The sensitivity and specificity of clinical measures of sport concussion: three tests are better than one

Jacob E Resch,[1] Cathleen N Brown,[2] Julianne Schmidt,[2] Stephen N Macciocchi,[3] Damond Blueitt,[4] C Munro Cullum,[5] Michael S Ferrara[6]

¹Exercise and Sport Injury Laboratory Department of Kinesiology The University of Virginia, Charlottesville, VA
²St. Mary's Athletic Training Research Laboratory, Department of Kinesiology, University of Georgia, Athens, Georgia, USA
³Atlanta Neuropsychology LLC Marble Hill, Georgia, USA
⁴Orthopedic Specialty Associates, Fort Worth, Texas, USA
⁵The University of Texas Southwestern Medical Center, Dallas, Texas, USA
⁶The University of New Hampshire, Durham, New Hampshire, UK

Correspondence to
Dr Jacob E Resch;
jer6x@virginia.edu

ABSTRACT

Context: A battery of clinical measures of neurocognition, balance and symptoms has been recommended for the management of sport concussion (SC) but is based on variable evidence.
Objective: To examine the sensitivity and specificity of a battery of tests to assess SC in college athletes.
Design: Cross-sectional.
Setting: Research laboratory.
Patients or other participants: Division 1 athletes diagnosed with a SC (n=40) who were 20.2±1.60 years of age and 180.5±11.12 cm tall and healthy athletes (n=40) who were 19.0±0.93 years of age and 179.1 ±11.39 cm tall were enrolled.
Intervention(s): Participants were administered Immediate Postconcussion Assessment and Cognitive Test (ImPACT), the Sensory Organization Test (SOT) and the Revised Head Injury Scale (HIS-r) prior to and up to 24 h following injury between the 2004 and 2014 sport seasons. Sensitivity and specificity were calculated using predictive discriminant analyses (PDA) and clinical interpretation guidelines.
Main outcome measures: Outcome measures included baseline and postinjury ImPACT, SOT and HIS-r composite scores.
Results: Using PDA, each clinical measure's sensitivity ranged from 55.0% to 77.5% and specificity ranged from 52.5% to 100%. The test battery possessed a sensitivity and specificity of 80.0% and 100%, respectively. Using clinical interpretation guidelines, sensitivity ranged from 55% to 97.5% individually, and 100% when combined.
Conclusions: Our results support a multidimensional approach to assess SC in college athletes which correctly identified 80–100% of concussed participants as injured. When each test was evaluated separately, up to 47.5% of our sample was misclassified. Caution is warranted when using singular measures to manage SC.

What are the new findings?

- A multidimensional assessment of concussed athletes that included a symptom checklist and computerised measures of cognitive function and balance resulted in a sensitivity of 80% or 100% using predictive discriminant analysis (PDA) or clinical interpretation guidelines, respectively.
- Individual sensitivities of the component measures using PDA ranged from 52.5% to 77.5%, with an overall sensitivity of 80%.
- Overall sensitivity of the battery based on clinical interpretation guidelines was 100%, with sensitivity of each individual measure ranging from 55% to 97.5%.
- Findings support the use of clinical interpretation of multidimensional assessment procedures in the management of SC.

Medicine) and consensus panels (ie, Concussion in Sport Group) have advocated a multidimensional approach to sport concussion (SC) management.[1–4] This approach consists of traditional and/or computerised neurocognitive testing (CNT), assessment of postural stability and self-report symptom instrument which are used to compliment the physical/neurological examination.[2–4] This multidimensional approach assists clinicians by providing more specific information about the subtle deficits associated with a SC that are not detectable via more gross assessments and may guide clinical decision-making during the acute phase of injury (eg, academic adjustments and modifications to activities of daily living). To date, there remains no consensus as to what specific tests should comprise such assessment batteries.[1] The aforementioned guidelines and/or position statements serve as a standard of care or recommendations for physicians, athletic trainers and neuropsychologists involved in the assessment and treatment of SC.[5] Despite

INTRODUCTION

Since 1997, numerous governing bodies (ie, American Academy of Neurology (AAN), National Athletic Trainers' Association, American Medical Society for Sports

these guidelines, surveyed athletic trainers and emergency department physicians often do not utilise the recommended multidimensional battery of tests to manage SC[6–8] due to limited time, personnel and fiscal resources.[7 8]

An increasing body of literature has addressed the reliability and validity of the multidimensional battery of tests used to evaluate SC.[9] This body of research has heightened awareness of the psychometric limitations of single SC metrics. For instance, multiple CNTs have been shown to possess highly variable levels of sensitivity to SC (9–93%) but moderate to high levels of specificity (69.1–100%) identifying healthy control participants.[10–16] Very few studies have addressed the sensitivity and specificity of balance measures for SC.[10 11] Investigations of the Sensory Organization Test (SOT), a computerised measure of postural stability, have observed very low to moderate levels of sensitivity (2–61.9%) but high levels of specificity (92.3–94.9%).[10 11] Finally, investigations of the frequency, duration and predictive values of one or more SC symptoms have been completed, but only a few of these studies have specifically addressed sensitivity and specificity.[10 11 17]

Given the variable methodology and results across studies, it is important to systematically investigate psychometric properties of these measures as they relate to SC assessment. The purpose of our study was to compare the sensitivity and specificity of clinical measures used to assess SC in college athletes via predictive discriminative analysis and examination of clinical interpretation guidelines. Additionally, we sought to determine which composite/summary scores of the Immediate Postconcussion Assessment and Cognitive Test (ImPACT) battery, the SOT, and the HIS-r demonstrated the strongest relationship(s) to SC. We hypothesised that a battery of tests consisting of ImPACT, the SOT and the HIS-r would demonstrate superior sensitivity and specificity compared with each measure independently. The results of our investigation may assist clinicians in identifying a combination of clinical measures which possess a high sensitivity and specificity to SC.

METHODS
Data collection occurred between the 2004 and 2014 sport seasons on two large metropolitan university campuses using the same SC assessment protocols. During 10-year data collection period, a minimum of one investigator associated with the current study assisted with data collection at each site to ensure consistency of the diagnosis and management of SC. The majority of injuries were recorded between the 2004 and 2009 sport seasons at the first participating university at which participants competed in division I football, gymnastics, basketball, equestrian, volleyball, soccer, cheerleading, softball and baseball. Starting with the 2010 sport season, the common investigator between institutions transitioned to the second participating university which

did not have football, soccer and equestrian which have been documented to be high-risk sports for SC. All participants were tested prior to the start of their respective sport season. A comparison group of healthy athletes were retrospectively matched by institution, gender, sport, position (if applicable), handedness, height and weight to establish specificity. Healthy participants needed to be within 10% of each concussed athlete's height and weight values. Healthy participants reported no prior history of concussion and again were assessed prior to the start of their respective sport season. Exclusion criteria consisted of English as a second language, diagnosis of a learning disability and/or attention deficit disorder, any other self-reported psychiatric disorder, and missing preinjury (baseline) values for the ImPACT, SOT or HIS-r values for all participants. Additionally, injured athletes were excluded from analysis if the athlete sustained a non-sport related concussion, if data were collected more than 24 h following SC diagnosis, or if the athlete was missing one or more scores from the ImPACT, the SOT or HIS-r prior to or following their concussion. Participants also had to have a valid baseline ImPACT test as determined by the manufacturer's automatic validity criteria and scored within normative values for the SOT to be included in our analyses.[18 19] Each institution's respective Institutional Review Boards approved the study and each participants signed an informed consent form prior to participation.

Clinical measures
Computerised neurocognitive test
The ImPACT (versions 2.3.813 to 6.7.723; ImPACT Applications, Pittsburgh, Pennsylvania, USA) is a popular CNT, which measures attention, memory, reaction time and information processing speed. The ImPACT consists of eight subtests including immediate and delayed word recall, immediate and delayed design recall, a symbol match test, a three letter working memory task, X's and O's attention test, and a choice reaction time colour match test with five alternate forms (ie, baseline and postinjury forms 2 through 5). We administered the ImPACT's baseline assessment (form 1) prior to the start of each of participant's sport season and postinjury 1 (form 2), following the diagnosis of a SC. Though varying versions of the ImPACT were used throughout the 10-year study period, all pertinent calculations and stimuli (eg, word and design memory) remained unchanged. Each participant's baseline ImPACT report was reviewed to assess for validity and inclusion in our analyses. The ImPACT includes invalidity criteria which assess for grossly inadequate effort during the baseline assessment and is described elsewhere.[18] The ImPACT took approximately 25 min to complete. In addition to the cognitive subtest scores, the ImPACT's Total Symptom Score (TSS), a 22-item checklist of various physical, cognitive and mood symptoms, was included in our analyses. On the TSS, participants

rate each individual symptom on a scale of '0' (not present) to '6' (severe).[20]

Computerised dynamic posturography

The NeuroCom Smart Balance Master SOT (NeuroCom, Clackamas, Oregon, USA) is a computerised assessment of balance. Each participant completed (3) 20 s trials of 6 different conditions (18 trials total) in order to determine a composite equilibrium (balance) score and somatosensory, visual, and vestibular ratios, which assist clinicians in determining the extent of postural instability.[21] The composite equilibrium score is calculated by summing the average equilibrium scores of SOT conditions 1 and 2 and then adding each individual SOT trial equilibrium score of the remaining 12 SOT trials from conditions 3 to 6. The summed value is then divided by fourteen. This value accounts for the 12 SOT trials for conditions 3–6. An additional two trials are added to represent the average scores from SOT conditions 1 and 2.[19] The SOT sensory ratios are calculated by dividing the average equilibrium score of one or more SOT conditions by the average equilibrium score of or more SOT conditions which are representative of the sensory input in question. For example, the somatosensory ratio consists of the average equilibrium score of SOT condition 2 (eyes closed, fixed support and surround) divided by the average equilibrium score of SOT condition 1 (eyes open, fixed support and surround).[19] The SOT sensory ratios are suggested to provide insight into the somatosensory, visual and vestibular interaction associated with postural stability.[22] The 18 SOT trials were administered in a randomised order which was determined by participants selecting one of four slips of paper, which contained a computer-generated randomised order of SOT trials. The total duration of the SOT was approximately 15 min.

Symptoms

To assess symptoms, we used an earlier version of the HIS-r that consists of 22 symptoms related to SC[23] which was later modified by Piland et al.[23] [24] Participants were first asked to circle yes/no if they had experienced one or more of the listed symptoms between the time of their injury and their assessment. If a participant responded 'yes' to any of the 22 items, they were then asked to rate each symptom on duration and severity using a six-point Likert scale. Duration of each symptom ranged from '1' *brief* to '6' *consistent* and severity ranged from '0' *not severe* to '6' *severe*. The sums of the symptom duration (0–132) and severity (0–132) columns for all 22 symptoms were included in our analyses. The total duration of the HIS-r was approximately 2 min.

Procedure

Individual baseline assessments took place prior to the start of all participants' respective sport seasons individually in quiet, controlled university research laboratories with minimal distractions. The baseline assessment consisted of the ImPACT, SOT and HIS-r along with the collection of demographic information, which included a self-reported concussion history. Once consent was obtained, participants completed the demographic/health history questionnaires and HIS-r. Next, athletes were with either ImPACT or the SOT and then switched to complete the other measure. The total completion time for the baseline assessment was approximately 60 min.

Athletes returned to the research laboratory within 24 h of a diagnosed SC. Concussed athletes completed a detailed medical history and were administered the aforementioned test battery. All SCs were diagnosed by certified athletic trainer(s) (ATC) and/or physician(s). For this study, concussion was defined in concordance with the AAN definition.[25]

Statistical procedures

Preliminary demographic (age, height, weight, years of education) and outcome variables for the ImPACT, SOT and HIS-r were examined using an analysis of variance. Paired t tests were used to assess differences for each outcome variable between baseline and postinjury time points for the concussed group. A multivariate analysis of variance (MANOVA), specifically predictive discriminant analysis (PDA) was used to calculate the sensitivity and specificity of each clinical measure separately and combined with the remaining tests.[26] PDA is one method to predict group membership (eg, concussed or healthy) based on calculated classification rules, in this case using ImPACT, SOT and HIS-r data. In order to develop classification rules, priors are needed to establish the likelihood of participants being classified into the concussed or healthy comparison group.[26] *Prior* values are based on related literature or theoretical reasoning and are used to minimise misclassification of participants. If limited or no evidence exists regarding which set of prior values are to be used, the values of (0.5, 0.5) are recommended.[26] Priors set at these values suggest participants will have an equal probability of being classified into either group. For our analyses, the external rule, which uses data from one set of participants to classify another set, was applied for PDA interpretation.[26] Specifically, we used the leave-one-out method,[26] which removes one participant and then calculates a linear classification function (LCF) based on the remaining sample's data. Using the calculated LCF, the removed participant is classified into one of the two groups (eg, healthy or concussed). The removed participant is then returned to the sample and the process is repeated multiple times resulting in the calculation of sensitivity, specificity and overall classification rate.[26]

Descriptive discriminant analysis (DDA) was used to determine which outcome variable(s) possessed the highest correlation to SC and discerned between concussed and healthy participants. Linear discriminant functions (LDFs) are used to describe which variable(s) derived from the ImPACT, SOT and HIS-r were able to

best distinguish between concussed and healthy participants. An outcome variable with a higher LDF value is interpreted as having the ability to discern between groups compared with variables with lower LDF values. Structured rs were used to determine which variable(s) demonstrated the highest correlation to the presence or absence of SC. Structured rs are interpreted similar to correlation coefficients which range from 0 to 1.[26]

Last, I scores were calculated for each test separately and when included in the battery to determine how much better than chance participants were correctly classified. An I score is defined as:

$$I = \frac{H_o - H_e}{1 - H_e}$$

Where H_o is equal to the observed hit rate, H_e is the hit rate expected by chance.[26] Results of the PDA and DDA were interpreted using the linear rule.[26] All statistical analyses were performed using IBM SPSS (Armonk, New York, USA) V.22 and with α=0.05.

To complement the results of our MANOVA, we used the methodology described by Broglio et al[10] to calculate sensitivity. For this analysis, only the concussed sample was used to calculate sensitivity as the healthy group was not reassessed following their baseline assessment. Sensitivity was calculated for each clinical measure independently and as a battery. For the HIS-r and the SOT, clinical change was defined as a concussed participant scoring beyond one SD of the concussed and healthy samples (N=80) baseline values.[10 27] For ImPACT, we used the software automated 80% reliable change index (RCI) for each neurocognitive domain (ie, verbal and visual memory, reaction time, and visual motor speed) with and without TSS values.[18] To calculate sensitivity, concussed participants that were observed to have significant clinical change(s) served as the numerator and the total number of concussed participants served as the denominator.

RESULTS

Over the 10-year study period, a total of 109 student-athletes were diagnosed with a SC. Of this sample, 40 concussed athletes met all of our inclusion criteria and were compared with 40 healthy controls similar in demographic characteristics. The primary rationale for exclusion of concussed participant data was the absence of multiple ImPACT, SOT and/or symptom scores and/or failure to be evaluated ≤24 h of injury. Concussed participants consisted of football (70%), women's basketball (10%), cheerleading (2.5%), women's soccer (5.0%), equestrian (5.0%), women's gymnastics (5.0%) and baseball (2.5%). Both groups were similar in terms of demographic variables with the exception of age, wherein the concussed participants were slightly older ($F_{(1,78)}$=17.09, p<0.001). Descriptive data may be found in table 1. Approximately 60.0% (n=24) of concussed participants had no prior history of concussion, 27.5% (n=11) had a history of one concussion, 10% (n=4) reported two prior

Table 1 Means and (SDs) for concussed and healthy participant demographic information

	n	Age (baseline)	Height (cm)	Years of education
Concussed				
Males	29	20.4 (1.62)	186.0 (5.90)	12.8 (1.32)
Females	11	19.6 (1.46)	165.79 (7.63)	12.5 (1.04)
Total	40	20.2 (1.60)*	180.47 (11.12)	12.7 (1.24)
Control				
Males	29	19.1 (0.94)	183.6 (6.51)	12.6 (0.94)
Females	11	18.7 (0.92)	167.2 (13.14)	12.5 (0.93)
Total	40	19.0 (0.93)*	179.1 (11.39)	12.6 (0.93)

*p≤0.001.

concussions, and 2.5% (n=1) had three or more previous concussions. Healthy control participants self-reported no prior history of SC.

Preinjury and postinjury comparison
In terms of the postinjury comparison, several significant differences were observed between injured and healthy participants. For ImPACT, the TSS was statistically different between groups ($F_{(1,78)}$=39.56, p<0.001). In terms of the SOT, concussed participants had an approximate 12% improvement for the vestibular ratio compared with their baseline values ($t_{(39)}$=−6.47) and when compared with the matched control group ($F_{(1,78)}$=26.90, p<0.001). As expected, concussed athletes also self-reported significantly higher total symptom duration ($F_{(1,78)}$=74.99, p<0.001) and total symptom severity ($F_{(1,109)}$=69.07, p<0.001) on the HIS-r compared with the healthy comparison group. When comparing the concussed participants baseline and postinjury data, we observed significantly higher levels of symptom duration ($t_{(39)}$=−8.66, p<0.001) and severity ($t_{(39)}$=−8.31, p<0.001), a significant decrease on the SOT vestibular ratio ($t_{(39)}$=−6.466, p<0.001) and ImPACT's Visual Memory Composite score ($t_{(39)}$=3.48, p=0.001), and a significant increase on ImPACT's TSS ($t_{(39)}$=−6.04, p<0.001) Descriptive data for the baseline HIS-r composite/summary scores are presented in table 2.

Descriptive discriminant analysis
Results of our DDA for the ImPACT, SOT and HIS-r may be found in table 3. For the battery of tests (HIS-r, ImPACT and SOT) the SOT vestibular ratio (0.95), HIS-r total symptom duration (0.59) and the SOT composite (−0.55) score distinguished between concussed and healthy athletes within 24 h of concussion diagnosis. The HIS-r total duration (r=0.68) and severity (0.65) were observed to have the strongest relationships to SC at the same time point. For ImPACT alone, the standardised LDFs suggest that the concussed and healthy control groups were most clearly differentiated based on TSS (0.99), and the TSS possessed the strongest relationship (r=0.97) to SC. On removing ImPACT's TSS, composite visual memory (−0.94) discriminated between

Table 2 Means and (SDs) for the ImPACT, SOT and HIS-r composite scores

Composite score control (n=40)	Concussed (n=40)		Baseline
	Baseline	Postinjury	
ImPACT			
Verbal memory	89.0 (8.56)	86.6 (10.21)	85.6 (8.21)
Visual memory	79.9 (12.78)	72.3 (13.26) *	75.9 (10.49)
Visual motor speed	36.7 (7.50)	37.3 (6.83)	37.0 (6.68)
Reaction time	0.57 (0.09)	0.60 (0.10)	0.58 (0.07)
Total symptom score	1.3 (1.62)	12.7 (12.11)*‡	0.65 (1.55)
SOT			
Composite	82.8 (4.06)	83.1 (5.27)	84.1 (3.61)
Somatosensory ratio	96.7 (2.75)	96.6 (4.86)	96.2 (2.44)
Visual ratio	93.7 (3.86)	93.8 (5.70)	92.1 (12.18)
Vestibular ratio	77.9 (7.04)	88.9 (8.88)*†	79.1 (8.05)
HIS-r			
Symptom severity	0 (0.00)	21.2 (15.47)*‡	0 (0.00)
Symptom duration	0.03 (0.16)	19.2 (14.61)*‡	0 (0.00)

*p≤0.05 for the paired t test comparing baseline to postinjury assessments for participants diagnosed with a SC.
†p≤0.05. ‡p≤0.001 and when comparing concussed participants to healthy controls.
ImPACT, Immediate Postconcussion Assessment and Cognitive Test; SC, sport concussion.

injured and healthy participants and had strongest relationship (r=−0.69) to the classification of concussion.

Table 3 Results of the descriptive discriminant analysis

	Standardised LDF		Structured r
Test battery			
Verbal memory	0.09	–	0.04
Visual memory	−0.24	–	−0.10
Visual motor speed	−0.01	–	0.02
Reaction time	−0.13	–	0.05
Total symptom score	0.17	–	0.49
Composite score	−0.55	–	−0.08
Somatosensory ratio	0.11	–	0.04
Visual ratio	0.02	–	0.06
Vestibular ratio	0.95	–	0.41
Total symptom duration	0.59	–	0.69
Total symptom severity	0.16	–	0.65
ImPACT composite score			
Verbal memory	0.18	0.54	0.07
Visual memory	−0.05	−0.94	−0.20
Visual motor speed	−0.05	0.39	0.04
Reaction time	−0.12	0.51	0.10
Total symptom score	0.99	NA	0.97
SOT composite score/ratios			
Composite score	−0.98	–	−0.13
Somatosensory	0.17	–	0.07
Visual	−0.01	–	0.10
Vestibular	1.30	–	0.66
HIS-r			
Total symptom duration	0.77	–	1.0
Total symptom severity	−0.24	–	0.96

A secondary analysis was performed for ImPACT without TSS.
ImPACT, Immediate Postconcussion Assessment and Cognitive Test; LDF, Linear discriminant function; NA, not included in analysis; TSS, Total Symptom Score.

The standardised LDFs for the SOT revealed the vestibular ratio (1.30) and the composite equilibrium score (−0.984) most clearly separated concussed and healthy participants. The SOT's structured rs revealed the vestibular ratio (0.66) possessed the strongest relationship to SC. Last, the LDFs for the HIS-r revealed total symptom duration (0.77) score most effectively differentiated between concussed healthy controls. Both the HIS-r's total symptom duration and severity possessed strong correlations (0.96–1.0) with SC.

Predictive discriminant analysis

Results of our PDA may be found in table 4. Overall, the combination of ImPACT, the SOT and the HIS-r scores demonstrated the highest overall classification rate (88.8%), sensitivity (80%) and specificity (97.5%) compared with any of the individual measures. Additionally, the I score revealed that the battery of tests correctly classified concussed athletes 55% better than chance. With regard to each individual clinical measure, the HIS-r demonstrated the highest level of sensitivity (77.5%) based on total symptom duration and severity. The ImPACT possessed the lowest sensitivity (55.0%) for all neurocognitive indices and the TSS composite scores. On removing ImPACT's TSS, sensitivity decreased slightly to 52.5%. Likewise, in terms of specificity, HIS-r composite scores possessed the highest value (100%) while ImPACT possessed the lowest (76.6%). Overall, each clinical measure and its corresponding composite/summary scores when administered independently correctly classified concussed and healthy participants approximately 5.0–55% better than chance. The entire battery of tests correctly classified concussed participants 60% better than chance. I score values may be found in table 4.

Table 4 The sensitivity and specificity of each clinical measure of sport concussion individually and as a battery using predictive discriminant analysis and clinical interpretation guidelines

PDA (N=80)							Clinical interpretation guidelines (n=40)
Clinical measure	Sensitivity (%)	Specificity (%)	Overall classification Rate (%)	I score (%)	Positive likelihood ratio	Negative likelihood ratio	Sensitivity (%)
Symptoms (HIS-r)	77.5	100.0	88.8	60.0‡	32.0	0.21	97.5
ImPACT	55.0	97.5	76.3	10.0‡	22.0	0.46	95.0
ImPACT without TSS	52.5	52.5	52.5	5.0	1.11	0.90	75.5
SOT	72.5	85.0	78.8	45.0‡	4.83	0.32	55.0
Test battery	80.0	97.5	88.8	55.0‡	0.86	2.25	100.0

‡p≤0.05.
ImPACT, Immediate Postconcussion Assessment and Cognitive Test; PDA, predictive discriminant analysis; TSS, Total Symptom Score.

Clinical interpretation guidelines

Our analysis using clinical interpretation rules to calculate sensitivity revealed the entire test battery to possess a sensitivity of 100%. Individually, the HIS-r possessed the highest sensitivity (97.5%) when correctly classifying athletes diagnosed with a SC followed by ImPACT with the TSS (95.0%) and without (75.5%) and the SOT (55%). The results of our sensitivity analysis using clinical interpretation guidelines may be found in table 4 and figure 1.

DISCUSSION

Our current study was novel because we used PDA to calculate sensitivity and specificity and to derive interpretive guidelines for each SC metric and our data show that combing scored from the ImPACT, SOT and HIS-r yields the highest sensitivity, specificity and classification rate values for discriminating between healthy college athletes and athletes who sustain a SC. Our current data are consistent with the existing body of literature that argues against using single, stand-alone measures to manage SC.[28] During the past two decades, several sports concussions advisory panels have recommended suing a battery of tests to assess athletes diagnosed with SC.[1–3 29] Despite these recommendations several factors including, limited resources, a lack of psychometric evidence,[30] observed misclassification rates,[31–33] and

random and/or systematic error have slowed the application of multifactorial assessment of SC.[28]

Our data are not entirely consistent with previous research on sensitivity of SC metrics. For instance, Broglio et al[10] reported CNTs (the ImPACT and the HeadMinder Concussion Resolution Index) possessed the highest sensitivity (79.2–78.6%) followed by self-reported symptoms (68%) and the SOT (61.9%). In order to examine the discrepancies between Broglio et al[10] and our study, we replicated the statistical approach the investigators used in order to calculate sensitivity using the current sample. In order to calculate ImPACT's sensitivity, the Broglio et al utilised reliable change indices to detect meaningful clinical change with and without TSS. Meaningful clinical change for both self-reported symptoms and postural stability were based on scoring beyond one SD of the average values of unpublished normative data.[10] Replication of this methodology in the current study again resulted in the recommended battery of tests possessing the highest sensitivity. When assessed independently using clinical interpretation guidelines, the HIS-r was observed to have the highest sensitivity, followed by the ImPACT with and without the TSS and then the SOT. Our sensitivity values are similar to those reported by Broglio et al[10] in terms of the sensitivity of ImPACT and the SOT. In contrast, we found the HIS-r possessed the highest sensitivity of the administered clinical measures. Rationale for the discrepancy is that

Figure 1 The sensitivity of ImPACT, the SOT and HIS-r administered individually and combined. (A) Sensitivity as calculated using predictive discriminant analysis. (B) Sensitivity as calculated using clinical interpretation guidelines (ImPACT, Immediate Postconcussion Assessment and Cognitive Test).
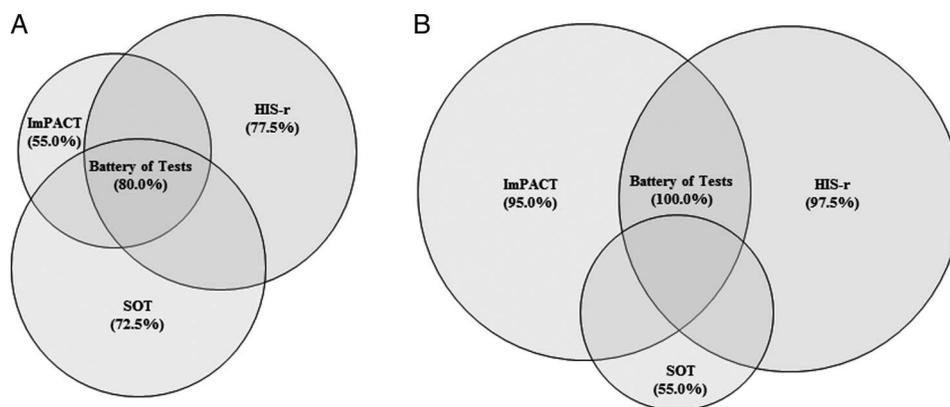
Broglio et al[10] used a revised nine-item HIS-r opposed to the 22-item inventory employed in the current study. The additional 13 items, which included somatic symptomology such as 'neck pain' and 'numbness or tingling' may have resulted in our observed sensitivity.

While a brief symptom inventory is most sensitive in detecting SC, clinicians must remember the definition of evidence-based practice and use their clinical expertise/experience to understand the limitations of this result. Evidence-based practice is defined as the 'integration of the best research evidence *with* clinical expertise and patient values to make clinical decisions'.[34] First, our participants completed the HIS-r after being removed from play and clinically diagnosed with a SC. Therefore, the athletes enrolled in the current study were potentially more forthright with their symptoms than if these data were collected prior to a diagnosis. Supplemental research and clinical evidence warrants extreme caution in the over-reliance on symptom questionnaires; however, since under-reporting of symptoms by athletes due to a lack of understanding of injury severity, not wanting to be withheld from competition (ie, sandbagging), and lack of awareness of injury can be an issue.[35] Furthermore, it is important to note that concussion-related symptomology have been demonstrated to resolve prior to the resolution of neurocognitive deficit.[36]

Time of evaluation following the diagnosis of SC may also partially explain differences between our findings and those of other studies such as Register-Mihalik et al, since athletes were assessed concussed athletes up to 5 days following injury in their study, whereas we conducted assessments within 24 h following injury. Those authors reported the SOT possessed greater sensitivity than the automated neuropsychological assessment metrics (ANAM),[11] which contrast somewhat with existing research addressing postural deficiencies following SC, which shows that college athletes return to baseline SOT values in ≤5 days.[22 37] In any case, the authors did not report average postinjury time, but based on existing research, the SOT composite score would most likely be less sensitive as time between SC and follow-up evaluation increases.

Our data show that the SOT composite equilibrium score and the vestibular sensory ratio distinguished between concussed and healthy participants possessed the strongest correlations (−0.28 to 0.41) to SC compared with the remaining SOT sensory ratios when administered within 24 h of diagnosis. In contrast, we found that concussed athletes evidenced a significantly higher value compared with their baseline assessment and when compared with healthy control participants, which diverges from existing research. One explanation for this finding may be increased effort during the postinjury evaluation. Though statistical significance was not observed, concussed participants achieved a slightly higher SOT composite equilibrium score which also may reflect increased motivation potentially as a result of the

desire to return-to-play. A more likely explanation for the improvement in the SOT vestibular ratio is practice effects associated with repeated administration.[38] Because we did not repeat SOT assessments in the control group, we are unable to empirically address this explanation for our findings; however, Peterson et al[38] reported concussed collegiate athletes returned to baseline vestibular ratio values within 48 h of their injury, and by day 3, concussed participants scored approximately five points above their baseline performance.

One of the largest discrepancies between our results and previous findings was the sensitivity of the ImPACT. ImPACT's sensitivity to SC effects has been reported to range from 79.2% to 91.4%.[39] When comparing the sensitivity calculated via PDA and clinical classification rules, ImPACT correctly classified 55–95% of athletes diagnosed with SC, respectively.[10 12 13] When removing the TSS and analysing solely ImPACT's cognitive indices, sensitivity ranged from 52.5% to 75.5%. For our PDA, our methodology was similar to that described by Schatz et al.[12 13] Though not specifically reported, Schatz employed DDA to calculate both discriminating factors between concussed and non-concussed participants as well as sensitivity and specificity.[12 13] In the current study, we employed DDA and PDA along specific rule criteria to determine which variables best discriminated between groups, possessed the strongest correlations with SC, and to calculate sensitivity and specificity. The omission of MANOVA interpretation rules, prior values and additional statistical information makes it difficult to determine why our results differed from those reported by Schatz et al.[12 13] When comparing our results to those of Broglio, we observed a similar sensitivity for the ImPACT with and without the TSS. An explanation for the discrepancy between the sensitivity calculated using PDA and the clinical interpretation guidelines is the omission of the reliable change indices with the latter statistical technique. Reliable change indices account for measurement error when comparing baseline-retest difference scores.[40 41] By not taking into account the CIs generated by this statistical technique, PDA provides a more conservative estimate of sensitivity. Despite this conservative analysis, ImPACT's specificity was similar between both PDA and when using ImPACT's clinical interpretation guidelines. Our results may also reflect the variability associated with systematic and/or random error.

Additional rationale for the discrepancies between our findings and related literature is sample composition. In terms of participants, we used a similarly matched control sample. Prior studies either used a control group which did not closely match their concussed participants in terms of gender, sport, and/or history of learning disabilities and/or attention deficit disorder (ADD)/attention deficit hyperactivity disorder (ADHD) or used no control group at all.[10 12 13] Factors such as gender, learning disabilities, sport, body mass index and ADD/ADHD may influence neurocognitive test performance,

postconcussion symptom reporting and postural stability. One or more of these variables may have influenced test performance and sensitivity.[42] [43] Based on this evidence, we were cognisant of these factors when selecting our control sample and were able to achieve statistical equivalence on all variables with the exception of age. Athletes diagnosed with SC were inherently older than the control sample since the comparison data were collected from baseline data. Despite a statistical difference in participant age between injured and healthy participants, on average age differed by no more than approximately 2 years.

Again, timing of the postinjury neurocognitive evaluation may have also contributed to the difference in our findings compared with related literature. In the current study, ImPACT data were collected on athletes within 24 h of diagnosis of a SC which was similar to the methodology employed by Broglio et al.[10] A limited number of studies have reported neurocognitive deficits may not be fully evident until 5 days following injury.[44] This may partially explain why ImPACT's TSS most effectively discerned between injured and healthy participants and showed the highest correlation to SC in the absence of statistical differences for ImPACT's cognitive indices. This may also explain why our sensitivity was significantly less for ImPACT compared with other studies when using PDA.[12] [13] Schatz employed a more liberal time frame (≤3 days of injury) when conducting his sensitivity analysis. The employed 72 h rather than 24 h testing period may explain the differences in our results.[11–13] These time frames may be more appropriate when using younger participants due to the delayed onset of symptoms and neurocognitive deficits in some cases.

In 2013, Lynall et al[45] reported 21% of surveyed certified athletic trainers abided by their governing body's position statement with regard to implementing neurocognitive testing, balance and self-reported symptoms to assess athletes diagnosed with SC. Our results emphasise the importance of utilising a multidimensional and interdisciplinary assessment of SC.[45] When considering the overall classification rate and/or solely the sensitivity of a battery of tests, the error rate was reduced by approximately 10–47.5% compared with administering any one clinical measure with the exception of a symptom inventory which possesses clinical limitations (eg, under-reporting, lack of concussion-related education, etc).[10] These findings may further support healthcare professionals, such as athletic trainers, in their requests for additional resources and/or community resources to implement all three types of measures into their concussion management protocol. Additionally, healthcare professionals such as athletic trainers and physicians must account for the limitations of each clinical measure in order to ensure proper interpretation. Accordingly, clinical neuropsychologists should be incorporated into a SC management protocol to assist with interpretation of CNT results in order to account for the

previously reported suboptimal reliability, clinical validity, misclassification rates and limited sensitivity.[10] [15] [31] [33] [46] [47] Ultimately, the incorporation of a multidimensional evaluation will assist clinicians in accounting for these caveats of each clinical measure.

Our study is not without its limitations. Our methodology included concussed participants who were evaluated within 24 h of their diagnosis. These criteria limited our sample size and ability to tightly match each concussed athlete to a healthy control. Another limitation is that only concussed participants were assessed at two different time points (ie, baseline and ≤24 h postinjury) as dictated within each institution's SC concussion management protocol. Future research addressing the sensitivity and specificity of the aforementioned measures should include administration of the investigated clinical measures to the control group at similar time points as injured participants to minimise the influence of practice effects. Additionally, we employed clinical measures that were available at each institution. These resources, particularly the ImPACT and SOT may not be available and/or feasible to implement at all venues due to limited resources such as cost and/or time. That said, a survey administered to certified athletic trainers revealed 93% of athletic trainers use the ImPACT as a CNT to manage SC while the remaining 7% used CogState, ANAM, or other computerised neurocognitive applications.[48] The incorporation of a balance measure and independent symptom scale in addition to a CNT such as ImPACT may result in a similar level of sensitivity and specificity as observed in our study. In terms of our balance assessment we used the SOT, a sophisticated and expensive measure of postural stability. Though multiple studies have supported the SOT's use for the management of SC,[22] the time and cost associated with the administration of this test may be prohibitive for the majority of clinicians who routinely assess SC. Concussion history may also have influenced the findings of the current study. In the current study, 40% of participants in the concussed group had history of one or more concussions. Schatz et al[49] reported an increased symptom burden on secondary school athletes with a history of one or more SCs. Provided the concussed group in the current study consisted of a convenience sample, it is possible prior history of concussion may have influenced our results. That said, no significant differences were observed between concussed and control groups in terms of any cognitive, motor or symptom score at the baseline assessment. Overall, future research should address the sensitivity and specificity of various combinations of clinical measures (ie, computerised neurocognitive and balance) of SC in adult and young athletes which are more cost-effective and time effective and overall more pragmatic for routine clinical use.

CONCLUSION

The purpose of the current study was to determine the sensitivity and specificity of a multidimensional approach to assess SC in collegiate athletes. Sensitivity and specificity were calculated using advanced statistical techniques and clinical interpretation guidelines. Our results demonstrate, regardless of the statistical technique employed, that a multidimensional approach consisting of ImPACT, the SOT and the HIS-r increased sensitivity by approximately 2.5–45% compared with the administration of any one clinical measure. Clinicians may use the results of the current study and related research to support requests for resources and policy development to implement a multimodal approach to SC management at all levels of sport across a variety of settings. The incorporation of clinical measures of neurocognition, balance and self-reported symptoms reduces the error associated with correctly classifying concussed and healthy athletes and may ultimately help prevent erroneous return-to-play decisions.

REFERENCES

1. Giza CC, Kutcher JS, Ashwal S, *et al*. Summary of evidence-based guideline update: Evaluation and management of concussion in sports: report of the Guideline Development Subcommittee of the American Academy of Neurology. *Neurology* 2013;80:2250–7.
2. Harmon KG, Drezner JA, Gammons M, *et al*. American Medical Society for Sports Medicine position statement: concussion in sport. *Br J Sports Med* 2013;47:15–26.
3. McCrory P, Meeuwisse WH, Aubry M, *et al*. Consensus statement on concussion in sport: the 4(th) International Conference on Concussion in Sport held in Zurich, November 2012. *J Am Coll Surg* 2013;216:e55–71.
4. Broglio SP, Cantu RC, Gioia GA, *et al*. National Athletic Trainers' Association position statement: management of sport concussion. *J Athl Train* 2014;49:245–65.
5. Anonymous. NATA Position Statement Disclaimer. Secondary NATA Position Statement Disclaimer 2014. http://www.nata.org/position-statements
6. Ferrara MS, McCrea M, Peterson CL, *et al*. A survey of practice patterns in concussion assessment and management. *J Athl Train* 2001;36:145–9.
7. Notebaert AJ, Guskiewicz KM. Current trends in athletic training practice for concussion assessment and management. *J Athl Train* 2005;40:320–5.
8. Zonfrillo MR, Master CL, Grady MF, *et al*. Pediatric providers' self-reported knowledge, practices, and attitudes about concussion. *Pediatrics* 2012;130:1120–5.
9. Crocker L, Algina J. *Introduction to classical and modern test theory*. Mason, Ohio: Cengage Learning, 2008.
10. Broglio SP, Macciocchi SN, Ferrara MS. Sensitivity of the concussion assessment battery. *Neurosurgery* 2007;60:1050–7; discussion 57–8.
11. Register-Mihalik JK, Guskiewicz KM, Mihalik JP, *et al*. Reliable change, sensitivity, and specificity of a multidimensional concussion assessment battery: implications for caution in clinical practice. *J Head Trauma Rehabil* 2013;28:274–83.
12. Schatz P, Pardini JE, Lovell MR, *et al*. Sensitivity and specificity of the ImPACT Test Battery for concussion in athletes. *Arch Clin Neuropsychol* 2006;21:91–9.
13. Schatz P, Sandel N. Sensitivity and specificity of the online version of ImPACT in high school and collegiate athletes. *Am J Sports Med* 2013;41:321–6.
14. Van Kampen DA, Lovell MR, Pardini JE, *et al*. The "value added" of neurocognitive testing after sports-related concussion. *Am J Sports Med* 2006;34:1630–5.
15. Schatz P, Ferris CS. One-month test-retest reliability of the ImPACT test battery. *Arch Clin Neuropsychol* 2013;28:499–504.
16. Schatz P, Moser RS. Current issues in pediatric sports concussion. *Clin Neuropsychol* 2011;25:1042–57.
17. Lau BC, Kontos AP, Collins MW, *et al*. Which on-field signs/symptoms predict protracted recovery from sport-related concussion among high school football players? *Am J Sports Med* 2011;39:2311–18.
18. Lovell M. Immediate Post-Concussion Assessment Testing (ImPACT) test clinical interpretive manual: online ImPACT 2007–2012. https://impacttest.com.
19. Thompson M. *Balance manager systems: clinical interpretation guide*. Clackamas, Oregon: NeuroCom International Inc., 2009.
20. Lovell M. *ImPACT 2007 (6.0) software user's manual*. Pittsburgh, PA: ImPACT Applications Inc., 2007.
21. Resch JE, May B, Tomporowski PD, *et al*. Balance performance with a cognitive task: a continuation of the dual-task testing paradigm. *J Athl Train* 2011;46:170–5.
22. Guskiewicz KM. Balance assessment in the management of sport-related concussion. *Clin Sports Med* 2011;30:89–102, ix.
23. Piland SG, Motl RW, Ferrara MS, *et al*. Evidence for the factorial and construct validity of a self-report concussion symptoms scale. *J Athl Train* 2003;38:104–12.
24. Piland SG, Motl RW, Guskiewicz KM, *et al*. Structural validity of a self-report concussion-related symptom scale. *Med Sci Sports Exerc* 2006;38:27–32.
25. [No authors listed]. Practice parameter: the management of concussion in sports (summary statement). Report of the Quality Standards Subcommittee. *Neurology* 1997;48:581–5.
26. Huberty CJ, Olejnik S. *Applied MANOVA and discriminant analysis*. 2nd edn. Hoboken, NJ: Jouhn Wiley and Sons, Inc., 2006.
27. Fritz JM, Wainner RS. Examining diagnostic tests: an evidence-based perspective. *Phys Ther* 2001;81:1546–64.
28. McCrory P, Meeuwisse WH, Kutcher JS, *et al*. What is the evidence for chronic concussion-related changes in retired athletes: behavioural, pathological and clinical outcomes? *Br J Sports Med* 2013;47:327–30.
29. Guskiewicz KM, Bruce SL, Cantu RC, *et al*. National Athletic Trainers' Association position statement: management of sport-related concussion. *J Athl Train* 2004;39:280–97.
30. Randolph C, McCrea M, Barr WB. Is neuropsychological testing useful in the management of sport-related concussion? *J Athl Train* 2005;40:139–52.
31. Broglio SP, Ferrara MS, Macciocchi SN, *et al*. Test-retest reliability of computerized concussion assessment programs. *J Athl Train* 2007;42:509–14.
32. Broglio SP, Macciocchi SN, Ferrara MS. Neurocognitive performance of concussed athletes when symptom free. *J Athl Train* 2007;42:504–8.
33. Resch J, Driscoll A, McCaffrey N, *et al*. ImPact test-retest reliability: reliably unreliable? *J Athl Train* 2013;48:506–11.
34. Sackett DL, Rosenberg WM, Gray JA, *et al*. Evidence based medicine: what it is and what it isn't. *BMJ* 1996;312:71–2.
35. McCrea M, Hammeke T, Olsen G, *et al*. Unreported concussion in high school football players: implications for prevention. *Clin J Sport Med* 2004;14:13–17.
36. Lovell MR, Collins MW, Iverson GL, *et al*. Recovery from mild concussion in high school athletes. *J Neurosurg* 2003;98:296–301.
37. Guskiewicz KM, Perrin DH, Gansneder BM. Effect of mild head injury on postural stability in athletes. *J Athl Train* 1996;31:300–6.
38. Peterson CL, Ferrara MS, Mrazik M, *et al*. Evaluation of neuropsychological domain scores and postural stability following cerebral concussion in sports. *Clin J Sport Med* 2003;13:230–7.
39. Resch JE, McCrea MA, Cullum CM. Computerized neurocognitive testing in the management of sport-related concussion: an update. *Neuropsychol Rev* 2013;23:335–49.
40. Iverson GL, Lovell MR, Collins MW. Interpreting change on ImPACT following sport concussion. *Clin Neuropsychol* 2003;17:460–7.

41. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol* 1991;59:12–19.

42. Covassin T, Elbin RJ, Harris W, et al. The role of age and sex in symptoms, neurocognitive performance, and postural stability in athletes after concussion. *Am J Sports Med* 2012;40:1303–12.

43. Elbin RJ, Kontos AP, Kegel N, et al. Individual and combined effects of LD and ADHD on computerized neurocognitive concussion test performance: evidence for separate norms. *Arch Clin Neuropsychol* 2013;28:476–84.

44. Field M, Collins MW, Lovell MR, et al. Does age play a role in recovery from sports-related concussion? A comparison of high school and collegiate athletes. *J Pediatr* 2003;142:546–53.

45. Lynall RC, Laudner KG, Mihalik JP, et al. Concussion-assessment and -management techniques used by athletic trainers. *J Athl Train* 2013;48:844–50.

46. Schatz P. Long-term test-retest reliability of baseline cognitive assessments using ImPACT. *Am J Sports Med* 2010;38:47–53.

47. Echemendia RJ, Iverson GL, McCrea M, et al. Role of neuropsychologists in the evaluation and management of sport-related concussion: an inter-organization position statement. *Arch Clin Neuropsychol* 2012;27:119–22.

48. Meehan WP III, d'Hemecourt P, Collins CL, et al. Computerized neurocognitive testing for the management of sport-related concussions. *Pediatrics* 2012;129:38–44.

49. Schatz P, Moser RS, Covassin T, et al. Early indicators of enduring symptoms in high school athletes with multiple previous concussions. *Neurosurgery* 2011;68:1562–7; discussion 1567.