OXFORD
UNIVERSITY PRESS

Archives of Clinical Neuropsychology 32 (2017) 168–183

Archives of CLINICAL NEUROPSYCHOLOGY

Normative Data for a Comprehensive Neuropsychological Test Battery used in the Assessment of Sports-Related Concussion

Victoria C. Merritt*, Jessica E. Meyer, Margaret H. Cadden, Cristina A.F. Roman, Dede M. Ukueberuwa, Michael D. Shapiro, Peter A. Arnett

Department of Psychology, Pennsylvania State University, University Park, PA, USA

*Corresponding author at: Department of Psychology, Pennsylvania State University, 372 Moore Building, University Park, PA 16802, USA. Tel.: 814-865-5578; fax: 814-863-1331.
E-mail address: vcabarnes@gmail.com (V.C. Merritt).

Accepted 4 October 2016

Abstract

Objective: The use of normative data is a hallmark of the neuropsychological assessment process. Within the context of sports-related concussion, utilizing normative data is especially essential when individualized baseline data are unavailable for comparison. The primary purpose of this study was to establish normative data for a comprehensive neuropsychological test battery used in the assessment of sports-related concussion. A secondary aim was to provide normative data for pertinent demographic variables relevant to the assessment of college athletes, including sex, previous head injuries (PHI), and history of attention deficit hyperactivity disorder (ADHD)/learning disability (LD).
Method: Participants included male and female college athletes ($N = 794$) who were involved in a concussion management program at an NCAA Division I university between 2002 and 2015. Athletes were administered a comprehensive neuropsychological test battery at baseline designed to assess the following cognitive domains: learning and memory, attention and concentration, processing speed, and executive functioning. The test battery primarily comprises paper-and-pencil measures.
Results: Normative data are presented for the overall athlete sample. Additional sub-norms are then provided for specified demographic populations (i.e., sex, PHIs, and history of ADHD/LD). Findings indicate that there are mild cognitive differences between men and women, as well as between those athletes with and without a history of ADHD/LD. Given these findings, additional norms are provided for men and women with and without a history of ADHD/LD.
Conclusions: In the absence of baseline testing, the normative data presented here can be used clinically to assess athletes' cognitive functioning post-concussion.

Keywords: Assessment; Head injury; Traumatic brain injury; Norms/normative studies

Introduction

The Role of Neuropsychology in Concussion Assessment

The impact of sports-related concussions on athletes' cognitive functioning has become a growing health concern over the past several years. Importantly, neuropsychologists are in a unique position to offer valuable services to those concussed. In particular, neuropsychological testing has been shown to be an advantageous tool in evaluating cognitive deficits following injury, appraising the functional impact of symptoms, and aiding in return to play decisions (Echemendia et al., 2013; McCrory et al., 2013; Moser et al., 2007). Despite these benefits, when and how neuropsychological testing should be used to manage sports-related concussions have been debated among health-care providers and researchers alike, with various concussion management models emerging over the last two decades.

doi:10.1093/arclin/acw090 Advance Access publication on 22 October 2016

The Sports Laboratory Assessment Model (SLAM), which incorporates the use of baseline cognitive testing, has served as a model for the clinical management of sports-related concussions since its development in the late 1980s (Barth et al., 1989). The use of baseline testing has been widely regarded as best practice for concussion management, as it allows for an individualistic approach that can account for pre-injury differences in cognitive profiles, as well as specific factors influencing baseline performance such as the presence of attention deficit hyperactivity disorder (ADHD) or learning disorders (LD), history of concussion, cultural/linguistic differences, and psychiatric issues (Barr, 2003; Elbin et al., 2013; Guskiewicz et al., 2003; McCrory et al., 2013; Moser et al., 2007). Thus, baseline testing should, in theory, provide a more sensitive approach for evaluating the degree of cognitive change following concussion. However, despite the potential benefits of the SLAM approach to concussion management, the model has not escaped criticism.

A major concern surrounding the use of baseline testing is that there is little empirical evidence to suggest that (a) baseline testing actually reduces the risk of concussion and (b) the use of such testing offers any clinical utility (for a review, see Randolph, 2011). A related criticism of the SLAM approach has been the lack of established test–retest reliabilities for the time intervals often seen in sports concussion assessments (Broglio, Ferrara, Macciocchi, Baumgartner, & Elliott, 2007; Ellemberg, Henry, Macciocchi, Guskiewicz, & Broglio, 2009; Randolph, 2011). While test–retest reliabilities for time intervals of weeks, months, and up to 1 year have been established for many measures—primarily computerized tests, post-concussion assessments can occur several years after a baseline assessment, bringing into question the stability of the measure(s) over the course of that time period (Arnett et al., 2014). In a university setting, for instance, it is possible that an athlete underwent baseline testing as a freshman and then sustained a concussion as a senior. In this scenario, a time interval of 3–4 years has elapsed between the initial test and the retest, raising concerns about the reliability of the measures. Additionally, as time elapses between the initial testing period and the retest, the internal and external validity of the instruments may be questioned.

As a result of the abovementioned issues related to test–retest time intervals, some recent research has focused on establishing test–retest reliabilities for time periods of greater than 1 year (Echemendia et al., 2016; Tsushima et al., 2016). However, even in the presence of established test–retest reliabilities over longer time periods, the use of baseline testing may not always be feasible. Significant time, money, and personnel are required to administer baseline assessments to all athletes, especially when administering paper-and-pencil measures, and this can serve as a major obstacle to the accessibility of baseline assessments (Arnett et al., 2014; Echemendia et al., 2012; Elbin, Schatz, & Covassin, 2011).

Use of Normative Data

In cases where baseline assessments are unavailable or inappropriate to use due to long test–retest intervals, it is critical that clinicians have appropriate normative data to use. Normative data can be general—the average performance for a diverse sample, or can be more specific—the average performances of groups divided by age, sex, education, ethnicity, and so forth (Strauss, Sherman, & Spreen, 2006). While there are benefits to the use of both general and specific normative data, in post-concussion assessments normative data that most closely match an athlete's demographics may provide the best comparison method. To illustrate, when considering the context of a sports concussion evaluation, a critical question to be addressed by the neuropsychologist is whether it is safe for the athlete to return to play. Given the individualistic nature of this question, applying specific normative data should improve the accuracy of the test results and interpretation of the data.

Demographics that may be particularly relevant in concussion assessments include sex, history of head injury, and diagnosis of ADHD or LD. Although other factors such as sport, psychiatric history, and ethnicity or culture may also be pertinent, for the purpose of the present study our focus is limited to the basic demographic variables that are commonly assessed in a concussion clinic and that have been shown to have an influence on neuropsychological outcomes. Specifically, differences in cognitive performance between men and women have been observed at both baseline and post-concussion (Broshek et al., 2005; Covassin et al. 2006; Covassin, Elbin, Harris, Parker, & Kontos, 2012; Covassin, Elbin, Kontos, & Larson, 2010; Covassin, Schatz, & Swanik, 2007). Additionally, athletes often have a history of head injury prior to completing a baseline assessment and these previous injuries have been found to affect cognitive performance at baseline (Covassin et al., 2010) and post-concussion (Colvin et al., 2009). However, these findings are not consistent across the literature (Collie, McCrory, & Makdissi, 2006; Iverson, Brooks, Collins, & Lovell, 2006), and thus warrant further investigation. Both ADHD and LD have also been found to affect cognitive performance of athletes at baseline (Elbin et al., 2013), and normative data for a popular computerized neuropsychological assessment (Immediate Post-Concussion Assessment and Cognitive Testing, ImPACT) have already been published for these populations (Zuckerman, Lee, Odom, Solomon, & Sills, 2013). Using normative data that account for factors known to impact baseline performance allows for the consideration of potential pre-injury differences even in the absence of baseline assessments.

Limitations of Currently Available Normative Data

The use of normative data is a cornerstone of neuropsychological assessment. A normative sample is often treated as if it were a population, a practice that is justifiable only when sample sizes are sufficiently large (Crawford & Howell, 1998). While many neuropsychological assessments include large sample sizes for the overall normative sample, when broken down by demographics, sample sizes are often significantly reduced. As an example, concussion assessments frequently use the Hopkins Verbal Learning Test-Revised (HVLT-R), a measure of verbal learning and memory that includes six alternate forms (Brandt & Benedict, 2001). The total normative sample comprises over 1,100 participants; however, the sample applicable to college athletes is significantly smaller. The normative sample for 16–19 year olds is only 29 participants (13 men and 16 women) and for 20–29 year olds, the sample is limited to 84 participants (42 men, 42 women; Brandt & Benedict, 2001). With respect to the Brief Visuospatial Memory Test-Revised (BVMT-R), a measure of visual learning and memory, the sample size applicable to college students is only slightly improved at 178 (Benedict, 1997). However, the breakdown of men and women included in this sample is not specified. As a final example, consider the Stroop Color-Word Test (SCWT), a measure of executive functioning (Trenerry, Crosson, DeBoe, & Leber, 1989). Normative data for the SCWT are divided into only two age groups: (a) 18–49 year olds and (b) 50+ year olds. The sample size for the 18–49 year-old group is 106 (43 men and 63 women). Given these figures, it would be beneficial to establish more robust normative data for use in the assessment of sports-related concussion, with an emphasis on generating normative values that take into account relevant demographics of college athletes. Generating such data has clear clinical implications as described earlier.

Paper-and-Pencil Versus Computerized Assessments

With the advent of computerized neuropsychological assessment tools, the sports concussion literature has seen a rise in the number of concussion management programs that utilize computerized testing—especially the ImPACT. While there may be several strengths associated with computerized testing, there are also many advantages associated with the use of more traditional neuropsychological tests. Although a complete discussion of the advantages and disadvantages associated with different testing methods is beyond the scope of this paper, it is noteworthy to highlight that paper-and-pencil testing allows for an individualized testing environment, as well as the ability to assess a breadth of cognitive domains. Additionally, recent work by Meyer and Arnett (under review) showed that paper-and-pencil tests were more sensitive to post-concussion decline than the ImPACT, and past research has demonstrated that several paper-and-pencil measures have well-established psychometric properties (see Measures section for more details).

Present Study

Given the earlier discussion, the purpose of the present study was to (a) provide normative data gathered from a large sample of collegiate athletes for a comprehensive paper-and-pencil test battery that can be used within the context of a sports concussion management program or a concussion clinic and (b) to provide normative data based on relevant demographic characteristics including sex, previous head injuries (PHI), and history of ADHD/LD.

Materials and Methods

Participants

Participants included 794 athletes who were involved in a concussion management program at an NCAA Division I university between 2002 and 2015. Athletes participating in the program are administered a battery of neuropsychological measures prior to any collegiate athletic activity, and in the event of a concussion, undergo neuropsychological testing post-concussion. The program is designed to function as a clinical service to the university's Sports Medicine Department; therefore, athletes are referred to our program for the purpose of a clinical evaluation and are offered research participation at the beginning of their testing appointment. Appointments are made prior to each team's athletic season and generally take place during the team's preparticipation physicals. Since our program's establishment, approximately 60–80 athletes have been enrolled in the program annually, representing the following athletic teams: Men's and Women's Basketball, Football, Men's Ice Hockey, Men's and Women's Lacrosse, Men's and Women's Soccer, and Wrestling. Occasionally, we may receive a post-concussion referral from a sports team that does not routinely participate in our program; in this event, testing is completed in the absence of a baseline evaluation. For the purpose of this study, only baseline data were examined.

Study participants were selected from a database comprising all athletes who enrolled in the concussion management program since 2002 ($N = 985$). The following inclusion criteria were used to select participants for the present study: (a) participated in baseline testing ($n = 920$; 93.4% of original sample); (b) had a complete neuropsychological data set at baseline ($n = 826$, 83.9% of original sample); and (c) demonstrated adequate effort on baseline testing ($n = 794$, 80.6% of original sample). Adequate effort was defined as having an ImPACT Impulse Control Composite (ICC) score of ≤30 (ImPACT Applications Inc, 2012). The rationale for selecting the ICC as an indicator of effort was twofold: (a) ICC data were available for all athletes in the selected sample and (b) the ICC is a commonly used indicator of effort in sports concussion assessments (Covassin, Elbin, Larson, & Kontos, 2012; Schatz, Moser, Covassin, & Karpf, 2011) and is easily generated from the ImPACT Clinical Report. Participant demographic characteristics are presented in Table 1. As noted in Table 1, approximately one-third of the sample had sustained at least one prior head injury; however, at the time of baseline testing, all athletes were considered non-injured.

Procedures

All athletes were administered a neuropsychological test battery at baseline, comprised primarily of paper-and-pencil measures. When available, alternate test forms were used (see Measures section for a detailed description of specific measures and forms used). Tests were administered by undergraduate research assistants or graduate students under the supervision of a Ph.D.-level clinical neuropsychologist. Testing was completed on an individual basis in a quiet room to provide an optimal testing environment. The purpose of testing was explained to all athletes and examinees were encouraged to put forth their best effort. The comprehensive baseline evaluations took approximately 2 h to complete, including the time required for administrative procedures and paperwork; required procedures for the present study took approximately 1-1.5 h. All participants signed an informed consent form prior to participation in research-related activities and the study was approved by the University's Institutional Review Board.

Table 1. Participant demographics ($N = 794$)

Variables	*M*	*SD*
Age	18.48	1.01
Education	12.17	0.78
Variables	*N*	%
Sex		
Men	577	72.7
Women	217	27.3
Ethnicity		
African American	159	20.0
Asian	7	0.9
Biracial/Multiracial	24	3.0
Caucasian	587	73.9
Hispanic	9	1.1
Other	8	1.0
Previous head injuries		
Absent	495	62.3
Present	299	37.7
History of ADHD/LD		
Absent	725	91.3
Present	69	8.7
Sport		
Basketball	92	11.6
Football	225	28.3
Hockey	66	8.3
Lacrosse	190	23.9
Soccer	188	23.7
Wrestling	27	3.4
Other	6	0.8

Notes: ADHD = attention deficit hyperactivity disorder; LD = learning disability.

Measures

The neuropsychological test battery was designed to assess a wide range of cognitive domains that are typically affected following concussion. Domains of interest included learning and memory, attention and concentration, processing speed, and executive functioning. A brief description of each measure used, along with information about alternate forms, follows.

Brief Visuospatial Memory Test-Revised. The BVMT-R is a test of visual learning and memory (Benedict, 1997). Examinees are presented with a display of six geometric figures for 10 s, and after the display has been removed, they are asked to draw the figures as accurately as possible, and in the same location, on a blank piece of paper. After the first display, there are two additional learning trials, and then a Delayed Recall Trial is administered 25 min after the third learning trial. The test has six alternate forms and Forms 1–4 were used in the current study. The BVMT-R has been shown to have high reliability (Benedict, 1997) and validity (Benedict, Schretlen, Groninger, Dobraski, & Shpritz, 1996).

Comprehensive Trail-Making Test. The CTMT is a measure of attention and concentration, visual-motor speed, and mental flexibility (Reynolds, 2002). The CTMT comprises five unique trials, and Trails 2–5 were used in the current study. Each participant was administered a set of two trails—Trails 2 and 4 were administered together and Trails 3 and 5 were administered together. Trails 2 and 3 involve serially connecting numbers as quickly as possible and Trails 4 and 5 involve the added component of a set-shifting task. The CTMT has been found to have high reliability and validity (Gray, 2006).

Hopkins Verbal Learning Test-Revised. The HVLT-R is a measure of verbal learning and memory (Brandt & Benedict, 2001). Examinees are read a list of 12 words and are then asked to repeat as many words as they can, in any order, from the list. The HVLT-R comprises three learning trials, a Delayed Recall Trial administered about 20–25 min after the third learning trial, and a delayed recognition trail. The test has six alternate forms; forms 1, 3, 4, and 5 were used in the current study. The reliability and validity of the HVLT-R has been adequately demonstrated (Benedict, Schretlen, Groninger, & Brandt, 1998; Brandt & Benedict, 2001; Shapiro, Benedict, Schretlen, & Brandt, 1999), and the test has been shown to be useful in the assessment of concussion (Bruce & Echemendia, 2003).

PSU Cancellation Task. The PSU Cancellation Task is a measure of attention and visual scanning (Echemendia & Julian, 2001) where examinees are presented with a display containing many symbols and a target symbol at the top of the display. Examinees are asked to draw a line through each symbol in the display that is identical to the target symbol and are directed to work as quickly as possible within an allotted time. The test has five alternate forms; Forms A–D were used in the current study. The PSU Cancellation Task has been used previously in the assessment of sports-related concussion (Echemendia & Julian, 2001; Echemendia, Putukian, Mackin, Julian, & Shoss, 2001; Rabinowitz & Arnett, 2012), and the reliability of the measure has been documented (Conder, Conder, Register-Mihalik, Conder, & Newton, 2015).

Symbol-Digit Modalities Test. The SDMT is a measure of memory, attention, and motor speed (Smith, 1991). Examinees are presented with a display containing a series of symbols, and a coding key at the top of the display with nine symbol–number pairs. Examinees work within an allotted time to pair the symbols in the display with the corresponding number in the coding key. Examinees are asked to record their answers as quickly as possible. Forms A–D (Rao & The Cognitive Function Study Group of the National Multiple Sclerosis Society, 1990) were used in the current study. The SDMT has been demonstrated to have high reliability (Strauss et al., 2006) and validity (Ponsford & Kinsella, 1992), and the test has been shown to be effective in assessing recovery following brain injury (Felmingham, Baguley, & Green, 2004; Smith, 1991).

Stroop Color-Word Test. The SCWT comprises two trials that are designed to assess processing speed, response inhibition, and set-shifting/cognitive flexibility (Trenerry et al., 1989). In both trials, examinees are presented with a stimulus sheet that contains 112 color-words (e.g., blue, green, red, and tan). Each color-word is printed in a non-matching color (e.g., the word "red" is printed in blue ink). During the first trial (SCWT-Word Task), examinees are asked to read the words aloud as quickly as possible. During the second trial (SCWT-Color/Word Task), examinees are asked to name the printed color of the word as quickly as possible. No alternate forms are available. The SCWT has been shown to have adequate reliability and validity (Strauss et al., 2006; Trenerry et al., 1989).

Vigil/W Continuous Performance Test. The Vigil is a computerized test designed to measure attention, concentration, and reaction time (Cegalis & Cegalis, 1994). During this task, a series of letters flash on a computer screen one at a time, and examinees are directed to press the space bar as quickly as possible each time the letter "K" appears. No alternate versions of this test were used. The Vigil has been previously used in the assessment of sports-related concussion (Bailey, Echemendia, & Arnett, 2006; Echemendia & Julian, 2001; Echemendia et al., 2001), and the measure has high reliability and validity (Cegalis & Cegalis, 1994).

Wechsler Adult Intelligence Scale-III Digit Span Test. The Digit Span Test is a subtest of the WAIS-III and was designed to measure working memory and attention (Wechsler, 1997). The Digit Span Test comprises Digits Forward and Digits Backward. During Digits Forward, examinees are read a series of numbers and are asked to repeat the number sequence in the same order as it was presented. During Digits Backward, examinees are read a series of numbers and are asked to repeat the number sequence in the reverse order. No alternate forms are available. The Digit Span Test has been shown to have adequate reliability and validity (The Psychological Corporation, 1997). In the present study, a modified version of the Digit Span Test was used. Specifically, for each trial, if the participant was correct on the first trial of a string (say, with three numbers), the examiner would move to the next trial without administering the second trial (so, would move straight to the four-number trial).

Additionally, athletes were asked to complete background paperwork that inquired about demographic characteristics, medical history, and social history. For example, we collected information about PHI, history of ADHD, and history of LD via a questionnaire.

Finally, the ImPACT program was also included as part of our comprehensive neuropsychological test battery. The ImPACT is a computerized test that was designed specifically for use in the assessment and management of sports-related concussions (Lovell et al., 2000). The test comprises six modules that assess attention, memory, visual scanning, reaction time, and processing speed. Five composite scores are derived from these modules, one of which is the ICC. This particular composite can serve as a measure of effort, with scores >30 suggesting poor test performance and/or an invalid baseline evaluation (ImPACT Applications Inc, 2012). For the purpose of the present study, the ICC was used to determine invalid baseline performance; as noted earlier, those with ICC scores >30 were excluded from data analysis. The remaining four composite scores were not analyzed in the present study because the primary focus of this study was to provide normative data on paper-and-pencil measures, and several studies have previously documented normative data for the ImPACT across several levels of sport (Bruce, Echemendia, Meeuwisse, Comper, & Sisco, 2014; Elbin et al., 2011; Iverson, Lovell, & Collins, 2003; Schatz, 2010).

Approach to Data Analysis

Descriptive statistics were calculated on the overall sample. Means and standard deviations are reported for continuous variables and frequencies and proportions are reported for categorical variables. Alternate test forms were used when available and ANOVAs were conducted to verify form equivalencies across each measure. Post hoc comparisons using the Tukey HSD test were consulted to examine any observed differences between form versions. Given the large sample size, it was determined *a priori* that if the overall ANOVA was significant and differences were observed between forms for a particular measure, effect sizes (Eta-squared) would be examined. Any measure showing significant differences between the forms but having a medium effect size or smaller (defined as $\eta^2 \leq .06$) would be interpreted as being equivalent, whereas any measure showing significant differences between forms but having a medium-large to large effect size (defined as $\eta^2 > .06$) would be interpreted as inequivalent. In the event of inequivalent forms, normative data would be reported independently by form.

After examining the sample as a whole, participants were divided into groups based on the following categories: (a) sex; (b) PHI; and (c) history of ADHD/LD. PHI were self-reported by the athletes at the time of their baseline evaluation. Athletes were provided with a definition of concussion, and then were asked how many concussions they have sustained throughout their life based on the given definition. Athletes reporting a history of 1 or more prior concussions were classified into the "PHI Present" group and athletes reporting no prior concussions were classified into the "PHI Absent" group. With respect to history of ADHD/LD, athletes were asked via questionnaire whether they had been diagnosed with ADHD or LD; athletes reporting either a diagnosed history of ADHD *or* LD were classified into the "ADHD/LD Present" group and athletes denying a history of ADHD *and* LD were classified into the "ADHD/LD Absent" group. Independent samples *t*-tests were used to compare neuropsychological test performance between the groups, and a significance level of $\alpha = 0.05$ was applied. (Given that the purpose of the study was to provide normative data on various samples of collegiate athletes, we were not concerned with correcting for multiple comparisons.) Effect sizes were reported using Cohen's *d*. All statistical analyses were conducted using IBM SPSS Statistics, Version 22 (IBM Corp., 2015).

Results

Means and standard deviations of the neuropsychological test battery for the overall sample are reported in Table 2. Alternate test forms were used for the BVMT-R, CTMT, HVLT-R, PSU Cancellation Test, and SDMT. The results of the ANOVA testing showed that the BVMT-R, HVLT-R, and SDMT forms were equivalent; thus, the normative data are reported altogether in Table 2. Conversely, the CTMT and PSU Cancellation Test forms were not equivalent; thus, normative data are reported by form in Table 2. Specifically, the four CTMT trials were all significantly different from one another and were reported in Table 2 as such. PSU Cancellation Test Forms A and B were significantly different from Forms C and D; data are reported accordingly in Table 2. The ANOVA results are available in the Supplementary Material online (see Table 1A).

Table 2. Normative data for comprehensive concussion test battery for overall sample

Neuropsychological Test Indices	Overall sample	
	($N = 794$)	
	M	*SD*
BVMT-R		
Trial 1	7.01	2.46
Trial 2	9.78	2.00
Trial 3	10.70	1.49
Total Recall	27.49	5.18
Delayed Recall	10.40	1.66
Percent Retained	96.96	10.98
CTMT		
Trail 2 ($n = 363$)	34.15	10.69
Trail 3 ($n = 431$)	41.94	14.05
Trial 4 ($n = 363$)	29.30	11.04
Trail 5 ($n = 431$)	50.18	18.55
HVLT-R		
Trial 1	6.98	1.59
Trial 2	9.41	1.55
Trial 3	10.31	1.36
Total Recall	26.72	3.76
Delayed Recall	9.49	1.82
Percent Retained	91.41	14.34
PSU Cancellation Test—Form A or B		
Omissions ($n = 360$)	4.01	4.05
Commissions ($n = 360$)	0.19	0.70
Total Correct ($n = 360$)	40.34	6.89
PSU Cancellation Test—Form C or D		
Omissions ($n = 434$)	1.43	1.86
Commissions ($n = 434$)	0.18	0.66
Total Correct ($n = 434$)	54.64	9.12
SDMT		
Total Correct	61.16	11.19
Incidental Memory	13.29	2.59
SCWT		
Word Time	53.63	9.38
Color-Word Time	111.95	21.27
Vigil		
Average Delay	414.04	38.62
Omissions	1.58	3.04
Commissions	2.55	3.02
Hit Rate	0.98	0.03
WAIS-III Digit Span		
Digits Forward	11.41	2.08
Digits Backward	7.76	2.41
Digits Total	19.16	3.77

Notes: BVMT-R = Brief Visuospatial Memory Test-Revised; CTMT = Comprehensive Trail-Making Test; HVLT-R = Hopkins Verbal Memory Test-Revised; SDMT = Symbol-Digit Modalities Test; SCWT = Stroop Color-Word Test; WAIS-III = Wechsler Adult Intelligence Scale—Third Edition.

Table 3 reports the means and standard deviations of the neuropsychological test battery for men and women. Independent samples *t*-test results comparing performance by sex are also listed in Table 3, along with an effect size for each comparison made. Results showed significant sex differences on measures of visual and verbal memory, attention and processing speed, and set-shifting/cognitive flexibility (see Table 3), with small-medium effect sizes (d = 0.20–0.45).

The means and standard deviations of the neuropsychological test battery for athletes with and without a history of concussion ("PHI Absent" vs. "PHI Present") as well as for athletes with and without a history of ADHD/LD ("ADHD/LD Absent" vs. "ADHD/LD Present") are presented in Tables 4 and 5, respectively. Independent samples *t*-test results comparing performance by PHI and ADHD/LD groups, along with an effect size for each comparison made, are also listed in Tables 4 and 5, respectively. Results for the PHI groups showed that there were largely no significant differences in test performance and effect sizes were generally small. Results for the ADHD/LD groups showed differences on measures of visual and verbal

Table 3. Normative data for comprehensive concussion test battery by gender

Baseline Neurocognitive Test Indices	Men			Women			*t*	*p*	Cohen's effect sizes (*d*)[a]
	N	*M*	*SD*	*N*	*M*	*SD*			
BVMT-R									
Trial 1	577	6.85	2.47	217	7.43	2.41	−2.98	.003	0.24
Trial 2		9.67	2.06		10.05	1.81	−2.52	.012	0.20
Trial 3		10.60	1.53		10.95	1.37	−3.11	.002	0.24
Total Recall		27.13	5.24		28.44	4.90	−3.20	.001	0.26
Delayed Recall		10.29	1.73		10.67	1.44	−3.14	.002	0.24
Percent Retained		96.65	10.58		97.77	11.97	−1.28	.200	0.10
CTMT									
Trail 2	253	34.81	11.03	110	32.65	9.74	1.77	.077	0.21
Trail 3	324	42.16	14.56	107	41.28	12.43	0.56	.577	0.07
Trail 4	253	29.96	11.44	110	27.79	9.95	1.72	.086	0.20
Trail 5	324	51.14	19.59	107	47.26	14.63	2.17	.031	0.22
HVLT-R									
Trial 1	577	6.95	1.58	217	7.06	1.61	−0.90	.369	0.07
Trial 2		9.30	1.57		9.72	1.46	−3.48	.001	0.28
Trial 3		10.23	1.38		10.51	1.28	−2.60	.009	0.21
Total Recall		26.52	3.80		27.26	3.60	−2.47	.014	0.20
Delayed Recall		9.34	1.87		9.81	1.62	−3.25	.001	0.27
Percent Retained		90.82	14.79		92.99	12.98	−1.91	.057	0.16
PSU Cancellation Test—Form A or B									
Omissions		4.03	4.25		3.96	3.57	0.16	.874	0.02
Commissions	250	0.22	0.80	110	0.13	0.39	1.42	.156	0.14
Total Correct		39.40	6.74		42.47	6.77	−3.98	<.001	0.46
PSU Cancellation Test—Form C or D									
Omissions		1.36	1.86		1.64	1.86	−1.31	.191	0.15
Commissions	327	0.20	0.73	107	0.10	0.39	1.80	.072	0.17
Total Correct		54.23	8.96		55.90	9.51	−1.65	.101	0.18
SDMT									
Total Correct	577	60.20	11.47	217	63.73	9.99	−4.26	<.001	0.33
Incidental Memory		13.29	2.61		13.29	2.53	0.02	.986	0.00
SCWT									
Word Time	577	54.22	9.83	217	52.06	7.87	3.20	.001	0.24
Color-Word Time		114.69	21.39		104.67	19.17	6.05	<.001	0.49
Vigil									
Average Delay	577	414.51	39.71	217	412.79	35.60	0.56	.577	0.05
Omissions		1.62	3.02		1.48	3.10	0.57	.569	0.05
Commissions		2.74	3.26		2.03	2.18	3.57	<.001	0.26
Hit rate		0.98	0.04		0.99	0.03	−0.80	.422	0.28
WAIS-III Digit Span									
Digits Forward	577	11.47	2.12	217	11.23	1.98	1.50	.134	0.12
Digits Backward		7.71	2.45		7.87	2.31	−0.83	.409	0.07
Digits Total		19.19	3.82		19.10	3.64	0.30	.764	0.02

Notes: BVMT-R = Brief Visuospatial Memory Test-Revised; CTMT = Comprehensive Trail-Making Test; HVLT-R = Hopkins Verbal Memory Test-Revised; SDMT = Symbol-Digit Modalities Test; SCWT = Stroop Color-Word Test; WAIS-III = Wechsler Adult Intelligence Scale—Third Edition.
[a]Cohen's effect sizes: small (0.2), medium (0.5), large (0.8).

Table 4. Normative data for comprehensive concussion test battery by previous head injury status.

Baseline Neurocognitive Test Indices	PHI absent			PHI present			*t*	*p*	Cohen's effect sizes (*d*)[a]
	N	*M*	*SD*	*N*	*M*	*SD*			
BVMT-R									
Trial 1	495	6.86	2.48	299	7.25	2.43	−2.17	.030	0.16
Trial 2		9.68	1.98		9.94	2.02	−1.79	.074	0.13
Trial 3		10.63	1.55		10.81	1.40	−1.71	.087	0.12
Total Recall		27.17	5.21		28.01	5.08	−2.21	.028	0.16
Delayed Recall		10.32	1.69		10.52	1.62	−1.66	.097	0.12
Percent Retained		96.86	11.74		97.12	9.61	−0.32	.752	0.02
CTMT									
Trail 2	229	34.65	11.29	134	33.31	9.57	1.15	.252	0.13
Trail 3	266	42.09	14.37	165	41.69	13.56	0.29	.771	0.03
Trail 4	229	29.66	11.94	134	28.69	9.33	0.80	.422	0.09
Trail 5	266	50.14	17.87	165	50.25	19.64	−0.06	.952	0.01
HVLT-R									
Trial 1	495	6.91	1.59	299	7.10	1.58	−1.65	.099	0.12
Trial 2		9.33	1.56		9.56	1.53	−2.06	.040	0.15
Trial 3		10.25	1.39		10.40	1.30	−1.52	.130	0.11
Total Recall		26.52	3.73		27.06	3.78	−1.95	.052	0.14
Delayed Recall		9.37	1.83		9.69	1.79	−2.38	.018	0.18
Percent Retained		90.95	14.92		92.18	13.31	−1.17	.243	0.09
PSU Cancellation Test—Form A or B									
Omissions	227	4.24	4.30	133	3.61	3.56	1.43	.155	0.16
Commissions		0.17	0.63		0.22	0.80	−0.61	.545	0.07
Total Correct		40.30	7.28		40.40	6.18	−0.13	.896	0.02
PSU Cancellation Test—Form C or D									
Omissions	268	1.40	1.85	166	1.49	1.89	−0.50	.616	0.05
Commissions		0.20	0.78		0.15	0.42	0.81	.416	0.08
Total Correct		54.59	9.35		54.72	8.76	−0.14	.891	0.01
SDMT									
Total Correct	495	61.18	11.46	299	61.13	10.76	0.07	.945	0.00
Incidental Memory		13.21	2.72		13.42	2.35	−1.17	.244	0.08
SCWT									
Word Time	495	53.84	10.00	299	53.28	8.26	0.82	.413	0.06
Color-Word Time		113.00	22.02		110.21	19.87	1.79	.074	0.13
Vigil									
Average Delay	495	414.18	39.74	299	413.81	36.74	0.13	.895	0.01
Omissions		1.71	3.41		1.38	2.31	1.63	.104	0.11
Commissions		2.59	3.29		2.48	2.51	0.53	.597	0.04
Hit rate		0.99	0.03		0.98	0.04	−0.47	.642	0.28
WAIS-III Digit Span									
Digits Forward	495	11.37	2.10	299	11.46	2.06	−0.61	.542	0.04
Digits Backward		7.69	2.40		7.87	2.43	−1.03	.301	0.07
Digits Total		19.06	3.78		19.33	3.77	−1.00	.318	0.07

Notes: PHI = Previous Head Injuries; BVMT-R = Brief Visuospatial Memory Test-Revised; CTMT = Comprehensive Trail-Making Test; HVLT-R = Hopkins Verbal Memory Test-Revised; SDMT = Symbol-Digit Modalities Test; SCWT = Stroop Color-Word Test; WAIS-III = Wechsler Adult Intelligence Scale—Third Edition.

[a]Cohen's effect sizes: small (0.2), medium (0.5), large (0.8).

memory, attention and concentration, processing speed, working memory, and set-shifting/cognitive flexibility. Among the measures showing significant differences, effect sizes were generally small-medium ($d = 0.24$–0.43).

As a result of the earlier findings, a supplementary analysis (two-way ANOVA) was conducted to determine whether the interaction between athletes' sex and ADHD/LD status influenced neuropsychological performance. The means and standard deviations for each test variable are reported in Table 6. The results showed that the interaction between sex and ADHD/LD status was significant for the following indices: BVMT-R Delayed Recall ($F(1, 790) = 4.53$, $p = .034$, $\eta_p^2 = 0.006$); HVLT-R Trial 2 ($F(1, 790) = 5.42$, $p = .020$, $\eta_p^2 = 0.007$); and HVLT-R Total Recall ($F(1, 790) = 5.86$, $p = .016$, $\eta_p^2 = 0.007$). Additionally, the interaction between sex and ADHD/LD status was trending toward significance for the following indices: BVMT-R Trial 3 ($F(1, 790) = 2.93$, $p = .088$, $\eta_p^2 = 0.004$) and Vigil Total Omissions ($F(1, 790) = 2.82$, $p = .094$, $\eta_p^2 = 0.004$).

Table 5. Normative data for comprehensive concussion test battery by ADHD/LD status

Baseline Neurocognitive Test Indices	ADHD/LD Absent			ADHD/LD Present			*t*	*p*	Cohen's effect sizes (*d*)[a]
	N	*M*	*SD*	*N*	*M*	*SD*			
BVMT-R									
Trial 1	725	7.06	2.45	69	6.49	2.63	−1.83	.068	0.22
Trial 2		9.83	1.96		9.26	2.29	−1.98	.051	0.27
Trial 3		10.74	1.47		10.29	1.71	−2.39	.017	0.28
Total Recall		27.62	5.09		26.04	5.89	−2.43	.015	0.29
Delayed Recall		10.43	1.65		10.01	1.78	−1.99	.047	0.25
Percent Retained		96.94	10.82		97.10	12.66	0.11	.911	0.01
CTMT									
Trail 2	340	34.17	10.81	23	33.90	8.98	−0.12	.906	0.03
Trail 3	385	41.69	13.93	46	44.00	15.05	1.06	.292	0.16
Trail 4	340	29.23	11.22	23	30.36	8.02	0.47	.637	0.12
Trail 5	385	49.52	18.21	46	55.67	20.56	2.13	.033	0.32
HVLT-R									
Trial 1	725	7.02	1.60	69	6.55	1.38	−2.34	.019	0.32
Trial 2		9.45	1.56		9.06	1.49	−1.99	.047	0.26
Trial 3		10.34	1.37		10.00	1.21	−2.18	.032	0.26
Total Recall		26.84	3.78		25.49	3.34	−2.86	.004	0.38
Delayed Recall		9.53	1.81		9.10	1.87	−1.86	.064	0.18
Percent Retained		91.49	14.19		90.58	15.96	−0.46	.648	0.06
PSU Cancellation Test—Form A or B									
Omissions	337	4.05	4.14	23	3.39	2.39	−0.75	.453	0.20
Commissions		0.19	0.71		0.13	0.46	−0.41	.679	0.10
Total Correct		40.43	6.94		38.96	6.06	−1.00	.321	0.23
PSU Cancellation Test—Form C or D									
Omissions	388	1.43	1.84	46	1.46	2.07	0.10	.922	0.02
Commissions		0.17	0.67		0.24	0.60	0.67	.504	0.11
Total Correct		54.85	9.03		52.87	9.77	−1.40	.164	0.21
SDMT									
Total Correct	725	61.50	11.08	69	57.67	11.81	−2.73	.007	0.33
Incidental Memory		13.31	2.63		13.10	2.12	−0.63	.530	0.09
SCWT									
Word Time	725	53.29	9.11	69	57.25	11.34	2.82	.006	0.39
Color-Word Time		111.07	20.13		121.16	29.41	2.79	.007	0.40
Vigil									
Average Delay	725	413.59	38.48	69	418.81	39.97	1.07	.283	0.13
Omissions		1.57	3.03		1.78	3.24	0.57	.572	0.07
Commissions		2.46	3.05		3.48	2.50	2.69	.007	0.37
Hit Rate		0.98	0.03		0.98	0.03	−0.43	.670	0.00
WAIS-III Digit Span									
Digits Forward	725	11.48	2.10	69	10.64	1.80	−3.23	.001	0.43
Digits Backward		7.80	2.44		7.29	2.05	−1.68	.093	0.23
Digits Total		19.28	3.80		17.93	3.18	−2.86	.004	0.39

Notes: BVMT-R = Brief Visuospatial Memory Test-Revised; CTMT = Comprehensive Trail-Making Test; HVLT-R = Hopkins Verbal Memory Test-Revised; SDMT = Symbol-Digit Modalities Test; SCWT = Stroop Color-Word Test; WAIS-III = Wechsler Adult Intelligence Scale—Third Edition; ADHD = attention deficit hyperactivity disorder; LD = learning disability.
[a]Cohen's effect sizes: small (0.2), medium (0.5), large (0.8).

Finally, given the significant differences observed between forms for several measures used in our concussion test battery, we provided additional tables in an appendix where normative data are listed independently by form for the BVMT-R, HVLT-R, PSU Cancellation Test, and SDMT. See Supplementary Material online (Tables 2A–5A, respectively).

Discussion

The primary purpose of the present study was to establish normative data for a comprehensive neuropsychological test battery used in the assessment of sports-related concussion. While comparison of post-concussion performance to the individual athlete's baseline performance has become somewhat common within sports concussion management programs, this practice

Table 6. Normative data for comprehensive concussion test battery by gender and ADHD/LD status

Gender	Baseline Neurocognitive Test Indices	ADHD/LD Absent			ADHD/LD Present		
		N	*M*	*SD*	*N*	*M*	*SD*
Men	BVMT-R						
	Trial 1	523	6.89	2.44	54	6.52	2.75
	Trial 2		9.72	2.02		9.24	2.35
	Trial 3		10.63	1.51		10.37	1.69
	Total Recall		27.23	5.15		26.13	5.97
	Delayed Recall		10.31	1.73		10.15	1.78
	Percent Retained		96.59	10.41		97.20	12.23
Women	BVMT-R						
	Trial 1	202	7.51	2.41	15	6.40	2.23
	Trial 2		10.10	1.77		9.33	2.16
	Trial 3		11.02	1.31		10.00	1.81
	Total Recall		28.64	4.78		25.73	5.76
	Delayed Recall		10.76	1.38		9.53	1.73
	Percent Retained		97.85	11.79		96.73	14.59
Men	CTMT						
	Trail 2	234	34.93	11.18	19	33.32	9.19
	Trail 3	289	41.97	14.45	35	43.67	15.55
	Trail 4	234	29.99	11.69	19	29.51	7.98
	Trail 5	289	50.58	19.24	35	55.80	22.07
Women	CTMT						
	Trail 2	106	32.50	9.78	4	36.67	8.47
	Trail 3	96	40.85	12.25	11	45.07	13.96
	Trail 4	106	27.54	9.97	4	34.37	7.94
	Trail 5	96	46.35	14.31	11	55.26	15.67
Men	HVLT-R						
	Trial 1	523	6.98	1.60	54	6.63	1.34
	Trial 2		9.31	1.58		9.19	1.48
	Trial 3		10.25	1.39		10.04	1.29
	Total Recall		26.59	3.84		25.89	3.34
	Delayed Recall		9.39	1.87		9.17	1.92
	Percent Retained		90.81	14.61		90.83	16.56
Women	HVLT-R						
	Trial 1	202	7.12	1.61	15	6.27	1.53
	Trial 2		9.81	1.43		8.60	1.50
	Trial 3		10.56	1.29		9.87	0.92
	Total Recall		27.50	3.53		24.07	3.01
	Delayed Recall		9.88	1.59		8.87	1.73
	Percent Retained		93.24	12.90		89.67	14.05
Men	PSU Cancellation Test—Form A or B						
	Omissions	231	4.12	4.37	19	2.89	2.11
	Commissions		0.23	0.83		0.05	0.23
	Total Correct		39.50	6.80		38.21	6.05
	PSU Cancellation Test—Form C or D						
	Omissions	292	1.34	1.82	35	1.54	2.24
	Commissions		0.19	0.74		0.29	0.67
	Total Correct		54.50	8.89		51.94	9.39
Women	PSU Cancellation Test—Form A or B						
	Omissions	106	3.89	3.59	4	5.75	2.50
	Commissions		0.11	0.35		0.50	1.00
	Total Correct		42.47	6.84		42.50	5.45
	PSU Cancellation Test—Form C or D						
	Omissions	96	1.69	1.90	11	1.18	1.47
	Commissions		0.10	0.40		0.09	0.30
	Total Correct		55.91	9.41		55.82	10.81
Men	SDMT						
	Total Correct	523	60.59	11.54	54	56.35	10.01
	Incidental Memory		13.29	2.67		13.26	1.91

(*continued on next page*)

Table 6. (*continued*)

Gender	Baseline Neurocognitive Test Indices	ADHD/LD Absent			ADHD/LD Present		
		N	*M*	*SD*	*N*	*M*	*SD*
Women	SDMT						
	Total Correct	202	63.83	9.41	15	62.40	16.36
	Incidental Memory		13.34	2.51		12.53	2.77
Men	SCWT						
	Word Time	523	53.83	9.54	54	58.02	11.76
	Color-Word Time		113.92	20.49		122.17	27.83
Women	SCWT						
	Word Time	202	51.89	7.74	15	54.45	9.48
	Color-Word Time		103.71	17.16		117.52	35.36
Men	Vigil						
	Average Delay	523	414.16	39.82	54	417.88	38.83
	Omissions		1.63	3.08		1.50	2.39
	Commissions		2.67	3.31		3.48	2.54
	Hit Rate		0.98	0.04		0.99	0.02
Women	Vigil						
	Average Delay	202	412.10	34.83	15	422.17	45.12
	Omissions		1.39	2.87		2.80	5.28
	Commissions		1.92	2.13		3.47	2.42
	Hit Rate		0.99	0.03		0.97	0.05
Men	WAIS-III Digit Span						
	Digits Forward	523	11.56	2.13	54	10.61	1.81
	Digits Backward		7.76	2.47		7.28	2.21
	Digits Total		19.32	3.85		17.89	3.34
Women	WAIS-III Digit Span						
	Digits Forward	202	11.26	1.99	15	10.73	1.83
	Digits Backward		7.91	2.37		7.33	1.40
	Digits Total		19.17	3.70		18.07	2.60

Notes: BVMT-R = Brief Visuospatial Memory Test-Revised; CTMT = Comprehensive Trail-Making Test; HVLT-R = Hopkins Verbal Memory Test-Revised; SDMT = Symbol-Digit Modalities Test; SCWT = Stroop Color-Word Test; WAIS-III = Wechsler Adult Intelligence Scale—Third Edition; ADHD = attention deficit hyperactivity disorder; LD = learning disability.

has not been universally adopted. As highlighted previously, collection of baseline data is time and resource intensive and may not be feasible for many athletic programs.

Even among programs that routinely utilize baseline testing, the data may not always be accessible or appropriate to use for one reason or another (i.e., athlete not referred for baseline testing, invalid test performance, test administration errors). Furthermore, the use of baseline data may not be ideal in some cases due to uncertain test–retest reliabilities over extended time periods.

Existing test norms are useful for the assessment of general populations, but sub-samples of groups more closely matched to the demographics and characteristics of collegiate athletes are small or unavailable. Here, we provided normative data for a comprehensive paper-and-pencil neuropsychological battery in a large sample of collegiate athletes who encompassed a wide variety of athletic teams. We also included additional sub-norms for alternative test forms and for special demographic populations, including sex, previous head injury history, and ADHD/LD history. All participants were screened for effort and only those providing sufficient effort were included in the normative sample.

Means and standard deviations for the overall sample were provided for general comparison of athletes' performance to established normative data. Alternate test forms were available for the BVMT-R, CTMT, HVLT-R, PSU Cancellation Test, and SDMT. Since alternate test forms for the BVMT-R, HVLT-R, and SDMT were deemed equivalent, the overall norms may be used for comparison. These tests may be particularly well suited for repeat assessment for athletes with extended recovery after concussion or who sustain multiple concussions throughout their collegiate career or beyond. In contrast, our data indicated non-equivalence of alternate forms for the CTMT and PSU Cancellation Test. Thus, we recommend the use of separate normative data for alternate forms of these tests.

Our goal in adding to general test norms was to provide information about cognitive tests in a sample with demographics similar to collegiate athletes. These demographics include pre-injury characteristics that could affect baseline testing performance. Given that men and women may perform differently on cognitive tests and may have different susceptibility to concussion, normative data were also reported by sex. Accordingly, our findings showed significant variation between men and women in performance on cognitive tests of visual and verbal memory, attention and processing speed, and set-shifting. These differences were significant enough to warrant reporting separate normative data, although the effect sizes were not large. These data are consistent with previous studies that have shown differences in cognitive performance at baseline between men and women (Covassin et al., 2006, 2012), and we recommend the consideration of gender-specific norms when evaluating post-injury test performance.

In addition, athletes with a history of ADHD/LD generally showed decreased performance on measures of visual and verbal memory, attention and processing speed, and set-shifting. Thus, comparison of individuals with this history to appropriate normative data provided here is recommended rather than use of general sample norms. Importantly, these findings are consistent with other studies evaluating ADHD/LD in college athletes, though these studies have been limited to the evaluation of computerized tests only (Elbin et al., 2013; Zuckerman et al., 2013). An additional set of norms was then provided that took into account both sex and ADHD/LD status. In situations where ADHD/LD is present, the clinician is advised to utilize the normative data that aligns with the athlete's sex and ADHD/LD status in order to most closely mirror the athlete's demographic. Although not all test indices varied as a function of sex and ADHD/LD status based on statistical significance, the normative values indicate that there may be meaningful differences depending on the athletes' sex and ADHD/LD status.

Finally, although many athletes have a prior history of head injury, in the present study, history of head injury did not appear to have a significant effect on neuropsychological test performance. This is consistent with previous work by Collie and colleagues (2006) and Iverson and colleagues (2006). Thus, utilizing the normative data provided for the overall sample is justified here.

The battery of neuropsychological tests used by our program covers a broad set of cognitive domains, including verbal and visual learning and memory, attention and concentration, processing speed, and executive functioning. The measures utilized in this study represent domains that are commonly affected by concussion. However, other programs may wish to use alternative tests based on accessibility or particular interests. Our objective was to emphasize the need for normative data for this population and to assess any effects of pre-injury demographic characteristics on baseline testing performance. Importantly, the measures utilized in this study can be easily adapted and used by other concussion management programs as a starting point for developing a comprehensive concussion battery, but can also be used independently. Either way, the present data are based on a large sample of athletes, which increases the confidence of using these data as a comparison group.

Limitations of this study should be addressed and kept in mind for future research. First, our study specifically focused on collegiate athletes; therefore, the generalizability of these results is limited to this specific demographic. Future research could address this limitation by taking a similar approach to establishing normative data for both younger and older athlete samples. Related, the data presented herein were derived from a Division I university where the model is to baseline test athletes during their freshman year, prior to their participation in athletics. Consequently, the applicability of this data to more senior colligate athletes is unknown, and thus another area for future research. Importantly, questions pertaining to *when* and *how often* baselines should be conducted have been raised, and this remains an ongoing area of research. It has been proposed that baselines should be conducted annually, although recent work from Bruce and colleagues (2016) showed that aggregating scores from two baselines improved temporal stability and resulted in greater ability to detect cognitive change. Clearly, more research is needed in this domain, as addressing these questions is critical for determining proper concussion management strategies moving forward.

Another limitation to this study is the method used to define adequate effort. As noted earlier, we excluded individuals who scored greater than 30 on the ImPACT ICC. Although this approach has been recognized as an appropriate method for identifying individuals exhibiting suspect effort/motivation (ImPACT Applications Inc, 2012), it is possible that participants not putting forth sufficient effort on their baseline testing were overlooked by this approach. Recent research by Nelson and colleagues (2015) compared validity indices across three computerized neurocognitive tests (ANAM, Axon, and ImPACT) and found that the ImPACT ICC was less sensitive than the ANAM and Axon at detecting invalid performance. Thus, the number of athletes in our study who were identified as having an invalid baseline may be on the low end. Given the complex issues surrounding motivation and effort in baseline testing, it is necessary for researchers to continue assessing and developing sensitive methods for gauging effort/validity. Finally, it is relevant to note that we did not consider whether athletes who self-identified as having ADHD/LD were taking medication at the time of baseline testing; consequently, the differences observed between the ADHD/LD present and absent groups may be slightly different if athletes with ADHD were untreated.

Conclusion

The normative data presented in this paper allow for the comparison of an individual's cognitive performance to a sample with similar demographic characteristics. The normative sample included a large group of college athletes of varied background and sport. We also accounted for pre-injury characteristics of athletes by characterizing testing performance by sex, concussion history, and ADHD/LD history. Sex and ADHD/LD had small effects on performance and sub-norms may be used for comparisons. These normative data address the growing need for neuropsychological information about collegiate athletes, while also minimizing demands on program resources. These data may be a valuable alternative to the use of individual baseline testing scores and improve options for effective concussion management.

Acknowledgments

The authors would like to thank Wayne Sebastianelli and Penn State Sports Medicine for their generous support of our research. We would also like to thank Gray Vargas, Amanda Rabinowitz, Fiona Barwick, Aaron Rosenbaum, and Chris Bailey for their help as project coordinators of the program over the years.

Conflict of Interest

None declared.

Supplementary material

Supplementary material is available at *Archives of Clinical Neuropsychology* online.

Funding

No formal funding was obtained for this project.

References

Arnett, P., Rabinowitz, A., Vargas, G., Ukueberuwa, D., Merritt, V., & Meyer, J. (2014). Neuropsychologial testing in sports concussion management: An evidence-based model when baseline is unavailable. In S. Slobounov & W. Sebastianelli (Eds.) *Concussion in athletics: Current understanding from basic brain science to clinical research* (pp. 35–48). New York: Springer.

Bailey, C. M., Echemendia, R. J., Arnett, P. A. (2006). The impact of motivation on neuropsychological performance in sports-related mild traumatic brain injury. *Journal of the International Neuropsychological Society*, *12*, 475–484.

Barr, W. B. (2003). Neuropsychological testing of high school athletes: Preliminary norms and test-retest indices. *Archives of Clinical Neuropsychology*, *18*, 91–101.

Barth, J. T., Alves, W. M., Ryan, T. V., Macciocchi, S. N., Rimel, R. W., Jane, J. A., & Nelson, W. E. (1989). *Mild head injury in sports: Neuropsychological sequelae and recovery of function*. New York: Oxford University Press.

Benedict, R. H. B. (1997). *Brief Visuospatial Memory Test - Revised: Professional manual*. Odessa, FL: Psychological Assessment Resources.

Benedict, R. H. B., Schretlen, D., Groninger, L., & Brandt, J. (1998). Hopkins Verbal Learning Test–Revised: Normative data and analysis of inter-form and test-retest reliability. *The Clinical Neuropsychologist*, *12*, 43–55.

Benedict, R. H. B., Schretlen, D., Groninger, L., Dobraski, M., & Shpritz, B. (1996). Revision of the Brief Visuospatial Memory Test: Studies of normal performance, reliability, and validity. *Psychological Assessment*, *8*, 145–153.

Brandt, J., & Benedict, R. H. B. (2001). *Hopkins Verbal Learning Test-Revised: Professional manual*. Odessa, FL: Psychological Assessment Resources.

Broglio, S. P., Ferrara, M. S., Macciocchi, S. N., Baumgartner, T. A., & Elliott, R. (2007). Test-retest reliability of computerized concusion assessment programs. *Journal of Athletic Training*, *42*, 509–514.

Broshek, D. K., Kaushik, T., Freeman, J. R., Erlanger, D., Webbe, F., & Barth, J. T. (2005). Sex differences in outcome following sports-related concussion. *Journal of Neurosurgery*, *102*, 856–863.

Bruce, J. M., & Echemendia, R. J. (2003). Delayed-onset deficits in verbal encoding strategies among patients with mild traumatic brain injury. *Neuropsychology*, *17*, 622–629.

Bruce, J., Echemendia, R., Meeuwisse, W., Comper, P., & Sisco, A. (2014). 1 year test-retest reliability of ImPACT in professional ice hockey players. *The Clinical Neuropsychologist*, *28*, 14–25.

Bruce, J., Echemendia, R., Tangeman, L., Meeuwisse, W., Comper, P., Hutchison, M., & Aubry, M. (2016). Two baselines are better than one: Improving the reliability of computerized testing in sports neuropsychology. *Applied Neuropsychology: Adult*, *23*, 336–342.

Cegalis, J. A., & Cegalis, S. (1994). *The Vigil/W Continuous Performance Test (manual)*. New York: ForThought.

Collie, A., McCrory, P., & Makdissi, M. (2006). Does history of concussion affect current cognitive status? *British Journal of Sports Medicine*, *40*, 550–551.

Colvin, A. C., Mullen, J., Lovell, M. R., West, R. V., Collins, M. W., & Groh, M. (2009). The role of concussion history and gender in recovery from soccer-related concussion. *The American Journal of Sports Medicine*, *37*, 1699–1704.

Conder, R. L., Conder, A. A., Register-Mihalik, J., Conder, L. H., & Newtown, S. (2015). Preliminary normative data on the Penn State University Symbol Cancellation Task with nonconcussed adolescents. *Applied Neuropsychology: Child*, *4*, 141–147.

Covassin, T., Elbin, R., Harris, W., Parker, T., & Kontos, A. (2012). The role of age and sex in symptoms, neurocognitive performance, and postural stability in athletes after concussion. *The American Journal of Sports Medicine*, *40*, 1303–1312.

Covassin, T., Elbin, R., Kontos, A., & Larson, E. (2010). Investigating baseline neurocognitive performance between male and female athletes with a history of multiple concussions. *Journal of Neurology, Neurosurgery & Psychiatry*, *81*, 597–601.

Covassin, T., Elbin, R. J., Larson, E., & Kontos, A. P. (2012). Sex and age differences in depression and baseline sport-related concussion neurocognitive performance and symptoms. *Clinical Journal of Sports Medicine*, *22*, 98–104.

Covassin, T., Schatz, P., & Swanik, C. B. (2007). Sex differences in neuropsychological function and post-concussion symptoms of concussed collegiate athletes. *Neurosurgery*, *61*, 345–351.

Covassin, T., Swanik, C. B., Sachs, M., Kendrick, Z., Schatz, P., Zillmer, E., & Kaminaris, C. (2006). Sex differences in baseline neuropsychological function and concussion symptoms of collegiate athletes. *British Journal of Sports Medicine*, *40*, 923–927.

Crawford, J. R., & Howell, D. C. (1998). Regression equations in clinical neuropsychology: An evaluation of statistical methods for comparing predicted and obtained scores. *Journal of Clinical and Experimental Neuropsychology*, *20*, 755–762.

Echemendia, R. J., Bruce, J. M., Bailey, C. M., Sanders, J. F., Arnett, P. A., & Vargas, G. (2012). The utility of post-concussion neuropsychological data in identifying cognitive change following sports-related mTBI in the absence of baseline data. *The Clinical Neuropsychologist*, *26*, 1077–1091.

Echemendia, R. J., Bruce, J. M., Meeuwisse, W., Comper, P., Aubry, M., & Hutchison, M. (2016). Long-term reliability of ImPACT in professional ice hockey. *The Clinical Neuropsychologist*, *30*, 328–337.

Echemendia, R. J., Iverson, G. L., McCrea, M., Macciocchi, S. N., Gioia, G. A., Putukian, M., & Comper, P. (2013). Advances in neuropsychological assessment of sport-related concussion. *British Journal of Sports Medicine*, *47*, 294–298.

Echemendia, R. J., & Julian, L. J. (2001). Mild traumatic brain injury in sports: Neuropsychology's contribution to a developing field. *Neuropsychology Review*, *11*, 69–88.

Echemendia, R. J., Putukian, M., Mackin, R. S., Julian, L., & Shoss, N. (2001). Neuropsychological test performance prior to and following sports-related mild traumatic brain injury. *Clinical Journal of Sport Medicine*, *11*, 23–31.

Elbin, R. J., Kontos, A. P., Kegel, N., Johnson, E., Burkhart, S., & Schatz, P. (2013). Individual and combined effects of LD and ADHD on computerized neurocognitive concussion test performance: evidence for separate norms. *Archives of Clinical Neuropsychology*, *28*, 476–484.

Elbin, R. J., Schatz, P., & Covassin, T. (2011). One-year test-retest reliability of the online version of ImPACT in high school athletes. *The American Journal of Sports Medicine*, *39*, 2319–2324.

Ellemberg, D., Henry, L. C., Macciocchi, S. N., Guskiewicz, K. M, & Broglio, S. P. (2009). Advances in sport concussion assessment: From behavioral to brain imaging measures. *Journal of Neurotrauma*, *26*, 2365–2382.

Felmingham, K. L., Baguley, I. J., & Green, A. M. (2004). Effects of diffuse axonal injury on speed of information processing following severe traumatic brain injury. *Neuropsychology*, *18*, 564–571.

Gray, R. (2006). Comprehensive Trail Making Test. *Journal of Psychoeducational Assessment*, *24*, 88–91.

Guskiewicz, K. M., McCrea, M., Marshall, S. W., Cantu, R. C., Randolph, C., Barr, W., & Kelly, J. P. (2003). Cumulative effects associated with recurrent concussion in collegiate football players: The NCAA concusion study. *JAMA*, *290*, 2549–2555.

IBM Corp. (2015). *IBM SPSS Statistics for Windows (Version 22.0)*. Armonk, NY: IBM Corp.

ImPACT Applications, Inc. (2012). *ImPACT Test Technical Manual: Online ImPACT 2007–2012*. Pittsburgh, PA: ImPACT Applications.

Iverson, G. L., Brooks, B. L., Collins, M. W., & Lovell, M. R. (2006). Tracking neuropsychological recovery following concussion in sport. *Brain Injury*, *20*, 245–252.

Iverson, G. L., Lovell, M. R., & Collins, M. W. (2003). Interpreting change on ImPACT following sport concussion. *The Clinical Neuropsychologist*, *17*, 460–467.

Lovell, M. R., Collins, M. W., Podell, K., Powell, J., & Maroon, J. (2000). *ImPACT: Immediate post-concussion assessment and cognitive testing*. Pittsburgh, PA: NeuroHealth Systems, LLC.

McCrory, P., Meeuwisse, W., Aubry, M., Cantu, B., Dvořák, J., Echemendia, R. J., & Raftery, M. (2013). Consensus statement on concussion in sport: The 4th International Conference on Concussion in Sport held in Zurich, November 2012. *British Journal of Sports Medicine*, *47*, 250–258.

Meyer, J. E. & Arnett, P. A. (2016). Clinical sensitivity of the ImPACT test and traditional neuropsychological measures for concussion in collegiate athletes. Manuscript submitted for publication.

Moser, R. S., Iverson, G. L., Echemendia, R. J., Lovell, M. R., Schatz, P., Webbe, F. M., & Broshek, D. K. (2007). Neuropsychological evaluation in the diagnosis and management of sports-related concussion. *Archives of Clinical Neuropsychology*, *22*, 909–916.

Nelson, L. D., Pfaller, A. Y., Rein, L. E., & McCrea, M. A. (2015). Rates and predictors of invalid baseline test performance in high school and collegiate athletes for 3 computerized neurocognitive tests: ANAM, Axon Sports, and ImPACT. *The American Journal of Sports Medicine*, *43*, 2018–2026.

Ponsford, J., & Kinsella, G. (1992). Attentional deficits following closed-head injury. *Journal of Clinical and Experimental Neuropsychology*, *14*, 822–838.

Rabinowitz, A. R., & Arnett, P. A. (2012). Reading based IQ estimates and actual premorbid cognitive performance: Discrepancies in a college athlete sample. *Journal of the International Neuropsychological Society*, *18*, 139–143.

Randolph, C. (2011). Baseline neuropsycholoigcal testing in managing sport-related concussion: Does it modify risk? *Current Sports Medicine Reports*, *10*, 21–26.

Rao, S. M., The Cognitive Function Study Group of the National Multiple Sclerosis Society. (1990). *Manual for the Brief Repeatable Battery of Neuropsychological Tests in Multiple Sclerosis*. New York: National Multiple Sclerosis Society.

Reynolds, C. R. (2002). *Comprehensive Trail Making Test (CTMT)*. Austin, TX: Pro-Ed.

Schatz, P. (2010). Long-term test-retest reliability of baseline cognitive assessments using ImPACT. *American Journal of Sports Medicine*, *38*, 47–53.

Schatz, P., Moser, R. S., Covassin, T., & Karpf, R. (2011). Early indicators of enduring symptoms in high school athletes with multiple previous concussions. *Neurosurgery*, *68*, 1562–1567.

Shapiro, A. M., Benedict, R. H. B., Schretlen, D., & Brandt, J. (1999). Construct and concurrent validity of the Hopkins Verbal Learning Test–Revised. *The Clinical Neuropsychologist*, *13*, 348–358.

Smith, A. (1991). *Symbol Digit Modalities Test*. Los Angeles, CA: Western Psychological Services.

Strauss, E., Sherman, E. M., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary*. USA: Oxford University Press.

The Psychological Corporation. (1997). *WAIS-III-WMS-III technical manual*. San Antonio, TX: The Psychological Corporation.

Trenerry, M. R., Crosson, B., DeBoe, J., & Leber, W. R. (1989). *Stroop Neuropsychological Screening Test*. Odessa, FL: Psychological Assessment Resources.

Tsushima, W. T., Siu, A. M., Pearce, A. M., Zhang, G., & Oshiro, R. S. (2016). Two-year test-retest reliability of ImPACT in high school athletes. *Archives of Clinical Neuropsychology*, *31*, 105–111.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale-III (WAIS-III)*. New York: Psychological Corporation.

Zuckerman, S. L., Lee, Y. M., Odom, M. J., Solomon, G. S., & Sills, A. K. (2013). Baseline neurocognitive scores in athletes with attention deficit–spectrum disorders and/or learning disability: Clinical article. *Journal of Neurosurgery: Pediatrics*, *12*, 103–109.