Journal of the International Neuropsychological Society (2001), **7**, 693–702.

Sensitivity and specificity of standardized neurocognitive testing immediately following sports concussion

WILLIAM B. BARR[1] AND MICHAEL McCREA[2,3]

[1]Departments of Neurology and Psychiatry, New York University Medical Center, New York, NY
[2]Neuropsychology Service, Waukesha Memorial Hospital, Waukesha, WI
[3]Department of Neurology, Medical College of Wisconsin, Milwaukee, WI

(Received April 4, 2000; Revised August 25, 2000; Accepted August 29, 2000)

Abstract

Neuropsychology, with its emphasis on standardized and empirically based methods, has made a number of scientific contributions to address growing concerns about concussions resulting from sports injuries. This study employs a test–retest paradigm to determine the immediate effects of concussion in high-school and college athletes. The Standardized Assessment of Concussion (SAC) was administered to 1,313 male athletes prior to the beginning of the competitive season. Reliable change indices and multiple regression models were computed on retest scores obtained from 68 noninjured athletes who were readministered the SAC at either 60 or 120 days following baseline testing. Receiver operating characteristic (ROC) curve analyses were used to test these models with data obtained on 50 athletes tested immediately following concussion. The results indicate that a decline of 1 point on the SAC at retesting classified injured and noninjured participants with a level of 94% sensitivity and 76% specificity. The RCI and multiple regression models provided comparable levels of group classification, but provided cut-offs that are conservative for use with this population. The results support and extend previous research findings indicating that the SAC is a valid instrument for detecting the immediate effects of mild traumatic brain injury. (*JINS*, 2001, *7*, 693–702.)

Keywords: Sports injuries, Concussion, Test–retest reliability, Traumatic brain injury

INTRODUCTION

Objective, quantitative measurement of neurocognitive functioning is considered perhaps the defining skill that sets the neuropsychologist apart from other clinicians and researchers in the neurosciences. In clinical settings, neuropsychological testing is now widely recognized as a sensitive and sophisticated means of detecting and characterizing neurocognitive impairment resulting from central nervous system trauma or disease (Levin, 1994). Research initiatives involving neuropsychological measures of cognition have also greatly benefited the neurosciences in advancing our understanding of the relation between brain and behavior (Keefe, 1995). In essence, neuropsychological testing makes more objective and quantifiable what other clinicians commonly report as a vague and subjective impression of neurocognitive abnormalities displayed by their patients.

Reliability and validity are the cornerstones of all forms of measurement, including neuropsychological testing. Interpretation of the neuropsychologist's test findings is based on the premise that a measure is affected minimally by measurement error or random influence, and that the measure can be used to support a specific inference. In basic terms, the psychometric properties of a given instrument impact directly on the neuropsychologist's ability to trust that the test result represents an accurate assessment of the neurocognitive construct in question. Similarly, the concepts of sensitivity and specificity provide an empirical basis for evaluating the use of neuropsychological testing in diagnosing neurologic disorders. The value of a neuropsychological test score rests on the instrument's ability to detect a positive test result in a reliable and valid manner while yielding a negative test result for the individual without actual impairment.

A relatively recent movement within neuropsychology involves the development of techniques to more precisely and empirically distinguishing between real neuropsychological change and performance variability due to psychometric or other extraneous factors. Using a test–retest paradigm, neuropsychologists have demonstrated the utility of reliable change indices (RCIs) and regression-based norms for change to identify and characterize meaningful and reliable change in neurocognitive performance by patients over time (Chelune et al.,

Reprint requests to: William B. Barr, Ph.D., NYU Comprehensive Epilepsy Center, 560 First Avenue, Rivergate 4th Floor, New York, NY 10016. E-mail: william.barr@med.nyu.edu

1993; Hermann et al., 1991, 1996; McSweeny et al., 1993; Sawrie et al., 1996;). A revised version of the original RCI calculation (Jacobson & Truax, 1991) now includes an adjustment for practice effects from serial testing, which has been demonstrated to considerably improve the predictive accuracy of an instrument (Chelune et al., 1993; Temkin et al., 1999). Simple and multiple regression methods have also been utilized to assess the concept of neuropsychological change (McSweeny et al., 1993). These regression-based approaches provide correction for both practice effects and regression toward the mean by taking into account baseline test performance and its impact on predicting performance on retesting.

The difficulties of interpreting change in performance on neuropsychological testing is perhaps no more complicated than in the case of mild traumatic brain injury (MTBI). Although numerous studies have used neuropsychological tests to document cognitive impairment following MTBI, there is debate about the role of injury-related and noninjury-related factors contributing to neuropsychological test performance in this population (Alexander, 1995, 1997; Binder, 1997; Binder et al., 1997). Additionally, cognitive functions such as attention, processing speed, and working memory, which appear to be most sensitive to change after MTBI, are considered to carry the least "hold" value in test–retest situations. In other words, these functions are not only subject to being affected by MTBI, but are also prone to the effects of numerous factors including anxiety, fatigue, and physical pain. As a result, the neuropsychologist evaluating a patient with MTBI in a period of days, weeks, or months post injury is often faced with the difficult task of teasing apart the effects of cognitive impairment from other possible confounding factors.

In recent years, neuropsychological investigation of MTBI in organized sports has revealed many advantages for prospective, controlled research not typically present in traditional studies on this population (Barth et al., 1989; Erlanger et al., 1999; Lovell & Collins, 1998). These advantages include access to a large at-risk population to undergo preinjury baseline testing, eye witness accounts of the injury, feasibility of conducting standardized assessment within minutes of injury, availability of participants for postinjury follow-up testing, access to a large pool of noninjured controls participants matched to injured subjects on several key variables. Furthermore, the issues of motivation and effort during testing, which are often possible confounds when studying patients who happen to be in some form of litigation, are not as much of a concern when evaluating individuals who are more likely attempting to "look good" in order to return to play.

These factors make the playing field a natural laboratory to study the effects of MTBI. As a result, neuropsychologists are playing an increasingly prominent role in assessing the effects of concussion resulting from sports competition. With its emphasis on empirically based methods of assessment, the field of neuropsychology is in an excellent position to provide information that can be used to monitor recovery following concussion and safety of injured athletes returning to competition after injury. Development of scientifically based methods and procedures for documenting the effects of injury in this population will have implications not only for sports injuries, but also for the effects of MTBI in the larger population.

Several studies have now utilized neuropsychological testing in sports settings to investigate the effects of MTBI days or weeks following the injury but none of these studies were designed to assess the immediate effects of injury (Collins et al., 1999; Hinton-Bayre et al., 1997; Lovell & Collins, 1998; Macciocchi et al., 1996; Maddocks et al., 1995). The lack of empirical data on the immediate effects of concussion is partly explained by situational constraints that significantly limit the capacity for extensive neuropsychological testing in assessing MTBI during the acute phase. The Standardized Assessment of Concussion (SAC) is a brief mental status and neurologic screening instrument originally developed to provide sports medicine clinicians with a standardized method of assessing athletes within minutes of having sustained MTBI during competition (McCrea et al., 1997, 1998). Earlier studies have demonstrated the SAC's utility as a sensitive and accurate method of detecting mental status and neurologic abnormalities immediately following sports-related concussion, but did not systematically examine the reliability of change in SAC performance as an indicator of clinically meaningful change in neurocognitive status resulting from injury (McCrea et al., 1997, 1998; Pottinger et al., 1999). Using a sports research model, combined with a test-retest design, the goal of the present study was to assess the reliability and validity of the SAC as an objective measure of the immediate neurocognitive effects of MTBI.

METHODS

Research Participants

A total of 1,313 male football players from 15 high schools and four universities underwent preseason baseline testing with a brief measure of cognitive functioning between 1997 and 1999. Thirty-two noninjured athletes from two high schools and 36 athletes from two colleges were retested at 60 and 120 days following baseline, respectively. These times corresponded roughly to the length of time between preseason baseline testing and the middle of the football season for each level of competition. Alternate forms were used for baseline and repeat testing of all subjects. Retesting of controls was conducted on the sideline during a practice session in order to control for fatigue, exertion, and exam setting. Previous reports (McCrea et al., 1997) have demonstrated that there is no significant difference in baseline performance on the SAC by noninjured controls testing during practice or on the sideline during actual sports competition, thereby supporting the collection of baseline data during practice in order to establish a valid and reliable benchmark for each subject against which to detect abnormalities resulting from sports-related concussion. The mean age of the 68 controls was 18.1 years (range 14–22 years).

Controls volunteered for retesting. Signed informed consent was obtained from all participants or their guardians prior to participation in this research.

Fifty injuries were documented over the course of the study, involving 35 high school and 15 collegiate athletes. All injured subjects completed baseline testing as part of the larger sample. Mean age of the injured athletes was 17.2 years (range 14–22 years). Athletes identified by the team's certified athletic trainer as having sustained a possible concussion were tested on the sideline immediately following injury. The mean interval between baseline testing and time of injury was 46.9 days with a range from 6 to 92 days. No players were independently diagnosed as having concussion without receiving follow-up testing. Alternate test forms were used for baseline and post-injury testing of all injured participants.

Concussion was defined according to the American Academy of Neurology (AAN) Practice Parameter (i.e., trauma-induced alteration in mental status with or without loss of consciousness; Kelly & Rosenberg, 1997) and the American Congress of Rehabilitation Medicine (ACRM) definition for mild traumatic brain injury (i.e., alteration in mental status, LOC for 30 min or less, PTA not greater than 24 hr; ACRM, 1993). Criteria contributing to the identification of an injured player included mechanism of injury (e.g., acceleration or rotational forces applied to the head), symptoms reported or signs exhibited (e.g., confusion, headache, dizziness, memory problems; Kelly & Rosenberg, 1997) by the player, and reports by teammates and other witnesses regarding the injured player's condition. These criteria for defining injury have been employed by most studies on sports-related concussion (Collins et al. 1999; Hinton-Bayre et al., 1997; Maddocks & Saling, 1996).

Measures

The Standardized Assessment of Concussion (SAC) was designed according to the recommendations of the American Academy of Neurology Practice Parameter (Kelly & Rosenberg, 1997) and the Colorado Guidelines for the Management of Concussion in Sports (Colorado Medical Society, 1991). The SAC assess four neurocognitive domains considered sensitive to change following mild traumatic brain injury, including Orientation, Immediate Memory, Concentration, and Delayed Recall (see Appendix). The instrument requires approximately 6 min to administer and is designed for use by a non-neuropsychologist with no prior expertise in psychometric testing.

The SAC is not intended as a substitute for medical, neurologic, or neuropsychological evaluation of the injured individual, but offers a quantifiable measure of deficits in orientation, concentration and memory that is feasible for use under the constraints of rapid, sideline evaluation of the injured player. Alternate Forms A, B, and C of the SAC were designed to allow follow-up testing of injured participants with minimal practice effects in order to track post-concussion recovery. The three forms differ only in the stimulus selection of digits in the Concentration section and words used to test Immediate Memory and Delayed Recall. Previous research (McCrea et al., 1997, 1998) has demonstrated the equivalence of these three forms for clinical use. The SAC is printed on pocket-sized cards for convenient use by athletic trainers and other medical personnel examining athletes on the sideline.

A standard line of questioning is used to assess orientation: the participant is asked to provide the day of the week, month, date, year and time of day within 1 hr. A 5-word list is used to measure immediate memory; the list is read to the participant for immediate recall and the procedure is repeated for three trials. Concentration is tested by having the participant repeat, in reverse order, strings of digits that increase in length from three to six numbers. Reciting the months of the year in reverse order is also utilized to assess concentration. Delayed recall of the original 5-word list is also ascertained. Total score (maximum = 30) is computed in order to derive a composite index of the overall level of impairment following concussion.

All participants completed baseline testing with the SAC as part of a larger physical examination prior to the beginning of the football season. Controls underwent repeat testing as scheduled by the protocol. All baseline testing and repeat testing of controls was conducted individually by trained research assistants or certified athletic trainers. Assessments following injury were conducted by trained professionals, research staff, or certified athletic trainers who had received training in administration of the SAC.

Data Analysis

Group differences in demographic and other variables were assessed with individual *t* tests. Changes in SAC scores over time were assessed with multivariate analysis of variance (MANOVA) for repeated measures. An alpha level of .05 was used to determine statistical significance for all analyses.

Reliable change indices (RCI) were computed using methodology described by Chelune and colleagues (Chelune et al., 1993; Sawrie et al., 1996). The standard error of measurement (SE_m) and standard error of difference (S_{Diff}) were calculated using the coefficient for test–retest reliability according to the formulae provided by Jacobson and Truax (1991). Reliable change is calculated as the difference in scores across time ($T1 - T2$) divided by the S_{Diff}. A confidence interval (*CI*) is calculated by multiplying the reliable change value by the desired *z*-score cut point. For example, to compute the most commonly used value, the 95th percentage point, one computes a 90% *CI* through multiplying the value by 1.64. Calculations of 80% and 70% confidence intervals require values of 1.30 and 1.05 respectively.

A multiple regression model was used to compute equations for predicting retest scores from a combination of baseline indices and demographic variables. The method used in this study is similar to what was described originally by McSweeney et al. (1993) and has been used in other studies (Chelune et al., 1993; Hermann et al., 1996; Sawrie et al., 1996; Temkin et al., 1999). Baseline SAC scores were used to predict scores at repeat testing through a stepwise analysis, using age and time interval as possible predictor values.

Receiver operating characteristic (ROC) curve analyses were conducted in a manner similar to those described in studies of epilepsy and dementia (Barr, 1997; Monsch et al., 1992). Each difference score was treated as a separate cut-off. The sum of injured or control athletes obtaining scores at or below these cut-offs was determined by examination of frequency distributions. Measures of sensitivity and specificity for distinguishing between these athlete groups were calculated from these values according with established formulae. Sensitivity (*Se*) refers to the probability that a patient with a certain diagnosis will be correctly identified. Specificity (*Sp*) refers to the probability that an individual will be correctly classified as not having the diagnosis. Cutting scores were evaluated by adding the *Se* and *Sp* values of each particular test score. The score with the greatest classification value was defined as the one with the greatest summed sensitivity and specificity values. Sensitivity and specificity ($1 - Sp$) values were plotted graphically to obtain ROC curves for each test. Statistical comparison of the areas under the curves were analyzed by a nonparametric method described by Hanley and McNeil (1982).

RESULTS

The results of a MANOVA for repeated testing with the SAC revealed no significant retest effect for the entire group of 68 control subjects [$F(1,66) = .87$, n.s.]. There was, however, a significant interaction effect indicating a difference in retest effects observed in the college and high school samples [$F(1,66) = 8.57$, $p = .005$]. An examination of mean difference scores (Test 2 − Test 1) indicates that college participants tested at 120 days following baseline exhibit more of a "learning effect" and less variability than high school participants tested following a 60 day-interval [college: $M = +.67$, $SD = 1.09$; high school: $M = -.34$, $SD = 1.72$).

The college and high school control samples were combined into one control group for comparison to injured athletes. Preliminary analyses indicate group differences in age [$t(116) = 2.45$, $p = .016$] and in the interval between baseline and subsequent testing [$t(116) = 8.95$, $p < .001$]. Control athletes were older and were retested after longer intervals than the injured group. Mean SAC scores at Time 1 and Time 2 for both groups are presented in Table 1. Injured athletes exhibit an approximate 4-point drop in SAC scores following concussion. Controls exhibit a less than 1 point increase in scores upon retesting. The results of a MANOVA for repeated measures indicates that this change differs significantly from what is observed in the control sample [$F(1,116) = 111.17$, $p < .001$]. The effect remains significant after accounting for the effects of age and time interval between baseline and repeat testing with a MANCOVA [$F(1,114) = 27.39$, $p < .001$].

Test–Retest Reliability and Reliable Change Cut-off Scores

Test–retest indices for the control group are included in Table 1. The level of test–retest reliability was only moderate, though statistically significant ($r = .55$, $p < .001$). Computation of a reliable change score at the 90% CI resulted in a value of ±2.38. After adding the mean change score from this group, as suggested by others (Chelune et al., 1993), the adjusted score would be ±2.59. Rounded to the nearest test score, this would indicate that an increase or decrease of 3 points with repeat testing on the SAC would represent a statistically reliable and clinically significant change in performance. Intervals based on 80% and 70% confidence intervals would be ±2.08 and ±1.71 respectively. The distribution of observed change scores, with sensitivity and specificity values, for the concussion and control groups is included in Table 2.

Regression Analysis of Predicted Change

The equation for a multiple regression model of predicted change in the control group is listed in Table 3. An initial analysis indicated that the baseline test score accounts for approximately 30% of the variance in follow-up test scores. The results of a stepwise analysis provided a significant model including effects of baseline scores and test interval [$F(2,65) = 25.18$, $p < .001$]. The combined model accounts for 44% of the variance. The effect of age did not meet statistical criteria for entry. Using this equation, the mean predicted value for SAC scores at Time 2 are 27.4 ($SD = 1.29$). The 90% confidence interval, based on the standard deviation of the residual, is ±1.99. Differences between predicted and observed scores at Time 2 are listed in Table 4.

ROC Curve Analysis of Cutting Scores

An examination of the distribution of observed change scores in Table 2 reveals that 72% of the injured sample exhibits a decrease in scores exceeding the adjusted RCI value of 3 or more points following concussion. Only 6% of the control

Table 1. SAC Scores at Time 1 and Time 2 for test–retest and concussion groups

	Time 1		Time 2		$T2 - T1$		$T1, T2$		
Group	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)	*r*	SE_m	S_{Diff}
Test–retest controls ($N = 68$)	27.78	(1.53)	27.97	(1.61)	+0.19	(1.50)	.55	1.03	1.45
Athletes with concussion ($N = 50$)	27.06	(1.94)	22.90	(3.32)	−4.16	(2.92)	—	—	—

https://doi.org/10.1017/S1355617701766052 Published online by Cambridge University Press

Table 2. Distribution of observed differences between SAC scores from Time 1 and Time 2

Difference score ($T1 - T2$)	Test–retest controls (Number)	Athletes with concussion (Number)	Sensitivity (*Se*)	Specificity (*Sp*)	Sum (*Se* + *Sp*)
+2	10	0	1.00	.04	1.04
+1	22	0	1.00	.15	1.15
0	20	3	1.00	.47	1.47
−1	5	5	.94	.76	1.70
−2	7	6	.84	.84	1.68
−3	3	9	.72	.94	1.66
−4	1	10	.54	.99	1.53
−5	0	17	.34	1.00	1.34

sample exhibited this decrease in test scores. Thirty-four percent of the injured group exhibited a drop of 5 or more points. Summed level of sensitivity and specificity values, used as gross indices of classification accuracy, indicate that cut-off scores ranging from a decrease of 1 to 3 points provide comparable levels of group categorization. In the sports setting, one might want to emphasize the sensitivity of the instrument over its specificity in order to identify as many injured participants as possible. In that context, a conservative cut-off score indicating a decline of 1 point provides the highest level of sensitivity to the effects of concussion. The combined sensitivity and specificity of this score was the highest at a value of 1.70.

Examination of the regression-based difference scores presented in Table 4 indicates that 76% of the injured athletes exhibit a drop of 2 or more points from their predicted score. Discrepancies of this magnitude were observed in only 7% of the control group. A 3-point difference between observed and predicted scores provided the highest degree of classification as determined by the summed values, though this score correctly identified only 72% of the injured sample.

Classification of participants, using scores exceeding the 90% confidence interval, were comparable for both models. The summed value of $Se + Sp$ for a 3-point decline with the RCI model was 1.66. The corresponding value for a 2-point drop, using the regression model, was 1.69. The statistically derived cutting score from the regression model appears to provide slightly higher sensitivity than the RCI model. Specificity values were nearly identical.

Comparison of Distributions From RCI and Regression Predictions

Figure 1 provides receiver operating characteristic (ROC) curves for group classification using the RCI and regression models. The area under the curve (AUC) for the RCI model, computed according to a nonparametric methods, was .939 ($SE = .021$). The AUC for the regression model was .911 ($SE = .028$). While a comparison of these areas indicates a greater AUC for the RCI model, the difference between the curves was not statistically significant ($p = .178$).

DISCUSSION

The results of this study demonstrate that the Standardized Assessment of Concussion (SAC) is a reliable and valid measure for evaluating the early neurocognitive effects of sports-related head injury. High school and college athletes tested within minutes of sustaining a concussion exhibited an average decrease of 4 points on a 30-point scale, while controls showed an average increase of less than 1 point when retested with the SAC. The decrease in test scores by injured participants indicates the presence of measurable neurocognitive impairment immediately following MTBI. These differences were not the result of age effects or differences in the interval between baseline and repeat testing.

Computation of reliable change indices (RCIs) for the SAC indicated that a decrease of 3 points represents a significant change when conservative criteria (i.e., 90% CI) are used. An examination of the distribution of observed change scores, however, indicates that decreases ranging from 1 to 3 points provide comparable levels of classification of injured athletes and controls. Ninety-four percent of the injured group showed a drop of 1 or more points as compared to only 24% of the control sample. Examination of sensitivity and specificity values indicates that this change score is optimal for identifying injured participants.

The use of multiple regression indicated that baseline test scores and testing interval provided a significant model for predicting scores at retesting. While 90% of the injured

Table 3. Regression equation for predicted SAC scores

	R	SE_{est}	Constant	Beta	
				SAC Score ($T1$)	Time Interval (days)
Predicted SAC score at Time 2	.661	1.23	11.89	.523	.014

https://doi.org/10.1017/S1355617701766052 Published online by Cambridge University Press

Table 4. Distribution of differences between predicted and observed SAC scores at Time 2

Difference score (Predicted − $T2$)	Test–retest controls (Number)	Athletes with concussion (Number)	Sensitivity (*Se*)	Specificity (*Sp*)	Sum (*Se* + *Sp*)
+2	8	0	1.00	.12	1.12
+1	15	1	1.00	.12	1.12
0	23	4	.98	.34	1.32
−1	17	7	.90	.68	1.58
−2	4	2	.76	.93	1.69
−3	0	13	.72	.99	1.71
−4	1	6	.46	.99	1.45
−5	0	17	.34	1.00	1.34

sample exhibited a 1-point difference between observed and predicted scores at the time of their injury, this difference in scores was also observed in 32% of the control sample. A comparison of receiver operating characteristic (ROC) curves indicates that classification of injured athletes and controls with the SAC is not necessarily enhanced by including the statistical effects of baseline test scores and the retest interval.

Assessment and interpretation of change in neuropsychological test performance is impacted by several factors other than injury or disease, including test–retest reliability issues, practice effects from serial testing, and regression to the mean (Bruggemans et al., 1997; Hermann et al., 1996; Sawrie et al., 1996). Unfortunately, less-than-perfect reliability of neuropsychological tests can result in varied results over repeated administrations, even in people who have not experienced any true change in neurobehavioral status (Temkin, et al., 1999). Obviously, this predicament confounds the neuropsychological interpretation of test findings, especially in the case of MTBI.

Much interest has focused on the use of RCI and regression methods for predicting performance in test–retest situations (Chelune et al., 1993; Hermann et al., 1996; McSweeny et al., 1993; Sawrie et al., 1996). In a recent study, Temkin et al. (1999) compared these methods in a large sample of controls undergoing retesting with a relatively large sample of neuropsychological tests. Similar to that study, we found that baseline performance and test interval contribute to the prediction of retest scores, though the level of variance attributed to SAC baseline scores was lower than what was seen with most tests. While that study found that the regression model outperformed the corrected RCI model, our results indicate that both methods are equivalent in terms of identifying injured athletes immediately following concussion.

Based on the relatively low test–retest reliability (r = .55) of the SAC in this study, it may be that instruments sensitive to the effects of MTBI, consisting primarily of "fluid" measures of attention, memory, and cognitive processing, are by definition susceptible to variability over time. This is, in fact, the case in a recent study where measures of digit span and recall, which are components of the SAC, are found to have less reliability than other measures of neuro-

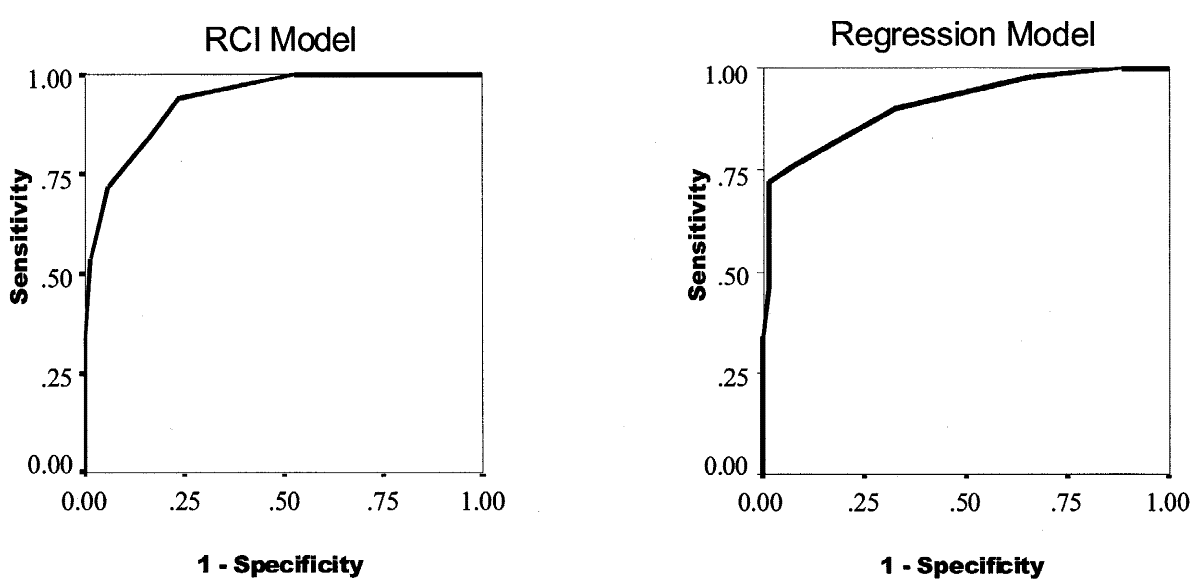

Fig. 1. Receiver operating characteristic (ROC) curves for reliable change index (RCI) and regression models with the SAC.

https://doi.org/10.1017/S1355617701766052 Published online by Cambridge University Press

psychological functioning (Dikmen et al., 1999). The issue of testing these functions in a sample of adolescents and young adults also needs to be addressed. The lack of reliability of such measures is likely to be an issue in future research relying on test–retest paradigms for assessment of neuropsychological change.

Although the persistent neuropsychological effects of MTBI weeks and months after injury have been extensively researched, very few studies have provided objective, empirical data on the immediate neurocognitive effects following MTBI (Binder, 1997; Maddocks et al., 1995; McCrea et al., 1997, 1998; Yarnell & Lynch, 1970). This gap in the literature is due in large part to situational circumstances that limit the capacity for prospective research implementing standardized assessment methods immediately following MTBI. Standardized measures for assessing mental status and neurologic changes beyond traditional injury classification criteria (e.g., Glasgow Coma Scale; Teasdale & Jennett, 1974) are quite uncommon in most trauma settings. Furthermore, injury severity grading systems may not be sensitive to subtle neurocognitive changes that present risks for more severe underlying neurologic complications after MTBI (Stein et al., 1993). In contrast, neuropsychological testing is considered a sensitive and sophisticated method for assessing concussion, but typically is not feasible immediately after injury in most acute care settings.

Results from this study and earlier reports (McCrea et al., 1997, 1998; Pottinger et al., 1999) indicate that the SAC provides not only a valid measure of neurocognitive changes immediately following head injury, but also one that is feasible for use under the constraints of the sports sideline and other acute or emergency care settings. Administration of screening instruments such as the SAC as soon as possible during the acute injury phase may have great clinical and research implications for clarifying the early natural history of MTBI and determining how the immediate neurocognitive effects of injury are predictive of eventual neuropsychological outcome. The design outlined in this study may also be helpful in determining the reliability and validity of other neuropsychological tests traditionally used in the assessment of MTBI patients, particularly in terms of deriving statistical indices of clinically meaningful change in neurocognitive status resulting from head injury.

The study of sports-related concussion provides one of the only opportunities for prospectively studying changes from baseline neurocognitive functioning following MTBI. Findings from this study are comparable to those found in other neuropsychological studies using different methods of assessment. In the only other published sports concussion study examining retest effects with RCI methods, Hinton-Bayre et al. (1999) found that 80% of their professional rugby players exhibited test scores exceeding the 90% CI. A similar level of classification of injured and control athletes was found in another study using different methods (Collins et al., 1999). Macciocchi et al. (1996) found that injured athletes could be distinguished from controls by exhibiting a lack of a "practice effect." All of these studies have demonstrated the utility of using neuropsychological test measures as a sensitive means for identifying persistent cognitive impairment in athletes.

This study extends previous work by evaluating classification rates with the use of ROC curve analyses. This method is utilized best for examining rates of diagnostic classification when well defined groups of patients and controls are identified. In this study ROC curve analyses revealed that while RCI and regression methods might provide the most reliable form of classification, they might be too conservative for use in the management of sports concussion. The use of ROC methods is recommended for future use in sports studies as a means to evaluate published guidelines regarding injury severity and return to play.

Our findings are perhaps most informative for the sports medicine clinician responsible for the assessment and management of concussion in young athletes. Our computation of RCI's indicates that a change score of 3 or more points is the most sensitive statistical index of change on the SAC, when the effects of error variance and practice are taken into account. However, when the actual distribution of change scores is observed, analyses of ROC curve data indicate that decreases of 1 to 3 points result in similar levels of classification. While a decrease of 1 point on the SAC identifies the largest number of injured athletes, it also provides incorrect classification of the highest number of athletes without injury. This raises an important issue. Should the clinician utilize the most psychometrically rigorous cutoff score or the most sensitive cutoff score? While parents and educators might opt for a classification criterion that emphasizes the sensitivity to detection of injury, the misclassification rate for other athletes might be unacceptable to coaches and teammates who are eager to have the player return to the playing field as soon as possible.

It should be made clear that the criteria for detecting statistical significance are arbitrary values that should be set by the clinician or experimenter depending on the question at hand. In the case of identifying young athletes with injuries, the use of a 90% confidence interval might be too conservative when identification of subtle impairments is desired and misclassification of uninjured players results in nothing more than missing the chance to return immediately to competition. While neuropsychological research can provide the empirical information for use in clinical decision-making, the ultimate responsibility must reside with the clinician, taking into account all available information. It is especially important to clarify that the SAC is neither intended as an independent return to play measure for sports medicine clinicians, nor a substitute for more extensive medical, neurologic or neuropsychological evaluation of the injured athlete following concussion. Clinicians should also be alerted that the mental status exam is just one facet of concussion assessment, and that a full survey of symptoms and physical indicators of injury should also be conducted. Objective data from standardized measures such as the SAC may aid the clinician in making an overall assessment of injury, recovery and the player's readiness to safely return to competition after concussion.

In summary, neuropsychology is likely to have a continuing impact on the development of methods for assessing sports-related head injury. Future research will need to address the empirical basis of return to play criteria. Attention will also need to focus on the use of multiple baselines to prevent practice effects following injury, the role that learning disability and prior concussions may have on the recovery of function, and the nature of practice effects resulting from repeated test administration during recovery (Collins et al., 1999; Hinton-Bayre et al., 1999). Future studies using athletes with orthopedic injuries as controls is also recommended to examine the more general effects of injury on neuropsychological status (Satz et al., 1999). Maintaining the scientist–practitioner model of neuropsychology will help us to extend findings obtained from the study of sports injury to the larger arena of clinical and forensic assessment of MTBI in the general population.

REFERENCES

Alexander, M.P. (1995). Mild traumatic brain injury: Pathophysiology, natural history, and clinical management. *Neurology*, *45*, 1253–1260.

Alexander, M.P. (1997). Minor traumatic brain injury: A review of physiogenesis and psychogenesis. *Seminars in Clinical Neuropsychiatry*, *2*, 177–187.

American Congress of Rehabilitation Medicine. (1993). Report of the Mild Traumatic Brain Injury Committee of the Head Injury Interdisciplinary Special Interest Group. Definition of mild traumatic brain injury. *Journal of Head Trauma Rehabilitation*, *8*, 86–87.

Barr, W.B. (1997). Receiver operating characteristic curve analysis of Wechsler Memory Scale–Revised scores in epilepsy surgery candidates. *Psychological Assessment*, *9*, 171–176.

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. In H.S. Levin, H.M. Eisenberg, & A.L. Benton (Eds.), *Mild head injury* (pp. 257–275). New York: Oxford University Press.

Binder, L.M. (1997). A review of mild head trauma: Part I. Meta-analytic review of neuropsychological studies. *Journal of Clinical and Experimental Neuropsychology*, *19*, 421–431.

Binder, L.M., Rohling, M.L., & Larrabee, G.J. (1997). A review of mild head trauma: Part II. Clinical implications. *Journal of Clinical and Experimental Neuropsychology*, *19*, 432–457.

Bruggemans, E.F., Van de Vijver F.J., & Huysmans, H.A. (1997). Assessment of cognitive deterioration in individual patients following cardiac surgery: Correcting for measurement error and practice effects. *Journal of Clinical and Experimental Neuropsychology*, *4*, 543–559.

Chelune, G.J., Naugle, R.I., Luders, H., Sedlak, J., & Awad, I.A. (1993). Individual change after epilepsy surgery: Practice effects and base-rate information. *Neuropsychology*, *7*, 41–52.

Collins, M.W., Grindel, S.H., Lovell, M.R., Dede, D.E., Moser, D.J., Phalin, B.R., Nogle, S. Wasik, M., Cordry, D., Daugherty, M.K., Sears, S.F., Nicolette, G., Indelicato, P., & McKeag, D.B.(1999). Relationship between concussion and neuropsychological performance in college athletes. *Journal of the American Medical Association*, *282*, 964–970.

Colorado Medical Society Sports Medicine Committee. (1991). *Guidelines of the management of concussion in sports*. Denver, CO: Colorado Medical Society.

Dikmen, S.S., Heaton, R.K., Grant, I., & Temkin, N.R. (1999). Test–retest reliability and practice effects of Expanded Halstead-Reitan Neuropsychological Test Battery. *Journal of the International Neuropsychological Society*, *5*, 346–356.

Erlanger, D.M, Kutner, K.C., Barth, J.T. & Barnes, R. (1999). Neuropsychology of sports-related head injury: Dementia pugilistica to post-concussion syndrome. *Clinical Neuropsychologist*, *13*, 193–209.

Hanley, J.A. & McNeil, B.J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Diagnostic Radiology*, *143*, 29–36.

Hermann, B.P., Seidenberg, M., Schoenfeld, J., Peterson, J., Leveroni, C., & Wyler, A.R. (1996). Empirical techniques for determining the reliability, magnitude, and pattern of neuropsychological change after epilepsy surgery. *Epilepsia*, *37*, 942–950.

Hermann, B.P., Wyler, A.R., VanderZwagg, R., LeBailly, R.K., Whitman, S., Somes, G., & Ward, A.R. (1991). Predictors of neuropsychological change following anterior temporal lobectomy: Role of regression toward the mean. *Journal of Epilepsy*, *4*, 139–148.

Hinton-Bayre, A.D., Geffen, G., & McFarland, K. (1997). Mild head injury and speed of information processing: A prospective study of professional rugby league players. *Journal of Clinical and Experimental Neuropsychology*, *19*, 275–289.

Hinton-Bayre, A.D., Geffen, G. M., Geffen, L.B., McFarland, K.A., & Friis, P. (1999). Concussion in contact sports: Reliable change indices of impairment and recovery. *Journal of Clinical and Experimental Neuropsychology*, *21*, 70–86.

Jacobson, N.S. & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*, 12–19.

Keefe, R.S.E. (1995). The contribution of neuropsychology to psychiatry. *American Journal of Psychiatry*, *152*, 6–15.

Kelly, J.P. & Rosenberg, J. (1997). Practice Parameter. The management of concussion in sports. Report of the Quality Standards Subcommittee. *Neurology*, *48*, 581–585.

Levin, H.S. (1994). A guide to clinical neuropsychological testing. *Archives of Neurology*, *51*, 854–859.

Lovell, M.R. & Collins, M.W. (1998). Neuropsychological assessment of the college football player. *Journal of Head Trauma Rehabilitation*, *13*, 9–26.

Macciocchi, S.N., Barth, J.T., Alves, W., Rimel, R.W., & Jane, J.A.(1996). Neuropsychological functioning and recovery after mild head injury in collegiate athletes. *Neurosurgery*, *39*, 510–514.

Maddocks, D. & Saling, M. (1996). Neuropsychological deficits following concussion. *Brain Injury*, *10*, 99–103.

Maddocks, D., Dicke, G.D., & Saling, M.M. (1995). The assessment of orientation following concussion in athletes. *Clinical Journal of Sports Medicine*, *5*, 32–35.

McCrea, M., Kelly, J.P., Kluge, J., Ackley, B., & Randolph, C. (1997). Standardized assessment of concussion in football players. *Neurology*, *48*, 586–588.

McCrea, M., Kelly, J.P., & Randolph, D. (1997). *The Standardized Assessment of Concussion (SAC): Manual for administration, scoring and interpretation*.

McCrea, M., Kelly, J.P., Randolph, C., Kluge, J., Bartolic, E., Finn, G., & Baxter, B. (1998). Standardized assessment of concussion: On-site mental status evaluation of the athlete. *Journal of Head Trauma Rehabilitation*, *13*, 27–35.

McSweeny, A.J., Naugle, R.I., Chelune, G.J., & Luders, H. (1993). T-scores for change: An illustration of a regression approach to depicting change in clinical neuropsychology. *The Clinical Neuropsychologist*, *7*, 300–312.

Monsch, A.U., Bondi, M.W., Butters, N., Salmon, D.P., Katzman, R. & Thal, L.J. (1992). Comparisons of verbal fluency tasks in the detection of dementia of the Alzheimer's type. *Archives of Neurology*, *49*, 1253–1258.

Pottinger, L., Cullum, M., & Stallings, R.L. (1999). Cognitive recovery following concussion in high school athletes. *Archives of Clinical Neuropsychology*, *14*, 39–40.

Rimel, R.W., Giordani, B., Barth, J.T., Boll, T.J., & Jane, J.A. (1981). Disability caused by minor head injury. *Neurosurgery*, *9*, 221–228.

Satz, P., Alfano, M.S., Light, R., Morgenstern, H., Zaucha, K., Asarnow, R.F., & Newton, S. (1999). Persistent post-concussive syndrome: A proposed methodology and literature review to determine the effects, if any, of mild head injury and other bodily injury. *Journal of Clinical and Experimental Neuropsychology*, *21*, 620–628.

Sawrie, M.S., Chelune, G.J., Naugle, R.I., & Luders, H.O. (1996). Empirical methods for assessing meaningful change following epilepsy surgery. *Journal of the International Neuropsychological Society*, *2*, 556–564.

Stein, S.C., Spettell, C., Young, G., & Ross, S.E. (1993). Limitations of neurological assessment in mild head injury. *Brain Injury*, *7*, 425–430.

Teasdale, G. & Jennett, B. (1974). Assessment of coma and impaired consciousness: A practical scale. *Lancet*, *2*(7872), 81–84.

Temkin N.R., Heaton, R.K., Grant I., & Dikmen, S.S. (1999). Detecting significant change in neuropsychological test performance: A comparison of four models. *Journal of the International Neuropsychological Society*, *5*, 357–369.

Yarnell, P.R. & Lynch, S. (1970). Retrograde memory immediately after concussion. *Lancet*, *1*(7652), 863–864.

APPENDIX

STANDARDIZED ASSESSMENT OF CONCUSSION (SAC)

1) ORIENTATION:

Month: ______________________ 0 1
Date: ______________________________ 0 1
Day of week: ________________________ 0 1
Year: ______________________________ 0 1
Time (within 1 hr.): ___________________ 0 1

Orientation Total Score ________ / 5

2) IMMEDIATE MEMORY: (all 3 trials are completed regardless of score on trial 1 & 2; total score equals sum across all 3 trials)

List	Trial 1	Trial 2	Trial 3
Word 1	0 1	0 1	0 1
Word 2	0 1	0 1	0 1
Word 3	0 1	0 1	0 1
Word 4	0 1	0 1	0 1
Word 5	0 1	0 1	0 1
Total			

Immediate Memory Total Score ________ / 15

(Note: Subject is not informed of Delayed Recall testing of memory)

NEUROLOGICAL SCREENING:

Loss of Consciousness: (occurrence, duration)

Pre- & Post-traumatic Amnesia: (recollection of events pre- and post-injury)

Strength:

Sensation:

Coordination:

3) CONCENTRATION:

Digits Backward (If correct, go to next string length. If incorrect, read trial 2. Stop after incorrect on both trials)

4-9-3 6-2-9 ________________ 0 1
3-8-1-4 3-2-7-9 ________ 0 1
6-2-9-7-1 1-5-2-8-6 ________________ 0 1
7-1-8-4-6-2 5-3-9-1-4-8 ________________ 0 1

Months in reverse order: (entire sequence correct for 1 point)
Dec-Nov-Oct-Sep-Aug-Jul
Jun-May-Apr-Mar-Feb-Jan ________________ 0 1

Concentration Total Score ________ / 5

EXERTIONAL MANEUVERS
(when appropriate):

5 jumping jacks **5 push-ups**
5 sit-ups **5 knee-bends**

4) DELAYED RECALL

Word 1 0 1
Word 2 0 1
Word 3 0 1
Word 4 0 1
Word 5 0 1

Delayed Recall Total Score ________ / 5

SUMMARY OF TOTAL SCORES:

Orientation________ / 5
Immediate Memory________ / 15
Concentration________ / 5
Delayed Recall________ / 5

Overall Total Score________ / **30**

https://doi.org/10.1017/S1355617701766052 Published online by Cambridge University Press