OXFORD
UNIVERSITY PRESS

Archives of Clinical Neuropsychology 29 (2014) 432–441

Archives of CLINICAL NEUROPSYCHOLOGY

Detecting Cognitive Impairment After Concussion: Sensitivity of Change From Baseline and Normative Data Methods Using the CogSport/Axon Cognitive Test Battery

Andrea G. Louey[1], Jason A. Cromer[2,3], Adrian J. Schembri[2], David G. Darby[4], Paul Maruff[2,4,*], Michael Makdissi[4,5], Paul Mccrory[4]

[1]*Melbourne School of Psychological Sciences, The University of Melbourne, Parkville, VIC, Australia*
[2]*CogState Ltd., Melbourne, VIC, Australia*
[3]*Child Study Center, Yale University, New Haven, CT, USA*
[4]*The Florey Institute of Neuroscience and Mental Health, Melbourne Brain Centre - Austin Campus, Heidelberg, VIC, Australia*
[5]*Centre For Health Exercise and Sports Medicine, Melbourne Physiotherapy Department, The University of Melbourne, Parkville, VIC, Australia*

*Corresponding author at: CogState Ltd, Melbourne, VIC 3000, Australia. Tel.: +61-3-9664-1300; fax: +61-3-9664-1301.
E-mail address: pmaruff@cogstate.com (P. Maruff).

Accepted 13 April 2014

Abstract

Concussion-related cognitive impairments are typically evaluated with repeated neuropsychological assessments where post-injury performances are compared with pre-injury baseline data (baseline method). Many cases of concussions, however, are evaluated in the absence of baseline data by comparing post-injury performances with normative data (normative method). This study aimed to compare the sensitivity and specificity of these two methods using the CogSport/Axon test battery. Normative data and reliable change indices were computed from a non-injured athlete sample (n = 235). Test-retest data from non-injured (n = 260) and recently concussed (n = 29) athlete samples were then used to compare the two methods. The baseline method was found to be more sensitive than the normative method, and both methods had high specificity and overall correct classification rates. This suggests that while the normative method identifies most cases of recent concussions, the baseline method remains a more precise approach to assessing concussion-related cognitive impairments.

Keywords: Concussion; mTBI; Cognition; Assessment; Injury management; Neuropsychological evaluation

Introduction

Concussion is a concerning injury for athletes participating in contact and collision sports (Koh, Cassidy, & Watkinson, 2003; Lincoln et al., 2011; Meehan, d'Hemecourt, & Comstock, 2010). Symptoms and cognitive impairments resulting from a concussion occur in the absence of any reliable evidence of structural injury or pathophysiological changes, and typically resolve spontaneously within several days (Makdissi et al., 2010). Consequently, the current consensus for concussion management is that athletes who have suffered a concussion should remain out of play until they are physically well, symptom free, and cognitively normal (McCrory et al., 2013). Given this recommendation, neuropsychological assessment is an important part of concussion management systems where decisions about recovery are based on concussion-related cognitive impairment. Furthermore, such decisions are currently recommended to be based on findings that cognitive performance has declined from some pre-injury baseline (Davis, Iverson, Guskiewicz, Ptito, & Johnston, 2009; Ellemberg, Henry, Macciocchi, Guskiewicz, & Broglio, 2009).

While comparison of post-injury cognitive data with a pre-injury baseline (hereafter termed *baseline method*) is the recommended approach to detecting concussion-related cognitive impairment (e.g., Aubry et al., 2002; Harmon et al., 2013; McCrory et al., 2005, 2009), there is growing interest in the extent to which decisions about concussion-related cognitive impairment can be based on comparisons of post-injury cognitive data with normative data alone (hereafter termed *normative*

doi:10.1093/arclin/acu020 Advance Access publication on 9 May 2014

method). Initial objections to the normative method asserted that conventional normative data did not provide a valid estimate of pre-injury cognitive function in elite athletes (Brown, Guskiewicz, & Bleiberg, 2007; Lovell & Solomon, 2011; Solomon & Haase, 2008). Furthermore, even when normative data are drawn from elite athletes, it may still contain uncontrolled variability from factors such as cultural and social background, native language, history of prior head injury, and potential learning difficulties (e.g., Lezak, Howieson, Bigler, & Tranel, 2012; Strauss, Sherman, & Spreen, 2006). Additionally, defining post-concussion cognitive impairment based on normative data may miss those individuals whose preinjury cognitive function is superior to the general population. In these athletes, concussion-related cognitive impairment may diminish test performance substantially, albeit insufficiently to be considered abnormal. Each of these factors can reduce the sensitivity of a test to true concussion-related cognitive impairment (Salinas & Webbe, 2012; Solomon, Haase, & Kuhn, 2013). However, all of the potential limitations associated with the use of normative data to identify post-concussion cognitive impairment may be outweighed by the reality that obtaining preinjury baseline assessments from athletes is practically very difficult, especially in recreational, student, or geographically dispersed competitions (Echemendia et al., 2012; Moser, Schatz, Neidzwski, & Ott, 2011). Accordingly, there is an imperative to determine whether the normative method can be optimized for use in concussion management systems.

Recently, Echemendia and colleagues (2012), using the Immediate Post-concussion Assessment and Cognitive Testing (ImPACT) battery, and Schmidt, Register-Mihalik, Mihalik, Kerr, and Guskiewicz (2012), using the Automated Neuropsychological Assessment Metrics (ANAM) battery, compared directly the normative method with the baseline method for sensitivity to concussion-related cognitive impairment. Both concluded that the two methods provided equivalent sensitivity to concussion-related cognitive impairment on individual measures of cognitive function. These studies provide a good foundation for considering the relative merits of the baseline and normative methods for identifying concussion-related cognitive impairment, although limitations in the methodology of each study suggest that further research is necessary.

First, both studies reported surprisingly small rates (2–24%) of abnormal decline from baseline on the individual cognitive tasks in samples of concussed athletes, especially given that >90% of assessments in these studies were conducted within a week of injury. While this low rate of decline from baseline might be related to the sensitivity of the cognitive tests used in the respective batteries, a more important issue is that the statistical estimates of agreement between the two methods become unreliable when the relevant event (i.e., concussion-related cognitive impairment) is low (Bruckner, Yoder, & MacLean, 2006). With low event rates, statistical analyses of frequency also become biased toward accepting null hypotheses (Glaros & Kline, 1988), which actually occurred in both studies. To optimize the comparison of the baseline and normative methods, it is important that a high rate of true concussion-related cognitive impairment be present. One way to increase the rate of cognitive impairment detected in concussed athletes would be to study recently concussed athletes who are symptomatic at the time of cognitive assessment (Collie, Makdissi, Maruff, & McCrory, 2006; Iverson, Brooks, Collins, & Lovell, 2006; Iverson, Gaetz, Lovell, & Collins, 2004; Makdissi et al., 2010).

Secondly, neither study considered explicitly the potential of their criteria for impairment to yield false-positive classifications of concussion-related cognitive impairment by assessing a noninjured group of athletes. For example, the 80% confidence interval used to define cognitive impairment in the study by Schmidt and colleagues (2012) for both the baseline and normative methods would generate a false-positive classification in 20% of otherwise healthy athletes. The criterion for impairment used by Echemendia and colleagues (2012) for the baseline method was more conservative (i.e., yielding a 10% false-positive classification rate) than that used by Schmidt et al.; however, this criterion was also more conservative than that used by Echemendia et al. for their normative method, where cognitive impairment was defined as post-injury scores of 1 or 1.5 *SD* units below the normative group mean (i.e., yielding false-positive classification probabilities of 31.7% and 13.4%, respectively). Thus, the different criteria used in the analysis by Echemendia et al. between their baseline and normative methods may have resulted in their normative method having an increased sensitivity to concussion-related cognitive impairment, albeit with a higher probability of false-positive classification.

Thirdly, although both studies considered the classification rates of abnormal task performances, neither classified concussion-related cognitive impairment at the level of the individual athlete (i.e., based on performance on an entire cognitive test battery). Cognitive test batteries will provide greater sensitivity to impairment than individual tests, although the use of multiple tests to classify cognitive impairment also requires that the family-wise Type I error rate be controlled (Ingraham and Aiken, 1996). Thus, demonstrations of the differences between the baseline and normative methods using individual tests do not provide information about how the different methods will influence the classification of concussion-related cognitive impairment in individual athletes.

The aim of the present study was to compare the sensitivity and specificity of the baseline and normative methods to concussion-related cognitive impairment in athletes who had recently sustained a concussion and were symptomatic at the time of assessment using the CogSport/Axon test battery. The presence of concussion-related cognitive impairment in the concussion sample was confirmed by group comparison with a sample of well-matched noninjured controls. To determine the extent to which the baseline

and normative methods gave rise to false-positive classifications, they were also tested in a large group of noninjured elite athletes. Because the CogSport/Axon approach to the identification of concussion-related cognitive impairment has been validated using the baseline method (Collie, Maruff, Makdissi et al., 2003), this approach was used as the reference. To ensure estimates of sensitivity and specificity were comparable, the criteria for the classification of cognitive impairment was the same for the baseline and normative methods. Furthermore, the effects of the two different methods were compared for classification of abnormality on the four individual cognitive tasks, as well as for individual athletes. Finally, normative data for the study were derived from a second large sample of elite noninjured athletes. The hypothesis was that recently concussed and symptomatic athletes would show greater cognitive impairment compared with noninjured athletes, which would be evident for each cognitive task in the CogSport/Axon test battery, as well as across all tasks of the CogSport/Axon test battery for individual athletes. Once this was established, we then explored the extent to which concussion-related cognitive impairment could be identified in the same athletes using the normative method, and the sensitivity and specificity of the two methods were compared.

Method

Participants

Three samples of male elite collegiate or professional athletes participating in Australian Rules Football (AFL) or Rugby Football Union (RFU) club competitions were used in the current study. Matching between the three samples was ensured through the use of athletes from the same level of expertise, professional status, and participation rates for their respective sports. All participants spoke English and the CogSport/Axon test battery was administered in English. Informed consent was provided by all participants, and this study was approved by institutional research and ethics committees.

First, the normative sample consisted of 235 healthy male athletes aged between 18 and 37 years ($M = 25.41$, $SD = 4.49$), who were participating in AFL or RFU competitions. None of these athletes reported having sustained a previous concussion. These athletes were assessed twice with the CogSport/Axon test battery within a 7-day interval between test and retest assessments at the beginning of their competitive season. This brief retest interval was used to minimize any potential for true cognitive change to occur (e.g., due to physical injury or illness, stressful life events, sleep disturbance, etc.; Lezak et al., 2012), in order to provide accurate estimates of reliability and stability for each CogSport/Axon task to be used in the calculation RCIs in the baseline method (Falleti, Maruff, Collie, & Darby, 2006; Hopkins, 2000). Second, a different sample of 272 noninjured male athletes underwent baseline assessments with the CogSport/Axon test battery at the beginning of two consecutive AFL or RFU seasons (i.e., a 1-year test–retest interval). Data from 12 of these athletes were excluded from the analysis due to the incompletion of all four tasks in the CogSport/Axon test battery, leaving a total sample of 260 male athletes aged between 18 and 39 years ($M = 25.75$, $SD = 4.78$). All games and training sessions for these athletes were monitored as part of their participation in a concussion management program and none of the athletes in this sample were reported as having sustained a concussion during the year. Of the athletes within both the normative and noninjured samples, 20% were currently enrolled or had completed a university undergraduate degree, 67% had completed high school to year 12 level but had not continued to university, 13% had completed 4 or more years at high school and were currently enrolled or had completed vocational or trade school training, and 4% had completed 4 or more years at high school and had not proceeded to vocational training. In terms of ethnicity, 9% of the normative sample identified themselves as being of indigenous (Australia, New Zealand, or Pacific Islands) heritage, with the remainder of the sample identifying themselves as being of English or European heritage.

The third sample was the concussion sample, which comprised 29 male athletes aged between 16 and 26 years ($M = 22.28$, $SD = 2.90$), who had suffered a concussion during their participation in an AFL or RFU competition. Of the concussion sample, 20% were currently enrolled or had completed a university undergraduate degree, 66% had complete high school to year 12 level but had not continued to university, and 14% had completed 4 or more years at high school and were currently enrolled or had completed vocational or trade school training. One of the athletes in this sample identified themselves as being of New Zealand indigenous heritage, with the remainder identifying themselves as being of English or European heritage. The team doctor for each AFL or RFU club who was present at the time of injury made the diagnosis of concussion in each case according to standard injury definitions and recommendations of the consensus statements from the 3rd and 4th International Conferences on Concussion in Sport (Collie et al., 2006; Makdissi et al., 2010; McCrory et al., 2009, 2013) where the diagnosis of concussion was not based on any score or rating scale alone, but with the consideration of factors across several clinical domains. Criteria contributing to the identification of concussion included symptoms reported by athletes or signs observed by medical staff after the injury. Symptoms that were reported by the concussed athletes in this sample included (but were not limited to) feeling dinged, dazed, stunned, woozy, foggy, “head full of cotton wool,” or “not quite right.” The concussed athletes in this sample also endorsed symptoms of post-traumatic headache, visual disturbance, confusion, memory disturbance, balance disturbance, vertigo, and light headedness. Signs that were observed by medical staff include confusion, loss of consciousness, disorientation, memory

disturbance, unsteadiness, attention deficit, and personality change. All athletes were symptomatic at the time of the post-concussion assessment, and were reporting between one and six total symptoms of concussion ($M = 3.66$, $SD = 1.11$). The average time after injury of the initial post-concussion cognitive assessment was 32.10 h ($SD = 4.06$ h; range [26, 42]).

Measures

Cognitive function was measured in all athletes using the CogSport/Axon test battery, which is a 10–15 min computerized test battery consisting of four cognitive tasks (Detection, Identification, One-Back, and One Card Learning) involving playing cards. Briefly, the Detection task was a reaction time test, in which athletes attended to the rule, “Has the card turned over?”; the Identification task was a choice reaction time test with the rule, “Is the card red?”; the One-Back task was a working memory test with the rule, “Is the previous card the same?”; and finally, the One Card Learning task was a learning test with the rule, “Have you seen this card before in this task?” The rationale, method of administration, and scoring of the tasks in the CogSport/Axon test battery have been described in detail elsewhere (Collie, Maruff, Darby et al., 2003; Falleti et al., 2006).

For each cognitive task, performance measures included the accuracy and speed of performance. Only the primary performance measures used in concussion management settings were analyzed for the purposes of the present study, as these are the variables that are reported to clinicians. Primary performance measures typically used to assess significant post-concussion cognitive decline from baseline include speed for the Detection, Identification, and One-Back tasks; and accuracy for the One Card Learning task.

In addition to the assessment of cognitive function, athletes in the concussion sample also completed a concussion symptom questionnaire prior to the test battery (Collie et al., 2006). Symptoms of concussion that were canvassed by the questionnaire include headache, confusion, visual disturbance, vertigo, amnesia, fatigue, fogginess, and sleep difficulties.

Data Analysis

The data for the normative, noninjured, and concussion samples were analyzed using the Statistical Package for the Social Sciences (SPSS; SPSS, Inc., Chicago, IL) for Windows, version 18.0. The data analysis proceeded in four stages following preliminary data analyses for each sample encompassing descriptive statistic calculations.

First, to allow for the computation of the reliable change indices (RCIs) necessary for the baseline method, and the cross-sectional means and standard deviations necessary for the normative method; data from the normative sample were used to calculate normative descriptive statistics for the primary performance measures as well as estimates of the test–retest reliability (intraclass correlations, ICC) and stability (within-subject standard deviations, WSD) for each of the four CogSport/Axon tasks. Age was not treated as a covariate in generating the normative data, as preliminary analyses (not reported presently) of the normative sample revealed no correlation between age and performance on any of the four CogSport/Axon tasks. As such, the normative data for each task were generated by collapsing performance estimates across the entire normative cohort (i.e., $n = 235$) to provide the most reliable estimate of population parameters possible.

Secondly, the nature and magnitude of cognitive change from baseline at the retest assessment (i.e., repeat baseline in noninjured athletes and post-injury assessment in concussed athletes) was examined at the group level for each task by comparing performance between the noninjured and concussion samples using a series of group × time analyses of covariance (ANCOVA), in which performance at the retest assessment was entered as the dependent variable and performance at the baseline assessment was treated as the covariate. For each measure, Cohen’s d was computed to express the magnitude of the group difference in the baseline adjusted means.

Thirdly, the frequency of classification of abnormal performance on each of the four CogSport/Axon tasks was computed using the baseline and normative methods. Concussion-related cognitive impairment in task performances for each participant was determined using a one-tailed 95% confidence (i.e., below the fifth percentile) or 1.65 SD criterion for both comparison methods. The one-tailed 95% confidence criterion was selected because this is in accord with the statistical convention of accepting a 5% false-positive rate of classification (i.e., $p < .05$, one tailed). For the baseline method, the performance of each athlete at the retest (i.e., repeat baseline or post-injury) assessment was compared with their own performance at baseline assessment using a RCI for each task. The Crawford and Howell (1998) RCI was used to express the magnitude of change in performance on each task because it corrects for any regression to the mean, as well as potential practice effects on an individualized basis (see Hinton-Bayre, 2010, 2012, for discussion). As RCIs are a standardized change score with a mean of zero and a SD of 1, a less than −1.65 SD cut-score was used as the criterion for classifying meaningful decline from baseline (Hinton-Bayre, 2010). For the normative method, each athlete’s performance on each task at retest assessment was compared with the mean and SD of performance computed from the normative sample in the first stage of analysis. From this comparison, a classification of impairment was then made when performance on a task was less than −1.65 SD units from the mean of the normative group for that same task. Chi-squared

analyses were then conducted to compare the classification rates of abnormal task performances by the baseline method and the normative method within the concussion sample and the noninjured sample.

Fourthly, the extent to which cognitive impairment could be classified in individual athletes on the basis of abnormal task performance across the CogSport/Axon test battery was analyzed. As there were four performance measures (one for each of the four tasks) and abnormality was classified when task performance was less than −1.65 *SD* units, the risk of family-wise error was controlled by requiring abnormal task performances on at least two of the four measures (Ingraham & Aiken, 1996). With the frequency of abnormal task performance computed for both the baseline method and the normative method, chi-squared analyses were then conducted to compare the frequency of classifications in the noninjured and concussion samples by the two methods.

Results

Normative Statistics, Stability, and Reliability of CogSport/Axon Tasks

Descriptive statistics for the stability and reliability of performance on each CogSport/Axon task in the normative sample are summarized in Table 1. There were no significant differences between athletes' performances at the initial test and the 1-week retest assessments on each of the tasks. Test–retest reliability (ICC) for each measure was very high, and the WSD values were low.

Nature and Magnitude of Concussion-Related Cognitive Change

Table 2 summarizes the test (i.e., baseline) and retest (i.e., repeat baseline or post-injury) means and *SD* for both the noninjured and concussion samples. The ANCOVAs indicated that after taking into account baseline performance, significant performance differences occurred between the noninjured and concussion samples on all four CogSport/Axon tasks (see Table 3) and the magnitude of these were by convention, large (e.g., Cohen, 1992).

Identifying Abnormal Performance on Individual Tasks

Table 4 shows the number of classifications of abnormal task performance arising from the baseline method and the normative method for each CogSport/Axon task. Within the concussion sample, the frequency of abnormal task performances classified by the baseline method was significantly higher than that classified by the normative method for only the detection task, χ^2 (1) = 6.03, $p = .014$. No significant differences in the frequency of classification of abnormal task performance between the baseline and normative methods were observed for the Identification, χ^2 (1) = 2.48, $p = .115$; One Card Learning, χ^2 (1) = 0.07, $p = .791$; or One-Back, χ^2 (1) = 0.00, $p = 1.000$, tasks. Within the noninjured sample, the frequency of abnormal task performances classified by the baseline method was significantly higher than those classified by the normative method for the Detection, χ^2 (1) = 15.62, $p < .001$, Identification, χ^2 (1) = 8.62, $p = .003$, and One Card Learning tasks, χ^2 (1) = 26.51, $p < .001$, but not the One-Back task, χ^2 (1) = 0.12, $p = .730$.

Identifying Individual Cases of Concussion-Related Cognitive Impairment

Table 5 shows the classification of cognitive impairment in athletes on the basis of the detection of declined or abnormal task performances on increasing numbers of the CogSport/Axon tasks. With the criterion for cognitive impairment requiring decline on two or more tasks, the baseline method classified cognitive impairment in 96.6% of concussed athletes. For the same criterion, the

Table 1. Normative data, test–retest stability and reliability of CogSport/Axon tasks for computation of RCI

	Baseline		Repeat baseline		Difference					
Task	*M* (*SD*)	CV	*M* (*SD*)	CV	*M* (*SD*)	WSD	ICC	*t*	*p*	*d* [95% CI]
Detection	2.43 (0.05)	2.13	2.42 (0.05)	2.22	0.00 (0.04)	0.03	0.85*	1.84	.07	0.20 [0.02, 0.38]
Identification	2.63 (0.05)	1.95	2.62 (0.06)	2.21	0.00 (0.04)	0.03	0.86*	1.26	.21	0.17 [0.00, 0.36]
One Card Learning	0.94 (0.14)	14.41	0.94 (0.14)	15.10	0.00 (0.07)	0.05	0.93*	0.47	.64	0.00 [−0.18, 0.18]
One-Back	2.74 (0.07)	2.72	2.74 (0.08)	2.90	0.00 (0.06)	0.04	0.83*	0.48	.64	0.00 [−0.18, 0.18]

Note: CV = coefficient of variation, WSD = within-subject standard deviation, ICC = intraclass correlation coefficient, CI = confidence interval. Estimates were derived from the normative sample data ($n = 235$).
*$p < .001$.

Table 2. Test and retest means and standard deviations for the CogSport/Axon tasks

Task	Noninjured ($n = 260$)		Concussion ($n = 29$)	
	Baseline	Repeat baseline	Baseline	Post-injury
Detection	2.43 (0.05)	2.43 (0.06)	2.42 (0.10)	2.59 (0.17)
Identification	2.63 (0.05)	2.62 (0.06)	2.65 (0.09)	2.79 (0.15)
One Card Learning	0.87 (0.15)	0.90 (0.15)	0.81 (0.19)	0.74 (0.16)
One-Back	2.75 (0.08)	2.74 (0.08)	2.82 (0.13)	2.91 (0.14)

Note: Values are mean (*SD*). The nature and magnitude of cognitive change from baseline was significantly different between the noninjured and concussion groups for all four tasks; the results of these contrasts are reported in Table 3.

Table 3. Contrasts of noninjured and concussion covariate-adjusted means across the four CogSport/Axon tasks

Task	Noninjured, Adj. *M* (*SD*)	Concussion, Adj. *M* (*SD*)	*F*	*P*	*d* [95% CI]
Detection	2.43 (0.06)	2.61 (0.07)	164.95	<.001	−2.95 [−3.39, −2.49]
Identification	2.63 (0.06)	2.78 (0.06)	166.67	<.001	−2.50 [−2.93, −2.06]
One Card Learning	0.90 (0.15)	0.76 (0.14)	26.26	<.001	−0.94 [−1.33, −0.55]
One−Back	2.74 (0.06)	2.87 (0.07)	85.83	<.001	−2.13 [−2.54, −1.70]

Note: CI = confidence interval. Cohen's *d* effect size and confidence intervals have been reversed for accuracy on the One Card Learning task so that negative values indicate a decline in performances.

Table 4. Percentage of cases identified as having cognitive impairment at retest using a 1.65 standard deviation cut-score criterion within each CogSport/Axon task

Task	Cut-score	Noninjured		Concussion	
		% (*n*)	[95% CI]	% (*n*)	[95% CI]
Baseline method[a]					
Detection	<−1.65	17.3 (45)	[12.7, 21.9]	89.7 (26)	[78.6, 100.7]
Identification	<−1.65	16.5 (43)	[12.0, 21.2]	86.2 (25)	[73.7, 98.8]
One Card Learning	<−1.65	24.2 (63)	[19.0, 29.4]	44.8 (13)	[26.7, 62.9]
One-Back	<−1.65	7.3 (19)	[4.1, 10.5]	58.6 (17)	[40.7, 76.5]
Normative method[b]					
Detection	>2.51[b]	6.2 (16)	[3.2, 9.1]	62.1 (18)	[44.4, 79.7]
Identification	>2.71[b]	8.1 (21)	[4.8, 11.4]	69.0 (20)	[52.1, 85.8]
One Card Learning	<0.71[b]	7.7 (20)	[4.5, 10.9]	41.4 (12)	[23.5, 59.3]
One-Back	>2.86[b]	6.5 (17)	[3.5, 9.5]	58.6 (17)	[40.7, 76.5]

Note: CI = confidence interval.
[a]Baseline comparisons were based on Crawford and Howell's (1998) calculation for reliable change using statistical parameters from the matched normative baseline data. As performance decline on speed (i.e., reaction time) tasks is represented by an increased score between test and retest assessments, the baseline comparisons for Detection, Identification, and One-Back tasks were reversed; negative values reflect a decline in performance.
[b]Cut-scores for normative data comparisons were calculated on the basis of normative baseline data ($M + 1.65\ SD$ for Detection, Identification, and One-Back; $M - 1.65SD$ for One Card Learning).

normative method classified significantly less cognitive impairment in the same sample, 69.0%, $\chi^2\,(1, N = 29) = 7.73, p = .005$. Importantly, even if the criterion for cognitive impairment was reduced to requiring abnormal performance on only one task, the baseline method remained sensitive to all cases of cognitive impairment in concussed and symptomatic athletes, while the normative method missed 17.2% of cases. The rate of false-positive classifications in the noninjured athlete sample using the same criterion of decline on two or more tasks was low for the two methods, and this was not significantly different between the baseline and normative methods, $\chi^2\,(1, N = 260) = 2.88, p = .090$. Thus, when given in terms of diagnostic accuracy, the criterion of 1.65 *SD* units below the mean on two or more tasks (of four) for the baseline method yielded a higher sensitivity and equal specificity to the normative method (see Table 6). Overall correct classification rates were not different between the baseline and normative methods, $\chi^2\,(1, N = 289) = 0.27, p = .601$.

Discussion

The hypothesis that recently concussed athletes would show greater cognitive impairment when compared with noninjured athletes on the basis of individual cognitive tasks and change in performance across the CogSport/Axon test battery within

Table 5. Percentage of cases identified as having cognitive impairment at retest using a 1.65 *SD* cut-score criterion on the basis of number of CogSport/Axon tasks meeting this criterion

Sample	No. of abnormal task scores, % (*n*)					Total *n*
	0	≥1	≥2	≥3	4	
Baseline method						
Noninjured	50.8 (132)	49.2 (128)	13.1 (34)	3.1 (8)	0.0 (0)	260
Concussion	0.0 (0)	100.0 (29)	96.6 (28)	65.5 (19)	17.2 (5)	29
Normative method						
Noninjured	81.9 (213)	18.1 (47)	8.5 (22)	1.9 (5)	0.0 (0)	260
Concussion	17.2 (5)	82.8 (24)	69.0 (20)	55.2 (16)	24.1 (7)	29

Table 6. Diagnostic classification accuracy of the baseline method and the normative method when two or more abnormal task scores are required on the CogSport/Axon test battery

	Sensitivity [95% CI]	Specificity [95% CI]	CCR [95% CI]
Baseline method	96.6 [82.8, 99.4]	86.9 [82.3, 90.5]	87.9 [84.1, 91.7]
Normative method	69.0 [50.8, 82.7]	91.5 [87.5, 94.3]	89.3 [85.7, 92.8]

Note: CI = confidence interval, CCR = correct classification rate.

individual athletes was supported. Taking into account preinjury cognitive function and comparing performance to a large group of healthy noninjured athletes, all aspects of cognition in the concussed athlete group declined significantly (see Table 3). The magnitude of decline (*d*) for each measure was, by convention, large (e.g., Cohen, 1992), but was greatest for psychomotor function (Detection task) and attention (Identification task). These data show that large impairment in psychomotor function, attention, working memory (One-Back task), and learning (One Card Learning task) is present in concussed and symptomatic athletes. This observation is consistent with the large cognitive impairment observed in previous studies of groups of concussed and symptomatic athletes (e.g., Collie et al., 2006; Iverson et al., 2004; Makdissi et al., 2010).

When considered for the individual cognitive tasks, the frequency of classification of abnormal performance in the concussed athletes was greater than that for the noninjured athletes on all CogSport/Axon tasks using the baseline method. The frequency of abnormal performances ranged between ∼45% and 90% across the four CogSport/Axon tasks. In contrast, abnormal performance was classified in less than 24.2% of the noninjured athletes on the same tasks (see Table 4). Finally, when cognitive impairment was classified for individual athletes on the basis of abnormal performance on two or more tasks across the complete CogSport/Axon test battery, over 96% of the concussed athletes met the criteria for cognitive impairment, while only 13% of the noninjured athletes were classified with cognitive impairment (see Table 5). These data, together with the concussion diagnosis by the team doctor and the athletes' self-report of concussion symptoms, suggest that the recently concussed athletes had meaningful cognitive impairment; the magnitude of which is greatest in the domains of psychomotor function and attention. Furthermore, classification of this impairment in individual athletes is most accurate—with the lowest levels of false-positive classification—when the post-concussion performance is considered across the entire CogSport/Axon test battery.

Compared with the baseline method, the rate of abnormal performance on individual tasks classified by the normative method in the concussed athletes was highly similar when measured with the CogSport/Axon test battery. Only the measure of psychomotor function showed significantly greater sensitivity to impairment under the baseline method than under the normative method. Equivalence in sensitivity to abnormal performance on individual cognitive tasks from the ImPACT and ANAM batteries between the baseline and normative methods was also observed by Echemendia and colleagues (2012) and Schmidt and colleagues (2012), albeit in concussed athletes that were asymptomatic at the time of testing. However, the very low rates of cognitive impairment identified using the baseline method within these concussed athlete samples is likely to have biased statistical tests to find no differences between the two methods (Bruckner et al., 2006). On the other hand, in the current sample and using the CogSport/Axon test battery, the baseline method did yield significantly greater rates of false-positive classifications in the noninjured sample than did the normative method for three of the individual tasks, suggesting that individual CogSport/Axon tasks may overestimate the rate of cognitive impairment in athletes. This supports the use of multiple tasks with adjustment for family-wise Type I error rate (e.g., impairment on two or more tasks) for identification of concussion-related cognitive impairment.

When classified on the basis of performance across all four tasks in the CogSport/Axon test battery, the rate of cognitive impairment in concussed athletes was substantially lower for the normative method than for the baseline method (see Table 6). Not only was the sensitivity of the baseline method greater than that of the normative method, but most importantly, 27.6% of the concussed athletes who had been classified as impaired using the baseline method were classified as unimpaired using the normative

method. Even if the criteria for impairment was reduced from requiring two abnormal test performances to one abnormal test performance, the normative method still classified 17.2% of athletes who were diagnosed with a concussion and were symptomatic at testing as being normal (see Table 5). Thus, whilst data from the analysis of classification rates in concussed athletes on the basis of individual tasks were equivalent on all tasks except Detection, the baseline method was far more sensitive than the normative method when performances on all CogSport/Axon tasks were considered simultaneously. The increased sensitivity of the baseline method is most likely due its ability to control interindividual differences that can increase estimates of variability in normative data, even when that normative data are derived from a well-matched group of noninjured elite athletes. The current normative group illustrates this with the within-subject variances being much smaller than the between-subject variances for performance on the same cognitive tasks (see Table 1). Applying the same criteria for cognitive impairment to the noninjured athletes tested with the CogSport/Axon test battery also showed the baseline and normative methods to both yield low and similar rates of false-positive classifications. This observation is consistent with the findings of a previous study using the ANAM battery that reported equivalent rates of false-positive classification rates between the baseline and normative methods using composite scores when equivalent criteria for impairment were used for both methods (Roebuck-Spencer, Vincent, Schlegel, & Gilliland, 2013). Thus, the strategy used to control family-wise error rates for decision-making on the basis of all four tasks in the CogSport/Axon test battery (e.g., Ingraham & Aiken, 1996) successfully limited the probability of false-positive classification of cognitive impairment based on a single task in the concussed athletes. Finally, the overall correct classification rates were similar between the baseline and normative methods in the current sample (see Table 6). For the normative method, the correct classification rate was similar to the rates observed in an independent study that used performance on the CogSport/Axon test battery to compare the rates of abnormal performance in noninjured and concussed athletes (Gardner, Shores, Batchelor, & Honan, 2012). Unfortunately, this study did not measure performance at baseline in the concussed athletes.

Considered together, the current results obtained with the CogSport/Axon test battery provide evidence that the baseline method for concussion management that is recommended by consensus guidelines (Aubry et al., 2002; McCrory et al., 2005, 2009) has both a high sensitivity and specificity for concussion-related cognitive impairment. Furthermore, the finding that the baseline method is more sensitive than the normative method suggests that it is the optimal approach to concussion management for the detection of concussion-related cognitive impairment. The baseline method offers an individualized approach, which means that it inherently controls for the many factors (e.g., age, previous history of concussion, learning disabilities) that can contribute to the increased variance that is associated with interindividual comparisons, such as the normative method. The current study did find, however, that while the sensitivity of the normative method was lower than that for the baseline method when considered at the level of the athlete and across cognitive tasks, the normative method did have moderate sensitivity and high specificity to concussion-related cognitive impairment.

In this respect, one potential limitation of the current study may be that the criterion for cognitive impairment (performance less than −1.65 *SD* units on two or more tests) was too stringent for the use of the normative method. This criterion is used currently to classify the presence of meaningful cognitive decline from baseline within established concussion management systems that use the CogSport/Axon test battery, where its validity as a measure of clinically relevant change is accepted (e.g., Makdissi et al., 2010). However, subtle cognitive impairment has also previously been classified in neuropsychological contexts on the basis of a less than 1 *SD* criterion below the performance of a well-matched control group (Lezak et al., 2012). The use of a less stringent criterion will, of course, increase the rate at which nonimpaired athletes will be classified as being impaired, and therefore withheld from returning to play. However, in concussion management systems that use the normative method for classifying cognitive impairment (e.g., where athletes who have had a concussion are likely to present with no pre-injury baseline), clinicians may tolerate the reduced specificity. This less stringent criterion will bias decisions toward keeping all athletes, those with and without true cognitive impairment, from returning to play. Such a strategy might be especially useful in nonelite and recreational sports settings, where the consequences of being withheld incorrectly from competition after a concussion are less than in elite sports. However, the current data suggest that more work is required before the normative method can be considered appropriate for classifying concussion-related cognitive impairment as we observed that normative comparisons failed to detect cognitive impairment in 31% of the concussed and symptomatic athletes.

Several aspects of the current study may limit the generalizability of the conclusions. First, the current samples consisted only of elite adult male athletes. We must now determine whether the baseline and normative methods would show the relative sensitivities and specificities observed here within other populations at risk of concussive injuries (e.g., recreational athletes, female athletes, paediatric athletes, military personnel). In addition, this study used only the CogSport/Axon test battery to compare the baseline and normative methods. As concussion management systems often employ other computerized cognitive assessment tools (e.g., ImPACT, ANAM), it will be necessary to investigate the sensitivity of the baseline and normative methods using these other test batteries. However, we propose that provided such studies use the statistical and methodological approaches described in the current study (i.e., ensure a high base rate of true concussion-related cognitive impairment, use an appropriately matched noninjured control group, use the same criteria used for classifying cognitive impairment across both the baseline and

normative methods, control Type I error associated with multiple tests, and analyze outcomes for individual athletes as well as groups), the baseline method will always provide greater sensitivity to concussion-related cognitive impairment than the normative method, irrespective of the cognitive test battery used. Importantly, as cognitive testing is typically conducted following the self-report of symptom resolution by concussed athletes, it will be necessary to assess the utility of both methods using cognitive test batteries within concussed athletes who are asymptomatic at post-injury assessment; although, given the rate of true concussion-related cognitive impairment in asymptomatic athletes is likely to be much lower than that which is observed for symptomatic athletes, much larger samples will be necessary to test the equivalence of the baseline and normative methods. In addition, further investigation using samples that are more representative of the populations typically assessed for concussion-related cognitive impairment will also provide important insights into the probability that the baseline and normative methods will give correct diagnoses in clinical settings (i.e., predictive values). Furthermore, the test–retest reliability estimates and normative data were computed from a separate sample of athletes who underwent two assessments across a 1 week test–retest interval in this study. This may be considered by some as not long enough to accurately represent a clinically relevant time frame in sports concussion (Broglio, Ferrara, Macciocchi, Baumgartner, & Elliott, 2007). In addition, Crawford and Howell (1998) recommend the use of a test–retest reliability estimate corresponding to the relevant test–retest interval for the calculation of RCIs. Despite this, the reliability coefficients used in the present study was based on the 1-week test–retest interval of the normative sample rather than the 1-year interval of the nonconcussed athlete sample, as the shorter interval is more likely to reflect only measurement error. It is therefore expected that the calculation of RCIs using nonconcussed athlete reliability coefficients will result in slight decreases in the sensitivity of the baseline method, as the longer test–retest interval will more likely reflect true variability in addition to measurement error.

This study aimed to compare the normative and baseline methods of classifying concussion-related cognitive impairment in a group of recently concussed athletes who were symptomatic at the time of assessment using the CogSport/Axon test battery. The findings of this study suggest that in concussion management systems using the CogSport/Axon test battery, the normative method can be used to identify the majority of elite male athletes who have recently sustained a concussive injury and are symptomatic at the time of assessment. On the other hand, the baseline method appears to improve the ability of clinicians to detect concussion-related cognitive impairments and it therefore remains as the optimal approach to concussion management. In the absence of baseline data, however, clinicians must be mindful of the sensitivity and specificity issues associated with selecting a criterion for classifying concussion-related cognitive impairment when using the normative method.

Conflict of interest

Paul Maruff and Adrian Schembri are full time employees of CogState the company who developed and provides the Axon/CogSport concussion management system.

Paul McCrory has received consultancy fees from CogState.

References

Aubry, M., Cantu, R., Dvorak, J., Graf-Baumann, T., Johnston, K., Kelly, J., et al. (2002). Summary and agreement statement of the first international conference on concussion in sport, Vienna 2001. *British Journal of Sports Medicine*, *36*, 6–10.

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

Brown, C. N., Guskiewicz, K. M., & Bleiberg, J. (2007). Athlete characteristics and outcome scores for computerized neuropsychological assessment: A preliminary analysis. *Journal of Athletic Training*, *42*, 515–523.

Bruckner, C. T., Yoder, P., & MacLean, W. E. (2006). Interpreting kappa in observational research: Base rate matters. *American Journal on Mental Retardation*, *111*, 433–441.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.

Collie, A., Makdissi, M., Maruff, P., & McCrory, P. (2006). Cognition in the days following concussion: Comparison of symptomatic versus asymptomatic athletes. *Journal of Neurology, Neurosurgery & Psychiatry*, *77*, 241–245.

Collie, A., Maruff, P., Darby, D. G., & McStephen, M. (2003). The effects of practice on the cognitive test performance of neurologically normal individuals assessed at brief test-retest intervals. *Journal of the International Neuropsychological Society*, *9*, 419–428.

Collie, A., Maruff, P., Makdissi, M., McCrory, P., McStephen, M., & Darby, D. (2003). CogSport: Reliability and correlation with conventional cognitive tests used in postconcussion medical evaluations. *Clinical Journal of Sport Medicine*, *13*, 28–32.

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.

Davis, G. A., Iverson, G. L., Guskiewicz, K. M., Ptito, A., & Johnston, K. M. (2009). Contributions of neuroimaging, balance testing, electrophysiology and blood markers to the assessment of sport-related concussion [Supplement]. *British Journal of Sports Medicine*, *43*, i36–i45.

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

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.

Falleti, M. G., Maruff, P., Collie, A., & Darby, D. G. (2006). Practice effects associated with the repeated assessment of cognitive function using the CogState battery at 10-minute, one week and one month test-retest intervals. *Journal of Clinical and Experimental Neuropsychology*, *28*, 1095–1112.

Gardner, A., Shores, E. A., Batchelor, J., & Honan, C. A. (2012). Diagnostic efficiency of ImPACT and CogSport in concussed rugby union players who have not undergone baseline neurocognitive testing. *Applied Neuropsychology: Adult*, *19*, 90–97.

Glaros, A. G., & Kline, R. B. (1988). Understanding the accuracy of tests with cutting scores: The sensitivity, specificity, and predictive value model. *Journal of Clinical Psychology*, *44*, 1013–1023.

Harmon, K. G., Dreznar, J. A., Gammons, M., Guskiewicz, K. M., Halstead, M., Herring, S. A., et al. (2013). American Medical Society for Sports Medicine position statement: Concussion in sport. *British Journal of Sports Medicine*, *47*, 15–26.

Hinton-Bayre, A. D. (2010). Deriving reliable change statistics from test-retest normative data: Comparison of models and mathematical expressions. *Archives of Clinical Neuropsychology*, *25*, 244–256.

Hinton-Bayre, A. D. (2012). Choice of reliable change model can alter decisions regarding neuropsychological impairment after sports-related concussion. *Clinical Journal of Sport Medicine*, *22*, 105–108.

Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. *Sports Medicine*, *30*, 1–15.

Ingraham, L. J., & Aiken, C. B. (1996). An empirical approach to determining criteria for abnormality in test batteries with multiple measures. *Neuropsychology*, *10*, 120–124.

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., Gaetz, M., Lovell, M. R., & Collins, M. W. (2004). Relation between subjective fogginess and neuropsychological testing following concussion. *Journal of the International Neuropsychological Society*, *10*, 904–906.

Koh, J. O., Cassidy, J. D., & Watkinson, E. J. (2003). Incidence of concussion in contact sports: A review of the evidence. *Brain Injury*, *17*, 901–917.

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). Neuropsychological assessment *(5th ed.)*. New York, NY: Oxford University Press.

Lincoln, A. E., Caswell, S. V., Almquist, J. L., Dunn, R. E., Norris, J. B., & Hinton, R. Y. (2011). Trends in concussion incidence in high school sports: A prospective 11-year study. *The American Journal of Sports Medicine*, *39*, 958–963.

Lovell, M. R., & Solomon, G. S. (2011). Psychometric data for the NFL neuropsychological test battery. *Applied Neuropsychology*, *18*, 197–209.

Makdissi, M., Darby, D., Maruff, P., Ugoni, A., Brukner, P., & McCrory, P. (2010). Natural history of concussion in sport: Markers in severity and implications for management. *The American Journal of Sports Medicine*, *38*, 464–471.

McCrory, P., Johnston, K., Meeuwisse, W., Aubry, M., Cantu, R., Dvorak, J., et al. (2005). Summary and agreement statement of the 2nd international conference on concussion in sport, Prague 2004. *Clinical Journal of Sport Medicine*, *15*, 48–55.

McCrory, P., Meeuwisse, W., Johnston, K., Dvorak, J., Aubry, M., Molloy, M., et al. (2009). Consensus statement on concussion in sport—The 3rd international conference on concussion in sport, held in Zurich, November 2008. *Journal of Clinical Neuroscience*, *16*, 755–763.

McCrory, P., Meeuwisse, W. H., Aubry, M., Cantu, B., Dvořák, J., Echemendia, R. J., et al. (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.

Meehan, W. P., III, d'Hemecourt, P., & Comstock, R. D. (2010). High school concussions in the 2008–2009 academic year: Mechanism, symptoms, and management. *The American Journal of Sports Medicine*, *38*, 2405–2409.

Moser, R. S., Schatz, P., Neidzwski, K., & Ott, S. D. (2011). Group versus individual administration affects baseline neurocognitive test performance. *American Journal of Sports Medicine*, *39*, 2325–2330.

Roebuck-Spencer, T. M., Vincent, A. S., Schlegel, R. E., & Gilliland, K. (2013). Evidence for added value of baseline testing in computer-based cognitive assessment. *Journal of Athletic Training*, *48*, 499–505.

Salinas, C. M., & Webbe, F. M. (2012). Sports neuropsychology with diverse athlete populations: Contemporary findings and special considerations. *Journal of Clinical Sports Psychology*, *6*, 363–384.

Schmidt, J. D., Register-Mihalik, J. K., Mihalik, J. P., Kerr, Z. Y., & Guskiewicz, K. M. (2012). Identifying impairments after concussion: Normative data versus individualized baselines. *Medicine and Science in Sports and Exercise*, *44*, 1621–1628.

Solomon, G. S., & Haase, R. F. (2008). Biopsychosocial characteristics and neurocognitive test performance in National Football League players: An initial assessment. *Archives of Clinical Neuropsychology*, *23*, 563–577.

Solomon, G. S., Haase, R. F., & Kuhn, A. (2013). The relationship among neurocognitive performances and biopsychosocial characteristics of elite national football league draft picks: An exploratory investigation. *Archives of Clinical Neuropsychology*, *28*, 9–20.

Strauss, E. H., Sherman, E. M. S., & Spreen, O. (2006). Norms selection in neuropsychological assessment. In *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed., pp. 44–58). New York, NY: Oxford University Press.