Long-Term Test-Retest Reliability of Baseline Cognitive Assessments Using ImPACT

Philip Schatz,* PhD
From the Department of Psychology, Saint Joseph's University, Philadelphia, Pennsylvania

Background: Computer-based assessment programs are commonly used to document baseline cognitive performance for comparison with postconcussion testing. There are currently no guidelines for how often baseline assessments should be updated, and no data documenting the test-retest stability of baseline measures over relevant time periods.

Purpose: To establish long-term test-retest reliability of baseline assessments using ImPACT, and to compare various statistical methods for establishing test-retest reliability.

Study Design: Case series; Level of evidence, 4.

Methods: Participants were 95 collegiate varsity athletes completing baseline cognitive testing at 2 time periods, approximately 2 years apart. No participant sustained a concussion between assessments. All athletes completed the ImPACT test battery; dependent measures were the composite scores and total symptom scale score.

Results: Intraclass correlation coefficient estimates for visual memory (.65), processing speed (.74), and reaction time (.68) composite scores reflected stability over the 2-year period, with greater variability in verbal memory (.46) and symptom scale (.43) scores. Using reliable change indices and regression-based methods, only a small percentage of participants' scores showed "reliable" or "significant" change on the composite scores (0%-6%), or symptom scale scores (5%-10%).

Conclusion: The current results suggest that college athletes' cognitive performance at baseline remains considerably stable over a 2-year period. These data help establish the effects of longer, clinically pragmatic testing intervals on test-retest reliability.

Clinical Implications: The current results suggest that stretching the time between baseline assessments from 1 to 2 years may have little effect on the clinical management of concussions in collegiate athletes. These results should not be generalized to collegiate football players, who were not included in this sample. Youth athletes (high school and younger) should continue to receive annually updated baseline assessments until prospective study of the stability of baseline assessments for this younger age group can be completed.

Keywords: concussion; neurocognitive testing; neuropsychological testing; baseline assessment

Repeated neuropsychological assessment is widely used for identifying the effects of concussion on cognitive functioning; baseline, or preseason, performance is compared with postconcussion abilities. The multicenter Virginia Football Study[2,6] was the first to use within-subjects comparison to assess changes from preseason baseline performance, along with between-groups comparisons to matched cohorts. This model was soon adopted in many colleges and universities,[32] as well as the National Hockey League and National Football League.[27,29] Computer-based tests were subsequently introduced as an alternative to paper-based measures for the purpose of assessing/screening athletes for effects of concussion on cognitive performance.[18,45] Expert consensus has since identified the need for, and benefit of, baseline preinjury testing and serial follow-up as necessary components of neuropsychological test batteries.[4,34]

Determining when to return an athlete to play after a concussion is of particular importance to sports medicine professionals, who have been termed the *final arbiters* of return to play,[29] and has been widely discussed over the past 2 decades.[25,51] Early practice parameters recommended development of neuropsychological test measures to detect impairment associated with concussion,[1] and position statements have since recommended neuropsychological test data to assist athletic trainers in evaluating

*Address correspondence to Philip Schatz, PhD, Department of Psychology, Saint Joseph's University, 222 Post Hall, 5600 City Avenue, Philadelphia, PA 19131 (e-mail: pschatz@sju.edu).

No potential conflict of interest declared.

The American Journal of Sports Medicine, Vol. 38, No. 1
DOI: 10.1177/0363546509343805

recovery.[20] Discussions of clinical management included the importance of neuropsychological screening batteries,[29,36] and inclusion of a brief neuropsychological test battery is considered a "cornerstone of concussion" evaluation by consensus experts, is a recommended part of preseason baseline evaluations,[4,34] and is recognized as an important component contributing to return to play decisions.[35]

Despite the advantages of computerized test batteries (eg, improved ease of administration, automated scoring, decreased practice effects, and increased test-retest reliability[14]), there has been considerable discussion of practice effects as a threat to reliability and validity.[2,6,13,28] As the interval between baseline and postconcussion testing can span several months, researchers have pointed out that establishing test-retest reliability over a very short interval (eg, days) may not be sufficient to establish reliability for this application.[41]

Currently, there are no guidelines for how often baseline assessments should be repeated, and the call for establishing test-retest reliability over time intervals that are clinically relevant (even up to a time interval of years), for the purpose of comparing postconcussion and preseason baseline data,[41] has gone unheeded. Similarly, there has been no response to the request that test developers consider and outline guidelines for repeat baseline assessment when revising their measures.[44]

There are several commercially available computer-based test programs for the assessment and management of sports-related concussion. Among the most widely used of these programs is the Immediate Post-Concussion Assessment and Cognitive Testing test battery (ImPACT). The ImPACT battery was developed for the assessment of sports-related concussion in high school, collegiate, and professional athletes,[29] and was designed to minimize practice effects through the use of several alternate forms.[21] Traditionally, researchers systematically evaluate the test-retest reliability of computer-based measures by focusing on test intervals as they relate to the time period between baseline and follow-up or serial assessments (eg, 1 week) or by identifying the temporal stability of the measure over longer periods of time. To this end, previous literature has documented the test-retest reliability of the ImPACT battery over a period of 7 to 50 days. Iverson and colleagues[21] established 7-day test-retest reliability coefficients for the ImPACT composite scores and symptom scale in 56 high school students. Pearson coefficients were as follows: verbal memory (.70), visual memory (.67), reaction time (.79), processing speed (.86), and symptom scale (.65). Broglio and colleagues[8] used intraclass correlation coefficients (ICCs) to document 45- and 50-day test-retest reliability among 188 student volunteers: verbal memory (.23/.40; ICC between baseline and 45- and 50-day follow-up, respectively), visual memory (.32/.39), reaction time (.39/.51), and processing speed (.38/.61). Miller and colleagues[37] documented no decrease in ImPACT scores from preseason to postseason (approximately 4 months) in non-concussed football players over a season of repetitive contact activity, with all postseason scores falling within 80% confidence intervals (the study used repeated measures analyses, so no reliability coefficients are provided).

There is a prevailing belief that baseline assessments should be repeated after a concussion and, in the case of younger athletes, at the start of a new academic year.[12,48] Researchers have identified the need for more frequent updating of baseline assessments for younger athletes (eg, high school and earlier) compared with older athletes,[9] as athletes between 8 and 15 years of age are thought to be in a period of rapid cognitive maturation, during which a twice-annual baseline test has been suggested.[30,33] Independent of an athletes' age, in the absence of long-term test-retest reliability data (ie, ≥1 year), annual baseline evaluations have been recommended.[41]

To date, there has been no systematic evaluation of the long-term reliability of ImPACT over a time period that might occur between updated baseline assessments in a college athletics program. The purpose of this investigation is to establish long-term test-retest reliability of baseline cognitive performance, using ImPACT, and to compare various statistical methods for establishing test-retest reliability.

MATERIALS AND METHODS

Participants

Participants were varsity collegiate athletes participating in baseball, basketball, field hockey, lacrosse, soccer, and softball (Table 1) who completed mandatory preseason cognitive assessments as required by the athletics program. Inclusion of athlete's data in research was optional, and the institutional review board approved the procedures for participant recruitment and informed consent. As part of the growth and ongoing development of a concussion testing and management program, a decision was made to update existing baseline assessments for varsity athletes. To accomplish this task, athletes previously assessed, primarily as freshman, sophomores, and transfer students, were retested during their junior or senior year.

One hundred seventeen athletes completed preseason baseline assessments approximately 2 years apart. Twenty-two of the 117 athletes were excluded from the main analysis: 15 sustained a concussion between baseline assessments and 7 were excluded because of erratic performance on baseline assessments, as denoted by impulse control composite scores over 22.[26] The resultant sample was composed of 95 athletes who completed baseline assessments at an initial session (mean age, 18.8 years; standard deviation [SD] = 0.6), and then at a second time period approximately 2 years later (mean age, 20.8 years; mean time difference of 1.9 years; SD = 0.6). Gender distribution was nearly equal, with 51 male (54%) and 44 female (46%) participants in the study.

The ImPACT version 3.0 (ImPACT Applications Inc., Pittsburgh, Pennsylvania) was used for the study. The ImPACT instrument is a computer-based program used to assess neurocognitive function and concussion symptoms. It consists of 6 tests that evaluate attention, working memory, and processing speed, yielding composite scores on the areas of verbal memory, visual memory, processing speed, reaction time, and impulse control (see Iverson et al[21]

TABLE 1
Participation by Sport

Sport	n	%
Baseball	13	13.7
Field hockey	9	9.5
Men's basketball	3	3.2
Men's lacrosse	25	26.3
Men's soccer	15	15.8
Softball	6	6.3
Women's basketball	2	2.2
Women's lacrosse	13	13.7
Women's soccer	9	9.5
Total	95	100

for more detail on the subscales, and Schatz et al[43] for more detail on the psychometric properties of ImPACT).

Design and Analyses

Participants completing 2 baseline assessments composed a within-subjects sample, allowing for the comparison between Time 1 and Time 2. Dependent measures included the verbal memory, visual memory, processing speed, reaction time composite scores from ImPACT, and the total symptom scores.

Pearson product moment correlations (r) were calculated as a general measure of the strength of linear association between variables at Times 1 and 2. A weakness of Pearson's r as a measure of test-retest reliability is when coefficients are high and group means are similar, but there is considerable individual variation in scores from Time 1 to Time 2.[42] As such, ICCs, considered a better measure of association than Pearson's r,[49] were also calculated as the primary indicator of test-retest reliability. The ICC can distinguish those sets of scores that are merely ranked in the same order from test to retest from those that are not only ranked in the same order but are in low, moderate, or complete agreement.[11] The ICC model "2-way mixed" type, "consistency," were used; ICC analyses also yield an unbiased estimate of reliability, which reflects the consistency of the baseline assessments.[19]

Reliable change indices (RCIs)[24] were calculated to assess whether a change between repeated assessments was reliable and meaningful. The RCI provides an estimate of the probability that a given difference score would not be obtained as a result of measurement error.[23] A modified RCI formula,[10] which includes an adjustment for practice effects, was also calculated (see Barr[5] for a more detailed discussion). Finally, regression-based methods (RBMs) were applied to the data. In RBM, the scores from the first assessment are placed into a regression analysis, using the score at Time 2 as the dependent variable, with the resulting equation providing an adjustment for the effect of initial performance level, as well as controlling for any regression to the mean.[31] With this technique, regression equations can be built to predict a participant's level of performance on a neuropsychological instrument at retest from the initial testing.[17]

After Bonferroni correction, alpha level for all analyses was set at $P < .01$.

RESULTS

Mean ImPACT composite and symptom scores showed little variation between the 2 assessments (Table 2). Pearson correlations between baseline assessments ranged from .27 to .60, with no significant differences noted on paired comparison t tests. The ICCs reflected higher reliability than Pearson's r across all measures. Processing speed scores showed the most stability (mean ICC = .74; lower and upper 95% confidence intervals [CIs]: .61-.83;), followed by reaction time (mean ICC = .68; lower and upper 95% CIs: .53-.78), visual memory (mean ICC = .65; lower and upper 95% CIs: .48-.77), verbal memory (mean ICC = .46; lower and upper 95% CIs: .19-.64), and total symptom scores (mean ICC = .43; lower and upper 95% CIs: .15-.62). Unbiased estimate of reliability was consistent with ICCs: processing speed (.75), reaction time (.68), visual memory (.66), verbal memory (.47), and total symptom scores (.44).

Regression-based measures revealed that follow-up baseline scores showed considerable stability (Table 3). At follow-up baseline assessments, 95% to 97% of all composite scores and 89% of symptom scale scores fell within an 80% CI; 97% to 98% of all composite scores and 95% of symptom scale scores fell within a 95% CI.

Reliable change indices were calculated for composite and total symptom scores. Reliable change indices are presented at 80% and 95% CIs (Table 4). There was little variation between the "traditional" method (lower line in Table 4) and the "corrected" method of Chelune et al[10] (top line in Table 4). Rates of impairment using RBM and RCI are presented in Table 5. After a time interval of approximately 2 years, only a small percentage of participants' scores fell outside the range of normality, as denoted by these techniques. The RBM served as a more conservative measure, with fewer follow-up baseline scores falling in the "impaired" range, as compared with RCI. Using RCI and RBM, only a small percentage of participants' scores showed "reliable" or "significant" change on the composite scores (0%-6%) and on the symptom scale scores (5%-10%). Of note, nearly equal numbers of participants showed improvement and declines over the time period across all measures.

DISCUSSION

This study assessed test-retest reliability of baseline cognitive performance measured by the ImPACT test battery over a period of approximately 2 years, a clinically relevant and practical interval. Intraclass correlation coefficients ranged from .47 to .75 for the composite scales, and .44 for the postconcussion symptom scale. Using reliable change and standard regression-based measures, only 0% to 6% of participants showed significant changes on symptom scale scores, when comparing performance at baseline assessments conducted over a 2-year period.

TABLE 2
Test-Retest Reliability[a]

Variable	Time 1 Mean (SD)	Time 2 Mean (SD)	r	ICC	ICC 95% CI Lower	ICC 95% CI Upper	UER	t^b	Sig. t
Verbal memory	87.6 (8.3)	87.8 (9.5)	.30	.459	.188	.640	.471	–0.2	.83
Visual memory	75.6 (12.1)	78.1 (11.4)	.49	.653	.479	.769	.660	–2.0	.045
Process speed	41.2 (5.9)	42.0 (6.9)	.60	.742	.612	.828	.747	–1.4	.16
Reaction time	.54 (.06)	.53 (.07)	.52	.676	.513	.784	.682	0.9	.36
Symptom scale	9.3 (11.9)	8.9 (11.4)	.27	.431	.145	.621	.443	0.3	.80

[a] ICC, intraclass correlation coefficient; CI, confidence interval; UER, unbiased estimate of reliability; Sig. t, significance of t.
[b] df = 93; Bonferroni corrected alpha $P < .01$.

TABLE 3
Regression-Based Methods[a]

Variable	Time 1 Mean (SD)	Time 2 Mean (SD)	α	β	Sxy	80% CI (%)	95% CI (%)
Verbal memory	87.6 (8.3)	87.8 (9.5)	.577	.343	.090	94.74	97.89
Visual memory	75.6 (12.1)	78.1 (11.4)	.435	.458	.099	94.74	97.89
Process speed	41.2 (5.9)	42.0 (6.9)	13.556	.691	5.505	94.74	96.84
Reaction time	.54 (.06)	.53 (.07)	.193	.632	.064	96.84	97.89
Symptom scale	9.3 (11.9)	8.9 (11.4)	6.495	.262	10.939	89.47	94.74

[a] α = intercept; β = slope; Sxy, standard error of estimate[17]; CI, confidence interval—numbers represent the percent of participants with change scores within cutoff (80% CI: 1.65; 90% CI: 1.96).

TABLE 4
Reliable Change Indices (RCI)

Variable		Time 1	Time 2	r^a	SEM[1b]	SEM[2b]	Sdiff[c]	RCI[d]	80% CI[e]	95% CI[e]
Verbal memory	(M)	87.6	87.8	.30	6.93	7.91	10.52	2.27	13.47	17.25
	(SD)	8.3	9.5				9.81	2.44	12.55	16.08
Visual memory	(M)	75.6	78.1	.49	8.70	8.20	11.96	2.08	15.31	19.61
	(SD)	12.1	11.4				12.31	2.02	15.76	20.19
Process speed	(M)	41.2	42.0	.60	3.76	4.36	5.75	1.48	7.36	9.44
	(SD)	5.9	6.9				5.31	1.60	6.80	8.71
Reaction time	(M)	.54	.53	.52	.04	.05	0.07	.096	0.09	0.11
	(SD)	.06	.07				0.06	.107	0.07	0.10
Symptom scale	(M)	9.3	8.9	.27	10.14	9.69	14.02	2.63	17.95	23.00
	(SD)	11.9	11.4				14.34	2.57	18.35	23.52

[a] r = Pearson's correlation between Time 1 and Time 2 scores.
[b] Standard error of measure at Time 1 and Time 2 = $[SD*\sqrt{(1-r_{xy})}]$.
[c] Standard error of difference scores based on (Chelune et al,[10] upper): $\sqrt{[(SEM1^2) + (SEM2^2)]}$ and (Jacobson and Truax,[24] lower): $\sqrt{[2*(SEM1^2)]}$.
[d] Reliable change: Chelune et al,[10] upper; Jacobson and Truax,[24] lower.
[e] CI, confidence interval; numbers represent reliable change scores at 80% (1.65) and 95% (1.96) CIs.

Among the benefits of computer-based testing is the ability to control for large practice effects sometimes seen on pencil-and-paper tests,[21] and the ImPACT developers documented stability of test scores for age-matched controls over a 1-week period. Test-retest reliability over a longer, 2-week period, was subsequently documented for verbal memory (.40), reaction time (.71), and processing speed (.80).[22] Broglio and colleagues[8] reported generally low ICCs on 3 of the composite scores of ImPACT, which were lower than those reported in the literature for computer-based tests. The 2-year ICCs from the current study were higher than the 45-day scores obtained by Broglio and colleagues[8] for all of the composite scores measured: verbal memory (.46 vs .23), visual memory (.65 vs .32), processing speed (.74 vs .39), and reaction time (.68 vs .38). It is important to note that these scores[8] were

TABLE 5
Rates of Impairment (%) using Reliable Change Indices (RCI) Versus Regression-Based Model (RBM)[a]

	RCI[b]						RBM					
	80% CI[c]			95% CI[c]			80% CI[c]			95% CI[c]		
Variable	Impr	Decl	Tot	Impr	Decl	Tot	Impr	Decl	Tot	Impr	Decl	Tot
Verbal memory	6	2	8	3	2	5	0	6	6	0	2	2
	7	5	12	3	2	5						
Visual memory	11	2	13	7	2	9	2	3	5	2	1	3
	8	2	10	0	7	7						
Process speed	4	3	7	2	2	4	3	2	5	2	1	3
	5	5	10	2	2	4						
Reaction time	4	3	7	3	2	5	0	3	3	0	2	2
	5	4	9	3	2	5						
Symptom scale	3	5	8	3	3	6	0	10	10	0	5	5
	3	5	8	3	3	6						

[a]Impr, improved; Decl, declined; Tot, total.
[b]RCI: Chelune et al,[10] upper; Jacobson and Truax,[24] lower.
[c]CI, confidence interval; numbers represent percent of participants scoring beyond cutoff values at 80% (1.65) and 95% (1.96) CIs.

derived from participants required to complete 3 commercially available computerized test batteries in succession (ImPACT, Headminder's CRI [Headminder, Inc, New York, New York], CogState's Concussion Sentinel [CogState Ltd, Victoria, Australia]) at baseline, day 45, and day 50. This increased variability in performance may be attributed to the error from having participants complete similar tasks, measuring nearly identical constructs, over a considerably demanding testing session. In this regard, 35% (40 of 113) of their sample were excluded because of poor effort in comparison to only 4% (5 of 117) of the current sample.

Annual baseline assessments have been recommended for younger athletes[12,48] (eg, childhood and adolescent ages), as compared with older athletes,[9] with some test developers suggesting twice-annual baseline testing[30,33] for athletes younger than 15 years of age. Others have recommended new annual baseline measures for each player, until data regarding very long-term (ie, >1 year) reliability can be established.[41] Although the efficacy and utility of repeated baseline have yet to be established, the current results suggest that baseline data obtained from college athletes remain considerably stable over a 2-year period, and demonstrate greater stability than data collected over a 1- to 2-week period. These data help fill in the gaps in the literature, especially calls for testing the effects of longer, clinically pragmatic testing intervals on test-retest reliability.[8]

Low correlations between baseline assessments, as measured by Pearson's r, reflect limited shared variance between at Times 1 and 2. Pearson's r is known to be smaller when group means are similar, but there is considerable individual variation in scores from Time 1 to Time 2,[42] so this is not an ideal measure of the strength of association between assessments. Rather, ICC has been shown to be a better indicator of scores that are not only ranked in the same order but are also in agreement. There is considerable variability in what constitutes "acceptable" agreement in ICC: .60[3] and .70[7] have been recommended as minimum acceptable ICCs, with coefficients measuring from .50 to .75 indicating moderate reliability, and coefficients greater than .75 indicating good reliability.[40] In this sample, it is unclear why verbal memory ICCs (.46) would fall outside the range of acceptable while the other composite measures (from .65 to .74) fell within the moderate range. Symptom scores have been found to show similar variability at baseline[15,16,46] so the lower ICC (.43) is not unexpected.

Reliable change indices base the significance of a change for an individual test score on the difference between the initial and follow-up scores for the normative subject sample. The RCI allows one to establish empirically derived cutoffs that can be used for evaluating meaningful differences in test scores, independent of psychometric issues, such as practice effects and other sources of variance.[5] In this study, CIs derived from RCIs showed greater variability than those established over a 1-week test-retest period.[21] Both RCI and RBM are designed so that no more than 5% of the cases should be outside the prediction interval indicating improvement and deterioration.[47] Using RCI, only visual memory scores showed significant improvement (95% CI) over a 1 -year period, with only a small percentage of participants (0%-3%) showing either improvement or decline over this time frame. With use of RBM, none of the participants' scores showed significant change.

This study documented the stability of baseline preseason cognitive assessments over a period of 2 years. Although no athletes were included in the sample if they sustained a concussion between baseline assessments, it is important to note that data were obtained from a Division I National Collegiate Athletic Association program with no football program. Football players are at risk for sustaining unique concussive injuries, with respect to velocity and translational acceleration (eg, average 9.3 m/sec velocity and 98 g for football concussions[39] vs average 3 m/sec and 21 g for elbow strikes and 8 m/sec and 20 g for wrist/forearm strikes in soccer collisions[50]). As well, football players may also experience a number of "subclinical collisions" over the course of their athletic careers, from which recovery is relatively rapid.[38] In this regard, these results should not be generalized to collegiate football players until results can be replicated on this population.

As colleges and universities implement concussion testing programs, there will be natural growth from the testing of the initial cohort. The current results suggest that stretching the time between baseline assessments to 2 years will have little deleterious effect on the management of student athletes. This is of practical importance to team physicians who are charged with making return to play decisions for concussed college athletes, as well as clinicians charged with interpreting postconcussion cognitive test data in the context of "change from baseline." In those cases for which postconcussive test data have not been updated annually, the current results suggest that comparisons with baseline data obtained 2 years prior remain clinically relevant.

It is important to note that while the current results apply to college-level athletes, they do not generalize to younger, developing youth athletes, such as those competing in junior high schools, Little League teams, and other organized leagues. Until prospective study of the stability of baseline assessments can be completed, youth athletes at the high school level and younger should continue to receive annually updated baseline assessments. Additional prospective studies comparing postconcussion data to initial (Time 1) and follow-up (Time 2) baseline data will help establish the efficacy and utility of repeat baseline assessments in diagnosing the effects of concussion on cognitive function.

ACKNOWLEDGMENT

The author acknowledges Jeffrey Trichon, FCAS, MAAA, for assistance with data preparation.

REFERENCES

1. AAN. Practice parameter: the management of concussion in sports (summary statement). Report of the Quality Standards Subcommittee. *Neurology.* 1997;48(3):581-585.
2. Alves W, Rimel RW, Nelson WE. University of Virginia prospective study of football-induced minor head injury: status report. *Clin Sports Med.* 1987;6:211-218.
3. Anastasi A. *Psychological Testing.* 7th ed. New York: MacMillan; 1998.
4. Aubry M, Cantu RC, Dvorak J, et al. Summary and agreement statement of the 1st international symposium on concussion in sport, Vienna. *Clin J Sport Med.* 2002;12:6-11.
5. Barr WB. Neuropsychological testing for assessment of treatment effects: methodologic issues. *CNS Spectr.* 2002;7(4):300-302, 304-306.
6. Barth JT, Alves W, Ryan T, Macciocchi SN, Rimel RW, Nelson WE. Mild head injury in sports: neuropsychological sequelae and recovery of function. In: Levin HS, Eisenberg HM, Benton AL, eds. *Mild Head Injury.* New York: Oxford University Press; 1989:257-275.
7. Baumgartner TA, Chung H. Confidence limits for intraclass reliability coefficients. *Measurement in Physical Education and Exercise Science.* 2001;5:179-188.
8. Broglio SP, Ferrara MS, Macciocchi SN, Baumgartner TA, Elliott R. Test-retest reliability of computerized concussion assessment programs. *J Athl Train.* 2007;42(4):509-514.
9. Buzzini SR, Guskiewicz KM. Sport-related concussion in the young athlete. *Curr Opin Pediatr.* 2006;18(4):376-382.
10. Chelune GJ, Naugle RI, Lüders H, Sedlak J, Awad IA. Individual change after epilepsy surgery: practice effects and base-rate information. *Neuropsychology.* 1993;7:41-52.
11. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment.* 1994;6(4):284-290.
12. Collie A, Darby D, Maruff P. Computerised cognitive assessment of athletes with sports related head injury. *Br J Sports Med.* 2001;35(5):297-302.
13. Collie A, Maruff P, Darby DG, McStephen M. The effects of practice on the cognitive test performance of neurologically normal individuals assessed at brief test-retest intervals. *J Int Neuropsychol Soc.* 2003;9(3):419-428.
14. Collie A, Maruff P, McStephen M, Darby DG. Psychometric issues associated with computerised neuropsychological assessment of concussed athletes. *Br J Sports Med.* 2003;37(6):556-559.
15. Covassin T, Schatz P, Swanik CB. Sex differences in neuropsychological function and post-concussion symptoms of concussed collegiate athletes. *Neurosurgery.* 2007;61(2):345-351.
16. Covassin T, Swanik CB, Sachs M, et al. Sex differences in baseline neuropsychological function and concussion symptoms of collegiate athletes. *Br J Sports Med.* 2006;40(11):923-927.
17. Crawford JR, Garthwaite PH. Comparing patients' predicted test scores from a regression equation with their obtained scores: a significance test and point estimate of abnormality with accompanying confidence limits. *Neuropsychology.* 2006;20(3):259-271.
18. Echemendia R, ed. *Sports Neuropsychology: Assessment and Management of Traumatic Brain Injury.* New York: Guilford Press; 2006.
19. Fleishman J, Benson J. Using LISREL to evaluate measurement models and scale reliability. *Educational and Psychological Measurement.* 1987;47:925-939.
20. Guskiewicz KM, Bruce SL, Cantu RC, et al. Recommendations on management of sport-related concussion: summary of the National Athletic Trainers' Association position statement. *Neurosurgery.* 2004;55(4):891-896.
21. Iverson GL, Lovell MR, Collins MW. Interpreting change on ImPACT following sport concussion. *Clin Neuropsychol.* 2003;17(4):460-467.
22. Iverson GL, Lovell MR, Collins MW, Norwig J. Tracking recovery from concussion using ImPACT [abstract]. *Arch Clin Neuropsychol.* 2002;17:770.
23. Iverson GL, Sawyer DC, McCracken LM, Kozora E. Assessing depression in systemic lupus erythematosus: determining reliable change. *Lupus.* 2001;10(4):266-271.
24. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol.* 1991;59(1):12-19.
25. Lehman LB. Nervous system sports-related injuries. *Am J Sports Med.* 1987;15(5):494-499.
26. Lovell M. *ImPACT Version 2.0 Clinical User's Manual*. Pittsburgh, PA: ImPACT Applications Inc; 2002.
27. Lovell M, Collins MW. Neuropsychological assessment of the college football player. *J Head Trauma Rehabilitation.* 1998;13(2):9-26.
28. Macciocchi SN. 'Practice makes perfect': retest effects in college athletes. *J Clin Psychol.* 1990;46(5):628-631.
29. Maroon JC, Lovell MR, Norwig J, Podell K, Powell JW, Hartl R. Cerebral concussion in athletes: evaluation and neuropsychological testing. *Neurosurgery.* 2000;47(3):659-669; discussion 669-672.
30. Maruff P, Collie A, Anderson V, Mollica C, McStephen M, McCrory P. Cognitive development in children: implications for concussion management [abstract]. *Br J Sports Med.* 2004;38:654-655.
31. McCrea M, Barr WB, Guskiewicz K, et al. Standard regression-based methods for measuring recovery after sport-related concussion. *J Int Neuropsychol Soc.* 2005;11(1):58-69.
32. McCrea M, Guskiewicz KM, Marshall SW, et al. Acute effects and recovery time following concussion in collegiate football players: the NCAA Concussion Study. *JAMA.* 2003;290(19):2556-2563.
33. McCrory P, Collie A, Anderson V, Davis G. Can we manage sport related concussion in children the same as in adults? *Br J Sports Med.* 2004;38(5):516-519.

34. McCrory P, Johnston K, Meeuwisse W, et al. Summary and agreement statement of the 2nd International Conference on Concussion in Sport, Prague. *Br J Sports Med.* 2005;39:196-204.
35. McCrory P, Meeuwisse W, Johnston K, et al. Consensus statement on concussion in sport: the 3rd International Conference on Concussion in Sport, held in Zurich, November 2008. *J Clin Neurosci.* 2009;16(6):755-763.
36. McCrory PR. Were you knocked out? A team physician's approach to initial concussion management. *Med Sci Sports Exerc.* 1997;29(7 Suppl):S207-S212.
37. Miller JR, Adamson GJ, Pink MM, Sweet JC. Comparison of preseason, midseason, and postseason neurocognitive scores in uninjured collegiate football players. *Am J Sports Med.* 2007;35(8):1284-1288.
38. Pellman EJ, Viano DC, Casson IR, et al. Concussion in professional football: repeat injuries—part 4. *Neurosurgery.* 2004;55(4):860-873.
39. Pellman EJ, Viano DC, Tucker AM, Casson IR, Waeckerle JF. Concussion in professional football: reconstruction of game impacts and injuries. *Neurosurgery.* 2003;53(4):799-812.
40. Portney LG, Watkins MP. *Foundations of Clinical Research: Application to Practice.* Upper Saddle River, NJ: Prentice Hall; 2000.
41. Randolph C, McCrea M, Barr WB. Is neuropsychological testing useful in the management of sport-related concussion? *J Athl Train.* 2005;40(3):139-152.
42. Rodgers JL, Nicewander WA. Thirteen ways to look at the correlation coefficient. *American Statistician.* 1998;42:59–66.
43. Schatz P, Pardini JE, Lovell MR, Collins MW, Podell K. Sensitivity and specificity of the ImPACT test battery for concussion in athletes. *Arch Clin Neuropsychol.* 2006;21(1):91-99.
44. Schatz P, Putz BO. Cross-validation of measures used for computer-based assessment of concussion. *Appl Neuropsychol.* 2006;13(3):151-159.
45. Schatz P, Zillmer EA. Computer-based assessment of sports-related concussion. *Appl Neuropsychol.* 2003;10(1):42-47.
46. Swanik CB, Covassin T, Stearne DJ, Schatz P. The relationship between neurocognitive function and noncontact anterior cruciate ligament injuries. *Am J Sports Med.* 2007;35(6):943-948.
47. Temkin NR, Heaton RK, Grant I, Dikmen SS. Detecting significant change in neuropsychological test performance: a comparison of four models. *J Int Neuropsychol Soc.* 1999;5(4):357-369.
48. Valovich TC, Perrin DH, Gansneder BM. Repeat administration elicits a practice effect with the balance error scoring system but not with the standardized assessment of concussion in high school athletes. *J Athl Train.* 2003;38(1):51-56.
49. Wilk CM, Gold JM, Bartko JJ, et al. Test-retest stability of the repeatable battery for the assessment of neuropsychological status in schizophrenia. *Am J Psychiatry.* 2002;159(5):838-844.
50. Withnall C, Shewchenko N, Gittens R, Dvorak J. Biomechanical investigation of head impacts in football. *Br J Sports Med.* 2005;39(Suppl 1):i49-57.
51. Wojtys EM, Hovda D, Landry G, et al. Concussion in sports. *Am J Sports Med.* 1999;27(5):676-687.