRT1, women participating in soccer scored lower than athletes participating in men's lacrosse, wrestling, or cheerleading. For MTH and CDS, wrestlers scored higher than football athletes. For ST6, field hockey athletes scored higher than football athletes. For MSP, wrestlers scored higher than football, field hockey, and women's basketball athletes. For SRT2, men's lacrosse athletes scored higher than women's soccer, women's lacrosse, and football athletes. While the PRO F value was statistically significant, post hoc Tukey tests revealed no significant differences ($P = .098–.99$). We report the range of the results of these tests because several tests were performed.

Table 6. Frequency of Stanford Sleep Scale Score Reports When Categorized by Season

Season	Sleep Scale Scores: High Alertness (n = 133)	Moderate Alertness (n = 144)	Low Alertness (n = 45)
In season (%)	56 (42.1)	72 (50.0)	32 (71.1)
Out of season (%)	77 (57.9)	72 (50.0)	13 (28.9)

A retrospective analysis comparing SAT scores between sports revealed numerous significant differences ($F_{9,236} = 7.88$, $P < .01$). Tukey post hoc tests revealed that football players reported significantly lower scores than wrestling, cheerleading, field hockey, men's and women's soccer, and men's and women's lacrosse athletes. Women's basketball athletes reported significantly lower SAT scores than wrestling and cheerleading athletes ($P < .01–.05$). These differences between sports were slightly more than 150 points, or approximately two-thirds of the SD.[41]

Table 7. Sport and 1-Way Analysis of Variance Results of Mean Throughput Automated Neuropsychological Assessment Metrics Scores by Subtest

Test	Sport											1-Way Analysis of Variance	
	Women's Soccer (23)	Women's Lacrosse (26)	Women's Basketball (14)	Women's Field Hockey (23)	Men's Football (84)	Men's Soccer (33)	Men's Lacrosse (45)	Men's Wrestling (37)	Track (17)	Cheerleading (25)		F	P
Simple reaction time test 1	201.36 ± 23.91	209.26 ± 27.35	226.68 ± 21.72	213.27 ± 33.80	216.82 ± 28.98	223.66 ± 24.42	227.67 ± 18.76	235.67 ± 19.38	226.61 ± 22.56	227.17 ± 30.50		4.28	<.001
Math processing	21.04 ± 7.16	21.33 ± 6.16	20.57 ± 6.33	21.84 ± 6.26	20.05 ± 5.88	22.19 ± 3.81	22.81 ± 5.68	24.17 ± 7.20	20.31 ± 4.00	24.25 ± 6.30		2.10	.030
Sternberg memory search	78.37 ± 9.29	79.01 ± 13.52	78.99 ± 16.25	83.53 ± 16.40	73.47 ± 15.70	80.28 ± 13.08	75.74 ± 15.78	80.37 ± 15.34	83.60 ± 13.55	81.71 ± 14.36		2.23	.020
Matching samples pairs	40.79 ± 8.29	39.41 ± 10.27	33.88 ± 6.41	35.02 ± 7.26	37.79 ± 10.23	40.63 ± 12.48	44.91 ± 11.45	46.33 ± 12.35	45.23 ± 7.70	39.69 ± 12.45		3.89	<.001
Procedural reaction time	86.69 ± 20.71	80.85 ± 13.95	86.75 ± 23.65	82.23 ± 14.67	81.36 ± 16.91	88.22 ± 14.46	82.64 ± 14.26	91.83 ± 16.94	88.47 ± 18.67	90.86 ± 14.81		1.94	.046
Code digit substitution	48.71 ± 5.42	47.91 ± 9.10	49.18 ± 8.67	49.04 ± 8.77	47.67 ± 8.77	52.13 ± 7.60	49.25 ± 5.29	53.53 ± 8.15	47.57 ± 6.89	53.81 ± 8.56		2.56	.008
Simple reaction time test 2	215.54 ± 30.75	207.58 ± 25.33	222.76 ± 16.66	216.31 ± 31.97	215.58 ± 24.44	224.35 ± 25.86	233.09 ± 20.34	235.10 ± 17.29	229.67 ± 24.61	232.70 ± 36.01		4.57	<.001
Stanford sleep scale	3.26 ± 1.28	2.83 ± 1.40	2.92 ± 1.32	3.35 ± 1.23	3.26 ± 1.49	3.11 ± 1.21	3.03 ± 1.33	2.71 ± 1.06	3.00 ± 0.85	2.33 ± 0.91		1.50	.146

DISCUSSION

The purpose of our research was to investigate the potential effects of sex, SAT score, alertness, and sport on baseline scores for nonsymptomatic collegiate athletes on the ANAM computerized NP test. Our most important finding was that ANAM subtest scores were different between sexes, among SAT score groups, among level-of-alertness groups, and among sports. We found that men had significantly higher throughput scores than women had on the SRT1, MSP, and SRT2. Women scored significantly higher than men on the ST6. Because throughput scores were tested, both speed and accuracy played roles. In a healthy military population, men demonstrated greater speed than women on MSP.[29] However, no other reaction-time differences on the other subtests were observed. In the same study, women tended to display greater accuracy than men on the ST6. In most subtests, women tended to be more accurate but slower than men, although the results were not statistically significant.[29] Historically, men have tended to display faster motor speed[42] and greater aptitude for manipulating spatial relationships.[43] Women, alternatively, have scored better on tests of verbal fluency and verbal memory.[43] Our results support these findings. Both sexes in the healthy military population had similar throughput scores, but a trend existed for men to have higher throughput scores on the MSP than women had.[29] Again, our results support the findings of these other researchers.

SAT Scoring Differences

Groupings based on self-reported SAT scores demonstrated significant differences on subtest performance. The highest-scoring SAT group outperformed the other 2 groups on all subtests except SRT1 and PRO. This group also reported being more alert during the test. Our results suggest some association between SAT scores and NP test results. Previous ANAM normative data from military subjects indicated education levels affected performance only on the MTH subtest, with higher levels of education correlating with faster and more accurate responses.[29] Little to no effect of education was noted on the other subtests.[29] All of our subjects had a high school degree and had been accepted to a major university; however, a large variability still existed in the SAT scores that subjects reported. Our preliminary results indicate that even subjects with similar levels of education can vary subtly in performance on multiple subtests. We could not control motivational levels, and this may have accounted for the differences in scores observed among different SAT groups.

Level of Alertness

Our results indicate that a connection exists between level of alertness and performance on select ANAM subtests; the low-alertness group scored significantly lower than the high-alertness group and the moderate-alertness group on MSP and CDS (Table 5). Authors of previous studies indicated poorer performance on NP tests when subjects were fatigued[44,45] and dehydrated.[46] In shift workers, NP testing demonstrated fatigue can negatively affect performance on visual memory[44] and speed of response on a vigilance task.[45] Response speeds also may be impaired by fatigue and dehydration,[46] which supports our findings regarding alertness.

Compared with out-of-season athletes, in-season athletes scored significantly lower on most subtests. In-season athletes

were those who were participating in heavy team-training sessions, and out-of-season athletes were those participating in limited off-season training. Compared with out-of-season athletes, in-season athletes also reported significantly more fatigue, suggesting that the systemic fatigue of heavy in-season training affected ANAM scores in healthy athletes without concussion. Our data included a sample of football players that was disproportionately larger than samples of athletes in other sports; however, when the sample size was randomly adjusted to match the sample size of athletes in other sports, similar differences were observed, and more in-season athletes reported lower levels of alertness.

The baseline scores of athletes tend to be dynamic over time and may change with fatigue, motivation, and other factors that we discussed. Based on our findings, the timing of the baseline test may be important. The performance of athletes may depend on many factors, including the time of day, whether they are in or out of season, and the intensity of their training. The level of alertness and fatigue should be recognized during baseline and postinjury testing. As discussed, we could not control motivation, which also may influence the result of NP testing. These factors can make interpretation of postinjury data more challenging. More importantly, clinicians must be aware that these factors can further complicate the interpretation of change scores when comparing preinjury with postinjury scores in athletes.

Sport Participation Differences

We observed differences among sports (Table 7) on all ANAM subtests except PRO and SLP. Overall, athletes participating in wrestling and men's lacrosse tended to score higher than athletes participating in women's soccer, lacrosse, and basketball. Football athletes demonstrated significantly lower throughput scores for MTH and MSP compared with wrestling athletes, for SRT2 compared with men's lacrosse athletes, and for ST6 compared with field hockey athletes. Few authors have reported the differences in NP test results based on sport played. In a study of adolescents, Barr[32] found no differences in traditional NP test scores between sports; however, the sample size was much smaller than ours, and the researcher only tested football athletes and field hockey and soccer athletes of both sexes. Obtaining individual baseline data is considered the best method for tracking recovery, but, in the absence of baseline scores, comparison with age-normative data is considered acceptable.[23,31] We found numerous significant differences among sports on many subtests except PRO and SLP, but many potential variables could influence the results.

A retrospective analysis of SAT scores revealed differences in scoring among sports. Football and women's basketball athletes reported significantly lower scores than athletes in a number of other sports. The initial difference in SAT scores among the sports could have accounted for some or all of the differences that we observed in throughput scores on the ANAM subtests. However, the clinician must note that, although women's soccer athletes reported high SAT scores, they demonstrated lower throughput scores than athletes in other sports with similar SAT scores. Further retrospective analyses using univariate analysis of covariance indicated that SAT was an influential covariate in several of the subtests (MTH, ST6, MSP, and CDS) but not in others (SRT1, SRT2, PRT, and SLP). We emphasize that, even with our data set of select, Division I athletes, we observed differences in performance on subtests among sports. This observation reinforces the rationale for individual baselines.

History of Previous Concussion Differences

We observed no differences on any ANAM subtest scores based on history of concussion. Absence of a significant difference in performance between 2 concussion groups is not surprising given the small sample size of the groups. In addition, our group of athletes who had sustained 2 or more concussions was very small compared with athletes who had sustained 1 or no concussion, and our findings can only be considered preliminary results. We speculate that, in fact, differences may exist if the existing data allowed for further stratification (eg, 0, 1 to 2, and 3 or more concussions).

Increased awareness of sport-related concussion is making diagnoses more common; however, some athletes in our study may have sustained concussions that were not diagnosed. Athletes also may have underreported previous concussion for fear of restrictions on their sport participation. Whether a history of concussion affects baseline NP test scores is a point of contention. Barr[32] found no difference in adolescent athletes' scores on traditional NP tests based on previous concussion. However, in their study of Division I athletes, Collins et al[11] found that those athletes reporting 2 or more concussions scored significantly worse on some traditional NP tests. Our sample size is comparable to that of Collins et al[11]; however, the average SAT score was 952 in their study population and 1143 in our study population, which may be a more select group. The potential interactive effects of previous concussion, learning disorder, and intelligence make comparisons difficult.

Limitations

Although our study included a large sample size, it had limitations. Data from female athletes were included in our results, but the data still were dominated by male athletes, particularly football players. The self-reporting of SAT scores was also a limitation. Approximately 25% of the overall sample did not report SAT scores. Higher-scoring subjects may have been more likely to report their scores, potentially skewing the average. Our sample was a homogenous group because they all graduated high school and were admitted to a selective Division I institution. They were potentially a more academically and socioeconomically select group. Confounding factors, including reluctance to report SAT scores and history of concussion, also may have existed. A difference in team SAT scores may have accounted for or confounded the differences that we observed in performance between sports. Fatigue and motivation also may have been confounding factors; these factors were hard to control in preseason baseline neurocognitive testing. In our testing protocol, athletes in 4 sports were baseline tested during heavy preseason training, while athletes in the 8 other sports were not yet practicing as teams. The in-season athletes reported significantly more fatigue than the out-of-season athletes, possibly confounding our results. Other limitations included a small sample size of subjects with 2 or more previous concussions.

CONCLUSIONS

The differences that we observed between sexes, among SAT-score groups, among level-of-alertness groups, and

among sports may be important in the interpretation of baseline and postinjury tests. The differences that we observed support the need for individualized baseline testing and serve as a caution to using normative data without accounting for any demographic factors.

Our study is a first attempt to establish ANAM reference scores. Our results were comparable to findings of baseline scores of military cadets.[15,29] However, minimal information is available for comparison with previous studies of female subjects.[32] Most previous NP test norms included only men, and these men typically were enrolled in college.[11,15,32] We believe our study is unique because it is the first to report nonmilitary collegiate baseline scores for both men and women. The sample size was large, and we had some interesting preliminary findings on associations between NP tests and intelligence. Future researchers should continue to explore the effect of heavy preseason training on NP test scores and to analyze associations between performance and intelligence and between performance and learning disorders. Other authors have not explored the potential effects of fatigue and alertness that we observed. Larger samples are needed, specifically samples involving athletes with 2 or more previous concussions. Analyses should include severity of concussions and a more standardized method of reporting.

CLINICAL RELEVANCE

Despite its preliminary nature, some recommendations can be made for using computerized NP testing in collegiate athletes. Establishing individual baselines is still considered the best model for identifying NP deficits after concussion. Using ANAM, we began establishing reference data for healthy male and female collegiate athletes. The cognitive domains tested by ANAM include attention and concentration, memory, information processing speed, and reaction time,[37] which are used alone or in combination in other computerized NP batteries.[27,28] Although we used only 1 computerized NP test format (ie, ANAM), we believe that these results may apply to other NP tests, which assess the same cognitive domains. Based on our findings, we recommend that future normative scores used for individuals without baseline data should include sex, sport, level of alertness, and possibly SAT scores or other standardized test scores. Baseline scores of athletes may be dynamic over time and may change with fatigue, motivation, intensity of training, and other factors that we discussed.

Computerized NP testing is advantageous because of its precision, multiple forms, and efficiency.[16,21] However, NP test scores are only 1 component of determining recovery from a sport-related concussion. Additional information regarding symptoms and balance deficits must be included in making the assessment and any return-to-play decisions. An individual baseline is still the best criterion to judge, but the reference data may be used with caution. Future studies need to be performed with larger and more diverse student-athletes and on other validated computerized NP test batteries.

ACKNOWLEDGMENTS

We thank Scott Ross, PhD, ATC, and Andrew Notebaert, MS, ATC, for their help with data collection. This study was funded by a grant to Kevin M. Guskiewicz, PhD, ATC, FACSM, from the Centers for Disease Control and Prevention, Atlanta, GA.

REFERENCES

1. McCrea M, Guskiewicz KM, Marshall SW, et al. Acute effects and recovery time following concussion in collegiate football players: the NCAA concussion study. *JAMA.* 2003;290:2556–2563.
2. Proctor MR, Cantu RC. Head and neck injuries in young athletes. *Clin Sports Med.* 2000;19:693–715.
3. Thurman DJ, Alverson C, Dunn KA, Guerrero J, Sniezek JE. Traumatic brain injury in the United States: a public health perspective. *J Head Trauma Rehabil.* 1999;14:602–615.
4. Sosin DM, Sniezek JE, Thurman DJ. Incidence of mild and moderate brain injury in the United States, 1991. *Brain Inj.* 1996;10:47–54.
5. McCrea M, Hammeke T, Olsen G, Leo P, Guskiewicz K. Unreported concussion in high school football players: implications for prevention. *Clin J Sport Med.* 2004;14:13–17.
6. Sefton JM, Pirog K, Capitao A, Harackiewicz D, Cordova ML. An examination of factors that influence knowledge and reporting of mild brain injuries in collegiate football. *J Athl Train.* 2004;39:S52–S53.
7. Guskiewicz KM, Weaver NL, Padua DA, Garrett WE Jr. Epidemiology of concussion in collegiate and high school football players. *Am J Sports Med.* 2000;28:643–650.
8. Injury Surveillance System: Sport Specific Injury Data 2003–04. Available at: http://www1.ncaa.org/membership/ed_outreach/health-safety/iss/Reports2003-04. Accessed April 3, 2004.
9. Guskiewicz KM, McCrea M, Marshall SW, et al. Cumulative effects associated with recurrent concussion in collegiate football players: the NCAA concussion study. *JAMA.* 2003;290:2549–2555.
10. Wojtys EM, Hovda DA, Landry G, et al. Current concepts: concussion in sports. *Am J Sports Med.* 1999;27:676–687.
11. Collins MW, Grindel SH, Lovell MR, et al. Relationship between concussion and neuropsychological performance in college football players. *JAMA.* 1999;282:964–970.
12. Echemendia RJ, Putukian M, Mackin RS, Julian L, Shoss N. Neuropsychological test performance prior to and following sports-related mild traumatic brain injury. *Clin J Sport Med.* 2001;11:23–31.
13. Guskiewicz KM, Ross SE, Marshall SW. Postural stability and neuropsychological deficits after concussion in collegiate athletes. *J Athl Train.* 2001;36:263–273.
14. McCrea M, Kelly JP, Randolph C, Cisler R, Berger L. Immediate neurocognitive effects of concussion. *Neurosurgery.* 2002;50:1032–1042.
15. Bleiberg J, Cernich AN, Cameron KL, et al. Duration of cognitive impairment after sports concussion. *Neurosurgery.* 2004;54:1073–1080.
16. Grindel SH, Lovell MR, Collins MW. The assessment of sport-related concussion: the evidence behind neuropsychological testing and management. *Clin J Sport Med.* 2001;11:134–143.
17. Collie A, Darby D, Maruff P. Computerised cognitive assessment of athletes with sports related head injury. *Br J Sports Med.* 2001;35:297–302.
18. Collins MW, Field M, Lovell MR, et al. Relationship between postconcussion headache and neuropsychological test performance in high school athletes. *Am J Sports Med.* 2003;31:168–173.
19. Makdissi M, Collie A, Maruff P, et al. Computerised cognitive assessment of concussed Australian rules footballers. *Br J Sports Med.* 2001;35:354–360.
20. Echemendia RJ, Julian LJ. Mild traumatic brain injury in sports: neuropsychology's contribution to a developing field. *Neuropsychol Rev.* 2001;11:69–88.
21. Bleiberg J, Halpern EL, Reeves D, Daniel JC. Future directions for the neuropsychological assessment of sports concussion. *J Head Trauma Rehabil.* 1998;13:36–44.
22. Wilken JA, Kane R, Sullivan CL, et al. The utility of computerized neuropsychological assessment of cognitive dysfunction in patients with relapsing-remitting multiple sclerosis. *Mult Scler.* 2003;9:119–127.
23. Collie A, Maruff P, Makdissi M, McCrory P, McStephen M, Darby D. CogSport: reliability and correlation with conventional cognitive tests used in postconcussion medical evaluations. *Clin J Sport Med.* 2003;13:28–32.
24. Collins MW, Iverson GL, Lovell MR, McKeag DB, Norwig J, Maroon J. On-field predictors of neuropsychological and symptom deficit following sports-related concussion. *Clin J Sport Med.* 2003;13:222–229.
25. Erlanger D, Feldman D, Kutner K, et al. Development and validation of

a web-based neuropsychological test protocol for sports-related return-to-play decision-making. *Arch Clin Neuropsychol.* 2003;18:293–316.
26. Erlanger D, Kaushik T, Cantu R, et al. Symptom-based assessment of the severity of a concussion. *J Neurosurg.* 2003;98:477–484.
27. Erlanger D, Saliba E, Barth J, Almquist J, Webright W, Freeman J. Monitoring resolution of postconcussion symptoms in athletes: preliminary results of a web-based neuropsychological test protocol. *J Athl Train.* 2001;36:280–287.
28. Iverson GL, Lovell MR, Collins MW. Interpreting change on ImPACT following sport concussion. *Clin Neuropsychol.* 2003;17:460–467.
29. Reeves DL, Bleiberg J, Roebuck-Spencer TM, et al. Reference values for performance on the Automated Neuropsychological Assessment Metrics V3.0 in an active duty military sample. *Mil Med.* 2006;171:982–994
30. Brown CN, Guskiewicz KM, Bleiberg J. Computerized neuropsychological assessment of collegiate athletes: establishing normative data. *J Athl Train.* 2004;39:S17.
31. Lovell MR, Collins MW. New developments in the evaluation of sports-related concussion. *Curr Sports Med Reports.* 2002;1:287–292.
32. Barr WB. Neuropsychological testing of high school athletes: preliminary norms and test-retest indices. *Arch Clin Neuropsychol.* 2003;18:91–101.
33. Bleiberg J, Garmoe WS, Halpern EL, Reeves DL, Nadler JD. Consistency of within-day and across-day performance after mild brain injury. *Neuropsychiatry Neuropsychol Behav Neurol.* 1997;10:247–253.
34. Daniel JC, Olesniewicz MH, Reeves DL, et al. Repeated measures of cognitive processing efficiency in adolescent athletes: implications for monitoring recovery from concussion. *Neuropsychiatry Neuropsychol Behav Neurol.* 1999;12:167–169.
35. Hoddes E, Zarcone V, Smythe H, Phillips R, Dement WC. Quantification of sleepiness: a new approach. *Psychophysiology.* 1973;10:431–436.
36. Brown CN, Guskiewicz KM, Bleiberg J, McCrea M, Marshall S, Matthews A. Comprehensive assessment of concussion in high school and collegiate athletes. *J Athl Train.* 2003;38:S24.
37. Bleiberg J, Reeves D. Procedural reaction time. *ANAM Sports Medicine Battery Research Edition: Software User's Guide.* Ft Detrick, MD: Army Medical Research and Materiel Command; January 17, 2005:12.
38. Marshall SW. Testing with confidence: the use (and misuse) of confidence intervals in biomedical research. *J Sci Med Sport.* 2004;7:135–137.
39. Wolfe R, Cumming G. Communicating the uncertainty in research findings: confidence intervals. *J Sci Med Sport.* 2004;7:138–143.
40. Poole C. Low P-values or narrow confidence intervals: which are more durable? *Epidemiology.* 2001;12:291–294.
41. The College Board. 2004 college-bound seniors: a profile of SAT program test takers. Available at: http://www.collegeboard.com/prod_downloads/about/news_info/cbsenior/yr2004/2004_CBSNR_total_group.pdf. Accessed June 5, 2005.
42. Heaton RK, Ryan L, Grant I, Matthews CG. Demographic influences on neuropsychological test performance. In: Grant I, Adams KM, eds. *Neuropsychological Assessment of Neuropsychiatric Disorders.* 2nd ed. New York, NY: Oxford University Press; 1996:141–163.
43. Beatty WW, Mold JW, Gontkovsky ST. RBANS performance: influences of sex and education. *J Clin Exp Neuropsychol.* 2003;25:1065–1069.
44. Rollinson DC, Rathlev NK, Moss M, et al. The effects of consecutive night shifts on neuropsychological performance of interns in the emergency department: a pilot study. *Ann Emerg Med.* 2003;41:400–406.
45. Purnell MT, Feyer AM, Herbison GP. The impact of a nap opportunity during the night shift on the performance and alertness of 12-h shift workers. *J Sleep Res.* 2002;11:219–227.
46. Zuri RE, Cleary MA, Lopez RM, Jones LC. Effects of active dehydration on cognitive performance. *J Athl Train.* 2004;39:S57–S58.

Cathleen N. Brown, PhD, ATC, and Kevin M. Guskiewicz, PhD, ATC, FACSM, participated in conception and design; acquisition and analysis and interpretation of the data; and drafting, critical revision, and final approval of the article. Joseph Bleiberg, PhD, participated in conception and design; analysis and interpretation of the data; and critical revision and final approval of the article.

Address correspondence to Cathleen N. Brown, PhD, ATC, Department of Kinesiology, University of Georgia, 330 River Road, Athens, GA 30602. Address e-mail to browncn@uga.edu.