PINK CONCUSSIONS

Citizen Petition to FDA

February 26, 2022
From Pink Concussions

The undersigned submits this petition pursuant to 21 CFR § 10.25 in the form of a citizen petition in 10.30 to the Federal Food & Drug Administration (FDA) to request the Commissioner revoke the clearance K202927, SyncThink, Inc.'s EYE-SYNC as an aid in diagnosis of concussion in the public interest.

A. Action requested: We respectfully request that FDA rescind SyncThink, Inc.'s 510(k) clearance K202927 for EYE-SYNC Indications for Use (IFU) as an aid in diagnosis of concussion.

Background:

SyncThink Inc. submitted a 510(k) application K202927 for EYE-SYNC as an aid in diagnosis of concussion on September 29, 2020. One year later, on October 2, 2021, it was granted a 510(k) clearance as an aid in diagnosis of concussion, product code GWN. The predicate device was the Oculogica EyeBOX, product code QEA.

The clinical trial associated with this submission must be listed on clinicaltrials.gov by the submitter, SyncThink Inc. It is listed here: https://clinicaltrials.gov/ct2/show/NCT04381767?lead=sync+think&draw=2&rank=2. This clinical trial began enrolling December 19, 2018, and completed June 23, 2020, prior to the September 29, 2020, submission of the 510(k).

For the reference of the reviewer: The SCAT5 is the Sports Concussion Assessment Tool 5. It is a standard assessment used to determine concussion via a symptom checklist and neurocognitive tests in athletes.

B. Statement of grounds:

1. The EYE-SYNC product does not provide any addition to sensitivity or specificity for assessment of concussion to the self-reported concussion symptom checklist that is required it be used in conjunction with. It is clear from the IFU, and SyncThink Inc.'s press releases, that EYE-SYNC only achieves an acceptable sensitivity and specificity in conjunction with a symptom checklist, and must be used in conjunction with a symptom checklist to reach a sensitivity/specificity of 82%/93%. However, the EYE-SYNC device's accuracy in diagnosing

concussion, when used alone, is much lower. Please see supporting material: a peer-reviewed article by Dr. Kimberly Harmon, *Diagnostic accuracy and reliability of sideline concussion evaluation: a prospective, case-controlled study in college athletes comparing newer tools and established tests*, April 4, 2021. One of the key findings of this paper (p.7) is “the EYE-SYNC Smooth Pursuit radial and tangential variability did not add to the diagnosis of concussion.” The sensitivity/specificity (area under the curve) of the EYE-SYNC Smooth Pursuit tangential variability is reported as 48%/58% and radial variability as 52%/61%. Area under the curve is reported as tangential variability 0.41 (0.30, 0.54) and radial variability 0.47 (0.34, 0.59). Within the same study, it was found that self-reported symptom checklists had the highest identified sensitivity/specificity of 81%/94%, which is very similar to the EYE-SYNC product in conjugation with symptom checklist. It is, therefore, unnecessary and inappropriate to clear a device for the diagnosis of a concussion when it does not provide any additive value in sensitivity and specificity. Clearance of the EYE-SYNC device - in that it does not provide additive value beyond a symptom checklist - is not in the public interest.

2. The clinical trial associated with the submission states that an exclusion criterion is: “Subjects without reported injury detailing alteration of consciousness (AOC) or altered mental status (AMS) defined as self-report, witness report or the following responses to the SCAT5 SSS: "Difficulty Remembering" ≥ 4 or "Confusion" ≥ 4, or scored less than 23 on the SCAT5 Standardized Assessment of Concussion (SAC) or greater than 25 on the SCAT5 Symptom Severity Score (SSS).” These exclusion criteria **<u>are</u>** actually the definition of concussion. Thus, only subjects with concussion are included in the subject population. This ensures a sensitivity of near 1 (or 100%) because the exclusion criteria, based on SCAT, exclude anyone without concussion. Such a test is not useful beyond the exclusion criteria, i.e., SCAT itself. Again, this shows that clearance of the EYE-SYNC device - in that it does not provide additive value beyond a symptom checklist - is not in the public interest.

3. The clinical reference standard in a concussion clinical study is either a) adjudication with physicians or b) some form of symptom checklist (e.g., Sports Concussion Assessment Tool 5 (SCAT5), etc.). In the case of adjudication, adjudicators must receive either a1) a score based on a symptom check list, or a2) the actual symptom check list itself in order to perform the adjudication. For the cases of a1), a2) or b) the clinical reference standard includes a symptom checklist. This SyncThink 510(k) is for a device which includes a symptom checklist. Thus, the clinical reference standard includes the very thing to which it is being compared, and is therefore circular. This is not good clinical trial design.

PINK CONCUSSIONS

C. Environmental Impact

“We claim categorical exclusion under 25.30, 25.31, 25.32, 25.33, or 25.34 of this chapter or an environmental assessment under 25.40 of this chapter.”

D. Economic Impact

Economic impact information will be submitted upon request of the commissioner.

E. Certification: "The undersigned certifies, that, to the best knowledge and belief of the undersigned, this petition includes all information and views on which the petition relies, and that it includes representative data and information known to the petition which are unfavorable to the petition."

Signature:

Katherine Price Snedaker, LCSW
CEO and Founder
PINK Concussions
203-984-0860
(She/Her)

15 Shorefront Park
Norwalk, CT 06854

Diagnostic accuracy and reliability of sideline concussion evaluation: a prospective, case-controlled study in college athletes comparing newer tools and established tests

Kimberly G Harmon [1], Bridget M Whelan,[1] Douglas F Aukerman,[2] Adam D Bohr,[3] J Matthew Nerrie,[4] Heather A Elkinton,[5] Marissa Holliday,[6] Sourav K Poddar,[7] Sara P D Chrisman [8,9] Matthew B McQueen[3]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bjsports-2020-103840).

For numbered affiliations see end of article.

Correspondence to
Dr Kimberly G Harmon, Family Medicine, University of Washington, Seattle WA 98195 USA, Washington, USA; kharmon@uw.edu

Accepted 4 April 2021

Check for updates

To cite: Harmon KG, Whelan BM, Aukerman DF, *et al*. *Br J Sports Med* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/bjsports-2020-103840

ABSTRACT

Objective To assess diagnostic accuracy and reliability of sideline concussion tests in college athletes.

Methods Athletes completed baseline concussion tests including Post-Concussion Symptom Scale, Standardised Assessment of Concussion (SAC), modified Balance Error Scoring System (m-BESS), King-Devick test and EYE-SYNC Smooth Pursuits. Testing was repeated in athletes diagnosed acutely with concussion and compared to a matched teammate without concussion.

Results Data were collected on 41 concussed athletes and 41 matched controls. Test–retest reliability for symptom score and symptom severity assessed using control athletes was 0.09 (−0.70 to 0.88) and 0.08 (−1.00 to 1.00) (unweighted kappa). Intraclass correlations were SAC 0.33 (−0.02 to 0.61), m-BESS 0.33 (−0.2 to 0.60), EYE-SYNC Smooth Pursuit tangential variability 0.70 (0.50 to 0.83), radial variability 0.47 (0.19 to 0.69) and King-Devick test 0.71 (0.49 to 0.84). The maximum identified sensitivity/specificity of each test for predicting clinical concussion diagnosis was: symptom score 81%/94% (3-point increase), symptom severity score 91%/81% (3-point increase), SAC 44%/72% (2-point decline), m-BESS 40%/92% (5-point increase), King-Devick 85%/76% (any increase in time) and EYE-SYNC Smooth Pursuit tangential variability 48%/58% and radial variability 52%/61% (any increase). Adjusted area under the curve was: symptom score 0.95 (0.89, 0.99), symptom severity 0.95 (95% CI 0.88 to 0.99), SAC 0.66 (95% CI 0.54 to 0.79), m-BESS 0.71 (0.60, 0.83), King-Devick 0.78 (0.69, 0.87), radial variability 0.47 (0.34, 0.59), tangential variability 0.41 (0.30, 0.54)

Conclusion Test–retest reliability of most sideline concussion tests was poor in uninjured athletes, raising concern about the accuracy of these tests to detect new concussion. Symptom score/severity had the greatest sensitivity and specificity, and of the objective tests, the King-Devick test performed best.

INTRODUCTION

The diagnosis of concussion on the sideline or shortly thereafter is challenging and relies heavily on subjective reporting of symptoms as well as assessments of cognition and balance.[1] Problems with the use of symptoms in the diagnosis of concussion have been well documented and include a reluctance of athletes to report symptoms due to internal or external pressure, an inability to recognise symptoms, or a delay in the development of symptoms after concussion.[2 3] Recent studies reported between 16% and 33% of college athletes had previously not reported a concussion and up to 68% of football players endorsed not disclosing a concussion.[4 5] Because of the problems associated with subjective reporting of symptoms, an accurate objective test for concussion is desirable.

Currently, the Concussion in Sport Group (CISG) recommends the use of the Sport Concussion Assessment Tool (SCAT) for the evaluation of concussion which is comprised of symptom, cognitive and balance assessments, and provides a standardised approach to the evaluation of concussion.[6 7] There have been four iterations of the SCAT which was first developed in 2004 with the most recent update in 2017. The core components of the SCAT, the Post-Concussion Symptom Score (PCSS), the Standardised Assessment of Concussion (SAC) and the modified Balance Error Scoring System (m-BESS), have remained relatively constant.[1 8–10]

The SCAT is the most widely used concussion assessment tool, however, warnings on the SCAT5 state ‘the diagnosis of concussion is a clinical judgement, made by a medical professional’ and that ‘the SCAT5 should not be used by itself to make, or exclude, the diagnosis of concussion’.[11] The elements of clinical judgement include an evaluation of all components of the SCAT and the clinical situation, however, there are not clear cut-off values for what constitutes a positive or negative test on any of the subcomponents leading to difficult decision making. In fact, it is widely recommended that a ‘multimodal’ evaluation, an evaluation which is inclusive of all aspects of the SCAT, be used when making a diagnosis of concussion.[12–15] Most studies recommending a multi-modal evaluation show that symptoms are the most accurate tool and that there is only incremental benefit to the addition of cognitive and balance assessments, leading some to recommend adding vestibular and oculomotor assessments to improve the acute diagnosis of concussion.[2 7 12–16]

This study evaluates all sub-components of the SCAT individually and two emerging tests, the King-Devick test EYE-SYNC Smooth Pursuit for diagnostic accuracy and reliability. Our aim is for

Table 1 Cohort characteristics

	Total (n=82)	Concussed (n=41)	Control (n=41)
Sex (%)			
Male	48 (59)	24 (59)	24 (59)
Female	34 (41)	17 (41)	17 (41)
Sport (%)			
Baseball	4 (5)	2 (5)	2 (5)
Cheer	4 (5)	2 (5)	2 (5)
Football	36 (44)	18 (44)	18 (44)
Gymnastics	2 (2)	1 (2)	1 (2)
Men's basketball	6 (7)	3 (7)	3 (7)
Men's soccer	2 (2)	1 (2)	1 (2)
Softball	12 (15)	6 (15)	6 (15)
Volleyball	6 (7)	3 (7)	3 (7)
Women's basketball	4 (5)	2 (5)	2 (5)
Women's soccer	6 (7)	3 (7)	3 (7)
Comorbidities (%)			
ADD/ADHD	6 (7)	3 (7)	3 (7)
Anxiety/depression	3 (4)	2 (5)	1 (2)
Headache/migraine	11 (13)	7 (17)	4 (10)
Year in school (%)			
1	24 (29)	12 (29)	12 (29)
2	26 (32)	15 (37)	11 (27)
3	20 (24)	10 (24)	10 (24)
4	11 (13)	4 (10)	7 (17)
5	1 (1)		1 (2)
Days between baseline and retest (median, IQR)			
SCAT (PCSS, SAC, m-BESS)	427 (152, 523)	429 (154, 521)	424 (149, 525)
King-Devick	383 (95, 462)	378 (103, 461)	392 (95, 476)
EYE-SYNC smooth pursuits	427 (152, 523)	429 (154, 521)	424 (149, 525)
Average Baseline Scores (average, range)			
SAC	27.0 (22–30)	26.9 (22–30)	27.1 (23–30)
King-Devick	44.3 (30–68)	43.9 (30–68)	44.8 (32–66)

ADD/ADHD, attention deficit disorder/attention deficit hyperactivity disorder; m-BESS, modified balance error scoring system; PCSS, post-concussion symptom scale; SAC, standardised assessment of concussion; SCAT, sport concussion assessment tool.

clinicians who evaluate acute concussion to possess an understanding of the psychometric properties of the individual tests in order to support clinical decision making.

METHODS

Cohort

Participants included National Collegiate Athletic Association (NCAA) Division I athletes from three universities who sustained a concussion during the 2018–2019 and 2019–2020 seasons, and matched controls. All athletes (both case and control) underwent baseline concussion testing per NCAA rules and institution protocol as part of routine clinical care. Testing included the SCAT, the King-Devick test and EYE-SYNC Smooth Pursuits. The baseline SCAT, in some cases, may have been performed at different time points than the King-Devick and EYE-SYNC Smooth Pursuits. For athletes suspected of a concussion, repeat sideline testing was performed within 24-hours. 'Sideline' tests refer to the battery of tests performed within minutes to 24 hours of an incident for the acute diagnosis of concussion, may have been performed in a locker room or private setting (not on the sideline) and excluded testing conducted after delayed reporting of concussion.

A diagnosis of concussion was based on the clinical assessment of the evaluating sports medicine physician using the definition from the CISG group.[1] All athletes diagnosed with concussion were managed using the local institution's concussion protocol. For each concussion diagnosis, a matched control-athlete was selected, and repeat testing completed within 2 weeks. Controls were matched based on team/sport, sex, 'concussion modifiers' (attention deficit disorder, learning disability, migraine headache disorder, mood disorder), and baseline test scores. Previous studies have suggested that concussion modifiers may affect test results on concussion tests and should be controlled for.[17] For matching baseline test scores, a similar SAC score was identified, followed by a King-Devick score that was within 3 s of the injured athlete.

Sport Concussion Assessment Tool

The SCAT assesses athletes using a standardised symptom scale (the PCSS) which includes both a symptom score (PCSS symptoms) and a symptom severity score (PCSS severity); the SAC, which includes measures of orientation, immediate memory, concentration and delayed recall; and the m-BESS which requires balancing in three stances for 20 s each with eyes closed, hands placed on hips while standing in socks on a firm surface while the assessor counts the number of times the athlete moves out of position.[18–21] The most recent version of the SCAT, the SCAT5 offers the option of using a 10-word list for the immediate and delayed recall portion

Table 2 Test–retest reliability among study controls (baseline to retest)

	Intraclass correlation		Weighted kappa	
	ICC (95% CI)	P value	Kappa (95% CI)	P value
King-Devick	0.71 (0.49 to 0.84)	**<0.001**	–	–
SAC	0.33 (–0.02 to 0.61)	**0.03**	–	–
Symptom Score	–	–	0.09 (–0.70 to 0.88)	0.61
Symptom Severity	–	–	0.08 (–1.00 to 1.00)*	0.64
m-BESS	0.33 (–0.02 to 0.60)	**0.03**	–	–
Tangential Variability	0.70 (0.50 to 0.83)	**<0.001**	–	–
Radial Variability	0.47 (0.19 to 0.69)	**<0.001**	–	–

Significant p values (<0.05) in bold
Trending p values (<0.10) in italic
*Weighted values of upper and lower bounds for 95% CI exceeded thresholds for Cohens Kappa. As such we report an inrteval of –1.00 o 1.00.
ICC, intraclass correlation coefficients; m-BESS, modified balance error scoring system; SAC, standardised assessment of concussion.

Table 3 Sensitivity, specificity, PPV, NPV and AAUC of concussion tests in the acute assessment of concussion

Test	Concussed (true positives)	Control (false positives)	Sensitivity, %	Specificity, %	PPV, %	NPV, %	AAUC
Symptoms							0.95 (0.89, 0.99)
Three or >increase in the no of reported symptoms	26	2	81	94	93	83	
Four or >increase in the no of reported symptoms	25	1	78	97	96	82	
Five or >increase in the no of reported symptoms	25	1	78	97	96	82	
Symptom Severity							0.95 (0.88, 0.99)
Three or >increase in the no of reported symptoms	29	6	91	81	83	90	
Six or >increase in the no of reported symptoms	29	1	85	97	97	86	
Ten or >increase in the no of reported symptoms	25	1	78	97	96	82	
Standardised Assessment of Concussion (5-word)							0.66 (0.54, 0.79)
1-point decline	15	12	47	63	56	54	
2-point declines	14	9	44	72	61	58	
3-point decline	12	2	38	9	86	64	
Modified Balance Error Scoring System							0.71 (0.60, 0.83)
Increase in three errors	12	11	41	61	52	55	
Increase in five errors	10	2	40	92	83	62	
Increase in eight errors	4	1	14	97	80	50	
King-Devick Test							0.78 (0.69, 0.87)
Any prolongation in time from baseline	28	5	85	76	78	84	
3 s or greater prolongation from baseline	22	2	68	88	85	86	
5 s or greater prolongation from baseline	35	15	54	95	92	87	
EYE-SYNC Smooth Pursuit							
Any increase in tangential variability from baseline	16	14	48	58	53	53	0.41 (0.30, 0.54)
Any increase in radial variability from baseline	13	9	52	61	59	54	0.47 (0.34, 0.59)

AAUC, adjusted area under the curve; NPV, negative predictive value; PPV, positive predictive value.

of the SAC or the 5-word list used in prior versions. In this study, the 5-word list was used as this was protocol at participating institutions.

King-Devick test

King-Devick methodology has been described in depth in previous manuscripts.[22–25] In brief, the King-Devick is a rapid number naming test, consisting of a demonstration card and three test 'cards' with rows of random single-digit numbers. All athletes in this study completed the King-Devick via electronic tablet. The athlete read the numbers on each screen from left to right as quickly as possible in English without any errors. The sum of the three test card times is recorded as the score. During the baseline testing, athletes complete the test twice and the fastest time is used. Postinjury, athletes read through the cards only once and time is recorded.

EYE-SYNC smooth pursuit

The EYE-SYNC Smooth Pursuit uses goggles embedded with eye tracking sensors to measure ocular-motor function by following a moving target. The associated tablet software uses an algorithm to analyse the position of the eye compared with the target over time to complete the assessment, measuring tangential variability and radial variability. The test provides clinicians with an objective assessment of slow, continuous eye movements while the patient's eyes follow a stimulus moving in a circular direction. Further detail on this assessment and programme has been described elsewhere.[26 27] This study examined only smooth pursuit. Other assessments of ocular motion and vestibular function can now be performed with EYE-SYNC, but these were not included as they became available only after the study started.

Diagnostic accuracy measures assessed

The test–retest reliability, sensitivity/specificity and area under a receiver operator characteristic (ROC) curve are important indicators of accuracy for tests used clinically.[28 29] Test–retest reliability measures the stability of a test, or the ability of the test to provide similar results in the same person at a repeated time point. With a reliable test, deviation in performance has a greater likelihood of being caused by a true difference in function rather than normal variation in the test itself. It is generally agreed, that for clinical utility, a test should have a test–retest reliability of 0.75 or higher.[29] A test with high sensitivity will be abnormal in nearly all individuals with the condition and have a low rate of false negatives. Conversely, a test with high specificity will be normal in nearly all individuals without the condition and have a low rate of false positives. The positive predictive value (proportion having a concussion among those with a positive test) and negative predictive value (proportion without concussion among those with a negative test) are also helpful in interpretation of clinical tests. The area under the curve (AUC) combines sensitivity and specificity and measures the ability of the test to correctly classify those with and without the condition. An AUC of 1.0 represents a perfect test while an AUC of 0.5 provides no information for correct classification. Excellent diagnostic value for AUC is represented at 0.90, 0.80 for good, 0.70 for fair and 0.60 for poor.[30]

Statistical analysis

Normality of all variables was assessed using the Shapiro-Wilk test. Test–retest reliability from baseline to follow-up was assessed using two-way intraclass correlation coefficients tests for SAC, BESS, King-Devick and smooth pursuit scores among control athletes. Quadratic weighted Cohen's

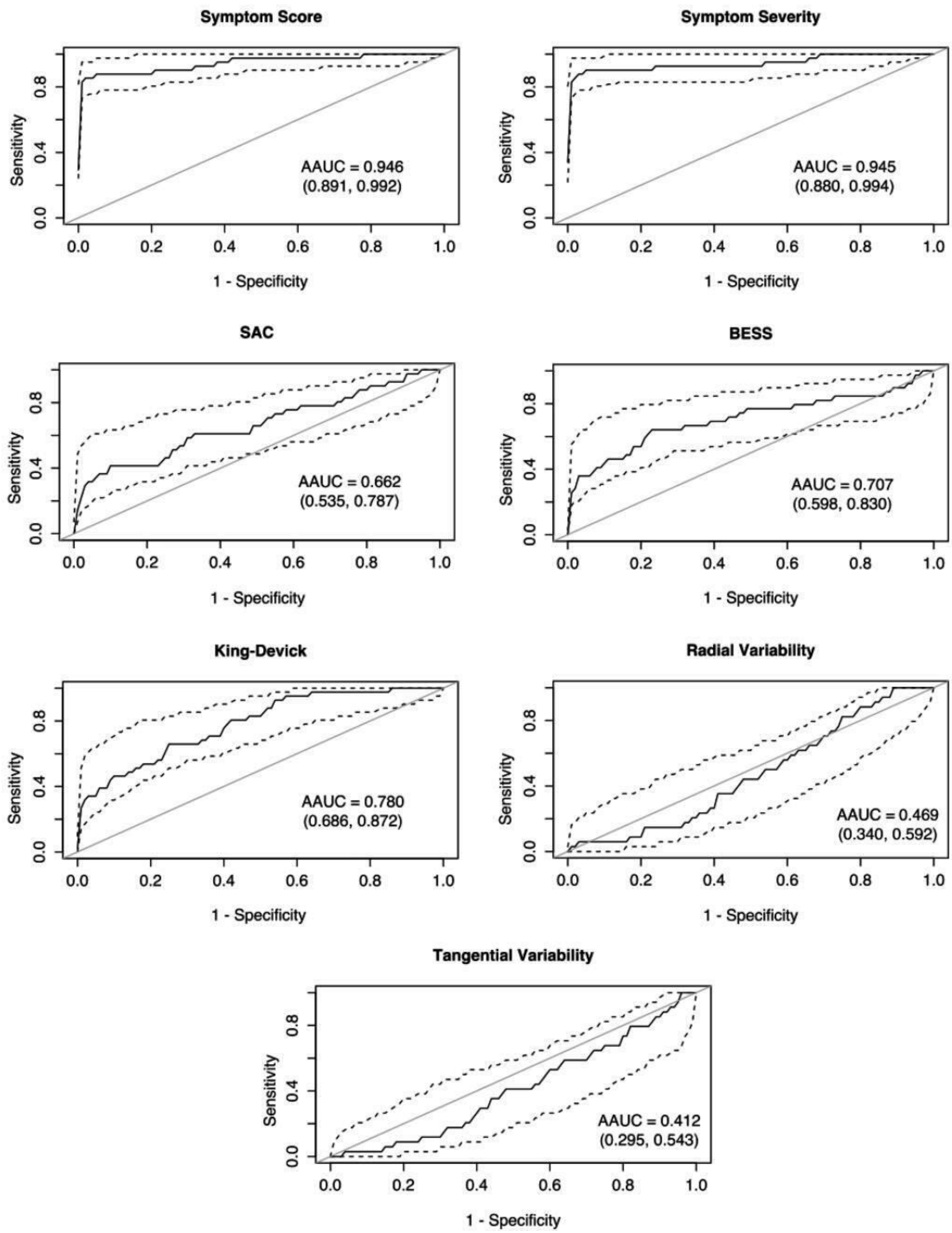

Figure 1 Receiver operator characteristic curves for concussion tests. AAUC, adjusted area under the curve; BESS, balance error scoring system; SAC, standardised assessment of concussion.

kappa coefficients were used to assess test–retest reliability of the SCAT symptom and severity score.[29 31] Sensitivity and specificity were calculated for all testing components. Semi-parametric covariate-adjusted ROC curves, adjusted AUC (AAUC) and 95% CI controlling for our matching variables, sport and year in school, were calculated to assess test validity for all postinjury assessments. All analyses were conducted in R Studio V.1.1.447. All athletes completed written informed consent. Patients or the public were not involved in study design, conduct, reporting or dissemination of this research.

RESULTS

A total of 41 concussions were diagnosed between September 2019 and February 2020. One institution had incomplete data for the SCAT (n=9 missing) and Smooth Pursuit (n=8 missing), and participants with incomplete data were removed from the component analysis. Demographics are shown in table 1. Twenty-four (59%) athletes were male with sport breakdown of football (n=18, 44%); softball (n=6, 15%); men's basketball, women's soccer, and volleyball (n=3 each, 7%); baseball, women's basketball and cheer (n=2 each, 5%); and gymnastics and men's soccer (n=1 each, 2%).

Diagnostic accuracy

For the SCAT analysis, baseline and postinjury data were collected on 32 concussed athletes and 32 matched controls and for the King-Devick and EYE-SYNC smooth pursuits 41 concussed and 41 control athletes were included. Reliability for all tests can be found in table 2. Sensitivity, specificity, positive and negative predictive value for various cut-offs and AAUC can be found in table 3. AAUC is also depicted in figure 1.

DISCUSSION

This study compares all subcomponents of the standard sideline evaluation tool, the SCAT, with emerging assessments of ocular function used for the acute evaluation of concussion in both concussed and control athletes matched for sex, sport, concussion modifiers and baseline testing characteristics. Our findings confirm the importance of symptoms in the evaluation of concussion and demonstrate the limitations of the objective portions of the SCAT, the SAC and m-BESS, as well as emerging assessments, the King-Devick and EYE-SYNC Smooth Pursuits. Some studies of the diagnostic accuracy of the SCAT have larger numbers, but have used aggregate differences in repeated baselines of non-concussed athletes as controls, use complicated statistics, and do not specifically address the individual components of the SCAT.[13 16] Comparing concussed to non-concussed athletes matched for sport, concussion modifiers and preconcussion baseline test scores at a similar point in the season, offers a direct measure of the clinical utility of sideline concussion tests.

Test–retest reliability is an important psychometric parameter. Symptoms, the SAC, the m-BESS, and radial variability of the EYE-SYNC smooth pursuit all had poor reliability indicating that there is significant variability from one test to a repeat even in non-concussed individuals. The King-Devick and tangential variability of EYE-SYNC smooth pursuit were the most reliable tests in the battery we tested, and approached, but did not exceed, the generally accepted threshold for clinical utility of 0.75. The time between the baseline administration and the repeat tests averaged over a year and may have contributed to the generally poor reliability. This interval, however, reflects how baselines are clinically used at universities in the USA. The NCAA requires baseline testing for concussion be done once, at matriculation. Athletes typically participate in a sport for four or more years and, if concussed later in their career, there may be a long time between administration of the initial baseline and the repeat test. Previous, larger studies have shown that test–retest reliability for the SAC remained stable between 7 days and 196 days and 1 year and 2 years.[18 29] Our values for test–retest reliability are similar to these larger studies supporting their validity (online supplemental table 1). A recent legal settlement by the NCAA requires annual baseline testing (the type is unspecified), thus, it is likely that universities will increase the frequency of baseline testing. The stability of reliability measurement across different studies at different time points suggests that more frequent baseline testing in this population is unlikely to improve diagnostic accuracy, except, perhaps in symptom score and severity.[18 29]

This study demonstrated symptoms which occur after a suspected concussion are the most accurate way to diagnosis a concussion with an AAUC estimate of 0.95 for both symptom score and symptom severity score. These findings remain consistent with other studies which have shown the sensitivity of symptoms to be between 84% and 98%, specificity between 97% and 100% and AUC 0.88-0.98[13 18 32–34] (table 4). The reliance of symptoms for the diagnosis of concussion is problematic as athletes may not be aware of, or report their symptoms, symptoms may be delayed in development, or they may deliberately not disclose their symptoms to avoid removal from sport. If an athlete develops concussion symptoms after a suspected concussive incident which cannot be explained by another injury (cervical strain, eye injury), the diagnosis of concussion can be made without additional testing.

The two most widely used objective portions of the SCAT, the SAC and the m-BESS have long been the primary evaluation tool for acute concussion. The SAC with 5-word recall performed relatively poorly in this study, with an AAUC of 0.66 (95% CI 0.54 to 0.79). Previous studies of the SAC in the early 2000s suggested sensitivity ranging from 72% to 80% and specificity ranging from 91% to 95%, although cut-off points were not always clear.[32 35–37] In more recent studies, sensitivity has ranged from 20% to 52% and specificity from 21% to 91%[33 38–40] (table 4).

Perhaps earlier studies reflected athlete naivety to the use of the SAC which has now become quite ubiquitous in sport although word and number lists should be varied with each administration. Many athletes take the SAC repeated times from middle school through college. The recent CISG recommended increasing both immediate and delayed recall sections of the test from 5 to 10 words.[41] This may decrease skew and potentially increase the sensitivity of the SAC as a diagnostic test for concussion, but it remains to be studied. The m-BESS also had relatively poor clinical performance with low sensitivity but higher specificity, yielding an AAUC of 0.71. This suggests that if balance (as assessed by the m-BESS) is impaired there is a high likelihood of concussion, but a lack of impairment does not necessarily exclude a concussion.

Among the objective tests assessed in this study, the King-Devick had the highest reliability (0.71), sensitivity (85%), specificity (76%) and AAUC (0.78). Most studies of the King-Devick test have not used controls, instead using return to baseline as a comparison.[13] All the studies with control athletes prior to ours included only males, while our sample included 41% female athletes. Galetta examined a small sample of 12 concussed and 14 control athletes and found an AUC of 0.92.[40] Hecimovich studied an Australian rules football team over a season and compared values of a small sample of concussed athletes (7) to non-concussed athletes (13) and calculated 98% sensitivity and 96% specificity.[42] A study of 84 Canadian Football League athletes diagnosed with concussion and compared with 63 controls who were not matched other than they were teammates found a 62% sensitivity and 84% specificity.[43] The reason for this large discrepancy in results from other studies is not clear. These were professional athletes, all male, and lack of motivation in controls to perform well was suggested as a possible explanation in the paper.[43] Our study suggests, that of the objective concussion assessments, the King-Devick may be the most accurate for the diagnosis of concussion. This finding needs to be confirmed.

The EYE-SYNC Smooth Pursuit radial (AAUC=0.47) and tangential variability (AAUC=0.41) both had similar sensitivity 48%/52%, specificity 58%/61%. EYE-SYNC Smooth Pursuit, without other measures, does not appear to be a good indicator of concussion. There are additional modalities that are currently part of the recommended EYE-SYNC protocol that may provide improved results.

Some studies have reported good success with the use of multimodal evaluation for concussion.[13 16] While sensitivity does

Table 4 Studies using controls examining the accuracy of sideline testing

Author	Year	Type of Athletes (concussed)	% female	No of Concussed Athletes	No of Control Athletes	Sensitivity/specificity of AUC* Symptoms	SAC	m-BESS	King-Devick	Eye-Sync	Matching of controls
McCrea[36]	2005	College football	0%	94	56	89/100	80/91	36/95			Matched on age, team, baseline performance on tests, controls only enrolled for first year of 3-year study
Barr[38]	2012	High school and college football	0%	59	31	94/72	46/87	31/71			31 non-injured athletes matched to injured group on the basis of age, years of education, cumulative grade point average and baseline performance on concussion assessment measures were selected as a control group. This sample of control subjects were 'yoked' to individual athletes based on the best fit to the aforementioned matching variables.
Putukian[33]	2015	College athletes (male 84%)	16%	32	23	84/100	41/91	25/100			Contact sport athletes, not matched individually, control testing done at 3 and 6 months after their baseline
Galetta[40]	2015	Youth ice hockey and lacrosse and college athletes	?	12	14		0.68*		0.92*		Control athletes playing a similar position (but not matched for age or baseline testing) had testing under the same conditions as the concussed athlete
Chin[18]	2016	High school (40%) and college athletes (60%)	16%	166	164	.88*	0.58*	0.62*			Control athletes were selected to match concussed athletes by school, sport team, Wechsler Test of Adult Reading performance, grade point average and age. Those with modifiers excluded.
Hecimovich[42]	2018	Australian rules football avg. age 19	0%	7	13				98/96		Postgame non-injured control players with no signs of SRC
Naidu[43]	2018	Canadian Football League	0%	84	63				62/84		The 64 control participants were randomly selected from team rosters, excluded those who had been diagnosed with a concussion during the competitive season. Tested after practice.
Harmon	2020	NCAA Division I athletes	41%	41	41	0.95*	0.65*	0.73*	0.80*	0.49*	Matched based on concussion modifiers, sex, team, baseline KD and SAC scores. Control testing done at same time of season

*Denotes area under the receiver operator characteristic curve
AUC, area under the curve; m-BESS, modified balance error scoring system; NCAA, National Collegiate Athletic Association; SAC, standardised assessment of concussion.

increase with multimodal evaluation, the primary contributor to the diagnosis of concussion in all these studies is symptoms.[13 16] In a large study inclusive of 1640 concussions in NCAA athletes, symptoms provided the best indicator of concussion with m-BESS incrementally adding to sensitivity.[16] In a study of the utility of a multimodal model of concussion, Garcia *et al* noted that the full model outperformed any single model, but, again, symptoms were the primary contributor to the diagnosis of concussion, resulting in a 93% sensitivity.[13] While the value of a multimodal evaluation cannot be dismissed, it should be recognised that if an athlete is not reporting the presence of symptoms accurately, the overall diagnostic accuracy of testing decreases considerably, and an understanding of the individual psychometric properties of each test becomes critical. Finally, concussions are heterogeneous, affecting different parts of the brain and leading to different manifestations and deficiencies making the use of various tests which challenge different systems important.

The limitations of this study include the relatively small sample size of 41 athletes with concussion and matched controls, and only 32 with complete SCAT data, however, it is the only study that compares the current standard, the SCAT, with emerging tests, the King-Devick and EYE-SYNC, breaks down the three major components of the SCAT individually, includes a heterogeneous mix of athletes with concussion and has controls matched for concussion modifiers and baseline testing scores. An a priori power analysis was not conducted for this study. Control athlete testing was not performed at the same time as concussed athlete testing (within 2 weeks rather than within 24 hours), possibly introducing bias if other aspects that could influence results had changed during this time (such as timing related to a game, travel or the end of the competitive season). Test–retest reliability also may be influenced by the duration between baseline and repeat testing although studies suggest these measures have been stable across different time frames.[18 29] While this study represents actual clinical practice within the requirements of the NCAA, reliability metrics could be altered if baseline testing was consistently updated before each season. In addition, the cohort included athletes at division I universities and findings may not generalise to athletes at other levels of play or younger ages. Strengths include the use of control athletes, and the broad sex and sport representation of athletes with concussions.

What are the findings?

- The most commonly used objective tests for the sideline diagnosis of concussion (Standardised Assessment of Concussion and modified Balance Error Scoring System) have poor diagnostic accuracy.
- The King-Devick test performed the best of the objective tests although it still misclassified 15% of concussions.
- An emerging tool, the EYE-SYNC Smooth Pursuit radial and tangential variability did not add to the diagnosis of concussion.

How might it impact on clinical practice in the future?

- Clinicians will have a better understanding of the strengths and weaknesses of the different sideline concussion tests.
- Clinicians may choose to use the 10-word Sport Concussion Assessment Tool (if they are not already) given the limitations of the 5-word version.
- Some clinicians may want to investigate use of the King-Devick as it was the most accurate diagnostic sideline concussion test in the study.

CONCLUSIONS

The acute diagnosis of concussion is difficult. Diagnosis currently relies on symptom report, given the lack of reliable objective tests for concussion. In this study, the diagnostic accuracy and reliability of traditional tests used for the diagnosis of concussion such as the SAC and the m-BESS was poor, the King-Devick had the best performance, and the EYE-SYNC Smooth Pursuit radial and tangential variability did not add to the diagnosis of concussion. It is important to understand the limitations and subtleties of the psychometric properties of individual sideline concussion tests to arrive at the best possible clinical decision for the diagnosis of concussion.

Author affiliations
[1]Family Medicine, University of Washington, Seattle, Washington, USA
[2]Department of Family Medicine, Oregon State University, Corvallis, Oregon, USA
[3]Department of Integrative Physiology, University of Colorado Boulder, Boulder, Colorado, USA
[4]Intercollegiate Athletics, University of Washington, Seattle, Washington, USA
[5]Intercollegiate Athletics, Oregon State University, Corvallis, Oregon, USA
[6]Intercollegiate Athletics, University of Colorado Boulder, Boulder, Colorado, USA
[7]Department of Family Medicine, University of Colorado - Anschutz Medical Campus, Aurora, Colorado, USA
[8]Center for Child Health, Behavior and Development, Seattle Children's Research Institute, Seattle, Washington, USA
[9]Harborview Injury Prevention and Research Center, Seattle, Washington, USA

Twitter Kimberly G Harmon @DrKimHarmon and Sara P D Chrisman @SPDChrismanMD

Contributors All authors contributed to study design, data collection, and writing and review of the manuscript. KGH, ADB, BMW, SPDC and MBM contributed to data analysis as well.

Funding Funding for this study was provided by a generous donation from University of Washington alumni Jack and Luellen Charneski.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval This study was approved by the Human Subjects Division at the University of Washington (STUDY00008085) with reliance agreements at the University of Colorado and Oregon State University.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. Data requests may be made to kharmon@uw.edu.

ORCID iDs
Kimberly G Harmon http://orcid.org/0000-0002-3670-6609
Sara P D Chrisman http://orcid.org/0000-0001-5373-7223

REFERENCES

1 McCrory P, Meeuwisse W, Dvořák J, *et al*. Consensus statement on concussion in sport-the 5th international conference on concussion in sport held in Berlin, October 2016. *Br J Sports Med* 2017;51:838–47.
2 Harmon KG, Clugston JR, Dec K, *et al*. American medical Society for sports medicine position statement on concussion in sport. *Br J Sports Med* 2019;53:213–25.
3 Kroshus E, Garnett B, Hawrilenko M, *et al*. Concussion under-reporting and pressure from coaches, teammates, fans, and parents. *Soc Sci Med* 2015;134:66–75.

4 Anderson M, Petit KM, Wallace J, *et al*. Predictors of concussion nondisclosure in collegiate Student-Athletes. *J Athl Train* 2021. doi:10.4085/1062-6050-0102.20. [Epub ahead of print: 15 Jan 2021].
5 Kerr ZY, Register-Mihalik JK, Kroshus E, *et al*. Motivations associated with nondisclosure of self-reported concussions in former collegiate athletes. *Am J Sports Med* 2016;44:220–5.
6 Echemendia RJ, Meeuwisse W, McCrory P, *et al*. The sport concussion assessment tool 5th edition (SCAT5): background and rationale. *Br J Sports Med* 2017;51:848–50.
7 Collins MW, Kontos AP, Okonkwo DO, *et al*. Statements of agreement from the targeted evaluation and active management (team) approaches to treating concussion meeting held in Pittsburgh, October 15-16, 2015. *Neurosurgery* 2016;79:912–29.
8 McCrory P*et al*. Summary and agreement statement of the 2nd International Conference on concussion in sport, Prague 2004. *Br J Sports Med* 2005;39:i78–86.
9 McCrory P, Meeuwisse W, Aubry M, *et al*. Consensus statement on Concussion in Sport - The 4th International Conference on Concussion in Sport held in Zurich, November 2012. *Phys Ther Sport* 2013;14:e1–13.
10 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. *Br J Sports Med* 2009;43 Suppl 1:i76–90.
11 Sport concussion assessment tool - 5th edition. *Br J Sports Med* 2017;51:851–8.
12 Broglio SP, Ferrara MS, Piland SG, *et al*. Concussion history is not a predictor of computerised neurocognitive performance. *Br J Sports Med* 2006;40:802–5.
13 Garcia G-GP, Broglio SP, Lavieri MS, *et al*. Quantifying the value of multidimensional assessment models for acute concussion: an analysis of data from the NCAA-DoD care consortium. *Sports Med* 2018;48:1739–49.
14 Garcia G-GP, Lavieri MS, Jiang R, *et al*. A data-driven approach to unlikely, possible, probable, and definite acute concussion assessment. *J Neurotrauma* 2019;36:1571–83.
15 Garcia G-GP, Yang J, Lavieri MS, *et al*. Optimizing components of the sport concussion assessment tool for acute concussion assessment. *Neurosurgery* 2020;87:971–81.
16 Broglio SP, Harezlak J, Katz B, *et al*. Acute sport concussion assessment optimization: a prospective assessment from the care consortium. *Sports Med* 2019;49:1977–87.
17 McCrory P, Meeuwisse WH, Aubry M, *et al*. Consensus statement on concussion in sport: the 4th International Conference on concussion in sport held in Zurich, November 2012. *Br J Sports Med* 2013;47:250–8.
18 Chin EY, Nelson LD, Barr WB, *et al*. Reliability and validity of the sport concussion assessment Tool–3 (SCAT3) in high school and collegiate athletes. *Am J Sports Med* 2016;44:2276–85.
19 McCrea M, Kelly JP, Randolph C, *et al*. Standardized assessment of concussion (sac): on-site mental status evaluation of the athlete. *J Head Trauma Rehabil* 1998;13:27–35.
20 Azad AM, Al Juma S, Bhatti JA, *et al*. Modified balance error scoring system (M-BESS) test scores in athletes wearing protective equipment and cleats. *BMJ Open Sport Exerc Med* 2016;2:e000117.
21 Yengo-Kahn AM, Hale AT, Zalneraitis BH, *et al*. The sport concussion assessment tool: a systematic review. *Neurosurg Focus* 2016;40:E6.
22 Breedlove KM, Ortega JD, Kaminski TW, *et al*. King-Devick test reliability in national collegiate athletic association athletes: a national collegiate athletic Association-Department of defense concussion assessment, research and education report. *J Athl Train* 2019;54:1241–6.
23 Chrisman SPD, Harmon KG, Schmidt JD, *et al*. Impact of factors that affect reading skill level on King-Devick baseline performance time. *Ann Biomed Eng* 2019;47:2122–7.
24 Clugston JR, Chrisman SPD, Houck ZM, *et al*. King-Devick test time varies by testing modality. *Clin J Sport Med* 2018;Publish Ahead of Print.
25 Clugston JR, Houck ZM, Asken BM, *et al*. Relationship between the King-Devick test and commonly used concussion tests at baseline. *J Athl Train* 2019;54:1247–53.
26 SYNCTHINK I. Assessment Battery Overview, 2020. Available: https://syncthink.com/ [Accessed 17 Aug 2020].
27 Sundaram V, Ding VY, Desai M, *et al*. Reliable sideline ocular-motor assessment following exercise in healthy student athletes. *J Sci Med Sport* 2019;22:1287–91.
28 Harmon KG, Clugston JR, Dec K, *et al*. American medical Society for sports medicine position statement on concussion in sport. *Clin J Sport Med* 2019;29:87–100.
29 Broglio SP, Katz BP, Zhao S, *et al*. Test-Retest reliability and interpretation of common concussion assessment tools: findings from the NCAA-DoD care consortium. *Sports Med* 2018;48:1255–68.
30 Li F, He H. Assessing the accuracy of diagnostic tests. *Shanghai Arch Psychiatry* 2018;30:207–12.
31 Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968;70:213–20.
32 McCrea M. Standardized mental status testing on the Sideline after sport-related concussion. *J Athl Train* 2001;36:274–9.
33 Putukian M, Echemendia R, Dettwiler-Danspeckgruber A, *et al*. Prospective clinical assessment using Sideline concussion assessment Tool-2 testing in the evaluation of sport-related concussion in college athletes. *Clin J Sport Med* 2015;25:36–42.
34 Resch JE, Brown CN, Schmidt J, *et al*. The sensitivity and specificity of clinical measures of sport concussion: three tests are better than one. *BMJ Open Sport Exerc Med* 2016;2:e000012.
35 Barr WB, McCrea M. Sensitivity and specificity of standardized neurocognitive testing immediately following sports concussion. *J Int Neuropsychol Soc* 2001;7:693–702.
36 McCrea M, Barr WB, Guskiewicz K, *et al*. Standard regression-based methods for measuring recovery after sport-related concussion. *J Int Neuropsychol Soc* 2005;11:58–69.
37 McCrea M, Kelly JP, Randolph C, *et al*. Immediate neurocognitive effects of concussion. *Neurosurgery* 2002;50:1032–40.
38 Barr WB, Prichep LS, Chabot R, *et al*. Measuring brain electrical activity to track recovery from sport-related concussion. *Brain Inj* 2012;26:58–66.
39 Marinides Z, Galetta KM, Andrews CN, *et al*. Vision testing is additive to the sideline assessment of sports-related concussion. *Neurol Clin Pract* 2015;5:25–34.
40 Galetta KM, Morganroth J, Moehringer N, *et al*. Adding vision to concussion testing: a prospective study of Sideline testing in youth and collegiate athletes. *J Neuroophthalmol* 2015;35:235–41.
41 Echemendia RJ, Broglio SP, Davis GA, *et al*. What tests and measures should be added to the SCAT3 and related tests to improve their reliability, sensitivity and/or specificity in sideline concussion diagnosis? A systematic review. *Br J Sports Med* 2017;51:895–901.
42 Hecimovich M, King D, Dempsey AR, *et al*. The King-Devick test is a valid and reliable tool for assessing sport-related concussion in Australian football: a prospective cohort study. *J Sci Med Sport* 2018;21:1004–7.
43 Naidu D, Borza C, Kobitowich T, *et al*. Sideline concussion assessment: the King-Devick test in Canadian professional football. *J Neurotrauma* 2018;35:2283–6.