What are the critical elements of sideline screening that can be used to establish the diagnosis of concussion? A systematic review

Patricios JS[1,2], Fuller G[3], Ellenbogen R[4], Herring S[5], Kutcher J[6], Loosemore M[7], Makdissi M[8,9], McCrea M[10], Putukian M[11,] Raftery M [12]

Corresponding author
Jon Patricios
Morningside Sports Medicine
PO Box 1267
Parklands
2121
South Africa
jpat@mweb.co.za
+27 11 8839000

1. Section Sports Medicine, Faculty of Health Sciences, University of Pretoria, Pretoria, South Africa
2. Department of Emergency Medicine, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, Sourth Africa
3. Centre for Urgent and Emergency Care Research, School of Health and Related Research, University of Sheffield, Sheffield, UK
4. Department of Neurological Surgery, University of Washington, Seattle, Washington, USA
5. Departments of Rehabilitation Medicine, Orthopedics and Sports Medicine and Neurological Surgery, University of Washington, Seattle, Washington, USA
6. The Core Institute, University of Michigan, Phoenix, Arizona, USA
7. Institute of Sport Exercise and Health, University College London, London, UK
8. Florey Institute of Neuroscience and Mental Health, Austin Campus, Melbourne Brain Centre, Heidelberg, Australia.
9. Olympic Park Sports Medicine Centre, Melbourne, Australia
10. Department of Neurosurgery, Medical College of Wisconsin, Milwaukee, WI, USA
11. Director of Athletic Medicine, Princeton University, Princeton, New Jersey, USA
12. Chief Medical Officer, World Rugby, Dublin, Ireland

ABSTRACT

BACKGROUND : Sideline detection is the first and most significant step in recognising a potential concussion and removing an athlete from harm. This systematic review aims to evaluate the critical elements aiding sideline recognition of potential concussions including screening tools, technologies and integrated assessment protocols.
DATA SOURCES : Bibliographic databases, grey literature repositories and relevant websites were searched from 1 January 2000 to 30 September 2016. A total of 3562 articles were identified.
STUDY SELECTION : Original research studies evaluating a sideline tool, technology or protocol for sports-related concussion were eligible, of which 27 studies were included.

DATA EXTRACTION : A standardised form was used to record information. The QUADAS-2 and Newcastle-Ottawa tools were used to rate risk of bias. Strength of evidence was assessed using the Grades of Recommendation, Assessment, Development and Evaluation Working Group system.
DATA SYNTHESIS : Studies assessing symptoms, the King-Devick test and multimodal assessments reported high sensitivity and specificity. Evaluations of balance and cognitive tests described lower sensitivity but higher specificity. However, these studies were at high risk of bias and the overall strength ofevidence examining sideline screening tools was very low. A strong body of evidence demonstrated that head impact sensors did not provide useful sideline concussion information. Low-strength evidence suggested a multimodal, multitime-based concussion evaluation process incorporating video review was important in the recognition of significant head impact events and delayed onset concussion.
CONCLUSION : Conclusion In the absence of definitive evidence confirming the diagnostic accuracy of sideline screening tests, consensus-derived multimodal assessment tools, such as the Sports Concussion Assessment Tool, are recommended. Sideline video review may improve recognition and removal from play of athletes who have sustained significant head impact events. Current evidence does not support the use of impact sensor systems for real-time concussion identification.

Introduction

Despite a consensus definition of sports-related concussion (SRC) having been well elucidated (McCrory 2013), its accurate and immediate recognition remains challenging. Central to effective concussion management is early detection of the condition which then facilitates removal of the affected player and referral to a clinical network capable of more intensive evaluation, management and guidance of the return-to-play process.

Suffering a concussion has repeatedly been shown to decrease reaction times (Covassin 2008), affect balance (McCrea 2003) and slow cognition (Matser 1999) making the likelihood of suffering an additional concussion or musculoskeletal injury greater (Herman 2015). Repeated concussions have a cumulative effect (Guskiewicz 2003, Iverson 2004) and may have long term consequences such as depression or neurodegenerative disorders (Guskiewicz 2005, Mckee 2009 and McCrory 2013). The aforementioned issues suggest that sideline diagnosis of concussion and subsequent removal from play are priorities in preventing potential adverse sequelae.

This systematic review evaluates critical elements for sideline recognition of potential concussions. The review focusses on three essential components: an assessment of existing clinical screening and diagnostic tools, an evaluation of emerging on-field and sideline technological instruments, and an assessment of integrated protocols used in professional collision sports.

Methods

Study design

Expert consensus guidelines for the conduct of systematic reviews were followed and a detailed protocol stating an a priori analysis plan was registered before data collection (Cochrane 2008, Liberati 2009, Leeflang 2008). The review question, and inclusion and exclusion criteria, are detailed in Table 1.

Identification of evidence

An extensive range of electronic information sources were examined including all major bibliographic databases, specialist sports medicine databases, grey literature repositories, and relevant websites. Additional information sources included forwards and backwards citation searching, author searching, reference checking and contact with experts. Search strategies for bibliographic databases were developed iteratively in conjunction with an information services specialist and underwent external peer review. Searches were conducted for original research published between xxxx (corresponding to the modern definition of concussion) and Week xx, xxxx 2016 and were otherwise unrestricted. Current awareness searches were conducted in MEDLINE and Embase (Week x, xxxx 2016) immediately prior to submission. Full details on information sources and search strategies are presented in the web appendix.

Selection of evidence and data extraction

All original research studies identified during searches were assessed in a four stage process. Firstly, initial screening of titles and abstracts for relevance was conducted by two independent reviewers. Secondly, these reviewers' examined full-text articles as required to assess eligibility. Thirdly, eligible studies meeting review inclusion criteria were classified into 3 domains pertaining to: sideline screening tests (comprising subtopics of clinical signs and symptoms, balance tests, oculomotor assessments, cognitive tests, and multi-modal testing strategies); sports-specific integrated diagnostic protocols; and technology (defined in Table 1). Finally, data extraction was performed separately for eligible studies within each sub-topic by separate teams consisting of two reviewers. A single unblinded reviewer extracted information on study characteristics, methodology and results using a standardised data extraction form; and a second reviewer independently checked data for consistency and accuracy. Although not eligible for inclusion, identified review articles were examined to provide a strategic overview and cross-check references. Where necessary study authors were contacted to provide

additional information. In cases of disagreement, consultation with a third author was planned, with consensus derived by arbitration.

Table 1. Review question and inclusion criteria

Review Question	
What are the critical elements of sideline screening that can be used to establish the diagnosis of concussion or suspected concussion?	
Inclusion Criteria	
Population	Athletes competing in sporting activity and sustaining a non-trivial head impact event [includes: any nationality, gender, age group, or level of performance].
Intervention / index tests	Any sideline screening assessment used to detect suspected concussion following sports-related significant head impact events [including: historical features, symptoms, physical findings, clinical tests, or technologies]
Outcome / reference standard	Concussion, clinically diagnosed by a registered medical practitioner.
Study design	Published or unpublished studies of any research design.
Review sub-topics	
Sideline screening tests	Utility of sideline clinical tests to detect suspected concussion, including: • Symptoms and clinical signs • Balance tests • Oculomotor tests • Cognitive tests • Multimodal assessments (either joint use of individual sideline tests, or multi-faceted instruments)
Integrated diagnostic protocols	Utility of system level interventions to detect and manage significant head impact events during sporting activities
Technology	Utility of technology to detect suspected concussion during sporting activites

Appraisal of quality, data synthesis and statistical analyses

Included studies were assessed for risk of bias using peer reviewed critical appraisal checklists appropriate to study design. The revised QUDAS-2 tool was used for diagnostic accuracy studies (Whiting 2011). Observational studies were assigned a level of evidence based on the hierarchal 'level of evidence' grading system established by the Cochrane Collaboration (Cochrane 2008). A single un-blinded reviewer within each sub-group team assessed risk of bias, with a second reviewer

independently checking the assessment for validity. Any disagreement between reviewers was resolved by consensus and consultation with a third author with expertise in epidemiology and critical appraisal. Results are presented descriptively, with reported point estimates and 95% confidence intervals. A narrative synthesis was pre-specified in the event that clinically and methodologically homogenous studies at low risk of bias were not identified. References were managed in EndNote (Thomson Reuters, CA, USA).

RESULTS

Study selection

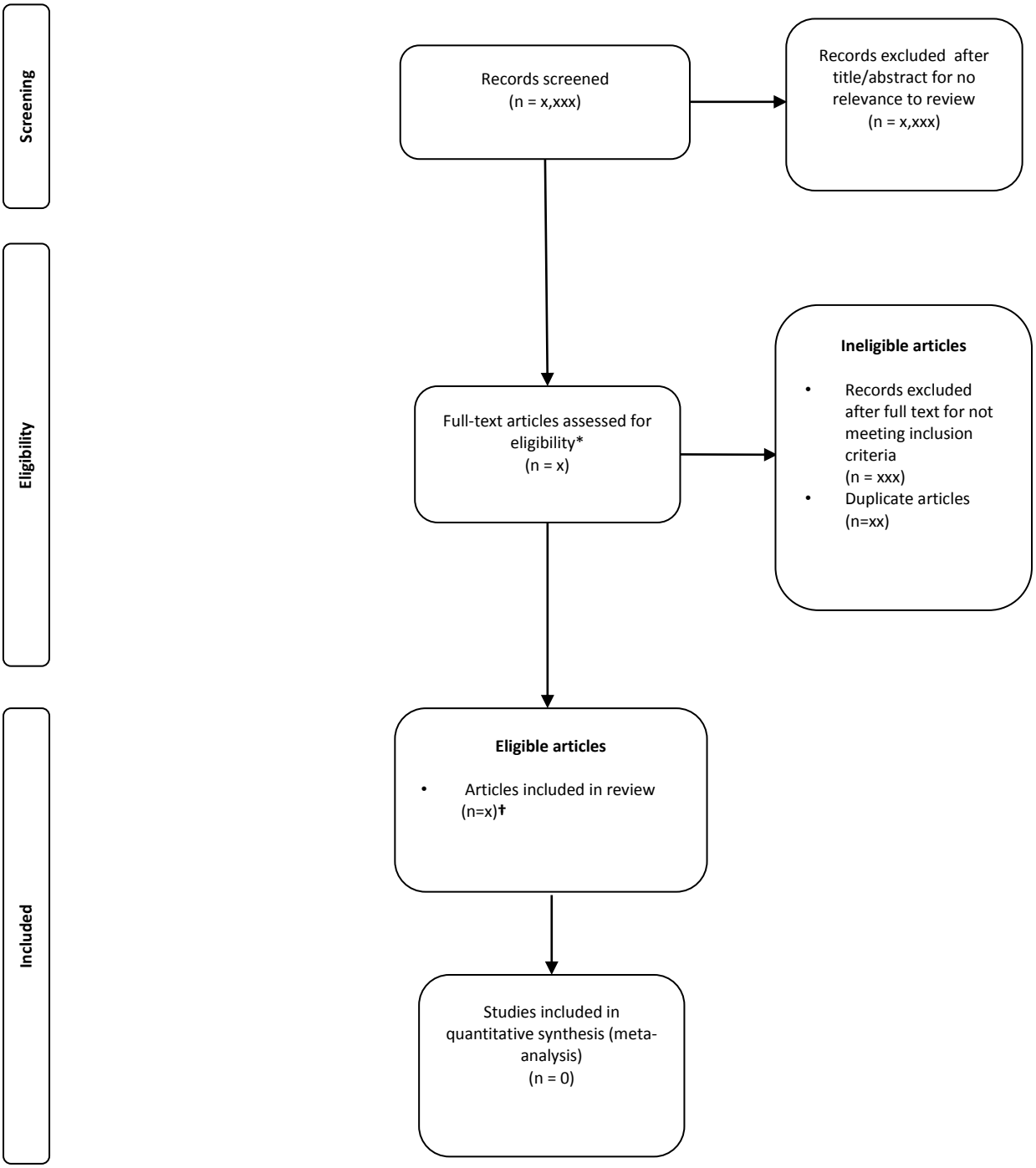
Xx,xxx citations were screened for eligibility, with the full text of xxx articles retrieved for detailed evaluation. During full text examination 27 studies were found meeting review inclusion criteria: sideline screening assessment (21 studies); integrated diagnostic protocols (1 study); and technology (5 studies). Figure 1 describes the selection of studies in detail.

Sideline screening tests

Characteristics of included studies

Twenty one studies met review inclusion criteria and reported interpretable data on the diagnostic accuracy of screening tests for the side-line identification of sport's concussion. These investigations included evaluation of sideline tests in a wide range of sports (American football, Australian football, soccer, ice hockey, field hockey, lacrosse, athletics, boxing, mixed martial arts, basketball, rugby league, rugby union, wrestling, athletics, crew, and sprint football); settings (Australia, New Zealand, South Africa, United States, and France); performance levels (high school, collegiate, amateur, professional); and age groups (children, adolescents, and young adults). The majority of studies were prospective cohort studies, with a single eligible retrospective cohort study identified (Marindes 2015). Overall source sample sizes were modest, ranging from n=27 to 337; however the number of participants included in diagnostic accuracy assessments were very low (median 30, interquartile range 2 – 337).

Fig. 1 Flow of identification, screening, eligibility and inclusion criteria for the literature review of sideline diagnosis of concussion.



A wide range of individual sideline tests was examined, comprising: Symptoms - Maddock 1995, McCory 2000, Erlanger 2003; Symptoms checklists - Graded Symptom Checklist (GSC, McCrea 2005), Concussion Symtpom Inventory (Barr 2012), Pitchside Concussion Assessment Tool symptom checklist (PSCA, Fuller

2014), Sports Concussion Assessment Tool 2 symptom checklist (SCAT, Putukian 2015); Clinical signs - Mental status evaluation (Fuller 2014); Oculomotor tests – King-Devick Test (KD, Galetta K 2011, Galetta K 2011b, King 2012, Galetta M 2013, Dhawan 2014, Leong 2014, Galetta K 2015, Leong 2015, Marinides 2015, Seidman 2015); Cognitive tests - Standardised Assessment of Concussion (SAC, Barr 2001, McCrea 2001, McCrea 2002, McCrea 2005, Echlin 2010, Barr 2012, Galetta M 2013, Marindes 2015, Galetta K 2015, Putukian 2015), Maddock's questions (Maddocks 1995, Fuller 2014), and orientation questions (Maddocks 1995); and balance assessments – Balance Error Scoring System (BESS, McCrea 2005, Echlin 2010, Barr 2012, Marindes 2015), Modified BESS (Putukian 2015), Tandem Stance Test (Fuller 2014), and Timed Tandem Gait (Galetta 2015).

Six eligible studies were identified which investigated either multi-faceted combined instruments or jointly used individual screening tests: SCAT 2 (Putukian 2015, Galetta M 2013); Pitchside Concussion Assessment Tool (PSCA, Fuller 2014); GSC, BESS, and SAC (McCrea 2005); SCAT2 and KD test (Galetta M 2013); SAC, Timed Tandem Gait Test and KD test (Galetta K 2015); and SAC, BESS and KD test (Marinides 2015). Cut-points and diagnostic thresholds varied significantly for individual tests across studies; further details are provided in the web appendix. Characteristics of the included studies examining sideline assessments are summarised in Table 2.

Seven potentially eligible studies were identified which recorded data on sideline tests and concussion, but did not report useable data on diagnostic accuracy (McCrory 2000 – Digital Subtraction Test and symptoms; Daniel 2002 – SAC; Nassiri 2002 –SAC; McCrea 1997 – SAC; McCrea 1998 – SAC; McCrea 2010 – Concussion Severity Inventory, BESS; Barr 2012 – Concussion Severity Inventory, BESS; McCrea 2013 – GSC, SAC). Where appropriate, data were extracted to allow analysis consistent with the review question (side-line testing following suspicious head impact events) and a standard diagnostic accuracy study design, rather than the investigators primary results (King 2012, Leong 2014, Leong 2015). Corresponding authors were contacted for clarification of methods and results if necessary.

Reference standards differed across studies in terms of content, timing, and type of assessor. The Sports Concussion Assessment Tool version 2 or 3 was the most commonly used formal assessment instrument (Echlin 2010, King 2012, Galetta M 2013, Fuller 2014, King 2015, Leong 2015, Seidman 2015) with the Military Acute Concussion Evaluation also utilised (Galetta K 2011, Leong 2014). The reference standard was unstructured clinical gestalt or unclear in remaining studies. The reference standard was generally

assessed at a single time point immediately after identification of a head impact event, but was delayed by 30 minutes or unclear in a minority of studies (King 2012, Galetta K 2015). Outcome assessors were generally qualified physicians, but comprised athletic trainers (Barr 2001, McCrea 2001, McCrea 2002, McCrea 2005, Galetta K 2011b, Barr 2012, Galetta K 2015, Leong 2015), lay people (Galetta K 2015) or was unclear (Galetta K 2013, Dhawan 2014) in some studies.

Table 2. Characteristics of included studies examining sideline screening assessments

Study	Setting	Study design	Sample Size (n=)*	Sport(s)	Level	Mean age (years±SE)	Gender (% male)	Index test(s)	Reference standard
Maddocks 1995	Aus	PCS	56	Australian Football	Professional	NR	100	Individual symptoms, Maddocks questions	Clinical diagnosis
McCrory 2000	Aus	PCS	303	Australian Football	Professional	NR	100	Individual symptoms	Clinical diagnosis
Barr 2001	US	PCS	118	American Football	Varsity High School	18.1 (NR)	NR	SAC	Clinical diagnosis
Erlanger 2003	US	PCS	47	American Football, Ice Hockey, Field Hockey, Wrestling, Soccer, Basketball	School Adolescents	17.6 (SD 2.23)	57	Individual symptoms	NR
McCrea 2001	US	PCS	118	American Football	Varsity High School	19.8±1.3	NR	SAC	Clinical diagnosis
McCrea 2002	US	PCS	91	American Football	Varsity High School	17.5±2.1	NR	SAC	Clinical diagnosis
McCrea 2005	US	PCS	150	American Football	Collegiate Adults	20.04 (SD 1.36)	100	GSC, BESS, SAC	Clinical diagnosis
Echlin 2010	US	PCS	67	Ice Hockey	Junior Adolescents	18.2 ± 1.2	100%	BESS, SAC	Clinical diagnosis + SCAT 2
Galetta K 2011	US	PCS	39	Boxing, mixed martial arts	Amateur - Adult	24	97	KD	MACE
Galetta K 2011b	US	PCS	219	American football, soccer, basketball	Collegiate athletics	20.3±1.4	83	KD	Clinical diagnosis
Barr 2012	US	PCS	90	American football	High school, collegiate	NR	100	CSI, SAC, BESS	Clinical diagnosis

King 2012	NZ	PCS	50	Rugby league	Amateur – Adult	22.4±4.1	100	KD	SCAT 2
Galetta M 2013	US	PCS	27	Ice hockey	Professional	25±5	100	KD, SAC	SCAT 2
Dhawan 2014	US	PCS	141	Hockey	High school athletics	NR	NR	KD	NR
Fuller 2014	UK, RSA, France	PCS	165	Rugby Union	Professional Adults	NR	100%	PSACA1 tool: Maddocks Questions, Symptoms checklist, Mental status assessment, Tandem Stance test	Clinical diagnosis + SCAT 3
Leong 2014	US	PCS	34	Boxing	Amateur - Adult	25.8±8.3	85	KD	MACE
Galetta K 2015	US	PCS	243	Ice hockey, lacrosse, Athletics	Amateur – Youth, Collegiate athletics	Youths: 11±3, Adults: 20±1	Youths: 84 Adults: 74	KD, Timed Tandem Gait, SAC	Clinical diagnosis
Leong 2015	US	PCS	127	American football, basketball	Collegiate athletics	19.6±1.2	94	KD	Modified SCAT 2
Marinides 2015	US	RCS	217	American football, lacrosse, soccer	Collegiate athletics	NR	70	KD, BESS, SAC	Clinical diagnosis
Putukian 2015	US	PCS	263	American Football, Rugby Union, Sprint Football, Crew	Collegiate Adults	20.33 (SD 1.74)	67%	SCAT2 symptom checklist, Modified BESS, SAC, SCAT2	Clinical diagnosis
Seidman 2015	US	PCS	337	American football	High school athletics	15.4 ± 1.3	100	KD	SCAT 3

AUS: Australia; GSC: Graded Symptom Checklist; PCS: prospective cohort study; RSA: Republic of South Africa; NR: Not reported; SCAT2: Sports Concussion Assessment Tool 2

Methodological quality of included studies

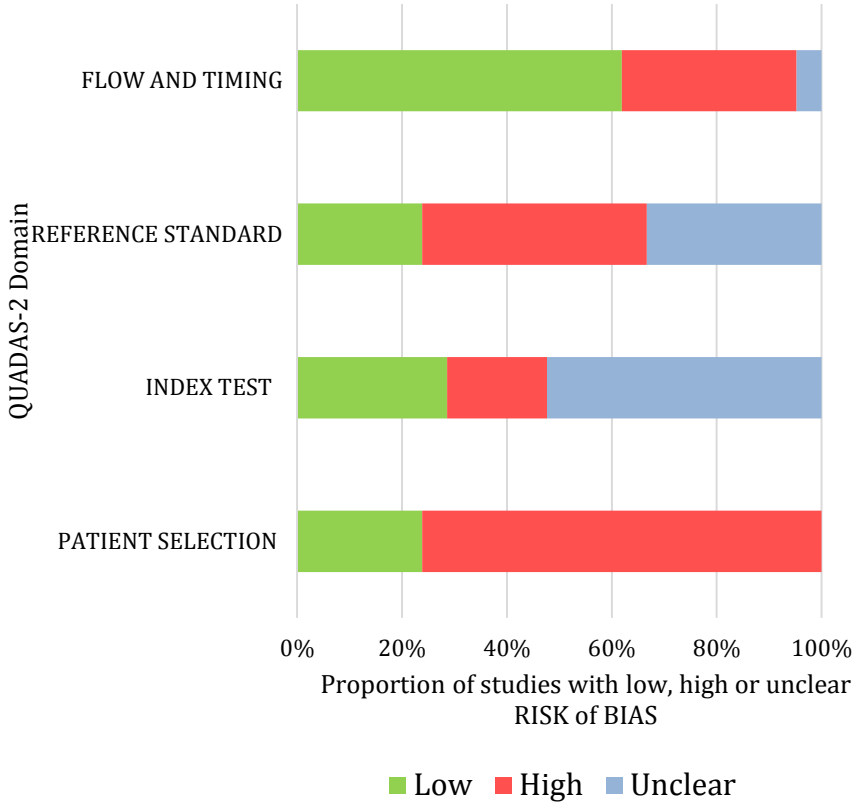
Assessment of methodological quality is summarised according to QUDAS-2 domains in Figure 2. Overall risk of bias was high or unclear for all included studies. The predominant limitation was the use of a case control study design, where sideline testing was performed separately on participants without head impact events and/or those with already diagnosed concussion. Exclusion of 'difficult to diagnose' cases with significant head impact events and possible concussion, the population forming the focus of the review question, could markedly exaggerate diagnostic accuracy metrics. Other systematic errors included delayed index testing, inaccurate reference standard assessment by a non-medically trained outcome assessors, and test and diagnostic review, incorporation and attrition biases.

Conversely there were no applicability concerns across included studies. The review question is broad in scope and although a wide range of settings, sports and age groups were investigated there were no concerns that these were not consistent with the review question. Furthermore, sideline testing was conducted according to standardised instructions, and the target condition defined by the included reference standards was consistent with the review inclusion criteria.

Figure 2. Summary of risk of bias across included studies

Study	Risk of Bias					APPLICABILITY CONCERNS			
	Patient selection	Index test	Reference standard	Flow and timing	Overall	Patient selection	Index test	Reference standard	Overall
Maddocks	☹	☺	☺	☺	☹	☺	☺	☺	☺
Barr 2001	☹	?	☹	☺	☹	☺	☺	☺	☺
McCrory	☹	?	☺	☺	☹	☺	☺	☺	☺
McCrea	☹	?	☹	☺	☹	☺	☺	☺	☺
McCrea	☹	?	☹	☺	☹	☺	☺	☺	☺
Erlanger	☹	?	?	☺	☹	☺	☺	☺	☺
McCrea	☹	☹	☹	☹	☹	☺	☺	☺	☺
Echlin	☹	☹	☹	☹	☹	☺	☺	☺	☺
Galetta K	☺	?	?	☹	☹	☺	☺	☺	☺
Galetta K	☹	☺	☹	☺	☹	☺	☺	☺	☺
Barr 2012	☹	☹	☹	☺	☹	☺	☺	☺	☺
King 2012	☺	?	?	?	☹	☺	☺	☺	☺
Galetta M	☹	?	?	☺	☹	☺	☺	☺	☺
Dhawan	☹	?	?	☺	☹	☺	☺	☺	☺
Fuller	☺	☺	☹	☺	☹	☺	☺	☺	☺
Leong	☺	☺	☺	☹	☹	☺	☺	☺	☺
Galetta K	☹	?	?	☺	☹	☺	☺	☺	☺
Leong	☺	☺	☹	☺	☹	☺	☺	☺	☺
Marinides	☹	?	?	☹	☹	☺	☺	☺	☺
Putukian	☹	☹	☺	☹	☹	☺	☺	☺	☺
Seidman	☹	☺	☺	☹	☹	☺	☺	☺	☺

☺ Low Risk ☹ High Risk ? Unclear Risk



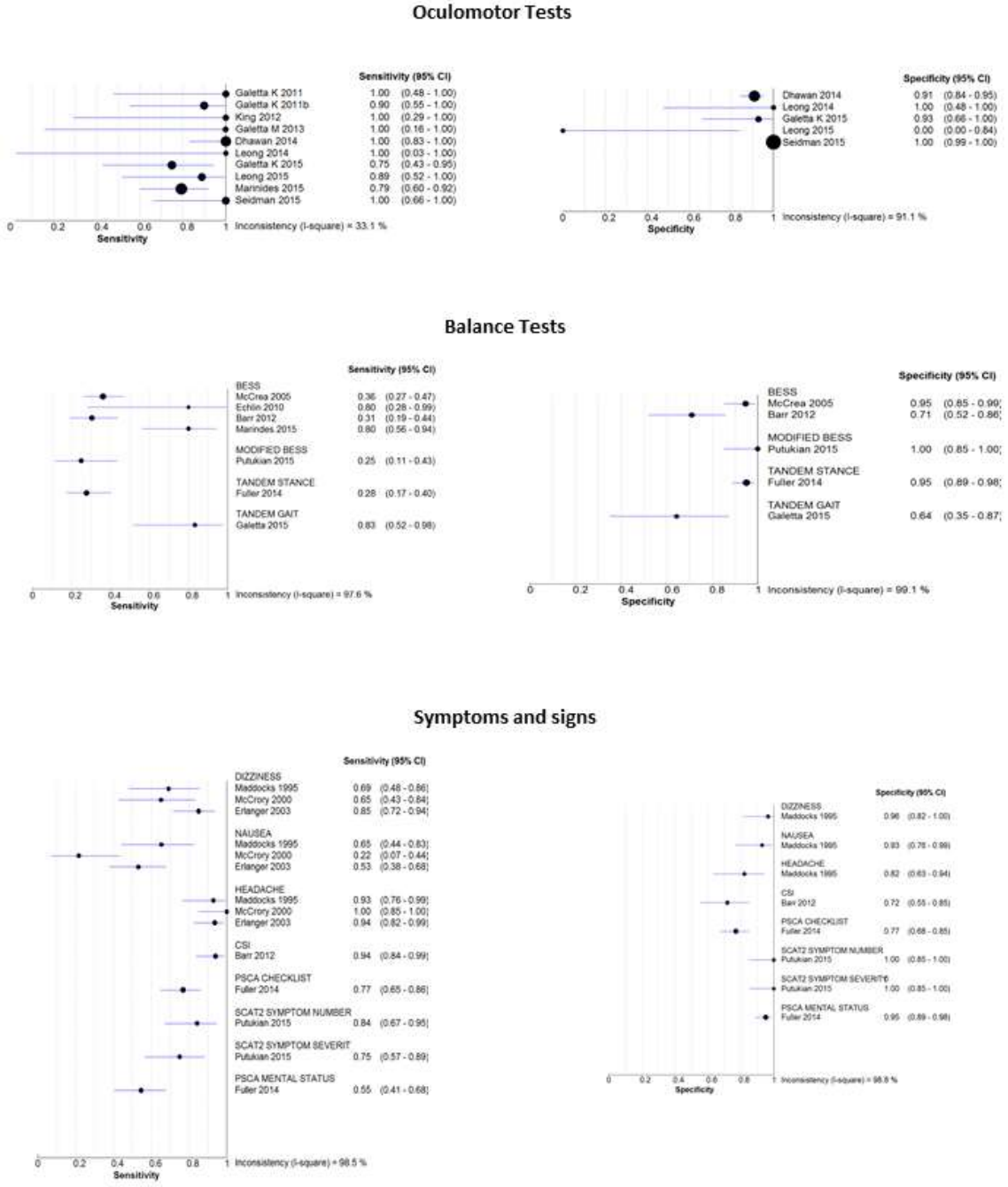
Results

The diagnostic accuracy of sideline assessments for detecting concussion is summarised in Figure 3, with detailed results available in the web appendix. In accordance with the pre-specified analysis plan a meta-analysis was not performed due to the absence of studies at low risk of bias. Reported results were imprecise and heterogenous for all types of sideline assessments. Notwithstanding the concerns regarding internal validity, the KD test, symptoms, and multimodal assessments demonstrated good sensitivity and specificity. Balance and cognitive tests appeared to have poorer sensitivity, but good specificity.

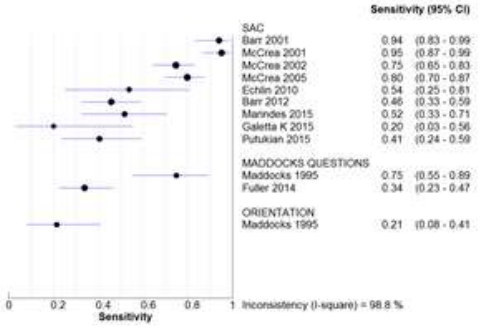
Overall quality of evidence

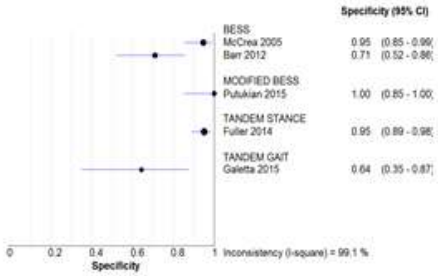
Assessment of the overall quality of evidence examining the performance of sideline tests in concussion screening is summarised according to GRADE criteria in Table 4. The final evidence rating was very low for all classes of sideline tests based on serious concerns regarding inconsistency, imprecision, and risk of bias. A more detailed evaluation of overall quality of evidence for individual tests is provided in the web appendix.

Figure 3. Forrest plots summarising diagnostic accuracy results of sideline screening tests

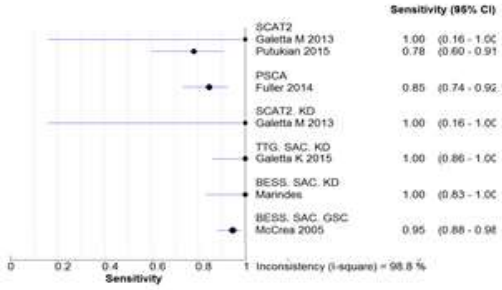


Cognitive Tests





Multi-modal assessments



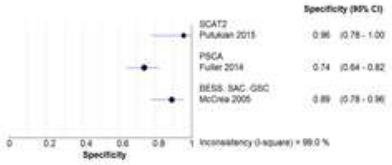


Table 3. GRADE quality of evidence table for sideline screening tests

Outcome	Study designs	Factors decreasing quality of evidence					Overall GRADE rating
		Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	
Balance Tests							
Sensitivity		Serious concerns	No concerns	Serious concerns	Serious concerns	Not detected	**Very Low**
Specificity		Serious concerns	No concerns	Serious concerns	Serious concerns	Not detected	**Very Low**
Oculomotor tests							
Sensitivity	1 PCS	Serious concerns	No concerns	Serious concerns	Serious concerns	Not detected	**Very Low**
Specificity	1 PCS	Serious concerns	No concerns	Serious concerns	Serious concerns	Not detected	**Very Low**
Symptoms and signs							
Sensitivity	1 PCS	Serious concerns	No concerns	Serious concerns	Serious concerns	Not detected	**Very Low**
Specificity	1 PCS	Serious concerns	No concerns	Serious concerns	Serious concerns	Not detected	**Very Low**
Cognitive tests							
Sensitivity	1 PCS	Serious concerns	No concerns	Serious concerns	Serious concerns	Not detected	**Very Low**
Specificity	1 PCS	Serious concerns	No concerns	Serious concerns	Serious concerns	Not detected	**Very Low**
Multimodal assessments							
Sensitivity	1 PCS	Serious concerns	No concerns	Serious concerns	Serious concerns	Not detected	**Very Low**
Specificity	1 PCS	Serious concerns	No concerns	Serious concerns	Serious concerns	Not detected	**Very Low**

Technology

Five studies met review inclusion criteria and reported interpretable data on the use of a technology in sideline screening for sport's related concussion. The technologies examined comprised head impact sensors (Guskiewicz 2007, Mihalak 2007, Greenwald 2008, Broglio 2010) and sideline video review (Fuller 2016). Six potentially eligible studies were also identified, which recorded data on technology use in concussed and non-concussed athletes, but did not report useable data on diagnostic accuracy or effectiveness, including: iPad

software applications for concussion screening (Alberts 2014, McKenzie 2014); Head Impact Telemetry Systems (Duma 2005, Brolinson 2006, Eckner 2011); and a portable computerised neuropsychological assessment tool (Espinoza 2014). Overall risk of bias according to QUDAS-2 domains was Low for Guskiewicz 2007, Greenwald 2008, and Broglio 2010, and unclear for Mihalek 2007. Fuller 2016 described the characteristics of sideline video review used in a single tournament and constituted level 2b evidence.

Full data allowing calculation of diagnostic accuracy metrics were not available for studies examining head impact sensors. Reported results indicated that no clinically significant relationship existed between impact magnitude or location and concussion. Greenwald 2008 reported that the sensitivity of linear acceleration impacts >98.1g for concussion was 61.5%, but no data was presented on the number of non-concussive impacts at this threshold. Greenwald 2008 and Mihalek 2007 demonstrated extremely low positive predictive values of 0.3% to 0.35% at similar impact thresholds. An improved, but still very low, positive predictive value of 13.5% was reported by Broglio 2010 using a statistical model including linear acceleration, rotational forces and impact location. Fuller 2016 reported that sideline video review contributed to identification of 61.5% of significant head impact events and helped sideline evaluation in 20.4% of cases. The overall GRADE quality of evidence was rated as high for head impact sensors and low for sideline video review. Table 4 summarises the characteristics, risk of bias and main results of included technology studies. Further details on risk of bias and GRADE quality ratings are provided in the web appendix.

Table 4. Characteristics, risk of bias and primary findings of included technology studies

Study	Setting	Design	Sample Size (n=)	Sport(s)	Level	Mean age (years± SE)	Technology	Risk of Bias / evidence level	Applicability concerns	Primary finding(s)
Guskiewicz 2007	US	PCS	81	American football	High school	20.2±1.8	HITS	Low	Low	61.5% sensitivity for concussion*
Mihalak 2007	US	PCS	102	American football	Collegiate	19.6±1.6	HITS	Unclear	Low	PPV of 0.35% for concussion†
Greenwald 2008	US	PCS	449	American football	High school	NR	HITS	Low	Low	PPV of 0.3% for concussion*
Broglio 2010	US	PCS	78	American football	High school	16.7±0.8	HITS	Low	Low	PPV of 13.4% for concussion**
Fuller 2016	UK	PCS	49	Rugby Union	Professional	26.5 (SD 3.5)	Sideline video review	Level 2b	Low	Contributed to identification of 61.% of significant head impact events

HITS: Head Impact Telemetry System
*Head impact threshold: linear acceleration >98.9g; **Threshold: >5582.3 rads/s^2 ± 96.1g linear acceleration ± front/side/top impact; † Threshold: linear acceleration >80g

Multi-modal and sports-specific diagnostic models

Comprehensive systems for the detection and assessment of significant head impact events, and subsequent management of suspected concussion, have been introduced in a number of sports. No experimental or comparative effectiveness research was identified evaluating the performance of alternative protocols. However, a single study was retrieved which evaluated a concussion management system used at the elite level in Rugby Union (Fuller 2016, Level 2b evidence). The major finding was the importance of a multimodal, multitime-based concussion evaluation process incorporating video review to identify significant head impact events and delayed onset concussion. Further details on existing concussion management protocols and the characteristics of Fuller 2016 are provided in the web appendix.

The systematic approaches to the on-field diagnosis and management of concussion currently used by professional sports (at the elite level of competition) are summarised in Table 5.

Table 5. Summary of the sideline head injury assessment protocols used in professional contact and collision sports

Sport	Tool / protocol	Person/s who can request test	Person/s conducting the assessment	Use of video review	Other key components
AFL/ NRL	Sport-specific Head Injury Assessment Form	Team doctor	Team doctor	Mandatory	Other club support staff must report observations to the team doctor. Any player requiring further assessment is removed from the field of play for a minimum of 15 minutes. SCAT3 used for further assessment. Head injury assessment forms are collected for audit and injury surveillance purposes.
FIFA	Immediate removal criteria				3-minute injury time following head impact. Pitch-Side assessment performed (based on a number of immediate removal criteria)
IIHF	Concussion protocol		Team doctor and/or athletic trainer/therapist. (However the team doctor is solely responsible for determining whether the player is diagnosed as having a concussion)		Observations made by team medical staff (or by any other team personnel and passed on to team medical staff). Player removed for assessment.
NFL	Side-line concussion assessment tool	Coach, player, teammate, official, team doctor, athletic trainer (ATC), ATC in the media booth or the unaffiliated neurotrauma consultants (UNC)	Team doctor, ATC or UNC	Mandatory	Booth ATC, UNC and the team doctor are connected by radio communication. The Booth ATC is also connected by radio communication with officials and has the ability to stop play and require that a player be evaluated. When a potential head injury is identified, the player is removed immediately from the field. The team doctor will review the video of the incident and (at a minimum) assess the player with a focussed

					neurological assessment (asking what happened, reviewing the "Go/No Go" signs and symptoms; and asking the Maddock's questions. If the diagnosis is unclear, the player will undergo a full NFL sideline Concussion Assessment in the team locker room.
World Rugby	Sport-specific HIA form	Match official, team doctor or independent match day doctor	Certified medical professional	Available	All medical professionals associated with teams or sideline care in professional rugby must successfully complete online education program for certification. Where the diagnosis is not immediately apparent, players removed & assessed. HIA forms are collected for audit.

AFL = Australian Football League; FIFA = Federation Internationale de Football Association; HIA + Head Injury Assessment; IIHF = International Ice Hockey Federation; NFL = National Football League; NRL = National Rugby League.

The criteria for immediate removal from play and no return (i.e. clear diagnosis of concussion) or further assessment (i.e. possible diagnosis of concussion) used by the various sports are summarised in Table 6. .

Table 6 . Summary of criteria for immediate removal from play or for further assessment used in professional sport.

Clinical criteria	AFL/ NRL	FIFA	IIHF	NFL	World Rugby
Confirmed loss of consciousness					
Definite confusion/disorientation					
Any balance disturbance (e.g. ataxia) or motor incoordination					
Impact seizure/convulsions or tonic posturing					
Player reports significant, new or progressive/persistent concussion symptoms					
Clearly dazed, "dinged", blank or vacant stare					
Behavioural change atypical of the player					
Any clinical impression that the player is not quite right following trauma (i.e. "physician's decision")					
Loss of responsiveness/suspected loss of consciousness					
Memory impairment/amnesia					
No protective action when falling to the ground (can be either tonic or hypotonic) – observed on video					
Dangerous mechanism of trauma					
Cross eyes (strabismus) or spontaneous nystagmus					
Possible impact seizure or tonic posturing on video review					
Possible balance disturbance					
Slow to get up following a hit to the head					
Possible behavioural changes					
Possible confusion					
Head impact event with the potential to result in concussion					
Diagnosis not apparent					

AFL = Australian Football League; IIHF = International Ice Hockey Federation; NRL = National Rugby League; NFL = National Football League, FIFA = Federation Internationale de Football Association

= Criteria for immediate removal and no return (i.e. diagnosis of concussion)

= Criteria for further assessment

= Criteria not specified

DISCUSSION

Summary of key findings and *narrative synthesis*

Symptom analysis and cognitve evaluation

Symptoms, both self-reported and as recorded on a Graded Symptom Checklist (GSC) provide valuable insight into the sideline diagnosis of concussion and should form part of any diagnostic rubric (McCrea, 2013). Results from the reviewed studies show strong evidence of elevated symptoms resulting from SRC, with a pattern of gradual resolution of

symptoms over a period of days in most athletes followed in prospective studies with preinjury baseline symptom assessment. Similarly, changes in cognition and other functional abilities are evident after concussion, with moderate to large effect sizes on some measures during the acute period. The pattern of cognitive and clinical recovery follows a similar time course to symptom recovery in prospective studies.

Any symptoms associated with a sports-associated collision should be interpreted as suspicious of concussion and the player removed from play. Confounding aspects of sideline symptom analysis include the spectrum of concussion-associated symptoms not being specific and the potential for delayed onset of symptoms. The absence of sideline symptoms should therefore not be interpreted as absolute proof of the absence of concussion. Symptom assessment and serial monitoring are useful markers of concussion and, together with evaluation of cognitive function and balance make up a useful composite asessment of brain function (Putkian, 2015). In the presence of the suspicion of concussion, further symptom analyses should be implemented and other parameters of brain function assessed. Conversely, formal, follow-up medical evaluation is required to confirm that any symptoms detected on the sideline are *not* concussion related.

Balance testing

Studies meeting inclusion criteria for this systematic review assessed two sideline postural stability protocols, the BESS (McCrea 2005 and Echlin 2010) and the Tandem Gait Test (Fuller 2014). Individual sensitivity estimates for the BESS were heterogenous and imprecise, with point estimates ranging from 34.0 to 80.0%, I^2 73.8%. BESS specificity, reported in a single study, was high 94.6% (95% CI 85.1 – 98.9). The Tandem Gait Test demonstrated poor sensitivity and good specificity (95.0%, 95% CI 88.7 – 98.4) in the single study available.

Balance testing enhances sensitivity when combined immediately post-injury with symptom assessment and cognitive testing (McCrea 2005, Echlin 2010)

Clinical research on sideline balance tests is challenging, and it is vital that researchers continue to explore the limits and potential of these tools. Biases present in the studies in this review underscore the importance of adequately powered enrollment, adequate matching of controls, independent blinded assessors, and other principles of rigorous research. Studies not included in this systematic review due to methodological limitations collectively suggest, but do not prove, that the circumstances and settings (barefoot vs cleats-on, the type of underfoot surface, quiet controlled environment vs live-sporting

scenarios) of BESS testing can affect test results (Azad 2014, Onate 2007 and Smith 2012). Larger and more rigorous studies will need to confirm these findings before they can be accepted but this area of research is significant if balance testing in any guise is to be incorporated into multi-modal sideline evaluations. Intuitively, footwear and environmental factors should be replicated for baseline and post-injury postural stability evaluations. Future studies following established and well-described methods for conducting high-quality research could help to answer key questions raised by these studies.

Oculomotor

Eleven studies met review inclusion criteria and reported interpretable data on the diagnostic accuracy of oculomotor screening test for the sideline identification of sports concussion. All eleven studies investigated the King-Devick (KD) Test; an objective clinical test of rapid eye movements, where worsening of test completion time from a baseline and/or errors committed is considered a positive finding indicating concussive injury (Figure 3).

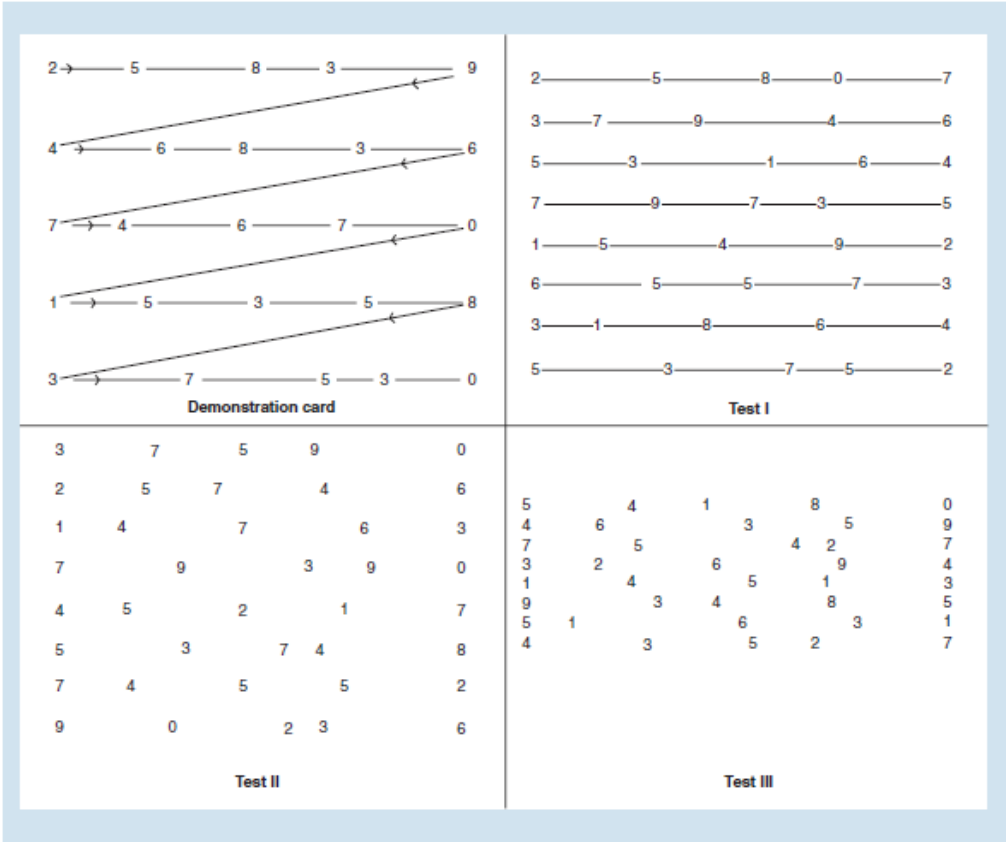


Figure 3 . King-Devick test cards.
Players begin at the top left number of Test card I and read from left to right across each row. The same procedure is repeated for Test II and III. The total time and number of errors are compared to a pre-season baseline assessment.

The investigations included evaluation of the KD test in a wide range of sports (American football, soccer, ice hockey, lacrosse, athletics, boxing, mixed martial arts, basketball, rugby union, rugby league); settings (New Zealand, United States); performance levels (high school, collegiate, amateur, professional); and age groups (children, adolescents, and young adults). The majority of studies were prospective cohort studies, with a single eligible retrospective cohort study identified (Marindes 2015). Overall source sample sizes were modest, ranging from n=27 to 337; however the number of participants included in diagnostic accuracy assessments were very low (median 11, interquartile range 7 – 33.5).

Data allowing calculation of sensitivity of the KD test for identifying concussion was measured in all included studies and varied widely from 71.4% to 100.0% . Individual estimates were very imprecise secondary to small sample sizes, with lower 95% confidence limits as low as 2.5% calculated . This diversity was reflected in a high I^2 statistic (52.1%). Data for specificity estimates was measured in six studies with similarly imprecise and heterogeneous results calculated, ranging from 0.0% to 100.0% (I^2 statistic 89.3%).

The KD test is a promising sideline screening test for sports related concussion that is simple, quick, acceptable to athletes, and has favourable reproducibility. However, there is an absence of valid research confirming its diagnostic accuracy and impact in improving outcomes. An adequately powered diagnostic accuracy study is therefore recommended, which avoids a case-control design and enrols a representative sample of athletes with suspected concussion.

Insufficient evidence currently exists to recommend the KD test as a side-line screening test for concussion. The overall body of evidence is rated as very low quality secondary to a high risk of bias, lack of consistency of results, and too much uncertainty in published studies.

A theme in SRC literature is that a single effective diagnostic entity does not exist but rather that the diagnosis of concussion is most appropriately made using a range of serially applied clinical and technological tools, some of which have been covered in this paper. Some sporting codes have evolved models utlising different sideline and post-incident interventions to more accurately make the diagnosis of concussion.

In order to deal with these challenges, sports have continued to evolve their approach to the sideline diagnosis and management of concussion. Professional sports have added components such as video analysis and concussion "spotters" to help identify head impact events and/or subtle clinical signs that may otherwise be missed from the side-lines; as well as independent practitioners and concussion experts to assist in the assessment of a potential concussion. Furthermore, sports have developed clear definitions and criteria for immediate removal from play and no return (i.e. clinical features consistent with a diagnosis of concussion) or for cases that require further clinical assessment (i.e. possible diagnosis of concussion). This combined approach facilitates identification of potential concussive incidents and improves the consistency of the side-line evaluation and management.

The operational definition of concussion is based largely on observable clinical signs, although most sports have added reporting of "significant", "new" or "progressive/persistent" post-concussion symptoms to the diagnostic criteria for immediate removal from play. In addition, sports have acknowledged that sometimes the player does not fit the specific criteria or may pass the assessment, but is still "not right". In these cases, the attending physician can still make a diagnosis of concussion based on their own clinical impressions ("physician's decision").

Any operational definition of concussion raises the question about a minimum threshold for diagnosis. Currently, any post-concussion symptom or sign is considered to fit the diagnosis of concussion (McCrory et al. BJSM 2013). Some signs, such as balance disturbance, have other potential causes (such as vestibular injury rather than brain injury per se).

Currently, there is little scientific data on the effectiveness of sideline head injury assessment protocols used in professional sports. This does not imply that they are irrelevant. On the contrary, multimodal diagnostic models are likely to be more effective screening and triage tools for head injury increasing the likelihood that concussed players

will be recognised, removed from the field of play and monitored. Further validation of the sideline protocols is recommended. Collaboration between sporting codes to rationlise multimodal diagnostic sideline protocols may help facilitate more efficient application and monitoring.

Review limitations

To maximise internal validity Cochrane Collaboration and PRISMA guidelines were followed aiming to ensure that all relevant evidence was included, accurately and precisely coded, validly assessed for risk of bias, and impartially analysed and interpreted. However, there are a number of potential methodological weaknesses which could limit the validity of this systematic review.

Due to time constraints hand searching of journals and conference proceedings was not performed, and regional bibliographic databases were not included. Furthermore, unclear reporting of non-randomised studies and poor indexing in databases may impair the detection of published information. However, given that several thousand research records were examined searching additional information sources, or utilising more sensitive search strategies, was impractical and unlikely to yield further relevant evidence. It can also be argued that harder to find observational evidence is likely to have lower internal validity. Extraction of data from primary analyses to allow consistency with the review question, and inclusion of unpublished data, should have provided more valid and applicable results; but the absence of peer-review may undermine their credibility.

Decisions on study relevance, information gathering, and validity were un-blinded and could potentially been influenced by pre-formed opinions. Masking by editing out information on journal, authors, institutions, and direction of results, is resource intensive and given the uncertain benefits was not performed. However, objective rating criteria and independent data collection should militate against the risk of reviewer bias.

Finally, assessment of reference standard bias was challenged by: the diverse clinical presentation of concussion, variable natural history and the lack of a convincing gold standard. It is therefore acknowledged that the accuracy of the reference standard to correctly classify concussion is uncertain in all included studies, and the risk of bias in the

reference standard domain is therefore arguably unclear or high. However, although a standardised multi-modal, multi-time point assessment has been recommended; pragmatically any clinical assessment by a qualified health professional proximate to the presentation of suspected concussion was considered to correctly classify concussion for the purposes of the current review.

CONCLUSIONS

Implications for clinical practice

Sideline diagnosis remains the first and perhaps most significant step in the process of recognising, managing and approriately advising the concussed athlete. This systematic review evaluated the most critical elements in the sideline dianosis of concussion most notably clinical screening and diagnostic tools, existing and emerging technology and multi-modal protocols.

- Insufficient evidence currently exists to recommend a particular sideline screening test or integrated management protocol for sports-related concussion.
- The overall body of evidence is rated as very low quality secondary to a high risk of bias, lack of consistency of results, and too much uncertainty in published studies.
- The limited available data suggest that joint use of individual sideline tests, or multifaceted multimodal assessment tools, are likely to be required to optimise sensitivity and specificity.
- Sideline video review may offer a promising approach to improve identification and evaluation of significant head impact events.
- Current evidence does not support the use of impact sensor systems for real-time concussion identification or screening.
- Multimodal sideline protocols incorporating symptom analysis, observation of key signs, balance evaluation, verbal cognitive screening and video review have the potential to more effectively identify potential cases of concussion; rationalising these models across sporting codes may help standardise levels of care. Moreover, a universally applied reference standard which includes multi-modal and multi-time point assessments, and blinded index test, would further increase research validity.

Implications for research

There is an absence of valid research confirming the diagnostic accuracy and impact on improving outcomes of currently used sideline screening tests and integrated management protocols. Adequately powered diagnostic accuracy studies are therefore recommended that enrol a representative sample of athletes with suspected concussion following non-trivial head impact events. A universally applied standardised reference which includes multi-modal and multi-time point assessments as well as blinded index test and reference standard assessment, would further increase internal validity. Ideally, once the diagnostic accuracy and optimal threshold of sideline tests or integrated management protocols have been validated, controlled trials would randomise athletes to competing sideline screening strategies and measure diagnostic accuracy, morbidity and acceptability to investigate whether important outcomes are improved. Further research is also recommended to: evaluate the comparative effectiveness of alternative integrated management protocols; investigate the impact of sideline video review on the identification and evaluation of head impact events; and examine the utility of tablet software applications as an adjunct to sideline concussion screening. Another key challenge for all sports is how the information is translated to lower levels of competition where resources and experience of medical staff may be limited. Given the reduced levels of medical support, and lower profile of these sports at amateur level, an operational definition is even more important to facilitate improved recognition and management of potential concussion. The key to achieving this objective involves developing a simple, cost-effective protocol, combined with an education program to help up-skill medical teams, as well as educate players, coaches, referees etc. about the signs and symptoms of concussion.

REFERENCES

Cochrane Handbook for Systematic Reviews of Interventions: The Cochrane Collaboration, 2008.

Azad AM, Juma SA, 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 doi:10.1136/bmjsem-2016-000117

Barr WB, McCrea M. Sensitivity and specificity of standardized neurocognitive testing immediately following sports concussion. Journal of the International Neuropsychological Society : JINS 2001;7(6):693-702

Barr WB, Prichep LS, Chabot R, et al. Measuring brain electrical activity to track recovery from sport-related concussion. Brain injury : [BI] 2012;26(1):58-66 doi: 10.3109/02699052.2011.608216[published Online First: Epub Date]|.

Broglio SP, Macciocchi SN, Ferrara MS. Sensitivity of the concussion assessment battery. Neurosurgery 2007;60(6):1050-7; discussion 57-8 doi: 10.1227/01.NEU.0000255479.90999.C0[published Online First: Epub Date]|.

Broglio, S. P., B. Schnebel, J. J. Sosnoff, et al. Biomechanical properties of concussions in high school football. Med. Sci. Sports Exerc. 42:2064–2071, 2010

Brolinson P.G., Manoogian S., McNeely D., Goforth M., Greenwald R., Duma S.. Analysis of linear head accelerations from collegiate football impacts. Curr Sports Med Rep [Internet]. Feb 2006 5(1):23-28

Covassin T, Stearne D, Elbin R. Concussion history and postconcussion neurocognitive performance and symptoms in collegiate athletes. Journal of athletic training 2008;43(2):119-24 doi: 10.4085/1062-6050-43.2.119[published Online First: Epub Date]|.

Daniel JC, Nassiri JD, Wilckens J, et al. The implementation and use of the standardized assessment of concussion at the U.S. Naval Academy. Military medicine 2002;167(10):873-6

Dhawan P SA, Tapsell L, Adler J, Galetta S, Balcer L, Dodick D. King-Devick Test Identifies Symptomatic Concussion in Real-time and Asymptomatic Concussion Over Time. Neurology Clinical Practice 2014;82(10):S11

Duma S.M., Manoogian S.J., Bussone W.R., Brolinson P.G., Goforth M.W., Donnenwerth J.J., Greenwald R.M., Chu J.J., Crisco J.J.. Analysis of real-time head accelerations in collegiate football players. Clin. J. Sport Med. [Internet]. January 2005 15(1):3-8.

Echlin PS, Tator CH, Cusimano MD, et al. Return to play after an initial or recurrent concussion in a prospective study of physician-observed junior ice hockey concussions: implications for return to play after a concussion. Neurosurgical focus 2010;29(5):E5 doi: 10.3171/2010.9.focus10210[published Online First: Epub Date]|.

Erlanger D, Kaushik T, Cantu R, et al. Symptom-based assessment of the severity of a concussion. J Neurosurg 2003;98(3):477-84 doi: 10.3171/jns.2003.98.3.0477[published Online First: Epub Date]|.

Espinoza T., Phelps S.E., Wright D.W., Bazarian J., Knezevic A., Gore R., Ciaravella N.M., Crooks C., Liu B., Smith S., LaPlaca M.C.. Display enhanced testing of cognitive impairment and mild traumatic brain injury (DETECT): A novel tool for concussion assessment. J. Neurotrauma [Internet].

Figler R., Hirsch J., Kana D., Linder S., Alberts J.. Field utilization of the cleveland clinic concussion ipad assessment tool. Clin. J. Sport Med. [Internet]. March 2013 23(2):128.

Fuller GW, Kemp SP, Decq P. The International Rugby Board (IRB) Pitch Side Concussion Assessment trial: a pilot test accuracy study. British journal of sports medicine 2015;49(8):529-35 doi: 10.1136/bjsports-2014-093498[published Online First: Epub Date]|.

Fuller CW, Fuller GW, Kemp SP, Raftery M. Evaluation of World Rugby's concussion management process: results from Rugby World Cup 2015. Br J Sports Med. 2016 Sep 1. pii: bjsports-2016-096461. doi: 10.1136/bjsports-2016-096461.

Galetta KM, Barrett J, Allen M, et al. The King-Devick test as a determinant of head trauma and concussion in boxers and MMA fighters. Neurology 2011;76(17):1456-62 doi: 10.1212/WNL.0b013e31821184c9[published Online First: Epub Date]|.

Galetta KM, Brandes LE, Maki K, et al. The King-Devick test and sports-related concussion: study of a rapid visual screening tool in a collegiate cohort. Journal of the neurological sciences 2011;309(1-2):34-9 doi: 10.1016/j.jns.2011.07.039[published Online First: Epub Date]|.

Galetta KM, Morganroth J, Moehringer N, et al. Adding Vision to Concussion Testing: A Prospective Study of Sideline Testing in Youth and Collegiate Athletes. Journal of neuro-ophthalmology : the official journal of the North American Neuro-Ophthalmology Society 2015;35(3):235-41 doi: 10.1097/wno.0000000000000226[published Online First: Epub Date]|.

Galetta MS, Galetta KM, McCrossin J, et al. Saccades and memory: baseline associations of the King-Devick and SCAT2 SAC tests in professional ice hockey players. Journal of the neurological sciences 2013;328(1-2):28-31 doi: 10.1016/j.jns.2013.02.008[published Online First: Epub Date]|.

Gavett BE, Stern RA, McKee AC. Chronic traumatic encephalopathy: a potential late effect of sport-related concussive and subconcussive head trauma. Clinics in sports medicine 2011;30(1):179-88, xi doi: 10.1016/j.csm.2010.09.007[published Online First: Epub Date]|.

Greenwald RM, Gwin JT, Chu JJ, Crisco JJ. Head impact severity measures for evaluating mild traumatic brain injury risk exposure. Neurosurgery [Internet]. 2008 [cited 2008 Apr];62(4):789-98; discussion 798.

Guskiewicz KM, Marshall SW, Bailes J, et al. Association between recurrent concussion and late-life cognitive impairment in retired professional football players. Neurosurgery 2005;57(4):719-26; discussion 19-26

Guskiewicz KM, McCrea M, Marshall SW, et al. Cumulative effects associated with recurrent concussion in collegiate football players: the NCAA Concussion Study. JAMA : the journal of the American Medical Association 2003;290(19):2549-55 doi: 10.1001/jama.290.19.2549[published Online First: Epub Date]|.

Guskiewicz K.M., Mihalik J.P., Shankar V., Marshall S.W., Crowell D.H., Oliaro S.M., Ciocca M.F., Hooker D.N.. Measurement of head impacts in collegiate football players: Relationship between head impact biomechanics and acute clinical outcome after concussion. Neurosurgery [Internet]. December 2007 61(6):1244-1252

Herman DC, Jones D, Harrison A, et al. Concussion May Increase the Risk of Subsequent Lower Extremity Musculoskeletal Injury in Collegiate Athletes. Sports medicine (Auckland, N.Z.) 2016 doi: 10.1007/s40279-016-0607-9[published Online First: Epub Date]|.

Iverson GL, Gaetz M, Lovell MR, et al. Cumulative effects of concussion in amateur athletes. Brain injury : [BI] 2004;18(5):433-43 doi: 10.1080/02699050310001617352[published Online First: Epub Date]|.

King D, Clark T, Gissane C. Use of a rapid visual screening tool for the assessment of concussion in amateur rugby league: a pilot study. Journal of the neurological sciences 2012;320(1-2):16-21 doi: 10.1016/j.jns.2012.05.049[published Online First: Epub Date]|.

Leong DF, Balcer LJ, Galetta SL, et al. The King-Devick test for sideline concussion screening in collegiate football. Journal of optometry 2015;8(2):131-9 doi: 10.1016/j.optom.2014.12.005[published Online First: Epub Date]|.

Leong DF, Balcer LJ, Galetta SL, et al. The King-Devick test as a concussion screening tool administered by sports parents. The Journal of sports medicine and physical fitness 2014;54(1):70-7

Maddocks DL, Dicker GD, Saling MM. The assessment of orientation following concussion in athletes. Clinical journal of sport medicine : official journal of the Canadian Academy of Sport Medicine 1995;5(1):32-5

Marinides Z, Galetta K, Andrews C, Wilson J, Herman D, Robinson C, Smith M, Bentley B, MD, Galetta D, Balcer L, Clugston J. Vision testing is additive to the sideline assessment of sports-related concussion. Neurology Clinical Practice 2015;5(1):24-34

Matser EJ, Kessels AG, Lezak MD, et al. Neuropsychological impairment in amateur soccer players. JAMA : the journal of the American Medical Association 1999;282(10):971-3

McCrea HJ, Perrine K, Niogi S, et al. Concussion in sports. Sports health 2013;5(2):160-4 doi: 10.1177/1941738112462203[published Online First: Epub Date]|.

McCrea M. Standardized Mental Status Testing on the Sideline After Sport-Related Concussion. Journal of athletic training 2001;36(3):274-79

McCrea M, Barr WB, Guskiewicz K, et al. Standard regression-based methods for measuring recovery after sport-related concussion. Journal of the International Neuropsychological Society : JINS 2005;11(1):58-69 doi: 10.1017/s1355617705050083[published Online First: Epub Date]|.

McCrea M, Guskiewicz K, Randolph C, et al. Incidence, clinical course, and predictors of prolonged recovery time following sport-related concussion in high school and college athletes. Journal of the International Neuropsychological Society : JINS 2013;19(1):22-33 doi: 10.1017/s1355617712000872[published Online First: Epub Date]|.

McCrea M, Guskiewicz KM, Marshall SW, et al. Acute effects and recovery time following concussion in collegiate football players: the NCAA Concussion Study. JAMA : the journal of the American Medical Association 2003;290(19):2556-63 doi: 10.1001/jama.290.19.2556[published Online First: Epub Date]|.

McCrea M, Kelly JP, Kluge J, et al. Standardized assessment of concussion in football players. Neurology 1997;48(3):586-8

McCrea M, Kelly JP, Randolph C, et al. Immediate neurocognitive effects of concussion. Neurosurgery 2002;50(5):1032-40; discussion 40-2

McCrea M, Kelly JP, Randolph C, et al. Standardized assessment of concussion (SAC): on-site mental status evaluation of the athlete. The Journal of head trauma rehabilitation 1998;13(2):27-35

McCrea M, Prichep L, Powell MR, et al. Acute effects and recovery after sport-related concussion: a neurocognitive and quantitative brain electrical activity study. The Journal of head trauma rehabilitation 2010;25(4):283-92 doi: 10.1097/HTR.0b013e3181e67923[published Online First: Epub Date]|.

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. British journal of sports medicine 2013;47(5):250-8 doi: 10.1136/bjsports-2013-092313[published Online First: Epub Date]|.

McCrory PR, Ariens T, Berkovic SF. The nature and duration of acute concussive symptoms in Australian football. Clinical journal of sport medicine : official journal of the Canadian Academy of Sport Medicine 2000;10(4):235-8

McKee AC, Cantu RC, Nowinski CJ, et al. Chronic traumatic encephalopathy in athletes: progressive tauopathy after repetitive head injury. Journal of neuropathology and experimental neurology 2009;68(7):709-35 doi: 10.1097/NEN.0b013e3181a9d503[published Online First: Epub Date]|.

McKenzie L. Evaluation of Spot Light: A Concussion Injury Management App for Youth Sports (SpotLight). ClinicalTrials.gov Identifier: NCT02249533

Mihalik J.P., Bell D.R., Marshall S.W., Guskiewicz K.M.. Measurement of head impacts in collegiate football players: An investigation of positional and event-type differences. Neurosurgery [Internet]. December 2007 61(6):1229-1235.

Nassiri JD DJ, Wilckens J, Land BC. . The implementation and use of the standardized assessment of concussion at the U.S. Naval Academy. . Military medicine 2002;167:873–76

Onate JA, Beck BC and Van Lunen BL. On-Field Testing Environment and Balance Error Scoring System Performance During Preseason Screening of Healthy Collegiate Baseball Players. J Athl Train. 2007 Oct-Dec; 42(4): 446–451.

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. Clinical journal of sport medicine : official journal of the Canadian Academy of Sport Medicine 2015;25(1):36-42 doi: 10.1097/jsm.0000000000000102[published Online First: Epub Date]|.

Rahn C, Munkasy BA, Joyner AB et al. Sideline Performance of the Balance Error Scoring System during a Live Sporting Event. Clin J Sport Med. 2015; 25(3): 248-253

Smith J.A.. Quantitative and objective assessment of the balance errors scoring system using accelerometer and gyroscope in the field following sport-related concussion. Clin. J. Sport Med. [Internet]. November 2012 22(6):519.

Seidman DH, Burlingame J, Yousif LR, et al. Evaluation of the King-Devick test as a concussion screening tool in high school football players. Journal of the neurological sciences 2015;356(1-2):97-101 doi: 10.1016/j.jns.2015.06.021[published Online First: Epub Date]|.

Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Annals of internal medicine 2011;155(8):529-36 doi: 10.7326/0003-4819-155-8-201110180-00009[published Online First: Epub Date]|.

Supplementary file

What are the critical elements of side-line screening that can be used to establish the diagnosis of concussion? A systematic review.

Web Appendix

CONTENTS

FURTHER METHODOLOGICAL DETAILS

Search strategy for identification of studies

Electronic information sources

1. Cochrane Database of Systematic Reviews (via Cochrane library)
2. Cochrane Injuries Group Specialised Register (via Cochrane library)
3. Database of Abstracts of Reviews of Effectiveness (via Cochrane library)
4. Cochrane Central Register of Controlled Trials (via Cochrane library)
5. metaRegister of Controlled Trials (mRCT)
6. ClinicalTrials.gov
7. MEDLINE (via OVID and PubMed platforms)
8. EMBASE (via OVID platform)
9. CINAHL (via OVID platform)
10. SPORTSDiscus (via EBSCO)
11. Science Citation Index (SCI, via Web of Science)
12. SCOPUS
13. ZETOC
14. Conference Proceedings Citation Index – Science (via Web of Science)
15. OpenGrey
16. New York Academy of Medicine Grey Literature Report
17. EThOS: UK E-Theses Online Service
18. ProQuest Dissertation & Theses Database
19. National Clinical Guidelines Clearing House website
20. World wide web

Non-electronic information sources

1. Checking reference lists of retrieved articles
2. Checking reference lists of existing literature and systematic reviews
3. Correspondence with experts in the field, and relevant study authors

Search terms

Database: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) <1946 to Present>

Search Strategy:

1 Athletic Injuries/
2 Sports Medicine/
3 exp Sports/
4 (athlete* or athletic* or sport* or player* or tennis or baseball or football* or basketball or boxing or boxer or gymnast* or hockey or soccer or volleyball or netball or wrestler or wrestling).mp.

5 1 or 2 or 3 or 4
6 Craniocerebral Trauma/
7 Brain Concussion/
8 Head Injuries, Closed/
9 Brain Injuries/
10 (blow adj3 head).mp.
11 ((head or brain) adj2 (trauma* or impact or injur*)).mp.
12 ((brain or cortical) adj2 contusion*).mp.
13 ((nonpenetrating or non-penetrating or blunt) adj3 (brain or head)).mp
14 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13
15 Brain Concussion/
16 (commotio cerebri or concuss*).mp.
17 Ataxia/ (6958)
18 (coordination adj3 (impair* or lack*)).mp.
19 (ataxia* or confusion or confused or dizziness or dizzy).mp.
20 Unconsciousness/
21 (loss ajd2 consciousness or unconscious*).mp.
22 headache.mp.
23 neurological dysfunction.mp.
24 (change* adj3 (behav* or attention or memory)).mp.
25 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24
26 (sideline* or side-line or side line or touch line or touch-line or touchline or pitch or pitch side or pitchside or pitch-side or court or courtside or court-side or court side or dug out or dugout or dug-out or bench or track or technical area or technical-area or ring or ringside or ring-side or ring side).mp.
27 (field or onfield or on-field or on field or in game or ingame or in-game or in match or inmatch or in-match or in play or inplay or in-play).mp.
28 26 or 27
29 (screen or screening or diagnos* or assess* or test*).mp.
30 Triage/
31 Early diagnosis/
32 Return to Sport/
33 Neuropsychological tests/
34 Vision tests/
35 Vestibular function tests/
36 ((return* or resume* or resumption) adj3 play).mp.
37 ((observable or visual) adj3 (sign or signs)).mp.
38 ((saccad* or psychometric or king-devick or KD or K-D or sensory organi#ation or immediate post-concussion or cognitive) adj2 test*).mp.
39 post-concussion symptom scale.mp.
40 (balance error scoring system or BESS).mp.
41 (standardi#ed assessment of concussion or SAC).mp.
42 (((sideline or side-line) adj2 concussion assessment tool) or SCAT2 or SCAT3 or SCAT-2 or SCAT-3).mp.
43 sport* concussion assessment tool or SAC.mp.
44 maddocks.mp.
45 **Add terms for any other sideline screening tests here**
46 29-45/or
47 5 and 14 and 25 and 28 and 46
48 Accelerometry/

49 (accelerometer* or video analysis or video-analysis or video review or video-review or impact sensor* or eye-trac advance or mobile app*).mp.
50 48 or 49
51 5 and 14 and 25 and 28 and 50

Development of search strategies

The search strategies were developed by the research team together with an information services expert from University College London based on expert subject knowledge and existing published search strategies. The search strategy was then further peer reviewed by librarians at the University of Sheffield. Searches were run research team members in conjunction with librarians from the University of Pretoria and University College London.

Study identification and data extraction

Although not eligible for inclusion, identified review articles were examined to provide a strategic overview and cross-check references. Where necessary study authors were contacted to provide additional information. Where appropriate, data were extracted to allow analysis consistent with the review questions and a standard diagnostic accuracy study design, rather than the investigators primary results. A single unblinded reviewer extracted information on study characteristics, methodology and results using a standardised data extraction form; and a second reviewer independently checked data for consistency and accuracy.

Summary of QUADAS-2 Risk of Bias Judgement criteria

Table 1. **Risk of Bias and Applicability Judgments in QUADAS-2**

Domain	Patient Selection	Index Test	Reference Standard	Flow and Timing
Description	Describe methods of patient selection Describe included patients (previous testing, presentation, intended use of index test, and setting)	Describe the index test and how it was conducted and interpreted	Describe the reference standard and how it was conducted and interpreted	Describe any patients who did not receive the index tests or reference standard or who were excluded from the 2 × 2 table (refer to flow diagram) Describe the interval and any interventions between index tests and the reference standard
Signaling questions (yes, no, or unclear)	Was a consecutive or random sample of patients enrolled? Was a case–control design avoided? Did the study avoid inappropriate exclusions?	Were the index test results interpreted without knowledge of the results of the reference standard? If a threshold was used, was it prespecified?	Is the reference standard likely to correctly classify the target condition? Were the reference standard results interpreted without knowledge of the results of the index test?	Was there an appropriate interval between index tests and reference standard? Did all patients receive a reference standard? Did all patients receive the same reference standard? Were all patients included in the analysis?
Risk of bias (high, low, or unclear)	Could the selection of patients have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could the reference standard, its conduct, or its interpretation have introduced bias?	Could the patient flow have introduced bias?
Concerns about applicability (high, low, or unclear)	Are there concerns that the included patients do not match the review question?	Are there concerns that the index test, its conduct, or its interpretation differ from the review question?	Are there concerns that the target condition as defined by the reference standard does not match the review question?	

Assessment of overall quality of evidence

The overall quality of evidence for each outcome was assessed using the consensus Grades of Recommendation, Assessment, Development and Evaluation Working Group (GRADE) approach. This specifies four outcome-specific levels of quality (high, moderate, low, and very low). For comparative effectiveness studies RCTs initially are initially rated as high quality, and observational studies as low quality evidence; for diagnostic accuracy studies cohort studies begin as high quality. The body of evidence is downgraded in the presence of within-study risk of bias, indirectness of evidence, heterogeneity, imprecision of effect/diagnostic accuracy estimates, and risk of publication bias; or up-graded due to large effect sizes, dose-response gradients, or plausible biases all working to undermine effect/accuracy estimates.

Protocol changes

There was a single protocol modification. The Newcastle-Ottawa risk of bias tool was used instead of a hierarchical level of evidence for non-diagnostic cohort studies in response to peer review.

RESULTS

Near miss articles

Seven potentially eligible sideline studies were identified which recorded data on sideline tests and concussion, but did not report useable data on diagnostic accuracy (McCrory 2000 – Digital Subtraction Test and symptoms; Daniel 2002 – SAC; Nassiri 2002 –SAC; McCrea 1997 – SAC; McCrea 1998 – SAC; McCrea 2010 – Concussion Severity Inventory, BESS; Barr 2012 – Concussion Severity Inventory, BESS; McCrea 2013 – GSC, SAC). Six potentially eligible technology studies were also identified, which recorded data on technology use in concussed and non-concussed athletes, but did not report useable data on diagnostic accuracy or effectiveness, including: iPad software applications for concussion screening (Alberts 2014, McKenzie 2014); Head Impact Telemetry Systems (Duma 2005, Brolinson 2006, Eckner 2011); and a portable computerised neuropsychological assessment tool (Espinoza 2014).

Diagnostic thresholds used in included sideline screening test studies

Study	Index tests	Test Threshold
Maddocks 1995	•Symptoms •Orientation, recent memory	•Present / not present •Correct / incorrect
McCrory 2000	Symptoms	Present / not present
Barr 2001	SAC	Any worsening from baseline
Erlanger 2003	Symptoms	Present / not present
McCrea 2001	SAC	Any worsening from baseline
McCrea 2002	SAC	$<10^{th}$ percentile of normal performance
McCrea 2005	GSC, SAC, BESS	Standardized regression based indices for detection of significant change in test scores
Echlin 2010	SAC, BESS	Any worsening from baseline
Galetta K 2011	KD	Any worsening from baseline
Galetta K 2011b	KD	Any worsening from baseline
Barr 2012	CSI, SAC, BESS	Any worsening from baseline
King 2012	KD	>3 seconds prolongation from baseline

Galetta M 2013	SCAT2, KD	Any worsening from baseline
Dhawan 2014	KD	Any worsening from baseline
Fuller 2014	•Symptom Checklist •Mental status evaluation •PSCA •Tandem Stance Test	•Any present •Any abnormality •Any abnormality •>4 errors in 20 seconds
Leong 2014	KD	Any worsening from baseline
Galetta K 2015	•SAC •Timed Tandem Gait, KD	•≥2 point drop in SAC compared to baseline •Any worsening from baseline
Leong 2015	KD	Any worsening from baseline
Marinides 2015	•SAC •KD •BESS	•≥2 point drop in SAC from baseline •Any worsening from baseline •≥3 point worsening form baseline
Putukian 2015	SCAT2 symptom checklist, SAC, SCAT 2, Modified BESS	<5th centile of normative performative.
Seidman 2015	KD	Any worsening from baseline

Detailed results for included sideline screening tests

Symptoms

Study	Index test	TP	FN	FP	TN	Sensitivity (%)	LCL	UCL	Specificity (%)	LCL	UCL
Maddocks 1995**	Dizziness	18	8	1	27	**69.2**	48.2	85.7	**96.4**	81.7	99.9
	Nausea	17	9	2	26	**65.4**	44.3	82.8	**92.9**	76.5	99.1
	Headache	26	2	5	23	**92.9**	76.5	99.1	**82.1**	63.1	93.9
McCrory 2000	Dizziness	15	8	NM	NM	**65.2**	42.7	83.6	**-**	-	-
	Nausea	5	18	NM	NM	**21.7**	7.5	43.7	**-**	-	-
	Headache	23	0	NM	NM	**100.0**	85.2	100..0	**-**	-	-
McCrea 2005*	GSC	84	10	0	56	**89.4**	81.3	94.8	**100.0**	93.6	100.0
Erlanger 2003	Dizziness	40	7	-	-	**85.1**	71.7	93.8	**-**	-	-
	Nausea	25	22	-	-	**53.2**	38.1	67.9	**-**	-	-
	Headache	44	3	-	-	**93.6**	82.5	98.7	**-**	-	-
Fuller 2014	Symptom Checklist	50	15	23	77	**76.9**	64.8	86.5	**77.0**	67.5	84.8
	Mental status evaluation	30	25	5	95	**54.5**	40.6	68.0	**95.0**	88.7	98.4
Putukian 2015†	SCAT2 symptom checklist – number	27	5	0	23	**84.4**	67.2	94.7	**100.0**	85.2	100.0
	SCAT2 symptom checklist – severity	24	8	0	23	**80.0**	61.4	92.3	**100.0**	85.2	100.0

* McCrea 2005 (i) Numbers for TP, FN, FP, TN, and 95% confidence intervals calculated from reported sensitivity and specificity estimates derived from standardized regression based indices for detection of significant change in test scores. † Numbers for TP, FN, FP, TN, and 95% confidence intervals calculated from reported diagnostic accuracy data for impairment in symptom number and severity <5th centile of normative performative. ** A range of symptoms studied, representative results for common symptoms presented.
TP: True positives; FN: False negatives; FP: False positives; TN: True negatives; LCI: Lower confidence interval; UCI: Upper confidence interval.

The presence of individual symptoms in concussed and non-concussed athletes was investigated by Maddocks 1995, McCrory 2000 and Erlanger 2003. Headache was a sensitive indicator of concussion with point estimates reported between 92.9% and 100.0%. Nausea and dizziness were less sensitive, but more specific (92.9% to 96.4% respectively). Diagnostic accuracy results for symptoms checklists were imprecise and heterogeneous. McCrea 2005 (GCS) and Putukian 2015 (SCAT2 symptom checklist) reported moderate sensitivity of 89.4% and 84.4% respectively for the presence of any symptoms, with excellent specificities of 100%. However, these results were not replicated in Fuller 2014 (PSCA symptom checklist) where sensitivity and specificity of 76.9% and 77.0% were reported. Clinical signs of abnormal mentation were found to be specific (95.0%), but not sensitive (54.5%) for concussion.

Cognition

Study	Index test	TP	FN	FP	TN	Sensitivity (%)	LCL	UCL	Specificity (%)	LCL	UCL
Orientation											
Maddocks 1995*	Orientation	6	22	2	26	**21.4**	8.3	41.0	**92.9**	76.5	99.1
Maddock's Questions											
Maddocks 1995*	Recent memory	21	7	4	24	**75.0**	55.1	89.3	**85.7**	67.3	96.0
Fuller 2014	Maddock's Questions	22	43	7	93	**33.8**	22.6	46.6	**93.0**	86.1	97.1
Standardised Assessment of Concussion											
Barr 2001**	SAC	47	3	16	52	**94.0**	83.5	98.7	**76.5**	64.6	85.9
McCrea 2001**	SAC	60	3	13	42	**95.2**	86.7	99.0	**76.4**	63.0	86.8
McCrea 2002†	SAC	68	23	NM	NM	**79.1**	69.3	86.9	**-**	-	-
McCrea 2005ǂ	SAC	75	19	5	51	**79.8**	70.2	87.4	**91.1**	80.4	97.0
Echlin 2010**	SAC	7	6	NM	NM	**53.8**	25.1	80.8			
Marindes 2015§	SAC	15	14	NM	NM	**55.6**	35.3	74.5	**-**	-	-
Galetta K 2015§	SAC	2	8	3	14	**20.0**	2.5	55.6	**82.4**	56.6	96.2
Putukian 2015***	SAC	13	19	2	20	**40.6**	23.7	59.4	**90.9**	70.8	98.9

* Diagnostic accuracy reported separately for a range of orientation and recent memory questions. Representative data for 'What month is it?' and 'How far in the quarter?' presented.

** Sensitivity and specificity presented for ≥1 point drop in SAC compared to baseline

† Sensitivity calculated for SAC score below 10th percentile of normal performance

ǂ Numbers for TP, FN, FP, TN, and 95% confidence intervals calculated from reported sensitivity and specificity estimates derived from standardized regression based indices for detection of significant change in test scores.
§ Sensitivity and specificity presented for ≥2 point drop in SAC compared to baseline
*** Numbers for TP, FN, FP, TN, and 95% confidence intervals calculated from reported diagnostic accuracy data for impairment in symptom number and severity <5th centile of normative performative.

TP: True positives; FN: False negatives; FP: False positives; TN: True negatives; LCI: Lower confidence interval; UCI: Upper confidence interval.

Diagnostic accuracy for orientation questions was available from Maddocks 1995, reporting a range of low and imprecise estimates for sensitivity between 3.6% and 57.1%, and 73.1% and 100% for specificity. Maddocks also provided estimates for individual sports-related recent memory questions ('Maddock's Questions) with sensitivity varying from 34.1% to 75.0%, and specificity of 85.7% to 100.0%. Fuller reported a contrasting sensitivity of 33.8% (95% CI 22.6 – 46.6) and specificity of 93.0% (95% CI 86.1 to 97.1) for all Maddock's Questions taken together. Studies examining the SAC used a wide variety of cut-points for positivity including a ≥1 or ≥2 drop in baseline score, regression based indices for detection of significant change in test scores, or scores <5^{th} or 10^{th} percentile of normal performance. Unsurprisingly, accuracy results varied widely , with lowest estimates for sensitivity and specificity of 20.0% and 76.4%, and highest estimates of 95.1% and 91.1% respectively (I^2 90.1%).

Balance

Study	Index test	TP	FN	FP	TN	Sensitivity (%)	LCL	UCL	Specificity (%)	LCL	UCL
McCrea 2005 (i)*	BESS	34	60	3	53	**36.0**	26.5	46.7	**94.6**	85.1	98.9
McCrea 2005 (ii)**	BESS					**34.0**	NR	NR	**91.0**	NR	NR
Echlin 2010§	BESS	4	1	-	-	**80.0**	28.4	99.5	**-**	-	-
Fuller 2014***	Tandem Stance	18	47	5	95	**27.7**	17.3	40.2	**95.0**	88.7	98.4
Putukian 2015†	Modified BESS	8	24	0	23	**25.0**	11.5	43.5	**100**	85.2	100.0
Marindes 2015ǂ	BESS	16	4	NM	NM	**80.0**	56.3	94.3	**-**	-	-
Galetta K 2015§	Timed Tandem Gait	10	2	5	9	**83.3**	51.6	97.9	**64.3**	35.1	87.2

* McCrea 2005 (i) Numbers for TP, FN, FP, TN, and 95% confidence intervals calculated from reported raw data for any impairment of BESS from baseline. ** McCrea 2005 (ii) Point estimates for sensitivity and specificity from standardized regression based indices for detection of significant change in test scores. ***>4 errors in 20 seconds. † Numbers for TP, FN, FP, TN, and 95% confidence intervals calculated from reported diagnostic accuracy data for impairment of modified BESS <5th centile of normative performative. ǂ≥3 point worsening in BESS. §Any worsening from baseline.
TP: True positives; FN: False negatives; FP: False positives; TN: True negatives; LCI: Lower confidence interval; UCI: Upper confidence interval.

Individual sensitivity estimates for the BESS were heterogenous and imprecise, with point estimates ranging from 34.0 to 80.0%, I^2 87.4%. BESS specificity, reported in a single study, was high 94.6% (95% CI 85.1 – 98.9). A range of accuracy results were calculated for the modified BESS by Putukian 2015 based on reliable change indices and comparison to normative performance. A representative sensitivity of 25.0% (95% CI 11.5 – 43.4) and specificity of 100.0% (95% CI 85.2 to 100.0) was reported for performance compared to normative values below the 5th percentile. The Tandem Stance Test demonstrated poor sensitivity (27.7%, 95% CI 17.3 – 40.2) and good specificity (95.0%, 95% CI 88.7 – 98.4) in the single study available. The Timed Tandem Gait demonstrated moderate sensitivity and specificity of 83.3% (95% CI 51.6-97.9) and 64.3% (95% CI 35.1-87.2) respectively.

Oculomotor

Study	TP	FN	FP	TN	Sensitivity (%)	LCL	UCL	Specificity (%)	LCL	UCL
Galetta K 2011	5	0	2	0	**100.0**	47.8	100.0	**0.0**	0.0	84.2
Galetta K 2011b	9	1	-	-	**90.0**	55.5	99.7	**-**	-	-
King 2012*	3	0	0	0	**100.0**	29.2	100.0	**-**	-	-
Galetta M 2013	2	0	-	-	**100.0**	15.8	100.0	**-**	-	-
Dhawan 2014	20	0	11	110	**100.0**	83.2	100	**90.9**	84.3	95.4
Leong 2014†	1	0	0	5	**100.0**	2.5	100.0	**100.0**	47.8	100.0
Galetta K 2015	9	3	1	13	**75.0**	42.8	94.8	**92.9**	66.1	100.0
Leong 2015†	8	1	2	0	**88.9**	51.8	99.7	**0.0**	0.0	84.2
Marinides 2015	23	6	NM	NM	**79.3**	60.3	92.0	**-**	-	-
Seidman 2015	9	0	0	328	**100.0**	66.4	100.0	**100.0**	98.9	100.0

-: No data available to allow calculation
* Data for witnessed head impact events undergoing side-line testing used only.
† Results reconstructed from side-line SCAT2 reference standard, not original case control study as per protocol.
TP: True positives; FN: False negatives; FP: False positives; TN: True negatives; LCI: Lower confidence interval; UCI: Upper confidence interval.

Data allowing calculation of sensitivity of the post-head impact event KD time for side-line identification of concussion was measured in all included studies and varied widely from 71.4% (Galetta K 2011) to 100.0% (King 2012, Galetta M 2013, Dhawan 2014, Leong 2014, King 2015, Seidman 2015). Individual estimates were very imprecise secondary to small sample sizes, with lower 95% confidence limits as low as 2.5% calculated (Leong 2014). This diversity was reflected in a high I2 statistic (52.1%). Data for specificity estimates was measured in six studies with similarly imprecise and heterogeneous results calculated, ranging from 0.0% (Leong 2015) to 100.0% (Leong 2014, Seidman 2015), I2 statistic 89.3%. KD test errors were reported in five studies (Galetta K 2011, Galetta K 2011b, Leong 2014, Leong 2015, Seidman 2015) and were found to be infrequent as shown in Table 5. Errors in isolation appeared to be specific, but non-sensitive, for the identification of concussion. However, results were very heterogeneous and imprecise with sensitivity point estimates ranging from 9.1 to 100.0%. 95% confidence limits for specificity varied from 47.8 to 100.0%. Insufficient data was reported to allow assessment of the diagnostic accuracy of both prolonged KD test times and errors in combination

Multimodal

Study	TP	FN	FP	TN	Sensitivity (%)	LCL	UCL	Specificity (%)	LCL	UCL
Sports Concussion Assessment Tool 2										
Galetta M 2013*	2	0	0	0	**100.0**	15.8	100.0	**-**	-	-
Putukian 2015†	25	7	1	22	**78.1**	60.0	90.7	**95.7**	78.1	99.9
Pitchside Concussion Assessment Tool										
Fuller 2014	55	10	26	74	**84.6%**	73.5	92.4	**74.0**	64.3	82.3
Sports Concussion Assessment Tool 2, King-Devick Test*										
Galetta M 2013	2	0	0	0	**100.0**	15.8	100.0	**-**	-	-
Timed Tandem Gait, Standardised Assessment of Concussion, King-Devick Test*										
Galetta K 2015	24	0	NR	NR	**100.0**	85.8	100.0	**-**	-	-
Balance Error Scoring System, Standardised Assessment of Concussion, King-Devick Test**										
Marinides 2015	20	0	NM	NM	**100.0**	83.2	100	**-**	-	-
Graded Symptom Checklist, Balance Error Scoring System, Standardised Assessment of Concussion										
McCrea 2005*	89	5	6	49	**94.7**	88.0	98.3	**89.1**	77.8	95.9

* Any worsening from baseline in any sub-test.

† Numbers for TP, FN, FP, TN, and 95% confidence intervals calculated from reported diagnostic accuracy data for impairment in symptom number and severity <15th centile of normative performative.

** From baseline: any increase in KD test, ≥2 points worsening on SAC, ≥3 points worsening on BESS

TP: True positives; FN: False negatives; FP: False positives; TN: True negatives; LCI: Lower confidence interval; UCI: Upper confidence interval.

Point estimates for the sensitivity of combined use of individual sideline screening tools were high, but imprecise, reaching 100% for combinations of SCAT2/KD, TTG/SAC/KD, and BESS/SAC/KD; and 94.7% for joint use of GCS/BESS/SAC. The specificity of joint use of individual screening tests was available for a single study (McCrea 2005, GCS/BESS/SAC), at 89.1% (95% CI 77.8-95.9). The diagnostic accuracy of multifaceted sideline screening tests appeared lower, with sensitivity and specificity of 78.1% and 95.7%, and 84.6% and 74.0% reported for the SCAT2 and PSCA instruments respectively.

Video analysis and integrated head injury assessment protocol

Characteristics of Fuller 2016

Study	Setting	Design	Sample Size (n=)	Sport(s)	Level	Mean age (years±SE)	Technology	Risk of Bias / evidence level	Applicability concerns	Primary finding(s)
Fuller 2016	UK	PCS	49	Rugby Union	Professional	26.5 (SD 3.5)	Sideline video review	Level 2b	Low	•Contributed to identification of 61.% of significant head impact events •21% of all diagnosed concussions presented post game

Detailed risk of bias assessments

Symptoms

Study	Risk of bias					Applicability concerns			
	Patient selection	**Index test**	**Reference standard**	**Flow and timing**	**Overall**	**Patient selection**	**Index test**	**Reference standard**	**Overall**
Maddocks 1995	High Case-control design	Low	Low	Low	**High**	Low	Low	Low	**Low**
McCrory 2000	High Case-control design	Unclear Test review bias?	Low	Low	**High**	Low	Low	Low	**Low**
McCrea 2005	High Case-control design	High Test review bias	High Non-physician assessment	Low	**High**	Low	Low	Low	**Low**
Erlanger 2003	High Case-control design	Unclear Test review bias?	Unclear Diagnostic review bias? Non-physician assessment?	Low	**High**	Low	Low	Low	**Low**
Fuller 2014	Low	Low	High Diagnostic review bias	Low	**High**	Low	Low	Low	**Low**
Putukian 2015	High Case-control design	High Test review bias	Low	High Delayed index test	**High**	Low	Low	Low	**Low**

Cognition

Study	Risk of bias					Applicability concerns			
	Patient selection	**Index test**	**Reference standard**	**Flow and timing**	**Overall**	**Patient selection**	**Index test**	**Reference standard**	**Overall**
Maddocks 1995	High Case-control design	Low	Low	Low	**High**	Low	Low	Low	**Low**
Barr 2001	High Case-control design	Unclear Test review bias?	High Non-physician assessment	Low	**High**	Low	Low	Low	**Low**
McCrea 2001	High Case-control design	Unclear Test review bias?	High Non-physician assessment	Low	**High**	Low	Low	Low	**Low**
McCrea 2002	High Case-control design	Unclear Test review bias?	High Non-physician assessment	Low	**High**	Low	Low	Low	**Low**
McCrea 2005	High Case-control design	High Test review bias	High Non-physician assessment	Low	**High**	Low	Low	Low	**Low**
Echlin 2010	High Case-control design	High Test review bias	High Incorporation bias	High Very high missing data levels	**High**	Low	Low	Low	**Low**
Galetta M 2013	High Case-control design	Unclear Test review bias?	Unclear Diagnostic review bias? Non-physician assessment?	Low	**High**	Low	Low	Low	**Low**
Fuller 2014	Low	Low	High Diagnostic review bias	Low	**High**	Low	Low	Low	**Low**
Marinides 2015	High Case-control design	Unclear Test review bias?	Unclear Diagnostic review bias?	High Delayed index test	**High**	Low	Low	Low	**Low**

			Non-physician assessment?						
Putukian 2015	High	High	Low	High	**High**	Low	Low	Low	**Low**
	Case-control design	Test review bias		Delayed index test					
Galetta K 2015	High	Unclear	Unclear	Low	**High**	Low	Low	Low	**Low**
	Case-control design	Test review bias?	Timing of reference standard?						

Balance

Study	Risk of bias					Applicability concerns			
	Patient selection	**Index test**	**Reference standard**	**Flow and timing**	**Overall**	**Patient selection**	**Index test**	**Reference standard**	**Overall**
McCrea 2005	High Case-control design	High Test review bias	High Non-physician assessment	Low	**High**	Low	Low	Low	**Low**
Echlin 2010	High Case-control design	High Test review bias	High Incorporation bias	High Very high missing data levels	**High**	Low	Low	Low	**Low**
Fuller 2014	Low	Low	High Diagnostic review bias	Low	**High**	Low	Low	Low	**Low**
Galetta K 2015	High Case-control design	Low	Unclear Diagnostic review bias? Timing of reference standard?	Low	**High**	Low	Low	Low	**Low**
Marinides 2015	High Case-control design	Unclear Test review bias?	Unclear Diagnostic review bias? Non-physician assessment?	High Delayed index test	**High**	Low	Low	Low	**Low**
Putukian 2015	High Case-control design	High Test review bias	Low	High Delayed index test	**High**	Low	Low	Low	**Low**

Oculomotor

Study	Risk of bias					Applicability concerns			
	Patient selection	**Index test**	**Reference standard**	**Flow and timing**	**Overall**	**Patient selection**	**Index test**	**Reference standard**	**Overall**
Galetta K 2011*	Low	Unclear Diagnostic review bias?	Unclear Test review bias?	High Delayed index test	**High**	Low	Low	Low	**Low**
Galetta K 2011b	High Case-control design	Low	High Non-physician assessment Test review bias?	Low	**High**	Low	Low	Low	**Low**
King 2012	Low	Unclear Diagnostic review bias?	Unclear Test review bias?	Unclear Timing of index test?	**Unclear**	Low	Low	Low	**Low**
Galetta M 2013	High Case-control design	Unclear Diagnostic review bias?	Unclear Test review bias? Non-physician assessment?	Low	**High**	Low	Low	Low	**Low**
Dhawan 2014	High Case-control design	Unclear Diagnostic review bias?	Unclear Test review bias? Non-physician assessment? Accurate reference standard?	Low	**High**	Unclear Sample not described	Low	Unclear Reference standard not described	**Unclear**
Leong 2014	Low	Low	Low	High Delayed index test	**High**	Low	Low	Low	**Low**
Galetta K 2015	High Case-control design	Low	Unclear Test review bias? Timing of reference standard?	Low	**High**	Low	Low	Low	**Low**

Leong 2015	Low	Low	High Test review bias Non-physician assessment?	Low	**High**	Low	Low	Low	**Low**
Marinides 2015	High Case-control design	Unclear Diagnostic review bias?	Unclear Test review bias? Non-physician assessment?	Low	**High**	Low	Low	Low	**Low**
Seidman 2015	High Case control design	Low	Low	High Delayed index test	**High**	Low	Low	Low	**Low**

Multimodal

Study	Risk of bias					Applicability concerns			
	Patient selection	**Index test**	**Reference standard**	**Flow and timing**	**Overall**	**Patient selection**	**Index test**	**Reference standard**	**Overall**
McCrea 2005	High	High	High	Low	**High**	Low	Low	Low	**Low**
	Case-control design	Test review bias	Non-physician assessment						
Galetta M 2013	High	Unclear	Unclear	Low	**High**	Low	Low	Low	**Low**
	Case-control design	Test review bias?	Diagnostic review bias? Non-physician assessment?						
Fuller 2014	Low	Low	High	Low	**High**	Low	Low	Low	**Low**
			Diagnostic review bias						
Putukian 2015	High	High	Low	High	**High**	Low	Low	Low	**Low**
	Case-control design	Test review bias		Delayed index test					
Marinides 2015	High	Unclear	Unclear	High	**High**	Low	Low	Low	**Low**
	Case-control design	Test review bias?	Diagnostic review bias? Non-physician assessment?	Delayed index test					
Galetta K 2015	High	Unclear	Unclear	Low	**High**	Low	Low	Low	**Low**
	Case-control design	Test review bias?	Timing of reference standard?						

Technology

Study	Risk of bias					Applicability concerns			
	Patient selection	**Index test**	**Reference standard**	**Flow and timing**	**Overall**	**Patient selection**	**Index test**	**Reference standard**	**Overall**
Guskiewicz 2007	Low	Low	Low	Low	**Low**	Low	Low	Low	**Low**
Mihalak 2007	Low	Low	Unclear Diagnostic review bias? Non-physician assessment?	Low	**Unclear**	Low	Low	Low	**Low**
Greenwald 2008	Low	Low	Low	Low	**Low**	Low	Low	Low	**Low**
Broglio 2010	Low	Low	Low	Low	**Low**	Low	Low	Low	**Low**

Video and integrated head injury assessment protocols

Study	Patient selection	Comparability	Outcome	Overall
Fuller 2016	Low •Census sample •Comprehensive identification of head impact events •Healthy athletes at start of study •No attrition	Not applicable •Not comparative effectiveness/diagnostic accuracy/aetiological study	Low •Comprehensive outcome assessment •Follow up beyond acute period	**Low**

Detailed quality of evidence assessments

These table summarise the strength of evidence for sensitivity and specificity estimates in each sub-topic domain according to GRADE criteria.

Symptoms

Outcome	Study designs	Factors decreasing quality of evidence					Overall GRADE rating
		Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	
Graded Symptom Scale							
Sensitivity	1 PCS	Serious concerns	No concerns	Unknown (single study)	Some concerns	Not detected	**Very Low**
Specificity	1 PCS	Serious concerns	No concerns	Unknown (single study)	No Concerns	Not detected	**Low**
Individual Symptoms							
Sensitivity	3 PCS	Serious concerns	No concerns	Serious concerns	Serious concerns	Not detected	**Very Low**
Mental Status Evaluation							
Sensitivity	1 PCS	Some concerns	No concerns	Unknown (single study)	Serious concerns	Not detected	**Low**
Specificity	1 PCS	Some concerns	No concerns	Unknown (single study)	Some concerns	Not detected	**Low**
PSCA symptom checklist							
Sensitivity	1 PCS	Some concerns	No concerns	Unknown (single study)	Serious concerns	Not detected	**Low**
Specificity	1 PCS	Some concerns	No concerns	Unknown (single study)	Some concerns	Not detected	**Low**
SCAT2 Symptom Checklist							
Sensitivity	1 PCS	Serious concerns	No concerns	Unknown (single study)	Serious concerns	Not detected	**Very Low**
Specificity	1 PCS	Serious concerns	No concerns	Unknown (single study)	Some concerns	Not detected	**Low**

Cognition

Outcome	Study designs	Factors decreasing quality of evidence					Overall GRADE rating
		Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	
Orientation Questions							
Sensitivity	1 PCS	Serious concerns	No concerns	Unknown (single study)	Serious concerns	Not detected	**Very Low**
Specificity	1 PCS	Serious concerns	No concerns	Unknown (single study)	Serious concerns	Not detected	**Very Low**
Maddock's Questions							
Sensitivity	2 PCS	Serious concerns	No concerns	Serious concerns	Serious concerns	Not detected	**Very Low**
Specificity	2 PCS	Serious concerns	No concerns	Serious concerns	Serious concerns	Not detected	**Very Low**
Standardised Assessment of Concussion							
Sensitivity	6 PCS 1 RCS	Serious concerns	No concerns	Serious concerns	Serious concerns	Not detected	**Very Low**
Specificity	5 PCS	Serious concerns	No concerns	Serious concerns	Serious concerns	Not detected	**Very Low**

Oculomotor

Outcome	Study designs	Factors decreasing quality of evidence					Overall GRADE rating
		Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	
King-Devick Test							
Sensitivity	10 PCS 1 RCS	Serious concerns	No concerns	Serious concerns	Serious concerns	Not detected	**Very Low**
Specificity	6 PCS	Serious concerns	No concerns	Serious concerns	Serious concerns	Not detected	**Very Low**

Balance

Outcome	Study designs	Factors decreasing quality of evidence					Overall GRADE rating
		Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	
Balance Error Scoring System							
Sensitivity	2 PCS 1RCS	Serious concerns	No concerns	Serious concerns	Serious concerns	Not detected	**Very Low**
Specificity	1 PCS	Serious concerns	No concerns	Unknown (single study)	Some concerns	Not detected	**Low**
Tandem Stance Test							
Sensitivity	1 PCS	Some concerns	No concerns	Unknown (single study)	Serious concerns	Not detected	**Low**
Specificity	1 PCS	Some concerns	No concerns	Unknown (single study)	Some concerns	Not detected	**Low**
Modified BESS							
Sensitivity	1 PCS	Serious concerns	No concerns	Unknown (single study)	Serious concerns	Not detected	**Very Low**
Specificity	1 PCS	Serious concerns	No concerns	Unknown (single study)	Some concerns	Not detected	**Low**
Timed Tandem Gait							
Sensitivity	1 PCS	Serious concerns	No concerns	Unknown (single study)	Serious concerns	Not detected	**Very Low**
Specificity	1 PCS	Serious concerns	No concerns	Unknown (single study)	Serious concerns	Not detected	**Very Low**

Multimodal tests

Outcome	Study designs	Factors decreasing quality of evidence					Overall GRADE rating
		Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	
Sports Concussion Assessment Tool 2							
Sensitivity	2 PCS	Serious concerns	No concerns	Serious concerns	Serious concerns	Not detected	**Very Low**
Specificity	2 PCS	Serious concerns	No concerns	Unknown (single study)	Serious concerns	Not detected	**Very Low**
Pitchside Concussion Assessment Tool							
Sensitivity	1 PCS	Serious concerns	No concerns	Unknown (single study)	Some concerns	Not detected	**Low**
Specificity	1 PCS	Serious concerns	No concerns	Unknown (single study)	Some concerns	Not detected	**Low**
Sports Concussion Assessment Tool 2, King-Devick Test*							
Sensitivity	1 PCS	Serious concerns	No concerns	Unknown (single study)	Serious concerns	Not detected	**Very Low**
Timed Tandem Gait, Standardised Assessment of Concussion, King-Devick Test*							
Sensitivity	1 PCS	Serious concerns	No concerns	Unknown (single study)	Some concerns	Not detected	**Low**
Balance Error Scoring System, Standardised Assessment of Concussion, King-Devick Test**							
Sensitivity	1 RCS	Serious concerns	No concerns	Unknown (single study)	Serious concerns	Not detected	**Very Low**
Graded Symptom Checklist, Balance Error Scoring System, Standardised Assessment of Concussion							
Sensitivity	1 PCS	Serious concerns	No concerns	Unknown (single study)	Some concerns	Not detected	**Very Low**
Specificity	1 PCS	Serious concerns	No concerns	Unknown (single study)	No concerns	Not detected	**Low**

Technology

Outcome	Study designs	Factors decreasing quality of evidence					Overall GRADE rating
		Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	
Head Impact Telemetry System							
Positive predictive value	4 PCS	No concerns	No concerns	No concerns	Unknown (not reported)	Not detected	**Moderate**
Side-line video review							
Identification of significant head impact events	1 PCS	No concerns	No concerns	Unknown (single study)	Some concerns (small sample size)	Not detected	**Low**

Integrated head injury assessment protocol

Outcome	Study designs	Factors decreasing quality of evidence					Overall GRADE rating
		Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	
Identification of significant head impact events and concussion	1 PCS	No concerns	No concerns	Unknown (single study)	Some concerns (small sample size)	Not detected	**Low**

Summary of the sideline head injury assessment protocols used in professional contact and collision sports

Sporting body	Tool / protocol	Person/s who can request sideline screening	Person/s conducting the assessment	Use of video review	Location /duration of testing	Other key components
AFL/ NRL	Sport-specific HIA Form	Team doctor	Team doctor	Mandatory	Off-field Minimum of 15 mins	Other club support staff must report observations to the team doctor. SCAT3 used for further assessment. HIA forms are collected for audit and injury surveillance purposes.
FIFA	Immediate removal criteria				On-field/pitchside	3-minute injury time following head impact. Pitch-Side assessment performed (based on a number of immediate removal criteria)
IIHF	Concussion protocol		Team doctor and/or AT (Team doctor solely responsible for determining concussion diagnosis)		Off-pitch	Observations made by team medical staff (or by any other team personnel and passed on to team medical staff).
NFL	Side-line concussion assessment tool	Coach, player, teammate, official, team doctor, AT, AT in the media booth or UNC	Team doctor, ATC or UNC	Mandatory	Off-pitch	Booth ATC, UNC, officials and the team doctor are connected by radio communication. The team doctor will review the video of the incident and (at a minimum) assess the player with a focussed neurological assessment (asking what happened, reviewing the "Go/No Go" signs and symptoms; and asking the Maddock's questions. If the diagnosis is unclear, the player will undergo a full NFL sideline Concussion Assessment in the team locker room.
World Rugby	HIA process	Match official, team doctor or independent match day doctor	Certified medical professional	Mandatory	Off-pitch 10 minutes	Mandatory online education program for relevant personnel. Where the diagnosis is not immediately apparent, players removed & assessed. HIA forms are collected for audit & research

AFL = Australian Football League; FIFA = Federation Internationale de Football Association; HIA + Head Injury Assessment; IIHF = International Ice Hockey Federation; NFL = National Football League; NRL = National Rugby League. AT=Athletic trainer. UNC= unaffiliated neurotrauma consultant. HIA= Head Injury Assessment

Summary of criteria for immediate removal from play or for further assessment used in professional sport.

Clinical criteria	AFL/ NRL	FIFA	IIHF	NFL	World Rugby
Confirmed loss of consciousness					
Definite confusion/disorientation					
Any balance disturbance (e.g. ataxia) or motor incoordination					
Impact seizure/convulsions or tonic posturing					
Player reports significant, new or progressive/persistent concussion symptoms					
Clearly dazed, “dinged”, blank or vacant stare					
Behavioural change atypical of the player					
Any clinical impression that the player is not quite right following trauma (i.e. “physician’s decision”)					
Loss of responsiveness/suspected loss of consciousness					
Memory impairment/amnesia					
No protective action when falling to the ground (can be either tonic or hypotonic) – observed on video					
Dangerous mechanism of trauma					
Cross eyes (strabismus) or spontaneous nystagmus					
Possible impact seizure or tonic posturing on video review					
Possible balance disturbance					
Slow to get up following a hit to the head					
Possible behavioural changes					
Possible confusion					
Head impact event with the potential to result in concussion					
Diagnosis not apparent					

AFL = Australian Football League; IIHF = International Ice Hockey Federation; NRL = National Rugby League; NFL = National Football League, FIFA = Federation Internationale de Football Association

= Criteria for immediate removal and no return (i.e. diagnosis of concussion)

= Criteria for further assessment

= Criteria not specified

GLOSSARY OF METHODOLOGICAL TERMS

Term	Definition	Ref
Grey Literature	Grey literature (or gray literature) are materials and research produced by organizations outside of the traditional commercial or academic publishing and distribution channels e.g. websites, conference proceedings, PhD theses, etc.	
Current awareness search	Literature searches conducted after the initial manuscript draft and just prior to submission to keep up-to-date with the most recently published information and developments.	
Forest plots	A graphical representation of the individual results of each study included in systematic review, presenting point estimates of effect estimates/diagnostic accuracy metrics (represented as squares) together with their precision (95% confidence intervals, represented as lines). The forest plot provides a quick visual representation of overall effect estimates, how certain these results are, and heterogeneity in results across studies.	
Imprecision	Imprecision is a measure of statistical variability. It is typically quantified by a confidence interval providing an estimated range of values which is likely to include the unknown population parameter in question, estimated from a given set of sample data. The width of the confidence interval indicates how uncertain we are about the unknown parameter. A very wide interval may indicate that more data should be collected before anything very definite can be said about the parameter.	
Heterogeneity	Statistical variability of results among studies included in a systematic review is termed heterogeneity. This may occur due to : • Variability in the participants, interventions and outcomes studied, described as clinical diversity or clinical heterogeneity. • Variability in study design and risk of bias, described as methodological diversity or methodological heterogeneity. Statistical heterogeneity manifests itself as the observed intervention results being more different from each other than one would expect due to random error (chance) alone.	
I^2 statistic	A useful statistic for quantifying inconsistency across studies included in a systematic review.The importance of the observed value of I2 depends on (i) magnitude and direction of effects and (ii) strength of evidence for heterogeneity e.g. a confidence interval for I^2.A rough guide to interpretation is as follows: I^2 0% to 40%: might not be important; I^2 30% to 60%: may represent moderate heterogeneity I^2 50% to 90%: may represent substantial heterogeneity I^2 75% to 100%: considerable heterogeneity	
Meta-analysis	A statistical analysis that combines the results of multiple scientific studies into a single weighted average.	
Narrative synthesis	The results of studies included in a systematic review are summarised, described, explained and interpreted qualitatively using words and text.	
Test review bias	Test review bias may be present when the results of the reference standard are known to those interpreting the index test. Results in overestimation of sensitivity. .	
Diagnostic review bias	Diagnostic review bias may be present when the results of the index test are known to those interpreting the reference standard. Results in overestimation of sensitivity and	

	specificity.	
Incorporation bias	Systematic error in calculated diagnostic accuracy metrics occurring when the result of the index test is used in establishing the final diagnosis (i.e. it forms part of the reference standard). Results in overestimation of sensitivity and specificity.	
Attrition bias	Non-random loss to follow up or withdrawal from the study can result in a non-representative sample and biased results if the withdrawal rate depends on the results of the index test or reference standard.	
Delayed index testing bias	A systematic error in diagnostic accuracy results arising from conducting the index test later than would be expected in practice (e.g. performing 'sideline' screening rests for concussion after completion of sporting participation). Could result in different estimates of diagnostic performance due to disease progression (e.g. transient concussions could have resolved).	
Inaccurate reference standard assessment	The error in diagnoses derived from an imperfect reference standard can result in underestimation of the performance of the index test.	