BRAIN INJURY

http://informahealthcare.com/bij
ISSN: 0269-9052 (print), 1362-301X (electronic)

Brain Inj, 2015; 29(2): 263–275

DOI: 10.3109/02699052.2014.965214

informa healthcare

ORIGINAL ARTICLE

Advanced topics in neuropsychological assessment following sport-related concussion

Grant L. Iverson[1,2,3,4,5] & Philip Schatz[5]

[1]*Department of Physical Medicine and Rehabilitation, Harvard Medical School, Boston, MA, USA,* [2]*Spaulding Rehabilitation Hospital, Charlestown, MA, USA,* [3]*MassGeneral Hospital for Children Sport Concussion Program, Boston, MA, USA,* [4]*Red Sox Foundation and Massachusetts General Hospital Home Base Program, Boston, MA, USA, and* [5]*Department of Psychology, Saint Joseph's University, Philadelphia, PA, USA*

Abstract

Objective: This study examined seven topics relating to neuropsychological assessment following sport-related concussion: (i) traditional vs. computerized tests; (ii) the value of baseline, pre-season testing; (iii) invalid baseline scores and poor effort; (iv) when to assess following injury; (v) the reliability of neuropsychological tests; (vi) reliable change methodology; and (vii) new methods for identifying cognitive impairment.

Main results: Baseline testing can be helpful for quantifying cognitive deficits following injury and for assessing recovery. At present, however, there is insufficient evidence to conclude that having baseline test results is clearly superior to not having baseline test results. Although invalid baseline test performance can be detected in some athletes, validity indicators cannot determine the extent to which the results were due to deliberately poor performance, confusion or misunderstanding regarding some aspect of the test, distractions in group testing environments or some combination of factors. When interpreting baseline and post-injury data, sophisticated psychometric methods (e.g. reliable change, multivariate base rates) are available to assist with more accurate identification of cognitive impairment and the serial monitoring of recovery.

Conclusions: The value of neuropsychological assessment in the management of sport-related concussion has a strong empirical foundation. Additional research is needed, however, to refine its use.

Keywords

Cognitive functioning, concussion, mild traumatic brain injury, neuropsychological assessment, psychometrics, reliable change

History

Received 15 January 2014
Revised 19 April 2014
Accepted 25 April 2014
Published online 14 October 2014

Introduction

Neuropsychological assessment has been an important part of sport-related concussion research and clinical practice since the 1980s [1]. There are dozens of studies showing that both traditional and computerized neuropsychological tests are sensitive to the acute effects of concussion [2–21]. According to meta-analyses of the literature, sport-related concussions have a large adverse effect on cognition in the first 24 hours, with resolution of these deficits occurring within ~1 week according to group studies [22, 23]. There is some evidence, however, that there is an embedded sub-group of athletes who do not experience rapid recovery in cognitive functioning [10] and this sub-group might be obscured in statistical analyses applied to larger groups of athletes [24]. In fact, in a prospective study of high school football players [25, 26], ~42–47% were deemed functionally recovered by 1 week (see [26], Figure 1, p. 503) and it was not until 4 weeks that 84–94% were considered recovered. These high school football players took considerably longer to recover than college [21] or professional [27–29] football players. It is increasingly clear that there are considerable individual differences in the rate of recovery following this injury.

The purpose of this article is to examine seven topics relating to neuropsychological assessment following sport-related concussion These topics are as follows: (i) the type of tests used (traditional vs. computerized); (ii) the value of baseline, pre-season testing; (iii) invalid baseline scores and poor effort; (iv) when to assess following injury; (v) the reliability of neuropsychological tests; (vi) reliable change methodology; and (vii) new methods for identifying cognitive impairment. Throughout this article, extensive data is presented relating to ImPACT® (Immediate Post-Concussion Assessment and Cognitive Testing). The psychometric principles, issues and approaches illustrated with these data are relevant to all batteries of neuropsychological tests used in concussion management programmes.

Traditional vs. computerized testing

In clinical practice and research, some people use traditional neuropsychological tests, others use computerized tests and

Correspondence: Grant Iverson, PhD, Neuropsychology Outcome Assessment Laboratory, Department of Physical Medicine and Rehabilitation, Harvard Medical School, 79/96 Thirteenth Street, Charlestown Navy Yard, Charlestown, MA 02129, USA. Tel: 617-952-6194. E-mail: giverson@mgh.harvard.edu

some use both (termed a 'hybrid' approach) [30]. The advantages of traditional testing include increased face-to-face interaction between the test taker and examiner, allowance for the examiner to 'test the limits' or customize test measures, the assessment of verbal fluency and verbal memory and the opportunity to take breaks between tests or sub-scales. The advantages of computerized testing include the ability to test groups of athletes simultaneously, automated randomization of test stimuli (i.e. for alternate forms), more accurate assessment of response time and automatic scoring and data storage. The advantages of a hybrid approach allow for the examiner to maximize the benefits of both approaches, with a broader number of test measures across a broader range of cognitive domains.

There have been very few studies that have used traditional and computerized tests within the first week following sport-related concussion. Broglio et al. [12] found that computerized test batteries (ImPACT® and Concussion Resolution Index - CRI) yielded significantly higher sensitivity to concussion in the first 24 hours (79%), as compared to symptom reports (62%), postural control data (44%) or pencil-and-paper measures (43%). With respect to computer-based measures alone, Schatz et al. [31] and Schatz and Sandel [32] documented the sensitivity of the ImPACT® test at 82% and 91% (using the 'desktop' and 'online' versions, respectively) within 72 hours of concussion. Register-Mihalik et al. [33] documented the sensitivity of the Automated Neuropsychological Assessment Metric (ANAM) test at 50% within 5 days of concussion. With respect to pencil-and-paper measures alone, McCrea et al. [34] documented the sensitivity of neuropsychological testing at 23% within 2 days of concussion and 56% when combined with symptom report and postural control data. More research is needed that compares and contrasts the sensitivity and specificity of computerized and traditional cognitive testing following sport-related concussion. The study by Broglio et al. [12] is important and worthy of follow-up research—it suggests that computerized testing might be superior to traditional paper–pencil testing. Computerized testing can measure processing speed and reaction time much more precisely than traditional testing and it might be possible to develop computerized tests that have smaller practice effects than traditional tests, adding further precision not only to measurement but to the clinical interpretation of serial test results.

The value of baseline testing

Few would doubt that a reliable, valid and accurate assessment of a person's pre-injury cognitive functioning would be useful for determining the nature and extent of post-injury cognitive deficits and the rate of recovery. There are multiple challenges and problems, however, associated with baseline testing [35, 36]. First, baseline computerized testing is often conducted in group settings and one study illustrated that athletes tested in group settings perform more poorly than athletes tested individually [37]. This study re-analysed that data, and found that, when considering the four primary ImPACT® composite scores simultaneously, 36% of those tested in a group and 26% of those tested individually had one or more unusually low scores (i.e. below the 10th percentile). When an athlete is injured, he or she is tested individually. Therefore, some athletes' baseline test scores might be artificially lowered by being tested in a group setting. As such, some authors have recommended individual baseline testing or testing in small, carefully monitored groups (e.g. 3–5 students) [38]. More research on this topic is clearly needed.

Second, baseline testing can be fairly expensive and labour intensive, depending on how it is done. Traditional neuropsychological tests require advanced training in administration and scoring and they must be administered individually. Therefore, it is impractical or not feasible to do baseline testing in some communities and settings. Third, some tests used in concussion management programmes have modest test–re-test reliability [36, 39–41]. Therefore, it can be difficult to accurately and precisely assess baseline and post-injury change scores. This topic is discussed in detail in a section below. Finally, the value of baseline testing is largely assumed and baseline testing is often encouraged in consensus [42] or agreement [43, 44] statements—at least with some athletes. Few studies have examined this issue empirically, however.

Echemendia et al. [38] examined the clinical usefulness of having baseline testing vs. no baseline testing in a sample of 266 concussed college athletes. Reliable change from baseline was compared to an approach relying on normative data only. Athletes with ImPACT® composite scores that were 1.5 SDs below the normative mean were classified as having clinically significant deficits. Both sensitivity (0.80–0.86) and specificity (0.95–0.97) were high using normative scores. These results suggested that a large percentage of concussed athletes can be readily identified by deviations from normative data, which may be particularly useful when baseline data are not available for comparison. These findings have been replicated in a sample of non-elite rugby players who did not complete baseline assessments, in which post-concussive neurocognitive test data alone were able to identify 87% of concussed athletes [45]. Schmidt et al. [46] also compared the baseline methodology to the normative methodology for interpreting post-injury test scores in a sample of 258 concussed college student athletes, using the ANAM Battery. The individual baseline method identified more athletes with impairment on the Simple Reaction Time sub-test and the normative method identified more athletes with impairment on the Mathematic Processing sub-test. No differences were present on the other ANAM sub-tests.

With advances in electrophysiology, neuroimaging and serum-based biomarker research, it is possible that baseline testing will have less value in the future. The direct measurement of post-injury physiology, combined with other measures such as balance, cognition and subjective symptoms, might ultimately prove to be highly accurate for identifying concussion, the severity of the injury in regards to its effect on physiology and behaviour and monitoring recovery. It will likely take many years, however, before those approaches are validated for clinical use and become feasible, from a time and cost perspective, to use with amateur athletes.

It is concluded that having an accurate measure of baseline cognitive functioning would be helpful for quantifying cognitive deficits following injury and for assessing recovery. This is especially true for athletes who have above average [47] or below average cognitive functioning at baseline. Moreover, it can be helpful for athletes with developmental conditions, such as attention-deficit hyperactivity disorder or a learning disability. Students with self-reported ADHD, learning disabilities or both perform more poorly on ImPACT® than students who do not have a developmental condition [48]. At present, however, there is insufficient evidence to conclude that having baseline test results is time- and cost-effective or clearly superior to not having baseline test results. More research on the advantages, disadvantages and appropriate use of baseline testing is needed.

Invalid baseline scores and poor effort

Baseline neurocognitive test performance is considered to be an important contributor to return-to-play decision-making, serving as a comparator for post-concussion test data. Therefore, it is essential to document a valid measure of an athlete's pre-season, baseline neurocognitive performance. To this end, test developers have identified built-in measures to identify athletes who score outside of expected ranges and may be providing less-than-optimal effort. Researchers have shown that invalid baseline scores occur in a small percentage of adolescents and young adults undergoing pre-season testing with ImPACT® [49] and younger athletes tested in a group setting are more likely to obtain invalid baselines than those tested individually [50]. Specifically, 5.4% of youth athletes, aged 10–12, who were tested individually obtained invalid baselines compared to 11.9% of those tested in groups [50].

There are references in the media and literature to athletes under-reporting post-concussion symptoms, in order to return to athletic competition [51–53]. However, this concern has not been studied systematically. McCrea et al. [54] documented that 53% of high school football players did not report concussions or symptoms, due to not thinking the injury was serious enough, not wanting to be withheld from competition and not knowing they had a concussion. A recent survey of 100 NFL players revealed 56% would not report a concussion, due to not wanting to be withheld from competition [55]. Some high-profile athletes have even stated they intentionally under-perform on baseline neurocognitive assessments (also referred to as 'sandbagging') so that post-concussion test data would compare more favourably to their lower baseline. However, researchers have shown that only a small percentage of students who deliberately under-performed on ImPACT® were able to bypass detection [56, 57].

The extent to which embedded validity indicators on baseline testing identify deliberately poor performance, confusion or misunderstanding regarding how to take some aspect of the test, situational distractions in a group testing environment or some combination of factors is unknown. There is a need for more analogue malingering studies to better understand how people under-perform on computerized testing. Realistic scenarios can be created in which athletes are given instructions to deliberately under-perform on the test in a manner that will not be obvious. In addition, more research is needed on how confusion or misunderstanding on the part of the subject, regarding the test instructions or procedures, influences the probability of being flagged by a validity indicator. The above issues can be studied systematically in groups of high school and university students. The goal of this research is not only to examine rates of 'invalid' scores as determined by the test publisher's criteria, but to document the magnitude of the effects on the cognitive composite scores and the percentages of athletes' scores that are artificially lowered by different situational factors.

When to test following injury

One management approach is to use neuropsychological assessment only after an athlete's symptoms have resolved, as part of a stepwise process for clearing an athlete to return to sports. This is certainly a practical suggestion, given the time, cost and expense of a brief neuropsychological assessment. Scheduling repeated evaluations can be difficult and practice effects associated with repeated testing can be challenging to interpret. There is some evidence, however, that early testing might have value for predicting recovery time. Iverson [7] reported that high school football players who took longer than 10 days to recover were much more likely to show evidence of cognitive impairment within the first 72 hours post-injury than athletes who recovered more swiftly. It might also be possible to use early symptom reporting, in the absence of neuropsychological testing, to predict recovery time. In one study [58], it was possible to identify most high school football players who recovered swiftly (i.e. in 10 days or less) based on how they reported their post-concussion symptoms in the first 24 hours (i.e. they had much higher total scores and they were far less likely to report specific combinations of symptoms such as headaches, dizziness, noise sensitivity and memory problems). Using methodologies similar to the two studies above, more research is needed to determine if early testing following injury can reasonably accurately predict recovery time—at least in a sub-group of athletes. If so, early testing could be of benefit as a component of an overall injury management strategy.

From a clinical perspective, testing within a short time period following concussion might be useful to assist with early management recommendations. A minority of athletes might have cognitive deficits, a few days post-injury, that would render them unsafe to drive. These athletes would also benefit from a greater duration of rest and activity limitations. In contrast, an athlete who appeared cognitively normal and only had very mild symptoms might be encouraged to engage in more activities in school and daily life, as tolerated without exacerbation of symptoms. In general, brief evaluations, while symptomatic, can be used to monitor recovery and to make recommendations regarding activity restrictions and academic accommodations.

Reliability of neuropsychological tests

Concerns have been expressed about relatively low test–re-test reliability of computerized neuropsychological tests used in concussion management programmes [36, 39, 40, 59]. Test–re-test reliability relates to the stability of test scores. According to classical test theory, it has been viewed in terms of the relation between 'true' scores and obtained scores.

Obtained scores are believed to contain an error component and test–re-test reliability is influenced by error resulting from time and situational variables. Thus, high test–re-test reliability may be viewed as the ability of a test to reflect an individual score that is minimally influenced by error. Reliability should not be considered a dichotomous concept; rather it falls on a continuum. One cannot say a test is reliable or unreliable, but more accurately should say it possesses a high or low degree of reliability for a specific purpose, with a specific population [60, 61].

Table I. Test–re-test reliability of ImPACT® across different assessment intervals.

	Interval between assessments				
Test score	7 days[a]	30 days[b]	45 days[c]	1 year[d]	2 years[e]
Verbal Memory (ICC)	—	0.79	0.23	0.62	0.46
r	0.70	0.66	—	0.45	0.30
Visual Memory (ICC)	—	0.60	0.32	0.70	0.65
r	0.67	0.43	—	0.55	0.49
Visual Motor Speed (ICC)	—	0.88	0.38	0.82	0.74
r	0.86	0.78	—	0.74	0.60
Reaction Time (ICC)	—	0.77	0.39	0.71	0.68
r	0.79	0.63	—	0.62	0.52

ICC, Intra-class correlation coefficient; *r*, Pearson or Spearman correlation coefficient.
[a]Iverson et al. [10], $n = 56$.
[b]Schatz and Ferris [94], $n = 25$.
[c]Broglio et al. [40], $n = 73$.
[d]Elbin et al. [95], $n = 369$.
[e]Schatz [96], $n = 95$.

The alternate forms test–re-test reliability of ImPACT® across studies using different re-test intervals is presented in Table I. For comparison, the test–re-test reliabilities of other neuropsychological tests are presented in Table II. As seen in these tables, there is considerable variability in the size of the correlation coefficients across tests, studies and retest intervals. Some tests of intellectual ability, such as the Vocabulary sub-test of the Wechsler Intellectual Scale for Children–Fourth Edition (WISC-IV), have very high test–re-test reliabilities (0.88–0.91), compared to lower reliabilities on tests of working memory (WISC-IV Letter-Number Sequencing, $r = 0.64$–0.72), processing speed (WISC-IV Symbol Search, $r = 0.57$–0.68) and verbal learning (California Verbal Learning Test – Children's Edition Trial 1–5 total score, $r = 0.61$–0.73). Some tests of fluid executive functioning have very low test–re-test reliability (e.g. Delis Kaplan Executive Function System (D-KEFS) Trail Making Test Number-Letter Sequencing, $r = 0.20$; D-KEFS Design

Table II. Test–re-test reliabilities from the test manuals of intellectual and neuropsychological tests.

	r_{12}		r_{12}		r_{12}
WISC-IV Age: 12–13		*CMS Age: 9–12*		*D-KEFS Ages 8–19*	
Block Design	0.84	Visual Immediate Index	0.65	TMT 1: Visual Scanning	0.50
Similarities	0.86	Visual Delayed Index	0.61	TMT 2: Number Sequencing	0.77
Digit Span	0.85	Verbal Immediate Index	0.82	TMT 3: Letter Sequencing	0.57
Picture Concepts	0.66	Verbal Delayed Index	0.79	TMT 4: Number-Letter Sequencing	0.20
Coding	0.79	General Memory Index	0.86	TMT 5: Motor Speed	0.82
Vocabulary	0.88	Attention/Concentration Index	0.88	Letter Fluency	0.67
Letter-Number Sequencing	0.72	Learning Index	0.67	Category Fluency	0.70
Matrix Reasoning	0.71	Delayed Recognition Index	0.57	Category Switching-Total Correct	0.65
Comprehension	0.80	*CMS Age: 13–16*		Category Switching-Accuracy	0.53
Symbol Search	0.57	Visual Immediate Index	0.26	Design Fluency 1: Filled Dots	0.66
Picture Completion	0.83	Visual Delayed Index	0.40	Design Fluency 2: Empty Dots	0.43
Cancellation	0.78	Verbal Immediate Index	0.85	Design Fluency 3: Switching	0.13
Information	0.80	Verbal Delayed Index	0.87	Color-Word: Color Naming	0.79
Arithmetic	0.79	General Memory Index	0.86	Color-Word: Word Reading	0.77
Word Reasoning	0.72	Attention/Concentration Index	0.86	Color-Word: Inhibition	0.90
WISC-IV Age: 14–16		Learning Index	0.78	Color-Word: Inhibition/Switching	0.80
Block Design	0.88	Delayed Recognition Index	0.56	Sorting: Confirmed Correct Sorts	0.49
Similarities	0.82	*CVLT-C Age 8 (n = 35)*		Sorting: Free Sorting Description	0.67
Digit Span	0.81	List A Trials 1–5	0.73	Sorting: Sort Recognition Description	0.56
Picture Concepts	0.62	Short Delay Free Recall	0.40	20 Quest.: Initial Abstraction	0.62
Coding	0.86	Long Delay Free Recall	0.59	20 Quest: Weighted Achievement	0.06
Vocabulary	0.91	*CVLT-C Age 12 (n = 40)*		Word Context Consecutively Correct	0.58
Letter-Number Sequencing	0.64	List A Trials 1–5	0.73	Tower Test Achievement	0.51
Matrix Reasoning	0.77	Short Delay Free Recall	0.77	Proverbs: Free Inquiry	0.90
Comprehension	0.82	Long Delay Free Recall	0.62		
Symbol Search	0.68	*CVLT-C Age 16 (n = 31)*			
Picture Completion	0.85	List A Trials 1–5	0.61		
Cancellation	0.76	Short Delay Free Recall	0.48		
Information	0.88	Long Delay Free Recall	0.60		
Arithmetic	0.84				
Word Reasoning	0.83				

WISC-IV, The test–re-test coefficients and SDs were adapted from tables on page 42 of Wechsler [97]. The test–re-test interval was 32 days (range = 13–63) and the total sample size was 243. Children's Memory Scale® (CMS), The test–re-test interval was 59.6 days (median = 65.3; SD = 29.2) and the total sample size was 125. CVLT-C, List A Trials 1–5 change estimates are for *T*-scores. The delayed recall change estimates are for raw scores. Median Re-test Interval = 28 days, range = 10–42. Data adapted from pages 88–90 of the test manual [98]. D-KEFS, These data were adapted from Tables 2.1–2.26 in the D-KEFS Technical Manual [99]. The ranges for internal consistency represent values across age groups from 8–19 years. The test–re-test coefficients are based on 28 individuals between the ages of 8–19 who were tested twice separated by ~25 days (SD = 12.8 days).

Fluency Switching, $r = 0.13$). The authors, for many years, have heard concerns expressed by some neuropsychologists and other professionals that computerized tests are 'unreliable'. Reliability, however, is a fundamental challenge and problem for both computerized and traditional neuropsychological tests. Those cognitive abilities that are more 'fluid' (e.g. processing speed, memory and executive functioning) and likely to be affected by a concussion are also much more difficult to assess in a reliable manner.

Reliable change methodology

The reliable change methodology (RCI), originally proposed by Jacobson and colleagues [62–64], can be helpful for interpreting serial assessment data in concussed athletes. It can be used to monitor change from baseline and to monitor recovery from injury. This methodology is commonly used in clinical neuropsychology [65–69] and sports neuropsychology [5, 70, 71] and allows the clinician to estimate measurement error surrounding test–re-test difference scores by creating confidence intervals for these difference scores. Specifically, the standard error of the difference (SE_{diff}) is used to create a confidence interval for the test–re-test difference score. Different formulae have been used across studies in clinical psychology and neuropsychology over the past 15 years, such as using two times the SEM for time one only in the formula for the SE_{diff}. The steps recommended for calculating the SE_{diff} are listed below.

(1) $SEM_1 = SD\sqrt{1 - r_{12}}$; Standard deviation from time 1 testing multiplied by the square root of 1 minus the test–re-test coefficient.
(2) $SEM_2 = SD\sqrt{1 - r_{12}}$; Standard deviation from time 2 testing multiplied by the square root of 1 minus the test–re-test coefficient.
(3) $SE_{diff} = \sqrt{SEM_1^2 + SEM_2^2}$; Square root of the sum of the squared SEMs for each testing occasion.
(4) Reliable Change Confidence Intervals = The SE_{diff} is multiplied by the following z-scores: $\pm$1.04 (70% CI), $\pm$1.28 (80% CI), $\pm$1.64 (90% CI) and $\pm$1.96 (95% CI).

Reliable change is used to determine if there has been improvement or deterioration in functioning that exceeds the probable range of measurement error. It is a practical psychometric strategy that supplements clinical judgement. Such a reliable change methodology, for example, is built-in to the ImPACT® scoring programme, whereby an asterisk is placed next to scores that are reliably different from baseline test data. However, there are several clinical and methodological challenges related to the best practice application of the reliable change methodology to cognitive test scores following sport-related concussion. First, it is possible to experience a 'real' improvement or decline in functioning that is not considered statistically reliable. Neuropsychological tests have considerable measurement error and there is a constant need to balance sensitivity and specificity in the context of detection of cognitive impairment and in the interpretation of serial testing. If measurement error is not accounted for, one is very likely to over-interpret change scores. It is likely that the fact that 'real' change has occurred is assumed, when it has not. However, if 90% or 95% confidence interval for change is rigidly applied to every test, one will invariably fail to identify change in some people. Thus, reliable change estimates are meant to supplement, not replace, clinical judgement.

Second, the reliable change methodology does not account for regression to the mean. In general, athletes who score very low or very high on a test at baseline are statistically more likely to score somewhat better or worse, respectively, at re-test due to regression to the mean. This is one of the reasons why the reliable change methodology can be less accurate when applied to athletes who score unusually low at baseline. Third, the reliable change methodology can be adjusted to account for 'average' practice effects in a group [65, 72], but in some contexts, such as testing with ImPACT®, practice effects are not well defined or routinely accounted for. Fourth, reliable change on ImPACT® is based on the assumption of construct and metric equivalence of the alternate forms. A recent study, however, illustrated that the different forms of ImPACT® do not appear to be equivalent [73]. This creates additional measurement error within the reliable change methodology because there is an assumption that the confidence interval for measurement error surrounding re-testing with Form 2 following Form 1 applies comparably across the other alternate forms (in comparison to Form 1).

Finally, the reliable change methodology rests on the assumption that a single pair of scores is being compared. This is rarely the case in neuropsychology. In contrast, multiple test–re-test comparisons are considered simultaneously. The estimates of reliable change apply to a single distribution of change scores. It might be uncommon, when considering one test in isolation, for there to be a large change in test scores. However, when considering a large battery of tests administered twice, it is statistically likely that healthy people will show a small number of large changes in test scores. This is an area of research that is conspicuously absent in clinical neuropsychology.

Data from Iverson et al. [5] can be used to illustrate the complexity of interpreting multiple test–re-test scores on ImPACT® simultaneously. Participants were 56 adolescents and young adults who completed the battery twice, with an average interval of 5.8 days (median = 7, SD = 3.0, range = 1–13). Assuming a symmetric 80% confidence interval for change (unadjusted for practice or regression), it would be predicted from the standard normal distribution of change scores that ~10% of healthy athletes would show a reliable decline and 10% would show a reliable improvement on re-testing. Athletes showing practice effects, however, would be statistically more likely to show reliable improvement (than predicted by the probable range of measurement alone). Considering the four cognitive composite scores and the Post-Concussion Scale score individually, 7.1–12.5% of healthy athletes obtained a reliably lower score (or greater symptom reporting) when each score was considered individually (and 7.1–16.1% showed improved scores). However, when all five scores were considered simultaneously, 39.3% had one decline, 1.8% had two declines, 0% had three declines and 1.8% had four declines. The important point is that having a statistically reliable improvement or deterioration on one test score, when only a single test is given, is statistically uncommon. However, having an improvement or decline on one test score, when multiple tests are given and interpreted simultaneously, is common.

The interpretation of post-injury neuropsychological test results, in individual athletes, is complex. In fact, it can be more complex, in some ways, than the interpretation of neuropsychological test results with other patient populations because athletes are sometimes tested several times over brief re-test intervals, so the reliability of the test results, situational factors affecting test performance, the prevalence of low scores and the reliable change methodology all need to be considered simultaneously. This is illustrated using three cases presented in Table. The ImPACT® test results for each athlete at baseline and three time periods following injury are presented. The arrows represent reliable differences (i.e. declined or improved) as compared to baseline testing. Reliable change difference scores were based on the 80% confidence interval from Table III in Iverson et al. [10]. The number of low scores was based on having a score below the 10th percentile, derived from the normative data for desktop version 2.0 of ImPACT®. Subject #1 appeared to recover cognitively and symptomatically at approximately the same time. At 7 days following injury, he had slower reaction time and ongoing post-concussion symptoms, but at 11 days he was asymptomatic, he had no low test scores and all scores were similar to baseline. Subject #2 appeared to recover cognitively before he recovered symptomatically. At 3 days following injury, he had no low test scores and all scores were similar or better than his baseline results. He was symptomatic, however, at that point (Post-Concussion Scale total score = 16). By 6 days following injury, he continued to have no low scores, his scores were comparable or better than baseline and he was mostly asymptomatic. Subject #3 appeared to recover symptomatically before he recovered cognitively. At 13 days following injury, he was asymptomatic but his two memory scores were reliably lower than baseline. He was tested for a fourth time 20 days post-injury (not shown in Table III) and his total symptom score was zero, he had no low cognitive composite scores and none of his cognitive scores were reliably lower than baseline. Notice that, for subjects #2 and #3, both showed post-injury test scores that were statistically reliably better post-injury compared to baseline (using the 80% confidence interval). This could reflect practice effects, situational factors artificially lowering baseline performance, non-specific measurement error or a combination of factors.

ImPACT® reliable change, 95% confidence interval, adjusted for practice effects

The confidence interval selected for reliable change influences the sensitivity and specificity for detecting change on cognitive test scores. Assuming a symmetric distribution of change scores, the 70% confidence interval will have 15% in each tail, the 80% confidence interval will have 10% in each tail and the 95% confidence interval will have 2.5% of healthy, uninjured subjects in each tail. Stated differently, somewhere between 2.5–15% of healthy subjects (based on the confidence interval selected), who have no clinical reason for obtaining a reliably higher or lower score, will show a statistically reliable change; the probable range of measurement error encompasses the change scores from 70–95% of healthy subjects. In the context of identifying a decline in performance, applying the 70% confidence interval will result in 15% false positives and applying the 95% confidence interval will result in 2.5% false positives. Of course, the lower false positive rate comes at the sacrifice of sensitivity: greater change scores are required for the 95% confidence interval, which means that fewer concussed athletes will be accurately classified as having reliably lower scores if this confidence interval is selected.

Reliable change, using the 95% confidence interval and adjusted for practice effects [65], for different time intervals, is illustrated in Table IV. Data from 25 healthy control athletes who were tested at an ∼1-month re-test interval and 369 athletes who were tested at an ∼1-year re-test interval were used. The percentages of athletes who improved or declined are presented. If the distributions of change scores were precisely normally distributed and practice effects could be precisely controlled, then there should be 2.5% who improved and 2.5% who worsened. The values in Table IV are fairly close to those theoretical values, although a greater percentage had improved scores over a 1-month re-test

Table III. Baseline and post-injury test results for three high school athletes.

	Subject 1				Subject 2				Subject 3			
	B	1	2	3	B	1	2	3	B	1	2	3
Verbal Memory	86	80	93	84	91	78↓	88	94	80	45↓	78	71↓
Visual Memory	92	75↓	86	94	78	74	79	81	71	51↓	42↓	56↓
Processing Speed	30.18	29.20	29.83	35.93	34.55	35.48	43.30↑	35.45	28.50	24.33↓	37.33↑	38.28↑
Reaction Time	0.55	0.58	0.73↓	0.57	0.54	0.56	0.57	0.48↑	0.63	1.10↓	0.56↑	0.52↑
Post-Concussion Scale	1	10	13↓	0	0	7	16↓	2	0	58↓	3	0
Days Following Injury	—	3	7	11	—	1	3	6	—	4	8	13
Number of Low Scores	0	1	1	0	0	0	0	0	1	4	1	2
Reliable Difference Scores	—	1	2	0	—	1	1	0	—	5	1	2
Cognitively Recovered	—	No	No	Yes	—	No	Yes	Yes	—	No	No	No
Asymptomatic	—	No	No	Yes	—	No	No	Probably	—	No	Maybe	Yes

B, Baseline testing; 1, Time 1 post-injury; 2, Time 2 post-injury; 3, Time 3 post-injury. The arrows represent reliable differences as compared to baseline testing. The following reliable change difference scores were applied, based on the 80% confidence interval from Table III in Iverson et al. [10]: Verbal Memory = ±9 points, Visual Memory = ±14 points, Reaction Time = ±0.06 seconds, Processing Speed = −3 points (declined) and +7 points (improved) and Post-Concussion Scale total scores = ±10 points.

Table IV. Reliable change on ImPACT® in healthy, uninjured athletes tested at different re-test intervals.

	30 day re-test interval ($n = 25$)		1 year re-test interval ($n = 369$)	
	% Improved	% Worsened	% Improved	% Worsened
Verbal Memory	8.0%	0%	2.7%	2.1%
Visual Memory	4.0%	0%	4.1%	1.4%
Reaction Time	4.0%	0%	5.7%	2.4%
Processing Speed	12.0%	4.0%	7.1%	1.4%
One score	32.0%	8.0%	17.9%	6.2%
Two scores	0%	0%	1.4%	0.3%
Three scores	0%	0%	0%	0%
Four scores	0%	0%	0%	0%

The 95% confidence interval adjusted for practice effects [65], for different time intervals, was used.

interval. When considering all composite scores simultaneously, a substantial minority will show at least one statistically reliable improvement (32% at 1 month and 18% at 1 year), but only a small percentage show a worsening in performance (6–8%). Having two or more scores that reliably improve or decline is rare at both time intervals.

New methods for identifying cognitive impairment

Following a sport-related concussion, an athlete could have severe cognitive impairment in the first 2 hours, moderate cognitive impairment in the first 24 hours, followed by mild cognitive impairment during days 2–5 post-injury (with full apparent recovery in cognitive functioning within 2 weeks). In contrast, another athlete might have mild cognitive impairment for the first day, followed by mild cognitive diminishment and a return to normal cognitive functioning within 72 hours of injury. Cognitive impairment is quantified based on neuropsychological test performance and interview results (i.e. how is the athlete functioning in daily life). It is normal and customary to apply the standard normal distribution (i.e. the 'bell curve') to the interpretation of individual cognitive test scores. For example, if a cut-off score of below one standard deviation from the mean is used (16th percentile), then only 15% of healthy people will obtain a score below the cut-off. Some clinicians will use 1 SD, the 10th percentile, 1.5 SDs or 2 SDs below the mean to define a low or 'impaired' score. It is important to note, however, that numerous studies have been published showing that a substantial minority of healthy children, adolescents and adults will obtain one or more unusually low scores when multiple tests are given [74–84]. Several review papers and book chapters illustrate the principles of *multivariate base rate analyses* (i.e. interpreting multiple test scores simultaneously vs. individually) as applied to batteries of cognitive test scores [85–88]. These articles and chapters illustrate some important points when trying to identify cognitive impairment in an individual patient: (i) the more tests that are given, the more likely it is for a healthy uninjured person to obtain one or more low scores; (ii) higher levels of intelligence and education are associated with fewer low scores; and (iii) race/ethnicity (e.g. African American or Hispanic), English as a second language, less education and lower intelligence are associated with obtaining more low cognitive test scores.

Classifying cognitive test scores using multivariate base rates

Iverson and Brooks [89] and Iverson [90] developed and evaluated evidence-based, psychometric criteria for defining cognitive impairment on ImPACT® in high school boys with sport-related concussions. The test performances from an archival normative database of healthy boys and a sample of concussed high school football players were used to define clinically- and theoretically-derived criteria for cognitive impairment. The clinical algorithm, set out below, was developed to represent the following classification ranges: broadly normal, below average, well below average, unusually low and extremely low.

(1) Broadly Normal: 2 or fewer scores below the 25th percentile, AND 1 or fewer scores below the 16th percentile, AND no scores below the 10th percentile.
(2) Below Average: 3 or more scores below the 25th percentile, OR 2 scores below the 16th percentile, OR 1 score below the 10th percentile.
(3) Well Below Average: 3 or more scores below the 16th percentile, OR 2 scores below the 10th percentile, OR 1 score at or below the 5th percentile.
(4) Unusually Low: 3 or more scores below the 10th percentile, OR 2 scores at or below the 5th percentile, OR 1 score at or below the 2nd percentile.
(5) Extremely Low: 3 or more scores at or below the 5th percentile, OR 2 or more scores at or below the 2nd percentile.

These classification ranges were developed on a sample of 341 adolescent boys who underwent pre-season baseline testing and then they were applied to a sample of 125 high school football players who completed testing within 5 days of sustaining a concussion. The breakdown of healthy control subjects, by classification range, was as follows: broadly normal = 73.0%, below average = 9.1%, well below average = 9.4%, unusually low = 7.9% and extremely low = 0.59%. The breakdown of concussed high school football players, by classification range, was as follows: broadly normal = 20.8%, below average = 10.4%, well below average = 12.8%, unusually low = 35.2% and extremely low = 20.8%. The majority of healthy subjects (73%) and a minority of concussed athletes (21%) were classified as broadly normal [$\chi^2(1, 466) = 103.1$, $p < 0.0001$]. In contrast, 56% of concussed athletes and only 8.5% of healthy subjects fell in the unusually low or extremely low classification ranges [$\chi^2(1, 466) = 123.3$, $p < 0.0001$].

The use of these new classification ranges is illustrated in three concussed high school football players presented in Table V. For subject #4, notice that at baseline he had three scores below the 25th percentile, placing his overall performance in the Below Average classification range. Nine days following his injury, he had three scores at or below the 5th percentile, placing his overall performance in the Extremely Low classification range. At 26 days following injury, he had only one score below the 25th percentile, placing his overall performance in the Broadly Normal range. His performance at that time was better than his baseline performance, suggesting that his baseline performance might have been an underestimate of his cognitive abilities, his post-injury performance

Table V. Baseline and post-injury test results for four concussed high school boys.

	Subject 4			Subject 5			Subject 6		
	B	1	2	B	1	2	B	1	2
Verbal Memory	67	45	78	91	54	74	94	45	97
Visual Memory	67	52	84	69	59	68	69	52	56
Processing Speed	34.58	28.95	35.58	27.80	24.85	32.63	38.00	28.95	49.50
Reaction Time	0.66	0.70	0.52	0.53	0.52	0.49	0.49	0.70	0.59
Post-Concussion Scale	3	24	7	0	6	2	0	10	2
Days Following Injury	—	9	26	—	3	29	—	6	18
Number of Low Scores									
25th percentile	3	4	1	1	3	2	0	4	1
16th percentile	2	3	0	0	3	1	0	3	1
10th percentile	1	3	0	0	3	0	0	3	1
5th percentile	1	3	0	0	1	0	0	3	0
2nd percentile	0	1	0	0	1	0	0	2	0
Classification Range Iverson & Brooks [89]	Below Average	Extremely Low	Broadly Normal	Broadly Normal	Well Below average	Broadly Normal	Broadly Normal	Extremely Low	Below Average

B, Baseline testing; 1, Time 1 post-injury; 2, Time 2 post-injury. The Iverson and Brooks [89] classification ranges were developed and evaluated on high school boys only. The original version 2.0 normative data for the desktop version of ImPACT was used for this table.

might have been artificially elevated due to practice effects or both. For subject #5, his overall performance was Broadly Normal at baseline, Well Below Average at 3 days following injury and Broadly Normal at 29 days following injury. However, at 29 days, his Verbal Memory score remained lower than expected based on his baseline score. It is not known whether this reflects a lingering cognitive deficit or situational factors adversely influencing his test performance (test results after day 29 were not available for review). For subject #6, his baseline performance was Broadly Normal, his profile of scores at 6 days following injury was Extremely Low and at 18 days following injury his profile of scores was Below Average because his Visual Memory score was low (i.e. below the 10th percentile). These cases illustrate that these classification ranges can be used as part of the overall clinical interpretation of ImPACT®, as a supplement to reliable change analyses and individual test score interpretations.

Multivariate base rates for online ImPACT®

For this article, multivariate base rate analyses were conducted on thousands of subjects who had taken online ImPACT®. These analyses are presented in Table VI. Participants were 46 679 adolescents and young adults who completed baseline computerized neurocognitive evaluations with ImPACT®. They had valid baseline scores and subjects with self-reported ADHD or academic problems were not included. The breakdown of the sample by gender-stratified age cohorts was as follows: females aged 14–18 = 14 426 and 19–22 = 7640 and males aged 14–18 = 14 487 and 19–22 = 10 126. The four primary composite scores used for the base rate analyses were Verbal Memory, Visual Memory, Processing Speed and Reaction Time.

The base rates of low scores in adolescents and young adults on the four composite scores vary by level of cut-off and do not conform to a standard normal distribution. Having one or more scores at or below the 2nd percentile occurs in 7–8% of the subjects in the four age cohorts and having one or more scores at or below the 10th percentile occurs in 26–28% of the subjects in the four age cohorts. Obtaining two or more scores below the 16th percentile occurs in ~15% of adolescents and young adults. Age and gender stratified base rate results for the following cut-off scores are provided in Table VI: 25th, 16th, 10th, 5th and 2nd percentiles.

As seen in Table VI, when multiple neurocognitive test scores are interpreted simultaneously, the metrics of the standard normal distribution cannot be relied upon. That is, the bell curve predicts that ~15% of healthy subjects will obtain a score that is greater than 1 SD below the mean, when a single score is considering in isolation, but, as seen in Table VI, 40% will obtain one score in this range if the four scores are considered simultaneously. This is very helpful to know in day-to-day clinical practice.

The multivariate base rates presented in Table VI are ready for clinical use. To illustrate their use, hypothetical data from four concussed athletes are presented in Table VII. The 18-year-old female had no low scores at baseline and three scores below the 25th percentile following injury. Only 9.9% of uninjured adolescent females, between 14–18, obtain three or more scores below the 25th percentile. The 20-year-old female had one score below the 10th percentile at baseline. Approximately 28% of young women her age have one or more scores below the 10th percentile (base rate = 27.7%). Following injury, however, she had three scores below the 10th percentile; this occurs in fewer than 2% of women her age (base rate = 1.8%). In the first three cases in Table VII, cognitive impairment secondary to concussion could be inferred from the post-injury data only, using the base rate results from Table VI (i.e. these three cases had post-injury scores associated with low base rates; 1.8–9.9%). For the 21-year-old man, however, having two scores below the 25th percentile following his injury is fairly common compared to healthy young men his age (base rate = 27.1%). In his case, having baseline data is particularly important. At baseline, he had no low scores and some of his post-injury scores would be reliably different compared to his baseline testing (if the reliable change methodology is used).

Table VI. Prevalence of low composite scores on ImPACT® (online version) in healthy, uninjured subjects.

Number of scores below cut-off	Females, ages 14–18 (*n* = 14 426)		Males, ages 14–18 (*n* = 14 487)		Females, ages 19–22 (*n* = 7640)		Males, ages 19–22 (*n* = 10 126)	
	%	C%	%	C%	%	C%	%	C%
<25th %ile								
4	2.5	2.4	2.9	2.9	2.2	2.2	2.9	2.9
3	7.4	9.9	7.9	10.9	7.1	9.3	7.3	10.2
2	16.8	26.6	16.0	26.8	17.7	27.0	16.9	27.1
1	29.4	56.1	28.5	55.4	30.1	57.1	28.0	55.1
0	43.9	100	44.7	100	42.9	100	44.9	100
<16th %ile								
4	0.9	0.9	1.2	1.2	0.7	0.7	1.1	1.1
3	3.5	4.4	3.6	4.8	3.2	3.9	3.4	4.5
2	10.8	15.2	10.0	14.8	10.9	14.8	10.7	15.2
1	25.7	40.9	24.3	39.1	25.8	40.6	24.4	39.6
0	59.1	100	60.9	100	59.4	100	60.4	100
<10th %ile								
4	0.3	0.3	0.4	0.4	0.3	0.3	0.4	0.4
3	1.6	1.9	1.7	2.1	1.5	1.8	1.7	2.1
2	6.1	8.0	5.9	7.0	5.8	7.6	6.0	8.0
1	19.8	27.8	18.8	25.8	20.1	27.7	18.8	26.8
0	72.2	100	73.2	100	72.3	100	73.2	100
≤5th %ile								
4	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
3	0.6	0.7	0.7	0.8	0.5	0.6	0.6	0.7
2	2.8	3.5	3.0	3.8	2.7	3.3	2.7	3.4
1	13.0	16.5	12.8	16.6	13.0	16.3	12.9	16.3
0	83.5	100	83.4	100	83.7	100	83.7	100
≤2nd %ile								
4	—	—	—	—	—	—	—	—
3	0.2	0.2	0.1	0.1	0.1	0.1	0.1	0.1
2	0.9	1.1	0.9	1.0	0.7	0.8	0.7	0.8
1	6.7	7.8	6.3	7.3	6.0	6.8	6.1	6.9
0	92.4	100	92.7	100	93.2	100	93.1	100

The cumulative percentages reflect the percentages of subjects who have a certain number of low scores or more. For example, having one or more scores below the 16th percentile is common (i.e. 24.3–40.6% of healthy subjects), two or more is uncommon (10.0–15.2%) and three or more is rare (3.2–4.8%).

Table VII. Hypothetical baseline and post-injury test scores illustrating the use of the multivariate base rates presented in Table VI.

Cognitive domain	18-year-old Female		15-year-old Male		20-year-old Female		21-year-old Male	
	B	P	B	P	B	P	B	P
Verbal Memory	77	21	63	34	35	7	92	30
Visual Memory	55	28	28	9	9	< 1	60	12
Visual Motor Speed	99	4	36	37	30	7	67	26
Reaction Time	78	19	37	5	78	23	73	17
Number of Low Scores								
<25th percentile	*0*	*3*	*0*	2	1	4	*0*	*2*
<16th percentile	0	1	0	2	1	3	0	1
<10th percentile	0	1	0	*2*	*1*	*3*	0	0
≤5th percentile	0	1	0	1	0	1	0	0
≤2nd percentile	0	0	0	0	0	1	0	0
Base Rate	43.9	9.9	44.7	7.0	27.7	1.8	44.9	27.1

B, Baseline; P, Post-Injury percentile ranks for online ImPACT®. The base rates represent the cumulative percentile ranks from Table VI. The scores in italics in the table are discussed in the text.

Conclusions

Several position and consensus statements have recommended neuropsychological assessment as a component of concussion management programmes [44, 91–93]. This review focused on a series of topics relating to the neuropsychological assessment of athletes who have sustained a sport-related concussion. First, in clinical practice and research, some use traditional neuropsychological tests, others use computerized tests and some use both (termed a 'hybrid' approach). There is insufficient evidence to conclude that one approach is clearly superior to the others, although some data suggest that the sensitivity of computer-based measures is superior to pencil-and-paper measures within 1–3 days of concussion. Second, in regard to pre-season testing, it is concluded that having an accurate measure of baseline cognitive functioning is helpful for quantifying cognitive deficits following injury and for assessing recovery. This is especially true for athletes who have above average or below average cognitive functioning at baseline or a developmental condition such as ADHD or a learning disability. At present, however, there is insufficient evidence to conclude that having baseline test results is time- and cost-effective or clearly superior to not having baseline test results. Third, some athletes perform poorly on baseline testing and are identified by embedded validity indicators as having invalid test results. The extent to which these embedded validity indicators on baseline testing identify deliberately poor performance, confusion or misunderstanding regarding how to take some aspect of the test, situational distractions in a group testing environment or some combination of factors is unknown. Fourth, concerns have been expressed about the reliability and validity of computerized cognitive screening batteries used in sport concussion management programmes. These concerns are not unique to computerized testing; traditional paper–pencil neuropsychological tests also have varying degrees of reliability and validity. Finally (and fortunately), sophisticated psychometric methods are available to assist clinicians and researchers with the identification of cognitive impairment and the serial monitoring of recovery, such as reliable change and multivariate base rates. These methods can improve accuracy, reduce false positive diagnoses and clinical inferences and strengthen the scientific underpinnings of clinical judgement.

Some directions for future research are listed below. These suggestions focus on advancing knowledge, with the goal of improving clinical practice.

(1) Conduct programmatic research relating to the strengths and limitations of baseline testing for improving the accuracy of neuropsychological assessment and determining whether improved accuracy contributes to improved management of this injury in athletes.
(2) Pursue analogue malingering studies to better understand how people under-perform on computerized testing. In addition, examine how confusion or misunderstanding on the part of the subject, regarding the test instructions or procedures, influences the probability of being flagged by a validity indicator.
(3) Evaluate and improve, if possible, the test–re-test reliability of traditional and computerized cognitive tests. Determine if there are differences in the magnitude of practice effects based on the domain of cognitive functioning assessed and whether the test is paper–pencil or computerized.
(4) Develop and evaluate more sophisticated methods for interpreting change on cognitive testing (e.g. refinements of the reliable change methodology, such as correction for practice and stratification of confidence intervals for change based on level of baseline performance; and use of standardized regression models). Apply multivariate base rate analyses to the reliable change methodology to quantify the likelihood of showing one or more reliable change scores when multiple change scores are considered simultaneously.
(5) Determine if cognitive functioning, assessed in the first 72 hours post injury, has prognostic value for predicting typical vs. slow recovery.
(6) Develop and evaluate clinical algorithms, with known false positive rates, for identifying and quantifying cognitive impairment following concussion.

Declaration of interest

Grant Iverson, PhD, has been reimbursed by the government, professional scientific bodies and commercial organizations for discussing or presenting research relating to mild TBI and sport-related concussion at meetings, scientific conferences and symposiums. He has a clinical and consulting practice in forensic neuropsychology involving individuals who have sustained mild TBIs (including professional athletes). He has received research funding from several test publishing companies, including ImPACT Applications, Inc., CNS Vital Signs and Psychological Assessment Resources (PAR, Inc.). He has not received research support from ImPACT Applications, Inc. in the past 3 years. He receives royalties from two books relating to neuropsychology and one test (WCST-64). Philip Schatz, PhD, has served as a consultant to the International Brain Research Foundation, the Sports Concussion Center of New Jersey and ImPACT Applications, Inc., to study the effects of concussion in high school and collegiate athletes. However, these entities had no role in the conceptualization or content of the current manuscript or the decision to submit for publication.

References

1. Barth JT, Alves W, Ryan T, Macciocchi S, Rimel RW, Jane JJ, Nelson W. Mild head injury in sports: Neuropsychological sequelae and recovery of function. In: Levin H, Eisenberg J, Benton A, editors. Mild head injury. New York: Oxford University Press; 1989. pp 257–275.
2. Collins MW, Iverson GL, Lovell MR, McKeag DB, Norwig J, Maroon J. On-field predictors of neuropsychological and symptom deficit following sports-related concussion. Clinical Journal of Sport Medicine 2003;13:222–229.
3. Lovell MR, Collins MW, Iverson GL, Johnston KM, Bradley JP. Grade 1 or ''ding'' concussions in high school athletes. American Journal of Sports Medicine 2004;32:47–54.
4. Lovell MR, Collins MW, Iverson GL, Field M, Maroon JC, Cantu R, Podell K, Powell JW, Belza M, Fu FH. Recovery from mild concussion in high school athletes. Journal of Neurosurgery 2003; 98:296–301.
5. Iverson GL, Lovell MR, Collins MW. Interpreting change on ImPACT following sport concussion. Clinical Neuropsychology 2003;17:460–467.
6. Fazio VC, Lovell MR, Pardini JE, Collins MW. The relation between post concussion symptoms and neurocognitive performance in concussed athletes. NeuroRehabilitation 2007;22:207–216.
7. Iverson GL. Predicting slow recovery from sport-related concussion: The new simple-complex distinction. Clinical Journal of Sport Medicine 2007;17:31–37.
8. Covassin T, Schatz P, Swanik CB. Sex differences in neuropsychological function and post-concussion symptoms of concussed collegiate athletes. Neurosurgery 2007;61:345–350; discussion 350–341.
9. McClincy MP, Lovell MR, Pardini J, Collins MW, Spore MK. Recovery from sports concussion in high school and collegiate athletes. Brain Injury 2006;20:33–39.
10. Iverson GL, Brooks BL, Collins MW, Lovell MR. Tracking neuropsychological recovery following concussion in sport. Brain Injury 2006;20:245–252.

11. Van Kampen DA, Lovell MR, Pardini JE, Collins MW, Fu FH. The ''value added'' of neurocognitive testing after sports-related concussion. American Journal of Sports Medicine 2006;34: 1630–1635.
12. Broglio SP, Macciocchi SN, Ferrara MS. Sensitivity of the concussion assessment battery. Neurosurgery 2007;60:1050–1057; discussion 1057–1058.
13. Broshek DK, Kaushik T, Freeman JR, Erlanger D, Webbe F, Barth JT. Sex differences in outcome following sports-related concussion. Journal of Neurosurgery 2005;102:856–863.
14. Erlanger D, Kaushik T, Cantu R, Barth JT, Broshek DK, Freeman JR, Webbe FM. Symptom-based assessment of the severity of a concussion. Journal of Neurosurgery 2003;98: 477–484.
15. Collie A, Makdissi M, Maruff P, Bennell K, McCrory P. Cognition in the days following concussion: Comparison of symptomatic vs asymptomatic athletes. Journal of Neurology, Neurosurgery & Psychiatry 2006;77:241–245.
16. Makdissi M, Collie A, Maruff P, Darby DG, Bush A, McCrory P, Bennell K. Computerised cognitive assessment of concussed Australian rules footballers. British Journal of Sports Medicine 2001;35:354–360.
17. Matser JT, Kessels AG, Lezak MD, Troost J. A dose-response relation of headers and concussions with cognitive impairment in professional soccer players. Journal of Clinical & Experimental Neuropsychology 2001;23:770–774.
18. Macciocchi SN, Barth JT, Alves W, Rimel RW, Jane JA. Neuropsychological functioning and recovery after mild head injury in collegiate athletes. Neurosurgery 1996;39:510–514.
19. Guskiewicz KM, Ross SE, Marshall SW. Postural stability and neuropsychological deficits after concussion in collegiate athletes. Journal of Athletic Training 2001;36:263–273.
20. Collins MW, Grindel SH, Lovell MR, Dede DE, Moser DJ, Phalin BR, Nogle S, Wasik M, Cordry D, Daugherty KM, et al. Relationship between concussion and neuropsychological performance in college football players. Journal of the American Medical Association 1999;282:964–970.
21. McCrea M, Guskiewicz KM, Marshall SW, Barr W, Randolph C, Cantu RC, Onate JA, Yang J, Kelly JP. Acute effects and recovery time following concussion in collegiate football players: The NCAA Concussion Study. Journal Of The American Medical Association 2003;290:2556–2563.
22. Broglio SP, Puetz TW. The effect of sport concussion on neurocognitive function, self-report symptoms and postural control: A meta-analysis. Sports Med 2008;38:53–67.
23. Belanger HG, Vanderploeg RD. The neuropsychological impact of sports-related concussion: A meta-analysis. Journal of the International Neuropsychological Society 2005;11:345–357.
24. Iverson GL. Mild traumatic brain injury meta-analyses can obscure individual differences. Brain Injury 2010;24:1246–1255.
25. Collins MW, Lovell MR, Iverson GL, Ide T, Maroon J. Examining concussion rates and return to play in high school football players wearing newer helmet technology: A three year prospective cohort study. Neurosurgery 2006;58:275–286.
26. Collins MW, Iverson GL, Gaetz M, Meehan WP, Lovell MR. Sport-related concussion. In: Zasler ND, Katz DI, Zafonte RD, editors. Brain injury medicine: Principles and practice. 2nd ed. New York: Demos Medical Publishing; 2012. pp 498–516.
27. Pellman EJ, Lovell MR, Viano DC, Casson IR, Tucker AM. Concussion in professional football: Neuropsychological testing-part 6. Neurosurgery 2004;55:1290–1305.
28. Pellman EJ, Lovell MR, Viano DC, Casson IR. Concussion in professional football: Recovery of NFL and high school athletes assessed by computerized neuropsychological testing-part 12. Neurosurgery 2006;58:263–274; discussion 263–274.
29. Pellman EJ, Viano DC, Casson IR, Arfken C, Powell J. Concussion in professional football: Injuries involving 7 or more days out-Part 5. Neurosurgery 2004;55:1100–1119.
30. Comper P, Hutchison M, Magrys S, Mainwaring L, Richards D. Evaluating the methodological quality of sports neuropsychology concussion research: A systematic review. Brain Inury 2010;24: 1257–1271.
31. Schatz P, Pardini JE, Lovell MR, Collins MW, Podell K. Sensitivity and specificity of the ImPACT test battery for concussion in athletes. Archives of Clinical Neuropsychology 2006;21:91–99.
32. Schatz P, Sandel N. Sensitivity and specificity of the online version of ImPACT in high school and collegiate athletes. American Journal of Sports Medicine 2013;41:321–326.
33. Register-Mihalik JK, Guskiewicz KM, Mihalik JP, Schmidt JD, Kerr ZY, McCrea MA. Reliable change, sensitivity, and specificity of a multidimensional concussion assessment battery: Implications for caution in clinical practice. Journal of Head Trauma Rehabilitation 2013;28:274–283.
34. McCrea M, Barr WB, Guskiewicz K, Randolph C, Marshall SW, Cantu R, Onate JA, Kelly JP. Standard regression-based methods for measuring recovery after sport-related concussion. Journal of the International Neuropsychological Society 2005;11:58–69.
35. Randolph C, Kirkwood MW. What are the real risks of sport-related concussion, and are they modifiable? Journal of the International Neuropsychological Society 2009;15:512–520.
36. Randolph C, McCrea M, Barr WB. Is neuropsychological testing useful in the management of sport-related concussion? Journal of Athletic Training 2005;40:139–152.
37. Moser RS, Schatz P, Neidzwski K, Ott SD. Group vs individual administration affects baseline neurocognitive test performance. American Journal of Sports Medicine 2011;39:2325–2330.
38. Echemendia RJ, Iverson GL, McCrea M, Macciocchi SN, Gioia GA, Putukian M, Comper P. Advances in neuropsychological assessment of sport-related concussion. British Journal of Sports Medicine 2013;47:294–298.
39. Mayers LB, Redick TS. Clinical utility of ImPACT assessment for postconcussion return-to-play counseling: Psychometric issues. Journal of Clinical & Experimental Neuropsychology 2012;34: 235–242.
40. Broglio SP, Ferrara MS, Macciocchi SN, Baumgartner TA, Elliott R. Test-retest reliability of computerized concussion assessment programs. Journal of Athletic Training 2007;42:509–514.
41. Resch J, Driscoll A, McCaffrey N, Brown C, Ferrara MS, Macciocchi S, Baumgartner T, Walpert K. ImPact test-retest reliability: Reliably unreliable? Journal of Athletic Training 2013; 48:506–511.
42. McCrory P, Meeuwisse W, Johnston K, Dvorak J, Aubry M, Molloy M, Cantu R. Consensus statement on concussion in sport: The 3rd International Conference on Concussion in Sport held in Zurich, November 2008. Journal of Athletic Training 2009;44: 434–448.
43. Guskiewicz KM, Bruce SL, Cantu RC, Ferrara MS, Kelly JP, McCrea M, Putukian M, McLeod TC. Research based recommendations on management of sport related concussion: Summary of the National Athletic Trainers' Association position statement. British Journal of Sports Medicine 2006;40:6–10.
44. Moser RS, Iverson GL, Echemendia RJ, Lovell MR, Schatz P, Webbe FM, Ruff RM, Barth JT. Neuropsychological evaluation in the diagnosis and management of sports-related concussion. Archivees of Clinical Neuropsychology 2007;22:909–916.
45. Gardner A, Shores EA, Batchelor J, Honan CA. Diagnostic efficiency of ImPACT and CogSport in concussed rugby union players who have not undergone baseline neurocognitive testing. Applied neuropsychology. Adult 2012;19:90–97.
46. Schmidt JD, Register-Mihalik JK, Mihalik JP, Kerr ZY, Guskiewicz KM. Identifying Impairments after Concussion: Normative data vs individualized baselines. Medicine & Science in Sports & Exercise 2012;44:1621–1628.
47. Putukian M. Panel discussion: Updating the SCAT2: Are there components that should be added or deleted? Do we a need a SCAT3?; Poster presented at the annual conference of the 4th International Consensus Conference on Concussion in Sport; 2012; November 1; Zurich, Switzerland.
48. Elbin RJ, Kontos AP, Kegel N, Johnson E, Burkhart S, Schatz P. Individual and combined effects of LD and ADHD on computerized neurocognitive concussion test performance: Evidence for separate norms. Archives of Clinical Neuropsychology 2013;28: 476–484.
49. Schatz P, Moser RS, Solomon GS, Ott SD, Karpf R. Prevalence of invalid computerized baseline neurocognitive test results in high school and collegiate athletes. Journal of Athletic Training 2012;47: 289–296.
50. Lichtenstein JD, Moser RS, Schatz P. Age and test setting affect the prevalence of invalid baseline scores on neurocognitive tests. American Journal of Sports Medicine 2014;42:479–484.

51. Echemendia RJ, Cantu RC. Return to play following sports-related mild traumatic brain injury: The role for neuropsychology. Applied Neuropsychology 2003;10:48–55.
52. Lovell MR, Collins MW. Neuropsychological assessment of the college football player. Journal of Head Trauma Rehabilitation 1998;13:9–26.
53. Lovell MR, Collins MW, Maroon JC, Cantu R, Hawn K, Burke CJ. Inaccuracy of symptom reorting following concussion in athletes. Medicine & Science in Sports & Exercise 2002;34:S298.
54. McCrea M, Hammeke T, Olsen G, Leo P, Guskiewicz K. Unreported concussion in high school football players: implications for prevention. Clinical Journal of Sport Medicine 2004;14:13–17.
55. SportingNews. NFL concussion poll: 56 percent of players would hide symptoms to stay on field. Volume 2012; http://aol.sportingnews.com/nfl/story/2012-11-11/nfl-concussions-hidesymptoms-sporting-news-midseason-players-poll. Accessed December 4, 2012.
56. Schatz P, Glatts C. ''Sandbagging'' baseline test performance on ImPACT, without detection, is more difficult than it appears. Archives of Clinical Neuropsychology 2013;28:236–244.
57. Erdal K. Neuropsychological testing for sports-related concussion: how athletes can sandbag their baseline testing without detection. Archives of Clinical Neuropsychology 2012;27:473–479.
58. Iverson GL, Collins MW, Lovell MR. Predicting recovery time from concussion in high school football players. Journal of the International Neuropsychological Society 2007;13:65.
59. Resch JE, Macciocchi SN, Ferrara MS. Equivalence of alternative forms of computerized neuropsychological test. St. Louis, MI: National Athletic Trainers' Association; 2012.
60. Franzen MD. Reliability and validity in neuropsychological assessment. New York, NY: Plenum Press; 1989.
61. Franzen MD. Reliability and validity in neurological assessment. New York, NY: Kluwer Academic/Plenum Press; 2000.
62. Jacobson NS, Revenstorf D. Statistics for assessing the clinical significance of psychotherapy issues: Issues, problems, and new developments. Behavioral Assessment 1988;10:133–145.
63. Jacobson NS, Truax P. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. Journal of Consulting & Clinical Psychology 1991;59:12–19.
64. Jacobson NS, Follette WC, Revenstorf D. Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. Behavioral Therapy 1984;15:336–352.
65. Chelune GJ, Naugle RI, Luders H, Sedlak J, Awad IA. Individual change after epilepsy surgery: Practice effects and base-rate information. Neuropsychology 1993;7:41–52.
66. Heaton RK, Temkin NR, Dikmen SS, Avitable N, Taylor MJ, Marcotte TD, Grant I. Detecting change: A comparison of three neuropsychological methods using normal and clinical samples. Archives of Clinical Neuropsychology 2001;16:75–91.
67. Temkin NR, Heaton RK, Grant I, Dikmen SS. Detecting significant change in neuropsychological test performance: A comparison of four models. Journal of the International Neuropsychological Society 1999;5:357–369.
68. Iverson GL. Interpretation of Mini-Mental State Examination scores in community-dwelling elderly and geriatric neuropsychiatry patients. International Journal of Geriatric Psychiatry 1998;13: 661–666.
69. Iverson GL. Interpreting change on the WAIS-III/WMS-III in clinical samples. Archives of Clinical Neuropsychology 2001;16: 183–191.
70. Barr WB, McCrea M. Sensitivity and specificity of standardized neurocognitive testing immediately following sports concussion. Journal of the International Neuropsychological Society 2001;7: 693–702.
71. Hinton-Bayre AD, Geffen GM, Geffen LB, McFarland KA, Friis P. Concussion in contact sports: Reliable change indices of impairment and recovery. Journal of Clinical & Experimental Neuropsychology 1999;21:70–86.
72. Chelune GJ. Assessing reliable neuropsychological change. In: Franklin RD, editor. Prediction in forensic and neuropsychology: sound statistical practices. Mahwah, NJ: Lawrence Erlbaum Associates; 2003. pp 65–88.
73. Resch JE, Macciocchi S, Ferrara MS. Preliminary evidence of equivalence of alternate forms of the ImPACT. Clinical Neuropsychology 2013;27:1265–1280.
74. Axelrod BN, Wall JR. Expectancy of impaired neuropsychological test scores in a non-clinical sample. International Journal of Neuroscience 2007;117:1591–1602.
75. Heaton RK, Grant I, Matthews CG. Comprehensive norms for an extended Halstead-Reitan Battery: Demographic corrections, research findings, and clinical applications. Odessa, FL: Psychological Assessment Resources, Inc.; 1991.
76. Heaton RK, Miller SW, Taylor MJ, Grant I. Revised comprehensive norms for an expanded Halstead-Reitan Battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults professional manual. Lutz, FL: Psychological Assessment Resources; 2004.
77. Ingraham LJ, Aiken CB. An empirical approach to determining criteria for abnormality in test batteries with multiple measures. Neuropsychology 1996;10:120–124.
78. Iverson GL, Brooks BL, Holdnack JA. Misdiagnosis of cognitive impairment in forensic neuropsychology. In: Heilbronner RL, editor. Neuropsychology in the courtroom: Expert analysis of reports and testimony. New York: Guilford Press; 2008. p 243–266.
79. Palmer BW, Boone KB, Lesser IM, Wohl MA. Base rates of ''impaired'' neuropsychological test performance among healthy older adults. Archives of Clinical Neuropsychology 1998;13: 503–511.
80. Schretlen DJ, Testa SM, Winicki JM, Pearlson GD, Gordon B. Frequency and bases of abnormal performance by healthy adults on neuropsychological testing. Journal of the International Neuropsychological Society 2008;14:436–445.
81. Brooks BL, Iverson GL, White T. Substantial risk of ''Accidental MCI'' in healthy older adults: Base rates of low memory scores in neuropsychological assessment. Journal of the International Neuropsychological Society 2007;13:490–500.
82. Crawford JR, Garthwaite PH, Gault CB. Estimating the percentage of the population with abnormally low scores (or abnormally large score differences) on standardized neuropsychological test batteries: A generic method with applications. Neuropsychology 2007; 21:419–430. Test Software. Available online at: http://homepages.abdn.ac.uk/j.crawford/pages/dept/PercentAbnormKtests.htm. Accessed September 30, 2014.
83. Brooks BL, Holdnack JA, Iverson GL. Advanced clinical interpretation of the WAIS-IV and WMS-IV: Prevalence of low scores varies by level of intelligence and years of education. Assessment 2011;18:156–167.
84. Brooks BL, Iverson GL, Holdnack JA, Feldman HH. The potential for misclassification of mild cognitive impairment: A study of memory scores on the Wechsler Memory Scale-III in healthy older adults. Journal of the International Neuropsychological Society 2008;14:463–478.
85. Brooks BL, Iverson GL, Holdnack JA. Understanding and using multivariate base rates with the WAIS-IV/WMS-IV. In: Holdnack JA, Drozdick LW, Weiss LG, Iverson GL, editors. WAIS-IV/WMS-IV/ACS: advanced clinical interpretation. San Diego, CA: Elsevier Science; 2013. pp 75–102.
86. Iverson GL, Brooks BL, Holdnack JA. Evidence-based neuropsychological assessment following work-related injury. In: Bush SS, Iverson GL, editors. Neuropsychological assessment of work-related injuries. New York: Guilford Press; 2012. p 360–400.
87. Iverson GL, Brooks BL. Improving accuracy for identifying cognitive impairment. In: Schoenberg MR, Scott JG, editors. The little black book of neuropsychology: a syndrome-based approach. New York: Springer; 2011. pp 923–950.
88. Binder LM, Iverson GL, Brooks BL. To err is human: ''Abnormal'' neuropsychological scores and variability are common in healthy adults. Archives of Clinical Neuropsychology 2009;24:31–46.
89. Iverson GL, Brooks BL. Development of preliminary evidence-based criteria for cognitive impairment associated with sport-related concussion. British Journal of Sports Medicine 2009;43: i100.
90. Iverson GL. Evidence-based neuropsychological assessment of sport-related concussion. In: Webbe FM, editor. Handbook of sport neuropsychology. New York, NY: Springer Publishing Company; 2011. pp 131–154.
91. Aubry M, Cantu R, Dvorak J, Graf-Baumann T, Johnston K, Kelly J, Lovell M, McCrory P, Meeuwisse W, Schamasch P. Summary and agreement statement of the First International Conference on Concussion in Sport, Vienna 2001. Recommendations for the

improvement of safety and health of athletes who may suffer concussive injuries. British Journal of Sports Medicine 2002;36:6–10.

92. McCrory P, Johnston K, Meeuwisse W, Aubry M, Cantu R, Dvorak J, Graf-Baumann T, Kelly J, Lovell M, Schamasch P. Summary and agreement statement of the 2nd International Conference on Concussion in Sport, Prague 2004. British Journal of Sports Medicine 2005;39:196–204.
93. McCrory P, Meeuwisse W, Johnston K, Dvorak J, Aubry M, Molloy M, Cantu R. Consensus statement on concussion in sport: The 3rd International Conference on Concussion in Sport held in Zurich, November 2008. British Journal of Sports Medicine 2009; 43:i76–i90.
94. Schatz P, Ferris CS. One-month test-retest reliability of the ImPACT test battery. Archives of Clinical Neuropsychology 2013; 28:499–504.
95. Elbin RJ, Schatz P, Covassin T. One-year test-retest reliability of the online version of ImPACT in high school athletes. American Journal of Sports Medicine 2011;39:2319–2324.
96. Schatz P. Long-term test-retest reliability of baseline cognitive assessments using ImPACT. American Journal of Sports Medicine 2010;38:47–53.
97. Wechsler D. Wechsler Intelligence Scale for Children – 4th ed: Technical and interpretive manual. San Antonio, TX: Psychological Corporation; 2003.
98. Delis D, Kramer J, Kaplan E, Ober BA. California Verbal Learning Test - Children's Version. San Antonio, TX: The Psychological Corporation; 1994.
99. Delis DC, Kaplan E, Kramer JH. The Delis Kaplan Executive Function System: Technical Manual. San Antonio, TX: The Psychological Corporation; 2001.